МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Г.В. Вавилова

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ РЕШЕНИЯ ИНЖЕНЕРНЫХ ЗАДАЧ

*Рекомендовано в качестве учебного пособия
Редакционно-издательским советом
Томского политехнического университета*

Издательство
Томского политехнического университета
2022

УДК 519.22(075.8)
ББК 22.172я73
В12

Вавилова Г.В.

В12 Применение математической статистики для решения инженерных задач : учебное пособие / Г.В. Вавилова ; Томский политехнический университет. – Томск : Изд-во Томского политехнического университета, 2022. – 197 с.

ISBN 978-5-4387-1073-8

В учебном пособии приведены необходимые сведения из теории вероятностей и математической статистики, применяемые для обработки результатов измерения. Рассмотрены основные положения теории оценивания и проверки статистических гипотез. Изложены основы корреляционного и регрессионного анализа. Подробно разобраны решения типовых задач, предложены задачи для самостоятельного решения.

Предназначено для обучающихся по направлению подготовки бакалавров «Приборостроение» и других смежных направлений. Может быть рекомендовано для самостоятельной работы студентов.

УДК 519.22(075.8)
ББК 22.172я73

Рецензенты

Кандидат технических наук
главный метролог ООО «НПО “Редвилл”»
Е.М. Федоров

Главный метролог АО «Сибкабель»
Г.И. Вакурова

ISBN 978-5-4387-1073-8

ВВЕДЕНИЕ

Цель любого измерительного эксперимента – получение сведений о количественной характеристике измеренной величины. Но результат необходимо не только получить, но и уметь проанализировать. Наша задача – научиться оценивать достоверность полученных результатов измерения и правильно представлять их.

Различают истинное значение измеренной величины, идеально отражающее свойства материального объекта, и действительное значение, найденное экспериментально, с помощью измерения. Разница между этими значениями называется погрешностью измерения. Погрешность измерения – величина неизвестная до проведения эксперимента, и какой она будет, точно сказать невозможно. Можно лишь предположить это с определенной долей вероятности. При оценке составляющих погрешности и неопределенности результата измерения используются такие характеристики, как среднее значение, среднее квадратическое отклонение, точечная и интервальная оценка и др. Для того, чтобы перейти к теории погрешностей и обработке результатов измерения, необходимы знания из теории вероятностей и математической статистики.

Учебное пособие состоит из 6 глав и справочного материала, размещенного в приложении. Каждая глава содержит теоретический материал из теории вероятностей и математической статистики, примеры решения задач в необходимом объеме для дальнейшего использования в решении прикладных задач (теория погрешностей, обработка результатов измерения, измерительная техника и т. д.).

Данное учебное пособие является продолжением и расширенной версией пособия автора «Математическая обработка результатов измерения», изданного в 2013 году.

Глава 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1.1. Краткая историческая справка

Человечество всегда стремилось к некоторого рода предсказаниям. Любая наука основана на этом. Однако предвидение фактов не может быть абсолютным, каким бы обоснованным оно не казалось. У нас не может быть абсолютной уверенности в том, что наше предвидение не будет опровергнуто опытом.

Теория вероятностей – это раздел математики, изучающий случайные события, случайные величины, их свойства и операции над ними и позволяющий по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом между собой.

Долгое время теория вероятностей считалась чисто опытной наукой и «не совсем математикой». Она возникла в Средние века при первых попытках математического анализа азартных игр (орлянка, кости, рулетка).

В истории развития теории вероятностей следует выделить несколько периодов [1–4].

Период 1. Предыстория (*до XVI века*). В античные времена и в Средневековье ученые ограничивались рассуждениями о происхождении случайности и её роли в природе. Математики пытались решать задачи, связанные с теорией вероятностей, но на данном этапе еще не появилось никаких общих методов и понятий. В это период начинают развивать комбинаторные методы [1].

Период 2. Начало формирования понятий (вторая половина XVII века). Основным стимулом развития теории вероятностей стало желание выиграть в азартных играх. Кроме того, теорию вероятностей стали применять для решения прикладных задач в области демографической статистики, страхового дела и теории приближённых вычислений.

Особый вклад в развитие теории вероятностей внесли известные математики Б. Паскаль (рис. 1.1) и П. Ферма (рис. 1.2), которые пытались решить задачи, предложенные профессиональным игроком в кости и карты шевалье де Мере [2]. В предложенных ими решениях можно увидеть зачатки *математического ожидания*, а также в еще не несовершенной форме «*теорему о сложении и умножении вероятностей*». Так как термин «*вероятность*» еще не было введен, Б. Паскаль и П. Ферма оперировали термином «*количество случаев*», благоприятствующих тому или иному событию [1].

Рис. 1.1. Блез Паскаль
(19 июня 1623 г. – 19 августа 1662 г.) – французский математик, механик, физик, литератор и философ

Рис. 1.2. Пьер де Ферма
(17 августа 1601 г. – 12 января 1665 г.) – французский математик-самоучка, один из создателей аналитической геометрии, математического анализа, теории вероятностей и теории чисел

Б. Паскаль существенно продвинул развитие комбинаторики, указав на её значение для теории вероятностей.

В 1657 г. появилась первая крупная работа по теории вероятностей, автором которой был Х. Гюйгенс (рис. 1.3). Он, наряду с Ферма и Паскалем, является родоначальником теории вероятностей [1, 3]. Гюйгенс ввёл фундаментальные термины «*математическое ожидание*» и *«числовая мера вероятности события»*.

Рис. 1.3. Христиан Гюйгенс
(14 апреля 1629 г. – 8 июля 1695 г.) – нидерландский механик, физик, математик, астроном и изобретатель

Рис. 1.4. Якоб Бернулли
(27 декабря 1654 г. – 16 августа 1705 г.) – швейцарский математик, один из основателей теории вероятностей и математического анализа

Гюйгенс проанализировал и решил задачу о разделе ставки: ставку надо разделить пропорционально вероятностям выигрыша при продолжении игры. Он также впервые применил вероятностные методы к демографической статистике и показал, как рассчитать среднюю продолжительность жизни [1, 3].

Период 3. Систематизация теории вероятностей (XVIII век). Первым трудом по теории вероятностей считается изданная посмертно монография Якоба Бернулли (рис. 1.4) «Искусство предположений» (1713 г.) [3, 4]. В ней Бернулли предложил классическое определение понятия «*вероятность случайного события*».

Период 4. Развитие теории вероятностей (начало XIX века). В это время теория вероятностей становится «модной наукой». Ее начинают применять даже там, где это совершенно не оправдано. Теорию вероятностей пытаются применить для изучения различных общественных явлений в так называемых «моральных» или «нравственных» науках [1–4].

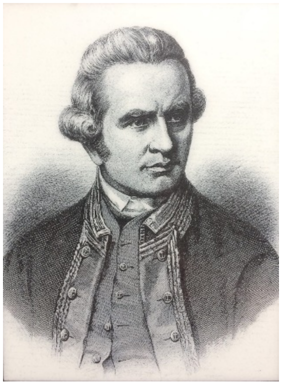


Рис. 1.5. Пьер-Симон, маркиз де Лаплас (23 марта 1749 г. – 5 марта 1827 г.) – французский математик, механик, физик и астроном

Рис. 1.6. Иоганн Карл Фридрих Гаусс (30 апреля 1777 г. – 23 февраля 1855 г.) – немецкий математик, механик, физик, астроном и геодезист

Рис. 1.7. Симеон Дени Пуассон (21 июня 1781 г. –25 апреля 1840 г.) – французский математик, механик и физик

Рис. 1.8. Пафнутий Львович Чебышев (16 мая 1821 г. – 8 декабря 1894 г.) – русский математик, механик

Идеи Бернулли нашли развитие в трудах П.-С. Лапласа (рис. 1.5), К.Ф. Гаусса (рис. 1.6), С.Д. Пуассона (рис. 1.7). Область применения вероятностных методов в прикладной статистике значительно расширилась.

Был определен термин «*вероятность для непрерывных случайных величин*», благодаря чему появилась возможность применения методов математического анализа. Появляются первые попытки применения теории вероятностей в физике. К концу XIX века формируются такие отрасли, как статистическая физика, строгая теория ошибок измерения. Вероятностные методы проникают в самые различные прикладные науки [2, 3].

Большой вклад в развитие теории вероятностей внесла Петербургская математическая школа. При содействии П.Л. Чебышева (рис. 1.8) и его учеников А.А. Маркова (рис. 1.9) и А.М. Ляпунова (рис. 1.10) теория вероятностей становится стройной математической наукой.

Рис. 1.9. Андрей Андреевич Марков (2 июня 1856 г. – 20 июля 1922 г.) – русский математик, академик

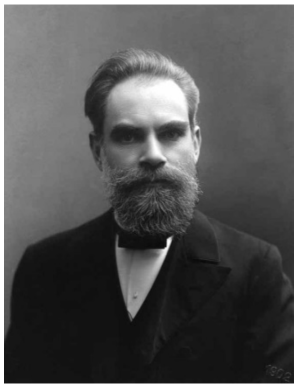


Рис. 1.10. Александр Михайлович Ляпунов (25 мая 1857 г. – 3 ноября 1918 г.) – русский математик и механик, академик Петербургской академии наук с 1901 г.

Чебышев получил общую формулировку закона больших чисел: «если математические ожидания серии случайных величин и квадраты этих математических ожиданий ограничены в совокупности, то среднее арифметическое этих величин с ростом сходится по вероятности к среднему арифметическому для их математических ожиданий». Марков внёс большой вклад в теорию вероятностей, доказав вариант закона больших чисел для некоторых распространённых типов зависимых величин.

Ляпунов, обобщая исследования Чебышева и Маркова, доказал центральную предельную теорему, устанавливающую общие достаточные условия для сходимости распределения сумм независимых случайных величин к нормальному закону [2–4].

Период 5. Применение теории вероятностей в других науках (XX век – наши дни). Сегодня теория вероятностей применяется в теории надежности, теории массового обслуживания, в теоретической физике, теории микромира, теории наследственности, геодезии, астрономии, теории стрельбы, теории ошибок наблюдений, теории автоматического управления, общей теории связи и во многих других теоретических и прикладных науках.

Теория вероятностей используется при планировании и организации производства, при анализе технологических процессов, предупредительном и приемочном контроле качества продукции и для многих других целей.

К. Пирсон (рис. 1.11) разработал методы математической статистики, широко и повсеместно применяемые для анализа прикладных измерений, проверки гипотез и принятия решений. А.Н. Колмогоров (рис. 1.12) сформулировал классическую аксиому, используемую в теории вероятностей [2, 3].

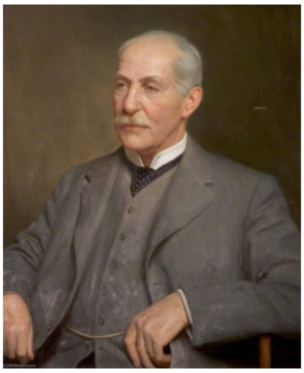


Рис. 1.11. Карл Пирсон (27 марта 1857 г. – 27 апреля 1936 г.) – английский математик, статистик, биолог и философ; основатель математической статистики, один из основоположников биометрики

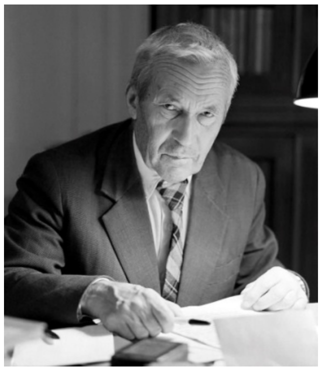


Рис. 1.12. Андрей Николаевич Колмогоров (25 апреля 1903 г. – 20 октября 1987 г.) – советский математик, один из крупнейших математиков XX века

В последние годы методы теории вероятностей все больше проникают в различные области науки и техники, способствуя их прогрессу.

1.2. Элементарные события

Теория вероятностей изучает вероятностные закономерности массовых однородных случайных событий.

Некоторая последовательность действий, которые выполняются при соблюдении определенных условий – это *опыт (эксперимент)*. Любой проводимый опыт имеет результат. В теории вероятностей результат опыта называется *исходом опыта*.

Пример 1[1]***.*** Рассмотрим различные примеры исходов опытов:

1. При бросании монеты возможны два исхода: появление «герба» – один исход, появление «цифры» – другой исход.
2. При бросании игральной кости возможны шесть исходов, а именно появление цифр: 1, 2, 3, 4, 5 и 6.
3. При вынимании шаров из урны, в которой находится два черных и три белых шара, возможны пять исходов: два исхода, когда вынимается черный шар, и три исхода, когда вынимается белый шар.

Событие – это исход или группа исходов, удовлетворяющих определённым требованиям.

Пример 2. Рассмотрим варианты события:

1. Появление «герба» при бросании монеты.
2. Появление двух «гербов» при двукратном бросании монеты.
3. Появления туза при вынимании игральной карты из колоды.
4. Попадание в цель с первого выстрела.

Пример 3. Если считать событием появление «герба» при бросании монеты, а появилась «цифра», то событие не произошло.

Если считать событием появление четного числа (2, 4, 6) при бросании игральной кости, а выпало одно из нечётных чисел (1, 3, 5), то событие не произошло.

Исходы опытов, соответствующие установленному событию, называются *благоприятными исходами*, в противном случае – *неблагоприятными исходами*.

События бывают достоверными, невозможными и случайными.

Достоверное событие – это событие, которое всегда происходит в рассматриваемых условиях.

Пример 4. Если в сосуде содержится вода при нормальном атмосферном давлении и температуре 20 °C, то событие «вода в сосуде находится в жидком состоянии» – достоверное при соблюдении постоянства атмосферного давления и температуры воды [6, 7].

Невозможное событие – это событие, которое никогда не может наступить в рассматриваемых условиях.

Пример 5. Для совокупности условий *примера 4* событие «вода в сосуде находится в твердом состоянии» – невозможное, так не может наступить при соблюдении заданных условий [6, 7].

---

[1] В главе 1 большинство примеров и их решения приводятся из учебного пособия Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике [5].

Случайное событие – это событие, которое при воспроизведении опыта может наступить, а может и не наступить.

Пример 6. При подбрасывании монеты может выпасть либо «герб», либо «цифра». Каждое из данных событий («выпадение герба» либо «выпадение цифры») – случайное.

Пример 7. В урне находится два белых шара и три черных. Вынимается один шар. *Достоверным событием* для данного эксперимента будет вынимание шара любого цвета; *невозможным событием* – вынимание шара зеленого цвета; *случайным событием* – вынимание шара белого цвета.

Каждое случайное событие есть следствие действия очень многих случайных причин (в *примере 6* с бросанием монеты: сила, с которой брошена монета, форма монеты, наличие дефектов и др.). Невозможно учесть влияние на результат всех этих причин, поскольку число их очень велико и законы их действия неизвестны. Поэтому теория вероятностей не ставит перед собой задачу предсказать, произойдет единичное событие или нет, – она просто не в силах это сделать.

Если рассматриваются случайные события, которые могут многократно наблюдаться в одних и тех же условиях (массовые однородные случайные события), то однородные случайные события независимо от их конкретной природы подчиняются определенным закономерностям. Установлением этих закономерностей и занимается теория вероятностей.

В результате опыта обязательно должно произойти хотя бы одно из событий, которые образуют *полную группу событий*.

Пример 8. Примеры событий, образующих полную группу событий:

1. Появление «герба» или появление «цифры» при бросании монеты.
2. Попадание или промах при выстреле в цель.
3. Появление цифр 1, 2, 3, 4, 5, 6 при бросании игральной кости.
4. Появление четного или нечётного числа при бросании игральной кости.

Несколько событий в опыте называют *несовместными*, если появление одного из них исключает появление других событий в одном и том же испытании. События, приведенные в *примере 8*, *несовместные*, так как они не могут появиться одновременно. Появление «герба» исключает появление «цифры», следовательно, данные события – несовместные.

Пример 9. Приобретены два билета денежно-вещевой лотереи. Обязательно произойдет только одно из следующих событий: «выигрыш выпал на первый билет и не выпал на второй», «выигрыш не выпал на первый билет и выпал на второй», «выигрыш выпал на оба билета», «на оба билета выигрыш не выпал». Эти события образуют полную группу попарно несовместных событий.

Пример 10. Стрелок произвел выстрел по цели. Обязательно произойдет одно из следующих событий: «попадание», «промах». Эти два несовместных события образуют полную группу.

События в опыте называются *равновозможными*, если есть веские основания считать, что ни одно из этих событий не является объективно более возможным, чем другое.

Пример 11. Появление «герба» и появление «цифры» при бросании монеты – события равновозможные. Действительно, если монета изготовлена из однородного материала, имеет правильную цилиндрическую форму и не имеет никаких дефектов, то выпадение «герба» или «цифры» – события равновозможные.

Пример 12. Появление того или иного числа очков на брошенной игральной кости – равновозможные события. Действительно, если игральная кость изготовлена из однородного материала, имеет форму правильного многогранника и не имеет дефектов, то выпадение каждого числа имеет одинаковую вероятность.

1.2.1. Вероятность событий

Можно утверждать, что одни события более возможны, а другие менее возможны.

Пример 13. Появление «герба» при одном бросании монеты более возможно, чем появление двух «гербов» при двух последовательных бросаниях монеты.

Чтобы количественно сравнивать между собой события по степени их возможности, нужно с каждым событием связать некоторое число, которое тем больше, чем более возможно это событие.

Вероятность события – это численная мера степени объективной возможности этого события, которую можно определить как отношение благоприятного числа исходов опыта к общему числу всех *равновозможных несовместных элементарных* исходов, образующих *полную группу*:

$$P(A) = \frac{m}{n}, \tag{1.1}$$

где m – благоприятное число исходов опыта; n – общее число исходов опыта.

Выражение (1.1) называется *классической формулой* для вычисления вероятности события.

Пример 14. Определить вероятность выпадения «герба» при подбрасывании одной монеты.

Решение

В этом опыте возможны два исхода: появление «герба» и появление «цифры». Общее число исходов $n = 2$.

Событие наступает при выпадении «герба», следовательно, число благоприятных исходов $m = 1$.

По классической формуле (1.1) вероятность события A равна:

$$P(A) = \frac{m}{n} = \frac{1}{2}.$$

Ответ: при бросании монеты вероятность появления «герба» равна 1/2.

Пример 15. В урне находятся шары: два белых и три черных. Из урны вынимается один шар. Определить вероятность того, что из урны вынут белый шар.

Решение

В этом опыте число возможных исходов (по общему числу шаров) $n = 5$. Число благоприятных исходов (по числу белых шаров) $m = 2$. Вероятность события A (вынуть белый шар) определяется по классической формуле для вычисления вероятности (1.1):

$$P(A) = \frac{m}{n} = \frac{2}{5}.$$

Ответ: вероятность наступления события A равна 2/5.

Из *классической формулы* определения вероятности вытекают следующие свойства.

Свойство 1. Вероятность достоверного события равна единице.

Если событие *достоверное*, то каждый элементарный исход испытания благоприятствует событию, т. е. число благоприятных исходов равно общему числу исходов ($m = n$). Следовательно, выражение (1.1) примет вид

$$P(A) = \frac{m}{n} = \frac{n}{n} = 1.$$

Свойство 2. Вероятность невозможного события равна нулю.

Если событие *невозможное*, то ни один из исходов испытания не благоприятствует событию, т. е. число благоприятных исходов равно нулю ($m = 0$). Следовательно, выражение (1.1) примет вид

$$P(A) = \frac{m}{n} = \frac{0}{n} = 0.$$

Свойство 3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

Случайному событию благоприятствует лишь часть из общего числа исходов испытания. В этом случае

$$0 < m < n,$$

следовательно,

$$0 < P(A) = \frac{m}{n} < 1. \tag{1.2}$$

Классическая формула для вычисления вероятности (1.1) имеет ограниченное применение, так как может быть использована только для определения вероятности события в опытах с *равновероятными* исходами.

Пример 16. Указать на ошибку в решении задачи. Задача: «Брошены две игральные кости. Найти вероятность того, что сумма выпавших очков равна 4 (событие A).

Решение

Всего возможны 2 исхода испытания: сумма выпавших очков равна 4, сумма выпавших очков не равна 4. Событию A благоприятствует один исход; общее число исходов равно двум. Следовательно, искомая вероятность равна:

$$P(A) = 1/2».$$

Правильное решение

Ошибка предложенного решения состоит в том, что рассматриваемые исходы *не являются равновозможными*.

Общее число равновозможных исходов $n = 6 \cdot 6 = 36$ (каждое число выпавших очков на одной кости может сочетаться со всеми числами очков другой кости). Среди этих исходов благоприятствуют событию A только три исхода: (1:3), (3:1), (2:2) (см. рис. 1.13).

1:1	2:1	3:1	4:1	5:1	6:1
1:2	2:2	3:2	4:2	5:2	6:2
1:3	2:3	3:3	4:3	5:3	6:3
1:4	2:4	3:4	4:4	5:4	6:4
1:5	2:5	3:5	4:5	5:5	6:5
1:6	2:6	3:6	4:6	5:6	6:6

Рис. 1.13. Варианты выпадения очков на двух игральных костях

Следовательно, искомая вероятность по формуле (1.1) равна:

$$P(A) = \frac{3}{36} = \frac{1}{12}.$$

Ответ: вероятность того, что сумма выпавших очков равна 4, составляет 1/12.

1.2.2. Относительная частота события

Термин «*Вероятность*» подразумевает конечное число элементарных исходов, что на практике встречается крайне редко. В реальности число возможных исходов бесконечно. В таких случаях термин «*Вероятность*» неприменим.

Очень часто результат испытания представить в виде совокупности элементарных событий вообще невозможно. А также сложно обосновать, почему представленные элементарные события можно считать равновозможными. Обычно о равновозможности элементарных исходов испытаний говорят из соображения симметрии.

Пример 17. При рассмотрении игральной кости на основании симметрии предполагается, что она имеет правильную форму (форму куба) и равномерную плотность по всему объему, без каких-либо внутренних и поверхностных дефектов.

Из-за наличия всевозможных факторов предполагать, что элементарные события равновозможны, следует лишь в исключительных случаях. Поэтому в теории вероятностей появляется термин «*Статистическая вероятность*».

Под *статистической вероятностью* события понимается относительная частота или число, близкое к ней.

Относительная частота события – это отношение числа испытаний, в которых событие произошло, к общему числу проведенных испытаний. Относительная частота события A рассчитывается по формуле

$$W(A) = \frac{m}{n}, \qquad (1.3)$$

где m – число испытаний, в которых произошло событие А; n – общее число проведенных испытаний.

Сопоставляя термины «*вероятность»* и «*относительная частота»*, следует отметить:

1. При использовании термина «вероятность» подразумевается, что испытание в действительности не производится, а лишь предполагается его проведение.

2. При использовании термина «относительная частота» подразумевается, что испытание произведено и есть его результат.

Другими словами, «*вероятность»* вычисляется *до опыта*, «*относительная частота» – после опыта.*

Пример 18. Отдел технического контроля обнаружил 3 нестандартных детали в партии из 80 случайно отобранных деталей. Найти относительную частоту обнаружения нестандартной детали.

Решение

Относительная частота появления нестандартных деталей рассчитывается по формуле (1.3):

$$W(A) = \frac{m}{n} = \frac{3}{80}.$$

Ответ: относительная частота появления нестандартных деталей равна 3/80.

Длительные наблюдения показали, что если в одинаковых условиях производить опыты, в каждом из которых число испытаний достаточно велико, то *относительная частота* проявляет *свойство устойчивости*. В различных опытах *относительная частота* изменяется мало (тем меньше, чем больше произведено испытаний), колеблясь около некоторого постоянного числа. Следует отметить, что это постоянное число есть *вероятность* появления события.

Таким образом, если *опытным путем* установлена «*относительная частота*», то полученное число можно принять за приближенное значение «*Вероятности*».

Недостатком понятия «*относительная частота*» является его неоднозначность. В *примере 18* в качестве статистической вероятности события можно принять не только число $\frac{3}{80} = 0{,}0375$, но и 0,04 или 0,038, а также 0,03 и т. д., все зависит от точности округления.

Пример 19. Английский ученый К. Пирсон (рис. 1.11), определяя относительную частоту появления герба при подбрасывании монеты 12 000 и 24 000 раз, получил значения этой частоты соответственно 0,5016 и 0,5005. Нетрудно для данного опыта установить, пользуясь обычными представлениями, что относительные частоты появления герба или цифры должны быть близки одна к другой, а их «точное» значение, около которого колеблются опытные данные, должно быть равно 0,5.

Задачи для самостоятельного решения

Задача 1. В урне три белых, четыре красных и пять черных шаров.

а) найти вероятность того, что из урны будет вытащен красный шар;

б) найти вероятность того, что из урны будет вытащен цветной (не белый) шар;

Задача 2. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера. Полученные части тщательно перемешаны.

Найти вероятность того, что наудачу извлеченный кубик будет иметь окрашенных граней:

а) одну;
б) две;
в) три.

Задача 3. Брошены две игральные кости.

Найти вероятность того, что

а) сумма очков на выпавших гранях нечетная;
б) сумма выпавших очков равна семи;
в) сумма выпавших очков равна восьми, а разность – четырем;
г) сумма выпавших очков равна пяти, а произведение – четырем.

Задача 4. При испытании партии приборов относительная частота годных приборов оказалась равной 0,9. Для проверки взято 200 приборов. Найти число годных приборов среди них.

Задача 5. В партии из 100 деталей отдел технического контроля обнаружил пять нестандартных деталей. Найти относительную частоту появления нестандартных деталей.

Задача 6. По цели произвели 24 выстрела, причем было зарегистрировано 19 попаданий. Найти относительную частоту поражения цели.

Задача 7. При стрельбе из винтовки относительная частота попадания в цель оказалась равной 0,85. Найти число попадания, если всего произведено 120 выстрелов?

Задача 8. При перевозке ящика, в котором содержались 21 стандартная и 10 нестандартных деталей, утеряна 1 деталь, причем неизвестно какая. Наудачу извлеченная (после перевозки) из ящика деталь оказалась стандартной.

Найти вероятность того, что была утеряна:

а) стандартная деталь;
б) нестандартная деталь.

Задача 9. Задумано двузначное число.

Найти вероятность того, что задуманным числом окажется:

а) случайное названное двузначное число;
б) случайное названное число, цифры которого различны.

1.3. Основные формулы комбинаторики

При подсчете числа элементарных исходов используются формулы комбинаторики. *Комбинаторика* изучает количество комбинаций, подчиненных определенным условиям, которое можно составить из элементов, заданного конечного множества. Рассмотрим наиболее часто используемые комбинации.

Перестановка – это комбинация, состоящая из одних и тех же n различных элементов и отличающаяся от предыдущей только порядком расположения элементов.

Число перестановок определяется по формуле

$$P_n = n! \tag{1.4}$$

где $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$.

Пример 1. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, если каждая цифра входит в изображение числа только один раз?

Решение

Искомое число комбинаций рассчитывается по формуле (1.4):

$$P_3 = 3! = 1 \cdot 2 \cdot 3 = 6.$$

Ответ: из трех цифр можно составить шесть трехзначных чисел. Методом простого перебора легко это проверить: 123, 132, 213, 231, 312, 321.

Размещение – это комбинация, составленная из n различных элементов по m элементов, которая отличается от предыдущей либо составом, либо порядком расположения элементов. Число размещений определяется по формуле

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1)(n-2)\ldots(n-m+1). \tag{1.5}$$

Пример 2. Сколько можно составить сигналов из шести флажков различного цвета, взятых по два?

Решение

Искомое число комбинаций из двух флажков (m = 2) разного цвета определяется по формуле (1.5), так как при составлении сигнала важен порядок (размещение) флажков.

$$A_6^2 = \frac{6!}{(6-2)!} = \frac{6!}{4!} = \frac{4! \cdot 5 \cdot 6}{4!} = 5 \cdot 6 = 30.$$

Ответ: из шести флажков различного цвета можно составить 30 комбинаций.

Сочетание – это комбинация, составленная из n различных элементов по m элементов, которая отличается от предыдущей хотя бы одним элементом. Число сочетаний определяем по формуле

$$C_n^m = \frac{n!}{m!(n-m)!}. \qquad (1.6)$$

Пример 3. Сколькими способами можно выбрать две детали из ящика, содержащего 10 деталей?

Решение

При выборе деталей не важен порядок извлечения деталей, поэтому для решения используется формула (1.6) для комбинации «сочетание»:

$$C_{10}^2 = \frac{10!}{2!\cdot 8!} = 45.$$

Ответ: выбрать две детали из ящика, содержащего 10 деталей, можно 45 способами.

Число комбинаций «*Размещение*» A_n^m, *«Перестановка»* P_m и «*Сочетание*» C_n^m связаны между собой равенством:

$$A_n^m = P_m C_n^m. \qquad (1.7)$$

При решении задач комбинаторики используются правило суммы и правило произведения.

Правило суммы. Если некоторый объект A может быть выбран из совокупности объектов m способами, а объект B может быть выбран n способами, то выбрать либо объект A, либо объект B можно $m + n$ способами.

Правило произведения. Если объект A можно выбрать из совокупности объектов m способами и после каждого такого выбора объект B можно выбрать n способами, то пара объектов (A, B) в указанном порядке может быть выбрана $m \cdot n$ способами.

Пример 4. В партии из 10 деталей 7 стандартных. Найти вероятность того, что среди 6 взятых наудачу деталей 4 стандартных.

Решение

Общее число возможных элементарных исходов испытания равно числу способов, которыми можно извлечь 6 деталей из 10, т. е. числу сочетаний из 10 элементов по 6 элементов (формула (1.6)): $n = C_{10}^6$.

Число исходов, благоприятствующих событию A (среди шести взятых деталей четыре стандартных), равно числу *сочетаний* (формула (1.6)): четыре стандартные детали можно взять из семи стандартных деталей C_7^4 способами. При этом остальные 6 – 4 = 2 детали должны быть нестандартными. Взять две нестандартные детали из 10 – 7 = 3 нестандартных деталей можно C_3^2 способами. Число благоприятствующих исходов, согласно правилу произведения, $m = C_7^4 \cdot C_3^2$.

Искомая вероятность определяется по классической формуле определения вероятностей (1.1) и равна отношению числа исходов, благоприятствующих событию A, к числу всех элементарных исходов ($n = C_{10}^{6}$):

$$P(A) = \frac{C_7^4 \cdot C_3^2}{C_{10}^6} = \frac{1}{2}.$$

Ответ: вероятность того, что среди шести взятых наудачу деталей четыре будут стандартными, равна 1/2.

Задачи для самостоятельного решения

Задача 1. Иногда номера трамваев обозначают двумя цветными фонариками. Какое количество различных маршрутов можно обозначить, используя фонарики восьми различных цветов?

Задача 2. Сколько сигналов можно подать пятью различными флажками, поднимая их в любом количестве и в произвольном порядке?

Задача 3. В ящике имеется 50 одинаковых деталей, из них пять окрашенных. Наудачу вынимают одну деталь. Найти вероятность того, что извлеченная деталь окажется окрашенной.

Задача 4. На каждой из шести одинаковых карточек напечатана одна из следующих букв: а, т, м, р, с, о. Карточки тщательно перемешаны. Найти вероятность того, что на четырех вынутых по одной и расположенных «в одну линию» карточках можно будет прочесть слово «трос».

1.4. Теоремы сложения и умножения вероятностей

Суммой $A + B$ двух событий A и B называется событие, состоящее в появлении события A или события B, или событий A и B одновременно.

Пример 1. Из орудия произведены два выстрела. Событие A – попадание при первом выстреле, событие B – попадание при втором выстреле, тогда события $A + B$ – попадание при первом выстреле или при втором, или при обоих выстрелах.

В частности, если два события – A и B – несовместные, то событие $A + B$ – событие, которое состоит в появлении одного из событий A или B, безразлично какого.

Теорема сложения вероятностей совместных событий. Вероятность появления хотя бы одного из двух совместных событий, безразлично какого, равна сумме вероятностей этих событий без учета вероятности их совместного появления:

$$P(A+B) = P(A) + P(B) - P(AB). \quad (1.8)$$

Справедливость формулы (1.8) наглядно показана на рис. 1.14, *а*.

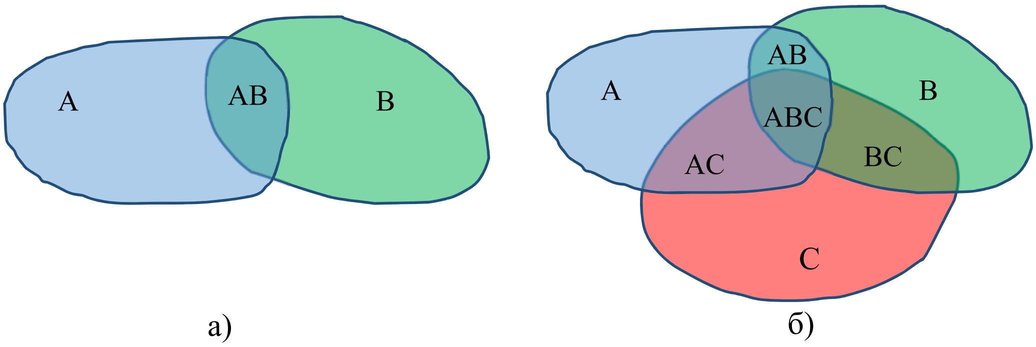


Рис. 1.14. Совместные события

Теорема сложения вероятностей несовместных событий. Вероятность появления одного из двух несовместных событий, безразлично какого, равна сумме вероятностей этих событий:

$$P(A+B)=P(A)+P(B). \qquad (1.9)$$

Вероятность появления одного из нескольких несовместных событий, безразлично какого, равна сумме вероятностей этих событий:

$$P(A_1+A_2+\ldots+A_n) = P(A_1)+P(A_2)+\ldots+P(A_n). \qquad (1.10)$$

Теорема сложения вероятностей событий, которые составляют полную группу. Сумма вероятностей событий A_1, A_2, …, A_n, образующих полную группу, равна единице:

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1. \qquad (1.11)$$

Вероятность суммы трех совместных событий вычисляется по формуле

$$\begin{aligned}P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-\\ -P(BC)+P(ABC).\end{aligned} \qquad (1.12)$$

Геометрическая интерпретация формулы (1.12) показана на рис. 1.14, *б*.

Получить формулу для расчета суммы вероятностей бо́льшего числа событий можно путем аналогичной геометрической интерпретации (рис. 1.14).

Пример 2. В урне 30 шаров: 10 красных, 5 синих и 15 белых. Найти вероятность появления цветного шара.

Решение

Появление цветного шара означает появление шара любого цвета, кроме белого. Следовательно, вероятность появления цветного шара рассматривается как сумма вероятностей появления шаров разных цветов (красного или синего).

Вероятность появления красного шара (событие A), согласно классической формуле определения вероятности (формула (1.1)), равна:

$$P(A)=\frac{10}{30}=\frac{1}{3}.$$

Вероятность появления синего шара (событие B) равна:

$$P(B)=\frac{5}{30}=\frac{1}{6}.$$

События A и B несовместны (появление шара одного цвета исключает появление шара другого цвета), поэтому теорема сложения вероятностей примет вид (формула (1.9))

$$P(A+B) = P(A)+P(B)=\frac{1}{3}+\frac{1}{6}=\frac{1}{2}.$$

Ответ: вероятность появления цветного шара (красного или синего) равна 1/2.

Пример 3. Стрелок стреляет по мишени, разделенной на 3 области. Вероятность попадания в первую область равна 0,45, во вторую – 0,35. Найти вероятность того, что стрелок при одном выстреле попадет либо в первую, либо во вторую область.

Решение

События «Стрелок попал в первую область» (событие A) и «Стрелок попал во вторую область» (событие B) – несовместны. Так как попадание в одну область исключает попадание в другую.

Искомая вероятность по теореме сложения вероятностей (формула (1.9)) равна:

$$P(A+B)=P(A)+P(B)=0{,}45+0{,}35=0{,}8.$$

Ответ: стрелок при одном выстреле попадет либо в первую, либо во вторую область с вероятностью 0,8 (80 %).

Пример 4. Образовательное учреждение получает пакеты с контрольными работами из городов A, B и C. Вероятность получения пакета из города A равна 0,7, из города B – 0,2. Найти вероятность того, что очередной пакет будет получен из города C.

Решение

События «Посылка из города A» (событие A), «Посылка из города B» (событие B), «Посылка из города C» (событие C) образуют *полную группу*, поэтому сумма вероятностей этих событий равна единице (формула (1.11)):

$$P(A)+P(B)+P(C)=1.$$

Отсюда искомая вероятность наступления события C равна:

$$P(C) = 1 - P(A) + P(B) = 1 - 0,7 - 0,2 = 0,1.$$

Ответ: вероятность, что «Посылка из города C» равна 0,1 (10 %).

Пример 5. Зашедший в магазин мужчина покупает что-нибудь с вероятностью 0,15, а зашедшая женщина – с вероятностью 0,7. У прилавка двое мужчин и одна женщина. Найти вероятность того, что по крайней мере одно лицо что-нибудь купит.

Решение

Событие A – покупку совершит первый мужчина, $P(A) = 0,15$;

Событие B – покупку совершит второй мужчина, $P(B) = 0,15$;

Событие C – покупку совершит женщина, $P(C) = 0,7$.

Вероятность совершения покупки определяется по формуле (1.12):

$$P(A + B + C) = 0,15 + 0,15 + 0,7 - 0,15^2 - 0,15 \cdot 0,7 - 0,15 \cdot 0,7 + \\ + 0,15^2 \cdot 0,7 = 0,78.$$

Ответ: вероятность совершения покупки 0,78 (78 %).

Произведением двух событий A и B называется событие AB, состоящее в совместном появлении событий (см. рис. 1.14, *а*).

Произведением нескольких событий называется событие, состоящее в появлении всех этих событий одновременно (см. рис. 1.14, *б*).

Пример 6. Рассматриваются события: событие A «Деталь годная», событие B «Деталь окрашенная». Тогда событие AB – «Деталь годна и окрашена».

Пример 7. Рассматриваются события: событие A «Выпадение “герба” при первом подбрасывании монеты», событие B «Выпадение “герба” при втором подбрасывании монеты», событие C «Выпадение “герба” при третьем подбрасывании монеты». Тогда событие ABC – «Выпадение “герба” во всех трех испытаниях».

Теорема умножения вероятностей. Вероятность совместного появления двух событий равна умножению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило:

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A), \tag{1.13}$$

где $P_A(B)$ – условная вероятность события B при условии, что событие A наступило; $P_B(A)$ – условная вероятность события A при условии, что событие B наступило.

Следует отметить, что порядок, в котором расположены события, может быть выбран любым, т. е. безразлично, какое событие считать первым.

Два события называются *независимыми*, если наступление одного из них не изменяет вероятность наступления другого, в противном случае события называют *зависимыми.*

Заключение о независимости событий между собой делается на основе анализа задачи, ее смысла.

Пример 8. Вероятности поражения цели каждым из двух орудий не зависят от того, поразило ли цель другое орудие, поэтому события «Первое орудие поразило цель» и «Второе орудие поразило цель» независимы.

Пример 9. У сборщика имеется 3 конусных и 7 эллиптических валиков. Сборщик взял один валик, а затем второй. Найти вероятность того, что первый из взятых валиков конусный, а второй – эллиптический.

Решение

Вероятность того, что «Первый валик окажется конусным» (событие A) вычислена по формуле (1.1), где число благоприятных исходов равно общему числу конусных валиков $m = 3$, а общее число исходов $n = 10$ (сумма всех имеющихся валиков):

$$P(A) = \frac{3}{10}.$$

Вероятность того, что «Второй валик окажется эллиптическим» (событие B), вычислена по формуле (1.1) из предположения, что первый конусный валик уже взят (условная вероятность). Число благоприятных исходов равно общему числу эллиптических валиков $m = 7$, а общее число исходов $n = 9$ (сумма всех оставшихся валиков, так как один валик уже вытащен):

$$P_B(A) = \frac{7}{9}.$$

По теореме умножения (формула (1.13)), искомая вероятность

$$P(A \cdot B) = P(A) \cdot P_A(B) = \frac{3}{10} \cdot \frac{7}{9} = \frac{7}{30}.$$

Для применения формулы (1.13) нет разницы, какое событие считать первым. Далее приводится доказательство этого утверждения.

Пусть событие А – «Первым вытаскивается эллиптический валик», событие B – «Вторым вытаскивается конусный валик». Используя размышления, приведенные выше, можно легко найти:

$$P(A \cdot B) = P(B) \cdot P_B(A) = \frac{7}{10} \cdot \frac{3}{9} = \frac{7}{30},$$

что наглядно иллюстрирует справедливость равенства (1.13).

Ответ: вероятность вытащить конусный валик, а затем эллиптический (и наоборот) равна 7/30.

Для *независимых* событий теорема умножения (формула (1.13)) примет вид

$$P(A \cdot B) = P(\mathrm{A}) \cdot P(B), \tag{1.14}$$

то есть вероятность совместного появления двух независимых событий равна произведению вероятностей этих событий.

Пример 10. Найти вероятность поразить цель двумя орудиями одновременно, если вероятность поражения цели первым орудием (событие A) равна 0,8, а вторым (событие B) – 0,7.

Решение

Так как вероятность поражения цели одним оружием не оказывает влияния на вероятность поражения цели другим оружием, то события A и B *независимые*. Для решения используется теорема умножения для независимых событий (формула (1.14)):

$$P(A \cdot B) = P(A) \cdot P(B) = 0{,}8 \cdot 0{,}7 = 0{,}56.$$

Ответ: вероятность совместного поражения цели двумя орудиями равна 0,56 (56 %).

Обратите внимание, что вероятность наступления хотя бы одного из двух событий всегда больше вероятности наступления каждого из этих событий по отдельности, а вероятность их совместного наступления меньше вероятности наступления каждого из этих событий по отдельности:

$$P(A + B) > P(A) \text{ или } P(B) > P(A \cdot B). \tag{1.15}$$

Пример 11. В цехе работают семь мужчин и три женщины. По табельным номерам наугад выбраны три человека. Найти вероятность того, что все выбранные лица окажутся мужчинами.

Решение

Способ 1. Применение теоремы умножения вероятностей.

Событие A – «Первый выбранный человек – мужчина»;

Событие B – «Второй выбранный человек – мужчина»;

Событие C – «Третий выбранный человек – мужчина».

Все события зависимые, так как каждый выбранный человек меняет количество оставшихся. Условные вероятности событий рассчитаны по формуле (1.1):

$$P(A) = \frac{7}{10};\ P_A(B) = \frac{6}{9};\ P_{AB}(C) = \frac{5}{8}.$$

Искомая вероятность рассчитывается по теореме умножения для независимых событий (формула (1.14)):

$$P(A \cdot B \cdot C) = P(A) \cdot P_A(B) \cdot P_{AB}(C) = \frac{7 \cdot 6 \cdot 5}{10 \cdot 9 \cdot 8} = \frac{7}{24} = 0,292.$$

Ответ: вероятность, что все отобранные люди – мужчины, 7/24 (29 %).

Способ 2. Применение формул комбинаторики.

Общее количество исходов, заключенных в выборе трех человек из 10 рассчитывается по формуле (1.6) (комбинация «С*очетание*», так как порядок выбора не важен):

$$n = C_{10}^3 = \frac{10!}{7! \cdot 3!} = 120.$$

Аналогично количество способов отбора трех мужчин из семи:

$$m = C_7^3 = \frac{7!}{4! \cdot 3!} = 35.$$

Вероятность того, что отобранные люди будут мужчинами равна (формула (1.1)):

$$P(A) = \frac{35}{120} = \frac{7}{24} = 0,292.$$

Ответ: вероятность, что все отобранные люди – мужчины, 7/24 (29 %).

Оба способа дают одинаковый результат.

Пример 12. Система S состоит из двух независимых подсистем – S_a и S_b (рис. 1.15). Неисправность хотя бы одной подсистемы ведет к неисправности всей системы (подсистемы соединены последовательно). Каждая подсистема состоит из двух независимых дублирующих блоков – a_k и b_k (блоки в подсистеме соединены параллельно).

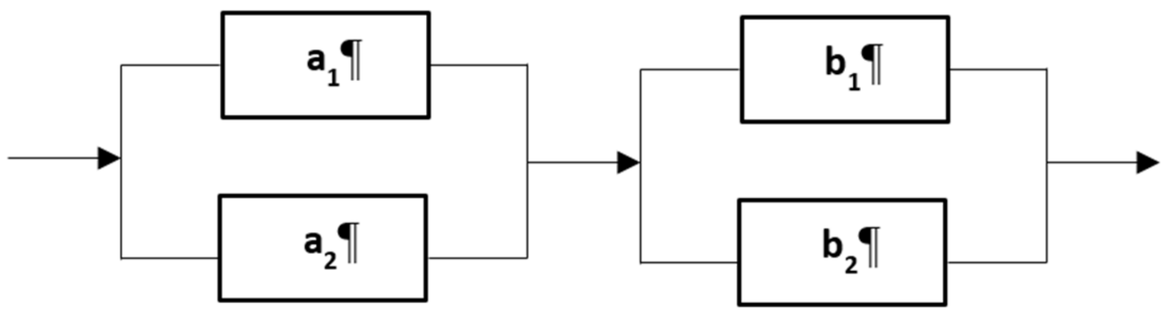


Рис. 1.15. Схема параллельного подсоединения блоков подсистем

«Найти надежность системы» – значит вычислить вероятность того, что система будет исправна в течение некоторого времени, если известна надежность каждого блока (вероятность безотказной работы): $P(a_k)=0{,}8$, $P(b_k)=0{,}9$.

Решение

Событие S – «Система S исправна».

Событие S_a – «Подсистема a исправна».

Событие S_b – «Подсистема b исправна».

Способ 1

Подсистема a исправна в том случае, когда исправен хотя бы один из блоков a_k (k = 1, 2), следовательно, вероятность безотказной работы подсистемы a рассчитывается по теореме сложения вероятностей (формула (1.8)):

$$P(S_a)=P(a_1)+P(a_2)-P(a_1)\cdot P(a_2)=0{,}8+0{,}8-0{,}8^2=0{,}96.$$

Аналогично для подсистемы b:

$$P(S_b)=P(b_1)+P(b_2)-P(b_1)\cdot P(b_2)=0{,}9+0{,}9-0{,}9^2=0{,}99.$$

Для исправности всей системы S необходима исправность обеих подсистем – S_a и S_b, т. е. необходимо определить вероятность наступления пары событий $S_a\cdot S_b$. Для этого используется теорема умножения вероятностей независимых событий (выход из строя одной подсистемы не влияет на работоспособность второй подсистемы) (формула (1.14)):

$$P(S)=P(S_a)\cdot P(S_b)=0{,}96\cdot 0{,}99=0{,}9504.$$

Ответ: вероятность того, что система S будет исправна, 0,9504 (95 %).

Способ 2

Надежность системы можно проконтролировать, используя формулу (1.11) (сумма вероятностей событий, образующих *полную группу событий*). Вероятность работоспособности всей системы противоположна вероятности неработоспособности всей системы:

$$P(S)=1-P(\overline{S}),$$

где событие $\overline{S}$ – «Система S неисправна» (событие, противоположное событию S).

Для неисправности системы S необходима неисправность хотя бы одной из подсистем (a или b), то есть по теореме сложения вероятностей (формула (1.8))

$$P(\overline{S})=P(\overline{S}_a)+P(\overline{S}_b)-P(\overline{S}_a)\cdot P(\overline{S}_b).$$

В свою очередь для неисправности подсистемы S_a необходима неисправность блоков a_1 и a_2, то есть по теореме умножения вероятностей (формула (1.14))

$$P(\overline{S}_a) = P(\overline{a}_1) \cdot P(\overline{a}_2) = (1-0{,}8) \cdot (1-0{,}8) = 0{,}04.$$

Аналогично для подсистемы S_b:

$$P(\overline{S}_b) = P(\overline{b}_1) \cdot P(\overline{b}_2) = (1-0{,}9) \cdot (1-0{,}9) = 0{,}01.$$

Следовательно, вероятность неисправности системы равна:

$$P(\overline{S}) = 0{,}04 + 0{,}01 - 0{,}04 \cdot 0{,}01 = 0{,}0496.$$

Таким образом, вероятность работоспособности системы равна:

$$P(S) = 1 - P(\overline{S}) = 1 - 0{,}0496 = 0{,}9504.$$

Ответ: вероятность того, что система S будет исправна, 0,9504 (95 %) (для двух способов решения ответ одинаковый).

Задачи для самостоятельного решения

Задача 1. Найти вероятность того, что наудачу взятое двухзначное число окажется кратным либо 2, либо 5, либо тому и другому одновременно.

Задача 2. В ящике лежат шары: 4 белых, 10 красных, 8 зеленых, 9 черных. Из ящика вынимают один шар.

1. Найти вероятность, что шар окажется цветным (не белым).
2. Найти вероятность вынуть:
   а) белый;
   б) красный;
   г) зеленый;
   д) черный.
3. Найти вероятность вынуть:
   а) белый или зеленый;
   б) красный или черный;
   в) белый, зеленый или красный.

Задача 3. Зашедший в магазин мужчина что-нибудь покупает с вероятностью 0,1, а зашедшая женщина – с вероятностью 0,6. У прилавка один мужчина и две женщины. Найти вероятность того, что покупка будет совершена.

Задача 4. Для сигнализации об аварии установлены два независимо работающих сигнализатора. Вероятность того, что при аварии сигнализатор сработает, равна 0,95 для первого сигнализатора и 0,90 для второго. Найти вероятность того, что при аварии сработает хотя бы один сигнализатор.

Задача 5. Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8. Найти вероятность того, что при одном выстреле в мишень попадает хотя бы один из стрелков.

Задача 6. Отдел технического контроля проверяет изделия на стандартность. Вероятность того, что изделие стандартное, равна 0,9. Найти вероятность того, что из двух проверенных изделий хотя бы одно стандартное.

Задача 7. Вероятность того, что при одном измерении некоторой величины будет допущена ошибка, превышающая заданную точность, равна 0,4. Произведены три независимых измерения. Найти вероятность того, что будет допущена ошибка, которая превысит заданную точность.

Задача 8. Студент знает ответы на 20 вопросов программы из 25. Найти вероятность того, что студент знает все вопросы, входящие в билет (билет состоит из трех вопросов).

Задача 9. Система S состоит из трех блоков – a_1, a_2, b, где a_1 и a_2 – дублирующие блоки (рис. 1.16). Неисправность хотя бы одной подсистемы ведет к неисправности всей системы (подсистемы соединены последовательно). Подсистема S_a состоит из двух независимых дублирующих блоков – a_k и b_k (схема параллельного подсоединения блоков подсистем).

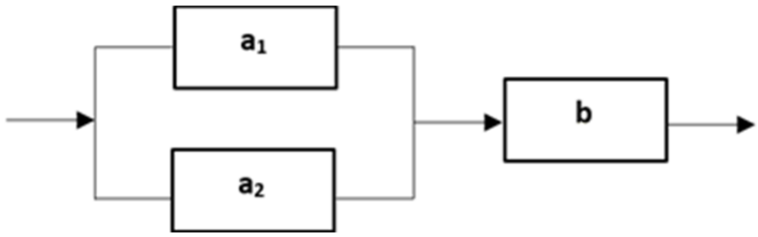


Рис. 1.16. Схема соединения блоков системы

Найти надежность системы (вероятность того, что система будет исправна в течение некоторого времени), если известна надежность каждого блока: $P(a_k) = 0{,}8$, $P(b_k) = 0{,}9$.

Задача 10. Дана система из трех подсистем – a, b, c (рис. 1.17), где b_1 и b_2 – дублирующие блоки.

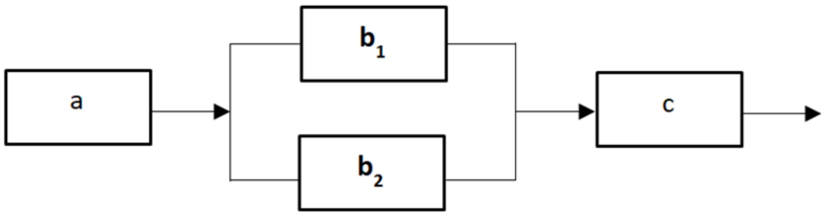


Рис. 1.17. Схема соединения блоков системы

Найти надежность системы (вероятность того, что система будет исправна в течение некоторого времени), если известна надежность каждого блока: $P(a_k)=0{,}85$, $P(b_k)=0{,}75$, $P(c_k)=0{,}9$.

1.5. Повторение испытаний

Если производится некоторое количество испытаний, в результате которых может наступить событие A, и вероятность появления этого события в каждом из испытаний не зависит от результатов остальных испытаний, то такие испытания называются *независимыми относительно события A*.

1.5.1. Формула Бернулли

Ставится задача вычислить вероятность того, что при n испытаниях событие A осуществится ровно k раз и, следовательно, не осуществится $n-k$ раз. Искомую вероятность обозначим $P_n(k)$.

Пример 1. Символ $P_5(3)$ означает вероятность того, что в пяти испытаниях событие A наступит три раза и, следовательно, не наступит два раза.

Важно подчеркнуть, что не требуется, чтобы событие A повторилось k раз в определенной последовательности.

Поставленную задачу можно решить с помощью *формулы Бернулли*:

$$P_n(k)=C_n^k\cdot p^k\cdot q^{n-k} \tag{1.16}$$

или согласно формуле (1.6):

$$P_n(k)=\frac{n!}{k!\cdot(n-k)!}\cdot p^k\cdot q^{n-k}.$$

Пример 2. Вероятность того, что расход электроэнергии в продолжение одних суток не превысит установленной нормы, составляет 75 % $\left(p=0{,}75\right)$. Найти вероятность того, что в ближайшие 6 суток расход электроэнергии в течение 4 суток не превысит нормы.

Решение

Вероятность нормального расхода электроэнергии в продолжение каждых из 6 суток постоянна $\left(p=0{,}75\right)$. Следовательно, вероятность перерасхода электроэнергии в каждые сутки также постоянна и равна

$$q=1-p=1-0{,}75=0{,}25.$$

Искомая вероятность по формуле Бернулли (1.16) равна:

$$P_6(4) = C_6^4 \cdot p^4 \cdot q^2 = \frac{6!}{4! \cdot (6-4)!} \cdot 0,75^4 \cdot 0,25^2 = 0,2966.$$

Ответ: вероятность того, что расход электроэнергии в течение четырех суток из шести не превысит нормы, равна 0,2966 (≈30 %).

Пример 3. Устройство, состоящее из пяти независимо работающих элементов, включено в течение времени *T*. Вероятность отказа каждого из элементов за это время равна 0,2. Найти вероятность того, что откажут:

1) три элемента;
2) не менее четырех элементов;
3) хотя бы один элемент.

Решение

По условию задачи имеется схема Бернулли с параметрами p = 0,2 (вероятность того, что элемент откажет), n = 5 (число испытаний, то есть число элементов), k (число «успехов» – число отказавших элементов). Для решения используется формула Бернулли (формула (1.16)).

1. Вероятность того, что откажут ровно три элемента из пяти, рассчитывается по формуле:

$$P_5(3) = \mathrm{C}_5^3 \cdot 0{,}2^3 \cdot 0{,}8^2 = 0{,}0512.$$

Ответ: вероятность того, что откажут три элемента, равна 0,0512 (≈ 5 %).

2. Вероятность того, что откажут не менее четырех элементов из пяти (т. е. рассматриваются случаи, когда откажут четыре или пять элементов):

$$P_5(\geq 4) = P_5(4) + P_5(5) = \mathrm{C}_5^4 \cdot 0{,}2^4 \cdot 0{,}8^1 + \mathrm{C}_5^5 \cdot 0{,}2^5 \cdot 0{,}8^0 = 0{,}0067.$$

Ответ: вероятность отказа четырех или пяти элементов равна 0,0067 (≈ 0,7 %).

3. Вероятность того, что откажет хотя бы один элемент. В этом случае можно рассчитать вероятность противоположного события – «ни один элемент не откажет» $\left(P_5(0)\right)$:

$$P_5(0) = \mathrm{C}_5^0 \cdot p^0 \cdot q^5 = \mathrm{C}_5^0 \cdot 0{,}2^0 \cdot 0{,}8^5 = 0{,}673.$$

Тогда вероятность того, что откажет хотя бы один элемент, будет равна:

$$P_5(\geq 0) = 1 - P_5(0) = 1 - 0{,}327 = 0{,}673.$$

Ответ: вероятность события отказа хотя бы одного элемента равна 0,673 (≈ 67 %).

1.5.2. Локальная и интегральная теоремы Лапласа

Формула Бернулли позволяет вычислить вероятность того, что событие появится в n испытаниях ровно k раз, но пользоваться ей при больших значениях n достаточно трудно, так как формула требует выполнения действий над очень большими числами.

Пример 3. Если $n = 50$, $k = 30$, $p = 0{,}1$, то для определения вероятности $P_{50}(30)$ надо вычислить значение выражения:

$$P_{50}(30) = \frac{50!}{30!\cdot 20!} \cdot 0{,}1^{30} \cdot 0{,}9^{20},$$

где $50! = 30414093 \cdot 10^{57}$; $20! = 24329020 \cdot 10^{11}$; $30! = 26525286 \cdot 10^{25}$.

Понятно, что операции с такими большими числами весьма затруднительны.

Естественно, возникает вопрос: можно ли вычислить эту вероятность, не прибегая к формуле Бернулли? Для этого используется *локальная теорема Лапласа*, которая позволяет приближенно найти вероятность появления события ровно k раз в n испытаниях, если число испытаний достаточно велико.

Для частного случая ($p = 0{,}5$) приближенная формула была найдена в 1730 г. Абрахамом де Муавром (рис. 1.18).

Рис. 1.18. Абрахам де Муавр (26 мая 1667 г. – 27 ноября 1754 г.) – английский математик французского происхождения

В 1783 г. Лаплас обобщил формулу Муавра для произвольного значения вероятности p. Поэтому используемая теорема иногда называется *теоремой Муавра–Лапласа*.

Локальная теорема Муавра–Лапласа. Если вероятность p появления события A в каждом испытании постоянна, то вероятность $P_n(k)$ того, что событие A появится в n испытаниях ровно k раз, приближенно равна (точность увеличивается при $n \to \infty$) значению функции:

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{npq}} \cdot \varphi(x), \qquad (1.17)$$

при

$$x = \frac{(k - np)}{\sqrt{npq}}. \qquad (1.18)$$

Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, соответствующие положительным значениям аргумента x, приведены в таблице значений локальной функции Лапласа (прил. 1, табл. П1.1). Так как функция $\varphi(x)$ четна, то ее значения для отрицательных значений аргумента можно найти в той же таблице.

Пример 4. Найти вероятность того, что событие A наступит ровно 80 раз в 400 испытаниях, если вероятность появления этого события в каждом испытании равна 0,2.

Решение

По условию задачи $n = 400$; $k = 80$; $p = 0{,}2$; $q = 0{,}8$. Так как числовые значения достаточно велики, то для решения используется асимптотическая формула Лапласа (1.17):

$$P_{400}(80) \approx \frac{1}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} \cdot \varphi(x) = \frac{1}{8} \cdot \varphi(x).$$

Значение x определяется по формуле 1.18:

$$x = \frac{(k - np)}{\sqrt{npq}} = \frac{80 - 400 \cdot 0{,}2}{8} = 0.$$

По таблице значений локальной функции Лапласа (прил. 1, табл. П1.1) находится значение функции $\varphi(0) = 0{,}3989$.

Искомая вероятность равна:

$$P_{400}(80) = \frac{1}{8} \cdot 0{,}3989 = 0{,}04986.$$

Формула Бернулли приводит примерно к такому же результату (выкладки ввиду их громоздкости опущены):

$$P_{400}(80) = 0{,}0498.$$

Ответ: вероятность того, что в 400 испытаниях событие A наступит ровно 80 раз, составляет 0,0498 (≈ 5 %).

Интегральная теорема Лапласа позволяет вычислить вероятность $P_n(k_1, k_2)$ того, что событие A появится в n испытаниях не менее k_1 и не более k_2 раз (для краткости будем говорить «от k_1 до k_2 раз»).

$$P_n(k_1, k_2) \cong \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-z^2/2}\, \mathrm{d}z, \qquad (1.19)$$

где

$$x' = \frac{(k_1 - np)}{\sqrt{npq}} \text{ и } x'' = \frac{(k_2 - np)}{\sqrt{npq}}. \qquad (1.20)$$

В результате преобразования соотношения (1.19) вероятность того, что в n независимых испытаниях событие A появится m раз в интервале от k_1 до k_2 раз, рассчитывается по формуле

$$P_n(k_1, k_2) \approx \Phi(x'') - \Phi(x'), \qquad (1.21)$$

где $\Phi(x')$ и $\Phi(x'')$ – значения интегральной функции Лапласа (прил. 1, табл. П1.2).

Пример 5. Вероятность того, что деталь не прошла проверку ОТК, $p = 0{,}2$. Найти вероятность того, что среди 400 случайно отобранных деталей окажется от 70 до 100 забракованных деталей.

Решение

По условию задачи $p = 0{,}2$; $q = 0{,}8$; $n = 400$; $k_1 = 70$; $k_2 = 100$. Искомая вероятность рассчитывается по интегральной теореме Лапласа (формула (1.21)):

$$P_{400}(70{,}100) \approx \Phi(x'') - \Phi(x').$$

Нижний и верхний пределы интегрирования вычисляются по формуле (1.20):

$$x' = \frac{(k_1 - np)}{\sqrt{npq}} = \frac{(70 - 400 \cdot 0{,}2)}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = -1{,}25;$$

$$x'' = \frac{(k_2 - np)}{\sqrt{npq}} = \frac{(100 - 400 \cdot 0{,}2)}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = 2{,}5.$$

Таким образом, при условии симметричности функции Лапласа

$$P_{400}(70{,}100) \approx \Phi(x'') - \Phi(x') = \Phi(2{,}5) - \Phi(-1{,}25) = \Phi(2{,}5) + \Phi(1{,}25).$$

Значения интегральной функции Лапласа (прил. 1, табл. П1.2) равны:

$$\Phi(2{,}5) = 0{,}4938,\ \Phi(1{,}25) = 0{,}3944.$$

Искомая вероятность

$$P_{400}(70{,}100) = 0{,}4938 + 0{,}3944 = 0{,}8882.$$

Ответ: вероятность того, что среди 400 случайно отобранных деталей окажется забракованных от 70 до 100 деталей, равна 0,89 (89 %).

1.5.3. Вероятность отклонения относительной частоты от постоянной вероятности в независимых испытаниях

При рассмотрении схемы Бернулли иногда необходимо найти вероятность того, что отклонение относительной частоты w от постоянной вероятности p по абсолютной величине не превышает заданного числа ε. Другими словами, необходимо найти вероятность существования неравенства

$$P\left(\left|w-p\right|\leq \varepsilon\right). \qquad (1.22)$$

Следует напомнить, что относительная частота $w=\dfrac{m}{n}$ (формула (1.13)).

В результате математических операций [8] с использованием *интегральной теоремы Лапласа* (формулы ((1.19)–(1.22)) вероятность осуществления неравенства $\left|w-p\right|\leq\varepsilon$ приближенно равна значению удвоенной функции Лапласа $2\Phi\left(x\right)$ при $x=\varepsilon\sqrt{\dfrac{n}{pq}}$:

$$P\left(\left|\frac{m}{n}-p\right|\leq\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \qquad (1.23)$$

Пример 6. Вероятность того, что деталь нестандартна, p = 0,1. Найти вероятность того, что среди случайно отобранных 400 деталей относительная частота появления нестандартных деталей отклонится от вероятности p = 0,1 по абсолютной величине не более чем на 0,03.

Решение

По условию n = 400; p = 0,1; q = 0,9; ε = 0,03. Требуется найти вероятность $P\left(\left|\dfrac{m}{400}-0{,}1\right|\leq 0{,}03\right)$, используя формулу (1,24):

$$P\left(\left|\frac{m}{400}-0{,}1\right|\leq 0{,}03\right)\approx 2\Phi\left(0{,}03\sqrt{\frac{400}{0{,}1\cdot 0{,}9}}\right)=2\Phi\left(2\right).$$

Интегральная функция Лапласа (прил. 1, табл. П1.2) $\Phi(2)=0{,}4772$, следовательно,

$$P\left(\left|\frac{m}{400}-0{,}1\right|\leq 0{,}03\right)\approx 2\Phi(2)=2\cdot 0{,}4772=0{,}9544.$$

Ответ: искомая вероятность приближенно равна 0,9544 (95,44 %).

Примечание. Смысл полученного результата таков: если взять достаточно большое число проб по 400 деталей в каждой, то примерно в 95,44 % этих проб отклонение относительной частоты от постоянной вероятности p = 0,1 по абсолютной величине не превысит 0,03 [5].

Пример 7. (Задача, обратная предыдущей.) Вероятность того, что деталь нестандартна, $p = 0{,}1$. Найти, сколько деталей надо отобрать, чтобы с вероятностью $P = 0{,}9544$ можно было утверждать, что относительная частота w появления нестандартных деталей (среди отобранных) отклонится от постоянной вероятности p по абсолютной величине не более чем на 0,03.

Решение

По условию задачи $p = 0{,}1$; $q = 0{,}9$; $\varepsilon = 0{,}03$; $P\left(\left|w - 0{,}1\right| \le 0{,}03\right) = 0{,}9544$.

Требуется найти n.

Для данной задачи формула (1.23) примет вид

$$2\Phi\left(0{,}03\sqrt{\frac{n}{0{,}1\cdot 0{,}9}}\right) = 2\Phi(0{,}1\sqrt{n}) = 0{,}9544,$$

следовательно, $\Phi(0{,}1\sqrt{n}) = 0{,}4772$.

Исходя из того, что интегральная функция Лапласа (прил. 1, табл. П1.2) $\Phi(x) = 0{,}4772$, значение $x = 2$.

Значение n определяется путем решения уравнения:

$$0{,}1\sqrt{n} = 2.$$

Отсюда искомое число деталей $n = 400$.

Ответ: для выполнения условия задачи необходимо отобрать 400 деталей.

Примечание. Смысл полученного результата таков: если взять достаточно большое число проб по 400 деталей, то в 95,44 % этих проб относительная частота появления нестандартных деталей будет отличаться от постоянной вероятности по абсолютной величине не более чем на 0,03. Это означает, что относительная частота заключена в границах от (0,1 – 0,03 = 0,07) до (0,1 + 0,03 = 0,13). Другими словами, число нестандартных деталей в 95,44 % проб будет заключено между 28 (7 % от 400) и 52 (13 % от 400).

Если взять лишь одну пробу из 400 деталей, то с большой уверенностью можно ожидать, что в этой пробе будет нестандартных деталей не менее 28 и не более 52. Возможно, хотя и маловероятно (< 3 %), что нестандартных деталей окажется меньше 28 либо больше 52 [5].

1.5.4. Наивероятнейшее число появлений события в независимых испытаниях

Число k_m (число наступления события в независимых испытаниях, в каждом из которых вероятность появления события равна p) называют наивероятнейшим, если вероятность того, что событие наступит в этих испытаниях k_m раз, превышает (или, по крайней мере, не меньше) вероятности остальных возможных исходов испытаний.

Число испытаний k_m может быть только целым числом. Наивероятнейшее число k_m определяется на основе формулы (1.20) двойным неравенством

$$np - q \le k_m \le np + q, \tag{1.24}$$

решением которого могут быть несколько вариантов:

а) если число «*np – q – дробное*», то наивероятнейшее число k_m – это ближайшее большее целое, удовлетворяющее неравенству

$$k_m > np - q; \tag{1.25}$$

б) если число «*np – q – целое*», то существует два наивероятнейших числа, а именно:

$$k_{m1} = np - q \text{ и } k_{m2} = np - q + 1; \tag{1.26}$$

в) если число «*np – целое*», то наивероятнейшее число k_m равно:

$$k_m = np. \tag{1.27}$$

Пример 8. Испытывается каждый из 15 элементов некоторого устройства. Вероятность того, что элемент выдержит испытания, равно $p = 0{,}9$. Найти наивероятнейшее число элементов, которые выдержат испытания.

Решение

По условию задачи n = 15; p = 0,9; q = 0,1. Наивероятнейшее число элементов k_m определяется неравенством (1.24). После подстановки данных задачи неравенство примет вид

$$15 \cdot 0{,}9 - 0{,}1 \le k \le 15 \cdot 0{,}9 + 0{,}1;$$

$$13{,}5 \le k \le 14{,}4.$$

Так как $np - q = 13{,}5$ – *дробное* число, то наивероятнейшее число k определяется неравенством (1.25): k_m = 14.

Ответ: наиболее вероятно, что испытания выдержат 14 элементов.

Пример 9. Товаровед осматривает 24 образца товаров. Вероятность того, что каждый из образцов будет признан годным к продаже, равна 0,6. Найти наивероятнейшее число образцов, которые товаровед признает годными к продаже.

Решение

По условию задачи n = 24; p = 0,6; q = 0,4. Наивероятнейшее число образцов товаров k_m определяется неравенством (1.24). Поставив данные задачи, неравенство примет вид

$$24 \cdot 0{,}6 - 0{,}4 \le k_m \le 24 \cdot 0{,}6 + 0{,}4;$$

$$14 \le k_m \le 15.$$

Так как $np - q = 14$ – *целое*, то существует два наивероятнейших числа (формула 1.26):

$$k_{m1} = 14 \text{ и } k_{m2} = 15.$$

Ответ: наиболее вероятно, что товаровед признает годными к продаже 14 или 15 образцов товаров.

Пример 10. Два стрелка стреляют по мишени. Вероятность промаха при одном выстреле для первого стрелка равна 0,2, а для второго – 0,4. Найти наивероятнейшее число залпов, при которых не будет ни одного попадания в мишень, если стрелки произведут 25 залпов.

Решение

Промахи стрелков – независимые события, поэтому для определения вероятности промахов двух стрелков используется теорема умножения вероятностей независимых событий (формула (1.14)). Вероятность того, что оба стрелка при одном залпе промахнутся,

$$p = 0,2 \cdot 0,4 = 0,08.$$

Наивероятнейшее число выстрелов k_m определяется неравенством (1.26). После подстановки данных задачи неравенство примет вид

$$25 \cdot 0,08 - 0,92 \le k_m \le 25 \cdot 0,08 + 0,92.$$

Так как $np = 2$ – *целое*, то наивероятнейшее число определяется формулой (1.27):

$$k_m = np = 2.$$

Ответ: 2 – наивероятнейшее число залпов, при которых не будет ни одного попадания в мишень.

Пример 11. Батарея произвела шесть выстрелов по объекту. Вероятность попадания в объект при одном выстреле равна 0,3.

Найти:

1) наивероятнейшее число попаданий;
2) вероятность наивероятнейшего числа попаданий;
3) вероятность того, что объект будет разрушен, если для этого достаточно хотя бы двух попаданий.

Решение

По условию задачи: $n = 6$; $p = 0,3$; $q = 0,7$.

1. Наивероятнейшее число попаданий рассчитывается по формуле (1.25). После подстановки данных задачи неравенство примет вид

$$6 \cdot 0,3 - 0,7 \le k_m \le 6 \cdot 0,3 + 0,7 \text{ или } 1,1 \le k_m \le 2,1.$$

Отсюда $k_m = 2$.

Ответ: 2 – наивероятнейшее число попаданий.

2. Вероятность наивероятнейшего числа попаданий (два попадания из шести возможных) можно рассчитать по формуле Бернулли (формулы (1.16), (1.17)):

$$P_6(2) = C_6^2 \cdot p^2 \cdot q^4 = \frac{6 \cdot 5}{1 \cdot 2} 0,3^2 \cdot 0,7^4 = 0,324.$$

Ответ: вероятность наивероятнейшего числа попаданий (два попадания из шести возможных) составляет 0,324 (34,2 %).

3. По условию задачи для того, чтобы объект был разрушен, достаточно, чтобы было совершено два, три, четыре, пять или шесть попаданий. Эти события не совместны, поэтому вероятность разрушения объекта равна сумме вероятностей этих событий (формула (1.10)):

$$P = P_6(2) + P_6(3) + P_6(4) + P_6(5) + P_6(6),$$

каждое слагаемое рассчитывается по формуле Бернулли (1.16). Однако проще сначала вычислить вероятность Q противоположного события (ни одного попадания или одно попадание):

$$Q = P_6(0) + P_6(1) = q^6 + C_6^1 \cdot p \cdot q^5 = 0,7^6 + 6 \cdot 0,3^2 \cdot 0,7^4 = 0,42.$$

Искомая вероятность того, что объект будет разрушен,

$$P = 1 - Q = 1 - 0,42 = 0,58.$$

Ответ: вероятность того, что объект будет разрушен, равна 0,58 (58 %).

Примечание. Доказательство теорем и более подробные математические выкладки преобразования формул, приведенных в этой главе, можно посмотреть в литературе по теории вероятностей и математической статистике [8].

Задачи для самостоятельного решения

Задача 1. Монету подбрасывают восемь раз. Какова вероятность того, что шесть раз выпадет «герб»?

Задача 2. В цехе шесть моторов. Для каждого мотора вероятность того, что он в данный момент включен, равна 0,8. Найти вероятность того, что в данный момент:

а) включено четыре мотора;
б) включены все моторы;
в) выключены все моторы.

Задача 3. Два равносильных шахматиста играют в шахматы. Что вероятнее:

а) выиграть одну партию из двух или две партии из четырех?

б) выиграть не менее двух партий из четырех или не менее трех партий из пяти?

Примечание. При решении задачи результат игры «ничья» во внимание не принимается, каждая игра завершается победой одного из игроков.

Задача 4. В семье пять детей. Найти вероятность того, что среди этих детей:

а) два мальчика;
б) не более двух мальчиков;
в) более двух мальчиков;
г) не менее двух и не более четырех мальчиков.

Вероятность рождения мальчика принять равной 0,51.

Задача 5. Вычислительное устройство состоит из 1000 элементов, работающих независимо друг от друга. Вероятность отказа каждого элемента за смену равна $p = 0{,}024$. Найти вероятность, что за смену откажут шесть элементов.

Задача 6. Производство электронно-лучевых трубок для телевизоров дает в среднем 12 % брака. Найти вероятность наличия 215 годных трубок в партии из 250 штук.

Задача 7. При установившемся технологическом процессе на ткацкой фабрике происходит 10 обрывов нити на 100 веретен в час.

Найти:

а) вероятность того, что в течение часа на 80 веретенах произойдет 7 обрывов нити;

б) наивероятнейшее число обрывов нити на 80 веретенах в течение часа.

Задача 8. Вероятность поражения мишени при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена 75 раз.

Задача 9. Вероятность рождения мальчика равна 0,51. Найти вероятность того, что среди 100 новорожденных окажется 50 мальчиков.

Задача 10. Вероятность появления события в каждом из 100 независимых испытаний постоянна ($p = 0{,}8$). Найти вероятность того, что событие появится:

а) не менее 75 раз и не более 90 раз;
б) не менее 75 раз;
в) не более 74 раз.

Задача 11. В жилом доме имеется 6400 ламп. Вероятность включения каждой из них в вечернее время равна 0,5. Найти вероятность того, что число одновременно включенных ламп будет между $k_1 = 3120$ и $k_2 = 3200$.

Задача 12. Найти вероятность того, что если бросить монету 200 раз, то «решка» выпадет от 90 до 110 раз.

Задача 13. Вероятность появления события в каждом из 10 000 независимых испытаний $p = 0{,}75$. Найти вероятность того, что относительная частота появления события отклонится от вероятности $p = 0{,}75$ по абсолютной величине не более чем на 0,01.

Задача 14. Французский ученый Бюффон (XVIII в.) бросил монету 4040 раз, причем «герб» появился 2048 раз. Найти вероятность того, что при повторении опыта Бюффона относительная частота появления «герба» отклонится от вероятности повеления «герба» по абсолютной величине не более чем в опыте Бюффона.

Задача 15. Сколько раз надо бросить монету, чтобы с вероятностью 0,6 можно было ожидать, что отклонение относительной частоты появлений «герба» от вероятности $p = 0{,}5$ не превысит 0,01?

Задача 16. Сколько надо произвести независимых испытаний с вероятностью появления события в каждом испытании равной 0,4, чтобы наивероятнейшее число появления события в этих испытаниях было равно 25?

Задача 17. Отдел технического контроля проверяет 475 изделий на брак. Вероятность того, что изделие бракованное, равна 0,05. Найти с вероятностью 0,95 границы, в которых будет заключено число m бракованных изделий среди проверенных.

Задача 18. Испытывается каждый из 15 элементов некоторого устройства. Вероятность того, что элемент выдержит испытание, равна 0,9. Найти наивероятнейшее число элементов, которые выдержат испытания.

Задача 19. Отдел технического контроля проверяет партию из 10 деталей. Вероятность того, что деталь стандартна, равна 0,75. Найти наивероятнейшее число деталей, которые будут признаны стандартными.

Контрольные вопросы и задания

1. Дайте определение термину «теория вероятностей».
2. Дайте определение основным понятиям теории вероятностей: «исход опыта», «группа исхода», «событие».
3. Достоверное, невозможное, случайное событие: дайте определения и приведите примеры.
4. Что такое «вероятность события»?
5. Приведите классическую формулу для вычисления вероятности.

6. Сравните понятия «вероятность события» и «относительная частота», укажите сходства и различия.

7. Приведите основные формулы комбинаторики: перестановка, размещение, сочетания. Приведите примеры их использования.

8. Сформулируйте «правило суммы», «правило произведения», используемые в комбинаторике.

9. Сформулируйте теорему сложения вероятностей для совместных событий; для несовместных событий. Приведите формулы и примеры их использования.

10. Сформулируйте теорему умножения вероятностей для зависимых событий; для независимых событий. Приведите формулы и примеры их использования.

11. Запишите формулу Бернулли. Приведите примеры использования данной формулы.

12. Сформулируйте локальную и интегральную теоремы Лапласа. Приведите примеры использования данных теорем.

13. Покажите, как рассчитать наивероятнейшее число появлений события в независимых испытаниях.

Глава 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Виды случайных величин

Одним из важнейших в теории вероятностей является понятие случайной величины. *Случайной величиной* (СВ) называется такая величина, которая в результате опыта может принимать то или иное значение, причём до завершения опыта неизвестно какое.

Случайные величины, как правило, обозначают заглавными латинскими буквами X, Y, Z, …, а их значения – соответствующими маленькими буквами с подстрочными индексами, например $x_1, x_2, \ldots, x_n$.

Пример 1. Число родившихся мальчиков среди 100 новорожденных – это случайная величина, которая имеет следующие возможные значения: $X = \{x_0 = 0, x_1 = 1, x_2 = 2, ..., x_{100} = 100\}$.

Пример 2. Расстояние, которое пролетит снаряд при выстреле из орудия, – это случайная величина. Действительно, расстояние зависит не только от установки прицела, но и от многих других причин (силы и направления ветра, температуры и т. д.), которые не могут быть полностью учтены. Возможные значения этой величины принадлежат некоторому промежутку $Y \in (a, b)$.

Случайную величину называют *дискретной* или *непрерывной* в зависимости от пространства элементарных событий, в котором она определена.

Дискретная случайная величина (*ДСВ*) может принимать конечное число определенных числовых значений.

Пример 3. Число попаданий при 10 выстрелах в цель является дискретной случайной величиной с возможными значениями: $X = \{x_0 = 0, x_1 = 1, x_2 = 2, ..., x_{10} = 10\}$.

Непрерывная случайная величина (*НСВ*) принимает непрерывный ряд значений и число таких значений бесконечно. Границы непрерывной случайной величины могут быть точно определены или быть неопределёнными, расплывчатыми.

Пример 4. Координаты пасущейся на огороженном участке коровы – непрерывная случайная величина с точно определенными границами (ограждение участка).

Пример 5. Погрешность измерения – непрерывная случайная величина с размытыми границами. Никогда нельзя точно определить возможные границы погрешности.

Основная задача теории вероятностей состоит в определении закона распределения СВ. *Закон распределения случайной величины* – любое правило, устанавливающее соответствие между значениями случайной величины и вероятностями ее наступления.

2.1.1. Дискретная случайная величина

Дискретная случайная величина может принимать отдельные, изолированные значения с определенными вероятностями.

Дискретная случайная величина X может быть задана рядом распределения или функцией распределения (интегральным законом распределения).

Рядом распределения называется совокупность всех возможных значений x_i и соответствующих им вероятностей $p_i = P(X = x_i)$. Ряд распределения может быть представлен в виде табл. 2.1:

Таблица 2.1

Ряд распределения дискретной случайно величины

Значение СВ	x_1	x_2	...	x_n
Вероятность	p_1	p_2	...	p_n

При этом вероятности p_i аналогично формуле (1.11) удовлетворяют условию

$$\sum_{i=1}^{n} p_n = p_1 + p_2 + \ldots + p_n = 1, \qquad (2.1)$$

так как события $X = x_1$, ..., $X = x_n$ несовместны и образуют *полную группу*.

Графическое изображение ряда распределения называется *многоугольником распределения.* Для его построения возможные значения случайной величины x_i откладываются по оси абсцисс, а вероятности p_i – по оси ординат. Точки A_i с координатами (x_i, p_i) соединяются ломаными линиями и образуют многоугольник распределения (рис. 2.1).

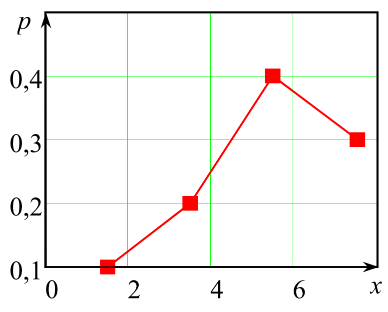


Рис. 2.1. Многоугольник распределения

Функцией распределения случайной величины X называется функция $F(x)$, значение которой в точке x равно вероятности того, что случайная величина X будет меньше этого значения x, т. е.

$$F(x) = P(X < x). \qquad (2.2)$$

Геометрически равенство (2.2) можно истолковать так: функция распределения $F(x)$ равна вероятности того, что случайная величина X примет значение, которое лежит на числовой оси левее точки x.

Функция распределения $F(x)$ для дискретной случайной величины вычисляется по формуле

$$F(x) = \sum_{x_i < x} p_i \,, \tag{2.3}$$

где суммирование ведется по всем значениям i, для которых $x_i < x$.

Для дискретной случайной величины *функция распределения* $F(x)$ графически представляет собой ступенчатую неубывающую функцию (рис. 2.2).

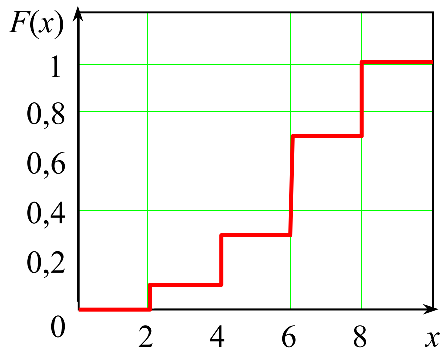


Рис. 2.2. Функция распределения

Пример 6. Партия состоит из 100 изделий, среди которых имеются 10 дефектных. Для проверки качества отобраны случайным образом пять изделий. Построить ряд распределений случайного числа X дефектных изделий, содержащихся в выборке.

Комментарий. Построить ряд распределений – значит рассчитать значения вероятностей для каждого значения случайной величины и привести их в виде таблицы.

Решение

Так как в выборке число дефектных изделий может быть любым целым числом в пределах от 0 до 5, то возможные значения случайной величины X равны:

$$x_1 = 0,\; x_2 = 1,\; x_3 = 2,\; x_4 = 3,\; x_5 = 4,\; x_6 = 5.$$

Вероятность того, что в выборке окажется x_i дефектных изделий, равна:

$$P(X = x_i) = \frac{C_{10}^{x_i} \cdot C_{90}^{5-x_i}}{C_{100}^{5}}.$$

С точностью до 0,001 значения вероятностей равны:

$$p_1 = P(X = 0) = 0{,}583,\ \ p_2 = P(X = 1) = 0{,}340,\ \ p_3 = P(X = 2) = 0{,}070,$$

$$p_4 = P(X = 3) = 0{,}007,\ \ p_5 = P(X = 4) = 0,\ \ p_6 = P(X = 5) = 0.$$

Используя для проверки равенство (2.1), можно убедиться, что расчеты и округление произведены правильно (табл. 2.2). В противном случае следует произвести повторное округление рассчитанных результатов с большей или меньшей точностью.

Таблица 2.2

Ряд распределения дефектных изделий в выборке

x_i	0	1	2	3	4	5
p_i	0,583	0,3401	0,070	0,007	0	0

Для большей наглядности ряд распределения можно изобразить на графике (т. е. *построить многоугольник распределения*). Для этого необходимо нанести на график точки с координатами (x_i, p_i) (рис. 2.3).

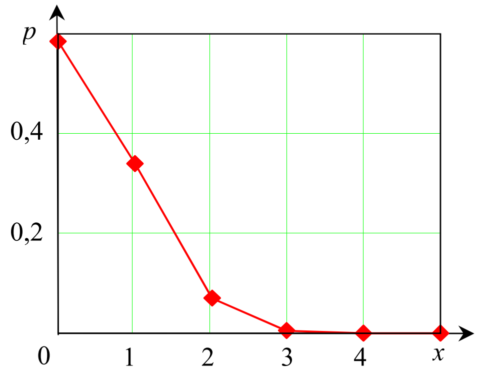


Рис. 2.3. Многоугольник распределения из примера 6

Ответ: ряд распределений представлен в табл. 2 и в виде многоугольника распределения на рис. 2.3.

2.1.2. Непрерывная случайная величина

Возможные значения непрерывной случайной величины непрерывно заполняют некоторый конечный или бесконечный промежуток.

Случайная величина называется *непрерывной*, если существует неотрицательная функция *f*(*x*), удовлетворяющая при любых *x* равенству

$$F(x) = \int_{-\infty}^{X} f(x)dx, \tag{2.4}$$

то есть является непрерывной кусочно-дифференцируемой функцией с непрерывной производной.

Каждому промежутку (a; b) из области значений непрерывной случайной величины (НСВ) соответствует определенная вероятность $P(a \le x < b)$ того, что значение этой случайной величины попало в указанный интервал.

На промежутке (a; b) невозможно перечислить все значения случайной величины, так как множество его точек несчетно, но можно утверждать, что вероятность каждого отдельно взятого значения НСВ есть бесконечно малая величина (из геометрического определения вероятности) [8]. Поэтому для НСВ имеет смысл говорить не о вероятности конкретного значения, а о вероятности попадания в некий интервал значений:

$$P(a \le x < b) = \int_a^b f(x)dx.$$

С увеличением x функция распределения $F(x)$ «накапливает» (суммирует) вероятности, а значит, является неубывающей и изменяется в пределах от 0 до 1. По этой причине её иногда называют *интегральной* функцией распределения.

В геометрической интерпретации функция распределения $F(x)$ равна площади, ограниченной сверху кривой плотности распределения $f(x)$ и лежащей левее точки x.

Плотностью распределения (или *плотностью вероятностей*) непрерывной случайной величины X в точке x называется производная ее функции распределения $F(x)$ в этой точке, обозначаемая $f(x)$. Функция $f(x)$ определяется соотношением

$$f(x) = \frac{dF(x)}{dx}. \tag{2.5}$$

Непрерывная случайная величина задается либо *функцией распределения* $F(x)$ (интегральным законом распределения), либо *плотностью вероятности* $f(x)$ (дифференциальным законом распределения).

2.1.3. Свойства функции распределения и плотности распределения случайной величины

Основные свойства функции распределения $F(x)$:

Свойство 1. Значения функции распределения принадлежат отрезку [0, 1]:

$$F(x) = 0 \text{ при } x \le x_{\min}; \tag{2.6}$$

$$F(x) = 1 \text{ при } x \le x_{\max}. \tag{2.7}$$

Свойство 2. Функция распределения – неубывающая функция:

$$F(x_1) \le F(x_2) \text{ при } x_1 < x_2. \tag{2.8}$$

Свойство 3. Вероятность попадания случайной величины в заданный интервал равна приращению функции на этом интервале:

$$P(a < X < b) = F(b) - F(a). \qquad (2.9)$$

Указанные свойства наглядно представлены на графике функции распределения НСВ (рис. 2.4).

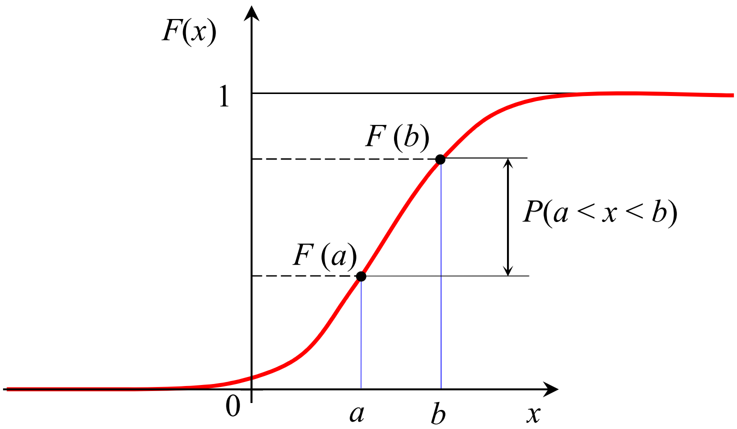


Рис. 2.4. Функция распределения НСВ

Основные свойства плотности распределения:

1. Плотность распределения – неотрицательная функция:

$$f(x) \geq 0. \qquad (2.10)$$

2. Плотность распределения подчиняется условию нормировки:

$$\int_{-\infty}^{\infty} f(x)dx = 1, \qquad (2.11)$$

другими словами, площадь фигуры, заключенной под *кривой распределения* (график плотности распределения), равна 1.

Указанные свойства плотности распределения наглядно представлены на рис. 2.5.

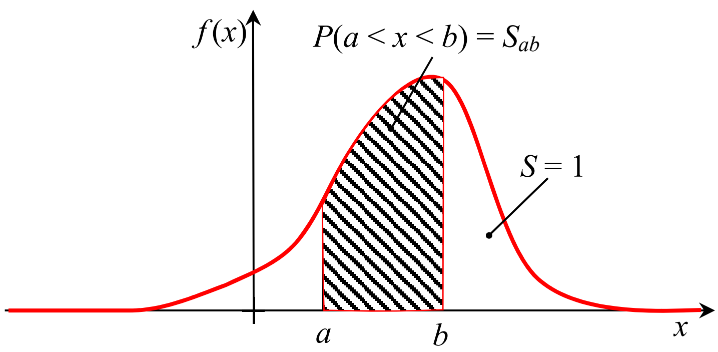


Рис. 2.5. Плотность распределения НСВ

Геометрический смысл равенства (2.9) состоит в том, что вероятность попадания НСВ X в интервал $[a; b]$ равна площади заштрихованной области S_{ab} (рис. 2.5).

Вероятность $P = P(a < x < b)$ называется *доверительной вероятностью*, а интервал $[a; b]$ – *доверительным интервалом*.

Пример 7. Изобразить графически плотность вероятности НСВ непрерывной случайной величины, заданной в следующем виде:

$$f(x) = \begin{cases} 0,5\cos(x) & \text{при} \quad -\pi/2 < x < \pi/2; \\ 0 & \text{при} \quad |x| > \pi/2. \end{cases}$$

Найти функцию распределения и построить ее.

Решение

Так как плотность распределения $f(x)$ задана в «кусочном» виде, то и функцию распределения $F(x)$ следует находить для каждого интервала в отдельности, используя формулу (2.4):

- для интервала $x < -\pi/2$:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{-\pi/2} 0dx = 0;$$

- для интервала $-\pi/2 \le x \le \pi/2$:

$$F(x) = \int_{-\infty}^{-\pi/2} 0dx + \int_{-\pi/2}^{x} \cos(x)dx = \frac{1+\sin(x)}{2};$$

- для интервала $x > \pi/2$:

$$F(x) = \int_{-\infty}^{-\pi/2} 0dx + \int_{-\pi/2}^{\pi/2} \cos x dx + \int_{\pi/2}^{+\infty} 0dx = \left.\frac{1+\sin(x)}{2}\right|_{-\pi/2}^{\pi/2} = \frac{1+1}{2} - \frac{1-1}{2} = 1.$$

Окончательно функция распределения примет вид

$$F(x) = \begin{cases} 0 & \text{при} \quad -\infty < x \le \dfrac{-\pi}{2}; \\ \dfrac{(\sin(x)+1)}{2} & \text{при} \quad \dfrac{-\pi}{2} < x \le \dfrac{\pi}{2}; \\ 1 & \text{при} \quad x > \dfrac{\pi}{2}. \end{cases}$$

На рис. 2.6 изображены графики исходной плотности распределения $f(x)$ (красная линия) и найденной функции распределения $F(x)$ (синяя линия).

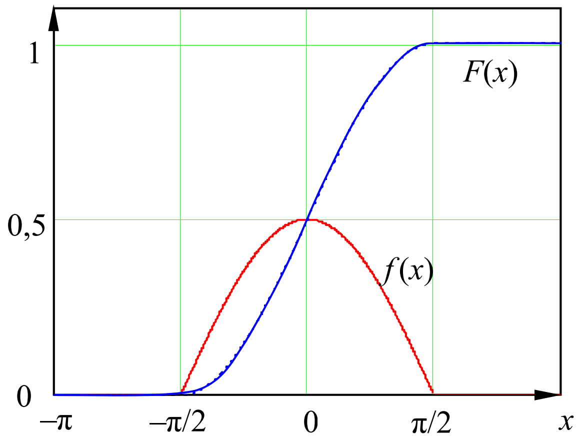


Рис. 2.6. Графики плотности распределения и функции распределения

Ответ: рассчитанная функция распределения показана на рис. 2.6 синей линией.

Пример 8. Случайная величина X задана функцией распределения

$$F(x)=\begin{cases} 0 & \text{при} \quad x\le 0; \\ x^2 & \text{при} \quad 0<x\le 1; \\ 1 & \text{при} \quad x>1. \end{cases}$$

Найти плотность распределения и вероятность того, что в результате четырех независимых испытаний величина X ровно три раза примет значение в интервале [0,25; 0,75].

Решение

Плотность распределения $f(x)$ определяется формулой (2.5), поэтому необходимо найти производные для каждого интервала x:

$$f(x)=\begin{cases} \dfrac{d}{dx}0 & \text{при} \quad x\le 0 \\ \dfrac{d}{dx}x^2 & \text{при} \quad 0<x\le 1 \\ \dfrac{d}{dx}1 & \text{при} \quad x>1 \end{cases} = \begin{cases} 0 & \text{при} \quad x\le 0; \\ 2x & \text{при} \quad 0<x\le 1; \\ 1 & \text{при} \quad x>1. \end{cases}$$

На рис. 2.7 показаны в виде графиков заданная функция распределения $F(x)$ (синяя линия) и плотность распределения $f(x)$ (красная линия).

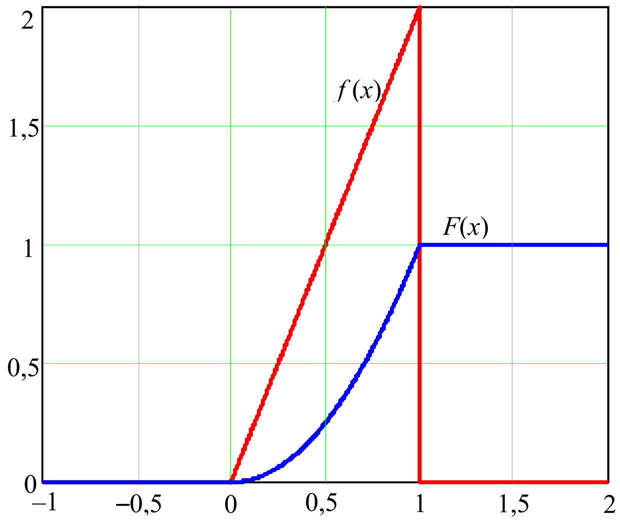


Рис. 2.7. Графики плотности распределения и функции распределения

Вероятность того, что в результате четырех независимых испытаний величина X ровно три раза примет значение в заданном интервале, определяется по формуле Бернулли (1.16).

По условию задачи n = 4; k = 3; p – вероятность попадания значения случайной величины в интервал [0,25; 0,75], которая рассчитывается по формуле (2.9). В указанном диапазоне функция распределения задана одной функцией: $F(x) = x^2$.

Вероятность попадания значения случайной величины в интервал [0,25; 0,75] с учетом условия задачи равна:

$$p = P(0,25 < X < 0,75) = F(0,75) - F(0,25) = 0,75^2 - 0,25^2 = 0,5.$$

Тогда по формуле Бернулли (формула (1.16)) вероятность наступления события A равна:

$$P_4(3) = C_4^3 \cdot p^3 \cdot q^{4-1} = \frac{4!}{3! \cdot 1!} \cdot 0,5^3 \cdot 0,5 = 0,25.$$

Ответ: вероятность того, что в результате четырех независимых испытаний величина X ровно три раза примет значение в интервале [0,25; 0,75], равна 0,25 (25 %).

Пример 9. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le -1; \\ \dfrac{x}{4} + \dfrac{1}{4} & \text{при} \quad -1 < x \le 3; \\ 1 & \text{при} \quad x > 3. \end{cases}$$

Найти вероятность того, что в результате испытания случайная величина X примет значение, принадлежащее интервалу [0; 2].

Решение

Вероятность того, что случайная величина примет значения, заключенные в интервале [0; 2], равна приращению функции распределения на этом интервале (свойство 3 функции распределения (формула (2.9)):

$$P(0 < X < 2) = F(2) - F(0).$$

Интервал [0; 2] входит в интервал [–1; 3], в котором по условию задачи *функция распределения* равна $F(x) = \frac{x}{4} + \frac{1}{4}$. После подстановки числовых значений формула (2.9) примет вид

$$F(2) - F(0) = \left(\frac{2}{4} + \frac{1}{4}\right) - \left(\frac{0}{4} + \frac{1}{4}\right) = \frac{1}{2}.$$

Ответ: вероятность попадания случайной величины X в интервал [0; 2] равна 1/2 (50 %).

Задачи для самостоятельного решения

Задача 1. В партии из шести деталей имеется четыре стандартных. Наудачу отобраны три детали.

Составить закон распределения числа стандартных деталей среди отобранных X.

Построить многоугольник распределения.

Задача 2. Дан ряд распределения случайной величины X (табл. 2.3).

Таблица 2.3

Ряд распределения случайной величины X

x_i	10	20	30	40	50
p_i	0,2	0,3	0,35	0,1	0,05

Найти функцию распределения вероятности $F(x)$ этой случайной величины и построить ее.

Задача 3. Дан ряд распределения случайной величины X (табл. 2.4).

Таблица 2.4

Ряд распределения случайной величины X

x_i	1	2	3	4
p_i	0,45	0,1	0,15	p_4

Найти значение вероятности появления значения случайной величины $x_4 = 4$.

Найти функцию распределения вероятности $F(x)$ этой случайной величины и построить ее.

Задача 4. Дан ряд распределения случайной величины X (табл. 2.5).

Таблица 2.5

Ряд распределения случайной величины X

x_i	10	15	20
p_i	0,1	0,7	0,2

Построить многоугольник распределения.

Найти функцию распределения вероятности $F(x)$ этой случайной величины и построить ее.

Задача 5. Непрерывная случайная величина X задана плотностью распределения) $f(x) = \frac{3}{2}\sin(3x)$ в интервале $[0; \pi/3]$, вне этого интервала $f(x) = 0$. Найти вероятность того, что X примет значение, принадлежащее интервалу $[\pi/6; \pi/4]$.

2.2. Числовые характеристики случайной величины

Функция распределения $F(x)$ и *плотность вероятности* $f(x)$ наиболее полно и наглядно описывают любую случайную величину. Но при решении многих практических задач не всегда нужно характеризовать СВ полностью, т. е. определять закон распределения, *функцию распределения* $F(x)$ или *плотность распределения* $f(x)$. Иногда достаточно указать отдельные числовые параметры, которые в сжатой форме выражают наиболее существенные особенности случайной величины. Такие параметры называются *числовыми характеристиками*.

К числовым характеристикам СВ относятся характеристики положения, характеристики рассеивания, характеристики симметрии и островершинности.

2.2.1. Характеристики положения случайной величины

К характеристикам положения относятся: математическое ожидание, мода, медиана. Они характеризуют положение случайной величины на числовой оси, т. е. указывают некоторые ориентировочные значения, около которых группируются все возможные значения случайной величины.

Одной из основных числовых характеристик СВ является *математическое ожидание (мат. ожидание)*. Мат. ожидание характеризует среднее значение СВ, представляет собой центр тяжести распределения СВ.

Обозначение: m, $M\left[X\right]$, где X – это обозначение случайной величины, для которой необходимо определить математическое ожидание.

Математическое ожидание вычисляют по формулам:

$$m = M\left[X\right] = \begin{cases} \sum\limits_i x_i \cdot p_i & \text{для ДСВ,} \\ \int\limits_{-\infty}^{\infty} x \cdot f\left(x\right)dx & \text{для НСВ.} \end{cases} \qquad (2.12)$$

Модой случайной величины называется ее наиболее вероятное значение, т. е. то значение, для которого вероятность p_i (для дискретной СВ) или плотность распределения $f(x)$ (для непрерывных СВ) достигают максимума.

Обозначение: Mo.

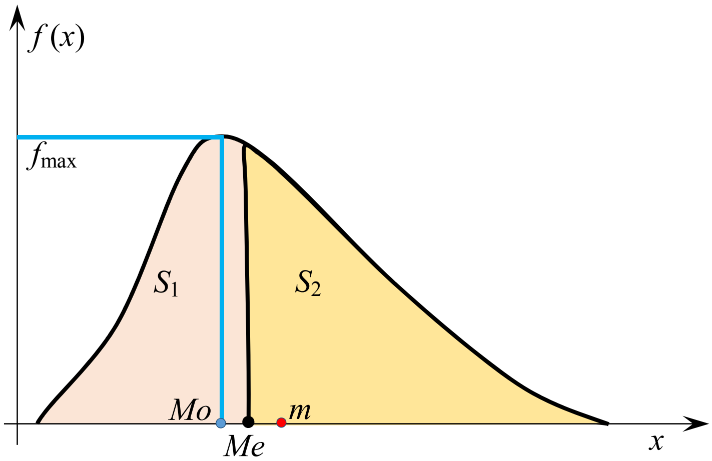


Рис. 2.8. Мода, медиана и математическое ожидание

Медиана СВ – это абсцисса точки, в которой ограниченная кривой плотности распределения $f(x)$ площадь S делится пополам $(S_1 = S_2 = S / 2)$, т. е. выполняется условие

$$P(X < Me) = P(X > Me) = \frac{1}{2}. \qquad (2.13)$$

Это означает, что значение случайной величины X с одинаковой вероятностью может оказаться меньше Me или больше Me.

Обозначение: Me.

Медиана, как правило, существует только для непрерывных случайных величин и указывает «центр» распределения случайной величины [3].

На рис. 2.8 обозначены математическое ожидание *m*, мода *Mo* и медиана *Me* для некоторой случайной величины *X*.

Для симметричного распределения все три характеристики (мода, медиана и математическое ожидание) совпадают.

2.2.2. Характеристики рассеивания случайной величины

Не менее важными являются числовые характеристики, описывающие рассеивание случайной величины. Такими характеристиками являются *дисперсия* и *среднее квадратическое отклонение* (СКО). На практике часто встречаются случайные величины с равными мат. ожиданиями, но различными распределениями. Дисперсия и СКО показывают, насколько тесно сгруппированы возможные значения СВ около ее мат. ожидания.

Дисперсия случайной величины – математическое ожидание квадрата отклонения СВ от ее мат. ожидания:

$$D\left[X\right]=M\left[(X-m)^2\right]. \tag{2.14}$$

Для вычисления дисперсии применяются следующие формулы:

$$D\left[x\right]=\begin{cases}\sum_i\left(x_i-m_x\right)^2\cdot P(x_i) & \text{для} \quad \text{ДСВ},\\ \int_{-\infty}^{\infty}\left(x-m\right)^2\cdot f\left(x\right)dx & \text{для} \quad \text{НСВ}.\end{cases} \tag{2.15}$$

Дисперсия является удобной характеристикой рассеивания значений СВ около ее мат. ожидания, но она лишена наглядности, так как имеет размерность квадрата СВ. Для удобства следует использовать величину, размерность которой совпадает с размерностью случайной величины. Такой величиной является *среднее квадратическое отклонение* случайной величины, которое представляет собой квадратный корень из ее дисперсии:

$$\sigma=\sigma\left[X\right]=\sqrt{D\left[X\right]}. \tag{2.16}$$

СКО характеризует ширину диапазона значений СВ.

2.2.3. Характеристики симметрии и островершинности случайной величины

На практике часто необходимо характеризовать геометрическую форму распределения. Для этого используется характеристики: *симметрия* и *островершинность*, которые можно представить с помощью моментов распределения.

В теории вероятностей различают моменты двух видов: начальные и центральные.

Начальный момент s-го порядка случайной величины $X(\alpha_s)$ – это мат. ожидание s-й степени этой случайной величины.

$$\alpha_s = M\left[X^s\right]. \qquad (2.17)$$

Начальный момент 1-го порядка α_1 – это *мат. ожидание* СВ:

$$\alpha_1 = M\left[X\right].$$

Центрированной случайной величиной $\overset{0}{X}$ называется отклонение СВ от ее математического ожидания, т. е. СВ, математическое ожидание которой равно нулю:

$$\overset{0}{X} = X - m, \quad M\left[\overset{0}{X}\right] = M\left[X - m\right] = 0. \qquad (2.18)$$

Моменты центрированной случайной величины – *центральные моменты*.

Центральный момент s-го порядка случайной величиной X (μ_s) – это мат. ожидание s-й степени центрированной случайной величины.

$$\mu_s = M\left[\overset{0}{X}{}^s\right] = M\left[\left(X - m\right)^s\right]. \qquad (2.19)$$

Для любой СВ центральный момент первого порядка равен 0.

Центральный момент 2-го порядка μ_2 – это *дисперсия* случайной величины (см. формулу (2.14)):

$$\mu_2 = M\left[\overset{0}{X}{}^2\right] = M\left[\left(X - m\right)^2\right].$$

Вычислить дисперсию можно и через второй начальный момент:

$$D\left[X\right] = M\left[X^2\right] - m^2. \qquad (2.20)$$

В этом случаи формулы для расчета дисперсии примут вид:

$$D[x] = \begin{cases} \sum_i x_i^2 \cdot P(x_i) - (m)^2 & \text{для} \quad \text{ДСВ}, \\ \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - (m)^2 & \text{для} \quad \text{НСВ}. \end{cases} \qquad (2.21)$$

Центральный момент 3-го порядка μ_3 служит характеристикой асимметрии *(«скошенности»*) распределения.

Коэффициентом асимметрии (или *«скошенности»*) распределения – мера сдвига распределения влево или вправо и рассчитывается по формуле

$$A = \frac{\mu_3}{\sigma^3}. \qquad (2.22)$$

Наличие асимметрии легко установить визуально по графику плотности распределения.

Если $A < 0$, то распределение имеет левостороннюю *«скошенность»*, график плотности распределения имеет вид, показанный на рис. 2.9, *а* (левая ветвь относительно максимальной ординаты вытянута больше, чем правая). В этом случае $m < Me < Mo$.

Если $A > 0$, то распределение имеет правостороннюю *«скошенность»*, график плотности распределения имеет вид, показанный на рис. 2.9, *в* (правая ветвь относительно максимальной ординаты вытянута больше, чем левая). В этом случае $m > Me > Mo$.

При симметричной СВ центральный момент $\mu_3 = 0$, а следовательно, и *коэффициент «скошенности»* $A = 0$ (рис. 2.9, *б*). В этом случае $m = Me = Mo$.

Центральный момент 4-го порядка μ_4 служит характеристикой *«островершинности»* или *«плосковершинности»* распределения. Это свойство распределения описывается с помощью эксцесса.

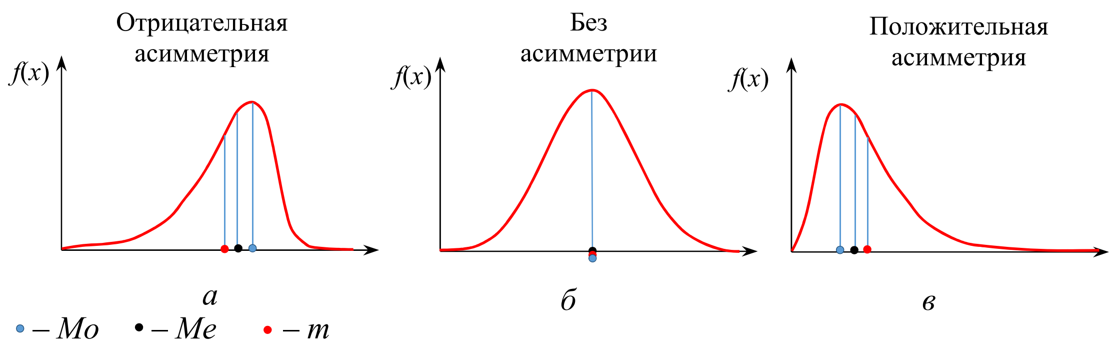


Рис. 2.9. Коэффициент асимметрии

Коэффициент эксцесса – это мера остроты пика распределения случайной величины и рассчитывается по формуле

$$E = \frac{\mu_4}{\sigma^4} - 3. \qquad (2.23)$$

Кривая нормального распределения (подробнее см. раздел 2.3.2) принята за эталон, остальные с ней сравниваются, поэтому эксцесс нормального распределения принят за $E = 0$ (именно для этого отнимается число 3 в формуле (2.23)), для более островершинных законов распределения $E > 0$, а для плосковершинных – $E < 0$ (см. рис. 2.10).

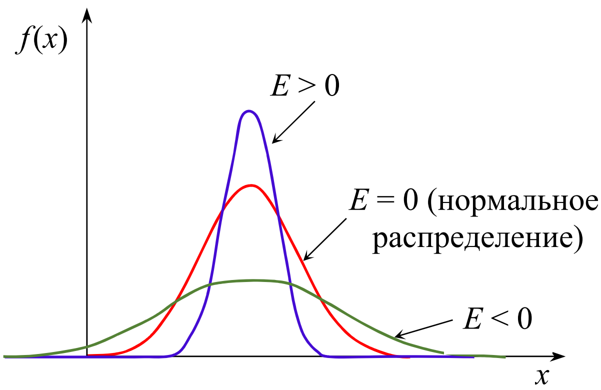


Рис. 2.10. Коэффициент эксцесса

Пример 1. Из партии численностью 25 изделий, среди которых имеется шесть нестандартных, случайным образом выбраны три изделия. Найти мат. ожидание и СКО нестандартных изделий, содержащихся в выборке.

Решение

По условию задачи СВ X принимает следующие значения: $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$. Вероятность того, что в этой выборке окажется i ($i = 0, 1, 2, 3$) нестандартных изделий, вычисляется по формуле

$$p_i = P(x_i) = \frac{C_6^i \cdot C_{19}^{3-i}}{C_{25}^i},$$

отсюда

$$p_1 = 0{,}41; \; p_2 = 0{,}43; \; p_3 = 0{,}11; \; p_4 = 0{,}05.$$

Ряд распределения для данной СВ представлен в табл. 2.6.

Таблица 2.6

Ряд распределения случайной величины X

x_i	0	1	2	3
p_i	0,41	0,43	0,15	0,05

Мат. ожидание вычисляется по формуле (2.12) для ДСВ:

$$m = \sum_i x_i \cdot p_i = 0 \cdot 0{,}41 + 1 \cdot 0{,}43 + 2 \cdot 0{,}11 + 3 \cdot 0{,}05 = 0{,}8.$$

Обратите внимание, что математическое ожидание всегда

$$x_{\min} < m < x_{\max}.$$

В данном случае $0 < m = 0{,}8 < 3$.

Для расчета СКО необходимо знать дисперсию, которая вычисляется по формуле (2.20). Для начала необходимо рассчитать начальный момент 2-го порядка (формула (2.17)):

$$M[X^2] = \sum_i x_i^2 \cdot p_i = 0^2 \cdot 0{,}41 + 1^2 \cdot 0{,}43 + 2^2 \cdot 0{,}11 + 3^2 \cdot 0{,}05 = 1{,}32.$$

Затем рассчитывается дисперсия:

$$D[X] = M[X^2] - (m)^2 = 1{,}32 - (0{,}8)^2 = 0{,}68.$$

СКО рассчитывается по формуле (2.16):

$$\sigma[X] = \sqrt{D[X]} = 0{,}82.$$

Ответ: для заданной выборки мат. ожидание равно 0,8, среднее квадратическое отклонение – 0,82.

Пример 2. Дан перечень возможных значений ДСВ *X*: $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, а также известны мат. ожидание этой величины $M[X] = 0{,}1$ и квадрат мат. ожидания $M[X^2] = 0{,}9$. Найти вероятности p_1, p_2, p_3, соответствующие заданным значениям x_1, x_2, x_3.

Решение

Для решения задачи необходимо составить систему уравнений.

Уравнение 1. Для любой ДСВ сумма вероятностей всех значений СВ равна 1 (формула (2.1)):

$$p_1 + p_2 + p_3 = 1.$$

Уравнение 2. Мат. ожидание вычисляется по формуле (2.12):

$$M[X] = \sum_i x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 = -1 \cdot p_1 + 0 \cdot p_2 + 1 \cdot p_3.$$

После преобразований:

$$p_3 - p_1 = 0{,}1.$$

Уравнение 3. Квадрат мат. ожидания $M[X^2]$ – это начальный момент 2-го порядка, вычисляется по формуле (2.17):

$$M[X^2] = \sum_i x_i^2 \cdot p_i = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + x_3^2 \cdot p_3 = (-1)^2 \cdot p_1 + 0^2 \cdot p_2 + 1^2 \cdot p_3.$$

После преобразований:

$$p_1 + p_3 = 0,9.$$

Далее необходимо решить систему из трех линейных уравнений относительно неизвестных вероятностей:

$$\begin{cases} p_1 + p_2 + p_3 = 1; \\ p_3 - p_1 = 0,1; \\ p_1 + p_3 = 0,9. \end{cases}$$

В результате получается $p_1 = 0,4$, $p_2 = 0,1$, $p_3 = 0,5$.

Ответ: результат можно представить в виде ряда распределения (табл. 2.7).

Таблица 2.7

Ряд распределения случайной величины X

x_i	–1	0	1
p_i	0,4	0,1	0,5

Пример 3. Непрерывная случайная величина распределена по закону Лапласа:

$$f(x) = b \cdot e^{-|x|}.$$

Найти коэффициент b, мат. ожидание, дисперсию, СКО.

Решение

Для нахождения коэффициента b можно использовать свойство нормировки плотности распределения (формула (2.11)):

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 2b\int_{-\infty}^{\infty} e^{-x}\mathrm{d}x = 2b = 1.$$

Отсюда

$$b = 1/2.$$

Мат. ожидание для НСВ рассчитывается по формуле (2.12):

$$M[x] = \int_{-\infty}^{\infty} 0,5 \cdot xe^{-|x|}dx.$$

Так как функция $xe^{-|x|}$ – нечетная, то ее интеграл равен нулю, следовательно, $m = M[x] = 0$.

Дисперсия $D[x]$ и СКО σ рассчитываются по формулам (2.21) и (2.16):

$$D[x] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - (m)^2 = \int_{-\infty}^{\infty} x^2 \cdot 0,5 \cdot e^{-|x|}\mathrm{d}x - 0.$$

Без приведения промежуточных выкладок:

$$D[x] = 2;\ \sigma = \sqrt{D[x]} = \sqrt{2} = 1,41.$$

Ответ: коэффициент b = 1/2, мат. ожидание $m = 0$, дисперсия $D[x]$ = 2, СКО σ = 1,41.

Пример 4. Дана дискретная случайная величина X, закон распределения которой представлен в виде табл. 2.8.

Таблица 2.8

Ряд распределения случайной величины X

x_i	1	2	5	10
p_i	0,6	0,2	0,19	0,01

Найти мат. ожидание, дисперсию, асимметрию и эксцесс.

Решение

Мат. ожидание для ДСВ рассчитывается по формуле (2.12):

$$M[X] = m = \sum_i x_i \cdot p_i = 1 \cdot 0,6 + 2 \cdot 0,2 + 5 \cdot 0,19 + 10 \cdot 0,01 = 2,05.$$

Дисперсию можно рассчитать по формуле (2.15). Для ДСВ она примет вид

$$D[X] = \sum_i (x_i - m)^2 \cdot p_i = (1 - 2,05)^2 \cdot 0,6 + (2 - 2,05)^2 \cdot 0,2 +$$
$$+ (5 - 2,05)^2 \cdot 0,19 + (10 - 2,05)^2 \cdot 0,01 = 2,95.$$

Коэффициент асимметрии (или «скошенности») рассчитывается по формуле (2.22).

Сначала необходимо вычислить центральный момент 3-го порядка по формуле (2.19):

$$\mu_3 = \sum_i \left(x_i - m\right)^3 \cdot p_i = (1 - 2,05)^3 \cdot 0,6 + (2 - 2,05)^3 \cdot 0,2 +$$
$$+ (5 - 2,05)^3 \cdot 0,19 + (10 - 2,05)^3 \cdot 0,01 = 9,21.$$

Среднеквадратическое отклонение расчитывается по формуле (2.15):

$$\sigma_x = \sqrt{D[X]} = \sqrt{2,95} = 1,72.$$

После подставления рассчитанных числовых значений коэффициент асимметрии равен:

$$A = \frac{\mu_3}{\sigma^3} = \frac{9,21}{1,72^3} = 3,11.$$

Так как $A > 0$, то распределение имеет правостороннюю *«скошенность»*, что означает смещение «центра тяжести» распределения влево.

Коэффициент эксцесса (или островершинности) рассчитывается по формуле (2.23), в состав которой входит *центральный момент* 4-го порядка (формула (2.19)):

$$\mu_4 = \sum_i \left(x_i - m\right)^4 \cdot p_i = (1-2{,}05)^4 \cdot 0{,}6 + (2-2{,}05)^4 \cdot 0{,}2 +$$

$$+ (5-2{,}05)^4 \cdot 0{,}19 + (10-2{,}05)^4 \cdot 0{,}01 = 55{,}06.$$

Коэффициент эксцесса равен:

$$E = \frac{\mu_4}{\sigma^4} - 3 = \frac{55{,}06}{1{,}72^4} - 3 = 3{,}29.$$

Так как эксцесс $E > 0$, распределение имеет более острую вершину, чем у нормального закона распределения.

Ответ: мат. ожидание $m = 2{,}05$, дисперсия $D[x] = 2{,}95$, коэффициент асимметрии $A = 3{,}11$, коэффициент эксцесса $E = 3{,}29$.

Пример 5. Найти математическое ожидание и дисперсию НСВ, распределенной равномерно в интервале (a, b).

Решение

Плотность равномерного закона распределения имеет вид (подробнее см. п. 2.3.2)

$$f(x) = \frac{1}{b-a}.$$

Мат. ожидание для НСВ рассчитывается по формуле (2.12):

$$m = \int_{-\infty}^{\infty} x \cdot f(x)dx = \frac{1}{b-a}\int_a^b x \cdot dx = \frac{1}{2(b-a)} x^2 \Big|_a^b = \frac{1}{2(b-a)}\left(b^2 - a^2\right) = \frac{(b+a)}{2}.$$

Дисперсию для НСВ удобнее рассчитать по формуле (2.21):

$$D[x] = \left(\int_{-\infty}^{\infty} x^2 \cdot f(x)dx\right) - m^2 = \left(\frac{1}{b-a}\int_a^b x^2 dx\right) - \left(\frac{(b+a)}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Ответ: для равномерного закона распределения мат. ожидание $m = \frac{(b+a)}{2}$, дисперсия $D[x] = \frac{(b-a)^2}{12}$. Полученные результаты соответствуют теоретическим данным [8], приведенным в разд. 2.3.2.

Пример 6. Круглая мишень установлена так, что может вращаться вокруг оси, расположенной в центре. Мишень разбита на 8 секторов. При достаточно большой угловой скорости вращения стрелок не в состоянии различить цифры, выписанные по одной на секторах. Он вынужден стрелять наугад. При попадании в сектор 1 стрелок выигрывает 10 руб., в сектор 2 – 20 руб. и т. д., в сектор 8 – 80 руб. Стоит ли ему участвовать в такой игре, если за право стрелять один раз надо платить 50 руб.?

Решение

Поскольку мишень вращается, то способности стрелка здесь не имеют никакого значения: попадание – чистая случайность, т. е. случайная величина X принимать значения: $x_1 = 10$, $x_2 = 20$, $x_3 = 30$, $x_4 = 40$, $x_5 = 50$, $x_6 = 60$, $x_7 = 70$, $x_8 = 80$ (возможные выигрыши).

Так как все секторы одинаковые, то вероятность попадания в каждый сектор одинаковая и согласно классической формуле определения вероятностей (1.1) $p_i = \frac{1}{8}$.

Среднее значение случайной величины (мат. ожидание), согласно формуле (2.12),

$$M[X] = 10 \cdot \frac{1}{8} + 20 \cdot \frac{1}{8} + 30 \cdot \frac{1}{8} + 40 \cdot \frac{1}{8} + 50 \cdot \frac{1}{8} + 60 \cdot \frac{1}{8} + 70 \cdot \frac{1}{8} + 80 \cdot \frac{1}{8} = 45 \text{ руб.}$$

Ответ: мат. ожидание выигрыша – 45 руб., а стоимость выстрела – 50 руб. Стрелять много раз явно невыгодно. На основании подобных расчетов организуются разнообразные азартные игры, приводящие игроков к разорению.

Задачи для самостоятельного решения

Задача 1. В партии из 10 деталей имеется 7 стандартных. Наудачу отобраны три детали.

Составить закон распределения числа стандартных деталей в выборке. Построить многоугольник распределения и функцию распределения. Рассчитать числовые характеристики: математическое ожидание, дисперсия, асимметрия и эксцесс.

Задача 2. Дана дискретная случайная величина X, закон распределения которой представлен в виде табл. 2.9.

Таблица 2.9

Ряд распределения случайной величины X

x_i	4,3	5,1	10,6
p_i	0,2	0,3	0,5

Найти дисперсию и среднеквадратическое отклонение случайной величины X.

Задача 3. Случайная величина X задана плотностью распределения $f(x) = 0{,}5x$ в интервале [0; 2]; вне этого интервала $f(x) = 0$. Найти математическое ожидание величины X.

Задача 4. Случайная величина X задана плотностью распределения $f(x) = c \cdot (x^2 + 2 \cdot x)$ в интервале [0; 2]; вне этого интервала $f(x) = 0$. Найти параметр c и мат. ожидание величины X.

Задача 5. Найти математическое ожидание случайной величины X, заданной функцией распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ \dfrac{x}{4} & \text{при} \quad 0 < x \le 4; \\ 1 & \text{при} \quad x > 4. \end{cases}$$

Задача 6. Найти дисперсию случайной величины X, заданной функцией распределения

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le -2; \\ \dfrac{x}{4} + \dfrac{1}{2} & \text{при} \quad -2 < x \le 2; \\ 1 & \text{при} \quad x > 2. \end{cases}$$

Задача 7. Даны дискретные случайные величины X и Y, законы распределения которых представлены в виде табл. 2.10 и 2.11.

Таблица 2.10

Ряд распределения случайной величины X

x_i	–1	1	2	3
p_{xi}	0,48	0,01	0,09	0,42

Таблица 2.11

Ряд распределения случайной величины Y

y_i	–1	1	2	3
p_{yi}	0,19	0,51	0,25	0,05

Сравнить мат. ожидания и дисперсии случайных величин X и Y.

2.3. Некоторые частные законы распределения случайных величин

В предыдущих разделах были рассмотрены общие формы законов распределения, а теперь следует уделить внимание наиболее часто встречающимся типовым законам распределения случайных величин.

2.3.1. Законы распределения дискретных случайных величин

Биномиальное распределение (схема Бернулли) характерно для серии из n независимых испытаний, каждое из которых заканчивается либо «успехом», либо «неуспехом». В каждом испытании (опыте) вероятность успеха неизменна и равна p, а вероятность неудачи, соответственно, – $q = 1 - p$. При биномиальном распределении случайная величина X принимает значения, равные числу успехов в серии из n испытаний $(k = 0, 1, \dots, n)$, с вероятностью, рассчитанной по *формуле Бернулли* (формула (1.16)):

$$p(k) = \frac{n!}{k!(n-k)!} \cdot p^k \cdot q^{n-k}.$$

Таким образом, параметрами биноминального распределения являются: n – число испытаний в серии; $k = 0, 1, \dots, n$ – количество успехов в серии испытании; p – вероятность успеха в каждом испытании.

Закон биноминального распределения X можно представить в виде ряда распределения (табл. 2.12).

Таблица 2.12

Ряд распределения для биноминального закона

$x_i = k$	0	1	…	k	…	n
$p(k)$	q^n	$n \cdot p \cdot q^{n-1}$		$C_n^k \cdot p^k \cdot q^{n-k}$		p^n

Плотность вероятности $p(k)$ и функция распределения $F(k)$ для различных параметров биноминального распределения показана на рис. 2.11.

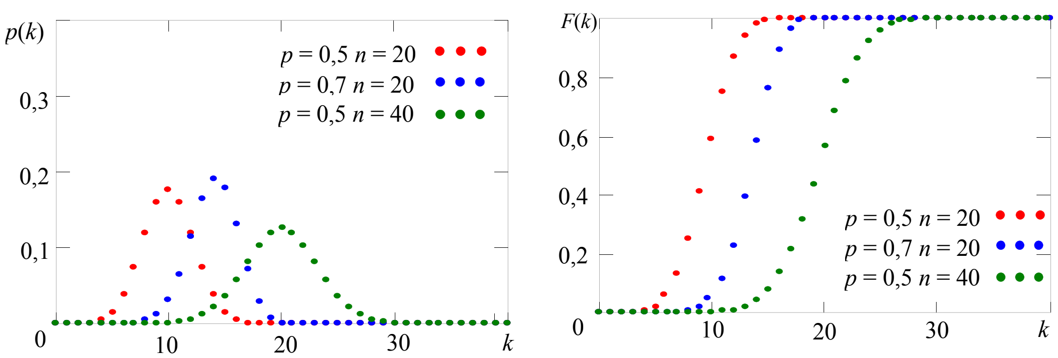


Рис. 2.11. Плотность вероятности p(k) и функция распределения F(k) для биноминального закона распределения

Основные числовые характеристики биноминального распределения с учетом его параметров примут вид: мат. ожидание – $m = np$, дисперсия – $D = npq$, асимметрия – $A = \dfrac{q-p}{\sqrt{npq}}$, эксцесс – $E = \dfrac{1-6pq}{npq}$.

Геометрическое распределение. Со схемой испытаний Бернулли можно связать еще одну случайную величину: испытание заканчивается, как только событие A наступает. Таким образом, если событие A появится в h-м испытании, то в предыдущих $h-1$ испытаниях оно не появилось. Аналогичным образом вероятность успеха в каждом испытании равна p, а вероятность неудачи – $q = 1 - p$.

Число испытаний в серии k принимает множество значений от 0 до $+\infty$, в этом случае вероятность наступления события A находится по преобразованной *формуле Бернулли*:

$$p(k) = p^k q. \tag{2.24}$$

Таким образом, параметрами геометрического распределения являются: $k = 0, 1, \ldots, +\infty$ – число неудач до первого успеха; p – вероятность успеха в каждом испытании.

Закон геометрического распределения X можно представить в виде ряда распределения (табл. 2.13).

Таблица 2.13

Ряд распределения для геометрического закона

$x_i = k$	0	1	2	…	k	…
$p(k)$	p	$q \cdot p$	$p^2 \cdot q$		$p^k \cdot q$	

Плотность вероятности $p(k)$ и функция распределения $F(k)$ для различных параметров геометрического распределения показана на рис. 2.12.

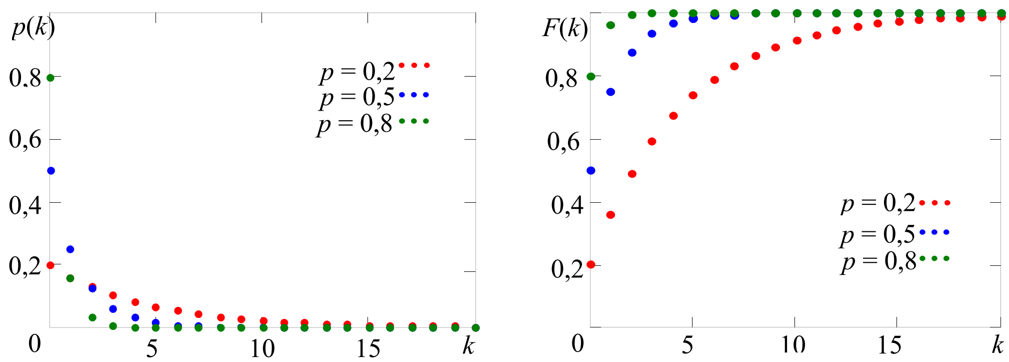


Рис. 2.12. Плотность вероятности p(k) и функция распределения F(k) для геометрического закона распределения

Основные числовые характеристики геометрического распределения с учетом его параметров примут вид: мат. ожидание – $m = \frac{1-p}{p}$, дисперсия – $D = \frac{q}{p^2}$, асимметрия – $A = \frac{2-p}{\sqrt{1-p}}$, эксцесс – $E = 6 + \frac{p^2}{1-p}$.

Пуассоновское распределение имеет место, если рассматривается задача – найти вероятность того, что при очень большом числе испытаний n, в каждом из которых вероятность события p очень мала, событие наступило ровно k раз. Считается, что произведение $n \cdot p$ сохраняет постоянное значение, а именно $\lambda = n \cdot p$. Закон распределения Пуассона является предельным случаем биноминального распределения.

Распределение Пуассона моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. Пуассоновское распределение используется при определении числа сбоев на автоматизированной линии, числа отказов сложной системы, анализе результатов маркетинговых обследований потребителей, для описания числа разладок при статически управляемом технологическом процессе, статических закономерностей несчастных случаев и т. д.

Вероятность наступления события A ровно k раз находится по закону Пуассона (закон редких событий):

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.25}$$

Таким образом, параметрами пуассоновского распределения являются: $k = 0, 1, \ldots,$ – число появлений события A; $\lambda = n \cdot p = \text{const}$, при условии $p \to 0, n \to \infty$.

Закон распределения Пуассона можно представить в виде ряда распределения (табл. 2.14).

Таблица 2.14

Ряд распределения для закона Пуассона

$x_i = k$	0	1	…	k	…
$p(k)$	$e^{-\lambda}$	$\lambda \cdot e^{-\lambda}$		$\frac{\lambda^k}{k!} e^{-\lambda}$	

Плотность вероятности $p(k)$ и функция распределения $F(k)$ для различных параметров пуассоновского распределения показаны на рис. 2.13.

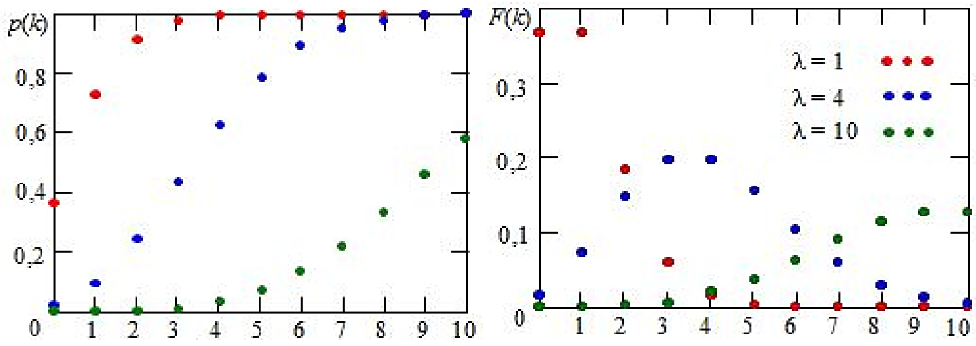


Рис. 2.13. Плотность вероятности p(k) и функция распределения F(k) для пуассоновского закона распределения

Основные числовые характеристики геометрического распределения с учетом его параметров примут вид: мат. ожидание – $m = \lambda$, дисперсия – $D = \lambda$, асимметрия – $A = \lambda^{-1/2}$, эксцесс – $E = \lambda^{-1}$.

2.3.2. Законы распределения непрерывных случайных величин

Равномерное распределение. Встречаются случайные величины, о которых заранее известно, что они могут принять какое-либо значение в строго определенных границах на отрезке [a; b], причем в этих границах все значения случайной величины имеют одинаковую вероятность (плотность вероятности).

Как показывают исследования, равномерному закону распределения подчиняются погрешности сухого трения в подвижных узлах электромеханических, измерительных устройств, погрешности округления, квантования в цифровых измерительных приборах, потенциометрических датчиков из-за дискретности намотки и др.

При снятии показаний измерительных приборов ошибка при округлении отсчёта до ближайшего целого деления шкалы является случайной величиной, которая может с постоянной плотностью вероятности принимать любые значения между двумя соседними делениями.

Так же равномерному закону распределения подчиняются погрешности округления, возникающие в результате отбрасывания у чисел одной или нескольких цифр.

Плотность распределения $p(x)$ и функция распределения $F(x)$ равномерного закона распределения имеют соответственно вид:

$$p(x) = \begin{cases} 0, & x \notin [a, b] \\ \dfrac{1}{b-a}, & x \in [a, b] \end{cases}; \quad F(x) = \begin{cases} 0, & x \le a; \\ \dfrac{x-a}{b-a}, & a < x \le b; \\ 1, & x > b. \end{cases} \qquad (2.26)$$

Параметрами равномерного распределения являются: a – начало интервала; b – конец интервала.

Плотность распределения $p(x)$ и функция распределения $F(x)$ равномерного распределения показаны на рис. 2.14.

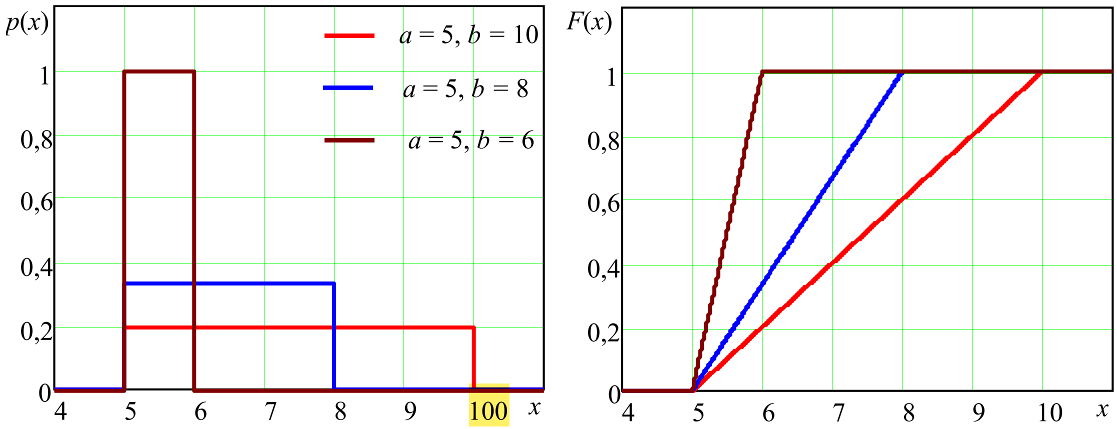


Рис. 2.14. Плотность распределения p(x) и функция распределения F(x) для равномерного закона распределения

Основные числовые характеристики равномерного распределения с учетом его параметров примут вид: мат. ожидание – $m=\frac{a+b}{2}$, дисперсия – $D=\frac{(b-a)^2}{12}$ (см. разд. 2.2.3, пример 5), асимметрия – $A=0$, эксцесс – $E=-\frac{6}{5}$.

Экспоненциальное (или показательное) распределение – это абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события. Параметр λ описывается среднее число наступлений события в единицу времени, Другими словами, $\frac{1}{\lambda}$ – средний промежуток времени между двумя последовательными событиями.

Обычно с помощью этого закона описывают продолжительность обслуживания покупателя, время жизни оборудования до отказа, промежуток времени между поломками и т. п. Данное распределение часто встречается в теории надёжности, в которой под параметром λ понимается срок службы аппаратуры или отдельных ее узлов.

Плотность распределения $p(x)$ и функция распределения $F(x)$ такой случайной величины X имеют соответственно вид:

$$p(x)=\begin{cases}0, & x<0;\\ \lambda\cdot e^{-\lambda x}, & x\geq 0.\end{cases}\quad F(x)=\begin{cases}0, & x\leq 0;\\ 1-\lambda\cdot e^{-\lambda x}, & x>0.\end{cases}\qquad (2.27)$$

Параметром экспоненциального распределения является λ – среднее число событий на единицу времени.

Плотность распределения $p(x)$ и функция распределения $F(x)$ для различных параметров экспоненциального распределения показаны на рис. 2.15.

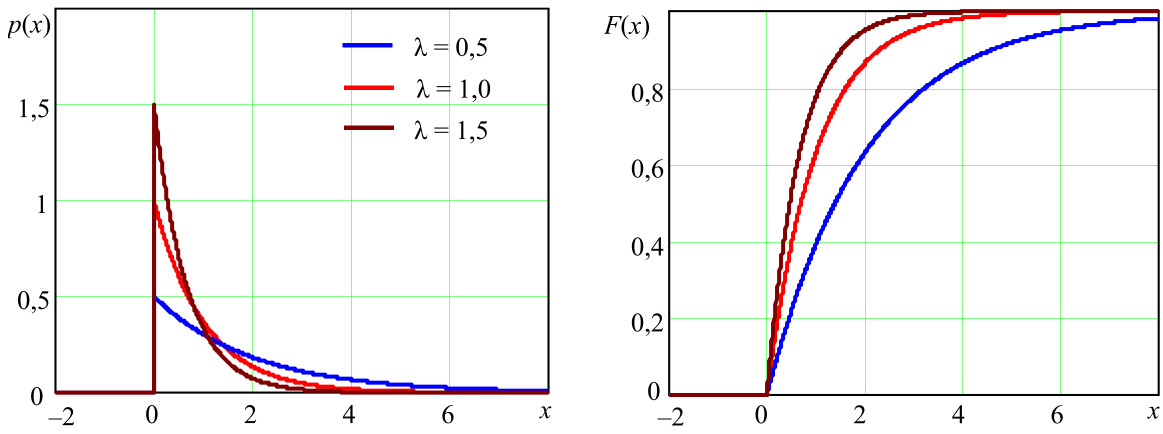


Рис. 2.15. Плотность распределения p(x) и функция распределения F(x) для экспоненциального закона распределения

Основные числовые характеристики экспоненциального распределения с учетом его параметра примут вид: мат. ожидание $-m=\dfrac{1}{\lambda}$, дисперсия – $D=\dfrac{1}{\lambda^2}$, асимметрия – $A=2$, эксцесс – $E=6$.

Нормальное распределение (распределение Гаусса) – распределение вероятностей, которое играет важнейшую роль в теории вероятностей и занимает среди других законов распределения особое положение.

Если случайная величина образуется в результате суммирования многочисленных составляющих, возникающих под действием случайных факторов, то, согласно предельной теореме Ляпунова, она подчиняется нормальному закону распределения. Такая ситуация крайне распространена, поэтому можно сказать, что из всех распределений в природе чаще всего встречается именно нормальное распределение. Главная особенность нормального закона распределения состоит в том, что он является предельным законом, к которому приближаются другие при типичных ситуациях.

Плотность распределения $p(x)$ и функция распределения случайной величины X имеют вид:

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);\ F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx. \quad (2.28)$$

Параметрами нормального распределения являются: m – мат. ожидание (среднее значение); $\sigma(\sigma>0)$ – СКО (разброс относительно среднего значения).

Нормальный закон распределения случайной величины с параметрами $m = 0$ и $\sigma = 1$ называется *стандартным*, или *нормированным*.

Плотность стандартного нормального распределения имеет вид

$$p(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right), \quad (2.29)$$

а его функция распределения $F(x)=\Phi(x)$ называется функцией Лапласа:

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{z^2}{2}\right)dz. \quad (2.30)$$

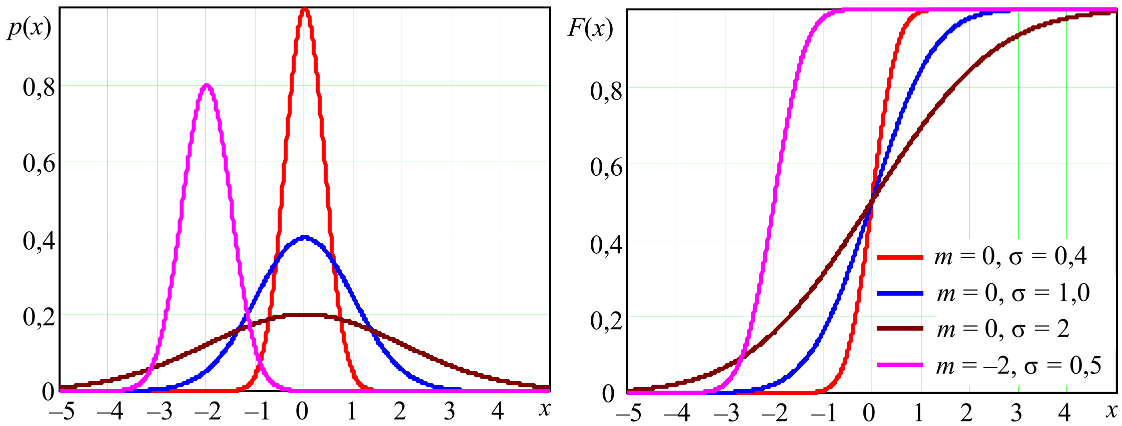


Рис. 2.16. Плотность распределения p(x) и функция распределения F(x) для нормального закона распределения

Функция распределения нормальной величины с параметрами m и σ также можно выразить через функцию Лапласа:

$$F(x)=\Phi\left(\frac{x-m}{\sigma}\right). \quad (2.31)$$

Плотность распределения $p(x)$ и функция распределения $F(x)$ для различных параметров нормального распределения показаны на рис. 2.16.

Основные числовые характеристики нормального распределения с учетом его параметров примут вид: мат. ожидание – m, дисперсия – $D=\sigma^2$, асимметрия – $A=0$, эксцесс – $E=0$.

Правило 3σ. Практически все значения нормально распределенной СВ находятся в интервале

$$[m - 3\sigma; m + 3\sigma]. \tag{2.32}$$

Вероятность отклонения значения случайной величины от мат. ожидания менее чем на 3σ (согласно формулы (2.31)) равна:

$$P(|X - m| < 3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3).$$

По таблице для интегральной функции Лапласа (прил. 1, табл. П1.2) получается:

$$2\Phi(3) = 2 \cdot 0,4978 = 0,9973, \text{ или } 99,73\ \%.$$

Другими словами, в партии из 1000 деталей отклонение размера более чем на 3σ будет наблюдаться всего лишь у 27 экземпляров.

Правило 3σ наглядно представлено на рис. 2.17.

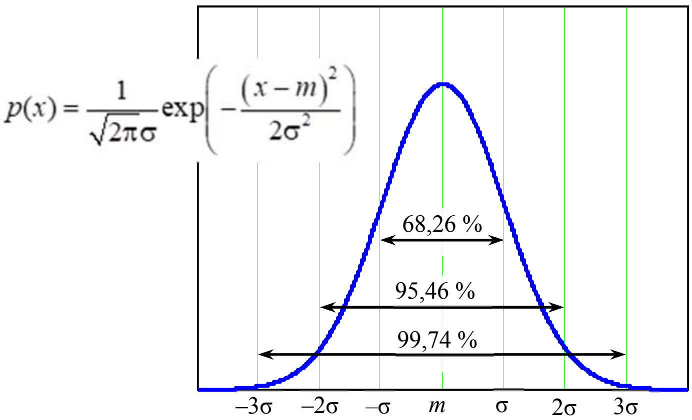


Рис. 2.17. Правило 3σ

С этим правилом связана часто используемая доверительная вероятность, равная 95 % (подробно в разд. 4.4).

С помощью нормального распределения определяются три других распределения: распределение Пирсона (хи-квадрат), распределение Стьюдента и распределение Фишера, которые получили широкое распространение при статистической обработке данных.

Распределение Стьюдента играет важную роль в статистическом анализе. Его применяют при оценивании математического ожидания, прогнозного значения и других характеристик с помощью доверительных интервалов (разд. 4.3), при проверке гипотез о значениях математических ожиданий (разд. 5.2.1), при расчете коэффициентов регрессионной зависимости (разд. 6.4) и т. д.

Распределение Стьюдента близко по форме к нормальному распределению, но отличается от него концентрацией отклонений в центральной части распределения. Малые выборки объемом n, полученные из нормального распределения, имеют распределение Стьюдента с k степенями свободы, где $k = n - 1$.

Плотность вероятности этой величины вычисляется по формуле

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in R, \tag{2.33}$$

где $\Gamma(x) = \int\limits_0^{\infty} t^{x-1} e^{-t} dt$ – гамма-функция.

При больших n распределение Стьюдента практически не отличается от стандартного нормального распределения с параметрами $m = 0$ и $\sigma = 1$.

Параметром распределения Стьюдента является n – число степеней свободы.

Плотность распределения $p(x)$ и функция распределения $F(x)$ для различных параметров распределения Стьюдента показаны на рис. 2.18.

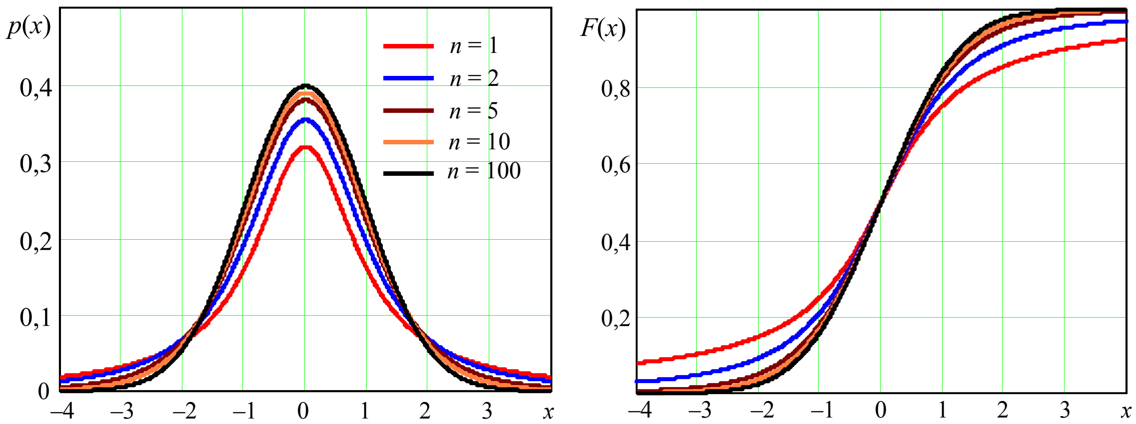


Рис. 2.18. Плотность распределения p(x) и функция распределения F(x) для закона распределения Стьюдента

Основные числовые характеристики нормального распределения с учетом его параметров примут вид: мат. ожидание $-m = 0$ (если $n > 1$), дисперсия – $D = \frac{n}{n-2}$ (если $n > 2$), асимметрия – $A = 0$ (если $n > 3$), эксцесс – $E = \frac{6}{n-4}$ (если $n > 4$).

Распределение Пирсона характерно для суммы независимых случайных величин $X_1, X_2, \ldots, X_n$, каждая из которых распределена по нормальному закону распределения с параметрами $m = 0$ и $\sigma = 1$, т. е. все имеют *стандартное нормальное распределение*.

Тогда сумма квадратов этих распределений $\chi^2 = \sum X_n^2$ будет иметь χ^2-*распределение* (распределение Пирсона, или распределение «хи-квадрат»).

Распределение Пирсона используют при оценивании дисперсии (с помощью доверительного интервала (разд. 4.4.2), при проверке гипотез согласия (разд. 5.3.2) и во многих других задачах статистического анализа данных.

Плотность вероятности этой величины вычисляется по формуле

$$f(x) = \frac{1}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)} e^{-x/2} \cdot x^{(n/2)-1}, \quad x > 0, \tag{2.34}$$

где $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ – гамма-функция.

Параметром распределения Пирсона является n – число степеней свободы.

Плотность распределения $p(x)$ и функция распределения $F(x)$ для различных параметров распределения Пирсона показаны на рис. 2.19.

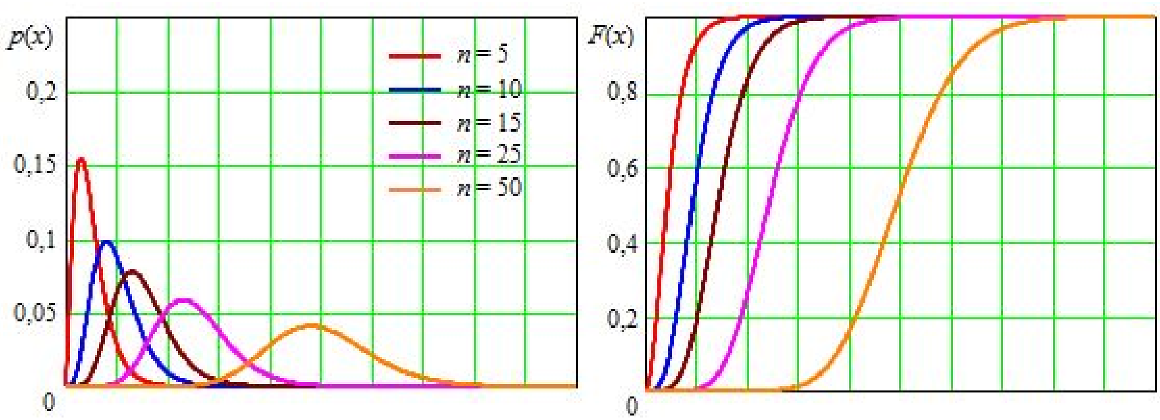


Рис. 2.19. Плотность распределения p(x) и функция распределения F(x) для закона распределения Пирсона

Распределение Фишера – это отношение независимых случайных величин X_1 и X_2, имеющих распределение Пирсона со степенями свободы k_1 и k_2 соответственно:

$$F = \frac{\frac{1}{k_1} X_1}{\frac{1}{k_2} X_2}. \tag{2.35}$$

Распределение случайной величины F названо в честь великого английского статистика Р. Фишера и используется при проверке гипотез о равенстве дисперсий (разд. 5.2.2) и в других задачах прикладной статистики.

Параметрами распределения Фишера являются: k_1 – число степеней свободы для случайной величины X_1; k_2 – число степеней свободы для случайной величины X_2.

Плотность распределения $p(x)$ и функция распределения $F(x)$ для различных параметров распределения Фишера показаны на рис. 2.20.

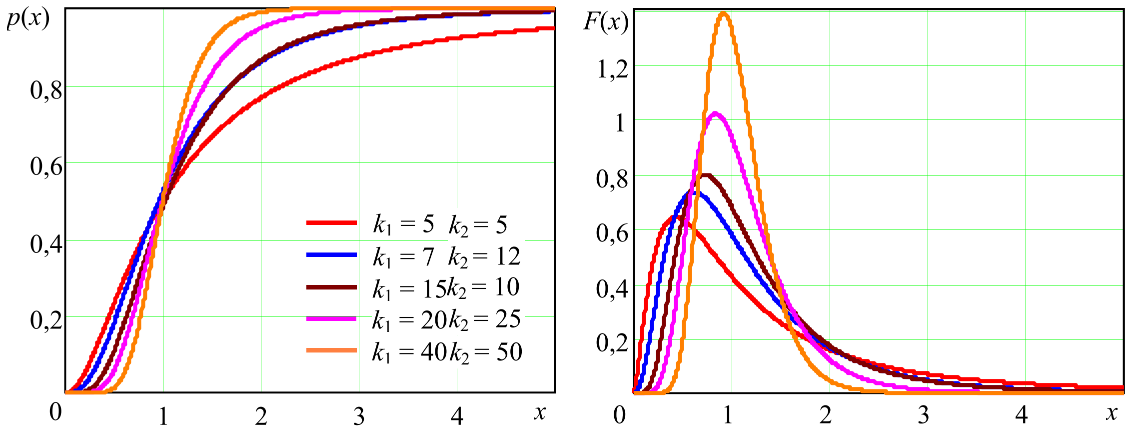


Рис. 2.20. Плотность распределения p(x) и функция распределения F(x) для закона распределения Фишера

Контрольные вопросы и задания

1. Что такое случайная величина: дискретная и непрерывная?
2. Как различить функции распределения и плотности вероятностей дискретных и непрерывных случайных величин?
3. Что такое числовые характеристики случайных величин?
4. Что такое математическое ожидание и дисперсия случайной величины? Приведите формулы для расчета математического ожидания и дисперсии непрерывной и дискретной случайных величин.
5. Дайте определение и приведите расчетные формулы начальных и центральных моментов.
6. Что такое коэффициент асимметрии, коэффициент эксцесса? Приведите расчетные формулы.
7. Какие величины являются характеристиками положения?
8. Какие величины являются характеристиками рассеивания?
9. Какие законы распределения случайных величин широко встречаются в измерительной практике?

Глава 3. Элементы математической статистики

3.1. Задачи математической статистики

Математическая статистика – раздел математики, посвященный математическим методам сбора, систематизации, обработки и интерпретации статистических данных, а также использованию их для научных или практических выводов.

Статистические данные – это сведения о числе объектов (или их параметрах), обладающих некоторыми признаками, в какой-либо совокупности. Статистические данные получаются в результате различных наблюдений, опросов или экспериментов.

Цель математической статистики состоит в исследовании свойств случайной величины на основе полученных статистических данных (результатов многократного наблюдения случайной величины), т. е. в установлении закономерностей, которым подчинены полученные данные. Правила и процедуры математической статистики опираются на теорию вероятностей, позволяют оценить точность и надежность выводов, получаемых в каждой задаче на основании имеющегося статистического материала.

При статистическом анализе в каждом конкретном случае возможно решение ряда задач, каждая из которых в общем случае включает два этапа:

- обработку статистических данных: указание способов сбора и группировки статистических данных;
- принятие решения: разработка, обоснование и применение различных статистических методов анализа с целью получения научных и практических выводов.

Принимаемое решение носит вероятностный характер, так как объем выборки ограничен (конечен) и содержимое ее случайно (присутствует условие неопределенности). Математическая статистика – это наука о принятии решений в *условиях неопределенности*.

Задачами современной математической статистики являются:

- разработка способов определения числа необходимых испытаний (наблюдений) на разных этапах: до начала исследования (планирование эксперимента) и в процессе исследования (последовательный анализ) (глава 3);
- определение законов распределения (оценка функции распределения или плотности распределения для неизвестного закона) (глава 3);
- проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен (глава 5);
- оценка параметров известного закона распределения (глава 4).

Эти задачи существенно различаются как по внутреннему содержанию, так и по методам их решения. Все они будут рассмотрены в соответствующих разделах.

3.2. Краткая историческая справка

Как наука математическая статистика возникла в XVII веке и развивалась параллельно с теорией вероятностей, но первые типовые примеры применения статистических методов описаны еще в Ветхом Завете (Библия). Там, в частности, приводится число воинов в различных племенах. С математической точки зрения дело сводилось к подсчёту числа попаданий значений наблюдаемых признаков в определённые градации.

В XVII веке вероятностные модели стали использоваться при обработке статистических данных [1].

Пример 1. При изучении частоты рождения мальчиков и девочек в Париже в XVII веке было установлено отличие вероятности рождения мальчика от 0,5. Был проведен анализ причины того, что в парижских приютах эта вероятность отличалась от данных по Парижу в целом [1].

Считается, что математическая статистика начинается с работ немецкого математика К.Ф. Гаусса (рис. 1.6). Он исследовал и обосновал *метод наименьших квадратов* (один из методов современной математической статистики), созданный им в 1795 г. и примененный для обработки астрономических данных (с целью уточнения орбиты малой планеты Церера).

В XIX веке значительный вклад в развитие практической статистики внёс бельгиец А. Кетле (рис. 3.1) – один из родоначальников научной статистики. Кетле усовершенствовал статистический метод, развил его, философски обосновал и применил с большим успехом в своих исследованиях.

Рис. 3.1. Адольф Кетле (22 февраля 1796 г. – 17 февраля 1874 г.) – бельгийский математик, астроном, метеоролог, социолог

Рис. 3.2. Рональд Эйлмер Фишер (17 февраля 1890 г. – 29 июля 1962 г.) – английский статистик, биолог-эволюционист и генетик

Рассматривая пример «доля самоубийств среди всех смертей», Кетле показал устойчивость относительных статистических показателей при анализе большого числа реальных данных.

В конце XIX – начале XX в. крупный вклад в математическую статистику внесли английские исследователи, прежде всего К. Пирсон (см. рис. 1.11), Р.А. Фишер (рис. 3.2), Е. Нейман (рис. 3.3). Они заложили основы теории анализа данных (параметрической статистики). *Параметрической статистикой* ее называют потому, что основным объектом изучения является выборка из распределений, описываемых одним или небольшим числом параметров.

Рис. 3.3. Е́жи Нейман (16 апреля 1894 г. – 5 августа 1981 г.) – польский и американский математик и статистик, член Национальной АН США (1963)

Рис. 3.4. Николай Васильевич Смирнов (17октября 1900 г. – 2 июня 1966 г.) – советский математик, член-корреспондент АН СССР

Пирсон разработал теорию корреляции, критерии согласия, алгоритмы принятия решений и оценки параметров. С его именем связаны множество широко используемых терминов и методов (*коэффициент корреляции*, *критерий «хи-квадрат»*, *нелинейная регрессия* и т. д.).

Фишер, наряду с Нейманом, разработал фундамент теории оценок параметров, статистических решений, планирования эксперимента и проверки гипотез. Большинство методов Фишера (*дисперсионный анализ*, *теория планирования эксперимента*, *метод максимального правдоподобия оценки параметров*) имеют общий характер и применяются в естественных науках, в экономике и других областях деятельности.

Нейман Е. развивал так называемую бихевиористскую статистику (методологию принятия решений в условиях неопределённости), которая нашла множество применений в научных исследованиях в астрономии, физике, биологии, медицине – везде, где необходимо снижать частоту ошибок.

Советские математики академик А.Н. Колмогоров (см. рис. 1.12) и член-корреспондент АН СССР Н.В. Смирнов (рис. 3.4) заложили основы непараметрической статистики. Непараметрическая статистика включает в себя описательную статистику и статистические выводы, т. е. обобщение информации из выборки для получения представления о свойствах генеральной совокупности.

Величайший вклад в развитие статистики внес не академический ученый, а скромный пивовар из компании «Гиннесс» (Guinness). Он применял свои знания в области статистики при варке пива и для выведения самого урожайного сорта ячменя.

Уильям Сили Госсет (рис. 3.5) приобретал эти знания на практике в биометрической лаборатории Карла Пирсона в период с 1906 по 1907 г. Пирсон помогал Госсету в математической части его исследований, которые были обращены к нуждам пивоваренной компании и проводились на малом количестве наблюдений, поэтому Пирсон не придавал им особого значения. Первым, кто понял значение работ Госсета по оценке параметров малой выборки, был биолог Рональд Фишер.

Самое важное открытие Госсета получило название «Распределение Стьюдента» (введено в 1908 г.), потому что он вынужден был работать под псевдонимом, чтобы скрыть свои научные публикации от работодателя. Гиннесс запретил своим работникам публикацию любых материалов, независимо от содержавшейся в них информации, чтобы предотвратить раскрытие сведений, составлявших коммерческую тайну пивоваренной компании «Гиннесс».

Госсет ввел в математическую статистику понятие «статистическая значимость», разработал систему промышленного контроля качества и метод эффективного планирования экспериментов.

Рис. 3.5. Уильям Сили Госсет (известный под псевдонимом Стьюдент) (13 июня 1876 г. – 16 октября 1937 г.) – британский ученый-статистик

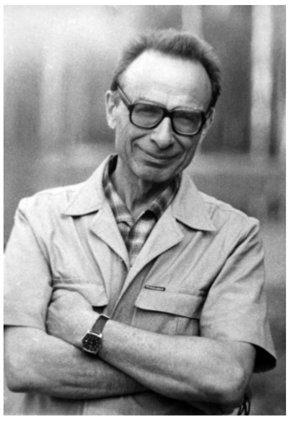


Рис. 3.6. Абрахам Вальд (31 октября 1902 г. – 13 декабря 1950 г.) – венгерский математик и статистик

В сороковые годы XX века венгерский математик и статистик Абрахам Вальд (рис. 3.6) построил теорию последовательного статистического анализа. Впервые последовательный подход был использован в задаче приемочного статистического контроля в 1929 г. В годы Второй мировой войны Вальд использовал статистические методы для решения проблемы уменьшения потерь американской боевой авиатехники. Он построил теорию последовательного анализа применительно к вопросу различения статистических гипотез и сформулировал общую задачу последовательного оценивания [1–4].

Математическая статистика бурно развивается и в настоящее время. На основе математической статистки особенно интенсивно разрабатываются статистические методы исследования и контроля массового производства, статистические методы в области физики, гидрологии, климатологии, звёздной астрономии, биологии, медицины и другие [2, 3].

В конце XX века можно выделить четыре принципиально новых направления исследований в области математической статистики:

- разработка и внедрение математических методов планирования экспериментов;
- развитие статистики объектов нечисловой природы как самостоятельного направления в прикладной математической статистике;
- развитие статистических методов, устойчивых по отношению к малым отклонениям от используемой вероятностной модели;
- широкое развитие компьютерных пакетов программ, предназначенных для проведения статистического анализа данных [2, 3].

3.3. Генеральная и выборочная совокупности

На практике часто требуется изучить совокупность однородных объектов относительно некоторого *качественного* или *количественного* признака, характеризующего эти объекты.

Признак[2] – это свойство, которое помогает идентифицировать или различать объекты данной совокупности. Каждое отдельное значение признака, полученное в результате наблюдения (измерения), называется *наблюдаемым значением.*

Пример 1. При рассмотрении партии деталей *качественным* признаком может служить стандартность (годность) детали, а *количественным* – числовое значение контролируемого размера детали.

---

[2] В этой главе большинство терминов приводятся в соответствии с ГОСТ Р 50779.10-2000 (ИСО 3534.1-93) «Статистические методы. Вероятность и основы статистики. Термины и определения» [9].

При изучении объектов иногда проводится *сплошное обследование*, т. е. исследуется интересуемый признак *каждого* объекта генеральной совокупности.

Генеральная совокупность – это множество всех рассматриваемых объектов.

Однако на практике сплошное обследование применяют сравнительно редко, потому что:

- сплошное обследование большого числа объектов (всей генеральной совокупности) иногда физически невозможно (требуется много временных, человеческих и/или материальных ресурсов);
- иногда исследование объекта связано с его уничтожением, поэтому проводить сплошное обследование не имеет смысла.

В случае невозможности проведения сплошного обследования изучению подвергают ограниченное число объектов (выборку).

Выборочная совокупность (выборка) – это совокупность объектов, взятых из генеральной совокупности и предназначенных для получения информации о ней.

При изучении выборки реализуется *выборочный метод*, сущность которого заключается в следующем: из *генеральной совокупности* случайным образом извлекается ограниченное число объектов, образующих *выборку*. Выборка подвергается детальному исследованию, по результатам которой определяются свойства (характеристики). Полученные данные применяются в отношении генеральной совокупности.

Объемом совокупности (выборочной или генеральной) называют число объектов совокупности, N обозначается объем генеральной совокупности, n – объем выборки.

Пример 2. Если из 1000 деталей отобрано для обследования 100 деталей, то $N = 1000$ рассматривается как объем генеральной совокупности, а $n = 100$ – объем выборки.

Если рассматривать объекты неслучайной природы (например, дефектность детали или ее габаритный размер, выраженный в см), то в математической статистике принято считать, что генеральная совокупность имеет конечный объем N.

Если генеральная совокупность представлена объектами случайной природы (результаты случайных процессов, например погрешности измерений каких-либо параметров), то объем генеральной совокупности принято считать бесконечным ($N \to \infty$).

После проведения наблюдения над объектом, взятым из выборки, он *возвращается* либо *не возвращается* в генеральную совокупность. Исходя из этого выборки подразделяют на *повторные* и *бесповторные*.

Повторной называют выборку, при которой отобранный объект (перед отбором следующего) возвращается в генеральную совокупность. Вероятность попасть в выборку для каждой единицы генеральной совокупности не меняется, независимо от числа отбираемых единиц.

Бесповторной называют выборку, при которой отобранный объект в генеральную совокупность не возвращается. Вероятность попасть в выборку для каждой единицы генеральной совокупности увеличивается по мере реализации отбора.

При достаточно большом объеме генеральной совокупности и незначительном объеме выборки различие между повторной и бесповторной выборками стирается.

В предельном случае, когда объем генеральной совокупности $N \to \infty$, а выборка имеет конечный объем n, это различие исчезает. Поэтому на практике обычно используется бесповторный случайный отбор, а повторная выборка, как правило, применяется лишь в теоретических исследованиях.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо при формировании выборки обеспечить:

- достаточный объем;
- репрезентативность (представительность).

Необходимый объем выборки зависит от решаемой задачи и от требуемой точности (надежности) получаемых результатов. В каждом конкретном случае определяется индивидуально.

Репрезентативность – соответствие характеристик выборки характеристикам генеральной совокупности в целом. В силу закона больших чисел можно утверждать, что выборка будет *репрезентативной*, если каждый объект выборки отобран случайным образом и все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку.

3.4. Способы отбора

На практике применяются различные способы отбора выборки из генеральной совокупности, которые можно подразделить на два вида:

1. Отбор, не требующий расчленения генеральной совокупности на части. К этому виду относится простой случайный отбор.
2. Отбор, при котором генеральная совокупность разбивается на части:
   а) типический отбор;
   б) механический отбор;
   в) серийный отбор.

Простым случайным называют отбор, при котором объекты извлекают по одному из генеральной совокупности. Осуществить простой отбор можно различными способами.

Пример 1. Если объекты достаточно велики (объем, вес и т. д.), чтобы их перемешать между собой, то поступают так: все объекты нумеруют от 1 до N, затем на карточках выписывают номера от 1 до N, тщательно перемешивают и наугад вынимают одну карточку. Объект, имеющий одинаковый номер с извлеченной карточкой, подвергают обследованию. Затем карточку возвращают в пачку, карточки перемешивают, наугад вынимают одну из них и т. д. Так поступают n раз. В итоге получают простую случайную повторную выборку объема n. Если карточку в пачку не возвращают, то получают простую случайную бесповторную выборку.

При большом объеме генеральной совокупности описанный процесс оказывается очень трудоемким. В этом случае пользуются готовыми таблицами «случайных чисел», в которых числа расположены в случайном порядке, или используют программы для генерации случайных чисел.

Пример 2. Необходимо отобрать 50 объектов из пронумерованной генеральной совокупности, для этого необходимо сгенерировать n случайных чисел. Лучше для этого использовать программу «Генератор случайных чисел» или по «старинке» использовать таблицы случайных чисел. В выборку попадают те объекты, номера которых совпадают с «выбранными» случайными числами.

Типическим называют отбор, при котором объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части (например, изделия, изготовленные разными сменами рабочих).

Пример 3. Если детали изготовляют на нескольких станках, то отбор производят не из всей совокупности деталей, произведенных всеми станками, а из продукции каждого станка в отдельности. Типическим отбором пользуются тогда, когда обследуемый признак заметно колеблется в различных типических частях генеральной совокупности: продукция изготовляется на нескольких машинах, имеющих разную степень износа.

Механическим называют отбор, при котором генеральную совокупность «механически» делят на столько групп, сколько объектов должно войти в выборку. Далее из каждой группы отбирают один объект, который подвергается обследованию.

Пример 4. Если нужно отобрать 20 % изготовленных станком деталей, то отбирают каждую пятую деталь; если требуется отобрать 5 % деталей, то отбирают каждую двадцатую деталь, и т. д.

Следует указать, что иногда механический отбор может не обеспечивать репрезентативности выборки.

Пример 5. Если отбирают каждый двадцатый обтачиваемый валик, причем сразу же после отбора производят замену резца, то отобранными окажутся все валики, обточенные затупленными резцами. В таком случае следует устранить совпадение ритма отбора с ритмом замены резца, для чего надо отбирать, скажем, каждый десятый валик из двадцати обточенных.

Серийным называют отбор, при котором объекты отбираются из генеральной совокупности не по одному, а «сериями». Каждую серию подвергают сплошному обследованию.

Пример 6. Если изделия изготовляются большой группой станков-автоматов, то подвергают сплошному обследованию продукцию (например, по 10 штук) только нескольких станков.

Серийным отбором пользуются тогда, когда обследуемый признак колеблется в различных сериях незначительно.

Подчеркнем, что на практике часто применяется комбинированный отбор, при котором используются различные сочетания указанных выше способов.

Пример 7. Иногда генеральную совокупность разбивают на серии одинакового объема, затем простым случайным отбором выбирают несколько серий. На следующем этапе из каждой серии простым случайным отбором извлекают отдельные объекты, которые подвергаются обследованию [8].

3.5. Статистическое распределение выборки

Пример 1. При проведении исследования могут быть получены данные:

- значения диаметра изготавливаемой детали;
- значения температуры при проведении серии измерений;
- число отказов аппаратуры за 1000 часов работы;
- и т. д.

Полученные значения представляют собой выборку из n элементов.

Представить выборку можно в виде *простого статистического ряда* – последовательность значений, перечисленная в порядке их получения.

После первичного просмотра выборки можно обнаружить, что значение x_1 наблюдалось n_1 раз, $x_2 - n_2$ раз, $x_i - n_i$ раз. Объем выборки равен:

$$n = \sum n_i \,. \tag{3.1}$$

Наблюдаемое значение x_i называется *вариантой.* Число повторений n_i называется *частотой*, а отношение частоты к объему выборки – *относительной частотой* W_i:

$$W_i = \frac{n_i}{n}\,. \tag{3.2}$$

Совокупность вариант и частот (относительных частот) полученных результатов исследования выборки называется *статистической совокупностью (или распределение (относительных) частот)*.

Последовательность расположенных в порядке возрастания вариант с указанием соответствующих им частот называется *вариационным рядом.*

Если закон распределения измеряемой случайной величины не известен заранее, то первым шагом к осмысливанию ряда наблюдений является построение вариационного ряда.

3.5.1. Дискретный вариационный ряд

Если выборка имеет выраженный дискретный характер, общее значение вариант невелико и достаточно легко посчитать частоту для каждой варианты, то выборку можно представить в виде дискретного вариационного ряда.

Дискретный вариационный ряд – это упорядоченное по возрастанию множество вариант (значений величины $x_1, x_2, \ldots, x_l$) и соответствующих им частот n_i либо относительных частот W_i, представленных в виде табл. 3.1.

Таблица 3.1

Дискретный вариационный ряд

x_i	x_1	x_2	…	x_l
n_i	n_1	n_2	…	n_l
W_i	W_1	W_2	…	W_l

Для проверки правильности заполнения дискретного вариационного ряда следует проверить выполнение условий:

$$\sum n_i = n; \ \sum W_i = 1. \tag{3.3}$$

Как правило, дискретный вариационный ряд применяется для представления дискретной СВ.

Пример 2. Дан ряд распределения частот для выборки объема $n = 20$ (табл. 3.2).

Таблица 3.2

Дискретный ряд распределения частот

x_i	2	6	12
n_i	20	20	20

Записать распределение относительных частот.

Решение

Относительные частоты выборки рассчитываются по формуле (3.2):

$$W_1 = \frac{3}{20} = 0{,}15; \quad W_2 = \frac{10}{20} = 0{,}5; \quad W_3 = \frac{7}{20} = 0{,}35.$$

Для контроля проверяется выполнение условия (3.3):

$$W = 0{,}15 + 0{,}5 + 0{,}35 = 1.$$

Ответ: результат представлен в виде табл. 3.3.

Таблица 3.3

Распределение относительных частот

x_i	2	6	12
W_i	0,15	0,50	0,35

Пример 3. Количество филиалов в городе у двадцати разных банков: 2, 4, 3, 5, 4, 4, 6, 5, 4, 3, 4, 3, 4, 5, 3, 4, 6, 3, 5, 4. Построить ряд распределения по имеющимся данным.

Решение

Вариация значений случайной величины имеет дискретный характер, число вариант невелико, поэтому целесообразно построить дискретный ряд распределения. Для его построения следует перечислить все встречающиеся варианты значений СВ x_i и подсчитать частоту их повторения n_i.

Также можно вычислить относительные частоты W_i по формуле (3.2), объем выборки $n = 20$.

Дискретный ряд распределения примет вид, приведенный в табл. 3.4.

Таблица 3.4

Распределение по количеству филиалов банков

Количество филиалов в городе, x_i	Число банков (или частота, n_i)	Относительная частота, W_i
2	1	1/20 = 0,05
3	5	5/20 = 0,25
4	8	8/20 = 0,40
5	4	4/20 = 0,20
6	2	2/20 = 0,10
Итого:	**20**	**1**

Для контроля проверяется выполнение условий (3.3):

$$n = 1 + 5 + 8 + 4 + 2 = 20;$$

$$W = 0,05 + 0,25 + 0,4 + 0,2 + 0,1 = 1.$$

Ответ: полученный ряд распределения представлен в табл. 3.4.

3.5.2. Интервальный вариационный ряд

Непрерывная случайная величина принимает слишком много различных значений, зачастую даже неповторяющихся, поэтому строить дискретный ряд трудоемко и нецелесообразно.

В подобных случаях следует применять *интервальный вариационный ряд* распределения. Для его построения весь интервал наблюдаемых значений НСВ разбивается на ряд частичных интервалов и подсчитывается частота попадания значений величины в каждый частичный интервал.

Интервальный вариационный ряд – это упорядоченная совокупность интервалов варьирования значений случайной величины X с соответствующими частотами n_i или относительными частотами W_i попаданий в каждый.

Алгоритм построения интервального вариационного ряда

Для построения интервального ряда необходимо:

1. *Определение размаха распределения.*

Размах выборки – разность между наибольшим и наименьшим наблюденными значениями количественного признака в выборке (вариантами).

Размах варьирования случайной величины R рассчитывается по формуле

$$R = x_{\max} - x_{\min}, \tag{3.4}$$

где $x_{\max}$ и $x_{\min}$ – максимальная и минимальная варианты выборки.

2. *Определение числа интервалов.*

Для определения оптимального числа интервалов L можно использовать формулу (правило) Стерджеса.

Правило Стерджеса – эмпирическое правило определения оптимального количества интервалов, на которые разбивается наблюдаемый диапазон изменения случайной величины при построении интервального вариационного ряда. Формула, названая по имени американского статистика Герберта Стерджеса (Herbert Arthur Sturges, 1882–1958), имеет вид

$$L = 1 + \log_2 n = 1 + 3,322 \cdot \lg(n), \tag{3.5}$$

где n – общее число элементов выборки.

Полученную величину L округляют обычно до целого большего числа, поскольку количество групп не может быть дробным числом.

Если интервальный ряд с таким количеством групп по каким-то критериям не устраивает, то можно построить другой интервальный ряд, округлив L до целого меньшего числа, и выбрать из двух рядов более подходящий.

Существуют и другие формулы для определения числа интервалов разбиения выборки. О них можно посмотреть в различной литературе [7, 8, 10].

Следует отметить, что правило Стерджеса носит *рекомендательный* характер. Нередко в условии задачи прямо сказано, на какое количество интервалов нужно проводить разбиение (на 4, 5, 6, 10 и т. д.), и тогда следует придерживаться именно этого указания. Если результат разбиения не удовлетворяет каким-либо требованиям, можно использовать любое другое правило.

3. *Определение ширины интервалов.*

Ширина интервала для интервального вариационного ряда с равными интервалами определяется по формуле

$$h = \frac{x_{\max} - x_{\min}}{L}. \qquad (3.6)$$

Величина интервала h обычно округляется до того разряда, который меняется в значениях выборки.

4. *Установление границ интервалов.*

Нижняя граница первого интервала выбирается равной минимальному значению выборки: $a_1 = x_{\min}$.

Промежуточные интервалы получаются путем прибавления к концу предыдущего интервала длины интервала h:

$$a_i = x_{\min} + (i-1) \cdot h, \qquad (3.7)$$

где $i = 1, \ldots, L + 1$ – номер интервала.

Построение шкалы интервалов на основе вычисления границ интервалов продолжается до тех пор, пока величина a_i не будет удовлетворять соотношению

$$a_i \geq x_{\max}.$$

5. *Группировка результатов наблюдения.*

В соответствии со шкалой интервалов производится группирование значений выборки – для каждого интервала вычисляется сумма частот n_i значений вариант (и/или сумма относительных частот W_i), попавших в i-й интервал. При этом в интервал включают значения случайной величины (варианты), *большие* или *равные нижней* границе и *меньшие верхней* границы интервала:

$$a_i \le x_i < a_{i+1}.$$

Результаты группировки заносятся в табл. 3.5, которая представляет собой *интервальный вариационный ряд*, и позволяют судить о распределении значений непрерывной случайной величины X между интервалами.

Таблица 3.5

Интервальный вариационный ряд

№ интервала	Границы i-го интервала	Абсолютные частоты	Относительные частоты
1	$a_1 - a_2$	n_1	W_1
2	$a_2 - a_3$	n_2	W_2
...	...	...	...
i	$a_i - a_{i+1}$	n_i	W_i

Для проверки правильности выполнения операций при построении интервального вариационного ряда следует проверить выполнение условий (3.3).

Обычно предпочтительны интервалы одинаковой ширины, но бывают исключения [7, 8, 10].

Пример 3. Общая численность сотрудников фирмы составляет 120 человек, а минимальный и максимальный доходы соответственно равны 500 и 6500 у. е. Пользуясь формулой Стерджеса, определить *частичные* интервалы группировки сотрудников фирмы по уровню доходов.

Решение

Оптимальное количество интервалов рассчитывается по формуле Стерджеса (3.5):

$$L = 1 + 3{,}322 \cdot \lg(120) = 7{,}9.$$

Поскольку количество интервалов не может быть дробным числом, число L округляется до ближайшего большего целого числа:

$$L = 8 \text{ интервалов}.$$

Ширина интервалов рассчитываются в соответствии с формулой (3.6):

$$h = \frac{x_{\max} - x_{\min}}{L} = \frac{6500 - 500}{8} = 750.$$

Нижняя граница первого интервала выбирается равной минимальному значению выборки: $a_1 = 500$ у. е. Остальные границы рассчитываются по формуле (3.7):

$$a_i = 500 + (i - 1) \cdot 750.$$

Интервалы группировки представлены в табл. 3.6.

Таблица 3.6

Группировка сотрудников по уровню доходов

№ интервала	Границы i-го интервала, у. е.
1	500…1250
2	1250…2000
3	2000…2750
4	2750…3500
5	3500…4250
6	4250…5000
7	5000…5750
8	5750…6500

Так как сама выборка в условии задачи не приводится, то распределение значений по рассчитанным интервалам невозможно.

Ответ: частичные интервалы группировки сотрудников фирмы по уровню доходов представлены в табл. 3.6.

Пример 4. Имеются следующие данные о размере прибыли (млн руб.) двадцати коммерческих банков: 3,7; 4,3; 6,7; 5,6; 5,1; 8,1; 4,6; 5,7; 6,4; 5,9; 5,2; 6,2; 6,3; 7,2; 7,9; 5,8; 4,9; 7,6; 7,0; 6,9. Построить ряд распределения по имеющимся данным.

Решение

Значения случайной величины носят непрерывный характер, значения в выборке не повторяются, поэтому целесообразно строить интервальный ряд распределения.

Так как изначально количество интервалов разбиения не задано, то их количество определяется по формуле Стерджеса (3.5):

$$L = 1 + 3,322 \cdot \lg(20) = 5,3.$$

Поскольку количество интервалов не может быть дробным числом, число L округляется до ближайшего большего целого числа:

$$L = 6 \text{ интервалов}.$$

Ширина интервала определяется по формуле (3.6):

$$h = \frac{x_{\max} - x_{\min}}{L} = \frac{8,1 - 3,7}{6} = 0,733.$$

Число, характеризующее величину интервала, округляется с той же точностью, что и исходные данные, следовательно, значение следует округлить до 0,1: $h = 0,8$.

Если округлить в меньшую сторону (h = 0,7), то размах интервального ряда будет равен: $R = h \cdot L = 0,7 \cdot 6 = 4,2$. Полученный размах не соответствует заданной выборке, которая, согласно формуле (3.4), имеет размах $R = x_{\max} - x_{\min} = 8,1 - 3,7 = 4,4$.

Нижняя граница первого интервала выбирается равной минимальному значению выборки: $a_1 = 3,7$ млн руб., остальные границы рассчитываются по формуле (3.7):

$$a_i = 3,7 + (i-1) \cdot 0,8 .$$

Интервалы группировки представлены во втором столбце табл. 21.

Распределение значений выборки по рассчитанным интервалам проводится путем сортировки (ст. 3 табл. 21). Поскольку значение 6,9 попало на границу четвертого и пятого интервалов, то его следует отнести к тому интервалу, где это значение является нижней границей интервала, т. е. к пятому интервалу.

Относительная частота W_i рассчитывается по формуле (3.2). Результаты представлены в четвёртом столбце табл. 3.7.

Таблица 3.7

Группировка банков по объему прибили

№ интервала	Границы интервалов, $a_i - a_{i+1}$, млн руб.	Число банков (частота), n_i	Относительная частота, W_i
1	3,7…4,5	2	0,1
2	4,5…5,3	4	0,2
3	5,3…6,1	4	0,2
4	6,1…6,9	4	0,2
5	6,9…7,7	4	0,2
6	7,7…8,5	2	0,1
Итого:		**20**	**1**

Ответ: интервальный ряд распределения двадцати коммерческих банков по размеру прибыли приведен в табл. 3.7.

3.6. Эмпирическая функция распределения

Относительная частота события «*Значение случайной величины X меньше некоторого значения x* ($X < x_i$)» равна $\frac{n_x}{n}$, где $n_x = \sum_{x_i < x} n_i$ – сумма всех частот для значений случайной величины $X < x_i$.

Если x изменяется, то изменяется и относительная частота $\frac{n_x}{n}$. Так как эта функция находится опытным путем, то ее называют эмпирической.

Эмпирической функцией распределения (функцией распределения выборки) называется функция $F^*(x)$, определяющая для каждого значения x относительную частоту события $X < x_i$:

$$F^*(x) = \frac{n_x}{n} = \frac{\sum\limits_{x_i < x} n_i}{n}. \qquad (3.8)$$

В отличие от *эмпирической* функции распределения $F^*(x)$ выборки, функция распределения $F(x)$ генеральной совокупности называется *теоретической функцией распределения* (см. разд. 2.1). Различие между этими функциями состоит в том, что теоретическая функция $F(x)$ определяется *вероятностью события* $X < x_i$, а эмпирическая функция $F^*(x)$ определяется *относительной частотой* этого события.

При больших значениях n функции $F^*(x)$ и $F(x)$ мало отличаются друг от друга, т. е. $\lim\limits_{n\to\infty} P\left[\left|F(x) - F^*(x)\right| < \varepsilon\right] = 1$ $(\varepsilon > 0)$. Поэтому целесообразно использовать эмпирическую функцию распределения $F^*(x)$ выборки для приближенного представления теоретической (интегральной) функции распределения $F(x)$ генеральной совокупности.

Такое заключение подтверждается и тем, что эмпирическая функция распределения $F^*(x)$ обладает всеми свойствами функции распределения $F(x)$ (см. разд. 2.1.3).

Свойство 1. Значение функции распределения принадлежат отрезку [0; 1]:

$$F^*(x_{\min}) = 0,\ F^*(x_{\max}) = 1. \qquad (3.9)$$

Свойство 2. Функции распределения – неубывающая функция:

$$F^*(x_1) \le F^*(x_2) \text{ при } x_1 < x_2. \qquad (3.10)$$

Свойство 3. Вероятность попадания случайной величины в заданный интервал равна приращению функции на этом интервале:

$$P(a < X < b) = F^*(b) - F^*(a). \qquad (3.11)$$

Итак, эмпирическая функция распределения выборки $F^*(x)$ служит для оценки теоретической функции распределении генеральной совокупности $F(x)$.

Пример 1. В табл. 3.8 представлен дискретный вариационный ряд.

Таблица 3.8

Распределение выборки

Варианта x_i	2	6	10
Частота n_i	12	18	30

Построить эмпирическую функцию по данной выборке.

Решение

Объем выборки определяется по формуле (3.1):

$$n = \sum n_i = 12 + 18 + 30 = 60.$$

Наименьшая варианта $x_{\min} = 2$, следовательно, согласно свойству 1 (формула 3.9), эмпирическая функция равна:

$$F^*(x) = 0 \text{ при } x \le 2.$$

Значение СВ $X < 6$, а именно $x_1 = 2$, наблюдалось $n_1 = 12$ раз, следовательно, по формуле (3.8)

$$F^*(x) = \frac{12}{60} = 0{,}2 \text{ при } 2 < x \le 6.$$

Аналогично значения СВ $X < 10$, а именно $x_1 = 2$ и $x_2 = 6$, наблюдались $n = 12 + 18 = 30$ раз, следовательно

$$F^*(x) = \frac{30}{60} = 0{,}5 \text{ при } 6 < x \le 10.$$

Так как $x_{\max} = 10$ – наибольшая варианта, то, согласно свойству 1 (формула 3.9), эмпирическая функция равна:

$$F^*(x) = 1 \text{ при } x > 10.$$

Искомая эмпирическая функция примет вид

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \le 2; \\ 0{,}2 & \text{при} \quad 2 < x \le 6; \\ 0{,}5 & \text{при} \quad 6 < x \le 10; \\ 1 & \text{при} \quad x > 10. \end{cases}$$

Также данные об эмпирической функции распределения можно представить в виде табл. 3.9.

Таблица 3.9

Распределение выборки

x_i		2	6	10
n_i		12	18	30
$F^*(x)$	0	0,2	0,5	1

График эмпирической функции распределения $F^*(x)$ изображен на рис. 3.7.

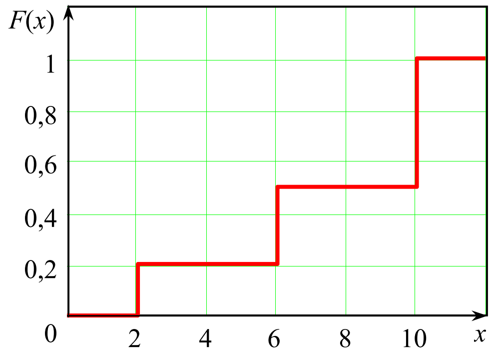


Рис. 3.7. Эмпирическая функция распределения

Ответ: эмпирическая функция распределения представлена в табл. 3.9 и на рис. 3.7.

3.7. Полигон и гистограмма

Графическое изображение выборки облегчает ее анализ и позволяет судить о форме распределения генеральной совокупности, которую представляет выборка.

Для наглядности в математической статистике выборку изображают в виде полигона или гистограммы, а также в виде графика эмпирической функции распределения (разд. 3.6).

Полигон – это ломаная линия, получаемая при соединении точек, абсциссы которых равны вариантам, а ординаты – либо абсолютным частотам, либо относительным.

Полигон частот соединяет отрезками точки $(x_1; n_1)$, $(x_2; n_2)$, …, $(x_i; n_i)$ (рис. 3.8, *а*).

Полигон относительных частот соединяет отрезками точки $(x_1;\ W_1)$, $(x_2;\ W_2), \ldots, (x_i;\ W_i)$ (рис. 3.8, *б*).

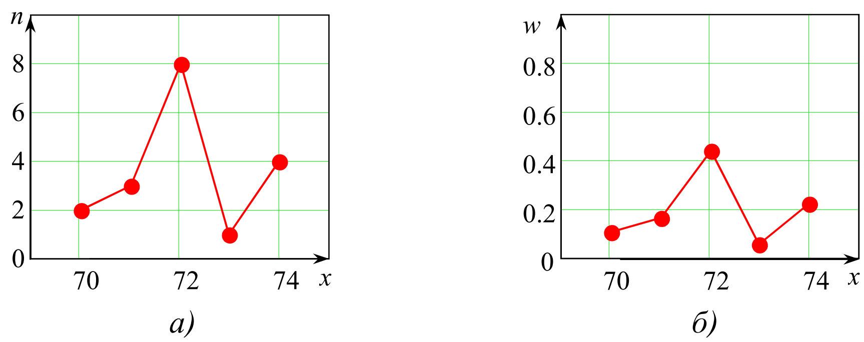


Рис. 3.8. Полигон частот и полигон относительных частот

Полигон удобно строить для дискретного вариационного ряда распределения.

Пример 1. В табл. 3.10 дано соотношение вариант и относительных частот для выборки.

Таблица 3.10

Данные выборки

x_i	1,5	3,5	5,5	7,5
W_i	0,1	0,2	0,4	0,3

Построить полигон относительных частот.

Решение

При построении полигона относительных частот по оси абсцисс откладываем значения случайной величины X, а по оси ординат – относительные частоты W. Точки с координатами $(x_i;\ W_i)$ соединяются. Результат показан на рис. 3.9.

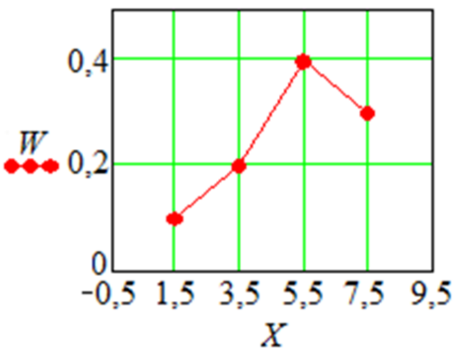


Рис. 3.9. Полигон относительных частот

Ответ: полигон относительных частот показан на рис. 3.9.

В случае *непрерывного признака* целесообразно строить гистограмму. *Гистограмма* – это графическое представление распределения частот для количественного признака, образуемое соприкасающимися прямоугольниками, основаниями которых служат частичные интервалы $(a_i - a_{i+1})$, а площади пропорциональны частотам. Как правило, при построении гистограммы за основу берется интервальный вариационный ряд.

Для построения гистограммы частот на оси абсцисс откладываются частичные интервалы $a_i - a_{i+1}$, а над ними проводят параллельно оси абсцисс отрезки на расстоянии n_i / h (плотность частоты) (рис. 3.10, *а*).

Площадь i-го частичного прямоугольника равна $h \cdot n_i / h = n_i$ – сумме частот вариант i-го интервала. Следовательно, *площадь гистограммы частот* равна *сумме всех частот* $\sum n_i = n$, т. е. объему выборки n.

Гистограмма относительных частот – это ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы $a_i - a_{i+1}$ длиною h, а высоты равны отношению W_i / h (плотность относительной частоты).

Для построения гистограммы относительных частот по оси абсцисс откладывают частичные интервалы $a_i - a_{i+1}$, а над ними проводят отрезки параллельно оси абсцисс на расстоянии W_i / h (плотность относительной частоты) (рис. 3.10, *б*).

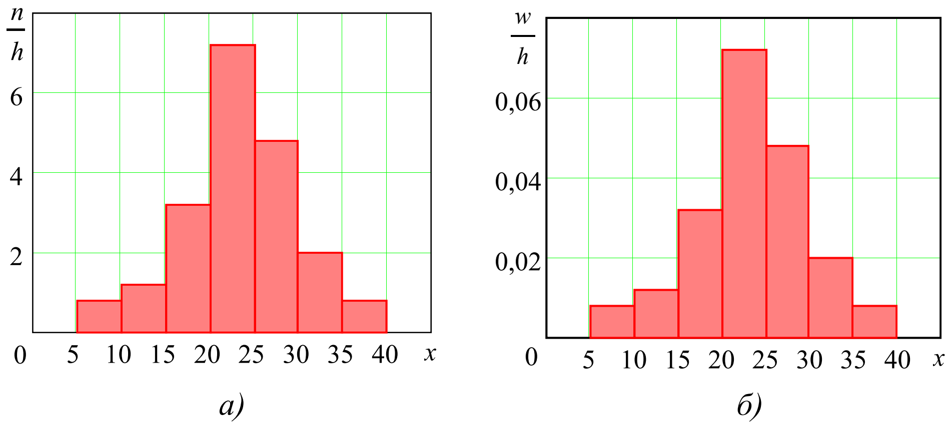


Рис. 3.10. Гистограмма частот и гистограмма относительных частот

Площадь i-го частичного прямоугольника равна $h \cdot W_i / h = W_i$ – сумме относительных частот вариант, попавших в i-й интервал. Следовательно, площадь гистограммы относительных частот равна сумме всех относительных частот. Согласно равенству (3.1) площадь равна единице.

Пример 2. Данные выборки объемом n = 100 представлены в табл. 3.11 в виде интервального ряда распределения.

Таблица 3.11

Данные для решения задачи

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант n_i
1	1…5	10
2	5…9	20
3	9…13	50
4	13…17	12
5	17…21	8

Построить гистограмму частот.

Решение

Для построения гистограммы частот по оси абсцисс откладываются отрезки, равные длине интервалов вариационного ряда h. Все интервалы данного ряда одинаковые: $h = 4$.

По оси ординат откладываются расстояния, равные соответствующим плотностям частоты n_i / h:

$$\frac{n_1}{h}=\frac{10}{4}=2{,}5;\ \frac{n_2}{h}=\frac{20}{4}=5;\ \frac{n_3}{h}=\frac{50}{4}=12{,}5;\ \frac{n_4}{h}=\frac{12}{4}=3;\ \frac{n_5}{h}=\frac{8}{4}=2.$$

Первый столбец имеет длину h = 4 и высоту $\frac{n_1}{h}=2{,}5$, остальные строятся аналогично. Гистограмма частот по данным выборки представлена на рис. 3.11.

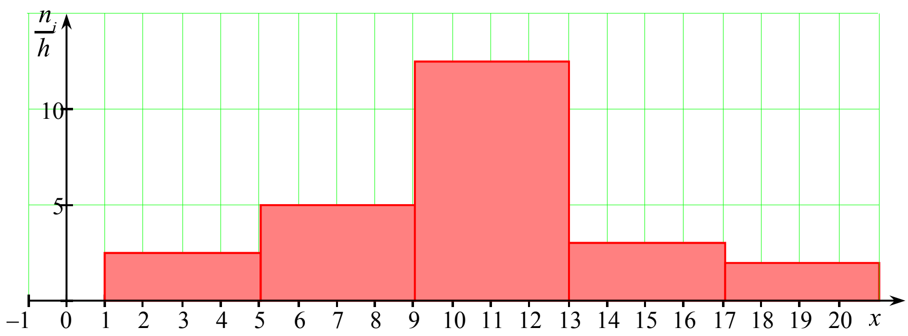


Рис. 3.11. Гистограмма частот

Ответ: гистограмма частот показана на рис. 3.11.

Очевидно, что гистограмма легко может быть преобразована в полигон распределения, если середины верхних сторон прямоугольников соединить отрезками прямых (рис. 3.12). И наоборот, полигон легко превратить в гистограмму, если принять узлы полигона за середины верхних сторон прямоугольников, из которых состоит гистограмма.

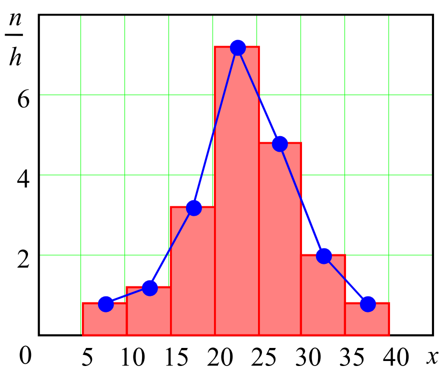


Рис. 3.12. Гистограмма и полигон частот

Когда ширина всех интервалов группировки *одинакова*, внешний вид гистограммы не изменится, если по оси ординат откладывать не величины n_i / h, а *частоты* интервалов n_i (*относительные частоты* интервалов W_i). Поэтому, если по графику гистограммы не проводится измерение, а требуется лишь оценить внешний вид распределения, то по оси y можно откладывать значения частот (или относительных частот), а не их плотности.

В случае неравенства интервалов между собой гистограмма строится по плотностям распределения.

Внешний вид гистограммы или полигона, сделанный по выборке, может позволить сделать предположение о законе распределения, которому подчиняется генеральная совокупность.

Пример 3. На рис. 3.13, *а* показаны гистограмма и предполагаемая плотность распределения, характерная для нормального закона распределения. По гистограмме на рис. 3.13, *б* можно сделать вывод, что выборка подчиняется экспоненциальному закону распределения.

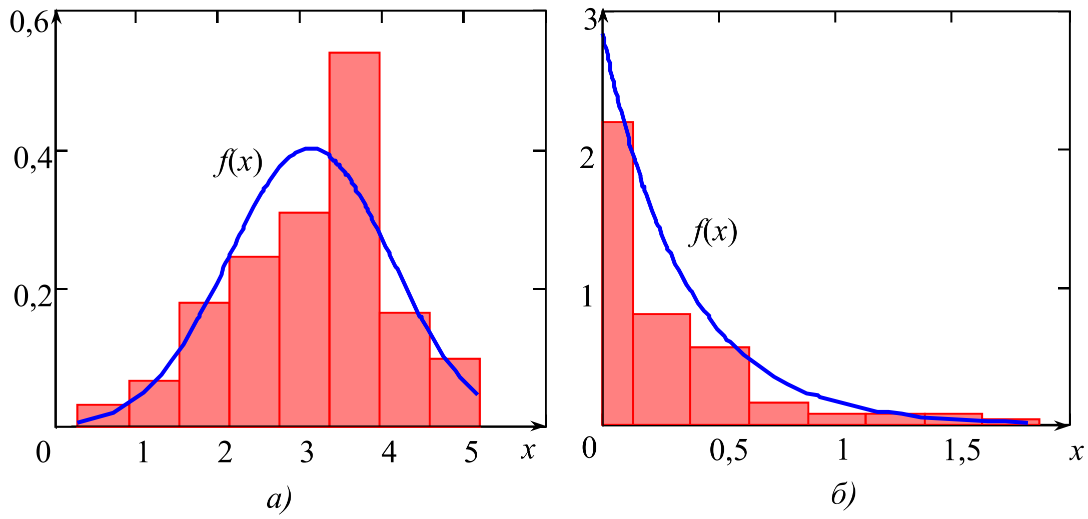


Рис. 3.13. Гистограммы и плотности распределений

Пример 4. Распределение населения России по возрастным группам по данным за 1997 г. показано в табл. 3.12.

Таблица 3.12

Численность населения России в 1997 г.

Все население	В том числе в возрасте								
	до 10	10…20	20…30	30…40	40…50	50…60	60…70	70 и старше	Всего
Численность населения, %	12,1	15,7	13,6	16,1	15,3	10,1	9,8	7,3	100,0

Построить гистограмму частот.

Решение

Для построения гистограммы на оси абсцисс указывают значения границ интервалов и на их основании строят прямоугольники, высота которых пропорциональна плотности частот $\frac{n_i}{h}$.

В этой задаче ширина всех интервалов группировки одинакова, вид гистограммы не изменится, если по оси ординат откладывать не величины $\frac{n_i}{h}$, а частоты интервалов n_i.

На рис. 3.14 показана гистограмма частот для данных табл. 3.12.

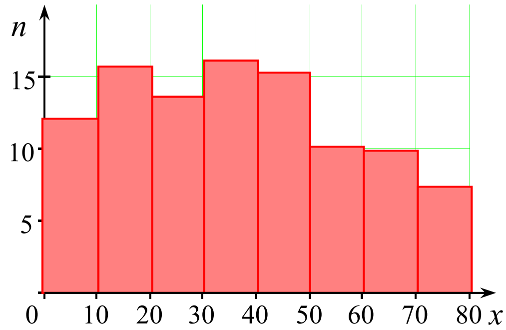


Рис. 3.14. Распределение населения России по возрастным группам

Форма гистограммы частот (рис. 3.14) говорит о равномерном распределении населения России по возрастным группам.

Ответ: гистограмма частот распределения населения России по возрастным группам показана на рис. 3.14.

Пример 5. Приведено распределение 30 работников фирмы по размеру месячной заработной платы (табл. 3.13).

Таблица 3.13

Распределение работников по уровню ежемесячного дохода

№	Размер з/п, тыс. руб/месяц	Численность работников, чел.
1	до 16	4
2	16–20	12
3	20–30	8
4	30–50	6
	Итого	**30**

Изобразить интервальный вариационный ряд графически в виде гистограммы.

Решение

В условии задачи неизвестна граница первого (открытого) интервала. Определить ее можно либо от начала отсчета (0), либо взяв более узкий интервал на основе имеющейся выборки. Для упрощения задачи и из-за отсутствия исходной выборки первый интервал принимается от 0 до 16 тыс. руб., т. е. $h_1 = 16$.

Длина остальных интервалов определяется из условия задачи: $h_2 = 4$; $h_3 = 10$; $h_4 = 20$.

Для построения гистограммы по оси абсцисс откладываем отрезки, длина которых соответствует интервалам вариационного ряда. При использовании различных программ автоматизированного построения графиков следует задавать середины частичных интервалов (обозначены на рис. 3.15 синими точками):

$$a_1 = 8;\ a_2 = 18;\ a_3 = 25;\ a_4 = 40.$$

Так как длина интервалов различна, то для сохранения общей площади гистограммы высота образуемых прямоугольников берется равной плотности частот $\frac{n_i}{h_i}$:

$$\frac{n_1}{h_1} = \frac{4}{16} = 0{,}25;\ \frac{n_2}{h_2} = \frac{12}{4} = 3;\ \frac{n_3}{h_3} = \frac{8}{10} = 0{,}8;\ \frac{n_4}{h_4} = \frac{6}{20} = 0{,}3.$$

Гистограмма частот изображена на рис. 3.15.

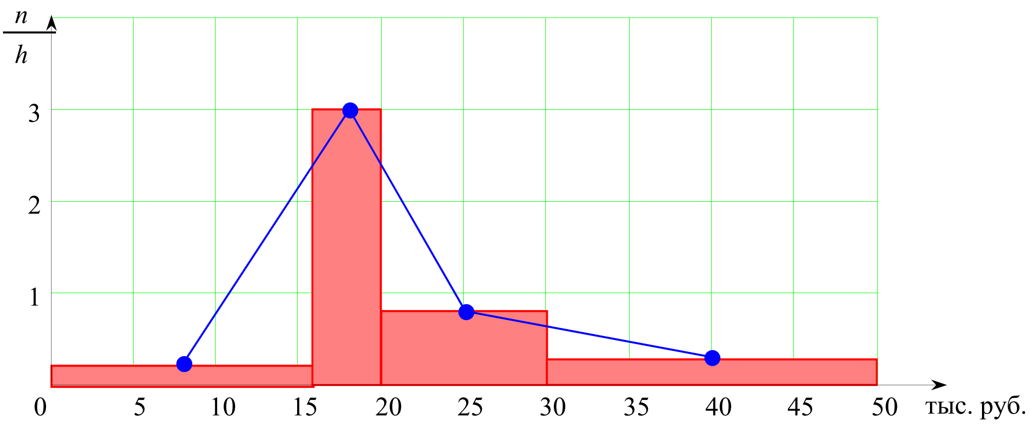


Рис. 3.15. Распределение работников по уровню ежемесячного дохода

Ответ: гистограмма частот распределения работников фирмы по размеру месячной заработной платы показана на рис. 3.15.

Задачи для самостоятельного решения

Задача 1. Пользуясь формулой Стерджеса, определите интервалы групп, полученных в результате группировки работников магазина по среднемесячной выработке, если общая численность работников составляет 22 человека, а минимальная и максимальная среднемесячная выработка соответственно равны 100 тыс. руб. и 250 тыс. руб.

Задача 2. По имеющимся данным о числе товарных секций по двадцати магазинам города (табл. 3.14) построить ряд распределения.

Таблица 3.14

Распределение товарных секций по магазинам города

2	4	3	5	6
5	6	4	6	3
2	2	4	3	3
4	5	5	4	4

Изобразить ряд распределения графически в виде полигона относительных частот.

Задача 3. Дана выборка о размере прибыли (млн руб.) двадцати коммерческих банков (табл. 3.15).

Таблица 3.15

Выборка о размере прибыли коммерческих банков

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

Составить ряд распределения и изобразить его графически в виде гистограммы и эмпирической функции распределения.

Задача 4. В табл. 3.16 представлен дискретный вариационный ряд.

Таблица 3.16

Данные распределения выборки

x_i	1	2	3	4
n_i	5	10	15	20

Построить эмпирическую функцию по данной выборке.

Задача 5. В табл. 3.17 представлен дискретный вариационный ряд.

Таблица 3.17

Данные распределения выборки

x_i	5	7	10	15
n_i	2	3	8	7

Построить эмпирическую функцию по данной выборке.

Задача 6. В табл. 3.18 представлен дискретный вариационный ряд.

Таблица 3.18

Данные распределения выборки

x_i	1	3	5	7	9
n_i	10	15	30	33	12

Построить полигоны частот и относительных частот распределения по данной выборке.

Задача 7. Данные выборки в виде интервального вариационного ряда представлены в табл. 3.19.

Таблица 3.19

Интервальный вариационный ряд

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант n_i
1	2…5	9
2	5…8	10
3	8…11	25
4	11…14	6

Построить гистограмму частот и относительных частот.

Задача 8. При проверке на надежность партии измерительных приборов были проведены испытания выборки из n = 50 приборов в течение T = 10000 часов и зарегистрировано в течение этого времени каждого из приборов следующее число отказов: 0, 3, 1, 0, 1, 1, 0, 0, 2, 0, 0, 0, 1, 1, 4, 0, 2, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 3, 2, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 2, 0, 0.

Составить дискретный ряд распределения числа отказов.

Построить полигон и гистограмму относительных частот.

Задача 9. В табл. 3.20 представлена выборка объемом n = 100 в виде интервального вариационного ряда.

Таблица 3.20

Интервальный вариационный ряд

Частичный интервал длиной h = 5	Сумма частот вариант частичного интервала n_i	Плотность частоты $\frac{n_i}{h}$	Сумма относительных частот $W_i = \frac{n_i}{n}$	Плотность относительных частот $\frac{W_i}{h}$
5–10	4			
10–15	6			
15–20	16			
20–25	36			
25–30	24			
30–35	10			
35–40	4			

Построить гистограмму частот и гистограмму относительных частот. Пустые колонки табл. 3.20 заполнить по мере решения задачи.

Контрольные вопросы и задания

1. Дайте определение термину «математическая статистика», перечислите её задачи.
2. Дайте определение термину «генеральная совокупность». В чем его отличие от термина «выборочная совокупность»?
3. Что такое «объем выборки»?
4. Перечислите признаки достоверности выборки.
5. Для чего необходимо обеспечить достаточный объём выборки и её репрезентативность?
6. В каком случае выборка будет являться репрезентативной?
7. Назовите способы отбора элементов в выборку. Приведите примеры реализации названных способов отбора.
8. Какая выборка называется повторной и бесповторной?
9. Дайте определение термину «вариационный ряд».
10. Для чего применяется формула Стерджеса? Приведите формулу Стерджеса с расшифровкой ее составляющих.
11. Какая функция является эмпирической?
12. Дайте определение терминам «полигон» и «гистограмма».
13. Что такое гистограмма относительных частот? В чем ее отличие от гистограммы частот?

Глава 4. ОСНОВЫ ТЕОРИИ ОЦЕНИВАНИЯ

4.1. Оценки параметров распределения

Обычно в распоряжении исследователя имеются лишь данные выборки, например значения количественного признака $x_1, x_2, \ldots, x_n$, полученные в результате n наблюдений. Используя эти данные, можно получить параметр $\tilde{\Theta}$. Значение параметра $\tilde{\Theta}$ позволяет сделать выводы о свойствах некоторой генеральной совокупности, из которой извлечена выборка.

Поскольку истинное значение искомой величины получить принципиально невозможно (из-за конечности выборки), то в качестве него принимается найденное по выборке значение Θ. Полученное значение $\tilde{\Theta}$ называется *оценкой истинного значения* параметра Θ.

Оценка $\tilde{\Theta}$ является функцией результатов наблюдений объема n:

$$\tilde{\Theta} = f(x_1, x_2, ..., x_n), \tag{4.1}$$

где $x_1, x_2, \ldots, x_n$ – независимые наблюдаемые значения случайных величин. Следовательно, найти статистическую оценку неизвестного параметра – значит найти функцию от наблюдаемых случайных величин (формула (4.1)), которая дает приближенное значение $\tilde{\Theta}$ оцениваемого параметра Θ.

Примечание. Оценка – это правило вычисления оцениваемого параметра, а термин «оценить» означает указать приближенное значение искомого параметра.

При оценивании определяются количественные характеристики случайных величин на основе имеющейся конечной выборки. В дальнейшем полученную информацию используют для описания свойств генеральной совокупности, поэтому к качеству полученных оценок предъявляют высокие требования.

При получении оценок $\tilde{\Theta}$ следует стремиться к тому, чтобы они были максимально близки к истинному значению искомой величины Θ.

Это достигается удовлетворением следующих требований.

1. *Несмещенность.*

Оценка называется *несмещенной*, если при любом объеме выборки n математическое ожидание оценки $\tilde{\Theta}$ равно истинному значению искомого параметра $\tilde{\Theta}$:

$$M\left[\tilde{\Theta}\right] = \Theta. \tag{4.2}$$

Во многих случаях увеличение объема выборки приводит к уменьшению смещения оценки.

2. *Эффективность.*

Оценка называется *эффективной*, если она обладает минимальной дисперсией:

$$D\left[\tilde{\Theta}\right] \to \min.$$

3. *Состоятельность.*

Оценка называется *состоятельной*, если при увеличении объема выборки ($n \to \infty$) значение оценки $\tilde{\Theta}$ стремится к оцениваемому параметру Θ. Это означает, что при достаточно большом объеме выборки оценка $\tilde{\Theta}$ практически совпадает с истинным значением Θ.

Если дисперсия оценки при $n \to \infty$ стремится к нулю, то такая оценка будет состоятельной:

$$D\left[\tilde{\Theta}\right] \to 0.$$

4. *Устойчивость.*

Устойчивость оценки характеризует ее чувствительность к аномальным результатам (промахам и выбросам).

4.2. Точечная оценка параметров распределения

Точечная оценка – это статистическая оценка, которая определяется одним числом. Это число в дальнейших вычислениях принимается за истинное значение искомого параметра.

Точечная оценка, как функция от выборки, является случайной величиной и меняется от выборки к выборке.

Точечную оценку целесообразно определять в тех случаях, когда объем экспериментальных данных (ЭД) достаточно велик. О достаточности объема выборки следует судить в каждом конкретном случае. Малый объем ЭД может значительно исказить оценку, что делает их непригодными для использования.

4.2.1. Точечная оценка математического ожидания

Математическое ожидание – одна из основных числовых характеристик, которая характеризует среднее значение совокупности данных (разд. 2.2.1). Ее применяют для обработки и представления результатов различных исследований. Для генеральной совокупности мат. ожиданием является *генеральное среднее* $\overline{x}_{\text{г}}$ (индекс «г» означает принадлежность к генеральной совокупности).

Генеральное среднее значение $\overline{x}_{\text{г}}$ (мат. ожидание генеральной совокупности) – среднее арифметическое значений искомого признака генеральной совокупности.

Если все значения $x_1, x_2, ..., x_N$ генеральной совокупности X объема N различны, то

$$\overline{x}_г = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{\sum x_N}{N}. \qquad (4.3)$$

Если же значения $x_1, x_2, ..., x_N$ имеют соответственно частоты $N_1, N_2, ..., N_i$, то

$$\overline{x}_г = \frac{x_1 N_1 + x_2 N_2 + ... + x_i N_i}{N} = \frac{\sum x_i N_i}{N},$$

т. е. генеральная средняя – это среднее взвешенное значение искомого признака с весами, равными соответствующим частотам.

Итак, если рассматривать обследуемый признак X генеральной совокупности как случайную величину, то математическое ожидание искомого признака равно генеральной средней этого признака $M(X) = \overline{x}_г$.

Если в распоряжении исследователя есть только выборка из генеральной совокупности, то путем вычислений можно получить математическое ожидание *только* для выборки. Мат. ожидание выборки – это *выборочное средней* $\overline{x}_в$ (индекс *«в»* означает принадлежность к выборке).

Выборочное среднее значение $\overline{x}_в$ – это среднее арифметическое значение, найденное для выборки.

Если все значения $x_1, x_2, ..., x_n$ выборки объема n различны, то

$$\overline{x}_в = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum x_n}{n}, \qquad (4.4)$$

иначе с учетом частоты повторения значений выборки $x_1, x_2, ..., x_i$ выборочное среднее равно:

$$\overline{x}_в = \frac{x_1 n_1 + x_2 n_2 + ... + x_i n_i}{n} = \frac{\sum x_i \cdot n_i}{n}. \qquad (4.5)$$

Выборочное среднее значение $\overline{x}_в$ – это среднее взвешенное значение признака с весами, равными соответствующим частотам.

Примечание. Обратите внимание, что формула (4.5) практически не отличается от формулы (2.11). Разница состоит лишь в том, что в формуле (2.11) используется значение вероятности p (найденное до опыта), а в формуле (4.5) используется относительная частота $W_i = \frac{n_i}{n}$ (формула (3.2)) (полученная в результате эксперимента).

Как правило, генеральная совокупность исследователю недоступна, следовательно, генеральное среднее $\overline{x}_г$ неизвестно. Получить сведения о генеральной совокупности, в частности о ее среднем значении, можно только в результате исследования некоторой выборки объема n.

Проведены измерения некоторого количественного признака *X* какого-либо объекта *n* раз. Это значит, что произведена повторная выборка объема *n* со значениями признака $x_1, x_2, ..., x_n$ из некоторой генеральной совокупности. Выборочное среднее значение $\overline{x}_в$ – это оценка значения этого признака, которая в дальнейшем используется для характеристики генерального среднего значения $\overline{x}_г$, т. е. принимается в качестве измеренного значения для характеристики количественного признака *X* исследуемого объекта.

В математической статистике [5] доказано, что математическое ожидание этой оценки $\overline{x}_в$ равно $\overline{x}_г$: $M(\overline{x}_в) = \overline{x}_г$, следовательно, $\overline{x}_в$ – *несмещенная* оценка.

Кроме того, доказано [5], что при увеличении объема выборки *n* выборочное среднее значение $\overline{x}_в$ стремится по вероятности к генеральной средней $\overline{x}_г$, а это и означает, что выборочная средняя является еще и *состоятельной* оценкой генерального среднего $\overline{x}_г$.

Это означает, что если по нескольким выборкам достаточно большого объема из одной и той же генеральной совокупности будут найдены выборочные средние, то они будут приближенно равны между собой. Этот факт доказывает еще и *устойчивость* полученной оценки.

4.2.2. Точечная оценка дисперсии

Еще одной важной числовой характеристикой является дисперсия, которая характеризует разброс относительного среднего значения (мат. ожидания) (разд. 2.2.2). Оценка дисперсии используется для характеристики погрешности определения среднего значения, полученного в результате исследований.

Генеральная дисперсия $D_г$ (дисперсия генеральной совокупности) – это среднее арифметическое квадратов отклонений значений генеральной совокупности от их среднего значения $\overline{x}_г$.

$$D_г = \frac{\sum(x_i - \overline{x}_г)^2}{N}. \tag{4.6}$$

Вместе с дисперсией для характеристики рассеяния значений признака генеральной совокупности вокруг своего среднего значения используется сводная характеристика – среднее квадратическое отклонение (СКО).

Генеральное среднее квадратическое отклонение – это квадратный корень из генеральной дисперсии:

$$\sigma_г = \sqrt{D_г}. \tag{4.7}$$

Выборочная дисперсия характеризует рассеянность наблюдаемых значений выборки вокруг своего среднего значения $\overline{x}_в$, найденного по выборке.

Аналогично *выборочная дисперсия* $D_{\text{в}}$ – это среднее арифметическое квадратов отклонений наблюдаемых значений искомого признака от их среднего значения $\overline{x}_{\text{в}}$. Если все значения выборки $x_1, x_2, ..., x_n$ различны, то

$$D_{\text{в}} = \frac{\sum (x_i - \overline{x}_{\text{в}})^2}{n}, \tag{4.8}$$

иначе учитывается количество повторений каждого из значений выборки:

$$D_{\text{в}} = \frac{\sum (x_i - \overline{x}_{\text{в}})^2 \cdot n_i}{n}. \tag{4.9}$$

Аналогично *выборочное среднее квадратическое отклонение* – это квадратный корень из выборочной дисперсии:

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}. \tag{4.10}$$

В математической статистике [5] доказано, что выборочная дисперсия $D_{\text{в}}$ является смещенной оценкой генеральной дисперсии $D_{\text{г}}$. Это означает, что мат. ожидание выборочной дисперсии не равно значению оцениваемой генеральной дисперсии, а равно

$$M\left[D_{\text{в}}\right] = \frac{n-1}{n} D_{\text{г}}. \tag{4.11}$$

Использование выборочной дисперсии $D_{\text{в}}$ будет приводить к систематическим ошибкам, давая заниженное значение генеральной дисперсии $D_{\text{г}}$. Для устранения этого недостатка легко «исправить» выборочную дисперсию так, чтоб ее математическое ожидание было равно генеральной дисперсии.

Исправленную дисперсию S^2 можно получить путем умножения $D_{\text{в}}$ на дробь $\frac{n}{n-1}$:

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{n}{n-1} \frac{\sum (x_i - \overline{x}_{\text{в}})^2}{n} = \frac{\sum (x_i - \overline{x}_{\text{в}})^2}{n-1}. \tag{4.12}$$

Исправленная дисперсия является *несмещенной* оценкой генеральной дисперсии [5], т. к. $M\left[S^2\right] = D_{\text{г}}$.

Для оценки среднего квадратического отклонения генеральной совокупности $\sigma_{\text{г}}$ используется *«исправленное» среднее квадратическое отклонение S*, которое равно квадратному корню из исправленной дисперсии:

$$S = \sqrt{\frac{\sum (x_i - \overline{x}_{\text{в}})^2}{n-1}} \tag{4.13}$$

или, с учетом повторения значений x в выборке,

$$S = \sqrt{\frac{\sum (x_i - \overline{x}_в)^2 \cdot n_i}{n-1}} \,. \tag{4.14}$$

Следует почеркнуть, что *«исправленное» среднее квадратическое отклонение* S не является *несмещенной* оценкой.

Формулы (4.8) и (4.13) отличаются лишь знаменателями. Очевидно, что при увеличении n *выборочная* и *исправленная* дисперсии различаются незначительно. На практике *исправленную* дисперсию следует использовать, только если $n < 30$. В остальных случаях можно использовать любую дисперсию, на итоговом результате это практически не отразится.

Так как в большинстве случаев в распоряжении исследователя есть только выборка из генеральной совокупности, то целесообразнее в дальнейшем *выборочное средней значение* $\overline{x}_в$ обозначать $\overline{x}$.

Также следует отметить важность обозначения единиц измерения для числовых значений, используемых для определения различных оценок.

Пример 1. В результате пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие данные (в мм): 92, 94, 103, 105, 106. Найти выборочную среднюю длину стержня, выборочную и исправленную дисперсии, характеризующие ошибку прибора.

Решение

Выборочное среднее значение рассчитывается по формуле (4.4):

$$\overline{x}_в = \overline{x} = \frac{92+94+103+105+106}{5} = 100 \text{ мм.}$$

Выборочная дисперсия рассчитывается по формуле (4.8):

$$D_в = \frac{(92-100)^2 + (94-100)^2 + (103-100)^2 + (105-100)^2 + (106-100)^2}{5},$$

$$D_в = 34 \text{ мм}^2.$$

Исправленную выборочную дисперсию можно вычислить напрямую (формула (4.13)) или используя связь между *выборочной* и *исправленной выборочной* дисперсиями (формула (4.12)):

$$S^2 = \sqrt{\frac{(92-100)^2 + (94-100)^2 + (103-100)^2 + (105-100)^2 + (106-100)^2}{5-1}} =$$

$$= 42{,}5 \text{ мм}^2,$$

или

$$S^2 = \frac{n}{n-1} D_в = \frac{5}{5-1} \cdot 34 = 42{,}5 \text{ мм}^2.$$

Обратите внимание, что выборочная дисперсия – это заниженное значение генеральной дисперсии, что приводит к меньшему, чем есть на самом деле, значению, характеризующему ошибку прибора.

Ответ: выборочная средняя длина стержня $\overline{x}$ = 100 мм, выборочная дисперсия $D_в$ = 34 мм^2, исправленная дисперсия S^2 = 42,5 мм^2.

Примечание. Так как *выборочная* дисперсия является смещённой оценкой, ее нельзя использовать для дальнейших вычислений, а вместо нее используют *исправленную выборочную* дисперсию. Поэтому в дальнейшем используемую *исправленную выборочную* дисперсию будем называть «*дисперсия*» (без пояснений).

Задачи для самостоятельного решения

Задача 1. В табл. 4.1 приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Таблица 4.1

Рост студентов (в см)

Рост, см	154–158	158–162	162–166	166–170	170–174	174–178	178–182
Число студентов	10	14	26	28	12	8	2

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

Указание. В качестве вариант принять середины интервалов.

Задача 2. Из генеральной совокупности извлечена выборка объема n = 50 (табл. 4.2).

Таблица 4.2

Данные выборки

Варианта x_i	2	5	7	10
Частота n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

Задача 3. В итоге четырех измерений некоторой величины одним прибором (без систематических ошибок) получены следующие результаты: 8, 9, 11, 12.

Найти выборочное среднее значение результатов измерений; выборочную и исправленную дисперсии ошибок прибора.

Задача 4. Из 1500 деталей отобрано 250, распределение которых по размеру приведено в табл. 4.3.

Таблица 4.3

Распределение деталей по размеру

Размер детали, мм	7,8…8,0	8,0…8,2	8,2…8,4	8,4…8,6	8,6…8,8	8,8…9,0
Количество деталей, n_i	5	20	80	95	40	10

Найти точечные оценки среднего значения размера деталей $\overline{x}$, выборочной и исправленной дисперсии D и S^2.

Задача 5. Вычислить несмещенные оценки среднего значения $\overline{x}$, дисперсии S^2 и стандартного отклонения S генеральной совокупности по двум выборкам: (43, 51, 44, 47, 34) и (52, 42, 40, 38, 37).

4.3. Методы получения точечных оценок

Количественные значения оценок искомых величин в общем случае являются функциями элементов выборки.

Правила, по которым проводится оценивание искомых параметров, называются *методами получения оценок*. Наиболее широкое распространение получили следующие методы: максимального правдоподобия, моментов, порядковых характеристик и др.

На практике при оценке параметров не всегда удается удовлетворить одновременно всем требованиям: состоятельность, несмещенность и эффективность. Однако выбору оценки всегда должно предшествовать критическое рассмотрение ее со всех точек зрения.

4.3.1. Метод максимального правдоподобия

Метод максимального правдоподобия, предложенный К.Ф. Гауссом (рис. 1.6) в 1848 г. и обобщенный Р.А. Фишером (рис. 3.1), с теоретической точки зрения является наиболее разработанным методом получения оценок. Этот метод чаще всего приводит к наилучшим оценкам, удовлетворяющим одновременно требование состоятельности, несмещенности и эффективности.

Идея метода основана на априорном (доопытном) распределении выборки:

- в результате наблюдения имеется выборка $x = x_1, x_2, ..., x_n$;
- вид плотности распределения данной случайной величины известен, но неизвестен его параметр Θ (или несколько параметров).

Метод максимального правдоподобия заключается в том, что в качестве оценки параметра Θ принимается то значение $\tilde{\Theta}$, при котором *функция правдоподобия* L достигает максимального значения.

Функция правдоподобия – это обобщенная функция, равная произведению плотностей распределения при каждом результате наблюдения:

$$L(\tilde{\Theta},x) = L(\tilde{\Theta}, x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(\tilde{\Theta}, x_1, x_2, ..., x_n), \quad (4.15)$$

где $x_1, x_2, \ldots, x_n$ – известные значения выборки (результаты наблюдения); $\tilde{\Theta}$ – искомый параметр распределения (Если параметров несколько, то вместо $\tilde{\Theta}$ записывает требуемое количество параметров: $\tilde{\Theta}_1, \tilde{\Theta}_2, ..., \tilde{\Theta}_i$).

Для определения максимального значения *функцию правдоподобия* L необходимо исследовать на экстремум (при экстремуме $\frac{\partial L}{\partial \tilde{\Theta}} = 0$). С целью упрощения вычислительной процедуры обычно максимизируют не саму функцию правдоподобия, а ее логарифм ($\ln L(\tilde{\Theta}, x)$, так как в этом случае проще найти частные производные.

Нахождение оценок сводится к решению системы уравнений правдоподобия относительно искомых параметров:

$$\frac{d \ln L(x, \tilde{\Theta}_i)}{d\tilde{\Theta}_i} = 0, \quad (4.16)$$

где $i = 1, 2, \ldots$ – число оцениваемых параметров.

Пример 1. Оценить методом максимального правдоподобия неизвестные параметры m и σ^2 нормального закона распределения.

Решение

Плотность распределения для нормального закона распределения имеет вид (формула (2.28))

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right),$$

где m и σ^2 – неизвестные параметры распределения.

В соответствии с формулой (4.15) функция правдоподобия примет вид

$$L\left(m, \sigma^2, x\right) = \prod_{i=1}^{n} f(x_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \cdot e^{-\frac{\sum_{i=1}^{n}(x_i - m)^2}{2\sigma^2}}. \quad (4.17)$$

После преобразования формулы (4.17) логарифм функции правдоподобия примет вид

$$\ln L\left(m,\sigma^2,x\right)=\ln\left(\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\cdot e^{-\frac{\sum_{i=1}^{n}(x_i-m)^2}{2\sigma^2}}\right)=-\frac{n}{2}\cdot\ln\left(2\pi\sigma^2\right)-\frac{\sum_{i=1}^{n}\left(x_i-m\right)^2}{2\sigma^2}.$$

Так как у нормального распределения два параметра, необходимо решить систему из двух уравнений:

$$\begin{cases}\dfrac{\partial\ln L\left(m,\sigma^2,x\right)}{\partial m}=\dfrac{\sum_{i=1}^{n}\left(x_i-m\right)}{\tilde{\sigma}^2}=0;\\[2ex] \dfrac{\partial\ln L\left(m,\sigma^2,x\right)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\cdot\sum_{i=1}^{n}\left(x_i-m\right)^2=0.\end{cases}$$

После преобразований:

$$\begin{cases}\sum_{i=1}^{n}\left(x_i-m\right)=0\\[2ex] \dfrac{1}{2(\sigma^2)^2}\cdot\sum_{i=1}^{n}\left(x_i-m\right)^2=\dfrac{n}{2\sigma^2}\end{cases}\Rightarrow\begin{cases}m=\dfrac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i\right)=\overline{x};\\[2ex] \sigma^2=\dfrac{\sum_{i=1}^{n}\left(x_i-m\right)^2}{n}.\end{cases}$$

Ответ: оценкой максимального правдоподобия математического ожидания нормального закона распределения является среднее арифметическое значение выборки $m=\overline{x}$, а оценкой дисперсии – среднее значение квадрата отклонения элементов выборки от среднего значения $\sigma^2=\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$.

Пример 2. Методом максимального правдоподобия найти точечную оценку параметра λ по выборке, приведенной в табл. 4.4.

Таблица 4.4

Данные выборки

x_i	1…3	3…5	5…7	7…9	9…11	11…13	13…15	15…17	17…19
n_i	5	6	7	15	22	27	30	34	35

Известно, что соответствующая непрерывная случайная величина имеет плотность распределения:

$$f(x)=\begin{cases}\lambda\cdot e^{\lambda(x-20)}, & x\le 20;\\ 0, & x>20.\end{cases}$$

Решение

Объем выборки $n=\sum n_i=181$. Выборка разбита на 9 интервалов, тогда x_i – это результат, попавший в i-й интервал.

Случайная величина подчиняется экспоненциальному закону, который имеет один параметр λ, поэтому функция правдоподобия примет вид (формула (4.15)):

$$L=\prod_{i=1}^{181} f(\lambda,\ x_i)=\prod_{i=1}^{5} f(\lambda, x_1)\cdot\prod_{i=1}^{6} f(\lambda, x_2)\cdot\ldots\cdot\prod_{i=1}^{35} f(\lambda, x_9)=$$

$$=\lambda^{181}\cdot e^{\lambda(5x_1+6x_2+\ldots+35x_9)-20\cdot 181\cdot\lambda}=\lambda^{181}\cdot e^{\lambda\sum\limits_{i=1}^{181}x_i-3620\cdot\lambda}.$$

Логарифм функции правдоподобия имеет вид

$$\ln L(\lambda, x)=\ln\left(\lambda^{181}\cdot e^{\lambda\sum\limits_{i=1}^{181}x_i-3620\cdot\lambda}\right)=181\cdot\ln(\lambda)+\lambda\sum_{i=1}^{181}x_i-3620\cdot\lambda.$$

Так как функция правдоподобия имеет одну неизвестную λ, то необходимо решить одно уравнение (формула (4.16)):

$$\frac{\partial\ln L(x,\lambda)}{\partial\lambda}=\frac{181}{\lambda}+\sum_{i=1}^{181}x_i-3620=0\Rightarrow\lambda=\frac{181}{3620-\sum\limits_{i=1}^{181}x_i}=\frac{1}{20-\frac{1}{181}\cdot\sum\limits_{i=1}^{181}x_i}=\frac{1}{20-\overline{x}}.$$

Итак, в качестве оценки максимального правдоподобия параметра λ принимается величина $\lambda=\dfrac{1}{20-\overline{x}}$.

Для получения точечной оценки (числового значения) необходимо вычислить выборочное среднее значение выборки, используя формулу (4.5):

$$\overline{x}=\frac{\sum x_i\cdot n_i}{n}.$$

В качестве значений следует принять середины интервалов, тогда

$$\overline{x}=\frac{2\cdot 5+4\cdot 6+6\cdot 7+8\cdot 15+10\cdot 22+12\cdot 27+14\cdot 30+16\cdot 34+18\cdot 35}{181}=$$

$$=12{,}895.$$

Искомая оценка равна:

$$\lambda=\frac{1}{20-\overline{x}}=\frac{1}{20-12{,}895}=0{,}141.$$

Ответ: точечная оценка параметра λ равна 0,141.

4.3.2. Метод моментов

Метод моментов предложен К. Пирсоном (см. рис. 1.11) в 1894 г. Сущность метода состоит в приравнивании теоретических моментов рассматриваемого распределения соответствующим эмпирическим моментам того же порядка и решении полученной системы уравнений относительно неизвестных параметров распределения Θ. Выбирается столько эмпирических моментов, сколько требуется оценить неизвестных параметров распределения. Желательно применять моменты младших порядков, так как погрешности вычисления оценок резко возрастают с увеличением порядка момента.

Теоретический момент – это момент, вычисленный аналитическим путем через плотность распределения $f(x)$ соответствующего закона распределения (см. разд. 2.3).

В соответствии с функцией (2.17) начальный момент соответствующего s-го порядка равен:

$$\alpha_s = M\left[X^s\right] = \sum x^s \cdot p = \int x^s \cdot f(x)dx .$$

Если записать плотность распределения $f(x)$ через ее параметры, которые в данном случае неизвестны, то формула (2.17) примет вид

$$\alpha_s(\Theta_1, \Theta_2, ..., \Theta_n) = \int x^s \cdot f(x, \Theta_1, \Theta_2, ..., \Theta_n)dx , \tag{4.18}$$

где $\Theta_1, \Theta_2, ..., \Theta_n$ – неизвестные параметры распределения.

Аналогично используя функцию (2.19), можно получить формулу для центрального момента s-го порядка:

$$\begin{aligned} \mu_s\left(\Theta_1, \Theta_2, ..., \Theta_n\right) &= M\left[\left(X - m\right)^s\right] = \\ &= \int \left(x - m\right)^s \cdot f\left(x, \Theta_1, \Theta_2, ..., \Theta_n\right)dx. \end{aligned} \tag{4.19}$$

Эмпирический момент – это момент, найденный по результатам наблюдения (по выборке через x_i).

Наиболее часто встречающиеся моменты:

- Начальный момент 1-го порядка – математическое ожидание или выборочное среднее значение. Рассчитывается по формуле (4.4) или формуле (4.5) (в зависимости от вида выборки):

$$\alpha_1 = m = \overline{x}. \tag{4.20}$$

- Центральный момент 2-го порядка – выборочная дисперсия. Рассчитывается по формуле (4.8) или (4.9) (в зависимости от вида выборки):

$$\mu_2 = D = \frac{\sum (x_i - \overline{x})^2}{n} . \tag{4.21}$$

Также центральный момент (дисперсия) может быть вычислен через начальный момент (с учетом формулы 2.20):

$$\mu_2 = \mathrm{D} = \frac{\sum (x_i)^2}{n} - \overline{x}\ . \qquad (4.22)$$

Метод моментов позволяет получить состоятельные оценки. Смещение удается устранить путем введения поправок. Эффективность оценок невысокая, т. е. даже при больших объемах выборок дисперсия оценок относительно велика (за исключением нормального распределения, для которого метод моментов дает эффективные оценки). В реализации метод моментов проще метода максимального правдоподобия. Его целесообразно применять для оценки не более чем четырех параметров, так как точность выборочных моментов резко падает с увеличением их порядка.

Пример 3. Дана выборка (табл. 4.5) из случайной величины, принадлежащей равномерному закону распределения.

Таблица 4.5

Данные выборки

x_i	3	5	7	9	11	13	15	17	19	21
n_i	21	16	15	26	22	14	21	22	18	25

Найти оценку параметров a и b методом моментов.

Решение

Зная значения мат. ожидания и дисперсии, выраженные через параметры равномерного распределения (разд. 2.3.2):

$$m = \frac{a+b}{2};\ D = \frac{(b-a)^2}{12},$$

можно приравнять их значения к соответствующим эмпирическим моментам (формулы (4.20) и (4.21)). В результате получается система уравнений:

$$\begin{cases} m = \dfrac{a+b}{2} = \alpha_1; \\ D = \dfrac{(b-a)^2}{12} = \mu_2. \end{cases} \qquad (4.23)$$

Эмпирические моменты на основе заданной выборки (табл. 4.5) соответственно равны:

$$\alpha_1 = \frac{\sum x_i \cdot n_i}{n} = 12{,}31 \text{ и } \mu_2 = \frac{\sum x_i^2 \cdot n_i}{n} - \overline{x}^2 = 33{,}78\,.$$

После подстановки система уравнений (4.23) примет вид

$$\begin{cases} \dfrac{a+b}{2} = 12{,}31; \\ \dfrac{(b-a)^2}{12} = 33{,}78. \end{cases}$$

В результате решения системы уравнений параметр a = 2,243 и b = 22,377.

Ответ: с учетом округления до десятых параметры равномерного распределения a = 2,2 и b = 22,4.

Пример 4. Найти методом моментов по выборке $x_1, x_2, \ldots, x_n$ точечную оценку неизвестного параметра λ экспоненциального распределения, плотность распределения которого

$$f(x) = \begin{cases} 0; & x < 0; \\ \lambda \cdot e^{-\lambda x}; & x \geq 0. \end{cases}$$

Решение

Зная значения (разд. 2.3.2) мат. ожидания и дисперсии, выраженные через параметры экспоненциального распределения:

$$m = \frac{1}{\lambda} \text{ и } D = \frac{1}{\lambda^2},$$

можно приравнять их значения к соответствующим эмпирическим моментам (формулы (4.20) и (4.22)). В результате получается система уравнений:

$$\begin{cases} m = \dfrac{1}{\lambda} = \overline{x}; \\ D = \dfrac{1}{\lambda^2} = \dfrac{\sum x_i^2 \cdot n_i}{n} - \overline{x}^2, \end{cases}$$

решить которую можно, подставив данные выборки.

В данном примере можно получить две оценки параметра λ:

$$\lambda = \frac{1}{\overline{x}} \text{ и } \lambda = \sqrt{\frac{n}{\left(\sum x_i^2 \cdot n_i\right) - \overline{x}^2 \cdot n}}.$$

Примечание. Какую оценку использовать в дальнейшем, принимает решение пользователь, исходя из предпочтений или дополнительных условий. Также можно взять усредненное значение.

Так как в условии задачи не даны числовые значения выборки, то нет возможности определить числовые значения искомого параметра λ.

Ответ: параметр экспоненциального распределения $\lambda = \dfrac{1}{\bar{x}}$ или

$$\lambda = \sqrt{\frac{n}{\left(\sum x_i^2 \cdot n_i\right) - \bar{x}^2 \cdot n}}.$$

Пример 5. Число нестандартных изделий в партии распределено по закону Пуассона. Проверено 200 партий изделий (выборка объема $n = 200$). Распределение нестандартных деталей в партии приведено в табл. 4.6.

Таблица 4.6

Распределение нестандартных деталей

x_i	0	1	2	3	4
n_i	132	43	20	3	2

Найти методом моментов точечную оценку неизвестного параметра λ распределения Пуассона.

Решение

В разд. 2.3.1 приведены значения мат. ожидания и дисперсии, выраженные через параметры пуассоновского распределения:

$$m = \lambda \text{ и } D = \lambda .$$

Для решения задачи необходимо приравнять значения данных параметров соответствующим эмпирическим моментам (формулы (4.20) и (4.22)), получается система уравнений:

$$\begin{cases} m = \lambda = \alpha_1; \\ D = \lambda = \mu_2, \end{cases}$$

решить которую можно, подставив данные выборки.

Эмпирические моменты на основе данной выборки соответственно равны:

$$\alpha_1 = \frac{\sum x_i \cdot n_i}{n} = \frac{0 \cdot 132 + 1 \cdot 43 + 2 \cdot 20 + 3 \cdot 3 + 4 \cdot 2}{200} = 0,5$$

и

$$\mu_2 = \frac{\sum x_i^2 \cdot n_i}{n} - \bar{x}^2 = \frac{0^2 \cdot 132 + 1^2 \cdot 43 + 2^2 \cdot 20 + 3^2 \cdot 3 + 4^2 \cdot 2}{200} - 0,5^2 = 0,66.$$

В данном примере можно получить две оценки параметра λ: $\lambda = 0,5$ и $\lambda = 0,66$. Какую именно оценку использовать в дальнейшем, решается в каждом конкретном случае, но следует помнить, что момент меньшего порядка дает меньшую ошибку.

Ответ: точечная оценка параметра распределения Пуассона $\lambda = 0,5$ или $\lambda = 0,66$.

Задачи для самостоятельного решения

Задача 1. Число семян сорняков в пробах зерна подчинено закону Пуассона. Имеется выборка проб зерна. Результаты представлены в табл. 4.7.

Таблица 4.7

Число семян сорняков в пробах зерна

Число семян сорняков, x_i	0	1	2	3	4	5
Количество проб, n_i	9	39	40	24	11	7

Найти параметр λ по выборке методом моментов.

Задача 2. Отклонение контролируемого размера изделия от номинального подчинено нормальному закону распределения с неизвестными параметрами m и σ. В табл. 4.8 приведены наблюдаемые отклонения от номинала для выборки из n = 200 изделий.

Таблица 4.8

Отклонение контролируемого размера изделия

Отклонения от номинала, мм, x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
Число наблюдений, n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки неизвестных параметров m и σ нормального распределения.

Задача 3. Число появлений событий A в k независимых испытаниях подчинено биноминальному закону распределения с неизвестным параметром p. В табл. 4.9 приведено эмпирическое распределение числа появлений события A в 10 опытах по 5 испытаний в каждом.

Таблица 4.9

Число появлений события А в испытаниях

Число появлений события A, x_i	0	1	2	3	4
Число наблюдений, n_i	5	2	1	1	1

Найти методом моментов точечную оценку неизвестного параметра p биноминального распределения.

Задача 4. Время безотказной работы элемента имеет экспоненциальное распределение (разд. 2.3.2):

$$p(x) = \begin{cases} 0; & x < 0; \\ \lambda \cdot e^{-\lambda x}; & x \geq 0. \end{cases}$$

В табл. 4.10 приведено эмпирическое распределение среднего времени работы 1000 элементов.

Таблица 4.10

Время безотказной работы элементов

Время безотказной работы A, x_i, ч	5	15	25	35	45	55	65
Кол-во элементов, n_i	365	245	150	100	70	45	25

Найти методом моментов точечную оценку неизвестного параметра λ экспоненциального распределения.

4.4. Интервальная оценка параметров распределения

Точечные оценки зависят от значений выборки, по которым они рассчитываются. Следовательно, точечные оценки являются случайными величинами, и их отклонения от оцениваемых параметров также случайны. Погрешности оценок принято характеризовать их дисперсиями или средними квадратическими отклонениями. Истинное значение искомого параметра может быть представлено точечной оценкой с учетом погрешности в виде

$$\Theta = \overline{\Theta} \pm \sigma_{\overline{\Theta}}, \qquad (4.24)$$

но данная оценка не всегда объективна.

Чтобы дать представление о точности и надежности оценки в математической статистике, пользуются понятиями *«доверительный интервал»* и *«доверительная вероятность»*.

Доверительный интервал – это интервал $[\Theta_1;\ \Theta_2]$, между границами которого с заданной вероятностью $P = 1 - \alpha$ будет заключено истинное значение случайной величины Θ, т. е.

$$P\left[\Theta_1 < \Theta < \Theta_2\right] = 1 - \alpha\,. \qquad (4.25)$$

Вероятность P принято называть *доверительной вероятностью*, а α – *уровнем значимости.*

Чем уже интервал, тем точнее оценка неизвестного параметра Θ. Задача состоит в нахождении по заданной доверительной вероятности границ этого интервала $[\Theta_1;\ \Theta_2]$. Выбор доверительной вероятности P зависит от конкретных условий и в измерительной практике обычно принимается равным 0,90; 0,95; 0,98; 0,99 и реже (в особо ответственных случаях) – 0,999 [10].

Для построения доверительного интервала в общем случае необходимо найти закон распределения оценки. Если оценка несмещенная, то плотность распределения оценки $f(\overline{\Theta})$ сосредоточена в окрестности неизвестного истинного значения Θ. По известной плотности распределения оценки можно установить границы доверительного интервала по заданной доверительной вероятности:

$$P = p_1 - p_2 = P(\bar{\Theta}_{p_1} \le \Theta \le \bar{\Theta}_{p_2}) = \int\limits_{\bar{\Theta}_{p2}}^{\bar{\Theta}_{p1}} f\left(\bar{\Theta}\right) d\bar{\Theta},$$

где $\bar{\Theta}_{p1}$, $\bar{\Theta}_{p2}$ – квантили распределения оценки порядка p_1 и p_2 (рис. 4.1).

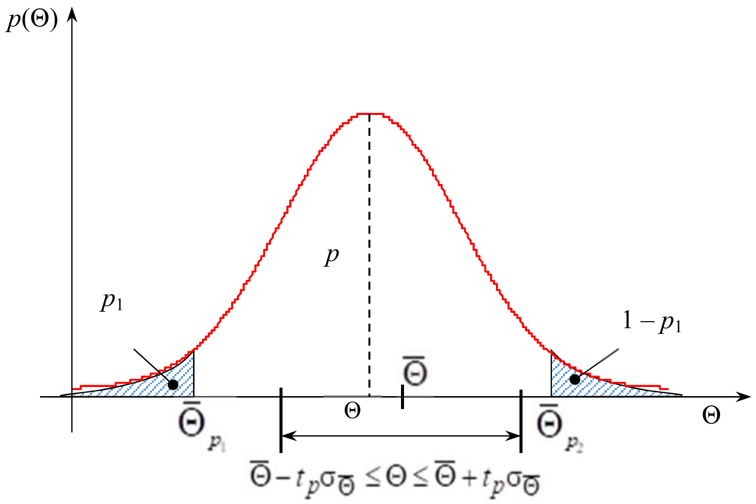


Рис. 4.1. Интервальное оценивание

При симметричном двухстороннем оценивании принимается: $p_1 = \dfrac{1-P}{2}$, $p_2 = \dfrac{1+P}{2}$. Для наиболее часто встречающейся доверительной вероятности $P = 0{,}95$ квантили равны: $p_1 = 0{,}025$ и $p_2 = 0{,}975$.

С вероятностью P справедливо неравенство

$$\bar{\Theta}_{p1} \le \Theta \le \bar{\Theta}_{p2}.$$

Представив отклонения квантилей от центра распределения через среднее квадратичное отклонение оценки (формула (4.24)), имеем интегральную оценку вида

$$\bar{\Theta} - t_p\sigma_{\bar{\Theta}} \le \Theta \le \bar{\Theta} + t_p\sigma_{\bar{\Theta}}.$$

Данное неравенство справедливо для симметричного распределения оценки. В противном случае имеют место два коэффициента: t_1 и t_2.

Полученный доверительный интервал с вероятностью P заключает в себе (накрывает) истинное значение Θ (рис. 4.1): истинное значение параметра постоянно, а доверительный интервал случайный по положению и ширине (зависит от случайных величин Θ и $\sigma_{\bar{\Theta}}$).

Определение законов распределения оценок – сложная и трудоемкая задача, поэтому желательно найти такую функцию $f(\overline{\Theta}, \Theta)$, точное распределение которой известно и не зависит от истинного значения Θ.

При интервальном оценивании параметров нормального распределения такими функциями являются переменные Стьюдента $t = \dfrac{(\overline{x} - m)}{\sqrt{\dfrac{S^2}{n}}}$ и Пирсона $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$, которые, как известно, распределены, соответственно, по законам Стьюдента и Пирсона (разд. 2.3.2) с $n - 1$ степенями свободы.

4.4.1. Построение доверительного интервала для математического ожидания

Построение доверительных интервалов для математического ожидания зависит от априорной информации о распределении генеральной совокупности. Для нормального закона с известной дисперсией среднее арифметическое распределено нормально с дисперсией $\sigma_{\overline{x}}^2 = \dfrac{\sigma^2}{n}$, т. е. известно точное распределение оценки, поэтому интервальная оценка примет вид

$$\overline{x} - U_p \sigma_{\overline{x}} \le m \le \overline{x} + U_p \sigma_{\overline{x}}, \tag{4.26}$$

где U_p – коэффициент стандартного нормального закона ($m = 0$, $\sigma = 1$) для двусторонней доверительной вероятности P.

Коэффициент U_p находится из соотношения $P = 2\Phi(U_p)$, где $\Phi(U_p)$ – значение интегральной функции Лапласа (прил. 1, табл. П1.2).

Пример 1. Найти доверительный интервал для оценки неизвестного математического ожидания m нормально распределенного признака X генеральной совокупности, если генеральное среднее квадратическое отклонение $\sigma = 5$, выборочная средняя $\overline{x} = 14$ и объем выборки $n = 25$. Доверительная вероятность $P = 0{,}95$.

Решение

По условию задачи известно генеральное среднее квадратическое отклонение σ, поэтому доверительный интервал можно представить в виде (4.26).

Коэффициент U_p находится из соотношения $\Phi(U_p) = \dfrac{P}{2} = \dfrac{0{,}95}{2} = 0{,}475$. По таблице значений функции Лапласа (прил. 1, табл. П1.2) коэффициент равен 1,96. После подстановки данных неравенство (4.26) примет вид

$$14-1{,}96\cdot\frac{5}{\sqrt{25}}\le m\le 4+1{,}96\cdot\frac{5}{\sqrt{25}}.$$

Искомый доверительный интервал

$$12{,}04\le m\le 15{,}96.$$

Обратите внимание. Количество разрядов в числах, обозначающих начало и конец интервала, должно быть одинаковое.

Ответ: доверительный интервал $12{,}04\le m\le 15{,}96$ накрывает неизвестное математическое ожидание m с вероятность $P = 0{,}95$.

Если дисперсия неизвестна, то для ряда наблюдений из нормальной совокупности среднее арифметическое распределено тоже по нормальному закону, но с неизвестной дисперсией. В этом случае по выборке находится точечная оценка дисперсии $S_{\bar{x}}^2=\dfrac{S^2}{n}$, т. е. точное распределение неизвестно. Поэтому в данном случае интервальная оценка строится на основе распределения Стьюдента:

$$\bar{x}-t_pS_{\bar{x}}\le m\le \bar{x}-t_pS_{\bar{x}}, \tag{4.27}$$

где t_p – коэффициент распределения Стьюдента для двусторонней доверительной вероятности P (прил. 2, табл. П2.1).

Пример 2. Из генеральной совокупности извлечена выборка объема $n = 10$ (табл. 4.11).

Таблица 4.11

Данные выборки

x_i	–2	1	2	3	4	5
n_i	2	1	2	2	2	1

Оценить неизвестное математическое ожидание m нормально распределенного признака X генеральной совокупности по выборочной средней при помощи доверительного интервала. Доверительная вероятность $P = 0{,}95$.

Решение

По условию задачи генеральное среднее квадратическое отклонение неизвестно, поэтому доверительный интервал необходимо представить в виде неравенства (4.27). Предварительно необходимо рассчитать выборочное среднее значение и исправленное среднее квадратическое отклонение.

Выборочное среднее значение (оценка мат. ожидания) рассчитывается по формуле (4.5):

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{2\cdot(-2)+1\cdot 1+2\cdot 2+3\cdot 2+4\cdot 2+5\cdot 1}{10} = 2.$$

Исправленное среднее квадратическое отклонение рассчитывается по формуле (4.14):

$$S = \sqrt{\frac{\sum (x_i - \overline{x})^2 \cdot n_i}{n-1}} =$$

$$= \sqrt{\frac{(-2-2)^2\cdot 2+(1-2)^2\cdot 1+(2-2)^2\cdot 2+(3-2)^2\cdot 2+(4-2)^2\cdot 2+(5-2)^2\cdot 1}{9}} =$$

$$= 2{,}4.$$

По таблице распределения Стьюдента (прил. 2, табл. П2.1) при доверительной вероятности $P = 0{,}95$ и числе степеней свободы $k = n - 1 = 9$ коэффициент Стьюдента $t = 2{,}26$.

Искомый доверительный интервал определяется по формуле (4.27):

$$2 - 2{,}26\frac{2{,}4}{\sqrt{10}} \le m \le 2 + 2{,}26\cdot\frac{2{,}4}{\sqrt{10}};$$

$$0{,}3 \le m \le 3{,}7.$$

Ответ: доверительный интервал $0{,}3 \le m \le 3{,}7$ покрывает неизвестное математическое ожидание m с надежностью 0,95.

4.4.2. Построение доверительного интервала для дисперсии

При построении доверительного интервала для дисперсии используется случайная величина, которая связана со значением выборочной дисперсия S^2 соотношением

$$\chi^2_{\frac{1-p}{2}} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\frac{1+p}{2}}. \quad (4.28)$$

Решив уравнение (4.28) относительно σ^2, получается

$$\frac{(n-1)S^2}{\chi^2_{\frac{1+p}{2}}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\frac{1-p}{2}}}, \quad (4.29)$$

где χ^2 – коэффициент распределения Пирсона (прил. 2, табл. П2.2).

Пример 3. По данным выборки объема $n = 16$ найдено «исправленное» среднее квадратическое отклонение $S = 1$ нормально распределенного количественного признака. Найти доверительный интервал для среднего квадратического отклонения σ с вероятность $P = 0{,}96$.

Решение

Задача сводится к отысканию доверительного интервала для дисперсии (формула (4.29)). По таблице χ^2-распределения (распределения Пирсона) (прил. 2, табл. П2.2) критические значения равны:

$$\lambda^2_{\frac{1-p}{2}} = \lambda^2_{\frac{1-0,96}{2}} = 5,2; \ \lambda^2_{\frac{1p}{2}} = \lambda^2_{\frac{1+0,96}{2}} = 28,3.$$

После подстановки найденных значений в формулу (4.29) интервал примет вид:

$$\frac{(16-1)\cdot 1^2}{28,3} \le \sigma^2 \le \frac{(16-1)\cdot 1^2}{5,2}; \ 0,53 \le \sigma^2 \le 2,88.$$

Ответ: доверительный интервал $0,53 \le \sigma^2 \le 2,88$ накрывает неизвестную дисперсию σ^2 с вероятность $P = 0,96$.

Пример 4. Определить число наблюдений над случайной величиной X для *примера 1* (разд. 4.4), если необходимо получить точность δ не менее 1,5.

Решение

Под *точностью* понимается *отклонение относительно точечной оценки*, т. е.

$$\overline{x} \pm U_p \sigma_{\overline{x}} = \overline{x} \pm \delta.$$

Следовательно, точность можно рассчитать по формуле

$$\delta = U_p \cdot \frac{\sigma}{\sqrt{n}}. \qquad (4.30)$$

Отсюда при известной точности можно определить требуемый объем выборки:

$$n \ge \left(\frac{U_p \cdot \sigma}{\delta}\right)^2 = \left(\frac{1,96\cdot 5}{1,5}\right)^2 \approx 42,68.$$

Так как объем выборки – это целое число, то по условию задачи необходимо выбрать ближайшее большее $n \ge 43$.

Ответ: выборка объема $n = 43$ позволит обеспечить точность $\delta = 1,5$.

Задачи для самостоятельного решения

Задача 1. Произведено 10 измерений одним прибором (без систематической ошибки) некоторой величины, причем «исправленное» среднеквадратическое отклонение S случайной ошибки измерений оказалось равным 0,8. Найти точность прибора с надежностью 0,95. Предполагается, что результаты измерений распределены нормально.

Задача 2. Найти доверительный интервал для оценки с надежностью 0,95 неизвестного математического ожидания m нормально распределенного признака X генеральной совокупности, если генеральное среднее квадратическое отклонение $\sigma = 2$, выборочная средняя $\overline{x} = 4,1$ и объем выборки $n = 16$.

Задача 3. Найти минимальный объем выборки, при котором с вероятностью $P = 0,975$ точность оценки математического ожидания m равна $\delta = 0,3$, если известно среднее квадратическое отклонение $\sigma = 1,2$ нормально распределенной генеральной совокупности.

Задача 4. Выборка из большой партии электроламп содержит 100 ламп. Средняя продолжительность горения лампы равна 1000 ч. Найти при доверительной вероятности $P = 0,9$ доверительный интервал для средней продолжительности горения лампы всей партии, если известно среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч. Предполагается, что продолжительность горения ламп распределена нормально.

Задача 5. По данным 16 независимых равноточных измерений некоторой величины найдены средний результат измерения $\overline{x} = 42,8$ и «исправленное» среднеквадратическое отклонение $S = 8$. Оценить истинное значение измеряемой величины с надежностью (доверительной вероятностью) $P = 0,95$.

Контрольные вопросы и задания

1. Что такое оценка? Какие требования предъявляют к оценкам?
2. Поясните понятия «точечная оценка», «генеральное и выборочное средние», «генеральная и выборочная дисперсии».
3. Какие методы оценивания применяют?
4. В чем заключается метод максимального правдоподобия?
5. В чем заключается метод моментов?
6. Что такое интервальное оценивание?
7. Поясните понятия «доверительный интервал», «доверительная вероятность».
8. Постройте доверительный интервал для математического ожидания при известной дисперсии генеральной совокупности; при неизвестной дисперсии генеральной совокупности.
9. Постройте доверительный интервал для дисперсии.

Глава 5. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Еще одной важнейшей задачей математической статистики является статистическая проверка гипотез. Несмотря на существенное различие задач оценивания (гл. 4) и проверки гипотез, между ними существует тесная взаимосвязь. На практике эти задачи могут предшествовать одна другой в зависимости от наличия априорной информации о свойствах случайных величин, объема статистического материала и конечной цели анализа.

В прикладных исследованиях часто возникает необходимость формулировки (высказывания) гипотезы, а затем и экспериментальной проверки предполагаемых утверждений (например, приводит ли к повышению производительности труда применение новой технологии). Далее детальному исследованию подвергаются полученные экспериментальные результаты (конечная выборка свойств какой-либо величины). Свойства величины могут выражаться как качественно, так и количественно. Цель исследования – определить, удовлетворяют ли они выдвинутым предположениям, на которых основаны вычислительные статистические процедуры (процедура оценивания): равенство дисперсий погрешностей, одинаковость законов распределения нескольких случайных величин и др.

Решение о справедливости гипотезы основывается на анализе выборки заданного размера (результатах наблюдений) и на знании некоторой априорной информации о законе распределения некоторых величин.

В математической статистике рассматриваются следующие гипотезы:

- о законе распределения некоторой случайной величины (генеральной совокупности);
- о числовых значениях параметров этого закона (например, математического ожидания или дисперсии);
- о наличии корреляционной зависимости между случайными величинами, определенными на множестве объектов одной и той же генеральной совокупности.

5.1. Статистические гипотезы. Ошибки первого и второго рода

После выдвижения предположения ставится задача, заключающаяся в том, чтобы подтвердить или опровергнуть это предположение, опираясь на результаты наблюдений, т. е. проверить статистическую гипотезу.

Статистическая гипотеза[3] – это любое предположение о виде неизвестного закона распределения или о параметрах известных распределений. Гипотезы о значениях параметров распределений или о сравнительной ве-

---

[3] В этой главе большинство терминов приводятся в соответствии с ГОСТ Р 50779.10-2000 (ИСО 3534.1-93) «Статистические методы. Вероятность и основы статистики. Термины и определения» [9].

личине параметров двух распределений называются параметрическими гипотезами. Гипотезы о виде распределения называются непараметрическими гипотезами.

Проверить статистическую гипотезу – значит проверить, согласуются ли выборочные данные с выдвинутой гипотезой.

Статистическая гипотеза, являющаяся утверждением о параметрах распределения некоторой случайной величины, называется *параметрической*.

Если речь идет о предположении о виде распределения (законе распределения) случайной величины, то такая гипотеза называется *непараметрической*.

Пример 1. Примеры статистических гипотез:

1) генеральная совокупность подчиняется нормальному закону распределения;

2) математические ожидания двух нормальных совокупностей равны между собой.

Первая гипотеза является непараметрической, а вторая – параметрической.

Нулевая (основная или проверяемая) гипотеза – предположение, которое подвергается проверке (обозначается H_0). Нулевая гипотеза рассматривается как утверждение, которое является особо важным. При наличии хотя бы одного противоречащего факта нулевая гипотеза отвергается, следовательно, теория, на которой она основана, должна быть *отвергнута*. Но стоит отметить, что при отсутствии противоречащих фактов нулевая гипотеза необязательно должна быть *принята*.

Другими словами, в основу проверки статистических гипотез положены принципы:

- маловероятные события считаются практически невозможными;
- события, вероятность которых близка к единице, считаются достоверными.

Наступление оцениваемых событий зависит от разного рода обстоятельств. Степень риска определяется тем, что событиями с малой вероятностью можно пренебречь [10].

Вместе с выдвигаемой основной гипотезой рассматривается противоречащая ей гипотеза. В том случае, если выдвинутая гипотеза будет отвергнута, противоречащая гипотеза окажется справедливой.

Конкурирующая (альтернативная) гипотеза – это утверждение, которое будет принято, если нулевая гипотеза отвергается (обозначается H_1).

Пример 2. Суть нулевой гипотезы H_0 состоит в предположении о том, что математическое ожидание m равно m_0. Тогда суть альтернативной гипотезы H_1 – в предположении, что математическое ожидание m не равно (больше или меньше) значению m_0

Запись будет выглядеть так:

$$H_0 : m = m_0 ;\ H_1 : m \neq m_0$$

или

$$H_1 : m > m_0 .$$

Первый вариант записи в *примере 2* является *простой* гипотезой, так как конкурирующая (альтернативная) гипотеза полностью определяет распределение случайной величины. Второй вариант – гипотеза сложная, так как нет полной определенности. В данном материале уделяется внимание в основном простым гипотезам. Этот материал дает лишь общее представление о процедуре проверки гипотез. В случае необходимости углубить знания в этом вопросе рекомендуется обратиться к литературе по математической статистике [5, 10].

Проверка любой статистической гипотезы осуществляется с помощью статистического критерия.

Статистический критерий – это статистический метод принятия решений о том, стоит ли отвергнуть нулевую гипотезу в пользу альтернативной или нет. В основу метода положена случайная величина (статистика), которая используется для проверки нулевой гипотезы. Числовое значение статистики позволяет сделать разумный выбор между нулевой и конкурирующей гипотезами.

Множество всех возможных значений выбранного статистического критерия разделяется на два непересекающихся подмножества: критическая область и область принятия гипотезы.

Критическая область – это множество возможных значений статистического критерия (статистики), при которых нулевая гипотеза отвергается.

Область принятия гипотезы – это множество возможных значений статистического критерия (статистики), при которых нулевая гипотеза принимается.

Критические точки (*квантили*) – это значения случайной величины статистики (точки), которые разграничивают критическую область и область принятия гипотезы. Обозначается: $K_{\text{кр}}$.

На рис. 5.1 показаны область принятия гипотезы и критическая область, разделенные между собой критической точкой $K_{\text{кр}}$. В том случае, если наблюдаемое значение статистического критерия (рассчитанное по выборочной совокупности) принадлежит критической области, нулевая гипотеза отвергается. Если же наблюдаемое значение статистического критерия принадлежит области принятия гипотезы, то нулевая гипотеза принимается.

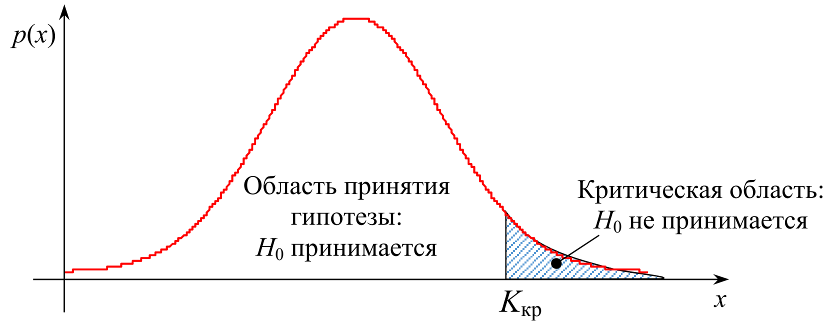


Рис. 5.1. Область принятия гипотезы и критическая область, разделенные критической точкой $K_{кр}$

В зависимости от попадания числового значения статистики в определенную область гипотеза может подтвердиться, а может и не подтвердиться.

При проверке гипотез приходится опираться на выборку, состав которой носит случайный характер. Случайность выборки подразумевает возможность ошибки в результатах статистических выводов.

В результате проверки статистических гипотез могут возникнуть ошибки двух видов: *ошибка первого рода* и *ошибка второго рода.*

Ошибка первого рода – это ошибка, состоящая в отбрасывании нулевой гипотезы H_0, поскольку статистика принимает значение, принадлежащее критической области.

Вероятность совершения ошибки первого рода обозначается α и называется уровнем значимости. У*ровень значимости* α – это заданное значение верхнего предела вероятности ошибки первого рода. Обычно задается близким к нулю (например, 0,05; 0,01; 0,02 и т. д.). Чем меньше уровень значимости α, тем меньше вероятность отвергнуть проверяемую гипотезу H_0, когда она верна, т. е. совершить ошибку первого рода.

При этом желательно, чтобы выбранный критерий минимизировал вероятность совершить ошибку второго рода.

Ошибка второго рода – это ошибка, состоящая в принятии нулевой гипотезы, поскольку статистика принимает значение, не принадлежащее критической области, в то время как нулевая гипотеза не верна (принимается гипотеза H_1 при верной гипотезе H_0).

Вероятность ошибки второго рода (не отклонить ложную гипотезу H_0) обозначается β, ее величина зависит от альтернативной гипотезы H_1.

Мощность критерия – это вероятность недопущения ошибки второго рода (правильного принятие гипотезы H_1). Числовое значение мощности критерия $\pi = 1 - \beta$ равно вероятности того, что наблюдаемое значение статистики попадет в критическую область при верной альтернативной гипотезе H_1.

Соотношение вероятностей совершения ошибок первого и второго рода и принятие/отклонение нулевой и альтернативной гипотез отражено в табл. 5.1 и на рис. 5.2.

Таблица 5.1

Соотношение ошибок первого и второго рода при проверке гипотез

	Принять нулевую гипотезу H_0	Принять альтернативную гипотезу H_1
Справедлива нулевая гипотеза H_0	Верное решение. Вероятность $1 - \alpha$	Неверное решение. Вероятность α. Ошибка первого рода
Справедлива альтернативная гипотеза H_1	Неверное решение. Вероятность β. Ошибка второго рода	Верное решение. Вероятность $1 - \beta$

Очевидно, что чем больше значение мощности используемого критерия $\pi = 1 - \beta$ при заданном значении уровня значимости α, тем лучше он различает гипотезы H_0 и H_1. Особенно важно, чтобы этот критерий хорошо различал близкие конкурирующие гипотезы. Графически это означает, что плотности распределения для каждой из гипотез должны быть максимально «раздвинуты» на числовой оси (рис. 5.2).

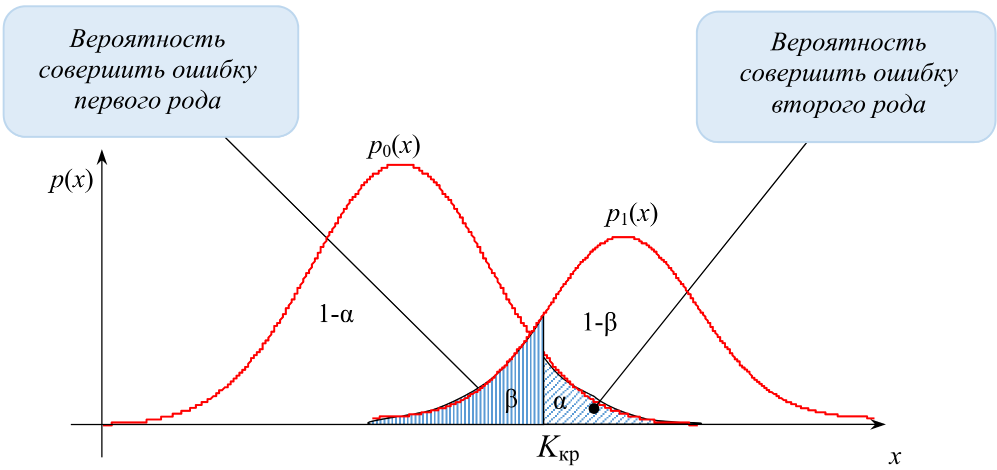


Рис. 5.2. Ошибки первого и второго рода

Какая из ошибок (первого или второго рода) является на практике более опасной, зависит от конкретной задачи.

Пример 3. При проверке правильности выбора метода лечения больного, ошибка первого рода означает отказ от правильной методики, что может замедлить лечение, а ошибка второго рода означает применение неправильной методики, что может привести к ухудшению состояния больного, и является более опасной.

Проверка статистической гипотезы осуществляется в следующем порядке:

1. Сформулировать нулевую H_0 и альтернативную H_1 гипотезы.
2. Задать уровень значимости α.
3. Выбрать статистический критерий K исходя из априорной информации и выдвигаемой гипотезы.
4. По выборке вычислить наблюдаемое значение критерия $K_{\text{набл}}$.
5. Вычислить критическую область и область принятия гипотезы, т. е. найти критическое значение $K_{\text{кр}}$ по уровню значимости в справочной таблице, соответствующей используемой статистике.
6. Сделать вывод о принятии гипотезы:
   - если значение $K_{\text{набл}}$ попадает в критическую область, нулевая гипотеза отвергается (можно ввести условное обозначение ($H_0 -$));
   - если значение $K_{\text{набл}}$ попадает в область принятия гипотезы, нулевая гипотеза принимается (можно ввести условное обозначение ($H_0 +$)).
7. Сделать содержательный вывод в соответствии с условием задачи.

5.2. Гипотезы о значениях числовых характеристик

Гипотезы о значениях числовых характеристик относятся к параметрическим гипотезам. Они применяются при сравнении информации о генеральной совокупности и числовых значений, найденных по выборкам. На практике уделяется внимание гипотезам о равенстве средних значений m и дисперсии σ^2 определенным числам m_0 и σ_0^2 или сравнению между собой числовых характеристик для нескольких выборок.

Пример 1. Задачи проверки гипотез о значениях числовых характеристик возникают при проверке качества функционирования измерительных устройств. Если m_0 – номинальное значение измеряемого параметра, а в результате проверки гипотезы обнаруживается, что $m \neq m_0$, то это означает, что прибор дает систематическую ошибку. Точность прибора определяется значением σ_0, и если $\sigma > \sigma_0$, то это означает, что качество прибора не отвечает стандартным требованиям.

Сравнение средних – это проверка гипотезы о равенстве средних значений совокупностей, из которых получены выборки.

На практике часто встречаются ситуации, когда среднее значение данных одного эксперимента отличается от среднего значения данных другого эксперимента, хотя условия эксперимента являются схожими. Тогда возникает вопрос, можно ли считать это расхождение незначимым, т. е. чисто случайным, или оно вызвано существенным различием двух генеральных совокупностей.

Пример 2. Необходимость проверки гипотезы о равенстве средних значений возникает: при исследовании надежности технических систем, где результаты сравниваются с предыдущими измерениями; при контроле качества изделий, изготовленных на разных предприятиях; в финансах – при сравнении уровня доходности различных активов.

При изучении влияния систематических погрешностей на результат измерений, при объединении нескольких серий результатов, полученных в схожих условиях, но разными исследователями, по различным методикам или в разные моменты времени возникает необходимость сравнения полученных выборок. В этих случаях проводится проверка гипотез о сравнении средних значений. Ставится вопрос о сравнении центров распределений двух случайных величин при различной априорной информации об их законах распределения и дисперсиях.

Сравнение дисперсий – это проверка гипотезы о равенстве значений дисперсии для различных выборок.

В науке, технике, экономике, медицине, социологических исследованиях гипотезы о дисперсиях играют очень большую роль. Дисперсия характеризует рассеяние и тем самым показывает изменяемость интересующего признака в определенных условиях. Среднее значение сглаживает, затушевывает эти изменения, а дисперсия их выявляет.

Пример 3. Дисперсия характеризует точность приборов, технологических процессов, риск, связанный с отклонением доходности от заданного уровня, и т. д.

Сравнение дисперсий в измерительной практике принципиально необходимо при объединении двух и более рядов измерений, выполнить которое можно только при одинаковой точности полученных результатов (при равенстве дисперсий)

В зависимости от априорной информации можно выделить несколько характерных случаев:

- Сравнение выборочной средней с гипотетической генеральной средней нормальной совокупности при известной и неизвестной дисперсиях (гипотеза о генеральной средней нормального распределения).

- Сравнение двух средних генеральных совокупностей с известными и неизвестными дисперсиями (гипотеза о равенстве генеральных средних двух распределений).
- Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности (гипотеза о генеральной дисперсии нормального распределения).
- Сравнение двух дисперсий нормальных генеральных совокупностей (гипотеза о равенстве генеральных дисперсий двух нормальных распределений).

Для лучшего понимания следует привести пояснения для каждого характерного случая.

Сравнение выборочной средней и гипотетической генеральной средней нормальной совокупности

Генеральная совокупность X распределена нормально, причем генеральная средняя m хотя и неизвестна, но имеются основания предполагать, что она равна гипотетическому (предполагаемому) значению m_0.

Таким образом, требуется проверить гипотезу о том, что математическое ожидание выборочной средней равно гипотетической генеральной средней. Другими словами, надо установить, значимо или незначимо различаются выборочная и генеральная средние.

Если нулевая гипотеза справедлива, это означает, что выборочная и гипотетическая генеральная средние различаются незначимо, отличия объясняются случайным отбором объектов выборки.

Если нулевая гипотеза отвергнута, это означает, что выборочная и гипотетическая генеральная средние различаются значимо, отличия объясняются принадлежностью выборки другой генеральной совокупности.

Пример 4. X – совокупность размеров x_i партии деталей, изготавливаемых станком-автоматом. Можно предположить, что математическое ожидание генеральной совокупности m (хотя и неизвестно наверняка) равно проектному размеру m_0 Чтобы проверить это предположение, необходимо проверить некоторую выборку и установить, значимо ли отличаются числовые значения выборочного среднего $\overline{x}$ и математическое ожидание генеральной совокупности m_0 Если различия окажутся незначительными, то в среднем станок обеспечивает проектный размер, если различие значимо, то станок требует наладки.

В зависимости от априорной информации возможны два варианта:

1. Дисперсия генеральной совокупности известна.

Предположение, что дисперсия генеральной совокупности известна, можно сделать, если:

- были проведены предварительные исследования;

- работа станка досконально изучена ранее;
- на основе документации найдено теоретическое значение;
- дисперсия вычислена по выборке большего объема (выборка большого объема дает достаточно хорошую оценку дисперсии).

2. Дисперсия генеральной совокупности неизвестна.

В противном случае считается, что дисперсия генеральной совокупности неизвестна.

Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны (независимые выборки)

Генеральные совокупности X и Y распределены нормально, причем их дисперсии известны (например, из предшествующего опыта или найдены теоретически). По извлеченным из этих совокупностей независимым выборкам, объемы которых соответственно равны n_1 и n_2, найдены выборочные средние $\overline{x}$ и $\overline{y}$.

Таким образом, требуется проверить гипотезу о том, что математические ожидания выборочных средних равны между собой. Такая задача ставится потому, что, как правило, выборочные средние оказываются различными. Возникает вопрос: значимо или незначимо различаются выборочные средние?

Если нулевая гипотеза справедлива (т. е. генеральные средние одинаковы), то различие выборочных средних незначимо и объясняется случайными причинами и, в частности, случайным отбором объектов выборки.

Если нулевая гипотеза отвергнута (т. е. генеральные средние неодинаковы), то различие выборочных средних значимо и объясняется тем, что сами генеральные средние (мат. ожидания) различны.

Пример 5. Если среднее арифметическое $\overline{x}$ результатов измерений величины A значимо отличается от среднего арифметического $\overline{y}$ результатов измерений величины B, то это означает, что истинные размеры (мат. ожидания) этих величин различны. Если величины A и B имеют одинаковые истинные размеры, а средние арифметические $\overline{x}$ и $\overline{y}$ результатов измерений этих величин различны, то это различие незначимое.

Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых неизвестны и одинаковы (малые независимые выборки)

Генеральные совокупности X и Y распределены нормально, причем их дисперсии неизвестны (например, при использовании выборки малого объема, так как по выборкам малого объема нельзя получить хорошие оценки генеральных дисперсий).

Таким образом, требуется проверить гипотезу о том, что математические ожидания выборочных средних равны между собой. Другими словами, требуется установить, значимо или незначимо различаются выборочные средние $\overline{x}$ и $\overline{y}$, найденные по независимым малым выборкам объемов n_1 и n_2.

Если нулевая гипотеза справедлива (т. е. генеральные средние одинаковы), то различие выборочных средних незначимо и объясняется случайными причинами и, в частности, случайным отбором объектов выборки.

Если нулевая гипотеза отвергнута (т. е. генеральные средние неодинаковы), то различие выборочных средних значимо и объясняется тем, что сами генеральные средние (мат. ожидания) различны.

В зависимости от априорной информации возможны два варианта.

1. Дисперсии генеральных совокупностей неизвестны, но равны между собой.

Предположение, что дисперсии генеральных совокупностей неизвестны, но равны между собой можно сделать, если выборки получены в одинаковых условиях, с использованием одинаковых инструментов, по одной методике и т. д.

Пример 6. Если сравниваются средние размеры двух партий деталей, изготовленных на одном и том же станке, то естественно допустить, что дисперсии контролируемых размеров одинаковы.

2. Дисперсии генеральных совокупностей неизвестны.

В этом случае нет никакой дополнительной информации, позволяющей считать, что выборки получены в одинаковых условиях, с использованием одинаковых инструментов, по одной методике и т. д.

В любом случае следует предварительно проверить гипотезу о равенстве дисперсий.

Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности

Генеральная совокупность распределена нормально, причем генеральная дисперсия хотя и неизвестна, но имеются основания предполагать, что она равна гипотетическому (предполагаемому) значению σ_0^2. На практике σ_0^2 устанавливается на основании предшествующего опыта или теоретически.

Итак, требуется проверить гипотезу о том, что математическое ожидание исправленной дисперсии равно гипотетическому значению генеральной дисперсии. Другими словами, требуется установить, значимо или незначимо различаются исправленная выборочная и гипотетическая генеральная дисперсии.

На практике рассматриваемая гипотеза проверяется, если нужно проверить точность приборов, инструментов, станков, методов исследования или устойчивость технологических процессов (определить, равно ли рассматриваемое значение установленному, например, в технической документации).

Пример 7. Если известна допустимая характеристика рассеяния контролируемого размера деталей, изготавливаемых станком-автоматом, равная σ_0^2, а найденная по выборке окажется значимо больше σ_0^2, то станок требует наладки.

Сравнение двух дисперсий нормальных генеральных совокупностей

На практике задача сравнения дисперсий возникает, если требуется сравнить между собой точность нескольких приборов, инструментов, методов измерений и т. д. Очевидно, предпочтительнее тот прибор, инструмент или метод, который обеспечивает наименьшее рассеяние результатов измерений, т. е. наименьшую дисперсию.

Генеральные совокупности X и Y распределены нормально. По независимым выборкам, объемы которых соответственно равны n_1 и n_2, извлеченным из этих совокупностей, найдены исправленные выборочные дисперсии S_x^2 и S_y^2.

Таким образом, требуется проверить гипотезу о том, что математические ожидания исправленных выборочных дисперсий равны между собой. Такая задача ставится потому, что обычно исправленные дисперсии оказываются различными. Возникает вопрос: значимо (существенно) или незначимо различаются исправленные дисперсии?

Если нулевая гипотеза справедлива (т. е. генеральные дисперсии одинаковы), то различие исправленных дисперсий незначимо и объясняется случайными причинами.

Если нулевая гипотеза отвергнута (т. е. генеральные дисперсии неодинаковы), то различие исправленных дисперсий значимо и является следствием того, что сами генеральные дисперсии различны.

Пример 8. Если различие исправленных выборочных дисперсий результатов измерений, выполненных двумя приборами, оказалось незначимым, то приборы имеют одинаковую точность. Если различие исправленных выборочных дисперсий результатов измерений, произведенных двумя приборами, оказалось значимым, то точность приборов различна.

Далее рассматриваются все отмеченные случаи в отдельности и предлагаются алгоритмы решения поставленных задач.

5.2.1. Сравнение средних

Сравнение выборочного среднего с гипотетической средней нормальной совокупности

1. Данные о генеральной совокупности: математическое ожидание m_0 и дисперсия σ^2 известны.

По выборки объема *n* получено выборочное среднее $\bar{x}$. Необходимо проверить гипотезу о равенстве математического ожидания *m* совокупности, из которой получена выборка объема *n*, генеральному среднему m_0.

Формализация гипотезы. Нулевая гипотеза H_0 : $m = m_0$, σ^2; альтернативная гипотеза H_1 : $m \neq m_0$.

Так как среднее арифметическое распределено нормально $N(m_0, \sigma/\sqrt{n})$ с известной дисперсией, то статистика примет вид

$$U = \frac{(\bar{x} - m)}{\sigma_{\bar{x}}} = \frac{(\bar{x} - m)\cdot\sqrt{n}}{\sigma}. \qquad (5.1)$$

Статистика подчиняется стандартному нормальному закону распределения для двусторонней доверительной вероятности $P = 1 - \alpha$, поэтому критическое значение определяется по таблице значений интегральной функции Лапласа (прил. 1, табл. П1.2) на основе формулы

$$P = 2\Phi(U_{\text{кр}}).$$

Решение о принятии гипотезы принимается на основе верности неравенства $|U| < U_{\text{кр}}$. Если неравенство верно, нулевая гипотеза (H_0: $m = m_0$) принимается, значит выборка принадлежит генеральной совокупности со средним значением m_0.

Пример 9. По результатам $n = 9$ замеров установлено, что выборочное среднее время (в секундах) изготовления детали $\bar{x} = 48$. Предполагая, что время изготовления – нормально распределенная случайная величина с дисперсией $\sigma^2 = 9$, рассмотреть гипотезу $H_0 : m = 49$ против конкурирующей гипотезы $H_1 : m \neq 49$. Доверительная вероятность P = 95 %.

Решение

Здесь рассматривается простая гипотеза $H_0 : m = 49$. На уровне $P = 0{,}95$ по таблице значений интегральной функции Лапласа (прил. 1, табл. П1.2) критическое значение $U_{\text{кр}}$ = 1,96. Значения статистики находится по формуле (5.1):

$$U = \frac{(\bar{x} - m)\sqrt{n}}{\sigma} = \frac{(48 - 49)\sqrt{9}}{3} = -1.$$

Так как $|U| < U_{кр}$, то гипотеза H_0 принимается. Это означает, что выборка принадлежит генеральной совокупности со средним значением $m = 49$.

Ответ: гипотеза H_0 принимается (H_0+): выборка принадлежит генеральной совокупности со средним значением $m = 49$, следовательно, заявленное среднее время изготовления деталей 48 с соответствует заявленному значению 49 с.

2. Данные о генеральной совокупности: математическое ожидание m_0 известно, дисперсия σ^2 неизвестна.

По выборки объема n получено выборочное среднее $\bar{x}$ и выборочная дисперсия S^2. Необходимо проверить гипотезу о равенстве математического ожидания m совокупности, из которой получена выборка объема n, генеральному среднему m_0.

Формализация гипотезы. Гипотезы $H_0: m = m_0$, S^2; альтернативная гипотеза $H_1: m \neq m_0$.

Так как дисперсия неизвестна, то статистика примет вид

$$t = \frac{(\bar{x} - m)}{S_{\bar{x}}} = \frac{(\bar{x} - m) \cdot \sqrt{n}}{S}. \tag{5.2}$$

Статистика подчиняется распределению Стьюдента с числом степеней свободы $k = n - 1$, поэтому критическое значение $t_{кр} = t_{P;\,k}$ – коэффициент Стьюдента для двусторонней доверительной вероятности $P = 1 - \alpha$ (прил. 2, табл. П2.1).

Решение о принятии гипотезы принимается на основе верности неравенства $|t| < t_{кр}$. Если неравенство верно, нулевая гипотеза ($H_0: m = m_0$) принимается, значит выборка принадлежит генеральной совокупности со средним значением m_0.

Пример 10. По утверждению руководства фирмы, средний размер дебиторского учета равен 187,5 тыс. руб. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборка равна 175 тыс. руб. при среднем квадратичном отклонении 35 тыс. руб. Может ли оказаться в действительности правильным объявленный размер дебиторского счета? Доверительная вероятность $P = 0{,}95$.

Решение

Здесь $m_0 = 187{,}5$ тыс. руб., $\bar{x} = 175$ тыс. руб., $n = 10$, $S = 35$, $P = 0{,}95$. Так как дисперсия неизвестна, то для проверки гипотезы

$H_0: m = 187{,}5$ тыс. руб. используется статистика, подчиняющаяся распределению Стьюдента (формула (5.2)). Тогда

$$t = \frac{(\bar{x} - m)\sqrt{n}}{S} = \frac{(175 - 187{,}5) \cdot \sqrt{10}}{35} = -1{,}129.$$

Число степеней свободы $k = n - 1 = 9$.

По таблице критических точек для распределения Стьюдента (прил. 2, табл. П2.1) при $P = 0{,}95$ критическое значение $t_{0{,}95;\,9} = 2{,}25$.

Так как $|t| < t_{\text{кр}}$, то гипотеза H_0 о среднем размере дебиторского счета принимается на уровне доверия $P = 0{,}95$.

Ответ: гипотеза H_0 принимается (H_0+): выборка принадлежит генеральной совокупности со средним значением $m_0 = 187{,}5$, следовательно, полученное ревизором среднее арифметическое, равное 175 тыс. руб., соответствует действительности.

Сравнение выборочных средних двух выборок

3. Исходные данные: выборочные средние $\bar{x}$ и $\bar{y}$, полученные по выборкам объема n_1 и n_2 из нормальных совокупностей с известными дисперсиями σ_1^2, σ_2^2.

Требуется по выборочным средним при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные средние (мат. ожидания) рассматриваемых совокупностей равны между собой.

Формализация гипотезы. Гипотезы $H_0 : m_1 = m_2$, σ_1^2, σ_2^2; альтернативная гипотеза $H_1 : m_1 \neq m_2$.

Так как средние арифметические $\bar{x}$ и $\bar{y}$ распределены нормально с параметрами m_1, $\sigma_1/\sqrt{n_1}$ и m_2, $\sigma_2/\sqrt{n_2}$, то разность средних $(\bar{x} - \bar{y})$ также распределена нормально с дисперсией $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. Поэтому статистика примет вид

$$U = \frac{|\bar{x} - \bar{y}|}{\sigma_{(\bar{x}-\bar{y})}} = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = |\bar{x} - \bar{y}|\sqrt{\frac{n_1 n_2}{n_1\sigma_2^2 + n_2\sigma_1^2}}. \tag{5.3}$$

Статистика подчиняется стандартному нормальному закону распределения для двусторонней доверительной вероятности $P = 1 - \alpha$, поэтому критическое значение определяется аналогично п. 1 по таблице значений интегральной функции Лапласа (прил. 1, табл. П1.2).

Решение о принятии гипотезы принимается на основе верности неравенства $|U| < U_{кр}$ аналогично п. 1. Если неравенство верно, нулевая гипотеза ($H_0 : m_1 = m_2$) принимается, значит, выборки принадлежат генеральным совокупностям с одинаковыми средними значениями.

Пример 11. Было произведено $n_1 = 12$ измерений диаметра вала (в мм). При этом оказалось, что среднее $\overline{x}_1 = 10,2$, а стандартное среднее квадратичное отклонение $\sigma_1 = 0,05$. Затем вал поместили в условия с высокой температурой и провели $n_2 = 8$ измерений диаметра его оси. Среднее на этот раз оказалось равным $\overline{x}_2 = 10,25$, а стандартное отклонение $\sigma_2 = 0,06$.

Можно ли сделать вывод, что диаметр вала существенно увеличивается при увеличении температуры? Доверительная вероятность $P = 0,95$.

Решение

Здесь рассматривается простая гипотеза $H_0 : m_1 = m_2$ против конкурирующей $H_1 : m_1 \neq m_2$.

По условию задачи следует считать, что результаты измерений диаметра вала являются нормально распределенными случайными величинами с известными дисперсиями σ_1^2 и σ_2^2.

Для проверки гипотезы H_0 используется статистика (формула (5.3)):

$$U = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{10,2 - 10,25}{\sqrt{\frac{0,05^2}{12} + \frac{0,06^2}{8}}} = -1,9487.$$

Для доверительной вероятности $P = 0,95$ по таблице значений интегральной функции Лапласа (прил. 1, табл. П1.2) критическое значение равно $U_{кр} = 1,96$.

Так как $|U| < U_{кр}$, то гипотеза H_0 принимается на уровне доверия 0,95.

Ответ: гипотеза H_0 принимается (H_0+), поэтому можно считать, что диаметр вала существенно не увеличивается в условиях повышенной температуры.

4. Исходные данные: выборочные средние $\overline{x}$ и $\overline{y}$, полученные по выборкам объема n_1 и n_2 из нормальных совокупностей с не известными, но равными дисперсиями $\sigma_1^2 = \sigma_2^2$.

Требуется по выборочным средним при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные средние (мат. ожидания) рассматриваемых совокупностей равны между собой.

Формализация гипотезы. Гипотеза $H_0 : m_1 = m_2$, $\sigma_1^2 = \sigma_2^2$ (числовое значение неизвестно); альтернативная гипотеза $H_1 : m_1 \neq m_2$.

Так как дисперсия неизвестна, то статистика примет вид

$$t = \frac{|\overline{x} - \overline{y}|}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}. \tag{5.4}$$

S^2 – исправленная выборочная дисперсия разности средних ($\overline{x} - \overline{y}$):

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}. \tag{5.5}$$

Статистика подчиняется распределению Стьюдента с числом степеней свободы $k = n_1 + n_2 - 2$, поэтому критическое значение $t_{\text{кр}} = t_{P;\,k}$ – коэффициент Стьюдента для двусторонней доверительной вероятности $P = 1 - \alpha$ (прил. 2, табл. П2.1).

Решение о принятии гипотезы принимается на основе верности неравенства $|t| < t_{\text{кр}}$. Если неравенство верно – нулевая гипотеза ($H_0 : m_1 = m_2$) принимается – значит выборки принадлежат генеральным совокупностям с одинаковыми средними значениями.

Пример 12. Необходимо проверить гипотезу о равенстве средних m_1 и m_2 нормальных генеральных совокупностей с равными дисперсиями на уровне значимости $\alpha = 0{,}05$, если по выборкам объема $n_1 = 12$ и $n_2 = 20$ найдены выборочные значения $\overline{x}_1 = 10{,}7$ и $\overline{x}_2 = 11{,}9$, исправленные выборочные дисперсии $S_1^2 = 3{,}55$ и $S_2^2 = 2{,}76$.

Решение

По условию задачи необходимо проверить гипотезу $H_0 : m_1 = m_2$ при неизвестных, но равных дисперсиях генеральной совокупности $\sigma_1^2 = \sigma_2^2$; альтернативная гипотеза $H_1 : m_1 \neq m_2$.

Так как дисперсии неизвестны, но равны между собой, то для проверки гипотезы используется статистика, подчиняющаяся распределению Стьюдента (формула (5.4)). Предварительно необходимо вычислить исправленную выборочную дисперсию разности средних ($\overline{x} - \overline{y}$) (формула (5.5)):

$$S^2 = \frac{(12-1)\cdot 3{,}55 + (20-1)\cdot 2{,}76}{12+20-2} = 3{,}05.$$

Наблюдаемое значение равно:

$$t = \frac{|10{,}7 - 11{,}9|}{\sqrt{3{,}05\cdot\left(\frac{1}{12}+\frac{1}{20}\right)}} = 1{,}882.$$

Уровень значимости α = 0,05 соответствует доверительной вероятности $P = 0{,}95$. По таблице критических точек для распределения Стьюдента (прил. 2, табл. П2.1) для числа степеней свободы $k = n_1 + n_2 - 2 = 12 + 20 - 2 = 30$ критическое значение $t_{0{,}95;\,30} = 2{,}042$.

Так как $|t| < t_{\text{кр}}$, то гипотеза ($H_0 : m_1 = m_2$) принимается.

Ответ: гипотеза H_0 принимается $(H_0 +)$: различие между средними значениями можно считать случайным. Можно предположить, что обе выборки принадлежат одной генеральной совокупности.

5. Исходные данные: выборочные средние $\bar{x}$ и $\bar{y}$, полученные по выборкам объема n_1 и n_2 из нормальных совокупностей с неизвестными дисперсиями $\sigma_1^2 \neq \sigma_2^2$.

Требуется по выборочным средним при заданном уровне значимости α проверить нулевую гипотезу, состоящую в том, что генеральные средние (мат. ожидания) рассматриваемых совокупностей равны между собой.

Формализация гипотезы. Гипотеза $H_0 : m_1 = m_2$, $\sigma_1^2 \neq \sigma_2^2$ (числовые значения неизвестны); альтернативная гипотеза $H_1 : m_1 \neq m_2$.

Так как дисперсии неизвестны, то статистика примет вид

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}. \qquad (5.6)$$

Статистика подчиняется распределению Стьюдента с числом степеней свободы $k = n_1 + n_2 - 2$, поэтому критическое значение $t_{\text{кр}} = t_{P;\,k}$ – коэффициент Стьюдента для двусторонней доверительной вероятности $P = 1 - \alpha$ (прил. 2, табл. П2.1).

Решение о принятии гипотезы принимается на основе верности неравенства $|t| < t_{\text{кр}}$ (аналогично п. 4). Если неравенство верно – нулевая гипотеза $H_0 : m_1 = m_2$ принимается – значит выборки принадлежат генеральным совокупностям с одинаковыми средними значениями.

Пример 13. Расходы сырья x_i и y_i на единицу продукции при использовании старой и новой технологий приведены в табл. 5.2. Предполагается, что генеральные совокупности X и Y имеют нормальные распределения со средними значениями m_1, m_2 Требуется проверить гипотезу $H_0: m_1 = m_2$ против $H_1: m_1 \neq m_2$. Доверительная вероятность $P = 90$ %.

Таблица 5.2

Расходы сырья на единицу продукции

	По старой технологии				По новой технологии				
Расход сырья	x_i	304	307	308	y_i	303	304	306	308
Число изделий	n_x	1	4	4	n_y	2	6	4	1

Решение

По данным табл. 5.2 следует вычислить средние значения (формула (4.5)) и выборочные среднеквадратические отклонения (формула (4.14)) для каждой выборки:

$$\overline{x} = \frac{1}{n_x}\sum_{i=1}^{3} x_i n_x = 307{,}11;\ S_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{3}\left(x_i - \overline{x}\right)^2 = 1{,}1610;$$

$$\overline{y} = \frac{1}{n_y}\sum_{i=1}^{4} y_i n_y = 304{,}77;\ S_y^2 = \frac{1}{n_y - 1}\sum_{i=1}^{4}\left(y_i - \overline{y}\right)^2 = 2{,}192.$$

Так как о значениях генеральных дисперсий ничего не известно, следует предположить, что $\sigma_x^2 \neq \sigma_y^2$ (строго говоря, выдвинутую гипотезу о равенстве дисперсий нужно проверять (разд. 5.2.2 п. 2)).

Тогда для проверки гипотезы используется статистика, подчиняющаяся распределению Стьюдента (формула 5.6):

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}} = \frac{307{,}11 - 304{,}77}{\sqrt{\dfrac{1{,}610}{9} + \dfrac{2{,}192}{13}}} = 4{,}29.$$

По таблице критических точек для распределения Стьюдента (прил. 2, табл. П2.1) при $P = 0{,}90$ для числа степеней свободы $k = n_1 + n_2 - 2 = 9 + 13 - 2 = 20$ критическое значение $t_{0{,}90;\,20} = 1{,}725$.

Так как неравенство $|t| < t_{\text{кр}}$ не выполняется, то гипотеза H_0 отвергается.

Ответ: гипотеза H_0 отвергается $(H_0 -)$, т. е. при переходе на новую технологию происходит изменение среднего расхода сырья.

Задачи для самостоятельного решения

Задача 1. Учет времени сборки узла машины бригадой из 10 слесарей показал, что среднее время (в минутах) сборки узла $\overline{x} = 76$, а исправленная выборочная дисперсия S^2 = 15. Предполагая распределение времени сборки нормальным, проверить на уровне значимости α = 0,1 гипотезу о том, что 75 мин. является нормативом (мат. ожиданием) трудоемкости сборки узла машины $(H_0 : m = m_0)$.

Задача 2. По техническим условиям средняя прочность троса составляет 2000 кг. В результате испытаний 20 кусков троса было установлено, что средняя прочность на разрыв равна 1955 кг при средней ошибке 25 кг. Удовлетворяет ли образец троса техническим условиям $(H_0 : m = m_0)$? Доверительная вероятность P = 0,95.

Задача 3. Для проверки эффективности нового лекарства были отобраны две случайные группы по 15 человек, страдающих гриппом. При применении старого лекарства средний срок выздоровления составлял 11 дней с выборочной дисперсией $S_x^2 = 3$, при применении нового – срок выздоровления составил 8 дней с $S_y^2 = 4$. Проверить при доверительной вероятности P = 0,99 гипотезу о преимуществе нового лекарства ($H_0 : m_1 = m_2$, дисперсии генеральных совокупностей неизвестны).

Задача 4. В двух фирмах, выпускающих детское питание, производилась оценка качества продукции. В фирме A, где проверялось 30 единиц продукции, средняя сумма баллов оказалась равной 52. Во второй фирме B проверялось 36 единиц продукции и их средняя сумма баллов оказалась равной 47. Считая дисперсию балльной оценкой, равной σ^2 = 12, определить на уровне значимости α = 0,05, какая фирма выпускает лучшую продукцию ($H_0 : m_1 = m_2$ при известной дисперсии генеральных совокупностей).

Задача 5. Для проверки новой технологии были выбраны две группы рабочих численностью n_1 = 40 и n_2 = 50 человек. В первой группе при применении старой технологии средняя выработка составила m_1 = 85 (изделий), во второй, где применялась новая технология, – m_2 = 95. Дисперсии по группам $\sigma_1^2 = 100$, $\sigma_2^2 = 75$ были известны заранее. Выяснить на уровне значимости α = 0,05 влияние новой технологии на производительность ($H_0 : m_1 = m_2$, дисперсии генеральных совокупностей известны).

Задача 6. Автомат, работающий со стандартным отклонением $\sigma = 1{,}5$ г, фасует чай в пачке со средним весом $m = 80$ г. В случайной выборке объема n = 16 средний вес в пачке составил $\bar{x} = 78{,}5$ г. Надо ли отрегулировать автомат ($H_0 : m = 80$)? Доверительная вероятность P = 99 %.

Задача 7. Станок, работающий со стандартным отклонением $\sigma = 0{,}4$ мм, производит детали, средняя длина которых составляет $m = 30$ мм. В случайной выборке объема $n = 25$ средняя длина деталей $\bar{x} = 30{,}1$ мм. Надо ли отрегулировать станок ($H_0 : m = 30$)? Доверительная вероятность P = 95 %.

Задача 8. Производитель утверждает, что средний вес плитки шоколада не меньше $m = 50$ г. Инспектор отобрал 10 плиток шоколада и взвесил. Их вес оказался 49, 50, 51, 52, 48, 47, 49, 52, 48, 51 г соответственно. Не противоречит ли это утверждению производителя ($H_0 : m_1 = m_2$, дисперсии генеральных совокупностей неизвестны, но равны между собой)? Доверительная вероятность P = 95 %.

Задача 9. С помощью двух измерительных установок проведены две серии измерений: $n_x = 10$, $n_y = 8$ (табл. 5.3).

Таблица 5.3

Результаты серии измерений

	Напряжения, В									
Серия измерения 1, x	10,5	12,6	11,4	9,8	10,8	12,1	11,9	10,9	10,2	11,6
Серия измерения 2, y	12,7	11,6	9,7	10,9	11,2	10,4	10,8	11,3	–	–

Требуется при уровне значимости α = 0,05 определить, значимо ли различаются между собой результаты двух серий измерений. Предполагается, что генеральные дисперсии между собой не равны ($H_0 : m_1 = m_2$, дисперсии генеральных совокупностей неизвестны).

5.2.2. Сравнение дисперсий

Сравнение выборочной дисперсии с гипотетической дисперсией нормальной совокупности

Исходные данные: по выборке объема n получено значение выборочной дисперсии S^2. Значение дисперсии генеральной совокупности σ^2 известно.

Требуется по исправленной дисперсии при заданном уровне значимости проверить нулевую гипотезу, состоящую в том, что генеральная дисперсия рассматриваемой совокупности равна гипотетическому значению σ_0^2.

Формализация гипотезы. Гипотеза $H_0 : \sigma = \sigma_0$; альтернативная гипотеза $H_1 : \sigma \neq \sigma_0$. При этом математическое ожидание m может быть произвольным.

В этом случае статистика примет вид

$$\psi = \frac{(n-1)S^2}{\sigma_0^2}. \qquad (5.7)$$

Статистика ψ подчиняется χ^2-распределению (распределение Пирсона) при доверительной вероятности $P = 1 - \alpha$ для числа степеней свободы $k = n - 1$, поэтому критическое значение $\psi_{\text{кр}} = \psi_{P;k}$ (прил. 2, табл. П2.2).

Решение о принятии гипотезы принимается на основе верности неравенства $|\psi| \le \psi_{\text{кр}}$. Если неравенство верно, нулевая гипотеза ($H_0 : \sigma = \sigma_0$) принимается, значит, выборка принадлежит генеральной совокупности с заданной дисперсией.

Пример 14. Точность работы станка-автомата проверяется по дисперсии σ^2 контролируемого размера изделий, которая не должна превышать 0,15. По данным из 25 отобранных изделий вычислена выборочная дисперсия $\overline{S}^2 = 0,25$. Необходимо определить, обеспечивает ли станок требуемую точность. Доверительная вероятность P = 90 %.

Решение

Предполагается, что размер изделия – нормальная случайная величина. На основе условия задачи формализуется нулевая гипотеза $H_0 : \sigma^2 = 0,15$ против альтернативной $H_1 : \sigma^2 > 0,15$.

Для проверки гипотезы используется статистика, подчиняющаяся χ^2-распределению (формула (5.7)):

$$\psi = \frac{(n-1)S^2}{\sigma_0^2} = \frac{24 \cdot 0,25}{0,15} = 40.$$

По таблице критических точек для χ^2-расперделения (распределения Пирсона) (прил. 2, табл. П2.2) при $P = 0,90$ для числа степеней свободы $k = n - 1 = 25 - 1 = 24$ критическое значение $\psi_{0,90;\,24} = 36,4$.

Так как неравенство $|\psi| \le \psi_{\text{кр}}$ не выполняется, то гипотеза H_0 отвергается.

Ответ: гипотеза H_0 отвергается (H_0-), т. е. станок требуемую точность не обеспечивает, требуется настройка.

Сравнение выборочных дисперсий для двух различных выборок (критерий Фишера)

Исходные данные: две выборки объемами n_1 и n_2, по их значениям определяются выборочные дисперсии S_1^2 и S_2^2.

Требуется проверить гипотезу о равенстве дисперсий каждой выборки между собой.

Формализация гипотезы. Гипотеза $H_0 : \sigma_1^2 = \sigma_2^2$; альтернативная гипотеза $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Известно, что величина $\left(\dfrac{S_1^2}{\sigma_1^2}\right) \Big/ \left(\dfrac{S_2^2}{\sigma_2^2}\right)$ имеет распределение Фишера. В условиях справедливости нулевой гипотезы $\sigma_1^2 = \sigma_2^2$ статистика критерия Фишера принимает вид

$$F = \frac{S_1^2}{S_2^2}, \tag{5.8}$$

где S_1^2 – б**о**льшая из двух сравниваемых дисперсий.

Статистика F подчиняется F-распределению (распределению Фишера) с $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ степенями свободы (прил. 2, табл. П2.3) при доверительной вероятности $P = 1 - \alpha$.

Решение о принятии гипотезы принимается на основе верности неравенства $F \le F_{P;\,k_1;\,k_2}$. Если неравенство верно, нулевая гипотеза $(H_0 : \sigma_1^2 = \sigma_2^2)$ принимается, значит, выборки принадлежат генеральным совокупностям с одинаковыми дисперсиями.

Пример 15. По независимым данным объемами $n_1 = 31$ и $n_2 = 25$ вычислены выборочные дисперсии $S_1^2 = 25{,}0$ и $S_2^2 = 16{,}0$, Необходимо проверить гипотезу о равенстве дисперсий исходных совокупностей на уровне значимости $\alpha = 0{,}01$.

Решение

Для сравнения дисперсий применяется критерий Фишера. Так как $S_1^2 > S_2^2$, то формула (5.8) примет вид

$$F = \frac{S_1^2}{S_2^2} = \frac{25{,}0}{16{,}0} = 1{,}563.$$

По таблице критических точек для распределения Фишера (прил. 2, табл. П2.3) при уровне значимости $\alpha = 0{,}01$ (доверительной вероятности $P = 0{,}99$) для числа степеней свободы $k_1 = n_1 - 1 = 31 - 1 = 30$, $k_2 = n_2 - 1 = 25 - 1 = 24$ критическое значение $F_{0,99;\,30;\,24} = 2{,}47$.

Так как неравенство $F < F_{0,99;\,30;\,24}$ выполняется, то гипотеза о равенстве дисперсий принимается.

Ответ: гипотеза H_0 принимается, т. е. выборки принадлежат совокупностям с одинаковыми дисперсиями.

Задачи для самостоятельного решения

Задача 1. Из нормальной генеральной совокупности извлечена выборка объема $n = 17$ и по ней найдена исправленная выборочная дисперсия $S^2 = 0{,}24$. Требуется при уровне значимости α = 0,05 проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 0{,}18$.

Задача 2. По двум независимым выборкам, объемы которых $n_x = 9$ и $n_y = 16$, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии $S_x^2 = 34{,}02$ и $S_y^2 = 12{,}05$. При уровне значимости α = 0,01 проверить нулевую гипотезу о равенстве генеральных дисперсий ($H_0 : \sigma_x^2 = \sigma_y^2$).

Задача 3. Для сравнения случайных погрешностей измерения электрического сопротивления мостов на разных диапазонах необходимо сравнить дисперсии. Были произведены измерения: на одном диапазоне число измерений $n_1 = 14$ и на втором – $n_2 = 10$. По данным выборок найдены исправленные выборочные дисперсии $S_1^2 = 0{,}84$ и $S_2^2 = 2{,}52$. При уровне значимости $\alpha = 0{,}1$ проверить нулевую гипотезу о равенстве генеральных дисперсий ($H_0 : \sigma_x^2 = \sigma_y^2$).

Задача 4. Для сравнения точности двух измерительных приборов каждым из них были произведены многократные измерения и получены два ряда наблюдений (табл. 5.4).

Таблица 5.4

Результаты измерений

Прибор x	1,08	1,10	1,12	1,14	1,15	1,15	1,13
Прибор y	1,11	1,12	1,35	1,18	1,22	1,35	

При уровне значимости α = 0,01 необходимо проверить гипотезу $H_0 : \sigma_x^2 = \sigma_y^2$.

Задача 5. Точность работы двух станков оценивалась отклонениями от номинала производимой продукции. Для 10 единиц продукции первого станка отклонение от номинала (выборочная дисперсия) составило $S_1^2 = 3{,}5$, для 15 единиц продукции второго станка – $S_2^2 = 4{,}5$. Можно ли считать на уровне значимости α = 0,05, что станки имеют одинаковую точность ($H_0 : \sigma_1^2 = \sigma_2^2$)?

Задача 6. Номинальная точность прибора равна σ_0 = 2 мкм. По 10 замерам детали была получена выборочная дисперсия показаний прибора, равная $S^2 = 0{,}9$. Для доверительной вероятности P = 0,9 проверить гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 0{,}9$.

Задача 7. При проверке размеров подшипников из двух партий по 10 штук в каждой, поставленных разными заводами, были обнаружены отклонения от номинала, характеризуемые выборочными дисперсиями $S_1^2 = 9$, $S_2^2 = 8{,}5$. Можно ли считать при уровне значимости α = 0,1 одинаковой точность изготовления подшипников разными заводами ($H_0 : \sigma_1^2 = \sigma_1^2$)?

5.3. Критерии согласия

При проверке гипотезы о предполагаемом законе неизвестного распределения используется специальная случайная величина – *критерий согласия*.

Критерий согласия – статистический критерий проверки гипотезы о предполагаемом законе неизвестного распределения (непараметрическая гипотеза). В математической статистике используются несколько критериев согласия: λ-критерий Колмогорова, χ^2-критерий Пирсона и другие [5, 10, 11].

5.3.1. λ-критерий Колмогорова

λ-критерий Колмогорова связан с расчетом эмпирических функций распределения (см. разд. 3.6) и применяется для проверки гипотезы о распределении непрерывной случайной величины. При его использовании сравниваются эмпирическая функция распределения $F^*(x)$ и предполагаемая функция распределения $F(x)$.

Алгоритм проведения проверки:

1. Произвести выборку объемом $n \geq 50$.
2. Выдвинуть гипотезу H_0 о предполагаемом законе распределения и определить теоретическую функцию распределения исходя из предположения, что параметры распределения известны заранее.
3. Найти эмпирическую функцию распределения $F^*(x)$ на основе полученной выборки.
4. Для предполагаемой функции распределения $F(x)$ по данным выборки x_i определить значения теоретической функции распределения $F(x_i)$.
5. Вычислить значение статистики:

$$\lambda = \sqrt{n} \cdot \max_{x_i} \left| F(x_i) - F^*(x_i) \right|. \tag{5.9}$$

6. Найти значение критической точки $\lambda_{кр}$ в зависимости от заданного уровня значимости по таблице значений для критерия Колмогорова (прил. 2, табл. П2.4).
7. Проверить верность неравенства $\lambda < \lambda_{кр}$ и принять решение о верности выдвинутой гипотезы. Если неравенство верно, то гипотеза принимается, т. е. различия между эмпирическим и предполагаемым теоретическим распределениями несущественны. В противном случае гипотеза отвергается, т. е. различия между эмпирическим и предполагаемым теоретическим распределениями существенны.

В критерии Колмогорова параметры закона распределения $F(x)$ считаются известными заранее, так что, заменяя их выборочными значениями, всегда привносится некоторая ошибка.

Для реализации λ-критерия Колмогорова удобно использовать расчётные таблицы, форма которых зависит от используемой выборки. Если выборка представлена в форме дискретного вариационного ряда (разд. 3.5.1, табл. 3.1), то результаты расчета записываются в табл. 5.5.

Таблица 5.5

λ-критерий Колмогорова. Расчетная таблица для дискретного вариационного ряда

i	x_i	n_i	$W_i = \frac{n_i}{n}$	$F^*(x_i)$	$F(x_i)$	$\lvert F(x_i) - F^*(x_i) \rvert$

Если выборка представлена в форме интервального вариационного ряда (см. разд. 3.5.2, табл. 3.5), то результаты расчета записываются в табл. 5.6.

Таблица 5.6

λ-критерий Колмогорова. Расчетная таблица для интервального вариационного ряда

i	a_i	n_i	$W_i = \frac{n_i}{n}$	$F^*(a_i)$	$F(a_i)$	$\lvert F(a_i) - F^*(a_i) \rvert$

Пример 1. Респондентам был задан вопрос о том, сколько порций мороженого они съедают за неделю. Результаты опроса приведены в табл. 5.7.

Таблица 5.7

Количество порций мороженого, съеденных за неделю

Число порций, x	1	2	3	4	5	6	7	8	9
Частота, n	10	11	8	9	12	10	13	15	12

Определить с помощью λ-критерия Колмогорова на уровне значимости $\alpha = 0{,}05$, согласуются ли данные выборки с равномерным распределением на отрезке [0, 10].

Решение

Выдвигаются нулевая и конкурирующая гипотезы:

H_0: различия между эмпирическим и предполагаемым равномерным распределением с параметрами a = 0, b = 10 несущественны.

H_1: различия между эмпирическим и предполагаемым равномерным распределением с параметрами a = 0, b = 10 существенны.

Функция распределения для равномерного закона имеет вид (формула 2.26)

$$F(x) = \begin{cases} 0, & x \le a; \\ \dfrac{x-a}{b-a}, & a < x \le b; \\ 1, & x > b. \end{cases}$$

Для данной задачи параметр a = 0, параметр b = 10, поэтому функция распределения примет следующий вид:

$$F(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{x}{10}, & 0 < x \le 10; \\ 1, & x > 10. \end{cases} \tag{5.10}$$

В процессе решения задачи необходимо заполнить табл. 5.8.

Значения первых двух столбцов берутся из условия задачи: столбец 1 – число порций мороженного x_i; столбец 2 – частота n_i. В столбец 3 записывается *относительная частота событий*, рассчитанная по формуле (3.2). Каждое значение столбца 4 равно сумме значения из этой же строки столбца 3 и значения из предыдущей строки столбца 4 (значение эмпирической функции распределения (формула (3.8)).

Каждое значение столбца 1 подставляется в формулу (5.10), и результат (значение теоретической функции распределения) записывается в столбец 5. Значение в столбце 6 – это результат разности значений столбцов 4 и 5, взятый по модулю.

На основе анализа столбца 6 определяется наибольшее значение: $\max\limits_{x_i}\left|F(x_i)-F_n(x_i)\right|=0{,}12$.

Таблица 5.8

Результаты вычисления

x_i	n_i	$W_i=\dfrac{n_i}{n}$	$F^*(x_i)$	$F(x_i)=0{,}1x_i$	$\left\|F(x_i)-F^*(x_i)\right\|$
1	2	3	4	5	6
1	10	0,10	0	0,1	0,1
2	11	0,11	0,10	0,2	0,1
3	8	0,08	0,21	0,3	0,09
4	9	0,09	0,29	0,4	0,11
5	12	0,12	0,38	0,5	0,12
6	10	0,10	0,50	0,6	0,1
7	13	0,13	0,60	0,7	0,1
8	15	0,15	0,73	0,8	0,07
9	12	0,12	0,88	0,9	0,02
10	0	0	1,00	1,00	0
Сумма	n = 100				max = 0,12

Тогда статистика (формула (5.9)) примет вид

$$\lambda=\sqrt{n}\cdot\max_{x_i}\left|F(x_i)-F_n(x_i)\right|=\sqrt{100}\cdot 0{,}12=1{,}12.$$

При уровню значимости $\alpha=0{,}05$ по таблице значений для критерия Колмогорова (прил. 2, табл. П2.4) критическая точка $\lambda_{\text{кр}}=1{,}358$.

Так как неравенство $\lambda<\lambda_{\text{кр}}$ (1,12 < 1,358) выполняется, то гипотеза H_0 на уровне значимости $\alpha=0{,}05$ принимается.

Ответ: гипотеза H_0 принимается, т. е. заданная выборка (табл. 5.8) согласуется с равномерным распределением на отрезке [0, 10].

Критерий Колмогорова достаточно часто применяется на практике благодаря своей простоте. Однако его применение возможно лишь тогда, когда теоретическая функция распределения $F(x)$ задана полностью. Но такой случай на практике встречается весьма редко.

Обычно из теоретических соображений известен лишь вид функции распределения, а ее параметры определяются по эмпирическим данным. При использовании оценок параметров неизвестных значений параметра может получиться завышенное значение вероятности $P(\lambda)$, а значит, большее критическое значение $\lambda_{кр} = \lambda$. В результате есть риск принять гипотезу о законе распределения случайной величины за правдоподобную, в то время как, на самом деле, она противоречит опытным данным.

5.3.2. χ^2-критерий Пирсона

На практике вычисление статистики λ для критерия Колмогорова – трудоемкая задача, поэтому часто применяют другой критерий, называемый χ^2-критерием. Его можно использовать для любых законов распределений.

χ^2-критерий Пирсона связан с расчетом теоретических частот.

Теоретические (выравнивающие) частоты – это частоты n_i', которые (в отличие от фактически наблюдаемых эмпирических частот) найдены путем вычислений по предполагаемой теоретической функции распределения.

Теоретические частоты n_i' рассчитываются по формуле

$$n_i' = n \cdot p_i, \tag{5.11}$$

где n – объем выборки; p_i – вероятность появления наблюдаемого значения x_i, рассчитанная при условии, что случайная величина X подчиняется предполагаемому закону распределения вероятностей.

Для реализации χ^2-критерия Пирсона выборочные данные необходимо представлять в виде интервального вариационного ряда (разд. 3.5.2, табл. 3.5).

Алгоритм проведения проверки:

1. Произвести выборку.
2. На основе выборке построить гистограмму.
3. По внешнему виду гистограммы предположить закон распределения случайной величины (вдвинуть нулевую гипотезу H_0).

4. Произвести оценку параметров предполагаемого закона распределения (получить точечные оценки (см. разд. 4.3)).

5. Записать предполагаемую функцию распределения $F(x)$ с учетом рассчитанных параметров данного закона.

6. Произвести расчет значений функции распределения для границ интервалов $F(a_i)$. Для удобства результаты расчета следует представлять в виде табл. 5.9.

Таблица 5.9

χ^2-критерий Пирсона. Расчетная таблица

i	a_i	$F(a_i)$	a_{i+1}	$F(a_{i+1})$	$p_i = F(a_i) - F(a_{i+1})$	n_i	$C_i = \frac{(n_i - np_i)^2}{np_i}$

Примечание: n_i – частоты попадания в i-й интервал значений выборки; np_i – это теоретические частоты.

7. Вычислить значение статистики:

$$\chi = \sum_{i=1}^{k} C_i = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i}. \tag{5.12}$$

8. Найти значение критической точки $\chi_{\text{кр}}$ по таблице значений для χ^2-распределения (прил. 2, табл. П2.2) в зависимости от уровня значимости α и числа степеней свободы k.

Число степеней свободы находят по формуле

$$k = s - r - 1, \tag{5.13}$$

где s – число интервалов вариационного ряда; r – число параметров предполагаемого закона распределения.

Например, если предполагаемый закон распределения – нормальное распределение, то оцениваются два параметра – математическое ожидание m и среднее квадратичное отклонение σ – и тогда число степеней свободы $k = s - r - 1 = s - 3$.

9. Проверить верность неравенства $\chi < \chi_{\text{кр}}$ и принять решение о верности выдвинутой гипотезы. Если неравенство верно, то гипотеза принимается, выборка соответствует предполагаемому закону распределения с рассчитанными параметрами. В противном случае гипотеза отвергается, т. е. выборка не соответствует предполагаемому закону распределения с рассчитанными параметрами.

Поскольку статистика χ имеет χ^2-распределение (распределение Пирсона) при $n \to \infty$, то:

1) критерий Пирсона следует применять только при больших n ($n > 30$);

2) необходимо, чтобы в каждом интервале вариационного ряда было достаточное количество наблюдений (не менее 5). Если в каком-нибудь интервале число наблюдений меньше пяти, имеет смысл объединить соседние интервалы. Поэтому при вычислении числа степеней свободы в качестве величины s берется соответственно уменьшенное число интервалов [12].

Пример 2. Погрешности 500 результатов измерений дальности до цели радиодальномером приведены в табл. 5.10. Пользуясь критерием Пирсона с уровнем значимости $\alpha = 0{,}01$, проверить согласованность теоретического и экспериментального распределений.

Таблица 5.10

Дальность до цели

Интервал погрешности измерения дальности Δ, м	–25; –15	–15; –5	–5; 5	5; 15	15; 25
Количество погрешностей в интервале n_i	50	130	200	100	20

Решение

По условию задачи число интервалов $s = 5$, а длина интервалов $\Delta_i = 10$ м. Для визуальной оценки следует построить вариационный ряд распределения в виде гистограммы (рис. 5.3).

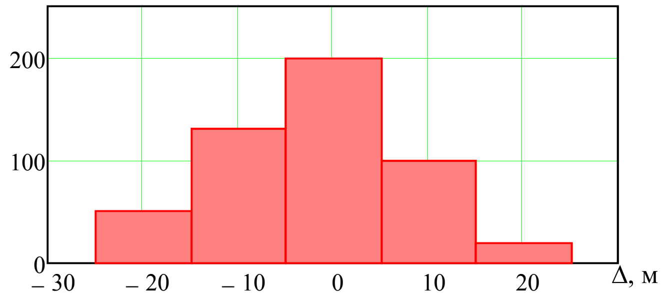


Рис. 5.3. Распределение результатов измерения дальности до цели

По внешнему виду гистограммы следует предположить, что экспериментальное распределение подчиняется нормальному закону распределения.

Формализация гипотезы

H_0: распределение результатов измерений дальности до цели подчиняется нормальному закону распределения.

H_1: распределение результатов измерений дальности до цели не подчиняется нормальному закону распределения.

Параметрами нормального распределения являются математическое ожидание m и среднеквадратическое отклонение σ. Оценить параметры можно, используя формулы (4.4) и (4.12) соответственно.

В качестве значений вариант взяты средние значения интервалов из табл. 5.10.

$$\overline{x} = \frac{1}{500}(-20 \cdot 50 - 10 \cdot 130 + 0 \cdot 200 + 10 \cdot 100 + 20 \cdot 20) = -1{,}8 \text{ м};$$

$$S^2 = \frac{1}{500}\sum_i \Delta_i^2 \cdot n_i - (\overline{x})^2 =$$

$$= \frac{1}{500}((-20)^2 \cdot 50 + (-10)^2 \cdot 130 + 0^2 \cdot 200 + 10^2 \cdot 100 + 20^2 \cdot 20) - (1{,}8)^2 =$$

$$= 98{,}76 \text{ м}^2;$$

$$S = \sqrt{S^2} = 9{,}93 \text{ м}.$$

Для вычисления значений функции распределения следует записать формулу с учетом рассчитанных значений параметров (формула (2.31)):

$$F(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

В качестве оценки мат. ожидания используется $\overline{x} = -1{,}8$ м, а среднеквадратического отклонения – $S = 9{,}93$ м.

Тогда

$$F(x) = \Phi\left(\frac{x+1{,}8}{9{,}93}\right).$$

Значение функции распределения для каждого значения x вычисляется по таблице значений интегральной функции Лапласа (прил. 2, табл. П2.2).

Все результаты вычисления сводятся в табл. 5.11.

Таблица 5.11

Расчетная таблица для примера 2

i	a_i	$F(a_i)$	a_{i+1}	$F(a_{i+1})$	$p_i = F(a_i) - F(a_{i+1})$	n_i	$C_i = \frac{(n_i - np_i)^2}{np_i}$
1	–25	0,0107	–15	0,0968	0,086	50	1,122
2	–15	0,0968	–5	0,3821	0,285	130	1,122
3	–5	0,3821	5	0,7580	0,376	200	0,773
4	5	0,7580	15	0,9554	0,197	100	0,017
5	15	0,9554	25	0,9965	0,041	20	0,015

Значение статистики рассчитывается по формуле (5.12):

$$\chi = \sum_{i=1}^{k} C_i = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} = 3{,}048.$$

Значение критической точки $\chi_{кр}$ находится по таблицы значений для χ^2-распределения (прил. 2, табл. П2.2) по уровню значимости $\alpha = 0,01$ и числа степеней свободы k (формула (5.13)).

Число параметров для нормального распределения $r = 2$, число интервалов по условию задачи $s = 5$. Тогда число степеней свободы $k = s - r - 1 = 5 - 2 - 1 = 2$.

$$\chi_{кр}(k, \alpha) = \chi_{кр}(2, 0,01) = 9,21.$$

Неравенство $\chi = 3,048 < \chi_{кр} = 9,21$ выполняется, следовательно, гипотеза принимается ($H_0 +$).

Ответ: гипотеза H_0 принимается, т. е. представленная выборка (табл. 5.11) соответствует нормальному распределению.

Задачи для самостоятельного решения

Задача 1. Значения наработки τ (в месяцах) на отказ некоторой технической системы приведены в табл. 5.12.

Таблица 5.12

Наработка на отказ. Данные выборки

0,173	0,748	0,393	0,638	1,669	0,720	1,089	1,139	0,784	0,414
0,285	5,479	1,885	1,075	1,632	1,374	0,337	2,457	2,795	0,271
4,623	0,642	4,092	5,641	0,766	1,418	0,747	0,414	0,946	2,132
0,384	1,281	0,909	0,226	0,697	0,943	0,583	0,925	1,257	1,214
1,514	0,251	1,138	0,904	0,337	0,338	1,374	1,443	0,371	1,697

Проверить критерий согласия выборки с теоретическим законом с помощью критерия Пирсона, предполагая, что генеральная совокупность подчиняется логарифмически нормальному закону распределения. Доверительная вероятность $P = 90$ %.

Задача 2. В ОТК с точностью до 1 мм проведены измерения длины 100 деталей, изготовленных на автоматическом станке. В табл. 5.13 приведены отклонения от номинального размера, разбитые на интервалы, количество деталей, попавших в каждый интервал. Оценить с помощью критерия Пирсона гипотезу о согласии выборочного распределения с нормальным законом распределения. Доверительная вероятность $P = 95$ %.

Таблица 5.13

Логарифмы данных ln x для выборки

Интервал $[a_i; a_{i+1}]$	–20; –15	–15; –10	–10; –5	–5; 0	0; 5	5; 10	10; 15	15; 20
Частоты, n	5	6	8	12	25	18	13	8

Задача 3. В больнице скорой помощи фиксировалось количество X вызовов специализированных бригад в час. Наблюдения велись в течение 100 ч, их результаты приведены в табл. 5.14.

Таблица 5.14

Количество вызовов специализированных бригад в час

Количество вызовов, x	0	1	2	3	4	5	6	7
Частота вызовов, n	6	27	26	20	10	5	5	1

Проверить, подчиняются ли закону Пуассона представленные экспериментальные значения на уровне доверия $\alpha = 0{,}05$.

Задача 4. Используя критерий Колмогорова, проверить на уровне значимости $\alpha = 0{,}1$ гипотезу о том, что выборка (табл. 5.15) подчиняется нормальному закону распределения с параметрами $m = 25$; $\sigma = 10$.

Таблица 5.15

Количество вызовов специализированных бригад в час

Диапазон изменения параметра, x	0–10	10–20	20–30	30–40	40–50
Частота, n	23	24	11	9	3

Контрольные вопросы и задания

1. Что такое статистическая гипотеза? Приведите примеры нулевой и конкурирующей гипотез.
2. Что означают ошибки 1-го и 2-го рода при проверке статистических гипотез?
3. Что такое непараметрические и параметрические гипотезы?
4. Приведите порядок проверки любой гипотезы.
5. Какие гипотезы рассматриваются при сравнении средних; при сравнении дисперсий?
6. Покажите, какая статистика используется при проверке гипотез в следующих случаях:
   а) сравнение выборочной средней с гипотетической генеральной средней нормальной совокупности при известной и неизвестной дисперсиях (гипотеза о генеральной средней нормального распределения);

б) сравнение двух средних генеральных совокупностей с известными и неизвестными дисперсиями (гипотеза о равенстве генеральных средних двух распределений);
в) сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности (гипотеза о генеральной дисперсии нормального распределения);
г) сравнение двух дисперсий нормальных генеральных совокупностей (гипотеза о равенстве генеральных дисперсий двух нормальных распределений).

7. Поясните, что такое критерий согласия. Приведите алгоритм проверки непараметрической гипотезы:
а) λ-критерия Колмогорова;
б) χ^2-критерий Пирсона.

Глава 6. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

6.1. Функциональная, статистическая и корреляционная зависимости

Во многих задачах требуется установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других величин.

Две случайные величины могут быть связаны функциональной зависимостью, статистической (стохастической) зависимостью или быть независимыми [8].

Если между значениями имеется такая связь, что любому x_i соответствует только одно значение y_i, то такая связь *называется функциональной.* Графически функциональная связь двух величин представляется какой-то кривой $y = f(x)$ или, в частности, прямой линией $y = k \cdot x + b$ (рис. 6.1).

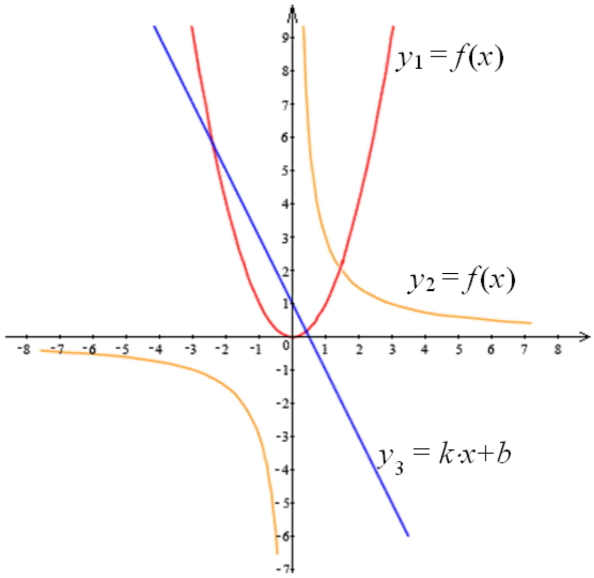


Рис. 6.1. Функциональная зависимость

Строгая функциональная зависимость реализуется редко, так как обе величины X и Y или одна из них подвержены воздействию случайных факторов. Эти факторы могут быть общими для обеих величин (под «общими» здесь подразумеваются такие факторы, которые воздействуют на СВ Y, и на СВ X). В этом случае возникает *статистическая зависимость.*

Пример 1. Величина Y зависит от случайных факторов Z_l, Z_2, V_1, V_2, а величина X зависит от случайных факторов Z_l, Z_2, U_1. Обе величины – X и Y – являются случайными, но так как имеются общие факторы Z_1 и Z_2, оказывающие влияние на обе величины, то значения X и Y обязательно будут взаимосвязаны. Связь носит вероятностный случайный характер, поэтому численное значение меняется от испытания к испытанию, но эта связь между значениями X и Y присутствует. При этом каждому значению X может соответствовать не одно значение Y, как при функциональной зависимости, а целое множество значений. Такая связь называется *статистической (стохастической)*.

Слово «стохастический» (от греческого слова στοχαστικός – «умеющий угадывать») означает «случайный».

Статистической (*стохастической*) называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой, в этом случае статистическую зависимость называют *корреляционной.* Корреляционная зависимость показана на рис. 6.2.

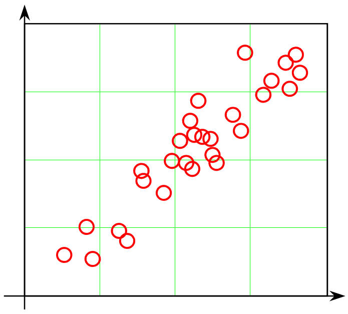
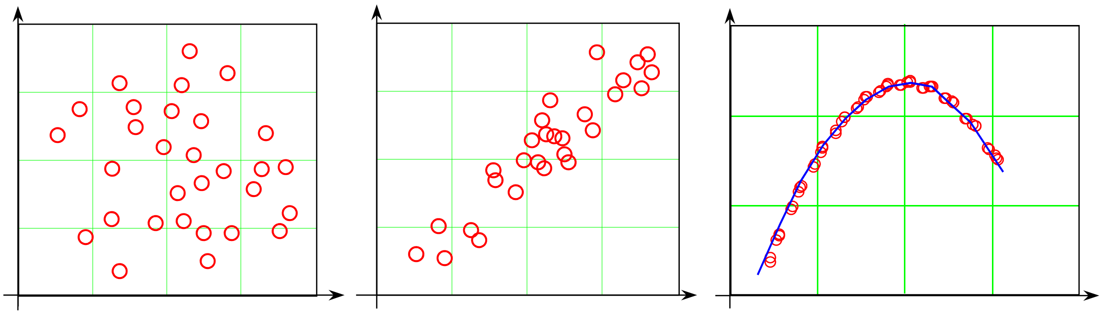


Рис. 6.2. Корреляционная зависимость

Пример 2. Величина Y – объем собранного зерна, величина X – количество внесенных за сезон удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т. е. величина Y не является функцией от величины X. Это объясняется влиянием случайных факторов (осадки, температура воздуха и др.). Однако очевидно, что *средний* урожай является функцией от количества удобрений, т. е. величина Y связана с величиной X *корреляционной зависимостью*.

На рис. 6.3 приведены разные виды связи между величинами X и Y. На рис. 6.3, *а* изменения величины Y вызваны случайными факторами и не связаны с величиной X, следовательно, величины X и Y независимы. На рис. 6.3, *б* показано, что величина Y изменяется не только от изменения величины X, но и под воздействием случайных факторов, следовательно, величины X и Y имеют статистическую связь.

При отсутствии случайной компоненты между величинами X и Y существует функциональная связь (рис. 6.3, *в*). Следует отметить, что строго функциональная зависимость на практике встречается крайне редко, так как при получении результатов исследования всегда присутствуют некоторые погрешности, которые являются случайными составляющими. В связи с этим незначительное отклонение результатов эксперимента от функциональной зависимости в расчет не берется. Поэтому зависимость, показанную на рис. 6.3, *в*, следует считать функциональной.



а) отсутствие зависимости *б) корреляционная зависимость* *в) функциональная зависимость*

Рис. 6.3. Возможные виды связей между величинами X и Y

6.2. Корреляционный анализ

Выявление статистической связи и оценка ее силы представляет собой важную и трудную задачу математической статистики. *Корреляционный анализ* (от лат. *correlatio* – «соотношение», «связь») – метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя случайными величинами.

Задача корреляционного анализа определить, существует ли статистически достоверная связь между двумя случайными величинами.

Пример 1. Корреляционный анализ применяется при попытке установить взаимосвязь между ростом и весом детей, взаимосвязь между успеваемостью и результатами выполнения теста IQ, между стажем работы и производительностью труда и т. д.

Если в исследуемой выборке между ростом и весом человека существует корреляционная зависимость, то это не означает, что вес является причиной роста человека. Иначе говоря, сброс лишних килограммов приводил бы к уменьшению роста человека (естественно, это не так). Корреляционная связь лишь говорит о взаимосвязанности данных параметров, причем в данной конкретной выборке, в другой выборке подобная зависимость может не наблюдаться.

Таким образом, корреляционная зависимость отражает только взаимосвязь между двумя случайными величинами и не говорит о причинно-следственных связях.

6.2.1. Коэффициент корреляции

Как известно из теории вероятностей, дисперсия суммы независимых случайных величин равна сумме их дисперсий [8, 10]:

$$D(x+y)=D(x)+D(y). \qquad (6.1)$$

Согласно формуле (2.14)

$$D(x)=M\left[\left(x-M\left[x\right]\right)^2\right];\ D(y)=M\left[\left(y-M\left[y\right]\right)^2\right].$$

Тогда формула (6.1) с учетом свойств математического ожидания примет вид

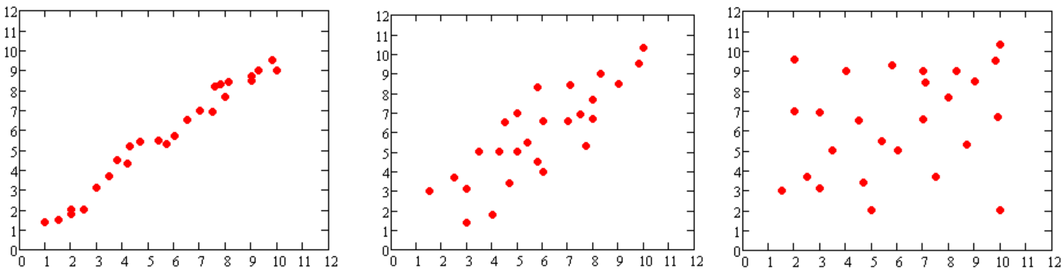
$$\begin{aligned} D(x+y)&=M\left[\left(x-M\left[x\right]\right)^2\right]+M\left[\left(y-M\left[y\right]\right)^2\right]= \\ &=M\left[\left((x+y)-M\left[x+y\right]\right)^2\right]= \\ &=D(x)+D(y)+2M\left[\left(x-M\left[x\right]\right)\cdot\left(y-M\left[y\right]\right)\right]. \end{aligned} \qquad (6.2)$$

Равенство (6.1) отличается от (6.2) наличием слагаемого $2M\left[\left(x-M\left[x\right]\right)\cdot\left(y-M\left[y\right]\right)\right]$. Это является признаком наличия корреляционной зависимости между величинами X и Y.

Следовательно, наличие зависимости между случайными величинами X и Y непосредственно вытекает из неравенства

$$M\left[\left(x-M\left[x\right]\right)\cdot\left(y-M\left[y\right]\right)\right]\neq 0.$$

Однако обратное утверждение не справедливо, т. е. при $M\left[\left(x-M\left[x\right]\right)\cdot\left(y-M\left[y\right]\right)\right]=0$ независимость величин X и Y не следует. Связь, при которой наблюдается отличие дисперсии суммы от суммы дисперсии, называется *корреляцией.*



а) сильная б) слабая в) корреляция отсутствует

Рис. 6.4. Сила корреляции между величинами X и Y

Для оценки степени корреляции нужно провести n испытаний, в каждом из которых регистрируются одновременно значения величин X и Y, и получить n пар значений (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$. Для наглядности эти пары значений можно отметить как координаты точек на плоскости. По внешнему виду полей корреляции можно сделать предварительный вывод о наличии и силе корреляции (рис. 6.4).

По внешнему виду можно примерно оценить силу взаимодействия величин X и Y. Для численного выражения силы связи следует ввести некоторую величину – *корреляционный момент* – величина, количественно зависящая от рассеяния аргументов величин X и Y:

$$K_{xy} = M\Big[\big(x - M[x]\big) \cdot \big(y - M[y]\big)\Big]. \qquad (6.3)$$

Корреляционный момент имеет размерность, зависящую от единиц измерения величин X и Y. На практике часто используется нормированная безразмерная величина ρ, называемая *коэффициентом корреляции:*

$$\rho = \frac{K_{xy}}{\sqrt{D(x) \cdot D(y)}} = \frac{K_{xy}}{\sigma_x \cdot \sigma_y}. \qquad (6.4)$$

Коэффициент корреляции определяется степенью тесноты корреляционной связи между двумя величинами – X и Y.

Основные свойства коэффициента корреляции ρ*:*

1. Коэффициент корреляции изменяется в диапазоне [–1; 1].
2. Если СВ X и Y независимы между собой, то коэффициент корреляции ρ = 0. (К сожалению, ρ = 0 для некоторых зависимых величин X и Y означает, что они *не коррелированы*). Для нормально распределенных величин некоррелированность одновременно означает и независимость.
3. Коэффициент характеризует силу связи: чем больше ρ по модулю, тем сильнее связь. При ρ = ±1 между величинами X и Y существует функциональная связь, но не любая, а *строгая линейная зависимость*.

В зависимости от значения коэффициента корреляции ρ можно ввести следующую градацию (одна из вариаций шкалы Чеддока [9, 10]), представленную в табл. 6.1.

Таблица 6.1

Значение коэффициента корреляции

Значение	**Интерпретация**
$\rho = 0$	Отсутствие корреляция
$0 < \lvert\rho\rvert \le 0{,}2$	Очень слабая корреляция
$0{,}2 < \lvert\rho\rvert \le 0{,}5$	Слабая корреляция
$0{,}5 < \lvert\rho\rvert \le 0{,}7$	Средняя корреляция
$0{,}7 < \lvert\rho\rvert \le 0{,}9$	Сильная корреляция
$0{,}9 < \lvert\rho\rvert \le 1$	Очень сильная корреляция

Таким образом, коэффициент корреляции показывает, насколько связь между величинами близка к строгой линейной зависимости. Коэффициент корреляции демонстрирует как долю случайности, так и степень близости к линейной зависимости.

Пример 2. На рис. 6.5 приведены различные виды поля корреляции. Прямая или кривая линия, проведенная через область значений (рис. 6.5, *а*, *б*, *в*), получила название *аппроксимирующей*.

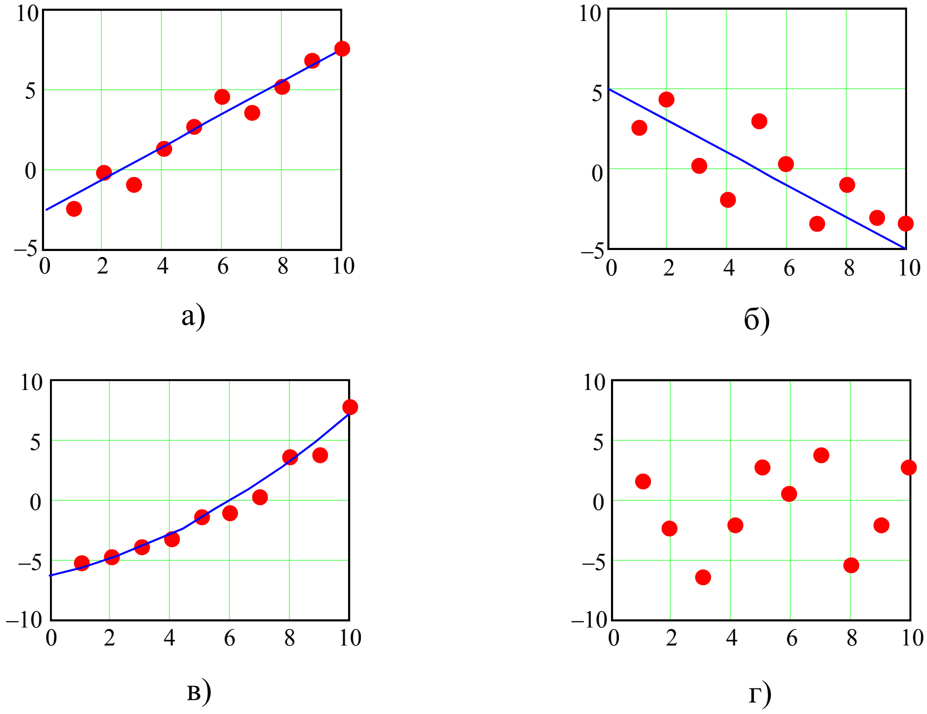


Рис. 6.5. Поля корреляции:
а – сильная положительная линейная корреляция; б – слабая отрицательная корреляция; в – сильная нелинейная корреляция; г – корреляции нет

Пример 3. Проведено исследование зависимости между среднемесячными доходами на семью (в тыс. у. е.) X и расходами на покупку кондитерских изделий (в у. е.) Y. Результаты представлены в табл. 6.2.

Таблица 6.2

Результаты исследования

x_i	4,8	3,8	5,4	4,2	3,4	4,6	3,4
y_i	75	68	78	71	64	73	66

Построить график зависимости $y = f(x)$ и сделать предварительный вывод о силе связи между среднемесячными доходами на семью и расходами на покупку кондитерских изделий.

Решение

Корреляционное поле, построенное по статистическим данным, приведено на рис. 6.6.

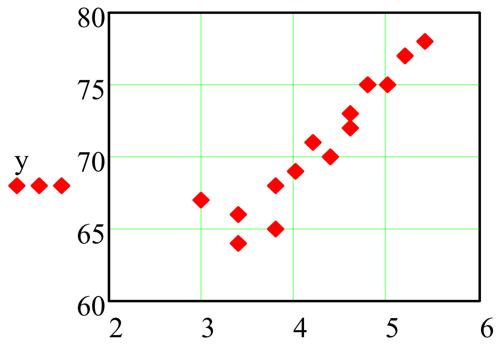


Рис. 6.6. График зависимости $y = f(x)$

Ответ: анализ внешнего вида зависимости (рис. 6.6) позволяет сделать вывод о том, что между среднемесячными доходами семьи и затратами на приобретение ею кондитерских изделий существует сильная линейная статистическая связь. При этом связь имеет положительную тенденцию, т. е. с ростом переменной X наблюдается увеличение отклика Y.

6.2.2. Оценивание коэффициента корреляции опытным путем

Анализ поля корреляции дает лишь общее представление о взаимосвязи величин, а для более детального изучения необходимо определить значение коэффициента корреляции. К сожалению, в распоряжении исследователя имеется только выборка из генеральной совокупности, состоящая из n пар значений (x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$. Оценку зависимости между случайными величинами можно провести только по данным имеющейся выборки.

Выборочный коэффициент корреляции r является оценкой коэффициента корреляции ρ и вычисляется по формуле (6.4), где вместо значений математических ожиданий и дисперсий берутся их оценки. После преобразования формула (6.3) для *выборочного корреляционного момента* примет вид

$$K_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}), \qquad (6.5)$$

а формула (6.4) для выборочного коэффициента корреляции –

$$r = \frac{1}{n-1}\frac{\sum_{i-1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{S_x \cdot S_y}, \qquad (6.6)$$

где оценки математических ожиданий и дисперсий рассчитываются по уже известным формулам (4.4) и (4.12). Для удобства использования в корреляционном и регрессионном анализе следует упомянутые формулы представить в несколько обновленном виде:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,;\ \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,; \tag{6.7}$$

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2;\ S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2. \tag{6.8}$$

Для практических вычислений удобнее использовать следующие формулы:

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n};$$

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}; \tag{6.9}$$

$$\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{n} y_i^{\,2} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}.$$

Пример 4. Для примера 3 провести корреляционный анализ (рассчитать выборочный коэффициент корреляции).

Решение

Выборочный коэффициент корреляции рассчитывается по формуле (6.6).

Для вычисления необходимо рассчитать оценки математических ожиданий и выборочных дисперсий для величин X и Y (формулы (6.7) и (6.8)). Расчет можно производить по указанным формулам или воспользоваться «удобными» формулами (6.9).

Все промежуточные результаты (для обоих случаев) удобно свести в расчетную таблицу (табл. 6.3).

Таблица 6.3

Расчетная таблица

								Сумма
x_i	4,8	3,8	5,4	4,2	3,4	4,6	3,4	29,6
y_i	75	68	78	71	64	73	66	495

								Сумма
$\overline{x}$	4,23							
$\overline{y}$	70,71							
$x_i - \overline{x}$	0,57	–0,43	1,17	–0,03	–0,83	0,37	–0,83	–0,01
$y_i - \overline{y}$	4,29	–2,71	7,29	0,29	–6,71	2,29	–4,71	0,03
$(x_i - \overline{x})(y_i - \overline{y})$	2,45	1,16	8,53	–0,01	5,56	0,85	3,91	22,46
$(x_i - \overline{x})^2$	0,33	0,18	1,37	0,00	0,69	0,14	0,69	3,39
$(y_i - \overline{y})^2$	18,37	7,37	53,08	0,08	45,08	5,22	22,22	151,43
S_x^2	0,57							
S_y^2	25,24							
r	0,99							

Ответ: выборочный коэффициент корреляции равен 0,99, что говорит о сильной положительной связи, близкой к строго линейной зависимости между среднемесячными доходами семьи и затратами на приобретение ею кондитерских изделий. Расчет подтвердил предположение, выдвинутое в *примере 3*.

Пример 5. Установить силу корреляционной связи, если известны результаты наблюдений (табл. 6.4).

Таблица 6.4

Результаты наблюдения

x	1,00	1,50	3,00	4,50	5,00
y	1,25	1,40	1,50	1,75	2,25

Решение

Предположить силу корреляции можно путем анализа поля корреляции. Для этого необходимо изобразить на плоскости результаты наблюдения (рис. 6.7).

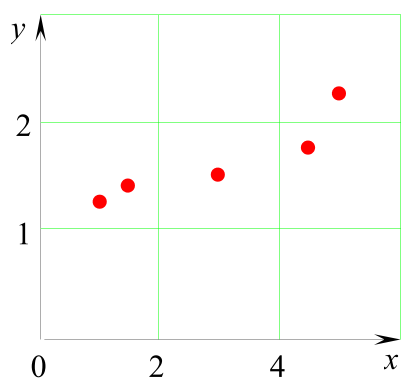


Рис. 6.7. Поле результатов наблюдения

Анализ рис. 6.7 показывает, что связь между величинами x и y присутствует, причем она достаточно сильная и близка к строго функциональной зависимости. Можно предположить, что значение коэффициента корреляции достаточно близко к 1 (r = 1).

Установить силу корреляции можно, рассчитав выборочный коэффициент корреляции по формуле (6.6). Для расчета необходимо выполнить промежуточные вычисления, используя формулы (6.7)–(6.9):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 3; \; \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 1{,}63;$$

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 3{,}125; \; S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = 0{,}15;$$

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = 2{,}525.$$

Полученные результаты подставляются в формулу (6.6) для вычисления выборочного коэффициента корреляции:

$$r = \frac{1}{n-1}\frac{\sum_{i-1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{S_x \cdot S_y} = \frac{2{,}525}{4 \cdot \sqrt{3{,}125} \cdot \sqrt{0{,}15}} = 0{,}912.$$

Ответ: выборочный коэффициент корреляции r = 0,912, что подтверждает сделанное на основе анализа поля корреляции предположение.

Пример 6. Данные по живой массе бычков при рождении X и последующая скорость роста Y приведены в табл. 6.5.

Таблица 6.5

Данные по бычкам

x_i, кг	40	42	35	36	45	47	40	43	41	38
y_i, г/сутки	1000	900	850	950	920	950	810	870	930	870

Построить график зависимости скорости роста бычка от его массы при рождении.

Рассчитать выборочный коэффициент корреляции.

Решение

График зависимости скорости роста бычка от его массы при рождении, построенный по данным условия задачи, приведен на рис. 6.8.

Анализ рис. 6.8 показывает, что связь между величинами x и y слабая. Можно предположить, что значение коэффициента корреляции близко к 0 (r = 1).

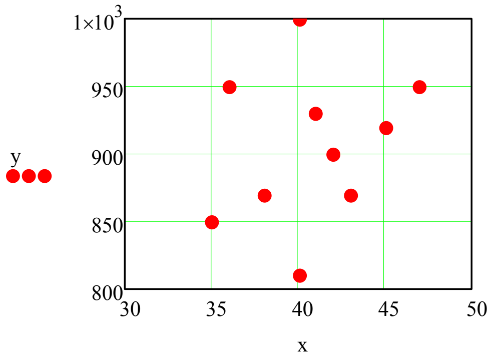


Рис. 6.8. График зависимости скорости роста бычка от его массы при рождении

Выборочный коэффициент корреляции рассчитывается по формуле (6.7).

Для удобства расчета можно воспользоваться формулами (6.9). Результаты промежуточные вычисления приведены в табл. 6.6.

Таблица 6.6

Результаты расчетов

Номер бычка (i)	x_i, кг	y_i, г/сутки	x_i^2	y_i^2	$x_i \cdot y_i$
1	40	1000	1600	1000000	40000
2	42	900	1764	810000	37800
3	35	850	1225	722500	29750
4	36	950	1296	902500	34200
5	45	920	2025	846400	41400
6	47	950	2209	902500	44650
7	40	810	1600	656100	32400
8	43	870	1849	756900	37410
9	41	930	1681	864900	38130
10	38	870	1444	756900	33060
Σ	**407**	**9050**	**16693**	**8218700**	**368800**
Среднее значение	**40,7**	**905**	–	–	–

В результате подстановки данных табл. 6.6 формулы (6.9) примут вид:

$$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)=\sum_{i=1}^{n}x_i y_i-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n}=368800-\frac{407\cdot 9050}{10}=465\,;$$

$$\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\sum_{i=1}^{n}x_i{}^2-\frac{\left(\sum\limits_{i=1}^{n}x_i\right)^2}{n}=16693-\frac{407^2}{10}=128{,}1;$$

$$\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2=\sum_{i=1}^{n}y_i{}^2-\frac{\left(\sum\limits_{i=1}^{n}y_i\right)^2}{n}=8218700-\frac{9050^2}{10}=28450\,.$$

Выборочные среднеквадратические отклонения рассчитываются по формулам (6.8):

$$S_x=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}=\sqrt{\frac{1}{10-1}\cdot128{,}1}=3{,}77\,;$$

$$S_y=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}=\sqrt{\frac{1}{10-1}\cdot28450}=56{,}22\,.$$

Тогда выборочный коэффициент корреляции равен:

$$r=\frac{1}{n-1}\frac{\sum\limits_{i-1}^{n}\left(x_i-\overline{x}\right)(y_i-\overline{y})}{S_xS_y}=\frac{1}{10-1}\cdot\frac{465}{3{,}77\cdot56{,}22}=0{,}24\,.$$

Ответ: выборочный коэффициент корреляции r = 0,24, что подтверждает сделанное на основе анализа поля корреляции предположение.

6.3. Проверка гипотезы о значимости выборочного коэффициента корреляции

При достаточно большом объеме выборочный коэффициент корреляции r приближается к истинному значению ρ. Но при небольшом объеме выборки полученная оценка неизбежно будет сопровождаться неизвестной погрешностью. Поэтому при ограниченном объеме выборки n необходимо проверить значимость отличия коэффициент корреляции r от нуля, т. е. проверить нулевую гипотезу (H_0: ρ = 0) о том, что генеральный коэффициент корреляции равен нулю (связь между величинами X и Y отсутствует).

В качестве критериев значимости при проверке нулевой гипотезы могут быть использованы r-критерий или t-критерий.

6.3.1. *r*-критерий

Проверка гипотезы о значимости коэффициента корреляции (H_0: ρ = 0) с помощью r-критерия осуществляется в следующем порядке:

1. По таблицам коэффициента корреляции r-распределения (прил. 2, табл. П2.5) определяется граничное значение критерия $r_{\text{гр}}$ по заданному уровню значимости α (доверительной вероятности P) и числу степеней свободы $k = n - 2$ (отнимается число 2, так как рассматриваются две величины – x и y).

2. Решение о принятии гипотезы принимается на основе верности неравенства $|r| < r_{\text{гр}}$. Если неравенство верно, нулевая гипотеза о равенстве нулю генерального коэффициента корреляции принимается ($H_0 +$) и считается, что случайные величины X и Y не коррелированы между собой. В противном случае гипотеза отвергается ($H_0 -$). Следовательно, между случайными величинами X и Y существует корреляционная связь.

6.3.2. *t*-критерий

Проверка гипотезы о значимости коэффициент корреляции (H_0: ρ = 0) с помощью t-критерия осуществляется в следующем порядке:

1. Вычисляется наблюдаемое значение критерия:

$$T_{\text{набл}} = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}}, \tag{6.10}$$

где r – выборочный коэффициент корреляции; n – объем выборки, количество пар значений (x_n, y_n).

2. По таблице критических точек для t-распределения (распределение Стьюдента) (прил. 2, табл. П2.1) при уровне значимости α (доверительной вероятности P) и числу степеней свободы $k = n - 2$ находится граничное значение критерия $t_{\text{кр}}$.

3. Решение о принятии гипотезы принимается на основе верности неравенства $|T_{\text{набл}}| < t_{\text{кр}}$. Если неравенство верно, нулевая гипотеза о равенстве нулю генерального коэффициента корреляции принимается ($H_0 +$): коэффициент корреляции r незначимо отличается от нуля, т. е. между случайными величинами отсутствует корреляционная связь. В противном случае случайные величины X и Y считаются коррелированными ($H_0 -$).

Пример 1. По выборке объема n = 122, извлеченной из двумерной совокупности, найден выборочный коэффициент корреляции r = 0,4.

При уровне значимости α = 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции ($H_0 : \rho = 0$).

Решение

Способ 1. t-критерий

Наблюдаемое значение критерия рассчитывается по формуле (6.10):

$$T_{\text{набл}} = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}} = 0,4\frac{\sqrt{122-2}}{\sqrt{1-0,4^2}} = 4,78.$$

По уровню значимости α = 0,05 и числу степеней свободы $k = 122 - 2 = 120$ по таблице критических точек для t-распределения (распределение Стьюдента) (прил. 2, табл. П2.1) определяется критическое значение: $t_{\text{кр}}(0,05;\ 120) = 1,98.$

Так как неравенство $|T_{\text{набл}}| < t_{\text{кр}}$ не выполняется, то гипотеза ($H_0 : \rho = 0$) отвергается.

Ответ: гипотеза H_0 отвергается (H_0 −), т. е. выборочный коэффициент корреляции r значимо отличается от нуля, следовательно, величины X и Y коррелированы.

Способ 2. r-критерий

По уровню значимости α = 0,05 и числу степеней свободы $k = 122 - 2 = 120$ по таблицам коэффициента корреляции r-распределения (прил. 2, табл. П2.5) определяется граничное значение критерия: $r_{\text{гр}}$ (0,05; 120) = 0,179.

Так как неравенство $|r| < r_{\text{гр}}$ не выполняется, то гипотеза ($H_0 : \rho = 0$) отвергается.

Ответ: гипотеза H_0 отвергается, т. е. выборочный коэффициент корреляции r значимо отличается от нуля, следовательно, величины X и Y коррелированы.

6.4. Регрессионный анализ

Корреляционный анализ одновременно оценивает факт случайности или причинности зависимости от фактора, но и одновременно степень неслучайности. В этом и состоит недостаток корреляционного анализа.

Регрессионный анализ является более высокой ступенью анализа по сравнению с корреляционным. Данный анализ позволяет оценить характер связи между величинами X и Y, получить количественные характеристики степени связи в виде математического описания зависимости среднего значения величины Y ($\overline{y}$) от случайной величины X . Эту зависимость называют *регрессией*, а кривую, описывающую эту зависимость, – *линией регрессии.* Термин «регрессия» впервые введен английским социологом Ф. Гальтоном и означает «воспроизводимость».

Таким образом, *регрессия* – это функция, которая позволяет по величине одной коррелируемой величины определить среднее значение другой величины.

При регрессионном анализе по парам экспериментальных данных (x_i, y_i) находится уравнение регрессии и оценивается адекватность полученного уравнения.

Таким образом, задача регрессии сводится к установлению уравнения, которое описывает экспериментальную зависимость между величинами X и Y, при наличии некоторой погрешности при получении экспериментальных данных. При проведении регрессионного анализа исследователь должен подобрать неизвестные коэффициенты аналитической зависимости $f(x)$.

В зависимости от вида уравнения различают следующие виды регрессии: линейная, полиномиальная, обобщённая.

6.4.1. Линейная регрессия

Линейная регрессия заключается в нахождении линии регрессии вида $y = \alpha + \beta \cdot x$.

Задача регрессионного анализа сводится к определению значения коэффициентов α и β по имеющимся экспериментальным данным (x_i, y_i) объема n.

Метод наименьших квадратов

Сущность метода заключается в том, чтобы подобрать коэффициенты для уравнения регрессии, при этом сумма квадратов отклонения экспериментальных данных от регрессии должна быть минимальной (рис. 6.9).

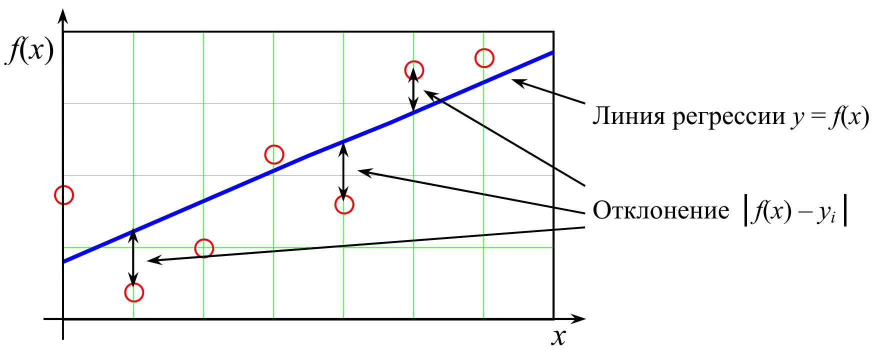


Рис. 6.9. Суть метода наименьших квадратов

Математические выкладки здесь не приводятся, с ними можно ознакомиться в литературе по математической статистике [8, 10].

В результате преобразований для метода наименьших квадратов получается система уравнений:

$$\begin{cases} n \cdot \alpha + \beta \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i; \\ \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \cdot y_i. \end{cases} \tag{6.11}$$

Необходимо определить значения коэффициента линейной регрессии β и свободного члена регрессии α.

В результате преобразований получаются следующие значения:

$$\beta = \frac{n \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \tag{6.12}$$

и

$$\alpha = \frac{\sum_{i=1}^{n} y_i - \beta \sum_{i=1}^{n} x_i}{n}. \tag{6.13}$$

Полученные коэффициенты полностью определяют линейную регрессию по заданной выборке.

С учетом формулы (6.7) выражение (6.13) можно представить в виде

$$\alpha = \overline{y} - \beta \cdot \overline{x}. \tag{6.14}$$

Отсюда уравнение регрессии:

$$\overline{y} = \alpha + \beta \cdot \overline{x}. \tag{6.15}$$

Таким образом, линия регрессии проходит через точку с координатами ($\overline{x}, \overline{y}$) – центр тяжести поля экспериментальных точек. Из этого следует, что для определения линии регрессии достаточно определить ее угловой коэффициент β. Это равнозначно тому, что исходные данные центрируются и начало координат переносится в точку ($\overline{x}, \overline{y}$). Тогда уравнение регрессии будет иметь вид

$$y = \overline{y} + \beta\left(x - \overline{x}\right). \tag{6.16}$$

Используя выборочный коэффициент корреляции r (формула (6.6)) и связав формулы (6.9) и (6.12), можно получить следующую формулу для расчета углового коэффициента β:

$$\beta = \frac{n\sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{n\left(\sum_{i-1}^{n} x_i y_i - \frac{\sum_{i-1}^{n} x_i \sum_{i-1}^{n} y_i}{n}\right)}{n\left(\sum_{i-1}^{n} x_i^2 - \frac{\left(\sum_{i-1}^{n} x_i\right)^2}{n}\right)} = \qquad (6.17)$$

$$= \frac{\sum_{i-1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum_{i-1}^{n}\left(x_i - \overline{x}\right)^2} = r\frac{S_y}{S_x}.$$

После преобразования формулы (6.17) выборочный коэффициент корреляции будет выглядеть следующим образом:

$$r = \beta\frac{S_x}{S_y} = \beta\frac{\sqrt{n\sum_{i-1}^{n} x_i^2 - \left(\sum_{i-1}^{n} x_i\right)^2}}{\sqrt{n\sum_{i-1}^{n} y_i^2 - \left(\sum_{i-1}^{n} y_i\right)^2}}. \qquad (6.18)$$

Уравнение линии регрессии, выраженное через выборочный коэффициент корреляции, имеет вид

$$y = \overline{y} + r\frac{S_y}{S_x}\left(x - \overline{x}\right). \qquad (6.19)$$

Уравнения (6.16) и (6.19) – уравнения регрессии y по x. Если же находится обратная регрессия x по y в виде выражения (6.16):

$$x = \overline{x} + r\frac{S_x}{S_y}\left(y - \overline{y}\right),$$

то коэффициент регрессии x по y находится по аналогии с уравнением (6.12):

$$\beta_{y/x} = \frac{n\sum_{i=1}^{n} y_i \cdot x_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} = r\frac{S_x}{S_y}. \qquad (6.20)$$

Очевидно, что полученные по уравнениям (6.12) и (6.20) коэффициенты регрессии не равны между собой ($\beta_{y/x} \neq \beta_{x/y}$). Выборочный коэффициент корреляции может быть найден по этим коэффициентам регрессии: $\beta_{y/x} \cdot \beta_{x/y} = r^2$. Следствием этого является то, что линии регрессии y по x $(y = \bar{y} + \beta_{y/x}(x - \bar{x}))$ и x по y $(x = \bar{x} + \beta_{x/y}(y - \bar{y}))$ проходят через центр тяжести поля экспериментальных данных ($\bar{x}$, $\bar{y}$) (рис. 6.10).

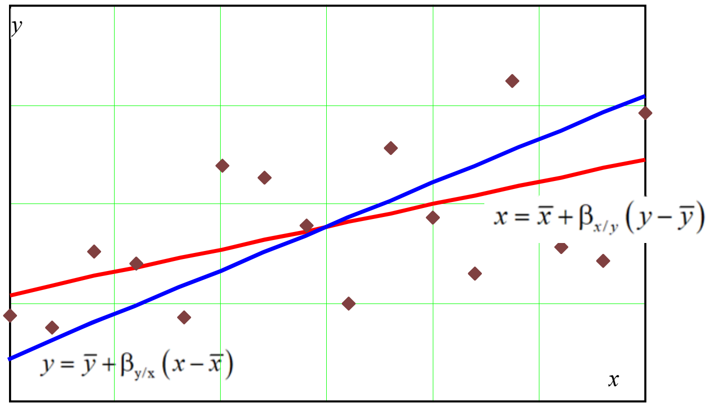


Рис. 6.10. Линейная регрессия

При коэффициенте корреляции ρ, близком к 1, т. е. при малом рассеянии экспериментальных точек и большой протяженности поля, обе линии регрессии практически совпадают.

Пример 1. Имеются данные средней выработки на одного рабочего y (тыс. руб.) и товарооборота x (тыс. руб.) в 10 магазинах за квартал (табл. 6.7).

Таблица 6.7

Средняя выработка на одного рабочего

Магазины	1	2	3	4	5	6	7	8	9	10
x, тыс. руб.	10	14	21	23	27	32	39	45	55	61
y, тыс. руб.	3,8	4,8	5,9	6,1	6,2	6,3	6,6	7,4	8,5	9,7

1. Определить зависимость (коэффициент корреляции) средней выработки на одного рабочего от товарооборота.
2. Составить уравнение регрессии этой зависимости.

Решение

Связь между изучаемыми признаками может быть выражена уравнением прямой линии регрессии Y на X: $\bar{y} = \alpha + \beta \cdot \bar{x}$. Для вычисления параметров α, β и коэффициента корреляции r используются формулы (6.12), (6.13), а также «удобные» формулы (6.9). Промежуточные вычисления удобно представить в виде расчетной табл. 6.8.

Таблица 6.8

Расчетная таблица

Магазины	x	y	x^2	y^2	xy
1	10	3,8	100	14,44	38
2	14	4,8	196	23,04	67,2
3	21	5,9	441	34,81	123,9
4	23	6,1	529	37,21	140,3
5	27	6,2	729	38,44	167,4
6	32	6,3	1024	39,69	201,6
7	39	6,6	1521	43,56	257,4
8	45	7,4	2025	54,76	333
9	55	8,5	3025	72,25	467,5
10	61	9,7	3721	94,09	591,7
Сумма:	**327**	**65,3**	**13311**	**452,3**	**2388**

Выборочный коэффициент корреляции найден по преобразованной формуле (6.6) с учетом формул (6.7)–(6.9):

$$r = \frac{1}{n-1}\frac{\sum\limits_{i-1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{S_x S_y} = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{\sqrt{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n\sum\limits_{i=1}^{n} y_i^2 - \left(\sum\limits_{i=1}^{n} y_i\right)^2}} =$$

$$= \frac{10 \cdot 2388 - 327 \cdot 65{,}3}{\sqrt{10 \cdot 13311 - 327^2} \cdot \sqrt{10 \cdot 452{,}3 - 65{,}3^2}} = 0{,}971.$$

Так как коэффициент корреляции практически равен 1, можно сделать вывод, что связь средней выработки на одного рабочего и товарооборота в магазине очень сильная, прямая.

На основе корреляционного анализа следует предположить линейную зависимость, выразить которую можно с помощью линейной регрессии. Параметры α, β находятся из системы уравнений (6.10):

$$\begin{cases} 10 \cdot \alpha + 327 \cdot \beta = 65{,}3; \\ 327 \cdot \alpha + 13311 \cdot \beta = 2388. \end{cases}$$

В результате расчета получаются следующие значения: α = 3,374, β = 0,097. Тогда уравнение регрессии примет вид

$$f(x) = 0{,}097 + 3{,}374 \cdot x.$$

На рис. 6.11 показана зависимости средней выработки на одного рабочего от товарооборота и рассчитанная линия регрессии.

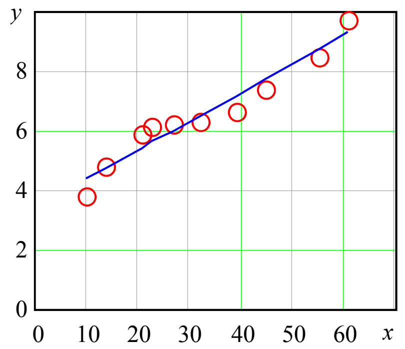


Рис. 6.11. График зависимости средней выработки на одного рабочего от товарооборота

Ответ: коэффициент корреляции для средней выработки на одного рабочего от товарооборота $r = 0{,}971$. Зависимость средней выработки на одного рабочего от товарооборота можно описать линейной регрессией вида $f(x) = 0{,}097 + 3{,}374 \cdot x$.

Общей характеристикой ошибки (проверка на адекватность), связанной с использование регрессионной зависимости, является величина, называемая *ошибкой регрессии*. Ошибка регрессии вычисляется по формуле

$$S = \sqrt{\frac{1}{n-2} \cdot \sum_i (y_i - f(x))^2}\,, \tag{6.21}$$

где y_i – экспериментальные данные; $f(x)$ – уравнение регрессии; $n-2$ – количество экспериментальных точек минус 2 (так как рассматриваются две случайные величины – X и Y).

Чем меньше ошибка регрессии, тем более адекватная получается модель регрессии.

6.4.2. Полиномиальная и обобщенная регрессии

Определение вида уравнения регрессии – очень сложная задача и не решается однозначно. Хорошо, если есть возможность использовать физические соображения и аналоги, в противном случае тип регрессии можно определить лишь путем кропотливого подбора. По экспериментальным данным осуществляется подгонка уравнения различного вида к истинной зависимости до тех пор, пока найденное уравнение регрессии не будет удовлетворять заданной точности.

Подбор вида уравнения начинается с простейшего – линейной регрессии. А если результат не удовлетворительный, то задается уравнение более сложной формы.

Очень многие эффекты и явления в физике и других точных науках не могут быть описаны линейным уравнением. Поэтому используется полиномиальная или обобщенная регрессии.

Полиномиальная регрессия описывает зависимость параметра y от параметра x в виде полинома:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n, \qquad (6.22)$$

где n – степень полиномиальной регрессии.

На рис. 6.12 показана полиномиальная регрессия.

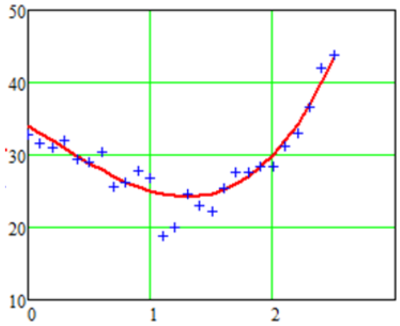


Рис. 6.12. Полиномиальная регрессия

Очевидно, что степень полинома качественно влияет на характер зависимости. Может сложиться впечатление, что чем больше степень полинома, тем лучше, но это не всегда так. При увеличении степени полинома увеличивается количество коэффициентов, при расчете которых неизбежно возникают ошибки из-за случайности исходных экспериментальных данных. Поэтому следует обдуманно подходить к определению степени полинома и проверять адекватность полученного уравнения регрессии. Применение различных программ значительно облегчает данный процесс.

В общем случае n может принимать совершенно любые значения, однако реально сталкиваться с полиномами выше 3–4 степени приходится редко, так как ошибки, связанные с расчетом коэффициентов регрессии, могут во много раз превышать значения коэффициентов.

Любую зависимость можно представить в виде полинома n-й степени, но предпочтительнее использовать функцию, которая описывает природу зависимости экспериментальных данных. Если есть информация об этой природе, то следует использовать *обобщенную регрессию*.

Обобщенная регрессия описывается линейным сочетанием $n + 1$ произвольных функций:

$$f(x) = C_1 f_1(x) + C_2 f_2(x) + \ldots + C_n f_n(x). \qquad (6.22)$$

Определить, какие функции стоит использовать для построения конкретной модели регрессии, можно только на основе априорной информации об экспериментальных данных или путем кропотливого подбора (но в это случае стоит проанализировать, не нарушена ли логика получения экспериментальных данных).

На рис. 6.13 показана обобщенная регрессия на основе сочетания линейной и тригонометрической функций: $y = C_1 \cdot x + C_2 \cdot \sin(C_3 \cdot x)$.

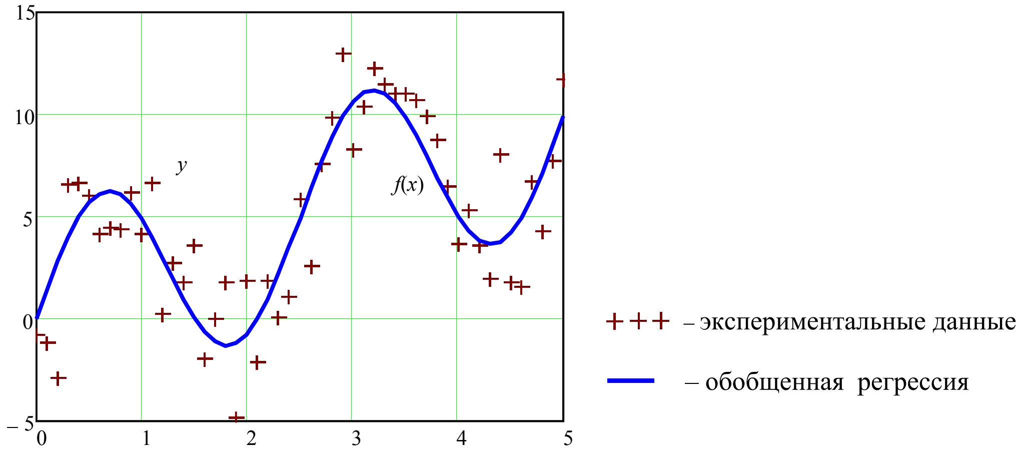


Рис. 6.13. Обобщенная регрессия

Задачи для самостоятельного решения

Задача 1. Результаты измерения электрического сопротивления R медного стержня при различной температуре приведены в табл. 6.9

Таблица 6.9

Зависимость сопротивления от температуры

t, °C	19,1	25,0	30,1	36,0	40.0	45,1	50,0
R, Ом	76,30	77,80	79,75	80,80	82,35	83,90	85,10

Полагая, что зависимость сопротивления от температуры имеет вид $R = a + b \cdot t$, найти параметры a и b в этой зависимости, а также проверить адекватность полученной регрессионной модели.

Задача 2. Исследовалась зависимость урожайности зерновых культур y (ц/га) от количества осадков x (см), выпавших в вегетационный период (в период роста и развития растений) (табл. 6.10).

Таблица 6.10

Зависимость урожайности зерновых культур от количества осадков

y, ц/га	25	27	30	35	36	38	39	41	42	45	46	47	50	52	53
x, см	23	24	27	27	32	31	33	35	34	32	29	28	25	24	25

Провести регрессионный анализ, получить полиномы 1-го и 2-го порядка и оценить адекватность каждого из них.

Контрольные вопросы и задания

1. Какую зависимость называют функциональной; статистической; корреляционной?
2. В чем состоит корреляционный анализ?
3. Что такое корреляционный момент?
4. Что такое коэффициент корреляции? Какими свойствами обладает коэффициент корреляции?
5. Приведите формулы для вычисления выборочного корреляционного момента и коэффициента корреляции.
6. Как проверить гипотезу о значимости выборочного коэффициента корреляции? Приведите алгоритм проверки, используя:
   a) r-критерий;
   b) t-критерий.
7. Дайте определения понятиям: регрессия, линия регрессии, коэффициент линии регрессии и свободная линия регрессии.
8. В чем состоит регрессионный анализ?
9. Приведите уравнения для линейной, полиномиальной и обобщенной регрессии.
10. Как оценить адекватность выбранной регрессионной модели?

ЗАКЛЮЧЕНИЕ

В данном учебном пособии представлен обобщенный материал, включающий в себя сжатый теоретический материал из разделов теории вероятностей и математической статистики без излишних, трудных для понимания математических выкладок. Автор попытался максимально подробно привести алгоритмы решения со ссылками на необходимый для понимания теоретический материал наиболее часто встречающихся задач. Задания для самостоятельного решения представляют собой аналогичные прорешенным задачам и позволяют проверить степень усвоения изложенного теоретического и практического материала.

Данное пособие можно использовать для методического обеспечения лекционных и практических занятий различных дисциплин, таких как «Теория измерений», «Теория погрешностей», «Математические основы измерений» и т. д., а также рекомендовать для самостоятельной работы студентов.

Автор считает, что данное пособие будет полезно студентам не только направления «Приборостроение», но и других технических специальностей для освоения основ обработки и представления результатов измерений применительно к любой технической области знаний.

Пособие является первой частью. В последующем автор планирует подготовить учебное пособие по применению изложенного здесь материала для решения конкретных инженерных задач по расчету погрешностей и по обработке результатов измерений.

УСЛОВНЫЕ СОКРАЩЕНИЯ И АББРЕВИАТУРЫ

СВ – случайная величина.
ДСВ – дискретная случайная величина.
НСВ – непрерывная случайная величина.
Мат. ожидание – математическое ожидание.
СКО – среднее квадратическое отклонение.
ЭД – экспериментальные данные.

СПИСОК ЛИТЕРАТУРЫ

1. Колмогоров А.Н. Роль русской науки в развитии теории вероятностей / А.Н. Колмогоров // Учёные записки МГУ. – 1947. – Т. 1, вып. 91, кн.1. – С. 53–64.

2. Гнеденко Б.В. Очерк по истории теории вероятностей / Б.В. Гнеденко // Курс теории вероятностей. – 8-е изд. – Москва : Едиториал УРСС, 2005. – 448 с.

3. Орлов А.И. Математика случая. Вероятность и статистика – основные факты : учебное пособие / А.И. Орлов. – Москва : МЗ-Пресс, 2004. – 110 с.

4. Реньи А. Трилогия о математике / А. Реньи. – Москва : Мир, 1980. – 376 с.

5. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие для бакалавров / В.Е. Гмурман. – 11-е изд., перераб. и доп. – Москва : Юрайт, 2013. – 404 с.

6. Бакалавр экономики. Хрестоматия в 3 томах. Том 1 / под общ. ред. В.И. Видяпина. – Москва : Информационно-издательская фирма «Триада», 1999. – 696 с.

7. Мойзес Б.Б. Статистические методы контроля качества и обработка экспериментальных данных : учебное пособие для вузов / Б.Б. Мойзес, И.В. Плотникова, Л.А. Редько. – 2-е изд. – Москва : Юрайт, 2021. – 118 с.

8. Гмурман В.Е. Теория вероятностей и математическая статистика : учебное пособие для бакалавров / В.Е. Гмурман. – 12-е изд. – Москва : Юрайт. 2013. – 479 с.

9. ГОСТ Р 50779.10-2000 (ИСО 3534.1-93) Статистические методы. Вероятность и основы статистики. Термины и определения. – Москва : Госстадарт России, 2008. – 42 с.

10. Садовский Г.А. Теоретические основы информационно-измерительной техники : учебное пособие / Г.А. Садовский. – Москва : Высшая школа, 2008. – 478 с.

11. Лемешко Б.Ю. Непараметрические критерии согласия. Руководство по применению / Б.Ю. Лемешко. – Москва : Инфра-М, 2014. – 163 с.

12. Пронкин Н.С. Основы метрологии : практикум по метрологии и измерениям / Н.С. Пронкин. – Москва : Логос, 2007. – 392 с.

ПРИЛОЖЕНИЕ 1

Функция Лапласа

Таблица П1.1

Таблица значений локальной функции Лапласа

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006

Окончание табл. П1.1

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Таблица П1.2

Таблица значений интегральной функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,00000	0,50	0,19146	1,00	0,34134	1,50	0,43319	2,00	0,47725	3,00	0,49865
0,01	0,00399	0,51	0,19497	1,01	0,34375	1,51	0,43448	2,02	0,47831	3,05	0,49886
0,02	0,00798	0,52	0,19847	1,02	0,34614	1,52	0,43574	2,04	0,47932	3,10	0,49903
0,03	0,01197	0,53	0,20194	1,03	0,34849	1,53	0,43699	2,06	0,48030	3,15	0,49918
0,04	0,01595	0,54	0,20540	1,04	0,35083	1,54	0,43822	2,08	0,48124	3,20	0,49931
0,05	0,01994	0,55	0,20884	1,05	0,35314	1,55	0,43943	2,10	0,48214	3,25	0,49942
0,06	0,02392	0,56	0,21226	1,06	0,35543	1,56	0,44062	2,12	0,48300	3,30	0,49952
0,07	0,02790	0,57	0,21566	1,07	0,35769	1,57	0,44179	2,14	0,48382	3,35	0,49960
0,08	0,03188	0,58	0,21904	1,08	0,35993	1,58	0,44295	2,16	0,48461	3,40	0,49966
0,09	0,03586	0,59	0,22240	1,09	0,36214	1,59	0,44408	2,18	0,48537	3,45	0,49972
0,10	0,03983	0,60	0,22575	1,10	0,36433	1,60	0,44520	2,20	0,48610	3,50	0,49977
0,11	0,04380	0,61	0,22907	1,11	0,36650	1,61	0,44630	2,22	0,48679	3,55	0,49981
0,12	0,04776	0,62	0,23237	1,12	0,36864	1,62	0,44738	2,24	0,48745	3,60	0,49984
0,13	0,05172	0,63	0,23565	1,13	0,37076	1,63	0,44845	2,26	0,48809	3,65	0,49987
0,14	0,05567	0,64	0,23891	1,14	0,37286	1,64	0,44950	2,28	0,48870	3,70	0,49989
0,15	0,05962	0,65	0,24215	1,15	0,37493	1,65	0,45053	2,30	0,48928	3,75	0,49991
0,16	0,06356	0,66	0,24537	1,16	0,37698	1,66	0,45154	2,32	0,48983	3,80	0,49993
0,17	0,06749	0,67	0,24857	1,17	0,37900	1,67	0,45254	2,34	0,49036	3,85	0,49994
0,18	0,07142	0,68	0,25175	1,18	0,38100	1,68	0,45352	2,36	0,49086	3,90	0,49995
0,19	0,07535	0,69	0,25490	1,19	0,38298	1,69	0,45449	2,38	0,49134	3,95	0,49996
0,20	0,07926	0,70	0,25804	1,20	0,38493	1,70	0,45543	2,40	0,49180	4,00	0,49997
0,21	0,08317	0,71	0,26115	1,21	0,38686	1,71	0,45637	2,42	0,49224	4,05	0,49997
0,22	0,08706	0,72	0,26424	1,22	0,38877	1,72	0,45728	2,44	0,49266	4,10	0,49998
0,23	0,09095	0,73	0,26730	1,23	0,39065	1,73	0,45818	2,46	0,49305	4,15	0,49998
0,24	0,09483	0,74	0,27035	1,24	0,39251	1,74	0,45907	2,48	0,49343	4,20	0,49999
0,25	0,09871	0,75	0,27337	1,25	0,39435	1,75	0,45994	2,50	0,49379	4,25	0,49999
0,26	0,10257	0,76	0,27637	1,26	0,39617	1,76	0,46080	2,52	0,49413	4,30	0,49999
0,27	0,10642	0,77	0,27935	1,27	0,39796	1,77	0,46164	2,54	0,49446	4,35	0,49999
0,28	0,11026	0,78	0,28230	1,28	0,39973	1,78	0,46246	2,56	0,49477	4,40	0,49999

Окончание табл. П1.2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,29	0,11409	0,79	0,28524	1,29	0,40147	1,79	0,46327	2,58	0,49506	4,45	0,50000
0,30	0,11791	0,80	0,28814	1,30	0,40320	1,80	0,46407	2,60	0,49534	4,50	0,50000
0,31	0,12172	0,81	0,29103	1,31	0,40490	1,81	0,46485	2,62	0,49560	4,55	0,50000
0,32	0,12552	0,82	0,29389	1,32	0,40658	1,82	0,46562	2,64	0,49585	4,60	0,50000
0,33	0,12930	0,83	0,29673	1,33	0,40824	1,83	0,46638	2,66	0,49609	4,65	0,50000
0,34	0,13307	0,84	0,29955	1,34	0,40988	1,84	0,46712	2,68	0,49632	4,70	0,50000
0,35	0,13683	0,85	0,30234	1,35	0,41149	1,85	0,46784	2,70	0,49653	4,75	0,50000
0,36	0,14058	0,86	0,30511	1,36	0,41309	1,86	0,46856	2,72	0,49674	4,80	0,50000
0,37	0,14431	0,87	0,30785	1,37	0,41466	1,87	0,46926	2,74	0,49693	4,85	0,50000
0,38	0,14803	0,88	0,31057	1,38	0,41621	1,88	0,46995	2,76	0,49711	4,90	0,50000
0,39	0,15173	0,89	0,31327	1,39	0,41774	1,89	0,47062	2,78	0,49728	4,95	0,50000
0,40	0,15542	0,90	0,31594	1,40	0,41924	1,90	0,47128	2,80	0,49744	5,00	0,50000
0,41	0,15910	0,91	0,31859	1,41	0,42073	1,91	0,47193	2,82	0,49760		
0,42	0,16276	0,92	0,32121	1,42	0,42220	1,92	0,47257	2,84	0,49774		
0,43	0,16640	0,93	0,32381	1,43	0,42364	1,93	0,47320	2,86	0,49788		
0,44	0,17003	0,94	0,32639	1,44	0,42507	1,94	0,47381	2,88	0,49801		
0,45	0,17364	0,95	0,32894	1,45	0,42647	1,95	0,47441	2,90	0,49813		
0,46	0,17724	0,96	0,33147	1,46	0,42785	1,96	0,47500	2,92	0,49825		
0,47	0,18082	0,97	0,33398	1,47	0,42922	1,97	0,47558	2,94	0,49836		
0,48	0,18439	0,98	0,33646	1,48	0,43056	1,98	0,47615	2,96	0,49846		
0,49	0,18793	0,99	0,33891	1,49	0,43189	1,99	0,47670	2,98	0,49856		

ПРИЛОЖЕНИЕ 2

Коэффициенты для различных распределений

Таблица П2.1

Значения для распределения Стьюдента

Число степеней свободы k	Уровень значимости α (доверительная вероятность P)				
	0,1 (0,9)	**0,05 (0,95)**	**0,02 (0,98)**	**0,01 (0,99)**	**0,001 (0,999)**
1	6,3138	12,7062	31,8205	63,6567	636,6192
2	2,9200	4,3027	6,9646	9,9248	31,5991
3	2,3534	3,1824	4,5407	5,8409	12,924
4	2,1318	2,7764	3,7469	4,6041	8,6103
5	2,0150	2,5706	3,3649	4,0321	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,9588
7	1,8946	2,3646	2,9980	3,4995	5,4079
8	1,8595	2,3060	2,8965	3,3554	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,4370
12	1,7823	2,1788	2,6810	3,0545	4,3178
13	1,7709	2,1604	2,6503	3,0123	4,2208
14	1,7613	2,1448	2,6245	2,9768	4,1405
15	1,7531	2,1314	2,6025	2,9467	4,0728
16	1,7459	2,1199	2,5835	2,9208	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,6460
35	1,6896	2,0301	2,4377	2,7238	3,5911
40	1,6839	2,0211	2,4233	2,7045	3,5510
45	1,6794	2,0141	2,4121	2,6896	3,5203

Окончание табл. П2.1

Число степеней свободы *k*	Уровень значимости α (доверительная вероятность *P*)				
	0,1 (0,9)	**0,05 (0,95)**	**0,02 (0,98)**	**0,01 (0,99)**	**0,001 (0,999)**
50	1,6759	2,0086	2,4033	2,6778	3,4960
55	1,6730	2,004	2,3961	2,6682	3,4764
60	1,6706	2,0003	2,3901	2,6603	3,4602
70	1,6669	1,9944	2,3808	2,6479	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,3905
110	1,6588	1,9818	2,3607	2,6213	3,3812
120	1,6577	1,9799	2,3578	2,6174	3,3735
∞	1,6448	1,9600	2,3263	2,5758	3,2905

Таблица П2.2

Значения для распределения Пирсона (χ^2-распределения)

Число степеней свободы *k*	Уровень значимости α (доверительная вероятность *P*)					
	0,01 (0,99)	**0,025 (0,975)**	**0,05 (0,95)**	**0,95 (0,05)**	**0,975 (0,025)**	**0,99 (0,01)**
1	6,63490	5,02389	3,84146	0,00393	0,00098	0,00016
2	9,21034	7,37776	5,99146	0,10259	0,05064	0,02010
3	11,34487	9,34840	7,81473	0,35185	0,21580	0,11483
4	13,2767	11,14329	9,48773	0,71072	0,48442	0,29711
5	15,08627	12,8325	11,0705	1,14548	0,83121	0,55430
6	16,81189	14,44938	12,59159	1,63538	1,23734	0,87209
7	18,47531	16,01276	14,06714	2,16735	1,68987	1,23904
8	20,09024	17,53455	15,50731	2,73264	2,17973	1,64650
9	21,66599	19,02277	16,91898	3,32511	2,70039	2,08790
10	23,20925	20,48318	18,30704	3,94030	3,24697	2,55821
11	24,72497	21,92005	19,67514	4,57481	3,81575	3,05348
12	26,21697	23,33666	21,02607	5,22603	4,40379	3,57057
13	27,68825	24,7356	22,36203	5,89186	5,00875	4,10692
14	29,14124	26,11895	23,68479	6,57063	5,62873	4,66043
15	30,57791	27,48839	24,99579	7,26094	6,26214	5,22935
16	31,99993	28,84535	26,29623	7,96165	6,90766	5,81221
17	33,40866	30,19101	27,58711	8,67176	7,56419	6,40776
18	34,80531	31,52638	28,86930	9,39046	8,23075	7,01491
19	36,19087	32,85233	30,14353	10,11701	8,90652	7,63273
20	37,56623	34,16961	31,41043	10,85081	9,59078	8,26040
21	38,93217	35,47888	32,67057	11,59131	10,2829	8,89720

Окончание табл. П2.2

Число степеней свободы *k*	Уровень значимости α (доверительная вероятность *P*)					
	0,01 (0,99)	**0,025 (0,975)**	**0,05 (0,95)**	**0,95 (0,05)**	**0,975 (0,025)**	**0,99 (0,01)**
22	40,28936	36,78071	33,92444	12,33801	10,98232	9,54249
23	41,63840	38,07563	35,17246	13,09051	11,68855	10,19572
24	42,97982	39,36408	36,41503	13,84843	12,40115	10,85636
25	44,31410	40,64647	37,65248	14,61141	13,11972	11,52398
26	45,64168	41,92317	38,88514	15,37916	13,84391	12,19815
27	46,96294	43,19451	40,11327	16,15140	14,57338	12,87850
28	48,27824	44,46079	41,33714	16,92788	15,30786	13,56471
29	49,58788	45,72229	42,55697	17,70837	16,04707	14,25645
30	50,89218	46,97924	43,77297	18,49266	16,79077	14,95346
31	52,19139	48,23189	44,98534	19,28057	17,53874	15,65546
32	53,48577	49,48044	46,19426	20,07191	18,29076	16,36222
33	54,77554	50,72508	47,39988	20,86653	19,04666	17,07351
34	56,06091	51,96600	48,60237	21,66428	19,80625	17,78915
35	57,34207	53,20335	49,80185	22,46502	20,56938	18,50893
36	58,61921	54,43729	50,99846	23,26861	21,33588	19,23268
37	59,89250	55,66797	52,19232	24,07494	22,10563	19,96023
38	61,16209	56,89552	53,38354	24,8839	22,87848	20,69144
39	62,42812	58,12006	54,57223	25,69539	23,65432	21,42616
40	63,69074	59,34171	55,75848	26,5093	24,43304	22,16426
41	64,95007	60,56057	56,94239	27,32555	25,21452	22,90561
42	66,20624	61,77675	58,12404	28,14405	25,99866	23,65009
43	67,45935	62,99035	59,30351	28,96472	26,78537	24,3976
44	68,70951	64,20146	60,48088	29,78748	27,57457	25,14803
45	69,95683	65,41016	61,65623	30,61226	28,36615	25,90127
46	71,2014	66,61653	62,82962	31,43900	29,16005	26,65724
47	72,44331	67,82065	64,00111	32,26762	29,9562	27,41585
48	73,68264	69,02258	65,17077	33,09808	30,75451	28,17701
49	74,91947	70,22241	56,94239	33,93031	31,55492	28,94065
50	76,15389	71,42020	67,50481	34,76425	32,35736	29,70668
60	88,37942	83,29767	79,08194	43,18796	40,48175	37,48485
70	100,42518	95,02318	90,53122	51,73928	48,75757	45,44172
80	112,32879	106,62857	101,87947	60,39148	57,15317	53,54008
90	124,11632	118,13589	113,14527	69,12603	65,64662	61,75408
100	135,80672	129,5612	124,34211	77,92947	74,22193	37,48485

Таблица П2.3

Значения для распределения Фишера (F-распределения)
(k_1 – число степеней свободы большей дисперсии,
k_2 – число степеней свободы меньшей дисперсии)

k_2	Уровень значимости α = 0,01												
	k_1												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55
6	13,75	10,93	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,67
14	8,86	6,52	5,56	5,04	4,70	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,90	3,81	3,67	3,52	3,37
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16
18	8,29	6,01	5,09	4,58	4,25	4,02	3,84	3,71	3,60	3,51	3,37	3,23	3,08
19	8,19	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94

k_2	Уровень значимости α = 0,05												
	k_1												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**
1	161	200	216	225	230	234	237	239	241	242	244	246	248
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12

Таблица П2.4

Значения λ_n для критерия Колмогорова

Доверительная вероятность, *P*	λ_n	Доверительная вероятность, *P*	λ_n	Доверительная вероятность, *P*	λ_n
0,01	0,44	0,40	0,77	0,90	1,22
0,05	0,52	0,50	0,83	0,95	1,36
0,10	0,57	0,60	0,89	0,98	1,52
0,20	0,65	0,70	0,97	0,99	1,63
0,30	0,71	0,80	1,07	0,999	1,95

Таблица П2.5

Значения коэффициента корреляции r

Число степеней свободы *k*	Уровень значимости α		Число степеней свободы *k*	Уровень значимости α	
	0,05	**0,01**		**0,05**	**0,01**
1	0,997	1,000	**24**	0,388	0,496
2	0,950	0,990	**25**	0,381	0,487
3	0,878	0,959	**26**	0,374	0,478
4	0,811	0,917	**27**	0,367	0,470
5	0,754	0,874	**28**	0,361	0,463
6	0,707	0,834	**29**	0,355	0,456
7	0,666	0,798	**30**	0,349	0,449
8	0,632	0,765	**35**	0,325	0,418
9	0,602	0,735	**40**	0,304	0,393
10	0,576	0,708	**45**	0,288	0,372
11	0,553	0,684	**50**	0,273	0,354
12	0,532	0,661	**60**	0,250	0,325
13	0,514	0,641	**70**	0,232	0,302
14	0,497	0,623	**80**	0,217	0,283
15	0,482	0,606	**90**	0,205	0,267
16	0,468	0,590	**100**	0,195	0,254
17	0,456	0,575	**125**	0,174	0,228
18	0,444	0,561	**150**	0,159	0,208
19	0,433	0,549	**200**	0,138	0,181
20	0,423	0,537	**300**	0,113	0,148
21	0,413	0,526	**400**	0,098	0,128
22	0,404	0,515	**500**	0,088	0,115
23	0,396	0,505	**1000**	0,062	0,081

Оглавление

Учебное издание

ВАВИЛОВА Галина Васильевна

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ РЕШЕНИЯ ИНЖЕНЕРНЫХ ЗАДАЧ

Учебное пособие

Корректура *Д.В. Заремба*

Компьютерная верстка *Д.В. Сотникова*

Дизайн обложки *Т.В. Буланова*

Подписано к печати 16.04.2022. Формат 60x84/16. Бумага «Снегурочка».
Печать CANON. Усл. печ. л. 11,46. Уч.-изд. л. 10,36.
Заказ 99-22. Тираж 100 экз.

Издательство
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ