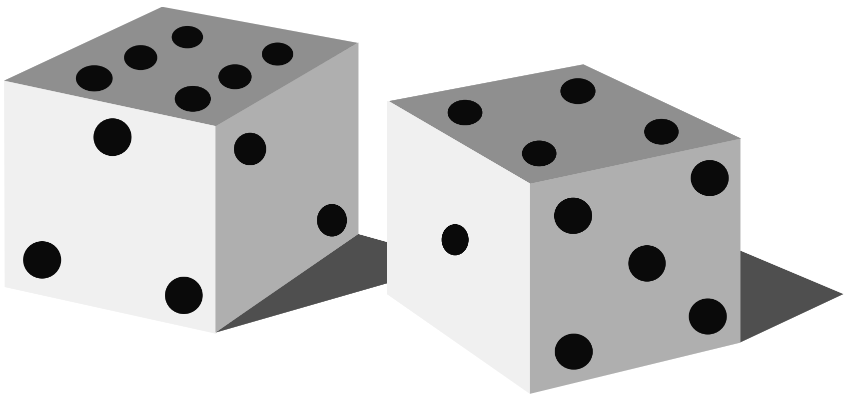
МИНИСТЕРСТВО НАУКИ И ОБРАЗОВАНИЯ УКРАИНЫ

ДОНЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Серия: «Математика для экономистов»

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Программа и методические указания по изучению курса



Заочная и ускоренная форма обучения

Донецк, ДонГУ 2000

УДК 519.217

Настоящие указания составлены в соответствии с программой курса "Математика для экономистов" высших учебных заведений.

(для студентов экономических специальностей заочной и ускоренной формы обучения) / Сост. В.Д. Породников, - Донецк: ДонГУ, 2000.

Приведены - программа курса «Теория вероятности и математическая статистика», задания для самоконтроля и общие указания по изучение этой дисциплины.

Р е ц е н з е н т ы:

заведующий кафедрой математики, информатики и вычислительной техники Донецкой государственной академии управления д-р физ-мат наук, проф. *Л.Е. Шайхет*,

доцент кафедры Алгебры и теории вероятностей, канд. физ-мат наук *В Н. Бандура.*

Введение

Настоящие методические указания для студентов-заочников экономических специальностей университета составлены в соответствии с программой курса «Математика для экономистов», которая утверждена научно-методической комиссией по экономическому образованию 14 октября 1997 года.

Методические указания содержат программу курса «Теория вероятностей и математическая статистика», которая утверждена кафедрой «Математики и математических методов экономике», контрольные задания по данному курсу и общие методические указания к изучению этой дисциплины.

По действующему учебному плану курс «Теория вероятностей и математическая статистика» изучается в течение одного семестра. В процессе изучения данного курса студент-заочник должен выполнить *контрольную работу*, главная цель которой - оказать помощь в его работе по изучению данного курса и приобретении навыков самостоятельной работы. Рецензии на эту работу и ее защита позволяют студенту судить о степени усвоения им соответствующего раздела курса, указывают на имеющиеся пробелы, на желательное направление дальнейшей работы; помогают сформулировать вопросы для консультации с преподавателем (письменной или устной).

Целью изучения курса является развитие у студентов навыков самостоятельного изучения учебной литературы, выработка навыков математического исследования прикладных вопросов и умения перевести экономическую задачу на математический язык. Освоение необходимого вероятностного аппарата поможет моделировать, анализировать и решать экономические задачи с применением, в случае необходимости, ЭВМ.

ОБЩИЕ РЕКОМЕНДАЦИИ

Основной формой обучения студента-заочника является самостоятельная работа над учебным материалом. По математическим курсам она складывается из чтения учебников, решения задач, выполнения контрольных заданий. В помощь заочникам университет организует чтение лекций и практические занятия. Кроме того, студент может обращаться к преподавателю с вопросами для письменной или устной консультации. Указания студенту по текущей работе даются также в процессе рецензирования контрольных работ. Однако студент должен помнить, что только при систематической и упорной самостоятельной работе помощь преподавателей университета будет достаточно эффективной. Если изучать материал формально (ввиду необходимости сдачи зачета или экзамена), то можно потерять драгоценное время, а потом чрезмерная поспешность при обучении значительно уменьшить его пользу. Результат обучения оценивается не количеством прочитанной информации, а качеством ее усвоения и развитием способностей к дальнейшему самостоятельному образованию.

Завершающим этапом изучения каждого из математических курсов или отдельных их частей является сдача зачетов и экзаменов в соответствии с учебным планом.

Чтение учебника

Изучая материал по учебнику или учебному пособию, следует переходить к следующему вопросу только после правильного понимания предыдущего, проделывая на бумаге все вычисления (в том числе и те, которые в учебнике опущены), воспроизводя имеющиеся в учебнике чертежи.

Особое внимание следует обращать на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно.

Необходимо помнить, что каждая теорема состоит из предположений и утверждения. Все предположения должны обязательно использоваться в доказательстве. Нужно добиваться точного представления о том, в каком месте доказательства использовано каждое предположение теоремы. Полезно составлять схемы доказательств сложных теорем. Правильному пониманию многих теорем помогает разбор примеров, обладающих и не обладающих свойствами, указанными в предположениях и утверждениях теорем.

При изучении материала по учебнику полезно вести конспект, в который рекомендуется выписывать определения, формулировки теорем, формул, уравнения и т.п. На полях конспекта следует отмечать вопросы, выделенные студентом для письменной или устной консультации с преподавателем.

Выводы, полученные в виде формул, рекомендуется в конспекте подчеркивать или обводить рамкой, чтобы при перечитывании конспекта они выделялись и лучше запоминались. Опыт показывает, что многим студентам помогает составление листа, содержащего важнейшие и наиболее часто употребляемые формулы курса. Такой лист не только помогает запомнить формулы, но и может служить постоянным справочником для студента.

Решение задач

Чтение учебника должно сопровождаться решением задач по соответствующей теме, для чего рекомендуется завести специальную тетрадь.

При решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения задачи, то он должен сравнить их и выбрать самый рациональный.

Полезно до начала вычислений составить краткий план решения. Решение задач и примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления. от основных. Чертежи можно выполнять от руки, но аккуратно и в соответствии с условием задачи. Если чертеж требует особо тщательного выполнения, например при графической проверке решения, полученного путем вычислений, то следует пользоваться линейкой, транспортиром, лекалом и указывать масштаб.

Решение каждой задачи должно доводиться до окончательного ответа, которого требует условие, и по возможности в общем виде с выводом формулы. Затем в полученную формулу подставляют числовые значения (если таковые даны) входящих в нее параметров, букв. В промежуточные вычисления не следует вводить приближенные значения корней, чисел.

Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно такте, если возможно, решить задачу несколькими способами и сравнить полученные результаты.
Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

Самопроверка

После изучения определенной темы по учебнику и решения достаточного количества соответствующих задач студенту рекомендуется воспроизвести по памяти определения, выводы формул, формулировки и доказательства теорем, проверяя себя каждый раз по учебнику.

В случае необходимости надо еще раз внимательно разобраться в материале учебника. и решить задачи. Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дельнейшего материала. В этом случае надо вернуться назад и повторить плохо усвоенный раздел.

Важным критерием усвоения теории является умение решать задачи на пройденный материал. Однако здесь следует предостеречь студента от весьма распространенной ошибки, заключающейся в том, что благополучное решение задач воспринимается им как признак усвоения теории. Иногда правильное решение задачи получается в результате применения механически заученных формул без понимания сущности теоретических положений.

Консультации

Если в процессе работы над изучением теоретического материала или при решении задач у студента возникают вопросы, разрешить которые самостоятельно не удается (неясность терминов, формулировок теорем, отдельных задач и др.), он может обратиться к преподавателю для получения от него указаний в виде письменной или устной консультации. В своих вопросах студент должен точно указывать, в чем он испытывает затруднение. Если он не разобрался в теоретических объяснениях, в доказательстве теоремы или в выводе формулы по учебнику, то нужно уточнить какой это учебник, год его издания, страницу, где рассмотрен затрудняющий вопрос, что именно его затрудняет. Если студент испытывает затруднение при решении задачи, то следует указать характер этого затруднения, привести предлагаемый план решения.

За консультацией следует обращаться и в случае, если возникнут сомнения в правильности ответов на вопросы для самопроверки.

Контрольные работы

В процессе изучения курса студент должен выполнить контрольную работу, главная цель которой - оказать студенту помощь в его работе. Рецензия на эту работу позволит студенту судить о степени усвоения им соответствующего раздела курса; укажет на имеющиеся у него пробелы, на желательное направление дальнейшей работы; поможет сформулировать вопросы для консультации с преподавателем (письменной или устной).

Не следует приступать к выполнению контрольного задания до решения достаточного количества задач по материалу, соответствующему этому заданию. Опыт показывает, что чаще всего неумение решить ту или иную задачу контрольного задания вызвано тем, что студент не выполнил это требование.

Контрольная работа должна выполняться самостоятельно. Несамостоятельно выполненная работа не дает возможности преподавателю-рецензенту указать студенту на недостатки в его работе, в усвоении им учебного материала, в результате чего студент не приобретает необходимых знаний и может оказаться неподготовленным к защите контрольной работы, а потом и к устному экзамену или зачету.

Не рекомендуется присылать контрольную работу на кануне защиты; это не дает возможности рецензенту своевременно указать студенту на допускаемые им ошибки и удлиняет срок рецензирования работ.

Прорецензированная контрольную работу вместе со всеми исправлениями и дополнениями, сделанными по требованию рецензента, следует сохранять. Без предъявления преподавателю прорецензированных контрольных работ студент не допускается к сдаче зачета (экзамена).

Сроки выполнения контрольной работы устанавливаются деканатом факультета в соответствии с распределением по семестрам материала и сообщаются студентам дополнительно.

Лекции и практические занятия

Лекции занимают в учебном процессе особое место: они направляют его и определяют его содержание и уровень. Во время установочных сессий для студентов-заочников проводятся лекции и практические занятия. Они носят по преимуществу обзорный характер. Их цель - обратить внимание на общую схему построения соответствующего раздела курса, подчеркнуть важнейшие факты, указать главные практические приложения, факты из истории науки. Кроме того, на этих занятиях могут быть более подробно разобраны отдельные вопросы курса (например, использование статистических выводов для обработки экономической информации методы описания наблюдений и др.); а также рассмотрены отдельные вопросы программы, отсутствующие или недостаточно полно освещенные в рекомендуемых пособиях. Для того чтобы практические занятия приносили пользу необходимо ознакомиться с теоретическим материалом до их начала.

Зачет (экзамен)

На зачете выясняется, прежде всего, отчетливое усвоение всех теоретических и прикладных вопросов программы, умение применять полученные знания к решению практических задач.

Определения, теоремы, правила должны формироваться точно и с пониманием существа дела, решение задач в простейших случаях проделываться без ошибок и уверенно; всякая письменная и графическая работа должна быть аккуратной и четкой. Только при выполнении этих условий знания могут быть признаны удовлетворяющими требованиям, предъявляемым программой.

При подготовке к зачету (экзамену) учебный материал рекомендуется повторять по учебнику и конспекту, обратив особое внимание на замечания лектора при обзоре программного материала.

Рекомендации

Настоящие методические указания составлены применительно к книгам, которые приведены в списке литературы после программы курса.

Ниже приводится план изучения курса в виде перечня разделов и тем, подлежащих последовательному изучению. После каждой темы даны номера глав и параграфы по учебникам, а также номера обязательных задач по сборникам задач. По каждой теме из указаний литературы можно использовать по выбору один из источников. Для краткости название учебника или сборника задач не повторяется, а указывается его номер.

ПРОГРАММА

<u>Тема 1</u>. *Основные понятия теории вероятностей*

Предмет курса, его содержание. Роль и место курса, как теоретической базы вероятностно-статистического моделирования. Стохастический эксперимент и событие. Вероятность события. Классическое и статистическое определение вероятноестей. Элементы комбинаторики в теории вероятностей. Представление о теоретико-множественных понятиях теории меры и их использование в построении общей теории вероятностей. Алгебра событий и их классификация. Вероятностное пространство. Свойства вероятностей. Теорема сложения вероятностей. Понятие зависимых и независимых случайных событий. Условные вероятности. Теорема умножения вероятностей. Формула полной вероятности и формулы Байеса.

Литература:
[1, §1-6] , [2 ., п. 1], [3, гл. 1-2, зад. 1-109], [4, гл. 1-4], [5; гл. 2]; [6, гл. 1-2], [7, гл. 25, п. А], [11,р.1, лаб. раб. 1-6].

Для усвоения этой темы нужно воспользоваться одним из предложенных учебников или сборников задач, разобрать задачи, решенные в них, а также самостоятельно решить рекомендованные задачи. При решении вероятностных задач используется раздел элементарной математики - комбинаторика. Перед изучением основных теорем необходимо хорошо усвоить действия над событиями их классификацию. Особое внимание нужно уделить формуле полной вероятности.

Вопросы для самопроверки

- Дайте классическое и статистическое определение вероятности.
- Сформулируйте принцип практической невозможности маловероятных событий.
- Дайте определение суммы событий.
- Докажите теорему сложения вероятностей несовместимых событий.
- Можно ли считать эту теорему частным случаем теоремы сложения вероятностей совместных событий?
- Дайте определение произведения событий.
- Докажите теорему умножения вероятностей независимых событий.
- Дайте определение условной вероятности.
- Докажите теорему умножения вероятностей зависимых событий.

- Выведите формулу Бернулли.
- Выведите формулу полной вероятности.
- Выведите формулу Байеса. Для чего служит эта формула?.

<u>Тема 2</u>. *Случайные величины и их характеристики*

Определение, примеры и основные типы случайной величины (сл.в.). Дискретные и непрерывные случайные величины. Функция распределения дискретной сл.в. и ее свойства. Непрерывная сл.в. и ее функция плотности распределения вероятностей с ее свойствами. Числовые характеристики случайных величин и их свойства. Начальные и центральные моменты сл. в., асимметрия и эксцесс.

Литература:
[1, §7-13] , [2 ., п. 2], [3, гл. 3,4,6, зад. 110-158, 164, 165, 188-235, 252-306], [4, гл. 5-8, 10, 11], [5; гл. 3-4]; [6, гл. 2, §2.2, гл. 3-4], [7, гл. 25, п. Б,В], [18 т. 1-2], [20, р. 2-3, лаб. раб. 7-16].

Нужно обратить внимание на определение случайной величины, ее функции распределения. Эти понятия являются одними из основных в теории вероятностей. При изучении числовых характеристик .следует уяснить вероятностный смысл математического ожидания и дисперсии, научиться уверенно вычислять эти характеристики.

Вопросы для самопроверки

- Приведите примеры дискретных и непрерывных случайных величин.
- Что называют законом распределения вероятностей дискретной случайной величины?
- Дайте определение интегральной функции и докажите ее свойства.
- Как, зная интегральную функцию, найти вероятность того, что случайная величина примет значение, заключенное в данном интервале?
- В чем состоит различие графиков интегральной функции непрерывной и дискретной случайных величин?
- Дайте определение дифференциальной функции и докажите ее свойства.
- Применима ли дифференциальная функция для задания дискретной случайной величины?
- Как найти интегральную функцию по известной дифференциальной функции?
- Дайте определение математического ожидания дискретной случайной величины и докажите его свойства.

- Дайте определение дисперсии дискретной случайной величины и докажите ее свойства.
- Дайте определение математического ожидания в дисперсии непрерывной случайной величины.
- В чем состоит преимущество среднего квадратического отклонения перед дисперсией?
- Чему равны математическое ожидание и дисперсия среднего арифметического одинаково распределенных независимых случайных величин?
- Дайте определения начального и центрального моментов случайной величины.

<u>**Тема 3.**</u> ***Основные законы распределения случайных величин***

Серия независимых испытаний Бернулли и биномиальный закон распределения вероятностей. Наивероятнейшее число. Предельные теоремы в схеме Бернулли (локальная и интегральная теоремы Муавра-Лапласа, а также теорема Пуассона) и их применения. Наиболее употребительные в социально-экономических приложениях модели законов распределения вероятностей и их основные свойства. Общая схема формирования, аналитические задания, графики и моменты законов распределения: нормальный (гауссовский) закон распределения и его значение в теории вероятностей; логарифмически нормальный; равномерный закон; экспоненциальный закон; распределения Стьюдента, χ^2 - Пирсона.

Литература:

[1, §12-14] , [2 ., п. 2], [3, гл. 4,6, зад. 166-187, 307-372], [4, гл. 6, 12, 13], [5; гл. 4]; [6, гл. 3, §3.2, 3.3, гл. 4, §4.1-4.3], [7, гл. 25, п. В], [18 т. 1-2], [20,р. 2-3, лаб. раб. 7,10,15,16].

При изучении этой темы необходимо обратить внимание на условия, когда применима та или иная теорема, что существенно при решении практических задач. Надо научиться уверенно находить вероятность отклонения частоты от вероятности случайного события.

Вопросы для самопроверки

- Дайте определение биномиального закона распределения вероятностей дискретной случайной величины.
- Как найти параметр λ распределения Пуассона?
- В чем состоит различие между локальной и интегральной теоремами Лапласа?
- Напишите дифференциальную функцию случайной величины, равномерно распределенной в интервале (а, Ь).
- Напишите дифференциальную функцию нормального распределения.

- Какими параметрами определяется нормальное распределение, каков их вероятностный смысл?
- Влияет ли изменение математического ожидания на форму нормальной кривой?
- Как влияет изменение среднего квадратического отклонения на форму нормальной кривой?
- Как вычислять вероятность попадания в заданный интервал нормально распределенной случайной величины.
- Случайные величины ξ_1 и ξ_2 распределены нормально, причем математические ожидания их равны а и дисперсия ξ_1 больше дисперсии ξ_2.
- Справедливо ли неравенство $P\{|\xi_1 - a| < \alpha\} < P\{|\xi_2 - a| < \alpha\}$?
- Среднее квадратическое отклонение случайной величины ξ равно σ. Абсолютная величина отклонения $|\xi - M(\xi)| = 10\sigma$. Можно ли считать, что величина ξ распределена нормально?
- Напишите дифференциальную и интегральную функции показательного распределения.
- Как найти математическое ожидание и дисперсию показательного распределения, зная параметр λ?
- Как найти распределение вероятностей функции одного случайного аргумента?
- Как найти распределение вероятностей суммы двух независимых случайных величин?
- Что называют композицией законов распределения?

<u>**Тема 4**</u> . ***Закон больших чисел и его следствие***

Задача оценки вероятностей заданных уклонений сл. в. От своих средних значений. Вывод и пояснение и точности неравенства Чебышева. Сущность закона больших чисел как выражения свойства устойчивости выборочного среднего значения (т-мы Чебышева, Маркова). Устойчивость относительных частот (т-ма Я. Бернулли) Особая роль нормального распределения. Формулировка центральной предельной теоремы (т-ма Ляпунова) и ее следствие (т-ма Муавра-Лапласса об асимптотической нормальности биномиальной сл.в.), а также ее использование в математической статистике.

Литература:

[1, §15-16] , [2 ., п. 5], [3, гл. 5, зад. 236-251], [4, гл. 9, 12, §8], [5; гл. 5]; [6, гл. 5], [18 т. 3], [20,п. 2.9].

Выражение закона больших чисел представлено в виде ряда теорем, которые указывают условия, при выполнении которых совокупное действие большого числа случайных причин приводит к результату, почти независимому от случая. Особо следует остановиться на важной роли теоремы Бернулли, которая позволяет нам находить вероятность события. До этой теоремы у нас не было средств определения этих вероятностей даже приближенно. Надо уяснить значение теоремы Чебышева для практики, обратить особое внимание на важное понятие сходимости по вероятности.

Вопросы для самопроверки

- Сформулируйте и докажите неравенство Чебышева.
- Почему неравенство Чебышева имеет для практики ограниченное значение?
- Каким условием должны удовлетворять случайные величины, чтобы к ним можно было применить теорему Чебышева?
- Сформулируйте и докажите частный случай теоремы Чебышева.
- Приведите примеры применения теоремы Чебышева на практике.
- Сформулируйте и докажите теорему Чебышева.
- В чем состоит различие величины рассеяния каждой из достаточно большого числа независимых случайных величин, имеющих ограниченные дисперсии, и их среднего арифметического?
- Сформулируйте теорему Бернулли и докажите ее как частный случай теоремы Чебышева.
- Сформулируйте теорему Бернулли, пользуясь понятием «сходимости по вероятности».
- Почему, исходя из теоремы Бернулли, нельзя заключить, что $\lim\limits_{n\to\infty}\frac{m}{n}=p$
- Сформулируйте частный случай теоремы Чебышева, пользуясь понятием «сходимости по вероятности».

<u>**Тема 5**</u> ***Элементы теории случайных процессов***

Определение случайного процесса (сл.пр.) и их классификация. Определение Марковского сл.пр. Цепи Маркова, примеры. Понятие матрицы вероятностей перехода, стохастической матрицы. Применение цепей Маркова для оценки эффективности функционирования систем. Элементы теории массового обслуживания. Простейшая математическая модель СМО.

Литература:

[2 ., п. 4.4], [3, гл. 16-17, зад. 756-918], [4, гл. 22-25], [6, гл. 6].

Вопросы для самопроверки

- Дайте определение случайного процесса.
- Реализация случайного процесса.

- Поток Пуассона и его свойства.
- Дайте определение простой однородной цепи Маркова.
- Что значит задать систему массового обслуживания?
- Основные характеристики СМО.

В данной теме важно уяснить определения случайного процесса и его задания, обратить внимание на классификацию случайных процессов.

Тема 6. *Основные понятия и задачи математической статистики*

Предмет и метод математической статистики (МС). МС и анализ данных. Генеральная и выборочная совокупности, основные способы организации выборки и ее характеристики. Методы статистического описания результатов наблюдений и графическое представление.

Понятие о статистической неизвестных параметров распределения. Числовые оценки параметров распределения. Основные свойства точечных оценок. Оценка матожидания и дисперсии по выборке. Понятие доверительного интервала. Доверительная вероятность. Методы получения оценок.

Литература:

[1, §17-20] , [2 ., п. 6], [3, гл. 9-11, зад. 439-449, 523-534], [4, гл. 15-17], [5; гл. 1,6]; [6, гл. 7], [8 лаб. раб.1], [18, т. 4-5], [20, п. 4.1-4.2].

Необходимо усвоить основные понятия и терминологию, которая принята в математической статистике. Обратить внимание на тот факт, что эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Следует научиться уверенно применять основные методы получения оценок, установить и хорошо усвоить различие между точечным и интервальным оцениванием, найти самостоятельно доверительные интервалы для оценки математического ожидания нормально распределенного признака, если известно среднее квадратическое отклонение значения признака генеральной совокупности от его среднего значения.

Вопросы для самопроверки

- В чем состоит различие между повторной и бесповторной выборками?
- В чем состоит характерная особенность графика эмпирической функция распределения?

- Дайте определение генеральной и выборочной средней, генеральной и выборочной дисперсии.
- Что такое групповая средняя?
- Как определяются несмещенная, эффективная и состоятельная оценки?
- Дайте определение групповой, внутригрупповой, межгрупповой и общей дисперсий.
- Сформулируйте теорему сложения дисперсий.
- Что называют «исправленной дисперсией»?
- Как найти доверительные интервалы для оценки математического ожидания нормально распределенного признака, если известном среднее квадратическое отклонение ?
- Что называют модой, медианой и размахом варьирования?

Тема 7. ***Статистическая проверка гипотез***

Понятие статистической гипотезы. Общая постановка задачи проверки гипотез. Виды статистических критериев. Проверка гипотез о законе распределения о значениях неизвестных параметров. Критерии согласия χ^2-Пирсона и Смирнова.

Литература:
[1, §21-27] , [2 ., п. 7-8], [3, гл. 10-13, зад. 450-522, 524-667], [4, гл. 16-19], [5; гл. 6,7]; [6, гл. 8-10], [8 лаб. раб. 2,3], [18, т. 6], [20, п. 4.3].

Вопросы для самопроверки

- Дайте определение статистической гипотезы.
- Приведите примеры простой и сложной гипотез.
- Приведите примеры нулевой и конструирующей гипотезы.
- Что называют ошибкой первого (второго) рода?
- Как находят критическую область?
- Как сравнивают среднее, дисперсии?
- Что называют критерием согласия?
- Как и для чего применяют критерий Пирсона?

Тема 8. ***Элементы дисперсионного анализа***

Основные понятия дисперсионного анализа. Случайная, детерминированная, смешанная модели. Формула разложения дисперсий. Общий метод проверки влияния фактора на признак способом сравнения дисперсий. Однофакторный, двухфакторный дисперсионный анализ.

Литература:
[2 ., п. 8], [3, гл. 14, зад. 668-678], [4, гл. 20], [6, гл. 12].

Вопросы для самопроверки

- Общая идея дисперсионного анализа.
- Построение однофакторного комплекса.
- Случай двухфакторного анализа.

Тема 9. *Основы регрессионного и корреляционного анализа*

О связях функциональных, стохастических, статистических и корреляционных. Определение формы связи. Понятие парной регрессии и множественной корреляции. Коэффициенты корреляции, детерминации и их свойства. Понятие о нелинейной регрессии.

Литература:
[1, §28-36] , [2 ., п. 9], [3, гл. 12, зад. 535-553], [4, гл. 18], [5; гл. 8]; [6, гл. 11], [8 лаб. раб. 4], [20, п. 4.4].

Вопросы для самопроверки

- Что такое условная средняя?
- Дайте определение корреляционной зависимости.
- В чем состоят две основные задачи теории корреляции?
- Какую корреляцию называют линейной?
- Выведите выборочные уравнения прямых регрессии.
- Дайте определение выборочного коэффициента корреляции и перечислите его свойства.
- Дайте определение корреляционного отношения и перечислите его свойства.
- В каком случае корреляцию называют криволинейной?
- Как называют корреляцию, если исследуется связь между несколькими признаками?

СПИСОК ЛИТЕРАТУРЫ

1. Бандура В.Н., Породников В.Д. Теория вероятностей и математическая статистика. - Донецк: ДонГУ, 1998.- 108с.
2. Булдык Г.М. Теория вероятностей и математическая статистика. - Минск: Высш. шк., 1998.- 285с.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высш. шк., 1977.- 400с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. -М.: Высш. шк., 1977.- 479с.
6. Колемаев В.А., Староверов О.В., Турундаевский А.Б. Теория вероятностей и математическая статистика. - М.: Высш. шк., 1991. -400с.
7. Кудрявцев В.А., Демидович Б.П. Краткий курс высшей математики. -М.: Наука, 1989.-656с.
8. Породников В.Д., Шкварченко Т.В. Лабораторный практикум по математической статистике. - Донецк: ДонГУ, - 1998. - 66с.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

9. Айвазян С.А., Енюков И.О., Мешалкин Л. Д. Прикладная **статистика. -** М.: Финансы и статистика, 1983. - 283 с.
10. Вентцель Е.С. Теория вероятностей. - М.: Наука, 1969. - 425 с.
11. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. - М.: Наука, 1973. -360 с..
12. Вентцель E.G., Овчаров А.А. Прикладные задачи теории вероятностей. - М.:Радио и связь, 1983. -230 с.
13. Виленкин Н.Я. Индукция, Комбинаторика. - М.: Просвещение, 1976.- 94 с.
14. Горский Е.И. Теория вероятностей с элементами математической статистики. - М.: В. шк., 1971. - 290 с.
15. Гутер Р.О., Овчинский Б.В, Основы теории вероятностей. - М.: Просвещение, 1967. -320с.
16. Ежов И.И., Скороход А.В., Ядренко М.И. Элементы комбинаторики. - М.: Наука, 1977. - 235 с.
17. Ефимов А.В. Сборник задач по математике. Т. 3. - М.: Наука, 1977. - 315с.

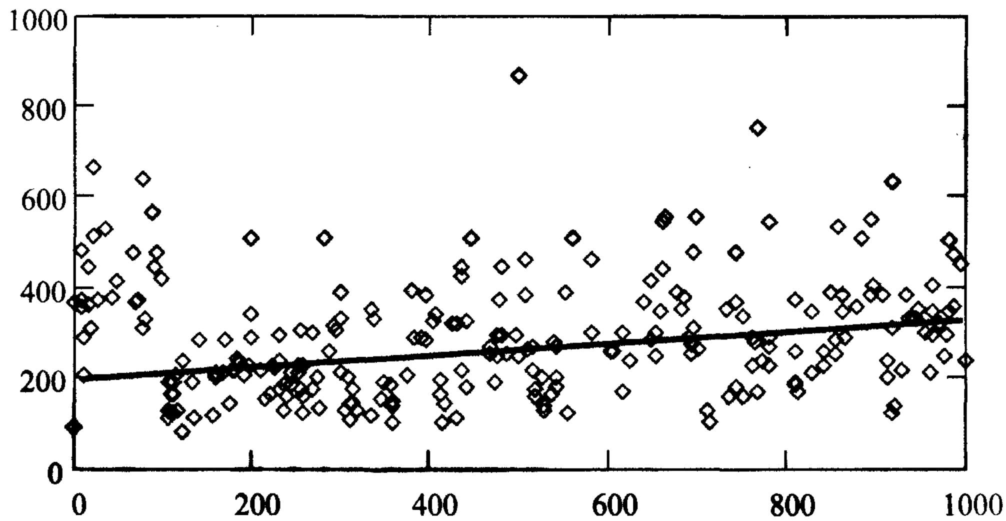
18. Породников В.Д., Румянцев Н.В., Шаташвили А.Д. Методические указания к решению задач по теории вероятностей и математической статистике. - Донецк, ДонГУ, 1986.-28 с.
19. Свешников А.А. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. - М.: Наука, 1970. - 655 с.
20. Сидорова В.М., Новожилова Е.Г. Методические рекомендации к элементам теории вероятностей и математической статистики. -Донецк, ДонГУ, 1990.-107с.
21. Чистяков В.П. Курс теории вероятностей. - М.: Наука, 1982. -.360 с.

Краткий курс

Серия: "Математика для экономистов"

В.Н. Бандура, В.Д. Породников

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА



Введение

Богатство и разнообразие применений теории вероятностей привлекают к ней многих людей. Одной из важнейших сфер приложения теории вероятностей является экономика. В настоящее время трудно представить исследование и прогнозирование экономических явлений без использования эконометрического моделирования, регрессионного анализа, трендовых и сглаживающей моделей и других методов, опирающихся на теорию вероятностей. Общая тенденция современной математики и ее приложений состоит в резком повышении роли тех разделов науки, которые анализируют явления, имеющие “случайный” характер, и основываются на теории вероятностей. Статистическая теория во многом основана на теории вероятностей, хотя здесь есть и обратная связь: при построении вероятностной модели также используются данные статистики. Теория вероятностей - это наука, а ее предмет- случайность, которая, казалось бы, не поддается никакому научному предсказанию. В дальнейшем мы увидим, что противоречие здесь кажущееся, так как теория вероятностей изучает свойства массовых случайных событий, способных многократно повторяться при воспроизведении определенного комплекса условий. Познавательная ценность теории вероятностей обусловлена тем, что массовые случайные явления в своем совокупном действии создают строгие закономерности.

Современная теория вероятностей начинается с установления аксиоматики, которую впервые в законченном виде сформулировал в 1933 году А.Н. Колмогоров в книге «Основные понятия теории вероятностей». Хотя возникновение теории вероятностей как науки относится к середине XVII века и связано с именами Паскаля, Ферма, Гюйгенса, но отдельные задачи, подсчета шансов в азартных играх, рассматривались ранее - XV-XVI в. итальянскими математиками Кардано, Пачоли, Тарталья и др.. Пожалуй, работы Паскаля и Ферма можно рассматривать лишь как предысторию теории вероятностей, а настоящая история начинается с закона больших чисел Я. Бернулли («Искусство предположений», 1713г.) и найденного вскоре Муавром нормального приближения к биномиальному распределению («Аналитическая смесь », 1730г.).

Дальнейшее развитие этой науки связано с именами Лапласа, Пуассона, Гаусса. Важный период в развитии теории вероятностей связан с именами П.Л. Чебышева, А.А. Маркова, А.М. Ляпунова, создавших эффективные методы предельных теорем. Первые попытки установления аксиоматики принадлежат С.Н. Бернштейну, Р. Мизесу, Э. Борелю.

Глава I
ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. СЛУЧАЙНЫЕ СОБЫТИЯ

1. Событие. Всевозможные результаты испытаний называются *элементарными событиями* ω. Совокупность всех элементарных событий $\Omega = \{\omega\}$ называется *множеством* (или *пространством*) элементарных событий.

В настоящем пособии будем считать, что пространство Ω конечно или счетно. В этом случае *событием* называется любое подмножество множества элементарных событий Ω. Говорят, что в результате испытания осуществилось событие А, если произошло элементарное событие $\omega \in A$. Пусть множество $\varnothing$ является *невозможным событием* Φ, которое при испытаниях произойти никогда не может. Само множество элементарных событий является *достоверным событием* Ω, которое произойдет в результате каждого испытания, потому что результат каждого испытания является элементарным событием. Остальные подмножества пространства элементарных событий Ω называются *случайными событиями* A, B, C, ... В результате испытания случайное событие может либо произойти, либо не произойти, при этом перед испытанием результат не определен.

2. Отношения между событиями. Отношения между событиями можно рассматривать как отношения между соответствующими подмножествами множества элементарных событий Ω.

1) **Следуемость**: *$A \subset B$, если из осуществления события A следует осуществление события B, или событие A влечет за собой событие B.*

2) **Равенство:** *$A = B$, если одновременно выполняются следующие условия: $A \subset B$ и $B \subset A$.*

3) **Сумма:** *$A \cup B$ выполняется тогда, когда происходит хотя бы одно из этих событий.*

4) **Произведение:** *$A \cap B$ выполняется тогда, когда происходят оба события (и A, и B).*

5) **Разность:** *$A \setminus B$ выполняется тогда, когда событие A происходит, а событие B не происходит.*

6) **Противоположность.** *Противоположное событие $\bar{A} = \Omega \setminus A$ осуществляется только тогда, когда не происходит событие A. Правильно и следующее утверждение: $\bar{\bar{A}} = A$.*

7) **Совместность.** *События называются совместными, если они могут одновременно произойти, и несовместными, если*

при осуществлении одного события другое произойти не может, т. е. они не могут произойти одновременно.

3. Системы событий. Множество событий $\{A_1, A_2, \ldots, A_n\}$ называется *системой событий* $\mathfrak{A}$, если это множество не содержит ни одной пары равных событий. Система событий $\mathfrak{A} = \{A_1, A_2, \ldots, A_n\}$ называется *системой несовместных событий*, если все события в системе попарно несовместны. Система несовместных событий называется *полной*, если сумма всех событий системы является достоверным событием Ω.

4. Классическая вероятность. Мера возможности наступления события называется его *вероятностью. Классическая вероятность* события A определяется как отношение количества элементарных событий m, входящих в состав события A, к количеству всех возможных элементарных событий n:

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

Вероятность невозможного события Φ равна нулю: $P(\Phi) = 0$, *вероятность достоверного события* Ω равна единице: $P(\Omega) = 1$, *а вероятность* произвольного *случайного события* A заключена между 0 и 1: $0 < P(A) < 1$.

Вероятность суммы событий $A \cup B$ определяется следующей формулой:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{I.2}$$

Если события несовместны, то формула (I.2) упрощается и принимает вид

$$P(A \cup B) = P(A) + P(B). \tag{I.3}$$

Эту формулу можно обобщить на случай суммы любого количества несовместных событий. Учитывая, что $A \cup \bar{A} = \Omega$, и принимая во внимание формулу (I.3), получаем

$$P(A) + P(\bar{A}) = 1. \tag{I.4}$$

События являются *независимыми*, если вероятность происхождения одного события не зависит от того, произошло ли другое событие или нет. При зависимых событиях A и B имеет смысл говорить об *условной вероятности* $P(A|B)$ события A при условии, что событие B уже произошло. При независимых событиях условная вероятность равна обычной вероятности $P(A|B) = P(A)$.

Вероятность произведения событий $A \cap B$ выражается следующей формулой:

$$P(A \cap B) = P(A|B) \cdot P(B) \quad \text{или} \quad P(A \cap B) = P(B|A) \cdot P(A). \tag{I.5}$$

Если события независимы, то формула (I.5) упрощается и принимает вид

$$P(A \cap B) = P(A) \cdot P(B). \tag{I.6}$$

Эту формулу можно обобщить на случай произведения любого количества независимых событий.

Если событие A зависит от событий (гипотез) полной системы событий $\mathscr{B} = \{B_1, B_2, \ldots, B_n\}$, то вероятность события A вычисляется по формуле полной вероятности

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i). \qquad (I.7)$$

В этом случае вместе с осуществлением события A происходит одно и только одно событие из системы $\mathscr{B}$.

Если событие A произошло, то можно вычислить условную вероятность того, что вместе с событием A осуществлялась гипотеза B_i:

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{P(A)}, \qquad (I.8)$$

где $P(A)$ — полная вероятность события A. Полученную формулу называют *формулой Бейеса*. С помощью формулы Бейеса можно после испытания уточнить вероятность происхождения гипотезы B_i. Сумма вероятностей гипотез B_i должна быть равна единице.

5. Понятия комбинаторики. При решении задач теории вероятностей часто используют следующие понятия комбинаторики: *перестановка, сочетание* и *размещение*, а также *правило умножения* и *правило сложения*.

Пусть дано множество $\mathscr{N}$ из n объектов. Всевозможные последовательности из всех n объектов называют **перестановками**. Общее число P_n различных перестановок из n объектов вычисляют по формуле

$$P_n = n! \qquad (I.9)$$

где $n! = 1 \cdot 2 \ldots n$, при этом считают, что $0! = 1$.

Сочетаниями называют подмножества множества $\mathscr{N}$. Общее число различных сочетаний C_n^m из n объектов по m вычисляют по формуле

$$C_n^m = \frac{n!}{m!(n-m)!}. \qquad (I.10)$$

Размещениями называют упорядоченные последовательности объектов подмножеств множества $\mathscr{N}$. Общее число различных размещений A_n^m из n объектов по m вычисляют по формуле

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1)\ldots(n-m+1). \qquad (I.11)$$

Правило умножения. Если требуется выполнить одно за другим какие-то k действий, которые можно выполнить соответственно $n_1, n_2, \ldots, n_k$ способами, то все k действий вместе могут быть выполнены $n_1 \cdot n_2 \ldots n_k$ способами.

Правило сложения. Если два взаимно исключающие друг друга действия могут выполняться соответственно m или n способами, то выполнить одно любое из этих действий можно $m + n$ способами.

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

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

Решение. Общее число возможных элементарных исходов испытания равно числу способов, которыми можно извлечь 4 карточки из 5, т. е. равно A_5^4 — числу размещений из 5 элементов по 4. Благоприятствует появлению слова «юрта» лишь один исход.

Искомая вероятность равна отношению числа исходов, благоприятствующих появлению события, к числу всех элементарных исходов:

$$p = \frac{1}{A_5^4} = \frac{1}{120}.$$

Задача 1. Вероятность поражения мишени при одном выстреле первым стрелком равна 0,8, а вторым стрелком — 0,9. Найти вероятность того, что оба стрелка поразят мишень.

Решение. События A (первый стрелок поразил мишень) и B (второй стрелок поразил мишень) независимые. Искомая вероятность того, что оба стрелка поразят мишень по теореме умножения вероятностей независимых событий равна

$$P(A \cdot B) = P(A) \cdot P(B) = 0{,}8 \cdot 0{,}9 = 0{,}72.$$

Задача 2. Для некоторой местности среднее число ясных дней в июле равно 25. Найти вероятность того, что первые два дня июля будут ясными.

Решение. Вероятность того, что первого июля будет ясный день (событие A), равна

$$P(A) = \frac{25}{31}.$$

Вероятность того, что второго июля будет ясный день (событие B), при условии, что первого июля также был ясный день, т. е. условная вероятность события B равна

$$P_A(B) = \frac{24}{30} = \frac{4}{5}.$$

Искомая вероятность того, что первые два дня июля будут ясными по теореме умножения вероятностей зависимых событий равна

$$P(A \cdot B) = P(A) \cdot P_A(B) = \frac{25}{31} \cdot \frac{4}{5} = \frac{20}{31}.$$

§ 2. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

1. Повторные испытания. Предположим, что событие A происходит в результате n независимых испытаний, притом в каждом испытании вероятность события A постоянна и равна p. Результатом каждого испытания является либо событие A, либо событие $\bar{A}$. Последнее происходит с вероятностью $q = 1 - p$.

Если рассматривать все n испытаний как одно испытание, то его результатом является произведение событий A и $\bar{A}$. Здесь ввиду независимости исходных испытаний важен не порядок событий, а число повторений события A. Частоту события A обозначим через k, $0 < k < n$. Вероятность появления события A k раз вычисляют по *формуле Бернулли:*

$$p_k = C_n^k p^k q^{n-k}. \tag{I.12}$$

Если нужно вычислить вероятности для всех значений k, $0 < k < n$, то можно воспользоваться формулой, с помощью которой p_k вычисляется по значению p_{k-1}:

$$p_k = \frac{n-k+1}{k} \cdot \frac{p}{q} \cdot p_{k-1}, \quad k = 1, \ldots, n. \tag{I.13}$$

Тогда p_0 следует вычислять по формуле (I.12), которая при $k = 0$ принимает вид $p_0 = q^n$, а все остальные p_k — по формуле (I.13). При больших значениях n и k вычисления по формуле Бернулли достаточно громоздки и, кроме того, на практике обычно не требуется такая высокая точность. Поэтому разработаны довольно точные приближенные методы вычисления вероятности p_k.

Иногда находят *наивероятнейшую частоту,* т. е. частоту, имеющую максимальную вероятность. Наивероятнейшая частота находится в интервале $np - q \leqslant k \leqslant np + p$. Длина этого интервала равна единице, поэтому если границы интервала — целые числа, то имеются две наивероятнейшие частоты, в противном случае — только одна.

2. Формулы Муавра — Лапласа и Пуассона. Вместо формулы Бернулли (I.12) можно использовать **локальную теорему Муавра — Лапласа:**

если при n независимых испытаниях событие A происходит с постоянной вероятностью p, которая не очень близка к нулю и единице ($0 < p < 1$), то при достаточно большом количестве испытаний n вероятность того, что событие A произойдет k раз, приближенно равна

$$P_n(k) \approx \frac{f(x)}{\sqrt{npq}}, \tag{I.14}$$

где $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x = \dfrac{k - np}{\sqrt{npq}}$.

Функция $f(x)$ — четная ($f(-x) = f(x)$) и принимает только неотрицательные значения (рис. 1). Для нее составлены таблицы (см. приложение 2). Так как график функции симметричен относительно оси ординат, то таблицы составлены только для положительных значений аргумента.

Если вероятность p реализации события A близка к нулю, то следует использовать следующую **теорему Пуассона**, которая в этом случае дает большую точность.

если при n независимых испытаниях событие A происходит с вероятностью p, близкой к 0, то при достаточно большом n вероятность осуществления события A k раз приближенно равна

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad (I.15)$$

где $\lambda = np$.

Для функции формулы (I.15) также составлены таблицы (см. приложение 1).

Часто нужно найти вероятность того, что частота появления события A находится в каком-то интервале. Эту проблему позволяет решить **интегральная теорема Муавра—Лапласа**:

если в n независимых испытаниях событие A происходит с постоянной вероятностью p, которая отличается от 0 и 1, то при достаточно большом n вероятность того, что частота k события A находится в интервале $[a, b]$, приближенно равна

$$P_n(a \leqslant k \leqslant b) \approx \Phi(x_2) - \Phi(x_1), \qquad (I.16)$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad x_1 = \frac{a-np}{\sqrt{npq}}, \quad x_2 = \frac{b-np}{\sqrt{npq}}.$$

Функция $\Phi(x)$ является интегралом от функции $f(x)$ [см. формулу (I.14)] и принимает значения в интервале [0, 1], при этом $\Phi(-\infty) = 0$; и $\Phi(\infty) = 1$ и $\Phi(0) = 0{,}5$. График функции $\Phi(x)$ приведен на рис. 2. Для функции $\Phi(x)$ составлены таблицы (см. приложение 3). Таблицы составлены только для положительных значений аргумента. Для отрицательных аргументов значения функции можно получить из этой же таблицы, используя соотношение

$$\Phi(-x) = 1 - \Phi(x). \qquad (I.17)$$

Иногда нужно решить следующую задачу. В n независимых испытаниях событие A происходит с постоянной вероятностью p. Найти вероятность того, что относительная частота k/n события A отличается от вероятности события A по абсолютной величине не больше чем на $\varepsilon > 0$. Решение этой задачи сводится к использованию интегральной формулы Муавра—Лапласа (I.16), с помощью которой для решения данной задачи получаем следующую формулу:

$$P_n\left(\left|\frac{k}{n} - p\right| \leqslant \varepsilon\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - 1. \qquad (I.18)$$

3. Понятие случайной величины. *Случайной величиной* X называют величину, которая случайно принимает какое-то значение

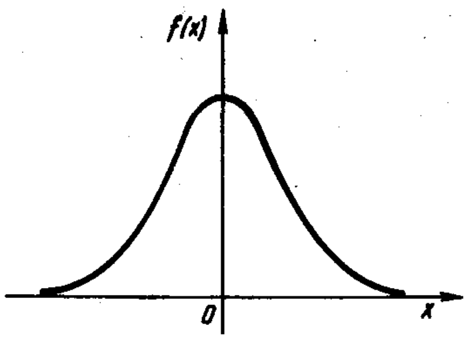


Рис. 1. График функции $f(x)$

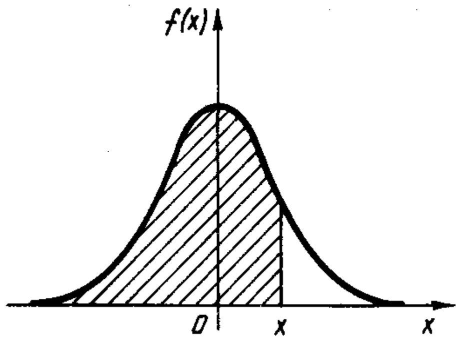


Рис. 2. Геометрический смысл функции $\Phi(x)$

из совокупности своих значений, ее *закон распределения* может быть задан *функцией распределения*

$$F(x) = P(X < x). \tag{I.19}$$

Функция распределения $F(x)$ — неубывающая, непрерывная слева функция, определенная на всей числовой оси, при этом $F(-\infty) = 0$ и $F(\infty) = 1$. Случайные величины будем обозначать обычно последними буквами латинского алфавита: X, Y, Z.

Практически случайную величину можно получить, сопоставляя событиям из полной системы событий вещественных чисел. Совокупность этих чисел образует совокупность значений случайной величины. По совокупностям значений различают случайные величины двух видов: дискретные и непрерывные. *Дискретными* называют случайные величины, значениями которых являются только отдельные точки числовой оси. Значениями непрерывной случайной величины могут быть любые точки какого-то интервала на числовой оси.

Закон распределения дискретной случайной величины X можно определить с помощью ряда распределения, заданного в виде следующей таблицы:

X	x_1	x_2	x_3	...	x_n
P	p_1	p_2	p_3	...	p_n

В первой строке этой таблицы указаны все значения x_i дискретной случайной величины X, а во второй строке — вероятности p_i принятия случайной величиной соответствующих значений x_i. Сумма всех вероятностей равна единице.

На основе ряда распределения можно получить *функцию распределения дискретной случайной величины* X. Эта функция выражается следующей формулой:

$$F(x) = \sum_{x_i < x} p_i. \tag{I.20}$$

Формулу (I.20) можно записать в следующем виде, наглядно иллюстрирующем непрерывность слева функции распределения:

$$F(x) = \begin{cases} 0, & x \leqslant x_1, \\ p_1, & x_1 < x \leqslant x_2, \\ p_1 + p_2, & x_2 < x \leqslant x_3, \\ \cdots\cdots\cdot \\ \sum\limits_{i=1}^{n-1} p_i, & x_{n-1} < x \leqslant x_n, \\ 1, & x > x_n. \end{cases} \tag{I.21}$$

График функции распределения дискретной случайной величины обычно представляет собой ступенчатую линию (рис. 3).

Закон распределения непрерывной случайной величины задается или *функцией распределения*, или *функцией плотности вероятности*. Функция распределения непрерывной случайной величины X представляется в виде интеграла

$$F(x) = \int\limits_{-\infty}^{x} f(t)\mathrm{d}t, \tag{I.22}$$

где $f(x) > 0$ — функция плотности вероятности. График этой функции всегда охватывает фигуру, площадь которой равна единице. Это следует из свойства функции распределения: $F(\infty) = 1$, так как

$$F(\infty) = \int\limits_{-\infty}^{\infty} f(t)\mathrm{d}t = 1. \tag{I.23}$$

Формула (I.22) определяет площадь под графиком функции $f(x)$ в интервале $]-\infty, x]$ (рис. 4).

Если заданы два значения x_1 и x_2 непрерывной случайной величины $X(x_1 < x_2)$, то вероятность того, что X принимает значение в интервале $[x_1, x_2]$, равна

$$P(x_1 \leqslant X \leqslant x_2) = \int\limits_{x_1}^{x_2} f(x)\mathrm{d}t. \tag{I.24}$$

Это можно доказать с помощью формул (I.19) и (I.22) (см. рис. 4). Подставив $x_1 = x_2$, получим

$$P(x_1 = X = x_2) = \int\limits_{x_1}^{x_1} f(x)\mathrm{d}t = 0. \tag{I.25}$$

Из формул (I.22) и (I.25) видно, что значение функции плотности вероятности $f(x)$ не точно совпадает с вероят-

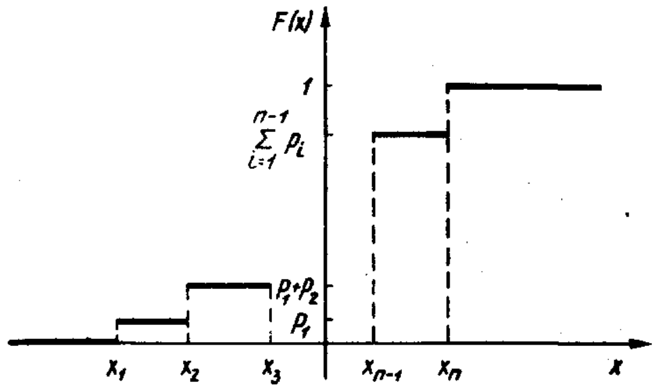


Рис. 3. График функции распределения $F(x)$ дискретной случайной величины X

ностью принятия значения дискретной случайной величины. Неравенство $f(x_1) > f(x_2)$ (см. рис. 4) означает, что при достаточно большом количестве испытаний вблизи точки x_1 окажется больше значений случайной величины X, чем вблизи точки x_2. Если все полученные значения изобразить в виде точек на числовой оси, то вокруг точки x_1 они будут находиться плотнее, чем вокруг точки x_2. Значит, чем больше значение функции плотности вероятности, тем больше вероятность того, что случайная величина примет значение вблизи этой точки (но не обязательно в самой точке).

График функции распределения $F(x)$ непрерывной случайной величины представляет собой непрерывную кривую (рис. 5). Из формулы (I.22) следует, что

$$f(x) = F'(x). \tag{I.26}$$

Случайные величины X и Y являются *независимыми*, если при всех парах чисел (x, y) независимы и соответствующие события $(X < x)$ и $(Y < y)$.

4. Числовые характеристики случайной величины. 4.1. Среднее значение. Среднее значение дает чаще всего серединное значение совокупности значений случайной величины и обозначается через EX или $E(X)$. Учитывая, что значения случайной величины разбросаны вблизи среднего значения, среднее значение иногда называют математическим ожиданием (обозначение: $M(X)$).

Среднее значения дискретной случайной величины вычисляют по следующей формуле:

$$EX = \sum_{i=1}^{n} x_i \cdot p_i, \tag{I.27}$$

где x_i — значение дискретной случайной величины, p_i — вероятность того, что случайная величина примет значение x_i, n — количество значений случайной величины, которое может быть равно и ∞, но тогда ряд (I.27) должен абсолютно сходиться.

Среднее значение непрерывной случайной величины вычисляется по формуле

$$EX = \int_{-\infty}^{\infty} x \cdot f(x) \mathrm{d}x, \tag{I.28}$$

где $f(x)$ — функция плотности вероятности случайной величины и интеграл сходится абсолютно.

Среднее значение непрерывной случайной величины имеет следующие свойства, общие как для дискретных, так и для непрерывных случайных величин.

1) $EC = C$, *если* C *— постоянная, т. е. среднее значение постоянной величины равно константе.*

2) $E(CX) = C \cdot EX$, *т. е. постоянную можно выносить из-под знака среднего значения.*

3) $E(X + Y) = EX + EY$. *Среднее значение суммы случайных величин равно сумме средних значений случайных величин.*

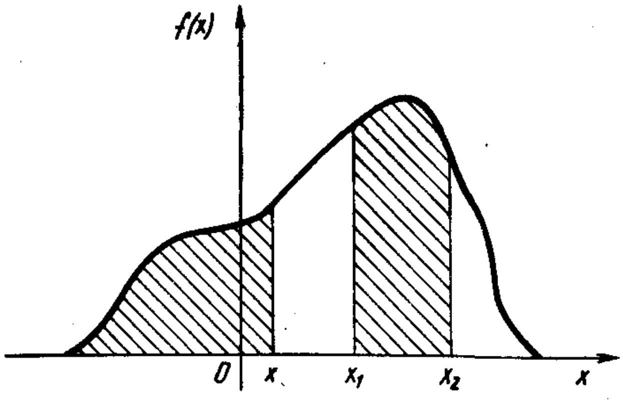


Рис. 4. График функции плотности вероятности $f(x)$ непрерывной случайной величины X

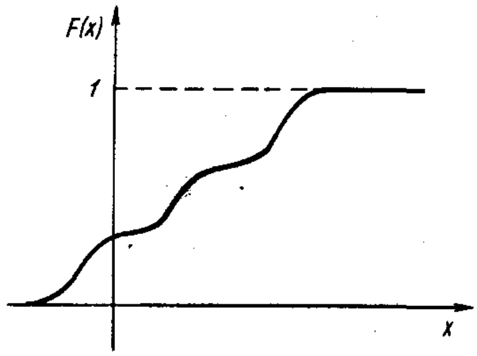


Рис. 5. График функции распределения $F(x)$ непрерывной случайной величины X

4) *Если случайная величина* $X \geqslant 0$, *то и среднее значение* $EX \geqslant 0$.

5) *Если случайные величины* X *и* Y *независимы, то* $E(X \cdot Y) = EX \cdot EY$, *т. е. среднее значение произведения независимых случайных величин равно произведению средних значений случайных величин.*

4.2. Дисперсия. Дисперсия характеризует степень рассеяния значений случайной величины от своего среднего значения и обозначается через DX или $D(X)$. *Дисперсия* определяется как среднее значение квадрата отклонений случайной величины от своего среднего значения

$$DX = E(X - EX)^2. \qquad (I.29)$$

Если эту формулу преобразовать, используя свойства среднего значения, то получаем формулу, которую обычно используют при вычислении дисперсии

$$DX = E(X^2) - (EX)^2. \qquad (I.30)$$

Дисперсия дискретной случайной величины вычисляется [на основе формул (I.27), (I.29) и (I.30)] по следующей формуле:

$$DX = \sum_{i=1}^{n} (x_i - EX)^2 \cdot p_i \qquad (I.31)$$

или

$$DX = \sum_{i=1}^{n} x_i^2 \cdot p_i - (EX)^2. \qquad (I.32)$$

Аналогично, согласно формулам (I.28) — (I.30), получаем формулы для вычисления дисперсии непрерывной случайной величины:

$$DX = \int_{-\infty}^{\infty} (x - EX)^2 f(x) \mathrm{d}x, \qquad (I.33)$$

или

$$DX = \int\limits_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - (EX)^2. \qquad (I.34)$$

Дисперсия имеет следующие свойства, которые можно доказать на основании свойств среднего значения.

1) *Дисперсия константы равна нулю:* $DC = 0$.

2) *Дисперсия всегда неотрицательна:* $DX \geqslant 0$.

3) *Постоянную можно вынести из-под знака дисперсии, возведя предварительно ее в квадрат:* $D(CX) = C^2 \cdot DX$.

4) *Изменение случайной величины на постоянную не изменяет ее дисперсию:* $D(C + X) = DX$.

5) *Если случайные величины X и Y независимы, то* $D(X \pm Y) = DX + DY$. *Дисперсия суммы или разности независимых случайных величин равна сумме дисперсий этих случайных величин.*

4.3. Среднее квадратичное отклонение. *Средним квадратичным отклонением* называется квадратный корень из дисперсии $\sqrt{DX}$. Среднее квадратичное отклонение означает абсолютное среднее отклонение случайной величины от своего среднего значения.

4.4. Моменты. *Моментом порядка r* называется величина

$$m_r = EX^r. \qquad (I.35)$$

Центральным моментом порядка r называется величина

$$\bar{m}_r = E(X - EX)^r. \qquad (I.36)$$

Из этих определений видно, что среднее значение является моментом первого порядка ($m_1 = EX^1$), а дисперсия — центральным моментом второго порядка ($\bar{m}_2 = E(X - EX)^2 = DX$).

4.5. Мода. *Модой (Mo)* называется значение случайной величины, которое встречается чаще всего, т. е. имеет максимальную вероятность (для дискретной случайной величины) или максимум функции плотности вероятности в данной точке (при непрерывной случайной величине).

Одна и та же случайная величина может иметь одну или несколько мод. Однако возможно, что случайная величина и не имеет моды (если все ее значения имеют одинаковую вероятность (равномерное распределение)).

4.6. Медиана. Определим сначала понятие *квантиля непрерывной случайной величины.* Корень уравнения $F(x) = p$, где $F(x)$ — функция распределения и $0 < p < 1$, называется p-квантилем x_p (рис. 6.); 1/2-квантиль называется *медианой (Me)*. Учитывая определение функции распределения $F(x)$ [формулу (I.19)], получаем $P(X < Me) = 1/2$ и отсюда $P(X > Me) = 1/2$. Таким образом, медиана делит область значений случайной величины на две равные по вероятности части.

5. Биномиальное распределение. Случайная величина, имеющая биномиальное распределение, получается при повторных независимых испытаниях (см. п. 1). Значениями случайной величины X

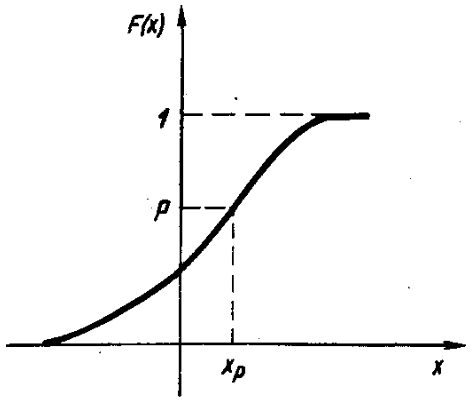


Рис. 6. p-квантиль

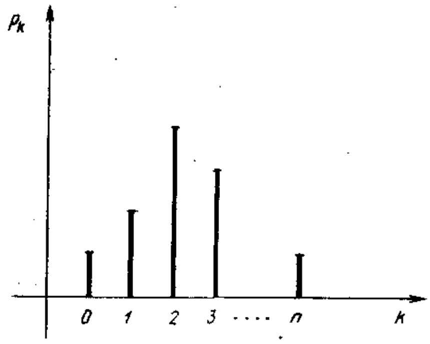


Рис. 7. График вероятностей значений случайной величины X с биномиальным распределением

являются частоты события A при независимых испытаниях, т. е. целые числа в интервале $[0, n]$. Это означает, что случайная величина с биномиальным распределением *дискретна.*

Вероятность каждого значения вычисляется по формуле Бернулли (I.12). Согласно формуле (I.20), можно записать *функцию распределения биномиальной случайной величины:*

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ \sum\limits_{k<x} C_n^k p^k q^{n-k}, & 0 < x \leqslant n, \\ 1, & x > n. \end{cases} \qquad (I.37)$$

График функции распределения похож на график, изображенный на рис. 3. Часто строят график вероятностей значений случайной величины с биномиальным распределением (рис. 7). Параметрами биномиального распределения являются n и p. Утверждение, что случайная величина X имеет биномиальное распределение с параметрами n и p, можно более кратко записывать в виде $X \in B(n, p)$. *Среднее значение* биномиального распределения $EX = np$ и *дисперсия* $DX = npq$. Модой является наивероятнейшая частота (см. п. 1).

6. Распределение Пуассона. Случайная величина, имеющая распределение Пуассона, принимает значения 0, 1, 2, ..., n, причем вероятность p_k того, что она принимает значения $k \geqslant 0$, вычисляется по формуле Пуассона (I.15). Ее *функция распределения* аналогично формулам (I.37) определяется соотношением

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ \sum\limits_{k<x} \frac{\lambda^k}{k!} e^{-\lambda}, & 0 < x \leqslant n, \\ 1, & x > n, \end{cases} \qquad (I.38)$$

где $\lambda = \text{const}$, $\lambda > 0$. Параметром распределения Пуассона является величина λ. Этому параметру равны и среднее значение, и дисперсия.

7. Равномерное распределение. Случайная величина X, имеющая равномерное распределение, принимает значения в интервале $[a, b]$, ее функция плотности вероятности $f(x)$ в этом интервале постоянна. По условию (I.23) можно определить эту константу и записать функцию плотности вероятности равномерного распределения:

$$f(x) = \begin{cases} 0, & x \leqslant a, \\ \frac{1}{b-a}, & a < x \leqslant b, \\ 0, & x > b. \end{cases} \tag{I.39}$$

Функцию распределения можно найти по формуле (I.22):

$$F(x) = \begin{cases} 0, & x \leqslant a, \\ \frac{x-a}{x-b}, & a < x \leqslant b, \\ 1, & x > b. \end{cases} \tag{I.40}$$

Графики этих функций изображены на рис. 8 и 9. *Среднее значение* EX получаем по формуле (I.28): $EX = (a + b)/2$, а *дисперсию* DX — по формуле (I.34), $DX = (b - a)/12$. Равномерное распределение не имеет моды, а медиана совпадает со средним значением.

8. Нормальное распределение. Нормальное распределение является самым распространенным распределением в природе, экономике и т. д. Случайная величина с нормальным распределением может принимать любые значения в интервале $]-\infty, +\infty[$ и имеет *функцию плотности вероятности*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{I.41}$$

где μ и σ — параметры нормального распределения, при этом $\sigma > 0$.

Как показывает исследование функции $f(x)$, функция определена на всей числовой оси, все ее значения неотрицательны, при $|x| \to \infty$ значения функции уменьшаются $f(x) \to 0$, т. е. ось x является асимптотой функции $f(x)$. Функция $f(x)$ достигает в точке $x = \mu$ максимума, равного $1/\sigma\sqrt{2\pi}$, и имеет точки перегиба в точках $x_1 = \mu - \sigma$ и $x_2 = \mu + \sigma$.

При изменении значения μ график функции $f(x)$ «жестко» смещается вдоль оси x (рис. 10). При изменении значения σ изменяется и вид графика: при увеличении значения σ в m раз максимальное значение функции уменьшается в m раз и график «вытягивается» в обе стороны вдоль оси x. При уменьшении значения σ происходит обратное (рис. 11).

На основании формул (I.22) и (I.41) получаем *функцию распределения* $\Phi(x)$ нормального распределения

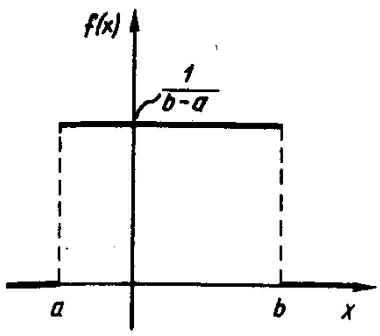


Рис. 8. График функции плотности вероятности $f(x)$ равномерного распределения

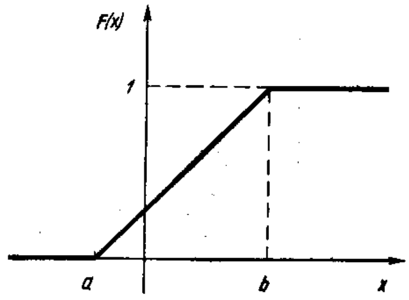


Рис. 9. График функции распределения $F(x)$ равномерного распределения

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (I.42)$$

По параметрам нормального распределения вычисляют и все числовые характеристики нормального распределения, а именно, $EX = \mu$, $DX = \sigma^2$, σ является средним квадратичным отклонением: $Mo = Me = \mu$. Утверждение «случайная величина X имеет нормальное распределение с параметрами μ и σ » кратко записывается так: $X \in N(\mu, \sigma)$.

Особое значение среди нормальных распределений имеет *нормированное нормальное распределение* с параметрами 0 и 1: $X \in N(0, 1)$. Если подставить $\mu = 0$ и $\sigma = 1$ в формулу (I.41) и (I.42), то получим уже знакомые формулы для $f(x)$ и $\Phi(x)$ [см. формулы (I.14) и (I.16)]. Для этих функций составлены таблицы (см. приложения 2 и 3). От произвольного нормального распределения $X \in N(\mu, \sigma)$ можно перейти к нормированному нормальному распределению, воспользовавшись заменой переменных

$$z = \frac{x - \mu}{\sigma}. \quad (I.43)$$

Случайные величины с нормальным распределением используют при решении задач двух типов. *Первая задача:* найти вероятность

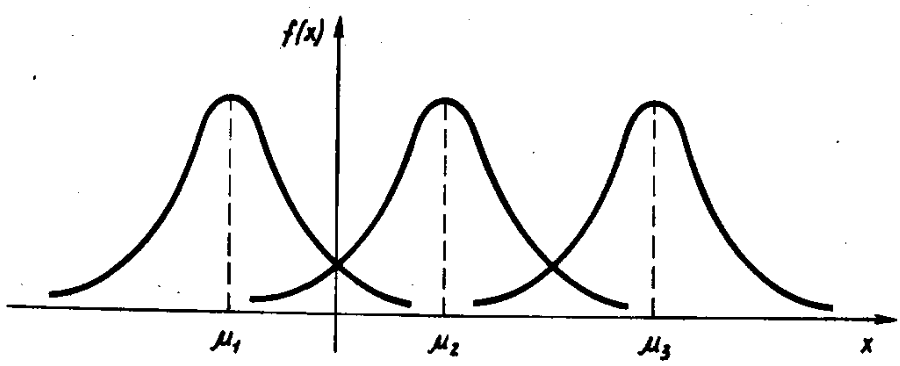


Рис. 10. Смещение графика $f(x)$ при изменении значения параметра μ

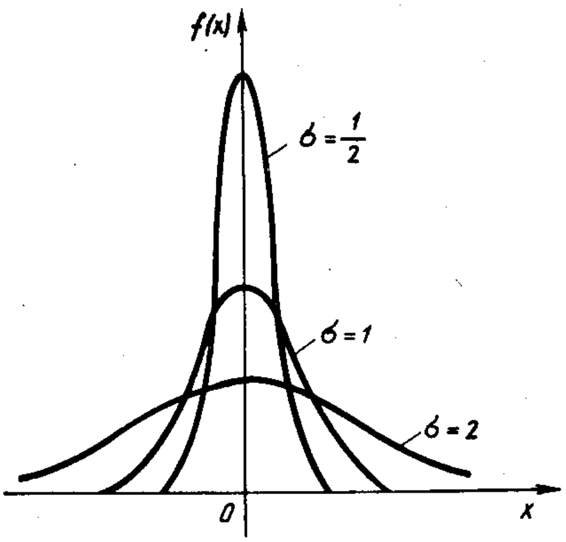


Рис. 11. Изменение графика $f(x)$ при изменении значения параметра σ

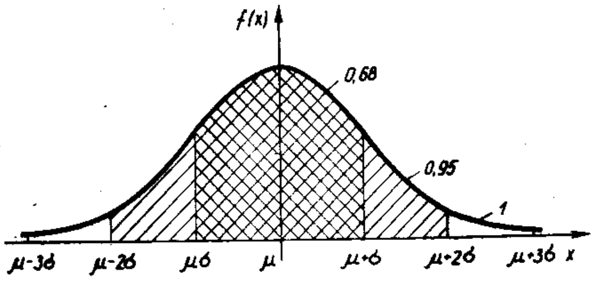


Рис. 12. Правило «трех сигм»

того, что случайная величина $X \in N(\mu, \sigma)$ принимает значение в интервале $[a, b]$. Найдем эту вероятность по формуле

$$P(a \leqslant X \leqslant b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \tag{I.44}$$

где $\Phi(x)$ — функция распределения нормированного нормального распределения [см. формулу (I.14)].

Вторая задача: найти вероятность того, что случайная величина $X \in N(\mu, \sigma)$ отличается от своего среднего значения μ по абсолютной величине не больше чем на

$$P(|X-\mu| \leqslant \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1. \tag{I.45}$$

Эта формула следует из формулы (I.44).

Если $\varepsilon = \sigma$, то по формуле (I.45) получаем $P = 0{,}68268$; если $\varepsilon = 2\sigma$, то $P = 0{,}95450$, если $\varepsilon = 3\sigma$, то $P = 0{,}99730 \approx 1$. Таким образом, случайная величина X с нормальным распределением практически не принимает значений, которые отличались бы от среднего значения по абсолютной величине больше чем на 3σ. Это утверждение называется *правилом «трех сигм»* (рис. 12).

Формулы (I.44) и (I.45) похожи на интегральные формулы Муавра — Лапласа (I.16) и (I.18), от которых можно перейти к первым формулам, заменяя параметры $EX = np$ на μ и $\sqrt{DX} = \sqrt{npq}$ на σ. При достаточно большом n такой переход от биномиального распределения к нормальному основывается на законе больших чисел.

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Задача. Вероятность появления события в каждом из 100 независимых испытании равна 0,9. Найти вероятность того, что относительная частота отклонится от вероятности появления события (по абсолютной величине) не более чем на 0,03.

Решение. Обозначим искомую вероятность через P. Воспользуемся формулой (§ 5)

$$P \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

По условию $n=100$, $\varepsilon=0{,}03$, $p=0{,}9$, $q=1-p=1-0{,}9=0{,}1$. Следовательно,

$$P \approx 2\Phi\left(0{,}03 \sqrt{\frac{100}{0{,}9\cdot 0{,}1}}\right) = 2\Phi(1).$$

По таблице (приложение 2 учебника [1]) найдем $\Phi(1)=0{,}3413$. Искомая вероятность

$$P = 2\cdot 0{,}3413 = 0{,}6826.$$

При решении задач на повторные независимые испытания, в которых вероятности появления события различны, удобно пользоваться производящей функцией вероятностей $P_n(k)$ (через $P_n(k)$ обозначена вероятность того, что в n испытаниях событие появится ровно k раз).

Пусть вероятность появления события в первом испытании равна p_1, во втором — p_2, ..., в n-м — p_n.

Производящей функцией вероятностей $P_n(k)$ называют функцию, определяемую равенством

$$\varphi_n(z) = (p_1 z + q_1)(p_2 z + q_2)\ldots(p_n z + q_n).$$

Пусть производят ряд испытаний, причем вероятность появления события в первом испытании равна p_1, во втором — p_2 и т. д. Тогда вероятность $P_n(k)$ того, что при n испытаниях события появятся ровно k раз, равна коэффициенту при z^k в разложении производящей функции по степеням z. Например, если $n=2$, то

$$\varphi_2(z) = (p_1 z + q_1)(p_2 z + q_2) = p_1 p_2 z^2 + (p_1 q_2 + p_2 q_1) z + q_1 q_2.$$

Здесь коэффициент $p_1 p_2$ при z^2 равен вероятности $P_2(2)$ того, что в двух испытаниях событие появится ровно два раза; коэффициент $p_1 q_2 + p_2 q_1$ при z равен вероятности $P_2(1)$ того, что событие появится ровно один раз; коэффициент при z^0, т. е. свободный член $q_1 q_2$, равен вероятности $P_2(0)$ того, что событие не появится ни одного раза.

Задача 2. Устройство состоит из двух независимо работающих элементов. Вероятность безотказной работы (за время t) первого элемента $p_1=0{,}8$, а второго $p_2=0{,}9$. Найти вероятности того, что за время t будут работать безотказно: а) 2 элемента, б) 1 элемент, в) ни один из элементов.

Решение. Так как вероятности безотказной работы элементов равны соответственно 0,8 и 0,9, то вероятности того, что элементы откажут, равны $q_1=1-0{,}8=0{,}2$; $q_2=1-0{,}9=0{,}1$.

Составим производящую функцию

$$\varphi_2(z) = (0{,}8z + 0{,}2)(0{,}9z + 0{,}1) = 0{,}72z^2 + 0{,}26z + 0{,}02.$$

Вероятность того, что два элемента будут работать безотказно равна коэффициенту при z^2:

$$P_2(2) = 0{,}72.$$

Вероятность того, что 1 элемент будет работать безотказно, равна коэффициенту при z:

$$P_2(1) = 0{,}26.$$

Вероятность того, что ни один из элементов не будет работать безотказно, равна свободному члену:

$$P_2(0) = 0{,}02.$$

Контроль: $0{,}72+0{,}26+0{,}02=1$.

Задача 1. Среднее число заявок, поступающих на предприятие бытового обслуживания за 1 ч, равно трем. Найти вероятность того, что за 2 ч поступит 5 заявок. Предполагается, что поток заявок простейший.

Решение. По условию $\lambda=3$, $t=2$, $k=5$. Воспользуемся формулой

$$P_t(k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Искомая вероятность того, что за 2 ч поступит 5 заявок, равна

$$P_2(5)=\frac{6^5 e^{-6}}{5!}=\frac{6^5\cdot 0{,}00248}{120}\approx 0{,}268.$$

Задача 2. Среднее число заявок, поступающих на АТС в 1 мин, равно двум. Найти вероятность того, что за 4 мин поступит: 1) три вызова, 2) менее трех вызовов, 3) не менее трех вызовов.

Решение. 1. По условию $\lambda=2$, $t=4$, $k=3$. Воспользуемся формулой

$$P_t(k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Подставив данные, получим

$$P_4(3)=\frac{8^3 e^{-8}}{3!}=\frac{512\cdot 0{,}000335}{6}\approx 0{,}03.$$

2. Найдем вероятность того, что за 4 мин поступит менее трех вызовов, т. е. ни одного вызова, или один вызов, или два вызова. Поскольку эти события несовместимы, применима теорема сложения

$$P_4(k<3)=P_4(0)+P_4(1)+P_4(2)=e^{-8}+8e^{-8}+\frac{8^2e^{-8}}{2!}=$$

$$=41\cdot 0{,}000335\approx 0{,}01.$$

3. Найдем вероятность того, что за 4 мин поступит не менее трех вызовов, так как события «поступило менее трех вызовов» и «поступило не менее трех вызовов» — противоположные, то сумма вероятностей этих событий равна единице:

$$P_4(k<3)+P_4(k\geqslant 3)=1.$$

Отсюда

$$P_4(k\geqslant 3)=1-P_4(k<3)=1-[P_4(0)+P_4(1)+P_4(2)]=1-0{,}01=0{,}99.$$

Так как полученная вероятность весьма близка к единице, полученный результат можно истолковать так: почти достоверно, что за 4 мин поступит не менее трех вызовов.

Числовые характеристики случайных величин

Задача 1. Дискретная случайная величина X задана законом распределения

X	1	2
p	0,8	0,2.

Пользуясь неравенством Чебышева, оценить вероятность того, что абсолютная величина отклонения $X-M(X)$ будет меньше, чем 1.

Решение. Найдем математическое ожидание X:

$$M(X) = 1 \cdot 0{,}8 + 2 \cdot 0{,}2 = 1{,}2.$$

Напишем закон распределения X^2:

X^2	1	4
p	0,8	0,2.

Найдем математическое ожидание X^2:

$$M(X^2) = 1 \cdot 0{,}8 + 4 \cdot 0{,}2 = 1{,}6.$$

Найдем дисперсию X:

$$D(X) = M(X^2) - [M(X)]^2 = 1{,}6 - (1{,}2)^2 = 0{,}16.$$

Воспользуемся неравенством Чебышева:

$$P(|X - M(X)| < \varepsilon) \geqslant 1 - \frac{D(X)}{\varepsilon^2}.$$

Подставив сюда $M(X)=1{,}2$, $D(X)=0{,}16$, $\varepsilon=1$, окончательно получим

$$P(|X - 1{,}2| < 1) \geqslant 1 - \frac{0{,}16}{1} = 0{,}84.$$

Задача 2. Последовательность независимых случайных величин $X_1, X_2, \ldots, X_n, \ldots$ задана законом распределения

$$\begin{array}{cccc} X_n & -na & 0 & na \\ p & \frac{1}{2n^2} & 1-\frac{1}{n^2} & \frac{1}{2n^2}. \end{array}$$

Применима ли к заданной последовательности теорема Чебышева?

Решение. Для того чтобы к последовательности случайных величин была применима теорема Чебышева, достаточно, чтобы эти величины были попарно независимы, имели конечные математические ожидания и равномерно ограниченные дисперсии. (Требование равномерной ограниченности дисперсий означает, что существует такое постоянное число C, что дисперсии всех величин последовательности не превышает C.)

Поскольку случайные величины независимы, то они тем более и попарно независимы, т. е. первое требование теоремы Чебышева выполняется.

Проверим, выполняется ли требование конечности математических ожиданий:

$$M(X_n) = -na\frac{1}{2n^2} + 0\cdot\left(1 - \frac{1}{n^2}\right) + na\frac{1}{2n^2} = 0.$$

Мы видим, что каждая случайная величина имеет конечное (равное нулю) математическое ожидание, т. е. второе требование теоремы выполняется.

Проверим, выполняется ли требование равномерной ограниченности дисперсий. Напишем закон распределения X_n^2:

$$\begin{array}{cccc} X_n^2 & n^2a^2 & 0 & n^2a^2 \\ p & \frac{1}{2n^2} & 1-\frac{1}{n^2} & \frac{1}{2n^2}, \end{array}$$

или, сложив вероятности одинаковых возможных значений, получим

$$\begin{array}{ccc} X_n^2 & n^2a^2 & 0 \\ p & \frac{1}{n^2} & 1-\frac{1}{n^2}. \end{array}$$

Найдем математическое ожидание $M(X_n^2)$:

$$M(X_n^2) = n^2a^2\frac{1}{n^2} + 0\cdot\left(1 - \frac{1}{n^2}\right) = a^2.$$

Найдем дисперсию $D(X_n)$, учитывая, что $M(X_n)=0$:

$$D(X_n) = M(X_n^2) - [M(X_n)]^2 = a^2.$$

Таким образом, дисперсии заданных случайных величин равномерно ограничены числом α^2, т. е. третье требование теоремы выполняется.

Итак, поскольку все требования теоремы выполняются, к рассматриваемой последовательности случайных величин теорема Чебышева применима.

Законы распределения случайных величин

Задача 1. Случайная величина X задана интегральной функцией

$$F(x)=\begin{cases} 0 \text{ при } x \leqslant 0, \\ \dfrac{x}{2} \text{ при } 0 < x \leqslant 2, \\ 1 \text{ при } x > 2. \end{cases}$$

Найти вероятность того, что в результате испытания X примет значение, принадлежащее интервалу $(0, 1)$.

Решение. Искомая вероятность равна приращению интегральной функции на заданном интервале:

$$P(0 < X < 1) = F(1) - F(0).$$

Так как на интервале $(0, 1)$ по условию $F(x)=x/2$, то

$$F(1) - F(0) = \frac{1}{2} - \frac{0}{2} = \frac{1}{2}.$$

Итак,

$$P(0 < X < 1) = \frac{1}{2}.$$

Задача 2. Случайная величина X задана интегральной функцией, указанной в задаче 1. Требуется: 1) найти дифференциальную функцию, 2) пользуясь дифференциальной функцией, найти вероятность того, что в результате испытания X примет значение, принадлежащее интервалу $(0, 1)$.

Решение. 1. Найдем дифференциальную функцию $f(x)$, для чего продифференцируем по x интегральную функцию $F(x)$:

$$f(x) = F'(x) = \begin{cases} 0 \text{ при } x < 0, \\ \dfrac{1}{2} \text{ при } 0 < x < 2, \\ 0 \text{ при } x < 2. \end{cases}$$

2. Искомая вероятность равна определенному интегралу в пределах от 0 до 1 от дифференциальной функции:

$$P(0 < X < 1) = \int_0^1 \frac{1}{2}\, dx = \frac{1}{2} x \Big|_0^1 = \frac{1}{2}(1 - 0) = \frac{1}{2}.$$

Студенту рекомендуется построить графики интегральной и дифференциальной функции.

Задача 1. Случайная величина X распределена нормально. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 2 и 5. Найти вероятность того, что в результате испытания X примет значение, принадлежащее интервалу (1, 4).

Решение. Воспользуемся формулой (§ 5)

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

По условию $\alpha=1$, $\beta=4$, $a=2$, $\sigma=5$, следовательно,

$$P(1 < X < 4) = \Phi\left(\frac{4-2}{5}\right) - \Phi\left(\frac{1-2}{5}\right) = \Phi(0{,}4) - \Phi(-0{,}2).$$

Так как функция Лапласа нечетна, то $\Phi(-0{,}2) = -\Phi(0{,}2)$. Таким образом,

$$P(1 < X < 4) = \Phi(0{,}4) + \Phi(0{,}2),$$

По таблице (приложение 2) учебника [1] находим:

$$\Phi(0{,}4) = 0{,}1554; \quad \Phi(0{,}2) = 0{,}0793.$$

Искомая вероятность равна

$$P(1 < X < 4) = 0{,}1554 + 0\,0793 = 0{,}2347.$$

Задача 2. Случайная величина X распределена нормально. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 10 и 5. Найти вероятность того, что абсолютная величина отклонения $X-a$ будет меньше двух.

Решение. Воспользуемся формулой (§ 6)

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

По условию $a=10$, $\sigma=5$, $\delta=2$, следовательно,

$$P(|X - 10| < 2) = 2\Phi\left(\frac{2}{5}\right) = 2\Phi(0{,}4).$$

По таблице (приложение 2 учебника [1]) находим:

$$\Phi(0{,}4) = 0{,}1554.$$

Искомая вероятность равна

$$P(|X - 10| < 2) = 2 \cdot 0{,}1554 = 0{,}3108.$$

Задачи для самостоятельного решения

Для повторения и закрепления пройденного материала рекомендуется решить приведенные ниже задачи.

1. Производится три независимых испытания, в каждом из которых вероятность появления события A равна 0,4. Составить закон распределения дискретной случайной величины X — числа появлений события A в указанных испытаниях. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение X.

Ответ: $M(X)=1{,}2$; $D(X)=0{,}72$; $\sigma(X)=0{,}85$.

2. Два автомата производят детали, которые поступают на общий конвейер. Производительность первого автомата втрое больше производительности второго. Вероятность изготовления годной детали первым автоматом равна 0,9, а вторым — 0,7. С конвейера взяты наудачу 5 деталей. Найти вероятность того, что 4 из них годные.

Ответ. $P \approx 0{,}39$.

3. Для поражения цели достаточно попадания хотя бы одного снаряда. Произведено два залпа из двух орудий. Найти вероятность поражения цели, если вероятность попадания в цель при одном выстреле из первого орудия равна 0,3, а из второго — 0,4.

Ответ. 0,8236.

4. Из пункта O ведется стрельба из орудия вдоль прямой Ox. Предполагается, что дальность полета снаряда распределена нормально с математическим ожиданием 1000 м и средним квадратическим отклонением 50 м. Найти, сколько процентов снарядов: 1) дадут перелет от 40 до 60 м; 2) пролетят расстояние меньшее средней дальности; 3) пролетят расстояние большее средней дальности.

Ответ. 1) 9,68%; 2) 50%; 3) 50%.

5. Даны законы распределения двух случайных величин:

$$X \quad 1 \quad 3 \qquad Y \quad 2 \quad 4$$
$$p \quad 0{,}4 \quad 0{,}6; \qquad g \quad 0{,}2 \quad 0{,}8.$$

Найти математическое ожидание суммы $X+Y$ двумя способами: 1) составив закон распределения $X+Y$; 2) пользуясь свойством $M(X+Y)=M(X)+M(Y)$.

Ответ. 5,8.

6. Даны законы распределения двух независимых случайных величин:

$$X \quad 1 \quad 4 \qquad Y \quad 0{,}5 \quad 2$$
$$p \quad 0{,}6 \quad 0{,}4; \qquad g \quad 0{,}8 \quad 0{,}2.$$

Найти дисперсию суммы $X+Y$ двумя способами: 1) составив закон распределения $X+Y$; 2) пользуясь свойством $D(X+Y)=D(X)+D(Y)$.

Ответ. 2,52.

7. Вероятность появления события в каждом из 100 независимых испытаний равна 0,8. Найти вероятность того, что событие появится в этих испытаниях: 1) ровно 90 раз; 2) не менее 80 и не более 90 раз.

Ответ: 1) 0,004; 2) 0,4938.

8. Три баскетболиста должны произвести по одному броску мяча. Вероятности попадания мяча в корзину первым, вторым и третьим баскетболистами соответственно равны 0,9; 0,8; 0,7. Найти вероятность того, что удачно произведет бросок только один баскетболист.

Ответ. 0,092.

9. Оценить вероятность отклонения $|X-M(X)|<2$, если $D(X)=0{,}004$.

Ответ. $P \geqslant 0{,}9$.

10. События A и B совместны. Доказать, что

$$P(A+B)=P(A)+P(B)-P(A \cdot B).$$

11. Среднее число вызовов, поступающих на АТС в 1 мин равно двум. Найти вероятность того, что за 3 мин поступит: 1) 5 вызовов; 2) более пяти вызовов.

Ответ. 1) $P_3(5)=0{,}16$; 2) $P_3(k>5)=0{,}84$.

12. Непрерывная случайная величина X распределена по показательному закону, который задан дифференциальной функцией $f(x)=3e^{-3x}$ при $x \geqslant 0$, $f(x)=0$ при $x<0$.

Найти: 1) математическое ожидание, дисперсию и среднее квадратическое отклонение X; 2) вероятность того, что X примет значение, принадлежащее интервалу (1, 2).

Ответ. 1. $M(X)=1/3$; $D(X)=1/9$; $\sigma(X)=1/3$; 2) $P\ (1<X<2)=0{,}05$.

Глава II
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

§ 3. ПЕРВИЧНАЯ ОБРАБОТКА ВЫБОРОК

1. Генеральная совокупность и выборка. Совокупность объектов, или, точнее, совокупность значений какого-то признака объектов, называется *генеральной совокупностью.* Основной задачей математической статистики является исследование генеральной совокупности статистически, т. е. выяснение вероятностных свойств совокупности: *распределения, числовых характеристик* и т. д.

Однако полное исследование генеральной совокупности обычно практически невозможно или неэкономно. Например, при проверке лампочек накаливания одним из ее качественных свойств считается время работы (до сгорания). Аналогичная ситуация имеет место и при проверке качества консервов, снарядов и т. д. Кроме того, всеобщая проверка генеральной совокупности требует больших материальных затрат. Поэтому всеобщее исследование применяют, как правило, редко. Например, всеобщую перепись населения в Советском Союзе производят примерно через 10 лет.

Обычно из генеральной совокупности делают выборку, т. е. исследуют только некоторые ее объекты. С помощью выборки оценивают генеральную совокупность по вероятностным свойствам. Чтобы оценки были достоверными, выборка должна быть *представительной*, т. е. ее вероятностные свойства должны совпадать или быть близкими к свойствам генеральной совокупности.

Представительную выборку можно получить, если выбирать объекты для исследований случайно, т. е. гарантировать всем объектам генеральной совокупности одинаковую вероятность подвергнуться исследованию. Далее предполагаем, что все выборки получены из генеральной совокупности случайно.

Случайно выбранный объект после проверки нужного признака можно возвратить (возвратная или повторная выборка) или не возвратить (*безвозвратная* или *бесповторная выборка*) обратно в генеральную совокупность. В первом случае получаем более независимую и представительную выборку.

Часто под генеральной совокупностью понимают и исследуемую случайную величину. Для исследования случайной величины при постоянных условиях выполняются испытания. Совокупность полученных значений также называется выборкой и обрабатывается статистически. Методы статистической обработки выборки аналогичны в обоих случаях.

При исследовании объектов можно фиксировать или измерять значение одного или нескольких признаков. Соответственно говорят об *одномерной, двумерной, трехмерной* и т. д. выборках. Вначале рассмотрим обработку одномерных выборок.

2. Вариационный ряд. Выбор объекта из генеральной совокупности и измерение значения признака называется статистическим наблюдением. Результаты наблюдений фиксируют в протоколе или дневнике наблюдений в порядке их появления.

Выборка будет намного нагляднее, если все ее элементы упорядочить по возрастанию или по убыванию. Но в выборке одно значение (вариант) может встречаться несколько раз, и поэтому целесообразно результаты записать в виде таблицы, в первом столбце которой находятся всевозможные значения (варианты) x_i генеральной совокупности (или случайной величины) X, а во втором — числа n_i, т. е. частоты появления i-го значения. Такую таблицу называют *вариационной таблицей* или *вариационным рядом*.

Для составления вариационного ряда нужно:

1) найти минимальное ($x_{\min}$) и максимальное ($x_{\max}$) значения выборки;

2) в первый столбец таблицы записать варианты значений случайной величины (генеральной совокупности), начиная с $x_{\min}$ и кончая $x_{\max}$;

3) просмотреть по одному все элементы выборки в протоколе наблюдений, и отметить каждое значение в соответствующем варианте во втором столбце таблицы;

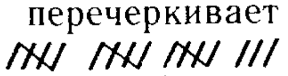
4) подсчитать количество меток в каждом варианте и записать соответствующее им число n_i;

5) подсчитать количество элементов в выборке (объем выборки) n, которое должно быть равно

$$n = \sum_{i=1}^{m} n_i, \qquad (\text{II}.1)$$

где m — количество вариантов в вариационном ряде. Если условие (II.1) не выполнено, то повторить все пункты, начиная с третьего.

Если объем выборки n большой, то строка меток может оказаться слишком длинной, т. е. подсчитать их неудобно. Более короткой получается запись при следующих способах подсчета: каждая пятая палочка перечеркивает предыдущие четыре:



из точек и палочек образуют фигуру из 10 элементов (см. табл. 7 и 9 на с. 47 и 49).

В обоих случаях упрощается подсчет количества пометок.

Если количество вариантов m слишком велико или близко к объему выборки, то целесообразно составить *вариационный ряд по интервалам значений генеральной совокупности.* По интервалам составляют вариационный ряд и из выборки непрерывной генеральной совокупности.

Вариационный ряд по интервалам значений можно получить с помощью приведенного выше алгоритма, где во втором пункте следует:

заполнить первый столбец таблицы интервалами значений генеральной совокупности. Все интервалы выбирать одинаковой длины таким образом, чтобы x_{min} вошло в первый, а x_{max} — в последний интервал. Обычно начало интервала входит в интервал, а его конец — не входит.

В остальных пунктах алгоритма следует слово «вариант» заменить словом «интервал». Пример вариационного ряда по интервалам см. в табл. 9 на с. 49.

3. Графики вариационных рядов. Своеобразными графиками являются строки меток, сделанных при составлении вариационных рядов. Сравнивая вариационные ряды в табл. 7 и 9, видим, что в обоих рядах встречаются одинаковые частоты n_i (16, 3). Но соответствующие значения имеют разный вес в выборках, так как объемы выборок различны (79 и 200). Значения генеральной совокупности будут сравнимыми, если использовать *относительные частоты или частости* n_i/n. При построении графиков обычно используют частости. Сумма частостей должна быть равна единице [см. формулу (II.1)]:

$$\sum_{i=1}^{m} \frac{n_i}{n} = 1. \tag{II.2}$$

Используют два вида графиков вариационных рядов: *полигон* и *гистограмму*. Если вариационный ряд составлен по значениям, то полигон строят из отрезков, соединяющих точки, координатами которых являются значения x_i и соответствующие частости n_i/n (см. рис. 17 на с. 48). При построении гистограммы над каждым значением x_i строят прямоугольник, высота которого пропорциональна соответствующей частости n_i/n (см. рис. 18 на с. 48).

Если вариационный ряд составлен по интервалам, то в качестве значений x_i следует рассматривать середины интервалов (см. рис. 20, 21 на с. 50).

4. Эмпирическая функция распределения. Каждая генеральная совокупность имеет функцию распределения $F(x)$ [см. формулу (I.19)], которая обычно неизвестна. По выборке можно найти эмпирическую функцию распределения $F^*(x)$, где на основании закона больших чисел Бернулли вместо вероятностей p_i берутся относительные частоты n_i/n. Процесс нахождения эмпрической функции распределения $F^*(x)$ аналогичен процессу нахождения функции распределения $F(x)$ дискретной случайной величины X [см. п. 3 § 2, формулы (I.20) и (I.21)]:

$$F^*(x) = \sum_{x_i<x} \frac{n_i}{n}, \tag{II.3}$$

$$F^*(x) = \begin{cases} 0, & x \leqslant x_1, \\ \dfrac{n_1}{n}, & x_1 < x \leqslant x_2, \\ \dfrac{n_1+n_2}{n}, & x_2 < x \leqslant x_3, \\ \dots\dots\dots\dots\dots \\ \sum\limits_{i=1}^{m-1} \dfrac{n_i}{n}, & x_{m-1} < x \leqslant x_m, \\ 1, & x > x_m. \end{cases} \qquad (II.4)$$

Значениями эмпирической функции распределения $F^*(x)$ [формула (II.4)] являются так называемые *накопленные частости* (см. табл. 7 и 9 на с. 47 и 49). График эмпирической функции распределения строят так же, как и график функции распределения $F(x)$ дискретной случайной величины (см. рис. 3).

Если вариационный ряд составлен по интервалам значений и в качестве представителя интервала берется его середина, то эмпирическая функция составляется так же, как по вариационному ряду по значениям. Но в качестве представителя интервала можно брать и правый конец интервала. Объединяя отрезками точки, координатами которых являются правые концы интервалов и накопленные частости соответствующих интервалов, получаем ломаную линию, являющуюся довольно хорошим приближением *графика функции распределения непрерывной случайной величины* (ср. рис. 5 и 22 на с. 25 и 50). Такой график является точным, если все значения в каждом интервале распределены равномерно. Аналитический вид этой функции довольно сложен.

5. Числовые характеристики выборки. 5.1. Среднее арифметическое. *Среднее арифметическое* $\bar{x}$ определяется по формуле

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad (II.5)$$

где x_i — элементы выборки, n — ее объем. Если объем выборки n небольшой и x_i не слишком велики, то расчет «вручную» по этой формуле не вызывает трудности. Для больших выборок необходимо прибегнуть к помощи микрокалькулятора или ЭВМ. С помощью формулы (II.5) вычисляют непосредственно по протоколу наблюдений.

Если составлен вариационный ряд, то следует использовать следующую формулу:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i, \qquad (II.6)$$

где x_i — варианты случайной величины, n_i — соответствующие частоты, m — количество вариантов, n — объем выборки.

Если при вычислении по этой формуле встречаются трудности то можно обратиться к микрокалькулятору или ЭВМ.

Для упрощения счета имеется следующая формула:

$$\bar{x} = \frac{\sum_{i=1}^{m} \frac{x_i - c}{k} n_i}{n} k + c, \tag{II.7}$$

где x_i, n_i, m и n имеют тот же смысл, что и в предыдущей формуле; k — шаг таблицы, т. е. интервал между соседними вариантами; c — произвольное число (но для простоты следует выбрать вариант, имеющий максимальную частоту). В результате при вычислениях приходится иметь дело с довольно малыми числами. Формулу (II.7) используют только в том случае, когда вариационный ряд имеет постоянный шаг таблицы k. При переменном шаге нужно использовать формулу (II.6).

Если вариационный ряд составлен по интервалам значений, то в роли x_i в формулах (II.6) и (II.7) используют середины интервалов.

Вычисления по формулам (II.5) и (II.6) можно упростить, если выполнить замену переменных $y_i = x_i - c$, где константа выбирается вблизи середины интервала, в котором находятся все значения выборки. Таким образом, новые значения вариантов y_i получаются как отклонения старых вариантов от «ложного нуля» c, т. е. они довольно малы по абсолютной величине. Формулы (II.5) и (II.6) принимают соответственно вид:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} (x_i - c) + c; \tag{II.5а}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} (x_i - c) n_i + c. \tag{II.6а}$$

5.2. Дисперсия выборки. Дисперсию выборки обозначим через $\bar{S}^2$. Для вычисления выборочной дисперсии $\bar{S}^2$ приведем такие же формулы, что и для нахождения среднего арифметического:

$$\bar{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2; \tag{II.8}$$

$$\bar{S}^2 = \frac{1}{n} \sum_{i=1}^{m} (x_i - \bar{x})^2 n_i = \frac{1}{n} \sum_{i=1}^{m} x_i^2 n_i - \bar{x}^2; \tag{II.9}$$

$$\bar{S}^2 = \frac{\sum_{i=1}^{m} \left(\frac{x_i - c}{k}\right)^2 n_i}{n} k^2 - (\bar{x} - c)^2. \tag{II.10}$$

Для упрощения расчетов формулы (II.8) и (II.9) можно преобразовать следующим образом:

$$\bar{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - c)^2 - (\bar{x} - c)^2; \tag{II.8а}$$

$$\bar{S}^2 = \frac{1}{n}\sum_{i=1}^{m}(x_i - c)^2 n_i - (\bar{x} - c)^2. \qquad \text{(II.9a)}$$

5.3. Стандартное отклонение. *Стандартное*, или *среднеквадратичное*, отклонение определяется как квадратный корень из дисперсии:

$$\bar{S} = \sqrt{\bar{S}^2}.$$

5.4. Мода. Если вариационный ряд составлен по значениям генеральной совокупности, то *модой выборки* является значение, имеющее максимальную частоту. Если вариационный ряд составлен по интервалам значений генеральной совокупности, то мода вычисляется по следующей приближенной формуле:

$$Mo = x_0 + k\frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}, \qquad \text{(II.11)}$$

где x_0 — начало модального интервала, т. е. интервала, имеющего максимальную частоту, k — длина модального интервала, n_i — частота модального интервала, n_{i-1} и n_{i+1} — частоты соответственно предшествующего и последующего за модальным интервалов.

5.5. Медиана. *Медианой* выборки является значение серединного элемента вариационного ряда. Если вариационный ряд составлен по значениям генеральной совокупности, то при нечетном объеме выборки n медиана — это действительное значение серединного элемента, а при n четном — среднее арифметическое двух серединных элементов.

Если вариационный ряд составлен по интервалам значений, то медиана вычисляется по следующей приближенной формуле:

$$Me = x_0 + k\frac{n/2 - T_{i-1}}{n_i}, \qquad \text{(II.12)}$$

где x_0 — начало медианного интервала, т. е. интервала, в котором содержится серединный элемента; k — длина медианного интервала; n — объем выборки, T_{i-1} — сумма частот интервалов, предшествующих медианному; n_i — частота медианного интервала.

§ 4. ТЕОРИЯ ОЦЕНОК

1. Понятие оценки. Генеральные совокупности характеризуются некоторыми постоянными числовыми характеристиками распределения. По выборкам можно найти оценки этих характеристик. Вследствие случайности выборок значения оценок одной числовой характеристики, вычисленные по разным выборкам из одной же генеральной совокупности, бывают, как правило, различными.

Обозначим неизвестный параметр распределения, т. е. числовую характеристику генеральной совокупности X, через θ, а оценку неизвестного параметра — через T_n. Оценка T_n — функция от выборки. Оценки неизвестного параметра можно находить различными способами. Например, если нужно оценить среднее значение $\theta = \mu$ нормального распределения, то можно использовать следующие оценки T_n:

1) x_1 — первый элемент выборки. На практике часто так и поступают: измеряют какую-то величину только один раз, и этот результат используют как значение этой величины;

2) $(x_{\max} + x_{\min})/2$ — среднее арифметическое максимального и минимального элементов выборки;

3) Mo — моду, которая при нормальном распределении равна среднему значению μ;

4) Me — медиану, которая при нормальном распределении также равна среднему значению μ;

5) $\bar{x}$ — среднее арифметическое.

Для того чтобы установить, какая из оценок лучше, надо знать основные свойства (виды) оценок.

2. Несмещенные оценки. *Несмещенной* называется оценка T_n, среднее значение которой равно оцениваемому параметру θ:

$$ET_n = \theta.$$

Если это условие не выполняется, то оценку называют *смещенной*, при этом смещение вычисляется как разность $ET_n - \theta$.

Другие свойства оценок в настоящем пособии не рассматриваются. *Несмещенной оценкой среднего значения* μ является среднее арифметическое $\bar{x}$.

Аналогично с помощью выборочной дисперсии $\bar{S}^2$ можно оценить дисперсию σ^2. Оказывается, что выборочная дисперсия $\bar{S}^2$ является *смещенной оценкой дисперсии* σ^2:

$$E(\bar{S}^2) = \frac{n-1}{n}\sigma^2,$$

т. е. $\sigma^2 - \frac{n-1}{n}\sigma^2 = \frac{1}{n}\sigma^2$. Отсюда видно, что при $n \to \infty$ смещение стремится к нулю. Значит, при достаточно большом объеме выборки n выборочную дисперсию можно приближенно принимать за несмещенную оценку дисперсии σ^2. Для оценки дисперсии, несмещенной при малом объеме выборки, используют *исправленную дисперсию*

$$S^2 = \frac{n}{n-1}\bar{S}^2. \qquad \text{(II.13)}$$

Если сравнить эту формулу с формулами для выборочной дисперсии из п. 5.2. § 3 [см. формулы (II.8) — (II.10)], то можно получить аналогичные формулы для вычисления несмещенной оценки S^2 дисперсии σ^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2; \qquad \text{(II.14)}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i. \qquad \text{(II.15)}$$

Преобразовать формулу (II.10) так просто не удается, поэтому нужно сначала по формуле (II.10) вычислить выборочную дисперсию $\bar{S}^2$, а затем с помощью формулы (II.13) найти несмещенную оценку S^2 дисперсии σ^2.

3. Доверительный интервал. Оценки T_n неизвестного параметра θ, рассмотренные выше, называют *точечными*, так как они определяют одно значение, одну точку на числовой оси. Все точечные оценки параметров распределения генеральной совокупности вычисляют по выборкам, но из-за случайности выборок *оценки являются случайными величинами*, отличающимися от постоянного истинного значения параметра θ. Обозначим точность оценки через Δ ($\Delta > 0$); тогда $|\theta - T_n| \leqslant \Delta$. Чем меньше Δ, тем точнее оценка.

Любую точность можно получить с определенной вероятностью (*надежностью*):

$$P(|\theta - T_n| \leqslant \Delta) = \gamma. \qquad \text{(II.16)}$$

Если преобразовать это выражение, то получим

$$P(-\Delta \leqslant \theta - T_n \leqslant \Delta) = \gamma,$$

или

$$P(T_n - \Delta \leqslant \theta \leqslant T_n + \Delta) = \gamma. \qquad \text{(II.17)}$$

Условие (II.17) означает, что интервал $[T_n - \Delta,\ T_n + \Delta]$ покрывает значение параметра θ с заданной *доверительной вероятностью* γ. Точность оценки Δ фактически определяет *длину доверительного интервала* (2Δ). Доверительная вероятность γ задается обычно значением, близким к единице, например, 0,95; 0,98; 0,99 и т. д.

Доверительная вероятность γ, точность оценки Δ и объем выборки n связаны между собой. Если определены две величины, то тем самым будет определена и третья.

4. Доверительный интервал для среднего значения μ нормального распределения при известном σ. Пусть задана генеральная совокупность с нормальным распределением $X \in N(\mu, \sigma)$, где значение стандартного отклонения σ известно. Для оценки параметра μ воспользуемся величиной $\bar{X}$. Заметим, что и *среднее арифметическое* $\bar{X}$, и *элементы выборки* $X_1, X_2, \ldots, X_n$ из-за случайности выборок являются *случайными величинами*. Все элементы выборки имеют то же распределение, что и генеральная совокупность: $X_i \in N(\mu, \sigma)$, $i = 1, 2, \ldots, n$. Среднее арифметическое также имеет нормальное распределение: $\bar{X} \in N(\mu, \frac{\sigma}{\sqrt{n}})$.

По формуле (II.16) получим

$$P(|\bar{X} - \mu| \leqslant \Delta) = \gamma. \tag{II.18}$$

С другой стороны, заменяя в (I.45) X на $\bar{X}$, σ на $\sigma/\sqrt{n}$ и ε на Δ, получим

$$P(|\bar{X} - \mu| \leqslant \Delta) = 2\,\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 = 2\,\Phi(u_\gamma) - 1 = \gamma, \tag{II.19}$$

где $u_\gamma = \frac{\Delta}{\sigma}\sqrt{n}$. Отсюда находим

$$\Delta = u_\gamma \frac{\sigma}{\sqrt{n}}. \tag{II.20}$$

Используя соотношения (II.17) и (II.18), можно записать формулу для вычисления доверительного интервала:

$$P\left(\bar{x} - u_\gamma \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + u_\gamma \frac{\sigma}{\sqrt{n}}\right) = \gamma, \tag{II.21}$$

где выражение в скобках определяет доверительный интервал. Вычислим значение переменной u_γ. На основании формулы (II.19) получим условие

$$\Phi(u_\gamma) = 1/2(1 + \gamma). \tag{II.22}$$

Согласно этому условию, из таблиц (см. приложение 3) найдем значение аргумента u_γ.

5. Доверительный интервал для среднего значения μ нормального распределения при неизвестном σ. Пусть задана генеральная совокупность с нормальным распределением $X \in N(\mu, \sigma)$, где значение стандартного отклонения σ неизвестно. Так как σ неизвестно, то непосредственно воспользоваться нормальным распределением $N(\mu, \sigma)$ нельзя. Однако известно, что случайная величина

$$t = \frac{\bar{x} - \mu}{S}\sqrt{n}, \tag{II.23}$$

где S — несмещенная оценка стандартного отклонения генеральной совокупности, n — объем выборки, имеет *распределение Стьюдента* (*t-распределение*) *с числом степеней свободы* $n - 1$.

Для получения интервальной оценки потребуем, чтобы выполнялось условие

$$P\left(\left|\frac{\bar{x}-\mu}{S}\sqrt{n}\right| \leqslant t_\gamma\right) = \gamma. \qquad (II.24)$$

Величина t_γ определяется по таблицам распределения Стьюдента (см. приложение 5). По нижней части головки таблицы; на основании условия $P(|X| > t_\gamma) = \gamma$ определяется t_γ, но в данном случае имеем противоположное неравенство, значит, нужно использовать условие

$$P\left(\left|\frac{\bar{x}-\mu}{S}\sqrt{n}\right| > t_\gamma\right) = 1-\gamma.$$

Число степеней свободы равно $n-1$.

Преобразовав условие (II.24), имеем

$$P\left(\bar{x}-t_\gamma\frac{S}{\sqrt{n}} \leqslant \mu \leqslant \bar{x}+t_\gamma\frac{S}{\sqrt{n}}\right) = \gamma, \qquad (II.25)$$

где доверительный интервал указан в скобках. Полученная формула аналогична формуле (II.20). Здесь

$$\Delta = t_\gamma\frac{S}{\sqrt{n}}. \qquad (II.26)$$

6. Доверительный интервал для дисперсии σ^2 нормального распределения. Предположим, что генеральная совокупность имеет нормальное распределение $X \in N(\mu, \sigma)$. Тогда случайная величина

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \qquad (II.27)$$

имеет *χ^2-распределение (распределение Пирсона) с числом степеней свободы $n-1$.* Случайная величина с χ^2-распределением принимает только неотрицательные значения. По таблицам χ^2-распределения (см. приложение 4) можно найти x_α, удовлетворяющее следующему условию: $P(\chi^2 > x_\alpha) = \alpha$ (рис. 23).

По таблицам χ^2-распределения всегда можно найти такие два числа u_1 и u_2, которые удовлетворяли бы условию

$$P(u_1 \leqslant \chi^2 \leqslant u_2) = \gamma. \qquad (II.28)$$

Таких пар чисел u_1 и u_2 существует бесконечное множество. Чтобы зафиксировать одну такую пару u_1, u_2, введем дополнительное условие (симметричность по вероятности) (рис. 24):

$$P(\chi^2 < u_1) = P(\chi^2 > u_2) = \frac{1}{2}(1-\gamma). \qquad (II.29)$$

Из таблиц, используя условие (II.29), получаем u_2. Для нахождения u_1 используем вероятность противоположного события

$$P(\chi^2 > u_1) = \frac{1}{2}(1+\gamma). \qquad (II.30)$$

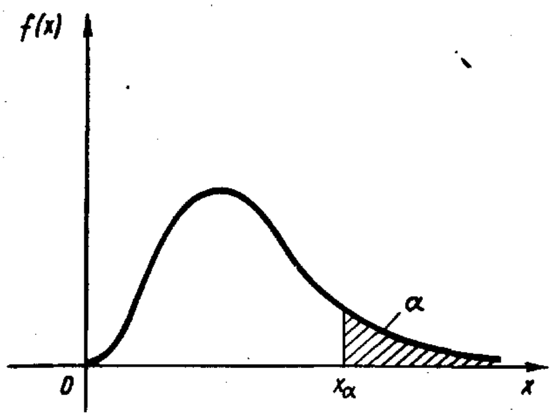


Рис. 23. Использование таблицы χ^2-распределения

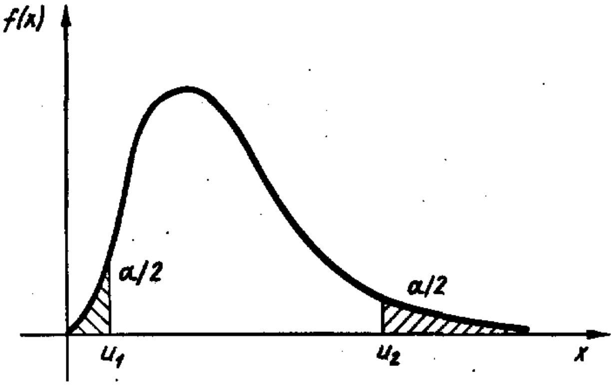


Рис. 24. Нахождение чисел u_1 и u_2

Заменяя в формуле (II.28) χ^2 его значением из формулы (II.27) и выполняя преобразования, получаем

$$P\left(\frac{(n-1)S^2}{u_2} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{u_1}\right) = \gamma, \tag{II.31}$$

где в скобках задан *доверительный интервал для дисперсии* σ^2.

Извлекая квадратный корень из обеих сторон неравенства, определяющего доверительный интервал для дисперсии σ^2, получаем *доверительный интервал для среднего квадратичного (стандартного) отклонения* σ:

$$\frac{S\sqrt{n-1}}{\sqrt{u_2}} \leqslant \sigma \leqslant \frac{S\sqrt{n-1}}{\sqrt{u_1}}. \tag{II.32}$$

7. Определение объема выборки *n*. До сих пор мы рассматривали обработку готовых выборок с фиксированным объемом n. Какой объем должна иметь выборка, чтобы было можно получить результаты нужной точности? По закону больших чисел предпочтение отдается выборкам с большим объемом. Но обычно больший объем выборки требует и больших затрат для ее получения и обработки. Поэтому на практике целесообразно использовать тот минимальный объем, который позволяет получить удовлетворительные результаты.

При вычислении доверительных интервалов среднего значения нормального распределения можно, используя формулу (II.20), определить *объем выборки*

$$n = u_\gamma^2 \frac{\sigma^2}{\Delta^2}. \tag{II.33}$$

Таким образом, объем выборки n прямо пропорционален σ^2 и u_γ^2 (при этом последнее значение зависит от γ) и обратно пропорционален σ^2. По этим данным можно задать значения вероятности и длину полуинтервала Δ. Если мы хотим получить интервал с большей доверительной вероятностью γ (вместе с тем увеличивается и u_γ), то следует увеличить объем выборки n. Если мы хотим укоротить интервал, то должны увеличить объем выборки n. Так, с помощью формулы (II.33) вычисляется объем выборки при известном σ.

Если σ неизвестно, то для выборки объема n из формулы (II.26) получаем

$$n = t_\gamma^2 \frac{S^2}{\Delta^2}. \qquad \text{(II.34)}$$

По формуле (II.33), задавая σ, можно определить соответствующий объем выборки n до получения самой выборки. По формуле (II.34) можно определить нужный объем выборки n после обработки уже имеющейся пробной выборки, по которой вычисляется несмещенная оценка дисперсии генеральной совокупности S^2.

§ 5. СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

1. Понятие статистической гипотезы. *Статистической* гипотезой называют любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин. Например, случайная величина X имеет распределение Пуассона, случайная величина с нормальным распределением имеет среднее значение $\mu = 5$ или $\mu \neq 5$ и т. д. Статистические гипотезы проверяются статистическими методами.

Гипотезы о неизвестном параметре θ распределения бывают простые и сложные; простая гипотеза утверждает, что параметр θ имеет одно конкретное значение $(\theta = \theta_0)$, а сложная гипотеза утверждает, что параметр θ имеет значение из совокупности значений $(\theta < \theta_0,\ \theta > \theta_0,\ \theta \neq \theta_0)$.

Проверяемую гипотезу обозначим H_0. Обычно вырабатывают еще и *альтернативную гипотезу* H_1, отрицающую или исключающую основную гипотезу H_1. Таким образом, в результате проверки можно принимать только одну из гипотез H_0 или H_1, отвергая в это же время другую.

Гипотезу проверяют на основании выборки, полученной из генеральной совокупности. Из-за случайности выборки в результате проверки могут возникать ошибки и приниматься неправильные решения. В принципе возможны ошибки первого и второго рода. *Ошибка первого рода* имеет место тогда, когда отвергается правильная гипотеза H_0. При *ошибке второго рода* принимается неправильная гипотеза H_0.

Таким образом, по одним выборкам принимается правильное решение, а по другим — неправильное. Решение принимается по значению некоторой функции выборки, называемой *статистикой* или *статистической характеристикой*. Множество значений этой статистики можно разделить на два непересекающихся подмножества:

подмножество значений статистики, при которых гипотеза H_0 принимается (не отклоняется), называют *областью принятия гипотезы (допустимой областью)*;

подмножество значений статистики, при которых гипотеза H_0 отвергается (отклоняется) и принимается гипотеза H_1, называют *критической областью*.

При проверке гипотез разумно уменьшить вероятности принятия неправильных решений. *Допустимая вероятность ошибки первого рода* обозначается через α и называется *уровнем значимости*. Значение α обычно мало. Но уменьшение вероятности ошибки первого рода обычно вызывает увеличение вероятности ошибки второго рода (β).

Статистика выбирается так, чтобы вероятности α и β были бы минимальными. Проверяемая гипотеза H_0 в настоящем пособии предполагается всегда простой, так что распределение статистики при правильной гипотезе H_0 известно. Методы выбора наилучшей статистики здесь не рассматриваются.

Для определения критической области статистики используют уровень значимости α и учитывают вид альтернативной гипотезы H_1. *Основная гипотеза* H_0 о значении неизвестного параметра θ распределения выглядит так:

$$H_0 : \theta = \theta_0.$$

Альтернативная гипотеза H_1 может при этом иметь следующий вид:

$$H_1 : \theta < \theta_0, \quad H_1 : \theta > \theta_0 \quad \text{или} \quad H_1 : \theta \neq \theta_0.$$

Соответственно можно получить *левостороннюю, правостороннюю* или *двустороннюю критические области* (рис. 25). Граничные точки критических областей определяют по таблицам распределения статистики.

Проверка статистической гипотезы состоит из следующих этапов:

1) определение гипотез H_0 и H_1;

2) выбор статистики и задание уровня значимости α;

3) определение по таблицам, по уровню значимости α и по альтернативной гипотезе H_1 критической области;

4) вычисление по выборке значения статистики;

5) сравнение значения статистики с критической областью;

6) принятие решения: если *значение статистики не входит в критическую область*, то принимается гипотеза H_0 и отвергается гипотеза H_1, а *если входит в критическую область*, то отвергается гипотеза H_0 и принимается гипотеза H_1.

Иногда целесообразно перед определением альтернативной гипотезы H_1 выполнить этап 4), где для получения значения статистики

Рис. 25. Критические области:
а — левосторонняя; *б* — правосторонняя; *в* — двусторонняя

нужно вычислить несмещенные оценки параметров генеральной совокупности. Например, если проверяется гипотеза $H_0: \mu \neq 5$ и несмещенная оценка среднего значения $\bar{x} = 7{,}2$, то имеют смысл только следующие альтернативные гипотезы $H_1: \mu > 5$ или $H_1: \mu = 5$.

Результаты проверки статистической гипотезы нужно интерпретировать так: *если приняли гипотезу* H_1, то можно считать ее *доказанной*, а *если приняли гипотезу* H_0, то признали, что *гипотеза* H_0 *не противоречит результатам наблюдений*. Однако этим свойством наряду с H_0 могут обладать и другие гипотезы. Например, если мы принимаем гипотезу $H_0: \mu = 5$, то может случиться, что по данной выборке можно принять и другие гипотезы, например, $H_0: \mu = 5{,}5$ или $H_0: \mu = 4$ и т. д. Вопрос о том, как найти среди них наилучшую гипотезу, в данном пособии не рассматривается. Следует помнить, что, *принимая гипотезу* H_0, *следует проводить еще дополнительные исследования.*

2. Гипотеза о среднем значении нормального распределения при известном σ. Предполагаем, что генеральная совокупность имеет нормальное распределение $X \in N(\mu, \sigma)$, где значение σ известно. *При уровне значимости* α нужно проверить гипотезу $H_0: \mu = \mu_0$. В качестве альтернативной можно использовать одну из следующих гипотез $H_1: \mu < \mu_0$, $H_1: \mu > \mu_0$ или $H_1: \mu \neq \mu_0$. В качестве статистики воспользуемся случайной величиной

$$Z = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}, \qquad \text{(II.35)}$$

которая при истинной гипотезе H_0 имеет *нормированное нормальное распределение* $Z \in N(0, 1)$.

Критическую область определяем с помощью таблицы функции распределения (см. приложение 3).

Если альтернативная гипотеза имеет вид $H_1: \mu < \mu_0$, то используем левостороннюю критическую область, которая удовлетворяет (рис. 26) следующему условию:

$$P(Z < -z_\alpha) = \Phi(-z_\alpha) = \alpha. \qquad \text{(II.36)}$$

Таблицы составлены только для положительных значений аргумента, поэтому из таблицы найдем z_α, учитывая [см. формулу (I.17)], что

$$\Phi(z_\alpha) = 1 - \alpha. \qquad \text{(II.37)}$$

Отсюда следует, что критическая область — это множество таких Z, для которых

$$Z < -z_\alpha. \qquad \text{(II.38)}$$

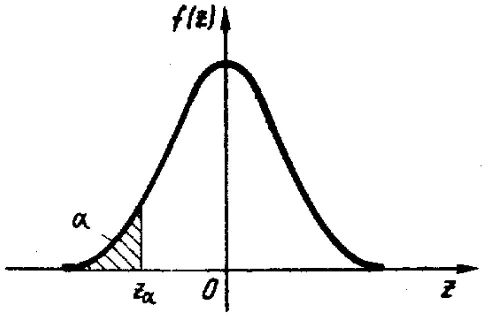


Рис. 26. Левосторонняя критическая область

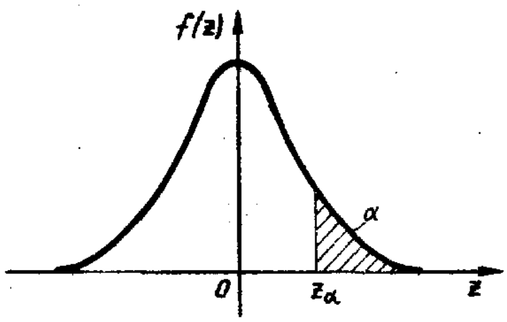


Рис. 27. Правосторонняя критическая область

Если альтернативная гипотеза имеет вид $H_1: \mu > \mu_0$, то используем правостороннюю критическую область, которая удовлетворяет (рис. 27) условию

$$P(Z > z_\alpha) = \alpha. \qquad \text{(II.39)}$$

Из таблицы получаем значение z_α, учитывая, что

$$P(Z < z_\alpha) = \Phi(z_\alpha) = 1 - \alpha. \qquad \text{(II.40)}$$

Отсюда находим критическую область

$$Z > z_\alpha. \qquad \text{(II.41)}$$

И наконец, при альтернативной гипотезе H_1: $\mu \neq \mu_0$ используем двустороннюю критическую область, удовлетворяющую (рис. 28) условию

$$P(|Z| > z_\alpha) = \alpha. \qquad \text{(II.42)}$$

Учитывая определение абсолютной величины, находим

$$P(Z < z_\alpha) = P(Z > z_\alpha) = \alpha/2.$$

По формулам (II.39) и (II.40) получаем условие использования таблицы:

$$\Phi(z_\alpha) = 1 - \alpha/2. \qquad \text{(II.43)}$$

Таким образом, критическая область имеет вид

$$|Z| > z_\alpha. \qquad \text{(II.44)}$$

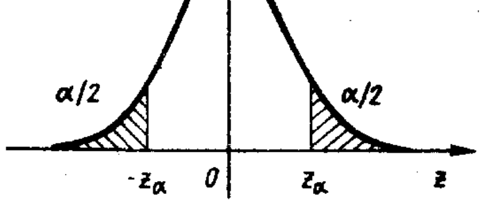


Рис. 28. Двусторонняя критическая область

Для вычисления значения статистики с помощью формулы (II.35) нужно по выборке найти среднее арифметическое $\bar{x}$.

3. Гипотеза о среднем значении нормального распределения при неизвестном σ. Предположения те же, что и в предыдущем пункте, но только σ неизвестно. В этом случае в качестве статистики используют случайную величину

$$t = \frac{\bar{X} - \mu_0}{S}\sqrt{n}, \qquad \text{(II.45)}$$

которая, если верна гипотеза H_0, имеет *t-распределение Стьюдента с числом степеней свободы* $n-1$, где n — объем выборки.

Критические области определяются так же, как и в предыдущем пункте. Но использование таблицы t-распределения Стьюдента проще, так как она составлена именно для определения критических областей. При нахождении левосторонней или правосторонней критических областей используем верхнюю головку таблицы, а для двусторонней — нижнюю.

Перед вычислением по формуле (II.45) значения статистики t нужно по выборке вычислить x и S.

4. Гипотеза о дисперсии нормального распределения. Предполагаем, что генеральная совокупность имеет нормальное распределение $X \in N(\mu, \sigma)$, где параметр σ неизвестен. Требуется *при уровне значимости α проверить гипотезу* H_0: $\sigma^2 = \sigma_0^2$. В качестве статистики используем случайную величину

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}. \qquad \text{(II.46)}$$

Если гипотеза H_0 верна, то случайная величина χ^2 имеет *χ^2-распределение Пирсона с числом степеней свободы* $n-1$, где n — объем выборки.

Критическая область определяется в зависимости от альтернативной гипотезы H_1 по таблице χ^2-распределения (см. приложение 4).

Если альтернативная гипотеза имеет вид $H_1: \sigma^2 < \sigma_0^2$, то используем *левостороннюю критическую область*, удовлетворяющую условию

$$P(\chi^2 < x_\alpha) = \alpha. \qquad \text{(II.47)}$$

Таблица χ^2-распределения составлена в соответствии с противоположным условием. Значит, для нахождения из таблицы x_α используем условие

$$P(\chi^2 > x_\alpha) = 1 - \alpha. \qquad \text{(II.48)}$$

При альтернативной гипотезе $H_1: \sigma^2 > \sigma_0^2$ находим *правостороннюю критическую область* исходя из условия

$$P(\chi^2 > x_\alpha) = \alpha, \qquad \text{(II.49)}$$

по которому x_α можно найти непосредственно из таблицы.

При альтернативной гипотезе $H_1: \sigma^2 \neq \sigma_0^2$ находим двустороннюю критическую область согласно условию

$$P((\chi^2 < x'_\alpha) \cup (\chi^2 > x''_\alpha)) = \alpha. \qquad \text{(II.50)}$$

Обычно принимают симметричную по вероятности критическую область, удовлетворяющую условию

$$P(\chi^2 < x'_\alpha) = P(\chi^2 > x''_\alpha) = \alpha/2. \qquad \text{(II.51)}$$

Для этого условия из таблицы можно сразу найти x''_α, а для получения x'_α следует преобразовать условия [см. формулы (II.47) и (II.48)]

$$P(\chi^2 > x_\alpha) = 1 - \alpha/2. \qquad (II.52)$$

По выборке нужно вычислить несмещенную оценку дисперсии генеральной совокупности S^2, а затем по формуле (II.46) можно найти значение статистики χ^2.

5. Гипотеза о равенстве двух средних значений. Предполагаем, что заданы две генеральные совокупности с нормальным распределением $X_1 \in N(\mu_1, \sigma_1)$, $X_2 \in N(\mu_2, \sigma_2)$, при этом стандартные отклонения σ_1 и σ_2, неизвестны, но должны быть равными. Сначала нужно проверить гипотезу о равенстве дисперсий. Из обеих генеральных совокупностей сделаны независимые выборки с параметрами n_1, $\bar{x}_1$, S_1 и n_2, $\bar{x}_2$, S_2 соответственно.

Обозначим разность средних значений через $\delta = \mu_1 - \mu_2$. *Зафиксировав уровень значимости* α, *проверим гипотезу* $H_0: \delta = \delta_0$, *используя статистику*

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \qquad (II.53)$$

Если гипотеза H_0 верна, то случайная величина t имеет *t-распределение с числом степеней свободы* $n_1 + n_2 - 2$. Обычно $\delta_0 = 0$, т. е. проверяется гипотеза о равенстве средних значений генеральных совокупностей.

Критическая область определяется в зависимости от вида альтернативной гипотезы H_1 ($\delta < \delta_0$, $\delta > \delta_0$ или $\delta \neq \delta_0$) по таблице t-распределения (см. приложение 5).

6. Гипотеза о равенстве двух дисперсий. Предполагаем, как и в предыдущем пункте, что заданы две генеральные совокупности X_1 и X_2 с нормальным распределением: $X_1 \in N(\mu_1, \sigma_1)$ и $X_2 \in N(\mu_2, \sigma_2)$. Из этих генеральных совокупностей сделаны независимые выборки с параметрами n_1, S_1^2 и n_2, S_2^2 соответственно. Требуется *при уровне значимости* α *проверить гипотезу* H_0: $\sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе H_1: $\sigma_1^2 > \sigma_2^2$. Обычно здесь другие альтернативные гипотезы не используют.

Предполагая, что $S_1^2 > S_2^2$, принимаем в качестве статистики величину

$$F = \frac{S_1^2}{S_2^2}. \qquad (II.54)$$

Если гипотеза H_0 верна, то случайная величина F имеет *F-распределение Фишера с числами степеней свободы* $n_1 - 1$ и $n_2 - 1$. Критическая область будет только *правосторонняя* и определяется условием

$$P(F > f_2) = \alpha. \qquad (II.55)$$

Значение f_α найдем из таблиц F-распределения (см. приложение 6). Значение f_α зависит от трех величин: уровня значимости α и двух чисел степеней свободы k_1 и k_2, поэтому таблицы составлены отдельно для каждого значения α (трехмерные таблицы). В таблицах число степеней свободы большей дисперсии k_1 находится в верхней части таблицы.

7. χ^2-критерий согласия. Рассмотрим, как можно проверить гипотезу о распределении генеральной совокупности X. Пусть генеральная совокупность имеет какое-то неизвестное распределение. Сделаем выборку из генеральной совокупности. На основании выборки или учитывая какие-то другие соображения составим гипотезу о конкретном распределении генеральной совокупности, выраженной через функцию распределения $F(x)$. Это распределение назовем *теоретическим.*

По выборке можем найти *эмпирическую функцию распределения* $F^*(x)$. Гипотезу H_0 о распределении генеральной совокупности принимаем тогда, когда эмпирическое распределение хорошо согласуется с теоретическим. Полного совпадения, конечно, ожидать не стоит. Для проверки таких гипотез разработаны несколько критериев согласия. Мы рассмотрим только χ^2-критерий согласия Пирсона.

При использовании χ^2-критерия согласия вся область изменения генеральной совокупности X делится на k интервалов, которые могут иметь различную длину. По выборке составляют вариационный ряд по этим же интервалам. Если в некотором интервале частота n_i слишком мала (меньше 5), то этот интервал объединяют с соседним интервалом. При дискретной генеральной совокупности интервал может содержать только одно значение генеральной совокупности.

По выборке вычисляют оценки параметров теоретического распределения. Тем самым теоретическое распределение будет полностью определено. Теперь по теоретическому распределению вычислим вероятности p_i того, что случайная величина X принимает значение из i-го интервала, при этом $\sum_{i=1}^{k} p_i = 1$. Затем найдем теоретические частоты $m_i = n \cdot p_i$.

Гипотеза H_0 верна, если теоретические и эмпирические частоты m_i и n_i достаточно мало отличаются друг от друга. Для проверки гипотезы H_0 используем следующую статистику:

$$Q^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}. \qquad (II.56)$$

Случайная величина Q^2 имеет *χ^2-распределение с числом степеней свободы* $k - r - 1$, где k — количество интервалов, r — количество параметров теоретического распределения, оценки которых вычислялись по выборке.

Чем больше Q^2, тем хуже согласованы теоретическое и эмпирическое распределения. При достаточно большом значении Q^2 нужно отвергнуть гипотезу H_0. Поэтому используем только правостороннюю критическую область.

§ 6. КОРРЕЛЯЦИОННЫЙ И ДИСПЕРСИОННЫЙ АНАЛИЗ

1. Понятие многомерной выборки. Во время статистических наблюдений для каждого объекта в ряде случаев можно измерить значения нескольких признаков. Таким образом получается *многомерная выборка.* Если многомерную выборку обработать по значениям отдельного признака, то получится обычная обработка одномерной выборки.

Смысл обработки многомерных выборок состоит в том, чтобы *установить связи между признаками.* Связи могут быть функциональными, т. е. каждому значению одной величины соответствует определенное значение другой величины.

Связь между случайными величинами часто носит случайный характер. Такая связь называется *стохастической* или *статистической*, если изменение одной величины вызывает изменение распределения другой величины. Если среднее значение одной случайной величины функционально зависит от значений другой случайной величины, то такая статистическая зависимость называется *корреляционной.*

Далее будем рассматривать в основном двумерные выборки.

2. Эмпирическая формула. Величины X и Y могут быть функционально зависимы, но на процесс измерения их значений влияют разные, в основном случайные, факторы. Установить по результатам измерений вид фактической зависимости не так просто.

Результаты измерений или наблюдений фиксируют в таблице наблюдений (табл. 25).

Таблица 25

X	x_1	x_2	x_3	...	x_n
Y	y_1	y_2	y_3	...	y_n

Представим эти результаты на координатной плоскости (корреляционном поле) в виде точек, координатами которых являются значения признаков X и Y одного объекта (x_i, y_i), $i=1, \ldots, n$ (рис. 29).

Из рис. 29 видно, что в случае а) следует искать линейную зависимость, в случае б) — нелинейную зависимость, а в случае в) вряд ли какая-то зависимость существует.

Конкретный вид функциональной зависимости между величинами X и Y, установленный по двумерной выборке, называют *эмпирической формулой*. Если построить график эмпирической формулы на корреляционном поле, то он не должен пройти через все точки (x_i, y_i) выборки, а *быть наилучшим приближением к этим точкам*. Среднее расстояние этих точек от графика должно быть минимальным. При этом предпочтение отдается эмпирической формуле, имеющей более простой вид.

Простейшим видом эмпирической формулы является *линейная функция*

$$y = ax + b. \tag{II.57}$$

Задача установления эмпирической формулы заключается в вычислении по выборке коэффициентов a и b в формуле (II.57). Аналогично можно получить и другие функции, например $y = ax^2 + bx + c$ и т. д.

3. Нахождение линейной эмпирической формулы. Для получения линейной эмпирической формулы (II.57) имеется несколько способов: метод «натянутой нити», метод сумм и метод наименьших квадратов.

В *методе «натянутой нити»* все результаты измерений изображают в виде точек на корреляционном поле (см. рис. 29). Поэтому следует мысленно натянуть между этими точками нить так, чтобы по обе стороны осталось примерно одинаковое количество точек (точнее, чтобы их суммарные отклонения в обе стороны были равными). Возьмем на прямой, совпадающей с направлением нити, две точки с координатами (x_1, y_1) и (x_2, y_2), которые не обязательно должны присутствовать в выборке, но быть достаточно удаленными друг от друга (рис. 30). Подставив эти координаты в формулу (II.57), имеем систему линейных уравнений

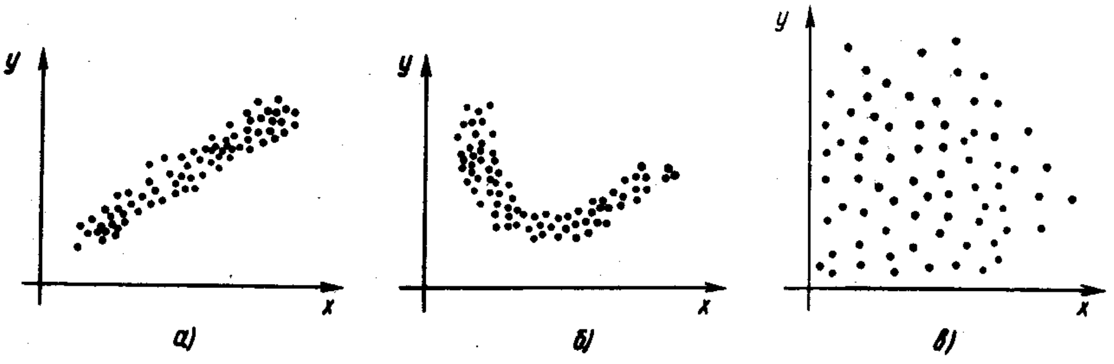


Рис. 29. Представление двумерной выборки на корреляционном поле

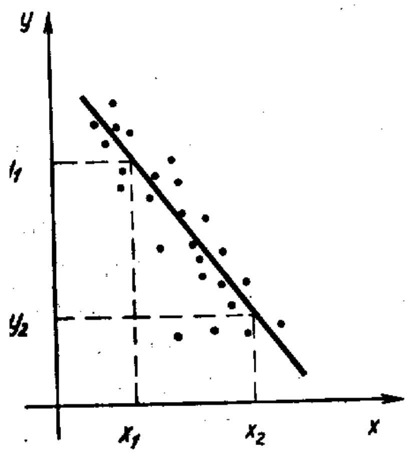


Рис. 30. Метод «натянутой нити»

$$\begin{cases} y_1 = ax_1 + b; \\ y_2 = ax_2 + b, \end{cases} \tag{II.58}$$

где неизвестными являются коэффициенты a и b. Решая систему (II.58), получаем эмпирическую формулу (II.57).

В *методе сумм* рассуждают следующим образом. Пусть имеется двумерная выборка, представленная в виде таблицы наблюдений (см. табл. 25). Предположим, что уже найдена эмпирическая формула (II.57). Подставим в нее значения случайной величины X из выборки и запишем соответствующие значения случайной величины Y в виде $\tilde{y} = ax_i + b$, $i = 1, \ldots, n$. Найдем отклонения между измеренными и вычисленными значениями Y:

$$\Delta_i = y_i - \tilde{y}_i = y_i - ax_i - b. \tag{II.59}$$

Разумно потребовать, чтобы для всей выборки *сумма этих отклонений была бы равна нулю*:

$$\sum_{i=1}^{n} \Delta_i = 0. \tag{II.60}$$

Получено одно уравнение для определения коэффициентов a и b, чтобы иметь два уравнения, разделим таблицу наблюдений (см. табл. 25) на две части. Пусть в первой из них k наблюдений, тогда во второй части $n-k$ наблюдений. Потребуем, чтобы в обеих частях таблицы выполнялось условие (II.60). В результате получаем следующую *систему линейных уравнений*:

$$\begin{cases} \sum\limits_{i=1}^{k} (y_i - ax_i - b) = 0; \\ \sum\limits_{i=k+1}^{n} (y_i - ax_i - b) = 0. \end{cases} \tag{II.61}$$

Если раскрыть скобки и просуммировать подобные члены, то получим

$$\begin{cases} \sum\limits_{i=1}^{k} y_i - a\sum\limits_{i=1}^{k} x_i - kb = 0; \\ \sum\limits_{i=k+1}^{n} y_i - a \sum\limits_{i=k+1}^{n} x_i - (n-k)b = 0. \end{cases} \tag{II.62}$$

Отсюда видно, что в обеих частях таблицы наблюдений нужно просуммировать значения x_i и y_i и затем решить систему (II.62).

Метод наименьших квадратов будет рассмотрен далее. Этот метод является самым точным, но в то же время и самым трудоемким. Метод «натянутой нити» самый простой, но и самый неточный.

4. Корреляционная таблица. По таблице наблюдений (см. табл. 25) невозможно судить о распределении случайных величин X и Y, а также об их общем распределении. По одномерной выборке мы составляли вариационные ряды, а в двумерном случае будем составлять *корреляционные таблицы.*

Варианты или интервалы значений одной случайной величины записывают в первый столбец корреляционной таблицы, а варианты или интервалы другой случайной величины — в первую строку. По каждой паре значений (x_i, y_i) решают, в какую строку попадает первое значение и в какой столбец — другое. Клетку, находящуюся на пересечении соответствующих строки и столбца, отметим палочкой. Так поступаем со всеми наблюдениями (x_i, y_i). Затем подсчитываем палочки во всех клетках и заменяем их соответствующим числом n_{XY}. Далее суммируем числа n_{XY} по строкам и столбцам и находим частоты n_X и n_Y. Объем выборки n можно получить с помощью одной из сумм $n = \sum n_X = \sum n_Y = \sum\sum n_{XY}$ (см. табл. 26—28). В корреляционной таблице могут остаться пустые клетки, это означает, что в выборке отсутствуют соответствующие сочетания значений X и Y.

5. Регрессия. Фиксируем в двумерной случайной величине значение одной случайной величины, например $Y = y$; тогда совокупность соответствующих значений другой случайной величины X можно рассматривать как отдельную случайную величину со своим законом распределения и своими числовыми характеристиками распределения, которые, как и само распределение, называют *условными.* Условное среднее значение одной случайной величины при условии, что вторая случайная величина фиксирована, обозначим так:

$$E(X \mid Y = y) \text{ или } E(Y \mid X = x).$$

Если рассмотреть условные средние значения одной случайной величины при всех значениях другой случайной величины, то получим следующие функции

$$f(x) = E(Y \mid X = x); \tag{II.63}$$

$$g(y) = E(X \mid Y = y), \tag{II.64}$$

называемые *функциями регрессии.* Если функции регрессии известны, то можно *по значению одной случайной величины прогнозировать значение другой случайной величины.*

Обычно конкретный вид функции регрессии неизвестен и определяется по двумерной выборке. При этом используют те же методы, что и при установлении эмпирической формулы.

6. Эмпирическая регрессия. Эмпирическая регрессия показывает зависимость условного среднего арифметического от значений другой случайной величины. Все вычисления выполняют по корреляционной таблице, используя следующие формулы:

$$\bar{x}_{Y=y} = \frac{1}{n_Y} \sum x_i n_{XY}; \qquad (II.65)$$

$$\bar{y}_{X=x} = \frac{1}{n_X} \sum y_j n_{XY}. \qquad (II.66)$$

Если при составлении корреляционной таблицы использованы интервалы значений случайной величины, то представителями интервалов являются их середины x_i или y_j (см. табл. 29—31).

Таким образом получаются пары чисел: значения одной случайной величины и условные средние арифметические другой случайной величины. Эти пары чисел изобразим в виде точек на координатном поле. Соединяя соседние точки плавной линией, получаем *графики эмпирической регрессии* (см. рис. 34—36).

7. Линейная регрессия. Будем искать функцию регрессии [см. формулы (II.63) и (II.66)] в самом простом — *линейном виде.* Имеем

$$y = f(x) = \alpha_1 x + \alpha_0; \qquad (II.67)$$

$$x = g(y) = \beta_1 y + \beta_0. \qquad (II.68)$$

Для определения этих функций, т. е. коэффициентов α_1, α_0, β_1, β_0, можно использовать приближенные методы, рассмотренные в п. 3. Здесь рассмотрим более точный способ — *метод наименьших квадратов.* Все рассуждения проводим для коэффициентов α_1 и α_0 первой функции. Формулы для вычисления β_1 и α_0 получаем из формул для α_1 и α_0, заменяя x на y, а y на x.

В методе наименьших квадратов первоначально рассуждаем так же, как и в методе сумм (см. п. 3). Находим отклонения между измеренными и вычисленными значениями Y [см. формулу (II.59)]:

$$\Delta_i = y_i - \tilde{y}_i = y_i - \alpha_1 x_i - \alpha_0, \quad i = 1, \dots, n.$$

В методе наименьших квадратов коэффициенты α_1 и α_0 определяют исходя из требования, состоящего в том, чтобы *сумма квадратов отклонений Δ_i была минимальной*:

$$T = \sum_{i=1}^{n} \Delta_i^2 = \sum_{i=1}^{n} (y_i - \alpha_1 x_i - \alpha_0)^2 = \min. \qquad (II.69)$$

Сумма (II.69) является функцией неизвестных коэффициентов α_1 и α_0. Для нахождения минимума запишем *частные производные*:

$$\frac{\partial T}{\partial \alpha_1} = 2\sum_{i=1}^{n} (y_i - \alpha_1 x_i - \alpha_0)(-x_i);$$

$$\frac{\partial T}{\partial \alpha_0} = 2\sum_{i=1}^{n} (y_i - \alpha_1 x_i - \alpha_0)(-1).$$

Приравняв эти частные производные нулю, получим систему линейных уравнений для определения коэффициентов α_1 и α_0. В результате несложных преобразований получаем решение системы.

$$\alpha_0 = \bar{y} - \alpha_1 \bar{x};$$

$$\alpha_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{S}_X^2}, \tag{II.70}$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\overline{S_X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$.

Эти формулы можно использовать для вычисления *коэффициентов функции линейной регрессии*, но последние часто вычисляют с помощью *коэффициента линейной корреляции*

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{S}_X \cdot \bar{S}_Y}, \tag{II.71}$$

где многие величины определены выше, а

$$\bar{S}_X = \sqrt{\bar{S}_X^2}, \ \bar{S}_Y^2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - (\bar{y})^2}.$$

Отсюда, подставляя в (II.70) формулу (II.71), получаем

$$\alpha_0 = \bar{y} - \alpha_1 \bar{x}, \ \alpha_1 = r \cdot \frac{\bar{S}_Y}{\bar{S}_X}. \tag{II.72}$$

Заменяя в этих формулах величины, связанные с x, соответствующими величинами, зависящими от y, и, наоборот, получим формулы для вычисления коэффициентов β_1 и β_0 функции линейной регрессии [см. формулы (II.68)]:

$$\beta_0 = \bar{x} - \beta_1 \bar{y}, \ \beta_1 = r \frac{\bar{S}_X}{\bar{S}_Y}. \tag{II.73}$$

Формулы (II.72) и (II.73) пригодны для вычислений с помощью ЭВМ или микрокалькулятора. На ЭВМ можно в одном цикле на основании массивов x и y, т. е. таблицы наблюдений (см. табл. 25), найти суммы $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum x_i y_i$, и затем — все величины, связанные с этими суммами. С помощью микрокалькулятора нужно также сначала вычислить указанные суммы. Некоторые микрокалькуляторы позволяют упростить эти вычисления, например микрокалькулятор «Электроника МК-51» при одном вводе допускает получение двух сумм $\sum x_i$ и $\sum x_i^2$.

Вычисления «вручную» по формулам (II.72) и (II.73) трудоемки; при этом используется корреляционная таблица. Для упрощения вычислений выполним замену переменных, аналогично тому, как это делалось при получении формул (II.7) и (II.10):

$$u_i = \frac{x_i - c_i}{k_1}, \quad v_j = \frac{y_j - c_2}{k_2}. \tag{II.74}$$

Здесь c_1 и c_2 — произвольные числа, но для простоты в качестве c_1 и c_2 обычно принимают соответствующие значения x_i или y_j, которые отвечают максимальной частоте n_{XY}. Желательно также, чтобы эта частота находилась в середине таблицы. Величины k_1 и k_2 — шаг таблицы для X и Y соответственно. Переменные u_i и v_j получаются как целые числа, близкие к нулю. Частоты n_{XY}, n_X, n_Y будем записывать в новых переменных соответственно так: n_{UV}, n_U, n_V. Далее вычисления производим по следующим формулам:

$$\bar{u} = \frac{1}{n}\sum u \cdot n_V, \quad \bar{v} = \frac{1}{n}\sum v \cdot n_V;$$

$$\bar{u^2} = \frac{1}{n}\sum u^2 \cdot n_V, \quad \bar{v^2} = \frac{1}{n}\sum v^2 \cdot n_V;$$

$$\bar{S}_U = \sqrt{\bar{u^2} - (\bar{u})^2}, \quad \bar{S}_V = \sqrt{\bar{v^2} - (\bar{v})^2};$$

$$\overline{uv} = \sum\sum u \cdot v \cdot n_{UV}; \tag{II.75}$$

$$r = \frac{\bar{u}\bar{v} - n \cdot \bar{u} \cdot \bar{v}}{n \cdot \bar{S}_U \cdot \bar{S}_V};$$

$$\bar{x} = \bar{u} \cdot k_1 + \bar{c}_1; \quad \bar{y} = \bar{v} \cdot k_2 + c_2;$$

$$\bar{S}_X = \bar{S}_U \cdot k_1; \quad \bar{S}_Y = \bar{S}_V \cdot k_2.$$

Затем для вычисления коэффициентов функций регрессии α_1, α_0, β_1, β_0 можно использовать формулы (II.72) и (II.73). Все вычисления по формулам (II.75) выполняем с помощью корреляционной таблицы (см. табл. 32—34). По найденным коэффициентам можно записать функции регрессии [см. формулы (II.67) и (II.68)] и построить их графики. Если все вычисления сделаны корректно и графики построены точно, то точка пересечения графиков имеет координаты (x, y).

8. Коэффициент линейной корреляции. Линейная функция регрессии — самый простой вид такой функции. Для оценки возможности применения линейной функции существует несколько способов: приближенных и более точных.

Самым простым способом является *оценка по расположению точек на корреляционном поле*, полученном на основании выборки (см. рис. 29). На рис. 29, *а* точки находятся вблизи воображаемой прямой и поэтому здесь разумно искать линейную функцию регрессии. На рис. 29, *б* и 29, *в* изображены ситуации, в которых линейная функция плохо соответствует действительности.

Такую же приближенную картину дает и *корреляционная таблица*. Если заполнены клетки вблизи той или другой диагонали, то стоит искать линейную функцию регрессии. Если заполнены большинство клеток или заполненные клетки образуют какую-то кривую, то использовать линейную функцию не следует.

Более точную оценку можно получить с помощью *коэффициента линейной корреляции r*. Сформулируем свойства этого коэффициента:

1) Если случайные величины X и Y независимы, то теоретически $r = 0$. Противоположное утверждение неверно и не всегда выполняется.

2) Если в формуле (II.71) поменять местами величины, зависящие от X и Y, то значение r не изменяется, значит, $r_{XY} = r_{YX}$.

3) Если случайные величины X и Y линейно зависимы, т. е. $Y = aX + b$, то

$$r = \begin{cases} +1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Если значение $|r|$ близко к единице, то надо найти линейную функцию регрессии. Если $|r| < 0{,}5$, то обычно не стоит использовать линейную функцию. В этом случае можно попробовать найти нелинейную функцию, например квадратичную функцию

$$y = \alpha_2 x^2 + \alpha_1 x + \alpha_0, \tag{II.76}$$

коэффициенты которой вычисляют также методом наименьших квадратов. Вычисление коэффициентов α_2, α_1, α_0 «вручную» трудоемко, но программа их вычисления на ЭВМ несложна.

9. Идея дисперсионного анализа. Случайные величины вызываются разными факторами, причинами. *В дисперсионном анализе исследуется значимость влияния факторов, сравнивается их влияние между собой* и т. д. С этой целью *разделяют дисперсию на независимые слагаемые*, которые потом сравнивают между собой.

Пусть, например, в результате измерения величины M получено значение X и пусть на процесс измерения влияют случайные независимые факторы A и B. Тогда отклонение $M - X = \alpha + \beta + \gamma$, где α — отклонение под влиянием фактора A, β — под влиянием фактора B, а γ — под влиянием остальных, неучтенных факторов, причем α, β и γ независимы.

Найдем дисперсию $D(M - X) = D(\alpha + \beta = \gamma)$. На основании свойств дисперсии (см. п. 4.2 § 2) получим $DX = D\alpha + D\beta + D\gamma$, где $D\alpha$ характеризует влияние фактора A, $D\beta$ — влияние фактора B, а $D\gamma$ — влияние остальных, неучтенных факторов. Дисперсия $D\gamma$ называется остаточной дисперсией. Для оценки значимости факторов A и B сравнивают соответствующие дисперсии $D\alpha$ и $D\beta$ с остаточной дисперсией $D\gamma$.

Если исследуется влияние одного фактора, то говорят об однофакторном (дисперсионном) анализе, при исследовании влияния двух факторов — о двухфакторном анализе и т. д.

10. Однофакторный анализ

10.1. Решение задачи однофакторного анализа. Рассмотрим пример. Пусть в каком-то цехе несколько станков выполняют одинаковые операции. Для планирования дальнейшей обработки деталей нужно знать, все ли станки дают одинаковую продукцию или нет. Иначе говоря, можно ли игнорировать влияние фактора (т. е. станков) на продукцию или нет?

Пусть фактор имеет m уровней (в цехе m станков). Из каждого уровня (продукции каждого станка) сделаем выборку из n элементов. Общее количество выбранных элементов обозначим $N = m \cdot n$. Вся выборка представляет собой матрицу

$$x = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

Полагая, что эта выборка сделана из нормально распределенной генеральной совокупности, и задавая уровень значимости α, нужно проверить гипотезу о равенстве средних значений на всех уровнях фактора: H_0: $\mu_1 = \mu_2 = \dots = \mu_m$. При альтернативной гипотезе H_1 не все средние значения μ_i не должны быть равными.

В качестве статистики используем величину

$$F = \frac{S_1^2}{S_2^2}, \qquad \text{(II.77)}$$

где S_1^2 — дисперсия, характеризующая влияние исследуемого фактора (*факторная дисперсия*): S_2^2 — дисперсия, характеризующая влияние остальных факторов (*остаточная дисперсия*). Если гипотеза H_0 верна, то случайная величина F имеет *F-распределение со степенями свободы* $m-1$ и $N-m=m(n-1)$.

При проверке гипотезы H_0 используем правостороннюю критическую область (см. п. 6 § 5), определяемую условием

$$P(F > f_\alpha) = \alpha.$$

Если значение статистики [см. формулу [II.77)] входит в критическую область, то гипотезу H_0 о равенстве средних значений на всех уровнях фактора отвергаем, т. е. *считаем влияние исследуемого фактора значимым*. В противном случае принимаем гипотезу H_0, т. е. считаем, что *значимость влияния фактора не установлена*.

Для нахождения величины F найдем сумму квадратов отклонений элементов выборки относительно общего среднего арифметического

$$Q = \sum_{j=1}^{m} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2, \qquad \text{(II.78)}$$

где $\bar{x} = \frac{1}{N} \sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}$. Эту общую сумму можно разделить на два независимых слагаемых:

$$Q_1 = n \sum_{j=1}^{m} (\bar{x}_j - \bar{x})^2; \qquad \text{(II.79)}$$

$$Q_2 = \sum_{j=1}^{m} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \qquad \text{(II.80)}$$

так, чтобы выполнялось равенство $Q = Q_1 + Q_2$. Здесь $\bar{x}_j = \frac{1}{n} \sum_{j=1}^{n} x_{ij}$, $j = 1, 2, \dots, m$ (групповые средние). ***Суммой квадратов*** *межгрупповых отклонений*, характеризующей влияние фактора, является сумма квадратов отклонений групповых средних относительно общей средней (сумма Q_1); Q_2 представляет собой сумму квадратов отклонений значений выборки относительно групповых средних и называется *суммой квадратов внутригрупповых отклонений*. Эта сумма характеризует влияние остальных, неучтенных факторов.

На основании сумм Q, Q_1, Q_2 можно вычислить соответствующие *дисперсии*:

$$S^2 = \frac{Q}{m \cdot n - 1}, \quad S_1^2 = \frac{Q_1}{m-1}, \quad S_2^2 = \frac{Q_2}{m(n-1)} \quad . \tag{II.81}$$

Две последние дисперсии используют при вычислении F [см. формулу (II.77)].

При практических вычислениях обычно по выборке находят Q и Q_1, а Q_2 определяют как разность Q и Q_1:

$$Q_2 = Q - Q_1. \tag{II.82}$$

10.2. Решение задачи при одинаковом количестве элементов на всех уровнях. В простейшем случае на каждом уровне фактора выбирают одинаковое количество объектов исследования. Но вычислять суммы по формулам (II.78)—(II.80) достаточно сложно. Более удобные формулы получаем преобразуя выражения (II.78) и (II.79). Имеем

$$Q = \sum_{j=1}^{m}\Big(\sum_{j=1}^{n} x_{ij}^2\Big) - \frac{1}{mn}\Big[\sum_{j=1}^{m}\Big(\sum_{i=1}^{n} x_{ij}\Big)\Big]^2;$$

$$Q_1 = \frac{1}{n}\sum_{j=1}^{m}\Big(\sum_{i=1}^{n} x_{ij}\Big)^2 - \frac{1}{mn}\Big[\sum_{j=1}^{m}\Big(\sum_{i=1}^{n} x_{ij}\Big)\Big]^2.$$

Обозначая $R_j = \sum_{i=1}^{n} x_{ij}^2$ и $L_j = \sum_{i=1}^{n} x_{ij}$, окончательно получаем

$$Q = \sum_{j=1}^{m} R_j - \frac{1}{mn}\Big(\sum_{j=1}^{m} L_j\Big)^2; \tag{II.83}$$

$$Q_i = \frac{1}{n}\sum_{j=1}^{m} L_j^2 - \frac{1}{mn}\Big(\sum_{j=1}^{m} L_j\Big)^2. \tag{II.84}$$

По исходной матрице (выборке) вычислим суммы элементов и их квадратов по столбцам (L_j и R_j, $j = 1, 2, \ldots, m$). Если x_{ij} — многоразрядные числа, то такие вычисления следует выполнять на микрокалькуляторе.

Воспользуемся заменой переменных $y_{ij} = x_{ij} - c$, где c целесообразно выбрать близким к общему среднему [в данном пособии c — целая часть $(x + 0{,}5)$]. В результате замены для y_{ij} получим следующие формулы:

$$Q = \sum_{j=1}^{m} P_j - \frac{1}{N}\Big(\sum_{j=1}^{m} T_j\Big)^2; \tag{II.85}$$

$$Q_1 = \frac{1}{n}\sum_{j=1}^{n} T_j^2 - \frac{1}{N}\left(\sum_{j=1}^{m} T_j\right)^2, \qquad (II.86)$$

где $N = m \cdot n$, $P_j = \sum_{i=1}^{n} y_{ij}^2$ и $T_j = \sum_{i=1}^{n} y_{ij}$, $j = 1, 2, ..., m$.

Все эти суммы можем вычислить «вручную» (см. табл. 36).

10.3. Решение задачи при неодинаковом количестве элементов на различных уровнях. На практике не всегда удается гарантировать одинаковое количество элементов на каждом уровне фактора. Обозначим количество элементов на j-м уровне через n_j, $j = 1, 2, ..., m$. Тогда объем выборки

$$N = \sum_{j=1}^{m} n_j,$$

формула (II.86) запишется в виде

$$Q_1 = \sum_{j=1}^{m} \frac{1}{n_j} T_j^2 - \frac{1}{N}\left(\sum_{j=1}^{m} T_j\right)^2, \qquad (II.87)$$

а суммы P_j и T_j — в виде $P_j = \sum_{i=1}^{n_j} y_{ij}^2$ **и** $T_j = \sum_{i=1}^{n_j} y_{ij}$. **Таблица для вычислений сумм аналогична табл. 36. Дисперсии [см. формулу (II.81)] таковы:**

$$S^2 = \frac{Q}{N-1}, \quad S_1^2 = \frac{Q_1}{m-1}, \quad S_2^2 = \frac{Q_2}{N-m}. \qquad (II.88)$$

Остальные формулы, используемые при решении задач, не изменяются.

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Задача. Случайная величина имеет нормальное распределение с известным средним квадратическим отклонением $\sigma=4$. Найти доверительные интервалы для оценки математического ожидания a по выборочным средним $\bar{x}$, если объем выборки $n=64$ и надежность $\gamma=0{,}95$.

Решение. Найдем t. Из отношения $2\Phi(t)=0{,}95$ получим $\Phi(t)=0{,}475$. По таблице (приложение 2 учебника [1]) находим $t=1{,}96$.

Найдем точность оценки:

$$\delta = \frac{t\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 4}{8} = 0{,}98.$$

Искомые доверительные интервалы $(\bar{x}-0{,}98;\ \bar{x}+0{,}98)$.

Смысл полученного результата таков: если будет произведено достаточно большое число выборок, то 95% из них определят такие доверительные интервалы, в которых математическое ожидание действительно будет заключено; лишь в 5% случаев оцениваемое математическое ожидание может выйти за границы доверительного интервала.

Задача. Даны выборочные варианты x_i и соответственные частоты n_i количественного признака X:

x_i	10	15	20	25	30
n_i	6	16	50	24	4

Найти методом произведений выборочные среднюю, дисперсию и среднее квадратическое отклонение.

Решение: Составим расчетную таблицу. Для этого:

1) запишем варианты x_i в первый столбец;

2) запишем частоты во второй столбец; сумму частот (100) поместим в нижнюю клетку столбца;

3) в качестве ложного нуля C выберем варианту 20 (эта варианта расположена в середине вариационного ряда); в клетке третьего столбца, которая принадлежит строке, содержащей варианту 20, пишем 0; над нулем последовательно записываем условные варианты -1, -2, а под нулем — последовательно 1, 2;

4) произведения частот на условные варианты u_i записываем в четвертый столбец; отдельно находим сумму отрицательных (-28) и отдельно сумму

положительных (32) чисел; сложив эти числа, их сумму (4) помещаем в нижнюю клетку столбца;

5) произведения частот на квадраты условных вариант запишем в пятый столбец; сумму чисел столбца (80) помещаем в нижнюю клетку столбца;

6) произведения частот на квадраты условных вариант, увеличенных на единицу запишем в шестой (контрольный) столбец; сумму чисел столбца (188) помещаем в нижнюю клетку столбца.

В итоге получим следующую расчетную таблицу:

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
10	6	−2	−12	24	6
15	16	−1	−16	16	0
20	50	0	$\boxed{-28}$	0	50
25	24	1	24	24	96
30	4	2	8	16	36
			$\boxed{32}$		
	$n=100$		$\sum n_i u_i = 4$	$\sum n_i u_i^2 = 80$	$\sum n_i (u_i+1)^2 = 188$

Контроль:

$$\sum (u_i + 1)^2 \cdot n_i = 188; \ \sum n_i u_i^2 + 2 \sum n_i u_i + n = 80 + 2 \cdot 4 + 100 = 188.$$

Совпадение найденных сумм свидетельствует о том, что вычисления произведены правильно.

Вычислим условные моменты первого и второго порядков:

$$M_1^* = \frac{\sum n_i u_i}{u} = \frac{4}{100} = 0{,}04; \quad M_2^* = \frac{\sum n_i u_i^2}{2} = \frac{80}{100} = 0{,}8.$$

Найдем шаг (разность между двумя соседними вариантами):

$$h = 15 - 10 = 5.$$

Найдем искомую выборочную среднюю:

$$\bar{x}_в = M_1^* h + C = 0{,}04 \cdot 5 + 20 = 20{,}2.$$

Найдем искомую выборочную дисперсию:

$$D_в = \left[M_2^* - \left(M_1^*\right)^2\right] \cdot h^2 = \left[0{,}8 - (0{,}04)^2\right] \cdot 5^2 = 19{,}96.$$

Найдем искомое выборочное среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в} = \sqrt{19{,}96} \approx 4{,}47.$$

Задача. Найти выборочное уравнение прямой линии

$$\overline{y}_x - \overline{y} = r_в \frac{\sigma_y}{\sigma_x} (x - \overline{x})$$

регрессии Y на X по данным корреляционной табл. 1.

Таблица 1

Y	X					
	20	25	30	35	40	n_y
16	4	6	—	—	—	10
26	—	8	10	—	—	18
36	—	—	32	3	9	44
46	—	—	4	12	6	22
56	—	—	—	1	5	6
n_x	4	14	46	16	20	$n=100$

Решение. Составим корреляционную табл. 2 в условных вариантах, выбрав в качестве ложных нулей $C_1=30$ и $C_2=36$ (каждая из этих вариант расположена в середине соответствующего вариационного ряда).

Таблица 2

v	u					
	−2	−1	0	1	2	n_v
−2	4	6	—	—	—	10
−1	—	8	10	—	—	18
0	—	—	32	3	9	44
1	—	—	4	12	6	22
2	—	—	—	1	5	6
n_u	4	14	46	16	20	$n=100$

Найдем $\bar{u}$ и $\bar{v}$:

$$\bar{u}=\frac{\sum n_u u}{n}=\frac{4(-2)+14\cdot(-1)+46\cdot 0+16\cdot 1+20\cdot 2}{100}=0{,}34;$$

$$\bar{v}=\frac{\sum n_v v}{n}=\frac{10\cdot(-2)+18\cdot(-1)+44\cdot 0+22\cdot 1+6\cdot 2}{100}=-0{,}04.$$

Найдем вспомогательные величины $\overline{u^2}$ и $\overline{v^2}$:

$$\overline{u^2}=\frac{\sum n_u u^2}{n}=\frac{4\cdot 4+14\cdot 1+16\cdot 1+20\cdot 4+46\cdot 0}{100}=1{,}26;$$

$$\overline{v^2}=\frac{\sum n_v v^2}{n}=\frac{10\cdot 4+18\cdot 1+0\cdot 44+22\cdot 1+6\cdot 4}{100}=1{,}04.$$

Найдем σ_u и σ_v:

$$\sigma_u=\sqrt{\overline{u^2}-(\bar{u})^2}=\sqrt{1{,}26-(0{,}34)^2}\approx 1{,}07;$$

$$\sigma_v=\sqrt{\overline{v^2}-(\bar{v})^2}=\sqrt{1{,}04-(0{,}04)^2}\approx 1{,}02.$$

Найдем $\Sigma n_{uv}uv$, для чего составим расчетную табл. 3.

Суммируя числа последнего столбца табл. 3, находим

$$\sum_v vU=\sum n_{uv}uv=82.$$

Для контроля вычислений находим сумму чисел последней строки:

$$\sum_u uV=\sum n_{uv}uv=82.$$

Совпадение сумм свидетельствует о правильности вычислений.

Указания к составлению табл. 3. Произведение частоты n_{uv} на варианту u, т. е. $n_{uv}u$ записывают в правом верхнем углу клетки, содержащей частоту. Например, в правых верхних углах клеток первой строки записаны произведения: $4\cdot(-2)=-8$; $6\cdot(-1)=-6$.

Складывают все числа, помещенные в правых верхних углах клеток одной строки, и их сумму помещают в клетку этой же строки «столбца U». Например, для первой строки $U=-8+(-6)=-14$.

Наконец, умножают варианту v на U и полученное произведение записывают в соответствующую клетку «столбца vU». Например, в первой строке таблицы $v=-2$, $U=-14$, следовательно, $vU=(-2)\ (-14)=28$.

Сложив все числа «столбца vU», получают сумму $\sum\limits_v vU$, которая равна искомой сумме $\Sigma n_{uv}uv$. Например, для табл. 3 $\sum\limits_v vU=82$; следовательно, искомая сумма $\Sigma n_{uv}uv=82$.

Для контроля расчета аналогичные вычисления производят по столбцам: произведения $n_{uv}v$ записывают в левый нижний угол клетки, содержащей частоту; все числа, помещенные в левых нижних углах одного столбца, складывают и их сумму помещают в «строку V»; наконец, умножают каждую варианту u на V и результат записывают в клетках последней строки.

Сложив все числа последней строки, получают сумму $\sum\limits_u uV$, которая также равна искомой сумме $\Sigma n_{uv}uv$. Например, для табл. 3 $\sum\limits_u uV=82$; следовательно, $\Sigma n_{uv}uv=82$.

Найдем искомый выборочный коэффициент корреляции:

$$r_в=\frac{\sum n_{uv}uv-n\bar{u}\bar{v}}{n\cdot\sigma_u\sigma_v}=\frac{82-100\cdot 0{,}34\cdot(0{,}04)}{100\cdot 1{,}07\cdot 1{,}02}=0{,}76.$$

Таблица 3

v	u = −2	u = −1	u = 0	u = 1	u = 2	$U=\Sigma n_{uv}u$	vU
−2	−8 4 −8	−6 6 −12	—	—	—	−14	28
−1	—	−8 8 −8	0 10 −10	—	—	−8	8
0	—	—	0 32 0	3 3 0	18 9 0	21	0
1	—	—	0 4 4	12 12 12	12 6 6	24	24
2	—	—	—	1 1 2	10 5 10	11	22
$V=\Sigma n_{uv}v$	−8	−20	−6	14	16		$\sum\limits_{v} vU=82$
uV	16	20	0	14	32	$\sum\limits_{u} uV=82$	контроль

Найдем шаги h_1 и h_2 (разности между любыми двумя соседними вариантами):

$$h_1 = 25 - 20 = 5; \quad h_2 = 26 - 16 = 10.$$

Найдем $\overline{x}$ и $\overline{y}$, учитывая, что $C_1=30$, $C_2=36$:

$$\overline{x} = \overline{u}h_1 + C_1 = 0{,}34\cdot 5 + 30 = 31{,}70;$$

$$\overline{y} = \overline{v}h_2 + C_2 = (-0{,}04)\cdot 10 + 36 = 35{,}60.$$

Найдем σ_x и σ_y:

$$\sigma_x = h_1\sigma_u = 5\cdot 1{,}07 = 5{,}35;$$

$$\sigma_y = h_2\sigma_v = 10\cdot 1{,}02 = 10{,}2.$$

Подставив найденные величины в соотношение (*), получим искомое уравнение прямой линии регрессии $У$ на X:

$$\overline{y}_x - 35{,}6 = 0{,}76\cdot\frac{10{,}2}{5{,}35}(x - 31{,}7)$$

или окончательно

$$\overline{y}_x = 1{,}45x - 10{,}36.$$

Задача. По двум независимым выборкам объемов $n=60$ и $m=50$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\overline{x}=1250$ и $\overline{y}=1275$. Генеральные дисперсии известны: $D(X)=120$, $D(У)=100$.

При уровне значимости $\alpha=0{,}01$ проверить нулевую гипотезу $H_0: M(X)=M(У)$ при конкурирующей гипотезе $H_1: M(X)\neq M(У)$.

Решение. Найдем наблюдаемое значение критерия:

$$Z_{набл} = \frac{\overline{x} - \overline{y}}{\sqrt{\dfrac{D(X)}{n} + \dfrac{D(Y)}{m}}}.$$

Подставив сюда данные условия, получим

$$Z_{набл} = -12{,}5.$$

Найдем правую критическую точку $z_{кр}$ по равенству

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0{,}01}{2} = 0{,}495.$$

По таблице функции Лапласа (приложение 2 учебника [1]) находим

$$z_{кр} = 2{,}58.$$

Так как $|Z_{набл}| > z_{кр}$, нулевую гипотезу отвергают. Другими словами, выборочные средние различаются значимо.

Правило. Для того чтобы при заданном уровне значимости проверить нулевую гипотезу H_0: генеральная совокупность распределена нормально, надо сначала вычислить теоретические частоты, а затем наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \sum \frac{(n_i - n_i')^2}{n_i'} \qquad (2)$$

и по таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $k=s-3$ найти критическую точку $\chi^2_{\text{кр}}(\alpha, k)$.

Если $\chi^2 < \chi^2_{\text{кр}}$, нет оснований отвергать нулевую гипотезу.

Если $\chi^2 > \chi^2_{\text{кр}}$, нулевую гипотезу отвергают.

Замечание 1. Объем выборки должен быть достаточно велик, во всяком случае не менее 50. Каждая группа должна содержать не менее 5—8 вариант; малочисленные группы следует объединить в одну, суммируя частоты.

Замечание 2. Поскольку возможны ошибки первого и второго рода, в особенности, если согласование теоретических и эмпирических частот «слишком хорошее», следует проявлять осторожность. Например, можно повторить опыт, увеличить число наблюдений, воспользоваться другими критериями, построить график распределения, вычислить асимметрию и эксцесс (гл. XVII, § 8).

Замечание 3. Для контроля вычислений формулу (2) приводят к виду

$$\chi^2_{\text{набл}} = \sum \frac{n_i^2}{n_i'} - n.$$

Рекомендуем читателю выполнить это преобразование, для чего надо в формуле (2) возвести в квадрат разность частот, сократить результат на n_i' и учесть, что $\sum n_i = n$, $\sum n_i' = n$.

Пример. При уровне значимости 0,05 проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

эмпирические частоты . . .	6	13	38	74	106	85	30	14
теоретические частоты . . .	3	14	42	82	99	76	37	13

Решение. Вычислим $\chi^2_{\text{набл}}$, для чего составим расчетную табл. 1:

Таблица 1

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	6	3	3	9	3	36	12
2	13	14	−1	1	0,07	169	12,07
3	38	42	−4	16	0,38	1444	34,38
4	74	82	−8	64	0,78	5476	66,78
5	106	99	7	49	0,49	11236	113,49
6	85	76	9	81	1,07	7225	95,07
7	30	37	−7	49	1,32	900	24,32
8	14	13	1	1	0,08	196	15,08
Σ	366	366			$\chi^2_{\text{набл}} = 7,19$		373,19

Контроль:

$$\chi^2_{\text{набл}} = 7{,}19; \quad \sum \frac{n_i^2}{n_i'} - n = 373{,}19 - 366 = 7{,}19.$$

Вычисления произведены правильно.

Учитывая, что число групп выборки (число различных вариант) $s=8$, найдем число степеней свободы:

$$k = s - 3 = 8 - 3 = 5.$$

По таблице критических точек распределения χ^2 (приложение 5 учебника [1]), по уровню значимости $\alpha=0{,}05$ и числу степеней свободы $k=5$ находим критическую точку $\chi^2_{\text{кр}}(0{,}05,\ 5)=11{,}1$. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отвергать нулевую гипотезу. Другими словами, расхождение эмпирических и теоретических частот незначимое. Следовательно, данные наблюдений согласуются с гипотезой о нормальном распределении генеральной совокупности.

Как следует из предыдущего, сущность критерия Пирсона состоит в сравнении эмпирических и теоретических частот. Ясно, что эмпирические частоты находят из опыта. Как найти теоретические частоты, если предполагается, что генеральная совокупность распределена нормально? Ниже указан один из способов решения этой задачи.

1. Весь интервал наблюдаемых значений X делят на s частичных интервалов $(x_i,\ x_{i+1})$ одинаковой длины. Находят середины частичных интервалов $x_i^* = \frac{x_i + x_{i+1}}{2}$; в качестве частоты n_i варианты x_i^* принимают число вариант, которые попали в i-й интервал. В итоге получают последовательность равноотстоящих вариант и соответствующих им частот:

$$x_1^*,\ x_2^*,\ \ldots,\ x_s^*$$

$$n_1,\ n_2,\ \ldots,\ n_s,$$

причем $\Sigma n_i = n$.

2. Вычисляют, например, методом произведений или сумм выборочную среднюю $\overline{x}^*$ и выборочное среднее квадратическое отклонение σ^*.

3. Нормируют случайную величину X, т. е. переходят к величине $Z = \frac{X - \overline{x}^*}{\sigma^*}$, и вычисляют концы интервалов $(z_i,\ z_{i+1})$:

$$z_i = \frac{x_i - \overline{x}_i^*}{\sigma^*},\quad z_{i+1} = \frac{x_{i+1} - \overline{x}^*}{\sigma^*},$$

причем наименьшее значение z, т. е. z_1, полагают равным $-\infty$, а наибольшее, т. е. z_s, полагают равным ∞. 4. Вычисляют теоретические вероятности p_i попадания X в интервалы $(x_i,\ x_{i+1})$ по равенству

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

($\Phi(z)$ — функция Лапласа) и, наконец, находят искомые теоретические частоты $n_i' = np_i$.

Задача. Найти теоретические частоты по заданному интервальному распределению выборки объема $n=200$, предполагая, что генеральная совокупность распределена нормально (табл. 2).

Как следует из предыдущего, сущность критерия Пирсона состоит в сравнении эмпирических и теоретических частот. Ясно, что эмпирические частоты находят из опыта. Как найти теоретические частоты, если предполагается, что генеральная совокупность распределена нормально? Ниже указан один из способов решения этой задачи.

1. Весь интервал наблюдаемых значений X делят на s частичных интервалов $(x_i,\ x_{i+1})$ одинаковой длины. Находят середины частичных интервалов $x_i^* = \dfrac{x_i + x_{i+1}}{2}$; в качестве частоты n_i варианты x_i^* принимают число вариант, которые попали в i-й интервал. В итоге получают последовательность равноотстоящих вариант и соответствующих им частот:

$$x_1^*,\ x_2^*,\ \ldots,\ x_s^*$$

$$n_1,\ n_2,\ \ldots,\ n_s,$$

причем $\Sigma n_i = n$.

2. Вычисляют, например, методом произведений или сумм выборочную среднюю $\overline{x}^*$ и выборочное среднее квадратическое отклонение σ^*.

3. Нормируют случайную величину X, т. е. переходят к величине $Z = \dfrac{X - \overline{x}^*}{\sigma^*}$, и вычисляют концы интервалов $(z_i,\ z_{i+1})$:

$$z_i = \frac{x_i - \overline{x}_i^*}{\sigma^*},\quad z_{i+1} = \frac{x_{i+1} - \overline{x}^*}{\sigma^*},$$

причем наименьшее значение z, т. е. z_1, полагают равным $-\infty$, а наибольшее, т. е. z_s, полагают равным ∞. 4. Вычисляют теоретические вероятности p_i попадания X в интервалы $(x_i,\ x_{i+1})$ по равенству

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

($\Phi(z)$ — функция Лапласа) и, наконец, находят искомые теоретические частоты $n_i' = np_i$.

Задача. Найти теоретические частоты по заданному интервальному распределению выборки объема $n=200$, предполагая, что генеральная совокупность распределена нормально (табл. 2).

Таблица 2

Номер интервала i	Границы интервала x_i	x_{i+1}	Частота n_i	Номер интервала i	Границы интервала x_i	x_{i+1}	Частота n_i
1	4	6	15	6	14	16	21
2	6	8	26	7	16	18	24
3	8	10	25	8	18	20	20
4	10	12	30	9	20	22	13
5	12	14	26				
							$n=200$

Решение. 1. Найдем середины интервалов $x_i^* = \frac{x_i + x_{i+1}}{2}$. Например, $x_1^* = \frac{4+6}{2} = 5$. Поступая аналогично, получаем последовательность равноотстоящих вариант и соответствующих им частот:

$$x_i^* \quad 5 \quad 7 \quad 9 \quad 11 \quad 13 \quad 15 \quad 17 \quad 19 \quad 21$$

$$n_i \quad 15 \quad 26 \quad 25 \quad 30 \quad 26 \quad 21 \quad 24 \quad 20 \quad 13.$$

2. Пользуясь методом произведений, найдем выборочную среднюю и выборочное среднее квадратическое отклонение:

$$\overline{x}^* = 12{,}63, \quad \sigma^* = 4{,}695.$$

3. Найдем интервалы $(z_i, \ z_{i+1})$, учитывая, что $\overline{x}^* = 12{,}63$; $\sigma^* = 4{,}695$; $1/\sigma^* = 0{,}213$, для чего составим расчетную табл. 3.

Таблица 3

Номер интервала i	Границы интервала x_i	x_{i+1}	$x_i - \overline{x}^*$	$x_{i+1} - \overline{x}^*$	Границы интервала $z_i = \frac{x_i - \overline{x}^*}{\sigma^*}$	$z_{i+1} = \frac{x_{i+1} - \overline{x}^*}{\sigma^*}$
1	4	6	—	−6,63	$-\infty$	−1,41
2	6	8	−6,63	−4,63	−1,41	−0,99
3	8	10	−4,63	−2,63	−0,99	−0,56
4	10	12	−2,63	−0,63	−0,156	−0,13
5	12	14	−0,63	1,37	−0,13	0,29
6	14	16	1,37	3,37	0,29	0,72
7	16	18	3,37	5,37	0,72	1,14
8	18	20	5,37	7,37	1,14	1,57
9	20	22	7,37	—	1,57	∞

4. Найдем теоретические вероятности p_i и искомые теоретические частоты $n'_i = np_i$, для чего составим расчетную табл. 4.

Таблица 4

Номер интервала i	Границы интервала z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$p_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = np_i = 200p_i$
1	$-\infty$	$-1,41$	$-0,5$	$-0,4207$	0,0793	15,86
2	$-1,41$	$-0,99$	$-0,4207$	$-0,3389$	0,0818	16,36
3	$-0,99$	$-0,56$	$-0,3389$	$-0,2133$	0,1266	25,32
4	$-0,56$	$-0,13$	$-0,2123$	$-0,0517$	0,1606	32,16
5	$-0,13$	0,29	$-0,0517$	0,1141	0,1658	33,16
6	0,29	0,72	0,1141	0,2642	0,1501	30,02
7	0,72	1,14	0,2642	0,3729	0,1087	21,74
8	1,14	1,57	0,3729	0,4418	0,0689	13,78
9	1,57	∞	0,4418	0,5	0,0582	11,64
					$\Sigma p_i = 1$	$\Sigma n'_i = 200$

Искомые частоты помещены в последнем столбце табл. 4.

Задачи для самостоятельного решения

Для повторения и закрепления пройденного материала рекомендуется решить приведенные ниже задачи.

1. Найти эмпирическую функцию по данному распределению выборки:

$$x_i \quad 2 \; 5 \; 7 \; 8$$
$$n_i \quad 1 \; 3 \; 2 \; 4.$$

О т в е т. $F^*(x) = \begin{cases} 0 & \text{при } x \leqslant 2, \\ 0{,}1 & \text{при } 2 < x \leqslant 5, \\ 0{,}4 & \text{при } 5 < x \leqslant 7, \\ 0{,}6 & \text{при } 7 < x \leqslant 8, \\ 1 & \text{при } x > 8. \end{cases}$

2. Построить полигоны частот и относительных частот по данному распределению выборки:

$$x_i \quad 2 \; 3 \; 5 \; 6$$
$$n_i \quad 10 \; 15 \; 5 \; 20.$$

3. Построить гистограммы частот и относительных частот по данному распределению выборки (см. табл. 5).

Таблица 5

Номер частичного интервала	Частичный интервал	Сумма частот вариант частичного интервала
1	2—7	5
2	7—12	10
3	12—17	25
4	17—22	6
5	22—27	4

4. По выборке объема $n=51$ найдена смещенная оценка $D_в=5$ генеральной дисперсии. Найти несмещенную оценку генеральной дисперсии.

О т в е т. $s^2 = \dfrac{n}{n-1} D_в = 5{,}1.$

5. В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: $x_1=8$, $x_2=9$, $x_3=11$, $x_4=12$. Найти: 1) выборочную среднюю результатов измерений; 2) выборочную и исправленную дисперсии ошибок прибора.

О т в е т: 1) $\overline{x}_в=10$; 2) $D_в=2{,}5$, $s^2=10/3$.

6. Произведено 5 равноточных измерений расстояния от орудия до цели одним прибором со средним квадратическим отклонением случайных ошибок измерений $\sigma=40$ м. Найти доверительный интервал для оценки истинного расстояния a до цели с надежностью 0,95, зная результаты измерений в метрах: $x_1=1950$, $x_2=1980$, $x_3=2000$, $x_4=2030$, $x_5=2050$.

О т в е т. $1960{,}8<a<2039{,}2$.

7. Станок-автомат штампует валики. По выборке объема $n=100$ вычислена выборочная средняя изготовленных валиков. Найти с надежностью 0,95 точность δ, с которой выборочная средняя оценивает математическое ожидание изготовляемых валиков, зная, что среднее квадратическое отклонение их размеров равно 2 мм.

О т в е т. $\delta = t\dfrac{\sigma}{\sqrt{n}} = 1{,}96\dfrac{2}{\sqrt{100}} = 0{,}392$ мм.

8. Найти минимальный объем выборки, при котором с надежностью 0,925 точность оценки математического ожидания нормально распределенной генеральной совокупности по выборочной средней будет равна 0,2, если известно среднее квадратическое отклонение генеральной совокупности $\sigma=1{,}5$.

Ответ. $n=179$.

9. По данным 16 независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\bar{x}_в=42{,}8$ и исправленное среднее квадратическое отклонение $s=8$. Оценить истинное значение a измеряемой величины с надежностью $\gamma=0{,}999$.

Ответ. $34{,}66<a<50{,}94$.

10. По двум независимым выборкам объемов $n_1=21$ и $n_2=16$, извлеченным из нормальных генеральных совокупностей X и $У$, найдены исправленные выборочные дисперсии $s_X^2=3{,}6$ и $s_Y^2=2{,}4$. При уровне значимости $\alpha=0{,}05$ проверить нулевую гипотезу $H_0: D(X)=D(У)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X)>D(У)$.

Указание. $F_{кр}$ (0,05; 20; 15) = 2,33.

Ответ: $F_{набл}=1{,}5$. Нет оснований отвергнуть нулевую гипотезу.

11. По двух независимым выборкам объемов $n_1=11$ и $n_2=14$, извлеченным из нормальных генеральных совокупностей X и $У$, найдены исправленные выборочные дисперсии $s_X^2=25{,}5$ и $s_Y^2=8{,}5$. При уровне значимости $\alpha=0{,}05$ проверить нулевую гипотезу $H_0: D(X)=D(У)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X)>D(У)$.

Указание. $F_{кр}$ (0,05; 10; 13) = 2,67

Ответ. $F_{набл}=3$. Нулевая гипотеза отвергается.

12. По двум независимым выборкам объемов $n=30$ и $m=20$, извлеченным из нормальных генеральных совокупностей X и $У$, найдены выборочные средние $\bar{x}=97$ и $\bar{y}=94$. Генеральные дисперсии известны: $D(X)=120$, $D(У)=100$. При уровне значимости $\alpha=0{,}05$ проверить нулевую гипотезу H_0: $M(X)=M(У)$ о равенстве математических ожиданий при конкурирующей гипотезе H_1: $M(X)\neq M(У)$.

Ответ. $Z_{набл}=1$; $z_{кр}=1{,}96$. Нет оснований отвергать нулевую гипотезу.

13. По двум независимым выборкам объемов $n=50$ и $m=40$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\bar{x}=370$ и $\bar{y}=350$. Генеральные дисперсии известны: $D(X)=50$, $D(У)=120$. При уровне значимости $\alpha=0{,}01$ проверить нулевую гипотезу $H_0: M(X)=M(У)$ о равенстве математических ожиданий при конкурирующей гипотезе $H_1: M(X)\neq M(У)$.

Ответ. $Z_{набл}=10$; $z_{кр}=2{,}58$. Нулевая гипотеза отвергается.

14. Установить, пользуясь критерием Пирсона, при уровне значимости $\alpha=0{,}05$, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n_i', которые вычислены, исходя из предположения, что генеральная совокупность распределена нормально:

1)	n_i	5	10	20	8	7		
	n_i'	6	14	18	7	5;		
2)	n_i	14	18	32	70	20	36	10
	n_i'	10	24	34	80	18	22	12.

Ответ: 1) случайно ($\chi^2_{кр}$ (0,05; 2) = 6,0; $\chi^2_{набл}=2{,}47$); 2) значимо ($\chi^2_{кр}$(0,05; 4) = 9,5; $\chi^2_{набл}=13{,}93$).

ПРИЛОЖЕНИЯ

Приложение 1

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \ell^{-\frac{x^2}{2}}$

x	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	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	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	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	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
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	0.0001	0.0001

Приложение 2

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int\limits_0^x \ell^{-\frac{z^2}{2}} dz$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0.00	*0.0000*	0.5	*0.1915*	1.00	*0.3413*	1.50	*0.4332*	2.00	*0.4772*	2.50	*0.4938*
0.01	*0.0040*	0.51	*0.1950*	1.01	*0.3438*	1.51	*0.4345*	2.01	*0.4778*	2.51	*0.4940*
0.02	*0.0080*	0.52	*0.1985*	1.02	*0.3461*	1.52	*0.4357*	2.02	*0.4783*	2.52	*0.4941*
0.03	*0.0120*	0.53	*0.2019*	1.03	*0.3485*	1.53	*0.4370*	2.03	*0.4788*	2.53	*0.4943*
0.04	*0.0160*	0.54	*0.2054*	1.04	*0.3508*	1.54	*0.4382*	2.04	*0.4793*	2.54	*0.4945*
0.05	*0.0199*	0.55	*0.2088*	1.05	*0.3531*	1.55	*0.4394*	2.05	*0.4798*	2.55	*0.4946*
0.06	*0.0239*	0.56	*0.2123*	1.06	*0.3554*	1.56	*0.4406*	2.06	*0.4803*	2.56	*0.4948*
0.07	*0.0279*	0.57	*0.2157*	1.07	*0.3577*	1.57	*0.4418*	2.07	*0.4808*	2.57	*0.4949*
0.08	*0.0319*	0.58	*0.2190*	1.08	*0.3599*	1.58	*0.4429*	2.08	*0.4812*	2.58	*0.4951*
0.09	*0.0359*	0.59	*0.2224*	1.09	*0.3621*	1.59	*0.4441*	2.09	*0.4817*	2.59	*0.4952*
0.10	*0.0398*	0.6	*0.2257*	1.10	*0.3643*	1.60	*0.4452*	2.10	*0.4821*	2.60	*0.4953*
0.11	*0.0438*	0.61	*0.2291*	1.11	*0.3665*	1.61	*0.4463*	2.11	*0.4826*	2.61	*0.4955*
0.12	*0.0478*	0.62	*0.2324*	1.12	*0.3686*	1.62	*0.4474*	2.12	*0.4830*	2.62	*0.4956*
0.13	*0.0517*	0.63	*0.2357*	1.13	*0.3708*	1.63	*0.4484*	2.13	*0.4834*	2.63	*0.4957*
0.14	*0.0557*	0.64	*0.2389*	1.14	*0.3729*	1.64	*0.4495*	2.14	*0.4838*	2.64	*0.4959*
0.15	*0.0596*	0.65	*0.2422*	1.15	*0.3749*	1.65	*0.4505*	2.15	*0.4842*	2.65	*0.4960*
0.16	*0.0636*	0.66	*0.2454*	1.16	*0.3770*	1.66	*0.4515*	2.16	*0.4846*	2.66	*0.4961*
0.17	*0.0675*	0.67	*0.2486*	1.17	*0.3790*	1.67	*0.4525*	2.17	*0.4850*	2.67	*0.4962*
0.18	*0.0714*	0.68	*0.2517*	1.18	*0.3810*	1.68	*0.4535*	2.18	*0.4854*	2.68	*0.4963*
0.19	*0.0753*	0.69	*0.2549*	1.19	*0.3830*	1.69	*0.4545*	2.19	*0.4857*	2.70	*0.4965*
0.20	*0.0793*	0.7	*0.2580*	1.20	*0.3849*	1.70	*0.4554*	2.20	*0.4861*	2.72	*0.4967*
0.21	*0.0832*	0.71	*0.2611*	1.21	*0.3869*	1.71	*0.4564*	2.21	*0.4864*	2.74	*0.4969*
0.22	*0.0871*	0.72	*0.2642*	1.22	*0.3888*	1.72	*0.4573*	2.22	*0.4868*	2.76	*0.4971*
0.23	*0.0910*	0.73	*0.2673*	1.23	*0.3907*	1.73	*0.4582*	2.23	*0.4871*	2.78	*0.4973*
0.24	*0.0948*	0.74	*0.2704*	1.24	*0.3925*	1.74	*0.4591*	2.24	*0.4875*	2.80	*0.4974*
0.25	*0.0987*	0.75	*0.2734*	1.25	*0.3944*	1.75	*0.4599*	2.25	*0.4878*	2.82	*0.4976*
0.26	*0.1026*	0.76	*0.2764*	1.26	*0.3962*	1.76	*0.4608*	2.26	*0.4881*	2.84	*0.4977*
0.27	*0.1064*	0.77	*0.2794*	1.27	*0.3980*	1.77	*0.4616*	2.27	*0.4884*	2.86	*0.4979*
0.28	*0.1103*	0.78	*0.2823*	1.28	*0.3997*	1.78	*0.4625*	2.28	*0.4887*	2.88	*0.4980*
0.29	*0.1141*	0.79	*0.2852*	1.29	*0.4015*	1.79	*0.4633*	2.29	*0.4890*	2.90	*0.4981*
0.30	*0.1179*	0.8	*0.2881*	1.30	*0.4032*	1.80	*0.4641*	2.30	*0.4893*	2.92	*0.4982*
0.31	*0.1217*	0.81	*0.2910*	1.31	*0.4049*	1.81	*0.4649*	2.31	*0.4896*	2.94	*0.4984*
0.32	*0.1255*	0.82	*0.2939*	1.32	*0.4066*	1.82	*0.4656*	2.32	*0.4898*	2.96	*0.4985*
0.33	*0.1293*	0.83	*0.2967*	1.33	*0.4082*	1.83	*0.4664*	2.33	*0.4901*	2.98	*0.4986*
0.34	*0.1331*	0.84	*0.2995*	1.34	*0.4099*	1.84	*0.4671*	2.34	*0.4904*	3.00	*0.4987*
0.35	*0.1368*	0.85	*0.3023*	1.35	*0.4115*	1.85	*0.4678*	2.35	*0.4906*	3.04	*0.4988*
0.36	*0.1406*	0.86	*0.3051*	1.36	*0.4131*	1.86	*0.4686*	2.36	*0.4909*	3.06	*0.4989*
0.37	*0.1443*	0.87	*0.3078*	1.37	*0.4147*	1.87	*0.4693*	2.37	*0.4911*	3.08	*0.4990*
0.38	*0.1480*	0.88	*0.3106*	1.38	*0.4162*	1.88	*0.4699*	2.38	*0.4913*	3.08	*0.4990*
0.39	*0.1517*	0.89	*0.3133*	1.39	*0.4177*	1.89	*0.4706*	2.39	*0.4916*	3.12	*0.4991*
0.40	*0.1554*	0.9	*0.3159*	1.40	*0.4192*	1.90	*0.4713*	2.40	*0.4918*	3.16	*0.4992*
0.41	*0.1591*	0.91	*0.3186*	1.41	*0.4207*	1.91	*0.4719*	2.41	*0.4920*	3.20	*0.4993*
0.42	*0.1628*	0.92	*0.3212*	1.42	*0.4222*	1.92	*0.4726*	2.42	*0.4922*	3.26	*0.4994*
0.43	*0.1664*	0.93	*0.3238*	1.43	*0.4236*	1.93	*0.4732*	2.43	*0.4925*	3.32	*0.4995*
0.44	*0.1700*	0.94	*0.3264*	1.44	*0.4251*	1.94	*0.4738*	2.44	*0.4927*	3.40	*0.4997*
0.45	*0.1736*	0.95	*0.3289*	1.45	*0.4265*	1.95	*0.4744*	2.45	*0.4929*	3.60	*0.4998*
0.46	*0.1772*	0.96	*0.3315*	1.46	*0.4279*	1.96	*0.4750*	2.46	*0.4931*	3.80	*0.49993*
0.47	*0.1808*	0.97	*0.3340*	1.47	*0.4292*	1.97	*0.4756*	2.47	*0.4932*	4.00	*0.499968*
0.48	*0.1844*	0.98	*0.3365*	1.48	*0.4306*	1.98	*0.4761*	2.48	*0.4934*	4.50	*0.499997*
0.49	*0.1879*	0.99	*0.3389*	1.49	*0.4319*	1.99	*0.4767*	2.49	*0.4936*	5.00	*0.4999997*

Таблица значений $\chi^2_{k\alpha}$, соответствующие вероятности $p = P\left\{\chi^2_k > \chi^2_{k,\alpha}\right\}$,

где χ^2_k имеет χ^2- распределение с k степенями свободы

k	α										
	0,99	**0,95**	**0,9**	**0,5**	**0,25**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**
1	0	0	0,02	0,45	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	0,02	0,1	0,21	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	0,11	0,35	0,58	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	0,3	0,71	1,06	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	0,55	1,15	1,61	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	0,87	1,64	2,2	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	1,24	2,17	2,83	6,35	9,04	12	14,1	16	18,5	20,3	24,3
8	1,65	2,73	3,49	7,34	10,2	13,4	15,5	17,5	20,1	22	26,1
9	2,09	3,33	4,17	8,34	11,4	14,7	16,9	19	21,7	23,6	27,9
10	2,56	3,94	4,87	9,34	12,5	16	18,3	20,5	23,2	25,2	29,6
11	3,05	4,57	5,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	3,57	5,23	6,3	11,3	14,8	18,5	21	23,3	26,2	28,3	32,9
13	4,11	5,89	7,04	12,3	16	19,8	22,4	24,7	27,7	29,8	34,5
14	4,66	6,57	7,79	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	5,23	7,26	8,55	14,3	18,2	22,3	25	27,5	30,6	32,8	37,7
16	5,81	7,96	9,31	15,3	19,4	23,5	26,3	28,8	32	34,3	39,3
17	6,41	8,67	10,1	16,3	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	7,01	9,39	10,9	17,3	21,6	26	28,9	31,5	34,8	37,2	42,3
19	7,63	10,1	11,7	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	8,26	10,9	12,4	19,3	23,8	28,4	31,4	34,2	37,6	40	45,3
21	8,9	11,6	13,2	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	9,54	12,3	14	21,3	26	30,8	33,9	36,8	40,3	42,8	48,3
23	10,2	13,1	14,8	22,3	27,1	32	35,2	38,1	41,6	44,2	49,7
24	10,9	13,8	15,7	23,3	28,2	33,2	36,4	39,4	43	45,6	51,2
25	11,5	14,6	16,5	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	12,2	15,4	17,3	25,3	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	12,9	16,2	18,1	26,3	31,5	36,7	40,1	43,2	47	49,6	55,5
28	13,6	16,9	18,9	27,3	32,6	37,9	41,3	44,5	48,3	51	56,9
29	14,3	17,7	19,8	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	15	18,5	20,6	29,3	34,8	40,3	43,8	47	50,9	53,7	59,7

Приложение 4

Таблица значений $t_{k\beta}$, соответствующие вероятности $\beta = P\{|t_k| > t_{k,\beta}\}$,

где случайная величина t_k имеет распределение *Стьюдента* с k степенями свободы

k	β							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	1,886	2,92	4,303	6,965	9,925	14,09	22,33	31,6
3	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,61
5	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	1,44	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	1,397	1,86	2,306	2,896	3,355	3,833	4,501	5,041
9	1,383	1,833	2,262	2,821	3,25	3,69	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,428	3,93	4,318
13	1,35	1,771	2,16	2,65	3,012	3,372	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,14
15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	1,337	1,746	2,12	2,583	2,921	3,252	3,686	4,015
17	1,333	1,74	2,11	2,567	2,898	3,222	3,646	3,965
18	1,33	1,734	2,101	2,552	2,878	3,197	3,61	3,922
19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,85
21	1,323	1,721	2,08	2,518	2,831	3,135	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	1,319	1,714	2,069	2,5	2,807	3,104	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	1,316	1,708	2,06	2,485	2,787	3,078	3,45	3,725
26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,66
30	1,31	1,697	2,042	2,457	2,75	3,03	3,385	3,646
40	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
60	1,296	1,671	2	2,39	2,66	2,915	3,232	3,46
120	1,289	1,658	1,98	2,358	2,617	2,86	3,16	3,373
10000	1,282	1,645	1,96	2,327	2,576	2,808	3,091	3,291

Содержание

Кафедра МММЭ
Экономический факультет

Учебное издание

Пороодников Виктор Дмитриевич

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие