МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ВОЛЖСКИЙ ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ)
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Т. А. Матвеева, В. Б. Светличная, Д. А. Мустафина, И. В. Ребро

МАТЕМАТИКА. ЧАСТЬ V.
ПРАКТИКУМ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Электронное учебное пособие

Волжский
2024

УДК 51(07)
ББК 22.1я73
М 34

Рецензенты:
доцент кафедры «Электроснабжение и энергетические системы» ФГБОУ ВО Волгоградский государственный аграрный университет, к. ф.-м. н. *Капля Е.В.;*
начальник отдела УВР образовательных программ цифровых технологий в промышленности, ГБПОУ «Волгоградский колледж управления и новых технологий им. Ю. Гагарина», к.т.н. *Александрина А.Ю.*

Издается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Матвеева, Т. А.
Математика. Часть V. Практикум по математической статистике [Электронный ресурс] : учебное пособие / Т. А. Матвеева, В. Б. Светличная, Д. А. Мустафина, И. В. Ребро ; Министерство науки и высшего образования Российской Федерации, ВПИ (филиал) ФГБОУ ВО ВолгГТУ. – Электрон. текстовые дан. (1 файл: 25,1 МБ). – Волжский, 2024. – Режим доступа: http://lib.volpi.ru. – Загл. с титул.экрана.
ISBN 978-5-9948-4889-0

Учебное пособие содержит изложение математической статистики в рамках изучения курса «Математика», «Теория вероятностей, математическая статистика и случайные процессы». Предназначено для студентов бакалавриата высших технических учебных заведений.

Илл. 26, табл. 6, библиограф.: 6 назв.

ISBN 978-5-9948-4889-0

Содержание

Введение

Предмет и задачи математической статистики

• ***Математической статистикой (МС)*** *называется наука, занимающаяся разработкой методов получения, описания и обработкой опытных данных с целью изучения закономерностей случайных массовых явлений.*

Задачи МС: 1. Описание и упорядочение статистического материала.
2. Оценка параметров распределения.
3. Проверка правдоподобия выдвигаемой гипотезы о соответствии статистического материала теоретическим выводам.

• *Множество всех* N *результатов измерения некоторой величины, которые могут быть получены в данных условиях, называется* ***генеральной совокупностью.*** N – объём генеральной совокупности ($N \to \infty$).

Аналогом генеральной совокупности в теории вероятностей является случайная величина X .

• *Множество случайно* n *отобранных измерений случайной величины* (*СВ*) *называется* ***выборочной совокупностью (выборкой).*** n – объём выборочной совокупности ($n << N$).

Распределение СВ X характеризуется рядом параметров (математическое ожидание, дисперсия, …). Эти параметры называются параметрами генеральной совокупности. Но их вычисление невозможно для больших значений N . Оцениваемые характеристики рассчитываются для выборки и объявляются точечными оценками характеристик всей совокупности. Чем больше $n \to N$, тем с большим основанием можно судить о свойствах генеральной совокупности.

1. Описание и упорядочение статистического материала

1.1. Статистический ряд. Его первичная обработка

- *Таблица, в которой содержатся номера опытов и соответствующие результаты измерений (значений признака), называется* ***статистическим рядом.***
- *Значение признака в i–ом опыте называется* ***вариантой*** (x_i)**.**

Значения вариант x_i есть значения случайной величины X .

- ***Частотой*** (m_i) *варианты называется количество опытов, в которых наблюдалось одно и то же значение варианты.*

$$\sum_{i=1}^{k} m_i = n, \quad k << n \qquad (n \text{ – объём выборки})$$

- ***Относительной частотой*** $\left(p_i^* = \omega_i\right)$ *варианты называется отношение частоты варианты к объёму выборки:*

$$p_i^* = \omega_i = \frac{m_i}{n} \quad \left(\sum_{i=1}^{k} p_i^* = 1\right).$$

- ***Дискретным вариационным рядом*** *называется упорядоченный ряд различных значений признака (в порядке возрастания или убывания) и соответствующих им кратностей или частот.*

Если объём выборки велик, интервал, которому принадлежат все варианты, разбивается на частичные интервалы (необязательно равные).

Интервальным вариационным рядом *называется ряд, который содержит интервалы разбиения значений признака и соответствующие им суммы частот вариант, которые попали в данный интервал разбиения.*

1.2. Графическая обработка статистического ряда

- ***Полигоном*** *относительных частот* (рис. 1) *называется ломаная, соединяющая точки вариационного ряда* $(x_i; p_i^*)$.

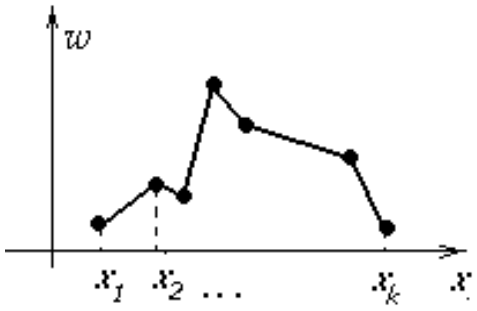


Рисунок 1

Замечание: Для интервального вариационного ряда в качестве абсциссы точки берут середину частичного интервала.

- ***Гистограммой*** *относительных частот* (рис. 2) *называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых являются интервалы разбиения* $(x_i; x_{i+1})$, *а высотой – числа* $f_i^* = \dfrac{p_i^*}{\Delta x_i}$.

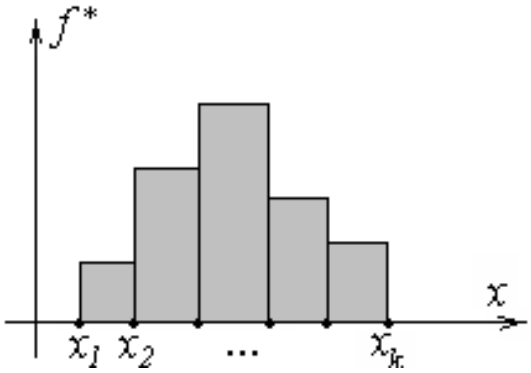


Рисунок 2

Замечание: Гистограмма строится только для интервального ряда.

Каждый прямоугольник гистограммы имеет площадь, равную p_i^*. Следовательно, вся площадь полученной фигуры будет равна 1.

f_i^* – статистический аналог дифференциальной функции распределения случайной величины (СВ), и гистограмма, являясь статистическим анало-

гом кривой распределения, в первом приближении указывает на вид теоретического распределения.

Полигон частот повторяет контур гистограммы. Поэтому он также указывает на вид теоретического распределения.

- ***Кумулятой*** (рис. 3) *называется ломаная, соединяющая точки с абсциссой* x_i *– значением варианты дискретного вариационного ряда (или правым концом частичного интервала интервального вариационного ряда) и ординатой, пропорциональной накопленной частоте.*

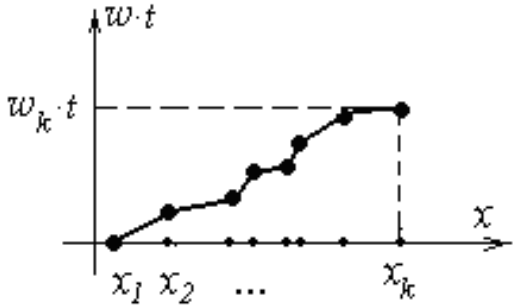


Рисунок 3

«Накопленные частоты» – значения ***эмпирической функции распределения*** (рис. 4)***:***

$$F^*(x)=\begin{cases}0, & x \le x_1,\\ p_1^*, & x_1 < x \le x_2,\\ p_1^* + p_2^*, & x_2 < x \le x_3,\\ p_1^* + p_2^* + p_3^*, & x_3 < x \le x_4,\\ \ldots & \\ 1, & x > x_k.\end{cases}$$

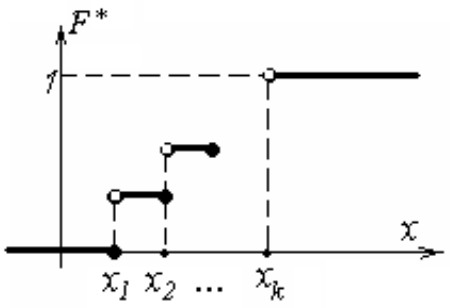


Рисунок 4

$F^*(x)$ – статистический аналог интегральной функции распределения СВ.

Свойства *эмпирической функции распределения:*

1. неубывающая;
2. неотрицательная;
3. область значения: $[0;1]$;
4. если x_1 – наименьшая варианта, то $(\forall x \le x_1)$: $F^*(x)=0$;

   если x_k – наибольшая варианта, то $(\forall x > x_k)$: $F^*(x)=1$.

1.3. Эмпирические характеристики выборки

- ***Модой*** M_0^* *называется такое значение варианты, которой соответствует максимальное значение частоты.*

В случае интервального вариационного ряда мода находится внутри частичного интервала $l_i = [x_i ; x_{i+1}]$, для которого соответствующая относительная частота p_i^* максимальна.

На чертеже гистограммы моду можно определить следующим образом (рис. 5):

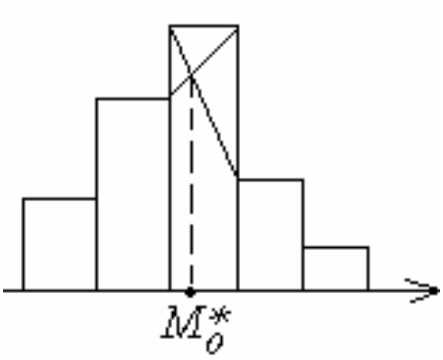


Рисунок 5

Значение моды $M_0^* \in [x_i ; x_{i+1}]$ вычисляется по формуле:

$$M_0^* = x_i + h_i \cdot \frac{p_i^* - p_{i-1}^*}{\left(p_i^* - p_{i-1}^*\right) + \left(p_i^* - p_{i+1}^*\right)},$$

где $h_i = (x_{i+1} - x_i)$ – длина частичного интервала l_i, p^*_{i-1}, p^*_{i+1} – относительные частоты, соответствующие предыдущему и последующему частичным интервалам.

- ***Медианой*** M^*_L *называется такое значение варианты* x, *для которой справедливы условия:* $\begin{cases} p^*(X \le M^*_L) \ge 0{,}5; \\ p^*(X \ge M^*_L) \ge 0{,}5. \end{cases}$

В случае интервального вариационного ряда медиана принадлежит тому частичному интервалу $l_i = [x_i; x_{i+1}]$, для которого накопленная частота составляет половину или больше половины всей суммы частот, а предыдущая накопленная частота меньше половины всей суммы частот.

Прямая $x = M^*_L$ делит площадь гистограммы пополам.

Значение медианы $M^*_L \in [x_i; x_{i+1}]$ вычисляется по формуле:

$$M^*_L = x_i + \frac{h_i}{p^*_i} \cdot \left(0{,}5 - \sum_{k=1}^{i-1} p^*_k\right).$$

- ***Выборочной средней*** $\bar{x}$ *называется среднее арифметическое вариант:*

$$\boxed{\bar{x} = m_в[X] = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i} = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i\, m_i = \sum_{i=1}^{k} x_i\, p^*_i .$$

В случае интервального ряда в качестве x_i берутся середины частичных интервалов, т. е. значения $\tilde{x}_i = \frac{x_{i-1} + x_i}{2}$, m_i – соответствующие им частоты, $i = 1, 2, ..., k$.

- ***Выборочной дисперсией*** $D_в$ *называется среднее арифметическое квадратов отклонений вариант от выборочной средней:*

$$\boxed{D_в[X] = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m_в)^2} = \frac{1}{n} \cdot \sum_{i=1}^{k} (x_i - m_в)^2 \cdot m_i =$$

$$= \sum_{i=1}^{k} (x_i - m_в)^2 \cdot p^*_i = \sum_{i=1}^{k} x_i^2 \cdot p^*_i - (\bar{x})^2 = \overline{x^2} - (\bar{x}_в)^2 .$$

- $\sigma_в[X]=\sqrt{D_в[X]}$ – ***выборочное среднее квадратическое отклонение.***

Чем меньше выборочное квадратическое отклонение, тем лучше выборочная средняя отражает собой всю представляемую совокупность.

Для характеристики меры колеблемости изучаемого признака относительно выборочной средней служит ***коэффициент вариации:***

$$V=\frac{\sigma_в}{\bar{x}}\cdot 100\%.$$

- ***Выборочный начальный момент k–ого порядка:*** $\nu_k^*[X]=\sum_i x_i^k\cdot p_i^*$.

В частности: $\nu_1^*=\bar{x}$

- ***Выборочный центральный момент k–ого порядка:***

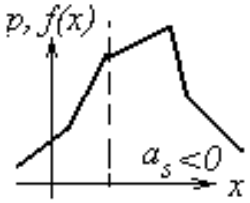
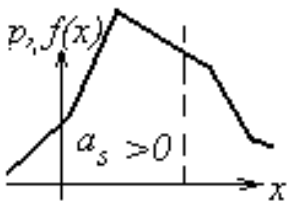
$$\mu_k^*[X]=\sum_i\left(x_i-\bar{x}\right)^k\cdot p_i^* \ .$$

В частности: $\mu_2^*=D_в[X]=\nu_2^*-\left(\nu_1^*\right)^2;$

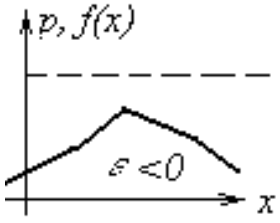
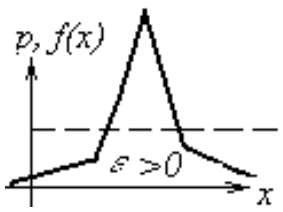
$$\mu_3^*=\nu_3^*-3\nu_1^*\nu_2^*+2\left(\nu_1^*\right)^3;$$

$$\mu_4^*=\nu_4^*+6\left(\nu_1^*\right)^2\cdot\nu_2^*-3\left(\nu_1^*\right)^4-4\nu_3^*\nu_1^*.$$

Выборочный коэффициент асимметрии

$a_s^*[X]=\frac{\mu_3^*[X]}{\sigma_в^3[X]}$		

Выборочный эксцесс

$\varepsilon_s^*[X]=\frac{\mu_4^*[X]}{\sigma_в^4[X]}-3$		

2. Оценка параметров распределения генеральной совокупности

2.1. Выравнивание статистических распределений

Во всяком статистическом распределении число опытов ограничено, что определяет случайный негладкий вид функциональных закономерностей (полигон, гистограмма, кумулята). Необходимо подобрать для данного статистического распределения аналитическую формулу, выражающую лишь существенные черты статистического материала. Такая задача называется задачей выравнивания статистического материала.

Обычно выравниванию подвергаются гистограммы. Принципиальный вид выравнивающей плавной кривой $f(x)$ выбирается исходя из условий возникновения СВ X и из соображений, связанных с внешним видом гистограммы.

В общих случаях, когда зависимость линейная $f(x)=a\,x+b$, квадратичная $f(x)=a\,x^2+b\,x+c,\ldots$, неизвестные параметры $a,b,c,...$ ищут методом наименьших квадратов. При этом аналитическая функция $f(x)$ должна обладать основными свойствами плотности распределения:

$$f(x)\geq 0,\quad \int_{-\infty}^{+\infty} f(x)\,dx=1.$$

Если предполагается, что распределение носит характер частных случаев распределения (биномиальное, нормальное, показательное, равномерное), то параметры выбираются так, чтобы важнейшие моменты (математическое ожидание, дисперсия) статистического и выравнивающего распределений совпадали.

2.2. Точечные оценки параметров распределения генеральной совокупности

Пусть по результатам n опытов необходимо определить (приближённо) некоторый параметр θ, связанный с законом распределения СВ X генеральной совокупности.

- *Приближённое значение параметра θ назовём его* ***оценкой*** θ^*.

Любая оценка θ^*, вычисляемая на основе выборки, сама является случайной величиной. Для того чтобы оценка неизвестного параметра давала хорошее приближение, она должна удовлетворять **требованиям:**

1. несмещённость;
2. состоятельность;
3. эффективность.

- ***Оценка*** *называется* ***несмещённой,*** *если её математическое ожидание по всевозможным выборкам данного объёма равно истинному значению определяемого параметра:* $M\left(\theta^*\right)=\theta$.

Несмещённая оценка обеспечивает близость в среднем значений оценки к значению оцениваемого параметра, то есть не даёт систематической ошибки.

- ***Оценка*** *называется* ***состоятельной,*** *если при увеличении объёма выборки оценка сходится по вероятности к истинному значению параметра:*

$$\lim_{n\to\infty} P\left(\left|\theta^*-\theta\right|<\varepsilon\right)=1 \quad \textit{или} \quad \lim_{n\to\infty}\theta^*=\theta,$$

где ε *– сколь угодно малое положительное число*.

Смысл понятия состоятельности заключается в том, что с увеличением объёма выборки оценка стремится к истинному значению параметра.

Из неравенства Чебышева следует, что для удовлетворения этого требования достаточно, чтобы оценка была несмещённой и $\lim\limits_{n\to\infty} D\left(\theta^*\right)=0$.

- *Оценка называется* ***эффективной****, если она имеет наименьшую дисперсию:* $D\left(\theta^*\right)=D_{min}$.

2.3.Точечная оценка вероятности события

Пусть произошло *n* независимых опытов. СВ X_i – индикатор события A в *i*– ом опыте:

$$X_i=\begin{cases}1, & \text{если } A \text{ произошло;}\\ 0, & \text{если } A \text{ не произошло.}\end{cases}$$

Тогда событие *A* произошло $X=\sum\limits_{i=1}^{n} X_i$ раз, и $(n-X)$ раз оно не произошло. X_i является случайной величиной, имеющей тот же закон распределения, что и СВ X .

Рассмотрим частоту $p^*=\dfrac{X}{n}=\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n} X_i$ как оценку вероятности p появления события.

1) Так как опыты не зависимы, то СВ X распределена по биномиальному закону с параметрами *n* и *p*. Тогда, учитывая формулу $M(X)=n\cdot p$, получим:

$$M\left(p^*\right)=M\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right)=\frac{1}{n}\cdot M\left(\sum_{i=1}^{n} X_i\right)=\frac{1}{n}\cdot M(X)=\frac{1}{n}\cdot n\,p=p \quad \Rightarrow$$

$$\Rightarrow \quad p^* - \text{несмещённая оценка } p.$$

2) По теореме Бернулли имеем

$$\lim_{k\to\infty} P\left(\left|\,p^*-p\,\right|<\varepsilon\right)=1 \qquad \Rightarrow \qquad p^* - \text{состоятельная оценка } p.$$

2.4. Точечная оценка математического ожидания

Рассмотрим выборочную среднюю $\bar{x} = m_в[X]$ как оценку генеральной средней $M_г$ – истинного значения распределения.

1) $M(\bar{x}) = M\left(\frac{1}{n}\cdot\sum_{i=1}^{n} x_i\right) = \frac{1}{n}M\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(x_i) = \frac{1}{n}\sum_{i=1}^{n} M_г =$

$$= \frac{1}{n}\cdot n\cdot M_г = M_г \qquad \Rightarrow \qquad m_в - \text{несмещённая оценка } M_г.$$

2) По теореме Чебышева имеем

$$\lim_{n\to\infty} P\left(|\bar{x} - M_г| < \varepsilon\right) = 1 \qquad \Rightarrow \qquad m_в - \text{состоятельная оценка } M_г.$$

2.5. Точечная оценка дисперсии при неизвестном математическом ожидании

Рассмотрим выборочную дисперсию $D_в[X]$ как оценку генеральной дисперсии $D_г$ – истинного значения распределения.

$$D_в[X] = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{x^2} - (\bar{x}_в)^2 = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\cdot\sum_{i=1}^{n} x_i\right)^2 =$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\cdot\left(\sum_{i=1}^{n} x_i^2 + 2\sum_{\substack{i,j=1\\ i<j}}^{n} x_i\, x_j\right) = \frac{n-1}{n^2}\cdot\sum_{i=1}^{n} x_i^2 - \frac{2}{n^2}\cdot\sum_{\substack{i,j=1\\ i<j}}^{n} x_i\, x_j.$$

Так как статистическая дисперсия не зависит от выбора начала координат, то выберем начало координат в точке $M_г$, т.е. отцентрируем СВ X: $\tilde{X} = X - \bar{x}$. Тогда

$$D_в = \frac{n-1}{n^2}\sum_{i=1}^{n}\tilde{x}_i^2 - \frac{2}{n^2}\sum_{\substack{i,j=1\\ i<j}}^{n}\tilde{x}_i\,\tilde{x}_j.$$

1) $M\left(D_в\right)=\dfrac{n-1}{n^2}\cdot M\left(\sum\limits_{i=1}^{n}\widetilde{x}_i^{\,2}\right)-\dfrac{2}{n^2}\cdot M\left(\sum\limits_{\substack{i,j=1\\ i<j}}^{n}\widetilde{x}_i\,\widetilde{x}_j\right)=$

$$=\frac{n-1}{n^2}\cdot\sum_{i=1}^{n}M\left(\widetilde{x}_i^{\,2}\right)-\frac{2}{n^2}\sum_{\substack{i,j=1\\ i<j}}^{n}M\left(\widetilde{x}_i\,\widetilde{x}_j\right).$$

Но $M\left(\widetilde{x}_i^{\,2}\right)=D_г,\quad M\left(\widetilde{x}_i\,\widetilde{x}_j\right)=0$.

Тогда имеем $M\left(D_в\right)=\dfrac{n-1}{n^2}\cdot\sum\limits_{i=1}^{n}D_г=\dfrac{n-1}{n^2}\cdot n\cdot D_г=\dfrac{n-1}{n}\cdot D_г \qquad \Rightarrow$

$\Rightarrow \qquad D_в$ – смещённая оценка $D_г$.

Смещение оценки произошло потому, что в формуле $D_{\text{в}}[X]=\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)^2$ отклонение X_i отсчитывается не от истинного математического ожидания, а от его статистического аналога $\bar{x}=m_в[X]$. Чтобы исправить этот недостаток, вводят

- ***исправленную выборочную дисперсию*** $s^2[X]=\dfrac{n}{n-1}\cdot D_в[X],$

где s – ***исправленное выборочное квадратическое отклонение***.

2) Учитывая, что $\lim\limits_{n\to\infty}s^2=\lim\limits_{n\to\infty}\dfrac{n}{n-1}\cdot D_в=\lim\limits_{n\to\infty}\dfrac{n}{n-1}\cdot\lim\limits_{n\to\infty}D_в=1\cdot D_г=D_г$,

получим, что $D_в$ – состоятельная оценка $D_г$.

2.6. Точность и надёжность оценок числовых характеристик СВ

При замене $M(X), D(X)$ на их точечные оценки $\bar{x}, s^2$ совершается ошибка.

- *Вероятность* $P\left(\left|\theta^*-\theta\right|<\varepsilon\right)=\gamma$, *что мы не ошибёмся, если поверим оценке, построенной с помощью выборки, называется* ***надёжностью*** *оцениваемого параметра, где* ε – ***точность оценки***.

- ***Доверительным*** *называется* ***интервал*** $\left(\theta^* - \varepsilon\ ;\ \theta^* + \varepsilon\right)$, *который покрывает неизвестный параметр* θ *с надёжностью* γ.

Исходя из центральной предельной теоремы:

$$P\left(\left| Y - M_Y \right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma(Y)}\right),$$

получаем, что чем больше надёжность (вероятность), тем больше аргумент функции Лапласа $\Phi(t)$ и тем больше значение ε (при заданном σ_Y).

Следовательно,

чем больше надёжность, тем шире доверительный интервал, тем больше вероятность ошибки, что $\theta = \theta^*$.

2.7. Односторонние доверительные интервалы

Если интерес представляет ситуация, когда важно сравнение только с одним критическим значением, то используют односторонние доверительные интервалы.

То есть для определённого уровня доверия γ строят двусторонний доверительный интервал, который затем расширяют за счёт одной из его границ.

Для двустороннего доверительного интервала имеем (рис. 6):

$$\gamma = P\left(\left| \theta^* - \theta \right| < \varepsilon\right) = \int_{\theta^* - \varepsilon}^{\theta^* + \varepsilon} f(x)\,dx$$

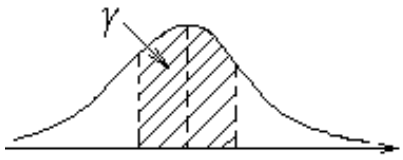


Рисунок 6

Тогда для одностороннего доверительного интервала (рис. 7):

$$P = \int_{-\infty}^{\theta^*+\varepsilon} f(x)dx = \int_{\theta^*-\varepsilon}^{+\infty} f(x)dx = \gamma + \frac{1-\gamma}{2} = \frac{1+\gamma}{2} = \gamma'$$

Рисунок 7

В результате получим односторонний интервал $(-\infty; u_p)$ или $(u_p; +\infty)$ с большей гарантией γ'. Таким образом «односторонний» подход позволяет вдвое снизить ошибку $(1-\gamma)$.

- *Значение* u_p*, для которого выполняется равенство:*

$$P = \int_{-\infty}^{u_p} f(x)dx \qquad \left(\text{или} \quad P = \int_{u_p}^{+\infty} f(x)dx \right)$$ *называется* ***квантилью.***

В статистике для обработки результатов эксперимента широко используют законы распределения Пирсона, Стьюдента, Фишера-Снедекора.

Квантили этих распределений табулированы (см. приложения).

3. Некоторые распределения функций случайных величин

3.1. Распределение χ^2 (хи-квадрат) Пирсона

Распределением Пирсона (χ^2) с k степенями свободы называется распределение суммы квадратов нормально распределённых независимых случайных величин X_1, X_2, …, X_k с параметрами $a = 0$, $\sigma = 1$:

$$\chi^2(k) = X_1^2 + X_2^2 + ... + X_k^2.$$

Распределение χ^2 определяется только одним параметром – числом степеней свободы *k*. Графики функции $f_{\chi_k^2}(x)$ для различных значений *k* представлены на рисунке 8. С увеличением числа степеней свободы *k* ($k \to \infty$) распределение χ^2 приближается к нормальному закону распределения (при $k > 30$ различий практически нет).

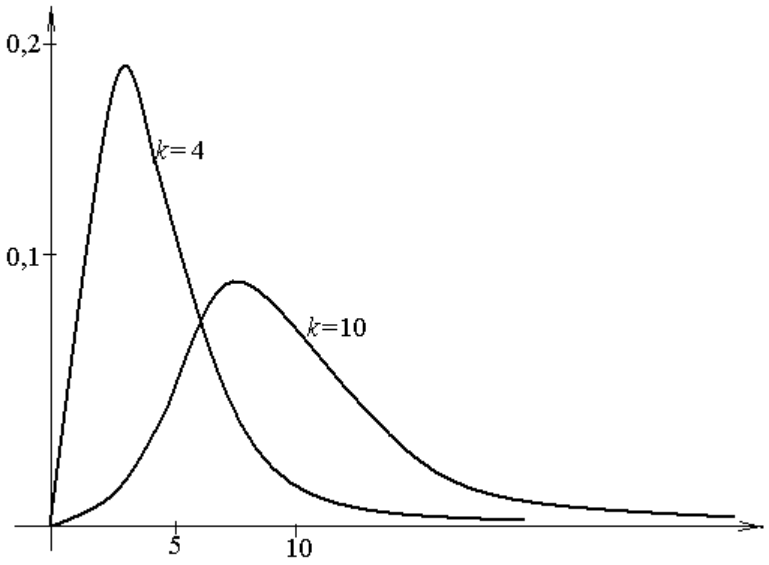


Рисунок 8

Числовые характеристики распределения χ^2:

$$M[\chi^2] = k,\ D[\chi^2] = 2k,\ a_s = \sqrt{\frac{8}{k}},\ \varepsilon_k = \frac{12}{k}.$$

На практике, как правило, используют не плотность вероятности, а квантили распределения χ_k^2.

3.2. Распределение Стьюдента

Пусть $X,\ X_1,\ X_2,\ \ldots,\ X_k$ – независимые случайные величины, имеющие стандартное нормальное распределение с параметрами $a = 0,\ \sigma = 1$.

Распределением Стьюдента (или *t-распределением*) с *k* степенями свободы называется распределение отношения

$$T = \frac{X \cdot \sqrt{k}}{\sqrt{X_1^2 + X_2^2 + ... + X_k^2}} = \frac{X \cdot \sqrt{k}}{\sqrt{\chi^2(k)}}.$$

Кривая распределения Стьюдента (рис. 9) внешне похожа на кривую стандартного нормального распределения.

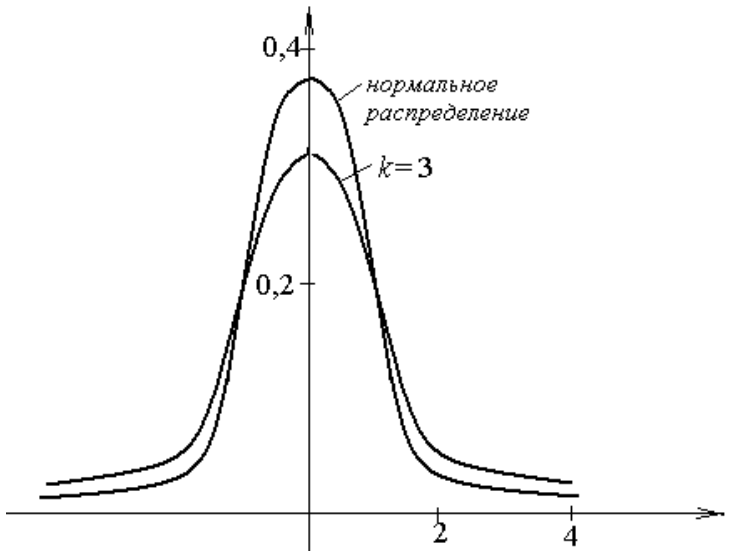


Рисунок 9

Числовые характеристики распределения Стьюдента:

$$M[T] = 0,\ D[T] = \frac{k}{k-2}\ (k > 2),\ a_s = 0,\ \varepsilon_k = \frac{6}{k-4}.$$

3.3. Распределение *Фишера*

Распределением Фишера (или *F-распределением*) со степенями свободы k_1 и k_2 называется распределение отношения

$$F = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}.$$

Кривая распределения Фишера внешне похожа на кривую распределения Пирсона. График плотности *F*-распределения при $k_1 = 10$ и $k_2 = 15$ представлен на рисунке 10.

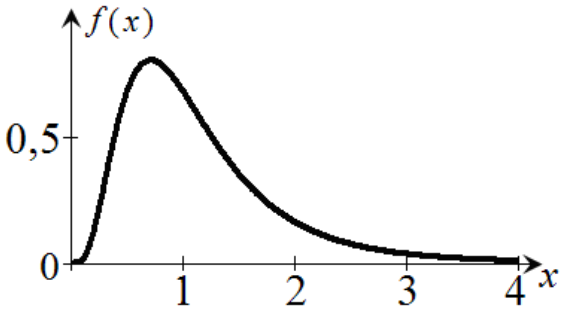


Рисунок 10

Числовые характеристики распределения Фишера-Снедекора:

$$M[F]=\frac{k_2}{k_2-2}\ (k_2>2),\quad D[F]=\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}\ (k_2>4).$$

Замечание. Квантили распределений Пирсона, Стьюдента, Фишера табулированы.

4. Доверительный интервал для оценки параметра

4.1. Доверительный интервал для оценки вероятности события A: $\boxed{p\in\left(p^*-\varepsilon\ ;\ p^*+\varepsilon\right)}$

I случай *(неизвестен объём генеральной совокупности N).*

Пусть событие A наступило m раз в n испытаниях (n – объём выборки). Тогда $p^*=\frac{m}{n}$ – точечная оценка вероятности p наступления события в одном испытании.

По следствию из центральной предельной теоремы, имеем:

$$\gamma=P\left(\left|p^*-p\right|<\varepsilon\right)=2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right)\quad\Rightarrow\quad t=\varepsilon\cdot\sqrt{\frac{n}{pq}}\quad\Rightarrow\quad\varepsilon=t\cdot\sqrt{\frac{pq}{n}},$$

где t – аргумент функции Лапласа: $\Phi(t)=\frac{\gamma}{2}$.

Так как $p\approx p^*$, $q=1-p\approx1-p^*$, то $\boxed{\varepsilon=t\cdot\sqrt{\frac{p^*\cdot\left(1-p^*\right)}{n}}}$.

II случай *(известен объём генеральной совокупности* N *)*:

$$\boxed{\varepsilon = t \cdot \sqrt{\frac{p^* \cdot (1 - p^*)}{n} \cdot \left(1 - \frac{n}{N}\right)}}.$$

Замечание. Формулы справедливы для n, при которых

$$n\,p\,q \approx n\,p^*\left(1 - p^*\right) > 9.$$

4.2. Доверительный интервал для оценки математического ожидания: $\boxed{M_г \in (m_в - \varepsilon\ ;\ m_в + \varepsilon)}$

I случай *(известно среднее квадратическое отклонение* σ *)*

Определим СВ $X = \sum_{i=1}^{n} X_i$. Её среднее выборочное $m_в[X] = \frac{1}{n}\sum_{i=1}^{n} X_i$ представляет собой сумму сравнительно большого числа n независимых величин и, согласно центральной предельной теореме, имеет распределение, близкое к нормальному. При этом:

$M(m_в) = M_г$ (так как $m_в$ – несмещённая оценка $M_г$);

$$D(m_в) = D\left(\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} D(X_i) = \frac{1}{n^2} \cdot n \cdot D_г = \frac{D_г}{n} \quad \Rightarrow$$

$$\sigma(m_в) = \frac{\sigma}{\sqrt{n}}.$$

Тогда

$$\gamma = P\left(|\, m_в - M_г \,| < \varepsilon\right) = 2\,\Phi\left(\frac{\varepsilon}{\sigma(m_в)}\right) = 2\,\Phi\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma}\right) \quad \Rightarrow$$

$$\Rightarrow \quad t = \frac{\varepsilon \cdot \sqrt{n}}{\sigma} \quad \Rightarrow \quad \boxed{\varepsilon = \frac{t \cdot \sigma}{\sqrt{n}}},$$

где t – аргумент функции Лапласа: $\Phi(t) = \frac{\gamma}{2}$.

Замечание:

$$t = \frac{\varepsilon \cdot \sqrt{n}}{\sigma} \quad \Rightarrow \quad \boxed{n = \left(\frac{t \cdot \sigma}{\varepsilon} \right)^2}$$

– данная формула позволяет оценить, каков должен быть объём выборки, чтобы точность оценки $(M_г \approx m_в)$ не превосходила заданного значения ε с заданным уровнем доверия γ.

II случай *(неизвестно σ, неизвестен объём генеральной совокупности N):*

$$\boxed{\varepsilon = \frac{t_\gamma \cdot s}{\sqrt{n}}},$$

где s – несмещённая оценка σ; $t_\gamma = t_{(n-1,\gamma)}$ – табулировано распределением Стьюдента.

- $\boxed{n = \left(\frac{t_\gamma \cdot s}{\varepsilon} \right)^2}$ – объём выборки, необходимый для определения точности оценки $(M_г \approx m_в)$, которая не превосходит заданного значения ε с заданным уровнем доверия γ.

III случай *(неизвестно σ, известен объём генеральной совокупности N)*:

$$\boxed{\varepsilon = t_\gamma \cdot \sqrt{\frac{s^2}{n} \cdot \left(1 - \frac{n}{N}\right)}}.$$

- $\boxed{n = \frac{N \cdot t_\gamma^2 \cdot s^2}{t_\gamma^2 \cdot s^2 + N \cdot \varepsilon^2}}$ – объём выборки, необходимый для определения точности оценки $(M_г \approx m_в)$, которая не превосходит заданного значения ε с заданным уровнем доверия γ.

4.3. Доверительный интервал для оценки среднего квадратического отклонения

Пусть для выборки объёма n задана надёжность, с которой нужно оценить отклонение найденного исправленного квадратического отклонения s от истинного σ: $\gamma = P\left(\left| s - \sigma \right| < \varepsilon\right)$.

Тогда

$$\left|\sigma - s\right| < \varepsilon \;\Rightarrow\; s - \varepsilon < \sigma < s + \varepsilon \;\Rightarrow\; s \cdot \left(1 - \frac{\varepsilon}{s}\right) < \sigma < s \cdot \left(1 + \frac{\varepsilon}{s}\right)$$

Введём $q = \frac{\varepsilon}{s} = q\left(n; \gamma\right)$, которое табулировано.

Учитывая не отрицательность среднего квадратического отклонения, окончательно получим:

Если $q < 1,$	то $\sigma \in \left(s \cdot (1 - q);\ s \cdot (1 + q)\right)$
Если $q > 1$	то $\sigma \in \left(0;\ s \cdot (1 + q)\right)$

5. Проверка статистических гипотез

5.1. Гипотезы. Постановка вопроса

Так как на основе статистического материала вычисляются лишь оценки истинных параметров распределения, то встаёт вопрос о значимости расхождений между статистическими вычислениями и действительностью.

- ***Нулевой гипотезой – H_o*** *называют гипотезу, выдвигаемую в качестве основной, которая отвергает значимость расхождений.*

- *Противоречащую ей гипотезу называют* ***альтернативной (конкурирующей) – H_1.***

Так как решение о справедливости гипотезы принимается на основании выборочных данных, то могут возникать ***ошибки двух родов***:

ошибка I рода	***ошибка II рода***
отвергается ***основная*** *(нулевая) гипотеза, хотя она* ***верна.***	*отвергается* ***конкурирующая*** *гипотеза, хотя она* ***верна.***
вероятность ошибки $P_{H_o}(H_1)=\alpha$ − уровень значимости α.	вероятность ошибки $P_{H_1}(H_o)=\beta$ − уровень значимости β.
вероятность принять верную *(нулевую)* гипотезу: $P_{H_o}(H_o)=1-\alpha$	вероятность принять верную (*конкурирующую)* гипотезу: $P_{H_1}(H_1)=1-\beta$

Вероятность ошибки β второго рода, как правило, не известна, но известно, что при уменьшении ошибки α первого рода ошибка β второго рода, как правило, возрастает.

- ***Статистическим критерием*** (статистикой) *называют СВ T, которая служит для проверки гипотезы.*

- ***Критической областью*** *называют совокупность значений критерия T, при которых нулевая гипотеза отвергается.*

- ***Критическими точками*** $T_{кр}$ *называют точки, отделяющие критическую область от области принятия нулевой гипотезы.*

СВ T имеет симметричный закон распределения с математическим ожиданием, равным 0. Положительные значения $T_{кр}$ табулированы с учётом уровня значимости.

Различают три типа критической области (рис. 11):

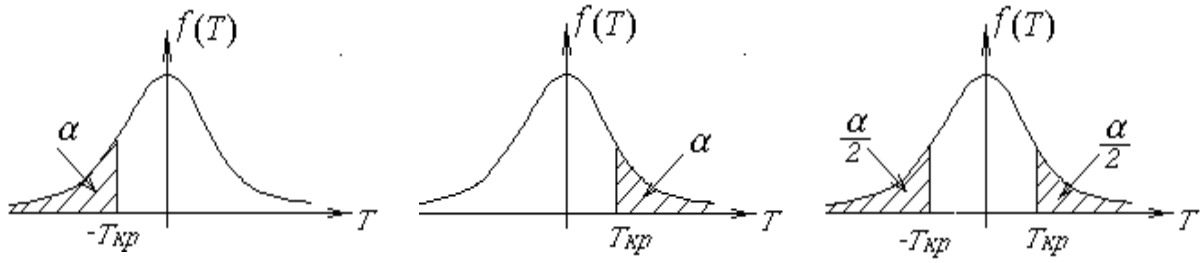


Рисунок 11

$$\omega = \begin{cases} (-\infty; -T_{кр}) & - \text{ левосторонняя критическая область;} \\ (T_{кр}; +\infty) & - \text{ правосторонняя критическая область;} \\ (-\infty; -T_{кр}) \cup (T_{кр}; +\infty) & - \text{ двусторонняя критическая область.} \end{cases}$$

<u>Схема проверки нулевой гипотезы</u>

Определяется критическая область с учётом гипотез:

Нулевая гипотеза	***Альтернативная гипотеза***	***Критическая область*** ω
$\theta = \theta^*$	$\theta \neq \theta^*$	$(-\infty; -T_{кр}) \cup (T_{кр}; +\infty)$
	$\theta < \theta^*$	$(-\infty; -T_{кр})$
	$\theta > \theta^*$	$(T_{кр}; +\infty)$

где θ – истинное значение исследуемого параметра.

По выборочным данным вычисляется число $T_{набл}$.

Из таблиц находится $T_{кр}$ с учётом уровня значимости α, исходя из следующих соотношений:

Для двусторонней ω	$P(T > T_{кр}) = P(T < -T_{кр}) = \frac{\alpha}{2}$
Для левосторонней ω	$P(T < -T_{кр}) = \alpha$
Для правосторонней ω	$P(T > T_{кр}) = \alpha$

Делают вывод: если $T \in \omega$, то H_o отвергается (принимается H_1);

если $T \notin \omega$, то H_o принимается (отвергается H_1).

5.2. Статистики сравнения точечных оценок неизвестных генеральных

	При не известном $\sigma_г$	*При известном* $\sigma_г$
Сравнение **выборочной средней и математического ожидания** $H_o : (M = a)$.	$T = \dfrac{(m_в - a)\cdot\sqrt{n-1}}{s}$ $T_{кр} = T\begin{pmatrix}\alpha\ \left(или\ \dfrac{\alpha}{2}\right)\\ l = n-1\end{pmatrix}$	$T = \dfrac{(m_в - a)\cdot\sqrt{n}}{\sigma}$ $\Phi(T_{кр}) = P(0 < t < T_{кр}) =$ $= 0{,}5 - \alpha\ \left(или\ \dfrac{\alpha}{2}\right)$
Сравнение **математических ожиданий двух выборок** с объёмами n_x и m_y $H_o : (M[X] = M[Y])$. Если принимается H_o, то общая средняя оценивается величиной $m_в = \dfrac{n\cdot m_в[X] + m\cdot m_в[Y]}{n+m}$	$T = \dfrac{m_в[X] - m_в[Y]}{S\cdot\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}$ где $S = \sqrt{\dfrac{s_x^2\cdot(n-1) + s_y^2\cdot(m-1)}{n+m-2}}$ $T_{кр} = T\begin{pmatrix}\alpha\ \left(или\ \dfrac{\alpha}{2}\right)\\ l = n+m-2\end{pmatrix}$	$T = \dfrac{m_в[X] - m_в[Y]}{\sqrt{\dfrac{D_x}{n}+\dfrac{D_y}{m}}}$ $\Phi(T_{кр}) = P(0 < t < T_{кр}) =$ $= 0{,}5 - \alpha\ \left(или\ \dfrac{\alpha}{2}\right)$
	$T_{кр}$ – из распределения Стьюдента	$T_{кр}$ – аргумент функции Лапласа
Сравнение **дисперсий двух выборок** с объёмами n_x и m_y $H_o : (D_x = D_y)$, где $s_x^2 > s_y^2$. Если принимается H_o, то генеральная дисперсии оценивается величиной $s^2 = \dfrac{(n-1)\cdot s_x^2 + (m-1)\cdot s_y^2}{n+m-2}$	$H_1 : D_x > D_y$ ***(критическая область – правосторонняя)*** $F = \dfrac{s_x^2}{s_y^2}$ $F_{кр} = F\begin{pmatrix}\alpha\\ l_1 = n-1\\ l_2 = m-1\end{pmatrix}$	$H_1 : D_x \neq D_y$ ***(критическая область – двусторонняя)*** $F = \dfrac{s_x^2}{s_y^2}$ $F_{кр} = F\begin{pmatrix}\alpha/2\\ l_1 = n-1\\ l_2 = m-1\end{pmatrix}$
	$F_{кр}$ – из распределения Фишера–Снедекора	

5.3. Построение теоретического закона распределения СВ

Пусть на основе статистических данных был сделан выбор теоретического закона распределения СВ. Для частных случаев распределения закон содержит параметры, которые имеют в качестве своих оценок выборочное среднее и исправленное среднее квадратическое отклонение.

Дискретная СВ (ДСВ)

Произведено n опытов. Каждый опыт состоит из k независимых испытаний, в каждом из которых вероятность (p) появления события A постоянна. ДСВ X – число появлений события A описывается рядом:

x_i	0	1	…	k
m_i	m_0	m_1		m_k

$$\sum_{i=0}^{k} m_i = n$$

$$m_в[X] = \frac{\sum_{i=1}^{n} x_i m_i}{n}, \qquad m_в[X^2] = \frac{\sum_{i=1}^{n} x_i^2 m_i}{n}, \qquad s^2 = \frac{n}{n-1} \cdot \left(m_в[X^2] - m_в{}^2 \right)$$

Биномиальный закон распределения $p \in (0{,}1;\ 0{,}9)$ $m_в \neq s^2$	k – невелико	$P_i = P_k(x_i) = C_k^{x_i} \cdot p^{x_i} \cdot q^{k-x_i}$ $p \approx p^* = \frac{\sum x_i \omega_i}{k} = \frac{\sum x_i m_i}{n \cdot k}$
	k – велико	$P_i = P_k(x_i) = \frac{\varphi(t)}{\sqrt{k\,p\,q}}, \qquad t = \frac{x_i - k\,p}{\sqrt{k\,p\,q}}$ $p \approx p^* = \frac{\sum x_i m_i}{n \cdot k}$
Закон распределения Пуассона $p \notin (0{,}1;\ 0{,}9)$ $m_в \approx s^2$	k – велико	$P_i = P_k(x_i) = \frac{\lambda}{x_i!} e^{-\lambda}$ $\lambda \approx \bar{x}$

Непрерывная СВ (НСВ)

Произведено n независимых испытаний.

В результате составлен интервальный вариационный ряд СВ X:

$(x_i ; x_{i+1})$	$(x_0 ; x_1)$	$(x_1 ; x_2)$	…	$(x_{k-1} ; x_k)$
m_i	m_0	m_1		m_{k-1}
$x_i^* = \frac{x_i + x_{i+1}}{2}$	x_0^*	x_1^*		x_{k-1}^*

$$\sum_{i=0}^{k-1} m_i = n$$

$$m_в[X] = \frac{\sum_{i=1}^{k-1} x_i^* m_i}{n}, \qquad m_в[X^2] = \frac{\sum_{i=1}^{n} (x_i^*)^2 m_i}{n}, \qquad s^2 = \frac{n}{n-1} \cdot \left(m_в[X^2] - (m_в)^2\right)$$

Равномерное распределение $\begin{cases} \frac{a+b}{2} = m_в \\ \frac{(b-a)^2}{12} = s^2 \end{cases} \Rightarrow \begin{cases} a^* = m_в - s\sqrt{3} \\ b^* = m_в + s\sqrt{3} \end{cases}$

$f(x) = \frac{1}{b^* - a^*}$ если $x \in [a^*; b^*]$	$F(x) = \begin{cases} 0, & x < a^* \\ \frac{x - a^*}{b^* - a^*}, & x \in [a^*; b^*] \\ 1, & x > b^* \end{cases}$	$P_i = P(x_{i-1} < x < x_i) = \frac{x_i - x_{i-1}}{b^* - a^*} = f(x) \cdot (x_i - x_{i-1})$ где $i = 1,2,...k$; $x_0 = a^*$, $x_k = b^*$

Показательное распределение $\lambda^* = \frac{1}{m_в}$

$f(x) = \lambda^* \cdot e^{-\lambda^* \cdot x}$, если $x > 0$	$F(x) = 1 - e^{-\lambda^* x}$, если $x > 0$	$P_i = P(x_{i-1} < x < x_i) = e^{-\lambda^* x_{i-1}} - e^{-\lambda^* x_i}$, где $i = 1,2,...k$

Нормальное распределение $a^* = m_в$, $\sigma^* = s$

$f(x) = \frac{1}{\sigma^*} \cdot \varphi(t)$, $t = \frac{x - a^*}{\sigma^*}$, $\varphi(t)$ – функция Гаусса	$F(x) = 0{,}5 + \Phi(t)$, $t = \frac{x - a^*}{\sigma^*}$, $\Phi(t)$ – функция Лапласа	$P_i = P(x_{i-1} < x < x_i) = \Phi(t_i) - \Phi(t_{i-1}) = \frac{x_i - x_{i-1}}{\sigma^*} \cdot \varphi(t_i^*)$, где $t_i^* = \frac{x_i^* - a^*}{\sigma^*}$, $i = 1,2,... k$

5.4. Критерий Колмогорова для проверки гипотезы о виде интегральной функции распределения непрерывной СВ

Алгоритм действий:

1. По выборке найти эмпирическую функцию распределения $F^*(x)$.
2. Выбрать закон распределения непрерывной СВ, для него по эмпирическим характеристикам составить теоретическую интегральную функцию $F(x)$.
3. Составить таблицу значений F^* и F для имеющихся значений вариант, вычислить модуль разности $(F^* - F)$.
4. Выбрать $\max |F^* - F| = D$ и найти статистику $\lambda = D\sqrt{n}$,

   где n – объём выборки.
5. По таблице статистики Колмогорова найти величину

$$P(\lambda) = 1 - \sum_{n=-\infty}^{+\infty} (-1)^n e^{-2n^2\lambda^2}.$$

- *Если $P(\lambda)$ имеет сравнительно большее значение, то закон теоретического распределения выбран правильно.*

5.5. Критерий Пирсона для проверки гипотезы о виде дифференциальной функции распределения СВ

- ***Эмпирические кратности*** *– это кратности m_i, наблюдаемые в эксперименте.*
- ***Выравнивающие кратности*** *– это кратности, которые находятся по формуле $m_i' = n \cdot P_i$*, где n – объём выборки, P_i – точечная вероятность варианты x_i дискретной СВ или интервальная вероятность для варианты $x \in (x_{i-1}; x_i)$ непрерывной СВ.

Алгоритм действий:

1. Выбрать закон распределения СВ и для него по эмпирическим характеристикам найти теоретическую дифференциальную функцию $f(x)$.
2. По соответствующей формуле вычислить точечную (или интервальную) вероятность P_i.
3. Вычислить выравнивающие кратности $m_i' = n \cdot P_i$, где n – объём выборки.
4. Найти статистику $\boxed{\chi^2_{набл} = \sum_{i=1}^{n} \frac{(m_i - m_i')^2}{m_i'}}$.
5. Определить число $l = k - r - 1$,

   где k – число частичных интервалов выборки;

   r – число параметров дифференциальной функции распределения.

В частности:

Биномиальный закон распределения СВ	$l = k - 1$, если p_A известно
	$l = k - 2$, если p_A неизвестно
Закон распределения Пуассона	$l = k - 2$
Равномерный закон распределения СВ	$l = k - 3$
Показательный закон распределения СВ	$l = k - 2$
Нормальный закон распределения СВ	$l = k - 3$

6. По таблице найти критическую величину $\chi^2_{кр} = \chi^2(\alpha; l)$, где α – уровень значимости.

- *Если $\chi^2_{набл} < \chi^2_{кр}$, то закон теоретического распределения выбран правильно.*

6. Теория корреляции

6.1. Понятие о корреляционной зависимости

Имеется две СВ X и Y. Данные их значений внесёны в таблицу:

	x_1	x_2	…	x_k	m_{y_j}
y_1	$m_{x_1 y_1}$	$m_{x_2 y_1}$		$m_{x_k y_1}$	$m_{y_1} = \sum_i m_{x_i y_1}$
y_2	$m_{x_1 y_2}$	$m_{x_2 y_2}$		$m_{x_k y_2}$	$m_{y_2} = \sum_i m_{x_i y_2}$
…					
y_l	$m_{x_1 y_l}$	$m_{x_2 y_l}$		$m_{x_k y_l}$	$m_{y_l} = \sum_i m_{x_i y_l}$
m_{x_i}	$m_{x_1} = \sum_j m_{x_1 y_{ji}}$	$m_{x_2} = \sum_j m_{x_2 y_j}$		$m_{x_k} = \sum_j m_{x_k y_j}$	$\sum_i m_{x_i} = \sum_j m_{y_j} = n$

где m_{y_j} – частота появления варианты y_j, m_{x_i} – частота появления варианты x_i, $m_{x y_j}$ – частота появления варианты y_j при заданном значении варианты x, $m_{y x_i}$ – частота появления варианты x_i при заданном значении варианты y, n – объём выборки.

Определим некоторые первичные понятия:

Для СВ Y	**Для СВ X**
Общее среднее	
– это среднее арифметическое всех значений СВ Y: $\overline{y} = \frac{\sum_j y_j \cdot m_{y_j}}{\sum_j m_{y_j}} = \frac{\sum_j y_j \cdot m_{y_j}}{n}$	– это среднее арифметическое всех значений СВ X: $\overline{x} = \frac{\sum_i x_i \cdot m_{x_i}}{\sum_i m_{x_i}} = \frac{\sum_i x_i \cdot m_{x_i}}{n}$
Условное среднее	
− это среднее арифметическое тех значений СВ Y, которые соответствуют значению СВ $X = x$: $\overline{y_x} = \frac{\sum_j y_j \cdot m_{x y_j}}{\sum_j m_{x y_j}}$	− это среднее арифметическое тех значений СВ X, которые соответствуют значению СВ $Y = y$: $\overline{x_y} = \frac{\sum_i x_i \cdot m_{y x_i}}{\sum_i m_{y x_i}}$

общее среднее квадратическое отклонение	
$\sigma_Y^2 = \overline{y^2} - \left(\overline{y}\right)^2$	$\sigma_X^2 = \overline{x^2} - \left(\overline{x}\right)^2$
межгрупповое среднее квадратическое отклонение	
$\delta_Y^2 = \dfrac{\sum\limits_i m_{x_i} \cdot \left(\overline{y_{x_i}} - \overline{y}\right)^2}{n}$	$\delta_X^2 = \dfrac{\sum\limits_j m_{y_j} \cdot \left(\overline{x_{y_j}} - \overline{x}\right)^2}{n}$

- ***Статистической зависимостью*** *называется зависимость, при которой изменение одной СВ влечёт изменение другой СВ.*
- ***Статистическую зависимость*** *называют* ***корреляционной,*** *если изменение одной СВ влечёт изменение среднего значения другой СВ.*
- *Уравнение, определяющее корреляционную зависимость* $\overline{y_x} = f(x)$**,** *называют* ***уравнением регрессии*** *СВ* Y *на СВ* X *.*
- $\overline{x_y} = g(y)$ **–** ***уравнение регрессии*** *СВ* X *на СВ* Y *.*
- *Если обе функции регрессии* $f(x)$ *и* $g(y)$ *линейны, то корреляционную* ***зависимость*** *называют* ***линейной****.*

6.2 Теснота корреляционной связи

- ***Корреляционным отношением*** *СВ* Y *на СВ* X *(или* X *на* Y*) называется отношение межгруппового среднего квадратического отклонения СВ* Y *(или* X*) к общему среднему квадратическому отклонению этой СВ:*

$$\boxed{\eta_{YX} = \frac{\delta_Y}{\sigma_Y}} \quad (\text{или } \eta_{XY} = \frac{\delta_X}{\sigma_X}).$$

Корреляционное отношение $\eta \in [0;1]$ служит для **оценки тесноты (наличия)** корреляционной связи между СВ X и Y:

Значения η	**Зависимость СВ X и Y**
$\eta = 0$	СВ X и Y не связаны корреляционной зависимостью
$0 < \eta < 0{,}3$	Функциональная корреляционная зависимость практически отсутствует
$0{,}3 < \eta < 0{,}5$	Слабая функциональная корреляционная зависимость
$0{,}5 < \eta < 0{,}7$	Умеренная функциональная корреляционная зависимость
$0{,}7 \le \eta < 1$	Умеренно сильная функциональная корреляционная зависимость
$\eta = 1$	СВ X и Y связаны функциональной корреляционной зависимостью

6.3. Линейная регрессия

I. Поиск уравнения связи

A) Различные значения СВ X и соответствующие значения СВ Y наблюдались по одному разу.

В результате n испытаний получены выборочные пары чисел

$$(x_1; y_1), (x_2; y_2), \ldots, (x_n; y_n),$$

которые располагаются на графике вдоль некоторой прямой $y = k x + b$.

- *Угловой коэффициент прямой линии регрессии $Y = k x + b$ называют* ***выборочным коэффициентом регрессии*** *Y на X* и обозначают: $k = \rho_{Y X}$.

Найдём неизвестные параметры b и $\rho_{Y X}$ по методу наименьших квадратов (МНК). Составим функцию:

$$F(\rho, b) = \sum_i \left(Y(x_i) - y_i \right)^2 = \sum_i \left(\rho_{Y X} x_i + b - y_i \right)^2$$

и решим систему относительно её частных производных:

$$\begin{cases} \dfrac{\partial F}{\partial b} = 0, \\ \dfrac{\partial F}{\partial \rho} = 0 \end{cases} \Rightarrow \begin{cases} 2 \cdot \sum\limits_i \left(\rho_{YX} x_i + b - y_i \right) \cdot 1 = 0, \\ 2 \cdot \sum\limits_i \left(\rho_{YX} x_i + b - y_i \right) \cdot x_i = 0. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \rho_{YX} \cdot \bar{x} + b = \bar{y}, \\ \rho_{YX} \cdot \overline{x^2} + b \cdot \bar{x} = \overline{xy} \end{cases} \Rightarrow \begin{cases} b = \bar{y} - \rho_{YX} \cdot \bar{x}, \\ \rho_{YX} \cdot \overline{x^2} + \left(\bar{y} - \rho_{YX} \cdot \bar{x} \right) \cdot \bar{x} = \overline{xy} \end{cases} \Rightarrow$$

$$\boxed{\Rightarrow \quad \rho_{YX} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \left(\bar{x} \right)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_X^2}}.$$

В итоге уравнение прямой линии регрессии примет вид:

$$\boxed{Y - \bar{y} = \rho_{YX} \cdot \left(x - \bar{x} \right)} \ ,$$

где $\boxed{\rho_{YX} = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_X^2}}$.

В) Сгруппированные значения СВ *X* и соответствующие значения СВ *Y*

В результате n испытаний получены выборочные пары чисел $\left(x_1 ; \overline{y_{x_1}} \right), \left(x_2 ; \overline{y_{x_2}} \right), ..., \left(x_n ; \overline{y_{x_n}} \right)$, которые располагаются на графике вдоль некоторой прямой линии регрессии $\overline{Y_X} = \rho_{YX} \ x + b$.

Проводя рассуждения по методу МНК, получим уравнение:

$\boxed{\overline{Y_X} - \bar{y} = \rho_{YX} \cdot \left(x - \bar{x} \right)}$, где $\rho_{YX} = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_X^2}$.

Видоизменим полученную зависимость.

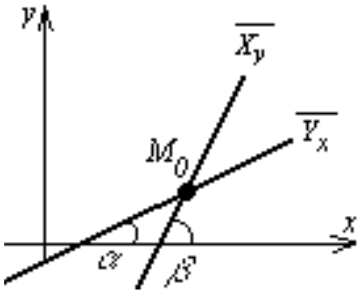
$$\rho_{YX} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_X^2} \ \Bigg| \cdot \frac{\sigma_X}{\sigma_Y} \quad \Rightarrow \quad \frac{\rho_{YX} \cdot \sigma_X}{\sigma_Y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_Y \cdot \sigma_X} \overset{\text{обозначим}}{=} r_в ;$$

$$\rho_{XY} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_Y^2} \ \Bigg| \cdot \frac{\sigma_Y}{\sigma_X} \quad \Rightarrow \quad \frac{\rho_{XY} \cdot \sigma_Y}{\sigma_X} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_Y \cdot \sigma_X} \overset{\text{обозначим}}{=} r_в \ .$$

Получили систему: $\begin{cases} \dfrac{\rho_{YX}\cdot\sigma_X}{\sigma_Y}=r_в, \\ \dfrac{\rho_{XY}\cdot\sigma_Y}{\sigma_X}=r_в. \end{cases} \Rightarrow \begin{cases} \rho_{YX}=\dfrac{r_в\cdot\sigma_Y}{\sigma_X}, \\ \rho_{XY}=\dfrac{r_в\cdot\sigma_X}{\sigma_Y}. \end{cases}$

В итоге уравнение прямой линии регрессии примет вид:

$$\boxed{\begin{array}{l} \overline{Y_X}-\bar{y}=\dfrac{r_в\cdot\sigma_Y}{\sigma_X}\cdot\left(x-\bar{x}\right), \\ \textit{или} \quad \overline{X_Y}-\bar{x}=\dfrac{r_в\cdot\sigma_X}{\sigma_Y}\cdot\left(y-\bar{y}\right) \end{array}}, \textit{ где } \boxed{r_в=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_Y\cdot\sigma_X}}.$$



$M_0\left(\bar{x};\bar{y}\right)$ – точка пересечения двух линий регрессий.

Для регрессии Y на X: $k_1=tg\,\alpha=\dfrac{r_в\cdot\sigma_Y}{\sigma_X}$;

для регрессии X на Y: $k_2=tg\,\beta=\dfrac{\sigma_Y}{r_в\cdot\sigma_X}$.

II. Теснота линейной корреляционной связи

Введённый коэффициент $r_в=\dfrac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_Y\cdot\sigma_X}$ является характеристикой линейности уравнения регрессии.

При этом: $\begin{cases} \dfrac{\rho_{YX}\cdot\sigma_X}{\sigma_Y}=r_в, \\ \dfrac{\rho_{XY}\cdot\sigma_Y}{\sigma_X}=r_в \end{cases} \Rightarrow \quad r_в^2=\rho_{YX}\cdot\rho_{XY}$

- ***Выборочным коэффициентом корреляции*** *переменных X и Y, между которыми предполагается линейная корреляционная связь, называется среднее геометрическое их коэффициентов регрессии и имеющее знак последних:* $r_в=\pm\sqrt{\rho_{YX}\cdot\rho_{XY}}$, $|r_в|\le\eta$.

Значения $r_в$	**Зависимость СВ X и Y**

$\lvert r_в \rvert = \eta$	Точная линейная корреляционная зависимость.

$0 < r_в < 1$	Прямая связь линейной зависимости (возрастающая функция)
$-1 < r_в < 0$	Обратная связь линейной зависимости (убывающая функция)
$r_в = 0$	Отсутствует линейная зависимость (графики уравнений регрессий взаимно перпендикулярны)

$0 < \lvert r_в \rvert < 0{,}3$	Линейная зависимость практически отсутствует
$0{,}3 < \lvert r_в \rvert < 0{,}5$	Слабая линейная зависимость
$0{,}5 < \lvert r_в \rvert < 0{,}7$	Умеренная линейная зависимость
$0{,}7 \le \lvert r_в \rvert < 1$	Сильная линейная зависимость
$\lvert r_в \rvert = 1$	Функциональной линейной зависимостью (графики уравнений регрессий совпадают)

III. Проверка гипотезы о наличии линейной зависимости между наблюдаемыми значениями СВ X и Y

Если двумерная генеральная совокупность распределена нормально и из этой совокупности извлечена выборка объёма n, для которой найден коэффициент корреляции $r_в$, то проверяют

− **нулевую гипотезу** $H_o : r_в = 0$ − линейная зависимость отсутствует

− при **альтернативн**ой $H_1 : r_в \neq 0$ − линейная зависимость присутствует.

<u>Для этого:</u>

1. Вычисляют статистику $T_{набл} = \dfrac{r_в \cdot \sqrt{n-2}}{\sqrt{1-r_в^2}}$.

2. Определяют по таблицам распределения Стьюдента $t_{кр} = t_{кр}(\alpha; l = n-2)$. Критической областью при этом есть промежуток

$$\omega = (-\infty; -t_{кр}) \cup (t_{кр}; +\infty).$$

3. Делают вывод:

$T_{набл} \in \omega$	гипотезу H_o отвергают, т. е. СВ X и Y линейно коррелированны
$T_{набл} \notin \omega$	гипотезу H_o принимают, т. е. СВ X и Y линейно не коррелированны

6.4. Нелинейные корреляционные связи

А) Параболическая корреляция

В результате n испытаний получены выборочные пары чисел

$$(x_1; \overline{y_{x_1}}), (x_2; \overline{y_{x_2}}), \ldots, (x_n; \overline{y_{x_n}}),$$

которые располагаются на графике вдоль некоторой кривой линии регрессии типа: $\overline{Y_x} = Ax^2 + Bx + C$.

Неизвестные параметры A, B, C найдём по методу МНК.

Составим функцию:

$$F(A,B,C) = \sum_i \left(\overline{Y}(x_i) - \overline{y_{x_i}}\right)^2 = \sum_i \left(A x_i^2 + B x_i + C - \overline{y_{x_i}}\right)^2$$

и решим систему относительно её частных производных:

$$\begin{cases} A \cdot \sum_i x_i^2 m_{x_i} + B \cdot \sum_i x_i m_{x_i} + C \cdot n = \sum_i \overline{y_{x_i}} \cdot m_{x_i}, \\ A \cdot \sum_i x_i^3 m_{x_i} + B \cdot \sum_i x_i^2 m_{x_i} + C \cdot \sum_i x_i m_{x_i} = \sum_i x_i \overline{y_{x_i}} \cdot m_{x_i}, \\ A \cdot \sum_i x_i^4 m_{x_i} + B \cdot \sum_i x_i^3 m_{x_i} + C \cdot \sum_i x_i^2 m_{x_i} = \sum_i x_i^2 \overline{y_{x_i}} \cdot m_{x_i} \end{cases}$$

Решение этой системы определяет искомые неизвестные A, B, C.

В) Гиперболическая корреляция

В результате n испытаний получены выборочные пары чисел

$$\left(x_1;\overline{y_{x_1}}\right),\left(x_2;\overline{y_{x_2}}\right),\ldots,\left(x_n;\overline{y_{x_n}}\right),$$

которые располагаются на графике вдоль некоторой кривой линии регрессии типа: $\overline{Y_x}=\dfrac{A}{x}+B.$

Неизвестные параметры A,B найдём по методу МНК.

Составим функцию: $F(A,B)=\sum\limits_i\left(\overline{Y}(x_i)-\overline{y_{x_i}}\right)^2=\sum\limits_i\left(\dfrac{A}{x_i}+B-\overline{y_{x_i}}\right)^2,$

и решим систему относительно её частных производных:

$$\begin{cases} A\cdot\sum\limits_i\dfrac{1}{x_i}\cdot m_{x_i}+B\cdot n=\sum\limits_i\overline{y_{x_i}}\cdot m_{x_i}, \\ A\cdot\sum\limits_i\dfrac{1}{x_i^2}\cdot m_{x_i}+B\cdot\sum\limits_i\dfrac{1}{x_i}\cdot m_{x_i}=\sum\limits_i\dfrac{\overline{y_{x_i}}}{x_i}\cdot m_{x_i} \end{cases}$$

Решение этой системы определяет искомые неизвестные A,B.

Замечание: Аналогично можно рассуждать и для другого вида корреляционной связи.

6.5. Множественная корреляция

- *Если исследуется связь между более, чем двумя СВ, то* ***корреляцию*** *называют* ***множественной.***

В простейшем случае, когда число СВ равно трём, анализируют линейную зависимость: $z=Ax+By+C$.

Для этого:

1. По выборочным данным наблюдений определяют (по методу МНК) неизвестные параметры A,B,C и составляют уравнение связи;

2. Находят выборочный совокупный коэффициент корреляции, который оценивает тесноту связи между Z и X, Y;
3. Находят частные выборочные коэффициенты корреляции, которые оценивают тесноту связи между Z и X при постоянном Y, и между Z и Y при постоянном X.

Эти коэффициенты имеют те же свойства и тот же смысл, что и выборочный коэффициент корреляции двух СВ, т. е. служат для оценки линейной связи между СВ.

7. Практикум по решению задач математической статистики

7.1. Первичная обработка выборки

Пример 1. (Дискретный случай). В супермаркете проводились наблюдения над числом X покупателей, обратившихся в кассу за один час. Наблюдения в течение 30 часов (15 дней с 9 до 10 и с 10 до 11 часов) дали следующие результаты: 80, 90, 100, 110, 90, 60, 100, 110, 80, 60, 70, 100, 70, 100, 80, 90, 60, 100, 100, 110, 80, 90, 80, 110, 70, 80, 90, 80, 100, 100. Требуется составить ряд распределения частот (вариационный ряд). Построить полигон относительных частот и эмпирическую функцию распределения.

Решение. Число X является дискретной случайной величиной, а полученные данные представляют собой выборку из $n = 30$ наблюдений.

Сначала составим ранжированный ряд: 60, 60, 60, 70, 70, 70, 80, 80, 80, 80, 80, 80, 80, 90, 90, 90, 90, 90, 100, 100, 100, 100, 100, 100, 100, 100, 110, 110, 110, 110.

Получено шесть различных значений (шесть вариант) случайной величины (СВ). Подсчитав частоту значений каждой варианты, составим таблицу, которая и будет представлять собой вариационный ряд.

Число обращений в кассу, x_i	60	70	80	90	100	110	Σ
Частота (кратность), m_i	3	3	7	5	8	4	$n = \sum_{i=1}^{6} m_i = 30$
Относительная частота, $\omega_i = p_i^* = \frac{m_i}{n}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{7}{30}$	$\frac{5}{30}$	$\frac{8}{30}$	$\frac{4}{30}$	$\sum_{i=1}^{6} p_i^* = 1$

Используя данные таблицы для точек $(x_i; p_i^*)$, построим полигон (рис.12).

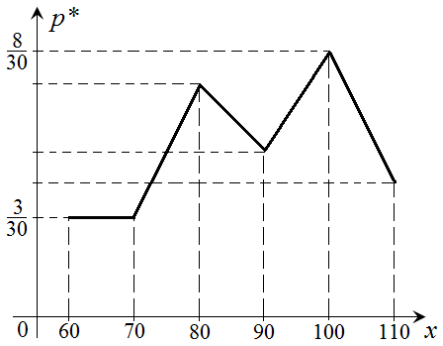


Рисунок 12

Вычислим значения эмпирической функции:

$F^*(x)=\frac{0}{30}=0$ при $x\le 60$ (наблюдений меньше 60 нет);

$F^*(x)=\frac{3}{30}$ при $60<x\le 70$;

$F^*(x)=\frac{3}{30}+\frac{3}{30}=\frac{6}{30}$ при $70<x\le 80$ и т. д.

$$\Rightarrow F^*(x)=\begin{cases}0, & \text{при } x\le 60,\\ \frac{3}{30}, & \text{при } 60<x\le 70,\\ \frac{6}{30}, & \text{при } 70<x\le 80,\\ \frac{13}{30}, & \text{при } 80<x\le 90,\\ \frac{18}{30}, & \text{при } 90<x\le 100,\\ \frac{26}{30}, & \text{при } 100<x\le 110,\\ 1, & \text{при } x>110.\end{cases}$$

График функции $F^*(x)$ представлен на рисунке 13.

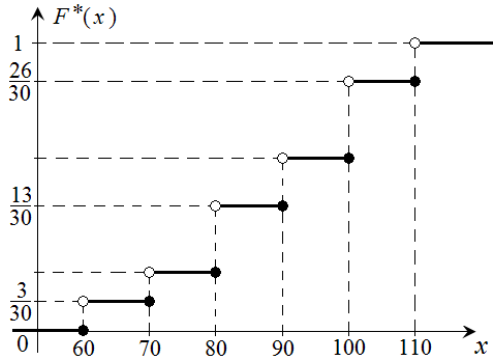


Рисунок 13

Пример 2. (Непрерывный случай). В таблице приведена выборка результатов измерения роста 105 студентов (юношей). Измерения проводились с точностью до 1 см. Требуется составить вариационный ряд. Построить гистограмму и эмпирическую функцию распределения.

155	170	185	180	188	152	173	178	178	168	185	172	170	183	175
173	170	183	175	180	175	193	178	183	180	196	178	181	187	168
174	179	184	183	178	180	178	163	166	178	175	182	190	167	170
178	183	170	178	181	173	168	185	175	170	155	169	186	179	189
156	174	179	179	169	186	174	171	184	175	193	178	184	180	196
175	181	188	168	179	178	183	184	178	181	177	163	166	178	175
183	190	167	170	178	183	170	178	182	173	168	186	176	171	188

Решение. Определим СВ «Рост юношей» как непрерывную, т.к. объём выборки $n = 105$ достаточно велик, при этом значения в выборке разнообразны.

Тогда количество интервалов: $k \approx 1 + \log_2 n = 1 + \log_2 105 = 7{,}714 \approx 8$.

Учитывая, что $x_{\min} = 152$, $x_{\max} = 196$, находим длину частичного интервала: $\dfrac{x_{\max} - x_{\min}}{k} = \dfrac{196 - 152}{8} \le 6 \quad \Rightarrow \quad h = 6$.

Примем $x_0 = x_{\min} - \dfrac{h}{2} = 152 - \dfrac{6}{2} = 152 - 3 = 149$.

Исходные данные разбиваем на 8 интервалов, начиная с 149:

$$[149;155),\ [155;161),\ \ldots,\ [191;197].$$

Подсчитав число студентов m_i, попавших в каждый из полученных промежутков, получим интервальный вариационный ряд.

Рост, x_i	149-155	155-161	161-167	167-173	173-179	179-185	185-191	191-197
m_i	1	3	4	21	31	28	13	4
$p_i^* = \dfrac{m_i}{n}$	0,0095	0,0286	0,0381	0,2	0,2952	0,2667	0,1238	0,0381

Здесь $n = \sum m_i = 105$.

Доопределим середины частичных интервалов

$$\tilde{x}_i = \frac{x_{i-1} + x_i}{2},\ i = 1, 2, ..., k.$$

Таблица примет вид:

Рост, $[x_{i-1}; x_i)$	149-155	155-161	161-167	167-173	173-179	179-185	185-191	191-197
середина, $\tilde{x}_i$	152	158	164	170	176	182	188	194
частота, m_i	1	3	4	21	31	28	13	4
$p_i^* = \frac{m_i}{n}$	0,0095	0,0286	0,0381	0,2	0,2952	0,2667	0,1238	0,0381

Гистограмма относительных частот интервального ряда представляет собой ступенчатую фигуру (рис. 14), состоящую из прямоугольников с основаниями длиной $h = 6$ и высотами $f_i^* = \frac{p_i^*}{6},\ i = 1, 2, ..., 8$.

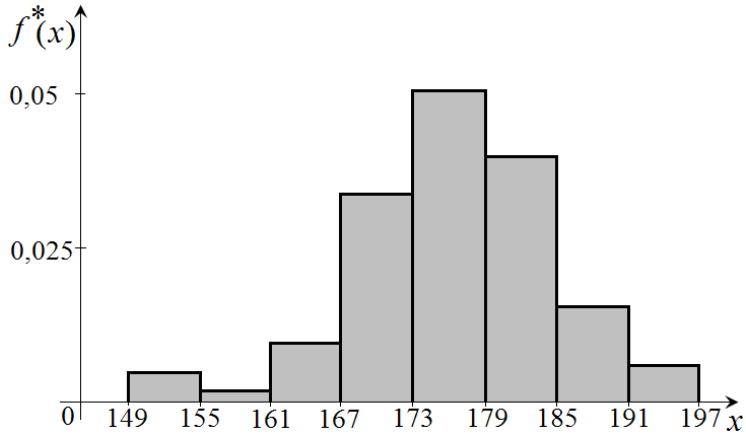


Рисунок 14

Накопленная частота интервала [173-179) равна сумме всех предшествующих частот, включая частоту данного интервала:

$$m^{\text{нак}}_{[173\text{-}179)} = 1 + 3 + 4 + 21 + 31 = 60.$$

Соответствующая накопленная относительная частота:

$$p^{\text{нак}}_{[173\text{-}179)} = 0{,}0095 + 0{,}0286 + 0{,}0381 + 0{,}2 + 0{,}2952 = 0{,}5714\,.$$

График накопленных относительных частот – кумулята.

Для интервального вариационного ряда эмпирическая функция распределения совпадает с кумулятой.

Найдем значения функции $F^*(x)$, равные значениям накопленных относительных частот, и запишем в таблице:

x	$x \le 149$	155	161	167	173	179	185	191	$x \ge 197$
$F^*(x)$	0	0,0095	0,0381	0,0762	0,2762	0,5714	0,8381	0,9619	1

Отметим на плоскости точки, соответствующие значениям $F^*(x)$, и соединим их отрезками прямых (рис. 15).

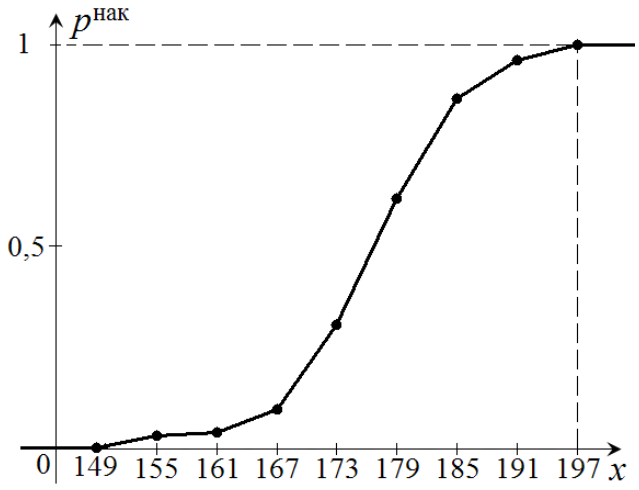


Рисунок 15

7.2. Числовые характеристики выборки

Пример 3. Данные о распределении 100 рабочих цеха по выработке в отчётном году (в процентах к предыдущему году) представлены в таблице. Найти среднюю выработку по цеху. Вычислить числовые характеристики.

Выработка в отчётном году в процентах к предыдущему, $[x_{i-1}; x_i)$	Середина интервала, $\tilde{x}_i$ (%)	Количество рабочих, m_i (чел.)
94 – 100	97	3
100 – 106	103	7
106 – 112	109	11
112 – 118	115	20
118 – 124	121	28
124 – 130	127	19
130 – 136	133	10
136 – 142	139	2
Σ	**–**	**100**

Решение. Вычислим среднее арифметическое всех значений выборочной совокупности – среднюю выборочную:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \frac{1}{100}(97\cdot 3 + 103\cdot 7 + \ldots + 139\cdot 2) = 119{,}2.$$

Вычислим оценку отклонений вариант от выборочной средней – выборочную дисперсию:

$$D_в = \overline{x^2} - (\overline{x}_в)^2 = \frac{1}{100}\left(97^2\cdot 3 + 103^2\cdot 7 + \ldots + 139^2\cdot 2\right) - 119{,}2^2 = 87{,}48.$$

Тогда: $\sigma_в = \sqrt{D_в} = \sqrt{87{,}48} \approx 9{,}35$.

Несмещенные оценки:

$$s^2 = \frac{n}{n-1}\cdot D_в = \frac{100}{99}\cdot 87{,}48 \approx 88{,}36, \quad s = \sqrt{88{,}36} \approx 9{,}4.$$

Мера колеблемости изучаемого признака относительно выборочной средней – коэффициент вариации:

$$V^* = \frac{\sigma_в}{\overline{x}_в}\cdot 100\,\% = \frac{9{,}35}{119{,}2}\cdot 100\,\% \approx 7{,}84\,\%.$$

Найдём моду, принадлежащую интервалу $[118;124)$, имеющему самую большую частоту $m = 28$ (частость $p_i^* = 0{,}28$):

$$M_O^* = 118 + 6 \cdot \frac{28 - 20}{(28 - 20) + (28 - 19)} = 118 + 6 \cdot \frac{8}{17} \approx 120{,}82.$$

Найдём медиану, принадлежащую интервалу $[118;124)$, т.к. для него накопленная частота $m^{нак}_{[118\text{-}124)} = \underbrace{3}_{1} + \underbrace{7}_{4} + \underbrace{11}_{2} + \underbrace{20}_{4} + \underbrace{28}_{43} = 69$ впервые превышает половину объёма выборки $\frac{n}{2} = \frac{100}{2} = 50$.

41

$$M_e^* = 118 + \frac{6}{28} \cdot (50 - (3 + 7 + 11 + 20)) \approx 119{,}93.$$

Пример 4. Методом моментов по выборке x_1, x_2, …, x_n найти точечную оценку неизвестного параметра λ показательного распределения с известной функцией плотности распределения $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$.

Решение. Формула $M[X] = \int_{-\infty}^{+\infty} x \cdot f(x, \lambda)\, dx$ при помощи метода интегрирования по частям даёт: $M[X] = \frac{1}{\lambda}$. Далее из равенства $M[X] = \overline{x}_в$ получаем, что $\lambda = \frac{1}{\overline{x}_в}$, т.е. искомая точечная оценка параметра λ показательного распределения равна обратной выборочной средней: $\lambda^* = \frac{1}{\overline{x}_в}$.

Пример 5. Найти методом моментов по выборке x_1, x_2, …, x_n точечные оценки неизвестных параметров a и σ нормального распределения с функцией плотности распределения вероятности

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

Решение. Из курса теории вероятностей известно, что для нормального распределения $M[X]=a$, $D[X]=\sigma^2$. Используя формулы: $M[X]=\bar{x}_в$, $D[X]=D_в$, получаем искомые точечные оценки параметров: $a^*=\bar{x}_в$, $\sigma^*=\sqrt{D_в}$.

Пример 6. Методом наибольшего правдоподобия найти оценку неизвестного параметра p биномиального распределения $P_n(m)=C_n^m p^m(1-p)^{n-m}$, если в n_1 испытаниях событие A наступило m_1 раз, а в n_2 испытаниях – m_2 раз.

Решение. Составим функцию правдоподобия, где $p=\theta$:

$$L=P_{n_1}(m_1)\cdot P_{n_2}(m_2)=C_{n_1}^{m_1}C_{n_2}^{m_2}p^{m_1+m_2}(1-p)^{n+m-m_1-m_2}.$$

Логарифмическая функция правдоподобия:

$$\ln L=\ln\left(C_{n_1}^{m_1}C_{n_2}^{m_2}\right)+(m_1+m_2)\ln p+(n_1-m_1+n_2-m_2)\ln(1-p).$$

Затем находим производную по p и, приравнивая её к нулю, получаем

$$\frac{m_1+m_2}{p}-\frac{n_1-m_1+n_2-m_2}{1-p}=0 \quad \Rightarrow \quad p=\frac{m_1+m_2}{n_1+n_2}.$$

Нетрудно убедиться, что вторая производная функции $\ln L<0$, т. е. полученное значение p является точкой максимума логарифмической функции правдоподобия, а, значит, эту величину нужно принять в качестве оценки наибольшего правдоподобия неизвестного параметра p^* биномиального распределения: $p^*=\frac{m_1+m_2}{n_1+n_2}$.

Пример 7. Методом наибольшего правдоподобия найти оценку неизвестного параметра λ показательного распределения $f(x) = \lambda e^{-\lambda x}$.

Решение. Составим функцию правдоподобия:

$$\theta = \lambda \quad \Rightarrow \quad L = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \ldots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda(x_1 + x_2 + \ldots + x_n)}.$$

Найдём логарифмическую функцию правдоподобия:

$$\ln L = n \ln \lambda - \lambda(x_1 + x_2 + \ldots + x_n).$$

Производную этой функции по λ приравняем к нулю. Получим:

$$\lambda = \frac{n}{x_1 + x_2 + \ldots + x_n} = \frac{1}{\overline{x}_в}.$$

Пример 8. Монету подбрасывают *n* раз. Монета выпала гербом *m* раз. Вероятность выпадения герба при каждом подбрасывании равна *p*. Показать несмещённость оценки $\theta^* = \frac{m}{n}$ вероятности $\theta = p$ выпадения герба в каждом опыте.

Решение. Так как имеем распределение Бернулли, то: $M[m] = np$.

Следовательно, $M[\theta^*] = M[m/n] = \frac{1}{n} M[m] = \frac{1}{n} \cdot n \cdot p = p = \theta$.

Таким образом, оценка $\theta^* = \frac{m}{n}$ – *несмещённая*.

7.3. Доверительные интервалы

Пример 9. Произведено пять независимых наблюдений над СВ $X \sim N(a, 20)$. Результаты наблюдений: $x_1 = -25$, $x_2 = 34$, $x_3 = -20$, $x_4 = 10$, $x_5 = 21$. Найти оценку для $a = M[X]$ при известном среднем квадратическом отклонении $\sigma_в = 20$ и построить для него 95%-й доверительный интервал.

Решение. Находим среднее выборочное

$$\bar{x}_в = \frac{1}{5}\cdot(-25+34-20+10+21) = 4.$$

Т. к. вероятность $\gamma = 0{,}95$, то $\Phi(t) = \frac{\gamma}{2} = 0{,}475$.

Тогда по таблице приложения «Функция Лапласа»: $t = 1{,}96$.

Значит:

$$\varepsilon = \frac{t\cdot\sigma_в}{\sqrt{n}} = \frac{1{,}96\cdot 20}{\sqrt{5}} \approx 17{,}5 \Rightarrow$$

$$(\alpha;\ \beta) = (4-17{,}5;\ 4+17{,}5) = (-13{,}5;\ 21{,}5).$$

Таким образом, оценкой $a = M[X]$ служит $\bar{x}_в = 4$, а доверительным интервалом для $a = M[X]$ является интервал: $(-13{,}5;\ 21{,}5)$.

Пример 10. По условию примера 9, считая, что СВ X *распределена нормально* $N(a,\sigma)$, построить для неизвестного $M[X] = a$ доверительный интервал при неизвестном среднем квадратическом отклонении для уровня надёжности $\gamma = 0{,}95$.

Решение. Ранее вычислено $\bar{x}_в = 4$. Находим s:

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 =$$

$$= \frac{1}{4}\cdot\left((-25-4)^2 + (34-4)^2 + (-20-4)^2 + (10-4)^2 + (21-4)^2\right) = 660{,}5$$

$$s = \sqrt{s^2} = \sqrt{660{,}5} \approx 25{,}7.$$

По таблице приложения ($\gamma = 0{,}95$, $n = 5$) находим $t_\gamma = 2{,}78$, и

$$\varepsilon = \frac{t_\gamma\cdot s}{\sqrt{n}} = \frac{2{,}78\cdot 25{,}7}{\sqrt{5}} \approx 31{,}9.$$

$$\Rightarrow (\alpha;\ \beta) = (4-31{,}9;\ \ 4+31{,}9) = (-27{,}9;\ \ 35{,}9) \ \Rightarrow\ a \in (-27{,}9;\ 35{,}9).$$

Пример 11. Количественный признак генеральной совокупности распределён нормально. По выборке объёма $n = 25$ найдено исправленное

среднее квадратическое отклонение $s = 0,8$. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение $\sigma_г$ с надёжностью $0,95$.

Решение. По таблице приложения ($\gamma = 0,95$, $n = 25$): $q = 0,32$.

Т. к. $q = 0,32 < 1$, то: $0,8 - 0,8 \cdot 0,32 < \sigma_г < 0,8 + 0,8 \cdot 0,32$

Таким образом, доверительный интервал для $\sigma_г$: $(0,544; 1,056)$.

7.4. Проверка гипотез о числовых характеристиках для одной выборки

Пример 12. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma_г = 5$ извлечена выборка объёма $n = 100$, и по ней найдено выборочное среднее $\bar{x}_в = 26,5$. Требуется на уровне значимости $0,05$ проверить гипотезу $H_0 : M_г = 25$ при альтернативной гипотезе $H_1 : M_г \neq 25$. Изменится ли результат, если изменить альтернативную гипотезу на $H_1 : M_г > 25$?

Решение. Найдём статистику критерия:

$$U_{набл} = \frac{\bar{x}_в - a_0}{\sigma_г}\sqrt{n} = \frac{26,5 - 25}{5}\sqrt{100} = 3.$$

При проверке гипотезы $H_1 : M_г \neq a_0$ по таблице приложения для двусторонней критической области из условия $\Phi(u_{кр}) = 0,5 - \frac{0,05}{2} = 0,475$ находим $u_{кр} = 1,96$.

Т.к. $|U_{набл}| > u_{кр}$, то основная гипотеза отвергается. Принимается гипотеза $H_1 : M_г \neq 25$.

При проверке гипотезы $H_1 : M_г > a_0$ для односторонней критической области из условия $\Phi(u_{кр}) = 0{,}5 - 0{,}05 = 0{,}45$ находим $u_{кр} = 1{,}65$.

Так как $U_{набл} > u_{кр}$, то основная гипотеза отвергается. Принимается гипотеза $H_1 : M_г > 25$.

В обоих случаях результат одинаков – основная гипотеза отвергается.

Пример 13. По выборке объёма $n = 16$, извлечённой из нормальной генеральной совокупности, найдены $\bar{x}_в = 12{,}4$ и $s = 1{,}2$. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0 : M_г = 11{,}8$ при конкурирующей гипотезе $H_1 : M_г \neq 11{,}8$.

Решение. Найдём статистику критерия:

$$T_{набл} = \frac{\bar{x}_в - a_0}{s}\sqrt{n-1} = \frac{12{,}4 - 11{,}8}{1{,}2}\sqrt{15} \approx 2.$$

Поскольку конкурирующая гипотеза имеет вид $H_1 : M_г \neq a_0$, то искомая критическая область двусторонняя. Из таблицы критических точек распределения Стьюдента по уровню значимости $\alpha = 0{,}05$ и числу степени свободы $k = n - 1 = 15$ найдём критическую точку $t_{кр} = t_{кр}(0{,}05;15) = 2{,}13$.

Т. к. $|T_{набл}| < t_{кр}$, то принимается нулевая гипотеза $H_0 : M_г = 11{,}8$.

Пример 14. Проверить нулевую гипотезу о том, что значение $a_0 = 40$ является математическим ожиданием нормально распределённой СВ при 5% уровне значимости для двусторонней и односторонней критической области, если в результате обработки выборки объёма $n = 10$ получено выборочное среднее $\bar{x}_в = 38$, а несмещённое среднее квадратичное отклонение равно $s = 3{,}6$.

Решение. Найдём статистику критерия:

$$T_{набл} = \frac{(\bar{x}_в - a_0)\cdot\sqrt{n-1}}{s} = \frac{(38-40)\cdot\sqrt{9}}{3{,}6} \approx -1{,}67.$$

1) Для двусторонней критической области:

основная гипотеза $H_0 : M_г = 40$; альтернативная гипотеза

$$H_1 : M_г \neq 40.$$

Критическая точка $t_{кр} = t_{\alpha, n-1} = t_{0{,}05;9} = 2{,}26$ определяется по таблице приложения (двусторонняя критическая область).

Т.к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу $H_0 : M_г = 40$.

2) Для левосторонней критической области: $H_0 : M_г = 40$; $H_1 : M_г < 40$

$t_{кр} = t_{\alpha, n-1} = t_{0{,}05;9} = 1{,}83$ (табл. для односторонней критической области).

Т. к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу $H_0 : M_г = 40$.

3) Для правосторонней критической области: $H_0 : M_г = 40; H_1 : M_г > 40$.

$t_{кр} = t_{\alpha, n-1} = t_{0{,}05;9} = 1{,}83$ (табл. для односторонней критической области).

Т. к. $|T_{набл}| < t_{кр}$, то принимаем основную гипотезу $H_0 : M_г = 40$.

Пример 15. Точность работы станка-автомата проверяется по дисперсии размеров изделий, которая не должна превышать 0,01 (мм^2). По выборке из 25 изделий получена исправленная выборочная дисперсия $s^2 = 0{,}02$ (мм^2). На уровне значимости $\alpha = 0{,}05$ проверить, обеспечивает ли станок необходимую точность?

Решение. $H_0 : D_г = 0{,}01$; $H_1 : D_г > 0{,}01$.

Статистика критерия:

$$\chi^2 = \frac{(n-1)\cdot s^2}{\sigma_0^2} = \frac{24\cdot 0{,}02}{0{,}01} = 48.$$

По таблице приложения распределения χ^2 – квадрат находим критическую точку: $\chi^2_{0,05;24} = 36{,}4$.

Т.к. $48 > 36{,}4$, то основная гипотеза отвергается. Следовательно, станок не обеспечивает необходимой точности.

7.5. Проверка гипотез о числовых характеристиках для двух независимых выборок

Пример 16. Для проверки эффективности новой технологии отобраны две группы рабочих: в первой группе численностью $n_1 = 50$ чел., где применялась новая технология, выборочная средняя выработка составила $\bar{x}_в = 85$ (изделий), во второй группе численностью $n_2 = 70$ чел. выборочная средняя – $\bar{y}_в = 78$ (изделий). Предварительно установлено, что дисперсии выработки в группах равны соответственно $\sigma_x^2 = 100$ и $\sigma_y^2 = 74$. На уровне значимости $\alpha = 0{,}05$ выяснить влияние новой технологии на среднюю производительность.

Решение. Проверяемая гипотеза H_0: $M_г[x] = M_г[y]$, т.е. средние выработки рабочих одинаковы по новой и старой технологиям. В качестве конкурирующей гипотезы можно взять H_1: $M_г[x] > M_г[y]$ или H_2: $M_г[x] \neq M_г[y]$. В данной задаче более естественна гипотеза H_1, т.к. её справедливость означает эффективность применения новой технологии.

Находим статистику критерия: $U = \dfrac{\bar{x}_в - \bar{y}_в}{\sqrt{\dfrac{D_г[x]}{n_x} + \dfrac{D_г[y]}{n_y}}} = \dfrac{85 - 78}{\sqrt{\dfrac{100}{50} + \dfrac{74}{70}}} = 4.$

При альтернативной гипотезе H_1 для односторонней критической области из условия $\Phi(u_{кр}) = 0{,}5 - 0{,}05 = 0{,}45$ найдём критическое значение $u_{кр} = 1{,}64$.

При конкурирующей гипотезе H_2 для двусторонней критической области из условия $\Phi(u_{кр}) = 0{,}5 - \frac{0{,}05}{2} = 0{,}475$ найдём $u_{кр} = 1{,}96$.

Т.к. $U > u_{кр}$ при любой из взятых конкурирующих гипотез, то гипотеза H_0 отвергается. Т.е. на 5%-ном уровне значимости можно сделать вывод, что новая технология позволяет повысить среднюю выработку рабочих.

Пример 17. Реклама утверждает, что из двух типов пластиковых карт «Русский экспресс» и «Супер-понт» богатые люди предпочитают первый. С целью проверки этого утверждения были обследованы среднемесячные платежи $n_1 = 16$ обладателей «Русского экспресса» и $n_2 = 11$ обладателей «Супер-понта». При этом выяснилось, что платежи по картам «Русский экспресс» составляют в среднем 563 долл. с исправленным средним квадратическим отклонением 178 долл., а по картам «Супер-понт» – в среднем 485 долл. с исправленным средним квадратическим отклонением 196 долл. Предварительный анализ законов распределения месячных расходов показал, что они достаточно хорошо описываются нормальным приближением. Проверить утверждение рекламы на уровне значимости 10 %.

Решение. В этом случае следует проверить гипотезу о средних при неизвестных дисперсиях (объёмы выборок малы). Поэтому, прежде всего, необходимо проверить гипотезу о равенстве дисперсий.

Имеем: $F = \frac{s^2_{\max}}{s^2_{\min}} = \frac{196^2}{178^2} = \frac{38416}{31684} = 1{,}21.$

Из таблицы критических значений Фишера-Снедекора по уровню значимости $\alpha/2 = 0{,}05$ и числам степеней свободы $k_1 = n_{\max} - 1 = 10$ и $k_2 = n_{\min} - 1 = 15$ ($n_{\max}$ и $n_{\min}$ соответствуют $s^2_{\max}$ и $s^2_{\min}$) находим критическую точку $F_{кр} = 2{,}55$. Поскольку $1{,}21 < 2{,}55$, принимаем гипотезу о равенстве дисперсий двух выборок.

Теперь можно воспользоваться критерием Стьюдента для проверки гипотезы о равенстве средних. Имеем:

$$s = \sqrt{\frac{(n_x - 1)\cdot s_x^2 + (n_y - 1)\cdot s_y^2}{n_x + n_y - 2}} = \sqrt{\frac{38416 \cdot 10 + 31684 \cdot 15}{11 + 16 - 2}} = 185{,}4.$$

Вычисляем статистику критерия:

$$T = \frac{\bar{x}_в - \bar{y}_в}{s \cdot \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{563 - 485}{185{,}4\sqrt{\frac{1}{11} + \frac{1}{16}}} = 1{,}07.$$

По таблице критических точек распределения Стьюдента для односторонней области по уровню значимости $\alpha = 0{,}1$ и числу степеней свободы k =25: $t_{кр} = 1{,}32$.

Поскольку $T < t_{кр}$, то принимается основная гипотеза (о равенстве средних). Т.о., утверждение рекламы не подтверждается имеющимися данными.

Пример 18. В партии из 500 деталей, изготовленных первым станком-автоматом, оказалось 60 нестандартных, из 600 деталей второго станка – 42 нестандартных. На уровне значимости $\alpha = 0{,}01$ проверить нулевую гипотезу $H_0 : p_x = p_y$ о равенстве вероятностей изготовления нестандартной детали обоими станками против конкурирующей гипотезы $H_1 : p_x \neq p_y$.

Решение. $p_1^* = \frac{60}{500} = 0{,}12$, $p_2^* = \frac{42}{600} = 0{,}07$, $p^* = \frac{60+42}{500+600} = 0{,}09$.

Находим значение статистики критерия:

$$U = \frac{p_1^* - p_2^*}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0{,}12-0{,}07}{\sqrt{0{,}09\cdot 0{,}91\cdot\left(\frac{1}{500}+\frac{1}{600}\right)}} = 2{,}85.$$

Критическую точку находим из соотношения $\Phi(u_{кр}) = 0{,}495$, откуда $u_{кр} = 2{,}57$. Так как $|U| > u_{кр}$, то гипотеза H_0 отвергается. Т.е. вероятности изготовления нестандартных деталей на двух станках различны.

Пример 19. При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве дисперсий двух нормально распределённых случайных величин *X* и *Y* на основе выборочных данных.

X		*Y*	
x_i	m_{x_i}	y_i	m_{y_i}
12,1	1	12,2	1
12,5	2	12,4	8
12,7	4	12,5	1
13,0	2	12,7	2
13,2	2	13,0	1

Решение. Проверим гипотезу $H_0 : D_г[x] = D_г[y]$ при альтернативной гипотезе $H_1 : D_г[x] \neq D_г[y]$.

Определим выборочные средние и выборочные дисперсии случайных величин *X* и *Y*, используя соответствующие ряды:

x_i	12,1	12,5	12,7	13,0	13,2
m_{x_i}	1	2	4	2	2

$$n_x = \sum_{i=1}^{5} m_{x_i} = 1+2+4+2+2 = 11.$$

$$\bar{x} = \frac{1}{n_x}\sum_{i=1}^{5} x_i m_{x_i} = \frac{1}{11}\left(12{,}1\cdot 1+12{,}5\cdot 2+12{,}7\cdot 4+13{,}0\cdot 2+13{,}2\cdot 2\right) = 12{,}75.$$

$$D_x = \frac{1}{n_x}\sum_{i=1}^{5} x_i^2 m_{x_i} - \bar{x}^2 =$$

$$= \frac{1}{11}\left(12{,}1^2\cdot 1+12{,}5^2\cdot 2+12{,}7^2\cdot 4+13{,}0^2\cdot 2+13{,}2^2\cdot 2\right) - 12{,}75^2 = 0{,}215.$$

$$s_x^2 = \frac{n_x}{n_x - 1} D_x = \frac{11}{10}\cdot 0{,}215 = 0{,}237.$$

y_i	12,2	12,4	12,5	12,7	13,0
m_{y_i}	1	8	1	2	1

$$n_y = \sum_{i=1}^{5} m_{y_i} = 1+8+1+2+1 = 13.$$

$$\bar{y} = \frac{1}{n_y}\sum_{i=1}^{5} y_i m_{y_i} = \frac{1}{13}\left(12{,}2\cdot 1+12{,}4\cdot 8+12{,}5\cdot 1+12{,}7\cdot 2+13{,}0\cdot 1\right) = 12{,}48.$$

$$D_y = \frac{1}{n_y}\sum_{i=1}^{5} y_i^2 m_{y_i} - \bar{y}^2 =$$

$$= \frac{1}{13}\left(12{,}2^2\cdot 1+12{,}4^2\cdot 8+12{,}5^2\cdot 1+12{,}7^2\cdot 2+13{,}0^2\cdot 1\right) - 12{,}48^2 = 0{,}153.$$

$$s_y^2 = \frac{n_y}{n_y - 1} D_y = \frac{13}{12}\cdot 0{,}153 = 0{,}166$$

Найдём значение критерия Фишера-Снедекора:

$$F_{набл} = \frac{s^2_{\max}}{s^2_{\min}} = \frac{s_x^2}{s_y^2} = \frac{0{,}237}{0{,}166} = 1{,}428,$$

здесь рассматривается отношение бóльшей дисперсии к меньшей.

По условию конкурирующая гипотеза имеет вид $D_г[x] \neq D_г[y]$, поэтому критическая область – двухсторонняя.

По таблице приложения, по уровню значимости, вдвое меньше заданного, т.е. при $\alpha/2 = 0{,}1/2 = 0{,}05$, и числам степеней свободы $k_1 = n_y - 1 = 12$ и $k_2 = n_x - 1 = 10$ находим критическое значение распределения Фишера-Снедекора: $F_{кр}(0{,}05;12;10) = 2{,}91$.

Т.к. $F_{набл} < F_{кр}$, то принимаем нулевую гипотезу о равенстве дисперсий.

7.6. Проверка гипотез о законе распределения

Пример 20. Для эмпирического распределения, заданного таблицей

Варианта, x_i	70	80	90	100	Σ
Частота, m_i	9	8	8	5	$n = 30$

при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности с помощью критерия χ^2 Пирсона.

Решение. Сформулируем гипотезы:

H_0: «СВ X распределена нормально».

H_1: «СВ X не распределена по нормальному закону».

Вычислим точечные оценки параметров a и σ:

$$a = \overline{x}_в = \frac{1}{30} \cdot (70 \cdot 9 + 80 \cdot 8 + 90 \cdot 8 + 100 \cdot 5) = 83,$$

$$\sigma_в^2 = \frac{1}{30} \cdot \left[(70-83)^2 \cdot 9 + (80-83)^2 \cdot 8 + (90-83)^2 \cdot 8 + (100-83)^2 \cdot 5\right] = 114{,}333.$$

Т. к. объём выборки невелик, то перейдём к исправленной дисперсии $s^2 = \frac{30}{29} \cdot \sigma_в^2$, тогда $\sigma = \sqrt{s^2} = 10{,}875$.

Таким образом, предполагаем нормальный закон распределения: $N(83;10{,}875)$.

Выравнивающие частоты m_i' найдём по формуле:

$$m_i' = n \cdot p_i = \frac{n h_i}{\sigma} \cdot \varphi(t_i),$$

где $n = 30$, $h_i = \Delta x_i = 10$, $t_i = \dfrac{x_i - a}{\sigma}$, $\varphi(t)$ – функция Гаусса.

Получаем $m_i' = \dfrac{300}{10{,}875} \cdot \varphi\left(\dfrac{x_i - 83}{10{,}875}\right)$.

i	x_i	m_i	$t_i = \dfrac{x_i - 83}{10{,}875}$	$\varphi(t_i)$	$m_i' = \dfrac{300}{10{,}875} \cdot \varphi(t_i)$	$(m_i - m_i')^2$	$\dfrac{(m_i - m_i')^2}{m_i'}$
1	70	9	– 1,20	0,1942	5,35	13,32	2,45
2	80	8	– 0,28	0,3836	10,58	6,66	0,63
3	90	8	0,64	0,3251	8,97	0,94	0,10
4	100	5	1,56	0,1182	3,26	3,03	0,93
Σ		**30**	**–**	**–**	**28,47**	**–**	**4,11**

Итак, фактически наблюдаемое значение статистики $\chi^2_{набл} = 4{,}11$.

Число наблюдений равно 4, то для нормального закона распределения число степеней свободы $k = 4 - 3 = 1$. По таблице приложения: $\chi^2_{кр} = \chi^2_{0{,}05;1} = 3{,}8$.

Так как $\chi^2_{набл} > \chi^2_{кр}$, то гипотеза о выбранном теоретическом нормальном законе $N(83;10{,}875)$ противоречит опытным данным, значит, принимаем альтернативную гипотезу: «СВ X не распределена по нормальному закону».

Пример 21. Для эмпирического распределения рабочих цеха по выработке в примере 3 на уровне значимости $\alpha = 0{,}05$ выдвинуть гипотезу о распределении генеральной совокупности и проверить её с помощью критерия χ^2 Пирсона.

Решение.

Выработка в отчётном году в процентах к предыдущему, $[x_{i-1}; x_i)$	Середина интервала, $\tilde{x}_i$ (%)	Количество рабочих, m_i (чел.)
94 – 100	97	3
100 – 106	103	7
106 – 112	109	11
112 – 118	115	20
118 – 124	121	28
124 – 130	127	19
130 – 136	133	10
136 – 142	139	2
Σ	**–**	**100**

По виду гистограммы распределения рабочих по выработке (рис. 16) можно предположить нормальный закон распределения признака.

Параметры a и σ^2 нормального закона распределения, являющиеся соответственно математическим ожиданием и дисперсией СВ X, неизвестны. Поэтому заменяем их «наилучшими» оценками по выборке – несмещёнными и состоятельными оценками соответственно выборочной средней $\overline{x}_в$ и исправленной выборочной дисперсией s^2.

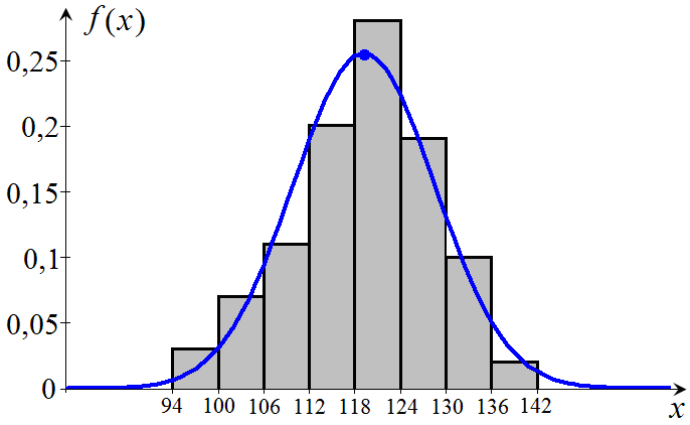


Рисунок 16

Т.к. число наблюдений $n=100$ достаточно велико, то вместо s^2 можно взять σ^2. В примере 3 были вычислены $\bar{x}=119{,}2\,(\%)$, $\sigma=9{,}35\,(\%)$.

Сформулируем основную гипотезу H_0: «СВ X – выработка рабочих цеха – распределена нормально с параметрами $a=119{,}2$, $\sigma=9{,}35$, т. е. $X \sim N(119{,}2;9{,}35)$».

Альтернативная гипотеза H_1: «СВ X не распределена по нормальному закону».

Число наблюдений в крайних интервалах меньше 5, поэтому объединим их с соседними.

$[x_{i-1};x_i)$	94-106	106-112	112-118	118-124	124-130	130-142	**Σ**
m_i	10	11	20	28	19	12	**100**

Для расчёта вероятностей $p_i = P\{x_{i-1} \le X \le x_i\}$ попадания СВ X в интервал $[x_{i-1};x_i)$ в соответствии со свойствами нормального распределения **можно использовать функцию Гаусса**

$$p_i = \frac{h_i}{\sigma}\cdot\varphi\left(\frac{\tilde{x}_i - a}{\sigma}\right),$$

где $\tilde{x}_i$ – середина интервала (пример 20), **или функцию Лапласа**

$$p_i = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

Найдём значения p_i, используя функцию Лапласа. Т. к. СВ $X \sim N(a;\sigma)$ определена на интервале $(-\infty,+\infty)$, то крайние промежутки в ряде распределения заменяем соответственно на $(-\infty,106)$ и $[130,+\infty)$.

$$p_1 = P\{-\infty \le X \le 106\} = \Phi\left(\frac{106-119,2}{9,35}\right) - \Phi(-\infty) =$$

$$= \Phi(-1,41) - \Phi(-\infty) = -0,4207 + 0,5 = 0,0793.$$

$$p_2 = P\{106 \le X \le 112\} = \Phi\left(\frac{112-119,2}{9,35}\right) - \Phi\left(\frac{106-119,2}{9,35}\right) =$$

$$= \Phi(-0,77) - \Phi(-1,41) = -0,2794 + 0,4207 = 0,1413.$$

$$p_3 = P\{112 \le X \le 118\} = \Phi\left(\frac{118-119,2}{9,35}\right) - \Phi\left(\frac{112-119,2}{9,35}\right) =$$

$$= \Phi(-0,13) - \Phi(-0,77) = -0,0517 + 0,2794 = 0,2277.$$

$$p_4 = P\{118 \le X \le 124\} = \Phi\left(\frac{124-119,2}{9,35}\right) - \Phi\left(\frac{118-119,2}{9,35}\right) =$$

$$= \Phi(0,51) - \Phi(-0,13) = 0,1950 + 0,0517 = 0,2467.$$

$$p_5 = P\{124 \le X \le 130\} = \Phi\left(\frac{130-119,2}{9,35}\right) - \Phi\left(\frac{124-119,2}{9,35}\right) =$$

$$= \Phi(1,16) - \Phi(0,51) = 0,3770 - 0,1950 = 0,1820.$$

$$p_6 = P\{130 \le X \le +\infty\} = \Phi(+\infty) - \Phi\left(\frac{130-119,2}{9,35}\right) =$$

$$= \Phi(+\infty) - \Phi(1,16) = 0,5 - 0,3770 = 0,1230.$$

Для определения статистики χ^2 удобно составить таблицу:

i	$[x_{i-1}; x_i)$	m_i	p_i	$m_i' = n \cdot p_i$	$(m_i - m_i')^2$	$\frac{(m_i - m_i')^2}{m_i'}$
1	$(-\infty,106)$	10	0,079	7,9	4,41	0,558
2	$[106,112)$	11	0,141	14,1	9,61	0,682
3	$[112,118)$	20	0,228	22,8	7,84	0,344

4	$[118,124)$	28	0,247	24,7	10,89	0,441
5	$[124,130)$	19	0,182	18,2	0,64	0,035
6	$[130,+\infty)$	12	0,123	12,3	0,09	0,007
Σ		**100**	**1**	**100**	**–**	**2,067**

Итак, фактически наблюдаемое значение статистики $\chi^2_{набл} = 2{,}067$.

Число интервалов $l = 6$, тогда для нормального закона распределения число степеней свободы $k = l - 3 = 3$. Критическое значение χ^2 находим по таблице приложения: $\chi^2_{кр} = \chi^2_{0{,}05;3} = 7{,}82$.

Т.к. $\chi^2_{набл} < \chi^2_{кр}$, то гипотеза о выбранном теоретическом нормальном законе $N(119{,}2;9{,}35)$ не противоречит опытным данным.

Пример 22. В результате опыта получена выборочная совокупность. По данной таблице составить вариационный ряд, разбив всю вариацию на интервалы (рекомендуется 8-10 интервалов).

88	104	91	97	77	103	86	79	86	100
82	68	71	87	89	89	81	81	70	79
84	91	87	83	90	69	83	96	79	94
93	86	81	83	84	92	93	85	84	88
77	85	93	85	87	100	76	79	90	91
84	74	76	75	93	103	80	96	72	95
81	102	75	80	90	85	82	77	94	102
87	95	99	83	80	93	90	79	93	105
95	85	84	90	93	95	98	88	79	91
86	88	93	80	88	88	90	68	89	90

1. По сгруппированным данным построить:

а) полигон частот; *б*) гистограмму частот; *в*) эмпирическую функцию $F^*(x)$. Выдвинуть гипотезу о распределении генеральной совокупности.

2. Найти числовые характеристики выборки: M_O^*, M_e^*, $\bar{x}_в$, $\sigma_в$, s, V^*, a_s^*, ε_k^*.

3. С учетом числовых характеристик определить формулы дифференциальной и интегральной функций предполагаемого теоретического распределения.

4. Найти доверительные интервалы для генеральной средней $x_г$ и генерального среднего квадратического отклонения $\sigma_г$ при уровне надёжности $\gamma = 0{,}99$.

5. Найти интервал, в котором практически окажутся все значения случайной величины X.

6. Применив критерий согласия Пирсона χ^2 при уровне значимости $\alpha = 0{,}01$, принять или отвергнуть выдвинутую гипотезу о распределении генеральной совокупности.

7. По сгруппированным данным на одном чертеже построить:

а) полигон относительных частот ω_i и кривую распределения p_i.

б) гистограмму относительных частот ω_i и теоретический аналог $f(x)$.

Сравнить гистограмму относительных частот ω_i с графиком идеально нормального распределения, используя значения a_s^*, ε_k^*.

в) эмпирическую функцию $F^*(x)$ и её теоретический аналог $F(x)$.

Решение.

Разобьем всю вариацию объёмом $n = 100$ на $k = 10$ частичных интервалов равной длины и вычислим частоты попадания в них наблюдаемых значений.

Длину интервала находим по формуле $h = \dfrac{x_{\max} - x_{\min}}{k} = \dfrac{105 - 68}{10} \approx 4$.

За начало первого интервала примем $x_0 = x_{\min} - \dfrac{h}{2} = 68 - \dfrac{4}{2} = 66$.

Составим вариационный ряд частот и относительных частот.

i	$[x_{i-1}; x_i)$	Середина интервала, $\tilde{x}_i$	m_i	$\omega_i = p_i^* = \frac{m_i}{n}$	$\frac{p_i^*}{h} = \frac{p_i^*}{4}$
1	[66; 70)	68	3	0,03	0,0075
2	[70; 74)	72	3	0,03	0,0075
3	[74; 78)	76	8	0,08	0,02
4	[78; 82)	80	14	0,14	0,035
5	[82; 86)	84	16	0,16	0,04
6	[86; 90)	88	17	0,17	0,0425
7	[90; 94)	92	18	0,18	0,045
8	[94; 98)	96	10	0,10	0,025
9	[98; 102)	100	4	0,04	0,01
10	[102; 106]	104	7	0,07	0,0175
Σ	–	–	**100**	**1**	

Отметим, что $\sum_{i=1}^{10} m_i = n = 100$ – объём выборки; $\sum_{i=1}^{10} p_i^* = \sum_{i=1}^{10} \frac{m_i}{n} = \frac{n}{n} = 1$.

Статистическое распределение выборки является оценкой неизвестного распределения. В частности, *относительные частоты ω_i являются статистическими аналогами вероятностей полной группы несовместных событий.*

1. ***а)*** Полигон относительных частот вариационного ряда – ломаная линия, соединяющая точки $(\tilde{x}_i;\ p_i^*)$ (рис. 17).

Полигон относительных частот является статистическим аналогом многоугольника распределения дискретной случайной величины X.

Вид полигона относительных частот напоминает многоугольник нормального распределения.

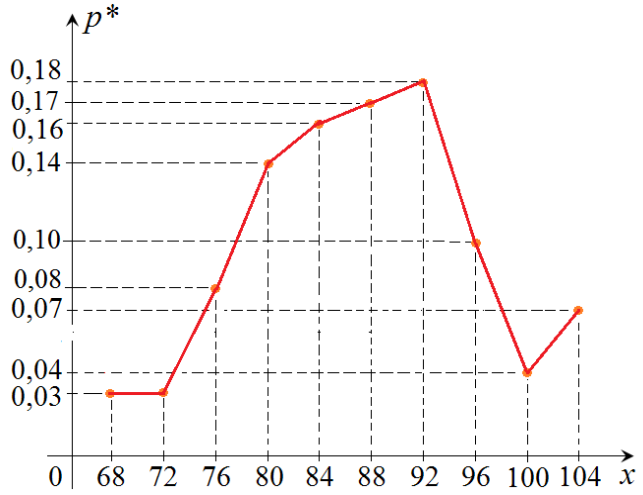


Рисунок 17

б) Гистограмма относительных частот имеет вид ступенчатой фигуры. На каждом частичном интервале строим прямоугольник высотой $\frac{p_i^*}{h}$ (рис. 18).

Гистограмма относительных частот является статистическим аналогом дифференциальной функции распределения (плотности) $f(x)$ непрерывной СВ X.

Вид гистограммы относительных частот напоминает дифференциальную функцию нормального распределения.

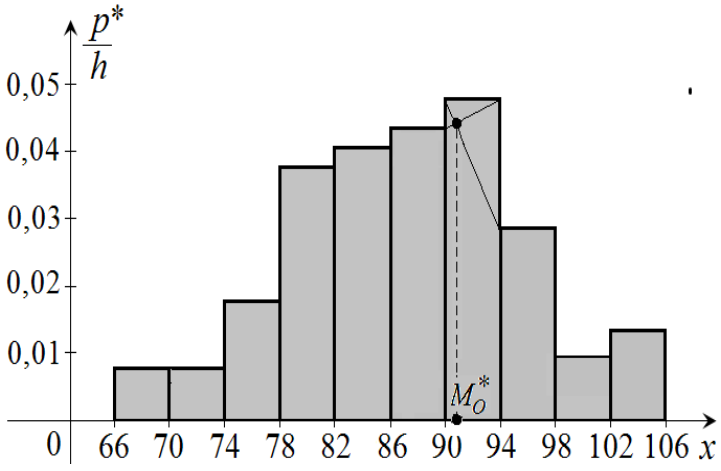


Рисунок 18

в) График эмпирической функции распределения $F^*(x)=\sum_{x_i<x}\frac{m_i}{n}$ непрерывной случайной величины X совпадает с кумулятой (графиком накопленных относительных частот).

Отметим на плоскости точки, соответствующие значениям функции $F^*(x)$ на концах интервалов, и соединим их отрезками прямых (рис. 19).

x	$x\le 66$	70	74	78	82	86	90	94	98	102	$x\ge 106$
$F^*(x)$	0	0,03	0,06	0,14	0,28	0,44	0,61	0,79	0,89	0,93	1

Эмпирическая функция распределения $F^*(x)$ является статистическим аналогом интегральной функции распределения $F(x)$ случайной величины X.

Вид эмпирической функции распределения $F^*(x)$ напоминает интегральную функцию нормального распределения.

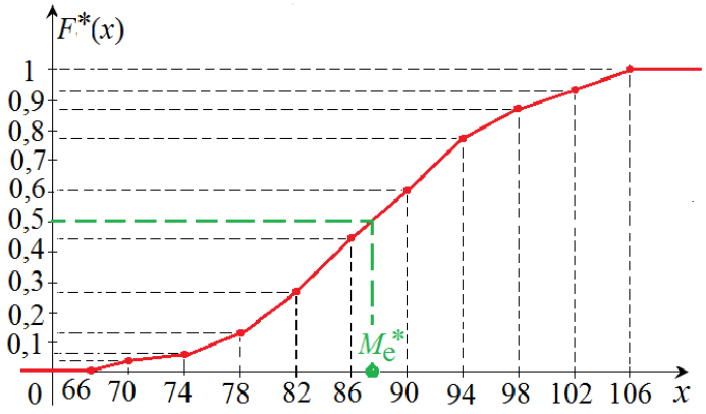


Рисунок 19

На основе графического анализа пункта 1 **выдвигаем основную (нулевую) гипотезу** H_0: «Генеральная совокупность распределена по нормальному закону» и **альтернативную гипотезу** H_1: «Генеральная совокупность не распределена по нормальному закону».

2. Найдем числовые характеристики выборки.

Выборочные характеристики – это функции наблюдений, приближённо оценивающие соответствующие генеральные числовые характеристики случайной величины.

1) Мода M_o^* находится внутри интервала, для которого соответствующая частота максимальна. В нашем случае $M_0^* \in [90;\ 94)$, при этом $p_7^* = 0{,}18$. Моду приближённо можно определить на чертеже гистограммы (рис. 7) или вычислить по формуле:

$$M_0^* = 90 + 4 \cdot \frac{0{,}18 - 0{,}17}{(0{,}18 - 0{,}17) + (0{,}18 - 0{,}10)} = 91{,}09.$$

2) Медиана M_e^* интервального вариационного ряда принадлежит тому частотному интервалу, для которого накопленная частота составляет половину или больше половины всей суммы частот, а предыдущая накопленная частота меньше половины всей суммы частот. Геометрически прямая $x = M_e^*$ делит площадь гистограммы пополам.

Медиана может быть приближённо найдена на чертеже графика $F^*(x)$ (рис. 8), как значение признака, для которого $F^*(M_e^*) = 0{,}5$. Для данного вариационного ряда $M_e^* \in (86;\ 90]$.

Вычисляем значение:

$$M_e^* = 86 + 4 \cdot \frac{0{,}5 - (0{,}03 + 0{,}03 + 0{,}08 + 0{,}14 + 0{,}16)}{0{,}17} = 86 + 4 \cdot \frac{0{,}5 - 0{,}44}{0{,}17} = 87{,}41.$$

3) Для нахождения выборочной средней $\overline{x}_в$, выборочной дисперсии $D_в$, выборочного среднего квадратического отклонения $\sigma_в$, коэффициента вариации V^*, асимметрии a_s^* и эксцесса ε_k^* вариационного ряда (*статистические аналоги соответствующих числовых характеристик случайной величины*) заполним вспомогательную таблицу.

i	$\widetilde{x}_i$	p_i^*	$\widetilde{x}_i p_i^*$	$\widetilde{x}_i^2 p_i^*$	$\widetilde{x}_i^3 p_i^*$	$\widetilde{x}_i^4 p_i^*$
1	68	0,03	2,04	138,72	9432,96	641441,28
2	72	0,03	2,16	155,52	11197,44	806215,68
3	76	0,08	6,08	462,08	35118,08	2668974,08
4	80	0,14	11,2	896	71680	5734400
5	84	0,16	13,44	1128,96	94832,64	7965941,76
6	88	0,17	14,96	1316,48	115850,24	10194821,12
7	92	0,18	16,56	1523,52	140163,84	12895073,28
8	96	0,10	9,6	921,6	88473,6	8493465,6
9	100	0,04	4	400	40000	4000000
10	104	0,07	7,8	757,12	78740,48	8189009,92
Σ	**–**	**1**	87,32	7700	685489,28	61589342,72

Найдем начальные моменты, используя результаты вспомогательной таблицы

$$\nu_1^*[x] = \sum \widetilde{x}_i p_i^* = 87{,}32; \qquad \nu_2^*[x] = \sum \widetilde{x}_i^2 p_i^* = 7700$$

$$\nu_3^*[x] = \sum \widetilde{x}_i^3 p_i^* = 685489{,}28; \qquad \nu_4^*[x] = \sum \widetilde{x}_i^4 p_i^* = 61589342{,}72$$

Находим выборочное среднее:

$$\overline{x}_{\text{в}} = \nu_1^*[x] = 87{,}32.$$

4) Находим выборочную дисперсию:

$$D_{\text{в}}[x] = \nu_2^*[x] - \left(\nu_1^*[x]\right)^2 = 7700 - (87{,}32)^2 = 75{,}216;$$

выборочное среднее квадратическое отклонение:

$$\sigma_{в}[x] = \sqrt{D_{в}} = \sqrt{75{,}216} = 8{,}673;$$

исправленную выборочную дисперсию:

$$s^2 = \frac{n}{n-1} \cdot D_в = \frac{100}{99} \cdot 75{,}2176 = 75{,}9774;$$

исправленное выборочное среднее квадратическое отклонение:

$$s = \sqrt{s^2} = 8{,}7165.$$

Т.к. число наблюдений $n = 100$ достаточно велико, то вместо s^2 можно использовать неисправленную выборочную дисперсию $\sigma_в^2$.

5) Для характеристики меры колеблемости изучаемого признака относительно выборочной средней вычислим коэффициент вариации:

$$V = \frac{\sigma_в}{m_в} \cdot 100\% = \frac{8{,}673}{87{,}32} \cdot 100\% = 9{,}93\%.$$

Он достаточно невелик, что говорит об однородности значений признака.

6) Выборочный коэффициент асимметрии:

$$a_s^*[x] = \frac{v_3^* - 3v_1^* \cdot v_2^* + 2\left(v_1^*\right)^3}{\sigma^3} =$$

$$= \frac{685489{,}28 - 3 \cdot 87{,}32 \cdot 7700 + 2\left(87{,}32\right)^3}{8{,}673^3} = -0{,}01644;$$

выборочный коэффициент эксцесса:

$$\varepsilon_k^*[x] = \frac{v_4^* - 4v_1^* \cdot v_3^* + 6v_2^* \cdot \left(v_1^*\right)^2 - 3 \cdot \left(v_1^*\right)^4}{\sigma^4} - 3 =$$

$$= \frac{61589342{,}2 - 4 \cdot 87{,}32 \cdot 685489{,}28 + 6 \cdot 7700 \cdot \left(87{,}32\right)^2 - 3 \cdot \left(87{,}32\right)^4}{8673^4} - 3 = -0{,}4064.$$

3. С учетом найденных числовых характеристик определим формулы дифференциальной

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 8{,}673} \cdot e^{-\frac{(x-87{,}32)^2}{2 \cdot 8{,}673^2}}, \quad x \in R$$

и интегральной функций предполагаемого теоретического (нормального) распределения: $F(x)=0{,}5+\Phi\left(\dfrac{x-87{,}32}{8{,}673}\right)$,

где точечной оценкой математического ожидания $\boldsymbol{a}$ *является средняя выборочная* $\overline{x}_в$, $(a=\overline{x}_в=87{,}32)$; *точечной оценкой генерального среднего квадратического отклонения* σ *является неисправленное выборочное среднее квадратическое отклонение* $(\sigma=\sigma_в=8{,}673)$.

4. Т.к. $\sigma_г$ неизвестно, то доверительный интервал для генеральной средней $x_г$ имеет вид:

$$\overline{x}_в-\varepsilon<\overline{x}_г<\overline{x}_в+\varepsilon\text{, где }\varepsilon=\frac{t_\gamma\cdot s}{\sqrt{n}}.$$

Значение $t_\gamma=t(\gamma,n)$ находим в таблице приложения по заданному уровню надёжности $\gamma=0{,}99$ и $n=100$: $t(0{,}99;100)=2{,}627$.

Тогда $\varepsilon=\dfrac{2{,}627}{\sqrt{100}}\cdot 8{,}673=2{,}28$.

Таким образом, получаем доверительный интервал для $\overline{x}_г$:

$$87{,}32-2{,}28\le\overline{x}_г\le 87{,}32+2{,}28\quad\Rightarrow\quad 85{,}04\le\overline{x}_г\le 89{,}6.$$

Это означает, что в 99 % случаев истинное значение генеральной средней $\overline{x}_г$ находится в промежутке $[85{,}04;\ 89{,}6]$.

Доверительный интервал для генерального среднего квадратического отклонения: $\sigma_в\cdot(1-q)<\sigma_г<\sigma_в\cdot(1+q)$.

По таблице приложения величина $q(\gamma,n)=q(0{,}99;100)=0{,}198<1$.

Отсюда:

$$8{,}673\cdot(1-0{,}198)\le\sigma_г\le 8{,}673\cdot(1+0{,}198)$$

$$\Rightarrow\quad 8{,}96\le\sigma_г\le 10{,}39.$$

Это означает, что в 99 % истинное значение генерального среднего квадратического отклонения $\sigma_г$ находится промежутке $[8{,}96;\ 10{,}39]$.

5. Найдём интервал, в котором практически окажутся все значения величины *X*. Для этого воспользуемся правилом «трёх сигм»: $P\{|X-a|\le 3\sigma\}=0{,}9973$, которое требует, чтобы в 99,73 % значения случайной величины, распределенной по нормальному закону, попадали на отрезок $[a-3\sigma;a+3\sigma]$.

В нашем случае: $\bar{x}_в-3\sigma_в=61{,}301$, $\bar{x}_в+3\sigma_в=113{,}339$.

Отсюда получаем, что интервал опытных данных $[66;\ 106]\in[61{,}301;\ 113{,}339]$.

Таким образом, найденный промежуток полностью накрыл наши статистические значения.

6. Проверим соответствие выдвинутой гипотезы H_0 опытным данным. Для этого вычислим теоретические вероятности p_i и выравнивающие частоты $m_i'=np_i$.

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений (т. е. $m_i\ge 5$). Т.к. число наблюдений в крайних интервалах меньше 5, то объединим их с соседними.

Получим следующий ряд распределения:

$[x_{i-1};x_i)$	[66; 74)	[74; 78)	[78; 8)]	[82; 86)	[86; 90)	[90; 94)	[94; 98)	[98; 106]
m_i	6	8	14	16	17	18	10	11

Используя таблицу приложения, найдём интервальные вероятности:

$$p_i=P\{x_{i-1}\le X\le x_i\}=F(x_i)-F(x_{i-1})=\Phi\left(\frac{x_i-a}{\sigma}\right)-\Phi\left(\frac{x_{i-1}-a}{\sigma}\right).$$

Т.к. случайная величина определена на интервале $(-\infty,+\infty)$, то крайние промежутки в ряде распределения заменяем, соответственно на $(-\infty;74]$ и $(98;+\infty)$.

$$p_1 = P\{-\infty < X \le 74\} = F(74) - F(-\infty) = \Phi\left(\frac{74-87{,}32}{8{,}673}\right) - \Phi(-\infty) =$$

$$= -\Phi(1{,}54) + \Phi(+\infty) = -0{,}4382 + 0{,}5 = 0{,}0618;$$

$$p_2 = P\{74 \le X \le 78\} = F(78) - F(74) = \Phi\left(\frac{78-87{,}32}{8{,}673}\right) - \Phi\left(\frac{74-87{,}32}{8{,}673}\right) =$$

$$= -\Phi(1{,}07) + \Phi(1{,}54) = -0{,}3537 + 0{,}4382 = 0{,}0845;$$

$$p_3 = P\{78 \le X \le 82\} = F(82) - F(78) = \Phi\left(\frac{82-87{,}32}{8{,}673}\right) - \Phi\left(\frac{78-87{,}32}{8{,}673}\right) =$$

$$= -\Phi(0{,}61) + \Phi(1{,}07) = -0{,}2291 + 0{,}3537 = 0{,}1246;$$

$$p_4 = P\{82 \le X \le 86\} = F(86) - F(82) = \Phi\left(\frac{86-87{,}32}{8{,}673}\right) - \Phi\left(\frac{82-87{,}32}{8{,}673}\right) =$$

$$= -\Phi(0{,}15) + \Phi(0{,}61) = -0{,}0596 + 0{,}2291 = 0{,}1695;$$

$$p_5 = P\{86 \le X \le 90\} = F(90) - F(86) = \Phi\left(\frac{90-87{,}32}{8{,}673}\right) - \Phi\left(\frac{86-87{,}32}{8{,}673}\right) =$$

$$= \Phi(0{,}31) + \Phi(0{,}15) = 0{,}1217 + 0{,}0596 = 0{,}1813;$$

$$p_6 = P\{90 \le X \le 94\} = F(94) - F(90) = \Phi\left(\frac{94-87{,}32}{8{,}673}\right) - \Phi\left(\frac{90-87{,}32}{8{,}673}\right) =$$

$$= \Phi(0{,}77) - \Phi(0{,}31) = 0{,}2794 - 0{,}1217 = 0{,}1577;$$

$$p_7 = P\{94 \le X \le 98\} = F(98) - F(94) = \Phi\left(\frac{98-87{,}32}{8{,}673}\right) - \Phi\left(\frac{94-87{,}32}{8{,}673}\right) =$$

$$= \Phi(1{,}23) - \Phi(0{,}77) = 0{,}3907 - 0{,}2794 = 0{,}1131;$$

$$p_8 = P\{98 \le X < +\infty\} = F(+\infty) - F(98) = \Phi(+\infty) - \Phi\left(\frac{98-87{,}32}{8{,}673}\right) =$$

$$= \Phi(+\infty) - \Phi(1{,}23) = 0{,}5 - 0{,}3907 = 0{,}1093.$$

Для дальнейших расчётов заполним вспомогательную таблицу:

i	$[x_{i-1}; x_i)$	m_i	p_i	$m_i' = np_i$	$\frac{(m_i - m_i')^2}{m_i'}$
1	$(-\infty; 74)$	6	0,0618	6,18	0,0052
2	[74; 78)	8	0,0845	8,45	0,0240
3	[78; 82)	14	0,1256	12,46	0,1903
4	[82; 86)	16	0,1695	16,95	0,0532
5	[86; 90)	17	0,1813	18,13	0,0704
6	[90; 94)	18	0,1577	15,77	0,3153
7	[94; 98)	10	0,1131	11,13	0,1147
8	$[98; +\infty)$	11	0,1093	10,93	0,0004
Σ	–	**100**	**1**	**100**	**0,7737**

Наблюдаемое значение критерия согласия Пирсона (итоговая строка таблицы): $\chi^2_{набл} = 0{,}7737$.

По таблице приложения для заданного уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $k = 8 - 3 = 5$ найдём: $\chi^2_{кр}(\alpha; k) = \chi^2_{кр}(0{,}05; 5) = 11{,}1$.

Т.к. $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть проверяемую нулевую гипотезу. Т.е. принимаем предположение, что **статистические данные распределены по нормальному закону с параметрами** $a = 87{,}32$ и $\sigma = 8{,}673$.

7. *а)* Построим на одном чертеже полигон эмпирических относительных частот ω_i и кривую распределения теоретических вероятностей p_i (рис. 20).

$[x_{i-1}; x_i)$	$\tilde{x}_i$	m_i	$\omega_i = p_i^* = \frac{m_i}{n}$	$p_i = P\{x_{i-1} \le X \le x_i\}$
[66; 70)	68	3	0,03	0,0618
[70; 74)	72	3	0,03	
[74; 78)	76	8	0,08	0,0845
[78; 82)	80	14	0,14	0,1246
[82; 86)	84	16	0,16	0,1695
[86; 90)	88	17	0,17	0,1813
[90; 94)	92	18	0,18	0,1577
[94; 98)	96	10	0,10	0,1131
[98; 102)	100	4	0,04	0,1093
[102; 106]	104	7	0,07	

Обозначения:

– кривая p_i^*.

– кривая p_i.

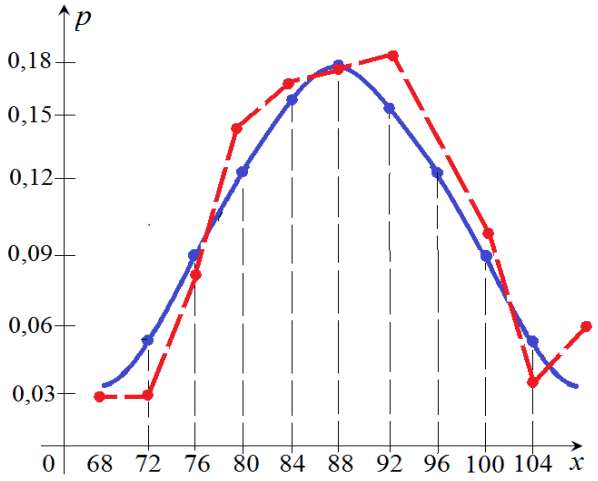


Рисунок 20

Кривая p_i выравнивает эмпирические данные p_i^*, тем самым приближая распределение генеральной совокупности к нормальному.

б) Построим на одном чертеже гистограмму относительных частот p_i^* и теоретический аналог (рис. 21):

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 8{,}673} \cdot e^{-\frac{(x-87{,}32)^2}{2 \cdot 8{,}673^2}} = \frac{1}{8{,}673} \cdot \varphi\left(\frac{x - 87{,}32}{8{,}673}\right).$$

$[x_{i-1}; x_i)$	$\tilde{x}_i$	$f_i^* = \frac{\omega_i}{h} = \frac{\omega_i}{4}$	$t_i = \frac{\tilde{x}_i - 87{,}32}{8{,}673}$	$\varphi(t_i)$	$f(x_i) = \frac{1}{8{,}673} \cdot \varphi(t_i)$
[66; 70)	68	0,0075	−2,23	0,0332	0,0038
[70; 74)	72	0,0075	−1,77	0,0833	0,0096
[74; 78)	76	0,02	−1,31	0,1691	0,0195
[78; 82)	80	0,035	−0,84	0,2803	0,0323
[82; 86)	84	0,04	−0,38	0,3712	0,0428
[86; 90)	88	0,0425	0,08	0,3977	0,0459
[90; 94)	92	0,045	0,54	0,3448	0,0398
[94; 98)	96	0,025	1,00	0,2420	0,0279
[98; 102)	100	0,01	1,46	0,1374	0,0158
[102; 106]	104	0,0175	1,92	0,0632	0,0073

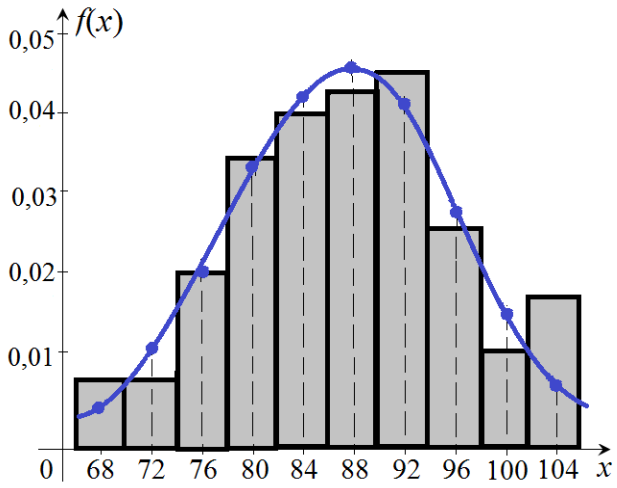


Рисунок 21

Т.к. $a_s^* < 0$, то «длинная» часть кривой теоретического распределения расположена несколько левее от $M_0^* = 91{,}09$. При этом кривая является бо-

лее пологой ($\varepsilon_k^* < 0$), т.е. менее островершинной, чем идеально нормальная кривая.

в) Построим на одном чертеже эмпирическую функцию $F^*(x)$ и её теоретический аналог (рис. 22):

$$F(x) = 0{,}5 + \Phi\left(\frac{x - 87{,}32}{8{,}673}\right).$$

$[x_{i-1}; x_i)$	F_i^*	$t_i = \frac{x_i - 87{,}32}{8{,}673}$	$\Phi(t_i)$	$F(x_i) = 0{,}5 + \Phi(t_i)$
$x \le 66$	0	−2,46	−0,4931	0,0069
70	0,03	−2,00	−0,4772	0,0228
74	0,06	−1,54	−0,4382	0,0618
78	0,14	−1,07	−0,3577	0,1423
82	0,28	−0,61	−0,2291	0,2709
86	0,44	−0,15	−0,0596	0,4404
90	0,61	0,31	0,1217	0,6217
94	0,79	0,77	0,2794	0,7794
98	0,89	1,23	0,3907	0,8907
102	0,93	1,69	0,4545	0,9545
$x \ge 106$	1	2,15	0,4842	0,9842

Обозначения:

– точки графика $F^*(x)$;

– график функции $F(x)$.

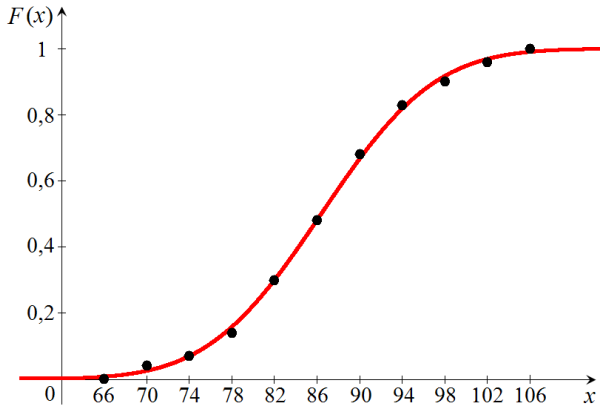


Рисунок 22

Пример 23. При обследовании 2000 теплиц было отобрано 110. Распределение их по объёму совокупных ежегодных продаж (ден. ед.) приведено в таблице:

Размер объёма совокупных ежегодных продаж, ден. ед.	менее 500	500-1000	1000-1500	1500-2000	2000-2500	Всего
Число теплиц	8	25	47	18	12	110

1. Используя критерий χ^2 Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что СВ X – размер объёма совокупных ежегодных продаж распределена по нормальному закону.

2. Найти:

а) вероятность того, что средний объём продаж во всех теплицах отличается от среднего объёма продаж в выборке не более чем на 100 ден. ед. (по абсолютной величине);

б) границы, в которых с вероятностью 0,97 заключена доля теплиц, объём продаж которых не более 1000 ден. ед.;

в) каким должен быть объём выборки, чтобы те же границы для доли теплиц, объём продаж которых не более 1000 ден. ед., можно было гарантировать с вероятностью 0,999?

Решение.

1. По условию $N = 2000$, $n = 110$. Найдём середины интервалов.

x_i	250	750	1250	1750	2250	Всего
m_i	8	25	47	18	12	110

Найдём числовые характеристики выборки:

Средняя выборочная

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{5} \tilde{x}_i m_i =$$

$$= \frac{1}{110}(250 \cdot 8 + 750 \cdot 25 + 1250 \cdot 47 + 1750 \cdot 18 + 2250 \cdot 12) = 1255;$$

выборочная дисперсия

$$D_в = \frac{1}{n}\sum_{i=1}^{5} x_i^2 m_i - \overline{x}_в^2 = 279524{,}7934;$$

выборочное среднее квадратическое отклонение

$$\sigma_в = \sqrt{D_в} = 528{,}701;$$

исправленное среднее квадратическое отклонение

$$s = \sqrt{\frac{n}{n-1}} \cdot \sigma_в = \sqrt{\frac{110}{109}} \cdot 528{,}701 = 531{,}121.$$

Используя критерий согласия Пирсона, при уровне значимости $\alpha = 0{,}05$ проверим гипотезу H_0: о «Нормальном распределении СВ X с параметрами $a = \overline{x}_в = 1255$ и $\sigma = s = 531{,}121$» при альтернативной гипотезе H_1: «СВ X не распределена по нормальному закону».

Вычислим вероятности p_i попадания СВ X в заданные интервалы с помощью функции Лапласа:

$$p_i = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right).$$

$$p_1 = P\{-\infty < x \le 500\} = \Phi(-1{,}42) - \Phi(-\infty) = -0{,}4222 + 0{,}5 = 0{,}0778;$$

$p_2 = P\{500 \le x < 1000\} = \Phi(-0,48) - \Phi(-1,42) = -0,1844 + 0,4222 = 0,2378$;

$p_3 = P\{1000 \le x < 1500\} = \Phi(0,46) - \Phi(-0,48) = 0,1772 + 0,1844 = 0,3616$;

$p_4 = P\{1500 \le x < 2000\} = \Phi(1,40) - \Phi(0,46) = 0,4192 - 0,1772 = 0,2420$;

$p_5 = P\{x \ge 2000\} = \Phi(+\infty) - \Phi(1,40) = 0,5 - 0,4192 = 0,0808$.

Для проведения расчётов заполним вспомогательную таблицу:

i	$[x_{i-1}; x_i)$	m_i	$m_i' = np_i$	$\frac{(m_i - m_i')^2}{m_i'}$
1	менее 500	8	8,558	0,0364
2	500-1000	25	26,158	0,0513
3	1000-1500	47	39,776	1,3120
4	1500-2000	18	26,62	2,7913
5	2000-2500	12	8,888	1,0896
Σ	–	**110**	**110**	**5,2806**

Наблюдаемое значение критерия согласия:

$$\chi^2_{набл} = \sum_{i=1}^{5} \frac{(m_i - m_i')^2}{m_i'} = 5,2806.$$

По таблице приложения для заданного уровня значимости $\alpha = 0,05$ и числа степеней свободы $k = 5 - 3 = 2$ найдём: $\chi^2_{кр}(0,05;\ 2) = 5,99$.

Т.к. $\chi^2_{набл} < \chi^2_{кр}$, то нулевая гипотеза о нормальном распределении принимается как не противоречащая опытным данным.

2. *а*) Для вычисления искомой вероятности применим формулу

$$P\{|\overline{x}_г - \overline{x}_в| \le \varepsilon\} = 2\Phi(t),$$

где $\varepsilon = 100$, t – аргумент функции Лапласа, который в случае неизвестного $\sigma_г$ и известного объёма генеральной совокупности N, определяется по формуле:

$$t = \frac{\varepsilon}{\sqrt{\frac{s^2}{n}\cdot\left(1-\frac{n}{N}\right)}} = \frac{100}{\sqrt{\frac{282089{,}241}{110}\cdot\left(1-\frac{110}{2000}\right)}} = \frac{100}{49{,}2281} = 2{,}03$$

Тогда $P\{|\overline{x}_г - \overline{x}_в| \le 100\} = 2\,\Phi(2{,}03) = 2\cdot 0{,}4788 = 0{,}9576.$

б) По исходной таблице найдём долю теплиц с объёмом продаж не более 1000 ден. ед.: $p^* = \frac{m}{n} = \frac{8+25}{110} = \frac{33}{110} = 0{,}3.$

Доля $p^* = 0{,}3$ с вероятностью 0,97 попадает в интервал $(p^* - \varepsilon, p^* + \varepsilon)$.

Т. к. $\gamma = 0{,}97 = P\{|p - p^*| < \varepsilon\} = 2\,\Phi(t)$, то $\Phi(t) = \frac{0{,}97}{2} = 0{,}485.$

По таблице приложения находим $t = 2{,}17$.

Тогда

$$\varepsilon = t\sqrt{\frac{p^*(1-p^*)}{n}\left(1-\frac{n}{N}\right)} = 2{,}17\cdot\sqrt{\frac{33}{110}\left(1-\frac{33}{110}\right)\cdot\frac{1}{110}\left(1-\frac{110}{2000}\right)} = 0{,}0922.$$

Окончательно находим доверительные границы:

$$0{,}3 - 0{,}0922 < p < 0{,}3 + 0{,}0922;$$

$$0{,}2078 < p < 0{,}3922.$$

в) Объём выборки:

$$n = \frac{p^*(1-p^*)}{\left(\frac{\varepsilon}{t}\right)^2 + \frac{p^*(1-p^*)}{N}},$$

где $\varepsilon = 0{,}0922$ (см. п. *б*), т.к. по условию задачи границы те же).

Для доверительной вероятности $2\Phi(t) = 0{,}999$ находим по таблице приложения значение аргумента $t = 3{,}3$.

Тогда: $n = \dfrac{0{,}3 \cdot 0{,}7}{\left(\dfrac{0{,}0922}{3{,}3}\right)^2 + \dfrac{0{,}3 \cdot 0{,}7}{2000}} = 237{,}2615.$

Окончательно получаем $n = 237$.

7.7. Элементы теории корреляции

Пример 24. Распределение 100 предприятий отрасли по объёму выпускаемой продукции X (тыс. единиц) и её себестоимости Y (руб.) представлено в таблице:

y / x	**48**	**58**	**68**	**78**	m_{x_i}
17	–	–	–	1	**1**
18	–	–	29	1	**30**
19	–	29	10	–	**39**
20	21	6	–	–	**27**
21	3	–	–	–	**3**
m_{y_j}	**24**	**35**	**39**	**2**	**100**

Необходимо:

1. Отметить на координатной плоскости точки $(x_i; y_i)$ данной выборки.

2. Вычислить групповые средние $\overline{x}_{y_j}$ и $\overline{y}_{x_i}$.

3. Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

а) вычислить коэффициент корреляции, на уровне значимости $\alpha = 0{,}05$ оценить его достоверность (значимость) и сделать вывод о тесноте и направлении связи;

б) найти уравнения прямых регрессии и построить их графики на одном чертеже с эмпирическими данными;

в) используя найденные уравнения, рассчитать показатели качества модели регрессии и сделать соответствующие выводы.

Решение.

1. Отметим на координатной плоскости точки данной выборки:

$(17;78)$, $(18;68)$, $(18;78)$, $(19;58)$,

$(19;68)$, $(20;48)$, $(20;58)$, $(21;48)$.

Заметим, что точки располагаются вдоль некоторой прямой линии (рис. 23).

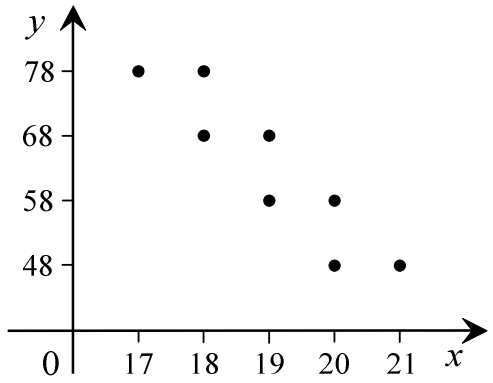


Рисунок 23

2. Для проведения расчётов заполним вспомогательные таблицы:

x	m_x	xm_x	x^2m_x
17	1	17	289
18	30	540	9720
19	39	741	14079
20	27	540	10800
21	3	63	1323
Σ	**100**	**1901**	**36211**

$$\bar{x} = \frac{1901}{100} = 19{,}01;$$

$$\overline{x^2} = \frac{36211}{100} = 362{,}11;$$

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 =$$

$$= 362{,}11 - 19{,}01^2 = 0{,}7299;$$

$$\sigma_x = \sqrt{D_x} = \sqrt{0{,}7299} = 0{,}854.$$

y	m_y	ym_y	y^2m_y
48	24	1152	55296
58	35	2030	117740
68	39	2652	180336
78	2	156	12168
Σ	**100**	**5990**	**365540**

$$\bar{y} = \frac{5990}{100} = 59{,}9;$$

$$\overline{y^2} = \frac{365540}{100} = 3655{,}4;$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 =$$

$$= 3655{,}4 - 59{,}9^2 = 67{,}39;$$

$$\sigma_y = \sqrt{D_y} = \sqrt{67{,}39} = 8{,}209.$$

x	y	m_{xy}	xym_{xy}
17	78	1	1326
18	68	29	35496
18	78	1	1404
19	58	29	31958
19	68	10	12920
20	48	21	20160
20	58	6	6960
21	48	3	3024
Σ		**100**	**113248**

$$\overline{xy} = \frac{113248}{100} = 1132{,}48$$

Найдём условные средние значения $\bar{x}_y$

$$\bar{x}_{y=48} = \frac{20 \cdot 21 + 21 \cdot 3}{24} = 20{,}13;$$

$$\bar{x}_{y=58} = \frac{19 \cdot 29 + 20 \cdot 6}{35} = 19{,}17;$$

$$\bar{x}_{y=68} = \frac{18 \cdot 29 + 19 \cdot 10}{39} = 18{,}26;$$

$$\bar{x}_{y=78} = \frac{17 \cdot 1 + 18 \cdot 1}{2} = 17{,}5.$$

y	m_y	$\bar{x}_{y_i}$	$(\bar{x}_{y_i} - 19{,}01)^2$	$m_y(\bar{x}_{y_i} - 19{,}01)^2$
48	24	20,13	1,24	29,84
58	35	19,17	0,03	0,91
68	39	18,26	0,57	22,15
78	2	17,5	2,28	4,56
	100			**Σ =57,46**

Тогда межгрупповые дисперсии:

$$\delta_x^2 = \frac{57{,}46}{100} = 0{,}5746 \quad \Rightarrow \quad \delta_x = \sqrt{0{,}5746} \approx 0{,}76.$$

Найдём условные средние значения $\bar{y}_x$

$$\bar{y}_{x=17} = \frac{78 \cdot 1}{1} = 78; \qquad \bar{y}_{x=18} = \frac{68 \cdot 29 + 78 \cdot 1}{30} = 68{,}33;$$

$$\bar{y}_{x=19} = \frac{58 \cdot 29 + 68 \cdot 10}{39} = 60{,}56; \qquad \bar{y}_{x=20} = \frac{48 \cdot 21 + 58 \cdot 6}{27} = 50{,}22;$$

$$\bar{y}_{x=21} = \frac{48 \cdot 3}{3} = 48.$$

x	m_x	$\bar{y}_{x_i}$	$(\bar{y}_{x_i} - 59{,}9)^2$	$m_x(\bar{y}_{x_i} - 59{,}9)^2$
17	1	78	327,61	327,61
18	30	68,33	71,12	2133,63
19	39	60,56	0,44	17,20
20	27	50,22	93,66	2528,80
21	3	48	141,61	424,83
	100			**Σ =5432,08**

$$\delta_y^2 = \frac{5432{,}08}{100} = 54{,}3208 \quad \Rightarrow \quad \delta_y = \sqrt{54{,}2878} \approx 7{,}37.$$

3. Найдём корреляционные отношения:

$$\eta_{XY} = \frac{\delta_x}{\sigma_x} = \frac{0{,}76}{0{,}854} = 0{,}89 \in [0{,}7;\ 1] \text{ и } \eta_{YX} = \frac{\delta_y}{\sigma_y} = \frac{7{,}37}{8{,}209} = 0{,}90 \in [0{,}7;\ 1].$$

Следовательно, между СВ X и Y имеется сильная корреляционная зависимость.

а) Найдём коэффициент корреляции:

$$r_в = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{1132{,}48 - 19{,}01 \cdot 59{,}9}{0{,}854 \cdot 8{,}209} = -0{,}89 \in [-1; -0{,}7].$$

Значит, имеется сильная убывающая линейная зависимость.

Проверим гипотезу $H_0 : r_г = 0$ (в генеральной совокупности линейная зависимость отсутствует) при альтернативе $H_1 : r_г \neq 0$ (имеется линейная зависимость).

Вычислим значение статистики:

$$T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1 - r_в^2}} = \frac{-0{,}89 \cdot \sqrt{100-2}}{\sqrt{1-(-0{,}89)^2}} = -19{,}32\,.$$

По условию $\alpha = 0{,}05$, $k = 100 - 2 = 98$, тогда по таблице приложения находим критическое значение $t_{кр} = t_{кр}(0{,}05; 98) = 1{,}98$.

Т. к. $|T_{набл}| > t_{кр}$, то нулевую гипотезу отвергаем и принимаем гипотезу H_1 о наличии линейной корреляции.

б) Составим уравнения линейной регрессии:

$y_x - \bar{y} = \rho_{YX}(x - \bar{x})$	$x_y - \bar{x} = \rho_{XY}(y - \bar{y})$
$\rho_{YX} = \frac{\sigma_y}{\sigma_x} \cdot r_в = \frac{8{,}209}{0{,}854} \cdot (-0{,}89) = -8{,}6$	$\rho_{XY} = \frac{\sigma_x}{\sigma_y} \cdot r_в = \frac{0{,}854}{8{,}209} \cdot (-0{,}89) = -0{,}1$
$y_x - 59{,}9 = -8{,}6 \cdot (x - 19{,}01)$	$x_y - 19{,}01 = -0{,}1(y - 59{,}9)$
$y_x = -8{,}6x + 223{,}4$	$x_y = -0{,}1y + 25$ (или $y = -10x + 250$)

Полученное уравнение показывает, что при увеличении объёма выпуска X на 1 тыс. единиц себестоимость Y уменьшается в среднем на 8,6 (руб.).	Полученное уравнение показывает, что для уменьшения себестоимости Y на 1 руб. необходимо в среднем увеличить объём выпуска X на 0,1 тыс. единиц.

Построим на одном чертеже (рис. 24) графики прямых регрессии:

$$y_x = -8{,}6x + 223{,}4 \text{ и } x_y = -0{,}1y + 25$$

и отметим экспериментальные данные (условные средние $\bar{x}_y$ и $\bar{y}_x$).

Обозначения:

.....	прямая $x_y = -0{,}1y + 25$
———	прямая $y_x = -8{,}6x + 223{,}4$
□□□	условные средние $\bar{x}_y$
○○○	условные средние $\bar{y}_x$
●	$M_0(\bar{x};\bar{y})$ – точка пересечения прямых регрессий, где $\bar{x} = 19{,}01$, $\bar{y} = 59{,}9$

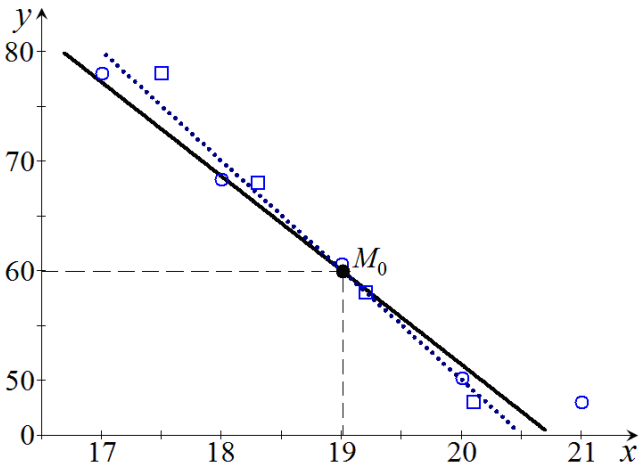


Рис. 24

в) Вычислим показатели качества найденной модели регрессии.

Заполним вспомогательные таблицы:

i	x_i	$\overline{y}_i$	$y_{x_i} = -8,6x_i + 223,4$	$\varepsilon_{y_i} = \overline{y}_i - y_{x_i}$	$\varepsilon_{y_i}^2$	$\left\|\frac{\varepsilon_{y_i}}{\overline{y}_i}\right\|$
1	17	78	77,2	0,8	0,64	0,010
2	18	68,3	68,6	$-0,3$	0,09	0,004
3	19	60,6	60	0,6	0,36	0,006
4	20	50,2	51,4	$-1,2$	1,44	0,010
5	21	48	42,8	5,2	27,04	0,108
Σ	**–**	**–**	**–**		**29,57**	**0,138**

j	y_j	$\overline{x}_j$	$x_{y_j} = -0,1y_j + 25$	$\varepsilon_{x_j} = \overline{x}_j - x_{y_j}$	$\varepsilon_{x_j}^2$	$\left\|\frac{\varepsilon_{x_j}}{\overline{x}_j}\right\|$
1	48	20,1	20,2	$-0,1$	0,01	0,005
2	58	19,2	19,2	0	0	0
3	68	18,3	18,2	0,1	0,01	0,005
4	78	17,5	17,2	0,3	0,09	0,017
Σ	**–**	**–**	**–**	**–**	**0,11**	**0,027**

Найдём теоретический коэффициент детерминации:

$$R^2 = r_в^2 = (-0,89)^2 = 0,7921.$$

Таким образом, $79,21\%$ *вариации себестоимости продукции объясняется уравнением линейной регрессии, остальные* $20,79\%$ *вариации себестоимости обусловлены влиянием не учтённых в модели факторов.*

Найдём средние квадратические ошибки:

$$S_{\varepsilon_y} = \sqrt{\frac{1}{5-2}\sum_{i=1}^{5}\varepsilon_{y_i}^2} = \sqrt{\frac{29,57}{3}} = 3,14 \text{ и } S_{\varepsilon_x} = \sqrt{\frac{1}{4-2}\sum_{i=1}^{4}\varepsilon_{x_i}^2} = \sqrt{\frac{0,11}{2}} = 0,23.$$

Поскольку $S_{\varepsilon_y} < \sigma_y$ и $S_{\varepsilon_x} < \sigma_x$, то *найденные модели линейной регрессии целесообразно использовать.*

Найдём средние ошибки аппроксимации:

$$A_y = \frac{1}{5}\sum_{i=1}^{5}\left|\frac{\varepsilon_{y_i}}{\overline{y}_i}\right| = \frac{0{,}138}{5} = 0{,}0276$$

$$\text{и} \quad A_x = \frac{1}{4}\sum_{i=1}^{4}\left|\frac{\varepsilon_{x_j}}{\overline{x}_j}\right| = \frac{0{,}027}{4} = 0{,}00675.$$

Средние ошибки составляют 2,76% и 0,675%, *что свидетельствует о незначительных погрешностях моделей.*

Пример 25. Дана двумерная выборка:

x \ y	**48**	**67**	**86**	m_{x_i}
0	1	–	–	**1**
1	2	29	–	**31**
2	–	2	30	**32**
3	–	29	6	**35**
4	1	–	–	**1**
m_{y_j}	**4**	**60**	**36**	**100**

1. вычислить корреляционное отношение η_{YX};

2. вычислить коэффициент корреляции $r_в$;

3. выдвинуть гипотезу о наличии или отсутствии линейной зависимости на уровне значимости $\alpha = 0{,}01$;

4. найти соответствующее уравнение регрессии;

5. вычислить теоретические значения условной средней y_x.

Результаты представить в виде таблицы:

x_i	x_1	x_2	$\cdots$	x_k
$\bar{y}_{x=x_i}$	$\bar{y}_1$	$\bar{y}_2$	$\cdots$	$\bar{y}_k$
y_{x_i}	y_1	y_2	$\cdots$	y_k

6. построить линию регрессии и выборочные средние значения $\bar{y}_x$.

Решение.

Для проведения расчётов заполним таблицы:

x	m_x	xm_x	x^2m_x
0	1	0	0
1	31	31	31
2	32	64	128
3	35	105	315
4	1	4	16
Σ	**100**	**204**	**490**

$$\bar{x} = \frac{204}{100} = 2{,}04.$$

$$\overline{x^2} = \frac{490}{100} = 4{,}9.$$

$$D_x = \overline{x^2} - \bar{x}^2 = 4{,}9 - 2{,}04^2 = 0{,}7384.$$

$$\sigma_x = \sqrt{D_x} = \sqrt{0{,}7384} = 0{,}859.$$

y	m_y	ym_y	y^2m_y
48	4	192	9216
67	60	4020	269340
86	36	3096	266256
Σ	**100**	**7308**	**544812**

$$\bar{y} = \frac{7308}{100} = 73{,}08.$$

$$\overline{y^2} = \frac{544812}{100} = 5448{,}12.$$

$$D_y = \overline{y^2} - \bar{y}^2 = 5448{,}12 - 73{,}08^2 = 107{,}4336.$$

$$\sigma_y = \sqrt{D_y} = \sqrt{107{,}4336} = 10{,}365.$$

x	y	m_{xy}	xym_{xy}
0	48	1	0
1	48	2	96
1	67	29	1943
2	67	2	268
2	86	30	5160
3	67	29	5829
3	86	6	1548
4	48	1	192
Σ		**100**	**15036**

Следовательно, $\overline{xy} = \frac{15036}{100} = 150{,}36$.

Найдём условные средние значения $\bar{y}_x$ и межгрупповую дисперсию:

$$\bar{y}_{x=0} = \frac{48 \cdot 1}{1} = 48; \qquad \bar{y}_{x=1} = \frac{48 \cdot 2 + 67 \cdot 29}{31} = 65{,}8$$

$$\bar{y}_{x=2} = \frac{67 \cdot 2 + 86 \cdot 30}{32} = 84{,}8; \qquad \bar{y}_{x=3} = \frac{67 \cdot 29 + 86 \cdot 6}{35} = 70{,}3$$

$$\bar{y}_{x=4} = \frac{48 \cdot 1}{1} = 48.$$

x	m_x	$\bar{y}_{x_i}$	$\left(\bar{y}_{x_i} - 73{,}08\right)^2$	$m_x\left(\bar{y}_{x_i} - 73{,}08\right)^2$
0	1	48	629,0064	629,0064
1	31	65,8	52,9984	1642,95
2	32	84,8	137,3584	4395,469
3	35	70,3	7,07284	270,494
4	1	48	629,0064	629,0064
	100			**Σ =7566,926**

$$\delta_y^2 = \frac{7566{,}926}{100} = 75{,}66926 \quad \Rightarrow \quad \delta_y = \sqrt{75{,}66926} \approx 8{,}699.$$

1. Вычислим корреляционное отношение:

$$\eta_{YX} = \frac{\delta_y}{\sigma_y} = \frac{8{,}699}{10{,}365} = 0{,}839.$$

Определяем, что имеется сильная корреляционная зависимость.

2. Найдём коэффициент корреляции:

$$r_в = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{150{,}36 - 2{,}04 \cdot 73{,}08}{0{,}859 \cdot 10{,}365} = 0{,}143.$$

Определяем, что линейная зависимость практически отсутствует.

3. Проверим гипотезу $H_0 : r_г = 0$ (в генеральной совокупности линейная зависимость отсутствует) при альтернативе $H_1 : r_г \neq 0$ (имеется линейная зависимость).

Вычислим значение статистики:

$$T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1 - r_в^2}} = \frac{0{,}143 \cdot \sqrt{100-2}}{\sqrt{1 - 0{,}143^2}} = 1{,}43.$$

$$\alpha = 0{,}01,\ l = 100 - 2 = 98 \ \Rightarrow\ t_{кр} = t_{кр}(\alpha; l) = 2{,}62.$$

Т.к. $|T_{набл}| < t_{кр}$, то гипотезу H_1 отвергаем и принимаем гипотезу H_0 о наличии нелинейной корреляции в генеральной совокупности.

4. По виду расположения условных средних значений $\bar{y}_x$ на плоскости (рис. 25) предполагаем квадратическую зависимость.

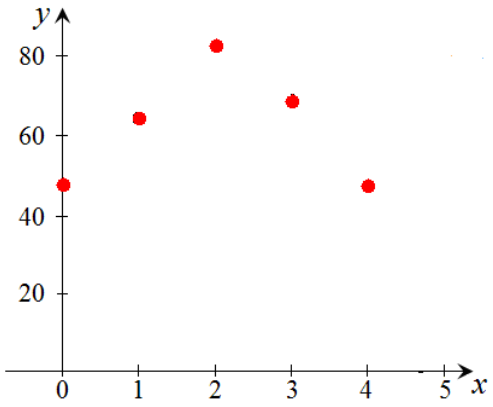


Рисунок 25

Составим уравнение нелинейной параболической регрессии: $y_x = ax^2 + bx + c$.

Заполним вспомогательную таблицу для вычисления коэффициентов:

x	$\bar{y}_x$	m_x	xm_x	x^2m_x	x^3m_x	x^4m_x	$\bar{y}_x m_x$	$x\bar{y}_x m_x$	$x^2\bar{y}_x m_x$
0	48	1	0	0	0	0	48	0	0
1	65,8	31	31	31	31	31	2039,8	2039,8	2039,8
2	84,8	32	64	128	256	512	2713,6	5427,2	10854,4
3	70,3	35	105	315	945	2835	2460,5	7381,5	22144,5
4	48	1	4	16	64	256	48	192	768
Σ	**–**	**100**	**204**	**490**	**1296**	**3634**	**7309,9**	**15040,5**	**35806,7**

$$\begin{cases} a\cdot\sum\limits_{i=1}^{n} x_i^{\,4} + b\cdot\sum\limits_{i=1}^{n} x_i^{\,3} + c\cdot\sum\limits_{i=1}^{n} x_i^{\,2} = \sum\limits_{i=1}^{n} x_i^{\,2}\, y_i, \\ a\cdot\sum\limits_{i=1}^{n} x_i^{\,3} + b\cdot\sum\limits_{i=1}^{n} x_i^{\,2} + c\cdot\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i\, y_i, \\ a\cdot\sum\limits_{i=1}^{n} x_i^{\,2} + b\cdot\sum\limits_{i=1}^{n} x_i + c\cdot n = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

$$\Rightarrow \begin{cases} 490a + 204b + 100c = 7309{,}9, \\ 1296a + 490b + 204c = 15040{,}5, \\ 3634a + 1296b + 490c = 35806{,}7. \end{cases}$$

Решая систему, например, методом Жордано-Гаусса, получим:

$$\begin{cases} a = -12{,}2; \\ b = 50{,}7; \\ c = 29{,}5. \end{cases}$$

Искомое уравнение регрессии принимает вид:

$$\boxed{y_x = -12{,}2x^2 + 50{,}7x + 29{,}5}.$$

5. Заполним таблицу:

X	0	1	2	3	4
$\bar{y}_x$	48	65,8	84,8	70,3	48
y_x	29,5	68	82,1	71,8	37,1

6. Построим на одном чертеже график параболической регрессии

$$y_x = -12{,}2x^2 + 50{,}7x + 29{,}5$$

и нанесём экспериментальные данные $\bar{y}_x$ (рис. 26).

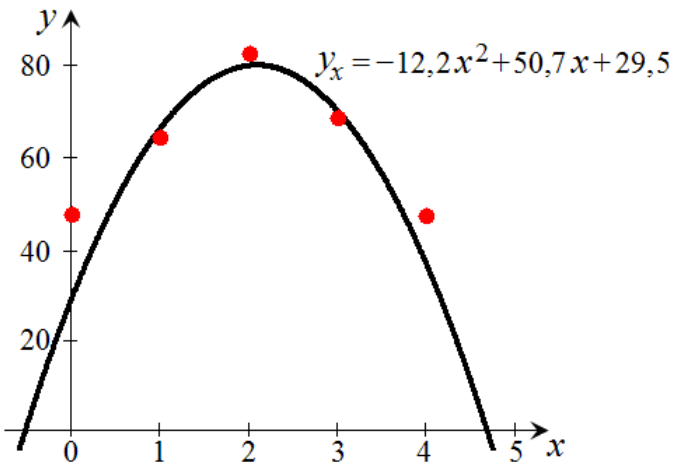


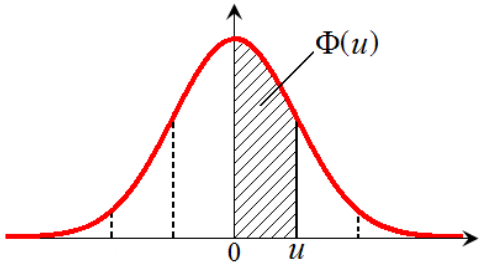
Рисунок 26

ПРИЛОЖЕНИЯ

Таблица 1

Таблица значений функции Лапласа

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$



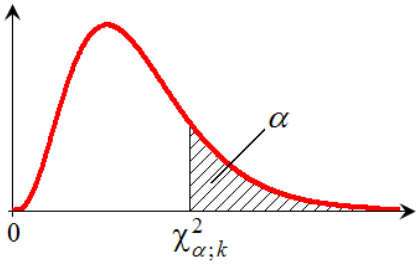
x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	**0,26**	0,1026	**0,52**	0,1985	**0,78**	0,2823
0,01	0,0040	**0,27**	0,1064	**0,53**	0,2019	**0,79**	0,2852
0,02	0,0080	**0,28**	0,1103	**0,54**	0,2054	**0,80**	0,2881
0,03	0,0120	**0,29**	0,1141	**0,55**	0,2088	**0,81**	0,2910
0,04	0,0160	**0,30**	0,1179	**0,56**	0,2123	**0,82**	0,2939
0,05	0,0199	**0,31**	0,1217	**0,57**	0,2157	**0,83**	0,2967
0,06	0,0239	**0,32**	0,1255	**0,58**	0,2190	**0,84**	0,2995
0,07	0,0279	**0,33**	0,1293	**0,59**	0,2224	**0,85**	0,3023
0,08	0,0319	**0,34**	0,1331	**0,60**	0,2257	**0,86**	0,3051
0,09	0,0359	**0,35**	0,1368	**0,61**	0,2291	**0,87**	0,3078
0,10	0,0398	**0,36**	0,1406	**0,62**	0,2324	**0,88**	0,3106
0,11	0,0438	**0,37**	0,1443	**0,63**	0,2357	**0,89**	0,3133
0,12	0,0478	**0,38**	0,1480	**0,64**	0,2389	**0,90**	0,3159
0,13	0,0517	**0,39**	0,1517	**0,65**	0,2422	**0,91**	0,3186
0,14	0,0557	**0,40**	0,1554	**0,66**	0,2454	**0,92**	0,3212
0,15	0,0596	**0,41**	0,1591	**0,67**	0,2486	**0,93**	0,3238
0,16	0,0636	**0,42**	0,1628	**0,68**	0,2517	**0,94**	0,3264
0,17	0,0675	**0,43**	0 1664	**0,69**	0,2549	**0,95**	0,3289
0,18	0,0714	**0,44**	0,1700	**0,70**	0,2580	**0,96**	0,3315
0,19	0,0753	**0,45**	0,1736	**0,71**	0,2611	**0,97**	0,3340
0,20	0,0793	**0,46**	0,1772	**0,72**	0,2642	**0,98**	0,3365
0,21	0,0832	**0,47**	0,1808	**0,73**	0,2673	**0,99**	0,3389
0,22	0,0871	**0,48**	0,1844	**0,74**	0,2703	**1,00**	0,3413
0,23	0,0910	**0,49**	0,1879	**0,75**	0,2734	**1,01**	0,3438
0,24	0,0948	**0,50**	0,1915	**0,76**	0,2764	**1,02**	0,3461
0,25	0,0987	**0,51**	0,1950	**0,77**	0,2794	**1,03**	0,3485

Продолжение таблицы 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,04	0,3508	**1,43**	0,4236	**1,82**	0,4656	**2,42**	0,4922
1,05	0,3531	**1,44**	0,4251	**1,83**	0,4664	**2,44**	0,4927
1,06	0,3554	**1,45**	0,4265	**1,84**	0,4671	**2,46**	0,4931
1,07	0,3577	**1,46**	0,4279	**1,85**	0,4678	**2,48**	0,4934
1,08	0,3599	**1,47**	0,4292	**1,86**	0,4686	**2,50**	0,4938
1,09	0,3621	**1,48**	0,4306	**1,87**	0,4693	**2,52**	0,4941
1,10	0,3643	**1,49**	0,4319	**1,88**	0,4699	**2,54**	0,4945
1,11	0,3665	**1,50**	0,4332	**1,89**	0,4706	**2,56**	0,4948
1,12	0,3686	**1,51**	0,4345	**1,90**	0,4713	**2,58**	0,4951
1,13	0,3708	**1,52**	0,4357	**1,91**	0,4719	**2,60**	0,4953
1,14	0,3729	**1,53**	0,4370	**1,92**	0,4726	**2,62**	0,4956
1,15	0,3749	**1,54**	0,4382	**1,93**	0,4732	**2,64**	0,4959
1,16	0,3770	**1,55**	0,4394	**1,94**	0,4738	**2,66**	0,4961
1,17	0,3790	**1,56**	0,4406	**1,95**	0,4744	**2,68**	0,4963
1,18	0,3810	**1,57**	0,4418	**1,96**	0,4750	**2,70**	0,4965
1,19	0,3830	**1,58**	0,4429	**1,97**	0,4756	**2,72**	0,4967
1,20	0,3849	**1,59**	0,4441	**1,98**	0,4761	**2,74**	0,4969
1,21	0,3869	**1,60**	0,4452	**1,99**	0,4767	**2,76**	0,4971
1,22	0,3883	**1,61**	0,4463	**2,00**	0,4772	**2,78**	0,4973
1,23	0,3907	**1,62**	0,4474	**2,02**	0,4783	**2,80**	0,4974
1,24	0,3925	**1,63**	0,4484	**2,04**	0,4793	**2,82**	0,4976
1,25	0,3944	**1,64**	0,4495	**2,06**	0,4803	**2,84**	0,4977
1,26	0,3962	**1,65**	0,4505	**2,08**	0,4812	**2,86**	0,4979
1,27	0,3980	**1,66**	0,4515	**2,10**	0,4821	**2,88**	0,4980
1,28	0,3997	**1,67**	0,4525	**2,12**	0,4830	**2,90**	0,4981
1,29	0,4015	**1,68**	0,4535	**2,14**	0,4838	**2,92**	0,4982
1,30	0,4032	**1,69**	0,4545	**2,16**	0,4846	**2,94**	6,4984
1,31	0,4049	**1,70**	0,4554	**2,18**	0,4854	**2,96**	0,4985
1,32	0,4066	**1,71**	0,4564	**2,20**	0,4861	**2,98**	0,4986
1,33	0,4082	**1,72**	0,4573	**2,22**	0,4868	**3,00**	0,49865
1,34	0,4099	**1,73**	0,4582	**2,24**	0,4875	**3,20**	0,49931
1,35	0,4115	**1,74**	0,4591	**2,26**	0,4881	**3,40**	0,49966
1,36	0,4131	**1,75**	0,4599	**2,28**	0,4887	**3,60**	0,499841
1,37	0,4147	**1,76**	0,4608	**2,30**	0,4893	**3,80**	0,499928
1,38	0,4162	**1,77**	0,4616	**2,32**	0,4898	**4,00**	0,499968
1,39	0,4177	**1,78**	0,4625	**2,34**	0,4904	**4,50**	0,499997
1,40	0,4192	**1,79**	0,4633	**2,36**	0,4909	**5,00**	0,5
1,41	0,4207	**1,80**	0,4641	**2,38**	0,4913	**> 5**	0,5
1,42	0,4222	**1,81**	0,4649	**2,40**	0,4918	**∞**	0,5

Таблица 2

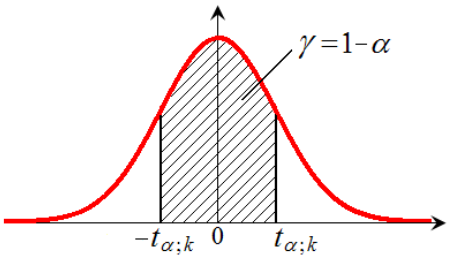
Критические точки распределения χ^2



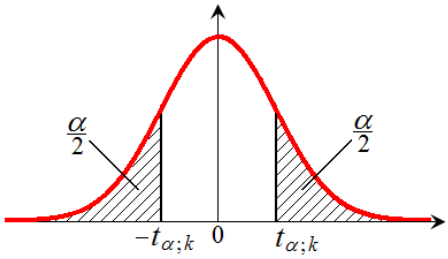
Число степеней свободы *k*	**Уровень значимости α**					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,89**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	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 3

Критические точки $t_{\alpha;k}$ распределения Стьюдента



Двусторонняя критическая область



Односторонние критические области

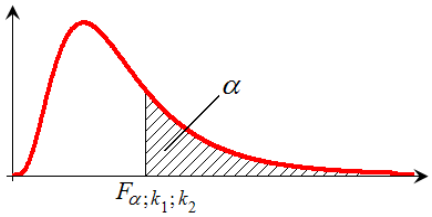
Число степеней свободы k	**Уровень значимости α (двусторонняя критическая область)**						
	0,2	**0,1**	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	3,08	6,31	12,7	31,82	63,7	318,3	637,0
2	1,89	2,92	4,30	6,97	9,92	22,33	31,6
3	1,64	2,35	3,18	4,54	5,84	10,22	12,9
4	1,53	2,13	2,78	3,75	4,60	7,17	8,61
5	1,48	2,01	2,57	3,37	4,03	5,89	6,86
6	1,44	1,94	2,45	3,14	3,71	5,21	5,96
7	1,42	1,89	2,36	3,00	3,50	4,79	5,40
8	1,40	1,86	2,31	2,90	3,36	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	4,03	4,44
12	1,36	1,78	2,18	2,68	3,05	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,69	4,01
17	1,33	1,74	2,11	2,57	2,90	3,65	3,96
18	1,33	1,73	2,10	2,55	2,88	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,58	3,88
20	1,33	1,73	2,09	2,53	2,85	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,49	3,77
24	1,32	1,71	2,06	2,49	2,80	3,47	3,74
25	1,32	1,71	2,06	2,49	2,79	3,45	3,72
26	1,32	1,71	2,06	2,48	2,78	3,44	3,71
27	1,31	1,71	2,05	2,47	2,77	3,42	3,69
28	1,31	1,70	2,05	2,46	2,76	3,40	3,66
29	1,31	1,70	2,05	2,46	2,76	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,28	1,64	1,96	2,33	2,58	3,09	3,29
Число степеней свободы k	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)						

Таблица 4

Критические точки $F_{\alpha;k_1;k_2}$ распределения Фишера

(k_1 – число степеней свободы бóльшей дисперсии,

k_2 – число степеней свободы мéньшей дисперсии)



Уровень значимости $\alpha = 0,01$

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	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	4,89	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	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости $\alpha = 0,05$

k_1 / k_2	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	3,84	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	3,29	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 5

Таблица значений $t_\gamma = t(\gamma, n)$

n \ γ	**0,95**	**0,99**	**0,999**	n \ γ	**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**	1,001	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 6

Таблица значений $q = q(\gamma, n)$

n \ γ	**0,95**	**0,99**	**0,999**	n \ γ	**0,95**	**0,99**	**0,999**
5	1,37	2,67	5,64	**20**	0,37	0,58	0,88
6	1,09	2,01	3,88	**25**	0,32	0,49	0,73
7	0,92	1,62	2,98	**30**	0,28	0,43	0,63
8	0,80	1,38	2,42	**35**	0,26	0,38	0,56
9	0,71	1,20	2,06	**40**	0,24	0,35	0,50
10	0,65	1,08	1,80	**45**	0,22	0,32	0,46
11	0,59	0,98	1,60	**50**	0,21	0,30	0,43
12	0,55	0,90	1,45	**60**	0,188	0,269	0,38
13	0,52	0,83	1,33	**70**	0,174	0,245	0,34
14	0,48	0,78	1,23	**80**	0,161	0,226	0,31
15	0,46	0,73	1,15	**90**	0,151	0,211	0,29
16	0,44	0,70	1,07	**100**	0,143	0,198	0,27
17	0,42	0,66	1,01	**150**	0,115	0,160	0,211
18	0,40	0,63	0,96	**200**	0,099	0,136	0,185
19	0,39	0,60	0,92	**250**	0,089	0,120	0,162

Список литературы

1. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для студентов вузов, обучающихся по экономическим специальностям / [Н.Ш. Кремер и др.]; под ред. проф. Н.Ш. Кремера. - 3-е изд. - М.: ЮНИТИ-ДАНА, 2012. - 551 с. — (Серия «Золотой фонд российских учебников»)
2. Кремер, Н. Ш. Высшая математика для экономистов: учебник для студентов вузов, обучающихся по экономическим специальностям / [Н.Ш. Кремер и др.]; под ред. проф. Н.Ш. Кремера. - 3-е изд. - М.: ЮНИТИ-ДАНА, 2010. - 479 с. — (Серия «Золотой фонд российских учебников»)
3. Агишева, Д.К., Зотова, С.А., Матвеева, Т.А., Светличная В.Б. Математическая статистика. Учебное пособие /Д.К. Агишева, С.А.Зотова, Т.А. Матвеева, В.Б. Светличная; ВПИ (филиал) ВолгГТУ.- Волгоград, 2010.- 160 с.
4. Владимирский, Б. М. Математика. Общий курс: учебник / Б. М. Владимирский, А. Б. Горстко, Я. М. Ерусалимский. — 4-е изд., стер. — Санкт-Петербург: Лань, 2022. — 960 с.
5. Письменный, Д. Т. Конспект лекций по теории вероятностей и математической статистики/ Д. Т. Письменный. - М.: Айрис-пресс, 2015. - 288 с.- (Высшее образование).
6. Письменный, Д. Т. Конспект лекций по высшей математике: полный курс / Д. Т. Письменный / 8-е изд. - М.: Айрис-пресс, 2022. - 608 с.- (Высшее образование).

Электронное учебное издание

Татьяна Александровна **Матвеева**
Виктория Борисовна **Светличная**
Джамиля Алиевна **Мустафина**
Ирина Викторовна **Ребро**

Математика. Часть V.

Практикум по математической статистики

Учебное пособие

Электронное издание сетевого распространения

Редактор Матвеева Н.И.

Темплан 2024 г. Поз. № 28.
Подписано к использованию 14.05.2024. Формат 60x84 1/16.
Гарнитура Times. Усл. печ. л. 6,4.

Волгоградский государственный технический университет.
400005, г. Волгоград, пр. Ленина, 28, корп. 1.

ВПИ (филиал) ВолгГТУ.
404121, г. Волжский, ул. Энгельса, 42а.