1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

Предположим, что некоторая величина X может принимать значения x_1, x_2, …, x_n в зависимости от некоторых случайных факторов, так что этим значениям можно сопоставить вероятности p_1, p_2, …, p_n. Такая величина называется **случайной величиной** (СВ). Свои значения СВ принимает случайным образом.

Рассмотрим два основных типа случайных величин: дискретные и непрерывные.

Случайная величина называется *дискретной*, если она принимает отдельные (изолированные) значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или счетным.

Случайная величина называется *непрерывной*, если она принимает возможные значения, которые сплошь заполняют некоторый интервал.

Значения функции $P(X = x_i) = p_i$, $i = 1, 2, \ldots$, называются рядом распределения вероятностей дискретной случайной величины X.

Законом распределения дискретной случайной величины называется таблица, в верхней строке которой указаны возможные (различные) значения случайной величины X (в порядке возрастания), а в нижней строке под каждым значением x_i – соответствующая вероятность $p_i = P(X = x_i)$, причем $p_1 + p_2 + \ldots + p_n = 1$.

Все свои значения случайная величина X принимает с некоторыми вероятностями, используя которые образуют закон распределения для СВ X:

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

Графически ряд распределения представляется в виде **полигона распределения**, причем по оси ОX откладывают отдельные значения величины X, а по оси ОY – соответствующие им вероятности. Полученные, таким образом точки с координатами (x_i, p_i), где $i = 1,\ 2,\ \ldots, n$, соединяют прямыми (рис.1).

Интегральной функцией распределения случайной величины X называется функция $F(x)$, выражающая вероятность того, что X примет значение меньшее, чем заданное x:

$$F(x) = P(X < x).$$

Следствие: $P(a \le x < b) = F(b) - F(a)$.

Функция $F(x)$ – неубывающая функция ($F(x_2) \ge F(x_1)$, если $x_2 \ge x_1$), кроме того, $F(-\infty) = 0$, $F(+\infty) = 1$. Для непрерывной случайной величины функция распределения непрерывна и существует производная $f(x) = F'(x)$, которая называется *плотностью распределения вероятностей или дифференциальной функцией распределения*. Плотность любой случайной величины неотрицательна и обладает свойствами:

1) $f(x) \ge 0$;

2) $\int\limits_{-\infty}^{\infty} f(x)\,dx = 1$.

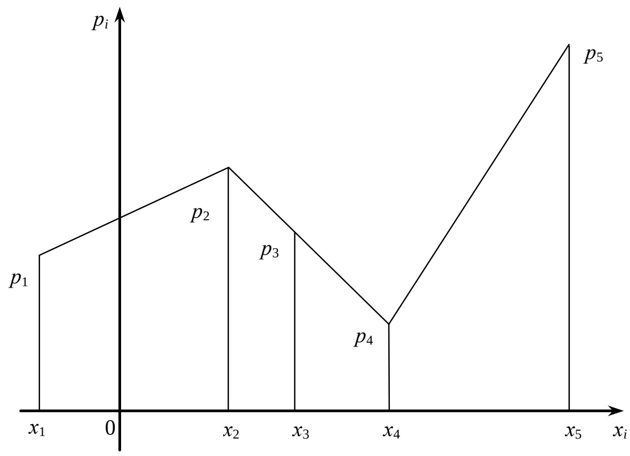

Рис.1. Полигон распределения

График дифференциальной функции распределения вероятностей называют *кривой распределения*. Интегральная функция распределения $F(x)$ выражается через функцию плотности $f(x)$ следующим образом:

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

Для непрерывной случайной величины вероятность принятия некоторого конкретного значения $P(x = a) = 0$. Тогда

$$P(a \le x \le b) = P(a < x < b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a).$$

Если непрерывная случайная величина $X \in [a, b]$, то $F(a) = 0$ и $F(b) = 1$ (условия нормировки). Для дискретной случайной величины функция распределения кусочно-постоянна и ступеньки (величины скачков) равны накопленным вероятностям p_1, $p_1 + p_2$, $p_1 + + p_2 + p_3$, … (рис.2).

Пример 1. Непрерывная случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0, & x < 1; \\ ax^2 + b, & 1 \le x \le 2; \\ 1, & x > 2. \end{cases}$$

Найти параметры a и b, плотность и вероятность попадания случайной величины в интервал (1, 5; 3).

Решение. По условиям нормировки имеем $F(1) = a+b = 0$ и $F(2) = 4a + b = 1$, т.е. $a = 1/3$ и

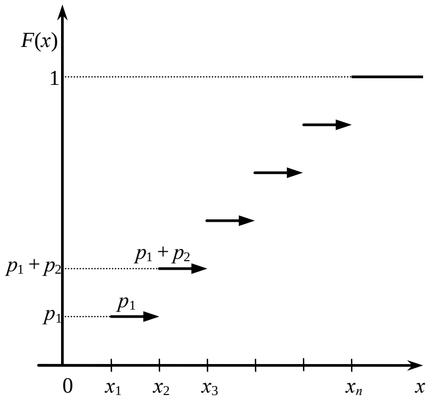

Рис.2. Интегральная функция распределения дискретной случайной величины

b = –1/3. Тогда функция распределения и ее плотность

$$f(x) = F'(x) = \begin{cases} 0, & x < 1; \\ 2x/3, & 1 \le x \le 2; \\ 0, & x > 2. \end{cases}$$

$$F(x) = \begin{cases} 0, & x < 1; \\ \dfrac{1}{3}x^2 - \dfrac{1}{3}, & 1 \le x \le 2; \\ 1, & x > 2. \end{cases}$$

Следовательно, получим

$P(1{,}5 < x < 3) = F(3) - F(1{,}5) = 1 - 1/3(9/4 - 1) = 7/12$.

Математическое ожидание или среднее значение случайной величины X есть величина, вычисляемая для дискретной и непрерывной случайных величин по формулам, соответственно:

$$\overline{X} = M(X) = \sum_{i=1}^{\infty} x_i p_i\,; \quad \overline{X} = M(X) = \int_{-\infty}^{\infty} x f(x)dx\,.$$

Дисперсия случайной величины X – это математическое ожидание квадрата ее отклонения от ее математического ожидания:

$$D(X) = M((X - \overline{X})^2) = M(X^2) - (M(X))^2 = M(X^2) - \overline{X}^2\,.$$

Дисперсия для дискретной и непрерывной случайной величин, соответственно

$$D(X) = \sum_{i=1}^{\infty} (x_i - M(X))^2 p_i = \sum_{i=1}^{\infty} x_i^2 p_i - (M(X))^2\,;$$

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - (M(x))^2\,.$$

Среднее квадратичное отклонение случайной величины X обозначим $\sigma(x)$, причем $\sigma(x) = \sqrt{D(X)}.$

Пример 2. Пусть заданное распределение имеет вид

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

Найти математическое ожидание, дисперсию и среднее квадратичное отклонение.

Решение. Математическое ожидание

$$M(X) = \sum_{i=1}^{4} x_i p_i = 1 \cdot 0{,}2 + 3 \cdot 0{,}3 + 4 \cdot 0{,}4 + 6 \cdot 0{,}1 =$$

$$= 0{,}2 + 0{,}9 + 1{,}6 + 0{,}6 = 3{,}3.$$

Дисперсия и среднее квадратичное отклонение

$$D(X) = \sum_{i=1}^{4} x_i^2 p_i - (M(X))^2 = 1\cdot 0{,}2+9\cdot 0{,}3+16\cdot 0{,}4+36\cdot 0{,}1 -$$

$$- 3{,}3^2 = 0{,}2+2{,}7+6{,}4+3{,}6-3{,}3^2 = 2{,}01;$$

$$\sigma(x) = \sqrt{D(X)} = \sqrt{2{,}01} \approx 1{,}42.$$

Пример 3. Для случайной величины X из примера 10 найти среднее квадратичное отклонение.

Решение. Математическое ожидание (см. пример 10)

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{1}^{2} x \frac{2x}{3} dx = \frac{2}{3} \frac{x^3}{3} \Big|_1^2 = \frac{14}{9}.$$

Дисперсия и среднее квадратичное отклонение

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (M(X))^2 =$$

$$=\int_1^2 x^2\frac{2x}{3}dx-\left(\frac{14}{9}\right)^2=\frac{2}{3}\frac{x^4}{4}\bigg|_1^2-\frac{196}{81}=\frac{5}{2}-\frac{196}{81}=\frac{13}{162};$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\frac{13}{162}}\approx 0{,}28.$$

Дискретная случайная величина X называется распределенной по **биномиальному закону**, если ее возможные значения равны 0, 1, 2, …, n, а вероятность того, что $X = m$, выражается формулой Бернулли:

$$P(X=m)=P_n(m)=C_n^m p^m q^{n-m},$$

где случайная величина X – число появлений некоторого события A в n испытаниях; p – вероятность появления события A в каждом испытании, не изменяющаяся от испытания к испытанию, $0 < p < 1$; q – вероятность отсутствия события A в каждом испытании, $q = 1 - p$.

Математическое ожидание и дисперсия случайной величины X, распределенной по биномиальному закону,

$$M(X) = np;\ \ D(X) = npq.$$

Дискретная случайная величина X называется распределенной по **закону Пуассона**, если ее возможные значения равны 0, 1, 2, …, m,…, а вероятность того, что $X = m$, выражается формулой

$$P(X=m)=P_n(m)=\frac{a^m}{m!}e^{-a},$$

где a – параметр закона Пуассона, $a > 0$.

Как было показано ранее, по этой формуле вычисляются вероятности редких событий, т.е. вероятность появления события A m раз при большом числе испытаний n, в каждом из которых вероятность p появления события A мала ($p \le 0{,}1$), однако, произведение $np = a$ постоянно. Математическое ожидание и дисперсия случайной величины X, распределенной по закону Пуассона,

$$M(X) = a;\ D(X) = a.$$

Непрерывная случайная величина X называется **равномерно распределенной** на отрезке $[a, b]$, если ее плотность распределения вероятностей постоянна (т.е. все значения на отрезке случайной величины X равновозможны):

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b]; \\ 0, & x \notin [a,b]. \end{cases}$$

Математическое ожидание и дисперсия СВ, равномерно распределенной на (a,b),

$$M(X) = \frac{a+b}{2};\ D(X) = \frac{(b-a)^2}{12}.$$

Непрерывная случайная величина X называется **распределенной по нормальному закону,** если ее плотность распределения вероятностей равна:

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

где a – математическое ожидание; σ^2 – дисперсия; σ – среднее квадратичное отклонение случайной величины X (a, σ – параметры нормального распределения).

Вероятность попадания случайной величины X, распределенной по нормальному закону, в интервал (α, β)

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

где $\Phi(x)$ – функция Лапласа.

Вероятность того, что абсолютная величина отклонения меньше постоянного числа ε, $P(|X-a| < \varepsilon) = 2\Phi(\varepsilon/\sigma)$.

<u>Следствие.</u> Правило «трех сигм»: $P(|X-a| < 3\sigma) = 0{,}9973 \approx 1$.

Пример 4. Случайная величина X распределена по нормальному закону с параметрами $a = 2$, $\sigma = 5$.

Найти вероятность того, что случайная величина X примет значение в интервале (1, 4).

Решение. По условию $\alpha = 1$, $\beta = 4$, $a = 2$, $\sigma = 5$. Тогда

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

$$= \Phi(0{,}4) + \Phi(0{,}2) = 0{,}1554 + 0{,}0793 = 0{,}2347.$$

2. СИСТЕМА ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН И РЕГРЕССИЯ

Система двух случайных величин – совокупность двух случайных величин (X,Y), которые рассматриваются одновременно.

Измерения обычно осуществляются попарно, а полученные значения случайных величин X и Y в определенном смысле взаимосвязаны.

Закон распределения двумерной случайной величины дискретного типа представляет собой перечень значений этой величины и их вероятностей, указанных в специальной таблице. В табл.1 представлены возможные значения (x_i, y_j) и их совместные вероятности:

$$p_{ij} = P(X = x_i,\ Y = y_j)\ ,\quad i = 1, 2, \ldots, n,\ \ j = 1, 2, \ldots, m\ .$$

Зная закон распределения дискретной двумерной случайной величины (X, Y), можно найти закон распределения каждой случайной величины X и Y:

$$p_{x_i} = P(X = x_i),\ p_{y_j} = P(Y = y_j);$$

$$p_{x_i} = \sum_{j=1}^{m} p_{ij}\,,\ i = 1,2,\ldots,n,\ \ p_{y_j} = \sum_{i=1}^{n} p_{ij}\,,\ j = 1,2,\ldots,m\,.$$

Таблица 1

Закон распределения двумерной случайной величины

Y	X						
	x_1	x_2	…	x_i	…	x_n	p_{yj}
y_1	p_{11}	p_{21}	…	p_{i1}	…	p_{n1}	p_{y1}
y_2	p_{12}	p_{22}	…	p_{i2}	…	p_{n2}	p_{y2}
…	…	…	…	…	…	…	…
y_j	p_{1j}	p_{2j}	…	p_{ij}	…	p_{nj}	p_{yj}
…	…	…	…	…	…	…	…
y_m	p_{1m}	p_{2m}	…	p_{im}	…	p_{nm}	p_{ym}
p_{xi}	p_{x1}	p_{x2}	…	p_{xi}	…	p_{xn}	

Интегральная функция распределения двумерной случайной величины (X, Y) есть вероятность совместного выполнения неравенств $X < x$ и $Y < y$, т.е.

$$F(x, y) = P(X < x, Y < y).$$

Двумерная случайная величина непрерывного типа может быть задана интегральной или дифференциальной функцией распределения. Если интегральная функция распределения всюду непрерывна и имеет непрерывную смешанную частную производную второго порядка, то дифференциальная функция распределения системы двух случайных величин (X, Y) определяется по формуле

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Плотность распределения отдельных случайных величин, входящих в систему, выражается через плотность системы случайных величин следующим образом:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)dy; \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Условный закон распределения случайной величины, входящей в систему, есть закон ее распределения, полученный в предположении, что другая случайная величина приняла определенное значение. Для системы случайных величин дискретного типа условные законы распределения имеют вид

$$p(x_i / y_j) = p_{ij} / p_{y_j}; \quad p(y_j / x_i) = p_{ij} / p_{x_i}.$$

Условные математические ожидания (условные средние) дискретных случайных величин

$$\overline{y}_{x_i} = M(Y / x_i) = \sum_{j=1}^{m} y_j p(y_j / x_i), \; i = 1, 2, \dots n,$$

$$\overline{x}_{y_j} = M(X / y_j) = \sum_{i=1}^{n} x_i p(x_i / y_j), \; j = 1, 2, \dots, m.$$

Условные распределения показывают, что одна СВ реагирует на изменение другой изменением своего закона распределения. Такая общая зависимость называется *стохастической* (*вероятностной*) и достаточно сложна для изучения. Однако зависимость условного среднего одной СВ от значений другой является функцией, которая называется *регрессией*: $\bar{y}_x = \varphi(x)$ – регрессия Y на X, $\bar{x}_y = \psi(y)$ – регрессия X на Y. Но и функции регрессии в общем случае достаточно сложны, поэтому используют различные их приближения, например линейной функцией (наилучшей в смысле наименьшего значения среднего квадрата отклонения). Это значит, что для регрессии Y на X функция $\overline{y}_x$ приближается линейной функцией $y = ax + b$:

$$M\left((\bar{y}_x - y)^2\right) \xrightarrow[a,b]{} \min.$$

Уравнения таких наилучших линейных регрессий
для регрессии Y на X

$$y - \bar{y} = r_{xy} \frac{\sigma_y}{\sigma_x}(x - \bar{x});$$

для регрессии X на Y

$$x-\bar{x}=r_{xy}\frac{\sigma_x}{\sigma_y}(y-\bar{y}),$$

где $\bar{y}=M(Y)$, $\bar{x}=M(X)$, $\sigma_x=\sigma(X)$, $\sigma_y=\sigma(Y)$.

Коэффициент корреляции

$$r_{xy}=\frac{M(XY)-\bar{x}\bar{y}}{\sigma_x\sigma_y}$$

характеризует близость (или тесноту) связи между случайными величинами к линейной.

Отметим, что всегда $|r_{xy}|\le 1$. Если $r_{xy}=0$, то СВ называются *некоррелированными* и в этом случае их условные средние значения являются постоянными, т.е. не зависят от значений другой СВ, что характеризует их слабую взаимозависимость. Если $r_{xy}\approx 0$ (угол между прямыми наилучших линейных регрессий близок к прямому), связь между случайными величинами достаточно слабая и нелинейная. Если $|r_{xy}|\approx 1$ (угол близок к нулю), то связь сильная и близка к линейной. В случае промежуточного значения r_{xy} (и угла) связь достаточно сильна и существенно нелинейная.

3. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Пусть для изучаемой случайной величины X получен ряд ее значений $x_1, x_2, \ldots, x_n$, который называют выборкой объема n из множества всех возможных значений X (генеральной совокупности). Эти значения x_i являются случайными величинами, так как меняются от выборки к выборке.

Важно, чтобы опыты для получения достоверных и правильно представляющих (репрезентативных) генеральную совокупность результатов проводились в одинаковых условиях и независимо друг от друга. Значит, случайные величины x_i будут независимы и одинаково распределены. Согласно **центральной предельной теореме**

(ЦПТ) распределение среднего значения $\bar{x}_в$ будет приближаться к нормальному распределению при $n \to \infty$.

Если число n невелико ($n \le 25$), то полученные значения можно упорядочить по величине и указать число повторений (частоту) каждого из значений: $x_1 < x_2 < \ldots < x_k$ с частотами $m_1, m_2, \ldots, m_k$, где $m_1 + m_2 + \ldots + m_k = n$ (вариационный ряд). При большом числе наблюдений вводятся интервалы группировки Δ_i, которые охватывают все значения вариационного ряда (причем, первое и последнее значения – с запасом). Интервалы выбираются равными, а их концы возможно более простыми (в целых точках или в целых десятках: 10, 20,…). Обычно удобно ввести не более двух-трех десятков таких интервалов. Например, если $x_1 = 0, \ldots, x_k = 20$, то вводим промежутки $[-10, 0]$, $[0, 10]$, $[10, 20]$, $[20, 30]$. Каждому интервалу сопоставляется его середина x_i^* и частота m_i^*, равная сумме частот значений ряда, попадающих в этот интервал. При этом для значения, попавшего на границу двух интервалов, частота делится пополам между ними. Таким образом, составляется сгруппированный вариационный ряд, для которого определяются относительные частоты (или эмпирические вероятности) $p_i^* = m_i^* / n$ и эмпирические плотности $f_i^* = p_i^* / \Delta_i$. По этим данным строятся полигон эмпирического распределения (см. рис.1), гистограмма $f_n^*(x)$ (рис.3) и эмпирическая функция распределения $F_n^*(x)$ по накопленной эмпирической вероятности (рис.4).

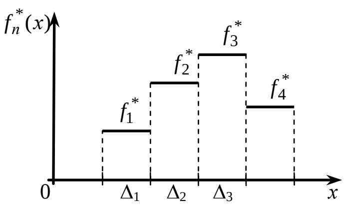

Рис.3. Гистограмма

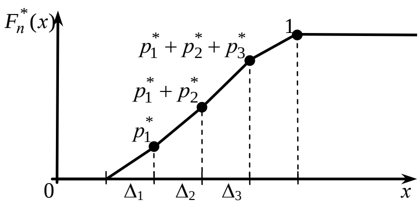

Рис.4. Эмпирическая функция распределения

По теореме Бернулли (или закону больших чисел) эмпирическая вероятность приближается к теоретической вероятности при $n \to \infty$, что справедливо и для значений эмпирической функции распределения и для гистограммы на интервалах группировки.

Пример 5. Выборка задана вариационным рядом:

$x_k =$	0	3	5	9	11	13	17	19	20	24	26	28	31	34	39	40	42
$m_k =$	4	2	5	3	6	3	3	4	3	2	1	3	3	2	1	3	2

($k = 1,2,\ldots,17$).

Произвести группировку значений и по сгруппированному вариационному ряду построить эмпирическую функцию распределения и гистограмму плотности.

Решение. Вводим интервалы группировки:

$\Delta_1 = [-10;\, 0]$ (чтобы первое значение $x_1 = 0$ включалось с запасом),

$\Delta_2 = [0;\, 10]$, $\Delta_3 = [10;\, 20]$, $\Delta_4 = [20;\, 30]$, $\Delta_5 = [30;\, 40]$, $\Delta_6 = [40;\, 50]$

(последнее значение $x_{17} = 42$ включается с запасом).

Для сгруппированного вариационного ряда значения равны серединам интервалов: $x_1^* = -5$, $x_2^* = 5$, $x_3^* = 15$, $x_4^* = 25$, $x_5^* = 35$, $x_6^* = 45$, а частоты m_i^* для этих значений (т.е. для интервалов Δ_i) получаем, складывая частоты значений x_k, попавшие в соответствующий интервал Δ_i группировки, причем для значения x_k, попавшего на границу двух интервалов, частота m_k делится между этими интервалами поровну:

$$m_1^* = \frac{m_1}{2} = \frac{4}{2} = 2;$$

$$m_2^* = \frac{m_1}{2} + m_2 + m_3 + m_4 = 2 + 2 + 5 + 3 = 12;$$

$$m_3^* = m_5 + m_6 + m_7 + m_8 + \frac{m_9}{2} = 6 + 3 + 3 + 4 + \frac{3}{2} = 17{,}5;$$

$$m_4^* = \frac{m_9}{2} + m_{10} + m_{11} + m_{12} = \frac{3}{2} + 2 + 1 + 3 = 7{,}5;$$

$$m_5^* = m_{13} + m_{14} + m_{15} + \frac{m_{16}}{2} = 3 + 2 + 1 + \frac{3}{2} = 7{,}5;$$

$$m_6^* = \frac{m_{16}}{2} + m_{17} = \frac{3}{2} + 2 = 3{,}5.$$

Объем выборки $n = \sum_{k=1}^{17} m_k = m_i^* = 50.$

Эмпирические вероятности равны:

$$p_1^* = \frac{m_1^*}{n} = \frac{2}{50} = 0{,}04;$$

$$p_2^* = \frac{m_2^*}{n} = \frac{12}{50} = 0{,}24;$$

$$p_3^* = \frac{m_3^*}{n} = \frac{17{,}5}{50} = 0{,}35;$$

$$p_4^* = \frac{m_4^*}{n} = \frac{7{,}5}{50} = 0{,}15;$$

$$p_5^* = \frac{m_5^*}{n} = \frac{7{,}5}{50} = 0{,}15;$$

$$p_6^* = \frac{m_6^*}{n} = \frac{3{,}5}{50} = 0{,}07;$$

Отметим, что $\sum_{i=1}^{6} p_i^* = 1.$ Если же значения p_i^* вычисляются приближенно, то при подсчете следует взять запасные знаки за запятой и, округляя p_i^* (до 0,01; затем до 0,001 и т.д.) найти те значения, при которых равенство $\sum_{i=1}^{6} p_i^* = 1$ выполнится. Именно эти приближенные значения p_i^* следует использовать в дальнейших вычислениях.

Накопленные вероятности:

за Δ_1: $p_1^* = 0{,}04$;

за Δ_1 и Δ_2: $p_1^* + p_2^* = 0{,}04 + 0{,}24 = 0{,}28$;

за $\Delta_1, \Delta_2, \Delta_3$: $p_1^* + p_2^* + p_3^* = 0{,}28 + 0{,}35 = 0{,}63$;

за $\Delta_1, \Delta_2, \Delta_3, \Delta_4$: $p_1^* + p_2^* + p_3^* + p_4^* = 0{,}63 + 0{,}15 = 0{,}78$;

за $\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5$: $p_1^* + p_2^* + p_3^* + p_4^* + p_5^* = 0{,}78 + 0{,}15 = 0{,}93$;

за все интервалы $\Delta_1 \div \Delta_6$: $0{,}93 + 0{,}07 = 1$.

Для построения графика эмпирической функции распределения, найденные значения накопленной вероятности следует отложить по вертикальной оси в правых концах соответствующих по номерам интервалов $\Delta_1 \div \Delta_6$. Полученные точки необходимо соединить отрезками, причем слева от Δ_1 функция $F_n^*(x) = 0$, а справа от Δ_6 функция $F_n^*(x) = 1$ (рис.5).

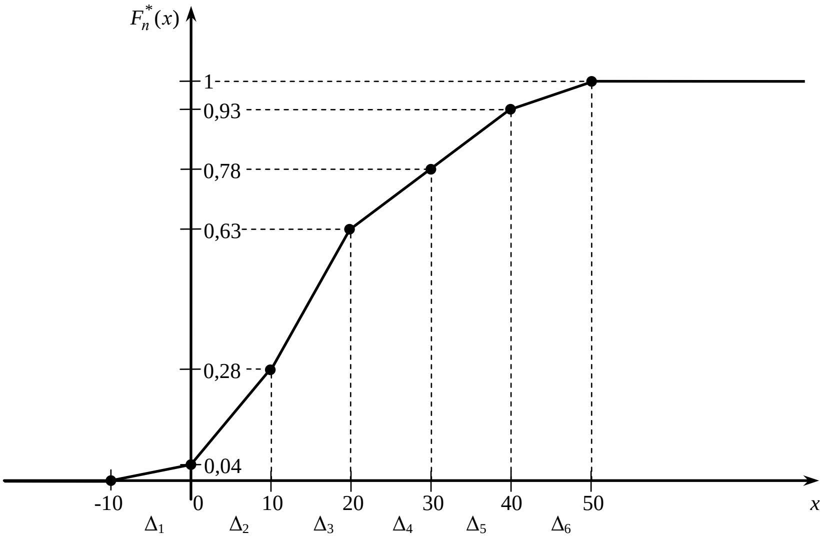

Рис.5. Эмпирическая функция распределения

Определяем эмпирические плотности:

$$f_1^* = \frac{p_1^*}{\Delta_1} = \frac{0{,}04}{10} = 0{,}004;$$

$$f_1^* = \frac{p_2^*}{\Delta_2} = \frac{0{,}24}{10} = 0{,}024;$$

$$f_3^* = \frac{p_3^*}{\Delta_3} = \frac{0{,}35}{10} = 0{,}035;$$

$$f_4^* = \frac{p_4^*}{\Delta_4} = \frac{0{,}15}{10} = 0{,}015;$$

$$f_5^* = \frac{p_5^*}{\Delta_5} = \frac{0{,}15}{10} = 0{,}015;$$

$$f_6^* = \frac{p_6^*}{\Delta_6} = \frac{0{,}07}{10} = 0{,}007.$$

Строим гистограмму (рис.6).

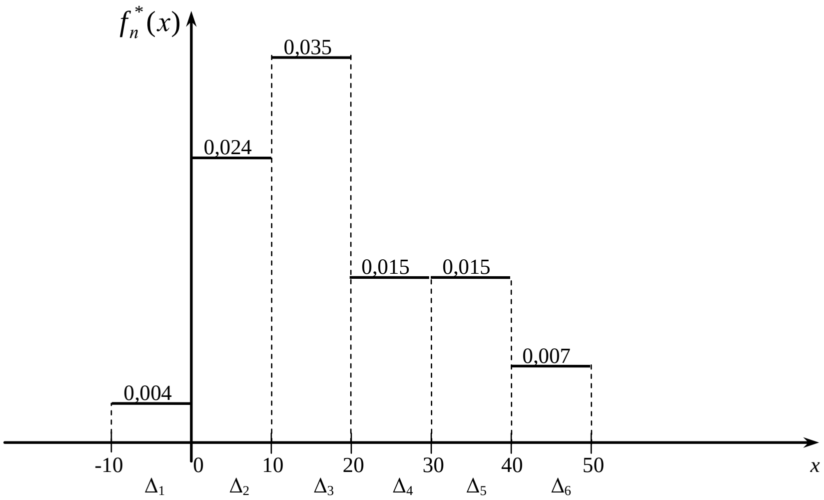

Рис.6. Гистограмма

По вариационному ряду (в том числе, сгруппированному) вычисляются основные эмпирические или выборочные характеристики: выборочное среднее $\bar{x}_{в}$, выборочная дисперсия $D_{в}$ и выборочное отклонение $\sigma_{в}$:

$$\bar{x}_{в} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i\,;\;\; D_{в} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 m_i - (\bar{x}_{в})^2\,;\;\; \sigma_{в} = \sqrt{D_{в}}\,.$$

Для каждой выборочной характеристики получается одно определенное значение (точка), которая является приближением соответствующей неизвестной характеристики $\bar{x}$ или D_x случайной величины X. Поэтому эти приближения называют *точечными оценками* характеристик (или параметров) неизвестного распределения. По закону больших чисел эти точечные оценки сходятся к соответствующим неизвестным значениям: $\bar{x}_{в} \to \bar{x}$, $D_{в} \to D_x$ при $n \to \infty$, т.е. эти оценки являются *состоятельными*. Кроме того, выборочное среднее является *несмещенной* оценкой, т.е. его математическое ожидание (среднее!) равно неизвестному значению $\bar{x}$: $M(\bar{x}_{в}) = \bar{x}$. Выборочная дисперсия является смещенной оценкой: $M(D_{в}) = \frac{n-1}{n} D$. В результате при небольших объемах ($n < 30$) часто рассматривают *исправленные* дисперсию $S_{в} = \frac{n}{n-1} D_{в}$ и отклонение $s_{в} = \sqrt{S_{в}}$ вместо $D_{в}$ и $\sigma_{в}$ соответственно.

Другой способ оценки неизвестных характеристик или параметров распределения заключается в указании интервала, куда попадает неизвестное значение с заданной вероятностью (или с заданной надежностью):

$$P(|\theta - \theta_{в}| < \varepsilon) = P(\theta_{в} - \varepsilon < \theta < \theta_{в} + \varepsilon) = \gamma\,,$$

где θ – неизвестное значение; $\theta_{в}$ – выборочное значение; γ – надежность (или доверительная вероятность); $(\theta_{в} - \varepsilon, \theta_{в} + \varepsilon)$ – доверительный интервал.

Такие оценки называются *интервальными*. Например, если распределение X является нормальным с неизвестным $a=\overline{X}$ и известным $\sigma=\sigma(X)$ параметрами, то радиус интервала $\varepsilon=t_\gamma\sigma/\sqrt{n}$, где $2\Phi(t_\gamma)=\gamma$, и доверительный интервал для a с надежностью γ

$$\bar{x}_{\text{в}}-\frac{t_\gamma\sigma}{\sqrt{n}}<a<\overline{x_{\text{в}}}+\frac{t_\gamma\sigma}{\sqrt{n}}.$$

Если вместо значения σ, которое может быть неизвестно, использовать точечную оценку $\sigma_{\text{в}}$, то получим приближенную интервальную оценку с $\varepsilon=t_\gamma\sigma_{\text{в}}/\sqrt{n}$, которая по ЦПТ может применяться и для любого X.

Вероятность $\alpha=1-\gamma$ задает вероятность ошибки, т.е. того, что значение a не попадает в доверительный интервал.

Отметим, что имеются и другие виды интервальных оценок для этих и других параметров распределения [1, 2].

В случае равноотстоящих друг от друга значений x_i (например, для сгруппированного вариационного ряда) можно упростить вычисления выборочных характеристик, если, выбрав значение $C=x_{i_0}$ (поближе к середине ряда и с большей частотой m_{i_0}), называемое «ложным» нулем, и определив величину шага h для значений ряда, ввести **условную** варианту по формуле:

$$u=\frac{x-C}{h}\Rightarrow u_i=\frac{x_i-C}{h}.$$

Тогда значения условной варианты будут целыми числами, причем, большой частоте m_{i_0} будет отвечать $u_{i_0}=0$, и поэтому выборочные характеристики для условной варианты вычисляются проще:

$$\bar{u}=\frac{1}{n}\sum_{i=1}^{k}u_i m_i\,;\ \ \overline{u^2}=\frac{1}{n}\sum_{i=1}^{k}u_i^2 m_i\,;$$

$$D_u = \overline{u^2} - (\bar{u})^2; \quad \sigma_u = \sqrt{D_u}\ .$$

Обратный пересчет производится по формулам

$$x = hu + C \Rightarrow \bar{x}_в = h\bar{u} + C, \quad D_в = h^2 D_u, \quad \sigma_в = h\sigma_u\ .$$

При изучении СВ возникает вопрос о возможном виде ее распределения, т.е. о соответствии (согласии) выборочных данных некоторому гипотетическому теоретическому распределению, что является одной из важных задач проверки статистических гипотез.

Основное предположение называется нулевой гипотезой H_0. Возможно рассмотрение и противоположной (альтернативной) гипотезы или каких-нибудь других гипотез. В нашем случае проверка гипотезы H_0 состоит в том, что эмпирические данные получены для нормально распределенной генеральной совокупности. Следовательно, при альтернативной гипотезе эмпирические данные не согласуются с ожидаемым нормальным распределением.

Проверка статистических гипотез осуществляется с помощью статистических критериев. Критерий – случайная величина, значение которой вычисляется по эмпирическим данным, т.е. по выборке. Статистический критерий определяет критическую область, при попадании в которую выборочного значения критерия нулевая гипотеза отвергается. Отвергая нулевую гипотезу (если она на самом деле верна), совершают ошибку первого рода; не отвергая нулевую гипотезу (если она на самом деле неверна), допускают ошибку второго рода. Критическая область определяется так, чтобы вероятность ошибки первого рода не превышала уровня значимости α, а вероятность совершить ошибку второго рода была бы наименьшей. Обычно в качестве α берут маленькое число (0,05; 0,01; 0,001; …), при этом следует учитывать, что при $\alpha \to 0$ будет увеличиваться критическая область, т.е. практически все гипотезы будут отвергаться.

Рассмотрим достаточно простой и эффективный критерий согласия – *критерий Пирсона хи-квадрат* (χ^2), для которого мерой расхождения между эмпирическим распределением (выборкой) и теоретическим распределением является разность между эмпирическими и теоретическими частотами для одного и того же значения дискретной случайной величины или, соответственно, для одного и того же интервала в случае непрерывной случайной величины. Для критерия Пирсона находят величину

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

где m_i – эмпирическая частота; p_i – соответствующая вероятность для теоретического распределения; np_i – теоретическая частота; $n = \sum_{i=1}^{k} m_i$ – объем выборки.

Распределение критерия χ^2 зависит от числа степеней свободы r и уровня значимости α. Число r определяется числом значений (или интервалов) k и числом наложенных связей ρ, равным числу соотношений для выборочных данных и теоретических параметров: $r = k - \rho$. Например, так как всегда $\sum_{i=1}^{k} m_i = 1$, то $\rho = 1$; если дополнительно положим $M(X) = \bar{x}_{\text{в}}$, то $\rho = 2$; если еще положим и $D(X) = D_{\text{в}}$ (т.е. $\sigma(X) = \sigma_{\text{в}}$), то $\rho = 3$ и т.д. По специальной таблице [2, 3], зная значения r и α, находят критическое значение $\chi^2_\alpha(r)$. Вычисленное по выборке значение χ^2 сравнивают с критическим значением: если $\chi^2 \le \chi^2_\alpha(r)$, то различие эмпирических данных с теоретическим распределением можно считать несущественным и гипотеза о согласии эмпирических данных с теоретическим распределением не отвергается; если $\chi^2 > \chi^2_\alpha(r)$, называемой критической областью, то различия существенны и гипотезу о согласии следует отвергнуть.

Отметим, что при использовании критерия Пирсона значения, частоты которых малы ($m_i < 5$), можно объединить (обычно это крайние значения или интервалы).

Пример 6. При проведении испытаний материала на разрыв получено 50 значений, характеризующих прочность на разрыв. По этим данным составлен сгруппированный вариационный ряд (масштаб 10^4 Па).

Интервал Δ_i	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
x_i	130	150	170	190	210	230	250	270
m_i	2	4	10	13	11	6	3	1

Оценить согласие полученных данных с нормальным распределением при уровне значимости $\alpha = 0{,}05$ и получить приближенную интервальную оценку для параметра $a = \overline{X}$ с надежностью $\gamma = 0{,}95$.

Решение. Введем условную варианту, определив шаг $h = 20$ и выбрав ложный нуль $C = 190$, и найдем $\bar{x}_в$ и $\sigma_в$ (табл.2).

Таблица 2

Интервал Δ_i	x_i	m_i	$u_i = \dfrac{x_i - C}{h}$	$u_i m_i$	u_i^2	$u_i^2 m_i$
1	130	2	–3	–6	9	18
2	150	4	–2	–8	4	16
3	170	10	–1	–10	1	10
4	190	13	0	0	0	0
5	210	11	1	11	1	11
6	230	6	2	12	4	24
7	250	3	3	9	9	27
8	270	1	4	4	16	16
Σ		50		12		122

По данным табл.2 имеем $n = 50$ и

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{8} u_i m_i = \frac{12}{50} = \frac{6}{25} \Rightarrow$$

$$\Rightarrow \bar{x}_в = h\bar{u} + C = 20 \cdot \frac{6}{25} + 190 = 194{,}8;$$

$$\overline{u^2} = \frac{1}{n}\sum_{i=1}^{8} u_i^2 m_i = \frac{122}{50} = \frac{61}{25} \Rightarrow$$

$$\Rightarrow D_u = \overline{u^2} - \bar{u}^2 = \frac{61}{25} - \frac{36}{625} = \frac{1489}{625};$$

$$\sigma_u = \sqrt{D_u} = \sqrt{\frac{1489}{625}} = \frac{\sqrt{1489}}{25} = 1{,}54 \Rightarrow$$

$$\Rightarrow \sigma_в = h\sigma_u = 20 \cdot 1{,}54 = 30{,}8.$$

Найдем теоретические частоты (табл.3) для интервалов $\Delta_i = (\alpha_i, \beta_i)$, используя формулу вероятности попадания значений в этот интервал (для нормального распределения с параметрами $a = \bar{x}_в$ и $\sigma = \sigma_в$):

$$p_i = P(\alpha_i < x < \beta_i) = \Phi\left(\frac{\beta_i - a}{\sigma}\right) - \Phi\left(\frac{\alpha_i - a}{\sigma}\right).$$

Таблица 3

Δ_i	$t_1 = \frac{\alpha_i - \bar{x}_в}{\sigma_в}$	$t_2 = \frac{\beta_i - \bar{x}_в}{\sigma_в}$	$\Phi(t_1)$	$\Phi(t_2)$	$P_i = \Phi(t_2) - \Phi(t_1)$	np_i
120-140	–2,43	–1,78	–0,4924	0,4624	0,0300	$1{,}5 \approx 1$
140-160	–1,78	–1,13	–0,4624	–0,3708	0,0916	$4{,}58 \approx 5$
160-180	–1,13	–0,48	–0,3708	–0,1844	0,1864	$9{,}34 \approx 9$

180-200	–0,48	0,17	–0,1844	0,0675	0,2519	$12{,}59 \approx 13$
200-220	0,17	0,81	0,0675	0,2910	0,2235	$11{,}17 \approx 11$
220-240	0,81	1,46	0,2910	0,4279	0,1369	$6{,}87 \approx 7$
240-260	1,46	2,11	0,4279	0,4826	0,0547	$2{,}78 \approx 3$
260-280	2,11	2,73	0,4826	0,4968	0,0142	$0{,}71 \approx 1$

Найдем выборочное значение χ^2, объединив крайние интервалы для маленьких частот m_i (табл.4). Это объединение не является необходимым, но вполне применимо для упрощения в случае маленьких частот.

Таблица 4

Номер интер-вала	Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	120-140	$\left.\begin{matrix}2\\4\end{matrix}\right\}6$	6	0	0	0
2	140-160					
3	160-180	10	9	1	1	0,111
4	180-200	13	13	0	0	0
5	200-220	11	11	0	0	0

Окончание табл.4

Номер интер-вала	Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
6	220-240	6	7	–	1	0,143
7	240-260	$\left.\begin{matrix}3\\1\end{matrix}\right\}4$	4	0	0	0
8	260-280					

Таким образом, если после объединения число интервалов $k = 6$, а число наложенных связей $\rho = 3$, то число степеней свободы $r = k - \rho = 6 - 3 = 3$. Поэтому по таблице критических значений $\chi^2_\alpha(r)$ (прил.3) имеем $\chi^2_{0,05}(3) = 7{,}82$. Сравнивая найденное значение

$\chi^2 = 0{,}254$ с критическим (0,254 < 7,82), получим, что рассматриваемые данные могут быть из нормально распределенной совокупности.

Для получения интервальной оценки найдем t_γ из условия $2\Phi(t_\gamma) = \gamma = 0{,}95$, т.е. $\Phi(t_\gamma) = \gamma = 0{,}475$ и $t_\gamma = 1{,}96$, и радиус интервала $\varepsilon = \dfrac{t_\gamma \sigma_{\text{в}}}{\sqrt{n}} = \dfrac{1{,}96 \cdot 30{,}8}{\sqrt{50}} = 8{,}54$.

Вычислим доверительный интервал для параметра $a = \overline{X}$ с надежностью $\gamma = 0{,}95$:

$$\bar{x}_{\text{в}} - \varepsilon = 194{,}8 - 8{,}54 = 186{,}26 < a < \bar{x}_{\text{в}} + \varepsilon = 194{,}8 + 8{,}54 = 203{,}34.$$

Пусть для изучаемой системы случайных величин (X, Y) получена выборка значений системы (x_i, y_i) с соответствующими совместными частотами $m_{ij} (1 \le i \le k,\ 1 \le j \le l)$. Объем выборки $n = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{l} m_{ij}$, каждое значение x_i встречается с частотой $m_i = \sum\limits_{j=1}^{l} m_{ij}$, а каждое значение y_j встречается, соответственно, с частотой $n_j = \sum\limits_{i=1}^{k} m_{ij}$. *Условные средние* $\bar{y}_{x_i}$ и $\bar{x}_{y_j}$ представляют отдельные значения для регрессий соответственно Y на X и X на Y:

$$\bar{y}_{x_i} = \frac{1}{m_i} \sum_{j=1}^{l} m_{ij} y_j, \quad \bar{x}_{y_j} = \frac{1}{n_j} \sum_{i=1}^{k} m_{ij} x_i \quad (1 \le i \le k;\ 1 \le j \le l).$$

По выборке системы СВ определяют выборочные наилучшие линейные регрессии, которые приближенно выражают регрессионную (или корреляционную) зависимость между рассматриваемыми в системе случайными величинами:

$$y - \bar{y}_{\text{в}} = r_{xy}^* \frac{\sigma_y^*}{\sigma_x^*} (x - \bar{x}_{\text{в}}), \quad x - \bar{x}_{\text{в}} = r_{xy}^* \frac{\sigma_x^*}{\sigma_y^*} (y - \bar{y}_{\text{в}}),$$

где σ_x^*, σ_y^* – выборочные отклонения для X и Y соответственно. Выборочный коэффициент корреляции

$$r_{xy}^* = \frac{\frac{1}{n}\sum_{i,j} x_i y_j m_{ij} - \bar{x}_{\text{в}}\bar{y}_{\text{в}}}{\sigma_x^*\sigma_y^*}.$$

Выборочные регрессии Y на X и X на Y приближают точки $(x_i, \bar{y}_{x_i})$ и $(y_j, \bar{x}_{y_j})$ соответствующих условных средних $\bar{y}_{x_i}$ и $\bar{x}_{y_j}$ соответственно. По величине коэффициента $\left|r_{xy}^*\right|$ или по значению угла между прямыми выборочных регрессий можно сделать вывод о качестве связи между случайными величинами. Для расчета σ_x^*, σ_y^* и r_{xy}^* удобно использовать условные варианты для X и Y:

$$u_i = \frac{x_i - C_1}{h_1}, \quad v_j = \frac{y_j - C_2}{h_2}, \quad r_{xy}^* = r_{uv}^* = \frac{\frac{1}{n}\sum_{i,j} u_i v_j m_{ij} - \bar{u}\bar{v}}{\sigma_u\sigma_v}.$$

Пример 7. Пусть имеется 100 сгруппированных наблюдений двух измеримых признаков X и Y, по которым составлена корреляционная таблица (x_i, y_j, m_{ij}) (табл.5).

Таблица 5

Y	X						n_j	$\bar{x}_{y_j}$
	30	35	40	45	50	55		
18	4	6					10	33,0
28		8	10				18	37,78
38			4	35	5		44	45,11
48			4	12	6		22	45,45
58				1	3	2	6	50,83
m_i	4	14	18	48	14	2	n = 100	
$\bar{y}_{x_i}$	18,00	23,71	34,67	40,92	46,57	58,0		

Найти выборочные регрессии и оценить качество связи признаков.

Решение. В табл.5 уже найдены отдельные частоты n_j для y_j (суммы частот m_{ij} по строкам), частоты m_i для x_i (сумма частот m_{ij} по столбцам) и условные средние. Например:

$$\bar{y}_{x_2} = \frac{1}{14}(6 \cdot 18 + 8 \cdot 28) = 23{,}71, \quad \bar{x}_{y_3} = \frac{1}{44}(4 \cdot 40 + 35 \cdot 45 + 5 \cdot 50) = 45{,}11.$$

Соответственно, точки этих условных средних $(x_2; \bar{y}_{x_2}) = (35; 23{,}71)$ и $(\bar{x}_{y_3}, y_3) = (45{,}11; 38)$.

Для определения выборочных регрессий перейдем к условным вариантам.

Наибольшая частота, ближайшая к центру таблицы, $m_{43} = 35$ и, следовательно, соответствующие ложные нули $C_1 = x_4 = 45$ и $C_2 = y_3 = 38$, шаг $h_1 = 5$ (для x_i) и $h_2 = 10$ (для y_j).

Составим новую таблицу в условных вариантах для расчета характеристик (табл.6).

Таблица 6

v_j	u_i								
	–3	–2	–1	0	1	2	n_j	$v_j n_j$	$v_j^2 n_j$
–2	4	6					10	–20	40
–1		8	10				18	–18	18
0			4	35	5		44	0	0
1			4	12	6		22	22	22
2				1	3	2	6	12	24
m_i	4	14	18	48	14	2	$n = 100$	$\Sigma_{1y} = -4$	$\Sigma_{2y} = 104$
$m_i u_i$	–12	–28	–18	0	14	4	$\Sigma_{1x} = -40$		

$m_i u_i^2$	36	56	18	0	14	8	$\Sigma_{2x} = 132$	

По данным табл.6 получим выборочные характеристики:

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{6} m_i u_i = \frac{1}{n}\Sigma_{1x} = \frac{-40}{100} = -0{,}4;$$

$$\bar{v} = \frac{1}{n}\sum_{j=1}^{5} n_j v_j = \frac{1}{n}\Sigma_{1y} = \frac{-4}{100} = -0{,}04;$$

$$\overline{u^2} = \frac{1}{n}\sum_{i=1}^{6} m_i u_i^2 = \frac{1}{n}\Sigma_{2x} = \frac{132}{100} = 1{,}32;$$

$$\overline{v^2} = \frac{1}{n}\sum_{j=1}^{5} n_j v_j^2 = \frac{1}{n}\Sigma_{2y} = \frac{104}{100} = 1{,}04;$$

$$\bar{x}_{\text{в}} = h_1\bar{u} + C_1 = 5(-0{,}4) + 45 = 43{,}0;$$

$$\bar{y}_{\text{в}} = h_2\bar{v} + C_2 = 10(-0{,}04) + 38 = 37{,}6;$$

$$\sigma_u = \sqrt{D_u} = \sqrt{\overline{u^2} - \bar{u}^2} = \sqrt{1{,}32 - (-0{,}4)^2} = 1{,}077 \Rightarrow$$

$$\Rightarrow \sigma_x^* = h_1\sigma_u = 5 \cdot 1{,}077 = 5{,}385;$$

$$\sigma_v = \sqrt{D_v} = \sqrt{\overline{v^2} - \bar{v}^2} = \sqrt{1{,}04 - (-0{,}04)^2} = 1{,}019 \Rightarrow$$

$$\Rightarrow \sigma_y^* = h_2\sigma_v = 10 \cdot 1{,}019 = 10{,}19.$$

Для вычисления r_{uv}^* найдем средние суммы всех произведений $u_i v_j m_{ij}$:

$$\frac{1}{n}\sum_{i,j} u_i v_j m_{ij} = \frac{1}{100}[(-2)(-3)\cdot 4 + (-2)(-2)\cdot 6 + (-1)(-2)\cdot 8 +$$

$$+(-1)(-1)\cdot 10 + 0\cdot(-1)\cdot 4 + 0\cdot 0\cdot 35 + 0\cdot 1\cdot 5 + 1\cdot(-1)\cdot 4 +$$

$$+1\cdot 0\cdot 12 + 1\cdot 1\cdot 6 + 2\cdot 0\cdot 1 + 2\cdot 1\cdot 3 + 2\cdot 2\cdot 2] = \frac{90}{100} = 0{,}9\,.$$

Выборочный коэффициент корреляции

$$r_{xy}^* = r_{uv}^* = \frac{\frac{1}{n}\sum_{i,j} u_i v_j m_{ij} - \bar{u}\bar{v}}{\sigma_u \sigma_v} = \frac{0{,}9 - (-0{,}4)(-0{,}04)}{1{,}077\cdot 1{,}019} = 0{,}805.$$

Уравнения выборочных регрессий имеют вид
для регрессии Y на X

$$y - \bar{y}_в = r_{xy}^* \frac{\sigma_y^*}{\sigma_x^*}(x - \bar{x}_в) \Rightarrow y - 37{,}6 =$$

$$= 0{,}805\cdot\frac{10{,}19}{5{,}385}\cdot(x - 43) \Rightarrow y = 1{,}523x - 27{,}902\,;$$

для регрессии X на Y

$$x - \bar{x}_в = r_{xy}^* \frac{\sigma_x^*}{\sigma_y^*}(y - \bar{y}_в) \Rightarrow x - 43 =$$

$$= 0{,}805\cdot\frac{5{,}385}{10{,}19}\cdot(y - 37{,}6) \Rightarrow x = 0{,}425y + 27{,}005.$$

Обе прямых регрессий проходят через точку средних $\bar{x}_в = 43$; $\bar{y}_в = 37{,}6$ и для построения последних достаточно найти еще по одной точке для каждой прямой.

Так как $\left|r_{xy}^{*}\right| = 0{,}805$ значительно отличаются от нуля, то связь между изучаемыми случайными величинами достаточно сильная, а так как это значение еще не близко к единице, связь нелинейная. Аналогичные выводы можно сделать и по величине угла между прямыми регрессий. Для наглядности точки условных средних $(x_i, \bar{y}_{x_i})$ и $(\bar{x}_{y_j}, y_j)$ вместе с прямыми выборочных регрессий изображаются на одном чертеже, причем во избежание искажений, масштабы на осях следует брать одинаковыми.

РАСЧЕТНО-ГРАФИЧЕСКАЯ РАБОТА

Задача 1. По двум последним цифрам шифра студента (… ab) определяется вариационный ряд из двадцати значений (с шагом h = 3) и соответствующих частот:

$$x_1 = b - a;\ \ x_2 = x_1 + 3,\ \ldots,\ x_{20} = x_{19} + 3;$$

$$m_i = \left|i - a\right| + 5 + (-1)^{b+i} \cdot 4\ \ (1 \le i \le 20)\,.$$

Произвести группировку значений и по сгруппированному вариационному ряду построить эмпирическую функцию распределения и гистограмму.

Задача 2. Сгруппированный вариационный ряд задан серединами интервалов x_i и соответствующими им частотами m_i (табл.10 по вариантам). Восстановить интервалы и оценить с помощью критерия Пирсона хи-квадрат согласие данных с нормальным распределением при уровне значимости $\alpha = 1 - (0{,}89 + 0{,}01(b + 1))$, где b – последняя цифра шифра.

Таблица 10

Вариант	x_i						m_i					
	x_1	x_2	x_3	x_4	x_5	x_6	m_1	m_2	m_3	m_4	m_5	m_6
1	11	21	31	41	51	61	6	8	14	11	7	4

2	2,5	3,5	4,5	5,5	6,5	7,5	5	8	14	12	6	5
3	11	21	31	41	51	61	5	7	9	14	9	6
4	26	36	46	56	66	76	6	8	13	12	7	4
5	1,5	2,5	3,5	4,5	5,5	6,5	7	8	13	10	7	5
6	1,5	2,5	3,5	4,5	5,5	6,5	6	7	10	13	9	5
7	2	3	4	5	6	7	5	8	11	14	7	5
8	16	17	18	19	20	21	5	8	14	11	7	5
9	1	2	3	4	5	6	6	9	10	14	6	5
10	3,5	4,5	5,5	6,5	7,5	8,5	7	7	12	10	9	5

Задача 3. Найти выборочные регрессии, построить их графики и точки условных средних на одном чертеже. Оценить качество связи. Корреляционная таблица (табл.11) определяется двумя последними цифрами шифра студента (… ab).

Таблица 11

Y	X					
	b	$b + (10 - a)$	$b + 2(10 - a)$	$b + 3(10 - a)$	$b + 4(10 - a)$	$b + 5(10 - a)$
a	8	$10 - a$		$12 - b$		
$a + 10$			$2b$		$20 - 2b$	1
$a + 20$				$30 - a - b$		
$a + 30$		6	a			b
$a + 40$	b	$a + b$		$10 - b$	3	

Замечание. Шифр студенту присваивает лектор потока, например можно положить:

$$ab = N+30(k-1),$$

где N = № по журналу группы, k = номер группы потока.

РЕКОМЕНДАТЕЛЬНЫЙ БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Основной:

1. *Гмурман В.Е.* Теория вероятностей и математическая статистика. М.: Высшая школа, 1998. 479 с.
2. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 1998. 400 с.
3. Математический практикум. Часть 5. Теория вероятностей и математическая статистика. Теория функций комплексного переменного. Операционное исчисление. Теория поля: Учебно-методическое пособие / А.П.Господариков, О.Е.Карпухина, Г.А.Колтон, И.А.Лебедев, С.Е.Мансурова, Т.С.Обручева, В.В.Тарабан; Санкт-Петербургский горный институт. СПб, 2003. 187 с.

Дополнительный:

4. *Вентцель Е.С.* Теория вероятностей / Е.С.Вентцель, Л.А.Овчаров. М.: Наука, 1973. 366 с.
5. *Данко П.Е.* Высшая математика в упражнениях и задачах / П.Е.Данко, А.Г.Попов, Т.Я.Кожевников. М.: Высшая школа, 1986. Ч.2. 415 с.

Приложение 1

Таблица значений функции $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-0,5x^2}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229

Окончание таблицы

x	0	1	2	3	4	5	6	7	8	9
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006

Приложение 2

Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-0,5x^2}dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,14	0,0557	0,28	0,1103	0,42	0,1628
0,01	0,0040	0,15	0,0596	0,29	0,1141	0,43	0,1664
0,02	0,0080	0,16	0,0636	0,30	0,1179	0,44	0,1700
0,03	0,0120	0,17	0,0675	0,31	0,1217	0,45	0,1736
0,04	0,0160	0,18	0,0714	0,32	0,1255	0,46	0,1772
0,05	0,0199	0,19	0,0753	0,33	0,1293	0,47	0,1808
0,06	0,0239	0,20	0,0793	0,34	0,1331	0,48	0,1844
0,07	0,0279	0,21	0,0832	0,35	0,1368	0,49	0,1879
0,08	0,0319	0,22	0,0871	0,36	0,1406	0,50	0,1915
0,09	0,0359	0,23	0,0910	0,37	0,1443	0,51	0,1950
0,10	0,0398	0,24	0,0948	0,38	0,1480	0,52	0,1985
0,11	0,0438	0,25	0,0987	0,39	0,1517	0,53	0,2019
0,12	0,0478	0,26	0,1026	0,40	0,1554	0,54	0,2054
0,13	0,0517	0,27	0,1064	0,41	0,1591	0,55	0,2088
0,56	0,2123	0,87	0,3078	1,18	0,3810	1,49	0,4319
0,57	0,2157	0,88	0,3106	1,19	0,3830	1,50	0,4332
0,58	0,2190	0,89	0,3133	1,20	0,3849	1,51	0,4345
0,59	0,2224	0,90	0,3159	1,21	0,3869	1,52	0,4357
0,60	0,2257	0,91	0,3186	1,22	0,3883	1,53	0,4370
0,61	0,2291	0,92	0,3212	1,23	0,3907	1,54	0,4382
0,62	0,2324	0,93	0,3238	1,24	0,3925	1,55	0,4394
0,63	0,2357	0,94	0,3264	1,25	0,3944	1,56	0,4406
0,64	0,2389	0,95	0,3289	1,26	0,3962	1,57	0,4418
0,65	0,2422	0,96	0,3315	1,27	0,3980	1,58	0,4429

Продолжение таблицы

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,66	0,2454	0,97	0,3340	1,28	0,3997	1,59	0,4441
0,67	0,2486	0,98	0,3365	1,29	0,4015	1,60	0,4452
0,68	0,2517	0,99	0,3389	1,30	0,4032	1,61	0,4463
0,69	0,2549	1,00	0,3413	1,31	0,4049	1,62	0,4474
0,70	0,2580	1,01	0,3438	1,32	0,4066	1,63	0,4484
0,71	0,2611	1,02	0,3461	1,33	0,4082	1,64	0,4495
0,72	0,2642	1,03	0,3485	1,34	0,4099	1,65	0,4505
0,73	0,2673	1,04	0,3508	1,35	0,4115	1,66	0,4515
0,74	0,2703	1,05	0,3531	1,36	0,4131	1,67	0,4525
0,75	0,2734	1,06	0,3554	1,37	0,4147	1,68	0,4535
0,76	0,2764	1,07	0,3577	1,38	0,4162	1,69	0,4545
0,77	0,2794	1,08	0,3599	1,39	0,4177	1,70	0,4554
0,78	0,2823	1,09	0,3621	1,40	0,4192	1,71	0,4564
0,79	0,2852	1,10	0,3643	1,41	0,4207	1,72	0,4573
0,80	0,2881	1,11	0,3665	1,42	0,4222	1,73	0,4582
0,81	0,2910	1,12	0,3686	1,43	0,4236	1,74	0,4591
0,82	0,2939	1,13	0,3708	1,44	0,4251	1,75	0,4599
0,83	0,2967	1,14	0,3729	1,45	0,4265	1,76	0,4608
0,84	0,2995	1,15	0,3749	1,46	0,4279	1,77	0,4616
0,85	0,3023	1,16	0,3770	1,47	0,4292	1,78	0,4625
0,86	0,3051	1,17	0,3790	1,48	0,4306	1,79	0,4633
1,80	0,4641	2,00	0,4772	2,40	0,4918	2,80	0,4974
1,81	0,4649	2,02	0,4783	2,42	0,4922	2,82	0,4976
1,82	0,4656	2,04	0,4793	2,44	0,4927	2,84	0,4977
1,83	0,4664	2,06	0,4803	2,46	0,4931	2,86	0,4979
1,84	0,4671	2,08	0,4812	2,48	0,4934	2,88	0,4980
1,85	0,4678	2,10	0,4821	2,50	0,4938	2,90	0,4981

Окончание таблицы

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,86	0,4686	2,12	0,4830	2,52	0,4941	2,92	0,4982
1,87	0,4693	2,14	0,4838	2,54	0,4945	2,94	0,4984
1,88	0,4699	2,16	0,4846	2,56	0,4948	2,96	0,4985
1,89	0,4706	2,18	0,4854	2,58	0,4951	2,98	0,4986
1,90	0,4713	2,20	0,4861	2,60	0,4953	3,00	0,49865
1,91	0,4719	2,22	0,4868	2,62	0,4956	3,20	0,49931
1,92	0,4726	2,24	0,4875	2,64	0,4959	3,40	0,49966
1,93	0,4732	2,26	0,4881	2,66	0,4961	3,60	0,49984
1,94	0,4738	2,28	0,4887	2,68	0,4963	3,80	0,49993
1,95	0,4744	2,30	0,4893	2,70	0,4965	4,00	0,49997
1,96	0,4750	2,32	0,4898	2,72	0,4967	4,50	0,499997
1,97	0,4756	2,34	0,4904	2,74	0,4969	5,00	0,499997
1,98	0,4761	2,36	0,4909	2,76	0,4971	∞	0,5
1,99	0,4767	2,38	0,4913	2,78	0,4973		

Приложение 3

Значения χ^2_α в зависимости от числа степеней свободы *m*
и уровня значимости $\alpha = 1 - \gamma$ (γ – доверительная вероятность)

m	α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,71	3,84	5,41	6,64
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,60	5,99	7,82	9,21
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37	3,66	4,64	6,25	7,82	9,84	11,34
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,67	13,28
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35	6,06	7,29	9,24	11,07	13,39	15,09
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81
7	1,239	1,564	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	14,07	16,62	18,48
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	19,68	21,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31	21,2	23,2
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68	22,6	24,7
12	3,57	4,18	5,23	6,3	7,81	9,03	11,34	14,01	15,81	18,55	21,0	24,1	26,2
13	4,11	4,70	5,88	7,04	8,63	9,93	12,34	15,12	16,98	19,81	22,4	25,5	27,7
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,1	23,7	26,9	29,1
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34	17,32	19,31	22,3	25,0	28,3	30,6

Окончание таблицы

m	α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34	18,42	20,5	23,5	26,3	29,6	32,0
17	6,41	7,26	8,67	10,08	12,00	13,53	16,34	19,51	21,6	24,8	27,6	31,0	33,4
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34	20,6	22,8	26,0	28,9	32,3	34,8
19	7,63	8,57	10,11	11,65	13,72	15,35	18,34	21,7	23,9	27,2	30,1	33,7	36,2
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34	22,8	25,0	28,4	31,4	35,0	37,6
21	8,90	9,92	11,59	13,24	15,44	17,18	20,3	23,8	26,2	29,6	32,7	36,3	38,9
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3	24,9	27,3	30,8	33,9	37,7	40,3
23	10,20	11,29	13,09	14,85	17,19	19,02	22,3	26,0	28,4	32,0	35,2	39,0	41,6
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3	27,1	29,6	33,2	36,4	40,3	43,0
25	11,52	12,70	14,61	16,47	18,94	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6
27	12,88	14,12	16,15	18,11	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3
29	14,26	15,57	17,71	19,77	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9

СОДЕРЖАНИЕ