ШКОЛЬНЫЙ КУРС МАТЕМАТИКИ

Я. С. Бродский

СТАТИСТИКА. ВЕРОЯТНОСТЬ. КОМБИНАТОРИКА

Москва
ОНИКС • Мир и Образование

УДК 519.2(075.3)
ББК 22.17я72
Б88

Бродский Я. С.

Б88 Статистика. Вероятность. Комбинаторика / Я. С. Бродский. — М.: ООО «Издательство Оникс»: ООО «Издательство «Мир и Образование», 2008. — 544 с.: ил. — (Школьный курс математики).

ISBN 978-5-488-01369-8 (ООО «Издательство Оникс»)
ISBN 978-5-94666-423-3 (ООО «Издательство «Мир и Образование»)

В данном учебном пособии подробно излагаются основы описательной и математической статистики, элементы теории вероятностей и комбинаторики. К каждому параграфу приводятся контрольные вопросы и задачи для самостоятельного решения. Кроме того, каждая глава содержит дополнительные задачи. В конце книги даны ответы и указания ко всем задачам.

Пособие предназначено старшеклассникам, студентам техникумов и младших курсов вузов, обучающихся на не математических специальностях.

УДК 519.2(075.3)
ББК 22.17я72

ISBN 978-5-488-01369-8(ООО «Издательство Оникс»)
ISBN 978-5-94666-423-3(ООО «Издательство «Мир и Образование»)

Оглавление

Предисловие

Книга адресована широкому кругу читателей — старшеклассникам, студентам техникумов и младших курсов вузов, обучающихся на не математических специальностях.

Для старшеклассников книга представляет интерес, так как вероятностно-статистическая, или, как ее еще называют, *стохастическая линия*, сравнительно недавно стала полноправной составляющей школьной программы по математике. Учащиеся получают первые представления об элементах статистики, вероятности и комбинаторики. Заметим, что решения, которые принимаются государственными органами в образовании, медицине, технике, бизнесе и других отраслях, зависят от анализа имеющихся или собранных данных. Поэтому анализ данных должен стать для старшеклассников предметом более глубокого изучения в школе. Статистическая культура человека является неотъемлемой составляющей его общей культуры. Каждый человек будет тем успешнее в жизни, чем полнее и глубже будет понимать статистическую природу окружающего мира и его законы. Надеемся, что данная книга поможет учащимся старших классов лучше ориентироваться в таких областях деятельности, как экономика, страхование, инвестиции, подверженных влиянию случая.

Комбинаторика, вероятность и статистика вошли также в программу по математике большинства техникумов. Студентам техникумов, безусловно, будет интересен современный взгляд на эти разделы математики и многие приложения вероятностно-статистических методов, о которых идет речь в книге.

Студенты многих высших учебных заведений изучают курс теории вероятностей и математической статистики. Для чего же им адресовать данную книгу? Дело в том, что изучение в вузе аналитической геометрии или математического анализа подготовлено школьным курсом математики: из геометрии они знакомы с векторами и методом координат, а из курса алгебры и начал анализа —

с функциями, производной и интегралом. Такая пропедевтическая подготовка способствует более сознательному усвоению разделов высшей математики, в том числе в области стохастики.

Пособие может быть использовано учителями общеобразовательных школ в классах с повышенной математической подготовкой, а также для проведения факультативных занятий в общеобразовательных старших классах, в классах экономического и других профилей.

Изложение учебного материала отличается от принятого в традиционных вузовских учебниках. Во-первых, в пособии допускаются различные уровни обоснования:

— уровень здравого смысла (на это направлены многочисленные примеры и сравнения);

— «прикладной» уровень;

— и конечно, формально-логический уровень.

Во-вторых, сделана попытка избежать рецептурного стиля, особенно при изложении элементов математической статистики. Главная цель, которую автор ставил перед собой, — довести до читателей основные идеи статистики, а не рецепты их применения. Поэтому идеи математической статистики реализуются при помощи сравнительно небольшого математического аппарата.

При написании книги автор придерживался следующих принципов:

— внедрение идеи математического моделирования и формирование навыков применения вероятности и статистики к описанию реальных процессов и явлений;

— формирование навыков предварительной обработки статистического материала с применением современной вычислительной техники;

— применение различных подходов к решению комбинаторных задач;

— использование различных определений понятия вероятности, позволяющих иметь различные источники введения вероятности события.

Книга состоит из введения, пяти глав, приложения, предметного и именного указателей. Первая глава посвящена описательной статистике. Наряду с обучением предварительной обработке экспериментальных данных она ознакомит читателя с основными идеями статистики — выборочным методом, оценкой параметров, проверкой статистических гипотез. Автор стремился к тому, чтобы при изучении этой главы читатель пришел к выводу о случайном характере многих явлений и к необходимости измерять эту случайность.

Во второй главе начинается систематическое изложение основ теории вероятностей. Здесь рассматриваются различные определения вероятности, основные свойства вероятности события, ее применение. В этой главе читатель подводится к необходимости иметь в своем распоряжении специальные приемы для вычисления вероятностей. Речь идет о комбинаторике, изложение которой и содержится в третьей главе. В ней представлены три различных подхода к решению комбинаторных задач.

В четвертой главе продолжается изучение элементов теории вероятностей, а именно случайных величин и их числовых характеристик.

Завершается книга пятой главой, посвященной элементам математической статистики.

Изложение теоретического материала сопровождается решением типовых задач. Начало доказательства теорем, а также решений примеров обозначено символом ❐, а их окончание — символом ■. Таким же знаком ■ обозначен конец рассмотрения примеров, иллюстрирующих некоторое понятие или факт. Иногда изложение прерывается вопросами ? к читателю, цель которых проверить, усвоено ли изложение теоретического материала, понятно ли решение примера.

К каждому параграфу приводятся контрольные вопросы, ориентированные на активное усвоение основных понятий, отношений и фактов в их взаимосвязи, и задачи, помогающие усвоить материал на основном уровне. В конце каждой главы приведены дополнительные задачи. Их назначение — обеспечить уровневую и профильную дифференциацию обучения стохастике. Задания базового уровня отмечены знаком °. Наиболее сложные задачи обозначены звездочкой *. Остальные задачи (их большинство) не отмечены никаким знаком.

Пособие содержит и материал, который можно опустить при первом прочтении: это могут быть доказательства утверждений, исторические экскурсы, материал, направленный на расширение и углубление основного текста. Исторические экскурсы отмечены знаком  и специальным шрифтом. Углубленный материал обозначен знаками ➤ ➤.

В конце пособия приводятся ответы практически ко всем задачам и указания к решению наиболее трудных задач.

Автор выражает глубокую благодарность А. Л. Павлову за поддержку при работе над книгой, за ценные замечания, способствовавшие ее улучшению.

Автор

Введение

На каждом шагу мы встречаемся с явлениями и действиями, исход которых зависит от случая. Их обычно называют ***случайными испытаниями***. Рассмотрим соответствующие примеры.

- Жребием определяют футбольную команду, начинающую игру ударом с центра поля; его исход случаен.
- Случайным испытанием является бросание игрального кубика; случаен выбор костей при игре в домино, результат раздачи карт при карточной игре, исход покупки лотерейного билета; случайно извлечение бочонка при игре в лото и т. д.
- Покупая булочку с изюмом, мы не можем заранее сказать, сколько изюминок окажется в ней; их число случайно.
- На городском перекрестке авария автотранспорта может произойти, а может и не произойти; это явление случайно.
- Мы посеяли на огороде какую-то культуру; исход наших ожиданий случаен: мы не знаем, каким будет урожай этой культуры, это зависит от многих факторов, которые мы не можем учесть.
- Случайными являются и такие явления, как смертность и рождаемость в определенном регионе.
- Невозможно точно на длительный период предсказать, какой будет погода в определенном населенном пункте — метеорологические явления случайны.

Перечень таких примеров можно продолжить (сделайте это самостоятельно!). Но и приведенных достаточно для обоснования утверждения, с которого начато изложение.

Анализ подобных ситуаций, *попытка обнаружить в них определенные закономерности* в большой степени стимулировали возникновение такой науки, как ***теория вероятностей и математическая статистика***.

Теория вероятностей и математическая статистика — один из наиболее увлекательных и вместе с тем доступных разделов математики. Он уже давно вошел в программу средней школы во многих странах мира. Наконец, он получил «право гражданства» и в отечественной школе.

Истоками появления теории вероятностей как науки послужили задачи, поставленные игроками в азартные игры. Для основателей теории вероятностей Б. Паскаля (1623—1662) и П. Ферма (1601—1665) эти задачи служили как бы отдыхом от других абстрактных занятий. Очевидно, первым, кто осознал значение теории вероятностей в различных прикладных науках, был С. Лаплас (1749—1827). Это значение быстро усиливалось по мере расширения области применения теории вероятностей. В настоящее время вероятности широко используются в современной физике, биологии, антропологии, страховании, бизнесе и других областях человеческой деятельности. Теория вероятностей стала чуть ли не самой первой по прикладному значению из всех математических дисциплин. А если к этому добавить, что теория вероятностей играет важную роль в становлении мировоззрения (понимании связи между «случайным» и «необходимым», понимании «детерминистских» и «статистических» закономерностей и т. п.), то станет ясно, что основы теории вероятностей должны входить в багаж каждого образованного человека.

Данная книга приглашает вас в мир вероятности и статистики. Этот мир необходим любому человеку, так как наилучшее решение в любой жизненной ситуации можно принять только тогда, когда имеется необходимая информация. Каким бы делом не занимался человек, он зачастую вынужден принимать решения с учетом многих обстоятельств, не имея полной и точной информации. Любую доступную информацию следует использовать как можно полнее. ***Статистический анализ*** помогает получать информацию на основе данных и оценивать качество этой информации. ***Вероятность*** позволяет оценить риски, влияние случайности, обеспечивает правдоподобные оценки получения различных возможных результатов.

Статистика занимается сбором, обработкой и анализом данных. Основное предназначение статистики состоит в том, чтобы помочь людям лучше понять проблемы, с которыми они сталкиваются.

Пример 1. Вы решили в свободное от занятий время заняться продажей газет. Этому вы можете посвятить только ограниченное время в течение дня. Чтобы принять разумное решение, нужно обладать определенной информацией. Желательно знать, какие газеты пользуются спросом; в какое время суток торговля идет более оживленно; в каких местах лучше организовать эту торговлю; какая торговая наценка окажется разумной; велика ли конкуренция со стороны других продавцов газет и т. д. ■

Пример 2. Вы решили приобрести качественный компьютер и по возможно невысокой цене. Чтобы принять разумное решение, нужно обладать определенной информацией. Желательно знать, где продаются компьютеры; на каких фирмах они собирались; какова их стоимость; каково их качество; какую гарантию предоставляют фирмы; каково качество обслуживания компьютеров фирмами-продавцами и т. д. ■

Пример 3. Вы выбираете вуз и специальность, по которой планируете продолжать образование. Вас должно интересовать, имеется ли у вуза лицензия на подготовку специалистов данного профиля; насколько эта специальность востребована в обществе; насколько сложно поступить на данную специальность; какова стоимость обучения; пользуются ли спросом специалисты, оканчивающие данный вуз, и т. д. ■

Перечень подобных примеров можно продолжить.

Статистический анализ осуществляется в несколько этапов.

1. ***Планирование исследования*** предусматривает составление плана сбора данных. В частности, решается вопрос о том, есть ли необходимость и возможность исследовать всю совокупность (ее называют ***генеральной совокупностью***) или можно ограничиться только ее частью, которую называют ***выборкой***. Подробно о построении выборки будет говориться в главе 5. Планирование предусматривает определение необходимого объема данных: их должно быть достаточно для получения обоснованных выводов. Но их не должно быть слишком много: это приводит к расточительному использованию времени и средств. Этот этап включает и подготовку специальных бланков для регистрации собранных данных.

Например, при подготовке к продаже газет можно спланировать наблюдение за бизнесом ваших друзей. В ходе наблюдения необходимо выяснить, какие издания они продают, в какое время есть наибольшее число покупателей, сколько времени они тратят на эту работу, по какой цене они получают различные издания, какова их торговая наценка, сколько времени затрачивается на получение газет и на сдачу выручки, как далеко место продажи газет от места вашего проживания, каковы транспортные расходы и т. д.

2. ***Сбор и регистрация данных*** предполагает реализацию составленного плана исследования. Этот этап предусматривает подбор людей, занимающихся сбором данных, их инструктаж, обеспечение их ответственности за своевременность и достоверность собранной информации, правильное заполнение подготовленных бланков.

3. ***Предварительная обработка данных*** предусматривает рассмотрение данных с различных точек зрения: их описание, обобщение, графическое представление, выяснение того, нет ли среди них грубых ошибок, соответствуют ли они запланированным методам анализа.

После сбора информации, необходимой для организации продажи газет, ее необходимо классифицировать, упорядочить, чтобы из хаотичного набора данных она приобрела форму, удобную для анализа и сравнения.

Предварительной обработке данных будет посвящена первая глава.

4. ***Оценка неизвестной величины*** представляет собой наиболее обоснованное, основанное на имеющихся данных, предположение о возможном значении интересующей нас величины. Рассмотрим некоторые примеры неизвестных величин для оценивания:

- количество газет, которое реально продать за сутки;
- время работы компьютера до выхода из строя;
- шансы трудоустроиться по специальности после окончания вуза.

Кроме того, есть возможность вычислить величину ошибки, которая возникает при использовании оценки вместо истинного, но неизвестного значения величины. Этой цели служат так называемые доверительные интервалы, которые позволяют с заданной вероятностью утверждать, что значение неизвестной величины лежит между двумя определенными числами.

5. ***Проверка статистических гипотез*** заключается в использовании данных для осуществления выбора одной из двух (или более) различных возможностей при решении вопроса в неоднозначной ситуации. Приведем примеры гипотез, которые можно было бы проверить с использованием данных:

- для продажи газет удобно время с 18 до 19 часов;
- уровень бракованных компьютеров определенной марки не превышает 0,001;
- на выбранную специальность поступает не менее 90% желающих.

Каждая из сформулированных гипотез может оказаться верной или нет. Результатом проверки гипотезы является заключение о том, что данные противоречат гипотезе или нет.

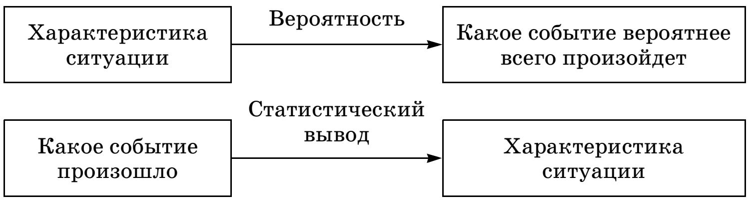
Оценка неизвестных параметров и проверка статистических гипотез подробно будут рассмотрены в главе 5.

Статистика помогает принимать решения. Она является одним из компонентов принятия решения. Но она не отвергает опыт, здравый смысл и интуицию. Если результаты статистического анализа расходятся с вашей интуицией, нужно разобраться, в чем причина. Статистический анализ может оказаться некорректным, если он проведен на основе неточных данных, если в его основу положены неверные допущения. С другой стороны, может подвести и интуиция, если она не базируется на фактах.

Теория вероятностей изучает правила оперирования со случайными явлениями и законы, которые руководят этими явлениями. Вероятность показывает возможность наступления в будущем каждого из потенциально возможных событий. Она может рассчитываться на основании информации о некоторой ситуации. Рассмотрим примеры различных ситуаций, где для принятия решения необходимо вычислить или оценить значение вероятности:

- Вы решили в свободное от занятий время заняться продажей газет. Вас интересует, какова вероятность получения прибыли в течение месяца.
- Вы решили купить компьютер. Интерес представляет вероятность того, что в течение пяти лет он не выйдет из строя.
- Вы закончили университет. Какова вероятность того, что вы трудоустроитесь по специальности?

Применение вероятности в некотором смысле обратно применению статистики. В то время как статистика помогает переходить от наблюдений к обобщениям рассматриваемой ситуации, вероятность позволяет, исходя из характеристики ситуации, выяснить шансы получения именно таких данных. Эта обратная связь может быть графически изображена следующим образом:



В условиях неопределенности нельзя точно знать, какое событие произойдет, всегда есть некоторая вероятность ошибки. Используя понятие вероятности, можно добиться того, чтобы ошибка наблюдалась не более чем в заданном проценте случаев.

Элементы теории вероятностей будут рассмотрены в главах 2 и 4. Глава 3 посвящена изучению комбинаторики, которая в определенных случаях помогает вычислять вероятности событий.

Контрольные вопросы

1. Почему следует изучать статистику?
2. В чем различие между теорией вероятностей и статистикой?
3. На каких данных основывалось последнее решение, принятое вами?
4. С какими статистическими данными вы встречались в газетах и других средствах массовой информации?
5. Какова роль вероятности при принятии того или иного решения?

Задачи

1. Какой из этапов статистического анализа представлен в каждой из следующих ситуаций?

а)° Администрация школы изучает вопрос об успеваемости учащихся выпускного класса по математике.

б)° Происходит отбор кандидатов для занятий в школе олимпийского резерва. Учителями физкультуры представлены количественные данные о спортивных достижениях учащихся школы.

в) Прогнозируется объем макулатуры, которая будет собрана в школе в следующем квартале. Высказывается некоторое предположение.

г)° Предстоит укомплектовать сборную школы по легкой атлетике для выступления на районных соревнованиях. У вас в распоряжении имеются данные о спортивных достижениях потенциальных кандидатов в сборную.

2. Школьная столярная мастерская заключила договор с мебельной фабрикой на изготовление отдельных деталей. Если в мастерскую привезут очень много древесины для выполнения заказа, то нужно будет искать место для ее хранения, она может испортиться от длительного хранения. Если привезут мало древесины, то будут простаивать некоторые учащиеся, привлеченные к выполнению заказа, заказ может быть не выполнен в срок. Однако вы точно не знаете, сколько необходимо древесины для обеспечения бесперебойной работы мастерской. Какой этап статистического анализа нужно реализовать для решения этой проблемы?

3. Проводится тендер на строительство некоторого объекта, в котором участвует несколько фирм. Собраны данные о качестве работ, выполняемых этими фирмами, о сроках выполнения подобных строительных работ, об их стоимости, о материалах, используемых при строительстве этими фирмами, и другие показатели. Какой этап статистического анализа реализуется при выборе фирмы?

4. Собраны данные опроса будущих избирателей об их отношении к двум партиям. Какой этап статистического анализа предстоит реализовать при выяснении вопроса, какой из партий отдают предпочтение избиратели?

5. Вам предстоит поменять рубли на доллары. Вы выбираете пункт обмена, где это выгоднее всего сделать. Какой этап статистического анализа вам нужно реализовать?

6. Фирма решила провести широкое рекламирование своей продукции. Исследуется эффективность рекламы.

а) В чем состоит планирование исследования?

б) Как собрать необходимые данные?

в) Что собой представляет предварительная обработка данных?

г) Какие величины нужно оценить?

д) Сформулируйте гипотезы, которые должны быть проверены в ходе исследования.

7. Опишите все этапы статистического анализа для каждой из следующих ситуаций.

а)° Исследуется качество автомобильных шин.

б) С целью уменьшения объема бракованной продукции, выпускаемой на некотором станке, проводится его переналадка. Исследуется ее эффективность.

в) Расшифровывается древняя рукопись.

г) Исследуется прибыль, получаемая от игры на игральных автоматах.

Глава 1

ОПИСАТЕЛЬНАЯ СТАТИСТИКА

О назначении описательной статистики можно судить по ее названию: она имеет дело с числами, характеризующими ту или иную интересующую нас ситуацию. Вот примеры статистической информации:

- уровень преступности в регионе;
- средняя зарплата в различных отраслях региона;
- уровень безработицы;
- число несчастных случаев на шахтах;
- число мобильных телефонов, проданных в текущем месяце;
- таблицы продолжительности жизни;
- уровень заболеваемости СПИДом;
- уровень достижений учащихся по математике;
- данные о доступности заданий единого государственного экзамена по математике;
- число граждан СНГ, обучающихся в Московском государственном университете, и т. п.

Ценность описательной статистики заключается прежде всего в том, что она дает сжатую и концентрированную характеристику изучаемого явления. Рассмотрим следующий пример. Пусть на некотором предприятии работает 1500 человек. Бухгалтерская ведомость на зарплату довольна большая. Информация о том, что средняя месячная зарплата работников этого предприятия составляет 8200 р., дает определенное, хотя и неполное представление об уровне заработной платы на этом предприятии.

Предмет исследований во многих сферах отличается исключительной сложностью, изменчивостью, индивидуальным многообразием явлений и процессов. Эти процессы происходят неоднозначно. Поэтому применение одинаковых подходов, средств, технологий дает в каждом конкретном случае различные результаты в зависимости от субъективных факторов, от обстоятельств, которые нельзя контролировать и которые влияют на протекание процесса. Неоднозначность протекания процесса порождается наличием присущего ему случайного. Но это не означает отсутствие общих закономерностей в изучаемых процессах и явлениях. Например, невысокая скорость чтения у отдельного учащегося является случайным событием, но у ученика, любящего читать, она встречается существенно реже, чем у того, кто редко берет книгу в руки. Эта устойчивость появления тех или других случайных событий уже является закономерностью.

Как мы отмечали во введении, теория вероятностей изучает правила оперирования со случайными явлениями и законы, которые руководят этими явлениями. Разработкой методов изучения свойств случайных событий и явлений занимается ***статистика***. Статистика имеет различные функции: информационную, прогностическую и аналитическую.

Информационная функция статистики состоит из сбора, обобщения и представления всем заинтересованным лицам достоверной, своевременной информации об исследуемом явлении. В связи с тем, что иногда исследованию подлежат тысячи объектов (детали, избиратели, учащиеся и т. п.), необходимым является переход от сплошного изучения к выборочному по многим показателям. Поэтому важное значение приобретают технологии сбора, обработки и анализа данных, которые позволяют использовать информационные возможности частичных первоначальных данных для разработки обобщенной информации о том или ином процессе.

Прогностическая функция статистики состоит в оценивании вероятностей тех или иных случайных событий, которые происходят в изучаемом процессе, показателей тех или иных случайных величин, связанных с этим процессом. Эта функция служит основой для принятия управленческих решений. С помощью этой функции можно получить сигнал о возможности появления кризисных явлений в изучаемом процессе, если не внести каких-то изменений в управление им.

Аналитическая функция статистики состоит, во-первых, в количественном исследовании тенденций развития процесса; во-вторых, в изучении этого процесса в динамике; в-третьих, в измерении связей между разными факторами, влияющими на процесс, и его результатами.

В данной главе начинается изложение вероятностно-статистических методов, которые используются при изучении объективных закономерностей явлений и процессов. В частности, будет рассмотрена первичная обработка статистического материала, статистические характеристики, графические методы представления информации, анализ информации, представленной в различных формах (табличной, графической и т. п.).

§ 1.1. Классификация данных и измерительные шкалы

Статистические данные могут быть представлены в различных формах. Набор данных содержит одно или несколько значений для каждого из отдельных объектов. В качестве таких объектов могут выступать люди, города, компьютеры, книги или все, что представляет интерес для изучения. Эти объекты называют ***элементарными единицами***. Для каждого объекта регистрируют один и тот же признак или признаки. Например, регистрируется рост и масса людей; численность населения, уровень рождаемости и смертности для городов; объем памяти, быстродействие, стоимость компьютеров; год издания, количество страниц, издательство для книг и т. д. Признак, который регистрируется для каждого из объектов, называют ***переменной***. Наборы данных классифицируют по следующим признакам:

- по количеству переменных (одномерные, двумерные или многомерные наборы данных);
- по типу данных (количественные или качественные);
- по тому, важна ли упорядоченность данных во времени или нет.

Одномерные наборы данных содержат только один признак для каждого объекта. Эти данные позволяют определить типичное значение признака, насколько значения отличаются друг от друга, требуют ли отдельные данные особого внимания. Примером одномерных данных является информация о средней зарплате в регионе по отраслям. Она позволяет назвать отрасль с самым высоким уровнем зарплаты, понять, насколько отличаются уровни средних зарплат в различных отраслях друг от друга, обратить внимание на отрасли, где уровень зарплаты самый низкий.

? Приведите пример одномерных данных.

Наборы ***двумерных*** данных содержат информацию о двух признаках для каждого из объектов. Кроме того, что они дают возможность получить два набора одномерных данных, двумерные данные позволяют установить, существует ли связь между двумя переменными, насколько сильно связаны переменные, можно ли предсказать значение одной переменной по значению другой и если да, то с какой надежностью. Например, данные опроса студентов о том, удовлетворены ли они уровнем теоретической и практической подготовки, получаемой в вузе (значения обеих переменных записываются в виде да/нет, или 1/0), позволяют установить, есть ли связь между уровнями теоретической и практической подготовки в вузе.

Приведите пример двумерных данных.

Многомерные данные содержат информацию о трех или более признаках для каждого объекта. В дополнение к той информации, которую можно извлечь из одномерных и двумерных наборов, многомерные данные можно использовать для получения информации о том, существует ли простая зависимость между этими признаками, насколько они взаимосвязаны (речь идет не только о попарной взаимосвязи признаков, но и о зависимости в совокупности), можно ли предсказать значение одной переменной на основании значений остальных. Примером многомерных данных является число граждан стран СНГ и Балтии, обучающихся в МГУ на 01.11.2003 г., по странам, приведенное в таблице 1.1[1]. Здесь содержатся данные о числе студентов, магистрантов, стажеров и аспирантов из республик бывшего СССР, обучающихся в МГУ.

Таблица 1.1

№ п/п	Страна	Студенты	Магистранты	Стажеры	Аспиранты	Всего
1	**2**	**3**	**4**	**5**	**6**	**7**
1	Азербайджан	57	2	0	4	63
2	Армения	32	2	0	1	36
3	Беларусь	89	0	1	1	91
4	Грузия	49	2	0	3	54
5	Казахстан	517	5	3	2	527
6	Кыргызстан	27	2	0	1	30
7	Латвия	36	1	0	3	40
8	Литва	37	0	0	0	37
9	Молдова	83	0	0	1	84
10	Таджикистан	13	0	0	1	14
11	Туркменистан	14	0	0	0	14
12	Узбекистан	59	2	1	0	62
13	Украина	1134	7	1	7	1149
14	Эстония	21	1	1	1	24
	Всего	2168	24	7	25	2224

[1] Данные взяты из газеты «Московский университет», № 39, 2003.

Даже беглый взгляд на данные, приведенные в таблице, позволяет сделать некоторые выводы: число граждан стран СНГ и Балтии, обучающихся в МГУ, не пропорционально численности населения этих стран; наличие большого числа граждан Казахстана и Украины, обучающихся в МГУ, объясняется наличием Казахстанского филиала МГУ и Черноморского филиала МГУ; не все страны бывшего СССР перешли на подготовку бакалавров и магистров и т. д.

Значения переменных, которые регистрируются с помощью чисел, имеющих содержательный смысл, называют ***количественными данными***. Данные, приведенные в столбцах 3—7 таблицы 1.1, являются количественными. В то же время числа, приведенные в первом столбце этой таблицы, не являются количественными данными: они только указывают на номер государства при их алфавитном расположении. Эти числа не имеют содержательной интерпретации. С количественными данными можно выполнять все обычные операции над числами, такие, как вычисление среднего (см. § 1.5) и оценку изменчивости (см. § 1.7).

В зависимости от того, какие значения может потенциально принимать переменная, выделяют два типа количественных данных: дискретные и непрерывные.

Дискретная — это такая переменная, которая может принимать значения только из некоторого списка определенных чисел. Примерами дискретной переменной являются число детей в семье; число вызовов «скорой помощи», поступающих в больницу; число отказов изделия; число клиентов, обратившихся в фирму за определенный промежуток времени, и т. д.

? Является ли дискретной переменной длина початка кукурузы, выращенной на некотором участке?

Непрерывной будем считать любую переменную, не являющуюся дискретной. Она принимает значения из некоторого промежутка. Примерами непрерывной переменной является рост взрослого человека (например, от 140 до 230 см), фактическая масса буханки хлеба (например, от 750 до 830 г), дальность полета снаряда, урожайность культуры, выращенной в хозяйстве, и т. п.

Есть данные, которые регистрируют определенное качество, которым обладает объект. Такие данные называют ***качественными***. Даже если значениям этого качества можно приписать числа (например, полу человека приписать соответственно числа 0 и 1), то обрабатывать эти числа как количественные данные нельзя. Примерами качественных данных являются тип школы, где обучается ребенок (лицей, гимназия, специализированная физико-математическая школа); должность, которую занимает сотрудник на предприятии; названия газет, которые читают в определенном городе, и т. п.

? Приведите пример качественных данных.

Качественные данные бывают двух типов: ***порядковые***, для которых существует имеющий содержательный смысл порядок, и ***номинальные***, для которых нет содержательно интерпретируемого порядка.

Порядковые данные можно ранжировать и использовать это ранжирование при проведении статистического анализа. Примером порядковых данных являются ответы на вопросы анкеты, содержащей следующие варианты ответов: да; больше да, чем нет; больше нет, чем да; нет. Хоть и можно выразить эти ответы числами (например, 4, 3, 2, 1), но предложенная шкала оценок носит субъективный характер. Нельзя считать, что разница между ответами 4 и 3 такая же, как и между ответами 2 и 1. Также нельзя считать, что ответ 3 в три раза лучше ответа 1.

Для номинальных данных нет числовых значений и нет основы для ранжирования. Примерами номинальных данных являются регионы России, из которых приехали студенты МГУ; названия фирм, изготавливающих моющие средства; пол работников фирмы.

? Оценки одного учащегося по пяти различным предметам являются порядковыми, номинальными или количественными данными?

Если порядок записи значений данных во времени имеет содержательный смысл, то говорят, что эти данные представляют собой ***временной ряд***. Эти данные представляют информацию об объекте в различные моменты времени. Если порядок записи данных во времени не существенен, то говорят об ***одном временном срезе***. Эти данные представляют информацию об объектах в определенный момент времени. Примерами временного ряда являются данные о преступности в регионе за несколько лет; об уровне безработицы за несколько лет; динамика успешности учащихся по математике и т. п. Примерами одного временного среза являются данные о числе преступлений в районах области в определенный год; число безработных среди различных возрастных групп населения по данным на определенный день; данные о выполнении теста учащимися класса и т. п.

? Табель ученика, где проставлены оценки за каждую четверть и итоговые оценки за год, является носителем временного среза или временного ряда? А аттестат о среднем образовании?

Иногда данные классифицируют также по тому, собирались ли они специально для запланированного анализа или собирались ранее для других нужд. В первом случае их называют ***первичными данными***, а во втором — ***вторичными данными***. Получение первичных данных сопряжено зачастую с большими затратами средств и времени, но зато исследователь получает то, что ему нужно. Вторичные данные, как правило, обходятся дешевле, их быстрее можно получить, но при этом не всегда в них можно найти то, что нужно.

Многие статистические данные получают в процессе измерений. Целью измерений является получение информации о признаках объектов, организмов, событий. Измеряется не сам объект, а только свойства или отличительные признаки объекта. Например, измеряется не ученик, а его рост, масса, его скорость чтения, достижения по математике, спортивные достижения и т. п. Измерения осуществляются путем установления соответствия между числами и объектами, которые являются носителями подлежащих измерению свойств. Измерения могут проводиться на разных уровнях. Различным уровням измерений соответствуют различные шкалы:

1) номинальная шкала;
2) порядковая, или ранговая, шкала;
3) шкала интервалов;
4) шкала отношений, или шкала пропорций;
5) логарифмическая шкала.

Номинальная шкала используется для регистрации самого низшего уровня измерений, предполагающего наличие минимальных предпосылок для измерения. При измерениях на данном уровне практически не используются числа. Здесь важно установить подобие или различие объектов по некоторому признаку, т. е. при этом имеют дело с качественными данными. Рассмотрим примеры.

▪ Распределения учащихся по классам, по половому признаку, по месту жительства, по видам спорта, которыми они занимаются, по числу детей в семье являются примерами величин номинальной шкалы. При этом возможно распределение учащихся по двум или более признакам (двумерные или многомерные данные).

- Перечень фирм, занимающихся производством грузовых и легковых автомобилей, автомобилей специального назначения, автобусов; отличительные признаки автомобилей являются примерами величин номинальной шкалы.

С помощью подсчета можно установить частоту той или иной категории (число мальчиков и девочек в школе; число учащихся, проживающих в каждом микрорайоне; число учащихся в каждом классе; число учащихся, занимающихся тем или иным видом спорта; количество фирм, занимающихся производством автобусов и т. д.). При этом можно определить наиболее часто встречающуюся величину (класс, в котором учится наибольшее число учащихся; вид спорта, пользующийся наибольшей популярностью у учащихся; тип автомобиля, производством которого занимается наибольшее число фирм). Категории данных номинальной шкалы обозначаются, как правило, словесно (вербально).

Порядковая, или ***ранговая***, ***шкала*** указывает лишь *последовательность* носителей признака или *направление* степени выраженности признака.

Например, учащихся можно ранжировать по количеству правильно выполненных тестовых заданий. Пусть учащиеся А, Б, В, Г, Д правильно выполнили соответственно 21, 16, 12, 9 и 3 задания. Графически это можно изобразить так, как показано на рис. 1. Эта порядковая шкала имеет величины от 1 до 5, и учащиеся на ней размещены в зависимости от количества правильно выполненных заданий: А — первый, Д — пятый. Из рисунка видно, что интервалы, разделяющие места в ряду, различны по величине. По этой причине нецелесообразно складывать, вычитать, умножать и делить порядковые места.

Шкала школьных оценок по одному предмету является порядковой шкалой, так как интервалы между отдельными баллами не отражают разрыва между реальными результатами. Мы знаем только, что ученик, получивший оценку «5» по какому-то предмету, знает этот предмет лучше того, кто получил «4». Но нельзя утверждать, что различие в знаниях этих учащихся такое же, как и в знаниях тех, кто получил «4» и «3». Так как шкала оценок яв-

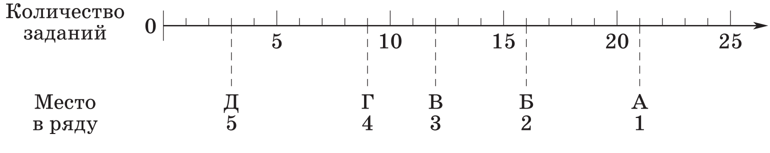


Рис. 1

ляется порядковой шкалой, то некорректно выставлять итоговую оценку как среднюю арифметическую текущих оценок.

Другим примером порядковой шкалы является шкала должностей, которые занимают работники на предприятии: директор, заместитель директора, начальник отдела, заведующий лабораторией, старший инженер, инженер, младший инженер, техник, старший лаборант, лаборант.

? Является ли шкала воинских званий примером порядковой шкалы?

На ***шкале интервалов*** равные интервалы отображают одинаковую меру величины измеряемого признака. Например, 1 см между 3-м и 4-м сантиметрами на шкале измерений длин имеет такой же смысл, как и 1 см между 82-м и 83-м сантиметрами. Другими словами, на шкале интервалов расстояния между соседними делениями равны. На интервальной шкале вполне осмысленным является вопрос «на сколько?». Но не всегда, пользуясь интервальной шкалой, можно формулировать вопрос «во сколько раз?». Дело в том, что на шкале интервалов устанавливаются произвольно начало отсчета (нуль шкалы), единица измерения и направление отсчета. Примером интервальной шкалы является температурная шкала по Цельсию. Разность между температурами воздуха +30 и +20 °С столь же велика, как и между −10 и −20 °С. Однако нельзя утверждать, что при температуре воздуха +30 °С в полтора раза теплее, чем при температуре +20 °С. Даже если температура воздуха равна 0 °С, нельзя утверждать, что тепла нет совсем: ведь начало отсчета выбрано произвольно.

Шкалы на большинстве физических приборов (амперметр, вольтметр и др.) являются интервальными. Шкала коэффициента интеллекта IQ является шкалой интервалов.

Шкала интервалов является метрической, с ее помощью можно выполнять сложение и вычитание. Она имеет значительные преимущества по сравнению с номинальной и порядковой шкалами.

Шкала отношений, или ***шкала пропорций***, дает возможность устанавливать отношения значений измеряемого признака благодаря тому, что значению шкалы «0» соответствует величина, для которой измеряемый признак отсутствует. Другими словами, начало отсчета на этих шкалах выбирают непроизвольно. Примерами шкалы отношений являются меры длины (м, см и т. д) и массы (кг, г и т. д.). Предмет длиной 100 см вдвое длиннее предмета длиной 50 см.

➤

Иногда данные нуждаются в преобразованиях. В частности, потребность в этом возникает, когда в ряду данных одно или несколько данных существенно превышают остальные. Если данные явно несимметричны, то заменяют каждое значение приведенного набора данных *логарифмом* этого значения с целью упростить статистический анализ. ***Логарифмирование*** преобразует «скошенные» (асимметричные) данные в более симметричные, так как происходит «растягивание» шкалы возле нуля, малые значения, сгруппированные вместе, распределяются вдоль шкалы. В то же время логарифмирование собирает вместе большие значения на правом конце шкалы. Наиболее часто применяют десятичные и натуральные логарифмы. Равным расстояниям на ***логарифмической шкале*** соответствует на исходной шкале равные процентные увеличения, а не равные увеличения значений.

Пример. В таблице 1.2 представлена численность населения (в тыс. чел.) в республиках бывшего СССР в 1976 г.

Таблица 1.2

Россия	Украина	Белоруссия	Узбекистан	Казахстан	Грузия	Азербайджан	Литва
134 650	49 075	9371	14 079	14 337	4954	5689	3315
Молдавия	**Латвия**	**Киргизия**	**Таджикистан**	**Армения**	**Туркмения**	**Эстония**	
3850	2497	3368	3486	2834	2581	1438	

Заменим все значения их десятичными логарифмами. В таблице 1.3 вместо численности населения представлены их десятичные логарифмы.

Таблица 1.3

Россия	Украина	Белоруссия	Узбекистан	Казахстан	Грузия	Азербайджан	Литва
8,13	7,69	6,97	7,15	7,16	6,69	6,76	6,52
Молдавия	**Латвия**	**Киргизия**	**Таджикистан**	**Армения**	**Туркмения**	**Эстония**	
6,59	6,40	6,53	6,54	6,45	6,41	6,16	

Как мы видим, данные симметрично группируются вокруг среднего значения 6,81. ■ ➤

Контрольные вопросы

1. Верно ли, что двумерные данные представляют собой два отдельных набора данных?
2. В чем различие между качественными и количественными данными?
3. В чем различие между порядковыми и номинальными данными?
4. Что легче анализировать — временные ряды или данные об одном временном срезе?
5. В чем различие между первичными и вторичными данными?
6. Какие возможности для анализа содержат в себе одномерные, двумерные и многомерные данные?
7. Какой смысл имеют результаты арифметических действий с переменной «номер» (№) в таблице 1.1?
8. К какому виду данных — временной ряд или один временной срез — относится цена различных типов бензина на автозаправочной станции?
9. Можно ли список чисел, значения из которого может принимать дискретная переменная, пополнить промежуточными значениями?
10. На какую величину могут отличаться друг от друга значения непрерывной переменной?

Задачи

8.° В таблице 1.4 содержатся рейтинги качества бритвенных лезвий[1].

Таблица 1.4

Изделие	Тип	Цена, $	Удобство в использовании
Gillette Sensor Excel	Бритвенный блок	4,50	Отличное
Schick Tracer	Бритвенный блок	3,34	Очень хорошее
Bic Twin Select for Sensible Skin	Одноразовые	3,68	Хорошее
Gillette Sensor for Women	Бритвенный блок	3,93	Отличное
Schick Silk Effects for Women	Бритвенный блок	4,64	Очень хорошее
Bic Twin Pastel	Одноразовые	2,03	Хорошее

[1] Данные взяты из книги: *Эндрю Ф. Сигел*. Практическая бизнес-статистика. — Москва — Санкт-Петербург — Киев, Вильямс, 2002.

а) Что является элементарной единицей в этом наборе данных?

б) Это одномерные, двумерные или многомерные данные?

в) Это временной ряд или один временной срез?

г) Переменная «тип» — количественная, порядковая или номинальная?

д) Переменная «цена» — количественная, порядковая или номинальная?

е) Переменная «удобство в использовании» — количественная, порядковая или номинальная?

9.° В таблице 1.5 приведены минимальные наборы из пяти продуктов питания для основных групп населения[1]. Нормы указаны в кг на 1 месяц.

Таблица 1.5

Продукт питания	Дети до 6 лет	Дети 6—18 лет	Взрослые	Пенсионеры
Говядина	0,975	1,525	1,17	1,0
Гречневая крупа	0,2	0,28	0,17	0,17
Капуста	1,37	1,825	2,33	2,29
Картофель	6,08	7,81	7,92	9,0
Молоко	7,61	6,84	5,0	6,92

а) Что является элементарной единицей в этом наборе данных?

б) Определите вид данных — одномерные, двумерные или многомерные?

в) Укажите качественные переменные, если они имеются.

г) Это временной ряд или один временной срез?

д) Есть ли в этих данных порядковая переменная?

10.° В отделе кадров учет работников ведется по следующей форме.

Номер работника	Месячная заработная плата, р.	Пол	Возраст, годы	Стаж работы, годы	Образование

а) Что является элементарной единицей в этом наборе данных?

б) Определите вид данных — одномерные, двумерные или многомерные.

[1] Данные взяты из газеты «Комсомольская правда. Украина» от 09.06.2005 г.

в) Укажите качественные и количественные переменные, если они имеются.

г) Это временной ряд или один временной срез?

д) Переменная «Образование» является порядковой или номинальной?

е) Можно ли выполнять арифметические действия с переменной «Номер работника»?

11. Для каждой переменной из данных, приведенных в задаче 10, определите, какие из указанных ниже операций можно применять к этой переменной:

а) сложение и вычитание;

б) подсчет числа работников в этой категории;

в) ранжирование по порядку;

г) вычисление процента работников в данной категории.

§ 1.2. Первичная обработка результатов измерений

Результаты измерения, как правило, регистрируют сначала в произвольном, хаотичном порядке, или в алфавитном порядке (например, при проверке скорости чтения учащихся вызывают в том порядке, в каком они записаны в классном журнале), или в том порядке, в каком поступают результаты измерения (например, результаты проверки контрольных работ или результаты тестирования поступают в том порядке, в каком лежат контрольные тетради или талоны для тестирования). В такой форме полученные данные неудобны для анализа и выявления закономерностей. Первичная обработка статистических данных состоит в упорядочении данных (по возрастанию или убыванию), подсчете некоторых показателей, характеризующих эти значения, в группировании данных.

? В каком порядке поступают данные при измерении длины прыжка учащихся класса?

Проиллюстрируем все сказанное на конкретном примере.

Пример. В таблице 1.6 представлены данные о скорости чтения учащихся 3-го класса, т. е. о количестве слов, которые ученик прочитывает за минуту.

Первый шаг обработки этих данных – это упорядочение по возрастанию или убыванию скорости чтения. В данном случае это удобно сделать по убыванию. В таблице 1.7 представлены те же 60 данных, но упорядоченные по убыванию от 110 до 27.

Таблица 1.6

№ ученика[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Скорость чтения	53	49	90	27	64	58	34	53	85	72	30	90	45	34	25
№ ученика	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Скорость чтения	61	49	39	45	56	72	34	82	47	64	29	78	58	32	64
№ ученика	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
Скорость Чтения	110	35	78	29	65	42	38	83	57	71	68	49	82	37	57
№ ученика	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Скорость чтения	55	29	43	78	39	34	67	92	28	55	52	85	49	34	56

[1] Номер ученика используется вместо его фамилии.

Таблица 1.7

№ ученика	31	53	3	12	9	57	38	23	43	27
Скорость чтения	110	92	90	90	85	85	83	82	82	78
Ранг ученика	1	2	3,5	3,5	5,5	5,5	7	8,5	8,5	11
№ ученика	49	33	10	21	40	41	52	35	5	25
Скорость чтения	78	78	72	72	71	68	67	65	64	64
Ранг ученика	11	11	13,5	13,5	15	16	17	18	20	20
№ ученика	30	16	6	28	39	45	20	60	55	46
Скорость чтения	64	61	58	58	57	57	56	56	55	55
Ранг ученика	20	22	23,5	23,5	25,5	25,5	27,5	27,5	28,5	28,5
№ ученика	1	8	56	2	17	42	58	24	13	19
Скорость чтения	53	53	52	49	49	49	49	47	45	45
Ранг ученика	31,5	31,5	33	35,5	35,5	35,5	35,5	38	39,5	39,5

Окончание табл. 1.7

№ ученика	48	36	18	50	37	44	32	7	14	22
Скорость чтения	43	42	39	39	38	37	35	34	34	34
Ранг ученика	41	42	43,5	43,5	45	46	47	50	50	50
№ ученика	51	59	29	11	26	34	47	54	4	15
Скорость чтения	34	34	32	30	29	29	29	28	27	25
Ранг ученика	50	50	53	54	56	56	56	58	59	60

Эта таблица позволяет установить ***ранг*** ученика, т. е. место, которое он занимает среди проверявшихся учеников, по скорости чтения. Чем меньше ранг, тем больше скорость чтения ученика. Ученик с рангом 1 имеет наибольшую скорость чтения. Поскольку имеются ученики с одинаковой скоростью чтения, то их ранги целесообразно считать одинаковыми, а именно — равными средним арифметическим соответствующих значений. Например, ученики со скоростью чтения 90 слов/мин занимают 3-е и 4-е места, ранг каждого из них равен $\frac{3+4}{2} = 3{,}5$. Ученики со скоростью чтения 78 слов/мин занимают 10-е, 11-е и 12-е места, ранг каждого из них равен $\frac{10+11+12}{3} = 11$. Ученики со скоростью чтения 49 слов/мин занимают 34-е, 35-е, 36-е и 37-е места, ранг каждого из них равен $\frac{34+35+36+37}{4} = 35{,}5$. ■

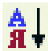
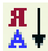
Почему ранг ученика со скоростью чтения 55 слов/мин равен 28,5?

Разумеется, если исследуется количество ошибок, допущенных учениками в диктанте, то целесообразно расположить полученные данные по возрастанию. Тогда чем меньше ранг, тем выше грамотность ученика. Ученик с рангом 1 допустил наименьшее количество ошибок, его можно считать самым грамотным.

Таблицу 1.7 можно сократить, заметив, что многие значения встречаются неоднократно. Число случаев, в которых встречается значение x_i, называют ***частотой*** значения x_i и обозначают n_i. Сокращенную таблицу можно представить в виде таблицы 1.8.

Таблица 1.8

x_i	110	92	90	85	83	82	78	72	71	68	67	65
n_i	1	1	2	2	1	2	3	2	1	1	1	1
x_i	64	61	58	57	56	55	53	52	49	47	45	43
n_i	3	1	2	2	2	2	2	1	4	1	2	1
x_i	42	39	38	37	35	34	32	30	29	28	27	25
n_i	1	2	1	1	1	5	1	1	3	1	1	1

Упорядочение данных и подсчет их частот можно выполнять с помощью редактора электронных таблиц Microsoft Excel. Для этого используются команды  или  и статистическая функция (ЧАСТОТА).

Сумму $n_1 + n_2 + \ldots + n_k$ называют ***объемом совокупности данных*** и обозначают n. Здесь k — число различных значений, которые принимает исследуемая величина. В рассмотренном примере $k = 36$, $n = 60$.

Понятно, что частота значения величины является ее определенной характеристикой, но недостаточной: в ней не учтен объем совокупности данных. Так, если скорость чтения 85 показали два ученика из 30 или из 60 учащихся, то эти показатели являются различными. В первом случае бóльшая доля учеников имеют приведенную скорость чтения. Поэтому целесообразно рассматривать такой показатель, как относительная частота значения.

Относительной частотой значения x_i *называют отношение частоты* n_i *этого значения к объему* n *совокупности данных.*

Будем обозначать относительную частоту значения x_i через ν_i:

$$\nu_i = \frac{n_i}{n}.$$

Для рассматриваемого примера частоты и относительные частоты скоростей чтения приведены в таблице 1.9.

Относительная частота значения является краткой и содержательной характеристикой рассматриваемой информации. Например, если считать, что для ученика 3-го класса нормой скорости чтения является 50 слов/мин, то относительная частота значений, не меньших нормы (т. е. от 110 до 52), в нашем примере равна $\frac{33}{60} = 0{,}55$.

Таблица 1.9

x_i	110	92	90	85	83	82	78	72	71	68	67	65
n_i	1	1	2	2	1	2	3	2	1	1	1	1
ν_i	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{60}$	$\frac{1}{30}$	$\frac{1}{20}$	$\frac{1}{30}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$
x_i	64	61	58	57	56	55	53	52	49	47	45	43
n_i	3	1	2	2	2	2	2	1	4	1	2	1
ν_i	$\frac{1}{20}$	$\frac{1}{60}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{60}$	$\frac{1}{15}$	$\frac{1}{60}$	$\frac{1}{30}$	$\frac{1}{60}$
x_i	42	39	38	37	35	34	32	30	29	28	27	25
n_i	1	2	1	1	1	5	1	1	3	1	1	1
ν_i	$\frac{1}{60}$	$\frac{1}{30}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{12}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{20}$	$\frac{1}{60}$	$\frac{1}{60}$	$\frac{1}{60}$

Этот показатель хорошо характеризует положение со скоростью чтения в 3-м классе: лишь немного больше половины третьеклассников достигли нормы. И не надо тщательно изучать, как были достигнуты такие результаты, очевидно, что нужна дополнительная работа для достижения нормы, или следует пересмотреть норму.

Чему равна относительная частота значений, не меньших 40 слов/мин?

Для большого количества данных на следующем этапе обработки статистических данных целесообразно их обобщение. В нашем примере можно не рассматривать в отдельности каждое значение скорости чтения, а разбить их на группы. Ведь сложно отличить скорость 32 от 33 слов/мин, но можно различить учащихся, скорость чтения которых находится в диапазоне от 30 до 40 и от 40 до 50 слов/мин. Тем самым приходим к необходимости ***группирования*** статистических данных. Имеются различные способы группирования. Рассмотрим некоторые из них.

Один из них реализуется в следующей последовательности действий.

1. ***Определение количества групп k***, на которые подразделяются все данные. Четкого правила выбора количества групп не существует. Они определяются содержанием рассматриваемой задачи. Обычно количество групп выбирают не меньше 12, но не больше 15. Малое количество групп (< 12) может исказить результаты, а большое количество (> 15) затрудняет роботу с таблицей.

2. ***Вычисление размаха данных*** ω — разности между наибольшим $x_{\max}$ и наименьшим $x_{\min}$ значениями, увеличенной на погрешность измерения величины. Если значения являются целыми числами, то $\omega = x_{\max} - x_{\min} + 1$. Объясним, почему прибавляется 1. Измерение любой величины приводит к приближенным значениям, ведь любой прибор имеет ограниченную точность; погрешность измерения, как правило, равняется половине значения деления прибора. В нашем примере скорость 110 фактически означает, что это значение находится в интервале 109,5—110,5. Поэтому в этом примере можно считать, что наибольшее значение равняется 110,5, а наименьшее — 24,5. Разность между ними: 110,5 − 24,5 = 86. Такой точно результат получим по формуле: $\omega = x_{\max} - x_{\min} + 1$. В самом деле, $\omega = 110 - 25 + 1 = 86$.

В общем случае к разности $x_{\max} - x_{\min}$ добавляется число, равное погрешности величины. Например, величина измеряется с точностью до десятых, ее наименьшее и наибольшее значения соответственно равны 3,2 и 6,7. Тогда размах данных равен 6,7 − 3,2 + + 0,1 = 3,6.

3. ***Нахождение длины группы.*** Для этого частное от деления размаха w на нижнюю границу количества групп (12) округляют с недостатком, а частное от деления ω на верхнюю границу количества групп (15) округляют с избытком. Округление проводят до разряда, соответствующего погрешности измерения. Если погрешность не превышает 1, то округление проводится до целых; если погрешность не превышает 0,1, то округление проводится до десятых и т. д. Длиной интервала удобнее выбирать нечетное число величин погрешности, поскольку в этом случае середина интервала будет целым количеством единиц погрешности. Так, серединой интервала от 30 до 34 длиной 34 − 29 = 5 (так же, как и при подсчете размаха, предполагается, что значения заключены в интервале (29,5; 34,5), длина которого равна 34,5 − 29,5 = 5) является число $\frac{30+34}{2} = 32$. Это число можно получить и другим способом: $30 + \frac{34-30}{2} = 32$. Если длина интервала является четным числом, то середина интервала будет дробным числом. Например, серединой интервала от 30 до 35 длиной 6 является число $\frac{30+35}{2} = 32{,}5$ или

$30 + \frac{35-30}{2} = 32{,}5$. В нашем примере $86 : 12 \approx 7$ (с недостатком), $86 : 15 \approx 6$ (с избытком). В качестве длины интервала можно принять число 7.

4. *Нахождение границ групп*. В крайние группы должны попасть наибольшее и наименьшее значения. Для этого следует начинать образовывать группы с величины, кратной длине интервала, причем нижняя группа должна содержать наименьшее значение. В нашем примере начинаем с числа 21. Тогда интервал 21—27 содержит наименьшее значение 24. Следующий интервал 28—34, дальше: 35—41, 42—48, и т. д., последний интервал 105—111. Всего получили 13 групп, последняя группа содержит наибольшее значение 110.

5. *Составление таблицы группированных значений величины*. Для этого нет надобности в упорядочении данных (табл. 1.10). В первой строке таблицы записывают полученные интервалы. Потом рассматривают подряд все полученные значения из таблицы 1.6. Первое значение — 53. В соответствующем столбце (49—55) ставят черточку. Следующее значение — 49. Черточку ставят в том же столбце. Следующее значение 90 попадает в интервал 84—97. Далее продолжают аналогично. Знак ~~////~~ (4 наклонные черточки, перечеркнутые горизонтальной черточкой) означает, что в соответствующий интервал попало 5 значений. Такие обозначения облегчают дальнейшие подсчеты частот.

В третьей строке приводятся значения частот для каждой группы. Они определяются по данным второй строки: подсчитывается количество черточек. Например, частота 9 означает, что 9 учеников имеют скорость чтения от 49 до 55 слов/мин.

В следующей строке приводятся значения так называемой ***накопленной частоты***, т. е. число значений, которые попали в этот интервал и все предшествующие. Значение накопленной частоты 21 означает, что 21 ученик имеет скорость чтения, не меньшую 63.

В четвертой строке приводятся значения относительных частот соответствующих интервалов, в последней — значения ***накопленных относительных частот***, т. е. отношений накопленных частот к объему совокупности данных. Эти значения приведены с тремя десятичными знаками. Может оказаться, что накопленная частота последней группы отличается от 1. Это объясняется наличием погрешности округлений.

Для первичной обработки статистических данных удобно использовать специальные компьютерные программы, например Microsoft Excel.

Таблица 1.10

Скорость чтения	105—111	98—104	91—97	84—90	77—83	70—76	63—69	56—62	49—55	42—48	35—41	28—34	21—27
Количество значений, попавших в группу	/	—	/	////	~~////~~ /	///	~~////~~ /	~~////~~ //	~~////~~ ///	~~////~~	~~////~~	~~////~~ ~~////~~ /	//
Частота	1	0	1	4	6	3	6	7	9	5	5	11	2
Накопленная частота	1	1	2	6	12	15	21	28	37	42	47	58	60
Относительная частота	0,017	0	0,017	0,067	0,1	0,05	0,1	0,117	0,15	0,083	0,083	0,183	0,033
Накопленная относительная частота	0,017	0,017	0,034	0,101	0,201	0,251	0,351	0,468	0,618	0,701	0,784	0,967	1,000

➤

Приведем несколько иной способ группирования данных. Он реализуется в следующей последовательности действий.

1. Выбирают количество интервалов k, на которые подразделяются все данные.

2. Вычисляют длину h каждого интервала по формуле

$$h = \left[\frac{x_{\max} - x_{\min}}{k - 1} \right],$$

где $x_{\max}$, $x_{\min}$ — соответственно наибольшее и наименьшее из наблюдаемых значений; $[x]$ — ближайшее целое число к x.

3. Левую и правую границы статистических данных $x_{\min}$ и $x_{\max}$ раздвигают на величину $\frac{h}{2}$ и получают соответственно $x'_{\min} = x_{\min} - \frac{h}{2}$, $x'_{\max} = x_{\max} + \frac{h}{2}$.

4. Определяют границы интервалов по формуле $x'_i = x'_{\min} + ih$, $i = 0, 1, \ldots, k$, $x'_0 = x'_{\min}$.

5. Группируют данные по интервалам, определяют количество значений n_i, попавших в каждый интервал (значения, которые совпадают с границами i-го и $(i+1)$-го интервалов, считают принадлежащими в одинаковой мере к обоим интервалам, т. е. к частотам n_i и n_{i+1} прибавляют по $\frac{1}{2}$).

6. Определяют относительную частоту наблюдений для каждого интервала.

Сгруппируем таким образом статистические данные, приведенные в нашем примере.

Пусть $k = 15$. Тогда $h = \left[\frac{110 - 25}{14} \right] = 6$, $x'_{\min} = 25 - \frac{6}{2} = 22$, $x'_{\max} = 110 + 3 = 113$.

Далее: $x'_0 = 22$, $x'_1 = 28$, $x'_2 = 34$, $x'_3 = 40$, $x'_4 = 46$, $x'_5 = 52$, $x'_6 = 58$, $x'_7 = 64$, $x'_8 = 70$, $x'_9 = 76$, $x'_{10} = 82$, $x'_{11} = 88$, $x'_{12} = 94$, $x'_{13} = 100$, $x'_{14} = 106$, $x'_{15} = 112$.

Сгруппированные данные представим в виде таблицы 1.11.

Таблица 1.11

x_i	22—28	28—34	34—40	40—46	46—52	52—58	58—64	64—70
n_i	2,5	8	7,5	4	5,5	9,5	3,5	4,5
ν_i	0,042	0,133	0,125	0,067	0,092	0,158	0,058	0,075
x_i	70—76	76—82	82—88	88—94	94—100	100—106	106—112	
n_i	3	4	4	3	0	0	1	
ν_i	0,05	0,067	0,067	0,05	0	0	0,017	

Первый способ выглядит несколько более предпочтительным, поскольку позволяет не иметь дело с дробями.

Иногда число групп n приближенно определяют по так называемой формуле Стерджесса: $n = 1 + 3{,}322 \lg N$, где N — численность совокупности. Тогда длину интервала h вычисляют по формуле

$$h = \frac{x_{\max} - x_{\min}}{n},$$

где $x_{\max}$ и $x_{\min}$ — максимальное и минимальное значения признака.

Применяя эту формулу к нашему примеру, получим: $N = 60$; $n = 1 + 3{,}322 \lg 60 \approx 7$; $h = \frac{110 - 24}{7} \approx 14$. Получили результат, существенно отличающийся от полученных при использовании других методов. При этом произойдет довольно большая потеря информации. ➤

Контрольные вопросы

1. Определите, к каким видам относятся данные, приведенные в примере (табл. 1.6): одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

2. Каковы цели предварительной обработки опытных данных?

3. Что показывает частота значения?

4. Что показывает относительная частота значения?

5. Что показывает накопленная частота значения?

6. Что показывает относительная накопленная частота значения?

7. Для чего проводят группирование данных?

8. Что показывает ранг объекта?

9. Как подсчитывают ранг объекта, если частота признака для этого объекта совпадает с частотами этого признака у других объектов?

10. Почему удобнее иметь дело с интервалами, длина которых выражается нечетным целым числом?

11. Верно ли, что при группировании данных происходит потеря информации?

Задачи

12.° Ниже приведены размеры проданной в магазине мужской обуви:

41	39	40	38	43	41	42	40	38	41	42	41	40
42	39	41	41	36	43	41	42	38	41	40	42	41
42	42	42	40	41	41	39	42	40	40	39	41	39
38	40	41	41	40	40	39	42	40	43	37	40	42
43	42	38	40	40	41	41	41	40	43	42	42	39
43	41	40	43	41	42	42	39	41	43	42	41	42
40	37											

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты значений.

в) Сгруппируйте данные.

13.° Ниже представлена продолжительность работы электронных ламп одного типа (в часах):

13,4	14,7	15,2	15,1	13,0	8,8	14,0	17,9	15,1	16,5	16,6	14,2
16,3	14,6	11,7	16,4	15,1	16,6	14,1	18,8	11,6	13,9	18,0	12,4
17,2	14,5	16,3	13,7	15,5	16,2	8,4	14,7	15,4	11,3	10,7	16,9
15,8	16,1	12,3	14,0	17,7	14,7	16,2	17,1	10,1	15,8	18,3	17,5
12,7	20,7	13,5	14,0	15,7	21,9	14,3	17,7	15,4	10,9	18,2	17,3
15,2	16,7	17,3	12,1	19,2	8,3	14,0					

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по убыванию, подсчитайте частоты, относительные частоты значений.

в) Сгруппируйте данные.

14. Средняя температура июня в Москве в течение 40 лет была такой (в °C):

12,0	13,8	14,0	14,9	15,9	16,9	18,0	20,0	12,0	13,1
14,0	15,0	16,0	17,0	18,1	20,2	12,0	13,0	13,9	15,0
16,9	16,8	18,4	14,0	12,0	13,9	15,0	15,9	17,2	16,0
17,5	19,2	14,0	12,8	14,2	14,9	16,9	18,0	19,3	15,8

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты значений.

в) Сгруппируйте данные.

15. Измерения чувствительности видеоканала у 50 телевизоров дали следующие результаты (в мкВ):

590	220	300	580	170	420	540	300	320	560
440	530	600	500	300	450	540	240	370	550
300	480	360	420	450	420	370	405	190	470
225	180	360	400	150	480	360	385	450	435
500	220	440	405	330	380	500	430	315	450

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по убыванию, подсчитайте частоты, относительные частоты значений.

в) Сгруппируйте данные.

§ 1.3. Вариационные ряды

Во многих исследованиях часто имеют дело с разнообразными совокупностями вещей или явлений, которые по одним признакам представляют собой единое целое, а по другим подразделяются на отдельные группы. Такие совокупности рассматривались выше.

Так, в примере предыдущего параграфа ученики 3-го класса — это определенная совокупность элементов (учеников), представляющих собой единое целое, поскольку элементы (ученики), которые ее составляют, объединены определенным признаком — все они учатся в 3-м классе. В то же время они подразделяются на отдельные группы по другим признакам: полу, скорости чтения, успешности обучения и т. п.

Участники областной олимпиады по математике образуют единое целое. В то же время они могут быть разделены на группы: по регионам, где они учатся; по успехам выступления на областной олимпиаде; по классам, в которых они учатся; по характеру математических способностей и т. п.

В задаче 12 § 1.2 приведены данные о размерах проданной мужской обуви. Мужская обувь — это определенная совокупность элементов, представляющих собой единое целое, поскольку элементы (обувь), которые ее образуют, объединены определенным признаком — все они предназначены для мужчин. В то же время они подразделяются на отдельные группы по другим признакам: размеру, фасону, производителям и т. п.

Совокупность, состоящая из однородных элементов, имеющих качественную общность, будем называть ***статистической совокупностью****. Элементы, из которых состоит данная совокупность, называют ее* ***членами****. Количество элементов в совокупности называют его* ***объемом****.* Объем совокупности будем обозначать через n.

Признак, по которому совокупность подразделяют на группы, называют ***аргументом****.* Признак (аргумент) будем обозначать прописными латинскими буквами X, Y, Z, ... *Отдельные числовые значения аргумента называют его* ***вариантами*** и обозначают через x_1, x_2, ..., x_k. (Скорость чтения — признак, его значения — $x_1 = 110$, $x_2 = 92$, ..., $x_{36} = 25$.) *Количество элементов совокупности, имеющих одинаковое числовое значение, мы назвали* ***частотой данной варианты***; частоты обозначили через n_1, n_2, ..., n_k; $n_1 + n_2 + ... + n_k = n$. *Отношение частоты варианты к объему совокупности мы назвали* ***относительной частотой варианты*** и обозначили через ν_1, ν_2, ..., ν_k; $\nu_1 + \nu_2 + ... + \nu_k = 1$.

? По какому аргументу разделена на группы совокупность, представленная в задаче 13 § 1.2? Каков объем этой совокупности?

В исследованиях, изучая тот или иной признак, часто приходится сталкиваться с такими совокупностями, члены которых принимают различные значения (наряду с одинаковыми). Такую переменчивость значений признака называют его ***варьированием***. Например, варьирование мы наблюдаем, изучая успешность учащихся по предмету, сформированность некоторого качества личности и т. п.

? Принимают ли члены совокупности, представленной в задаче 14 § 1.2, одинаковые значения или только различные?

Если две варианты признака в данной совокупности могут отличаться одна от другой не менее чем на определенное число или вообще совпадают, то такие данные называют ***дискретными*** (число учеников в классах школы; количество баллов, которые набирает ученик при тестировании, и т. п.). Если же две варианты признака могут отличаться одна от другой на произвольно малую величину, то такие данные называют ***непрерывными*** (процент учеников, которые имеют достаточный уровень подготовки по предмету в разных классах; время, за которое ученики пробежали 60 м на соревнованиях; продолжительность работы электронных ламп; температура воздуха и т. п.).

Ряд значений признака, или вариант, полученных вследствие массового обследования однородных вещей или явлений, размещенных в порядке возрастания или убывания их величин, вместе с соответствующими частотами (или относительными частотами) называют ***вариационным рядом***. Примерами вариационного ряда являются таблицы 1.8, 1.9, 1.11.

Если в вариационном ряде значения признака (варианты) заданы в виде отдельных конкретных чисел, то такой ряд называют ***дискретным*** (например, табл. 1.8).

Если в вариационном ряде значения признака заданы в виде интервалов, то такой ряд называют ***интервальным*** (например, табл. 1.10, 1.11).

Если в интервальном вариационном ряде в двух последовательных интервалах верхнее предельное значение признака одного интервала равняется нижнему предельному значению второго, условно будем считать, что это число принадлежит второму интервалу. *Разность между верхней и нижней границами интервала называют* ***шириной*** *этого* ***интервала***.

Рассматриваются еще так называемые ***кумулятивные*** вариационные ряды. В таких рядах вместо частот или относительных частот определенных вариант (или интервалов) записаны накопленные частоты или относительные частоты. Например, строки 1 и 4 или 1 и 6 в таблице 1.10 образуют кумулятивный вариационный ряд.

Для анализа статистических данных, содержащихся в вариационном ряде, целесообразно ввести такую числовую характеристику, как плотность распределения.

Если в интервальном вариационном ряде ширина интервала отлична от единицы, то определяют абсолютную и относительную плотности распределения.

Отношение частоты n_i интервала к ширине h_i этого интервала называют ***абсолютной плотностью распределения*** *для i-го интервала*. Будем обозначать ее символом p_i: $p_i = \dfrac{n_i}{h_i}$. Абсолютная плотность распределения — это частота, приходящаяся на единицу ширины интервала. Например, по данным таблицы 1.10 плотность распределения в интервале (70; 76) равна $p = \dfrac{3}{6} = 0{,}5$.

Относительной плотностью распределения π_i *для i-го интервала называют отношение относительной частоты интервала к его ширине*: $\pi_i = \dfrac{\nu_i}{h_i}$. По данным таблицы 1.10 относительная плотность распределения для интервала (77; 83) равна $\pi = \dfrac{0{,}1}{7} \approx 0{,}014$.

Пример. Измерения диаметров 50 валиков, выточенных на станке, дали следующие результаты (в мм):

14,51	14,42	14,56	14,47	14,46	14,35	14,48	14,53
14,21	14,31	14,35	14,68	14,56	14,28	14,36	14,21
14,52	14,23	14,41	14,46	14,69	14,54	14,36	14,15
14,37	14,51	14,25	14,55	14,51	14,36	14,62	14,55
14,38	14,33	14,40	14,52	14,48	14,51	14,55	14,39
14,54	14,58	14,48	14,37	14,38	14,51	14,36	14,15
14,24	14,32						

а) Построить дискретный вариационный ряд.
б) Построить интервальный вариационный ряд.
в) Для интервального вариационного ряда вычислить абсолютные и относительные плотности распределения.

❒ а) Объем совокупности равен 50. Выпишем все различные значения диаметра валика и вычислим их частоты и относительные частоты, накопленные частоты и относительные накопленные частоты. Результаты представим в таблице 1.12.

Построен дискретный вариационный ряд.

Таблица 1.12

x_i	14,15	14,21	14,23	14,24	14,25	14,28
n_i	2	2	1	1	1	1
Накопленная частота	2	2	5	6	7	8
ν_i	0,04	0,04	0,02	0,02	0,02	0,02
Накопленная относительная частота	0,04	0,08	0,10	0,12	0,14	0,16
x_i	14,31	14,32	14,33	14,35	14,36	14,37
n_i	1	1	1	2	4	2
Накопленная частота	9	10	11	13	17	19
ν_i	0,02	0,02	0,02	0,04	0,08	0,04
Накопленная относительная частота	0,18	0,20	0,22	0,26	0,34	0,38
x_i	14,38	14,39	14,40	14,41	14,42	14,46
n_i	2	1	1	1	1	2
Накопленная частота	21	22	23	24	25	27
ν_i	0,04	0,02	0,02	0,02	0,02	0,04
Накопленная относительная частота	0,42	0,44	0,46	0,48	0,50	0,54
x_i	14,47	14,48	14,51	14,52	14,53	14,54
n_i	1	3	5	2	1	2
Накопленная частота	28	31	36	38	39	41
ν_i	0,02	0,06	0,1	0,04	0,02	0,04
Накопленная относительная частота	0,56	0,62	0,72	0,76	0,78	0,82
x_i	14,55	14,56	14,58	14,62	14,68	14,69
n_i	3	2	1	1	1	1
Накопленная частота	44	46	47	48	49	50
ν_i	0,06	0,04	0,02	0,02	0,02	0,02
Накопленная относительная частота	0,88	0,92	0,94	0,96	0,98	1

б) Интервальный вариационный ряд можно построить группированием членов дискретного вариационного ряда. В соответствии с рекомендациями число интервалов выбираем между 12 и 15. Определяем размах данных по формуле $\omega = x_{\max} - x_{\min} + 0{,}01$ (добавляется единица наименьшего разряда, в которых приведены данные). Получим $\omega = 14{,}69 - 14{,}15 + 0{,}01 = 0{,}55$. Делим размах на 12 и результат округляем с недостатком. Получим 0,045. Делим теперь 0,55 на 15 и результат округляем с избытком. Получим 0,037. В качестве длины интервала можно принять число 0,05 (нечетное число единиц последнего разряда). В качестве нижней границы первого интервала возьмем число 14,15 (число 1415 кратно 5). Интервалы ряда имеют вид: 14,15—14,19; 14,20—14,24; ... . Получим 11 интервалов. Найдем частоты и относительные частоты каждого интервала.

в) Вычислим абсолютные плотности интервалов по формуле $p_i = \dfrac{n_i}{h_i}$ и относительные плотности по формуле $\pi_i = \dfrac{\nu_i}{h_i}$. Здесь n_i — частота i-го интервала, h_i — ширина i-го интервала, ν_i — относительная частота i-го интервала. Полученный ряд представлен в таблице 1.13.

Таблица 1.13

x_i	14,15—14,19	14,20—14,24	14,25—14,29	14,30—14,34	14,35—14,39	14,40—14,44
n_i	2	4	2	3	11	3
ν_i	0,04	0,08	0,04	0,06	0,22	0,06
p_i	40	80	40	60	220	60
π_i	0,8	1,6	0,8	1,2	4,4	1,2

x_i	14,45—14,49	14,50—14,54	14,55—14,60	14,60—14,64	14,65—14,69
n_i	6	10	6	1	2
ν_i	0,12	0,20	0,12	0,02	0,04
p_{i}	120	200	120	20	40
π_i	2,4	4	2,4	0,4	0,8

■

? Как получены значения 2; 0,04; 40; 0,8 для первого интервала?

Контрольные вопросы

1. Может ли накопленная частота значения превышать объем совокупности?

2. Может ли накопленная относительная частота значения быть больше 1?

3. Для какого значения накопленная частота равна объему совокупности?

4. Для какого значения накопленная относительная частота равна 1?

5. Для какого значения накопленная частота совпадает с его частотой?

6. Для какого значения накопленная относительная частота совпадает с его относительной частотой?

7. Чему равна сумма абсолютных плотностей всех интервалов, если все интервалы имеют одинаковую длину?

8. Чему равна сумма относительных плотностей всех интервалов, если все интервалы имеют одинаковую длину?

9. Почему абсолютную и относительную плотности распределения определяют для тех интервальных вариационных рядов, у которых длины интервалов отличны от 1?

Задачи

16. Ниже представлено время химической реакции (в с):

8,5	7,1	6,7	6,2	2,9	4,4	6,0	5,8	5,4	8,2	6,9
6,5	6,1	3,8	6,0	6,0	5,6	5,3	7,7	6,8	6,5	6,1
4,2	4,7	5,6	5,4	5,3	7,4	6,7	6,4	6,1	4,5	6,0
5,8	5,6	5,1								

а)° Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда вычислите абсолютные и относительные плотности распределения.

17. Получены следующие данные о плотности ткани (в условных единицах):

197,4	192,7	193,6	191,1	191,3	188,3	194,3	195,4	196,6	177,7
197,8	196,7	190,8	191,7	195,2	182,7	200,2	196,1	189,9	184,8
190,3	198,6	190,2	183,5	187,2					

а)° Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда подсчитайте абсолютные и относительные плотности распределения.

18. Получены данные о прочности 50 образцов стального троса (в условных единицах):

1800	1700	1740	1640	1780	1860	1600	1840	1880	1980
1540	1620	1860	1660	1840	1820	1840	1740	1780	1960
1840	1880	1740	1560	1740	1640	1740	1640	1660	1840
1940	1740	1800	1680	1800	1700	1640	1560	1740	1840
1740	1860	1740	1620	1740	1560	1600	1720	1840	1800

а)° Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда вычислите абсолютные и относительные плотности распределения.

19. Имеются результаты измерения содержания кремния (в %) в чугуне:

0,27	0,42	0,28	0,37	0,32	0,38	0,40	0,52	0,40	0,51	0,33	0,54	0,47
0,53	0,52	0,43	0,55	0,48	0,49	0,45	0,98	0,35	0,27	0,43	0,32	

а)° Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда вычислите абсолютные и относительные плотности распределения.

§ 1.4. Графическое изображение вариационных рядов

Графическое изображение зависимости между величинами дает возможность представить эту зависимость наглядно. Графики могут служить основой для открытия новых свойств, соотношений и закономерностей.

Наиболее употребительными графиками для изображения вариационных рядов, т. е. соотношений между значениями признака и соответствующими частотами или относительными частотами, являются полигон, гистограмма и кумулята.

Полигон чаще всего используют для изображения дискретных рядов. Для построения полигона в прямоугольной системе координат на оси абсцисс в произвольно выбранном масштабе откладывают значения аргумента, т. е. варианты, а на оси ординат также в произвольно выбранном масштабе — значения частот или относительных частот. Масштаб выбирают такой, чтобы была обеспечена необходимая наглядность, и чтобы рисунок имел желательный размер. Далее в этой системе координат строят точки, координатами которых являются пары соответствующих чисел из вариационного ряда. Полученные точки последовательно соединяют отрезками прямой. Крайнюю «левую» точку соединяют с точкой оси абсцисс, абсцисса которой находится слева от рассматриваемой точки на таком же расстоянии, как абсцисса ближайшей справа точки. Аналогично крайнюю «правую» точку также соединяют с точкой оси абсцисс.

Пример 1. Учебные достижения учащихся некоторого класса по математике характеризуются данными, представленными в таблице 1.14.

Построить полигон частот.

Таблица 1.14

Количество баллов x	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся n	1	1	2	3	4	4	6	5	3	3	2	1

❒ Строим точки (1; 1), (2; 1), (3; 2), (4; 3), (5; 4), (6; 4), (7; 6), (8; 5), (9; 3), (10; 3), (11; 2), (12; 1). Полученные точки соединяем отрезками прямой. Обратите внимание на точки (0; 0) и (13; 0), расположенные на оси абсцисс и имеющие своими абсциссами числа, на 1 меньшее и большее, чем соответственно абсциссы самой левой и самой правой точек. Полигон частот изображен на рис. 2[1]. ■

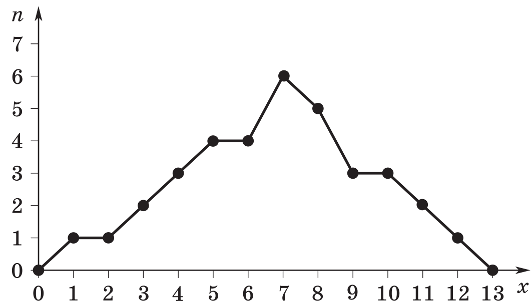


Рис. 2

Если полигон строят по данным интервального ряда, то в качестве абсцисс точек берут середины соответствующих интервалов. Крайние левую и правую точки соединяют с точками оси абсцисс — серединами ближайших интервалов, частоты которых равны ну-

[1] Здесь и далее полигоны, гистограммы и кумуляты выполнены в программе Microsoft Excel.

лю. Конечно, в этом случае полигон лишь приближенно отображает зависимость частот от значений аргумента.

Гистограмму используют для изображения интервальных рядов. Для построения гистограммы по данным вариационного ряда с равными интервалами, как и для построения полигона, на оси абсцисс откладывают значения аргумента, а на оси ординат — значения частот или относительных частот. Далее строят прямоугольники, основаниями которых служат отрезки оси абсцисс, длины которых равны длинам интервалов, а высотами — отрезки, длины которых пропорциональны частотам или относительным частотам соответствующих интервалов.

В результате получают ступенчатую фигуру в виде сдвинутых друг к другу прямоугольников, площади которых пропорциональны частотам (или относительным частотам).

Если интервалы неравные, то на оси ординат следует откладывать в произвольно выбранном масштабе значения плотности распределения (абсолютной или относительной). Таким образом, высоты прямоугольников, которые мы строим, должны равняться плотностям соответствующих интервалов.

При графическом изображении вариационного ряда с помощью гистограммы плотность изображается так, как если бы она оставалась постоянной внутри каждого интервала. На самом деле, как правило, это не так. Если построить распределение по частям интервалов, то можно убедиться в том, что плотность распределения на различных участках интервала не остается постоянной. Плотность, полученная ранее, представляла лишь некоторую среднюю плотность. Итак, гистограмма изображает не фактическое изменение плотности распределения, а лишь средние плотности распределения на каждом интервале.

Если построена гистограмма интервального распределения, то полигон того же распределения можно получить, если соединить прямолинейными отрезками середины верхних оснований прямоугольников.

Пример 2. По результатам тестирования по математике учащихся 7-го класса получены данные о доступности заданий теста (отношение числа учащихся, правильно выполнивших задания, к числу тестировавшихся учащихся), представленные в таблице 1.15.

Тест содержал 25 заданий. Построить гистограмму.

Таблица 1.15

Доступность задания x, %	25—35	35—45	45—55	55—65	65—75	75—85	85—95
Количество задач n	1	1	5	7	7	3	1

❒ Откладываем на оси абсцисс 7 отрезков длиной 10. На них, как на основаниях, строим прямоугольники, высоты которых соответственно равны 1, 1, 5, 7, 7, 3, 1. Полученная ступенчатая фигура и является искомой гистограммой (рис. 3). ■

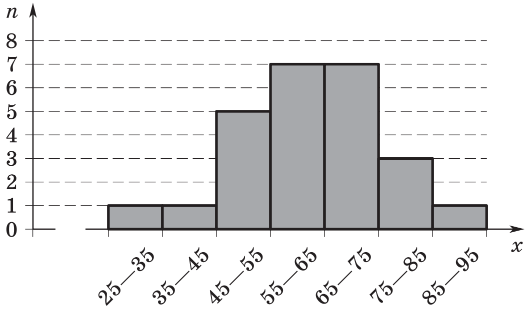


Рис. 3

Пример 3. Данные, приведенные в таблице 1.15, представлены более подробно в таблице 1.16 (интервалы имеют длину не 10, а 5). На рис. 4 построена гистограмма по этим данным. Получено изображение более подробной информации о распределении данных.

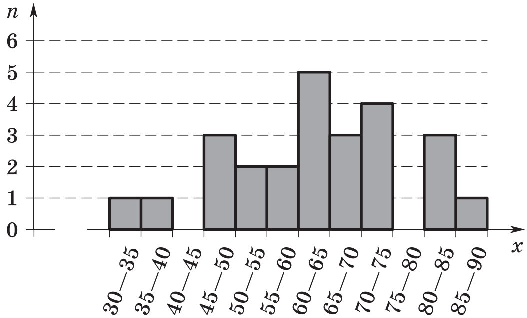


Рис. 4

Таблица 1.16

Доступность задачи x, %	30—35	35—40	40—45	45—50	50—55	55—60
Количество задач n	1	1	0	3	2	2
Доступность задачи x, %	60—65	65—70	70—75	75—80	80—85	85—90
Количество задач n	5	3	4	0	3	1

■

? По гистограммам, изображенным на рис. 3 и 4, постройте полигоны частот.

Вычислите площади гистограмм.

Кумулята служит для графического изображения кумулятивного вариационного ряда. Для ее построения на оси абсцисс откладывают значения аргумента, а на оси ординат — накопленные частоты или накопленные относительные частоты. Масштаб на каждой оси выбирают произвольно. Далее строят точки, абсциссы которых равны вариантам (в случае дискретных рядов) или верхним границам интервалов (в случае интервальных рядов), а ординаты — соответствующим частотам (накопленным частотам). Эти точки соединяют отрезками прямой. Полученная ломаная и является кумулятой.

? Если кумуляту строят по накопленным относительным частотам, то чему равна ордината крайней правой точки?

По данным, приведенным в таблице 1.14, составим кумулятивный вариационный ряд (табл. 1.17), для которого построим кумуляту (рис. 5).

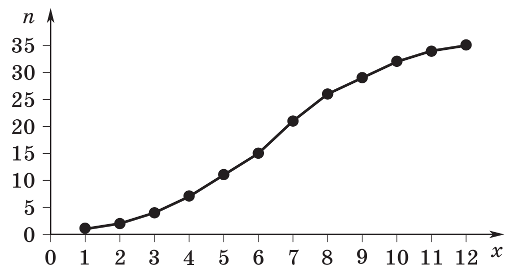


Рис. 5

Таблица 1.17

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Частота	1	1	2	3	4	4	6	5	3	3	2	1
Накопленная частота n	1	2	4	7	11	15	21	26	29	32	34	35

Контрольные вопросы

1. На рис. 6 изображен полигон частот, построенный по данным о распределении рабочих цеха по тарифным разрядам. По оси абсцисс отложены значения тарифных разрядов x, по оси ординат — число рабочих n.

а) Каковы наименьший и наибольший тарифные разряды рабочих?

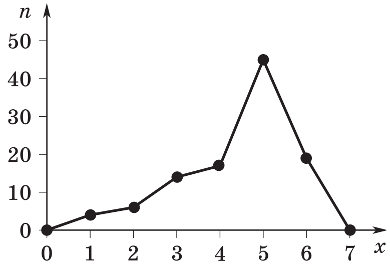


Рис. 6

б) Какой тарифный разряд имеет наибольшее число рабочих?

в) Какой разряд — четвертый или шестой — имеет большее число рабочих?

г) Сколько примерно рабочих работает в цехе?

д) Каков приближенно дискретный вариационный ряд, по которому построен полигон?

2. На рис. 7 изображена гистограмма, построенная по данным о распределении образцов волокон хлопка по прочности. По оси абсцисс отложены пределы прочности x, по оси ординат — число образцов m.

а) В каких пределах заключены наименьшая и наибольшая прочности волокон хлопка?

б) В каких пределах находится прочность волокон, которую имеет наибольшее число волокон?

в) Какую прочность — от 3 до 4 или от 5 до 6 — имеет большее число волокон?

г) Сколько примерно волокон исследовалось на прочность?

д) Каков приближенно интервальный вариационный ряд, по которому построена гистограмма?

3. На рис. 8 изображена кумулята, построенная по данным о распределении образцов волокон хлопка по прочности. По оси абсцисс указаны пределы прочности x, по оси ординат — накопленные частоты образцов S.

а) Сколько примерно волокон исследовалось на прочность?

б) Сколько примерно волокон имеет прочность от 7 до 8?

в) В каких пределах заключены наименьшая и наибольшая прочности волокон хлопка?

4. Как, имея гистограмму, построить полигон того же распределения?

5. Как меняется гистограмма и полигон при увеличении ширины интервалов?

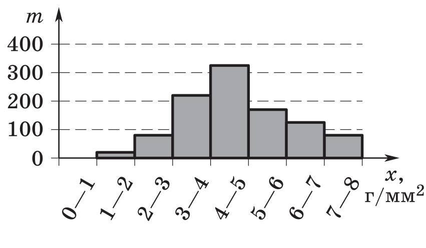


Рис. 7

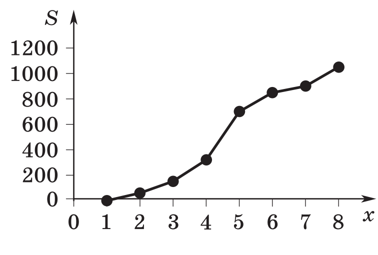


Рис. 8

Задачи

20.° Постройте полигоны частот для дискретных вариационных рядов, составленных в заданиях 16—19 § 1.3.

21.° Постройте гистограммы для интервальных вариационных рядов, составленных в заданиях 16—19 § 1.3.

22. В таблице 1.18 приведено распределение числа взрослых рабочих-мужчин цеха по росту.

а)° Постройте по этим данным гистограмму и полигон частот.

б) Опишите форму этого распределения. В частности, укажите, является это распределение симметричным или нет.

Таблица 1.18

Рост, см	143—146	146—149	149—152	152—155	155—158	158—161	161—164	164—167
Число мужчин	1	2	8	26	65	120	181	201
Рост, см	167—170	170—173	173—176	176—179	179—182	182—185	185—188	Всего
Число мужчин	170	120	64	28	10	3	1	1000

23.° В таблице 1.19 приведены данные о массе новорожденных детей при рождении.

Таблица 1.19

Масса, г	1000—1500	1500—2000	2000—2500	2500—3000	3000—3500
Число детей	84	205	502	1723	3752
Масса, г	3500—4000	4000—4500	4500—5000	5000—5500	Всего
Число детей	2747	852	124	11	10 000

а)° Вычислите относительные частоты и накопленные относительные частоты.

б) Постройте по этим данным гистограмму и кумулятивную кривую.

24.° Получены результаты измерения отклонений диаметров детали от номинального размера (в мк):

42	46	41	39	42	40	43	40	41	44
42	42	41	40	42	42	41	43	39	40

а) Постройте дискретный вариационный ряд.

б) Постройте полигон частот.

25. Рассмотрите данные, представленные в таблице 1.2 § 1.1, о численности населения республик бывшего СССР и их логарифмы, представленные в таблице 1.3. Постройте гистограммы по тем и другим данным. Чем они отличаются?

§ 1.5. Среднее арифметическое — показатель центральной тенденции

В результате исследований, связанных с массовыми явлениями, получают много числовых данных. Возникает проблема — найти такие характеристики, которые довольно полно характеризовали бы полученный числовой материал. Характеристики, которые базируются на данных массовых наблюдений, называют ***обобщающими показателями***. Эти показатели характеризуют значения признака, его вариацию. Их вычисляют с помощью вариант и соответствующих частот (относительных частот). Важнейшие среди обобщающих показателей — ***средние величины***, т. е. такие значения признака, вокруг которых группируются отдельные наблюдаемые значения элементов. Отсюда и название — ***меры центральной тенденции***.

В зависимости от характера задачи пользуются тем или иным видом средней величины. К ним принадлежат среднее арифметическое, мода, медиана, степенные средние (среднее гармоническое, среднее геометрическое и т. п.).

Изучая и используя обобщающие показатели, следует иметь в виду, что они только тогда объективно будут соответствовать своему назначению, если применяются к однородным совокупностям. В противном случае можно получить неправильные выводы. Например, едва ли правильно характеризовать средние учебные достижения учащихся одного региона, вычисленные по данным совокупности, к которой относятся наряду с учащимися элитных учебных заведений (лицеев, гимназий и т. п.) ученики общеобразовательных школ, специализированных школ для умственно отсталых детей и др.

Неправильное использование средних показателей приводит к тому, что в ряде случаев за «благополучным» показателем успешности скрываются проблемы, связанные с обучением отдельных категорий учащихся.

В настоящем параграфе будет рассмотрено среднее арифметическое и его свойства.

С понятием среднего арифметического вы уже знакомы с младших классов. Напомним его. Пусть имеется n объектов, для которых измерена некоторая характеристика, и получены значения x_1,

$x_2, \ldots, x_n$. ***Среднее арифметическое*** этих n значений обозначают через $\bar{x}$ и определяют как $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$, или $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Символом $\sum\limits_{i=1}^{n} a_i$ с переменным индексом i (читается так: «сумма по i от 1 до m») обозначают следующую сумму: $\sum\limits_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n$.

Заметим, что выражение $\sum\limits_{k=1}^{n} a_k$ означает то же самое, что и $\sum\limits_{i=1}^{n} a_i$, т. е. переменный индекс можно обозначать любой буквой.

Это сокращенное обозначение суммы обладает следующими простыми свойствами, вытекающими из известных свойств сложения:

1. $\sum\limits_{i=1}^{n} a = \underbrace{a + a + \ldots + a}_{n \text{ слагаемых}} = na.$

2. $\sum\limits_{i=1}^{n} ca_i = c\sum\limits_{i=1}^{n} a_i$, так как общий множитель можно выносить за знак суммы.

3. $\sum\limits_{i=1}^{n} a_i = \sum\limits_{i=1}^{m} a_i + \sum\limits_{i=m+1}^{n} a_i$ $(m < n)$, так как для сложения справедливо сочетательное свойство.

4. $\sum\limits_{i=1}^{n} (a_i + b_i) = \sum\limits_{i=1}^{n} a_i + \sum\limits_{i=1}^{n} b_i$ на основе одновременного применения сочетательного и переместительного свойств сложения.

Сущность среднего арифметического состоит в следующем. Если каждое наблюдение заменить средним, то общая сумма не изменится. Это среднее можно интерпретировать еще и так: если все наблюдения будут равны между собой, а сумма наблюдений останется неизменной, то каждое наблюдение будет равно среднему. Поскольку среднее сохраняет неизменной сумму при равномерном распределении значений, то оно наиболее полезно в качестве обобщающего показателя при отсутствии резко выделяющихся наблюдений, или как их называют, выбросов, т. е. когда набор данных представляет собой более менее однородную группу.

Пример 1. Рассмотрим среднюю месячную зарплату работников некоторого предприятия. Пусть, например, в фирме работает 20 человек, зарплата 19 из них составляет 10 000 р., а зарплата 20-го, руководителя, — 1 000 000 р. Тогда средняя зарплата одного работника на этой фирме будет равна $\frac{19 \cdot 10\,000 + 1\,000\,000}{20} =$

= 59 500. Хотя среднее и сохранило общую сумму заработной платы, но оно является в данном случае плохим обобщающим показателем: оно плохо характеризует зарплату одного работника на этой фирме. Причина этого кроется в том, что набор данных содержит выброс — 1 000 000 р. Среднее оказалось слишком большим для большинства работников и слишком малым для высокооплачиваемого руководителя. ■

Среднее арифметическое, как указывалось выше, является обобщающим показателем, сохраняющим общую сумму при замене на него каждого значения. Это свойство особенно полезно в тех ситуациях, когда необходимо планировать общую сумму для большой группы. Сначала вычисляют среднее арифметическое для меньшей выборки данных, хорошо представляющей большую группу. Затем полученное значение умножают на количество элементов в большой группе. В результате получают приближенное значение суммы для всей группы. С помощью среднего арифметического решается одна из задач статистики — *оценка неизвестного параметра совокупности*.

Пусть дан дискретный вариационный ряд:

x_1	x_2	...	x_m
n_1	n_2	...	n_m

Тогда среднее арифметическое вычисляют с учетом частот следующим образом:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_m n_m}{n_1 + n_2 + \ldots + n_m}, \text{ или } \bar{x} = \frac{\sum_{i=1}^{m} x_i n_i}{\sum_{i=1}^{m} n_i}, \text{ или } \bar{x} = \sum_{i=1}^{m} x_i \frac{n_i}{n}.$$

Здесь m — количество различных значений, которые принимает признак. Такую форму среднего арифметического иногда называют ***средним взвешенным***.

Среднее взвешенное можно интерпретировать как среднюю величину для значений $x_1, x_2, \ldots, x_m$, используемую в ситуациях, когда одни значения более важны по сравнению с другими. Более важные значения вносят больший вклад в значение среднего взвешенного. Роль весов играют отношения $\frac{n_i}{n}$: чем больше частота элемента, тем больший вклад вносит этот элемент в значение среднего взвешенного. Сумма всех весов равна 1.

Пример 2. Вычислить среднее арифметическое по данным о доступности заданий теста по математике для 7-го класса. Эти данные приведены в таблице 1.20.

Таблица 1.20

№ задания	1	2	3	4	5	6	7	8	9
Доступность, %	69	81	62	59	71	70	52	61	61
№ задания	10	11	12	13	14	15	16	17	18
Доступность, %	69	73	72	60	31	36	63	87	80
№ задания	19	20	21	22	23	24	25		
Доступность, %	47	82	48	50	56	49	66		

❒ Сумма вариант равна: $69 + 81 + 62 + 59 + 71 + 70 + 52 + 61 + 61 + 69 + 73 + 72 + 60 + 31 + 36 + 63 + 87 + 80 + 47 + 82 + 48 + 50 + 56 + 49 + 66 = 1555$. Разделив это число на количество задач (25), получим $\bar{x} = 62{,}2$. Итак, средняя доступность одного задания равна 62,2 %. ■

Пример 3. В таблице 1.21 представлены данные о количестве баллов, которые набрали на олимпиаде представители одного района. Вычислить по этим данным среднее арифметическое.

Таблица 1.21

Варианта x	2	3	5	6	8	9	10	11	15	18
Частота n	1	2	4	3	2	2	2	3	1	1

❒ Проведем вычисления по схеме, представленной в таблице 1.22.

Таблица 1.22

№	1	2	3	4	5	6	7	8	9	10	Σ
x_i	2	3	5	6	8	9	10	11	15	18	
n_i	1	2	4	3	2	2	2	3	1	1	21
$x_i n_i$	2	6	20	18	16	18	20	33	15	18	166

$$\bar{x} = \frac{166}{21} \approx 7{,}9.$$

В среднем, один представитель района набрал примерно 7,9 балла. ■

Также вычисляют среднее арифметическое по данным интервального вариационного ряда. За значение признака для всех элементов в данном интервале берут середину интервала. При этом допускается определенная неточность, но обычно в различных интервалах погрешности будут разных знаков, а потому при большом количестве наблюдений они в значительной мере «гасят» друг друга.

Пример 4. Вычислить среднее арифметическое результатов контрольной работы по математике, проведенной в 9-х классах школ некоторой области. Результаты представлены в первых двух столбцах таблицы 1.23 (работа оценивалась по 12-балльной шкале).

Таблица 1.23

Количество баллов	Число учащихся n_i	x_i	$x_i n_i$
1—3	26	2	52
4—6	478	5	2390
7—9	369	8	2952
10—12	127	11	1397
Σ	1000		6791

❒ В последних двух столбцах таблицы 1.23 указаны середины интервалов и их произведения на частоты соответствующих интервалов. В результате получим

$$\bar{x} = \frac{6791}{1000} \approx 6{,}8.$$

В среднем, контрольная работа одного учащегося оценена примерно на 6,8 балла. ■

Пример 5. Социологическая лаборатория интересуется, сколько тратят жители города N в месяц на молочные продукты. Население этого города составляет 160 500 человек. Опросить такое количество жителей трудоемко. Опросили 400 человек. Оказалось, что в среднем каждый из них в месяц на молочные продукты тратит 470 р. Оценить, т. е. указать приближенно, сколько тратят жители города N в месяц на молочные продукты.

❒ Чтобы найти требуемую оценку, нужно среднее значение расходов одного человека из выборки умножить на численность населения города N: $470 \cdot 160\,500 = 75\,435\,000$ р. Этот прогноз является приемлемым (если выборка была представительной; см. подробнее об этом в главе 5) и полезным. Но это значение не является точным. ■

Рассмотрим некоторые свойства среднего арифметического, которые позволяют упростить его вычисление и которые понадобятся при дальнейшем изучении математической статистики.

СВОЙСТВО 1. *Среднее арифметическое постоянной величины равно этой постоянной.*

❒ Пусть при исследовании признака x он n раз принимал одно и то же значение c. Тогда

$$\bar{x} = \frac{c + c + \ldots + c}{n} = \frac{nc}{n} = c. \blacksquare$$

СВОЙСТВО 2. *Если каждое значение признака Z равно сумме (разности) значений признаков X и Y, то среднее арифметическое признака Z равно сумме (разности) средних арифметических признаков X и Y.*

❒ Обозначим i-е варианты признаков X, Y, Z через x_i, y_i, z_i. По условию $x_i + y_i = z_i$. Тогда

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}(x_i + y_i) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{x} + \bar{y}.$$

Аналогично доказывается свойство и в случае разности. ■

?

Проведите доказательство этого свойства в случае разности.

Верно ли, что средняя зарплата работника цеха за два месяца равна сумме средних зарплат работника этого цеха за каждый из этих двух месяцев?

Например, из этого свойства вытекает, что если контрольная работа по геометрии состоит из двух сюжетных задач, то среднее время, которое идет на выполнение контрольной работы, равно сумме средних времен, которые расходуются на выполнение первой и второй задач.

СВОЙСТВО 3. *Если ко всем вариантам прибавить одно и то же число, то и к среднему арифметическому будет прибавлено то же число.*

❒ Пусть $x'_i = x_i + c$ — новые варианты, полученные после прибавления к каждой первоначальной варианте x_i одного и того же числа c. Тогда

$$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n} x'_i = \frac{1}{n}\sum_{i=1}^{n}(x_i + c) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} c =$$

$$= \bar{x} + \frac{nc}{n} = \bar{x} + c. \blacksquare$$

Рассмотренное свойство позволяет значительно упростить вычисление среднего арифметического без использования вычислительных средств, особенно тогда, когда варианты принимают большие значения.

Это свойство обосновывает произвольный выбор начала отсчета.

? Верно ли, что если зарплату всех работников некоторой фирмы увеличить на одну и ту же сумму, то и средняя зарплата работника этой фирмы увеличится на ту же сумму?

СВОЙСТВО 4. *Если все варианты умножить (разделить) на одно и то же число, то среднее арифметическое умножится (разделится) на то же число.*

❐ Пусть $x_i' = x_i \cdot c$ — новые варианты, полученные после умножения каждой первоначальной варианты x_i на одно и то же число c. Тогда

$$\bar{x}' = \frac{1}{n}\sum_{i=1}^{n} x_i' = \frac{1}{n}\sum_{i=1}^{n}(x_i \cdot c) = c \cdot \frac{1}{n}\sum_{i=1}^{n} x_i = c \cdot \bar{x}\,. \blacksquare$$

На основании этого свойства можно изменять единицы, в которых выражаются данные.

? Как изменится средняя зарплата в цехе, если заработная плата каждого работника будет удвоена?

СВОЙСТВО 5. *Если все частоты умножить (разделить) на одно и то же число, то среднее арифметическое не изменится.*

❐ Пусть $n_i' = n_i \cdot c$ — новые частоты, полученные после умножения каждой первоначальной частоты n_i на одно и то же число c. Тогда

$$\bar{x}' = \frac{\sum_{i=1}^{m} x_i \cdot (n_i \cdot c)}{\sum_{i=1}^{m} (n_i \cdot c)} = \frac{c\sum_{i=1}^{m} (x_i \cdot n_i)}{c\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} (x_i \cdot n_i)}{\sum_{i=1}^{m} n_i} = \bar{x}\,. \blacksquare$$

На основании этого свойства при вычислении среднего частоты можно заменять, например, относительными частотами.

? В фирме несколько отделов. Все сотрудники одного отдела получают одну и ту же зарплату. Как изменится средняя зарплата одного работника фирмы, если численность работников в каждом отделе удвоить?

СВОЙСТВО 6. *Сумма отклонений вариант от их среднего арифметического равна нулю.*

❐ Отклонение варианты x_i от среднего арифметического $\bar{x}$ равно разности $x_i - \bar{x}$. Тогда

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x} = n\bar{x} - n\bar{x} = 0. \blacksquare$$

СВОЙСТВО 7. *Сумма квадратов отклонений вариант от их среднего арифметического меньше суммы квадратов отклонений вариант от произвольного числа с на величину* $(c - \bar{x})^2 n$.

❐ В самом деле,

$$\sum_{i=1}^{n}(x_i - c)^2 - \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}((x_i - c)^2 - (x_i - \bar{x})^2) =$$

$$= \sum_{i=1}^{n}(x_i^2 - 2cx_i + c^2 - x_i^2 + 2x_i\bar{x} - \bar{x}^2) =$$

$$= -2cn\bar{x} + nc^2 + 2n\bar{x}^2 - n\bar{x}^2 = n(\bar{x}^2 - 2c\bar{x} + c^2) = n(c - \bar{x})^2.$$

Разность оказалась положительной (при $\bar{x} \neq c$), поэтому сумма $\sum_{i=1}^{n}(x_i - c)^2$ больше суммы $\sum_{i=1}^{n}(x_i - \bar{x})^2$. ■

? Что больше: сумма квадратов отклонений всех значений совокупности от среднего арифметического этих значений или сумма квадратов отклонений всех значений этой совокупности от 1?

СВОЙСТВО 8. *Среднее арифметическое, вычисленное по данным всех элементов совокупности, равно взвешенному среднему для так называемых частичных средних, т. е. средних, найденных для отдельных частей совокупности, причем частота для каждого частичного среднего равна количеству элементов в соответствующей части совокупности.*

❐ Пусть совокупность состоит из таких элементов:

$$x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_l, z_1, z_2, \ldots, z_m,$$

причем $k + l + m = n$.

Поскольку частичные средние соответственно равны

$$\bar{x} = \frac{1}{k}\sum_{i=1}^{k} x_i; \quad \bar{y} = \frac{1}{l}\sum_{i=1}^{l} y_i; \quad \bar{z} = \frac{1}{m}\sum_{i=1}^{m} z_i,$$

то общее среднее равно

$$\frac{\sum_{i=1}^{k} x_i + \sum_{i=1}^{l} y_i + \sum_{i=1}^{m} z_i}{k+l+m} = \frac{k\bar{x} + l\bar{y} + m\bar{z}}{n}. \ \blacksquare$$

Например, это свойство дает возможность упростить вычисление среднего арифметического результатов тестирования учащихся классов одной параллели нескольких школ. Для этого достаточно вычислить среднее прифметическое для каждого класса, а затем вычислить среднее этих частичных средних, приняв в качестве их частот количество учащихся в соответствующих классах.

? Как можно вычислить среднюю зарплату работников трех цехов, если известны средние зарплаты и число работников в каждом цехе?

В примере 5 было показано, как среднее арифметическое можно использовать для оценки неизвестных параметров совокупности. Кроме того, среднее арифметическое позволяет решать задачи, связанные с проверкой гипотез.

Пример 6. Два стрелка сделали по 100 выстрелов. Первый выбил 8 очков 40 раз, 9 очков — 10 раз и 10 очков — 50 раз. Второй выбил 8, 9 и 10 очков соответственно — 10, 60 и 30 раз. Какой из стрелков стреляет лучше?

❒ Вычислим средние арифметические $\bar{x}$ и $\bar{y}$ числа очков, которые выбил при 100 выстрелах каждый из двух стрелков.

$$\bar{x} = \frac{8 \cdot 40 + 9 \cdot 10 + 10 \cdot 50}{100} = 9{,}1;$$

$$\bar{y} = \frac{8 \cdot 10 + 9 \cdot 60 + 10 \cdot 30}{100} = 9{,}2.$$

Среднее число очков, которое выбивает из 100 выстрелов второй стрелок, несколько выше, чем тот же показатель у первого стрелка. Естественно признать второго стрелка лучшим. ■

Пример 7. В конце XVIII века в американской газете «The Federalist» была опубликована серия статей, подписанных псевдонимом «Публий». Было установлено, что их авторами могли быть Александр Гамильтон и Джеймс Медисон, однако, кто именно, неизвестно. Позиции, которых придерживались в своих статьях Гамильтон и Медисон, не давали ключа к разгадке тайны. Не помогал и анализ общего стиля письма обоих авторов. Единственное, за что можно зацепиться, так это частота, с которой встречаются те или иные предлоги. В таблице 1.24 представлено число статей, в которых предлог «by» встречается с той или иной частотой (в расчете на 1000 слов) для спорных статей и статей А. Гамильтона и Д. Медисона.

Таблица 1.24

Частота предлога «by»	1—3	3—5	5—7	7—9	9—11	11—13	13—15	15—17	17—19
Статьи Гамильтона	2	9	12	18	4	5			
Статьи Медисона			5	7	8	16	6	5	3
Спорные статьи			2	1	2	4	2	1	

Подсчитать средние арифметические для частот предлога «by» в спорных статьях, в статьях Гамильтона и Медисона и высказать свои соображения по поводу авторства спорных статей.

❒ Обозначим средние частоты предлога «by» в спорных статьях, в статьях Гамильтона и Медисона соответственно через $\bar{x}$, $\bar{y}$, $\bar{z}$. Будем иметь:

$$\bar{x} = \frac{6 \cdot 2 + 8 \cdot 1 + 10 \cdot 2 + 12 \cdot 4 + 14 \cdot 2 + 16 \cdot 1}{12} = 11;$$

$$\bar{y} = \frac{2 \cdot 2 + 4 \cdot 9 + 6 \cdot 12 + 8 \cdot 18 + 10 \cdot 4 + 12 \cdot 5}{50} = 7{,}12;$$

$$\bar{z} = \frac{6 \cdot 5 + 8 \cdot 7 + 10 \cdot 8 + 12 \cdot 16 + 14 \cdot 6 + 16 \cdot 5 + 18 \cdot 3}{50} = 11{,}9.$$

Средняя частота предлога «by» в спорных статьях (11) значительно ближе к средней частоте этого предлога в статьях Медисона (11,9), чем в статьях Гамильтона (7,12). Это говорит в пользу того, что автором спорных статей является Медисон. В пользу того же вывода говорит и сравнение гистограмм, построенных по приведенным дан-

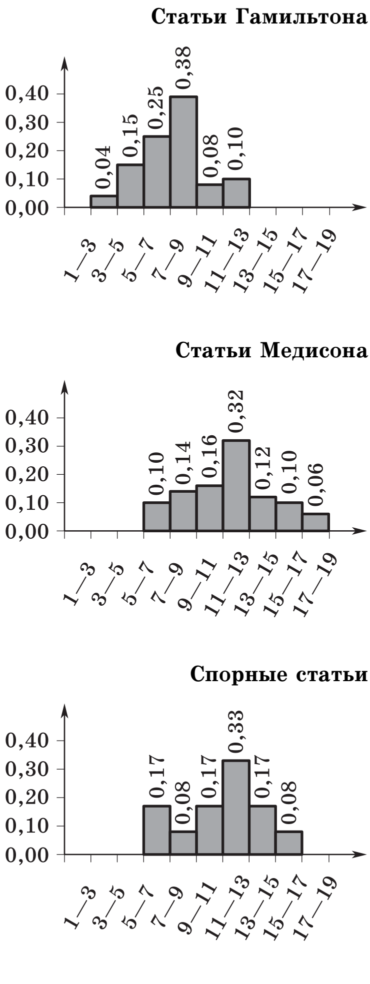


Рис. 9

ным (рис. 9). Распределения показывают, что употребление предлога «by» в спорных статьях в большей степени напоминает соответствующее распределение Медисона, а не Гамильтона. ■

Для вычисления среднего арифметического с помощью электронных таблиц Excel можно использовать функцию (СРЗНАЧ). Выбирают ячейку, куда помещают среднее. В главном меню выбирают (Вставка ⇒ функция), затем в качестве категории функции выбирают (Статистические) и в качестве названия функции — (СРЗНАЧ). Появится диалоговое окно. Перетаскивая курсор мыши, выделяют необходимый список чисел, а затем нажимают клавишу ⟨Enter⟩ для завершения процесса.

Чтобы найти среднее взвешенное с помощью Excel, используем команду (Вставка ⇒ Имя ⇒ Создать) и кнопку ОК. Далее используем функцию (СУММПРОИЗВ), которая умножает значения на соответствующие частоты и складывает результаты. Полученный результат делим на сумму частот, в итоге имеем среднее взвешенное.

Контрольные вопросы

1. Какой смысл имеет среднее арифметическое с точки зрения суммы всех значений набора данных?

2. Что нужно знать, кроме среднего числа отказов прибора, чтобы оценить общее количество отказов приборов?

3. Как используют среднее арифметическое для оценки суммы значений большой совокупности?

4. Достаточно ли знать среднее количество бракованных изделий, выпускаемых рабочим в каждом из трех цехов предприятия, чтобы найти среднее число бракованных изделий, выпускаемых рабочим на всем предприятии, если на этом предприятии три цеха?

5. Как, имея данные о количестве заказов, поступающих в две мастерские ежедневно в течение месяца, сравнить загруженность этих мастерских работой?

6. Как учитываются резко выделяющиеся значения совокупности при вычислении среднего арифметического?

7. Почему средняя зарплата одного работающего в Москве плохо характеризует состояние с заработной платой в столице?

8. Почему бессмысленным считается выражение «средняя температура больного по палате»?

9. Есть ли смысл находить среднее арифметическое для порядковых данных?

10. Требуется выяснить потребность населения некоторого города в определенном товаре. Как может помочь среднее арифметическое в решении этой проблемы?

Задачи

26. Число бракованных изделий на некотором производстве в течение 20 дней равнялось

24	28	5	10	22	5	19	0	3	18
20	4	0	1	21	17	10	20	4	7

а)° Найдите среднее арифметическое дневного выпуска бракованных изделий.

б) Если бы регистрация числа бракованных изделий продолжалась еще три дня и если бы характеристика числа бракованных изделий была такой же, как и в остальные дни, то какое число бракованных изделий можно было бы ожидать за 23 дня?

27. В таблице 1.25 приведены данные о размере последних заказов потребителей фирмы (в десятках тыс. р.).

а)° Найдите средний размер заказа.

б) Если бы к приведенным данным о 24 последних заказах потребителей фирмы добавили еще 6 предыдущих заказов и если бы характеристика размера этих заказов была такой же, как и у последующих 24 заказов, то каким бы примерно был суммарный размер последних 30 заказов?

Таблица 1.25

Размер заказа, десятки тыс. р.	1—4	5—8	9—12	13—16	16—19	20—23	24—27	28—31
Число заказов	10	2	4	1	2	1	1	3

28. В таблице 1.26 представлено число статей, в которых предлог «on» встречается с той или иной частотой (в расчете на 1000 слов) для спорных статей и статей А. Гамильтона и Д. Медисона (см. пример 7).

а) Подсчитайте средние арифметические для частот предлога «on» в спорных статьях, в статьях Гамильтона и Медисона и выскажите свои соображения по поводу авторства спорных статей.

б) Постройте гистограммы по приведенным данным, сравните их.

Таблица 1.26

Частота предлога «on»	0,4	0,4—0,8	0,8—1,2	1,2—1,6	1,6—2,0	2—3
Статьи Гамильтона			2	3	6	11
Статьи Медисона	41	2	4	1	2	
Спорные статьи	11			1		

Частота предлога «on»	3—4	4—5	5—6	6—7	7—8
Статьи Гамильтона	11	10	3	1	1
Статьи Медисона					
Спорные статьи					

29°. Контрольная работа десяти учащихся проверялась двумя преподавателями и оценивалась ими по 12-балльной шкале. Результаты оценивания представлены в таблице 1.27.

Какой из преподавателей строже?

Таблица 1.27

№ учащегося	1	2	3	4	5	6	7	8	9	10
1-й преподаватель	7	11	6	8	3	10	8	5	7	10
2-й преподаватель	5	12	6	7	2	11	9	3	6	10

30. В ряде американских университетов средний балл результатов обучения студента вычисляют как взвешенное среднее. Это связано с тем, что некоторые курсы оцениваются бо́льшим количеством баллов по сравнению с другими. Пусть система оценок в некотором университете включает оценки от 1,0 (незачет) до 5,0 (отлично). Вес каждой оценки пропорционален количеству баллов, присвоенных курсу. Карточка студента N имеет вид таблицы 1.28.

а) Найдите среднюю оценку студента N.

б) Оценка по какому предмету внесла наибольший вклад в среднюю оценку?

в) Существенно ли повлияла на среднюю оценку оценка по спецкурсу?

Таблица 1.28

Курс	**Баллы**	**Оценка**
Статистика	6	4,5
Экономика	6	4,3
Математика	5	4,7
Маркетинг	5	3,8
Спецкурс	2	2,2
Итого	24	

31. В двух бригадах насчитывается по 7 человек. В первой бригаде месячная зарплата двух рабочих составляет по 12 600 р., трех рабочих — по 14 400 р., один рабочий зарабатывает 15 300 р. и один рабочий — 16 200 р. Во второй бригаде трое рабочих зарабатывают по 12 600 р., один — 14 580 р., двое — по 15 300 р. и один — 15 390 р. В какой бригаде выше оплата труда?

32. В результате испытаний двух однотипных приборов А и В установили число помех, которые оценивались по трехбалльной системе (табл. 1.29).

Меньшему уровню помех соответствует меньшее количество баллов. В случае отсутствия помех их уровень считался равным нулю. Какой из приборов более чувствителен к помехам?

Таблица 1.29

Уровень помех в баллах		1	2	3
Число появлений помех данного уровня в 100 испытаниях приборов	A	20	6	4
	B	7	3	10

§ 1.6. Другие меры центральной тенденции

В предыдущем параграфе была рассмотрена наиболее широко используемая мера центральной тенденции — среднее арифметическое. Оно применяется в том случае, когда количественные данные имеют содержательный смысл. Кроме среднего арифметического мерой центральной тенденции может служить:

1) *медиана*, или *срединная точка*, которую можно вычислять как для порядковых, так и для количественных данных;

2) *мода* — наиболее часто встречающаяся категория, которую можно вычислять для номинальных данных, упорядоченных категорий и количественных данных.

1.6.1. Медиана

В предыдущем параграфе при рассмотрении примера 1 мы уже видели, что среднее арифметическое плохо характеризует состояние с заработной платой на предприятии. Средний арифметический заработок оказался слишком высоким для работников предприятия. Более правильную картину даст то значение, которое делит данные на две равные части. Таким значением является медиана.

Если все элементы совокупности размещены в порядке возрастания или убывания числовых значений признака, то ***медиана*** *— это такое значение признака, которое делит всю совокупность пополам.*

Итак, количество элементов совокупности, имеющих значение признака, меньшее медианы, равно количеству элементов со значением признака, большим медианы. Будем обозначать медиану символом Me.

При нахождении медианы дискретного вариационного ряда следует различать два случая:

1) объем совокупности нечетный;

2) объем совокупности четный.

Если объем совокупности нечетный и равен $2n + 1$, и варианты размещены в порядке возрастания их значений:

$$\underbrace{x_1, x_2, \ldots, x_n}_{n\ \text{значений}}, x_{n+1}, \underbrace{x_{n+2}, \ldots, x_{2n+1}}_{n\ \text{значений}},$$

то $\mathrm{Me} = x_{n+1}$.

Если же количество элементов четное и равно $2n$, то нет варианты, которая бы делила совокупность на две равные по объему части:

$$\underbrace{x_1, x_2, \ldots, x_n}_{n\ \text{значений}}, \underbrace{x_{n+1}, \ldots, x_{2n}}_{n\ \text{значений}}.$$

Поэтому в качестве медианы условно берется полусумма вариант, находящихся в середине вариационного ряда:

$$\mathrm{Me} = \frac{x_n + x_{n+1}}{2}.$$

Пример 1. Вычислить медиану по данным таблицы 1.30, в которой приведена информация об успеваемости по математике 100 учащихся 7-х классов (успеваемость оценивается по 12-балльной шкале).

Таблица 1.30

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	3	4	4	9	11	12	18	14	9	8	6	2

❐ Поскольку совокупность содержит 100 элементов (учащихся), упорядоченных по значению признака, и количество элементов четно, то надо найти полусумму числовых значений 50-го и 51-го элементов. Складывая последовательно частоты, находим накопленные частоты (табл. 1.31).

Таблица 1.31

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Накопленные частоты	3	7	11	20	31	43	61	75	84	92	98	100

Устанавливаем первую накопленную частоту, большую половины общего количества элементов. В данном примере это 61. Итак, и 50-му и 51-му элементам отвечает значение 7, поэтому и Me = 7.

Полученное значение медианы означает, что примерно половина семиклассников по математике учатся на 7 и меньше баллов, а половина — на 7 и больше баллов. ■

Обращаем внимание на ошибку, часто встречающуюся при вычислении медианы. Иногда не учитывают ни частоты вариант, ни общего количества элементов и в качестве медианы берут полусумму средних вариант. В примере 1 это полусумма 6-й и 7-й вариант, т. е. 6,5, что не верно.

Рассмотрим вычисление медианы интервального упорядоченного вариационного ряда. Интервал, в котором находится медиана, назовем ***медианным***. Вывод формулы для вычисления медианы базируется на предположении, что плотность распределения признака на медианном интервале является постоянной.

❒ Введем обозначения:

x_n — начало медианного интервала;

h — ширина медианного интервала;

n_{Me} — частота медианного интервала;

$S_{\text{Me}-1}$ — сумма частот интервалов, предшествующих медианному;

n — объем совокупности;

$\frac{n}{2}$ — накопленная частота до значения медианы;

$\frac{n}{2} - S_{\text{Me}-1}$ — частота интервала от x_n до Me, ширина которого равна $\text{Me} - x_n$.

Абсолютная плотность распределения на медианном интервале равна $\frac{n_{\text{Me}}}{h}$. Абсолютная плотность распределения на интервале от x_n до Me равна $\frac{\frac{n}{2} - S_{\text{Me}-1}}{\text{Me} - x_n}$. По условию эти плотности равны друг другу, поэтому

$$\frac{n_{\text{Me}}}{h} = \frac{\frac{n}{2} - S_{\text{Me}-1}}{\text{Me} - x_n}.$$

Отсюда

$$\mathrm{Me} = x_n + \frac{\frac{n}{2} - S_{\mathrm{Me}-1}}{n_{\mathrm{Me}}} h.$$

Если плотность распределения на медианном интервале не является постоянной, то полученная формула будет приближенной. ■

Пример 2. Вычислить медиану по данным таблицы 1.23 из предыдущего параграфа.

❐ Из таблицы имеем:

$$x_n = 4,\ h = 6 - 4 = 2,\ n_{\mathrm{Me}} = 478,\ S_{\mathrm{Me}-1} = 26,\ n = 1000;$$

$$\mathrm{Me} = 4 + \frac{\frac{1000}{2} - 26}{478} \cdot 2 \approx 5{,}98.$$

Полученный результат означает, что примерно половина девятиклассников области написали контрольную работу на 6 и меньше баллов, а половина — на 6 и больше баллов. ■

? Как в примере 2 установлен медианный интервал?

Медиана обладает важными свойствами, которые в некоторых случаях дают ей преимущество перед другими средними величинами. Например, если при упорядоченном размещении некоторого признака «крайние» значения сомнительные и к тому же резко отличаются от основной массы данных, то в качестве меры центральной тенденции целесообразно использовать медиану, поскольку на ее величину эти «крайние» значения никакого влияния не оказывают, и в то же время они могут существенным образом повлиять на значение среднего арифметического.

В предыдущем параграфе рассматривался пример с зарплатой работников некоторой фирмы, в которой работает 20 человек, зарплата 19 из них составляет 10 000 р., а зарплата 20-го, руководителя, — 1 000 000 р. Как мы видели, средняя зарплата одного работника на этой фирме составляет 59 500 р. Среднее арифметическое явилось в данном случае плохой мерой центральной тенденции. Медиана данной совокупности равна 10 000 руб. Она лучше характеризует совокупность, состоящую из размеров зарплат работников фирмы.

? Что означает значение медианы в этом примере?

Медиана обладает следующим важным **СВОЙСТВОМ**.

Сумма модулей отклонений вариант признака от медианы не больше суммы модулей отклонений от любого другого числа:

$$\sum_{i=1}^{m} |x_i - \mathrm{Me}| n_i \leqslant \sum_{i=1}^{m} |x_i - a| n_i.$$

В случае нечетного объема совокупности $n = 2l + 1$ при $a \neq \mathrm{Me}$ имеет место строгое неравенство, а в случае четного $n = 2l$ равенство имеет место для любого значения a, находящегося между двумя средними значениями x_l и x_{l+1} и только при этих значениях.

➤

Докажем это утверждение.

❒ Пусть $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ — упорядоченный набор данных. Рассмотрим сначала случай, когда n — нечетное число, $n = 2l + 1$. Заметим, что из неравенства треугольника вытекает, что для любых трех чисел x, x_1, x_2 выполняется неравенство $|x - x_1| + |x - x_2| \geqslant$ $\geqslant |x_2 - x_1|$, причем равенство имеет место тогда и только тогда, когда $x_1 \leqslant x \leqslant x_2$. Поэтому для произвольного числа a имеем

$$\begin{gathered}|a - x_1| + |a - x_2| + \ldots + |a - x_{2l}| + |a - x_{2l+1}| = \\ = (|a - x_1| + |a - x_{2l+1}|) + (|a - x_2| + |a - x_{2l}|) + \ldots \\ \ldots + (|a - x_l| + |x - x_{l+2}|) + |a - x_{l+1}|) \geqslant \\ \geqslant |x_1 - x_{2l+1}| + |x_2 - x_{2l}| + \ldots + |x_l - x_{l+2}| + |a - x_{l+1}|.\end{gathered}$$

Так как медиана $\mathrm{Me} = x_{l+1}$ находится между каждой парой значений x_1 и x_{2l+1}, x_2 и x_{2l}, ..., x_l и x_{l+2}, то при $a = x_{l+1}$ предыдущее неравенство обращается в равенство. Если же $a \neq x_{l+1}$, то будет иметь место строгое неравенство. Итак, сумма $|a - x_1| + |a - x_2| + \ldots$ $\ldots + |a - x_{2l}| + |a - x_{2l+1}|$ принимает наименьшее значение при $a =$ $= \mathrm{Me} = x_{l+1}$.

Аналогично доказывается утверждение и в случае, когда имеется четное число значений $n = 2l$. При этом наименьшее значение сумма $|a - x_1| + |a - x_2| + \ldots + |a - x_{2l}|$ принимает при любом значении a, лежащем между двумя средними значениями x_l и x_{l+1}, в частности при $a = \mathrm{Me} = \dfrac{x_l + x_{l+1}}{2}$. ■ ➤

Пример 3. Семеро друзей живут вдоль шоссе, которое проложено в лесу. Расположение их домов показано на рис. 10. Они являются членами клуба туристов. Стоимость бензина оплачивается из казны клуба. В каком месте шоссе им необходимо собраться на пикник, чтобы израсходовать на путешествие минимальное количество денег на бензин? Любое место в лесу у шоссе является прекрасным местом для пикника.

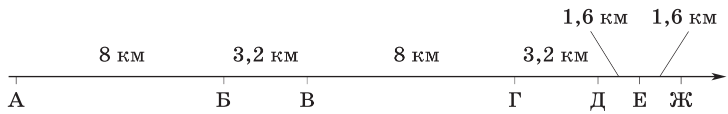


Рис. 10

❐ Нужно выбрать на шоссе такую точку, чтобы сумма расстояний от точек А, Б, В, Г, Д, Е и Ж до этой точки была минимальной. Введем координатную прямую, направив ее вдоль прямой, изображающей шоссе, приняв за начало точку А, направление — в сторону точки Б, единицей масштаба будем считать 1 км. Тогда имеем следующие координаты отмеченных точек: А(0), Б(8), В(11,2), Г(19,2), Д(22,4), Е(24), Ж(25,6). Согласно приведенному свойству сумма расстояний от всех точек до точки с координатой, равной медиане координат всех точек, является наименьшей. Медианой является точка Г(19,2) (четвертая из семи точек, координаты которых расположены в возрастающем порядке). В этом месте друзьям целесообразно собраться на пикник.

В этом случае сумма расстояний, которую должны проехать друзья, равна $19{,}2 + 11{,}2 + 8 + 0 + 3{,}2 + 4{,}8 + 6{,}4 = 52{,}8$ (км). Для сравнения найдем среднее арифметическое координат и подсчитаем сумму расстояний до соответствующей точки. Имеем

$$\bar{x} = \frac{0 + 8 + 11{,}2 + 19{,}2 + 22{,}4 + 24 + 25{,}6}{7} \approx 15{,}8.$$

Сумма расстояний от всех точек до точки с координатой 15,8 равна $15{,}8 + 7{,}8 + 3{,}4 + 3{,}4 + 6{,}6 + 8{,}2 + 9{,}8 = 55$ (км). Как и следовало ожидать, эта сумма оказалась больше предыдущей. ■

? Возьмите любую точку на координатной прямой и найдите сумму расстояний от всех точек А, Б, В, Г, Д, Е и Ж до этой точки. Сравните эту сумму с суммой расстояний до точки Г.

? Изменится ли результат решения примера 3, если начало координат выбрать в точке Б?

Что больше — сумма модулей отклонений вариант признака от медианы или сумма модулей отклонений вариант признака от среднего арифметического?

Медиану определяют и для порядковых качественных данных. Рассмотрим в качестве примера совокупность пяти военнослужащих, имеющих воинские звания: рядовой, ефрейтор, младший сержант, сержант, старший сержант. Эти данные упорядочены по возрастанию званий рядового и сержантского состава. В этой совокупности 5 элементов. Медианой является среднее, третье, т. е. «младший сержант».

Если же в подобной совокупности четное число данных, причем средние данные различны, то считают, что медианой является пара средних данных: ведь найти их среднее арифметическое нельзя. Если к перечисленным военнослужащим добавить одного с воинским званием старшина, то медианой совокупности, состоящей из 6 элементов, является пара «младший сержант и сержант».

Пример 4. В таблице 1.32 представлено распределение личного состава подразделения по приведенным воинским званиям.

Найти медиану приведенной совокупности.

Таблица 1.32

Звание	Число военнослужащих
Рядовой	25
Ефрейтор	18
Младший сержант	7
Сержант	5
Старший сержант	2

❐ Объем совокупности равен 57. Этот набор данных является порядковым, так как для военнослужащих существует естественный порядок — старшинство воинского звания. Медианой будет среднее значение — 29-е. В подразделении 25 рядовых, поэтому медиана будет находиться за пределами этого звания. Рядовых и ефрейторов в подразделении $25 + 18 = 43$. Таким образом, медиана находится между военнослужащими с номерами 26 и 43, т. е. медианой этого подразделения является «ефрейтор». Это означает, что около половины личного состава подразделения имеют воинское звание ефрейтор и ниже его, и примерно половина — ефрейтор и выше этого звания. ■

Медиану можно определить и в терминах *рангов* (см. § 1.2). В предыдущем примере можно подсчитать ранг каждого военнослужащего. Все рядовые имеют один и тот же ранг, равный

$$\frac{1+2+\ldots+25}{25}=\frac{1}{25}\cdot\frac{1+25}{2}\cdot 25=13$$

(мы применили формулу суммы первых n членов арифметической прогрессии). Все ефрейторы имеют ранг, равный

$$\frac{26+27+\ldots+43}{18}=\frac{1}{18}\cdot\frac{26+43}{2}\cdot 18=34{,}5.$$

Ранг медианы равен 29, поэтому медианой является «ефрейтор».

Для вычисления медианы в Excel можно использовать функцию (МЕДИАНА).

1.6.2. Мода

Напомним, что среднее арифметическое является хорошей мерой центральной тенденции для количественных данных, не имеющих выбросов; медиана — для порядковых данных и для количественных данных, в том числе и при наличии выбросов. Подобная характеристика нужна и для номинальных данных. Такой характеристикой является мода. Она применяется как для неупорядоченных категорий, так и для упорядоченных, и для количественных данных. При этом для количественных данных может иметь место и некоторая неопределенность.

Мода *— это такое значение признака, которое встречается наиболее часто.* В случае дискретных рядов вычислить моду нетрудно. Достаточно найти варианту, которая имеет наибольшую частоту или относительную частоту, это и будет мода. Будем обозначать моду символом Мо.

Пример 5. Во время выборов подсчитывают число голосов, отданных за ту или иную партию, за того или иного кандидата. У каждого избирателя, у каждого политолога может быть свое мнение по поводу упорядочения кандидатов. Так как общего мнения нет, то список кандидатов или партий можно считать неупорядоченным. В таблице 1.33 представлены данные о результатах выборов пяти партий, условно обозначенных буквами А, Б, В, Г и Д.

Указать моду этого распределения.

Таблица 1.33

Название партии	**Число полученных голосов**	**Процент**
А	7515	15
Б	14 028	28
В	3507	7
Г	17 034	34
Д	8016	16
Итого	50 100	100

❒ Ясно, что модой в этом наборе данных будет партия Г: она набрала больше всех голосов. ■

Пример 6. Важным приемом при обучении решению задач является анализ ошибок, допущенных на различных этапах решения. Сбор и анализ такой информации является важным моментом в процессе обучения. В таблице 1.34 представлены результаты регистрации причин ошибок, приведших к неправильному решению задачи.

Указать моду этого распределения.

Таблица 1.34

Причина ошибки	**Число случаев**
Построение модели	42
Преобразования	16
Решение уравнений	12
Вычисления	18
Проверка решения	12

❒ Ясно, что модой в этом наборе является построение модели, поскольку эта причина ошибок встречается чаще всего. Мода помогает сосредоточить внимание на самой важной категории, уточнить имеющуюся проблему. ■

? Можно ли считать приведенный в примере 6 набор данных порядковым?

Пример 7. В таблице 1.35 приведены итоговые оценки учащихся некоторого класса по математике.

Найти моду данного распределения.

Таблица 1.35

Количество баллов	1	2	3	4	5	6	7	8	9	10	11	12
Число учащихся	1	1	2	3	4	4	6	5	3	3	2	1

❐ Из всех оценок чаще всего встречается 7 баллов: шесть раз. Поэтому Мо = 7. Этот результат имеет вполне определенный смысл — больше всего учащихся класса имеют по математике 7 баллов. ■

Если все значения в вариационном ряде встречаются одинаково часто, то считают, что этот ряд *не имеет моды*.

Если два соседних значения вариационного ряда имеют одинаковую частоту и она больше частоты любого другого значения, то считают, что мода равняется среднему арифметическому этих значений.

Если два не соседних значения вариационного ряда имеют одинаковую частоту и она больше частоты любого другого значения, то считают, что вариационный ряд имеет *две моды*, а соответствующее распределение называют ***бимодальным***.

В случае интервальных рядов с равными интервалами за приближенное значение моды можно взять центр ***модального интервала***, т. е. интервала с наибольшей частотой или относительной частотой. Точнее значение моды можно получить по формуле

$$\mathrm{Mo} = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)},$$

где x_0 — начальное значение модального интервала, т. е. интервала, который содержит моду; h — длина модального интервала; n_2 — частота модального интервала; n_1 — частота интервала, предшествующего модальному; n_3 — частота интервала, следующего за модальным.

Эту формулу доказывают средствами математического анализа.

➤

❐ Обозначим через $F_n(x)$ относительную частоту элементов совокупности, принимающих значения, меньшие x. Разность $F_n(x + \Delta x) -$

$-\ F_n(x)$ равна относительной частоте элементов, содержащихся в интервале $(x,\ x + \Delta x)$. Тогда средняя плотность распределения на этом интервале равна $\dfrac{F_n(x + \Delta x) - F(x)}{\Delta x}$. Предположим, что функция $F_n(x)$ имеет производную, т. е. существует предел

$$\lim_{\Delta x \to 0} \frac{F_n(x + \Delta x) - F(x)}{\Delta x} = F'_n(x) = \varphi(x).$$

Полученную функцию $\varphi(x)$ называют *теоретической плотностью распределения в точке* x. Для этой функции функция $F_n(x)$ является первообразной. Согласно формуле Ньютона—Лейбница получим

$$F_n(x_2) - F_n(x_1) = \int\limits_{x_1}^{x_2} \varphi(x)\,\mathrm{d}x.$$

Разность, стоящая в левой части равенства, равна относительной частоте w элементов на интервале $(x_1,\ x_2)$. Поэтому

$$w = \int\limits_{x_1}^{x_2} \varphi(x)\,\mathrm{d}x.$$

Для величин с непрерывной плотностью распределения $y = \varphi(x)$ мода может быть определена как точка максимума функции $y = \varphi(x)$.

Для нахождения такого значения $x_1 = \text{Мо}$ кривую распределения $y = \varphi(x)$ в границах модального интервала и двух соседних с ним интервалов — слева и справа — приближенно заменим параболой второго порядка $y = ax^2 + bx + c$, т. е. будем считать, что в упомянутых интервалах $\varphi(x) = ax^2 + bx + c$, где $a < 0$.

Почему $a < 0$?

По условию ширина h всех интервалов одинакова. Для упрощения изложения временно примем за нуль начало модального интервала. Тогда модальный интервал будет от 0 до h. Интервал, предшествующий модальному, — от $-h$ до 0, интервал, следующий за модальным, — от h до $2h$. Относительные частоты этих интервалов обозначим соответственно w_2, w_1, w_3. Неизвестные параметры a, b, c определяются из равенств:

$$w_1 = \int_{-h}^{0} \varphi(x)\,\mathrm{d}x; \quad w_2 = \int_{0}^{h} \varphi(x)\,\mathrm{d}x; \quad w_3 = \int_{h}^{2h} \varphi(x)\,\mathrm{d}x,$$

где $\varphi(x) = ax^2 + bx + c$. Вычислим эти интегралы:

$$w_1 = \int_{-h}^{0} (ax^2 + bx + c)\,\mathrm{d}x = \frac{ah^3}{3} - \frac{bh^2}{2} + ch;$$

$$w_2 = \int_{0}^{h} (ax^2 + bx + c)\,\mathrm{d}x = \frac{ah^3}{3} + \frac{bh^2}{2} + ch;$$

$$w_3 = \int_{h}^{2h} (ax^2 + bx + c)\,\mathrm{d}x = \frac{7ah^3}{3} + \frac{3bh^2}{2} + ch.$$

Решив систему уравнений

$$\begin{cases} \dfrac{ah^3}{3} - \dfrac{bh^2}{2} + ch = w_1; \\ \dfrac{ah^3}{3} + \dfrac{bh^2}{2} + ch = w_2; \\ \dfrac{7ah^3}{3} + \dfrac{3bh^2}{2} + ch = w_3, \end{cases}$$

получим

$$a = \frac{w_3 - 2w_2 + w_1}{2h^3}; \quad b = \frac{w_2 - w_1}{h^2}; \quad c = \frac{2w_1 + 5w_2 - w_3}{6h}.$$

Как известно, точкой максимума квадратичной функции $\varphi(x) = ax^2 + bx + c$ при $a < 0$ является точка $x = -\dfrac{b}{2a}$. В полученную формулу вместо a и b подставим их значения. После упрощения получим

$$x = -\frac{h(w_2 - w_1)}{w_3 - 2w_2 + w_1} = h\frac{w_2 - w_1}{(w_2 - w_1) + (w_2 - w_3)}.$$

Последняя формула получена при условии, что начало модального интервала условно принято в качестве нуля. В общем случае абсцисса точки максимума, т. е. мода Мо, функции $y = \varphi(x)$ определяется по формуле Mo = $x + x_0$, где x_0 — начало модального интервала. Итак,

$$\text{Mo} = x_0 + h\frac{w_2 - w_1}{(w_2 - w_1) + (w_2 - w_3)}.$$

После замены $w_i = \frac{n_i}{n}$ (i = 1, 2, 3) получим искомую формулу для моды:

$$\text{Mo} = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)}. \ \blacksquare$$ ◄

Пример 8. Вычислить моду по данным таблицы 1.23 из предыдущего параграфа.

❐ Здесь модальным является интервал (4—6), так как он имеет наибольшую частоту; $x_0 = 4$, $h = 2$, $n_2 = 478$, $n_1 = 26$, $n_3 = 369$. Поэтому

$$\text{Mo} = 4 + 2\frac{478 - 26}{(478 - 26) + (478 - 369)} \approx 5{,}61. \ \blacksquare$$

? Что означает полученное значение моды?

Если набор данных представляет собой описание причин выхода из строя сложного устройства с соответствующими частотами, то мода помогает сосредоточить внимание на самой важной категории.

Если набор данных представляет собой описание последовательных этапов производства сложного устройства (например, автомобиля) с указанием количества блоков, находящихся на разных стадиях производства, то мода указывает на стадию производства, на которой находится наибольшее количество блоков, т. е. на «узкое» место в производстве.

Следует осторожно относиться к использованию моды для характеристики степени центрирования данных. Например, пусть в классе, в котором 22 учащихся, выполняется тест, состоящий из 25 заданий. На тестирование явилось 20 человек, двое не явились; все тестировавшиеся показали различные результаты. Модой является результат «не явился». Конечно, этот результат является плохой характеристикой результатов тестирования.

Для вычисления моды в Excel можно использовать функцию (МОДА).

Выясним, в каких случаях мода, медиана и среднее арифметическое дают близкие значения и от чего зависят различия между этими показателями.

Если «сгладить» гистограмму гладкой кривой, т. е. провести гладкую кривую через середины верхних оснований прямоугольников, то получим так называемую ***кривую распределения***. Мода является абсциссой точки максимума кривой распределения. Графически медиану можно определить как точку на оси абсцисс, в которой ордината разделяет площадь под графиком распределения на две равные части.

Если график распределения имеет симметричную форму и точку максимума, то прямая, разделяющая площадь под кривой пополам, и центр тяжести лежат на оси симметрии (рис. 11). Итак, для такого симметричного распределения мода, медиана и среднее арифметическое совпадают.

Если график распределения имеет правостороннюю асимметрию («хвост» вправо), то в этом случае мода размещена левее, а среднее арифметическое — правее медианы (рис. 12).

В самом деле, левее и правее медианы размещены одинаковые площади, но левой части соответствует меньшее основание, а правой — большее основание. Поэтому левая часть кривой выше правой, и мода будет находиться левее медианы. По этой самой причине медиана не может служить точкой равновесия, так как площадь части, соответствующей большему основанию, больше площади, размещенной на меньшем основании. Итак, среднее арифметическое находится правее медианы.

Таким образом, при правосторонней асимметрии левее расположена мода, далее медиана и правее — среднее арифметическое. Обратное расположение имеет место при левосторонней асимметрии графика. При этом, чем больше асимметричен график, тем больше расстояние между его средними точками.

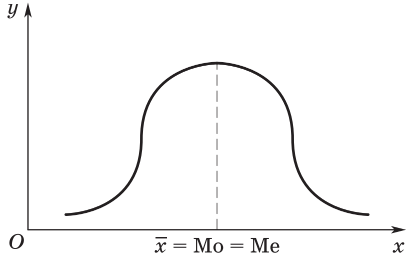


Рис. 11

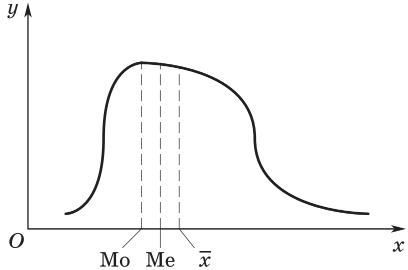


Рис. 12

1.6.3. Некоторые виды степенного среднего

Среднее арифметическое входит в группу средних величин, объединенных названием — *степенное среднее*. В эту группу, наряду со средним арифметическим, входят среднее гармоническое, среднее геометрическое, среднее квадратичное и т. д.

Среднее гармоническое. Рассмотрим две задачи, которые вы наверняка решали в младших классах.

ЗАДАЧА 1. Первую половину времени, затраченного на все путешествие, турист двигался со скоростью 4 км/ч, а вторую половину времени — со скоростью 6 км/ч. Какова средняя скорость движения туриста на протяжении всего путешествия?

❒ Если обозначить расстояние, пройденное туристом, через s, а время, затраченное на все путешествие, через t, то $s = 4 \cdot \frac{t}{2} + 6 \cdot \frac{t}{2}$, и средняя скорость будет равна отношению всего пути ко времени:

$$v_{\text{ср}} = \frac{s}{t} = \frac{4 \cdot \frac{t}{2} + 6 \cdot \frac{t}{2}}{t} = \frac{4+6}{2} = 5 \text{ км/ч}.$$

В этом случае средняя скорость равна среднему арифметическому скоростей, с которыми двигался турист на двух одинаковых временных промежутках. ■

ЗАДАЧА 2. Первую половину пути турист двигался со скоростью 4 км/ч, а вторую половину — со скоростью 6 км/ч. Какова средняя скорость движения туриста на протяжении всего пути?

❒ При тех же обозначениях имеем

$$t = \frac{s}{2 \cdot 4} + \frac{s}{2 \cdot 6}; \quad v_{\text{ср}} = \frac{s}{t} = \frac{s}{\frac{s}{2 \cdot 4} + \frac{s}{2 \cdot 6}} = \frac{2}{\frac{1}{4} + \frac{1}{6}} = 4{,}8 \text{ км/ч}.$$

Величину $\frac{2}{\frac{1}{4} + \frac{1}{6}}$ называют средним гармоническим чисел 4 и 6. ■

В общем случае *среднее гармоническое* значений $x_1, x_2, \ldots, x_n$ определяется по формулам

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}} \quad \text{или} \quad \bar{x}_{\text{гарм}} = \frac{\sum\limits_{i=1}^{m} n_i}{\sum\limits_{i=1}^{m} \frac{n_i}{x_i}}.$$

Пример 9. Вычислить среднемесячную производительность труда одного рабочего на основании данных о производительности труда на трех шахтах (табл. 1.36).

Таблица 1.36

№ шахты	Среднемесячная производительность труда одного рабочего, т	Общая добыча угля на шахте за месяц, т
1	20	20 100
2	30	37 800
3	40	62 100

❒ На основании данных среднее находится как среднее гармоническое по формуле

$$\bar{x}_{\text{гарм}} = \frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m} \frac{n_i}{x_i}}.$$

$$\text{Искомое среднее} = \frac{\text{Общая месячная добыча на всех шахтах}}{\text{Общее количество рабочих на всех шахтах}} =$$

$$= \frac{\sum \text{Общая месячная добыча на шахте}}{\sum \dfrac{\text{Общая месячная добыча на шахте}}{\text{Среднемесячная производительность труда на этой шахте}}}.$$

Данные представим в виде таблицы 1.37.

Таблица 1.37

x	n	$\frac{n}{x}$
20	20 100	1005
30	37 800	1260
40	62 100	1552,5
Σ	120 000	3817,5

$$x_{\text{гарм}} = \frac{120\,000}{3817{,}5} \approx 31{,}4. \ \blacksquare$$

Среднее гармоническое необходимо в том случае, когда наблюдения, для которых мы хотим получить среднее арифметическое, заданы обратными значениями.

В примере 9 нам нужно было вычислить среднее арифметическое для производительности труда рабочего на трех шахтах. Если бы мы знали, сколько человек работает на каждой шахте, то искомое среднее вычисляли бы по формуле среднего арифметического: общую добычу угля на трех шахтах мы делили бы на общее количество рабочих.

Но нам дополнительно пришлось находить количество рабочих, которые работают на каждой шахте. Для этого известны общая добыча угля на каждой шахте и среднемесячная производительность труда на каждой шахте. Аналогично с помощью среднего гармонического вычисляется средняя скорость на эстафете, если известны скорости на каждом этапе и длины всех этапов. Среднее гармоническое используется при расчете средней продолжительности жизни.

Для вычисления среднего гармонического в Excel можно использовать функцию (СРГАРМ).

? Приведите примеры применения среднего гармонического для вычисления средних показателей.

Среднее геометрическое. Со средним геометрическим двух чисел вы встречались и в школьном курсе геометрии (высота, проведенная из вершины прямого угла на гипотенузу прямоугольного треугольника, является средним геометрическим между отрезками, на которые она делит гипотенузу), и в курсе алгебры (неравенство между средним арифметическим и средним геометрическим).

? Какая связь существует между средним арифметическим и средним геометрическим двух положительных чисел?

Среднее геометрическое значений x_1, x_2, ..., x_n определяется по формулам:

$$\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \quad \text{или} \quad \bar{x}_{\text{геом}} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}}.$$

Пример 10. На протяжении трех лет производительность труда в некотором цехе увеличивалась соответственно на 10, 15 и 30% по сравнению с предыдущим годом. Найти ежегодный средний процент увеличения производительности труда за эти три года.

❒ Если обозначить через a производительность труда в году, предшествующему увеличению, то через год она станет равной $1{,}1a$, через два года — $1{,}1a \cdot 1{,}15 = 1{,}265a$, через три года — $1{,}265a \cdot 1{,}3 = 1{,}6445a$. Средний процент p увеличения производительности труда за три года позволяет найти производительность труда, если ежегодно она будет увеличиваться на одинаковое число процентов, а именно на $p\%$. В этом случае по формуле сложных процентов будем иметь, что через три года производительность труда будет составлять $a\left(1+\frac{p}{100}\right)^3$, т. е. имеем уравнение $a\left(1+\frac{p}{100}\right)^3 =$ $= 1{,}6445a$. Отсюда $p = 100\cdot(\sqrt[3]{1{,}6445}-1) \approx 18{,}0$. Число $\sqrt[3]{1{,}6445}$ равно среднему геометрическому чисел 1,1; 1,15; 1,3, т. е. искомый средний ежегодный процент p увеличения производительности труда находится по формуле: $p = 100\cdot(\sqrt[3]{1{,}1\cdot 1{,}15\cdot 1{,}3}-1)$. ■

Среднее геометрическое используют прежде всего тогда, когда среднее значение вычисляют для значений, заданных через некоторые равные промежутки времени (рост или снижение успеваемости, заработной платы, вклада в банке за несколько лет). Среднее геометрическое применяют тогда, когда переменная с течением времени изменяется примерно с одинаковым соотношением между измерениями. Среднее геометрическое применяют также тогда, когда отдельные значения в статистической совокупности удалены от других значений; это меньше влияет на среднее геометрическое по сравнению со средним арифметическим, а потому дает более правильное представление о среднем.

Для вычисления среднего геометрического в Excel можно использовать функцию (СРГЕОМ).

Среднее гармоническое и среднее геометрическое относятся к так называемым степенным средним.

Среднее степенное k-го порядка определяется при помощи формул:

$$\bar{x}^{(k)}_{\text{степ}} = \left(\frac{\sum_{i=1}^{n} x_i^k}{n}\right)^{\frac{1}{k}} \quad \text{или} \quad \bar{x}^{(k)}_{\text{степ}} = \left(\frac{\sum_{i=1}^{m} x_i^k n_i}{\sum_{i=1}^{m} n_i}\right)^{\frac{1}{k}}.$$

Среднее арифметическое является степенным средним порядка 1, среднее гармоническое можно считать степенным средним порядка −1.

➤

Если под средним степенным нулевого порядка понимать $\lim\limits_{k\to 0} \bar{x}^{(k)}_{\text{степ}}$, то можно доказать, что среднее геометрическое является степенным средним нулевого порядка.

❐ В самом деле,

$$\ln \bar{x}^{(k)}_{\text{степ}} = \frac{1}{k}\left(\ln \sum_{i=1}^{n} x_i^k - \ln n\right);$$

$$\lim_{k \to 0} \ln \bar{x}^{(k)}_{\text{степ}} = \lim_{k \to 0} \frac{\ln \sum\limits_{i=1}^{n} x_i^k - \ln n}{k}.$$

Применяя правило Лопиталя, будем иметь

$$\lim_{k \to 0} \ln \bar{x}^{(k)}_{\text{степ}} = \lim_{k \to 0} \frac{\ln \sum\limits_{i=1}^{n} x_i^k - \ln n}{k} = \lim_{k \to 0} \frac{\sum\limits_{i=1}^{n} x_i^k \ln x_i}{\sum\limits_{i=1}^{n} x_i^k} =$$

$$= \frac{\sum\limits_{i=1}^{n} \ln x_i}{n} = \ln\left(x_1 \cdot x_2 \cdot \ldots \cdot x_n\right)^{\frac{1}{n}}.$$

Отсюда

$$\lim_{k \to 0} \bar{x}^{(k)}_{\text{степ}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \bar{x}_{\text{геом}}. \blacksquare$$

Среднее степенное второго порядка называют ***средним квадратичным***, его используют при вычислении среднего квадратичного отклонения. Мы рассмотрим его в следующем параграфе. Среднее степенное третьего порядка называют ***средним кубическим*** и т. д.

В § 1.1 были рассмотрены различные шкалы, используемые при измерении величин. Каждому уровню шкалы соответствует определенная мера центральной тенденции. Информация об этом приведена в таблице 1.38.

Контрольные вопросы

1. Какой характеристикой центральной тенденции целесообразно воспользоваться для номинальных данных?

2. Какими характеристиками центральной тенденции целесообразно воспользоваться для порядковых данных?

3. Какими характеристиками центральной тенденции целесообразно воспользоваться для количественных данных?

Таблица 1.38

Уровень шкалы	**Условия**	**Показатель центральной тенденции**	**Примеры**
Номинальная	Признаки отождествляются или различаются	Мода	Перечень причин выхода из строя сложного устройства
Порядковая, или ранговая	Возможность провести градацию по степени выраженности признака	Мода и медиана	Последовательность поступления деталей автомобиля на сборочный конвейер; школьные оценки
Шкала интервалов	Возможность установить равные интервалы, точка отсчета условная	Среднее арифметическое	Шкала температур
Шкала отношений, или пропорций	Возможность определить отношение, точка отсчета абсолютная	Среднее арифметическое и среднее геометрическое	Меры длины; меры массы

4. Почему определение моды содержит неоднозначность?

5. Что можно сказать о кривой распределения, если среднее арифметическое и медиана совпадают?

6. Изучалась статистика пострадавших вследствие дорожно-транспортных происшествий. Выяснилось, что в среднем 61% погибших составляют водители — не профессионалы, 15% — пешеходы, 15% — мотоциклисты, 5% — водители грузовых или служебных машин, 4% — велосипедисты. Какая из статистических характеристик наилучшим образом описывает средние показатели этих данных?

7. После окончания олимпиады подсчитаны баллы, набранные 20 ее участниками, определены места, занятые ими. Какая из статистических характеристик наилучшим образом описывает средние показатели участников олимпиады?

8. Есть данные о спаде производства на некоторой фирме на протяжении 5 лет: 20, 30, 25, 20, 15%. Какая из статистических характеристик наилучшим образом описывает средний ежегодный спад производства?

9. Есть данные о движении 10 туристов, которые прошли одно и то же расстояние соответственно со скоростями: 3; 3; 3,5; 4; 4; 4,5; 4,5; 5; 5,5; 6 км/ч. Какая из статистических характеристик наилучшим образом описывает средние показатели этих данных?

10. Собраны данные о тарифном разряде рабочих цеха: 4 рабочих имеют первый разряд, 6 — второй, 12 — третий, 16 — четвертый, 44 — пятый и 18 рабочих — шестой. Какая из статистических характеристик наилучшим образом описывает средние показатели этих данных?

11. Как изменится медиана совокупности, если ко всем ее значениям прибавить одно и то же число?

12. Как изменится мода совокупности, если ко всем ее значениям прибавить одно и то же число?

13. Как изменится медиана совокупности, если все ее значения умножить на одно и то же число?

14. Как изменится мода совокупности, если все ее значения умножить на одно и то же число?

Задачи

33.° В рыбном магазине имеется 26 наименований черноморских рыбных консервов, 33 наименования балтийских, 47 наименований тихоокеанских и 15 наименований атлантических рыбных консервов. Средние цены черноморских, балтийских, тихоокеанских и атлантических консервов соответственно составляют 26, 35, 28, 41 р.

а) Определите моду места производства рыбных консервов, имеющихся в магазине.

б) Найдите среднее значение цены для всех этих консервов.

34. Служащий получал на протяжении трех последовательных лет надбавки к заработной плате соответственно 10, 20 и 30%. Процентная надбавка относится к заработной плате предыдущего года. Чему равна средняя ежегодная надбавка к заработной плате?

35. В трех различных магазинах продается определенный предмет по цене: 10, 6 и 8 штук за 100 р. Сколько в среднем единиц этого предмета можно приобрести в этих магазинах за 1000 р.?

36.° Вычислите медиану по данным о распределении работников цеха по тарифным разрядам (табл. 1.39).

Таблица 1.39

Тарифные разряды	1	2	3	4	5	6	Всего
Число рабочих	4	6	12	16	44	18	100

37.° Вычислите моду и медиану по данным о процентном содержании золы в угле (табл. 1.40).

Таблица 1.40

Зольность, %	9—12	12—15	15—18	18—21	21—24	24—27
Число проб	17	29	22	18	9	5

38. В некоторой культуре число бактерий в единице объема за три дня увеличилась со 100 до 500. Определите средний прирост за день в процентах.

39. Издательство первоначально разослало свой каталог представительной выборке из 1000 организаций, взятых из имеющегося списка рассылок, и получило заказы на общую сумму 1 084 030 р.

а) Определите средний размер заказа (в рублях) на одну организацию из этой первоначальной рассылки.

б) Какой общий объем заказов (в рублях) следует ожидать издательству в случае рассылки каталога всем 12 346 организациям, включенным в список рассылки?

в) В выборке из 1000 организаций, объем заказов для которой составил 1 084 030 р., реально только 128 организаций сделали заказ. Найдите средний объем заказа для тех, кто действительно сделал заказ.

г) Исходя из задания в), определите число заказов, которые следует ожидать издательству после того, как каталог будет послан каждой из 12 346 организаций, включенных в список рассылки.

40.° По данным, приведенным в таблице 1.41, вычислите моду и медиану. Постройте гистограмму частот. Отметьте на ней моду и медиану.

Таблица 1.41

x	5—7	7—9	9—11	11—13	13—15	15—17	17—19
n_i	4	8	11	7	5	3	2

§ 1.7. Показатели вариации

Средние величины являются важными характеристиками статистических совокупностей. Они говорят нам о концентрациях совокупности значений на числовой шкале. Каждая мера центральной тенденции дает такое значение, которое «представляет» в определенном смысле все значения совокупности. В этом случае пренебрегают различиями, существующими между отдельными значениями. Для измерения разброса и вариации значений внутри совокупности нужны другие показатели.

Пример 1. Пусть средний рост учащихся 8 «А» класса равен 1,65 м, а 8 «Б» — 1,7 м. По этим данным нельзя определить, в каком из этих двух классов учится самый высокий ученик. ■

Пример 2. Учащиеся различаются между собою по успеваемости. На успеваемость влияет много факторов. Различия в успеваемости у школьников, обучающихся примерно в одинаковых условиях, значительно меньшие, чем среди учащихся вообще. При разработке мероприятий по повышению успеваемости школьников необходимо учитывать и их индивидуальные особенности. Некоторые программы повышения успеваемости могут быть ориентированы на всех учащихся, другие — на учащихся, имеющих самую худшую или самую лучшую успеваемость. Определение изменчивости успеваемости дает возможность выявить разброс таких индивидуальных различий и получить полезную информацию для планирования повышения общей успеваемости. ■

Пример 3. Недостаточно знать среднюю производительность труда на любом предприятии. Необходимо учитывать и индивидуальные особенности работников. Некоторые программы повышения производительности труда могут быть ориентированы на всех работников, другие — на самых «медленных» или самых «быстрых». Определение изменчивости производительности труда дает возможность выявить разброс таких индивидуальных различий и получить полезную информацию для планирования повышения общей производительности труда. ■

1.7.1. Размах вариации

О размахе данных уже говорилось в § 1.2. Размах измеряет на числовой шкале расстояние, в пределах которого изменяются значения совокупности. Различают два вида размаха.

Исключающий размах *— это разность между максимальным и минимальным значениями в совокупности.*

Включающий размах *— это разность между естественной верхней границей интервала, содержащего максимальное значение совокупности, и естественной нижней границей интервала, содержащего минимальное значение.*

Естественные верхняя и нижняя границы в окрестности любого значения устанавливаются путем добавления к этому значению и вычитания из этого значения половины погрешности измерения рассматриваемой величины.

Итак, включающий размах равен исключающему размаху плюс погрешность измерения величины.

Пример 4. Вычислить включающий и исключающий размах по данным таблицы 1.8 из § 1.2 о скорости чтения третьеклассников.

❐ Исключающий размах равен 110 − 25 = 85. Естественная верхняя граница в окрестности наибольшей скорости чтения равна 110,5 (к наибольшему значению 110 добавлено 0,5), естественная нижняя граница в окрестности наименьшей скорости чтения равна 24,5 (от наименьшего значения 25 вычли 0,5), поэтому включающий размах равен 110,5 − 24, 5 = 86. ■

Пример 5. Вычислить включающий и исключающий размах по приведенным данным о диаметрах валиков (в мм):

14,51	14,42	14,56	14,47	14,46	14,35	14,48	14,53
14,21	14,31	14,35	14,68	14,56	14,28	14,36	14,21
14,52	14,23	14,41	14,46	14,69	14,54	14,36	14,15
14,37	14,51	14,25	14,55	14,51	14,36	14,62	14,55
14,38	14,33	14,40	14,52	14,48	14,51	14,55	14,39
14,54	14,58	14,48	14,37	14,38	14,51	14,36	14,15
14,24	14,32						

❐ Наименьшее значение диаметра равно 14,15 мм, наибольшее — 14,69 мм. Исключающий размах равен 14,69 − 14,15 = 0,54. Погрешность измерения диаметра равна 0,01 мм. Естественная верхняя граница в окрестности наибольшего диаметра равна 14,695 (к наибольшему значению 14,69 добавлено 0,005), естественная нижняя граница в окрестности наименьшего диаметра равна 14,145 (от наименьшего значения 14,15 вычли 0,005), поэтому включающий размах равен 14,695 − 14,445 = 0,55. ■

Преимуществом этого показателя является очевидная простота его вычисления. Но часто он дает лишь приближенную характеристику вариации. Это особенно заметно в случае достаточно больших по объему совокупностей, когда подавляющее большинство элементов имеют значения признака, компактно сгруппированные вокруг некоторой средней величины, и только некоторые из них в силу случайных обстоятельств принимают значения (наибольшее и наименьшее), существенно отличающиеся от основной массы. При этом размах вариации будет значительным, а вариация по существу малой. Дело в том, что при вычислении размаха не учитывается каждое отдельное значение.

? В каком случае размах может быть хорошей мерой разброса?

1.7.2. Дисперсия

Необходима мера вариации, учитывающая каждое отдельное значение совокупности. Естественно было бы рассмотреть сумму отклонений всех значений совокупности от среднего арифметического. Но, как известно, эта сумма равняется нулю (см. свойство 6 среднего арифметического в § 1.5) и потому ее нельзя использовать в качестве показателя вариации. А почему эта сумма всегда равна нулю? Имеются положительные и отрицательные отклонения, они взаимно компенсируют друг друга и потому их сумма является нулевой. Чтобы избежать этого, следует учитывать лишь величину отклонения и не учитывать его знак. Это можно сделать, рассмотрев квадраты отклонений или их модули[1].

Дисперсией статистической совокупности *называют среднее арифметическое квадратов отклонений вариант от их среднего арифметического значения.*

Обозначают дисперсию через s^2:

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{или} \quad s_x^2 = \frac{1}{n}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i.$$

Дисперсия имеет размерность, равную квадрату размерности элементов совокупности. Чтобы иметь показатель вариации с той же размерностью, что и размерность элементов данной совокупности, рассматривают также так называемое ***среднее квадратичное***

[1] Величину $E_x = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$, или $E_x = \frac{1}{n}\sum_{i=1}^{m}|x_i - \bar{x}|n_i$ называют ***средним отклонением***.

отклонение, *или* ***стандартное отклонение s***, *которое равно арифметическому корню квадратному из дисперсии*:

$$s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{или} \quad s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i}.$$

? Какую размерность имеет среднее квадратичное отклонение?

Стандартное отклонение имеет простую и понятную интерпретацию: эта величина описывает *типичное расстояние от среднего значения* для отдельных значений набора данных.

Пример 6. Найти s^2 и s по данным таблицы 1.20 из § 1.5 о количестве баллов, которые набрали на олимпиаде представители одного района.

❐ Ранее мы нашли, что $\bar{x} \approx 7{,}9$. Вычисления удобно проводить по схеме, представленной в таблице 1.42.

Таблица 1.42

x_i	n_i	$x_i n_i$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
2	1	2	5,9	34,81	34,81
3	2	6	4,9	24,01	48,02
5	4	20	2,9	8,41	33,64
6	3	18	1,9	3,61	10,83
8	2	16	0,1	0,01	0,02
9	2	18	1,1	1,21	2,42
10	2	20	2,1	4,41	8,82
11	3	33	3,1	9,61	28,83
15	1	15	7,1	50,41	50,41
18	1	18	10,1	102,01	102,01
Σ	21	166 $\bar{x} = \frac{166}{21} \approx 7{,}9$			319,81 $s_x^2 = \frac{319{,}81}{21} \approx 15{,}23$

Итак, $s_x^2 \approx 15{,}23$, $s_x \approx 3{,}90$. ■

Из определения дисперсии вытекает ряд свойств, которые применяют как для упрощения вычислений, так и для осознания сущности этого показателя. Рассмотрим некоторые из них.

СВОЙСТВО 1. *Дисперсия постоянной величины равна нулю.*

❐ В самом деле, поскольку все варианты одинаковы ($x_i = c$), то и их среднее значение равняется c: $\bar{x} = c$. Поэтому все отклонения от среднего равны нулю: $x_i - \bar{x} = c - c = 0$, а значит, дисперсия постоянной равняется нулю. ■

СВОЙСТВО 2. *Если ко всем элементам совокупности прибавить или из всех элементов вычесть одно и то же число a, то дисперсия не изменится.*

$$❐\ s_{x+a}^2 = \frac{1}{n}\sum_{i=1}^{n}\left((x_i + a) - (\overline{x+a})\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i + a - \bar{x} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s_x^2.\ ■$$

Верно ли, что если зарплату всех работников некоторой фирмы увеличить на одну и ту же сумму, то стандартное отклонение не изменится?

СВОЙСТВО 3. *Если все элементы совокупности умножить (разделить) на одно и то же число $h \neq 0$, то дисперсия умножится (разделится) на h^2.*

$$❐\ s_{hx}^2 = \frac{1}{n}\sum_{i=1}^{n}(hx_i - \overline{hx})^2 = \frac{1}{n}\sum_{i=1}^{n}(hx_i - h\bar{x})^2 =$$

$$= \frac{h^2}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = h^2 s_x^2.\ ■$$

Как изменится стандартное отклонение зарплаты в цехе, если заработная плата каждого работника будет удвоена?

СВОЙСТВО 4. *Дисперсия равна разности среднего арифметического квадратов вариант и квадрата среднего арифметического значений вариант*:

$$s_x^2 = \overline{x^2} - \bar{x}^2.$$

❐ В самом деле,

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2}{n}\bar{x}\sum_{i=1}^{n}x_i + \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2. \blacksquare$$

Полученную формулу можно использовать для вычисления дисперсии.

Пример 7. По данным таблицы 1.21 (с. 57) вычислить дисперсию, воспользовавшись этой формулой.

❐ Вычисления представлены в таблице 1.43.

Таблица 1.43

x_i	n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
2	1	2	4	4
3	2	6	9	18
5	4	20	25	100
6	3	18	36	108
8	2	16	64	128
9	2	18	81	162
10	2	20	100	200
11	3	33	121	363
15	1	15	225	225
18	1	18	324	324
Σ	21	166 $\bar{x} = \frac{166}{21} \approx 7{,}9$		1632 $\overline{x^2} = \frac{1632}{21} \approx 77{,}7$

$$s_x^2 \approx 77{,}7 - 7{,}9^2 \approx 15{,}29.$$

Небольшое различие в результатах, полученных при использовании различных способов вычисления, объясняется наличием вычислительных погрешностей в промежуточных вычислениях. ■

Стандартное отклонение, как указывалось выше, отражает типичную картину отклонения отдельных значений от среднего значения совокупности. Для одних значений отклонение будет меньше стандартного, для других — больше. Схематически связь между отклонениями отдельных значений от среднего и стандартным отклонением показана на рис. 13.

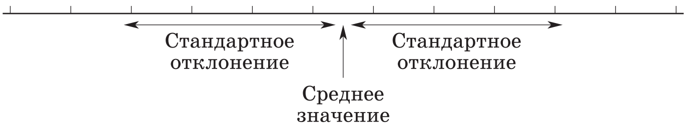


Рис. 13

Пример 8. Пусть в классе проведено тестирование по математике. Балл учащегося N оказался равным 17. Результаты тестирования учащихся класса следующие: 9, 17, 19, 18, 23, 25, 25, 19, 13, 10, 19, 18, 19, 25, 10, 15, 16, 15, 20, 12.

Требуется определить, типичен ли результат учащегося N для всего класса.

❐ Средний балл в классе равен 17,4, стандартное отклонение — 4,94. Разность между результатом учащегося N и средним баллом значительно меньше стандартного отклонения. Таким образом, результат учащегося N, несмотря на то что он меньше среднего, является типичным для класса. ■

Дисперсия статистической совокупности может быть использована для сравнения двух совокупностей, для оценивания параметров, для предварительной проверки статистических гипотез. Если арифметические средние у двух совокупностей окажутся одинаковыми, то в некоторых случаях вопрос о том, какой совокупности отдать предпочтение, может быть решен с помощью дисперсии статистической совокупности. Например, если у двух стрелков совпадают средние числа выбитых очков, то лучшим естественно признать того, у кого меньший разброс числа очков относительно среднего, т. е. того из стрелков, у кого «кучнее» попадания. Например, пусть каждый из двух стрелков при всех трех выстрелах попал в «десятку». Их попадания изображены на рис. 14 («десятка» изображена в увеличенном виде). Средние числа выбитых очков у них одинаковы и равны 10. Но ни у кого не вызывает сомнения утверждение о том, что второй стрелок стреляет лучше.

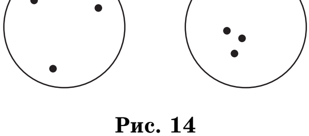


Рис. 14

Пример 9. Два стрелка сделали по 100 выстрелов. Первый выбил 8 очков 40 раз, 9 очков — 10 раз и 10 очков — 50 раз. Второй выбил 8, 9 и 10 очков соответственно — 10, 70 и 20 раз. Какой из стрелков стреляет лучше?

❐ Вспоминая решение примера 5 из § 1.5, можно прийти к выводу, что для ответа на вопрос достаточно вычислить средние числа очков $\bar{x}$ и $\bar{y}$, выбиваемых каждым из стрелков при 100 выстрелах. Однако оказывается, что $\bar{x} = \bar{y} = 9{,}1$. Средние арифметические не позволили отдать предпочтение одному из стрелков. Вычислим меру разброса данных — дисперсию статистической совокупности: ведь при равенстве средних естественно отдать предпочтение тому из стрелков, у которого попадания группируются кучнее вокруг среднего, т. е. тому, для которого мера разброса числа выбитых очков принимает меньшее значение. Вычисления выполним по схеме, представленной в таблицах 1.44 и 1.45.

Таблица 1.44

x_i	n_i	$x_i n_i$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
8	40	320	1,1	1,21	48,4
9	10	90	0,1	0,01	0,1
10	50	500	0,9	0,81	40,5
Σ	100	910 $\bar{x} = 9{,}1$			89 $s_1^2 = 0{,}89$

Таблица 1.45

x_i	n_i	$x_i n_i$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
8	10	80	1,1	1,21	12,1
9	70	630	0,1	0,01	0,7
10	20	200	0,9	0,81	16,2
Σ	100	910 $\bar{y} = 9{,}1$			29,0 $s_2^2 = 0{,}29$

При равенстве средних арифметических дисперсия статистической совокупности у второго стрелка оказалась меньшей. Естественно его считать лучшим. ■

Иногда дисперсию определяют при помощи формул

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{или} \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i,$$

т. е. сумма квадратов отклонений вариант от их среднего арифметического делится не на n, а на $n - 1$. Причины этого будут разъяснены далее в главе 5. Все свойства дисперсии остаются одинаковыми при обоих определениях.

Дисперсия s_x^2 и стандартное отклонение s_x могут быть вычислены с помощью Excel. Для их вычисления в Excel можно использовать соответственно функции (ДИСП) и (СТАНДОТКЛОН).

1.7.3. Коэффициент вариации

Коэффициент вариации представляет собой относительную меру изменчивости.

Коэффициентом вариации V называют отношение среднего квадратичного отклонения s к среднему арифметическому вар믐ант $\bar{x}$: $V = \frac{s}{\bar{x}}$ или в процентах $V = \frac{s}{\bar{x}} \cdot 100\%$.

В отличие от всех рассмотренных выше показателей вариации коэффициент вариации является числом относительным, безразмерным. Это свойство позволяет использовать его для сравнения вариаций не только однородных признаков, но и различных.

? Может ли коэффициент вариации превысить 100%?

1.7.4. Стандартизированные данные

Часто целесообразно преобразовать статистические данные так, чтобы они имели нулевое среднее арифметическое и единичное стандартное отклонение. Такие преобразования называют *переходом к стандартизированным показателям*. Одной из главных причин преобразования первоначальных данных в стандартизированные показатели является желание иметь возможность сопоставлять результаты исследований, проведенных разными методами и средствами. Новые преобразованные данные будут непосредственно выражаться в отклонениях первоначальных значений от среднего арифметического, которые измеряются в единицах стандартного отклонения.

Преобразованные значения определяют с помощью равенства $z_i = \frac{x_i - \bar{x}}{s_x}$ и называют *z-показателями*. Убедимся в том, что $\bar{z} = 0$, $s_z = 1$.

В самом деле,

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s_x} = \frac{1}{s_x}\cdot\frac{1}{n}\left(\sum_{i=1}^{n} x_i - n\bar{x}\right) = 0;$$

$$s_z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^2 =$$

$$= \frac{1}{s_x^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{s_x^2}\cdot s_x^2 = 1.$$

Пример 10. Стандартизировать данные, приведенные в таблице 1.21 в § 1.5.

❒ При решении этого примера было найдено, что $\bar{x} \approx 7{,}9$, $s_x \approx 3{,}90$. Стандартизация данных проводится по схеме, представленной в таблице 1.46.

Таблица 1.46

x_i	2	3	5	6	8
n_i	1	2	4	3	2
$x_i - \bar{x}$	−5,9	−4,9	−2,9	−1,9	0,1
$z_i = \frac{x_i - \bar{x}}{s_x}$	−1,51	−1,26	−0,74	−0,49	0,026
x_i	9	10	11	15	18
n_i	2	2	3	1	1
$x_i - \bar{x}$	1,1	2,1	3,1	7,1	10,1
$z_i = \frac{x_i - \bar{x}}{s_x}$	0,282	0,538	0,795	1,82	2,59

В последней строке получены стандартизированные данные. ■

Контрольные вопросы

1. Как изменится размах совокупности, если ко всем ее значениям прибавить одно и то же число?

2. Как изменится размах совокупности, если все ее значения умножить на одно и то же число?

3. Как изменится стандартное отклонение совокупности, если ко всем ее значениям прибавить одно и то же число?

4. Как изменится стандартное отклонение совокупности, если все ее значения умножить на одно и то же число?

5. Как изменится коэффициент вариации совокупности, если все ее значения умножить на одно и то же положительное число?

6. Верно ли, что коэффициент вариации совокупности не изменится, если ко всем ее значениям прибавить одно и то же число?

7. Что больше — размах совокупности или среднее квадратичное отклонение?

8. Как изменится дисперсия совокупности s^2, если все частоты умножить на одно и то же число?

9. Как изменится среднее абсолютное отклонение совокупности, если ко всем ее значениям прибавить одно и то же число?

10. Верно ли, что если все элементы совокупности умножить на одно и то же число, то среднее абсолютное отклонение умножится на то же число?

11. Верно ли, что расстояние от значений совокупности до среднего значения не превосходит стандартного отклонения?

Задачи

41. Некто решил стать фермером, заняться производством зерна. Изучая положение дел на соседних фермах, он получил данные об урожайности (табл. 1.47).

Таблица 1.47

Урожайность, ц/га	20—22	22—24	24—26	26—28	28—30
Посевная площадь, га	100	150	250	150	150

а)° Определите среднюю урожайность на фермах.

б)° Найдите дисперсию и стандартное отклонение урожайности на фермах. Дайте краткую интерпретацию стандартного отклонения для исследования различий между фермами.

в)° Определите размах и дайте краткую интерпретацию размаха для исследования различий между фермами.

г) Найдите коэффициент вариации и дайте краткую интерпретацию коэффициента вариации для исследования различий между фермами.

д) Постройте гистограмму для приведенного набора данных, укажите на графике среднее значение, стандартное отклонение и размах.

42. В таблице 1.48 приведены данные о прочности волокон хлопка.

Таблица 1.48

Пределы прочности, г/мм^2	1—2	2—3	3—4	4—5	5—6	6—7	7—8
Количество образцов	30	80	210	320	160	120	80

а)$^\circ$ Определите среднюю прочность волокон хлопка.

б)$^\circ$ Найдите дисперсию и стандартное отклонение прочности волокон. Дайте краткую интерпретацию стандартного отклонения для исследования различий между образцами.

в) Сколько образцов не выходит за пределы одного стандартного отклонения от среднего?

г) Сколько образцов находится в пределах двух стандартных отклонений от среднего?

д) Сколько образцов находится в пределах трех стандартных отклонений от среднего?

е) Определите размах и дайте краткую интерпретацию размаха для исследования различий между образцами.

ж) Найдите коэффициент вариации и дайте краткую интерпретацию коэффициента вариации для исследования различий между образцами.

з) Постройте гистограмму для приведенного набора данных, укажите на графике среднее значение, стандартное отклонение и размах.

43. Все 20 сотрудников фирмы получили 20%-ную надбавку к зарплате. Как изменились при этом:

а) средняя зарплата на фирме;

б) стандартное отклонение зарплат;

в) размах зарплаты;

г) коэффициент вариации зарплат?

44. Контроль прошли 40 партий по 250 конденсаторов. Оказалось, что в них число бракованных изделий соответственно составляет:

2	25	5	3	6	4	1	3	3	2
7	5	3	0	1	6	3	5	41	0
1	1	3	4	1	7	1	2	3	1
2	4	5	0	5	3	3	1	2	6

а) Вычислите стандартное отклонение и размах.

б) Вычислите стандартное отклонение и размах, не рассматривая два крайних значения.

в) Чувствительны ли эти характеристики к исключению из рассмотрения двух крайних значений?

45. В таблице 1.49 приведены данные о настриге шерсти с одной овцы.

Таблица 1.49

Настриг шерсти с одной овцы, кг	3	4	5	6	7
Количество овец	100	150	300	350	100

а) Вычислите стандартное отклонение и дайте его интерпретацию для приведенных данных.

б) Вычислите размах и дайте его интерпретацию.

в) На сколько величин стандартного отклонения превышает среднее то значение, которое имеет наибольшее положительное отклонение?

г) На сколько величин стандартного отклонения наименьшее значение ниже среднего значения?

46. Цена всех 1000 видов продукции, находящейся в магазине, увеличилась на 10%. Как изменились при этом:

а) средняя цена продукции;

б) стандартное отклонение цен;

в) размах цен;

г) коэффициент вариации цен?

§ 1.8. Квантили

Ранее мы видели, что медиана делит совокупность данных на две равные части или в отношении 1 : 1. В статистике также рассматривают показатели, делящие совокупность данных в произвольном отношении. Такие показатели называют ***квантилями***.

Пусть $0 < p < 1$ *и статистические данные размещены по возрастанию:* $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. *Если* np *не является целым числом, то* ***квантилем уровня p*** *(обозначим его через* z_p*) называют варианту* x_{k+1}, *где* $k = [np]$ *означает наибольшее целое число, не превышающее* np. *Если* np *является целым числом, то квантиль* z_p *считают равным полусумме вариант* x_{np} *и* x_{np+1}.

Понятно, что медиана является квантилем уровня 0,5. Напомним, если объем совокупности является нечетным числом и равен

$2n + 1$, то медиана $z_{0,5}$ равна x_{n+1}, если объем совокупности — четное число, равное $2n$, то $\text{Me} = z_{0,5} = \dfrac{x_n + x_{n+1}}{2}$.

Квантили $z_{0,25}$, $z_{0,5}$, $z_{0,75}$ делят совокупность данных на 4 равные части (кварты). Их называют ***квартилями***. Четвертая часть наблюдений лежит ниже $z_{0,25}$, половина наблюдений — ниже $z_{0,5}$, три четверти наблюдений — ниже $z_{0,75}$. Таким образом, три квартили делят совокупность наблюдений на четыре равные части.

Квантили $z_{0,1}$, $z_{0,2}$, ..., $z_{0,9}$ делят совокупность наблюдений на 10 равных частей. Их называют ***децилями***.

99 квантилей $z_{0,01}$, $z_{0,02}$, ..., $z_{0,99}$ делят совокупность наблюдений на 100 частей с равным количеством наблюдений в каждой. Их называют ***перцентилями***.

Если по результатам тестирования упорядочить учащихся по убыванию их успеваемости, то квартили $z_{0,25}$, $z_{0,5}$, $z_{0,75}$ делят их на четыре равные части: в первую входят лучшие по успеваемости 25% учащихся, во вторую — следующие 25% по успеваемости учащихся и т. д.

Квантили являются удобным показателем для обобщения данных. Даже простая информация о том, что $z_{0,05} = 15$, $z_{0,15} = 24$, говорит, что 5% наблюдений меньше 15, а 10% размещены между 15 и 24.

Так же как медиану интервального ряда вычисляли по формуле $\text{Me} = x_n + \dfrac{\frac{n}{2} - S_{\text{Me}-1}}{n_{\text{Me}}} h$, так и квантиль z_p уровня p вычисляют по формуле

$$z_p = x_{z_p} + \frac{np - S_{z_p - 1}}{n_{z_p}} h,$$

где x_{z_p} — нижняя граница интервала, содержащего квантиль z_p; $S_{z_p - 1}$ — частота, накопленная к интервалу, содержащему квантиль z_p; n_{z_p} — частота интервала, содержащего квантиль z_p; h — ширина интервала, содержащего квантиль z_p.

Пример 1. В таблице 1.50 представлены данные о возрасте преподавателей некоторого университета. Вычислить квантили $z_{0,25}$, $z_{0,5}$, $z_{0,75}$.

Таблица 1.50

Интервал	22—27	27—32	32—37	37—42	42—47	47—52
Частота	8	34	89	198	245	126
Накопленная частота	8	42	131	329	574	680
Интервал	52—57	57—62	62—67	67—72	72—77	
Частота	77	48	29	17	6	
Накопленная частота	757	805	834	851	857	

❐ Вычислим сначала $z_{0,25}$. Имеем: $n = 857$, $p = 0,25$, $[np] = 214$; $z_{0,25} = x_{215}$ помещается в интервале 37—42, $x_{z_{0,25}} = 37$, $S_{z_p - 1} = 131$, $n_{z_p} = 198$, $h = 5$. Поэтому

$$z_{0,25} = 37 + \frac{857 \cdot 0,25 - 131}{198} \cdot 5 \approx 39,1.$$

Аналогично для $z_{0,5}$ имеем: $n = 857$, $p = 0,5$, $[np] = 428$; $z_{0,5} = x_{429}$ помещается в интервале 42—47, $x_{z_{0,5}} = 42$, $S_{z_p - 1} = 329$, $n_{z_p} = 245$, $h = 5$. Поэтому

$$z_{0,5} = 42 + \frac{857 \cdot 0,5 - 329}{245} \cdot 5 \approx 44,0.$$

Для $z_{0,75}$ имеем: $n = 857$, $p = 0,75$, $[np] = 642$; $z_{0,75} = x_{643}$ помещается в интервале 47—52, $x_{z_{0,75}} = 47$, $S_{z_p - 1} = 574$, $n_{z_p} = 126$, $h = 5$. Поэтому

$$z_{0,75} = 47 + \frac{857 \cdot 0,75 - 574}{126} \cdot 5 \approx 49,7.$$

Итак, квартили 39,1; 44,0; 49,7 делят совокупность преподавателей университета по возрасту на четыре одинаковые по количеству группы: 22—39,1; 39,1—44,0; 44,0—49,7; 49,7—77. ■

? Укажите, сколько примерно наблюдений находится левее числа 39,1; правее числа 49,7.

Рассмотренной формулой можно воспользоваться для вычисления квантилей дискретного вариационного ряда.

Пример 2. В таблице 1.51 представлены данные о результатах тестирования 143 учащихся. Вычислить квантили $z_{0,2}$, $z_{0,4}$, $z_{0,6}$, $z_{0,8}$.

Таблица 1.51

Баллы	8	9	10	11	12	13	14	15	16
Частота	1	1	3	5	9	8	17	23	24
Накопленная частота	1	2	5	10	19	27	44	67	91
Баллы	17	18	19	20	21	22	23	24	25
Частота	18	10	6	6	5	4	0	2	1
Накопленная частота	109	119	125	131	136	140	140	142	143

❒ Для $z_{0,2}$ имеем: $n = 143$, $p = 0{,}2$, $[np] = 28$; $z_{0,2} = x_{29}$ помещается в той группе, где варианты равны 14, $x_{z_{0,2}} = 14$, $S_{z_p - 1} = 27$, $n_{z_p} = 17$, $h = 1$. Поэтому

$$z_{0,2} = 14 + \frac{143 \cdot 0{,}2 - 27}{17} \cdot 1 \approx 14{,}1.$$

Для $z_{0,4}$ имеем: $n = 143$, $p = 0{,}4$, $[np] = 57$; $z_{0,4} = x_{58}$ помещается в той группе, где варианты равны 15, $x_{z_{0,4}} = 15$, $S_{z_p - 1} = 44$, $n_{z_p} = 23$, $h = 1$. Поэтому

$$z_{0,4} = 15 + \frac{143 \cdot 0{,}4 - 44}{23} \cdot 1 \approx 15{,}6.$$

Для $z_{0,6}$ имеем: $n = 143$, $p = 0{,}6$, $[np] = 85$; $z_{0,6} = x_{86}$ помещается в той группе, где варианты равны 16, $x_{z_{0,6}} = 16$, $S_{z_p - 1} = 67$, $n_{z_p} = 24$, $h = 1$. Поэтому

$$z_{0,6} = 16 + \frac{143 \cdot 0{,}6 - 67}{24} \cdot 1 \approx 16{,}8.$$

Для $z_{0,8}$ имеем: $n = 143$, $p = 0{,}8$, $[np] = 114$; $z_{0,8} = x_{115}$ помещается в той группе, где варианты равны 18, $x_{z_{0,6}} = 18$, $S_{z_p - 1} = 109$, $n_{z_p} = 10$, $h = 1$. Поэтому

$$z_{0,8} = 18 + \frac{143 \cdot 0{,}8 - 109}{10} \cdot 1 \approx 18{,}5.$$

Итак, квантили 14,1; 15,6; 16,8; 18,5 делят совокупность тестировавшихся учащихся на пять одинаковых по количеству групп: 8—14,1; 14,1—15,6; 15,6—16,8; 16,8—18,5; 18,5—25. ■

При заданном значении p разность между квантилями z_{1-p} и z_p используют как показатель вариации. Этот показатель лучше характеризует вариацию совокупности по сравнению с размахом. Но, как и для размаха, при вычислении этого показателя не учитывается каждое отдельное значение.

Для вычисления квантилей с помощью Excel можно использовать статистические функции (КВАРТИЛЬ) и (ПЕРСЕНТИЛЬ).

Контрольные вопросы

1. В каких единицах выражен квантиль распределения?
2. Какому значению соответствует 0-й перцентиль, 100-й перцентиль, 50-й перцентиль?
3. Чему равен ранг медианы, если объем совокупности равен 25?
4. Чему равен ранг нижнего квартиля, если объем совокупности равен 13?
5. Чему равен ранг верхнего квартиля, если объем совокупности равен 13?

Задачи

47. На металлургическом заводе исследовали выход стали в 15 плавках определенного вида стали. Получили следующие результаты (в т):

70,3	85,0	100,0	78,1	77,9	98,4	59,2	86,8
70,1	42,2	81,9	97,1	68,2	92,1	91,2	

а)° Определите средний выход стали.

б)° Определите медиану выхода стали.

в) Постройте кумуляту распределения.

г) Найдите квартили этого набора данных.

д) Найдите 10-й и 90-й перцентили.

е) Руководство завода хотело бы, чтобы по крайней мере 90% плавок давали 75 т и больше стали. Соответствуют ли эти данные такому требованию? С каким перцентилем нужно проводить сравнение?

48. В баскетбольной секции университета измерили рост участников секции. Получили следующие результаты (в см):

193	180	186	188	186	190	186	195	194	184	178	184	199	182
185	177	189	186	179	188	179	181	180	187	186	189	184	186
198	188	182	186	177	176	190	193	186	192	188	182	195	193
181	179	185	171	174	176	183	174	175	184	176	183	178	182

а)° Определите средний рост участника секции.

б)° Определите медиану роста участников секции.

в) Постройте кумуляту распределения.

г) Найдите квартили этого набора данных.

д) Найдите 20-й и 80-й перцентили.

е) Руководство секции хотело бы, чтобы по крайней мере 80% участников секции имели рост 180 см и больше. Соответствуют ли эти данные такому требованию? С каким перцентилем нужно проводить сравнение?

Дополнительные задачи к главе 1

К § 1.1. Классификация данных и измерительные шкалы

49. В таблице 1.52 представлены данные прогноза погоды на каждый из трех дней.

Таблица 1.52

Дата	1.10.2005	8.10.2005	12.10.2005
Температура воздуха ночью, °C	11—13	7—9	6—8
Температура воздуха днем, °C	14—16	15—17	14—16
Давление, мм ртутного столба	743	748	745
Относительная влажность воздуха,%	46	48	77
Направление ветра	восточный	северо-восточный	восточный
Скорость ветра, м/с	3—6	1—3	1—3
Время восхода Солнца	6.24	6.34	6.40
Время заката Солнца	18.11	17.57	17.50
Влияние на здоровье	благоприятное	благоприятное	неблагоприятное

а) Что является элементарной единицей в этом наборе данных?

б) Это одномерные, двумерные или многомерные данные?

в) Это временной ряд или один временной срез?

г) Переменная «Направление ветра» — количественная, порядковая или номинальная?

д) Переменная «Скорость ветра» — количественная, порядковая или номинальная?

е) Переменная «Влияние на здоровье» — количественная, порядковая или номинальная?

ж) Переменная «Дата» — количественная, порядковая или номинальная?

з) Какие шкалы соответствуют переменным?

50. Ниже представлена таблица футбольного чемпионата Украины по состоянию на 5 октября 2005 г. (табл. 1.53).

Таблица 1.53

Место	Команда	Игры	Выигрыши	Ничьи	Поражения	Мячи	Очки
1	Шахтер	11	9	2	0	28—5	29
2	Динамо	11	8	3	0	30—7	27
3	Черноморец	11	6	2	3	17—9	20
4	Металлург Д	11	5	4	2	19—9	19
5	Волынь	11	5	2	4	15—21	17
6	Металлист	11	5	1	5	16—21	16
7	Харьков	11	4	3	4	18—24	15
8	Ильичевец	11	4	3	4	9—15	15
9	Арсенал К	11	3	5	3	11—13	14
10	Кривбасс	11	3	4	4	7—7	13
11	Таврия	11	4	0	7	10—16	12
12	Днепр	11	3	2	6	8—13	11
13	Сталь	11	2	5	4	8—17	11
14	Ворскла	11	2	4	5	10—17	10
15	Металлург З	11	2	3	6	14—22	9
16	Закарпатье	11	0	3	8	3—17	3

а) Что является элементарной единицей в этом наборе данных?

б) Это одномерные, двумерные или многомерные данные?

в) Это временной ряд или один временной срез?

г) Переменная «Место» — количественная, порядковая или номинальная?

д) Переменная «Очки» — количественная, порядковая или номинальная?

е) Переменная «Команда» — количественная, порядковая или номинальная?

ж) Переменная «Мячи» — количественная, порядковая или номинальная?

з) Какие шкалы соответствуют переменным? Имеются ли переменные, соответствующие шкале отношений?

и) На какие вопросы можно найти ответы при детальном анализе этого набора данных?

51. Для каждой из переменных, приведенных в предыдущем задании, определите, какие из перечисленных ниже операций можно применять к этой переменной:

а) сложение, вычитание;

б) подсчет количества;

в) ранжирование по порядку;

г) вычисление процента.

52. Найдите в Интернете, в статье из экономического журнала или в газете таблицу данных. Скопируйте эту таблицу.

а) Классифицируйте набор данных в соответствии с количеством переменных.

б) Что является элементарной единицей в этом наборе данных?

в) Это временной ряд или один временной срез?

г) Определите тип каждой переменной.

д) Укажите, какие из операций можно применять к каждой переменной.

е) Сформулируйте, на какие вопросы можно найти ответы при детальном анализе набора данных.

К § 1.2. Первичная обработка результатов измерений

53. Собраны данные о возрасте 25 сотрудников, занимающихся реализацией продукции некоторой компании:

50	42	32	35	41	44	24	46	31	47	36	32	30
44	22	47	31	56	28	37	49	28	42	38	45	

а) Определите, к какому виду относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты, накопленные частоты и относительные накопленные частоты значений.

в) Сгруппируйте данные.

г) Какую информацию можно извлечь из анализа этих данных?

54. Ниже приведены результаты (в %) выполнения 25 заданий основной части теста по математике 2400 выпускников начальной школы из 37 регионов России в 1999 году[1]:

89	94	95	76	56	83	84	72	86	94	77	82	69
56	65	90	75	83	70	70	52	74	72	90	82	

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты, накопленные частоты и относительные накопленные частоты значений.

в) Сгруппируйте данные.

г) Какую информацию можно извлечь из анализа этих данных?

55. Ниже приведены данные о числе ежегодных (каждый год состоялось по 52 розыгрыша) розыгрышей спортлото «6 из 49», в которых среди выигрышных номеров не было соседних:

26	28	26	23	24	18	29	24	22	21
19	24	23	17	19	23	27	20	14	20
25	24	25	21	15	21	24	22	26	26

а) Определите, к какому виду относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты, накопленные частоты и относительные накопленные частоты значений.

---

[1] Данные взяты из газеты «Математика», приложение к газете «Первое сентября», № 37, 2000 г.

в) Сгруппируйте данные.

г) Какую информацию можно извлечь из анализа этих данных?

56. Ниже приведены данные о численности 20 семейств:

4	6	2	8	3	5	6	4	7	2
5	6	4	6	9	4	7	5	6	3

а) Определите, к каким видам относятся данные, приведенные в условии: одномерные, двумерные или многомерные; качественные или количественные; временной ряд или один временной срез.

б) Упорядочьте эти данные по возрастанию, подсчитайте частоты, относительные частоты, накопленные частоты и относительные накопленные частоты значений.

в) Сгруппируйте данные.

К § 1.3. Вариационные ряды

57. Ниже приведены данные о числе ежегодных (каждый год состоялось по 52 розыгрыша) розыгрышей спортлото «6 из 49», в которых среди выигрышных номеров была точно одна пара соседних номеров:

27	16	19	21	18	22	17	22	24	23
19	24	23	24	21	24	19	24	28	22
20	21	19	24	30	21	26	21	17	17

а) Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда подсчитайте абсолютные и относительные плотности распределения.

г) Какую информацию можно извлечь из анализа этих данных?

58. Ниже приведены данные об объемах продаж (в условных денежных единицах) некоторой фирмы молочной продукции за неделю в течение 26 недель:

65 012	58 792	61 177	64 599	63 146	80 997	95 904
83 577	91 258	65 615	68 933	66 557	69 628	70 545
93 921	81 883	86 662	92 663	94 687	92 794	90 877
76 600	80 749	70 359	77 073	59 633		

а) Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда подсчитайте абсолютные и относительные плотности распределения.

г) Какую информацию можно извлечь из анализа этих данных?

59. По данным, приведенным в таблице 1.54, вычислите относительные и абсолютные плотности распределения работников цеха по возрасту.

Таблица 1.54

Возраст, годы	До 20	20—30	30—40	40—50	50 и более	Всего
Число работников	2	18	10	8	2	40

60. Ниже приведены результаты тестирования 75 взрослых по определению коэффициента интеллектуальности[1]:

141	92	100	132	97	110	106	107	105	83	127	95	109	108	104
104	87	133	118	124	111	135	110	110	127	114	105	102	92	94
101	115	124	98	118	138	97	101	116	112	113	95	102	131	121
130	91	92	101	146	121	108	129	113	114	106	105	102	86	107
148	96	123	107	107	129	108	105	123	105	139	106	89	134	103

а) Постройте дискретный вариационный ряд.

б) Постройте интервальный вариационный ряд.

в) Для интервального вариационного ряда подсчитайте абсолютные и относительные плотности распределения.

г) Какую информацию можно извлечь из анализа этих данных?

К § 1.4. Графическое изображение вариационных рядов

61. Постройте полигоны частот для дискретных вариационных рядов, составленных в заданиях 57 и 58.

62. Постройте гистограммы для интервальных вариационных рядов, составленных в заданиях 57 и 58.

63. По данным, представленным в задании 59 (табл. 1.54):

а) Постройте гистограмму, полигон частот, кумуляту;

б) Укажите, является ли распределение симметричным.

---

[1] Данные взяты из книги: *Дж. Гласс*, *Дж. Стэнли*. Статистические методы в педагогике и психологии. — М.: Прогресс, 1976.

64. Были собраны следующие данные о коррозии сплавов, которая определялась по потере веса в миллиграммах на квадратный дециметр в день[1]:

127,6	124,0	110,8	103,9	101,5	130,1	122,0
92,3	113,1	83,7	128,0	91,4	86,2	

а) Постройте дискретный вариационный ряд.
б) Постройте полигон частот и кумуляту.
в) Охарактеризуйте распределение данных.

К § 1.5. Среднее арифметическое — показатель центральной тенденции

65. Активные сборщики макулатуры составляют 15% учащихся школы и собирают ежегодно в среднем 23 кг макулатуры. Пассивные сборщики макулатуры составляют 27% учащихся школы и собирают ежегодно в среднем 8 кг макулатуры. Остальные сборщики макулатуры собирают ежегодно в среднем 13 кг макулатуры. Сколько килограммов макулатуры собирает в среднем учащийся школы?

66. В производстве мыла имеет важное значение высота мыльной пены в лотках. Проведен эксперимент, в котором варьировалась масса мыла и измерялась высота пены в стандартном лотке при определенной степени перемешивания. Получены данные[1] (табл. 1.55).

Таблица 1.55

Масса мыла, г	4,0	4,5	5,0	5,5	6,0	6,5	7,0
Высота пены	33	42	45	51	53	61	62

а) Какова примерная зависимость между массой мыла и высотой пены? Постройте график по точкам, координатами которых являются приведенные данные.

б) Подсчитайте среднюю массу мыла в лотке и среднюю высоту пены. Отражают ли средние значения наблюдаемую зависимость между рассматриваемыми параметрами?

---

[1] Данные взяты из книги: *Н. Драйпер, Г. Смит.* Прикладной регрессионный анализ. — М.: Статистика, 1973.

67. Считается, что стоимость эксплуатации транспортных винтовых самолетов растет с возрастом самолетов. Собраны данные (табл. 1.56).

Таблица 1.56

Возраст, годы	4,5	4,5	4,5	4,0	4,0	4,0	5,0	5,0	5,5
Стоимость эксплуатации, у. е.	619	1049	1033	495	723	681	890	1522	987
Возраст, годы	5,0	0,5	0,5	6,0	6,0	1,0	1,0	1,0	
Стоимость эксплуатации, у. е.	1194	163	182	764	1373	978	466	549	

а) Какова примерная зависимость между возрастом самолетов и стоимостью их эксплуатации? Постройте график по точкам, координатами которых являются приведенные данные.

б) Подсчитайте средний возраст самолетов и среднюю стоимость эксплуатации. Отражают ли средние значения наблюдаемую зависимость между рассматриваемыми параметрами?

К § 1.6. Другие меры центральной тенденции

68. В некотором рабочем процессе для пяти рабочих определены расходы времени на одно изделие в минутах (табл. 1.57). Четверо рабочих работают по 8 ч, а пятый — 4 ч. Рассчитайте среднее время, необходимое для производства одного изделия для группы из пяти рабочих.

Таблица 1.57

№	**Рабочее время, мин**	**Время на одно изделие, мин**
1	480	0,8
2	480	1,0
3	480	1,2
4	480	1,2
5	240	1,5

69. Начальный вклад в 4 млн р. за 4 года возрос до 5 млн р. Определите среднюю скорость роста вклада.

70. За сколько лет банковский вклад при процентной ставке $p\%$ удваивается?

71. Рассмотрим данные о размере платы за ссуду под залог дома, представленные в таблице 1.58. Плата за ссуду указана как процент от величины ссуды и представляет собой одноразовый платеж при возврате ссуды.

Таблица 1.58

Фирма	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Плата за ссуду	2,5	4	2	4	4	3	4	2	4	2,5	4	2	3	4

а) Найдите среднее значение платы за ссуду.

б) Найдите медиану платы за ссуду.

в) Найдите моду.

г) Какой из показателей (среднее, медиана, мода) наиболее полезен при описании меры центральной тенденции значения платы за ссуду?

72. Имеются данные о длине пробега 25 автомобильных шин (в км):

1420	1443	1448	1472	1512	1481	1488	1492	1508	1456
1528	1535	1538	1543	1549	1551	1557	1568	1571	1586
1592	1595	1587	1603	1628					

а) Постройте интервальный вариационный ряд.

б) Постройте гистограмму и кривую распределения.

в) Вычислите среднее и медиану для дискретного и интервального вариационного ряда.

г) Какой из показателей (среднее, медиана) наиболее полезен для описания типического значения длины пробега?

73. Восемь друзей живут вдоль шоссе, которое проложено в лесу. Расположение их домов показано на рис. 15. Они являются членами клуба туристов. Стоимость бензина оплачивается из казны клуба. В каком месте шоссе им необходимо собраться на пикник,

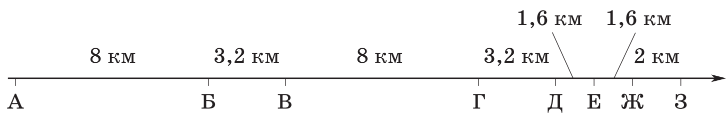


Рис. 15

чтобы израсходовать на путешествие минимальное количество денег на бензин? Любое место в лесу у шоссе является прекрасным местом для пикника.

К § 1.7. Показатели вариации

74. По данным, приведенным в задании 71:

а) Найдите стандартное отклонение и дайте краткую интерпретацию этой величины.

б) Найдите размах данных и дайте краткую интерпретацию этой величины.

в) Найдите коэффициент вариации и дайте краткую интерпретацию этой величины.

г) Определите, сколько данных не выходит за пределы одного стандартного отклонения.

д) Определите, сколько данных не выходит за пределы двух стандартных отклонений.

75. По данным, приведенным в задании 72:

а) Найдите стандартное отклонение и дайте краткую интерпретацию этой величины.

б) Найдите размах данных и дайте краткую интерпретацию этой величины.

в) Найдите коэффициент вариации и дайте краткую интерпретацию этой величины.

г) Определите, сколько данных не выходит за пределы одного стандартного отклонения.

д) Определите, сколько данных не выходит за пределы двух стандартных отклонений.

76. В двух бригадах насчитывается по 7 человек. В первой бригаде месячная зарплата двух рабочих составляет по 12 600 р., трех рабочих — по 14 400 р., один рабочий зарабатывает 15 300 р. и один рабочий — 16 200 р. Во второй бригаде трое рабочих — по 12 600 р., один — 14 580 р., двое — по 15 300 р. и один — 15 390 р. Какая бригада более однородна по оплате труда?

К § 1.8. Квантили

77. По данным, приведенным в задании 72:

а) Определите среднюю длину пробега шин.

б) Определите медиану длины пробега шин.

в) Постройте кумуляту распределения.

г) Найдите квартили этого набора данных.

д) Найдите 30-й и 70-й перцентили.

е) Выясните, соответствуют ли эти данные требованию о том, чтобы не менее 70% шин имело бы длину пробега не менее 1580 км.

ж) Укажите, с каким перцентилем нужно проводить сравнение для ответа на предыдущий вопрос.

78. По данным, приведенным в задании 71:

а) Определите среднюю плату за ссуду.

б) Определите медиану платы за ссуду.

в) Постройте кумуляту распределения.

г) Найдите квартили этого набора данных.

д) Найдите 20-й и 80-й перцентили.

Глава 2

СЛУЧАЙНЫЕ СОБЫТИЯ

Во введении было рассмотрено достаточное количество примеров реальных явлений и ситуаций, исход которых зависит от случая. Мир, в котором мы живем, очень сложен. Присущие каждому явлению закономерности прокладывают себе путь через хаос случайностей. Для многих явлений влияние случая настолько существенно, что их изучение и практическое использование невозможны без исследования и количественной оценки этого влияния.

Человеку необходимо уметь разбираться в случайных ситуациях, по крайней мере настолько, насколько это возможно. По всей видимости, мы никогда не сможем «совершенно точно» сказать, что случится в будущем. Однако те, кто осознают, что одни возможности являются более правдоподобными, чем другие, и умеют количественно описывать эти отношения, имеют преимущества по сравнению с теми, кто полагается лишь на свои ощущения, не имея для этого объективных предпосылок. Лучше объединять понимание вероятностей со всеми доступными знаниями и опытом.

Цель настоящей главы — научиться количественно измерять случайность. Случайность будет измеряться подобно тому, как измеряется длина, угол, площадь, скорость, сила и т. д. Мерой случайности является *вероятность*. Интуитивно часто можно ответить на вопрос, какова вероятность того или иного события. Значительно трудней ответить на вопрос о том, что же такое вероятность. И это не случайно — физическую или геометрическую величину легче измерить, чем объяснить, а что же такое длина, площадь, скорость, сила тока и т. д.

В этой главе исследование случайных событий начнется с тщательного определения и ограничения ситуации. Результатом такого определения является понятие *случайного опыта*. При этом составляют перечень всех возможных результатов — *выборочное* *пространство*, или *пространство элементарных исходов*. Выделяют ряд особых случаев, которые следует рассмотреть и которые либо будут наблюдаться, либо нет; их называют *событиями*. Правдоподобие наступления события характеризуют конкретным числом, которое называют *вероятностью*.

Итак, в данной главе рассматриваются три основные понятия теории вероятностей — *случайный опыт, случайное событие и вероятность*. Будут строиться математические модели явлений, позволяющие учитывать влияние случая. Умение создавать математические, в частности вероятностные, модели реальных явлений и процессов составляет важный аспект математической культуры современного человека. Будут рассмотрены различные источники получения вероятностей событий, различные подходы к определению этого понятия. Это могут быть и статистические данные, и комбинаторные знания, и суждение эксперта.

При этом может возникнуть потребность объединить информацию о более чем одном событии (*операции над событиями* и *соответствующие теоремы сложения и умножения вероятностей*), необходимость с течением времени обновлять имеющиеся значения вероятностей с помощью так называемых *условных вероятностей*, отражающих доступную информацию. При исследовании случайных явлений будет установлена зависимость между некоторыми из них.

§ 2.1. Статистическая вероятность

2.1.1. Случайный опыт

Для многих процессов, происходящих в экономике, в природе, на производстве и в повседневной жизни, невозможно заведомо с уверенностью предсказать исход того или иного явления. Для подтверждения этого приведем примеры таких сфер деятельности, где наблюдаются подобные процессы:

- страховое дело;
- инвестиционная деятельность банков;
- грузовые работы в морском порту;
- игра в рулетку;
- распространение эпидемий;
- контроль качества продукции;
- оказание скорой медицинской помощи.

Этот перечень можно продолжать сколько угодно. Но и этих примеров достаточно для того, чтобы увидеть, что общего в таких сферах деятельности. Во-первых, соответствующие действия можно повторять многократно, примерно в одинаковых условиях (застраховать человека; вложить средства в определенный проект; разгружать судна одного класса; играть в рулетку; распространять инфекционное заболевание; проверять качество изделий, изготовленных на одном станке; принимать заявки на оказание скорой медицинской помощи). Во-вторых, исходы этих действий нельзя предсказать однозначно (страхование может закончиться тем, что произойдет или не произойдет событие, от которого страхуется человек; вложенные средства могут быть возвращены или не возвращены; различно время разгрузки суден; инфекционное заболевание может быть локализовано быстро или эпидемия продлится длительное время; деталь может оказаться бракованной или удовлетворяющей стандарту; различно время поездки машины «скорой помощи» к больному).

Все приведенные в этих примерах явления имеют между собой нечто общее, а именно: результат отдельных испытаний здесь предсказать невозможно. Поэтому их называют *случайными*.

Под ***случайным опытом****, или* ***случайным испытанием****, будем понимать любое действие, которое можно повторить большое количество раз в приблизительно одинаковых условиях и результаты которого предсказать невозможно.*

Приведем еще несколько примеров случайных опытов: одноразовое или двукратное подбрасывания монеты; приобретение лотерейного билета; стрельба по мишени.

В дальнейшем будут использоваться как синонимы следующие термины: ***случайный опыт, случайный эксперимент*** и ***случайное испытание***.

Проанализируем требования к случайным испытаниям. Во-первых, их можно проводить *многократно*. Так, можно много раз покупать лотерейный билет, много раз бросать игральный кубик, много раз извлекать шары из ящика, стрелять по мишени. Однако к таким испытаниям нельзя отнести, например, войну между двумя государствами. Вряд ли она будет происходить много раз. Многократно нельзя повторить запуск космического корабля: это очень дорогое мероприятие. Эти испытания не относят к случайным.

Во-вторых, случайные испытания проходят примерно *в одинаковых условиях*. Не меняется центр тяжести кубика при повторных его бросках. В одной и той же лотерее заранее определено число выигрышей, и оно не меняется по мере распространения билетов до проведения тиража. По мишени стреляет один и тот же стрелок из одного и того же оружия, положение мишени не меняется. В отличие от них, повторные футбольные матчи между двумя командами могут проходить при измененных условиях: может измениться степень готовности той или иной команды к игре, могут измениться погодные условия, состояние футбольного поля и т. п. Меняются условия, при которых один и тот же ученик пишет контрольные работы по математике в течение учебного года: может измениться сложность заданий, их количество, готовность ученика к выполнению контрольной работы. Последние два испытания не относят к случайным.

В-третьих, исходы случайных испытаний *неоднозначны*. Так, при бросании игрального кубика заранее неизвестно, какое число очков выпадет. При покупке лотерейного билета заранее неизвестно, выпадет или не выпадет на него выигрыш, а если выпадет, то какой. При извлечении шара из ящика неизвестно, какого цвета шар будет извлечен. Однако есть такие испытания, исходы которых заранее известны. Так, заранее можно сказать, что при нагревании воды в чайнике при 100 °С она закипит. Точно также заранее известно, что на смену дню приходит ночь, после лета наступает осень. При ходьбе человек после шага левой ногой обязательно делает шаг правой. Нагревание воды, смену дня и ночи, ходьбу не считают случайными испытаниями.

Обратите внимание на то, что *только наличие всех трех условий делает испытание случайным*. Например, результаты олимпийских игр по спортивной гимнастике мы не можем предсказать однозначно, многое зависит от случая. Однажды во время выступления претендентки на первое место развалились брусья. Могут быть неожиданные травмы, болезни, кто-то не обрел еще свою лучшую форму. И все же олимпийские игры не являются случайными испытаниями: их нельзя повторять многократно в одних и тех же условиях.

?

Какие из следующих испытаний можно считать случайными:
а) стрельба по мишени;
б) нагревание воды в чайнике;
в) покупка лотерейного билета;
г) вращение рулетки в игре «Поле чудес»;
д) поступление юноши в лицей;
е) многолетние наблюдения за погодой в один и тот же день в одной и той же местности;
ж) подбрасывание кнопки;
з) участие команды «Локомотив» в первенстве России по футболу?

Таким образом, случайный эксперимент — это четко определенная процедура, результат которой можно наблюдать, но невозможно точно предсказать заранее. Каждый случайный эксперимент характеризуется ***выборочным пространством***, представляющим собой перечень всех возможных результатов этого эксперимента, описанных заранее, без знания того, что действительно произойдет в ходе эксперимента. Такой подход позволяет сделать случайную ситуацию более определенной и часто помогает также прояснить свое представление о ней. Приведем ряд примеров.

Пример 1. Планируется исследование доходов семей, проживающих вблизи места, где предполагается открыть новый супермаркет. Случайным образом выбирают телефонный номер, по которому звонят в семью и записывают ее доход с точностью до 10 денежных единиц. В случае, если никто не ответил на звонок или если собеседник не захотел ответить на заданный вопрос о доходе, выбирают новый номер и звонки повторяют до тех пор, пока не будет получен конкретный ответ о размере дохода. Выборочное пространство представляет собой список возможных значений дохода: 0, 10, 20, 30, ..., 150 000, 150 010, ... ■

Пример 2. Исследуется вопрос о скорости чтения выпускника начальной школы, т. е. о количестве слов, прочитываемых учащимся за минуту. Выборочное пространство представляет собой перечень целых неотрицательных чисел от 0 до 300. ■

Пример 3. В финале конкурса «Мисс Европа» участвуют 10 девушек, из которых победительницей будет только одна. Выборочное пространство состоит из 10 участниц финала: А, Б, В, Г, Д, Е, Ж, З, И, К. ■

Пример 4. Качество пяти бизнес-проектов оценивается по четырехбалльной системе целым числом от 1 до 4. Выборочное пространство представляет собой набор всех возможных вариантов оценок качества. Это набор $4^5 = 1024$ списков, каждый из которых содержит пять чисел: (1, 1, 1, 1, 1), (1, 1, 1, 1, 2), (1, 1, 1, 1, 3), (1, 1, 1, 1, 4), ..., (4, 4, 4, 4, 4). ■

2.1.2. Случайное событие

Выше мы показали, что каждый случайный опыт может закончиться различными исходами. Эти исходы могут указываться описательно или с помощью элементов выборочного пространства.

Рассмотрим, например, опыт, состоящий в однократном бросании игрального кубика. Его исходами является выпадение любого числа очков от 1 до 6. Вместе с тем этот опыт может закончиться выпадением четного или нечетного числа очков, его результатом можно считать выпадение числа очков, меньшего 3, или большего 2, кратного 3, являющегося простым или составным числом. Все эти исходы будем считать *случайными событиями*.

Любой исход случайного опыта назовем ***случайным событием.***

В результате такого опыта случайное событие может или произойти, или не произойти. Случайные события будем обозначать прописными латинскими буквами A, B, C, ... Случайными событиями являются, например, «выпадение двух гербов» при подбрасывании двух монет, «попадание в цель» при выстреле, «выигрыш» при приобретении лотерейного билета.

Итак, каждый раз выполнение случайного эксперимента дает ровно один результат из возможных элементов выборочного пространства. Поскольку выборочное пространство содержит все возможные результаты, то не должно быть никаких неожиданностей: результат эксперимента должен содержаться в выборочном пространстве. Вот результаты одного выполнения каждого из рассмотренных в примерах 1—4 случайных экспериментов:

- в исследовании дохода семьи после одного звонка, в результате которого был получен ответ «это вас не касается», удалось дозвониться до человека, который назвал доход семьи 15 300;
- при изучении скорости чтения был получен результат 115 слов в минуту;
- в конкурсе «Мисс Европа» победила участница 3;
- при исследовании качества бизнес-проектов был получен список оценок качества для каждого из пяти оцененных проектов (3, 2, 1, 4, 2).

?

Для случайных испытаний из предыдущего задания приведите примеры случайных событий.

Акционерное общество в ближайшие дни должно объявить о размере своих доходов за последний квартал. Но рядовые акционеры не знают о размере доходов.

а) Опишите соответствующий данному случаю случайный эксперимент.

б) Что представляет собой выборочное пространство?

в) О чем могут свидетельствовать результаты эксперимента?

г) Точно определите событие «заявленная величина доходов выше ожидаемой», предварительно предсказав ожидаемую величину дохода.

2.1.3. Относительная частота события

Если много раз повторять случайный эксперимент, то одни события при этом будут происходить чаще других. Поэтому при одноразовом проведении эксперимента с бо́льшим основанием можно надеяться на появление первых событий, чем вторых. Так, для отличного стрелка ожидать попадания при одном выстреле можно с бо́льшим основанием, чем промаха.

Предположим, что при неизменных условиях проведено n случайных опытов и в $n(A)$ из них произошло событие A. Число $n(A)$ называют ***частотой события** A*. Отношение $\frac{n(A)}{n}$ называют ***относительной частотой события** A*. (Сравните эти понятия с частотой и относительной частотой, с которыми встречается в вариационном ряде та или иная варианта.) Например, если в некотором тексте из 2000 знаков буква О встретилась 182 раза, то частота и относительная частота события «встретилась буква О» составляет соответственно 182 и $\frac{182}{2000} = 0{,}091$.

Относительную частоту события иногда выражают в процентах: $\frac{n(A)}{n} \cdot 100\%$. В нашем примере она равна 0,091, или $0{,}091 \cdot 100\% = 9{,}1\%$.

?

Обозначим число проведенных опытов через n, количество появлений события в этих опытах — через m, относительную частоту события $\frac{m}{n}$ — через r. Найдите:

а) r, если $n = 400$, $m = 160$;
б) n, если $r = 0{,}3$, $m = 45$;
в) m, если $r = 0{,}85$, $n = 180$.

Известно, что после 200 подбрасываний монеты относительная частота выпадения герба равнялась 0,5. В каких пределах могла находиться относительная частота того же события после первых 100 подбрасываний?

При 100 бросках игрального кубика «шестерка» выпала 15 раз. Укажите, в каких пределах может находиться относительная частота появления «шестерки» после того, как будет выполнено еще 100 бросков кубика?

Относительная частота является короткой и содержательной характеристикой рассматриваемой информации. Например, если баскетболист за игру сделал 40 бросков мячом в корзину, 30 из которых оказались результативными, то относительная частота события «баскетболист при броске попал в корзину» (она равняется в этом примере 0,75) является показателем результативности этого спортсмена. Полное описание всех бросков со всеми обстоятельствами, которые сопровождали каждый бросок, заняло бы много страниц текста. Относительная частота попаданий в мишень для стрелка описывает его мастерство. Относительная частота бракованных деталей в некоторой партии оценивает качество продукции.

Если для некоторого стрелка относительная частота попадания в мишень равна, например, 93% , то с уверенностью мы можем сказать, что это хороший стрелок. Если в некоторой партии электрических лампочек, выпущенных предприятием, 1,6% бракованных, то с таким качеством продукции смириться можно.

? Зная относительную частоту события, охарактеризуйте соответствующее явление:

а) относительная частота реализованных пенальти для данного футболиста равна 0,8;

б) относительная частота пенальти, взятых данным вратарем равна 0,2;

в) относительная частота рабочих дней на предприятии, в которых достигалась намеченная цель, равна 0,88;

г) относительная частота обнаружения бракованной детали с помощью рентгена равна 0,8.

2.1.4. Статистическая устойчивость опытов

Относительная частота события, вообще говоря, изменяется, если изменяется количество опытов или выполняется другая серия из такого же количества опытов. Известны результаты эксперимента с подбрасыванием монеты, проведенного исследователем Дж. Э. Беррихом. Он провел 10 серий, каждая из которых состояла из 1000 подбрасываний. Результаты эксперимента приведены в таблице 2.1.

Видно, что от серии к серии относительная частота рассматриваемого события изменяется. Но она колеблется вокруг числа 0,5, хотя ни одно ее значение не равняется в точности 0,5.

Итак, есть немало случайных опытов, в которых *относительная частота события при повторении серии из достаточно большого количества опытов колеблется около одного и того же числа. Такие опыты называют* ***статистически устойчивыми.***

Классическими примерами статистически устойчивых опытов являются эксперименты, связанные с подбрасыванием монеты. Их результаты приведены в таблице 2.2.

Эти данные показывают, что предположение о равенстве 0,5 относительной частоты появления любой стороны монеты хорошо согласуется с опытом.

Еще в XVIII веке было замечено, что среди обычной корреспонденции письма без адреса обладают определенной устойчивостью. Данные, приведенные в таблице 2.3[1], собранные по результатам русской почтовой статистики, свидетельствуют о том, что на протяжении нескольких лет на каждый миллион писем приходилось в среднем 25—27 писем без адреса.

Таблица 2.1

№ серии	1	2	3	4	5
Число выпадений герба	502	511	497	529	504
Относительная частота события «выпал герб»	0,502	0,511	0,497	0,529	0,504
№ серии	6	7	8	9	10
Число выпадений герба	476	507	528	504	529
Относительная частота события «выпал герб»	0,476	0,507	0,528	0,504	0,529

Таблица 2.2

Исследователь	**Число подбрасываний монеты**	**Число выпадений герба (событие *A*)**	**Относительная частота события *A***
Ж. Бюффон	4040	2048	0,5069
Де Морган	4092	2048	0,5005
К. Пирсон	12 000	6019	0,5016
В. Феллер	10 000	4979	0,4979
К. Пирсон	24 000	12 012	0,5005
В. Романовский	80 640	40 151	0,4979

[1] Данные взяты из книги: *А. Н. Колмогоров, И. Г. Журбенко, А. В. Прохоров.* Введение в теорию вероятностей. — М.: Наука, 1982.

Таблица 2.3

Год	Число писем	Число писем без адреса	Относительная частота писем без адреса
1906	983 000 000	26 112	0,0000266
1907	1 076 000 000	26 977	0,0000251
1908	1 214 000 000	33 515	0,0000276
1909	1 357 000 000	33 643	0,0000248
1910	1 507 000 000	40 101	0,0000266
Всего	6 137 000 000	160 348	0,0000261

Приведенные данные свидетельствуют о том, что относительные частоты события «случайно взятое письмо не содержит адреса» ежегодно колебалось около числа 0,0000261.

Случайным испытанием, имеющим статистически устойчивый характер, является рождение ребенка. Ниже в таблице 2.4[1] приведены данные о распределении пола детей, родившихся в Швеции в 1935 г.

Можно утверждать, что относительные частоты рождения мальчиков в Швеции в 1935 г. ежемесячно колебались вокруг числа 0,518.

Устойчивую частоту имеют метеорологические явления. Так, наблюдения над числом солнечных дней на протяжении года в Саратове дают следующие результаты (табл. 2.5)[2].

Видно, что относительные частоты события «день солнечный» колеблются вокруг числа 0,583.

Статистически устойчивыми опытами являются эксперименты Г. Менделя. По теории Менделя при скрещивании желтого гороха с желтым примерно в одном случае из четырех получают зеленый горох. Для проверки этой теории опыт по скрещиванию желтого гороха с желтым был проведен 34 153 раза. В 8506 случаях получали зеленый горох. Относительная частота события «появление зеленого гороха» в проведенной серии экспериментов равна $\frac{8506}{34\,153} \approx$ $\approx 0{,}249 \approx 0{,}25$. Подобные опыты повторялись неоднократно, и каж-

[1] Данные взяты из книги: *Г. Крамер*. Математические методы статистики. — М.: Мир, 1975.

[2] Данные взяты из книги: *Г. П. Боев*. Теория вероятностей. — М.-Л.: Гостехиздат, 1950.

Таблица 2.4

Месяц	Общее число детей	Мальчики		Девочки	
		Число	Относительная частота	Число	Относительная частота
1	7280	3743	0,514	3537	0,486
2	6957	3550	0,510	3407	0,490
3	7883	4017	0,510	3866	0,490
4	7884	4173	0,529	3711	0,471
5	7892	4117	0,522	3775	0,478
6	7609	3944	0,518	3665	0,482
7	7585	3964	0,523	3621	0,477
8	7393	3797	0,514	3596	0,486
9	7203	3712	0,515	3491	0,485
10	6903	3512	0,509	3391	0,491
11	6552	3392	0,518	3160	0,482
12	7132	3761	0,527	3371	0,473
Всего	88 273	45 682	0,518	42 591	0,482

Таблица 2.5

Год	1920	1921	1922	1923	1924
Число солнечных дней	203	215	243	194	210
Относительная частота события «день солнечный»	0,555	0,589	0,666	0,522	0,574

дый раз, когда их количество было достаточно большим, относительная частота события «появление зеленого гороха» мало отличалась от 0,25.

Исследования различных произведений русской литературы позволили сделать выводы об относительных частотах употребления различных букв русского алфавита. Эти данные представлены в таблице 2.6 (тире в ней означает пробел)[1].

[1] Данные взяты из книги: *А. М. Яглом*, *И. М. Яглом*. Вероятность и информация. — М.: Наука, 1973.

Таблица 2.6

Буква	—	о	е, ё	а	и	т	н	с
Относительная частота	0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
Буква	р	в	л	к	м	д	п	у
Относительная частота	0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
Буква	я	ы	з	ь, ъ	б	г	ч	и
Относительная частота	0,018	0,016	0,016	0,014	0,014	0,013	0,012	0,010
Буква	х	ж	ю	ш	ц	щ	э	ф
Относительная частота	0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

Таким образом, исследования показывают, что в среднем из 1000 наудачу выбранных в тексте пробелов и букв на трех местах будет стоять буква «э», на тридцати пяти — буква «л», на девяноста — буква «о».

Оказывается, что у каждого автора имеется своя таблица распределения относительных частот использования букв, слов, словосочетаний и предлогов. По этим таблицам можно установить автора примерно так же точно, как и преступника по отпечаткам пальцев (см. § 1.5).

Рассмотренные и подобные опыты позволяют сделать вывод о том, что относительная частота событий, подсчитанная по результатам статистически устойчивых опытов, мало отличается от некоторого числа. И чем больше число испытаний, тем реже встречаются значительные отклонения относительной частоты события от этого числа.

Число, около которого колеблется относительная частота события в сериях из достаточно большого количества статистически устойчивых опытов, называют ***вероятностью*** *этого события.*

Приведенное определение вероятности часто называют ***статистическим***, а саму вероятность — ***статистической***. Вероятность события A будем обозначать символом $P(A)$ (от латинского слова *probabilitas* — «вероятность»).

Итак, статистическую вероятность находят с помощью большого числа опытов. Проводя опыты и наблюдения, мы принимаем относительную частоту за приближенное значение вероятности. При большом числе статистически устойчивых опытов это приближение вполне достаточно для оценивания риска и шансов. Кстати, и большинство физических величин измеряется приближенно.

Во многих физических, астрономических и других естественнонаучных задачах число выполняемых испытаний очень велико, и опыт может дать значение вероятности события с довольно высокой точностью. Например, для распада радия, где число испытаний огромно — равно числу атомов в испытуемом образце и составляет примерно 10^{23}—10^{24}, — можно утверждать, что вероятность распада атома радия за 100 лет равна 0,04184.

При статистической оценке вероятности события необходимо, чтобы условия испытаний не менялись. Если, например, вероятность выдачи конвейером бракованной детали определяется при помощи относительной частоты бракованных изделий на протяжении ряда лет, то нужно иметь в виду, что, несмотря на кажущееся сохранение условий эксперимента, на самом деле они будут меняться: отдельные части конвейера со временем изнашиваются, меняются условия обработки детали. В этом случае нельзя говорить о статистической устойчивости частот, вероятность появления бракованной детали будет меняться со временем.

Наличие у события при определенных условиях вероятности, равной p, проявляется в том, что почти в каждой достаточно длинной серии опытов относительная частота события приближенно равна p.

?

По результатам какой серии опытов можно оценить вероятность изготовления бракованной детали?

а) проверено 200 деталей, изготовленных на одном станке в течение одной рабочей смены;

б) проверено 100 деталей, изготовленных на 20 станках в течение рабочей смены;

в) проверено 150 деталей, изготовленных на одном станке в течение месяца;

г) проверено 8 деталей, изготовленных на одном станке в течение месяца.

По результатам какой серии опытов можно оценить вероятность попадания в цель при одном опыте?

а) 100 стрелков по одному разу выстрелили в мишень;

б) стрелок в течение короткого времени 50 раз выстрелил в мишень, расположенную в закрытом помещении;

в) стрелок в течение короткого времени 5 раз выстрелил в мишень, расположенную в закрытом помещении;
г) стрелок в течение светового дня на открытом стрельбище сделал 300 выстрелов в мишень.

Укажите, кто сделал верный вывод:
а) Купив два лотерейных билета, среди которых оказался один выигрышный, Сергей сделал вывод о том, что вероятность выигрыша в этой лотерее равна 0,5.
б) Подбросив 4 раза монету и увидев, что герб выпал 3 раза, Анатолий сделал вывод о том, что вероятность выпадения герба при одном бросании монеты равна 0,75.
в) Узнав, что в последних 20 тиражах лотереи «5 из 36» все пять номеров угадывались 6 раз, Евгений сделал вывод о том, что вероятность угадать 5 номеров из 36 равна 0,3.
г) Подбросив 100 раз кнопку и подсчитав, что острием вверх она упала 45 раз, Владимир сказал, что вероятность того, что кнопка упадет острием вверх, приближенно равна 0,45.

На рис. 16 изображен график зависимости относительной частоты $\frac{n(\text{Г})}{n}$ появления герба при 12 подбрасываниях монеты с результатами ГГЦГЦЦГГЦЦЦГ (Г — появление герба, Ц — появление цифры) от числа подбрасываний n. При $n > 10$ относительная частота колеблется около числа 0,5, которое и принимают в качестве вероятности события «выпадение герба при одном подбрасывании монеты».

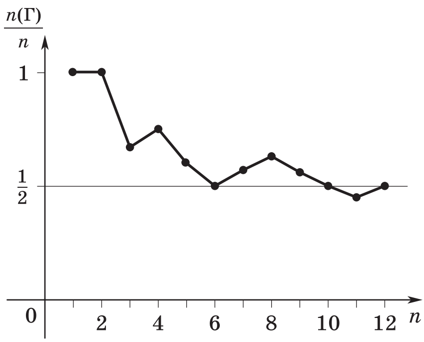


Рис. 16

Проведите опыт с бросанием игрального кубика 30 раз, каждый раз фиксируя исход опыта, и постройте график зависимости относительной частоты события «выпало четное число очков» от числа бросков.

Итак, *если говорят, что вероятность некоторого события равна, например, 0,82, то это практически означает, что в среднем в каждых 100 опытах, проведенных примерно в одинаковых условиях, это событие происходит примерно 82 раза.*

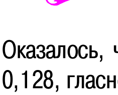
Таким образом, каждому событию соответствует число от 0 до 1, называемое вероятностью и описывающее, насколько правдоподобно наступление данного события при каждом выполнении случайного эксперимента.

2.1.5. Применение статистической вероятности

Статистическое определение вероятности может быть использовано для получения практически важных выводов.

Пример 5. Некоторое предприятие производит массовую продукцию. Если изделие, поступившее в продажу, выходит из строя на протяжении одного года, то его заменяют запасным. Сколько запасных изделий необходимо изготовить, если в течение года продается N изделий? Для упрощения предположим, что выход из строя запасного изделия невозможен.

❒ Будем считать, что событие A — «отказ изделия» — имеет некоторую вероятность p. Пусть в течение года вышло из строя m изделий. Тогда $\frac{m}{N}$ — относительная частота события A. Если N — достаточно большое число, то $\frac{m}{N} \approx p$. Отсюда $m \approx Np$. На практике довольно часто вероятность p неизвестна. Для ее оценки (приближенного вычисления) в течение года проводят некоторое число n испытаний изделий. Если за это время откажет k изделий, то $\frac{k}{n}$ — относительная частота события A. При достаточно большом значении n можем считать, что $\frac{k}{n} \approx p$. Тогда $m \approx \frac{Nk}{n}$. ■



Всемирно известный математик, наш соотечественник А. А. Марков (1856–1922) исследовал чередование гласных и согласных букв в последовательности, содержащей 20 000 букв из «Евгения Онегина» А. С. Пушкина. Оказалось, что относительная частота появления гласной буквы после гласной равна 0,128, гласной буквы после согласной – 0,663. Эти относительные частоты можно принять в качестве вероятностей соответствующих событий для произведений А. С. Пушкина.

В качестве примера рассмотрим страхование жизни. Представим себе, что некий мужчина в возрасте 55 лет хочет застраховать свою жизнь, например, на 10 лет. Чтобы знать, какой взнос потребовать от застрахованного и какую сумму выплатить родственникам в случае его смерти, надо рассчитать, сколько мужчин в возрасте 55 лет останется в живых через 10 лет и сколько из них умрет в течение этого срока. Кажется, что здесь ничего нельзя предвидеть – ведь никто не знает, когда он умрет. Но оказывается, что при рассмотрении большого числа случаев можно уверенно предсказать то число мужчин, которое останется в живых после 65 лет. Для этого изучается так называемая статистика смертности, и предсказание тем точнее, чем большее количество людей охвачено страхованием.

Аналогично этому можно заранее рассчитать необходимое в данном городе число пожарных команд или предсказать запасы зерна, которые надо иметь на случай неурожая.

Мы уже отмечали, что рождение детей является статистически устойчивым испытанием. Во всех странах и среди всех народов всегда на 1000 родившихся в среднем приходится 515 мальчиков и 485 девочек. Это поразительное постоянство рождений мальчиков и девочек отмечалось многими учеными, среди которых был и один из основателей теории вероятностей – французский математик Симон Лаплас (1749–1827). Поэтому в тех случаях, когда нас интересует число мальчиков среди очень большого числа новорожденных, мы можем уверенно предсказать это число с большой степенью точности: ведь эти числа 0,515 и 0,485 принимаются в качестве вероятностей рождения мальчика и девочки соответственно. Просматривая[1] в свое время списки рождений по городу Парижу за 1745–1784 годы, С. Лаплас обнаружил, что отношение числа мальчиков к общему числу рождений здесь оказалось равным 0,510, т. е. чуть меньше, чем 0,515. Несмотря на то что разница была очень мала, Лаплас заключил, что должна быть какая-то специальная причина, увеличивающая число девочек: ведь число рождений в Париже за 39 лет и в те годы было уже довольно большим; поэтому даже такое малое отклонение от обычного отношения нельзя было объяснить только действием случая. И действительно, Лаплас обнаружил причину отклонения: она заключалась в том, что в число детей, рожденных в Париже, включались также и дети, подкинутые в специальный приют – единственный на всю Францию. Так как французские крестьяне ценили в сыновьях будущих работников, то они чаще подкидывали девочек, чем мальчиков. Исключив подкидышей из числа родившихся (многие из них на самом деле родились не в Париже), Лаплас получил обычное отношение числа мальчиков к числу девочек.

Задачи и проблемы, существенно повлиявшие на зарождение и первичное развитие теории вероятностей, возникали при обработке статистических данных и результатов наблюдений в разных науках, из практики страховых компаний, в связи с анализом азартных игр. Первые морские страховые компании возникают в XI столетии в Италии и Нидерландах. В этих компаниях подсчитывали шансы аварий, поскольку при большом риске собирались большие страховые взносы. В связи с бурным развитием естествознания в эпоху Возрождения, в связи с тем, что увеличивалась роль наблюдений и эксперимента, возник вопрос о методах обработки результатов наблюдений, в частности об оценивании случайных ошибок, которые возникают при наблюдениях. Статистические закономерности исторически впервые привлекли внимание ученых, старавшихся выяснить суть азартных игр – таких как игра в кости, «орлянку» и т. п. Эти наблюдения проложили путь к статистическому подходу при числовом определении вероятности.

[1] Детская энциклопедия, т. 3. АПН РСФСР, 1959.

Статистические закономерности в демографических процессах были открыты в XVIII столетии при изучении статистики рождаемости, смертности, несчастных случаев и т. п. В конце XIX – начале XX столетия были выявлены новые статистические закономерности в физике, химии, биологии, экономике и других науках. Все это стимулировало развитие теории вероятностей. Довольно удачное объяснение статистическому понятию вероятности дал Я. Бернулли (1654–1705) в своей книге «Искусство предположений», опубликованной в 1713 г.

Контрольные вопросы

1. Игральный кубик бросали трижды, при этом выпало соответственно 2, 2, 5 очков. Можно ли по этим данным указать приближенное значение вероятности события «при броске игрального кубика выпало два очка»?

2. Проводится последовательное подбрасывание монеты, после каждого из них подсчитывается относительная частота события «выпал герб». Какие из приведенных ниже числовых последовательностей могут соответствовать указанному опыту:

а) $\frac{1}{1}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots$;

б) $\frac{1}{1}, \frac{1}{2}, 1, \frac{3}{4}, \frac{3}{5}, \dots$;

в) $0, 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{5}, \dots$;

г) $\frac{1}{2}, 1, \frac{1}{3}, \frac{3}{4}, \frac{4}{5}, \dots$?

3. Вероятность наступления события A в некотором опыте равна 0,72. Можно ли утверждать, что в 100 таких же опытах, проведенных в тех же условиях, это событие произойдет ровно 72 раза?

4. Для контроля качества продукции некоторого завода из каждой партии готовых изделий отбирают для проверки 150 деталей. Проверку не выдерживают в среднем 6 деталей. Как оценить вероятность выпуска бракованных деталей заводом:

а) в данное время;

б) после усовершенствования технологии производства?

Задачи

79.° Для проверки качества было исследовано 200 деталей, среди которых 5 оказались бракованными. Какой можно считать вероятность того, что наугад взятая деталь будет:

а) пригодной; б) бракованной?

Сколько бракованных деталей окажется в среднем в партии из 1000 деталей?

80.° Вероятность того, что в наугад взятой семье некоторого поселка имеется телевизор, равна 0,998.

а) Сколько в среднем телевизоров будет в 500 семьях этого поселка?

б) Сколько приблизительно опросили семей в этом поселке, если насчитали 1497 телевизоров?

81.° Сколько выстрелов было сделано, если относительная частота попаданий равна 0,7, а количество промахов равно 12?

82.° Относительная частота пар обуви для взрослых, которые продали в магазине за день, равна 0,6. В этот день продали 24 пары детской обуви. Сколько всего пар обуви продали в этот день?

83. Выполните опыт с бросанием игрального кубика 100 раз, записывая все исходы опыта. Исходом будем считать число выпавших очков. Их можно записать в таблицу следующей формы:

№ опыта	1	2	3	4	5	...
Число очков	3	5	2	3	1	...

Вычислите относительную частоту появления каждого числа очков. Подтверждают ли рассчеты предположение о том, что вероятность каждого исхода должна равняться $\frac{1}{6}$?

84. В любой книге на произвольно взятой странице выберите 10 полных строк текста. Вычислите количество гласных и, в частности, буквы «е» в каждой из этих строк. Какой процент от всех букв строки составляет:

а) буква «е»; б) гласная?

Выполните подобный опыт с другой книгой и сравните результаты.

85. Выберите из колоды в 36 игральных карт карты двух мастей — пиковой и червовой. Расположите карты червовой масти в следующем порядке: туз, 6, 7, 8, 9, 10, валет, дама, король. Перетасуйте карты пиковой масти и выложите их рядом с червями.

а) Подсчитайте число совпадений, т. е. сколько карт одинакового достоинства оказалось на одних и тех же местах в обеих мастях.

б) Повторите этот опыт 20 раз, и по результатам этих опытов найдите относительные частоты событий: «совпадений не было», «имело место одно совпадение», «имело место два совпадения», «число совпадений равно 3», «число совпадений больше 3».

в) Могло ли число совпадений равняться 9?

г) Какое число совпадений вы считаете наиболее вероятным в одном опыте?

д) Подсчитайте среднее арифметическое числа совпадений для 20 опытов.

86.* На листе начерчены параллельные прямые, расстояния между которыми равны длине иглы (рис. 17). Подбросьте иглу 100 раз и подсчитайте количество ее пересечений с любой из прямых. Подсчеты выполните для 10, 20, 30, ..., 100 подбрасываний. Результаты занесите в таблицу следующей формы:

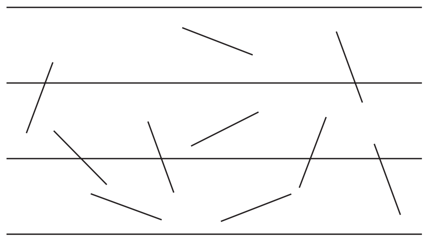


Рис. 17

Число подбрасываний	**Частота пересечения иглой линии**	**Относительная частота пересечения иглой линии**
10		
20		
...		
100		

Оцените вероятность события «игла пересечет линию».

87. Зная относительную частоту события, охарактеризуйте соответствующее явление:

а) относительная частота солнечных дней в июне равна 0,4;

б) относительная частота дождливых дней в сентябре равна 0,6;

в) относительная частота бракованных парашютов, производимых некоторым предприятием, равна 0,01;

г) относительная частота обнаружения некоторого заболевания с помощью рентгена равна 0,8.

88. Тщательно перемешайте колоду из 36 карт. Вытяните наугад одну карту. Возвратите ее в колоду. Снова тщательно перемешайте колоду и вытяните наугад карту. Повторите этот опыт 100 раз. Результаты опытов поместите в таблицу следующей формы:

№ вытягивания	1	2	...	100
Масть	Пиковая	Бубновая	...	Бубновая
Название карты	Туз	Девятка	...	Дама

Вычислите:

а) относительную частоту события «вытянули даму»;

б) относительную частоту события «вытянули карту бубновой масти»;

в) относительную частоту события «вытянули бубновую даму».

Сравните свои результаты с результатами ваших одноклассников. Объедините результаты опытов, проведенных всеми одноклассниками. Выполните по объединенным результатам задачи а—в. Какие выводы вы можете сделать?

89. Начальник цеха, выпускающего принтеры, будет оценивать в конце завтрашнего дня число произведенных за месяц устройств и число дефектных среди них.

а) Опишите соответствующий случайный эксперимент.

б) Каким будет его выборочное пространство?

в) О чем могут свидетельствовать результаты эксперимента?

г) Точно определите событие «достигнута цель выпуска не менее 50 исправных (не имеющих дефектов) устройств при не более одном дефектном устройстве» в терминах результатов, составляющих выборочное пространство.

д) В течение 22 дней из последних 25 дней эта цель достигалась. Найдите соответствующую относительную частоту.

§ 2.2. Классическая вероятность

Наличие вероятности у некоторого события до сих пор мы связывали с существованием статистически устойчивых опытов. Проведение массовых экспериментов в некоторых случаях или невозможно, или чрезвычайно трудоемко. В таких случаях иногда целесообразно использовать для характеристики шансов наступления события так называемую *классическую вероятность*. Фактически на ней базируется доэкспериментальный расчет вероятностей.

2.2.1. Равновозможность

Этот подход к подсчету вероятностей базируется на понятии *равновозможности*, или *равновероятности*, *исходов* некоторого опыта. Интуитивно эти термины понятны.

Пример 1. При подбрасывании монеты разумно предположить, что монета «одинаково» может упасть вверх гербом или цифрой; эти два исхода опыта — герб и цифра — рассматриваются как равновозможные. Иначе говоря, мы считаем, что выпадение герба имеет такие же шансы осуществиться, как и выпадение цифры. ■

Пример 2. Большинство людей считают достаточно разумным представление о том, что при бросании игрального кубика (рис. 18) выпадение одной грани столь же возможно, как и выпадение другой. Поэтому опыт с бросанием игрального кубика имеет шесть равновозможных исходов: на верхней грани выпадает одно очко, два очка, три очка и т. д. Говорят, что для осуществления каждого из этих исходов имеется один шанс из шести. ■

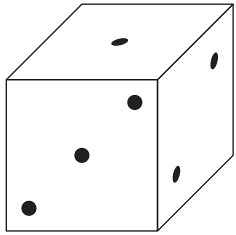


Рис. 18

Пример 3. В колоде из 36 игральных карт все карты таковы, что невозможно отличить одну из них от другой, если смотреть на них с тыльной стороны. Предположим, что из тщательно перетасованной колоды вытаскивают наугад одну карту. Говорят, что каждая карта имеет такие же шансы быть вытянутой, как и любая другая, — один шанс из тридцати шести. ■

Пример 4. Предположим, что 15 имеющих одинаковую квалификацию человек подали заявление на некоторую вакансию. Если в этой группе выбор при приеме на работу осуществляют случайным образом, то каждый из претендентов имеет одинаковый шанс быть принятым на работу — один из пятнадцати. ■

Пример 5. Представим себе, что на складе имеется 74 коробки передач, из них одна имеет дефект. Наудачу выбирают одну коробку передач для потребителя. Дефектная коробка имеет такой же шанс попасть к потребителю, как и любая другая — один из семидесяти четырех. ■

Равновозможность исходов означает по существу сохранение «равноправия» одного исхода по отношению к другому. Она является проявлением симметрии исходов опыта. Симметрия в природе — это и геометрическая симметрия, и различного рода однородность, например однородность материала. На вопрос о том, какие исходы считать равновозможными, математика ответа не дает. Если монета изготовлена из однородного материала, если она не сточена, если она непредвзято брошена, то мы вправе считать исходы «выпадение герба» и «выпадение цифры» равновозможными. Точно так же при бросании «правильного» игрального кубика все шесть исходов данного опыта естественно считать равновозможными. Если из ящика с шарами наугад, наудачу, случайным образом извлекают шар, причем после каждого извлечения шар возвращают в ящик, содержимое ящика тщательно перемешивают, а потом извлекают следующий шар, то исходы этого эксперимента мы вправе считать равновозможными.

Заметим, что на практике процедура перемешивания, особенно при большом числе шаров, не всегда является эффективной. Тщательно перемешать не так просто, как это кажется. И об этом хорошо знает каждый, кто, например, занимался перемешиванием теста, составлением кулинарных смесей и т. д. А вот для «неправильного» кубика (например, на одну грань прикрепили кусочек пластилина) исходы опыта не равновозможные. Точно так же нельзя считать равновозможными исходы подбрасывания кнопки, или пуговицы несимметричной формы, или «неправильной» монеты. Если в ящике шары различаются по весу, то при перемешивании, по-видимому, более тяжелые шары окажутся внизу ящика, и исходы опыта с извлечением шара вряд ли можно будет считать равновозможными.

? Можно ли считать равновозможными следующие исходы:
а) «на купленный билет выпал выигрыш» и «на купленный билет не выпал выигрыш»;
б) «из ящика с пятью одинаковыми шарами, пронумерованными числами 1, 2, 3, 4, 5, извлечены наугад шары № 1 и № 2»;
в) «при выборе случайным образом представителя от коллектива фабрики в 100 человек выбранным окажется работник цеха, насчитывающего 35 человек, и человек, не работающий в этом цехе»;
г) «посетитель в казино выиграл» и «посетитель в казино не выиграл»?

2.2.2. Вероятность события

Рассмотрим несколько примеров, связанных с подбрасыванием монет, на подсчет шансов наступления тех или иных исходов.

Пример 6. Подброшена монета. Каковы шансы того, что она упадет гербом вверх?

❐ Давайте подумаем, сколько может быть вариантов того, что монета упадет вверх гербом или цифрой, если считать невозможным, что монета встала на ребро. У монеты две стороны. И если она симметрична, то шансы того, что она упадет на любую из сторон одинаковы. Итак, шансы равны $\frac{1}{2}$, т. е. составляют 50% от числа всех возможных результатов подбрасывания монеты. ■

Пример 7. Подброшены две монеты. Каковы шансы того, что обе монеты упадут гербом вверх?

❒ Здесь уже сложней перебрать все варианты. Попробуем сделать это, воспользовавшись таблицей 2.7.

Таблица 2.7

№ варианта	Первая монета	Вторая монета
1	**Герб**	**Герб**
2	Герб	Цифра
3	Цифра	Герб
4	Цифра	Цифра

Всего существует 4 варианта. И только при одном из них (он в таблице отмечен полужирным шрифтом) обе монеты упадут гербом вверх. Если монета не деформирована, то у нас нет оснований считать, что какой-то из этих вариантов предпочтительней остальных. Другими словами, они имеют равные шансы. Мы будем говорить, что *вероятность* того, что обе монеты упадут гербом вверх, можно считать равной $\frac{1}{4}$, или 25%. А вероятность того, что монета в примере 6 упадет гербом вверх, — $\frac{1}{2}$, или 50%. ■

Продолжим рассматривать примеры на подсчет шансов, постепенно усложняя рассмотренную выше ситуацию.

Пример 8. Подброшены три монеты. Какова вероятность того, что все три монеты упадут гербом вверх?

❒ Снова попробуем перебрать все возможные варианты. Будем, для краткости, выпадение монеты гербом вверх обозначать буквой Г, а цифрой вверх — буквой Ц (табл. 2.8).

Таблица 2.8

№ варианта	1	2	3	4	5	6	7	8
Первая монета	Г	Г	Г	Г	Ц	Ц	Ц	Ц
Вторая монета	Г	Г	Ц	Ц	Г	Г	Ц	Ц
Третья монета	Г	Ц	Г	Ц	Г	Ц	Г	Ц

Всего 8 вариантов. Все они имеют равные шансы. И только при одном из них (он выделен полужирным шрифтом) все три монеты упадут гербом вверх. Итак, вероятность того, что все три монеты упадут гербом вверх, равна $\frac{1}{8}$. ■

Продолжим усложнять задачу.

Пример 9. Подброшены четыре монеты. Какова вероятность того, что все четыре монеты окажутся гербом вверх?

❐

Таблица 2.9

№ варианта	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Первая монета	Г	Г	Г	Г	Г	Г	Г	Г	Ц	Ц	Ц	Ц	Ц	Ц	Ц	Ц
Вторая монета	Г	Г	Г	Г	Ц	Ц	Ц	Ц	Г	Г	Г	Г	Ц	Ц	Ц	Ц
Третья монета	Г	Г	Ц	Ц	Г	Г	Ц	Ц	Г	Г	Ц	Ц	Г	Г	Ц	Ц
Четвертая монета	Г	Ц	Г	Ц	Г	Ц	Г	Ц	Г	Ц	Г	Ц	Г	Ц	Г	Ц

Аналогично предыдущему примеру получим, что вероятность того, что все четыре монеты упадут гербом вверх, равна $\frac{1}{16}$. ■

Теперь вам нетрудно будет подсчитать, какова вероятность того, что из пяти и шести подброшенных монет все окажутся упавшими гербом вверх.

? Вычислите эти вероятности.

Примеры, которые мы рассмотрели, подсказывают нам определенный метод количественного определения шансов наступления некоторого исхода опыта. Рассмотрим, например, подбрасывание монеты. Мы ранее на повседневном языке говорили, что осуществление выпадения герба имеет «один шанс из двух»; на математическом языке говорят, что «вероятность выпадения герба равна $\frac{1}{2}$». Символически это записывают так:

$$P(\text{герб}) = \frac{1}{2}.$$

Точно так же при бросании игрального кубика грань, на которой изображено 5 точек, имеет один шанс из шести оказаться наверху; поэтому мы скажем, что «вероятность выпадения пяти очков равна $\frac{1}{6}$». Символически это записывают так:

$$P(\text{пять очков}) = \frac{1}{6}.$$

Аналогично при вытягивании из колоды в 36 карт одной карты есть только один шанс из 36 вытянуть пикового туза. Говорят, что «вероятность появления пикового туза равна $\frac{1}{36}$» и записывают:

$$P(\text{пиковый туз}) = \frac{1}{36}.$$

Рассмотрим еще несколько примеров.

Пример 10. Из колоды в 36 карт наугад вынимают одну карту. Какова вероятность того, что будет извлечен: а) пиковый туз; б) туз?

❐ На первый вопрос ответ мы получили выше: $P(\text{пиковый туз}) = \frac{1}{36}$. А извлечь туз безразлично какой масти можно с бóльшими шансами, ведь в колоде 4 туза, т. е. при вытягивании карты извлеченным может оказаться туз любой из четырех мастей, шансы увеличиваются в 4 раза. Поэтому $P(\text{туз}) = \frac{4}{36} = \frac{1}{9}$. ■

Пример 11. В классе внеочередной дежурный определяется жребием среди 10 человек, еще не дежуривших в этом месяце. На 10 карточках пишутся имена претендентов, после чего все карточки тщательно перемешиваются в коробке и одну карточку вынимают наугад. Тот, чье имя написано на этой карточке, назначается дежурным. Если имя «Сергей» написано на одной карточке, то шансы Сергею быть дежурным равны 1 из 10, так как все исходы жеребьевки можно считать равновозможными. Поэтому

$$P(\text{дежурить будет Сергей}) = \frac{1}{10}.$$

Представьте себе, что кто-то имя «Сергей» вписал на трех карточках. Тогда его шансы дежурить были бы равны 3 из 10, и

$$P(\text{дежурить будет Сергей}) = \frac{3}{10}.$$

Обратите внимание на то, что в знаменателе, как и в предыдущем примере, стоит общее число исходов рассматриваемого опыта (вытягивание одной карты из 36 или одной карточки из 10), а в числителе — число тех исходов, при которых происходит интересующее нас событие (извлечен туз, дежурить будет Сергей). Так как карточки, по условию, вытягивались наугад, после тщательного перемешивания, то можно считать исходы опыта равновозможными.

Итак, для нахождения вероятности рассматриваемого события (дежурить будет Сергей) нужно из всех равновозможных исходов выделить так называемые «благоприятствующие исходы». Затем вероятность рассматриваемого события вычисляется по следующему правилу:

$$P(\text{события}) = \frac{\text{число благоприятствующих исходов}}{\text{число всех равновозможных исходов}}. \blacksquare$$

Пример 12. Игральный кубик бросают один раз. Чему равна вероятность того, что на верхней грани кубика окажется: а) четное число очков; б) число очков, большее 2?

❒ а) Выпадения всех шести граней равновозможны. Четное число очков на трех гранях: 2, 4, 6. Поэтому вероятность выпадения четного числа очков равна

$$P(\text{четное число очков}) =$$
$$= \frac{\text{число благоприятствующих исходов}}{\text{число всех равновозможных исходов}} = \frac{3}{6} = \frac{1}{2}.$$

б) Число очков, большее 2, имеется на четырех гранях: 3, 4, 5, 6. Поэтому

$$P(\text{число очков больше 2}) = \frac{4}{6} = \frac{2}{3}. \blacksquare$$

Пример 13. Подбрасывают две монеты. Какова вероятность того, что: а) обе они упадут вверх гербом; б) одна упадет вверх гербом, а другая — вверх цифрой?

❒ Всего возможны четыре исхода: ГГ, ГЦ, ЦГ, ЦЦ. Все они равновозможны. Обе монеты упадут гербом вверх при одном исходе ГГ. Поэтому $P(\text{ГГ}) = \frac{1}{4}$. Этот результат мы ранее получили в примере 7.

Одна монета упадет гербом вверх, а другая — вверх цифрой при двух исходах ГЦ или ЦГ. $P(\text{ГЦ или ЦГ}) = \frac{2}{4} = \frac{1}{2}$. ■

Пример 14. В классе учатся 15 мальчиков и 10 девочек. По жребию выбирают одного ученика для выполнения поручения классного руководителя. Какова вероятность того, что выбранным окажется мальчик.

❒ Всего 25 исходов, так как выбранным может оказаться любой учащийся в классе. Все исходы равновозможны. Мальчик окажется выбранным при любом из 15 исходов. Поэтому $P(\text{мальчик}) = \frac{15}{25} = \frac{3}{5}$. ■

Пример 15. Пусть в примере 4 о приеме на работу из 15 человек, подавших заявление, 6 претендентов — женщины. Какова вероятность того, что предпочтение будет отдано женщине?

❒ Всего 15 исходов, так как на работу могут принять любого из 15 человек, подавших заявление. Все исходы равновозможны (выбор осуществляется случайным образом). Женщина окажется выбранной при любом из 6 исходов. Поэтому $P(\text{женщина}) = \frac{6}{15} = \frac{2}{5}$. ■

Вероятностью события *называют отношение числа исходов опыта, благоприятствующих событию, к общему числу равновозможных исходов опыта.*

Если эксперимент может привести к любому из N различных равновозможных исходов и если в точности $N(A)$ из этих исходов соответствуют событию A, то вероятность события A равна

$$P(A) = \frac{N(A)}{N}.$$

?

Петя купил один билет лотереи, в которой разыгрывается 10 призов и продано 120 билетов.

а) Какова вероятность того, что он выиграет приз?

б) Какова вероятность того, что он не выиграет приз?

В корзине лежат 28 яблок красного и зеленого цветов. Известно, что вероятность вытянуть яблоко зеленого цвета, не глядя в корзину, равняется $\frac{3}{7}$. Сколько красных яблок в корзине?

В корзине лежат яблоки красного и зеленого цветов, красного — 8 штук. Известно, что вероятность вытянуть яблоко зеленого цвета, не глядя в корзину, равняется $\frac{2}{5}$. Сколько всего яблок в корзине?

Андрей и Олег договорились, что если при бросании двух игральных кубиков в сумме выпадет число очков, кратное 5, то выигрывает Андрей, а если в сумме выпадет число очков, кратное 6, то выигрывает Олег. У кого из мальчиков больше шансов выиграть?

На вопрос, какие исходы опыта считать равновозможными, как мы отмечали выше, математика ответа не дает. Проверить предположение о равновозможности исходов опыта можно экспериментально, проведя довольно большое число опытов.

Как проверить, является ли игральный кубик правильным?

Как проверить, равновозможны ли все исходы при вращении рулетки?

Говорят, что известный французский математик Д'Аламбер (1717—1783) при решении задачи о вероятности выпадения двух гербов при подбрасывании двух монет рассуждал так: при подбрасывании двух монет есть три возможных исхода опыта: 1) «выпало два герба»; 2) «выпали один герб и одна цифра»; 3) «выпали две цифры». Он думал, что ни один из исходов не имеет по сравнению с другими больше шансов наступить, и поэтому каждый из этих исходов наступает с вероятностью $\frac{1}{3}$. Если бы это было верно, то каждый из этих исходов встречался бы примерно в $\frac{1}{3}$ всех проведенных опытов с подбрасыванием двух монет. Для проверки опыт с подбрасыванием двух монет был проведен 300 раз. Оказалось, что дважды герб выпал 86 раз, один герб и одна цифра – 141 раз, две цифры – 73 раза. Эти результаты свидетельствуют в пользу решения, приведенного при решении примера 7, согласно которому мы вправе были ожидать, что выпадение одного герба и одной цифры встретится приблизительно в два раза чаще, чем выпадение двух гербов или двух цифр.

Известны и другие примеры, когда неравновозможные исходы принимались за равновозможные, вследствие чего получали неправильные результаты[1].

Один богатый джентльмен – предположительно Великий герцог Тосканский – пожаловался Галилею (1564—1642), что игральные кости не всегда ведут себя, как бы это

[1] См. книгу: *С. Даймена*. Мир вероятностей. — М.: Статистика, 1970.

следовало согласно логике. Он знал по опыту (а это, по-видимому, был достаточно богатый опыт), что при игре тремя кубиками легче получить 10, чем 9, т. е. при бросании трех игральных кубиков чаще выпадает 10 очков по сравнению с 9 очками. Это его удивляло, так как, по его мнению, имеется ровно столько же возможностей получить сумму 9, как и сумму 10. Он составил из трех игральных кубиков шесть комбинаций, каждая из которых давала в сумме по 9 очков, и шесть других комбинаций, составляющих каждая по 10 очков. Эти комбинации представлены в таблице 2.10.

Таблица 2.10

Комбинации, которые дают 9 очков	**Комбинации, которые дают 10 очков**
6 + 2 + 1	6 + 3 + 1
5 + 3 + 1	6 + 2 + 2
5 + 2 + 2	5 + 4 + 1
4 + 4 + 1	5 + 3 + 2
4 + 3 + 2	4 + 4 + 2
3 + 3 + 3	4 + 3 + 3

В чем же заключается, спросил Великий герцог Галилея, нарушение правильности хода действий. Галилей указал ему на ошибку. Давайте вместе разберемся в этой ошибке.

Сначала рассмотрим опыт с бросанием двух игральных кубиков. При этом можно получить любую сумму очков от 2 до 12. Можно было бы думать, что в задаче имеется 11 возможных случаев и вероятность появления каждого из них равна $\frac{1}{11}$. Но это не так. Опыт показывает, что, например, сумма 7 появляется много чаще, чем сумма 12. Это и понятно, так как 12 можно получить только в виде: $6 + 6 = 12$, а 7 – многими способами: $1 + 6 = 2 + 5 = 3 + 4 = 4 + 3 = 5 + 2 = 6 + 1 = 7$. Первым записано число очков на первом кубике, а вторым – на втором. Так что 11 названных случаев нельзя считать равновозможными. Для подсчета вероятностей появления любой суммы очков приходится исходами опыта считать пары чисел, первое из которых характеризуется определенным числом очков, выпавших на первом кубике, второе – определенным числом очков, выпавших на втором кубике. Все исходы представлены в таблице 2.11.

Естественно, эти 36 случаев считать равновозможными. Опыт показывает, что в случае правильных кубиков, сделанных из однородного материала, не налитых, например свинцом, и надлежащих приемов бросания (например, после встряхивания в стаканчике) эти $6 \cdot 6 = 36$ случаев появляются при большом числе бросков примерно одинаково часто. Для сумм очков на двух кубиках получаем следующие вероятности (табл. 2.12). Обратите внимание на то, что исходы, благоприятствующие каждой сумме очков, расположены вдоль диагоналей в таблице: $2 = 1 + 1$; $3 = 1 + 2 = 2 + 1$; $4 = 1 + 3 = 2 + 2 = 3 + 1$; $9 = 3 + 6 = 4 + 5 = 5 + 4 = 6 + 3$ и т. д.

Таблица 2.11

Число очков на 1-м кубике	**Число очков на 2-м кубике**					
	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Таблица 2.12

Сумма	2	3	4	5	6	7	8	9	10	11	12
Число благоприятствующих случаев	1	2	3	4	5	6	5	4	3	2	1
Вероятность	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Обратите внимание на то, что в таблицах 2.11 и 2.12 представлены все исходы опыта, состоящего в бросании двух игральных кубиков. Из них можно выделить исходы, благоприятствующие, например, следующим событиям:

A – «сумма очков на кубиках меньше 5»;

B – «сумма очков на кубиках кратна 3».

Однако исходы, при которых наступают, например, события:

C – «на первом кубике выпало вдвое больше очков, чем на втором»;

D – «на кубиках выпало одинаковое число очков»

можно выделить из исходов, представленных в таблице 2.11, и нельзя выделить из исходов, представленных в таблице 2.12: ведь по сумме очков нельзя определить, какое число очков выпало на каждом кубике. Говорят, что исходы в таблице 2.11 «мельче» по сравнению с исходами в таблице 2.12. Чем «мельче» описаны исходы опыта, тем больше событий можно выразить через них, т. е. выделить те исходы, которые благоприятствуют им.

Более «мелкие» исходы чаще, по сравнению с «крупными», оказываются равновозможными.

В задаче с бросанием трех игральных кубиков исходами опыта будем считать тройки чисел, первое из которых характеризуется числом очков, выпавших на первом кубике, второе – числом очков, выпавших на втором кубике, третье – числом очков, выпавших на третьем кубике. Ясно, что всего имеется $6 \cdot 6 \cdot 6 = 216$ равновозможных случаев. Галилей заметил, что любая комбинация из трех различных чисел, например

6 + 2 + 1, встречается в 6 исходах: 6 + 2 + 1, 6 + 1 + 2, 2 + 6 + 1, 2 + 1 + 6, 1 + 6 + 2, 1 + 2 + 6. Комбинация, в которую входит только два различных числа, например 4 + 4 + 1, встречается при трех исходах: 4 + 4 + 1, 4 + 1 + 4, 1 + 4 + 4. А комбинация, в которой все числа одинаковые, например 3 + 3 + 3, встречается только в одном исходе. Другими словами, комбинации, представленные в таблице 2.10, не являются равновозможными. Используя вышеприведенные рассуждения, дополним эту таблицу (табл. 2.13).

Таблица 2.13

Комбинации, которые дают 9 очков	**Число случаев, в которых встречается комбинация**	**Комбинации, которые дают 10 очков**	**Число случаев, в которых встречается комбинация**
6 + 2 + 1	6	6 + 3 + 1	6
5 + 3 + 1	6	6 + 2 + 2	3
5 + 2 + 2	3	5 + 4 + 1	6
4 + 4 + 1	3	5 + 3 + 2	6
4 + 3 + 2	6	4 + 4 + 2	3
3 + 3 + 3	1	4 + 3 + 3	3
Всего	25	Всего	27

Следовательно, вероятность выпадения суммы 9 равна $\frac{25}{216}$, а вероятность выпадения суммы 10 равна $\frac{27}{216}$. Теперь понятно, почему при игре тремя кубиками число 9 у Великого герцога появлялось реже, чем 10.

Иногда, чтобы получить удовлетворительные результаты для опытов, исходы которых, как мы знаем заранее, не являются равновозможными, можно допустить их равновозможность. Например, несмотря на то, что известно, что вероятности рождения мальчика и девочки соответственно равны 0,515 и 0,485, часто допускают, что новорожденный имеет одинаковые шансы оказаться мальчиком или девочкой. И для многих практических задач такое допущение вполне приемлемо.

Вплотную к классическому определению вероятности через отношение количеств равновозможных событий подошел Д. Кардано (1501–1576) в XVI столетии. Он сформулировал правило для подсчета размера ставок, которое было близким к этому определению. Фактически классическим определением пользовались в своих работах Г. Галилей (1564–1642), Б. Паскаль (1623–1662), П. Ферма (1601–1665), Х. Гюйгенс (1629–1695), Я. Бернулли (1654–1705), Т. Бейес (1702–1761) и другие ученые. Классическое определение вероятности в современном виде предложил П. Лаплас (1749–1827) в начале XIX столетия. Его используют в том случае, если существует возможность предсказания вероятности на основе симметрии условий, при которых происходит опыт.

Контрольные вопросы

1. Можно ли считать равновозможными следующие исходы:

а) «промах» и «попадание» у отличного стрелка;

б) «выпал герб» и «выпала цифра» при подбрасывании деформированной монеты;

в) «выпало 1—6 очков» при бросании правильного игрального кубика;

г) «выпало 1—6 очков» при бросании деформированного игрального кубика?

2. Исходы какого случайного эксперимента можно считать равновозможными:

а) подбрасывание правильного тетраэдра — номер грани, на которую упал тетраэдр;

б) стрельба первоклассного стрелка по мишени — попадание или непопадание в мишень;

в) проверка стандартности детали — деталь стандартна или деталь бракована;

г) подбрасывание кнопки — кнопка упала острием вверх или вниз?

3. При проведении эксперимента могут наступить 10 равновозможных исходов, взаимно исключающих друг друга. Чему равна вероятность события, которое происходит:

а) только при одном исходе;

б) при любом из двух определенных исходов?

4. В результате эксперимента происходят равновозможные исходы, взаимно исключающие друг друга. Вероятность любого из них равна 0,05. Найдите число этих исходов.

5. При 10 подбрасываниях правильной монеты каждый раз выпадал герб. Что вероятнее при следующем подбрасывании: выпадет герб или цифра?

Задачи

90.° В лотерее из одиннадцати билетов выигрывают три. Какова вероятность того, что наугад взятый билет:

а) выиграет; б) не выиграет?

91.° Из полного набора костей домино (28 штук) наугад выбирают одну. Найдите вероятность того, что:

а) выбранная кость является шестеркой;

б) сумма очков на кости равна 5.

92. Числа от 1 до 15 написаны на 15 карточках по одному на каждой. Выбирают наугад одну карточку. Чему равна вероятность того, что написанное на этой карточке число:

а)° делится на 5;

б)° является четным;

в)° является нечетным;

г) является точным квадратом;

д) двузначное;

е) простое?

93.° В ящике в три раза больше красных шаров, чем черных (шары одинаковы во всем, кроме цвета). Наугад вынимают один шар. Какова вероятность того, что он:

а) красный; б) черный?

94. Числа от 1 до 20 написаны на листах бумаги, которые поместили в коробку и перемешали. Из коробки наугад взяли один лист. Какова вероятность того, что число на взятом листе будет или простым, или кратным трем?

95. В вазе 20 хризантем, из них четыре белых, пять синих, восемь красных, остальные желтые. Из вазы наудачу 30 раз вынимают цветок, возвращая его после каждого извлечения. При этом в шести случаях цветок оказался белым, в семи — синим, в 12 — красным и в пяти — желтым.

а)° Найдите вероятность и относительную частоту события «извлеченный цветок — желтый».

б) Сравните вероятность и относительную частоту события «извлеченный цветок — желтый».

96.° Два игрока одновременно показывают друг другу либо один, либо два, либо три пальца на правой руке. Если для каждого игрока равновозможно показать один, два или три пальца, то чему равна вероятность того, что общее число показанных пальцев:

а) четно; б) нечетно; в) больше четырех;

г) меньше двух; д) простое; е) составное?

97. Из тщательно перетасованной колоды в 36 карт наугад выбирают одну карту. Какова вероятность того, что она окажется:

а) пиковой масти; б) тузом; в) красной масти;

г) картинкой (т. е. или валетом, или дамой, или королем, или тузом)?

98. Из набора домино, состоящего из 28 костей, наугад выбирают одну. Какова вероятность того, что эта кость:

а) будет содержать 6 очков;

б) окажется дублем;

в) не окажется дублем?

99. Из букв слова «математика» наугад выбирают одну букву. Какова вероятность того, что это окажется:

а) буква «м»; б) буква «а»;

в) гласная буква; г) согласная буква?

100. Бросают два игральных кубика. Какова вероятность того, что сумма выпавших очков окажется равной:

а) 2; б) 3; в) 4; г) 7; д) 12?

101. Грани кубика окрашены в синий и зеленый цвета. Вероятность выпадения синей грани равна $\frac{1}{3}$, вероятность выпадения зеленой грани равна $\frac{2}{3}$. Сколько синих и сколько зеленых граней у кубика?

102. Поставим 120 фишек на кружок **А** (рис. 19). Будем подбрасывать монету и перемещать фишки одну за другой с **А** на **Б**, если выпал герб, и с **А** на **В**, если выпала цифра. С фишками на **Б** поступим так: будем подбрасывать игральный кубик и перемещать фишки одна за другой — если выпала 1, то на **Д**, если не выпала 1, то на **Г** — и так до тех пор, пока кружок **Б** не опустеет. Тогда перейдем к фишкам на **В**: при выпадении 1 будем отправлять их на **Д**, при выпадении остальных цифр — на **Е**, и так до тех пор, пока кружок **В** тоже не опустеет. Сколько примерно фишек окажется на кружках **Г**, **Д** и **Е**?

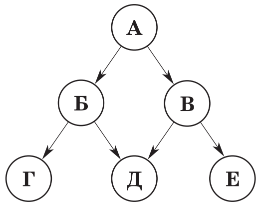


Рис. 19

103. Найдите ошибку в следующих рассуждениях. Чтобы определить вероятность того, что наугад выбранный гражданин Украины живет в определенном регионе (одна автономная республика, 24 области и два города государственного подчинения), число случаев, благоприятствующих наступлению рассматриваемого события (1), делят на общее число регионов (27); при этом получают $\frac{1}{27}$.

104. Одновременно бросают два игральных кубика, развертки которых изображены на рис. 20. Принимаются ставки на суммы очков, которые выпадут на обоих кубиках. Найдите вероятность того, что сумма будет равняться 0, 1, 2. На какое число очков вы поставите?

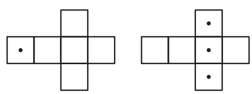


Рис. 20

105. Чему равна вероятность того, что при бросании двух игральных кубиков:

а) число очков на одном кубике будет в 2 раза больше числа очков на другом кубике;

б) выпадет разное число очков на обоих кубиках;

в) число очков на одном кубике будет на 2 меньше числа очков на другом;

г) число очков на одном кубике будет по крайней мере на 2 больше, чем число очков на другом кубике?

106. Одновременно подбрасывают две монеты. Составьте для этого опыта:

а) систему равновозможных исходов;

б) систему неравновозможных исходов;

в) событие, которое описывается исходами системы а), но не описывается исходами системы б).

107. В кошельке 25 монет, из них 15 никелевых, остальные — медные. Из него наудачу 50 раз вынимают монету, возвращая ее после каждого извлечения. При этом в 19 случаях монета оказалась медной.

а) Найдите вероятность того, что при некотором извлечении вынута медная монета.

б) Найдите относительную частоту события «извлечена медная монета» и сравните ее с вероятностью этого события.

108. Из ящика, в котором три белых и два черных шара, случайно извлекают два шара. Миша и Маша договорились, что если шары будут одного цвета, то победителем будет Маша, если разного — Миша. У кого из детей больше шансов стать победителем?

109. Из колоды в 36 карт наугад выбирают одну карту. Какова вероятность того, что это не будет картинка (шестерка, семерка, восьмерка, девятка, десятка)?

§ 2.3. Субъективная вероятность

В предыдущих параграфах о вероятности некоторого события говорилось в двух случаях: при существовании большого числа статистически устойчивых опытов и при конечном числе равновозможных исходов эксперимента.

Иногда вероятность рассматривают как некоторую меру личного доверия к какому-либо утверждению, например к утверждению о том, что сегодня будет снег. Такие вероятности часто называют

субъективными. Сторонники такого подхода к оценке вероятности считают, что различные лица могут проявлять различную степень доверия к тому или иному утверждению, исходя из одних и тех же фактов, поэтому субъективные вероятности какого-либо события могут оказаться различными. Они применяют теорию вероятностей не только к анализу статистически устойчивых экспериментов и экспериментов с конечным числом равновозможных исходов, но и многих других. Например, они могут говорить о вероятности свершения террористического акта в том или ином регионе, в той или иной стране.

Субъективную оценку вероятности получают на основе суждения определенного лица о вероятности некоторого события. Такой подход может показаться не научным, однако часто оказывается, что это есть лучшее, что можно сделать в отсутствие предыдущего опыта (т. е. не имея возможности использовать относительную частоту) и в отсутствие равновозможности исходов эксперимента (т. е. без возможности вычислить теоретическое значение вероятности). Один из путей улучшения качества подхода на основе субъективной оценки вероятности состоит в использовании мнения эксперта в данной области. Например, можно воспользоваться мнением специалиста по банковским инвестициям для оценки вероятности того, что слияние конкурирующих фирм окажется успешным, или мнением инженера о технической осуществимости нового технологического подхода в области энергетики, или мнением спортивного специалиста о шансах той или иной команды в соревновании и т. д.

Когда человек говорит, что с вероятностью 0,9 он выйдет завтра утром на прогулку, то фактически он учитывает ряд условий, к которым он пришел на основании размышлений. Например, этими условиями могут быть следующие: к нему никто не должен прийти, погода позволит выйти на прогулку, дела не помешают этой прогулке, состояние здоровья не будет препятствовать прогулке и т. д. При этом не исключается возможность того, что одно из перечисленных условий не даст осуществить прогулку. Ясно, что приведенная вероятность носит субъективный характер: у другого человека по поводу вероятности того же события могут быть другие суждения.

Когда человек ставит себе задачу оценить вероятность события A, он учитывает как природу события A, так и все, что он знает относительно различных возможностей, которые могут благоприятствовать или не благоприятствовать осуществлению события A.

Многие считают, что количественные оценки для вероятностей тех или иных поступков отдельных лиц (физических или юридических) едва ли можно считать интересными в научном плане. Эти вероятности имеют отношение лишь к данному лицу и в большой степени зависят от его психологических и физиологических особенностей и даже от его состояния в данный момент. Выводы, которые при этом будут получены, несправедливы не только для других людей, но и для того же лица в другое время. В то же время выводы, сделанные на основании субъективных оценок вероятностей, могут представлять определенный интерес.

Пример. Против некоторой фирмы возбуждено судебное дело. Предъявлен иск на 50 000 денежных единиц. Нужно выбрать оптимальную стратегию для защиты интересов фирмы. Для этого требуется оценить, насколько правдоподобными являются различные возможные исходы дела. Юридическая служба фирмы предлагает уладить дело без судебного разбирательства, путем переговоров. Возможны следующие представляющие интерес события:

1) урегулирование вопроса с затратами менее 1000 денежных единиц;

2) урегулирование вопроса с затратами от 1000 до 10 000 денежных единиц;

3) урегулирование вопроса с затратами более 10 000 денежных единиц.

Как определить вероятности этих событий? Достаточное количество аналогичных случаев, которое позволило бы применить статистический подход, отсутствует. Исходы неравновозможны, поэтому неприменима и классическая модель. Остается единственный путь — провести субъективную оценку вероятности. Изучив все обстоятельства дела, юридическая служба представила следующие значения субъективной вероятности:

1) урегулирование вопроса с затратами менее 1000 денежных единиц — 0,10;

2) урегулирование вопроса с затратами от 1000 до 10 000 денежных единиц — 0,65;

3) урегулирование вопроса с затратами более 10 000 денежных единиц — 0,15.

Обратите внимание на то, что сумма этих вероятностей равна 0,90, при этом остается вероятность в 0,10, что дело будет решаться в суде.

Эти субъективные вероятности представляют наилучшую из доступных оценок правдоподобия различных вариантов развития событий, и из них следует, что, вероятно, вопрос можно будет урегулировать со средними затратами, существенно меньшими 50 000 денежных единиц. Теперь можно воспользоваться этими значениями вероятности для того, чтобы принять непростые решения относительно типа и объема доводов, которые можно привести в защиту интересов фирмы. ■

Итак, можно говорить о трех источниках получения вероятностей для их использования в реальной жизни: найти относительную частоту (с помощью эксперимента), вычислить теоретическое значение вероятности (используя формулы) или воспользоваться субъективной оценкой вероятности (на основе экспертных заключений).

Контрольные вопросы

1. На основе данных за прошлый год было установлено, что 40% посетителей некоторого магазина не бывали в нем ранее. В то время как некоторые пришли просто посмотреть, 30% посетителей что-либо купили. Однако среди тех, кто раньше в магазине не был, покупку совершили только 20%. Какие типы вероятностей здесь указаны — с точки зрения того, каков их источник?

2. Руководитель считает, что график работ на строительстве объекта можно выполнить при условии, что вовремя удастся принять на работу нового прораба, однако, несмотря на это, ситуация останется рискованной. По его мнению, вероятность своевременно нанять нового прораба равна 70%. Если прораб будет принят на работу вовремя, то вероятность успеха составит 80%, в противном случае вероятность успеха составит только 40%. Какие типы вероятностей здесь указаны — с точки зрения того, каков их источник?

Задачи

110. Укажите, какой вы считаете вероятность победы вашей школьной футбольной команды в первенстве района среди школьных команд? На основании чего вы пришли к такому выводу?

111. Укажите, какими вы считаете вероятности поступления в три выбранные вами вуза? На основании чего вы пришли к такому выводу? Как вы используете эту информацию в процессе поступления в вуз?

112. Предполагается открыть новое кафе. Есть некоторые шансы получить место для него в одном из двух районов. Для каждого района различны шансы успеха проекта. Как можно оценить те и другие шансы? Попытайтесь составить соответствующий бизнес-план проекта.

§ 2.4. Вероятностная модель случайного опыта

В предыдущих трех параграфах были рассмотрены три способа нахождения вероятностей случайных событий: статистический, классический и субъективный. Каждый из них имеет свою сферу применения, свои достоинства и недостатки.

Статистическое определение вероятности имеет широкую сферу применения, оно тесно связано с практикой. Однако требуются большие усилия по проведению экспериментов, оно имеет много «слабых мест». Что значит «достаточно большое количество опытов»? Насколько может отклониться частота от вероятности при данном числе опытов? Поэтому статистическое определение не является строгим определением вероятности.

Классическое определение вероятности более строго с математической точки зрения, оно понятно для применения. Однако имеет очень ограниченную сферу использования: оно применимо только к анализу экспериментов с конечным числом равновозможных исходов. Такие случайные опыты на практике встречаются не часто.

Понятие субъективной вероятности вообще не поддается математической формализации. Но оно ценно тем, что иллюстрирует, как в тех случаях, когда неприменимы ни статистическая, ни классическая вероятности для случайных событий, могут быть определены вероятности.

Что же объединяет рассмотренные подходы к определению вероятности события? Ответ на этот вопрос и составляет содержание настоящего параграфа. Кратко этот ответ можно сформулировать так: каждый из этих подходов позволяет таким образом приписывать событиям положительные числа, не превышающие 1 (их называют вероятностями), что они обладают свойствами, подобными свойствам длин, площадей, объемов и вообще чисел, полученных при измерении величин. Другими словами, все эти подходы приводят к построению математической модели случайного опыта, которую называют ***вероятностной моделью***.

2.4.1. Пространство элементарных исходов опыта

В дальнейшем мы ограничимся рассмотрением опытов с *конечным числом исходов*.

Чтобы построить вероятностную модель такого случайного опыта, проанализируем еще раз структуру решения задачи, основанного на классическом определении вероятности. Решение начиналось с описания всех возможных исходов опыта, взаимно исключающих друг друга. Из тех или иных соображений предполагалось, что они равновозможны. Подсчитывалось число всех исходов опыта. Описывались все исходы опыта, при которых наступает интересующее нас событие, и подсчитывалось их количество. И в конце концов вычислялась вероятность события. Описанную модель и соответствующую вероятность называют *классической*.

Первым шагом при построении классической вероятностной модели было описание совокупности его простейших исходов, которую называют ***пространством элементарных исходов опыта***.

Пример 1. Описать совокупность простейших исходов опыта, состоящего в подбрасывании двух различных монет.

❒ Дать такое описание однозначно нельзя. Ответ зависит от того, что нас интересует в эксперименте.

Если нас интересует, что выпало при каждом подбрасывании: герб или цифра, то множество

$$U_1 = \{\text{ГГ, ГЦ, ЦГ, ЦЦ}\},$$

где, например, ГЦ означает, что при подбрасывании первой монеты выпал герб, при подбрасывании второй — цифра, представляет собой список всех возможных исходов нашего опыта.

Если нас интересует число выпавших гербов, то список всех возможных исходов опыта представляет собой множество

$$U_2 = \{0, 1, 2\},$$

где каждый элемент соответствует числу выпавших гербов.

Можно считать простейшими исходами опыта и такие его результаты, как «монеты упали одинаково» (обе гербом или обе цифрой вверх) и «монеты упали по-разному», т. е. совокупность всех возможных исходов опыта есть множество

$$U_3 = \{\text{О, Р}\},$$

где О означает «одинаково», Р — «различно».

Исходы, которые содержатся в U_1 и U_3, равновозможны, исходы, входящие в U_2, неравновозможны, в чем можно убедиться с помощью проведения экспериментов. Имеем две классические модели эксперимента, вторую модель нельзя отнести к классическим, в ней не равны вероятности простейших исходов. Первая из построенных моделей позволяет вычислять вероятности большего числа событий. Так, вероятности событий «монеты упали одинаково», «монеты упали не одинаково» можно вычислить и в модели U_1, и в модели U_3, а вероятности событий «герб выпал по крайней мере один раз» или «цифра выпала два раза» можно вычислить только в модели U_1.

Обратите внимание на то, что каждому простейшему исходу во множестве U_2 соответствует подмножество множества U_1: 0 → {ЦЦ}, 1 → {ГЦ, ЦГ}, 2 → {ГГ}. Аналогично устанавливается соответствие между элементами U_3 и подмножествами U_1: О → {ГГ, ЦЦ}, Р → {ГЦ, ЦГ}.

Следовательно, с помощью множества U_1 можно построить и U_2, и U_3. В дальнейшем будем отдавать предпочтение описанию «самого большого» множества исходов опыта, если не будет дополнительных требований. Такой выбор в приложениях часто диктуется потребностями.

Таким образом, в данном примере дано описание трех совокупностей элементарных исходов данного случайного опыта. И в этом нет противоречий. То, что считается простейшим элементарным исходом с одной точки зрения, с другой — может состоять из более простых элементов. ■

?

Можно ли считать множеством элементарных исходов опыта, состоящего в бросании игрального кубика, множество:

а) $U = \{1, 2, 3, 4, 5, 6\}$;

б) $U = \{Ч, Н\}$, где Ч означает, что выпало четное число очков, а Н — нечетное;

в) $U = \{П, С\}$, где П означает, что число выпавших очков является простым, а С — составным;

г) $U = \{< 4, > 2\}$, где < 4 означает, что выпало менее 4 очков, а > 2 — более 2?

Рассмотренный пример позволяет осознать, что причиной ошибок, допущенных и Д'аламбером, и герцогом Тосканским, является попытка применить классическое определение вероятности к модели, которая не является классической. Эту ошибку допускают многие начинающие изучать вероятность.

Пример 2. Описать совокупность элементарных исходов опыта, состоящего в двукратном бросании игрального кубика.

❒ Как и в примере 1, такое описание выполняется неоднозначно. Исходами опыта могут служить пары цифр (i, j), $i, j = 1, 2, \ldots, 6$, где первая цифра — результат первого броска, вторая — второго. Соответствующее множество элементарных исходов представим в виде таблицы 2.14.

Таблица 2.14

Результат первого броска	**Результат второго броска**					
	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Если в качестве элементарных исходов опыта выбрать суммы выпавших очков, то множество элементарных исходов примет вид

$$U = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}. \blacksquare$$

? Могут ли совпадать множества элементарных исходов следующих опытов:
а) бросание игрального кубика;
б) бросание дартса (маленькой стрелы) в мишень, изображенную на рис. 21?

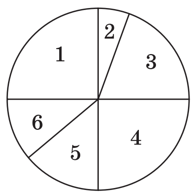


Рис. 21

Рассмотренные примеры позволяют дать описание *пространства элементарных исходов опыта* с конечным числом исходов.

Пусть имеется случайный опыт с конечным числом исходов $u_1, \ldots, u_N$, таких, что при любом осуществлении опыта происходит один и только один из них. Множество $U = \{u_1, \ldots, u_N\}$ называют ***пространством элементарных исходов*** *(для краткости будем его обозначать ПЭИ), а его элементы $u_1, \ldots, u_N$ —* ***элементарными исходами****.*

Основными свойствами ПЭИ являются:

1) полнота: ПЭИ должно содержать каждый исход, которым может закончиться опыт;

2) взаимная исключаемость элементарных исходов: никакой исход опыта не должен дважды фигурировать среди элементов ПЭИ.

? Может ли ПЭИ случайного опыта состоять из одного элемента?

Составление ПЭИ опыта требует умения кодировать информацию с помощью символов.

Пример 3. Производится опрос, связанный с планами улучшения жилищных условий работников большого предприятия. Каждому из опрашиваемых задают два вопроса:

— Удовлетворены ли вы качеством жилья?

— Удовлетворены ли вы удаленностью квартиры от места работы?

Описать ПЭИ опыта, заключающегося в опросе одного человека.

❒ Будем обозначать утвердительный ответ на вопрос цифрой 1, а отрицательный — 0. Тогда запись 10 означает, что опрашиваемый на первый вопрос ответил утвердительно, а на второй — отрицательно.

ПЭИ опыта, заключающегося в опросе одного человека, имеет вид

$$U = \{11, 10, 01, 00\}. \blacksquare$$

Пример 4. Пусть из коробки, содержащей 2 простых (П), 2 синих (С) и 2 зеленых (З) карандаша, наугад одновременно вынимают два карандаша. Составить ПЭИ опыта, считая, что исходами опыта является:

а) состав извлеченных карандашей без учета порядка извлечения;

б) состав извлеченных карандашей с учетом порядка извлечения;

в) состав, описываемый количеством простых и цветных (Ц) карандашей, с учетом порядка извлечения;

г) состав, описываемый количеством простых и цветных карандашей, без учета порядка извлечения.

❒ Соответствующие ПЭИ имеют вид:

U_1 = {2П, 2С, 2З, 1П 1С, 1П 1З, 1С 1З};
U_2 = {ПП, ПС, ПЗ, СП, СС, СЗ, ЗП, ЗС, ЗЗ};
U_3 = {ПП, ПЦ, ЦП, ЦЦ};
U_4 = {ПП, ПЦ, ЦЦ}.

Исходы ни одного из этих ПЭИ не являются равновозможными. Если мы хотим иметь равновозможные исходы, то можно пронумеровать, например, числами 1 и 2 простые карандаши, 3 и 4 — синие, 5 и 6 — зеленые. Пары номеров с учетом порядка извлечения описывают все исходы рассмотренного опыта. Они равновозможны. Нет оснований один из них предпочесть другому. Соответствующее ПЭИ имеет вид

$$U_5 = \{12, 13, 14, 15, 16, 21, 23, 24, 25, 26, 31, 32, 34, 35, 36,\\ 41, 42, 43, 45, 46, 51, 52, 53, 54, 56, 61, 62, 63, 64, 65\}.$$

Если же исходами опыта считать пары номеров извлеченных карандашей без учета порядка их извлечения, то ПЭИ будет иметь вид

$$U_6 = \{12, 13, 14, 15, 16, 23, 24, 25, 26,\\ 34, 35, 36, 45, 46, 56\}.$$ ■

2.4.2. Вероятности элементарных исходов

Для полного математического описания случайного опыта недостаточно построить его ПЭИ. Например, ПЭИ опыта, состоящего в двукратном подбрасывании монеты, может иметь вид $U = \{0, 1, 2\}$, где каждый элемент соответствует числу выпавших гербов. Такой же вид имеет ПЭИ опыта, состоящего в извлечении двух шаров из мешка, содержащего три белых и два красных шара. ПЭИ этого опыта имеет тот же вид $U = \{0, 1, 2\}$, где 0 — оба шара белые; 1 — один шар белый, другой красный; 2 — оба шара красные. Но это различные опыты: в первом случае вероятности исходов соответственно равны $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ (эти вероятности получены при решении примера 13 § 2.2); во втором случае — $\frac{3}{10}$, $\frac{6}{10}$, $\frac{1}{10}$ (их можно получить с помощью классического определения вероятности, перебрав возможные исходы опыта). Вероятности элементарных исходов оказались различными. Поэтому для полного описания опыта необходимо, кроме перечисления всех возможных его элементарных исходов, еще и указать, как часто может наступить тот или иной элементарный исход.

Второй шаг в построении вероятностной модели случайного опыта состоит в приписывании элементарным исходам их вероятностей.

В опыте с бросанием игрального кубика, ввиду его симметричности, у каждой грани равные шансы оказаться сверху, т. е. можно считать вероятность p_i выпадения i очков равной $\frac{1}{6}$, $i = 1, 2, 3, 4, 5, 6$. При этом

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1.$$

В ПЭИ примера 1 в связи с равновозможностью исходов пространства $U_1 = \{ГГ, ГЦ, ЦГ, ЦЦ\}$ также естественно положить вероятность наступления любого исхода равной $\frac{1}{4}$. Точно так же в первом ПЭИ примера 2 благодаря равновозможности исходов естественно положить вероятность выпадения любой пары очков равной $\frac{1}{36}$. Однако не всегда есть основания считать исходы данного опыта равновозможными. Например, ПЭИ опыта, состоящего в одном выстреле по мишени, содержит два, вообще говоря, неравновозможных исхода. В этом случае вероятности исходов можно принять равными относительным частотам соответствующих исходов при большом числе повторений опыта.

Описанным способом можно ввести вероятности исходов в примере 3. Пусть в этом опыте 50% опрошенных положительно ответили на оба вопроса, 25% — положительно на первый и отрицательно на второй, 15% — отрицательно на первый и положительно на второй, 10% — отрицательно на оба вопроса. Вероятности элементарных исходов естественно положить равными

$$p_1 = P(11) = 0{,}5;$$
$$p_2 = P(10) = 0{,}25;$$
$$p_3 = P(01) = 0{,}15;$$
$$p_4 = P(00) = 0{,}1;$$
$$p_1 + p_2 + p_3 + p_4 = 1.$$

В опыте, заключающемся в бросании стрелы в мишень, изображенную на рис. 21, элементарные вероятности можно ввести, положив их равными дроби, выражающей отношение площади того или иного сектора к площади круга, а именно:

$$p_1 = 0{,}25; \quad p_2 = 0{,}05; \quad p_3 = 0{,}2;$$
$$p_4 = 0{,}25; \quad p_5 = 0{,}15; \quad p_6 = 0{,}1;$$
$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1.$$

Здесь предполагается, что при каждом бросании стрела попадает в мишень.

Теперь можно завершить описание построения вероятностной модели случайного опыта.

Пусть случайному опыту поставлено в соответствие пространство элементарных исходов (ПЭИ) $U = \{u_1, u_2, \ldots, u_N\}$, *а каждому элементарному исходу* u_i *— некоторое число* p_i, *которое удовлетворяет условиям:*

1) $0 < p_i < 1$ *для всех* $i = 1, 2, \ldots, N$;

2) $p_1 + p_2 + \ldots + p_N = 1$.

Числа p_i *называют* ***элементарными вероятностями****, или* ***вероятностями элементарных исходов*** u_i, *а множество* U *вместе с вероятностями* p_i *—* ***вероятностной моделью случайного опыта****.*

Почему элементарная вероятность не может быть равна: а) 0; б) 1?

Классическая модель получается в том случае, если все элементарные вероятности p_i равны между собой и, в силу условия 2, равны $\frac{1}{N}$.

Статистическая вероятностная модель строится на основе исследования относительных частот элементарных исходов случайного опыта.

Вернемся к рассмотрению примера 4. Повторим описанный там опыт 100 раз, каждый раз возвращая карандаши в коробку и тщательно перемешивая их, результаты представим в таблицах 2.15—2.20, соответствующих ПЭИ U_1—U_6.

Таблица 2.15

Исходы U_1	2П	2С	2З	1П 1С	1П 1З	1С 1З
Частота	6	7	8	28	26	25
Относительная частота	0,06	0,07	0,08	0,28	0,26	0,25

Таблица 2.16

Исходы U_2	ПП	ПС	ПЗ	СП	СС
Частота	6	16	12	12	7
Относительная частота	0,06	0,16	0,12	0,12	0,07
Исходы U_2	СЗ	ЗП	ЗС	ЗЗ	
Частота	12	14	13	8	
Относительная частота	0,12	0,14	0,13	0,08	

Таблица 2.17

Исходы U_3	ПП	ПЦ	ЦП	ЦЦ
Частота	6	28	26	40
Относительная частота	0,06	0,28	0,26	0,4

Таблица 2.18

Исходы U_4	ПП	ПЦ	ЦЦ
Частота	6	54	40
Относительная частота	0,06	0,54	0,4

Таблица 2.19

Исходы U_5	12	13	14	15	16	21	23	24	25	26
Частота	4	3	4	3	2	2	5	4	4	3
Относительная частота	0,04	0,03	0,04	0,03	0,02	0,02	0,05	0,04	0,04	0,03
Исходы U_5	31	32	34	35	36	41	42	43	45	46
Частота	1	4	3	2	2	3	4	4	5	3
Относительная частота	0,01	0,04	0,03	0,02	0,02	0,03	0,04	0,04	0,05	0,03
Исходы U_5	51	52	53	54	56	61	62	63	64	65
Частота	3	4	5	3	4	4	3	3	2	4
Относительная частота	0,03	0,04	0,05	0,03	0,04	0,04	0,03	0,03	0,02	0,04

Таблица 2.20

Исходы U_6	12	13	14	15	16
Частота	6	4	7	6	6
Относительная частота	0,06	0,04	0,07	0,06	0,06
Исходы U_6	23	24	25	26	34
Частота	9	8	8	6	7
Относительная частота	0,09	0,08	0,08	0,06	0,07
Исходы U_6	35	36	45	46	56
Частота	7	5	8	5	8
Относительная частота	0,07	0,05	0,08	0,05	0,08

Как мы раньше и предполагали, пары номеров с учетом порядка извлечения, указанные в таблице 2.19, представляют собой равновозможные исходы рассмотренного опыта. Любой из 30 описанных в таблице исходов появлялся приблизительно одинаковое число раз: от 1 до 5.

Вероятности исходов рассматриваемого опыта для каждого пространства элементарных исходов естественно положить равными относительным частотам, представленным в соответствующих таблицах. Во всех рассмотренных случаях сумма вероятностей всех исходов равна 1. ◄

Контрольные вопросы

1. Приведите пример опыта с неравновозможными исходами.

2. В ящике 6 карточек, на которых написаны числа 1, 2, 3, 4, 5, 6. Наудачу извлекают одну карточку. Образуют ли ПЭИ исходы опыта, состоящие в извлечении карточки с числом:

а) четным, нечетным;

б) большим 3, меньшим 3;

в) большим 3, не большим 3;

г) простым, составным;

д) кратным 3, четным, нечетным?

3. Приведите пример опыта с тремя исходами, не образующими ПЭИ.

4. Бросают игральный кубик. ПЭИ этого опыта построено двумя способами: $U_1 = \{1, 2, 3, 4, 5, 6\}$ (исход совпадает с числом выпавших очков) и $U_2 = \{Ч, Н\}$ (Ч — выпало четное число очков, Н — нечетное число). Какое из этих ПЭИ «богаче» событиями?

Задачи

113.° Числа 1, 2, 3, 4 записывают на четырех листках бумаги. После этого листки кладут в коробку и перемешивают. Человек с завязанными глазами вынимает один за другим два листка. Опишите ПЭИ опыта, введите элементарные вероятности.

114. Два шара — красный и синий — помещают наугад в два ящика, пронумерованных числами 1 и 2. Постройте ПЭИ этого опыта, если:

а) оба шара можно положить в один ящик;

б) ни один ящик не должен быть пустым.

115. ПЭИ опыта состоит из 10 исходов, причем девять из них имеют равные вероятности, а последний имеет вероятность, равную сумме вероятностей пяти других. Найдите элементарные вероятности данного опыта.

116.° В ящике пять пронумерованных шаров: три белых и два черных. Наудачу вынимают три шара. Постройте ПЭИ этого опыта.

117. Производится два выстрела по мишени. Постройте ПЭИ этого опыта и введите элементарные вероятности, если известно, что при многократном его повторении в 50% случаев имелось два, а в 40% — одно попадание.

118. Монета искривлена и поэтому вероятность выпадения цифры вдвое меньше вероятности выпадения герба. Чему равны эти вероятности?

119. Наугад выбирают одну букву из слова «закон». Какие из следующих совокупностей являются ПЭИ рассмотренного опыта:

а) {з, а, к, о, н}; б) {гласная, з, к, н}; в) {а, к, о, н};

г) {гласная, согласная}; д) {согласная, о}?

120.° Подбрасывают монету, а после этого бросают игральный кубик. Постройте различными способами ПЭИ этого опыта.

121. Подбрасывают две монеты, а после этого бросают игральный кубик. Составьте для этого опыта ПЭИ с:

а) 24 исходами; б) 18 исходами; в) 12 исходами;

г) 8 исходами; д) 6 исходами; е) 4 исходами.

В каких из этих случаев исходы опыта будут равновозможными, а в каких — нет?

122. В ящике красные, белые и синие шары. Шары одинаковы во всем, кроме цвета. Один за другим вынимают два шара. Постройте различные ПЭИ этого опыта.

123. Некая фирма выпускает шоколадки. В каждую шоколадку вкладывается одна из двух фотографий известных спортсменов. Фотографий каждого спортсмена одинаковое количество. Тому, кто соберет полный комплект фотографий, следующая шоколадка выдается бесплатно. Постройте ПЭИ опыта, состоящего в том, что некто соберет полный комплект фотографий или купит не более трех шоколадок.

124. Постройте ПЭИ следующих опытов и введите элементарные вероятности:

а) Иглу бросают на отрезок длиной 20 см, разделенный на четыре части, имеющие длины 8, 6, 4 и 2 см.

б) Скаковая лошадь «Фаворит» может в забеге занять первое, второе, третье или четвертое место.

в) Из лекарств a, b, c, применяемых для лечения хронической болезни, отбирают случайным образом два для проведения сравнительного исследования.

г) Телевизионная программа предложила телезрителям ответить на вопрос, поддерживается ли некоторая акция правительства, при этом предложено три варианта ответа: «да», «нет», «не знаю».

д) На ткацкой фабрике из каждой сотни остановок ткацкого станка, требующих последующей работы ткачихи, в среднем 22 остановки происходит из-за обрыва нитей основы, 31 — из-за обрыва нитей утка, 27 — из-за смены человека, 3 — из-за поломки погонялок, а остальные остановки происходят по другим причинам.

125. ПЭИ опыта состоит из пяти элементарных исходов. Элементарные вероятности p_1, p_2, p_3, p_4, p_5 образуют арифметическую прогрессию с разностью 0,05. Найдите эти элементарные вероятности.

§ 2.5. Случайные события и их вероятности

Построение вероятностной модели случайного опыта отвлекло нас от главной цели — уточнить смысл выражения «вероятность случайного события». Но именно ради этой цели мы занимались построением математической модели случайного опыта. В настоящем параграфе дается определение понятия вероятности случайного события, связанного со случайным опытом.

2.5.1. Случайное событие

Вы уже имеете представление о случайном событии. Теперь уточним это понятие. Предварительно рассмотрим некоторые примеры.

Пример 1. Для опыта, заключающегося в подбрасывании двух различных монет, описать событие A — «герб выпал один раз».

❒ Описание этого события зависит от выбора пространства элементарных исходов опыта. Если ПЭИ есть множество $U_2 = \{1, 2, 3\}$, то событие A является элементом этого множества, т. е. элементарным исходом, обозначенным цифрой 1: $A = \{1\}$.

Если ПЭИ данного опыта есть множество $U_1 = \{ГГ, ГЦ, ЦГ, ЦЦ\}$, то событие A наступает при каждом из исходов ГЦ, ЦГ. Справедливо и обратное утверждение: если наступает событие A, то наступает один из исходов ГЦ, ЦГ. Естественно это событие отождествить с множеством {ГЦ, ЦГ}, которое является подмножеством ПЭИ. ■

Точно так же подмножеству {ГГ, ЦЦ} соответствует событие B — «монеты упали одинаково». Событие B наступает при каждом элементарном исходе, входящем в это подмножество и только при этих исходах. Подмножеству {ГЦ, ЦГ} соответствует событие «монеты упали по-разному», подмножеству {ГГ, ГЦ, ЦГ} — «герб выпал по крайней мере один раз».

? Можно ли описать словами событие {ГЦ, ЦГ} способом, отличным от приведенного в тексте?

Пример 2. Для опыта, заключающегося в бросании игрального кубика, описать:

а) событие A, состоящее в том, что выпало нечетное число очков;

б) событие B, состоящее в том, что выпало более 4 очков.

❒ ПЭИ данного опыта является множество $\{1, 2, 3, 4, 5, 6\}$.

а) Событие A наступает при реализации одного из исходов множества $\{1, 3, 5\}$. И обратно, если наступает событие A, то наступает один из исходов 1, 3, 5. В связи с этим событие A отождествляется с множеством $\{1, 3, 5\}$.

б) Событие B наступает при реализации одного из исходов множества $\{5, 6\}$. И обратно, если наступает событие B, то наступает один из исходов 5, 6. В связи с этим событие B отождествляется с множеством $\{5, 6\}$. ■

Пример 3. Для опыта, заключающегося в бросании игрального кубика дважды, описать:

а) событие A, состоящее в том, что сумма выпавших очков равна 4;

б) событие B, состоящее в том, что при обоих бросаниях выпало одинаковое число очков.

❒ ПЭИ данного опыта является множество, представленное таблицей 2.14 в § 2.4.

а) Событие A — «сумма выпавших очков равна 4», есть множество $\{(1, 3), (2, 2), (3, 1)\}$. Оно наступает при любом исходе этого множества, и если оно наступает, то наступает один из исходов этого множества. Событие A отождествляется с множеством $\{(1, 3), (2, 2), (3, 1)\}$.

б) Событие B — «выпало одинаковое число очков», есть множество $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. Оно наступает при любом исходе этого множества, и если оно наступает, то наступает один из исходов этого множества. Событие B отождествляется с множеством $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$. ■

Приведенные примеры показывают, что случайное событие можно описывать с помощью подмножеств ПЭИ.

Любое множество исходов ПЭИ можно понимать как некоторое событие. Так, множество $\{2, 4, 6\}$ в опыте примера 2 составляет событие D — «выпало четное число очков», множество $\{1, 2, 3\}$ означает событие E — «выпало менее 4 очков».

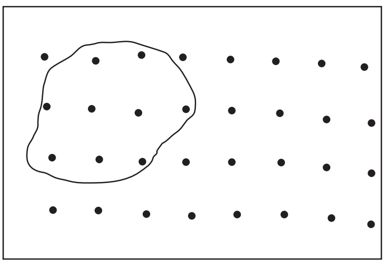
***Случайным событием** (или короче, **событием**) называют произвольное подмножество (часть) пространства элементарных исходов опыта.*

В частности, каждый элементарный исход ПЭИ является событием, всё ПЭИ является событием.

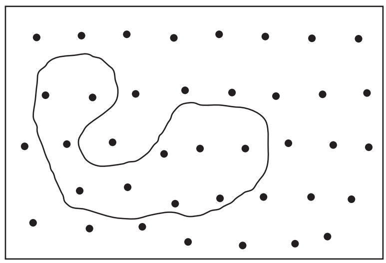
Графически пространство элементарных исходов изображают в виде конечного множества точек, находящихся внутри некоторой фигуры (прямоугольника или круга). Тогда событие изображается в виде подмножества этого множества точек[1].

На рис. 22, *а* ПЭИ опыта состоит из 32 элементарных исходов, а событие, связанное с соответствующим опытом, — из 9 исходов. Аналогично, на рис. 22, *б* ПЭИ опыта состоит из 39 элементарных исходов, а событие, связанное с соответствующим опытом, — из 10 исходов.

Если $U = \{u_1, u_2, ..., u_N\}$ — ПЭИ опыта, то случайное событие A, связанное с этим опытом, записывают перечислением элементарных исходов, входящих в A: $\{u_{i_1}, u_{i_2}, ..., u_{i_k}\}$.



а)



б)

Рис. 22

[1] Подобные рисунки называют диаграммами Венна в честь английского математика и логика Д. Венна (1834—1923). Они имеют вид прямоугольника, точки внутри которого изображают элементы ПЭИ. Внутри прямоугольника находятся изображения событий в виде кругов, овалов, прямоугольников.

2.5.2. Вероятность случайного события

Теперь мы вплотную подошли к поиску ответа на вопрос: как определить вероятность произвольного события, связанного с опытом, т. е. вероятность любого подмножества пространства элементарных исходов?

Вернемся к рассмотрению примера 1. Если в качестве ПЭИ взять множество U = {ГГ, ГЦ, ЦГ, ЦЦ} и все элементарные вероятности p_i положить равными $\frac{1}{4}$, то событие A — «герб выпал хотя бы один раз», состоящее из трех исходов, A = {ГГ, ГЦ, ЦГ}, при проведении большого числа опытов будет очевидно происходить в три раза чаще, чем каждый из исходов. Поэтому его вероятность естественно считать в три раза большей по сравнению с вероятностью одного исхода, она, очевидно, равна $\frac{3}{4}$:

$$P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

Вероятность события «герб выпал два раза» совпадает с вероятностью исхода ГГ и равна $\frac{1}{4}$.

Пример 4. Производится опрос, связанный с планами улучшения жилищных условий работников большого предприятия. Каждому из опрашиваемых задают два вопроса:

— Удовлетворены ли вы качеством жилья?

— Удовлетворены ли вы удаленностью квартиры от места работы?

ПЭИ опыта, заключающегося в опросе одного человека, имеет вид $U = \{11, 10, 01, 00\}$. Вероятности элементарных исходов положили равными $p_1 = P(11) = 0{,}5$; $p_2 = P(10) = 0{,}25$; $p_3 = P(01) = 0{,}15$; $p_4 = P(00) = 0{,}1$. Найти вероятность:

а) события A — «опрашиваемый удовлетворен хотя бы одним из исследуемых параметров: качеством жилья или неудаленностью от места работы»;

б) события B — «опрашиваемый удовлетворен одним и только одним из исследуемых параметров».

❒ а) Событие A состоит из трех исходов, $A = \{11, 10, 01\}$. Так как проведенный опрос показал, что в среднем $50\% + 25\% + 15\% = 90\%$ опрошенных положительно ответили по крайней мере на один из поставленных вопросов, то вероятность события A естественно считать равной сумме соответствующих элементарных вероятностей, т. е. $P(A) = p_1 + p_2 + p_3 = 0{,}5 + 0{,}25 + 0{,}15 = 0{,}9$.

б) Событие B состоит из двух исходов $B = \{10,\ 01\}$. В среднем 25% + 15% = 40% опрошенных положительно ответили на один и только один из поставленных вопросов. Поэтому вероятность события C естественно считать равной сумме соответствующих элементарных вероятностей, т. е. $P(B) = p_2 + p_3 = 0{,}25 + 0{,}15 = 0{,}4$. ■

Рассмотренные примеры показывают, что вероятностью случайного события естественно считать сумму вероятностей исходов, образующих это событие.

Пусть имеется вероятностная модель случайного опыта, т. е. пространство элементарных исходов (ПЭИ) $U = \{u_1, u_2, ..., u_N\}$ и элементарные вероятности $p_1, p_2, ..., p_N$. Пусть A — некоторое событие, т. е. подмножество множества U: $A = \{u_{i_1}, u_{i_2}, ..., u_{i_k}\}$.

***Вероятностью события** A называют сумму вероятностей исходов, составляющих это событие.*

Вероятность события обозначают $P(A)$. По определению

$$P(A) = p_{i_1} + p_{i_2} + ... + p_{i_k}.$$

Сумму, стоящую в правой части этого равенства, для краткости обозначают так: $\sum\limits_{i:\ u_i \in A} p_i = \sum\limits_{m=1}^{k} p_{i_m}$. Сумма, стоящая в левой части этого равенства, читается так: сумма вероятностей тех элементарных исходов, которые образуют событие A. Тогда последнее равенство принимает вид

$$P(A) = \sum_{i:\ u_i \in A} p_i = \sum_{m=1}^{k} p_{i_m}.$$

?

Чему равна вероятность события U?

Чему равна вероятность события $A = \{u_1,\ u_2\}$?

Может ли вероятность события быть больше 1?

Приведем ряд примеров на вычисление вероятности события.

Пример 5. Найти вероятности следующих событий, связанных с опытом, заключающимся в двукратном бросании игрального кубика и представленным в примере 3:

A — «число очков, выпавших на верхней грани первого кубика, превышает число очков на верхней грани второго кубика»,

B — «сумма выпавших очков меньше 5».

❐ Эти события можно выразить через исходы ПЭИ, рассмотренного в примере 3:

$$A = \{(2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (3, 2), (4, 2), (5, 2), (6, 2), (4, 3), (5, 3), (6, 3), (5, 4), (6, 4), (6, 5)\};$$

$$B = \{(1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1)\}.$$

Так как вероятность каждого исхода равна $\frac{1}{36}$, и эти события состоят соответственно из 15 и 6 исходов, то их вероятности равны:

$$P(A) = \frac{1}{36} \cdot 15 = \frac{5}{12}; \quad P(B) = \frac{1}{36} \cdot 6 = \frac{1}{6}.$$

Событие B можно выразить и через исходы другого ПЭИ $U = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, где в качестве элементарных исходов опыта выбраны суммы выпавших очков. Однако для него пока не введены элементарные вероятности. Воспользовавшись введенным определением, теперь можно и для этого ПЭИ ввести элементарные вероятности. Обратите внимание на то, что в таблице 2.14 исходы, для которых суммы выпавших очков одинаковы, расположены на диагоналях этой таблицы. Например, событие «сумма выпавших очков равна 5» состоит из исходов (1, 4), (2, 3), (3, 2), (4, 1). Его вероятность равна $\frac{4}{36} = \frac{1}{9}$. Элементарные исходы второго ПЭИ рассматриваемого опыта представлены в таблице 2.21.

Таблица 2.21

Сумма (исход 2-го ПЭИ)	2	3	4	5	6	7	8	9	10	11	12
Число исходов 1-го ПЭИ	1	2	3	4	5	6	5	4	3	2	1
Вероятность	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

■

Пример 6. Вычислить вероятности событий, связанных с опытом, рассмотренным в примере 4 § 2.4:

A — «вынуты карандаши разного цвета»;

B — «вынуты только цветные карандаши»;

C — «вынуты простой и цветной карандаши, причем цветной карандаш появился раньше простого»;

D — «первым вынут синий карандаш».

❐ Исходы, из которых состоит событие A, можно охарактеризовать ПЭИ, представленными в таблицах 2.15 и 2.16, событие B — во всех таблицах 2.15—2.20, событие C — в таблицах 2.16 и 2.17, события D — в таблице 2.16. Ни в одной из этих таблиц не представлены равновозможные исходы. Равновозможные исходы представлены в таблице 2.19.

Фактически в этом примере построено 6 различных ПЭИ одного и того же опыта. Вычислим вероятности этих событий, приняв, как мы раньше указывали, за элементарные вероятности соответствующие относительные частоты. По данным таблицы 2.15 (ПЭИ U_1) имеем:

$$P(A) = P(1\text{П } 1\text{С}, 1\text{П } 1\text{З}, 1\text{С } 1\text{З}) = 0{,}28 + 0,\ 26 + 0{,}25 = 0{,}79;$$

$$P(B) = P(1\text{С } 1\text{З}, 2\text{С}, 2\text{З}) = 0{,}25 + 0{,}07 + 0{,}08 = 0{,}4.$$

По данным таблицы 2.16 (ПЭИ U_2) имеем:

$$P(C) = P(\text{СП}, \text{ЗП}) = 0{,}12 + 0{,}14 = 0{,}26;$$

$$P(D) = P(\text{СП}, \text{СС}, \text{СЗ}) = 0{,}12 + 0{,}07 + 0{,}12 = 0{,}31.\ \blacksquare$$

2.5.3. Классическая вероятность и ее связь со статистической вероятностью

Рассмотрим некоторые следствия из приведенного определения вероятности события.

Пусть дан опыт с N *равновозможными* исходами: $U = \{u_1, \dots, u_N\}$. Так как сумма элементарных вероятностей равна 1, то каждая из них равна $\frac{1}{N}$:

$$p_1 = p_2 = \dots = p_N = \frac{1}{N}.$$

По определению вероятность события A есть сумма вероятностей элементарных исходов, составляющих это событие:

$$P(A) = \frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N},$$

где число слагаемых равно числу исходов $N(A)$, составляющих событие A. Отсюда

$$P(A) = \frac{N(A)}{N}.$$

Итак, *вероятность события в опыте с N равновозможными исходами равна отношению числа исходов, образующих событие A, к общему числу исходов.*

Последнее утверждение мы ранее назвали *классическим определением вероятности.* Теперь мы его получили как частный случай общего определения.

Пример 7. В урне находится пять шаров, среди которых три белых и два черных. Из нее наудачу извлекают один. Какова вероятность того, что извлеченный шар окажется черным?

❒ Пронумеруем шары числами 1, 2, 3, 4, 5, причем первые три номера припишем белым шарам. ПЭИ опыта имеет вид $U = \{1, 2, 3, 4, 5\}$, исходы равновозможны, $N = 5$. Событие A составляют те исходы, которые совпадают с номерами черных шаров: $A = \{4, 5\}$, $N(A) = 2$. Таким образом, $P(A) = \dfrac{2}{5}$. ■

Чему равна вероятность того, что извлеченный шар окажется белым?

Вернемся к примеру 6. Ранее мы фактически подсчитали относительные частоты событий A, B, C, D, они соответственно равны: 0,79; 0,4; 0,26; 0,31. Вычислим вероятности этих событий по данным таблицы 2.19 с помощью классического определения.

❒ Эти события можно выразить через исходы ПЭИ U_5 рассматриваемого опыта следующим образом:

$$A = \{13, 14, 15, 16, 23, 24, 25, 26, 31, 32, 35, 36, 41, 42, 45, 46, 51, 52, 53, 54, 61, 62, 63, 64\};$$

$$B = \{34, 35, 36, 43, 45, 46, 53, 54, 56, 63, 64, 65\};$$

$$C = \{31, 32, 41, 42, 51, 52, 61, 62\};$$

$$D = \{31, 32, 34, 35, 36, 41, 42, 43, 45, 46\}.$$

Так как $U = \{12, 13, 14, 15, 16, 21, 23, 24, 25, 26, 31, 32, 34, 35, 36, 41, 42, 43, 45, 46, 51, 52, 53, 54, 56, 61, 62, 63, 64, 65\}$ и число всех исходов опыта равно 30, то вероятности этих событий соответственно равны:

$$P(A) = \frac{24}{30} = \frac{4}{5} = 0{,}8; \quad P(B) = \frac{12}{30} = \frac{2}{5} = 0{,}4;$$

$$P(C) = \frac{8}{30} = \frac{4}{15} \approx 0{,}27; \quad P(D) = \frac{10}{30} = \frac{1}{3} \approx 0{,}33.$$

Как видим, имеет место неплохое согласование между теоретическими расчетами и результатами опыта; другими словами, относительные частоты событий близки к их вероятностям. Точно такой же результат получим при вычислении вероятностей событий A и B, если воспользуемся моделью, зафиксированной в таблице 2.20: $U = \{12, 13, 14, 15, 16, 23, 24, 25, 26, 34, 35, 36, 45, 46, 56\}$, число всех исходов опыта равно 15; исходы равновозможны;

$$A = \{13, 14, 15, 16, 23, 24, 25, 26, 35, 36, 45, 46\};$$

$$B = \{34, 35, 36, 45, 46, 56\}.$$

$$P(A) = \frac{12}{15} = 0{,}8; \quad P(B) = \frac{6}{16} = 0{,}4. \ \blacksquare$$

В следующем примере используется связь между вероятностью события и его относительной частотой.

Пример 8. Ихтиолог хотел определить, сколько в пруду рыбы, пригодной для вылавливания. Для этого он закинул сеть с заведомо заданными размерами ячеек и, вытянув ее, обнаружил 30 рыб. Пометив каждую из них меткой, он бросил всю рыбу назад в пруд. На следующий день ихтиолог закинул ту же сеть и поймал 40 рыб, на двух из которых были его метки. Каким образом он по этим данным найдет приближенное количество рыбы, пригодной для вылавливания?

❒ Пусть в пруду N рыб, пригодных для вылавливания, тогда вероятность события «наугад взятая рыба помечена» равна $\frac{30}{N}$. Соответственно результатам 40 опытов (опытом считаем вылавливание одной рыбы), проведенных на следующий день, можно подсчитать относительную частоту этого события. Она равна $\frac{2}{40} = \frac{1}{20}$. На основании приближенного равенства относительных частот и вероятностей событий имеем $\frac{30}{N} \approx \frac{1}{20}$. Итак, можно утверждать, что $N \approx 600$. ■

Подобные вероятностные оценки широко используют в физике, биологии, социологии, языкознании, экономике, политике, спорте и повседневной жизни каждого человека. Например, такие методы используют при оценивании урожайности культуры на поле по урожайности той же культуры на некотором количестве небольших участков; плотности минералов по результатам взвешивания некоторого количества образцов; влажности зерна на приемных пунктах; волокнистости хлопка и т. п.

Теперь уже можно подвести итог, ответив на главный вопрос: «Что такое вероятность события?» Для его получения мы математически описали случайный опыт, построили его вероятностную модель. С ее помощью определили вероятность события. Сказанное можно четко описать следующим образом.

Пусть $U = \{u_1, \ldots, u_N\}$ — произвольное конечное множество. Будем называть его ***пространством элементарных исходов (ПЭИ)*** *случайного опыта, а его элементы $u_1, \ldots, u_N$ — исходами этого опыта. Произвольное подмножество (часть) ПЭИ будем называть* ***случайным событием*** *(или короче,* ***событием****). Каждому элементарному исходу u_i из ПЭИ $U = \{u_1, u_2, \ldots, u_N\}$ поставим в соответствие некоторое число p_i, которое удовлетворяет условиям:*

1) $0 < p_i < 1$ *для всех* $i = 1, 2, \ldots, N$;

2) $p_1 + p_2 + \ldots + p_N = 1$.

Числа p_i называют ***элементарными вероятностями****, или* ***вероятностями элементарных исходов*** u_i.

Вероятностью события *A называют сумму вероятностей исходов, составляющих это событие*:

$$P(A) = \sum_{i:\, u_i \in A} p_i.$$

Это определение иногда называют *аксиоматическим определением вероятности*. Напоминаем, что речь идет об определении вероятности в частном случае: для опытов с конечным числом исходов.

Введение ПЭИ и элементарных вероятностей означает построение вероятностной модели случайного опыта. Для одного и того же опыта могут быть построены разные модели. Теория вероятностей не учит тому, как «правильно» определять вероятности p_i элементарных исходов u_i. Она также не занимается поиском ответа на вопрос о том, правильно ли построена вероятностная модель. Соответствие модели реальному опыту проверяют на практике. Теория вероятностей лишь отвечает на вопрос: «Как вычислять вероятности различных событий, связанных с построенной моделью?»

В дальнейшем, говоря о вероятностях событий, мы можем не задумываться над тем, связано ли это событие со статистически устойчивыми опытами, или с опытом с конечным числом равновозможных исходов, или эти вероятности являются оценками экспертов. Наличие вероятности события будет означать, что для опыта с конечным числом элементарных исходов выполняются условия 1 и 2.

Система аксиом теории вероятностей была построена на основе теории множеств в 1933 г. выдающимся российским математиком А. Н. Колмогоровым (1903–1992). Более ранним является иной (и несколько менее удачный) вариант аксиоматики теории вероятностей, предложенный в 1917 г. отечественным математиком С. Н. Бернштейном (1880–1968). Известной книгой А. Н. Колмогорова «Основные понятия теории вероятностей», вышедшей впервые на немецком языке в 1933 г., история аксиоматических подходов к теории вероятностей также не закончилась.

Контрольные вопросы

1. Сумма всех элементарных вероятностей ПЭИ, кроме одной, равна 0,8. Чему равна эта последняя вероятность?

2. Случайное событие A образуют все элементарные исходы ПЭИ, кроме одного. Вероятность этого элементарного исхода равна 0,15. Чему равна вероятность события A?

3. Вероятность события A равна 0,45. Чему равна сумма вероятностей элементарных исходов, не входящих в A?

4. Два события состоят из всех элементарных исходов ПЭИ, причем ни один исход не входит в оба события. Чему равна сумма вероятностей этих событий?

5. Два события состоят из всех элементарных исходов ПЭИ, причем один элементарный исход входит в оба события. Чему равна сумма вероятностей этих событий, если вероятность этого исхода равна 0,05?

6. Может ли сумма вероятностей трех событий, связанных с одним опытом, быть больше 1, если никакой исход не входит в два из этих трех событий?

Задачи

126.° Производится опрос, связанный с планами улучшения обслуживания населения зрелищными мероприятиями. Каждому из опрашиваемых задают два вопроса:

1) Регулярно ли вы посещаете кинотеатры?

2) Регулярно ли вы смотрите телевизор?

Оказалось, что 40% опрошенных регулярно посещают кинотеатры и смотрят телевизор; 20% посещают кинотеатры, но телевизор не смотрят; 30% не ходят в кино, но смотрят телевизор и 10% не посещают кино и не смотрят телевизор. Постройте ПЭИ опыта, заключающегося в опросе одного человека. Введите элементарные вероятности. Найдите вероятность того, что наудачу выбранный человек регулярно: а) посещает кино; б) смотрит телевизор.

127. Для лечения некоторой болезни применяют четыре лекарства: a, b, c, d. Проводится сравнительное исследование трех лекарств, выбираемых случайным образом. Постройте ПЭИ опыта и выпишите элементарные исходы, составляющие события:

а) «лекарство a исследовано»;

б) «исследованы лекарства a и b»;

в) «исследовано по крайней мере одно из лекарств a или b».

Вычислите вероятности этих событий.

128. Одновременно бросают два игральных кубика. Постройте для этого опыта:

а)° ПЭИ с равновозможными исходами;

б)° ПЭИ с неравновозможными исходами;

в) событие, описываемое исходами ПЭИ а), но не описываемое исходами ПЭИ б).

129. Подбрасывают монету, а после этого бросают игральный кубик. Составьте для этого опыта:

а) ПЭИ с равновозможными исходами;

б) ПЭИ с неравновозможными исходами;

в) событие, описываемое исходами ПЭИ а), но не описываемое исходами ПЭИ б).

130. Из колоды в 36 карт наугад вынимают одну карту. Постройте ПЭИ этого опыта, принимая в качестве исходов опыта:

а) масть извлеченной карты;

б) достоинство извлеченной карты (шестерка, валет, туз и т. д.);

в) масть и достоинство извлеченной карты;

г) тот факт, является ли извлеченная карта картинкой (валет, дама, король, туз) или не является.

В каких из этих случаев исходы опыта будут равновозможными, а в каких — неравновозможными? Приведите примеры событий, которые можно описать элементами этих пространств.

131.° В ящике пять пронумерованных шаров: три белых и два черных. Наудачу вынимают три шара.

а) Постройте ПЭИ этого опыта.

б) Укажите два события, связанные с этим опытом. Опишите их элементами ПЭИ.

в) Введите элементарные вероятности.

г) Найдите вероятность того, что извлечены два белых и один черный шар.

132. Некая фирма выпускает шоколадки. В каждую шоколадку вкладывается одна из двух фотографий известных спортсменов. Фотографий каждого спортсмена одинаковое количество. Тому, кто соберет полный комплект фотографий, следующая шоколадка выдается бесплатно.

а) Постройте ПЭИ опыта, состоящего в том, что некто соберет полный комплект фотографий или купит не более трех шоколадок.

б) Укажите два события, связанные с этим опытом. Опишите их элементами ПЭИ.

в) Введите элементарные вероятности.

г) Найдите вероятность того, что будет получен приз.

133. Выпущено 100 лотерейных билетов с 11 денежными выигрышами, из которых восемь — по 10 денежных единиц, два — по 50 денежных единиц и один — 100 денежных единиц. Из купленных 25 билетов три выиграли по 10 денежных единиц и один выиграл 50 денежных единиц. Остальные остались без выигрыша. Найдите вероятность и относительную частоту события:

а) «купленный билет невыигрышный»;

б) «на приобретенный билет выпадает выигрыш 10, 50, 100 денежных единиц».

§ 2.6. Операции над событиями

Вычисление вероятности события по определению не всегда удобно и даже не всегда осуществимо. Поэтому для вычисления вероятностей событий часто пользуются правилами, дающими возможность по известным вероятностям одних событий находить вероятности других событий, которые получают из первых с помощью некоторых операций. Введение этих операций и составляет содержание данного параграфа.

2.6.1. Достоверное и невозможное события

Пусть $U = \{u_1, \ldots, u_N\}$ — ПЭИ некоторого случайного опыта.

Все ПЭИ U естественно назвать ***достоверным событием***. Оно наступает при любом исходе эксперимента.

Для удобства через V обозначим подмножество множества U, не содержащее ни одного элемента U. Такое подмножество в математике называют *пустым*. В теории вероятностей оно моделирует событие, которое не наступает ни при каком исходе эксперимента. Его называют ***невозможным событием***.

Например, в опыте с бросанием игрального кубика событие «число выпавших очков не превышает 6» достоверно, а событие «выпало 7 очков» — невозможно.

События, отличные от достоверных и невозможных, считают случайными. Обращаем внимание на то, что одно и то же событие в одной ситуации может быть достоверным, а в другой — случайным и даже невозможным. Например, если в ящике находится пять белых, пять черных и пять красных шаров, то при извлечении семи, восьми или девяти шаров мы можем получить три цвета, а можем и не получить, т. е. событие «извлечены шары трех цветов» является случайным. Если же извлекают не менее 11 шаров, то событие «извлечены шары трех цветов» является достоверным. Если извлекают менее трех шаров, то событие «извлечены шары трех цветов» является невозможным.

Причина такого парадокса кроется в том, что эти события связаны с различными ПЭИ и, по существу, являются различными событиями.

?

Какие из следующих событий — случайные, достоверные, невозможные:

а) «извлечь из ящика, содержащего белые и красные шары, три шара различных цветов»;

б) «сумма очков, выпавших при бросании двух игральных кубиков, больше 1»;

в) «сумма очков, выпавших при бросании трех игральных кубиков, кратна 3»;

г) «при подбрасывании трех монет герб не выпал ни разу»;

д) «при подбрасывании монета стала на ребро»;

е) «при подбрасывании двух монет герб выпал хотя бы один раз»;

ж) «при бросании игрального кубика выпало по крайней мере одно очко»;

з) «при раздаче колоды из 36 карт один из игроков получил шесть тузов»?

В ящике лежит несколько одинаковых по размеру катушек с нитками трех цветов: черные, белые и коричневые. Из ящика берут четыре катушки ниток. Какие из следующих событий, связанных с этим опытом, — случайные, достоверные, невозможные:

а) «среди взятых катушек хотя бы две с черными нитками»;

б) «среди взятых катушек хотя бы две с нитками одного цвета»;

в) «среди взятых катушек хотя бы три с черными нитками»;

г) «среди взятых катушек хотя бы три с нитками одного цвета»?

Основная цель теории вероятностей — изучение случайных событий. Достоверные и невозможные события являются «крайними» случаями. Случайное событие — промежуточное между достоверным и невозможным. В дальнейшем все события мы будем называть *случайными*, а невозможные и достоверные рассматривать как их частные, крайние разновидности.

Будем предполагать, что на ПЭИ заданы элементарные вероятности, т. е. построена вероятностная модель случайного опыта.

Из определения вероятности события вытекает, что вероятность случайного события всегда является числом положительным, кроме случая, когда оно невозможно. Она не может быть больше единицы (сумма всех элементарных вероятностей равна 1). Вероятность достоверного события равна 1, поскольку $P(U) = P\{u_1, u_2, ..., u_N\} = p_1 + p_2 + ... + p_N = 1$. Вероятность невозможного события равна 0: $P(V) = 0$.

Таким образом, имеем следующие свойства вероятности:

1) $P(U) = 1$;
2) $P(V) = 0$;
3) $0 \leqslant P(A) \leqslant 1$ для любого события A.

➤

Если вероятность события равна 0 или близка к 0, то это событие или невозможно, или происходит очень редко и его можно считать ***практически невозможным***. Возникает вопрос, насколько малой должна быть вероятность события, чтобы считать его практически невозможным? Понятно, что это зависит от характера события, о котором идет речь. Например, если вероятность наличия дефекта для электрической лампы равна 0,001, то событие «лампа дефектна» можно считать практически невозможным: такой процент брака можно проигнорировать. В то же время, если вероятность наличия дефекта для парашюта равна 0,001, то это означает, что из каждых 1000 парашютов приблизительно один откажет в работе. Такой результат нельзя считать удовлетворительным. В этом случае нельзя игнорировать возможность осуществления события «парашют дефектный». Другими словами, в каждой конкретной ситуации на основе ее анализа можно установить, насколько малой (или большой) должна быть вероятность, чтобы событие можно было бы считать ***практически невозможным*** (***практически достоверным***). Выиграть большой приз на тотализаторе — событие, практически невозможное.

Пример 1. Представим, что вам предложили купить электронные часы за 1 р. на следующих условиях: заплатив 1 р., вы получаете право вынуть один шар из урны, содержащей один черный шар и 149 белых. Если шар окажется черным, то часы ваши. Если шар окажется белым, то 1 р. остается у продавца.

а) Можно ли считать, что событие «вы выиграете часы» является практически невозможным?

б) Можно ли считать, что событие «часы останутся у продавца» является практически достоверным?

❐ В первую очередь нужно выбрать критерий, по которому событие будет признаваться практически невозможным или практически достоверным. Иногда пользуются таким критерием (если это не относится к жизненно важным событиям): если вероятность события не превосходит 0,01, то событие считают практически невозможным; если же вероятность события не меньше 0,99, то — практически достоверным. Воспользуемся и мы таким критерием для ответа на вопросы, поставленные в задании.

а) Вероятность события «вы выиграете часы» равна $\frac{1}{150} \approx 0{,}0067$. Оно практически невозможно.

б) Вероятность события «часы останутся у продавца» равна $\frac{149}{150} \approx 0{,}9933$. Оно практически достоверно. ■ ◀

2.6.2. Противоположные события

Часто возникает необходимость по информации, содержащейся в одних событиях, делать выводы о шансах наступления других событий, каким-то образом связанных с первыми. Точнее, необходимо уметь одни события выражать через другие, или по одним событиям строить новые. Конечно, речь идет о событиях, связанных с одним случайным опытом, т. е. о событиях, являющихся подмножествами одного и того же ПЭИ.

Если опыт завершился исходом, не содержащимся в некотором событии, то это событие не наступило. Событие $\overline{A}$, составленное из тех и только тех элементов ПЭИ, которые не содержатся в событии A, называют *противоположным* событию A. Оно наступает тогда и только тогда, когда A не наступает. В опыте с бросанием игрального кубика событие $\overline{A} = \{1, 2, 3, 4\}$ противоположно событию A — «число выпавших очков больше 4», $A = \{5, 6\}$. На рис. 23 событие $\overline{A}$ изображается совокупностью 20 точек, не входящих в событие A.

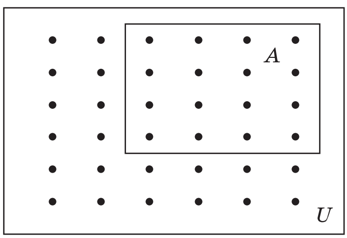


Рис. 23

Событие $\overline{A}$, которое происходит тогда и только тогда, когда событие A не происходит, называют **противоположным** *событию A.*

Ясно, что если событие $\overline{A}$ противоположно событию A, то событие A противоположно событию $\overline{A}$. События A и $\overline{A}$ называют *противоположными*.

Так, в опыте с одним выстрелом по мишени событию A — «попадание в цель» противоположно событие $\overline{A}$ — «непопадание в цель». В опыте с приобретением лотерейных билетов событию A — «ни один из билетов не выиграл» противоположно событие $\overline{A}$ — «хотя бы один из билетов выиграл». Противоположными являются выпадение единицы и невыпадение единицы при подбрасывании игрального кубика; непопадание в цель ни при одном из трех выстрелов и попадание в цель хотя бы при одном выстреле и т. д.

? Из ящика, содержащего белые и красные шары, извлекли пять шаров. Какие из приведенных событий являются парами противоположных событий:
а) «извлекли по крайней мере один белый шар»;
б) «извлекли более одного белого шара»;
в) «среди извлеченных шаров нет белых»;
г) «извлекли один белый шар»?

Вероятность события $\overline{A}$, противоположного событию A, равна сумме вероятностей всех исходов ПЭИ, не входящих в A. Так как сумма вероятностей всех элементарных исходов ПЭИ равна 1, то

$$P(\overline{A}) = 1 - P(A).$$

2.6.3. Пересечение, или произведение, событий

Многие случайные события образуются из двух событий A, B с помощью союза «и»: A и B. Например:

- событие «сумма выпавших очков равна 2» при двукратном бросании игрального кубика образуется из событий «при первом броске выпала 1» и «при втором броске выпала 1»;
- событие «только первое изделие браковано» при контроле двух изделий образуется из событий «первое изделие браковано» и «второе изделие не браковано»;

- событие «два промаха» при двух бросках мяча в корзину образуется из событий «промах при первом броске» и «промах при втором броске»;
- событие «электрическая цепь, составленная из двух элементов, соединенных параллельно, не работает» образуется из двух событий «первый элемент не работает» и «второй элемент не работает»;
- событие «дважды выпала цифра» при двух подбрасываниях монеты образуется из двух событий «при первом подбрасывании выпала цифра» и «при втором подбрасывании выпала цифра».

Как же строится событие «A и B» в вероятностной модели случайного опыта? Пусть каждое из событий A, B состоит из некоторых элементов ПЭИ. Тогда событие, заключающееся в том, что произошло и событие A, и событие B, должно содержать общие элементарные исходы, входящие и в A, и в B и только их, т. е. состоять из общей части множеств элементарных исходов, образующих эти события. Эту общую часть в математике называют ***пересечением*** двух множеств.

Событие C, составленное из тех и только тех элементов ПЭИ, которые принадлежат одновременно и событию A, и событию B, называют ***пересечением****, или* ***произведением****, событий A и B; его обозначают*:

$$C = A \cap B, \text{ или } C = AB.$$

Использование двух терминов (пересечение и произведение) для результата рассмотренной операции над событиями вполне оправдано. На языке множеств — пересечение, на языке алгебры — произведение. Мы будем пользоваться обоими терминами.

Графическая иллюстрация операции пересечения (или умножения) событий A и B приведена на рис. 24. На нем четыре точки являются общими для событий A и B. Если опыт закончится одним из этих четырех исходов, то наступают оба события A и B, поскольку этот исход содержится и в A, и в B.

Событие $C = A \cap B$ наступает тогда и только тогда, когда наступает и событие A, и событие B одновременно. Нетрудно видеть, что

$$A \cap A = A; \quad A \cap V = V;$$
$$A \cap \overline{A} = V; \quad A \cap U = A$$

для любого события A (здесь U — достоверное событие, V — невозможное событие).

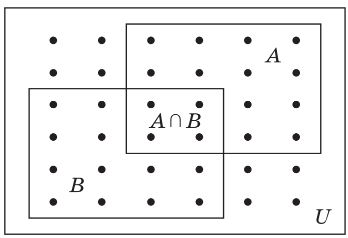


Рис. 24

Пример 2. Так, в опыте с бросанием игрального кубика произведением событий A — «выпало четное число очков» и B — «число выпавших очков кратно 3» является событие C — «выпало 6 очков». ■

Пример 3. В опыте, заключающемся в том, что контролю подвергают два изделия, событие «оба изделия бракованы» является произведением двух событий: «первое изделие браковано» и «второе изделие браковано». ■

Пример 4. Отец играет с сыном в шашки до первого поражения сына. Событие A_i, $i = 1, 2, 3, \ldots$, означает, что сын выиграл i-ю партию.

а) Что означает событие $B = A_1A_2$?

б) Выразить через события A_i событие «состоялись только две игры».

❒ а) Событие $B = A_1A_2$, равное произведению событий A_1 и A_2, состоит из тех исходов, которые входят и в A_1, и в A_2; происходит тогда и только тогда, когда происходит и событие A_1, и событие A_2, т. е. когда сын выиграл первую и вторую партии. Итак, событие B означает, что сын выиграл первые две партии.

б) Событие C — «состоялись только две игры» происходит тогда и только тогда, когда сын выиграл первую партию (игра продолжается) и проиграл вторую партию (игра прекращена). Другими словами, когда происходит событие A_1 и событие $\overline{A}_2$, т. е. $C = A_1\overline{A}_2$. ■

В каждой вероятностной модели существуют события, не имеющие общих элементарных исходов в своем составе, т. е. события, которые не могут произойти одновременно. Например:

- в опыте с бросанием игрального кубика не могут одновременно наступить события: «число выпавших очков меньше 3» и «число выпавших очков есть число составное». Первое состоит из элементарных исходов $\{1, 2\}$, второе — $\{4, 6\}$;
- в опыте с двукратным подбрасыванием монеты события «герб выпал ровно один раз» и «герб не выпал ни разу» не происходят одновременно, они состоят из различных исходов: {ГЦ, ЦГ} и {ЦЦ};
- в опыте с опросом по поводу жилищных условий работников предприятия события «опрошенный удовлетворен качеством жилья и удаленностью квартиры от места работы» и «опрошенный не удовлетворен качеством жилья и не удовлетворен удаленностью квартиры от места работы» не происходят одновременно, они состоят из различных исходов $\{11\}$ и $\{00\}$ соответственно.

Тот факт, что два события не происходят одновременно, означает, что множества их элементарных исходов не содержат общих элементов, т. е. что их пересечение есть пустое множество: $A \cap B = V$.

Если $A \cap B = V$, то события A и B называют ***несовместными*** (рис. 25). События A, $\bar{A}$ — несовместны.

В опыте с бросанием игрального кубика события A — «выпало больше 4 очков» и B — «выпало меньше 3 очков» — несовместны.

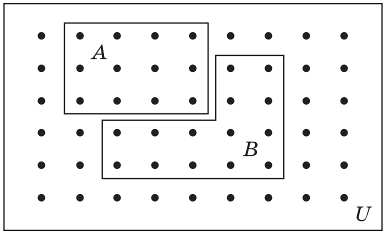


Рис. 25

Какие пары событий из перечисленных ниже являются несовместными в опыте с бросанием игрального кубика:

а) «число выпавших очков кратно 3»;
б) «число выпавших очков простое»;
в) «число выпавших очков меньше 3»;
г) «число выпавших очков составное»;
д) «число выпавших очков четно»;
е) «число выпавших очков больше 4»;
ж) «число выпавших очков не делится на 3»?

2.6.4. Объединение, или сумма, событий

Еще один способ образования нового события из двух событий A, B связан с использованием союза «или»: A или B. Например:

- событие «сумма выпавших очков равна 3» при двукратном бросании игрального кубика образуется из событий «при первом броске выпала 1, при втором — 2» или «при первом броске выпало 2 очка, при втором — 1»;
- событие «хотя бы одно изделие браковано» при контроле двух изделий образуется из событий «первое изделие браковано», или «второе изделие браковано», или «оба изделия бракованы»;
- событие «ровно один промах» при двух бросках мяча в корзину образуется из событий «промах при первом броске и попадание при втором» или «попадание при первом броске и промах при втором»;
- событие «электрическая цепь, составленная из двух элементов, соединенных последовательно, не работает» образуется из двух событий «первый элемент не работает» или «второй элемент не работает» или «оба элемента не работают»;

- событие «ровно один раз выпала цифра» при двух подбрасываниях монеты образуется из двух событий «при первом подбрасывании выпала цифра, а при втором — герб» или «при втором подбрасывании выпала цифра, а при первом — герб».

В русском языке союз «или» употребляют в двух смыслах: исключающем и не исключающем. Мы будем употреблять этот союз в не исключающем смысле.

Как же строится событие «A или B» в вероятностной модели случайного опыта? Пусть каждое из событий A, B состоит из некоторых элементов ПЭИ. Тогда событие, заключающееся в том, что произошло или событие A, или событие B, или оба события A и B, должно содержать элементарные исходы, входящие или в A, или в B, или в оба этих события и только их, т. е. состоять из объединения множества A и той части множества B, которая не входит в A. Такое множество в математике называют ***объединением*** двух множеств.

Событие D, составленное из тех и только тех элементов ПЭИ, которые принадлежат или событию A, или событию B, или событиям A и B одновременно, называют ***объединением****, или* ***суммой****, событий A и B; его обозначают:*

$$D = A \cup B, \quad \text{или} \quad D = A + B.$$

Другими словами, событие $A \cup B$ наступает тогда и только тогда, когда или наступает событие A, или наступает событие B, или A и B одновременно.

28 точек на рис. 24 принадлежат по крайней мере одному из событий или A, или B. Если опыт закончится одним из этих исходов, то наступает или событие A, или B, или оба эти события одновременно.

Ясно, что для любого события A имеют место равенства:

$$A \cup U = U; \quad A \cup A = A; \quad A \cup \overline{A} = U; \quad A \cup V = A.$$

Пример 5. В опыте с бросанием игрального кубика суммой событий A — «выпало четное число очков» и B — «число выпавших очков кратно 3» является событие C — «выпало любое число очков, кроме 1 и 5». ■

Пример 6. В опыте, заключающемся в том, что контролю подвергают два изделия, событие «ровно одно изделие браковано» является суммой двух событий: «первое изделие браковано, а второе пригодно», «первое изделие пригодно, а второе изделие браковано». ■

Пример 7. Отец играет с сыном в шашки до первого поражения сына. Событие A_i, $i = 1, 2, 3, \ldots$, означает, что сын выиграл i-ю партию.

а) Что означает событие $B = A_1 + A_2$?

б) Выразить через события A_i событие «число проведенных партий отлично от двух».

❒ а) Событие $B = A_1 + A_2$, равное сумме событий A_1 и A_2, происходит тогда и только тогда, когда происходит или событие A_1 или событие A_2, т. е. когда сын выиграл первую или вторую партию. Итак, событие B означает, что проведено более одной партии.

б) Событие C — «число проведенных партий отлично от двух» происходит тогда и только тогда, когда сын проиграл первую партию (игра состоит из одной партии) или выиграл вторую партию (игра продолжается). Другими словами, когда происходит событие $\overline{A}_1$ или событие A_2, т. е. $C = \overline{A}_1 + A_2$. ■

2.6.5. Обобщение операций объединения и пересечения событий

Введенные операции сложения и умножения событий обобщаются на любое конечное число событий. Пусть $A_1, A_2, \ldots, A_n$ — события, связанные с данным случайным опытом, т. е. подмножества ПЭИ U.

***Пересечением**, или **произведением**, событий $A_1, A_2, \ldots, A_n$ называют событие, состоящее из тех и только тех исходов, которые являются общими для всех данных событий, и обозначают*:

$$A_1 \cap A_2 \cap \ldots \cap A_n, \quad \text{или} \quad \bigcap_{k=1}^{n} A_k, \quad \text{или} \quad A_1 \cdot A_2 \cdot \ldots \cdot A_n.$$

Оно происходит тогда и только тогда, когда происходит и событие A_1, и событие A_2, и т. д., и событие A_n.

***Объединением**, или **суммой**, событий $A_1, A_2, \ldots, A_n$ называют событие, состоящее из всех тех и только тех исходов, которые принадлежат хотя бы одному из данных событий и обозначают:*

$$A_1 \cup A_2 \cup \ldots \cup A_n, \quad \text{или} \quad \bigcup_{k=1}^{n} A_k, \quad \text{или} \quad A_1 + A_2 + \ldots + A_n.$$

Оно происходит тогда и только тогда, когда происходит хотя бы одно из событий $A_1, A_2, \ldots, A_n$.

Пример 8. Баскетболист сделал три броска мяча в корзину. Пусть событие A_i означает попадание в корзину при i-м броске ($i = 1, 2, 3$).

а) Что означают события:

$$B = A_1 \cup A_2; \quad C = A_1 \cap A_2 \cap A_3; \quad D = A_1 \cap A_2 \cap \overline{A}_3\,; \quad E = A_1 \cup A_2 \cup A_3?$$

б) Выразить через события A_1, A_2 и A_3 следующие события: F — «спортсмен попал в корзину лишь при первом броске»; G — «имело место три попадания»; H — «попадание произошло только при одном из бросков».

❒ а) B — «попадание имело место при первом или втором броске»; C — «попадание имело место при всех трех бросках»; D — «попадание имело место только при первых двух бросках»; E — «произошло хотя бы одно попадание».

б) Событие F означает, что спортсмен попал при первом броске, и не попал при втором, и не попал при третьем; событие G означает, что спортсмен попал и при первом, и при втором, и при третьем броске; событие H означает, что имело место попадание или только при первом, или только при втором, или только при третьем броске. Поэтому:

$$F = A_1 \cap \bar{A}_2 \cap \bar{A}_3; \quad G = A_1 \cap A_2 \cap A_3;$$

$$H = (A_1 \cap \bar{A}_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap A_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap \bar{A}_2 \cap A_3). \blacksquare$$

Контрольные вопросы

1. Из ящика, который содержит белые и черные шары, вынимают наугад четыре шара. Какое событие противоположно событию:

а) «вынут хотя бы один белый шар»;

б) «вынуто более двух белых шаров»;

в) «среди вынутых шаров белых нет»?

2. Бросают игральный кубик. Пусть событие A — «выпало 4 очка», B — «выпало четное число очков», C — «выпало нечетное число очков». Будут ли несовместными события A и B; B и C?

3. Может ли сумма двух событий совпадать с одним из слагаемых? Что можно тогда сказать о другом из событий?

4. Что можно сказать о событиях A и B, если их произведение есть:

а) достоверное событие;

б) невозможное событие?

5. Известно, что всякий исход, принадлежащий событию B, содержится и в A. Чему равно:

а) $A \cup B$; б) $A \cap B$?

6. Может ли быть достоверным событием сумма двух несовместных событий?

7. Рассмотрим такую игру. Ведущий записывает на листе бумаги любое число от 1 до 1000. Если вы его угадаете, то получаете приз. Можно ли считать практически невозможным событие «число угадано»?

Задачи

134.° Электрическая цепь состоит из двух элементов. Пусть A_i — событие, заключающееся в том, что выйдет из строя i-й элемент, $i = 1, 2$. Что означает событие:

а) $A_1 \cap A_2$; б) $\overline{A}_1 \cap \overline{A}_2$; в) $A_1 \cap \overline{A}_2$; г) $A_1 \cup A_2$?

135. Прибор состоит из блока I типа и двух блоков II типа. Событие A — «исправен блок I типа», B_i ($i = 1, 2$) — «исправен i-й блок II типа». Прибор работает, если исправен блок I типа и хотя бы один из блоков II типа. Выразите через A и B_i событие C, означающее, что прибор работает.

136. Электрическая цепь составлена по схеме, приведенной на рис. 26. Выход из строя элемента a_i ($i = 1, 2$) — событие A_i, элемента b_i ($i = 1, 2$) — событие B_i. Выразите через A_i и B_i события C и $\overline{C}$, если C означает разрыв цепи.

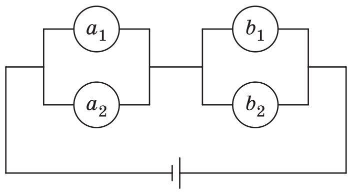


Рис. 26

137.° Произведено два выстрела по мишени. Рассмотрим следующие события:

B_1 — «попадание имело место при обоих выстрелах»;

B_2 — «попадание имело место только при первом выстреле»;

B_3 — «попадание имело место только при втором выстреле»;

B_4 — «при обоих выстрелах имел место промах».

а) Выразите через эти события следующие события: A_1 — «попадание имело место при первом выстреле»; A_2 — «попадание имело место при втором выстреле»;

б) Выразите события B_i ($i = 1, 2, 3, 4$) через события A_1 и A_2.

в) Что означают события: $B_2 + B_3$; B_1B_2; $B_1\overline{B}_4$?

§ 2.7. Шансы в пользу события

Мы часто прибегаем к заключению пари, когда не можем аргументами убедить своего оппонента в справедливости своего прогноза или своей точки зрения. Не будем рассматривать те пари, когда один из спорящих наверняка знает, что возобладает его точка зрения. Например, обсуждается вопрос о том, как закончился матч между двумя футбольными командами, репортаж о котором будет транслироваться только поздно вечером. Но один из спорящих из Интернета уже знает исход этого матча и все-таки идет на заключение пари. Такие пари будем называть *недобросовестными* и в дальнейшем рассматривать не будем. Будем рассматривать только такие пари, когда каждый из спорящих считает, что наступление любого из исходов есть дело случая, его нельзя заранее предсказать, но у спорящих различные представления о вероятностях наступления этих исходов.

Как правило, спорящие отстаивают различные, противоположные точки зрения. Например, один утверждает, что дождь сегодня пойдет, другой — что не пойдет; при извлечении карты из колоды появится красная карта, черная карта; останется в живых герой фильма или погибнет и т. д. Такие события являются противоположными.

Если проведено n опытов, и в m из них произошло событие A, то в $n - m$ опытах событие A не произошло, т. е. произошло событие $\bar{A}$. Поэтому относительные частоты событий A и $\bar{A}$ соответственно равны $\frac{m}{n}$ и $\frac{n-m}{n} = 1 - \frac{m}{n}$. Тогда вероятности событий A и $\bar{A}$, связанные с проведением достаточно большого количества статистически устойчивых опытов, естественно принять равными их относительным частотам:

$$P(A) = \frac{m}{n};\quad P(\bar{A}) = 1 - \frac{m}{n}.$$

Если же рассматривается опыт с N равновозможными исходами, в котором наступлению события A благоприятствует M исходов, то наступлению события $\bar{A}$ будет благоприятствовать $N - M$ исходов, и вероятности этих событий соответственно равны:

$$P(A) = \frac{M}{N};\quad P(\bar{A}) = \frac{N-M}{N} = 1 - \frac{M}{N}.$$

В любом случае, как мы установили в предыдущем параграфе, имеет место равенство:

$$P(\bar{A}) = 1 - P(A).$$

Относительные возможности событий A и $\bar{A}$ по сравнению друг с другом будем выражать термином «*шансы в пользу* A», или «*шансы против* $\bar{A}$»:

$$\textit{шансы в пользу}\ A = \frac{P(A)}{P(\bar{A})}.$$

Понятно, что *шансы в пользу* $\overline{A} = \dfrac{P(\overline{A})}{P(A)}$.

Так, например, вероятность события 0,5 соответствует шансу $\dfrac{0{,}5}{1-0{,}5} = 1$. Это иногда формулируют в виде «шансы 1 к 1». Вероятность 0,75 соответствует шансу $\dfrac{0{,}75}{1-0{,}75} = 3$, или 3 к 1. Больший шанс соответствует более высокой вероятности и большему правдоподобию. Обратите внимание, что, несмотря на то что вероятность не может выходить за пределы промежутка от 0 до 1, шанс может принимать любые неотрицательные значения.

В случае большого числа n статистически устойчивых опытов имеем

$$\text{шансы в пользу } A = \frac{P(A)}{P(\overline{A})} = \frac{\frac{m}{n}}{\frac{n-m}{n}} = \frac{m}{n-m}.$$

Для опытов с N равновозможными исходами аналогично получим

$$\text{шансы в пользу } A = \frac{P(A)}{P(\overline{A})} = \frac{\frac{M}{N}}{\frac{N-M}{N}} = \frac{M}{N-M}.$$

Пример 1. Рассмотрим опыт с бросанием игрального кубика. Пусть событие A означает, что число выпавших очков кратно 3. Тогда событие $\overline{A}$ означает, что число выпавших очков не кратно 3. Так как событию A благоприятствуют два исхода $\{3, 6\}$ из шести возможных $\{1, 2, 3, 4, 5, 6\}$, значит событию $\overline{A}$ — четыре исхода $\{1, 2, 4, 5\}$, то $P(A) = \frac{2}{6} = \frac{1}{3}$; $P(\overline{A}) = 1 - \frac{1}{3} = \frac{2}{3}$. Поэтому

$$\text{шансы в пользу } A = \frac{2}{4} = \frac{1}{2};$$

$$\text{шансы в пользу } \overline{A} = \frac{4}{2} = \frac{2}{1}. \blacksquare$$

Пример 2. Пусть некоторый водитель путем многократных наблюдений установил, что в 85% тех случаев, когда он моет свой автомобиль, на следующий день идет дождь. Другими словами, можно считать, что вероятность того, что на следующий день после мытья машины пойдет дождь, равна 0,85, а вероятность того, что на следующий день после мытья машины не пойдет дождь, равна $1 - 0{,}85 = 0{,}15$. Тогда шансы в пользу дождя $\frac{0{,}85}{0{,}15} = \frac{17}{3}$. ■

Пример 3. Пусть принято, что вероятность того, что некоторая лошадь выиграет на скачках, равна $\frac{3}{4}$. Тогда шансы сторон в пари в пользу победы этой лошади относятся как $\frac{3}{4} : \frac{1}{4} = 3 : 1$. Возникает вопрос, на каких условиях следует заключать пари, чтобы оно было честным? Ясно, что тот, кто ставит на эту лошадь, рискует меньше, чем тот, кто ставит против этой лошади. Можно считать, что этот риск в 3 раза меньше риска второго участника пари. Поэтому ставки при честном пари должны быть обратно пропорциональны вероятностям выигрыша и проигрыша лошади. Для того чтобы пари было честным, надо условиться платить тому, кто ставит против этой лошади, например, 30 (или 60, или 90) денежных единиц, если эта лошадь проиграет, и взимать с него 10 (соответственно, 20, 30) денежных единиц, если она выиграет. Подсчитаем суммарный выигрыш при большом количестве таких пари для того, кто ставит против этой лошади. В $\frac{1}{4}$ случаев (а именно в такой доле случаев проигрывает рассматриваемая лошадь) выигрыш составляет 30 денежных единиц, а в $\frac{3}{4}$ случаев, когда эта лошадь выигрывает, проигрыш составляет 10 денежных единиц. Суммарный выигрыш в среднем равен $30 \cdot \frac{1}{4} - 10 \cdot \frac{3}{4} = 0$. Для другого участника пари выигрыш составляет 10 денежных единиц (лошадь выигрывает в $\frac{3}{4}$ случаев), а проигрыш — 30 денежных единиц (лошадь не выигрывает в $\frac{1}{4}$ случаев). Суммарный выигрыш в среднем для него равен $10 \cdot \frac{3}{4} - 30 \cdot \frac{1}{4} = 0$. ■

В дальнейшем *пари будем считать честным, если средняя величина выигрыша, приходящегося на одно пари, равна нулю.*

Если шансы в пользу события A равны $p : q$, выигрыш игрока, поставившего на событие A, составляет a денежных единиц, а проигрыш этого игрока — b денежных единиц, то суммарный выигрыш в среднем для этого игрока, по аналогии с вышерассмотренным примером, равен $ap - bq = ap - b(1 - p)$. Если эта величина равна нулю, то пари является честным. Если $ap - bq = 0$, то $\frac{p}{q} = \frac{b}{a}$, т. е. отношение $b : a$ также равно шансам в пользу события A. Об этом отношении также говорят, что оно определяет *условия, при которых пари будет честным*.

В примере 2 с мытьем машины пари будет честным, если водитель в случае дождя на следующий день после мытья машины получит, например, 15 денежных единиц, а в случае, если дождя не будет, заплатит 85 денежных единиц.

Если известно, что шансы в пользу события A равны $\frac{p}{q}$, то

$$P(A) = \frac{p}{p+q}; \quad P(\bar{A}) = \frac{q}{p+q}.$$

❒ Действительно, пусть $P(A) = h$, $P(\bar{A}) = 1 - h$, тогда шансы в пользу события A равны $\frac{h}{1-h}$. Имеем уравнение $\frac{h}{1-h} = \frac{p}{q}$. Решим это уравнение относительно h. Согласно основному свойству пропорции имеем:

$$hq = (1 - h)p, \text{ или } hq = p - hp,\ hq + hp = p,\ h(p + q) = p,\ h = \frac{p}{p+q},$$

что и требовалось доказать. ■

Пример 4. Сергей готов заключить пари на условиях 5 : 3, что при выстреле по мишени он попадет в нее. Какой должна быть вероятность попадания в мишень для того, чтобы пари было честным?

❒ Здесь $p = 5$, $q = 3$, A — «попадание в мишень при выстреле». Согласно полученной формуле $P(A) = \frac{5}{5+3} = \frac{5}{8}$. ■

Формулой для подсчета шансов в пользу некоторого события пользовались итальянские математики Л. Пачиоли (1445–1514), Д. Кардано (1501–1576), Н. Тарталья (1499–1557), французские ученые Б. Паскаль (1623–1662), П. Ферма (1601–1665) еще до того, как было сформулировано определение вероятности.

Контрольные вопросы

1. Вероятность некоторого события равна 0,65. Каковы шансы в пользу этого события?

2. Известно, что шансы в пользу некоторого события равны 1 : 3. Какова вероятность этого события?

3. Известно, что шансы в пользу некоторого события меньше 1. Что можно сказать о вероятности наступления этого события?

4. На каких условиях должно заключаться честное пари о том, что наступит событие A, если вероятность его наступления равна 0,3?

5. Будет ли пари о том, что наступит событие A, честным, если вероятность этого события равна 0,6 и пари заключено на условиях 3 : 2?

Задачи

138.° Из колоды в 36 карт наугад выбирают одну карту. Каковы шансы в пользу того, что это не будет картинка (шестерка, семерка, восьмерка, девятка, десятка)?

139. В тесте по истории учащемуся названы три даты — 1825, 1812 и 1861 гг. — и сказано, что это даты трех исторических событий: начала первой Отечественной войны, отмены крепостного права и восстания декабристов. Предложено указать дату каждого из этих трех событий. Учащийся не знает, когда произошли эти события и называет даты наугад.

а) Постройте ПЭИ опыта, заключающегося в том, что учащийся наугад выбирает из предложенных дату каждого из этих исторических событий.

б) Введите элементарные вероятности исходов этого опыта.

в) Каковы шансы в пользу наступления события «ученик не угадал ни одной даты»?

140. Лена во время пребывания в летнем лагере написала три письма: маме, дедушке и подруге. Она положила их в конверты, не сверяя с заранее написанными адресами, не проверив, соответствует ли письмо адресу на конверте, и опустила письма в почтовый ящик.

а) Постройте ПЭИ опыта, заключающегося в выборе наугад конвертов для писем.

б) Введите элементарные вероятности исходов этого опыта.

в) Каковы шансы в пользу наступления события «точно одно письмо будет отправлено своему адресату»?

141.° Подбрасывают три монеты.

а) Постройте ПЭИ этого опыта.

б) Введите элементарные вероятности исходов этого опыта.

в) Каковы шансы в пользу наступления события «выпадут три герба или три цифры»?

§ 2.8. Вероятность суммы событий

Теперь рассмотрим правила, с помощью которых можно по вероятностям одних случайных событий вычислять вероятности других событий, каким-то образом связанных с первыми. Необходимость подобных правил иллюстрирует следующий пример.

Пример 1. Прибор, состоящий из двух блоков, выходит из строя, если выходят из строя оба блока. Вероятность безотказной работы за определенный промежуток времени первого блока составляет 0,9, второго — 0,8, обоих блоков — 0,75. К вычислению вероятности какого события сводится нахождение вероятности безотказной работы прибора за указанный промежуток времени?

❒ В этом задании даны вероятности трех событий: A — «первый блок работает безотказно в течение определенного промежутка времени», B — «второй блок работает безотказно в течение определенного промежутка времени», AB — «оба блока работают безотказно в течение определенного промежутка времени». Требуется найти вероятность события C — «прибор работает безотказно в течение определенного промежутка времени», являющегося суммой событий A и B: $C = A + B$. ■

Чтобы найти правило, позволяющее по вероятностям двух событий найти вероятность их суммы, рассмотрим следующий пример.

Пример 2. Рассмотрим опыт с бросанием дважды игрального кубика и события A — «число очков, выпавших при первом броске, больше 4» и B — «при втором броске выпало более 3 очков». Найти $P(A + B)$.

❒ Событие A образуют 12 исходов, расположенных в последних двух строках ПЭИ (табл. 2.22), событие B составляют 18 исходов, записанных в последних трех столбцах той же таблицы. Событие $A + B$ содержит 24 исхода, следовательно, $P(A + B) = \frac{24}{36} = \frac{2}{3}$. Вычисление числа исходов, составляющих событие $A + B$, можно выполнить следующим образом: $24 = 12 + 18 - 6$, где $N(A + B) = 24$, $N(A) = 12$, $N(B) = 18$, $N(AB) = 6$. Отсюда получаем, что

$$P(A + B) = \frac{12 + 18 - 6}{36} = \frac{N(A)}{N} + \frac{N(B)}{N} - \frac{N(AB)}{N} =$$

$$= P(A) + P(B) - P(AB). \blacksquare$$

Таблица 2.22

I	II					
	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Этот результат имеет место и в общем виде. Обычно его называют ***теоремой сложения вероятностей***.

ТЕОРЕМА 1. *Для любых двух событий A и B справедливо равенство*

$$P(A + B) = P(A) + P(B) - P(AB),$$

т. е. вероятность суммы двух событий равна сумме их вероятностей минус вероятность их произведения.

❒ Докажем это утверждение. По определению вероятности события (см. § 2.5) вероятность $P(A + B)$ равна сумме вероятностей элементарных исходов, входящих в $A + B$. В то же время $P(A) + P(B)$ — это сумма вероятностей элементарных исходов, составляющих событие A, и вероятностей элементарных исходов, образующих событие B. При этом вероятности исходов, входящих в AB, суммируются дважды: эти исходы входят и в событие A, и в событие B. Рис. 27 наглядно иллюстрирует этот факт. Если вычесть $P(AB)$ из $P(A) + P(B)$, то получится сумма вероятностей исходов, составляющих событие $A + B$. ■

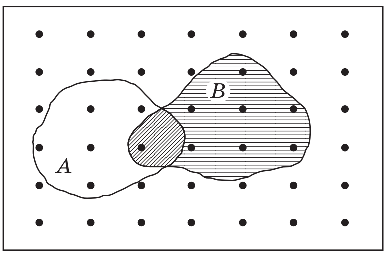


Рис. 27

?

Верно ли, что вероятность суммы двух событий не больше суммы вероятностей этих событий?

Теперь вернемся к рассмотрению примера 1 и вычислим вероятность события C.

❐ По условию, $P(A) = 0{,}9$, $P(B) = 0{,}8$, $P(AB) = 0{,}75$. Требуется найти $P(C) = P(A + B)$. По теореме сложения вероятностей

$$P(C) = P(A + B) = 0{,}9 + 0{,}8 - 0{,}75 = 0{,}95. \blacksquare$$

Пример 3. Предыдущий опыт работы мастерской свидетельствует о том, что вероятность перегорания предохранителя в неисправном приборе составляет 6%, а вероятность обрыва провода — 4%. Также в 1% случаев приборы поступали с перегоревшим предохранителем и обрывом провода. Найти вероятность того, что в конкретном сданном в ремонт приборе присутствует по крайней мере одна из этих неисправностей.

❐ Обозначим через A и B соответственно события: «перегорел предохранитель» и «обрыв провода». Событие «перегорел предохранитель или есть обрыв провода» является их суммой. Его вероятность, по теореме сложения вероятностей, равна $0{,}06 + 0{,}04 - 0{,}01 = 0{,}09$.

Таким образом, в 9% случаев обращений в мастерскую прибор имеет одну из этих неисправностей или обе сразу. ■

Пример 4. Числа 1, 2, 3, ..., 20 написаны на листах бумаги, которые помещены в коробку и тщательно перемешаны. Из коробки наугад вынимают один лист. Какова вероятность того, что число на вынутом листе окажется либо простым, либо делящимся на 3?

❐ Пространство элементарных исходов данного опыта имеет вид $U = \{1, 2, 3, \ldots, 20\}$, A — «на вынутом листе простое число», $A = \{2, 3, 5, 7, 11, 13, 17, 19\}$; B — «на вынутом листе число, кратное 3», $B = \{3, 6, 9, 12, 15, 18\}$.

Так как исходы опыта равновозможны, то $P(A) = \frac{8}{20} = \frac{2}{5}$; $P(B) = \frac{6}{20} = \frac{3}{10}$. Событие AB заключается в том, что на вынутом листе простое число, кратное 3: $AB = \{3\}$, $P(AB) = \frac{1}{20}$. Таким образом,

$$P(A + B) = \frac{2}{5} + \frac{3}{10} - \frac{1}{20} = \frac{13}{20}.$$

Впрочем, эту вероятность можно найти и непосредственно:

$$A + B = \{2, 3, 5, 6, 7, 9, 11, 12, 13, 15, 17, 18, 19\},$$

$$P(A + B) = \frac{13}{20}. \blacksquare$$

Из теоремы сложения вероятностей следует, что, зная любые три из четырех вероятностей ($P(A)$, $P(B)$, $P(A + B)$, $P(AB)$), можно найти четвертую неизвестную вероятность. В частности, $P(AB) = P(A) + P(B) - P(A + B)$.

➤

В каких случаях оказываются полезными эти формулы? Один из случаев их применения состоит в том, чтобы взять в качестве исходных известные сведения о вероятностях и вычислить по соответствующей формуле вероятность другого события, возможно, представляющего больший интерес или являющегося более важным. Другой случай — если мы хотим убедиться, что информация, на которой основано решение, логически непротиворечива. Предположим, например, что у нас есть вероятности событий A и B, вычисленные как относительные частоты на основе данных прошлых наблюдений. Планируется использовать субъективную оценку вероятностей событий AB и $A + B$. При этом может оказаться полезным убедиться в том, что связь между четырьмя рассматриваемыми вероятностями не противоречит приведенным формулам. ➤

? Вероятность попадания в цель одним стрелком равна p_1, вторым — p_2, вероятность того, что хотя бы один стрелок попадет в цель равна p_3. Вероятность того, что оба стрелка попадут в цель равна p_4. Выразите через остальные данные каждую из вероятностей p_1, p_2, p_3, p_4.

Приведенные примеры показывают, что трудности применения теоремы сложения вероятностей связаны с нахождением вероятности произведения событий (либо эта вероятность задавалась, либо находилась по классической модели из предположения о равновозможности исходов опыта). Эти трудности исчезают, если известно, что события несовместны, т. е. нет элементарных исходов, входящих в оба события; другими словами, если $AB = V$. В этом случае $P(AB) = 0$ и теорема сложения вероятностей принимает следующий вид.

ТЕОРЕМА 2. *Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий*:

$$P(A + B) = P(A) + P(B).$$

ЗАМЕЧАНИЕ 1. Полученная формула напоминает известное равенство

$$S(F_1 + F_2) = S(F_1) + S(F_2),$$

где F_1 и F_2 — фигуры, не имеющие общих внутренних точек, а $S(F_1)$ и $S(F_2)$ — их площади. Эта общность на самом деле очень глубокая, она отражает тот факт, что и площадь фигуры, и вероятность события являются мерами величин.

ЗАМЕЧАНИЕ 2. Точно так же, как равенство $S(F_1 + F_2) = S(F_1) + S(F_2)$ используют при аксиоматическом определении площади фигуры, так и равенство $P(A + B) = P(A) + P(B)$ может быть принято в качестве одной из аксиом при определении вероятности события.

Пример 5. Найти вероятность того, что сумма очков, выпавших при бросании двух игральных кубиков, равна 3 или 4.

❒ ПЭИ опыта представлено в таблице 2.22. Пусть событие A — «сумма выпавших очков равна 3», B — «сумма выпавших очков равна 4».

$$A = \{(1, 2), (2, 1)\}; \quad B = \{(1, 3), (2, 2), (3, 1)\},$$

A и B — несовместные события.

$$P(A + B) = P(A) + P(B) = \frac{2}{36} + \frac{3}{36} = \frac{5}{36}. \ ■$$

? Найдите ошибку в следующих рассуждениях.
Вероятность попадания в цель одним стрелком равна 0,8, вторым — 0,7. Стрелки сделали по одному выстрелу. Какова вероятность того, что хотя бы один из них попал в цель? Применение предыдущего следствия приводит к нелепому результату: $0{,}8 + 0{,}7 = 1,5$, ведь вероятность не может быть больше 1.

При решении задач нередко приходится вычислять вероятности событий, являющиеся объединением более чем двух событий. Правило сложения вероятностей имеет место для суммы любого конечного числа попарно несовместных событий:

Если $A_1, A_2, \ldots, A_m$ — попарно несовместные события, то

$$P(A_1 + A_2 + \ldots + A_m) = \\ = P(A_1) + P(A_2) + \ldots + P(A_m).$$

Рис. 28 является иллюстрацией этого утверждения.

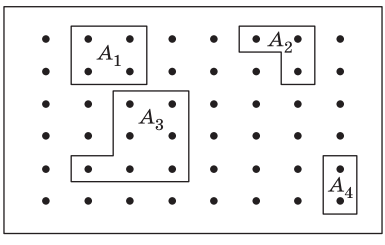


Рис. 28

?

Верно ли, что попарно несовместные события несовместны в совокупности?

Верно ли, что события, несовместные в совокупности, являются попарно несовместными?

Обращаем внимание на то, что теорема сложения вероятностей справедлива для событий, относящихся к одному ПЭИ.

➤

Предыдущее утверждение позволяет по вероятностям событий находить вероятность суммы нескольких событий в случае их попарной несовместности. А как быть, если это условие не выполняется? Ответ на этот вопрос дадим для суммы трех событий.

По определению вероятности события (см. § 2.5) вероятность $P(A + B + C)$ равна сумме вероятностей элементарных исходов, входящих в $A + B + C$. В то же время $P(A) + P(B) + P(C)$ — это сумма вероятностей элементарных исходов, образующих событие A, и вероятностей элементарных исходов, образующих событие B, и вероятностей элементарных исходов, образующих событие C. При этом вероятности исходов, входящих в AB, AC, BC, суммируются дважды: исходы AB входят и в событие A, и в событие B; исходы AC — и в событие A, и в событие C; исходы BC — и в событие B, и в событие C. Вычтя $P(AB)$, $P(AC)$, $P(BC)$ из $P(A) + P(B) + P(C)$, получим, что оставшаяся разность не будет содержать вероятностей исходов, входящих одновременно и в A, и в B, и в C. Дело в том, что эти вероятности трижды входят в сумму $P(A) + P(B) + P(C)$ и входят в каждую из вероятностей $P(AB)$, $P(AC)$, $P(BC)$, т. е. $P(ABC)$ мы трижды сложили, а затем трижды вычли. Поэтому, прибавив $P(ABC)$ к выражению $P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC)$, получим сумму вероятностей исходов, образующих событие $A + B + C$.

Таким образом, доказано следующее утверждение.

ТЕОРЕМА 3. *Для любых трех событий A, B, C имеет место равенство*

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC),$$

т. е. вероятность суммы трех событий равна сумме вероятностей этих событий без суммы вероятностей их попарных произведений, сложенной с вероятностью произведения этих событий.

Пример 6. Игральный кубик бросают дважды. Какова вероятность того, что сумма выпавших очков кратна 2, 3 или 5?

❒ Обозначим через A, B, C соответственно события «сумма выпавших очков кратна 2», «сумма выпавших очков кратна 3», «сумма выпавших очков кратна 5». В § 2.4 и 2.5 строились ПЭИ для данного опыта и вводились элементарные вероятности. Если в качестве элементарных исходов опыта выбрать суммы выпавших очков, то множество элементарных исходов примет вид $U = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Элементарные вероятности для этого ПЭИ соответственно равны:

$$p(2) = \frac{1}{36}; \ \ p(3) = \frac{2}{36}; \ \ p(4) = \frac{3}{36}; \ \ p(5) = \frac{4}{36}; \ \ p(6) = \frac{5}{36}; \ \ p(7) = \frac{6}{36};$$

$$p(8) = \frac{5}{36}; \ \ p(9) = \frac{4}{36}; \ \ p(10) = \frac{3}{36}; \ \ p(11) = \frac{2}{36}; \ \ p(12) = \frac{1}{36}.$$

Отсюда

$$P(A) = p(2) + p(4) + p(6) + p(8) + p(10) + p(12) = \frac{1}{2};$$

$$P(B) = p(3) + p(6) + p(9) + p(12) = \frac{1}{3};$$

$$P(C) = p(5) + p(10) = \frac{7}{36}.$$

События AB, AC, BC состоят в том, что сумма выпавших очков кратна соответственно 6, 10 и 15.

$$P(AB) = p(6) + p(12) = \frac{6}{36}; \ \ P(AC) = p(10) = \frac{3}{36}; \ \ P(BC) = P(V) = 0.$$

Событие ABC состоит в том, что сумма выпавших очков кратна 30, $ABC = V$, $P(ABC) = P(V) = 0$. Поэтому

$$P(A + B + C) = \frac{1}{2} + \frac{1}{3} + \frac{7}{36} - \frac{6}{36} - \frac{3}{36} - 0 + 0 = \frac{7}{9}.$$

Впрочем, эту вероятность можно было найти, не прибегая к общей теореме сложения, а воспользовавшись определением вероятности события. Так как $A + B + C = \{2, 3, 4, 5, 6, 8, 9, 10, 12\}$, то

$$P(A + B + C) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{1}{36} = \frac{7}{9}.$$

Можно было перейти к событию, противоположному событию $A + B + C$:

$$P(A + B + C) = 1 - P(\overline{A + B + C}) = 1 - P(7, 11) = 1 - \left(\frac{6}{36} + \frac{2}{36}\right) = \frac{7}{9}. \ \blacksquare$$

Теорему 3 можно обобщить на сумму любого конечного числа событий. ◄

Фактически еще Д. Кардано (1501–1576) в XVI столетии, складывая шансы, пользовался теоремой сложения вероятностей. Он понимал, что при этом события должны быть несовместными. Понятно, что он не рассматривал вероятности, а подсчитывал ставки в «справедливых» играх, которые пропорциональны вероятностям. Теорему сложения вероятностей применял Х Гюйгенс (1629–1695) в своей работе «О подсчетах в азартных играх» (1657 г.) при решении известной задачи о распределении ставки. Я. Бернулли в (1654–1705) своей книге «Искусство предположений» (1713 г.) в оригинальной форме разъясняет применение теоремы сложения вероятностей, в частности невозможность ее применения для совместных событий. Английский ученый Т. Бейес (1702–1761) доказал теорему сложения вероятностей.

Контрольные вопросы

1. Может ли вероятность суммы двух событий быть:

а) меньше суммы вероятностей этих событий;

б) равной сумме вероятностей этих событий;

в) больше суммы вероятностей этих событий?

2. Может ли вероятность суммы двух событий равняться вероятности одного из слагаемых?

3. Как с помощью теоремы сложения вероятностей установить связь между вероятностями противоположных событий?

4. Чему равна вероятность суммы событий, если в каждом опыте наступает одно и только одно из этих событий?

5. Верно ли, что вероятность суммы трех событий не превосходит сумму вероятностей этих событий?

6. Известно, что во избежание больших затруднений на строительстве некоторого объекта необходимо, чтобы цемент был доставлен не позднее 27 июля, а финансирование организовано до 6 августа. На основе предыдущего опыта и анализа аналогичных ситуаций с применением субъективной оценки вероятности этим двум событиям приписаны соответственно вероятности 0,83 и 0,91. Предположим также, что вероятность выполнения хотя бы одного из этих сроков составляет 0,96.

а) Чему равна вероятность возникновения «больших затруднений»?

б) Являются ли данные события несовместными?

7. Найдите ошибку в следующих рассуждениях: в Москве и Санкт-Петербурге одновременно проходят соревнования. Вероятность того, что спортсмен выиграет соревнование в Москве, равна 0,9, а в Санкт-Петербурге — 0,6. Вероятность того, что спортсмен победит в Москве или в Санкт-Петербурге, по теореме сложения вероятностей равна $0{,}9 + 0{,}6 = 1{,}5$.

Задачи

142.° Проводится стрельба по мишени, изображенной на рис. 29. Вероятность попадания для некоторого стрелка в участок 1 составляет 0,35, в участок 2 — 0,21. Выстрел считается отличным, если пуля попала в участок 1, и хорошим, если пуля попала в участок 2. Какова вероятность того, что выстрел будет или хорошим, или отличным?

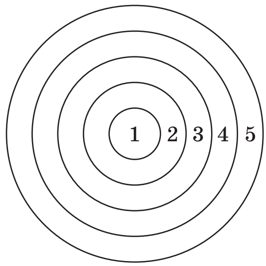


Рис. 29

143.° При многократном повторении двух выстрелов по мишени в среднем в 30% случаев было два, а в 50% — одно попадание. Найдите вероятность того, что имело место:

а) хотя бы одно попадание;

б) не более одного попадания.

144.° Продолжительными наблюдениями установлено, что в магазине из каждых 100 проданных пар мужской обуви в среднем 12 пар имеют 44-й размер, 5 пар — 45-й или больший размер. Какова вероятность того, что следующей будет продана пара мужской обуви, размер которой не меньше, чем 44-й?

145. Пассажир ждет автобус 32-го или 35-го маршрута на остановке, где проходят три автобусных маршрута: 32-й, 35-й, 37-й. Считая, что автобусы всех маршрутов появляются в среднем с одинаковой частотой, найдите вероятность того, что первый из автобусов, который подойдет к остановке, будет нужного маршрута.

146. Подбрасывают три монеты. Найдите вероятность выпадения не менее двух гербов.

147. Три события попарно несовместны, их объединение есть достоверное событие. Вероятности этих событий относятся как 3 : 2 : 1. Найдите вероятности каждого из этих событий.

148. Бросают два игральных кубика. Найдите вероятность того, что:

а) сумма выпавших очков равна 4 или 11;

б) наступит хотя бы одно из двух событий: «на первом кубике выпало не менее 3 очков», «на втором кубике — не менее 2 очков».

149. В урне десять красных и шесть синих шаров. Наудачу вынимают два шара. Какова вероятность того, что оба шара одного цвета?

150. Найдите вероятность того, что наудачу выбранное двузначное число окажется кратным либо 2, либо 5, либо тому и другому одновременно.

151. Проводится опрос, связанный с улучшением работы службы быта. Каждый должен ответить на два вопроса:

1) Удовлетворены ли вы качеством ремонта телевизора?

2) Удовлетворяет ли вас срок выполнения заказа на его ремонт?

Оказалось, что 30% опрошенных положительно ответили на оба вопроса; 20% — положительно на первый и отрицательно на второй вопрос; 10% — отрицательно на первый и положительно на второй вопрос; 40% — отрицательно на оба вопроса. Найдите вероятность того, что наугад выбранный человек:

а) удовлетворен качеством ремонта;

б) удовлетворен качеством ремонта или сроком его выполнения;

в) не удовлетворен сроком выполнения ремонта.

152. Вероятность получения крупного заказа некоторой фирмой равна 0,4. Вероятность финансовых потерь этой фирмы в текущем квартале составляет 0,5. Указанные события несовместны.

а)° Найдите вероятность получения заказа или финансовых потерь.

б) Исключена ли возможность того, что заказ не будет получен и не удастся заработать денег?

153. Продолжительными наблюдениями установлено, что из каждых 100 студентов высших учебных заведений приблизительно 60 изучают английский язык, 30 — французский, 10 — английский и французский. Найдите вероятность того, что наугад выбранный студент:

а) изучает по крайней мере один из двух указанных языков;

б) не изучает ни одного из этих языков.

§ 2.9. Условные вероятности

При проведении случайных опытов часто появляется дополнительная информация, которая существенно влияет на шансы наступления событий, связанных с этим опытом. Рассмотрим примеры, иллюстрирующие эти соображения.

Пример 1. Игральный кубик бросают дважды. Пусть событие A означает, что «сумма выпавших очков равна 11», а событие B — «при первом броске выпало 6 очков». Ясно, что шансы наступления события A изменяются в зависимости от того, произошло или не произошло событие B. ■

Пример 2. Если события A и B несовместны, то вероятность наступления события A, если известно, что событие B наступило, равна 0. ■

Пример 3. Если каждый исход из B входит и в A, то вероятность наступления A при условии, что B наступило, равна 1. ■

? Чему равно произведение событий A и B, если каждый исход из B входит и в A?

Пример 4. Предположим, что местная футбольная команда может выиграть важную игру с вероятностью 70%. Поступила новая информация, соответствующая событию «после окончания первого тайма команда выигрывает». В зависимости от того, реализуется ли это событие, вероятность победы изменяется. Вероятность победы команды при условии, что она действительно выигрывает после первого тайма, окажется выше и будет равна, например, 85%. Эта вероятность 85% представляет собой вероятность события «команда одержала победу» при условии наступления события «команда выигрывает после первого тайма». Вероятность выигрыша при условии, что команда проигрывает после первого тайма, будет меньше, чем вероятность победы, составляющая 70%; пусть, например, эта вероятность оценивается в 35%. Данная вероятность представляет собой вероятность события «команда одержала победу» при условии наступления события «команда проигрывает после первого тайма». ■

Пример 5. На успех нового коммерческого проекта влияет много факторов, таких, как благоприятные или неблагоприятные экономические условия, действия конкурентов. Экономический рост будет повышать шансы на успех; это означает, что вероятность успеха при условии экономического роста будет больше, чем вероятность успеха вообще (мы будем говорить: чем безусловная вероятность). ■

Примеры 1—5 позволяют сделать вывод, что целесообразно рассматривать вероятность одного события при условии, что другое произошло.

Пример 6. В опыте с бросанием игрального кубика дважды найдем вероятность того, что «при первом броске выпала 1» (событие A), если известно, что «сумма выпавших очков меньше 4» (событие B).

❐ ПЭИ рассматриваемого опыта представлено в таблице 2.22. Исходы этого опыта будем считать равновозможными. Другими словами, имеем классическую модель случайного опыта. По условию событие B считается наступившим. Оно состоит из трех равновозможных исходов $\{(1, 1), (1, 2), (2, 1)\}$, причем в двух из них можно наблюдать и событие $A = \{(1, 1), (1, 2)\}$. Таким образом, вероятность события A при условии, что B наступило, равна $\frac{2}{3}$. Это число будем называть ***условной вероятностью события A при условии B*** и обозначать $P(A \mid B)$. Так как $P(B) = \frac{3}{36}$, $P(AB) = \frac{2}{36}$, то

$$P(A \mid B) = \frac{2}{3} = \frac{\frac{2}{36}}{\frac{3}{36}} = \frac{P(AB)}{P(B)}. \ ■$$

Это равенство получено для опытов с равновозможными исходами.

Рассмотрим подобную задачу для статистически устойчивых опытов.

Пример 7. С первого станка в сборочный цех поступает 40%, а со второго — 60% всех деталей. 3% всех деталей, изготовленных на первом станке, и 5% всех деталей, изготовленных на втором, — бракованы. Вероятность детали быть бракованной зависит от того, на каком станке ее изготовили.

Поскольку существует вероятность события «деталь бракована», то опыты являются статистически устойчивыми и вероятность события можно считать равной относительной частоте этого события. Поэтому легко найти, что из каждой тысячи деталей, поступивших на сборку, в среднем 42 бракованы: $400 \cdot 0{,}03 + 600 \cdot 0{,}05 = 42$. Другими словами, вероятность того, что наугад взятая деталь будет бракованной (событие A), можно считать равной 0,042: $P(A) = 0{,}042$.

Вместе с тем, если деталь изготовлена на первом станке (событие B_1), то, как следует из условия, вероятность того, что она бракована, составляет 0,03. Эту вероятность будем обозначать $P(A \mid B_1)$ и говорить: *условная вероятность события A при условии, что событие B_1 произошло*: $P(A \mid B_1) = 0{,}03$.

Аналогично если через B_2 обозначить событие «деталь изготовлена на втором станке», то $P(A \mid B_2) = 0{,}05$.

Вычислим вероятность одновременного осуществления событий A и B_1, т. е. события AB_1. Из любой 1000 деталей, поступивших на сборку, в среднем 400 изготовлены на первом станке, из них примерно 3% (т. е. 12) бракованы. Поэтому

$$P(AB_1) = \frac{12}{1000} = \frac{400}{1000} \cdot \frac{12}{400} = P(B_1)P(A \mid B_1).$$

Отсюда

$$P(A \mid B_1) = \frac{P(AB_1)}{P(B_1)}.$$

Опять пришли к тому же равенству для условной вероятности. Примем его в качестве определения для произвольных опытов. ■

Условная вероятность *события A при условии B есть число, определяемое равенством*:

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

Конечно, здесь подразумевается, что $P(B) > 0$, т. е. событие B не является невозможным. В противном случае событие AB невозможно и $P(AB) = 0$.

?

В ящике 15 шаров, из них 10 белых. Из ящика наудачу последовательно один за другим, не возвращая обратно, извлекают два шара. Какова вероятность того, что второй шар белый, если первый белый?

Вероятность того, что деталь, изготовленная на станке А, является бракованной равна 0,01. Что такое 0,01: вероятность одновременного наступления событий «деталь изготовлена на станке А» и «деталь бракована» или условная вероятность того, что деталь бракована при условии, что она изготовлена на станке А»?

Введение условной вероятности $P(A \mid B)$ для каждого события B в ПЭИ U приводит к построению новой вероятностной модели с ПЭИ U_B, состоящем из тех и только тех элементов U, которые образуют событие B, и элементарными вероятностями $\frac{p(u_i)}{p(B)}$.

Пример 8. Вероятность того, что в неисправном приборе перегорел предохранитель, составляет 6%, вероятность обрыва провода равна 4%, а вероятность наличия обеих этих неисправностей равна 1%. Найти условную вероятность обрыва провода при условии того, что в приборе перегорел предохранитель.

❐ Обозначив через A и B соответственно события «обрыв провода» и «предохранитель перегорел», будем иметь

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{0{,}01}{0{,}06} \approx 0{,}167.$$

В этом случае перегорание предохранителя означает повышение вероятности того, что в неисправном приборе присутствует также и обрыв провода.

Такая условная вероятность свидетельствует о том, что из всех приборов, в которых сгорел предохранитель, $16{,}7\%$ обычно имеют еще и обрыв провода. Обратите внимание на то, насколько эта условная вероятность больше, чем безусловная вероятность обрыва провода (4%). Это связано с тем, что при рассмотрении приборов со сгоревшим предохранителем больше не идет речь обо «всех приборах», а только о тех, в которых сгорел предохранитель, т. е. о 6% всех приборов. Вероятность «обрыва провода» при этом возрастает с 4 до $16{,}7\%$, что и отражает учет дополнительной информации.

На рис. 30 схематично изображена рассматриваемая ситуация. ■

Вычислим условную вероятность $P(A \mid B)$ для событий, рассмотренных в примере 1.

$$B = \{(6,\ 1),\ (6,\ 2),\ (6,\ 3),\ (6,\ 4),\ (6,\ 5),\ (6,\ 6)\}; \quad P(B) = \frac{1}{6};$$

$$AB = \{(6{,}5)\}, \quad P(AB) = \frac{1}{36}; \quad P(A \mid B) = \frac{1}{36} : \frac{1}{6} = \frac{1}{6}.$$

$$P(\overline{B}) = 1 - \frac{1}{6} = \frac{5}{6}; \quad A\overline{B} = \{(5,\ 6)\};$$

$$P(A\overline{B}) = \frac{1}{36}; \quad P(A \mid \overline{B}) = \frac{1}{36} : \frac{5}{6} = \frac{1}{30}.$$

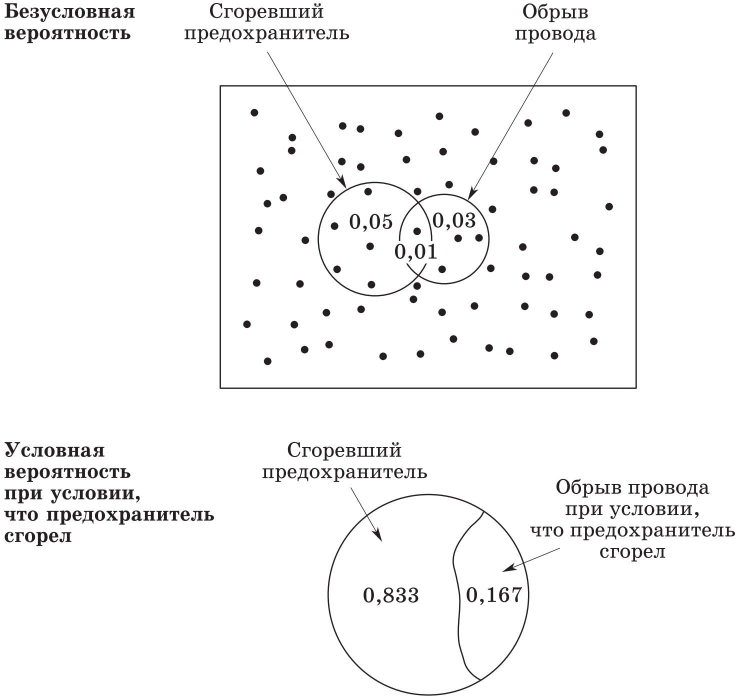


Рис. 30

Как видим, подтвердилась мысль о том, что вероятность наступления события A изменяется в зависимости от того, произошло или не произошло событие B.

Из определения условной вероятности следует, что если $P(B) > 0$, то

$$P(AB) = P(B)\,P(A \mid B).$$

Аналогично если $P(A) > 0$, то

$$P(AB) = P(A)\,P(B \mid A).$$

Таким образом, имеет место так называемая ***теорема умножения вероятностей***.

ТЕОРЕМА. *Если* $P(B) > 0$, $P(A) > 0$, *то*

$$P(AB) = P(A)\;P(B \mid A) = P(B)\;P(A \mid B).$$

Другими словами, *если вероятности каждого из двух событий отличны от нуля, то вероятность одновременного наступления* *двух событий равна произведению вероятности одного из этих событий и условной вероятности второго при условии, что первое наступило.*

Эту теорему используют для нахождения вероятности произведения двух событий.

Пример 9. Цех изготовляет кинескопы для телевизоров, причем 70% всех кинескопов предназначены для цветных телевизоров. Из общего количества кинескопов, предназначенных для цветных телевизоров, 40% отправляют на экспорт. Найти вероятность того, что наугад взятый для контроля кинескоп предназначен для цветного телевизора и будет отправлен на экспорт.

❐ Через A обозначим событие «наугад взятый кинескоп изготовлен для цветного телевизора», через B — «кинескоп будет отправлен на экспорт».

Известно, что $P(A) = 0{,}7$; $P(B \mid A) = 0{,}4$. Требуется найти $P(AB)$. Согласно теореме умножения вероятностей

$$P(AB) = P(A)\,P(B \mid A) = 0{,}7 \cdot 0{,}4 = 0{,}28. \blacksquare$$

Пример 10. В урне пять белых и три черных шара. Наудачу один за другим извлекают два шара, причем извлеченный шар в урну не возвращается. Какова вероятность, что:

а) оба шара белые;

б)* второй шар белый;

в)* первый шар белый, если второй оказался черным?

❐ а) Пусть событие A_i, $i = 1, 2$, означает, что «i-й извлеченный шар белый»; A — «оба шара белые»; $A = A_1A_2$.

Известно, что в урне восемь шаров, извлечение любого из них — равновозможно (выбор производится наугад). Событие A_1 происходит при любом из пяти исходов опыта (пять белых шаров), тогда $P(A_1) = \frac{5}{8}$. Если первым вынут белый шар, то после его извлечения в урне остается семь шаров, из которых четыре белых. Итак,

$$P(A_2 \mid A_1) = \frac{4}{7}; \quad P(A) = P(A_1)\,P(A_2 \mid A_1) = \frac{5}{8} \cdot \frac{4}{7} = \frac{5}{14}.$$

б) Событие A_2 — «второй извлеченный шар белый» — наступает в двух случаях: если или оба извлеченных шара белые, или первый черный, а второй белый, т. е.

$$A_2 = A_1A_2 + \bar{A}_1 A_2.$$

Применяя последовательно теоремы сложения и умножения вероятностей, получим

$$P(A_2) = P(A_1A_2) + P(\bar{A}_1 A_2) = P(A_1)\, P(A_2 \mid A_1) + P(\bar{A}_1)\, P(A_2 \mid \bar{A}_1).$$

? Почему вероятность события $A_2 = A_1A_2 + \bar{A}_1 A_2$ равна сумме вероятностей событий A_1A_2 и $\bar{A}_1 A_2$?

Ранее мы установили, что $P(A_1) = \frac{5}{8}$, а $P(A_2 \mid A_1) = \frac{4}{7}$. Аналогично находим, что $P(\bar{A}_1) = \frac{3}{8}$, $P(A_2 \mid \bar{A}_1) = \frac{5}{7}$. Отсюда $P(A_2) = \frac{5}{8} \cdot \frac{4}{7} + \frac{3}{8} \cdot \frac{5}{7} = \frac{5}{8}$. Заметим, что

$$P(A_1) = P(A_2) = \frac{5}{8}.$$

? Как найдено, что $P(A_2 \mid \bar{A}_1) = \frac{5}{7}$?

в) Надо найти $P(A_1 \mid \bar{A}_2)$. В соответствии с определением условной вероятности и теоремой умножения вероятностей,

$$P(A_1 \mid \bar{A}_2) = \frac{P(A_1\bar{A}_2)}{P(\bar{A}_2)} = \frac{P(A_1)\, P(\bar{A}_2 \mid A_1)}{1 - P(A_2)} = \frac{\frac{5}{8} \cdot \frac{3}{7}}{1 - \frac{5}{8}} = \frac{5}{7}. \blacksquare$$

При решении задач нередко приходится вычислять вероятности событий, являющихся пересечением более чем двух событий. Обобщим теорему умножения вероятностей на произведение любого конечного числа событий. Вначале рассмотрим вероятность произведения трех событий $P(ABC)$. Обозначив $AB = D$ и применив теорему умножения для двух событий дважды, получим

$$P(ABC) = P(DC) = P(D)\, P(C \mid D) = P(AB)\, P(C \mid AB) =$$
$$= P(A)\, P(B \mid A)\, P(C \mid AB),$$

т. е.

$$P(ABC) = P(A)\, P(B \mid A)\, P(C \mid AB).$$

Аналогично для n событий имеет место формула

$$P(A_1A_2A_3 \dots A_n) = P(A_1)\, P(A_2\,|\,A_1)\, P(A_3\,|\,A_1A_2) \dots P(A_n\,|\,A_1A_2 \dots A_{n-1}).$$

Пример 11. Из полной колоды в 36 карт вынимают наугад последовательно без возвращения одну за другой три карты. Найдите вероятность того, что среди них не будет ни одного туза.

❐ Обозначим через A_i событие «i-я вынутая карта не туз» ($i = 1, 2, 3$). Тогда искомое событие B — «среди трех вынутых карт нет туза» можно представить в виде $B = A_1A_2A_3$. Применяя теорему умножения для трех событий, получим

$$P(B) = P(A_1A_2A_3) = P(A_1)\, P(A_2\,|\,A_1)\, P(A_3\,|\,A_1A_2).$$

Так как всех карт 36, а тузов из них четыре, то при извлечении первой карты для A_1 имеем классическую модель с $N = 36$, $N(A_1) = 32$, поэтому $P(A_1) = \frac{32}{36}$. При условии, что первая извлеченная карта не туз для события A_2 снова имеем классическую модель с $N = 35$, $N(A_2) = 31$, поэтому $P(A_2\,|\,A_1) = \frac{31}{35}$. Аналогично $P(A_3\,|\,A_1A_2) = \frac{30}{34}$, и следовательно,

$$P(A_1A_2A_3) = \frac{32}{36} \cdot \frac{31}{35} \cdot \frac{30}{34} = 0{,}695. \blacksquare$$

Контрольные вопросы

1. Почему $P(A \mid B) < 1$?

2. Пусть A и B — несовместные события и $P(B) \neq 0$. Найдите $P(A \mid B)$.

3. Пусть каждый элементарный исход события B входит также и в событие A. Чему равна: а) $P(A \mid B)$; б) $P(B \mid A)$?

4. Известно, что события A и B несовместны. Верно ли, что $P(A_1 + A_2|B) = P(A_1|B) + P(A_2|B)$?

5. Чему равна: а) $P(A \mid U)$; б) $P(U \mid A)$; в) $P(V \mid A)$ для произвольного события A?

6. Имеет ли смысл $P(A \mid V)$?

7. Верно ли, что $P(A \mid B) \geqslant P(A)$?

8. Вам сообщили хорошие новости: опытный образец нового товара выпущен с опережением графика, и его функциональные качества выше, чем ожидалось. Следует ли ожидать, что «условная вероятность того, что этот товар будет иметь успех в случае хороших новостей», окажется больше, меньше или равной безусловной вероятности успеха?

Задачи

154.° В ящике десять деталей, среди которых семь окрашенных. Сборщик наудачу достает две детали. Найдите вероятность того, что вторая деталь окрашена, если первая:

а) окрашена; б) не окрашена.

155. Из цифр 1, 2, 3, 4, 5 выбирают одну, а из оставшихся — еще одну. Найдите вероятность того, что во второй раз будет выбрана нечетная цифра, если в первый раз была выбрана:

а) четная цифра; б) нечетная цифра.

156. Урна содержит пять черных и десять красных мячей. Из урны (без возвращения) вынимают один за другим два мяча.

а) Найдите вероятность того, что оба мяча черные.

б) Постройте ПЭИ этого опыта. Определите вероятности его элементов.

157.° Для условия задачи 154 найдите вероятность того, что обе детали:

а) окрашены; б) не окрашены.

158.° Для условия задачи 155 найдите вероятность того, что оба раза выбраны

а) нечетные цифры; б) четные цифры.

159.° На некотором предприятии 96% изделий считаются пригодными. Из пригодных изделий в среднем 75% составляют изделия первого сорта, остальные — второго сорта. Найдите вероятность того, что изделие, изготовленное на этом предприятии, окажется второсортным.

160. В ящичке десять красных и шесть синих пуговиц. Вынимают наудачу две пуговицы. Какова вероятность того, что обе пуговицы будут:

а) красными; б) синими?

161. Вероятность для деталей некоторого производства удовлетворять стандарту равна 0,96. Предлагается упрощенная система испытаний, которая дает положительный результат с вероятностью 0,98 для деталей, удовлетворяющих стандарту, и с вероятностью 0,05 — для тех, которые не удовлетворяют стандарту. Найдите вероятность того, что наугад взятая деталь:

а) удовлетворяет стандарту и выдерживает испытание;

б) удовлетворяет стандарту и не выдерживает испытания;

в) не удовлетворяет стандарту, но выдерживает испытание;

г) не удовлетворяет стандарту и не выдерживает испытания.

162. По данным переписи населения в Англии и Уэльсе (1891 г.) установлено, что темноглазые родители и темноглазые дети составляют 5% от общего количества обследованных лиц; темноглазые родители и светлоглазые дети — 7,9%; светлоглазые родители и темноглазые дети — 8,9%; светлоглазые родители и светлоглазые дети — 78,2%. Найдите связь между цветом глаз родителей и детей.

163. В техникуме обучается 500 студентов, причем известно, что 300 студентов изучают английский язык, 200 — французский, 50 — немецкий, 20 — английский и немецкий языки, 30 — французский и немецкий, 20 — английский и французский, 10 человек изучают все три языка. Какова вероятность того, что наугад выбранный студент изучает французский язык, если:

а)° он изучает английский язык;

б) он изучает английский и немецкий языки;

в)* он не изучает ни английского, ни немецкого языка?

164. Число грузовых автомобилей, проезжающих вдоль шоссе, на котором находится бензоколонка, относится к числу легковых автомобилей, проезжающих вдоль того же шоссе, как 3 : 2. Известно, что в среднем один из 30 грузовых и два из 50 легковых автомобилей подъезжают к бензоколонке для заправки. Чему равна вероятность того, что:

а) к бензоколонке подъедет грузовой автомобиль, и он будет заправляться;

б) к бензоколонке подъедет легковой автомобиль, и он будет заправляться;

в) автомобиль, который подъедет к бензоколонке, будет заправляться?

165. В ящике имеется десять деталей, среди которых семь окрашенных. Рабочий наугад по одной вынимает две детали. Найдите вероятность того, что:

а)° вторая деталь окрашена, если первая окрашена;

б)° вторая деталь окрашена, если первая не окрашена;

в) обе детали не окрашены;

г) обе детали окрашены;

д)* вторая деталь окрашена;

е)* первая деталь окрашена, если вторая оказалась окрашенной.

166. Следующий год для фирмы ожидается удачным с вероятностью 0,7. При условии, что год удачный, с вероятностью 0,9 ожидается выплата дивидендов. Однако, если год окажется неудачным, выплата дивидендов произойдет с вероятностью 0,2.

а) Найдите вероятность того, что год удачный и дивиденды выплачиваются.

б) Найдите вероятность того, что дивиденды выплачиваются.

в) Найдите условную вероятность того, что год удачный при условии, что дивиденды выплачиваются.

167.* В ящичке десять красных и шесть синих пуговиц. Вынимают наудачу последовательно без возвращения одну за другой три пуговицы. Какова вероятность того, что все пуговицы будут:

а) красными; б) синими?

§ 2.10. Независимые события

При исследовании случайных событий иногда важно установить зависимость между некоторыми из них. Понятие независимости является очень важным в теории вероятностей. Оно довольно точно отображает понятие независимости явлений в обычном понимании, т. е. в понимании отсутствия влияния одних событий на другие. В повседневной речи часто говорят о том, что какие-то два события «не имеют отношения друг к другу». Фактически речь идет о понятии, которое на математическом языке называют независимостью событий. Вместе с тем обычное понятие независимости и понятие вероятностной независимости имеют отличия.

Пример 1. Два исследователя проводят измерения некоторой физической величины. Вероятности сделать ошибку при снятии показателей прибора для них соответственно составляют 0,1 и 0,15. Какова вероятность того, что при одноразовом измерении оба исследователя допустят ошибку?

❒ Обозначим через A_i, $i = 1, 2$, событие, которое состоит в том, что i-й исследователь сделает ошибку. Нужно найти $P(A_1A_2)$. По теореме умножения вероятностей

$$P(A_1A_2) = P(A_1)\ P(A_2 \mid A_1).$$

По условию $P(A_1) = 0{,}1$. А что такое $P(A_2 \mid A_1)$? Это условная вероятность того, что второй исследователь ошибется при снятии показателей прибора, при условии, что первый ошибся. Но мы вправе считать, что вероятность события A_2 не зависит от того, произошло событие A_1 или нет. Ведь второй исследователь мог снимать данные прибора, не зная результата работы первого. Итак, $P(A_2 \mid A_1) = P(A_2) = 0{,}15$. Поэтому

$$P(A_1A_2) = P(A_1)\ P(A_2) = 0{,}1 \cdot 0{,}15 = 0{,}015. \ ■$$

В рассмотренном примере условная вероятность $P(A_2 \mid A_1)$ равна безусловной вероятности $P(A_2)$. В этом случае говорят, что *событие* A_2 *не зависит от события* A_1.

Пример 2. Обратимся снова к опыту с бросанием игрального кубика дважды (см. табл. 2.22 на с. 198).

Рассмотрим следующие события, связанные с этим опытом:

A — «при первом броске выпало менее 3 очков»;

B — «при втором броске выпало более 3 очков».

Ясно, что $P(A) = \frac{12}{36} = \frac{1}{3}$ (исходы, образующие событие A, стоят в первых двух строках таблицы); $P(B) = \frac{18}{36} = \frac{1}{2}$ (исходы, образующие событие B, стоят в последних трех столбцах таблицы); $P(AB) = \frac{6}{36} = \frac{1}{6}$ (исходы, образующие событие AB, стоят на пересечении первых двух строк и последних трех столбцов таблицы), т. е. $P(AB) = \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{2} = P(A)\ P(B)$.

По теореме умножения $P(AB) = P(A)\ P(B \mid A)$. Сравнивая два последних равенства, получим $P(B \mid A) = P(B)$. ■

Этот результат можно объяснить на интуитивном уровне. Результат первого броска никак не влияет на результат второго броска, значит, вероятность того, что при втором броске выпало более 3 очков, не зависит от того, выпало ли при первом броске менее 3 очков или нет. Как и в предыдущем примере, говорят, что событие B не зависит от события A.

Однако не всегда условная вероятность наступления одного события при условии, что другое наступило, совпадает с безусловной вероятностью наступления первого. Например, в том же опыте с бросанием игрального кубика дважды рассмотрим то же событие A и событие C — «сумма выпавших очков равна 5»:

$$P(A) = \frac{1}{3}; \quad P(C) = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \frac{4}{36} = \frac{1}{9};$$

$$P(C \mid A) = P\{(1, 4), (2, 3)\} = \frac{2}{36} = \frac{1}{18} \neq P(C).$$

Интуитивно ясно, что сумма очков зависит от числа очков, выпавших при первом броске.

Рассмотренные примеры показывают, что естественно считать одно событие не зависящим от другого, если условная вероятность наступления первого при условии, что второе произошло, совпадает с безусловной вероятностью наступления первого.

Будем говорить, что событие A не зависит от события B, если $P(A \mid B) = P(A)$.

Покажем, что *если событие A не зависит от B и $P(A) \neq 0$, то и событие B не зависит от A.*

❐ В самом деле, в соответствии с теоремой умножения вероятностей

$$P(A \mid B)\, P(B) = P(B \mid A)\, P(A).$$

По условию $P(A \mid B) = P(A)$ и $P(A) \neq 0$. Отсюда

$$P(B \mid A) = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

А это означает, что событие B не зависит от события A. ■

Если событие A не зависит от события B, а событие B не зависит от события A, то события A и B называют ***независимыми****.*

ТЕОРЕМА 1. *События A и B независимы тогда и только тогда, когда имеет место равенство*

$$P(AB) = P(A)\, P(B).$$

Если $P(A) \neq 0$, $P(B) \neq 0$, то это утверждение вытекает из предшествующих соображений. Это равенство остается справедливым и в том случае, если $P(A) = 0$ или $P(B) = 0$.

? Проведите самостоятельно полное доказательство.

Утверждение, содержащееся в теореме 1, может быть принято в качестве определения независимых событий.

Пример 3. Симметричную монету подбрасывают трижды. Показать, что события A — «при первом подбрасывании выпал герб» и B — «при последних двух подбрасываниях выпала цифра» независимы.

❐ ПЭИ рассматриваемого опыта, события A и B имеют вид:

$$U = \{\text{ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ}\},$$

$$A = \{\text{ГГГ, ГГЦ, ГЦГ, ГЦЦ}\}, \quad B = \{\text{ГЦЦ, ЦЦЦ}\}, \quad AB = \{\text{ГЦЦ}\};$$

$$P(A) = \frac{4}{8} = \frac{1}{2}; \quad P(B) = \frac{2}{8} = \frac{1}{4}; \quad P(AB) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = P(A)\, P(B).$$

События независимы. ■

Пример 4. В опыте со статистикой жилищных условий (см. пример 3 из § 2.4) выяснить, зависимы ли события A — «опрошенный удовлетворен качеством жилья» и B — «опрошенный удовлетворен удаленностью квартиры от места работы».

❒ Напомним, что ПЭИ опыта имеет вид

$$U = \{11,\ 10,\ 01,\ 00\}, \text{ причем } p_1 = 0{,}5;\ p_2 = 0{,}25;\ p_3 = 0{,}15;\ p_4 = 0{,}1;$$
$$P(A) = P\{11,\ 10\} = p_1 + p_2 = 0{,}75;$$
$$P(B) = P\{11,\ 01\} = p_1 + p_3 = 0{,}65;$$
$$P(AB) = P\{11\} = p_1 = 0{,}5 \neq 0{,}75 \cdot 0{,}65 = P(A) \cdot P(B).$$

События зависимы. ■

?

Вероятность получения крупного заказа некоторой фирмой равна 0,4. Вероятность финансовых потерь этой фирмы в текущем квартале составляет 0,5, вероятность получения заказа и финансовых потерь составляет 0,1. Являются ли события «получение заказа» и «финансовые потери» независимыми?

Вероятность получения крупного заказа некоторой фирмой равна 0,4. Вероятность финансовых потерь этой фирмы в текущем квартале составляет 0,5. Если эти события независимы, найдите вероятность получения заказа и финансовых потерь.

➤

С помощью определения установить независимость событий можно далеко не всегда, особенно если речь идет о построении вероятностной модели какой-либо реальной ситуации. Существует способ, основанный на гипотезе о физической независимости событий. Если какие-либо события описывают пренебрежимо мало связанные процессы, то их считают физически независимыми, а из физической независимости следует независимость в вероятностном смысле. Это утверждение является эмпирическим фактом, но многовековой опыт человечества подтверждает возможность такой точки зрения и ее практическую пользу. Так, например, отсутствие ошибок при программировании не зависит от отсутствия сбоев ЭВМ при решении задач: эти события — результат действия несвязанных физических процессов. ➤

Рассмотрим важное свойство независимых событий.

ТЕОРЕМА 2. *Если события A и B независимы, то независимыми будут и события A и $\overline{B}$.*

❒ Поскольку событие A происходит тогда и только тогда, когда происходят события A и B или события A и $\overline{B}$, то

$$A = AB + A\overline{B}, \quad P(A) = P(AB) + P(A\overline{B}).$$

Отсюда

$$P(A\overline{B}) = P(A) - P(AB) = P(A) - P(A)\,P(B) =$$
$$= P(A)(1 - P(B)) = P(A)\,P(\overline{B}).$$

А это и означает, что события A и $\overline{B}$ независимы. ■

Это свойство дает возможность заменять любое событие в обеих частях равенства $P(AB) = P(A)\,P(B)$ событием, ему противоположным. Доказанное утверждение означает, что независимость событий A и B эквивалентна независимости событий A и $\overline{B}$, а в связи с симметрией также и событий $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.

?

Правильную монету подбрасывают до первого появления герба. Какова вероятность того, что потребуется два подбрасывания?

Вероятность того, что в течение одной смены возникнет неполадка станка, равна 0,05. Какова вероятность того, что не произойдет ни одной неполадки за две смены?

Обобщим понятие независимости на любое конечное число событий.

Случайные события A_1, A_2, A_3 независимы в совокупности, если они попарно независимы и

$$P(A_1A_2A_3) = P(A_1)\,P(A_2)\,P(A_3).$$

Случайные события A_1, A_2, A_3, A_4 независимы в совокупности, если независимы в совокупности любые три из них и

$$P(A_1A_2A_3A_4) = P(A_1)\,P(A_2)\,P(A_3)\,P(A_4).$$

Аналогично определяют независимость в совокупности событий A_1, A_2, ..., A_n.

Можно доказать, что если мы имеем некоторую совокупность независимых в совокупности событий, то, заменив некоторые из них противоположными им событиями, снова получим независимые в совокупности события.

Например, покажем, что если события A_1, A_2, A_3 независимы в совокупности, то и события A_1, A_2, $\overline{A}_3$ независимы в совокупности.

❐ Из теоремы 2 следует, что события A_1, A_2, $\overline{A}_3$ попарно независимы. Аналогично доказательству теоремы 2 имеем:

$$A_1 A_2 = A_1A_2A_3 + A_1A_2\overline{A}_3; \quad P(A_1A_2) = P(A_1A_2A_3) + P(A_1A_2\overline{A}_3);$$

$$P(A_1A_2\overline{A}_3) = P(A_1A_2) - P(A_1A_2A_3) = P(A_1A_2) - P(A_1A_2)\, P(A_3) =$$

$$= P(A_1A_2)\,(1 - P(A_3)) = P(A_1A_2)\, P(\overline{A}_3) = P(A_1)\, P(A_2)\, P(\overline{A}_3). \blacksquare$$

? * Докажите, что если события A_1, A_2, A_3 независимы в совокупности, то и события A_1, $\overline{A}_2$, $\overline{A}_3$ независимы в совокупности.

Пример 5. Прибор, работающий в течение суток, состоит из трех узлов, каждый из которых независимо друг от друга может за это время выйти из строя. Неисправность хотя бы одного узла приводит к отказу прибора. Вероятность безотказной работы в течение суток первого узла равна 0,9, второго — 0,95, третьего — 0,85. Чему равна вероятность того, что:

а) в течение суток прибор будет работать безотказно;

б) на протяжении суток будет работать по крайней мере один узел?

❐ а) Пусть событие A_i, $i = 1, 2, 3$, означает, что «i-й узел исправен»; событие A — «прибор на протяжении суток работает безотказно». Поскольку прибор работает безотказно тогда и только тогда, когда исправны все три узла, то $A = A_1A_2A_3$. Так как события A_1, A_2, A_3 независимы в совокупности, то

$$P(A) = P(A_1A_2A_3) = P(A_1)\, P(A_2)\, P(A_3) = 0{,}9 \cdot 0{,}95 \cdot 0{,}85 \approx 0{,}73.$$

б) Пусть событие B означает, что на протяжении суток будет работать по крайней мере один узел. Это событие противоположно событию «на протяжении суток откажут все узлы», т. е. событию $\overline{A}_1\overline{A}_2\overline{A}_3$. Согласно сделанному замечанию, поскольку события A_1, A_2, A_3 независимы в совокупности, то события $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$ также независимы в совокупности. Поэтому

$$P(B) = 1 - P(\overline{A}_1\,\overline{A}_2\,\overline{A}_3) = 1 - P(\overline{A}_1)\,P(\overline{A}_2)\,P(\overline{A}_3) =$$
$$= 1 - 0{,}1 \cdot 0{,}05 \cdot 0{,}15 = 0{,}99925. \blacksquare$$

? Правильную монету подбрасывают до первого появления герба. Какова вероятность того, что потребуется три подбрасывания?

Вероятность того, что в течение одной смены возникнет неполадка станка, равна 0,05. Какова вероятность того, что не произойдет ни одной неполадки за три смены?

Из независимости событий в совокупности вытекает их попарная независимость. Обратное утверждение неверно даже для трех событий, в чем можно убедиться на следующем примере.

Пример 6. Рассмотрим опыт с подбрасыванием двух монет. ПЭИ этого опыта имеет вид {ГГ, ГЦ, ЦГ, ЦЦ}, исходы опыта будем считать равновозможными, т. е. P(ГГ) = P(ГЦ) = P(ЦГ) = P(ЦЦ) = 0,25. Введем следующие обозначения: A_1 — «на первой монете выпал герб»; A_2 — «на второй монете выпал герб»; A_3 — «монеты упали одинаково». Являются ли события A_1, A_2, A_3:

а) попарно независимыми; б) независимыми в совокупности?

❐ Поскольку A_1 = {ГГ, ГЦ}, то $P(A_1)$ = P(ГГ, ГЦ) = P(ГГ) + P(ГЦ) = = 0,25 + 0,25 = 0,5.

Аналогично $P(A_2)$ = P(ГГ) + P(ЦГ) = 0,5; $P(A_3)$ = P(ГГ) + + P(ЦЦ) = 0,5.

Очевидно, что A_1A_2 = ГГ. Поэтому $P(A_1A_2)$ = P(ГГ) = 0,25 = = 0,5 · 0,5 = $P(A_1)P(A_2)$.

Аналогично $P(A_1A_3)$ = P(ГГ) = 0,25 = 0,5 · 0,5 = $P(A_1)\,P(A_3)$.

$$P(A_2A_3) = P(\text{ГГ}) = 0{,}25 = P(A_2)P(A_3).$$

События A_1, A_2, A_3 попарно независимы. Проверим их независимость в совокупности: $P(A_1A_2A_3)$ = P(ГГ) = 0,25 $\neq$ 0,5 · 0,5 · 0,5, т. е. эти события не являются независимыми в совокупности. ■

Теоремой умножения для независимых событий фактически пользовался еще Д. Кардано (1501–1576). Но, как и для теоремы сложения, он не рассматривал вероятности, а подсчитывал ставки в «справедливых» играх, которые пропорциональны вероятностям. В работах Т. Бейеса (1702–1761) содержится доказательство теоремы умножения для зависимых событий и теоремы умножения для нескольких независимых событий.

Контрольные вопросы

1. Будут ли независимыми произвольное событие A и:

а) достоверное событие U;

б) невозможное событие V;

в) противоположное ему событие $\overline{A}$?

2. Будут ли несовместными независимые события с положительными вероятностями?

3. Верны ли следующие утверждения:

а) если события A и B независимы, то они и несовместны;

б) если события A и B несовместны, то они и независимы:

в) если события A и B независимы, то они совместны?

4. Следует ли из независимости событий в совокупности их попарная независимость? А наоборот?

5. Если события A и B независимы, то, что можно сказать о независимости событий:

а) A и $\overline{B}$; б) $\overline{A}$ и B; в) $\overline{A}$ и $\overline{B}$; г) $A + B$ и B;

д) AB и A?

6. Чему равна вероятность суммы двух независимых событий?

7. Верно ли, что вероятность произведения трех попарно независимых событий равна произведению вероятностей этих событий?

8. Известно, что во избежание больших затруднений на строительстве некоторого объекта необходимо, чтобы цемент был доставлен не позднее 27 июля, а финансирование организовано до 6 августа. На основе предыдущего опыта и анализа аналогичных ситуаций с применением субъективной оценки вероятности этим двум событиям приписаны соответственно вероятности 0,83 и 0,91. Предположим также, что вероятность выполнения хотя бы одного из этих сроков составляет 0,96. Являются ли данные события независимыми?

Задачи

168.° В урне лежат пять черных, четыре красных и три белых шара. Последовательно вынимают три шара, причем каждый шар возвращают в урну перед тем, как вынимать следующий. Найдите вероятность того, что первый шар окажется черным, второй — красным и третий — белым.

169. Вероятность безотказной работы прибора равна 0,9. При выходе его из строя происходит мгновенное переключение на дублирующий прибор, причем вероятность его безотказной работы — 0,9, а переключающего устройства — 1. Найдите вероятность безотказной работы системы, состоящей из:

а) двух приборов; б) трех приборов; в)* n приборов.

170. Решите предыдущую задачу в случае, если вероятность безотказной работы переключающего устройства равна 0,8.

171. Бросают два правильных игральных кубика. Пусть событие A_1 означает, что «на первом кубике выпало четное число очков», A_2 — «на втором кубике выпало нечетное число очков», A_3 — «кубики упали одинаково». Будут ли события A_1, A_2, A_3:

а) попарно независимыми; б) независимыми в совокупности?

172. Для условий задачи 126 о статистике зрелищных мероприятий выясните, зависимы ли события A — «опрошенный регулярно посещает кинотеатр», B — «опрошенный регулярно смотрит телевизор».

173. На мебельной фабрике на отдельных участках производятся спинки, сиденья и ножки стульев. Дефектными оказываются 1% спинок, 2% сидений и 0,5% ножек. Произведенные детали случайно комбинируются на участке, где собирают стулья. Какой примерно процент стульев:

а) не будет испорчен; б) имеет дефектные спинки и сиденья?

174. Пусть $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ — ПЭИ некоторого опыта. Пусть $p_1 = \frac{1}{8}$; $p_2 = \frac{5}{16}$; $p_3 = \frac{1}{16}$; $p_4 = \frac{3}{8}$; $p_5 = p_6 = \frac{1}{16}$, где $p_i = P(u_i)$, $i = 1, 2, \ldots, 6$. Пусть $A = \{u_1, u_4\}$, $B = \{u_1, u_2, u_5\}$, $C = \{u_1, u_2, u_3\}$. Будут ли события A, B, C:

а) попарно независимы; б) независимы в совокупности?

175. Вероятность попадания в мишень для первого стрелка равна 0,8, а для второго — 0,6. Стрелки независимо друг от друга сделали по одному выстрелу. Какова вероятность того, что а)° в мишень попадает только первый стрелок;

б)° оба стрелка попадут в мишень;

в)° ни один из стрелков не попадет в мишень;

г) только один стрелок попадет в мишень;

д) хотя бы один из стрелков попадет в мишень?

176. Устройство состоит из двух элементов, работающих независимо друг от друга. Вероятности отказа элементов соответственно равны 0,05 и 0,08. Найдите вероятность того, что:

а)° будет работать только второй элемент;

б)° откажут оба элемента;

в)° будут работать оба элемента;

г) будет работать только один элемент;

д) откажет устройство, если для этого достаточно, чтобы отказал хотя бы один элемент.

177. На исследовательском стенде проходят испытание 180 приборов. Вероятность того, что в течение часа откажет любой из этих приборов, равна 0,05. Найдите вероятность того, что в течение часа:

а) ни один из приборов не откажет;

б) откажет хотя бы один из приборов.

§ 2.11. Формула полной вероятности

Рассмотрим формулу, которая позволит по вероятностям попарно несовместных событий H_1, ..., H_n, охватывающих всевозможные случаи, найти вероятность события A, если известны условные вероятности $P(A \mid H_i)$, $i = 1, \ldots, n$.

Пример 1. С первого станка на сборку поступает 40% , со второго — 30% и с третьего — 30% всех деталей. Вероятности изготовления бракованной детали равны для каждого станка соответственно 0,01; 0,03; 0,05. Найдите вероятность того, что наудачу взятая деталь, поступившая на сборку, бракованная.

❒ Обозначим через A событие, заключающееся в том, что деталь бракованная, H_i — деталь изготовлена на i-м станке, $i = 1, 2, 3$. Из условия следует:

$$P(H_1) = 0{,}4; \quad P(H_2) = 0{,}3; \quad P(H_3) = 0{,}3;$$
$$P(A \mid H_1) = 0{,}01; \quad P(A \mid H_2) = 0{,}03; \quad P(A \mid H_3) = 0{,}05.$$

Деталь может быть изготовлена или на первом станке, или на втором, или на третьем станке. Таким образом, $H_1 + H_2 + H_3 = U$

и события H_1, H_2, H_3 попарно несовместны. Рис. 31 служит иллюстрацией этой ситуации. Событие A наступает тогда и только тогда, когда или деталь изготовлена на первом станке и она бракованная, или деталь изготовлена на втором станке и она бракованная, или на третьем станке и бракованная, т. е.

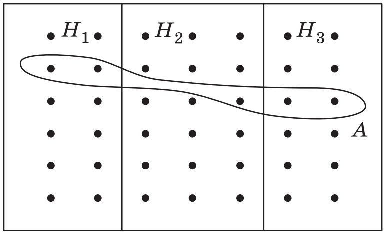


Рис. 31

$$A = H_1A + H_2A + H_3A.$$

Так как события H_i, $i = 1, 2, 3$, попарно несовместны, то несовместны попарно и события H_1A, H_2A, H_3A. Рис. 31 иллюстрирует это утверждение. По теореме сложения вероятностей для попарно несовместных событий

$$P(A) = P(H_1A) + P(H_2A) + P(H_3A).$$

Используя теорему умножения вероятностей, получим

$$P(A) = P(H_1)\,P(A \mid H_1) + P(H_2)\,P(A \mid H_2) + P(H_3)\,P(A \mid H_3) =$$
$$= 0{,}4 \cdot 0{,}01 + 0{,}3 \cdot 0{,}03 + 0{,}3 \cdot 0{,}05 = 0{,}028. \blacksquare$$

Рассуждения, проведенные при решении примера 1, могут быть использованы и в общей ситуации. Предварительно введем новое понятие, которое фактически рассматривалось при решении примера 1.

Пусть U — произвольное ПЭИ и события $H_1, H_2, \ldots, H_n$ такие, что:

1) $H_1 + H_2 + \ldots + H_n = U$, т. е. по крайней мере одно из событий H_i, $i = 1, 2, \ldots, n$, обязательно наступает;

2) $H_iH_k = V$, $i \neq k$, т. е. события H_i попарно несовместны.

Такую систему событий $H_1, H_2, \ldots, H_n$ называют ***полной группой несовместных событий***. События $H_1, H_2, \ldots, H_n$ называют иногда *гипотезами*. Произвольное событие A и ему противоположное событие $\overline{A}$ образуют полную группу несовместных событий.

В примере 1 события H_1, H_2, H_3 образовывали полную группу попарно несовместных событий.

Пример 2 (задача о телевизионном шоу). Вы участвуете в игровом телевизионном шоу. Приз спрятан за одной из трех дверей. Вы выбираете одну. Прежде чем открыть выбранную дверь, открывают другую, за которой приза нет. После этого вам разрешается изменить свой выбор. Закрытыми остались две двери: названная вами и еще одна. Измените ли вы свой выбор?

В описанной ситуации события «дверь, за которой скрывается приз, угадана при первом выборе» и «дверь, за которой скрывается приз, не угадана при первом выборе» образуют полную группу несовместных событий. ■

Какие из перечисленных ниже систем событий, связанных с указанным опытом, образуют полную группу несовместных событий:

а) наугад выбирают студента дневного отделения математического факультета университета; события A_i — «выбранный студент учится на i-м курсе», $i = 1, 2, 3, 4, 5$;

б) учащийся выполняет тестовое задание, в котором предлагается выбрать один из пяти указанных ответов; события A_i — «учащийся выбрал i-й ответ», $i = 1, 2, 3, 4, 5$;

в) бросают игральный кубик; события A_1 — «число выпавших очков меньше 2», A_2 — «число выпавших очков простое», A_3 — «число выпавших очков составное»;

г) бросают игральный кубик; события A_1 — «число выпавших очков меньше 3», A_2 — «число выпавших очков больше 3».

Проводя рассуждения, аналогичные решению примера 1, получим

$$P(A) = P(H_1)\, P(A \mid H_1) + P(H_2)\, P(A \mid H_2) + \ldots$$
$$\ldots + P(H_n)\, P(A \mid H_n).$$

Это равенство называют ***формулой полной вероятности***. Приведем ее доказательство.

❒ Имеем

$$A = AU = A(H_1 + H_2 + \ldots + H_n)\ .$$

Можно показать, что операции сложения (объединения) и умножения (пересечения) событий подчиняются распределительным законам:

$$A(B + C) = AB + AC; \quad A + BC = (A + B)(A + C).$$

Поэтому

$$A = AH_1 + AH_2 + \ldots + AH_n.$$

Из попарной несовместности событий $H_1, H_2, \ldots, H_n$ следует и попарная несовместность событий $AH_1, AH_2, \ldots, AH_n$. Применяя теорему сложения для попарно несовместных событий, получим

$$P(A) = P(AH_1) + P(AH_2) + \ldots + P(AH_n).$$

Использование теоремы умножения вероятностей завершает доказательство формулы полной вероятности. ■

Пример 3. Решим рассмотренную в примере 2 задачу о телевизионном шоу.

❐ Введем следующие обозначения: A_1 — «местонахождение приза угадано при первом выборе», A_2 — «местонахождение приза угадано при втором выборе».

$P(A_1) = \frac{1}{3}$ (выбор производится наугад из трех дверей).

$P(A_2)$ найдем по формуле полной вероятности, приняв в качестве полной группы несовместных событий события A_1 и $\overline{A}_1$. Будем иметь

$$P(A_2) = P(A_1)\,P(A_2 \mid A_1) + P(\overline{A}_1)\,P(A_2 \mid \overline{A}_1) = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{3}.$$

Обратите внимание на то, что условная вероятность того, что при втором выборе дверь будет угадана, если она была угадана при первом, равна 1, а условная вероятность того, что при втором выборе дверь будет угадана, если она не была угадана при первом, равна $\frac{1}{2}$ (выбор производится наугад из двух дверей).

Итак, правильным является решение об изменении выбора: оно удваивает шансы на получение приза с $\frac{1}{3}$ до $\frac{2}{3}$. А какое решение вы приняли до рассмотрения этого решения?

Как и большинство приведенных вероятностных задач, данная задача может быть решена неформально.

Представим себе, что у участника шоу есть двойник, который меняет выбор в то время, когда сам участник свой выбор не меняет. Поскольку осталось только две двери и приз должен находиться за одной из них, двойник выиграет в каждом случае, когда участник проиграет. Поскольку общий шанс участника на выигрыш остается неизменным и составляет $\frac{1}{3}$, получается, что при изменившемся выборе двойник получает остальной шанс, равный $\frac{2}{3}$. ■

Пример 4. В урне 15 шаров, из них 10 белых. Наугад один за другим извлекают два шара, не возвращая их в урну. Какова вероятность того, что второй шар будет белый?

❐ Обозначим через A_i, $i = 1, 2$, событие «i-й извлеченный шар белый». События A_1 и $\overline{A}_1$ образуют полную группу несовместных событий. Поэтому применима формула полной вероятности:

$$P(A_2) = P(A_1)\,P(A_2|A_1) + P(\overline{A}_1)\,P(A_2|\overline{A}_1).$$

Так как

$$P(A_1) = \frac{10}{15} = \frac{2}{3}, \quad P(\overline{A}_1) = \frac{5}{15} = \frac{1}{3},$$

$$P(A_2|A_1) = \frac{9}{14},\ P(A_2|\overline{A}_1) = \frac{10}{14} = \frac{5}{7},$$

то $P(A_2) = \frac{2}{3} \cdot \frac{9}{14} + \frac{1}{3} \cdot \frac{5}{7} = \frac{2}{3}$.

Обратите внимание на то, что $P(A_1) = P(A_2) = \frac{2}{3}$! ■

Вероятность того, что мама возьмет своего сына с собой в магазин, равна 0,4. Если это происходит, то вероятность покупки новой игрушки равна 0,8. Если же она не взяла сына с собой в магазин, то игрушка покупается с вероятностью 0,3. Какова вероятность того, что сыну будет куплена новая игрушка?

Обратите внимание на то, что пример 10 из § 2.9, аналогичный предыдущему примеру, был решен без использования формулы полной вероятности, фактически там была реализована идея ее вывода.

Контрольные вопросы

1. Приведите пример n попарно несовместных событий, объединение которых есть достоверное событие, если:
а) $n = 2$; б) $n = 3$; в) $n = 4$.

2. Чему равна сумма вероятностей попарно несовместных событий, объединение которых есть достоверное событие?

3. Известно, что события B_1 и B_2 несовместны. Будут ли несовместны события AB_1 и AB_2?

4. О событиях B_1, B_2, B_3, B_4 известно, что $P(B_1) = 0{,}1$, $P(B_2) = 0{,}25$, $P(B_3) = 0{,}15$, $P(B_4) = 0{,}4$. Могут ли они образовывать полную группу событий?

5. О событиях B_1, B_2, B_3, B_4 известно, что $P(B_1) = 0{,}1$, $P(B_2) = 0{,}25$, $P(B_3) = 0{,}15$, $P(B_4) = 0{,}5$. Обязательно ли они образуют полную группу событий?

Задачи

178. Имеются два одинаковых ящика с шарами. В первом ящике два белых и один черный шар, во втором — один белый и четыре черных шара. Наудачу выбирают один ящик и вынимают из него шар. Какова вероятность того, что вынутый шар окажется белым?

179. На фабрике, изготовляющей болты, первая машина производит 25%, вторая — 35%, третья — 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%. Каковая вероятность того, что случайно выбранный болт дефектный?

180. В школе 60% учащихся — мальчики. 80% мальчиков и 75% девочек имеют билеты на футбольный матч. Какова вероятность того, что наугад выбранный учащийся школы имеет билет на футбольный матч?

181. На трех дочерей — Веру, Надю и Любу — в семье возложена обязанность мыть посуду. Старшая дочь Вера выполняет 50% всей работы. Остальные 50% Надя и Люба делят между собой поровну. Когда Вера моет посуду, вероятность для нее разбить хотя бы одну тарелку равна 0,01; для Нади и Любы эта вероятность соответственно равна 0,02 и 0,03. Какова вероятность того, что при мытье посуды будет разбита по крайней мере одна тарелка?

182. Вероятность получения патента равна 0,6. Если патент будет получен, условная вероятность получения дохода от него составит 0,9. Однако, если патент не будет получен, условная вероятность получения дохода составит только 0,3. Найдите вероятность получения дохода.

Дополнительные задачи к главе 2

К § 2.1. Статистическая вероятность

183. Выберите в телефонном справочнике подряд 100 телефонных номеров. Найдите распределение частот последних цифр этих номеров. Результаты занесите в таблицу следующего вида:

Цифра	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Частота										
Относительная частота										

Какой вывод можно сделать о распределении частот? Сравните свои результаты с результатами ваших одноклассников. Можно ли считать проведенные опыты статистически устойчивыми? Объедините результаты опытов, проведенных всеми одноклассниками. Какие выводы вы можете сделать?

184. Руководителю производства в конце рабочего дня докладывают, сколько выпущено чехлов для автомобильных сидений.

а) Опишите соответствующий случайный эксперимент.

б) Каким будет его выборочное пространство?

в) О чем могут свидетельствовать результаты эксперимента?

г) Точно определите событие «в соответствии с дневным планом выпущено 750 единиц продукции, плюс-минус 5 единиц» в терминах результатов, составляющих выборочное пространство.

д) За последние 15 дней в течение 8 дней это событие наблюдалось. Найдите соответствующую относительную частоту.

К § 1.2. Классическая вероятность

185. Игральный кубик бросают дважды. Какие значения суммы очков при двух бросках имеют наибольшую и наименьшую вероятность?

186. Есть пять отрезков длиной 1, 3, 4, 7 и 9 см. Найдите вероятность того, что из трех наугад выбранных отрезков можно построить треугольник.

187. В некоторой семье есть три ребенка. Считая вероятности рождения мальчика и девочки одинаковыми, найдите вероятность того, что все дети в этой семье мальчики.

188. Из 28 костей домино наугад выбирают одну. До вытягивания кости вы предсказываете сумму очков на ней. Если вы угадаете, то получите приз. Какой сумме вы отдадите предпочтение?

189. Рассматривается упрощенный вариант игры в спортлото. Случайно выбирают два шарика из урны, в которой находится пять шариков, пронумерованных числами 1, 2, 3, 4, 5. Игрок вычеркивает два числа на бланке (рис. 32). Найдите вероятность того, что игрок:

а) угадает оба номера извлеченных шариков;

б) угадает лишь один номер;

в) угадает по крайней мере один номер;

г) не угадает ни одного номера.

1	2	~~3~~	~~4~~	5

Рис. 32

190. Рассматривается игра в рулетку. Есть вращающийся диск (он разделен на 37 одинаковых секторов, пронумерованных числами от 0 до 36; рис. 33), металлический шарик, панель с полями (рис. 34), жетоны. На панели обозначены поля с отдельными числами (от 0 до 36), а также другие поля, обозначающие части совокупности $\{0, 1, 2, ..., 36\}$: поле Pair — четные числа, поле Impair — нечетные числа, поле Manque — числа от 1 до 18, поле Passe — числа от 19 до 36, поле 12^P — числа от 1 до 12 (первая дюжина), поле 12^M — числа от 13 до 24 (вторая дюжина), поле 12^D — числа от 25 до 36 (третья дюжина). Есть на диске «черные» числа (они на диске вписаны в темные клетки, на панели — поле с черным ромбом),

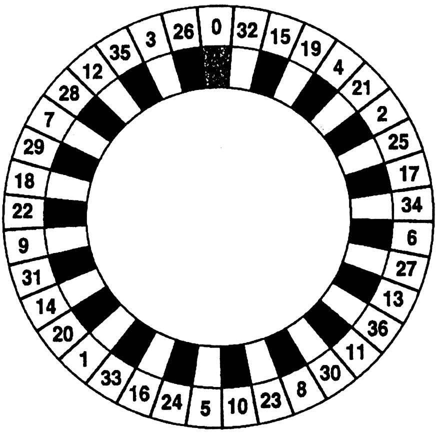


Рис. 33

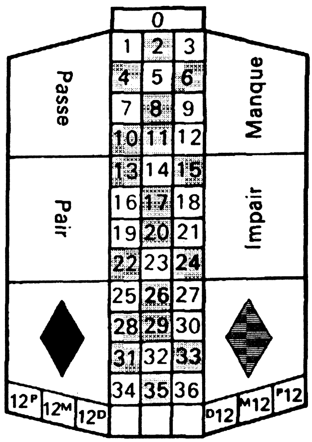


Рис. 34

есть красные числа (они на диске вписаны в белые клетки, на панели — поле с серым ромбом). С началом игры игрок кладет свои жетоны на выбранные поля панели. На вращающийся диск бросают шарик. Какова вероятность того, что выиграет игрок, который положил жетон:

а) на число 13; б) на «черное»; в) на 12^{P}; г) на Pair?

К § 2.3. Субъективная вероятность

191. Укажите, к какому типу вероятностей (имеется в виду источник ее получения) относится вероятность события:

а) «скаковая лошадь Чемпион победит на скачках»;

б) «наудачу выбранный головной убор, выпущенный фабрикой, имеет дефекты», если известно, что из 925 выпущенных фабрикой головных уборов 13 имеют дефекты;

в) «деталь имеет дефект»;

г) «заявленная величина дохода выше ожидаемой».

К § 2.4. Вероятностная модель случайного опыта

192. ПЭИ опыта состоит из n исходов, взаимно исключающих друг друга, причем $n - 1$ из этих исходов имеют равные вероятности, а последний имеет вероятность, равную сумме вероятностей $r + 1$ других, $1 \leqslant r \leqslant n - 1$. Найдите вероятности элементарных исходов данного ПЭИ.

193. Четверо друзей Юрий, Андрей, Владимир и Игорь сообща приобрели лодку, причем Юрий внес 10% стоимости лодки, Андрей — 20%, Владимир — 30% и Игорь — 40% стоимости. В праздничный день они жребием решают, кому достанется лодка в этот день. Постройте вероятностную модель этого эксперимента, считая, что:

а) шансы воспользоваться в этот день лодкой не зависят от внесенного взноса;

б) шансы воспользоваться в этот день лодкой пропорциональны внесенному взносу.

194. Из колоды в 36 карт вынимают карты одну за другой до появления первого туза. Постройте ПЭИ этого эксперимента.

195. Тщательно перетасованы четыре карточки, причем на двух из них одна сторона окрашена в красный цвет. Карточки сложены в стопку. Последовательно одну за другой открывают верхние карточки.

а) Постройте ПЭИ, соответствующее этому эксперименту.

б) Введите элементарные вероятности.

196.* В каждую единицу времени к прилавку с вероятностью $p = \frac{1}{2}$ подходит новый покупатель. Если у прилавка стоят другие покупатели, то он станет в очередь. В противном случае продавец обслужит его в течение $m = 2$ единиц времени. В исходный момент у прилавка нет ни одного покупателя.

а) Постройте ПЭИ U движения покупателей в течение 4 единиц времени.

б) Введите элементарные вероятности p_k. Проверьте, выполняется ли равенство $\sum\limits_k p_k = 1$.

К § 2.5. Случайные события и их вероятности

197. Наугад выбирают по одной букве из слов «дама» и «мама».

а) Постройте ПЭИ опыта и введите вероятности его элементов.

б) Какова вероятность того, что элементы ПЭИ будут состоять из одинаковых букв?

198. Известно, что 5% новорожденных мальчиков и 0,25% новорожденных девочек страдают дальтонизмом. Предположим, что 50% новорожденных — мальчики, и 50% — девочки. Наугад выбирают одного новорожденного. Отмечают его пол и наличие или отсутствие у него дальтонизма.

а) Постройте ПЭИ, соответствующее этому эксперименту.

б) Введите элементарные вероятности.

в) Найдите вероятность того, что новорожденный:

— мальчик и страдает дальтонизмом;

— девочка и страдает дальтонизмом;

— страдает дальтонизмом.

199. Два игрока играют серию партий, причем в каждой партии какой-то из игроков выигрывает очко. Шансы на выигрыш для каждого игрока одинаковы. Тот из игроков, который первым наберет 4 очка, получает приз. Игру вынуждены были прекратить в момент, когда первый игрок набрал 3 очка, а второй — 1 очко.

а) Постройте ПЭИ U опыта, заключающегося в том, что игра доигрывается.

б) Введите элементарные вероятности p_k. Проверьте, выполняется ли равенство $\sum\limits_k p_k = 1$.

в) Выпишите все исходы, образующие события:

A — «игра закончилась победой первого игрока»,

B — «игра закончилась победой второго игрока»,

C — «для завершения игры понадобилось сыграть более одной партии».

г) Вычислите вероятности этих событий.

д)* Как следует разделить приз между игроками?

200. Два игрока обладают начальным капиталом в 3 и 2 единицы соответственно. Чтобы участвовать в одной партии, каждый должен поставить на кон 1 единицу своего капитала. Шансы на выигрыш для каждого игрока одинаковы. Выигрывающий партию забирает обе ставки. Игра продолжается до разорения какого-то игрока, но не более четырех партий.

а) Постройте ПЭИ U этой игры.

б) Введите элементарные вероятности p_k. Проверьте, выполняется ли равенство $\sum\limits_k p_k = 1$.

в) Выпишите все исходы, образующие события:

A — «разорится первый игрок»,

B — «разорится второй игрок»,

C — «игра продлится не более трех партий».

г) Вычислите вероятности этих событий.

201. В точке O числовой оси находится некоторая частица. Каждую секунду она с равной вероятностью сдвигается на единицу либо вправо, либо влево.

а) Постройте ПЭИ U движения частицы в течение 4 с.

б) Введите элементарные вероятности p_k. Проверьте, выполняется ли равенство $\sum\limits_k p_k = 1$.

в) Выпишите все исходы, образующие события:

A — «частица через 4 с будет находиться на расстоянии 2 единиц от точки O»,

B — «частица за 4 с удалится на 2 единицы вправо от точки O»,

C — «частица через 4 с будет находиться в точке O».

г) Вычислите вероятности этих событий.

К § 2.6. Операции над событиями

202.° Что означают сумма и произведение указанных ниже событий, связанных с соответствующими испытаниями:

а) сделано два выстрела по мишени; A — «попадание при первом выстреле», B — «попадание при втором выстреле»;

б) бросают игральный кубик; A — «выпала единица», B — «выпала двойка»;

в) купили два лотерейных билета; A — «первый билет выигрышный», B — «второй билет выигрышный».

203. Трижды бросают мяч в баскетбольную корзину. Пусть события A_i ($i = 1, 2, 3$) означают, что при i-м броске мяч попадает в корзину. Выразите через события A_1, A_2, A_3 событие:

а)° B — «мяч ни разу не попал в корзину»;

б)° C — «мяч по крайней мере один раз попал в корзину»;

в)° D — «мяч попал в корзину при всех бросках»;

г) E — «мяч попал в корзину лишь при первом броске»;

д) F — «мяч попал в корзину лишь при первом и втором бросках»;

е) G — «мяч попал в корзину ровно один раз»;

ж) H — «мяч попал в корзину ровно два раза».

204. Прибор состоит из двух элементов. Рассмотрим следующие события: B_1 — «работают оба элемента»; B_2 — «работает только первый элемент»; B_3 — «работает только второй элемент»; B_4 — «оба элемента вышли из строя».

а) Выразите через эти события следующие события: A_1 — «первый элемент работает»; A_2 — «второй элемент работает»; A_3 — «один элемент работает»; A_4 — «хотя бы один элемент работает».

б) Выразите события B_i ($i = 1, 2, 3, 4$) через события A_1 и A_2.

в) Что означают события $B_2 + B_3$; B_1B_2; $B_1\overline{B}_4$?

К § 2.7. Шансы в пользу события

205. Из урны, содержащей два белых и один красный шар, наугад вынимают два шара. Каковы шансы в пользу события:

а) «вынутые шары одноцветны»;

б) «вынутые шары разноцветны»?

206. Как изменятся шансы в пользу событий, рассмотренных в предыдущей задаче, если в урну добавить белый шар?

207. Дважды бросают правильный игральный кубик. Каковы шансы в пользу события:

а) «сумма выпавших очков четная»;

б) «сумма выпавших очков меньше 5»;

в) «шестерка не выпадет ни разу»?

208. Водитель путем продолжительных наблюдений сделал вывод, что в 80% тех случаев, когда он моет свой автомобиль, на следующий день идет дождь. На каких условиях следует идти на пари, что это состоится и в следующий раз?

209. Григорий предлагает честное пари на условиях 3 : 1, что настанет событие A. Какими он считает вероятности событий $\bar{A}$ и A?

210. Из колоды карт взяты шесть карт: два туза, два короля, две дамы. Карты разложены в две стопки, причем в каждой стопке по одному тузу, одному королю и одной даме. Обе стопки карт тщательно перемешаны. Затем открывают последовательно верхние карты, и регистрируют, лежат ли на одинаковых местах карты одинакового достоинства.

а) Постройте ПЭИ опыта, приняв в качестве его исходов расположение тройки карт.

б) Введите элементарные вероятности.

в) Каковы шансы в пользу события «хотя бы на одном месте расположены карты одинакового достоинства»?

г)* На каких условиях заключается честное пари о том, что «точно на двух местах расположены карты одинакового достоинства»?

211. В урне четыре шара с номерами 1, 2, 3, 4. Случайно извлекают два шара, и подсчитывают сумму их номеров.

а) Постройте ПЭИ опыта, приняв в качестве его исходов суммы номеров извлеченных шаров.

б) Введите элементарные вероятности.

в) Каковы шансы в пользу события «сумма номеров извлеченных шаров нечетна»?

г) На каких условиях заключается честное пари о том, что «сумма номеров извлеченных шаров окажется нечетной»?

К § 2.8. Вероятность суммы событий

212. Игральный кубик налили свинцом так, что вероятности выпадения каждой грани стали пропорциональными числу очков на них. Чему равна вероятность выпадения четного числа очков при одном броске кубика?

213. Продолжительные наблюдения показывают, что приблизительно 60% студентов первого курса математического факультета сдают с первой попытки экзамен по математическому анализу; 80% — по крайней мере один из двух экзаменов по математическому анализу и линейной алгебре; 50% — по обоим этим предметам. Чему равна вероятность того, что наугад выбранный студент сдаст экзамен с первой попытки по курсу линейной алгебры?

214. Сергей предлагает пари на условиях 1 : 3, что настанет событие A и пари на условиях 1 : 2, что настанет событие B. Он знает, что события A и B не могут произойти одновременно. На каких условиях он пойдет на пари, что произойдет по крайней мере одно из двух событий A или B?

215. Цех дает в среднем 32% продукции высшего сорта и 60% — первого сорта. Найдите вероятность того, что наугад взятое изделие будет:

а)° первого или высшего сорта;

б) худшим по сравнению с первым сортом.

216. В зрительном зале кинотеатра 9 рядов, пронумерованных подряд числами от 1 до 9, а в каждом ряду по 9 кресел, также пронумерованных числами от 1 до 9. Зритель наудачу занимает место. Какова вероятность того, что сумма номеров ряда и ме́ста в ряду окажется:

а) четной; б) нечетной?

217.* Не менее чем 60% студентов лингвистического университета изучают немецкий язык, не менее чем 70% — французский язык; не менее чем 80% — английский язык. Какова наименьшая доля студентов, изучающих одновременно немецкий, французский и английский языки?

218.* Из первой сотни натуральных чисел наугад выбирают одно. Какова вероятность того, что оно не делится ни на одно из чисел 2, 3, 5?

К § 2.9. Условные вероятности

219. Два игрока играют в следующую игру. Есть кубик, развертка которого изображена на рисунке, и урна с пронумерованными шарами (рис. 35). Выбирают число с помощью кубика (номер

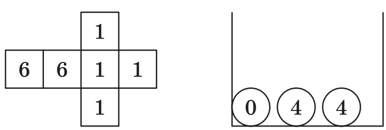


Рис. 35

на верхней грани после броска кубика) или с помощью урны (номер наугад извлеченного шара). Начинающий игру получает право выбора прибора (кубик или урну). Побеждает тот, кто выберет большее число. Что лучше выбрать: кубик или урну, если вам предложат начать игру?

220. В мае 20% дождливых дней. Продолжительными наблюдениями было подмечено, что некоторая футбольная команда в ясный день побеждает с вероятностью 0,4, а в дождливый — с вероятностью 0,7.

а)° Найдите вероятность того, что в некоторый день мая будет дождь, и эта команда победит.

б) Какова вероятность того, что команда победит в наугад выбранный майский день?

в) Известно, что команда выиграла игру в мае. Какова вероятность того, что в этот день шел дождь?

221. Тест с выбором одного ответа из некоторых предложенных содержит для каждой задачи четыре варианта ответа. Таким образом, если ученик знает правильный ответ, то для него вероятность правильного ответа равняется 1; если он просто угадывает, то вероятность правильного ответа составляет 0,25. Старательный ученик знает 90% ответов, посредственный — лишь 50%.

а) Чему равна вероятность того, что старательный (посредственный) ученик ответил правильно на случайно выбранный вопрос?

б) Старательный (посредственный) ученик правильно ответил на вопрос. Какова вероятность того, что он угадал ответ?

222. Группе лиц вручают четыре конверта, в каждом из которых помещено условие одной задачи. Группе предлагают открыть один конверт и постараться решить помещенную там задачу в течение 10 минут. Из предшествующего опыта известно, что вероятность того, что за 10 минут будет решена самая трудная задача, равна 0,1, а для других задач эти вероятности равняются 0,3; 0,5; 0,8.

а) Чему равна вероятность того, что будет открыт конверт с самой трудной задачей, и она будет решена за 10 минут?

б) Чему равна вероятность того, что за 10 минут будет решена задача из открытого конверта?

в) Группа справилась с задачей за установленное время. Какова вероятность того, что был открыт конверт с самой трудной задачей?

223. На протяжении длительного времени Анна заметила, что с вероятностью 0,55 она ходит в кино, если идет дождь, и с вероятностью 0,30, если дождя нет. Метеорологическая служба прогнозирует на следующий день дождь с вероятностью 0,4.

а)° Какова вероятность того, что в этот день пойдет дождь, и Анна пойдет в кино?

б) Какова вероятность того, что Анна в этот день пойдет в кино?

в) Анна пошла в этот день в кино. Найдите вероятность того, что дождя не было.

224.* На каждые 100 000 мужчин приходится в среднем пять больных раком легких. Курят 75% всех больных раком легких и 60% тех, кто не болеет этой болезнью. Вычислите долю больных раком легких среди тех, кто курит, и среди тех, кто не курит.

225. В 1998 году палата представителей Конгресса США подготовила к выходу в свет видеозапись показаний Большому жюри президента Клинтона; его рейтинг одобрения был таким: 36% одобряли его как личность, 63% одобряли его как президента, 30% одобряли его как президента, но не как личность. Найдите процент людей, которые одобряли его как личность, но не как президента.

226. Вы принимаете участие в телевизионном шоу и боретесь за получение приза, спрятанного за одной из пяти дверей. Есть только один приз, и он спрятан за одной из этих дверей. После того как вы сделали свой выбор, устроители шоу открывают три двери (но не ту, которую вы выбрали), за которыми нет приза. У вас есть возможность изменить свой первоначальный выбор и выбрать другую неоткрытую дверь. Измените ли вы свое решение?

К § 2.10. Независимые события

227. Необходимо разжечь костер при наличии двух спичек. Можно попробовать разжечь костер одной спичкой, а потом, если это не удалось, — второй, или обеими спичками одновременно. Продолжительными наблюдениями замечено, что вероятность разжечь костер одной спичкой равна 0,7, а двумя одновременно — 0,95. Какое решение оптимальное? Ответ обоснуйте.

228. Результаты экзаменов в некотором техникуме показали, что 8% студентов не смогли сдать экзамен по математике, 6% — по физике и 2% не сдали экзамен ни по математике, ни по физике. Будут ли события «наугад выбранный студент не сдал экзамен по математике» и «наугад выбранный студент не сдал экзамен по физике» независимыми?

229. Симметричную монету подбрасывают три раза. Будут ли независимыми события «при первом подбрасывании выпадает герб» и «выпадет не меньше двух гербов»?

230. Каждая из трех электрических схем состоит из четырех выключателей. Каждый из выключателей с вероятностью 0,5 может быть включен или выключен. Выясните, для какой из схем, изображенных на рис. 36, вероятность того, что ток будет проходить от точки A к точке B, будет наибольшей?

231. Каждая из двух электрических схем состоит из шести выключателей. Каждый из выключателей с вероятностью 0,5 может быть включен или выключен. Выясните, для какой из схем, изображенных на рис. 37, вероятность того, что ток будет проходить от точки A к точке B, будет наибольшей?

232. Два игрока играют серию партий, причем в каждой партии какой-то из игроков выигрывает очко. Шансы на выигрыш относятся как 1 : 2. Тот из игроков, который первым наберет 4 очка, получает приз. Игру вынуждены были прекратить в момент, когда первый игрок набрал 2 очка, а второй 1 очко.

а)° Найдите вероятности выигрышей для обоих игроков в одной партии.

б) Постройте ПЭИ опыта, заключающегося в том, что игра доигрывается.

в) Введите элементарные вероятности исходов этого опыта.

г) Каковы шансы в пользу наступления события «приз получит первый игрок»?

д) Как разделить приз между игроками?

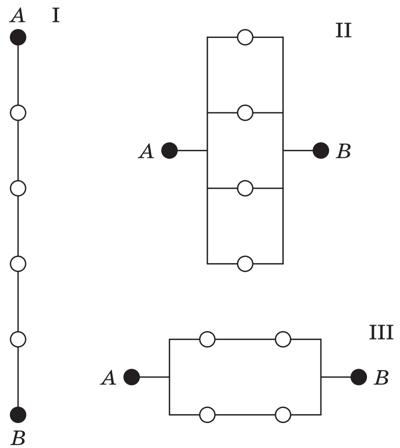


Рис. 36

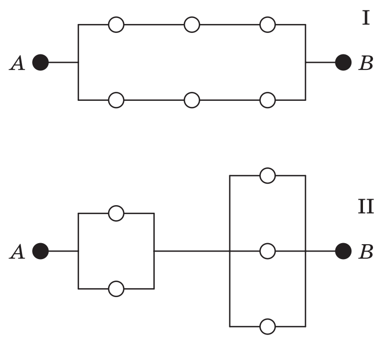


Рис. 37

е) На каких условиях должно заключаться честное пари на победу первого игрока?

233. Два игрока обладают начальным капиталом в 3 и 1 единицы соответственно. Чтобы участвовать в одной партии, каждый должен поставить на кон 1 единицу своего капитала. Шансы на выигрыш относятся как 2 : 3. Выигрывающий партию забирает обе ставки. Игра продолжается до разорения какого-то игрока, но не более четырех партий.

а)° Найдите вероятности выигрышей для обоих игроков в одной партии.

б) Постройте ПЭИ этой игры.

в) Введите элементарные вероятности исходов этого опыта.

г) Каковы шансы в пользу наступления события «разорится первый игрок»?

д) На каких условиях должно заключаться честное пари на разорение первого игрока?

К § 2.11. Формула полной вероятности

234.* При переливании крови надо учитывать группу крови донора и больного. Человеку, имеющему четвертую группу крови, можно перелить кровь любой группы, человеку со второй или третьей группой можно перелить кровь либо той же группы, либо первой. Человеку с первой группой крови можно перелить кровь только первой группы. Среди населения 33,7% имеют первую, 37,5% — вторую, 20,9% — третью и 7,9% — четвертую группу крови. Найдите вероятность того, что случайно взятому больному можно перелить кровь случайно взятого донора.

235.* За некоторый промежуток времени амеба может погибнуть с вероятностью $\frac{1}{4}$, выжить с вероятностью $\frac{1}{4}$, разделиться на две с вероятностью $\frac{1}{2}$. В следующий промежуток времени с каждой амебой происходит то же самое. Сколько амеб, и с какими вероятностями будут существовать к концу второго промежутка времени, если в момент начала отсчета времени число амеб равнялось 1?

236.* Студент знает не все экзаменационные билеты. В каком случае вероятность вытащить неизвестный билет будет для него наименьшей: когда он тащит билет первым или последним?

Глава 3

ЭЛЕМЕНТЫ КОМБИНАТОРИКИ

В опытах, которые рассматривались в предыдущей главе, число исходов можно было подсчитать непосредственно. Но довольно часто этого сделать невозможно без использования специальных приемов.

Область математики, в которой изучается вопрос о том, сколько различных конфигураций, удовлетворяющих тем или иным условиям, можно составить из заданных объектов, называют ***комбинаторикой***.

В повседневной жизни нередко возникают проблемы, которые имеют не один, а несколько вариантов решения. Чтобы сделать правильный выбор, очень важно не упустить ни один из них. Для этого надо уметь осуществлять перебор всех возможных вариантов или хотя бы подсчитывать их число. Такого рода задачи называют ***комбинаторными***.

Комбинаторика — важный раздел математики, знание которого необходимо представителям самых разных специальностей. С комбинаторными задачами приходится иметь дело физикам, химикам, биологам, лингвистам, экономистам, специалистам по кодам, компьютерам, информационным технологиям и т. д. Комбинаторные методы лежат в основе решения многих задач теории вероятностей, математической статистики и их приложений.

В настоящей главе рассмотрены различные методы решения комбинаторных задач. Прежде всего предлагается освоить метод перебора различных вариантов. При этом появляется возможность различать случаи, когда конфигурации составляются из одинаковых элементов или из различных; когда конфигурации содержат данный элемент ровно один раз, не более одного раза, хотя бы один раз, более одного раза; когда порядок извлечения элементов будет существенным или несущественным при составлении конфигураций; когда конфигурация образуется выбором элементов с возвращением или без возвращения. Метод перебора вариантов будет полезен и в том случае, когда вы овладеете другими методами решения комбинаторных задач. Иногда другие методы или оказываются неприменимыми, или приводят к сложным рассуждениям. Но в любом случае метод перебора вариантов можно использовать для самоконтроля.

Далее рассмотрен метод решения комбинаторных задач, основанный на применении правил умножения и сложения. Этот метод позволяет решать разнообразные комбинаторные задачи, практически не прибегая к использованию формул. Он дает возможность овладеть основными комбинаторными идеями.

И наконец, комбинаторные задачи будут решаться с помощью формул для числа размещений, сочетаний, перестановок как с повторениями, так и без повторений.

Элементы комбинаторики будут применены для обобщения известных формул квадрата и куба двучлена, т. е. для получения так называемой формулы Ньютона n-й степени двучлена.

Основное применение комбинаторики связано с решением вероятностных задач, приводящих к классической вероятностной модели.

§ 3.1. Перебор возможных вариантов

Сущность *метода перебора вариантов* рассмотрим на примерах.

Пример 1. Проводится игра. Из коробочки, содержащей три белых и два красных шара, наугад вынимают два.

а) Ведущий перед извлечением шаров принимает у зрителей ставки на число вынутых белых шаров. На сколько белых шаров вы поставите?

б) Ведущий принимает ставки на два исхода игры: шары одинакового цвета, шары разного цвета. На какой исход вы поставите?

Чтобы ответить на поставленные вопросы, нужно или подсчитать число всех возможных вариантов исхода игры, или перебрать их, а затем подсчитать их число. Перебор можно осуществить различными способами.

Первый способ — ***способ кодировки.***

❐ Обозначим белые шары цифрами 1, 2, 3, а красные — буквами а, б, т. е. закодируем предметы. Возможные варианты:

1 2	1 3	2 3			
1 а	1 б	2 а	2 б	3 а	3 б
а б					

Заметим, что вариант 1 2 означает, что извлечены два белых шара (1 и 2). Вариант 2 1 мы не записываем, так как он совпадает с вариантом 1 2. Всего получили 10 вариантов извлечения двух шаров, в трех из них два белых шара (1 2, 1 3, 2 3), в шести — один белый шар (1 а, 1 б, 2 а, 2 б, 3 а, 3 б) и в одном — нет белых шаров (а б).

а) Ясно, что если игру повторить много раз, то чаще будут появляться варианты с одним белым шаром. Теперь понятно, что нужно ставить на один белый шар. Фактически мы подсчитали, что вероятность появления двух белых шаров равна 0,3, одного белого шара — 0,6, отсутствия белых шаров — 0,1.

б) Из двух исходов — шары одинакового цвета и шары разного цвета — при многократном повторении опыта чаще будет иметь место второй исход: шесть вариантов из десяти против четырех из десяти, т. е. вероятности этих исходов равны соответственно 0,6 и 0,4. Ставить нужно на шары разного цвета. ■

Второй способ перебора основан на *построении* так называемого ***дерева возможных вариантов***.

❐ Звездочка (*) на рис. 38 изображает корень дерева, ветви дерева — различные варианты. Чтобы вынуть два шара, нужно сначала вынуть первый, а для этого есть пять вариантов (1, 2, 3, а, б). Поэтому от звездочки проведены пять отрезков и на их концах поставлены обозначения 1, 2, 3, а, б. Затем нужно вынуть второй шар из оставшихся четырех. Поэтому от конца каждого отрезка проведены по четыре отрезка, на концах которых написаны обозначения оставшихся шаров. Получено 20 вариантов извлечения шаров. Но среди них каждый вариант повторяется дважды: 1 2 и 2 1, 1 3 и 3 1, а б и б а и т. д. Итак, различных вариантов 10 и других вариантов извлечения шаров нет. Получили тот же результат, что и способом кодировки. ■

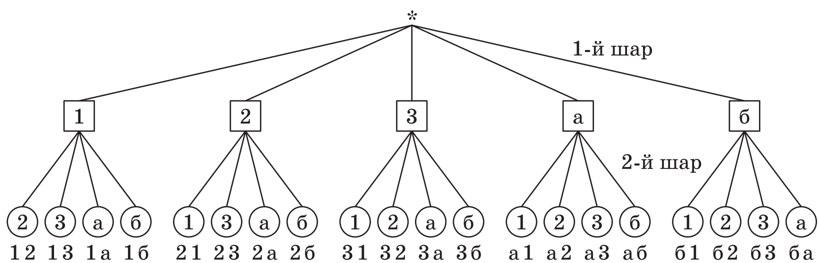


Рис. 38

Третий способ (***набор точек и отрезков***) применим в том случае, когда из некоторой совокупности предметов выбирают два. Изобразим шары в виде точек, расположенных так, что никакие три точки не лежат на одной прямой (рис. 39). Затем соединяем каждые две точки отрезком прямой. Всего получено 10 отрезков. Каждый из них изображает вариант извлечения двух шаров. Снова получили тот же результат.

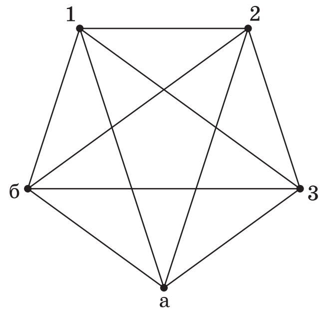


Рис. 39

Четвертый способ — ***табличный***. Его можно применять как в случае подсчета числа вариантов, с помощью которых можно извлечь из данной совокупности некоторое количество элементов, удовлетворяющих определенным условиям, так и в случае нахождения числа способов разбиения совокупности различных или одинаковых элементов на заданное число групп.

Пример 2. Сколькими способами можно распределить четыре одинаковых карандаша между тремя детьми?

❐ Решение представлено в таблице 3.1.

Таблица 3.1

№ ребенка	№ варианта														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	4	0	0	3	3	1	1	0	0	2	2	0	2	1	1
2	0	4	0	1	0	3	0	3	1	2	0	2	1	2	1
3	0	0	4	0	1	0	3	1	3	0	2	2	1	1	2

Целесообразно заполнять эту таблицу по столбцам. Первый столбец означает, что первый ребенок получил все четыре карандаша, а двум другим карандаши не достались. Обратите внимание на то, что карандаши одинаковые. Поэтому, например, для варианта (3, 1, 0), стоящего в четвертом столбце, не следует рассматривать различные способы выбора трех карандашей из четырех. Из четырех одинаковых элементов три (и любое другое число) можно выбрать единственным способом.

Ответ: 15. ■

Пример 3. Сколькими способами можно распределить два различных карандаша между тремя детьми?

❒ Обозначив карандаши цифрами 1 и 2, можно рассмотреть все варианты распределения карандашей (табл. 3.2). В клетках таблицы указаны номера карандашей, полученных соответствующим ребенком.

Таблица 3.2

№ ребенка	№ варианта								
	1	2	3	4	5	6	7	8	9
1	1, 2	—	—	1	1	—	2	2	—
2	—	1, 2	—	2	—	1	1	—	2
3	—	—	1, 2	—	2	2	—	1	1

Обратите внимание на то, что, например, в четвертом и седьмом столбцах стоят различные варианты распределения карандашей: ведь карандаши разные.

Ответ: 9. ■

Пример 4. Сколькими способами из пяти шахматистов можно выбрать трех для участия в соревнованиях и, кроме того, одного капитана?

❒ Обозначив шахматистов цифрами 1, 2, 3, 4, 5, можно рассмотреть все варианты выбора (табл. 3.3). В клетках таблицы указаны номера выбранных шахматистов.

Таблица 3.3

Выбор	№ варианта									
	1	2	3	4	5	6	7	8	9	10
Трех шахматистов	1 2 3	1 2 3	1 2 4	1 2 4	1 2 5	1 2 5	1 3 4	1 3 4	1 3 5	1 3 5
Капитана	4	5	3	5	3	4	2	5	2	4
Выбор	**№ варианта**									
	11	12	13	14	15	16	17	18	19	20
Трех шахматистов	1 4 5	1 4 5	2 3 4	2 3 4	2 3 5	2 3 5	2 4 5	2 4 5	3 4 5	3 4 5
Капитана	2	3	1	5	1	4	1	3	1	2

Обратите внимание на то, что сначала выбирались три шахматиста для участия в соревнованиях, а потом из двух оставшихся выбирался капитан. Задачу можно решить и иначе: сначала выбрать четырех спортсменов из пяти, а потом из них — капитана, или сначала выбрать капитана, а потом из четырех оставшихся шахматистов — трех участников соревнований.

Ответ: 20. ■

Постройте соответствующие таблицы для каждого из этих способов и убедитесь в том, что они приводят к одному и тому же результату.

Пример 5. В кондитерском кафе имеются пирожные четырех сортов: «Наполеон», «Эклер», песочные и слоеные. Сергей и Оксана решили купить по одному пирожному.

а) Сколько существует вариантов такой покупки?

б) Сколько существует вариантов покупки, при которой Сергей и Оксана не хотят покупать пирожные одного вида?

в) Сколько существует вариантов покупки, если Сергей оба пирожных покупает Оксане?

г) Сколько существует вариантов покупки, если оба пирожных покупаются Оксане, и она не хочет, чтобы они были одного сорта?

❐ Закодируем сорта пирожных буквами Н, Э, П, С (соответственно).

а) Н Н Э Н П Н С Н
Н Э Э Э П Э С Э
Н П Э П П П С П
Н С Э С П С С С

Всего 16 вариантов и все они различны.

б) Если Сергей и Оксана не хотят покупать пирожные одного сорта, то нужно исключить варианты Н Н, Э Э, П П, С С. Остается 12 различных вариантов.

в) Если оба пирожных покупают для Оксаны, то, например, варианты Н Э и Э Н, П Н и Н П не различаются и получаются следующие 10 вариантов:

Н Н Н Э Н П Н С
Э Э Э П Э С
П П П С
С С

г) Если оба пирожных покупают для Оксаны, и она не хочет, чтобы они были одного сорта, то надо исключить варианты Н Н, Э Э, П П, С С и останется шесть вариантов:

Н Э Н П Н С Э П Э С П С

Ответ: 16, 12, 10, 6. ■

Обратите внимание, что в заданиях а)—г) речь шла о выборе двух предметов из четырех. В первом случае никаких ограничений на выбор не делалось. Во втором предполагалось, что один предмет не может быть выбран дважды. В задании в не различался порядок выбранных предметов, однако предметы могли быть и одинаковыми. В последнем задании снова не различался порядок выбранных предметов, и ни один предмет не мог быть выбран дважды.

? Попробуйте решить задачу, пользуясь другими способами перебора возможных вариантов. Сравните их.

В большинстве из выше рассмотренных задач элементы, из которых осуществлялся выбор, или которые распределялись по различным группам, были различными. Решение задачи существенным образом меняется, если элементы считать одинаковыми. Например, из пяти различных элементов *а*, *б*, *в*, *г*, *д* два элемента можно выбрать 10 способами: *а б, а в, б в, а г, б г, в г, а д, б д, в д, г д*, а из пяти одинаковых *а*, *а*, *а*, *а*, *а* — только одним способом (выбранные два элемента *а а* ничем не будут отличаться от другой пары *а а*). Рассмотрим следующую задачу.

Пример 6. Сколькими способами три карандаша можно распределить между двумя детьми, если карандаши:

а) различные; б) одинаковые?

❒ а) Обозначим карандаши цифрами 1, 2, 3. Все способы их распределения между тремя детьми представим в таблице 3.4.

Таблица 3.4

№ ребенка	**№ способа**							
	1	2	3	4	5	6	7	8
1	1 2 3	—	1 2	1 3	2 3	1	2	3
2	—	1 2 3	3	2	1	2 3	1 3	1 2

Итак, восемью способами можно распределить три различных карандаша между двумя детьми. Например, 4-й и 5-й способы различны, хотя первый ребенок получил при каждом из этих способов два карандаша, а второму достался один карандаш, но первый ребенок получил при этом различные карандаши.

б) Если карандаши одинаковые, то различные способы распределения карандашей между двумя детьми будут отличаться только количеством карандашей, которые достанутся детям. Представим эти способы в таблице 3.5. В клетках на пересечении строк и столбцов стоит количество карандашей.

Таблица 3.5

№ ребенка	№ способа			
	1	2	3	4
1	3	—	2	1
2	—	3	1	2

Таким образом, есть четыре способа распределения трех одинаковых карандашей между двумя детьми. Заменили в условии различные предметы одинаковыми — получили другую задачу. Поэтому при решении подобных задач обращайте внимание на то, о каких предметах — одинаковых или различных — идет речь. ■

Пример 7. Сколькими способами три карандаша можно разделить на две группы, если карандаши:

а) различные; б) одинаковые?

❐ В предыдущей задаче мы тоже делили три карандаша на две группы, но при этом группы были различны: карандаши распределялись между детьми. Поэтому распределение, при котором первому ребенку достались карандаши 1 2, а второму — карандаш 3, отличалось от распределения, при котором первому ребенку достался карандаш 3, а второму — карандаши 1 2. А вот если мы раскладываем карандаши на две «кучки», то распределения 1 2 + 3 и 3 + 1 2 ничем не отличаются друг от друга.

Поэтому понятным будет представленные в таблицах 3.6 и 3.7 решения задач а и б соответственно.

Таблица 3.6

№ группы	№ способа			
	1	2	3	4
1	1 2 3	1 2	1 3	2 3
2	—	3	2	1

Таблица 3.7

№ группы	№ способа	
	1	2
1	3	2
2	—	1

Итак, имеем соответственно четыре (два) способа распределения на две группы трех различных (одинаковых) карандашей. Обратите внимание на то, что по сравнению с примером 7, а (где группы были различимы) в примере 7, б число способов уменьшилось вдвое, т. е. в число способов, с помощью которых можно переставить два элемента (число групп). ■

Пример 8. Сколькими способами шесть шахматистов из клуба «Данко», среди которых есть два мастера спорта, могут разделиться на две команды по три спортсмена в каждой так, чтобы в любой из них было по одному мастеру спорта, если:

а) эти команды будут представлять свой клуб на одновременно проходящих командных соревнованиях в двух городах;

б) если эти команды будут играть между собой?

❐ Обозначим шахматистов — мастеров спорта буквами *м*, *н*, а остальных — цифрами 1, 2, 3, 4.

а) Их разбиение на две группы представим в таблице 3.8.

Таблица 3.8

№ группы	№ способа					
	1	2	3	4	5	6
1	1 2 *м*	1 2 *н*	1 3 *м*	1 3 *н*	1 4 *м*	1 4 *н*
2	3 4 *н*	3 4 *м*	2 4 *н*	2 4 *м*	2 3 *н*	2 3 *м*
№ группы	**№ способа**					
	7	8	9	10	11	12
1	2 3 *м*	2 3 *н*	2 4 *м*	2 4 *н*	3 4 *м*	3 4 *н*
2	1 4 *н*	1 4 *м*	1 3 *н*	1 3 *м*	1 2 *н*	1 2 *м*

Заметьте, например 1-й и 12-й способы различны: в первом случае команда 1 2 *м* едет, скажем, на соревнования в Пермь, а команда 3 4 *н* — в Казань, а во втором случае — наоборот: команда 1 2 *м* — в Казань, а команда 3 4 *н* — в Пермь. Поэтому эти два разбиения естественно считать различными. Всего имеем 12 разбиений шести шахматистов на две команды.

б) В этом случае разбиения 1 и 12 одинаковы, так как безразлично, как считать: команда 1 2 *м* играет с командой 3 4 *н* или команда 3 4 *н* играет с командой 1 2 *м*. Точно также не различаются разбиения 2 и 11, 3 и 10, 4 и 9, 5 и 8, 6 и 7. Всего получили шесть различных разбиений.

Обратите внимание на то, что в первом случае рассматривалось разбиение шахматистов на две различимые группы, а во втором — на две неразличимые группы. Число разбиений уменьшилось в 2 раза. ■

? Как отличается друг от друга число способов разбиения элементов на три различимые группы и три неразличимые группы?

Контрольные вопросы

1. Сколькими способами из двух различных подарков можно выбрать один?

2. Сколькими способами из двух различных подарков можно выбрать два?

3. Сколькими способами из трех одинаковых подарков можно выбрать два?

4. Сколькими способами двоих детей можно послать вдвоем в лес за ягодами?

5. Сколькими способами двоих детей можно послать по одному в магазин и на почту?

6. Сколькими способами двоих человек можно построить в колонну по одному?

7. Можно ли два различных подарка распределить между тремя детьми так, чтобы каждому достался хотя бы один подарок?

8. Из двух различных предметов выбирают один так, что каждый выбранный элемент фиксируется и возвращается в прежний набор элементов. Следующий предмет выбирают из данного набора. Обязательно ли он отличается от первого выбранного? Обязательно ли он совпадает с первым выбранным?

Задачи

237.° Сколько можно составить из цифр 1, 2, 3, 4:

а) двузначных чисел;

б) двузначных чисел с различными цифрами;

в) двузначных нечетных чисел;

г) двузначных нечетных чисел с различными цифрами;

д) двузначных чисел с нечетными цифрами;

е) двузначных чисел с различными нечетными цифрами;

ж) двузначных чисел, содержащих хотя бы одну четную цифру?

238.° Четыре друга собрались на футбольный матч. Им удалось купить два билета. Сколькими способами их можно распределить между четырьмя друзьями, если:

а) билеты различны и каждый может получить оба билета (например, для своей девушки);

б) билеты различны, но каждый может получить не более одного билета;

в) билеты одинаковы (например, на соседние места), но каждый может получить оба билета;

г) билеты одинаковы и каждый может получить не более одного билета?

239.° Сколькими способами можно распределить четыре карандаша:

а) между двумя детьми, если карандаши различны;

б) между двумя детьми, если карандаши одинаковы;

в) между двумя детьми, если карандаши одинаковы, и каждый ребенок должен получить хотя бы один карандаш?

240.° Четыре человека обменялись:

а) рукопожатиями; б) фотографиями.

Сколько было сделано рукопожатий? Сколько понадобилось фотографий?

241.° Сколькими способами коллектив из пяти человек может выбрать из своего состава:

а) руководителя и казначея;

б) делегацию из двух человек для возложения цветов?

242.° Сколькими способами два фломастера можно распределить между тремя детьми, если фломастеры:

а) различны; б) одинаковы?

243. Шесть туристов, среди которых двое уроженцев тех мест, где проходит туристский поход, распределились на две равные группы, в каждой из которых было по одному уроженцу этих мест. Сколькими способами они могут это сделать, если группы:

а) должны заняться поисками одного исчезнувшего туриста;

б) отправятся на экскурсии в различные населенные пункты?

244. Группа детей, среди которых два мальчика и две девочки, случайно разделилась на две равные группы. Какова вероятность того, что в каждой группе будет по одному мальчику и по одной девочке, если распределение на группы проведено с целью:

а) организации парной встречи между ними по теннису;

б) пополнения двух школьных команд по стрельбе?

§ 3.2. Правила умножения и сложения

3.2.1. Комбинаторное правило умножения

Метод перебора позволяет решать комбинаторные задачи при небольших значениях данных величин (параметров). Если же параметры принимают большие числовые или буквенные значения, то этот метод или становится громоздким, или совсем неприменим. В этих случаях часто применяют *правило умножения*, или *основное правило комбинаторики*.

Пример 1. Из города A в город B ведут пять дорог, а из B в C — три дороги. Сколько путей, проходящих через B, ведут из A в C?

❒ Обозначим дороги из A в B цифрами 1, 2, 3, 4, 5, а дороги из B в C — буквами а, б, в. Построим дерево возможных вариантов (рис. 40).

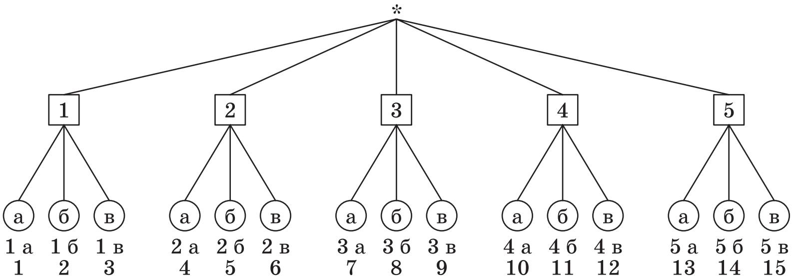


Рис. 40

На первом уровне дерева пять «узлов» (дорог из A в B). Из каждого узла выходит три ветки (дороги из B в C). Всего получилось $5 \cdot 3 = 15$ путей из A в C. Если число дорог из A в B будет равно 10, а из B в C — 8, то изобразить дерево возможных вариантов сложно. Задачу легче решить рассуждением. Из A в B можно добраться по любой из 10 дорог. По какой бы дороге мы ни прибыли в B, есть 8 путей, по которым можно добраться из B в C. Всего получим $10 \cdot 8 = 80$ различных путей из A в C через B. ■

Мы использовали так называемое ***правило умножения***. Сформулируем его:

Если объект A можно выбрать m способами и если после ***каждого*** *такого выбора объект B можно выбрать n способами, то выбор пары (A, B) в указанном порядке можно осуществить $m \cdot n$ способами.*

Можно было описать решение примера 1 иначе. Требовалось осуществить последовательно одно за другим два действия: выбор пути из A в B и выбор пути из B в C. Первое действие можно осуществить пятью способами и при любом способе его осуществления второе действие можно выполнить тремя способами. Тогда оба действия (выбор пути из A в B и выбор пути из B в C) можно осуществить $5 \cdot 3 = 15$ способами.

Обратите внимание, каким бы способом ни был выбран один объект, второй объект должен выбираться одним и тем же числом способов. Другими словами, из каждого узла одного уровня дерева должно выходить одно и то же число веток.

Правило умножения справедливо для выбора любого конечного числа объектов. В общем виде его можно сформулировать так:

Если объект A_1 может быть выбран n_1 различными способами, A_2 — n_2 различными способами и т. д., A_k — n_k различными способами, то k объектов $A_1, A_2, \ldots, A_k$ в указанном порядке можно выбрать $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Пример 2. Составляются дорожные знаки, состоящие из геометрической фигуры (круга, квадрата, треугольника или шестиугольника), буквы и цифры. Сколько таких знаков можно составить?

❒ Сначала нужно выбрать геометрическую фигуру. Этот выбор может быть совершен четырьмя способами (всего четыре фигуры). При любом выборе фигуры можно выбрать одну из 33 букв. Поэтому фигуру и букву можно выбрать $4 \cdot 33 = 132$ способами. Для каждого из этих 132 вариантов есть 10 способов выбора цифры. Всего получается $4 \cdot 33 \cdot 10 = 1320$ знаков. ■

В предыдущих примерах фактически в условиях указывалось количество способов, с помощью которых можно было осуществить тот или иной выбор. Однако так бывает не всегда, в некоторых задачах это число способов еще нужно определить.

Пример 3. В классе 25 человек. Сколькими способами:

а) можно распределить между ними два различных учебника;

б) можно распределить между ними два различных учебника так, чтобы никто не получил оба учебника;

в) можно выбрать в этом классе старосту и его заместителя?

❒ а) Первый учебник может получить любой из 25 учащихся. Кто бы ни получил первый учебник, второй может достаться снова любому из 25, ведь в условии не сказано, что каждый должен получить не более одного учебника. Всего имеем $25 \cdot 25 = 625$ способов.

б) В отличие от предыдущего задания, никто не должен получить оба учебника. Поэтому для каждого из 25 вариантов выбора обладателя первого учебника есть 24 способа выбора обладателя второго учебника. Всего имеем $25 \cdot 24 = 600$ способов.

в) Эта задача подобна предыдущей. Старостой может быть любой из 25 учащихся. После выбора старосты на роль заместителя могут претендовать 24 оставшихся учащихся. Таким образом, всего есть $25 \cdot 24 = 600$ различных вариантов выбора. ■

Этот пример отличается от предыдущих тем, что в условии фактически указано число способов выбора только первого объекта. Число способов выбора второго объекта нужно устанавливать, исходя из условия задачи. В первом случае оно совпадает с числом способов выбора первого объекта (в этом случае говорят, что совершается ***выбор с возвращением***: выбранный объект возвращается в данную совокупность и следующий объект выбирается из прежней совокупности). Во втором случае выбранный объект не возвращается в исходную совокупность: второй учебник не может достаться тому же учащемуся. Точно также в третьем случае выбор старосты сокращает круг кандидатов на роль его заместителя: выбранный старостой уже не может быть его заместителем. В этом случае говорят, что осуществляется ***выбор без возвращения***.

? Можно ли из некоторой совокупности выбрать такое количество элементов, которое превышает численность данной совокупности, если:

а) выбор осуществляется без возвращения;

б) выбор осуществляется с возвращением?

В отличие от предыдущих двух задач, где при выборе дорог или составлении дорожных знаков все действия были независимыми, выборы старосты и его заместителя не являются независимыми.

В следующем примере правило умножения будет применено в различных ситуациях.

Пример 4. Сколько можно составить из цифр 0, 1, 2, 3, 4, 5:

а) четырехзначных чисел;

б) четырехзначных чисел с различными цифрами;

в) нечетных четырехзначных чисел;

г) нечетных четырехзначных чисел с различными цифрами;

д) нечетных четырехзначных чисел, содержащих хотя бы две одинаковые цифры;

е) четных четырехзначных чисел;

ж) четырехзначных чисел, составленных из нечетных цифр;

з) четырехзначных чисел, делящихся на 4?

❒ Задача сводится к выбору четырех цифр из данных шести.

а) На первое место можно поставить любую цифру, кроме нуля, т. е. имеется пять способов ее выбора. Для любого способа выбора первой цифры на второе место может претендовать любая из данных шести цифр. Продолжая и т. д., по правилу умножения будем иметь $5 \cdot 6^3 = 1080$ четырехзначных чисел.

? Какой осуществляется выбор: с возвращением или без возвращения?

б) Первую цифру, как и в предыдущей задаче, можно выбрать пятью способами. При любом способе ее выбора на второе место можно поставить одну из пяти цифр (кроме цифры, стоящей на первом месте). Две следующие цифры можно выбрать соответственно четырьмя и тремя способами. Итак, количество четырехзначных чисел с различными цифрами равно $5 \cdot 5 \cdot 4 \cdot 3 = 300$.

? Какой осуществляется выбор: с возвращением или без возвращения?

в) Четыре цифры, начиная с первой, можно выбрать соответственно пятью, шестью, шестью и тремя способами (последней цифрой может быть одна из трех нечетных цифр 1, 3, 5). Таким образом, имеем $5 \cdot 6 \cdot 6 \cdot 3 = 540$ нечетных четырехзначных чисел.

? Какой осуществляется выбор: с возвращением или без возвращения?

г) Выбор начинаем с последней цифры, в противном случае после выбора первых трех цифр может не остаться нечетной цифры. Для нее есть три возможности: 1, 3, 5. Первой цифрой может быть одна из четырех цифр (любая из данных, кроме нуля и цифры, стоящей на последнем месте). Вторую и третью цифру можно выбрать соответственно четырьмя и тремя способами. Всего получим $3 \cdot 4 \cdot 4 \cdot 3 = 144$ нечетных четырехзначных числа с различными цифрами.

д) Чтобы подсчитать количество нечетных четырехзначных чисел, содержащих по крайней мере две одинаковые цифры, можно от общего количества нечетных четырехзначных чисел (см. задание в) отнять количество нечетных четырехзначных чисел с различными цифрами (см. задание г): $540 - 144 = 396$. Мы фактически воспользовались так называемым ***правилом дополнения***:

Чтобы найти количество элементов некоторой совокупности, удовлетворяющих определенному условию, можно из общего количества элементов этой совокупности вычесть количество тех ее элементов, которые не удовлетворяют этому условию.

Задание д) можно выполнить и иначе, не пользуясь правилом дополнения. При этом придется рассмотреть много возможных случаев.

е) Аналогично заданию в) имеем $5 \cdot 6 \cdot 6 \cdot 3 = 540$ четных четырехзначных чисел. Можно было воспользоваться правилом дополнения — от общего количества четырехзначных чисел отнять количество нечетных четырехзначных чисел: $1080 - 540 = 540$.

ж) Воспользовавшись правилом умножения и учтя, что среди данных цифр есть три нечетные, получим $3 \cdot 3 \cdot 3 \cdot 3 = 81$ четырехзначное число с нечетными цифрами.

? Чему равно количество четырехзначных чисел, образованных из данных цифр и содержащих хотя бы одну четную цифру?

з) Воспользуемся признаком делимости на 4: *на 4 делятся те и только те числа, в которых число, образованное двумя последними цифрами* (*в том же порядке*), *делится на* 4. Таких двузначных чисел, образованных из данных цифр и делящихся на 4, девять: 00, 04, 12, 20, 24, 32, 40, 44, 52 (мы воспользовались методом перебора). Для каждого из них первую цифру можно выбрать пятью способами, вторую — шестью. Итак, имеем $9 \cdot 5 \cdot 6 = 270$ четырехзначных чисел, делящихся на 4.

Обратите внимание на то, что из 540 четырехзначных четных чисел каждое второе делится на 4. ■

3.2.2. Комбинаторное правило сложения

Выше мы отмечали, что задание 4, д) можно выполнить, не прибегая к правилу дополнения, рассмотрев несколько возможных случаев. Это связано с применением так называемого *комбинаторного правила сложения*. Чтобы получить его, вернемся к примеру 1.

Пример 5. Пусть по-прежнему из города A в город B ведут пять дорог, а из города B в город C — три дороги. Пусть, кроме того, из города A в город D можно попасть двумя путями, из D в C — четырьмя (рис. 41). Сколькими способами можно добраться из A в C?

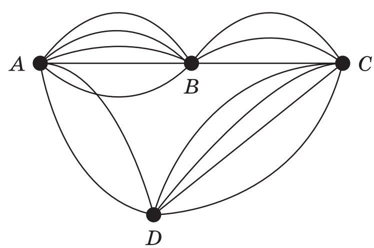


Рис. 41

❐ Возможны два случая: путь из A в C проходит через город B или через город D. В каждом из этих случаев число возможных маршрутов легко подсчитать, воспользовавшись правилом умножения. В первом случае имеется $5 \cdot 3 = 15$ маршрутов; во втором — $2 \cdot 4 = 8$. Складывая, получаем общее число маршрутов: $15 + 8 = 23$. ■

В задаче все рассматриваемые варианты разбиты на два класса, причем каждый вариант входит в один и только в один класс. В этом случае общее число вариантов равно сумме количеств вариантов в обоих классах. Это утверждение называют ***правилом сложения***. Его можно сформулировать так:

Если некоторый объект A можно выбрать m способами, а другой объект B можно выбрать n способами, причем ни один из способов выбора A не совпадает с каким-нибудь способом выбора B, то выбор «либо A, либо B» можно осуществить $m + n$ способами.

? Как, воспользовавшись правилом сложения, найти:

а) сколькими способами из набора монет можно выбрать некоторое количество одноцветных монет;

б) сколькими способами из набора шаров можно выбрать шары в количестве, не превышающем заданного числа?

Вернемся к условию примера 4.

Пример 6. Сколько можно составить из цифр 0, 1, 2, 3, 4, 5:

а) четных четырехзначных чисел с различными цифрами;

б) четырехзначных чисел, делящихся на 4, с различными цифрами;

в)* нечетных четырехзначных чисел, содержащих хотя бы две одинаковые цифры?

❒ а) Возможны два случая: число заканчивается нулем или цифрой, отличной от нуля. Количество четных четырехзначных чисел с различными цифрами, оканчивающихся нулем, равно $5 \cdot 4 \cdot 3 = 60$, а оканчивающихся цифрой, отличной от нуля, — $2 \cdot 4 \cdot 4 \cdot 3 = 96$. Всего $60 + 96 = 156$ вариантов.

б) Для двух последних цифр есть семь вариантов (см. пример 4, з): 04, 12, 20, 24, 32, 40, 52. Из них в трех случаях имеется нуль, в четырех — нет нуля. Искомое количество чисел равно $3 \cdot 4 \cdot 3 + 4 \cdot 3 \cdot 3 = 72$.

в)* С помощью правила дополнения мы получили, что количество нечетных четырехзначных чисел, содержащих хотя бы две одинаковые цифры, равно 396.

➤ Однако это задание можно выполнить используя правило сложения. Искомая совокупность чисел состоит из:

1) нечетных трехзначных чисел, не содержащих в записи нуль, у которых одна из двух одинаковых цифр стоит на последнем месте (количество таких чисел равно $3 \cdot 3 \cdot 4 \cdot 3 = 108$);

2) нечетных трехзначных чисел, не содержащих в записи нуль, у которых одна из двух одинаковых цифр не стоит на последнем месте (количество таких чисел равно $3 \cdot 4 \cdot 3 \cdot 3 = 108$);

3) нечетных трехзначных чисел, у которых одна из двух одинаковых цифр есть нуль (количество таких чисел равно $3 \cdot 1 \cdot 1 \cdot 4 = 12$);

4) нечетных трехзначных чисел, содержащих в записи нуль, у которых одна из двух одинаковых цифр не стоит на последнем месте (количество таких чисел равно $3 \cdot 4 \cdot 2 = 24$);

5) нечетных трехзначных чисел, не содержащих в записи нуль, у которых одна из трех одинаковых цифр стоит на последнем месте (количество таких чисел равно $3 \cdot 3 \cdot 4 = 36$);

6) нечетных трехзначных чисел, не содержащих в записи нуль, у которых одна из трех одинаковых цифр не стоит на последнем месте (количество таких чисел равно $3 \cdot 4 \cdot 1 = 12$);

7) нечетных трехзначных чисел, содержащих в записи нуль, у которых одна из двух одинаковых цифр стоит на последнем месте (количество таких чисел равно $3 \cdot 3 \cdot 4 = 36$);

8) нечетных трехзначных чисел, содержащих в записи нуль, у которых одна из трех одинаковых цифр стоит на последнем месте (количество таких чисел равно $3 \cdot 2 = 6$);

9) нечетных трехзначных чисел, не содержащих в записи нуль, у которых две пары одинаковых цифр (количество таких чисел равно $3 \cdot 3 \cdot 4 = 36$);

10) нечетных трехзначных чисел, не содержащих в записи нуль, у которых одна из трех одинаковых цифр не стоит на последнем месте (количество таких чисел равно $3 \cdot 4 = 12$);

11) нечетных трехзначных чисел, содержащих в записи нуль, у которых две пары одинаковых цифр (количество таких чисел равно 3);

12) нечетных трехзначных чисел, у которых четыре одинаковые цифры (количество таких чисел равно 3).

Общее количество нечетных трехзначных чисел, содержащих хотя бы две одинаковые цифры, равно $108 + 108 + 12 + 24 + 36 + 12 + 36 + 6 + 36 + 12 + 3 + 3 = 396$.

Вы убедились, как усложняется решение без использования правила дополнения. ■ ◄

При использовании правила сложения нужно следить, чтобы ни один из способов выбора объекта A не совпадал с каким-нибудь способом выбора объекта B, т. е. чтобы ни одна комбинация не попала сразу в два класса. Если такие совпадения имеются, правило сложения в ранее сформулированной форме утрачивает силу, и мы получаем $m + n - k$ способов выбора, где k — число совпадений.

Пример 7. В классе каждый ученик знает хотя бы один иностранный язык: английский или немецкий. 25 человек знают английский, 10 учащихся — немецкий, а пятеро знают оба языка. Сколько учеников в классе? Сколько из них знают только немецкий язык?

❒ В классе 25 человек знают английский язык, 10 человек — немецкий. При этом пятеро из них знают оба языка, т. е. они вошли и в число тех, кто знает английский, и в число тех, кто знает немецкий язык. Другими словами, если сложим $25 + 10$, мы их учтем дважды (сравните с выводом теоремы сложения вероятностей в § 2.8). Поэтому в классе $25 + 10 - 5 = 30$ учеников. Немецкий язык знают 10 человек, из них и английский знают пятеро. Значит, только немецкий знают $10 - 5 = 5$ человек. ■

Мы воспользовались так называемым ***общим правилом сложения***. Его можно сформулировать следующим образом:

Если объект A можно выбрать m способами, а объект B — n способами, причем при k способах одновременно выбираются и A и B, то выбор «A или B» можно осуществить $m + n - k$ способами.

Пример 8. Сколько чисел в первой сотне, не делящихся ни на 2, ни на 3?

❒ Легче вычислить сначала количество чисел первой сотни, *делящихся* на 2 или на 3. Каждое второе число в натуральном ряде делится на 2, каждое третье — на 3. Поэтому в первой сотне есть 50 чисел, делящихся на 2, и 33 числа (неполное частное от деления 100 на 3), делящихся на 3. Но среди первых и вторых имеются числа, делящиеся и на 2, и на 3, т. е. делящиеся на 6. На 6 делится каждое шестое число в натуральном ряде. Если 100 разделить на 6, то неполное частное будет равняться 16, т. е. 16 чисел в первой сотне делится на 6. Итак, количество чисел в первой сотне, делящихся на 2 или на 3, равно $50 + 33 - 16 = 67$. Все остальные не делятся ни на 2, ни на 3. Этих чисел $100 - 67 = 33$. ■

3.2.3. Перестановки

Рассмотрим задачи, в которых приходится подсчитывать число способов, с помощью которых n различных предметов можно расположить на n местах. Такие расположения называют ***перестановками***.

Пример 9. Сколькими способами можно расставить на пятиместной полке пять различных книг?

❒ Будем рассуждать так же, как и при решении предыдущих задач. На первое место можно поставить любую из пяти книг, на второе место — любую из четырех оставшихся, на третье — любую из трех оставшихся и т. д. Таким образом, всего получается $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ способов. ■

Произведение $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$ является произведением пяти первых натуральных чисел, его обычно обозначают $5!$ и читают «5-факториал». И вообще произведение первых n ($n > 1$) натуральных чисел $1 \cdot 2 \cdot 3 \cdot ... \cdot n$ обозначают $n!$ и читают «n-факториал». Ясно, что $2! = 1 \cdot 2 = 2$; $3! = 1 \cdot 2 \cdot 3 = 6$; $4! = 24$.

?

Чему равно: а) $20! \cdot 21$; б) $n! \cdot (n + 1)$?

Чему равно: а) $\frac{50!}{48!}$; б) $\frac{n!}{(n-1)!}$?

Так как $n! = (n - 1)! \cdot n$ для всех натуральных чисел $n > 1$, то чтобы это равенство оставалось справедливым и для $n = 1$, полагают $1! = 1$, $0! = 1$.

Аналогично предыдущей задаче получаем, что n различных предметов можно расположить в ряд $n(n-1)(n-2) \cdot \ldots \cdot 2 \cdot 1$ различными способами, т. е. *число перестановок из n элементов равно n!*

Пример 10. Сколько перестановок можно составить из букв слова: а) «линейка»; б) «тетрадь»; в) «перешеек»; г) «калькулятор»?

❒ а) В слове «линейка» семь букв, все они различны. Поэтому число перестановок равно $7! = 5040$.

б) В слове «тетрадь» тоже семь букв. Если бы все они были различны, то из них можно было бы составить $7!$ перестановок. Однако те перестановки, которые получаются только переменой местами букв «т» и «т», на самом деле одинаковы. Таким образом, полученные $7!$ перестановок разбиваются на пары одинаковых. Поэтому число различных перестановок равно $\frac{7!}{2} = 2520$.

в) В слове «перешеек» восемь букв. Если бы они были различны, то из них можно было получить $8!$ перестановок. Поскольку четыре буквы «е» можно переставлять $4!$ способами, то все $8!$ перестановок разбиваются на группы по $4!$ одинаковых. Поэтому различных перестановок всего $\frac{8!}{4!} = 1680$.

г) В слове «калькулятор» имеем 11 букв, среди которых две пары одинаковых. Отождествляя перестановки, отличающиеся лишь перестановкой буквы «к», но не «л», получим $\frac{11!}{2!}$ различных перестановок. Отождествляя теперь перестановки, отличающиеся переменой мест букв «л», получим $\frac{11!}{2! \cdot 2!}$. ■

Рассмотренный метод можно применить к решению задач, где явно не идет речь о подсчете числа перестановок.

Пример 11. Сколькими способами можно расставить белые фигуры (два коня, два слона, две ладьи, ферзя и короля) на первой линии шахматной доски?

❐ Если бы все фигуры были различны, это можно было бы сделать 8! способами. Так как способы, получающиеся перестановкой двух коней между собой, двух слонов между собой и двух ладей между собой, можно отождествить, то число способов будет равно $\frac{8!}{2! \cdot 2! \cdot 2!} = 5040$. ■

В некоторых задачах удобно объединять несколько элементов и рассматривать их в качестве одного элемента.

Пример 12. Сколькими способами можно переставить буквы слова «задача» так, чтобы три буквы «а» шли подряд?

❐ Чтобы три буквы «а» шли подряд, их нужно рассматривать как один элемент и переставлять только соединенными между собой. «Свяжем» три буквы «а», тогда искомое число способов равно числу перестановок из четырех элементов, т. е. $4! = 24$. ■

Обобщением предыдущего примера служит следующий.

Пример 13. Сколько существует перестановок из n ($n \geqslant 2$) различных элементов, в которых данные два элемента не стоят рядом?

❐ Число всех перестановок из n элементов равно $n!$ Вычтем из него число тех перестановок, в которых данные два элемента стоят рядом. Для подсчета этого числа объединим эти два элемента в один. Полученные $n - 1$ элементов можно переставить $(n - 1)!$ способами. Однако в каждой из этих перестановок данные элементы можно поменять местами, т. е. всего имеется $2(n - 1)!$ перестановок, в которых данные два элемента стоят рядом. Искомое число перестановок равно $n! - 2(n - 1)! = n(n - 1)! - 2(n - 1)! = (n - 1)!\,(n - 2)$. ■

3.2.4. Зависит ли результат выбора от порядка следования элементов?

Ответ на вопрос, поставленный в заголовке, попытаемся найти при решении следующей задачи.

Пример 14. Сколькими способами в классе из 25 человек можно избрать:

а) старосту, заместителя, казначея;

б) трех членов культурно-массовой комиссии?

❐ а) Подобную задачу мы уже решали, используя правило умножения. Старосту можно выбрать 25 способами, на роль его заместителя могут претендовать 24 человека, а казначеем может стать любой из 23 оставшихся человек. Всего, по правилу умножения, имеем $25 \cdot 24 \cdot 23 = 13\,800$ способов. Заметим, что если учащихся закодируем номерами 1, 2, 3, ..., 25, то результаты выбора имеют, например, вид (1, 2, 3), (5, 7, 23), и т. д. или вообще (a_1, a_2, a_3), где a_1, a_2, a_3 принимают значения от 1 до 25, причем все числа a_1, a_2, a_3 различны. Запись (1, 2, 3) означает, что старостой выбран учащийся № 1, заместителем — № 2, казначеем — № 3. Выбор (2, 1, 3) отличается от предыдущего, так как теперь старостой является № 2, а его заместителем — № 1.

б) Эта задача внешне похожа на предыдущую. Казалось бы, тоже можно было бы рассуждать так: первого члена комиссии можно выбрать 25 способами, при любом способе его выбора второго можно выбрать 24 способами, третьего — 23 способами. Однако при этом результатами выбора является и набор (1, 2, 3), и набор (2, 1, 3), а вместе с тем, это один и тот же состав комиссии. Возникает вопрос: сколько раз мы учли каждый состав комиссии? Очевидно, столько, сколько перестановок можно составить из трех элементов, т. е. $3! = 6$. Чтобы исключить одинаковые способы, нужно произведение $25 \cdot 24 \cdot 23$ разделить на 3! Получим $\dfrac{25 \cdot 24 \cdot 23}{3!} = 2300$ способов. ■

При решении подобных задач не забывайте определять, зависят ли результаты от порядка следования элементов или нет. Для этого можно закодировать все элементы, рассмотреть результаты выбора, состоящие из одних и тех же элементов, но отличающиеся порядком следования элементов, и выяснить, описывают они один и тот же результат выбора или различные. В первом случае необходимо число способов выбора разделить на число перестановок из

k элементов, где k — число отбираемых элементов. В этом случае результаты выбора называют ***неупорядоченными выборками*** (результат не зависит от порядка следования элементов), в противном случае — ***упорядоченными выборками*** (результат зависит от порядка следования элементов).

Пример 15. Сколькими способами из колоды в 36 карт можно выбрать:

а) четыре карты;

б) четыре карты разных мастей и достоинств;

в) такие четыре карты так, чтобы среди них были ровно две бубновой масти?

❐ а) Карты можно пронумеровать числами 1, 2, 3, ... , 36. Первую карту можно выбрать 36 способами; вторую — 35; третью — 34, четвертую — 33. Однако результаты выбора (1, 2, 3, 4) и (4, 2, 1, 3) совпадают. Поэтому имеем $\frac{36 \cdot 35 \cdot 34 \cdot 33}{4!} = 58\,905$ способов.

? Результаты выбора являются упорядоченными или неупорядоченными выборками?

б) *1-й способ*. Первую карту можно выбрать 36 способами; при любом ее выборе для второй остаются 24 возможности (исключаются девять карт той же масти, три карты того же достоинства, что и извлеченная карта). Для третьей остаются 14 вариантов (исключаем восемь карт той же масти и две карты того же достоинства, которые имела вторая карта). Для четвертой карты остаются шесть вариантов (еще нужно исключить семь карт той же масти и одну карту того же достоинства, которые имела третья карта). По правилу умножения имеем $36 \cdot 24 \cdot 14 \cdot 6$. При этом результаты выбора не меняются от перестановки элементов в наборе, например (1, 2, 3, 4). Окончательно имеем $\frac{36 \cdot 24 \cdot 14 \cdot 6}{4!} = \frac{36}{4} \cdot \frac{24}{3} \cdot \frac{14}{2} \cdot \frac{6}{1} = 9 \cdot 8 \cdot 7 \cdot 6 =$ $= 3024$.

2-й способ. Выбор будем производить следующим образом. Колоду разделим на четыре части, каждая из которых содержит карты одной масти. Первую карту извлечем из первой части, это можно сделать девятью способами; вторую — из второй части. Это можно сделать восемью способами (нельзя брать карту того же достоинства, что и первая). Аналогично для третьей карты остается семь вариантов, а для четвертой — шесть. Фактически мы зафиксировали порядок частей колоды, из которых производили последовательный выбор карт. При изменении порядка меняется количество возможностей выбора из каждой части. Поэтому результаты выбора при изменении порядка следования элементов меняются. Искомое число способов равно $9 \cdot 8 \cdot 7 \cdot 6 = 3024$.

? Результаты выбора (при втором способе) являются упорядоченными или неупорядоченными выборками?

в) Сначала извлечем две карты из девяти бубновой масти. Это можно сделать $\frac{9 \cdot 8}{1 \cdot 2} = 36$ способами. Осталось извлечь еще две карты из оставшихся 27. Это можно сделать $\frac{27 \cdot 26}{1 \cdot 2} = 351$ способом. Обратите внимание: вначале выбрали две карты бубновой масти (первый объект), затем еще две не бубновой масти (второй объект). По правилу умножения имеем $36 \cdot 351 = 12\,636$ способов. ■

Пример 16. Сколькими способами из десяти различных карандашей можно выбрать семь?

❒ Решая аналогично предыдущим задачам, получим

$$\frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{7!} = 120 \text{ способов.}$$

Однако эту задачу можно решить и иначе. При любом способе выбора семи карандашей остается $10 - 7 = 3$ карандаша. Поэтому вместо того, чтобы выбирать для каких-то целей семь карандашей, можно выбрать три, остающиеся после выбора семи. Тем самым однозначно определятся и семь карандашей, которые необходимо было выбрать. Таким образом, число способов выбора семи карандашей из десяти различных равно числу способов выбора трех карандашей из десяти различных, т. е. $\frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} = 120$. И вообще, число способов выбора k ($k < n$) элементов из n различных, при которых результат выбора не меняется при изменении порядка следования элементов, равно числу способов выбора $n - k$ элементов из n различных. ■

Этот прием удобно применять, если требуется подсчитать число способов выбора из некоторой совокупности различных элементов больше половины ее элементов.

3.2.5. Распределение n одинаковых предметов по m ячейкам

В большинстве ранее рассмотренных задач элементы, из которых осуществлялся выбор, или которые распределялись по различным ячейкам, были различны. Решение задачи существенно меняется, если элементы считать одинаковыми. Например, из шести различных элементов два элемента можно выбрать $\frac{6 \cdot 5}{2} = 15$ способами, а из шести одинаковых — только одним (выбранные два элемента ничем не будут отличаться от другой пары). Для сравнения рассмотрим следующую задачу.

Пример 17. Сколькими способами шесть карандашей можно распределить между тремя детьми, если карандаши:

а) разные; б) одинаковые?

❒ а) Первый карандаш может получить любой из трех детей, второй также любой из трех и т. д. до шестого карандаша. По правилу умножения имеем

$$3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^6 = 729 \text{ способов.}$$

б) Если карандаши одинаковые, то различные способы распределения карандашей между детьми будут отличаться лишь количеством карандашей у детей. Если детей мы закодируем цифрами 1, 2, 3, то возможны, например, такие способы распределения: (1 1 1 2 2 3), (1 1 1 1 2 2), (3 3 3 3 3 3) и т. д. Первая запись означает, что первому ребенку досталось три карандаша (их порядок несущественен), второму — два, третьему — один. Во втором из приведенных вариантов первый ребенок получил четыре карандаша, второй — два. В третьем все шесть карандашей получил третий ребенок. Укажем способ подсчета числа вариантов такого распределения. Изобразим карандаши в виде точек. Чтобы распределить их между тремя детьми, поставим две перегородки (на одну меньше, чем детей). Для трех вышеприведенных примеров распределений имеем следующее положение перегородок:

$$...|..|.; \quad ....|..|; \quad ||......$$

Левее первой перегородки отмечены карандаши, доставшиеся первому ребенку, между первой и второй — второму, правее второй — третьему. Число способов распределения шести одинаковых карандашей между тремя детьми совпадает с числом способов выбора из восьми мест (шесть точек + две перегородки) двух мест для перегородок. Так как от перестановки перегородок результат не меняется, то это число способов равно $\frac{8 \cdot 7}{2} = 28$. ■

Каждое распределение карандашей между детьми можно рассматривать как выбор детей для вручения им карандашей. Для каждого из рассмотренных случаев определите:

а) этот выбор является выбором с возвращением или без возвращения?

б) результаты выбора являются упорядоченными или неупорядоченными?

Пример 18. Трое ребят собрали с яблони 40 яблок. Яблоки считаются одинаковыми. Сколькими способами их можно разделить:

а) между тремя сборщиками;

б) между тремя сборщиками так, чтобы каждому досталось хотя бы одно яблоко;

в) между тремя сборщиками так, чтобы каждому досталось по крайней мере по 10 яблок?

❒ а) Аналогично предыдущему примеру число способов распределения 40 одинаковых яблок между тремя сборщиками равно числу способов выбора из 42 мест (40 точек + 2 перегородки) двух мест для перегородок, т. е. $\frac{42 \cdot 41}{2} = 861$.

б) Начнем с того, что каждому дадим по одному яблоку. Так как яблоки одинаковые, то это сделать можно одним способом. Оставшиеся $40 - 3 = 37$ яблок нужно распределить произвольным образом между тремя сборщиками. Число способов такого распределения равно числу способов выбора из 39 мест $(37 + 2)$ двух мест для перегородок, т. е. $\frac{39 \cdot 38}{2} = 741$.

в) Опять сначала нужно каждому из трех дать по 10 яблок. Это можно осуществить единственным способом. Оставшиеся $40 - 30 = 10$ яблок нужно распределить между тремя сборщиками произвольным образом. Число способов такого распределения равно $\frac{12 \cdot 11}{2} = 66$. ■

Заметим, что в примерах 17, б) и 18 по существу шла речь о распределении n одинаковых предметов в m ячейках. В примере 17, б) шесть одинаковых карандашей (предметы) распределялись между тремя детьми (ячейки). В примере 18 сорок яблок (предметы) распределялись между тремя сборщиками (ячейки).

Рассмотренные примеры подсказывают метод решения задачи в общем виде.

➤

Пример 19. Сколькими способами n одинаковых предметов можно распределить по k ячейкам?

❒ n одинаковых предметов изображаем точками, между которыми ставим $k - 1$ перегородок. (Обратите внимание: предметы изображаем точками, а между перегородками — содержимое ячеек.) Число способов распределения n одинаковых предметов по k ячейкам равно числу способов выбора из $(n + k - 1)$ мест $(k - 1)$-го места для перегородок. Первое место можно выбрать $(n + k - 1)$ способом, второе — $(n + k - 2)$ способами и т. д., $(k - 1)$-е — $(n + 1)$ способом. После выбора $(k - 2)$ мест останется $n + k - 1 - (k - 2) = n + 1$ место. По правилу умножения получим $(n + k - 1)(n + k - 2) \ldots (n + 1)$ способов. Но не все они различны. Каждый способ повторяется столько раз, сколько перестановок можно составить из $(k - 1)$ элементов, т. е. $(k - 1)!$ раз. Число повторений для каждого способа равно $(k - 1)!$

Искомое число способов равно $\frac{(n+k-1)(n+k-2)\ldots(n+1)}{(k-1)!}$. ■ ➤

Пример 20. Сколько решений имеет уравнение $x_1 + x_2 + x_3 + x_4 = 9$:

а) в целых неотрицательных числах;

б) в целых положительных числах?

❒ а) Неизвестные x_1, x_2, x_3, x_4 могут принимать значения 0, 1, 2, ..., 9. Напрашивается мысль: записать решение уравнения в виде (A_1, A_2, A_3, A_4), где A_j, $j = 1, 2, 3, 4$, принимают значения от 0 до 9. Но если набор (1, 2, 3, 3) является решением, так как $1 + 2 + 3 + 3 = 9$, то набор (1, 2, 3, 4) уже им не является: сумма чисел, стоящих в скобках, не равна 9. Поэтому число искомых решений нельзя определить, подсчитав число способов выбора с возвращением из 10 элементов четырех. Будем искать другой путь. Закодируем неизвестные числами 1, 2, 3, 4 и тогда сможем записать примеры решений в виде (1 1 2 2 2 3 3 4 4), (1 1 1 1 1 4 4 4 4), (2 2 2 2 2 2 2 2 2) и т. д. Эти записи соответственно означают: $x_1 = 2$, $x_2 = 3$, $x_3 = 2$, $x_4 = 2$; $x_1 = 5$, $x_2 = x_3 = 0$, $x_4 = 4$; $x_1 = x_3 = x_4 = 0$, $x_2 = 9$. Здесь роль ячеек играют неизвестные, правая часть указывает на число одинаковых предметов. Решение можно изобразить также в виде девяти точек (см. правую часть уравнения) с тремя перегородками (4 неизвестных – 1 = 3):

$$..|...|..|..; \quad .....|||....; \quad |.........||$$

Число решений равно числу способов распределения девяти точек по четырем ячейкам, т. е. $\frac{12 \cdot 11 \cdot 10}{3!} = 220$.

б) Эта задание отличается от предыдущего тем, что ни одно из неизвестных не может принимать значение, равное нулю. Другими словами, в каждой из четырех ячеек должно быть хотя бы по одному предмету. Для этого понадобится четыре предмета. Остается пять предметов распределить по четырем ячейкам произвольным образом. Число способов такого распределения равно числу способов выбора из $5 + 3 = 8$ мест трех мест для перегородок, т. е. $\frac{8 \cdot 7 \cdot 6}{3!} = 56$. ■

Контрольные вопросы

1. Сколько существует двузначных чисел, в которых обе цифры четные?

2. Сколько существует двузначных чисел, в которых обе цифры имеют различную четность?

3. У одного ученика семь книг по математике, у второго — девять детективов. Сколькими способами можно обменять одну книгу первого ученика на одну книгу второго?

4. Сколькими способами можно отправить четыре срочных письма, если для этого использовать трех курьеров и каждое письмо можно дать любому из курьеров?

5. В магазине продается пять разных видов вилок, три разных типов ножей и четыре разных видов ложек. Сколькими способами можно купить в этом магазине комплект из вилки, ножа и ложки?

6. Сколькими способами шесть различных карандашей можно разделить между двумя детьми?

Задачи

245.° В классе 25 человек. Каждый день назначают одного дежурного. Сколькими способами можно составить расписание на пять дней так, чтобы никто не дежурил более одного раза?

246. Четверо учащихся Олег, Андрей, Сергей и Руслан за упражнение на гимнастическом снаряде могут получить 2, 3, 4, 5 баллов.

а)° Сколькими способами можно оценить их выступление на этом снаряде?

б)° Сколькими способами можно оценить их выступление так, чтобы никакие два ученика не получили одинаковое число баллов?

в) Сколькими способами можно оценить их выступление так, чтобы все получили 4 или 5 баллов?

г) Сколькими способами можно оценить их выступление так, чтобы никакие два ученика не получили одинаковое число баллов и Олег получил более высокую оценку, чем Андрей?

247. Алфавит племени Пинг-Понг состоит из пяти букв П, И, Н, Г, О. Словом является любая последовательность, состоящая не более чем из четырех букв. Сколько слов в языке племени Пинг-Понг?

248. Сколькими способами можно поставить на шахматную доску белого и черного королей так, чтобы получилась допустимая правилами игры позиция?

249. Сколько существует шестизначных чисел, все цифры которых имеют одинаковую четность?

250. Футбольная команда определяется составом игроков и ролью, которую играет в команде каждый отдельный игрок. Сколькими способами тренер может определить стартовый состав, если в его распоряжении 13 футболистов (общее число игроков 11) и:

а) каждый из этих футболистов может занимать в команде любое место;

б) двое игроков могут играть только вратарями?

251.° Сколько четырехбуквенных слов, состоящих только из согласных или только из гласных, можно составить из букв слова «карниз»?

252. В колледже студенты изучают математику или экономику. 200 студентов изучают математику, 150 — экономику. Сколько студентов в колледже, если:

а) 20 студентов в колледже изучают оба предмета;

б) ни один из студентов-математиков не изучает экономику;

в) каждый студент-экономист изучает математику?

253.° В лифт восьмиэтажного дома на первом этаже вошли пять человек. Любой из них с одинаковой вероятностью может выйти на любом этаже, начиная со второго. Найдите вероятность того, что все пятеро выйдут:

а) на третьем этаже;

б) на одном этаже;

в) на разных этажах.

254. В соревнованиях по гимнастике участвует 10 человек. Трое судей должны независимо друг от друга расположить их в порядке, отражающем их успехи в соревновании по мнению судей. Победителем считается тот, кого назовут первым хотя бы двое судей.

а) Сколькими способами трое судей могут назвать гимнаста, занявшего, по их мнению, первое место?

б) При скольких способах победитель не будет определен?

в) При скольких способах победитель будет определен?

255. Сколькими способами можно переставить буквы слова:

а) «перешеек» так, чтобы четыре буквы «е» не шли подряд;

б) «огород» так, чтобы три буквы «о» не стояли рядом?

256. Каких семизначных чисел больше: тех, в записи которых есть 1, или остальных?

257. Сколькими способами можно поселить семь студентов в три комнаты: одноместную, двухместную и четырехместную?

258. Футбольная команда школы должна сыграть за сезон шесть игр с командами других школ. Сколькими различными способами может пройти сезон для этой команды, если в результате она выиграет два матча, три проиграет и один сведет вничью?

259. Сколькими способами группу из восьми человек можно расположить за круглым столом так, чтобы два определенных лица оказались:

а) сидящими рядом;

б) сидящими не рядом?

Два положения считаются одинаковыми, если у каждого человека совпадают соседи слева и справа.

260. В спортивном клубе 20 человек. Сколькими способами из них можно выбрать команду из четырех человек:

а) для участия в беге на 100 м;

б) для участия в эстафете 100 + 200 + 400 + 800 м;

в) для участия в беге на 100 м и кроме них одного капитана?

261.° Учащемуся надо выбрать два факультативных курса из шести возможных.

а) Сколькими способами он может это сделать?

б) Сколько имеется способов выбора, если чтение каких-то двух курсов совпадает по времени?

в) Сколько имеется способов выбора, если чтение двух курсов начинается в 10 часов, чтение двух других — в 12 часов, а остальные курсы не пересекаются по времени?

262. Сколькими способами из 28 костей домино можно выбрать:

а) две кости;

б)* две кости так, чтобы их можно было приложить друг к другу?

263. Рассматривается спортивная лотерея «5 из 36». Сколькими способами можно отобрать из 36 видов спорта:

а) пять видов спорта;

б) пять видов спорта так, чтобы все они были выигрышными (таких 5);

в) пять видов спорта так, чтобы среди них не было ни одного выигрышного;

г) пять видов спорта так, чтобы среди них было ровно три выигрышных;

д) пять видов спорта так, чтобы среди них было не менее трех выигрышных?

264. Рассмотрим прямоугольную сеть квадратов («шахматный город», состоящий из 10×8 квадратов, которые разделены семью горизонтальными и девятью вертикальными улицами; рис. 42). Сколько различных кратчайших путей ведут из левого нижнего угла (точка (0; 0)) в правый верхний угол (точка (10; 8))?

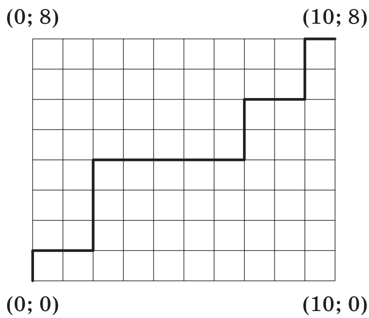


Рис. 42

265.* Сколько существует четырехзначных чисел, у которых каждая следующая цифра:

а) больше предыдущей;

б) меньше предыдущей?

266. Имеется домино с числом очков на костях от 0 до 8. Сколько таких костей?

267. Сколькими способами 12 экземпляров одной книги можно разделить между:

а) тремя призерами олимпиады;

б) тремя призерами олимпиады так, чтобы каждому досталась хотя бы одна книга;

в) тремя призерами олимпиады так, чтобы каждому досталось по крайней мере по две книги?

268. Сколько решений имеет уравнение $x_1 + x_2 + x_3 = 8$:

а) в целых неотрицательных числах;

б) в целых положительных числах?

269. Сколько решений в целых неотрицательных числах имеет неравенство $x_1 + x_2 + x_3 \leqslant 5$?

§ 3.3. Основные комбинаторные схемы

В предыдущем параграфе были рассмотрены некоторые общие правила (правила умножения, сложения, дополнения) решения комбинаторных задач. С их помощью, как мы видели, можно решать самые разнообразные комбинаторные задачи. Однако как и в геометрии неудобно сводить решение задач и доказательство теорем к аксиомам, а удобнее пользоваться теоремами, так и в комбинаторике вместо решения задач по общим правилам часто удобнее пользоваться готовыми формулами. Их выводу и применению посвящен настоящий параграф.

Основываясь на основных комбинаторных правилах, рассмотрим важные комбинаторные схемы.

Имеем случайный опыт, который состоит в том, что из урны с n пронумерованными шарами последовательно один за другим наугад вынимают r шаров. Исходы этого опыта будем называть ***выборками из n элементов по r***. Любая из них имеет вид $(a_1, a_2, \ldots, a_r)$, где a_i — номер шара, вынутого на i-м шаге, $i = 1, 2, \ldots, r$. Как мы уже отмечали в предыдущем параграфе, различают ***выборки с возвращением и без возвращения***. В первом случае на каждом шаге вынутый шар возвращают в урну, т. е. каждый элемент a_i может принимать любое из n значений $1, 2, \ldots, n$. Если выбор осуществляется без возвращения, т. е. извлеченный шар не возвращают в урну ни на одном шаге, то все элементы выборки $(a_1, a_2, \ldots, a_r)$ различны.

Если выборки, состоящие из одних и тех же элементов, но отличающиеся порядком следования элементов, характеризуют различные исходы опыта, то эти выборки называют ***упорядоченными***. В противном случае — ***неупорядоченными***. Например, если собрание из 20 человек наугад избирает президиум в составе трех человек, то исходами этого опыта являются неупорядоченные выборки без возвращения из 20 элементов по три: например, выборки (1, 2, 3) и (2, 3, 1) определяют один и тот же состав президиума. Если это собрание должно избрать председателя собрания, секретаря и члена президиума, то выборки (1, 2, 3) и (2, 3, 1) характеризуют различные составы президиума: в первом случае председателем избран участник собрания № 1, во втором — № 2. Здесь исходами опыта являются упорядоченные выборки без возвращения из 20 элементов по три.

Четырехзначное число, не содержащее в своей записи нулей, можно рассматривать как выборку из девяти элементов 1, 2, ..., 9 по четыре элемента в каждой. Упорядочена эта выборка или нет? В каком случае она будет выборкой с возвращением и в каком — выборкой без возвращения?

3.3.1. Упорядоченные выборки (размещения)

Подсчитаем число упорядоченных выборок с возвращением из n элементов по r. Первый элемент выборки можно получить n способами, при любом способе его выбора второй элемент выбирается также n способами (напомним: выбор с возвращением!) и т. д. По правилу умножения ***число упорядоченных выборок с возвращением из n элементов по r равняется n^r*** (такие выборки также называют ***размещениями с повторениями,*** а их число обозначают $\overline{A}_n^r$). Итак,

$$\overline{A}_n^r = n^r.$$

Можно ли образовать упорядоченные выборки с возвращением из n элементов по r, если $r > n$?

Пусть исходами опыта являются упорядоченные выборки без возвращения из n элементов по r ($r \leqslant n$). Их называют также ***размещениями из n элементов по r,*** а их число обозначают A_n^r. Найдем, чему равняется A_n^r. Первый элемент можно выбрать n различными способами, для выбора второго элемента уже существует $n - 1$ возможность и т. д. Последний r-й элемент выбирают после извлечения ($r - 1$)-го элемента, т. е. из ($n - (r - 1)$) оставшихся элементов. Итак, применяя правило умножения, получим, что число упорядоченных выборок без возвращения из n элементов по r равно

$$A_n^r = n(n-1)(n-2) \dots (n-r+1).$$

Чему равно A_n^1; A_5^2 ?

Пример 1. Наугад набирают телефонный номер, состоящий из пяти цифр. Найти вероятность того, что этот номер состоит из различных цифр, если ни один из телефонных номеров не начинается с нуля.

❐ Исходы опыта, рассматриваемого в этом примере, имеют вид

$$(a_1, a_2, a_3, a_4, a_5); \quad a_1 \neq 0; \quad a_i = 0, 1, \ldots, 9; \quad i = 1, 2, 3, 4, 5,$$

где a_i — i-я цифра номера. По правилу умножения их число равно $N = 9 \cdot 10^4$. Все исходы равновозможны. Событие A — «телефонный номер состоит из различных цифр» — происходит при любом из $N(A) = 9 \cdot 9 \cdot 8 \cdot 7 \cdot 6$ исходов. Поэтому

$$P(A) = \frac{N(A)}{N} = \frac{9 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{9 \cdot 10^4} = 0{,}3024. \blacksquare$$

? Можно ли рассматривать в этом примере телефонные номера как упорядоченные выборки (с возвращением или без возвращения) из 10 элементов по пять?

Пример 2. Поезду, в одном из вагонов которого находится 10 пассажиров, предстоит сделать 15 остановок. Какова вероятность того, что все пассажиры выйдут на разных остановках?

❐ Для решения задачи можно обратиться к классической вероятностной модели. Нужно найти общее число исходов опыта, допустить, что исходы равновозможны и подсчитать число исходов, при которых наступает искомое событие.

Проиллюстрируем на этой задаче общую схему решения комбинаторных задач, сводящихся к выбору. Эта схема имеет следующий вид:

1) выяснить, что представляет собой опыт, о котором идет речь в условии задачи;

2) привести несколько примеров выборок — исходов этого опыта, закодировав объекты, фигурирующие в условии;

3) установить, упорядочены эти выборки или нет;

4) установить, являются ли эти выборки с возвращением или без возвращения;

5) выяснить, из какого числа элементов они образованы;

6) выяснить, сколько элементов содержит каждая выборка;

7) применить соответствующую формулу.

В данной задаче речь идет о выборе остановки каждым из 10 пассажиров. Обозначим остановки числами 1, 2, 3, ..., 15. Исхо-

дами этого опыта являются, например, выборки $\left(\underbrace{1,\,1,\,\ldots,\,1}_{10}\right)$,
$\left(\underbrace{1,\,2,\,\ldots,\,10}_{10}\right)$, $\left(\underbrace{1,\,1,\,2,\,2,\,3,\,3,\,1,\,4,\,8,\,12}_{10}\right)$. Первая из них означает, что все 10 пассажиров вышли на первой остановке, вторая — первый пассажир вышел на первой остановке, второй — на второй, и т. д., десятый — на десятой. Смысл последнего примера теперь понятен: первый, второй и седьмой пассажиры вышли на первой остановке и т. д. Эти выборки упорядочены, так как, например, две выборки (1, 2, 3, ..., 10), (2, 1, 3, ..., 10), состоящие из одних и тех же элементов, но отличающиеся порядком их следования, описывают разные исходы опыта. Эти исходы являются выборками с возвращением: на одной остановке может выйти более одного пассажира. Каждая выборка образована из 15 элементов и содержит 10 элементов. Итак, исходы данного опыта — упорядоченные выборки с возвращением из 15 элементов по 10. Их число равно $N = \bar{A}_{15}^{10} = 15^{10}$.

Событие A — «все выйдут на различных остановках» — состоит из упорядоченных выборок без возвращения (на разных остановках!) из 15 элементов по 10. Их число равно $N(A) = A_{15}^{10} = 15 \cdot 14 \times \times 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6$. Искомая вероятность равна

$$P(A) = \frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{15^{10}} \approx 0{,}0189. \blacksquare$$

Число размещений из n элементов по r можно вычислить, воспользовавшись редактором электронных таблиц Microsoft Excel. Для этого применяют статистическую функцию ПЕРЕСТ(число; выбранное_число), где число равно n, выбранное_число равно r.

3.3.2. Перестановки

Если в опыте с извлечением r шаров из n пронумерованных $r = n$, то упорядоченные выборки без возвращения называют ***перестановками из n элементов***. Число всех различных перестановок из n элементов вычисляют по формуле

$$n(n-1)(n-2)\ldots(n-n+1) = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n.$$

Напомним, что произведение $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ коротко обозначают $n!$ (читается «n-факториал»). Так $2! = 1 \cdot 2$; $3! = 1 \cdot 2 \cdot 3 = 6$. По определению $0! = 1$; $1! = 1$.

Сколько перестановок можно образовать из букв слова «слон»?

Ранее мы доказали, что число перестановок из n различных элементов равно $n!$. Так, из букв слова «линейка» можно образовать $7! = 5040$ перестановок. В слове «элемент» тоже семь букв, но не все они различны, это слово содержит дважды букву «е». Если эти буквы «е» переставлять, то будем получать ту же перестановку, т. е. любая из 7! перестановок будет повторяться дважды. Поэтому число перестановок, которые можно получить из букв этого слова, равно $\frac{7!}{2} = 2520$. Используем эти соображения для доказательства общего утверждения.

ТЕОРЕМА 1. *Число перестановок, которые можно образовать из n элементов, среди которых n_1 элементов 1-го типа, n_2 — 2-го типа, ..., n_k — k-го типа, равно*

$$\frac{n!}{n_1!\, n_2! \ldots n_k!}.$$

Такие перестановки называют иногда ***перестановками с повторениями.***

➤

❒ Если бы все n элементов были различными, то число перестановок равнялось бы $n!$ Поскольку некоторые элементы совпадают, число перестановок уменьшится. Рассмотрим, например, перестановку $\underbrace{a, a, \ldots, a}_{n_1};\ \underbrace{b, b, \ldots, b}_{n_2};\ \ldots;\ \underbrace{z, z, \ldots, z}_{n_k}$. Элементы 1-го типа можно переставлять $n_1!$ способами. Но, поскольку эти элементы одинаковые, получим ту же перестановку из n элементов. Так же ничего не изменяют $n_2!$ перестановок элементов 2-го типа, ..., $n_k!$ перестановок k-го типа. Перестановки элементов 1-го типа, 2-го типа и т. д. можно выполнять независимо друг от друга. Поэтому одинаковые элементы любой перестановки из n элементов можно переставлять $n_1!\, n_2! \ldots n_k!$ способами так, что она не изменяется. Значит, совокупность всех перестановок распадается на части, состоящие из $n_1!\, n_2! \ldots n_k!$ одинаковых перестановок. Таким образом, число различных перестановок с повторениями, которые можно образовать из данных элементов (обозначим это число $P_n(n_1, n_2, ..., n_k)$), равно

$$P_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!n_2! \dots n_k!}. \blacksquare$$

Сколько перестановок можно образовать из букв слова «мама»?

Пример 3. Что вероятнее: при образовании всех возможных перестановок из букв слова «задача» получить «слово», где три буквы «а» будут идти подряд, или «слово», где эти буквы не будут идти подряд?

❒ Если бы все буквы были бы различными, то общее число перестановок равнялось бы $6! = 720$. Но слово «задача» содержит три одинаковых буквы «а». Поэтому число всех перестановок из букв слова «задача» равно $\frac{6!}{3!} = 120$. Это общее число исходов опыта — число всех перестановок. Обозначим через A событие «в перестановке три буквы «а» идут подряд». Вычислим число исходов опыта, благоприятствующих наступлению этого события. «Свяжем» три буквы «а», фактически будем иметь перестановки из четырех различных элементов, их число равно $4! = 24$. Итак, $P(A) = \frac{24}{120} = 0{,}2$. Вероятность противоположного события $\bar{A}$ равна $1 - 0{,}2 = 0{,}8$. Итак, $P(\bar{A}) > P(A)$. ■

3.3.3. Неупорядоченные выборки (сочетания)

Выведем теперь формулу для вычисления числа неупорядоченных выборок без возвращения из n элементов по r ($r \leqslant n$). Такие выборки называют ***сочетаниями из n элементов по r***. Их число обозначают C_n^r (читается «число сочетаний из n элементов по r»). Идею вывода объясним сначала на примере. Имеем четыре элемента a, b, c, d. Выпишем все неупорядоченные выборки без возвращения по три элемента в каждой, т. е. сочетания из четырех элементов по три:

$$(a, b, c); \quad (a, b, d); \quad (a, c, d); \quad (b, c, d).$$

Из каждой из этих выборок (их C_4^3) образуем все перестановки (их 3!) из трех элементов. Тем самым мы построим все упорядоченные выборки из четырех элементов по три:

$$\begin{array}{cccccc} abc & acb & bac & bca & cab & cba \\ abd & adb & bad & bda & dab & dba \\ acd & adc & cad & cda & dac & dca \\ bcd & bdc & cbd & cdb & dbc & dcb \end{array}$$

Все перестановки, например, из элементов a, b, c записаны в первой строке. Как известно, число упорядоченных выборок без возвращения из четырех элементов по три равно $4 \cdot 3 \cdot 2$. Таким образом, $4 \cdot 3 \cdot 2 = = C_4^3 \cdot 3!$, отсюда

$$C_4^3 = \frac{4 \cdot 3 \cdot 2}{3!}.$$

Аналогичные соображения остаются справедливыми и в общем случае.

Из каждой неупорядоченной выборки, состоящей из r различных элементов, можно получить ровно $r!$ упорядоченных выборок. Итак, по правилу умножения число всех упорядоченных выборок из n по r равно

$$n(n-1)(n-2)...(n-r+1) = C_n^r \cdot r!.$$

Таким образом,

$$C_n^r = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!}. \tag{1}$$

По определению считают, что

$$C_n^0 = 1.$$

Умножив числитель и знаменатель правой части равенства (1) на $(n-r)! = 1 \cdot 2 \cdot ... \cdot (n-r)$, получим

$$C_n^r = \frac{n(n-1)(n-2)\ldots(n-r+1) \cdot 1 \cdot 2 \cdot \ldots \cdot (n-r)}{r!\,(n-r)!}.$$

Итак,

$$C_n^r = \frac{n!}{r!\,(n-r)!}. \tag{2}$$

Заменив в равенстве (2) r на $n-r$, получим

$$C_n^{n-r} = \frac{n!}{(n-r)!\,(n-(n-r))!} = \frac{n!}{(n-r)!\,r!},$$

т. е.

$$C_n^r = C_n^{n-r}.$$

Содержание последнего равенства понятно и без преобразований: каждая выборка из n элементов по r однозначно определяет выборку из n элементов по $n - r$ оставшихся элементов.

Пример 4. Вычислить C_5^2; C_{10}^4; C_n^1; C_n^n; C_9^6; C_{100}^{98}.

❐ $C_5^2 = \dfrac{5 \cdot 4}{2!} = 10;\quad C_{10}^4 = \dfrac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = \dfrac{10 \cdot 9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4} = 210;$

$$C_n^1 = \frac{n!}{1!(n-1)!} = n;\quad C_n^n = \frac{n!}{n!(n-n)!} = \frac{1}{0!} = 1;$$

$$C_9^6 = C_9^3 = \frac{9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3} = 84;\quad C_{100}^{98} = C_{100}^2 = \frac{100 \cdot 99}{1 \cdot 2} = 4950. \blacksquare$$

Число сочетаний из n элементов по r можно вычислить, воспользовавшись редактором Excel. Для этого применяют математическую функцию ЧИСЛКОМБ(число; выбранное_число), где число равно n, выбранное_число равно r.

? Придумайте задание, ответом которого будет число C_{10}^4.

Пример 5. Из ящика, содержащего 20 пригодных и пять бракованных изделий, наугад вынимают три изделия. Чему равна вероятность того, что:

а) все изделия пригодны;

б) пригодны лишь два изделия;

в) пригодно лишь одно изделие;

г) все изделия бракованы?

❐ а) Поскольку порядок извлечения изделий не имеет значения, то исходами опыта являются неупорядоченные выборки без возвращения из 25 элементов по три. Исходы опыта равновозможны, их число равно

$$N = C_{25}^3 = \frac{25 \cdot 24 \cdot 23}{1 \cdot 2 \cdot 3} = 2300.$$

Событие A — «все изделия пригодны» — состоит из выборок, элементами которых являются пригодные изделия, т. е. из неупорядоченных выборок без возвращения из 20 элементов по три:

$$N(A) = C_{20}^3 = \frac{20 \cdot 19 \cdot 18}{1 \cdot 2 \cdot 3} = 1140;\quad P(A) = \frac{1140}{2300} \approx 0{,}496.$$

б) Событие B — «пригодны лишь два изделия» — наступает тогда и только тогда, когда извлекли два пригодных изделия (это можно сделать $C_{20}^2 = \dfrac{20 \cdot 19}{1 \cdot 2} = 190$ способами) и одно бракованное (для этого существует $C_5^1 = 5$ способов). По правилу умножения вычислим число исходов опыта, благоприятствующих наступлению события B: $N(B) = 190 \cdot 5 = 950$. Итак, $P(B) = \dfrac{950}{2300} \approx 0{,}413$.

в) Аналогично предыдущему заданию, для события C — «пригодно лишь одно изделие» — имеем $P(C) = \dfrac{C_{20}^1 \cdot C_5^2}{C_{25}^3} = \dfrac{20 \cdot 10}{2300} \approx 0{,}087$.

г) Аналогично выполнению задания а) для события D — «все изделия бракованы» — имеем $P(D) = \dfrac{C_5^3}{C_{25}^3} = \dfrac{10}{2300} \approx 0{,}004$. ■

Обратите внимание на то, что $0{,}496 + 0{,}413 + 0{,}087 + 0{,}004 = 1{,}000$. Дело в том, что сумма событий $A + B + C + D$ является достоверным событием: обязательно наступает одно и только одно из попарно несовместных событий A, B, C, D, поэтому $P(A + B + C + D) = P(A) + P(B) + P(C) + P(D) = 1$.

Выше выведены формулы для числа упорядоченных выборок с возвращением и без возвращения и для числа неупорядоченных выборок без возвращения. Осталось получить формулу для числа неупорядоченных выборок из n элементов по r с возвращением. Их еще называют ***сочетаниями с повторениями из n элементов по r***. Любая такая выборка содержит r элементов, причем каждый элемент принадлежит к одному из n типов элементов: так как выборки с возвращением, то каждый данный из n элементов можно тиражировать в любом количестве экземпляров.

Пример 6. Образуем, например, всевозможные неупорядоченные выборки с возвращением из трех элементов a, b, c по два элемента в каждой:

$$\begin{array}{lll} aa & & \\ ab & bb & \\ ac & bc & cc \end{array} \quad ■$$

Выпишите все неупорядоченные выборки с возвращением из трех элементов a, b, c по три.

Можно ли образовать неупорядоченные выборки с возвращением из n элементов по r, если $r > n$?

Пример 7. В качестве примера неупорядоченных выборок с возвращением можно рассматривать кости домино: каждую кость домино можно рассматривать как выборку, образованную из семи элементов 0, 1, 2, 3, 4, 5, 6, содержащую два элемента, причем это выборки с возвращением (среди костей домино есть дубли 0 : 0, 1 : 1 и т. д.), неупорядоченные (кости 1 : 2 и 2 : 1 неразличимы). ■

? Подсчитайте число указанных выборок, т. е. число костей домино, с помощью «точек и перегородок».

Пример 8. Неупорядоченными выборками с возвращением являются, например, различные распределения r одинаковых предметов в n ячейках. Например, распределения трех одинаковых карандашей между двумя детьми можно представить в виде: (1 1 1), (1 1 2), (1 2 2), (2 2 2). Эти представления означают: первый ребенок получил все три карандаша; первый получил два карандаша, второй — один; первый получил один карандаш, второй — два; все три карандаша получил второй ребенок. Эти представления можно рассматривать как неупорядоченные (распределения (1 1 2) и (1 2 1) не различаются) выборки с возвращением (они могут содержать одинаковые элементы) из двух элементов 1 и 2 по три элемента в каждой. ■

? Опишите на языке выборок распределения n одинаковых предметов в r ячейках.

Обозначим число неупорядоченных выборок без возвращения из n элементов по r через f_n^r. В предыдущем параграфе мы подсчитывали это число с помощью «точек и перегородок». Оно равнялось числу способов, с помощью которых можно выбрать $n - 1$ место для перегородок из общего числа $n + r - 1$ мест. Как мы видели выше, оно равняется $C_{n+r-1}^{n-1} = C_{n+r-1}^{r}$, т. е.

$$f_n^r = C_{n+r-1}^{n-1} = C_{n+r-1}^{r}.$$

➤

❐ Докажем эту формулу. Рассмотрим следующий метод для подсчета таких выборок.

Каждой неупорядоченной выборке с возвращением из n элементов по r поставим в соответствие последовательность из нулей и единиц, составленную по следующему правилу. Запишем подряд столько единиц, сколько элементов 1-го типа входит в выборку, затем поставим нуль, далее запишем столько единиц, сколько элементов 2-го типа содержит выборка, затем ставим нуль и т. д., наконец, записываем столько единиц, сколько элементов n-го типа содержит выборка.

Для примера 6 эти последовательности имеют следующий вид:

1100
1010 0110
1001 0101 0011

Таким образом, каждой неупорядоченной выборке с возвращением из n элементов по r поставлена в соответствие последовательность из $n - 1$ нулей и r единиц, т. е. последовательность длиной $n + r - 1$. Верно и обратное: каждая последовательность из $n - 1$ нулей и r единиц однозначно определяет такую выборку. Между выборками и последовательностями установлено взаимно однозначное соответствие. Поэтому число неупорядоченных выборок без возвращения из n элементов по r равно числу таких последовательностей, что в свою очередь равно числу способов выбора r мест для единиц или, что то же самое, $n - 1$ мест для нулей из общего числа $n + r - 1$ мест, т. е.

$$f_n^r = C_{n+r-1}^{n-1} = C_{n+r-1}^{r}. \blacksquare$$

? Выпишите все неупорядоченные выборки с возвращением из трех элементов a, b, c по три и поставьте каждой выборке в соответствие последовательность из единиц и нулей по описанному выше правилу.

Вернемся к примеру 7 и подсчитаем число костей домино. Их число равно

$$f_7^2 = C_{7+2-1}^2 = C_8^2 = \frac{8 \cdot 7}{2} = 28.$$

Вернемся к примеру 8 и подсчитаем, сколькими способами можно распределить три одинаковых карандаша между двумя детьми. Число способов равно $f_2^3 = C_{2+3-1}^3 = C_4^3 = C_4^{4-3} = C_4^1 = 4$. Все эти четыре способа были выписаны выше.

Вообще, r одинаковых предметов в n ячейках можно распределить $f_n^r = C_{n+r-1}^{n-1} = C_{n+r-1}^{r}$ способами.

Пример 9. Сколькими способами можно выбрать три из 12 букв: А, А, А, Т, Т, Т, Г, Г, Г, Ц, Ц, Ц?

❒ Исходы опыта, заключающегося в выборе трех букв из 12 заданных, имеют, например, вид: (А, А, А), (А, Ц, Ц), (Т, Г, Ц) и т. д. Они представляют собой неупорядоченные выборки (исходы (А, Ц, Ц) и (Ц, А, Ц) неразличимы) с возвращением (буквы могут повторяться) из четырех элементов (А, Т, Г, Ц) по три в каждой. Их число равно $f_4^3 = C_{4+3-1}^{3} = C_6^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} = 20$. ■

➤

Задача имеет отношение к теории белкового кода, предложенной американским физиком Г. Гамовым (1904—1968)[1]. Буквы А, Т, Г, Ц обозначают нуклеотиды: аденин, тимин, гуанин, цитозин. Число троек нуклеотидов оказывается равным 20 — числу стандартных аминокислот, на которые разлагаются молекулы белка. ➤

3.3.4. Свойства сочетаний

Числа C_n^r обладают целым рядом замечательных свойств. Одно из них было рассмотрено выше: $C_n^r = C_n^{n-r}$. Фактически оно было доказано двумя способами. Один способ сводился к применению формулы $C_n^r = \frac{n!}{r!\,(n-r)!}$. Другой основан на комбинаторных соображениях: всякая неупорядоченная выборка без возвращения из n элементов по r порождает такую же выборку из n элементов по $n - r$, и наоборот, каждая неупорядоченная выборка без возвращения из n элементов по $n - r$ порождает такую же выборку из n элементов по r. Значит, $C_n^r = C_n^{n-r}$.

Точно так же могут доказываться и другие свойства сочетаний.

---

[1] Георгий Антонович Гамов родился в Одессе, закончил Ленинградский университет и до 1933 г. работал в Физико-техническом институте в Ленинграде.

ТЕОРЕМА 2. *Имеет место равенство*

$$C_n^r = C_{n-1}^r + C_{n-1}^{r-1}.$$

❐ *1-й способ.* Составим все неупорядоченные выборки без возвращения из n элементов $a_1, a_2, \ldots, a_{n-1}, a_n$ по r элементов в каждой и разобьем их на два класса. В первый из них войдут выборки, содержащие элемент a_n, а во второй — выборки, не содержащие этого элемента. Если из любой выборки первого класса отбросить элемент a_n, то останется выборка, содержащая $r-1$ элемент и составленная из элементов $a_1, a_2, \ldots, a_{n-1}$. Число этих выборок равно C_{n-1}^{r-1}. Поэтому в первый класс входит C_{n-1}^{r-1} выборок. Выборки второго класса являются неупорядоченными выборками без возвращения из $n-1$ элемента $a_1, a_2, \ldots, a_{n-1}$ по r элементов в каждой. Поэтому их число равно C_{n-1}^r. Так как каждая выборка принадлежит одному и только одному из этих классов, а общее число этих выборок равно C_n^r, то приходим к требуемому равенству $C_n^r = C_{n-1}^r + C_{n-1}^{r-1}$.

2-й способ.

$$C_{n-1}^r + C_{n-1}^{r-1} = \frac{(n-1)!}{r!\,(n-r-1)!} + \frac{(n-1)!}{(r-1)!\,(n-r)!} =$$

$$= \frac{(n-1)!}{(r-1)!\,(n-r-1)!}\left(\frac{1}{r} + \frac{1}{n-r}\right) = \frac{(n-1)!}{(r-1)!\,(n-r-1)!}\,\frac{n}{r(n-r)} =$$

$$= \frac{n!}{r!\,(n-r)!} = C_n^r.$$

Еще один способ доказательства основан на применении геометрической интерпретации числа сочетаний, рассмотренной в задаче 264 к § 3.2. ■

ТЕОРЕМА 3. *Имеет место равенство*

$$C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n.$$

❐ 2^n — это число всех упорядоченных выборок с возвращением из двух элементов по n. Разобьем эти выборки на $n+1$ класс, отнеся к k-му классу ($k = 0, 1, 2, \ldots, n$) те выборки, в которые первый элемент входит k раз, а второй — $n-k$ раз. Число таких выборок равно $\frac{n!}{k!\,(n-k)!}$, т. е. C_n^k. Значит, общее число выборок всех классов равно $C_n^0 + C_n^1 + C_n^2 + ... + C_n^n$. Теорема доказана. ■

Некоторые элементы комбинаторики были известны еще во II веке до н. э. Термин «комбинаторика» происходит от латинского слова *combina* – «соединять». Этот термин впервые ввел Г. Лейбниц (1646–1716), который обосновал формулы числа сочетаний и перестановок. Изучением сочетаний занимались М. Штифель (1486–1567), Н. Тарталья (ок. 1500–1557), Б. Паскаль (1623–1662), П. Ферма (1601–1665) и другие математики. Размещения впервые изучил Я. Бернулли (1654–1705). Он же ввел термин «размещение» и употребил в современном понимании термин «перестановка». Термин «сочетание» использовал еще Б. Паскаль. Современную символику сочетаний предложили лишь в XIX столетии. К началу XX века комбинаторика считалась практически завершенным разделом математики, лежащим вне основного русла развития математики и ее приложений. В XX веке комбинаторику стали рассматривать как раздел теории множеств, изучающий различные проблемы, возникающие при изучении конечных множеств. Такая точка зрения привела к естественной классификации понятий и задач комбинаторики. В последнее время роль комбинаторики возросла в связи с развитием теории вычислительных машин, теории информации, изучающей методы оптимального кодирования, декодирования и передачи информации. Ныне комбинаторика является одним из наиболее интенсивно развивающихся разделов математики.

Контрольные вопросы

1. Чему равно: C_{15}^1; C_{10}^0; 1!; 0!?

2. Четырехзначный номер автомобиля можно рассматривать как выборку из 10 цифр. Упорядочена эта выборка или нет? С возвращением или нет?

3. Ящик содержит как пригодные, так и бракованные изделия. Для проверки отбирают без возвращения 10 изделий. Упорядочена полученная выборка или нет?

4. Назовите все выборки по два элемента из двух элементов a и b:

а) упорядоченные, с возвращением;

б) упорядоченные, без возвращения;

в) неупорядоченные, с возвращением;

г) неупорядоченные, без возвращения.

5. Сколько перестановок можно образовать из букв слова «рама»?

6. Сколько перестановок можно образовать из букв слова «рама» так, чтобы две буквы «а» стояли рядом; не стояли рядом?

Задачи

270.° Вычислите: C_7^2; C_7^6; C_{15}^3; C_{50}^{48}.

271.° Сколькими способами из 10 спортсменов можно отобрать четыре человека для участия:

а) в соревнованиях по бегу на 100 метров;

б) в эстафете 100 м + 200 м + 400 м + 800 м?

272. Из урны, содержащей три белых шара и семь черных шаров, наугад вынимают сразу два шара. Найдите вероятность того, что:

а)° оба шара черные;

б)° оба шара белые;

в) один шар белый, второй — черный.

273.° Контролеру принесли 100 одинаковых изделий, среди которых есть 10 бракованных. Но контролер этого не знает и наугад берет для проверки 10 изделий. Если все проверенные изделия окажутся доброкачественными, то подконтрольная партия изделий принимается, в противном случае она подлежит дополнительной проверке. Какова вероятность того, что эту партию изделий контролер примет?

274. Партия из 100 изделий подлежит выборочному контролю. Условием непригодности всей партии является наличие по крайней мере одного бракованного изделия среди 5 проверенных. Какова вероятность для данной партии быть принятой, если она содержит 5% неисправных изделий?

275. Двенадцать мест для стоянки автомобилей расположены в один ряд. На стоянке случайно разместились восемь автомобилей. Найдите вероятность того, что четыре пустые места расположены рядом.

276. На пятиместную скамью случайно садятся пять человек. Какова вероятность того, что три определенных человека окажутся рядом?

277. Найдите вероятность того, что среди наугад выбранных m человек у кого-то совпадут дни рождения.

278. Футбольная команда лицея может сыграть за сезон шесть игр с командами других лицеев города. Все результаты игр — победа, ничья, поражение — равновозможны. Найдите вероятность того, что эта команда выиграет две игры, три проиграет и одну сведет к ничьей.

279. Для награждения шести призеров математической олимпиады выделено три экземпляра одной книги, два экземпляра второй и один экземпляр третьей книги. Сколькими способами могут быть вручены премии, если никому не дают более одной книги?

280. Сколькими способами можно выбрать шесть одинаковых или различных сортов пирожных в кондитерской, где есть 11 разных сортов пирожных?

281.* Дано уравнение $x_1 + x_2 + ...+ x_m = n$. Сколько оно имеет решений:

а) в целых неотрицательных числах;

б) в целых положительных числах?

§ 3.4. Бином Ньютона

Как известно из предыдущих параграфов, n различных шаров в k ячейках можно разместить C_n^k способами. Впишем в таблицу 3.9 значения C_n^k при различных значениях k и n.

Таблица 3.9

n	k						
	0	1	2	3	4	5	6
1	$C_1^0 = 1$	$C_1^1 = 1$					
2	$C_2^0 = 1$	$C_2^1 = 2$	$C_2^2 = 1$				
3	$C_3^0 = 1$	$C_3^1 = 3$	$C_3^2 = 3$	$C_3^3 = 1$			
4	$C_4^0 = 1$	$C_4^1 = 4$	$C_4^2 = 6$	$C_4^3 = 4$	$C_4^4 = 1$		
5	$C_5^0 = 1$	$C_5^1 = 5$	$C_5^2 = 10$	$C_5^3 = 10$	$C_5^4 = 5$	$C_5^5 = 1$	

Попробуем изучить закономерности, присущие элементам этой таблицы. Сразу бросается в глаза то, что первый столбец состоит из единиц, и точно так же последняя диагональ составлена из единиц. Второй столбец состоит из последовательных натуральных чисел от 1 до 5. Внимательно присмотревшись к этой таблице, можно обнаружить и другие закономерности. Обратите внимание на числа во второй и третьей строках: $2 + 1 = 3$ (тройка стоит под единицей). Точно так же $3 + 3 = 6$ (шестерка — под тройкой), $3 + 1 = 4$ (четверка под единицей) и т. д. Теперь нетрудно заполнить всю шестую строку: $1, 1 + 5 = 6, 5 + 10 = 15, 10 + 10 = 20, 10 + 5 = 15, 5 + 1 = 6, 1$.

?

Заполните седьмую строку последней таблицы.

Пользуясь составленной таблицей, ответьте, сколькими способами можно разместить пять одинаковых шариков в шести ячейках.

Построенная таблица давно известна в математике, ее называют ***треугольником Паскаля.***

Рассмотрим еще раз внимательно треугольник Паскаля. Во второй строке стоят числа 1, 2, 1. Вам это ничего не напоминает? Давайте вспомним формулу для квадрата суммы двух выражений

$$(a + b)^2 = a^2 + 2ab + b^2.$$

Заметили сходство? Коэффициенты в правой части последнего равенства тоже равны 1, 2, 1. Обратимся к следующей строке треугольника Паскаля. Она состоит из чисел 1, 3, 3, 1. Сравним с формулой куба суммы двух выражений

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Опять коэффициенты 1, 3, 3, 1 в правой части равенства совпали с числами, стоящими в третьей строке треугольника Паскаля. По аналогии можно записать:

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.$$

?

Проверьте правильность этого равенства, представив $(a + b)^4$ в виде $(a + b)^3(a + b)$.

Аналогично формула записывается в общем виде:

$$(a + b)^n = a^n + C_n^1 a^{n-1}b + C_n^2 a^{n-2}b^2 + \ldots + C_n^{n-1} a^1b^{n-1} + b^n.$$

Коротко ее можно записать так:

$$(a + b)^n = \sum_{k=0}^{n} C_n^k a^{n-k}b^k.$$

Эта формула носит название ***формулы бинома Ньютона.*** Она служит для возведения в натуральную степень суммы двух слагаемых (бином — это двучлен). Коэффициенты C_n^k в этой формуле называют ***биномиальными коэффициентами.***

Пример 1. Раскрыть скобки в выражении $(1 - x)^7$.

$$❐\ (1 - x)^7 = 1^7 + C_7^1 \cdot 1^{7-1} \cdot (-x)^1 + C_7^2 \cdot 1^{7-2} \cdot (-x)^2 +$$
$$+ C_7^3 \cdot 1^{7-3} \cdot (-x)^3 + C_7^4 \cdot 1^{7-4} \cdot (-x)^4 + C_7^5 \cdot 1^{7-5} \cdot (-x)^5 +$$
$$+ C_7^6 \cdot 1^{7-6} \cdot (-x)^6 + (-x)^7 =$$
$$= 1 - 7x + 21x^2 - 35x^3 + 35x^4 - 21x^5 + 7x^6 - x^7. \blacksquare$$

? Найдите разложение $(1 + x)^7$ с помощью треугольника Паскаля.

Пример 2. Найти n, если известно, что в разложении $(1 + x)^n$ коэффициенты при x^5 и x^{12} равны.

❐ Коэффициенты при x^5 и x^{12} соответственно равны C_n^5 и C_n^{12}. Равенство $C_n^5 = C_n^{12}$ имеет место, если $n - 12 = 5$, т. е. при $n = 17$. ■

Пример 3. Сколько рациональных членов содержит разложение $(\sqrt{2} + \sqrt[4]{3})^{100}$?

❐ $(k + 1)$-й член этого разложения имеет вид

$$C_{100}^k (\sqrt{2})^{100-k} (\sqrt[4]{3})^k = C_{100}^k \cdot 2^{\frac{100-k}{2}} \cdot 3^{\frac{k}{4}}.$$

Это выражение будет рациональным, если числа $\frac{100-k}{2}$ и $\frac{k}{4}$, $k = 0, 1, \ldots, 100$, будут целыми. Каждое четвертое значение k из совокупности чисел от 0 до 100, начиная с 0, удовлетворяет этому условию. Таких чисел 26: 0, 4, 8, ..., 100. ■

? Совпадают ли биномиальные коэффициенты с коэффициентами членов в разложении степени бинома?

Приведем теперь доказательство формулы бинома Ньютона.

➤

❐ Перемножим последовательно $(a + b)$ n раз. Получим сумму 2^n слагаемых вида $d_1 d_2 \ldots d_n$, где d_i $(i = 1, 2, \ldots, n)$ равно либо a, либо b. Разобьем все слагаемые на $n + 1$ группу $B_0, B_1, \ldots, B_n$, отнеся к B_k $(k = 0, 1, 2, \ldots, n)$ все те произведения, в которых b встречается множителем k раз, а a встречается $(n - k)$ раз. Число произведений в B_k равно C_n^k (таким числом способов среди n множителей $d_1, d_2, \ldots, d_n$

можно выбрать k множителей, которые будут равны b), а каждое слагаемое в B_k равно $a^{n-k}b^k$. Поэтому $(a+b)^n = \sum\limits_{k=0}^{n} C_n^k a^{n-k}b^k$. ■ ◄

Отметим некоторые **СВОЙСТВА** биномиальных коэффициентов.

1. *Коэффициенты членов, равноудаленных от концов разложения, равны между собой.* Это свойство вытекает из тождества $C_n^r = C_n^{n-r}$.

2. *Для получения биномиального коэффициента следующего члена разложения $(a+b)^n$ достаточно биномиальный коэффициент предыдущего члена умножить на показатель буквы a в этом члене и разделить на число членов, предшествующих определяемому.* Это свойство вытекает из тождества $C_n^r = C_n^{r-1}\dfrac{n-r}{r}$.

3. *Сумма всех биномиальных коэффициентов разложения $(a+b)^n$ равна 2^n.* Это свойство вытекает из тождества $C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n$, доказанного в предыдущем параграфе.

4. *Сумма биномиальных коэффициентов, стоящих на нечетных местах, равна сумме биномиальных коэффициентов, стоящих на четных местах.*

❒ Положив $a = 1$, $b = -1$ в равенстве

$$(a+b)^n = a^n + C_n^1 a^{n-1}b + C_n^2 a^{n-2}b^2 + \ldots + C_n^{n-1} a^1b^{n-1} + b^n,$$

получим

$$0 = 1 - C_n^1 + C_n^2 - C_n^3 + \ldots + (-1)^n,$$

откуда

$$1 + C_n^2 + C_n^4 + \ldots = C_n^1 + C_n^3 + C_n^5 + \ldots . \blacksquare$$

Формулу бинома Ньютона можно использовать для приближенных вычислений степеней.

Положив в формуле бинома Ньютона $a = 1$, $b = x$, получим

$$(1+x)^n = 1 + C_n^1 x + C_n^2 x^2 + \ldots + C_n^{n-1} x^{n-1} + C_n^n x^n.$$

Если значение x мало, то значения $x^2, x^3, \ldots, x^n$ тем более малы. Поэтому если в последнем равенстве отбросить все слагаемые, начиная с третьего, и учесть, что $C_n^1 = n$, то получим приближенную формулу

$$(1+x)^n \approx 1 + nx.$$

При малых значениях x она дает вполне удовлетворительный результат.

Пример 4. Вычислить без использования вычислительных средств:

а) $(1{,}02)^6$; б) $(0{,}97)^8$.

❒ Применяя полученную формулу, будем иметь:

$$(1{,}02)^6 = (1 + 0{,}02)^6 \approx 1 + 0{,}02 \cdot 6 = 1{,}12;$$
$$(0{,}97)^8 = (1 + (-0{,}03))^8 \approx 1 + (-0{,}03) \cdot 8 = 0{,}76.$$

Применение вычислительных средств дает соответственно результаты: 1,1262 и 0,784. ■

Хотя доказанную формулу называют биномом Ньютона, исторически это не является справедливым. Формулу для $(a + b)^n$ знали еще среднеазиатские математики Омар Хайам (1048–1131) и Гийас ад-Дин Джемшид ал-Каши (ум. 1429). В западной Европе до Ньютона ее знал Паскаль (1623–1662). Строго доказана она была швейцарским математиком Я. Бернулли (1654–1705). Заслуга Исаака Ньютона (1643–1727) в том, что он обобщил эту формулу для нецелого и отрицательного показателя n.

Контрольные вопросы

1. Как изменяются показатели степеней букв a и b в разложении $(a + b)^n$?

2. Справедлива ли формула бинома Ньютона для $n = 1$, $n = 0$?

3. Чему равно число слагаемых в разложении $(a + b)^n$?

4. Почему приближенная формула $(1 + x)^n \approx 1 + nx$ применима только для малых значений x?

5. Чему равен показатель степени в выражении $(a + b)^n$, если сумма всех биномиальных коэффициентов равна 64?

6. Биномиальный коэффициент третьего от начала члена разложения $(a + b)^n$ равен 28. Чему равен:

а) показатель степени n;

б) биномиальный коэффициент третьего от конца члена этого разложения;

в) биномиальный коэффициент четвертого от начала члена этого разложения?

7. Каков наибольший биномиальный коэффициент разложения $(a + b)^n$, если сумма всех биномиальных коэффициентов равна 128?

8. Может ли биномиальный коэффициент разложения $(a + b)^n$ равняться 24?

9. Чему равно отношение биномиального коэффициента шестого члена разложения $(a + b)^{15}$ к биномиальному коэффициенту пятого члена этого разложения?

Задачи

282.° Раскройте скобки в выражении $(x + y)^6$.

283.° Запишите первые три члена выражения $(a + b)^{100}$.

284.° Найдите показатель степени n в выражении $(3a - 2)^n$, если известно, что коэффициент при a^2 в разложении этой степени равен 216.

285.° В разложении $\left(a + \frac{1}{a}\right)^{12}$ найдите коэффициент при a^8.

286. Найдите пятый член разложения $(\sqrt{x} - \sqrt{y})^{12}$.

287. Найдите тот член разложения $(\sqrt{x} + a)^9$, который содержит x^3.

288. Найдите без использования вычислительных средств приближенное значение выражения:

а) $(1{,}001)^{10}$; б) $(0{,}98)^6$.

289.* Найдите разложение $(a + b + 2c)^4$.

Дополнительные задачи к главе 3

К § 3.1. Перебор возможных вариантов

290. Сколько из цифр 1, 2, 3 можно составить:

а) трехзначных чисел;

б) трехзначных чисел с различными цифрами;

в) нечетных трехзначных чисел;

г) нечетных трехзначных чисел с различными цифрами;

д) четных трехзначных чисел;

е) четных трехзначных чисел с различными цифрами;

ж) нечетных трехзначных чисел, содержащих хотя бы две одинаковые цифры?

з) четных трехзначных чисел, содержащих хотя бы две одинаковые цифры?

291. Какова вероятность того, что два человека, выбирая наугад вагон из трех имеющихся, окажутся в разных вагонах?

292. Три разных подарка случайным образом распределяют между тремя детьми. Какова вероятность того, что все дети получат по подарку?

293. Три одинаковых подарка случайным образом распределяют между тремя детьми. Какова вероятность того, что все дети получат по подарку?

294. Три письма рассылают по трем адресам. Сколькими способами можно осуществить рассылку писем, если никакие два письма нельзя посылать по одному адресу? Сколькими способами можно разослать письма, если по одному адресу разрешается посылать более одного письма?

295. Имеется пять карточек, на которых написаны числа 1, 2, 3, 4, 5. Наугад вынимают две карточки.

а) Какова вероятность того, что на обеих карточках написаны простые числа?

б) Какова вероятность того, что двузначное число, образованное из цифр, написанных на карточках, делится на 6?

296. Сколькими способами из четырех учеников, которые увлекаются работой на компьютере, учитель информатики может выбрать:

а) трех учеников для переноски компьютеров;

б) трех учеников, из которых одного — для создания алгоритма решения задачи, второго — для разработки компьютерной программы, третьего — для набора текста?

297. В чемпионате школы по футболу принимают участие четыре команды.

а) Сколько игр будет сыграно в этом чемпионате, если каждая команда играет с каждой из других по одной игре?

б) Сколькими способами в этом чемпионате могут определиться чемпион и команда, занявшая второе место?

298. Организационный комитет по проведению новогоднего утренника состоит из трех человек, каждый из которых может выполнять любую работу, связанную с проведением праздника.

а) Сколькими способами из них можно выбрать двух человек для поездки за елкой?

б) Сколькими способами из них можно выбрать Деда Мороза и Снегурочку?

299. Имеется четыре различные карточки: две красные и две черные.

а) Сколькими способами из них можно выбрать две карточки?

б) Сколькими способами из них можно выбрать две карточки так, чтобы среди них была одна красная и одна черная?

300. Имеется четыре различных карандаша: два простых и два красных. Наугад вынимают два карандаша. Какова вероятность того, что будут вынуты:

а) два простых карандаша;

б) два одноцветных карандаша;

в) один простой и один красный карандаш?

301. Сколькими способами три фломастера можно распределить между двумя детьми так, чтобы каждому достался хотя бы один фломастер, если фломастеры:

а) разные; б) одинаковые?

302. Сколькими способами четыре монеты можно распределить между двумя детьми, если монеты:

а) разные; б) одинаковые?

303. Сколькими способами четыре монеты можно разложить на две группы, если монеты:

а) разные; б) одинаковые?

К § 3.2. Правила умножения и сложения

304.° В вашем распоряжении есть три разных флага. На флагштоке поднимают сигнал, который состоит не менее чем из двух флагов. Сколько различных сигналов можно поднять на флагштоке, если порядок флагов:

а) учитывать; б) не учитывать?

305.* Сколько существует различных 2002-значных чисел, в которых произведение цифр, стоящих на четных местах, является нечетным числом, а произведение цифр, стоящих на нечетных местах, является четным числом?

306. В школе есть три выпускных класса, в каждом из которых по 30 учащихся. На каких условиях вы согласились бы заключить пари, что по крайней мере два выпускника родились в один и тот же день?

307. Из 28 костей домино наугад выбирают две кости. Какова вероятность того, что:

а)° среди них ровно один дубль;

б)° среди них нет дублей;

в)* их можно приставить друг к другу?

308. Из цифр 0, 1, 2, 3, 4, 5 образуют четырехзначные числа. Чему равна вероятность того, что наугад образованное число:

а) является четным и состоит из разных цифр;

б) делится на 4;

в) содержит по крайней мере одну нечетную цифру;

г) делится на 4 и состоит из разных цифр?

309. Из 33 букв русского алфавита образуют слова из шести букв, т. е. произвольные последовательности букв длиной 6. Найдите вероятность того, что любые две буквы, стоящие рядом, различны.

310. На шахматной доске наугад выбрали два квадрата. Какова вероятность того, что выбраны:

а) черный и белый квадраты;

б) два белых квадрата;

в) квадраты одного цвета?

311. Игральный кубик бросают четыре раза. Найдите вероятность того, что среди результатов бросков встретится по крайней мере одна единица?

312. Поезд состоит из восьми вагонов. Каждый из пяти пассажиров выбирает себе вагон наугад. Сколькими способами:

а) может быть произведен этот выбор;

б) они могут выбрать вагоны так, чтобы все пассажиры оказались в разных вагонах;

в) они могут выбрать вагоны так, чтобы хотя бы в одном вагоне оказалось более одного пассажира;

г) они могут выбрать вагоны так, чтобы все пассажиры оказались не более чем в трех вагонах;

д) они могут выбрать вагоны так, чтобы в первом вагоне оказалось ровно два пассажира?

313. Алфавит некоторого языка содержит 15 букв, из которых пять гласных и 10 согласных. Словом будем называть любую последовательность букв.

а) Сколько из этих букв можно составить пятибуквенных слов?

б) Сколько из этих букв можно составить пятибуквенных слов, если гласные и согласные должны чередоваться? Сколько из этих слов начинаются с гласной и сколько — с согласной?

в) Сколько из этих букв можно составить пятибуквенных слов, если ни одна буква не должна повторяться и гласные и согласные должны чередоваться?

г) Сколькими способами можно составить пятибуквенные слова, в которые входят две различные гласные и три различные согласные буквы?

д) Сколькими способами можно составить пятибуквенные слова, в которые входят две различные гласные и три различные согласные буквы, если никакие две согласные не стоят рядом?

е)* Сколько пятибуквенных слов содержат ровно три различные буквы?

314. Сколькими способами можно разделить девять экземпляров одной книги между:

а) тремя детьми;

б) тремя детьми так, чтобы каждому досталась хотя бы одна книга;

в) тремя детьми так, чтобы каждому досталось по три книги;

г) тремя детьми А, Б, В так, чтобы А и Б обязательно достались книги;

д)* тремя детьми так, чтобы двум из них обязательно достались книги;

е) тремя детьми так, чтобы каждому достались по крайней мере две книги?

315.* Имеется колода карт, которая содержит четыре масти по девять карт в каждой, пронумерованных цифрами 1, 2, 3, ..., 9. Сколькими способами можно из них выбрать пять карт так, что среди них окажутся:

а) пять последовательных карт одной масти;

б) четыре карты из пяти с одинаковыми номерами;

в) три карты с одним номером и две карты с другим;

г) пять карт одной масти;

д) пять последовательно пронумерованных карт;

е) три карты из пяти с одним и тем же номером;

ж) две карты из пяти с одинаковыми номерами, а три остальные с разными?

316.* В некоторых селах России когда-то существовало следующее гадание. Девушка зажимает в руке шесть травинок так, чтобы их концы торчали сверху и снизу, а ее подружка связывает концы травинок попарно сверху и снизу по отдельности. Если при этом травинки оказываются связанными в одно кольцо, то это должно означать, что в текущем году девушка выйдет замуж. Какова вероятность того, что связанные наугад травинки образуют кольцо?

К § 3.3. Основные комбинаторные схемы

317. Группа состоит из девяти лиц. Сколько можно образовать различных подгрупп при условии, что в подгруппу входит не менее двух лиц?

318. Рассмотрим игру спортлото «6 из 49». Лототрон случайно выбирает шесть шаров из урны, в которой содержится 49 шаров, пронумерованных числами 1, 2, 3, ..., 49. Сначала игрок вычеркивает шесть чисел на бланке. Найдите вероятность того, что игрок:

а) угадает все номера извлеченных шаров;

б) угадает ровно четыре номера;

в) угадает не менее чем четыре номера;

г) не угадает ни одного номера.

319. Шесть различных конфет наугад распределяют между двумя детьми. Найдите вероятность того, что:

а) каждый ребенок получит по три конфеты;

б) каждый ребенок получит по крайней мере одну конфету.

320. Из колоды в 52 карты наугад выбирают пять карт. Какова вероятность того, что среди них будут:

а) десятка, валет, дама, король, туз одной масти;

б) пять последовательных карт одной масти за исключением набора, приведенного в предыдущем задании;

в) четыре карты с одним и тем же названием;

г) две карты с одним названием и еще три карты тоже с одним названием?

321. Найдите вероятность того, что у наугад выбранного четырехзначного числа каждая следующая цифра будет:

а) больше предыдущей;

б) меньше предыдущей.

322. Известно, что учащийся подготовил ответы не на все 16 вопросов, вынесенных на зачет.

а) Сколько вопросов он выучил, если известно, что вероятность того, что он сможет ответить на оба вопросы зачетного задания не меньше, чем $\frac{7}{8}$?

б) Сколько вопросов он выучил, если вероятность того, что он сможет ответить лишь на один из двух наугад выбранных вопросов, равна $\frac{1}{2}$?

в) В каком случае вероятность того, что он сможет ответить на один случайно выбранный вопрос, больше, чем вероятность того, что ему удастся ответить на два, по его выбору, из трех случайно выбранных вопросов?

г) Учитель распределил наугад вопросы по восьми билетам (по два вопроса в каждом билете). Какова вероятность того, что ученик в состоянии ответить по крайней мере на один вопрос любого билета, если он подготовил ответы на 10 вопросов?

323. Бросают 12 игральных кубиков. Чему равна вероятность того, что каждое из чисел 1, 2, 3, 4, 5, 6 выпадет дважды?

324. В вещевой лотерее, содержащей 50 билетов, разыгрывается восемь предметов. Первый подошедший к урне вынимает из нее пять билетов. Какова вероятность того, что:

а) ровно два из них окажутся выигрышными;

б) по крайней мере два из них окажутся выигрышными?

325. В лифте семь пассажиров. Лифт останавливается на 10 этажах, начиная со второго. Какова вероятность того, что:

а) все пассажиры выйдут на одном и том же этаже;

б) никакие два пассажира не выйдут на одном этаже?

К § 3.4. Бином Ньютона

326. В разложении $(1 + x)^n$ четвертый член равен 120. Найдите значения x и n, если сумма биномиальных коэффициентов равна 1024.

327. Каков наибольший коэффициент разложения $(a + b)^n$, если сумма всех коэффициентов равна 4096?

328. В какую натуральную степень следует возвести двучлен $\frac{1}{\sqrt{2}} + 3$, чтобы отношение четвертого слагаемого к третьему было равно $3\sqrt{2}$?

329. Сумма третьего от начала и третьего от конца биномиальных коэффициентов разложения $(\sqrt[4]{3} + \sqrt[3]{4})^n$ равна 9900. Сколько рациональных членов содержится в этом разложении?

330. Коэффициент при x во втором члене разложения $\left(x^2 - \frac{1}{4}\right)^n$ равен 31. Найдите n.

331. Сумма коэффициентов первого, второго и третьего слагаемых разложения $\left(x^2 + \frac{1}{x}\right)^n$ равна 46. Найдите член разложения, не содержащий x.

332. Сумма коэффициентов первого, второго и третьего слагаемых разложения $\left(x^2 - \frac{2}{x}\right)^n$ равна 97. Найдите член разложения, содержащий x^4.

333. Докажите, что в разложении $(1 + a)^n$ по степени a $(a \neq 0)$ не могут содержаться три равных последовательных слагаемых.

334.* Докажите неравенство $n^{n+1} > (n + 1)^n$, $n \geqslant 3$, $n \in N$.

Глава 4

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

В обыденной жизни, на производстве, в деловой жизни и научных исследованиях постоянно приходится сталкиваться с такими ситуациями, когда интересующая нас величина может принимать различные значения в зависимости от случайных обстоятельств. Сколько автомобильных аварий произойдет в некотором населенном пункте за предстоящие сутки? На этот вопрос нельзя дать строго определенного ответа, так как число аварий подвержено случайным колебаниям. Точно так же нет возможности указать, сколько дефектных изделий обнаружится при контроле n изделий. Это число является случайным. В каждом случайном испытании (контроль n изделий является случайным испытанием) оно принимает вполне определенное значение, но при повторении эксперимента число дефектных изделий меняется случайным образом. Объем продаж в следующем месяце характеризуется некоторым числом, значение которого точно неизвестно, но находится среди некоторого набора значений. Число космических частиц, регистрируемых счетчиком, — случайно. Все рассмотренные величины — число аварий, число дефектных изделий, объем продаж, число космических частиц — являются примерами ***случайных величин***, т. е. величин, значения которых различны в зависимости от случая.

В главе 2 рассматривались случайные испытания, строились их вероятностные модели. Каждый раз, когда частью результата случайного эксперимента является число, мы имеем дело со случайными величинами. Они и станут предметом изучения данной главы. Для нас важно уметь вычислять и анализировать некоторые обобщающие характеристики (такие, как типичное значение, риск), а также вероятности событий, которые зависят от значений, принимаемых случайной величиной, т. е. от результата наблюдений. Эти вопросы будут рассмотрены на основании изучения распределения случайной величины, т. е. вероятностей принимать ею определенные значения. Некоторые из этих распределений важны для приложений. Их мы изучим более тщательно. К ним относятся ***биномиальное*** и ***нормальное*** распределения.

Случайные величины можно также рассматривать в качестве источника данных. Многие из тех наборов данных, которые рассматривались в главе 1, были получены в результате наблюдения и регистрации значений некоторых случайных величин. В этом смысле сама случайная величина уже представляет некоторую генеральную совокупность, в то время как наблюдаемые значения случайной величины представляют собой результат выборки. В главе 5 генеральные совокупности и выборки будут рассмотрены более детально.

Мы будем рассматривать случайные величины двух типов: ***дискретные*** и ***непрерывные*** (в основном первого типа). С дискретными случайными величинами работать проще, так как для них можно составить перечень всех возможных значений. Кроме того, математический аппарат, который необходим для их исследования, может быть ограничен материалом, изучаемым в средней школе.

§ 4.1. Случайная величина, закон ее распределения

В главе 2 рассмотрение случайных величин связывалось с некоторой вероятностной моделью случайного опыта, которая представляет собой пространство элементарных исходов (ПЭИ) опыта $U = \{u_1, \ldots, u_N\}$ и совокупность элементарных вероятностей $p_i = P(u_i)$, $i = 1, 2, \ldots, N$. Всякое случайное событие A есть подмножество ПЭИ: $A = \{u_{i_1}, u_{i_2}, \ldots, u_{i_k}\}$, а вероятность этого события определяется равенством $P(A) = p_{i_1} + p_{i_2} + \ldots + p_{i_k}$. Точнее

$$P(A) = \begin{cases} 0, \text{ если } A = V, \\ p(u), \text{ если } A = \{u\}, \\ p(u_{i_1}) + \ldots + p(u_{i_k}), \text{ если } A = \{u_{i_1}, \ldots, u_{i_k}\}. \end{cases}$$

Теперь в рамках такой же модели мы будем изучать случайные величины.

Рассмотрим примеры, которые помогут ввести понятие случайной величины.

Пример 1. Рассмотрим опыт, заключающийся в трехкратном подбрасывании правильной монеты. Результатом этого опыта можно считать число выпавших гербов. Эта величина зависит от исходов опыта и меняется случайно. Для ее исследования построим ПЭИ опыта, приняв в качестве элементарных исходов результаты трех подбрасываний. ПЭИ опыта имеет вид

$$U = \{\text{ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ}\}.$$

Число выпавших гербов вполне определяется исходом эксперимента. Каждому элементу ПЭИ можно поставить в соответствие число выпавших гербов (табл. 4.1).

Таблица 4.1

Исходы	ГГГ	ГГЦ	ГЦГ	ГЦЦ	ЦГГ	ЦГЦ	ЦЦГ	ЦЦЦ
Число гербов	3	2	2	1	2	1	1	0

Эта таблица задает некоторую функцию $X(u)$ на ПЭИ. Ее область определения — множество U, множество значений $\{0, 1, 2, 3\}$. ■

Пример 2. Продолжительными наблюдениями установлено, что данный стрелок при 100 независимых выстрелах примерно 20 раз выбивает восемь очков, 50 раз — девять и 30 раз — десять очков. Что можно сказать о числе очков, выбиваемых стрелком при двух независимых выстрелах?

❒ ПЭИ опыта, состоящего из двух независимых выстрелов, имеет вид $U = \{(8, 8), (8, 9), (8, 10), (9, 8), (9, 9), (9, 10), (10, 8), (10, 9), (10, 10)\}$. Например, событие (8, 9) означает, что первым выстрелом выбито восемь очков, а вторым — девять. Сумма очков полностью определяется результатом опыта. Каждому элементу ПЭИ можно поставить в соответствие число, равное сумме выбитых очков (табл. 4.2).

Таблица 4.2

Результаты опыта	(8, 8)	(8, 9)	(8, 10)	(9, 8)	(9, 9)
Сумма очков	16	17	18	17	18
Результаты опыта	(9, 10)	(10, 8)	(10, 9)	(10, 10)	
Сумма очков	19	18	19	20	

Таблица 4.2 задает некоторую функцию $X(u)$, определенную на ПЭИ. Она отображает область определения — множество U — на множество значений $\{16, 17, 18, 19, 20\}$. ■

Случайной величиной *называют числовую функцию, определенную на пространстве элементарных исходов.*

Обозначают случайные величины прописными латинскими буквами: $X = X(u)$, $Y = Y(u)$, $Z = Z(u)$, а их значения — соответствующими строчными буквами.

Постройте ПЭИ опыта, заключающегося в подбрасывании двух монет, и каждому исходу опыта поставьте в соответствие число выпавших цифр.

Постройте ПЭИ опыта, заключающегося в бросании игрального кубика, и каждому исходу опыта поставьте в соответствие число выпавших очков.

Постройте ПЭИ опыта, заключающегося в стрельбе стрелка, имеющего три патрона, до первого попадания в цель, и каждому исходу опыта поставьте в соответствие число использованных патронов.

Тот факт, что случайная величина X принимает значение x, является случайным событием. Будем обозначать его $(X = x)$.

Случайная величина является ***дискретной***, если можно перечислить все возможные значения, которые она может принимать в результате наблюдений. В примерах 1 и 2 рассматривались дискретные случайные величины — число гербов, выпавших при трех подбрасываниях монеты, и число очков, выбиваемых стрелком при двух независимых выстрелах. Количество неисправных устройств, которые будут выпущены за определенный промежуток времени, число дефектных изделий из n проверенных, число аварий на перекрестке, число космических частиц, зарегистрированных счетчиком, являются примерами дискретных случайных величин.

Случайная величина является ***непрерывной***, если она может принимать любое значение из некоторого интервала. Результат измерения физической величины (массы, напряжения, температуры), урожайность культуры у определенного фермера, размеры и массы отдельных представителей биологического рода, моменты солнечных вспышек, энергия космических частиц являются примерами непрерывных случайных величин.

На случайные величины распространяются все правила действий над функциями, заданными на числовом множестве: их можно складывать, вычитать, перемножать и т. п. Например, если каждое значение случайной величины X уменьшить на 3, то в результате получим случайную величину $X - 3$, определяемую тем же опытом. Если значения X возвести в квадрат, то получим значения случайной величины X^2. Если случайная величина X означает число очков, полученных первым стрелком при двух выстрелах, а случайная величина Y — число очков, выбитых вторым стрелком при двух выстрелах, то прибавляя к каждому значению X значения Y, получим все значения случайной величины $X + Y$ — суммы очков, выбитых двумя стрелками, каждый из которых сделал по два выстрела. Заметим, что обе случайные величины X и Y заданы на одном ПЭИ — на ПЭИ опыта, состоящем из четырех выстрелов, сделанных по два каждым из двух стрелков.

Что нужно знать о случайной величине для ее изучения? Очевидно, нужно знать все значения, которые она может принимать. Однако этого недостаточно. Например, пусть проводится два тиража лотереи, каждый из которых содержит по 100 билетов. В первой лотерее выигрыши по 10, 60 и 100 р. приходятся соответственно на 20, 5 и 1 билет, а во второй — те же выигрыши приходятся на 40, 10 и 5 билетов соответственно. Ясно, что выигрыши, которые приходятся на один билет, в обеих лотереях — случайные величины, принимающие одинаковые значения, но существенным образом отличающиеся друг от друга. Важно знать, как часто случайная величина принимает то или другое значение, т. е. вероятность, с которой она принимает каждое свое значение. Здесь рассматриваемый опыт состоит в проведении двух описанных лотерей, а его исходами являются пары чисел, равные размерам выигрышей в этих лотереях.

Итак, для описания случайной величины нужно иметь вероятностную модель соответствующего опыта. В рамках этой модели можно найти все значения $x_1, x_2, ..., x_n$, которые она может принимать. Тем самым будут определены события $(X = x_k)$, $k = 1, 2, \dots, n$. Каждое из этих событий имеет вероятность. Тем самым будем знать, как часто случайная величина принимает то или иное значение, т. е. вероятность, с которой она принимает каждое свое значение.

В опыте с трехкратным подбрасыванием правильной монеты элементарные исходы можно считать равновозможными, т. е. их вероятности равны $\frac{1}{8}$. Значит, вероятность появления нуля гербов или трех гербов равна $\frac{1}{8}$: $P(X = 0) = P(X = 3) = \frac{1}{8}$. Событие $(X = 1)$ означает, что число выпавших гербов равно 1, оно состоит из исходов ГЦЦ, ЦГЦ, ЦЦГ. Его вероятность по определению вероятности события равна $P(X = 1) = \frac{3}{8}$. Аналогично $P(X = 2) = P(\text{ГГЦ, ГЦГ, ЦГГ}) = \frac{3}{8}$. Итак, получим следующую таблицу 4.3.

Таблица 4.3

Число гербов	0	1	2	3
Вероятность	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

В верхней строке приведены все возможные различные значения, принимаемые случайной величиной X, в нижней — соответствующие вероятности. Эта таблица носит название *закона распределения дискретной случайной величины X*.

Законом распределения дискретной случайной величины X *называют числовую функцию, которая каждому значению x случайной величины X ставит в соответствие вероятность $P(X = x)$.*

В общем случае закон распределения случайной величины записывают следующим образом:

Значения x	x_1	x_2	...	x_n
Вероятности p	p_1	p_2	...	p_n

Здесь $x_1, x_2, \ldots, x_n$ — все различные значения случайной величины X, а $p_k = P(X = x_k)$, $k = 1, 2, \ldots, n$ — вероятности, с которыми X принимает эти значения.

Всякая ли таблица вида

x_1	x_2	...	x_n
p_1	p_2	...	p_n

задает закон распределения дискретной случайной величины? Чтобы дать ответ на этот вопрос, заметим, что события $(X = x_1), \ldots, (X = x_n)$ попарно несовместны и сумма их является достоверным событием. Поэтому сумма вероятностей этих событий $p_1 + p_2 + \ldots + p_n$ равна 1. Данное равенство можно использовать для проверки правильности составления закона распределения случайной величины.

Какая из таблиц не задает закон распределения случайной величины?

а)

x	0	1	2	3
p	0,1	0,5	0,1	0,3

б)

x	1	2	3	4
p	0,3	0,1	0,2	0,3

в)

x	−1	1	2	3
p	0,1	0,2	0,3	0,4

Закон распределения дискретной случайной величины представляет собой список ее возможных значений, снабженный вероятностями появления этих значений. Числа p_k для x_k, $k = 1, 2, \ldots, n$, можно рассматривать как массы, сосредоточенные в точках x_k на числовой оси (рис. 43). Закон распределения случайной величины можно рассматривать как распределение единичной массы в точках $x_1, x_2, \ldots, x_n$ на числовой оси.

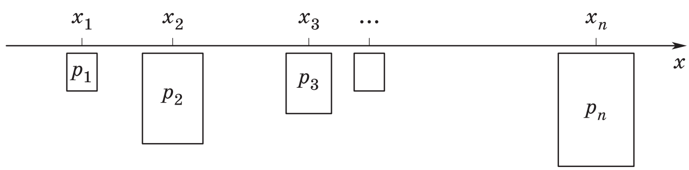


Рис. 43

Пример 3. Составить закон распределения числа очков, выбиваемых стрелком при двух независимых выстрелах в опыте, описанном в примере 2.

❒ ПЭИ опыта построено. Зададим вероятности исходов этого опыта. Событие (8, 8) является произведением двух независимых событий: «при первом выстреле выбито восемь очков» и «при втором выстреле выбито восемь очков». Вероятность каждого из этих событий равна 0,2. По теореме умножения вероятностей $P(8, 8) = 0{,}2 \cdot 0{,}2 = 0{,}04$. Аналогично:

$$P(8, 9) = 0{,}2 \cdot 0{,}5 = 0{,}1; \quad P(8, 10) = 0{,}2 \cdot 0{,}3 = 0{,}06;$$
$$P(9, 8) = 0{,}5 \cdot 0{,}2 = 0{,}1; \quad P(9, 9) = 0{,}5 \cdot 0{,}5 = 0{,}25;$$
$$P(9, 10) = 0{,}5 \cdot 0{,}3 = 0{,}15; \quad P(10, 8) = 0{,}3 \cdot 0{,}2 = 0{,}06;$$
$$P(10, 9) = 0{,}3 \cdot 0{,}5 = 0{,}15; \quad P(10, 10) = 0{,}3 \cdot 0{,}3 = 0{,}09.$$

Если через X обозначить сумму очков, выбитых при двух выстрелах (закон распределения этой случайной величины нам и нужно составить), то $P(X = 16) = P(8, 8) = 0{,}04$, $P(X = 20) = P(10, 10) = 0{,}09$. Событие $(X = 17)$ происходит при одном из результатов: (8, 9) или (9, 8). Поскольку эти события несовместны, то по теореме сложения вероятностей $P(X = 17) = P(8, 9) + P(9, 8) = 0{,}1 + 0{,}1 = 0{,}2$. Аналогично:

$$P(X = 18) = P(8, 10) + P(9, 9) + P(10, 8) = \\ = 0{,}06 + 0{,}25 + 0{,}06 = 0{,}37;$$

$$P(X = 19) = P(9, 10) + P(10, 9) = 0{,}15 + 0{,}15 = 0{,}3.$$

Итак, получаем таблицу 4.4.

Таблица 4.4

Сумма выбитых очков	16	17	18	19	20
Вероятность	0,04	0,2	0,37	0,3	0,09

Снова в верхней строке приведены все возможные различные значения, которые принимает случайная величина X, в нижней — соответствующие вероятности. Эта таблица задает закон распределения случайной величины X. ■

Итак, при составлении закона распределения случайной величины можно действовать в следующей последовательности:

1. Построить для данного эксперимента ПЭИ и задать в нем вероятности.
2. Выписать значения случайной величины, соответствующие каждому элементарному исходу.
3. Выписать все возможные различные значения $x_1, x_2, \ldots, x_n$ случайной величины и вычислить соответствующие им вероятности $p_1, p_2, \ldots, p_n$. Вероятность p_i находится сложением вероятностей всех элементарных исходов, отвечающих значению x_i.

? Составьте законы распределения следующих случайных величин, рассмотренных выше:

а) число цифр, выпавших при подбрасывании двух правильных монет;

б) число очков, выпавших при бросании симметричного игрального кубика;

в) число патронов, использованных стрелком, имеющим три патрона, при стрельбе до первого попадания в цель, если вероятность попадания при каждом выстреле равна 0,6.

Закон распределения, как и другие числовые функции, может задаваться таблицей, графически, аналитически и описательно. Первый из указанных способов уже рассмотрен. По таблицам, полученным при решении предыдущих примеров, можно построить соответствующие графики, являющиеся совокупностью точек с координатами $(x; p)$. Например, график закона распределения числа гербов, выпавших при трехкратном подбрасывании монеты, изображен на рис. 44. Он представляет собой совокупность четырех точек.

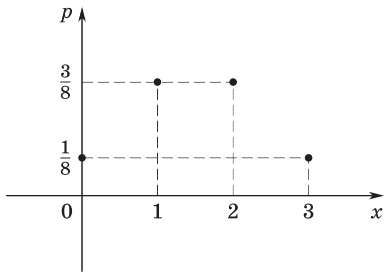


Рис. 44

Постройте график закона распределения числа очков, выбитых стрелком при двух независимых выстрелах (см. пример 3).

Примеры аналитического представления закона распределения будут приведены в следующих параграфах.

Говоря, что случайная величина принимает значения 1, 2, 3 с одинаковыми вероятностями, тем самым мы задаем закон ее распределения описательно. Постоянную величину C можно рассматривать как случайную, которая принимает лишь одно значение C с вероятностью 1. Ее закон распределения также задан описательно.

Закон распределения случайной величины дает возможность вычислять вероятности событий, связанных со случайной величиной, т. е. вероятности того, что случайная величина принимает значения из заданных промежутков.

Пример 4. Случайная величина X принимает значения 1, 2, 3, 4, причем $P(X = 1) = 0{,}4$; $P(X = 2) = 0{,}3$; $P(X = 3) = 0{,}2$. Найти $P(X = 4)$, $P(X > 2)$, $P(X \leqslant 3)$, $P(1{,}5 < X < 3{,}7)$.

❒ Поскольку $P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1$, то

$$P(X = 4) = 1 - (0{,}4 + 0{,}3 + 0{,}2) = 0{,}1.$$

Событие $(X > 2)$ наступает тогда и только тогда, когда случайная величина X принимает значение 3 или 4, т. е. это событие является суммой событий $(X = 3)$ и $(X = 4)$. Поэтому

$$P(X > 2) = P((X = 3) + (X = 4)) = P(X = 3) + P(X = 4) =$$
$$= 0{,}2 + 0{,}1 = 0{,}3.$$

Аналогично:

$$P(X \leqslant 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0{,}9;$$

$$P(1{,}5 < X < 3{,}7) = P(X = 2) + P(X = 3) = 0{,}5. \blacksquare$$

В следующем примере рассматривается частный случай распределения дискретной случайной величины, которое находит применение в задачах контроля за качеством продукции.

Пример 5. Партия из 10 телевизоров содержит четыре неисправных телевизора. Из этой партии наугад выбирают три телевизора. Составить закон распределения числа неисправных телевизоров в выборке.

❒ Строить закон распределения по приведенной схеме довольно громоздко. Если пронумеровать телевизоры числами от 1 до 10, причем последние четыре номера приписать неисправным телевизорам, и в качестве элементарных исходов принять тройки номеров выбранных телевизоров, то ПЭИ будет состоять из довольно большого числа элементов (позже убедимся в этом). Ясно, что выписывание всех этих исходов — путь неприемлемый. Но в главе 3 у нас появился такой инструмент для решения подобных задач, как комбинаторика. Вот и воспользуемся им.

Рассматривается опыт, состоящий в выборе трех элементов из совокупности 1, 2, ..., 10. Исходами опыта будем считать наборы троек номеров. Выписывать их не будем, а подсчитаем их число. Воспользуемся схемой решения комбинаторных задач, приведенной в § 3.3. Выпишем несколько примеров результатов выбора: (1, 5, 6), (2, 3, 9). Ясно, что это выборки неупорядоченные (результаты (1, 5, 6) и (5, 1, 6) ничем не отличаются друг от друга), без возвращения (незачем дважды проверять один и тот же телевизор), образованы из 10 различных элементов и содержат три элемента. Их число равно $C_{10}^3 = \dfrac{10 \cdot 9 \cdot 8}{3!} = 120$. Все эти исходы можно считать равновозможными и каждому можно приписать вероятность, равную $\dfrac{1}{120}$.

Очень трудоемко каждому из 120 исходов ставить в соответствие число неисправных телевизоров. Ясно, что число X неисправных телевизоров среди трех выбранных может принимать значения 0, 1, 2, 3. Теперь нужно вычислить вероятности событий $(X = 0)$, $(X = 1)$, $(X = 2)$, $(X = 3)$. Можно обратиться к классической вероятности. Число всех исходов подсчитано. Они равновозможны. Подсчитаем число исходов, образующих каждое из этих событий.

Событие $(X = 0)$ происходит тогда и только тогда, когда все три телевизора выбираются из шести исправных, оно состоит из неупорядоченных выборок без возвращения из шести элементов 1, 2, ..., 6 (номера исправных телевизоров) по три. Их число равно $C_6^3 = \dfrac{6 \cdot 5 \cdot 4}{3!} = 20$. Аналогично событие $(X = 3)$ наступает тогда и только тогда, когда все телевизоры выбираются из четырех неисправных, оно состоит из неупорядоченных выборок без возвращения из четырех элементов 7, 8, 9, 10 (номера неисправных телевизоров) по три. Их число равно $C_4^3 = C_4^1 = 4$. Событие $(X = 1)$ происходит тогда и только тогда, когда один телевизор выбирается из четырех неисправных (число способов выбора равно $C_4^1 = 4$) и два телевизора — из шести исправных (число способов выбора равно $C_6^2 = \dfrac{6 \cdot 5}{2!} = 15$). Применяя комбинаторное правило умножения, получим, что событию $(X = 1)$ благоприятствует $C_4^1 = C_6^2 = 4 \cdot 15 = 60$ исходов. Аналогично событие $(X = 2)$ состоит из $C_4^2 \cdot C_6^1 = \dfrac{4 \cdot 3}{2} \cdot 6 = 36$ исходов. Применяя теперь формулу классической вероятности, будем иметь:

$$P(X = 0) = \frac{C_6^3}{C_{10}^3} = \frac{20}{120} = \frac{1}{6}; \quad P(X = 1) = \frac{C_6^2 \cdot C_4^1}{C_{10}^3} = \frac{60}{120} = \frac{1}{2};$$

$$P(X = 2) = \frac{C_6^1 \cdot C_4^2}{C_{10}^3} = \frac{36}{120} = \frac{3}{10}; \quad P(X = 3) = \frac{C_4^3}{C_{10}^3} = \frac{4}{120} = \frac{1}{30}.$$

Как проверить правильность вычислений? Подсчитаем сумму полученных вероятностей $\dfrac{1}{6} + \dfrac{1}{2} + \dfrac{3}{10} + \dfrac{1}{30}$. Она равна 1. Это говорит в пользу правильности решения задачи.

Ответ:

x_i	0	1	2	3
p_i	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

■

При решении этой задачи в качестве элементарных исходов опыта, состоящего в выборе трех телевизоров из 10, можно было принять события $(X = 0)$, $(X = 1)$, $(X = 2)$, $(X = 3)$: ведь при построении вероятностной модели случайного опыта в главе 2 мы говорили о том, что элементарные исходы опыта можно выбирать различными способами. Тогда в роли элементарных вероятностей выступают числа $\dfrac{1}{6}, \dfrac{1}{2}, \dfrac{3}{10}, \dfrac{1}{30}$. Получена новая вероятностная модель рассматриваемого опыта. Эта модель «беднее» событиями той, где в качестве элементов ПЭИ выступают неупорядоченные выборки без возвращения из 10 элементов по три. В последней модели выбор тройки (3, 6, 8) является событием, а в рамках новой модели это не является событием, так как не является подмножеством нового ПЭИ.

Как известно, ПЭИ вместе с вероятностями элементарных событий есть не что иное, как вероятностная модель эксперимента. Закон распределения случайной величины X фактически является новой вероятностной моделью опыта. Роль ПЭИ играют события $(X = x_1)$, $(X = x_2)$, ..., $(X = x_n)$, где $x_1, \ldots, x_n$ — значения случайной величины X, а в роли элементарных вероятностей выступают вероятности наступления событий $(X = x_k)$, равные $p_k = P(X = x_k)$, $k = 1, 2, \ldots, n$.

? Представьте закон распределения числа выпавших гербов при трехкратном подбрасывании правильной монеты в виде вероятностной модели соответствующего случайного опыта.

Контрольные вопросы

1. Приведите примеры случайных величин и случайных событий, связанных со следующими опытами:

а) игральный кубик бросают трижды;

б) четыре лампочки, среди которых есть и бракованные, последовательно вкручиваются в патрон до обнаружения недефектной;

в) монета подбрасывается четыре раза;

г) из ящика с тремя белыми и двумя черными шарами наугад вынимаются два шара;

д) из шести спичек, среди которых у одной сломана головка, наугад вытягивают одну.

2. Какая из таблиц задает закон распределения случайной величины?

а)

x	0	1	2	3
p	0,1	0,5	0,1	0,3

б)

x	1	2	3	4
p	0,3	0,1	0,2	0,3

в)

x	1	1	2	3
p	0,1	0,2	0,3	0,4

г)

x	1	2	3	4
p	0,4	0,3	0,2	0,2

3. Пусть a и b — соответственно наименьшее и наибольшее значения случайной величины X. Чему равна $P(a \leqslant X \leqslant b)$?

4. Пусть $P(X = 3) = 0{,}7$. Верно ли равенство $P(X \geqslant 3) = 0{,}6$?

5. Пусть $P(X = 3) = 0{,}7$. В каком случае верно равенство:

а) $P(X \geqslant 3) = 0{,}7$;

б) $P(X \leqslant 3) = 0{,}7$?

Задачи

335.° На рис. 45 показано распределение случайной величины X.

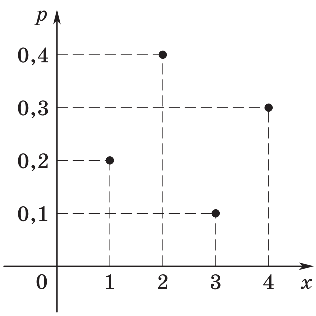


Рис. 45

а) Запишите закон ее распределения в виде таблицы.

б) Найдите $P(X > 1)$; $P(X \leqslant 2)$; $P(1{,}5 < X < 4{,}1)$.

336. Дан закон распределения случайной величины X:

x	-1	0	1	2
p	$0{,}3a^2$	a	$0{,}1a^2$	$0{,}4$

а) Найдите a; $P(X \leqslant 0)$; $P(X \geqslant 0)$.

б) Составьте закон распределения для X^2; $(X - 0{,}5)^2$.

337. Составьте законы распределения:

а)° числа гербов, выпавших при двух подбрасываниях монеты;

б) суммы очков, выпавших при бросании двух правильных кубиков;

в) квадрата числа очков, выпавших при бросании игрального кубика;

г) разности очков, выпавших при первом и втором бросках игрального кубика.

338.° В урне пять белых и 25 черных шаров. Наугад вынимают один шар. Составьте закон распределения числа вынутых белых шаров.

339. Составьте закон распределения числа попаданий мячом в корзину при двух бросках, если вероятность попадания при каждом броске равна 0,7.

340. Длительными наблюдениями установлено, что некоторый рабочий на протяжении рабочего дня может изготовить не более двух бракованных деталей. При этом на протяжении 100 рабочих дней брака не бывает приблизительно в 60 случаях, одна бракованная деталь встречается в 30, а две — в 10 случаях. Составьте закон распределения числа бракованных деталей, изготовленных рабочим:

а)° за сутки; б)* за двое суток.

341. Круговая мишень разделена на три одинаковых сектора. Она установлена так, что может вращаться вокруг горизонтальной оси. При довольно большой угловой скорости вращения мишени стрелок не может видеть цифр, вписанных в секторы. Поэтому он стреляет наугад. При любом выстреле стрелок попадает в мишень, причем при попадании в первый сектор он выиграет 1 денежную единицу, во второй сектор — 2 денежные единицы, и в третий сектор — 4 денежные единицы. За право сделать один выстрел он платит 2,5 денежной единицы. Составьте закон распределения выигрыша стрелка при:

а)° одном выстреле; б)* двух выстрелах.

342. Рабочий обслуживает три станка, размещенных на одной прямой на расстоянии a друг от друга. Станки обслуживаются при отказе в их работе. Вероятность отказа каждого станка равна $\frac{1}{3}$, причем станки выходят из строя независимо друг от друга. В начале работы рабочий находится возле среднего станка. После ремонта какого-либо станка рабочий остается возле него. Составьте закон распределения длины перехода, осуществляемого рабочим для ликвидации:

а)° первого отказа; б)* первых двух отказов.

§ 4.2. Математическое ожидание случайной величины

Закон распределения случайной величины содержит всю вероятностную информацию о ней, т. е. он дает возможность вычислить вероятность любого события, связанного с этой случайной величиной. Однако в ряде случаев исключительно полезными бывают некоторые постоянные числовые характеристики, дающие представление о случайной величине. Среди таких характеристик особенно большую роль играет ***математическое ожидание***.

Пример 1. Пусть число очков, выбиваемых при одном выстреле каждым из двух стрелков, имеет соответственно закон распределения:

x	8	9	10
p	0,4	0,1	0,5

x	8	9	10
p	0,1	0,6	0,3

Какой стрелок стреляет лучше?

Только по виду законов распределения затруднительно ответить на этот вопрос: восемь очков первый стрелок выбивает чаще, чем второй, вероятность попадания в «десятку» у него также выше, но в «девятку» чаще попадает второй. Нужна числовая характеристика, которая помогла бы оценить качество стрельбы каждого стрелка. ■

Чтобы получить представление о виде этой числовой характеристики, рассмотрим следующий пример.

Пример 2. Выпущено 500 лотерейных билетов, причем на 40 билетов выпадает выигрыш по 10 р., 10 билетов принесут их владельцам выигрыш по 50 р., 5 билетов — по 100 р. Остальные билеты безвыигрышные. Какой средний выигрыш выпадает на один билет?

❒ Средний выигрыш, приходящийся на один билет, равен сумме всех выигрышей, деленной на число выпущенных билетов:

$$\frac{10\cdot 40+50\cdot 10+100\cdot 5+0\cdot 445}{500}=$$
$$=10\cdot\frac{40}{500}+50\cdot\frac{10}{500}+100\cdot\frac{5}{500}+0\cdot\frac{445}{500}=$$
$$=0\cdot 0{,}89+10\cdot 0{,}08+50\cdot 0{,}02+100\cdot 0{,}01=2{,}8\text{ (р.)}.$$

В этом примере речь идет о случайной величине — выигрыше, выпадающем на один выбранный наугад лотерейный билет. Ее закон распределения

x	0	10	50	100
p	0,89	0,08	0,02	0,01

Обратите внимание, что средний выигрыш равен сумме произведений значений случайной величины на вероятности, с которыми она принимает эти значения. ■

Обобщение этого наблюдения позволяет определить среднее значение произвольной случайной величины X, имеющей закон распределения

x	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Сумму произведений значений случайной величины на соответствующие вероятности называют ***математическим ожиданием (или средним значением) случайной величины X.***

Математическое ожидание, или среднее значение, дискретной случайной величины представляет собой некоторое определенное число, характеризующее типичное значение этой величины, подобно тому, как некоторый набор данных характеризуется средним арифметическим значением.

Обозначается математическое ожидание символом MX:

$$MX = x_1p_1 + x_2p_2 + \ldots + x_np_n, \quad \text{или} \quad MX = \sum_{i=1}^{n} x_i p_i.$$

Теперь можно ответить на вопрос, заданный в примере 1. Если X_i $(i = 1, 2)$ — число очков, выбиваемых i-м стрелком при одном выстреле, то

$$MX_1 = 8 \cdot 0{,}4 + 9 \cdot 0{,}1 + 10 \cdot 0{,}5 = 9{,}1;$$
$$MX_2 = 8 \cdot 0{,}1 + 9 \cdot 0{,}6 + 10 \cdot 0{,}3 = 9{,}2.$$

Среднее число очков, выбиваемых при одном выстреле вторым стрелком, несколько выше, чем у первого. Поэтому естественно признать его лучшим.

Можно заметить, что и по смыслу, и по своему виду математическое ожидание случайной величины имеет много общего со средним арифметическим статистических данных. В главе 1 с помощью среднего арифметического сравнивались стили журналистов и писателей, строгость преподавателей, другими словами, оценивались параметры, проверялись гипотезы. Ту же роль для случайных величин играет математическое ожидание. Похожи и формулы для вычисления математического ожидания и среднего арифметического:

$$MX = \sum_{i=1}^{n} x_i p_i; \quad \bar{x} = \sum_{i=1}^{m} x_i \frac{n_i}{n}.$$

Есть сходство и в терминах: среднее арифметическое и среднее значение случайной величины. Кстати, последний термин лучше отражает сущность понятия. Термин «математическое ожидание» является данью традиции, он восходит ко временам азартных игр, где игроков интересовало *математическое ожидание выигрыша*.

Различие этих показателей состоит в том, что среднее арифметическое является обобщающим показателем для наблюдаемых значений случайной величины, а математическое ожидание — для всех возможных.

Понятие математического ожидания является обобщением понятия среднего арифметического. К связи между этими понятиями мы будем возвращаться неоднократно и в этой, и в следующей главах.

? Чему равно математическое ожидание случайной величины, все значения которой имеют одинаковую вероятность?

Как можно сравнить:
а) загруженность работой двух однотипных мастерских, имея информацию о количестве заказов, поступающих в эти мастерские;
б) качество работы двух рабочих, имея информацию о количестве бракованных деталей, производимых ими;
в) безопасность движения на двух перекрестках, имея информацию о количестве аварий, случающихся на них;
г) возможность выиграть в двух лотереях, имея информацию о размерах выигрышей в каждой из них?

Мы видели, что для вычисления математического ожидания случайной величины нужно знать ее закон распределения. На практике установить закон распределения случайной величины (т. е. указать все ее значения и соответствующие вероятности) часто невозможно. Приходится довольствоваться результатами большого числа независимых наблюдений за этой величиной, проведенных примерно в одинаковых условиях. Такая совокупность наблюдений является ***выборкой из значений рассматриваемой случайной величины***.

Пример 3. Куплено 500 лотерейных билетов (не выпущено, а куплено!). На 40 билетов выпал выигрыш по 10 р., 10 билетов принесли их владельцам выигрыш по 50 р., 5 билетов — по 100 р. Остальные билеты оказались безвыигрышными. Найти средний выигрыш, выпавший на один билет.

❐ Прежде всего обращаем внимание на различия в условиях примеров 2 и 3. Если в примере 2 были известны все значения, которые может принимать случайная величина — выигрыш, выпадающий на один билет во всей лотерее (в рублях), то в примере 3 речь идет о выборке из 500 наблюдений за этой случайной величиной. Эта величина 445 раз приняла значение 0, 40 раз — 10, 10 раз — 50 и 5 раз — 100. Среднее арифметическое выигрыша на один билет равно 2,8 р. (см. вычисления в примере 2). ■

Вообще, пусть в выборке из n наблюдений за случайной величиной X эта величина n_1 раз приняла значение x_1, n_2 раз — значение x_2, ..., n_m раз — значение x_m. Таблица относительных частот событий $(X = x_i)$, $i = 1, 2, \ldots, m$, имеет вид таблицы 4.5.

Таблица 4.5

Значение	x_1	x_2	...	x_m
Относительная частота	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_m}{n}$

Выборочным средним *называют величину*

$$\bar{x} = x_1\frac{n_1}{n} + x_2\frac{n_2}{n} + \ldots + x_m\frac{n_m}{n}, \quad \text{или} \quad \bar{x} = \frac{1}{n}\sum_{k=1}^{m} x_k n_k.$$

Как видим, выборочное среднее — это не что иное, как среднее арифметическое статистических данных. Приведенная выше таблица распределения частот значений случайной величины в выборке похожа на закон распределения случайной величины. Выражения для $\bar{x}$ и MX хотя и похожи, но не идентичны. Дело в том, что отношения $\frac{n_k}{n}$ изменяются от выборки к выборке. Однако при достаточно большом числе наблюдений $\frac{n_k}{n} \approx p_k$ (вероятности того, что $X = x_k$). Поэтому

$$\bar{x} \approx MX. \qquad (1)$$

Это приближение тем точнее, чем больше n.

Приближенное равенство (1) позволяет, с одной стороны, по значению математического ожидания предсказать среднее значение случайной величины в достаточно большом числе экспериментов.

Результат, полученный при решении примера 1, означает, что в результате 100 выстрелов, произведенных примерно в одних и тех же условиях, стрелки вправе ожидать примерно 910 и 920 очков соответственно.

Обратимся снова к примеру 2. Понятно, что никакой выигрыш не может быть равным 2,8 р. Полученный результат означает, что при покупке, например, 100 билетов (число билетов должно быть достаточно велико) выигрыш составит примерно 280 р.

? Известно, что математическое ожидание некоторой случайной величины равно 12,5. Что можно сказать о значениях этой случайной величины, которые она примет в 100 опытах, в которых наблюдается эта случайная величина?

С другой стороны, равенство (1) позволяет приближенно оценить математическое ожидание по выборочному среднему.

Пример 4. При сборке прибора для точной подгонки некоторой детали требуется произвести ряд проб, причем деталь, забракованная при сборке одного прибора, уже не используется при сборке других. Для установления числа деталей, которыми необходимо снабдить сборщика, было проведено 100 наблюдений. Оказалось, что в семи случаях понадобилась только одна проба, в 16 — две, в 55 — три, в 21 — четыре и в одном случае — пять проб. Найти среднее число деталей, необходимое для сборки одного прибора.

❒ Ответом на вопрос задачи служит математическое ожидание случайной величины — числа деталей, необходимых для сборки одного прибора. По результатам 100 наблюдений можно подсчитать выборочное среднее числа деталей, затрачиваемых для сборки одного прибора:

$$\bar{x} = \frac{1}{100}(1 \cdot 7 + 2 \cdot 16 + 3 \cdot 55 + 4 \cdot 21 + 5 \cdot 1) = 2{,}93.$$

Таким образом, данному сборщику для сборки одного прибора требуется примерно три детали. ■

? Как можно оценить среднюю потребность в бензине для обслуживания машин «скорой помощи» в некотором регионе в определенный промежуток времени?

Как можно оценить среднюю потребность в определенной детали для ремонта автомобилей некоторого класса в заданном регионе?

В дальнейшем для обоснования некоторых свойств математического ожидания случайной величины нам понадобится другое выражение для математического ожидания.

❒ Для его вывода обратимся снова к определению математического ожидания:

$$MX = \sum_{k=1}^{n} x_k p_k.$$

Значения x_k случайной величины X являются значениями функции $X(u)$ от элементов u ПЭИ. Ясно, что одно и то же значение x_k случайная величина X может принимать при различных исходах опыта. В примере 2 один и тот же размер выигрыша выпадает на различные лотерейные билеты. Пусть, например, событие $(X = x_1)$ состоит из исходов $u_1, \ldots, u_k$, т. е. $X(u_1) = \ldots = X(u_k) = x_1$. Тогда $p_1 = P(X = x_1) = p(u_1) + \ldots + p(u_k)$. Отсюда

$$x_1 p_1 = x_1(p(u_1) + \ldots + p(u_k)) = x_1 p(u_1) + \ldots + x_1 p(u_k) =$$
$$= X(u_1)\, p(u_1) + \ldots + X(u_k)\, p(u_k).$$

Аналогичные соотношения имеют место для всех событий $(X = x_k)$ и их вероятностей p_k, причем различные события $(X = x_k)$ и $(X = x_i)$, $k \neq i$, составлены из различных исходов ПЭИ. Поэтому для математического ожидания случайной величины справедливо следующее соотношение:

$$MX = \sum_{k=1}^{N} X(u_k)\, p(u_k), \qquad (2)$$

где $u_1, \ldots, u_N$ — все элементы ПЭИ, $p(u_1), \ldots, p(u_N)$ — их элементарные вероятности. ■

? Случайная величина X задана на ПЭИ $U = \{u_1, u_2, u_3\}$, причем $X(u_1) = 1$, $X(u_2) = 0$, $X(u_3) = 1$. Вероятности элементарных исходов u_1 и u_3 одинаковы и вдвое меньше, чем вероятность исхода u_2. Найдите MX.

Понятие математического ожидания было введено в науку в XVII веке под названием «справедливая цена шанса». Фактически впервые его ввел и использовал Х. Гюйгенс (1629–1695). Он считал, что математическое ожидание – это цена шанса на выигрыш в «справедливой» игре и пришел к выводу, что справедливая цена – это средняя цена. Сам Гюйгенс не называет математическое ожидание ожиданием, оно у него фигурирует как стоимость шанса. Впервые термин «ожидание» появляется в переводе работы Гюйгенса, который выполнил Ф. Схоутен (1615–1660). Математическое ожидание является обобщением понятия среднего арифметического. Оно широко использовалось в торговле и промышленности для определения средних цен, средней прибыли и т. п.

Контрольные вопросы

1. Известны законы распределения числа бракованных изделий, выпущенных каждым из двух цехов за смену. Как можно сравнить качество работы этих цехов в рассматриваемую смену?
2. Случайная величина принимает два значения 0 и 1. Чему равно ее математическое ожидание?
3. Куплено два билета некоторой лотереи. Можно ли по размерам выигрышей, выпавших на эти билеты, судить о среднем выигрыше, выпадающем на один билет в этой лотерее?
4. Является ли выборочное среднее случайной величиной или нет?
5. Известно, что $MX = 3{,}7$. Что можно сказать о значениях, которые примет случайная величина X в 100 опытах?

Задачи

343.° Для некоторого дня вероятность отсутствия заказов равна 30%, вероятность поступления одного заказа — 50%, вероятность поступления двух заказов — 15%, а вероятность поступления трех заказов — 5%. Найдите среднее число ожидаемых заказов.

344.° Вычислите математическое ожидание числа приборов:

а) вышедших из строя;

б) не вышедших из строя за время испытаний на надежность, если испытанию подвергается один прибор, а вероятность его отказа равна p.

345.° Время, в течение которого система находится в неработающем состоянии, описывается приведенным в таблице 4.6 распределением.

Таблица 4.6

Неисправность	Время простоя, мин	Вероятность
Незначительная	5	0,60
Существенная	30	0,30
Катастрофическая	120	0,10

а) Найдите среднее значение времени простоя по причине неисправности.

б) Чему равна вероятность того, что время простоя превысит 10 мин?

346.° Кредитный отдел банка установил, что один из заемщиков испытывает финансовые затруднения и, возможно, не сможет произвести текущий платеж, срок которого истекает на следующей неделе. Руководство отдела считает, что с вероятностью 60% он внесет всю подлежащую выплате сумму в 100 000 р., с вероятностью 30% внесет только половину, а с вероятностью 10% не внесет ничего. Найдите среднюю ожидаемую кредитную выплату.

347. Предположим, что выпущено 100 000 билетов денежной лотереи. Разыгрывается два выигрыша по 5000 р., восемь — по 1000 р., 170 — по 100 р., 350 — по 50 р. и 750 выигрышей — по 10 р. Вычислите «справедливую цену» одного билета.

348. Проводят независимые испытания, в каждом из которых с вероятностью 0,8 может произойти некоторое событие A. Испытания проходят до первого появления события A; общее число испытаний не превосходит четырех. Найдите среднее число проведенных испытаний.

349. Правильную монету подбрасывают до тех пор, пока она не выпадет цифрой вверх, либо до трех последовательных выпадений герба. Найдите математическое ожидание числа подбрасываний при одном выполнении этого эксперимента.

350. Найдите математическое ожидание случайных величин, рассмотренных в задаче:

а) 337, а; б) 339; в) 340, а; г) 342, б.

351. Случайная величина X задана на ПЭИ $U = \{u_1, \ldots, u_6\}$, причем $X(u_1) = 1$, $X(u_2) = 0$, $X(u_3) = 1$, $X(u_4) = 1$, $X(u_5) = 0$. Вероятности элементарных исходов образуют арифметическую прогрессию, причем $p(u_1) = 0{,}3$. Найдите MX.

§ 4.3. Свойства математического ожидания

Мы уже знаем, что случайные величины, как и все числовые функции, можно складывать, перемножать, умножать на число. Довольно часто возникает необходимость вычислять среднее значение суммы, произведения случайных величин, если известны средние значения каждой из этих величин. Например, известно среднее число бракованных деталей, выпускаемых в цехе за один день. Как найти среднее число бракованных деталей, выпускаемых в этом цехе за 10 дней? Или известны средние выигрыши, выпадающие на один билет в двух лотереях. Как найти средний выигрыш, выпадающий владельцу одного билета первой лотереи и одного билета второй лотереи?

Ответы на эти и подобные вопросы помогут найти свойства математического ожидания. Эти свойства аналогичны ранее рассмотренным свойствам арифметического среднего.

СВОЙСТВО 1. *Для произвольной случайной величины X и произвольного числа C имеет место равенство*

$$M(CX) = CMX.$$

Интуитивно это свойство понятно: если X — число бракованных деталей, выпускаемых в цехе за один день, то CX — число бракованных деталей, выпускаемых в этом цехе за C дней. Понятно, что среднее значение CX в C раз больше среднего значения X. Сравните это свойство со свойством 4 среднего арифметического (§ 1.5): если все варианты умножить (разделить) на одно и то же число, то среднее арифметическое умножится (разделится) на то же число.

? Если случайная величина X означает число опечаток на одной странице книги, то что означает случайная величина $100X$?

Пример 1. Пусть случайная величина X имеет закон распределения

x_k	−2	−1	0	1	2
p_k	0,2	0,1	0,3	0,2	0,2

Вычислить $M(3X)$ двумя способами:

а) пользуясь свойством 1;

б) предварительно составив закон распределения случайной величины $3X$.

❐ а) $MX = -2 \cdot 0{,}2 + (-1) \cdot 0{,}1 + 0 \cdot 0{,}3 + 1 \cdot 0{,}2 + 2 \cdot 0{,}2 = 0{,}1;$

$$M(3X) = 3 \cdot MX = 3 \cdot 0{,}1 = 0{,}3.$$

б) Составим закон распределения случайной величины $Y = 3X$:

y_k	−6	−3	0	3	6
p_k	0,2	0,1	0,3	0,2	0,2

$$M(3X) = -6 \cdot 0{,}2 + (-3) \cdot 0{,}1 + 0 \cdot 0{,}3 + 3 \cdot 0{,}2 + 6 \cdot 0{,}2 = 0{,}3. \blacksquare$$

СВОЙСТВО 2. *Математическое ожидание постоянной величины С равно самой постоянной*:

$$MC = C.$$

И это свойство интуитивно понятно: среднее значение одного числа равно этому числу.

Иногда приходится вычислять среднее значение суммы двух (а порой и большего количества) случайных величин, средние значения которых известны. Пусть, например, на двух станках вырабатываются одинаковые изделия. Известно, что в среднем на первом станке ежедневно выпускается три бракованные изделия, а на втором — два. Достаточно ли этих данных, чтобы найти среднее значение количества бракованных изделий, которое следует ожидать ежедневно от обоих станков? Возможно, нужно еще что-то знать об этих случайных величинах, например, их законы распределения? Ответ на эти вопросы дает следующее свойство.

СВОЙСТВО 3. *Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий этих величин*:

$$M(X + Y) = MX + MY.$$

Смысл этого свойства фактически объяснен предыдущими рассуждениями. Сравните его со свойством 2 арифметического среднего (см. § 1.5): если каждое значение признака Z равно сумме (разности) значений признаков X и Y, то среднее арифметическое признака Z равно сумме (разности) средних арифметических признаков X и Y.

? Пусть случайная величина X_i, $i = 1, 2$, означает число очков, выпавших при бросании i-го игрального кубика. Что означает случайная величина $X_1 + X_2$?

Пример 2. В цехе установлены два станка, на которых изготовляют одинаковые детали. Число бракованных изделий, которые могут быть изготовлены на этих станках за сутки, имеет соответственно законы распределения:

x	0	1	2	3
p	0,6	0,2	0,15	0,05

y	0	1	2	3
p	0,5	0,25	0,2	0,05

а) Найти среднее число бракованных деталей, изготовленных на первом станке за 10 суток.

б) Каково среднее число изготовленных цехом за сутки бракованных деталей?

❐ Пусть X и Y — число бракованных деталей, которые изготовляются соответственно на первом и втором станках за сутки; $MX = 0{,}65$; $MY = 0{,}8$.

а) За 10 суток на первом станке изготовляется $10X$ бракованных деталей; $M(10X) = 10 \cdot MX = 6{,}5$.

б) Цех за сутки изготовляет $X + Y$ бракованных деталей; $M(X + Y) = MX + MY = 0{,}65 + 0{,}8 = 1{,}45$. ■

СВОЙСТВО 4. *Математическое ожидание отклонения случайной величины X от ее математического ожидания равно нулю:*

$$M(X - MX) = 0.$$

Сравните это свойство со свойством 6 среднего арифметического (см. § 1.5): сумма отклонений вариант от их среднего арифметического равна нулю.

? Верно ли утверждение, обратное свойству 4, т. е. верно ли, что если $M(X - a) = 0$, то $MX = a$?

СВОЙСТВО 5. *Если $X \geqslant 0$, то $MX \geqslant 0$.*

Смысл этого свойства ясен: если случайная величина принимает неотрицательные значения, то ее среднее значение также неотрицательно.

Справедливость свойства сразу следует из определения математического ожидания. Действительно, так как $X \geqslant 0$, т. е. случайная величина X принимает неотрицательные значения $x_1, x_2, \ldots, x_n$ с положительными вероятностями $p_1, p_2, \ldots, p_n$, то из определения математического ожидания $MX = x_1p_1 + x_2p_2 + \ldots + x_np_n$ следует, что $MX \geqslant 0$.

? Как вы понимаете утверждение $X > 0$? Означает ли оно, что случайная величина X принимает только положительные значения?

Справедливо ли утверждение, обратное свойству 5?

СВОЙСТВО 6. *Если $X \geqslant Y$, то $MX \geqslant MY$.*

Смысл этого свойства понятен: если при каждом элементарном исходе случайная величина X принимает значение, не меньшее, чем случайная величина Y, то среднее значение X не меньше среднего значения Y.

? Как вы понимаете утверждение $X \geqslant Y$? Означает ли оно, что каждое значение случайной величины X не меньше любого значения случайной величины Y?

Справедливо ли утверждение, обратное свойству 6?

Теперь приведем строгие доказательства рассмотренных свойств.

ДОКАЗАТЕЛЬСТВО СВОЙСТВА 1

❐ Пусть случайная величина X принимает значения $x_1, \ldots, x_n$ с вероятностями соответственно $p_1, \ldots, p_n$. Тогда величина CX при $C \neq 0$ принимает значения $Cx_1, \ldots, Cx_n$ с теми же самыми вероятностями. Поэтому

$$M(CX) = (Cx_1)\, p_1 + \ldots + (Cx_n)\, p_n = C(x_1p_1 + \ldots + x_np_n) = CMX.$$

Свойство справедливо и при $C = 0$. ■

ДОКАЗАТЕЛЬСТВО СВОЙСТВА 2

❐ Это свойство проверяется непосредственно: постоянную C, как отмечалось в § 4.1, можно рассматривать как случайную величину, принимающую одно значение C с вероятностью 1:

$$MC = C \cdot 1 = C. \blacksquare$$

ДОКАЗАТЕЛЬСТВО СВОЙСТВА 3

❐ Пусть случайные величины X и Y определены на ПЭИ $U = \{u_1, \ldots, u_m\}$. Обозначим $Z = X + Y$. В соответствии с равенством (2) § 4.2, $MZ = Z(u_1)\, p(u_1) + \ldots + Z(u_m)\, p(u_m)$. Поскольку сложение случайных величин осуществляется, как сложение функций, т. е. $Z(u_k) = X(u_k) + Y(u_k)$, то

$$MZ = (X(u_1) + Y(u_1))\, p(u_1) + \ldots + (X(u_m) + Y(u_m))\, p(u_m) =$$
$$= (X(u_1)\, p(u_1) + \ldots + X(u_m)\, p(u_m)) + (Y(u_1)\, p(u_1) + \ldots + Y(u_m)\, p(u_m)) =$$
$$= MX + MY. \blacksquare$$

Свойство 3 имеет место для суммы произвольного конечного числа слагаемых.

? Докажите свойство 3 для суммы трех слагаемых.

ДОКАЗАТЕЛЬСТВО СВОЙСТВА 4

❐ Обозначим MX через a. По свойствам 3 и 1 имеем $M(X - a) = MX - Ma$. Так как $a = MX$ является величиной постоянной, то по свойству 1 $Ma = a$. Итак, $M(X - a) = a - a = 0$. ■

➤

Математическое ожидание случайной величины $X - c$ часто называют первым моментом случайной величины X относительно c. Причина использования такого названия состоит в следующем. Выражению

$$(x_1 - c)p_1 + (x_2 - c)p_2 + \ldots + (x_n - c)p_n$$

можно придать физическую интерпретацию. Пусть имеется легкий, но абсолютно жесткий стержень, к которому в точке x_1 прикреплен груз весом p_1, в точке x_2 — груз весом p_2 и т. д., в точке x_n прикреплен груз весом p_n. На стержне выбран некоторый масштаб, а веса выражены в некоторой системе единиц. Тогда предыдущее выражение представляет собой сумму произведений веса каждого груза на длину «плеча» от точки приложения этого груза до точки c (рис. 46).

В физике приведенное выражение называют первым моментом системы масс (сил) относительно c. Если c является центром тяжести системы, то первый момент относительно c равен нулю. В этом случае система, на которую не действуют никакие силы, кроме сил тяжести, приложенных к изображенным на рисунке массам, будет находиться в равновесии, если она закреплена в точке c. Из свойства 4 следует, что если точка опоры помещается в точку $a = MX$, то первый момент относительно a равен нулю. Обратно, если $M(X - c) = 0$, то $c = MX$; поэтому среднее значение является единственной точкой, относительно которой первый момент равен нулю.

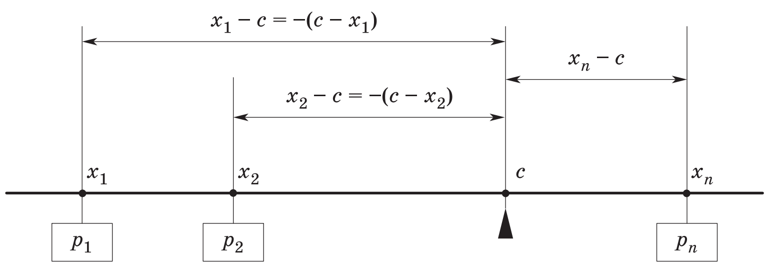


Рис. 46

Это одна из причин, по которой полезно использовать среднее значение для представления «расположения» значений случайной величины. Среднее значение не обязательно расположено на равных расстояниях от крайних значений случайной величины — подобно тому, как два мальчика, качающиеся на доске, располагают точку опоры не в середине доски, а ближе к тому из них, который тяжелее. Аналогично если «все вероятности» группируются вблизи одного из крайних значений случайной величины, то среднее значение случайной величины будет также расположено ближе к этому значению.

? Почему эквилибрист в цирке сохраняет равновесие, располагая точку опоры доски, на которой он стоит, прямо под центром тяжести?

ДОКАЗАТЕЛЬСТВО СВОЙСТВА 6

❐ Действительно, так как $X \geqslant Y$, то $X - Y \geqslant 0$, и согласно свойству 4 $M(X - Y) \geqslant 0$. Применяя свойства 3 и 1, получим $MX + M(-Y) \geqslant 0$, $MX - MY \geqslant 0$. Отсюда следует то, что требовалось доказать. ■

Контрольные вопросы

1. Как изменится математическое ожидание случайной величины, если ко всем ее значениям прибавить одно и то же число?

2. Как изменится математическое ожидание случайной величины, если все ее значения умножить на одно и то же число?

3. $MX = 3$. Чему равно $M(-X)$?

4. Верно ли, что математическое ожидание разности двух случайных величин равно разности их математических ожиданий?

5. Может ли случайная величина Y принимать большее значение, чем случайная величина X, если $MX > MY$?

Задачи

352.° Найдите математическое ожидание случайной величины Z, если:

а) $Z = 2X - 3Y$, $MX = 3$, $MY = 1$;

б) $Z = X + 3Y + 1$, $MX = 2$, $MY = 0$.

353.° Случайная величина X имеет закон распределения

x_k	1	2	3	4	5
p_k	0,2	0,3	0,2	0,2	0,1

Найдите: а) MX; б) $M(3X - 7)$; в) $M(10X)$; г) $M(X - 2{,}7)$.

354. Дан закон распределения случайной величины X:

x	−2	0	1	3
p	0,3	0,2	0,4	0,1

Найдите $M(-2X + 3)$:

а)° пользуясь свойствами математического ожидания;

б) составив сначала закон распределения случайной величины $-2X + 3$.

355. На некотором предприятии установлено n станков и с каждого станка отобрано по одному изделию. Определите среднее число бракованных изделий, если известно, что вероятность изготовления бракованного изделия:

а) на каждом станке равна p;

б) на первом станке равна p_1, на втором станке — p_2, ..., на n-м станке — p_n.

356. Случайные величины X_1, ..., X_n имеют одно и то же математическое ожидание a. Найдите MY, если $Y = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$.

357. Вероятности попадания мячом в кольцо при одном броске для двух баскетболистов соответственно равны 0,7 и 0,8. Найдите среднее число попаданий, если баскетболисты сделали:

а) по одному броску;

б) по 10 бросков.

358. Выигрыши, выпадающие на один билет в двух лотереях, имеют соответственно законы распределения:

x, р.	0	10	50	100
p	0,9	0,06	0,03	0,01

y, р.	0	10	50	100
p	0,85	0,12	0,02	0,01

а)° Какой из лотерей вы бы отдали предпочтение?

б)° Найдите средний выигрыш для обладателя 10 билетов в первой лотерее.

в) Какой средний выигрыш будет иметь человек, приобретший два билета первой лотереи и три билета второй?

359.* На двух столах расположено по две коробки с фантами. Коробки внешне абсолютно одинаковы. На первом столе в одной коробке имеется один фант, а в другой — семь фантов. На втором столе в одной коробке имеются два фанта, а в другой — пять фантов. Ребенок сначала выбирает стол, а затем наудачу берет коробку с этого стола. После того как коробка выбрана, игра начинается сначала и повторяется n раз. Какой стол лучше выбирать, чтобы в среднем за n игр получить большее число фантов?

360. Докажите, что для произвольной случайной величины X и любых чисел a и b имеет место равенство $M(aX + b) = aMX + b$.

§ 4.4. Формула Бернулли

В различных сферах деятельности мы встречаемся с последовательными испытаниями. Вот примеры таких испытаний: последовательные партии, сыгранные шахматистами; последовательная проверка изделий на длительность безотказной работы; заказы, поступающие в фирму; выпуск изделий. Перечень подобных примеров можно продолжать.

Математической моделью многих реальных опытов, представляющих собой последовательность нескольких независимых испытаний, являются так называемые *испытания Бернулли*.

Чтобы определить испытания Бернулли, рассмотрим примеры последовательных испытаний:

- четыре подбрасывания симметричной монеты;
- пять независимых выстрелов одного стрелка по мишени с вероятностью попадания при каждом выстреле, равной 0,8;
- шесть бросаний правильного игрального кубика;
- последовательное извлечение с возвращением трех шаров из урны, содержащей 10 шаров, среди которых четыре белых.

Эти испытания обладают следующими свойствами:

1. Они состоят из фиксированного числа опытов n (в приведенных примерах n соответственно равно 4, 5, 6, 3).

2. В результате каждого опыта может наступить («успех») или не наступить («неудача») некоторое случайное событие (в рассмотренных примерах в качестве «успеха» можно соответственно принять, например, выпадение герба, попадание в цель, выпадение «шестерки», извлечение белого шара).

3. Вероятность «успеха» от опыта к опыту не изменяется (в приведенных примерах она для указанных событий соответственно равна 0,5; 0,8; $\frac{1}{6}$; 0,4).

4. Испытания независимы, т. е. произвольные случайные события $A_1, A_2, \ldots, A_n$, связанные соответственно с 1-м, 2-м, ..., n-м опытом, независимы в совокупности. Другими словами, вероятность того или иного результата в каждом из этих испытаний не зависит от того, какие результаты наступили или наступят в остальных испытаниях.

Последовательные испытания, удовлетворяющие свойствам 1—4, называют ***испытаниями Бернулли.***

Приведем еще несколько примеров. Два шахматиста условились сыграть 10 партий. Вероятность выигрыша каждой отдельной партии первым игроком равна 0,6. Каждую партию можно считать отдельным испытанием. Всего производится 10 испытаний. Предполагается, что результат каждой сыгранной партии не влияет на результаты остальных партий. Эти испытания являются испытаниями Бернулли. Интерес представляет, например, вероятность того, что вся игра будет выиграна первым игроком.

Известно, что в некотором городе в течение месяца родилось 500 детей. Вероятность рождения мальчика при каждом рождении будем считать равной 0,5. Под испытанием здесь понимается рождение ребенка. Очевидно, что эти испытания можно считать независимыми. Снова имеем дело с испытаниями Бернулли. Конечно, рождение ровно 250 мальчиков является практически невозможным. Но разумно поставить вопрос, какова вероятность того, что число родившихся мальчиков окажется, например, между 225 и 275.

Однако не все последовательные испытания являются испытаниями Бернулли. Так, последовательные выстрелы 10 стрелков в мишень не являются испытаниями Бернулли, так как от испытания к испытанию меняются вероятности «успеха» (попадания в цель), ведь у различных стрелков могут быть различными вероятности попадания. Не удовлетворяет всем требованиям испытаний Бернулли и последовательное извлечение без возвращения трех шаров из урны, содержащей, например, четыре белых и шесть черных шаров, так как эти испытания зависимы. Последовательное извлечение с возвращением шаров из урны, содержащей, например, четыре белых и шесть черных шаров, до появления первого белого шара не являются испытаниями Бернулли, так как число испытаний не фиксировано, а случайно.

? Какие из следующих последовательных испытаний являются, и какие не являются испытаниями Бернулли?

а) Бросание трех игральных кубиков;

б) пять выстрелов одного стрелка в мишень;

в) проверка пяти деталей, изготовленных на одном станке одним рабочим;

г) опрос пары, состоящей из юноши и девушки, пришедших вместе в музей, относительно их мнения по поводу пяти картин, выставленных в музее;

д) подбрасывание правильной монеты до выпадения герба;

е) освещение елки десятью электрическими лампочками, изготовленными на одном заводе, включенными в цепь параллельно;

ж) обращение в фирму шести ее клиентов, одинаково часто пользующихся услугами фирмы.

В испытаниях Бернулли часто интерес представляет вопрос, какова вероятность того, что во всех этих испытаниях «успех» произойдет заданное число раз, или вероятность того, что число «успехов» находится в некоторых пределах. Например, статистические правила приема готовой продукции часто имеют следующий вид: если число бракованных изделий в проверяемой партии определенного объема не превышает заданного числа, то вся партия принимается, в противном случае — отправляется на переработку. Поэтому необходимо уметь вычислять вероятность того, что число бракованных изделий в проверяемой партии равно заданному числу.

Для ответа на этот вопрос получим формулу для вычисления вероятности того, что в n испытаниях Бернулли некоторое событие произойдет ровно m раз. Покажем идею вывода этой формулы сначала на примере.

Пример 1. Стрелок производит три независимых выстрела по мишени. Вероятность попадания при каждом выстреле равна 0,9. Найти вероятность того, что стрелок ровно два раза попадет в цель.

❒ ПЭИ рассматриваемого эксперимента имеет вид U = {ППП, ППН, ПНП, ПНН, НПП, НПН, ННП, ННН}. Здесь, например, запись ППН означает, что при первых двух выстрелах стрелок попал в мишень, а при третьем — не попал. Обратите внимание, что построенное ПЭИ состоит из неравновозможных исходов. Событие «стрелок ровно два раза попал в цель» содержит три исхода: ППН, ПНП, НПП. Число исходов, составляющих это событие, совпадает с числом способов выбора двух мест для буквы П из трех, т. е. с числом неупорядоченных выборок без возвращения из трех элементов по два — C_3^2. Вероятность каждого из этих исходов ввиду независимости испытаний равна $0{,}9^2 \cdot 0{,}1$. Искомая вероятность, которую мы обозначим $P_3(2)$, по определению вероятности события, равна

$$P_3(2) = 0{,}9^2 \cdot 0{,}1 + 0{,}9^2 \cdot 0{,}1 + 0{,}9^2 \cdot 0{,}1 =$$
$$= 3 \cdot 0{,}9^2 \cdot 0{,}1 = C_3^2 \cdot 0{,}9^2 \cdot 0{,}1 = 0{,}243.$$

Аналогично

$$P_3(0) = P(\text{ННН}) = 0{,}1^3 = C_3^0 \cdot 0{,}9^0 \cdot 0{,}1^3 = 0{,}001;$$
$$P_3(1) = P(\text{ПНН, НПН, ННП}) = 3 \cdot 0{,}9 \cdot 0{,}1^2 = C_3^1 \cdot 0{,}9^1 \cdot 0{,}1^2 = 0{,}027;$$
$$P_3(3) = P(\text{ППП}) = 0{,}9^3 = C_3^3 \cdot 0{,}9^3 \cdot 0{,}1^0 = 0{,}729.$$

Ясно, что

$$P_3(0) + P_3(1) + P_3(2) + P_3(3) =$$
$$= 0{,}1^3 + 3 \cdot 0{,}1^2 \cdot 0{,}9 + 3 \cdot 0{,}1 \cdot 0{,}9^2 + 0{,}9^3 = (0{,}1 + 0{,}9)^3 = 1. \blacksquare$$

Рассмотрим обобщение примера 1.

Пусть проводится n испытаний Бернулли, в каждом из которых может наступить один из двух взаимно исключающих друг друга исходов: У («успех») и Н («неудача»). Вероятности этих исходов в каждом испытании равны соответственно p и $q = 1 - p$. Тогда вероятность того, что произойдет ровно m «успехов», равна

$$P_n(m) = C_n^m p^m q^{n-m}, \quad m = 0, 1, 2, \ldots, n. \tag{1}$$

Равенство (1) носит название ***формулы Бернулли*** по имени швейцарского математика Я. Бернулли (1654—1705). Докажем эту формулу.

❒ ПЭИ n испытаний Бернулли состоит из выборок вида (УННУН...Н) длины n. Приведенная запись означает, что в первом испытании наступил «успех», во втором и третьем — «неудача», в четвертом — «успех» и т. д. Нас интересует событие «“успех” наступает ровно m раз ($0 \leqslant m \leqslant n$)». Это событие образуют те исходы, которые содержат m букв У и $n - m$ букв Н. Число исходов, образующих это событие, равно числу способов выбора m мест из n для буквы У, т. е. C_n^m. Вероятность наступления каждой такой выборки ввиду независимости испытаний равна $p^m q^{n-m}$ (вероятность пересечения взаимно независимых событий равна произведению вероятностей). Искомая вероятность по определению вероятности события равна сумме вероятностей исходов, образующих событие «ровно m “успехов” в n испытаниях», т. е.

$$P_n(m) = \underbrace{p^m q^{n-m} + p^m q^{n-m} + \ldots + p^m q^{n-m}}_{C_n^m \text{ слагаемых}} = C_n^m p^m q^{n-m}. \blacksquare$$

? Почему исходы (УННУН...Н), образующие ПЭИ, попарно несовместны?

Рассмотрим примеры на применение полученной формулы.

Пример 2. Расход электроэнергии на протяжении суток не превышает установленной нормы с вероятностью $p = \frac{3}{4}$. Найти вероятность того, что:

а) в ближайшие шесть суток расход энергии не будет превышать нормы в течение каких-либо четырех суток;

б) в ближайшие шесть суток расход энергии не будет превышать нормы менее четырех суток.

❒ а) Расход электроэнергии в течение суток считаем испытанием. «Успехом» в этом испытании является отсутствие перерасхода электроэнергии. Испытания независимы. Имеем:

$$P(\text{У}) = p = \frac{3}{4}, \quad P(\text{Н}) = q = 1 - \frac{3}{4} = \frac{1}{4}.$$

Требуется найти $P_6(4)$. Имеем

$$P_6(4) = C_6^4 p^4 q^2 = \frac{6 \cdot 5}{1 \cdot 2} \cdot \left(\frac{3}{4}\right)^4 \cdot \left(\frac{1}{4}\right)^2 \approx 0{,}297.$$

б) Событие «в ближайшие шесть суток расход энергии не будет превышать нормы менее четырех суток» означает, что расход энергии вообще не будет превышен в ближайшие шесть суток, или не будет превышен в течение каких-либо суток, или не будет превышен в течение каких-либо двух суток, или не будет превышен в течение каких-либо трех суток. Вероятность этого события по теореме сложения вероятностей для попарно несовместных событий (см. § 2.8) равна

$$P_6(0) + P_6(1) + P_6(2) + P_6(3) = C_6^0 p^0 q^6 + C_6^1 p^1 q^5 + C_6^2 p^2 q^4 + C_6^3 p^3 q^3 =$$

$$= \left(\frac{1}{4}\right)^6 + 6 \cdot \left(\frac{3}{4}\right) \cdot \left(\frac{1}{4}\right)^5 + 15 \cdot \left(\frac{3}{4}\right)^2 \cdot \left(\frac{1}{4}\right)^4 + 20 \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^3 \approx 0{,}169. \blacksquare$$

Пример 3. Два шахматиста условились сыграть 10 результативных партий. Вероятность выигрыша каждой отдельной партии первым игроком равна 0,6, вероятность выигрыша каждой отдельной партии вторым игроком равна 0,4 (ничьи не считаются). Какова вероятность того, что:

а) вся игра будет выиграна первым игроком;

б) будет достигнут общий ничейный результат;

в) вся игра будет выиграна вторым игроком?

❒ Выше мы уже объясняли, почему 10 партий, сыгранных шахматистами, являются испытаниями Бернулли.

а) Для того чтобы игру выиграл первый игрок, ему необходимо выиграть 6, 7, 8, 9 или 10 партий. Вероятность этого в силу формулы Бернулли и теоремы сложения вероятностей для попарно несовместных событий равна

$$P_{10}(6) + P_{10}(7) + P_{10}(8) + P_{10}(9) + P_{10}(10) =$$

$$= C_{10}^{6} \cdot 0{,}6^6 \cdot 0{,}4^4 + C_{10}^{7} \cdot 0{,}6^7 \cdot 0{,}4^3 + C_{10}^{8} \cdot 0{,}6^8 \cdot 0{,}4^2 +$$

$$+ C_{10}^{9} \cdot 0{,}6^9 \cdot 0{,}4^1 + C_{10}^{10} \cdot 0{,}6^{10} \cdot 0{,}4^0 =$$

$$= \frac{3^6}{5^{10}} \cdot (3360 + 2880 + 1620 + 540 + 81) \approx 0{,}633.$$

б) Вероятность ничейного результата равна

$$P_{10}(5) = C_{10}^{5} \cdot 0{,}6^5 \cdot 0{,}4^5 = \frac{252 \cdot 243 \cdot 32}{5^{10}} \approx 0{,}201.$$

в) Вероятность выигрыша игры вторым игроком равна

$$P_{10}(0) + P_{10}(1) + P_{10}(2) + P_{10}(3) + P_{10}(4) =$$

$$= C_{10}^{0} \cdot 0{,}6^0 \cdot 0{,}4^{10} + C_{10}^{1} \cdot 0{,}6^1 \cdot 0{,}4^9 + C_{10}^{2} \cdot 0{,}6^2 \cdot 0{,}4^8 +$$

$$+ C_{10}^{3} \cdot 0{,}6^3 \cdot 0{,}4^7 + C_{10}^{4} \cdot 0{,}6^4 \cdot 0{,}4^6 =$$

$$= \frac{2^6}{5^{10}} \cdot (16 + 240 + 1620 + 6480 + 17\,010) \approx 0{,}166.$$

Впрочем, эту вероятность можно было найти, воспользовавшись связью между вероятностями противоположных событий, т. е. вычитанием из единицы суммы вероятностей выигрыша игры первым игроком и ничейного результата: $1 - (0{,}633 + 0{,}201) = 0{,}166$. ■

? Монета подброшена четыре раза. Найдите вероятность выпадения:

а) ровно трех гербов;

б) не менее двух гербов.

При решении примера 1 мы фактически нашли закон распределения случайной величины X — числа попаданий при трех выстрелах:

x_i	0	1	2	3
p_i	0,001	0,027	0,243	0,729

? Найдите закон распределения случайной величины — числа суток в течение ближайших шести суток, на протяжении которых расход энергии не превышает нормы, рассмотренной в примере 2.

О случайной величине — числе «успехов» в n испытаниях Бернулли — говорят, что она имеет ***биномиальное распределение с параметрами n и p.***

Наличие двух возможных результатов в каждом испытании Бернулли и определяет приставку *би-* (*лат.* bi — «двух») в слове *биномиальное*.

Случайная величина — число попаданий при трех выстрелах, рассмотренная в примере 1, имеет биномиальное распределение с параметрами 3 и 0,9. Число суток в ближайшие шесть суток, в течение которых расход энергии не превышает нормы (пример 2), имеет биномиальное распределение с параметрами 6 и $\frac{3}{4}$.

? Какое распределение имеет число выигрышей первого шахматиста, рассмотренное в примере 3?

Планируется позвонить в восемь фирм с предложением продать им некоторую продукцию. Считается, что каждый такой звонок с вероятностью 20% приводит к продаже. Фирмы принимают решения о покупке независимо друг от друга.

а) О какой случайной величине можно говорить, что она имеет биномиальное распределение?

б) Каковы параметры этого биномиального распределения?

Закон биномиального распределения с параметрами n и p имеет вид

$$P(X = m) = C_n^m p^m q^{n-m}, \quad m = 0, 1, 2, \ldots, n.$$

Сумма $\sum\limits_{m=0}^{n} C_n^m p^m q^{n-m}$ представляет собой вероятность того, что число «успехов» в n испытаниях Бернулли равно или 0, или 1, или 2, ..., или n, т. е. вероятность достоверного события. Поэтому $\sum_{m=0}^{n} C_n^m p^m q^{n-m} = 1$.

С другой стороны,

$$\sum_{m=0}^{n} C_n^m p^m q^{n-m} =$$

$$= C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \ldots + C_n^n p^n q^0 = (p+q)^n = 1,$$

т. е. равенство $\sum_{m=0}^{n} C_n^m p^m q^{n-m} = 1$ является частным случаем биномиальной формулы Ньютона.

Опираясь на предыдущие выкладки, выведите формулу Ньютона

$$(a+b)^n = \sum_{m=0}^{n} C_n^m a^m b^{n-m}.$$

Для вычисления биномиальных вероятностей можно воспользоваться редактором электронных таблиц Microsoft Excel. Чтобы получить вероятность $P(X = a)$ того, что случайная величина X принимает значение, равное a, необходимо использовать формулу (=БИНОМРАСП(a; n; π; ЛОЖЬ)), а для вычисления вероятности $P(X \leqslant a)$ того, что случайная величина X принимает значение меньшее или равное a, использовать формулу (=БИНОМРАСП (a; n; π; ИСТИНА)). Вычисления для примера 2 приведены на рис. 47 и 48.

Здесь n — число испытаний Бернулли, π — вероятность «успеха» в каждом испытании. На рис. 47 $n = 6$, $a = 4$, $\pi = 0{,}75$. Вычисляется $P_6(4)$. На рис. 48 $n = 6$, $a = 3$, $\pi = 0{,}75$. Вычисляется $P(X \leqslant 3)$.

Выполните вычисления в примере 3 с помощью программы Excel.

В § 4.1 мы отмечали, что, как и для произвольной числовой функции, имеются различные способы задания закона распределения случайной величины: табличный, графический, аналитический.

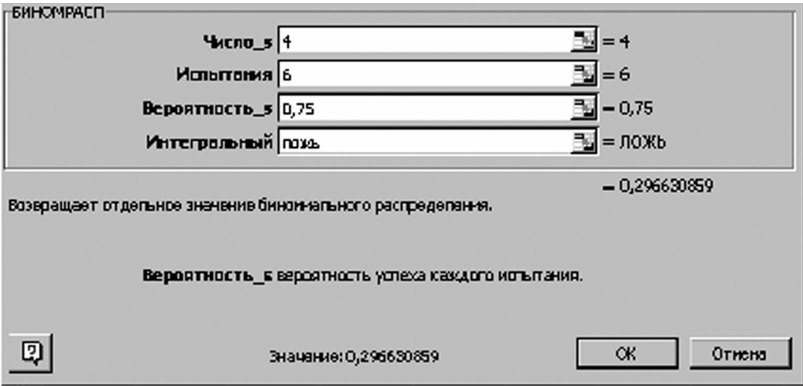


Рис. 47

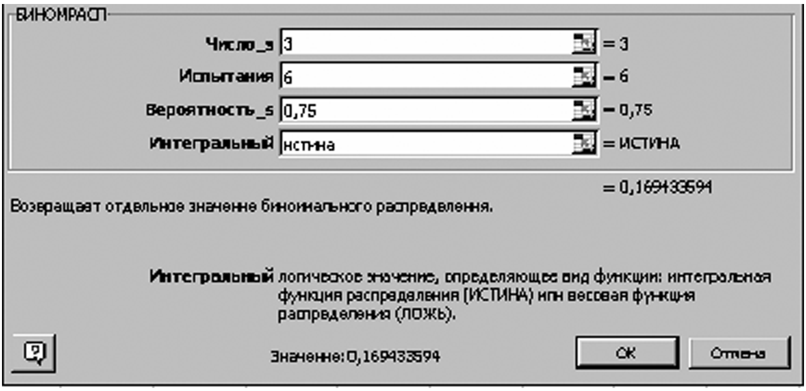


Рис. 48

Равенство (1) является примером аналитического задания закона распределения.

На рис. 49—51 приведены графики «закона биномиального распределения» при различных n и p.

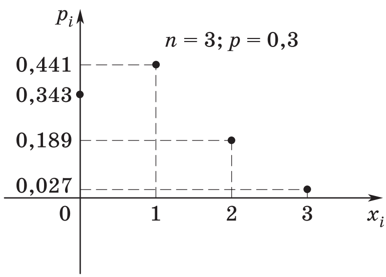


Рис. 49

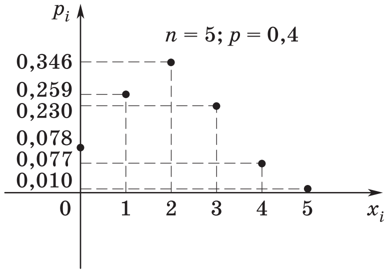


Рис. 50

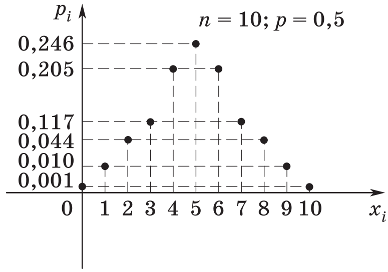


Рис. 51

? Постройте график закона биномиального распределения при $n = 6$ и $p = 0{,}75$, пользуясь результатами решения примера 2.

➤

Использование формулы Бернулли часто связано с колоссальными вычислениями. Если значения n велики, а p малы, то имеет место так называемая приближенная ***формула Пуассона***, названная по имени французского математика С. Пуассона (1781—1840):

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

где $\lambda \approx np$, $e \approx 2{,}72$. Формулу Пуассона применяют для нахождения вероятностей «редких» событий, точнее если $np(1 - p) \leqslant 9$. Подобные события могут порождаться такими случайными величинами, как число дефектов в произведенной продукции, число вызовов, поступающих на телефонную станцию, число заказов, поступающих на фирму в течение дня, и т. д. Другими словами, формула Пуассона, в частности, применяется в статистическом контроле качества, задачах надежности, расчете телефонных сетей.

Пример 4. Фирма выпустила 1000 изделий. Вероятность того что на гарантийный ремонт вернется определенное изделие, равна 0,0013. Какова вероятность того, что в определенный день на гарантийный ремонт:

а) не поступит ни одного изделия;
б) поступит одно изделие;
в) поступит два изделия;
г) поступит три изделия;
д) поступит не более двух изделий?

❐ В задаче речь идет об испытаниях Бернулли. Проведено $n = 1000$ испытаний; испытанием считаем получение информации об определенной детали, поступит она на гарантийный ремонт или нет. Вероятность «успеха» (на гарантийный ремонт поступит определенное изделие) равна $p = 0{,}0013$. Так как n велико, а p мало ($np(1-p) \approx 1{,}3 < 9$), то применима приближенная формула Пуассона. Применяя ее, получим:

$$P(X = 0) \approx e^{-1,3} \cdot \frac{1{,}3^0}{0!} \approx 0{,}272;$$

$$P(X = 1) \approx e^{-1,3} \cdot \frac{1{,}3^1}{1!} \approx 0{,}354;$$

$$P(X = 2) \approx e^{-1,3} \cdot \frac{1{,}3^2}{2!} \approx 0{,}230;$$

$$P(X = 3) \approx e^{-1,3} \cdot \frac{1{,}3^3}{3!} \approx 0{,}100.$$

Вероятность того, что на гарантийный ремонт поступит не более двух изделий, равна $P(X = 0) + P(X = 1) + P(X = 2) \approx 0{,}856$. ■

Для вычисления этих вероятностей с помощью программы Excel применяют функцию (=ПУАССОН(значение; среднее; ЛОЖЬ)), которая вычисляет некоторое конкретное значение $P(X = k)$, а также функцию (=ПУАССОН(значение; среднее; ИСТИНА)), вычисляющая вероятность $P(X \leqslant k)$ того, что случайная величина принимает значение меньшее или равное заданному. На рис. 52 приведены соответствующие вычисления. ◂

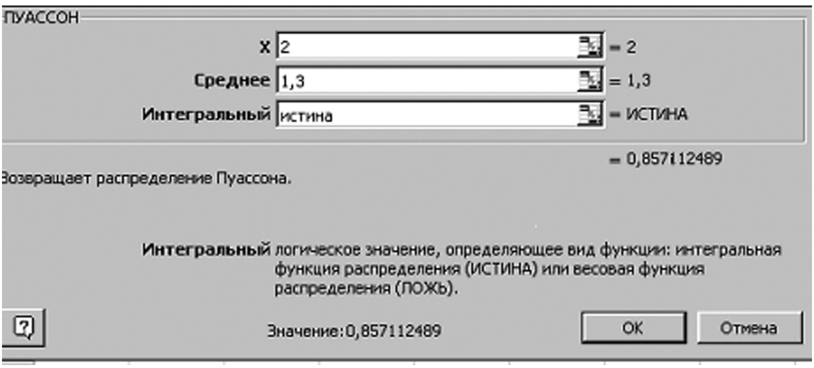


Рис. 52

Решите следующую задачу, пользуясь формулой Бернулли (с помощью Excel) и формулой Пуассона. Сравните результаты.

Вероятность изготовления бракованной детали равна 0,02. Найдите вероятность того, что среди 100 изготовленных деталей бракованных будет: 0, 1, 2, 3.

Формула Бернулли была получена Я. Бернулли (1654—1705) в его работе «Искусство предположений» (1713). Задача отыскания приближенной формулы для подсчета вероятности m «успехов» в n независимых испытаниях для больших значений n и малых значений p была решена в книге С. Пуассона (1781—1840) «Исследования о вероятности судебных приговоров по уголовным и гражданским делам» (1837). Польский статистик Л. Борткевич (1868—1931) в конце XIX назвал формулу Пуассона законом малых чисел. Борткевич применял формулу Пуассона к очень редко встречающимся явлениям: смерть от удара копытом лошади, рождение троен и т. п.

Контрольные вопросы

1. Верно ли, что буквой p в формуле Бернулли

$$P(m) = C_n^m p^m q^{n-m}$$

обозначают вероятность наступления хотя бы одного «успеха» в n испытаниях Бернулли?

2. При каком значении n вероятность «успеха» в каждом испытании совпадает с вероятностью хотя бы одного «успеха» в испытаниях Бернулли. Какая из этих вероятностей больше при остальных значениях n?

3. Как, пользуясь формулой Бернулли, найти вероятность того, что в n испытаниях Бернулли некоторое событие наступит n раз? Можно ли вычислить эту вероятность, не прибегая к формуле Бернулли?

4. Какие из следующих случайных величин имеют биномиальное распределение:

а) число извлеченных последовательно (без возвращения) нестандартных деталей из партии в 100 деталей, среди которых 10 нестандартных;

б) число попаданий в мишень при трех независимых выстрелах, если вероятность попадания при каждом выстреле одна и та же;

в) число очков, выпавших при бросании одного игрального кубика;

г) число гербов, выпавших при четырехкратном подбрасывании монеты?

5. Какие значения может принимать случайная величина, имеющая биномиальное распределение с параметрами n и p?

6. Почему случайные величины — число «успехов» в 1-м, 2-м, ..., n-м испытаниях Бернулли — имеют одинаковые законы распределения?

Задачи

361.° Изделия некоторого производства содержат 5% брака. Найдите вероятность того, что среди пяти взятых наугад изделий:

а) нет ни одного дефектного;

б) будут ровно два дефектных.

362.° Для прядения смешаны поровну белый и окрашенный хлопок. Какова вероятность среди пяти случайно выбранных волокон смеси обнаружить менее двух окрашенных?

363.° Наблюдениями установлено, что в некоторой местности в сентябре в среднем бывает 12 дождливых дней. Какова вероятность того, что из случайно взятых в этом месяце восьми дней ровно три дня окажутся дождливыми?

364. Проводится три испытания Бернулли. При каких значениях p имеет место равенство:

а) $P_3(0) = P_3(1)$; б) $P_3(0) = P_3(3)$?

365. В мастерской работает независимо друг от друга четыре мотора. Для каждого мотора вероятность перегрева к обеденному перерыву равна 0,8. Составьте закон распределения числа моторов, перегретых к перерыву.

366. Всхожесть ржи составляет 90%. Составьте закон распределения числа взошедших семян из пяти посеянных.

367. Эксперимент состоит в выборе с возвращением пяти шаров из урны, содержащей равное количество черных и белых шаров. Этот эксперимент повторяют 819 раз. Полученные результаты представлены в таблице 4.7.

Таблица 4.7

Число черных шаров	0	1	2	3	4	5
Наблюдаемая частота	30	125	277	224	136	27

Найдите соответствующие вероятности и сравните их с относительными частотами наблюдаемых значений.

368. Вероятность допустить ошибку при наборе некоторого текста, состоящего из 800 знаков, равна 0,005. Найдите вероятность того, что при наборе этого текста:

а) не будет допущено ошибок;

б) будет допущена ровно одна ошибка;

в) будут допущены ровно две ошибки;

г) будут допущены ровно три ошибки;

д) будет допущено не более двух ошибок.

§ 4.5. Дисперсия случайной величины

В главе 1 мы отмечали, что для характеристики статистических данных показателей центральной тенденции недостаточно. Они характеризуют концентрацию совокупности значений на числовой шкале и не позволяют учесть различия, существующие между отдельными значениями. Для измерения разброса, вариации значений внутри совокупности использовались различные показатели вариации и в первую очередь дисперсия статистической совокупности. Аналогичная ситуация имеет место при рассмотрении случайных величин, описывающих не выборочную, а генеральную совокупность.

Среднее значение случайной величины является важной, но ориентировочной ее характеристикой. Иногда практически важные свойства случайной величины не определяются ее средним значением, а требуют более детального изучения ее закона распределения.

Пусть случайная величина X — отклонение полученного значения измеряемой величины от ее истинного значения. Если погрешности измерений носят случайный характер, то они будут как положительными, так и отрицательными. Тогда, если их шансы приблизительно одинаковы, то математическое ожидание равно нулю и говорят, что измерения не имеют систематических ошибок. При этом, однако, неизвестно, как будут расположены в большинстве своем результаты измерений, близко к истинному значению измеряемой величины, или рассеяны далеко от него.

Итак, возникает необходимость в специальной числовой характеристике разброса значений случайной величины. Такой характеристикой и является ***дисперсия случайной величины***.

Рассмотрим еще один пример. Пусть испытываются на урожайность два сорта пшеницы. В зависимости от случайных обстоятельств (количество осадков, распределение удобрений, глубина вспашки и др.) урожайность подвержена значительным колебаниям и представляет собой случайную величину. Предположим, что средняя урожайность для каждого сорта одна и та же. Можно ли судить о качестве испытываемого сорта только по значению средней урожайности? Очевидно, нет, так как наибольший хозяйственный интерес представляет тот сорт, урожайность которого меньше подвержена случайным влияниям метеорологических и других факторов, иными словами, разброс урожайности меньше. Таким образом, при испытании того или иного сорта пшеницы на урожайность не меньшее значение, чем средняя урожайность, имеют возможные ее колебания. И вообще, часто показателем качества продукции является ее однородность.

Аналогично при анализе успеваемости по математике в двух классах, имеющих одинаковую среднюю успеваемость, часто предпочтение будет отдано тому классу, где разброс успеваемости меньше, т. е. наблюдается более стабильная успеваемость учащихся.

? Даны законы распределения двух случайных величин:

x	$-0{,}01$	$0{,}01$
p	$0{,}5$	$0{,}5$

y	-100	100
p	$0{,}5$	$0{,}5$

Чему равны их средние значения? Для какой из них значения относительно среднего значения рассеяны больше?

Для введения меры разброса значений случайной величины вокруг среднего рассмотрим следующий пример.

Пример 1. Пусть орудие ведет прицельный огонь по мишени, удаленной от нее на расстояние a км (рис. 53).

Если обозначить дальность стрельбы через X (км), то ее среднее значение, как правило, будет равно $MX = a$. Отклонение среднего значения от a свидетельствовало бы о наличии *систематической* погрешности стрельбы (систематического перелета или недолета снарядов), которую можно было бы устранить, изменив соответствующим образом наклон ствола орудия. Однако отсутствие систематической ошибки нисколько не гарантирует высокую точность стрельбы: чтобы оценить точность, необходимо также знать, насколько близко ложатся снаряды к цели. ■

? Верно ли, что в рассмотренной ситуации равенство $MX = a$ означает, что перелет снаряда встречается в среднем так же часто, как и недолет?

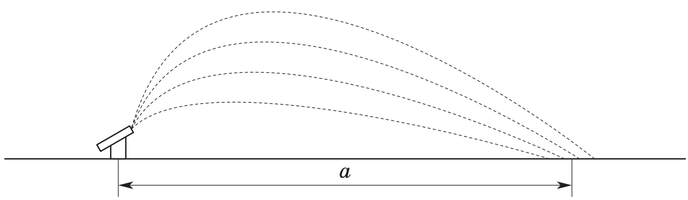


Рис. 53

Итак, нужно оценить отклонение случайной величины X от ее математического ожидания MX, т. е. случайную величину $X - MX$. В качестве меры такого отклонения естественно было бы взять величину $M(X - MX)$. Однако $M(X - MX) = 0$ (см. свойство 4 математического ожидания в § 4.3), и этот факт интуитивно понятен, ведь имеют место отклонения обоих знаков, их среднее равно нулю. Поэтому желательно было бы не учитывать знак отклонения. Это можно сделать, рассмотрев величину $M|X - MX|$ или $M(X - MX)^2$. Первую из них называют ***средним абсолютным отклонением***, вторую — ***дисперсией случайной величины***. Оказывается, что математически работать удобнее со второй величиной $M(X - MX)^2$. Позже мы увидим, почему следует отдать предпочтение ей по сравнению с другими числовыми характеристиками разброса. А именно, позже мы докажем, что дисперсия суммы двух независимых случайных величин равна сумме их дисперсий. Даже если случайные величины зависимы, для дисперсии суммы существует простое выражение. Другие характеристики разброса, в частности среднее абсолютное отклонение, таким свойством не обладают.

Дисперсией случайной величины *X называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания (обозначается дисперсия символом DX):*

$$DX = M(X - MX)^2.$$

? Может ли дисперсия случайной величины принимать отрицательные значения; обращаться в нуль?

Верно ли, что дисперсия случайной величины совпадает с квадратом самого большого отклонения значений случайной величины от ее среднего значения?

Дисперсию выражают в квадратах тех единиц, в которых измеряют саму случайную величину (или ее математическое ожидание). Поэтому наряду с дисперсией часто рассматривают характеристику, которую выражают в тех же единицах, что и случайную величину, и, также служащую мерой «разброса» ее значений.

Средним квадратичным отклонением случайной величины *называют корень квадратный из ее дисперсии:*

$$\sigma(X) = \sqrt{DX}.$$

Пример 2. Случайная величина X имеет закон распределения:

x	-3	-2	-1	0	1	2
p	0,1	0,1	0,2	0,1	0,2	0,3

Найти дисперсию и среднее квадратичное отклонение случайной величины X.

❐ Предварительно найдем MX:

$$MX = (-3) \cdot 0{,}1 + (-2) \cdot 0{,}1 + (-1) \cdot 0{,}2 + \\ + 0 \cdot 0{,}1 + 1 \cdot 0{,}2 + 2 \cdot 0{,}3 = 0{,}1.$$

Составим закон распределения случайной величины $Y = (X - 0{,}1)^2$:

y	$(-3{,}1)^2$	$(-2{,}1)^2$	$(-1{,}1)^2$	$(-0{,}1)^2$	$(0{,}9)^2$	$(1{,}9)^2$
p	0,1	0,1	0,2	0,1	0,2	0,3

Воспользовавшись определением, вычислим дисперсию и среднее квадратичное отклонение:

$$DX = M(X - 0{,}1)^2 = 9{,}61 \cdot 0{,}1 + 4{,}41 \cdot 0{,}1 + 1{,}21 \cdot 0{,}2 + \\ + 0{,}01 \cdot 0{,}1 + 0{,}81 \cdot 0{,}2 + 3{,}61 \cdot 0{,}3 = 2{,}89;$$

$$\sigma(X) = \sqrt{DX} = 1{,}7. \blacksquare$$

➤

При составлении закона распределения случайной величины $(X - a)^2$, где a — математическое ожидание случайной величины X, нужно учесть, что не всегда вероятность того, что $(X - a)^2$ принимает значение $(x - a)^2$, совпадает с вероятностью того, что случайная величина X принимает значение x. Например, если случайная величина X имеет закон распределения

x	-2	0	2	4
p	0,25	0,25	0,25	0,25

,

то $MX = 1$ и $P((X - 1)^2 = (2 - 1)^2) = P((X - 1)^2 = 1) \neq P(X = 2)$, так как $P(X = 2) = 0{,}25$, а $P((X - 1)^2 = 1) = P(X = 2) + P(X = 0) = 0{,}5$. Случайная величина $(X - 1)^2$ имеет закон распределения

x	1	9
p	0,5	0,5

Иногда, особенно в военном деле, для характеристики меры рассеяния употребляется так называемое ***срединное (вероятное) отклонение***, т. е. такое число, что отклонение $X - MX$ с одинаковой вероятностью может оказаться по модулю как больше, так и меньше этого числа.

Как мы уже упоминали, в качестве меры разброса значений случайной величины относительно среднего значения применяется ***среднее абсолютное значение*** $M|X - MX|$. ◄

В коммерческой деятельности, в сфере производства дисперсия и стандартное отклонение характеризуют риск, показывая, насколько неопределенной является ситуация.

Пример 3. Анализ различных сценариев развития экономики и их влияния на получение дохода фирмы привел к результатам, представленным в таблице 4.8.

Таблица 4.8

Сценарий развития экономики	Доход, млн денежных единиц	Вероятность
Прекрасный	10	0,20
Хороший	5	0,40
Нормальный	1	0,25
Плохой	−4	0,15

Найти среднее значение ожидаемого дохода, его дисперсию и среднее квадратичное отклонение.

❒ Обозначим через X доход фирмы.

$$MX = 10 \cdot 0{,}20 + 5 \cdot 0{,}40 + 1 \cdot 0{,}25 + (-4) \cdot 0{,}15 = 3{,}65;$$
$$DX = (10 - 3{,}65)^2 \cdot 0{,}20 + (5 - 3{,}65)^2 \cdot 0{,}40 +$$
$$+ (1 - 3{,}65)^2 \cdot 0{,}25 + (-4 - 3{,}65)^2 \cdot 0{,}15 \approx 19{,}3;$$
$$\sigma(X) \approx \sqrt{19{,}3} \approx 4{,}4.$$

Стандартное отклонение в размере 4 400 000 денежных единиц показывает, что в данном случае присутствует значительный риск. Доход вполне может оказаться на 4 400 000 денежных единиц выше или ниже среднего значения в 3 650 000 денежных единиц. ■

? Как сравнить качество стрельбы двух стрелков, если у них одинаковы средние значения числа выбитых очков?

Как сравнить качество измерений двух исследователей, если у них средняя величина ошибки одна и та же?

Для вычисления дисперсии можно получить другую формулу, которая часто упрощает вычисления.

Пусть $MX = a$, тогда $DX = M(X - a)^2 = M(X^2 - 2ax + a^2)$.

Используя свойства математического ожидания, получим

$$DX = MX^2 - 2aMX + a^2 = MX^2 - a^2 = MX^2 - (MX)^2.$$

Итак, доказано следующее свойство.

СВОЙСТВО 1. $DX = MX^2 - (MX)^2$.

Решим пример 2, воспользовавшись этим свойством.

❐ Закон распределения случайной величины $Y = X^2$ имеет вид

y	0	1	4	9
p	0,1	0,4	0,4	0,1

$$MX^2 = 0 \cdot 0{,}1 + 1 \cdot 0{,}4 + 4 \cdot 0{,}4 + 9 \cdot 0{,}1 = 2{,}9;$$
$$DX = 2{,}9 - (0{,}1)^2 = 2{,}89. \blacksquare$$

? Всегда ли вероятность того, что случайная величина X^2 принимает значение x^2, совпадает с вероятностью того, что случайная величина X принимает значение x?

СВОЙСТВО 2. *Если C — постоянная величина, то*

$$D(X + C) = DX; \quad D(CX) = C^2DX; \quad DC = 0.$$

Рассмотрим смысл этих равенств. Добавление к случайной величине постоянной C означает сдвиг всех ее значений (а значит, и среднего значения) на одну и одну и ту же величину C, поэтому разброс значений около среднего остается неизменным. Умножение случайной величины на C эквивалентно изменению масштаба измерений на C. При этом математическое ожидание умножается на C, а дисперсия — на C^2.

Смысл равенства $DC = 0$ состоит в том, что у случайной величины, принимающей одно значение C с вероятностью, равной 1, математическое ожидание равно C и разброса значений около среднего нет.

Докажем свойство 2:

$$❐\ D(X + C) = M((X + C) - M(X + C))^2 = M(X - MX)^2 = DX;$$
$$D(CX) = M(CX - M(CX))^2 = M(C(X - MX))^2 =$$
$$= C^2M(X - MX)^2 = C^2DX;$$
$$DC = M(C - MC)^2 = 0. \blacksquare$$

Следующий пример иллюстрирует, как доказанные свойства могут упростить вычисления дисперсии, если вычисления выполняются без применения вычислительных средств.

Пример 4. Случайная величина X имеет закон распределения

x	475	500	525
p	0,6	0,3	0,1

Найти DX.

❒ Случайная величина X принимает довольно большие значения. Вычитая из всех значений одно и то же число 500, мы фактически изменяем начало отсчета. Разделив все значения на 25, по существу изменим единицу масштаба. Затем обратными преобразованиями можно вернуться к исходной случайной величине.

Итак, введем случайную величину $Y = \dfrac{X-500}{25}$. Законы распределения Y и $Z = Y^2$ имеют соответственно вид:

y	−1	0	1
p	0,6	0,3	0,1

z	0	1
p	0,3	0,7

$$MY = -1 \cdot 0{,}6 + 0 \cdot 0{,}3 + 1 \cdot 0{,}1 = -0{,}5;$$
$$MY^2 = 1 \cdot 0{,}7 + 0 \cdot 0{,}3 = 0{,}7;$$
$$DY = MY^2 - (MY)^2 = 0{,}7 - (-0{,}5)^2 = 0{,}45.$$

Так как $X = 25Y + 500$, то, воспользовавшись свойством 2, получим

$$DX = D(25Y + 500) = D(25Y) = 625DX = 625 \cdot 0{,}45 \approx 281. \blacksquare$$

Как и для вычисления математического ожидания случайной величины, так и для вычисления ее дисперсии нужно знать ее закон распределения. На практике установить закон распределения случайной величины (т. е. указать все ее значения и соответствующие вероятности) часто невозможно. Приходится довольствоваться результатами большого числа независимых наблюдений за этой величиной, проведенных примерно в одинаковых условиях, т. е. *выборкой из значений рассматриваемой случайной величины*.

Выборочное среднее дает представление о математическом ожидании случайной величины. Аналогично по выборке можно определить приближенное значение дисперсии случайной величины.

Выборочной дисперсией *называют среднее всех квадратов отклонений результатов наблюдений от их среднего значения*:

$$s^2 = \sum_{i=1}^{m} (x_i - \bar{x})^2 \frac{n_i}{n},$$

где $x_1, \ldots, x_m$ — различные наблюдаемые значения случайной величины; $n_1, \ldots, n_m$ — их частоты, $n = n_1 + \ldots + n_m$ — общее число наблюдений; $\bar{x}$ — выборочное среднее; s^2 — выборочная дисперсия.

Как видим, выборочная дисперсия есть не что иное, как дисперсия статистической совокупности, рассмотренная в § 1.7. Сравнивая выражения для дисперсии случайной величины $DX = \sum_{i=1}^{n} (x_i - MX)^2 p_i$ и выборочной дисперсии и учитывая, что при достаточно большом числе наблюдений $\frac{n_i}{n} \approx p_i$ (вероятности того, что $X = x_i$), получим приближенное равенство

$$DX \approx s^2.$$

В тех случаях, когда неизвестен закон распределения случайной величины, полученное приближенное равенство позволяет оценивать дисперсию случайной величины по наблюдаемым ее значениям.

?

В каком случае выборочная дисперсия равна нулю?

Какими свойствами дисперсии обладает выборочная дисперсия?

Величину s, равную корню квадратному из выборочной дисперсии, называют ***выборочным средним квадратичным****, или* ***стандартным отклонением****.*

Следующий пример иллюстрирует применение выборочной дисперсии для сравнения двух величин.

Пример 5. Два стрелка выполнили по 100 выстрелов. Первый 8 очков выбил 40 раз, 9 очков — 10 раз и 10 очков — 50 раз. Второй — 8, 9, 10 очков выбил соответственно 10, 70, 20 раз. Кто из стрелков стреляет лучше?

❐ Имеем две выборки, составленные из наблюдений за случайной величиной X_i (i = 1, 2) — числом очков, которое выбивает i-й стрелок. Подсчитаем выборочные средние:

$$\bar{x}_1 = \frac{1}{100} \cdot (8 \cdot 40 + 9 \cdot 10 + 10 \cdot 50) = 9{,}1;$$

$$\bar{x}_2 = \frac{1}{100} \cdot (8 \cdot 10 + 9 \cdot 70 + 10 \cdot 20) = 9{,}1.$$

Итак, выборочные средние у обоих стрелков одинаковы. Можно говорить и о приближенном совпадении математических ожиданий этих величин. Значит, по этим данным стрелков можно было бы признать одинаково умелыми. Тем не менее определим меру разброса результатов стрельбы каждого из стрелков, для чего подсчитаем выборочные дисперсии. Вычисления удобно провести по плану, приведенному в п. 1.7.2, заполняя таблицы: для первого стрелка — таблица 4.9, для второго стрелка — таблица 4.10.

Таблица 4.9

x_k	n_k	$x_k n_k$	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 n_k$
8	40	320	−1,1	1,21	48,4
9	10	90	− 0,1	0,01	0,1
10	50	500	0,9	0,81	40,5
	$\sum = 100$	$\sum = 910$			$\sum = 89$

Таблица 4.10

x_k	n_k	$x_k n_k$	$x_k - \bar{x}$	$(x_k - \bar{x})^2$	$(x_k - \bar{x})^2 n_k$
8	10	80	− 1,1	1,21	12,1
9	70	720	−0,1	0,01	0,7
10	20	200	0,9	0,81	16,2
	$\sum = 100$	$\sum = 910$			$\sum = 29$

$$\bar{x}_1 = \frac{1}{100} \cdot 910 = 9{,}1; \quad s_1^2 = \frac{1}{100} \cdot 89 = 0{,}89;$$

$$\bar{x}_2 = \frac{1}{100} \cdot 910 = 9{,}1; \quad s_2^2 = \frac{1}{100} \cdot 29 = 0{,}29.$$

Поскольку $s_2^2 < s_1^2$, то можно считать, что $DX_2 < DX_1$, т. е. результаты стрельбы второго стрелка менее рассеяны по сравнению с первым. Итак, сделаем вывод, что второй стрелок стреляет лучше первого. ■

Контрольные вопросы

1. Какое из следующих неравенств верно:

$$DX \leqslant MX^2;$$
$$DX > MX^2;$$
$$DX \geqslant MX^2;$$
$$DX < MX^2?$$

2. Может ли выполняться равенство $DX = MX^2$?

3. Как изменится дисперсия случайной величины, если ко всем ее значениям прибавить одно и то же число?

4. Как изменится дисперсия случайной величины, если все ее значения умножить на одно и то же число?

5. Является ли выборочная дисперсия случайной величиной?

6. Изменится ли среднее квадратичное отклонение случайной величины, если ко всем ее значениям прибавить одно и то же число?

7. Как изменится среднее квадратичное отклонение случайной величины, если все ее значения умножить на одно и то же число?

Задачи

369.° Найдите дисперсию случайной величины, если ее закон распределения имеет вид

x	−2	−1	0	2
p	0,1	0,2	0,5	0,2

370. Найдите $D(2X + 3)$, если закон распределения случайной величины X имеет вид

x	−1	0	2	3
p	0,2	0,1	0,1	0,6

371. Случайная величина X имеет закон распределения

x	2025	2050	2075
p	0,1	0,4	0,5

Найдите DX:

а)° непосредственно по закону распределения X;

б) воспользовавшись свойствами дисперсии и предварительно преобразовав X.

372. Число бракованных изделий из 1000 производимых на двух однотипных станках, имеет соответственно законы распределения:

x	0	1	2	3	4
p	0,1	0,2	0,4	0,2	0,1

x	0	1	2	3	4
p	0,15	0,2	0,25	0,3	0,1

Сравните средние значения и дисперсии числа бракованных изделий, выпущенных на первом и втором станках. Какому станку вы отдадите предпочтение? Продукция какого станка является более «стабильной»?

373. Требуется оценить три разных проекта и разработать рекомендации. По каждому из проектов необходимы инвестиции в объеме 12 000 денежных единиц, а возврат средств планируется на следующий год. По первому проекту гарантированный возврат составит 14 000 денежных единиц. По второму проекту может быть получено либо 10 000 денежных единиц, либо 20 000 денежных единиц, вероятность в каждом случае составляет 0,5. Третий проект не даст ничего с вероятностью 0,98 или принесет 1 000 000 денежных единиц с вероятностью 0,02. Какой проект вы выберете?

374. Для условий задачи 345 (с. 322) найдите:

а) дисперсию и среднее квадратичное отклонение времени простоя;

б) вероятность того, что время простоя будет отличаться от среднего значения не более чем на одну величину среднего квадратичного отклонения;

в) вероятность того, что время простоя будет отличаться от среднего значения не более чем на две величины среднего квадратичного отклонения.

375. Для условий задачи 346 (с. 323) найдите уровень риска в данной ситуации.

376. На основе анализа выборки, составленной из результатов измерения 800 початков кукурузы, получены данные о длине початков, измеренной с точностью до 1 см (табл. 4.11).

Таблица 4.11

Длина, см	10	11	12	13	14	15	16	17
Число початков	1	1	2	6	23	35	50	95
Длина, см	18	19	20	21	22	23	24	25
Число початков	166	162	114	67	40	22	14	2

Вычислите среднее значение и стандартное отклонение длины початка.

§ 4.6. Независимые случайные величины

Так же, как и для суммы случайных величин, для их произведения приходится находить среднее значение по известным средним значениям сомножителей. Произведением случайных величин X и Y является случайная величина, все возможные значения которой равны произведениям вида x_iy_j, где x_i $(i = 1, \ldots, n)$, y_j $(j = 1, \ldots, m)$ — все значения соответственно случайных величин X и Y.

Рассмотрим пример.

Пример 1. Число яиц, присылаемых фермером на рынок ежедневно, имеет следующий закон распределения:

***x*, десятков**	5	6	7	8
p	0,1	0,4	0,3	0,2

Цена одного десятка яиц может равняться 30; 35; 40 р. с вероятностями 0,1; 0,6; 0,3 соответственно. Найти среднедневную выручку фермера от реализации яиц.

Мы можем найти среднее значение числа яиц, присылаемых фермером на рынок ежедневно (обозначим эту случайную величину через X), среднее значение цены одного десятка яиц (эту случайную величину обозначим через Y). А требуется найти математическое ожидание случайной величины XY — дневной выручки фермера от реализации яиц. Как это сделать? Хотелось бы, чтобы $M(XY) = MX \cdot MY$. Оказывается, что это равенство выполняется не всегда.

Уже простой пример заставляет отказаться от такого предположения. ■

Пример 2. Пусть случайная величина X имеет закон распределения

x	-1	1
p	0,5	0,5

Подсчитаем MX и MX^2.

❐ По определению математического ожидания $MX = -1 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0$. Случайная величина X^2 принимает единственное значение 1 с вероятностью, равной 1, т. е. она является константой. $MX^2 = 1$. С другой стороны, $MX^2 = M(XX) \neq MX \cdot MX = 0$. ■

Оказывается, что в общем случае нельзя однозначно выразить $M(XY)$ через MX и MY. Но есть очень важный частный случай, когда можно выразить $M(XY)$ через MX и MY, причем связь между этими средними значениями имеет довольно простой вид. Этот случай связан с ***независимыми случайными величинами***.

Независимость двух случайных величин означает отсутствие взаимосвязи между ними, когда знание значения одной случайной величины не помогает прогнозировать значение другой случайной величины. Для двух независимых случайных величин вероятности, с которыми одна принимает определенные значения, не зависят от того, какие значения в этот момент (т. е. при тех же исходах опыта) приняла другая величина.

Каждая случайная величина X определяет некоторый класс событий, которые описывают ее значения. В этот класс входят, например, такие события, как $(X = x)$, $(X < x)$, $(X > x)$ и т. д. Естественно независимость случайных величин отождествить с независимостью классов определяемых ими событий.

Случайные величины X и Y называют ***независимыми***, *если для любых x и y выполняется равенство:*

$$P((X = x) \cdot (Y = y)) = P(X = x) \cdot P(Y = y).$$

Другими словами, случайные величины X и Y независимы, если для любых x и y независимы события $(X = x)$ и $(Y = y)$.

Пример 3. Рассмотрим опыт, состоящий в двукратном подбрасывании симметричной монеты и случайные величины: X — число гербов, выпавших при первом подбрасывании монеты; Y — число гербов, выпавших при втором подбрасывании монеты. Выясним, являются ли эти случайные величины независимыми.

❐ Для ответа на поставленный вопрос нужно вычислить $P((X = x) \cdot (Y = y))$ для любых x и y и сравнить с произведением $P(X = x) \cdot P(Y = y)$. ПЭИ опыта имеет вид {ГГ, ГЦ, ЦГ, ЦЦ}. Каждая из случайных величин X и Y может принимать значения 0 или 1. Если x или y отлично от этих значений, то $P((X = x) \cdot (Y = y)) = P(X = x) \cdot P(Y = y) = 0$. Остается проверить четыре равенства: каждое значение X комбинировать с каждым значением Y (вспомните комбинаторное правило умножения). Имеем:

$$P((X = 1) \cdot (Y = 1)) = P(\text{ГГ}) = 0{,}25;$$
$$P(X = 1) \cdot P(Y = 1) = P\{\text{ГГ, ГЦ}\} \cdot P\{\text{ГГ, ЦГ}\} = 0{,}5 \cdot 0{,}5 = 0{,}25.$$

Аналогично:

$$P((X = 1) \cdot (Y = 0)) = P(\text{ГЦ}) = 0{,}25;$$
$$P(X = 1) \cdot P(Y = 0) = P\{\text{ГГ, ГЦ}\} \cdot P\{\text{ЦЦ, ЦГ}\} = 0{,}5 \cdot 0{,}5 = 0{,}25.$$

Точно так же проверяются остальные два равенства. ■

? Если случайные величины X и Y независимы, то чему равна $P(X = a \mid Y = b)$?

Вернемся к задаче выражения математического ожидания произведения двух случайных величин через математические ожидания сомножителей. Попробуем решить эту задачу для независимых случайных величин. Предварительно рассмотрим пример.

Пример 4. Пусть случайные величины X и Y независимы и имеют соответственно законы распределения:

x	−1	0	1
p	0,2	0,3	0,5

x	1	2
p	0,7	0,3

а) Вычислить MX и MY.

б) Составить закон распределения случайной величины $X \cdot Y$.

в) Вычислить $M(XY)$ и выразить это математическое ожидание через MX и MY.

❐ а) Первое задание выполняется сразу:

$$MX = -1 \cdot 0{,}2 + 0 \cdot 0{,}3 + 1 \cdot 0{,}5 = 0{,}3;$$
$$MY = 1 \cdot 0{,}7 + 2 \cdot 0{,}3 = 1{,}3.$$

б) Чтобы составить закон распределения случайной величины XY, нужно знать все ее значения и соответствующие вероятности. Эта случайная величина принимает значения: -2, -1, 0, 1, 2. Вычислим вероятности, с которыми она принимает эти значения. Значение -2 она принимает тогда и только тогда, когда $X = -1$ и $Y = 2$, т. е. $(XY = -2) = (X = -1) \cdot (Y = 2)$. Ввиду независимости событий, $P(XY = -2) = P(X = -1) \cdot P(Y = 2) = 0{,}2 \cdot 0{,}3 = 0{,}06$. Аналогично:

$$P(XY = -1) = P(X = -1) \cdot P(Y = 1) = 0{,}2 \cdot 0{,}7 = 0{,}14;$$
$$P(XY = 2) = P(X = 1) \cdot P(Y = 2) = 0{,}5 \cdot 0{,}3 = 0{,}15;$$
$$P(XY = 1) = P(X = 1) \cdot P(Y = 1) = 0{,}5 \cdot 0{,}7 = 0{,}35.$$

Осталось найти $P(XY = 0)$. Конечно, можно найти эту вероятность вычитанием из 1 суммы найденных вероятностей: $P(XY = 0) = 1 - (0{,}06 + 0{,}14 + 0{,}15 + 0{,}35) = 0{,}3$. Можно найти ее и другим способом. Значение 0 случайная величина XY принимает тогда и только тогда, когда $X = 0$, а вероятность этого события равна 0,3. Итак, закон распределения $Z = XY$ имеет следующий вид:

z	-2	-1	0	1	2
p	0,06	0,14	0,3	0,35	0,15

в) $M(XY) = -2 \cdot 0{,}06 + (-1) \cdot 0{,}14 + 0 \cdot 0{,}3 + 1 \cdot 0{,}35 + 2 \cdot 0{,}15 = 0{,}39 = 0{,}3 \cdot 1{,}3 = MX \cdot MY$. ■

Можно высказать предположение о том, что для независимых случайных величин математическое ожидание произведения равно произведению математических ожиданий сомножителей. Оказывается, этот результат имеет место и в общем случае.

ТЕОРЕМА 1. *Если случайные величины X и Y независимы, то*

$$M(XY) = MX \cdot MY.$$

Это свойство доказывается аналогично свойству математического ожидания суммы двух случайных величин.

❐ Пусть случайные величины X и Y определены на ПЭИ $U = \{u_1, \ldots, u_m\}$. Обозначим $Z = X \cdot Y$. В соответствии с равенством (2) § 4.2,

$$MZ = Z(u_1) \cdot p(u_1) + \ldots + Z(u_m) \cdot p(u_m).$$

Поскольку $Z(u_k) = X(u_k) \cdot Y(u_k)$, $k = 1, 2, \ldots, m$, то

$$MZ = X(u_1) \cdot Y(u_1) \cdot p(u_1) + \ldots + X(u_m) \cdot Y(u_m) \cdot p(u_m).$$

Пусть, например, событие $(X = x_1) \cdot (Y = y_1)$ происходит при одном из исходов опыта $u_1, \ldots, u_k$, т. е.

$$X(u_1) = \ldots = X(u_k) = x_1,\ Y(u_1) = \ldots = Y(u_k) = y_1.$$

Тогда $p_{11} = P((X = x_1) \cdot (Y = y_1)) = p(u_1) + \ldots + p(u_k)$. Используя независимость случайных величин X и Y, будем иметь

$$p_{11} = P((X = x_1) \cdot (Y = y_1)) = P(X = x_1) \cdot P(Y = y_1).$$

Отсюда

$$X(u_1) \cdot Y(u_1) \cdot p(u_1) + \ldots + X(u_k) \cdot Y(u_k) \cdot p(u_k) =$$
$$= x_1 y_1 (p(u_1) + \ldots + p(u_k)) = x_1 y_1 P((X = x_1) \cdot (Y = y_1)) =$$
$$= x_1 P(X = x_1) \cdot y_1 P(Y = y_1).$$

Аналогичные соотношения имеют место для всех событий $(X = x_i) \cdot (Y = y_j)$ и их вероятностей p_{ij}, причем различные события $(X = x_i) \cdot (Y = y_j)$ происходят при различных исходах эксперимента. Продолжая таким образом и далее, получим $M(XY) = MX \cdot MY$. ■

Как видно из примера 2, для зависимых случайных величин X и X теорема 1, вообще говоря, не имеет места. В том, что случайные величины X и X, рассмотренные в примере 2, зависимы, убедиться несложно. В самом деле, $P((X = 1) \cdot (X = 1)) = P(X = 1) = 0{,}5$, а $P(X = 1) \cdot P(X = 1) = 0{,}5 \cdot 0{,}5 = 0{,}25$.

Вернемся к примеру 1.

❐ Пусть X — число яиц (в десятках), которые присылает фермер каждую неделю на рынок, Y — цена одного десятка. Тогда XY — ежедневная выручка фермера. Нужно найти $M(XY)$. Случайные величины X и Y можно считать независимыми, так как естественно принять, что количество яиц, присылаемых на рынок одним фермером, не влияет на цену одного десятка яиц на рынке. Поэтому $M(XY) = MX \cdot MY$. Поскольку

$$MX = 5 \cdot 0{,}1 + 6 \cdot 0{,}4 + 7 \cdot 0{,}3 + 8 \cdot 0{,}2 = 6{,}6,$$
$$MY = 30 \cdot 0{,}1 + 35 \cdot 0{,}6 + 40 \cdot 0{,}3 = 36 \text{ (р.)},$$

то $M(XY) = 6{,}6 \cdot 36 \approx 237{,}6$ (р.) ■

Так же как понятие независимости двух случайных событий обобщается на любое конечное число событий, так и для нескольких случайных величин вводится понятие независимости в совокупности.

Случайные величины $X_1, \ldots, X_n$ называют ***независимыми в совокупности****, если события $(X_1 = x_1), \ldots, (X_n = x_n)$ независимы в совокупности для любых $x_1, \ldots, x_n$.*

Независимость случайных величин в совокупности означает, что задание каких-либо определенных значений для части этих величин не влияет на законы распределения остальных величин.

?

Рассмотрим случайные величины:

X_1 — число гербов, выпавших при первом подбрасывании симметричной монеты;

X_2 — число гербов, выпавших при втором подбрасывании симметричной монеты;

X_3 — число совпадений результатов, имевших место при двух подбрасываниях симметричной монеты.

Выясните, являются ли эти случайные величины:

а) попарно независимыми;

б) независимыми в совокупности.

Теорема 1 выполняется для произведения любого конечного числа *независимых в совокупности* случайных величин.

?

* Докажите, что математическое ожидание произведения трех независимых в совокупности случайных величин равно произведению математических ожиданий этих величин.

* Справедливо ли утверждение, сформулированное в предыдущем задании, для трех попарно независимых случайных величин?

При определении дисперсии давалось обещание в дальнейшем объяснить, почему было отдано предпочтение принятому определению. Настало время его выполнять. Попробуем догадаться о полезном свойстве дисперсии, рассмотрев вначале пример.

Пример 5. Законы распределения числа бракованных деталей, выпускаемых за смену двумя незнакомыми между собой рабочими, работающими в разных цехах, изготавливающими различные детали, не связанные между собой технологиями изготовления, используемым сырьем, средствами изготовления, имеют соответственно следующий вид:

x	0	1
p	0,9	0,1

x	0	1
p	0,95	0,05

а) Вычислить дисперсии числа бракованных деталей, изготовленных каждым из рабочих.

б) Составить закон распределения числа бракованных деталей, изготовленных двумя рабочими за смену.

в) Вычислить дисперсию числа бракованных деталей, изготовленных двумя рабочими за смену, сравнить ее с дисперсиями числа бракованных деталей, изготовленных каждым из рабочих.

❒ а) Обозначим число бракованных деталей, изготовленных каждым из рабочих, соответственно через X и Y. Имеем:

$$MX = 0{,}1,\quad MX^2 = 0{,}1,\quad DX = MX^2 - (MX)^2 = 0{,}09;$$
$$MY = 0{,}05,\quad MY^2 = 0{,}05,\quad DY = MY^2 - (MY)^2 = 0{,}0475.$$

б) Число бракованных деталей, изготовленных двумя рабочими за смену, есть случайная величина $X + Y$. Она принимает значения 0, 1, 2. Значение 0 она принимает тогда и только тогда, когда X и Y равны нулю. Благодаря условиям, перечисленным при формулировке задания (рабочие не знакомы между собой, работают в разных цехах, изготавливают различные детали, не связанные между собой технологиями изготовления, используемым сырьем, средствами изготовления), случайные величины X и Y можно считать независимыми. Поэтому

$$P(X + Y = 0) = P((X = 0) \cdot (Y = 0)) =$$
$$= P(X = 0) \cdot P(Y = 0)) = 0{,}9 \cdot 0{,}95 = 0{,}855.$$

Аналогично:

$$P(X + Y = 2) = P((X = 1) \cdot (Y = 1)) =$$
$$= P(X = 1) \cdot P(Y = 1)) = 0{,}1 \cdot 0{,}05 = 0{,}005;$$

$$P(X + Y = 1) = P((X = 0) \cdot (Y = 1) + (X = 1) \cdot (Y = 0)) =$$
$$= P(X = 0) \cdot P(Y = 1) + P(X = 1) \cdot P(Y = 0) =$$
$$= 0{,}9 \cdot 0{,}05 + 0{,}1 \cdot 0{,}95 = 0{,}14.$$

Итак, случайная величина $Z = X + Y$ имеет закон распределения

z	0	1	2
p	0,855	0,14	0,005

в) По свойству 1 дисперсии,

$$DZ = D(X + Y) = M(X + Y)^2 - (M(X + Y))^2.$$

Так как $M(X + Y) = 0{,}15$, $M(X + Y)^2 = 0{,}16$, то

$$D(X + Y) = 0{,}1375 = 0{,}09 + 0{,}0475 = DX + DY. \blacksquare$$

Можно высказать предположение о том, что для независимых случайных величин дисперсия суммы равна сумме дисперсий слагаемых. Оказывается, этот результат имеет место и в общем случае.

ТЕОРЕМА 2. *Если случайные величины X и Y независимы, то*

$$D(X + Y) = DX + DY.$$

❒ Обозначим $MX = a$, $MY = b$. Тогда $M(X + Y) = a + b$. По определению дисперсии $D(X + Y) = M((X + Y) - (a + b))^2$. Выполняя несложные преобразования, будем иметь

$$\begin{gathered} D(X + Y) = M((X - a) + (Y - b))^2 = \\ = M(X - a)^2 + M(Y - b)^2 + 2M(X - a)(Y - b) = \\ = DX + DY + 2M(XY) - 2bMX - 2aMY + 2ab = \\ = DX + DY + 2M(XY) - 2ab - 2ab + 2ab. \end{gathered}$$

Поскольку X и Y — независимы, то в соответствии с теоремой 1 $M(XY) = MX \cdot MY = ab$. Поэтому $D(X + Y) = DX + DY$. ■

Это утверждение выполняется и для суммы *n попарно независимых случайных величин* $X_1, \ldots, X_n$:

$$D(X_1 + \ldots + X_n) = DX_1 + \ldots + DX_n.$$

? Проверьте, выполняется ли последнее утверждение для следующих трех случайных величин:
X_1 — число гербов, выпавших при первом подбрасывании симметричной монеты;
X_2 — число гербов, выпавших при втором подбрасывании симметричной монеты;
X_3 — число совпадений результатов, имевших место при двух подбрасываниях симметричной монеты.

Выразите среднее квадратичное отклонение суммы n независимых в совокупности случайных величин через средние квадратичные отклонения слагаемых.

В теореме 2 условие независимости случайных величин существенно: если оно не выполняется, то теорема, вообще говоря, не имеет места. Убедимся в этом на примере.

Пример 6. Пусть случайная величина X имеет закон распределения

x	0	1
p	0,9	0,1

Вычислить $D(2X)$ двумя способами:

а) воспользовавшись свойствами дисперсии;

б) представив $2X$ в виде $X + X$ и воспользовавшись теоремой 2.

❐ а) Из решения примера 5 известно, что $DX = 0{,}09$; согласно свойству дисперсии, $D(2X) = 4DX = 0{,}36$.

б) Если бы мы воспользовались теоремой 2, то получили бы $D(2X) = D(X + X) = DX + DX = 0{,}18$.

Почему получились различные результаты? Дело в том, что случайные величины X и X, как мы видели выше, зависимы и применение теоремы 2 было неправомерным. Итак, $D(2X) = 0{,}36$. ■

Пример 7. Пусть $X_1, \ldots, X_n$ — независимые в совокупности случайные величины, для которых $MX_k = a$, $DX_k = \sigma^2$. Найти математическое ожидание и дисперсию их среднего арифметического ($k = 1, 2, \ldots, n$).

❐ Пусть $Y = \frac{1}{n}(X_1 + \ldots + X_n)$. Тогда

$$MY = M\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{n}na = a;$$

$$DY = D\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) =$$

$$= \frac{1}{n^2}(DX_1 + \ldots + DX_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \blacksquare$$

Таким образом, среднее значение среднего арифметического n независимых в совокупности случайных величин с одинаковыми средними значениями и дисперсиями равняется среднему значению любой из этих величин; дисперсия же среднего арифметического в n раз меньше дисперсии каждой из рассматриваемых случайных величин.

Этот вывод имеет широкое применение.

Пусть требуется с наибольшей точностью найти значение некоторой физической величины a. Пусть α — результат измерения, X — погрешность измерения (α и X — случайные величины). Тогда $a = \alpha + X$. Если нет систематической погрешности измерения, то $MX = 0$. Отсюда $a = M\alpha$. Проведем, например, n независимых измерений $\alpha_1, \ldots, \alpha_n$. Пусть $\bar{\alpha} = \frac{1}{n}(\alpha_1 + \ldots + \alpha_n)$. Тогда $M\bar{\alpha} = \frac{1}{n}nM\alpha_1 = a$, т. е. значения величины $\bar{\alpha}$, ра́вно как и значения α_i, группируются вокруг истинного значения данной величины. Это значение $\bar{\alpha}$ можно принять в качестве приближенного значения данной величины. Поскольку $D\bar{\alpha} = \frac{1}{n^2} nD\alpha_i = \frac{D\alpha_i}{n}$, то рассеяние значений $\bar{\alpha}$ заметно меньше рассеяния значений α_i, поэтому, взяв за a значение $\bar{\alpha}$, имеем основание надеяться, что большая ошибка будет менее вероятной, чем тогда, когда в качестве a берут результат α_i одного измерения.

В качестве показателя погрешности измерения обычно берут среднее квадратичное отклонение. Поскольку $D\bar{\alpha} = \frac{1}{n^2} D\alpha$, то $\sigma(\bar{\alpha}) = \frac{1}{\sqrt{n}}\sigma(\alpha_i)$, и погрешность среднего арифметического n результатов измерений в $\sqrt{n}$ раз меньше погрешности каждого измерения.

➤

Пример 8. Для условий примера 5 проверим, будет ли среднее абсолютное отклонение суммы двух независимых случайных величин равняться сумме средних абсолютных отклонений слагаемых, т. е. будет ли иметь место равенство

$$M|(X+Y) - M(X+Y)| = M|X - MX| + M|Y - MY|.$$

❐ Напомним, что случайные величины X, Y, $Z = X + Y$ имеют соответственно законы распределения:

x	0	1
p	0,9	0,1

y	0	1
p	0,95	0,05

z	0	1	2
p	0,855	0,14	0,005

Их математические ожидания соответственно равны: $MX = 0{,}1$; $MY = 0{,}05$; $MZ = 0{,}15$. Случайные величины $|X - MX|$, $|Y - MY|$, $|Z - MZ|$ имеют соответственно законы распределения:

x	0,1	0,9
p	0,9	0,1

y	0,05	0,95
p	0,95	0,05

z	0,15	0,85	1,85
p	0,855	0,14	0,005

Поэтому $M|X - MX| = 0{,}18$; $M|Y - MY| = 0{,}095$; $M|Z - MZ| = 0{,}2565 \neq 0{,}18 + 0{,}095 = M|X - MX| + M|Y - MY|$. Итак, «второй кандидат» на меру отклонения значений случайной величины от среднего желаемым свойством не обладает. ■ ➤

Контрольные вопросы

1. Могут ли быть независимыми случайные величины X и $-X$?

2. Найдите ошибку в следующих вычислениях. Пусть $DX = 2$, тогда $D(2X) = 4DX = 8$, с другой стороны, $D(2X) = D(X + X) = DX + DX = 4$.

3. Верно ли, что дисперсия разности двух независимых случайных величин равняется разности дисперсий этих величин?

4. Известно, что X и Y — независимые случайные величины; $DX = 1$, $DY = 2$. Найдите ошибку в следующих вычислениях: $D(2X - 3Y) = 4DX - 9DY = 4 - 18 = -14$.

5. Как выразить среднее квадратичное отклонение суммы двух независимых случайных величин через средние квадратичные отклонения слагаемых?

6. Чему равна дисперсия суммы двух произвольных случайных величин?

Задачи

377. Пусть случайная величина X равна числу очков, выпавшему при первом бросании игрального кубика, а Y — числу очков, выпавшему при втором ее бросании. Докажите, что случайные величины X и Y независимы, если все исходы двух бросаний равновозможны.

378.° Случайные величины X и Y независимы, причем $DX = 1$, $DY = 2$. Найдите DZ, если:

а) $Z = 3X + Y$; б) $Z = 2X - Y - 2$; в) $Z = aX + bY + c$, где a, b, c — постоянные величины.

379.° По проводнику, сопротивление которого зависит от случайных обстоятельств, проходит электрический ток, сила которого также зависит от случая. Известно, что среднее значение сопротивления проводника R равно 25 Ом, а средняя сила тока I равна 6 А. Найдите среднее значение электродвижущей силы, если известно, что сопротивление и сила тока независимы.

380. Случайные величины X и Y независимы и имеют, соответственно, законы распределения:

x	−1	0	1
p	0,4	0,1	0,5

y	−2	−1	0	1
p	0,3	0,1	0,4	0,2

Найдите:

а)° $D(3X + 2Y)$; б)° $D(3X - 2Y)$; в)° $M(XY)$; г) закон распределения XY.

Докажите, что случайные величины $aX + b$ и $cY + d$ (a, b, c, d — постоянные), а также X^2 и Y^2 независимы.

381. Два стрелка независимо друг от друга сделали по одному выстрелу в одну и ту же мишень. Вероятность попадания для первого стрелка равна p_1, а для второго — p_2. Найдите MX, DX, если X — общее число попаданий в мишень.

382. Пусть X — случайная величина, равная числу очков, выпавших при бросании первого игрального кубика, а Y — случайная величина, равная числу очков, выпавших при бросании второго игрального кубика. Докажите, что $M(XY) = MX \cdot MY$. Игральные кубики считать правильными.

383. Сколько независимых измерений (проведенных с одинаковой точностью) достаточно сделать, чтобы результат среднего арифметического этих измерений был в 10 раз точнее каждого измерения?

384*. Предприятие выпускает некоторые изделия, причем для каждого отдельного изделия существует определенная вероятность $p = 0{,}002$ оказаться бракованным. Считая, что все изделия из некоторой тысячи изделий независимо друг от друга могут оказаться бракованными с вероятностью p, найдите среднее число бракованных изделий на 1000 выпущенных изделий и дисперсию этой случайной величины.

§ 4.7. Числовые характеристики биномиального распределения

Ранее уже отмечалось, что биномиальное распределение служит математической моделью многих реальных явлений. В связи с этим представляют интерес математическое ожидание и дисперсия случайной величины, имеющей биномиальное распределение.

В примере 1 § 4.4 получен следующий закон распределения числа попаданий при трех выстрелах ($n = 3$), если вероятность попадания при каждом выстреле равна $p = 0{,}9$:

x	0	1	2	3
p	0,001	0,027	0,243	0,729

Вычислим математическое ожидание и дисперсию этой величины:

$$MX = 0 \cdot 0{,}001 + 1 \cdot 0{,}027 + 2 \cdot 0{,}243 + 3 \cdot 0{,}729 = 2{,}7;$$
$$MX^2 = 0 \cdot 0{,}001 + 1 \cdot 0{,}027 + 4 \cdot 0{,}243 + 9 \cdot 0{,}729 = 7{,}56;$$
$$DX = 7{,}56 - (2{,}7)^2 = 0{,}27.$$

Обратите внимание на то, что

$$MX = 2{,}7 = 3 \cdot 0{,}9 = np; \quad DX = 0{,}27 = 3 \cdot 0{,}9 \cdot 0{,}1 = np(1-p),$$

где n — число испытаний Бернулли, p — вероятность попадания при каждом выстреле. Оказывается, этот результат имеет место и в общем случае.

ТЕОРЕМА. *Если случайная величина X имеет биномиальное распределение с параметрами n и p, то*

$$MX = np; \quad DX = np(1-p).$$

?

Пусть случайная величина X имеет биномиальное распределение с параметрами $n = 10$, $p = 0{,}2$. Найдите MX, DX и $\sigma(X)$.

Пусть случайная величина X имеет биномиальное распределение с параметрами n, $p = 0{,}3$. На сколько сдвигается вправо среднее значение случайной величины X при возрастании n на 1?

❐ Докажем сформулированную теорему. Применим свойства математического ожидания и дисперсии к доказательству этих равенств. Обозначим через X_k — число успехов в k-м испытании Бернулли, $k = 1, \ldots, n$. Тогда

$$X = X_1 + \ldots + X_n.$$

Из свойства 3 математического ожидания (см. § 4.3) следует, что

$$MX = MX_1 + \ldots + MX_n.$$

Составим закон распределения X_k. Каждая из этих случайных величин принимает значения 0 или 1 («успех» наступил или нет) с вероятностями $q = 1 - p$ и p (вероятности исходов в испытаниях Бернулли от опыта к опыту не меняются). Отсюда

$$MX_k = 0 \cdot (1-p) + 1 \cdot p = p; \quad MX = \underbrace{p + p + \ldots + p}_{n \text{ раз}} = np.$$

Аналогично вычислим DX. Согласно свойству дисперсии суммы попарно независимых случайных величин $X_1, \ldots, X_n$,

$$DX = DX_1 + \ldots + DX_n.$$

Так как случайная величина X_k^2 имеет тот же закон распределения, что и X_k ($0^2 = 0$, $1^2 = 1$), то

$$MX_k^2 = p, \quad DX_k = MX_k^2 - (MX_k)^2 = p - p^2 = p(1-p);$$

$$DX = \underbrace{p(1-p) + \ldots + p(1-p)}_{n \text{ раз}} = np(1-p). \blacksquare$$

Верно ли, что случайная величина X_k (см. доказательство) имеет биномиальное распределение? Каковы его параметры?

Из доказанной теоремы следует, что среднее квадратичное отклонение случайной величины, имеющей биномиальное распределение с параметрами n и p, равно $\sqrt{np(1-p)}$. Поскольку среднее квадратичное отклонение является мерой разброса значений случайной величины X и его значение при постоянном p пропорционально $\sqrt{n}$, то при возрастании n биномиальное распределение имеет разброс, пропорциональный $\sqrt{n}$.

Пример. Установлено, что вероятность ошибки со стороны контролера, проверяющего на заводе соответствие фактических размеров деталей техническим стандартам, равна 0,02. Сколько ошибок в среднем допустит за день контролер, к которому поступает на проверку 200 деталей?

❐ В задаче речь идет о 200 независимых испытаниях (проверка качества деталей), являющихся испытаниями Бернулли. Пусть X — число ошибок контролера при проверке 200 деталей. Эта случайная величина имеет биномиальное распределение с параметрами $n = 200$ и $p = 0{,}02$. Поэтому $MX = 200 \cdot 0{,}02 = 4$. Таким образом, при проверке 200 деталей контролер допустит примерно четыре ошибки. ■

Контрольные вопросы

1. Можно ли найти дисперсию биномиально распределенной случайной величины, зная только ее математическое ожидание?

2. Что нужно знать о биномиально распределенной случайной величине, кроме математического ожидания, чтобы вычислить ее дисперсию?

3. Пусть X — биномиально распределенная случайная величина. Что больше MX или DX?

4. Пусть X_k, $k = 1, 2, \ldots, n$, — число наступлений события A в k-м испытании Бернулли. Почему случайные величины $X_1, \ldots, X_n$:

а) независимы в совокупности;

б) имеют одинаковые законы распределения?

5. Пусть X — число наступлений события A в n независимых испытаниях, в которых это событие наступает соответственно с вероятностями $p_1, p_2, \ldots, p_n$. Чему равно:

а) MX; б) DX?

Задачи

385. Вероятность попадания в мишень при одном выстреле для данного стрелка равна $p = 0{,}8$; X — число попаданий в мишень в 100 независимых выстрелах. Найдите MX, DX, $\sigma(X)$.

386. Приобретено 40 лотерейных билетов, на каждый из которых выпадает выигрыш с вероятностью 0,05. Сколько в среднем следует ожидать выигрышей?

387. Случайная величина X имеет биномиальное распределение с параметрами n и p; $MX = 10$, $\sigma(X) = 3$. Найдите n и p.

388. Сколько игральных кубиков необходимо бросить для того, чтобы математическое ожидание числа кубиков, на которых выпало одно очко, равнялось пяти?

389. Пусть m — число наступлений события A в n независимых испытаниях, p — вероятность наступления события A в каждом испытании. Найдите:

а) $M\dfrac{m}{n}$; б) $D\dfrac{m}{n}$.

390. Известно, что 86% потребителей одобряют изменения, которые фирма намерена внести в некоторый потребительский товар. Принято решение опросить 10 потребителей для выяснения вопроса о том, следует ли вносить эти изменения.

а) Какое распределение будет иметь число потребителей, одобряющих изменения?

б) Чему равно среднее число людей из 10 опрошенных, которые одобрят изменения?

в) Чему равно среднее квадратичное отклонение числа людей из числа 10 опрошенных, которые одобрят изменения?

391. Запланировано проведение голосования по вопросу о профсоюзной забастовке. Пусть число голосов, поданных за проведение забастовки, подчиняется биномиальному распределению. Ожидается, что в голосовании примет участие 300 человек, а вероятность того, что каждый из них будет голосовать «за», составляет 0,53.

а) Укажите параметры биномиального распределения.

б) Найдите среднее и среднее квадратичное отклонение для числа людей, которые выскажутся за проведение забастовки.

в) Найдите приближенно вероятность того, что забастовка будет проведена, т. е. вероятность того, что большинство участников голосования выскажутся за ее проведение.

392.* Случайная величина X имеет биномиальное распределение с параметрами $n = 10$ и $p = 0{,}2$. Найдите вероятность того, что значения X отличаются от среднего значения на величину, не превосходящую:

а) величины одного среднего квадратичного отклонения;

б) величины двух средних квадратичных отклонений;

в) величины трех средних квадратичных отклонений.

§ 4.8. Неравенство Чебышёва

Знание среднего квадратичного отклонения случайной величины позволяет создать ориентировочное представление о том, насколько велики ожидаемые отклонения фактических значений этой величины от ее среднего значения. Из смысла дисперсии интуитивно ясно, что при малых значениях дисперсии большие отклонения значений случайной величины от среднего значения довольно редки. Другими словами, если дисперсия случайной величины мала, то вероятность того, что случайная величина сильно отклоняется от своего среднего значения, мала. Однако эти замечания сами по себе не содержат еще никаких количественных оценок и не дают возможности хотя бы приближенно рассчитать, сколь вероятными могут оказаться большие отклонения. Следующее утверждение уточняет содержание сказанного.

ТЕОРЕМА 1. *При произвольном положительном* ε *для любой случайной величины* X *выполняется неравенство*

$$P(|X - MX| > \varepsilon) \leqslant \frac{DX}{\varepsilon^2}. \qquad (1)$$

Неравенство (1) читается так: вероятность того, что отклонение случайной величины от ее среднего значения по модулю больше ε, не превосходит дисперсии этой случайной величины, деленной на ε^2.

Это неравенство получил один из крупнейших математиков XIX века П. Л. Чебышёв (1821–1894). Поэтому оно носит название неравенства Чебышёва. Идея вывода этого неравенства содержалась в работе Ж. Бьенэме (1796–1878), посвященной методу наименьших квадратов. За этим неравенством надолго закрепилось имя неравенства Бьенэме–Чебышёва. Известный российский математик А. А. Марков (1856–1922) объяснял это следующим образом: «Мы соединяем с этим замечательным, простым неравенством два имени Бьенэме и Чебышёва по той причине, что оно впервые ясно высказано и доказано Чебышёвым, но основная идея доказательства была значительно раньше указана Бьенэме, в мемуаре которого можно найти и самое неравенство, обставленное только некоторыми частными предложениями». Позднее это неравенство нашло многочисленные применения как в теории вероятностей, так и в других разделах математики.

❐ Для доказательства неравенства Чебышёва рассмотрим случайную величину Y, которую определим следующим образом:

$$Y = \begin{cases} 0, \text{ если } |X - MX| \leqslant \varepsilon, \\ \varepsilon^2, \text{ если } |X - MX| > \varepsilon. \end{cases}$$

Другими словами, для тех элементов u ПЭИ U, для которых $|X(u) - MX| \leqslant \varepsilon$, случайная величина $Y(u)$ равна 0, для остальных элементов ПЭИ $Y(u) = \varepsilon^2$. Тогда для всех $u \in U$ значения случайной величины Y не превосходят значений $(X - MX)^2$, т. е.

$$Y \leqslant (X - MX)^2.$$

Из свойства 6 математического ожидания (см. § 4.3) следует, что $MY \leqslant M(X - MX)^2$. Так как $M(X - MX)^2 = DX$, то $MY \leqslant DX$. Воспользовавшись определением математического ожидания, вычислим MY:

$$MY = 0 \cdot P(Y = 0) + \varepsilon^2 P(Y = \varepsilon^2) = \varepsilon^2 P(|X - MX| > \varepsilon).$$

Поэтому

$$\varepsilon^2 P(|X - MX| > \varepsilon) \leqslant DX.$$

Отсюда следует доказываемое неравенство. ■

Неравенство Чебышёва позволяет оценить вероятность отклонений от среднего, чисел бо́льших, чем любое заданное число, если только известно среднее квадратичное отклонение. Правда, оценка, даваемая неравенством Чебышёва, часто оказывается весьма грубой; все же иногда она может быть использована практически. Что касается его теоретического значения, то оно чрезвычайно велико.

? Можно ли получить с помощью неравенства Чебышёва содержательную оценку для вероятности $P(|X - MX| > \varepsilon)$, если $\varepsilon = \sigma(X)$; $\varepsilon < \sigma(X)$?

Неравенство Чебышёва позволяет не только оценить сверху вероятность того, что отклонение случайной величины от ее среднего значения по модулю больше ε, но и оценивать снизу вероятность того, что отклонение случайной величины от ее среднего значения по модулю не превосходит ε.

Действительно,

$$P(|X - MX| \leqslant \varepsilon) = 1 - P(|X - MX| > \varepsilon) \geqslant 1 - \frac{DX}{\varepsilon^2}.$$

Мы воспользовались свойством вероятностей противоположных событий, неравенством (1) и свойствами неравенств. Итак, из неравенства (1) вытекает:

$$P(|X - MX| \leqslant \varepsilon) \geqslant 1 - \frac{DX}{\varepsilon^2}. \qquad (2)$$

Неравенство Чебышёва в форме (2) позволяет оценить вероятность попадания значений случайной величины X в заданный интервал с центром в математическом ожидании этой случайной величины, а также указать ширину интервала с тем же центром, в который с заданной вероятностью попадают значения X.

Пример 1. Пусть $MX = 5$, $DX = 0{,}9$.

а) Оценить $P(|X - 5| > 1)$ и $P(1 \leqslant X \leqslant 9)$.

б) При каких значениях k выполняется неравенство $P(|X - 5| \leqslant k) \geqslant 0{,}9$?

❐ а) В соответствии с неравенствами (1) и (2):

$$P(|X - 5| > 1) = P(|X - MX| > 1) \leqslant \frac{DX}{1^2} = \frac{0{,}9}{1} = 0{,}9;$$

$$P(1 \leqslant X \leqslant 9) = P(-4 \leqslant X - 5 \leqslant 4) = P(|X - 5| \leqslant 4) \geqslant 1 - \frac{0{,}9}{16} \approx 0{,}94.$$

б) $P(|X - 5| \leqslant k) \geqslant 1 - \dfrac{0{,}9}{k^2}$.

Для выполнения требуемого неравенства достаточно потребовать, чтобы выполнялось неравенство $1 - \dfrac{0{,}9}{k^2} \geqslant 0{,}9$. Решая его, получим, что $k \geqslant 3$. ■

? Пусть $MX = 3$, $DX = 1$. Оцените с помощью неравенства Чебышёва $P(|X - 3| > 2)$ и $P(|X - 3| \leqslant 2)$.

Пример 2. Среднее значение измеряемой величины равно 200 м, а среднее квадратичное отклонение равно 5 м. Эту величину измерили 100 раз.

а) Найти среднее значение и среднее квадратичное отклонение среднего арифметического результатов измерений.

б) Оценить с помощью неравенства Чебышёва вероятность получения отклонения результата одного измерения от среднего больше 3 м.

в) Оценить с помощью неравенства Чебышёва вероятность получения отклонения среднего арифметического результатов 100 измерений от среднего, большего 3 м.

❐ а) Обозначим через X_i, $i = 1, 2, \ldots, 100$, результат i-го измерения, $Y = \dfrac{X_1 + X_2 + \ldots + X_{100}}{100}$. Тогда по свойствам математического ожидания

$$MY = M\frac{X_1 + X_2 + \ldots + X_{100}}{100} = \frac{MX_1 + MX_2 + \ldots + MX_{100}}{100} = 200 \text{ (м)}.$$

Применяя теорему о дисперсии суммы попарно независимых случайных величин (результаты измерения естественно считать независимыми), получим:

$$DY = D\frac{X_1 + X_2 + \ldots + X_{100}}{100} = \frac{DX_1 + DX_2 + \ldots + DX_{100}}{100^2} =$$

$$= \frac{25 \cdot 100}{100^2} = 0{,}25;$$

$$\sigma(Y) = \sqrt{DY} = 0{,}5 \text{ (м)}.$$

б) Применяя неравенство Чебышёва для случайной величины X_1, получим $P(|X_1 - 200| > 3) \leqslant \dfrac{DX_1}{9} = \dfrac{25}{9}$, т. е. содержательного результата мы не получили (и заранее было известно, что вероятность любого события не превосходит 1). На самом деле эта вероятность значительна, точное ее значение может быть найдено только тогда, когда полностью известен закон распределения результатов измерений.

в) В силу неравенства Чебышёва,

$$P(|Y - 200| > 3) \leqslant \frac{DY}{9} = \frac{0{,}25}{9} \approx 0{,}03.$$

Таким образом, для среднего арифметического из 100 измерений вероятность получить отклонение от среднего более 3 м уже очень мала (на самом деле она еще значительно меньше полученной оценки, так что практически можно совсем не считаться с возможностью такого отклонения). ■

Пример 3. Пусть $MX = a$, $DX = \sigma^2$. Оценить $P(|X - a| \leqslant 3\sigma)$.

❐ В соответствии с неравенством (2)

$$P(|X - a| \leqslant 3\sigma) \geqslant 1 - \frac{DX}{9\sigma^2} = \frac{8}{9}. \ ■$$

Полученное неравенство часто называют ***правилом трех сигм.*** Оно имеет следующий смысл: *с вероятностью, близкой к* 1 $\left(\geqslant \frac{8}{9}\right)$, *значения случайной величины должны находиться в интервале* **$(a - 3\sigma;\ a + 3\sigma)$.**

Можно ли событие «значение случайной величины содержится в интервале $(a - 3\sigma;\ a + 3\sigma)$» считать практически достоверным? Если да, то по какому критерию?

Пусть теперь X — случайная величина, имеющая биномиальное распределение с параметрами n и p. Так как $MX = np$, $DX = np(1 - p)$, $\sigma = \sqrt{np(1-p)}$, то с вероятностью, не меньшей $\frac{8}{9}$, число «успехов» в n испытаниях Бернулли должно находиться в интервале $\left(np - 3\sqrt{np(1-p)};\ np + 3\sqrt{np(1-p)}\right)$.

Если X — число выпадений двух гербов при 50 подбрасываниях трех симметричных монет, то вероятность выпадения двух гербов при подбрасывании трех монет можно вычислить по формуле Бернулли

$$P_3(2) = C_3^2 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^1 = \frac{3}{8}.$$

Тогда

$$MX = np = 50 \cdot \frac{3}{8} = 18{,}75;$$

$$DX = np(1 - p) = 50 \cdot \frac{3}{8} \cdot \frac{5}{8} \approx 11{,}72;\ \sigma(X) \approx 3{,}42.$$

Итак, с вероятностью, близкой к 1 $\left(\geqslant \frac{8}{9}\right)$, число выпадений двух гербов при 50 подбрасываниях трех монет должно находиться в интервале (9; 29). Опыт с подбрасыванием трех симметричных монет должен подтвердить этот прогноз.

Проведите 50 опытов с подбрасыванием трех монет. Подтвердился ли этот прогноз?

Правило трех сигм может быть использовано для проверки того, согласуется ли результат проведения испытаний Бернулли с допущением о том, что вероятность «успеха» в каждом испытании равна заранее заданному числу p_0.

Если хотя бы одно из наблюдаемых значений биномиальной случайной величины с параметрами n и p_0 не попадает в интервал

$$\left(np_0 - 3\sqrt{np_0(1-p_0)}\,;\, np_0 + 3\sqrt{np_0(1-p_0)}\,\right),$$

т. е. наблюдается практически невозможное событие (по какому критерию?), то гипотезу о значении вероятности «успеха» следует отвергнуть. В противном случае нет оснований для отклонения рассматриваемой гипотезы.

? Почему событие «значение биномиально распределенной случайной величины с параметрами n и p_0 не содержится в интервале $\left(np_0 - 3\sqrt{np_0(1-p_0)}\,;\, np_0 + 3\sqrt{np_0(1-p_0)}\,\right)$» можно считать практически невозможным? По какому критерию?

Пример 4. Перед началом футбольного матча одной из команд было высказано сомнение в симметричности монеты, с помощью которой судья определял, какая из команд начинает игру с центра поля. Для проверки этих сомнений монета была подброшена 100 раз, при этом герб выпал 42 раза. Опровергают ли эти данные предположение о правильности монеты?

❐ Для получения ответа воспользуемся правилом трех сигм. Если предположение о правильности монеты верно, то случайная величина «число выпадений герба при 100 подбрасываниях монеты» имеет биномиальное распределение с параметрами 100 и 0,5:

$$MX = np = 100 \cdot 0{,}5 = 50;$$
$$DX = np(1-p) = 100 \cdot 0{,}5 \cdot 0{,}5 \approx 25;$$
$$\sigma(X) = 5.$$

Интервал, построенный по правилу трех сигм, имеет вид $(50 - 3 \cdot 5;\ 50 + 3 \cdot 5)$, или $(35;\ 65)$. Значение 42 рассматриваемой случайной величины содержится в этом интервале. Данных, противоречащих предположению о правильности монеты, не получено. Следовательно, нет оснований отвергать это предположение. ■

Контрольные вопросы

1. Известно, что $MX = a$, $DX = \sigma^2$. Можно ли с помощью неравенства Чебышёва получить содержательную оценку для $P\left(|X - a| \leqslant \frac{1}{2}\sigma\right)$?

2. Пусть $MX = 0$, $DX = 1$. Укажите число:
а) меньшее $P(-3 < X < 3)$;
б) большее $P(|X| > 2)$.

3. Пусть $x_1, \ldots, x_r$ — все значения, которые принимает случайная величина X на отрезке $[-2; 2]$, $MX = 0$, $DX = 1$. Что можно сказать о величине $p_1 + p_2 + \ldots + p_r$, где $p_k = P(X = x_k)$, $k = 1, 2, \ldots, r$?

4. Оцените $P(|X - MX|) \leqslant 2\sigma(X))$ (правило «двух сигм»).

5. Имеет ли содержательный смысл правило «одной сигмы»?

Задачи

393. Пусть случайная величина X имеет закон распределения:

x	0	1	2	3
p	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Найдите вероятность того, что значения случайной величины X удалены от среднего значения на расстояние, не превышающее:

а) одного среднего квадратичного отклонения;

б) двух средних квадратичных отклонений;

в) трех средних квадратичных отклонений.

394. Пусть $MX = a$, $DX = \sigma^2$. При каком значении h вероятность попадания значения случайной величины X на отрезок $[a - h\sigma; a + h\sigma]$ не меньше 0,9; 0,99?

395. Известно, что $MX = 5$, $DX = 1$. Оцените

а) величину $P(2 \leqslant X \leqslant 8)$ снизу;

б) величину $P(|X - 5| > 2)$ сверху;

в) при каком значении k выполняется неравенство $P(|X - 5| \leqslant k) \geqslant 0{,}99$?

396.* Сколько раз достаточно бросить правильный игральный кубик, чтобы с вероятностью 0,99 можно было утверждать, что относительная частота выпадения «шестерки» отличается от $\frac{1}{6}$ не более чем на 0,05?

397. Частица пролетает сквозь поглощающий экран с вероятностью 0,01.

а) Чему равны математическое ожидание и дисперсия числа пролетевших сквозь экран частиц, если их было выпущено 100?

б) Оцените вероятность того, что сквозь экран пролетит более 11 частиц?

398. Среднее значение числа бракованных изделий при проверке 60 000 изделий оказалось равным 2400, а среднее квадратичное отклонение — 48. Оцените вероятность того, что фактическое число бракованных изделий будет заключено между 2300 и 2500.

399. Оцените с помощью неравенства Чебышёва суммарную вероятность всех значений, удаленных от среднего значения не более чем на h средних квадратичных отклонений. При каких значениях h полученная оценка не будет содержательной?

400. Предварительная оценка нового лекарства показала, что его эффективность равна 0,9. Для подтверждения или опровержения этого предположения оно было испытано на 100 больных. В 83 случаях оно оказалось эффективным. Какое решение об этом лекарстве можно принять на основании правила трех сигм?

401. Симметричную монету подбрасывают 1600 раз. Оцените вероятность получения при этом выпадения герба:

а) более 1200 раз;

б) более 900 раз.

§ 4.9. Закон больших чисел

В предыдущих параграфах на основании эмпирических фактов утверждалось, что относительная частота события приближенно равна его вероятности при достаточно большом числе испытаний. Аналогично среднее арифметическое независимых наблюдений случайной величины при большом числе наблюдений приближенно равно математическому ожиданию этой величины. Оказывается, что эти факты можно доказать строго математически. Они и составляют содержание *закона больших чисел.*

Пусть X_1, ..., X_n — независимые измерения некоторой величины с одним и тем же математическим ожиданием a и одним и тем же средним квадратичным отклонением σ. Если $Y = \frac{1}{n}(X_1 + ... + X_n)$, то, как установлено ранее, $MY = a$, $\sigma(Y) = \frac{\sigma}{\sqrt{n}}$. При любом $\varepsilon > 0$ в соответствии с неравенством Чебышёва имеем

$$P(|Y - a| \leqslant \varepsilon) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Если число n велико, то правая часть последнего неравенства близка к 1, т. е. при достаточно большом n с вероятностью, близкой к 1, среднее арифметическое n случайных величин как угодно ма́ло (меньше, чем на произвольное $\varepsilon > 0$) отличается от их математического ожидания.

Итак, доказана теорема.

ТЕОРЕМА 1. *Если случайные величины X_1, ..., X_n независимы и имеют одно и то же математическое ожидание a и одну и ту же дисперсию, то их среднее арифметическое при достаточно большом n с вероятностью как угодно близкой к* 1, *будет как угодно мало отличаться от a.*

?

Где в ходе доказательства теоремы 1 использовалась независимость случайных величин?

Почему $MY = a$, $\sigma(Y) = \frac{\sigma}{\sqrt{n}}$?

Это простейший частный случай одной из главных теорем теории вероятностей, так называемого ***закона больших чисел***. Его доказал П. Л. Чебышёв (1821—1894).

На действии закона больших чисел основано исключение случайных ошибок измерения. Пусть измеряется некоторая величина x. Обычно с помощью измерительных приборов нельзя определить ее абсолютно точно. Существует некоторая погрешность измерения, обозначим ее через X. Если $MX \neq 0$, то говорят, что прибор имеет *систематическую погрешность*. Изменив шкалу, можно добиться того, чтобы систематическая ошибка отсутствовала, т. е. чтобы $MX = 0$. Пусть x_1, ..., x_n — результаты измерения величины x, а $X_1 = x_1 - x$, ..., $X_n = x_n - x$ — погрешности соответствующих измерений. Их можно считать независимыми случайными величинами. Поскольку $MX_k = 0$, то из теоремы Чебышёва следует, что

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \leqslant \varepsilon\right) \to 1, \text{ при } n \to \infty.$$

Итак, вычисляя среднее арифметическое значений измерений

$$\frac{1}{n}(x_1 + \ldots + x_n) = x + \frac{1}{n}(X_1 + \ldots + X_n),$$

можно с большой точностью получить значение измеряемой величины, хотя все измерения проводились с погрешностями.

Происходит это, разумеется, потому, что при замене отдельных результатов измерения их средним арифметическим случайные отклонения в ту или другую сторону взаимно уничтожаются, вследствие чего суммарное отклонение оказывается малым.

➤

Важное и часто встречающееся на практике использование результатов теоремы Чебышёва состоит в том, что по сравнительно небольшой пробе (выборке) судят о качестве однородного материала. Так, например, о качестве хлопка, находящегося в кипе, судят по нескольким его маленьким пучочкам (штапелям), выхваченным случайно из разных мест кипы. О качестве большой партии зерна судят по нескольким небольшим пуркам (меркам), наполненным случайно захваченными в пурку зернами из разных мест оцениваемой партии. Суждения о качестве продукции, сделанные на основании такой выборки, обладают большой точностью, так как, скажем, число зерен, захваченных в пурку, хотя и мало по сравнению со всем запасом зерна, но само по себе велико и позволяет, согласно закону больших чисел, достаточно точно судить о средней массе одного зерна и, следовательно, о качестве всей партии зерна. Точно так же и о двадцатипудовой кипе хлопка судят по маленькому штапелю, содержащему несколько сотен волокон, весящих всего-навсего какую-нибудь десятую долю грамма.

Особенность теоремы Чебышёва состоит в том, что она применима к любому распределению вероятностей с конечными средним значением и дисперсией.

Закон больших чисел является статистическим законом, т. е. таким законом, который справедлив не для одного какого-либо явления, а для большого количества явлений, среди которых хаотично распределен некоторый общий признак.

Проиллюстрируем на примерах применение закона больших чисел.

Пример 1. При определении средних характеристик некоторого биологического рода берут достаточно большое количество особей и проводят соответствующие измерения. Средние арифметические, в силу закона больших чисел, дают достаточно хорошее приближение для средних величин по всему виду. Это значит, что для определения, например, средней массы особей данного вида не нужно взвешивать всех представителей, достаточно взвесить 100 независимо отобранных особей. ■

Пример 2. Давление газа на стенку сосуда определяется импульсами молекул газа, соударяющихся со стенкой. Для простоты возьмем молекулы, летящие перпендикулярно к поверхности стенки. Если молекула массы m соударяется со стенкой со скоростью v, то в результате упругого удара ее скорость, оставаясь по модулю той же, меняет знак. Изменение импульса этой молекулы равно $2mv$. Если за некоторое время t со стенкой столкнулось n молекул и их скорости были $v_1, v_2, \ldots, v_n$, то общее изменение импульса молекул газа равно $\sum\limits_{i=1}^{n} 2mv_i$. На основании второго закона Ньютона он равен импульсу силы со стороны стенки на газ; т. е. произведению силы на время. Сила воздействия стенки сосуда на газ равна на основании третьего закона Ньютона силе воздействия газа на стенку, т. е. силе давления газа. Если площадь стенки S, давление p, то

$$\sum_{i=1}^{n} 2mv_i = tSp; \quad p = \frac{2m}{tS} \sum_{i=1}^{n} v_i.$$

Скорости отдельных молекул случайны, справа стоит случайная величина. Тем не менее давление газа постоянно. Это результат действия закона больших чисел. Вместо суммы справа можно рассматривать ее математическое ожидание. ■ ◄

Как следствие из теоремы Чебышёва, может быть получена ***теорема Бернулли.***

ТЕОРЕМА 2. *Пусть событие A наступает k раз в n независимых испытаниях, p — вероятность наступления этого события в каждом испытании. Тогда для любого $\varepsilon > 0$ $P\left(\left|\frac{k}{n} - p\right| \leqslant \varepsilon\right) \to 1$ при $n \to \infty$.*

Эта теорема утверждает, что с вероятностью, сколь угодно близкой к 1, при достаточно большом числе независимых опытов относительная частота события A как угодно мало отличается от ее вероятности в каждом опыте. Утверждение теоремы Бернулли целиком согласуется с экспериментально установленным свойством устойчивости частот.

Докажем эту теорему.

❐ Для доказательства обозначим через X_m число появлений события A в m-м испытании ($m = 1, 2, \ldots, n$). Тогда

$$k = X_1 + X_2 + \ldots + X_m, \quad MX_m = p, \quad DX_m = p(1 - p).$$

Случайные величины $X_1, X_2, \ldots, X_m$ независимы. Выполняются все условия теоремы Чебышёва. Отсюда следует утверждение теоремы Бернулли. ■

? Вспомните, как выводятся соотношения $MX_m = p$, $DX_m = p(1 - p)$.

Какие условия теоремы Чебышёва надо проверить при доказательстве теоремы Бернулли?

Теорема Бернулли является простейшей формой закона больших чисел. Она принадлежит швейцарскому математику Я. Бернулли (1654–1705), и была опубликована в 1713 г. Обобщение теоремы Бернулли получил С. Пуассон (1781–1840). В отличие от теоремы Бернулли, Пуассон допускает, что вероятность наступления события в отдельных испытаниях может быть различной. Во второй половине XIX века создатель российской школы теории вероятностей П. Л. Чебышёв (1821–1894) при помощи неравенства, носящего теперь его имя, получил в более общей форме закон больших чисел – теорему Чебышёва. Из этой теоремы как следствие получаются теоремы Бернулли и Пуассона. Условия применимости закона больших чисел в дальнейшем расширили русские математики А. А. Марков (1856–1922), А. Н. Колмогоров (1903–1992), А. Я. Хинчин (1894–1959).

Контрольные вопросы

1. Пусть событие A наступает m раз в n независимых испытаниях, p — вероятность события A в каждом испытании.

а) Что означает неравенство:

$$\left|\frac{m}{n} - p\right| > \varepsilon; \quad \left|\frac{m}{n} - p\right| \leqslant \varepsilon?$$

б) Что необходимо знать, чтобы с помощью теоремы Бернулли оценить:

— точность приближенного равенства $\frac{m}{n} \approx p$;

— надежность приближенного равенства $\frac{m}{n} \approx p$?

в) Верно ли, что:

— $P\left(\left|\frac{m}{n} - p\right| \leqslant 1\right) \geqslant 1 - \frac{1}{4n}$;

— $P\left(\left|\frac{m}{n} - p\right| > 1\right) \leqslant \frac{1}{4n}$?

2. Пусть $MX = 0$, $DX = 1$. Укажите число из промежутка $(0; 1)$, которое меньше $P(-3 \leqslant X \leqslant 3)$.

Задачи

402. Пусть $MX = 7$, $DX = 4$.

а) Оцените с помощью неравенства Чебышева $P(|X - 7| > 2{,}5)$; $P(1 \leqslant X \leqslant 13)$.

б) При каком значении k выполняется неравенство $P(|X - 7| \leqslant k) \geqslant 0{,}99$?

403. Пусть случайная величина X принимает значения $-c$, 0, c с вероятностями, равными p, $1 - 2p$, p.

а) Найдите MX и DX.

б) Найдите зависимость между c и $\sigma = \sqrt{DX}$ при условии $P(|X - MX| \geqslant \sigma) = 1$.

404. Имеем следующие результаты измерений: $-8, -1, -1, 0, 0, 0, 0, 1, 1, 8$. Проверьте, что по крайней мере $\frac{3}{4}$ этих результатов удалены от выборочного среднего $\bar{x}$ на расстояние, не превышающее двух стандартных отклонений, и что по крайней мере $\frac{8}{9}$ этих результатов удалены от $\bar{x}$ на расстояние, не превышающее трех стандартных отклонений.

405.* Среднее значение измеряемой величины равно 200 м, а среднее квадратичное отклонение равно 5 м. Эту величину измерили n раз. При каких значениях n вероятность того, что среднее арифметическое результатов измерений отличается по модулю от его среднего значения более чем на:

а) 0,5 м; б) 0,1 м,

не превосходит 0,01?

406. Чтобы определить прочность волокон хлопка в данной партии, случайным образом отобрали 1000 образцов и испытали их на разрыв. При этом оказалось, что средняя прочность образцов была 5,2 г/мм^2. Оцените вероятность того, что ошибка при определении средней прочности волокон не превзойдет 0,3 г/мм^2, если известно, что дисперсия прочности образцов не превосходит 1,2. Среднюю прочность образцов волокон во всей партии принять равной математическому ожиданию средней прочности отобранных образцов.

407. Вероятность того, что автомат, продающий воду, при опускании в него монеты будет работать безошибочно, равна 0,96. Воспользовавшись теоремой Бернулли, найдите, сколько следует провести испытаний правильности работы автомата, чтобы с вероятностью не меньшей 0,99 можно было утверждать, что отклонение относительной частоты правильной работы автомата от вероятности его правильной работы по модулю не превосходит 0,03.

408. Станок-автомат изготовляет валики, средний диаметр которых равен 10 см. Диаметры отдельно взятых валиков отличаются от среднего в ту или иную сторону. Известно, что дисперсия диаметров валиков не превосходит 0,007. У скольких изготовленных валиков необходимо проверить величину диаметров, чтобы с вероятностью не меньшей 0,95 можно было утверждать, что их средний диаметр отличается от математического ожидания этого среднего (10 см) не более чем на 0,04 см?

409. Дана последовательность независимых случайных величин $X_1, X_2, \ldots, X_n, \ldots$, каждая из которых имеет закон распределения

x	$-\sqrt{3}$	0	$\sqrt{3}$
p	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Применима ли к этой последовательности теорема Чебышёва?

§ 4.10. Нормальное распределение

В § 4.1 уже упоминалось о том, что различают дискретные и непрерывные случайные величины. До сих пор рассматривались дискретные случайные величины. Непрерывная случайная величина может принимать любое значение из некоторого интервала. Например, дальность полета снаряда, величина ошибки измерения являются примерами непрерывных случайных величин. Для таких случайных величин предпочтительнее указывать вероятности не отдельных их значений, а целых участков таких значений, например вероятность того, что ошибка измерения заключена в пределах от $-0,5$ до $+0,5$ мм. Но в любом случае, чтобы иметь представление о случайной величине, нужно знать ее распределение.

Получить представление о распределении случайной величины можно чисто опытным путем, анализируя, например, гистограммы частот. Но зачастую эта задача становится очень трудоемкой, требует проведения большого числа опытов. Поэтому естественными были попытки чисто теоретически установить общие типы распределений, наличие которых можно было бы предсказывать, а потом подтверждать эти предположения с помощью экспериментов.

В частности, этим путем пришли к ***нормальным распределениям*** случайных величин. Оказалось, что закон распределения случайной величины, являющейся суммой очень большого числа независимых в совокупности случайных величин, какова бы ни была природа слагаемых, лишь бы каждое из них было мало по сравнению со всей суммой, должен быть близок к нормальному закону.

➤

Приведем примеры явлений, протекающих по только что описанной схеме.

При стрельбе из орудия по цели неизбежны отклонения точки попадания снаряда от точки прицеливания. Это — хорошо известное явление рассеяния снарядов. Так как рассеяние является результатом воздействия огромного числа независимо действующих факторов (например, неправильности в обточке стакана снаряда, головки снаряда, колебания в плотности материала, из которого выточена головка снаряда, ничтожные колебания количества взрывчатого вещества в различных снарядах, малые, незаметные для глаза ошибки в наводке орудия, ничтожные изменения состояния атмосферы при различных стрельбах и многие другие), каждый из которых лишь ничтожно мало влияет на траекторию снаряда, то из приведенных выше соображений оно должно подчиняться нормальному закону.

Когда производится какое-нибудь наблюдение с целью измерить ту или иную физическую константу, то на результат наблюдения неизбежно влияет огромное количество факторов, каждый из которых в отдельности невозможно учесть, но которые порождают ошибки в измерении. Сюда относятся ошибки в состоянии измерительного прибора, показания которого могут нечувствительно меняться под влиянием различных атмосферных, тепловых, механических и других причин. Сюда также относятся ошибки наблюдателя, вызываемые особенностями его зрения или слуха и также нечувствительно меняющиеся в зависимости от психического или физического состояния наблюдателя. Таким образом, фактическая ошибка измерения является результирующей огромного количества ничтожных по величине, независимых между собой ошибок, зависящих от случая. Поэтому можно ожидать, что ошибки наблюдений будут подчинены нормальному закону распределения. ◄

Ранее биномиальные законы распределения изображались графически, в виде диаграмм. Точно также можно изображать законы распределения произвольных дискретных случайных величин. На горизонтальной оси откладывают все различные значения данной случайной величины. Против каждого возможного значения откладывают по вертикали вверх вероятность этого значения. Масштаб в обоих направлениях выбирают такой, чтобы вся диаграмма имела удобную и легко обозримую форму. Из рис. 54 легко установить, например, что наивероятнейшим значением случайной величины является x_5. Вероятность того, что случайная величина примет значение, заключенное в интервале $(\alpha; \beta)$, по теореме сложения вероятностей равна сумме вероятностей всех возможных значений, лежащих в этом интервале, и геометрически изображается суммой длин вертикальных отрезков, расположенных над этим отрезком: $P(\alpha < X < \beta) = p_1 + p_2 + p_4 + p_5$.

? Чему равна вероятность $P(x_6 < X < x_2)$?

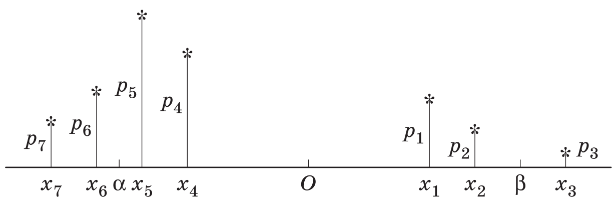


Рис. 54

Если число возможных значений случайной величины очень велико, то, для того чтобы чертеж не сильно растянулся по горизонтали, выберают большой масштаб в горизонтальном направлении, вследствие чего возможные значения располагаются достаточно густо, так что верхние концы вертикальных отрезков практически сливаются в одну сплошную линию, которую называют ***кривой распределения*** данной случайной величины (рис. 55).

Предположим, что расстояния между двумя возможными соседними значениями случайной величины равны 1. Тогда длина каждого вертикального отрезка численно равна площади прямоугольника, высотой которого служит этот отрезок, а основанием — равное 1 расстояние от него до соседнего отрезка (рис. 56). Тогда вероятность $P(\alpha < X < \beta)$ графически может быть изображена суммой площадей прямоугольников, расположенных над интервалом $(\alpha; \beta)$.

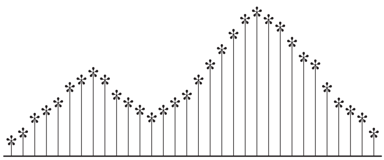


Рис. 55

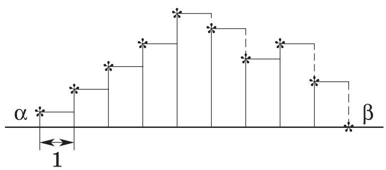


Рис. 56

Если возможные значения случайной величины расположены очень густо, то сумма площадей таких прямоугольников практически не будет отличаться от площади криволинейной фигуры, ограниченной снизу отрезком $[\alpha; \beta]$, сверху — кривой распределения, а с боков — вертикальными отрезками, проведенными из точек α и β (рис. 57).

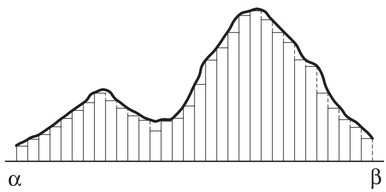


Рис. 57

Если закон распределения случайной величины задан такой криволинейной диаграммой, то вопрос о вероятности отдельных значений теряет свою актуальность: если значений очень много, то вероятности отдельных значений будут практически равны нулю. Так, при измерении дальности стрельбы орудия совсем несущественно знать, что место попадания отклонится от цели ровно на 156 см. Более важной является вероятность того, что отклонение места попадания от цели заключено, например, в промежутке от −1 до 1 м. Другими словами, интерес представляют вероятности того, что случайная величина примет значения на заданных отрезках. Но эти вероятности задаются наглядно и непосредственно на криволинейных диаграммах.

Случайная величина, распределенная по *нормальному закону*, принимает бесконечное множество значений, поэтому нормальные законы удобно графически изображать криволинейными диаграммами. На рис. 58—60 представлены кривые распределения по нормальному закону. Несмотря на различия, они имеют общие черты.

1) все кривые имеют наивысшую точку, при удалении от которой точки кривой понижаются;

2) все кривые симметричны относительно вертикальной прямой, проведенной через наивысшую точку;

3) все кривые имеют колоколообразную форму.

Для всякой кривой распределения площадь, расположенная под ней, равна 1, так как эта площадь равна вероятности того, что

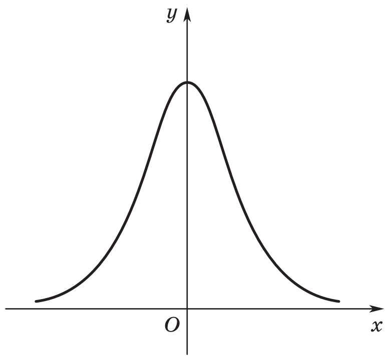


Рис. 58

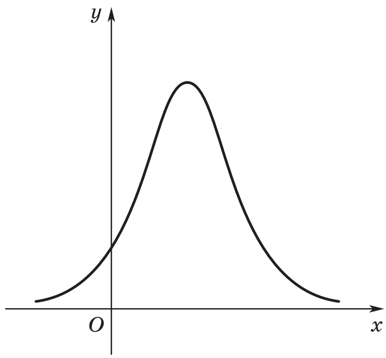


Рис. 59

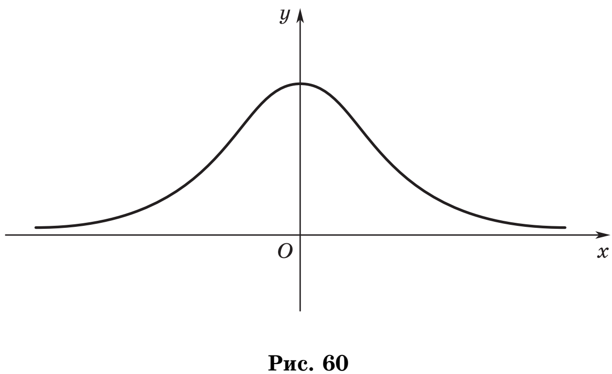


Рис. 60

данная случайная величина примет какое бы то ни было из своих значений, т. е. вероятности достоверного события.

Более вероятным оказывается наблюдение значений вблизи центра кривой, там, где она поднимается выше. Вблизи краев, где кривая проходит ниже, наблюдение соответствующих значений оказывается менее вероятным.

Вероятность того, что значение попадет в некоторый интервал на числовой прямой, равна площади соответствующей области под кривой (см. рис. 61).

Особенностью нормального распределения является то, что, зная только среднее значение и среднее квадратичное отклонение, можно вычислить любую представляющую интерес вероятность (конечно, при условии, что распределение — действительно нормальное).

Для любой комбинации значений среднего и среднего квадратичного отклонения можно построить свой график. Кривая сдвигается вправо или влево таким образом, что вершина «колокола» располагается над средним значением, и растягивается или сжимается так, что масштаб по горизонтали соответствует среднему квадратичному отклонению.

На рис. 61 изображено нормальное распределение со средним a и средним квадратичным отклонением σ. Обратите внимание на то, что среднее может быть любым действительным числом, среднее квадратичное отклонение — любым положительным числом. На рис. 62 изображены два различных нормальных распределения. Кривой, расположенной левее, соответствует среднее значение 20 и среднее квадратичное отклонение 5. Расположенной спра-

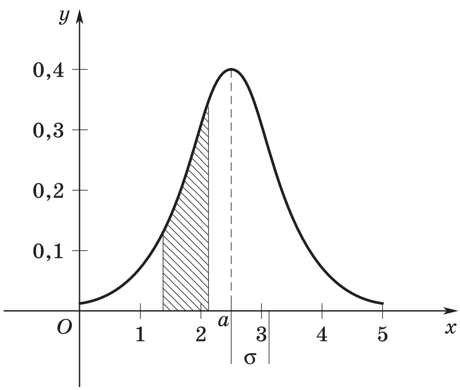


Рис. 61

ва кривой соответствует среднее значение 40 и среднее квадратичное отклонение 10.

Нормальное распределение со средним значением $a = 0$ и средним квадратичным отклонением $\sigma = 1$ называют ***стандартным нормальным распределением***. Случайную величину, имеющую стандартное нормальное распределение, будем обозначать через Z. Составлены таблицы вероятностей того, что случайная величина Z, имеющая стандартное нормальное распределение, принимает значение меньше заданного числа z (см. Приложение, табл. П.1). Например, вероятность того, что величина Z меньше 1,54, равна $P(Z < 1{,}54) = 0{,}9382$. Имея эту таблицу, можно вычислить следующие вероятности:

$P(Z < z)$ — вероятность того, что Z меньше z — находится по таблице;

$P(Z > z)$ — вероятность того, что Z больше z — из 1 вычитается предыдущий результат, как вероятность противоположного события;

$P(z_1 < Z < z_2)$ — вероятность того, что Z лежит между z_1 и z_2 — из вероятности $P(Z < z_2)$ вычитается вероятность $P(Z < z_1)$, которые находятся по таблице (на основании теоремы сложения вероятностей);

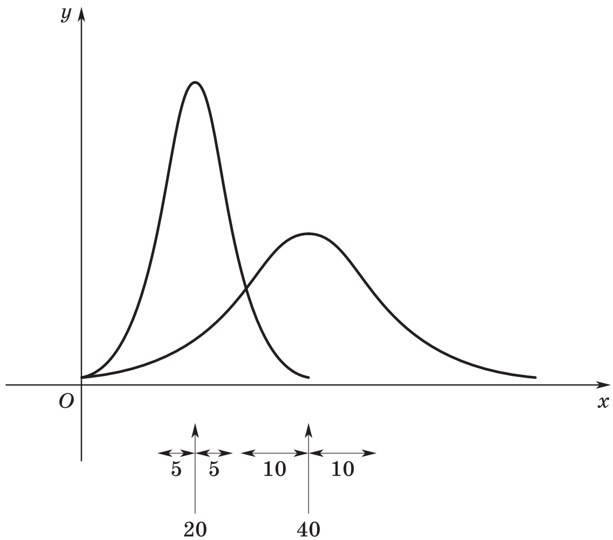


Рис. 62

$P((Z < z_1) + (Z > z_2))$ — вероятность того, что Z лежит за пределами интервала от z_1 до z_2 — из 1 вычитается предыдущий результат (свойство вероятностей противоположных событий).

? Противоположным событию $(Z < z)$ является не событие $(Z > z)$, а событие $(Z \geqslant z)$. Почему можно считать, что $P(Z > z) = 1 - P(Z < z)$?

Противоположным событию $(z_1 < Z < z_2)$ является не событие $((Z < z_1) + (Z > z_2))$, а событие $((Z \leqslant z_1) + (Z \geqslant z_2))$. Почему можно считать, что $P((Z < z_1) + (Z > z_2)) = 1 - P(z_1 < Z < z_2)$?

Случайную величину, имеющую произвольное нормальное распределение, можно выразить через случайную величину, имеющую стандартное нормальное распределение, если только известны среднее значение данной случайной величины и среднее квадратичное отклонение. Поэтому нет необходимости составлять таблицы вероятностей для каждой возможной комбинации среднего значения и среднего квадратичного отклонения.

Мы не будем доказывать сформулированное утверждение, а покажем на примерах, как это делается.

Пример 1. На станке изготовляют некоторую деталь. Ее длина представляет собой случайную величину, распределенную по нормальному закону со средним значением 20 см и средним квадратичным отклонением 0,2 см.

а) Найти вероятность того, что длина детали будет заключена между 19,7 и 20,3 см, т. е. что отклонение от среднего в ту или иную сторону не превзойдет 0,3 см.

б) Какую точность длины изделия можно гарантировать с вероятностью 0,95?

❐ а) Обозначим длину детали через X. Требуется найти $P(19{,}7 < X < 20{,}3)$, или $P(|X - 20| < 0{,}3)$.

Перейдем от случайной величины X к случайной величине $Z = \dfrac{X - a}{\sigma} = \dfrac{X - 20}{0{,}2}$. Эта случайная величина имеет стандартное нормальное распределение (как раз это утверждение мы не будем доказывать). Нетрудно доказать, что математическое ожидание случайной величины Z равно 0, а дисперсия — 1. Воспользуемся свойствами математического ожидания и дисперсии случайной величины. Эти свойства были доказаны для дискретных случайных величин, они остаются справедливыми и для непрерывных.

Действительно,

$$MZ = M\frac{X-a}{\sigma} = \frac{1}{\sigma}(MX - a) = \frac{1}{\sigma}(a - a) = 0;$$

$$DZ = D\frac{X-a}{\sigma} = \frac{1}{\sigma^2}DX = \frac{\sigma^2}{\sigma^2} = 1.$$

Основную трудность составляет доказательство того, что случайная величина Z имеет нормальное распределение.

Далее, вычитая из всех трех частей неравенства $19{,}7 < X < 20{,}3$ среднее значение 20 и разделив все три части полученного неравенства на среднее квадратичное отклонение 0,2, имеем:

$$P(19{,}7 < X < 20{,}3) = P(19{,}7 - 20 < X - a < 20{,}3 - 20) =$$
$$= P(-0{,}3 < X - a < 0{,}3) = P\left(-\frac{0{,}3}{0{,}2} < \frac{X-a}{\sigma} < \frac{0{,}3}{0{,}2}\right) =$$
$$= P(-1{,}5 < Z < 1{,}5) = P(Z < 1{,}5) - P(Z < -1{,}5) =$$
$$= 0{,}9332 - 0{,}0668 = 0{,}8664.$$

Итак, около 87% всех деталей, изготовленных при данных условиях, будут иметь длины между 19,7 и 20,3 см; остальные 13% деталей будут иметь большие отклонения от среднего.

Заметим, что $P(Z < 1{,}5) = 1 - P(Z < -1{,}5) = 0{,}9332 = 1 - 0{,}0668$. Оказывается, этот факт имеет место для произвольного числа c:

$$P(Z < c) = 1 - P(Z < -c).$$

В этом можно убедиться, рассматривая кривую стандартного нормального распределения и используя ее симметричность относительно оси ординат.

Из рис. 63 видно, что $P(Z < -c) = P(Z > c)$. По свойству вероятностей противоположных событий, $P(Z > c) = 1 - P(Z < c)$. Поэтому $P(Z < -c) = 1 - P(Z < c)$.

б) Требуется найти такое положительное число c, для которого

$$P(|X - 20| < c) > 0{,}95.$$

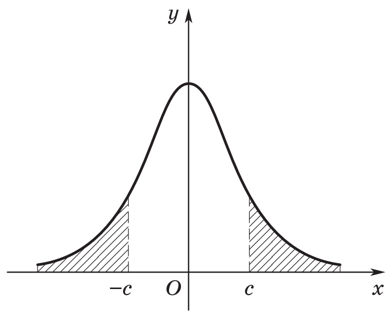


Рис. 63

Число c и называют *точностью* длины изделия, которую можно гарантировать с вероятностью 0,95 (или с надежностью 0,95). Вычисления, проведенные при решении первой части данного примера, говорят о том, что 0,3 является точностью длины детали, но с надежностью 0,866.

Разделив обе части неравенства $|X - 20| < c$ на среднее квадратичное отклонение $\sigma = 0{,}2$ и переходя от модуля к двойному неравенству, получим:

$$0{,}95 < P(|X - 20| < c) = P\left(\frac{|X-20|}{\sigma} < \frac{c}{0{,}2}\right) =$$

$$= P\left(-\frac{c}{0{,}2} < \frac{X-20}{\sigma} < \frac{c}{0{,}2}\right) = P\left(-\frac{c}{0{,}2} < Z < \frac{c}{0{,}2}\right) =$$

$$= P\left(Z < \frac{c}{0{,}2}\right) - P\left(Z < -\frac{c}{0{,}2}\right) = P\left(Z < \frac{c}{0{,}2}\right) - 1 + P\left(Z < \frac{c}{0{,}2}\right) =$$

$$= 2P\left(Z < \frac{c}{0{,}2}\right) - 1.$$

Отсюда $P\left(Z < \dfrac{c}{0{,}2}\right) > 0{,}975.$

Воспользовавшись таблицей для стандартного нормального распределения, получим, что $\dfrac{c}{0{,}2} = 5c > 1{,}96$, или $c > 0{,}392$.

Таким образом, с вероятностью, превосходящей 0,95, можно гарантировать, что отклонение длины детали от среднего не превзойдет 0,4 см. ■

? Имеет ли место равенство $P(Z < -c) = 1 - P(Z < c)$ для нормального распределения, отличного от стандартного?

Пример 2. Руководство фирмы высказало претензии к отделу прогнозов: объем продаж на текущий квартал прогнозировался на уровне 18 млн р., однако достигнутый уровень составил 21 300 000 р. На следующий квартал прогнозируется объем продаж в 20 млн р. со средним квадратичным отклонением (которое установлено на основании прошлого опыта работы) в 3 млн р. Предполагая, что объем продаж имеет нормальное распределение с центром в прогнозируемом значении, найти вероятность того, что в следующем квартале объем продаж составит:

а) меньше 15 млн р., т. е. квартал будет очень неудачным;

б) больше 24 млн р., т. е. квартал будет очень удачным;

в) от 16 млн р. до 23 млн р., т. е. квартал будет типичным;

г) либо меньше 16 млн р., либо больше 23 млн. р., т. е. квартал будет необычным.

❒ От случайной величины X — объема продаж — перейдем к случайной величине $Z = \frac{X-a}{\sigma}$, имеющей стандартное нормальное распределение.

а) Требуется найти $P(X < 15\ 000\ 000)$. Аналогично решению предыдущего примера имеем (в дальнейшем будем выражать данные в млн р.):

$$P(X < 15) = P\left(\frac{X-a}{\sigma} < \frac{15-20}{3}\right) = P(Z < -1{,}67) = 0{,}0475,$$

т. е. квартал окажется неудачным с вероятностью примерно равной 5%. Эта вероятность невелика, но и ее не следует недооценивать.

б) Требуется найти $P(X > 24)$. Воспользовавшись свойством вероятностей противоположных событий, получим:

$$P(X > 24) = 1 - P(X < 24) = 1 - P\left(\frac{X-a}{\sigma} < \frac{24-20}{3}\right) =$$
$$= 1 - P(Z < 1{,}33) = 1 - 0{,}9082 = 0{,}0918,$$

т. е. с вероятностью, примерно равной 9%, квартал будет удачным.

в) Требуется найти $P(16 < X < 23)$.

$$P(16 < X < 23) = P\left(\frac{16-20}{3} < \frac{X-a}{\sigma} < \frac{23-20}{3}\right) =$$
$$= P(-1{,}33 < Z < 1{,}00) = P(Z < 1{,}00) - P(Z < -1{,}33) =$$
$$= 0{,}8413 - 0{,}0918 = 0{,}7495,$$

т. е. с вероятностью, примерно равной 75%, квартал будет типичным.

г) Требуется найти $P(X < 16$ или $X > 23)$. Снова применив свойство вероятностей противоположных событий, получим:

$$P(X < 16 \text{ или } X > 23) = 1 - P(16 < X < 23) =$$
$$= 1 - 0{,}7495 = 0{,}2505,$$

т. е. с вероятностью, примерно равной 25%, квартал будет необычным. ■

Почему можно считать, что объем продаж имеет нормальное распределение?

На рис. 64 изображена кривая нормального распределения со средним 20 и средним квадратичным отклонением 3. Выделите (различной штриховкой) области под этой кривой, соответствующие вероятностям, рассмотренным в заданиях а)—г).

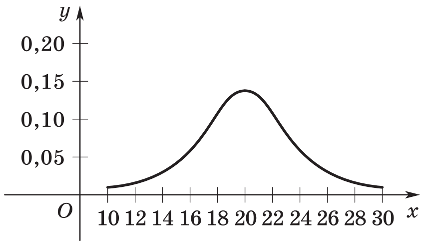


Рис. 64

Для вычисления вероятностей, рассмотренных в примерах 1 и 2, можно воспользоваться программой Excel — функция (=НОРМРАСП(Значение; Среднее; СтандОткл; Истина)). Эта функция дает возможность найти вероятность того, что случайная величина, имеющая нормальное распределение с некоторыми средним и средним квадратичным отклонением, окажется меньше некоторого значения.

Выполните вычисления, необходимые для решения примеров 1 и 2 с помощью программы Excel.

Контрольные вопросы

1. Какие из следующих случайных величин являются дискретными, а какие — непрерывными:

а) число попыток, которые нужно сделать, чтобы запустить двигатель;

б) число аварий на перекрестке;

в) дальность полета снаряда;

г) диаметр валика, изготовленного на станке;

д) число людей, обратившихся в фирму по объявлению о наличии вакансии?

2. Почему можно считать, что урожайность пшеницы у определенного фермера может иметь нормальное распределение?

3. Если известно, что урожайность пшеницы у фермера имеет нормальное распределение, то может ли она иметь стандартное нормальное распределение?

4. Случайная величина X имеет нормальное распределение со средним 10. Сравните:

а) $P(X > 10)$ и $P(X < 10)$;

б) $P(X > 10)$ и $P(X < 9)$;

в) $P(X > 11)$ и $P(X < 10)$;

г) $P(X > 2)$ и $P(X \geqslant 2)$.

5. Если случайная величина имеет нормальное распределение со средним 10 и дисперсией 4, то какое распределение имеет случайная величина $\frac{X-10}{2}$?

Задачи

410.° В обычных условиях нефтеперегонный завод может перерабатывать в среднем 135 000 баррелей сырой нефти в день со средним квадратичным отклонением 6000 баррелей в день. Предполагая, что объем обрабатываемой нефти подчиняется нормальному распределению, найдите вероятность того, что за один день будет переработано:

а) более чем 135 000 баррелей нефти;
б) более чем 130 000 баррелей нефти;
в) более чем 150 000 баррелей нефти;
г) менее чем 125 000 баррелей нефти;
д) менее чем 100 000 баррелей нефти.

411.° По стандарту требуется, чтобы диаметр клапана находился в пределах от 2,53 до 2,57 см. Оборудование, на котором производятся клапаны, налажено так, что средний диаметр клапана равен 2,56 см, а среднее квадратичное отклонение составляет 0,01 см. Предполагая, что диаметр клапана имеет нормальное распределение, найдите, какой процент произведенных в течение длительного времени клапанов будет соответствовать стандарту.

412. В момент закрытия биржи значение индекса активности было равно 9246 пунктам. На следующий день ожидается подъем в среднем на 4 пункта со средним квадратичным отклонением 115 пунктов. Предполагая, что величина индекса имеет нормальное распределение, найдите вероятность того, что на следующий день будет наблюдаться:

а) снижение индекса активности;
б) повышение индекса активности более чем на 50 пунктов;
в) повышение индекса активности более чем на 100 пунктов;
г) снижение индекса активности более чем на 150 пунктов;
д) колебание, превышающее 200 пунктов в любую сторону.

413.* Для некоторых практических вопросов считают, что случайная величина X, распределенная по нормальному закону, не обнаруживает отклонения от среднего, больше, чем на три средних квадратичных отклонений. Какие для этого имеются основания?

414. Какова вероятность того, что случайная величина X, распределенная по нормальному закону, отклоняется от среднего не больше, чем на два средних квадратичных отклонения?

415. Стрельбу ведут из некоторой точки вдоль некоторой прямой. Средняя дальность полета снаряда равна 1200 м. Предполагая, что дальность полета распределена по нормальному закону со средним квадратичным отклонением 40 м, найдите, какой процент выпускаемых снарядов даст перелет от 60 до 80 м.

416.* Предположим, что случайная величина X имеет биномиальное распределение с параметрами 100 и 0,1, а случайная величина Y — нормальное распределение, причем ее среднее значение и среднее квадратичное отклонение совпадают соответственно со средним значением и средним квадратичным отклонением случайной величины X. Сравните:

а) $P(X = 8)$ и $P(7{,}5 < Y < 8{,}5)$;

б) $P(15 < X < 23)$ и $P(14{,}5 < Y < 23{,}5)$.

417.* Решите задачу 416, а в предположении, что случайная величина X имеет биномиальное распределение с параметрами 10 и 0,1. Чем вы объясните существенное различие результатов?

Дополнительные задачи к главе 4

К § 4.1. Случайная величина, закон ее распределения

418. Проводят независимые испытания, в каждом из которых с вероятностью 0,6 может наступить событие A. Испытания проводят до первого наступления события A, но общее число испытаний не превышает четырех. Найдите:

а) закон распределения числа испытаний;

б) вероятность того, что будет проведено более двух испытаний.

419. Стрелок, имея три патрона, стреляет до первого попадания в цель. Вероятность попадания в цель при каждом выстреле равна 0,7. Найдите закон распределения числа выстрелов, выполненных стрелком.

420. Два правильных тетраэдра, грани каждого из которых пронумерованы цифрами 1, 2, 3, 4, брошены на пол. Составьте закон распределения:

а) суммы номеров нижних граней двух тетраэдров;

б) разности номеров нижних граней первого и второго тетраэдров;

в) наибольшего из номеров нижних граней двух тетраэдров;

г) наименьшего из номеров нижних граней двух тетраэдров.

421.* Три человека сдали свои шляпы в гардероб. В связи с тем что потух свет, шляпы возвращались наугад. Составьте закон распределения числа людей, которые получат свои собственные шляпы.

422. Подброшены три симметричные монеты. Найдите закон распределения:

а) числа выпавших гербов;

б) разности между числом выпавших гербов и числом выпавших цифр.

Постройте графики этих законов распределения.

423. Из чисел от 1 до 20 наугад выбирают одно число.

а) Составьте закон распределения числа его делителей.

б) Вычислите вероятность того, что выбранное число имеет не менее трех делителей.

424. В урне пять белых и 10 красных шаров. Наугад вынимают два шара. Составьте закон распределения числа извлеченных красных шаров.

425.* Игрок, заплатив за вход в игральный салон, получает право трижды бросить игральный кубик. После броска он получает выигрыш. Составьте закон распределения выигрыша, состоящего из стольких денежных единиц, сколько составляет:

а) сумма очков, не выпавших ни разу;

б) максимальное число очков, не выпавшее ни разу;

в) минимальное число очков, не выпавшее ни разу.

426. Иногда для принятия решения используют жребий с помощью пальцев. Два игрока одновременно показывают один или больше пальцев правой руки. Составьте закон распределения суммы количеств пальцев, показанных игроками.

427. Партия из восьми телевизоров содержит три неисправных телевизора. Из этой партии наугад выбирают два телевизора. Составьте закон распределения числа неисправных телевизоров среди выбранных.

К § 4.2. Математическое ожидание случайной величины

428. В соответствии со статистическими данными вероятность того, что 25-летний человек проживет еще один год, равняется 0,998. Страховая кампания предлагает 25-летнему человеку застраховаться на сумму 10 000 р., страховой взнос равняется 30 р. Какую прибыль ожидает получить страховая кампания при страховании одного 25-летнего человека?

429. Завод выпускает массовую продукцию. Если изделие, изготовленное заводом и реализованное, выходит из строя на протяжении года, то его заменяют запасным. Для определения необходимого числа запасных изделий были проведены наблюдения в 50 пунктах, в каждом из которых продали по 100 изделий. Выяснилось, что на протяжении года в 15 случаях ни одно изделие не вышло из строя, в 15 — одно, в 10 — два, в шести — три, в трех — четыре и в одном пять изделий вышли из строя. Считая, что выход из строя запасного изделия невозможен, найдите среднее число запасных изделий, которое необходимо выпускать заводу на каждые 100 изделий.

430. Два игрока играют в такую игру. Первый игрок дает второму 50 р., а потом бросает игральный кубик. Если выпадет n очков, то первый получает $\max(10n, 70 - 10n)$ р. от второго. Для кого из игроков выгодна эта игра?

431. При каждом ходе игрок бросает игральный кубик и получает столько очков, сколько выпадет. Вдобавок, если выпадет шесть очков, он тут же бросает кубик вторично и получает дополнительно столько очков, сколько выпадет. Сколько в среднем очков получает игрок за один ход?

432. Для условий задачи 425 установите, какую плату за вход в игральный салон можно считать справедливой.

433.* Рассмотрим такую игру. Сначала нужно купить любое количество жетонов по 20 р. за жетон. Потом игрок бросает игральный кубик и фиксирует число выпавших очков. Если выпало a очков, то за a жетонов он получает по 30 р. за жетон, а за остальные — по 10 р. Сколько жетонов целесообразно покупать?

434. Игральный автомат имеет два окошка, в каждом из которых появляются две картинки. В каждом окошке может появиться одна из трех картинок, называемых «звонком», «яблоком» и «вишней». Машина построена таким образом, что картинки в окошках появляются независимо одна от другой. После того как автомат запустили, в каждом окошке появляется одна картинка. Вероятности появления картинок представлены в таблице 4.12.

Таблица 4.12

Результат	Звонок	Вишня	Яблоко
Вероятность	0,4	0,5	0,1

Для запуска автомата необходимо заплатить 5 р. Игрок получает в случае появления двух картинок «яблоко» — 50 р., двух картинок «звонок» — 10 р., двух картинок «вишня» — 5 р.

Во всех остальных случаях игрок ничего не получает. Найдите математическое ожидание чистого выигрыша для игрока, заплатившего 5 р.

435. Четыре одинаковые электрические лампочки временно выкручивают из патронов и убирают в ящик. Потом их наугад вынимают из него и вкручивают в патроны. Чему равняется математическое ожидание числа лампочек, которые попали в тот патрон, из которого они были выкручены?

436. Случайная величина X принимает значения -1, 0, 1. Известно, что $\boldsymbol{M}X = 0{,}1$; $\boldsymbol{M}X^2 = 0{,}9$. Найдите вероятности, с которыми X принимает свои значения.

437. Рассмотрим такую игру. В урне пять белых шаров и один черный. Игрок наугад извлекает два шара и за каждый извлеченный белый шар получает 10 р., за черный — 100 р. За участие в этой игре надо платить. При какой стоимости участия в игре она принесет ее организатору прибыль?

438. Вероятность того, что 50-летний человек будет жить еще один год, равняется 0,988. Какую страховую премию может назначить наследникам этого человека страховая кампания в случае его смерти, если он купил страховой полис стоимостью 10 000 р., без учета административных затрат, прибыли и т. п.?

439. Пусть в игре в спортлото «5 из 36» заранее известно, что при 5, 4, 3 угаданных номерах выигрыш равняется соответственно 100 000, 1750, 80 р. Аналогично пусть в спортлото «6 из 49» заранее известно, что при 6, 5, 4, 3 угаданных номерах выигрыш равняется соответственно 100 000, 27 300, 420, 30 р. Какая из игр оказывается более выгодной для игрока, который собирается играть довольно много раз, если цена одного билета в обеих лотереях одинаковая?

К § 4.3. Свойства математического ожидания

440.° Найдите математическое ожидание суммы числа очков, которые выпадают при бросании двух игральных кубиков.

441. Производят три выстрела с вероятностями попадания в цель, равными соответственно 0,4; 0,3; 0,6. Найдите среднее число попаданий.

442.° Для условий задачи 418 (с. 396) найдите среднюю стоимость испытаний, если стоимость одного испытания составляет 450 р.

443. За выход оборудования из строя предприятие, изготовившее это оборудование, должно уплатить штраф за простой, половину стоимости материала, расходуемого на ремонт, и четвертую часть трудозатрат на ремонт. Математические ожидания штрафа за простой, стоимости материала, трудозатрат на ремонт равны соответственно 10 000, 5000 и 3000 р. Найдите математическое ожидание общей суммы затрат при выходе оборудования из строя.

444. Число заявок, которые поступают в две прачечные за один час, имеет соответственно законы распределения:

x	0	1	2	3	4
p	0,05	0,1	0,2	0,25	0,4

x	0	1	2	3	4
p	0,1	0,15	0,15	0,25	0,35

а) Какая прачечная больше загружена работой?

б) Найдите среднее число заявок, которые поступают в первую прачечную за семь часов.

в) Какое среднее число заявок поступает в обе прачечные за один час?

445. Игральный кубик бросают 10 раз. Пусть X — число появлений шестерки. Вычислите среднее число появлений шестерки.

446. В промежутке времени от 16 до 17 часов в понедельник может произойти или 0, или 1, или 2, или 3 автомобильные аварии; вероятности этого соответственно равны 0,94; 0,03; 0,02; 0,01. Найдите среднее число аварий:

а) в указанный промежуток времени;

б) на протяжении 100 таких промежутков.

447. Испытывают техническое устройство, состоящее из трех приборов. Вероятности выхода из строя для этих приборов соответственно равны 0,1; 0,05; 0,15. Найдите математическое ожидание числа приборов, вышедших из строя.

448.* В урне две монеты по 5 р., две — по 10 р. и две — по 1 р. Игрок имеет право наугад вытащить для себя три монеты. Сколько можно заплатить за это право, т. е. чему равняется цена игры?

К § 4.4. Формула Бернулли

449.° На испытательный стенд поставлено 4 конденсатора. Вероятность пробоя конденсатора до истечения 1000 часов равна 0,01. Найдите вероятность того, что в течение испытания откажет:

а) ровно три конденсатора;

б) не более одного конденсатора;

в) хотя бы один конденсатор.

450.° Вратарь парирует в среднем 30% всех одиннадцатиметровых штрафных ударов. Какова вероятность того, что он:

а) возьмет ровно два из четырех мячей;

б) возьмет хотя бы один из четырех мячей;

в) не возьмет ни одного из четырех мячей?

451. Что вероятнее выиграть у равносильного противника: три партии из четырех или пять из восьми?

452.° Из 10 выстрелов стрелок поражает цель в среднем восемь раз. Какова вероятность того, что из трех независимых выстрелов он точно два раза попадет в цель?

453. Подбрасывают пять симметричных монет. Какова вероятность того, что выпадет:

а) ровно три герба;

б) более одного герба;

в) хотя бы один герб?

454. Промышленную продукцию определенного вида изготавливают большими партиями. Из каждой партии случайным образом выбирают 20 изделий. Партию принимают, если выборка содержит не более трех дефектных изделий. Какова вероятность принятия партии, если в процессе производства в среднем 10% изделий получаются дефектными?

455. Стрелок производит три независимых выстрела по мишени. Вероятность попадания при каждом выстреле равна 0,9. Составьте закон распределения числа попаданий.

456. Устройство состоит из трех взаимно независимых элементов. Вероятность отказа каждого элемента равна 0,1.

а) Составьте закон распределения числа отказавших элементов в одном опыте.

б) Вычислите вероятность того, что отказавших элементов будет не менее двух.

457. В урне один красный и два белых шара, одинаковые во всем, кроме цвета. Из урны извлекаются три шара так, что перед извлечением следующего шара предыдущий возвращают в урну. Составьте закон распределения числа белых шаров среди извлеченных.

458. На фондовой бирже в неудачный день падает стоимость 80% ценных бумаг. Оценивая портфель, содержащий 15 ценных бумаг, вы предполагаете, что число понижающихся в цене бумаг имеет биномиальное распределение.

а) Какие допущения вы при этом делаете?

б) Найдите вероятность падения в цене всех 15 бумаг.

в) Найдите вероятность падения в цене 10 бумаг.

г) Найдите вероятность падения в цене 13 или более ценных бумаг.

459. В трех испытаниях Бернулли вероятность ровно двух «успехов» в 12 раз больше вероятности трех «успехов». Найдите вероятность «успеха» в каждом испытании.

460. Проводят три испытания Бернулли, вероятность наступления «успеха» в каждом испытании равна p. Докажите, что вероятность того, что число «успехов» равно 1 или 2, равна $3p(1-p)$.

461.* Сколько надо произвести независимых выстрелов, чтобы с вероятностью не менее 0,9 хоть один раз попасть в цель, если вероятность попадания при каждом выстреле равна 0,7?

462.* Какова должна быть вероятность попадания при каждом из 10 независимых выстрелов, чтобы с вероятностью не менее 0,9 имело место хотя бы одно попадание?

463.* Пользуясь средствами дифференциального исчисления, найдите значение p, при котором вероятность $P_n(m)$ того, что в n испытаниях Бернулли число «успехов» равно m, достигает наибольшего значения, и вычислите это наибольшее значение.

464.* Пользуясь средствами дифференциального исчисления, найдите значение λ, при котором пуассоновская вероятность $P(m) = \frac{\lambda^m}{m!}e^{-\lambda}$ достигает наибольшего значения, и вычислите это наибольшее значение.

465. На факультете обучается 1000 студентов. Какова вероятность того, что 1 сентября является днем рождения одновременно для k студентов данного факультета? Вычислите эту вероятность для $k = 0, 1, 2, 3$.

К § 4.5. Дисперсия случайной величины

466. Пусть X — число гербов, выпавших при подбрасывании одной симметричной монеты, а Y — число гербов, выпавших при подбрасывании двух симметричных монет. Сравните дисперсии случайных величин X и Y.

467. Найдите математическое ожидание и дисперсию числа делителей произвольного наугад выбранного натурального числа из совокупности 1, 2, ..., 20.

468. Найдите математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины X, имеющей закон распределения

x	9998	9999	10 000	10 001
p	0,1	0,2	0,3	0,4

469. Для лечения некоторой болезни используют пять лекарств a, b, c, d, e. Врач проводит сравнительное исследование трех из этих пяти лекарств, выбирая три лекарства наугад. Обозначим через X случайную величину, принимающую значение 0, если a не попадает в выбранные лекарства, и 1 — в противном случае. Вычислите MX и DX.

470. Вычислите математические ожидания и дисперсии случайных величин, рассмотренных в задаче 420.

471.* Из колоды в 36 карт наугад выбирают пять карт.

а) Найдите математическое ожидание и дисперсию числа красных карт среди выбранных.

б) Опыт с выбором пяти карт повторили 30 раз. Результаты представлены в таблице 4.13.

Таблица 4.13

Число красных карт	0	1	2	3	4	5
Число опытов	1	6	10	7	5	1

Вычислите выборочные среднее и дисперсию числа красных карт при одном выборе.

в) Сравните выборочные и теоретические характеристики.

472. В ходе опыта с разведением мышей были получены данные, представленные в таблице 4.14.

Таблица 4.14

Число мышей в одном помете n	1	2	3	4	5	6	7	8	9
Число случаев, когда в помете n мышей	7	11	16	17	26	31	11	1	1

Найдите среднее значение, дисперсию и стандартное отклонение этого выборочного распределения.

473.* Найдите математическое ожидание и дисперсию случайной величины, которая принимает значения 1, 2, 3, ..., n, каждое с вероятностью $\frac{1}{n}$.

474.* Пусть X — число очков, выпавших при бросании первого игрального кубика, Y — число очков, выпавших при бросании второго игрального кубика. Найдите закон распределения, математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины:

а) $Z = \max(X, Y)$; б) $T = \min(X, Y)$.

К § 4.6. Независимые случайные величины

475. Пусть случайная величина X равна числу очков, выпавших при первом броске игрального кубика, а Y — числу очков, выпавших при втором броске. Докажите, что случайные величины X и Y независимы при условии, что все исходы двух подбрасываний кубика равновозможны.

476. Трижды бросают или правильный игральный кубик с обычными числами очков 1, 2, 3, 4, 5, 6 на гранях или правильный кубик, на гранях которого обозначены числа очков 1, 1, 1, 6, 6, 6.

а) Найдите для обоих кубиков математические ожидания и дисперсии суммы числа выпавших очков.

б)* Каким кубиком лучше играть, чтобы с большей вероятностью набрать в сумме не менее 15 очков?

477. Вычислите дисперсии случайных величин, рассмотренных в задачах 440, 441, 445, 446, б (с. 400) при условии, что соответствующие опыты являются независимыми.

478. Дисперсия каждой из девяти попарно независимых случайных величин равна 36. Найдите дисперсию среднего арифметического этих величин.

479. Число бракованных изделий в любой партии, поставляемой заказчику, имеет закон распределения

x	0	1	2	3
p	0,6	0,25	0,1	0,05

Стоимость восстановления любого из бракованных изделий может принимать значения 1000, 1500, 2000 р. с вероятностями соответственно 0,5; 0,3; 0,2. Найдите средние потери потребителей одной партии изделий, если число бракованных изделий и стоимость восстановления изделий независимы.

К § 4.7. Числовые характеристики биномиального распределения

480. Техническая система содержит пять однотипных деталей, дублирующих работу друг друга. Каждая из них выходит из строя с вероятностью 0,1. Пусть X — число деталей, вышедших из строя. Найдите:

а) MX; б) DX; в) $\sigma(X)$.

481. Игральный кубик бросают 15 раз. Сколько раз в среднем может появиться 4 очка?

482. Найдите дисперсию числа появлений события в трех независимых испытаниях, если математическое ожидание этой случайной величины равно 0,9; вероятность наступления события от испытания к испытанию не меняется.

483. Производят независимые испытания с одинаковой вероятностью появления события A в каждом испытании. Найдите эту вероятность, если дисперсия числа появлений события A в трех независимых испытаниях равна 0,63.

484.* Устройство состоит из четырех элементов. Вероятность отказа любого элемента за время опыта равна 0,2. Найдите математическое ожидание числа опытов, в каждом из которых откажет ровно один элемент, если всего проведено 100 независимых опытов.

485. Для условий задачи 458 (с. 402) найдите:

а) Для какого среднего числа ценных бумаг, входящих в портфель, ожидается снижение стоимости?

б) Чему равно среднее квадратичное отклонение числа таких ценных бумаг, стоимость которых снижается?

486.* Случайная величина X имеет биномиальное распределение с параметрами $n = 10$ и $p = 0{,}2$. При каких значениях k значения X отличаются от среднего значения на величину, не превосходящую величины k средних квадратичных отклонений?

К § 4.8. Неравенство Чебышёва

487. Вероятность изготовления бракованного изделия равна 0,01. Оцените вероятность того, что среди 10 000 изделий число бракованных будет находиться в пределах между 85 и 115.

488.* Устройство состоит из 10 независимо работающих элементов. Вероятность отказа на протяжении суток для каждого элемента равна 0,05. С помощью неравенства Чебышёва оцените вероятность того, что модуль разности между числом отказавших элементов и средним числом отказов за сутки окажется меньше 2.

489.* В осветительную сеть включено параллельно 20 лампочек. Вероятность того, что на протяжении суток лампочка будет включена, равна 0,8. Пользуясь неравенством Чебышёва, оцените вероятность того, что модуль разности между числом включенных лампочек и средним числом включенных лампочек за сутки окажется:

а) не большим трех;

б) большим трех.

490.* Правильный игральный кубик бросают 2500 раз. Оцените вероятности получить при этом выпадение шестерки:

а) свыше 2000 раз;

б) свыше 1300 раз.

491.* Сколько следует сделать независимысых опытов, чтобы равенство $p \approx \frac{m}{n}$ с точностью до 0,05 выполнялась с вероятностью 0,95?

492.* С какой вероятностью выполняется неравенство $\left|\frac{m}{n} - p\right| \leqslant \leqslant 0{,}1$ при 100 независимых опытах?

493.* Укажите приближенно точность равенства $p \approx \frac{m}{n}$, если оно получено при 50 независимых опытах с вероятностью 0,9.

494. В некотором регионе в течение года родилось 39 200 детей, среди них мальчиков было 20 300. С какой вероятностью можно гарантировать, что относительная частота рождений мальчиков отличается по модулю от неизвестной вероятности рождения мальчика не более чем на 0,05?

495. Чтобы определить среднюю урожайность пшеницы на поле в 10 000 га, случайным образом отобрали по 1 $м^2$ с каждого гектара, определили урожайность на каждом из них и вычислили по этим данным среднюю (выборочную) урожайность. Оцените вероятность того, что выборочная средняя урожайность по модулю отличается от средней урожайности на всем поле не более чем на 0,1 ц/га, если известно, что среднее квадратичное отклонение урожайности для каждого гектара не превосходит 3 ц/га и если среднюю урожайность на всем поле считать равной математическому ожиданию выборочной средней урожайности.

К § 4.9. Закон больших чисел

496. На опытном участке выборочно обследовали диаметры 200 саженцев и по этим данным определили их выборочный средний диаметр. Известно, что среднее квадратичное отклонение диаметров не превышает 6 мм. Какое наибольшее отклонение выборочного среднего от среднего диаметра всех саженцев на участке можно гарантировать с вероятностью не меньшей:

а) 0,82; б) 0,96?

Средний диаметр всех саженцев принять равным математическому ожиданию выборочного среднего диаметра.

497. Вероятность изготовления бракованной радиолампы равна 0,02. Какое наименьшее число радиоламп нужно отобрать, чтобы с вероятностью не меньшей 0,8 можно было утверждать, что относительная частота бракованных среди них отличается от вероятности изготовления бракованной радиолампы по модулю не более чем на 0,003?

498.* Проведено n измерений, для результатов которых среднее арифметическое равно 0, а стандартное отклонение — 1. Каково минимальное число результатов измерений, лежащих в пределах от -3 до 3?

499. Вероятность изготовления прибора с тем или иным дефектом равна 0,2. Оцените вероятность того, что относительная частота приборов с дефектами среди 300 изготовленных отличается по модулю от математического ожидания этой относительной частоты не более чем на:

а) 0,15; б) 0,99.

500. Дана последовательность независимых случайных величин, каждая из которых имеет закон распределения

x	$-n\alpha$	0	$n\alpha$
p	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

Применима ли к заданной последовательности теорема Чебышёва?

К § 4.10. Нормальное распределение

501. При средней массе некоторого изделия 8,4 кг найдено, что отклонения, по модулю превосходящие 50 г, встречаются в среднем три раза на 100 изделий. Допуская, что масса изделий распределена по нормальному закону, найдите его среднее квадратичное отклонение.

502. Случайная величина X распределена по нормальному закону со средним значением a и средним квадратичным отклонением σ. Найдите вероятность того, что отклонение случайной величины от среднего по модулю будет находиться между числами c и d.

503.* Случайная величина X распределена по нормальному закону со средним значением a и средним квадратичным отклонением σ. Найдите вероятность того, что значения случайной величины X будут находиться между числами c и d.

504.* Случайная величина X распределена по нормальному закону со средним значением a и средним квадратичным отклонением σ. Найдите вероятность того, что отклонение случайной величины от среднего значения будет находиться между числами c и d.

Глава 5

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Как мы увидели выше, в теоретико-вероятностных задачах обычно по вероятности одних событий вычисляют вероятности других, более сложных; находят законы распределения случайных величин, их числовые характеристики. При этом вычисленная вероятность события позволяет предсказать значение частоты его появления в серии проведенных опытов. Значение математического ожидания случайной величины дает возможность оценить среднее значение, которое может принимать эта величина в результате ряда наблюдений. Более того, при проведении статистических экспериментов можно сравнить полученные результаты с предсказанными теоретически, оценить согласованность математической модели с изученным явлением. С другой стороны, во многих опытах неизвестны вероятности случайных событий, или законы распределения рассматриваемых случайных величин. В статистических задачах об этих вероятностных характеристиках судят по результатам реальных наблюдений, измерений, другими словами, на основании некоторой выборки.

В частности, вероятность случайного события полагают примерно равной его относительной частоте, математическое ожидание случайной величины считают приближенно равным выборочному среднему.

Разработка методов, позволяющих делать выводы на основе статистических данных, обобщить результаты выборочных наблюдений, — основная задача математической статистики.

В настоящей главе будут рассмотрены некоторые задачи математической статистики. Речь пойдет, во-первых, о построении ***случайной выборки***, анализ которой дает возможность делать статистические выводы обо всей совокупности. Например, как организовать опрос по политическим или маркетинговым вопросам, чтобы предсказать, как поведут себя в той или иной ситуации жители некоторого региона.

В математической статистике обосновывается тот факт, что выводы, полученные на основании анализа данных выборки наблюдений, можно распространить на всю исследуемую совокупность. При этом методы теории вероятностей позволяют оценивать надежность этих выводов. В задачи математической статистики входит также разработка требований к выборке, обеспечивающих достоверность полученных выводов.

Во-вторых, будет рассмотрена задача ***оценивания параметров совокупности***. Параметр представляет собой значение некоторой характеристики исследуемой совокупности. Простейший пример: событие A имеет вероятность p, значение которой неизвестно. Требуется оценить p по результатам испытаний. Обратите внимание на то, что речь идет не о нахождении вероятности события A, а об ее оценке. Результаты испытаний случайны, поэтому и оценка, которую мы можем получить с их помощью, также случайна. С помощью статистических правил можно найти приближенное значение неизвестного параметра в некотором обобщенном смысле: постоянная величина приближенно равна некоторой случайной величине, построенной по результатам наблюдений по некоторому правилу.

Примерами параметров, которые требуется оценить, могут служить также средний доход и дисперсия уровней дохода работающего населения того или иного региона.

Часто имеется информация только о некоторой выборке, являющейся лишь небольшой частью исходной совокупности. Характеристики, полученные для выборок, называют ***статистиками***. По их значениям можно дать оценку параметрам исходной совокупности.

При этом будем различать ***точечные оценки***, указывающие точное значение параметра, оцененного для исходной совокупности (например, оценкой неизвестной вероятности события является число 0,3), и ***интервальные оценки***, указывающие диапазон изменения значений по обе стороны от полученной точечной оценки, содержащей с определенной вероятностью истинное значение параметра (например, интервал (0,28; 0,32) содержит с достаточно большой вероятностью вероятность события A).

В-третьих, будет дано представление о ***проверке статистических гипотез***, которые могут существенно помочь в оценке той или иной ситуации, в обосновании принятия того или иного решения.

Примерами статистических гипотез являются гипотезы о значении вероятности события, о значении среднего случайной величины, о равенстве средних двух случайных величин, о равенстве вероятностей двух событий и т. п.

Две основные задачи математической статистики — оценка параметров и проверка статистических гипотез — будут решаться в основном для выборок из биномиального распределения: простой математический аппарат позволит на этом распределении познакомиться с основными идеями математической статистики, что дает возможность более менее сознательно пользоваться соответствующими статистическими процедурами для других широко распространенных распределений, например для нормального.

При этом нужно иметь в виду, что в статистике главное состоит не столько в использовании математических формул и проведении расчетов, сколько в определении последовательности хода рассуждений. Для каждой из решаемых далее задач будет дана схема этой последовательности.

§ 5.1. Генеральная совокупность и выборка

Необходимость проводить выборочные исследования может быть вызвана различными причинами.

1. Часто полное исследование изучаемого явления слишком дорогостоящее и длительное.

Пример 1. Фирмы, специализирующиеся на продаже женской одежды, не имеют возможности учесть вкус абсолютно всех покупательниц. Применяют другой метод учета вкуса покупателей: предсезонный каталог, в котором представлены несколько типов одежды разного стиля, рассылают некоторой выборке потенциальных покупательниц. На основе заказов, полученных от них, составляют основной каталог, содержащий модели одежды, которым было отдано предпочтение большинством опрошенных покупательниц. ■

Пример 2. Работник фирмы получает срочное задание: отследить реакцию клиентов фирмы на предлагаемые нововведения. За выделенное время обзвонить всех не представляется возможным. Приходится довольствоваться обзваниванием отобранного реального количества клиентов. При этом нужно иметь в виду, что руководство фирмы интересует реакция *всех* клиентов, а не только тех, кто попал в отобранный список. От способа составления этого списка зависит, будет ли полученная информация полезной или нет. ■

Пример 3. При аудиторской проверке фирм с большим числом сделок приходится довольствоваться изучением отобранного числа сделок. ■

2. Иногда возможность использовать полученную информацию при полном исследовании может исчерпаться раньше, чем завершится процесс его подготовки.

Пример 4. Измерение роста всех призывников с целью обеспечения соответствующей информацией швейные объединения, изготавливающие солдатскую форму одежды, мероприятие бессмысленное. Сбор этой информации обойдется слишком дорого, займет много времени, а сама информация практически будет устаревшей. В связи с этим о распределении роста всех призывников судят по некоторой выборке наблюдений, достаточно представительной и правильно организованной. По ней делают выводы относительно среднего роста и разброса его значений. ■

3. В некоторых случаях в результате проверки качества изделия происходит уничтожение исследуемого объекта.

Пример 5. Пусть электролампы проверяют на продолжительность горения, вплоть до выхода из строя. Если бы подобным образом испытывались все изготовленные лампы, то пришлось бы уничтожить всю произведенную продукцию. Поэтому для установления среднего времени горения лампы исследуют лишь некоторую ограниченную часть всех ламп. Аналогично расчет ожидаемого числа бракованных ламп во всей партии делается на основании проверки некоторой выборочной группы ламп.

Подобная ситуация будет иметь место при проверке качества фотобумаги. ■

?

В чем польза использования выборок?

Почему приходится прибегать к формированию выборок?

Почему прибегают к выборочному исследованию при контроле качества ампул для инъекций?

Рассмотрим, какие требования предъявляют к выборке.

Генеральная совокупность — это набор объектов, о которых необходимо получить информацию.

Выборка — это небольшой набор объектов, извлеченных из генеральной совокупности.

- В примере 1, связанном с фирмой по продаже женской одежды, генеральная совокупность состоит из всех потенциальных покупателей, которые могли обратиться к обычному каталогу данной фирмы, чтобы выбрать интересующую их модель одежды. Покупательницы, получившие предварительный каталог, представляют собой выборку.
- В примере 2, связанном с выяснением реакции клиентов фирмы на нововведения, все клиенты фирмы представляют собой генеральную совокупность. Те клиенты, которых обзвонили, образуют выборку.
- В примере 3 все сделки фирмы образуют генеральную совокупность, отобранные — выборку.
- В примере 4 генеральную совокупность образуют все призывники определенного года. Те из них, у которых снимали мерку для пошива одежды и обуви, образуют выборку.
- В примере 5 все лампы, изготовленные за определенное время на некотором предприятии, образуют генеральную совокупность. Те лампы, которые отобраны для контроля, — выборку.

Существует много различных способов построения выборок. Каждый способ может иметь преимущества для определенных целей. Рассмотрим несколько примеров генеральных совокупностей и построения выборок из них.

Пример 6. В городе, насчитывающем 253 000 жителей, имеющих право голосовать, исследуют политические симпатии будущих избирателей.

Выборку можно построить, опрашивая каждого 15-го покупателя, выходящего из крупного торгового центра. Такая выборка будет отражать мнение посетителей торгового центра, но вряд ли будет представлять точку зрения всех жителей города.

Другой метод построения выборки — провести опрос по телефону каждого 100-го жителя города, взяв номера из телефонного справочника. Такая систематическая выборка даст информацию о точке зрения группы людей, имеющих телефон, находящихся дома и отвечающих на телефонные звонки. Но она не отражает мнения всех жителей города.

Еще один метод построить выборку может заключаться в том, чтобы опросить участников митинга, организованного несколькими политическими партиями. Такая выборка даст информацию о жителях, активно участвующих в политической жизни города. ■

Итак, нужны такие способы образования выборки, которые представляли бы всю генеральную совокупность, т. е. выборка должна быть *репрезентативной* (представительной).

Пример 7. Генеральная совокупность — 756 ящиков с оборудованием, содержимое которых нужно проверить на соответствие документам.

Удобный способ заключается в том, чтобы отобрать 10 ближайших ящиков и проверить их содержимое. Такая выборка вряд ли будет представлять всю генеральную совокупность, т. е. вряд ли будет репрезентативной, так как поставщик мог позаботиться об укомплектовании какого-то числа первых и какого-то числа последних ящиков.

Можно выбрать для проверки три больших, три средних и три небольших по размеру ящика. Такой вариант может не дать желаемого результата, например, в случае, если большинство ящиков в генеральной совокупности окажется большого размера. Другими словами, и в этом случае выборка окажется нерепрезентативной.

Можно взять накладную и случайно отобрать ящики для проверки из перечня, указанного в накладной. Случайность отбора гарантирует, что поставщики не смогут предугадать, какие ящики будут отобраны для проверки. ■

Таким образом, для образования выборки целесообразно иметь перечень элементов генеральной совокупности и из него каким-то случайным образом организовывать выборку.

? Результаты переписи населения, проводимой в стране, являются генеральной совокупностью или выборкой?

Информация, полученная в результате построения выборки, будет только тогда надежной основой для принятия решения относительно тех или иных свойств исходной совокупности, когда структура образующих выборку элементов будет аналогична структуре элементов в генеральной совокупности. Такую выборку называют ***репрезентативной***.

Пример 8. Менеджер обувной фабрики большого города хочет выяснить, в каком количестве нужно шить обувь тех или иных размеров. Он должен составить представительную выборку жителей этого города. Ее объем может быть и не очень большим (например, 500 человек), но в качестве такой выборки нельзя брать только учащихся какой-то школы, или баскетболистов города. Очевидно, неплохую выборку могут составить жильцы многоквартирного дома или домов, расположенных в одном микрорайоне. Дело в том, что в многоквартирном доме живут люди разных возрастов, различных социальных слоев, различных профессий, поэтому жильцы такого дома могут в определенной мере представлять жителей большого города. Конечно, при этом не будут учтены потребности наиболее состоятельной части населения города, живущей в отдельных особняках, элитных квартирах. ■

Отдельно нужно поговорить об объеме выборки. Понятно, что она должна быть достаточно большого объема. Нельзя утверждать, что три четверти жителей Калуги по утрам пьют кофе на основании того, что из четырех калужан, которых мы рано утром встретили в кафе, трое пили кофе. Но и ошибочным является мнение о том, что объем выборки должен быть очень большим. Там, где выборка невелика, необходимо использовать наиболее тонкие математико-статистические методы; такие методы существуют, разрабатываются, но здесь мы их рассматривать не будем. Каким должен быть оптимальный объем выборки? Это основной вопрос, который приходится решать при практическом использовании выборочного метода. На численность выборки главным образом влияют следующие факторы:

1) степень колебаний значений изучаемого признака, которая характеризуется дисперсией: чем больше дисперсия, тем больше должна быть численность выборки;

2) значение допустимой ошибки случайной выборки: чем меньше допустимая ошибка, тем большей должна быть численность выборки, чтобы обеспечить требуемую высокую точность оценки;

3) уровень надежности (вероятности), с которой требуется гарантировать результаты, полученные по выборке: чем выше будет выбран уровень надежности, тем больше и должна быть численность выборки.

Ярким историческим примером неудачного применения выборочного метода являются результаты опроса, проведенного в 1936 году американским журналом «Literary Digest». Редакция журнала разослала 10 млн бюллетеней, в которых просила получивших их людей ответить, за кого они будут голосовать на предстоящих выборах – за кандидата республиканской партии А. Лэндона или за демократа Ф. Рузвельта. Было возвращено более 2 млн заполненных бюллетеней. Опубликованные в журнале результаты опроса предсказывали, что президентом станет А. Лэндон. Однако оказалось, что с большим преимуществом победу на выборах одержал Ф. Рузвельт, за которого проголосовало более 60% избирателей. Причина столь существенной ошибки журнала кроется в том, что полученная в результате проведения опроса выборка, на данных которой основывался прогноз, не была репрезентативной выборкой из генеральной совокупности избирателей. Бюллетени были разосланы подписчикам журнала, людям, чьи фамилии и адреса были взяты из телефонных справочников, а также владельцам автомобилей. Следовательно, в выборке были слишком плохо представлены менее состоятельные люди, которые в своей массе поддерживали «новый курс» Ф. Рузвельта. Кроме того, ответы прислали не все, а люди, не только достаточно уверенные в своем мнении, но и привыкшие отвечать на письма, т. е. в значительной мере представители делового мира, которые и поддерживали А. Лэндона.

В то же время американские социологи Дж. Гэллап и Э. Роупер правильно предсказали победу Ф. Рузвельта, основываясь всего на 4000 анкет. Причиной такого успеха, сделавшего славу[1] его авторам, было не только правильное составление выборки. Они учли, что общество распадается на социальные группы, которые более однородны по отношению к кандидатам в президенты. Поэтому выборка из слоя может быть относительно малочисленной с тем же результатом точности. Имея результаты исследования по слоям, можно характеризовать общество в целом.

Этот пример в политической жизни не единственный. В начале 70-х годов XX века много шума вызвала история о том, как «специалисты по изучению общественного мнения» подвели британскую партию консерваторов: они предсказали кандидатам этой партии успех на парламентских выборах. Правительство Э. Хита, доверившись этим прогнозам, провело выборы за полтора года до окончания полномочий и потерпело фиаско.

В конце 60-х годов XX века канадские институты по изучению общественного мнения предсказывали неудачу либеральной партии П. Э. Трюдо, а либералы одержали внушительную победу.

Чем объяснить такие провалы использования выборочного метода? На самом деле такие опросы зачастую далеки от использования научных методов прогнозирования. Реально фабрикуются нужные ответы, подтасовываются статистические данные, сказывается влияние тех, кто финансирует работу таких институтов.

Нерепрезентативные выборки называют ***смещенными***. Смещение — один из источников ошибок при использовании выборочного метода. Важно иметь в виду, что чем больше объем смещенной выборки, тем с большей вероятностью полученные данные приведут исследователя к ошибочному выводу. В этом смысле рассмотренный выше пример с опросом, проведенным американским журналом, является очень показательным.

Существует два основных типа выборок. После того как объект извлечен из генеральной совокупности для включения в выборку, его либо возвращают в генеральную совокупность, либо не возвращают. ***Выборка без возвращения*** имеет место, когда любой объект не может попасть в выборку более одного раза, т. е. когда все элементы выборки различны. ***Выборка с возвращением*** имеет место, если объект генеральной совокупности может попасть в выборку более одного раза. Мы будем в основном использовать *выборки без возвращения*.

[1] Именем Дж. Геллапа назван Американский институт общественного мнения, проводящий регулярные опросы населения по проблемам внутренней и внешней политики.

?

Чем отличается выборка с возвращением от выборки без возвращения?

Верно ли, что при построении выборки с возвращением все элементы генеральной совокупности имеют одинаковую вероятность попасть в выборку?

Верно ли, что при построении выборки без возвращения все элементы генеральной совокупности перед началом построения выборки имеют одинаковую вероятность попасть в выборку?

Через N будем обозначать объем генеральной совокупности, а через n — объем выборки.

Получить репрезентативную выборку можно с помощью случайного отбора входящих в эту выборку элементов.

Случайная выборка представляет собой выборку, при которой каждый отдельный элемент и каждая комбинация отдельных элементов, принадлежащих исходной совокупности, имеет одинаковые шансы попасть в выборку.

Создать случайную выборку не так просто. Покажем это на ряде примеров.

Пример 9. При создании выборки из некоторой генеральной совокупности людей может найтись некоторое число людей, которые бы хотели минимизировать вероятность попадания в эту выборку (преступники, лица, уклоняющиеся от кредиторов, от уплаты алиментов и т. д.). Некоторых людей трудно застать дома (находящихся в частых командировках, в отпусках и т. д.). Иногда трудно определить, принадлежит ли человек к интересующей исследователя генеральной совокупности (например, не все потенциальные избиратели примут участие в выборах). Перечень таких причин и подтверждающих их примеров можно было бы продолжить. ■

Трудности случайного отбора заключаются в том, что при таком отборе должна быть исключена всякая тенденциозность. Поэтому случайный отбор не может быть представлен как беспорядочный отбор. «Если беспорядочно втыкать булавки в карту, то это не даст случайного распределения точек на карте. Если отбирать данные для обследования, просто идя по улицам города, то это не будет случайным отбором домов»[1].

---

[1] *Ф. Йейтс.* Выборочный метод в переписях и обследованиях. — М.: Статистика, 1965.

Случайную выборку строят таким образом, что:

1) каждый объект генеральной совокупности имеет одинаковую вероятность быть отобранным;

2) объекты отбирают независимо друг от друга.

Независимость отбора обеспечивает сбор максимально возможного объема независимой информации. Рассмотрим пример, который поможет понять проблемы, возникающие, когда элементы совокупности отбирают не независимо.

Пример 10. В школе с 30 классами, в каждом из которых учится по 25 учеников, нужно составить случайную выборку. Если это сделать путем случайного выбора класса и проведения в нем опроса всех учащихся, то полученная выборка не будет случайной. Следует отметить, что каждый учащийся имеет одинаковую вероятность быть опрошенным $\left(\frac{1}{30}\right)$. Однако, так как вместо независимого отбора учащихся респондентов опроса отбирают целой группой, в этой выборке будет отсутствовать важная информация обо всех учащихся школы. ■

Одним из способов построения случайной выборки является применение ***таблиц случайных чисел*** для получения номера элемента генеральной совокупности, включаемого в выборку. Таблица случайных чисел представляет собой организованную в виде таблицы последовательность цифр, в которой каждая из цифр от 0 до 9 встречается независимо друг от друга с вероятностью 0,1 (см. Приложение, табл. П.2).

Схема образования случайной выборки может иметь следующий вид:

1. По какому-либо признаку пронумеровать элементы генеральной совокупности числами от 1 до N.

2. Выбрать точку начала считывания случайных чисел из таблицы случайным образом.

3. Начав с выбранной точки, последовательно записать цифры обычным образом, например слева направо с переходом на следующую строку.

4. Объединить эти цифры в группы, размер которых равен количеству цифр в числе N. Например, при объеме генеральной совокупности $N = 834$ нужно объединять случайные цифры в группы по три, так как число 834 содержит три цифры.

5. Следующие действия нужно выполнять до тех пор, пока не образуется выборка из n элементов:

— если получено случайное число между 1 и N и элемент с таким номером еще не извлекался, то его следует включить в выборку;

— если полученное число равно 0 или больше N, то его следует отбросить, так как для него нет соответствующего элемента генеральной совокупности;

— если получено такое случайное число, что элемент с соответствующим номером уже был извлечен ранее, то его нужно отбросить, так как строится выборка без возвращения.

Пример 11. Построить случайную выборку объемом $n = 15$ из генеральной совокупности, содержащей 105 элементов.

❐ Начнем с цифр 81677 (строка 21, столбец 1, табл. П.2). Так как число 105 состоит из трех цифр, то объединим последовательность случайных чисел в группы, состоящие из трех цифр, следующим образом:

816, 776, 263, 452, 794, $\underline{014}$, 668, 593, 814, 565, 799, 934, 495, 682, 254, 652, 234, 584, 901, 177, 137, 734, 352, 369, 825, 322, 258, 458, 774, 635, 852, 107, 273, 972, 529, 225, 790, 419, $\underline{012}$, 415, 251, 666, 293, 145, 362, 387, $\underline{078}$, 402, 417, 592, 623, 277, 422, 762, 895, 758, 742, 831, 870, 472, $\underline{009}$, 292, 676, 120, 174, 355, 487, 799, 336, $\underline{020}$, 193, 166, 913, 630, $\underline{080}$, 374, 593, 939, $\underline{071}$, 787, 000, 318, 158, 461, 206, 229, 869, 129, $\underline{078}$, 627, 673, 158, 527, 393, 424, 274, 957, $\underline{050}$, 917, 255, 329, 255, 652, 118, 344, 758, 125, 682, 640, 852, 658, 792, 289, 418, 533, 835, 481, 606, 560, $\underline{090}$, 602, 198, 392, 404, 508, 772, 150, 917, 169, 783, 947, 223, 505, 678, 194, 731, 496, 988, 899, 314, 939, 537, $\underline{071}$, 726, 585, 394, 711, 996, 646, 315, $\underline{045}$, 820, 350, 873, 628, 399, 686, 422, 586, 947, 181, 397, 695, 288, $\underline{045}$, 852, 359, 772, 270, $\underline{009}$, 780, 525, $\underline{042}$, $\underline{099}$, 167, 756, 971, 347, 626, 679, 330, $\underline{021}$, 529, 475, 291, $\underline{056}$, $\underline{089}$.

Первые пять чисел из этой последовательности отбрасываем, так как они больше 105. Первым попавшим в выборку будет число 014, далее исключаем 32 числа, в выборку попадает число 012. Следующие семь чисел снова исключаем. В выборку попадает число 078 (в приведенном перечне числа, не превосходящие 105, подчеркнуты). При этом надо исключить числа 078, 071, 045, так как они уже встречались ранее, и число 000, так как оно меньше 1. Процедура завершается, когда получено 15 чисел. Итак, выбираем первые 15 чисел:

$$14, 12, 78, 9, 20, 80, 71, 50, 90, 45, 42, 99, 21, 56, 89. \blacksquare$$

Другой способ получения случайной выборки можно осуществить с помощью компьютерной программы, предназначенной для работы с электронными таблицами. Идея заключается в том, чтобы перемешать элементы генеральной совокупности случайным образом и затем отобрать в выборку необходимое количество элементов.

В одном столбце располагают числа от 1 до N. Следующий столбец с помощью генератора случайных чисел заполняют равномерно распределенными случайными числами из интервала от 0 до 1 таким образом, чтобы эти случайные числа находились рядом с числами первого столбца. На следующем шаге оба столбца сортируют так, чтобы упорядочить числа во втором столбце. В результате все элементы генеральной совокупности будут перемешаны случайным образом. Осталось взять первые n элементов из этой перемешанной генеральной совокупности.

Пример 12. Построим с помощью программы Excel случайную выборку объемом $n = 5$ из генеральной совокупности объемом $N = 20$.

❐ В одном столбце расположим числа от 1 до 20, в верхнюю ячейку столбца случайных чисел введем формулу (=СЛЧИС()) и нажмем клавишу <ENTER>; затем скопируем результат вниз по столбцу, чтобы получить столбец случайных чисел. После выделения обоих столбцов (с номерами элементов и со случайными числами), выполняем команду (Данные ⇒ Сортировка) из меню, чтобы выполнить сортировку строк, используя значения из столбца со случайными числами. После этого числа в первом столбце будут упорядочены случайным образом, и для получения искомой случайной выборки достаточно взять первые пять номеров элементов. Ниже представлены окончательные результаты выполненной работы. В данном примере в выборку попали элементы с номерами 19, 2, 16, 11, 12:

19	0,015999
2	0,045645
16	0,106258
11	0,124938
12	0,227005
14	0,264088
4	0,300515
20	0,337087
7	0,358024
8	0,363328

9	0,454697
3	0,494867
6	0,569611
5	0,573705
13	0,624993
1	0,743619
17	0,771466
18	0,782283
10	0,786673
15	0,821596

Полученная таким образом случайная выборка обладает теми же свойствами, что и выборка, построенная с использованием таблицы случайных чисел. ■

Случайная выборка — не единственный метод построения выборки из генеральной совокупности. Существуют и другие методы, каждый из которых имеет свои преимущества и недостатки.

Часто генеральная совокупность определенным образом разбита на *слои* (группы). Например, при опросе родителей учащихся одной из школ генеральная совокупность — родители всех учащихся — может быть разбита на две группы: родители, которые могут профессионально судить о работе школы, и родители, которые такой профессиональной подготовкой не обладают. Покупателей магазина одежды — генеральную совокупность — можно разделить на тех, кто покупает обычную одежду, и тех, кто покупает модную одежду.

В этом случае часто используют так называемую ***стратифицированную случайную выборку***. Ее получают путем осуществления случайной выборки в каждом слое генеральной совокупности. Если генеральная совокупность однородна внутри каждого слоя, но отдельные слои заметно отличаются друг от друга, стратификация может увеличить точность статистического анализа. Размеры выборок для каждого из слоев могут быть различными. Не обязательно отбирать одинаковое число элементов из каждого слоя или планировать размер выборки для слоя пропорционально его объему в генеральной совокупности. Это позволяет определять размеры выборок для слоев исходя из затрат и ресурсов. Для одних слоев процесс отбора может быть сложнее и дороже чем для других, и для этих слоев можно использовать меньшие по объему выборки. Одни слои могут иметь больший разброс по сравнению с другими, и для них необходимо использовать бóльшие по объему выборки.

Иногда стратифицированную выборку называют ***серийной***.

Систематическую выборку строят путем целенаправленного отбора данных в генеральной совокупности, элементы которой пронумерованы по какому-то признаку. Выбирают одну случайную точку, а затем производят отбор элементов генеральной совокупности с некоторым постоянным шагом.

Если необходимо построить систематическую выборку объемом n из генеральной совокупности объемом N, то интервал между выбираемыми элементами будет равен наибольшему целому числу, не превосходящему $\frac{N}{n}$. В качестве начальной точки можно выбрать число между 1 и $\frac{N}{n}$. Систематическая выборка имеет ограниченное применение.

Иногда систематическую выборку называют ***механической***.

Фактически выше не было дано строгого определения выборки. Мы сформировали представление о выборке, достаточное для практического использования. На естественнонаучном языке тот факт, что некоторые числа образуют выборку, означает, что все они являются результатом измерения одной и той же величины, причем эти измерения проводились в неизменных условиях и независимо друг от друга. На математическом языке имеет место следующее определение:

Выборкой объемом n *из некоторого распределения называют совокупность взаимно независимых случайных величин* X_1, X_2, ..., X_n, *каждая из которых имеет это распределение.*

Никакого противоречия с предыдущими определениями нет. Пусть, например, измеряют продолжительность работы n приборов, испытываемых на стенде. Каждый прибор снабжен номером от 1 до n. Время безотказной работы прибора является случайной величиной. X_1 — это время безотказной работы прибора № 1, это случайная величина. В ходе испытания она принимает определенное значение, обозначим его через x_1. Аналогично X_2 — это время безотказной работы прибора № 2, x_2 — это значение, которое принимает случайная величина X_2 в результате испытания и т. д. Числа x_1, x_2, ..., x_n называют ***реализацией*** выборки X_1, X_2, ..., X_n.

В дальнейшем при рассмотрении задач математической статистики мы будем оперировать с числами — реализацией выборки.

Контрольные вопросы

1. На фабрике по производству поршней возникли определенные проблемы с качеством. Для проведения тщательного анализа принято решение собрать информацию о продукции, выпущенной в некоторый определенный день. Для каждого из указанных ниже методов извлечения выборки определите, является ли такая процедура хорошей, приемлемой или необоснованной. Ответы аргументируйте.

а) Первые пять произведенных поршней.

б) 10 поршней, находящихся вне фабрики, так как они никогда не использовались.

в) Каждый 15-й произведенный поршень.

г) Образованная в конце дня случайная выборка.

д) Все явно бракованные поршни вместе со случайной выборкой из очевидно стандартных поршней.

2. Какая из приведенных ниже выборок будет наиболее репрезентативной для совокупности всех зарегистрированных избирателей России?

а) Случайная выборка из 1000 избирателей г. Челябинска.

б) Случайная выборка из 1000 студентов Петербургского университета.

в) Выборка из 1000 человек, образованная на основе случайных телефонных номеров.

г) Выборка из 1000 человек — сторонников определенной политической партии.

3. Была ли выборка репрезентативной, если при изучении времени, которое затрачивают на выполнение уроков десятиклассники:

а) опрашивали только девочек;

б) опрос проводили только по средам;

в) опрашивали только учащихся лицеев;

г) опрашивали только неуспевающих?

4. В ходе опроса предстоит выяснить отношение жителей региона к введению Единого государственного экзамена. Какие категории жителей должны быть включены, на ваш взгляд, в составляемую выборку?

5. Для определения числа собак, больных чумкой, из всех бездомных собак города образовали две выборки:

а) одна собачья стая;

б) по несколько случайно отловленных собак из каждого района города.

Какую из них можно считать репрезентативной?

Задачи

505. Проводят исследование по оценке числа детей в семьях, проживающих в некотором микрорайоне, расположенном недалеко от школы. С этой целью опрашивают детей из разных классов о том, сколько у них братьев и сестер. Укажите ошибки, допущенные при формировании выборки, а также направление смещения, полученного в результате опроса (завышение или занижение данных о количестве детей в семье).

506.° Постройте случайную выборку без возвращения объемом 3 из следующей совокупности фирм: A, B, C, D, E, F, G, H. Используйте следующую последовательность случайных цифр: 55105972709475963962282 11.

507.° Постройте случайную выборку без возвращения объемом 4 из следующей совокупности фирм: A, B, C, D, E, F, G, H, K, L, M, N, P, Q, R, S. Используйте таблицу случайных чисел (табл. П.2), начните с 7-й строки и 2-го столбца.

508.° Из генеральной совокупности, состоящей из 85 учащихся школы, постройте случайную выборку из шести учащихся. Используйте таблицу случайных чисел, начните с 10-й строки и 4-го столбца.

509. Постройте случайную выборку объемом 5 из генеральной совокупности объемом 15, используя программу Excel.

510. Для исследования числа слов в одном предложении русского текста надо выбрать 10 абзацев из 100. Какие абзацы целесообразно выбрать?

§ 5.2. Оценивание параметров

Одной из основных задач математической статистики является оценивание неизвестных параметров генеральной совокупности. Фактически с этой задачей мы сталкивались в главах 1, 2 и 4, когда оценивали число изделий, которое необходимо дополнительно выпустить предприятию для замены вышедших из строя; когда оценивали число рыб в закрытом водоеме; траты населения города на молочные продукты и т. д. Изучая случайные величины, мы фактически в качестве оценки (приближенного значения) математического ожидания указывали выборочное среднее, а в качестве оценки дисперсии случайной величины — выборочную дисперсию. Вспомните приближенные равенства: $MX \approx \bar{x}$, $DX \approx s^2$.

В этом параграфе задача оценивания параметров будет рассмотрена в общей постановке.

Параметр генеральной совокупности (или просто ***параметр***) — это показатель, вычисленный для всей генеральной совокупности. В качестве примера можно привести математическое ожидание, дисперсию или среднее квадратичное отклонение случайной величины (генеральной совокупности). Параметр является фиксированным, но неизвестным числом.

На основе данных выборки вычисляют некоторый показатель, который называют ***параметром выборки***, или ***выборочным параметром***, или ***статистикой***. В качестве примера можно привести выборочное среднее, медиану, выборочную дисперсию, стандартное отклонение выборки, перцентили. Статистика является случайной величиной, ее значение — известное число.

Часто существует естественное соответствие между статистиками и параметрами. Для каждого параметра совокупности (показателя, значение которого хотелось бы знать, но которое точно неизвестно) существует выборочная статистика, рассчитанная на основе данных, представляющих наилучшую доступную информацию о неизвестном параметре. Такую выборочную статистику называют *оценочной функцией* параметра генеральной совокупности, а ее фактическое значение, рассчитанное по данным выборки, называют *оценкой* параметра совокупности.

Сущность задачи оценивания неизвестных параметров рассмотрим на следующем примере.

Пример 1. Пусть требуется оценить неизвестное число рыб, пригодных для вылавливания, в закрытом водоеме. Отловим некоторое число рыб, пометим их и выпустим обратно в водоем. Затем поймаем еще некоторое число рыб, подсчитаем число меченых среди них и с помощью этих данных оценим общее число рыб в пруду, пригодных для вылавливания.

❒ Здесь генеральной совокупностью является множество всех рыб в пруду, пригодных для вылавливания, параметром — общее число таких рыб в пруду, выборкой — рыбы, пойманные во второй раз. На основе данных выборки можно вычислить параметр выборки — долю меченых рыб среди отловленных. Эта величина случайная, в конкретном опыте она принимает определенное значение. Важным здесь является способ извлечения репрезентативной случайной выборки: неприменимы ни таблицы случайных чисел, ни компьютерные программы, так как мы не в состоянии перенумеровать рыб в пруду. Повысить репрезентативность выборки можно тем, что второй отлов рыб будет проводиться той же сетью, что и в первый раз, причем она будет заброшена не сразу после возвращения рыб в водоем (иначе меченые рыбы не успеют перемешаться с остальными) и не через продолжительное время (иначе может произойти нерест рыб или сброс вредных отходов в пруд: и то и другое существенно повлияет на общее число рыб в водоеме), а, например, на следующий день, в том же месте.

В рассматриваемом примере доля меченых рыб среди отловленных является оценочной функцией для доли меченых рыб среди всех рыб водоема. Если в первый раз отловлено 40 рыб и все они помечены и выпущены в пруд, а во второй раз поймано 30 рыб и среди них две оказались мечеными, то оценкой для доли меченых рыб в пруду является число $\frac{2}{30}$. Она позволяет оценить число всех рыб в пруду: оно приближенно равно 600. ■

Вообще, выборочное среднее является оценочной функцией для среднего генеральной совокупности, или для математического ожидания случайной величины. А в конкретном случае оно может равняться, например числу 5,3. Это и есть оценка параметра.

Оценочной функцией *для параметра называют произвольную функцию от элементов выборки. Ее значение от реализации выборки называют* ***оценкой параметра****.*

Оценочная функция — случайная величина, ее оценка — число, равное значению этой случайной величины.

В этом определении речь идет о функциях нескольких переменных. С такими функциями вы встречались. Например, левые части уравнений плоскости $ax + by + cz + d = 0$ и сферы $x^2 + y^2 + z^2 = R^2$ являются функциями трех переменных.

Пример 2. Пусть $X_1, X_2, \ldots, X_n$ — выборка из некоторого распределения. Тогда выборочное среднее $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$, выборочная дисперсия $S^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$, выборочное среднее квадратичное отклонение $S = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}$ являются функциями от $X_1, X_2, \ldots, X_n$ и поэтому являются оценочными функциями для параметров рассматриваемого распределения. Их значения

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n},\ s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2,\ s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

подсчитанные по значениям $x_1, x_2, \ldots, x_n$ элементов выборки, т. е. по реализации выборки, являются оценками этих параметров. Каждый элемент выборки X_i, как функция от $X_1, X_2, \ldots, X_n$, также является оценочной функцией для параметра, а его значение x_i — оценкой. ■

Для одного и того же параметра можно построить различные оценочные функции, а значит, и получить различные оценки. Например, чтобы оценить математическое ожидание случайной величины, можно в качестве оценочных функций взять либо среднее арифметическое элементов выборки (выборочное среднее), либо полуразность между наибольшим и наименьшим наблюдениями, либо какую-либо другую функцию от результатов наблюдений.

Каким критерием пользоваться при выборе оценочной функции? Ясно, что ответ зависит от поставленных целей. На практике на ответ влияют и экономические факторы. Например, одна оценка может оказаться лучше другой, но, чтобы получить первую, нужно затратить 500 000 р., а затраты на получение второй оценки составляют 5000 р. Если разница, существующая между оценками, не столь существенна для исследователя, то естественно, что он остановит свой выбор на второй оценке.

Рассмотрим общие подходы к выбору оценки параметра.

Ошибкой оценочной функции (оценки) *называют разность между оценочной функцией (оценкой) и параметром генеральной совокупности*. Ошибка оценки, как правило, неизвестна. Ошибка оценочной функции — случайная величина, ошибка оценки — число.

? Почему ошибка оценки, как правило, неизвестна?

Желательно, чтобы оценка не была систематически слишком завышенной или слишком заниженной по сравнению с соответствующим параметром генеральной совокупности, т. е. чтобы ее смещение от истинного значения параметра было незначительным.

Оценочную функцию некоторого параметра генеральной совокупности называют ***несмещенной****, если ее математическое ожидание равно этому параметру. В противном случае ее называют* ***смещенной****.*

Пример 3. Пусть $X_1, X_2, \ldots, X_n$ — выборка из некоторой генеральной совокупности. Покажем, что выборочное среднее $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$ является несмещенной оценочной функцией математического ожидания генеральной совокупности a.

❐ Действительно, тот факт, что $X_1, X_2, \ldots, X_n$ образуют выборку из генеральной совокупности, означает, что случайные величины $X_1, X_2, \ldots, X_n$ независимы в совокупности, и их числовые характеристики равны соответствующим числовым характеристикам генеральной совокупности, в частности $MX_i = a$, $i = 1, 2, \ldots, n$. Используя свойства математического ожидания случайной величины (см. § 4.3), получим

$$M\overline{X} = M\frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}(MX_1 + MX_2 + \ldots + MX_n) =$$

$$= \frac{1}{n}\underbrace{(a + a + \ldots + a)}_{n \text{ раз}} = a.$$

Оказалось, что математическое ожидание $\overline{X}$ равно математическому ожиданию a генеральной совокупности, поэтому $\overline{X}$ является несмещенной оценкой математического ожидания. ■

Несмещенность $\overline{X}$ означает, что для любого конкретного набора данных (значений $X_1, X_2, ..., X_n$) значение $\overline{X}$ обычно будет больше или меньше среднего генеральной совокупности a, но если многократно производить измерения, т. е. повторять процесс извлечения выборки и для каждой выборки вычислять значение $\overline{X}$, то полученные результаты будут в среднем близки к a и, следовательно, ошибка оценки не будет систематически слишком отличаться по модулю от нуля. Другими словами, несмещенность оценки означает, что оценка не содержит систематической ошибки. Значение $\overline{X}$ почти наверняка не равно точно a, оно содержит в себе некоторую неизвестную ошибку $\overline{X} - a$. В среднем величина этой ошибки равна нулю.

В главе 1 на с. 98 говорилось, что иногда дисперсию статистической совокупности определяют с помощью формулы $s_x^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$. Объясним причину различных определений этого понятия.

Дело в том, что выборочная дисперсия $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ является смещенной оценочной функцией дисперсии случайной величины: $MS^2 = \frac{n}{n-1}\sigma^2$, где $\sigma^2 = DX$. Математическое ожидание $\frac{n}{n-1}\sigma^2$ оценочной функции не равно оцениваемому параметру σ^2. Для устранения смещения рассматривают оценочную функцию $\hat{S}^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. Так как $\hat{S}^2 = \frac{n}{n-1}S^2$, то $M\hat{S}^2 = \frac{n}{n-1}MS^2 = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2$. Таким образом, $\hat{S}^2$ является несмещенной оценочной функцией дисперсии, и поэтому дисперсию статистической совокупности иногда определяют вышеприведенной формулой.

➤

Докажем равенство $MS^2 = \frac{n}{n-1}\sigma^2$. Пусть $MX_i = a$, $DX_i = \sigma^2$, тогда $M\overline{X} = a$, $D\overline{X} = \frac{\sigma^2}{n}$. Преобразуем выражение для выборочной дисперсии:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a + a - \overline{X})^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 - \frac{2}{n}\sum_{i=1}^{n}(X_i - a)(\overline{X} - a) + \frac{1}{n}\sum_{i=1}^{n}(\overline{X} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 - \frac{2}{n}(\overline{X} - a)\sum_{i=1}^{n}(X_i - a) + \frac{1}{n}\cdot n(\overline{X} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 - \frac{2}{n}(\overline{X} - a)(n\overline{X} - na) + (\overline{X} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 - (\overline{X} - a)^2.$$

Применяя свойства математического ожидания, получим

$$MS^2 = M\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 - (\overline{X} - a)\right)^2 =$$

$$= \frac{1}{n}nM(X_i - a)^2 - M(\overline{X} - a)^2 = DX_i - D\overline{X} =$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2. \quad ◄$$

Другое желательное свойство для оценочных функций (оценок) некоторого параметра состоит в том, чтобы с ростом числа наблюдений она приближалась к истинному значению параметра.

Оценочную функцию некоторого параметра генеральной совокупности называют ***состоятельной****, если с вероятностью, сколь угодно близкой к 1, при достаточно большом n разность между ней и параметром по модулю можно сделать меньше любого наперед заданного положительного числа* ε*. Оценочную функцию, не обладающую этим свойством, называют* ***несостоятельной****.*

❐ Покажем, что выборочное среднее является состоятельной оценкой математического ожидания. Из теоремы Чебышёва (см. § 4.9, теорема 1) следует, что $P\left(\left|\overline{X} - a\right| \leqslant \varepsilon\right) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}$, $\sigma^2 = Dx_i$, и если число n велико, то правая часть последнего неравенства близка к 1, т. е. при достаточно большом n с вероятностью, близкой к 1, выборочное среднее как угодно ма́ло отличается от их математического ожидания. ■

Как отмечалось во введении к этой главе, основные задачи математической статистики будут рассмотрены для биномиального распределения.

Напомним, что через p ранее (см. § 4.4) обозначалась вероятность «успеха» в каждом испытании Бернулли. Проводят n испытаний Бернулли, пусть m — число «успехов» в них. Тогда в качестве оценочной функции параметра p генеральной совокупности можно принять относительную частоту «успеха» $\bar{p} = \frac{m}{n}$. Напомним, что оценочной функцией для вероятности p является случайная величина $\bar{p} = \frac{m}{n}$, а оценкой неизвестной вероятности — ее значение при некотором значении m. Например, если шахматист А выиграл у шахматиста В четыре партии из 10, то вероятность выигрыша партии шахматистом А у В можно оценить отношением $\frac{4}{10} = 0{,}4$. Это оценка. А оценочной функцией является относительная частота. В приведенном примере она приняла значение 0,4, в другой раз шахматист А смог выиграть три партии из 10 и тогда оценкой неизвестной вероятности служит значение 0,3. А оценочная функция одна и та же — это относительная частота партий, выигранных этим шахматистом.

? Укажите, что в этом примере играет роль генеральной совокупности, что представляет собой извлеченная выборка, что является параметром генеральной совокупности, статистикой, оценкой параметра.

Рассмотрим свойства оценочной функции $\bar{p} = \frac{m}{n}$, в частности исследуем ее на несмещенность и состоятельность.

Если X_i означает число «успехов» в i-м испытании Бернулли, $i = 1, 2, \ldots, n$ (т. е. X_i принимает значение 1, если в i-м испытании наступил «успех», и 0 — в противном случае), то случайная величина $m = X_1 + X_2 + \ldots + X_n$ имеет биномиальное распределение с параметрами n и p, поэтому $Mm = np$ (см. § 4.7). Отсюда

$$M\bar{p} = M\frac{m}{n} = \frac{1}{n}Mm = \frac{1}{n}np = p,$$

т. е. $\bar{p} = \frac{m}{n}$ — несмещенная оценочная функция вероятности p.

Из теоремы Бернулли (см. § 4.9) следует, что $P\left(\left|\frac{m}{n} - p\right| \leqslant \varepsilon\right) \to 1$ при $n \to \infty$, т. е. $\bar{p} = \frac{m}{n}$ — состоятельная оценочная функция вероятности p.

В то же время оценочная функция X_1 для вероятности p не является состоятельной, так как она не зависит от n и поэтому с ростом n не может приближаться к p. Заметим, что эта оценочная функция означает, что если в первом испытании произошел «успех», то вероятность p оценивается числом 1, в противном случае — числом 0.

? Докажите, что оценочная функция X_1 является несмещенной для вероятности p.

Тот факт, что оценочная функция параметра является несмещенной, еще не гарантирует, что она оказывается достаточно часто близкой к оцениваемому параметру. Для иллюстрации рассмотрим 100 испытаний Бернулли. В качестве оценочной функции для вероятности p выберем X_1. Она является несмещенной (см. вопрос). Эта оценочная функция вряд ли может удовлетворить исследователя. Он скорее согласился бы использовать такие оценочные функции, как $\frac{m}{99}$, где m — число «успехов» в 100 испытаниях Бернулли, или $\frac{m}{101}$, хотя они и не являются несмещенными:

$$M\left(\frac{m}{99}\right)=\frac{100p}{99};\quad M\left(\frac{m}{101}\right)=\frac{100p}{101}.$$

Ясно, что необходимо какое-нибудь представление о точности оценочной функции. Из нескольких оценочных функций нужно отдать предпочтение той, которая в каком-либо смысле ближе к истинному значению оцениваемого параметра.

На практике нет возможности работать непосредственно с распределением статистики, выбранной в качестве оценочной функции, ибо его параметры определяются свойствами всей генеральной совокупности, а информация имеется только о выборке. Каждое распределение характеризуется средним квадратичным отклонением, поэтому выборочное распределение любой статистики также имеет среднее квадратичное отклонение. Если оно известно, то будет приближенно известно и то, насколько значение статистики отличается от среднего значения статистики. Точное значение среднего квадратичного отклонения неизвестно, оно зависит от генеральной совокупности. Поэтому используют информацию о выборке, чтобы сделать предположение или оценить среднее квадратичное отклонение выборочного распределения статистики. Полученное таким образом приближение среднего квадратичного отклонения статистики, основанное только на данных выборки, называют ***выборочным средним квадратичным отклонением***, или ***стандартной ошибкой статистики***. Стандартная ошибка показывает приближенно, насколько наблюдаемое значение статистики отличается от ее среднего значения. Более точно, стандартная ошибка приближенно показывает, каким будет стандартное отклонение в том случае, если взять большое число выборок, определить для каждой из этих выборок среднее и рассмотреть эти выборочные средние как набор данных.

? Верно ли, что стандартная ошибка статистики является оценочной функцией для среднего квадратичного отклонения этой статистики?

Вернемся к рассмотрению биномиального распределения. Вычислим дисперсию и среднее квадратичное отклонение оценочной функции $\bar{p} = \frac{m}{n}$ неизвестной вероятности p. Используя свойства дисперсии случайной величины (см. § 4.5, 4.6, 4.7), получим:

$$Dm = np(1-p); \quad D\bar{p} = D\frac{m}{n} = \frac{1}{n^2}Dm = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

Таким образом,

$$\sigma(m) = \sqrt{np(1-p)}, \quad \sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

Полученный результат показывает, что, хотя стандартная ошибка частоты m возрастает с ростом n, стандартная ошибка относительной частоты $\bar{p} = \frac{m}{n}$ при возрастании n убывает.

Недостаток полученных формул состоит в том, что вероятность p, как правило, неизвестна. На практике произведение $p(1-p)$ заменяют дробью $\frac{1}{4}$ на том основании, что $p(1-p) \leqslant \frac{1}{4}$ или заменяют p его оценочной функцией $\bar{p}$ и тогда выборочные средние квадратичные отклонения или стандартные ошибки, вычисленные по выборке, принимают следующий вид:

$$S_m = \sqrt{n\bar{p}(1-\bar{p})}, \quad S_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}.$$

Пример 4. При проверке 50 изделий шесть оказались бракованными. Вычислить выборочные средние квадратичные отклонения для числа бракованных изделий и их доли среди всех изделий.

❒ В данном примере $n = 50$, $m = 6$, $\bar{p} = 0{,}12$. Поэтому $S_m = \sqrt{50 \cdot 0{,}6 \cdot (1 - 0{,}6)} \approx 3{,}5$; $S_{\bar{p}} = \sqrt{\frac{0{,}6 \cdot (1 - 0{,}6)}{50}} \approx 0{,}069$. Полученный результат означает, что наблюдаемое значение 6 на 3,5 изделия отклоняется от неизвестного значения бракованных изделий в генеральной совокупности. Наблюдаемая доля бракованных изделий, равная 12%, примерно на 6,9% отличается от истинной неизвестной процентной доли бракованных изделий во всей генеральной совокупности. ■

? Пользуясь выделением полного квадрата, докажите, что $p(1 - p) \leqslant \frac{1}{4}$.

Пример 5. Служба контроля изучает 1000 деталей с тем, чтобы выявить в них наличие или отсутствие брака. В качестве оценочной функции вероятности p того, что деталь бракована, принимается относительная частота $\bar{p}$. На сколько могут отличаться p и $\bar{p}$?

❒ Применим неравенство Чебышёва (см. § 4.8) к случайной величине $\bar{p}$, среднее значение которой равно p, а среднее квадратичное отклонение $\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}}$. Получим

$$P\left(|p - \bar{p}| \leqslant h\sigma(\bar{p})\right) \geqslant 1 - \frac{D\bar{p}}{h^2\sigma^2(\bar{p})} = 1 - \frac{1}{h^2}.$$

Трудность заключается в том, что $\sigma(\bar{p})$ неизвестно, так как неизвестно p. Воспользуемся неравенством $p(1 - p) \leqslant \frac{1}{4}$. Тогда

$$\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}} \leqslant \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}.$$

Предыдущее неравенство можно переписать в виде:

$$P\left(|p - \bar{p}| \leqslant \frac{h}{2\sqrt{n}}\right) \geqslant 1 - \frac{1}{h^2}.$$

Если $h = 2$, то $P\left(|p - \bar{p}| \leqslant \frac{1}{\sqrt{n}}\right) \geqslant 1 - \frac{1}{4} = 0{,}75$. Если $h = 3$, то $P\left(|p - \bar{p}| \leqslant \frac{3}{2\sqrt{n}}\right) \geqslant 1 - \frac{1}{9} = \frac{8}{9}$. В рассматриваемом примере $n = 1000$. Поэтому при $h = 2$ можно утверждать, что вероятность выполнимости неравенства $|p - \bar{p}| \leqslant \frac{1}{\sqrt{n}}$, или неравенства $|p - \bar{p}| \leqslant \frac{1}{\sqrt{1000}}$, или неравенства $|p - \bar{p}| \leqslant 0{,}032$ не меньше $0{,}75$. Другими словами со степенью уверенности (или с надежностью) не менее 75% можно утверждать, что различие между p и $\bar{p}$ не превосходит $0{,}032$. ■

? Пользуясь методом, примененном при решении примера 3, найдите, на сколько могут отличаться p и $\bar{p}$, если $n = 1000$, $m = 30$. Сравните полученный результат с результатом решения примера 5. В чем принципиальное различие?

Фактически решение этого примера базируется на законе больших чисел (§ 4.9), который утверждает, что для любого положительного ε при $n \to \infty$ вероятность выполнения неравенства $|p - \bar{p}| \leqslant \varepsilon$ стремится к 1. Поэтому использованный метод применим для выборок большого объема. То же относится и к следующему примеру.

Пример 6. Для оценивания неизвестной вероятности p используют относительную частоту $\bar{p} = \frac{m}{n}$, причем требуется, чтобы вероятность выполнения неравенства $|p - \bar{p}| \leqslant 0{,}05$ была не меньше $0{,}85$. Каким должен быть объем выборки?

❒ Используя неравенство $P\left(|p - \bar{p}| \leqslant \frac{h}{2\sqrt{n}}\right) \geqslant 1 - \frac{1}{h^2}$, получим:

$$\frac{h}{2\sqrt{n}} = 0{,}05; \quad h = 0{,}1\sqrt{n}; \quad 1 - \frac{1}{h^2} = 1 - \frac{1}{0{,}01n} = 0{,}85.$$

Решая полученное уравнение, будем иметь $n = \frac{1}{0{,}15 \cdot 0{,}01} \approx 667$.

Итак, выборка должна содержать не менее 667 наблюдений. ■

➤

ЗАМЕЧАНИЕ. Использование неравенства Чебышёва приводит к грубым оценкам вероятности, завышенным оценкам для n. Дело в том, что неравенство Чебышёва справедливо для любой случайной величины, а всем, изучающим математику, занимающимся ею, известно, чем шире условия для какого-то утверждения, тем слабее результаты. Есть другие методы, которые приводят к более точным результатам. Но здесь мы практически не будем на них останавливаться. В частности, в данном примере эти методы дают значительно меньшее значение $n \geqslant 218$. ◄

Для выборок небольшого объема с помощью программы Excel можно решать задачи о нахождении вероятности того, что вероятность события в каждом испытании Бернулли отклоняется от относительной частоты этого события не более чем на заданную величину. Впрочем, Excel позволяет решать эту задачу и для выборок большого объема.

Пример 7. Вычислить вероятности:

а) $P\Big(\,|p-\bar{p}|\leqslant 0{,}1\Big)$, $P\Big(\,|p-\bar{p}|\leqslant 0{,}2\Big)$, если $n = 10$, $p = 0{,}2$;

б) $P\Big(\,|p-\bar{p}|\leqslant 0{,}03\Big)$, если $n = 1000$, $p = 0{,}1$.

❒ а) Преобразуем искомую вероятность:

$$P(|p-\bar{p}|\leqslant 0{,}1) = P\left(\left|p-\frac{m}{n}\right|\leqslant 0{,}1\right) = P(|np-m|\leqslant 0{,}1n) =$$

$$= P(|m-2|\leqslant 1) = P(1\leqslant m\leqslant 3) = P(m = 1) + P(m = 2) + P(m = 3).$$

Каждое слагаемое можно вычислить с помощью программы Excel, введя формулу (=БИНОМРАСП(a; n; π; ЛОЖЬ)) (см. § 4.2), а затем просуммировать полученные результаты с помощью автосуммы Σ. В результате получим $P(|p-\bar{p}|\leqslant 0{,}1) = 0{,}7718$. Аналогично

$$P(|p-\bar{p}|\leqslant 0{,}2) = P\left(\left|p-\frac{m}{n}\right|\leqslant 0{,}2\right) =$$

$$= P(|np-m|\leqslant 0{,}2n) = P(|m-2|\leqslant 2) = P(0\leqslant m\leqslant 4).$$

Для вычисления этой вероятности с помощью Excel можно воспользоваться формулой (=БИНОМРАСП(a; n; π; ИСТИНА)). Получим $P(|p-\bar{p}|\leqslant 0{,}2) = 0{,}9672$.

б) $P(|p-\bar{p}|\leqslant 0{,}03) = P\left(\left|p-\frac{m}{n}\right|\leqslant 0{,}03\right) = P(|np-m|\leqslant 0{,}03n) =$

$$= P(|m-100|\leqslant 30) = P(-70\leqslant m\leqslant 130) = P(0\leqslant m\leqslant 130).$$

Воспользовавшись опять формулой (=БИНОМРАСП(a; n; π; ИСТИНА)) при $a = 130$, $n = 1000$, $\pi = 0{,}1$, получим $P(|p-\bar{p}|\leqslant 0{,}03) = 0{,}9990$.

Заметим, что $P(-70\leqslant m\leqslant 130) = P(0\leqslant m\leqslant 130)$, так как случайная величина m принимает только неотрицательные значения. ■

? Почему при решении первой части примера 7, а удобнее было воспользоваться формулой (=БИНОМРАСП(a; n; π; ЛОЖЬ)), а при решении второй — формулой (=БИНОМРАСП(a; n; π; ИСТИНА))?

➤

Обратите внимание на то, что формулы $S_m = \sqrt{n\bar{p}(1-\bar{p})}$, $S_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ получены из формул $\sigma(m) = \sqrt{np(1-p)}$, $\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}}$ заменой теоретических характеристик σ и p соответствующими выборочными характеристиками S и $\bar{p}$. Точно так же выражения для выборочного среднего $\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i$ и выборочного среднего квадратичного отклонения $S = \sqrt{\frac{1}{n}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i}$ получаются из выражений для математического ожидания $MX = \sum_{i=1}^{m} x_i p_i$ и среднего квадратичного отклонения $\sigma(X) = \sqrt{M(X-MX)^2} = \sqrt{\sum_{i=1}^{m}(x_i - MX)^2 p_i}$ заменой вероятности p_i на относительную частоту $\frac{n_i}{n}$. Выражения для S_m и $S_{\bar{p}}$, впрочем, как и произвольные функции от элементов выборки, являются оценочными функциями соответственно для $\sigma(m)$, $\sigma(\bar{p})$. ◄

Пусть имеется теперь генеральная совокупность, состоящая из k слоев соответственно объемами $N_1, N_2, ..., N_k$, из которой извлечена стратифицированная выборка с размерами слоев $n_1, n_2, ..., n_k$. Пусть $N = N_1 + N_2 + ... + N_k$ — объем генеральной совокупности. Через $\bar{X}_1, \bar{X}_2, ..., \bar{X}_k$ обозначим выборочные средние слоев, а через $\bar{S}_1, \bar{S}_2, ..., \bar{S}_k$ — их выборочные средние квадратичные отклонения. Выборочное среднее и его стандартная ошибка для стратифицированной ошибки вычисляются по следующим формулам:

$$\bar{X} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + ... + N_k\bar{X}_k}{N_1 + N_2 + ... + N_k} = \frac{\sum_{i=1}^{k} N_i\bar{X}_i}{N},$$

$$S_{\bar{X}} = \frac{1}{N}\sqrt{\frac{N_1^2\bar{S}_1^2}{n_1} + \frac{N_2^2\bar{S}_2^2}{n_2} + ... \frac{N_k^2\bar{S}_k^2}{n_k}} = \frac{1}{N}\sqrt{\sum_{i=1}^{k}\frac{N_i^2\bar{S}_i^2}{n_i}}.$$

Пример 8. Проводят опрос среди 1400 родителей учащихся школы, которым предложено оценить по нескольким параметрам работу школы. Всех родителей естественно разделить на две группы в зависимости от их возможностей профессионально судить о работе школы. Решено использовать стратифицированную случайную выборку. Это разумно, так как можно ожидать, что группа «профессионалов» качественнее оценит работу школы. Стратификация позволит уменьшить общую вариацию возможных оценок работы школы. Основа выборки — это уже классифицированный список 1400 родителей: в списке 253 «профессионала» и 1147 «непрофессионалов». Решено отобрать 100 «профессионалов» и 50 «непрофессионалов», сделав акцент на слое подготовленных родителей. В таблице 5.1 приведены результаты.

Таблица 5.1

Слой	Объем совокупности	Объем выборки	Выборочное среднее	Стандартное отклонение выборки
«Профессионалы»	$N_1 = 253$	$n_1 = 100$	$\overline{x_1} = 125$	$\overline{s_1} = 45$
«Непрофессионалы»	$N_2 = 1147$	$n_1 = 50$	$\overline{x_2} = 29$	$\overline{s_2} = 8$

Полученные для двух слоев оценки интересны. Эти результаты подтвердили предположение о том, что «профессионалы» выше оценят работу школы. Вычислим среднюю оценку работы школы, данную одним родителем:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{253 \cdot 125 + 1147 \cdot 29}{253 + 1147} = \frac{64\,888}{1400} = 46.$$

Значение результирующей средней оценки, 46, ближе к оценке «непрофессионалов» (29), чем к оценке «профессионалов» (125). Это следствие того, что слой «непрофессионалов» составляет значительно большую часть генеральной совокупности, чем слой «профессионалов».

Неопределенность полученной оценки можно определить вычислением стандартной ошибки:

$$S_{\bar{x}} = \frac{1}{N}\sqrt{\frac{N_1^2 s_1^2}{n_1} + \frac{N_2^2 s_2^2}{n_2}} = \frac{1}{1400}\sqrt{\frac{253^2 \cdot 45^2}{100} + \frac{1147^2 \cdot 8^2}{50}} =$$

$$= \frac{1}{1400}\sqrt{84\,198\,976 + 129\,618\,225} \approx 10{,}4.$$

Можно ожидать, что оценка, полученная от опроса всех родителей, будет отличаться в ту или иную сторону примерно на 10. ■

? Придумайте ситуацию, когда удобно воспользоваться стратифицированной выборкой.

Контрольные вопросы

1. В чем сущность задачи оценивания неизвестных параметров генеральной совокупности?
2. Стандартное отклонение выборки равно 8,5. Является ли это число оценочной функцией или оценкой среднего квадратичного отклонения генеральной совокупности?
3. Известна ли оценка ошибки, если оценивается неизвестный параметр генеральной совокупности?
4. Каким образом стандартная ошибка характеризует качество информации, полученной в результате оценивания?
5. Как изменится стандартная ошибка при увеличении объема выборки n?
6. Как вы понимаете утверждение: «Относительная частота события в испытаниях Бернулли является несмещенной оценочной функцией вероятности наступления этого события в каждом испытании»?
7. Как вы понимаете утверждение: «Относительная частота события в испытаниях Бернулли является состоятельной оценочной функкией вероятности наступления этого события в каждом испытании»?
8. В чем преимущества стратификации?
9. Какие дополнительные соображения, кроме требования несмещенности, влияют на выбор метода оценивания?

Задачи

511.° Найдите:

а) $P(|p-\bar{p}| \leqslant 0{,}1)$, если $n = 20$, $p = 0{,}01$;

б) $P(|p-\bar{p}| \leqslant 0{,}2)$, если $n = 5$, $p = 0{,}2$.

512.° С помощью неравенства Чебышёва оцените величину $P(|p-\bar{p}| \leqslant d)$, если:

а) $d = 0{,}1$; $n = 1000$; б) $d = 0{,}05$; $n = 100$;
в) $d = 0{,}2$; $n = 10$; г) $d = 0{,}01$; $n = 400$.

513. Каким должен быть объем выборки, чтобы вероятность выполнения неравенства $P(|p-\bar{p}| \leqslant d)$ была не меньшей 0,92, если:

а) $d = 0{,}1$; б) $d = 0{,}05$; в) $d = 0{,}2$?

514. На сколько могут отличаться вероятность p и ее оценочная функция $\bar{p}$, чтобы при объеме выборки $n = 500$ это можно было гарантировать с вероятностью, не меньшей 0,88?

515.° Обозначим неизвестное число овец в отаре через N. Из этой отары наугад отбирают M овец, которых затем клеймят и возвращают в отару. В следующий раз отбирают n овец, среди которых m оказываются клеймеными. Предложите оценочную функцию величины N.

516. С какой вероятностью можно утверждать, что для 100 опытов вероятность события отличается от относительной частоты этого события по модулю не больше, чем на 0,1?

517. Опрос 756 случайно отобранных взрослых жителей Москвы показал, что 63% из них поддерживают текущую политику правительства Москвы. Найдите показатель, который приближенно характеризовал бы отличие этого выборочного процента от значения, которое могло бы быть получено при опросе всего взрослого населения Москвы.

518. В тексте 20 000 слов. Отобрали n слов, среди них было m глаголов.

а) Оцените вероятность того, что вероятность появления глагола в тексте отличается от $\frac{m}{n}$ по модулю не более чем на 0,05, если $n = 900$.

б) Сколько слов следует отобрать из текста, чтобы с вероятностью не менее 0,8 ожидать, что отклонение относительной частоты от вероятности появления глагола по модулю не будет превосходить 0,01?

519. В некотором городе живет 25 тыс. взрослых жителей. Социологи отобрали для опроса n взрослых жителей. Оказалось, что m из них читает газету А.

а) Оцените вероятность того, что доля всех жителей города, читающих газету А, отличается от доли читающих эту газету в выборке не более чем на 0,02, если $n = 2000$.

б) Сколько жителей надо отобрать для опроса, чтобы с вероятностью не менее 0,83 процент читающих газету А в выборке отличался от процента таких взрослых жителей в городе не более чем на 0,01?

§ 5.3. Доверительные интервалы

В предыдущем параграфе были введены понятия оценочной функции и оценки параметра генеральной совокупности, обсуждались некоторые желательные свойства оценочных функций такие, как несмещенность и состоятельность. Рассмотренные оценочные функции часто называют ***точечными оценками*** (от слова «точка»; параметр оценивается некоторой точкой). Оценочная функция есть случайная величина, имеющая некоторый разброс около истинного значения параметра, а поэтому, приравнивая истинное значение параметра числовому значению оценочной функции или оценке, мы допускаем определенную ошибку. Другими словами, построить оценку неизвестного параметра по результатам наблюдений — значит, найти приближенное значение этого неизвестного параметра. С приближенными вычислениями и понятием приближения вы знакомы. Но, говоря о приближениях или пользуясь приближенными значениями, надо ясно представлять себе и точность приближения, представлять границы абсолютной погрешности. Например, 1 м может считаться приближенным значением и для длины 910 мм, и для 1007 мм, и для 999,3 мм. Границы абсолютной погрешности составят соответственно 100 мм, 10 мм и 1 мм. Без указания, с какой точностью взяты приближенные значения, сами по себе они практически смысла не имеют.

Эта общая идея находит свое применение и в статистике. Здесь понятие точности приближения реализуется в виде ***доверительного интервала***.

В этом параграфе будет рассмотрен вопрос получения ***интервальных оценок***, т. е. возможность построения некоторого интервала, содержащего истинное значение параметра с заданной вероятностью. Сам метод получения таких интервалов сводится к построению так называемых ***доверительных пределов***, а полученные интервалы называют ***доверительными***. Процедура, которая будет здесь описана, состоит в построении верхнего и нижнего доверительных пределов уровня $1 - \alpha$, обладающих следующим свойством: если говорят, что истинное значение параметра лежит между этими пределами, то это утверждение верно с вероятностью $1 - \alpha$ (и неверно с вероятностью α). Очевидно, что α следует выбирать достаточно малым. Число $1 - \alpha$ называют ***коэффициентом доверия***, или ***доверительной вероятностью***. Величина $1 - \alpha$ отражает «степень готовности мириться с возможностью ошибки».

С идеей построения доверительных интервалов познакомимся прежде всего на примере. Пусть проведено n измерений некоторой неизвестной величины a с помощью прибора, точность которого известна. Получены значения $x_1, x_2, \ldots, x_n$. Будем считать, что измерения проведены без систематической ошибки, так что ошибки измерения только случайные. Наблюдения $x_1, x_2, \ldots, x_n$ будем считать реализацией (значениями) независимых случайных величин, имеющих одно и то же распределение. Их математическое ожидание a неизвестно, и его необходимо оценить. Качество измерений характеризуется дисперсией случайных ошибок наблюдений: чем больше дисперсия, тем менее стабильны результаты. В значительной степени дисперсия результатов измерений определяется качеством прибора. Так как качество прибора мы предположили известным, то можем считать известной дисперсию наблюдений.

Основываясь на законе больших чисел, в качестве оценочной функции математического ожидания a можно принять среднее арифметическое значений наблюдений, т. е. $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

Задача состоит в нахождении двух случайных величин $\overline{X} - k(n)$ и $\overline{X} + k(n)$, зависящих от $\overline{X}$ и n, таких, что интервал $(\overline{X} - k(n); \overline{X} + k(n))$ с вероятностью $1 - \alpha$ содержит истинное значение параметра. Вся трудность состоит в нахождении $k(n)$.

Для аналогии напомним, что в приближенных вычислениях говорят, что число x является приближением к числу a с точностью до h, если $x - h \leqslant a \leqslant x + h$. Отличие статистических приближений состоит в том, что гарантируется эта точность лишь с определенной вероятностью $1 - \alpha$, меньшей 1. Другими словами, при массовых вычислениях подобного рода точность приближения к параметру a среднего арифметическего n наблюдений $\bar{x}$ будет не более чем $k(n)$ примерно в $100 \cdot (1 - \alpha)\%$ случаев. В оставшейся доле случаев приближение $\bar{x}$ к a может оказаться худшим.

Воспользуемся неравенством Чебышёва для построения доверительного интервала для неизвестного математического ожидания a случайной величины X. Согласно неравенству Чебышёва

$$P(|\overline{X} - a| \leqslant \varepsilon) > 1 - \frac{D\overline{X}}{\varepsilon^2}.$$

Так как $D\overline{X} = \dfrac{DX}{n} = \dfrac{\sigma^2}{n}$ (σ^2 — дисперсия одного наблюдения), то неравенство примет вид

$$P(|\overline{X} - a| \leqslant \varepsilon) > 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Решая неравенство, стоящее под знаком вероятности относительно a, получим $\overline{X} - \varepsilon \leqslant a \leqslant \overline{X} + \varepsilon$. Если дисперсия была бы известна, то по заданной доверительной вероятности $1 - \alpha$ из равенства $1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \alpha$ можно было бы найти ε ($\varepsilon = \frac{\sigma}{\sqrt{n\alpha}}$) и построить искомый доверительный интервал:

$$\overline{X} - \frac{\sigma}{\sqrt{n\alpha}} \leqslant a \leqslant \overline{X} + \frac{\sigma}{\sqrt{n\alpha}}.$$

Вероятность того, что этот доверительный интервал содержит истинное значение a больше $1 - \alpha$. Если же дисперсия неизвестна, то в предыдущем неравенстве ее заменяют выборочной дисперсией S и получают приближенный доверительный интервал

$$\overline{X} - \frac{S}{\sqrt{n\alpha}} \leqslant a \leqslant \overline{X} + \frac{S}{\sqrt{n\alpha}}.$$

Пример 1. В школах района учится 500 десятиклассников. Для проверки усвоения некоторой темы был предложен тест из 10 вопросов. Для испытания отобраны 62 человека. Результаты испытаний представлены в таблице 5.2.

Таблица 5.2

Число правильных ответов	3	4	5	6	7	8	9	10
Число учащихся	2	18	13	8	10	6	4	1

Найти границы, в которых с вероятностью не менее 0,84 лежит среднее число правильных ответов всех десятиклассников.

Дисперсию генеральной совокупности принять равной дисперсии выборки.

❐ По данным выборки вычислим выборочное среднее и выборочную дисперсию: $\bar{x} = 5{,}73$; $s = 1{,}81$. По условию $\alpha = 1 - 0{,}84 = 0{,}16$. Тогда $\frac{s}{\sqrt{n\alpha}} = \frac{1{,}81}{\sqrt{62 \cdot 0{,}16}} = 0{,}58$. Доверительный интервал будет иметь вид $(5{,}73 - 0{,}58;\ 5{,}73 + 0{,}58)$, или $(5{,}15;\ 6{,}31)$. В таких пределах содержится среднее число правильных ответов десятиклассников района на задания теста с доверительной вероятностью 0,84. ■

Подробно мы рассмотрим построение доверительного интервала для неизвестной вероятности p наступления события в каждом из n испытаний Бернулли. Требуется при заданном числе испытаний

n и заданном числе m наступлений события («успеха») в n испытаниях Бернулли найти две такие случайные величины $p_1(n)$ и $p_2(n)$, чтобы с заданной вероятностью (например, 0,9) можно было утверждать, что имеет место неравенство

$$p_1(n) \leqslant p \leqslant p_2(n).$$

При некоторых исходах эксперимента (n испытаний Бернулли) это неравенство окажется справедливым, при других — нет. При большом числе экспериментов истинными эти неравенства окажутся примерно в 90% случаев.

Напомним, что несмещенной и состоятельной оценочной функцией неизвестной вероятности p является относительная частота $\bar{p} = \frac{m}{n}$. Из неравенства Чебышёва следует, что

$$P(|\bar{p} - p| \leqslant h\sigma(\bar{p})) \geqslant 1 - \frac{1}{h^2}.$$

Отсюда нетрудно получить, что

$$P(\bar{p} - h\sigma(\bar{p}) \leqslant p \leqslant \bar{p} + h\sigma(\bar{p})) \geqslant 1 - \frac{1}{h^2}.$$

Казалось бы, что доверительный интервал построен. Но значения его пределов $p - h\sigma(\bar{p})$, $p + h\sigma(\bar{p})$ не могут быть вычислены по результатам испытаний, так как неизвестно среднее квадратичное отклонение $\sigma(\bar{p})$ относительной частоты $\bar{p} = \frac{m}{n}$. Напомним, что $\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}}$. Вероятность p не известна. На практике произведение $p(1-p)$ заменяют дробью $\frac{1}{4}$ на том основании, что $p(1-p) \leqslant \frac{1}{4}$. В этом случае получим

$$\sigma(\bar{p}) = \sqrt{D\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \leqslant \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}.$$

Итак, с вероятностью не меньшей, чем $1 - \frac{1}{h^2}$ выполняется неравенство $|p - \bar{p}| \leqslant \frac{h}{2\sqrt{n}}$, или $\bar{p} - \frac{h}{2\sqrt{n}} \leqslant p \leqslant \bar{p} + \frac{h}{2\sqrt{n}}$. Словами эти неравенства можно сформулировать так: расстояние между p и $\bar{p}$ не превосходит $\frac{h}{2\sqrt{n}}$.

Если $h = 2$, то вероятность выполнения неравенства $|p - \bar{p}| \leqslant \dfrac{1}{\sqrt{n}}$ не меньше 0,75. Аналогично, положив $h = 3$, получим, что вероятность выполнения неравенства $|p - \bar{p}| \leqslant \dfrac{3}{2\sqrt{n}}$ не меньше 0,88.

Итак, *интервал* $\left(\bar{p} - \dfrac{h}{2\sqrt{n}};\ \bar{p} + \dfrac{h}{2\sqrt{n}}\right)$ *является* ***доверительным интервалом*** *для вероятности* p *с доверительной вероятностью* $1 - \dfrac{1}{h^2}$. *Число* $\dfrac{h}{2\sqrt{n}}$ *называют* ***точностью оценки*** $\bar{p}$, *а число* $1 - \dfrac{1}{h^2}$ — ***надежностью этой оценки***, *число* $2 \cdot \dfrac{h}{2\sqrt{n}} = \dfrac{h}{\sqrt{n}}$ — ***длиной доверительного интервала***.

Пример 2. Завод производит электрические лампочки. Вероятность лампочке быть бракованной считается равной p. Для контроля отобрано n ламп, среди которых бракованных оказалось m. Требуется с доверительной вероятностью 0,83 оценить вероятность того, что лампа бракована, если $n = 1000$, $m = 120$.

❒ Относительная частота события «лампа бракована» равна $\bar{p} = \dfrac{m}{n} = \dfrac{120}{1000} = 0{,}12$. По условию доверительная вероятность не меньше, чем 0,83, т. е. $1 - \dfrac{1}{h^2} = 0{,}83$. Решая это уравнение относительно h, получим: $\dfrac{1}{h^2} = 0{,}17$, $\dfrac{1}{h} \approx 0{,}41$, $h \approx 2{,}44$. Тогда точность оценки будет равна $\dfrac{h}{2\sqrt{n}} \approx \dfrac{2{,}44}{2\sqrt{1000}} \approx 0{,}04$. Итак, $(0{,}12 - 0{,}04;\ 0{,}12 + 0{,}04)$, или $(0{,}08;\ 0{,}16)$, — искомый доверительный интервал для неизвестной вероятности p с доверительной вероятностью 0,83. ■

?

Чему равна длина построенного доверительного интервала?

Чему равна точность оценки $\bar{p} = 0{,}12$ неизвестной вероятности p?

Чему равна надежность оценки $\bar{p} = 0{,}12$ неизвестной вероятности p?

Напоминаем, что примерно в 83% случаев доверительный интервал, вычисленный по правилу $\left(\bar{p} - \frac{h}{2\sqrt{n}};\ \bar{p} + \frac{h}{2\sqrt{n}}\right)$, будет содержать при приведенных данных истинное значение вероятности.

В предыдущем параграфе мы уже говорили о том, что применение неравенства Чебышёва приводит к довольно грубым результатам. Обратите внимание на то, что в рассмотренном интервале доверительная вероятность невысокая — 0,83, т. е. мы согласны идти на 17%-й риск. Платой за более высокую доверительную вероятность является более широкий, а значит, и менее полезный интервал. Например, в примере 2 для доверительной вероятности, равной 0,95, мы получили бы доверительный интервал (0,05; 0,19), а для доверительной вероятности, равной 0,99, мы получили бы доверительный интервал (−0,02; 0,28), т. е. фактически односторонний интервал: значение вероятности лампочке быть бракованной с вероятностью, не меньшей 0,99, не превышает 0,28.

? Какой из двух доверительных интервалов больше: 90%-ный или 85%-ный?

Более точный доверительный интервал для неизвестной вероятности события можно получить при достаточно большом объеме выборки, если воспользоваться нормальным приближением к биномиальному распределению (см. § 4.10). По заданной доверительной вероятности $1 - \alpha$ с помощью таблиц стандартного нормального распределения можно найти такое число h, что $2P(Z < h) - 1 = 1 - \alpha$.

Пример 3. Решим пример 2, воспользовавшись нормальным приближением для биномиального распределения, если доверительная вероятность равна:

а) 0,83; б) 0,95.

❐ а) Если $2P(Z < h) - 1 = 0{,}83$, то $h = 1{,}37$, тогда точность оценки будет равна $\frac{h}{2\sqrt{n}} \approx \frac{1{,}37}{2\sqrt{1000}} \approx 0{,}022$. Итак, $(0{,}12 - 0{,}022;\ 0{,}12 + 0{,}022)$, или (0,098; 0,142), — искомый доверительный интервал с доверительной вероятностью 0,83. Как видим, получен более узкий, по сравнению с результатом применения неравенства Чебышёва, доверительный интервал.

б) Если $2P(Z < h) - 1 = 0{,}95$, то $h = 1{,}96$, тогда точность оценки будет равна $\frac{h}{2\sqrt{n}} \approx \frac{1{,}96}{2\sqrt{1000}} \approx 0{,}031$. Итак, $(0{,}12 - 0{,}031; 0{,}12 + 0{,}031)$, или $(0{,}089; 0{,}151)$, — искомый доверительный интервал с доверительной вероятностью $0{,}95$. ■

➤

В некоторых ситуациях нет необходимости утверждать, что вероятность события находится между двумя границами доверительного интервала. Доверительный интервал может строиться на основе утверждения о том, что вероятность события по крайней мере не меньше, чем некоторое число, или что вероятность события по крайней мере не больше, чем некоторое число. В этом случае мы имеем дело с так называемым односторонним доверительным интервалом. ***Односторонний доверительный интервал*** устанавливает с доверительной вероятностью, что параметр генеральной совокупности либо не меньше, либо не больше некоторого вычисленного значения.

В каких случаях нас может удовлетворить односторонний доверительный интервал? Например, известно, что изменение технологии может привести к уменьшению процента брака. Тогда нам достаточно знать, не меньше какого числа вероятность брака в генеральной совокупности. Другой пример. Известно, что социальная политика правительства привела к улучшению жизненного уровня населения страны. Исследуют долю людей, доход которых превышает заданный уровень. Нам достаточно знать, не больше какого числа эта доля в генеральной совокупности. ➤

Контрольные вопросы

1. Какую дополнительную информацию о генеральной совокупности дает доверительный интервал по сравнению с оценкой параметра?

2. Как вы понимаете фразу: «Доверительные интервалы позволяют, по возможности, избавиться от неопределенности и сделать как можно более точный вывод»?

3. Можно ли при построении доверительного интервала для неизвестной вероятности события при помощи неравенства Чебышёва полагать $h \leqslant 1$?

4. Что происходит с длиной доверительного интервала при увеличении объема выборки с фиксированной доверительной вероятностью?

5. Верно ли, что с увеличением доверительной вероятности при фиксированном объеме выборки длина доверительного интервала увеличивается?

6. Что происходит с доверительной вероятностью при уменьшении длины доверительного интервала, если объем выборки остается неизменным?

7. Как влияет на длину доверительного интервала оценка неизвестной вероятности, если объем выборки и доверительная вероятность остаются неизменными?

8. Представьте, что вы строите доверительный интервал для неизвестной вероятности выпуска бракованного парашюта. Удовлетворит ли вас доверительная вероятность, равная 0,99?

9. Верно ли, что 90%-ный доверительный интервал, построенный для неизвестной вероятности события, ровно в 10% случаев не будет содержать истинной вероятности события?

Задачи

520. Предварительный опрос 961 случайно отобранного избирателя показал, что кандидат, за которого вы проголосовали, идет впереди с 52,4% голосов. Постройте 90%-й доверительный интервал для процента голосов за вашего кандидата в генеральной совокупности.

521. В некотором большом городе образована выборка из 25 жителей. Оказалось, что 10 из них являются читателями газеты А. Постройте 75%-ный доверительный интервал для вероятности того, что каждый житель города читает газету А.

522. Из большого числа школьников образована выборка, состоящая из 100 человек. Оказалось, что для 80 из них способность запомнить изучаемый материал существенно повышается в том случае, если занятиям предшествует 8-часовой сон. Постройте 75%-ный доверительный интервал для оценки доли в выборке лиц, способность которых к запоминанию существенно повышается в результате 8-часового сна, предшествующего занятиям.

523. Каким должен быть объем выборки, приводящий к получению 84%-ного доверительного интервала для неизвестной вероятности события, длина которого не превышает 0,25?

524. Случайная выборка 100 семей, проживающих в Краснодаре, показала, что в 20 из этих семей месячный доход на одного члена семьи превышает 15 000 р. Постройте 91%-ный доверительный интервал для доли p семей в Краснодаре, чей месячный доход на одного члена семьи превышает 15 000 р.

525. Случайная выборка 100 семей, проживающих в Краснодаре, показала, что в 20 из этих семей месячный доход на одного члена семьи превышает 15 000 р. С какой вероятностью можно утверждать, что интервал $\left(\frac{1}{30}; \frac{11}{30}\right)$ является доверительным для доли p семей в Краснодаре, чей месячный доход на одного члена семьи превышает 15 000 р.?

526. Анализ выборки объемом 240 человек, случайно отобранных из населения города численностью 250 300 человек, показал, что 150 из них поддерживают программу дорожного строительства в городе. Постройте 85%-ный доверительный интервал для процента поддерживающих программу жителей всего города:

а) пользуясь неравенством Чебышёва;

б) используя нормальное приближение для биномиального распределения.

527. Выпущено 1000 радиоприемников.

Проводились испытания чувствительности 40 радиоприемников из этой партии. Результаты представлены в таблице 5.3.

Таблица 5.3

Чувствительность, мкВ	75—175	175—275	275—375	375—475	475—575	575—675
Число радиоприемников	1	14	14	8	2	1

По этим данным определите границы, в которых с вероятностью не менее 0,91 лежит средняя чувствительность радиоприемника во всей партии. Дисперсию генеральной совокупности примите равной дисперсии выборки.

§ 5.4. Проверка статистических гипотез

Человеку часто приходится принимать то или иное решение. В большинстве принимаемых решений содержится элемент риска. Решив заняться бизнесом, человек должен принять решение, каким бизнесом ему нужно заняться. Очень важно угадать, какой бизнес принесет прибыль, где меньше конкуренции. Учащийся, отвечая на вопрос тестового задания, к которому приведено несколько ответов, должен принять решение, на каком варианте остановиться. Он может быть не знаком с данным вопросом и поэтому остановится на том варианте, который ему покажется наиболее правдоподобным.

В рассмотренных примерах элемент случайности не велик, и при принятии решения значительную роль играет сознание человека. Однако во многих случаях статистика может существенно помочь обосновать принятие того или иного решения. Например, принятию решения о переходе на новую технологию производства какого-либо изделия должна предшествовать экспериментальная проверка этой технологии, сбор необходимой информации, ее обработка, проверка того, говорят ли собранные данные в пользу новой технологии. Аналогично перед введением нового учебника в школу, при изменении содержания обучения по какому-либо предмету должна проводиться апробация учебника, проверка необходимости и возможности изменения содержания обучения, при этом собранная информация должна обрабатываться с помощью методов математической статистики, в частности на основе этих методов должно приниматься решение, какой из выдвинутых гипотез отдать предпочтение.

В данном параграфе мы познакомимся с основными идеями, на которых базируется раздел статистики, называемый ***проверкой статистических гипотез***. Эти идеи будут применены в основном к проверке гипотез о вероятности «успеха» в каждом испытании Бернулли.

Проверка статистических гипотез позволяет на основе имеющейся информации сделать выбор между двумя предположениями, называемыми ***гипотезами***. Эта процедура дает ответ на вопрос, являются ли наблюдаемые результаты (об эффективности новой технологии, нового учебника, нового содержания обучения) случайными или они реальны. Проверку статистических гипотез можно рассматривать как один из компонентов принятия решений. Сама по себе проверка гипотез не может быть использована для принятия решения о покупке, например, мобильного телефона той или иной фирмы (ведь решение принимает человек), но она дает важную информацию об эффективности того или иного мобильного телефона.

Сущность идей, лежащих в основе проверки статистических гипотез, рассмотрим сначала на следующем примере. Кратко эти идеи рассматривались в § 4.8 при изложении правила трех сигм.

? Какие гипотезы проверялись в § 4.8 с помощью правила трех сигм?

Пример 1. Родители Миши живут раздельно. Он их одинаково любит (впрочем, так же, как и они его), и после школы ежедневно едет к одному из родителей. В выходные он остается у того из родителей, к которому приехал накануне выходных. Мишины родители проживают рядом со станциями метро, расположенными на одинаковом расстоянии от станции метро, находящейся рядом с Мишиной школой, но по разные стороны от этой станции. Придя на станцию метро, Миша садился в первый подошедший к станции поезд и, если этот поезд отправлялся в одном направлении, он приезжал к маме, если в другом — к отцу. В предположении, что вероятности уехать в каждом из этих двух направлений примерно одинаковы, можно ожидать, что число визитов Миши к маме и папе за достаточно длительный промежуток времени должно оказаться примерно одинаковым. Однако через некоторое время мама упрекнула Мишу в том, что он отдает предпочтение папе, так как за последние 30 рабочих дней он 21 раз ездил к папе и только 9 раз — к ней.

Какова же причина подобной несправедливости по отношению к Мишиной маме? С помощью каких аргументов Миша обосновывал решение доверить выбор направления своей поездки случаю? Он рассуждал так: за достаточно длительный промежуток времени число визитов к папе и маме должно оказаться примерно одинаковым, так как вероятность попасть на поезд, идущий в одном направлении, равна вероятности уехать на поезде, идущем в противоположном направлении. Таким образом, можно говорить об *отсутствии различий* в числе встреч с одним и другим родителем. Подобное предположение об «отсутствии различий» получило в статистике название «***нулевая гипотеза***». В данном примере нулевая гипотеза состоит в том, что Миша в течение 30 рабочих дней 15 раз поедет к маме и 15 раз — к папе. Точнее, нулевая гипотеза состоит в том, что вероятности p_1 и p_2 попасть на поезд каждого направления одинаковы, $p_1 = p_2$.

Но фактическое соотношение 21 : 9 поездок к папе и маме приводит к мысли отвергнуть данную гипотезу и считать предположения, из которых исходил Миша, неверными. А что было бы в случае, если соотношение между визитами составило 17 : 13? Возможно, этот результат покажется нам правдоподобным. А как поступить в случае, если это соотношение будет равняться 18 : 12, 19 : 11, 20 : 10? Случайно ли оно оказалось именно таким? По-видимому, где-то лежит граница, после которой уже нельзя сделать вывод о случайности полученного соотношения, и имеются какие-то скрытые факторы, в силу которых Миша чаще попадал на поезд, отвозивший его к папе.

Соотношение 21 : 9 могло произойти случайно и в случае $p_1 = p_2$. Но вероятность этого мала. Она совпадает с вероятностью того, что при 30 подбрасываниях симметричной монеты герб выпадет ровно 21 раз. Согласно формуле Бернулли (см. § 4.4), она равна $C_{30}^{21} \cdot \left(\frac{1}{2}\right)^{30} = C_{30}^{9} \cdot \left(\frac{1}{2}\right)^{30} \approx 0{,}013$. Это событие редкое, но все же оно может произойти и по этой причине не существует однозначно определенного положительного или отрицательного ответа для случаев, сходных с описанной ситуацией. ■

Итак, есть две возможности:

1) анализируемая ситуация не удовлетворяет первоначально выдвинутому предположению (нулевая гипотеза оказалась ошибочной), и полученный результат является правильным отражением реальной действительности;

2) анализируемая ситуация соответствует первоначально выдвинутому предположению (нулевая гипотеза оказалась верной), и произошла та редкая случайность, которая в принципе возможна, но практически не происходит.

Столкнувшись с подобной альтернативой, дальше можно рассуждать так: если $p_1 = p_2$, то такое соотношение между фактическими поездками, как 21 : 9, может встретиться лишь в одном или двух случаях из 100. Поскольку подобный результат является малоправдоподобным, видимо, есть смысл *отбросить основную гипотезу* и рассмотреть ее альтернативы. Гипотеза $p_1 = p_2$ могла оказаться неверной, так как участки пути в оба направления на самом деле могут иметь разную длину, или они имеют одинаковую длину, но на одном направлении больше остановок, возможно, в одном из направлений поезд должен по какой-то причине снижать скорость движения и т. д. Учитывая эти возможности и располагая фактическими данными, мы со статистической точки зрения имеем все основания полагать, что первоначально выдвинутая гипотеза $p_1 = p_2$ неверна и ее следует отбросить. Точнее данная нулевая гипотеза отвергается при определенном *уровне значимости*. Мы можем ошибиться, отвергая нулевую гипотезу, однако это может случиться лишь в одном или двух случаях из 100. Таким образом, нулевая гипотеза отвергается при 1%-ном или 2%-ном уровне значимости.

Сформулируйте нулевую гипотезу, которая проверялась в § 4.8 с помощью правила трех сигм?

Чему равен уровень значимости, на котором проверяется гипотеза с помощью правила «двух сигм»; правила трех сигм?

Следует понимать, что есть определенный риск и в том случае, когда нулевая гипотеза отвергается (она может оказаться верной), и в том случае, когда нулевая гипотеза не отвергается (она может оказаться неверной).

Если отвергается нулевая гипотеза в случае, когда она верна, то говорят, что совершается ***ошибка первого рода***.

Нулевая гипотеза верна. При проверке с помощью правила трех сигм она отвергнута. Чему равна вероятность ошибки первого рода?

Если же нулевая гипотеза не отвергнута в то время, когда она неверна, то совершается ***ошибка второго рода***.

Итак, на основании рассмотренного примера сформулируем ***обобщенную схему последовательности рассуждений***, которыми пользуются в статистике.

1. *Формулируют нулевую гипотезу.*

Понятие статистической гипотезы, с которым имеет дело математическая статистика значительно уже, чем общее понятие научной гипотезы. Научные гипотезы о происхождении Вселенной, о существовании жизни на Марсе, о наличии нефти на глубине 15 км под поверхностью Земли не являются статистическими, их нельзя подвергнуть проверке с помощью математической статистики. Статистические гипотезы касаются поведения наблюдаемых случайных величин или случайных событий. *Средний рост юношей за последние 20 лет увеличился на 10 см*; или *погрешности при изготовлении болтов на двух исследуемых станках одинаковы*; или *за данного кандидата проголосует более половины избирателей* — вот примеры статистических гипотез.

Нулевую гипотезу обозначают через H_0. Она представляет собой такое утверждение, которое принимается тогда, когда нет убедительных аргументов для его отклонения.

Альтернативную гипотезу обозначают через H_1. Ей отдают предпочтение только тогда, когда есть убедительное статистическое доказательство, которое отвергает приемлемость нулевой гипотезы.

2. *Получают фактические данные о событиях, относительно которых была сформулирована нулевая гипотеза.*

Эти данные получают в результате опыта. Проверку статистических гипотез осуществляют путем сопоставления случайных событий или случайных величин, о которых идет речь в гипотезе, с результатами наблюдений. Здесь как раз и используют выборочный метод. Например, по результатам контроля качества некоторой выборки фармацевтических ампул судят о качестве всех ампул.

3. *Определяют вероятность того, что полученный результат мог быть получен при условии, что нулевая гипотеза верна.*

Результаты наблюдений зависят от случая. Поэтому статистические гипотезы носят не категорический характер, а характер правдоподобного утверждения, которое имеет вполне определенную вероятность.

4. *Если вероятность получения данного результата при условии, что нулевая гипотеза верна, мала, нулевую гипотезу отвергают на уровне значимости, равном этой вероятности.*

Если вероятность получения данного результата мала, то он практически невозможен, т. е. в однократном эксперименте практически не может быть получен. В этом состоит сущность ***принципа практической уверенности***, на котором основана проверка статистических гипотез. Другими словами, соответствующее событие практически не должно произойти, если верна нулевая гипотеза. Но в конкретном эксперименте оно произошло. Следовательно, нулевая гипотеза неверна, и ее отвергают на определенном уровне значимости.

5. *Признают, что и в том случае, когда отвергают и когда не отвергают нулевую гипотезу, возможен определенный риск.*

Понять смысл ошибок первого и второго рода поможет рассмотрение ситуации с вынесением судебного приговора. В соответствии с презумпцией невиновности, подсудимый считается невиновным (первый этап, нулевая гипотеза) до тех пор, пока не будет доказано противное в результате показаний свидетелей, экспертиз, вещественных доказательств и т. д. (второй этап, получение фактических данных). На третьем и четвертом этапах не должно остаться места для сомнений. Если суд считает невиновного человека преступником, то совершается ошибка первого рода (нулевая гипотеза отвергается в то время, когда она верна). Если же суд признает преступника невиновным, то совершается ошибка второго рода (нулевая гипотеза не отвергается в то время, когда она неверна). Два других решения, которые могут быть приняты судом, являются верными и следовательно, справедливыми.

В таблице 5.4 представлены верные решения и типы ошибок при проверке статистических гипотез.

Таблица 5.4

Статистическое решение	Фактическая оценка нулевой гипотезы	
	Верна	**Неверна**
Не отвергать нулевую гипотезу	Правильное решение	Ошибка второго рода
Отвергнуть нулевую гипотезу	Ошибка первого рода	Правильное решение

Процедура проверки статистических гипотез в определенной степени напоминает проведение обоснования математических утверждений: чтобы опровергнуть утверждение, достаточно привести один пример, подтверждающий, что оно не имеет места (ср.: наступление события в однократном эксперименте); однако даже десятки примеров, подтверждающих справедливость утверждения, не гарантируют его правильность, они говорят лишь о том, что пока не получены факты, опровергающие его. Поэтому в статистике в том случае, когда событие, характеризующее исследуемую ситуацию, имеет достаточно большую вероятность, если нулевая гипотеза верна, то не говорят, что нулевая гипотеза принимается, а говорят осторожнее: нулевая гипотеза не отвергается, так как полученные данные не противоречат ей. Ведь пока мы только знаем, что нулевая гипотеза не отвергнута по результатам конкретного эксперимента на данном уровне значимости. Но она может быть отвергнута по результатам других экспериментов или на другом уровне значимости.

Пример 2. Рассматривают вопрос о влиянии нового рациона кормления коров на жирность молока. Исследовали 20 коров.

а) Указать генеральную совокупность и выборку в данной ситуации.

б) Сформулировать нулевую и альтернативную гипотезы.

в) Привести пример, когда нулевая гипотеза будет отвергнута.

г) Охарактеризовать ошибки первого и второго рода.

д) Построить таблицу, аналогичную таблице 5.4.

❐ а) Генеральную совокупность составляют все коровы фермы, на которой проводится исследование, выборку — 20 исследованных коров.

б) Нулевая гипотеза — средняя жирность молока остается неизменной (если бы были данные о жирности молока до введения нового рациона, нулевую гипотезу можно было бы сформулировать так: средняя жирность молока равна известной величине); альтернативная гипотеза — средняя жирность молока изменяется (если мы уверены в том, что от введения нового рациона кормления жирность молока может только увеличиться, то альтернативную гипотезу можно было бы сформулировать так: средняя жирность молока больше известной величины).

в) Определяют среднюю жирность молока у 20 коров, попавших в выборку. Вычисляют вероятность того, что жирность молока равна полученному по выборке значению при условии, что верна нулевая гипотеза. Для вычисления этой вероятности нужно знать закон распределения случайной величины — жирности молока. Если эта вероятность окажется малой, равной, например, 0,05, то нулевую гипотезу можно отвергнуть.

г) Ошибка первого рода: нулевая гипотеза верна, но она отвергается, т. е. на самом деле жирность молока не изменилась, но жирность молока, полученная по выборке, значительно отличается от прежней.

Ошибка второго рода: нулевая гипотеза неверна, но она не отвергается, т. е. на самом деле жирность молока изменилась, но жирность молока, полученная по выборке, незначительно отличается от прежней.

д)

Таблица 5.5

Принятое решение	**Истинное положение**	
	Жирность молока не изменилась	**Жирность молока изменилась**
Жирность молока не изменилась	*Правильное решение*	*Ошибка второго рода* Принято решение о неэффективности нового рациона кормления — возможен возврат к прежнему рациону
Жирность молока изменилась	*Ошибка первого рода* Принято решение об эффективности нового рациона кормления — будет повсеместно введен новый рацион, но он не оправдает надежд	*Правильное решение*

■

Пример 3. На учете 2000 больных, страдающих от хронических приступов головной боли. Вероятность того, что некоторое лекарство снимает боль, равна p. Проверяют нулевую гипотезу: $p = 0{,}8$. Для проверки этой гипотезы отобрали $n = 100$ больных, после принятия лекарства у $m = 75$ из них приступ прошел.

а) Вычислить вероятность того, что при случайном отборе 100 больных 75 из них лекарство поможет.

б) Принять решение, касающееся нулевой гипотезы.

в) Указать уровень значимости, на котором нулевая гипотеза отвергается или не отвергается.

г) Вычислить вероятность ошибки первого рода.

❐ а) Обозначим через m число больных из 100, которым лекарство помогло. Нетрудно проверить, что случайная величина m имеет биномиальное распределение с параметрами 100 и $p = 0{,}8$, если верна нулевая гипотеза. Вычислим искомую вероятность $P(m = 75 \mid p = 0{,}8)$. Ввиду больших значений параметров ($n = 100$, $m = 75$), вычисление значения выражения $C_{100}^{75} \cdot (0{,}80)^{75} \cdot (1 - 0{,}8)^{100-75}$ громоздко, поэтому воспользуемся программой Excel. Получим $P(m = 75 \mid p = 0{,}8) = 0{,}044$.

б) Эта вероятность мала, т. е. событие «у 75 больных из 100 после принятия лекарства приступ прошел» является практически невозможным. Произошло событие, которое практически не должно было наступить, если верна нулевая гипотеза. Поэтому нулевая гипотеза отвергается.

в) Нулевая гипотеза отвергается на уровне значимости 0,044. Напоминаем, что уровень значимости — это вероятность отвергнуть нулевую гипотезу, если она верна.

г) Вероятность ошибки первого рода равна вероятности отвергнуть нулевую гипотезу, если она верна. В данном случае она равна 0,044. ■

Иногда используют другой ***порядок проверки статистических гипотез***. А именно:

1. Формулируют нулевую и альтернативную гипотезы H_0 и H_1.

2. Выбирают некоторое событие S (называемое критическим или критерием для гипотезы H_0).

3. Назначают уровень значимости, т. е. число α, равное вероятности наступления события S при условии, что верна нулевая гипотеза. Это число должно быть малым: если мы дорожим нулевой гипотезой и не хотим ее отвергнуть понапрасну, то α должно быть малым. Выбор α лежит вне математической статистики и вообще вне математики. Его выбирают исходя из того риска, на который согласен пойти исследователь, отвергнув верную гипотезу.

Каким условиям должно удовлетворять событие S? Хорошо было бы, чтобы вероятность наступления события S при условии, что верна нулевая гипотеза, равнялась 0, а вероятность наступления события S при условии, что верна альтернативная гипотеза, равнялась 1. Но это невозможно. Поэтому приходится допускать и ненулевые значения вероятности наступления события S при условии, что верна нулевая гипотеза.

4. Проводят опыт. Если в результате этого опыта происходит событие S, то гипотезу H_0 отвергают. Если событие S не происходит, то гипотезу на уровне значимости α не отвергают.

Основная сложность состоит в выборе события S. Например, если проверяют гипотезу о доле брака в некоторой партии деталей, то событие S может выглядеть так: «число бракованных деталей в партии из 100 деталей превышает 2».

Что представляет событие S при проверке гипотезы о вероятности некоторого события с помощью правила трех сигм?

В следующих параграфах мы покажем, как строится событие S при проверке гипотезы о значении среднего генеральной совокупности и о значении биномиальной вероятности.

Контрольные вопросы

1. В чем заключается цель проверки статистических гипотез?
2. Является ли статистическая гипотеза утверждением о генеральной совокупности или о выборке?
3. В чем различие между результатом проверки статистической гипотезы и утверждением о доверительном интервале?
4. В чем отличие роли нулевой гипотезы от роли альтернативной?
5. Верно ли, что уровень значимости совпадает с вероятностью ошибки второго рода?
6. Может ли вероятность ошибки первого рода равняться нулю?
7. Что показывает уровень значимости?
8. Верно ли, что если нулевая гипотеза не отвергается, то она верна?
9. Верно ли, что если нулевая гипотеза отвергается, то она неверна?

10. Почему в статистике никогда не говорят об отклонении альтернативной гипотезы?

11. Применение новой технологии производства некоторого продукта в течение нескольких дней привело к увеличению объема выпускаемой продукции. Является ли это достаточным основанием для внедрения новой технологии?

12. Какая из двух гипотез (нулевая или альтернативная) требует доказательства?

Задачи

528. Группа из 100 случайно отобранных людей посмотрела рекламу товара. После этого регистрируют число людей из этой группы, которые в течение следующей недели купили рекламируемый товар.

а)° Укажите генеральную совокупность и выборку в данной ситуации.

б)° Сформулируйте нулевую и альтернативную гипотезы.

в)° Приведите пример, когда нулевая гипотеза будет отвергнута.

г) Охарактеризуйте ошибки первого и второго рода.

д) Постройте таблицу, аналогичную таблице 5.4.

529. Отобрали группу учащихся, плохо написавших контрольную работу, и провели с ними дополнительные занятия по устранению пробелов, выявленных контрольной. Затем зарегистрировали число учащихся, показавших при повторном выполнении контрольной работы по той же теме лучший по сравнению с предыдущим результат.

а)° Укажите генеральную совокупность и выборку в данной ситуации.

б)° Сформулируйте нулевую и альтернативную гипотезы.

в)° Приведите пример, когда нулевая гипотеза будет отвергнута.

г) Охарактеризуйте ошибки первого и второго рода.

д) Постройте таблицу, аналогичную таблице 5.4.

530. Для контроля отобрали группу инъекционных ампул.

а) Укажите генеральную совокупность и выборку в данной ситуации.

б) Сформулируйте нулевую и альтернативную гипотезы.

в) Приведите пример, когда нулевая гипотеза будет отвергнута.

г) Охарактеризуйте ошибки первого и второго рода.

д) Постройте таблицу, аналогичную таблице 5.4.

531. Завод выпустил 10 000 электрических ламп с матовым покрытием. Вероятность лампе иметь дефект в покрытии считается равной p. Проверяют нулевую гипотезу: $p = 0{,}1$. Для проверки этой гипотезы отобрали $n = 500$ ламп, среди которых $m = 55$ имеют дефект в покрытии.

а) Вычислите вероятность того, что при случайном отборе 500 ламп у 55 из них будет дефект в покрытии.

б) Примите решение, касающееся нулевой гипотезы.

в) Укажите уровень значимости, на котором нулевая гипотеза отвергается или не отвергается.

г) Вычислите вероятность ошибки первого рода.

532. Перед началом футбольного матча одной из команд было высказано сомнение в симметричности монеты, с помощью которой судья определял, какая из команд начнет игру с центра поля. Для проверки этих сомнений монету подбросили 100 раз, при этом герб выпал 42 раза.

а) Вычислите вероятность того, что при 100 подбрасываниях монеты 42 раза выпадет герб.

б) Примите решение, касающееся симметричности монеты.

в) Укажите уровень значимости, на котором гипотеза отвергается или не отвергается.

г) Вычислите вероятность ошибки первого рода.

§ 5.5. Проверка гипотезы о равенстве среднего генеральной совокупности некоторому заданному значению

Один из самых простых случаев проверки статистической гипотезы заключается в проверке равенства между средним генеральной совокупности и некоторым заданным значением. Заданное значение представляет собой фиксированное число a_0, полученное не из выборочных данных, а, например, из предыдущих исследований. Нулевая гипотеза H_0: $a = a_0$ утверждает, что неизвестное среднее значение генеральной совокупности a в точности равно заданному значению a_0. Альтернативная гипотеза H_1: $a \neq a_0$ утверждает, что неизвестное среднее значение генеральной совокупности a не равно заданному значению a_0.

Фактически здесь фигурируют три различные величины, имеющие отношение к среднему:

1) a — неизвестное среднее генеральной совокупности, которое нас интересует;

2) a_0 — заданное значение, в отношении которого проверяют гипотезу;

3) $\overline{X}$ — известное выборочное среднее, которое используют для проверки нулевой гипотезы. Это случайная величина, являющаяся оценочной функцией a.

Проверка гипотезы заключается в сравнении двух известных величин $\overline{X}$ и a. Если эти значения отличаются больше, чем можно было бы ожидать, исходя их случайности, то нулевую гипотезу $a = a_0$ отклоняют, так как $\overline{X}$ дает информацию о неизвестном среднем a. Если значения $\overline{X}$ и a достаточно близки, то нулевую гипотезу $a = a_0$ не отвергают. Фактически сравниваются значение $\bar{x}$ случайной величины $\overline{X}$ и значение a_0 математического ожидания a. Что означает «близость»? Где находится необходимая граница? Близость $\overline{X}$ и a определяется средним квадратичным отклонением $\sigma(\overline{X})$ случайной величины $\overline{X}$ (если она известна) или ее стандартной ошибкой $S_{\overline{X}}$ (в противном случае). Для простоты будем предполагать, что $\sigma(\overline{X})$ известно. Таким образом, если $\overline{X}$ и a отстоят друг от друга на расстоянии достаточного числа средних квадратичных отклонений, то это является доказательством того, что $a \neq a_0$.

? Почему приходится предполагать, что $\sigma(\overline{X})$ известно?

Почему предположение о том, что $\sigma(\overline{X})$ известно, ограничивает применимость процедур проверки гипотез?

Существуют различные методы проверки рассматриваемой гипотезы. *1-й метод* использует доверительные интервалы, о которых речь шла в § 5.3.

2-й метод основан на статистике $\dfrac{\overline{X} - a}{\sigma(\overline{X})}$, имеющей при достаточно большом объеме выборки приближенно стандартное нормальное распределение (см. § 4.10), и на применении таблиц стандартного нормального распределения (см. Приложение, табл. П.1). Обратите внимание на то, что числитель этой статистики характеризует отклонение среднего генеральной совокупности от выборочного среднего, а знаменатель показывает, в каких единицах описывается это отклонение.

3-й метод, более грубый, метод проверки гипотезы $a = a_0$ основан на применении неравенства Чебышёва.

Разберем каждый из этих методов.

1-й метод. Рассмотрим проверку нулевой гипотезы H_0: $a = a_0$ против альтернативной H_1: $a \neq a_0$ на основе данных случайной выборки из генеральной совокупности. Для этого строят доверительный интервал для среднего a, исходя из значений $\overline{X}$ и $\sigma(\overline{X})$, с доверительной вероятностью $1 - \alpha$. Если значение a_0 находится вне этого доверительного интервала, то a_0 не может рассматриваться как допустимое значение среднего генеральной совокупности, и гипотезу H_0 следует отвергнуть на уровне значимости α в пользу альтернативной гипотезы. В противном случае нулевую гипотезу на уровне значимости α не отвергают.

Зачем в предложении «гипотезу H_0 следует отвергнуть» добавляют слова «на уровне значимости α»?

Покажем, как строится соответствующий доверительный интервал. Как отмечалось в § 4.10, закон распределения случайной величины, являющейся суммой очень большого числа независимых в совокупности случайных величин, какова бы ни была природа слагаемых, лишь бы каждое из них было мало по сравнению со всей суммой, должен быть близок к нормальному закону. На этом основании мы делаем вывод, что выборочное среднее $\overline{X}$ при достаточно большом объеме выборки имеет приближенно нормальное распределение. В том же параграфе указывалось, что если случайная величина X имеет нормальное распределение со средним a и с дисперсией σ^2, то случайная величина $Z = \frac{X - a}{\sigma}$ имеет стандартное нормальное распределение. Так как $M\overline{X} = a$, $M\overline{X} = \frac{\sigma^2}{n}$, где $\sigma^2 = DX_i$, то случайная величина $\frac{\overline{X} - a}{\sigma}\sqrt{n}$ имеет стандартное нормальное распределение.

Почему $M\overline{X} = a$, $D\overline{X} = \frac{\sigma^2}{n}$?

По таблице нормального распределения можно по заданной доверительной вероятности $1-\alpha$ найти такое число l, что $P\left(\left|\frac{\overline{X}-a}{\sigma}\sqrt{n}\right| \leqslant l\right) = 1-\alpha$. Решая неравенство, стоящее под знаком вероятности относительно a, получим искомый доверительный интервал для среднего генеральной совокупности с доверительной вероятностью $1-\alpha$:

$$P\left(\overline{X} - l\frac{\sigma}{\sqrt{n}} \leqslant a \leqslant \overline{X} + l\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Напомним, что по заданной вероятности $1-\alpha$ число l по таблицам нормального распределения находится из условия $2P(Z < l) - 1 = 1 - \alpha$, или $P(Z < l) = 1 - \frac{\alpha}{2}$. Здесь Z — случайная величина, имеющая стандартное нормальное распределение.

? Почему число l по таблицам нормального распределения находится из условия $P(Z < l) = 1 - \frac{\alpha}{2}$?

Поясним, почему метод доверительных интервалов работает при проверке гипотез. По определению доверительного интервала вероятность того, что доверительный интервал со случайными границами содержит a, близка к 1 и равна $1-\alpha$. Если нулевая гипотеза верна, т. е. $a = a_0$, то вероятность того, что доверительный интервал содержит a, также равна $1-\alpha$. Таким образом, если нулевая гипотеза верна, то принятое решение будет правильным примерно в $(1-\alpha)\cdot 100\%$ случаев и неверным примерно в $\alpha \cdot 100\%$ случаев.

? Верно ли, что если нулевая гипотеза $a = a_0$ верна, то событие «доверительный интервал не содержит a_0» является практически невозможным?

Изменяется ли доверительный интервал в зависимости от заданного значения среднего генеральной совокупности?

Пример 1. Настриг шерсти с одной овцы в некотором хозяйстве есть случайная величина со средним 5,2 кг и средним квадратичным отклонением 1,1 кг. В этом хозяйстве изменили условия содержания и кормления овец. Для проверки эффективности новой технологии содержания овец из отары случайным образом отобрали 200 овец. Средний настриг с одной овцы в этой выборке составил 5,3 кг. Дают ли эти данные достаточно оснований для внедрения новых условий содержания и кормления овец?

❒ В качестве нулевой гипотезы примем, что средний настриг шерсти с одной овцы остался неизменным, т. е. H_0: $a = 5{,}2$; в качестве альтернативной — H_1: $a \neq 5{,}2$. Нулевая гипотеза утверждает, что в новых условиях кормления и содержания овец средний настриг шерсти в отаре большого размера точно равен заданному значению 5,2 кг. Альтернативная гипотеза утверждает, что в новых условиях кормления и содержания овец средний настриг шерсти в отаре большого размера не равен заданному значению 5,2 кг. Построим доверительный интервал для среднего a генеральной совокупности с доверительной вероятностью $1 - \alpha = 0{,}95$, базируясь на данных выборки. Известно, что $\bar{x} = 5{,}3$; $n = 200$; $\sigma = 1{,}1$. Из условия $P(Z < l) = 1 - \frac{\alpha}{2} = 1 - 0{,}025 = 0{,}975$ по таблицам нормального распределения найдем l: $l = 1{,}96$. Доверительный интервал будет иметь вид $\left(5{,}3 - \frac{1{,}96 \cdot 1{,}1}{\sqrt{200}}; 5{,}3 + \frac{1{,}96 \cdot 1{,}1}{\sqrt{200}}\right)$, или (5,14; 5,46). Полученный результат означает, что мы на 95% уверены, что в новых условиях кормления и содержания овец средний настриг шерсти в отаре большого размера находится между 5,14 и 5,46 кг. Остается проверить, находится ли заданное значение 5,2 в пределах доверительного интервала. Это значение находится в интервале. Иными словами, утверждение $5{,}14 \leqslant 5{,}2 \leqslant 5{,}46$ является справедливым. Нулевая гипотеза на уровне значимости 0,05 не отвергается.

Средний настриг шерсти с одной овцы в новых условиях кормления и содержания овец, равный 5,3 кг, несущественно отличается от среднего настрига шерсти в отаре большого размера при старых условиях, он равен 5,2 кг. Мы не получили убедительного доказательства в пользу введения новой технологии кормления и содержания овец. Доказала ли проверка гипотезы неэффективность новой технологии? Нет. Технология может быть и эффективной. Но мы не получили убедительных доказательств ни ее эффективности, ни ее неэффективности. ■

? Нужно ли проверять, принадлежит ли доверительному интервалу выборочное среднее $\overline{X}$?

Что еще следует предпринять для решения проблемы об эффективности новой технологии содержания и кормления овец?

2-й метод. Другой метод проверки гипотезы о среднем генеральной совокупности состоит в том, что сначала вычисляют статистику $\frac{\overline{X}-a_0}{\sigma}\sqrt{n}$, а затем по таблицам нормального распределения по заданному уровню значимости α находят такое значение l_α, что $P(Z < l_\alpha) = \frac{\alpha}{2}$. Здесь Z — случайная величина, имеющая стандартное нормальное распределение. Далее модуль значения статистики $\frac{\overline{X}-a_0}{\sigma}\sqrt{n}$ сравнивают со значением l_α. Если $\left|\frac{\overline{X}-a_0}{\sigma}\sqrt{n}\right| > l_\alpha$, то гипотезу H_0: $a = a_0$ отвергают в пользу альтернативной гипотезы H_1: $a \neq a_0$, в противном случае ее не отвергают (на уровне значимости α).

? В каких единицах выражается отклонение среднего генеральной совокупности от выборочного среднего в статистике $\frac{\overline{X}-a_0}{\sigma}\sqrt{n}$?

Зависит ли значение статистики $\frac{\overline{X}-a_0}{\sigma}\sqrt{n}$ от заданного значения среднего генеральной совокупности?

Пример 2. Проверить гипотезу, сформулированную в примере 1, по данным этого примера рассмотренным методом.

❒ В примере 1: $\bar{x} = 5{,}3$; $n = 200$; $\sigma = 1{,}1$; $a_0 = 5{,}2$; значение $\left|\frac{\overline{X}-a_0}{\sigma}\sqrt{n}\right|$ равно 1,28; l_α при $\alpha = 0{,}05$ равно 1,96, $1{,}28 < 1{,}96$, следовательно, гипотеза H_0 на уровне значимости 0,05 не отвергается. Как и следовало ожидать, этот метод дает тот же результат, что и метод доверительных интервалов. ■

3-й метод. Рассмотрим теперь применение неравенства Чебышёва для проверки гипотезы о значении среднего генеральной совокупности при известной дисперсии.

Согласно неравенству Чебышёва

$$P\left(\left|\overline{X}-M\overline{X}\right| \leqslant h\sigma(\overline{X})\right) \geqslant 1 - \frac{D\overline{X}}{h^2\sigma^2(\overline{X})} = 1 - \frac{1}{h^2}.$$

Это неравенство эквивалентно неравенству

$$P\left(\left|\overline{X}-M\overline{X}\right| > h\sigma(\overline{X})\right) \leqslant \frac{1}{h^2}.$$

?

На каком основании утверждается, что неравенства $P\left(\left|\overline{X}-M\overline{X}\right| \leqslant h\sigma(\overline{X})\right) \geqslant 1-\frac{1}{h^2}$ и $P\left(\left|\overline{X}-M\overline{X}\right| > h\sigma(\overline{X})\right) \leqslant \frac{1}{h^2}$ эквивалентны?

Далее по заданному уровню значимости $\alpha = \frac{1}{h^2}$ вычисляют h, $h\sigma(\overline{X})$ и сравнивают значение $\left|\overline{X}-M\overline{X}\right| = \left|\overline{X}-a_0\right|$ с $h\sigma(\overline{X})$. Если окажется, что $\left|\overline{X}-a_0\right| > h\sigma(\overline{X})$, т. е. произошло практически невозможное событие, то нулевую гипотезу на уровне значимости α нужно отвергнуть. В противном случае она не отвергается.

?

Как, зная среднее квадратичное отклонение σ одного наблюдения, вычислить среднее квадратичное отклонение выборочного среднего $\sigma(\overline{X})$?

Пример 3. С помощью неравенства Чебышёва проверить гипотезу, сформулированную в примере 1, по данным этого примера, приняв в качестве уровня значимости $\alpha = 0{,}1$.

❒ Так как $\alpha = \frac{1}{h^2} = 0{,}1$, то $h = 3{,}2$; $h\sigma(\overline{X}) = 3{,}2 \cdot \frac{1{,}1}{\sqrt{200}} = 0{,}25$; $\left|\overline{X}-M\overline{X}\right| = 0{,}1$; $0{,}1 < 0{,}25$, нулевая гипотеза не отвергается. ■

Альтернативная гипотеза H_1: $a \neq a_0$, которую мы до сих пор рассматривали, была двусторонней, так как, согласно ей, среднее генеральной совокупности может быть как больше, так и меньше заданного значения a_0. Однако часто интерес представляет не проверка того, отличается ли среднее генеральной совокупности от заданного значения, а проверка более определенного вопроса: является ли среднее генеральной совокупности больше заданного значения или меньше его. Например, в рассмотренных примерах 1—3 интерес представлял вопрос, повышается ли средний настриг шерсти с одной овцы в новых условиях кормления и содержания овец, ведь исследователи были уверены в том, что новая технология не могла привести к уменьшению среднего настрига шерсти. Аналогично при введении новой технологии производства некоторого изделия исследователи уверены в том, что она не может привести к увеличению брака, и тогда разумно в качестве альтернативной гипотезы принять следующую: средний процент бракованных изделий меньше заданного значения (того, что был длительное время, до введения новой технологии, на этом производстве). Если подобная информация о среднем значении генеральной совокупности отсутствует, то следует воспользоваться двусторонней проверкой и выяснить, значимо ли $\overline{X}$ отличается от a_0. Двустороннюю проверку можно использовать и для того, чтобы утверждать, что среднее выборочное значение $\overline{X}$ значимо больше a_0 или значимо меньше a_0.

Однако часто выгодно применить одностороннюю проверку. Концентрируя внимание только на одной стороне и пренебрегая другой, одностороння проверка может лучше определить различие между средним и заданным значениями именно на этой стороне. При односторонней проверке нулевая гипотеза утверждает, что значение a находится по одну сторону от a_0, а альтернативная гипотеза утверждает, что значение a находится по другую сторону от a_0.

Гипотезы для двух видов односторонней проверки формулируются следующим образом.

Односторонняя проверка того, что а больше a_0

H_0: $a \geqslant a_0$

Нулевая гипотеза утверждает, что неизвестное среднее значение генеральной совокупности не меньше, чем известное заданное значение a_0.

H_1: $a < a_0$

Альтернативная гипотеза утверждает, что неизвестное среднее значение генеральной совокупности меньше, чем известное заданное значение a_0.

Односторонняя проверка того, что а меньше a_0

H_0: $a \leqslant a_0$

Нулевая гипотеза утверждает, что неизвестное среднее значение генеральной совокупности не больше, чем известное заданное значение a_0.

H_1: $a > a_0$

Альтернативная гипотеза утверждает, что неизвестное среднее значение генеральной совокупности больше, чем известное заданное значение a_0.

Приведем алгоритм проверки каждой из этих двух нулевых гипотез.

Алгоритм односторонней проверки того, что а больше a_0

Проверяется нулевая гипотеза H_0: $a \leqslant a_0$ против альтернативной H_1: $a > a_0$.

1. Задают уровень значимости α.

2. Строят односторонний доверительный интервал с доверительной вероятностью $1 - \alpha$: $a \geqslant \overline{X} - l_{1-\alpha} \cdot \sigma(\overline{X})$, где $l_{1-\alpha}$ находится по таблицам стандартного нормального распределения из условия $P(Z < l_{1-\alpha}) = 1 - \alpha$.

3. Выясняют, находится ли заданное значение a_0 внутри этого доверительного интервала.

Если да, то нулевую гипотезу не отвергают: выборочное среднее $\overline{X}$ не является значимо большим, чем заданное значение a_0.

Если нет, то нулевую гипотезу отвергают: выборочное среднее $\overline{X}$ значимо больше, чем заданное значение a_0.

Аналогично поступают в случае использования статистики $\frac{\overline{X} - a_0}{\sigma}\sqrt{n}$.

Проверяют, выполняется ли неравенство $\frac{\overline{X} - a_0}{\sigma}\sqrt{n} \leqslant l_{1-\alpha}$.

Если да, то нулевую гипотезу не отвергают: выборочное среднее $\overline{X}$ не является значимо большим, чем заданное значение a_0.

Если нет, то нулевую гипотезу отвергают: выборочное среднее $\overline{X}$ значимо больше, чем заданное значение a_0.

Алгоритм односторонней проверки того, что а меньше a_0

Проверяется нулевая гипотеза H_0: $a \geqslant a_0$ против альтернативной H_1: $a < a_0$.

1. Задают уровень значимости α.

2. Строят односторонний доверительный интервал с доверительной вероятностью $1 - \alpha$: $a \geqslant \overline{X} + l_{1-\alpha} \cdot \sigma(\overline{X})$, где $l_{1-\alpha}$ находится по таблицам стандартного нормального распределения из условия $P(Z < l_{1-\alpha}) = 1 - \alpha$.

3. Выясняют, находится ли заданное значение a_0 внутри этого доверительного интервала.

Если да, то нулевую гипотезу не отвергают: выборочное среднее $\overline{X}$ не является значимо меньшим, чем заданное значение a_0.

Если нет, то нулевую гипотезу отвергают: выборочное среднее $\overline{X}$ значимо меньше, чем заданное значение a_0.

Аналогично поступают в случае использования статистики $\frac{\overline{X} - a_0}{\sigma}\sqrt{n}$.

Проверяют, выполняется ли неравенство $\frac{\overline{X} - a_0}{\sigma}\sqrt{n} < -l_{1-\alpha}$.

Если да, то нулевую гипотезу не отвергают: выборочное среднее $\overline{X}$ не является значимо меньшим, чем заданное значение a_0.

Если нет, то нулевую гипотезу отвергают: выборочное среднее $\overline{X}$ значимо меньше, чем заданное значение a_0.

Пример 4. Внедряют новую технологию производства некоторого изделия, направленную на снижение переменных издержек производства одного изделия (т. е. стоимости изготовления одного изделия за вычетом постоянных расходов таких, как арендная плата, заработная плата административного персонала и т. п.). Переменные издержки производства одной детали являются случайной величиной со средним значением, равным 854 р. и средним квадратичным отклонением 15 р. Изучение 50 случайно отобранных изделий, изготовленных по новой технологии, показало, что средние переменные издержки производства одной детали составили 817 р. Похоже, что они снизились. Требуется выяснить, значимо ли это снижение.

❒ Нулевая гипотеза H_0: $a \geqslant 854$; альтернативная H_1: $a < 854$. Пусть уровень значимости $\alpha = 0{,}05$. Строим правосторонний доверительный интервал для среднего генеральной совокупности: $a \geqslant \overline{X} + l_{0,95} \cdot \sigma(\overline{X})$, где $l_{0,95}$ находится по таблицам стандартного нормального распределения из условия $P(Z < l_{0,95}) = 0{,}95$; $l_{0,95} = 1{,}64$. Доверительный интервал имеет вид $a \geqslant 817 + 1{,}64 \cdot 15$, или $a \geqslant 842$. Заданное значение 854 находится внутри этого интервала. Нулевую гипотезу на уровне значимости 0,05 не отвергаем: выборочное среднее $\overline{X}$ не является значимо меньшим, чем заданное значение a_0. Такой же результат получим при использовании статистики $\frac{\overline{X} - a_0}{\sigma}\sqrt{n}$. Ее значение равно $\frac{817 - 854}{15}\sqrt{30} = -13{,}5$. Сравним это значение с $-1{,}64$: $-13{,}5 < -1{,}64$. Нулевую гипотезу на уровне значимости 0,05 не отвергаем. Снижение переменных издержек производства незначимо.

Можно ли теперь принять решение о том, стоит ли повсеместно внедрять новую технологию? Статистика ответила только на поставленный вопрос о значимости снижения переменных издержек производства. Окончательное решение за теми, кто управляет производством: они могут учесть и другие факторы. ■

➤

Одним из ограничений рассмотренных методов проверки гипотезы о среднем генеральной совокупности является предположение о том, что дисперсия генеральной совокупности известна. На практике это бывает очень редко и тогда значение статистики $\frac{\overline{X}-a_0}{\sigma}\sqrt{n}$ не может быть вычислено, доверительный интервал $\left(\overline{X} - l\frac{\sigma}{\sqrt{n}} \leqslant a \leqslant \leqslant \overline{X} + l\frac{\sigma}{\sqrt{n}}\right)$ не может быть построен. Эта трудность может быть устранена, если воспользоваться статистикой $\frac{\overline{X}-a_0}{S}\sqrt{n-1}$, где S — стандартная ошибка выборки. Об этой случайной величине говорят, что она имеет *распределение Стьюдента* с $n-1$ степенью свободы. Составлены таблицы этого распределения (см. Приложение, табл. П.3). Снова обратите внимание на то, что числитель этой статистики характеризует отклонение среднего генеральной совокупности от выборочного среднего, а знаменатель $\frac{S}{\sqrt{n-1}}$ показывает, в каких единицах описывается это отклонение. Напомним, что

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{m}(X_i-\overline{X})^2\,n_i,$$

где X_1, X_2, ..., X_m — все различные элементы выборки; n_1, n_2, ..., n_m — их частоты; $n = = n_1 + n_2 + \ldots + n_m$ — объем выборки.

Представляет интерес история открытия этого распределения. Первые идеи применения методов математической статистики в массовом производстве принадлежат одному из директоров крупных пивоваренных заводов Гиннеса в Англии. В начале XX века он прочитал книгу по теории вероятностей и подумал, что «из этого можно делать деньги». Пригласив к себе Уильяма Госсета (1876–1937), младшего служащего завода, директор предложил ему поехать в единственный в то время центр статистических исследований в Лондоне для учебы под руководством крупнейшего статистика, биолога и философа Карла Пирсона (1857–1936). Госсет проявил инициативу и выдающиеся способности и вскоре приступил к самостоятельным исследованиям. Их результаты были весьма значительны: одни представляли несомненную ценность для пивоварения, другие – большой теоретический интерес. В результате научный мир был изумлен рядом первоклассных статей в журнале «Биометрика», опубликованных начиная с 1908 г. под псевдонимом «Student» («Стьюдент»), что значит «Студент». Эти работы совершили переворот в статистике.

Пример 5. В некотором районе 9-й класс заканчивают 1000 школьников. Для диагностики достижения стандарта математического образования был составлен тест из 30 вопросов, правильный ответ на каждый вопрос оценивался одним баллом. Средний балл девятиклассников по результатам тестирования прошлых лет был принят равным 15. Проведены специальные занятия для подготовки учащихся к тестированию. Результаты выборочного испытания 50 учащихся, прошедших эти занятия, представлены в таблице 5.6.

Подтверждают ли эти данные эффективность специальных занятий?

Таблица 5.6

Количество баллов	0—5	6—10	11—15	16—20	21—25	26—30
Число учащихся	1	5	16	20	6	2

❒ Нулевая гипотеза — средний балл девятиклассника равен 15, альтернативная — средний балл девятиклассника не равен 15. Для проверки этой гипотезы используем статистику $t = \dfrac{\overline{X} - a_0}{S}\sqrt{n-1}$, имеющую распределение Стьюдента с $n - 1$ степенью свободы. Вычислим по выборке ее значение. Воспользуемся программой Excel. В результате получим: $\overline{X} = 16{,}1$; $S = 5{,}15$; $\left|\dfrac{\overline{X} - a_0}{S}\sqrt{n-1}\right| =$ $= \dfrac{16{,}1 - 15}{5{,}15} \cdot \sqrt{49} = 1{,}5$. Примем уровень значимости равным 0,05. По таблице распределения Стьюдента найдем такое число t_α, которое удовлетворяет условию $P(|t| > t_\alpha) = \alpha$, или $2P(t < t_\alpha) - 1 = 1 - \alpha$, или $P(t < t_\alpha) = 1 - \frac{\alpha}{2}$. Получим $t_{0,05} = 1{,}96$. Это значение находится в строке «бесконечность» (число степеней свободы $49 > 40$), в столбце

95% ($1 - \alpha = 0{,}95$). Так как $1{,}5 < 1{,}96$, то нулевая гипотеза не отвергается: не получены убедительные данные в пользу эффективности специальных занятий. ■ ◀

Контрольные вопросы

1. Согласны ли вы с тем, что следующие утверждения, касающиеся проверки нулевой гипотезы H_0: $a = a_0$, эквивалентны:

а) заданное значение a_0 не находится в пределах доверительного интервала;

б) значение статистики

$$\frac{\overline{X} - a_0}{\sigma}\sqrt{n} > l_\alpha;$$

в) нулевая гипотеза отвергается;

г) выборочное среднее $\overline{X}$ значимо отличается от a_0;

д) наблюдаемую разность $\overline{X} - a_0$ нельзя объяснить лишь случайностью;

е) результат проверки гипотезы является статистически значимым?

2. Согласны ли вы с тем, что следующие утверждения, касающиеся проверки нулевой гипотезы H_0: $a = a_0$, эквивалентны:

а) заданное значение a_0 находится в пределах доверительного интервала;

б) значение статистики

$$\frac{\overline{X} - a_0}{\sigma}\sqrt{n} < l_\alpha;$$

в) нулевая гипотеза не отвергается;

г) выборочное среднее $\overline{X}$ незначимо отличается от a_0;

д) наблюдаемая разность $\overline{X} - a_0$ может быть обусловлена просто случайностью;

е) результат проверки гипотезы не является статистически значимым?

3. Сформулируйте утверждения, подобные утверждениям а)—е) из вопроса 1, для случая односторонней проверки нулевой гипотезы H_0: $a \geqslant a_0$, если она отвергается.

4. Сформулируйте утверждения, подобные утверждениям а)—е) из вопроса 1, для случая односторонней проверки нулевой гипотезы H_0: $a \leqslant a_0$, если она отвергается.

5. Сформулируйте утверждения, подобные утверждениям а)—е) из вопроса 2, для случая односторонней проверки нулевой гипотезы H_0: $a \geqslant a_0$, если она не отвергается.

6. Сформулируйте утверждения, подобные утверждениям а)—е) из вопроса 2, для случая односторонней проверки нулевой гипотезы H_0: $a \leqslant a_0$, если она не отвергается.

7. Опишите этапы выполнения двусторонней проверки гипотезы о среднем генеральной совокупности с использованием доверительного интервала.

8. Опишите этапы выполнения двусторонней проверки гипотезы о среднем генеральной совокупности с использованием статистики $\frac{\overline{X} - a_0}{\sigma}\sqrt{n}$.

9. На заседании совета директоров акционерной компании было высказано предложение о предоставлении работникам компании права получать часть прибыли в зависимости от результатов работы. Было опрошено 256 экспертов, которым предложили оценить влияние передачи акций на качество продукции по пятибалльной шкале: −2, −1, 0, 1, 2, где −2 означало «резко отрицательное влияние», а 2 — «сильное положительное влияние». Среднее значение оценки составило 0,39. Можно ли считать, что этот результат свидетельствует о том, что эксперты считают высказанное предложение полезным?

Задачи

533. Количество деловой древесины в одном дереве на некотором участке леса есть случайная величина, математическое ожидание которой равно 1,3 м^3. Проведено выборочное исследование 1000 деревьев на другом участке леса. Результаты исследования представлены в таблице 5.7.

Таблица 5.7

Количество деловой древесины в одном дереве, м^3	0,25—0,75	0,75—1,25	1,25—1,75
Число деревьев	208	484	308

Можно ли утверждать, что второй участок столь же богат деловой древесиной, как и первый? Сформулируйте гипотезу и проверьте ее на уровне значимости 0,05 с использованием:

а) доверительного интервала;

б) соответствующей статистики;

в) неравенства Чебышёва.

Дисперсию генеральной совокупности положите равной выборочной дисперсии, вычисленной по данным выборочного исследования.

534. Отклонение диаметров валиков, изготавливаемых на некотором станке, является случайной величиной со средним, равным 10 мк. После переналадки станка отобрали для контроля 250 валиков. Результаты измерений приведены в таблице 5.8.

Таблица 5.8

Отклонение диаметра от номинала, мк	0—5	5—10	10—15	15—20	20—25
Число валиков	15	75	100	50	10

Можно ли утверждать, что переналадка станка оказалась эффективной? Сформулируйте гипотезу и проверьте ее на уровне значимости 0,05 с использованием:

а) доверительного интервала;

б) соответствующей статистики;

в) неравенства Чебышёва.

Дисперсию генеральной совокупности положите равной выборочной дисперсии, вычисленной по данным выборочного исследования.

535. Чувствительность радиоприемников, выпущенных некоторым заводом, является случайной величиной со средним, равным 300 мкВ. После усовершенствования технологии производства отобрали для контроля 40 радиоприемников. Результаты измерений чувствительности приведены в таблице 5.9.

Таблица 5.9

Чувствительность, мкВ	75—175	175—275	275—375	375—475	475—575	575—675
Число радиоприемников	1	14	14	8	2	1

Можно ли утверждать, что усовершенствование технологии оказалось эффективным? Сформулируйте гипотезу и проверьте ее на уровне значимости 0,05 с использованием:

а) доверительного интервала;

б) соответствующей статистики;

в) неравенства Чебышёва.

Дисперсию генеральной совокупности положите равной выборочной дисперсии, вычисленной по данным выборочного исследования.

536. Пекарня выпекает буханки хлеба, на этикетках которых указана масса 800 г. Ниже приведены значения массы (в г) случайно отобранных буханок из продукции одного дня:

816, 776, 784, 880, 880, 816, 784, 824, 824, 840, 816, 848.

а) Постройте 95%-ный доверительный интервал для средней массы всех буханок, изготовленных в тот день, когда производилось выборочное исследование.

б) Какое число нужно использовать в качестве заданного значения при проверке гипотезы о среднем значении массы буханок, выпекаемых пекарней?

в) Сформулируйте гипотезы H_0 и H_1.

г) Выполните двустороннюю проверку гипотезы на уровне значимости 0,05.

д) Какие ошибки могут быть допущены?

537. На некоторой фирме принято решение о запуске нового изделия в производство, если потенциальные потребители согласятся платить в среднем 600 р. за изделие. Изучение ответов 290 случайно отобранных потенциальных потребителей показало, что они готовы платить в среднем 545 р. за данное изделие. Стандартное отклонение составляет 60 р.

а) Какое число нужно использовать в качестве заданного значения при проверке гипотезы о среднем для всех потенциальных покупателей?

б) Выполните двустороннюю проверку гипотезы на уровне значимости 0,05.

в) Выполните двустороннюю проверку гипотезы на уровне значимости 0,01.

г) Почему в данной задаче приемлема односторонняя проверка?

д) Сформулируйте словесно и на математическом языке нулевую и альтернативную гипотезы для односторонней проверки.

е) Выполните одностороннюю проверку и опишите полученный результат.

§ 5.6. Проверка гипотез о биномиальной вероятности

Во многих практических задачах возникает потребность проверить, согласуются ли результаты испытаний Бернулли с допущением о том, что вероятность «успеха» в каждом из этих испытаний равна заранее заданному числу p_0.

Например, в задачах выборочного контроля качества произведенной продукции, как правило, принимается, что какая-то доля, например p_0, всех производимых изделий бракована. Это может быть сделано на основании длительных наблюдений за качеством продукции. Далее приемочные правила контроля выглядят примерно так: если в выборке определенного объема доля дефектных изделий превышает заданное число, то вся партия бракуется, в противном случае — не бракуется. Возникает вопрос, насколько это правило согласуется с принятым $100p_0$ процентом дефектных изделий.

Фактически речь идет о проверке нулевой гипотезы $p = p_0$ при альтернативной $p \neq p_0$. Для проверки этой гипотезы будут использованы:

1) правило трех сигм;

2) метод доверительных интервалов;

3) неравенство Чебышёва;

4) статистика $\dfrac{m-np}{\sqrt{np(1-p)}}$, имеющая при достаточно большом n приближенно стандартное нормальное распределение;

5) односторонняя проверка.

Рассмотрим все эти методы.

1) *Правило трех сигм.*

Согласно правилу трех сигм (см. § 4.8), если хотя бы одно из наблюдаемых значений биномиальной случайной величины с параметрами n и p_0 не попадает в интервал

$$(np_0 - 3\sqrt{np_0(1-p_0)}\,;\ np_0 + 3\sqrt{np_0(1-p_0)}\,),$$

т. е. наблюдается практически невозможное событие, то гипотезу о значении вероятности «успеха» следует отвергнуть. В противном случае нет оснований для отклонения рассматриваемой гипотезы.

? Чему равна вероятность ошибки первого рода при использовании правила трех сигм?

2) *Метод доверительных интервалов.*

Строят доверительный интервал для вероятности p, исходя из значения $\bar{p}$, с доверительной вероятностью $1-\alpha$. Если значение p_0 находится вне этого доверительного интервала, то p_0 не может рассматриваться как допустимое значение вероятности, и гипотезу H_0 следует отвергнуть на уровне значимости α в пользу альтернативной гипотезы. В противном случае нулевую гипотезу на уровне значимости α не отвергают.

Как отмечалось в § 4.10, закон распределения случайной величины, являющейся суммой очень большого числа независимых в совокупности случайных величин, какова бы ни была природа слагаемых, лишь бы каждое из них было мало по сравнению со всей суммой, должен быть близок к нормальному закону. На этом основании мы делаем вывод, что число наступлений m «успехов» в n испытаниях Бернулли при достаточно большом n имеет приближенно нормальное распределение. В том же параграфе указывалось, что если случайная величина X имеет нормальное распределение со средним a и с дисперсией σ^2, то случайная величина $Z = \dfrac{X-a}{\sigma}$ имеет стандартное нормальное распределение. Так как $Mm = np$, $Dm = np(1-p)$, то случайная величина $\dfrac{m-np}{\sqrt{np(1-p)}}$ имеет приближенно при достаточно большом n стандартное нормальное распределение.

По таблице нормального распределения (см. Приложение, табл. П.1) можно по заданной доверительной вероятности $1-\alpha$ найти такое число l, что $P\left(\left|\dfrac{m-np}{\sqrt{np(1-p)}}\right| \leqslant l\right) = 1-\alpha$.

Так как p неизвестно и $p(1-p) \leqslant \dfrac{1}{4}$, то произведение $p(1-p)$ заменяют дробью $\dfrac{1}{4}$. Тогда неравенство, стоящее под знаком вероятности, можно переписать в следующем виде:

$$l \geqslant \left|\frac{m-np}{\sqrt{np(1-p)}}\right| \geqslant |m-np| \cdot \frac{2}{\sqrt{n}} = \left|\frac{m}{n}-p\right| \cdot \frac{2n}{\sqrt{n}} = |\bar{p}-p| \cdot 2\sqrt{n}.$$

Напомним, что $\bar{p} = \dfrac{m}{n}$.

Решая это неравенство относительно p, получим

$$\bar{p} - \frac{l}{2\sqrt{n}} \leqslant p \leqslant \bar{p} + \frac{l}{2\sqrt{n}}.$$

Это и есть искомый доверительный интервал для неизвестной вероятности. Напомним, что по заданной вероятности $1-\alpha$ число l по таблицам нормального распределения находится из условия $2P(Z < l) - 1 = 1 - \alpha$, или $P(Z < l) = 1 - \dfrac{\alpha}{2}$. Здесь Z — случайная величина, имеющая стандартное нормальное распределение.

? Почему можно считать, что число успехов в n испытаниях Бернулли имеет при достаточно большом n приближенно нормальное распределение?

➤

Существуют и более точные методы, связанные с построением доверительных вероятностей для неизвестной вероятности. Не будем заменять $p(1-p)$ дробью $\frac{1}{4}$. Решим неравенство, стоящее под знаком вероятности в равенстве $P\left(\left|\frac{m-np}{\sqrt{np(1-p)}}\right| \leqslant l\right) = 1-\alpha$, относительно p. Имеем:

$$\frac{|\bar{p}-p|}{\sqrt{\frac{p(1-p)}{n}}} \leqslant l; \quad \bar{p}^2 - 2p\bar{p} + p^2 \leqslant \frac{l^2 p}{n} - \frac{l^2 p^2}{n};$$

$$p^2\left(1 + \frac{l^2}{n}\right) - 2p\left(\bar{p} + \frac{l^2}{2n}\right) + \bar{p}^2 \leqslant 0.$$

В левой части последнего неравенства стоит квадратный трехчлен относительно p. Найдя его корни и решив неравенство относительно p, получим искомый доверительный интервал для вероятности p:

$$\left(\frac{\bar{p} + \frac{l^2}{2n} - l\sqrt{\frac{\bar{p}(1-\bar{p})}{n} + \frac{l^2}{4n^2}}}{1+\frac{l^2}{n}}; \frac{\bar{p} + \frac{l^2}{2n} + l\sqrt{\frac{\bar{p}(1-\bar{p})}{n} + \frac{l^2}{4n^2}}}{1+\frac{l^2}{n}}\right).$$

? Почему этот доверительный интервал при той же доверительной вероятности будет более узким по сравнению с доверительным интервалом $\bar{p} - \frac{l}{2\sqrt{n}} \leqslant p \leqslant \bar{p} + \frac{l}{2\sqrt{n}}$? ➤

3) ***Неравенство Чебышёва.***

Применим теперь для проверки сформулированной гипотезы неравенство Чебышёва: $P\left(|\bar{p}-p| \leqslant \frac{h}{2\sqrt{n}}\right) \geqslant 1 - \frac{1}{h^2}$. По заданному уровню значимости $\frac{1}{h^2}$ можно найти h. Если хотя бы одно из наблюдаемых значений $\bar{p}$ не попадает в интервал $\left(p_0 - \frac{h}{2\sqrt{n}}; p_0 + \frac{h}{2\sqrt{n}}\right)$, то гипотезу $p = p_0$ следует отвергнуть.

? В каких случаях удобно пользоваться неравенством Чебышева для проверки гипотезы о значении биномиальной вероятности?

4) *Статистика* $\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}}$.

Этот способ проверки гипотезы о среднем генеральной совокупности состоит в том, что сначала вычисляют статистику $\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}}$, а затем по таблицам нормального распределения по заданному уровню значимости α находят такое значение l_α, что $P(Z > l_\alpha) = \frac{\alpha}{2}$. Здесь Z — случайная величина, имеющая стандартное нормальное распределение. Далее модуль значения статистики $\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}}$ сравнивают со значением l_α. Если $\left|\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}}\right| > l_\alpha$, то гипотезу H_0: $p = p_0$ отвергают в пользу альтернативной гипотезы H_1: $p \neq p_0$, в противном случае ее не отвергают (на уровне значимости α).

5) *Односторонняя проверка.*

Нулевая гипотеза H_0: $p \geqslant p_0$ утверждает, что неизвестная вероятность не менее заданного значения p_0. Альтернативная гипотеза H_1: $p < p_0$ утверждает, что неизвестная вероятность меньше заданного значения p_0. По заданному уровню значимости α строят односторонний доверительный интервал с доверительной вероятностью $1 - \alpha$: $p \geqslant \bar{p} + l_{1-\alpha} \cdot \sqrt{\dfrac{\bar{p}(1 - \bar{p})}{n}}$, где $l_{1-\alpha}$ находится по таблицам стандартного нормального распределения из условия $P(Z < l_{1-\alpha}) = 1 - \alpha$. Проверяют, находится ли заданное значение p_0 внутри этого доверительного интервала. Если да, то нулевую гипотезу не отвергают: относительная частота $\bar{p}$ не является значимо меньшим, чем заданное значение p_0. Если нет, то нулевую гипотезу отвергают: относительная частота $\bar{p}$ значимо меньше, чем заданное значение p_0.

Аналогично поступают в случае использования статистики $\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}}$.

Проверяют, выполняется ли неравенство $\dfrac{m - np_0}{\sqrt{np_0(1 - p_0)}} < -\,l_{1-\alpha}$. Если да, то нулевую гипотезу не отвергают: относительная частота $\bar{p}$ не является значимо меньшим, чем заданное значение p_0. Если нет, то нулевую гипотезу отвергают: выборочное среднее $\overline{X}$ значимо меньше, чем заданное значение a_0.

Точно так же проводится односторонняя проверка, если нулевая гипотеза H_0: $p \leqslant p_0$ утверждает, что неизвестная вероятность не более заданного значения p_0, а альтернативная гипотеза H_1: $p > p_0$ утверждает, что неизвестная вероятность больше заданного значения p_0.

Проиллюстрируем все рассмотренные методы на следующем примере.

Пример 1. Производство микросхем принято считать успешным, если не менее 10% произведенных микросхем имеют особенно высокое качество. После модернизации производства в выборке из 500 микросхем оказалось, что 58 имеют особенно высокое качество. Можно ли считать, что модернизация производства значимо повысила заданное граничное значение 10% или этот результат является случайным?

❒ Нулевая гипотеза H_0: $p = 0{,}1$ утверждает, что микросхемы особенно высокого качества составляют 10% от всего объема продукции. Альтернативная гипотеза H_1: $p \neq 0{,}1$ утверждает, что доля микросхем особенно высокого качества отличается от 10%: либо она выше, либо ниже. По условию размер выборки $n = 500$, наблюдаемая частота $m = 58$, относительная частота $\bar{p} = \dfrac{58}{500} = 0{,}116$.

1) Сначала применим для проверки нулевой гипотезы правило трех сигм. В нашем случае

$$np_0 - 3\sqrt{np_0(1 - p_0)} = 500 \cdot 0{,}1 - 3 \cdot \sqrt{500 \cdot 0{,}1 \cdot 0{,}9} =$$
$$= 50 - 3 \cdot 6{,}7 = 29{,}9; \quad np_0 + 3\sqrt{np_0(1 - p_0)} = 70{,}1.$$

Наблюдаемая частота $m = 58$ попадает в интервал (29,9; 70,1), т. е. на уровне значимости $\frac{1}{9}$ нет оснований отвергать нулевую гипотезу. Значение $\bar{p} = \frac{58}{500} = 0{,}116$ незначимо отличается от 0,1. Другими словами, мы не получили доказательства того, что модернизация производства привела к изменению процента микросхем особенно высокого качества.

2) Применим метод доверительных интервалов для проверки той же гипотезы. Подставив значения $n = 500$, $\bar{p} = 0{,}116$, $l = 1{,}96$ (значение l найдено по таблицам нормального распределения при доверительной вероятности $1 - \alpha = 0{,}95$) в неравенство $\bar{p} - \frac{l}{2\sqrt{n}} \leqslant p \leqslant \bar{p} - \frac{l}{2\sqrt{n}}$, получим следующий доверительный интервал для вероятности p: (0,072; 0,160). Он содержит заданное значение $p_0 = 0{,}1$, поэтому нулевую гипотезу на уровне значимости 0,05 не отвергаем. Получили тот же результат, что и с помощью правила трех сигм.

➤

Подставив значения $n = 500$, $\bar{p} = 0{,}116$, $l = 1{,}96$ в неравенство

$$\frac{\bar{p} + \frac{l^2}{2n} - l\sqrt{\frac{\bar{p}(1-\bar{p})}{n} + \frac{l^2}{4n^2}}}{1 + \frac{l^2}{n}} \leqslant p \leqslant \frac{\bar{p} + \frac{l^2}{2n} + l\sqrt{\frac{\bar{p}(1-\bar{p})}{n} + \frac{l^2}{4n^2}}}{1 + \frac{l^2}{n}},$$

получим следующий доверительный интервал для вероятности p: (0,1189; 0,1198). Он не содержит заданное значение $p_0 = 0{,}1$, поэтому нулевую гипотезу на уровне значимости 0,05 отвергаем, т. е. более точные статистические методы говорят в пользу эффективности новой технологии производства микросхем. ➤

3) Применим неравенство Чебышёва для проверки гипотезы. В нашем случае значение $\bar{p} = 0{,}116$ попадает в интервал $\left(0{,}1 - \frac{3{,}2}{2\sqrt{500}};\ 0{,}1 + \frac{3{,}2}{2\sqrt{500}}\right)$, или (0,028; 0,172). Поэтому нет оснований отвергать гипотезу $p = p_0$.

4) Используем статистику $\frac{m - np_0}{\sqrt{np_0(1-p_0)}}$. Модуль ее значения равен $\frac{58 - 500 \cdot 0,1}{\sqrt{500 \cdot 0,1 \cdot 0,9}} = 1,192$. Сравниваем это значение со значением $l_\alpha = 1,96$: $1,192 < 1,96$. Поэтому нулевую гипотезу на уровне значимости 0,05 не отвергаем.

5) Применим одностороннюю проверку. Так как от модернизации производства ожидалось повышение процента микросхем особенно высокого качества, то нулевую гипотезу можно сформулировать так: $p \leqslant p_0$, а альтернативную: $p > p_0$. Уровень значимости выбираем равным 0,05. Строим левосторонний доверительный интервал для среднего генеральной совокупности: $p \geqslant \bar{p} - l_{1-\alpha} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$, где $l_{0,95}$ находится по таблицам стандартного нормального распределения из условия $P(Z < l_{0,95}) = 0,95$; $l_{0,95} = 1,64$. Доверительный интервал имеет вид $p \geqslant 0,116 - 1,64 \cdot \sqrt{\frac{0,116 \cdot 0,884}{500}}$, или $p \geqslant 0,0925$. Заданное значение 0,1 находится внутри этого интервала. Нулевая гипотеза на уровне значимости 0,05 не отвергается: относительная частота $\bar{p} = 0,116$ не является значимо большей, чем заданное значение $p_0 = 0,1$. ■

Обращаем внимание на то, что все процедуры проверки нулевой гипотезы о значении вероятности базируются на том, что если *хотя бы одно из наблюдаемых значений* случайной величины, имеющей биномиальное распределение с параметрами n и p_0, не попадает в интервал, построенный по правилу трех сигм, или приводит к доверительному интервалу (двустороннему или одностороннему), не содержащему заданное значение вероятности, или область, построенная с помощью той или иной статистики и имеющая малую вероятность, содержит хотя бы одно значение относительной частоты события или числа «успехов», то нулевую гипотезу на соответствующем уровне значимости отвергают.

Пример 2. Изменили технологию подготовки семян к севу. Для проверки эффективности новой технологии пять раз отобрали по 500 семян. Оказалось, что взошло соответственно 405, 410, 410, 415, 405 семян. Можно ли утверждать, что новая технология эффективнее прежней, если средний процент взошедших семян при старой технологии равнялся 80% ?

❒ Речь идет о проверке нулевой гипотезы H_0: $p = 0{,}8$ при альтернативной гипотезе H_1: $p \neq 0{,}8$. Применим различные способы проверки этой гипотезы.

1) Правило трех сигм. Если нулевая гипотеза верна, то интервал $\left(np_0 - 3\sqrt{np_0(1-p_0)}\,;\, np_0 + 3\sqrt{np_0(1-p_0)}\right)$ с вероятностью, не меньшей $\frac{8}{9}$, содержит наблюдаемое значение числа «успехов» в n испытаниях Бернулли. В нашем случае указанный интервал имеет вид $\left(500\cdot 0{,}8 - 3\sqrt{500\cdot 0{,}8\cdot(1-0{,}8)}\,;\, 500\cdot 0{,}8 + 3\sqrt{500\cdot 0{,}8\cdot(1-0{,}8)}\right)$, или (373; 427). Все наблюдаемые значения 405, 410, 410, 415, 405 попали в этот интервал. Нет оснований отвергать нулевую гипотезу и считать новую технологию более эффективной.

2) Метод доверительных интервалов. Задаем доверительную вероятность $1 - \alpha = 0{,}99$. Строим доверительный интервал $\bar{p} - \frac{l}{2\sqrt{n}} \leqslant p \leqslant \bar{p} + \frac{l}{2\sqrt{n}}$, где число l находится по таблице нормального распределения из условия $P(Z < l) = 1 - \frac{0{,}01}{2} = 0{,}995$; $l = 2{,}58$. Доверительный интервал имеет вид $\left(\bar{p} - \frac{2{,}58}{2\sqrt{500}}\,;\, \bar{p} + \frac{2{,}58}{2\sqrt{500}}\right)$, или $(\bar{p} - 0{,}057;\ \bar{p} + 0{,}058)$. Наблюдаемые относительные частоты события «семя взошло» равны 0,810, 0,820, 0,820, 0,830, 0,810. Они приводят к следующим доверительным интервалам: (0,753; 0,868), (0,763; 0,878), (0,763; 0,878), (0,773; 0,888), (0,753; 0,868). Все они содержат заданное число 0,8. Поэтому отличие наблюдаемых частот от 0,8 является незначимым и нулевую гипотезу не отвергаем.

3) Неравенство Чебышёва. Задаем уровень значимости $\frac{1}{h^2} = 0{,}05$. Отсюда $h = 4{,}47$. Строим интервал $\left(0{,}8 - \frac{4{,}47}{2\sqrt{500}}\,;\, 0{,}8 + \frac{4{,}47}{2\sqrt{500}}\right)$, или (0,701; 0,900). Все наблюдаемые относительные частоты 0,810, 0,820, 0,820, 0,830, 0,810 попадают в этот интервал. Нулевую гипотезу на уровне значимости 0,05 не отвергаем.

Аналогично можно применить и другие методы. ■

Развитие статистики было обусловлено задачами общественной значимости. С накоплением в развитых европейских странах многочисленной информации по вопросам динамики народонаселения, развития торговли, экономики и хозяйственной деятельности, здравоохранения и др. возникла потребность поиска способов анализа статистических данных и их теоретического осмысления. Первой на путь разработки аппарата математической статистики стала английская школа так называемых «политических арифметиков», которую возглавлял В. Петти (1623–1687). Представители этого направления пытались по данным статистических наблюдений сформулировать законы общественных явлений. О существовании таких законов свидетельствовали факты повторения соотношений между количеством новорожденных мальчиков и девочек, динамика рождаемости и смертности. Первые попытки использовать количественные измерения как инструмент научного познания живых существ были реализованы Г. Галилеем (1564–1642), Санторио (1561–1636), Дж. А. Борелли (1608–1679) и др. Биологические исследования послужили в XIX веке толчком для постановки многочисленных вопросов, приведших в начале XX века к выделению математической статистики в особую науку. Первым, кто соединил методы антропометрии и социальной статистики с теорией вероятностей, был А. Кетле (1796–1874). Существенные результаты в области теории вероятностей и математической статистики были получены П. Лапласом (1749–1827) и К. Ф Гауссом (1777–1855). Первый из них доказал, что нормальное распределение служит приближением для биномиального. Гаусс вывел нормальный закон распределения случайных ошибок наблюдений. Почти одновременно с ним и независимо от него американский математик Р. Эдрейн (1775–1843) получил тот же результат. Большой вклад в дальнейшее развитие прикладной и теоретической статистики внесли Ф. Гальтон (1822–1911) и К. Пирсон (1857–1936). Гальтон впервые применил статистический подход Кетле к решению проблем наследования и изменчивости. Его идеи развил Пирсон. Гальтона и Пирсона считают основоположниками современной биометрии и психометрии. Дальнейшее развитие прикладная статистика получила в работах В. Госсета (1876–1937), который публиковал свои работы под псевдонимом «Стьюдент». Значительный вклад в развитие теории малых выборок внес Р. Фишер (1890–1962). Кроме того, он разработал метод, называемый дисперсионным анализом, для интерпретации результатов агрономических опытов. Он также исследовал так называемый метод наибольшего правдоподобия оценивания неизвестных параметров. В отечественной науке одним из первых, кто составил подборку статистических прикладных методов еще в 1909–1911 гг., был А. В. Леонтович (1869–1943), тогда же появились и «Очерки по теории статистики» А. А. Чупрова (1874–1926). В 1916 г. вышло пособие по статистическим методам А. А. Кауфмана (1864–1919). Известными исследователями в области математической статистики были С. Н. Бернштейн (1880–1968), А. Я. Хинчин (1894–1959), Е. Е. Слуцкий (1880–1948), В. И. Романовский (1879–1954), А. Н. Колмогоров (1903–1992), Н. В. Смирнов (1900–1966), Б. В. Гнеденко (1912–1996), И. И. Гихман (1918–1985) и многие другие. Исследования в области математической статистики продолжаются и в настоящее время.

Контрольные вопросы

1. Есть подозрение, что некоторая монета неправильная. Для того чтобы это проверить, эту монету подбрасывают четыре раза. Если при всех четырех подбрасываниях она падает гербом вверх, то гипотезу о правильности монеты отвергают. Каким может быть уровень значимости этого критерия?

2. Есть подозрение, что некоторая монета неправильная. Для того чтобы это проверить, эту монету подбрасывают четыре раза. Если при всех четырех подбрасываниях она падает гербом вверх, то гипотезу о правильности монеты отвергают. Сформулируйте нулевую гипотезу. p — это вероятность того, что монета упадет гербом вверх.

3. Некто утверждает, что всегда может отличить по вкусу китайский чай от цейлонского. Ему предлагают девять пар чашек чая, причем каждая пара состоит из одной чашки китайского и одной чашки цейлонского чая. Он правильно указывает сорт чая в восьми случаях из девяти и ошибается в одном. Сформулируйте проверяемую гипотезу. p — это вероятность того, что правильно будет определен сорт чая из пары чашек.

4. Некто утверждает, что всегда может отличить по вкусу китайский чай от цейлонскиого. Ему предлагают девять пар чашек чая, причем каждая пара состоит из одной чашки китайского и одной чашки цейлонского чая. Он правильно указывает сорт чая в восьми случаях из девяти и ошибается в одном. Каким может быть уровень значимости этого критерия?

5. У Василия есть игральный кубик, по поводу которого он считает, что кубик чаще падает вверх гранью с 6 очками. Он бросает кубик 4 раза и все 4 раза на кубике выпадает 6 очков. Используя 9% -ный уровень значимости, следует ли ему отклонить гипотезу о том, что вероятность выпадения 6 очков равна $p = \frac{1}{6}$?

Задачи

538. Установлено, что в публицистическом тексте процент глаголов составляет 9% . Заменили публицистический текст на технический. Отобрали пять раз по 500 слов. Число глаголов в них составило 43, 42, 44, 41, 43 соответственно. Дают ли эти данные основания утверждать, что технический текст менее насыщен глаголами по сравнению с публицистическим?

539. Процент брака при производстве некоторого изделия на заводе составляет 5%. Высказано подозрение, что ухудшилось качество сырья. Для проверки этого отобрали пять раз по 200 деталей, и бракованных среди них оказалось соответственно 13, 12, 11, 12, 14. Подтверждают ли эти данные высказанное подозрение?

540. Завод выпускает электрические лампочки. В связи с эпидемией для работы были привлечены менее квалифицированные рабочие, поэтому было высказано опасение, что возрастет процент бракованных ламп. Для контроля пять раз отбирали по 500 ламп. Нестандартных среди них оказалось соответственно 80, 79, 78, 78, 81. Подтверждают ли эти данные предположение о квалификации привлеченных рабочих, если ранее нестандартные лампы составляли в среднем 15%?

541. Опрос случайно отобранных 1163 зарегистрированных избирателей показал, что 592 человека планируют проголосовать за определенного кандидата. Значимо ли наблюдаемый процент превышает 50%?

542. При раздаче колоды из 36 карт по шесть карт шести игрокам вероятность того, что при тщательной перетасовке карт игрок получит хотя бы одного туза, равна приближенно 0,53. Игрок А при пяти раздачах карт получил только одного туза. Он высказал подозрение, что карты недостаточно хорошо тасовались. Достаточно ли имеющихся данных для того, чтобы отвергнуть нулевую гипотезу $p = 0{,}53$ на 5%-ном уровне значимости?

Дополнительные задачи к главе 5

К § 5.1. Генеральная совокупность и выборка

543. Исследуется число рыб в закрытом водоеме. С этой целью заброшена сеть с фиксированными размерами ячеек. Все пойманные рыбы помечены и возвращены в водоем. Затем снова заброшена сеть, пойманные рыбы образуют выборку. По числу всех меченых рыб и меченых рыб в выборке судят об общем числе рыб в водоеме. Будет ли эта выборка репрезентативной, если:

а) сеть вторично заброшена сразу после того, как пойманные рыбы возвращены в водоем;

б) сеть вторично заброшена через год после того, как пойманные рыбы возвращены в водоем;

в) вторично заброшена другая сеть?

544. Вам поручено выяснить, сколько учащихся школы хотели бы заниматься в спортивных секциях. Эту информацию нужно получить срочно. Вы организуете выборочный опрос.

а) Как вы извлечете случайную выборку?

б) С помощью таблицы случайных чисел создайте выборку, равную 0,1 числа учащихся школы.

в) Нецелесообразно ли прибегнуть к стратифицированной выборке?

545. Какая из приведенных ниже выборок будет наиболее репрезентативной для совокупности всех учащихся вашей школы?

а) 20 отличников старших классов.

б) Случайная выборка из 20 учащихся, занимающихся в спортивных секциях.

в) 20 так называемых «середнячков», отобранных из разных классов.

г) Случайная выборка 20 учащихся, отобранных из списка всех учащихся школы.

546. Сколько различных выборок без возвращения объемом n можно составить из генеральной совокупности, содержащей N элементов?

547. Вам поручено организовать случайную выборку учащихся школы для проверки их достижений по математике. Вы выяснили процент учащихся школы, растущих в полных и неполных семьях, имеющих и не имеющих дома компьютер, имеющих и не имеющих домашней библиотеки. Эти проценты вы сохранили в выборке. Получена ли таким образом репрезентативная случайная выборка?

К § 5.2. Оценивание параметров

548. Газета провела опрос 983 своих читателей и оказалось, что 39,8% из них готовы проголосовать за данного кандидата. Выборы состоятся через три недели.

а) Какой примерно процент всех жителей ответил бы о готовности проголосовать за данного кандидата, если бы все население было опрошено при тех же условиях?

б) Приведите две причины, почему фактический результат выборов может отличаться от этих 39,8% больше, чем на величину стандартной ошибки?

549. Имеется 8000 деталей. Из них отобрали n штук, среди которых m бракованных. Вероятность детали быть бракованной считается равной p.

а) Оцените вероятность того, что доля нестандартных деталей во всей партии отличается от доли таких деталей в выборке по модулю не более чем на 0,05, если $n = 1000$.

б) Сколько деталей нужно выбрать, чтобы с вероятностью не менее 0,88 средний процент бракованных деталей в выборке отличался от среднего процента бракованных деталей во всей совокупности по модулю не более чем на 0,02?

550. На учете стоят 2000 больных, страдающих хроническими приступами головной боли. Вероятность того, что некоторое лекарство снимет боль, равна p. Для контроля эффективности лекарства отобрали n больных, m из которых лекарство помогло.

а) Оцените вероятность того, что относительная частота события «лекарство снимает боль» отличается от p не более чем на 0,07, если $n = 100$.

б) Сколько больных нужно отобрать для исследования, чтобы с вероятностью не менее 0,74 ожидать, что отклонение относительной частоты от вероятности того, что лекарство снимает боль по модулю не превзойдет 0,04?

551. Из списка 1247 участников круиза случайным образом было опрошено 50 человек. Из них 43 человека заявили, что удовлетворены обслуживанием. Если бы была возможность опросить все 1247 человек, то насколько отличался бы процент довольных обслуживанием от процента удовлетворенных обслуживанием в выборке?

552.* В школах района учится 500 десятиклассников. Для проверки усвоения некоторой темы был предложен тест из 10 заданий. Для испытания случайным образом отобрано 62 учащихся. Результаты испытания представлены в таблице 5.10.

Таблица 5.10

Число правильных ответов	3	4	5	6	7	8	9	10
Число учащихся	2	18	13	8	10	6	4	1

а) Оцените вероятность того, что среднее значение числа правильных ответов во всех школах отличается от среднего числа правильных ответов в выборке не более чем на 0,5.

б) Сколько десятиклассников нужно выбрать для контроля, чтобы с вероятностью не меньшей 0,8 среднее число правильных ответов всех десятиклассников района отличалось по модулю от среднего числа правильных ответов в выборке не более чем на 0,42?

Примите дисперсию генеральной совокупности равной выборочной дисперсии, вычисленной по данным таблицы 5.10.

К § 5.3. Доверительные интервалы

553. Имеется 8000 деталей. Из них отобрали n штук, среди которых m бракованных. Вероятность детали быть бракованной считается равной p. Найдите границы, в которых с вероятностью не менее 0,85 лежит средний процент бракованных деталей во всей партии, если $n = 1000$, $m = 50$.

554. В некотором хозяйстве собрали 10 000 плодов. Биолог случайным образом отобрал n плодов с тем, чтобы обнаружить у них наличие или отсутствие некоторого определенного полезного признака. Оказалось, что m из отобранных n плодов обладают этим признаком. Найдите границы, в которых с вероятностью не меньшей 0,88 лежит процент плодов, обладающих изучаемым признаком во всей совокупности плодов, если $n = 1000$, $m = 120$.

555. Для условий задачи 552 найдите пределы, которые с вероятностью не менее 0,84 содержат среднее число правильных ответов всех десятиклассников района.

556.* На некотором станке изготовлено 2000 валиков. Измерены отклонения диаметра 250 валиков от заданного размера. Результаты измерения приведены в таблице 5.11.

Таблица 5.11

Отклонение диаметра валика от номинала, мк	0—5	5—10	10—15	15—20	20—25
Число валиков	15	75	100	50	10

а) Оцените вероятность того, что среднее отклонение диаметра валика от номинала во всей партии отличается от среднего отклонения диаметра валика в выборке не более чем на 0,5 мк.

б) Постройте 78%-ный доверительный интервал, который содержит среднее отклонение диаметра валика от номинала во всей партии валиков.

в) Сколько валиков необходимо отобрать для их проверки, чтобы с вероятностью не меньшей 0,75 среднее отклонение диаметра валика от номинала во всей партии отличалось от среднего отклонения в выборке не более чем на 0,8 мк?

Примите дисперсию генеральной совокупности равной выборочной дисперсии, вычисленной по данным таблицы 5.11.

557. Из 763 случайно отобранных студентов вуза 152 студента не знакомы с сущностью Болонского процесса создания Единого европейского образовательного процесса.

а) Оцените процент студентов во всем вузе, не знакомых с сущностью Болонского процесса создания Единого европейского образовательного процесса.

б) Вычислите стандартную ошибку оценки, вычисленной в предыдущем задании.

в) Постройте 90%-ный доверительный интервал для процента таких студентов во всем вузе, пользуясь неравенством Чебышёва.

г) Постройте 90%-ный и 95%-ный доверительные интервалы для процента таких студентов во всем вузе, пользуясь нормальным приближением для биномиального распределения.

558.* При проведении общенационального опроса требуется, чтобы границы ошибки не превышали 3% в каждом направлении (т. е. плюс или минус) при доверительной вероятности 0,95.

а) Проверьте выполнение этого требования для ситуации, когда 309 из 1105 зарегистрированных избирателей утверждают, что они поддержат на выборах определенного кандидата, пользуясь нормальным приближением для биномиального распределения.

б) Постройте 95%-ный доверительный интервал для выраженной в процентах доли зарегистрированных избирателей, которые поддержат на выборах определенного кандидата, по данным предыдущего задания, пользуясь нормальным приближением для биномиального распределения.

К § 5.4. Проверка статистических гипотез

559. Имеется 8000 деталей. Вероятность детали быть бракованной считается равной p. Из них отобрали n штук, среди которых m бракованных. Проверяется нулевая гипотеза: $p = 0{,}05$. Для проверки этой гипотезы отобрали $n = 500$ деталей, среди которых $m = 28$ оказались бракованными.

а) Вычислите вероятность того, что при случайном отборе 500 деталей 28 из них будут бракованы (если нулевая гипотеза верна).

б) Примите решение, касающееся нулевой гипотезы.

в) Укажите уровень значимости, на котором гипотеза отвергается или не отвергается.

г) Вычислите вероятность ошибки первого рода.

560. При игре в кости было высказано сомнение в правильности игрального кубика. Для проверки этих сомнений кубик был брошен 300 раз, при этом «тройка» выпала 58 раз.

а) Вычислите вероятность того, что при бросании кубика 300 раз 58 раз выпадет «тройка».

б) Примите решение, касающееся правильности кубика.

в) Укажите уровень значимости, на котором гипотеза отвергается или не отвергается.

г) Вычислите вероятность ошибки первого рода.

561. В главе 1 методами описательной статистики устанавливалось авторство статей, опубликованных в конце XIX века в американской газете «The Federalist», подписанных псевдонимом «Публий». Решение принималось на основе вычисления частоты, с которой встречается предлог «on» в спорных статьях и в статьях А. Гамильтона и Д. Медисона, претендентов на авторство спорных статей. В таблице 5.12 представлено число статей, в которых предлог «on» встречается с той или иной частотой (в расчете на 1000 слов).

Таблица 5.12

Частота предлога «on»	0,4	0,4—0,8	0,8—1,2	1,2—1,6	1,6—2,0	2—3
Статьи Гамильтона			2	3	6	11
Статьи Медисона	41	2	4	1	2	
Спорные статьи	11			1		
Частота предлога «on»	3—4	4—5	5—6	6—7	7—8	
Статьи Гамильтона	11	10	3	1	1	
Статьи Медисона						
Спорные статьи						

а) Сформулируйте нулевую гипотезу.

б) Интуитивно оцените величину вероятности, при которой нулевая гипотеза может быть с уверенностью отвергнута.

в) Примите решение об авторстве спорных статей.

562. Решение об авторстве спорных статей принималось на основе вычисления частоты, с которой встречается предлог «by» в спорных статьях и в статьях А. Гамильтона и Д. Медисона, претендентов на авторство спорных статей. В таблице 5.13 представлено число статей, в которых предлог «by» встречается с той или иной частотой (в расчете на 1000 слов).

Таблица 5.13

Частота предлога «by»	1—3	3—5	5—7	7—9	9—11
Статьи Гамильтона	2	9	12	18	4
Статьи Медисона			5	7	8
Спорные статьи			2	1	2
Частота предлога «by»	11—13	13—15	15—17	17—19	
Статьи Гамильтона	5				
Статьи Медисона	16	6	5	3	
Спорные статьи	4	2	1		

а) Сформулируйте нулевую гипотезу.

б) Интуитивно оцените величину вероятности, при которой нулевая гипотеза может быть с уверенностью отвергнута.

в) Примите решение об авторстве спорных статей.

563. За роман «Тихий Дон» М. А. Шолохов в 1965 г. получил Нобелевскую премию. С 1928 г. в СССР и в эмигрантских кругах курсировали слухи о том, что роман принадлежит не ему. Анонимным советским критиком сообщалось, что произведение написано в 1920 г. умершим от тифа писателем-казаком Ф. Крюковым. Перед принятием решения о присуждении Нобелевской премии шведские ученые провели анализ рассказа «Шагать немедленно», который, безусловно, принадлежал Крюкову, рассказа «Путь и тропа», который, безусловно, принадлежал Шолохову, и романа «Тихий Дон». Вычислялась частота различных слов (словарь) в отрывке из 1000 слов. Данные приведены в таблице 5.14.

а) Сформулируйте нулевую гипотезу.

б) Интуитивно оцените величину вероятности, при которой нулевая гипотеза может быть с уверенностью отвергнута.

в) Примите решение об авторстве «Тихого Дона».

Таблица 5.14

Текст	Число слов	Число различных слов
«Шагать немедленно»	$a + c = 1000$	$a = 589$
«Тихий Дон»	$b + d = 1000$	$b = 646$
«Путь и тропа»	$a_1 + c_1 = 1000$	$a_1 = 656$

К § 5.5 Проверка гипотезы о равенстве среднего генеральной совокупности некоторому заданному значению

564. На заседании совета директоров акционерной компании было высказано предложение о предоставлении работникам компании возможности получать часть прибыли в зависимости от результатов работы. Было опрошено 256 экспертов, которым предложили оценить влияние передачи акций на качество продукции по пятибалльной шкале: −2, −1, 0, 1, 2, где −2 означало «резко отрицательное влияние», а 2 — «сильное положительное влияние». Результаты опроса представлены в таблице 5.15.

Таблица 5.15

Балл	−2	−1	0	1	2
Число экспертов	17	42	67	83	47

Можно ли утверждать, что все эксперты (руководящие работники всех фирм компании) считают высказанное предложение полезным? Сформулируйте нулевую и альтернативную гипотезы и проверьте нулевую гипотезу на уровне значимости 0,05 с использованием:

а) доверительного интервала;

б) соответствующей статистики;

в) неравенства Чебышёва.

565. Средний балл, получаемый учащимися за выполнение некоторого теста в течение нескольких лет, равнялся 80. Ниже приведены оценки за этот тест у 12 случайно отобранных учащихся в нынешнем году:

89, 98, 96, 65, 99, 81, 76, 51, 82, 90, 96, 76.

Значимо ли отличается наблюдаемое среднее значение оценки от средней оценки в предыдущие годы? Ответ обоснуйте.

566. Завод может сбрасывать загрязняющие окружающую среду отходы производства в количестве, не превышающем 125 мг вредного вещества в неделю. Последняя проверка, проведенная экологами, дала следующие результаты недельных загрязнений:

65, 61, 49, 114, 72, 53, 78, 47, 52, 38, 61, 43, 98, 85, 91, 63, 71 мг.

а) Уложился ли завод в норму? Поясните ответ, исходя из односторонней проверки гипотезы на уровне значимости 0,05.

б) Сформулируйте нулевую и альтернативную гипотезы.

в) К какому результату приведет двусторонняя проверка гипотезы при тех же условиях?

567. Для условий задачи 561 решите вопрос об авторстве спорных статей, проверив гипотезы о равенстве средних, приняв в качестве выборочного среднего среднюю частоту встречаемости предлога «on» в спорных статьях, а в качестве среднего и дисперсии генеральной совокупности среднюю частоту и выборочную дисперсию встречаемости предлога «on»:

а) в статьях Гамильтона;

б) в статьях Медисона.

568. Для условий задачи 562 решите вопрос об авторстве спорных статей, проверив гипотезы о равенстве средних, приняв в качестве выборочного среднего среднюю частоту встречаемости предлога «by» в спорных статьях, а в качестве среднего и дисперсии генеральной совокупности среднюю частоту и выборочную дисперсию встречаемости предлога «by»:

а) в статьях Гамильтона;

б) в статьях Медисона.

569. Для условий задачи 563 решите вопрос об авторстве «Тихого Дона», проверив гипотезы о равенстве средних, приняв в качестве выборочного среднего среднюю частоту различных слов в «Тихом Доне», а в качестве среднего и дисперсии генеральной совокупности среднюю частоту и дисперсию числа различных слов:

а) в рассказе Ф. Крюкова;

б) в рассказе М. А. Шолохова.

К § 5.6. Проверка гипотез о биномиальной вероятности

570. Некоторое изделие производится по двум технологиям: старой и новой. Результаты выборочной проверки качества изделий представлены в таблице 5.16.

Таблица 5.16

Числовые характеристики выборки	Старая технология	Новая технология
Средний уровень брака	0,047	0,023
Стандартное отклонение	0,068	0,050
Размер выборки (число дней)	50	44

а) На сколько уменьшится уровень брака при переходе от старой технологии к новой? Какова стандартная ошибка средней разности уровней брака?

б) Проверьте, улучшает ли переход на новую технологию качество изделий.

в) Постройте 95%-ный доверительный интервал для снижения уровня брака.

г) Является ли улучшение качества продукции статистически значимым?

571. В больнице на учете есть больные, страдающие хроническими приступами головной боли. Вероятность того, что некоторое лекарство снимает боль, равна 0,6. Начали применять для снятия головных болей новое лекарство. Для его исследования пять раз отобрали по 20 больных, среди которых новое лекарство сняло боль у 13, 14, 13, 15, 14 больных. Дают ли эти данные основания считать, что новое лекарство более эффективно?

572. На уровне значимости 0,05 проверяется нулевая гипотеза H_0: $p = 0{,}8$ при альтернативной гипотезе H_1: $p = 0{,}9$. Наблюдаемое значение числа «успехов» в 1000 наблюдений равно 830.

а) Какой гипотезе следует отдать предпочтение?

б) Какова вероятность ошибки второго рода?

573. 60-ваттные лампы разложены в коробки по восемь штук в каждой. Каждая коробка проверяется испытанием двух ламп из нее, выбранных наугад. Правило приемки заключается в том, что если обе испытываемые лампы горят, то вся коробка считается принятой; если же не горит хотя бы одна лампа, то коробка бракуется. На контроль поступает коробка, содержащая две неисправные и шесть исправных ламп. Какова вероятность того, что эта коробка окажется принятой?

Ответы и указания к задачам

Глава 1

8. а) Изделие; б) многомерные; в) временной срез; г) номинальная; д) количественная; е) порядковая. **9.** а) Продукт питания; б) многомерные; в) нормы потребления для различных категорий населения; г) временной срез; д) нет. **10.** а) Работник; б) многомерные; в) качественные переменные: образование, пол, возраст, стаж работы; количественная переменная: месячная заработная плата; г) временной срез; порядковая; е) нет. **11.** а) Месячная заработная плата; б) пол, возраст, стаж работы, образование; в) месячная заработная плата, возраст, стаж работы, образование; г) пол, возраст, стаж работы, образование. **12.** а) Одномерные, количественные, временной срез;

б)

x_i	36	37	38	39	40	41	42	43
n_i	1	2	5	8	17	21	18	8
ν_i	0,0125	0,025	0,0625	0,1	0,2125	0,2625	0,225	0,1

в) в каждую группу входит один размер обуви, получим восемь групп.
13. а) Одномерные, количественные, временной срез;

б)

x_i	21,9	20,7	19,2	18,8	18,3	18,2	18,0
n_i	1	1	1	1	1	1	1
ν_i	0,015	0,015	0,015	0,015	0,015	0,015	0,015
x_i	17,9	17,7	17,5	17,3	17,2	17,1	16,9
n_i	1	2	1	2	1	1	1
ν_i	0,015	0,03	0,015	0,03	0,015	0,015	0,015
x_i	16,7	16,6	16,5	16,4	16,3	16,2	16,1
n_i	1	2	1	1	2	2	1
ν_i	0,015	0,03	0,015	0,015	0,03	0,03	0,015

x_i	15,8	15,7	15,5	15,4	15,2	15,1	14,7
n_i	2	1	1	2	2	3	3
ν_i	0,03	0,015	0,015	0,03	0,03	0,044	0,044
x_i	14,6	14,5	14,3	14,2	14,1	14,0	13,9
n_i	1	1	1	1	1	4	1
ν_i	0,015	0,015	0,015	0,015	0,015	0,059	0,015
x_i	13,7	13,5	13,4	13,0	12,7	12,4	12,3
n_i	1	1	1	1	1	1	1
ν_i	0,015	0,015	0,015	0,015	0,015	0,015	0,015
x_i	12,1	11,7	11,6	11,3	10,9	10,7	10,1
n_i	1	1	1	1	1	1	1
ν_i	0,015	0,015	0,015	0,015	0,015	0,015	0,015
x_i	8,8	8,4	8,3				
n_i	1	1	1				
ν_i	0,015	0,015	0,015				

в) сгруппированные данные:

x_i	21,0—22,0	19,9—20,9	18,8—19,8	17,7—18,7	16,6—17,6
n_i	1	1	2	6	9
x_i	15,5—16,5	14,4—15,4	13,3—14,3	12,2—13,2	11,1—12,1
n_i	11	12	11	4	4
x_i	10,0—11,0	8,9—9,9	7,8—8,8		
n_i	3	0	3		

14. а) Одномерные, количественные, временной ряд;

б)

x_i	12,0	12,8	13,0	13,1	13,8	13,9	14,0
n_i	4	1	1	1	1	2	4
x_i	14,2	14,9	15,0	15,8	15,9	16,0	16,8
n_i	1	2	3	1	2	2	1
x_i	16,9	17,0	17,2	17,5	18,0	18,1	18,4
n_i	3	1	1	1	2	1	1
x_i	19,2	19,3	20,0	20,2			
n_i	1	1	1	1			

в) сгруппированные данные:

x_i	11,6—12,2	12,3—12,9	13,0—13,6	13,7—14,3	14,4—15,0
n_i	4	1	2	8	5
x_i	15,1—15,7	15,8—16,4	16,5—17,1	17,2—17,8	17,9—18,5
n_i	0	5	5	2	4
x_i	18,6—19,2	19,3—19,9	20,0—20,6		
n_i	1	1	2		

15. а) Одномерные, количественные, временной срез;

б)

x_i	600	590	580	560	550	540	530	500
n_i	1	1	1	1	1	2	1	3
x_i	480	470	450	440	435	430	420	405
n_i	2	1	4	2	1	1	3	2
x_i	400	385	380	370	360	330	320	315
n_i	1	1	1	2	3	1	1	1
x_i	300	240	225	220	190	180	170	150
n_i	4	1	1	2	1	1	1	1

в) сгруппированные данные:

x_i	580—615	540—575	500—535	460—495	420—455	380—415
n_i	3	4	4	3	11	5
x_i	340—375	300—335	260—295	220—255	180—215	140—175
n_i	5	7	0	4	2	2

16. а) Дискретный вариационный ряд:

x_i	2,9	3,8	4,2	4,4	4,5	4,7	5,1	5,3
n_i	1	1	1	1	1	1	1	2
x_i	5,4	5,6	5,8	6,0	6,1	6,2	6,4	6,5
n_i	2	3	2	4	3	1	1	2
x_i	6,7	6,8	6,9	7,1	7,4	7,7	8,2	8,5
n_i	2	1	1	1	1	1	1	1

б), в) интервальный вариационный ряд:

x_i	2,8—3,2	3,3—3,7	3,8—4,2	4,3—4,7	4,8—5,2	5,3—5,7
n_i	1	0	2	3	1	7
p_i	2	0	4	6	2	14
π_i	$\frac{1}{18}$	0	$\frac{2}{18}$	$\frac{3}{18}$	$\frac{1}{18}$	$\frac{7}{18}$
x_i	5,8—6,2	6,3—6,7	6,8—7,2	7,3—7,7	7,8—8,2	8,3—8,7
n_i	10	5	3	2	1	1
p_i	20	10	6	4	2	2
π_i	$\frac{10}{18}$	$\frac{5}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{1}{18}$	$\frac{1}{18}$

17. а) Дискретный вариационный ряд: 177,7; 182,7; 183,5; 184,8; 187,2; 188,3; 189,9; 190,2; 190,3; 190,8; 191,1; 191,3; 191,7; 192,7; 193,6; 194,3; 195,2; 195,4; 196,1; 196,6; 196,7; 197,4; 197,8; 198,6; 200,2; б), в) интервальный вариационный ряд (число групп уменьшено, так как небольшое число данных):

x_i	177,5—180,7	180,8—184,1	184,2—187,5	187,6—190,9
n_i	1	2	2	5
p_i	0,303	0,606	0,606	1,52
π_i	0,012	0,024	0,024	0,061
x_i	191,0—194,3	194,4—197,7	197,8—201,7	
n_i	6	6	3	
p_i	1,82	1,82	0,909	
π_i	0,073	0,073	0,036	

18. а) Дискретный вариационный ряд:

x_i	1540	1560	1600	1620	1640	1660	1680	1700	1720	1740
n_i	1	3	2	2	4	2	1	2	1	10
x_i	1780	1800	1820	1840	1860	1880	1940	1960	1980	
n_i	2	4	1	7	3	2	1	1	1	

б), в) интервальный вариационный ряд (число групп уменьшено, так как небольшое число данных):

x_i	1520—1565	1565—1610	1610—1655	1655—1700
n_i	4	2	6	5
ν_i	0,08	0,04	0,12	0,10
p_i	0,0889	0,0444	0,133	0,111
π_i	0,00178	0,000889	0,00267	0,00222
x_i	1700—1745	1745—1790	1790—1835	1835—1880
n_i	11	2	5	11
ν_i	0,22	0,04	0,10	0,22
p_i	0,244	0,0444	0,111	0,244
π_i	0,00542	0,000889	0,00222	0,00542
x_i	1880—1925	1925—1970	1970—2015	
n_i	1	2	1	
ν_i	0,02	0,04	0,02	
p_i	0,0222	0,0444	0,0222	
π_i	0,000444	0,000889	0,000444	

19. а) Дискретный вариационный ряд:

x_i	0,27	0,28	0,32	0,33	0,35	0,37	0,38	0,40	0,42	0,43
n_i	2	1	2	1	1	1	1	2	1	2
x_i	0,45	0,47	0,48	0,49	0,51	0,52	0,53	0,54	0,55	0,98
n_i	1	1	1	1	1	2	1	1	1	1

б), в) интервальный вариационный ряд (число групп уменьшено, так как небольшое число данных, резко выделяющееся значение 0,98 исключено из рассмотрения):

x_i	0,235—0,305	0,305—0,375	0,375—0,445	0,445—0,515	0,515—0,585
n_i	3	5	6	5	5
ν_i	0,125	0,208	0,250	0,208	0,208
p_i	43	71	86	71	71
π_i	1,79	2,97	3,57	2,97	2,97

20. Для задачи 16 — рис. 65; для задачи 18 — рис. 66.

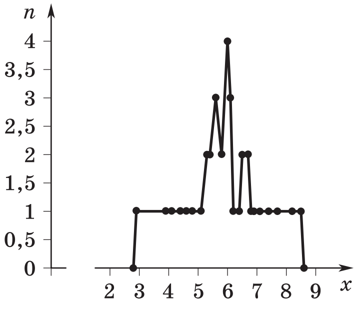


Рис. 65

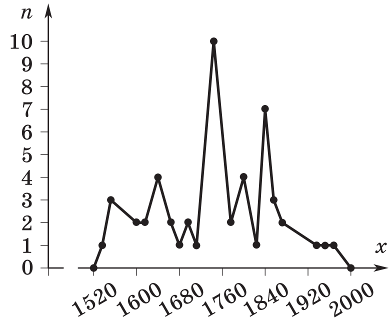


Рис. 66

21. Рис. 67. **22.** а) Рис. 68.

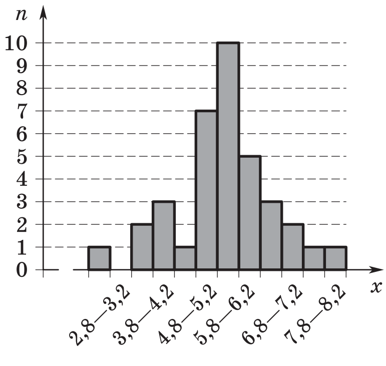


Рис. 67

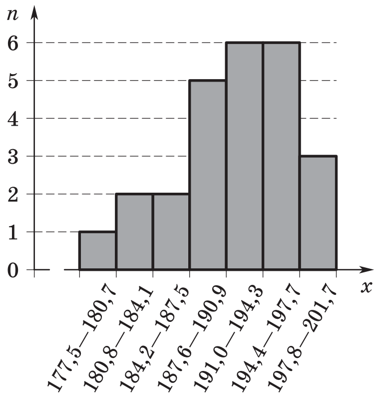


Рис. 68

23. а)

Масса, г	1000—1500	1500—2000	2000—2500
Число детей	84	205	502
ν_i	0,0084	0,0205	0,0502
Накопленная относительная частота	0,0084	0,0289	0,0791

Продолжение табл.

Масса, г	2500—3000	3000—3500	3500—4000
Число детей	1723	3752	2747
ν_i	0,1723	0,3752	0,2747
Накопленная относительная частота	0,2514	0,6266	0,9013
Масса, г	4000—4500	4500—5000	5000—5500
Число детей	852	124	11
ν_i	0,0852	0,0124	0,0011
Накопленная относительная частота	0,9865	0,9989	1,0000

б) рис. 69 и 70.

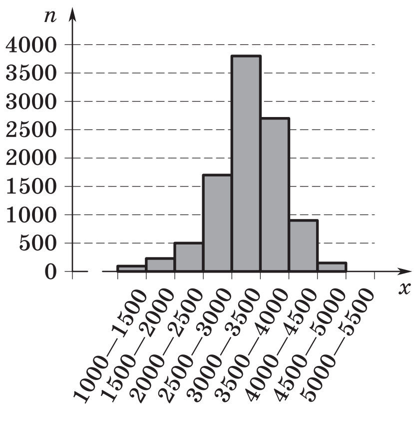


Рис. 69

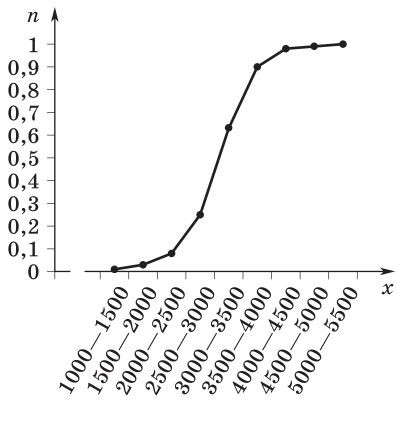


Рис. 70

24. а) Дискретный вариационный ряд:

x_i	39	40	41	42	43	44	46
n_i	2	4	4	6	2	1	1

б) полигон частот — рис. 71. **25.** Рис. 72 и 73. **26.** а) 11,9; б) $\approx$ 274. **27.** а) $\approx$ 11; б) 331. **28.** а) $\approx$ 0,48; 3,30; 0,532; автором спорных статей является, по-видимому, Медисон; б) рис. 74, 75 и 76.

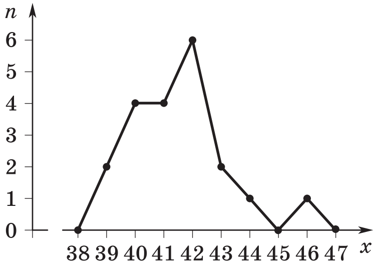


Рис. 71

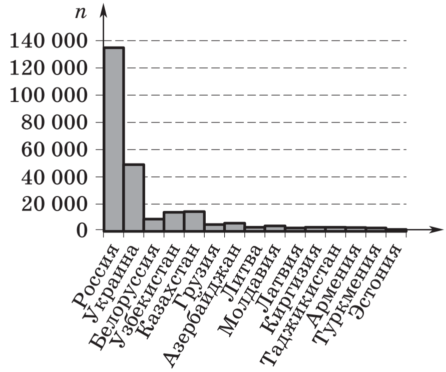


Рис. 72

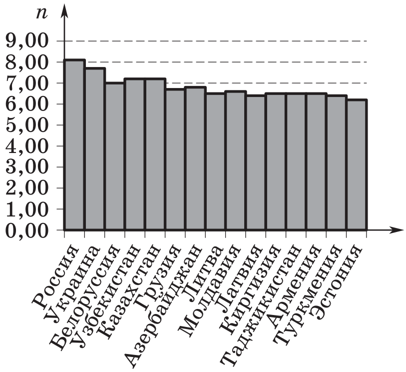


Рис. 73

Статьи Гамильтона

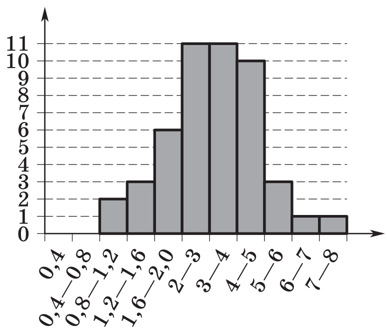


Рис. 74

Статьи Медисона

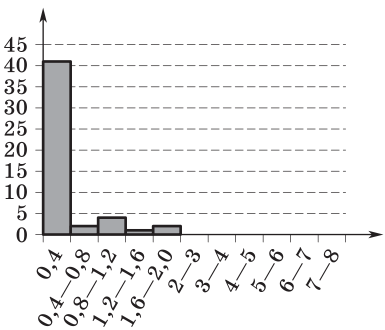


Рис. 75

Спорные статьи

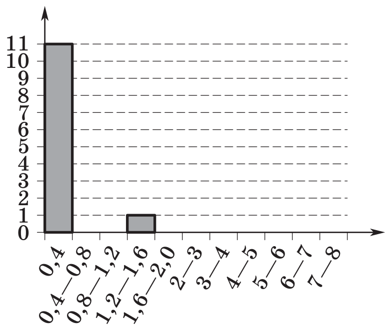


Рис. 76

29. 1-й. **30.** а) 4,2; б) статистика; в) нет. **31.** В первой. **32.** А. **33.** а) Тихоокеанские; б) 28,7 р. **34.** 19,7%. **35.** 76,5. **36.** 5. **37.** ≈ 13,9; ≈ 15,5. **38.** 70,1%. **39.** а) 1084 р.; б) 13 383 064 р.; в) 8469 р.; г) 1580. **40.** ≈ 9,85; ≈ 9,73; рис. 77. **41.** а) 25,25 ц/га; б) ≈ 6,44; ≈ 2,54 ц/га; типичное расстояние от средней урожайности для значений урожайности на отдельных фермах примерно равно 2,54 ц/га; в) 10 ц/га; расстояние между наименьшей и наибольшей урожайностью на фермах равна 10 ц/га; г) ≈ 0,1; урожайность на отдельных фермах отличается от средней урожайности примерно на 10%; д) рис. 78. **42.** а) ≈ 4,68 г/мм^2; б) ≈ 2,1; ≈ 1,4 г/мм^2; в) ≈ 600; г) ≈ 930; д) 1000; е) 7 г/мм^2; ж) ≈ 0,3; з) рис. 79. **43.** а) Увеличилась на 20%; б) не изменилось; в) не изменился; г) уменьшился на 16,7%. **44.** а) 7,04; 41; б) 2,2; 7; в) обе характеристики чувствительны к исключению крайних значений. **45.** а) 1,12 кг; б) 5 кг; в) на 1,6 величины стандартного отклонения; г) на 1,96 величины стандартного отклонения. **46.** а) Увеличилась на 10%; б) не изменилось; в) не изменился; г) уменьшился на 9%. **47.** а) 78,7 т; б) 80,0 т; в) рис. 80; г) $z_{0,25} = 70{,}1$; $z_{0,75} = 91{,}2$; д) $z_{0,1} = 59{,}2$; $z_{0,9} = 98{,}4$; е) не соответствует; с 10-м.

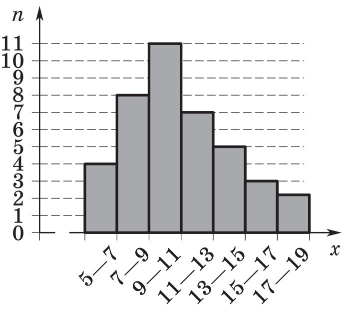


Рис. 77

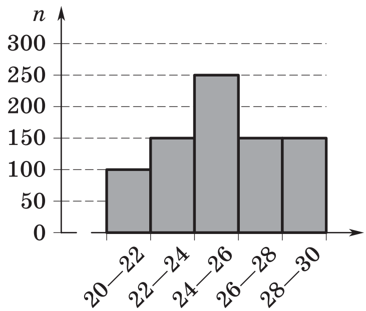


Рис. 78

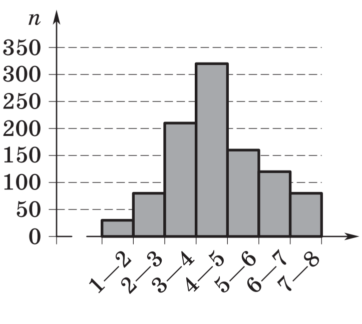


Рис. 79

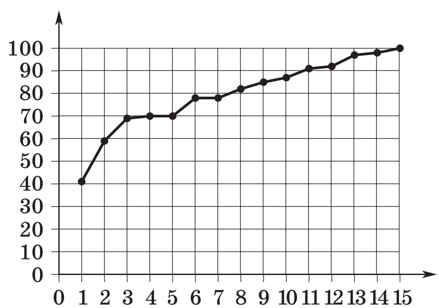


Рис. 80

48. а) 188 см; б) 184,5 см; в) рис. 81; г) 180; 185; 188,75; д) 179; 190; е) почти соответствуют; с 20-м. **49.** а) День; б) многомерные; в) временной ряд; г) номинальная; д) количественная; е) номинальная; ж) порядковая; з) дате — порядковая; температурам, давлению, влажности воздуха, скорости ветра, времени — шкалы интервалов; направлению ветра, влиянию на здоровье — номинальная. **50.** а) Футбольная команда; б) многомерные; в) временной срез; г) порядковая; д) количественная; е) номинальная; ж) количественная; з) месту — порядковая; названию команды — номинальная; количеству игр, выигрышей, ничьих, поражений, забитых и пропущенных мячей, очков — шкалы интервалов; нет; и) например, какие команды претендуют на призовые места, каким командам угрожает опасность покинуть высшую лигу, у каких команд есть проблема с организацией защиты ворот и т. д. **53.** а) Одномерные, количественные, временной срез;

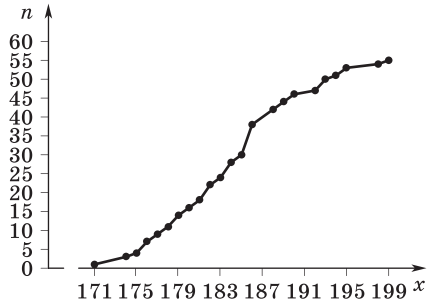


Рис. 81

б)

Возраст	22	24	28	30	31	32	35
n_i	1	1	2	1	2	2	1
ν_i	0,04	0,04	0,08	0,04	0,08	0,08	0,04
Накопленная частота	1	2	4	5	7	9	10
Относительная накопленная частота	0,04	0,08	0,16	0,20	0,28	0,36	0,40
Возраст	36	37	38	41	42	44	45
n_i	1	1	1	1	2	2	1
ν_i	0,04	0,04	0,04	0,04	0,08	0,08	0,04
Накопленная частота	11	12	13	14	16	18	19
Относительная накопленная частота	0,44	0,48	0,52	0,56	0,64	0,72	0,76

Продолжение табл.

Возраст	46	47	49	50	56
n_i	1	2	1	1	1
ν_i	0,04	0,08	0,04	0,04	0,04
Накопленная частота	20	22	23	24	25
Относительная накопленная частота	0,80	0,88	0,92	0,96	1,00

в) сгруппированные данные (число групп уменьшено, так как небольшое число данных):

Возраст	21—25	26—30	31—35	36—40
n_i	2	3	5	3
ν_i	0,08	0,12	0,20	0,12
Накопленная частота	2	5	10	13
Относительная накопленная частота	0,08	0,20	0,40	0,52
Возраст	41—45	46—50	51—55	56—60
n_i	6	5	0	1
ν_i	0,24	0,20	0	0,04
Накопленная частота	19	24	24	25
Относительная накопленная частота	0,76	0,96	0,96	1,00

г) например, что большинство сотрудников среднего возраста, пенсионного возраста — не более одного сотрудника, желательно привлечение молодежи и т. д. **60.** а) Дискретный вариационный ряд:

x_i	83	86	87	89	91	92	94	95	96
n_i	1	1	1	1	1	3	1	2	1
x_i	97	98	100	101	102	103	104	105	106
n_i	2	1	1	3	3	1	2	5	3
x_i	107	108	109	110	111	112	113	114	115
n_i	4	3	1	3	1	1	2	2	1

Продолжение табл.

x_i	116	118	121	123	124	127	129	130	131
n_i	1	2	2	2	2	2	2	1	1
x_i	132	133	134	135	138	139	141	146	148
n_i	1	1	1	1	1	1	1	1	1

б, в) интервальный вариационный ряд:

x_i	83—87	88—92	93—97	98—102	103—107
n_i	3	5	6	8	15
ν_i	0,04	0,07	0,08	0,11	0,20
p_i	0,6	1	1,2	1,6	3
π_i	0,008	0,014	0,016	0,022	0,008
x_i	108—112	113—117	118—122	123—127	128—132
n_i	9	6	4	6	5
ν_i	0,12	0,08	0,05	0,08	0,07
p_i	1,8	1,2	0,8	1,2	1
π_i	0,024	0,016	0,01	0,016	0,014
x_i	133—137	138—142	143—147	148—152	
n_i	3	3	1	1	
ν_i	0,04	0,04	0,01	0,01	
p_i	0,6	0,6	0,2	0,2	
π_i	0,008	0,008	0,002	0,002	

г) например, наиболее распространенный, наименьший, наибольший уровень интеллектуальности и т. д. **64.** а) Дискретный вариационный ряд: 83,7; 86,2; 91,4; 92,3; 101,5; 103,9; 110,8; 113,1; 122,0; 124,0; 127,6; 128,0; 130,1; б) рис. 82 и 83; в) распределение данных — равномерное.

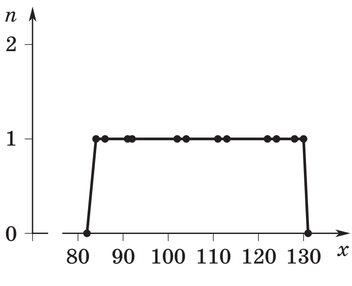


Рис. 82

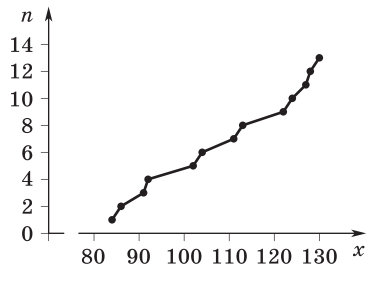


Рис. 83

66. а) Зависимость примерно линейная, рис. 84; б) 5,5 г и 49,6; примерно отражают. **71.** а) 3,2; б) 3,5; в) 4; г) медиана. **73.** *Указание*: воспользуйтесь свойством медианы распределения. **74.** а) 0,84; в среднем размер платы за ссуду в различных фирмах отличается от средней платы за ссуду во всех фирмах на 0,84%; б) 2%; разность между наименьшей и наибольшей платой за ссуду в фирмах составляет 2%; в) 0,26; коэффициент вариации показывает, что плата за ссуду у отдельных фирм отличается от средней платы в среднем на 26%; г) 11; д) все 14. **78.** а) 3,2; б) 3,5; в) рис. 85; г) 2,625; 3,5; 4; д) 2,5 и 4.

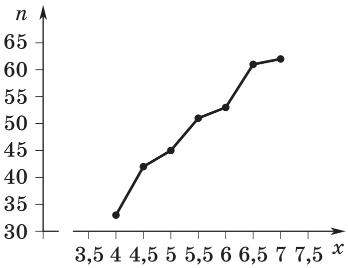


Рис. 84

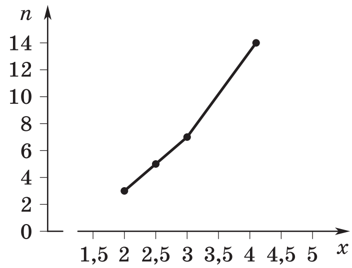


Рис. 85

Глава 2

79. а) 0,975; б) 0,025; 25. **80.** а) 499; б) 1500. **81.** 40. **82.** 60. **87.** а) В июне погода была, преимущественно, пасмурной; б) сентябрь был довольно дождливым; в) предприятие выпускает парашюты очень низкого качества; г) рентген не является очень надежным средством обнаружения этого заболевания. **89.** а) Прогнозирование пары данных: число произведенных за месяц устройств и число дефектных из них; б) множество пар целых неотрицательных чисел (x_i, y_i), где x_i может принимать значения, скажем, от 1 до 10 000, а y_i — от 0 до тех же 10 000; в) результаты эксперимента могут свидетельствовать, например, о том, насколько хорошо начальник цеха знает свое производство; г) $\{(x_i, y_i)\}$, где $x_i \geqslant 50$, $y_i = 0$ или 1; д) 0,88. **90.** а) $\frac{3}{11}$; б) $\frac{8}{11}$. **91.** а) 0,25; б) $\frac{3}{28}$. **92.** а) 0,2; б) $\frac{7}{15}$; в) $\frac{8}{15}$; г) 0,2; д) 0,4; е) 0,4. **93.** а) 0,75; б) 0,25. **94.** 0,65. **95.** а) 0,15; $\approx 0{,}17$; б) $\left|p - \frac{m}{n}\right| \approx 0{,}02$. **96.** а) $\frac{5}{9}$; б) $\frac{4}{9}$; в) $\frac{1}{3}$; г) 0; д) $\frac{5}{9}$; е) $\frac{4}{9}$. **97.** а) $\frac{1}{4}$; б) $\frac{1}{9}$; в) $\frac{1}{2}$; г) $\frac{4}{9}$. **98.** а) $\frac{1}{4}$; б) $\frac{1}{4}$; в) $\frac{3}{4}$. **99.** а) 0,2; б) 0,3; в) 0,5; г) 0,5. **100.** а) $\frac{1}{36}$; б) $\frac{1}{18}$; в) $\frac{1}{12}$; г) $\frac{1}{6}$; д) $\frac{1}{36}$. **101.** 2 и 4. **102.** 50, 20 и 50. **103.** Исходы опыта неравновозможны. **104.** $\frac{5}{12}$; $\frac{1}{2}$; $\frac{1}{12}$; на 1.

105. а) $\frac{1}{6}$; б) $\frac{5}{6}$; в) $\frac{2}{9}$; г) $\frac{5}{9}$. **106.** а) {ГГ, ГЦ, ЦГ, ЦЦ}; б) {0; 1; 2} (указано число выпавших гербов); в) обе монеты упали одинаково. **107.** а) 0,4; б) 0,38. **108.** У Миши. **109.** $\frac{5}{9}$. **113.** $U = \{12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43\}$; $p_i = \frac{1}{12}$. **114.** а) $U = \{К_1С_1, К_1С_2, К_2С_1, К_2С_2\}$; б) $U = \{К_1С_1, К_2С_1\}$, где, например, $К_1С_2$ означает, что красный шар — в первом ящике, а синий — во втором. **115.** $p_1 = \ldots = p_9 = \frac{1}{14}$, $p_{10} = \frac{5}{14}$. **116.** $U = \{123, 124, 125, 134, 135, 145, 234, 235, 245, 345\}$, где 1, 2, 3, 4, 5 — номера шаров. **117.** $U = \{0, 1, 2\}$; $P(0) = 0{,}1$; $P(1) = 0{,}4$; $P(2) = 0{,}5$. **118.** $P(\text{Г}) = \frac{2}{3}$; $P(\text{Ц}) = \frac{1}{3}$. **119.** а), б), г). **120.** Например, U = {Г1, Г2, Г3, Г4, Г5, Г6, Ц1, Ц2, Ц3, Ц4, Ц5, Ц6}; U = {ГЧ, ГН, ЦЧ, ЦН}. **121.** а) U = {ГГ1, ГГ2, ГГ3, ГГ4, ГГ5, ГГ6, ГЦ1, ГЦ2, ГЦ3, ГЦ4, ГЦ5, ГЦ6, ЦГ1, ЦГ2, ЦГ3, ЦГ4, ЦГ5, ЦГ6, ЦЦ1, ЦЦ2, ЦЦ3, ЦЦ4, ЦЦ5, ЦЦ6}; б) U = {01, 02, 03, 04, 05, 06, 11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26}, где на первом месте стоит число гербов, выпавших при подбрасывании двух монет; в) U = {О1, О2, О3, О4, О5, О6, Н1, Н2, Н3, Н4, Н5, Н6}, где О означает, что монеты упали одинаково, Н — неодинаково; г) U = {ГГЧ, ГГН, ГЦЧ, ГЦН, ЦГЧ, ЦГН, ЦЦЧ, ЦЦН}, где Ч означает, что на верхней грани кубика выпало четное число очков, Н — нечетное; д) U = {0Ч, 0Н, 1Ч, 1Н, 2Ч, 2Н}, где на первом месте стоит число гербов, выпавших при подбрасывании двух монет, Ч означает, что на верхней грани кубика выпало четное число очков, Н — нечетное; е) U = {ОЧ, ОН, НЧ, НН}, где О означает, что монеты упали одинаково, Н на первом месте — неодинаково, Ч означает, что на верхней грани кубика выпало четное число очков, Н на втором месте — нечетное. **122.** Например, U = = {ББ, БК, БС, КБ, КК, КС, СБ, СК, СС}, где буквы Б, К, С означают, что вынут соответственно белый, красный, синий шар; U = {2Б, 1Б1К, 1Б1С, 2К, 1К1С, 2С}, где 1Б1С означает, что извлечен один белый шар, один синий. **123.** U = {12, 112, 111, 21, 221, 222}, где, например, 112 означает, что первые две купленные шоколадки содержат фотографию первого спортсмена, а третья – фотографию второго. **124.** а) U = {2, 4, 6, 8}, где числа обозначают длину отрезка, на которую попала игла; $p_i = \frac{1}{4}$; б) U = {1, 2, 3, 4}, где числа обозначают место, занятое лошадью в забеге; нет данных для введения элементарных вероятностей; в) $U = \{ab, ac, bc\}$; $p_i = \frac{1}{3}$; г) U = {да, нет, не знаю}; нет данных для введения элементарных вероятностей; д) U = {О, У, Ч, П, Пр}, где О означает, что остановка станка произошла из-за обрыва нитей основы, У — из-за обрыва нитей утка, Ч — из-за смены человека, П — из-за поломки погонялок, Пр — из-за прочих причин; $P(\text{О}) = 0{,}22$, $P(\text{У}) = 0{,}31$, $P(\text{Ч}) = 0{,}27$, $P(\text{П}) = 0{,}03$, $P(\text{Пр}) = 0{,}17$. **125.** 0,1; 0,15; 0,2; 0,25; 0,3. **126.** U = {11, 10, 01, 00}, где, например, 10 означает, что опрошенный положительно ответил на первый вопрос и отрицательно — на второй;

$P(11) = 0{,}4$; $P(10) = 0{,}2$; $P(01) = 0{,}3$; $P(00) = 0{,}1$; а) 0,6; б) 0,7. **127.** $U = \{abc, abd, acd, bcd\}$; а) $A = \{abc, abd, acd\}$; $P(A) = 0{,}75$; б) $B = \{abc, abd\}$; $P(B) = 0{,}5$; в) $C = \{abc, abd, acd, bcd\}$; $P(C) = 1$. **128.** а) См. таблицу 2.14; б) $U = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, где исходами являются суммы выпавших очков; в) например, «на первом кубике выпало больше очков, чем на втором». **129.** а) $U =$ = {Г1, Г2, Г3, Г4, Г5, Г6, Ц1, Ц2, Ц3, Ц4, Ц5, Ц6}; б) $U =$ {Г1, Ц1, ГП, ГС, ЦП, ЦС}, где, например, ГП означает, что монета упала гербом вверх, а на верхней грани игрального кубика выпало простое число очков; Г1 — монета упала гербом вверх, а на верхней грани игрального кубика выпало одно очко; ГС — монета упала гербом вверх, а на верхней грани игрального кубика выпало составное число очков; в) например, «монета упала гербом вверх, а на верхней грани игрального кубика выпало четное число очков». **130.** а) $U =$ {Б, Ч, Т, П}, где исходами опыта служат названия мастей карт: бубны, червы, трефы, пики; б) $U =$ {6, 7, 8, 9, 10, В, Д, К, Т}, где исходами опыта служат названия карт: шестерка, семерка, ..., десятка, валет, ..., туз; в) $U =$ {Б6, Б7, ..., БТ, Ч6, Ч7, ..., ЧТ, Т6, Т7, ..., ТТ, П6, П7, ..., ПТ}, где первый символ означает название масти, второй — достоинство карты; г) $U = \{К, \overline{К}\}$, где К означает, что извлечена картинка, $\overline{К}$ — извлечена карта, не являющаяся картинкой. **131.** а) См. ответ к задаче **116**; б) A — «белых шаров извлечено больше, чем «черных» (белые шары обозначены числами 1, 2, 3, черные — 4, 5), $A = \{123, 124, 125, 134, 135, 234, 235\}$, B — «все извлеченные шары — белые», $B = \{123\}$; в) $p_i = 0{,}1$; г) 0,6. **132.** а) См. ответ к задаче 123; б) A — «покупки двух шоколадок достаточно для получения приза», $A =$ $= \{12, 21\}$, B — «получен приз», $B = \{12, 21, 112, 221\}$; в) $p(12) = p(21) =$ $= 0{,}25$; $p(112) = p(221) = p(111) = p(222) = 0{,}125$; г) 0,75. **133.** а) 0,91 и 0,84; б) 0,08 и 0,12, 0,02 и 0,04, 0,01 и 0. **134.** а) Выйдут из строя оба элемента; б) по крайней мере один элемент работает; в) из строя вышел только первый элемент; г) по крайней мере один элемент вышел из строя. **135.** $C =$ $= A \cap (B_1 \cup B_2)$. **136.** $\overline{C} = (\overline{A_1} \cap \overline{A_2}) \cap (\overline{B_1} \cap \overline{B_2})$; $C = (A_1 \cup A_2) \cup (B_1 \cap B_2)$. **137.** а) $A_1 = B_1 + B_2$; $A_2 = B_1 + B_3$; б) $B_1 = A_1 \cdot A_2$; $B_2 = A_1 \cdot \overline{A_2}$; $B_3 = \overline{A_1} \cdot A_2$; $B_4 = \overline{A_1} \cdot \overline{A_2}$; в) попадание имело место только при одном выстреле; невозможное событие; попадание имело место при обоих выстрелах. **138.** 5 : 4. **139.** а) $U =$ {(1812, 1825, 1861), (1812, 1861, 1825), (1825, 1812, 1861), (1825, 1861, 1812), (1861, 1825, 1812), (1861, 1812, 1825)}, где в круглых скобках на первом месте стоит названная учащимся дата первой Отечественной войны, на втором — дата восстания декабристов, на третьем — дата отмены крепостного права; б) $p_i = \frac{1}{6}$; в) 1 : 2. **140.** а) $U =$ {(м, д, п), (м, п, д), (д, м, п), (д, п, м), (п, д, м), (п, м, д)}, где в круглых скобках на первом месте стоит обозначение конверта, куда попало письмо маме, на втором — дедушке, на третьем — подруге; м — конверт с маминым адресом, д — с адресом дедушки, п — с адресом подруги; б) $p_i = \frac{1}{6}$; в) 1 : 1. **141.** а) $U =$ {ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ}; б) $p_i = \frac{1}{8}$; в) 1 : 3. **142.** 0,56. **143.** а) 0,8; б) 0,7. **144.** ≈ 0,17. **145.** $\frac{2}{3}$. **146.** 0,5. **147.** $\frac{1}{2}$; $\frac{1}{3}$; $\frac{1}{6}$. **148.** а) $\frac{5}{36}$; б) $\frac{17}{18}$. **149.** 0,5. **150.** 0,6. **151.** а) 0,5; б) 0,6; в) 0,6. **152.** а) 0,9; б) нет. **153.** а) 0,8; б) 0,2. **154.** а) $\frac{2}{3}$; б) $\frac{7}{9}$. **155.** а) 0,75; б) 0,5. **156.** а) $\frac{2}{21}$; б) например, U = {ББ, БЧ, ЧБ, ЧЧ}; p(ББ) = $\frac{3}{7}$; p(БЧ) = p(ЧБ) = $\frac{10}{21}$; p(ЧЧ) = $\frac{2}{21}$. **157.** а) $\frac{7}{15}$; б) $\frac{1}{15}$. **158.** а) 0,3; б) 0,1. **159.** 0,24. **160.** 1) $\frac{9}{24}$; б) $\frac{1}{8}$. **161.** а) 0,941; б) 0,019; в) 0,002; г) 0,038. **162.** Если событие A означает, что у родителей темные глаза, а событие B — у детей темные глаза, то $P(B \mid A) \approx 0{,}388$; $P(B \mid \bar{A}) \approx 0{,}102$; $P(\bar{B} \mid A) \approx 0{,}612$; $P(\bar{B} \mid \bar{A}) \approx 0{,}898$. **163.** а) $\frac{1}{15}$; б) $\frac{1}{2}$; в) $\frac{14}{15}$. **164.** а) 0,016; б) 0,02; в) 0,036. **165.** а) $\frac{2}{3}$; б) $\frac{7}{9}$; 3) $\frac{1}{15}$; г) $\frac{7}{15}$; д) 0,7; е) $\frac{2}{3}$. **166.** а) 0,63; б) 0,69; в) $\frac{63}{69}$. **167.** а) $\frac{3}{14}$; б) $\frac{1}{28}$. **168.** $\frac{5}{144}$. **169.** а) 0,99; б) 0,999; в) $1 - (0{,}1)^n$. **170.** а) 0,972; б) ≈ 0,9778; в) $\approx 0{,}978 \cdot (1 - (0{,}08)^n)$. **171.** а) Да; б) нет. **172.** Да. **173.** а) ≈ 92,2%; б) 0,005%. **174.** а) Нет; б) нет. **175.** а) 0,32; б) 0,48; в) 0,08; г) 0,44; д) 0,92. **176.** а) 0,046; б) 0,004; в) 0,874; г) 0,122; д) 0,126. **177.** а) ≈ 0; б) ≈ 1. **178.** $\frac{13}{30}$. **179.** 0,0345. **180.** 0,78. **181.** 0,0175. **182.** 0,66. **185.** Наименьшую — 2 и 12, наибольшую — 7. **186.** 0,2. **187.** $\frac{1}{8}$. **188.** 6. **189.** а) 0,1; б) 0,6; в) 0,7; г) 0,3. **190.** а) $\frac{1}{37}$; б) $\frac{18}{37}$; в) $\frac{12}{37}$; г) $\frac{19}{37}$. **191.** а) Субъективная вероятность; б) классическая вероятность; в) статистическая вероятность; г) субъективная вероятность. **192.** $p_1 = \ldots = p_{n-1} = \frac{1}{n+r}$; $p_n = \frac{r+1}{n+r}$. **193.** U = {Ю, А, В, И}, где буквами обозначены первые буквы имен друзей, кому по жребию может достаться лодка; а) p_i = 0,25; б) p(Ю) = 0,1; p(А) = 0,2; p(В) = 0,3; p(И) = 0,4. **194.** Например, $U = \{T,\ \bar{T}\,T,\ \bar{T}\,\bar{T}\,T,\ \bar{T}\,\bar{T}\,\bar{T}\,T,\ \ldots,\ \underbrace{\bar{T}\,\bar{T}\ldots\bar{T}}_{32}\,T\}$, где T означает, что извлечен туз, а $\bar{T}$ — не туз. **195.** а) U = {1234, 1243, 1324, 1342, 1423, 1432, 2134, 2143, 2314, 2341, 2413, 2431, 3214, 3241, 3124, 3142, 3421, 3412, 4231, 4213, 4321, 4312, 4123, 4132}, где числами 1, 2, 3, 4 пронумерованы карточки; б) $p_i = \frac{1}{24}$. **196.** а) U = {1111, 1110, 1101, 1100, 1011, 1010, 1001, 1000, 0111, 0110, 0101, 0100, 0011, 0010, 0001, 0000}, где числами 0 и 1 обозначены события «покупатель не подошел к прилавку» и «покупатель подошел к прилавку»; б) $\frac{1}{16}$. **197.** а) U = {дм, да, ам, аа, мм, ма}; p(дм) = 0,125; p(да) = 0,125; p(ам) = 0,25; p(аа) = 0,25; p(мм) = 0,125;

p(ма) = 0,125; б) 0,375. **198.** а) U = {МД, М$\overline{Д}$, ЖД, Ж$\overline{Д}$}; б) p(МД) = 0,125; p(М$\overline{Д}$) = 0,375; p(ЖД) = 0,00125; p(Ж$\overline{Д}$) = 0,49875. Здесь буквами М и Ж обозначен пол новорожденного — мужской и женский, буквой Д — новорожденный страдает дальтонизмом, символом $\overline{Д}$ — не страдает дальтонизмом; в) 0,125; 0,00125; 0,12625. **199.** а) U = {1, 21, 221, 222}, где, например, символ 21 означает, что при доигрывании первую партию выиграл второй игрок, а вторую — первый; б) p(1) = 0,5; p(21) = 0,25; p(221) = = 0,125; p(222) = 0,125; в) A = {1, 21, 221}, B = {222}, C = {21, 221, 222}; г) $P(A)$ = 0,875; $P(B)$ = 0,125; $P(C)$ = 0,5; д) 7 : 1. *Указание*: приз делится пропорционально вероятностям выигрышей игроков при доигрывании. **200.** а) U = {11, 1211, 2111, 222, 1212, 1222, 1221, 2112, 2122, 2121, 2211, 2212}, где, например, символ 1211 означает, что в первой, третьей и четвертой партиях выиграл первый игрок, а во второй — второй; б) p(11) = 0,25; p(222) = 0,125; p(1211) = p(2111) = p(1212) = p(1222) = p(1212) = p(2112) = = p(2122) = p(2121) = p(2211) = p(2212) = 0,0625; в) A = {222}, B = {11, 1211, 2111}, C = {11, 222}; г) $P(A)$ = 0,125; $P(B)$ = 0,375; $P(C)$ = 0,375. **201.** U = = {пппп, пппл, пплп, пплл, плпп, плпл, пллп, пллл, лппп, лппл, лплп, лплл, ллпп, ллпл, ллпп, лллл}, где, например, символ «пппл» означает, что первые 3 с частица двигалась вправо, а в четвертую — влево; б) $p_i = \frac{1}{16}$; в) A = {пппл, пплп, плпп, пллл, лппп, лплл, ллпп, ллпл}, B = {пппл, пплп, плпп, лппп}, C = {пплл, плпл, пллп, лппл, лплп, ллпп}; г) $P(A)$ = 0,5; $P(B)$ = = 0,25; $P(C)$ = 0,375. **202.** а) $A + B$ — «попадание по крайней мере при одном выстреле», $A \cdot B$ — «попадание при двух выстрелах»; б) $A + B$ — «выпало не более двух очков», $A \cdot B = V$; в) $A + B$ — «по крайней мере один билет выигрышный»; $A \cdot B$ — «оба билета выигрышные». **203.** а) $B = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$; б) $C = A_1 + A_2 + A_3$; в) $D = A_1 \cdot A_2 \cdot A_3$; г) $E = A_1 \cdot \overline{A_2} \cdot \overline{A_3}$; д) $F = A_1 \cdot A_2 \cdot \overline{A_3}$; е) $G = A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3$; ж) $H = A_1 \cdot A_2 \cdot \overline{A_3} + A_1 \cdot \overline{A_2} \cdot A_3 + \overline{A_1} \cdot A_2 \cdot A_3$. **204.** а) $A_1 = B_1 + B_2$; $A_2 = B_1 + B_3$; $A_3 = B_2 + B_3$; $A_4 = B_1 + B_2 + B_3$; б) $B_1 = A_1 \cdot A_2$, $B_2 = A_1 \cdot \overline{A_2}$, $B_3 = \overline{A_1} \cdot A_2$; $B_4 = \overline{A_1} \cdot \overline{A_2}$; в) работает ровно один элемент; невозможное событие; работают оба элемента. **205.** а) 1 : 2; б) 2 : 1. **206.** 1 : 1. **207.** а) 1 : 1; б) 1 : 5; в) 25 : 11. **208.** 4 : 1. **209.** 0,25; 0,75. **210.** а) U = {ТКД, ТДК, КТД, КДТ, ДТК, ДКТ}, где, например, символ ТКД означает, что верхняя карта — туз, следующая — король и самая нижняя — дама; б) $p_i = \frac{1}{6}$; в) 2 : 1; г) честное пари в пользу этого события заключено быть не может, так как это событие невозможно. **211.** а) U = {3, 4, 5, 6, 7}; б) $p(3) = \frac{1}{6}$, $p(4) = \frac{1}{6}$, $p(5) = \frac{1}{3}$, $p(6) = \frac{1}{6}$, $p(7) = \frac{1}{6}$; в) 1 : 2; г) 2 : 1. **212.** $\frac{12}{21}$. **213.** 0,7. **214.** 7 : 5; **215.** а) 0,92; б) 0,08.

216. а) $\frac{41}{81}$; б) $\frac{40}{81}$. **217.** 10%. *Указание*: воспользуйтесь тем, что вероятность суммы событий не превышает суммы вероятностей этих событий. **218.** 0,26. **219.** Урну. **220.** а) 0,14; б) 0,46; в) $\frac{7}{23}$. **221.** а) 0,925; 0,625; б) $\frac{1}{37}$; 0,2. **222.** а) 0,025; б) 0,425; в) $\frac{1}{17}$. **223.** а) 0,22; б) 0,4; в) 0,45. **224.** $0{,}00624\%$; $0{,}00312\%$. **225.** 3%. **226.** Целесообразно изменить. **227.** Двумя одновременно. **228.** Нет. **229.** Нет. **230.** II. **231.** II. **232.** а) $\frac{1}{3}$ и $\frac{2}{3}$; б) $U = \{11, 121, 1221, 1222, 211, 2121, 2122, 2211, 2212, 222\}$, где, например, символ 211 означает, что при доигрывании первую партию выиграл второй игрок, а вторую и третью — первый; в) $p(11) = \frac{1}{9}$; $p(121) = \frac{2}{27}$; $p(1221) = \frac{4}{81}$; $p(1222) = \frac{8}{81}$; $p(211) = \frac{2}{27}$; $p(2121) = \frac{4}{81}$; $p(2122) = \frac{8}{81}$; $p(2211) = \frac{4}{81}$; $p(2212) = \frac{8}{81}$; $p(222) = \frac{8}{27}$; г) 11 : 16; д) 11 : 16; е) 11 : 16. **233.** а) $\frac{2}{5}$ и $\frac{3}{5}$; б) $U = \{1, 211, 2211, 222, 2212, 2122, 2121\}$, где, например, символ 211 означает, что во второй и третьей партиях выиграл первый игрок, а в первой — второй; в) $p(1) = \frac{2}{5}$; $p(211) = \frac{12}{125}$; $p(2211) = p(2121) = \frac{36}{625}$; $p(222) = \frac{27}{125}$; $p(2212) = p(2122) = \frac{54}{625}$; г) 243 : 382; д) 243 : 382. **234.** $\approx 0{,}574$. **235.** Число амеб к концу второго промежутка может равняться 0, 1, 2, 3, 4 с вероятностями $\frac{11}{32}$, $\frac{4}{32}$, $\frac{9}{32}$, $\frac{4}{32}$, $\frac{4}{32}$. **236.** Вероятности одинаковы.

Глава 3

237. а) 16; б) 12; в) 8; г) 6; д) 4; е) 2; ж) 12. **238.** а) 16; б) 12; в) 10; г) 6. **239.** а) 16; б) 5; в) 3. **240.** а) 6; б) 12. **241.** а) 20; б) 10. **242.** а) 9; б) 4. **243.** а) 6; б) 12. **244.** а) $\frac{1}{3}$; б) $\frac{2}{3}$. **245.** 6 375 600. **246.** а) 256; б) 24; в) 16; г) 12. **247.** 780. **248.** 3612. **249.** 28 125. **250.** а) 3 113 510 400; б) 119 750 400. **251.** 260. **252.** а) 330; б) 350; в) 200. **253.** а) $\frac{1}{7^5}$; б) $\frac{1}{7^4}$; в) $\frac{360}{7^4}$. **254.** а) 1000; б) 720; в) 280. **255.** а) 1560; б) 96. **256.** Тех, в записи которых есть 1. **257.** 217 945 728 000. **258.** 60. **259.** а) 10 720; б) 4320. **260.** а) 4845; б) 116 280; в) 77 520. **261.** а) 15; б) 14; в) 13. **262.** а) 378; б) 147. **263.** а) 376 992; б) 1; в) 169 911; г) 4650; д) 4806. **264.** 43 758. **265.** а) 84; б) 120. **266.** 45. **267.** а) 91; б) 55; в) 10. **268.** а) 45; б) 21. **269.** 56. **270.** 21; 7; 455; 1225. **271.** а) 210; б) 5040. **272.** а) $\frac{7}{15}$; б) $\frac{1}{15}$; в) $\frac{7}{15}$. **273.** $\approx 0{,}33$. **274.** $\approx 0{,}77$. **275.** $\frac{1}{55}$. **276.** 0,3. **277.** Если $m > 365$, то 1; если $m \leqslant 365$, то $\frac{365 \cdot 364 \cdot \ldots \cdot (365 - m + 1)}{365^m}$.

278. $\frac{20}{243}$. **279.** 60. **280.** 8008. **281.** а) C_{n+m-1}^{m-1}; б) C_{n-1}^{m-1}. **282.** $x^6 + 6x^5y + 15x^4y^2 + 20x^3y^3 + 15x^2y^4 + 6xy^5 + y^6$. **283.** $a^{100} + 100a^{99}b + 4950a^{98}b^2$. **284.** 4. **285.** 66. **286.** $495x^4y^2$. **287.** $84x^3a^3$. **288.** а) 1,01; б) 0,88. **289.** *Указание*: примените формулу бинома Ньютона к выражению $((a + b) + 2c)^4$. **290.** а) 27; б) 64; в) 184; г) 4; д) 9; е) 2; ж) 14; з) 7. **291.** $\frac{2}{3}$. **292.** $\frac{6}{27}$. **293.** 0,1. **294.** 6; 27. **295.** а) 0,3; б) 0,15. **296.** а) 4; б) 24. **297.** а) 6; б) 12. **298.** а) 3; б) 6. **299.** а) 6; б) 4. **300.** а) $\frac{1}{6}$; б) $\frac{1}{3}$; в) $\frac{2}{3}$. **301.** а) 6; б) 2. **302.** а) 16; б) 5. **303.** а) 36; б) 2. **304.** а) 12; б) 4. **305.** $5^{2001} \cdot (9 \cdot 2^{1000} - 5)$. **306.** $(365^{90} - 365 \cdot 364 \cdot \ldots \times 276) : (365 \cdot 364 \cdot \ldots \cdot 276)$. **307.** а) $\frac{7}{18}$; б) $\frac{5}{9}$; в) $\frac{7}{18}$. **308.** а) $\frac{13}{90}$; б) $\frac{5}{24}$; в) 0,95; г) $\frac{1}{15}$. **309.** $\left(\frac{32}{33}\right)^5$. **310.** а) $\frac{32}{63}$; б) $\frac{31}{126}$; в) $\frac{31}{63}$. **311.** $\frac{671}{1296}$. **312.** а) 32 768; б) 6720; в) 26 048; г) 13 608; д) 3430. **313.** а) 759 375; б) 37 500; в) 19 800; г) 144 000; д) 14 400; е) 68 250. **314.** а) 55; б) 28; в) 1; г) 36; д) 52; е) 10. **315.** а) 20; б) 288; в) 1728; г) 504; д) 5120; е) 17 856; ж) 193 536. **316.** $\approx 0{,}53$. *Указание*: шесть концов травинок можно попарно связать между собой 15 различными способами. Так как верхние и нижние концы независимо можно связать 15 различными способами, то общее число равновозможных исходов опыта равно $15 \cdot 15 = 225$. Для того чтобы получилось кольцо, нижний конец первой травинки должен быть связан с нижним концом третьей, четвертой, пятой или шестой травинки, т. е. имеем четыре возможности. Если, например, нижний конец первой травинки связан с концом третьей, то нижний конец второй травинки надо связать с концом или пятой или шестой травинки, т. е. имеем две возможности. После этого останется связать еще два не связанных конца. **317.** 502. **318.** а) $\approx 0{,}0000000736$; б) $\approx 0{,}000969$; в) $\approx 0{,}000986$; г) $\approx 0{,}436$. **319.** а) $\approx 0{,}0274$; б) 0,957. **320.** а) 0,00000154; б) 0,0000123; в) 0,00001; г) 0,00144. **321.** а) $\frac{7}{750}$; б) $\frac{1}{75}$. **322.** а) 15 или 16; б) 6 или 10; в) < 8; г) $\frac{1}{45\,045}$. **323.** $\frac{12!}{6^{12} \cdot 2^6}$. **324.** а) $\approx 0{,}152$; б) $\approx 0{,}176$. **325.** а) 0,000001; б) 0,06048. **326.** 1 и 10. **327.** 924. **328.** 5. **329.** 8. **330.** 124. **331.** Седьмой член, равный 120. **332.** Пятый член, равный $56x^4$. **334.** *Указание*: данное неравенство равносильно неравенству $\left(1 + \frac{1}{n}\right)^n < n$. Докажите, что $\left(1 + \frac{1}{n}\right)^n < 3$.

Глава 4

335. а)

1	2	3	4
0,2	0,4	0,1	0,3

б) 0,8; 0,6; 0,8. **336.** а) 0,5; 0,575; 0,925; б)

0	1	4
0,5	0,1	0,4

0,25	2,25
0,525	0,475

337. а)

0	1	2
0,25	0,5	0,25

б)

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

в)

1	4	9	16	25	36
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

г)

−5	−4	−3	−2	−1	0	1	2	3	4	5
$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

338.

0	1
$\frac{5}{6}$	$\frac{1}{6}$

339.

0	1	2
0,09	0,42	0,49

340. а)

0	1	2
0,6	0,3	0,1

б)

0	1	2	3	4
0,36	0,36	0,21	0,06	0,01

341. а)

−1,5	−0,5	1,5
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

б)

−3	−2	−1	0	1	3
$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

342. а)

0	a
$\frac{1}{3}$	$\frac{2}{3}$

б)

0	a	$2a$	$3a$
$\frac{1}{9}$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

343. 0,95. **344.** а) p; б) $1 - p$. **345.** а) 24; б) 0,4. **346.** 75 000 р. **347.** 0,6 р. **348.** 1,248. **349.** 1,75. **350.** а) 1; б) 1,4; в) 0,5; г) $\frac{14a}{9}$. **351.** 0,65. **352.** а) 3; б) 3. **353.** 1) 2,7; б) 1,1; в) 27; г) 0. **354.** 2,8. **355.** 1) np; б) $p_1 + p_2 + \ldots + p_n$. **356.** a. **357.** а) 1,5; б) 15. **358.** а) Второй; б) 31 р.; в) 15,8 р. **359.** Первый. **361.** а) $\approx 0{,}774$; б) $\approx 0{,}0214$. **362.** $\frac{3}{16}$. **363.** $\approx 0{,}279$. **364.** а) $\approx 0{,}25$; б) $\approx 0{,}5$.

365.

0	1	2	3	4
0,0016	0,0256	0,1536	0,4096	0,4096

366.

0	1	2	3	4	5
10^{-5}	$4{,}5 \cdot 10^{-4}$	$8{,}1 \cdot 10^{-3}$	$7{,}29 \cdot 10^{-2}$	$\approx 3{,}28 \cdot 10^{-1}$	$\approx 5{,}9 \cdot 10^{-1}$

367.

x_k	0	1	2	3	4	5
p_k	0,03125	0,15625	0,3125	0,3125	0,15625	0,03125
ν_k	0,03663	0,1526	0,3382	0,2735	0,1661	0,03297

368. а) 0,0183; б) 0,0733; в) 0,147; г) 0,195; д) 0,238. **369.** 1,4. **370.** 11,04. **371.** 275. **372.** $MX = MY = 2$; $DX = 1{,}2$; $DY = 1{,}5$. Предпочтение нужно отдать первому станку. Его продукция «стабильнее». **373.** Наибольшую среднюю прибыль может принести третий проект, наиболее надежный — первый. **374.** а) 1129; 33,6; б) 0,9; в) 0,9. **375.** Уровень риска характеризуется средним квадратичным отклонением, равным 43 300 000 р. **376.** 18,6; 5,0. **378.** а) 11; б) 6; в) $a^2 + 2b^2$. **379.** 150 В. **380.** а) 13,01; б) 13,01; в) – 0,07;

г)

−2	−1	0	1	2
0,15	0,14	0,46	0,13	0,12

381. $p_1 + p_2$; $p_1(1 - p_1) + p_2(1 - p_2)$.

383. 100. **384.** 2; 1,996. **385.** 80; 16; 4. **386.** 2. **387.** 100; 0,1. **388.** 30. **389.** а) p; б) $\frac{p(1-p)}{n}$. **390.** а) Биномиальное, с параметрами 10 и 0,86; б) 8,6; в) $\approx 1{,}1$. **391.** а) 300 и 0,53; б) 159; 8,64; в) 0,833. **392.** а) $\approx 0{,}570$; б) $\approx 0{,}967$; в) $\approx 0{,}994$. **393.** а) $\frac{27}{32}$; б) $\frac{63}{64}$; в) 1. **394.** $\geqslant 3{,}17$; $\geqslant 10$. **395.** а) $\geqslant \frac{8}{9}$; б) $\leqslant 0{,}25$; в) $k \geqslant 10$. **396.** 5556. **397.** а) 1; 0,99; б) $\leqslant 0{,}0099$. **398.** $\geqslant 0{,}76$. **399.** $\geqslant 1 - \frac{1}{h^2}$; при $0 \leqslant h \leqslant 1$. **400.** Нет оснований отвергать предположение. **401.** а) $\leqslant \frac{1}{800}$; б) $\leqslant 0{,}02$. **402.** а) $\leqslant \frac{16}{25}$; $\geqslant \frac{8}{9}$; б) $\geqslant 20$. **403.** а) 0 и $2c^2p$; 2) $\sigma \geqslant c$. **405.** 1) $\geqslant 10\ 000$; б) $\geqslant 250\ 000$. **406.** $\geqslant 0{,}986$. **407.** $\geqslant 4267$. **408.** $\geqslant 87$. **409.** Применима. **410.** а) 0,5; б) 0,203; в) 0,006; г) 0,0475; д) 0. **411.** $\approx 84\%$. **412.** а) 0,3632; б) 0,3446; в) 0,2005; г) 0,102; д) 0,9554. **413.** Вероятность того, что случайная величина, имеющая стандартное нормальное распределение, отклоняется от своего среднего значения больше, чем на три средних квадратичных отклонения, равна 0,0017, т. е. это событие практически невозможно. **414.** 0,9545. **415.** 4,4%. **416.** а) 0,1148 и 0,1052; б) 0,0398 и 0,0668.

417. $3{,}645 \cdot 10^{-7}$ и 0,0072. **418.** а)

1	2	3	4
0,6	0,24	0,096	0,064

б) 0,16.

419.

1	2	3
0,7	0,21	0,09

420. а)

2	3	4	5	6	7	8
$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

б)

−3	−2	−1	0	1	2	3
$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

в)

1	2	3	4
$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$

г)

1	2	3	4
$\frac{7}{16}$	$\frac{5}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

421.

0	1	2
$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

422. а)

0	1	2	3
$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

; рис. 86; б)

−3	−1	1	3
$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

; рис. 87. **423.** а)

1	2	3	4	5	6
0,05	0,4	0,1	0,25	0,05	0,15

б) 0,55. **424.**

0	1	2
$\frac{2}{21}$	$\frac{10}{21}$	$\frac{9}{21}$

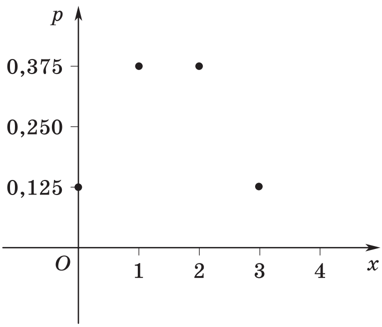


Рис. 86

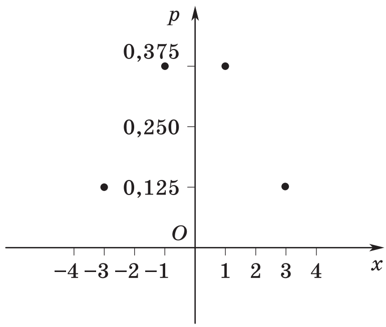


Рис. 87

425. а)

6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\frac{6}{216}$	$\frac{6}{216}$	$\frac{12}{216}$	$\frac{18}{216}$	$\frac{24}{216}$	$\frac{24}{216}$	$\frac{30}{216}$	$\frac{24}{216}$	$\frac{24}{216}$	$\frac{19}{216}$	$\frac{13}{216}$	$\frac{7}{216}$	$\frac{7}{216}$	$\frac{1}{216}$	$\frac{1}{216}$

б)

3	4	5	6
$\frac{6}{216}$	$\frac{24}{216}$	$\frac{61}{216}$	$\frac{125}{216}$

в)

1	2	3	4
$\frac{125}{216}$	$\frac{6}{216}$	$\frac{24}{216}$	$\frac{61}{216}$

426.

2	3	4	5	6	7	8	9	10
0,04	0,08	0,12	0,16	0,2	0,16	0,12	0,08	0,04

427.

0	1	2
$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

428. 10 р. **429.** $\approx 1{,}4$. **430.** Никто из игроков не имеет преимущества. **431.** $6\frac{1}{12}$. **432.** а) $\approx 12{,}16$ денежных единиц; б) $\approx 5{,}47$ денежных единиц; в) $\approx 1{,}57$ денежных единиц. **433.** 3 или 4. **434.** $-1{,}65$ р. **435.** 1. **436.** 0,4; 0,1; 0,5. **437.** $>$ 50 р. **438.** 10 120 р. **439.** «5 из 36». *Указание*: вычислите математические ожидания выигрышей для обеих игр. **440.** 7. *Указание*: X_i — число очков, выпавших при бросании i-го кубика, $i = 1, 2$; X — сумма очков; $X = X_1 + X_2$. **441.** 1,3. См. указание к задаче 440. **442.** $\approx$ 731 р. **443.** 13 250 р. **444.** а) Первая; б) 18,2; в) 5,45. **445.** $\approx 1{,}7$. **446.** 1) 0,1; б) 10. **447.** 0,3. **448.** 16 р. *Указание*: обозначьте через X_i случайную величину, равную стоимости i-й вынутой монеты, $i = 1, 2, 3$. Тогда сумма $X_1 + X_2 + X_3$ равна общей стоимости вынутых монет. **449.** а) 0,00000396; б) 0,9994; в) 0,0395. **450.** а) 0,2646; б) 0,7599; в) 0,2401. **451.** Три из четырех. **452.** 0,384. **453.** а) 0,3125; б) 0,8125; в) 0,96875. **454.** $\approx 0{,}867$.

455.

0	1	2	3
0,001	0,027	0,243	0,729

456. а)

0	1	2	3
0,729	0,243	0,027	0,001

б) 0,028. **457.**

0	1	2	3
$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$

458. а) Стоимость ценных бумаг падает независимо друг от друга; б) $\approx 0{,}035$; в) $\approx 0{,}103$; г) $\approx 0{,}398$. **459.** 0,2. **461.** Не менее двух. **462.** $\approx 0{,}206$. **463.** $\frac{m}{n}$; $\frac{C_n^m m^n (n-m)^{n-m}}{n^n}$. **464.** m; $\left(\frac{m}{e}\right)^m \cdot \frac{1}{m!}$. **465.** $\approx \frac{\left(\frac{1000}{365}\right)^k}{k!} e^{-\frac{1000}{365}}$; 0,065; 0,177; 0,242; 0,221. **466.** $DX < DY$. **467.** 3,3; 2,31. **468.** 10 000; 1; 1. **469.** 0,6; 0,24. **470.** а) 5 и 2,5; б) 0 и 2,5; в) $\frac{25}{8}$ и $\frac{55}{64}$; г) $\frac{15}{8}$ и $\frac{55}{64}$. **471.** а) 2,5; 1,11; б) 2,4; $\approx 1{,}37$. *Указание*: если X — число красных карт среди пяти выбранных, то $P(X = i) = \frac{C_{18}^i \cdot C_{18}^{5-i}}{C_{36}^5}$, $i = 0, 1, 2, 3, 4, 5$. **472.** $\approx 4{,}59$; $\approx 3{,}07$; $\approx 1{,}75$. **473.** $\frac{n+1}{2}$; $\frac{n^2-1}{12}$. *Указание*: воспользуйтесь формулой $1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

474. а)

1	2	3	4	5	6
$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

; $\frac{161}{36}$; $\frac{2555}{1296}$; $\approx \frac{51}{36}$.

б)

1	2	3	4	5	6
$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

; $\frac{91}{36}$; $\frac{2555}{1296}$; $\approx \frac{51}{36}$. **476.** а) 10,5; 10,5; 8,75; 18,75. б) Вторым. **477.** $\frac{35}{6}$; 0,69; $\frac{25}{18}$; 10. **478.** 4. **479.** 810. **480.** а) 0,5; б) 0,45; в) $\approx$ 0,67. **481.** 2,5. **482.** 0,63. **483.** 0,7 или 0,3. **484.** 40,96. **485.** а) 12; б) $\approx$ 1,55. **486.** $\geqslant$ 7. **487.** $\geqslant$ 0,56. **488.** $\geqslant$ 0,525. **489.** а) $\geqslant$ 0,64; б) $\leqslant$ 0,36. **490.** а) $\leqslant \frac{1}{1800}$; б) $\leqslant$ 0,125. **491.** $\geqslant$ 2000. *Указание*: используйте неравенство Чебышева и неравенство $p(1-p) \leqslant \frac{1}{4}$. **492.** $\geqslant$ 0,75. См. указание к задаче 491. **493.** $\left|p - \frac{m}{n}\right| \leqslant 0{,}22$. См. указание к задаче 491. **494.** $\geqslant \frac{391}{392}$. **495.** $\geqslant$ 0,91. **496.** а) 1 мм; б) 3 мм. **497.** 3920. **498.** 89%. **499.** а) $\geqslant$ 0,976; б) $\geqslant$ 0,934. **500.** Применима. **501.** 1,68. **502.** $2P\left(Z < \frac{d}{\sigma}\right) - 2P\left(Z < \frac{c}{\sigma}\right)$. **503.** $2P\left(Z < \frac{d-a}{\sigma}\right) - 2P\left(Z < \frac{c-a}{\sigma}\right)$. **504.** $P\left(Z < \frac{d}{\sigma}\right) - P\left(Z < \frac{c}{\sigma}\right)$.

Глава 5

506. E, A, G. **507.** R, B, M, Q. **508.** Учащиеся с номерами: 48, 40, 38, 63, 79, 61. **510.** Например, абзацы с номерами: 23, 43, 45, 50, 58, 62, 63, 72, 79, 91. **511.** а) 0,999; б) 0,942. **512.** а) $\geqslant$ 0,975; б) нет содержательных оценок; в) $\geqslant$ 0,375; г) нет содержательных оценок. **513.** а) $\geqslant$ 313; б) $\geqslant$ 1250; в) $\geqslant$ 79. **514.** 0,065. **515.** $\frac{nM}{m}$. **516.** $\geqslant$ 0,75. **517.** Таким показателем может быть стандартная ошибка, равная 1,8%. **518.** а) $\geqslant \frac{8}{9}$; б) $\geqslant$ 50 000. **519.** а) $\geqslant$ 0,6875; б) $\geqslant$ 14 706. **520.** (47,4; 57,4). **521.** (0,2; 0,6). **522.** (0,7; 0,9). **523.** 100. **524.** $\left(\frac{1}{30}; \frac{11}{30}\right)$. **525.** 0,91. **526.** а) (0,52; 0,73); б) (0,58; 0,67).

527. (318; 328). **528.** а) Генеральная совокупность — потенциальные покупатели товара, выборка — отобранные 100 человек; б) нулевая гипотеза — число покупателей товара не изменится после просмотра рекламы; альтернативная — это число покупателей изменится; в) например, если из 100 человек, посмотревших рекламу, 80 купят рекламируемый товар, то нулевую гипотезу отвергают; г) ошибка первого рода — из числа посмотревших рекламу большая доля купит рекламируемый товар (нулевая гипотеза отвергается), а общее число покупателей этого товара не изменилось (нулевая гипотеза верна); ошибка второго рода — из числа посмотревших рекламу небольшая доля купит рекламируемый товар (нулевая гипотеза не отвергается), а общее число покупателей этого товара существенно изменилось (нулевая гипотеза неверна);

д)

Принятое решение	Истинное положение	
	Число покупателей не изменилось	**Число покупателей изменилось**
Число покупателей не изменилось	Правильное решение	*Ошибка второго рода* Принято решение о неэффективности рекламы
Число покупателей изменилось	*Ошибка первого рода* Принято решение об эффективности рекламы — она будет повсеместно вводиться, но не оправдает надежд	*Правильное решение*

531. а) 0,043; б) нулевая гипотеза отвергается; в) 0,043; г) 0,043. **532.** а) 0,023; б) нулевая гипотеза отвергается; в) 0,023; г) 0,023. **533.** Нулевая гипотеза — среднее количество древесины в одном дереве на другом участке леса равно 1,3 м^3 — на уровне значимости 0,05 отвергается. **534.** Нулевая гипотеза — среднее отклонение диаметров валиков от номинала равно 10 мк — на уровне значимости 0,05 отвергается. Переналадка станка дала отрицательный эффект: отклонение увеличилось после переналадки. **535.** Нулевая гипотеза — средняя чувствительность радиоприемника равна 300 мкВ — на уровне значимости 0,05 не отвергается. Нельзя утверждать, что усовершенствование технологии оказалось эффективным. **536.** а) (805; 843); б) 800; в) H_0 — истинная масса буханки хлеба равна 800 г; H_1 — истинная масса буханки хлеба не равна 800 г; г) гипотеза H_0 на уровне значимости 0,05 отвергается; д) масса буханки хлеба в отобранной выборке значимо отличается от 800 г, а в целом она близка к 800 г (ошибка первого рода); масса буханки хлеба в отобранной выборке близка к 800 г, а в целом она значимо отличается от 800 г (ошибка второго рода). **537.** а) 600; б) нулевая гипотеза — потенциальные покупатели готовы в среднем платить за изделие 600 р. — на уровне значимости 0,05 отвергается; в) нулевая гипотеза — потенциальные покупатели готовы в среднем платить за изделие 600 р. — на уровне значимости 0,01 отвергается; г) ввиду того, что выборочное исследование привело к результату, значительно меньшему, чем 600 р.; д) нулевая гипотеза для односторонней проверки того, что покупатели готовы платить меньше 600 р., утверждает, что покупатели готовы в среднем платить не менее 600 р., H_0: $a \geqslant 600$; альтернативная гипотеза для проверки этой гипотезы утверждает, что покупатели готовы в среднем платить менее 600 р., H_1: $a < 600$; е) нулевая гипотеза на уровне значимости 0,05 при односторонней проверке не отвергается: выборочное среднее 545 не является значимо меньшим, чем заданное значение 600. **538.** Нет. **539.** Нет. **540.** Нет. **541.** Нет. **542.** Нет. **543.** а) Нет; б) нет; в) нет. **545.** г). **546.** C_N^n. **547.** Нет.

548. а) 39,8%; б) выборка может оказаться нерепрезентативной; за три недели могли измениться намерения избирателей благодаря проведенной рекламной кампании. **549.** а) $\geqslant 0{,}9$; б) $\geqslant 5209$. **550.** а) $\geqslant 0{,}48$; б) $\geqslant 7$. **551.** Примерно на 5%. **552.** а) $\geqslant 0{,}77$; б) $\geqslant 105$. **553.** (1; 9). **554.** (0,7; 0,17). **555.** (4,8; 6,0). **556.** а) $\geqslant 0{,}65$; б) (11,4; 12,2); в) $\geqslant 138$. **557.** а) $\approx 29\%$; б) $\approx 0{,}015$; в) (0,14; 0,26); г) (0,175; 0,225); (0,17; 0,23). **558.** а) Требование выполняется; б) (0,25; 0,31). **559.** а) 0,0645; б) нулевая гипотеза не отвергается; в) 0,0645; г) 0,0645. **560.** а) 0,028; б) гипотеза о правильности кубика отвергается; в) 0,028; г) 0,028. **561.** а) Средние частоты употребления предлога «on» в спорных статьях и в статьях Гамильтона совпадают. **563.** а) Средние частоты различных слов в рассказе «Шагать немедленно» и в романе «Тихий Дон» совпадают. **564.** Нулевая гипотеза — средняя оценка влияния передачи акций на качество продукции равна 0 — на уровне значимости 0,05 отвергается. **565.** Незначимо. **566.** а) Уложился; б) H_0: $a \geqslant 125$, H_1: $a < 125$; в) нулевая гипотеза H_0: $a \geqslant 125$ при двусторонней проверке отвергается. **567.** Нет данных, которые противоречили бы тому, что автором спорных статей является Медисон. **568.** Нет данных, которые противоречили бы тому, что автором спорных статей является Медисон. **569.** Нет данных, которые противоречили бы тому, что автором «Тихого Дона» является М. А. Шолохов. **570.** а) На 51%; 0,013; б) гипотеза о том, что уровень бака остается на том же уровне, не отвергается; в) (−0,002; 0,05); г) нет. **571.** Нет. **572.** а) Гипотезе H_0; б) 0,002. **573.** $\frac{15}{28}$.

Приложение

Основные понятия, факты, формулы

Описательная статистика

Число случаев n_i, в которых встречается значение x_i, называют ***частотой*** значения x_i.

Сумму $n = n_1 + n_2 + \ldots + n_k$, где k — число различных значений, которые принимает исследуемая величина, называют ***объемом совокупности данных***.

Относительной частотой ν_i значения x_i называют отношение частоты n_i этого значения к объему n совокупности данных: $\nu_i = \dfrac{n_i}{n}$.

Накопленной частотой интервала называют число значений, которые попали в этот интервал и все предшествующие.

Накопленной относительной частотой интервала называют отношение накопленной частоты этого интервала к объему совокупности данных.

Отношение частоты n_i интервала к ширине h_i этого интервала называют ***абсолютной плотностью распределения*** для i-го интервала: $p_i = \dfrac{n_i}{h_i}$.

Относительной плотностью распределения π_i для i-го интервала называют отношение относительной частоты интервала к его ширине: $\pi_i = \dfrac{\nu_i}{h_i}$.

Среднее арифметическое n значений $x_1, x_2, \ldots, x_n$: $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Среднее арифметическое дискретного вариационного ряда

x_1	x_2	...	x_m
n_1	n_2	...	n_m

: $\bar{x} = \sum\limits_{i=1}^{m} x_i \frac{n_i}{n}$.

Среднее арифметическое по данным интервального вариационного ряда вычисляют по той же формуле, но за значение x_i берут середину i-го интервала.

Медиана совокупности — это такое значение признака, которое делит всю совокупность пополам.

Медиана дискретного вариационного ряда $\underbrace{x_1, x_2, \ldots, x_n}_{n \text{ значений}}$, x_{n+1}, $\underbrace{x_{n+2}, \ldots, x_{2n+1}}_{n \text{ значений}}$, варианты которого размещены в порядке возрастания их значений и объем нечетный, равна $\mathrm{Me} = x_{n+1}$.

Медиана дискретного вариационного ряда $\underbrace{x_1, x_2, \ldots, x_n}_{n \text{ значений}}$, $\underbrace{x_{n+1}, \ldots, x_{2n}}_{n \text{ значений}}$, варианты которого размещены в порядке возрастания их значений и объем четный, равна $\mathrm{Me} = \frac{x_n + x_{n+1}}{2}$.

Медиана интервального упорядоченного вариационного ряда равна $\mathrm{Me} = x_n + \frac{\frac{n}{2} - S_{\mathrm{Me}-1}}{n_{\mathrm{Me}}} h$, где x_n — начало медианного интервала, т. е. интервала, в котором находится медиана; h — ширина медианного интервала; n_{Me} — частота медианного интервала; $S_{\mathrm{Me}-1}$ — сумма частот интервалов, предшествующих медианному; n — объем совокупности.

Мода — это такое значение признака, которое наиболее часто встречается.

Мода интервального вариационного ряда с равными интервалами вычисляется по формуле $\mathrm{Mo} = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)}$, где x_0 — начальное значение модального интервала, т. е. интервала, который содержит моду; h — длина модального интервала; n_2 — частота модального интервала; n_1 — частота интервала, предшествующего модальному; n_3 — частота интервала, следующего за модальным.

Среднее гармоническое значений $x_1, x_2, \ldots, x_n$ определяется формулой $\bar{x}_{\text{гарм}} = \dfrac{n}{\sum\limits_{i=1}^{n} \dfrac{1}{x_i}}$.

Среднее гармоническое дискретного вариационного ряда

x_1	x_2	...	x_m
n_1	n_2	...	n_m

: $x_{\text{гарм}} = \dfrac{\sum\limits_{i=1}^{m} n_i}{\sum\limits_{i=1}^{m} \dfrac{n_i}{x_i}}$.

Среднее геометрическое значений $x_1, x_2, \ldots, x_n$ определяется по формулам: $\bar{x}_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$ или $\bar{x}_{\text{геом}} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}}$.

Размах совокупности — это разность между максимальным и минимальным значениями в совокупности.

Дисперсия статистической совокупности определяется по формулам:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{или} \quad s_x^2 = \frac{1}{n} \sum_{i=1}^{m} (x_i - \bar{x})^2 n_i.$$

Среднее квадратичное, или ***стандартное, отклонение***:

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{или} \quad s_x = \sqrt{\frac{1}{n} \sum_{i=1}^{m} (x_i - \bar{x})^2 n_i}.$$

Дисперсия статистической совокупности может быть вычислена по формуле $s_x^2 = \overline{x^2} - \bar{x}^2$.

Коэффициент вариации V определяется по формулам:

$$V = \frac{s}{\bar{x}} \quad \text{или} \quad V = \frac{s}{\bar{x}} \cdot 100\%.$$

Пусть $0 < p < 1$ и статистические данные размещены по возрастанию: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. Если np не является целым числом, то ***квантилем*** z_p ***уровня*** p называют варианту x_{k+1}, где $k = [np]$. Если np является целым числом, то квантиль z_p считают равным полусумме вариант x_{np} и x_{np+1}.

Квантиль z_p ***уровня*** p для интервального вариационного ряда вычисляют по формуле $z_p = x_{z_p} + \dfrac{np - S_{z_p - 1}}{n_{z_p}} h$, где x_{z_p} — нижняя граница интервала, содержащего квантиль z_p; S_{z_p-1} — частота, накопленная к интервалу, содержащему квантиль z_p; n_{z_p} — частота интервала, содержащего квантиль z_p; h — ширина интервала, содержащего квантиль z_p.

Случайные события

Классическое определение вероятности

Если эксперимент может привести к любому из N различных равновозможных исходов и если в точности $N(A)$ из этих исходов соответствуют событию A, то ***вероятность события*** A равна $P(A) = \frac{N(A)}{N}$.

Вероятностная модель случайного опыта

Пусть $U = \{u_1, \ldots, u_N\}$ — произвольное конечное множество. Будем называть его ***пространством элементарных исходов*** (ПЭИ) случайного опыта, а его элементы $u_1, \ldots, u_N$ — исходами этого опыта. Произвольное подмножество (часть) ПЭИ будем называть ***случайным событием*** (или короче, ***событием***). Каждому элементарному исходу u_i из ПЭИ $U = \{u_1, u_2, \ldots, u_N\}$ поставим в соответствие некоторое число p_i, которое удовлетворяет условиям:

1) $0 < p_i < 1$ для всех $i = 1, 2, \ldots, N$; 2) $p_1 + p_2 + \ldots + p_N = 1$.

Числа p_i называют ***элементарными вероятностями,*** или ***вероятностями элементарных исходов*** u_i.

Вероятностью события A называют сумму вероятностей исходов, составляющих это событие: $P(A) = \sum\limits_{i:\, u_i \in A} p_i$.

Вероятность суммы двух событий:

$$P(A + B) = P(A) + P(B) - P(AB).$$

Вероятность суммы двух несовместных событий:

$$P(A + B) = P(A) + P(B).$$

Вероятность суммы m попарно несовместных событий:

$$P(A_1 + A_2 + \ldots + A_m) = P(A_1) + P(A_2) + \ldots + P(A_m).$$

Вероятность суммы трех событий:

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Условная вероятность события A при условии B:

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

Теорема умножения вероятностей:

Если $P(B) > 0$, $P(A) > 0$, то $P(AB) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B)$.

Теорема умножения вероятностей для n событий:

$$P(A_1A_2A_3 \ldots A_n) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1A_2) \ldots P(A_n \mid A_1A_2 \ldots A_{n-1}).$$

События A и B независимы тогда и только тогда, когда $P(AB) = P(A)\,P(B)$.

Случайные события A_1, A_2, A_3 независимы в совокупности, если они попарно независимы и

$$P(A_1A_2A_3) = P(A_1)\,P(A_2)\,P(A_3).$$

Случайные события A_1, A_2 A_3, A_4 независимы в совокупности, если независимы в совокупности любые три из них и

$$P(A_1A_2A_3A_4) = P(A_1)\,P(A_2)\,P(A_3)\,P(A_4).$$

Система событий H_1, H_2, ..., H_n называется ***полной группой несовместных событий,*** если:

1) $H_1 + H_2 + \ldots + H_n = U$; 2) $H_iH_k = V$, $i \neq k$.

Формула полной вероятности:

$$P(A) = P(H_1)\,P(A \mid H_1) + P(H_2)\,P(A \mid H_2) + \ldots + P(H_n)\,P(A \mid H_n),$$

где H_1, H_2, ..., H_n — полная группа несовместных событий.

Элементы комбинаторики

Комбинаторное правило умножения:

Если объект A можно выбрать m способами и если после ***каждого*** такого выбора объект B можно выбрать n способами, то выбор пары (A, B) в указанном порядке можно осуществить $m \cdot n$ способами.

Обобщенное комбинаторное правило умножения:

Если объект A_1 может быть выбран n_1 различными способами, A_2 — n_2 различными способами и т. д., A_k – n_k различными способами, то k объектов A_1, A_2, ..., A_k в указанном порядке можно выбрать $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Комбинаторное правило сложения:

Если некоторый объект A можно выбрать m способами, а другой объект B можно выбрать n способами, причем ни один из способов выбора A не совпадает с каким-нибудь способом выбора B, то выбор «либо A, либо B» можно осуществить $m + n$ способами.

Общее комбинаторное правило сложения:

Если объект A можно выбрать m способами, а объект B – n способами, причем при k способах одновременно выбираются и A и B, то выбор «A или B» можно осуществить $m + n - k$ способами.

Основные комбинаторные схемы

Число упорядоченных выборок с возвращением из n элементов по r (размещений с повторениями из n элементов по r) равно $\overline{A_n^r} = n^r$.

Число упорядоченных выборок без возвращения из n элементов по r (размещений из n элементов по r) равно $A_n^r = n(n-1)(n-2)\,\ldots\,(n-r+1)$.

Число всех различных перестановок из n элементов вычисляют по формуле $P_n = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n = n!$

Число перестановок, которые можно образовать из n элементов, среди которых n_1 элементов 1-го типа, n_2 — 2-го типа, ... , n_k — k-го типа (перестановок с повторениями), равно $P_n(n_1, n_2, \ldots, n_k) = \dfrac{n!}{n_1!n_2!\,\ldots\,n_k!}$.

Число неупорядоченных выборок без возвращения из n элементов по r (сочетаний из n элементов по r) равно

$$C_n^r = \frac{n(n-1)(n-2)\,\ldots\,(n-r+1)}{r!}, \quad \text{или} \quad C_n^r = \frac{n!}{r!(n-r)!}.$$

Число неупорядоченных выборок с возвращением из n элементов по r (сочетаний с повторениями из n элементов по r) равно

$$f_n^r = C_{n+r-1}^{n-1} = C_{n+r-1}^r.$$

Формула бинома Ньютона:

$$(a+b)^n = a^n + C_n^1 a^{n-1}b + C_n^2 a^{n-2}b^2 + \ldots + C_n^{n-1} a^1 b^{n-1} + b^n,$$

$$\text{или} \quad (a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k}b^k.$$

Коэффициенты C_n^k в этой формуле называют ***биномиальными коэффициентами.***

Случайные величины

Случайной величиной называют числовую функцию, определенную на пространстве элементарных исходов.

Законом распределения дискретной случайной величины X называют числовую функцию, которая каждому значению x случайной величины X ставит в соответствие вероятность $P(X = x)$:

Значения x	x_1	x_2	...	x_n
Вероятности p	p_1	p_2	...	p_n

Сумму произведений значений случайной величины на соответствующие вероятности называют ***математическим ожиданием (или средним значением) случайной величины X***:

$$MX = x_1p_1 + x_2p_2 + \ldots + x_np_n, \quad \text{или} \quad MX = \sum_{i=1}^{n} x_ip_i.$$

При достаточно большом числе наблюдений ***математическое ожидание случайной величины приближенно равно выборочному среднему*** наблюдаемых значений этой величины: $\bar{x} \approx MX$.

Случайные величины X и Y называют ***независимыми***, если для любых x и y выполняется равенство $P((X = x) \cdot (Y = y)) = P(X = x) \cdot P(Y = y)$.

Дисперсия случайной величины X определяется равенством

$$DX = M(X - MX)^2.$$

Среднее абсолютное отклонение случайной величины X:

$$M|X - MX|.$$

Среднее квадратичное отклонение случайной величины X:

$$\sigma(X) = \sqrt{DX}\,.$$

При достаточно большом числе наблюдений ***дисперсия случайной величины приближенно равна выборочной дисперсии*** наблюдаемых значений этой величины: $DX \approx s^2$.

Формула Бернулли

Вероятность того, что в n испытаниях Бернулли произойдет ровно m «успехов», равна $P_n(m) = C_n^m\, p^m(1 - p)^{n - m}$, $m = 0, 1, 2, \ldots, n$, p — вероятность «успеха» в каждом испытании.

Случайная величина — число «успехов» в n испытаниях Бернулли – имеет ***биномиальное распределение с параметрами n и p***.

Закон биномиального распределения с параметрами n и p имеет вид:

$$P(X = m) = C_n^m p^m (1 - p)^{n-m}, \quad m = 0, 1, 2, \ldots, n.$$

Если значения n велики, а p малы, то имеет место так называемая приближенная ***формула Пуассона***: $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$, где $\lambda \approx np$, $e \approx 2{,}72$.

Числовые характеристики биномиального распределения

Если случайная величина X имеет биномиальное распределение с параметрами n и p, то $MX = np$; $DX = np(1 - p)$.

Неравенство Чебышёва

При произвольном положительном ε для любой случайной величины X выполняется неравенство $P(|X - MX| > \varepsilon) \leqslant \frac{DX}{\varepsilon^2}$.

Правило трех сигм: $P(|X - a| \leqslant 3\sigma) \geqslant \frac{8}{9}$.

Закон больших чисел

Теорема Чебышёва. Если случайные величины $X_1, \ldots, X_n$ независимы и имеют одно и то же математическое ожидание a и одну и ту же дисперсию, то их среднее арифметическое при достаточно большом n с вероятностью как угодно близкой к 1, будет как угодно мало отличаться от a.

Теорема Бернулли. Пусть событие A наступает k раз в n независимых испытаниях, p — вероятность наступления этого события в каждом испытании. Тогда для любого $\varepsilon > 0$ $P\left(\left|\frac{k}{n} - p\right| \leqslant \varepsilon\right) \to 1$ при $n \to \infty$.

Элементы математической статистики

Выборкой объема n из некоторого распределения называют совокупность взаимно независимых случайных величин $X_1, X_2, \ldots, X_n$, каждая из которых имеет это распределение.

Оценочной функцией для параметра называют произвольную функцию от элементов выборки. Ее значение от реализации выборки называют ***оценкой параметра***.

Оценочную функцию некоторого параметра генеральной совокупности называют ***несмещенной***, если ее математическое ожидание равно этому параметру. В противном случае ее называют ***смещенной***.

Выборочное среднее $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ является несмещенной оценочной функцией математического ожидания генеральной совокупности a.

Относительная частота события является несмещенной оценочной функцией вероятности этого события.

Оценочную функцию некоторого параметра генеральной совокупности называют ***состоятельной***, если с вероятностью, сколь угодно близкой к 1, при достаточно большом n разность между ней и параметром по модулю можно сделать меньше любого наперед заданного положительного числа ε. Оценочную функцию, не обладающую этим свойством, называют ***несостоятельной***.

Выборочное среднее является состоятельной оценочной функцией математического ожидания.

Относительная частота события является состоятельной оценочной функцией вероятности этого события.

Интервал $\left(\bar{p} - \dfrac{h}{2\sqrt{n}};\ \bar{p} + \dfrac{h}{2\sqrt{n}}\right)$ является ***доверительным интервалом для вероятности р*** некоторого события ***с доверительной вероятностью*** $1 - \dfrac{1}{h^2}$. Число $\dfrac{h}{2\sqrt{n}}$ называют ***точностью оценки*** $\bar{p}$, а число $1 - \dfrac{1}{h^2}$ — ***надежностью этой оценки,*** число $2 \cdot \dfrac{h}{2\sqrt{n}} = \dfrac{h}{\sqrt{n}}$ — ***длиной доверительного интервала*** ($\bar{p}$ — относительная частота события).

Схема проверки статистических гипотез

1. Формулируют нулевую гипотезу.

2. Получают фактические данные о событиях, относительно которых была сформулирована нулевая гипотеза.

3. Определяют вероятность того, что полученный результат мог быть получен при условии, что нулевая гипотеза верна.

4. Если вероятность получения данного результата при условии, что нулевая гипотеза верна, мала, нулевую гипотезу отвергают на уровне значимости, равном этой вероятности.

5. Признают, что и в том случае, когда отвергают и когда не отвергают нулевую гипотезу, возможен определенный риск.

Если отвергается нулевая гипотеза в случае, когда она верна, то говорят, что совершается ***ошибка первого рода***.

Если же нулевая гипотеза не отвергнута в то время, когда она неверна, то совершается ***ошибка второго рода***.

Таблица П.1

Вероятности для стандартного нормального распределения

z	P	z	P	z	P	z	P	z	P	z	P
−2,00	0,0228	−1,00	0,1587	−0,00	0,5000	0,00	0,5000	1,00	0,8413	2,00	0,9772
−2,01	0,0222	−1,01	0,1562	−0,01	0,4960	0,01	0,5040	1,01	0,8438	2,01	0,9778
−2,02	0,0217	−1,02	0,1539	−0,02	0,4920	0,02	0,5080	1,02	0,8461	2,02	0,9783
−2,03	0,0212	−1,03	0,1515	−0,03	0,4880	0,03	0,5120	1,03	0,8485	2.03	0,9788
−2,04	0,0207	−1,04	0,1492	−0,04	0,4840	0,04	0,5160	1,04	0,8508	2,04	0,9793
−2,05	0,0202	−1,05	0,1469	−0,05	0,4801	0,05	0,5199	1,05	0,8531	2,05	0,9798
−2,06	0,0197	−1,06	0,1446	−0,06	0,4761	0,06	0,5239	1,06	0,8554	2,06	0,9803
−2,07	0,0192	−1,07	0,1423	−0,07	0,4721	0,07	0,5279	1,07	0,8577	2,07	0,9808
−2,08	0,0188	−1,08	0,1401	−0,08	0,4681	0,08	0,5319	1,08	0,8599	2,08	0,9812
−2,09	0,0183	−1,09	0,1379	−0,09	0,4641	0.09	0,5359	1,09	0,8621	2,09	0,9817
−2,10	0,0179	−1,10	0,1357	−0,10	0,4602	0,10	0,5398	1,10	0,8643	2,10	0,9821
−2,11	0,0174	−1,11	0,1335	−0,11	0,4562	0,11	0,5438	1,11	0,8665	2,11	0,9826
−2,12	0,0170	−1,12	0,1314	−0,12	0,4522	0,12	0,5478	1,12	0,8686	2,12	0,9830
−2,13	0,0166	−1,13	0,1292	−0,13	0,4483	0,13	0,5517	1,13	0,8708	2,13	0,9834
−,2,14	0,0162	−1,14	0,1271	−0,14	0,4443	0,14	0,5557	1,14	0,8729	2,14	0,9838
−2,15	0,0158	−1,15	0,1251	−0,15	0,4404	0,15	0,5596	1,15	0,8749	2,15	0,9842
−2,16	0,0154	−1,16	0,1230	−0,16	0,4364	0,16	0,5636	1,16	0,8770	2,16	0,9846
−2,17	0,0150	−1,17	0,1210	−0,17	0,4325	0,17	0,5675	1,17	0,8790	2,17	0,9850
−2,18	0,0146	−1,18	0,1190	−0,18	0,4286	0,18	0,5714	1,18	0,8810	2,18	0,9854
−2,19	0,0143	−1,19	0,1170	−0,19	0,4247	0,19	0,5753	1,19	0,8830	2,19	0,9857
−2,20	0,0139	−1,20	0,1151	−0,20	0,4207	0,20	0,5793	1,20	0,8849	2,20	0,9861
−2,21	0,0136	−1,21	0,1131	−0,21	0,4168	0,21	0,5832	1,21	0,8869	2,21	0,9864

Продолжение табл. П.1

z	*P*	*z*	*P*	*z*	*P*	*z*	*P*	*z*	*P*	*z*	*P*
−2,22	0,0132	−1,22	0,1112	−0,22	0,4129	0,22	0,5871	1,22	0,8888	2,22	0,9868
−2,23	0,0129	−1,23	0,1093	−0,23	0,4090	0,23	0,5910	1,23	0,8907	2,23	0,9371
−2,24	0,0125	−1,24	0,1075	−0,24	0,4052	0,24	0,5948	1,24	0,8925	2,24	0,9375
−2,25	0,0122	−1,25	0,1056	−0,25	0,4013	0,25	0,5987	1,25	0,8944	2,25	0,9878
−2,26	0,0119	−1,26	0,1038	−0,26	0,3974	0,26	0,6026	1,26	0,8962	2,26	0,9861
−2,27	0,0116	−1,27	0,1020	−0,27	0,3936	0,27	0,6064	1,27	0,8980	2,27	0,9884
−2,28	0,0113	−1,28	0,1003	−0,28	0,3897	0,28	0,6103	1,28	0,8997	2,28	0,9887
−2,29	0,0110	−1,29	0,0985	−0,29	0,3859	0,29	0,6141	1,29	0,9015	2,29	0,9890
−2,30	0,0107	−1,30	0,0968	−0,30	0,3821	0,30	0,6179	1,30	0,9032	2,30	0,9893
−2,31	0,0104	−1,31	0,0951	−0,31	0,3783	0,31	0,6217	1,31	0,9049	2,31	0,9896
−2,32	0,0102	−1,32	0,0934	−0,32	0,3745	0,32	0,6255	1,32	0,9066	2,32	0,9898
−2,33	0,0099	−1,33	0,0918	−0,33	0,3707	0,33	0,6293	1,33	0,9082	2,33	0,9901
−2,34	0,0096	−1,34	0,0901	−0,34	0,3669	0,34	0,6331	1,34	0,9099	2,34	0,9904
−2,35	0,0094	−1,35	0,0885	−0,35	0,3632	0,35	0,6368	1,35	0,9115	2,35	0,9906
−2,36	0,0091	−1,36	0,0869	−0,36	0,3594	0,36	0,6406	1,36	0,9131	2,36	0,9909
−2,37	0,0089	−1,37	0,0853	−0,37	0,3557	0,37	0,6443	1,37	0,9147	2,37	0,9911
−2,38	0,0087	−1,38	0,0838	−0,38	0,3520	0,38	0,6480	1,38	0,9162	2,38	0,9913
−2,39	0,0084	−1,39	0,0823	−0,39	0,3483	0,39	0,6517	1,39	0,9177	2,39	0,9916
−2,40	0,0082	−1,40	0,0808	−0,40	0,3446	0,40	0,6554	1,40	0,9192	2,40	0,9918
−2,41	0,0080	−1,41	0,0793	−0,41	0,3409	0,41	0,6591	1,41	0,9207	2,41	0,9920
−2,42	0,0078	−1,42	0,0778	−0,42	0,3372	0,42	0,6628	1,42	0,9222	2,42	0,9922
−2,43	0,0075	−1,43	0,0764	−0,43	0,3336	0,43	0,6664	1,43	0,9236	2,43	0,9925
−2,44	0,0073	−1,44	0,0749	−0,44	0,3300	0,44	0,6700	1,44	0,9251	2,44	0,9927
−2,45	0,0071	−1,45	0,0735	−0,45	0,3264	0,45	0,6736	1,45	0,9265	2,45	0,9929
−2,46	0,0069	−1,46	0,0721	−0,46	0,3228	0,46	0,6772	1,46	0,9279	2,46	0,9931
−2,47	0,0068	−1,47	0,0708	−0,47	0,3192	0,47	0,6808	1,47	0,9292	2,47	0,9932

Продолжение табл. П.1

z	*P*	*z*	*P*	*z*	*P*	*z*	*P*	*z*	*P*	*z*	*P*
−2,48	0,0066	−1,48	0,0694	−0,48	0,3156	0,48	0,6844	1,48	0,9306	2,48	0,9934
−2,49	0,0064	−1,49	0,0681	−0,49	0,3121	0,49	0,6879	1,49	0,9319	2,49	0,9936
−2,50	0,0062	−1,50	0,0668	−0,50	0,3085	0,50	0,6915	1,50	0,9332	2,50	0,9938
−2,51	0,0060	−1,51	0,0655	−0,51	0,3050	0,51	0,6950	1,51	0,9345	2,51	0,9940
−2,52	0,0059	−1,52	0,0643	−0,52	0,3015	0,52	0,6985	1,52	0,9357	2,52	0,9941
−2,53	0,0057	−1,53	0,0630	−0,53	0,2981	0,53	0,7019	1,53	0,9370	2,53	0,9943
−2,54	0,0055	−1,54	0,0618	−0,54	0,2946	0,54	0,7054	1,54	0,9382	2,54	0,9945
−2,55	0,0054	−1,55	0,0606	−0,55	0,2912	0,55	0.7088	1,55	0,9394	2,55	0,9946
−2,56	0,0052	−1,56	0,0594	−0,56	0,2877	0,56	0,7123	1,56	0,9406	2,56	0,9948
−2,57	0,0051	−1,57	0,0582	−0,57	0,2843	0,57	0,7157	1,57	0,9418	2,57	0,9949
−2,58	0,0049	−1,58	0,0571	−0,58	0,2810	0,58	0,7190	1,58	0,9429	2,58	0,9951
−2,59	0,0048	−1,59	0,0559	−0,59	0,2776	0,59	0,7224	1,59	0,9441	2,59	0,9952
−2,60	0,0047	−1,60	0,0548	−0,60	0,2743	0,60	0,7257	1,60	0,9452	2,60	0,9953
−2,61	0,0045	−1,61	0,0537	−0,61	0,2709	0,61	0,7291	1,61	0,9463	2,61	0,9955
−2,62	0,0044	−1,62	0,0526	−0,62	0,2676	0,62	0,7324	1,62	0,9474	2,62	0,9956
−2,63	0,0043	−1,63	0,0516	−0,63	0,2643	0,63	0,7357	1,63	0,9484	2,63	0,9957
−2,64	0,0041	−1,64	0,0505	−0,64	0,2611	0,64	0,7389	1,64	0,9495	2,64	0,9959
−2,65	0,0040	−1,65	0,0495	−0,65	0,2578	0,65	0,7422	1,65	0,9505	2,65	0,9960
−2,66	0,0039	−1,66	0,0485	−0,66	0,2546	0,66	0,7454	1,66	0,9515	2,66	0,9961
−2,67	0,0038	−1,67	0,0475	−0,67	0,2514	0,67	0,7486	1,67	0,9525	2,67	0,9962
−2,68	0,0037	−1,68	0,0465	−0,68	0,2483	0,68	0,7517	1,68	0,9535	2,68	0,9963
−2,69	0,0036	−1,69	0,0455	−0,69	0,2451	0,69	0,7549	1,69	0,9545	2,69	0,9964
−2,70	0,0035	−1,70	0,0446	−0,70	0,2420	0,70	0,7580	1,70	0,9554	2,70	0,9965
−2,71	0,0034	−1,71	0,0436	−0,71	0,2389	0,71	0,7611	1,71	0,9564	2,71	0,9966
−2,72	0,0033	−1,72	0,0427	−0,72	0,2358	0,72	0,7642	1,72	0,9573	2,72	0,9967
−2,73	0,0032	−1,73	0,0418	−0,73	0,2327	0,73	0,7673	1,73	0,9582	2,73	0,9968

Окончание табл. П.1

z	***P***	***z***	***P***	***z***	***P***	***z***	***P***	***z***	***P***	***z***	***P***
−2,74	0,0031	−1,74	0,0409	−0,74	0,2296	0,74	0,7704	1,74	0,9591	2,74	0,9969
−2,75	0,0030	−1,75	0,0401	−0,75	0,2266	0,75	0,7734	1,75	0,9599	2,75	0,9970
−2,76	0,0029	−1,76	0,0392	−0,76	0,2236	0,76	0,7764	1,76	0,9608	2,76	0,9971
−2,77	0,0028	−1,77	0,0384	−0,77	0,2206	0,77	0,7794	1.77	0,9616	2,77	0,9972
−2,78	0,0027	−1,78	0,0375	−0,78	0,2177	0,78	0,7823	1,78	0,9625	2,78	0,9973
−2,79	0,0026	−1,79	0,0367	−0,79	0,2148	0,79	0,7852	1,79	0,9633	2,79	0,9974
−2,80	0,0026	−1,80	0,0359	−0,80	0,2119	0,80	0,7881	1,80	0,9641	2,80	0,9974
−2,81	0,0025	−1,81	0,0351	−0,81	0,2090	0,81	0,7910	1,81	0,9649	2,81	0,9975
−2,82	0,0024	−1,82	0,0344	−0,82	0,2061	0,82	0,7939	1,82	0,9656	2,82	0,9976
−2,83	0,0023	−1,83	0,0336	−0,83	0,2033	0,83	0,7967	1,83	0,9664	2,83	0,9977
−2,84	0,0023	−1,84	0,0329	−0,84	0,2005	0,84	0,7995	1,84	0,9671	2,84	0,9977
−2,85	0,0022	−1,85	0,0322	−0,85	0,1977	0,85	0,8023	1,85	0,9678	2,85	0,9978
−2,86	0,0021	−1,86	0,0314	−0,86	0,1949	0,86	0,8051	1,86	0,9686	2,86	0,9979
−2,87	0,0021	−1,87	0,0307	−0^7	0,1922	0,87	0,8078	1,87	0,9693	2,87	0,9979
−2,88	0,0020	−1,88	0,0301	−0,88	0,1894	0,88	0,8106	1,88	0,9699	2,88	0,9980
−2,89	0,0019	−1,89	0,0294	−0,89	0,1867	0,89	0,8133	1,89	0,9706	2,89	0,9981
−2,90	0,0019	−1,90	0,0287	−0,90	0,1841	0,90	0,8159	1,90	0,9713	2.90	0,9981
−2,91	0,0018	−1,91	0,0281	−0,91	0,1814	0,91	0,8186	1,91	0,9719	2,91	0,9982
−2,92	0,0018	−1,92	0,0274	−0,92	0,1788	0,92	0,8212	1,92	0,9726	2,92	0,9982
−2,93	0,0017	−1,93	0,0268	−0,93	0,1762	0,93	0,8238	1,93	0,9732	2,93	0,9983
−2,94	0,0016	−1,94	0,0262	−0,94	0,1736	0,94	0,8264	1,94	0,9738	2,94	0,9984
−2,95	0,0016	−1,95	0,0256	−0,95	0,1711	0,95	0,8289	1,95	0,9744	2,95	0,9984
−2,96	0,0015	−1,96	0,0250	−0,96	0,1685	0,96	0,8315	1,96	0,9750	2.96	0,9985
−2,97	0,0015	−1,97	0,0244	−0,97	0,1660	0,97	0,8340	1,97	0,9756	2,97	0,9985
−2,98	0,0014	−1,98	0,0239	−0,98	0,1635	0,98	0,8365	1,98	0,9761	2,98	0,9986
−2,99	0,0014	−1,99	0,0233	−0.99	0,1611	0,99	0,8389	1,99	0,9767	2,99	0,9986
−3,00	0,0013	−2,00	0,0228	−1,00	0,1587	1,00	0,8413	2,00	0,9772	3,00	0,9987

Таблица П.2

Случайные числа

	1	2	3	4	5	6	7	8	9	10
1	51449	39284	85527	67168	91284	19954	91166	70918	85957	19492
2	16144	56830	67507	97275	25982	69294	32841	20861	83114	12531
3	48145	48280	99481	13050	81818	25282	66466	24461	97021	21072
4	83780	48351	85422	42978	26088	17869	94245	26622	48318	73850
5	95329	38482	93510	39170	63683	40587	80451	43058	81923	97072
6	11179	69004	34273	36062	26234	58601	47159	82248	95968	99722
7	94631	52413	31524	02316	27611	15888	13525	43809	40014	30667
8	64275	10294	35027	25604	65695	36014	17988	02734	31732	29911
9	72125	19232	10782	30615	42005	90419	32447	53688	36125	28456
10	16463	42028	27927	48403	88963	79615	41218	43290	53618	68082
11	10036	66273	69506	19610	01479	92338	55140	81097	73071	61544
12	85356	51400	88502	68267	73943	25828	38219	13268	09016	77465
13	84076	82087	55053	75370	71030	92275	55497	97123	40919	57479
14	76731	39755	78537	51937	11680	78820	50082	56068	36908	55399
15	19032	73472	79399	05549	14772	32746	38841	45524	13535	03113
16	72791	59040	61529	74437	74482	76619	05232	28616	98690	24011
17	11553	00135	28306	65571	34465	47423	39198	54456	95283	54637
18	71405	70352	46763	64002	62461	41982	15933	46942	36941	93412
19	17594	10116	55483	96219	85493	96955	89180	59690	82170	77643
20	09584	23476	09243	65568	89128	36747	63692	09986	47687	46448
21	81677	62634	52794	01466	85938	14565	79993	44956	82254	65223
22	45849	01177	13773	43523	69825	3222	58458	77463	58521	7273
23	97252	92257	90419	01241	52516	66293	14536	23870	78402	41759
24	26232	77422	76289	57587	42831	87047	20092	92676	12017	43554
25	87799	33602	01931	66913	63008	03745	93939	07178	70003	18158

Окончание табл. П.2

	1	2	3	4	5	6	7	8	9	10
26	46120	62298	69129	07862	76731	58527	39342	42749	57050	91725
27	53292	55652	11834	47581	25682	64085	26587	92289	41853	38354
28	81606	56009	06021	98392	40450	87721	50917	16978	39472	23505
29	67819	47314	96988	89931	49395	37071	72658	53947	11996	64631
30	50458	20350	87362	83996	86422	58694	71813	97695	28804	58523
31	59772	27000	97805	25042	09916	77569	71347	62667	09330	02152
32	94752	91056	08939	93410	59204	04644	44336	55570	21106	76588
33	01885	82054	45944	55398	55487	56455	56940	68787	36591	29914
34	85190	91941	86714	76593	77199	39724	99548	13827	84961	76740
35	97747	67607	14549	08215	95408	46381	12449	03672	40325	77312
36	43318	84469	26047	86003	34786	38931	34846	28711	42833	93019
37	47874	71365	76603	57440	49514	17335	71969	58055	99136	73589
38	24259	48079	71198	95859	94212	55402	93392	31965	94622	11673
39	31947	64805	34133	03245	24546	48934	41730	47831	26531	02203
40	37911	93224	87153	54541	57529	38299	65659	00202	07054	40168
41	82714	15799	93126	74180	94171	97117	31431	00323	62793	11995
42	82927	37844	74411	45887	36713	52339	68421	35968	67714	05883
43	65934	21782	35804	36676	35404	69987	52268	19894	81977	87764
44	56953	04356	68903	21369	35901	86797	83901	68681	02397	55359
45	16278	17165	67843	49349	90163	97337	35003	34915	91485	33814
46	96339	95028	48468	12279	81039	56531	10759	19579	00015	22829
47	84110	49661	13988	75909	35580	18426	29038	79111	56049	96451
48	49017	60748	03412	09880	94091	90052	43596	21424	16584	67970
49	43560	05552	54344	69418	01327	07771	25364	77373	34841	75927
50	25206	15177	63049	12464	16149	18759	95184	15968	89446	07168

Таблица П.3

Распределение Стьюдента

Доверительный уровень								
Двусторонний		80%	90%	95%	98%	99%	99,80%	99,90%
Односторонний		90%	95%	97,5%	99%	99,50%	99,90%	99,95%
Уровень проверки гипотезы								
Двусторонняя		0,2	0,10	0,05	0,02	0,01	0,002	0,001
Односторонняя		0,1	0,05	0,025	0,01	0,005	0,001	0,0005
Для одной выборки: *n*	**В целом число степеней свободы**	**Критические значения**						
2	1	3,078	6,314	12,706	31,821	63,657	318,309	636,619
3	2	1,886	2,92	4,303	6,965	9,925	22,327	31,599
4	3	1,638	2,353	3,182	4,541	5,841	10,215	12,924
5	4	1,533	2,132	2,776	3,747	4,604	7,173	8,61
6	5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
7	6	1,44	1,943	2,447	3,143	3,707	5,208	5,959
8	7	1,415	1,896	2,365	2,998	3,499	4,785	5,408
9	8	1,397	1,86	2,305	2,896	3,355	4,501	5,041
10	9	1,383	1,833	2,262	2,821	3,25	4,297	4,781
11	10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
12	11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
13	12	1,356	1,782	2,179	2,681	3,055	3,93	4,318
14	13	1,35	1,771	2,16	2,65	3,012	3,852	4,221
15	14	1,345	1,761	2,145	2,624	2,977	3,787	4,14
16	15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
17	16	1,337	1,746	2,12	2,583	2,921	3,686	4,015
18	17	1,333	1,74	2,11	2,567	2,898	3,646	3,965
19	18	1,33	1,734	2,101	2,552	2,878	3,61	3,922

Окончание табл. П.3

Для одной выборки: *n*	**В целом число степеней свободы**	**Критические значения**						
20	19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
21	20	1,325	1,725	2,086	2,528	2,845	3,552	3,85
22	21	1,323	1,721	2,08	2,518	2,831	3,527	3,819
23	22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
24	23	1,319	1,714	2,069	2,5	2,807	3,485	3,768
25	24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
26	25	1,316	1,708	2,06	2,485	2,787	3,45	3,725
27	26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
28	27	1,314	1,703	2,052	2,473	2,771	3,421	3,69
29	28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
30	29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
31	30	1,31	1,697	2,042	2,457	2,75	3,385	3,646
32	31	1,309	1,696	2,04	2,453	2,744	3,375	3,633
33	32	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	33	1,308	1,692	2,035	2,445	2,733	3,356	3,611
35	34	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	35	1,306	1,69	2,03	2,438	2,724	3,340	3,591
37	36	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	37	1,305	1,687	2,026	2,431	2,715	3,326	3,574
39	38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	39	1,304	1,685	2,023	2,426	2,708	3,313	3,558
Бесконечность		1,282	1,645	1,96	2,326	2,576	3,09	3,291

Предметный указатель

Именной указатель

Список использованной литературы

1. *Афанасьева О. Н.*, *Бродский Я. С.*, *Павлов А. Л.* Математика для техникумов на базе среднего образования: Учеб. пособие. — М.: Издательство физико-математической литературы, 2005.

2. *Афанасьєва О. М.*, *Бродський Я. С.*, *Павлов О. Л.*, *Сліпенко А. К.* Алгебра і початки аналізу. 10 клас. — Тернопіль: Навчальна книга — Богдан, 2004.

3. *Виленкин Н. Я.* Комбинаторика. — М.: Наука, 1969.

4. *Виленкин Н. Я.* Популярная комбинаторика. — М.: Наука, 1975.

5. *Гнеденко Б. В.*, *Хинчин А. Я.* Элементарное введение в теорию вероятностей. — М.: Наука, 1964.

6. *Дайменд С.* Мир вероятностей. — М.: Статистика, 1970.

7. *Ежов И. И.*, *Скороход А. В.*, *Ядренко М. И.* Элементы комбинаторики. — М.: Наука, 1977.

8. *Кимбл Г.* Как правильно пользоваться статистикой. — М.: Финансы и статистика, 1982.

9. *Колмогоров А. Н.*, *Журбенко И. Г.*, *Прохоров А. В.* Введение в теорию вероятностей. — М.: Наука, Главная редакция физико-математической литературы, 1982.

10. *Майстров Л. Е.* Теория вероятностей. Исторический очерк. — М.: Наука, 1967.

11. *Мостеллер Ф.*, *Рурке Р.*, *Томас Дж.* Вероятность. — М.: Мир, 1969.

12. *Сигел*, *Эндрю*. Практическая бизнес-статистика.: Пер. с англ. — М.: Издательский дом «Вильямс», 2002.

13. *Скороход А. В.* Вероятность вокруг нас. — Киев: Наукова думка, 1980.

14. *Яглом А. М.*, *Яглом И. М.* Вероятность и информация. — М.: Наука, 1973.

Учебное издание

ШКОЛЬНЫЙ КУРС МАТЕМАТИКИ

Бродский Яков Соломонович

**СТАТИСТИКА.
ВЕРОЯТНОСТЬ.
КОМБИНАТОРИКА**

Редактор *Н. В. Валуева*
Корректор *Е. В. Морозова*
Технический редактор *Е. А. Вишнякова*
Компьютерная верстка *Е. Ю. Пучковой*

Подписано в печать 27.09.2007. Формат 60x90 1/16.
Гарнитура «Школьная». Печать офсетная.
Усл. печ. л. 34,00. Тираж 5000 экз. Заказ № .

Общероссийский классификатор продукции
ОК-005-93, том 2; 953005 — учебная литература

ООО «Издательство Оникс».
127422, Москва, ул. Тимирязевская, д. 38/25.
Почтовый адрес: 117418, Москва, а/я 26.
Отдел реализации: тел. (499) 619-02-20, 610-02-50.
Internet: www.onyx.ru; e-mail: mail@onyx.ru

ООО «Издательство «Мир и Образование».
Изд. лиц. ИД № 05088 от 18.06.2001.
109193, Москва, ул. 5-я Кожуховская, д. 13, стр. 1.
Тел./факс (495) 129-09-60, 120-51-47, 742-43-54.
E-mail: mir-obrazovanie@onyx.ru

Издание осуществлено при техническом содействии
ООО «Издательство АСТ»