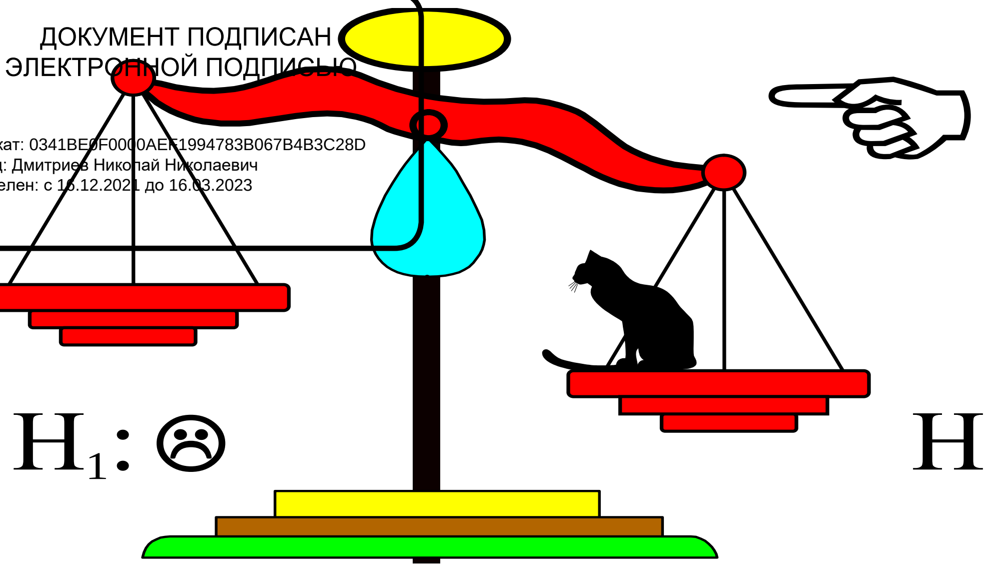
Министерство сельского хозяйства Российской Федерации
ФГОУ ВО Иркутский государственный аграрный университет им. А.А. Ежевского
Кафедра Математики

Н.И. Овчинникова,
Е.В. Елтошкина,
С.Е. Васильева



Практикум по математической статистике
часть I

учебное пособие

Иркутск - 2015

УДК 519. 22 (076.5)
ББК 22.172.73

Рекомендовано к изданию Научно-методическим советом Иркутского государственного аграрного университета им. А.А. Ежевского (протокол № 2 от 26 октября 2015 г.

Составители: Н.И. Овчинникова - доктор технических наук, профессор, заведующий кафедрой Математики;
Е.В. Елтошкина – к.т.н., доцент кафедры Математики;
С.Е. Васильева - старший преподаватель кафедры Математики.

Рецензенты: Ю.М. Краковский - д.т.н., член-корреспондент СО АН ВШ, профессор кафедры Информатики и математического моделирования Иркутского ГАУ им. А.А. Ежевского;
В.Г. Власов - д.ф.-м.н., профессор, заведующий кафедрой Математики Иркутского научно-исследовательского технического университета

Практикум предназначен для студентов биологических, экономических и технических направлений подготовки бакалавриата в Иркутском аграрном университете им. А.А. Ежевского, а также для магистрантов, аспирантов и преподавателей нематематических кафедр, применяющих в своей практике методы математической статистики и интересующихся анализом экспериментальных данных.

Учебное пособие составлено на основе действующих стандартов и рабочих программ по математической статистике. Каждый раздел издания содержит краткие теоретические сведения, образцы решения типовых заданий, задачи для самостоятельного решения при выполнении лабораторных и контрольных работ. В конце практикума приведены типовой расчет, статистические таблицы и рекомендуемая литература. Разработанное учебное пособие можно использовать на практических и лабораторных занятиях по математической статистке, при выполнении аудиторных и домашних контрольных работ, при подготовке к итоговому экзамену или зачету, а также к Федеральному тестированию.

Внимательная работа с представленным практикумом послужит студентам усвоению основных понятий математической статистики, привитию навыков решения различных типов статистических задач, построению линейных моделей на основе статистических данных.

С О Д Е Р Ж А Н И Е

В В Е Д Е Н И Е

Изучение дисциплины "Математическая статистика" является продолжением курса "Теория вероятностей" и строится, исходя из требуе-мого уровня базовой подготовки студентов, обучающихся по биологическим, экономическим и техническим направлениям подготовки бакалавриата. При разработке методов и формул математической статистики непосредственно используют результаты из теории вероятностей. В то же время информация, полученная путем статистического анализа, служит исходной базой для вероятностных расчетов. Эти два раздела математики очень тесно взаимосвязаны и переплетены между собой, что составляют неразрывное целое. В то же время, многие разделы математической статистики, используемые в первую очередь на практике, исторически создавались параллельно, а иногда и независимо от теории вероятностей. Поэтому сложилась своя терминология, которая иногда немного отличается от той, которая использовалась в теории вероятностей.

При изучении теории вероятностей мы рассматривали два различных типа случайностей: *случайные события* и *случайные величины*. Выяснили, что во многих случаях можно *теоретическим путем* рассчитывать вероятности случайных событий (*классическое определение вероят-ности*). В то же время существуют ситуации (и таких, вообще говоря, гораздо больше), когда классическое определение применить невоз-можно. Тогда единственным путем для определения вероятности случай-ных событий остается эксперимент, наблюдения. При изучении *случай-ных величин* оказалось, что мы могли прогнозировать их поведение, предсказывать вероятности попаданий случайных величин в любые ин-тересующие нас интервалы, если только нам известен *закон распределе-ния*. При этом во многих случаях теория может предсказать, с каким законом распределения мы имеем дело в данной конкретной ситуации. А как быть в тех случаях, когда о характере закона распределения мы ничего сказать не можем? Единственный выход – проведение эксперимента, наблюдений над случайной величиной и построение закона распределения по результатам этих наблюдений.

Математическая статистика позволяет обосновать ответ на вопросы: случайно или закономерно изучаемое явление, сколько необходи-мо провести наблюдений для объективного суждения об изучаемом явлении, как зависит результативный признак от факторного, какой фактор сильнее влияет на результат и т.д. Основная цель математической статистики – получение и обработка данных для принятия решений в задачах планирования, управления, прогнозирования и др.

Целью изучения дисциплины «математическая статистика» явля-ется обучение студентов основам теории математической статистики и ее методам. *Задачи* изучения дисциплины состоят в реализации требова-ний, установленных в Федеральных государственных образовательных стандартах высшего образования к подготовке специалистов агропромышленного комплекса.

В результате изучения курса студенты должны:

– *овладеть:* знаниями общих основ математической статистики; общими навыками и умениями проведения статистического исследования и анализа его результатов;

– *знать:* принципы и методы обработки результатов статистического наблюдения; сущность обобщающих статистических показателей – абсолютных статистических величин, средних показателей вариации, динамики, взаимосвязи, основы анализа статистических данных;

– *уметь:* организовать и провести выборочное наблюдение; строить статистические модели явлений; исчислять различные статистические показатели (абсолютные и относительные, средние, вариации, тесноты связи); анализировать статистические данные и формулировать выводы, вытекающие из анализа данных.

1 ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.1 Генеральная совокупность и выборка

Одним из фундаментальных понятий математической статистики является понятие генеральной совокупности. Под *генеральной совокупностью* понимают множество качественно однородных элементов (объектов, изделий) самой различной природы. Генеральная совокупность называется *конечной* или *бесконечной* в зависимости от того конечна или бесконечна совокупность всех мыслимых наблюдений. Бесконечную генеральную совокупность можно истолковывать как предельный случай конечной, когда число объектов, порождаемых данным реальным комплексом условий, неограниченно возрастает. *Объем генеральной совокупности* – это число объектов этой совокупности.

Вторым основным понятием математической статистики является понятие выборочной совокупности (выборки). *Выборка* из данной генеральной совокупности – это результат ограниченного ряда наблюдений случайной величины, т.е. специально организованная часть генеральной совокупности. Выборку можно рассматривать как некий эмпирический аналог генеральной совокупности, поскольку обследование всей генеральной совокупности либо слишком трудоемко, либо принципиально невозможно. Число наблюдений, образующих выборку, называют *объемом выборки.* Выборку осуществляют двумя способами. Если исследуемый объект возвращается в генеральную совокупность после исследования, то такая выборка называется *повторной*. Если исследуемый объект не возвращается в генеральную совокупность после исследования, то такая выборка называется *бесповторной*. *Простая выборка* – это значения, расположенные в порядке получения.

Сущность *выборочного метода* математической статистики состоит в том, что по некоторой части генеральной совокупности (т.е. по выборке) выносится суждение о ее свойствах в целом. Для того, чтобы выводы, получаемые в результате статистической обработки данных, были *достоверными*, т.е. полно и адекватно представляли интересующие нас свойства генеральной совокупности, выборка должна быть репрезентативной или *представительной*. Выборка будет *репрезентативной*, если каждый элемент выборки отобран из генеральной совокупности случайно и все элементы этой совокупности имеют одинаковую вероятность попасть в выборку.

Признак - свойство элементов генеральной совокупности. Признаки делятся на качественные и количественные. *Качественные признаки* - это те признаки, которыми объект либо обладает, либо не обладает. К ним относятся: пол, цвет волос или национальность и т.д. Такие признаки не являются физически измеримыми, однако они могут быть двузначными или многозначными. *Количественные признаки* являются измеримыми и определяются путем измерений, взвешиваний и подсчетов. В соответствии с этим различают дискретные и непрерывные количественные признаки. *Дискретные признаки* могут принимать лишь изолированные значения, отличающиеся друг от друга на некоторую конечную величину. Примером таких признаков является академическая система успеваемости: 5 - отлично, 4 - хорошо и т.д. Совокупность возможных значений, среди которых изменяется (варьируется) дискретный признак называется системой вариант. *Непрерывные признаки* могут принимать любые значения на некотором числовом интервале, отличающиеся друг от друга на сколь угодно малую величину. К таким признакам относятся, например, возраст, рост и вес человека.

Множество допустимых значений признаков как качественного, так и количественного вида характеризуются типом шкалы в которой они изменяются. Различают три основных типа шкал: *номинальная* или шкала наименований, *порядковая* и *количественная*.

1.2 Статистические ряды и их графические представления

Вариационный ряд – это выборка, в которой данные расположены в порядке возрастания. Каждый член вариационного ряда называется *вариантой*. *Статистический ряд* – это выборка, в которой указано сколько раз встречается каждая варианта. *Сгруппированный статистический ряд* – выборка из непрерывной генеральной совокупности или выборка с большим объемом, разбитая на интервалы.

Пусть X - одномерный количественный признак и в результате его измерения наблюдалось n его значений (вариант) $x_1, \ldots, x_n$, среди которых могут быть одинаковые. Пусть среди имеющихся n вариант имеется k различных $x_1, \ldots, x_k, k \le n$, причем x_1 встречается n_1 раз, x_2 – n_2 раз и т.д. x_k - n_k раз. Количество повторений каждого из значений выборки называется *частотами*. Сумма всех частот должна быть равна объему выборки

$$\sum_{i=1}^{k} n_i = n . \qquad (1.1)$$

Точечным или дискретным вариационным рядом (*распределением частот* или *частотным распределением*) называется n различных вариант, записанных в возрастающем порядке вместе с соответствующими частотами.

x_i	x_1	x_2	. . .	x_k
n_i	n_1	n_2	. . .	n_k

(1.2)

Частотное распределение обычно записывается в одном из видов: в таблице с частотами n_i , (2) и через *относительные частоты*, ω_i , заданных в виде доли или в виде процента:

$$\omega_i = \frac{n_i}{n}, \omega_i = \frac{n_i}{n} \cdot 100\ \%. \qquad (1.3)$$

x_i	x_1	x_2	. . .	x_k
ω_i	ω_1	ω_2	. . .	ω_k

(1.4)

Очевидно, что $$\sum_{i=1}^{k} \omega_i = 1\ (100\ \%) . \qquad (1.5)$$

Вариационный ряд вида (4), построенный по относительным частотам, является статистической аппроксимацией ряда распределения вероятностей исследуемой дискретной случайной величины.

Пример 1. Привести оценки 45 студентов по курсу «математика» в порядке сдачи экзамена: 5, 3, 3, 4, 2, 4, 4, 3. 5, 4, 4, 5, 5, 4, 4, 3, 3, 3, 2, 5, 5, 4, 4, 4, 3, 4, 3, 4, 5, 4, 4, 4, 4, 3, 3, 4, 3, 4, 3, 2, 3, 2, 3, 3, 3 к точечному вариационному ряду и оценить успеваемость.

Решение. При таком представлении информации трудно делать какие-либо выводы об успеваемости. Произведем группировку данных путем подсчета количества различных оценок.

оценки, x_i	2	3	4	5
количество, n_i	4	16	18	7

Как видим, вместо 45 чисел осталось 8, при этом повысилась информативность таблицы, более 50% студентов сдали предмет на «хорошо» и «отлично». Данный пример показывает, что эти данные лучше сгруппировать, то есть разделить их на однородные группы по некоторому признаку.

В зависимости от объема исходных данных и области допустимых значений одномерного количественного признака, частотные распределения также подразделяются на дискретные и интервальные. Если объем выборки n велик ($n \geq 50$) и мы имеем дело с одномерной непрерывной величиной (или с одномерной дискретной, число возможных значений достаточно велико, больше 10), то эти варианты группируют. При этом выбирается определенное число интервалов группировки и, получают, таким образом, *интервальное частотное распределение*, т.е. переходят к так называемым «группированным» выборочным данным. Благодаря группировке данные приобретают систематизированный вид. Алгоритм группировки массива данных $x_1, \ldots, x_n$ состоит из следующих шагов:

1) находят минимальную, x_{min} и максимальную, x_{max} варианты;

2) весь диапазон значений признака $[x_{min}, x_{max}]$ разбивают на k интервалов (количество интервалов не должно быть меньше 8-10 и больше 20-25). Для примерной ориентации в выборе k можно пользоваться приближенной формулой Стерджесса:

$$k = 1 + \log_2 n, \ \ k = 1 + 3{,}332 \lg n, \ k = \sqrt{n}. \qquad (1.6)$$

Полученное число нужно округлить до ближайшего целого (число интервалов дробным быть не может);

3) определяют длину интервала по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}; \qquad (1.7)$$

4) находят граничные точки каждого из интервалов $(x_i; x_{i+1})$

$$x_1 = x_{\min}, x_2 = x_{\min} + h, \ldots \text{ и т.д.}; \qquad (1.8)$$

5) подсчитывают число вариант n_i, попавших в интервал $(x_i; x_{i+1})$, причем варианты, попавшие на границы интервалов, относят только к одному из

интервалов, результат заносят в таблицу, представляющую собой *интервальный вариационный ряд* (*интервальное частотное распределение*)

$[x_i, x_{i+1})$	$[x_1, x_2)$	$[x_2, x_3)$	...	$[x_k, x_{k+1})$
n_i	n_1	n_2	...	n_k

(1.9)

От интервального ряда (9) можно вновь перейти к точечному ряду вида (2), если в качестве значений случайной величины, соответствующего i-ому интервалу, взять его середину

$$x_i^0 = \frac{x_i + x_{i+1}}{2}. \tag{1.10}$$

Несмотря на видимую несхожесть, ряды (2), (4) и (9) отражают одно и то же фактическое распределение признака.

Пример 2. Случайную выборку объема $n = 60$ привести к интервальному вариационному ряду.

42,7	37,6	45,1	55,4	50,7	30,7	31,9	43,8	47,5	42,1	57,7	21,3
45,5	45,3	46,2	50,9	33,2	40,4	40,0	59,6	46,0	44,0	37,0	44,7
64,6	58,9	31,3	59,2	45,5	53,3	43,6	37,5	33,0	42,6	39,6	51,5
47,4	48,6	33,8	29,2	33,7	48,5	44,4	37,6	45,1	36,0	26,4	38,0
49,7	52,1	42,7	49,0	31,9	52,2	60,6	44,6	43,9	59,4	53,7	45,9.

Решение. Определим наименьшее и наибольшее значение варианты

$$x_{\max} = 64,6; \quad x_{\min} = 21,3.$$

Найдем число интервалов по формуле Стерджесса и ширину каждого интервала: $k = 1 + 3{,}332 \lg 60 \approx 7$, $h = \frac{64,6 - 21,3}{7} = \frac{43,3}{7} = 6,2$.

Составим таблицу, определяющую интервальный вариационный ряд и соответствующий ему точечный где: $[x_i; x_{i+1})$- полуоткрытые интервалы для распределения вариант, концы которых определяются по формуле (8); x_i^0 - середина полученных интервалов; n_i - частота; ω_i - относительная частота.

$[x_i; x_{i+1})$	x_i^0	n_i	ω_i
[21,3; 27,5)	24,4	2	0,0333
[27,5; 33,7)	30,6	7	0,1167
[33,7; 39,9)	36,8	9	0,15
[39,9; 46,1)	43,0	20	0,3333
[46,1; 52,3)	49,2	12	0,2
[52,3; 58,5)	55,4	4	0,0667
[58,5; 64,7)	61,6	6	0,1
$\sum_{i=1}^{7} n_i(\omega_i)$		60	1

Обычно табличное распределение частот дополняют его графическим представлением. Схематически все множество графических представлений статистических данных разделяют на два класса: диаграммы и линейные изображения. *Диаграмма* (от греческого diagramma - изображение, чертеж, рисунок) - это графическое изображение, наглядно показывающее соотношение между сравниваемыми величинами. Диаграммы бывают различных видов: полосовые (ленточные), столбиковые, квадратные, круговые, секторные, фигурные, радиальные, знак Варзара.

Полосовые диаграммы особенно наглядны при сравнении величин, связанных между собой в единое целое. Ширина полос должна быть одинаковой. По длине полосы разбиваются на части, пропорциональные изображаемым величинам. Основным видом *столбиковых диаграмм* являются гистограммы. При построении *квадратных и круговых диаграмм* площади квадратов или кругов выражают изображаемые величины. Круговые *секторные диаграммы* применяют для графического изображения составных частей целого. Для их построения необходимо изображаемые данные выразить в градусах, т.к. 1% составляет 3,6 градусов, то соответствующие показатели для определения центральных улов надо умножить на 3.6. Чтобы легче различать сектора используют различную раскраску или штриховку. *Радиальные диа-*

граммы строятся в полярной системе координат и используются для изображения признаков, периодически изменяющихся во времени (в большинстве своем сезонных колебаний). Вычисляется среднее арифметическое, затем строится окружность радиуса равного среднему арифметическому. Данная окружность делится на нужное число секторов (обычно 12) и на каждом радиальном направлении откладываются точки в соответствии со значениями x_i. *Фигурные диаграммы* строятся двумя основными способами: данные изображаются либо фигурами различных размеров, либо разной численностью фигур одинакового размера. Второй способ чаще используется, каждая фигура содержит определенное число единиц признака и сравнение осуществляется по числу фигурок. При этом допускается дробление знака до половины. *Stem & leaf* - данные можно представить в виде десятков и единиц, где десятки - это стебли, единицы - лепестки. *Диаграмма "знак Варзара"* названа в честь русского статистика. С помощью данной диаграммы можно изображать многомерные признаки на плоскости посредством прямоугольников с разным соотношением между основанием и высотой. Одна из компонент признака изображается основанием прямоугольника, вторая его высотой, третья - равная произведению двух других размером получившейся площади.

К классу *линейных графиков* относятся полигон, кумулятивная кривая (кумулята), огива. *Полигоном частот* (*относительных частот*) называют ломаную, отрезки которой соединяют точки $(x_1, n_1),\ldots, (x_k, n_k)$ или $(x_1, \omega_1),\ldots, (x_k, \omega_k)$ (рис. 1). Полигон относительных частот выглядит точно также, как полигон частот, только меняется масштаб на оси ординат.

Кумулятивная кривая или *кумулята* (кривая сумм) - ломаная, составленная по последовательно суммированным, т.е. накопленным частотам или относительным частотам.

n_i

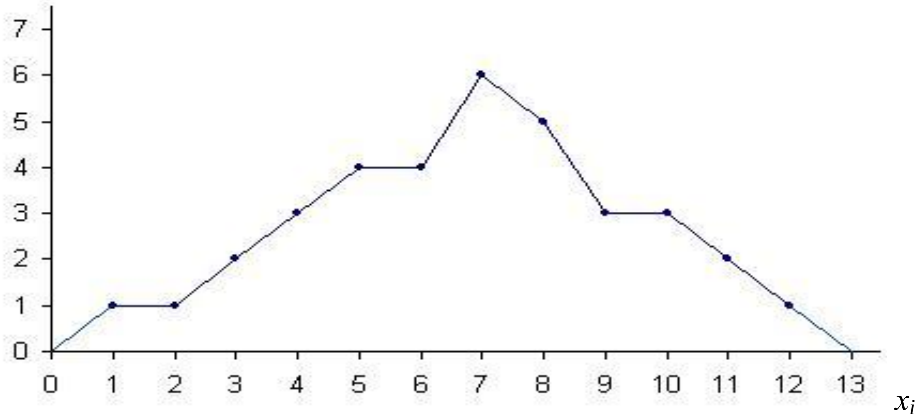


Рисунок 1 - Полигон частот

Накопленная частота n_i^* получается суммированием частот значений, предшествующих данному, с частотой n_i, т.е.

$$n_1^* = n_1, n_2^* = n_1 + n_2, n_3^* = n_1 + n_2 + n_3, \ldots, \; n_i^* = n_1 + n_2 + \ldots + n_i. \qquad (1.11)$$

Отсюда, накопленная частота крайнего правого значения (или максимального элемента выборки) равна объему выборки n. При построении кумулятивной кривой дискретного признака на ось абсцисс наносятся значения признака, а ординатами служат нарастающие итоги частот. Соединением вершин ординат прямыми линиями получают кумуляту, (рис. 2). Кумулятивную кривую называют *полигоном накопленных частот*. Если на ось ординат нанести значение признака, а накопленные частоты - на ось абсцисс, то получим кривую, называемую *огивой*.

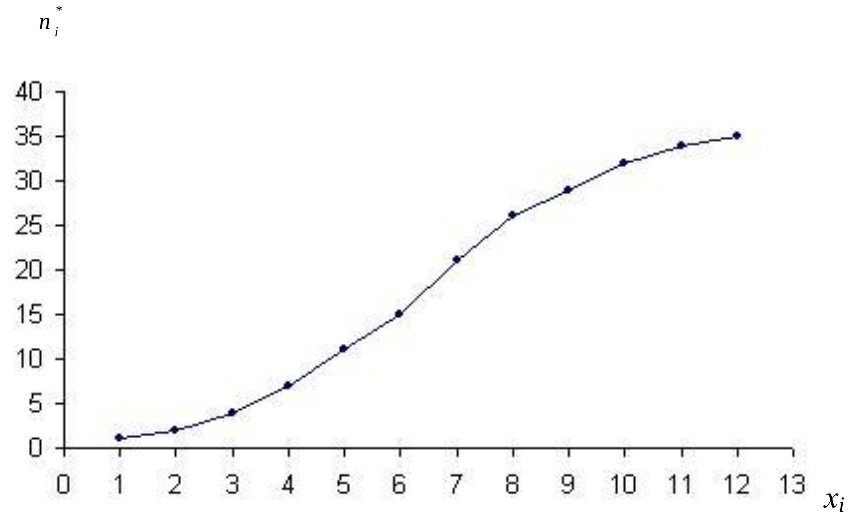


Рисунок 2 - Кумулята

Для непрерывной случайной величины закон распределения задается в виде плотности распределения $f(x)$. Вероятность попадания случайной величины в любой интервал $(x_i; x_{i+1})$ - это площадь под графиком плотности, опирающаяся на интервал $(x_i; x_{i+1})$.

Когда по данным наблюдений найдены относительные частоты ω_i попадания в разные интервалы, можно и построить прямоугольники соответствующей площади. Тогда полученная фигура, состоящая из последовательности примыкающих друг к другу прямоугольников, называется *гистограммой.* Ширина этих прямоугольников равна ширине h интервалов группировки и откладывается по оси абсцисс, высота измеряется по оси ординат и пропорциональна частоте n_i или относительной частоте ω_i. В первом случае имеем *гистограмму частот* с высотами прямоугольников $f = n_i/h$ (рис. 3) и общей площадью, равной объему выборки n. Во втором – *гистограмму относительных частот* с высотами прямоугольников n_i/nh и общей площадью, равной 1.

Ступенчатая ломаная, ограничивающая в этом случае сверху постоянную фигуру, является статистической аппроксимацией функции плотности вероятности генеральной совокупности. Если соединить плавной кривой середины верхних оснований прямоугольников, то получим также приближенное представление графика функции плотности распределения исследуемой непрерывной величины.

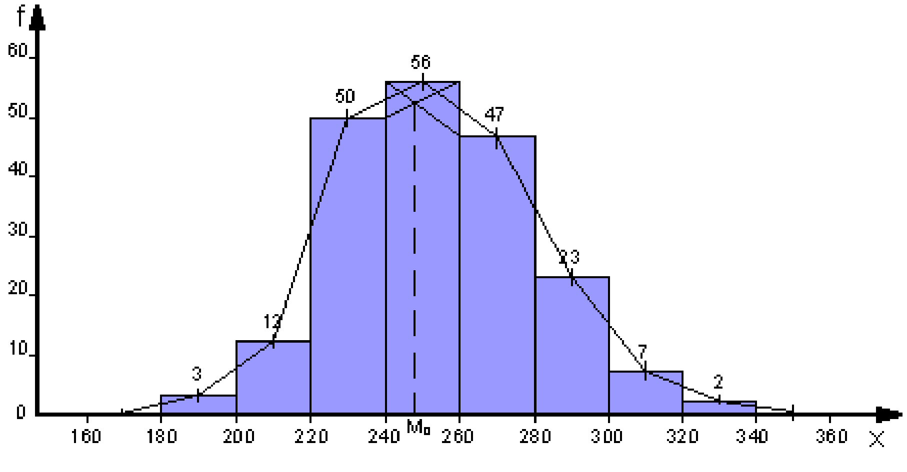


Рисунок 3 – Гистограмма частот

Чем больше опытов проводится, тем ближе построенная гистограмма к теоретической плотности распределения.

Иногда высоты прямоугольников в гистограмме не делят на h, но указывают над столбиками значение высоты и над осью ординат пишут, что ее значение надо делить на h. Такую гистограмму называют *масштабированной*.

Пример 3. Из крупного стада коров произведена случайная выборка. Получено 30 вариант удоя коров за 300 дней лактации (в *ц*): 1,55; 1,57; 2,01; 2,15; 1,55; 1,56; 1,63; 1,71; 1,57; 1,63; 1,55; 1,63; 1,56; 1,71; 1,71; 2,01; 1,56; 1,71; 1,63; 1,57; 1,63; 1,55; 1,56; 1,71; 1,72; 1,80; 1,81; 1,57; 2,15; 2,01.
Построить дискретный вариационный ряд и изобразить его графически.

Решение. Занесем исходные данные в таблицу (в первую строку x_i), располагая их в порядке возрастания. Во второй строке укажем соответствующие им частоты n_i. Расположим варианты по возрастанию

1,55; 1,55; 1,55; 1,55; 1,56; 1,56; 1,56; 1,56; 1,57; 1,57; 1,57; 1,57; 1,63; 1,63; 1,63; 1,63; 1,63; 1,71; 1,71; 1,71; 1,71; 1,71; 1,72; 1,80; 1,81; 2,01; 2,01; 2,01; 2,15; 2,15.

Таблица к примеру 3

x_i	1,55	1,56	1,570	1,63	1,71	1,72	1,80	1,81	2,01	2,15
n_i	4	4	4	5	5	1	1	1	3	2
n_i^*	4	8	12	17	22	23	24	25	28	30

Таким образом, таблица 1 является дискретным вариационным рядом, где $n = \sum_{i=1}^{k} n_i = 30$ – объем выборки.

В третьей строке таблицы 1 указаны накопленные частоты n_i^*, показывающие, во скольких наблюдениях признак принял значения, меньшие заданного значения x_i.

Построим полигон частот в системе координат $x_i O n_i$ и кумулятивную кривую в системе координат $x_i O n_i^*$ для полученного дискретного вариационного ряда.

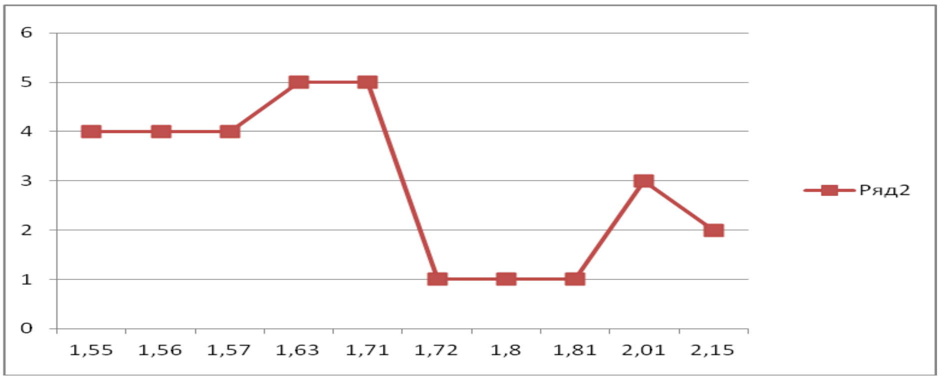


Рисунок 4 - Полигон частот (*Пример* 3)

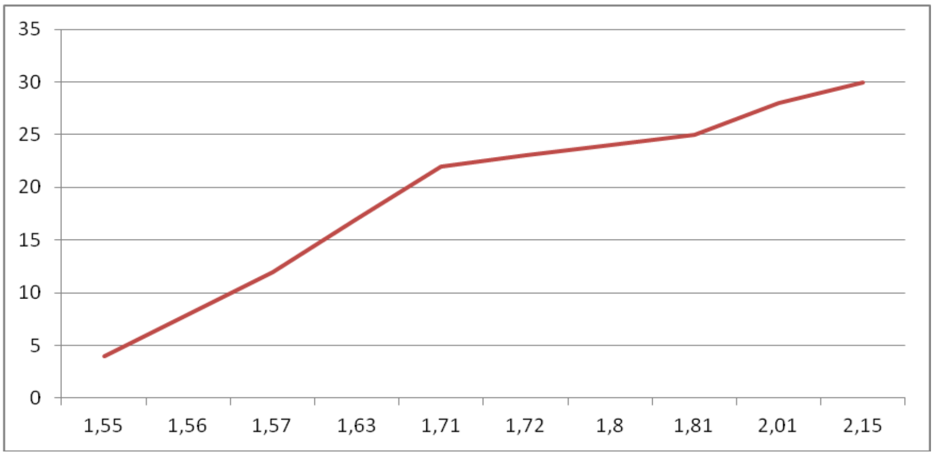


Рисунок 5 - Кумулята (*Пример* 3)

Пример 4. Обследование оплаты труда (у.е.) 100 работников фирмы дало следующие результаты

338	336	312	322	342	302	296	358	342	334
348	304	323	310	340	314	298	312	322	350
304	302	336	334	304	292	324	331	324	334
314	338	324	292	298	342	338	331	325	324
326	314	312	342	342	321	352	304	302	332
314	304	312	340	290	322	326	316	328	340
324	320	358	304	340	290	318	332	354	324
304	321	356	358	328	332	304	282	330	314
342	322	350	298	316	298	332	342	316	326
308	321	302	304	322	296	322	338	324	323

Составить интервальный вариационный ряд, построить гистограмму и полигон относительных частот. Перейти к соответствующему дискретному ряду и построить полигон относительных частот.

Решение. Определим оптимальную длину интервала:

$$h = \frac{x_{max} - x_{\min}}{1 + 3,3\lg n} = \frac{358 - 282}{1 + 3,3\lg 100} = \frac{76}{1 + 3,3 \cdot 2} \approx 10$$

Запишем вариационный ряд:

$[x_i, x_{i+1})$	n_i	ω_i
[280,290)	1	0,01
[290,300)	10	0,1
[300,310)	14	0,14
[310,320)	14	0,14
[320,330)	25	0,25
[330,340)	16	0,16
[340,350)	12	0,12
[350,360)	8	0,08
Контроль	$\sum_{i=1}^{8} n_i = 100$	$\sum_{i=1}^{8} \omega_i = 1$

Построим гистограмму относительных частот в системе координат $x_i O \omega_i$.

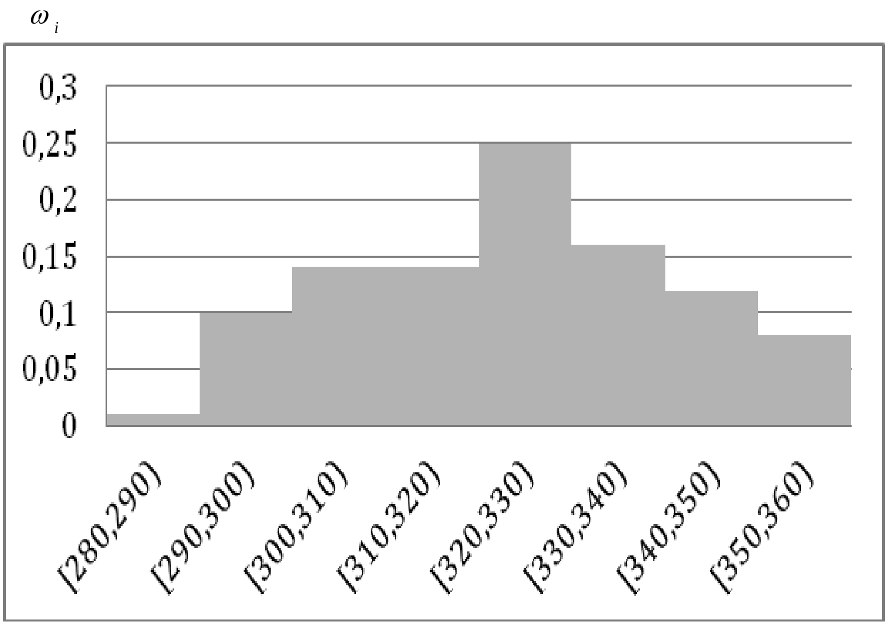


Рисунок 6 – Масштабированная гистограмма относительных частот (частостей) (*Пример* 4)

Для преобразования полученного интервального статистического ряда в дискретный в качестве значений x_i признака X возьмем середины интервалов, а относительные частоты оставим прежние.

x_i	285	295	305	315	325	335	345	355
ω_i	0,01	0,1	0,14	0,14	0,25	0,16	0,12	0,08

Полигон относительных частот (частостей) имеет вид:

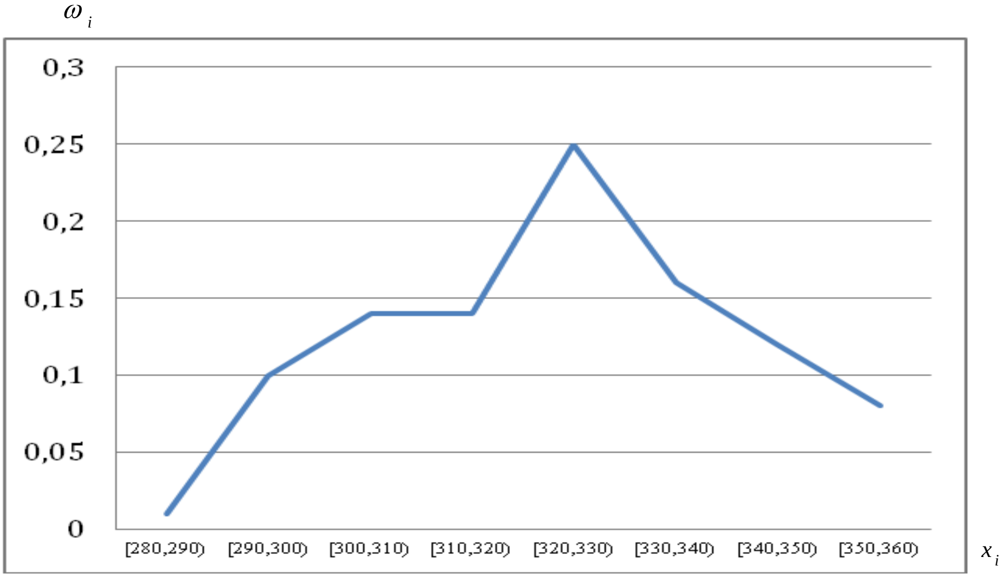


Рисунок 7 - Полигон относительных частот (частостей) (*Пример* 4)

1.3 Эмпирическая функция распределения

Для любой случайной величины универсальным способом задания закона распределения является *функция распределения* $F(x)$. По определению, функция распределения - это вероятность попадания случайной величины в область, лежащую слева от аргумента x: $F(x) = P(X<x)$.

Из эксперимента мы можем найти *относительную частоту* попадания в область, лежащую слева от аргумента, и это будет эмпирическая или статистическая функция распределения. *Статистическая (эмпирическая) функция*, $F^*(x)$ – это функция, которая для каждого значения аргумента равна относительной частоте попадания опытных данных в область, лежащую слева от аргумента:

$$F^*(x) = \frac{\sum_{x_i < x} n_i}{n} = \frac{n_x}{n} = \frac{n^*(x)}{n}. \tag{1.12}$$

График статистической функции распределения $F^*(x)$ представляет собой ступенчатую фигуру со скачками в точках, определяемых элементами выборки. Технология построения статистической функции распределения $F^*(x)$ такая же, как и для теоретической функции распределения в случае дискретных случайных величин: суммируются относительные частоты для всех опытных значений, лежащих слева от аргумента (как раньше суммировались вероятности). На основании закона больших чисел статистическая функция распределения сходится по вероятности к теоретической функции распределения $F(x)$ генеральной совокупности, когда объем выборки неограниченно возрастает (рис. 8). Так как с увеличением числа опытов относительные частоты приближаются к вероятностям, делаем вывод, что с увеличением количества наблюдений ступенчатая статистическая функция распределения приближается к теоретической функции распределения.

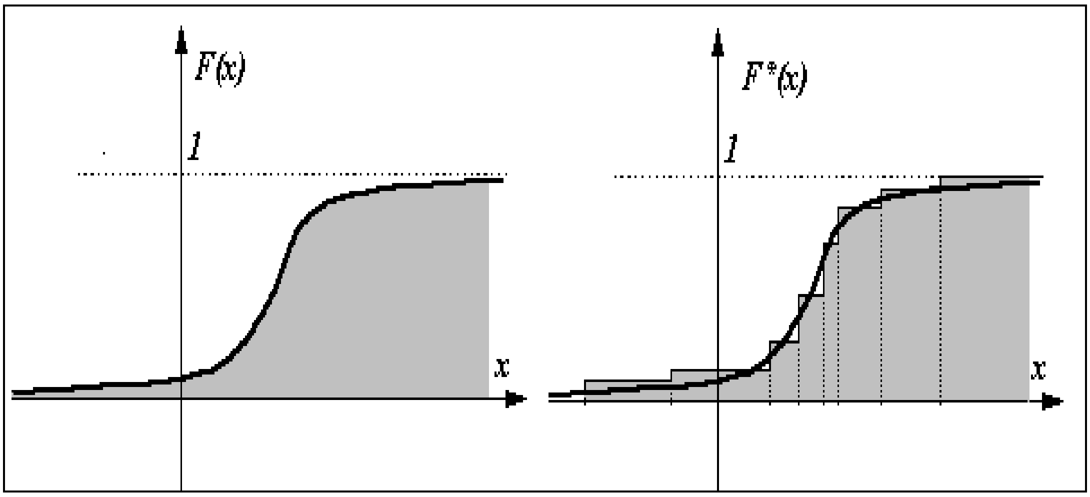


Рисунок 8 – Графики теоретической и статистической функций распределения

Следовательно, $F^*(x)$ является статистической аппроксимацией функции распределения, ее приближенным значением и обладает следующими свойствами:

1) значения функции $F^*(x)$ принадлежат отрезку [0;1];

2) $F^*(x)$ – неубывающая функция;

3) если $x_{\max}$, $x_{\min}$ - наибольший и наименьший элементы выборки, то

$$F^*(x) = \begin{cases} 0, x < x_{\min} \\ 1, x > x_{\max} \end{cases};$$

4) $F^*(x)$ – непрерывна слева.

Пример 5. Ниже приводятся результаты наблюдений за количеством обрывов нити в определенные промежутки времени:

0, 2, 1, 0, 3, 1, 1, 0, 1, 2, 1, 0, 1, 0, 0, 2, 3, 1, 0, 1.

Построить график эмпирической функции распределения.

Решение. Эмпирическая функция распределения определена на всей числовой оси. Числовую ось разобьем не несколько интервалов в которых функция сохраняет постоянное значение.

Пусть $x \in (-\infty;0]$. По определению: $F^* = \frac{n_x}{n} = \frac{0}{20} = 0$,

где n_x - число вариант, меньших любого x из данного интервала.

Если $x \in (0;1]$, то $F^* = \frac{7}{20}$;

если $x \in (1;2]$, то $F^* = \frac{15}{20}$;

если $x \in (2;3]$, то $F^* = \frac{18}{20}$;

если $x \in (3;5]$, то $F^* = \frac{20}{20} = 1$.

Итак, имеем :

$$F^*(x) = \begin{cases} 0, & -\infty < x \le 0 \\ 7/20, & 0 < x \le 1 \\ 15/20, & 1 < x \le 2 \\ 18/20, & 2 < x \le 3 \\ 1, & x > 3 \end{cases}$$

График эмпирической функции распределения изображен на рисунке 10.

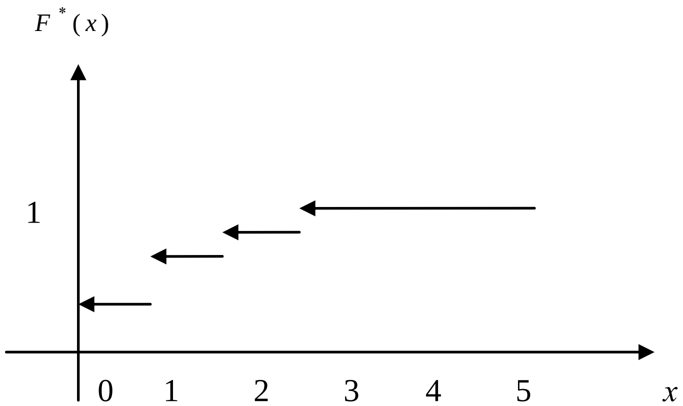


Рисунок 9 - График эмпирической функции распределения (Пример 5)

Пример 6. Из генеральной совокупности извлечена выборка:

-0,269;	-0,786;	0,585;	-1,107;	1,574;	0,341;
-1,309;	-0,165;	-0,483;	0,525;	1,620;	0,206;
0,346;	-0,973;	-0,363;	0,660;	1,084;	0,903;
1,387;	1,261;	0,786;	1,107;	0,341;	0,525.

Построить сгруппированный статистический ряд, гистограмму частот и эмпирическую функцию распределения.

Решение. Занесем данные в таблицу по возрастанию значений

№	x_i	По возрастанию	№	x_i	По возрастанию
1	-0,269	-1,309	11	1,620	0,525
2	0,786	-1,107	12	0,206	0,585
3	0,585	-0,973	13	0,346	0,660
4	-1,107	-0,483	14	-0,973	0,786
5	1,574	-0,363	15	-0,363	0,903
6	0,341	-0,269	16	0,660	1,084
7	-1,309	-0,165	17	1,084	1,261
8	-0,165	0,206	18	0,903	1,387
9	-0,483	0,341	19	1,387	1,574
10	0,525	0,346	20	1,261	1,620

Количество интервалов рассчитаем по формуле Стерджесса, получим $k = 1 + 3{,}3 \cdot 1{,}301 = 5{,}293$, округляем в большую сторону, $k = 6$. Наименьшее значение равно –1,309, наибольшее значение равно 1,620. Длина интервала находится по формуле $h = \frac{x_{max} - x_{min}}{k}$. «Упакуем» выборку в интервале [-1,311;1,623], который разобьем на 6 интервалов длиной $h = \frac{1{,}623 - (-1{,}31)}{6} = 0{,}489$. Подсчитаем частоту n_i для каждого интервала и получим сгруппированный статистический ряд.

Интервалы	Частоты, n_i	Высоты, $h_i = n_i/h$	Накопленные частоты, n_i *	Относительные накопленные частоты, n_i * $/n$
[-1,311; -0,822)	3	0,307	3	0,15
[-0,822; -0,333)	2	0,204	5	0,25
[-0,333; 0,156)	2	0,204	7	0,35
[0,156; 0,645)	4	0,409	11	0,55
[0,645; 1,134)	5	0,511	16	0,80
[1,134; 1,623)	4	0,409	20	1,00

Тогда гистограмма частот будет иметь вид:

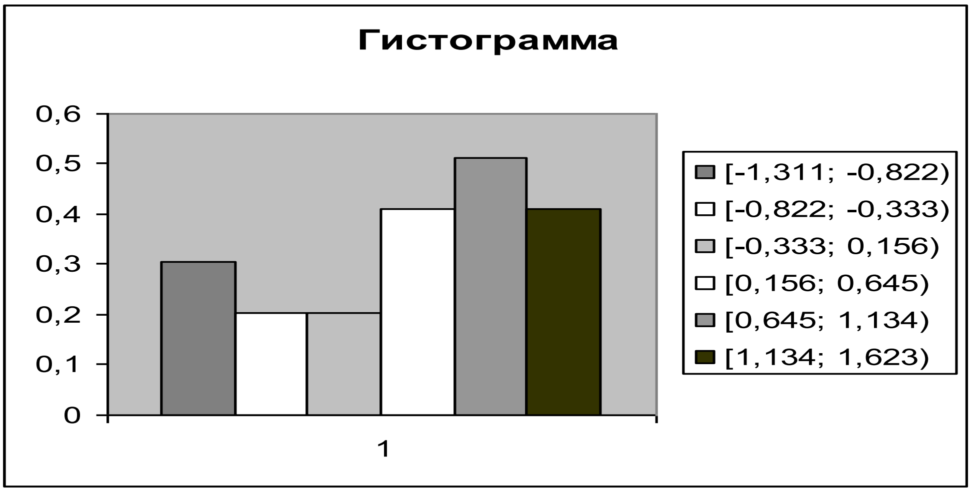


Рисунок 10 – Гистограмма частот (*Пример* 6)

Эмпирическая функция распределения определится:

$$F^*(x)=\begin{cases} 0 & \text{при} \ , x \le -1{,}311; \\ 0{,}15 & \text{при} \ -1{,}311 < x \le -0{,}822; \\ 0{,}25 & \text{при} \ -0{,}822 < x \le -0{,}333; \\ 0{,}35 & \text{при} \ -0{,}333 < x \le 0{,}156; \\ 0{,}55 & \text{при} \ 0{,}156 < x \le 0{,}645; \\ 0{,}80 & \text{при} \ 0{,}645 < x \le 1{,}134; \\ 1 & \text{при} \ 1{,}134 < x \le 1{,}623. \end{cases}$$

График эмпирической функции распределения представлен на рисунке 12.

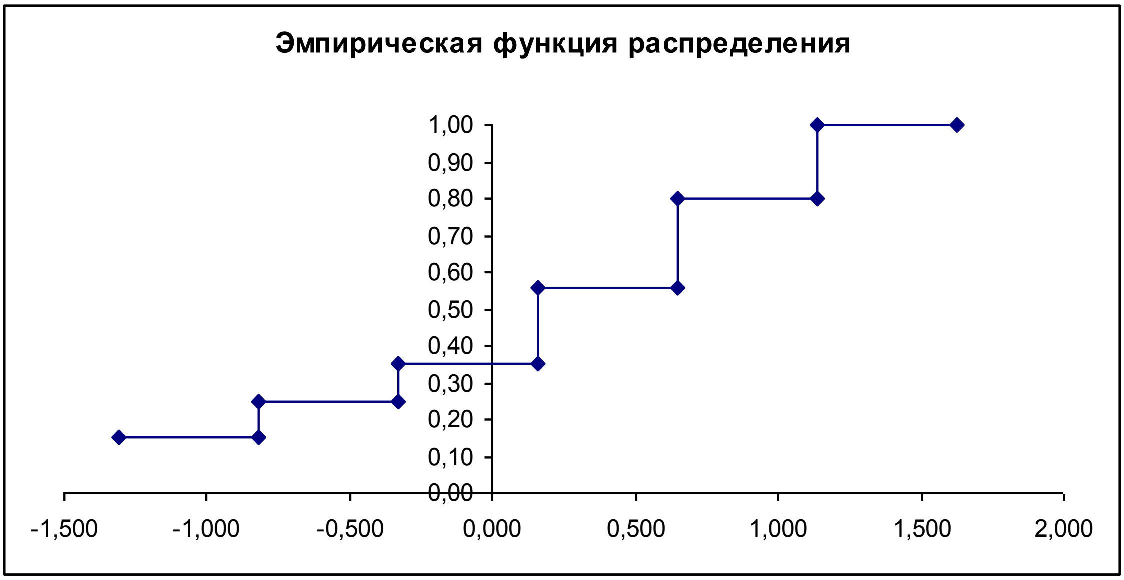


Рисунок 11 – График эмпирической функции распределения (*Пример* 6)

Задачи для самостоятельного решения

1. В результате опроса студентов одного потока, их возраст представился следующими данными: 17, 20, 18, 19, 18, 17, 20, 21, 24, 22, 20, 21, 20, 19, 18, 20, 21, 22, 25, 20. Составить вариационный ряд, построить полигон частот.

2. По данному распределению выборки найти эмпирическую функцию распределения и построить ее график

x	2	5	7	8
n	1	3	2	4

3. В результате проверки партии деталей получены результаты по сор-

там: 1, 2, 1, 1, 4, 2, 1, 1, 1, 2, 1, 2, 1, 1, 1, 3, 4, 1, 1, 2, 2, 3, 1, 1. Составить распределение частостей, построить полигон частостей, найти $F^*(x)$ и построить её график.

4. Составить дискретный вариационный ряд для случайной величины - длины заготовок, отобранных случайным образом: 39, 41, 40, 40, 43, 41, 44, 42, 41, 41, 43, 42, 39, 40, 42, 43, 41, 42, 41, 39, 42, 42, 41, 42, 40, 41, 43, 41, 39, 40. Построить полигон частот и кумуляту.

5. Построить интервальный вариационный ряд и начертить гистограмму частот распределения 56 абитуриентов по числу баллов, полученных ими на приемных экзаменах: 20, 29, 42, 61, 44, 28, 30, 57, 10, 60, 50, 30, 61, 29, 30, 21, 28, 10, 44, 21, 10, 57, 42, 10, 80, 42, 21, 21, 57, 29, 10, 10, 21, 28, 42, 21, 50, 42, 10, 10, 42, 21, 29, 42, 21, 29, 10, 30, 42, 70, 21, 21, 44, 21, 50, 28.

6. Данные об урожайности ржи на полях приведены в следующей таблице:

Урожайность в ц/га	12-14	14-16	16-18	18-20	20-22	22-24
Доля участка от общей площади, %	15	20	30	15	10	10

Построить гистограмму частот и график эмпирической функции распределения.

7. Для определения прочности нити проведено 50 испытаний, давших следующие результаты:

Прочность нити, в г	180-200	200-220	220-240	240-260	260-280
Количество испытаний	2	11	23	13	1

Построить гистограмму относительных частот и график эмпирической функции распределения. Перейти от интервального вариационного ряда к точечному, построить полигон относительных частот и кумуляту.

8. Имеются следующие данные о жирности молока у коров фермы

Жирность, в %	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2
Число коров	8	11	28	43	71	52	29	8

Преобразовать данный вариационный ряд в интервальный с равными интервалами так, чтобы приведенные значения признака были центрами интервалов и построить гистограмму относительных частот.

9. Через каждый час измерялось напряжение тока в электросети. При этом были получены следующие значения (в вольтах): 227, 219, 215, 230, 232, 223, 220, 226, 209, 211, 215, 218, 220, 216, 219, 221, 218, 220, 222, 220, 217, 219, 229, 227, 212, 226, 217, 221, 225, 230. Найти интервальный статистический ряд с равными интервалами. Построить гистограммы частот и относительных частот.

10. Длина волокон хлопка (в мм) представлена некоторой выборкой:

34,0	26,0	14,1	35,0	19,6	25,9	30,8	25,9	26,4	37,1
28,1	23,5	39,0	28,7	31,2	20,5	26,0	18,5	25,3	31,5
27,2	28,5	29,5	34,4	15,2	22,1	30,3	30,2	28,3	25,4
22,5	26,5	17,2	34,3	16,0	34,2	30,6	34,1	24,7	26,2
31,9	34,0	27,0	28,0	12,8	30,3	31,3	22,3	30,6	19,8
28,2	25,7	22,7	31,8	36,8	28,4	16,1	33,0	31,1	23,2
26,1	26,5	23,5	25,2	32,0	27,8	13,2	28,5	31,2	24,8
30,5	19,1	19,0	33,5	28,0	23,0	16,8	26,1	30,2	34,0
20,4	21,1	29,1	29,0	24,0	37,0	17,1	27,5	20,7	20,5
30,0	36,1	30,9	24,6	21,5	24,5	28,7	28,5	32,2	28,4

Построить интервальный вариационный ряд, представить его графически гистограммой частот. Найти эмпирическую функцию распределения и построить ее график.

1.4 Числовые характеристики статистических рядов

С целью обеспечения обработки частотных распределений и представления информации, заключенной в статистических данных, некоторыми количественными показателями вариационные ряды описывают с помощью определенных числовых характеристик. Такими характеристиками для одномерных статистических рядов являются *характеристики* (*меры*) *положения*, *характеристики рассеяния и характеристики формы.*

1.4.1 Характеристики или меры положения

Для *описания характера расположения распределений* применяют три группы мер: выборочная средняя (арифметическая, геометрическая, степенная и гармоническая), медиана и мода.

Выборочной средней $\overline{x}_в$ называют среднее арифметическое значение признака выборочной совокупности. Формула (1.13) определяет *невзвешенную* или *простую выборочную среднюю*, если все значения вариант x_i объема n различны (для *несгруппированных данных*), а (1.13') - *взвешенную выборочную среднюю*, если варианты $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$, причем $\sum_{i=1}^{k} n_i = n$ (для *сгруппированных данных*).

$$\overline{x}_в = \frac{\sum_{i=1}^{n} x_i}{n}. \quad (1.13) \qquad \overline{x}_в = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}. \quad (1.13')$$

Выборочная средняя представляет собой значение, относительно которого может быть «сбалансировано» все эмпирическое распределение, фактически она является абсциссой центра масс гистограммы частот.

Пример 7. Рассчитать средний возраст студентов в группе из 20 человек: 18, 18, 19, 20, 19, 20, 19, 19, 19, 20, 22, 19, 19, 20, 20, 21, 19, 19, 19, 19.

Решение: Применим формулу (1.13)

$$\overline{x}_в = \frac{18+18+19+\ldots+21+19+19+19+19}{20} = 19{,}4.$$

Если сгруппировать данные, то получим ряд распределения:

x_i	18	19	20	21	22
n_i	2	11	5	1	1

Согласно формуле (1.13') средний возраст студентов определится:

$$\overline{x}_в = \frac{18\cdot 2+19\cdot 11+20\cdot 5+21\cdot 1+22\cdot 1}{20} = 19{,}4 \text{ года.}$$

Выборочная средняя имеет те же единицы измерения, что и варианты, и *обладает свойствами*, более полно раскрывающими ее сущность:

1) выборочная средняя постоянной величины равна самой постоянной $\overline{C} = C$;

2) произведение выборочной средней на сумму частот равно сумме произведений отдельных вариант на соответствующие им частоты

$$\overline{x}_в \cdot \sum_{i=1}^{k} n_i = \sum_{i=1}^{k} x_i \cdot n_i ;$$

3) сумма взвешенных отклонений значений признака от выборочной средней равна нулю

$$\sum_{i=1}^{k} (x_i - \overline{x}_в) \cdot n_i = 0 ;$$

4) если все значения признака увеличить (уменьшить) на одно и то же постоянное число *C*, то и выборочная средняя увеличится (уменьшится) на это число *C*

$$\frac{\sum_{i=1}^{k} (x_i \pm C) \cdot n_i}{\sum_{i=1}^{k} n_i} = \overline{x}_в \pm C$$

5) если все значения признака увеличить (уменьшить) в одно и то же постоянное число *A*, то и выборочная средняя увеличится (уменьшится) в это же число *A*

$$\frac{\sum_{i=1}^{k} Ax_i \cdot n_i}{\sum_{i=1}^{k} n_i} = A \cdot \overline{x}_в ;$$

6) если все частоты вариант умножить (разделить) на одно и то же постоянное число *A*, то выборочная средняя не изменится

$$\frac{\sum\limits_{i=1}^{k} x_i (n_i \cdot A)}{\sum\limits_{i=1}^{k} n_i} = \overline{x}_в;$$

7) выборочная средняя алгебраических сумм, соответствующих друг другу значений, принадлежащим двум группам наблюдений, равна алгебраической сумме выборочных средних арифметических этих групп

$$\overline{(x \pm y)}_в = \overline{x}_в \pm \overline{y}_в;$$

8) если ряд наблюдений состоит из двух непересекающихся групп наблюдений, то выборочная средняя всего ряда равна выборочной средней взвешенной групповых средних $\overline{x}_в, \overline{y}_в$, причем весами являются объемы групп $\sum\limits_{i=1}^{k} n_i = n_1, \sum\limits_{j=1}^{p} m_j = n_2$

$$\overline{(x \pm y)}_в = \frac{\sum\limits_{i=1}^{k} x_i \cdot n_i + \sum\limits_{j=1}^{p} y_j \cdot m_j}{n_1 + n_2}.$$

Выборочная средняя интервального вариационного ряда вычисляется по формуле (1.13'), где в качестве вариант x_i принимаются середины соответствующих интервалов.

Средняя геометрическая (*невзвешенная* и *взвешенная*) применяется при оценке темпов изменения величин (в частности, при расчете индексов цен) и находится по формулам (1.14) и (1.14'):

$$\overline{x}_{геом} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdot \ldots \cdot x_n}. \quad (1.14) \qquad \overline{x}_{геом} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \cdot \ldots \cdot x_k^{n_k}}. \quad (1.14')$$

В практических расчетах удобнее использовать *десятичный* (*натуральный*) *логарифм средней геометрической*

$$lg\,\overline{x}_{геом} = \frac{1}{n}\sum_{i=1}^{n} lg\,x_i . \quad (15) \qquad lg\,\overline{x}_{геом} = \frac{1}{n}\sum_{i=1}^{n} (lg\,x_i)\cdot n_i . \quad (1.15')$$

Средняя гармоническая (*невзвешенная и взвешенная*) определяется по формулам:

$$\overline{x}_{гарм.} = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}\frac{1}{x_i}} . \quad (1.16) \qquad \overline{x}_{гарм.} = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}\frac{n_i}{x_i}} . \quad (1.16')$$

Область применения гармонической средней ограничена. В экономике гармонические средние используют при анализе средних норм времени, а также в тех случаях, когда суммируемый признак выражен обратной величиной исследуемого признака.

Между выборочной средней, средней геометрической и средней гармонической существует соотношение

$$\overline{x}_{гарм.} \le \overline{x}_{геом.} \le \overline{x}_{в} . \quad (1.17)$$

Средняя степенная (*невзвешенная и взвешенная*) определяется по формулам:

$$\overline{x^P} = \frac{\sum\limits_{i=1}^{n} x_i^p}{n} . \quad (1.18) \qquad \overline{x^P} = \frac{\sum\limits_{i=1}^{k} x_i^p \cdot n_i}{\sum\limits_{i=1}^{k} n_i} . \quad (1.18')$$

При $p = 2$ имеем *среднюю квадратичную,* при $p = 3$ - *среднюю кубическую* и т.д.

Пример 8. Известны данные по валовому сбору и урожайности некоторой сельскохозяйственной культуры по районам области:

Район	Валовый сбор, тыс. т.	Урожайность ц/га
А	52	10
Б	40	14
В	31	15
Г	67	8

Найти среднюю урожайность по указанным районам.

Решение. Средняя урожайность может быть определена на основе следующего соотношения:

$$\frac{\text{общий валовый сбор}}{\text{общая посевная площадь}}.$$

Общий валовый сбор (тыс. ц.) мы получим простым суммированием валового сбора по районам: 520 + 400 + 310 + 670 = 1900. Данные же о посевной площади отсутствуют, но их можно получить, разделив валовый сбор каждого района на урожайность. С учетом этого найдем среднюю урожайность по формуле взвешенной средней гармонической (1.16'), переведя для сопоставимости тонны в центнеры:

$$\bar{x}_{гарм.} = \frac{520 + 400 + 310 + 670}{\frac{520}{10} + \frac{400}{14} + \frac{310}{15} + \frac{670}{8}} = 10{,}3 \text{ (ц/га).}$$

Таким образом, общая посевная площадь данной культуры в целом по области составляет 185,2 тыс. га, а средняя урожайность – 10, 3 ц с одного гектара.

Наряду с приведенными средними величинами в качестве характеристик положения используют *структурные средние*: моду и медиану. *Мода* (модальное значение $x_{мод} = M_o$) - наиболее часто встречающееся в статистической совокупности значение признака. Для дискретного вариационного ряда мода определяется по частотам вариант и соответствует варианте с максимальной частотой. При определении моды обычно применяют следующие соглашения:

1) если все значения вариационного ряда имеют одинаковую частоту, то говорят, что этот вариационный ряд не имеет моды;

2) если две соседних варианты имеют одинаковую доминирующую частоту, что мода вычисляется как среднее арифметическое этих вариант;

3) если две не соседних варианты имеют одинаковую доминирующую частоту, то такой вариационный ряд называется *бимодальным*.

4) если таких вариант более двух, то ряд - *полимодальный*.

В случае интервального вариационного ряда с равными интервалами *модальный интервал* (интервал, содержащий моду) определяется по наибольшей частоте, а при неравных интервалах - по наибольшей плотности. Вычисление моды производится по следующей формуле:

$$M_0 = x_0 + h\frac{n_{M_0} - n_{M_0-1}}{(n_{M_0} - n_{M_0-1}) + (n_{M_0} - n_{M_0+1})}, \qquad (1.19)$$

где x_o- левый конец модального интервала; h- ширина модального интервала; n_{Mo}- частота модального интервала; $n_{Mo\text{-}1}$ - частота интервала, предшествующего модальному; n_{Mo+1} - частота интервала, следующего за модальным.

При вычислении моды в формуле (1.19) можно использовать не только частоты, но и относительные частоты. Для *графического определения моды* используют 3 соседних столбца гистограммы (самый высокий и 2 прилегающих к нему). Затем правую вершину модального прямоугольника соединяют с правым верхним углом предыдущего прямоугольника. А левую вершину модального прямоугольника – с левым верхним углом последующего прямоугольника. Далее из точки их пересечения опускают перпендикуляр на ось абсцисс (рис. 13).

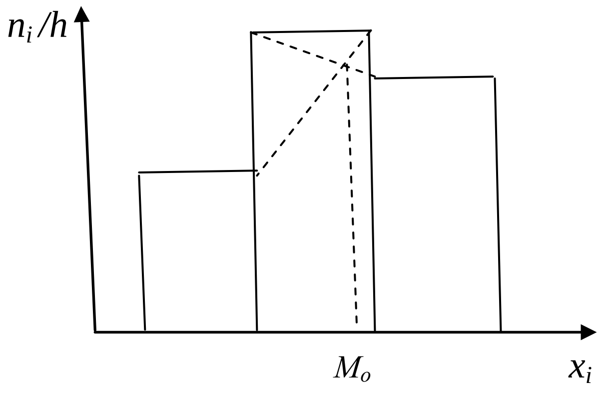


Рисунок 12 – Графическое определение моды

Медиана (медианное значение $x_{мед} = M_e$) - значение изучаемого признака, приходящееся на середину ранжированной совокупности. Для дискретного вариационного ряда медиану определяют по формуле:

$$M_e = \begin{cases} \dfrac{x_l + x_{l+1}}{2}, n = 2l; \\ x_{l+1}, n = 2l + 1. \end{cases} \qquad (1.20)$$

При вычислении медианы интервального вариационного ряда, сначала находят *медианный интервал* $[x_{Me}; x_{Me} + h]$ (интервал, содержащий медиану), где h - длина медианного интервала, путем использования накопленных частот. *Медианному интервалу* соответствует тот, в котором содержится накопленная частота, *превышающая половину объема выборки*. Расчет медианы при постоянной плотности внутри интервала производится по формуле:

$$M_e = x_o + h \cdot \frac{\frac{n}{2} - n^*_{M_e - 1}}{n_{M_e}}, \qquad (1.21)$$

где x_o - левый конец медианного интервала; h - ширина медианного интервала; n^*_{Me-1} – накопленная частота интервала, предшествующего медианному; n_{Me} – частота медианного интервала.

Медиана может быть определена *графически* по *кумуляте*. Для этого последнюю ординату, равную сумме всех частот, т.е. объему выборки n, делят пополам. Из полученной точки восстанавливают перпендикуляр до пересечения с кумулятой. Абсцисса точки пересечения и дает значение медианы (рис.14).

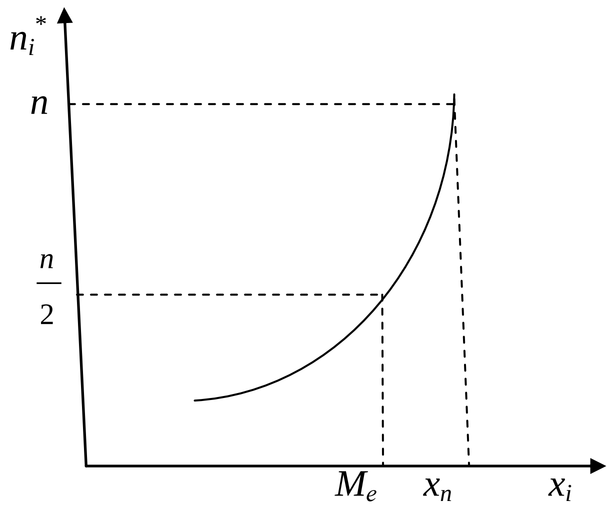


Рисунок 13 – Графическое определение медианы

Замечание. Известно, что множество результатов наблюдений имеет свойство группироваться относительно некоторого центра. Центр может быть описан мерами центральной тенденции: модой, медианой, средней арифметической. При выборе меры центральной тенденции следует иметь в виду следующее:

1) в малой группе наблюдений мода не стабильна, при небольших изменениях она может сильно измениться;

2) на медиану не влияет величина «больших» и «малых» значений;

3) медиана – лучшая характеристика центральной тенденции, когда гистограмма исходных данных *унимодальна* (имеет одну моду);

4) на величину средней арифметической влияет каждое значение;

5) некоторые множества не имеют центральной тенденции.

Пример 9. На основе интервального вариационного ряда распределения

Затраты времени (мин.) клиентов на оформление потребительского кредита в крупных салонах электроники города	10,8-14,6	14,6-18,4	18,4-22,2	22,2-26,0	26,0-29,8
Количество клиентов в отдельной группе	6	4	5	7	12

определить: 1) среднее значение затрат времени на оформление потребительского кредита в крупных салонах электроники города;

2) структурные средние для интервального и соответствующего дискретного вариационных рядов.

Решение. 1) Для определения среднего значения затрат времени на оформление потребительского кредита для клиента крупного салона электроники города необходимо перейти от интервального вариационного ряда к дискретному ряду, определив середины интервалов.

Затраты времени (мин.) клиентов на оформление потребительского кредита в крупных салонах электроники города	12,7	16,5	20,3	24,1	27,9
Количество клиентов в	6	4	5	7	12

отдельной группе					

Применим формулу (1.13′):

$$\bar{x}_в = \frac{12,7 \cdot 6 + 16,5 \cdot 4 + 20,3 \cdot 5 + 24,1 \cdot 7 + 27,9 \cdot 12}{6 + 4 + 5 + 7 + 12} = \frac{747,2}{34} = 21,976 \approx 22.$$

Таким образом, среднее значение затрат времени на оформление потребительского кредита для клиента крупного салона электроники города составило около 22 минут.

2) Определим структурные средние: моду и медиану по формулам (1.19) и (1.21). Модальным интервалом является интервал (26,0; 29,8) с наибольшей частотой 12.

$$M_o = 26,0 + 3,8 \cdot \frac{12 - 7}{(12 - 7) + (12 - 0)} = 27,12.$$

Чаще всего затраты времени на оформление потребительского кредита составляют в количестве 27,12 минут. Для соответствующего дискретного ряда M_o = 27,9 мин.

Медианным интервалом является интервал (22,2; 26,0), так как его накопленная частота равна 6 + 5 + 4 + 7 = 22, что превосходит половину объема выборки 17.

$$M_e = 22,2 + 3,8 \cdot \frac{\frac{34}{2} - 15}{7} = 23,29 \text{ мин.}$$

У половины клиентов на оформление потребительского кредита уходит 23,29 минуты. Для дискретного вариационного ряда согласно формуле (1.20) M_e = 20,3.

1.4.2 Характеристики или меры рассеяния

Средние величины, характеризующие вариационный ряд одним числом, не учитывают вариацию признака. Для измерения вариации применяются следующие характеристики: вариационный размах, среднее линейное отклонение, выборочная дисперсия, выборочное среднее квадратическое отклонение, коэффициент вариации и др.

Вариационный размах (*размах вариации*) R представляет собой разность между наибольшим и наименьшим значением наблюдений:

$$R = x_{max} - x_{min}. \tag{1.22}$$

Вариационный размах вариации применяется в качестве приблизительной оценки вариации значений признака, широко используется в ряде отраслей промышленности при статистическом изучении качества продукции.

Среднее линейное отклонение (*невзвешенное* и *взвешенное*) определяется по формулам:

$$d = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x_в} \right|}{n}. \quad (1.23) \qquad d = \frac{\sum_{i=1}^{k} \left| x_i - \overline{x_в} \right| \cdot n_i}{\sum_{i=1}^{k} n_i}. \quad (1.23')$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\overline{x}_в$. Формулы (24) и (24') служат для определения выборочной дисперсии для негруппированных и сгруппированных данных:

$$D_в = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x_в} \right)^2}{n}. \quad (1.24) \qquad D_в = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_в} \right)^2 \cdot n_i}{\sum_{i=1}^{k} n_i} \quad (1.24')$$

На практике для вычисления $D_в$ удобнее использовать формулу:

$$D_в = \overline{x_в^2} - (\overline{x}_в)^2. \tag{1.24''}$$

Выборочное среднее квадратическое отклонение $\sigma_в$ представляет собой квадратный корень из выборочной дисперсии:

$$\sigma_в = \sqrt{D_в}. \tag{1.25}$$

Выборочное среднее квадратическое отклонение выражается в тех же единицах, что и значения исследуемого признака.

Выборочная дисперсия и среднее квадратическое отклонение обладают следующими свойствами:

1) выборочная дисперсия постоянной величины C равна нулю

$$D_в(C) = 0;$$

2) если все результаты наблюдений увеличить (уменьшить) на одно и то же постоянное число C, то дисперсия и среднее квадратическое отклонение не изменятся

$$D_в(X \pm C) = D_в(X), \qquad \sigma_в(X \pm C) = \sigma_в(X);$$

3) если все результаты наблюдений умножить на одно и то же постоянное число, то имеют место следующие равенства:

$$D_в(CX) = C^2 D_в(X), \qquad \sigma_в(CX) = |C|\sigma_в(X);$$

4) если все частоты вариант умножить на одно и тоже постоянное число, то выборочная дисперсия и среднее квадратическое отклонение не изменятся

$$\frac{\sum_{i=1}^{k}\left(x_i - \overline{x_в}\right)^2 \cdot Cn_i}{\sum_{i=1}^{k} Cn_i} = D_в(X) = \sigma_в^2(X);$$

5) выборочная дисперсия принимает минимальное значение, когда выборочная средняя принимает постоянное значение (*свойство минимальности дисперсии*)

$$D_в(X) = \frac{\sum_{i=1}^{k}\left(x_i - C\right)^2 \cdot n_i}{n} \to \min \qquad \text{при } \overline{x_в} = C;$$

6) выборочное среднее квадратическое отклонение равно среднему квадрату отклонений значений исследуемого признака от произвольной по-

стоянной a минус квадрат разности между выборочной средней арифметической и этой произвольной постоянной

$$\sigma_в^{\ 2}(X) = \sigma_a^{\ 2} - \left(\overline{x_в} - a\right)^2,$$

где $$\sigma_a^2 = \frac{\sum_{i=1}^{k}\left(x_i - a\right)^2 \cdot n_i}{n};$$

7) если объединяются несколько распределений в одно, то общая дисперсия $\sigma_о^{\ 2}$ нового распределения равна сумме внутригрупповой и межгрупповой дисперсий (*правило сложения дисперсий*)

$$\sigma_о^2 = \overline{\sigma_в^2} + \delta_о^2, \tag{1.26}$$

или $$\sigma_о^2 = \frac{\sum_{i,j=1}^{k.m}\left(x_{ij} - \overline{x_о}\right)^2 \cdot n_{ij}}{N} = \frac{\sum_{j=1}^{m} x_j^2 \cdot N_j}{N} - \left(\overline{x_о}\right)^2, \tag{1.26'}$$

где n_{ij}- частота j-го варианта i-го частного распределения (j = 1,2, …m; i = 1,2, … , k);

x_{ij}- j-й вариант i-го частного распределения;

n_i – объем i-го частного распределения;

$N_j = \sum_{i=1}^{k} n_{ij}$ - частота j-го варианта нового распределения;

$N = \sum_{i=1}^{k} n_i = \sum_{j=1}^{m} N_j$ - объем нового распределения;

$\overline{x}_i = \frac{\sum_{j=1}^{m} x_i \cdot n_{ij}}{n_i}$ - средняя арифметическая i-го частного распределения;

$\overline{x}_о = \frac{\sum_{j=1}^{m} x_j \cdot N_i}{N}$ - средняя арифметическая нового распределения;

$$\overline{\sigma}_i^2 = \frac{\sum\limits_{j=1}^{m} x_{ij}^2 \cdot n_{ij}}{n_i} - \left(\overline{x}_i\right)^2$$ - дисперсия i-го частного распределения;

$$\overline{\sigma_в^2} = \frac{\sum\limits_{i=1}^{k} \sigma_i^2 \cdot n_i}{N}$$ - внутригрупповая дисперсия;

$$\delta_o^2 = \frac{\sum\limits_{i=1}^{k} \left(\overline{x}_i - \overline{x}_o\right)^2 \cdot n_i}{N}$$ - межгрупповая дисперсия.

Общая дисперсия σ_o^2 измеряет вариацию признака по всей совокупности под влиянием всех факторов, обусловивших эту вариацию. Внутригрупповая дисперсия $\overline{\sigma_в^2}$ отражает часть вариации, происходящей под влиянием неучтенных факторов, и не зависит от признака-фактора, положенного в основу группировки. Межгрупповая дисперсия δ_o^2 характеризует систематическую вариацию, т.е. различия в величине изучаемого признака, возникающие под влиянием признака-фактора, положенного в основу группировки. Правило сложения дисперсий используется при расчете ошибок выборочного наблюдения и при измерении тесноты связи между признаками.

*Пример*10. По данным распределения студентов по результатам сдачи экзаменов

Оценка на экзамене	Число студентов, получивших оценку по предметам			
	Философия, x_1	Физика, x_2	Экономика, x_3	Математика, x_4
Неудовлетворительно,2	2	1	4	3
Удовлетворительно, 3	6	10	8	8
Хорошо, 4	10	8	9	9
Отлично, 5	7	6	4	5

определить: а) средний балл успеваемости студентов по каждому предмету и по всем предметам;

б) дисперсии балла успеваемости по каждому предмету и в целом по всем предметам;

в) внутригрупповую и межгрупповую дисперсии;

г) общую дисперсию успеваемости, используя правило сложения дисперсий.

Решение. а) Подсчитаем количество студентов, сдававших соответственно философию, физику, экономику и математику:

$$2 + 6 + 10 + 7 = 25 = N_1; \qquad 1 + 10 + 8 + 6 = 25 = N_2;$$

$$4 + 8 + 9 + 4 = 25 = N_3; \qquad 3 + 8 + 9 + 5 = 25 = N_4.$$

Тогда средний балл по этим предметам определится:

$$\bar{x}_1 = \frac{2 \cdot 2 + 3 \cdot 6 + 4 \cdot 10 + 5 \cdot 7}{25} = 3{,}88; \qquad \bar{x}_2 = \frac{2 \cdot 1 + 3 \cdot 10 + 4 \cdot 8 + 5 \cdot 6}{25} = 3{,}76;$$

$$\bar{x}_3 = \frac{2 \cdot 4 + 3 \cdot 8 + 4 \cdot 9 + 5 \cdot 4}{25} = 3{,}52; \qquad \bar{x}_4 = \frac{2 \cdot 3 + 3 \cdot 8 + 4 \cdot 9 + 5 \cdot 5}{25} = 3{,}64.$$

Подсчитаем количество студентов, получивших соответственно«2», «3», «4» и «5»:

$$2 + 1 + 4 + 3 = 10 = n_1; \qquad 6 + 10 + 8 + 8 = 32 = n_2;$$

$$10 + 8 + 9 + 9 = 36 = n_3; \qquad 7 + 6 + 4 + 5 = 22 = n_4.$$

Средний балл по всем предметам составит:

$$\bar{x}_o = \frac{2 \cdot 10 + 3 \cdot 32 + 4 \cdot 36 + 5 \cdot 22}{100} = 3{,}7.$$

б) Дисперсии балла успеваемости по каждому предмету определятся:

$$\bar{\sigma}_1^2 = \frac{2^2 \cdot 2 + 3^2 \cdot 6 + 4^2 \cdot 10 + 5^2 \cdot 7}{25} - (3{,}88)^2 = \frac{397}{25} - 15{,}05 = 15{,}88 - 15{,}05 = 0{,}83;$$

$$\bar{\sigma}_2^2 = \frac{2^2 \cdot 1 + 3^2 \cdot 10 + 4^2 \cdot 8 + 5^2 \cdot 6}{25} - (3{,}76)^2 = \frac{372}{25} - 14{,}14 = 14{,}88 - 14{,}14 = 0{,}74;$$

$$\bar{\sigma}_3^2 = \frac{2^2 \cdot 4 + 3^2 \cdot 8 + 4^2 \cdot 9 + 5^2 \cdot 4}{25} - (3{,}52)^2 = \frac{332}{25} - 12{,}39 = 13{,}28 - 12{,}39 = 0{,}89;$$

$$\bar{\sigma}_4^2 = \frac{2^2 \cdot 3 + 3^2 \cdot 8 + 4^2 \cdot 9 + 5^2 \cdot 5}{25} - (3{,}64)^2 = \frac{353}{25} - 13{,}25 = 14{,}12 - 13{,}25 = 0{,}87.$$

В целом по всем предметам дисперсия балла успеваемости составит:

$$\sigma_o^2 = \frac{2^2 \cdot 10 + 3^2 \cdot 32 + 4^2 \cdot 36 + 5^2 \cdot 22}{100} - (3,7)^2 = \frac{1454}{100} - 13,69 = 14,54 - 13,69 = 0,85 .$$

в) Определим внутригрупповую и межгрупповую дисперсии:

$$\overline{\sigma_в^2} = \frac{0,83 \cdot 10 + 0,74 \cdot 32 + 0,89 \cdot 36 + 0,87 \cdot 22}{100} = \frac{83,16}{100} = 0,83 \cdot$$

$$\delta_o^2 = \frac{(3,88 - 3,7)^2 \cdot 10 + (3,76 - 3,7)^2 \cdot 32 + (3,52 - 3,7)^2 \cdot 36 + (3,64 - 3,7)^2 \cdot 22}{100} = \frac{2,009}{100} = 0,2 \cdot$$

г) По правилу сложения дисперсий общая дисперсия определится:

$$\sigma_o^2 = \overline{\sigma_в^2} + \delta_o^2 = 0,83 + 0,2 = 0,85 .$$

Результат совпал со значением общей дисперсии, вычисленной в б).

Кроме абсолютных величин для характеристики вариации статистической совокупности значений признаков применяются относительные показатели: линейный коэффициент вариации, коэффициент вариации, эмпирический коэффициент детерминации, эмпирическое корреляционное отношение.

Линейным коэффициентом вариации V_d выраженное в процентах отношение среднего линейного отклонения d к к выборочной средней $\overline{x}_в$:

$$V_d = \frac{d}{\overline{x}_в} \cdot 100\ \% \quad . \qquad (1.27)$$

Коэффициентом вариации V называют выраженное в процентах отношение выборочного среднего квадратического отклонения $\sigma_в$ к выборочной средней $\overline{x}_в$:

$$V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\ \% . \qquad (1.28)$$

Коэффициенты вариации используются для сравнения размеров вариации в вариационных рядах с различными средними, а также для сравнения вариаций разных показателей в одной и той же совокупности. В коэффициентах вариации устраняются не только несопоставимость, связанная с различными измерениями изучаемого признака, но и несопоставимость, возникающая вследствие различий в величине средних арифметических. Кроме того, показатели вариации дают характеристику однородности статистических данных. Статистическая совокупность считается однородной, если коэффициент вариации не превышает 33%. Состояние между вариацией выборки и ее центром можно также оценить по значению коэффициента вариации:

1) $V \le 30\,\%$ - выборка имеем довольно большую степень концентрации относительного среднего;

2) $30\,\% \le V \le 100\,\%$ - степень концентрации допустимая;

3) $V \ge 100\,\%$ - делается вывод о неоднородности выборки.

Коэффициенты вариации служат для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней $\bar{x}_в$, у которого коэффициент вариации больше.

Каждому закону распределения соответствует определенный приближенный диапазон значений коэффициента вариации, поэтому он также может служить в качестве критерия при выборе закона распределения. Для нормального закона распределения коэффициент вариации находится в пределах 8%-40%.

Пример 11. По данной выборке вычислить характеристики рассеяния $R, d, D_в, \sigma_в, V$.

x_i	17	18	19	20	21	22	23
n_i	7	7	3	1	3	2	2

Решение. Согласно формуле (1.22) вариационный размах определится:

$$R = x_{max} - x_{min} = 23 – 17 = 6.$$

Для нахождения остальных характеристик определим выборочную среднюю

$$\bar{x}_в = \frac{17 \cdot 7 + 18 \cdot 7 + 19 \cdot 3 + 20 \cdot 1 + 22 \cdot 2 + 23 \cdot 2}{25} = 19 .$$

Тогда выборочная дисперсия по формуле (1.24′) равна:

$$D_в = \frac{7\cdot(-2)^2 + 7\cdot(-1)^2 + 0\cdot 3 + 1^2\cdot 1 + 2^2\cdot 3 + 3^2\cdot 2 + 4^2\cdot 2}{25} = 3,92 .$$

Среднее квадратическое отклонение найдем как корень квадратный из выборочной дисперсии:

$$\sigma_в = \sqrt{3,92} \approx 1,98 .$$

Коэффициент вариации определим по формуле (26):

$$V = \frac{1,98}{19}\cdot 100\ \% = 10,42\ \%,$$

что свидетельствует о большой концентрации значений признака около выборочной средней.

Эмпирический коэффициент детерминации η^2 представляет собой долю межгрупповой дисперсии в общей дисперсии и рассчитывается по формуле:

$$\eta^2 = \frac{\delta_0^2}{\sigma_o^2}. \tag{1.29}$$

Этот коэффициент показывает долю (удельный вес) общей вариации изучаемого признака, обусловленного вариацией группировочного признака.

Эмпирическим корреляционным отношением η называется корень квадратный из эмпирического коэффициента детерминации, определяемое по формуле:

$$\eta = \sqrt{\frac{\delta_0^2}{\sigma_o^2}}. \tag{1.30}$$

Эмпирическое корреляционное отношение характеризует влияние признака, положенного в основу группировки, на вариацию результативного признака и изменяется в пределах от 0 до 1. Если $\eta = 0$, признак не оказывает влияния на результативный. Если $\eta = 1$, то результативный признак изменяется только в зависимости от признака, положенного в основу группировки, а влияние прочих факторных признаков равно нулю.

1.4.3 Характеристики формы эмпирического распределения

При анализе вариационных рядов смещение от центра и крутизну распределения характеризуют специальные показатели, называемые *характеристиками формы*. Эмпирические распределения, как правило, смещены от центра распределения вправо или влево, асимметричны. Нормальное распределение строго симметрично относительно средней арифметической, что обусловлено четностью функции плотности вероятности. *Асимметрия распределения* возникает вследствие того, что какие-либо факторы действуют в одном направлении сильнее, чем в другом, или процесс развития явления таков, что доминирует какая-то причина. Кроме того, природа некоторых явлений такова, что имеет место асимметричное распределение.

Моментом порядка p распределения вариационного ряда называется величина, определяемая по формуле:

$$\mu_p = \frac{1}{n}\sum_{i=1}^{k}(x_i - a)^p \cdot n_i. \tag{1.31}$$

В зависимости от значения a общая система моментов разбивается на 3 подсистемы. Если $a = 0$, получаем *систему начальных моментов*:

$$m_p = \frac{1}{n}\sum_{i=1}^{k} x_i^{\,p} \cdot n_i\ . \tag{1.32}$$

При $a = \overline{x}_в$ получаем *систему центральных моментов*:

$$\mu_p = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^p \cdot n_i. \tag{1.33}$$

Если $a = C = const$ (обычно C близко к середине вариационного ряда, причем, если k – четное, то $C = x_{k\backslash 2}$, если k – нечетное, то $C = x_{(k+1)\backslash 2}$), получаем *систему условных моментов*:

$$M_p^* = \frac{1}{n}\sum_{i=1}^{k} u_i^p \cdot n_i = \frac{1}{n}\sum_{i=1}^{k} (x_i - C)^p \cdot n_i, \tag{1.34}$$

вычисленных для *условных вариант* u_i. Эта система применяется для упрощения расчетов.

Между центральными и начальными моментами имеют место следующие соотношения:

$$\mu_1 = 0;\ \mu_2 = m_2 - \left(m_1\right)^2;\ \mu_3 = m_3 - 3m_1 \cdot m_2 + 2(m_1)^3; \tag{1.35}$$

$$\mu_4 = m_4 - 4m_1 m_3 + 6m_2 (m_1)^2 - 3(m_1)^4. \tag{1.35}$$

Имеют место удобные для вычислений формулы, выражающие центральные моменты через условные:

$$\mu_2 = \left[M_2^* - (M_1^*)^2\right] \cdot h^2; \quad \mu_3 = \left[M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3\right] \cdot h^3;$$

$$\mu_4 = \left[M_4^* - 4M_1^* M_3^* + 6M_2^* (M_1^*)^2 - 3(M_1^*)^4\right] \cdot h^4. \tag{1.36}$$

Центральные моменты 3 и 4 порядков используются для определения коэффициентов асимметрии и эксцесса эмпирических распределений.

Коэффициент асимметрии определяет направление и величину смещения (асимметрию) распределения и рассчитывается по формуле:

$$A_s = \frac{\mu_3}{\sigma_в^3} = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^3 \cdot n_i}{n \cdot \sigma_в^3}. \tag{1.37}$$

Если вершина кривой распределения сдвинута вправо и левая часть ветви оказывается длиннее правой, то такая асимметрия является *правосторонней*, в противоположном случае *левосторонней*. При *левосторонней* (*отрицательной*) *асимметрии* коэффициент асимметрии $A_s < 0$, при *правосто-*

ронней (*положительной*) $A_s > 0$. Если $A_s = 0$, то распределение имеет симметричную форму, т.е. варианты, равноудаленные от выборочной средней, имеют одинаковую частоту.

Эксцесс (или *коэффициент крутости*) *эмпирического распределения* находится по формуле:

$$E_k = \frac{\mu_4}{\sigma_в^4} - 3 = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^4 \cdot n_i}{n \cdot \sigma_в^4} - 3. \qquad (1.38)$$

Эксцесс у нормальной кривой равен нулю. Кривые, у которых $E_k < 0$, по сравнению с нормальной, менее крутые и имеют более *плоскую вершину*. Кривые, у которых $E_k > 0$ более крутые, чем нормальная кривая и имеют более *острую вершину*.

Коэффициент асимметрии и эксцесс могут быть использованы в качестве предварительного выбора закона для эмпирического распределения. В частности, если $\left|A_s\right| < 3\sigma_{A_s}$ и $\left|E_k\right| < 3\sigma_{E_k}$, можно выдвинуть гипотезу о нормальном законе распределения случайной величины *X*. Здесь σ_{A_s} и σ_{E_k} - эмпирические средние квадратические отклонения коэффициента асимметрии и эксцесса, определяемые по формулам:

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)\cdot(n+3)}}. \qquad (1.39) \qquad \sigma_{E_k} = \sqrt{\frac{24n\cdot(n-2)\cdot(n-3)}{(n-1)^2\cdot(n+3)\cdot(n+5)}}. \qquad (1.40)$$

Пример 12. По данному дискретному вариационному ряду найти коэффициент асимметрии и эксцесс:

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

Решение. Найдем выборочную среднюю по формуле (1.13′):

$$\overline{x}_в = \frac{10 \cdot 6 + 15 \cdot 16 + 20 \cdot 50 + 25 \cdot 24 + 30 \cdot 4}{6 + 16 + 50 + 24 + 4} = \frac{2020}{100} = 20,2.$$

Для нахождения выборочной дисперсии, а значит и среднего квадратического отклонения, определим квадратичную среднюю:

$$\overline{x_в^2} = \frac{10^2 \cdot 6 + 15^2 \cdot 16 + 20^2 \cdot 50 + 25^2 \cdot 24 + 30^2 \cdot 4}{6 + 16 + 50 + 24 + 4} = \frac{42820}{100} = 428.$$

Тогда $D_в = 428 - (20,2)^2 = 428 - 408,04 = 19,96$, $\sigma_в = \sqrt{19,96} \approx 4,47$.

Найдем центральные эмпирические моменты 3-го и 4-го порядков по формуле (1.36):

$$\mu_3 = \frac{1}{100}\begin{bmatrix}(10 - 20,2)^3 \cdot 6 + (15 - 20,2)^3 \cdot 16 + (20 - 20,2)^3 \cdot 50 + \\ + (25 - 20,2)^3 \cdot 24 + (30 - 20,2)^3 \cdot 4\end{bmatrix} = -21,99.$$

$$\mu_4 = \frac{1}{100}\begin{bmatrix}(10 - 20,2)^4 \cdot 6 + (15 - 20,2)^4 \cdot 16 + (20 - 20,2)^4 \cdot 50 + \\ + (25 - 20,2)^4 \cdot 24 + (30 - 20,2)^4 \cdot 4\end{bmatrix} = 1262,8.$$

Тогда коэффициент асимметрии и эксцесс, вычисленные по формулам (1.37) и (1.38) соответственно, будут равны:

$$A_s = \frac{\mu_3}{\sigma_в^3} = \frac{-21,9}{\sqrt{19,96}^{\,3}} \approx -0,246\,, . \qquad E_k = \frac{\mu_4}{\sigma_в^4} - 3 = \frac{1262,8}{398,4} - 3 = 0,17.$$

Найденные значения коэффициента асимметрии и эксцесса указывают на то, что распределение не симметрично: имеет левостороннюю асимметрию и кривая распределения более крутая, чем нормальная кривая.

Задачи для самостоятельного решения

11. Дана некоторая выборка

38	60	41	51	33	42	45	21	53	60
68	52	47	46	42	43	57	44	54	59
77	47	28	27	49	49	14	28	61	30
61	35	47	46	58	45	42	21	30	40

67	65	39	35	41	60	54	42	59	60

Составить интервальный вариационный ряд, изобразить его графически. Определить все характеристики положения.

12. Найти выборочную среднюю, моду и медиану распределения количества бракованных изделий, имеющихся среди ста изделий, сошедших с конвейера:

Количество бракованных изделий	0	1	2	3	4	5
Количество партий по 100 изделий	7	15	14	10	0	4

13. Известно распределение работников некоторого предприятия по стажу работы:

Стаж работы, лет	До 1	1-5	5-10	10-20	20-40	Всего
Число работников	8	12	16	14	10	60

Найти средний стаж работы, среднее квадратическое отклонение и коэффициент вариации.

14. Сгруппированные значения выборки признака X даны в таблице

Интервалы	Середина интервала x_i^o	Частота, n_i
27,5; 29,5	28,5	3
29,5; 31,5	30,5	9
31,5; 33,5	32,5	23
33,5; 35,5	34,5	33
35,5; 37,5	36,5	38
37,5; 39,5	38,5	34
39,5; 41,5	40,5	21
41,5; 43,5	42,5	8
43,5; 45,5	44,5	1

Вычислить выборочное среднее значение, моду и медиану (аналитически и графически), среднее квадратическое отклонение, асимметрию и эксцесс выборки.

15. По выборке признака X , заданной следующей таблицей:

x_i	45	50	55	60	65	70	75

n_i	4	6	10	40	20	12	8

найти все характеристики рассеяния и формы распределения.

16. Имеется распределение хозяйств по количеству рабочих на 100 га сельскохозяйственных угодий:

Группы хозяйств по численности работников на 100 га с\х угодий	4-5,6	5,61-7,2	7,21-8,8	8,81-10,4	10,41-12,0	12,01-13,6	13,61-15,2
Число хозяйств в группе	5	17	9	15	10	1	3

Определить: среднюю численность работников на 100 га сельскохозяйственных угодий по исследуемой совокупности хозяйств, дисперсию и среднее квадратическое отклонение, модальное и медианное значение числа числа работников на 100 га с\х угодий, коэффициенты вариации и асимметрии, а также эксцесс. Сделать выводы по результатам вычислений.

17. Имеются следующие данные о площади посевов овощей в 7 хозяйствах 10 районов Иркутской области:

Район	Номер хозяйства						
	1	2	3	4	5	6	7
1	8	10	12	6	15	30	21
2	32	16	26	41	44	38	-
3	101	165	230	144	84	76	260
4	22	30	44	18	16	31	-
5	10	7	4	3	12	7	6
6	255	366	384	273	450	510	-
7	121	84	96	110	161	143	-
8	34	16	84	71	36	8	17
9	46	41	48	52	50	58	-
10	15	24	86	144	34	14	24

Определить дисперсию площади посевов овощей по каждому району и в целом по всем приведенным районам. Дать оценку колеблемости площади посевов овощей в хозяйствах указанных районов с применением правила сложения дисперсий.

18. Дана некоторая выборка

14	16	20	15	18	12	13	10	17	15	12
17	11	17	13	11	17	16	12	16	13	16
12	19	18	9	7	18	15	11	11	14	13
10	16	15	17	21	13	17	16	13	15	15
18	14	16	14	10	14	8	15	16	16	14

Найти и построить эмпирическую функцию распределения, начальные и центральные моменты 1-4 порядков (проверить связь между ними по формулам (35)), вычислить коэффициент асимметрии и эксцесс. Сделать выводы по результатам расчетов.

19. Работники предприятия сгруппированы по возрасту:

Категории работников	Возраст работников, лет				
	До 30	30-40	40-50	50-60	Свыше 60
Рабочие	43	141	216	127	118
Руководители	2	4	6	8	4
Специалисты	48	163	252	169	144

Определить: средний возраст работников предприятия в целом и по отмеченным категориям; модальное и медианное значения возраста работников по категориям и предприятию в целом; дисперсию и среднее квадратическое отклонение возраста работников по категориям и предприятию; межгрупповую дисперсию возраста работников; эмпирическое корреляционное отношение; общую дисперсию возраста работников, используя правило сложения дисперсий.

20. Проведите анализ данных годовых уровней прибыли (у.е.) 3-х компаний:

Год	“Cherry Computers”	“Lemon Motors”	“Orange Electronics”
2002	14,2	-6,2	37,5

2003	12,3	13,3	-10,6
2004	-16,2	-8,4	40,3
2005	15,4	27,3	5,4
2006	17,2	28,2	6,2
2007	10,3	14,5	10,2
2008	-6,3	-2,4	13,8
2009	-7,8	-3,1	11,5
2010	3,4	15,6	-6,2
2011	12,2	18,2	27,5

Найти среднее значение и стандартное отклонение прибыли для каждой из компаний. Сравните результаты их деятельности за 10 лет. Деятельность какой из компаний, по вашему мнению, более успешна?

1.4.4 Метод произведений для расчета выборочной средней и дисперсии

Для упрощения вычислений средних характеристик и показателей рассеяния, а также для уменьшения погрешности округления используют *метод произведений*. Если выборка задана в виде распределения *равноотстоящих* вариант и соответствующих им частот, то удобно находить выборочную среднюю и дисперсию по формулам:

$$\overline{x}_в = M_1^* \cdot h + C. \qquad (1.41) \qquad D_в = \left[M_2^* - \left(M_1^* \right)^2 \right] \cdot h^2, \qquad (1.42)$$

где h – длина интервала группировки, C – ложный ноль (варианта, у которой самая большая частота), M_1^* и M_2^* – условные моменты первого и второго порядка, определяемые по формуле (1.34).

Пример 13. По данной выборке методом произведений вычислить $\overline{x}_в, D_в$.

x_i	12	14	16	18	20	22

n_i	5	15	50	16	10	4

Решение. Данное распределение представляет собой распределение равностоящих вариант и соответствующих им частот. Для вычисления $\overline{x}_в, D_в$ методом произведений составим расчётную таблицу:

1) запишем варианты 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, которая принадлежит строке, содержащей наибольшую частоту, пишем 0;

4) над нулем последовательно запишем -1, -2;

5) под нулём последовательно запишем 1, 2, 3;

6) в четвёртый столбец записываем произведение частот n_i на *условные варианты* $u_i = \dfrac{x_i - C}{h}$, в нижнюю клетку этого столбца запишем сумму произведений $u_i \cdot n_i$;

7) произведение частот на квадраты условных вариант, то есть $n_i \cdot u_i^2$ запишем в пятый столбец. Сумму чисел $n_i \cdot u_i^2$ (127) записываем в нижнюю клетку столбца;

8) произведения частот на квадраты условных вариант, увеличенных на единицу, то есть $n_i \cdot (u_i + 1)^2$, запишем в шестой контрольный столбец. Сумму чисел столбца (273) помещаем в нижнюю клетку шестого столбца.

Таблица заполнена. Для *контроля вычислений* воспользуемся тождеством:

$$\sum_{i=1}^{к} n_i(u_i + 1)^2 = \sum_{i=1}^{к} n_i u_i^2 + 2\sum_{i=1}^{к} n_i u_i + n$$

$$\sum_{i=1}^{6} n_i(u_i + 1)^2 = 273, \qquad \sum_{i=1}^{6} n_i u_i^2 + 2\sum_{i=1}^{6} n_i u_i + n = 127 + 2 \cdot 23 + 100 = 273.$$

Таблица к примеру 11

x_i	n_i	u_i	$n_i \cdot u_i$	$n_i \cdot u_i^2$	$n_i \cdot (u_i+1)^2$
12	5	-2	-10	20	5
14	15	-1	-15	15	-
16	50	0	0	-	50
18	16	1	16	16	64
20	10	2	20	40	90
22	4	3	12	36	64
	n =100		$\sum n_i u_i = 23$	$\sum n_i u_i^2 = 127$	$\sum n_i (u_i+1)^2 = 273$

Для вычисления выборочной средней и дисперсии воспользуемся формулами (1.41) и (1.42):

$$C = 16,\ h = 14 - 12 = 2,\quad M_1^* = \frac{23}{100} = 0{,}23,\quad M_2^* = \frac{127}{100} = 1{,}27.$$

$$\overline{x}_в = M_1^* \cdot h + C = 0{,}23 \cdot 2 + 16 = 16{,}46;$$

$$D_в = \left[M_2^* - (M_1^*)^2\right] \cdot h^2 = \left[1{,}27 - 0{,}23^2\right] \cdot 2^2 = 4{,}87.$$

Пример 14. Даны выборочные варианты и соответствующие им частоты количественного признака *X* (данные *Примера* 12):

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

Найти методом произведений: а) выборочную среднюю $\overline{x}_в$, дисперсию $D_в$ и среднее квадратическое отклонение $\sigma_в$; б) асимметрию и эксцесс.

Решение. Составим расчетную таблицу, выполнив следующие действия:

1) запишем варианты x_i в первый столбец;

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

3) в качестве ложного нуля *C* выберем варианту 20 (она расположена в середине вариационного ряда); в столбце $u_i = \dfrac{x_i - C}{h}$, напротив строки, со-

держащей варианту 20, пишем 0; над нулем последовательно записываем условные варианты –1, -2, а под нулем – последовательно 1, 2;

4) произведения частот на условные варианты записываем в четвертый столбец – столбец $n_i u_i$; 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 – столбец $n_i u_i^2$, сумму чисел столбца помещаем в нижнюю клетку столбца;

6) произведения частот на квадраты условных вариант, увеличенных на единицу, запишем в шестой (контрольный) столбец – столбец $n_i (u_i + 1)^2$, сумму чисел столбца помещаем в нижнюю клетку столбца.

7) произведения частот на условные варианты в кубе $n_i u_i^3$; сумму чисел столбца помещаем в нижнюю клетку столбца;

8) произведения частот на условные варианты в четвертой степени $n_i u_i^4$; сумму чисел столбца помещаем в нижнюю клетку столбца;

9) произведения частот на условные варианты, увеличенных на единицу в четвертой степени запишем в девятый (контрольный) столбец – столбец $n_i (u_i + 1)^4$; сумму чисел столбца помещаем в нижнюю клетку столбца.

В итоге получим следующую расчетную таблицу:

Таблица к примеру 14

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i (u_i + 1)^4$
10	6	-2	-12	24	6	-48	96	6
15	16	-1	-16	16	0	-16	16	0
20	50	0	<u>0</u>	0	50	0	0	50
25	24	1	24	24	96	24	24	384
30	4	2	8	16	36	32	64	324
	n=10		$\sum n_i u_i = 4$	$\sum n_i u_i^2 = 80$	$\sum n_i (u_i + 1)^2 = 188$	$\sum n_i u_i^3 = -8$	$\sum n_i u_i^4 = 200$	$\sum n_i (u_i + 1)^4 = 764$

Контроль:

$$\sum_{i=1}^{5} n_i(u_i+1)^2 = 188, \qquad \sum_{i=1}^{5} n_i u_i^2 + 2\sum_{i=1}^{5} n_i u_i + n = 80 + 2\cdot 4 + 100 = 188.$$

Совпадение найденных сумм свидетельствует о правильности вычислений. Вычислим условные моменты первого, второго, третьего и четвертого порядков по формуле (1.34):

$$M_1^* = \frac{\sum_{i=1}^{5} n_i \cdot u_i}{n} = \frac{4}{100} = 0{,}04; \qquad M_2^* = \frac{\sum_{i=1}^{5} n_i \cdot u_i^2}{n} = \frac{80}{100} = 0{,}8;$$

$$M_3^* = \frac{\sum_{i=1}^{5} n_i \cdot u_i^3}{n} = \frac{-8}{100} = -0{,}08; \qquad M_4^* = \frac{\sum_{i=1}^{5} n_i \cdot u_i^4}{n} = \frac{200}{100} = 2.$$

Шаг h (разность между двумя соседними вариантами) равен:

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

Выборочная средняя и дисперсия согласно формулам (1.41) и (1.42) определятся:

$$\overline{x}_в = 0{,}04\cdot 5 + 20 = 20{,}2. \qquad D_в = \left[0{,}8 - (0{,}04)^2\right]\cdot 5^2 = 19{,}96.$$

Выборочное среднеквадратическое отклонение находится как корень квадратный из выборочной дисперсии:

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

Найдем центральные эмпирические моменты 3-го и 4-го порядков по формулам (1.33):

$$\mu_3 = \left[-0{,}08 - 3\cdot 0{,}04\cdot 0{,}8 + 2(0{,}04)^3\right]\cdot 5^3 = -21{,}99.$$

$$\mu_4 = \left[2 - 4\cdot 0{,}04\cdot(-0{,}08) + 6(0{,}04)^2\cdot 0{,}8 - 3(0{,}04)^4\right]\cdot 5^4 = 1262{,}8.$$

Тогда коэффициент асимметрии и эксцесс, вычисленные по формуле (1.37) и (1.38) соответственно, будут равны:

$$A_s = \frac{\mu_3}{\sigma_в^3} = \frac{-21{,}9}{\sqrt{19{,}96}^{\,3}} \approx -0{,}246 \; .$$

$$E_k = \frac{\mu_4}{\sigma_в^4} - 3 = \frac{1262{,}8}{398{,}4} - 3 = 0{,}17 \; .$$

Полученные результаты вычислений для выборочной средней, выборочной дисперсии, коэффициента асимметрии и эксцесса полностью совпадают с решением *Примера* 12, где не использовался метод произведений.

1.4.5 Метод сумм для расчета выборочной средней и дисперсии

Метод сумм, как и метод произведений для расчета выборочной средней и дисперсии, применяется для упрощения вычислений и уменьшения погрешности округления. При использовании *метода сумм* условные моменты первого, второго, третьего и четвертого порядков находят по формулам:

$$M_1^* = \frac{d_1}{n}; \quad M_2^* = \frac{S_1 + 2S_2}{n}; \quad M_3^* = \frac{d_1 + 6d_2 + 6d_3}{n};$$

$$M_4^* = \frac{S_1 + 14S_2 + 36S_3 + 24S_4}{n}, \qquad (1.43)$$

где $d_1 = a_1 - b_1,\; d_2 = a_2 - b_2,\; d_3 = a_3 - b_3,$

$$S_1 = a_1 + b_1,\; S_2 = a_2 + b_2,\; S_2 = a_2 + b_2, S_3 = a_3 + b_3, S_4 = a_4 + b_4. \quad (*)$$

Практическое вычисление чисел a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4 покажем на примере.

Пример 15. Методом сумм вычислить выборочную среднюю арифметическую, дисперсию, коэффициент асимметрии и эксцесс по заданному распределению выборки объема $n = 100$.

x_i	48	52	56	60	64	68	72	76	80	84
n_i	2	4	6	8	12	30	18	8	7	5

Решение. Так как выборка состоит из равноотстоящих вариант с шагом h, то выборочная средняя арифметическая и дисперсия определятся по формулам (1.41) и (1.42), а условные моменты – по формулам (1.43). Числа a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4 находим по определенному правилу. Составим расчетную таблицу:

1) в первый столбец записываем значения вариант;

2) во второй столбец помещаем частоты и считаем, сумму частот (100) разместим в нижнюю клетку столбца;

3) в качестве ложного нуля выбираем значение варианты наибольшей частоты (C = 68), на пересечении третьего и четвертого столбцов и строки, содержащей ложный нуль, ставим нули; в четвертом столбце над ложным нулем и под ним пишем нули;

4) в оставшихся незаполненных над нулем клетках третьего столбца (исключая самую верхнюю) записываем накопленные частоты: 2; 2 + 4 = 6; 6 + 6 = 12; 12 + 8 = 20; 20 + 12 = 32; находим их суммы, получим число $b_1 = 72$, которое поместим в верхнюю клетку третьего столбца. Аналогично в клетках третьего столбца, расположенных под нулем (исключая самую нижнюю), записываем последовательно накопленные частоты (снизу вверх): 5; 5 + 7 = 12; 12 + 8 = 20; 20 + 18 = 38; вычисляем их сумму, получим число $a_1 = 75$, которое поместим в нижнюю клетку третьего столбца;

5) аналогично заполняется четвертый столбец, причем суммируются частоты третьего столбца; сложив все накопленные частоты, расположенные над нулем, получим число $b_2 = 70$,которое поместим в верхнюю клетку четвертого столбца; сумма накопленных частот, расположенных под нулем равна числу a_2 = 59, которое поместим в нижнюю клетку четвертого столбца;

6) для заполнения пятого столбца запишем нуль в клетке строки, содержащей ложный нуль (68), над этим нулем и под ним поставим еще два нуля;

7) в клетках над нулями запишем накопленные частоты, для чего просуммируем частоты четвертого столбца (сверху вниз): 2; 2 + 8 = 10; 10 + 20 = 30; сложив накопленные частоты, получим число $b_3 = 2 + 10 + 30 = 42$, которое поместим в верхнюю клетку пятого столбца; в клетках под нулями запишем накопленные частоты, для чего просуммируем частоты четвертого столбца (снизу вверх): 5; 5 + 17 = 22; сложив накопленные частоты, получим число $a_3 = 5 + 22 = 27$, которое поместим в нижнюю клетку пятого столбца;

8) аналогично заполняется шестой столбец, причем суммируют частоты пятого столбца; сложив накопленные частоты, расположенные над нулями, получим число $b_4 = 2 + 12 = 14$, которое запишем в верхнюю клетку шестого столбца; сложив числа, расположенные под нулями (у нас только одно слагаемое), получим число $a_4 = 5$, которое поместим в нижнюю клетку шестого столбца.

В итоге получим расчетную таблицу:

Таблица к примеру 15

x_i	n_i	$b_1 = 72$	$b_2 = 70$	$b_3 = 42$	$b_4 = 42$
48	2	2	2	2	2
52	4	6	8	10	12
56	6	12	20	30	0
60	8	20	40	0	0
64	12	32	0	0	0
68	30	0	0	0	0
72	18	38	0	0	0
76	8	20	37	0	0
80	7	12	17	22	0
84	5	5	5	5	5

	100	$a_1 = 75$	$a_2 = 59$	$a_3 = 27$	$a_3 = 27$

Контроль: сумма чисел, расположенных непосредственно над нулем третьего столбца, слева от него и под ним, должна быть равна объему выборки (32 + 30 + 38 = 100). Сумма двух чисел, расположенных над двумя ступеньками ступенчатой линии (начерчена жирными отрезками), должна быть равна соответственно числам b_i, стоящим над *предшествующей* ступенькой (при движении по «лесенке» вверх): $32 + 40 = 72 = b_1$; $40 + 30 = 70 = b_2$; $30 + 12 = 42 = b_3$. Аналогично проверяется совпадение сумм двух чисел, стоящих под ступеньками лесенки, ведущей вниз: $38 + 37 = 75 = a_1$;
$37 + 22 = 59 = a_2$; $22 + 5 = 27 = a_3$. При несовпадении хотя бы одной из указанных сумм следует искать ошибку расчета.

Из значений таблицы найдем числа d_i (i =1, 2, 3) и S_i (i =1, 2, 3, 4):

$$d_1 = a_1 - b_1 = 75 - 72 = 3;\ d_2 = a_2 - b_2 = 59 - 70 = -11;$$

$$d_3 = a_3 - b_3 = 27 - 42 = -15;\ S_1 = a_1 + b_1 = 75 + 72 = 147;$$

$$S_2 = a_2 + b_2 = 59 + 70 = 129;\ S_3 = a_3 + b_3 = 27 + 42 = 69;$$

$$S_4 = a_4 + b_4 = 5 + 14 = 19.$$

Тогда условные моменты (*моменты условных вариант*) определятся:

$$M_1^* = \frac{d_1}{n} = \frac{3}{100} = 0{,}03; \qquad M_2^* = \frac{S_1 + 2S_2}{n} = \frac{147 + 2 \cdot 129}{100} = 4{,}05;$$

$$M_3^* = \frac{d_1 + 6d_2 + 6d_3}{n} = \frac{3 + 6\cdot(-11) + 6\cdot(-15)}{100} = -1{,}53;$$

$$M_4^* = \frac{S_1 + 14S_2 + 36S_3 + 24S_4}{n} = \frac{147 + 14\cdot 129 + 36\cdot 69 + 24\cdot 19}{100} = 48{,}9.$$

Так как $h = 4$ и $C = 68$, то выборочная средняя и дисперсия будут равны:

$$\bar{x}_в = 0{,}03 \cdot 4 + 68 = 68{,}12;\ D_в = \sigma_в^2 = \left[4{,}05 - (0{,}03)^2\right] \cdot 4^2 = 64{,}78.$$

Для нахождения коэффициента асимметрии и эксцесса, найдем центральные эмпирические моменты третьего и четвертого порядков по формулам (1.36):

$$\mu_3=\left[M_3^*-3M_1^*M_2^*+2(M_1^*)^3\right]\cdot h^3=\left[-1{,}53-3\cdot 0{,}03\cdot 4{,}05+2\cdot(0{,}03)^3\right]\cdot 4^3=-121{,}248\ ;$$

$$\mu_4=\left[M_4^*-4M_1^*M_3^*+6M_2^*(M_1^*)^2-3(M_1^*)^4\right]\cdot h^4=$$

$$=\left[48{,}93-4\cdot 0{,}03\cdot(-1{,}53)+6\cdot(0{,}03)^2\cdot 4{,}05-3\cdot(0{,}03)^4\right]\cdot 4^4=49{,}135\ .$$

Найдем искомые коэффициент асимметрии и эксцесс, используя формулы (1.37) и (1.38):

$$A_s=\frac{\mu_3}{\sigma_в^3}=\frac{-121{,}248}{\sqrt{64{,}78^3}}\approx-0{,}23;\qquad E_k=\frac{\mu_4}{\sigma_в^4}-3=\frac{49{,}134}{\sqrt{64{,}78^4}}-3=0{,}01\ .$$

Ответ: $\overline{x_в}=68{,}12;\ D_в=64{,}78;\ A_s=-0{,}23;\quad E_k=0{,}01$.

Если первоначальные варианты *не являются равноотстоящими*, то интервал, в котором заключены все варианты выборки делят на несколько равных, длиною h, частичных интервалов (каждый частичный интервал должен содержать не менее 8-10 вариант). Затем находят середины частичных интервалов, которые и образуют последовательность равноотстоящих вариант. В качестве частоты каждой середины интервала принимают сумму частот вариант, которые попали в соответствующий частичный интервал.

При вычислении выборочной дисперсии для уменьшения ошибки, вызванной группировкой (особенно при малом числе интервалов), делают *поправку Шеппарда*, вычитая из вычисленной дисперсии одну двенадцатую квадрата длины частичного интервала. Таким образом, с учетом поправки Шеппарда дисперсия вычисляется по формуле:

$$D_{в}^{*} = D_{в} - \frac{1}{12}h^{2}. \qquad (1.44)$$

Пример 16. Найти методом произведений выборочное среднее и выборочную дисперсию по заданному распределению выборки:

x_i	2	3	7	9	11	12,5	16	18	23	25	26
n_i	3	5	10	6	10	4	12	13	8	20	9

Решение. Разобьем интервал 2-26 на 4 частичные интервала длиною $h = 6$: 2 – 8; 8 – 14; 14 – 20; 20 - 26 . Приняв середины частичных интервалов в качестве новых вариант y_i, получим равноотстоящие варианты:

$$y_1 = 5;\ y_2 = 11;\ y_3 = 17;\ y_4 = 23.$$

В качестве частоты n_1 варианты y_1 примем сумму частот вариант, попавших в первый интервал: $n_1 = 3 + 5 + 10 = 18$. Найдя аналогично частоты остальных вариант, получим распределение равноотстоящих вариант:

y_i	5	11	17	23
n_i	18	20	25	37

Пользуясь методом произведений, найдем

$$\overline{y_{в}} = 15{,}86, \qquad D_{в} = 45{,}14 .$$

Учитывая, что число частичных интервалов мало (4), учтем поправку Шеппарда:

$$D_{в}^{*} = D_{в} - \frac{1}{12}h^{2} = 45{,}14 - \frac{6^{2}}{12} = 42{,}14 .$$

Задачи для самостоятельного решения

Задачи 21-25 решить, используя *метод произведений*.

21. На молочной ферме крестьянского хозяйства имеются сведения о величине удоя коров за лактационный период:

Величина удоя, кг	Количество коров	Величина удоя, кг	Количество коров
400-600	1	1600-1800	14
600-800	3	1800-2000	12
800-1000	6	2000-2200	10
1000-1200	11	2200-2400	6
1200-1400	15	2400 и выше	2
1400-1600	20		

Найти величину среднего удоя и выборочную дисперсию.

22. По данной выборке распределения:

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

Найти выборочную среднюю арифметическую, дисперсию, коэффициент асимметрии и эксцесс.

23. Имеются следующие данные о жирности молока в фермерском хозяйстве:

Жирность, %	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2
Число коров	8	11	28	43	71	52	29	8

Определить величину средней жирности и выборочную дисперсию.

24. По заданному распределению неравноотстоящих вариант выборки объема $n = 50$

x_i	6	8	11	13	15,5	17,5	20	23,5	24,5	26
n_i	1	9	6	6	4	6	8	5	4	1

найти выборочную дисперсию с учетом поправки Шеппарда.

25. По заданному распределению неравноотстоящих вариант выборки объема $n = 100$

x_i	10	13	15	17	19	23	24	26	28	32	34	35
n_i	2	4	6	8	9	6	20	15	10	8	7	5

найти: а) выборочную среднюю арифметическую;

б) выборочную дисперсию с учетом поправки Шеппарда;

г) коэффициент асимметрии и эксцесс.

Задачи 26-30 решить, используя *метод сумм.*

26. Имеются данные дискретного ряда распределения студентов одного из факультетов по возрасту:

Группы студентов по возрасту, лет	17	18	19	20	21	22	23	24
Число студентов	20	80	90	110	130	170	90	60

Определить средний возраст студентов, выборочную дисперсию и среднее квадратическое отклонение, коэффициент вариации.

27. Определите среднюю арифметическую, дисперсию, коэффициент асимметрии и эксцесс по данным таблицы:

Группы магазинов по размерам товарооборота, млн. руб.	50-60	60-70	70-80	80-90
Число магазинов	7	15	6	4

28. По заданному распределению выборки объема $n = 100$

x_i	10,2	10,4	10,6	10,8	11	11,2	11,4	11,6	11,8	12
n_i	2	3	8	13	25	20	12	10	6	1

найти: а) выборочную среднюю арифметическую и дисперсию;

б) коэффициент асимметрии и эксцесс.

29. По заданному распределению неравноотстоящих вариант выборки объема $n = 20$

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

найти: а) выборочную среднюю арифметическую;

б) выборочную дисперсию с учетом поправки Шеппарда;

г) коэффициент асимметрии и эксцесс.

30. По заданному распределению неравноотстоящих вариант выборки объема $n = 150$

x_i	18,4	18,9	19,3	19,6	20,2	20,7	30,2
n_i	5	10	20	35	15	7	3

найти выборочную дисперсию с учетом поправки Шеппарда.

2 СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Как правило, в математической статистике редко о проводимом эксперименте совсем ничего нельзя сказать. Обычно, во многих задачах, тип распределения известен заранее, но неизвестны параметры или часть параметров этого распределения. Например, ошибки измерения предполагаются распределенными по нормальному закону с математическим ожиданием равным нулю (если систематическая ошибка равна нулю) и неизвестной дисперсией; число покупателей в магазине в течении часа имеет распределение Пуассона с неизвестным параметром λ и т.д. В результате мы имеем *задачу статистического оценивания параметров* этого распределения на основе выборочных данных.

Пусть мы располагаем исходными статистическими данными - выборкой $(x_1, x_2, \dots , x_n)$ объема n из исследуемой генеральной совокупности. Пусть изучаемая случайная величина распределена по закону $P(x;\theta)$, где θ - параметр распределения, значение которого неизвестно до получения выборки. *Задача оценивания* неизвестного параметра θ состоит в построении такой функции $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$ от имеющихся в нашем распоряжении данных, которая давала бы в определенном смысле наиболее точное приближенное значение истинного (неизвестного нам) параметра θ . Любую функцию выборки называют *статистикой.* Статистика $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$, используемая для оценки неизвестного параметра θ , называется *статистической оценкой параметра* θ . Оценка, полученная в виде одного числа - точки на числовой прямой, называется *точечной*. Оценка, полученная в виде интервала, называется *интервальной*.

2.1 Несмещенные, эффективные и состоятельные оценки

Все статистики и статистические оценки являются случайными величинами, принимающими различные значения при переходе от одной выборки к другой (даже в рамках одной генеральной совокупности). Однако значения оценки, подсчитанные по разным выборкам и подтвержденные случайному разбросу, должны концентрироваться около истинного значения оцениваемого параметра. Это обеспечивается требованиями, предъявляемыми к точечным оценкам, которые формулируются с помощью трех свойств оценок: состоятельности, несмещенности и эффективности.

Статистика θ^* называется *состоятельной* оценкой параметра θ , если по мере роста числа наблюдений (при $n \to \infty$), она стремится по вероятности к оцениваемому значению θ , т.е. при $\varepsilon > 0$ выполняется равенство:

$$\lim_{n \to \infty} \left(\left| \theta^* - \theta \right| < \varepsilon \right) = 1 . \tag{2.1}$$

Это означает, что для любых положительных чисел ε и γ найдется такое число $n_{\varepsilon\gamma}$, что для всех чисел n, удовлетворяющих неравенству $n > n_{\varepsilon\gamma}$ выполняется условие:

$$P\left(\left\|\theta^{*} - \theta\right\| < \varepsilon\right) > 1 - \gamma \ . \tag{2.1'}$$

Статистика θ^{*} называется *несмещенной* оценкой пара метра θ, если при любом объеме выборки n результат ее усреднения по всем возможным выборкам приводит к точному истинному значению оцениваемого параметра, т.е. выполняется условие:

$$M\,(\theta^{*}) = \theta, \quad \forall\,\theta \in \Theta \ . \tag{2.2}$$

Свойство несмещенности означает отсутствие ошибки в среднем, при систематическом использовании оценки.

Статистика θ^{*} называется *эффективной* оценкой параметра θ, если она среди всех прочих оценок того же самого параметра обладает наименьшей мерой случайного разброса (наименьшей дисперсией) относительно истинного значения оцениваемого параметра, т.е. имеет место равенство:

$$D\,(\theta^{*}) = min\ D\,(\theta_i^{*})\ , \tag{2.3}$$

где θ_i^{*} - несмещенные оценки параметра θ .

Эффективность является решающим свойством, определяющим качество оценки, и состоящим в том, что из всех возможных несмещенных оценок параметра θ, вычисленных по выборкам одного и того же объема, выбирается та, которая имеет наименьшую дисперсию.

На практике при оценке параметров не всегда оказывается возможным одновременно выполнение требований несмещенности, эффективности и состоятельности оценки. Определяют оценки *асимптотически*, т.е. при $n \to \infty$, удовлетворяющие указанным требованиям.

2.2 Точечное оценивание параметров выборки и генеральной совокупности

Если по выборке (x_1, x_2, … , x_n) требуется оценить математическое ожидание $M(X) = a$ и дисперсию $D(X) = \sigma^2$ случайной величины X, то естественно в качестве оценки математического ожидания взять выборочную среднюю арифметическую, определяемую по формулам (1.13) или (1.13'), т.е.

$$a^* = \overline{x}_в = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad a^* = \overline{x}_в = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}, \tag{2.4}$$

а в качестве оценки дисперсии – выборочную дисперсию, определяемую по формулам (1.24) и (1.24'):

$$D^* = D_в = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}_в\right)^2}{n}, \qquad D^* = D_в = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x}_в\right)^2 \cdot n_i}{\sum_{i=1}^{k} n_i} \tag{2.5}$$

Выборочная средняя арифметическая является состоятельной, несмещенной и эффективной оценкой математического ожидания, а выборочная дисперсия обладает этими свойствами лишь асимптотически.

Для доказательства несмещённости некоторых точечных оценок будем рассматривать выборку объема n как систему n независимых случайных величин $X_1, X_2, .. , X_n$, каждая из которых имеет тот же закон распределения с теми же параметрами, что и случайная величина X, представляющая генеральную совокупность. При таком подходе становятся очевидными равенства:

$$M(x_i) = M(X_i) = M(X);$$

$$D(x_i) = D(X_i) = D(X) \text{ для всех } i = \overline{1, n}.$$

Теперь можно показать, что выборочная средняя $\overline{x}_в$ - есть несмещенная оценка средней генеральной совокупности или, что то же самое, математического ожидания интересующей нас случайной величины X:

$$M(\overline{x}_в) = M\frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\left(M(X_1) + M(X_2) + ... + M(X)_n\right) = \frac{1}{n}nM(X) = M(X).$$

Выведем формулу для дисперсии выборочной средней:

$$D(\overline{x}_в) = D\left(\frac{x_1 + x_2 + ... + x_n}{n}\right) = \frac{1}{n^2}\left[D(X_1) + D(X)_2 + ... + D(X_n)\right] = \frac{1}{n^2}nD(X) = \frac{D(X)}{n}.$$

Найдем математическое ожидание выборочной дисперсии σ^2. Сначала преобразуем σ^2 следующим образом:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_в)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M(X) + M(X) - \overline{x}_в)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left((x_i - M(X))^2 - 2(x_i - M(X))(\overline{x}_в - M(X)) + (\overline{x}_в - M(X))^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - M(X))^2 - (\overline{x}_в - M(X))^2.$$

Здесь использовано преобразование:

$$\sum_{i=1}^{n} 2(x_i - M(X))(\overline{x}_в - M(X)) = 2(\overline{x}_в - M(X))\sum_{i=1}^{n}(x_i - M(X)) =$$

$$= 2(\overline{x}_в - M(X))\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} M(X)\right) = 2(\overline{x}_в - M(X))(n\overline{x}_в - nM(X)) = 2n(\overline{x}_в - M(X))^2$$

Теперь, используя полученное выше выражение для величины σ^2, найдем ее математическое ожидание.

$$M(\sigma^2) = M\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - M(X))^2 - (\overline{x}_в - M(X))^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} M(x_i - M(X))^2 - M(\overline{x}_в - M(X))^2 = \frac{1}{n}nD(X) - D(\overline{x}_в) =$$

$$= D(X) - \frac{D(X)}{n} = \frac{n-1}{n}D(X)..$$

Так как $M(\sigma^2) \neq D(X)$, то *выборочная дисперсия не является несмещенной оценкой дисперсии генеральной совокупности*.

Чтобы получить несмещенную оценку дисперсии генеральной совокупности, нужно умножить выборочную дисперсию на $\frac{n}{n-1}$. Тогда получится величина

$$s^2 = \frac{n}{n-1}\sigma^2 = \frac{n}{n-1}D_в, \qquad (2.6)$$

называемая *исправленной выборочной дисперсией.*

Соответственно, смещенной и несмещенной оценками среднего квадратического отклонения служат величины:

$$\sigma_в = \sqrt{D_в} \text{ и } s_в = \sqrt{s_в^2}. \qquad (2.7)$$

Пример 17. По данным микропереписи 1999 г. получено следующее распределение населения, проживающего в месте постоянного жительства не с рождения:

Продолжительность проживания в месте постоянного жительства, лет	менее 2	2-5	6-9	10-14	15-24	25 и более
Доля населения в % к итогу	7,5	11,0	10,5	12,3	21,1	37,6

Найти несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения продолжительности проживания в месте постоянного жительства.

Решение. Так как несмещенной оценкой математического ожидания является выборочная средняя арифметическая, а для дисперсии - исправленная выборочная дисперсия, то для их расчета построим таблицу:

Таблица к примеру 17

Продолжительность проживания в месте постоянного жи-	Доля населения в % к	Середина интервала, x_i	$x_i n_i$	$(x_i - \bar{x}_в)$	$(x_i - \bar{x}_в)^2$	$(x_i - \bar{x}_в)^2 n_i$

тельства, лет (x_i,; x_{i+1})	итогу, n_i					
1	2	3	4	5	6	7
менее 2	7,5	0,5	3,8	-17,4	302,8	2271
2 - 5	11,0	3,5	38,5	14,4	207,4	2281,4
6 - 9	10,5	7,5	78,8	-10,4	108,2	1136,1
10 - 14	12,3	12,0	147,6	-5,9	34,8	428,0
15 - 24	21,1	19,5	411,5	1,6	2,6	54,6
25 и более	37,6	29,5	1109,2	11,6	134,6	5060,9
Итого	100,0		1789,4			11232

Определим выборочную среднюю арифметическую взвешенную (графа 3) согласно (2.4):

$$\overline{x}_в = \frac{1789,4}{100} = 17,9 \text{ (года).}$$

Выборочная дисперсия определится по формуле (2.5):

$$D_в = \sigma_в^2 = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x_в}\right)^2 \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{11232}{100} = 112,3.$$

Несмещенные оценки дисперсии и среднего квадратического отклонения определим по формулам (2.6) и (2.7):

$$S^2 = \frac{n}{n-1} D_в = \frac{100}{99} \cdot 112,3 = 113,45; \qquad S = \sqrt{113,45} = 10,65.$$

2.3 Методы статистического оценивания неизвестных параметров

Точечные оценки, удовлетворяющие свойствам несмещенности, эффективности и состоятельности, получают с помощью метода максимального правдоподобия и метода моментов.

2.3.1 Метод максимального правдоподобия

Пусть $(x_1, x_2, \dots, x_n)$ – независимая выборка из распределения с плотностью $p(x,\theta_1,..., \theta_k)$, зависящей от неизвестных параметров $\theta_1,\theta_2,..., \theta_k$.

Функция вида

$$L(x,\theta) = \prod_{i=1}^{n} p(x_i,\theta_1,\theta_2,..., \theta_k) = p(x_1,\theta_1,..., \theta_k)\cdot p(x_2,\theta_1,..., \theta_k)\cdot ... \cdot p(x_n,\theta_1,..., \theta_k) = \qquad (2.8)$$
$$= L(\theta_1,\theta_2,..., \theta_k).$$

называется *функцией правдоподобия,* а функция

$$\ln L(x,\theta) = \ln \prod_{i=1}^{n} p(x_i,\theta_1,\theta_2,..., \theta_k) = \sum_{i=1}^{n} \ln p(x_i,\theta_1,\theta_2,..., \theta_k) \qquad (2.9)$$

называется *логарифмической функцией правдоподобия*.

При составлении функции правдоподобия (2.8) следует иметь в виду, что $x_1, x_2, \dots, x_n$ – фиксированы, а параметры $\theta_1,\theta_2,..., \theta_k$ являются переменными величинами.

Метод максимального правдоподобия состоит в том, что в качестве оценок $\theta_1^*,\theta_2^*,\dots,\theta_k^*$ неизвестных параметров $\theta_1,\theta_2,..., \theta_k$ берутся такие значения аргументов функции правдоподобия, которые обращают ее в максимум. *Оценкой параметра* θ *по методу максимального правдоподобия* называется значение θ^*, при котором функция максимального правдоподобия достигает наибольшего значения, как функция многих переменных $\theta_1,\theta_2,..., \theta_k$, в области Θ. Так как функция $\ln L(x,\theta)$ монотонна, то максимумы функции $L(x,\theta)$ совпадают с максимумами функции $\ln L(x,\theta)$, которая более проста для исследования на экстремумы.

Для нахождения оценок максимального правдоподобия необходимо решить *уравнения правдоподобия*:

$$\frac{\partial L(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i}=0,\quad i=1,2,\ \ldots,k. \qquad (2.10)$$

или

$$\frac{\partial lnL(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i}=0, \qquad (2.11)$$

Так как частные производные от логарифической функции правдоподобия определятся по формулам:

$$\frac{\partial lnL(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i}=\frac{1}{\theta_i}\cdot\frac{\partial L(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i},$$

то равенство (2.11) запишется:

$$\frac{\partial lnL(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i}=\frac{1}{\theta_i}\cdot\frac{\partial L(\theta_1,\theta_2,\ldots,\ \theta_k)}{\partial\theta_i}=0. \qquad (2.11')$$

Пример 18. Пусть $(x_1, x_2, \ldots, x_n)$ - выборка из распределения Пуассона с неизвестным параметром λ. Найти оценку параметра λ по методу максимального правдоподобия .

Решение. Для случайной величины, распределенной по *закону Пуассона*, функция плотности имеет вид:

$$P(X=m)=\frac{\lambda^m e^{-\lambda}}{m!}. \qquad (2.12)$$

Тогда логарифическая функция правдоподобия запишется:

$$ln\,L(x,\lambda)=\sum_{i=1}^{n} ln\left(\frac{\lambda^{x_i}e^{-\lambda}}{(x_i)!}\right)=\sum_{i=1}^{n}\left(x_i ln\,\lambda-\lambda-ln\,(x_i)!\right)=ln\,\lambda\sum_{i=1}^{n}x_i-\lambda n-\sum_{i=1}^{n}ln\,(x_i)!\,.$$

Найдем максимум функции, используя методы дифференциального исчисления. Вычислим производную от полученной логарифической функции правдоподобия и приравняем ее к нулю:

$$\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0.$$

Тогда оценка правдоподобия параметра λ определится в виде выборочной средней арифметической

$$\lambda^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x_в}.$$

Пример 19. Найти оценки максимального правдоподобия параметров нормального распределения.

Решение. Пусть изучаемая величина X распределена по *нормальному закону* с математическим ожиданием a и средним квадратическим отклонением σ, плотность распределения которой задается функцией Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi D}} \cdot e^{-\frac{(x-a)^2}{2D}}. \tag{2.13}$$

Функция правдоподобия в этом случае является функцией двух параметров a и σ (или a и D), которая определится как

$$L(a, D) = \frac{1}{\sqrt{2\pi D}} \cdot e^{-\frac{(x_1-a)^2}{2D}} \cdot \frac{1}{\sqrt{2\pi D}} \cdot e^{-\frac{(x_2-a)^2}{2D}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi D}} \cdot e^{-\frac{(x_n-a)^2}{2D}} =$$

$$= \frac{1}{(2\pi D)^{n/2}} \cdot e^{-\frac{1}{2D}\cdot\sum\limits_{i=1}^{n}(x_i-a)^2}.$$

Воспользуемся логарифмической функцией правдоподобия:

$$\ln L(a, D) = \ln\left[\frac{1}{(2\pi D)^{n/2}} \cdot e^{-\frac{1}{2D}\cdot\sum\limits_{i=1}^{n}(x_i-a)^2}.\right] =$$

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln D - \frac{1}{2D}\cdot\sum_{i=1}^{n}(x_i-a)^2.$$

Теперь составим систему уравнений правдоподобия:

$$\begin{cases} \dfrac{\partial lnL(a,\sigma)}{\partial a} = \dfrac{1}{D} \cdot \sum\limits_{i=1}^{n} \left(x_i - a \right) = 0, \\ \dfrac{\partial lnL(a,\sigma)}{\partial \sigma} = -\dfrac{n}{2D} + \dfrac{1}{2D^2} \cdot \sum\limits_{i=1}^{n} \left(x_i - a \right)^2 = 0 \end{cases}$$

Из первого уравнения системы найдем оценку максимального правдоподобия a^* для параметра a:

$$a^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x_в}.$$

Подставляя это значение во второе уравнение системы и, решая его относительно D, получим:

$$D^* = \frac{\sum\limits_{i=1}^{n}\left(x_i - \overline{x_в}\right)^2}{n} = D_в .$$

Таким образом, оценками математического ожидания и дисперсии нормально распределенной случайной величины являются выборочная средняя арифметическая и выборочная дисперсия соответственно. Однако выборочная дисперсия будет смещенной оценкой для теоретической дисперсии. Устраняя смещение, получают несмещенную оценку дисперсии вида (2.6).

Задачи для самостоятельного решения

31. Проведено 150 независимых наблюдений случайной величины X – срока службы некоторого оборудования:

x_i	2	3	5	7	8	10
n_i	7	19	34	57	24	9

Считая, что случайная величина X распределена по закону с плотностью

$$p(x,a) = a \cdot e^{-ax}, x \geq 0$$

Найти оценку параметра a по методу максимального правдоподобия.

32. Проведено 100 независимых наблюдений случайной величины X – дохода на душу населения (в денежных единицах):

x_i	2	4	6	8	10
n_i	8	13	42	25	12

Считая, что случайная величина X распределена по закону Парето с плотностью

$$p(x,\theta)=\begin{cases}\dfrac{\theta}{2}\cdot\left(\dfrac{2}{x}\right)^{\theta+1}, & x\ge 2,\ (\theta>0),\\ 0, & x<2.\end{cases}$$

Найти по методу максимального правдоподобия оценку параметра θ.

33. Контролируемый размер детали есть случайная величина X, распределенная по закону с плотностью

$$p(x,a)=\frac{2x}{1{,}5\cdot 2\pi}e^{-\frac{(x^2-a)^2}{4{,}5}},\quad x\ge 0.$$

Найти по методу максимального правдоподобия оценку параметра a, если результаты 30 измерений представлены в виде ряда:

x_i	2	7	12	17	22	27	40
n_i	4	10	7	3	2	2	2

34. Проведено 50 независимых наблюдений случайной величины X - коэффициента использования пробега для автотранспорта некоторой области. Результаты наблюдений представлены в виде ряда:

x_i	2	4	5	6	8	10	11
n_i	3	7	30	4	3	2	1

Считая, что случайная величина X распределена по закону с плотностью $p(x,a)=\dfrac{a^2x^2}{2}e^{-ax},\ x\ge 0.$ Найти оценку параметра a, используя метод максимального правдоподобия.

35. Произведено 50 независимых наблюдений случайной величины X - уровня заработной платы одного работающего на данном предприятии (тыс.руб.) Результаты наблюдений представлены в виде ряда:

x_i	20	30	60	80	90
n_i	6	12	16	10	6

Считая, что случайная величина X - распределена по закону с плотностью

$$p(x,a)=\frac{x}{2\pi}e^{-\frac{(x^2-a)^2}{8}}, \quad x\ge 0.$$

Найти по методу максимального правдоподобия оценку параметра a.

36. Результаты 50 независимых наблюдений случайной величины X - фондоотдачи предприятий данной отрасли (млн.руб.) представлены в виде ряда:

x_i	1	2	4	6	7	8
n_i	5	24	8	7	4	2

Полагая, что величина X - распределена по закону с плотностью

$$p(x,a)=\frac{1}{2}a^3x^2e^{-ax}, \quad x>0.$$

Найти по методу максимального правдоподобия оценку параметра a.

37. Найти по методу максимального правдоподобия оценку параметра a распределения с плотностью

$$p(x,a)=\begin{cases} 3x^2a^{-3}e^{-\left(\frac{x}{a}\right)^3}, & x\ge 0, \\ 0, & x<0. \end{cases},$$

если результаты 60 независимых исследований представлены в виде ряда:

x_i	100	120	140	160	170	200
n_i	8	14	28	6	3	1

38. Найти по методу максимального правдоподобия оценку параметра $\theta>0$, если распределение выборки имеет плотность:

$$p(x,\theta)=\begin{cases} \theta x^{\theta-1}, & x\in[0,1], \\ 0, & x\notin[0,1]. \end{cases}$$

39. Найти по методу максимального правдоподобия оценку параметра $\theta>0$, если распределение выборки имеет плотность:

$$p(x,\theta) = \begin{cases} \dfrac{2x}{\theta^2}, & x \in [0,\,\theta], \\ 0, & x \notin [0,\,\theta]. \end{cases}$$

40. Пусть (x_1, x_2, … , x_n) - выборка из двухпараметрического показательного распределения с плотностью

$$p(x,a,b) = \begin{cases} a^{-1} e^{-\frac{(x-b)}{a}}, & x \ge b, \\ 0, & x < b, \end{cases}$$

где $a > 0, b \notin R$. Найти оценку максимального правдоподобия двух параметров a, b .

2.3.2 Метод моментов

Метод моментов был предложен известным английским статистиком К.Пирсоном и отличается простотой получения оценок. *Метод моментов* состоит в том, что оценки неизвестных параметров $\theta_1^*, \theta_2^*, \ldots, \theta_k^*$, определяющих заданное распределение с плотностью распределения $p(x, \theta_1, \theta_2, \ldots, \theta_k)$, находятся как решение системы уравнений, полученной из равенства начальных моментов генеральной совокупности (теоретических начальных моментов) и выборочных или эмпирических начальных моментов, найденных по заданной выборке:

$$\begin{cases} m_1(\theta_1, \theta_2, \ldots, \theta_k) = m_1^*, \\ m_2(\theta_1, \theta_2, \ldots, \theta_k) = m_2^*, \\ \ldots\ldots\ldots\ \ \ldots\ldots\ldots\ \ \ldots\ldots\ldots \\ m_r(\theta_1, \theta_2, \ldots, \theta_k) = m_r^*. \end{cases} , \tag{2.14}$$

где $m_r(\theta_1, \theta_2, ..., \theta_k)$, $r = 1, 2, \ldots, k$ - теоретические начальные моменты, определяемые по формулам:

$$m_r(\theta_1, \theta_2, ..., \theta_k) = M(X^r) = \int\limits_{-\infty}^{-\infty} x^r \cdot p(x, \theta_1, \theta_2, ..., \theta_k) dx, \qquad (2.15)$$

а $m_r^*(\theta_1, \theta_2, ..., \theta_k)$ - эмпирические начальные моменты, вычисляемые по формулам (1.32). Использование начальных моментов необязательно, могут использоваться центральные теоретические моменты и соответствующие им эмпирические моменты.

Пример 20. При тестировании группы студентов есть основание считать, что средний балл X - равномерно распределенная на отрезке [a;b] случайная величина. Результаты исследований представлены в виде ряда:

x_i	1	2	3	4	5
n_i	12	10	9	9	10

Найти методом моментов оценки параметров a и b.

Для *равномерного* на отрезке [a;b] *распределения* плотность вероятности имеет вид:

$$p(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & a < x \le b, \\ 0, x \le a, x \rangle b. \end{cases} \qquad (2.16)$$

Теоретические начальные моменты согласно формуле (2.15) равны:

$$m_1(a, b) = M(X) = \int\limits_a^b x \cdot p(x, a, b) dx = \frac{1}{b-a} \cdot \int\limits_a^b x dx =$$

$$= \frac{1}{b-a} \cdot \frac{x^2}{2}\bigg|_a^b = \frac{b^2 - a^2}{b-a} = \frac{b+a}{2};$$

$$m_2(a, b) = M(X^2) = \int\limits_a^b x^2 \cdot p(x, a, b) dx = \frac{1}{b-a} \cdot \int\limits_a^b x^2 dx =$$

$$= \frac{1}{b-a} \cdot \frac{x^3}{3}\bigg|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

Эмпирические начальные моменты вычисляются по заданной выборке объема $n = 50$:

$$m_1^* = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i = \frac{1}{50}(1\cdot 12 + 2\cdot 10 + 3\cdot 9 + 4\cdot 9 + 5\cdot 10) = 2{,}9;$$

$$m_2^* = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot n_i = \frac{1}{50}\left(1^2\cdot 12 + 2^2\cdot 10 + 3^2\cdot 9 + 4^2\cdot 9 + 5^2\cdot 10\right) = 10{,}54.$$

По методу моментов оценки двух неизвестных параметров a и b находятся из системы уравнений (2.14):

$$\begin{cases} \dfrac{b+a}{2} = 2{,}9 \\ \dfrac{b^2 + ab + a^2}{3} = 10{,}54. \end{cases}$$

Отсюда
$$\begin{cases} b + a = 5{,}8 \\ b^2 + ab + a^2 = 31{,}62. \end{cases}$$

Из решения этой системы уравнений получаем оценки параметров a и b: $a^* \approx 0{,}37$; $b^* \approx 5{,}43$.

Задачи для самостоятельного решения

41. При обследовании продолжительности стоянки автобуса на промежуточной станции получены данные (мин.):

x_i	4	5	6	7	8	9
n_i	12	15	10	12	11	12

Считая, что продолжительность стоянки X подчиняется закону распределения с плотностью

$$p(x; a, b) = \frac{1}{b-a}, \quad a < x \le b.$$

Найти с помощью метода моментов оценки параметров a и b.

42. При обследовании работы билетного кассира получены следующие данные по времени обслуживания пассажиров (мин.):

x_i	2	3	4	5	6	7	8
n_i	1	2	1	3	8	4	1

Считая, что случайная величина X - время обслуживания распределена по так называемому «двойному» закону Пуассона, то есть

$$p(m;a,b)=\frac{1}{2}\frac{a^m}{m!}e^{-a}+\frac{1}{2}\frac{b^m}{m!}e^{-b},\quad m=1,\ 2,\ldots\ .$$

Найти по методу моментов оценки параметров a и b.

43. Имеются следующие данные о размерах основных фондов предприятий (млн.руб.):

x_i	220	221	224	225	233
n_i	4	3	5	4	4

Считая, что случайная величина X - размер основных фондов распределена по «двойному» закону Пуассона, то есть

$$p(m;a,b)=\frac{1}{2}\frac{a^m}{m!}e^{-a}+\frac{1}{2}\frac{b^m}{m!}e^{-b},\quad m=0,1,\ 2,\ldots\ .$$

Найти с помощью метода моментов оценки параметров a и b.

44. Результаты тестирования руководителей ряда предприятий приведены в следующей таблице:

x_i	220	221	224	225	233
n_i	4	3	5	4	4

Считая, что случайная величина X - суммарный балл подчиняется закону распределения с плотностью

$$p(x;a,b)=\frac{1}{b-a},\quad a<x\le b.$$

Найти по методу моментов оценки параметров a и b.

45. Используя метод моментов с пробной функцией $g(x)=x$, оценить параметр сдвига $b\in R$ показательного распределения с плотностью

$$p(x,\ b)=\begin{cases}e^{b-x}, & x\ge b,\\ 0, & x<b.\end{cases}$$

46. Пусть дана выборка из двухпараметрического показательного распределения с плотностью

$$p(x,a,b) = \begin{cases} a^{-1} e^{-\frac{(x-b)}{a}}, & x \geq b, \\ 0, & x < b. \end{cases}$$

Используя метод моментов, оценить параметры масштаба $a > 0$ и сдвига $b \in R$.

47. Случайная величина X (число семян сорняков в пробе зерна) распределена по закону Пуассона. Ниже приведено распределение семян сорняков в $n = 1000$ пробах зерна (в первой строке указано количество x_i сорняков в одной пробе; во второй строке указана частота n_i - число проб, содержащих х семян сорняков):

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

48. Случайная величина X (время работы элемента) имеет показательное распределение. Ниже приведено эмпирическое распределение среднего времени работы $n = 200$ элементов (в первой строке приведено среднее время x_i работы элемента в часах; во второй строке указана частота n_i - количество элементов» проработавших в среднем x часов):

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Найти методом моментов точечную оценку неизвестного параметра показательного распределения.

49. Найти методом моментов по выборке (x_1, x_2, … , x_n) точечные оценки неизвестных параметров a и σ нормального распределения, плотность которого задается формулой (2.13).

50. Случайная величина X распределена по «двойному» закону Пуассона. Ниже приведено эмпирическое распределение числа появлений события в $n = 327$ испытаниях (в первой строке указано число x_i - появлений события; во второй строке приведена частота n_i — количество испытаний, в которых появилось x_i событий):

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	28	47	81	67	53	24	13	8	3	2	1

Найти методом моментов точечные оценки неизвестных параметров «двойного распределения» Пуассона.

2.4 Интервальные оценки

В практических задачах наряду с точечными оценками необходимо указывать интервалы, которые бы с практической достоверностью, т.е. с вероятностью, близкой к единице, покрывали бы истинное неизвестное значение параметра. Такие интервалы называются *доверительными*, или *интервальными оценками*, а вероятность, с которой доверительный интервал покрывает неизвестное значение параметра, называется *доверительной вероятностью* или *надежностью*.

В математической статистике доверительные интервалы используются для определения точности оценки θ^* неизвестного параметра θ, а доверительные вероятности – для определения надежности. Интервальное оценивание особенно необходимо при малом числе наблюдений (n < 30), когда точечная оценка мало надежна.

Пусть по результатам выборки найдена точечная оценка θ^* неизвестного параметра θ. Очевидно, что чем меньше величина $\left|\theta - \theta^*\right|$, тем точнее оценка. Число δ, для которого $\left|\theta - \theta^*\right| < \delta$, характеризует *точность оценки*. Очевидно, чем больше n - объем выборки, тем точнее будет оценка. Однако, нельзя точно определить, при каком значении n будет достигнута заданная точность. Можно говорить только о вероятности, с которой данное неравенство будет выполняться.

Вероятность

$$P\left(\left|\theta - \theta^*\right| < \delta\right) = \gamma \qquad (2.17)$$

называется *доверительной вероятностью* или *надежностью оценки*. Используя определение модуля, формула (2.17) может быть записана:

$$P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma .$$

Полученная формула читается: вероятность того, что интервал $(\theta^* - \delta ; \theta^* + \delta)$, называемый *доверительным интервалом*, заключает в себе (покрывает) неизвестный параметр θ, равна γ. Обычно надежность γ, задается наперед числом, близким к единице (например $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99;\ 0{,}999$). Событие с вероятностью γ считается практически достоверным. Чем меньше длина доверительного интервала, тем точнее оценка.

2.4.1 Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии

Пусть случайная величина X распределена по нормальному закону (2.13), причем дисперсия, а, следовательно, и среднее квадратическое отклонение известно. Требуется оценить неизвестное математическое ожидание a с помощью интервальной оценки, т.е. найти интервал, покрывающий a с надежностью γ. Будем считать, что случайная величина принимает значения x_1, x_2, … , x_n. Если X подчиняется нормальному закону распределения, то и средняя арифметическая ее значений распределена нормально. Найдем математическое ожидание $M(\overline{x}_в)$ и дисперсию средней арифметической $D(\overline{x}_в)$:

$$M\left(\overline{x_в}\right) = M\left(\frac{x_1 + x_2 + \ldots x_n}{n}\right) = \frac{1}{n}\left(M(x_1) + M(x_2 + \ldots + M(x_n)\right) = \frac{n \cdot a}{n} = a,$$

$$D\left(\overline{x_в}\right)=D\left(\frac{x_1+x_2+\ldots x_n}{n}\right)=\frac{1}{n^2}\left(D(x_1)+D(x_2+\ldots+D(x_n))\right)=\frac{n\cdot D(X)}{n^2}=\frac{D(X)}{n}.$$

Так как $D(X)=\sigma^2$, то из последнего равенства имеем:

$$\sigma\left(\overline{x_в}\right)=\sqrt{\frac{D(X)}{n}}=\frac{\sigma}{\sqrt{n}}. \qquad (2.18)$$

Воспользуемся формулой, определяющей вероятность попадания нормально распределенной случайной величины в интервал, симметричный относительно математического ожидания:

$$\mathrm{P}\left(\left|X-a\right|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right). \qquad (2.19)$$

Заменим в формуле (2.19) X на $\overline{x_в}$, σ на $\sigma\left(\overline{x_в}\right)$ из формулы (2.18):

$$\mathrm{P}\left(\left|\overline{x_в}-a\right|<\delta\right)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right). \qquad (2.20)$$

Обозначим

$$t_\gamma=\frac{\delta\sqrt{n}}{\sigma}. \qquad (2.21)$$

Выразим параметр δ, характеризующий точность оценки:

$$\delta=t_\gamma\frac{\sigma}{\sqrt{n}}. \qquad (2.22)$$

Из формулы (2.21) можно получить минимальный объем, который обеспечивает заданную точность δ и заданную надежность γ:

$$n=\frac{t_\gamma^2\sigma^2}{\delta^2}. \qquad (2.23)$$

Так как надежность задается, то равенство (2.20) с учетом (2.22) можно записать:

$$\mathrm{P}\left(\left|\overline{x_в}-a\right|<t_\gamma\frac{\sigma}{\sqrt{n}}\right)=\gamma,$$

или $$P\left(\overline{x_в} - t_\gamma \frac{\sigma}{\sqrt{n}} < a < \overline{x_в} + t_\gamma \frac{\sigma}{\sqrt{n}}\right) = \gamma .$$

Таким образом, получен доверительный интервал для оценки математического ожидания a нормально распределенной случайной величины при известном среднем квадратическом отклонении σ :

$$\overline{x_в} - t_\gamma \frac{\sigma}{\sqrt{n}} < a < \overline{x_в} + t_\gamma \frac{\sigma}{\sqrt{n}} , \qquad (2.24)$$

где $\overline{x}_в$ - выборочная средняя, n - объем выборки. Число t_γ - число определяется из равенства:

$$2\Phi(t_\gamma) = \gamma , \quad \Phi(t_\gamma) = \frac{\gamma}{2} , \qquad (2.25)$$

по таблице функции Лапласа $\Phi(x)$.

Пример 21. Случайные ошибки, даваемые измерительным прибором, распределены нормально с параметром σ = 30 м. Измерения, проведенные 9 раз, дали выборочную среднюю арифметическую $\overline{x}_в$ =1580 м.

1. Построить 95 % -ный доверительный интервал для измеряемой величины.

2. С какой вероятностью полученный результат будет отличаться от истинного значения величины не более, чем на 10 м в ту или иную сторону?

Решение.

1) Истинное значение измеряемой величины равно ее математическому ожиданию a . Известно, что γ = 0,95; n = 9; σ = 30 м. Согласно формуле (2.25) имеем:

$$2\Phi(t) = \gamma = 0{,}95; \ \Phi(t) = 0{,}475 \ .$$

По таблице 2 значений функции Лапласа $\Phi(t)$ найдем t_γ = 1,96.

По формуле (2.24) определим доверительный интервал:

$$1580 - 1,96 \cdot \frac{30}{\sqrt{9}} < a < 1580 + 1,96 \cdot \frac{30}{\sqrt{9}},$$

$$1560,4 < a < 1599,6.$$

2) Задана точность оценки δ =10 м, тогда искомая вероятность определится по формуле (2.20). Найдем число t_γ согласно формуле (2.21):

$$t_\gamma = \frac{\delta\sqrt{n}}{\sigma} = \frac{10\sqrt{9}}{30} = 1,$$

по таблице $\Phi(1) = 0,3413$, тогда $\gamma = 2\Phi(t) = 0,6826$.

2.4.2 Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии

Пусть случайная величина X распределена по нормальному закону (2.13), причем среднее квадратическое отклонение неизвестно. Требуется оценить неизвестное математическое ожидание a с помощью доверительного интервала. По данным выборки $x_1, x_2, \dots, x_n$ построим случайную величину:

$$T = \frac{\overline{x}_в - a}{S/\sqrt{n}}, \tag{2.26}$$

где $\overline{x}_в$ - выборочная средняя, n - объем выборки, S - исправленное выборочное среднее квадратическое отклонение, определяемое по формуле (2.6). Случайная величина T имеет *распределение Стьюдента* с k степенями свободы и плотностью распределения, выражаемую через гамма-функцию $\Gamma(x)$, которая табулирована:

$$p(x, n) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)} \cdot \Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}. \tag{2.27}$$

Распределение Стьюдента определяется одним праметром n - объемом выборки или, что то же самое *числом степеней свободы*:

$$k = n - 1, \tag{2.28}$$

и не зависит от параметров a и σ.

Пользуясь распределением Стьюдента, получим:

$$\mathrm{P}\left(\overline{x_в} - t_\gamma \frac{S}{\sqrt{n}} < a < \overline{x_в} + t_\gamma \frac{S}{\sqrt{n}}\right) = \gamma ,$$

или доверительный интервал для неизвестного математического ожидания a при неизвестной дисперсии σ запишется в виде:

$$\bar{x}_в - t_\gamma \frac{S_в}{\sqrt{n}} < a < \bar{x}_в + t_\gamma \frac{S_в}{\sqrt{n}} . \tag{2.29}$$

Сравнивая полученный доверительный интервал (2.29), который покрывает неизвестное математическое ожидание a с надежностью γ , с доверительным интервалом (2.24), легко убедиться, что доверительные границы для неизвестного параметра a определяются аналогичным образом, за исключением того, что вместо σ ипользуется несмещенная оценка среднего квадратического отклонения $S_в$ - исправленное выборочное среднее квадратическое отклонение (2.6). Коэффициент t_γ в форомуле (2.29) находится из таблицы распределения Стьюдента по заданной доверительной вероятности и числу степеней свободы k согласно формуле (2.28) или объему выборки. Величина

$$\delta = t_\gamma \frac{S_в}{\sqrt{n}} \tag{2.30}$$

также определяет точность оценки, иногда ее называют *предельной погрешностью оценки*.

Из формул (2.24) и (2.29) видно, что увеличение объема выборки n приводит к уменьшению длины доверительного интервала. Очевидно, что увеличение надежности интервальной оценки влечет за собой увеличение ее точности. Если задать точность δ и доверительную вероятность γ, то из соотношения (2.30), можно найти минимальный объем выборки n, который обеспечит заданную точность:

$$n = \frac{t_\gamma^{\ 2} \cdot S_в^{\ 2}}{\delta^2}. \tag{2.31}$$

Пример 22. По данным 16 неизвестных равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\bar{x}_в = 42{,}8$ и исправленное среднее квадратическое отклонение $S_в = 8$. Оценить истинное значение измеряемой величины с надежностью $\gamma = 0{,}999$.

Решение. Из таблицы 7 найдем $t_\gamma = 4{,}07$ и подставим в формулу (2.29):

$$42{,}8 - 4{,}07\frac{8}{\sqrt{16}} < a < 42{,}8 + 4{,}07\frac{8}{\sqrt{16}},$$

$$34{,}66 < a < 50{,}94.$$

Смысл полученного результата: если будет проведено достаточно большое число выборок по 16 измерений этой величины, то в 99,9% из этих выборок доверительный интервал $34{,}66 < a < 50{,}94$ покроет математическое ожидание a измеряемой величины.

Пример 23. Результаты 10 наблюдений над случайной величиной заданы таблицей:

2,485	2,386	2,415	2,471	2,432	2,442	2,491	2,418	2,462	2,425
1	2	3	4	5	6	7	8	9	10

Найти интервальные оценки математического ожидания с надежностью 0,95 и 0,99, предполагая, что случайная величина распределена нормально. Определить минимальное число измерений, которые надо выполнить с надежностью 0,95, чтобы предельная погрешность точечной оценки не превышала 0,01.

Решение. Находим характеристики $\bar{x}_в$, $S_в$ по формулам:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}\cdot 24{,}427 = 2{,}4427 ,$$

$$S_в = \sqrt{\frac{n}{n-1}D_в} = \sqrt{\frac{(x_i - \bar{x}_в)^2}{n-1}} = 0{,}0338 .$$

Из таблицы 7 находим значения коэффициента t_γ при разной доверительной вероятности:

$$n = 10,\ t_\gamma^1 = 2{,}26,\ \gamma = 0{,}95; \qquad n = 10,\ t_\gamma^2 = 3{,}25,\ \gamma = 0{,}99 .$$

Следовательно, согласно формуле (2.30), будем иметь:

$$\delta_1 = 2{,}26\,\frac{0{,}0338}{\sqrt{10}} = 0{,}0253 , \qquad \delta_2 = 3{,}25\,\frac{0{,}0338}{\sqrt{10}} = 0{,}0364 .$$

Подставляя значения полученных точностей в формулу (2.29) доверительного интервала для математического ожидания a найдем:

1) $2{,}417 < a < 2{,}468$,

2) $2{,}406 < a < 2{,}479$.

Для расчета минимального числа измерений, необходимых для определения истинного значения случайной величины (математического ожидания) с погрешностью, не превышающей δ = 0,01, применим формулу (2.31):

$$n = \frac{(2,26)^2 \cdot (0,0338)^2}{(0,01)^2} \approx 56 .$$

Таким образом, при $n \geq 56$ предельная погрешность оценки не превысит 0,01. Следовательно, для того, чтобы предельная точность не превышала $\delta = 0,01$, следует помимо проделанных 10 измерений выполнить еще 46 измерений.

2.4.3 Доверительный интервал для оценки среднего квадратического отклонения нормального распределения

Доверительный интервал для оценки среднего квадратического отклонения σ нормально распределенной случайной величины X с надежностью γ находится из двойных неравенств:

$$S_в (1 - q) < \sigma < S_в (1 + q) \text{ при } q < 1, \tag{2.32}$$

$$0 < \sigma < S_в (1 + q) \text{ при } q > 1,$$

где q находится из таблицы 6 по заданной надежности γ.

Пример 24. Произведено 12 измерений одним прибором (без систематических ошибок) некоторой физической величины, причем исправленное среднее квадратическое отклонение $S_в$ случайных ошибок измерения оказалось равным 0,6. Найти точность прибора с надежностью 0,99. Предполагается, что результаты измерений распределены нормально.

Решение. Точность прибора характеризуется средним квадратическим отклонением случайных ошибок измерений. Поэтому задача сводится к отысканию доверительного интервала, покрывающего σ с заданной надежностью γ = 0,99. По данным γ =0,99 и n =12 по таблице 6 найдем q =0,9. Подставим в формулу (2.32) при $q < 1$, получим: $0,06 < \sigma < 1,14$.

Задачи для самостоятельного решения

51. Выборка из большой партии электроламп содержит 100 ламп. Средняя продолжительность горения лампы выборки оказалась равной 100 часам. Найти с надежностью 0,95 доверительный интервал для средней продолжительности a горения лампы всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы σ = 40 часов. Предполагается, что продолжительность горения ламп распределена нормально.

52. На основании опытной носки 100 пар обуви получено, что средний срок службы испытываемой обуви равен 180 дням, а исправленная дисперсия $s_в^2 = 196$ дн^2. Найти 90%-й доверительный интервал для среднего срока службы обуви.

53. Для изучения случайных ошибок, даваемых измерительным прибором, было произведено 15 измерений одной и той же величины. Выборочная дисперсия результатов измерений оказалась равной $D_в = 20$ мм^2. Найти доверительный интервал для среднего квадратического отклонения ошибок прибора с доверительной вероятностью 0,98.

54. Произведено 10 наблюдений случайной величины X, результаты занесены в таблицу:

№ наблюдений	1	2	3	4	5	6	7	8	9	10
X	7,322	7,336	7,298	7,284	7,413	7,307	7,346	7,337	7,299	7,406

Найти доверительные интервалы для неизвестного математического ожидания и среднего квадратического отклонения (предполагается, что случайная величина распределена нормально), покрывающие эти параметры с вероятностью 0,95 и 0,99.

55. Глубина дорог и подшипника измеряется оптическим глубиномет-

ром, систематическая ошибка которого равна нулю, а случайные ошибки распределены нормально со средним квадратическим отклонением $\sigma = 20$ мк. Сколько надо сделать независимых измерений, чтобы определить глубину с ошибкой не более 15 мк, при доверительной вероятности 90%?

56. Случайная величина имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 4$. Найти доверительные интервалы для оценки математического ожидания *a*, если объем выборки $n = 64$, $\overline{x}_в$ =20,38 и надежность $\gamma = 0{,}95$.

57. Из генеральной совокупности извлечена выборка объема n =130, по которой определен душевой доход на одного члена семьи, представленный в виде интервального вариационного ряда:

Группы семей по месячному доходу на одного члена семьи, руб	До 500	500-1000	1000-1500	1500-2000	Свыше 2000
Число семей	23	36	44	17	101

С доверительной вероятностью $\gamma = 0{,}95$ определить границы, в которых будет находиться средний месячный доход на одного члена семьи.

58. Систематические ошибки измерительного прибора практически равны нулю, а случайные ошибки распределены нормально со средним квадратическим отклонением $\sigma = 20$ м. Необходимо, чтобы абсолютное значение разности между оценкой измеряемой величины и истинным её значением не превосходило 10 м. Определить, с какой вероятностью будет выполняться это условие, если произведено 25 измерений.

59. Задана выборка значений признака *X*, имеющего нормальное распределение:

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Требуется: а) найти выборочную среднюю $\overline{x}_в$ и исправленное среднее квадратическое отклонение $s_в$; б) указать доверительные интервалы, покрывающие с надежностью 0,95 неизвестные математическое ожидание a и среднее квадратическое отклонение σ признака X.

60. Для определения влияния микроэлементов на результаты откорма свиней произведен опыт на 8 группах животных. Рационы отличаются набором и дозами микроэлементов. Результаты откорма свиней в опыте представлены таблицей:

Рацион	Поголовье свиней, гол.	Среднесуточный прирост живой массы, г	Среднее квадратическое отклонение, г
1	90	500	40
2	75	575	45
3	100	610	54
4	50	450	52
5	70	590	65
6	60	650	70
7	110	490	48
8	80	540	62

С доверительной вероятностью 0,95 определить границы, в которых будет находиться среднесуточный прирост свиней по каждому рациону и опыту в целом.

3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

3.1. ПОСТАНОВКА ЗАДАЧИ И ОБЩАЯ СХЕМА ПРОВЕРКИ ГИПОТЕЗ

При решение многих практических задач результаты наблюдений используются для проверки предположений или гипотез относительно тех или иных свойств распределения генеральной совокупности. В частности, при сравнении способов управления производством, выполнении различных технологических процессов, использовании методов обработки по определенным измеряемым признакам и т.д.

Пусть $x_1, x_2, ..., x_n$ - случайная выборка объёма n из некоторой генеральной совокупности (конечной или бесконечной). Каждое значение x_i в этой выборке само является случайной величиной, даже если генеральная совокупность состоит из конечного числа элементов. Необходимо также иметь в виду, что случайная выборка из какой-либо генеральной совокупности должна соответствовать некоторой схеме испытаний, при реализации которой выявляется искомая случайная величина X. При этом полученные в вышеупомянутой серии испытаний значения случайной величины X должны быть *независимыми* и *распределены по тому же закону*, что и сама генеральная совокупность X (хотя бы и приближённо). Мы будем рассматривать гипотезы о виде и параметрах распределения некоторой генеральной совокупности, а также о сравнении выборок из различных генеральных совокупностей.

Статистической гипотезой называется любое предположение относительно вида или параметров генерального распределения. Гипотеза называется *простой*, если она однозначно определяет распределение случайной величины, в противном случае гипотеза *сложная*. Статистическая гипотеза называется *параметрической*, если она содержит утверждение о значении конечного числа параметров распределения, которое считается неизвестным. *Непараметрическая* гипотеза - это утверждение о виде распределения. Например, простой параметрической гипотезой является предположение о том, что наблюдаемая случайная величина имеет нормальное распределение с параметрами 0 и 1.

Та гипотеза, относительно которой ведется проверка, называется *основной* или *нулевой* и обозначается H_0. Наряду с гипотезой H_0 рассматривается *конкурирующая* или *альтернативная* гипотеза H_1, которая должна быть принята в случае отклонения H_0, т.е. H_0 и H_1 - две взаимно исключающие гипотезы. В качестве *базисного предположения* принимается утверждение о справедливости одной их этих гипотез. Отметим, что для одной основной гипотезы может быть выдвинуты несколько альтернативных

Так, например, пусть случайная величина X имеет нормальное распределение со средним a и дисперсией σ^2. Рассмотрим основную гипотезу: $H_0: a = 0,\ \sigma^2 = 1$. В качестве альтернативных могут быть выдвинуты такие гипотезы: 1). H_1: $a = 0,\ \sigma^2 = 2$; 2) H_1: $a \neq 0,\ \sigma^2 = 1$. Рассмотрим их подробнее.

1. Альтернативная гипотеза H_1 по структуре такая же, как и основная. Базисное предположение в этом случае состоит в том, что случайная величина имеет нормальный закон распределения $N\left(0, \sigma^2\right)$, причем значение дисперсии либо 1, либо 2.

2. Альтернативная гипотеза H_1 более сложная, т.к. a может принимать различные значения. Базисное предположение состоит в том, что генеральное распределение имеет вид $N(a,1)$, причем значение a неизвестно. Гипотеза такого вида называется двусторонней.

Можно было бы выдвинуть альтернативные гипотезы - H_1: $a < 0$ (левосторонняя гипотеза); или H_1: $a > 0$ (правосторонняя гипотеза).

Критерием проверки гипотез или *статистическим критерием* называется правило, по которому решают, принять или отклонить нулевую гипотезу H_0 (соответственно, отклонить или принять альтернативную гипотезу H_1). Решение здесь выносится в зависимости от значения специальным образом случайной величины, называемой *статистикой критерия* или просто *критерием*, распределение которой известно и затабулировано. Критерии значимости подразделяются на три типа:

1. Критерии значимости, которые служат для проверки гипотез о параметрах распределений генеральной совокупности (чаще всего нормального распределения). Эти *критерии* называются *параметрическими*.
2. Критерии, которые для проверки гипотез не используют предположений о распределении генеральной совокупности. Эти *критерии* не требуют знаний параметров распределения, поэтому называются *непараметрическими.*
3. Особую группу критериев составляют *критерии согласия*, служа-

щие для проверки гипотез о согласии распределения генеральной совокупности, из которой получена выборка, с ранее принятой теоретической моделью (чаще всего нормальным распределением).

Множество значений критерия разбивается на две непересекающиеся области: *область принятия нулевой гипотезы*, при попадании в которую нулевая гипотеза принимается, и *критическую область*, при попадании в которую нулевая гипотеза отвергается (рис. 14).

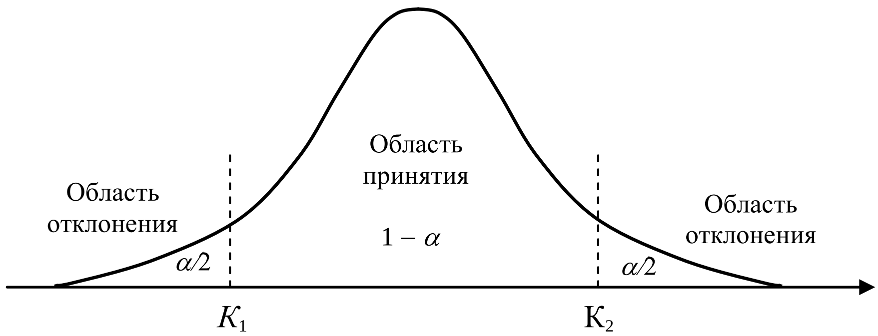
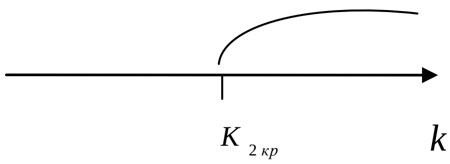


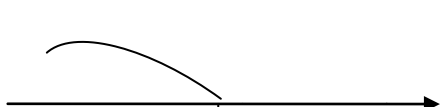
Рисунок 14 - Область отклонения и принятия гипотезы

Критической точкой $K_{кр}$ или *квантилем* называют значение критерия, отделяющее критическую область от области принятия гипотезы. Для нахождения критической точки задаются *уровнем значимости* α - вероятностью попадания случайной величины в критическую область. Различают *односторонние* (*правосторонняя* и *левосторонняя*) и *двухсторонняя* критические области. Вид критической области определяется по виду альтернативной гипотезы: если альтернативная гипотеза имеет вид $H_1: \theta > \theta_0$, то область *правосторонняя* (рис. 15); если $H_1: \theta < \theta_0$ – *левосторонняя* (рис. 16); если $H_1: \theta = \theta_0$ – *двусторонняя* (рис. 17).



$P(k > K_{2кр}) = \alpha$

Рисунок 15 – Правосторонняя критическая область



$P(k < K_{1кр}) = \alpha$

$K_{1кр}$ k

Рисунок 16 – Левостороння критическая область

$P(k < K_{1кр}$ и $k > K_{2кр}) = \alpha$

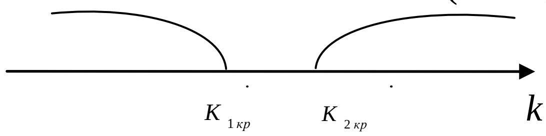


Рисунок 17 – Двухстороння критическая область

Последнее равенство в практических задачах разделяют на два, считая вероятности попадания в правую и левую критические подобласти равными $\alpha/2$. В этом случае

$$P(k < K_{1кр}) = P(k > K_{2кр}) = \alpha/2. \qquad (3.1)$$

В процессе проверки гипотезы можно прийти либо к правильному решению, либо совершить одну из двух ошибок: отвергнуть гипотезу H_o, когда она верна (*ошибка первого рода*) или принять гипотезу H_o , когда она неверна (*ошибка второго рода*), табл. 3.1 .

Таблица 3.1 – Ошибки первого и второго рода

		Верная гипотеза	
		H_o	H_1
Результат применения гипотезы	H_o	H_0 верно принята	H_1 неверно отвергнута (*ошибка второго рода*)
	H_1	H_0 неверно отвергнута (*ошибка первого рода*)	H_1 верно принята

Вероятность ошибки первого рода при проверке статистических гипотез называют *уровнем значимости* и обычно обозначают α (отсюда название α-errors). Уровень значимости α определяет «размер» критической области.

Чем меньше α , тем меньше вероятность отклонения верной гипотезы H_o, т.е. меньше вероятность совершить ошибку первого рода – отвергнуть верную гипотезу. Вероятность ошибки второго рода обозначается β (отсюда β-errors). Величина $(1 - \beta)$— *мощность критерия*. Таким образом, чем выше мощность, тем меньше вероятность совершить ошибку второго рода. Это позволяет сделать следующее *утверждение*: *при выбранном уровне значимости α для рассматриваемого критерия критическую область следует определять таким образом, чтобы мощность критерия была максимальной*.

Желательно провести проверку таким образом, чтобы свести к минимуму вероятности обоих типов ошибок. Однако при фиксированном объеме выборки одновременно уменьшить α и β невозможно. Единственный путь уменьшения ошибок первого и второго рода – увеличить объем выборки. Последствия совершения ошибок первого и второго рода, как правило, различны и могут привести к далеко неравноценным последствиям.

В статистических тестах обычно приходится идти на компромисс между приемлемым уровнем ошибок первого и второго рода. Зачастую для принятия решения используется пороговое значение, которое может варьироваться с целью, сделать тест более строгим или, наоборот, более мягким. Этим пороговым значением и является уровень значимости. Например, в случае металлодетектора повышение чувствительности прибора приведёт к увеличению риска *ошибки первого рода* (ложная тревога), а понижение чувствительности - к увеличению риска *ошибки второго рода* (пропуск запрещённого предмета).

По своему назначению, и характеру решаемых задач статистические критерии чрезвычайно разнообразны. Однако их объединяет общность логической схемы, по которой они строятся. Коротко эту схему можно описать следующим образом:

1) формулируются основная H_0 и конкурирующая H_1 гипотезы;

2) задается уровень значимости α (обычно принимают α = 0,1; 0,01;0,05;0,001;

3) выбирается критерий проверки гипотезы H_0, который являясь функцией выборки $x_1, x_2, \ldots, x_n$, будет случайной величиной с известным законом распределения вероятностей (обычно один из перечисленных ниже: нормальное распределение; распределение Пирсона – хи-квадрат; t - распределение Стьюдента; F распределение Фишера-Снедекора);

4) из таблиц распределения критерия по заданному уровню значимости α находится критическая точка, которая делит множество допустимых значений критерия на область принятия критерия и критическую область;

5) по данным выборки подсчитывается экспериментальное значения критерия; если окажется, что вычисленное значение критерия принадлежит области принятия нулевой гипотезы, то H_0 следует принять, т.е. считать ее непротиворечивой выборочным данным; противном случае гипотезу H_0 следует отвергнуть.

3.2 Проверка гипотез о параметрах нормального распределения

Пусть $x_1, x_2, \ldots, x_n$ – независимая выборка объёма n извлечена из некоторой генеральной совокупности, распределённой по нормальному закону с параметрами: математическим ожиданием a и дисперсией закону σ^2. Здесь возможны следующие предположения о значениях неизвестных параметров:

- гипотезы о математическом ожидании нормальной генеральной совокупности при а) σ –известном; б) σ – неизвестном;

- гипотезы о неизвестной дисперсии нормальной генеральной совокуп-

№ п/п	Задача	Статистика критерия	Теоретическое распределение

ности при а) известном математическим ожиданием; б) неизвестным математическим ожиданием.

В таблице 3.2 приведена сводка процедур проверки гипотез о средних и дисперсиях для нормального распределения случайной величины. В качестве критических точек взяты квантили соответствующих распределений.

Рассмотрим наиболее часто встречающиеся задачи о проверке согласованности статистических оценок и параметров генеральной совокупности нормального закона распределения.

3.2.1 Сравнение выборочной средней с математическим ожиданием нормальной генеральной совокупности при известной дисперсии

Пусть $x_1, x_2, ..., x_n$ - случайная выборка объёма n из некоторой генеральной совокупности, распределённой по нормальному закону с известной . Необходимо сравнить выборочную среднюю с генеральной средней.

Нулевая гипотеза H_0: $\overline{x}_в = M(X) = a$.

Построим критерий проверки этой гипотезы. Рассмотрим величину:

$$\overline{x}_в = \frac{x_1 + x_2 + ... + x_n}{n}.$$

Таблица 3. 2 – Критерии значимости для проверки гипотез о средних и дисперсиях нормально распределенной генеральной совокупности

1	Сравнение выборочной средней с математическим ожиданием нормальной ГС (генеральной совокупности)	$K_{расч.} = \frac{\bar{x} - a}{\sigma} \cdot \sqrt{n}$	$N(0;1)$ H_0 принимается, если $K_{расч.} \in (-K_{кр}, K_{кр})$
2	Сравнение двух выборочных средних из нормальных ГС с равными n и σ	$K_{расч.} = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{2}} \cdot \sqrt{n}$	- // - // - // - // -
3	Сравнение выборочных средних (две выборки) из нормальных ГС с разными, но заданными σ_x и σ_y	$K_{расч.} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$	- // - // - // - // -
4	Сравнение выборочных средних (две выборки) с большими и независимыми выборками любого распределения	$K_{расч.} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D_в(X)}{m} + \frac{D_в(У)}{n}}}$	Приближенный $N(0;1)$ H_0 принимается, если $K_{расч.} \in (-K_{кр}, K_{кр})$
5	Сравнение выборочной исправленной дисперсии с заданной σ^2 нормальной ГС	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	«хи-квадрат» $k = n - 1$
6	Сравнение выборочного распределения долей признаков с теоретическим распределением	$\chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - np_i\right)^2}{np_i}$	«хи-квадрат» $k = l - 1$
7	Сравнение выборочной гистограммы с плотностью нормального закона	$\chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - np_i\right)^2}{np_i}$	«хи-квадрат» $k = r - 3$
8	Сравнение средних (две малые независимые выборки из нормальных ГС) с неизвестными, но равными дисперсиями	$T = \frac{(\bar{x} - \bar{y})\sqrt{mn(m+n-2)}}{\sqrt{(m+n)[(m-1)s_x^2 + (n-1)s_y^2]}}$	Распределение Стьюдента $k = m + n - 2$
9	Сравнение двух дисперсий нормальных ГС	$F = \frac{s_x^2}{s_y^2}$ $(s_x > s_y)$	Распределение Фишера-Снедекора $k_1 = n_x - 1$ $k_2 = n_y - 1$

Её математическое ожидание $M\left(\bar{x}\right) = a$, дисперсия $D\left(\bar{x}\right) = \frac{\sigma^2}{n}$ (стр.85), следовательно $\sigma\left(\bar{x}\right) = \frac{\sigma}{\sqrt{n}}$ (формула (2.18)).

Введем статистику

$$K_{\text{расч.}} = \frac{\bar{x}_в - a}{\sigma} \cdot \sqrt{n}. \qquad (3.2)$$

Рассмотрим альтернативные гипотезы:

1) $H_1 : \overline{x_в} \neq M(X) \neq a$; 2) $H_1 : \overline{x_в} > a = M(X)$; 3) $H_1 : \overline{x_в} < a = M(X)$.

В первом случае имеем двустороннюю критическую область (двусторонний критерий) и гипотеза $H_о$ принимается, если

$$\left|K_{расч.}\right| < K_{кр.}, \qquad (3.3)$$

где $K_{кр.}$ определяется согласно заданному уровню значимости из равенства:

$$\Phi(K_{кр.}) = (1 - \alpha)/2. \qquad (3.4)$$

с использованием статистической таблицы 2 значений интеграла Лапласа, $\Phi(x)$.

Во втором случае имеем правостороннюю критическую область (правосторонний критерий) и гипотеза $H_о$ принимается, если

$$K_{расч.} < K_{кр.}, \qquad (3.5)$$

где $K_{кр.}$ определяется согласно заданному уровню значимости из равенства:

$$\Phi(K_{кр.}) = (1 - 2\alpha)/2 = 1/2 - \alpha. \qquad (3.6)$$

В третьем случае имеем левостороннюю критическую область (левосторонний критерий) и гипотеза $H_о$ принимается, если

$$K_{расч.} > -K_{кр.}, \qquad (3.7)$$

где $K_{кр.}$ определяется по равенству (3.6).

Пример 25. Рост абитуриентов среди поступающих в Иркутский аграрный университет распределён по нормальному закону с математическим ожиданием a = 181 см и среднеквадратическим отклонением σ = 3 см. Для выдачи медицинских справок об основных физиологических показателях были случайно отобраны 9 абитуриентов, полученные данные об их росте приведены в следующей таблице:

№ п/п	1	2	3	4	5	6	7	8	9
X (рост, см)	185,5	180,3	182,7	177,7	178,8	181,9	174,2	180,7	180,7

Проверить гипотезу о равенстве средней по выборке и математического ожидания по этому показателю у обследованных абитуриентов при уровне значимости $\alpha = 0{,}1$.

Решение. Вычислим среднее значение по выборке :

$$\overline{x}_в = \frac{x_1 + x_2 + ... + x_n}{n} =$$

$$= \frac{185{,}5 + 180{,}3 + 182{,}7 + 177{,}7 + 178{,}8 + 181{,}9 + 174{,}2 + 180{,}7 + 180{,}7}{9} = 182{,}5.$$

Применяя формулу (3.2), вычислим расчетное значение критерия:

$$K_{расч.} = \frac{182{,}5 - 181}{3} \cdot \sqrt{9} = 1{,}5.$$

Пусть альтернативной гипотезой будет $H_1 : \overline{x}_в \neq M(X) \neq 181$.

Используя уравнение (3.4), найдем значение функции Лапласа:

$$\Phi(K_{кр.}) = (1 - 0{,}1)/2 = 0{,}45,$$

и по таблице 2 значений функции Лапласа определим критическое значение $K_{кр.} = 1{,}65$. Поскольку выполняется неравенство (3.3) $1{,}5 < 1{,}65$, то нулевая гипотеза $H_о$ принимается.

Пример 26. Техническая норма предусматривает в среднем 40 с. на выполнение определенной технологической операции на конвейере по производству часов. От работниц поступили жалобы, что они в действительности затрачивают на эту операцию больше времени. Для проверки жалобы проведены хронометрические измерения времени выполнения операции у 36 работниц, и получено среднее время выполнения этой операции $\overline{x} = 42$ с. Можно ли по имеющимся данным на уровне значимости $\alpha = 0{,}01$ отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, если известно, что среднее квадратическое отклонение генеральной совокупности $\sigma = 3{,}5$ с?

Решение. Сформулируем основную и альтернативную гипотезы.

$H_0: a = 40$ – неизвестное генеральное среднее равно заданному значению (время выполнения технологической операции соответствует норме).

$H_1: a > 40$ – время выполнения технологической операции больше нормы. По условию задачи $\alpha = 0{,}01$, $n = 36 > 30$, $\sigma = 3{,}5$, значит, воспользуемся статистикой (3.2). Ее расчетное значение равно:

$$K_{расч.} = \frac{\bar{x} - a}{\sigma} \cdot \sqrt{n} = \frac{42 - 40}{3{,}5} \cdot \sqrt{36} = \frac{12}{3{,}5} = 3{,}43 .$$

Так как альтернативная гипотеза правосторонняя, то и критическая область тоже правосторонняя. Из равенства (3.6) определим значение функции Лапласа:

$$\Phi(K_{кр}) = \frac{1}{2} - \alpha = 0{,}5 - 0{,}01 = 0{,}49 .$$

По таблице 2 значений функции Лапласа $K_{кр} = 2{,}33$. Получили, что наблюдаемое значение $K_{расч.} > K_{кр}$, т.е. попадает в критическую область (рис. 15). Следовательно, на данном уровне значимости основная гипотеза отвергается в пользу альтернативной.

Уровень значимости характеризует надежность нашего утверждения: более чем с 99% надежностью можно утверждать, что среднее время выполнения технологической операции превышает норму. Следовательно, жалобы работниц обоснованы. Таким образом, можно утверждать, что среднее время выполнения технологической операции превышает норму.

3.2.2 Сравнение выборочной средней с математическим ожиданием нормальной генеральной совокупности при неизвестной дисперсии

Пусть случайная величина X распределена по нормальному закону (2.13), причем среднее квадратическое отклонение неизвестно. Произведена случайная выборка $x_1, x_2, ..., x_n$ объёма n из генеральной значений случайной величины. Требуется сравнить выборочную среднюю с генеральной средней

Рассмотрим нулевую (основную) и альтернативные гипотезы:

Нулевая гипотеза $H_0 : \overline{x_в} = M(X) = a$. Альтернативные гипотезы:

1) $H_1 : \overline{x_в} \neq M(X) \neq a$; 2) $H_1 : \overline{x_в} > a = M(X)$; 3) $H_1 : \overline{x_в} < a = M(X)$.

Статистика (критерий), по которой будем осуществлять проверку гипотез, имеет вид:

$$K_{расч.} = t_{расч.} = \frac{\overline{x}_в - a}{S} \cdot \sqrt{n} \, , \tag{3.8}$$

где S – исправленная выборочная дисперсия, вычисляемая по формуле (2.6). $K_{кр.} = t_{кр.}$ при заданном уровне значимости α (или надежности $\gamma = 1 - \alpha$) и объеме выборки n будем искать из таблицы 4 – Критические точки распределения Стьюдента, определив число степеней свободы по формуле (2. 28). Нулевая гипотеза будет принята в случае двусторонней критической области, если выполняется неравенство:

$$\left| t_{расч.} \right| < t_{кр.} \text{ или } -t_{кр.} < t_{расч.} < t_{кр.} \tag{3.9}$$

В случае односторонних критических областей (правосторонней и левосторонней) основная гипотеза принимается при выполнении соответствующих неравенств:

$$t_{расч.} < t_{кр.} \tag{3.10}$$

$$t_{расч.} > - t_{кр.} \tag{3.11}$$

Пример 27. Из нормальной генеральной совокупности сельскохозяйственных предприятий, рассматриваемых по показателю урожайности пшеницы, с неизвестной дисперсией и генеральной средней $a = 38,1$, извлечена выборка объема $n = 50$. По ней найдена выборочная средняя арифметическая $\bar{x}_в = 42$ и исправленное среднее квадратическое отклонение $S = 9{,}4$. Требуется при уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу H_0: $\bar{x}_в = a = 38{,}1$ при конкурирующей гипотезе $H_1: M(X) > 38{,}1$.

По условию конкурирующая гипотеза с правосторолнней критической областю. Найдем критическую точку из равенства (3.8):

$$K_{\text{расч.}} = t_{\text{расч.}} = \frac{42 - 38{,}1}{9{,}4} \cdot \sqrt{50} = 2{,}93 .$$

По таблице 7 определим критическую точку $t_{\text{кр.}}(0{,}05;\ 49) = 2{,}009$. На основании невыполнения неравенства (3.10) у нас $t_{\text{расч.}} > t_{\text{кр}}$, т.е. 2,93>2,009. Делаем вывод об отклонении нулевой гипотезы, т.е. выборочная и гипотетическая генеральная средняя статистически различаются значимо.

3.2.3 Сравнение нормальных генеральных средних по разным выборкам при известных дисперсиях

Пусть $x_1, x_2, ..., x_n$ и $y_1, y_2, ..., y_n$ – выборки одного объёмов n_1 и n_2 из нормальных распределений $N(a_x, \sigma^2)$ и $N(a_y, \sigma^2)$ соответственно, причем значение σ известно.

Далее будем считать, что случайные величины X и Y независимы. В этих предположениях проверим нулевую гипотезу $H_0 : a_x = a_y$. Построим критерий проверки $K_{\text{расч.}}$ этой гипотезы:

$$K_{расч.} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} . \tag{3.12}$$

Рассмотрим нулевую и альтернативные гипотезы:

Нулевая гипотеза H_0: $M(X) = M(У)$. Альтернативные гипотезы:

1) $H_1 : M(X) \neq M(У)$; 2) $H_1 : M(X) > M(У)$; 3) $H_1 : M(X) < M(У)$.

При заданном уровню значимости $K_{кр.}$ в случае двусторонней критической области при первой альтернативной гипотезе определяется из равенства (3.4); в случае одностронних критических областей (правосторонней, альтернативная гипотеза –вторая и левосторонней, альтернативная гипотеза - третья) $K_{кр.}$ находится согласно формуле (3.6) с использованием таблицы 2 значений функции Лапласа.

Если нулевая гипотеза верна, то выполняются соответственно неравенства (3.3), (3.5) и (3.7). А вновь полученная случайная величина (3.12) имеет стандартное нормальное распределение $N(0,1)$.

Пример 28. Экономический анализ производительности труда предприятий отрасли позволил выдвинуть гипотезу о наличии двух типов предприятий с различной средней величиной показателя производительности труда. Выборочное обследование 42 предприятий первой группы показало: средняя производительность труда $\bar{x} = 119$ деталей. На 35 предприятиях второй группы средняя производительность оказалась $\bar{y} = 107$ деталей. Генеральные дисперсии известны: $\sigma_1^2(X) = 126{,}9$; $\sigma_2^2(Y) = 136{,}1$. Считая, что выборки произведены из нормально распределенных совокупностей X и Y, на уровне значимости $\alpha = 0{,}05$ проверить, случайно ли полученное различие средних показателей производительности труда в группах или же имеются два типа предприятий с различной средней величиной производительности труда.

Решение. Сформулируем основную и альтернативную гипотезы. $H_0 : a_1 = a_2$ или $H_0 : \overline{X} = \overline{Y}$ – генеральные средние двух нормально распределенных совокупностей равны (предприятия двух обследованных групп относятся к одному типу предприятий; выборочные средние отличаются незначимо; средняя производительность в двух группах одинакова).

$H_1 : a_1 \neq a_2$ – генеральные средние различны (предприятия двух групп относятся к разному типу предприятий; средняя производительность труда в двух группах различна). Так как генеральные дисперсии известны, то для проверки гипотезы используем статистику, имеющую нормальное стандартное распределение (3.12). Ее расчетное значение равно:

$$K_{\text{расч.}} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{119 - 107}{\sqrt{\frac{126,9}{42} + \frac{136,1}{35}}} = 4,57 .$$

Альтернативная гипотеза двусторонняя, поэтому критическое значение $K_{кр}$ находим по таблице значений функции Лапласа из соотношения:

$$K_{\text{кр.}} = 1{,}96 \text{ и } -K_{\text{кр.}} = -1{,}96.$$

Область допустимых значений основной гипотезы $(-1{,}96;\ 1{,}96)$. Наблюдаемое (расчетное) значение $K_{набл} = 4{,}57$ лежит за пределами этого интервала и не является допустимым (принадлежит критической области) на заданном уровне значимости. Основная гипотеза отвергается в пользу альтернативной: полученное различие средних показателей производительности труда в группах неслучайно, имеется два типа предприятий с различной средней производительностью труда.

Пример 29. Количество продаж молока по неделям (в тыс. литров), реализуемого в супермаркетах "Просто продукты" (ПП) и «Крестовский» (К), заданы в следующих таблицах:

	1	2	3	4	5	6	7	8	9
ПП	15,5	10,3	12,7	7,7	8,8	11,9	4,2	4,2	10,7

	1	2	3	4	5	6	7	8	9
К	10,8	11,1	13,6	12,5	13.7	13.7	12,4	13.7	8,5

Проверить гипотезу H_0 о равенстве математических ожиданий при альтернативной гипотезе, что $H_1: M(X) < M(Y)$. Предполагается, что для этих супермаркетов стандартные отклонения продаж молока известны и равны $\sigma = 2$. Зададим уровень значимости $\alpha = 0{,}1$.

Решение. Применив смещение обеих случайных величин на $x_0 = 10$, т.е. введя переменные $u = x - 10$, $v = y - 10$, составим расчетные таблицы для новых переменных:

№	1	2	3	4	5	6	7	8	9	Σ
u	5,5	0,3	2,7	-2,3	-1,2	1,9	-5,8	-5,8	0,7	-4

№	1	2	3	4	5	6	7	8	9	Σ
v	0,8	1,1	3,6	2,5	3,7	3,7	2,4	3,7	-1,5	20,0

Последовательно получим:

$$\bar{u} = -\frac{4}{9} \Rightarrow \bar{x} = \frac{86}{9} \cong 9{,}556, \quad \bar{v} = \frac{20}{9} \approx 2{,}222 \Rightarrow \bar{y} = \frac{110}{9} \cong 12{,}222.$$

Вычислим статистику $K_{\text{расч.}}$, применив формулу (3.12). Учитывая, что объемы выборок одинаковы и равны 9, получим:

$$K_{\text{расч.}} = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{2}} \cdot \sqrt{n} = \frac{9{,}556 - 12{,}222}{2\sqrt{2}} \cdot \sqrt{9} = -2{,}86.$$

Из уравнения для левостронней критической области имеем:

$$\Phi(K_{\text{кр.}}) = (1 - 2\alpha)/2 = 1/2 - \alpha = 0{,}5 - 0{,}1 = 0{,}4.$$

По таблице 2 значений функции Лапласа находим критическое значение $K_{\text{кр}} = 0{,}16$. Неравенство (3.11) для левостронней критической области не выполняется $t_{\text{расч.}} > - t_{\text{кр}}$, у нас $-2{,}86 < -0{,}16$ и гипотеза H_0 отвергается. Таким образом отличие средних продаж молока в этих супермаркетах значимо.

3.2.4 Сравнение нормальных генеральных средних по разным выборкам при неизвестных дисперсиях

Пусть по двум независимым выборкам объема n и m извлеченным из нормальных генеральных совокупностей $X \sim N(a_1, \sigma_1)$ и $Y \sim N(a_2, \sigma_2)$ требуется сравнить генеральные средние, т.е. основная гипотеза имеет вид:

$$H_0 : a_1 = a_2 \left(H_0 : \overline{X} = \overline{Y}\right).$$

Если обе генеральные дисперсии неизвестны, то применяется статистика, имеющая распределение Стьюдента:

$$K_{\text{расч.}} = t_{\text{расч.}} = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)\cdot s_x^2 + (m-1)\cdot s_y^2}} \cdot \sqrt{\frac{m \cdot n \cdot (m+n-2)}{m+n}}, \qquad (3.13)$$

где s_x^2 и s_y^2 – исправленные дисперсии каждой выборки, определяемые по формуле (2.6). Для проверки приемлемости гипотезы H_0 по заданному уровню значимости α (или надежности $\gamma = 1 - \alpha$) и числу степеней свободы

$$k = n + m - 2 \qquad (3.14)$$

из таблицы 4 – Критические точки распределения Стьюдента находим значение $K_{\text{кр.}} = t_{\text{кр}}$. Далее проверка аналогична изложенной методике в разделе 3.2.2. Вывод о приемлемости или об отвержении гипотезы H_0 осуществляется на основании формул (3.9) – (3.11).

Пример 30. Для двух нормальных случайных величин X и Y получены выборки

x_i	3,7	4,9	5,6	6,1
n_i	2	4	2	3

y_i	3,4	4,2	5,1
n_i	4	3	2

Проверить гипотезу о равенстве математических ожиданий при уровне значимости $\alpha = 0{,}05$.

Решение. Объемы выборок соответственно равны $n = 11$, $m = 9$. Для каждой выборки подсчитываем выборочные средние: $\overline{x} = 5{,}1363$; $\overline{y} = 4{,}0444$. Формулируем основную гипотезу о равенстве математических ожиданий H_0: $a_1 = a_2$ или $M(X) = M(У)$. Альтернативная гипотеза H_1: $M(X) \neq M(У)$. Для использования статистики (3.13) подсчитаем исправленные дисперсии для обеих выборок :

$$\overline{x^2} = 27{,}0700\ ;\ \overline{y^2} = 16{,}7978\ ;\ D_x = 0{,}6884\ ;\ D_y = 0{,}4403\ ;$$

$$s_x^2 = 0{,}7572\ ;\quad s_y^2 = 0{,}4953\ .$$

Подсчитываем расчетное (наблюдаемое) значение критерия:

$$t_{\text{расч.}} = \frac{5{,}1363 - 4{,}0444}{\sqrt{10 \cdot 0{,}7572 + 8 \cdot 0{,}4953}} \cdot \sqrt{\frac{9 \cdot 11 \cdot (9 + 11 - 2)}{9 + 11}} = \frac{1{,}0919}{3{,}3962} \cdot 9{,}4393 = 3{,}0348\ .$$

По таблице 4 критических точек распределения Стьюдента находим:

$$t_{\text{кр.}}(\alpha\ , n+m - 2) = t_{\text{кр.}}(0{,}05\,; 18) = 2{,}10\ .$$

Сравниваем расчетное значение с критическим : $t_{\text{расч.}} > t_{\text{кр.}}$. Это значит, что выборочные средние различаются значимо. Гипотезу о равенстве математических ожиданий генеральных совокупностей следует отвергнуть.

3.2.5 Проверка гипотезы о равенстве математических ожиданий при равных неизвестных дисперсиях

Пусть теперь для тех же выборок обе генеральные дисперсии неизвестны, но одинаковы, т.е. $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Рассмотрим их выборочные средние и исправленные дисперсии:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n},\ s_X^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1},$$

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_m}{m},\ s_Y^2 = \frac{\sum_{j=1}^{m} (y_j - \bar{y})^2}{m - 1}.$$

Известно, что $\bar{x} \sim N\left(a_X, \dfrac{\sigma}{\sqrt{n}}\right)$, $\bar{y} \sim N\left(a_Y, \dfrac{\sigma}{\sqrt{m}}\right)$ - нормально распределенные случайные величины. Величины s_X^2, s_Y^2 подчинены χ^2–распределению соответственно с ($n-1$) и ($m-1$) степенями свободы. Поскольку случайные величины X и Y независимы, то величина

$$U = \frac{(m-1)s_X^2}{\sigma^2} + \frac{(n-1)s_Y^2}{\sigma^2}$$

имеет χ^2–распределение с (m+n–2) степенями свободы, а величина $\bar{x} - \bar{y}$ распределена нормально:

$$\bar{x} - \bar{y} \sim N\left(a_X - a_Y, \sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}\right) = N\left(a_X - a_Y, \sigma\sqrt{\frac{1}{m} + \frac{1}{n}}\right).$$

Поэтому нормализованная случайная величина

$$V = \frac{(\bar{x} - \bar{y}) - (a_X - a_Y)}{\sigma\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}}. \qquad (3.15)$$

имеет стандартное нормальное распределение $N(0, 1)$, а отношение

$$\frac{V}{\sqrt{U/(m+n-2)}} = \frac{(\bar{x} - \bar{y}) - (a_X - a_Y)}{\sigma\sqrt{\dfrac{1}{m} + \dfrac{1}{n}}\sqrt{\left(\dfrac{(m-1)s_X^2}{\sigma^2} + \dfrac{(n-1)s_Y^2}{\sigma^2}\right)\dfrac{1}{m+n-2}}}$$

имеет распределение Стьюдента с (m+n–2) степенями свободы. Таким образом, если гипотеза H_0: $a_x = a_y$ верна, то величина

$$K_{\text{расч.}} = t_{\text{расч.}} = T = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\dfrac{1}{m} + \dfrac{1}{n}\right)\dfrac{(m-1)s_X^2 + (n-1)s_Y^2}{m+n-2}}} \qquad (3.16)$$

имеет распределение Стьюдента с (m+n–2) степенями свободы. Эта величина используется в качестве критерия для проверки гипотезы H_0.

В качестве альтернативных гипотез могут выступать:

1) $H_1: a_x \neq a_y$; 2) $H_1: a_x > a_y$; 3) $H_1: a_x < a_y$.

Зададим уровень значимости α и определим число степеней свободы по формуле (3.14). Воспользуемся таблицей 4 – Критические точки распределения Стьюдента для нахождения $K_{кр.} = t_{кр.}$

Нулевая гипотеза будет принята в случае двусторонней критической области, если выполняется неравенство (3.9). В случае односторонних критических областей (правосторонней и левосторонней) основная гипотеза принимается при выполнении соответствующих неравенств (3.10) и (3.11).

Пример 31. Для того чтобы проверить технологию изготовления нового кваса, периодически отбирают случайным образом 10 бутылок и находят концентрацию сахара. В следующей таблице приведены данные по стандартной партии (X) и по очередной проверяемой (Y).

№	1	2	3	4	5	6	7	8	9	10
X	11,93	9,43	10,43	8,93	9,93	9,43	7,43	8,93	8,43	9,93
Y	10,24	9,74	10,74	8,24	11,24	9,74	8,74	11,24	9,74	9,24

При уровне значимости $\alpha = 0{,}1$ проверить нулевую гипотезу H_0: $a_x = a_y$ при конкурирующей гипотезе H_1: $a_x \neq a_y$.

Решение. Введём новые переменные $u = x – 9{,}43$; $v = y - 9{,}24$. Составим расчетные таблицы для новых переменных:

№	1	2	3	4	5	6	7	8	9	10	Σ
u	2,5	0	1	-0,5	0,5	0	-2	-0,5	-1	0,5	0,5
u^2	6,5	0	1	0,25	0,25	0	4	0,25	1	0,25	13,5

№	1	2	3	4	5	6	7	8	9	10	Σ
v	1	0,5	1,5	-1	2	0	-0,5	0	0,5	0	4
v^2	1	0,25	2,25	1	4	0	0,25	0	0,25	0	9,5

Вычисление средних значений и стандартных отклонений дают следующие результаты $\bar{u} = 0{,}05 \Rightarrow \bar{x} = 9{,}48$; $\bar{v} = 0{,}4 \Rightarrow \bar{y} = 9{,}64$, $s_x = 1{,}22$,

$s_y = 0{,}89$. Учитывая, что в данном примере $n = m = 10$, мы можем вычислить критерий $K_{расч.} = t_{расч.}$ по формуле (3.16):

$$K_{расч.} = t_{расч.} = \frac{9{,}48 - 9{,}64}{\sqrt{\left(\frac{1}{10} + \frac{1}{10}\right)\frac{9 \cdot 1{,}22 + 9 \cdot 0{,}89}{10 + 10 - 2}}} = -0{,}335.$$

По данному значению α и по числу степеней свободы $(n+m-2) = 18$ находим по таблице критическое значение $t_{кр.}(0{,}1;18) = 1{,}73$. Следовательно, область принятия имеет вид $(-1{,}73;\ 1{,}73)$. Поскольку найденное значение $t_{расч.}$ попадает в область принятия, то гипотеза H_0 принимается.

3.2.6 Сравнение дисперсий двух нормальных распределений

Пусть случайные величины X и Y распределены по нормальному закону. По выборкам значений X объема n и Y объема m требуется проверить нулевую гипотезу H_0 о равенстве дисперсий этих случайных величин: $\sigma^2(X) = \sigma^2(Y)$.

Как обычно предположим вначале, что математические ожидания X и Y известны и рассмотрим случайную величину:

$$K_{расч.} = F_{расч.} = \frac{s^2(X)}{s^2(Y)},\ s_x > s_y. \tag{3.17}$$

Указанная случайная величина распределена по закону Фишера-Снедекора со степенями свободы:

$$k_1 = n - 1 \text{ и } k_2 = m - 1. \tag{3.18}$$

Альтернативными гипотезами являются: H_1: $\sigma^2(X) \neq \sigma^2(Y)$ (двусторонняя критическая область) и H_1: $\sigma^2(X) > \sigma^2(Y)$ (правосторонняя критическая область. Зададим уровень значимости α и определим число степеней свободы по формуле (3.18). Для определения $K_{кр.} = F_{кр.}$ воспользуемся таблицей 8 – Критические точки распределения Фишера-Снедекора. Нулевая гипотеза будет принята в случае двусторонней критической области, если выполняется неравенство:

$$F_{\text{расч.}} < F_{\text{кр.}}(k_1, k_2, \alpha/2). \tag{3.19}$$

В случае правосторонней основная гипотеза принимается при выполнении неравенства:

$$F_{\text{расч.}} < F_{\text{кр.}}(k_1, k_2, \alpha). \tag{3.20}$$

Пример 32. Двумя методами проведены измерения одной и той же физической величины. В первом случае по выборке объема n = 7 получена несмещенная оценка дисперсии $s_1^2 = 15,6$. Для второго метода соответственно m = 6 и $s_2^2 = 14,2$. Можно ли считать, что оба метода обеспечивают одинаковую точность измерений на уровне значимости $\alpha = 0,05$. Предполагается, что результаты измерений распределены нормально и выборки независимы.

Решение. Нулевая гипотеза имеет вид $H_0 : \sigma_1^2 = \sigma_2^2$

Согласно формуле (3.17) расчетное (наблюдаемое) значение статистики равно:

$$K_{\text{расч.}} = \frac{15,6}{14,2} = 1,099 .$$

Границу критической области определим по таблице 8 распределения Фишера на уровне значимости $\alpha = 0,05$ и степеням свободы

$$k_1 = n - 1 = 7 - 1 = 6,\ k_2 = m - 1 = 6 - 1 = 5 :\ K_{кр.} = 4,39 .$$

Наблюдаемое значение статистики не попадает в критическую область, значит нет оснований отвергать нулевую гипотезу. Таким образом, выборочные оценки дисперсий отличаются незначимо, и оба метода обеспечивают одинаковую точность измерений.

3.2.7 Распределение долей признаков

Пусть генеральная совокупность разбита на l непересекающихся классов A_1, A_2, …, доли которых в генеральной совокупности составляют p_1, p_2, … Имеется выборка объема n, и выполняются условия:

$$np_1 \geq 5,\ np_2 \geq 5,\ \ldots\ . \tag{3.21}$$

Обозначим

$$m_i = np_i\ , \tag{3.22}$$

где m_i – число представителей выборки, попадающих в категорию i (*теоретические частоты*);

n_i – ожидаемое число в каждом классе (*эмпирические частоты*), вычисленное согласно нулевой гипотезе.

Выдвигается нулевая гипотеза H_o: $m_i = n_i$ (теоретические и эмпирические частоты достаточно мало отличаются друг от друга). Альтернативными гипотезами являются: 1) H_1: $m_i \neq n_i$; 2) H_1: $m_i > n_i$; 3) H_1: $m_i < n_i$.

Отклонение выборочных значений от ожидаемых вычисляется по формуле:

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{l} \frac{\left(n_i - np_i\right)^2}{np_i} = \sum_{i=1}^{l} \frac{\left(n_i - m_i\right)^2}{m_i}, \tag{3.23}$$

где суммирование производится по всем классам l.

Эта статистика подчиняется распределению «хи-квадрата» χ^2 с q

$$q = l - 1 \tag{3.24}$$

степенями свободы, где l – число классов. Задавая уровень значимости α, нулевая гипотеза принимается, если

$$\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}\left((1 - \alpha / 2),\ q\right) \tag{3.25}$$

в случае двусторонней критической области, где $\chi^2_{\text{кр.}}$ находится по таблице 5 - Критические точки распределения χ^2, и

$$\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}\left(\alpha,\ q\right) \tag{3.26}$$

в случае правосторонней критической области, и

$$\chi^2_{\text{расч.}} > - \chi^2_{\text{кр.}}(\alpha, q) \qquad (3.27)$$

в случае левосторонней критической области.

Пример 33. Данные о различных заболеваниях, определенные на основе сводных данных за год по всем поликлиникам Иркутска, представлены в следующей таблице:

Виды забол.	Грипп	Сердечно-сосудистые	Кишечные	Онкологические	Травматологические	Иммунные
Доля	34%	18%	12%	10%	15%	11%

В некоторой районной поликлинике Иркутска аналогичные данные о частоте заболеваний за тот же период представлены ниже.

Виды забол.	Грипп	Сердечно-сосудистые	Кишечные	Онкологические	Травматологические	Иммунные
n_i	1116	502	341	313	415	413
m_i	1054	558	372	310	465	341

Проверить, соответствует ли распределение заболеваний в данной поликлинике общей структуре заболеваний с уровнем значимости α = 0,01?

*Решение***.** Вычислим общее количество заболеваний за год:

$$n = 1116 + 502 + 341 + 313 + 415 + 413 = 3100.$$

Найдем предполагаемые значения теоретических частот по каждой категории согласно формуле (3.22):

$m_1 = 3100 \cdot 0,34 = 1054$; $m_2 = 3100 \cdot 0,18 = 558$; $m_3 = 3100 \cdot 0,12 = 372$ и т.д.

и внесём их 3-ю строку таблицы. Для проверки нулевой гипотезы вычислим критерий χ^2 по формуле (3.23):

$$\chi^2_{\text{расч.}} = \frac{(1116 - 1054)^2}{1116} + \frac{(502 - 558)^2}{558} + \frac{(341 - 372)^2}{372} + \frac{(313 - 310)^2}{313} +$$

$$+\frac{(415-465)^2}{465}+\frac{(413-341)^2}{341}=32{,}25\ .$$

По статистической таблице 5 согласно заданного уровня значимости α = 0,01 и числу степеней свободы q = 6–1 = 5 определим критическое значение $\chi^2_{\text{крит.}}$. Поскольку $\chi^2_{\text{крит}}$ = 15,1 и $\chi^2_{\text{расч.}} > \chi^2_{\text{кр.}}(\alpha, q)$, то наша гипотеза не принимается. Более подробный анализ слагаемых, из которых состоит сумма наблюдаемых значений критерия, показывает, что основной вклад даёт последнее слагаемое. Районным властям следует обратить внимание на неблагополучную экологическую обстановку, приводящую к росту иммунных заболеваний.

При проверке гипотез о генеральной доле предполагают, что испытания проводятся по схеме Бернулли (независимы, вероятность p появления события A в каждом опыте постоянна). По выборке объема n определяют относительную частоту p^* появления события A:

$$p^*=\frac{m}{n}\ , \tag{3.28}$$

где m – количество появления события A в серии из n испытаний. При проверке гипотез о генеральной доле используется статистика, имеющая при достаточно большом объеме выборки стандартное нормальное распределение (табл. 3.3), критическое значение определяется из статистической таблицы 2 значений функции Лапаса и применяются формулы (3.3) – (3.7).

Таблица 3.3 – Проверка гипотез о генеральной доле

Гипотеза	$H_0: p = p_0$	$H_0: p_1 = p_2$
Предположение	Схема испытаний Бернулли	
Оценки по выборке	$p^*=\frac{m}{n}$	$p_1^*=\frac{m_1}{n_1}$; $p_2^*=\frac{m_2}{n_2}$;

		$\bar{p}=\frac{m_1+m_2}{n_1+n_2}$
Статистика $K_{\text{расч.}}$	$\frac{p^*-p_0}{\sqrt{p_0(1-p_0)}}\cdot\sqrt{n}$	$\frac{p_1^*-p_2^*}{\sqrt{\bar{p}(1-\bar{p})}}\cdot\sqrt{\frac{n_1 n_2}{n_1+n_2}}$
Распределение статистики $K_{\text{расч.}}$	Стандартное нормальное $N(0;1)$	Стандартное нормальное $N(0;1)$

Пример 34. Партия изделий принимается, если вероятность того, что изделие окажется стандартным, составляет не менее 0,97. Среди случайно отобранных 200 изделий проверяемой партии оказалось 193 стандартных. Можно ли на уровне значимости $\alpha = 0{,}02$ принять партию?

Решение. Сформулируем основную и альтернативную гипотезы.

$H_0: p = p_0 = 0{,}97$ – неизвестная генеральная доля p равна заданному значению $p_0 = 0{,}97$. Вероятность того, что деталь окажется стандартной, равна 0,97, т.е. партию изделий можно принять.

$H_1: p < 0{,}97$ – вероятность того, что деталь окажется стандартной, меньше 0,97, т.е. партию изделий нельзя принять.

По условию задачи: $p_0 = 0{,}97$, $n = 200$, $m = 193$, $p^* = \frac{m}{n} = \frac{193}{200} = 0{,}965$.

$$K_{\text{расч.}} = \frac{p^*-p_0}{\sqrt{p_0(1-p_0)}}\cdot\sqrt{n} = \frac{0{,}965-0{,}97}{\sqrt{0{,}97\cdot(1-0{,}97)}}\cdot\sqrt{200} =$$

$$= -0{,}005\cdot 10\cdot\sqrt{\frac{2}{0{,}97\cdot 0{,}03}} = -0{,}05\cdot 8{,}290 = -0{,}4145 .$$

Так как критическая область левосторонняя, критическое значение найдем по статистической таблице 2 значений функций Лапласа из равенства:

$$\Phi(K_{кр}) = \frac{1}{2}-\alpha = 0{,}5-0{,}02 = 0{,}48 \text{, отсюда } K_{кр} = 2{,}05 .$$

Критическая область является интервалом $(-\infty; -2{,}05)$. Расчетное значение $K_{\text{расч.}} = -0{,}4145$ не принадлежит критической области, значит, на данном уровне значимости нет оснований отвергать основную гипотезу и партию изделий можно принять.

Пример 35. Менеджер торговой фирмы должен заключить контракт на поставку изделий с заводом, производящим данные изделия. Таких заводов два. Одним из критериев выбора служит качество, производимой продукции. Для оценки качества сделаны выборки из продукции этих заводов и получены следующие результаты. Среди отобранных 200 изделий первого завода оказалось 20 бракованных, среди 300 изделий второго завода – 15 бракованных. На уровне значимости $\alpha = 0{,}025$ выяснить, имеется ли существенная разница в качестве изготовляемых этими заводами изделий. Т.е. может ли менеджер отдать предпочтение одному из заводов, в виду того, что качество изделий у этого завода лучше?

Решение. Задача на сравнение генеральных долей. Сформулируем основную и альтернативную гипотезы.

$H_0 : p_1 = p_2$ – генеральные доли равны (качество продукции одинаково).

$H_0 : p_1 \neq p_2$ – заводы изготавливают изделия разного качества.

По условию задачи $n_1 = 200$; $m_1 = 20$; $n_2 = 300$; $m_2 = 15$.

$$p_1^* = \frac{m_1}{n_1} = \frac{20}{200} = 0{,}1; \quad p_2^* = \frac{15}{300} = 0{,}05; \quad \vec{p} = \frac{m_1 + m_2}{n_1 + n_2} = \frac{20 + 15}{200 + 300} = \frac{35}{500} = 0{,}07;$$

$$K_{набл} = \frac{p_1^* - p_2^*}{\sqrt{\vec{p}(1 - \vec{p})}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{0{,}1 - 0{,}05}{\sqrt{0{,}07 \cdot (1 - 0{,}07)}} \cdot \sqrt{\frac{200 \cdot 300}{200 + 300}} =$$

$$= 0{,}05 \cdot 100 \cdot \sqrt{\frac{6}{0{,}07 \cdot 0{,}93 \cdot 500}} = 5 \cdot \sqrt{\frac{6}{0{,}35 \cdot 93}} = 5 \cdot 0{,}429 = 2{,}147.$$

Так как критическая область двусторонняя, то критическое значение статистики $K \sim N(0;1)$ найдем по статистической таблице 2 функции Лапласа из соотношения:

$$\Phi\left(K_{кр}\right) = \frac{1 - \alpha}{2} = \frac{1 - 0{,}025}{2} = \frac{0{,}975}{2} = 0{,}4875; \; K_{кр} = 2{,}24.$$

При двусторонней критической области область допустимых значений основной гипотезы имеет вид $(-K_{кр}; K_{кр}) = (-2{,}24; 2{,}24)$. Наблюдаемое значение $K_{расч.} = 2{,}147$ попадает в этот интервал, т.е. на данном уровне значимости нет оснований отвергать основную гипотезу. Заводы изготавливают продукцию одинакового качества. Менеджер отдать предпочтение одному из заводов, в виду того, что качество изделий у этого завода лучше не может.

3.2.8 Сравнение выборочной исправленной дисперсии с заданной дисперсией нормальной генеральной совокупности

Рассмотрим ещё один класс задач, связанный с параметрической проверкой статистических гипотез. в которых применяется критерий χ^2.

Рассматривается выборка, предположительно извлечённая из нормально распределённой генеральной совокупности с заданной дисперсией σ_0^2. Однако случайная величина, которой в данном случае является дисперсия, не подчиняется нормальному закону распределения. В этом случае применяется критерий χ^2.

На практике нулевая гипотеза H_0: $M(S^2) = \sigma_0^2$ проверяется, если нужно проверить точность приборов, методики контроля ритмичности работы и т.д. В качестве критерия проверки нулевой гипотезы принимается случайная величина

$$K_{расч.} = \chi^2_{расч.} = \frac{(n-1)\cdot S^2}{\sigma_0^2}, \qquad (3.29)$$

где S^2 - выборочная исправленная дисперсия. Это случайная величина имеет теоретическое распределение χ^2 с $(n$ - 1)-й степенью свободы. Схема принятия или отвержения нулевой гипотезы аналогична той, которая описана в главе 3.2.8 (формулы (3.24) – (3.27)). По статистической таблице 5 χ^2 по заданному уровню значимости и числу степеней свободы определяется критическое значение $\chi^2_{кр}$. Если $\chi^2_{расч.} > \chi^2_{кр.}$, то нулевую гипотезу отвергают, т.е. принимают двустороннюю альтернативную гипотезу H_1: $S^2 \neq \sigma_0^2$ или односторонние альтернативные гипотезы H_1: $S^2 > \sigma_0^2$ и H_1: $S^2 < \sigma_0^2$. В противном случае можно считать различие исправленной выборочной дисперсии S^2 и гипотетической дисперсии σ_0^2 незначимым.

Пример 36. Ритмичность работы кассира сбербанка по приёму коммунальных платежей определяется дисперсией времени обслуживания клиентов, которая не должна превышать величины D = 1 мин^2. Результаты 30 наблюдений за работой нового кассира приведены в таблице:

Время обслуживания клиента t_i	5,5	6,5	7,5	8,5	9,5	10,5	11,5
Число наблюдений n	2	10	11	3	2	1	1

Требуется проверить нулевую гипотезу о допустимой ритмичности работы новичка при уровне значимости 0,05.

Решение. Введем новую переменную и составим вспомогательную таблицу:

n_i	2	10	11	3	2	1	1	Сумма	Среднее
$x_i = t_i$-7,5	-2	-1	0	1	2	3	4	0	0
x_i^2	4	1	0	1	4	9	16	54	1,8

$$D_в(t) = 1{,}8 \qquad s_t^2 = \frac{30}{29} \cdot 1{,}8 = 1{,}86.$$

Вычислим статистику по формуле (3.29) и применим критерий χ^2:

$$\chi^2_{расч.} = \frac{(n-1)\cdot s_t^2}{\sigma_0^2} = \frac{29 \cdot 1{,}86}{1} = 53{,}94 .$$

$$\chi^2_{кр.}(0{,}05;29) = 42{,}6 , \ \chi^2_{расч.} > \chi^2_{кр.} .$$

Нулевая гипотеза о равенстве дисперсий должна быть отвергнута, следовательно, новый кассир пока ещё не вошёл в требуемый ритм работы.

Пример 37. Точность работы станка-автомата проверяется по дисперсии контролируемого размера деталей, которая не должна превышать $\sigma^2 = 0{,}1$. По выборке из 25 случайно отобранных деталей рассчитаны оценки генерального среднего и генеральной дисперсии, при этом $s^2 = 0{,}2$. На уровне значимости 0,05 проверить, обеспечивает ли станок требуемую точность.

Решение. Сформулируем основную и альтернативную гипотезы:

$H_0: \sigma^2 = 0{,}1$ – станок обеспечивает требуемую точность.

$H_1: \sigma^2 > 0{,}1$ – точность не обеспечивается.

По условию задачи $n = 25 < 30$, $\alpha = 0{,}05$. Так как генеральное среднее неизвестно (оценивается по выборке), то будем находить статистику по формуле (3.29). Ее расчетное значение равно:

$$K_{\text{расч.}} = \frac{(25-1)\cdot 0{,}2}{0{,}1} = 48\,.$$

Критическая область правосторонняя, ее границу $K_{\text{кр.}}$ определим по таблице распределения «хи-квадрат»:

$$K_{\text{кр.}} = \chi^2(n-1;\alpha) = \chi^2(24;\,0{,}05) = 36{,}4\,.$$

Получили $K_{\text{расч.}} > K_{\text{кр}}$., т.е. расчетное (наблюдаемое) значение попадает в критическую область. Значит, основную гипотезу нужно отвергнуть в пользу альтернативной. Станок не обеспечивает требуемую точность и требует наладки.

Рассмотренные примеры применения основных правил проверки статистических гипотез не исчерпывают всех сфер применения указанной теории, являющейся частью более общей теории планирования эксперимента. Дополним вышеперечисленные правила таблицами 3.4 – 3.8, позволяющими легче ориентироваться в многообразных задачах проверки параметрических и непараметрических гипотез.

Таблица 3.4 - Сравнение выборочной средней с генеральной средней (гипотетической) при известной дисперсии

	Большой объем выборки, n		
Генеральная дисперсия известна	$K_{\text{расч.}} = \dfrac{\left(\overline{X} - \overline{X_0}\right) \cdot \sqrt{n}}{\sigma}$ $H_0\text{: } \overline{X} = \overline{X_0}$; H_0 принимается если:		
	Двусторонний критерий	Односторонний критерий	
	$H_1\text{: } \overline{X} \neq \overline{X_0}$	$H_1\text{: } \overline{X} > \overline{X_0}$	$H_1\text{: } \overline{X} < \overline{X_0}$
	$\left\lvert K_{\text{расч.}} \right\rvert < K_{\text{кр., } \alpha/2}$	$K_{\text{расч.}} < K_{\text{кр., } \alpha}$	$K_{\text{расч.}} > -K_{\text{кр., } \alpha}$
	$\Phi\left(K_{\text{кр., } \alpha/2}\right) = \dfrac{1-\alpha}{2}$	$\Phi\left(K_{\text{кр., } \alpha}\right) = \dfrac{1-2\alpha}{2}$	
	$K_{\text{кр.}}$ определяется по таблице значений интеграла Лапласа, $\Phi(x)$		

Таблица 3.5 - Сравнение выборочной средней с генеральной средней (гипотетической) при неизвестной дисперсии

Ге не	Малый объем выборки, n

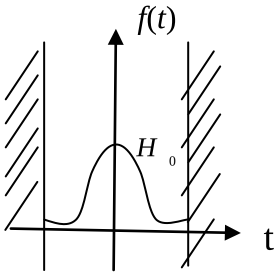
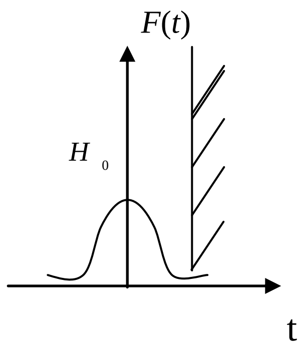
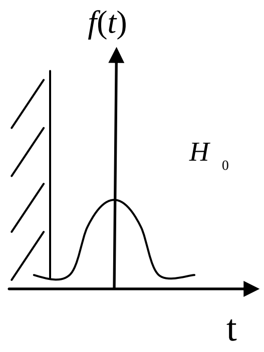
<table>
<tr><td rowspan="6"></td><td colspan="3">$$K_{\text{расч.}} = t_{\text{расч.}} = \frac{\left(\overline{X} - \overline{X}_0\right)\cdot\sqrt{n}}{S},$$ $$S^2 = \frac{\sum X_j^2 n_j - \frac{\left(\sum X_j n_j\right)^2}{n}}{n-1} = \frac{\sum\left(\overline{X}_j - \overline{X}\right)}{n-1}$$ $H_0\colon \overline{X} = \overline{X}_0$; H_0 принимается если:</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>$H_1\colon \overline{X} \neq \overline{X}_0$</td><td>$H_1\colon \overline{X} > \overline{X}_0$</td><td>$H_1\colon \overline{X} < \overline{X}_0$</td></tr>
<tr><td>$\left|t_{\text{расч.}}\right| < t_{кр,\,\nu,\,\frac{\alpha}{2}}$</td><td>$t_{\text{расч.}} < t_{\text{кр.},\,\nu,\,\alpha}$</td><td>$t_{\text{расч.}} > -t_{\text{кр.},\,\nu,\,\alpha}$</td></tr>
<tr><td>$-t_{\text{кр.}} < t_{\text{расч.}} < +t_{\text{кр.}}$</td><td>$t_{\text{расч.}} < +t_{\text{кр.}}$</td><td>$-t_{\text{кр.}} < t_{\text{расч.}}$</td></tr>
<tr><td colspan="3">$t_{\text{кр.}}$ определяется по таблице распределения Стьюдента</td></tr>
</table>

Таблица 3.6 - Сравнение двух средних 2-х генеральных совокупностей при известных дисперсиях

<table>
<tr><td rowspan="2"></td><td>2 выборки большого объема</td></tr>
<tr><td>Независимые объемы выборок n и m</td></tr>
</table>

<table>
<tr><td rowspan="6">Генеральные дисперсии известны</td><td colspan="3">$$K_{\text{расч.}} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{D(X_1)}{n} + \frac{D(X_2)}{m}}},$$
$$H_0 \; M(X_1) = M(X_2)$$
H_0 принимается если:</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>$H_1\colon \; M(X_1) \neq M(X_2)$</td><td>$H_1\colon \; M(X_1) > M(X_2)$</td><td>$H_1\colon \; M(X_1) < M(X_2)$</td></tr>
<tr><td>$\lvert K_{\text{расч.}} \rvert < K_{\text{кр.},\, \alpha/2}$</td><td>$K_{\text{расч.}} < K_{\text{кр.},\, \alpha}$</td><td>$K_{\text{расч.}} > -K_{\text{кр.},\, \alpha}$</td></tr>
<tr><td>$\Phi\left(K_{\text{кр.},\, \alpha/2}\right) = \frac{1-\alpha}{2}$</td><td colspan="2">$\Phi\left(K_{\text{кр.},\, \alpha}\right) = \frac{1-2\alpha}{2}$</td></tr>
<tr><td colspan="3">$K_{\text{кр.}}$ определяется по таблице значений интеграла Лапласа, $\Phi(x)$</td></tr>
</table>

Таблица 3.7 - Сравнение двух средних 2-х генеральных совокупностей при неизвестных дисперсиях

<table>
<tr><td rowspan="2">Ге</td><td>2 выборки малого объема</td></tr>
<tr><td>Независимые объемы выборок n и m</td></tr>
</table>

$$D(X_1) \approx D(X_2)$$

$$K_{расч} = t_{расч} = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{(n-1)\cdot S_1^2 + (m-1)\cdot S_2^2}} \cdot \sqrt{\frac{n \cdot m(n+m-2)}{n+m}}$$

$$v = n + m - 2\,,$$

$H_0: M(X_1) = M(X_2)$, H_0 принимается если:

Двусторонний критерий	Односторонний критерий	
$H_1: M(X_1) \neq M(X_2)$	$H_1: M(X_1) > M(X_2)$	$H_1: M(X_1) < M(X_2)$
$\lvert t_{\text{расч.}} \rvert < t_{\text{кр.}, \text{v}, \frac{\alpha}{2}}$	$t_{\text{расч.}} < t_{\text{кр.}, \text{v}, \alpha}$	$t_{\text{расч.}} > -t_{\text{кр.}, \text{v}, \alpha}$

$t_{\text{кр.}}$ - определяется по таблице распределения Стьюдента

Продолжение таблицы 3.7

Гене-	2 выборки малого объема
	Зависимые объемы выборок $n = m$

<table>
<tr><td rowspan="5"></td><td colspan="3">$$K_{расч} = t_{расч} = \frac{\overline{d} \cdot \sqrt{n}}{S_d}, \text{ где } \overline{d} = \frac{\sum d_j}{n}; \quad d_j = x_{1j} - x_{2j};$$
$$S_d = \sqrt{\frac{\sum d_j^2 - \frac{\left(\sum d_j\right)^2}{n}}{n-1}};$$
$$v = n - 1;$$
$H_0: M(X_1) = M(X_2)$, H_0 принимается если:</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>$H_1: M(X_1) \neq M(X_2)$</td><td>$H_1: M(X_1) > M(X_2)$</td><td>$H_1: M(X_1) < M(X_2)$</td></tr>
<tr><td>$\left|t_{\text{расч.}}\right| < t_{\text{кр.}, v, \frac{\alpha}{2}}$</td><td>$t_{\text{расч.}} < t_{\text{кр.}, v, \alpha}$</td><td>$t_{\text{расч.}} > -t_{\text{кр.}, v, \alpha}$</td></tr>
<tr><td colspan="3">$t_{\text{кр.}}$ определяется по таблице распределения Стьюдента</td></tr>
</table>

Таблица 3.8 – Сравнение дисперсий

1-выборка

<table>
<tr><td colspan="3">$H_0: \sigma^2 = \sigma_0^2$ (σ_0^2 генеральная дисперсия)
$H_1: \sigma^2 < \sigma_0^2$</td></tr>
<tr><td colspan="3">Дано: n, S^2, σ_0^2</td></tr>
<tr><td colspan="3">$K_{\text{расч.}} = \chi^2_{\text{расч.}} = \frac{(n-1)S^2}{\sigma_0^2}$, $v = n - 1$.
а) $n < 30$, $\chi^2_{\text{кр.}}$ при α, q - по таблице распределении Пирсона;
б) $n > 30$, $\chi^2_{\text{кр.}}$ при α, q $\chi^2_{\text{кр.}} = v\left(1 - \frac{2}{9v} + u_\alpha\sqrt{\frac{2}{9v}}\right)^3$,
$u_i = \frac{x_i - \bar{x}}{\sigma}$, $u_\alpha = \Phi(u_i)$ определяется по таблице распределения Лапласа</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>$H_1: \sigma^2 \neq \sigma_0^2$</td><td>$H_1: \sigma^2 > \sigma_0^2$</td><td>$H_1: \sigma^2 < \sigma_0^2$</td></tr>
<tr><td rowspan="2">$\chi^2_{\text{кр. лев.}} \leq \chi^2_{\text{расч.}} \leq \chi^2_{\text{кр. прав.}}$
Тогда H_0 принимается
$\chi^2_{\text{кр. лев.}}$ при $\left(1 - \frac{\alpha}{2}\right)$, q
$\chi^2_{кр.прав}$ при $\frac{\alpha}{2}$, q
$\Phi\left(u_{\frac{\alpha}{2}}\right) = \frac{1-\alpha}{2}$</td><td>$\chi^2_{\text{расч.}} < \chi^2_{\text{кр. прав.}, \alpha, q}$</td><td>$\chi^2_{\text{расч.}} > \chi^2_{\text{кр. лев.}(1-\alpha), q}$</td></tr>
<tr><td colspan="2">$\Phi(u_\alpha) = \frac{1 - 2\alpha}{2}$</td></tr>
</table>

Продолжение таблицы 3.8– Сравнение дисперсий

<table>
<tr><td colspan="2">2-выборки</td></tr>
<tr><td colspan="2">$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$</td></tr>
<tr><td colspan="2">Дано: n, m, S_1^2, S_2^2</td></tr>
<tr><td colspan="2">$K_{\text{расч.}} = F_{\text{расч.}} = \dfrac{S_1^2}{S_2^2}$, при v_1, v_2
$S_1^2 > S_2^2,\ v_1 = n - 1,\ v_2 = m - 1$</td></tr>
<tr><td>Двусторонний критерий</td><td>Односторонний критерий</td></tr>
<tr><td>$H_1 : \sigma_1^2 \neq \sigma_2^2$</td><td>$H_1 : \sigma_1^2 > \sigma_2^2$</td></tr>
<tr><td>$F_{\text{расч.}} < F_{\text{кр.}}$
при $v_1, v_2, \dfrac{\alpha}{2}$</td><td>$F_{\text{расч.}} < F_{\text{кр.прав.}}$
при v_1, v_2, α</td></tr>
<tr><td colspan="2">$F_{\text{кр.}}$ определяется по таблице распределения Фишера-Снедекора</td></tr>
</table>

Продолжение таблицы 3.8– Сравнение дисперсий

<table>
<tr><td colspan="2">k – выборок</td></tr>
<tr><td>$H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$
$H_1: \sigma_1^2 > \sigma_2^2 > \ldots > \sigma_k^2$</td><td>$H_0: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$
$H_1: \sigma_1^2 > \sigma_2^2 > \ldots > \sigma_k^2$</td></tr>
<tr><td>Дано: $n_1 = n_2 = \ldots = n_k$</td><td>Дано: $n_1 \neq n_2 \neq \ldots \neq n_k$
$n_i \geq 4$</td></tr>
<tr><td>$K_{\text{расч.}} = \dfrac{S^2_{\max}}{\sum\limits_{i=1}^{k} S_i^2}$,
$S^2_{\max} = \max\left\{S_i^2\right\}$</td><td rowspan="2">$K_{\text{расч.}} = B_{\text{расч.}} = \dfrac{V}{C}$,
$V = 2{,}303 \cdot \left(\lg S^2 - \sum\limits_{i=1}^{k} \nu_i \lg S_i^2\right)$
$C = 1 + \dfrac{1}{3(k-1)}\left[\sum\limits_{i=1}^{k} \dfrac{1}{q_i} - \dfrac{1}{q}\right]$, где $q_i = n_i - 1$, $q = \sum\limits_{i=1}^{k} q_i$,
$S_k^2 = \dfrac{\sum\limits_{i=1}^{k} q_i S_i^2}{q}$</td></tr>
<tr><td>Односторонний критерий</td></tr>
<tr><td rowspan="2">По критерию Кочрена $K_{кр}$, α, ν, k.
H_0 - принимается, если $K_{\text{расч.}} < K_{\text{кр}, \alpha, \nu, k}$.
Если H_0 принята то в качестве оценки дисперсии генеральной совокупности принимают
$S_k^2 = \dfrac{\sum\limits_{i=1}^{k} S_i^2}{k}$</td><td>Односторонний критерий</td></tr>
<tr><td>H_0 принимается, если $B_{\text{расч.}} < \chi^2_{\text{кр. прав.}}$ при α, q по таблице распределения Пирсона</td></tr>
</table>

Задачи для самостоятельного решения

61. Произведено выборочное обследование 10% приусадебных участков восьми районов случайным бесповоторным способом. Получены следующие результаты об урожайности овощей:

№ п/п	Урожайность, ц/га	Среднее квадратическое отклонение, ц/га	Доля овощей в площади участков, %	Число обследованных участков
1	215	30	30	100
2	246	35	35	80
3	305	32	40	150
4	220	24	50	120
5	164	20	36	60
6	280	23	65	70
7	340	40	45	90
8	316	36	53	100

При уровне значимости 0,05 по двум районам проверить гипотезы о равенстве средних выборочных урожайностей, дисперсий, долей посевов овощей в площади приусадебных участков.

62. Из нормальной ГС извлечена случайная выборка объёма $n = 20$, по этой выборке найдена исправленная выборочная дисперсия $s^2 = 16,3$. При уровне значимости 0,01 проверить нулевую гипотезу H_0 о равенстве дисперсии $\sigma^2 = 14$, приняв в качестве конкурирующей гипотезы H_1: $\sigma^2 > 14$.

63. По двум независимым выборкам объёмов $m = 9$ и $n = 15$, извлеченных из нормальных ГС соответственно X и Y, найдены исправленные выборочные дисперсии $s_x^2 = 2,61$ и $s_y^2 = 0,77$. При уровне значимости 0,1 проверить нулевую гипотезу H_0 о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: $D(X) \neq D(Y)$.

64. Из двух нормальных генеральных совокупностей с неизвестными (предположительно равными) дисперсиями извлечены выборки, объёмы которых $m = 11$ и $n = 17$, выборочные средние $\overline{x} = 131,2$ и $\overline{y} = 127,2$; исправ-

ленные дисперсии $s_x^2 = 0{,}87$ и $s_y^2 = 0{,}65$. Требуется при уровне значимости 0,05 проверить нулевую H_0 о равенстве математических ожиданий.

65. Из двух партий тортов, изготовленных на одном хлебозаводе, в булочную завезли 10 тортов первой партии и 12 тортов второй партии. Был произведен замер веса каждого из тортов и получены следующие результаты:

Вес тортов 1-й партии x_i кг	1,34	1,35	1,37	1,39
Число тортов первой партии	2	3	4	1

Вес тортов 2-й партии y_i кг	1,32	1,34	1,36
Число тортов второй партии	2	2	2

Требуется при уровне значимости 0,02 проверить гипотезу H_0 о равенстве математических ожиданий. Предполагается, что случайные величины X и Y распределены нормально.

66. В двух цехах в течение нескольких дней проводился выборочный контроль производительности труда. Результаты отражены в таблицах.

Цех № 1	1	2	3	4	5
Производительность труда	22,6	23,0	22,8	23,2	23,6

Цех № 2	1	2	3	4
Производительность труда	23,4	22,7	23,0	23,3

При уровне значимости 0,1 проверить нулевую гипотезу о равенстве средней производительности труда в этих цехах в предположении равенства их дисперсий.

67. Вероятность приобрести выигрышный билет в некоторой лотерее $p = 0{,}2$. Предприятие приобрело 100 лотерейных билетов, из которых в результате очередного розыгрыша лотереи оказалось 14 выигрышных билетов. При уровне значимости 0,05 требуется проверить правильность случайной выборки.

68. Ритмичность работы городского автобуса определяется дисперсией времени ожидания пассажиров, которая не должна превышать величины D = 9 мин. Результаты 30 наблюдений за работой нового маршрута приведены в таблице:

Время ожидания клиента, t_i	1	2	3	4
Производительность труда	23,4	22,7	23,0	23,3

Время ожидания клиента, t_i	3	5	7	9	11	13	15
Производительность труда	2	3	11	10	2	1	1

Проверить нулевую гипотезу о ритмичности работы нового автобусного маршрута по сравнению со средней нормой при уровне значимости 0,05.

69. Рекламное агенстство рассылает своим клиентам каталоги. Вероятность того, что клиент приобретёт одно из рекламируемых изделий p = 0,07. Агентство разослало серию из 1000 каталогов улучшенной формы. В результате клиентами было приобретено 97 изделий по новому каталогу. При уровне значимости 0,05 требуется проверить, эффективность новой серии по сравнению с прежней.

70. Крупный коммерческий банк заказал маркетинговое исследование по выявлению эффекта «премирования» (калькулятор, набор ручек и др.) как стимула для открытия счета в банке. Для проверки случайным образом было отобрано 200 «премированных» посетителей и 200 «непремированных». В результате выяснилось, что 79% посетителей, которым не предлагалась премия и 89% посетителей, которым премия предлагалась, открыли счет в банке в течение 6 месяцев. Используя эти данные, проверьте гипотезу о том, что доля «премированных» посетителей, открывших счет в банке, существенно отличается от удельного веса «непремированных», открывших счет. Уровень значимости α = 0,05.

3.3 Проверка гипотезы о виде закона распределения

Если закон распределения генеральной совокупности неизвестен, но имеются основания предположить, что он имеет определенный вид (назовем его A), то проверяют нулевую гипотезу H_0: генеральная совокупность распределена по закону A. Проверка гипотезы о предполагаемом законе неизвестного распределения производится с помощью критерия согласия.

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Имеется несколько критериев согласия: Пирсона К., Стьюдента, Фишера, Колмогорова - Смирнова и др. Для каждого критерия имеются соответствующие таблицы, по которым и находят критическую точку, удовлетворяющую этому требованию. Когда критическая точка уже найдена, вычисляют по данным выборок наблюдаемое значение критерия и сравнивают со статистикой. По результатам сравнения делают соответствующий вывод.

Замечание 1. Статистический вывод *неверно формулировать* так: генеральная совокупность имеет нормальный (показательный, равномерный и др.) закон распределения. Можно лишь утверждать, что *данная выборка согласуется с гипотезой о нормальном* (показательном, равномерном и др.) *распределении генеральной совокупности* с параметрами $a = \overline{x_в}$, $\sigma^2 = \sigma_в^2$ на уровне значимости α.

Ограничимся описанием некоторых критериев согласия.

3.3.1 Критерий Пирсона

Критерием Пирсона χ^2 («*хи-квадрат*») называют критерий, статистикой которого является случайная величина

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} = \sum_{i=1}^{k} \frac{\left(n_i - m_i\right)^2}{m_i}, \tag{3.30}$$

где k – число интервалов разбиения выборки (число степеней свободы); n_i – частота i – го интервала; m_i - теоретические частоты, определяемые по формуле (3.22); n – объем выборки; p_i – теоретическая вероятность попадания значения случайной величины X в i–й интервал.

Критерий согласия Пирсона служит для проверки гипотезы H_0: закон распределения выборки согласуется с гипотезой о нормальном распределении генеральной совокупности с параметрами $M(X) = a$ и $D(X) = \sigma^2$ (критерий аналогично применяется и для других распределений, в этом состоит его преимущество перед другими критериями).

Пусть эмпирическое распределение (выборка объема n) задано в виде равноотстоящих вариант и соответствующих им частот:

x_i	x_1	x_2	…	x_m
n_i	n_1	n_2	…	n_m

Для того, чтобы проверить согласуются ли данные выборки с нулевой гипотезой – гипотезой о нормальном распределении генеральной совокупности, поступают по следующему правилу:

1) выдвигают нулевую гипотезу о нормальном законе распределения случайной величины X и находят несмещенные оценки его параметров $\overline{x_в}$ и S;

2) определяют теоретические частоты m_i, соответствующие опытным частотам, по формуле (3.22) (если среди опытных частот имеются малочисленные, то их необходимо объединить с соседними):

$$m_i = n \cdot p_i = n \cdot \left[\Phi\left(\frac{x_i - \overline{x_в}}{S} \right) - \Phi\left(\frac{x_{i-1} - \overline{x_в}}{S} \right) \right], \qquad (3.31)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ - - стандартная функция Лапласа, значения которой находят по статистической таблице 2;

3) по формуле (3.30) вычисляют статистику критерия;

4) определяют число степеней свободы:

$$q = k - r - 1, \tag{3.32}$$

где k – число интервалов разбиения выборки; r – число оцениваемых независимых параметров выбранного закона распределения (для нормального закона $r = 2$, параметры a и σ);

5) задают уровень значимости α и критическую точку критерия определяют по статистической таблице 5 при заданных α и q;

6) если

$$\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}, \tag{3.33}$$

то нулевую гипотезу принимают, в противном случае нулевую гипотезу отвергают. Но это вовсе не означает, что H_0 является единственно подходящей гипотезой: просто расхождение между выборочными данными и гипотезой H_0 невелико, или иначе H_0 *не противоречит результатам наблюдений*; однако таким же свойством наряду с H_0 могут обладать и другие гипотезы.

Замечание 2. Объем выборки должен быть достаточно велик, т.е. $n > 50$ и все теоретические частоты удовлетворяют неравенству (3.21), т.е. $m_i = np_i \geq 5$ для каждого интервала. Интервалы, для которых это условие не выполняется, следует объединить с соседними, суммируя частоты.

Замечание 3**.** Для контроля вычислений формулу (3.30) приводят к виду:

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n\,. \tag{3.30'}$$

Пример 38. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты

m_i	4	27	73	135	128	78	50	5
n_i	5	27	70	125	137	82	48	6

Решение. Вычислим значение критерия Пирсона по формуле (3.30):

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{8} \frac{(5-4)^2}{4} + \frac{(27-27)^2}{27} + \frac{(70-73)^2}{73} + \frac{(125-135)^2}{135} +$$

$$+ \frac{(137-128)^2}{128} + \frac{(82-78)^2}{78} + \frac{(48-50)^2}{50} + \frac{(6-5)^2}{5} = 2{,}457.$$

Число степеней свободы в данном случае $q = 8 - 2 - 1 = 5$. По статистической таблице 5 критических точек распределения χ^2 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $q = 5$ находим $\chi^2_{\text{кр.}} = 11{,}1$. Итак, $\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}$, поэтому можно утверждать, что данная выборка согласуется с гипотезой о нормальном распределении генеральной совокупности.

Пример 39. Задана выборка, полученная для непрерывной случайной величины *X*. Обработав ее, записать гипотезу о виде закона распределения. Проверить гипотезу, используя критерий Пирсона ($\alpha = 0.05$).

$(x_i\,;\,x_{i+1})$	(0;5)	(5;10)	(10;15)	(15;20)	(25; 30)	(30;35)	(35;40)
n_i	67	49	24	17	9	3	2

Решение. В примере приведены сгруппированные опытные данные , полученные для непрерывной случайной величины. Для нее удобнее всего задавать *плотность распределения* . Для того, чтобы получить представление о плотности распределения надо строить гистограмму относительных частот . Подсчитываем относительные частоты попаданий в каждый интервал и высоты прямоугольников, из которых состоит гистограмма . Объем выборки равен $n = 67 + 49 + 24 + 17 + 9 + 3 + 2 = 171$.

$(x_i\,;\,x_{i+1})$	(0;5)	(5;10)	(10;15)	(15;20)	(25; 30)	(30;35)	(35;40)
n_i	67	49	24	17	9	3	2
w_i	0,3918	0,2865	0,1404	0,1994	0,0526	0,0175	0,0117
w_i/h_i	0,0784	0,0583	0,0281	0,0199	0,0105	0,0035	0,0023

Построим гистограмму относительных частот. Полученный график (рис. 18) позволяет предположить, что случайная величина подчиняется *показательному закону распределения* с плотностью распределения

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

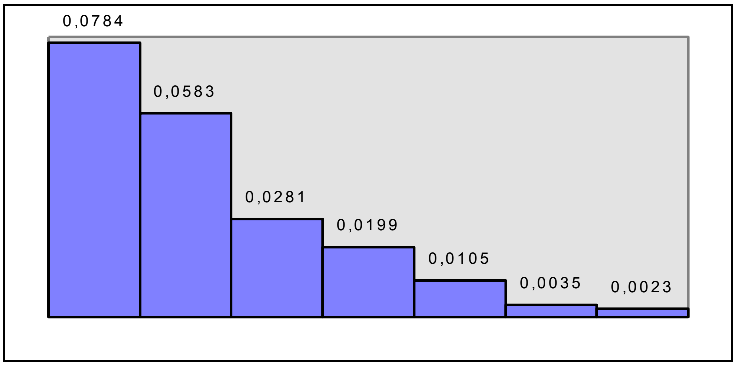


Рисунок 18 – Гистограмма относительных частот (к примеру 39)

Определим точечную оценку параметра λ по выборке, который связан с выборочной средней арифметической соотношением:

$$\lambda = \frac{1}{\overline{x}_в}. \tag{3.31}$$

Для этого перейдем от интервального ряда к дискретному и определим выборочную среднюю:

x_i^o	2,5	7,5	12,5	17,5	22,5	27,5	32,5
n_i	67	49	24	17	9	3	2

Выборочная средняя равна:

$$\overline{x}_в = \frac{\sum_{i=1}^{k} x_i^0 \cdot n_i}{n} = \frac{2,5 \cdot 67 + 7,5 \cdot 49 + 12,5 \cdot 24 + 17,5 \cdot 17 + 22,5 \cdot 9 + 27,5 \cdot 3 + 32,5 \cdot 2}{171} = 8,6696,$$

тогда оценка для параметра λ определится:

$$\lambda^* = \frac{1}{8,6696} = 0,1153.$$

Нулевая гипотеза о виде закона распределения сформулируется следующим образом: H_0: случайная величина распределена по показательному закону с параметром λ = 0,1153. Проверяем гипотезу по критерию Пирсона . Для этого нужно подсчитать вероятности попаданий в каждый из интервалов с использованием формул предполагаемого распределения. Для *показательного распределения* считаем их по известной формуле:

$$p_i = P(x_i < X < x_{i+1}) = e^{-\lambda \cdot x_i} - e^{-\lambda \cdot x_{i+1}}. \qquad (3.32)$$

Найдем значения показательной функции и занесем в таблицу

x_i	0	5	10	15	20	25	30	35
$e^{-0,1153\,x}$	1	0,5617	0,3155	0,1773	0,0996	0,0559	0,0314	0,0176

Теперь находим вероятности p_i попаданий значений в интервалы согласно формуле (3.32), а по ним определяем теоретические частоты $m_i = n \cdot p_i$. Результаты заносим в таблицу:

$(x_i; x_{i+1})$	(0;5)	(5;10)	(10;15)	(15;20)	(20; 25)	(25;30)	(30;35)
n_i	67	49	24	17	9	3	2
p_i	0,4383	0,2462	0,1382	0,0777	0,0437	0,0245	0,0138
m_i	79,95	42,10	23,63	13,29	7,47	4,19	2,36

При использовании критерия Пирсона малочисленные группы (с теоретическими частотами, меньшими, либо равными, 5) надо объединять с соседними. Здесь надо объединить две последние группы в один интервал и сложить соответствующие частоты, получим:

$(x_i; x_{i+1})$	(0;5)	(5;10)	(10;15)	(15;20)	(20; 25)	(25;35)
n_i	67	49	24	17	9	5
m_i	79,95	42,10	23,63	13,29	7,47	6,55

Подставляем теоретические и эмпирические частоты в формулу критерия Пирсона (3.30) и находим расчетное значение критерия:

$$\chi^2_{\text{расч.}} = \frac{(67-79{,}95)^2}{79{,}95} + \frac{(49-42{,}10)^2}{42{,}10} + \frac{(24-23{,}63)^2}{23{,}63} +$$

$$+ \frac{(17-13{,}29)^2}{13{,}29} + \frac{(9-7{,}47)^2}{7.47} + \frac{(5-6{,}55)^2}{6{,}55} = 4{,}9501 \ .$$

По таблице критических точек критерия χ^2 находим $\chi^2_{\text{кр}}(\alpha,\ q)$, где α = 0,05 - уровень значимости; q - число степеней свободы, определяемое по формуле (3.32) q = 5-1-1=3: $\chi^2_{\text{кр.}}(0{,}05\ ;\ 3) = 7{,}8$. Так как $\chi^2_{\text{расч.}} = 4{,}9501$ меньше $\chi^2_{\text{кр}} = 7{,}8$, то H_0 не противоречит результатам наблюдений и гипотезу о показательном распределении генеральной совокупности следует принимать.

Пример 40. Задана выборка , полученная для дискретной случайной величины X:

x_i	0	1	2	3	4	5
n_i	355	186	59	22	8	3

Используя критерий Пирсона, проверить гипотезу о том, что случайная величина подчиняется *распределению Пуассона* (уровень значимости α = 0.05) .

Решение. В пуассоновском распределении вероятности каждого из возможных значений подсчитываются по формуле Пуассона:

$$P(k) = \frac{a^k}{k!} \cdot e^{-a} \ . \qquad (3.33)$$

Входящий в формулу (3.33) *параметр* a *совпадает с математическим ожиданием.* Поэтому его можно оценить по выборке как выборочную среднюю:

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n} = \frac{0\cdot 355 + 1\cdot 186 + 2\cdot 59 + 3\cdot 22 + 4\cdot 8 + 5\cdot 3}{633} = 0{,}6588 \ . ,$$

Сформулируем нулевую гипотезу: H_o: случайная величина распределена по закону Пуассона с параметром $a = 0,6588$.

Находим по формуле Пуассона (3.33) вероятности p_i каждого из значений x_i, по ним теоретические частоты $m_i = n \cdot p_i$. Результаты заносим в таблицу:

x_i	0	1	2	3	4	5
n_i	355	186	59	22	8	3
p_i	0,5175	0,3409	0,1123	0,0247	0,0041	0,0005
m_i	327,58	215,79	71,09	15,64	2,60	0,32

Объединяем малочисленные три последних группы в одну:

n_i	355	186	59	33
m_i	327,58	215,79	71,09	18,56

Подставляем частоты в формулу критерия Пирсона (3.30) и находим расчетное значение критерия:

$$\chi^2_{\text{расч.}} = \frac{(355-327,58)^2}{327,58} + \frac{(186-215,79)^2}{215,79} + \frac{(33-18,56)^2}{18,56} = 17,64.$$

По таблицам критических точек критерия χ^2 находим:

$$\chi^2_{\text{кр.}}(\alpha, q) = \chi^2_{\text{кр.}}(0,05\ ;\ 4\text{-}1\text{-}1) = \chi^2_{\text{кр}}(0,05\ ;\ 2) = 6,0.$$

Так как $\chi^2_{\text{расч.}} = 17,64$ больше $\chi^2_{\text{кр.}} = 6,0$, то гипотезу о распределении Пуассона следует отвергнуть.

71. При уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты выборки объёма 1000 наблюдений по вкладам сбербанка.

m_i	57	99	150	195	191	147	93	68
n_i	66,8	91,9	149,8	191,5	191,5	149,8	91,9	66,8

72. Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$, проверить, согласуется ли, гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объёма $n = 200$.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

73. Из генеральной совокупности X извлечена выборка объемом 50.

Интервалы	$[-2{,}06; -1{,}46)$	$[-1{,}46; -0{,}86)$	$[-0{,}86; -0{,}26)$	$[-0{,}26; 0{,}34)$
Частоты n_i	2	6	11	15

Интервалы	$[0{,}34; 0{,}94)$	$[0{,}94; 1{,}54)$	$[1{,}54; 2{,}14)$
Частоты n_i	11	3	2

Требуется, используя критерий Пирсона, проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки при уровне значимости $\alpha = 0{,}05$.

74. В результате испытания 200 элементов на длительность работы получено эмпирическое распределение:

Время работы, лет	(0; 5)	(5; 10)	(10; 15)	(15; 20)	(20;25)	(20;25)
Число ламп, n_i	133	45	15	4	2	1

При уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о показательном распределении генеральной совокупности X - времени работы элементов с эмпирическим распределением выборки.

75. В итоге регистрации времени прихода 800 посетителей выставки (в качестве начала отсчета времени принят момент открытия выставки) получено эмпирическое распределение:

$(x_i; x_{i+1})$	(0;1)	(1;2)	(2;3)	(3;4)	(4;5)	(5;6)	(6;7)	(7;8)
n_i	259	167	109	74	70	47	40	34

В первой строке указаны интервалы времени, во второй – количество посетителей, пришедших в течение соответствующего интервала времени. При уровне значимости $\alpha = 0{,}01$ проверить, согласуется ли гипотеза о показательном распределении генеральной совокупности X с эмпирическим распределением выборки времени прихода посетителей выставки.

76. В итоге испытаний 1000 элементов технической системы на время безотказной работы получено эмпирическое распределение:

$(x_i; x_{i+1})$	(0;10)	(10;20)	(20;30)	(30;40)	(40;50)	(50;60)	(60;70)
n_i	365	245	150	100	70	45	25

В первой строке указаны интервалы времени часах, во второй - количество элементов, отказавших в i –ом интервале. При уровне значимости $\alpha = 0{,}01$ проверить, согласуется ли гипотеза о показательном распределении генеральной совокупности X с эмпирическим распределением времени безотказной работы элементов.

77. Для определения засоренности партии семян клевера семенами сорняков проверено 1000 случайно отобранных проб и получено эмпирическое распределение:

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

При уровне значимости $\alpha = 0{,}01$ проверить, согласуется ли гипотеза о Пуассоновском распределении генеральной совокупности X с эмпирическим распределением числа семян.

78. В итоге проверки на нестандартность 200 ящиков консервов получено следующее эмпирическое распределение (в первой строке указано количество нестандартных коробок консервов в одном ящике, во второй – число ящиков, содержащих число коробок, содержащих нестандартные консервы):

x_i	0	1	2	3	4
n_i	132	43	20	3	2

При уровне значимости $\alpha = 0{,}05$ проверить согласуется ли гипотеза о Пуассоновском распределении генеральной совокупности X с эмпирическим распределением числа нестандартных коробок.

79. Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема $n = 200$:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

80. Известно эмпирическое распределение 1000 экземпляров северной сосны по диаметру ствола:

Диаметр ствола, см	Количество сосен
14-18	16
18-22	35
22-26	109
26-30	183
30-34	214
34-38	197
38-42	115
42-46	71
46-50	36
50-54	19
54-58	5

Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить согласуется ли гипотеза о нормальном распределении генеральной совокупности X - диаметра северной сосны с эмпирическим распределением выборки.

3.3.2 Критерий Колмогорова

Критерий Колмогорова является наиболее простым критерием проверки гипотезы о виде закона распределения, однако его можно применять только в том случае, когда гипотетическое распределение полностью известно заранее из каких-либо теоретических соображений, т.е. когда *известен не только вид функции распределения $F(x)$, но и все входящие в нее параметры.*

В отличие от критерия Пирсона, где сравниваются теоретические и эмпирические частоты, критерий Колмогорова использует сравнение теоретической $F(x)$ и эмпирической (статистической) $F^*(x)$ функций распределения. Общая схема применения критерия Колмогорова состоит в следующем:

1. По имеющейся выборке $(x_1, x_2, x_3, \ldots, x_n)$ строится эмпирическая функция распределения $F^*(x)$.

2. В этих же экспериментальных точках x_i подсчитывается значение теоретической функции распределения $F(x_i)$ по предполагаемой формуле.

3. Определяется значение Δ - критерия Колмогорова как наибольшая из разностей:

$$\Delta = max \left| F^*(x_i) - \mathrm{F}(x_i) \right| . \tag{3.34}$$

Чем меньше значение этой разности, тем лучше согласуется выбранный теоретический закон с данными выборки. Для точной оценки вычисляют параметр $\lambda_{\text{расч.}}$:

$$\lambda_{\text{расч}} = \Delta \cdot \sqrt{n} . \tag{3.35}$$

4. Задают тот или иной уровень значимости α.

5. По статистической таблице 9 – Критические значения λ_α распределения Колмогорова находят соответствующее значение λ_α. Если

$$\lambda_{\text{расч.}} < \lambda_\alpha , \tag{3.36}$$

то гипотеза о предполагаемом законе распределения принимается, в противном случае – гипотеза отвергается.

Замечание 4. *Вид теоретического закона распределения выбирается* или на основании данных о механизме образования случайной величины, или *путем качественного анализа гистограммы распределения*. Если закон распределения не удается установить из общих соображений, то применяется его аппроксимация таким законом распределения, соответствующее число первых моментов которого равно их оценкам, полученным из рассматриваемой выборки.

Пример 41. Задана выборка:

0,45	2,51	1,64	3,35	0,54	3,90	1,36	2,65	1,32	3,06

Используя критерий Колмогорова, проверить гипотезу о том, что это выборка согласуется с гипотезой о равномерном распределении генеральной совокупности на интервале (0; 4).

Решение. Найдем эмпирическую функцию распределения $F^*(x)$. Для этого сначала запишем вариационный ряд, расположив значения вариант в возрастающем порядке:

0,45	0,54	1,32	1,36	1,64	2,51	2,65	3,06	3,35	3,90

Данные не повторяются, поэтому относительная частота каждого из них равна $1/n$. Объем выборки равен $n = 10$, значит $w_i = 0{,}1$. Так как эмпирическая функция распределения накапливает относительные частоты, то каждый раз при переходе через очередную экспериментальную точку она будет увеличиваться на 0,1. При этом в самой экспериментальной точке она разрывна, происходит скачок на 0,1. Значение функции в точке разрыва совпадает с пределом слева. Занесем в таблицу значения $F^*(x_i)$ причем будем записывать значения справа и слева от экспериментальной точки, т.е. $F^*(x_i\text{-}0)$ и $F^*(x_i+0)$.

Функция распределения равномерного распределения хорошо известна:

$$F(x) = \begin{cases} 0, & x < 0, \\ x/4, & 0 \le x < 4, \\ 1, & x \ge 4. \end{cases}$$

Если задана плотность распределения $f(x)$, то нужно найти функцию распределения по известной формуле:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \quad . \tag{3.37}$$

Подсчитаем функцию распределения в экспериментальных точках, т.е., $F(x_i)$. Результаты занесем в таблицу.

x_i	0,45	0,54	1,32	1,36	1,64	2,51	2,65	3,06	3,35	3,90
$F(x_i)$	0,11	0,14	0,33	0,34	0,41	0,63	0,66	0,77	0,84	0,98
$F^*(x_i\text{-}0)$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$F^*(x_i+0)$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Δ_i	0,11	0,06	**0,13**	0,06	0,09	0,13	0,06	0,07	0,06	0,08

Для каждого значения вариант x_i подсчитываем разность между значением теоретической функции распределения $F(x_i)$ и значениями эмпирической функции распределения справа и слева от экспериментальной точки, т.е. $F^*(x_i-0)$ и $F^*(x_i+0)$. Из этих разностей выбираем большую по модулю Δ_i и записываем ее в последний ряд таблицы. Затем выбираем наибольшее из чисел Δ_i в последней строке, т.е., находим $\Delta = max\ \Delta_i = 0{,}13$. Этим числом и измеряется отклонение.4

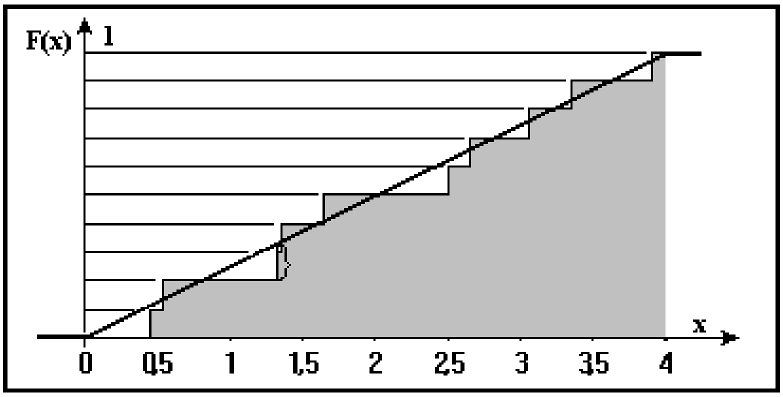


Рисунок 19 – Сравнение теоретической и эмпирической функций распределения (к примеру 41)

По таблице 9 критерия Колмогорова находим при $\alpha = 0{,}05$ значение $\lambda_\alpha = 1{,}358$. Вычислим расчетное значение $\lambda_{\text{расч.}}$ по формуле (3.35), используя объем выборки $n = 10$:

$$\lambda_{\text{расч.}} = 0{,}13 \cdot \sqrt{10} = 0{,}411$$

Так как $\lambda_{\text{расч.}} = 0{,}411$ меньше критического значения $\lambda_\alpha = 1{,}358$, то гипотезу о том, что это выборка согласуется с гипотезой о равномерном распределении генеральной совокупности на интервале (0; 4).

Пример 42. Результаты измерения 1000 деталей представлены в виде группированной выборки:

x_i	98,0	98,5	99,0	99,5	100,0	100,5	101,0	101,5	102,0	102,5
n_i	21	47	87	158	181	201	142	97	41	25

Проверить, пользуясь критерием Колмогорова, согласие полученных наблюдений с гипотезой о том, что генеральная совокупность X – длина диаметра деталей подчиняется закону нормального распределения с математическим ожиданием a = 100,25 мм и средним квадратическим отклонением σ = 1 мм, при уровне значимости α = 0,1.

Решение. Теоретическая функция распределения нормального закона определяется формулой:

$$F(x) = 0{,}5\cdot[1 + \Phi(x - a)]. \tag{3.38}$$

Статистическая функция распределения $F^*(x)$, согласно формуле (1.12), будет вычисляться по формуле:

$$F^*(x_i) = \frac{1}{1000}\left[\sum_{i=1}^{k-1} n_i + 0{,}5 n_k\right]. \tag{3.39}$$

Составим для каждого значения x_i разности $F^*(x_i)$ - $F(x_i)$ и выберем из них наибольшую по абсолютной величине. Все результаты вычислений занесем в таблицу.

i	$x_i - a$	$0{,}5\Phi(x_i - a)$	$F(x_i)$	$F^*(x_i)$	$\lvert F^*(x_i) - F(x_i) \rvert$
1	-2,25	-0,4877	0,0123	0,0105	0,0018
2	-1,75	-0,4599	0,0401	0,0445	0,0044
3	-1,25	-0,3944	0,1056	0,1115	0,0059
4	-0,75	-0,2734	0,2266	0,2340	0,0074
5	-0,25	-0,0987	0,4013	0,4035	0,0022
6	0,25	0,0987	0,5987	0,5945	0,0042
7	0,75	0,2734	0,7734	0,7660	0,0074
8	1,25	0,3944	0,8944	0,8855	**0,0089**
9	1,75	0,4599	0,9599	0,9545	0,0054
10	2,25	0,4877	0,9877	0,9875	0,0002

Наибольшее из разностей Δ_i = 0,0089. Вычислим расчетное значение $\lambda_{\text{расч.}}$ по формуле (3.35), используя объем выборки n = 1000:

$$\lambda_{расч.} = 0{,}0089 \cdot \sqrt{1000} = 0{,}281.$$

Находим из таблицы 9 - Критические значения λ_α распределения Колмогорова значение λ_α = 1,224. Видим, что $\lambda_{расч.} < \lambda_\alpha$. Следовательно, отклонения незначимы и можно считать, что гипотеза о согласии эмпирических данных с законом нормального распределения генеральной совокупности X – длины диаметра деталей с параметрами a = 100,25 мм и σ = 1 мм, не опровергается.

Задачи для самостоятельного решения

81. Дано распределение тысячи женщин по росту. Пользуясь критерием Колмогорова, установить согласуются ли опытные данные с гипотезой о распределении случайной величины по нормальному закону с параметрами $a = \overline{x_в}$, $\sigma^2 = \sigma_в^2$ при уровне значимости α = 0,05.

Рост, см.	Число женщин
134-137	1
137-140	4
140-143	16
143-146	53
146-149	121
149-152	193
152-155	229
155-158	186
158-161	121
161-164	53
164-167	17
167-170	5
170-173	1

82. Отдел технического контроля проверил 200 партий одинаковых изделий и получил следующее эмпирическое распределение:

x_i	0	1	2	3	4
n_i	116	56	22	4	2

В первой строке указано количество x_i нестандартных изделий в одной партии, во второй строке n_i – количество партий, содержащих x_i нестандартных изделий. Применив критерий Колмогорова, при уровне значимости α = 0,02 проверить, согласуются ли эмпирические данные с гипотезой о том, что генеральная совокупность - число нестандартных изделий распределено по закону Пуассона с параметром $a = \overline{x_в}$ (см. формулу (3.33).

83. Распределение 500 рабочих предприятия по времени, затрачиваемому на обработку детали, приведено в таблице:

Время, мин	Число рабочих
4,0-4,5	4
4,5-5,0	14
5,0-5,5	55
5,5-6,0	92
6,0-6,5	160
6,5-7,0	96
7,0-7,5	66
7,5-8,0	11
8,0-8,5	2

Требуется по критерию Колмогорова проверить гипотезу о том, что время, затрачиваемое на обработку детали, согласуется с гипотезой о нормальном распределении генеральной совокупности с параметрами $a = \overline{x_в}$, $\sigma^2 = \sigma_в^2$ при уровне значимости α = 0,1.

84. В некоторой местности в течение 300 суток регистрировалась среднесуточная температура воздуха. В итоге наблюдений было получено эмпирическое распределение:

Температура в градусах	Количество дней
-40 – (-30)	25
-30 – (-20)	40
-20 – (-10)	30
-10 - 0	45
0 - 10	40
10 - 20	46
20 - 30	48
30 - 40	26

При уровне значимости $\alpha = 0{,}1$ проверить, согласуется ли гипотеза о равномерном распределении генеральной совокупности X - среднесуточной температуры воздуха с эмпирическим распределением, применив критерий Колмогорова.

85. В течение 10 часов регистрировали прибытие автомашин к бензоколонке и получили эмпирическое распределение:

Время, час.	Количество машин
8 - 9	12
9 - 10	40
10 - 11	22
11- 12	16
12 - 13	28
13 - 14	6
14 - 15	11
15 - 16	33
16 - 17	18
17 - 18	14

Требуется, используя критерий Колмогорова, при уровне значимости $\alpha = 0{,}01$ проверить, согласуется ли гипотеза о равномерном распределении генеральной совокупности X - времени прибытия машин с эмпирическими данными.

86. Известно распределение двухсот радиоламп по сроку службы

Срок службы, ч	Количество ламп
300-400	1
400-500	9
500-600	18
600-700	33
700-800	40
800-900	52
900-1000	29
1000-1100	14
1100-1200	4

Пользуясь критерием Колмогорова, установить согласуются ли опытные данные с гипотезой о распределении случайной величины по нормальному закону с параметрами $a = \overline{x_в}$, $\sigma^2 = \sigma_в^2$ при уровне значимости α = 0,025.

87. Распределение нитей пряжи по крепости приведено в таблице

Крепость нити, г	Количество нитей
170-180	159
180-190	152
190-200	184
200-210	128
210-220	187
220-230	225
230-240	174
240-250	207
250-260	234

При уровне значимости α = 0,2 проверить, согласуется ли гипотеза о равномерном распределении генеральной совокупности X - числа нитей пряжи по крепости с эмпирическим распределением, применив критерий Колмогорова.

88 . Известно распределение тысячи волокон хлопка по длине:

Длина волокна, мм	Количество волокон
4,5-7,5	2
7,5-10,5	28
10,5-13,5	51
13,5-16,5	66
16,5-19,5	86
19,5-22,5	128
22,5-25,5	140
25,5-28,5	170
28,5-31,5	136
31,5-34,5	100
34,5-37,5	72
37,5-40,5	21

Пользуясь критерием Колмогорова, установить согласуются ли опытные данные с гипотезой о нормальном распределении случайной величины X - числа волокон хлопка по длине с параметрами $a = \overline{x_в}$, $\sigma^2 = \sigma_в^2$ при уровне значимости $\alpha = 0{,}01$.

89. В таблице приведена группированная выборка значений случайной величины X объема $n = 300$:

Границы интервала	Частота	Границы интервала	Частота
50 - 60	2	110 - 120	61
60 -70	3	120 - 130	49
70 - 80	9	130 - 140	25
80 - 90	23	140 - 150	19
90 - 100	33	150 - 160	16
100 - 110	56	160 - 170	4

Проверить, используя критерий Колмогорова, при уровне значимости $\alpha = 0{,}05$, согласуются ли эмпирические данные с нормальным распределением генелальной совокупности, оценки параметров которого подобрать на основании опытных данных.

90. Отсчет по шкале измерительного прибора оценивается приблизительно в долях деления шкалы. Приведено 200 результатов отсчета последней цифры между соседними делениями шкалы:

x_i	0	1	2	3	4	5	6	7	8	9
n_i	35	16	15	17	17	19	11	16	30	24

Установить, используя критерий Колмогорова, согласуются ли эмпирические данные с законом равномерного распределения генеральной совокупности X - отсчета последней цифры между соседними делениями шкалы, при котором вероятность появления любой цифры равна 0,1 при уровне значимости $\alpha = 0{,}05$.

ТИПОВОЙ РАСЧЕТ

Приведена выборка 120 результатов измерений случайной величины *X*.

1. Построить интервальный вариационный ряд (ряд 1) по частотам, относительным частотам и накопленным частотам.

2. От ряда 1 перейти к точечному вариационному ряду (ряд 2).

3. Начертить полигоны частот и относительных частот, кумуляту (по ряду 2) и гистограммы частот и относительных частот (по ряду 1).

4. Записать аналитически и построить графически статистическую функцию распределения (по ряду 2).

5. Найти выборочные средние: среднюю арифметическую, среднюю геометрическую, среднюю гармоническую;среднюю квадратичную и среднюю кубическую.

6. Определить моду и медиану графически и аналитически (по рядам 1 и 2).

7. Определить показатели вариации: вариационный размах, среднее линейное отклонение, выборочную дисперсию, выборочное среднеквадратическое отклонение, линейный коэффициент вариации и коэффициент вариации. При вычислении выборочной дисперсии сделать поправку Шеппарда.

8. Используя методы произведения и суммы найти выборочную среднюю и выборочную дисперсию.

9. Найти коэффициент асимметрии и эксцесс (по ряду 2), дать характеристику формы кривой распределения. Использовать коэффициент асимметрии и эксцесс в качестве выдвижения гипотезы о нормальном законе распределения генеральной совокупности по выборке.

10. На основе анализа гистограммы и статистической функции распределения оценить близость эмпирического распределения к нормальному закону (к любому другому известному закону распределения).

11. В предположении, что выборка извлечена из генеральной совокупности, подчиненной нормальному закону, при заданной надежности $\gamma = 0{,}95$ построить доверительные интервалы для неизвестного математического ожидания $\boldsymbol{a}$ при известной и неизвестной дисперсии, неизвестного среднего квадратического отклонения $\boldsymbol{\sigma}$ случайной величины X.

12. В предположении, что выборка извлечена из генеральной совокупности, подчиненной нормальному закону, проверить гипотезу при уровне значимости $\alpha = 0{,}01$ о равенстве выборочной средней с гипотетической генеральной средней (математическим ожиданием) при известной и неизвестной дисперсиях.

13. По критериям Пирсона (χ^2) и Колмогорова при уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении (или каком-то другом) генеральной совокупности X с эмпирическими данными.

№ задания	Значения случайной величины X											
1	1,55	1,79	1,64	1,72	1,76	1,82	1,90	1,56	1,60	1,62	1,84	1,68
	1,75	1,78	1,80	1,68	1,78	1,92	1,59	1,75	1,77	1,80	1,78	1,65
	1,66	1,56	1,77	1,74	1,66	1,71	1,67	1,70	1,54	1,69	1,82	1,64
	1,81	1,73	1,54	1,63	1,75	1,68	1,56	1,88	1,84	1,77	1,60	1,72
	1,63	1,69	1,66	1,75	1,69	1,54	1,56	1,79	1,52	1,66	1,70	1,81
	1,69	1,72	1,68	1,56	1,58	1,65	1,73	1,55	1,64	1,56	1,55	1,57
	1,56	1,80	1,74	1,81	1,64	1,67	1,62	1,60	1,55	1,72	1,66	1,82
	1,64	1,68	1,70	1,62	1,59	1,73	1,70	1,65	1,63	1,65	1,58	1,71
	1,71	1,59	1,67	1,58	1,71	1,66	1,59	1,62	1,73	1,69	1,68	1,52
	1,48	1,71	1,93	1,67	1,73	1,74	1,69	1,90	1,70	1,73	1,47	1,88
2	11,2	21,0	13,3	21,5	20,6	8,0	19,7	13,4	14,5	18,3	25,1	5,3
	14,2	12,0	17,7	15,9	11,1	18,5	20,0	19,1	9,6	20,5	13,8	21,3
	20,9	7,7	20,1	13,1	14,8	18,0	26,1	5,9	14,6	11,7	18,1	15,6
	11,6	20,2	18,8	8,2	12,0	13,5	22,0	14,4	15,2	15,1	11,0	11,8
	18,1	14,0	12,5	20,8	13,4	21,2	20,9	8,5	17,7	14,2	21,8	7,9
	19,5	13,6	14,3	19,6	25,4	12,3	19,9	19,7	20,2	17,8	6,9	25,8
	8,2	11,7	18,4	18,8	11,3	22,0	14,4	10,2	9,1	21,8	13,7	6,5
	17,9	16,5	16,9	17,3	15,8	14,6	16,8	8,8	12,5	22,3	15,7	16,9
	15,8	14,6	6,2	10,4	7,5	9,4	17,9	16,7	18,9	21,5	24,6	19,9
	15,0	13,9	15,4	6,0	15,7	14,9	15,6	16,1	14,8	15,7	5,9	15,3

3	13,6	12,2	13,2	12,8	12,3	13,3	13,0	13,1	14,9	13,8	12,7	11,9
	13,7	13,3	13,0	14,3	13,4	12,8	13,3	13,1	13,2	11,8	12,8	12,3
	13,0	13,9	14,5	12,2	13,0	12,8	13,6	13,2	12,6	12,4	11,7	13,9
	14,1	13,2	14,4	13,8	12,7	13,3	15,0	14,4	13,3	13,4	12,5	14,0
	13,5	12,9	13,8	14,8	15,0	12,6	13,6	13,5	15,2	13,5	13,8	14,2
	13,3	12,6	14,3	14,6	13,4	12,9	14,9	12,7	13,2	11,8	13,9	14,8
	12,6	13,7	14,4	13,8	14,3	12,0	14,4	14,2	14,1	12,8	13,7	12,5
	13,9	13,5	13,9	14,3	12,8	14,6	13,8	12,8	12,5	12,3	14,7	14,9
	12,8	14,6	13,2	14,4	12,5	12,4	12,9	14,7	13,9	14,5	14,6	12,9
	13,0	13,2	15,4	15,8	13,6	11,2	15,3	12,3	11,8	15,9	13,6	14,0
4	9,45	9,99	8,74	7,79	8,82	7,43	8,87	9,56	8,57	9,72	7,04	9,85
	8,50	8,68	9,80	8,47	9,89	8,18	8,59	8,24	8,07	8,63	8,48	8,69
	8,86	8,65	8,86	8,12	8,16	8,85	8,68	7,70	8,48	8,09	8,22	7,64
	7,11	8,73	7,15	8,08	8,05	7,18	8,96	7,88	9,84	8,57	9,30	7,58
	7,64	9,42	8,23	9,75	8,02	9,78	9,86	7,07	7,27	9,66	8,70	9,83
	8,69	8,32	9,68	8,06	9,96	8,35	8,42	8,25	9,34	9,46	8,75	8,57
	8,75	8,05	8,54	9,81	8,34	8,27	8,22	7,18	7,55	9,92	8,66	8,23
	8,97	9,67	8,62	8,20	7,59	8,73	7,79	8,65	8,63	7,65	7,78	7,09
	7,78	7,89	9,60	7,58	8,77	7,66	8,50	8,72	8,03	8,97	7,68	8,52
	8,05	8,32	8,28	9,98	7,32	7,55	7,46	9,95	7,38	7,02	8,05	7,59
5	150	154	148	149	160	147	158	164	153	135	152	150
	168	158	138	151	147	136	160	163	141	148	139	153
	171	141	143	156	164	161	159	149	146	156	130	152
	139	153	154	136	166	169	147	152	156	154	166	135
	155	148	138	173	136	150	159	142	173	144	150	145
	150	158	168	139	164	154	150	151	154	158	132	160
	146	154	148	139	158	146	167	155	144	141	158	146
	155	144	147	141	166	140	159	160	163	133	154	163
	157	152	168	157	149	135	158	165	146	158	169	143
	149	131	153	172	174	155	173	152	130	151	155	132
6	26,9	78,6	58,5	22,1	62,3	36,3	93,0	83,4	44,8	63,8	42,8	61,9
	63,8	43,5	63,0	44,6	63,9	52,8	83,7	43,1	53,2	41,8	24,8	32,3
	33,0	55,9	54,5	52,7	43,0	62,3	53,6	60,2	52,6	52,4	31,7	63,8
	44,1	53,7	44,2	53,8	62,7	73,3	75,5	34,4	73,3	69,4	52,5	54,9
	53,5	62,9	73,3	64,5	55,0	52,6	33,9	43,5	45,0	63,0	63,8	44,2
	83,3	92,5	64,3	54,6	82,9	42,9	54,8	62,7	53,2	51,8	53,9	54,6
	32,6	83,7	54,4	63,8	74,3	82,0	64,4	74,6	64,1	62,8	83,7	92,5
	73,9	53,5	63,7	74,3	52,8	64,6	73,8	51,8	62,5	42,3	34,7	64,1
	62,8	64,6	73,2	44,4	52,4	92,4	72,9	64,7	53,9	54,8	64,6	62,9
	60,3	64,8	27,0	89,5	90,2	65,0	91,9	62,8	64,5	92,1	65,1	63,7
7	4,25	2,05	4,02	6,71	4,52	2,82	5,80	2,87	7,35	4,44	6,64	2,68
	6,23	7,18	5,80	8,84	3,75	4,96	7,59	6,73	5,74	3,80	5,78	7,45
	3,66	6,56	2,77	3,74	6,63	8,71	6,67	5,72	6,57	5,69	6,02	5,64
	4,81	5,77	7,54	7,63	5,79	6,48	5,96	3,88	4,89	5,77	6,34	4,72
	2,63	4,69	6,02	4,75	5,05	7,54	6,76	4,79	6,52	6,83	5,70	6,85
	3,99	7,72	5,68	8,56	6,38	6,65	4,73	6,55	5,69	5,96	4,55	7,57
	5,59	5,80	6,73	6,81	5,64	6,07	5,62	5,90	6,55	6,72	2,66	6,80
	6,67	5,68	5,79	5,62	4,59	7,73	6,92	6,65	5,63	4,95	3,58	4,71

7	5,91	6,59	7,67	6,58	6,76	5,46	5,09	3,62	6,78	6,69	6,28	6,52
	4,98	5,04	5,21	4,88	7,42	2,12	4,78	3,92	3,88	4,15	2,32	5,23
8	260	786	108	585	574	325	165	484	620	535	206	339
	428	660	438	903	347	566	261	786	741	689	439	693
	381	292	543	606	494	261	559	447	396	458	520	452
	639	403	354	537	296	416	348	352	556	354	566	494
	255	418	286	673	536	750	559	542	571	544	350	546
	650	358	568	739	964	554	602	451	654	708	632	567
	246	354	648	639	808	646	567	559	504	441	358	241
	495	544	587	541	569	741	559	706	863	536	854	925
	557	652	468	257	549	835	458	565	646	758	599	543
	555	898	956	549	560	558	601	213	542	556	930	209
9	42,3	47,4	58,5	51,2	52,3	43,9	77,0	36,4	43,5	83,9	92,8	71,9
	33,8	46,5	63,0	64,5	63,9	52,8	53,8	43,6	53,3	41,5	84,1	35,4
	53,1	57,9	64,5	62,7	53,0	60,3	43,6	65,2	52,8	62,4	35,7	63,6
	44,9	54,7	60,2	83,8	42,7	73,7	70,5	44,4	73,6	59,3	52,6	58,9
	93,5	62,0	74,3	54,3	55,0	65,8	53,9	63,5	75,0	53,0	63,8	49,2
	73,3	72,5	64,3	54,6	72,9	30,9	44,7	60,7	53,4	58,8	56,9	64,6
	31,6	73,7	84,4	63,8	74,0	80,5	64,4	34,6	54,1	62,1	61,7	92,5
	73,9	53,5	63,7	74,3	52,8	64,6	73,8	51,8	62,5	42,3	34,7	64,1
	52,4	64,8	73,9	41,4	52,4	72,4	62,9	64,1	57,7	50,8	64,0	62,7
	31,9	53,9	55,1	54,2	32,0	78,1	77,6	57,3	56,8	78,3	54,9	52,8
10	1,25	2,86	0,32	6,92	4,50	1,82	0,85	8,07	5,31	3,94	4,84	0,67
	4,21	7,32	5,84	3,89	3,76	4,97	7,50	4,23	5,78	4,81	5,78	7,96
	2,66	1,59	4,71	5,04	6,23	8,21	4,67	5,75	4,57	5,62	6,02	5,64
	1,82	4,77	4,50	3,03	4,99	7,43	3,99	4,80	1,82	4,79	5,74	4,99
	2,96	4,25	5,02	4,61	5,15	6,44	7,06	8,39	4,02	5,83	4,97	3,81
	3,59	4,72	5,27	8,56	0,98	4,15	4,53	3,95	5,29	5,16	4,05	3,35
	5,22	5,81	4,73	3,94	5,64	5,07	4,68	5,91	6,05	6,92	2,16	6,88
	6,67	5,68	5,79	5,62	4,59	7,73	6,92	6,65	5,63	4,95	3,58	4,71
	4,31	5,29	7,16	6,58	5,16	4,62	5,19	3,62	1,84	3,89	4,27	3,92
	4,05	4,12	4,56	7,92	8,17	0,99	4,23	4,41	8,13	4,22	4,51	0,42
11	41,5	42,3	47,4	51,6	52,3	43,9	49,2	46,6	41,8	57,5	52,3	45,8
	48,0	49,4	57,5	44,4	51,1	49,8	43,8	50,3	49,7	40,9	51,9	50,1
	47,6	44,7	59,6	41,0	61,4	46,7	39,2	49,8	36,6	54,7	40,5	55,9
	60,2	51,8	38,9	43,7	51,7	55,4	45,6	44,9	59,0	42,8	47,6	30,2
	31,4	43,9	42,7	39,2	45,0	43,8	31,1	51,8	58,5	44,9	32,7	44,7
	35,9	40,2	37,3	50,1	62,5	46,1	50,3	62,7	39,9	45,0	44,8	45,3
	50,3	61,6	48,1	45,5	30,2	45,6	54,9	52,5	60,4	61,2	45,1	60,4
	44,9	45,1	62,4	37,4	44,8	50,3	44,9	30,4	44,8	39,6	50,4	43,9
	37,8	39,5	44,0	41,3	52,0	46,2	53,7	54,5	51,2	44,7	52,2	51,6
	61,8	44,8	45,1	42,9	31,6	62,4	45,0	44,7	62,3	45,2	31,9	44,8
12	0,56	1,99	2,94	3,32	3,36	5,82	2,91	3,66	4,60	1,65	1,81	2,98
	4,97	6,56	4,81	5,68	4,18	2,98	4,59	4,75	5,71	2,87	3,78	1,35
	3,16	2,95	1,70	4,14	5,67	3,77	5,27	3,79	3,56	3,09	3,12	3,94
	4,84	4,78	2,97	3,93	4,31	4,61	2,96	2,88	1,84	2,71	4,63	1,70
	6,03	5,09	1,60	2,85	2,69	4,04	1,06	4,31	0,92	0,66	5,70	0,81

12	5,49	2,77	2,88	4,06	3,08	2,95	2,83	2,97	1,64	6,56	3,05	4,57
	2,06	4,01	5,74	1,71	5,65	1,60	3,62	1,60	2,85	2,92	2,26	2,89
	3,95	3,12	4,96	0,62	1,59	5,73	4,75	2,95	3,63	3,15	1,18	3,78
	0,74	2,98	3,69	0,98	2,91	3,16	1,19	0,62	4,73	2,99	2,98	5,52
	3,08	3,11	3,48	3,15	6,60	6,49	3,24	3,55	3,47	6,61	6,59	3,38
13	650	700	742	769	564	712	681	643	753	538	752	710
	771	629	685	652	691	726	670	731	754	767	719	778
	668	725	663	630	740	691	702	797	711	656	703	821
	679	683	754	773	669	745	811	652	786	674	736	695
	735	691	738	693	642	740	800	684	753	731	695	649
	613	789	846	687	724	758	613	751	754	558	532	863
	686	624	798	819	768	726	711	715	654	761	785	776
	765	840	447	541	666	742	659	560	463	493	654	423
	557	652	468	657	449	535	558	665	646	558	469	843
	810	625	648	650	442	822	440	452	461	448	446	451
14	1,50	1,48	1,49	1,54	1,60	1,27	1,85	1,77	1,25	1,94	1,54	1,37
	1,51	1,47	1,40	1,89	1,61	1,70	1,60	1,63	1,78	1,41	1,48	1,36
	1,64	1,56	1,71	1,64	1,63	1,59	1,47	1,49	1,57	1,62	1,62	1,64
	1,82	1,27	1,59	1,83	1,49	1,46	1,59	1,80	1,42	1,39	1,74	1,69
	1,96	1,66	1,32	1,60	1,54	1,44	1,52	1,37	1,30	1,83	1,87	1,31
	1,55	1,72	1,27	1,56	1,68	1,50	1,38	1,75	1,49	1,36	1,50	1,35
	1,29	1,81	1,74	1,54	1,64	1,70	1,68	1,31	1,58	1,62	1,61	1,88
	1,67	1,68	1,79	1,62	1,59	1,73	1,29	1,65	1,43	1,50	1,58	1,71
	1,31	1,29	1,65	1,58	1,61	1,42	1,37	1,62	1,84	1,80	1,27	1,42
	1,98	1,55	1,59	1,60	1,58	1,61	1,63	1,97	1,32	1,66	1,65	1,59
15	31,9	29,8	11,4	10,6	10,3	17,9	39,2	26,0	21,8	37,5	22,3	25,3
	18,0	39,4	17,5	24,4	21,1	40,5	33,8	20,3	30,7	20,9	21,9	10,1
	27,6	14,7	29,6	21,0	21,4	36,7	19,2	19,8	26,1	24,7	30,5	15,9
	20,2	21,8	28,9	23,7	11,7	25,4	25,6	24,9	19,0	12,8	27,6	30,2
	31,4	23,9	22,7	29,2	25,0	23,8	31,1	21,8	28,5	14,9	22,7	24,7
	25,9	20,2	27,3	20,1	22,5	26,1	30,3	32,7	31,9	25,0	24,8	25,3
	20,3	21,6	28,1	25,5	30,2	25,6	24,9	22,5	24,4	21,2	25,1	30,4
	24,9	15,1	32,4	37,4	24,8	20,3	14,8	20,4	19,8	29,6	20,4	23,9
	34,8	36,0	24,0	21,3	22,0	26,2	23,7	24,5	30,2	34,7	22,2	21,6
16	94,5	99,8	71,8	79,6	80,3	77,9	99,2	66,0	71,8	77,5	82,3	85,3
	80,1	69,4	77,5	74,4	81,1	60,5	73,8	80,3	66,7	70,9	61,9	90,1
	87,6	74,7	69,6	71,0	71,4	86,7	79,2	69,8	86,1	82,7	81,5	75,9
	80,2	71,8	68,9	66,7	67,7	75,4	85,6	74,9	90,9	82,8	82,6	80,2
	71,4	73,9	72,7	79,2	75,0	73,8	81,1	71,8	68,5	66,9	82,7	84,7
	85,9	83,2	77,3	80,1	82,5	83,1	70,3	72,7	71,9	85,0	74,8	65,3
	80,3	81,6	78,1	75,5	80,2	85,6	74,9	82,5	84,4	71,2	85,1	80,4
	68,9	75,1	82,4	87,4	74,8	80,3	84,8	80,4	79,8	69,5	80,4	83,9
	74,8	66,5	74,0	81,3	82,0	86,2	83,7	74,5	80,2	84,7	82,2	81,6
	75,0	94,2	74,7	62,0	74,8	75,1	76,0	74,9	95,0	75,1	74,8	75,0
17	0,15	0,74	0,78	1,07	2,59	1,21	0,77	0,14	1,71	0,36	0,43	0,77
	0,43	1,20	0,19	0,77	0,41	1,06	1,19	0,67	0,74	0,14	0,26	0,56
	0,17	0,86	1,94	1,23	1,09	0,67	1,95	0,26	1,37	1,39	1,89	1,25

17	1,87	1,79	1,82	1,66	1,58	1,75	1,84	2,65	1,99	2,33	2,37	2,59
	1,76	1,64	1,53	0,91	2,02	2,39	1,80	1,87	2,48	2,01	1,69	1,34
	1,83	1,78	2,04	1,82	1,86	1,79	2,05	1,33	1,22	1,17	2,03	1,87
	1,88	1,41	1,86	1,77	1,35	2,11	1,87	1,66	1,05	2,00	1,88	2,11
	0,19	1,85	0,21	0,65	0,69	1,87	0,73	0,18	2,17	1,88	1,51	0,44
	0,56	1,67	0,34	1,79	1,80	0,45	1,84	0,31	1,86	0,95	1,76	1,87
	1,90	1,04	1,12	1,87	1,13	1,15	0,16	1,11	1,90	1,09	1,10	0,14
18	1,90	3,18	1,39	0,74	3,20	1,17	2,95	2,77	4,05	1,74	3,24	0,07
	0,81	3,12	1,20	2,69	1,91	2,30	3,20	4,13	1,38	2,41	4,58	2,26
	1,44	0,96	1,61	2,54	3,13	1,51	1,11	2,39	4,37	2,12	0,72	1,24
	1,52	1,82	2,92	0,83	0,98	1,76	4,19	4,30	2,62	0,91	0,74	1,19
	2,60	0,96	0,82	1,20	2,54	3,24	2,92	1,17	3,20	4,53	2,17	3,13
	5,05	1,12	1,97	0,96	3,68	0,99	3,38	2,85	2,59	4,06	4,30	1,15
	2,09	3,81	4,64	3,84	2,34	3,70	2,68	4,31	3,88	2,62	4,61	4,98
	0,97	2,68	3,87	1,62	3,59	2,73	1,19	1,65	1,93	3,50	2,58	2,71
	3,31	2,29	1,65	1,98	2,61	1,42	1,27	1,62	5,04	4,80	3,27	3,42
	0,08	2,45	2,39	2,50	5,02	2,51	2,48	5,11	5,10	2,53	2,49	2,52
19	702	471	715	721	724	482	491	740	493	508	702	410
	611	529	685	552	791	726	810	731	614	667	719	738
	608	705	519	630	743	611	702	793	701	656	713	800
	479	483	614	773	569	705	615	602	786	774	736	615
	535	651	538	793	442	540	800	684	753	731	695	649
	613	789	846	687	724	758	613	610	454	558	532	763
	786	424	498	519	468	626	621	615	654	767	781	776
	665	810	447	541	666	742	629	568	463	499	614	423
	539	652	478	617	449	535	758	715	616	558	469	643
	550	553	499	548	782	770	551	498	544	560	548	551
20	17,7	21,2	18,4	18,6	20,3	19,9	21,2	18,0	17,8	17,5	18,3	22,3
	15,3	20,4	22,5	18,4	18,1	14,5	19,8	14,3	16,7	20,9	21,9	19,1
	16,6	17,7	19,6	18,7	20,4	21,7	19,2	19,8	21,1	14,7	20,5	15,9
	20,2	21,8	22,0	13,9	14,7	22,4	21,6	14,9	19,0	22,8	27,6	30,2
	21,4	17,9	12,7	20,2	21,0	14,8	18,1	19,8	21,5	14,9	21,7	17,7
	20,9	20,2	22,3	18,1	18,5	16,1	20,3	22,7	11,9	15,0	14,8	15,3
	20,3	21,6	18,1	15,5	20,2	20,6	14,9	18,5	14,4	21,2	20,1	18,4
	16,9	15,1	22,4	17,4	18,8	20,3	17,8	20,4	16,8	20,6	18,4	21,9
	14,8	16,0	20,0	21,3	22,0	16,2	18,7	14,5	20,2	14,7	22,2	21,6
	21,9	22,0	19,8	18,6	18,7	18,4	18,5	14,0	14,1	18,5	18,6	18,7
21	0,27	0,72	0,58	078	0,21	19,9	21,2	18,0	17,8	17,5	18,3	22,3
	15,3	20,4	22,5	18,4	18,1	14,5	19,8	14,3	16,7	20,9	21,9	19,1
	16,6	17,7	19,6	18,7	20,4	21,7	19,2	19,8	21,1	14,7	20,5	15,9
	20,2	21,8	22,0	13,9	14,7	22,4	21,6	14,9	19,0	22,8	27,6	30,2
	21,4	17,9	12,7	20,2	21,0	14,8	18,1	19,8	21,5	14,9	21,7	17,7
	20,9	20,2	22,3	18,1	18,5	16,1	20,3	22,7	11,9	15,0	14,8	15,3
	20,3	21,6	18,1	15,5	20,2	20,6	14,9	18,5	14,4	21,2	20,1	18,4
	16,9	15,1	22,4	17,4	18,8	20,3	17,8	20,4	16,8	20,6	18,4	21,9
	14,8	16,0	20,0	21,3	22,0	16,2	18,7	14,5	20,2	14,7	22,2	21,6
	0,22	10,4	10,9	11,2	11,0	21,8	10,8	11,0	11,3	0,25	10,6	11,1

22	108	76	73	100	117	91	92	88	107	84	108	85
	102	94	91	96	82	86	95	89	69	100	83	114
	73	86	92	102	96	114	82	91	110	92	85	95
	96	94	112	78	77	81	93	96	105	85	94	113
	72	90	91	94	117	85	99	116	111	78	96	101
	99	95	101	114	109	97	95	94	115	74	97	115
	74	96	115	86	104	94	110	98	96	88	89	93
	72	112	108	94	99	96	118	97	93	116	109	95
	97	100	96	118	95	97	116	72	76	93	94	96
	95	93	115	110	98	96	102	75	97	94	98	94
23	27,9	35,7	25,9	36,8	33,5	41,6	28,2	24,3	21,0	36,1	22,7	24,8
	28,4	27,8	32,0	26,2	32,1	32,9	30,4	27,9	30,3	15,9	19,8	16,6
	33,1	27,8	24,5	35,4	21,2	32,3	28,4	24,0	36,1	23,5	24,6	35,7
	36,1	34,3	39,0	20,8	30,1	27,7	30,6	33,4	34,2	27,3	33,6	27,7
	32,0	27,8	26,6	29,4	30,8	27,1	22,8	41,4	32,5	34,8	21,1	35,6
	35,2	35,0	30,1	33,8	29,7	26,8	29,4	30,9	28,4	27,6	28,9	23,4
	29,8	26,4	30,2	32,4	27,2	28,4	31,5	34,0	23,6	28,5	41,3	28,1
	24,3	21,0	36,2	22,7	24,8	29,3	27,6	32,0	32,8	30,4	27,7	30,3
	21,3	26,7	33,1	42,0	31,2	31,4	41,9	30,9	31,4	32,0	21,1	21,2
	30,8	31,0	21,2	32,5	30,9	41,8	31,1	30,8	31,2	21,2	31,0	41,8
24	178	185	152	153	162	156	168	198	194	184	157	168
	164	161	153	189	166	162	157	135	173	188	167	185
	154	183	171	153	199	167	158	143	175	139	186	172
	165	159	164	158	176	154	170	175	130	145	167	160
	180	154	183	151	146	163	174	145	173	152	183	179
	179	146	149	163	196	163	161	182	154	170	181	147
	164	175	178	159	177	168	167	160	170	153	172	169
	151	168	165	164	172	172	145	169	167	176	173	198
	154	166	182	164	170	188	161	163	177	142	162	168
	164	166	180	171	154	198	167	142	151	170	172	171
25	40	25	40	35	40	45	40	35	39	40	40	40
	42	40	37	49	35	44	35	30	40	31	34	33
	40	41	30	38	40	30	34	31	33	28	29	42
	31	33	36	30	29	36	40	41	37	34	40	40
	28	27	41	38	42	31	41	38	39	32	42	43
	44	40	32	33	31	30	45	34	35	34	32	45
	40	29	36	40	25	29	32	30	26	40	40	40
	32	38	36	35	40	40	33	48	51	52	46	43
	26	38	40	50	37	35	32	40	40	43	32	31
	40	40	34	32	38	50	52	44	30	26	35	30
26	4,25	4,35	4,45	4,35	4,39	4,40	4,42	4,37	4,35	4,44	4,35	4,30
	4,30	5,10	5,32	4,15	4,86	4,32	4,55	4,80	5,22	5,21	2,89	2,92
	4,12	2,84	3,12	3,26	4,07	5,11	5,13	5,28	5,06	4,29	3,07	3,12
	2,65	2,87	3,05	3,44	4,15	4,32	4,03	5,15	4,23	4,08	3,11	3,33
	3,34	3,45	4,28	4,29	2,05	2,13	2,98	3,15	3,18	3,86	2,28	2,45
	3,06	2,55	4,87	3,98	5,25	4,31	2,88	3,08	4,26	5,26	5,42	2,61
	4,26	4,02	3,87	3,03	2,26	2,29	5,27	5,13	2,11	2,65	2,45	5,14
	5,16	5,32	2,09	5,63	5,65	3,08	3,56	3,24	3,26	3,67	4,03	5,20

26	4,02	4,32	3,15	3,96	3,69	5,15	2,11	4,39	5,64	3,46	5,60	4,02
	2,33	2,06	5,26	5,31	3,38	3,67	3,71	3,69	3,80	3,68	3,56	3,66
27	702	739	760	699	688	713	700	719	706	696	697	703
	714	755	838	748	715	831	709	734	736	700	722	715
	730	712	716	700	798	787	695	713	730	697	799	689
	729	785	777	711	780	784	755	706	756	678	880	755
	683	696	741	705	741	696	706	719	751	682	655	739
	718	707	737	702	752	706	710	721	717	779	793	695
	697	681	709	831	838	742	804	818	809	680	725	729
	765	836	847	822	653	660	752	824	837	661	781	714
	852	861	837	700	719	708	787	737	695	775	697	694
	751	744	818	683	704	711	654	729	709	800	814	644
28	20,7	17,1	24,6	19,1	17,2	25,5	21,3	26,1	20,3	24,2	24,8	22,1
	19,5	17,2	18,6	19,7	18,2	22,6	20,7	20,1	24,1	21,9	18,3	23,9
	20,9	20,1	20,8	15,9	18,3	19,7	17,6	19,2	20,4	19,5	22,0	20,5
	20,7	20,1	21,5	19,7	20,6	24,1	19,7	19,6	21,6	18,0	21,4	18,5
	19,6	20,8	22,0	23,1	23,4	22,5	19,8	21,7	22,5	22,4	23,7	23,6
	15,4	18,9	22,2	21,4	16,0	27,0	26,1	25,9	21,2	25,3	26,4	21,6
	27,6	24,8	25,3	26,6	25,2	15,4	24,3	17,3	17,8	19,2	17,1	15,6
	21,4	21,9	20,2	23,5	26,7	20,0	26,8	27,0	21,3	23,7	22,2	21,8
	18,1	17,4	21,7	21,0	23,5	25,5	24,1	24,3	24,5	25,2	21,9	16,3
	16,5	16,8	20,9	19,9	21,0	27,0	15,2	15,9	26,8	17,5	16,8	26,8
29	577	548	575	573	519	530	443	555	502	506	554	535
	478	495	635	564	583	534	518	574	454	566	522	599
	554	542	588	537	550	559	479	446	565	538	520	478
	554	542	566	558	480	486	500	560	517	558	559	496
	580	542	462	504	627	607	618	521	497	546	582	538
	495	554	552	511	555	496	495	530	534	437	446	460
	551	465	543	550	497	558	520	435	564	541	424	516
	413	505	514	479	490	483	577	569	493	461	486	492
	630	456	601	567	475	450	516	468	602	623	472	538
	454	471	527	516	564	445	518	460	436	504	498	506
30	2,91	2,70	2,79	2,82	2,54	2,64	2,16	2,76	2,48	2,53	2,76	2,48
	3,13	2,92	2,19	2,81	2,62	2,52	3,00	2,41	2,97	3,20	2,64	2,05
	2,44	2,77	2,79	2,53	2,75	2,45	2,48	2,43	2,64	2,67	2,13	2,18
	2,23	2,78	2,36	2,63	2,68	2,43	2,71	2,41	2,43	2,45	2,82	2,53
	2,76	2,75	2,46	2,84	2,71	2,27	2,56	3,09	3,00	3,10	2,55	2,52
	2,74	2,91	2,19	2,81	2,62	3,02	2,75	2,67	2,90	2,66	2,70	2,83
	2,37	2,22	2,81	2,76	2,82	2,77	2,91	2,56	2,71	2,92	2,78	3,13
	2,99	2,09	2,69	2,57	2,35	2,76	2,07	2,46	2,13	2,61	2,03	2,42
	2,48	2,21	2,25	2,07	2,01	2,41	2,19	2,39	2,09	2,02	2,20	2,08
	2,37	2,51	2,34	2,25	2,17	2,29	2,24	2,37	2,22	2,17	2,50	2,49

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Таблица 1 - Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1	2420	2396	2371	2347	2323	22989	2275	2251	2227	2203
1,1	2179	2155	2331	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290

Продолжение таблицы 1

2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблица 2 - Значения функции $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	0	1	2	3	4
0,0	0,00000	0,00399	0,00798	0,01197	0,01595
0,1	0,03983	0,04338	0,04776	0,05172	0,05567
0,2	0,07926	0,08317	0,08706	0,09095	0,09483
0,3	0,11791	0,12172	0,12552	0,12930	0,13307
0,4	0,15542	0,15910	0,16276	0,16640	0,17003
0,5	0,19146	0,19497	0,19847	0,20194	0,20540
0,6	0,22575	0,22907	0,23237	0,23565	0,23891
0,7	0,25804	0,26115	0,26424	0,26730	0,27035
0,8	0,28814	0,29103	0,29389	0,29673	0,29955
0,9	0,31594	0,31859	0,32121	0,32381	0,32639
1,0	0,34134	0,34375	0,34614	0,34850	0,35083
1,1	0,36433	0,36650	0,36864	0,37076	0,37286
1,2	0,38493	0,38686	0,38877	0,39065	0,39251
1,3	0,40320	0,40490	0,40658	0,40824	0,40988
1,4	0,41924	0,42073	0,42220	0,42364	0,42507
1,5	0,43319	0,43448	0,43574	0,43699	0,43822
1,6	0,44520	0,44630	0,44738	0,44845	0,44950
1,7	0,45543	0,45637	0,45728	0,45818	0,45907
1,8	0,46407	0,46485	0,46582	0,46638	0,46712
1,9	0,47128	0,47193	0,47257	0,47320	0,47381
2,0	0,47725	0,47778	0,47831	0,47882	0,47932
2,1	0,48214	0,48257	0,48300	0,48341	0,48382
2,2	0,48610	0,48645	0,48679	0,48713	0,48745
2,3	0,48928	0,48956	0,48983	0,49010	0,49036
2,4	0,49180	0,49202	0,48224	0,49245	0,49266
2,5	0,49379	0,49396	0,49413	0,49430	0,49446
2,6	0,49534	0,49547	0,49560	0,49573	0,49585
2,7	0,49653	0,49664	0,49674	0,49683	0,49693
2,8	0,49744	0,49752	0,49760	0,49767	0,49774
2,9	0,49813	0,49819	0,49825	0,49831	0,49836
3,0	0,49865	0,49869	0,49874	0,49878	0,49882
3,1	0,49903	0,49906	0,49910	0,49913	0,49916
3,2	0,49931	0,49934	0,49936	0,49938	0,49940
3,3	0,49952	0,49953	0,49955	0,49957	0,49958
3,4	0,49966	0,49968	0,49969	0,49970	0,49971
3,5	0,49977	0,49978	0,49979	0,49980	0,49981
3,6	0,49984	0,49985	0,49985	0,49986	0,49987
3,7	0,49989	0,49990	0,49990	0,49991	0,49991
3,8	0,49993	0,49993	0,49993	0,49994	0,49994
3,9	0,49995	0,49995	0,49996	0,49996	0,49996

Продолжение таблицы 2

x	5	6	7	8	9
0,0	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,42647	0,42786	0,42922	0,43056	0,43189
1,5	0,43943	0,44062	0,441 79	0,44295	0,44408
1,6	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,495098	0,49609	0,49621	0,49632	0,49643
2,7	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49931	3,30000	0,49952	3,40000	0,49986
3,1	0,49989	3,80000	0,49993	3,90000	0,49995
3,2	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49979	0,49980	0,49981
3,6	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49996	0,49996	0,49996	0,49997	0,49997

Таблица 3 – Значения функции $y = e^{-x}$

x	e^{-x}	x	e^{-x}	x	e^{-x}	x	e^{-x}
0,00	1,000	0,40	0,670	0,80	0,449	3,00	0,050
0,01	0,990	0,41	0,664	0,81	0,445	3,10	0,040
0,02	0,980	0,42	0,657	0,82	0,440	3,20	0,041
0,03	0,970	0,43	0,650	0,83	0,436	3,30	0,037
0,04	0,961	0,44	0,644	0,84	0,432	3,40	0,033
0,05	0,951	0,45	0,638	0,85	0,427	3,50	0,030
0,06	0,942	0,46	0,631	0,86	0,423	3,60	0,027
0,07	0,932	0,47	0,625	0,87	0,419	3,70	0,025
0,08	0,923	0,48	0,619	0,88	0,415	3,80	0,022
0,09	0,914	0,49	0,613	0,89	0,411	3,90	0,020
0,10	0,905	0,50	0,606	0,90	0,407	4,00	0,0183
0,11	0,896	0,51	0,600	0,91	0,403	4,10	0,0166
0,12	0,887	0,52	0,595	0,92	0,399	4,20	0,0150
0,13	0,878	0,53	0,589	0,93	0,395	4,30	0,0136
0,14	0,869	0,54	0,583	0,94	0,391	4,40	0,0123
0,15	0,861	0,55	0,577	0,95	0,387	4,50	0,0111
0,16	0,852	0,56	0,571	0,96	0,383	4,60	0,0101
0,17	0,844	0,57	0,565	0,97	0,379	4,70	0,0091
0,18	0,835	0,58	0,560	0,98	0,375	4,80	0,0082
0,19	0,827	0,59	0,554	0,99	0,372	4,90	0,0074
0,22	0,819	0,60	0,549	1,00	0,368	5,00	0,0067
0,21	0,811	0,61	0,543	1,10	0,333	5,10	0,0061
0,22	0,803	0,62	0,538	1,20	0,302	5,20	0,0055
0,23	0,795	0,63	0,533	1,30	0,273	5,30	0,0050
0,24	0,787	0,64	0,527	1,40	0,247	5,40	0,0045
0,25	0,779	0,65	0,522	1,50	0,223	5,50	0,0041
0,26	0,771	0,66	0,517	1,60	0,202	5,60	0,0037
0,27	0,763	0,67	0,512	1,70	0,183	5,70	0,0033
0,28	0,756	0,68	0,507	1,80	0,165	5,80	0,0030
0,29	0,748	0,69	0,502	1,90	0,150	5,90	0,0027
0,30	0,741	0,70	0,497	2,00	0,135	6,00	0,0025
0,31	0,733	0,71	0,492	2,10	0,122	6,10	0,0022
0,32	0,726	0,72	0,487	2,20	0,111	6,20	0,0020
0,33	0,719	0,73	0,482	2,30	0,100	6,30	0,0018
0,34	0,712	0,74	0,477	2,40	0,091	6,40	0,0017
0,35	0,705	0,75	0,472	2,50	0,082	6,50	0,0015
0,36	0,698	0,76	0,468	2,60	0,074	6,60	0,0014
0,37	0,691	0,77	0,463	2,70	0,067	6,70	0,0012
0,38	0,684	0,78	0,458	2,80	0,061	6,80	0,0011
0,39	0,677	0,79	0,454	2,90	0,055	7,00	0,0009

Таблица 4 - Критические точки распределения Стьюдента (t-критерия)

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,313	12,706	31,82	63,656	318,31	636,62
2	2,920	4,303	6,964	9,924	22,327	31,599
3	2,353	3,182	4,540	5,840	10,214	12,924
4	2,132	2,776	3,746	4,604	7,173	8,610
5	2,015	2,571	3,649	4,032	5,893	6,863
6	1,943	2,447'	3,142	3,707	5,207	5,958
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,249	4,297	4,780
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,105	4,024	4,437
12	1,782	2,179	2,680	3,085	3,929	4,321
13	1,771	2,160	2,650	3,112	3,852	4,220
14	1,761	2,145	2,625	2,976	3,787	4,140
15	1,753	2,131	2,603	2,947	3,732	4,072
16	1,746	2,120	2,583	2,920	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,551	2,878	3,611	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,849
21	1,721	2,080	2,521	2,837	3,534	3,824
22	1,717	2,074	2,514	2,824	3,513	3,790
23	1,714	2,069	2,506	2,815	3,501	3,775
24	1,711	2,064	2,498	2,802	3,475	3,742
25	1,708	2,060	2,490	2,792	3,452	3,723
26	1,706	2,056	2,485	2,784	3,446	3,718
27	1,703	2,052	2,471	2,771	3,422	3,697
28	1,701	2,048	2,469	2,769	3,407	3,669
29	1,699	2,045	2,467	2,758	3,401	3,658
30	1,697	2,042	2,462	2,754	3,395	3,645
40	1,684	2,021	2,427	2,708	3,312	3,575
60	1,671	2,000	2,395	2,665	3,234	3,460
120	1,658	1,980	2,362	2,627	3,172	3,378
∞	1,645	1,960	2,338	2,580	3,091	3,290
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Таблица 5 - Критические точки распределения χ^2

Число степеней свободы, q	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,039	0,0982	0,0157
2	9,21	7,38	5,99	0,103	0,0506	0,0201
3	11,3	9,35	7,81	0,352	0,216	0,115
4	13,3	11,1	9,49	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Таблица 6 – Значения чисел q для определения доверительного интервала среднего квадратического отклонения

Объем выборки, n	Надежность, γ			Объем выборки, n	Надежность, γ		
	0,95	0,99	0,999		0,95	0,99	0,999
7	0,92	-	-	25	0,32	0,49	0,73
8	0,80	-	-	30	0,28	0,43	0,63
9	0,71	-	-	35	0,26	0,38	0,56
10	0,65	-	-	40	0,24	0,35	0,50
11	0,59	0,98	-	45	0,22	0,32	0,46
12	0,55	0,90	-	50	0,21	0,30	0,43
13	0,52	0,83	-	60	0,188	0,269	0,38
14	0,48	0,78	-	70	0,174	0,245	0,34
15	0,46	0,73	-	80	0,161	0,227	0,31
16	0,44	0,70	-	90	0,151	0,211	0,29
17	0,42	0,66	-	100	0,143	0,198	0,27
18	0,39	0,60	0,92	150	0,115	0,160	0,211
19	0,39	0,60	0,92	200	0,099	0,136	0,185
20	0,37	0,58	0,88	250	0,089	0,120	0,161

Таблица 7 – Значения $t_\gamma = t(\gamma, n)$

Объем выборки, n	Надежность, γ			Объем выборки, n	Надежность, γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Таблица 8 – Критические точки распределения Фишера-Снедекора

Уровень значимости α = 0,01

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5989	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,87	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,44	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости α = 0,05

k_2	k_1											
	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	4,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,39	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Таблица 9 – Критические значения λ_α распределения Колмогорова

α	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
λ_α	0,828	0,895	0,974	1,073	1,224	1,358	1,520	1,627	1,950

Таблица 10 – Критические точки распределения Кочрена
(ν – число степеней свободы, k – количество выборок)

k \ ν	Уровень значимости α = 0,01						
	1	2	3	4	5	6	7
2	0,9999	0,9950	0,9794	0,9586	0,9373	0,9172	0,8988
3	9933	9423	8831	8335	7933	7606	7335
4	9676	8643	7814	7212	6761	6410	6129
5	9279	7885	6957	6329	5875	5531	5259
6	8828	7218	6258	5635	5195	4866	4608
7	8376	6644	5685	5080	4659	4347	4105
8	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704
9	7544	5727	4810	4251	3870	3592	3378
10	7175	5358	4469	3934	3572	3308	3106
12	6528	4751	3919	3428	3099	2861	2680
15	5747	4069	3317	2882	2593	2386	2228
20	4799	3297	2654	2288	2048	1877	1748
24	0,4247	0,2871	0,2295	0,1970	0,1759	0,1608	0,1495
30	3632	2412	1913	1635	1454	1327	1232
40	2940	1915	1508	1281	1135	1033	0957
60	2151	1371	1069	0902	0796	0722	0668
120	1225	0759	0585	0489	0429	0387	0357
∞	0000	0000	0000	0000	0000	0000	0000

k \ ν	Уровень значимости α = 0,01						
	8	9	10	16	36	144	∞
2	0,8823	0,8674	0,8539	0,7949	0,7067	0,6062	0,5000
3	7107	6912	6743	6059	5153	4230	3333
4	5897	5702	5536	4884	4057	3251	2500
5	5037	4854	4697	4094	3351	2644	2000
6	4401	4229	4084	3529	2858	2229	1667
7	3911	3751	3616	3105	2494	1929	1429
8	0,3522	0,3373	0,3248	0,2779	0,2214	0,1700	0,1250
9	3207	3067	2950	2514	1992	1521	1111
10	2945	2813	2704	2297	1811	1376	1000

Продолжение таблицы 10

k \ ν	Уровень значимости $\alpha = 0{,}01$						
	8	9	10	16	36	144	∞
12	0,2535	0,2419	0,2320	0,1961	0,1535	0,1157	0,0833
15	2104	2002	1918	1612	1251	0934	0667
20	1646	1567	1501	1248	0960	0709	0500
24	1406	1338	1283	1060	0810	0595	0417
30	0,1157	0,1100	0,1054	0,0867	0,0658	0,0480	0,0333
40	0898	0853	0816	0668	0503	0363	0250
60	0625	0594	0567	0461	0344	0245	0167
120	0334	0316	0302	0242	0178	0125	0083
∞	0000	0000	0000	0000	0000	0000	0000

k \ ν	Уровень значимости $\alpha = 0{,}05$						
	1	2	3	4	5	6	7
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332
3	9669	8709	7977	7457	7071	6771	6530
4	9065	7679	6841	6287	5895	5598	5363
5	8412	6338	5981	5440	5063	4783	4564
6	7808	6161	5321	4803	4447	4184	3980
7	7271	5157	4800	4307	3974	3726	3535
8	0,6798	0,4775	0,4337	0,3910	0,3595	0,3362	0,3185
9	6385	4450	4027	3584	3286	3067	2901
10	6020	3924	3733	3311	3029	2823	2666
12	5410	3346	3624	2880	2624	2439	2299
15	4709	2705	2758	2419	2195	2034	1911
20	3894	2354	2205	1921	1735	1602	1501
24	0,3434	0,2295	0,1907	0,1656	0,1493	0,1374	0,1286
30	2929	1980	1593	1377	1237	1137	1061
40	2370	1576	1259	1082	0968	0887	0827
60	1737	1131	0895	0765	0682	0663	0583
120	0998	0632	0495	0419	0371	0337	0312
∞	0000	0000	0000	0000	0000	0000	0000

k \ ν	Уровень значимости α = 0,05						
	8	9	10	16	36	144	∞
2	0,8159	0,8110	0,7880	0,7341	0,6602	0,5813	0,5000
3	6333	6167	6025	5466	4748	4031	3333
4	5175	5017	4884	4366	3720	3093	2500
5	4387	4241	4118	3645	3066	2013	2000
6	3817	3682	3568	3135	2612	2119	1667
7	3384	3259	3154	2756	2278	1833	1429
8	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0,1250
9	2768	2659	2558	2226	1820	1446	1111
10	2541	2439	2353	2032	1655	1308	1000
12	2187	2098	2020	1737	1403	1100	0833
15	1815	1736	1671	1429	1144	0889	0667
20	1422	1357	1303	1108	0879	0675	0500
24	0,1216	0,1160	0,1113	0,0942	0,0743	0,0567	0,0417
30	1002	0958	0921	0771	0604	0457	0333
40	0780	0745	0713	0595	0462	0347	0250
60	0552	0520	0497	0411	0316	0234	0167
120	0292	0279	0266	0218	0165	0120	0083
∞	0000	0000	0000	0000	0000	0000	0000

ЛИТЕРАТУРА

1. Баврин, И.И. Теория вероятностей и математическая статистика – М.: Высшая школа, 2005. – 160 с.

2. Балдин, К. В. Теория вероятностей и математическая статистика [Электронный ресурс] : Учебник / К. В. Балдин, В. Н. Башлыков, А. В. Рукосуев. - 2-е изд. - М.: Дашков и К, 2010. - 473 с. - http://znanium.com/bookread.php?book=414902.

3. Боровков, А.А. Математическая статистика. Оценка параметров. Проверка гипотез / А. А. Боровков. – М. : ФИЗМАТЛИТ, 2007. – 704 с.

4. Бось, В. Ю. Математическая статистика [Электронный ресурс]: метод. указ. / В. Ю. Бось. – Саратов: СГАУ, 2012.-60 с.- Электрон. текстовые дан. // Руконт: электронно-библиотечная система. - Режим доступа: http://rucont.ru/efd/192900.

5. Вентцель, Е. С. Теория вероятностей и ее инженерные приложения : [учебное пособие для втузов] / Е. С. Вентцель, Л. А. Овчаров. - М., 2007. - 490, [1] с.

6. Вуколов, Э. Л. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учеб. пособие / Э. Л. Вуколов. – 2-е изд., испр. и доп. – М. : ФОРУМ, 2008. – 464 с

7. Горелова, Г.В., Кацко, И.А. Теория вероятностей и математическая статистика в примерах и задачах с применением EXCEL. – Ростов на Дону «Феникс», 2005. – 475 с.

8. Горлач, Б.А. Теория вероятностей и математическая статистика: Учебное пособие / Б.А. Горлач. - СПб.: Лань, 2013. - 320 с.

9. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике – М.: Высшее образование, 2007.- 403 с.

10. Гмурман, В.Е. Теория вероятностей и математическая статистика: Учебное пособие для бакалавров / В.Е. Гмурман. - М.: Юрайт, 2013. - 479 с.

11. Гусева, Е. Н. Теория вероятностей и математическая статистика [Электронный ресурс] : Уч. пособ. / Е. Н. Гусева. - 5-е изд., стереотип. - М. : Флинта, 2011. - 220 с. http://znanium.com/bookread.php?book=406064.

12. Ивченко, Г.И., Медведев, И.Ю. «Введение в математическую статистику», Учебное пособие. – М.: Издательство ЛКИ, , 2010. - 310 с.

13. Калинина, В.Н. Теория вероятностей и математическая статистика: Учебник для бакалавров / В.Н. Калинина. - М.: Юрайт, 2013. - 472 с.

14. Кобзарь, А.И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. - М.: ФИЗМАТЛИТ, 2012. - 816 с.

15. Колемаев, В.А. Теория вероятностей и математическая статистика: Учебник / В.А. Колемаев, В.Н. Калинина. - М.: КноРус, 2013. - 376 с.

16. Кочетков, Е.С. Теория вероятностей и математическая статистика: Учебник / Е.С. Кочетков, С.О. Смерчинская, В.В. Соколов. - М.: Форум, НИЦ ИНФРА-М, 2013. - 240 с.

17. Краснов, М.Л. Вся высшая математика. Т. 5. Теория вероятностей. Математическая статистика. Теория игр: Учебник / М.Л. Краснов, А.И. Киселев, Г.И. Макаренко [и др.]. - М.: ЛКИ, 2013. - 296 с.

18. Кремер, Н.Ш. Теория вероятностей и математическая статистика: Учебник для студентов вузов / Н.Ш. Кремер. - М.: ЮНИТИ-ДАНА, 2012. – 551 с.

19. Лебедев, А.В. Теория вероятностей и математическая статистика: Учебное пособие / Л.Н. Фадеева, А.В. Лебедев; Под ред. проф. Л.Н. Фадеева. - М.: Рид Групп, 2011. - 496 с.

20. Павлов, С.В. Теория вероятностей и математическая статистика: Учебное пособие / С.В. Павлов. - М.: ИЦ РИОР: ИНФРА-М, 2010. - 186 с. http://znanium.com/bookread.php?book=217167.

21. Семенов, В.А. Теория вероятностей и математическая статистика: Учебное пособие / В.А. Семенов. - СПб.: Питер, 2013. - 192 с.

22. Сидняев, Н.И. Теория вероятностей и математическая статистика: Учебник для бакалавров / Н.И. Сидняев. - М.: Юрайт, ИД Юрайт, 2011. - 219 с.

23. Статистика в Excel: учеб. пособие для студентов, обучающихся по спец. 061700 "Статистика" и др. спец. / Н.В. Макарова, В.Я. Трофимец . - Москва: Финансы и статистика, 2006. – 364 с.

24. Чашкин, Ю.Р. Математическая статистика. Анализ и обработка данных: Учебное пособие / Ю.Р. Чашкин; Под ред. С.Н. Смоленский. - Рн/Д: Феникс, 2010. - 236 с.

25. Чудесенко, В.Ф. Сборник заданий по специальным курсам высшей математики. Типовые расчеты : учеб. пособие для вузов / В. Ф. Чудесенко, 2007. - 190 с.

26. Шириков, В.Ф. Математическая статистика : учеб. пособие для вузов : рек. Учеб.-метод. об-нием / В. Ф. Шириков, С. М. Зарбалиев, 2009. - 479 с.

27. Щербакова, Ю.В. Теория вероятностей и математическая статистика : конспект лекций / Ю. В. Щербакова, 2007. - 159 с.