

Г.Н. Кокурина

Математические методы  
обработки экспериментальных данных

Учебное пособие

ИВАНОВО

2019

Министерство науки и высшего образования Российской Федерации  
Ивановский государственный химико-технологический университет

Г.Н. Кокурина

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ  
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ**

Учебное пособие

Иваново 2019

**УДК 519.25**

**Кокурина, Г.Н.**

Математические методы обработки экспериментальных данных: учебное пособие / Г.Н. Кокурина; Иван. гос. хим.-технол. ун-т. Иваново, 2019. 123 с.

Учебное пособие содержит основные теоретические положения, примеры и задания для выполнения лабораторных работ, связанных с математическими методами обработки экспериментальных данных.

Предназначено для студентов и магистрантов, обучающихся по направлению 15.04.02 «Технологические машины и оборудование», но будет полезно также для студентов других технических специальностей и направлений при написании магистерской диссертации и дальнейшего применения при решении профессиональных задач, связанных с проведением научных и технических экспериментов.

Печатается по решению редакционно-издательского совета Ивановского государственного химико-технологического университета.

**Рецензенты:**

кандидат физико-математических наук, доцент О.В. Хонгорова  
(ФГБОУ ВО Ивановская пожарно-спасательная академия ГПС МЧС России);

кандидат психологических наук, доцент И.Б. Румянцева  
(Шуйский филиал ФГБОУ ВО Ивановского государственного университета)

© Кокурина Г.Н., 2019

© ФГБОУ ВО «Ивановский государственный  
химико-технологический университет», 2019

## ВВЕДЕНИЕ

В современном быстро меняющемся мире большую роль играют информационные потоки. Умение и готовность обработки и анализа данных становятся очень важными как для профессиональной работы инженера, руководителя, так и для научного работника.

Для эффективного управления и прогнозирования развития любой отрасли промышленности необходим постоянный контроль и анализ состояния ее систем. Это означает, что математическому моделированию подлежит информационный поток – *статистическая совокупность* – в виде случайных событий и случайных величин. Изучение подобного рода *массовых явлений*, выявление их статических и динамических закономерностей становится *предметом математической статистики*.

Для результативной работы над магистерской диссертацией студент должен иметь сформированные базовые знания и навыки в области методов обработки информации, анализа экспериментальных данных, направленных на решение профессиональных задач, связанных с проведением научных и технических экспериментов.

В данном учебном пособии рассмотрены основы теории статистического анализа результатов измерений. В каждом разделе содержатся теоретические положения, задания к лабораторным работам, варианты для студентов и пример выполнения лабораторной работы.

Первый раздел пособия, включающий расчет выборочных характеристик результатов наблюдений и интервальные оценки параметров распределения, посвящен вопросам первичной обработки статистических данных.

Во втором разделе рассмотрена процедура проверки соответствия выборки нормальному закону распределения.

Третий и четвертый разделы посвящены корреляционному анализу сгруппированных и несгруппированных данных.

# 1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННЫХ РЯДОВ

## 1.1. Первичная обработка результатов наблюдений

Анализ любой совокупности инженерных данных начинается с их первичной обработки. Результаты наблюдений или экспериментов при этом и представляют собой статистическую совокупность, которую записывают, согласно методам математической статистики, в виде ряда  $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$  числовых значений интересующего нас признака (случайной величины  $X$ ). При этом *точность* статистических методов повышается с ростом числа статистических данных  $n$ .

Основой статистической обработки большого объема информации является так называемый *выборочный метод*. Пусть  $A$  – некоторое множество (например, множество работников конкретного предприятия),  $B \in A$  – случайно выбранное подмножество (например, множество случайно выбранных служащих, при этом некий наблюдатель измерил у них производительность). Выборочным методом называется метод исследования общих свойств множества  $A$  на основе изучения так называемых *статистических свойств* лишь множества  $B$ . Множество  $A$  называется *генеральной совокупностью*, а множество  $B$  – *выборочной совокупностью* или *выборкой* [1]. Число  $N = |A|$  элементов множества  $A$  называется *объемом генеральной совокупности*, а число  $n = |B|$  – *объемом выборки*. При изучении некоторого признака  $X$  (в нашем примере – производительности) выборки производят *испытания* или *наблюдения*.

Пусть в результате независимых испытаний, проведенных в одинаковых условиях, получены числовые значения  $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ , где  $n$  – объем выборки. При обработке статистических данных строятся так называемые *статистики*.

Статистика – это функция от результатов наблюдений [2].

Более развернутое определение указано у Губина В.И. [1]: *статистикой* называется функция  $R^n \xrightarrow{f} R, (x_{(1)}, x_{(2)}, \dots, x_{(n)}) \rightarrow f(x)$ , которая набору значений

$(x_{(1)}, x_{(2)}, \dots, x_{(n)})$  случайной величины ставит в соответствие по некоторому правилу  $f$  действительное число. Статистика является числовой функцией на множестве реализаций случайной величины. Значения  $x_{(i)}$  располагают в порядке возрастания:  $x_1, x_2, \dots, x_n$  ( $x_1 < x_2 < \dots < x_n$ ).

Чаще всего некоторые варианты  $x_i$  в выборке встречаются несколько раз. Число  $n_i$ , показывающее, сколько раз встречается варианта  $x_i$  в выборочной совокупности, называется ее *частотой* (эмпирической частотой).

Частоты вариантов называются их *весами*. Отношение  $w_i = n_i/n$  частоты  $n_i$  к объему  $n$  выборки называют *относительной частотой* (частотью) варианты  $x_i$ .

*Статистическим распределением* (или *вариационным рядом*) называется ранжированный в порядке возрастания или убывания ряд вариантов с соответствующими им частотами.

Различают дискретные и непрерывные вариационные ряды. Дискретный вариационный ряд записывают в виде табл. 1.

Таблица 1

<i>Варианты <math>x_i</math></i>	$x_1$	$x_2$	....	$x_k$
<i>Частоты <math>n_i</math></i>	$n_1$	$n_2$	...	$n_k$

Здесь  $n_i$  – частота появления значения  $x_i$ , причем  $\sum_{i=1}^n n_i = n$ .

Если объем  $n$  выборки большой ( $n > 30$ ), то результаты наблюдений сводят в *интервальный вариационный ряд*, который формируется следующим образом. Вычисляют *размах  $R$  варьирования признака  $X$*  как разность между наибольшим  $x_{\max}$  и наименьшим  $x_{\min}$  значениями признака:  $R = x_{\max} - x_{\min}$ .

Размах  $R$  варьирования признака  $X$  делится на  $k$  равных частей, и таким образом определяется число столбцов (интервалов) в таблице. Согласно [1] число  $k$  частичных интервалов выбирают, пользуясь одним из следующих правил: 1)  $6 \leq k \leq 20$ , 2)  $k \approx \sqrt{n}$ , 3)  $k \approx 1 + \log_2 n \approx 1 + 3,221 \lg n$ .

При небольшом объеме  $n$  выборки число  $k$  интервалов принимают равным от 6 до 10. Длина  $h$  каждого частичного интервала определяется по

формуле:  $h = R / k$ . Величину  $h$  обычно округляют. Затем подсчитывается частота  $n_i$ , с которой попадают значения  $x_i$  признака  $X$  в  $i$ -й интервал.

Значение  $x_i$ , которое попадает на границу интервала, относят к какому-либо определенному концу, например, к левому. За начало  $x_0$  первого интервала рекомендуется брать величину  $x_0 = x_{\min} - 0,5h$ . Конец  $x_k$  последнего интервала находят по формуле  $x_k = x_{\max} + 0,5h$ . Сформированный интервальный вариационный ряд записывают в виде табл. 2.

Таблица 2

Варианты-интервалы $(x_{i-1}; x_i)$	$(x_0; x_1)$	$(x_1; x_2)$	...	$(x_{k-1}; x_k)$
Частоты $n_i$	$n_1$	$n_2$	...	$n_k$

Интервальный вариационный ряд изображают в виде гистограммы частот  $n_i$  или гистограммы относительных частот  $w_i = n_i / n$ .

*Гистограмма* – столбчатая диаграмма, один из видов графического представления эмпирического распределения [2].

*Гистограммой* называется ступенчатая фигура, для построения которой по оси абсцисс откладывают отрезки, изображающие частичные интервалы  $(x_{i-1}; x_i)$  варьирования признака  $X$ , и на этих отрезках, как на основаниях, строят прямоугольники с высотами, равными частотам или частостям соответствующих интервалов [1].

Для расчета статистик (выборочной средней, выборочной дисперсии, асимметрии и эксцесса) переходят от интервального вариационного ряда к дискретному. В качестве вариант  $x_i$  этого ряда берут середины интервалов  $(x_i; x_{i+1})$ .

*Дискретный вариационный ряд* записывается в виде табл. 3 или табл. 4.

Таблица 3

Варианты $x_i$	$x_1$	$x_2$	....	$x_k$
Частоты $n_i$	$n_1$	$n_2$	...	$n_k$

Здесь  $\sum n_i = n$  – объем выборки.

Таблица 4

Варианты $x_i$	$x_1$	$x_2$	....	$x_k$
Относительная частота $\omega_i=n_i/n$	$\omega_1$	$\omega_2$	...	$\omega_k$

Здесь  $\sum_{i=1}^k \omega_i = 1$ .

Графически дискретный вариационный ряд изображают в виде *полигона частот* (соответственно в виде полигона относительных частот) следующим образом. Сначала на числовой плоскости строят точки  $(x_i; n_i)$  (точки  $(x_i; w_i)$ ), где  $x_i$  –  $i$ -я варианта, число  $n_i$  (число  $w_i$ ) называют *частотой* (*частостью*). Затем строят ломаную, соединяющую построенные точки, которую и называют *полигоном*.

Вариационные ряды графически можно изобразить в виде *кумулятивной кривой* (*кривой сумм – кумуляты*). При построении кумуляты дискретного вариационного ряда на оси абсцисс откладывают варианты  $x_i$ , а по оси ординат – соответствующие им *накопленные частоты*  $W_i$ . Соединяя точки  $(x_i; W_i)$  отрезками, получаем ломаную, которую называют *кумулятой*. Для получения накопленных частот и дальнейшего построения точек  $(x_i; W_i)$  составляется расчетная табл. 5.

Таблица 5

Варианты $x_i$	$x_1$	$x_2$	....	$x_k$
Относительная частота, $\omega_i=n_i/n$	$\omega_1=n_1/n$	$\omega_2=n_2/n$	...	$\omega_k=n_k/n$
Накопленные относительные частоты, $W_i=W_{i-1}+ \omega_i$	$W_1= \omega_1$	$W_2= W_1+\omega_2$	...	$W_k= W_{k-1}+ \omega_k$

При построении кумуляты интервального вариационного ряда левому концу первого интервала сопоставляется частота, равная нулю, а правому – частота этого интервала. Правому концу второго интервала соответствует накопленная частота первых двух интервалов, то есть сумма частот этих интервалов и т.д. Правая граница последнего интервала равна сумме всех частот, то есть объему  $n$  выборки.



Для характеристики свойств статистического распределения в математической статистике вводится понятие эмпирической функции распределения.

Эмпирической функцией распределения или функцией распределения называется функция  $F_B(x)$ , определяемая равенством:

$$F_B(x) = \frac{n_x}{n}; \quad (1)$$

где  $n$  – объем выборки;  $n_x$  – число вариант  $x_i$ , меньших  $x$ .

Эмпирическая функция  $F_B(x)$  служит для оценки теоретической функции распределения генеральной совокупности [1]. Различие между ними состоит в том, что теоретическая функция  $F(x)$  определяет вероятность события  $X < x$ , а эмпирическая функция  $F_B(x)$  определяет относительную частоту этого события. Из теоремы Бернулли следует, что при больших  $n$  числа  $F_B(x)$  и  $F(x)$  мало отличаются одно от другого в том смысле, что  $\lim_{n \rightarrow \infty} P[|F(x) - F_B(x)| < \varepsilon] = 1$ ,  $\varepsilon > 0$ . Таким образом, при больших объемах выборки  $n$ , согласно закону больших чисел, функция  $F_B(x)$  сходится по вероятности к теоретической функции  $F(x)$  признака  $X$ .

Аналогом этой функции в теории вероятностей является интегральная функция распределения  $F(x)$ . Функция  $F_B(x)$  отличается от функции  $F(x)$  тем, что вместо вероятности  $P(X < x)$  берется накопленная  $W_i = \frac{n_x}{n}$ . Значение функции  $F_B(x)$  равно числу  $n_x$  принятых признаком  $X$  вариант, меньших  $x$ , деленному на объем  $n$  выборки.

Эмпирическая функция  $F_B(x)$  служит для оценки теоретической функции распределения генеральной совокупности. Значения эмпирической функции  $F_B(x)$  принадлежат промежутку  $[0;1]$ ; ее графиком служит кусочно-постоянная кривая (рис. 1).

Она имеет скачки в точках, которые соответствуют вариантам  $x_i$ . При обработке результатов эксперимента, например, результатов механических испытаний, целесообразно вместо ступенчатой кривой вычерчивать плавную

кривую, которая проходит через точки, расположенные посередине вертикальных частей ступенчатой кривой.

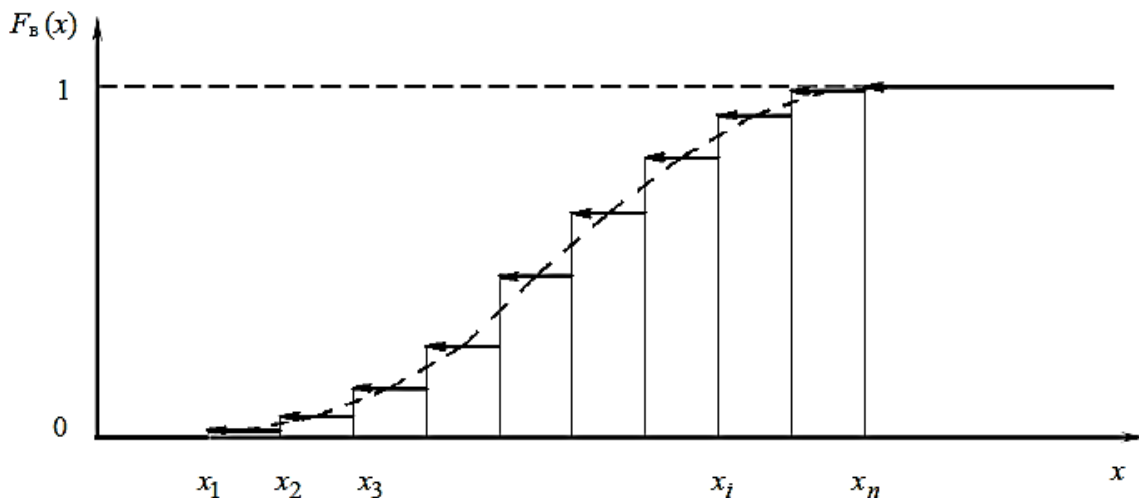


Рис.1. Кумулята и эмпирическая функция распределения

Абсциссами этих точек служат значения механической характеристики  $x_i$ , а ординатами – эмпирическая функция  $F_B(x)$ , характеризующая оценку вероятности события  $X \in x_i$ .

## 1.2. Расчет выборочных характеристик статистического распределения

Пусть дана выборка объема  $n$  со значениями  $x_1, x_2, \dots, x_n$  признака  $X$ . Для характеристики важнейших свойств статистического распределения используют средние показатели, называемые *выборочными числовыми характеристиками*. Если значения  $x_i$  признака  $X$  не сгруппированы в вариационные ряды (табл. 2, 3, 4) и объем выборки  $n$  небольшой, то оценки для неизвестных математического ожидания  $a$  и дисперсии  $s$  находят по формулам:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

для математического ожидания и

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (3)$$

для дисперсии [3].

Если результаты наблюдений сгруппированы в дискретный вариационный ряд (табл. 3), то те же оценки находят по формулам:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i, n = \sum_{i=1}^k n_i, \quad (4)$$

$$S^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2. \quad (5)$$

Формулы (2) и (4), как (3) и (5), дают одинаковые результаты соответственно для  $\bar{x}$  и  $S^2$ .

По формуле (5) вычисляют  $S^2$  в случае, если объем выборки  $n \geq 50$ .

Если же  $n < 50$ , то вычисляют исправленную дисперсию  $\hat{s}^2$  по формуле:

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 \quad (6)$$

для простой выборки или

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 n_i \quad (7)$$

для взвешенной выборки.

Выборочное среднее квадратическое отклонение находят по формулам

$$s = \sqrt{S^2} \quad \text{или} \quad \hat{s} = \sqrt{\hat{s}^2} \quad (8)$$

при различных объемах выборки.

Для анализа вариационных рядов вычисляют такие статистики, как *моду* и *медиану*.

*Модой*  $M_o(X)$  называют варианту, которая имеет наибольшую частоту.

Например, для вариационного ряда

$x_i$	4	9	14	19
$n_i$	3	7	2	5

мода равна  $M_o(X)=9$ .

*Медианой*  $M_e(X)$  называют варианту, которая делит вариационный ряд на равные по числу вариант части [4].

При нечетном объеме выборки  $n = 2k + 1$  медиана равна  $M_e(X) = x_{k+1}$ .

Например, для вариационного ряда

$x_i$	3	5	8	12	15
$n_i$	6	2	4	5	8

медиана равна  $M_e(X) = x_{13} = 12$ .

При четном объеме выборки  $n = 2k$  медиана находится по формуле:

$$M_e(x) = \frac{x_k + x_{k+1}}{2}. \quad (9)$$

Здесь  $x_k$  – варианта, которая находится слева от середины вариационного ряда, а  $x_{k+1}$  – справа от нее. Например, для следующего вариационного ряда:

$x_i$	2	5	7	10	12	14
$n_i$	3	4	8	2	3	6

медиана равна  $M_e(X) = 7$ .

Для вычисления выборочной средней  $\bar{x}$  и выборочной дисперсии  $S^2$ , асимметрии  $A_s$  и эксцесса  $E_x$  при достаточно большом объеме выборки ( $n > 30$ ) применяют метод произведений [5]. При этом вводят условные варианты  $u_i$ , которые вычисляют по формуле:

$$u_i = \frac{x_i - C}{h}, \quad (10)$$

где  $C = M_o(X)$ ,  $h$  – шаг (длина интервала).

Составляется расчетная табл. 6.

Таблица 6

$x_i$	$n_i$	$u_i$	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	Контрольный столбец $n_i(u_i + 1)^2$
$x_1$	$n_1$	$u_1$	$n_1 u_1$	...			
$x_2$	$n_2$	$u_2$	$n_2 u_2$	...			
....	...	...	...	...			
$x_n$	$n_n$	$u_n$	$n_n u_n$	...			
Строка сумм	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$

Контроль вычислений ведут по формуле:

$$\Sigma n_i + 2\Sigma n_i u_i + \Sigma n_i u_i^2 = \Sigma n_i (u_i + 1)^2.$$

Пользуясь табл. 6, вычисляют, согласно [5], условные начальные моменты по формулам:

$$M_1^* = \frac{1}{n} \sum n_i u_i, \quad (11)$$

$$M_2^* = \frac{1}{n} \sum n_i u_i^2, \quad (12)$$

$$M_3^* = \frac{1}{n} \sum n_i u_i^3, \quad (13)$$

$$M_4^* = \frac{1}{n} \sum n_i u_i^4 \quad (14)$$

Тогда выборочную среднюю находят по формуле:

$$\bar{x} = M_1^* h + C. \quad (15)$$

Выборочную дисперсию находят по формуле:

$$S^2 = (M_2^* - M_1^{*2}) h^2. \quad (16)$$

Выборочное среднее квадратическое отклонение находят по формуле:

$$S = \sqrt{S^2}. \quad (17)$$

Асимметрию и эксцесс находят по формулам:

$$A_s = \frac{m_3}{S^3}, \quad (18)$$

$$E_x = \frac{m_4}{S^4} - 3, \quad (19)$$

$$\text{где } m_3 = (M_3^* - 3M_2^*M_1^* + 2M_1^{*3})h^3 \quad (20)$$

– условный центральный момент третьего порядка;

$$m_4 = (M_4^* - 4M_3^*M_1^* + 6M_2^*M_1^{*2} - 3M_1^{*4})h^4, \quad (21)$$

– условный центральный момент четвертого порядка.

Для характеристики *колеблемости* признака  $X$  используют [4] относительный показатель – коэффициент вариации  $V$ , который для положительной случайной величины  $X$  вычисляют по формуле:

$$V = \frac{S}{\bar{x}}. \quad (22)$$

### 1.3. Доверительные интервальные оценки параметров распределения

Выборочные характеристики  $\bar{x}$  и  $S^2$  являются надежными количественными оценками генеральных характеристик  $a$  и  $\sigma^2$  только при большом объеме выборки. При ограниченных объемах выборки возникает необходимость указать степень точности и надежности оценок генеральных характеристик. При решении практических задач, связанных со статистическим анализом характеристик изучаемого признака  $X$ , например, механических свойств конструкционных материалов, несущей способности элементов конструкций, пропускной способности нефтегазопроводов, себестоимости единицы производимой продукции и т.д., как правило, значения генеральной дисперсии и математического ожидания неизвестны.

Для оценки генеральной средней  $M(X)=a$  и генерального среднеквадратического отклонения  $\sigma$  по выборочной средней  $\bar{x}$  и выборочному среднеквадратическому отклонению  $S$  находят, согласно [6], доверительные интервалы по формулам:

$$\bar{x} - \frac{S}{\sqrt{n}} t_{\gamma} < a < \bar{x} + \frac{S}{\sqrt{n}} t_{\gamma}, \quad (23)$$

где  $t_{\gamma}$  находят из таблицы (см. прил. 3) по заданным  $n$  и  $\gamma$  ( $\gamma$  – уровень доверия или надежность, которая задается заранее).

Для генерального среднего квадратического отклонения доверительные интервалы находят [6] по формулам:

$$S(1-q) < \sigma < S(1+q) \quad (\text{при } q < 1) \quad (24)$$

или

$$0 < \sigma < S(1+q) \quad (\text{при } (q > 1)). \quad (25)$$

Величину  $q$  находят по таблице значений  $q=(\gamma, n)$  (прил. 4) по заданным  $n$  и  $\gamma$ .

## 1.4. Лабораторная работа №1

### Первичная обработка экспериментальных данных

*Цель работы.* Овладение способами построения вариационных рядов распределения и методами расчета основных числовых характеристик.

#### *Задание к лабораторной работе №1*

1. Для предложенных данных построить интервальный и дискретный вариационные ряды:

а) построить интервальный вариационный ряд. Для этого найти:

– размах варьирования  $R=x_{max}-x_{min}$ , где  $x_{max}$  и  $x_{min}$  – наибольшая и наименьшая варианты в данной выборочной совокупности соответственно;

– разбить ряд на интервалы, пользуясь одним из приведенных соотношений:  $k \approx \sqrt{n}$ ,  $6 < k < 12$ ,  $k \approx 1 + \log_2 n \approx 1 + 3,2 \lg n$ , где  $n$  – объем выборки;

– найти длину  $h$  частичных интервалов по формуле  $h = \frac{R}{k}$ , если необходимо найти заданную длину, округлив это значение до четырех знаков после запятой;

– создать и заполнить свою таблицу интервального вариационного ряда, взяв за образец табл. 2;

б) построить дискретный вариационный ряд (см. табл. 3), взяв в качестве вариант середины вариант-интервалов интервального вариационного ряда;

в) построить полигон и гистограмму частот.

2. Построить кумуляту (график накопленных частот), т.е. ломанную, проходящую через точки с координатами  $x_i$  с соответствующими накопленными частотами, предварительно составив табл. 5.

3. Найти эмпирическую функцию распределения и изобразить ее графически.

4. Вычислить основные числовые характеристики вариационного ряда:

а) найти моду  $M_0(x)$  и медиану  $M_e(x)$ ;

б) применяя метод произведений и используя формулу (10), заполнить табл. 6;

в) найти выборочную среднюю по формуле (15) и выборочную дисперсию по формуле (16). Для этого вычислить начальные моменты (формулы (11) (14));

г) вычислить выборочное среднее квадратическое отклонение – формула (17);

д) найти коэффициент вариации, воспользовавшись формулой (22);

е) вычислить асимметрию (18) и эксцесс (19), предварительно найдя центральные моменты (20) и (21).

5. Определить доверительные интервалы (23) и (24) для оценки генеральной средней  $\mu$  и генерального среднее квадратического отклонения  $\sigma$ .

6. Раскрыть смысловую сторону каждой характеристики. Сделать вывод.



### 1.4.1. Пример выполнения лабораторной работы №1

*Задача.* Имеются данные об обводненности нефти из насосных скважин (в %):

61,2	61,4	60,2	61,2	61,3	60,4	61,4	60,8	61,2	60,6
61,6	60,2	61,3	60,3	60,7	60,9	61,2	60,5	61,0	61,4
61,1	60,9	61,5	61,4	60,6	61,2	60,1	61,3	61,1	61,3
60,3	61,3	60,6	61,7	60,6	61,2	60,8	61,3	61,0	61,2
60,5	61,4	60,7	61,3	60,9	61,2	61,1	61,3	60,9	61,4
60,7	61,2	60,3	61,1	61,0	61,5	61,3	61,9	61,4	63,3
61,6	61,0	61,7	61,1	60,9	61,5	61,6	61,4	61,5	61,2
61,6	61,3	61,8	61,1	61,7	60,9	62,2	61,1	62,1	61,0
61,5	61,7	62,3	62,3	61,7	62,8	62,5	62,8	62,6	61,5
62,1	62,6	61,6	62,5	62,4	62,3	62,1	62,3	62,2	62,1

1. Обозначим через  $X$  обводненность нефти из насосных скважин.

Для предложенных данных строим интервальный ряд распределения:

а) Ранжируем вариационный ряд по возрастанию или по убыванию для удобства дальнейшей обработки. Если при этом обработку информации производите в электронных таблицах, например в MS Excel, то ранжирование можно произвести при помощи кнопки «Сортировка по возрастанию» во вкладке «Данные», выделив нужный диапазон для сортировки - столбец или строку;

–  $x_{max} = 62,8$  и  $x_{min} = 60,1$ , то размах варьирования признака  $X$  равен  $R = x_{max} - x_{min} = 62,8 - 60,1 = 2,7$ . Вручную считать размах так же нет необходимости, можно использовать для этого строку функций. Для этого в выбранной для вычисления размаха ячейке нажмите «=» и в появившемся окошечке выберите последнюю из ранжированных ячеек, знак «-» и первую, например «=A51-A1» и клавишу «Enter»;

– разобьем ряд на интервалы, положим  $k=10$ ;

– длину  $h$  каждого частичного интервала рассчитаем по формуле

$$h = \frac{R}{k} = \frac{2,7}{10} = 0,27;$$

– подсчитаем число вариантов, попадающих в каждый интервал по данным выборки. Значение  $x_i$ , попадающее на границу интервала, относим к правому интервалу. За начало  $x_0$  первого интервала берем величину  $x_0$ :

$x_0 = x_{min} - 0,5 h = 60,1 - 0,15 \cdot 0,3 = 59,95 \approx 60$ . Конец  $x_k$  последнего интервала находим  $x_k = x_{min} + 0,5 h = 62,5 + 0,15 \cdot 0,3 = 62,95 \approx 63,0$ . Сформированный интервальный вариационный ряд записываем в виде табл. 7.

Таблица 7

Варианты-интервалы	60- 60,3	60,3- 60,6	60,6- 60,9	60,9- 61,2	61,2- 61,5	61,5- 61,8	61,8- 62,1	62,1- 62,4	62,4- 62,7	62,7- 63,0
Частоты, $n_i$	3	6	9	18	29	16	2	10	5	2

Контроль  $\sum n_i = 100$ .

б) Записываем дискретный вариационный ряд (табл. 8). В качестве вариант  $x_i$  берем середины интервалов интервального вариационного ряда. Середину каждого интервала вычисляем в таблице автоматически, например, в первом интервале « $=(60+60,3)/2$ ».

Таблица 8

Варианты	60,15	60,45	60,75	60,05	61,35	61,65	61,95	62,25	62,55	62,85
Частоты, $n_i$	3	6	9	18	29	16	2	10	5	2

в) Строим гистограмму и полигон частот (рис. 2.). Для этого, выделив нужные ячейки во вкладке меню «Вставка», выбираем «Гистограмма». Для того, чтобы в одной системе координат у Вас был построен и полигон, нужно, выделив область графика добавить данные ряда из табл.8, а затем в получившейся диаграмме изменить вид диаграммы на «Точечную».



Рис. 2. Гистограмма и полигон частот

2. Строим график накопленных частот – кумуляту (рис.3).

Предварительно составляем расчетную табл. 9.

3. Для нахождения функции распределения строим табл.9.

Таблица 9

Варианты	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
Относительная частота, $\omega_i = n_i/n$	0,03	0,06	0,09	0,18	0,29	0,16	0,02	0,1	0,05	0,02
Накопленные относительные частоты, $W_i = W_{i-1} + \omega_i$	0,03	0,09	0,18	0,36	0,65	0,81	0,83	0,93	0,98	1

Построение кумуляты производится по данным третьей строки табл.9, процесс построения аналогичен построению полигона.

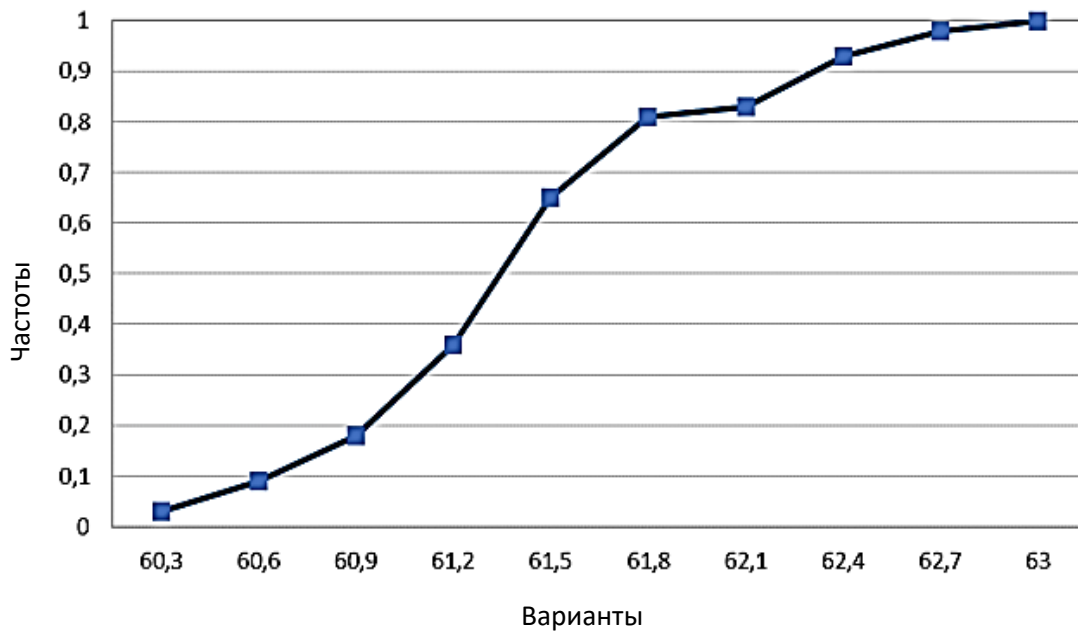


Рис.3. Кумулята

4. Находим эмпирическую функцию распределения по формуле (1).

Если  $x \leq 60,15$ , то  $F_B(x) = 0$ , по свойству эмпирической функции распределения.

$$\text{Если } 60,15 < x \leq 60,45, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3}{100} = 0,03.$$

$$\text{Если } 60,55 < x \leq 60,75, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6}{100} = 0,09.$$

$$\text{Если } 60,75 < x \leq 61,05, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9}{100} = 0,18.$$

$$\text{Если } 61,05 < x \leq 61,35, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18}{100} = 0,36.$$

$$\text{Если } 61,35 < x \leq 61,65, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18+29}{100} = 0,65.$$

$$\text{Если } 61,65 < x \leq 61,95, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18+29+16}{100} = 0,81.$$

$$\text{Если } 61,95 < x \leq 62,25, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18+29+16+2}{100} = 0,83.$$

$$\text{Если } 62,25 < x \leq 62,55, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18+29+16+2+10}{100} = 0,93.$$

$$\text{Если } 62,55 < x \leq 62,85, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9+18+29+16+2+10+5}{100} = 0,98.$$

Если  $x > 62,85$ , то  $F_B(x) = 1$ , по свойству эмпирической функции распределения.

$$F_B(x) = \begin{cases} 0, & x \in (-\infty; 60,15], \\ 0,03, & x \in (60,15; 60,45], \\ 0,09, & x \in (60,45; 60,75], \\ 0,18, & x \in (60,75; 61,05], \\ 0,36, & x \in (61,05; 61,35], \\ 0,65, & x \in (61,35; 61,65], \\ 0,81, & x \in (61,65; 61,95], \\ 0,83, & x \in (61,95; 62,25], \\ 0,93, & x \in (62,25; 62,55], \\ 0,98, & x \in (62,55; 62,85], \\ 1, & x \in (62,85; +\infty). \end{cases}$$

График функции  $F_B(x)$  представлен на рис. 1.

Соединив середины вертикальных частей ступенчатой кусочно-постоянной кривой, являющейся графиком функции  $F_B(x)$ , получаем плавную кривую.

5. Вычислим основные числовые характеристики вариационного ряда.

Составим табл. 10.

а) Так как варианта  $x = 61,35$  в табл. 8 встречается с наибольшей частотой  $n_5 = 29$ , то  $M_0(x) = 61,35$ .

Находим медиану  $M_e(x)$ . Так как табл. 8 содержит четное число столбцов, то  $M_e(x) = (61,35 + 61,65) / 2 = 61,5$ . Это значение делит данные выборки признака  $X$  на равные части.

б) Для нахождения других статистик применим метод произведений.

Введем, согласно (9), условные варианты  $u_i = (\bar{x}_i - M_0(x)) / h$ ;  $h = 0,3$ .

Таблица 10

$x_i$	$n_i$	$u_i$	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	Контр.столбец $n_i(u_i + 1)^2$
60,15	3	-4	-12	48	-192	768	27
60,45	6	-3	-18	54	-162	486	24
60,75	9	-2	-18	36	-72	144	9
61,05	18	-1	-18	18	-18	18	0
61,35	29	0	0	0	0	0	29
61,65	16	1	16	16	16	16	64
61,95	2	2	4	8	16	32	18
62,25	10	3	30	90	270	810	160
62,55	5	4	20	80	320	1280	125
62,85	2	5	10	50	250	1250	72
	100		14	400	428	4804	528

Контроль вычислений проводим по формуле:

$$\sum n_i + 2\sum n_i u_i + \sum n_i u_i^2 = \sum n_i (u_i + 1)^2.$$

То есть  $100+28+400=528=27+24+9+29+64+18+160+125+72$ .

Следовательно, вычисления произведены верно.

в) пользуясь результатами табл. 10, находим условные начальные моменты (11-14):

$$M_1^* = \frac{1}{n} \sum n_i u_i = 14/100=0,14,$$

$$M_2^* = \frac{1}{n} \sum n_i u_i^2 = \frac{400}{100} = 4,$$

$$M_3^* = \frac{1}{n} \sum n_i u_i^3 = 428/100=4,28,$$

$$M_4^* = \frac{1}{n} \sum n_i u_i^4 = \frac{4804}{100} = 48,04.$$

Находим выборочную среднюю (15):

$$\bar{x} = M_1^* h + Mo(x) = 0,14 \cdot 0,3 + 61,35 = 61,392.$$

Находим выборочную дисперсию (16):

$$S^2 = (M_2^* - M_1^{*2})h^2 = (M_2^* - M_1^{*2})h^2;$$

г) Вычисляем среднее квадратическое отклонение (17):

$$S = \sqrt{S^2} = 0,598553.$$

д) Величина  $S$  характеризует степень рассеивания значений обводненности нефти относительно средней обводненности. Для определения колеблемости значений обводненности нефти в процентном отношении вычисляем коэффициент вариации (22):

$$V = \frac{S}{\bar{x}} = 0,599/61,39=0,00976.$$

Так как величина коэффициента вариации мала, то это означает тесную сгруппированность обводненности нефти около центра рассеивания.

е) Для предварительной оценки отклонения значений выборки от нормального распределения вычисляем асимметрию и эксцесс. Для этого сначала вычисляем центральные моменты третьего и четвертого порядков (20), (21):

$$m_3 = (M_3^* - 3M_2^*M_1^* + 2M_1^{*3})h^3 =$$

$$= (4,28 - 3 \cdot 4 \cdot 0,14 + 2 \cdot 0,14^3) \cdot 0,3^3 = 0,07035.$$

$$\begin{aligned} m_4 &= (M_4^* - 4M_3^*M_1^* + 6M_2^*M_1^{*2} - 3M_1^{*4})h^4 = \\ &= (48,04 - 4 \cdot 4,28 \cdot 0,14 + 6 \cdot 4 \cdot 0,14^2 - 3 \cdot 0,14^4) \cdot 0,3^4 = 0,37023. \end{aligned}$$

В соответствии с (18), (19):

$$A_s = \frac{m_3}{S^4} = \frac{0,07}{0,599^3} = 0,328.$$

$$E_x = \frac{m_4}{S^4} - 3 = \frac{0,37}{0,599^4} - 3 = -0,115.$$

6. Произведем оценку генеральной средней  $M(X)=a$  и генерального среднего квадратического отклонения  $S=\sigma$  по выборочным статистикам, используя теорию доверительных интервалов.

По таблице прил. 3 при  $n=100$  и  $\gamma=0,95$  находим  $t_\gamma=1,984$ .

Доверительный интервал можно записать так:

$$\begin{aligned} 61,39 - \frac{0,599}{10} 1,984 < a < 61,39 + \frac{0,599}{10} 1,984, \\ 61,27 < a < 61,51. \end{aligned}$$

Таким образом, средняя обводненность нефти по данным выборки должна находиться в промежутке (61,27; 61,51).

Доверительный интервал для генерального среднеквадратического отклонения  $S=\sigma$ . По таблице прил. 4 при  $n=100$  и  $\gamma=0,95$  находим  $q=0,143$ . Так как  $q < 1$ , то доверительный интервал записываем в виде (24):

$$\begin{aligned} S(1-q) < \sigma < S(1+q), \\ 0,599(1-0,143) < \sigma < 0,599(1+0,143), \\ 0,51 < \sigma < 0,68. \end{aligned}$$

Следовательно, отклонения истинных значений обводненности нефти не должны выходить за пределы промежутка (0,51; 0,68).

## 1.4.2. Варианты заданий лабораторной работы № 1

### Вариант № 1

Имеются данные о производительности труда (количество деталей в смену):

71	76	79	86	78	76	84	78
76	99	87	82	78	84	81	76
82	76	79	76	86	86	86	89
80	77	80	75	85	75	88	83
85	87	84	81	85	93	85	79
74	85	85	75	74	83	74	88
74	74	93	95	77	86	79	82
80	84	86	91	76	88	95	92

### Вариант № 2

Имеются данные о пропускной способности 50 участков нефтепровода (м<sup>3</sup>/сут):

19,8	19,1	19,3	18,8	20,2	20,8	20,7	19,7	19,6	19,2
20,9	20,9	20,2	19,6	20,4	20,2	20,4	20,4	18,9	19,7
19,8	20,6	20,7	19,7	20,3	19,8	20,4	20,3	20,6	20,5
20,5	20,4	20,6	21,0	20,4	20,4	20,3	19,7	19,9	20,1
20,4	20,5	20,3	20,5	20,2	20,5	20,7	21,0	20,4	20,8

### Вариант № 3

Имеются данные о суточной добыче нефти в одном из районов страны (в тоннах):

85	76	90	91	84	74	77	83	86	93
84	85	89	89	77	80	83	89	78	90
88	75	85	87	82	84	89	94	81	83
91	82	91	74	83	91	89	96	88	79
89	86	87	76	84	93	91	88	92	86
80	89	81	88	90	90	90	95	94	90
84	88	78	81	87	87	88	99	86	79
89	95	89	90	98	93	84	96	92	88
87	85	95	87	85	88	97	91	88	90
88	84	85	88	86	88	81	92	90	82



#### Вариант № 4

Имеются данные о вводе в эксплуатацию новых газовых скважин за год по различным газодобывающим районам страны:

52	33	10	22	28	34	39	29	57	7
29	21	51	11	31	17	31	50	28	32
27	28	44	29	31	32	42	55	23	18
31	40	32	38	24	47	15	37	20	29
12	46	16	44	9	41	21	19	45	25

#### Вариант № 5

Имеются энергетические затраты на 1 метр проходки при эксплуатационном бурении нефтяных скважин в различных нефтеносных районах страны (руб.):

14	13	18	15	12	13	14
13	16	16	15	12	13	13
16	18	13	15	14	15	14
15	12	13	12	14	16	12
12	14	13	13	15	15	13
15	15	15	15	17	17	14
15	18	15	12	15	13	13

#### Вариант № 6

Имеются данные о суточном дебите газа в наблюдаемой скважине ( $\text{м}^3/\text{сут}$ ):

30	19	21	28	27	29	31	24	25	30	25	18	28	33	21	30	34	26	25	32	20	26	20
28	28	32	34	26	24	19	27	23	29	18	24	31	18	26	32	29	23	27	23	21	22	27

#### Вариант № 7

Имеются данные о себестоимости 1 тонны нефти и нефтяного попутного газа (тыс. руб.):

0,3	0,4	0,8	0,7	1,0	0,9	1,0	0,6	1,0	1,0	1,1
1,2	1,4	1,9	1,3	0,5	1,2	1,3	0,5	1,2	1,0	1,4
1,6	1,1	1,1	1,8	1,3	1,5	0,9	1,3	1,1	0,9	1,1
0,5	0,6	1,1	0,8	1,6	0,7	1,5	1,1	1,0	1,5	1,2
1,2	0,3	1,4	0,9	2,7	0,8	1,2	1,8	1,2	1,3	1,3

### Вариант № 8

Имеются данные о числе рабочих дней без простоя для пятидесяти пяти цехов нескольких предприятий одного из регионов страны:

261	260	258	257	263	260	259	264	261	260	264
261	265	261	260	263	260	259	259	262	259	262
260	258	265	259	262	259	261	262	259	262	261
259	265	261	259	264	263	266	261	262	266	262
260	259	258	258	258	266	262	266	261	259	259

### Вариант № 9

Приведено количество деталей, выработанных за смену различными рабочими:

75	88	74	80	76	82	86	76	93	74	82
72	82	71	82	87	81	87	79	78	87	93
82	87	82	74	77	83	86	85	86	76	76
81	86	76	71	80	85	73	75	88	89	83
84	85	81	83	87	76	76	78	87	80	80
85	82	85	76	87	73	78	72	83	91	92

### Вариант № 10

Имеются данные о рабочих дебитах газовой скважины (тыс. м<sup>3</sup>/сут):

550	538	552	550	545	560	549	549
550	551	546	557	552	554	539	539
552	547	557	547	558	563	562	569
568	552	559	552	554	547	554	567
550	551	550	551	562	550	562	561
530	542	535	542	534	558	548	541
539	537	556	556	548	560	540	549
543	540	546	556	533	558	551	551

### Вариант № 11

Имеются данные о коэффициенте эксплуатации насосных скважин в различных нефтеносных районах страны:

0,90	0,79	0,84	0,86	0,88	0,90	0,89	0,85	0,91	0,98
0,91	0,80	0,87	0,89	0,88	0,78	0,81	0,85	0,88	0,94
0,86	0,80	0,86	0,91	0,78	0,86	0,91	0,95	0,97	0,88
0,79	0,82	0,84	0,90	0,81	0,87	0,91	0,90	0,82	0,85
0,90	0,82	0,85	0,90	0,96	0,98	0,89	0,87	0,99	0,85

### Вариант № 12

50 сверл были подвергнуты испытанию на твердость. При этом фиксировалась твердость лапки. Результаты испытаний следующие:

14,5	14,6	15,1	15,5	16,3	16,8	17,9	16,3	14,5	14,9
13,6	15,4	16,9	15,4	14,3	15,5	11,3	15,5	17,1	16,8
12,2	15,2	15,7	16,9	15,7	17,7	16,6	16,2	15,5	14,2
15,5	16,1	14,5	17,9	16,9	11,7	13,2	14,9	12,8	19,8
14,3	16,5	11,6	17,8	16,6	17,9	14,9	15,2	17,3	16,9

### Вариант № 13

Даны значения обследуемого признака  $X$  – себестоимости единицы продукции (в руб.):

88	73	79	89	75	87	71	83	82	95
84	83	93	92	83	89	75	87	88	91
77	91	84	96	84	79	79	83	92	80
87	81	83	80	82	76	81	90	78	81
79	92	86	84	77	92	93	91	85	84
75	91	83	82	84	85	78	85	85	95
73	77	78	86	85	88	81	79	76	76
88	76	78	87	86	74	87	71	76	90
77	90	84	83	84	77	85	88	82	79
75	88	79	79	73	74	76	83	73	89

### Вариант № 14

Имеются данные о суточном дебите газа в наблюдаемой скважине:

39	19	21	28	26	27	29	28	28	27
23	26	32	34	26	24	22	19	23	27
30	29	25	18	18	25	20	22	24	28
31	33	23	18	21	26	30	32	34	29
26	21	20	23	25	27	30	32	29	27

### Вариант № 15

Даны замеры толщины резца (в мм):

24,5	26,8	23,6	25,5	22,2	26,9	25,3	24,1	28,5	25,3
24,1	28,5	25,3	24,6	27,9	25,4	21,3	25,2	27,7	23,6
25,2	26,8	25,9	25,1	26,3	25,4	21,3	25,2	25,5	25,7
26,6	28,2	25,4	23,2	26,6	25,7	24,3	26,8	25,8	27,1
26,2	25,9	21,6	25,3	25,1	24,8	26,3	24,9	24,3	26,8

### Вариант № 16

Имеются данные о расходах, связанных с монтажом и демонтажем оборудования на предприятии (в тыс. руб.):

4,7	7,2	6,2	6,7	7,2	5,7	7,7	8,2	6,2	5,2
7,2	5,7	6,2	5,7	8,2	5,7	6,7	6,2	5,7	6,2
6,7	5,2	7,7	6,2	7,2	7,7	6,7	7,2	8,2	6,2
5,7	6,2	7,7	6,7	7,2	5,7	6,7	8,2	7,7	8,2
4,7	8,7	4,2	8,7	6,2	6,7	6,2	7,2	4,9	5,5

### Вариант № 17

Даны значения обследуемого признака  $X$  – себестоимости одной детали:

82	83	73	79	89	95	92	93	88	84
76	76	88	81	78	86	84	84	86	85
87	84	74	83	87	73	76	73	78	76
76	74	88	82	73	85	79	77	79	97
84	80	75	81	73	78	83	75	90	83
77	84	85	90	92	91	85	71	85	87
82	94	92	76	93	90	73	92	84	93
88	84	81	93	81	91	78	85	83	96
84	95	79	79	89	82	79	77	83	88
81	88	82	77	92	76	84	83	87	89

### Вариант № 18

Даны значения диаметров шестерен, обрабатываемых на станке:

29	27	29	29	29	33	29	25	25	27
31	29	29	27	33	29	31	29	29	29
35	27	29	35	29	33	21	29	27	29
27	29	31	29	31	29	29	23	31	29
31	29	31	29	31	33	31	31	31	27
23	27	33	29	25	29	19	29	31	23
31	29	27	33	29	31	29	31	23	35
27	29	29	27	29	29	21	35	27	33
29	27	31	31	27	29	29	27	29	25
27	31	37	25	31	27	27	29	25	24

### Вариант № 19

Даны значения внутреннего диаметра гайки (в мм):

4,25	4,38	4,48	4,53	4,54	4,41	4,52	4,39	4,16	4,27
4,59	4,48	4,56	4,13	4,51	4,31	4,27	4,87	4,32	4,49
4,74	4,17	4,66	4,92	4,48	4,68	4,45	4,12	4,69	4,28
4,74	4,55	4,28	4,54	4,51	4,77	4,71	4,78	4,13	4,51
4,42	4,36	4,45	4,32	4,17	4,79	4,13	4,52	4,73	4,95

### Вариант № 20

Даны значения ширины пера круглой плашки (в мм):

3,69	3,56	3,52	3,68	3,49	3,58	3,59	3,54	3,35	3,69
3,87	3,67	3,79	3,75	3,43	3,50	3,57	3,53	3,49	3,68
3,36	3,63	3,51	3,99	3,90	3,53	3,50	3,55	3,40	3,73
3,72	3,53	3,42	3,72	3,68	3,46	3,46	3,36	3,37	3,53
3,48	3,70	3,48	3,68	3,46	3,61	3,57	3,47	3,74	3,47

### Вариант № 21

Имеются данные замеров температуры воздуха в цехе (в °С):

19	29	21	39	25	26	32	25	28	26
36	30	31	29	35	23	32	27	27	26
26	30	27	25	28	28	36	29	35	26
32	29	38	28	25	29	34	28	29	32
34	28	28	29	33	27	34	25	28	26
30	38	39	32	29	29	34	35	32	27
26	25	26	35	36	30	28	33	26	28
26	28	27	33	33	29	32	25	38	26
36	23	24	27	26	34	30	25	24	33

### Вариант № 22

Результаты измерения температуры раздела фракции бензин-авиакеросин на установке первичной переработки нефти (в °С):

133	133	142	135	145	144	145	147	146	134
130	134	138	144	141	141	134	141	136	140
143	139	141	137	140	145	145	141	144	138
139	143	141	141	146	143	140	139	143	143
139	140	139	138	138	135	141	141	140	138
145	135	148	136	139	142	143	143	137	138
138	139	138	144	143	138	142	138	140	140
137	139	140	139	137	136	136	135	135	141
142	136	140	136	137	138	138	137	139	139
140	139	140	140	139	139	139	140	140	146

### Вариант № 23

Имеются данные о суточном дебите нефти наблюдаемой скважины (вт/сут.):

16	13	11	15	18	19	21	18	17	15
13	16	18	17	19	15	13	12	14	16
17	20	17	17	20	19	18	22	24	15
14	10	12	16	18	18	19	21	23	20
22	24	17	16	14	15	18	15	11	16
17	15	13	16	17	18	14	15	19	17
18	16	13	15	17	21	23	26	19	22
24	25	20	21	24	19	22	23	20	25
21	20	22	26	19	22	25	28	23	20
21	27	19	15	22	23	18	22	22	18

### Вариант № 24

Имеются результаты испытания – твердости лапки сверла:

36,8	32,0	39,4	36,3	35,4	37,3	34,7	39,0	28,3	41,3
36,1	37,3	32,2	38,5	34,2	37,2	30,6	37,3	35,2	36,9
34,3	35,2	30,8	36,0	39,3	32,7	34,6	36,8	39,1	29,5
30,4	35,2	36,5	38,2	40,2	36,8	39,3	32,7	37,1	29,3
28,4	40,2	34,8	37,2	32,6	41,0	40,4	28,3	34,8	39,2

### Вариант № 25

Даны значения овальности валика (в мк):

25	29	33	21	29	25	29	28	31	23
31	27	29	27	27	29	31	27	29	29
29	31	25	29	29	27	29	31	29	27
25	28	27	31	31	29	27	27	33	29
33	31	33	25	27	35	37	35	27	27
29	27	29	31	29	27	29	31	29	21
29	37	31	29	31	29	31	29	39	29
23	29	39	39	27	31	37	29	31	29
27	23	29	27	31	29	29	31	29	35
29	19	29	27	29	29	31	33	29	25

## 2. ПРОВЕРКА СООТВЕТСТВИЯ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Одним из важных результатов теории вероятностей и математической статистики, который широко используется в обработке экспериментальных данных, является центральная предельная теорема. Она очень важна для физических приложений, в частности, потому, что позволяет утверждать, что в подавляющем большинстве экспериментов результаты мы можем описать *нормальным распределением* или, как его еще называют – *распределением Гаусса* [7].

Строгая формулировка этой теоремы следующая:

*Пусть случайная величина  $x$  имеет среднее  $\bar{x}$  и дисперсию  $\sigma^2$ . Если величина  $\sigma^2$  конечна, то распределение выборочного среднего при стремлении числа измерений  $n$  к бесконечности ( $n \rightarrow \infty$ )*

$$X = \frac{1}{n} \sum_{i=1}^n x_i$$

*стремится к распределению Гаусса со средним  $\bar{x}$  дисперсией  $\sigma^2/n$ .*

Другая формулировка теоремы: *если некоторая величина подвержена влиянию большого числа независимых факторов, влияние каждого из которых в отдельности на физическую величину пренебрежимо мало, то распределение этой величины будет нормальным.*

Доказательство этой теоремы выходит за рамки нашего пособия, его подробное объяснение можно найти, например у К. В. Протасова [7]. Нас интересует применение этой теоремы к обработке экспериментальных данных, поскольку она гарантирует, что если число измерений достаточно, то рано или поздно мы получим физическую величину с хорошо известным распределением вероятности. Но обращаем внимание на то, что теорема предельная, следовательно, переход к нормальному распределению осуществляется только в пределе достаточно больших  $n$ . В конкретной экспериментальной ситуации желательно знать, насколько распределение



измеряемой величины близко к Гауссову и достаточно ли число проведенных измерений. Для человека, анализирующего результаты наблюдений, основное заключение теоремы состоит в том, что почти все величины могут быть проанализированы с помощью этого распределения.

### 2.1. Построение кривой распределения по опытным данным

Проверку соответствия опытных данных предполагаемому закону распределения в первом приближении можно осуществить графическим методом. Опытные данные наносят на вероятностную бумагу и сравнивают с графиком принятой функции распределения, которая на вероятностной сетке изображается прямой линией. Если экспериментальные точки ложатся вблизи прямой со случайными отклонениями вправо или влево, то опытные данные соответствуют рассматриваемому закону распределения [1].

Систематическое и значительное отклонения экспериментальных точек от аппроксимирующей прямой свидетельствует о несоответствии данной выборки предполагаемому закону распределения.

Возможен другой вариант применения графического метода для проверки соответствия опытных данных предполагаемому закону распределения [8].

Пусть требуется определить соответствие опытных данных нормальному закону распределения. С этой целью за основу берут дискретный вариационный ряд и в системе координат строят эмпирическую кривую распределения – полигон частот. Затем в этой же системе координат строят точки с координатами  $(x_i; n_i)$ , через которые проводят теоретическую кривую нормального распределения, характеристики  $a, s$  совпадают со статистиками  $x, S$ . Для нахождения теоретических частот  $n_i$  составляется табл. 11:

Таблица 11

$x_i$	$n_i$	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$y_i = \frac{nh}{s} \varphi(u_i)$	$n_i$

где  $x_i$  – варианты дискретного вариационного ряда;

$n_i$  – частоты вариант  $x_i$ ;

$\bar{x}$  – выборочная средняя;

$S$  – выборочное среднее квадратическое отклонение;  
 $h$  – шаг (разность между соседними вариантами);  
 $\varphi(u_i)$  – функция, значения которой находят по прил 1;  
 $y_i$  – выровненные частоты (ординаты) теоретической кривой;  
 $n_i^{\wedge}$  – округленные частоты  $y_i$  до ближайшего целого числа.

## 2.2. Статистические оценки параметров распределения

С помощью статистик  $\bar{x}$  и  $S^2$  можно производить оценку неизвестных параметров предполагаемого закона распределения [1]. Так, например, для нормального закона распределения признака  $X$  при большом объеме выборки его параметры можно считать равными  $\bar{x}$  и  $S^2$ , т.е.  $M(X) = \bar{x}$  и  $\sigma^2 = S^2$ . При малом объеме выборки  $\bar{x}$  и  $S^2$  являются случайными и не совпадают с  $M(X)$  и  $\sigma^2$ . В этом случае оценку неизвестных параметров  $M(X) = a$  и  $\sigma^2$  осуществляют с помощью точечных статистических оценок.

Статистической оценкой неизвестного параметра  $\theta$  теоретического распределения называют его приближенное значение  $\theta^*$ .

Статистическая оценка  $\theta$  является случайной величиной. Чтобы оценка имела практическое значение, она должна быть *несмещенной, состоятельной и эффективной*.

Оценка параметра называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, т.е.  $M(\theta^*) = \theta$ , и – *смещенной*, если  $M(\theta^*) \neq \theta$ .

Оценка называется *состоятельной*, если при увеличении объема выборки она сходится по вероятности к оцениваемому параметру, т.е.

$$\lim_{n \rightarrow \infty} P[|\theta^* - \theta| < \varepsilon] (\varepsilon > 0).$$

Оценка называется *эффективной*, если при заданном объеме выборки  $n$  она имеет наименьшую дисперсию.

Согласно [5], средняя выборочная  $\bar{x}$  является несмещенной и состоятельной оценкой генеральной средней  $M(X)$ , а также доказано, что

исправленная дисперсия  $S^2$  является несмещенной и состоятельной оценкой генеральной дисперсии  $D_r$ , а выборочная дисперсия  $S^2$  – смещенной оценкой  $D_r$ .

Несмещенные оценки  $\bar{x}$ ,  $\hat{s}^2$  и смещенную оценку  $S^2$  называют *точечными*, так как они указывают точку на числовой оси, в которой должно находиться значение оцениваемого параметра.

### 2.3. Проверка статистических гипотез

При изучении генеральной совокупности нам неизвестен либо закон ее распределения, либо параметры распределения. В таком случае в математической статистике выдвигается предположение относительно свойств генеральной совокупности. Такое предположение [4] носит название *статистической гипотезы*.

Гипотезу, имеющую наиболее важное значение в проводимом исследовании, называют нулевой и обозначают через  $H_0$ . При рассмотрении, например, свойств продукции разных машиностроительных предприятий нулевая гипотеза заключается в предположении о независимости характеристик механических свойств профилей от уровня технологии производства.

Нулевую гипотезу выдвигают и затем проверяют с помощью статистических критериев с целью выявления оснований для ее отклонения и для принятия альтернативной гипотезы  $H_A$ . Если имеющийся статистический материал не позволяет отвергнуть нулевую гипотезу, то ее принимают и используют в качестве рабочей гипотезы до тех пор, пока новые накопленные результаты испытаний не позволят ее отклонить. Нулевая гипотеза отвергается, если на основании выборочных испытаний получается маловероятный результат для случая истинности выдвинутой нулевой гипотезы. Границами между высокой и малой вероятностями служат так называемые уровни значимости. Для большинства областей научного исследования в качестве уровней значимости принимают уровни в 5 % и 1 %. Значения статистики, при которых для выбранного уровня значимости отвергается нулевая гипотеза, образуют так называемую критическую область критерия, а значения, при

которых гипотеза не отвергается, – область допустимых значений критерия. Таким образом, статистическая проверка гипотез заключается в построении критической области критерия для выбранного уровня значимости. Если статистика, вычисленная на основании выборки, попадает в критическую область, нулевая гипотеза отвергается, что означает несоответствие проверяемой гипотезы опытными данным [9].

При проверке нулевой гипотезы  $H_0$  могут быть допущены ошибки двух видов: 1) ошибка первого рода, когда верная гипотеза отвергается; 2) ошибка второго рода, когда альтернативная гипотеза принимается.

Наличие таких ошибок объясняется тем, что проверка гипотезы осуществляется с помощью случайной конечной выборки, которая может оказаться «неудачной», приводящей к ложному выводу. Однако преимущество статистической гипотезы состоит в том, что мы можем оценить вероятность, с которой принимается то или иное решение. Вероятность совершить ошибку первого рода, т.е. забраковать верную гипотезу, обозначают через  $\alpha$  и называют уровнем значимости. Чем меньше  $\alpha$ , тем меньше вероятность отвергнуть верную гипотезу. На практике в качестве  $\alpha$  чаще берут значение  $\alpha = 0,05 = 5\%$ . Реже принимают  $\alpha = 0,1$  и  $\alpha = 0,01$ . Если  $\alpha = 5\%$ , то это означает, что существует вероятность ошибочно отвергнуть правильную гипотезу в одном случае из 20.

Вероятность совершить ошибку второго рода обозначают через  $\beta$ . Величину  $1-\beta$  называют *мощностью критерия*. Между уровнем значимости и мощностью критерия  $1-\beta$  существует связь: с уменьшением уровня значимости  $\alpha$ , а значит, с уменьшением вероятности появления ошибки первого рода, падает мощность критерия. В этом случае он все меньше улавливает различие между нулевой и альтернативной гипотезами. Поэтому нельзя беспрестанно уменьшать риск ошибки первого рода, так как суждения становятся все менее определенными.

Для проверки нулевой гипотезы  $H_0$  в математической статистике используют следующие критерии:  $\chi^2$  (хи-квадрат) Пирсона, Романовского, Колмогорова, Ястремского, Стьюдента, Фишера и др.

Статистическая гипотеза может быть проверена на основании результатов случайной выборки. Правило, устанавливающее условия, при которых проверяемая гипотеза принимается или отвергается, называется *статистическим критерием*.

Обработка экспериментальных данных с помощью любого критерия осуществляется по следующей схеме [1].

1. Берется один или два ряда наблюдений (одна или две выборки) и по элементам этих рядов по определенным формулам (для каждого критерия свои формулы) вычисляют статистику.

2. По заданному уровню значимости  $\alpha$  и числу степеней свободы  $k$  находят по таблицам (для каждого критерия свои таблицы), приводимым в приложении любого учебника по теории вероятностей и математической статистике, граничные значения для полученной в п. 1 статистики.

3. Если полученная в п. 1 статистика не выходит за пределы найденных границ, то принимается следующее утверждение: «Нет достаточных оснований отвергнуть выдвинутую гипотезу» [6]. В противном случае нулевая гипотеза отвергается. В математической статистике наиболее употребительными являются следующие критерии: Стьюдента, Фишера (эти критерии исходят из предположения о близости эмпирического распределения признака  $X$  к нормальному закону распределения),  $\chi^2$  Пирсона, Колмогорова, Смирнова, Романовского, Ястремского и другие (эти критерии применяют для проверки близости эмпирического распределения признака  $X$  к нормальному или другим распределениям признака  $X$ , когда не используются конкретные значения параметров этих распределений).

### 2.3.1. Критерий согласия Пирсона $\chi^2$

Критерий согласия Пирсона  $\chi^2$  (хи-квадрат) применяют для проверки гипотезы о соответствии эмпирического распределения предполагаемому теоретическому распределению при большом объеме выборки ( $n > 100$ ) и больших частотах ( $n_i > 5$ ) вариант  $x_i$ .

За меру расхождения эмпирического и теоретического распределений английский математик К. Пирсон принял величину  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^l \frac{(n_i - n'_i)^2}{n'_i}, \quad (26)$$

где  $n_i$  – эмпирические частоты;  $n'_i$  – теоретические частоты.

Применение критерия  $\chi^2$  к проверке гипотезы о нормальном распределении генеральной совокупности значений признака  $X$  осуществляется по следующему правилу:

#### **правило применения критерия $\chi^2$ :**

1. По имеющейся выборке сделать предположение о нормальном законе распределения признака  $X$  генеральной совокупности.

Затем найти оценки параметров этого закона, т.е. найти  $\bar{x}$  и  $S^2$ .

2. Вычислить теоретические частоты  $n'_i$  по формуле

$$n'_i = \frac{nh}{s} \varphi(u_i), \quad (27)$$

где  $n$  – объем выборки;

$h$  – шаг;

$S$  – выборочное среднее квадратическое отклонение;

$u_i = \frac{x_i - \bar{x}}{S}$ ;  $\varphi(u_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$  находится по таблице приложения 1.

Для вычисления теоретических частот  $n'_i$  составить табл. 12.

Таблица 12

$x_i$	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$n'_i = \frac{nh}{s} \varphi(u_i)$

Полученные частоты  $n'_i$  округлить до целых.

3. Вычислить величину  $\chi^2$  по формуле (26) и обозначить ее через  $\chi_0^2$ .

Расчет вести, пользуясь табл. 13.

Таблица 13

$n_i$	$n'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
				$\chi_0^2$

4. Найти число степеней свободы  $k$  (параметр распределения Пирсона) по формуле  $k = s - r = s - 3$ , где  $s$  – число интервалов вариационного ряда;  $r$  – сумма числа параметров теоретического закона распределения. Для нормального распределения признака  $X$  принято  $r = 3$  (учитываются параметры нормального распределения  $a$  и  $\sigma$ , а также объем выборки  $n$ ).

5. Выбрать уровень значимости  $\alpha$ .

6. По найденному числу степеней свободы  $k$  и уровню значимости  $\alpha$ , пользуясь приложением 5, определить критическое значение  $\chi_{кр}^2$ .

Если  $\chi_0^2 < \chi_{кр}^2$ , то нет достаточных оснований отвергнуть выдвинутую гипотезу о нормальном распределении признака  $X$ . Если  $\chi_0^2 > \chi_{кр}^2$ , то гипотеза о нормальном распределении признака  $X$  отвергается.

Критерий Пирсона можно применять для проверки гипотезы о том, что данная выборка взята из генеральной совокупности, распределенной по биномиальному закону, закону Пуассона, экспоненциальному закону.

Рассмотрим гипотезу  $H_0$  о близости эмпирического распределения признака  $X$  к распределению Пуассона:

$$H_0: F(x) = \sum_0^m \frac{\lambda^i}{i!} e^{-\lambda}, \quad (28)$$

где  $\lambda$  – параметр распределения Пуассона.

Для применения критерия  $\chi^2$  надо рассчитать теоретические частоты  $n'_i$ , а также получить по значениям выборки оценку параметра  $\lambda$ . Методом максимального правдоподобия доказывается, что выборочная средняя  $\bar{x}$

является пригодной оценкой для  $\lambda$ , т.е.  $\lambda \approx \bar{x}$ . Теоретические частоты  $n'_i$  вычисляются по формуле:

$$n'_i = n_p(X = x_i). \quad (29)$$

Вероятности  $P(X = x_i)$  вычисляются по формуле:

$$P(X = x_i) = \frac{\bar{x}^i}{i!} e^{-\bar{x}}. \quad (30)$$

Так как при расчете теоретических частот  $n'_i$  используется один параметр  $\lambda$ , то число степеней свободы  $k$  находят по формуле:

$$k = s - 2. \quad (31)$$

Затем вычисляют величину  $\chi^2$  по формуле (26), обозначают ее через  $\chi_0^2$ . По заданному уровню значимости  $\alpha$  и найденному числу степеней свободы  $k$  по приложению 5 находят  $\chi_{кр}^2$ .

Если  $\chi_0^2 < \chi_{кр}^2$ , то нет достаточных оснований отвергнуть выдвинутую гипотезу  $H_0$ .

Если  $\chi_0^2 > \chi_{кр}^2$ , то гипотеза  $H_0$  отвергается.

### 2.3.2. Критерий Романовского

Для оценки близости эмпирического распределения признака  $X$  к нормальному теоретическому Романовский предложил вычислять отношение:

$$\left| \frac{\chi^2 - k}{\sqrt{2k}} \right|, \quad (32)$$

где  $\chi^2$  – статистика критерия Пирсона, вычисленная по формуле (26), используя опытные данные;  $k = s - 3$  – число степеней свободы.

Если указанное отношение по модулю меньше трех, то расхождение между теоретическим и эмпирическим распределениями считается несущественным, т.е. можно принять, что данное эмпирическое распределение моделируется нормальным распределением. Если отношение (32) больше трех, то у нас нет оснований считать, что эмпирическое распределение признака  $X$  подчиняется нормальному закону распределения [4].



### 2.3.3. Критерий Колмогорова

Критерий Колмогорова в своем классическом виде является более мощным, чем критерий Пирсона, и может быть использован для проверки гипотезы о соответствии эмпирического распределения любому теоретическому непрерывному распределению  $F(x)$  с заранее известными параметрами.

Однако параметры функции распределения  $F(x)$ , как правило, нам неизвестны, и их оценка производится по данным самой выборки. Это обстоятельство накладывает ограничения на возможность широкого практического применения критерия: он может быть использован только для проверки соответствия опытных данных лишь некоторым конкретным функциям распределения.

Для проверки соответствия эмпирического распределения теоретическому нормальному распределению критерий Колмогорова применяют следующим образом.

Вычисляется статистика  $\lambda$  критерия Колмогорова по формуле:

$$\lambda = \frac{D}{\sqrt{n}}, \quad (33)$$

где  $D = \max|M - M'|$  – максимум абсолютного значения разности между накопленными эмпирическими частотами  $M$  и накопленными теоретическими частотами  $M'$ ;  $n$  – объем выборки.

По вычисленному  $\lambda$  находят значение функции

$$K(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 \lambda^2}$$

– вероятность того, что  $\lambda$  достигает данной величины.

Значения  $K(\lambda)$  находят, пользуясь табл. 14.

Если найденному значению  $\lambda$  соответствует очень малая вероятность, т.е.  $K(\lambda) < 0,05$ , то имеет место существенное расхождение между эмпирическим и теоретическим распределениями, которое нельзя считать случайным.

Следовательно, рассматриваемая выборка не может быть смоделирована нормальным законом распределения.

Таблица 14

$\lambda$	$K(\lambda)$	$\lambda$	$K(\lambda)$
0,30	1,0000	1,10	0,1777
0,35	0,9997	1,20	0,1122
0,40	0,9972	1,30	0,681
0,45	0,9874	1,40	0,397
0,50	0,9639	1,50	0,222
0,55	0,9228	1,60	0,120
0,60	0,8643	1,70	0,052
0,70	0,7112	1,90	0,015
0,75	0,6272	2,00	0,007
0,80	0,5441	2,10	0,0003
0,85	0,4653	2,20	0,0001
0,90	0,3927	2,30	0,0001
0,95	0,3275	2,40	0,0000
1,00	0,2700	2,50	0,0000

Если вероятность  $K(\lambda) > 0,05$ , то расхождение между частотами может быть случайным и распределения хорошо соответствуют одно другому

Если проверяют гипотезу по критерию Колмогорова о соответствии выборки экспоненциальному распределению, параметры которого оценивают по опытным данным, то сначала вычисляют статистики [9]:

$$D_n^+ = \max_{1 \leq i \leq n} \left[ \frac{i}{n} - 1 + \exp(-x_i / \bar{x}) \right], \quad (34)$$

$$D_n^- = \max_{1 \leq i \leq n} \left[ 1 - \exp(-x_i / \bar{x}) - \frac{i-1}{n} \right], \quad (35)$$

$$D_n = \max[D_n^+, D_n^-], \quad (36)$$

Затем составляют неравенство:

$$\left( D_n - \frac{0,2}{n} \right) \left( \sqrt{n} + 0,25 + \frac{0,5}{n} \right) \leq \lambda_\alpha. \quad (37)$$

Критические значения  $\lambda_\alpha$  для этого случая:  $\lambda_{0,1} = 0,99$ ,  $\lambda_{0,05} = 1,09$  и  $\lambda_{0,01} = 1,31$ . Если неравенство (37) при выбранном  $\lambda_\alpha$  выполняется, то это означает, что эмпирическое распределение можно изучать на математической модели, подчиняющейся экспоненциальному закону распределения.

### 2.3.4. Критерий Ястремского

Для проверки соответствия данной выборочной совокупности признака  $X$  нормальному распределению составляется [4] неравенство:

$$J \leq 3\sqrt{2l + 4\theta}, \quad (38)$$

где  $J = |c - l|$ ;

$$c = \sum \frac{(n_i - n'_i)^2}{n'_i q'_i};$$

$n_i$  – эмпирические частоты;

$n'_i$  – теоретические частоты;

$l$  – число столбцов дискретного вариационного ряда;

$q'_i = 1 - p_i$ ,  $p'_i = \frac{n'_i}{n}$ ,  $n$  – объем выборки.

Если  $l < 20$ , то  $\theta = 0,6$ .

Если вычисленное значение  $J$  меньше  $3\sqrt{2l + 4\theta}$ , то гипотеза о близости эмпирического распределения признака  $X$  к нормальному закону распределения принимается. При значениях  $J$ , больших  $3\sqrt{2l + 4\theta}$ , расхождения между эмпирическим и теоретическим распределениями существенно. В этом случае данные выборки не будут подчиняться нормальному закону распределения.

Для вычисления величины  $c$  составляется табл. 15.

Таблица 15

$x_i$	$n_i$	$n'_i$	$p'_i$	$q'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$n'_i q'_i$	$\frac{(n_i - n'_i)^2}{n'_i q'_i}$
$\Sigma$								

### 2.3.5. Приближенные критерии нормальности распределения

Для проверки гипотезы о соответствии данной выборки нормальному закону распределения используют выборочные статистики: асимметрию и эксцесс. В этом случае названные статистики вычисляются по формулам (18) и (19). Затем вычисляются [9] их средние квадратические отклонения по формулам:

$$S_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}, \quad (39)$$

$$S_{E_x} = \sqrt{\frac{24(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}. \quad (40)$$

Если  $|A_s| \leq S_{A_s}$  и  $|E_x| \leq S_{E_x}$ , то выборочная совокупность подчиняется нормальному закону распределения.

Если  $|A_s|$  и  $|E_x|$  заметно больше своих средних квадратических отклонений, то выборочная совокупность не будет распределена по нормальному закону.

Проверку выборочной совокупности на близость ее к нормальному распределению можно производить, используя статистики  $\chi^2$ ,  $A_s$ ,  $E_x$ .

Сначала вычисляют статистику  $\chi^2$  по формуле:

$$\chi^2 = \frac{A_s^2}{S_{A_s}^2} + \frac{E_x^2}{S_{E_x}^2}. \quad (41)$$

Затем при заданном уровне значимости  $\alpha$  и числе степеней свободы  $k = 2$  (используют в расчетах две статистики  $A$  и  $E$ ) по приложению 5 для распределения  $\chi^2$  Пирсона находят  $\chi_{кр}^2$ . Если выполняется неравенство  $\chi^2 < \chi_{кр}^2$ , то гипотезу о нормальном распределении выборочной совокупности принимают. В противном случае, т.е. когда  $\chi^2 > \chi_{кр}^2$ , гипотезу о нормальном распределении выборки отвергают [1].

## 2.4. Лабораторная работа № 2

### Построение кривой нормального распределения по опытным данным. Проверка гипотезы о нормальном распределении выборки

*Цель работы:* овладение студентом способами построения эмпирической и теоретической (нормальной) кривой распределения; выработка умения и навыков применения критериев согласия для проверки выдвинутой статистической гипотезы.

#### *Задание к лабораторной работе № 2*

На основе дискретного вариационного ряда, полученного в лабораторной работе № 1, выполнить следующее:

- Построить эмпирическую (полигон) и теоретическую (нормальную) кривую распределения.
- Проверить согласованность эмпирического распределения с теоретическим нормальным, применяя три критерия:
  - а) критерий Пирсона;
  - б) один из критериев: Колмогорова, Романовского или Ястремского;
  - в) приближенный критерий.

#### 2.4.1. Пример выполнения лабораторной работы №2

Продолжим вероятностно-статистическую обработку результатов эксперимента, предложенных в лабораторной работе № 1, то есть обводненности нефти из насосных скважин. За основу берем дискретный вариационный ряд из лабораторной работы № 1 и значения  $\bar{x} = 61,39$  и  $S = 0,599$ .

Варианты, $x_i$	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
Частоты, $m_i$	3	6	9	18	29	16	2	10	5	2

Эмпирическая кривая распределения представляет собой полигон частот (см. лабораторную работу № 1). Для построения теоретической (нормальной) кривой найдем координаты точек  $(x_i, n_i)$ , для чего рассчитаем теоретические частоты  $n'_i$  (табл. 16).

Таблица 16

$x_i$	$n_i$	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$y_i = \frac{nh}{s} \varphi(u_i)$	$n'_i$
60,15	3	-1,24	-2,07	0,0468	2,3	2
60,45	6	-0,94	-1,57	0,1163	5,8	6
60,75	9	-0,64	-1,07	0,2251	11,3	11
61,05	18	-0,34	-0,57	0,3391	17,0	17
61,35	29	-0,04	-0,07	0,3980	19,9	20
61,65	16	0,26	0,43	0,3637	18,2	18
61,95	2	0,56	0,93	0,2589	13,0	13
62,25	10	0,86	1,44	0,1415	7,1	7
62,55	5	1,16	1,94	0,0608	3,0	3
62,85	2	1,46	2,44	0,0203	1,0	1

Строим эмпирическую и теоретическую кривую (рис.4).

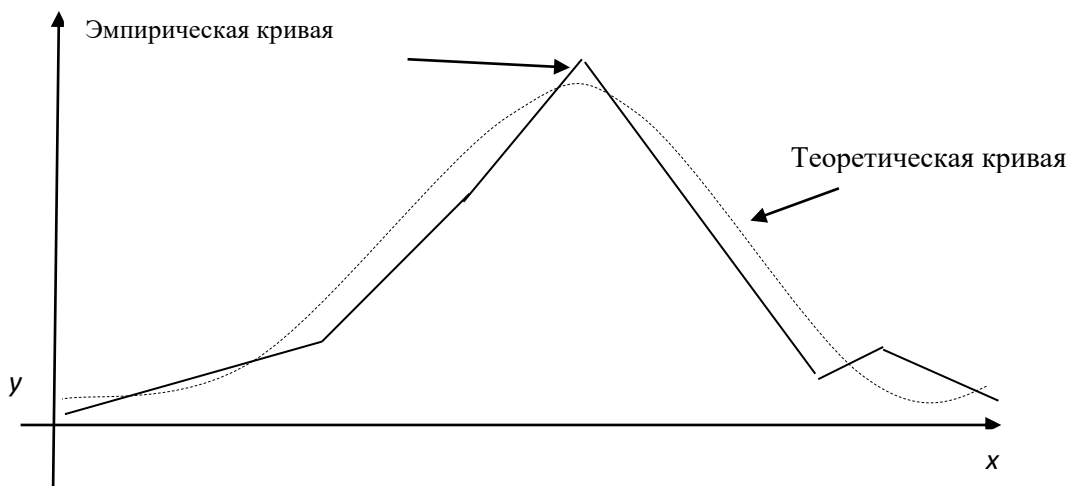


Рис. 4 Эмпирическая и теоретическая кривые

Проверим согласованность эмпирического распределения (обводненности нефти из насосных скважин) с теоретическим нормальным по критерию Пирсона. Вычислим величину  $\chi^2$  по формуле (26):

$$\chi^2 = \sum_{i=1}^l \frac{(n_i - n'_i)^2}{n'_i},$$

Для нахождения суммы составляем расчетную табл. 17.

Таблица 17

$n_i$	$n'_i$	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
3	2	1	1	0,5
6	6	0	0	0
9	11	-2	4	0,363636
18	17	1	1	0,058824
29	20	9	81	4,05
16	18	-2	4	0,222222
2	13	-11	121	9,307692
10	7	3	9	1,285714
5	3	2	4	1,333333
2	1	1	1	1
				$\chi_0^2=18,12$

Находим число степеней свободы  $k = s - r = s - 3 = 10 - 3 = 7$ .

Выбираем уровень значимости  $\alpha=0,95$ . По таблице критических точек распределения (приложение 5) находим  $\chi_{кр}^2 = 2,17$ . Так как  $\chi_{кр}^2 < \chi_0^2$  ( $2,17 < 18,12$ ), то делаем вывод, что данные выборки, характеризующие обводненность нефти из насосных скважин, не подчиняются нормальному закону распределения.

Проведём проверку близости эмпирического распределения к нормальному по критерию Романовского. Вычислим, согласно (32), величину  $\left| \frac{\chi^2 - k}{\sqrt{2k}} \right|$ . Так как  $\chi^2 = \chi_0^2 = 18,12$ ,  $k = 7$ , то  $\left| \frac{\chi^2 - k}{\sqrt{2k}} \right| = \left| \frac{18,12 - 7}{\sqrt{14}} \right| = 2,97 < 3$ , т.е. расхождение между эмпирическим и теоретическим распределением несущественно, что позволяет утверждать, что данные выборки, характеризующие обводненность нефти из насосных скважин по критерию Романовского, подчиняются нормальному закону распределения. К такому же выводу мы приходим, применяя критерий Колмогорова (проверить самостоятельно).

Наконец, проведём проверку близости рассматриваемой выборки к нормальному распределению по приближенному критерию, используя выборочные статистики: асимметрию, эксцесс и их средние квадратические отклонения. В лабораторной работе № 1 были найдены  $A_s = 0,328$ ,  $E_x = -0,115$ .

Средние квадратические отклонения для асимметрии и эксцесса находим по формулам (39) и (40):

$$S_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} = \sqrt{\frac{6 \cdot 99}{101 \cdot 103}} = 0,24,$$

$$S_{E_x} = \sqrt{\frac{24(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} = \sqrt{\frac{24 \cdot 100 \cdot 98 \cdot 97}{99^2 \cdot 103 \cdot 105}} = 0,46.$$

Так как  $|A_s| = 0,328 > S_{A_s}$ ,  $|E_x| = 0,115 < S_{E_x}$ , то делаем вывод, что  $A_s$  и  $E_x$  данной выборки, характеризующие обводненность нефти из насосных скважин, не подчиняются нормальному закону распределения.

Итак, для проверки согласованности эмпирического распределения с теоретическим нормальным мы применили четыре критерия; два из них подтвердили близость выборочной совокупности к нормальному распределению. Однако, учитывая, что критерий Колмогорова является более мощным, чем критерий  $\chi^2$  Пирсона, и подтверждает близость рассматриваемой выборки к нормальному распределению, окончательно заключаем, что за закон распределения признака  $X$  – обводненности нефти из насосных скважин – можно принять нормальное распределение.

*Замечание.* В качестве вариантов заданий для выполнения лабораторной работы № 2 следует брать дискретные вариационные ряды из лабораторной работы № 1, а также значения основных статистик.



### 3. ПАРНАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

#### 3.1. Понятие корреляционной зависимости. Задачи теории корреляции

При анализе хозяйственной деятельности предприятий возрастает роль экономико-математических методов для управления производством. Управление производством – это сложный динамический процесс. Поэтому при выработке оптимального решения по управлению производственно-хозяйственной деятельностью предприятия необходимо не только учитывать изменения параметров и характеристик, описывающих эту деятельность, но и уметь их прогнозировать, основываясь на экономических законах, которые наиболее полно отражают взаимосвязи основных показателей предприятия и его подразделений. Математическая формализация этих связей создает условия для экономического обоснования целесообразных объемов производимой продукции, определения ее качественных показателей и условий эффективного использования ресурсов.

Для решения этих задач применяют методы корреляционного анализа.

Корреляция — статистическая взаимосвязь двух или более случайных величин. При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

При выявлении зависимостей между производственными показателями методами корреляционного анализа выделяют два основных типа переменных количественных признаков: независимые переменные (факторные признаки) и зависимые переменные (результативные признаки).

При изучении взаимосвязей между переменными признаками надо, прежде всего, установить, к какому типу зависимостей относится эта связь. Зависимость между признаками  $X$  и  $Y$  называется корреляционной, если каждому возможному значению  $x_i$  признака  $X$  сопоставляется условная средняя соответствующего распределения признака  $Y$ .

Среднее арифметическое значение признака  $Y$ , вычисленное при условии, что признак  $X$  принимает фиксированное значение  $x_i$ , называется условным средним, обозначается через  $\bar{y}_{x_i}$  и вычисляется по формуле:

$$\bar{y}_{x_i} = \frac{\sum n_{ij} \cdot y_j}{n_{x_i}}, \quad (42)$$

где  $n_{ij}$  – частоты, показывающие, сколько раз повторяются парные значения  $x_i$ ,  $y_j$  в данной выборке;  $n_{x_i}$  – частота появления значения  $x_i$ .

Теория корреляции изучает такую зависимость между признаками  $X$  и  $Y$ , при которой с изменением одного признака меняется распределение другого. Она применяется для того, чтобы при сложном взаимодействии посторонних факторов выяснить, какова должна быть зависимость между признаками  $X$  и  $Y$ , если бы посторонние факторы не изменялись и своим изменением не искажали истинную статистическую зависимость [1].

В теории корреляции решается триединая задача, методологической основой которой является триада:

*Модель – Свойства – Адекватность.*

Первая задача – поиск подходящей модели. На основе опытных данных выявляется характер корреляционной зависимости между признаками  $X$  и  $Y$ . При парной корреляции для ее решения применяют графический метод. Если в корреляционном поле точки  $(x_i, y_j)$  хорошо ложатся на прямую, то можно предположить, что связь между признаками  $X$  и  $Y$  носит линейный характер. Если точки не ложатся на прямую, то связь будет нелинейной. Исходя из геометрических соображений, выбирают уравнение линии, которое называют уравнением регрессии, и находят неизвестные параметры, входящие в уравнение.

Вторая задача – изучение свойств модели. Определяется теснота связи между признаками, включенными в модель, по коэффициенту  $r$  корреляции (в случае линейной корреляции) или по корреляционным отношениям  $\eta_{yx}$ ,  $\eta_{xy}$  (в случае криволинейной корреляции).

Третья задача – выявление степени адекватности построенной корреляционной модели (проверяется соответствие полученного уравнения регрессии опытными данным). Если данная модель оказалась не адекватной, то всё начинается сначала – строят новую модель.

### 3.2. Парная линейная корреляционная зависимость

Предположим, что на основе геометрических, физических или других соображений установлено, что между двумя количественными признаками  $X$  и  $Y$  существует линейная корреляционная зависимость. Тогда уравнение регрессии записывают в виде:

$$\hat{y}_x = a_0 + a_1x. \quad (43)$$

Пусть опытные данные не сгруппированы в корреляционную таблицу, т.е. заданы в виде табл. 18.

Таблица 18

$x_i$	$x_1$	$x_2$	$x_3$	...	$x_k$
$y_i$	$y_1$	$y_2$	$y_3$	...	$y_k$

В этом случае значения  $a_0$ ,  $a_1$ , являющиеся оценками истинных величин уравнения регрессии, находят по методу наименьших квадратов [1], решая систему линейных алгебраических уравнений (СЛАУ) относительно  $a_0$ ,  $a_1$ :

$$\begin{cases} na_0 + [x]a_1 = [y], \\ [x]a_0 + [x^2]a_1 = [xy], \end{cases} \quad (44)$$

где  $[x] = \sum_{i=1}^k x_i$ ,  $[y] = \sum_{i=1}^k y_i$ ,  $[x^2] = \sum_{i=1}^k x_i^2$ ,  $[xy] = \sum_{i=1}^k x_i y_i$ .

Для нахождения сумм, входящих в систему (44), составляется табл. 19

Таблица 19

$x_i$	$y_i$	$x_i y_i$	$x_i^2$
$[x]$	$[y]$	$[xy]$	$[x^2]$

Если опытные данные сгруппированы в корреляционную таблицу, то значения  $a_0$  и  $a_1$  уравнения регрессии (43) находят по методу наименьших квадратов, решая СЛАУ:

$$\begin{cases} na_0 + [n_x x]a_1 = [n_y y], \\ [n_x x]a_0 + [n_x x^2]a_1 = [n_{xy} xy], \end{cases} \quad (45)$$

где  $n_x$  и  $n_y$  – частоты признаков  $X$  и  $Y$ ;  $n_{xy}$  – частота совместного появления признаков  $X$  и  $Y$ .

Для нахождения сумм, входящих в систему (45), составляется табл. 20.

Таблица 20

y	x					
	$x_1$	$x_2$	...	$x_k$	$n_y$	$n_{xy}$
$y_1$			...			
$y_2$			...			
...	...	...	...	...	...	...
$y_m$			...			
$n_x$			...			$[n_{xy}]$
$n_x x$			...		$[n_x x]$	
$n_x x^2$			...		$[n_x x^2]$	
$n_{xy} xy$			...		$[n_{xy} xy]$	

Суммы  $[n_x x]$ ,  $[n_x x^2]$ ,  $[n_{xy} xy]$  в табл. 20 находятся по строкам, а сумма  $[n_{xy}]$  – по последнему столбцу табл. 20.

В уравнении регрессии (43) параметр  $a_0$  характеризует усредненное влияние на результативный признак  $Y$  неучтенных (не выявленных для исследования) факторных признаков  $X_i$ . Параметр  $a_1$  показывает, насколько изменяется в среднем значение результативного признака  $Y$  при увеличении факторного признака на единицу.

Согласно [12], используя параметр  $a_1$ , вычисляют коэффициент эластичности  $K_y$  по формуле:

$$K_y = a_1 \frac{\bar{x}}{\bar{y}}. \quad (46)$$

Коэффициент эластичности  $K_y$  показывает, на сколько процентов изменяется результативный признак  $Y$  при изменении факторного признака  $X$  на 1 %.

В случае линейной корреляционной зависимости между признаками  $X$  и  $Y$ , уравнения регрессий находят [1] по формулам:

$$\hat{y}_x = \bar{y} + r \frac{S_y}{S_x} (x - \bar{x}), \quad (47)$$

$$\hat{x}_y = \bar{x} + r \frac{S_x}{S_y} (y - \bar{y}), \quad (48)$$

где  $\bar{x}$  и  $\bar{y}$  – выборочные средние признаков  $X$  и  $Y$ ;  $S_x, S_y$  – выборочные средние квадратические отклонения признаков  $X$  и  $Y$ , вычисляемые по формулам (при  $n < 50$ ):

$$S_x = \sqrt{S_x^2}, \text{ где } S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (49)$$

$$S_y = \sqrt{S_y^2}, \text{ где } S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (50)$$

При  $n \geq 50$   $S_x$  и  $S_y$  находят по формулам:

$$S_x = \sqrt{S_x^2}, \text{ где } S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (51)$$

$$S_y = \sqrt{S_y^2}, \text{ где } S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (52)$$

Коэффициент линейной корреляции  $r$  находят по формуле:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x S_y}, \quad (53)$$

где  $\overline{xy}$  – средняя произведения значений признаков  $X$  и  $Y$ ,  $\bar{x}, \bar{y}$  – средние значения признаков  $X$  и  $Y$ ,  $S_x, S_y$  – выборочные средние квадратические отклонения признаков  $X$  и  $Y$ , вычисленные по формулам (49) и (50), если  $n < 50$ , или по формулам (51) и (52), если  $n \geq 50$ .

Уравнение (47) называют уравнением регрессии  $y$  на  $x$ , а уравнение (48) – уравнением регрессии  $x$  на  $y$ .

Если данные выборки для признаков  $X$  и  $Y$  заданы в виде корреляционной таблицы и объем выборки  $n > 30$ , то для нахождения величин, входящих в уравнения линий регрессий (47) и (48), переходят к вспомогательному распределению с условными вариантами  $u_i$  и  $v_j$ , вычисляемых по формулам:

$$u_i = \frac{x_i - C_1}{h_1}, \quad (54)$$

$$v_j = \frac{y_j - C_2}{h_2}, \quad (55)$$

где  $C_1 = M_0 X$ ,  $C_2 = M_0 Y$ ,  $h_1$  и  $h_2$  – шаги значений признаков  $X$  и  $Y$ .

Выборочный коэффициент линейной корреляции  $r$  в этом случае находят по формуле:

$$r = \frac{\sum n_{uv}uv - n\bar{u}\bar{v}}{nS_uS_v}, \quad (56)$$

$$\text{где } S_u = \sqrt{\bar{u}^2 - \bar{u}^2}, S_v = \sqrt{\bar{v}^2 - \bar{v}^2}. \quad (57)$$

Для нахождения суммы  $\sum n_{uv}uv$  составляется расчетная табл. 21.

Таблица 21

$u$	$v$				
	$v_1$	$v_2$	...	$v_k$	$n_u$
$u_1$	$u_1v_1$ $n_{u_1v_1}$	$u_1v_2$ $n_{u_1v_2}$	...	$u_1v_k$ $n_{u_1v_k}$	$n_{u_1}$
$u_2$	$u_2v_1$ $n_{u_2v_1}$	$u_2v_2$ $n_{u_2v_2}$	...	$u_2v_k$ $n_{u_2v_k}$	$n_{u_2}$
...	...	...	...	...	...
$u_n$	$u_nv_1$ $n_{u_nv_1}$	$u_nv_2$ $n_{u_nv_2}$	...	$u_nv_k$ $n_{u_nv_k}$	$n_{u_n}$
$n_v$	$n_{v_1}$	$n_{v_2}$	...	$n_{v_k}$	$\sum n_{uv}uv$

Статистики  $\bar{x}$ ,  $\bar{y}$ ,  $S_x$ ,  $S_y$  находят по формулам:

$$\bar{x} = \bar{u}h_1 + C_1, \bar{y} = \bar{v}h_2 + c_2, S_x = S_uh_1, S_y = S_vh_2. \quad (58)$$

### 3.3. Коэффициент корреляции, его свойства и значимость

После выбора функции как формы корреляционной зависимости между признаками  $X$  и  $Y$  решается задача, состоящая в определении тесноты связи между ними, в оценке рассеяния относительно линии регрессии значений одного признака для различных значений другого. Для этого используют выборочный коэффициент  $r$  корреляции, который вычисляют по формуле (53). Линейный коэффициент корреляции изменяется на отрезке  $[-1; 1]$ , т.е.  $|r| \leq 1$ . Если  $r = \pm 1$ , то корреляционная зависимость становится функциональной. В случае  $r > 0$  говорят о положительной корреляции величин  $X$ ,  $Y$ , например, вес и рост человека связаны положительной корреляцией; в случае  $r < 0$  – об

отрицательной корреляции. Положительная корреляция между случайными величинами означает, что при возрастании одной из них другая имеет тенденцию в среднем возрастать; отрицательная корреляция означает, что при возрастании одной из случайных величин другая имеет тенденцию в среднем убывать.

Если  $r = 0$ , то линейная связь между признаками  $X$  и  $Y$  отсутствует, но может существовать криволинейная корреляционная связь или нелинейная функциональная.

Оценку тесноты линейной корреляционной связи определяют, пользуясь табл. 22.

Таблица 22

Теснота связи	Величина $r$	
	положительная	отрицательная
Линейной связи нет	$0 \div 0,2$	$0 \div -0,2$
Слабая	$0,2 \div 0,5$	$-0,2 \div -0,5$
Средняя	$0,5 \div 0,75$	$-0,5 \div -0,75$
Сильная	$0,75 \div 0,95$	$-0,75 \div -0,95$
Функциональная	$0,95 \div 1$	$-0,95 \div -1$

Значимость выборочного коэффициента корреляции проверяют по критерию Стьюдента. По опытным данным вычисляют расчетную статистику  $t_p$ , пользуясь формулой:

$$t_p = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (59)$$

Затем по таблице критических точек распределения Стьюдента (приложение 6) по заданному уровню значимости  $\alpha$  и числу степеней свободы  $k=n-2$  находят табличное значение двусторонней критической области  $t_{кр}$ . Если  $t_p < t_{кр}$ , то коэффициент корреляции  $r$  – незначимый (мало отличается от нуля) и

признаки  $X$  и  $Y$  некоррелированные. Если  $t_p > t_{кр}$ , то приходят к выводу о наличии линейной корреляционной связи [1].

### 3.4. Определение надежности

#### (доверительного интервала) коэффициента корреляции

Коэффициент корреляции, как правило, рассчитывается по данным выборки. Чтобы полученный результат распространить на генеральную совокупность, возможно возникновение некоторой ошибки, которую оценивают с помощью средней квадратичной ошибки  $\sigma_r$ . С помощью  $\sigma_r$  производят оценку надежности коэффициента корреляции, построив доверительные интервалы для различных объемов выборки [4]. Пусть число  $n$  наблюдений пар чисел  $(x; y)$  меньше 50 ( $n < 50$ ). В этом случае средняя квадратическая ошибка  $\sigma_r$  вычисляется по формуле:

$$\sigma_r = \frac{1-r^2}{\sqrt{n-2}}, \quad (60)$$

где  $r$  – коэффициент парной линейной корреляции;  $n$  – объем выборки.

Доверительный интервал для оценки  $r$  находят по формуле:

$$r - t_\gamma \sigma_r \leq \hat{r} \leq r + t_\gamma \sigma_r, \quad (61)$$

где  $t_\gamma$  находят по таблице значений функции Лапласа  $\Phi(x)$  (приложение 2). Если задать надежность  $\gamma = 0,95$ , то  $\Phi(x) = \gamma/2 = 0,475$  и  $t_\gamma = 1,96$ .

Если объем выборки  $n > 50$ , то погрешность  $\sigma_r$  для коэффициента корреляции  $r$  находят также по формуле (60). Затем вычисляют отношение  $r/\sigma_r$ . Если это отношение больше 3, то можно считать, что найденный коэффициент корреляции  $r$  отражает истинную зависимость между признаками  $X$  и  $Y$ . Величина  $r-3\sigma_r$  является, как правило, гарантийным  $r$  минимумом, а величина  $r+3\sigma_r$  – гарантийным максимумом коэффициента корреляции  $r$ , и доверительный интервал для оценки  $r$  запишется в виде:

$$r-3\sigma_r \leq r \leq r+3\sigma_r. \quad (62)$$



### 3.5. Коэффициент детерминации

Линейный коэффициент корреляции оценивает тесноту взаимосвязи между признаками и показывает, является ли эта корреляция положительной или отрицательной. Однако понятия тесноты взаимосвязи бывает недостаточно при содержательном анализе взаимосвязей. В частности, коэффициент корреляции не показывает степень воздействия факторного признака  $X$  на результативный  $Y$ . Таким показателем является коэффициент детерминации [1].

Пусть по опытным данным для признаков  $X$  и  $Y$  получены уравнения регрессий  $\hat{y}_x = a_0 + a_1x$  и  $\hat{x}_y = b_0 + b_1y$ . Величину  $a_1b_1 = r^2$  называют *коэффициентом детерминации*. Этот коэффициент детерминации можно находить и по формуле:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_{x_i})^2}{\sum(y_i - \bar{y})^2}, \quad (63)$$

где  $y_i$  – опытные значения признака  $Y$ ;  $\hat{y}_{x_i}$  – значения  $y$ , найденные по уравнению регрессии;  $\bar{y}$  – средняя признака  $Y$ .

Формулой (63) пользуются тогда, когда общее число значений  $y_i$  равно числу значений  $x_i$  признака  $X$ .

Коэффициент детерминации используется, во-первых, для контроля вычислений, проводимых при получении уравнений регрессий ( $r^2 = a_1b_1$ ) и, во-вторых, он показывает, какую часть рассеяния результативного признака  $Y$  можно объяснить принятой регрессионной моделью.

### 3.6. Проверка адекватности модели

Для проверки соответствия уравнения регрессии  $\hat{y}_x = a_0 + a_1x$  опытными данным применяют критерий Фишера – Снедекора. Вычисляют статистику  $F_H$  по формуле:

$$F_H = \frac{R^2(n-2)}{1-R^2}, \quad (64)$$

где  $R^2$  – коэффициент детерминации,  $n$  – объем выборки.

Чем ближе значение  $R^2$  к единице, тем лучше модель согласуется с опытными данными. Затем при заданном уровне значимости  $\alpha$  и числах степеней свободы  $k_1=1$ ,  $k_2= n-2$  находят по таблице критических точек распределения Фишера – Снедекора (приложение 7)  $F_T = F_{\alpha; k_1; k_2}$ . Если окажется, что  $F_H > F_T$ , то полученное уравнение линейной регрессии согласуется с опытными данными. Если  $F_H < F_T$ , то модель регрессии не согласуется с данными опыта. Формулой (64) пользуются тогда, когда исходные данные заданы не в виде корреляционной таблицы.

Если опытные данные заданы в виде корреляционной таблицы, то проверку модели на адекватность можно выполнить тогда, когда общее число значений  $y_i$  больше числа значений  $x_j$ . В этом случае [1] находят остаточную сумму квадратов  $Q_e$ , характеризующую влияние неучтенных в модели факторов, по формуле:

$$Q_e = Q - Q_R, \quad (65)$$

где  $Q = \sum_{i=1}^n (y_i - \bar{y})^2$  – сумма квадратов отклонений значений  $y_i$  от средней  $\bar{y}$ ,  $Q_R = \sum_{i=1}^n (\bar{y}_{x_i} - \bar{y})^2$  – сумма квадратов отклонений условных средних  $\bar{y}_{x_i}$  от средней  $\bar{y}$ .

Затем вычисляется статистика  $F_H$  по формуле:

$$F_H = \frac{Q_R(n-2)}{Q_e(k-1)}. \quad (66)$$

По таблице критических точек распределения Фишера – Снедекора (приложение 7) при заданном уровне значимости  $\alpha$  и числах степеней свободы  $k_1 = 1$ ,  $k_2 = n - 2$  находят  $F_T = F_{\alpha; k_1; k_2}$ . Если  $F_H > F_T$ , то модельное уравнение

регрессии значимо описывает опытные данные, в противном случае ( $F_H < F_T$ ) – нет.

### 3.7. Оценка величины погрешности

После проверки модельного уравнения линейной корреляции на адекватность находят относительную погрешность уравнения по формуле, %:

$$\delta = \frac{\sigma_u}{\bar{y}} 100, \quad (67)$$

где  $\sigma_u = \sqrt{D_u}$ ,  $\sigma_u$  – стандартная ошибка уравнения регрессии;

$$D_u = \frac{\sum(u_i - \bar{u})^2}{n-2} \text{ – остаточная дисперсия;}$$

$$u_i = y_i - \hat{y}_{x_i}, \text{ } y_i \text{ – опытные значения } y;$$

$$\hat{y}_{x_i} \text{ – значения } y, \text{ полученные по уравнению регрессии;}$$

$$\bar{u} = \frac{1}{n} \sum (y_i - \hat{y}_{x_i})^2 \text{ – среднее значение } u_i;$$

$n$  – объем выборки.

Если величина  $\delta$  мала, то прогнозные качества оцененного регрессионного уравнения высоки. Одновременно производят оценку коэффициентов уравнения регрессии  $\hat{y}_x = a_0 + a_1 x$ . Пусть  $S_{a_0}$  и  $S_{a_1}$  – стандартные ошибки соответственно коэффициентов  $a_0$  и  $a_1$  уравнения регрессии. Их вычисление производят по формулам:

$$S_{a_0} = S_y \sqrt{1 - r^2} \sqrt{\frac{[x^2]}{n[x^2] - [x]^2}}, \quad (68)$$

$$S_{a_1} = S_y \sqrt{1 - r^2} \sqrt{\frac{n}{n[x^2] - [x]^2}}. \quad (69)$$

Коэффициенты  $a_0$  и  $a_1$  считаются значимыми, если  $2S_{a_i} < |a_i|$ . Если же коэффициенты  $a_0$  и  $a_1$  незначимы, то ситуацию можно поправить путем увеличения объема выборки  $n$ , увеличения числа факторов, включаемых в модель или изменения формы уравнения связи [1].

### 3.8. Лабораторная работа № 3

#### Построение модели линейной корреляции по несгруппированным данным

*Цель работы:* овладение способами построения моделей линейной корреляции для несгруппированных данных, выработка умения и навыков оценки надежности коэффициента корреляции, уравнения регрессии и его коэффициентов.

#### *Задание к лабораторной работе №3*

На основании опытных данных требуется:

1. Построить корреляционное поле. По характеру расположения точек в корреляционном поле выбрать общий вид регрессии.
2. Вычислить числовые характеристики  $\bar{x}$ ,  $\bar{y}$ ,  $S_x$ ,  $S_y$ ,  $r$ ,  $\sigma_r$ .
3. Определить значимость коэффициента корреляции  $r$  и найти для него доверительный интервал с надежностью  $\gamma = 0,95$ .
4. Написать эмпирические уравнения линий регрессий  $y$  на  $x$  и  $x$  на  $y$ .
5. Вычислить коэффициент детерминации  $R^2$  и объяснить его смысловое значение.
6. Проверить адекватность уравнения регрессии  $y$  на  $x$ .
7. Провести оценку величины погрешности уравнения регрессии  $y$  на  $x$  и его коэффициентов.
8. Построить уравнение регрессии  $y$  на  $x$  в первоначальной системе координат.

#### 3.8.1. Пример выполнения лабораторной работы №3

*Задача.* Результаты наблюдений зависимости средней заработной платы  $Y$  тыс. руб. от производительности труда  $X$  тыс. руб. по цеху технологической связи ТПЭУС № 1 по кварталам приведены в табл. 23.

Таблица 23

X	24,3	24,9	28,1	30,5	31,5	39,3	40,2	43,5	45,4	45,9
Y	8,2	8,6	8,7	8,9	9,1	10,6	11,3	11,8	12,9	13,1

1. Для решения поставленной задачи методами корреляционного анализа определим, какой из указанных в условии показателей выбрать за факторный признак, а какой за результативный. На основании, например, экономического анализа производственной деятельности и взаимосвязи производительности труда и средней заработной платы следует, что за факторный признак  $X$  следует принять производительность труда, а среднюю зарплату за результативный признак  $Y$ .

Для определения формы связи между признаками  $X$  и  $Y$  строим на координатной плоскости точки  $(x_i, y_i)$ , пользуясь табл. 23. Около построенных точек проводим так называемую линию тренда (на рис. 5 – пунктирная линия) так, чтобы построенные точки были расположены как можно ближе к ней. По расположению точек около линии тренда делаем вывод о том, что связь между производительностью труда и средней зарплатой может носить линейный характер.

2. Произведем расчет статистик  $\bar{x}$ ,  $\bar{y}$ ,  $S_x$ ,  $S_y$ ,  $r$ , которые войдут в уравнения линий регрессий. Составим расчетную табл. 24.

Таблица 24

$x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$x^2$	$xy$
24,3	- 11,06	122,3236	8,2	-2,12	4,4944	590,49	199,26
24,9	- 10,46	109,4116	8,6	- 1,72	2,9584	620,01	214,14
28,1	- 7,26	52,7076	8,7	- 1,62	2,6244	789,61	244,47
30,5	- 4,86	23,6196	8,9	- 1,42	2,0164	930,25	271,45
31,5	- 3,86	14,8996	9,1	- 1,22	1,4884	992,25	286,65
39,3	3,94	15,5236	10,6	0,28	0,0784	1544,49	416,58
40,2	4,84	23,4256	11,3	0,98	0,9604	1616,04	454,26
43,5	8,14	66,2596	11,8	1,48	2,1904	1892,25	513,3
45,4	10,04	100,8016	12,9	2,58	6,6564	2061,16	585,66
45,9	10,54	11,0916	13,1	2,78	7,7284	2106,81	601,29
353,6		640,064	103,2		31,196	13143,36	3787,06

Пользуясь результатами последней строки табл. 24, согласно (2), (49), (50) – (53), находим:

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} 353,6 = 35,36$  – средняя производительность труда;

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{10} 103,2 = 10,32$  – средняя зарплата сотрудников цеха

технологической связи.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{9} 640,064 = 71,118, \quad S_x = \sqrt{S_x^2} = 8,43;$$

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{9} 31,196 = 3,466 \quad S_y = \sqrt{S_y^2} = 1,87;$$

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{10} \cdot 3787,06 = 378,71;$$

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x S_y} = \frac{378,71 - 35,36 \cdot 10,32}{8,43 \cdot 1,87} = 0,87.$$

Проверяем значимость коэффициента корреляции. Вычислим статистику  $t_p$  по формуле (59):

$$t_p = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,87 \sqrt{10-2}}{\sqrt{1-0,87^2}} = 5,02.$$

По таблице критических точек распределения Стьюдента (приложение 6) по уровню значимости  $\alpha = 0,05$  и числу степеней свободы  $k = n - 2 = 10 - 2 = 8$  находим  $t_{\alpha} = t_{\alpha;k} = t_{0,05;8} = 2,31$ . Так как  $t_p = 5,02 > t_{\alpha}$ , то выборочный коэффициент корреляции значимо отличается от нуля. Следовательно, можно предположить, что средняя зарплата  $Y$  и производительность  $X$  труда в цехе связаны линейной корреляционной зависимостью, и провести приблизительную прямую  $c$  на числовой плоскости (рис. 5).

Находим доверительный интервал для выборочного коэффициента корреляции  $r$  с надежностью  $\gamma = 0,95$ . Так как объем выборки  $n = 10 < 50$ , то доверительный интервал находим по формуле (61):

$$r - t_{\gamma} \cdot \sigma_r \leq \hat{r} \leq r + t_{\gamma} \cdot \sigma_r$$

Так как по условию надежность (доверительная вероятность) равна  $\gamma = 0,95$ , то по таблице функции Лапласа (приложение 2) находим  $t_{\gamma} = 1,96$ .

Вычисляем среднюю квадратическую ошибку  $\sigma_r$  по формуле (60):

$$\sigma_r = \frac{1-r^2}{\sqrt{n-2}} = \frac{1-0,87^2}{\sqrt{10-2}} = 0,08.$$

Записываем доверительный интервал:

$$0,87 - 1,96 \cdot 0,08 \leq \hat{r} \leq 0,87 + 1,96 \cdot 0,08.$$

Следовательно, с вероятностью 0,95 линейный коэффициент корреляции генеральной совокупности находится в пределах от 0,71 до 1.

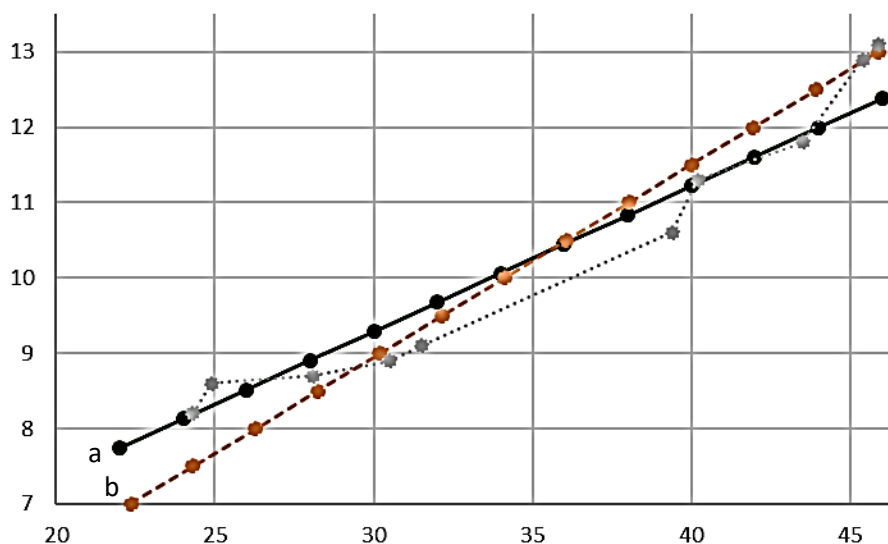


Рис. 5. Линейная регрессия  $Y$  на  $X$  – прямая  $a$  с уравнением  $y = 0,193x + 3,5$  и линейная регрессия  $X$  на  $Y$  – прямая  $b$  с уравнением  $x = 3,922y - 5,1$ . Точечная ломанная проведена «от руки»

Применительно к решаемой задаче полученный результат означает, что по имеющейся выборке следует ожидать влияние производительности труда на рост средней зарплаты работников цеха технологической связи не менее чем на 71 %.

Найдем эмпирические линейные уравнения регрессии  $y$  на  $x$  и  $x$  на  $y$ , которые являются приближенными уравнениями для истинных уравнений регрессий.

Уравнение регрессии  $y$  на  $x$  по (47):

$$\hat{y}_x = \bar{y} + r \frac{s_y}{s_x} (x - \bar{x}), \quad \hat{y}_x = 10,32 + 0,87 \frac{1,87}{8,43} (x - 35,36);$$

$$\hat{y}_x = 0,192989x + 3,495898.$$

Уравнение регрессии  $x$  на  $y$  по (48):

$$\hat{x}_y = \bar{x} + r \frac{s_x}{s_y} (y - \bar{y}), \quad \hat{x}_y = 35,36 + 0,87 \frac{8,43}{1,87} (y - 10,32);$$

$$\hat{x}_y = 3,921979y - 5,114819.$$

Контроль вычислений:

$$a_1 b_1 = 0,192989 \cdot 3,921979 = 0,7569,$$

$$r^2 = (0,87)^2 = 0,7569.$$

Так как условие  $a_1 b_1 = r^2$  выполняется, то вычисления выполнены верно.

Из уравнения  $\hat{y}_x = 0,192989x + 3,495898$  следует, что при увеличении производительности труда на 1 тыс. руб. средняя зарплата работников цеха технологической связи возрастает на 192,989 рублей. Этот результат следует учесть на предприятии при разработке мероприятий по стимулированию производственной деятельности работников цеха в условиях рыночных отношений.

Подставляя  $\bar{x} = 35,36$ ,  $\bar{y} = 10,32$  в уравнения регрессий, получаем точки, координаты которых совпадают с координатами центра распределения  $C(\bar{x}, \bar{y})$ . Следовательно, линии регрессий пересекаются в точке  $C(\bar{x}, \bar{y})$ .

Найдем коэффициент детерминации. Для линейной корреляции при вычисленном коэффициенте  $r$  он равен  $r^2$ . У нас  $r^2 = 0,76$ . Это означает, что 76 % рассеивания средней зарплаты работников технологического цеха связи объясняется линейной корреляционной зависимостью между средней зарплатой и производительностью труда и только 24 % рассеивания средней зарплаты работников технологического цеха остались необъяснимыми. Такое положение могло произойти из-за того, что в модель не включены другие факторы, влияющие на изменение средней зарплаты работников технологического цеха связи, либо опытных данных в данной выборке недостаточно, чтобы построить более надежное уравнение регрессии.

Проверим адекватность уравнения линейной регрессии  $y$  на  $x$  по критерию Фишера – Снедекора. Вычислим статистику  $F_H$  по формуле (64):

$$F_H = \frac{R^2(n-2)}{1-R^2}, \text{ где по (63) } R^2 = 1 - \frac{\sum(y_i - \hat{y}_{x_i})^2}{\sum(y_i - \bar{y})^2}.$$



Для нахождения суммы  $\sum(y_i - \hat{y}_{x_i})^2$  составляем табл. 25. Из табл. 24 и 25 находим:  $\sum(y_i - \bar{y})^2 = 31,196$ ,  $\sum(y_i - \hat{y}_{x_i})^2 = 1,8729$ .

Тогда

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_{x_i})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{1,8729}{31,196} = 0,94, F_H = \frac{0,94(10-2)}{1-0,94} = 125,3.$$

При уровне значимости  $\alpha = 0,05$  и числах степеней свободы  $k_1 = 1$ ,  $k_2 = n - 2$   $k_2 = 10 - 2 = 8$  по таблице критических точек распределения Фишера – Снедекора (приложение 7) находим  $F_T = F_{\alpha; k_1; k_2} = F_{0,05; 1; 8} = 5,32$ .

Так как  $F_H = 125,3 > 5,32$ , то заключаем, что уравнение линейной регрессии  $\hat{y}_x = 0,192989x + 3,495898$  статистически значимо описывает результаты эксперимента.

Таблица 25

$y_i$	$\hat{y}_x$	$y_i - \hat{y}_x$	$(y_i - \hat{y}_x)^2$
8,2	8,18	0,02	0,0004
8,6	8,3	0,3	0,09
8,7	8,9	-0,2	0,04
8,9	9,4	-0,5	0,25
9,1	9,6	-0,5	0,25
10,6	11,1	-0,5	0,25
11,3	11,25	0,05	0,0025
11,8	11,9	-0,1	0,01
12,9	12,2	0,7	0,49
13,1	12,4	0,7	0,49
			<b>1,8729</b>

Проведем оценку величины погрешности уравнения регрессии  $\hat{y}_x = 0,192989x + 3,495898$ . Найдем относительную погрешность  $\delta$  уравнения по формуле (67), % :

$$\delta = \frac{\sigma_u}{\bar{y}} 100,$$

где  $\sigma_u = \sqrt{D_u} = \sqrt{\frac{\sum(u_i - \bar{u})^2}{n-2}}$ ,  $u_i = y_i - \hat{y}_{x_i}$ ,  $\bar{u} = \frac{1}{n} \sum(y_i - \hat{y}_{x_i})^2$ .

Так как  $\sum(y_i - \hat{y}_{x_i})^2 = 1,8729$ , то  $\bar{u} = 0,19$ .

Для нахождения суммы  $\sum(u_i - \bar{u})^2$  составляем табл. 26.

Тогда  $\sigma_u = \sqrt{\frac{2,2588}{8}} = 0,53$ ,  $\delta = \frac{0,53}{10,32} 100 = 5 \%$ .

Так как величина  $\delta$  мала, то уравнение линейной регрессии  $\hat{y}_x = 0,192989x + 3,495898$  хорошо описывает опытные данные.

Таблица 26

$u_i = y_i - \hat{y}_{x_i}$	$(u_i - \bar{u})$	$(u_i - \bar{u})^2$
0,02	- 0,17	0,0289
0,03	- 0,16	0,0256
- 0,2	- 0,39	0,1521
- 0,5	- 0,69	0,4761
- 0,5	- 0,69	0,4761
- 0,5	- 0,69	0,4761
0,05	- 0,14	0,0196
- 0,1	- 0,29	0,0841
0,7	0,51	0,2601
0,7	0,51	0,2601
		2,2588

Оценим коэффициенты уравнения регрессии. У нас  $a_0 = 3,495898$ ,  $a_1 = 0,192989$ .

Для нахождения отношений  $S_{a_0}/|a_0|$  и  $S_{a_1}/|a_1|$  вычислим средние квадратические ошибки коэффициентов по формулам (68) и (69):

$$S_{a_0} = S_y \sqrt{1 - r^2} \sqrt{\frac{[x^2]}{n[x^2] - [x]^2}}, S_{a_1} = S_y \sqrt{1 - r^2} \sqrt{\frac{n}{n[x^2] - [x]^2}}.$$

По табл. 24 находим:  $[x] = 353,6$ ,  $[x^2] = 13143,36$ . Учитывая, что  $n = 10$ ,  $r^2 = 0,7569$  и  $S_y = 1,87$ , находим:

$$S_{a_0} = 1,87 \sqrt{1 - 0,7569} \sqrt{\frac{13143,36}{10 \cdot 13143,36 - (353,6)^2}} = 0,71;$$

$$S_{a_1} = 0,4931 \sqrt{\frac{10}{10 \cdot 13143,36 - (353,6)^2}} = 0,04.$$

Так как  $S_{a_0}/|a_0| = 0,71/3,495898 < 0,5$  и  $S_{a_1}/|a_1| = 0,04/0,192989 < 0,5$ , то коэффициенты  $a_0$  и  $a_1$  уравнения регрессии  $y$  на  $x$  значимы. Графиками найденных регрессий являются прямые  $a$ ,  $b$ , представленные на рис. 5.

Таким образом, уравнение регрессии  $\hat{y}_x = 0,192989x + 3,495898$ , описывающее зависимость средней зарплаты работников цеха технологической связи от производительности труда, значимо описывает опытные данные и может быть принято для практического руководства.

### 3.8.2. Варианты заданий лабораторной работы № 3

#### Вариант № 1

Имеются данные, характеризующие зависимость между стоимостью  $X$  основных производственных фондов и объемом  $Y$  валовой продукции по десяти однотипным предприятиям:

$X$	1	2	3	4	5	6	7	8	9	10
$Y$	20	25	31	31	40	56	52	60	60	70

Найти эмпирическую формулу функциональной зависимости стоимости основных производственных фондов и объема валовой продукции на предприятиях. Построить эмпирическую и теоретическую линии.

#### Вариант № 2

Зависимость между стоимостью  $X$  тыс. млн руб. основных средств предприятий и месячным выпуском  $Y$  тыс. руб. продукции характеризуется следующими данными:

$X$	1	2	3	4	5	6	7
$Y$	10	12	28	40	42	52	54

Найти эмпирическую формулу функциональной зависимости месячного выпуска продукции и стоимости основных средств предприятий. Построить эмпирическую и теоретическую линии.

#### Вариант № 3

Имеются данные наблюдений изменения средней заработной платы  $Y$  руб. в зависимости от изменения производительности труда  $X$  шт. за 4 месяца 1992 года по девяти токарям цеха № 23 электромеханического завода:

$X$	406	660	914	1168	1422	1676	1930	2184	2438
$Y$	518,5	813,5	1108,5	1403,5	1698,5	1993,5	2288,5	2583,5	2878,5

Найти эмпирическую формулу функциональной зависимости изменения средней заработной платы и изменения производительности труда.

#### Вариант № 4

Дано распределение заводов по производственным средствам  $X$  млн руб. и по суточной выработке  $Y$  тыс. руб.:

X	50	49	48	51	52	53	54	57	59	60	61	55	60	62	63
Y	10	8	10	9	10	12	13	15	16	18	20	17	21	25	24

Найти эмпирическую формулу функциональной зависимости производственных средств и суточной выработки.

#### Вариант № 5

Данные нормы расхода моторного масла на угар и замену  $Y$  л/100 л.т. от максимальной мощности двигателя  $X$  л.с. приведены в таблице:

X	39	42	53	70	73,5	75	90	98	110	115
Y	1,3	1,3	0,8	2,2	1,8	2	2,2	1,8	2,8	2,1

Найти эмпирическую формулу функциональной зависимости расхода моторного масла и максимальной мощности двигателя.

#### Вариант № 6

Результаты наблюдений изменения диаметра  $Y$  мм вала и износа  $X$  мм резца приведены в следующей таблице:

X	0,01	0,02	0,024	0,03	0,032	0,035	0,037	0,042	0,048	0,057
Y	20	20,01	20,014	20,022	20,024	20,027	20,029	20,034	20,04	20,049

Найти эмпирическую формулу функциональной зависимости диаметра вала и износа резца.

#### Вариант № 7

Компанию по прокату автомобилей интересует зависимость между пробегом  $X$  автомобилей и стоимостью  $Y$  ежемесячного технического обслуживания. Для выяснения характера этой связи было отобрано 15 автомобилей. Данные приведены в таблице:

X	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	13	16	15	20	19	21	26	24	30	32	30	35	34	40	39

Найти эмпирическую формулу функциональной зависимости пробега автомобилей и стоимости ежемесячного технического обслуживания.

#### Вариант № 8

При исследовании зависимости между выпуском готовой продукции  $Y$  тыс. руб. и энерговооруженностью труда  $X$  (кВт-ч) получены следующие данные:

X	1201	1300	1375	1412	1443	1500	1526	1516	1718	1783	1819	1877
Y	14	21	27	29	30	31,5	35	34	41	38	39	46

Найти эмпирическую формулу функциональной зависимости между выпуском готовой продукции и энерговооруженностью труда.

#### Вариант № 9

В таблице приведены данные, характеризующие зависимость израсходованных долот  $Y$  шт. при бурении 8 скважин в зависимости от механической скорости  $X$  м/с проходки:

X	10	15	8	12	16	18	22	25
Y	55	40	60	50	40	30	25	20

Найти эмпирическую формулу зависимости  $Y$  от  $X$ .

#### Вариант № 10

Скорость  $Y$  (м/час) бурения в твердых породах от нагрузки  $X$  (атм.) на долото характеризуется следующими данными:

X	10	10,5	11	11,5	12	12,5	13	13,5
Y	3,5	2,5	2,5	1,5	1	0,5	1	0,5

Найти эмпирическую формулу зависимости  $Y$  от  $X$ .

#### Вариант № 11

Зависимость между выработкой продукции  $X$  тыс. руб. и затратами топлива  $Y$  в условных единицах характеризуется следующими данными:

X	3	4	5	6	7	8	9	10	11	12
Y	2	2,3	2,2	2,3	2,5	2,6	2,5	2,6	2,8	2,9

Найти эмпирическую формулу функциональной зависимости между выработкой продукции и затратами топлива.

#### Вариант №12

Имеются данные распределения заводов по производственным средствам  $X$  млн руб. и по суточной выработке  $Y$  млн руб.:

X	48	49	50	52	53	54	57	60	63	65	68	80
Y	11	8	10	10	12	13	15	18	24	26	25	38

Найти эмпирическую формулу функциональной зависимости распределения заводов по производственным средствам и суточной выработки.

### Вариант № 13

Найти формулу, устанавливающую зависимость между коэффициентом  $Y$  сменности техники и ее средним возрастом  $X$  по предприятию ПМК-7 объединения Сибкомплектмонтаж на основании следующих данных:

$Y$	1,18	1,21	1,25	1,26	1,3	1,32	1,33	0,69	0,72	0,8
$X$	6,31	5,8	5,1	5,6	6,1	6,5	6,55	3,8	3,41	4

### Вариант № 14

Имеются данные о реализации продукции  $X$  млн руб. и накладных расходах  $Y$  тыс. руб. на реализацию:

$X$	9	13	17	22	29	36	44	51	60	65
$Y$	27	36	29	41	54	71	65	81	90	95

Найти эмпирическую формулу функциональной зависимости реализации продукции и накладными расходами по реализации.

### Вариант № 15

Зависимость линейной нормы расхода топлива  $Y$  л. от максимальной мощности двигателя автомобиля  $X$  л.с. характеризуется следующими данными:

$X$	39	53	70	75	90	98	110	115	120	150
$Y$	12	11	21,5	22,8	18	21	31	25	30	36

Найти эмпирическую формулу функциональной зависимости нормы расхода топлива и мощности двигателя автомобиля.

### Вариант № 16

Имеются данные нормы расхода моторных масел на угар и замену  $Y$  л/100 л.т. в зависимости от максимального крутящего момента  $X$ :

$X$	7,4	7,6	8,2	10,8	17	21	35	41	42	44
$Y$	1,3	1,3	1,5	1,8	2,2	2,1	2,8	2,6	2,3	2,4

Найти эмпирическую формулу функциональной зависимости нормы расхода моторных масел от максимального крутящего момента.

### Вариант № 17

Фазовая проницаемость нефти  $Y$  и насыщенность породы нефтью  $X$  характеризуются следующими данными:

X	0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
Y	0,35	0,45	0,55	0,65	0,75	0,8	0,85	0,95	1	1,25

Найти эмпирическую формулу функциональной зависимости фазовой проницаемости нефти и насыщенности породы нефтью.

#### Вариант № 18

Имеются данные нормы расхода моторных масел на угар и замену  $Y$  л/100 л.т. в зависимости от максимальной мощности двигателя автомобиля  $X$  л.с.:

X	39	42	53	70	75	90	110	115	150	170
Y	1,3	1,3	0,9	2,2	2,2	2,2	2,8	2,4	2,5	2,6

Найти эмпирическую формулу функциональной зависимости нормы расхода моторных масел и максимальной мощности двигателя автомобиля.

#### Вариант № 19

Зависимость между среднегодовой стоимостью основных производственных фондов  $X$  млн руб. и стоимостью товарной продукции  $Y$  млн руб. характеризуется следующими данными:

X	1,94	2,68	3,47	4,12	4,77	5,34	5,85	6,65
Y	0,82	0,97	1,06	1,08	1,1	1,14	1,21	1,25

Найти эмпирическую формулу функциональной зависимости между среднегодовой стоимостью основных производственных фондов и стоимостью товарной продукции.

#### Вариант № 20

Найти формулу, устанавливающую зависимость себестоимости одной тонны нефти  $Y$  в руб. от затрат  $X$  в тыс. руб. на одну тонну по следующим данным:

X	1,44	1,6	1,85	2,1	2,25	2,42	2,55	2,65
Y	161,5	165	170	175	178	182	186	190

#### Вариант № 21

Ниже приводятся данные о производительности труда  $Y$  м на одного чел/ч и стаже рабочих  $X$  (в годах):

$X$	1	2	3	4	5	6	7	8
$Y$	9,8	15	16	19	20	22	23	27

Найти эмпирическую формулу функциональной зависимости производительности труда и стажа рабочих.

Вариант № 22

Найти формулу зависимости электрического сопротивления  $R$  Ом проводника от температуры  $Q$  °C по следующим данным:

$Q$	19,1	25,0	30,1	36,0	40,0	45,1	50,0
$R$	76,30	77,80	79,75	80,80	82,35	83,90	85,10

Построить теоретическую и эмпирическую линии.

Вариант № 23

Найти зависимость израсходованных долот  $Y$  при бурении 10 скважин в зависимости от механической скорости  $X$  проходки на основании следующих данных:

$X$	8	10	12	14	15	16	18	19	20	21
$Y$	60	55	50	48	45	40	30	31	29	32

Вариант № 24

Зависимость между пористостью  $Y$  пород и их газонасыщенностью  $X$  задана таблицей:

$X$	10	18	25	36	43	54	62	70	75	80
$Y$	12	16	21	24	27	29	31	34	37	42

Найти эмпирическую формулу функциональной зависимости между пористостью и газонасыщенностью пород.

Вариант № 25

В таблице приведены данные, характеризующие зависимость растворимости азотно-натриевой соли  $S$  в зависимости от температуры  $T$ :

$T$	0	4	10	15	21	29	36	51	68
$S$	66,7	71	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Найти эмпирическую формулу функциональной зависимости растворимости азотно-натриевой соли  $S$  и температуры.



### Вариант № 26

При исследовании зависимости хода  $Y$  мм поршня двигателя автомобиля от максимального крутящего момента  $X$  получены следующие данные:

$X$	7,4	8,2	10,8	17	18,4	21	29	41	46	47,5
$Y$	66	67	70	92	92	110	85	95	88	88

Найти эмпирическую формулу функциональной зависимости хода поршня двигателя и максимального крутящего момента.

### Вариант № 27

Результаты исследования зависимости между средней месячной выработкой  $Y$  млн руб. продукции на одного рабочего и стоимостью  $X$  млн руб. основных производственных средств приведены в таблице:

$X$	9,9	10,0	10,1	10,2	10,3	10,4	10,5	10,6
$Y$	0,8	0,9	0,95	1,1	1,25	1,2	1,28	1,32

Найти эмпирическую формулу функциональной зависимости между средней месячной выработкой продукции и стоимостью основных производственных средств.

### Вариант № 28

При исследовании зависимости между выработкой  $Y$  тыс. руб. и энерговооруженностью  $X$  кВт/ч труда получены следующие данные:

$X$	5	3	6	7	10	4	11	9	2	8
$Y$	6,3	3,6	7,5	8,5	12,5	6,2	12,6	10,7	2,6	7,7

Найти эмпирическую формулу функциональной зависимости между выработкой и энерговооруженностью труда

### Вариант № 29

При изучении зависимости между производительностью  $Y$  тн/чел. труда и дебитом  $X$  тн/сут. скважин получены следующие результаты:

$X$	51	40	46	42	43	44	45	47	49	50
$Y$	30	33	30,8	32	31,5	33	33,5	32,5	31,2	31

Найти эмпирическую формулу функциональной зависимости между производительностью труда и дебитом скважин.

Вариант № 30.

При исследовании зависимости времени  $t$  с, затрачиваемого на закрепление детали, от ее веса  $P$  кг получены следующие результаты:

$P$	51	40	46	42	43	44	45	47	49	50
$t$	30	33	30,8	32	31,5	33	33,5	32,5	31,2	31

Найти эмпирическую формулу функциональной зависимости времени, затрачиваемого на закрепление детали, от ее веса.

### 3.9. Лабораторная работа № 4

**Построение выборочного уравнения линии регрессии по сгруппированным данным**

*Цель работы:* овладение способами построения моделей линейной корреляции для сгруппированных данных по методу наименьших квадратов с использованием коэффициента линейной корреляции, выработка умения и навыков оценки надежности уравнения регрессии и его коэффициентов.

Задание к лабораторной работе: по опытным данным требуется:

1. Построить корреляционное поле. По характеру расположения точек в корреляционном поле выбрать общий вид регрессии.
2. Написать уравнение линии регрессии  $y$  на  $x$  по методу наименьших квадратов и  $s$  с использованием коэффициента корреляции  $r$ . Сравнить полученные уравнения и сделать вывод о выборе одного из них.
3. Оценить тесноту связи между признаками  $X$  и  $Y$  с помощью выборочного коэффициента корреляции  $r$  и его значимость.
4. Проверить адекватность модельного уравнения регрессии  $y$  на  $x$ , записанного через коэффициент корреляции  $r$ .
5. Проверить надежность уравнения регрессии  $y$  на  $x$ , записанного через коэффициент корреляции  $r$ , и его коэффициентов.
6. Построить уравнения регрессий в первоначальной системе координат.

### 3.9.1. Пример выполнения лабораторной работы №4

*Задача.* Валики при черновой обработке на станке № 1 передаются последовательно на станок № 2 для чистовой обработки. Экспериментатор, изучающий зависимость между отклонениями размеров валиков от номинала при черновой обработке (мкм), от номинала при чистовой обработке (мкм), произвел измерения отклонений у 50 случайно отобранных валиков. Результаты измерений сведены в табл. 27.

Таблица 27

$x$	$y$				$n_y$
	-30	-20	-10	0	
	1				1
-4	4	1			5
0	1	15	1		17
4		2	13		15
8			2	1	3
12				9	9
$n_x$	6	18	16	10	50

Выполнение работы

Пусть признак  $X$  характеризует отклонение размеров валиков от номинала при черновой обработке, а признак  $Y$  отклонение размеров валиков от номинала при чистовой обработке. Используя данные табл. 27, строим корреляционное поле.

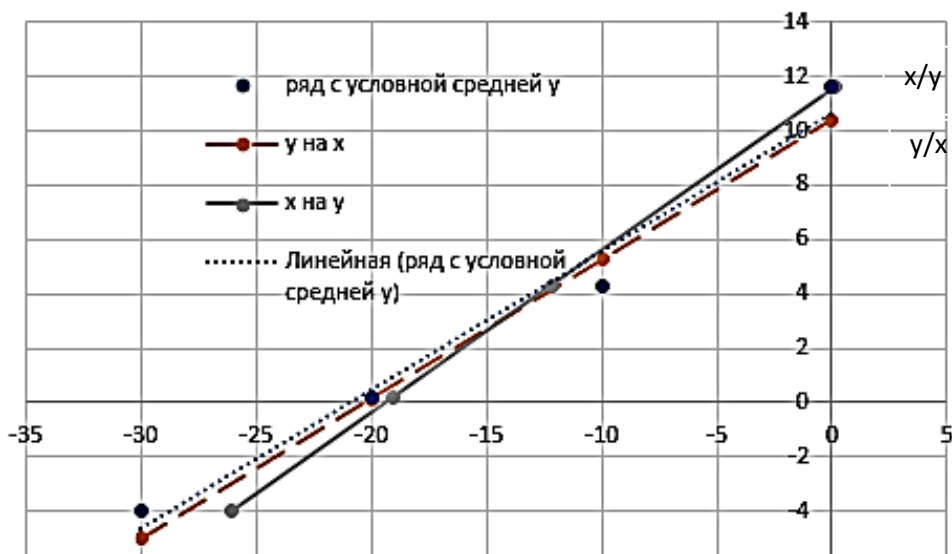


Рис. 6. Выявление линейной регрессии. От руки проводится линия тренда (штриховая прямая). Прямая  $(y/x)$  – регрессия  $y$  на  $x$ . Прямая  $(x/y)$  – регрессия  $x$  на  $y$

Проведя линию тренда (пунктирная линия), видим, что число точек, расположенных над и под ней, практически одинаково, причем расстояния этих точек до линии тренда одинаковые. Это дает основание предположить наличие линейной зависимости между признаками  $X$  и  $Y$ . Для подтверждения этой гипотезы перейдем от данного распределения к новому, найдя для каждого значения признак  $X$  условное среднее признака по формуле (42):

$$\bar{y}_{x_j} = \frac{\sum n_{ij}y_i}{n_{x_j}}.$$

При  $x_1 = -30$ ,  $\bar{y}_{x_1} = \frac{1 \cdot (-8) + 4 \cdot (-4) + 1 \cdot 0}{6} = -46 = -4$ .

При  $x = -20$ ,  $\bar{y}_{x_2} = \frac{1 \cdot (-4) + 15 \cdot 0 + 2 \cdot 4}{18} = 0,2$ .

При  $x = -10$ ,  $\bar{y}_{x_3} = \frac{1 \cdot 0 + 13 \cdot 4 + 2 \cdot 8}{16} = 4,3$ .

При  $x = 0$ ,  $\bar{y}_{x_4} = \frac{1 \cdot 8 + 9 \cdot 12}{10} = 11,6$ .

Строим точки с координатами  $(x_j; \bar{y}_{x_j})$  (рис. 7).

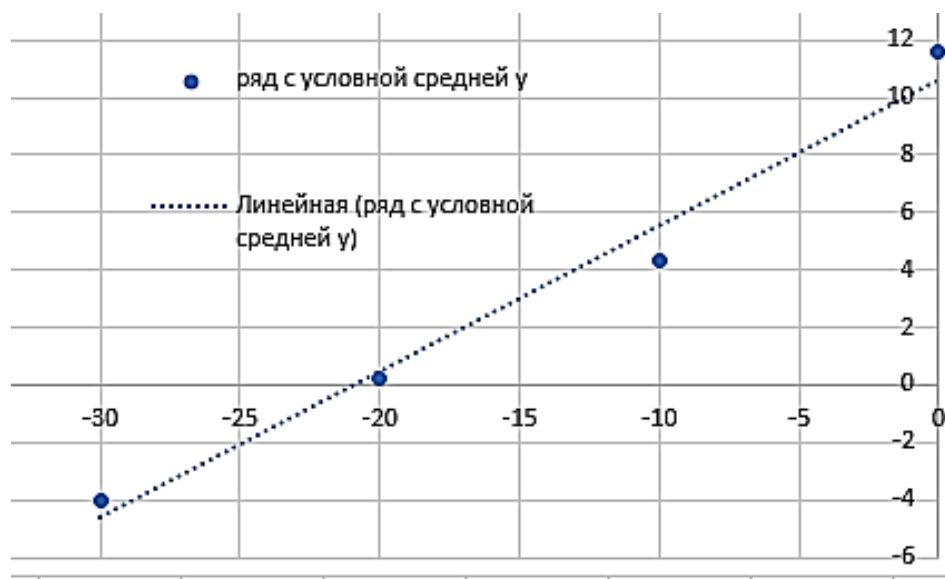


Рис.7 Выявление корреляционной зависимости

Из рис. 7 видно, что отклонения точек от построенной прямой незначительны. Следовательно, связь между признаками  $X$  и  $Y$  может носить линейный характер. Составим уравнения линий регрессий  $y$  на  $x$  по методу наименьших квадратов и через коэффициент линейной корреляции  $r$ .

Применим метод наименьших квадратов к нахождению коэффициентов  $a_0$  и  $a_1$  уравнения линейной регрессии  $\bar{y}_x = a_1 x + a_0$ . Решаем систему нормальных уравнений (45):

$$\begin{cases} na_0 + [n_x x]a_1 = [n_y y], \\ [n_x x]a_0 + [n_x x^2]a_1 = [n_{xy} xy], \end{cases}$$

Для нахождения сумм, входящих в систему, составляем табл. 28.

Таблица 28

x	y					$n_y$	$n_y y$
	-30	-20	-10	0			
-8	1					1	-8
-4	4	1				5	-20
0	1	15	1			17	0
4		2	13			15	60
8			2	1		3	24
12				9		9	108
$n_x$	6	18	16	10		50	164
$n_x x$	-180	-360	-160	0		-700	
$n_x x^2$	5400	7200	1600	0		14200	
$n_{xy} xy$	720	-80	-680	0		-40	

Полученная из табл. 28 система

$$\begin{cases} 50a_0 - 700a_1 = 164, \\ -700a_0 + 14200a_1 = -40 \end{cases}$$

имеет решение  $(a_0, a_1) = (0,512728; 10,458192)$ . Тогда уравнение линейной регрессии запишется в виде:

$$\bar{y}_x = 0,512728x + 10,458192. \quad (70)$$

Найдем уравнение линейной регрессии  $y$  на  $x$  по формуле (47), используя коэффициент линейной корреляции:

$$\hat{y}_x = \bar{y} + r \frac{S_y}{S_x} (x - \bar{x}). \quad (71)$$

Так как данные выборки для признаков  $X$  и  $Y$  заданы в виде корреляционной таблицы и объем выборки  $n=50$ , то для нахождения величин, входящих в уравнение регрессии, переходим к вспомогательному распределению с условными вариантами  $u_i$  и  $v_j$ . По корреляционной табл. 27

находим наибольшую частоту совместного появления признаков  $X$  и  $Y$ :  $n_{32}=15$ . Тогда  $C_1 = M_0 X = -20$ ,  $C_2 = M_0 Y = 0$ ,  $h_1 = 10$ ,  $h_2 = 4$ . Составляем корреляционную табл. 29 в условных вариантах.

Таблица 29

$u$	$v$				$n_{y}$
	-1	0	1	2	
-2	1				1
-1	4	1			5
0	1	15	1		17
1		2	13		15
2			2	1	3
3				9	9

По табл. 29 находим:

$$\bar{u} = \frac{1}{n} \sum n_{u_j} u_j = \frac{1}{50} (6(-1) + 18 \cdot 0 + 16 \cdot 1 + 10 \cdot 2) = 0,6,$$

$$\bar{v} = \frac{1}{n} \sum n_{v_i} \cdot v_i = \frac{1}{50} (1(-2) + 5(-1) + 17 \cdot 0 + 15 \cdot 1 + 3 \cdot 2 + 9 \cdot 3) = 0,8,$$

$$\overline{u^2} = \frac{1}{n} \sum n_{u_j} \cdot u_j^2 = \frac{1}{50} (6(-1)^2 + 18 \cdot 0^2 + 16 \cdot 1^2 + 10 \cdot 2^2) = 1,24,$$

$$\overline{v^2} = \frac{1}{n} \sum n_{v_i} \cdot v_i^2 = \frac{1}{50} (1(-2)^2 + 5(-1)^2 + 17 \cdot 0^2 + 15 \cdot 1^2 + 3 \cdot 2^2 + 9 \cdot 3^2) = 2,34.$$

Тогда

$$S_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1,24 - 0,36} = 0,94; S_v = \sqrt{\overline{v^2} - (\bar{v})^2} = \sqrt{2,34 - 0,64} = 1,3.$$

Для нахождения суммы  $\sum n_{u_j v_i} u_j v_i$  составляем табл. 30.

Таблица 30

$v$	$u$				
	-1	0	1	2	
-2	1	2			2
-1	4	1			4
0					
1			13	1	13
2			2	2	8
3				9	54
	6		17	58	81

Тогда, согласно (56), (58),

$$r_B = \frac{81 - 50 \cdot 0,6 \cdot 0,8}{50 \cdot 0,94 \cdot 1,3} = 0,93,$$

$$\bar{x} = \bar{u} h_1 + C_1 = 0,6 \cdot 10 - 20 = -14,$$

$$\bar{y} = \bar{v} h_2 + C_2 = 0,8 \cdot 4 + 0 = 3,2,$$

$$S_x = S_u h_1 = 0,94 \cdot 10 = 9,4,$$

$$S_y = S_v h_2 = 1,3 \cdot 4 = 5,2.$$

Отсюда и из (47), (48) следует уравнение линии регрессии  $y$  на  $x$ :

$$\hat{y}_x = 3,2 + 0,93 \frac{5,2}{9,4} (x + 14)$$

или

$$\hat{y}_x = 0,514468x + 10,40255$$

и уравнение линии регрессии  $x$  на  $y$ :

$$\hat{x}_y = -14 + 0,93 \frac{5,2}{9,4} (y - 3,2).$$

или

$$\hat{x}_y = 1,681154y - 19,37969.$$

Проверяем тесноту связи между признаками  $X$  и  $Y$ . Для этого, используя критерий Стьюдента, вычисляем статистику:

$$t_H = \frac{|r_B| \sqrt{n-2}}{\sqrt{1-r_B^2}} = \frac{0,93 \sqrt{50-2}}{\sqrt{1-0,93^2}} = 17,5.$$

При уровне значимости  $\alpha=0,05$  и числе степеней свободы  $k=n-2=50-2=48$  находим по таблице распределения Стьюдента  $t_{\alpha,k} = t = 2,01$ . Так как  $t_H = 17,5 > 2,01$ , то выборочный коэффициент  $r_B$  линейной корреляции значимо отличается от нуля. Следовательно, можно считать, что отклонение размеров валиков от номинала при черновой обработке на станке № 1 и отклонение размеров валиков от номинала при чистовой обработке на станке № 2 связаны линейной корреляционной зависимостью. Дадим интерпретацию, например, уравнению регрессии на  $x$ . Из уравнения регрессии видно, что при отклонении от нормальных размеров валиков при черновой обработке ( $x_j$ ), например, на 10 мкм на станке № 1 отклонение от нормального размера валиков при последующей чистовой обработке на станке № 2 составит  $y_i = 10 \cdot 0,514468 = 5,14468 \approx 5,1$  мкм. Это результат воздействия отклонений при черновой обработке валиков на станке № 1. Фактически отклонение может составить  $y = 5,14468 + 10,40255 = 15,54723$  мкм, что является результатом воздействия неучтенных в модели

факторов, не зависящих от отклонений при черновой обработке. Уравнения линий регрессий построены на рис. 6.

Проверим полученное уравнение регрессии  $y$  на  $x$  на адекватность по критерию Фишера-Снедекора. Вычислим статистику:

$$F_H = \frac{Q_R(n-2)}{Q_e(k-1)}.$$

Составим расчетные табл. 31 и 32.

Таблица 31

$y_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
-8	-11,2	125,44
-4	-7,2	51,84
0	-3,2	10,24
4	0,8	0,64
8	4,8	23,04
12	8,8	77,44
		$Q=288,$

Таблица 32

$\bar{y}_{x_i}$	$\bar{y}_{x_i} - \bar{y}$	$(\bar{y}_{x_i} - \bar{y})^2$
-4	-7,2	51,84
0,2	-3	9
4,3	1,1	1,21
11,6	8,4	70,56
		$Q_R=132,61$

Находим  $Q_e = Q - Q_R = 288,64 - 132,61 = 156,03$ . По условию  $n = 50, k = 4$ .

Тогда  $F_H = \frac{132,61(50-2)}{156,03(4-1)} = 13,03$ .

При уровне значимости  $\alpha = 0,05$  и числах степеней свободы  $k_l = 1, k = n - 2 = 50 - 2 = 48$  по таблице критических точек распределения Фишера-Снедекора находим  $F_T = 4,05$ . Так как  $F_H = 13,03 > 4,05$ , то модель линейной регрессии  $\hat{y}_x = 0,514468x + 10,40255$  согласуется с опытными данными.

Итак, мы получили два уравнения линий регрессии  $y$  на  $x$ , описывающих зависимость между признаками  $X$  и  $Y$ . При подстановке в каждое из них опытных значений признака  $X$  убеждаемся в том, что уравнение (70), полученное по методу наименьших квадратов без использования коэффициента линейной корреляции  $r$ , дает лучшие значения признака  $Y$ , чем уравнение (71). Однако практика решения инженерных задач, связанных с построением моделей парной линейной корреляции [1], показывает, что не всегда уравнения, полученные без применения коэффициента корреляции  $r$ , лучше уравнений с



его применением. Поэтому, проводя моделирование в случае сгруппированных экспериментальных данных, из двух линий регрессий следует выбирать лучшую из них.

### 3.9.2. Варианты заданий лабораторной работы № 4

#### Вариант № 1

Зависимость объема  $Y$  см<sup>3</sup> разрушенной породы от глубины  $X$  мм внедрения зуба при постоянном давлении приведены в корреляционной таблице:

$X$	$Y$					
	0,5	1,5	2,5	3,5	4,5	$n_y$
0,05	4	3				7
0,15		5	3			8
0,25			8	1		9
0,35			4	6		10
0,45				5	1	6
0,55				6		6
0,65				1	3	4
$n_x$	4	8	15	19	4	50

#### Вариант № 2

Распределение предприятий по объему продукции  $X$  тыс. руб. и по ее себестоимости  $Y$  руб. приведено в корреляционной таблице:

$X$	$Y$					
	1	2	3	4	5	$n_y$
2				3	4	7
2,5			5	3	5	13
3		3	6	4		13
3,5	4	4	2	2		12
4	1	4				5
$n_x$	5	11	13	12	9	50

#### Вариант № 3

Результаты исследования зависимости среднегодового перевыполнения нормы  $Y$  % от стажа  $X$  в годах работы приведены в корреляционной таблице:

$X$	$Y$						
	2	3	4	5	6	7	$n_y$
5	1	1	1				3
6	4	2	1				7
7		2	8	6			16
8			3	9	2		14
9				1	5	2	8
10					1	1	2
$n_x$	5	5	13	16	8	3	50

Вариант № 4

Фонтанную скважину исследовали на приток  $Y$  нефти при различных режимах работы с замерами забойных давлений  $X$  глубинным манометром.

Данные замеров приведены в корреляционной таблице:

$X$	$Y$									$n_y$
	125	135	145	155	165	175	185	195	205	
11	3	4								7
12		5	4							9
13			3	5						8
14			5	6						11
15				2	18					20
16					4	14				18
17						7	2			9
18							4	6		10
19								2	6	8
$n_x$	3	9	12	13	22	21	6	8	6	100

Вариант № 5

Зависимость коэффициента обрабатываемости  $Y$  от ударной вязкости  $X$  кг/мм<sup>2</sup> инструментальных быстродайствующих сталей задана корреляционной таблицей:

$X$	$Y$					$n_y$
	0,7	0,9	1,1	1,3	1,5	
0,5	1	2				3
0,6	7	10	3			20
0,7		1	12	1		14
0,8				4	3	7
0,9				2	4	6
$n_x$	8	13	15	7	7	50

Вариант № 6

Результаты исследования зависимости между среднемесячной выработкой продукции на одного рабочего  $Y$  тыс. руб. и стоимостью основных производственных средств  $X$  млн руб. приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	0,8	0,9	1	1,1	1,2	1,3	
9,9	2						2
10	2	7	3			1	10
10,1		2	1	1			3
10,2		1	3	9			13
10,3			4	3	6		13
10,4					4	5	9
$n_x$	4	10	8	12	10	6	50

### Вариант № 7

Распределение цилиндрических болванок по длине  $X$  см и по весу  $Y$  кг приведено в корреляционной таблице:

$X$	$Y$							$n_y$
	1,05	1,15	1,25	1,35	1,45	1,55	1,65	
19,5	4	1						5
22,5		2	4	1				7
25,5			3	8	1			12
28,5				2	3	2		7
31,5					2	4		6
34,5						3	3	6
37,5							4	4
$n_x$	4	3	7	11	6	9	7	47

### Вариант № 8

Результаты замера температуры  $X$  °С смазочного масла в двигателе и температуры  $Y$  °С масла в коробке передач автомобиля приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	6	10	14	18	22	26	
15	4	3					7
21	1	5	1				7
27		2	5				7
33			6	3			9
39				5	4		9
45					2	9	11
$n_x$	5	10	12	8	6	9	50

### Вариант № 9

Результаты измерений сверл по твердости  $Y$  (HRC) и по стойкости  $X$  ч приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	20,5	25,5	30,5	35,5	40,5	45,5	
25	2	3					5
35	1	7					8
43			2	4			6
51			1	9	4		14
59				4	1	1	6
67					3	8	11
$n_x$	3	10	3	17	8	9	50

### Вариант № 10

Результаты измерений времени  $X$  ч непрерывной работы и количества  $Y$  шт. полностью обработанных деталей приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	6	10	14	18	22	26	
0,05	5	5					10
0,15		4	2				6
0,25			6	8			14
0,35				1	6		7
0,45				5	3	5	13
$n_x$	5	9	8	14	9	5	50

Вариант № 11

Результаты измерений температуры смазочного масла  $X$  °С в коробке передач и скорости  $Y$  км/ч автомобиля приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	30,5	40,5	50,5	60,5	70,5	80,5	
15,5	5	3					8
22,5	4	3					7
29,5		2	5	3			10
36,5				8	5		13
43,5				1	3	4	8
50,5					2	2	4
$n_x$	9	8	5	12	10	6	50

Вариант № 12

Найти зависимость между средней скоростью  $Y$  км/ч прохождения пути и температурой  $X$  °С смазочного масла в коробке передач у 100 автомобилей:

$X$	$Y$						$n_y$
	20	25	30	35	40	45	
0,05	3	3					6
0,15		5	4				9
0,25			12	30			42
0,35			5	10	14		29
0,45					5	9	14
$n_x$	3	8	21	40	19	9	100

Вариант № 13

Распределение заводов по основным фондам  $X$  млн руб. и по готовой продукции  $Y$  млн руб. приведено в корреляционной таблице:

$X$	$Y$					
	20	30	40	50	60	$n_y$
15	5	7				12
25	8	5				13
35		5	7	2		14
45			13	16	12	41
55				8	12	20
$n_x$	13	17	20	26	24	100

Вариант № 14

Распределение цехов по изменению средней заработной платы  $Y$  руб. в зависимости от изменения производительности труда  $X$  руб. по кварталам приведено в корреляционной таблице:

$X$	$Y$							
	2700	2800	2900	3000	3100	3200	3300	$n_y$
4600	1	3	2					6
5100		2	5	2				9
5600		1	3	1	2			7
6100				1	8			9
6600					2	5		7
7100						4	4	8
7600						1	3	4
$n_x$	1	6	10	4	12	10	7	50

Вариант № 15.

Найти зависимость скорости  $Y$  м/ч бурения в твердых породах от нагрузки  $X$  атм на долото:

$X$	$Y$					
	0,5	1,5	2,5	3,5	4,5	$n_y$
0,05	4	3				7
0,15		5	3			8
0,25			8	1		9
0,35			4	6		10
0,45				5	1	6
0,55				6		6
0,65				1	3	4
$n_x$	4	8	15	19	4	50

Вариант № 16

Результаты исследования зависимости длительности  $T$  ч непрерывной работы двигателей и расхода  $V$  л топлива заданы корреляционной таблицей:

$V$	$T$							$n_T$
	20	50	80	110	140	170	200	
50	10							10
100	4	12						16
150	1	4	6					11
200			10	16				26
250				4	17			21
300				1	1			2
350					3	8		11
400						1	2	3
$n_V$	15	16	16	21	21	9	2	100

Вариант № 17

Результаты измерений диаметра  $X$  мм трубы скважины и производительности  $Y$  м<sup>3</sup>/ч скважины приведены в корреляционной таблице:

$X$	$Y$						$n_Y$
	20	22	24	26	28	30	
100,5	6	6					12
105,5		3	2	1			6
110,5			4	6			10
115,5				3	5		8
120,5				3	4		7
125,5					2	5	7
$n_X$	6	9	6	13	11	5	50

Вариант № 18

Имеются данные мощности  $Y$  кВт на долоте и осевой статической нагрузки  $X$  тс. на забое, полученные при бурении пород на одном из месторождений нашей страны:

$X$	$Y$									$n_Y$
	1	3	5	7	9	11	13	15	17	
12,5	3	3								6
17,5		7	1	4						12
22,5		4	7	1						12
27,5			4	9						13
32,5			1	7	5	6	7			26
37,5					1	2		1	2	6
$n_X$	3	14	13	21	6	8	7	1	2	75

Вариант № 19

Найти зависимость между средней скоростью  $Y$  км/ч прохождения пути и температурой  $X$  °С смазочного масла в коробке передач у 100 автомобилей.

$X$	$Y$						$n_y$
	15	20	25	30	35	40	
30	3	3					6
40		5	4				9
50			8	40	2		50
60			5	10	6		21
70				4	7	3	14
$n_x$	3	8	17	54	15	3	100

Вариант № 20

Распределение 50 рабочих по выполнению сменного задания  $X$  и по повышению производительности труда  $Y$  % дается следующей таблицей:

$X$	$Y$					$n_y$
	25	28	31	34	37	
110	1	3				4
115		2	6	1		9
120		1	5	5		11
125		1	6	7	2	16
130			1	4	2	7
135				1	2	3
$n_x$	1	7	18	18	6	50

Вариант № 21

Результаты исследования зависимости между средней месячной выработкой продукции  $Y$  тыс. руб. на одного рабочего и стоимостью основных производственных средств  $X$  млн руб. приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	1	1,52	1,4	1,6	1,8	2	
15	2	1					3
16	2	2	1				5
17		2	10				12
18		1	3	2			6
19			2	2	1		5
20					2	3	5
$n_x$	4	6	16	4	3	3	36

### Вариант № 22

Результаты исследования зависимости относительной проницаемости нефти  $Y$  от насыщенности пород нефтью  $X$  приведены в корреляционной таблице:

$Y$	$X$							$n_y$
	0,05	0,15	0,25	0,35	0,45	0,55	0,65	
0,25	1							1
0,35	5	3						8
0,45	3	2	1					6
0,55		5	3	4				12
0,65				6	5			11
0,75					2	3		5
0,85						3	4	7
$n_x$	9	10	4	10	7	6	4	50

### Вариант № 23

Зависимость скорости отскока инструмента  $Y$  м/с при ударно-вращательном бурении от коэффициента пластичности долот  $X$  заданы в корреляционной таблице:

$X$	$Y$						$n_y$
	2,5	3,5	4,5	5,5	6,5	7,5	
0,5					6	3	9
0,7			2	2	5		9
0,9			9	7			16
1,1		6	5				11
1,3	3	2				1	5
$n_x$	3	8	16	9	11	3	50

### Вариант № 24

Результаты исследования зависимости выпуска валовой продукции  $Y$  % в отчетном году и выработки на одного рабочего  $X$  % приведены в корреляционной таблице:

$X$	$Y$						$n_y$
	75	85	95	105	115	125	
90	4	5					9
100	5	2	22				29
110		5	6	10			21
120			2	2	4		8
130				12	5		17
140					12	4	16
$n_x$	9	12	30	24	21	4	100



Вариант № 25

Распределение предприятий по основным фондам  $X$  млн руб. и себестоимости единицы продукции  $Y$  тыс. руб. приведено в корреляционной таблице:

$X$	$Y$					$n_y$
	1,5	2,0	2,5	3,0	3,5	
10				2	4	6
15			1	4	3	8
20		4	7	1		12
25	3	6	5			14
30	6	4		5	1	10
$n_x$	9	14	13	7	7	50

Вариант № 26

Распределение 100 автомобилей по температуре смазочного масла в двигателе  $X$  °С и по температуре масла в коробке передач  $Y$  °С даны в корреляционной таблице:

$X$	$Y$						$n_y$
	5	10	15	20	25	30	
15	2	4					6
25		6	2				8
35			3	50	2		55
45			1	10	6		17
55				4	7	3	14
$n_x$	2	10	6	64	15	3	100

Вариант № 27

Результаты измерения твердости  $Y$  (HRS) и стойкости  $X$  ч фрез приведены в корреляционной таблице:

$X$	$Y$							$n_y$
	30	34	38	42	46	50	54	
21,5	4	2						6
28,5		4	3					7
35,5			5	4				9
42,5				8	2	4		14
49,5						7	2	9
56,5						2	3	5
$n_x$	4	6	8	12	2	13	5	50

### Вариант № 28

Фонтанная скважина исследована на приток изменением диаметра штуцера с замером глубинных давлений регистрирующим манометром. Результаты зависимости скорости изменения  $Y$  т/сутки дебита скважины от скорости изменения давления  $X$  атм/сутки приведены в корреляционной таблице:

$X$	$Y$									
	150	200	250	300	350	400	450	500	550	$n_y$
5	2	1				3				3
10	2	3	1							6
15	1	2	4	3						10
20		1	5	7	6					19
25			3	4	9	6				22
30			2	2	4	5	3			16
35				4	4	2	1			11
40					1	4	3	2		10
45								1	2	3
$n_x$	4					8	15	19	4	100

### Вариант № 29

Имеют данные распределения 100 автомобилей по температуре  $Y$  °C смазочного масла в двигателе и по скорости  $X$  движения км/ч:

$X$	$Y$						$n_y$
	10	15	20	25	30	35	
20	1	5					6
30		6	4				10
40			7	40	3		50
50			2	10	8		20
60				5	6	3	14
$n_x$	1	11	13	55	17	3	100

### Вариант № 30

Результаты зависимости между дебитом  $Q$  м<sup>3</sup>/ч скважины и диаметром штуцера  $D$  мм приведены в корреляционной таблице:

$D$	$Q$							$n_Q$
	18	21	24	27	30	33	36	
100	6	3						9
105	1	5	4					10
110		7	6	2				15
115		1	5	3				9
120			4	9	2	1		16
125				4	3			7
130					2	1	1	4
$n_D$	7	16	19	18	7	2	1	70

## 4 ПАРНАЯ НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ

### 4.1. Нелинейная корреляционная зависимость

Между изучаемыми признаками  $X$  и  $Y$  может существовать нелинейная корреляционная зависимость. Если линия регрессии с уравнениями  $\hat{y}_x = f(x)$   $\hat{x}_y = \varphi(x)$  не является прямой, то зависимость между показателями называют нелинейной корреляционной зависимостью. Среди видов нелинейной корреляции обыкновенно выделяют: полиномиальную (в частности, параболическую), степенную (в частности, гиперболическую), экспоненциальную, гармоническую, фрактальную и другие. Рассмотрим некоторые из них, начав с параболической корреляции.

Пусть зависимость между признаками  $X$  и  $Y$  задана в виде корреляционной таблицы. Для определения типа нелинейной зависимости на координатной плоскости строят точки  $M_i(x_i; y_i)$ .

Параболическая корреляция. Если точки в корреляционном поле располагаются вблизи некоторой параболы, то уравнение регрессии записывают [1] в виде:

$$\hat{y}_x = a_0 + a_1x + a_2x^2 \quad . \quad (72)$$

Оценки  $a_0$ ,  $a_1$ ,  $a_2$  для неизвестных параметров истинного уравнения регрессии находят по методу наименьших квадратов. Если опытные данные не сгруппированы в корреляционную таблицу, то оценки находят, решая систему уравнений:

$$\begin{cases} na_0 + [x]a_1 + [x]^2a_2 = [y], \\ [x]a_0 + [x^2]a_1 + [x^3]a_2 = [xy], \\ [x^2]a_0 + [x^3]a_1 + [x^4]a_2 = [x^2y]. \end{cases} \quad (73)$$

Для сгруппированных значений признаков  $X$  и  $Y$  оценки  $a_0$ ,  $a_1$ ,  $a_2$  находят, решая СЛАУ:

$$\begin{cases} na_0 + [n_x x]a_1 + [n_x x^2] a_2 = [n_x \bar{y}_x], \\ [n_x x]a_0 + [n_x x^2]a_1 + [n_x x^3]a_2 = [n_x x \bar{y}_x], \\ [n_x x^2]a_0 + [n_x x^3]a_1 + [n_x x^4]a_2 = [n_x x^2 \bar{y}_x]. \end{cases} \quad (74)$$

**Гиперболическая корреляция 1.** Зависимость между  $X$  и  $Y$  может быть близкой к гиперболической. В этом случае уравнение регрессии ищут в виде

$$\hat{y}_x = a_0 + \frac{a_1}{x}. \quad (75)$$

Оценки  $a_0$  и  $a_1$  неизвестных параметров истинного уравнения регрессии находят по методу наименьших квадратов, решая систему нормальных уравнений:

$$\begin{cases} na_0 + \left[\frac{1}{x}\right] a_1 = [y], \\ \left[\frac{1}{x}\right] a_0 + \left[\frac{1}{x^2}\right] a_1 = \left[\frac{y}{x}\right], \end{cases} \quad (76)$$

где  $\left[\frac{1}{x}\right] = \sum_{i=1}^n \frac{1}{x_i}$ ,  $\left[\frac{1}{x^2}\right] = \sum_{i=1}^n \frac{1}{x_i^2}$ ,  $[y] = \sum_{i=1}^n y_i$ ,  $\left[\frac{y}{x}\right] = \sum_{i=1}^n \frac{y_i}{x_i}$ .

**Гиперболическая корреляция 2.** Если гиперболическая зависимость между признаками  $X$  и  $Y$  имеет вид:

$$\hat{y}_x = \frac{1}{a_0 + a_1 x}, \quad (77)$$

то оценки  $a_0$  и  $a_1$  находят, решая систему уравнений

$$\begin{cases} na_0 + [x]a_1 = \left[\frac{1}{y}\right], \\ [x]a_0 + [x^2]a_1 = \left[\frac{x}{y}\right]. \end{cases} \quad (78)$$

Если зависимость между признаками имеет **экспоненциальный характер**, то уравнение регрессии ищут в виде:

$$\hat{y}_x = a \cdot b^x. \quad (79)$$

Для определения оценок  $a$  и  $b$ , входящих в уравнение регрессии, решают СЛАУ:

$$\begin{cases} [x] \lg a + [x^2] \lg b = [x \lg y] \\ n \lg a + [x] \lg b = [\lg y] \end{cases}. \quad (80)$$

Если в корреляционном поле около построенных точек предполагается проведение разных по типу линий (параболы, гиперболы, экспоненты, логарифмики), то для выбора одной из них, характеризующей наилучшим образом зависимость между признаками  $X$  и  $Y$ , применяют либо метод конечных разностей, либо производят проверку необходимых условий.

## 4.2. Проверка необходимых условий

Проверку необходимых условий для выбора одной из предполагаемых нелинейных зависимостей проводят [11], пользуясь табл. 33.

Таблица 33

Необходимые условия	Регрессия	Способ выравнивания (приведения к линейной зависимости $Y=AX+B$ )
$y\left(\frac{x_1 + x_n}{2}\right) = \frac{y(x_1) + y(x_n)}{2}$	Линейная $y=ax+b$	Тождественное преобразование
$y(\sqrt{x_1 \cdot x_n}) = \sqrt{y(x_1)y(x_n)}$	Степенная $y=ax^b$	$Y=\ln y$ , $X=\ln x$ , $A=\ln a$ , $B=b$ .
$y\left(\frac{x_1 + x_n}{2}\right) = \sqrt{y(x_1)y(x_n)}$	Экспоненциальная $y=ae^{cx}(c=\ln b)$	$Y=\ln y$ , $X=x$ , $A=\ln a$ , $B=\ln b$ .
$y\left(\frac{2x_1x_n}{x_1 + x_n}\right) = \frac{y(x_1) + y(x_n)}{2}$	Гиперболическая $y=a+\frac{b}{x}$	$Y=xy$ , $X=x$
$y\left(\frac{x_1 + x_n}{2}\right) = \frac{2y(x_1)y(x_n)}{y(x_1) + y(x_n)}$	Гиперболическая $y = \frac{1}{ax + b}$	$Y=\frac{1}{y}$ , $X=x$
$y(\sqrt{x_1 \cdot x_n}) = \frac{y(x_1) + y(x_n)}{2}$	Логарифмическая $y = a \ln x + b$	$Y=y$ , $X=\ln x$

Если выполняется одно из условий первого столбца таблицы, то выбирают в качестве предполагаемой формулы соответствующую формулу, стоящую во втором столбце таблицы рассматриваемой строки. В третьем столбце указывается способ выравнивания, то есть приведения изучаемой зависимости к линейной. Если выровненные точки  $(X, Y_i)$  хорошо ложатся на прямую, то указанную во втором столбце таблицы зависимость принимаем в качестве предполагаемой.

Если значения функции, вычисленные в первом столбце таблицы при выбранных значениях аргумента, отсутствуют в таблице опытных данных, то их находят линейным интерполированием по формуле:

$$y(x) = y(x_1) + \frac{y(x_2) - y(x_1)}{x_2 - x_1} (x - x_1), \quad (81)$$

где  $x_1$  и  $x_2$  – два рядом стоящих значения признака  $X$  в таблице опытных данных, между которыми находится значение  $x$ , вычисленное по табл. 33 первого столбца.

Для всех предполагаемых формул по результатам первого столбца табл. 33 вычисляют отклонения  $\Delta$  правой части от левой необходимого условия. Вычисленные отклонения  $\Delta_i$  сравнивают и по наименьшему из них выбирают окончательно одну из формул.

### 4.3. Метод конечных разностей

Пусть в корреляционном поле могут быть проведены линии, описываемые уравнениями  $y = ax^b$ ,  $y = ae^{cx}$  ( $c = \ln b$ ),  $y = a + \frac{b}{x}$ ,  $y = \frac{1}{ax+b}$ ,  $y = a \ln x + b$ . Все эти формулы содержат по два параметра и могут быть приведены к формуле  $Y = AX + B$  посредством табл. 33. Так как все зависимости, приведенные в табл. 33, сводятся к линейной, то для обоснования выбора формулы  $Y = AX + B$  вычисляют конечные разности первого порядка  $\Delta X$ ,  $\Delta Y$  и отношения  $\Delta Y / \Delta X$ . Аналитическим критерием выбора формулы по этому методу служит тот факт, что отношения  $\Delta Y / \Delta X$  мало отличаются друг от друга для выбранной формулы.

Таблица 34

Таблица конечных разностей

x	y	$\Delta y$	$\Delta^2 y$
$x_0$	$y_0$		
		$\Delta y_0$	
$x_1$	$y_1$		$\Delta^2 y_0$
		$\Delta y_1$	
$x_2$	$y_2$		$\Delta^2 y_1$
		$\Delta y_2$	
$x_3$	$y_3$		$\Delta^2 y_2$
		$\Delta y_3$	
$x_4$	$y_4$		$\Delta^2 y_3$
		$\Delta y_4$	
$x_5$	$y_5$		
...	...	...	...

Если предполагаемая формула имеет вид  $y=ax^2+bx+c$ , то критерием выбора этой формулы являются незначительные отклонения по модулю конечных разностей второго порядка  $\Delta^2 y$  от среднего значения этих разностей  $\overline{\Delta^2 y}$ . Конечные разности находят, пользуясь табл. 34.

#### 4.4. Определение силы криволинейной связи

Для определения тесноты связи между признаками  $X$  и  $Y$  при нелинейной корреляции используют корреляционные отношения и индекс корреляции.

Корреляционным отношением  $\eta_{yx}$  называется величина, определяемая равенством:

$$\eta_{yx} = \sqrt{\frac{(\sum n_x (\bar{y}_{x_i} - \bar{y})^2)/n}{(\sum n_y (y - \bar{y})^2)/n}}, \quad (82)$$

где  $n$  – объем выборки;  $n_x$  – частота значения  $x$  признака  $X$ ;  $n_y$  – частота значения  $y$  признака  $Y$ ;  $\bar{y}$  – общая средняя признака  $Y$ ;  $\bar{y}_{x_i}$  – условная средняя признака  $Y$ .

Аналогично определяется выборочное корреляционное отношение  $\eta_{xy}$ :

$$\eta_{xy} = \sqrt{\frac{(\sum n_y (\bar{x}_{y_i} - \bar{x})^2)/n}{(\sum n_x (x - \bar{x})^2)/n}}. \quad (83)$$

Корреляционные отношения обладают следующими свойствами (сформулируем свойства для  $\eta_{yx}$ , так как для  $\eta_{xy}$  они аналогичны).

- 1) корреляционное отношение заключено между 0 и 1, т.е.  $0 \leq \eta_{yx} \leq 1$ ;
- 2) корреляционная связь между признаками  $X$  и  $Y$  отсутствует тогда и только тогда, когда  $\eta_{yx} = 0$ ;
- 3) если  $\eta_{yx} = 1$ , то между признаками  $X$  и  $Y$  существует обычная функциональная связь;
- 4) чем ближе значение  $\eta_{yx}$  к 1, тем сильнее корреляционная связь между признаками  $X$  и  $Y$ , а чем ближе  $\eta_{yx}$  к 0, тем слабее эта зависимость;
- 5) если  $\eta_{yx} = |r|$ , то регрессия  $y$  на  $x$  является линейной;

б) коэффициент линейной корреляции  $r$  не превосходит по модулю  $\eta_{yx}$ , т.е.  $|r| \leq \eta_{yx}$ .

По коэффициенту корреляции  $r$  можно судить о наличии и тесноте линейной корреляционной связи между признаками  $X$  и  $Y$ . По корреляционным отношениям можно судить только о наличии и силе корреляционной связи между признаками  $X$  и  $Y$ , но не о форме связи, которая устанавливается из геометрических соображений.

Теснота связи между признаками  $X$  и  $Y$  при любой форме корреляции может быть измерена также с помощью индекса корреляции  $i$ . Если опытные данные не сгруппированы в корреляционную таблицу, то индекс корреляции находят, согласно [1], по формуле:

$$i = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}}, \quad (84)$$

где  $S_{yx}^2 = (\sum(y_i - \hat{y}_{xi})^2)/n$  – средний квадрат отклонений фактических значений  $y$  от значений  $\hat{y}$ , вычисленных по уравнению регрессии;

$S_y^2 = (\sum(y_j - \bar{y})^2)/n$  – средний квадрат отклонений фактических значений  $y$  от их средней арифметической.

Если опытные данные сгруппированы в корреляционную таблицу, то индекс корреляции находят по формуле:

$$i = \sqrt{1 - \frac{Q_e}{Q}} \quad (85)$$

или

$$i = \sqrt{\frac{\sigma_y^2}{S_y^2}}, \quad (86)$$

где

$$Q_e = Q - Q_R, \quad Q = \sum(y_j - \bar{y})^2, \quad Q_R = (\sum(\bar{y}_{xi} - \bar{y})^2);$$

$$S_y^2 = \frac{\sum(y_j - \bar{y})^2 n_j}{n}; \quad \sigma_y^2 = (\sum(\bar{y}_{xi} - \bar{y})^2 n_i)/n.$$



Индекс корреляции по величине изменяется от 0 до 1. По индексу корреляции можно определять, как правило, тесноту связи между признаками  $X$  и  $Y$ , но не форму криволинейной связи.

#### 4.5. Проверка адекватности модели

Проверить адекватность модели – значит установить, соответствует ли построенное уравнение регрессии опытными данным и достаточно ли включенных в уравнение факторных признаков  $X_i$  для описания результативного признака. Оценка значимости уравнения регрессии производится на основе *дисперсионного анализа*. В случае нелинейной парной корреляции находят [1] статистику  $F_H$  по формуле:

$$F_H = \frac{Q_R(n-2)}{Q_e} \quad (87)$$

или

$$F_H = \frac{i^2(n-2)}{1-i^2}. \quad (87a)$$

Если оценке подвергаются не два, а  $k$  параметров уравнения регрессии, то находят статистику

$$F_H = \frac{Q_R(n-k)}{Q_e(k-1)}. \quad (88)$$

Затем при заданном уровне значимости  $\alpha$  и числах степеней свободы  $k_1=1$ ,  $k_2 = n - 2$  по таблице критических точек распределения Фишера – Снедекора находят  $F_T = F_{\alpha, k_1, k_2}$ . Если  $F_H > F_T$ , то уравнение регрессии согласуется с опытными данными, в противном случае – нет.

## 4.6. Лабораторная работа № 5

### Построение модельного уравнения нелинейной регрессии

*Цель работы:* овладение способами выбора модельного уравнения нелинейной регрессии, выработка умения и навыков расчета параметров уравнения, проверка его надежности.

#### *Задание к лабораторной работе № 5*

На основании опытных данных требуется:

- 1) построить корреляционное поле. По характеру расположения точек в корреляционном поле подобрать вид функции регрессии;
- 2) написать уравнение функции регрессии;
- 3) определить тесноту корреляционной связи между рассматриваемыми признаками;
- 4) проверить адекватность модели;
- 5) построить линию регрессии в системе координат.

*Задача.* При обработке металлов резанием устанавливается зависимость резания металла от различных характеристик резца и стружки. Зависимость скорости резания  $x$  (м/мин) и площади поперечного сечения стружки  $y$  мм<sup>2</sup> при обработке хромоникелевой стали задана табл. 35.

Таблица 35

$x$	1,1	1,4	1,7	2,1	2,6	4,7	6,1	7,0	10	12,8
$y$	25	22,7	22,1	19,8	17	12,3	10,7	10	8,2	6,7

#### 4.6.1. Пример выполнения лабораторной работы №5

В системе координат  $Oxy$  строим корреляционное поле (рис. 9).

По расположению точек в корреляционном поле видно, что около них можно провести ветвь гиперболы. Следовательно, уравнение регрессии будем искать в виде (75) или (77). Для выбора одного из этих уравнений применим необходимые условия.

Для формулы  $y = a + \frac{b}{x}$  по табл. 33 проверяем выполнение равенства

$$y \left( \frac{2x_1 x_n}{x_1 + x_n} \right) = \frac{y(x_1) + y(x_n)}{2}.$$

$$y\left(\frac{2x_1x_{10}}{x_1+x_{10}}\right) = y\left(\frac{2 \cdot 1,1 \cdot 12,8}{1,1+12,8}\right) = y(2,02).$$

Значение  $y(2,02)$  находим линейным интерполированием по формуле (81):

$$y(2,02) = y(1,7) + \frac{y(2,1) - y(1,7)}{2,1 - 1,7} (2,02 - 1,7) = 22,1 + \frac{19,8 - 22,1}{0,4} 0,32 = 20,26.$$

$$\frac{y(x_1)+y(x_n)}{2} = \frac{y(1,1)+y(12,8)}{2} = \frac{25+6,7}{2} = 15,85.$$

Вычисляем отклонение  $\Delta_1$ :  $\Delta_1 = |15,85 - 20,26| = 4,41$ .

Для формулы  $y = \frac{1}{ax+b}$  проверяем выполнение равенства

$$y\left(\frac{x_1 + x_n}{2}\right) = \frac{2y(x_1)y(x_n)}{y(x_1) + y(x_n)}$$

$$y\left(\frac{x_1+x_n}{2}\right) = y\left(\frac{x_1+x_{10}}{2}\right) = y\left(\frac{1,1+12,8}{2}\right) = y(6,95).$$

Значение  $y(6,95)$  находим линейным интерполированием:

$$y(6,95) = y(6,1) + \frac{y(7) - y(6,1)}{7 - 6,1} (6,95 - 6,1) = 10,7 + \frac{10 - 10,7}{0,9} 0,85 = 10,04.$$

$$\frac{2y(x_1)y(x_n)}{y(x_1)+y(x_n)} = \frac{2y(1,1)y(12,8)}{y(1,1)+y(12,8)} = \frac{2 \cdot 25 \cdot 6,7}{25+6,7} = 10,57.$$

Вычисляем отклонение  $\Delta_2$ :  $\Delta_2 = 10,57 - 10,04 = 0,53$ . Так как  $\Delta_2 < \Delta_1$ , то по методу необходимых условий выбираем формулу  $y = \frac{1}{ax+b}$ .

Произведем выбор одной из вышеприведённых формул по методу конечных разностей. Пусть  $y = a + \frac{b}{x}$ . Сводим эту зависимость к линейной

$Y = AX + B$ , где  $X = x$ ,  $Y = xy$  (см. табл. 35).

Вычисляем отношения  $\Delta Y / \Delta X$ . Для этого составляем расчетную табл. 36.

Таблица 36

$X = x$	1,1	1,4	1,7	2,1	2,6	4,7	6,1	7,0	10	12,8
$y$	25	22,7	22,1	19,8	17	12,3	10,7	10	8,2	6,7
$Y = xy$	27,5	31,78	37,57	41,58	44,2	57,81	65,27	70	82	85,76
$\Delta Y$	4,28	5,79	4,01	2,62	13,61	7,46	4,73	12	3,76	
$\Delta X$	0,3	0,3	0,4	0,5	2,1	1,4	0,9	3	2,8	
$\Delta Y / \Delta X$	14,27	19,3	10,03	5,24	6,48	5,33	5,26	4	1,34	

Рассмотрим зависимость  $y = \frac{1}{ax+b}$ . Пользуясь табл. 33, сводим нелинейную зависимость к линейной  $Y = AX + B$ , где  $X = x$ ,  $Y = \frac{1}{y}$ . Для нахождения отношений  $\Delta Y/\Delta X$  составляем расчетную табл. 37.

Таблица 37

$X=x$	1,1	1,4	1,7	2,1	2,6	4,7	6,1	7,0	10	12,8
$y$	25	22,7	22,1	19,8	17	12,3	10,7	10	8,2	6,7
$Y = 1/y$	0,040	0,044	0,045	0,051	0,059	0,081	0,093	0,100	0,122	0,149
$\Delta Y$	0,004	0,001	0,005	0,008	0,022	0,012	0,007	0,022	0,027	
$\Delta X$	0,3	0,3	0,4	0,5	2,1	1,4	0,9	3	2,8	
$\Delta Y/\Delta X$	0,0135	0,0040	0,0131	0,0166	0,0107	0,0087	0,0073	0,0073	0,0098	

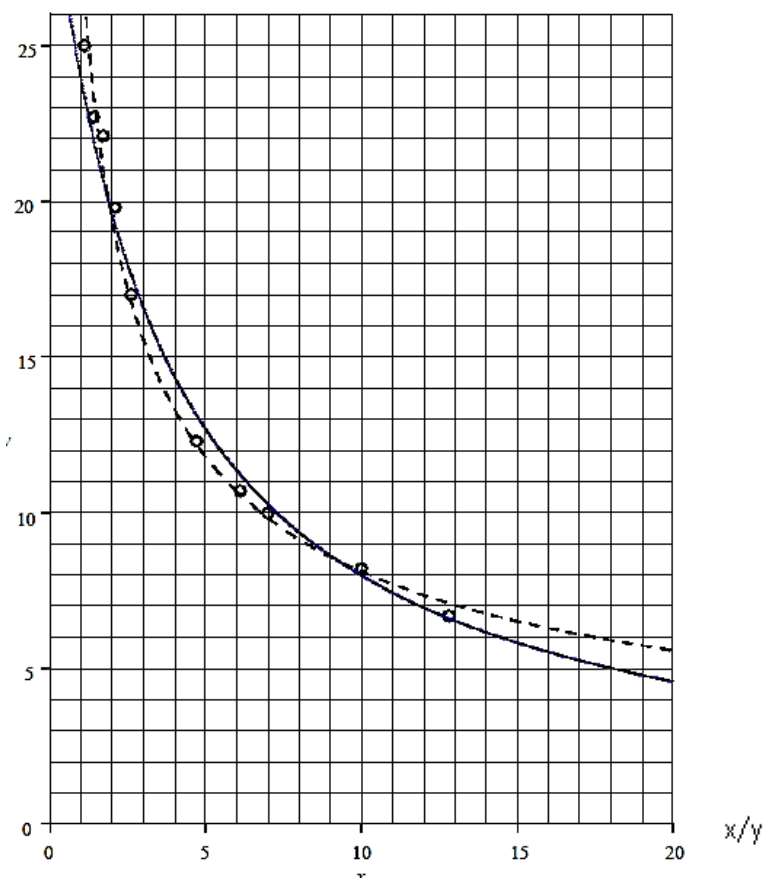


Рис. 9. Десять точек корреляционного поля позволяют предположить, что искомая функция регрессии является степенной. Сплошная линия – график функции  $y = 1/(0,00928x + 0,0325)$ , а штриховая – график функции  $y = 28,1x^{-0,54}$

Отношения  $\Delta Y/\Delta X$ , полученные для формулы  $y = \frac{1}{ax+b}$  и для  $y = a + \frac{b}{x}$  мало отличаются друг от друга. Поэтому по методу конечных разностей в качестве

лучшей выбираем формулу  $y = \frac{1}{ax+b}$ . К такому же выводу мы пришли, применяя метод необходимых условий.

Итак, зависимость скорости резания от площади поперечного сечения стружки при обработке хромоникелевой стали выражается формулой:  $y = \frac{1}{ax+b}$ . Оценки  $a$  и  $b$  неизвестных параметров истинного уравнения регрессии находим, решая систему уравнений (78):

$$\begin{cases} na_0 + [x]a_1 = \left[\frac{1}{y}\right], \\ [x]a_0 + [x^2]a_1 = \left[\frac{x}{y}\right]. \end{cases}$$

Для вычисления сумм, входящих в систему, составляем расчетную табл.

38.

Решением системы

$$\begin{cases} 10a_0 + 49,5a_1 = 0,784595, \\ 49,5a_0 + 389,37a_1 = 5,223764. \end{cases}$$

является точка  $(a_0, a_1) = (0,03251; 0,0092835)$ . Поэтому уравнение регрессии примет вид:

$$\hat{y}_x = \frac{1}{0,0092835x + 0,03251}.$$

Таблица 38

$x$	$y$	$1/y$	$x/y$	$x^2$
1,1	25	0,04	0,044	1,21
1,4	22,7	0,044053	0,061674	1,96
1,7	22,1	0,045549	0,076923	2,89
2,1	19,8	0,050505	0,106061	4,41
2,6	17	0,058824	0,152941	6,76
4,7	12,3	0,081301	0,382114	22,09
6,1	10,7	0,093458	0,570093	37,21
7	10	0,1	0,7	49
10	8,2	0,121951	1,21951	100
12,8	6,7	0,149254	1,910448	163,84
		0,784595	5,223764	389,37

Оценим силу корреляционной связи между скоростью резания и площадью поперечного сечения стружки хромоникелевой стали. Вычислим индекс корреляции по формуле (84):

$$i = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}}, \text{ где } S_{yx}^2 = (\sum (y_i - \hat{y}_{xi})^2) / (n - 1) \text{ и } S_y^2 = \frac{\sum (y_j - \bar{y})^2}{n - 1}, \text{ т.к. в нашей}$$

задаче  $n=10 < 50$ .

Для нахождения  $S_{yx}^2$  и  $S_y^2$  составляем расчетную табл. 39.

Таблица 39

x	y	$\hat{y}_x$	$(y_i - \hat{y}_{xi})^2$	$(y_j - \bar{y})^2$
1,1	25	19	36	91,2025
1,4	22,7	18	22,09	52,5625
1,7	22,1	17,2	24,01	44,2225
2,1	19,8	16,2	12,96	18,9225
2,6	17	15,1	3,61	2,4025
4,7	12,3	11,7	0,36	9,9225
6,1	10,7	10,2	0,25	22,5625
7	10	9,4	0,36	29,7025
10	8,2	7,5	0,49	52,5625
12,8	6,7	6,3	0,16	76,5625
			100,29	400,0625

Тогда  $i = \sqrt{1 - \frac{11,14333}{44,45139}} = 0,86$ . Связь между скоростью резания и

площадью поперечного сечения стружки хромоникелевой стали сильная.

Проверяем адекватность полученного уравнения регрессии по критерию Фишера – Снедекора. Находим статистику

$$F_H = \frac{i^2(n-2)}{1-i^2} = \frac{0,86^2(10-2)}{1-0,86^2} = 22,7.$$

При уровне значимости  $\alpha = 0,05$  и числах степеней свободы  $k_1 = 1$ ,  $k_2 = n - 2 = 10 - 2 = 8$  по таблице критических точек распределения Фишера – Снедекора находим  $F_m = F_{\alpha, k_1, k_2} = F_{0,05; 1; 8} = 5,32$ .

Так как  $F_H = 22,7 > 5,32$ , то модель адекватна. Следовательно, зависимость скорости резания от площади поперечного сечения стружки при обработке хромоникелевой стали по данным выборки описывается уравнением:

$$\hat{y}_x = \frac{1}{0,0092835x + 0,03251}.$$

Во многих инженерных задачах функцию регрессии можно предсказать, исходя из особенностей исследуемого реального процесса. Так, при резании металла для инженера очевидно, что искомая функция имеет степенной характер, т.е.  $y = kx^{-D}$ ,  $D > 0$ . На рис. 9 для сравнения показаны гипербола (сплошная линия) и график функции  $y = 28,1x^{-0,54}$ .

#### 4.6.2. Варианты заданий лабораторной работы № 5

##### Вариант № 1

При исследовании зависимости между средней заработной платой  $Y$  тыс. руб. на одного работника и выпуском  $X$  тыс. руб. продукции на одного работника по заводу получены следующие данные:

$X$	21,07	23,07	28,69	22,42	21,41	18,49	21,64	39,19	51,96	42,36	51,80	50,45
$Y$	50,6	40,0	29,6	32,5	34,9	47,6	42,0	18,7	12,5	20,1	15,1	15,2

##### Вариант № 2

Данные о производстве  $X$  тыс. руб. дизтоплива и себестоимости  $Y$  тыс. руб. единицы продукции приведены в таблице:

$X$	5	6	8	13	34	72	95	113	127	90
$Y$	143	125	87	45	33	27	16	25	24	27

##### Вариант № 3

Зависимость скорости  $v$  м/мин резания от площади  $S$  мм<sup>2</sup> поперечного сечения стружки при обработке стали задана таблицей:

$v$	1	1,2	1,7	2,4	3,6	4,7	6	8	11	14
$S$	26	22	22,6	19,1	15,2	12,8	11,3	9,5	7,8	6,5

##### Вариант № 4

Данные зависимости мощности на долоте  $S$  кВт от осевой статической нагрузки на забой  $P_c$  тс. при бурении пород приведены в таблице:

$P_c$	1	3	5	7	9	11	13	15	17
$S$	12,5	17,8	37	41,9	45	47	39	32	28

### Вариант № 5

Зависимость скорости отскока инструмента  $v_0$  м/с при ударно-вращательном бурении от коэффициента  $K$  пластичности долот задана таблицей:

$K$	1,5	2,5	3,5	4,5	5,5	6,5
$v_0$	1,2	0,6	0,21	0,9	1,4	1,9

### Вариант № 6

Данные о количестве  $X$  тыс. шт. выпускаемых деталей и полных затратах  $Y$  сотни руб. на их изготовление на однотипных предприятиях приведены в таблице:

$X$	1	2	4	9	13	18	20
$Y$	26	22	19	12	9	8	6

### Вариант № 7

При исследовании зависимости времени  $Y$  мин на обработку одной детали от стажа  $X$  в годах работы на одном из моторостроительных объединений в цехе резиново-технических и пластмассовых изделий на слесарном участке получены следующие данные:

$X$	1	2	3	4	5	6	7
$Y$	5	3,33	2,9	2,2	2,1	2	2

### Вариант № 8

Зависимость удельного момента  $M_y$  кгс·м/тс. на долоте от осевой статической нагрузки  $P_c$  тс. на забой при бурении пород задана таблицей:



$P_c$	1	3	5	7	9	11	13	15
$M_y$	22,5	11,5	6	5,5	2,6	2,4	2,1	2

### Вариант № 9

Результаты измерений зависимости фазовой проницаемости  $K_B$  воды от нефтенасыщенности  $S_H$  породы приведены в таблице:

$K_B$	0,25	0,35	0,45	0,55	0,65	0,75	0,85
$S_H$	0,65	0,45	0,25	0,15	0,10	0,05	0,07

### Вариант № 10

В результате исследований установлено, что между овальностью  $X$  колец после их обработки и овальностью  $Y$  термической обработки, существует связь, которая задана таблицей:

$X$	5	10	15	20	25
$Y$	12,1	29,3	36	42	45

### Вариант № 11

При исследовании зависимости между выпуском  $Y$  тыс.руб. готовой продукции и коэффициентом  $X$  % использования техники получены следующие данные:

$X$	73	75	79	82	83	86	80	85	95	93	97	77
$Y$	14	21	29	30	31	35	34	41	38	39	46	27

### Вариант № 12

Давление  $P$  кг воздуха на парашют возрастает при увеличении скорости  $v$  м/с падения следующим образом:

$P$	2,23	3,28	4,65	6,5	8,1
$v$	0,3	0,6	1,2	2,4	4,2

Вариант № 13

Прочность  $Y$  кг/см<sup>2</sup> бетона при испытании цилиндрических образцов в зависимости от отношения  $X = h / \alpha$  высоты  $h$  к диаметру  $\alpha$  оказалась равной:

$X$	0,5	1,0	2,0	3,0	4,0	5,0	6,0
$Y$	290	250	216	206	200	195	190

Вариант № 14

Зависимость между размером предприятия по стоимости  $X$  млн руб. основных средств и себестоимостью  $Y$  руб. единицы продукции характеризуется следующими данными:

$X$	0,5	1,5	2,5	3,5	4,5	5,5	7,5
$Y$	15	11	12	10,8	10	9	8

Вариант № 15

Зависимость между ростом  $X$  тыс. руб. производительности труда на одного работающего и выпуском  $Y$  тыс. руб. товарной продукции ремонтного цеха машиностроительного завода характеризуется следующими данными:

$X$	1,5	2,9	3,0	3,1	3,2	3,4	3,5	3,6	4,2
$Y$	580	618	658	670	662	699	717	775	786

Вариант № 16

Зависимость себестоимости  $Y$  тыс. руб. продукции от затрат  $X$  тыс. руб. на единицу продукции по объединению «Сибкомплектмонтаж» характеризуется следующими данными:

$X$	0,1	0,4	1	4	6	10	20	26
$Y$	2248	1950	1500	1020	906	290	175	121

### Вариант № 17

Компрессорную скважину исследовали на приток  $Q$  т/сут. нефти при различных режимах работы с величиной  $\Delta P$  атм забойных давлений глубинным манометром. Результаты исследований приведены в таблице:

$Q$	5	15	25	35	45	55
$\Delta P$	1,25	1,3	5,25	11,25	17,25	21,25

### Вариант № 18

Зависимость между стоимостью  $X$  млн руб. основных средств предприятия и выработкой  $Y$  тыс. руб. продукции на одного работника характеризуется следующими данными:

$X$	1	1,5	2,5	3,5	4,5	5,5	6,5
$Y$	4	6	6,8	7,9	8,7	9	9,5

### Вариант № 19

Ниже приводятся данные удельного момента  $Y$  кг·м/тс на долото и осевой статической нагрузки  $X$  тс. на забой при бурении пород на одном из месторождений:

$X$	1	3	5	7	9	11	13	15	17
$Y$	25	15	12	8	10	5	4,5	3	2,8

### Вариант № 20

Зависимость между мощностью  $X$  млн ед. продукции в год предприятия и фактическими капитальными вложениями  $Y$  млн руб. характеризуется следующими данными:

$X$	1	2	3	4	5	6
$Y$	1,2	2,6	3,8	4,6	4,9	5,4

### Вариант № 21

Результаты изучения зависимости между среднемесячной производительностью  $X$  руб. труда рабочего и себестоимостью  $Y$  руб. одной тонны продукции приведены в следующей таблице:

$X$	21	24	30	34	35	36	39	40
$Y$	20	13	12	13	11	10	11	10

### Вариант № 22

Энерговооруженность  $X$  тыс. кВт-ч труда на одного рабочего и производительность  $Y$  тыс. штук изделий труда одного рабочего на ряде предприятий характеризуется следующими данными:

$X$	3	3,05	3,6	4,25	4,45	4,55
$Y$	1	1,5	1,8	2,5	3	4

### Вариант № 23

Зависимость между стоимостью  $X$  млн руб. основных средств предприятий и месячным выпуском  $Y$  тыс. руб. продукции характеризуется следующими данными:

$X$	1	4	7	9	11	12	15
$Y$	10	12	26	30	40	42	58

### Вариант № 24

Зависимость между капитальными вложениями  $Y$  млн руб. и мощностью  $X$  млн тонн продукции предприятий данного типа задана таблицей:

$X$	1	2	3	4	5	6	7	8
$Y$	0,9	2,59	3,67	4,45	4,95	5,20	4,7	4,5

### Вариант № 25

Распределение однотипных предприятий по объему  $X$  произведенной за день продукции и себестоимости  $Y$  единицы продукции в условных единицах приведено в таблице:

$X$	50	100	150	200	250	300
$Y$	140	120	118	110	115	100

### Вариант № 26

Результаты исследования зависимости объема  $V$  см<sup>3</sup> зоны разрушения от предела  $P_0$  кг/мм<sup>2</sup> текучести известняков приведены в таблице:

$P_0$	12,5	37,5	62,5	87,5	112,5	137,5	187,5
$V$	0,19	0,13	0,11	0,10	0,08	0,07	0,06

### Вариант № 27

Зависимость перепада  $\Delta P$  кг/см<sup>2</sup> давления (разность между гидростатическим и пластовым давлением) от времени  $t$  с при бурении в песчанике задана таблицей:

$t$	0,025	0,074	0,125	0,175	0,225	0,275	0,325
$\Delta P$	95	73	52	45	35	33	31

### Вариант № 28

Зависимость среднемесячной заработной платы  $Y$  тыс. руб. рабочих химического завода от их квалификации  $X$  (разряд) характеризуется следующими данными:

$X$	1	2	3	4	5	6	7
$Y$	0,8	1,2	1,8	2,9	4,2	5,9	12,5

Вариант № 29

Зависимость между размером  $X$  млн руб. предприятия по стоимости основных средств и себестоимостью  $Y$  тыс. руб. единицы продукции характеризуется следующими данными:

$X$	0,5	1,5	2,5	3,5	4,5	6,0	9	10
$Y$	14	11	10	8	6,5	5	4,5	4

Вариант № 30

Зависимость между фазовой проницаемостью  $K_n$  нефти и насыщенностью  $S_n$  породы нефтью характеризуется следующими данными:

$S_n$	0,35	0,45	0,55	0,65	0,75	0,85
$K_n$	0,05	0,1	0,15	0,45	0,55	0,75

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Губин В.И., Осташков В.Н. Статистические методы обработки экспериментальных данных: учебное пособие для студентов технических вузов. Тюмень: Изд-во «ТюмГНГУ», 2007. 202 с.
2. Математика: Энциклопедия / под ред. Ю.В. Прохорова. М.: Большая Российская энциклопедия, 2003. 847 с.
3. Климова Е.З., Миллионщикова А.П., Фастовец Н.О. Задачи по математической статистике. М.: МИНХ И ГП им. И.М. Губкина, 1976. 139 с.
4. Венецкий, И.Г. Кильдышев Г.С. Основы теории вероятностей и математической статистики. М.: Статистика, 1968. 320 с.
5. Гмурман В.Е. Теория вероятностей и математическая статистика: учебник для СПО пособие для вузов / В.Е. Гмурман. 12-е изд. М.: Издательство Юрайт, 2018. 479 с.
6. Калинина В.Н., Панкин В.Ф. Математическая статистика: учебник для студ. сред. спец. учеб. заведений. М.: Высшая школа, 2001. 429 с.
7. Протасов К.В. Статистический анализ экспериментальных данных. М.: Мир, 2005. 142 с.
8. Карасев А.И., Аксютин З.М., Савельева Т.И. Курс высшей математики для экономических вузов: учеб. пособие для студентов вузов. Ч. 2. М.: Высшая школа, 1982. 528 с.
9. Степнов М.Н. Статистические методы обработки результатов механических испытаний: справочник. М.: Машиностроение, 1985. 411 с.
10. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. М.: «Инфра-М», 1997. 399 с.
11. Янпольский А.Р. Лабораторный вычислительный практикум. М.: Изд-во ВИА, 1968. 478 с.

12. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // УФН. 1996. Т. 166. № 11.
13. Добеши И. Десять лекций по вейвлетам. Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001 (пер. с англ. Daubechies I., Ten Lectures on Wavelets, 1992).
14. Ферстер Э., Рёнц Б. Методы корреляционного и регрессивного анализа. М.: «Статистика», 1987. 481 с.



## ОБОЗНАЧЕНИЯ

$a$  – математическое ожидание;

$c$  – статистика критерия Ястремского;

$\det C$  – детерминант, определитель матрицы;

$h$  – шаг, длина частичного интервала;

$i$  – индекс корреляции;

$k$  – число степеней свободы (параметр распределения Пирсона);

$l$  – число вариантов дискретного вариационного ряда;

$n_i$  – частота;

$n'_i$  – округленные частоты;

$m_3, m_4$  – условные центральные моменты 3-го, 4-го порядка;

$n$  – объем выборки;

$n_{ij}$  – частота совместного появления пары  $(x_i, y_j)$ ;

$n_{xi}$  – частота появления признака  $X$ ;

$p$  – число факторных признаков;

$q$  – табличное значение;

$r$  – сумма числа параметров теоретического закона распределения;

и коэффициент линейной корреляции;

$\hat{r}$  – оценка для коэффициента  $r$  линейной корреляции;

$r_{xy}$  – коэффициент парной корреляции;

$r_{xy(z)}$  – коэффициент частной корреляции;

$s$  – число интервалов вариационного ряда;

$t_p$  – расчетная статистика;

$t_{кр}$  – табличная статистика распределения Стьюдента;

$t_\gamma$  – табличное значение и табличная статистика функции Лапласа;

$u_i$  – условная варианта;

$\bar{x}$  – выборочная средняя;

$x_i$  – значение признака, варианта;

$x_{max}$  – наибольшая варианта;

$x_{min}$  – наименьшая варианта;

$\overline{xy}$  – средняя произведения значений признаков  $X, Y$ ;  
 $y_i$  – выровненные частоты;  
 $\bar{y}_{xi}$  – условное среднее признака  $Y$ ;  
 $A_s$  – асимметрия;  
 $C$  – ложный нуль;  
 $D_{\Gamma}$  – генеральная дисперсия;  
 $D_u$  – остаточная дисперсия;  
 $F_{\varepsilon}(x)$  – функция распределения выборочная;  
 $F_H$  – вычисленная статистика критерия Фишера – Снедекора;  
 $F_T$  – табличная статистика критерия Фишера – Снедекора;  
 $E_x$  – эксцесс;  
 $F(x)$  – интегральная функция распределения;  
 $H_0$  – нулевая гипотеза;  
 $J$  – статистика критерия Ястремского;  
 $K(l)$  – функция вероятности;  
 $K_{\varepsilon}$  – коэффициент эластичности;  
 $M_e X$  – медиана;  
 $M_o X$  – мода;  
 $M_i^*$  – условный начальный  $i$ -й момент;  
 $M(X)$  – генеральная средняя;  
 $Q$  – сумма квадратов отклонений;  
 $Q_e$  – остаточная сумма квадратов;  
 $Q_R$  – сумма квадратов отклонений условных средних;  
 $R$  – размах варьирования признака;  
 $R$  – множество всех действительных чисел;  
 $R^2$  – коэффициент детерминации;  
 $S, \hat{S}$  – выборочные средние квадратические отклонения;  
 $S_{A_s}$  – среднее квадратическое асимметрии;  
 $S_{E_x}$  – среднее квадратическое эксцесса;  
 $S_R$  – среднеквадратическая ошибка  $R$ ;

$S_u$  – среднее квадратическое отклонение условных вариант  $u_i$ ;  
 $S_x$  – выборочное среднее квадратическое отклонение признака  $X$ ;  
 $\widehat{S^2}$  – исправленная выборочная дисперсия;  
 $S^2$  – выборочная дисперсия;  
 $S_{yx}^2$  – средний квадрат отклонений фактических значений  $y$  от вычисленных;  
 $S_y^2$  – средний квадрат отклонений опытных значений  $y$  от их средней  $y$ ;  
 $V$  – коэффициент вариации;  
 $W_i$  – накопленная частота;  
 $\alpha$  – уровень значимости;  
 $\gamma$  – уровень доверия, надежность;  
 $\delta$  – относительная погрешность;  
 $\varepsilon$  – средняя ошибка аппроксимации;  
 $\eta_{xy}$  – корреляционное отношение;  
 $\lambda$  – параметр распределения Пуассона и статистика критерия Колмогорова;  
 $\lambda_\alpha$  – критическое значение статистики критерия Колмогорова;  
 $\sigma$  – генеральная средняя квадратическая;  
 $\sigma^2$  – дисперсия;  
 $\sigma_r$  – средняя квадратическая ошибка;  
 $\sigma_y$  – стандартная ошибка уравнения регрессии;  
 $\varphi(u_i)$  – табличное значение функции;  
 $\chi^2$  – статистика критерия Пирсона;  
 $\chi_{кр}^2$  – табличное значение;  
 $\chi_0^2$  – вычисленное значение;  
 $w$  – частота;  
 $w_i$  – относительная частота;  
 $\Delta_y^2$  – конечные разности второго порядка;  
 $\Phi(x)$  – функция Лапласа.

# ПРИЛОЖЕНИЯ

## Приложение 1

Таблица значений функции  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	39986	3984	3982	3980	3977	3973
0,1	0,3970	3965	3961	3956	3951	3945	3939	3032	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3244
0,7	0,3123	3101	3079	3056	3034	3011	2989	1966	2943	2920
0,8	0,2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0,0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0,0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0,0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0,0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0,0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0,0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0,0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0,0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0,0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0,0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0,0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0,0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0,0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0,0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0,0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0,0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0,0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0,0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0,0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

$$\text{Значение функции } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	00000	0,52	0,1985	1,04	0,3508	1,56	0,4406	2,16	0,4846
0,01	0,0040	0,53	0,2019	1,05	0,3531	1,57	0,4418	2,18	04854
0,02	0,0080	0,54	0,2054	1,06	0,3554	1,58	0,4429	2,20	0,4861
0,03	0,0120	0,55	0,2088	1,07	0,3577	1,59	0,4441	2,22	0,4868
0,04	0,0160	0,56	0,2123	1,08	0,3599	1,60	0,4452	2,24	0,4875
0,05	0,0199	0,57	0,2157	1,09	0,3621	1,61	0,4463	2,26	0,4881
0,06	0,0239	0,58	0,2190	1,10	0,3643	1,62	0,4474	2,28	0,4887
0,07	0,0279	0,59	0,2224	1,11	0,3665	1,63	0,4484	2,30	0,4893
0,08	0,0319	0,60	0,2257	1,12	0,3686	1,64	0,4495	2,32	0,4898
0,09	0,0359	0,61	0,2291	1,13	0,3708	1,65	0,4505	2,34	0,4904
0,10	0,0398	0,62	0,2324	1,14	0,3729	1,66	0,4515	2,36	0,4908
0,11	0,0438	0,63	0,2357	1,15	0,3749	1,67	0,4525	2,38	0,4913
0,12	0,0478	0,64	0,2389	1,16	0,3770	1,68	0,4535	2,40	0,4918
0,13	0,0517	0,65	0,2422	1,17	0,3790	1,69	0,4545	2,42	0,4922
0,14	0,0557	0,66	0,2454	1,18	0,3810	1,70	0,4554	2,44	0,4927
0,15	0,0596	0,67	0,2486	1,19	0,3830	1,71	0,4564	2,46	0,4931
0,16	0,0636	0,68	0,2517	1,20	0,3849	1,72	0,4573	2,48	0,4934
0,17	0,0675	0,69	0,2549	1,21	0,3869	1,73	0,4582	2,50	0,4938
0,18	0,0714	0,70	0,2580	1,22	0,3888	1,74	0,4591	2,52	0,4941
0,19	0,0753	0,71	0,2611	1,23	0,3907	1,75	0,4599	2,54	0,4945
0,20	0,0793	0,72	0,2642	1,24	0,3925	1,76	0,4608	2,56	0,4948
0,21	0,0832	0,73	0,2673	1,25	0,3944	1,77	0,4616	2,58	0,4951
0,22	0,0871	0,74	0,2703	1,26	0,3962	1,78	0,4625	2,60	0,4953
0,23	0,0910	0,75	0,2734	1,27	0,3980	1,79	0,4633	2,62	0,4956
0,24	0,0948	0,76	0,2764	1,28	0,3997	1,80	0,4641	2,64	0,4959
0,25	0,0987	0,77	0,2794	1,29	0,4015	1,81	0,4649	2,66	0,4961
0,26	0,1026	0,78	0,2823	1,30	0,4032	1,82	0,4656	2,68	0,4963
0,27	0,1064	0,79	0,2852	1,31	0,4049	1,83	0,4664	2,70	0,4965
0,28	0,1103	0,80	0,2881	1,32	0,4066	1,84	0,4671	2,72	0,4967
0,29	0,1141	0,81	0,2910	1,33	0,4082	1,85	0,4678	2,74	0,4969
0,30	0,1179	0,82	0,2939	1,34	0,4099	1,86	0,4686	2,76	0,4971
0,31	0,1217	0,83	0,2967	1,35	0,4115	1,87	0,4693	2,78	0,4973
0,32	0,1255	0,84	0,2995	1,36	0,4131	1,88	0,4699	2,80	0,4974
0,33	0,1293	0,85	0,3023	1,37	0,4147	1,89	0,4706	2,82	0,4976
0,34	0,1331	0,86	0,3051	1,38	0,4162	1,90	0,4713	2,84	0,4977
0,35	0,1368	0,87	0,3078	1,39	0,4177	1,91	0,4719	2,86	0,4979
0,36	0,1406	0,88	0,3106	1,40	0,4192	1,92	0,4726	2,88	0,4980
0,37	0,1443	0,89	0,3133	1,41	0,4207	1,93	0,4732	2,90	0,4981
0,38	0,1480	0,90	0,3159	1,42	0,4222	1,94	0,4738	2,92	0,4982
0,39	0,1517	0,91	0,3186	1,43	0,4236	1,95	0,4744	2,94	0,4984
0,40	0,1554	0,92	0,3112	1,44	0,4251	1,96	0,4750	2,96	0,4985
0,41	0,1591	0,93	0,3238	1,45	0,4265	1,97	0,4756	2,98	0,4986
0,42	0,1628	0,94	0,3264	1,46	0,4279	1,98	0,4761	3,00	0,49865
0,43	0,1664	0,95	0,3289	1,47	0,4292	1,99	0,4767	3,20	0,49931
0,44	0,1700	0,96	0,3315	1,48	0,4306	2,00	0,4772	3,40	0,49966
0,45	0,1736	0,97	0,3340	1,49	0,4319	2,02	0,4783	3,60	0,499841
0,46	0,1772	0,98	0,3365	1,50	0,4332	2,04	0,4793	3,80	0,499928
0,47	0,1808	0,99	0,3389	1,51	0,4345	2,06	0,4803	4,00	0,499968
0,48	0,1844	1,00	0,3413	1,52	0,4357	2,08	0,4812	4,50	0,499997
0,49	0,1879	1,01	0,3438	1,53	0,4370	2,10	0,4821	5,00	0,4999997
0,50	0,1915	1,02	0,3461	1,54	0,4382	2,12	0,4830		
0,51	0,1950	1,03	0,3485	1,55	0,4394	2,14	0,4838		

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

$n$	$\gamma$			$n$	$\gamma$		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,361	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	$\infty$	1,960	2,576	3,291
19	2,10	2,88	3,92				

Таблица значений  $q_\gamma = q(\gamma, n)$ 

$n$	$\gamma$			$n$	$\gamma$		
	0,95	0,99	0,999		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

Критические точки распределения  $\chi^2$ 

Число степеней свободы $k$	Уровень значимости $\alpha$					
	0,01	0,025	0,05	0,95	0,975	0,98
1	6,6	5,024	3,841	0,0039	0,00098	0,00016
2	9,2	7,378	5,991	0,103	0,051	0,020
3	11,3	9,348	7,815	0,352	0,216	0,115
4	13,3	11,143	9,488	0,711	0,484	0,297
5	15,1	12,832	11,070	1,15	0,831	0,554
6	16,8	14,449	12,592	1,64	1,24	0,872
7	18,5	16,013	14,067	2,17	1,69	1,24
8	20,1	17,535	15,507	2,73	2,18	1,65
9	21,7	19,023	16,919	3,33	2,70	2,09
10	23,2	20,483	18,307	3,94	3,25	2,56
11	24,7	21,920	19,676	4,57	3,82	3,05
12	26,2	23,336	21,026	5,23	4,40	3,57
13	27,7	24,736	22,362	5,89	5,01	4,11
14	29,1	26,129	23,685	6,57	5,63	4,66
15	30,6	27,488	24,996	7,26	6,26	5,23
16	32,0	28,845	26,296	7,96	6,91	5,81
17	33,4	30,191	27,587	8,67	7,56	6,41
18	34,8	31,536	28,869	9,39	8,23	7,01
19	36,2	32,852	30,144	10,1	8,91	7,63
20	37,6	34,170	31,410	10,9	9,59	8,26
21	38,9	35,479	32,671	11,6	10,3	8,90
22	40,3	36,781	33,924	12,3	11,0	9,54
23	41,6	38,076	35,172	13,1	11,7	10,2
24	43,0	39,364	36,415	13,8	12,4	10,9
25	44,3	40,646	37,652	14,6	13,1	11,5
26	45,6	41,923	38,885	15,4	13,8	12,2
27	47,0	43,194	40,113	16,2	14,6	12,9
28	48,3	44,461	41,337	16,9	15,3	13,6
29	49,6	45,722	42,557	17,7	16,0	14,3
30	50,9	46,979	43,773	18,5	16,8	15,0

## Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости $\alpha$ (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,3138	12,7062	6,97	63,7	318,3	637,0
2	2,9200	4,3037	4,54	9,92	22,33	31,6
3	2,3534	3,1824	3,75	5,84	10,22	12,9
4	2,1318	2,7764	3,37	4,60	7,17	8,61
5	2,0150	2,5706	3,14	4,03	5,89	6,86
6	1,9432	2,4469	3,00	3,71	5,21	5,96
7	1,8946	2,3646	2,90	3,50	4,79	5,40
8	1,8595	2,3060	2,82	3,36	4,50	5,04
9	1,8331	2,2622	2,76	3,25	4,30	4,78
10	1,8125	2,2281	2,72	3,17	4,14	4,59
11	1,7959	2,2010	31,82	3,11	4,03	4,44
12	1,7823	2,1788	2,68	3,05	3,93	4,32
13	1,7709	2,1604	2,65	3,01	3,85	4,22
14	1,7613	2,1448	2,62	2,98	3,79	4,14
15	1,7530	2,1314	2,60	2,95	3,73	4,07
16	1,7459	2,1190	2,58	2,92	3,69	4,01
17	1,7396	2,1098	2,57	2,90	3,65	3,96
18	1,7341	2,1009	2,55	2,88	3,61	3,92
19	1,7291	2,0930	2,54	2,86	3,58	3,88
20	1,7247	2,0860	2,53	2,85	3,55	3,85
21	1,7207	2,0796	2,52	2,83	3,53	3,82
22	1,7171	2,0739	2,51	2,82	3,51	3,79
23	1,7139	2,0687	2,50	2,81	3,49	3,77
24	1,7109	2,0639	2,49	2,80	3,47	3,74
25	1,7081	2,0595	2,49	2,79	3,45	3,72
26	1,7056	2,0555	2,48	2,78	3,44	3,71
27	1,7033	2,0518	2,47	2,77	3,42	3,69
28	1,7011	2,0484	2,46	2,76	3,40	3,66
29	1,6991	2,0452	2,46	2,76	3,40	3,66
30	1,6973	2,0423	2,46	2,75	3,39	3,65
40	1,6839	2,0211	2,42	2,70	3,31	3,55
60	1,6706	2,0003	2,39	2,66	3,23	3,46
120	1,6577	1,9840	2,36	2,62	3,17	3,37
$\infty$	1,6479	1,9647	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости $\alpha$ (односторонняя критическая область)					



**Критические точки распределения Фишера – Снедекора** $(k_1$  – число степеней свободы большей дисперсии, $k_2$  – число степеней свободы меньшей дисперсии)

$k_1$	$k_2$											
	1	2	3	4	5	6	7	8	9	10	11	12
Уровень значимости $\alpha = 0,01$												
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,0	90,1	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	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
$k_1$	$k_2$											
	1	2	3	4	5	6	7	8	9	10	11	12
Уровень значимости $\alpha = 0,05$												
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

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННЫХ РЯДОВ	
1.1. Первичная обработка результатов наблюдений.....	4
1.2. Расчет выборочных характеристик статистического распределения.....	9
1.3. Доверительные интервальные оценки параметров распределения .....	13
1.4 Лабораторная работа № 1. Первичная обработка экспериментальных данных.....	14
1.4.1. Пример выполнения лабораторной работы №1.....	16
1.4.2. Варианты заданий лабораторной работы №1.....	22
2. ПРОВЕРКА СООТВЕТСТВИЯ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ	
2.1. Построение кривой распределения по опытным данным.....	32
2.2. Статистические оценки параметров распределения.....	33
2.3. Проверка статистических гипотез.....	34
2.3.1. Критерий согласия Пирсона $\chi^2$ .....	37
2.3.2. Критерий Романовского.....	39
2.3.3. Критерий Колмогорова.....	40
2.3.4. Критерий Ястремского.....	42
2.3.5. Приближенные критерии нормальности распределения.....	42
2.4. Лабораторная работа № 2. Построение кривой нормального распределения по опытным данным. Проверка гипотезы о нормальном распределении выборки.....	44
2.4.1 Пример выполнения лабораторной работы №2.....	44
3. ПАРНАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ	
3.1. Понятие корреляционной зависимости. Задачи теории корреляции.....	48
3.2. Парная линейная корреляционная зависимость.....	50
3.3. Коэффициент корреляции, его свойства и значимость.....	53
3.4. Определение надежности (доверительного интервала) коэффициента корреляции.....	55

3.5. Коэффициент детерминации.....	56
3.6. Проверка адекватности модели.....	57
3.7 Оценка величины погрешности.....	58
3.8 Лабораторная работа № 3. Построение модели линейной корреляции по несгруппированным данным.....	59
3.8.1. Пример выполнения лабораторной работы №3.....	59
3.8.2. Варианты заданий лабораторной работы №3.....	66
3.9. Лабораторная работа № 4. Построение выборочного уравнения линии регрессии по сгруппированным данным .....	73
3.9.1. Пример выполнения лабораторной работы №4.....	74
3.9.2. Варианты заданий лабораторной работы №4.....	80
<b>4. ПАРНАЯ НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ</b>	
4.1. Нелинейная корреляционная зависимость.....	90
4.2. Проверка необходимых условий.....	92
4.3. Метод конечных разностей.....	93
4.4. Определение силы криволинейной связи.....	94
4.5. Проверка адекватности модели.....	96
4.6. Лабораторная работа № 5. Построение модельного уравнения нелинейной корреляции.....	97
4.6.1. Пример выполнения лабораторной работы №5.....	97
4.6.2. Варианты заданий лабораторной работы №5.....	102
<b>БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....</b>	<b>110</b>
<b>ОБОЗНАЧЕНИЯ.....</b>	<b>112</b>
<b>ПРИЛОЖЕНИЯ.....</b>	<b>115</b>

Учебное издание

**Кокурина** Галина Николаевна

Математические методы  
обработки экспериментальных данных

Учебное пособие

Редактор О.А. Соловьева

Подписано в печать 29.10 2018 Формат 60x84 1/16. Бумага писчая.

Усл.печ.л. 7,21. Тираж 50 экз. Заказ

ФГБОУ ВО «Ивановский государственный  
химико-технологический университет»

Отпечатано на полиграфическом оборудовании  
редакционно-издательского центра ФГБОУ ВО «ИГХТУ»  
153000, г.Иваново, Шереметевский пр., 7