

Сайт: [www.resolventa.ru](http://www.resolventa.ru) , E-mail: [resolventa@list.ru](mailto:resolventa@list.ru)

**Доктор физико-математических наук, профессор**

**К. Л. САМАРОВ**

**МАТЕМАТИКА**

**Учебно-методическое пособие по разделу**

**МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

**© К. Л. Самаров, 2009**

Сайт: [www.resolventa.ru](http://www.resolventa.ru) , E-mail: [resolventa@list.ru](mailto:resolventa@list.ru)

## СОДЕРЖАНИЕ

1. Понятия генеральной совокупности и выборки.....	3
2. Группировка выборочных данных. Интервальный вариационный ряд. Дискретный вариационный ряд .....	3
3. Вычисление характеристик выборки.....	6
4. Точечные и интервальные оценки параметров распределения изучаемой случайной величины. Статистика. Доверительный интервал.....	9
5. Распределения $\chi^2$ , Стьюдента и Фишера.....	11
6. Проверка гипотезы о нормальном распределении изучаемой случайной величины. Критерий согласия $\chi^2$ Пирсона.....	12
7. Группировка выборочных данных двумерной выборки. Корреляционная таблица.....	14
8. Вычисление характеристик двумерной выборки. Прямая линия регрессии.....	15
9. Примеры.....	18
10. Статистические таблицы.....	24
ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ .....	27
ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ .....	28
ЛИТЕРАТУРА .....	30

## 1. Понятия генеральной совокупности и выборки

Предположим, что  $\xi$  – случайная величина, о которой нет полной информации, однако известно некоторое конечное множество принимаемых ею значений. Основная задача математической статистики заключается в том, чтобы на основании этих данных исследовать свойства случайной величины  $\xi$ .

Определение. Известное множество значений  $X = \{x_1, x_2, \dots, x_n\}$ , принимаемых случайной величиной  $\xi$ , называется *выборкой* из *генеральной совокупности*  $\Gamma_\xi$  всех значений случайной величины, а количество  $n$  элементов выборки – *объемом* выборки.

Для того чтобы получить достоверные сведения об изучаемой случайной величине, значения  $x_1, x_2, \dots, x_n$  должны быть выбраны из генеральной совокупности случайным образом, а объем выборки должен быть достаточно большим. Выборка, удовлетворяющая таким условиям, называется *репрезентативной* (представительной).

Определение. Функция распределения  $F_\xi$  изучаемой случайной величины  $\xi$  называется *теоретической функцией распределения*.

## 2. Группировка выборочных данных. Интервальный вариационный ряд.

### Дискретный вариационный ряд

Изучение свойств случайной величины  $\xi$ , начинается с *группировки выборочных данных*. Для проведения группировки выборочные данные сначала нужно перенумеровать так, чтобы они оказались расположенными в порядке возрастания. В результате возникает выборка  $X = \{x_1, x_2, \dots, x_n\}$ , элементы которой удовлетворяют неравенствам  $x_1 \leq x_2 \leq \dots \leq x_n$ . Такая выборка называется *вариационным рядом*. После того, как вариационный ряд получен, осуществляется переход к построению групп. С этой целью вводится в рассмотрение произвольный полуинтервал  $[a; b)$ , содержащий все выборочные данные  $x_1, x_2, \dots, x_n$ . Если разделить этот полуинтервал точками  $a_1, a_2, \dots, a_{k-1}$  на  $k$  полуинтервалов

$$[a; a_1), [a_1; a_2), [a_2; a_3), \dots, [a_{k-1}; b),$$

имеющих одинаковую длину  $h$  (число  $k$  может быть любым натуральным числом), и воспользоваться обозначениями  $a_0 = a$  и  $a_k = b$ , то полуинтервалы примут вид

$$[a_0; a_1), [a_1; a_2), [a_2; a_3), \dots, [a_{k-1}; a_k),$$

и можно ввести следующее определение.

*Интервальным вариационным рядом* называется таблица

$[a_0; a_1)$	$[a_1; a_2)$	...	$[a_{k-1}; a_k)$
$m_1$	$m_2$	...	$m_k,$

элементы  $m_1, m_2, \dots, m_k$  которой, называемые *частотами*, являются целыми числами, причем число  $m_1$  равно количеству выборочных данных, попавших в полуинтервал  $[a_0; a_1)$ , число  $m_2$  равно количеству выборочных данных, попавших в полуинтервал  $[a_1; a_2)$ , и т.д.

Операция построения интервального вариационного ряда по выборке называется *группировкой выборочных данных*, число  $k$  является *количеством групп* выборочных данных, а число  $h$  называется *шагом вариации*.

Замечание. Частоты  $m_1, m_2, \dots, m_k$  удовлетворяет соотношению

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

Числа

$$W_i = \frac{m_i}{n}, i = 1, 2, \dots, k$$

называются *относительными частотами* и удовлетворяют соотношению

$$\sum_{i=1}^k W_i = 1.$$

Числа

$$f_i = \frac{m_i}{n \cdot h} = \frac{W_i}{h}, i = 1, 2, \dots, k$$

называются *плотностями относительных частот*.

Подсчитав плотности относительных частот, можно построить *гистограмму* плотностей относительных частот, являющуюся удобным графическим представлением интервального вариационного ряда. С этой целью сначала на оси абсцисс координатной плоскости изобразим точки с координатами

$$a_0, a_1, a_2, \dots, a_{k-1}, a_k.$$

После этого над каждым из  $k$  отрезков

$$[a_0; a_1], [a_1; a_2], [a_2; a_3], \dots, [a_{k-1}; a_k]$$

построим прямоугольник, нижним основанием которого является сам этот отрезок, а высота является соответствующей плотностью относительной частоты. В результате получится гистограмма, состоящая из  $k$  прямоугольников, общая площадь которых равна 1. Гистограмма дает представление о плотности распределения измеряемой случайной величины.

*Дискретным вариационным рядом* называется таблица

$x'_1$	$x'_2$	...	$x'_k$
$m_1$	$m_2$	...	$m_k$

в которой число  $x'_1$  является серединой отрезка  $[a_0; a_1]$ , число  $x'_2$  является серединой отрезка  $[a_1; a_2]$ , ..., число  $x'_k$  является серединой отрезка  $[a_{k-1}; a_k]$ .

Поскольку все отрезки

$$[a_0; a_1], [a_1; a_2], [a_2; a_3], \dots, [a_{k-1}; a_k]$$

имеют одну и ту же длину  $h$ , то и все отрезки

$$[x'_1; x'_2], [x'_2; x'_3], \dots, [x'_{k-1}; x'_k]$$

имеют ту же длину  $h$  (*шаг вариации*). Другими словами, справедливо соотношение

$$x'_i = x'_1 + (i-1)h, \quad i = 1, 2, \dots, k .$$

*Полигоном (многоугольником) выборочного распределения* называется ломаная линия на координатной плоскости, вершины которой имеют в указанном порядке следующие координаты:

$$(x'_1; m_1), (x'_2; m_2), \dots, (x'_k; m_k) .$$

*Полигоном (многоугольником) относительных частот* называется ломаная линия на координатной плоскости, вершины которой имеют в указанном порядке следующие координаты:

$$(x'_1; W_1), (x'_2; W_2), \dots, (x'_k; W_k) .$$

Полигоны выборочного распределения и относительных частот являются удобными графическими представлениями дискретного вариационного ряда.

### 3. Вычисление характеристик выборки

Предположим, что выборка объема  $n$  из генеральной совокупности значений изучаемой случайной величины  $\xi$  задана дискретным вариационным рядом

$$X: \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_k \\ \hline m_1 & m_2 & \dots & m_k, \\ \hline \end{array}$$

и рассмотрим дискретную случайную величину  $\xi_{\text{выб}}$ , закон распределения которой имеет вид

$$\xi_{\text{выб}}: \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & \dots & x_k \\ \hline W_1 & W_2 & \dots & W_k, \\ \hline \end{array}$$

где  $W_1, W_2, \dots, W_k$  – относительные частоты.

- Функция распределения  $F(x)$  дискретной случайной величины  $\xi_{\text{выб}}$  называется *эмпирической (выборочной) функцией распределения* изучаемой случайной величины  $\xi$  и, в соответствии с определением, введенным в Модуле 7, удовлетворяет условию

$$F(x) = P\{\xi_{\text{выб}} \leq x\}.$$

- *Модой* распределения называется значение случайной величины  $\xi_{\text{выб}}$ , обладающее наибольшей относительной плотностью (наибольшей вероятностью).

Замечание. У распределения может существовать несколько мод.

- *Квантилью* уровня  $l$ , где  $l$  – произвольное число, заключенное в пределах  $0 < l < 1$ , называется число  $x_l$ , являющееся корнем уравнения

$$F(x) = l.$$

Замечание. Квантили могут определяться неоднозначно.

- *Нижней квантилью, медианой и верхней квантилью* называются квантили уровней 0.25, 0,5 и 0.75, соответственно.

- Математическое ожидание случайной величины  $\xi_{\text{выб}}$  называется *выборочным средним значением* и вычисляется по формуле

$$M\xi_{\text{выб}} = \frac{1}{n} \sum_{i=1}^k m_i x_i$$

или по эквивалентной формуле

$$M\xi_{\text{выб}} = \sum_{i=1}^k W_i x_i.$$

- Дисперсия случайной величины  $\xi_{\text{выб}}$  называется *выборочной дисперсией* и вычисляется по формуле

$$D\xi_{\text{выб}} = \frac{1}{n} \sum_{i=1}^k m_i x_i^2 - (M\xi_{\text{выб}})^2$$

или по эквивалентной формуле

$$D\xi_{\text{выб}} = \sum_{i=1}^k W_i x_i^2 - (M\xi_{\text{выб}})^2.$$

• *Исправленной выборочной дисперсией* называется число, обозначаемое  $s^2$  и определяемое по формуле

$$s^2 = \frac{n}{n-1} \cdot D\xi_{\text{выб}}.$$

• Число  $\sigma_{\text{выб}}$ , равное квадратному корню из выборочной дисперсии, называется *выборочным средним квадратическим отклонением*.

• *Моментом порядка  $p$  (моментом  $p$ -го порядка)*, где  $p$  – натуральное число, называется число, определяемое по формуле

$$M\xi_{\text{выб}}^p = \sum_{i=1}^k W_i x_i^p.$$

Момент  $p$ -го порядка, деленный на  $p$ -ю степень выборочного среднего квадратического отклонения, называется *нормированным моментом  $p$ -го порядка*.

• *Центральным моментом порядка  $p$*  называется число, определяемое по формуле

$$M(\xi_{\text{выб}} - M\xi_{\text{выб}})^p = \sum_{i=1}^k W_i (x_i - M\xi_{\text{выб}})^p.$$

Центральный момент  $p$ -го порядка, деленный на  $p$ -ю степень выборочного среднего квадратического отклонения, называется *нормированным центральным моментом  $p$ -го порядка*.

Замечание. Дисперсия является центральным моментом 2-го порядка.

• *Асимметрией* распределения называется число, определяемое по формуле

$$\mu = \frac{M(\xi_{\text{выб}} - M\xi_{\text{выб}})^3}{\sigma_{\text{выб}}^3}.$$

- *Экссесом* распределения называется число, определяемое по формуле

$$\varepsilon = \frac{M(\xi_{\text{выб}} - M\xi_{\text{выб}})^4}{\sigma_{\text{выб}}^4}.$$

Замечание. Асимметрия и эксцесс являются центральными нормированными моментами 3-го и 4-го порядков, соответственно.

В ряде работ по экономике используется нормированный момент 1-го порядка, равный выборочному среднему значению, деленному на выборочное среднее квадратическое отклонение.

#### **4. Точечные и интервальные оценки параметров распределения изучаемой случайной величины. Статистика. Доверительный интервал**

Предположим, что *теоретическое распределение* (распределение изучаемой случайной величины  $\xi$ ) зависит от числового параметра  $\lambda$ , значение которого неизвестно. Таким параметром может быть, например, математическое ожидание случайной величины  $\xi$ .

Задача, состоящая в том, чтобы по выборочным данным определить *неизвестные значения* параметров теоретического распределения, называется задачей *точечного оценивания* параметров, а её решение основано на понятии *статистики*.

Рассмотрим выборку  $x_1, x_2, \dots, x_n$  из генеральной совокупности значений случайной величины  $\xi$ , и  $n$  независимых случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ , каждая из которых, во-первых, распределена по тому же закону, как и случайная величина  $\xi$ , а, во-вторых, выполняются соотношения

$$\xi_1(\omega) = x_1, \xi_2(\omega) = x_2, \dots, \xi_n(\omega) = x_n.$$

*Статистикой*  $\Lambda$  параметра  $\lambda$  называется произвольная случайная величина

$$\Lambda = \Lambda(\xi_1, \xi_2, \dots, \xi_n),$$

являющаяся функцией от случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ .

- Статистика  $\Lambda$  параметра  $\lambda$  называется *несмещенной*, если математическое ожидание случайной величины  $\Lambda$  равно  $\lambda$ , т.е. выполнено соотношение

$$\lambda = M[\Lambda(\xi_1, \xi_2, \dots, \xi_n)].$$

- Статистика  $\Lambda$  параметра  $\lambda$  называется *состоятельной*, если она *сходится по вероятности* к параметру  $\lambda$ , т.е. для любого положительного числа  $\varepsilon$  выполнено соотношение

$$\lim_{n \rightarrow \infty} P(|\Lambda(\xi_1, \xi_2, \dots, \xi_n) - \lambda| > \varepsilon) = 0.$$

- Статистика  $\Lambda$  параметра  $\lambda$  называется *эффективной*, если она обладает наименьшей дисперсией.

*Точечной оценкой  $\tilde{\lambda}$  параметра  $\lambda$*  называется число, полученное в результате подстановки выборочных данных в статистику  $\Lambda$  параметра  $\lambda$ , т.е.

$$\tilde{\lambda} = \Lambda(x_1, x_2, \dots, x_n).$$

Замечание. Среднее выборочное значение и *исправленная* выборочная дисперсия являются несмещенными точечными оценками математического ожидания и дисперсии теоретического распределения, соответственно. Выборочная дисперсия является *смещенной* точечной оценкой дисперсии изучаемой случайной величины.

После того, как точечная оценка параметра теоретического распределения получена, возникает задача определения *надежности* этой оценки.

Поставленная задача решается при помощи построения *интервальной оценки параметра распределения* изучаемой случайной величины.

Для построения интервальной оценки параметра зададим достаточно большую вероятность  $\gamma$ , например,  $\gamma=0.97$ , называемую *доверительной вероятностью (надежностью)*, и построим *доверительный интервал*, т.е. такой интервал, в котором с вероятностью  $\gamma$  находится параметр  $\lambda$ .

В случае, когда изучаемая случайная величина  $\xi$  распределена по нормальному закону с известной дисперсией  $D = \sigma^2$ , доверительный интервал с надежностью  $\gamma$  для оценки её математического ожидания имеет вид

$$M\xi_{\text{выб}} - u_\alpha \frac{\sigma}{\sqrt{n}} < M\xi < M\xi_{\text{выб}} + u_\alpha \frac{\sigma}{\sqrt{n}},$$

где число  $u_\alpha$  определяется по числу  $\alpha = \frac{1-\gamma}{2}$  с помощью таблицы значений функции  $u_\alpha$  (Таблица 10.2), приведенной в пункте 10 Модуля.

## 5. Распределения $\chi^2$ , Стьюдента и Фишера

Введем три важных распределения, близко связанных с нормальным распределением. Пусть  $\xi_0, \xi_1, \xi_2, \dots, \xi_n$  – независимые случайные величины, каждая из которых распределена по нормальному закону с математическим ожиданием 0 и дисперсией 1.

Распределение случайной величины, заданной с помощью формулы

$$\chi_n^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2,$$

называется *распределением  $\chi^2$  (хи-квадрат) с  $n$  степенями свободы*.

Плотность распределения  $\chi^2$  с  $n$  степенями свободы задается функцией

$$f(x, n) = \begin{cases} Ke^{-\frac{x}{2}} x^{\frac{n}{2}-1} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где  $K$  – число, удовлетворяющее условию

$$K = \frac{1}{\int_0^\infty e^{-\frac{x}{2}} x^{\frac{n}{2}-1} dx}.$$

В математическом анализе интеграл

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$$

называется Гамма-функцией, следовательно,

$$K = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}.$$

Распределение случайной величины, заданной с помощью формулы

$$t_n = \frac{\sqrt{n} \cdot \xi_0}{\sqrt{\chi_n^2}},$$

называется *распределением Стьюдента* ( $t$  – *распределением*) с  $n$  *степенями свободы*.

Распределение случайной величины, заданной с помощью формулы

$$F_{n_1, n_2} = \frac{n_2 \cdot \chi_{n_1}^2}{n_1 \cdot \chi_{n_2}^2},$$

называется *распределением Фишера* ( $F$  – *распределением*) с  $n_1$  и  $n_2$  *степенями свободы*.

Для каждого, из приведенных выше распределений, существуют специальные вероятностные таблицы. В пункте 10 Модуля содержится таблица квантилей распределения  $\chi^2$  (Таблица 10.3).

## **6. Проверка гипотезы о нормальном распределении изучаемой случайной величины. Критерий согласия $\chi^2$ Пирсона**

Завершающим этапом обработки выборочных данных обычно является выдвижение и проверка *гипотезы* (предположения) о законе распределения изучаемой случайной величины. Выдвижение такой гипотезы происходит на основе построенной гистограммы и полученных оценок основных характеристик распределения. Проверка выдвинутой гипотезы осуществляется с помощью *статистических критериев*.

Для применения статистического критерия задается какое-либо малое значение  $\beta$  вероятности *отклонить верную гипотезу*. Число  $\beta$  называется *уровнем значимости критерия*. По уровню значимости с помощью таблиц находится *критическая область*, при попадании в которую некоторой *статистической*

*характеристики (статистики)*, гипотеза принимается. В противном случае гипотеза отвергается. Статистика и таблица для определения критической области зависят от выбранного критерия.

Одним из наиболее распространенных критериев проверки гипотез о виде закона распределения изучаемой случайной величины является *критерий согласия  $\chi^2$  Пирсона*. Статистикой для этого критерия является выражение

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - m_i^T)^2}{m_i^T},$$

где  $k$  – число групп выборочных данных,  $m_i$  – частоты, полученные при группировке выборочных данных, а  $m_i^T$  – *теоретические частоты*, которые *вычисляются* в соответствии с предполагаемым законом распределения.

В частности, *при проверке гипотезы о нормальном распределении* изучаемой случайной величины теоретические частоты для *дискретного вариационного ряда* вычисляются по формуле

$$m_i^T = \frac{nh}{\sigma_X} \cdot \varphi\left(\frac{x_i - \bar{X}}{\sigma_X}\right),$$

а для *интервального вариационного ряда* – по формуле

$$m_i^T = n \cdot \left[ \Phi\left(\frac{a_i - \bar{X}}{\sigma_X}\right) - \Phi\left(\frac{a_{i-1} - \bar{X}}{\sigma_X}\right) \right],$$

где символами  $\bar{X}$  и  $\sigma_X$  обозначены соответственно математическое ожидание и дисперсия нормального закона распределения, а с функциями

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{и} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{x^2}{2}} dx$$

(см. Таблицы 10.1 и 10.4) мы уже не один раз встречались. Для определения критической области по заданному уровню значимости  $\beta$  с помощью Таблицы 10.3. (Квантили распределения  $\chi^2$ ) найдем значение  $\chi_{кр}^2$  (квантиль уровня  $1-\beta$ ).

Число степеней свободы  $\nu = k - r - 1$ , где  $k$  – число групп выборочных данных, а  $r$  – число неизвестных параметров предполагаемого закона распределения (у нормального распределения таких параметров 2: математическое ожидание и дисперсия).

Гипотеза о законе распределения принимается, если  $\chi^2 < \chi_{кр}^2$ . В противном случае она отвергается.

## 7. Группировка выборочных данных двумерной выборки.

### Корреляционная таблица

В случае, когда изучаются свойства *случайного двумерного вектора*, выборка  $XU$  из генеральной совокупности значений является *двумерной* и представляет собой набор *пар* значений  $(\tilde{x}, \tilde{y})$  двух *количественных характеристик (признаков)*  $X$  и  $Y$ .

Действуя по аналогии с тем, как это было сделано в случае одномерной выборки, построим  $k$  групп данных

$$[a_0; a_1], [a_1; a_2], [a_2; a_3], \dots, [a_{k-1}; a_k]$$

по признаку  $X$  и  $l$  групп данных

$$[b_0; b_1], [b_1; b_2], [b_2; b_3], \dots, [b_{l-1}; b_l]$$

по признаку  $Y$ .

- *Двумерным интервальным вариационным рядом* называется Таблица

	$[b_0; b_1]$	$[b_1; b_2]$	...	$[b_{l-1}; b_l]$
$[a_0; a_1]$	$m_{11}$	$m_{12}$	...	$m_{1l}$
$[a_1; a_2]$	$m_{21}$	$m_{22}$	...	$m_{2l}$
...	...	...	...	...
$[a_{k-1}; a_k]$	$m_{k1}$	$m_{k2}$	...	$m_{kl}$

В этой Таблице символами

$$m_{ij}, i = 1, 2, \dots, l, j = 1, 2, \dots, k$$

обозначены *частоты*, показывающие, сколько пар выборочных данных  $(\tilde{x}, \tilde{y})$  при группировке попали, соответственно, в прямоугольники

$$[a_{i-1}; a_i] \times [b_{j-1}; b_j].$$

- *Двумерным дискретным вариационным рядом (корреляционной таблицей)* называется следующая Таблица 1.

Т а б л и ц а 1

	$y_1$	$y_2$	...	$y_l$
$x_1$	$m_{11}$	$m_{12}$	...	$m_{1l}$
$x_2$	$m_{21}$	$m_{22}$	...	$m_{2l}$
...	...	...	...	...
$x_k$	$m_{k1}$	$m_{k2}$	...	$m_{kl}$

В Таблице 1 *середины* групп (отрезков) выборочных данных обозначены символами

$$x_i = \frac{a_{i-1} + a_i}{2}, i = 1, 2, \dots, k$$

$$y_j = \frac{b_{j-1} + b_j}{2}, j = 1, 2, \dots, l$$

- Справедливы следующие соотношения:

$$x_i = x_1 + (i-1)h_x, i = 1, 2, \dots, k,$$

$$y_j = y_1 + (j-1)h_y, j = 1, 2, \dots, l,$$

где через  $h_x$  и  $h_y$  обозначены *шаги вариации* по признакам  $X$  и  $Y$  соответственно.

- Объем выборки  $n$  удовлетворяет соотношению:

$$n = \sum_{i,j} m_{ij}.$$

## 8. Вычисление характеристик двумерной выборки.

### Прямая линия регрессии

Пусть двумерная выборка  $XU$  задана корреляционной Таблицей 1. Рассмотрим следующую Таблицу 2, которая является расширением Таблицы 1:

Т а б л и ц а 2

	$y_1$	$y_2$	...	$y_l$	$SX$
$x_1$	$m_{11}$	$m_{12}$	...	$m_{1l}$	$\sum_{j=1}^{j=l} m_{1j}$
$x_2$	$m_{21}$	$m_{22}$	...	$m_{2l}$	$\sum_{j=1}^{j=l} m_{2j}$
...	...	...	...	...	...
$x_k$	$m_{k1}$	$m_{k2}$	...	$m_{kl}$	$\sum_{j=1}^{j=l} m_{kj}$
$SU$	$\sum_{i=1}^{i=k} m_{i1}$	$\sum_{i=1}^{i=k} m_{i2}$	...	$\sum_{i=1}^{i=k} m_{il}$	

В последнем столбце  $SX$  Таблицы 2 записаны суммы элементов Таблицы 1, расположенных по строкам  $x_1, x_2, \dots, x_k$  соответственно, а в последней строке  $SU$  записаны суммы элементов Таблицы 1, расположенных по столбцам  $y_1, y_2, \dots, y_l$  соответственно.

Введем в рассмотрение одномерную выборку  $SX$ , заданную дискретным вариационным рядом

Т а б л и ц а 3

	$x_1$	$x_2$	...	$x_k$
$SX :$	$\sum_{j=1}^{j=l} m_{1j}$	$\sum_{j=1}^{j=l} m_{2j}$	...	$\sum_{j=1}^{j=l} m_{kj}$ ,

и одномерную выборку  $SU$ , заданную дискретным вариационным рядом

Т а б л и ц а 4

$SU :$	$y_1$	$y_2$	...	$y_l$
--------	-------	-------	-----	-------

	$\sum_{i=1}^{i=k} m_{i1}$	$\sum_{i=1}^{i=k} m_{i2}$	...	$\sum_{i=1}^{i=k} m_{il}$
--	---------------------------	---------------------------	-----	---------------------------

Для контроля правильности вычислений отметим, что сумма элементов, расположенных в нижней строке Таблицы 3, должна совпадать с суммой элементов, расположенных в нижней строке Таблицы 4.

- Введем следующее обозначение

$$\overline{XY} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^l m_{ij} x_i y_j .$$

- *Выборочным коэффициентом корреляции* называется число, которое определяется по формуле

$$r = \frac{\overline{XY} - \overline{SX} \cdot \overline{SY}}{\sigma_{SX} \cdot \sigma_{SY}} ,$$

где через  $\overline{SX}, \overline{SY}, \sigma_{SX}, \sigma_{SY}$  обозначены средние выборочные значения и выборочные квадратические отклонения одномерных выборок  $SX$  и  $SY$  соответственно.

Замечание. Выборочный коэффициент корреляции оценивает *тесноту линейной корреляционной зависимости* признаков  $X$  и  $Y$  измеряемой случайной двумерной величины и по модулю не превосходит 1. Чем ближе  $|r|$  к 1, тем сильнее линейная корреляционная зависимость.

- *Прямой линией регрессии* называется прямая линия на координатной плоскости  $XOY$  , имеющая уравнение

$$y = cx + d ,$$

где

$$c = r \cdot \frac{\sigma_{SX}}{\sigma_{SY}} ,$$

$$d = \frac{\sigma_{SX} \cdot \overline{SY} - r \cdot \sigma_{SX} \cdot \overline{SX}}{\sigma_{SX}} .$$

Замечание. Прямая линия регрессии *наилучшим среди всех прямых образом* приближает выборочные данные. Приведенные формулы для коэффициентов уравнения прямой линии регрессии можно получить с помощью *метода наименьших квадратов*.

• При решении задач *уравнение прямой линии регрессии* удобно использовать в следующем виде:

$$\frac{y - \overline{SY}}{\sigma_{SY}} = r \cdot \frac{x - \overline{SX}}{\sigma_{SX}}.$$

## 9. Примеры

**Пример 1.** Одномерная выборка задана интервальным вариационным рядом:

Группа	[2,3 ; 3,1]	[3,1 ; 3,9]	[3,9 ; 4,7]	[4,7 ; 5,5]	[5,5 ; 6,3]	[6,3 ; 7,1]
$m_i$	11	18	30	21	15	5

Требуется:

- а) построить гистограмму плотностей относительных частот;
- б) перейти к дискретному вариационному ряду и построить полигон относительных частот;
- в) вычислить среднее выборочное значение и среднее выборочное квадратическое отклонение;
- г) при уровне значимости  $\beta = 0,05$  проверить гипотезу о нормальном распределении измеряемой случайной величины.

**Решение.**

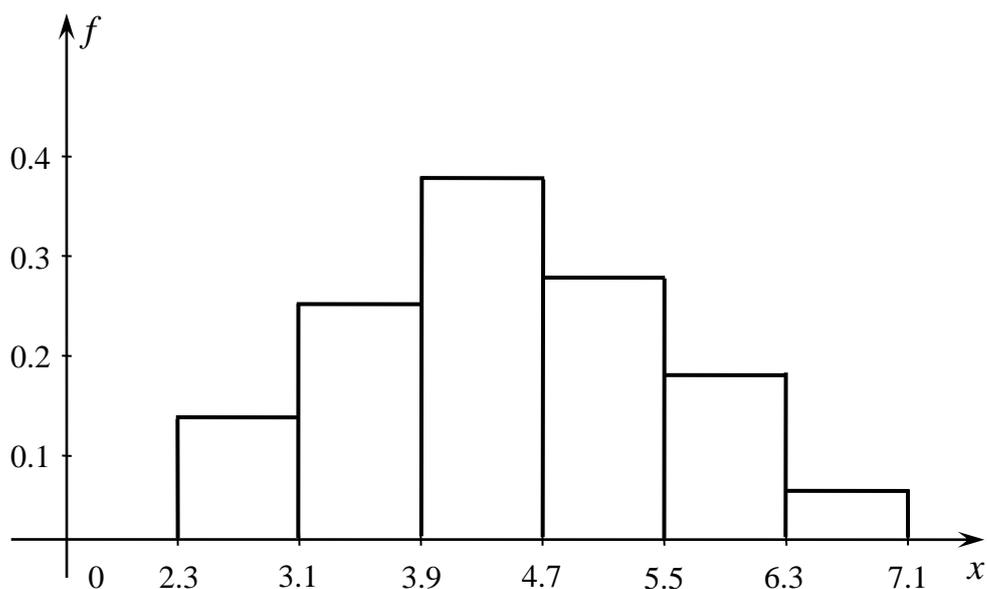
а) Поскольку объем выборки  $n = 100$ , а шаг вариации  $h = 0,8$ , то, воспользовавшись для плотностей относительных частот формулой

$$f_i = \frac{m_i}{n \cdot h} = \frac{m_i}{80},$$

преобразуем интервальный вариационный ряд к виду

<b>Группа</b>	[2,3 ; 3,1]	[3,1 ; 3,9]	[3,9 ; 4,7]	[4,7 ; 5,5]	[5,5 ; 6,3]	[6,3 ; 7,1]
$f_i$	0,14	0,23	0,38	0,26	0,19	0,06

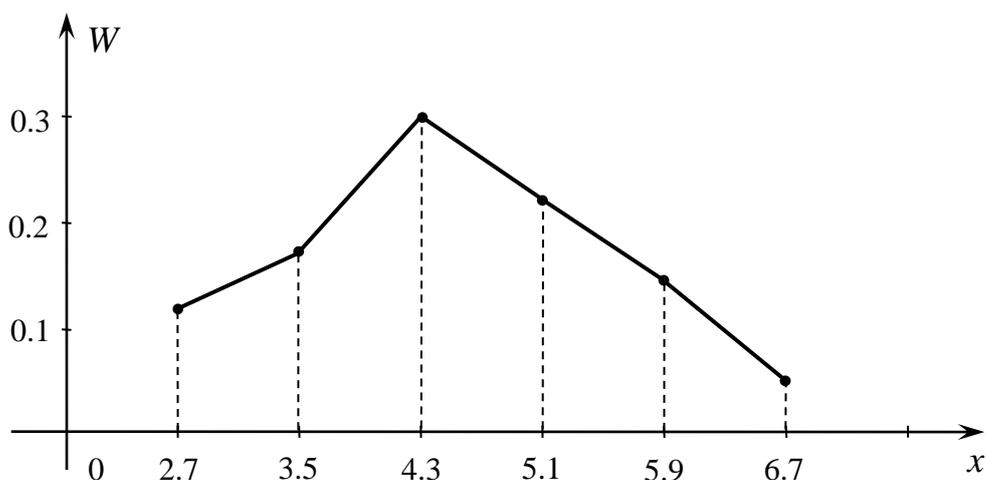
Гистограмма плотностей относительных частот  $f_i$ :



б) Заменяя в первой строке исходного интервального вариационного ряда каждый из отрезков на его середину и разделив все элементы второй строки на объем выборки  $n = 100$ , получим дискретный вариационный ряд:

$x_i$	2,7	3,5	4,3	5,1	5,9	6,7
$W_i$	0,11	0,18	0,3	0,21	0,15	0,05

Полигон относительных частот  $W_i$ :



в) Вычислим сначала среднее выборочное значение, воспользовавшись дискретным вариационным рядом:

$$\bar{X} = \sum_{i=1}^6 W_i x_i = 2,7 \cdot 0,11 + 3,5 \cdot 0,18 + 4,3 \cdot 0,3 + 5,1 \cdot 0,21 + 5,9 \cdot 0,15 + 6,7 \cdot 0,05 = 4,508.$$

Перейдем к расчету выборочной дисперсии:

$$D_X = \sum_{i=1}^6 W_i x_i^2 - (4,508)^2 = (2,7)^2 \cdot 0,11 + (3,5)^2 \cdot 0,18 + (4,3)^2 \cdot 0,3 + (5,1)^2 \cdot 0,21 + (5,9)^2 \cdot 0,15 + (6,7)^2 \cdot 0,05 - (4,508)^2 \approx 1,16$$

Теперь найдем выборочное среднее квадратическое отклонение:

$$\sigma_X = \sqrt{D_X} = \sqrt{1,16} \approx 1,08.$$

г) Для проверки гипотезы о нормальном распределении изучаемой случайной величины нам потребуется Таблица 10.4. (Значения функции  $y = \varphi(x)$ ) и Таблица 10.3. (Квантили распределения  $\chi^2$ ).

На основе формулы для теоретических частот

$$m_i^T = \frac{nh}{\sigma_X} \cdot \varphi\left(\frac{x_i - \bar{X}}{\sigma_X}\right) = \frac{100 \cdot 0,8}{1,08} \varphi\left(\frac{x_i - 4,508}{1,08}\right)$$

с помощью таблицы значений функции  $y = \varphi(x)$  составим следующую таблицу:

$i$	$x_i$	$\frac{x_i - 4,508}{1,08}$	$m_i^T$	$m_i$	$\frac{(m_i - m_i^T)^2}{m_i^T}$
1	2,7	-0,68	7,25	11	1,94
2	3,5	-0,94	19,08	18	0,06
3	4,3	-0,19	29,02	30	0,03
4	5,1	0,55	25,45	21	0,78
5	5,9	1,29	12,91	15	0,34
6	6,7	2,03	3,78	5	0,39

Просуммировав последний столбец этой таблицы, определим для рассматриваемой задачи значение  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^6 \frac{(m_i - m_i^T)^2}{m_i^T} = 1,94 + 0,06 + 0,03 + 0,78 + 0,34 + 0,39 = 3,54.$$

Теперь найдем число степеней свободы в рассматриваемой задаче:

$$v = k - r - 1 = 6 - 2 - 1 = 3.$$

Далее с помощью таблицы распределения  $\chi^2$  для уровня значимости  $\beta = 0,05$  и 3-х степеней свободы находим критическое значение:  $\chi_{кр}^2 = 7,8$ .

Поскольку  $\chi^2 = 3,54 < \chi_{кр}^2 = 7,8$ , то гипотеза о нормальном распределении изучаемой случайной величины принимается.

**Пример 2.** Двумерная выборка  $XU$  задана следующей корреляционной таблицей:

$x_i \backslash y_j$	5,3	7,4	9,5	11,6	13,7
2,7	6	5	-	-	-
3,5	-	4	7	7	-
4,3	-	9	12	9	-
5,1	-	-	15	6	-
5,9	-	-	2	7	6
6,7	-	-	-	3	2

Построить уравнение прямой линии регрессии  $y = cx + d$ .

**Решение.** Найдем сначала объем выборки:

$$n = 6 + 5 + 4 + 7 + 7 + 9 + 12 + 9 + 15 + 6 + 2 + 7 + 6 + 3 + 2 = 100.$$

Теперь вычислим значение выражения  $\overline{XU}$ :

$$\overline{XY} = \frac{1}{100} \sum_{i=1}^6 \sum_{j=1}^5 m_{ij} x_i y_j = 0,01 \cdot (6 \cdot 2,7 \cdot 5,3 + 5 \cdot 2,7 \cdot 7,4 + 4 \cdot 3,5 \cdot 7,4 + 7 \cdot 3,5 \cdot 9,5 + 7 \cdot 3,5 \cdot 11,6 + 9 \cdot 4,3 \cdot 7,4 + 12 \cdot 4,3 \cdot 9,5 + 9 \cdot 4,3 \cdot 11,6 + 15 \cdot 5,1 \cdot 9,5 + 6 \cdot 5,1 \cdot 11,6 + 2 \cdot 5,9 \cdot 9,5 + 7 \cdot 5,9 \cdot 11,6 + 6 \cdot 5,9 \cdot 13,7 + 3 \cdot 6,7 \cdot 11,6 + 2 \cdot 6,7 \cdot 13,7) = 0,01 \cdot 4606,42 \approx 46,06.$$

Просуммировав в корреляционной таблице частоты по строкам, составим дискретный вариационный ряд для одномерной выборки  $SX$  :

$SX$ :	2,7	3,5	4,3	5,1	5,9	6,7
	11	18	30	21	15	5

Заметим, что дискретный вариационный ряд  $SX$  полностью совпадает с дискретным вариационным рядом из примера 1, все характеристики которого мы уже вычислили при решении примера 1. Следовательно,

$$\overline{SX} = 4,508 \approx 4,51,$$

$$D_{SX} = 1,16,$$

$$\sigma_{SX} = 1,08.$$

Просуммировав в корреляционной таблице частоты по столбцам, составим дискретный вариационный ряд для одномерной выборки  $SY$  :

$SY$ :	5,3	7,4	9,5	11,6	13,7
	6	18	36	32	8

Вычислим среднее выборочное значение для выборки  $SY$  :

$$\overline{SY} = \frac{1}{100} (5,3 \cdot 6 + 7,4 \cdot 18 + 9,5 \cdot 36 + 11,6 \cdot 32 + 13,7 \cdot 8) = 0,01 \cdot 987,8 \approx 9,88.$$

Вычислим выборочную дисперсию для выборки  $SY$  :

$$D_{SY} = \frac{1}{100} \left[ (5,3)^2 \cdot 6 + (7,4)^2 \cdot 18 + (9,5)^2 \cdot 36 + (11,6)^2 \cdot 32 + (13,7)^2 \cdot 8 \right] - (9,878)^2 \approx 102,11 - 97,57 = 4,54.$$

Найдем выборочное среднее квадратическое отклонение  $\sigma_{SY}$  :

$$\sigma_{SY} = \sqrt{D_{SY}} = \sqrt{4,54} = 2,13.$$

Теперь найдем выборочный коэффициент корреляции  $r$  :

$$r = \frac{\overline{XY} - \overline{SX} \cdot \overline{SY}}{\sigma_{SX} \cdot \sigma_{SY}} = \frac{46,06 - 4,51 \cdot 9,88}{1,08 \cdot 2,13} = 0,65.$$

Теперь можно составить уравнение прямой линии регрессии. С этой целью подставим в соотношение

$$\frac{y - \overline{SY}}{\sigma_{SY}} = r \cdot \frac{x - \overline{SX}}{\sigma_{SX}}$$

полученные в ходе решения примера числовые данные:

$$\frac{y - 9,88}{2,13} = 0,65 \cdot \frac{x - 4,51}{1,08}.$$

Следовательно,  $y = 1,28x + 4,11$ .

*Замечание.* Вычисления, подобные проведенным при решении примеров 1 и 2, очень удобно и просто осуществлять, используя, например, программный пакет Excel.

## 10. Статистические таблицы

**Т а б л и ц а 10.1.** Значения функции Лапласа  $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-t^2/2} dt$

x	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4985	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

**Т а б л и ц а 10.2.** Значения функции  $u_\alpha$ , определяемой равенством

$$\alpha = \frac{1}{\sqrt{2\pi}} \cdot \int_{u_\alpha}^{+\infty} e^{-t^2/2} dt$$

$\alpha$	0,001	0,005	0,010	0,015	0,020	0,025	0,030	0,035	0,040	0,045	0,050
----------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

$u_{\alpha}$	3,0902	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954	1,6449
--------------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

**Т а б л и ц а 10.3. Квантили распределения  $\chi^2$  с  $n$  степенями свободы**

$n$	Уровень квантиля						
	0,01	0,05	0,10	0,50	0,90	0,95	0,99
1	0,000157	0,00393	0,0158	0,455	2,71	3,84	6,64
2	0,0201	0,103	0,211	1,39	4,61	5,99	9,21
3	0,115	0,352	0,584	2,37	6,25	7,81	11,3
4	0,297	0,711	1,06	3,36	7,78	9,49	13,3
5	0,554	1,15	1,61	4,35	9,24	11,1	15,1
6	0,872	1,64	2,20	5,35	10,6	12,6	16,8
7	1,24	2,17	2,83	6,35	12,0	14,1	18,5
8	1,65	2,73	3,49	7,34	13,4	15,5	20,1
9	2,09	3,33	4,17	8,34	14,7	16,9	21,7
10	2,56	3,94	4,87	9,34	16,0	18,3	23,2
11	3,05	4,57	5,58	10,3	17,3	19,7	24,7
12	3,57	5,23	6,30	11,3	18,5	21,0	26,2
13	4,11	5,89	7,04	12,3	19,8	22,4	27,7
14	4,66	6,57	7,79	13,3	21,1	23,7	29,1
15	5,23	7,26	8,55	14,3	22,3	25,0	30,6
16	5,81	7,96	9,31	15,3	23,5	26,3	32,0
17	6,41	8,67	10,1	16,3	24,8	27,6	33,4
18	7,01	9,39	10,9	17,3	26,0	28,9	34,8
19	7,63	10,1	11,7	18,3	27,2	30,1	36,2
20	8,26	10,9	12,4	19,3	28,4	31,4	37,6
21	8,90	11,6	13,2	20,3	29,6	32,7	38,9
22	9,54	12,3	14,0	21,3	30,8	33,9	40,3
23	10,2	13,1	14,8	22,3	32,0	35,2	41,6
24	10,9	13,8	15,7	23,3	33,2	36,4	43,0
25	11,5	14,6	16,5	24,3	34,4	37,7	44,3

**Т а б л и ц а 10.4. Значения функции  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$**

	0	1	2	3	4	5	6	7	8	9
0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3322	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518

<b>1,4</b>	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
<b>1,5</b>	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
<b>1,6</b>	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,957
<b>1,7</b>	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
<b>1,8</b>	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
<b>1,9</b>	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
<b>2</b>	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
<b>2,1</b>	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
<b>2,2</b>	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
<b>2,3</b>	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
<b>2,4</b>	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
<b>2,5</b>	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
<b>2,6</b>	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
<b>2,7</b>	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
<b>2,8</b>	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
<b>2,9</b>	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
<b>3</b>	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
<b>3,1</b>	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
<b>3,2</b>	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
<b>3,3</b>	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
<b>3,4</b>	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
<b>3,5</b>	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
<b>3,6</b>	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
<b>3,7</b>	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
<b>3,8</b>	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
<b>3,9</b>	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
<b>4</b>	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

## ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Что называется генеральной совокупностью и выборкой?
2. Как по выборке построить вариационный ряд, интервальный вариационный ряд, дискретный вариационный ряд?
3. Что такое многоугольник выборочного распределения?
4. Что такое гистограмма и полигон относительных частот?
5. Какие выборочные характеристики Вы знаете?
6. Что называется квантилью выборочного распределения?
7. Что называется статистикой параметра случайной величины?
8. Что называется точечной оценкой параметра случайной величины?
9. Что называется интервальной оценкой параметра случайной величины?
10. Какая статистика называется несмещенной?
11. Какая статистика называется состоятельной?
12. Какая статистика называется эффективной?
13. Что называется доверительным интервалом?
14. Как построить доверительный интервал для математического ожидания нормального распределения, если известна дисперсия?
15. В чем состоит критерий согласия  $\chi^2$  Пирсона?
16. Какое распределение называется распределением  $\chi^2$ ?
17. Какое распределение называется распределением Стьюдента?
18. Какое распределение называется распределением Фишера?
19. Что называется двумерной выборкой?
20. Что называется корреляционной таблицей?
21. Что называется коэффициентом корреляции?
22. В чем состоит корреляционная (регрессионная) зависимость?
23. В чем состоит схема построения уравнения прямой линии регрессии для двумерной выборки?

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Выборка из генеральной совокупности значений случайной величины задана дискретным вариационным рядом

$x_i$	0,1	0,4	0,7	1	1,3	1,6	1,9
$m_i$	5	17	24	26	15	10	3

Найти:

1. Моду распределения;
2. Эмпирическую функцию распределения и построить ее график;
3. Медиану распределения;
4. Нижнюю и верхнюю квартили распределения;
5. Квантиль распределения уровня 0,9;
6. Выборочное среднее значение;
7. Выборочную дисперсию;
8. Исправленную выборочную дисперсию;
9. Выборочное среднее квадратическое отклонение;
10. Эксцесс распределения;
11. Асимметрию распределения.
12. При уровне значимости  $\beta = 0,05$  проверить гипотезу о нормальном распределении изучаемой случайной величины.

Двумерная выборка  $XU$  задана корреляционной таблицей:

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
$x_1$	1	2	1	0	0	0
$x_2$	0	3	5	2	0	0
$x_3$	0	2	5	12	0	0
$x_4$	0	0	1	7	3	0
$x_5$	0	0	0	2	3	1

где  $x_i = 1,2 + 0,5 \cdot (i - 1)$ ,  $y_j = 0,5 + 0,9 \cdot (j - 1)$ .

13. Найти выборочный коэффициент корреляции;

14. Построить уравнение прямой линии регрессии вида  $y = cx + d$ .

## ЛИТЕРАТУРА

### Основная:

1. Бочаров П.П., Печинкин А.В. Теория вероятностей и математическая статистика: Учебное пособие. – М.: Физматлит, 2005.
2. Кибзун А.И., Горяинова Е.Р., Наумов А.В., Сиротин А.Н. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами: Учебное пособие. – М.: Физматлит, 2002.
3. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика: Учебник. – М.: ИНФРА-М, 2002.
4. Сборник задач по теории вероятностей, математической статистике и теории случайных функций./Под ред. Свешникова А.А. – СПб.: Издательство «Лань», 2006.

### Дополнительная:

5. Бородин А.Н. Элементарный курс теории вероятностей и математической статистики. – СПб.: Издательство «Лань», 2006.
6. Кремер Н.Ш. Теория вероятностей и математическая статистика. Учебник. – М.: ЮНИТИ-ДАНА, 2001.
7. Общий курс высшей математики для экономистов. Учебник. / Под ред. В.И.Ермакова. – М.: ИНФРА-М, 2001.
8. Сборник задач по высшей математике для экономистов: Учебное пособие. / Под ред. В.И.Ермакова. – М.: ИНФРА-М, 2004.
9. Чистяков В.П. Курс теории вероятностей. – СПб.: Издательство «Лань», 2006.