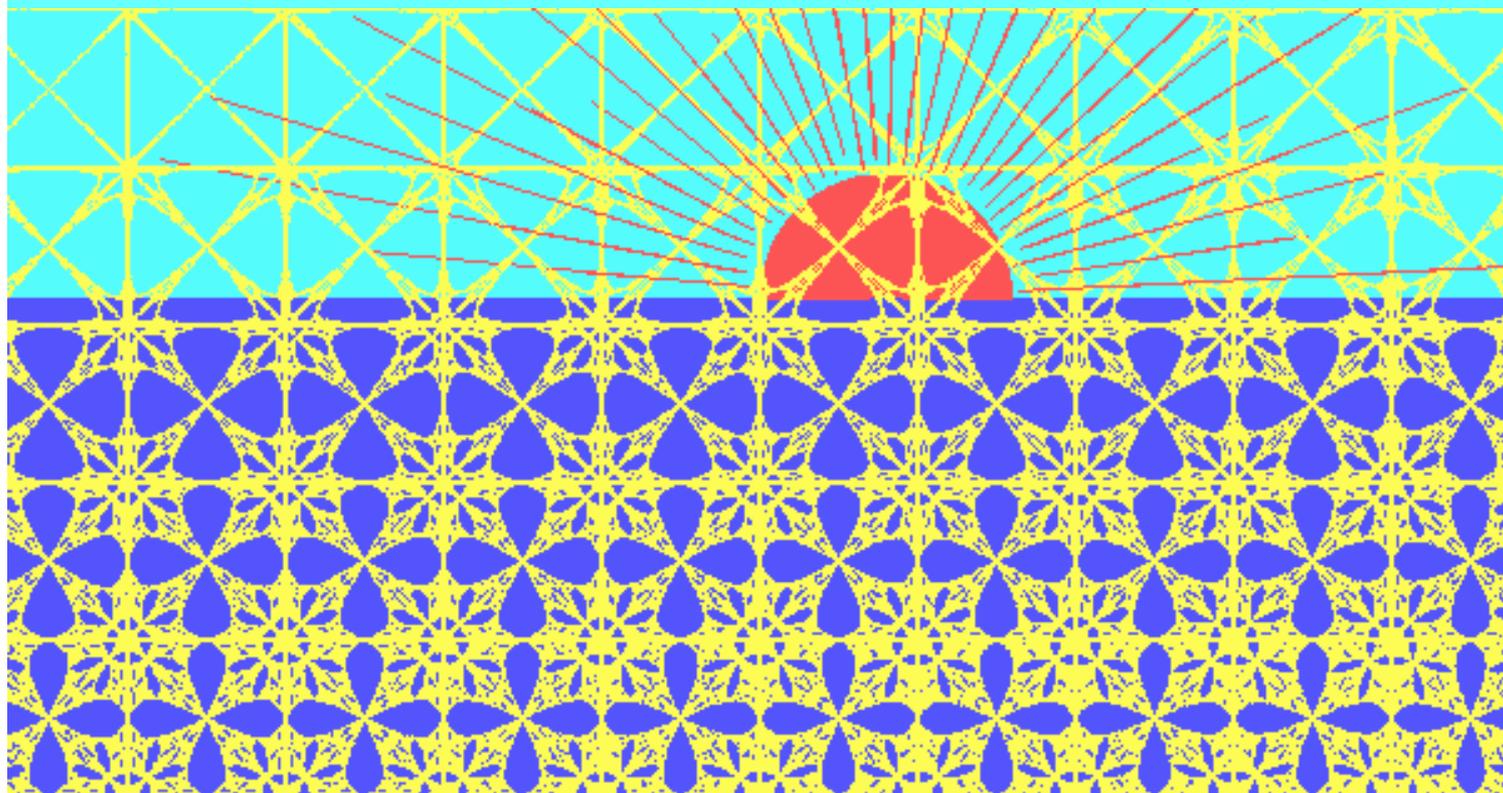


В. И. Губин, В. Н. Осташков

**Статистические методы
обработки
экспериментальных данных**

Учебное пособие



Тюмень 2007

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ НЕФТЕГАЗОВЫЙ УНИВЕРСИТЕТ»

В. И. Губин, В. Н. Осташков

Статистические методы обработки экспериментальных данных

*Рекомендовано
Министерством образования и науки
Российской Федерации в качестве учебного
пособия для студентов высших учебных
заведений, обучающихся по специальностям
«Технология машиностроения»,
«Машины и аппараты химических производств»*

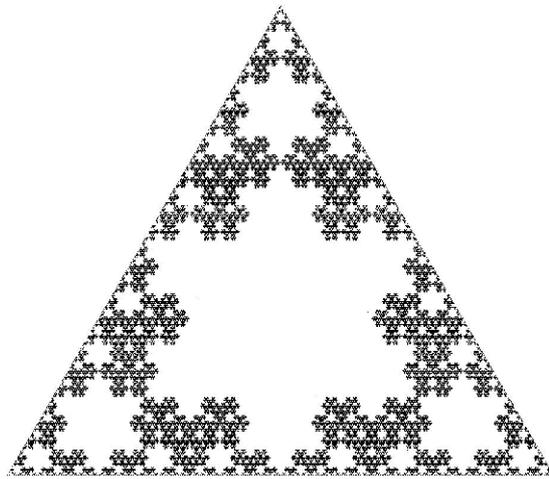
Тюмень
Издательство «ТюмГНГУ»
2007

В. И. Губин, В. Н. Осташков. Статистические методы обработки экспериментальных данных: Учеб. пособие для студентов технических вузов.— Тюмень: Изд-во «ТюмГНГУ», 2007.— 202 с.

Учебное пособие состоит из трех частей: теоретической, методической и практической. В первой части содержится теоретический материал справочного характера по разделу «Статистические методы обработки экспериментальных данных» курса математики. Во второй части приведены образцы выполнения лабораторных работ по первичной обработке результатов экспериментальных данных из различных сфер производственной деятельности, проверке статистических гипотез, построению однофакторных и многофакторных моделей. Третья часть включает варианты заданий для выполнения шести лабораторных работ.

По содержанию данное пособие соответствует требованиям ГОС ВПО для специальностей технических вузов. Может быть использовано преподавателями при организации и проведении лабораторных занятий по математической статистике во всех формах обучения.

Рецензенты: *В.Н. Кутрунов* — д.ф.-м.н., проф., зав. кафедрой математического моделирования Тюменского гос. ун-та; *С.Г. Обухов* — д.ф.-м.н., проф., зав. кафедрой математики и информатики Тюменского гос. нефтегазового ун-та.



ПРЕДИСЛОВИЕ

Математическая статистика — раздел математики, в котором изучаются математические методы сбора, обработки и использования информации, полученной путем наблюдений массовых случайных явлений, — статистических данных — для научных и практических выводов путем выявления существующих закономерностей.

Статистическими данными называется информация об объектах в какой-либо достаточно обширной совокупности, обладающий теми или иными признаками.

Считается, что математическая статистика — это теория принятия решений в условиях неопределенности.

Математическая статистика, возникнув в XVIII веке, получила свое развитие в работах Б. Паскаля, П. Ферма, Я. Бернулли, П. Лапласа. В современном развитии определяющую роль сыграли труды К. Пирсона, К. Крамера, Р. Фишера, Ю. Неймана и др. Значимый вклад в математическую статистику внесли русские ученые П. Л. Чебышёв, А.М. Ляпунов, А. Н. Колмогоров, Б. В. Гнеденко и другие.

Для изучения *случайных* массовых явлений строятся математические модели в терминах и понятиях математической статистики, фундаментом которой является *теория вероятностей*, увенчанная предельными теоремами. При этом теория вероятностей выводит из математической модели свойства, интерпретирующие реальный процесс, а математическая статистика устанавливает свойства математической модели, исходя из данных наблюдений (говорят: «из статистических данных»).

Предметом математической статистики является изучение сложных динамических систем путем объяснения (реальных наблюдений), управления (заключающегося в выборе решения), предсказания (возможных альтернатив, катастроф). *Математически* это приводит к необходимости построения модели, позволяющей решить триединую задачу: упорядочить данные (полученные в результате наблюдения, сначала надо каким-либо образом обработать, представить в удобном для обозрения и анализа виде), оценить статистику (хотя бы приблизительно, интересующие нас характеристики наблюдаемой случайной величины; например, дать оценку неиз-

вестной вероятности события, оценку неизвестной функции распределения, оценку математического ожидания, оценку дисперсии, случайной величины, оценку параметров распределения, вид которого неизвестен, и т. д.), проверить статистические гипотезы (т. е. решение вопроса согласования результатов оценивания с опытными данными; например, выдвигается гипотеза, что наблюдаемая случайная величина подчиняется нормальному закону; математическое ожидание наблюдаемой случайной величины равно нулю; случайное событие обладает данной вероятностью и т.д.).

Настоящее учебное пособие возникло на основе чтения авторами раздела «Вероятность и статистика», предусмотренного ГОСом, и написано в соответствии с государственным образовательным стандартом для студентов, обучающихся по специальностям 151001 — «Технология машиностроения» и 240801 — «Машины и аппараты химических производств».

Предлагаемое учебное пособие является руководством к лабораторному практикуму по следующим разделам математической статистики:

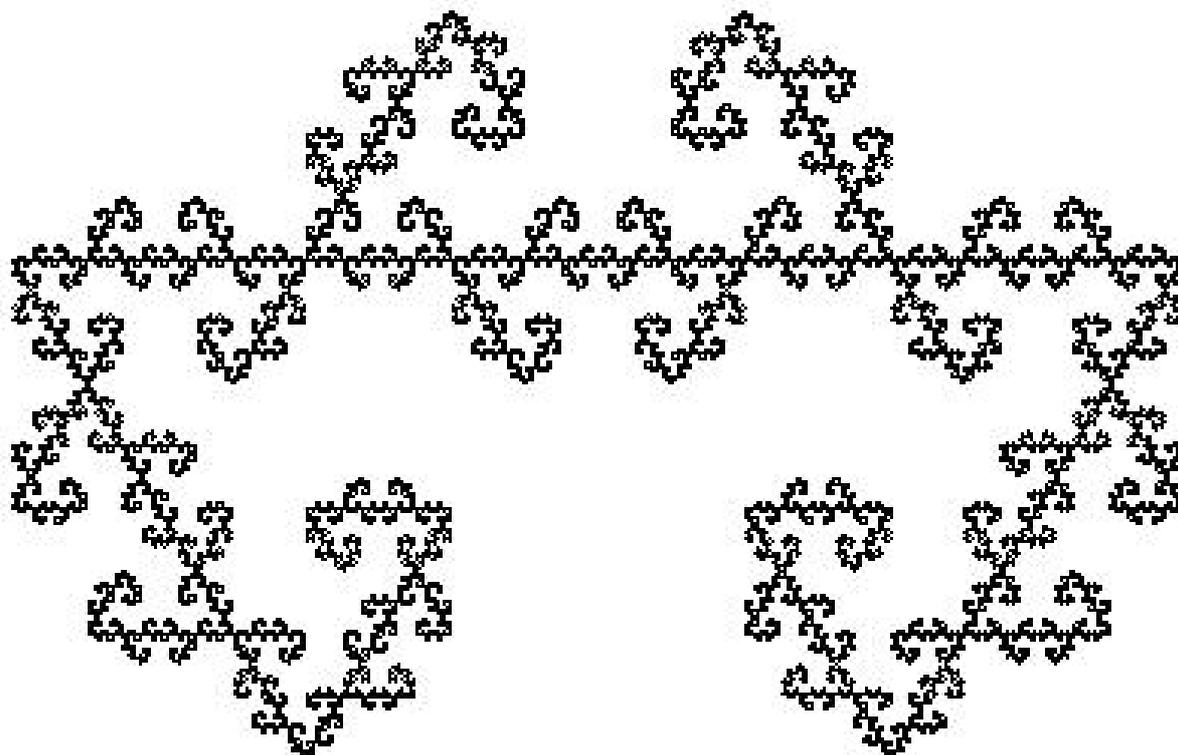
- построение вариационных рядов статистических распределений и расчет числовых характеристик;
- построение эмпирических, теоретических кривых распределений и проверка согласованности эмпирического распределения с нормальным теоретическим в соответствии с критериями согласия;
- построение математических моделей парных линейной и нелинейной корреляций;
- построение множественных линейных корреляционных моделей.

Пособие составлялось с учетом реализации в учебном процессе межпредметных связей математики с инженерными дисциплинами; применения математического моделирования для анализа производственных процессов и их прогнозирования; формирования знаний основных сведений математической статистики и умение использовать статистические методы реальных процессов для решения инженерных задач различной степени сложности. Кроме того, в этом пособии учтены рекомендации, изложенные в документе «Стандарты и Процедуры аккредитации инженерных программ, разработанные в рамках проекта EUR — ACE», которые предусматривают двухуровневую подготовку специалистов, формирование у студентов знаний инженерных дисциплин, математики, навыков анализа производственных процессов, принятия оптимальных решений и способности применения их в практической деятельности. Целью проекта EUR — ACE является содействие созданию Европейской Зоны Высшего образования (Болонский процесс), в котором участвует и Россия.

Указанное пособие может быть полезно студентам инженерно-технических и экономических специальностей высших учебных заведений, преподавателям, начинающим свою профессиональную деятельность, всем интересующимся рассматриваемыми вопросами математической статистики.

ГЛАВА 1

ВАРИАЦИОННЫЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ



§ 1. Первичная обработка результатов наблюдений

Информация о работе любой отрасли производственной сферы (добыча нефти и газа, ремонт скважин, нефтехимическое производство, машиностроение и т.д.) ставит перед ее руководством и наукой задачу: как, сведя к минимуму расходы по использованию природных, материальных и людских ресурсов, эффективно анализировать работу отрасли, управлять ею, прогнозировать развитие возможных сценариев поведения отрасли как сложной системы. Это означает, что математическому моделированию подлежит (дискретный, непрерывный, фрактальный) информационный по-

ток — *статистическая совокупность* — в виде случайных событий и случайных величин. Изучение подобного рода *массовых явлений*, выявление их статических и динамических закономерностей становится *предметом математической статистики*. Среди полезной информации о статистической совокупности особый интерес представляют *статистические данные*, которые можно записать в виде ряда $\{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ числовых значений интересующего нас *признака* (случайной величины X). Обработку этого ряда производят посредством *методов* математической статистики; при этом *точность* статистических методов повышается с ростом n .

Пусть A — некоторое множество (например, множество всех жителей данного города), $B \subset A$ — случайно выбранное подмножество (например, множество случайно выбранных жителей, при этом некий наблюдатель измерил у них рост, скажем, в сантиметрах). *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*. Число $N = |A|$ элементов множества A называется *объемом генеральной совокупности*, а число $n = |B|$ — *объемом выборки*. При изучении некоторого *признака* X (в нашем примере — роста) выборки производят *испытания* или *наблюдения* (измерение роста).

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 статистики. *Статистикой* называется функция

$$\begin{array}{ccc} \mathbf{R}^n & \xrightarrow{f} & \mathbf{R} \\ \Psi & & \Psi \\ (x^{(1)}, x^{(2)}, \dots, x^{(n)}) = x & \mathbf{a} & f(x), \end{array}$$

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:

$$x_1, x_2, \dots, x_n \quad (x_1 < x_2 < \dots < x_n).$$

Может оказаться, что 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 вариант называются их *весами*. Отношение $w_i = n_i / n$ частоты n_i к объему n выборки называют *относительной частотой* (*частотью*) варианты x_i .

Вариационным рядом (или *статистическим распределением*) называется ранжированный в порядке возрастания или убывания ряд вариант с соответствующими им весами.

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

Т а б л и ц а 1

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

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

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. Число k частичных интервалов выбирают, пользуясь одним из следующих правил:

$$1) 6 \leq k \leq 20, \quad 2) k \approx \sqrt{n}, \quad 3) k \approx 1 + \log_2 n \approx 1 + 3,221 \cdot \lg n.$$

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 до некоторого значения d . Например, если результаты x_i признака X — целые числа, то h округляют до целого значения, если x_i содержат десятичные знаки, то h округляют до значения d , содержащего такое же число десятичных знаков. 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* $w_i = n_i / n$.

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 *частотам* соответствующих интервалов.

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 $\sum n_i = n$, где n — объем выборки.

Таблица 4

Варианты, x_i	x_1	x_2	...	x_k
относительные частоты, $w_i = n_i / n$	w_1	w_2	...	w_k

Здесь $\sum_{i=1}^k w_i = 1$.

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* (*частотью*). 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 . Со-

единия 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.

Т а б л и ц а 5

Варианты, x_i	x_1	x_2	...	x_k
Относительные частоты, $w_i = n_i / n$	$w_1 = n_1 / n$	$w_2 = n_2 / n$...	$w_k = n_k / n$
Накопленные относительные частоты, $W_i = W_{i-1} + w_i$	$W_1 = w_1$	$W_2 = W_1 + w_2$...	$W_k = W_{k-1} + w_k$

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. 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

$$\lim_{n \rightarrow \infty} \mathbf{P} [|F(x) - F_B(x)| < \varepsilon] = 1 \quad (\varepsilon > 0).$$

Другими словами, при больших объемах выборки n , согласно закону больших чисел, функция $F_B(x)$ сходится по вероятности к теоретической функции $F_B(x)$ 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 $\mathbf{P}(X < x)$ берется накопленная частота $W_i = \frac{n_x}{n}$. 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 ре-

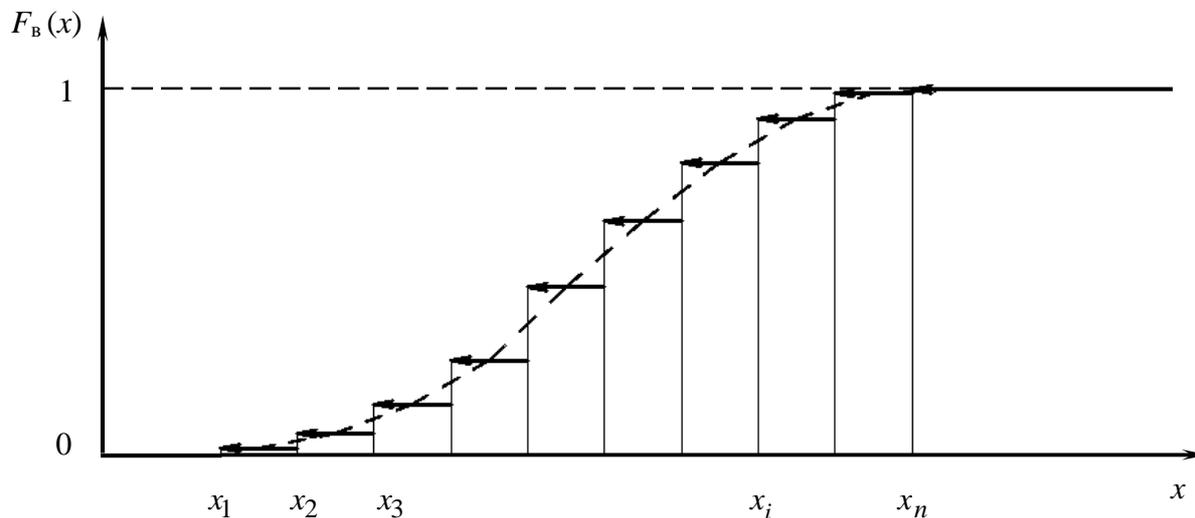
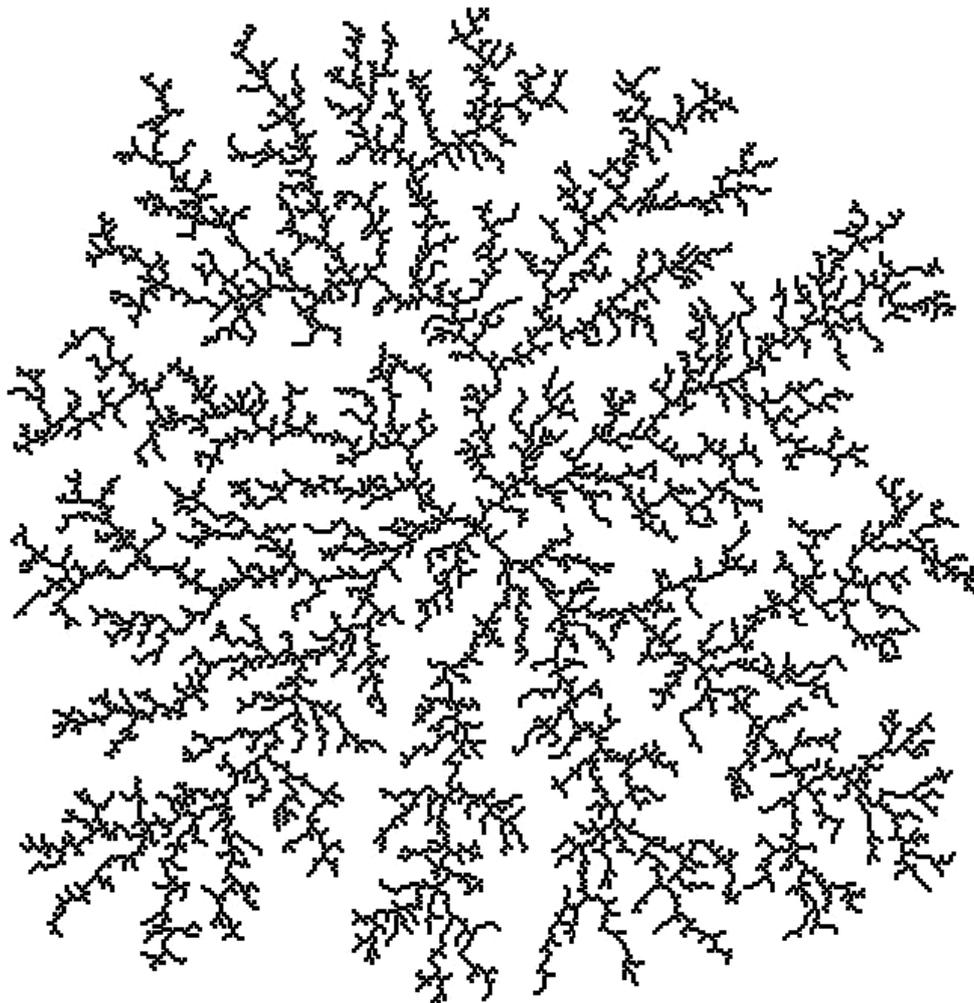


Рис. 1. Кумулята и эмпирическая функция распределения.

зультатов эксперимента, например, результатов механических испытаний, целесообразно вместо ступенчатой кривой вычерчивать плавную кривую (на рис. 1 это штриховая линия), которая проходит через точки, расположенные посередине вертикальных частей ступенчатой кривой [11]. Абсциссами этих точек служат значения механической характеристики x_i , а ординатами — эмпирическая функция $F_B(x)$, характеризующая оценку вероятности события $X \leq x_i$.



§ 2. Расчет выборочных характеристик статистического распределения

Рассмотрим выборку 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 σ^2 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 [8].

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

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

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

Несомненно, что формулы (2) и (4), как и (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:

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

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_oX называют варианту, которая имеет наибольшую частоту. Например, для вариационного ряда

x_i	4	9	14	19
n_i	3	7	2	5

мода равна $M_oX = 9$.

Медианой M_eX называют варианту, которая делит вариационный ряд на равные по числу вариант части [2].

При нечетном объеме выборки $n = 2k + 1$ медиана равна $M_eX = x_{k+1}$. Например, для вариационного ряда

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

медиана равна $M_eX = x_{13} = 12$.

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

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

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 $M_eX = 7$.

Для вычисления выборочной средней \bar{x} , выборочной дисперсии S^2 , асимметрии A_s и эксцесса E_x при достаточно большом объеме выборки ($n > 30$) применяют метод произведений [3]. При этом вводят условные варианты u_i , которые вычисляют по формуле:

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

где $C = M_oX$, h — шаг (длина интервала).

Составляется расчетная табл. 6.

Т а б л и ц а 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 $n_i(u_i + 1)^2$
строка сумм:	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$

Контроль вычислений ведут по формуле:

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

Пользуясь табл. 6, вычисляют [3] 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:

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

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

где

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

— условный центральный момент третьего порядка, а

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

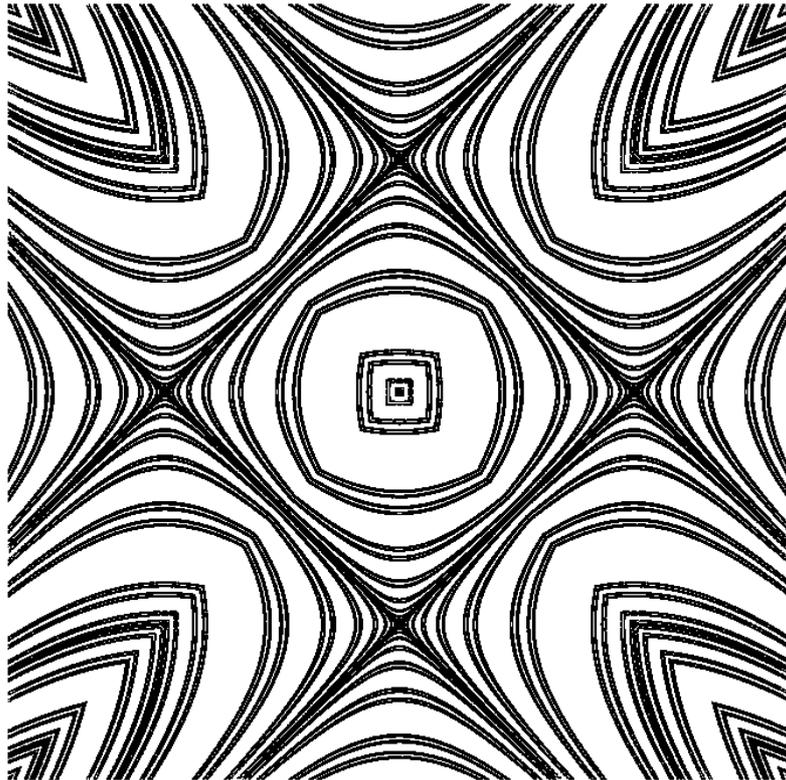
— условный центральный момент четвертого порядка.

Для характеристики колеблемости признака X используют [2] относительный показатель — *коэффициент вариации* V , который для положительной случайной величины X вычисляют по формуле:

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

Коэффициент вариации подобного вида был предложен Пирсоном (1895) в несколько иной форме:

$$V' = 100S/\bar{x}.$$



§ 3. Интервальные (доверительные) оценки параметров распределения

Выборочные характеристики \bar{x} и S^2 являются надежными количественными оценками генеральных характеристик a и σ^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 σ по выборочной средней \bar{x} и выборочному среднеквадратическому отклонению S находят [6] *доверительные интервалы* по формулам:

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

где t_γ находят из таблицы (см. приложение 3) по заданным n и γ (γ — *уровень доверия* или *надежность*, которая задается заранее).

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

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

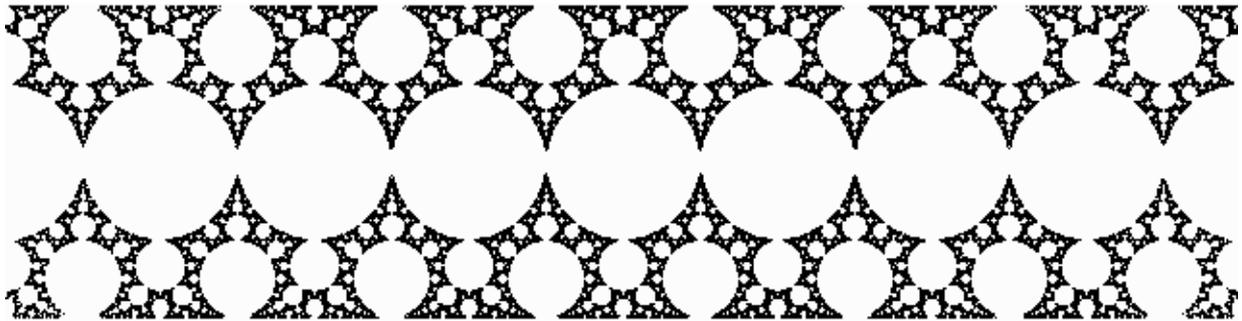
или

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

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

Контрольные вопросы

1. Что называется статистической совокупностью?
2. Что понимается под генеральной совокупностью?
3. Что называется выборкой?
4. Что называется вариационным рядом?
5. Описать алгоритм построения непрерывного вариационного ряда.
6. Рассказать о графическом изображении дискретного и непрерывного вариационных рядов.
7. Что называется эмпирической функцией распределения? Сформулировать ее свойства и рассказать о ее назначении.
8. По каким формулам находятся выборочные средние статистического распределения?
9. Дать определение выборочной дисперсии и рассказать о ее назначении.
10. Записать формулы для вычисления дисперсии для простой и взвешенной выборки.
11. Записать формулы для вычисления исправленной дисперсии и рассказать для чего она вводится.
12. Что называется модой и медианой вариационного ряда?
13. Рассказать о нахождении медианы при различном объеме выборки.
14. Сформулировать алгоритм вычисления \bar{x} и S^2 по методу произведений.
15. Дать определения асимметрии и эксцесса статистического распределения и рассказать об их назначении.
16. Записать доверительные интервалы для оценки генеральных математического ожидания и среднего квадратического отклонения.



§ 4. Лабораторная работа № 1. Построение вариационных рядов. Расчет числовых характеристик

Ц е л ь р а б о т ы: овладение способами построения рядов распределения и методами расчета числовых характеристик.

Выполнение лабораторной работы № 1 рассмотрим на примере следующей задачи.

З а д а ч а. 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	61,3
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	62,9	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

С о д е р ж а н и е р а б о т ы: на основе совокупности данных опыта выполнить следующее:

1. Построить ряды распределения (интервальный и дискретный вариационные ряды). Изобразить их графики.
2. Построить график накопительных частот — кумуляту.
3. Составить эмпирическую функцию распределения и изобразить ее графически.
4. Вычислить моду, медиану, выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение, коэффициент вариации, асимметрию, эксцесс.

5. Построить доверительные интервалы для истинного значения измеряемой величины и среднего квадратического отклонения генеральной совокупности.

6. Раскрыть смысловую сторону каждой характеристики.

Методические указания по выполнению работы

1.1. Построить интервальный вариационный ряд. Для этого найти:

а) размах варьирования признака по формуле $R = x_{\max} - x_{\min}$, где x_{\min} — наименьшая, x_{\max} — наибольшая варианты в данной выборочной совокупности;

б) число интервалов вариационного ряда, пользуясь одним из приведенных ниже соотношений:

$k \approx \sqrt{n}$, $6 < k < 12$, $k \approx 1 + \log_2 n \approx 1 + 3,2 \lg n$, где n — объем выборки;

в) длину h частичных интервалов по формуле $h = \frac{R}{k}$ и, если необходимо, округлить это значение до некоторого числа;

г) записать полученный интервальный вариационный ряд, заполнив табл. 2, § 1. Сделать контроль, убедившись, что $\sum n_i = n$.

1.2. Построить дискретный вариационный ряд, взяв в качестве вариант середины вариант-интервалов непрерывного вариационного ряда, а в качестве частот — частоты непрерывного вариационного ряда (табл. 3).

1.3. Изобразить графически интервальный и дискретный вариационные ряды (построить гистограмму и полигон частот).

2. Построить график накопленных частот — *кумуляту*, т.е. ломаную, проходящую через точки с координатами x_i и соответствующими накопленными частотами. Предварительно составить табл. 5 § 1.

3. Найти эмпирическую функцию распределения и изобразить ее графически.

4.1. Найти моду $M_o X$ и медиану $M_e X$.

4.2. Для вычисления остальных статистик воспользоваться методом произведений. Ввести условные варианты $u_i = \frac{x_i - C}{h}$, где $C = M_o X$, h — шаг (длина интервала). Составить расчетную табл. 6, § 2.

Контроль вычислений произвести по формуле:

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

4.3. Пользуясь табл. 6, вычислить начальные моменты (11) — (14):

$$M_1^* = \frac{1}{n} \sum n_i u_i, \quad M_2^* = \frac{1}{n} \sum n_i u_i^2, \quad M_3^* = \frac{1}{n} \sum n_i u_i^3, \quad M_4^* = \frac{1}{n} \sum n_i u_i^4.$$

4.4. Найти выборочную среднюю $\bar{x} = M_1^* h + C$.

4.5. Найти выборочную дисперсию $D(X) = (M_2^* - M_1^{*2})h^2$.

4.6. Найти выборочное среднее квадратическое отклонение (17):

$$\sigma_B = S = \sqrt{D(X)}.$$

4.7. Найти коэффициент вариации (22): $V = S / \bar{x}$.

4.8. Найти центральные моменты (20), (21):

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

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

4.9. Вычислить асимметрию (18): $A_s = \frac{m_3}{S^3}$ и эксцесс (19): $E_x = \frac{m_4}{S^4} - 3$.

5. Доверительные интервалы для a и σ найти, согласно (23), (24):

$$\bar{x} - t_\gamma \cdot \frac{S_x}{\sqrt{n}} < a < \bar{x} + t_\gamma \cdot \frac{S_x}{\sqrt{n}} \text{ при } \gamma = 0,95.$$

Величину t_γ найти по приложению 3, а по приложению 4 найти величину q , удовлетворяющую одному из условий:

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

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

6. Раскрыть смысловую сторону каждой характеристики.

Выполнение работы

Обозначим через X обводненность нефти из рассматриваемых насосных скважин.

1.1. По данным выборки строим интервальный вариационный ряд.

а) Поскольку, как легко выяснить, $x_{\max} = 62,8$, $x_{\min} = 60,1$, то размах варьирования признака X равен $R = x_{\max} - x_{\min} = 62,8 - 60,1 = 2,7$.

б) Определяя число k интервалов (число столбцов в таблице) вариационного ряда, положим $k = 10$.

в) Длина h каждого частичного интервала равна $h = \frac{R}{k} = \frac{2,7}{10} = 0,27$.

Так как исходные данные мало отличаются друг от друга и содержат один

десятичный знак, то величину h округляем до одного десятичного знака: $h = 0,3$.

г) Подсчитываем 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 $x_0 = x_{\min} - 0,5h = 60,1 - 0,5 \cdot 0,3 = 59,95 \approx 60$. Конец x_k последнего интервала находим по формуле $x_k = x_{\max} + 0,5h = 62,8 + 0,15 = 62,95 \approx 63,0$. 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: $\sum n_i = 100$, и объем выборки $n = 100$.

1.2. Записываем дискретный вариационный ряд (табл. 8). В качестве вариант x_i берем середины интервалов интервального вариационного ряда.

Таблица 8

варианты, x_i	60,15	60,45	60,75	61,05	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

1.3. Изображаем интервальный и дискретный вариационные ряды графически, построив гистограмму и полигон частот в одной системе координат (рис. 2).

2. Строим график накопленных частот — кумуляту (рис. 3). Предварительно составляем расчетную табл. 9.

Таблица 9

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

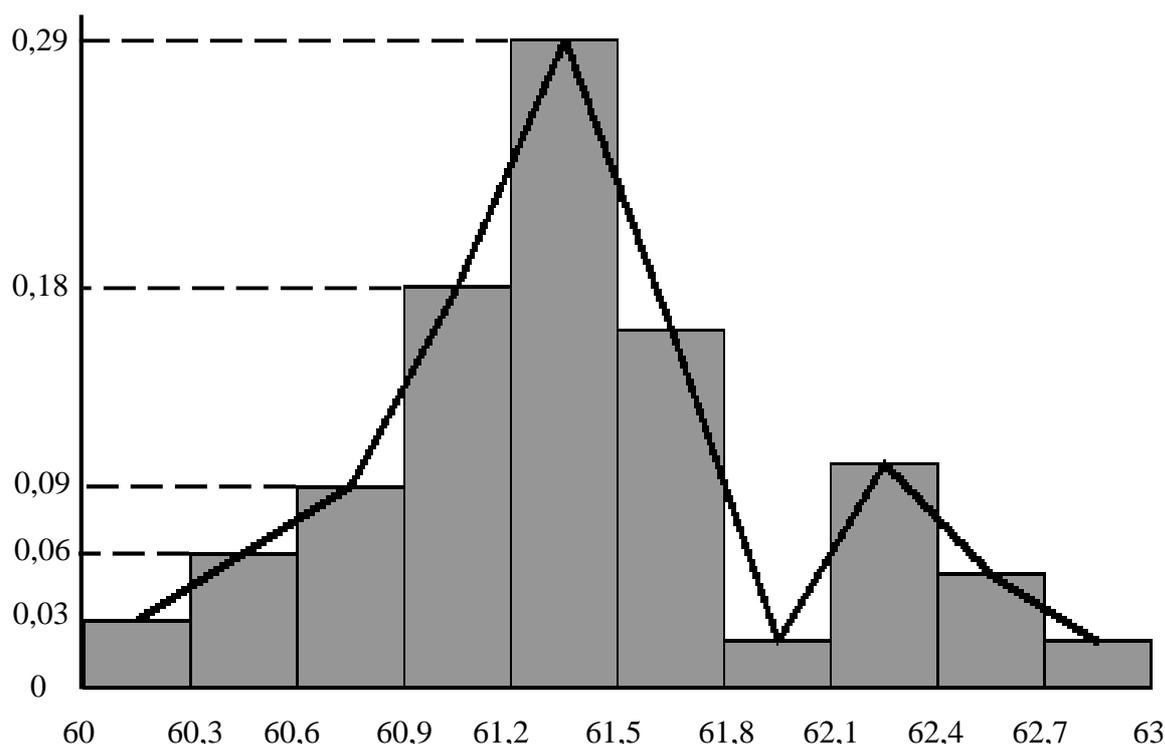


Рис.2. Гистограмма и полигон.

3. Находим эмпирическую функцию распределения. Воспользуемся формулой (1): $F_B(x) = \frac{n_x}{n}$.

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

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

$$\text{Если } 60,45 < x \leq 60,75, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6}{100} = 0,09.$$

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

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

$$\text{Если } 61,35 < x \leq 61,65, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{36+29}{100} = 0,65.$$

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

$$\text{Если } 61,95 < x \leq 62,25, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{81+2}{100} = 0,83.$$

Если $62,25 < x \leq 62,55$, то $F_B(x) = \frac{n_x}{n} = \frac{83+10}{100} = 0,93$.

Если $62,55 < x \leq 62,85$, то $F_B(x) = \frac{n_x}{n} = \frac{93+5}{100} = 0,98$.

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

Записываем полученную эмпирическую функцию в виде:

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

График функции $F_B(x)$ представлен на рис.4.

Соединив середины вертикальных частей ступенчатой кусочно-постоянной кривой, являющейся графиком функции $F_B(x)$, получаем плавную кривую (на рис. 4 это штриховая линия). Абсциссами точек этой кривой служат значения обводненности нефти, добываемой насосным способом из скважин, а ординатами — значения эмпирической функции распределения, характеризующей оценку вероятности события $X \leq x_i$, т.е. вероятности попадания возможных значений обводненности нефти на промежуток $(-\infty, x_i]$.

Для нахождения числовых характеристик признака X — обводненности нефти (*несмещенных оценок* для $M(X) = a$, $D(X)$, а также $M_e X$, $M_o X$, A_s , E_x) воспользуемся табл. 8.

4.1. Так как варианта $x = 61,35$ в табл. 8 встречается с наибольшей частотой $n_5 = 29$, то $M_o X = 61,35$, т.е. это значение обводненности нефти, встречающееся в данной выборке с наибольшей частотой.

Находим $M_e X$. Так как табл. 8 содержит четное число столбцов, то $M_e X = \frac{61,35+61,65}{2} = 61,5$. Это значение обводненности нефти, которое делит данные выборки признака X на равные части.

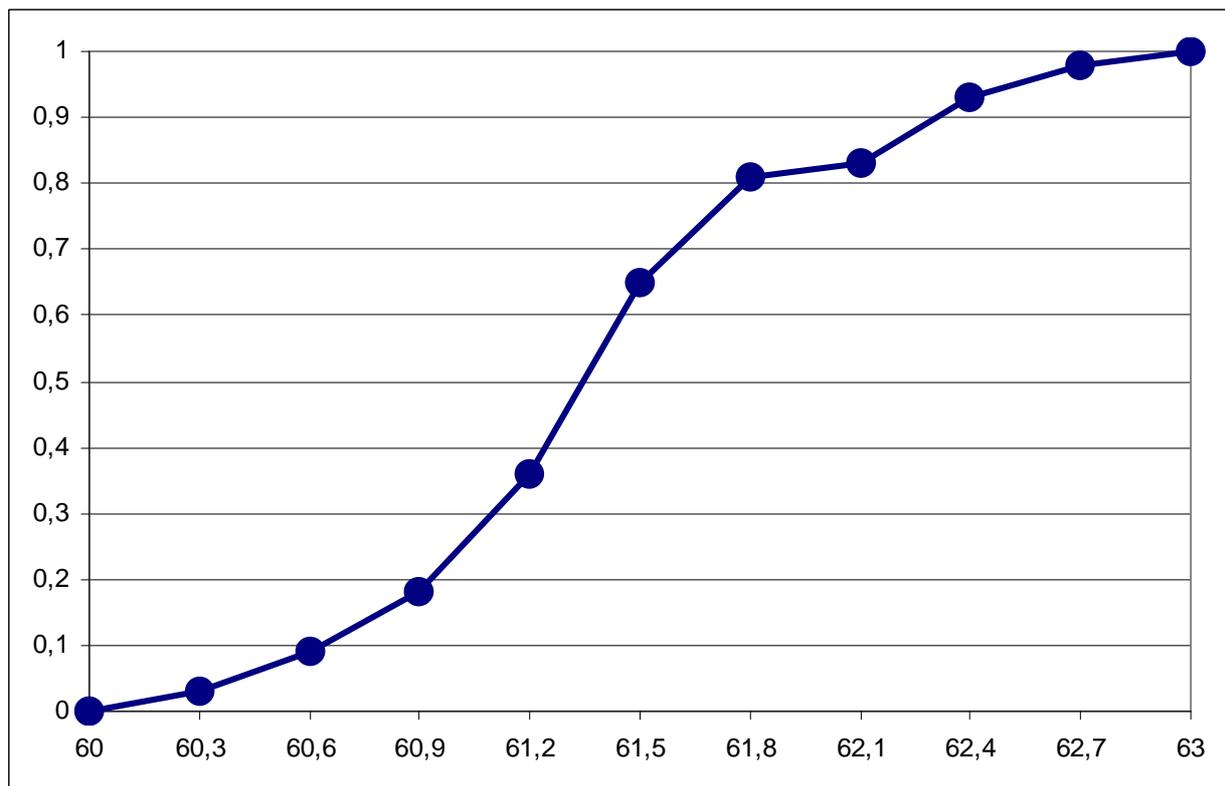


Рис. 3. Кумулятивная кривая.

4.2. Для нахождения остальных статистик, характеризующих обводненность нефти, воспользуемся методом произведений. Введем, согласно (9), условные варианты $u_i = \frac{x_i - C}{h}$; $C = M_o X = 61,35$, $h = 0,3$.

Составим расчетную табл. 10.

Таблица 10

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	контрольный 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:

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

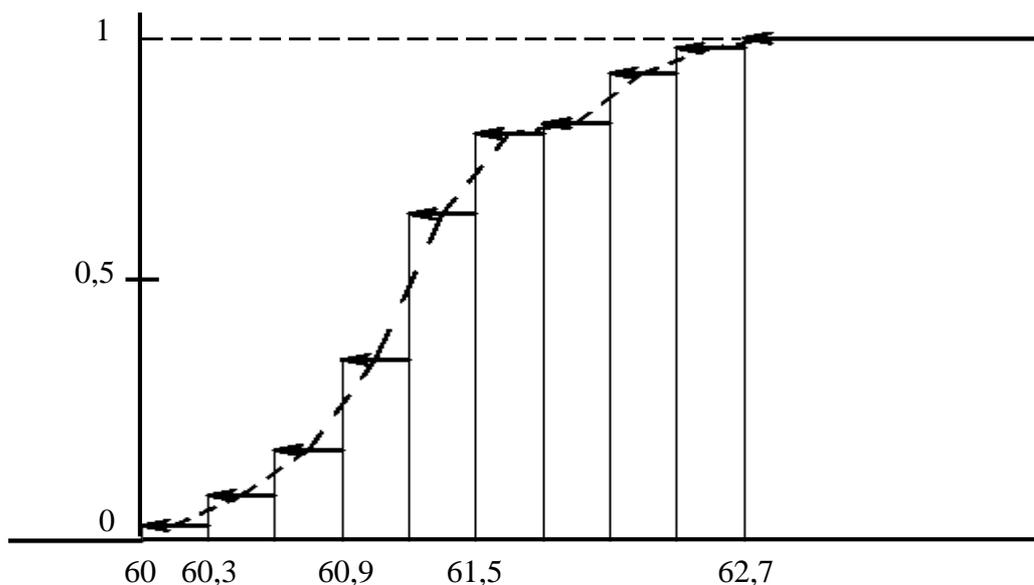


Рис.4. Кумулята и эмпирическая функция распределения.

т.е.

$$100 + 2 \cdot 14 + 400 = 528 = 27 + 24 + 9 + 29 + 64 + 18 + 160 + 125 + 72.$$

Следовательно, вычисления проведены верно.

4.3. Пользуясь результатами последней строки табл. 10, находим условные начальные моменты (11) — (14):

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

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

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

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

4.4. Находим выборочную среднюю (15):

$$\bar{x} = M_1^* h + C = 0,14 \cdot 0,3 + 61,35 = 61,392 \approx 61,39,$$

которая характеризует среднюю обводненность нефти из насосных скважин в данной выборке, составляющую 61,39 %.

4.5. Находим выборочную дисперсию (16):

$$S^2 = (M_2^* - M_1^{*2}) h^2 = (4 - 0,14^2) \cdot 0,09 = 0,358236$$

4.6. Вычисляем выборочное среднее квадратичное отклонение (17):

$$S = \sqrt{S^2} = \sqrt{0,358236} = 0,59853.$$

4.7. Величина $S = 0,599$ характеризует степень рассеяния значений обводненности нефти относительно средней обводненности. Для определения колеблемости значений обводненности нефти в процентном отношении вычисляем коэффициент вариации (22):

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

Величина коэффициента вариации мала (составляет 0,01), что означает тесную сгруппированность значений обводненности нефти около центра рассеяния, т.е. около средней обводненности нефти.

4.8. Для предварительной оценки отклонения значений обводненности нефти от нормального распределения вычисляем асимметрию и эксцесс. Сначала находим центральные моменты третьего и четвертого порядков (20), (21):

$$m_3 = (M_3^* - 3M_2^*M_1^* + 2M_1^{*3})h^3 = (4,28 - 3 \cdot 4 \cdot 0,14 + 2 \cdot 0,14^3) \cdot 0,3^3 = 0,07035.$$

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

4.9. Тогда в соответствии с (18), (19), находим:

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

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

Резюме. Значения A_s и E_x мало отличаются от нуля. Поэтому можно предположить близость данной выборки, характеризующей обводненность нефти, к нормальному распределению. Эта гипотеза будет проверяться в лабораторной работе № 2.

5. Произведем оценку генеральной средней $M(X) = a$ и генерального среднеквадратического отклонения $\sigma = S$ по выборочным статистикам \bar{x} и S , используя теорию доверительных интервалов для нормального распределения.

Доверительный интервал для истинного значения обводненности нефти с надежностью $\gamma = 0,95$ находим, согласно (23):

$$\bar{x} - \frac{S}{\sqrt{n}} \cdot t_\gamma < a < \bar{x} + \frac{S}{\sqrt{n}} \cdot t_\gamma.$$

Согласно приложению 3, при $n=100$ и $\gamma=0,95$ находим $t_\gamma=1,984$.
Записываем доверительный интервал:

$$61,39 - \frac{0,599}{10} \cdot 1,984 < a < 61,39 + \frac{0,599}{10} \cdot 1,984,$$

или $61,27 < a < 61,51$.

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

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

$$S(1 - q) < \sigma < S(1 + q),$$

или

$$0,599(1 - 0,143) < \sigma < 0,599(1 + 0,143),$$

или

$$0,51 < \sigma < 0,68;$$

следовательно, отклонения истинных значений обводненности нефти из насосных скважин не должны выходить за пределы промежутка (0,51;0,68).

Варианты заданий по лабораторной работе № 1

Вариант № 1. Имеются данные о производительности труда (количество деталей в смену):

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

Вариант № 2. Имеются данные о пропускной способности 50 участков нефтепровода ($\text{м}^3/\text{сут.}$):

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

Вариант № 8. Имеются данные о числе рабочих дней без простоя для пятидесяти буровых бригад одного из районов страны:

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

Вариант № 9. Приведено количество деталей, выработанных за смену различными рабочими:

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

Вариант № 10. Имеются данные о рабочих дебитах газовой скважины (тыс. м³/сут.):

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

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

Вариант № 13. Даны значения обследуемого признака X — себестоимости единицы продукции (в руб.):

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

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	18,5	20
22	24	28	31	33	25	18	21	26	30	32	34	29
26	21	20	23	25	27	30	32	29	27	23		

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 демонтажом 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 (в руб.):

82	83	73	76	79	89	95	92	93	84	88	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
84	95	79	79	83	96	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:

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

Вариант № 19. Даны значения израсходованных долот на 100 скважинах при механической скорости проходки 18 м/сек.:

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

Вариант № 20. 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 № 21. 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 № 22. Имеются данные об энергетических затратах на 1 м проходки при разведочном бурении нефтяных скважин в различных нефтяных районах страны (в тыс. руб.):

48	29	6	18	24	30	35	25	17	24	36	42	47
40	28	12	7	25	23	33	28	19	14	8	40	27
20	27	15	6	16	25	34	17	25	46	6	51	13
28	37	43	27	38	53	24	41	21	11	26		

Вариант № 23. Имеются данные о пластовом давлении (в атм.) при насосном способе эксплуатации 100 скважин:

95	57	15	26	35	46	52	55	59	47	42	48	58
55	102	96	45	54	56	60	10	16	20	49	48	43
12	19	51	103	62	61	38	29	10	39	40	18	14
41	58	63	59	60	63	68	70	71	75	82	87	92
99	65	68	78	91	94	77	65	79	67	74	80	89
69	81	83	100	90	36	64	97	50	76	72	31	55
28	57	85	69	13	53	11	61	90	76	17	37	

Вариант № 24. Имеются данные о продолжительности (в мес.) 50 фонтанирующих скважин:

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

Вариант № 25. Имеются данные замеров температуры масла двигателя автомобиля ГАЗ-53А:

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	30	34	25	24	33	

Вариант № 26. Результаты измерения температуры раздела фракции бензин-авиакеросин на установке первичной переработке 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 № 27. Имеются данные о суточном дебите нефти наблюдаемой скважины (в т/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	1	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				

Вариант № 28. 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 № 29. Имеются данные о расходах, связанных с подготовительными работами, на 1 м проходки при разведочном бурении нефтяных скважин в различных нефтеносных районах страны (в тыс. руб.):

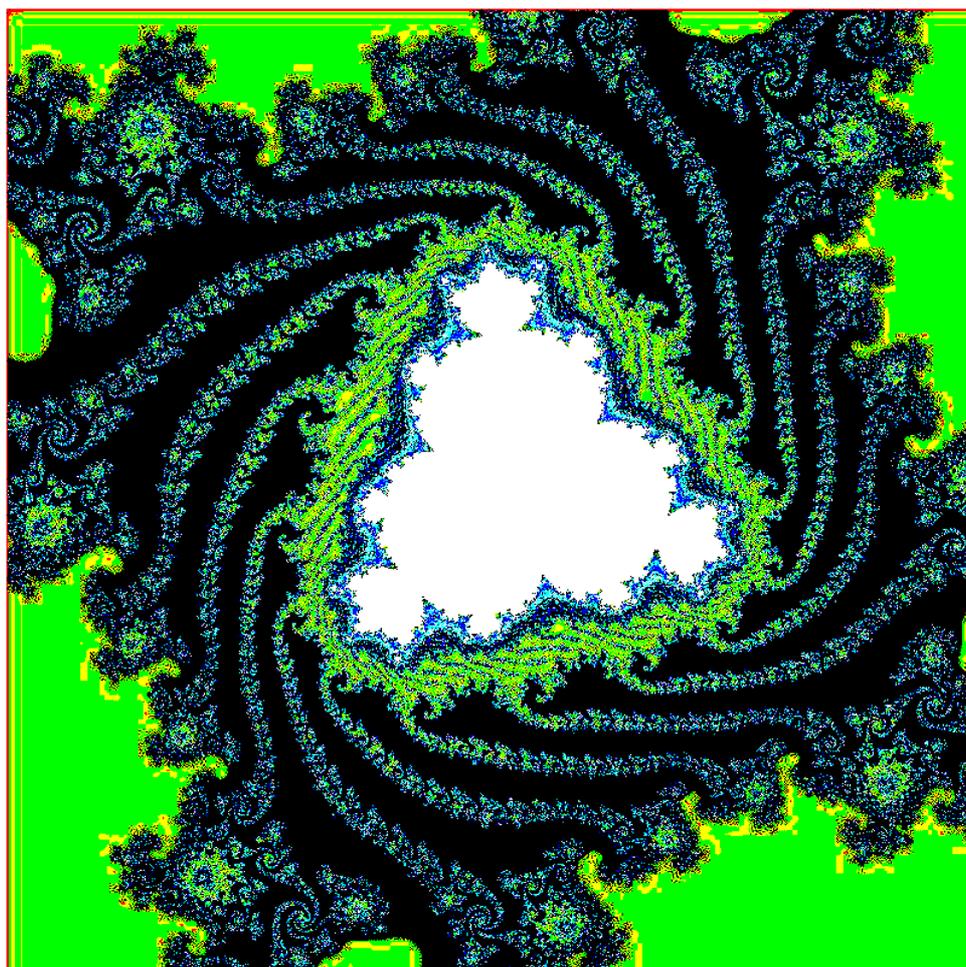
11	15	20	25	29	34	19	25	16	21	29	20	21
22	23	26	28	30	18	13	17	22	29	26	39	14
16	24	27	25	31	32	23	37	23	27	37	36	42
32	34	39	38	44	28	33	23	35	36	34		

Вариант № 30. 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	23	29	37	29	31
29	31	29	31	29	39	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				

ГЛАВА 2

ПРОВЕРКА СООТВЕТСТВИЯ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ



§ 5. Построение кривой нормального 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 вероятност-

ной сетке изображается прямой линией. Если экспериментальные точки ложатся вблизи прямой со случайными отклонениями вправо или влево, то опытные данные соответствуют рассматриваемому закону распределения. 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 [7]. 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 μ , σ совпадают с статистиками \bar{x} , S . Для нахождения теоретических частот n'_i составляется табл. 11:

Т а б л и ц а 1 1

x_i	n_i	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\Phi(u_i)$	$y_i = \frac{nh}{S} \Phi(u_i)$	n'_i

где x_i — варианты дискретного вариационного ряда,

n_i — частоты вариант x_i ,

\bar{x} — выборочная средняя,

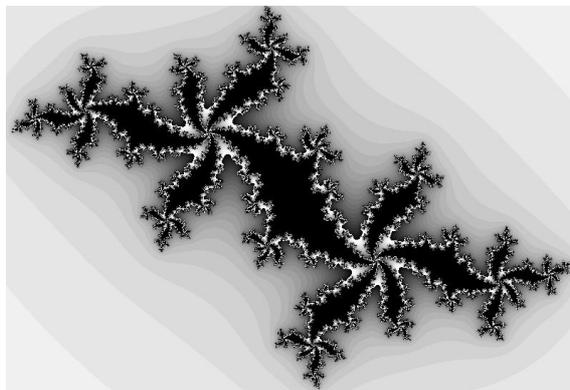
S — выборочное среднее квадратическое отклонение,

h — шаг (разность между соседними вариантами),

$\Phi(u_i)$ — функция, значения которой находят по приложению 1,

y_i — выровненные частоты (ординаты) теоретической кривой,

n'_i — округленные частоты y_i до ближайшего целого числа.



§ 6. Статистические оценки параметров распределения

С помощью статистик \bar{x} и S^2 можно производить оценку неизвестных параметров предполагаемого закона распределения. 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 σ^2 . В этом случае оценку неизвестных параметров $M(X) = a$ и σ^2 осуществляют с помощью точечных статистических оценок.

Статистической оценкой неизвестного параметра Θ теоретического распределения называют его приближенное значение Θ^* .

Статистическая оценка Θ 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 $M(\Theta^*) \neq \Theta$.

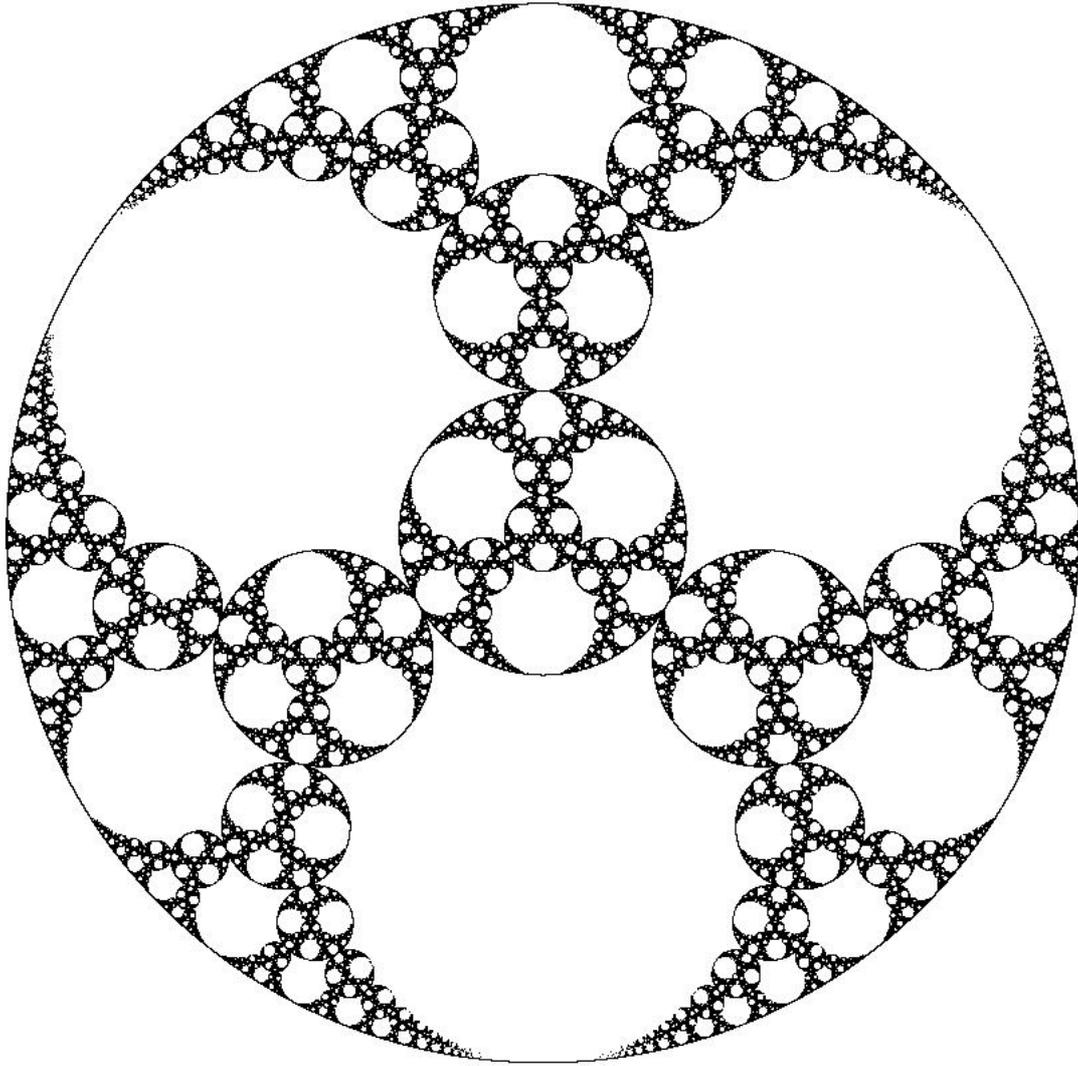
Оценка называется *состоятельной*, если при увеличении объема выборки она сходится по вероятности к оцениваемому параметру, т.е.

$$\lim_{n \rightarrow \infty} \mathbf{P} [|\Theta^* - \Theta| < \varepsilon] = 1 \quad (\varepsilon > 0).$$

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

Известно [3], что средняя выборочная \bar{x} является несмещенной и состоятельной оценкой генеральной средней $M(X)$, а также доказано, что исправленная дисперсия \hat{S}^2 является несмещенной и состоятельной оценкой генеральной дисперсии D_{Γ} , а выборочная дисперсия S^2 — смещенной оценкой D_{Γ} .

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.



§ 7. Проверка статистических гипотез

При изучении той или иной генеральной совокупности нам неизвестен либо закон ее распределения, либо параметры распределения. В подобных случаях в математической статистике выдвигается некоторое предположение относительно свойств генеральной совокупности. Такое предположение 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 статистиче-

ский 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. Границы 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 опытным данным [11].

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 α и называют *уровнем значимости*. Чем меньше α , тем меньше вероятность отвергнуть верную гипотезу. На практике в качестве α 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 β . Величину $1 - \beta$ называют *мощностью критерия*.

Между уровнем значимости α и мощностью критерия $1 - \beta$ существует связь: с уменьшением уровня значимости α , 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.

В математической статистике для проверки нулевой гипотезы H_0 используют следующие критерии: χ^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. 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 α 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), χ^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).

Критерий согласия Пирсона χ^2

Критерий согласия Пирсона χ^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 [9].

За меру расхождения эмпирического и теоретического распределений английский математик Пирсон принял величину χ^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 χ^2 к проверке гипотезы о нормальном распределении генеральной совокупности значений признака X осуществляется по следующему правилу.

Правило применения критерия χ^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, $u_i = \frac{x_i - \bar{x}}{S}$; $\varphi(u_i) = \frac{1}{\sqrt{2\pi}} e^{-u_i^2/2}$ находится по таблице приложения 1.

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

Т а б л и ц а 1 2

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 χ^2 по формуле (26) и обозначить ее через χ_0^2 . Расчет вести, пользуясь табл. 13.

Т а б л и ц а 1 3

n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
				χ_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 σ , а также объем выборки n).

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

Если $\chi_0^2 < \chi_{\text{кр}}^2$, то нет достаточных оснований отвергнуть выдвинутую гипотезу о нормальном распределении признака X . Если $\chi_0^2 > \chi_{\text{кр}}^2$, 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, по закону Пуассона, по 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 λ — параметр распределения Пуассона. Для применения критерия χ^2 надо рассчитать теоретические частоты n'_i , а также получить по значениям выборки оценку параметра λ . Методом максимального правдоподобия доказывается, что выборочная средняя \bar{x} является пригодной оценкой для λ , т.е. $\lambda \approx \bar{x}$. Теоретические частоты n'_i вычисляются по формуле:

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

Вероятности $\mathbf{P}(X = x_i)$ вычисляются по формуле:

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

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

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

Затем вычисляют величину χ^2 по формуле (26), обозначают ее через χ_0^2 . По заданному уровню значимости α и найденному числу степеней свободы k по приложению 5 находят $\chi_{\text{кр}}^2$. Если $\chi_0^2 < \chi_{\text{кр}}^2$, то нет достаточных оснований отвергнуть выдвинутую гипотезу H_0 . Если $\chi_0^2 > \chi_{\text{кр}}^2$, то гипотеза H_0 отвергается.

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 χ^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 [2].

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 λ критерия Колмогорова по формуле:

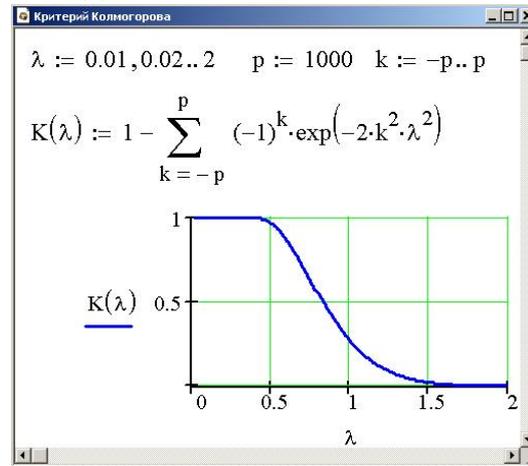
$$I = D / \sqrt{n}, \quad (33)$$

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

По вычисленному λ находят значение функции

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

— вероятность того, что λ достигает данной величины. График функции $K(\lambda)$ имеет следующий вид:



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

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

Таблица 14

λ	$K(\lambda)$	λ	$K(\lambda)$
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 проверяют гипотезу по критерию Колмогорова о соответствии выборки экспоненциальному распределению, параметры которого оценивают по опытным данным, то сначала вычисляют статистики [11]:

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

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

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

Затем составляют неравенство:

$$(D_n - \frac{0,2}{n})(\sqrt{n} + 0,26 + \frac{0,5}{n}) \leq \lambda_\alpha. \quad (37)$$

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

Критерий Ястремского

Для проверки соответствия данной выборочной совокупности признака X нормальному распределению составляется [2] 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.

Т а б л и ц а 15

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

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 вычисляют по формулам (18) и (19). Затем вычисляют [11] их средние квадратические отклонения по формулам:

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

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

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 χ^2 , A_s и E_x . Сначала вычисляют статистику χ^2 по формуле:

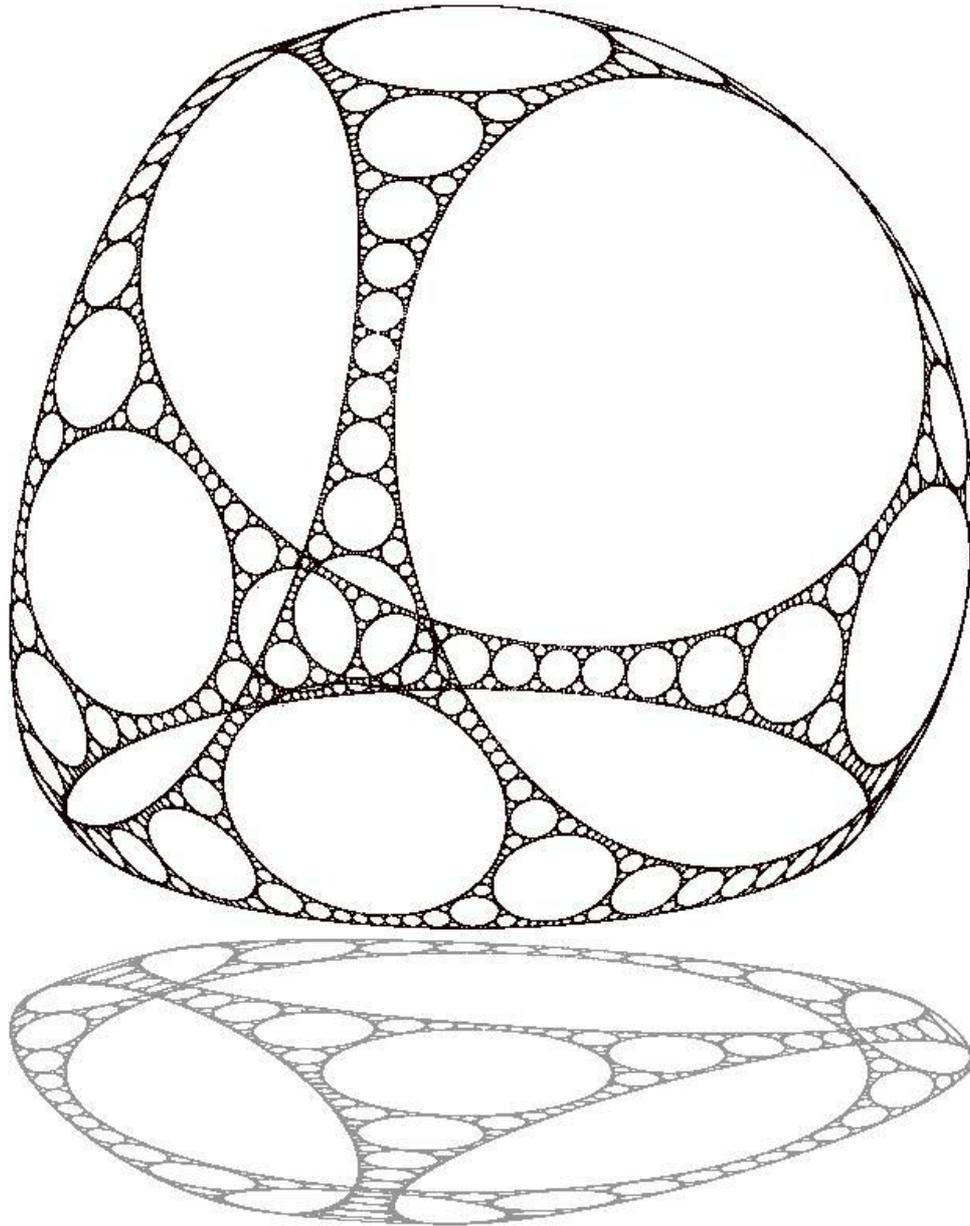
$$\chi^2 = \frac{A_s^2}{S_{A_s}^2} + \frac{E_x^2}{S_{E_x}^2}. \quad (41)$$

Затем при заданном уровне значимости α и числе степеней свободы $k = 2$ (используют в расчетах две статистики A_s и E_x) по приложению 5 для распределения χ^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.

Покажем применение рассмотренной теории на примере выполнения лабораторной работы № 2, являющейся продолжением лабораторной работы № 1.

Контрольные вопросы:

1. Рассказать о возможных вариантах построения кривой нормального распределения по опытным данным.
2. Дать определение статистической гипотезы.
3. Что называется статистическим критерием?
4. Описать алгоритм применения любого статистического критерия для обработки экспериментальных данных.
5. Сформулировать правило применения критерия согласия χ^2 Пирсона для проверки гипотезы согласованности эмпирического распределения с теоретическим нормальным.
6. Рассказать о применении критерия согласия Романовского для оценки близости эмпирического распределения к теоретическому нормальному.
7. Описать алгоритм применения критерия Колмогорова для проверки соответствия эмпирического распределения нормальному теоретическому распределению.
8. Рассказать о применении критерия Ястремского для проверки соответствия данной выборочной совокупности нормальному распределению.
9. Рассказать о приближенных критериях, применяемых для проверки гипотезы о нормальном распределении выборочной совокупности.



**§ 8. Лабораторная работа № 2.
Построение кривой нормального
распределения по опытным данным.
Проверка гипотезы о нормальном рас-
пределении 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.

Содержание работы: на основе дискретного вариационного

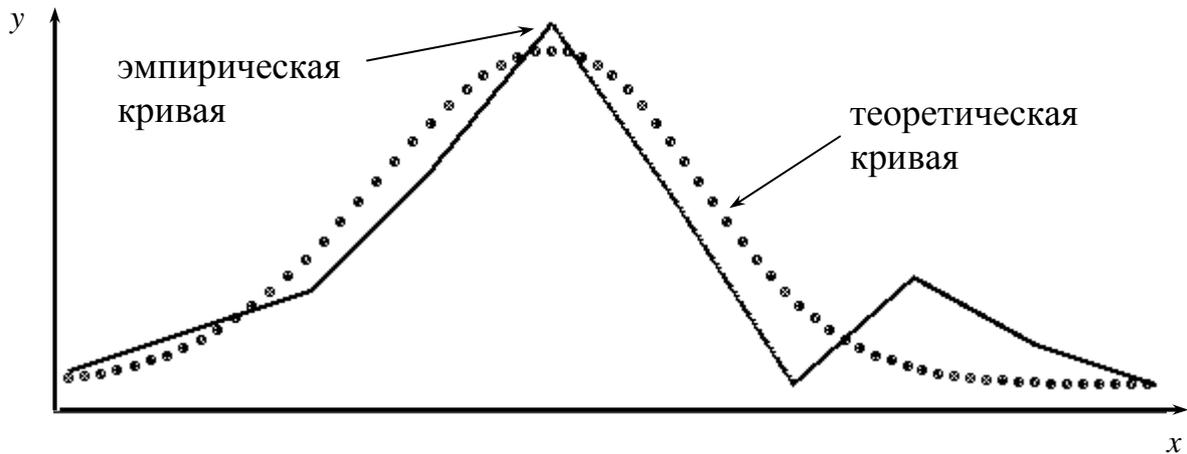


Рис. 5.

ряда, полученного в лабораторной работе № 1, выполнить следующее:

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

Методика выполнения работы

Продолжим вероятностно-статистическую обработку результатов эксперимента, предложенных в лабораторной работе № 1, то есть обводненности нефти из насосных скважин. За основу берем дискретный вариационный ряд в табл. 8

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

и значения $\bar{x} = 61,39$ и $S = 0,599$.

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 кривые (рис. 5).

Проверим согласованность эмпирического распределения (обводненности нефти из насосных скважин) с теоретическим нормальным по критерию Пирсона. Вычислим величину χ^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 χ^2 (приложение 5) находим $\chi_{кр}^2 = 2,17$. Так как $\chi_{кр}^2 < \chi_0^2$ ($2,7 < 18,12$), 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 не существенно, 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):

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

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

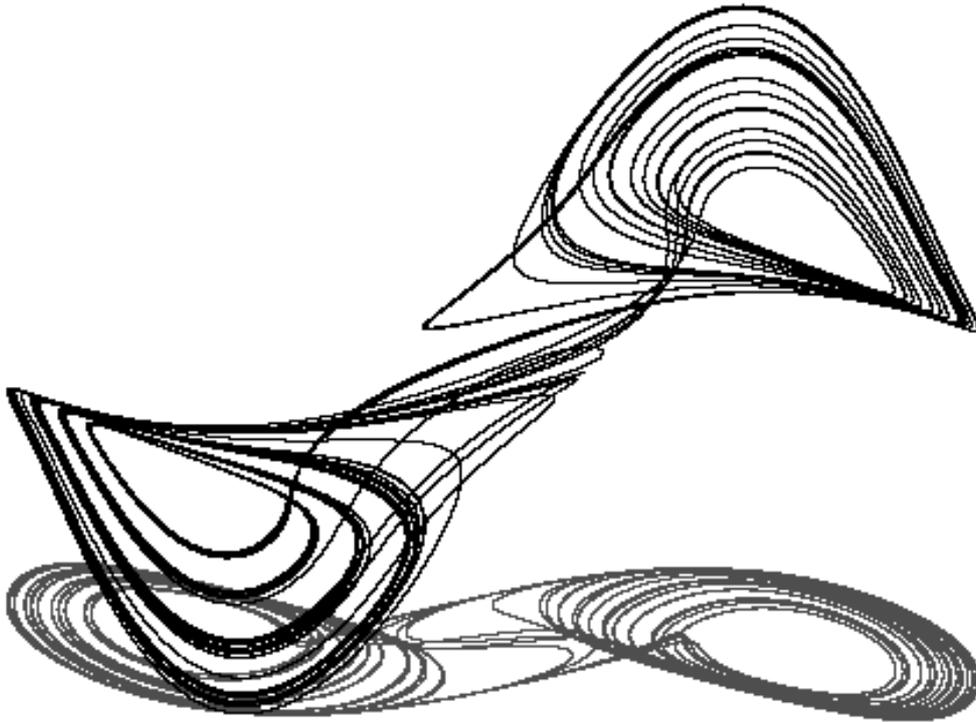
Так как $|A_s| = 0,328 > S_{A_s}$ и $|E_x| = 0,115 < S_{E_x}$, то делаем вывод, что данные выборки, характеризующие обводненность нефти из насосных скважин, не подчиняются нормальному закону распределения.

Итак, для проверки согласованности эмпирического распределения с теоретическим нормальным мы применили 4 критерия, два из них подтвердили близость выборочной совокупности к нормальному распределению. Однако, учитывая, что критерий Колмогорова является более мощным, чем критерий χ^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.

З а м е ч а н и е. В качестве вариантов заданий для выполнения лабораторной работы № 2 следует брать дискретные вариационные ряды из лабораторной работы № 1, а так же значения статистик \bar{x} , S , A_s , E_x .

ГЛАВА 3

ПАРНАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ



§ 9. Понятие корреляционной зависимости. Задачи теории корреляции

В новых условиях 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. При анализе 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:

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

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 [4].

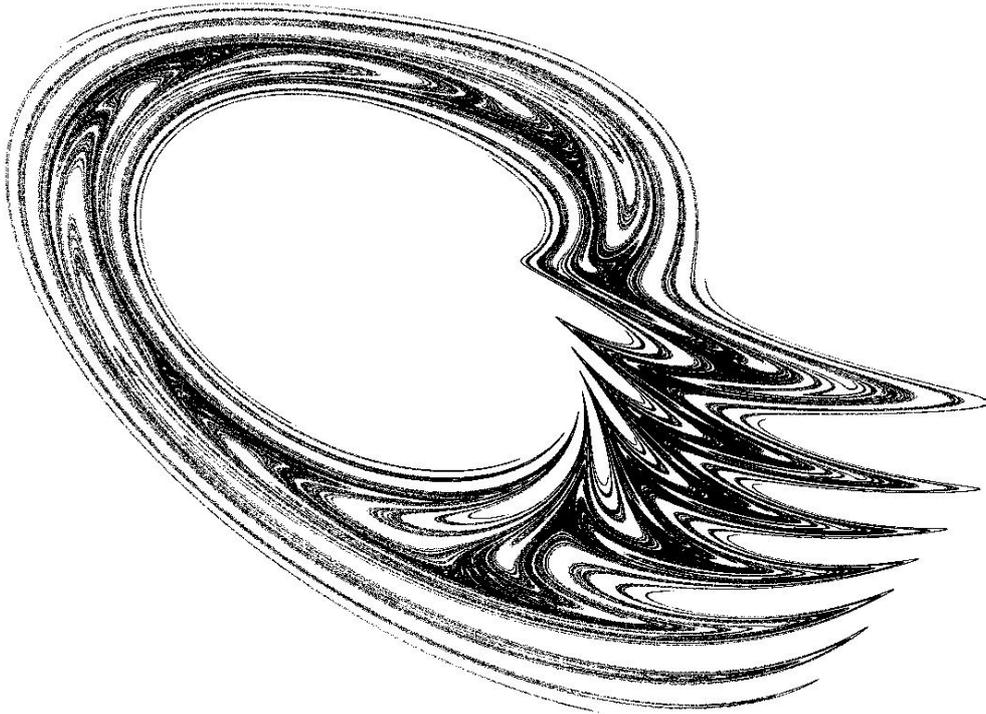
В теории корреляции решается триединая задача, методологической основой которой является триада:

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 η_{yx}, η_{xy} (в случае криволинейной корреляции).

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



§ 10. Парная линейная корреляция

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:

$$\hat{y}_x = a_0 + a_1 x. \quad (43)$$

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

Т а б л и ц а 1 8

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 [4], 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

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

Суммы $[n_x x]$, $[n_x x^2]$, $[n_{xy} xy]$ в табл. 20 находятся по строкам, а сумма $[n_y y]$ — 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, на сколько изменяется в среднем значение результативного признака Y при увеличении факторного признака на единицу.

Используя параметр a_1 , вычисляют [9] коэффициент эластичности $K_э$ по формуле:

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

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

В случае линейной корреляционной зависимости между признаками X и Y , уравнения регрессий находят [4] 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} , \bar{y} — выборочные средние признаков X и Y ; S_x , S_y — выборочные средние квадратические отклонения признаков X и Y , вычисляемые по формулам:

$$\hat{S}_x = \sqrt{\hat{S}_x^2}, \text{ где } \hat{S}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (n < 50), \quad (49)$$

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

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:

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

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 $C_1 = M_0X$, $C_2 = M_0Y$, h_1 и h_2 — шаги значений признаков X и Y .

Выборочный коэффициент линейной корреляции r в этом случае находят по формуле

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

где

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

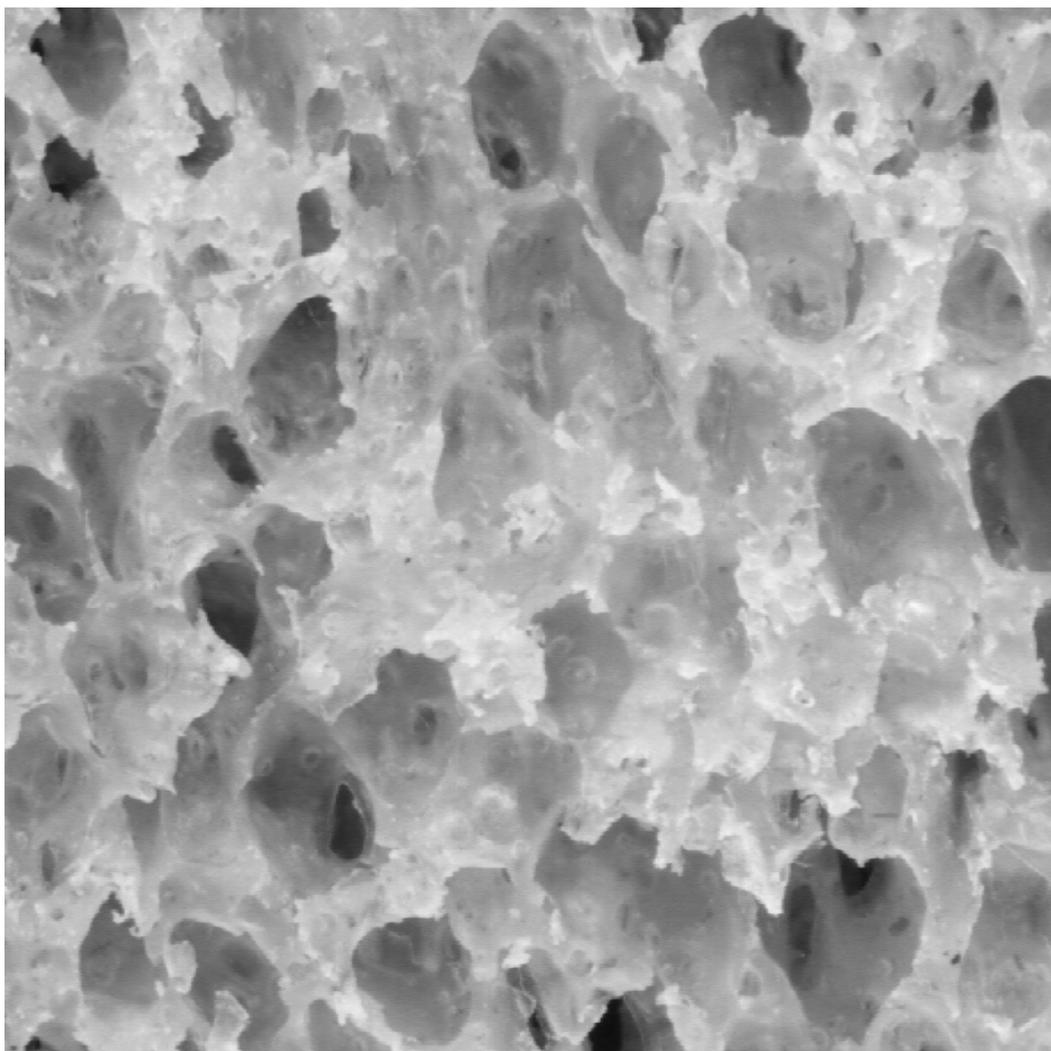
Для нахождения суммы $\sum n_{uv}uv$ составляется расчетная табл. 21.

Таблица 21

$u \backslash v$	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:

$$\bar{x} = \bar{u}h_1 + C_1, \quad \bar{y} = \bar{v}h_2 + C_2, \quad S_x = S_uh_1, \quad S_y = S_vh_2. \quad (58)$$



§ 11. 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

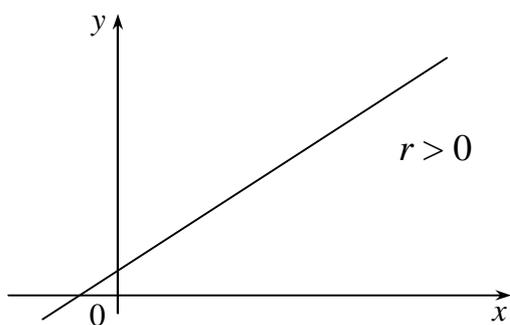


Рис. 6.

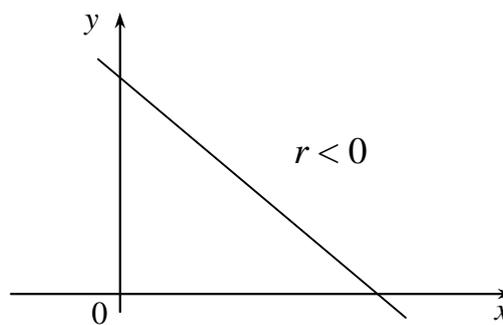


Рис. 7.

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 (рис. 6); например, вес и рост человека связаны положительной корреляцией; в случае $r < 0$ — об *отрицательной корреляции* (рис. 7). 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 [4], пользуясь табл. 22.

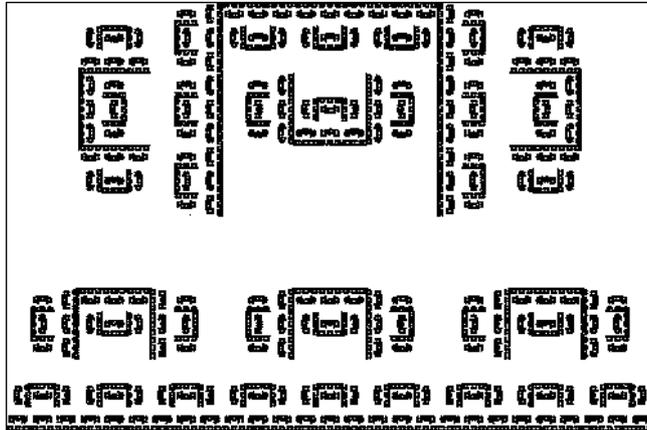
Таблица 22

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 б) по заданному уровню значимости α и числу степеней свободы $k = n - 2$ находят табличное значение $t_{кр}$ двусторонней критической области. Если $t_p < t_{кр}$, то коэффициент корреляции r — незначимый (мало отличается от нуля) и признаки X и Y некоррелированы. Если $t_p > t_{кр}$, то приходят к выводу о наличии линейной корреляционной связи [4].



§ 12. 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 σ_r . С помощью σ_r производят оценку надежности коэффициента корреляции, построив доверительные интервалы для различных объемов выборки [2]. Пусть число n наблюдений пар чисел $(x; y)$ меньше 50 ($n < 50$). В этом случае средняя квадратическая ошибка σ_r вычисляется по формуле

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

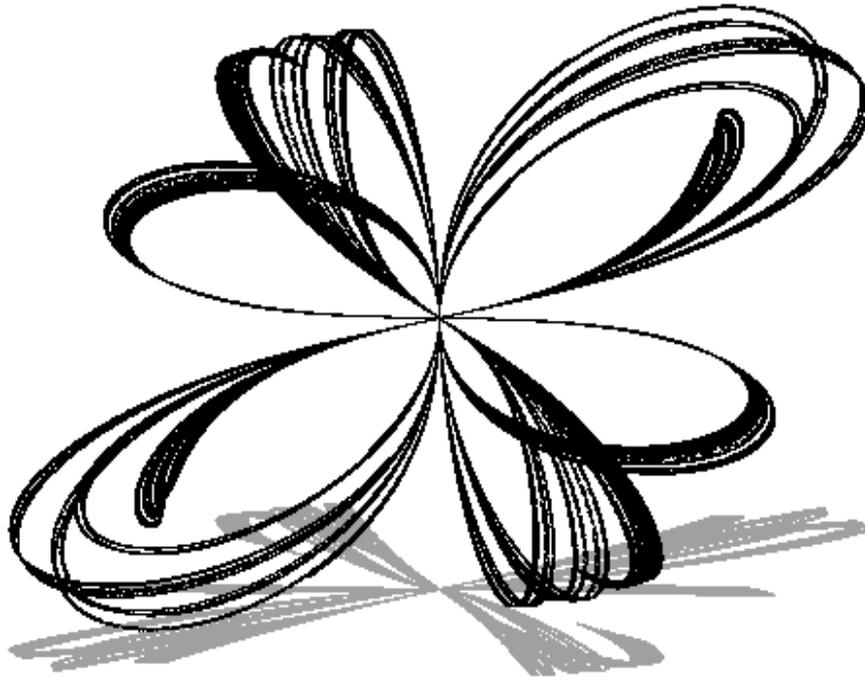
где r — коэффициент парной линейной корреляции, n — объем выборки. Доверительный интервал для оценки r находят по формуле

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

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

Если объем выборки $n > 50$, то погрешность σ_r для коэффициента корреляции r находят также по формуле (60). Затем вычисляют отношение r/σ_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 минимумом, а величина $r + 3\sigma_r$ — гарантийным максимумом коэффициента корреляции r и доверительный интервал для оценки r запишется в виде

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



§ 13. 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 [4].

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.



§ 14. Проверка адекватности модели

Для проверки соответствия уравнения регрессии $\hat{y}_x = a_0 + a_1x$ опыт-ным данным применяют критерий Фишера — Снедекора. Вычисляют [4] 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 дан-ными. Затем при заданном уровне значимости α и числах степеней свобо-ды $k_1 = 1$, $k_2 = n - 2$ находят по таблице критических точек распределения

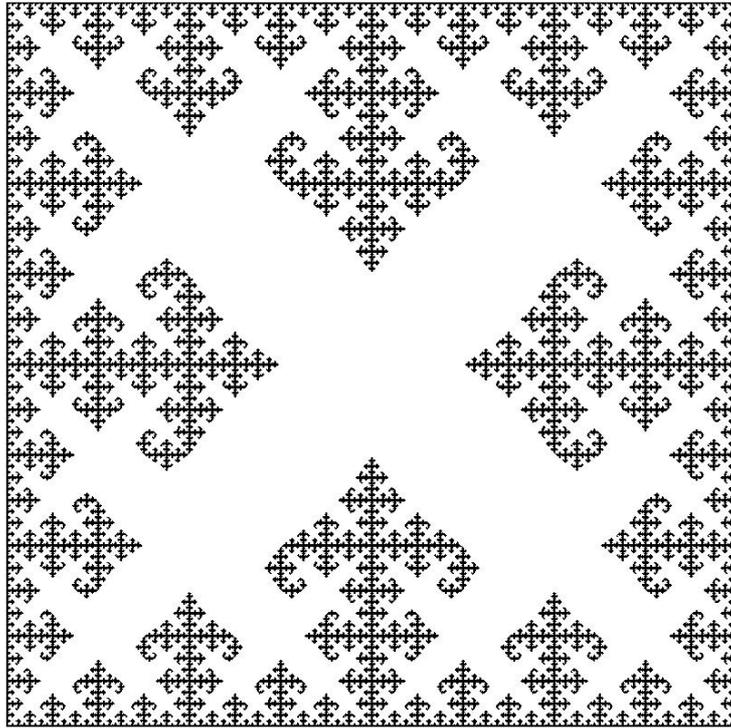
Фишера — Снедекора (приложение 7) $F_T = F_{\alpha; k_1; k_2}$. Если окажется, что $F_H > F_T$, то полученное уравнение линейной регрессии согласуется с опытными данными. Если $F_H < F_T$, то модель регрессии не согласуется с данными опыта. Формулой (64) пользуются тогда, когда исходные данные заданы 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 находят [4] остаточную сумму квадратов Q_e , характеризующую влияние неучтенных в модели факторов, по формуле

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

где $Q = \sum (y_i - \bar{y})^2$ — сумма квадратов отклонений значений y_i от средней \bar{y} , $Q_R = \sum (\bar{y}_{x_i} - \bar{y})^2$ — сумма квадратов отклонений условных средних \bar{y}_{x_i} от средней \bar{y} . Затем вычисляется статистика F_H по формуле

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

По таблице критических точек распределения Фишера – Снедекора (приложение 7) при заданном уровне значимости α 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.



§ 15. Оценка величины погрешности

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

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

где $\sigma_u = \sqrt{D_u}$, σ_u — стандартная ошибка уравнения регрессии,

$D_u = \frac{\sum (u_i - \bar{u})^2}{n-2}$ — остаточная дисперсия,

$u_i = y_i - \hat{y}_{x_i}$, y_i — опытные значения y ,

\hat{y}_{x_i} — значения y , полученные по уравнению регрессии,

$\bar{u} = \frac{1}{n} \sum (y_i - \hat{y}_{x_i})$ — среднее значение u_i ,

n — объем выборки.

Если величина δ 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:

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

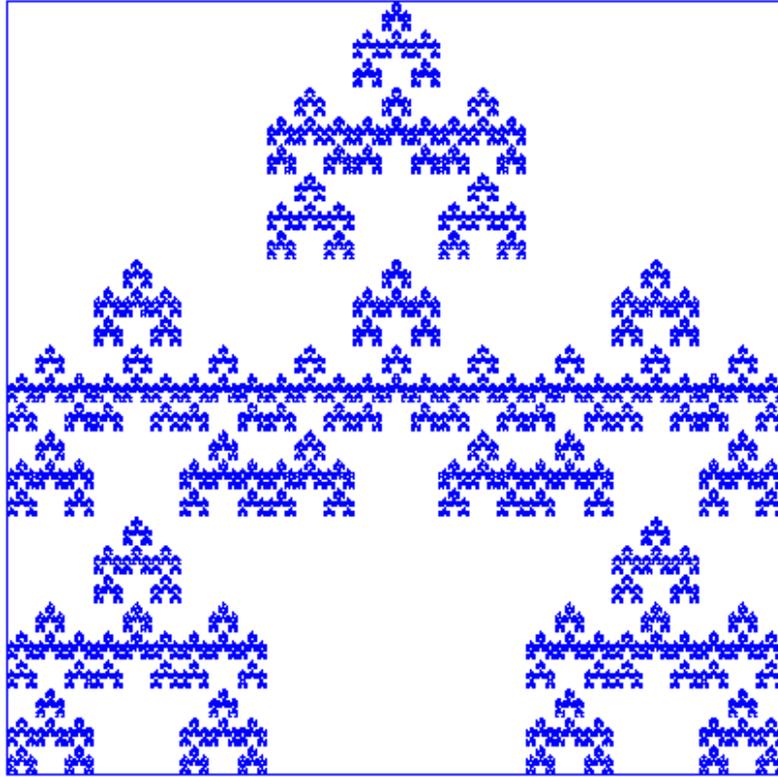
$$S_{a_1} = S_y \sqrt{1 - r^2} \cdot \sqrt{\frac{n}{n[x^2] - [x]^2}}. \quad (69)$$

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 [4].

Для более глубокого понимания теоретической части главы 3 настоятельно рекомендуем читателю выполнить лабораторные работы № 3 и № 4.

Контрольные вопросы:

1. Дать определение корреляционной зависимости между двумя признаками X и Y .
2. Дать определение условной средней признака Y и записать формулу для ее нахождения.
3. Сформулировать задачи, решаемые в теории корреляции.
4. Записать систему нормальных уравнений для нахождения параметров a_0 и a_1 уравнения линии регрессии $\hat{y}_x = a_0 + a_1x$ в случае, когда опытные данные не сгруппированы в корреляционную таблицу.
5. Записать уравнения регрессий y на x и x на y , используя коэффициент линейной корреляции r .
6. Дать определение коэффициента линейной корреляции, сформулировать его свойства.
7. Рассказать о том, как определяется теснота линейной корреляционной связи между двумя признаками с помощью коэффициента линейной корреляции.
8. Как определяется значимость коэффициента линейной корреляции?
9. Записать доверительные интервалы для оценки коэффициента линейной корреляции при различных объемах выборки.
10. Записать формулу для нахождения коэффициента детерминации в случае парной линейной корреляции и рассказать о его назначении.
11. Рассказать о проверке адекватности уравнения линейной регрессии y на x для случая несгруппированных опытных данных.
12. Рассказать о нахождении относительной погрешности линейного уравнения регрессии $\hat{y}_x = a_0 + a_1x$.
13. Как производится оценка коэффициентов a_0 и a_1 уравнения линейной регрессии $\hat{y}_x = a_0 + a_1x$?



§ 16. 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.

Содержание работы: 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 , σ_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.

Суть лабораторной работы отражает следующая задача.

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

Выполнение работы

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. 8 — 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. 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:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \cdot 353,6 = 35,36 \text{ — средняя производительность труда.}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{10} \cdot 103,2 = 10,32 \text{ — средняя зарплата сотрудников цеха техно-}$$

логический связи,

$$\hat{S}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{9} \cdot 640,064 = 71,118 \Rightarrow \hat{S}_x = 8,43,$$

$$\hat{S}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{9} \cdot 31,196 = 3,466 \Rightarrow \hat{S}_y = 1,87,$$

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

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

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

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

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 $t_T = t_{\alpha; k} = t_{0,05; 8} = 2,31$. Так как $t_p = 5,02 > t_T$, 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 s на числовой плоскости (рис. 8).

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):

$$r - t_g \cdot S_r \leq \hat{r} \leq r + t_g \cdot S_r.$$

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 σ_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:

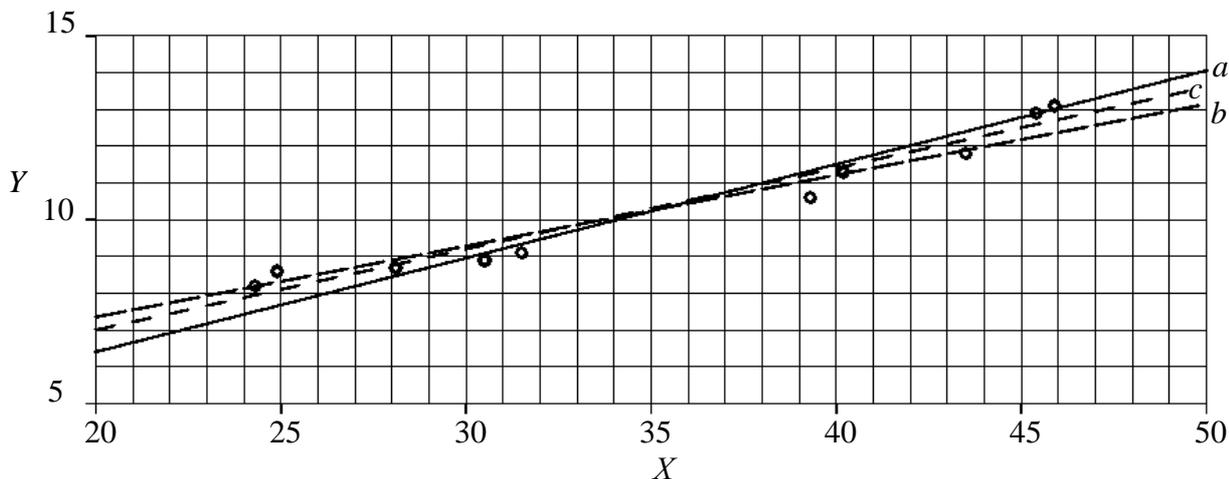


Рис. 8. Линейная регрессия Y на X — прямая a с уравнением $y = 0,193x + 3,5$ и линейная регрессия X на Y — прямая b с уравнением $x = 3,922y + 5,1$. Пунктированная прямая c проведена «от руки».

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

или $\hat{r} \in [0,71; 1]$. Следовательно, с вероятностью 0,95 линейный коэффициент корреляции генеральной совокупности находится в пределах от 0,71 до 1. 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 :

$$\hat{y}_x = \bar{y} + r \frac{\hat{s}_y}{\hat{s}_x} (x - \bar{x}) \Rightarrow \hat{y}_x = 10,32 + 0,87 \cdot \frac{1,87}{8,43} (x - 35,36) \Rightarrow$$

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

Уравнение регрессии x на y :

$$\hat{x}_y = \bar{x} + r \frac{\hat{s}_x}{\hat{s}_y} (y - \bar{y}) \Rightarrow \hat{x}_y = 35,36 + 0,87 \cdot \frac{8,43}{1,87} (y - 10,32) \Rightarrow$$

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

Контроль вычислений:

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

$$r^2 = 0,87 \cdot 0,87 = 0,7569.$$

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 не достаточно, 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):

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

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

$$R^2 = 1 - \frac{1,8729}{31,196} = 0,94, \quad F_H = \frac{0,94 \cdot (10-2)}{1-0,94} = 125,3.$$

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 (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

Т а б л и ц а 2 5

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
			1,8729

$\hat{y}_x = 0,192989x + 3,495898$ статистически значимо описывает результаты эксперимента.

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

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

где

$$\sigma_u = \sqrt{Du} = \sqrt{\frac{\sum (u_i - \bar{u})^2}{n-2}}, \quad u_i = y_i - \hat{y}_x, \quad \bar{u} = \frac{1}{n} \sum (y_i - \hat{y}_x)^2.$$

Так как $\sum (y_i - \hat{y}_x)^2 = 1,8729$, то $\bar{u} = 0,19$. Для нахождения суммы $\sum (u_i - \bar{u})^2$ составляем табл. 26.

Тогда $\sigma_u = \sqrt{\frac{2,2588}{8}} = 0,53$, $\delta = \frac{0,53}{10,32} \cdot 100\% = 5\%$. Так как величина δ мала, то уравнение линейной регрессии $\hat{y}_x = 0,192989x + 3,495898$ хорошо описывает опытные данные.

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):

Т а б л и ц а 2 6

$u_i = y_i - \hat{y}_x$	$u_i - \bar{u}$	$(u_i - \bar{u})^2$
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

$$S_{a_0} = S_{y/x} \cdot \sqrt{\frac{[x^2]}{n[x^2] - ([x])^2}}, S_{a_1} = S_{y/x} \sqrt{\frac{n}{n[x^2] - ([x])^2}}, S_{y/x} = \hat{S}_y \sqrt{1 - r^2}.$$

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

$$S_{y/x} = 1,87 \cdot \sqrt{1 - 0,7569} = 0,4931,$$

$$S_{a_0} = 0,4931 \cdot \sqrt{\frac{13143,36}{10 \cdot 13143,36 - (353,6)^2}} = 0,71,$$

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

Так как $S_{a_0}/|a_0| = \frac{0,71}{3,495898} < 0,5$ и $S_{a_1}/|a_1| = \frac{0,04}{0,192989} < 0,5$, то коэффициенты a_0 и a_1 уравнения регрессии y на x значимы. Графиками найденных регрессий являются прямые a , b , представленные на рис. 8.

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.

Варианты заданий к лабораторной работе № 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

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 от 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 электро-механического 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

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

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

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 зависимость диаметра вала от 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

Вариант № 8. При исследовании зависимости между выпуском готовой продукции Y (тыс. руб.) и энерговооруженностью труда X (кВт-час) 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

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

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

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

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

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

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

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

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

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

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/час 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

Вариант № 22. Найти формулу зависимости электрического сопротивления R (Ом) проводника от температуры Θ (°C) по следующим данным:

Θ	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

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

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

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

Вариант № 28. При исследовании зависимости между выработкой Y (тыс. руб.) и энерговооруженностью X (кВт / час) труда получены следующие данные:

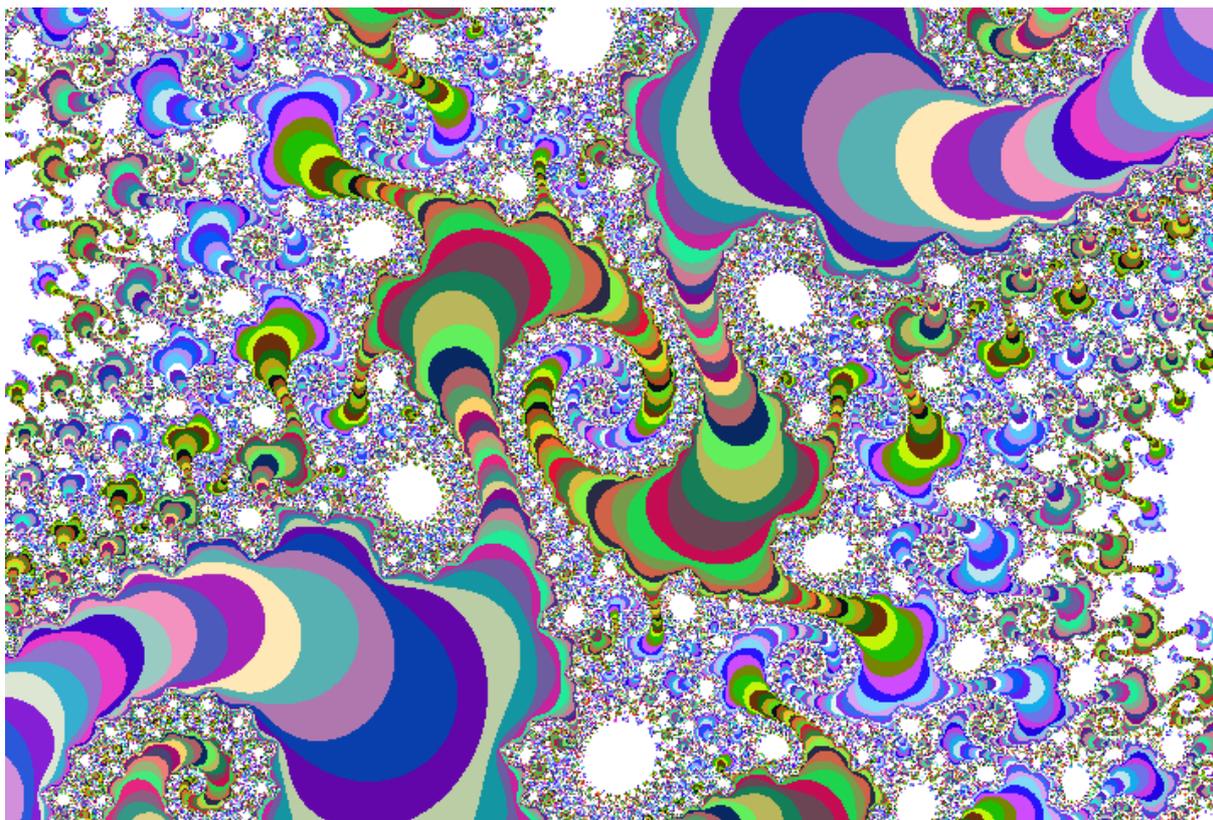
X	5	3	6	7	10	4	11	9	2	8
Y	6,3	3,6	7,5	8,5	12,5	6,2	12,6	10,7	2,6	7,7

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

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



§ 17. 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.

Содержание работы: 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 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.

З а д а ч а. 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

$y \backslash x$	-30	-20	-10	0	n_y
-8	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. 9).

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 признака X условное среднее признака Y по формуле (42):

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

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

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

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

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

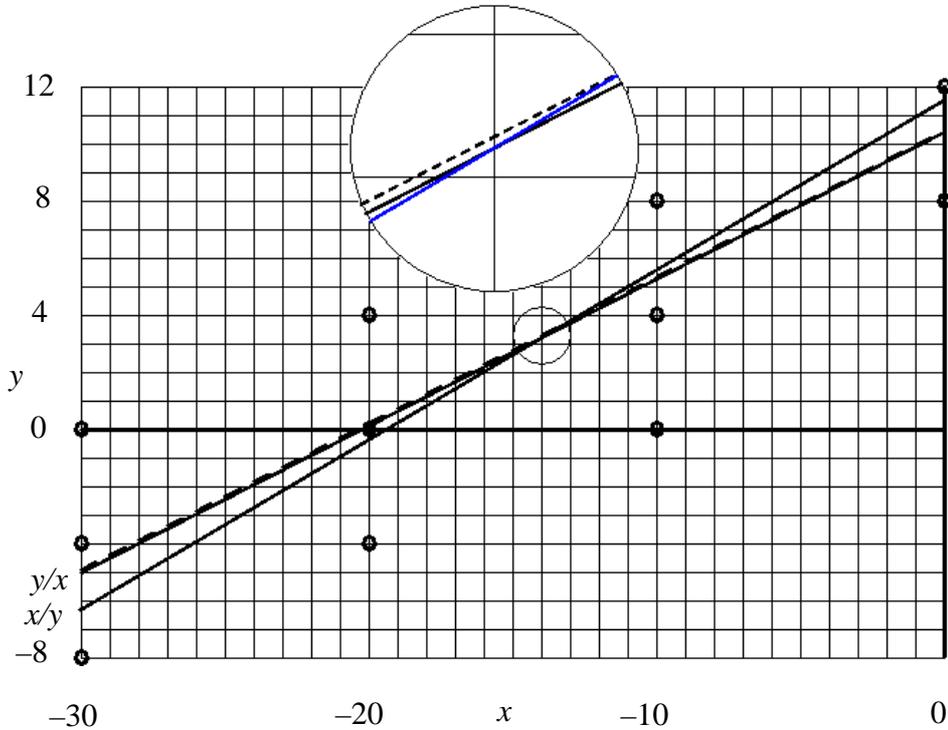


Рис. 9. Выявление линейной регрессии. «От руки» проводится линия тренда (штриховая прямая). Прямая (y/x) — регрессия y на x . Прямая (x/y) — регрессия x на y . 5-кратный «микроскоп» показывает картину в окрестности точки (\bar{x}, \bar{y}) .

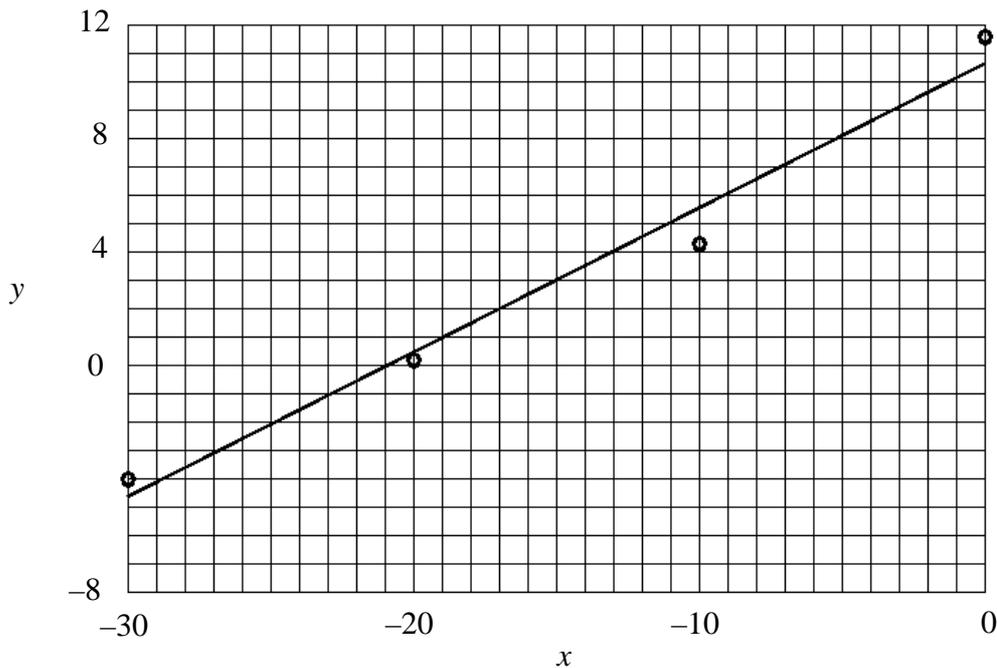


Рис. 10.

Строим точки с координатами $(x_j; \bar{y}_{x_j})$ (рис. 10).

Из рис. 10 видно, что отклонения точек от построенной прямой незначительны. Следовательно, связь между признаками X и Y может носить линейный характер. Составим уравнения линий регрессии 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 $\hat{y}_x = a_1x + a_0$. 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.

Т а б л и ц а 28

$y \backslash x$	-30	-20	-10	0	n_y	$n_y y$
-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:

$$\hat{y}_x = \bar{y} + r \frac{s_y}{s_x} (x - \bar{x}).$$

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 $C_1 = M_0X = -20$, $C_2 = M_0Y = 0$, $h_1 = 10$, $h_2 = 4$. Составляем корреляционную табл. 29 в условных вариантах.

Таблица 29

$\mathbf{v} \backslash u$	-1	0	1	2	n_v
-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:

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

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

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

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

Тогда

$$S_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1,24 - 0,36} = 0,94,$$

$$S_{\mathbf{v}} = \sqrt{\overline{\mathbf{v}^2} - (\bar{\mathbf{v}})^2} = \sqrt{2,34 - 0,64} = 1,3.$$

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

Таблица 30

$\mathbf{v} \backslash u$	-1	0	1	2	
-2	1	2			2
-1	4	1			4
0					
1			13	1	13
2			2	2	4
3				9	6
	6		17	58	81

Тогда, согласно (56), (58),

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

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

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

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

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

Отсюда и из (47), (48) следуют 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 :

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

или

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

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

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

При уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = n - 2 = 50 - 2 = 48$ находим по таблице распределения Стьюдента $t_{\alpha;k} = t_{0,05;48} = 2,01$. Так как $t_H = 17,5 > 2,01$, то выборочный коэффициент линейной корреляции r_g 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 y

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. 9.

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
		$Q = 288,64$

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

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

При уровне значимости $\alpha = 0,05$ и числах степеней свободы $k_1 = 1$, $k_2 = n - 2 = 50 - 2 = 48$ 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, связан-

ных с построением моделей парной линейной корреляции 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 линейных регрессий следует выбирать лучшую из них.

Варианты заданий по лабораторной работе № 4

Вариант № 1. Зависимость объема Y (см³) разрушенной породы от глубины X (мм) внедрения зуба при постоянном давлении приведены в корреляционной таблице:

$Y \backslash X$	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:

$Y \backslash X$	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:

$Y \backslash X$	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 замерах забойных давлений X глубинным манометром. Данные замеров приведены в корреляционной таблице:

$Y \backslash X$	125	135	145	155	165	175	185	195	205	n_y
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/мм²) инструментальных быстрорежущих сталей задана корреляционной таблицей:

$Y \backslash X$	0,7	0,9	1,1	1,3	1,5	n_y
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 средне-месячной выработкой продукции на одного рабочего Y (тыс. руб.) и стоимостью основных производственных средств X (млн. руб.) приведены в корреляционной таблице:

$Y \backslash X$	0,8	0,9	1,0	1,1	1,2	1,3	n_y
9,9	2						2
10	2	7				1	10
10,1		2	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:

$Y \backslash X$	1,05	1,15	1,25	1,35	1,45	1,55	1,65	n_y
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 ($^{\circ}\text{C}$) смазочного масла в двигателе и температуры Y ($^{\circ}\text{C}$) масла в коробке передач автомобиля приведены в корреляционной таблице:

$Y \backslash X$	6	10	14	18	22	26	n_y
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 (час) приведены в корреляционной таблице:

$Y \backslash X$	20,5	25,5	30,5	35,5	40,5	45,5	n_y
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 (час) непрерывной работы и количества Y (шт.) полностью обработанных деталей приведены в корреляционной таблице:

$Y \backslash X$	6	10	14	18	22	26	n_y
0,95	5	5					10
1,85		4	2				6
2,75			6	8			14
3,65				1	6		7
4,55				5	3	5	13
n_x	5	9	8	14	9	5	50

Вариант № 11. Результаты измерений температуры смазочного масла X ($^{\circ}\text{C}$) в коробке передач и скорости Y (км/час) автомобиля приведены в корреляционной таблице:

$Y \backslash X$	30,5	40,5	50,5	60,5	70,5	80,5	n_y
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/час) прохождения пути и температурой X ($^{\circ}\text{C}$) смазочного масла в коробке передач у 100 автомобилей:

$Y \backslash X$	20	25	30	35	40	45	n_y
20	3	3					6
30		5	4				9
40			12	30			42
50			5	10	14		29
60					5	9	14
n_x		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:

$Y \backslash X$	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:

$Y \backslash X$	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/час) бурения в твердых породах от нагрузки X (атм.) на долото:

$Y \backslash X$	0	0,5	1,0	1,5	2,0	2,5	3,0	3,5	n_y
10							2	1	3
10,5					1	1	2		4
11				3	2	2			7
11,5			5	7	4				16
12		3	2	8					13
12,5	2	1	1						4
13	1	1							2
13,5	1	0							1
n_x	4	5	8	18	7	6	4	1	50

Вариант № 16. Результаты исследования зависимости длительности T (час) непрерывной работы двигателей и расхода V (литров) топлива заданы корреляционной таблицей:

$T \backslash V$	20	50	80	110	140	170	200	n_T
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 (м³/час) скважины приведены в корреляционной таблице:

$Y \backslash X$	20	22	24	26	28	30	n_y
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 Тюменской области:

$Y \backslash X$	1	3	5	7	9	11	13	15	17	n_y
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.

$Y \backslash X$	15	20	25	30	35	40	n_y
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:

$Y \backslash X$	25	28	31	34	37	n_y
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	1	2
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:

$Y \backslash X$	1,0	1,2	1,4	1,6	1,8	2	n_y
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:

$Y \backslash X$	0,05	0,15	0,25	0,35	0,45	0,55	0,65	n_y
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/сек) при ударно-вращательном бурении от коэффициента пластичности долот X заданы в корреляционной таблице:

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

Вариант № 24. Результаты исследования зависимости выпуска валовой продукции Y (%) в отчетном году и выработкой на одного рабочего X (%) приведены в корреляционной таблице:

$Y \backslash X$	75	85	95	105	115	125	n_y
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:

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

Вариант № 26. Распределение 100 автомобилей по температуре смазочного масла в двигателе X и по температуре масла в коробке передач Y даны в корреляционной таблице:

$Y \backslash X$	5	10	15	20	25	30	n_y
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 (час) фрез приведены в корреляционной таблице:

$Y \backslash X$	30	34	38	42	46	50	54	n_y
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:

$Y \backslash X$	150	200	250	300	350	400	450	500	550	n_y
5	2	1								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	5	7	15	20	24	17	7	3	2	100

Вариант № 29. Имеют данные распределения 100 автомобилей по температуре Y смазочного масла в двигателе и по скорости X движения (км/час):

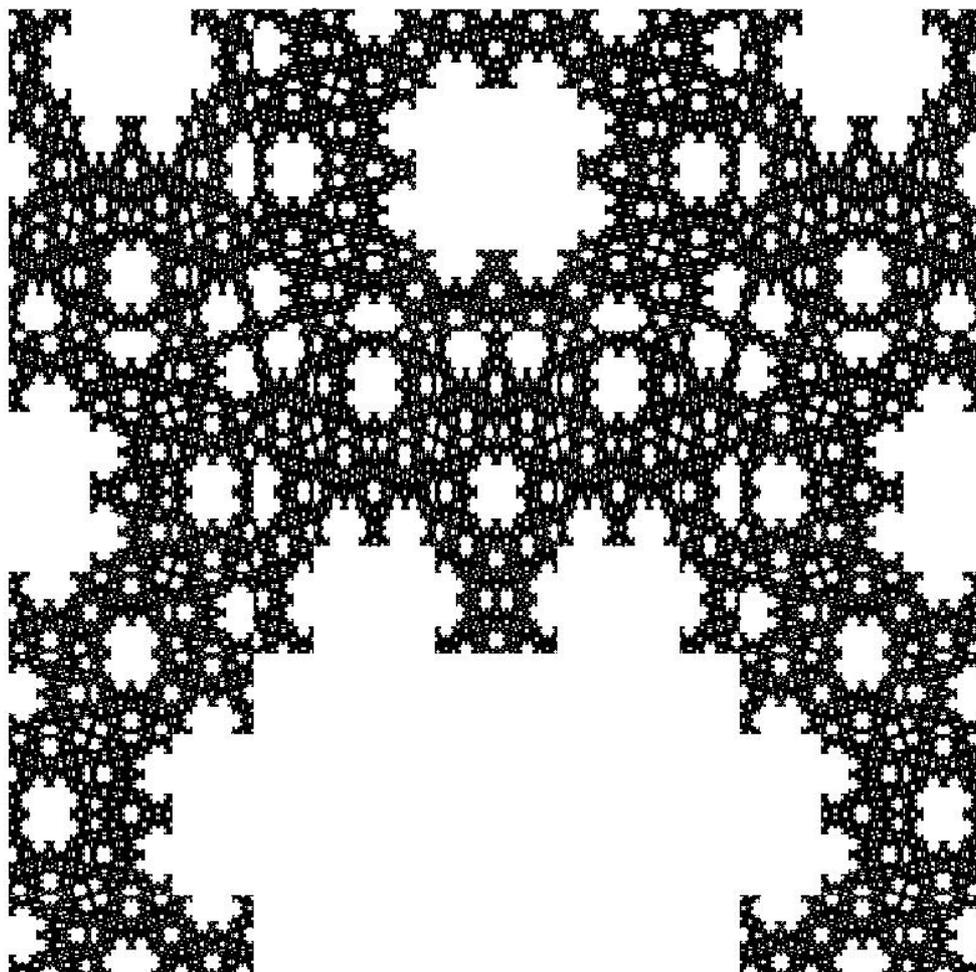
$Y \backslash X$	10	15	20	25	30	35	n_y
20	1	5					6
30		6	4				10
40			7	40	3		50
50			2	10	8		20
60				5	6	3	14
n_x	1	11	13	55	17	3	100

Вариант № 30. Результаты зависимости между дебитом Q (м³/час) скважины и диаметром штуцера D (мм) приведены в корреляционной таблице:

$Q \backslash D$	18	21	24	27	30	33	36	n_Q
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

ГЛАВА 4

ПАРНАЯ НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ



§ 18. Нелинейная корреляционная зависимость

Между изучаемыми признаками X и Y может существовать нелинейная корреляционная зависимость. Если линия регрессии с уравнением $\hat{y}_x = f(x)$ или $\hat{x}_y = j(y)$ 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 [4] в виде

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

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 нормальных уравнений

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

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:

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

где

$$[1/x] = \sum_{i=1}^n 1/x_i, [1/x^2] = \sum_{i=1}^n 1/x_i^2, [y] = \sum_{i=1}^n y_i, [y/x] = \sum_{i=1}^n y_i/x_i.$$

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 нормальных уравнений

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

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.

Проверка необходимых условий

Проверку необходимых условий для выбора одной из предполагаемых нелинейных зависимостей проводят [13], пользуясь табл. 33.

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_i, 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

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
$y(\sqrt{x_1 x_n}) = \sqrt{y(x_1)y(x_n)}$	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 $y = ab^x$	$Y = \ln y, X = x,$ $A = \ln a, B = \ln b$
$y\left(\frac{2x_1 x_n}{x_1+x_n}\right) = \frac{y(x_1)+y(x_n)}{2}$	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 $y = \frac{1}{ax+b}$	$Y = 1/y, X = x$
$y(\sqrt{x_1 x_n}) = \frac{y(x_1)+y(x_n)}{2}$	логарифмическая $y = a \ln x + b$	$Y = y, X = \ln x$

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

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 Δ правой части от левой необходимого условия. Вычисленные отклонения Δ_i сравнивают и по наименьшему из них выбирают окончательно одну из формул.

Метод конечных разностей

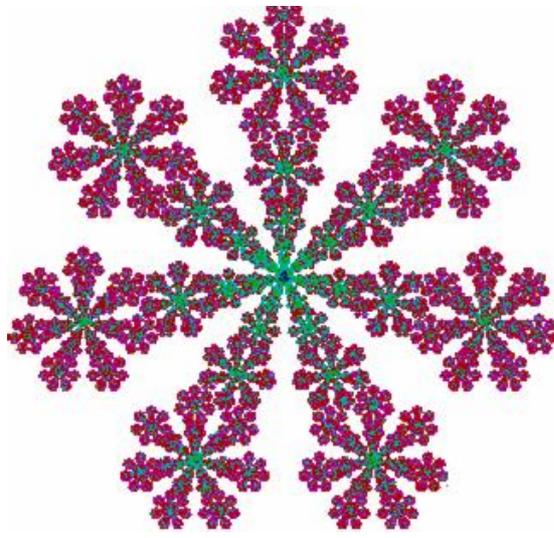
Пусть в корреляционном поле могут быть проведены линии, описываемые уравнениями $y = ax^b$, $y = ae^{cx}$ ($c = \ln b$), $y = a + \frac{b}{x}$, $y = \frac{1}{ax+b}$, $y = a \ln x + b$. Все эти формулы содержат по два параметра и могут быть приведены к формуле $Y = AX + B$ посредством табл. 33. Так как все зависимости, приведенные в табл. 33, сводятся к линейной, то для обоснования выбора формулы $Y = AX + B$ вычисляют конечные разности первого порядка ΔX , ΔY 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.

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.

Таблица 34

Таблица конечных разностей

x	y	Δy	$\Delta^2 y$
x_0	<u>y_0</u>		
		<u>Δy_0</u>	
x_1	y_1		<u>$\Delta^2 y_0$</u>
		Δy_1	
x_2	y_2		$\Delta^2 y_1$
		Δy_2	
x_3	y_3		$\Delta^2 y_2$
		Δy_3	
x_4	y_4		$\Delta^2 y_3$
		Δy_4	
x_5	y_5		
...



§ 19. Определение силы криволинейной СВЯЗИ

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

Корреляционным отношением η_{yx} называется величина, определяемая равенством:

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

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}_{xi} — условная средняя признака Y .

Аналогично определяется выборочное корреляционное отношение η_{xy} :

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

Корреляционные отношения обладают следующими свойствами (сформулируем свойства для η_{yx} , так как для η_{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 η_{yx} к 1, тем сильнее корреляционная связь между признаками X и Y , а чем ближе η_{yx} к 0, тем слабее эта зависимость.

5. Если $\eta_{yx} = |r|$ регрессия y на x является линейной.

6. Коэффициент линейной корреляции r не превосходит по модулю η_{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 [4] по формуле:

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

где $S_{yx}^2 = (\sum (y_j - \hat{y}_{x_j})^2) / n$ — средний квадрат отклонений фактических значений y от значений 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

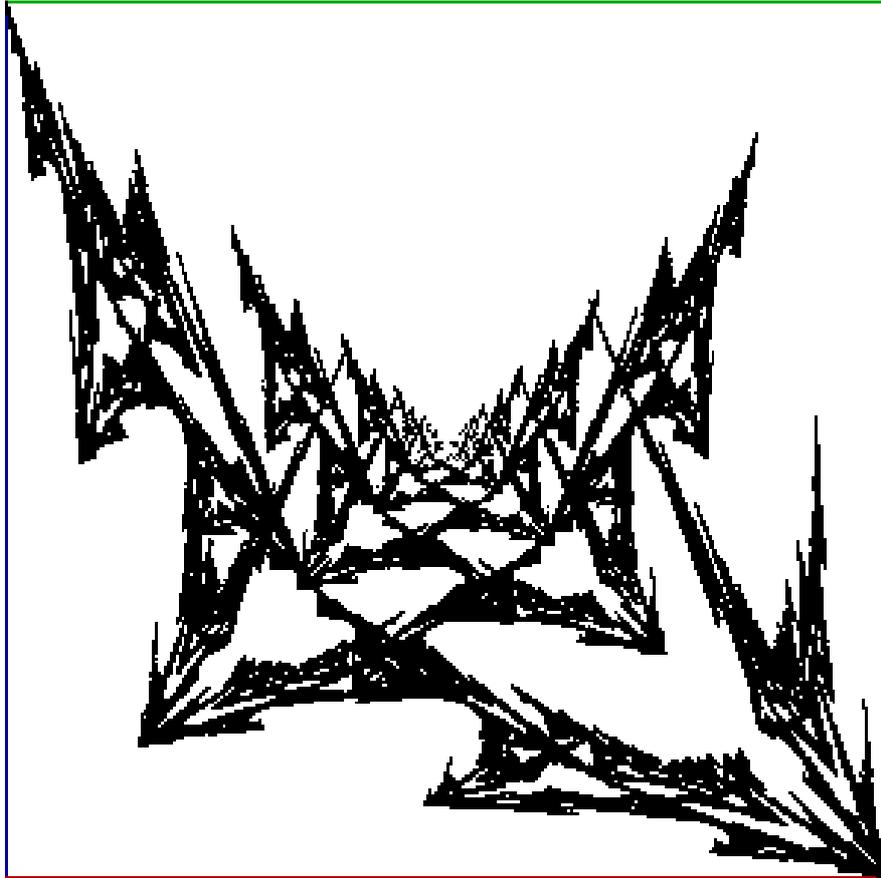
$$i = \sqrt{\frac{\sigma_y^2}{S_y^2}}, \quad (86)$$

где

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

$$S_y^2 = (\sum (y_j - \bar{y})^2 n_j) / n, \quad \sigma_y^2 = \sum (\bar{y}_{x_i} - \bar{y})^2 \cdot n_i / n.$$

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.



§ 20. Проверка адекватности модели

Проверить адекватность модели — значит установить, соответствует ли построенное уравнение регрессии опытным данным и достаточно ли включенных в уравнение факторных признаков X_i для описания резуль- тативного признака. Оценка значимости уравнения регрессии производится на основе дисперсионного анализа. В случае нелинейной парной корреля- ции находят [4] 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 регрес- сии, то находят статистику

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

Затем при заданном уровне значимости α 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.

Контрольные вопросы:

1. Записать систему нормальных уравнений для нахождения коэффициентов a_0 , a_1 , a_2 уравнения регрессии $\hat{y}_x = a_0 + a_1x + a_2x^2$ в случае не сгруппированных опытных данных.

2. Записать систему нормальных уравнений для нахождения коэффициентов a_0 , a_1 , a_2 уравнения регрессии $\hat{y}_x = a_0 + a_1x + a_2x^2$ в случае сгруппированных опытных данных.

3. Записать системы нормальных уравнений для нахождения коэффициентов a_0 и a_1 уравнений регрессий $\hat{y}_x = a_0 + \frac{a_1}{x}$ и $\hat{y}_x = \frac{1}{a_0 + a_1x}$.

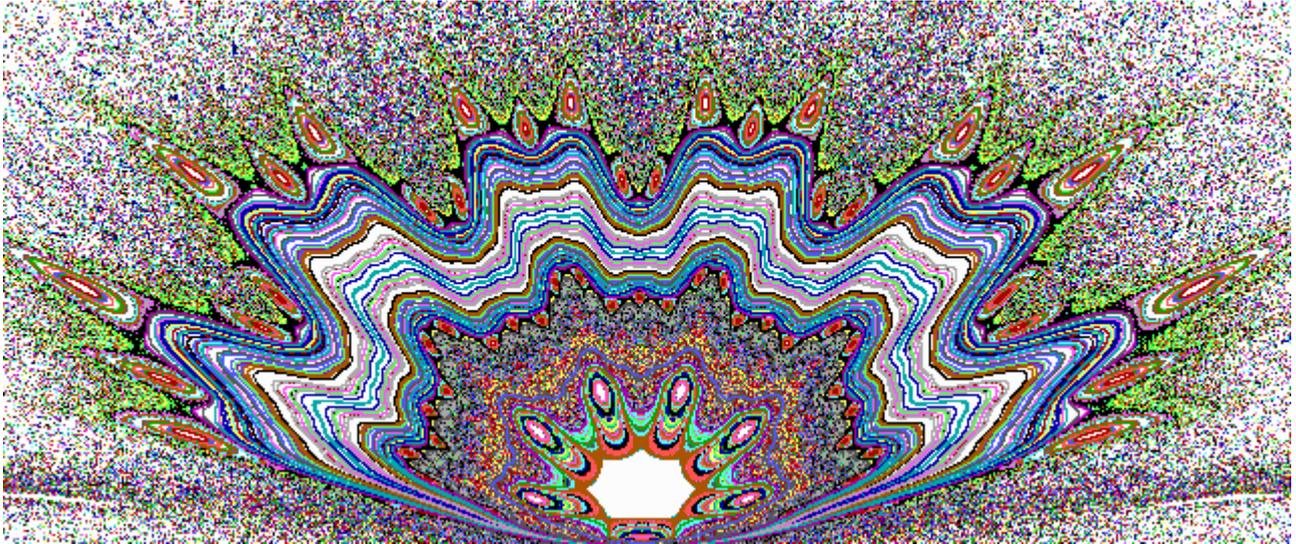
4. Рассказать о применении необходимых условий выбора одной из предполагаемых нелинейных зависимостей.

5. Записать формулу нахождения значения y методом линейного интерполирования для значения x , отсутствующего в таблице опытных данных.

6. Рассказать о применении метода конечных разностей для выбора одной из предполагаемых нелинейных зависимостей.

7. Рассказать об установлении тесноты связи между признаками в случае нелинейной зависимости с помощью корреляционного отношения и индекса корреляции.

8. Как осуществляется проверка адекватности нелинейной корреляционной модели?



§ 21. 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.

Содержание работы: 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^2) 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

Выполнение работы

В системе координат Oxy строим корреляционное поле (рис. 11).

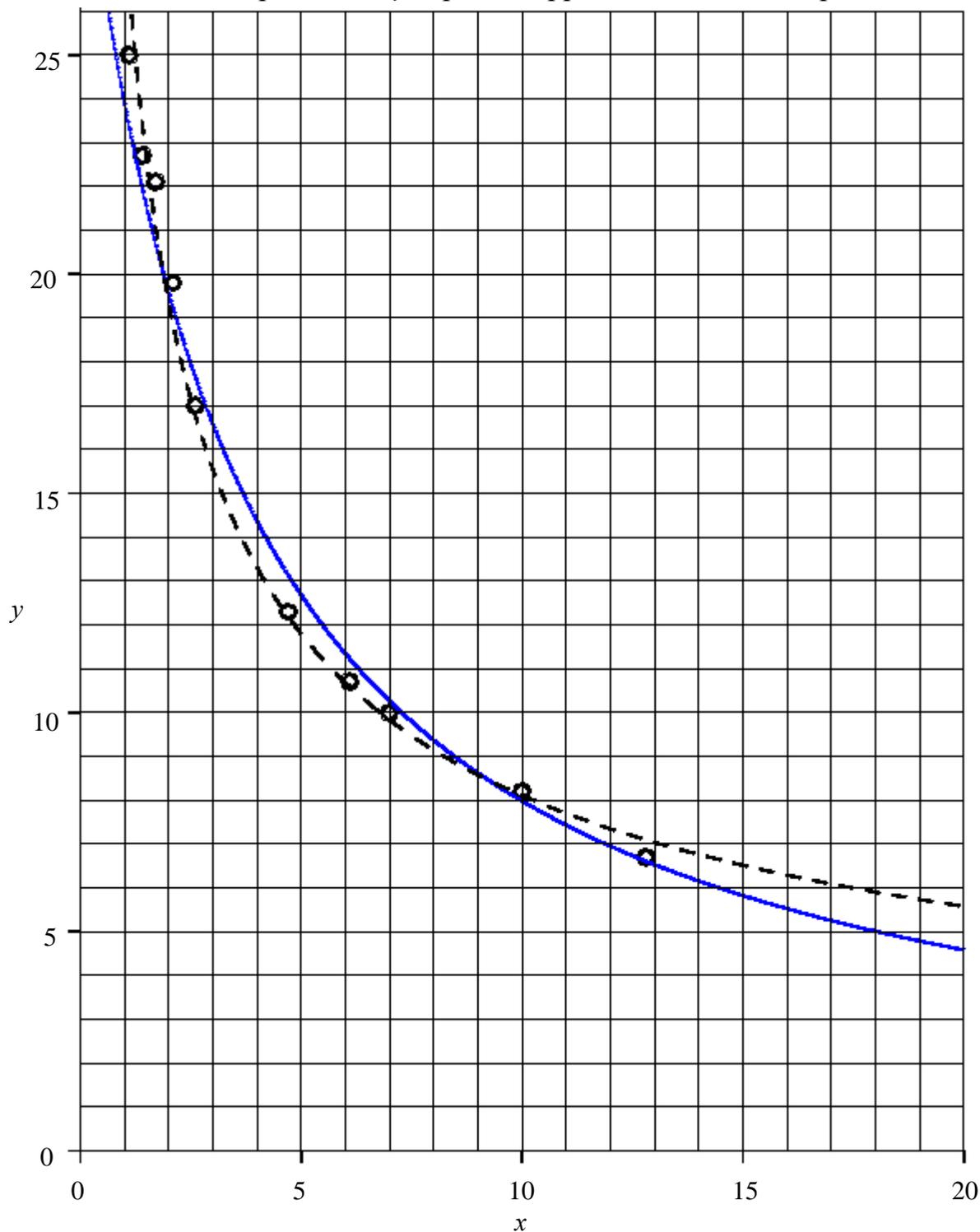


Рис. 11. Десять точек корреляционного поля позволяют предположить, что искомая функция регрессии является степенной. Сплошная линия — график функции $y = 1/(0,00928x + 0,0325)$, а штриховая — график функции $y = 28,1x^{-0,54}$.

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 проверим выполнение равенства $y\left(\frac{2x_1x_n}{x_1+x_n}\right) = \frac{y(x_1)+y(x_n)}{2}$.

$$y\left(\frac{2x_1x_n}{x_1+x_n}\right) = y\left(\frac{2x_1x_{10}}{x_1+x_{10}}\right) = y\left(\frac{2 \cdot 1,1 \cdot 12,8}{1,1+12,8}\right) = y(2,02).$$

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

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

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

Вычисляем отклонение Δ_1 : $\Delta_1 = |15,85 - 20,26| = 4,41$. Для формулы $y = \frac{1}{ax+b}$ проверяем выполнение равенства $y\left(\frac{x_1+x_n}{2}\right) = \frac{2y(x_1) \cdot y(x_n)}{y(x_1)+y(x_n)}$.

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

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

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

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

Вычисляем отклонение Δ_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 выше рассматриваемых 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 $\Delta Y / \Delta X$. Составляем расчетную табл. 36.

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 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
ΔY	4,28	5,79	4,01	2,62	13,61	7,46	4,73	12	3,76	
ΔX	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	

Отношения $\frac{\Delta Y}{\Delta X}$, полученные для формулы $y = \frac{1}{ax+b}$, мало отличаются друг от друга, чем для формулы $y = a + \frac{b}{x}$. 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 $\hat{y}_x = \frac{1}{ax+b}$. Оценки a и b неизвестных параметров истинного уравнения регрессии находим, решая систему нормальных уравнений (78):

$$\begin{cases} na_0 + [x]a_1 = [1/y], \\ [x]a_0 + [x^2]a_1 = [x/y]. \end{cases}$$

Таблица 37

$X = x$	1,1	1,4	1,7	2,1	2,6	4,7	6,1	7	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
ΔY	0,004	0,001	0,005	0,008	0,022	0,012	0,007	0,022	0,027	
ΔX	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	

Для вычисления сумм, входящих в систему, составляем расчетную табл. 38.

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
49,5		0,784595	5,223764	389,37

Решением системы

$$\begin{cases} 10a_0 + 49,5a_1 = 0,784594, \\ 49,5a_0 + 389,37a_1 = 5,223761 \end{cases}$$

является точка $(a_0, a_1) = (0,03251; 0,0092835)$. Поэтому уравнение регрессии примет вид:

$$\hat{y}_x = \frac{1}{0,00928x+0,0325}.$$

Оценим силу корреляционной связи между скоростью резания и площадью поперечного сечения стружки хромоникелевой стали. Вычислим индекс корреляции по формуле (84):

$$i = \sqrt{1 - \frac{\hat{S}_{yx}^2}{\hat{S}_y^2}},$$

где

$$\hat{S}_{yx}^2 = \frac{1}{n-1} \sum (y_i - \hat{y}_{xi})^2, \quad \hat{S}_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

(так как $n=10 < 50$). Для нахождения \hat{S}_{yx}^2 и \hat{S}_y^2 составляем расчетную табл. 39.

Т а б л и ц а 39

x_i	y_i	\hat{y}_x	$(y_i - \hat{y}_{xi})^2$	$(y_i - \bar{y})^2$
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
7,0	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 $i = \sqrt{1 - 11,14333/44,45139} = 0,86$. 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

$$F_T = F_{\alpha; k_1; k_2} = F_{0,05; 1; 8} = 5,32.$$

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 $\hat{y}_x = \frac{1}{0,009x+0,04282}$.

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. $y = k x^{-D}$, $D > 0$. Рекомендуем читателю самостоятельно убедиться, что в этом случае $k \approx 28,1122$, $D \approx 0,5399$. На рис. 11 для сравнения показаны гипербола (сплошная линия) и график функции $y = 28,1x^{-0,54}$.

Варианты заданий по лабораторной работе № 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 Пластмасс 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 (мм²) 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 N (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 Тюменском моторостроительном объединении 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 / сек) 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/см²) бетона при испытании цилиндрических образцов в зависимости от отношения $X = h / \alpha$ высоты h к диаметру α 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 «Сибком-плектмонтаж» 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 ΔP (атм) забойных давлений глубинным манометром. Результаты исследований приведены в таблице:

Q	5	15	25	35	45	55
ΔP	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 Тюменской области:

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·час) 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 (см³) зоны разрушения от предела P_0 (кг/мм²) 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 ΔP (кг/см²) 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
ΔP	95	73	52	45	35	33	31

Вариант № 28. Зависимость среднемесячной заработной платы Y (тыс. руб.) рабочих нефтеперерабатывающего 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:

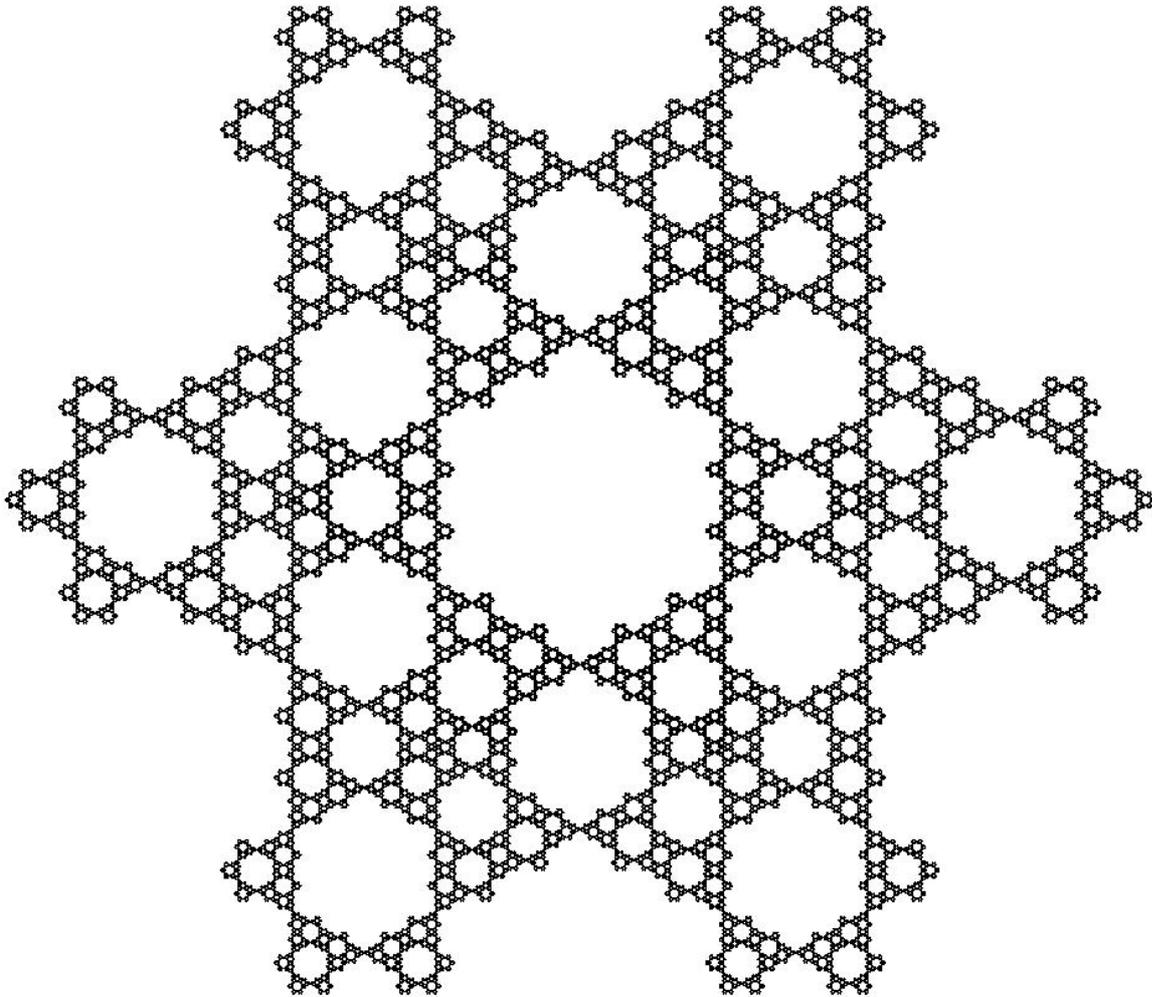
X	0,5	1,5	2,5	3,5	4,5	6,0	9	10
Y	14	11	10	8	6,5	5	4,5	4

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

ГЛАВА 5

МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ



§ 22. Понятие множественной корреляции

Производственные взаимосвязи, как правило, определяются большим числом одновременно и совокупно действующих факторов. Например, овальность после чистового шлифования зависит от припуска на чистовое шлифование и от овальности после предварительного шлифования. Себестоимость продукции зависит от стоимости материала, основной заработной платы рабочих, премиальных, расходов на содержание оборудования, отчислений на соцстрахование. В связи с этим возникает задача исследования зависимости между признаками X_1, X_2, \dots, X_n , называемыми *фак-*

торными признаками (а также регрессорами, предикторами) и признаком Y , называемым *результативным*. Эта задача решается методами так называемого *множественного корреляционного анализа*. Его характерной особенностью является то, что для измерения совместного влияния ряда показателей — факторов на величину анализируемого показателя строятся модели множественной корреляции.

Построение многофакторной корреляционной модели начинается с установления формы связи, используя графический метод для пространства R^2 и метод перебора различных уравнений. От правильности выбора вида уравнения зависит, насколько построенная модель будет адекватна не только имеющимся экспериментальным данным, но и истинной зависимости между изучаемыми показателями. При прочих равных условиях предпочтение отдается модели, зависящей от меньшего числа параметров, так как для их оценки требуется меньшее количество эмпирических данных.

После выбора формы многофакторной корреляционной модели проводят отбор факторных признаков и включение их в модель. Принято считать, что в уравнение множественной регрессии можно включать только независимые друг от друга факторные признаки X_i . Практически факторные признаки зависят либо слабо, либо сильно. Поэтому вопрос о включении факторных признаков в уравнение регрессии решают следующим образом. Пусть, например, имеется три факторных признака X_1, X_2, X_3 , влияющих на результативный признак Y и модель является линейной. Чтобы выяснить, какие факторные признаки включить в модель, находят коэффициенты парной корреляции $r_{X_1X_2}, r_{X_1X_3}, r_{X_2X_3}$. Если их значения меньше 0,8, то их можно включить в модель. Если же их значение больше 0,8, то следует какие-то из этих факторов исключить из модели. Пусть, например, $r_{X_1X_3} = 0,85$. Ясно, что какой-то из признаков X_1 или X_3 надо исключить из модели. Для этого находят парные коэффициенты корреляции между каждым из факторов X_1 и X_3 и результативным признаком Y , то есть вычисляют r_{YX_1} и r_{YX_3} . Затем сравнивают r_{YX_1} и r_{YX_3} . Пусть оказалось, что $r_{YX_3} > r_{YX_1}$. Это означает, что факторный признак X_3 сильнее связан с результативным признаком Y , чем признак X_1 . Поэтому фактор X_3 следует включить в модель, а X_1 исключить из нее. Этот вывод подтверждаем путем вычисления коэффициентов частной корреляции $r_{YX_3(X_1)}$ и $r_{YX_1(X_3)}$. При исключении факторов из модели можно руководствоваться правилом. Если $|r_{ij}| > 1 - 3\sigma_{r_{ij}}$, где

$$S_{r_{ij}} = \frac{1 - r_{ij}^2}{\sqrt{n-1}}, \quad (89)$$

Для того чтобы проще было запомнить систему (96), запишем ее в матричной форме, предварительно введя следующие условные обозначения (только для этого конкретного случая):

$$[X_0] = n, [X_r] = \sum_{i=1}^n x_{ri}, [Y] = \sum_{i=1}^n y_i, [X_r Y] = \sum_{i=1}^n x_{ri} y_i, r, s = 1 \dots k,$$

$$[X_0][X_0] = [X_0^2] = n, \quad [X_0][X_s] = [X_s], \quad [X_r][X_s] = [X_r X_s],$$

$$[X_0][Y] = [Y], \quad [X_r][Y] = [X_r Y].$$

Рассмотрим вектор

$$X = \begin{pmatrix} [X_0] \\ [X_1] \\ \mathbf{M} \\ [X_k] \end{pmatrix} = ([X_0] \ [X_1] \ \mathbf{L} \ [X_k])^T.$$

Тогда (96) примет вид:

$$XX^T A = XY, \quad (97)$$

или, в развернутой форме,

$$\begin{pmatrix} [X_0] \\ [X_1] \\ \mathbf{M} \\ [X_k] \end{pmatrix} ([X_0] \ [X_1] \ \mathbf{L} \ [X_k]) \begin{pmatrix} a_0 \\ a_1 \\ \mathbf{M} \\ a_k \end{pmatrix} = \begin{pmatrix} [X_0] \\ [X_1] \\ \mathbf{M} \\ [X_k] \end{pmatrix} [Y],$$

или

$$\begin{pmatrix} [X_0 X_0] & [X_0 X_1] & \mathbf{L} & [X_0 X_k] \\ [X_1 X_0] & [X_1 X_1] & \mathbf{L} & [X_1 X_k] \\ \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} \\ [X_k X_0] & [X_k X_1] & \mathbf{L} & [X_k X_k] \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \mathbf{M} \\ a_k \end{pmatrix} = \begin{pmatrix} [X_0 Y] \\ [X_1 Y] \\ \mathbf{M} \\ [X_k Y] \end{pmatrix},$$

или, наконец,

$$\begin{pmatrix} n & [X_1] & \mathbf{L} & [X_k] \\ [X_1] & [X_1^2] & \mathbf{L} & [X_1 X_k] \\ \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} \\ [X_k] & [X_k X_1] & \mathbf{L} & [X_k^2] \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \mathbf{M} \\ a_k \end{pmatrix} = \begin{pmatrix} [Y] \\ [X_1 Y] \\ \mathbf{M} \\ [X_k Y] \end{pmatrix}.$$

Для оценки надежности коэффициентов a_0, a_1, \dots, a_k находят средние квадратические ошибки этих коэффициентов, т.е. S_{a_r} , $r = 0..k$. Пусть

$C = XX^T$ — матрица коэффициентов при переменных a_r системы (97). Эта матрица является симметрической, т.е. $C^T = C$, но, — существенно для дальнейшего, — она не является вырожденной: $\det C = k + 1$. Тогда средние квадратические ошибки коэффициентов a_i вычисляются по формуле:

$$S_{a_r} = \hat{S}_e \sqrt{c'_{rr}}, \quad (98)$$

где $\hat{S}_e = \sqrt{\sum (Y - \hat{Y}_X)^2}$, c'_{rr} — диагональный элемент матрицы C^{-1} , соответствующий факторному признаку X_r .

Пр и м е р. По опытным данным найдена матрица

$$(XX^T)^{-1} = \begin{pmatrix} 70,31 & -1,65 & 0,56 & 0,21 \\ -1,65 & 0,08 & -0,03 & -0,02 \\ 0,56 & -0,03 & 0,02 & 0,01 \\ 0,21 & -0,02 & 0,01 & 0,003 \end{pmatrix},$$

величина $\hat{S}_e = 3,08$ и уравнение регрессии

$$\hat{Y}_X = 59,5 - 0,81X_1 - 0,29X_2 + 0,13X_3.$$

Требуется вычислить средние квадратические ошибки коэффициентов уравнения регрессии.

Для решения воспользуемся формулой (98). Учитывая условие задачи, находим:

$$c'_{00} = 70,31, \quad c'_{11} = 0,08, \quad c'_{22} = 0,02, \quad c'_{33} = 0,003.$$

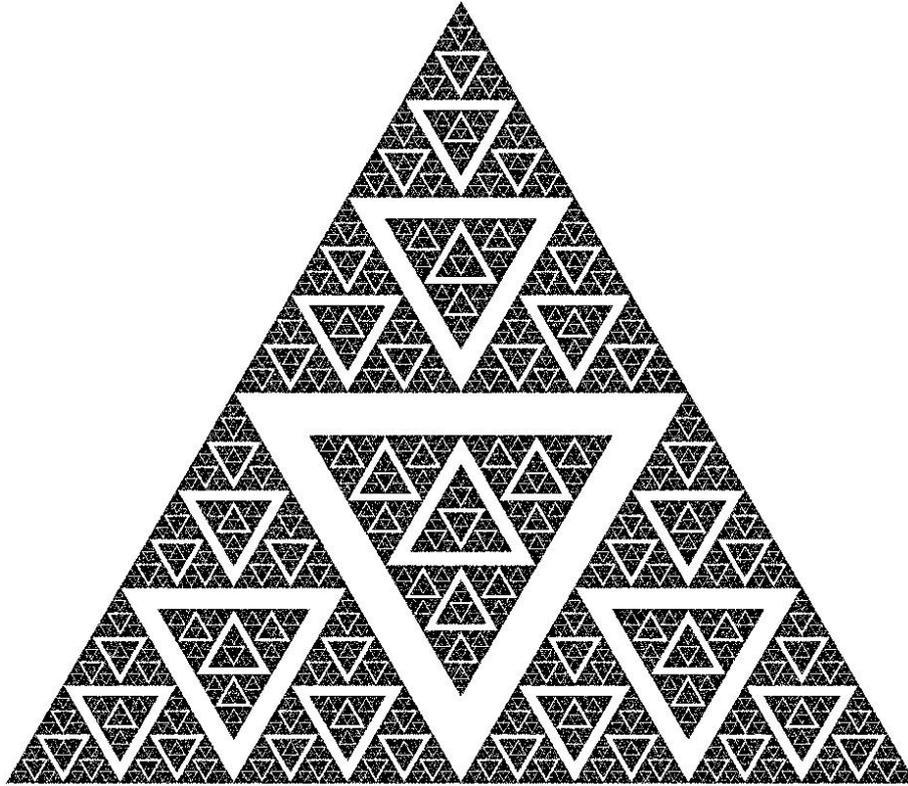
Тогда

$$S_{a_0} = 3,08\sqrt{70,31} = 25,826, \quad S_{a_1} = 3,08\sqrt{0,08} = 0,871,$$

$$S_{a_2} = 3,08\sqrt{0,02} = 0,436, \quad S_{a_3} = 3,08\sqrt{0,003} = 0,533.$$

Записываем уравнение регрессии и под коэффициентами их полученные ошибки:

$$\hat{Y}_X = 59,5 - 0,81X_1 - 0,29X_2 + 0,13X_3, \\ (25,826), (0,871), (0,436), (0,533).$$



§ 23. Измерение тесноты множественной линейной корреляционной связи

За меру тесноты линейной связи между факторными и результативным признаками в совокупности принимают множественный или совокупный коэффициент корреляции R , который вычисляют [4] по формуле:

$$R_{1.2.L.k} = \sqrt{1 - \frac{\hat{S}_{1.2.L.k}^2}{\hat{S}_Y^2}}, \quad (99)$$

где $\hat{S}_{1.2.L.k}^2 = \frac{1}{n-k-1} \sum (Y_i - \hat{Y}_{1.2.L.k})^2$ — остаточная дисперсия;

$\bar{S}_Y^2 = \frac{1}{n-1} \sum (Y_i - \bar{Y})^2$ — общая дисперсия результативного признака.

Множественный коэффициент корреляции можно рассчитать, используя парные коэффициенты корреляции. Так, например, для линейной множественной корреляции между Y , X_1 , X_2 коэффициент R вычисляют по формуле:

$$R_{1.2.3} = R_{Y.X_1.X_2} = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{X_1X_2} \cdot r_{YX_2} r_{YX_1}}{1 - r_{X_1X_2}^2}}. \quad (100)$$

Множественный коэффициент корреляции можно получить на основе вычисления определителей, составленных из парных коэффициентов корреляции:

$$\Delta^* = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{vmatrix}, \Delta = \begin{vmatrix} 1 & r_{23} & r_{24} & \dots & r_{2n} \\ r_{32} & 1 & r_{34} & \dots & r_{3n} \\ r_{42} & r_{43} & 1 & \dots & r_{4n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n2} & r_{n3} & r_{n4} & \dots & 1 \end{vmatrix}, R = \sqrt{\frac{\Delta^*}{\Delta}}. \quad (101)$$

Множественный коэффициент корреляции R обладает следующими свойствами:

1. $R \in [0, 1]$.
2. Если $R = 0$, то линейная корреляционная связь между признаками X_i и Y отсутствует, но другая зависимость (функциональная или нелинейная корреляционная) между ними может существовать.
3. Если $R = 1$, то между факторами X_i и Y существует функциональная линейная зависимость.

Величину множественного коэффициента корреляции корректируют, так как при малом числе наблюдений значение R получается завышенным. Корректировку осуществляют по формуле:

$$\hat{R}_{1.2L.k} = \sqrt{1 - (1 - R^2) \frac{n-1}{n-k}}, \quad (102)$$

где \hat{R} — скорректированное значение R , n — число наблюдений, k — число факторных признаков. Корректировка R не производится при условии, если $(n-k)/k \geq 20$. Для коэффициента множественной корреляции определяют среднеквадратическую ошибку по формуле:

$$S_R = \frac{1}{\sqrt{n-1}}. \quad (103)$$

Если выполняется неравенство $R/S_R > 3$, то с вероятностью $p = 0,99$ можно считать R значимым.

Наряду с определением показателя, отражающего тесноту связи результативного признака Y с факторными, вместе взятыми, определяют степень влияния каждого фактора в отдельности на изменение результативного фактора с помощью коэффициентов частной корреляции. Если уравнение множественной линейной регрессии между факторами X_1, X_2 и Y имеет вид (90), то коэффициенты частной корреляции рассчитывают по формулам:

$$r_{YX_1(X_2)} = \frac{r_{YX_1} - r_{X_1X_2} \cdot r_{YX_2}}{\sqrt{(1-r_{X_1X_2}^2)(1-r_{YX_2}^2)}}, \quad (104)$$

$$r_{YX_2(X_1)} = \frac{r_{YX_2} - r_{X_1X_2} \cdot r_{YX_1}}{\sqrt{(1-r_{X_1X_2}^2)(1-r_{YX_1}^2)}}. \quad (105)$$

Если линейная регрессия имеет вид

$$\hat{Y}_{1.2.3} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3, \quad (106)$$

то частные коэффициенты корреляции находят по формуле:

$$r_{ij.kh} = \frac{r_{ij.k} - r_{ih.k} \cdot r_{jh.k}}{\sqrt{(1-r_{ih.k}^2)(1-r_{jh.k}^2)}}, \quad (107)$$

где $i \neq j \neq k \neq h = 1, 2, 3, 4$. Для общего случая

$$r_{1.2...p} = \frac{r_{12.34...(p-1)} - r_{1p.34...(p-1)} \cdot r_{2p.34...(p-1)}}{\sqrt{(1-r_{1p.34...(p-1)}^2)(1-r_{2p.34...(p-1)}^2)}}. \quad (108)$$

Если в корреляционную модель включено k факторных признаков, воздействующих на результирующий признак Y , то коэффициент частной корреляции, например, для первого фактора, можно определить по формуле:

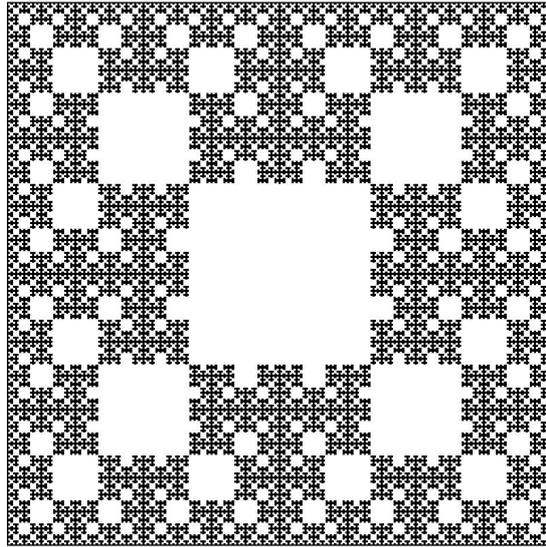
$$r_{Y1.2...k} = \sqrt{1 - \frac{\hat{S}_{Y.1.2...k}^2}{\hat{S}_{Y.2.3...k}^2}}, \quad (190)$$

где

$\hat{S}_{Y.1.2...k}^2 = \frac{1}{n-1} \sum (y_i - \hat{y}_{1.2...k})^2$ — средний квадрат отклонений фактических значений признака Y от значений, вычисленных по формуле с учетом всех факторных признаков;

$\hat{S}_{Y.2.3...k}^2 = \frac{1}{n-1} \sum (y_i - \hat{y}_{2.3...k})^2$ — средний квадрат отклонений фактических значений признака Y от значений, вычисленных по формуле, включающей все факторы кроме первого.

Коэффициенты частной корреляции изменяются от 0 до 1 и обладают всеми свойствами парного коэффициента корреляции. Коэффициенту частной корреляции приписывается тот же знак, который имеет в уравнении множественной линейной регрессии коэффициент регрессии a_i при соответствующем факторном признаке X_i .



§ 24. Проверка адекватности модели множественной линейной корреляции

Адекватность модели означает не только количественное, но, прежде всего, качественное соответствие описания объекта. Для проверки соответствия полученного уравнения множественной линейной регрессии опытными данными используют коэффициент множественной регрессии R . Если $R = 0$, то модель полностью не адекватна. Если $R = 1$, то модель в общем и в целом воспроизводит свойства моделируемого объекта. Количественным показателем адекватности модели служит коэффициент детерминации R^2 , который показывает долю дисперсии, объясняемой данной моделью в общей дисперсии. Адекватность построенной модели может быть проверена по критерию Фишера — Снедекора. Для этого вычисляют [4] статистику F_H по формуле:

$$F_H = \frac{R^2(n-p-1)}{(1-R^2)p}, \quad (110)$$

где n — объем выборки; p — число факторных признаков, включенных в модель. Затем находят при заданном уровне значимости α и числах степеней свободы $k_1 = p$, $k_2 = n - p - 1$ по таблице критических точек распределения Фишера F_T . Если $F_H > F_T$, то уравнение регрессии согласуется с опытными данными, если же $F_H < F_T$, то уравнение регрессии не согласуется с данными эксперимента.

Адекватность модели множественной корреляции можно определять по средней ошибке аппроксимации

$$e = \frac{1}{p} \left[\frac{|Y - \hat{Y}_{1,2,L,p}|}{Y} \right] \cdot 100\%. \quad (111)$$



§ 25. Экономическая интерпретация уравнения регрессии

Заключительным этапом, завершающим построение корреляционной модели, является интерпретация полученного уравнения регрессии, т.е. перевод его с языка статистики и математики на язык экономики. Интерпретация начинается с выяснения, как каждый факторный признак, входящий в модель, влияет на величину результативного признака. Чем больше величина коэффициента a_i регрессии, тем сильнее фактор X_i влияет на результативный признак Y . Знаки коэффициентов регрессии a_i говорят о характере влияния на результативный признак. Если коэффициент a_i имеет знак (+), то с увеличением данного фактора X_i результативный фактор Y возрастает. Если коэффициент a_i имеет знак (-), то с увеличением данного фактора X_i результативный признак уменьшается. Интерпретация знаков зависит от экономической сущности результативного признака. Если величина результативного признака должна изменяться в сторону увеличения (объем реализованной продукции, фондоотдача, производительность труда и т.д.), то плюсовые знаки коэффициентов a_i свидетельствуют о положительном влиянии соответствующих факторов. Если величина результатив-

ного признака изменяется в сторону снижения (себестоимость продукции, материалоемкость, простой оборудования и т.д.), то в этом случае положительное влияние на результативный признак будут оказывать факторы, коэффициенты которых отрицательны.



Рис. 12. Алгоритм построения многофакторной модели: объяснение — управление — предсказание.

Если экономический анализ подсказывает, что факторный признак должен влиять положительно, а коэффициент при нем имеет знак $(-)$, то необходимо проверить расчеты. Так получается за счет допущенных ошибок при решении и в силу наличия взаимосвязей между факторными признаками, включенными в модель, влияющих в совокупности на результативный признак.

При построении регрессионной модели можно рекомендовать алгоритм выполнения операций, представленный на рис.12.

Контрольные вопросы:

1. Рассказать о механизме включения факторных признаков в модель множественной линейной корреляции.

2. Как найти коэффициенты a_0 , a_1 , a_2 уравнения регрессии

$$\hat{Y}_{1.2} = a_0 + a_1 X_1 + a_2 X_2?$$

3. Записать модельное уравнение множественной линейной регрессии для случая, когда в модель включено четыре фактора.

4. Записать систему нормальных уравнений для уравнения

$$\hat{Y}_{1.2.3} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3.$$

5. Как определяется надежность коэффициентов уравнения множественной линейной регрессии?

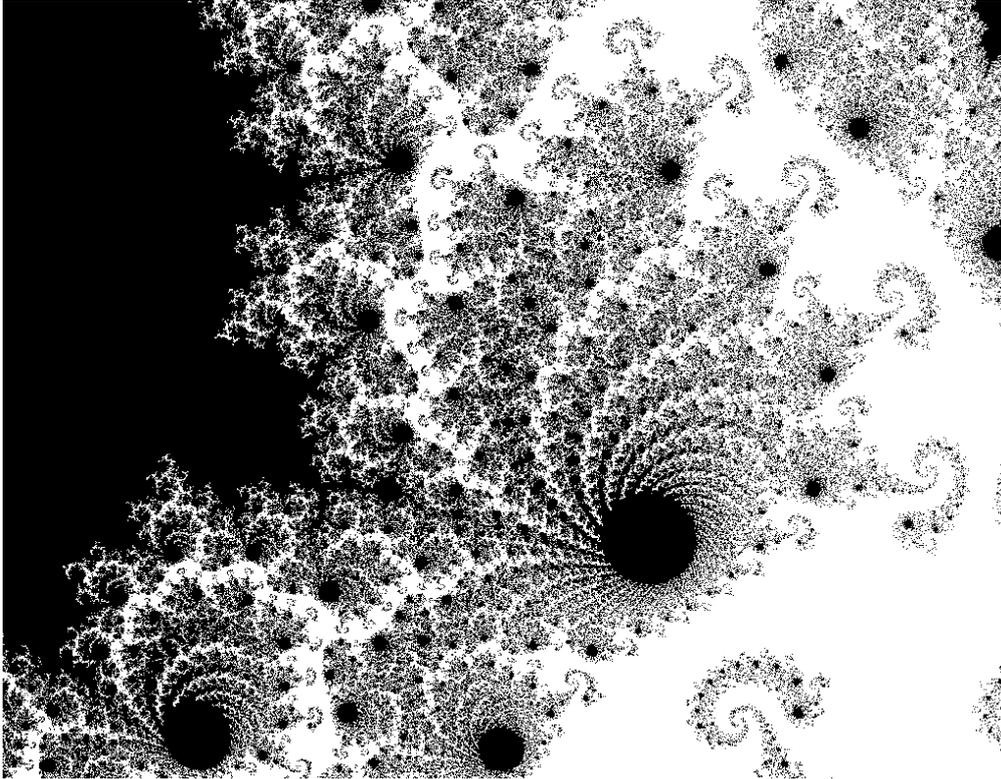
6. Как решается вопрос об измерении тесноты связи между факторными и результативными признаками в случае множественной линейной корреляции?

7. Как осуществляется корректировка множественного коэффициента корреляции?

8. Как определить степень влияния каждого факторного признака в отдельности, включенного в модельное уравнение множественной линейной регрессии, на изменение результативного признака?

9. Рассказать, как осуществляется проверка адекватности модели множественной линейной корреляции.

10. Рассказать об экономической интерпретации уравнения множественной линейной регрессии.



§ 26. Лабораторная работа № 6. Построение модели множественной линейной корреляции

Цель работы: овладение способами построения модели множественной линейной корреляции, выработка умений и навыков нахождения параметров уравнения, оценки надежности уравнения регрессии и его параметров, проведения экономической интерпретации полученных результатов.

Содержание работы: на основании опытных данных требуется:

1. Определить форму связи между факторными и результативными признаками, построив корреляционные поля на плоскости для каждой пары факторов. Записать уравнение множественной регрессии.
2. Произвести отбор факторов, включаемых в модель.
3. Определить тесноту связи между факторами, включенными в модель множественной линейной корреляции.
4. Найти оценки уравнения регрессии по методу наименьших квадратов.
5. Проверить адекватность полученного модельного уравнения регрессии тремя способами:
 - с помощью коэффициента детерминации R^2 ;

- по критерию Фишера;
 - с помощью средней ошибки аппроксимации.
6. Определить воздействие неучтенных в модели факторов.
7. Дать экономическую интерпретацию найденных оценок уравнения регрессии.

З а д а ч а. Исходные данные для признаков X_1, X_2, X_3, Y — для различных нефтегазодобывающих управлений — приведены в табл. 40:

Т а б л и ц а 4 0

Признаки	Значение признаков на различных НГДУ									
X_1	0,92	0,93	0,89	0,90	0,90	0,89	0,92	0,91	0,93	0,89
X_2	45	47	42	46	43	45	48	46	48	44
X_3	69	71	64	66	65	63	68	66	69	65
Y	35	36	31	33	34	32	38	34	37	33

В таблице обозначено: X_1 — коэффициент эксплуатации скважин; X_2 — дебит скважин (тн/сут.); X_3 — уровень автоматизации труда (%); Y — производительность труда (тн/чел.).

Определим форму связи. Для чего строим корреляционные поля (рис. 13 – 18), по которым можно предположить, что зависимость между факторными признаками X_1, X_2, X_3 и результативным признаком Y может носить линейный характер. Решим вопрос о включении факторных признаков X_1, X_2, X_3 в уравнение линейной регрессии. Найдем коэффициенты парной корреляции по формуле (53). Предварительно составим расчетную табл. 41. Пользуясь табл. 41 и формулами (49) — (50), находим:

$$\hat{S}_{X_1}^2 = \frac{1}{n-1} [(X_1 - \bar{X}_1)^2] = 0,00236/9 = 0,000262, \hat{S}_{X_1} = 0,0162.$$

$$\hat{S}_{X_2}^2 = \frac{1}{n-1} [(X_2 - \bar{X}_2)^2] = 37,8/9 = 4,2, \hat{S}_{X_2} = 2,0494.$$

$$\hat{S}_{X_3}^2 = \frac{1}{n-1} [(X_3 - \bar{X}_3)^2] = 58,4/9 = 6,4(8), \hat{S}_{X_3} = 2,5473.$$

$$\hat{S}_Y^2 = \frac{1}{n-1} [(Y - \bar{Y})^2] = 44,1/9 = 4,9, \hat{S}_Y = 2,2136.$$

$$r_{X_1 X_2} = \frac{\overline{X_1 X_2} - \bar{X}_1 \bar{X}_2}{\hat{S}_{X_1} \cdot \hat{S}_{X_2}} = \frac{41,246 - 0,908 \cdot 45,4}{0,0162 \cdot 2,0494} = 0,69.$$

Таблица 41

Факторы	Значения факторов												Сумма	Среднее
	0,92	0,93	0,89	0,9	0,9	0,89	0,92	0,91	0,93	0,89	0,93	0,89		
X_1	0,92	0,93	0,89	0,9	0,9	0,89	0,92	0,91	0,93	0,89	0,93	0,89	9,08	0,908
X_2	45	47	42	46	43	45	48	46	48	44	48	44	454	45,4
X_3	69	71	64	66	65	63	68	66	69	65	69	65	666	66,6
Y	35	36	31	33	34	32	38	34	37	33	37	33	343	34,3
X_1Y	32,2	33,48	27,59	29,7	30,6	28,48	34,96	30,94	34,41	29,37	34,41	29,37	311,73	31,173
X_2Y	1575	1692	1302	1518	1462	1440	1824	1564	1776	1452	1776	1452	15605	1560,5
X_3Y	2415	2556	1984	2178	2210	2016	2584	2244	2553	2145	2553	2145	22885	2288,5
X_1X_2	41,4	43,71	37,38	41,4	38,7	40,05	44,16	41,86	44,64	39,16	44,64	39,16	412,46	41,246
X_2X_3	3105	3337	3688	3036	2795	2835	3264	3036	3312	2860	3312	2860	302683	3026,8
X_3X_1	63,48	66,03	56,96	59,4	58,5	56,07	62,56	60,06	64,17	57,85	64,17	57,85	605,08	60,508
X_1^2	0,8464	0,6849	0,7921	0,81	0,81	0,7921	0,8464	0,8281	0,6849	0,7921	0,6849	0,7921	8,247	
X_2^2	2025	2209	1764	2116	1849	2025	2304	2116	2304	1936	2304	1936	20648	
X_3^2	4761	5041	4096	4356	4225	3969	4624	4356	4761	4225	4761	4225	44414	
$10^6 (X_1 - \bar{X}_1)^2$	144	484	324	64	64	324	144	4	484	324	484	324	2360	
$(X_2 - \bar{X}_2)^2$	0,16	2,56	11,56	0,36	5,76	0,16	6,76	0,36	6,76	1,96	6,76	1,96	36,4	
$(X_3 - \bar{X}_3)^2$	5,76	19,36	6,76	0,36	2,56	12,96	1,96	0,36	5,76	2,56	5,76	2,56	58,4	
$(Y - \bar{Y})^2$	0,49	2,89	10,89	1,69	0,09	5,29	13,69	0,09	7,29	1,69	7,29	1,69	44,1	

$$r_{X_1 X_3} = \frac{\overline{X_1 X_3} - \overline{X_1} \cdot \overline{X_3}}{\hat{s}_{X_1} \hat{s}_{X_3}} = \frac{60,508 - 0,908 \cdot 66,6}{0,0162 \cdot 2,5473} = 0,85,$$

$$r_{X_2 X_3} = \frac{\overline{X_2 \cdot X_3}}{\hat{s}_{X_2} \cdot \hat{s}_{X_3}} = \frac{3026,8 - 45,4 \cdot 66,6}{2,0494 \cdot 2,5473} = 0,60.$$

По найденным коэффициентам парной корреляции видно, что сильно коррелируют между собой факторы X_1 или X_3 . Для решения вопроса о

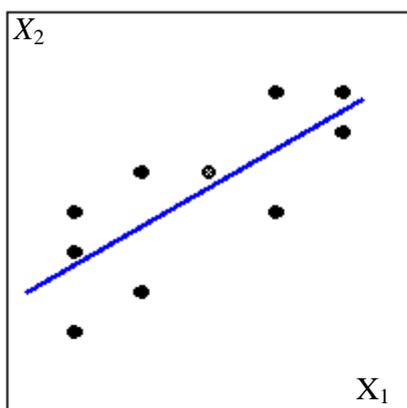


Рис.13.

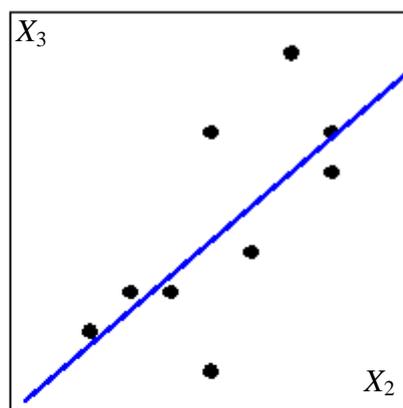


Рис.16.

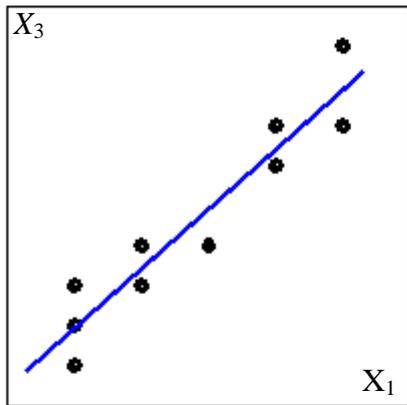


Рис.14.

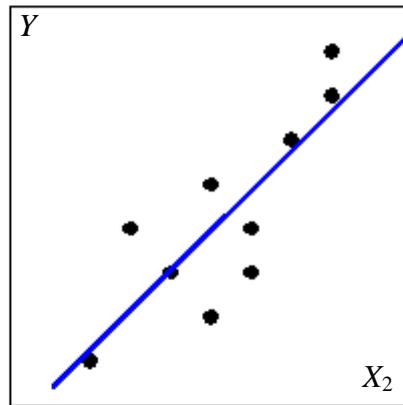


Рис.17.

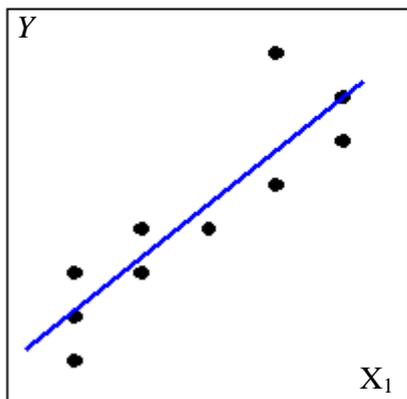


Рис.15.

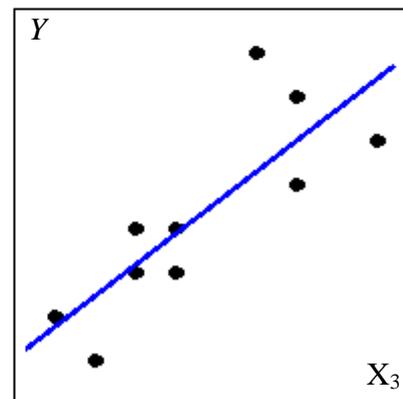


Рис.18.

том, какой из факторов X_1 или X_3 следует исключить из модели множественной линейной корреляции, вычислим коэффициенты парной корреляции r_{YX_1} и r_{YX_3} :

$$r_{YX_1} = \frac{\overline{YX_1} - \bar{Y} \cdot \bar{X_1}}{\hat{S}_Y \cdot \hat{S}_{X_1}} = \frac{31,173 - 34,3 \cdot 0,908}{2,2136 \cdot 0,0162} = 0,80,$$

$$r_{YX_3} = \frac{\overline{YX_3} - \bar{Y} \cdot \bar{X_3}}{\hat{S}_Y \cdot \hat{S}_{X_3}} = \frac{2288,5 - 34,3 \cdot 66,6}{2,2136 \cdot 2,5473} = 0,73.$$

Так как $r_{YX_1} > r_{YX_3}$, то между признаками X_1 и Y связь сильнее, чем между X_3 и Y . Этот факт подтверждается путем вычисления коэффициентов частной корреляции $r_{YX_1(X_3)}$ и $r_{YX_3(X_1)}$ по формуле (104):

$$r_{YX_1(X_3)} = \frac{r_{YX_1} - r_{X_1X_2} \cdot r_{YX_3}}{\sqrt{(1-r_{X_1X_3}^2)(1-r_{YX_3}^2)}} = \frac{0,80 - 0,85 \cdot 0,73}{\sqrt{(1-0,85^2)(1-0,73^2)}} = 0,50,$$

$$r_{YX_3(X_1)} = \frac{r_{YX_3} - r_{X_1X_3} \cdot r_{YX_1}}{\sqrt{(1-r_{X_1X_3}^2)(1-r_{YX_1}^2)}} = \frac{0,73 - 0,85 \cdot 0,80}{\sqrt{(1-0,85^2)(1-0,80^2)}} = 0,16.$$

Поэтому из модели множественной линейной корреляции исключаем фактор X_3 . Тогда в модель будут включены факторы X_1 и X_2 и уравнение регрессии запишется в виде

$$\hat{Y}_{1,2} = a_0 + a_1 X_1 + a_2 X_2.$$

Включение фактора X_2 в модель обосновано значимостью коэффициента парной корреляции r_{YX_2} :

$$r_{YX_2} = \frac{\overline{YX_2} - \bar{Y} \cdot \bar{X_2}}{\hat{S}_Y \cdot \hat{S}_{X_2}} = \frac{1560,5 - 34,3 \cdot 45,4}{2,2136 \cdot 2,0494} = 0,72.$$

Для выяснения вопроса о силе линейной связи между факторами, включенными в модель, вычисляем множественный коэффициент корреляции R по формуле (100):

$$R = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{X_1X_2} \cdot r_{YX_1} \cdot r_{YX_2}}{1 - r_{X_1X_2}^2}} = \sqrt{\frac{0,80^2 + 0,72^2 - 2 \cdot 0,69 \cdot 0,80 \cdot 0,72}{1 - 0,69^2}} = 0,83.$$

Так как в нашем примере объем выборки небольшой ($n = 10$), то произведем корректировку R по формуле (102):

$$\hat{R} = \sqrt{1 - (1 - R^2)^{\frac{n-1}{n-k}}} = \sqrt{1 - (1 - 0,83^2)^{\frac{9}{8}}} = 0,81.$$

Проверяем значимость $R = \hat{R}$ по критерию Стьюдента. Вычисляем средне-квадратическую ошибку S_R по формуле (103):

$$S_R = \frac{1}{\sqrt{n-1}} = \frac{1}{3} = 0,3,$$

Вычисляем статистику

$$t_{\text{н}} = \frac{\hat{R}}{S_R} = \frac{0,81}{0,3} = 2,7$$

По таблице критических точек распределения Стьюдента при уровне значимости $\alpha = 0,05$ с числом степеней свободы $k = n - 2 = 10 - 2 = 8$ находим $t_{\text{Т}} = 1,86$. Так как $t_{\text{н}} = 2,7 > t_{\text{Т}}$, то делаем вывод, что $R = \hat{R}$ значим.

Для нахождения оценок a_0 , a_1 , a_2 уравнения регрессии $\hat{Y}_{2,3} = a_0 + a_1 X_1 + a_2 X_2$ решаем систему нормальных уравнений по формуле (91):

$$\begin{cases} 10a_0 + 9,08a_1 + 454a_2 = 343, \\ 9,08a_0 + 8,247a_1 + 412,46a_2 = 311,73, \\ 454a_0 + 412,46a_1 + 20648a_2 = 15605. \end{cases} \quad (112)$$

Решив эту систему, получаем $a_2 = 0,360611$, $a_1 = 86,3271$, $a_0 = -60,45674$. Тогда уравнение регрессии, устанавливающее зависимость производительности труда Y от коэффициента эксплуатации X_1 и дебита скважин X_2 запишется в виде $\hat{Y}_{1,2} = -60,45674 + 86,3271X_1 + 0,360611X_2$.

Проверяем адекватность уравнения регрессии. Используем коэффициент детерминации R^2 , полагая $R = \hat{R}$. Для полученной модели $\hat{R}^2 = 0,81^2 = 0,66$. Это означает, что полученная модель приблизительно на 66% объясняет изменение производительности труда в зависимости от изменения включенных в модель факторов X_1 и X_2 , что является не плохим показателем.

Проведем проверку модели на адекватность по критерию Фишера — Снедекора. Найдем статистику $F_{\text{н}}$ по формуле (110), полагая в ней $R = \hat{R}$:

$$F_{\text{н}} = \frac{\hat{R}^2 \cdot (n-p-1)}{(1-\hat{R}^2) \cdot p} = \frac{0,81^2 \cdot (10-2-1)}{(1-0,81^2) \cdot 2} = 3,5.$$

По таблице критических точек распределения Фишера — Снедекора при

уровне значимости $\alpha = 0,05$ и числах степеней свободы $k_1 = p = 2$, $k_2 = n - p - 1 = 10 - 2 - 1 = 7$ (p — число факторов X_i , включенных в модель, n — объем выборки) находим $F_T = F_{\alpha; k_1; k_2} = F_{0,05; 2; 7} = 3,26$. Так как $F_H > F_T$, то найденное уравнение регрессии, устанавливающее зависимость производительности труда на десяти нефтегазодобывающих управлениях (НГДУ) от коэффициента эксплуатации скважин X_1 и дебита скважин X_2 , значимо описывает опытные данные и может быть принято для руководства.

Оценим адекватность уравнения регрессии по средней ошибке аппроксимации ε , которую вычислим по формуле (111):

$$e = \frac{1}{p} \sum \frac{|Y - \hat{Y}_{1,2}|}{Y} \cdot 100\% .$$

Для нахождения суммы, входящей в формулу, составляем расчетную табл. 42.

Таблица 42

Y	$\hat{Y}_{1,2}$	$ Y - \hat{Y}_{1,2} $	$\frac{ Y - \hat{Y}_{1,2} }{Y}$
35	35,2	0,2	0,000114
36	36,8	0,8	0,017778
31	31,5	0,5	0,005952
33	33,8	0,8	0,019394
34	32,7	1,3	0,049706
32	32,6	0,6	0,008
38	36,3	1,7	0,076053
34	34,7	0,7	0,014412
37	37,1	0,1	0,00027
33	32,2	0,8	0,019394
			0,211073

По табл. 42 находим:

$$\varepsilon = (0,211073 : 2) \cdot 100\% = 10,6\% .$$

Среднеквадратическая ошибка небольшая, что дает основание считать, что построенная модель адекватно описывает опытные данные.

Итак, все три метода проверки модели на адекватность подтвердили гипотезу о том, что уравнение регрессии

$$\hat{Y}_{1,2} = -60,45674 + 86,3271X_1 + 0,360611X_2$$

в целом статистически значимо и хорошо соответствует данным наблюдений.

Дадим экономическую интерпретацию найденных коэффициентов уравнения регрессии. Значение свободного члена $a_0 = -60,45674$ характе-

ризует влияние неучтенных в модели факторов, в частности фактора X_3 (уровень автоматизации труда). Знак минус говорит о том, что отсутствие этого фактора в модели отрицательно сказывается на повышении производительности труда. Величина коэффициента $a_1 = 86,3271$ показывает, что при увеличении коэффициента эксплуатации на 0,01 производительность труда увеличивается в среднем на 86,3271 тн/чел. Коэффициент $a_2 = 0,360611$ показывает, что при увеличении дебита скважин на одну тонну производительность труда увеличивается в среднем на 0,360611 тн/чел.

Варианты заданий к лабораторной работе № 6

Варианты №1 – №10. Данные экспериментального определения производительности Y труда в зависимости от коэффициента X_1 эксплуатации скважин, дебита X_2 скважин, уровня X_3 автоматизации труда приведены в табл. 43. Пользуясь данными табл. 43, выполнить задание (по образцу приведенного выше примера) по вариантам, номера предприятий (НГДУ) для которых указаны в табл. 44.

Таблица 43

Факторы	Значения факторов на различных НГДУ															
	Номера НГДУ															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_1	0,93	0,95	0,94	0,89	0,91	0,90	0,92	0,93	0,89	0,90	0,90	0,89	0,92	0,91	0,9	0,89
X_2	51	40	46	40	49	43	45	44	42	46	40	49	50	46	48	51
X_3	71	74	72	65	68	67	69	72	65	68	65	66	71	67	70	65
Y	35	32	30	31	33	30	34	35	31	33	32	32	31	34	35	30

Здесь: X_1 — коэффициент эксплуатации скважин (в долях),

X_2 — дебит скважин (тн/сутки),

X_3 — уровень автоматизации труда (%),

Y — производительность труда (тн/чел.).

Таблица 44

Варианты	Номера предприятий	Варианты	Номера предприятий
1	1-3, 7-12, 16	6	4-6, 10-16
2	1-3, 7-9, 13-16	7	1-6, 13-16
3	1-6, 10-12, 16	8	7-16
4	1-3, 10-16	9	4-9, 13-16
5	4-12, 16	10	1-9, 16

Вариант № 11. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области, — характеризующие зависимость среднего дебита Y действующих скважин по нефти, от фонда X_1 действующих нагнетательных скважин на конец года, средней приемистости X_2 нагнетательных скважин и фонда X_3 механизированных скважин

на конец года,— приведены в табл. 45.

Таблица 45

Y (т/сут.)	X_1 (шт.)	X_2 (м ³ /сут.)	X_3 (шт.)
3,5	3	31	26
3,5	5	30	27
3,6	6	29	26
3,6	6	24	26
3,5	7	23	25
3,5	7	20	25
3,4	7	20	25
3,3	8	20	24
3,4	8	17	24
3,3	8	17	24
3,2	8	17	23
3,2	8	17	23
3,1	7	16	22
3,2	7	19	22
3,1	8	18	21
3,1	8	16	21
3,0	8	16	20
3,1	8	16	20
3,0	8	15	19
3,0	8	15	19

Вариант № 12. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость объема Y добычи жидкости с начала разработки, от годовой добычи X_1 жидкости из перешедших скважин, среднегодовой обводненности X_2 и от среднего дебита X_3 действующих скважин по жидкости,— приведены в табл. 46.

Таблица 46

Y (тыс. т)	X_1 (тыс. т)	X_2 (%)	X_3 (т/сут.)
107	34,5	2,8	3,9
142	34,4	2,8	4
176	34,3	2,7	4
210	34,2	2,6	4,1
244	34,1	2,5	4,1
278	34	2,4	4,1
312	33,8	2,4	4,3
346	33,7	2,3	4,3
379	33,6	2,2	4,2
413	33,4	2,2	4,4
446	33,2	2,1	4,4
479	33,1	2,0	4,6
512	32,9	2,0	4,5
545	32,7	1,9	4,7
577	32,5	1,8	4,7

Вариант № 13. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость добычи Y нефти с начала разработки от суммарной добычи X_1 нефти из скважин предыдущего года, падения X_2 добычи нефти и фонда X_3 добывающих скважин на конец года,— приведены в табл. 47.

Таблица 47

Y (тыс.т)	X_1 (тыс.т)	X_2 (тыс.т)	X_3 (шт.)
100,5	30,4	-0,5	27
102	33	-0,9	26
133,1	32,1	-1	26
163,1	31,1	-0,9	25
192,6	30,2	-0,9	25
220,9	29,3	-0,9	25
248,5	28,4	-0,9	24
275,1	27,5	-0,8	24
301	26,7	-0,8	23
326,1	25,9	-0,8	23
350,4	25,1	-0,7	22
374	24,3	-0,7	21
379,5	23,6	-0,6	20

Вариант № 14. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость добычи Y нефти с начала разработки от суммарной добычи X_1 нефти из скважин предыдущего года, падение X_2 добычи нефти и коэффициента X_3 нефтеизвлечения,— приведены в табл. 48.

Таблица 48

Y (тыс.т)	X_1 (тыс. т)	X_2 (тыс.т)	X_3 (%)
286	22	- 0,7	0,9
360	22,7	- 0,9	1,2
86,9	31,7	- 0,9	1,9
117,3	31,2	- 0,8	2,5
147,1	30,5	- 0,8	3,2
176,1	29,7	- 0,8	3,8
204,5	29	- 0,7	4,4
232,2	28,4	- 0,7	5,0
259,2	27,7	- 0,6	5,6
285,6	26,4	- 0,6	6,2
311,4	25,8	- 0,6	6,7
336,6	25,2	- 0,5	7,3
361,2	24,6	- 0,5	7,8
385,3	24	- 0,5	8,3
408,7	23,5	- 0,5	8,8

Вариант № 15. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость добычи Y жидкости с начала разработки от годовой добычи X_1 жидкости из перешедших скважин, среднегодовой обводненности X_2 и от

среднего дебита X_3 действующих скважин по жидкости,— приведены в табл. 49.

Таблица 49

Y (тыс.т)	X_1 (тыс. т)	X_2 (%)	X_3 (т/сут.)
90	32,7	4,8	2,7
123	32,7	6,8	2,8
155	32,6	8,8	2,8
188	32,5	10,7	2,8
220	32,4	12,6	2,9
253	32,5	14,4	2,9
285	32,3	16,2	3,0
317	32,2	17,9	3,0
349	32,1	19,5	2,9
381	32	21,2	3,0
413	31,8	22,8	3,0
445	31,7	24,3	3,1
476	31,6	25,8	3,1
508	31,5	27,2	3,2
539	31,4	28,7	3,2

Вариант № 16. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость среднего дебита Y действующих скважин по нефти от фонда X_1 действующих нагнетательных скважин на конец года, средней приемистости X_2 нагнетательных скважин и темпа X_3 отбора от начальных извлекаемых запасов,— приведены в табл. 50.

Таблица 50

Y (т/сут.)	X_1 (шт.)	X_2 (м ³ /сут.)	X_3 (%)
3	1	50	2,5
2,7	2	43	2
2,6	3	31	2,3
2,7	5	27	2,2
2,6	6	22	2,1
2,6	6	22	2,2
2,5	6	22	2,1
2,5	7	19	2
2,5	7	19	2,1
2,5	8	16	1,9
2,4	8	16	1,9
2,4	8	16	1,8
2,4	7	15	1,8
2,3	8	15	1,7
2,3	7	18	1,7
2,3	7	18	1,9
2,2	8	16	2,3
2,3	7	15	1,9
2,3	8	15	2,1
2,2	8	15	2

Вариант № 17. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области, — характеризующие зависимость добычи Y нефти с начала разработки от коэффициента X_1 нефтеизвлечения, темпа X_2 отбора от начальных извлекаемых запасов и среднего дебита X_3 действующих скважин по нефти, — приведены в табл. 51.

Таблица 51

Y (тыс.т)	X_1 (%)	X_2 (%)	X_3 (т/сут.)
102	1,5	2,8	3,6
133,1	1,9	2,8	3,6
163,3	2,4	2,7	3,5
192,6	2,8	2,6	3,5
220,9	3,2	2,5	3,4
275,1	4	2,4	3,4
301	4,4	2,4	3,3
326,1	4,8	2,3	3,2
350,4	5,1	2,2	3,2
374	5,5	2,2	3,1
396,9	5,8	2,1	3,2
419,9	6,1	2	3,1
440,6	6,4	2	3,1
461,5	6,7	1,9	3
481,7	7	1,8	3,1

Вариант № 18. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость добычи Y жидкости сначала разработки от закачки X_1 агента за год, закачки X_2 агента с начала разработки и среднего дебита X_3 действующих скважин по жидкости,— приведены в табл. 52.

Таблица 52

Y (тыс. т)	X_1 (тыс. т/год)	X_2 (тыс. т)	X_3 (т/сут.)
123	44,5	161	2,8
155	44,3	205	2,8
188	44	249	2,8
220	43,7	293	2,9
253	43,4	336	2,9
285	43,1	380	3
317	42,8	422	3
349	42,5	465	2,9
381	42,3	507	3
413	42	549	3
445	41,7	591	3,1
476	41,4	632	3,1
508	41,4	673	3,2
539	40,8	714	3,2
570	40,5	755	3,3

Вариант № 19. Прогнозные показатели разработки по нефти на одном из месторождений Тюменской области,— характеризующие зависимость суммарной добычи Y нефти из скважин предыдущего года от падения X_1 добычи нефти, процента X_2 падения добычи нефти и фонда X_3 добывающих скважин на конец года,— приведены в табл. 53.

Таблица 53

Y (тыс. т)	X_1 (тыс. т)	X_2 (%)	X_3 (шт.)
35	- 1,1	- 2,4	27
33	- 0,9	- 2,6	26
32,1	- 1	- 2,6	26
31,1	- 0,9	- 2,6	25
29,3	- 0,9	- 2,7	25
28,4	- 0,9	- 2,7	24
27,5	- 0,8	- 2,7	24
25,9	- 0,8	- 2,8	23
24,3	- 0,7	- 2,8	22
22,9	- 0,7	- 2,8	21
21,5	- 0,7	- 2,9	20
20,2	- 0,6	- 2,9	19
20	- 0,5	- 3,1	18

Вариант № 20. Исходные данные по цеху акционерного объединения машиностроительного профиля за 10 месяцев,— характеризующие зависимость между себестоимостью Y произведенной продукции, стоимостью X_1 материалов, основной зарплатой X_2 и расходами X_3 по содержанию и эксплуатации оборудования,— представлены в табл. 54.

Таблица 54

Y (тыс. руб.)	X_1 (тыс. руб.)	X_2 (тыс. руб.)	X_3 (тыс. руб.)
82,3	36,4	11,5	14,2
83,8	36,6	11,5	13,9
81,5	37,9	11,6	15,2
83,1	38,2	11,8	16,5
84,3	39,4	12	16,7
82,6	39,8	12,2	17,2
85,4	40,1	12,5	18,3
84,6	41,5	12,6	18,6
86,8	42,6	12,8	19,4
88,3	45,7	13,2	20,7

Вариант № 21. Исходные данные по цеху акционерного объединения машиностроительного профиля за 10 месяцев,— характеризующие зависимость между себестоимостью произведенной продукции Y , стоимостью X_1 материалов, основной зарплатой X_2 и цеховыми расходами X_3 ,— представлены в табл. 55.

Т а б л и ц а 5 5

Y (тыс. руб.)	X_1 (тыс. руб.)	X_2 (тыс. руб.)	X_3 (тыс. руб.)
81,5	37,9	11,6	9,5
82,3	36,5	11,5	10,6
83,8	36,6	11,5	7,8
83,1	38,2	11,8	9,1
84,3	39,4	12	13,6
82,6	39,8	12,2	14,1
85,4	40,1	12,5	14,6
84,6	41,5	12,6	15,1
86,8	42,6	12,8	16
88,3	45,7	13,2	17,2

Вариант № 22. Имеются данные, характеризующие зависимость нормы расхода Y моторного масла на угар и замену марки $Y_0 = 0,55$ от максимальной мощности X_1 двигателя, максимального крутящего момента X_2 , линейной нормы X_3 расхода топлива, и скорости автомобиля X_4 (табл. 56).

Т а б л и ц а 5 6

Y (л/100 л. т)	X_1 (л. с)	X_2 ()	X_3 (л)	X_4 (км/ч)
1,3	39	7,4	12	40
1,3	42	7,6	8	75
0,8	53	8,2	8	90
1,3	53	8,2	11	70
2,2	70	20,5	21,5	40
2,2	72	17	17	80
1,8	73,5	10,8	10	90
2,2	75	17	16	30
2	75	21	22,8	60
2,1	75	17	15	40
2,2	90	17,5	16	70
2,3	90	17	17	60
1,8	98	18,4	15	60
2,8	110	35	39	35
2,2	115	29	27	70
2,1	115	29	29	60
2,1	120	29	35	50
2	150	41	36	45
2	180	47,3	44	40
1,8	175	48	54	40

Вариант № 23. Имеются данные, характеризующие зависимость нормы расхода Y моторного масла на угар и замену марки $Y_0 = 0,55$ от максимальной мощности X_1 двигателя, линейной нормы X_2 расхода топлива, скорости X_3 двигателя и контрольного расхода топлива при данной скорости X_4 (табл. 57).

Таблица 57

Y (л/100 л. т.)	X_1 (л. с.)	X_2 (л)	X_3 (км/ч)	X_4 (л)
1,3	39	12	40	10
1,3	42	8	75	8
0,8	53	8	90	9,45
1,3	53	11	70	8,85
2,2	70	21,5	40	21
2,2	72	17	80	13
1,8	73,5	10	90	9,3
2,2	75	16	30	10,6
2	75	22,8	60	22
2,1	75	15	40	12
2,2	90	16	70	13,1
2,3	90	17	60	10,6
1,8	98	15	60	11,8
2,8	110	39	35	38,5
2,2	115	27	70	23
2,1	115	29	60	20
2,1	120	35	50	35
2	150	36	45	35,9
2	180	44	40	41
1,8	175	54	40	39

Вариант № 24. Имеются данные, характеризующие зависимость нормы расхода Y моторного масла на угар и замену марки $Y_0 = 0,55$ от максимальной мощности X_1 двигателя, диаметра X_2 цилиндра, линейной нормы X_3 расхода топлива и скорости X_4 (табл. 58).

Таблица 58

Y (л/100 л. т.)	X_1 (л. с.)	X_2 (мм)	X_3 (л)	X_4 (км/ч)
1,3	39	76	12	40
1,3	42	76	8	75
0,8	53	72	8	90
1,3	53	76	11	70
2,2	70	82	21,5	40
2,2	72	92	17	80
1,8	73,5	82	10	90
2,2	75	92	16	30
2	75	82	22,8	60
2,1	75	92	15	40
2,2	90	92	16	70
2,3	90	92	17	60
1,8	98	92	15	60
2,8	110	101,6	39	35
2,2	115	92	27	70
2,1	115	92	29	60
2,1	120	92	35	50
2	150	100	36	45
2	180	108	44	40
1,8	175	108	54	40

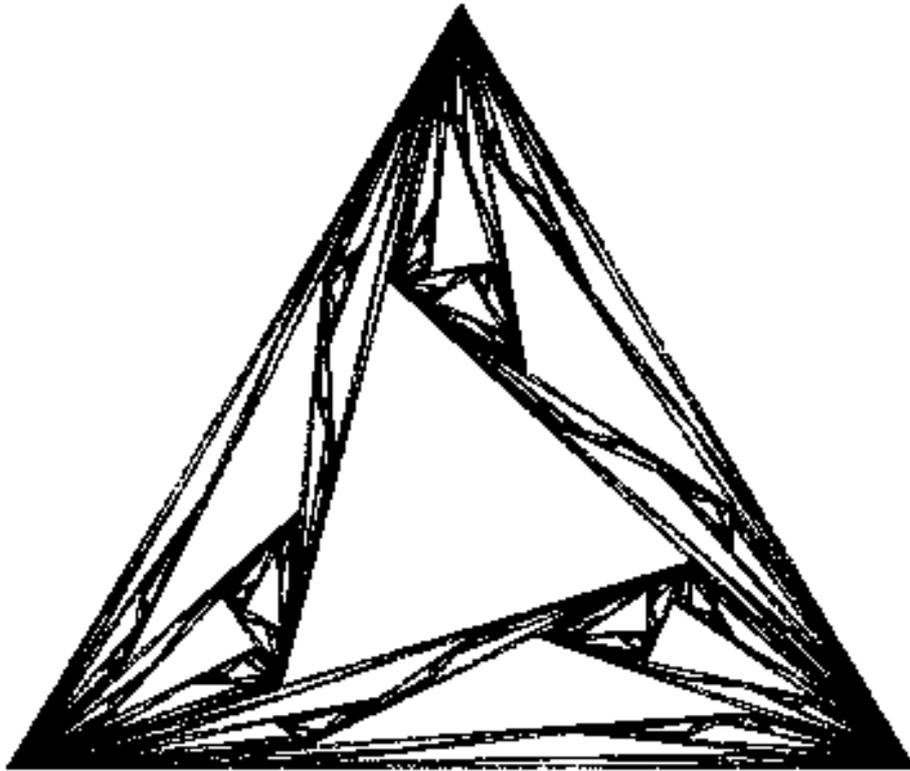
Вариант № 25. Имеются данные, характеризующие зависимость нормы Y расхода моторного масла на угар и замену марки $Y_0 = 0,55$ от максимальной мощности двигателя X_1 , оборотов при максимальной мощности X_2 , линейной нормы X_3 расхода топлива и скорости X_4 (табл. 59).

Таблица 59

Y (л/100 л. т.)	X_1 (л. с.)	X_2 (об/мин)	X_3 (л)	X_4 (км/ч)
1,3	39	4200	12	40
1,3	42	4400	8	75
0,8	53	5400	8	90
1,3	53	5400	11	70
2,2	70	2800	21,5	40
2,2	72	4000	17	80
1,8	73,5	5800	10	90
2,2	75	2600	16	30
2	75	2600	22,8	60
2,1	75	2600	15	40
2,2	90	4000	16	70
2,3	90	4000	17	60
1,8	98	4500	15	60
2,8	110	2800	39	35
2,2	115	3200	27	70
2,1	115	3200	29	60
2,1	120	3300	35	50
2	150	3200	36	45
2	180	3200	44	40
1,8	175	3200	54	40

ГЛАВА 6

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ



§ 27. Пакет MathCAD и математическая статистика

В предыдущих параграфах мы изучили, как обрабатывать статистические данные «вручную», т.е. средневековым способом. Для *понимания* языка математической статистики, для того чтобы хорошо *знать* ее «душу», для *умелого обращения* с математическими моделями в инженерной или иной научной практике,— для формирования статистических компетенций очень значим каждый штрих «ручной» работы. Но сегодня этого уже недостаточно будущему инженеру. Как лопату сменил экскаватор, как каурые и буланные уступили дорогу автомобилям, так и та математическая статистика, которая веками исполнялась «вручную», уступила с настороженной радостью той математической статистике, которая сегодня шагу ступить не может без компьютера, без новых информационных технологий. Это означает, что появились новые — мощные, быстрые, эффективные — способы получения статистик: их обеспечивают вычислительные системы высокого уровня, без которых сегодня немислимы инженерные

расчеты. К таковым относятся, например, системы: Maple, Mathcad, MathLab, Mathematica. У них есть свои достоинства (это радует) и недостатки (это настораживает). Настораживает то, что, предавшись компьютерным вычислениям, скрытым от него на 99,99 %, инженер, особенно тот, который не работал «вручную», может не понять языка математической статистики и может неправильно истолковать полученные статистики, а, следовательно, ошибется при принятии решения.

Мы рекомендуем студентам, обратиться к пакету MathCAD, который не предназначен для программирования сложных задач (для этого есть традиционные языки программирования, системы MathLab, Mathematica), но который завоевал популярность во всем мире благодаря главным своим достоинствам:

-] легкость и наглядность программирования задач;
-] запись сложных математических выражений в том виде, в котором они обычно записываются инженерами на листе бумаги (т.е. отсутствие специального языка программирования) — следствие того, что создатели пакета сделали достаточно удобный интерфейс;
-] возможность создания встроенными средствами высококачественных технических отчетов с таблицами, графиками, текстом.

Кроме мелких недоработок есть два существенных недостатка, ограничивающих применение MathCAD в качестве средства программирования:

- отсутствие встроенных средств отладки программ, которые есть в других средах программирования;
- относительно недостаточная скорость расчетов.

Пакет MathCAD создан разработчиками как инструмент для работы расчетчиков-инженеров. Он не предназначен для профессиональных математиков. Пакет MathCAD создавался как суперкалькулятор, позволяющий легко справляться с рутинными задачами инженерной практики, ежедневно встречающимися в работе. Сюда можно отнести решение алгебраических и дифференциальных уравнений с постоянными и переменными параметрами, анализ функций, поиск их экстремумов, численное и аналитическое дифференцирование и интегрирование, вывод таблиц и графиков при анализе найденных решений.

MathCAD имеет богатый набор функций математической статистики, позволяющих вычислять характеристики выборки данных (средние величины, дисперсию, коэффициенты корреляции и др.), функция плотности вероятности, функции распределения вероятности, квантили вероятности для 17 различных видов распределения случайных величин. Кроме того, в MathCAD есть генераторы случайных чисел, соответствующие всем 17 видам распределения, позволяющие эффективно проводить моделирование методом Монте-Карло.

MathCAD содержит 16 различных функций для оценки параметров

выборки данных:

1. `mean(A)` — возвращает среднее значение элементов массива A ;
2. `hmean(x)` — возвращает среднее гармоническое значение элементов массива A ;
3. `gmean(A)` — возвращает среднее геометрическое значение элементов массива A ;
4. `var(A)` — возвращает дисперсию элементов массива A ;
5. `Var(A)` — возвращает несмещенную дисперсию элементов массива A ;
6. `stdev(A)` — возвращает среднеквадратическое отклонение элементов массива A ;
7. `Stdev(A)` — возвращает несмещенное среднеквадратическое отклонение элементов массива A ;
8. `median(A)` — возвращает медиану массива A , которая делит гистограмму плотности вероятностей на две равные части;
9. `mode(A)` — возвращает моду массива A (наиболее часто встречающееся значение выборки данных);
10. `skew(A)` — возвращает асимметрию массива A (степень асимметричности гистограммы плотности вероятности относительно оси, проходящей через ее центр масс);
11. `kurt(x)` — возвращает эксцесс массива A (степень сглаженности плотности вероятности в окрестности главного максимума);
12. `stderr(A, B)` — возвращает стандартную ошибку при линейной регрессии массивов A, B ;
13. `cvar(A, B)` — возвращает ковариацию элементов двух массивов A, B ;
14. `corr(A, B)` — возвращает коэффициент корреляции двух массивов A, B ;
15. `hist(int, y)` — строит гистограмму массива A ;
16. `histogram(n, y)` — тоже строит гистограмму массива A .

Все приведенные статистические функции могут использоваться для работы как с векторами, так и с матрицами. При этом статистические характеристики рассчитываются для совокупности значений всех элементов матрицы, без деления ее на строки и столбцы. Так, для матрицы размером $m \times n$ объем выборки равен mn .

На рис. 19 показано, как построить гистограмму по данным лабораторной работы № 1 на с. 18.

В MathCAD имеются три функции для оценки связи двух векторов или матриц:

- 】 `stderr(A, B)` — возвращает стандартную ошибку при линейной регрессии массивов A, B ;
- 】 `cvar(A, B)` — возвращает ковариацию элементов двух масси-

вов A, B ;

] $\text{corr}(A, B)$ — возвращает коэффициент корреляции двух массивов A, B .

Чаще на практике используется коэффициент корреляции, дающий относительную, а не абсолютную (как ковариация) оценку связи двух массивов. Чем ближе к единице коэффициент корреляции, тем

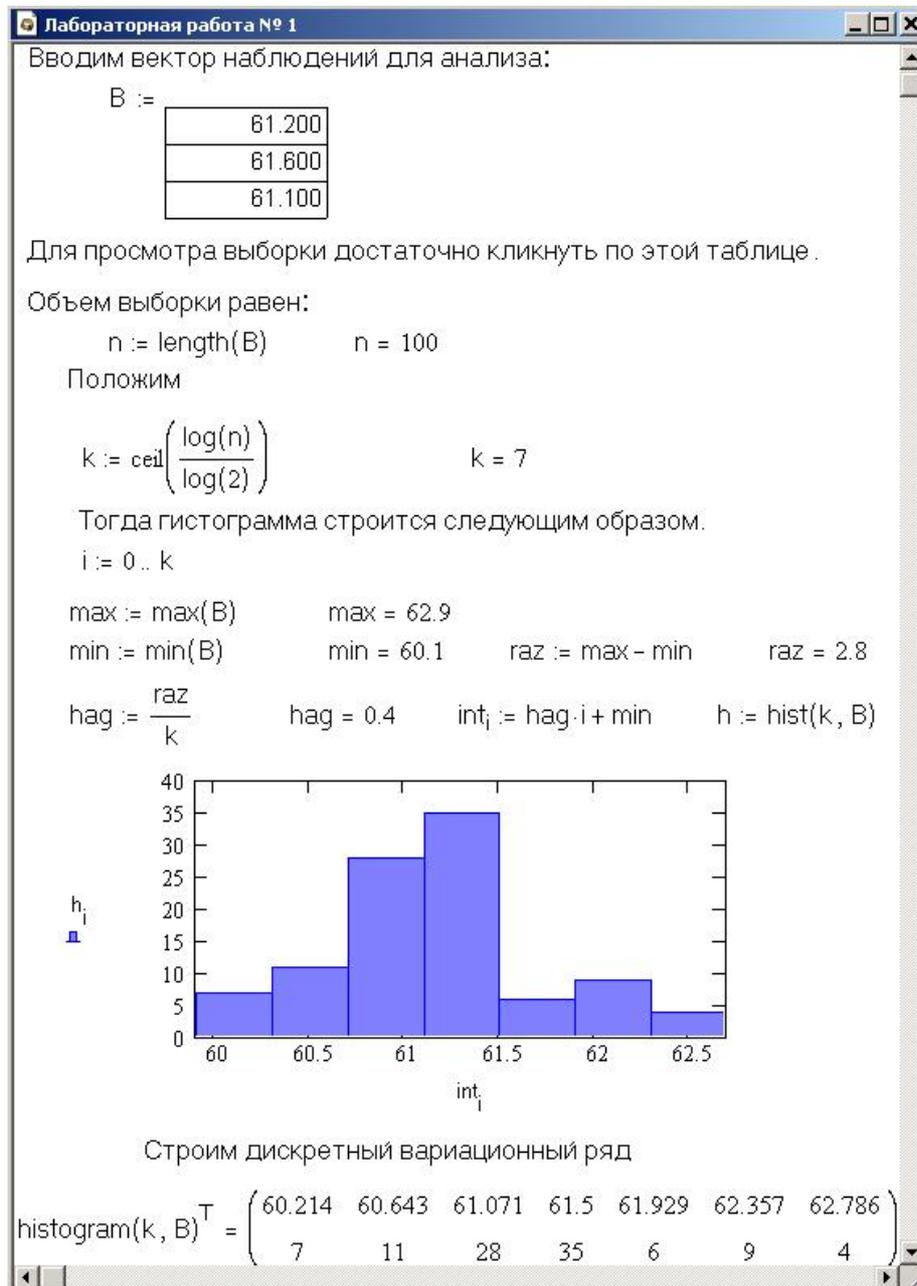


Рис. 19. Построение гистограммы и вариационного ряда.

теснее связь.

В § 4 статистические расчеты были выполнены нами вручную, при этом пришлось проделать огромный объем рутинных вычислений. Пример на рис. 19 показывает, как легко рассчитывается очень трудоемкая статистическая задача. Для построения гистограммы используем встроенную функцию $\text{histogram}(k, B)^T$, возвращающую матрицу из

двух строк. Первая из них — значения середины интервалов, вторая — частоты.

Лабораторная работа № 1_

Вектор наблюдений
 $X := (60.15 \ 60.45 \ 60.75 \ 61.05 \ 61.35 \ 61.65 \ 61.95 \ 62.25 \ 62.55 \ 62.85)$

Частоты $m := (3 \ 6 \ 9 \ 18 \ 29 \ 16 \ 2 \ 10 \ 5 \ 2)$ $X := X^T$ $m := m^T$

Выборочное среднее $X_{cp} := \left| \frac{1}{n} \cdot X^T \cdot m \right|$ $X_{cp} = 61.392$

$\max := \max(X)$ $\max = 62.85$ $\min := \min(X)$ $\min = 60.15$ $Raz := \max - \min$ $Raz = 2.7$

Шаг $h := \frac{Raz}{9}$ $h = 0.3$ $M0X := 61.35$

$i := 0..9$ $u_i := \frac{X_i - M0X}{h}$ $u2_i := u_i \cdot u_i$ $u3_i := u2_i \cdot u_i$ $u4_i := u3_i \cdot u_i$

Начальные моменты

$Mo1 := \frac{1}{n} \cdot \sum_i m_i \cdot u_i$ $Mo2 := \frac{1}{n} \cdot \sum_i m_i \cdot u2_i$ $Mo3 := \frac{1}{n} \cdot \sum_i m_i \cdot u3_i$ $Mo4 := \frac{1}{n} \cdot \sum_i m_i \cdot u4_i$

$Mo1 = 0.14$ $Mo2 = 4$ $Mo3 = 4.28$ $Mo4 = 48.04$

Выборочная дисперсия $S2 := (Mo2 - Mo1^2) \cdot h^2$ $S2 = 0.358$

Выборочное среднее квадратическое отклонение $S := \sqrt{S2}$ $S = 0.599$

Коэффициент вариации $V := \frac{S}{X_{cp}} \cdot 100$ $V = 0.975$

Центральные моменты

$m3 := (Mo3 - 3Mo2Mo1 + 2Mo1^3) \cdot h^3$ $m3 = 0.070348$

$m4 := (Mo4 - 4Mo3Mo1 + 6 \cdot Mo2Mo1^3 - 3 \cdot Mo1^4) \cdot h^4$ $m4 = 0.37$

Асимметрия $As := \frac{m3}{S^3}$ $As = 0.328$

Экцесс $Ex := \frac{m4}{S^4} - 3$ $Ex = -0.115$

Рис. 20. Лабораторная работа № 1.

Функции распределения вероятностей

В MathCAD имеются встроенные функции для оценки 17 видов распределения случайных величин. Эти функции рассчитывают функцию плотности вероятности, функции распределения вероятности, квантиль вероятности, генерируют вектор случайных чисел, распределенных по любому из 17 видов распределения.

Функции распределения вероятностей различаются написанием первой буквы, а оставшаяся часть имени функции указывает на выбранный вид распределения (в списке функций эта часть обозначена звездочкой).

$d^*(x, \text{par})$ — плотность вероятности;

$p^*(x, \text{par})$ — функция распределения;

$q^*(P, \text{par})$ — квантиль распределения;

$r^*(M, \text{par})$ — вектор из M независимых чисел, распределенных по выбранному закону.

Здесь:

- x — значение случайной величины;
- P — значение вероятности;
- par — список параметров распределения.

Чтобы получить функции, относящиеся, например, к нормальному распределению, надо вместо звездочки поставить `norm`. Список параметров для нормального распределения содержит две величины:

- μ — математическое ожидание;
- σ — среднее квадратическое отклонение.

Указанные функции приобретают вид: $dnorm(x, \mu, \sigma)$, $pnorm(x, \mu, \sigma)$, $qnorm(P, \mu, \sigma)$, $gnorm(M, \mu, \sigma)$.

Перечислим все типы распределения вместе с их параметрами, обозначив звездочкой первую букву встроенных функций.

- $*beta(x, s1, s2)$ — бета-распределение, x — значение случайной величины ($0 < x < 1$), $s1, s2$ — параметры ($s1, s2 > 0$).
- $*binom(k, n, p)$ — биномиальное распределение, n — целый параметр ($0 < k < n$), p — вероятность успеха единичного испытания ($0 < p < 1$).
- $*cauchy(x, l, s)$ — распределение Коши, l — параметр разложения, s — параметр масштаба ($s > 0$).
- $*chisq(x, d)$ — χ^2 -распределение (хи-квадрат-распределение), d — число степеней свободы ($d > 0$).
- $*exp(x, r)$ — экспоненциальное распределение, r — показатель экспоненты ($r > 0$).
- $*F(x, d1, d2)$ — распределение Фишера, $d1, d2$ — числа степеней свободы ($d1, d2 > 0$).
- $*gamma(k, s)$ — гамма-распределение, s — параметр формы ($s > 0$).
- $*geom(k, p)$ — геометрическое распределение, p — параметр, равный вероятности успеха единичного испытания ($0 \leq p \leq 1$), k — целое число ($k > 0$).
- $*hypergeom(k, a, b, n)$ — гипергеометрическое распределение, a, b, n — целые параметры.
- $*lnorm(x, \mu, \sigma)$ — логарифмически нормальное распределение, μ — натуральный логарифм математического ожидания, σ — натуральный логарифм среднеквадратичного отклонения ($\sigma > 0$).

- *logis(x, l, s) — логистическое распределение, l — математическое ожидание, s — параметр масштаба (s > 0).
- *nbinom(k, n, p) — отрицательное биномиальное распределение, n — целый параметр (n > 0), 0 < p < 1.
- *norm(x, μ, σ) — нормальное распределение, μ — математическое ожидание, σ — среднеквадратическое отклонение (σ > 0).
- *pois(k, λ) — распределение Пуассона, λ — параметр (λ > 0).
- *t(x, d) — распределение Стьюдента, d — число степеней свободы (d > 0).
- *unif(x, a, b) — равномерное распределение, a, b — границы интервала (a > b).
- *weibull(x, s) — распределение Вейбулла, s — параметр (s > 0).

Теперь мы можем легко найти доверительный интервал в лабораторной работе № 1, для этого достаточно выполнить операции, указанные на рис. 21.

Доверительный интервал

Вводим варианты для проверки $X :=$

61.200
61.600
61.100

$n := \text{length}(X)$ $n = 100$

Зададим уровень значимости $\alpha := 0.05$

Уровень доверия $1 - \alpha = 95\%$

Стандартная ошибка $s := \sqrt{\frac{n}{n-1} \cdot \text{var}(X)}$ $s = 0.609$

Степень свободы $df := n - 1$ $df = 99$

Определяем

критическое значение $t_\gamma := qt\left(1 - \frac{\alpha}{2}, df\right)$ $t_\gamma = 1.984$

верхний предел $U := \text{mean}(X) + t_\gamma \cdot \frac{s}{\sqrt{n}}$ $U = 61.459$

нижний предел $L := \text{mean}(X) - t_\gamma \cdot \frac{s}{\sqrt{n}}$ $L = 61.217$

Рис. 21. Определение доверительного интервала в лабораторной работе № 1.

Примеры определения плотности вероятности нормального распределения для различных σ приведены на рис. 22.

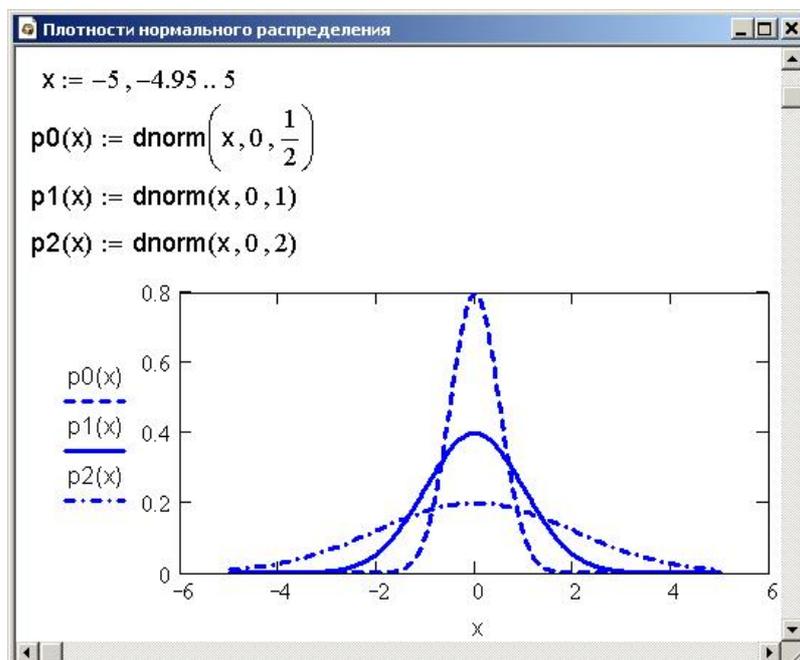


Рис. 22. Графики плотности нормального распределения для $\sigma \in \{1/2, 1, 2\}$.

На рис. 23 – 31 на фоне графика плотности нормального распределения $n(x, 0, 1)$ приведены графики плотности различных распределений: распределения Стьюдента, β -распределения, χ^2 -распределения, экспоненциального распределения, распределения Фишера, логарифмически нормального распределения, логистического распределения, распределение Вейбулла, распределения Коши.

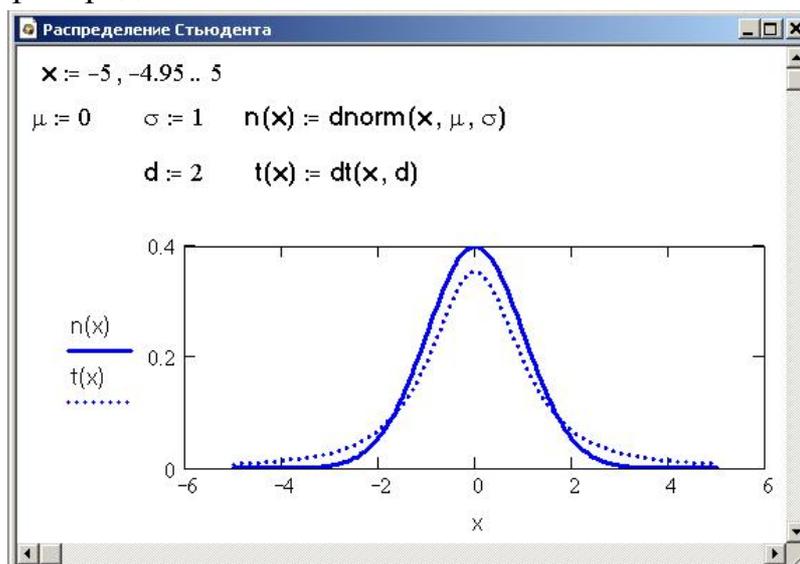
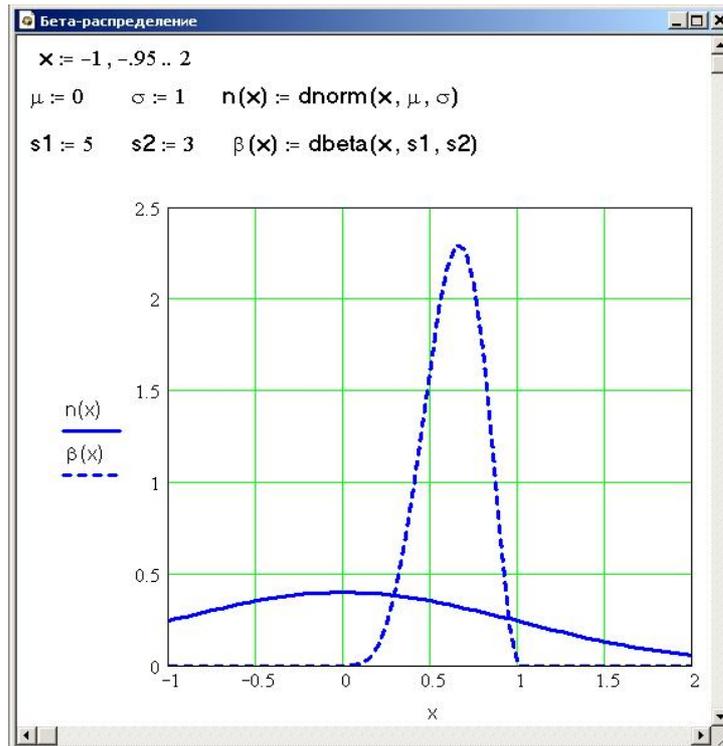
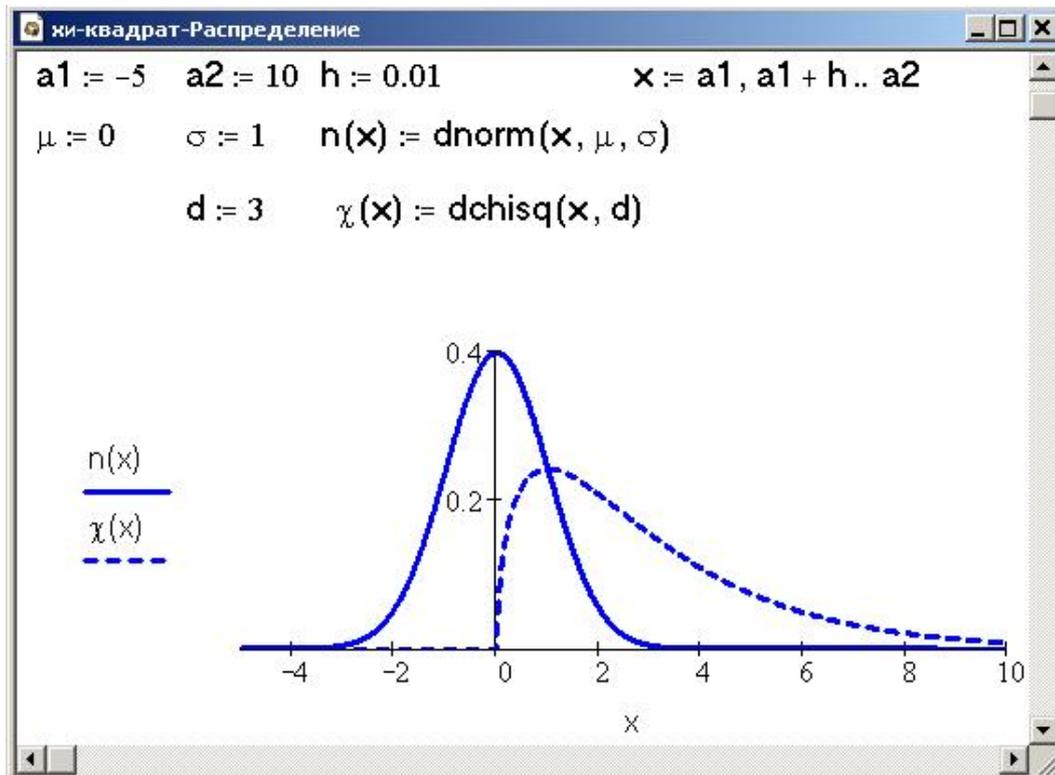


Рис. 23. Графики плотности нормального распределения и распределения Стьюдента.

Рис. 24. Нормальное распределение и β -распределение.Рис. 25. Графики плотности нормального распределения и χ^2 -распределения.

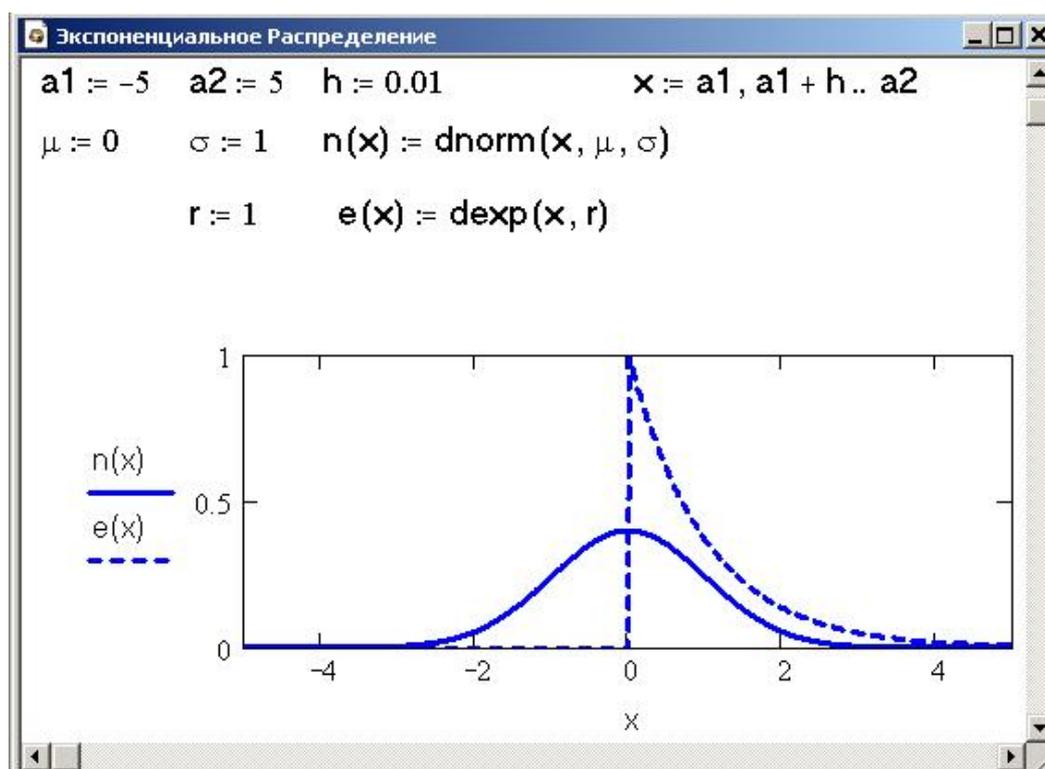


Рис. 26. Графики плотности нормального распределения и экспоненциального распределения.

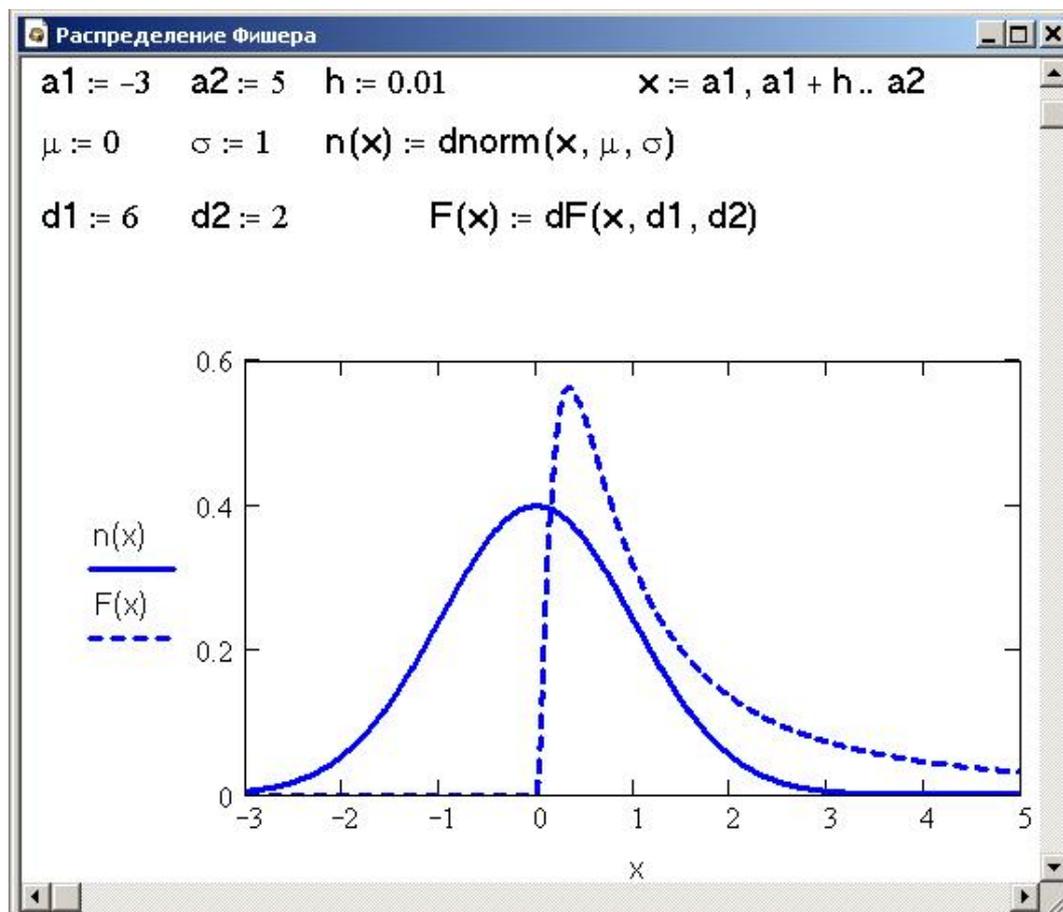


Рис. 27. Графики плотности нормального распределения и распределения Фишера.

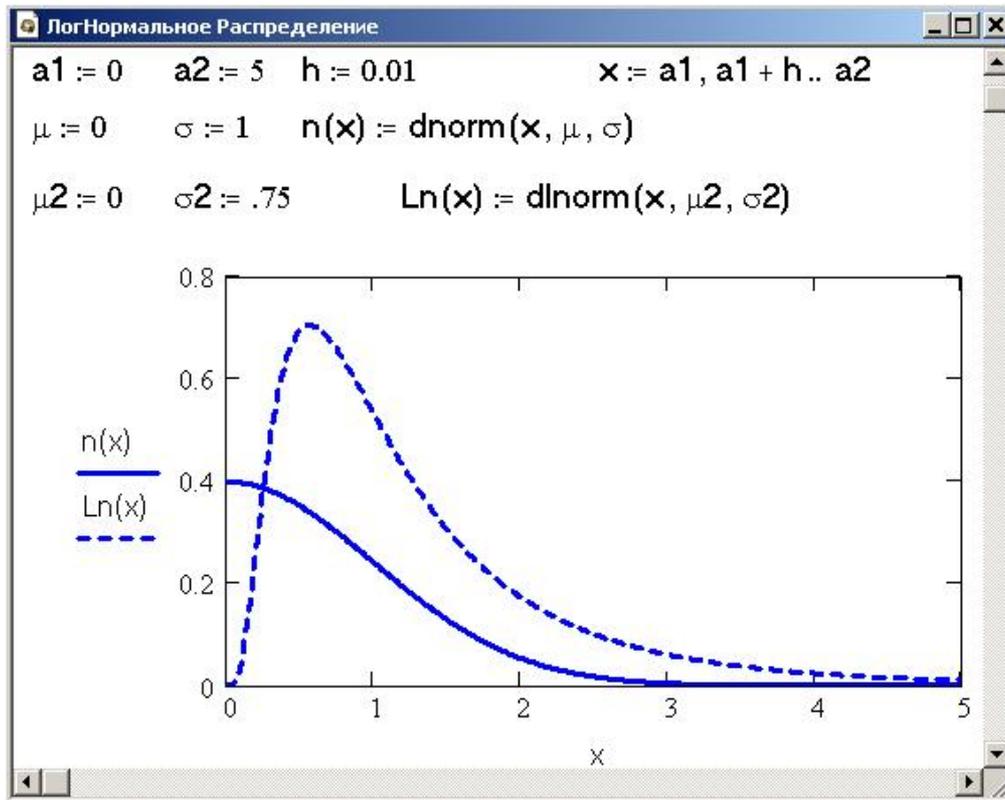


Рис. 28. Графики плотности нормального распределения и логарифмически нормального распределения.

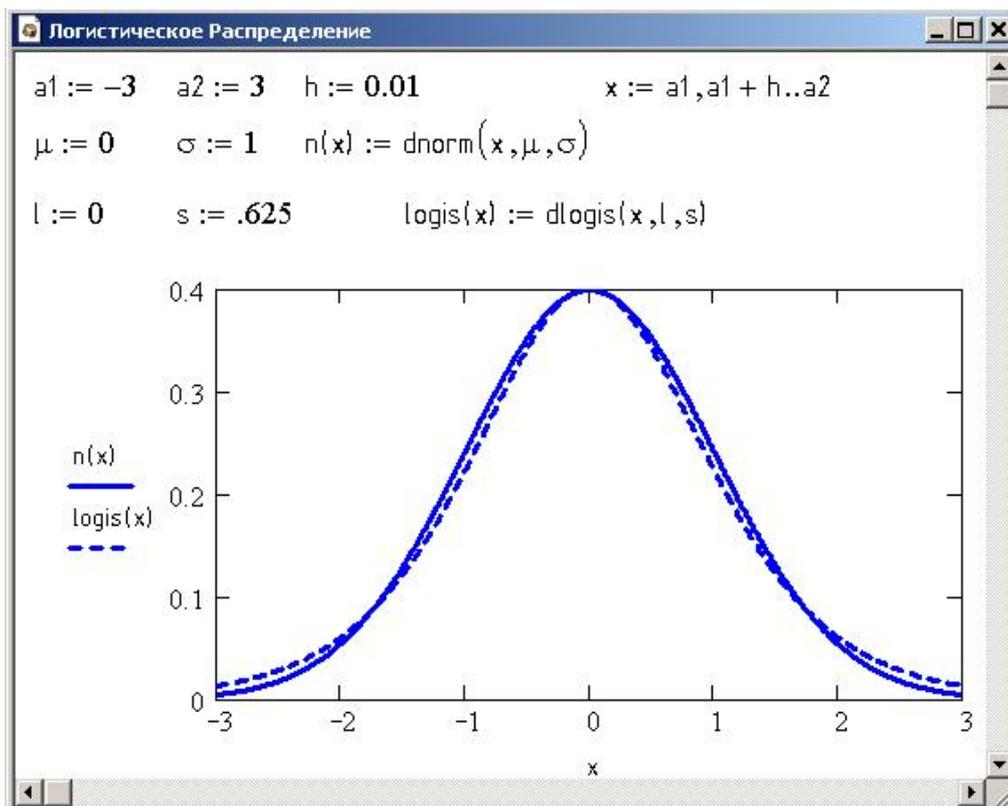


Рис. 29. Графики плотности нормального распределения и логистического распределения.

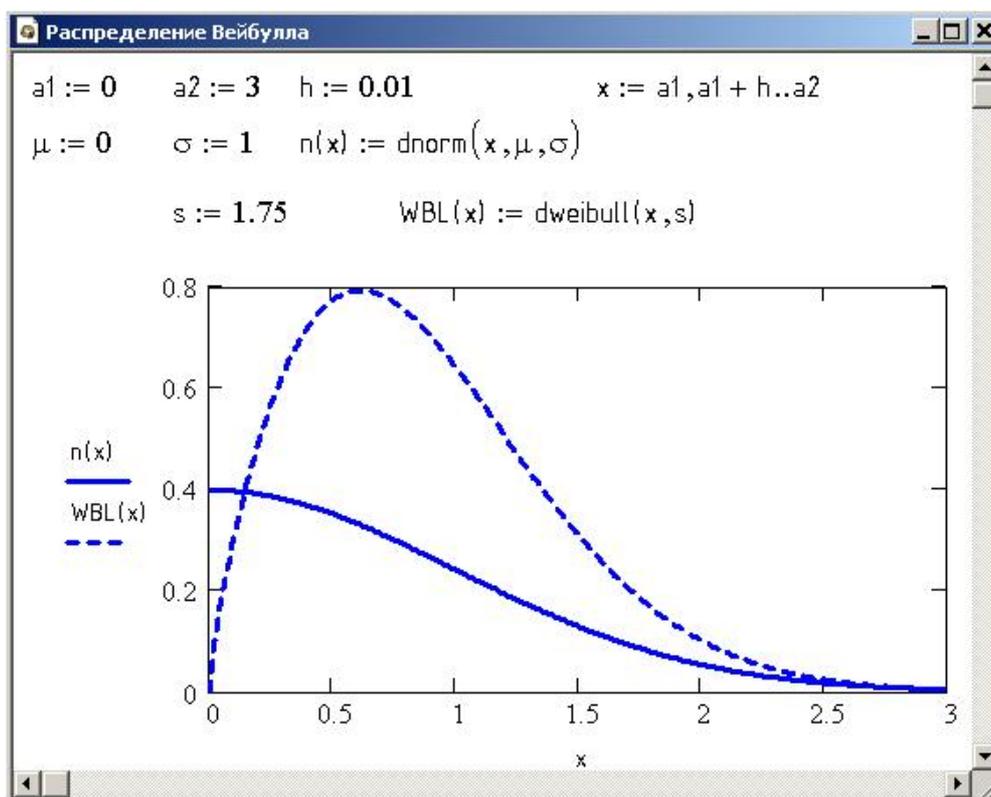


Рис. 30. Графики плотности нормального распределения и распределения Вейбулла.

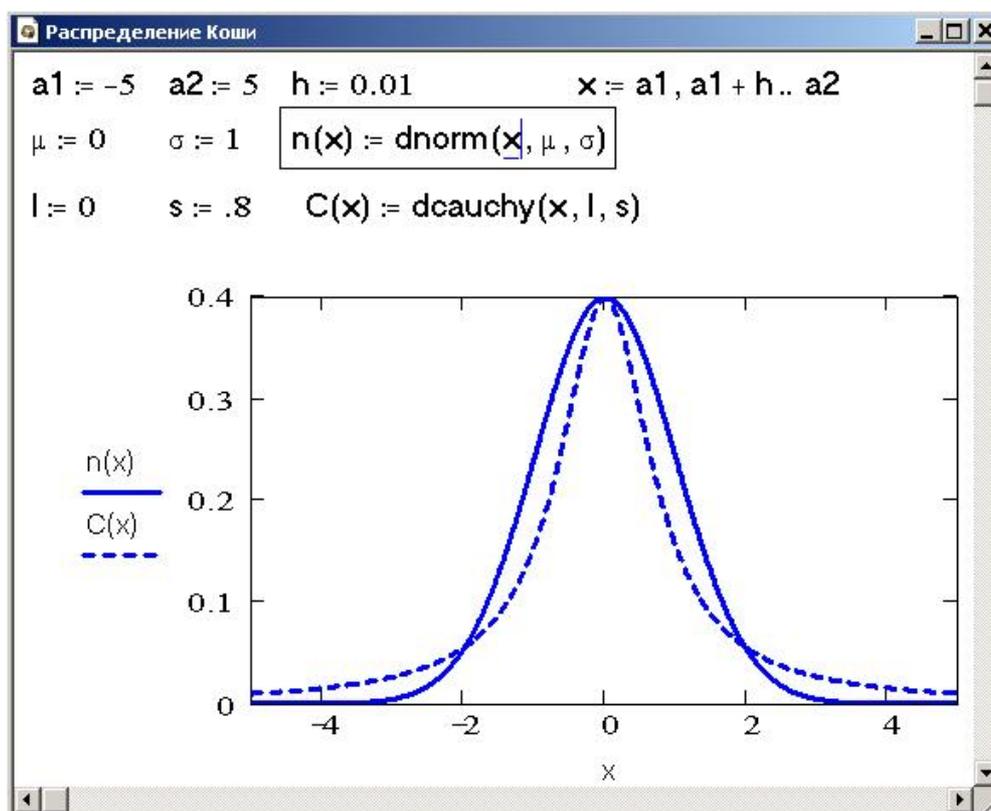
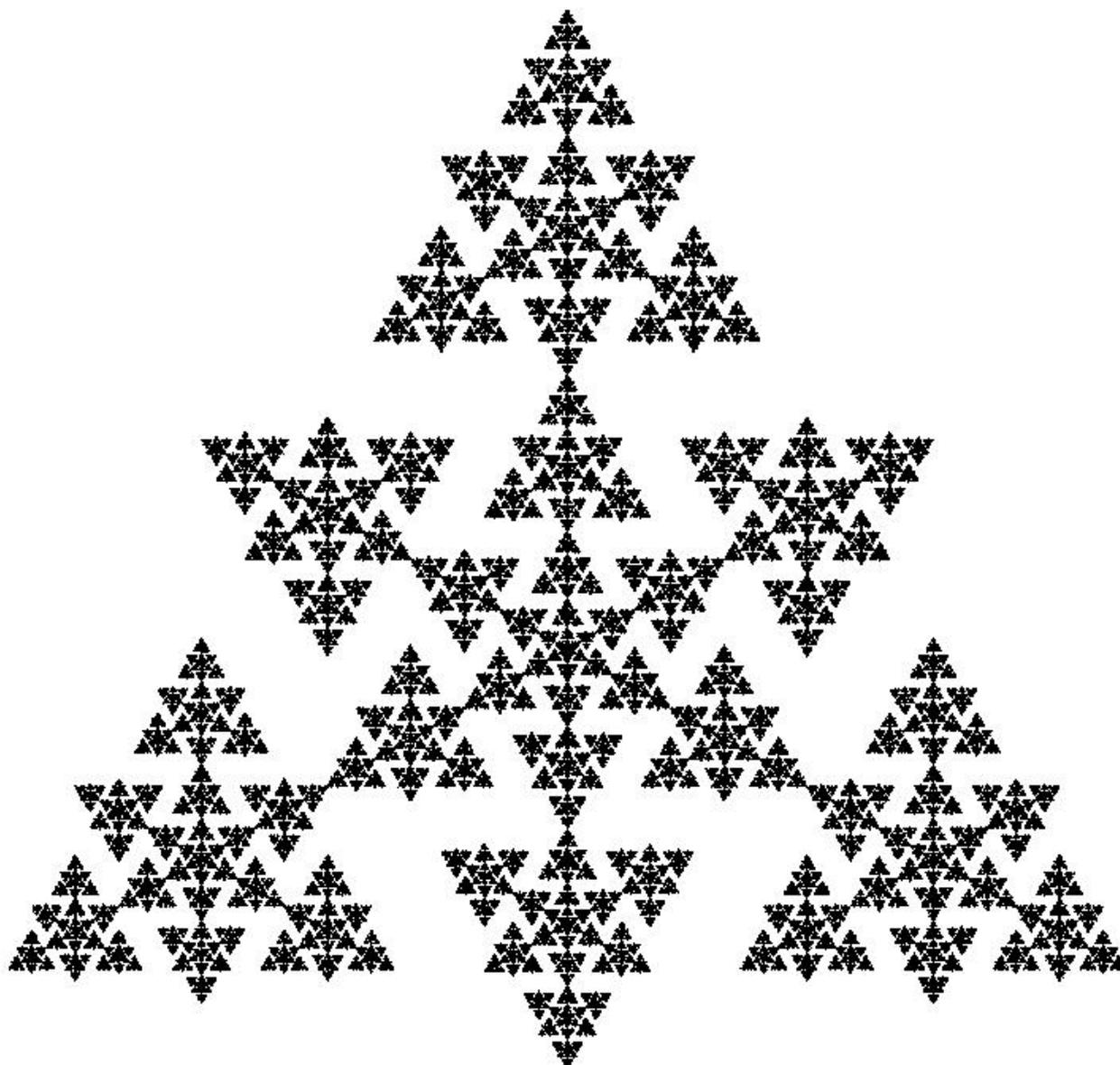


Рис. 31. Графики плотности нормального распределения и распределения Коши.



§ 28. Исследование временных рядов средствами вейвлет-анализа

Временной ряд — это функция от времени. Примеры временных рядов доставляют нам многие наблюдения природных и технических процессов: температура, вибрационные сигналы различных узлов металлообрабатывающего станка, количество осадков, курс рубля по отношению к евро.

Математическая подготовка будущих инженеров, как одна из составляющих их профессиональной компетентности, необходима, прежде всего, для того чтобы инженер, перед которым производство все время ставит практические и теоретические проблемы, мог научно

- о б ъ я с н я т ь происходящие явления,
- у п р а в л я т ь технологическими процессами,

- прогнозировать их последствия.

Рациональное осмысление триединства целей науки [объяснение + управление + прогноз] опирается, в конечном счете, на математическую модель динамического процесса, каковым является, например, вибрация сложных агрегатов (металлорежущих станков, буровых установок, мостов, геофлюидных потоков, нефтегазопроводов и т.п.). В последнюю четверть века стало понятным, что окружающий нас мир

- сильно нелинейный (для его описания используются нелинейные дифференциальные уравнения и нелинейные отображения)

- турбулентный (нет лапласовского детерминизма, но есть детерминированный хаос)

- иерархический (образно говоря, «в каждой матрешке лежит две или более подобных ей матрешек»; это немедленно привело к появлению фрактальной геометрии, изучающей самоподобные объекты и использующей новые методики вычисления — как правило, дробных — размерностей этих объектов).

Понятием самоподобия человечество владеет с тех пор, как люди стали понимать красоту, гармонию. Еще Платон, которого, по словам Павла Флоренского, «равно чтит и христиане, и язычники» и которого «древние апологеты называли *христианином до Христа*», упоминал о самоподобии «сущего — зримого и безвидного», вкладывая в «Диалогах» в уста Сократа слова о том, что гармония несет в себе еще одну гармонию, а дисгармония не содержит никакой гармонии.

Фрактальная геометрия, собрав на начальной стадии своего развития обширную коллекцию фракталов и мультифракталов и научившись их классифицировать, сегодня переходит к объяснению физических причин и следствий фрактальности в сложных системах — природных, технических, абстрактных. При этом еще более возросла роль методики определения фрактальных размерностей, так как теперь они уже являются параметрами конкретных физических моделей, что предъявляет особые требования к точности соответствующей методики.

Вибросигнал представляет собой некую функцию $s = f(t)$ от времени t . Как показывает практика, $f(t)$ обладает тем или иным самоподобием. Это означает, что график самоподобной функции инвариантен относительно некоторого класса непрерывных преобразований плоскости (t, s) . Например, одной из многих используемых модельных функций является функция Ван дер Вардена $f(x) = \sum_{n=0}^{\infty} \varphi_n(x)$, где $\varphi_n(x) = \frac{1}{\pi 2^{n+1}} |\arcsin \sin 2^n \pi x|$, $n \in \mathbb{N}$. Она инвариантна относительно аффинных преобразований

$$A: (x, y) \mathbf{a} \left(\frac{1}{2}x, \frac{1}{2}(x+y) \right),$$

$$B: (x, y) \mathbf{a} \left(\frac{1}{2}(x+1), \frac{1}{2}(-x+y+1) \right)$$

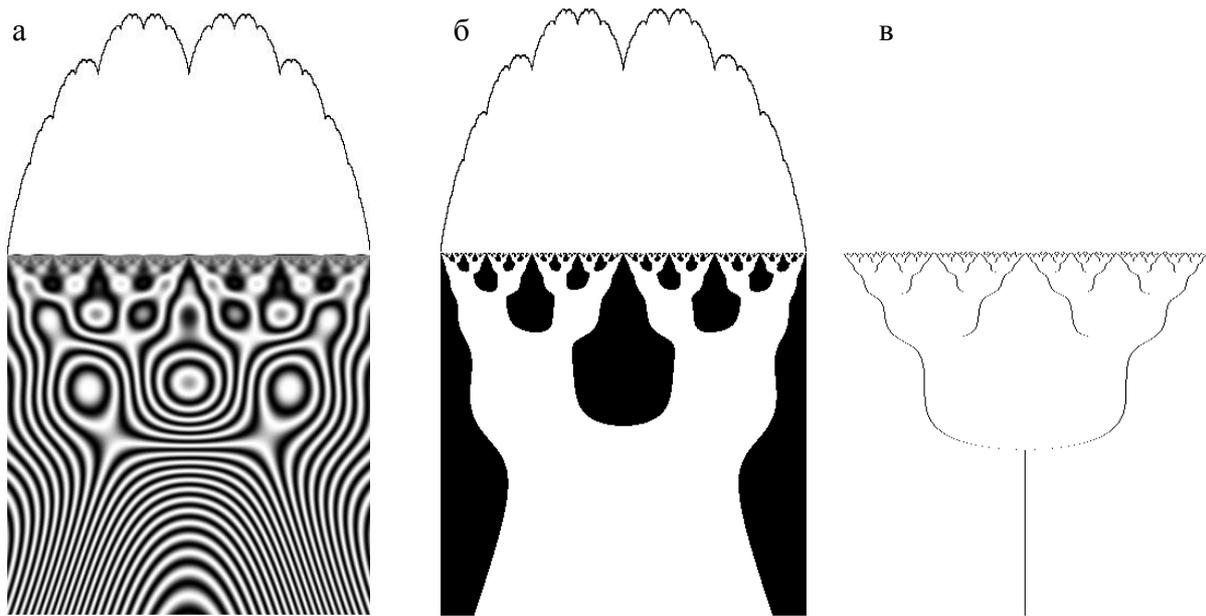


Рис. 32. а — график функции Ван дер Вардена $f(x)$ на отрезке $[0; 1]$ и рельеф поверхности $W(a, b)$, выполненных в 256 оттенках серого цвета; б — график функции $f(x)$ и поверхность W (положительным значениям отвечают белые точки, отрицательным — черные); в — скелетон функции Ван дер Вардена.

[10]. График этой функции представлен на рис. 32 а, б.

Идеи вейвлетов созвучны идеям Фурье-преобразования. Вместо базисной функции в виде комплексной экспоненты вейвлет-преобразование использует базисный вейвлет

$$\psi(t) = \frac{d^8}{dt^8} e^{-t^2},$$

или

$$\psi(t) = 16 \exp(-t^2) (16t^8 - 224t^6 + 840t^4 - 840t^2 + 105),$$

который удовлетворяет требованиям одновременной локализации в частотном и временном пространствах, ограниченности, нулевого среднего (график функции ψ представлен на рис. 33). В этом случае одномерное непрерывное вейвлет-преобразование функции $f(t)$, которая принадлежит пространству $L^2(\mathbf{R})$ (т.е. определена на всей действительной оси и обладает конечной нормой), имеет следующий вид:

$$\left[W^\psi f \right](a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (113)$$

Параметры a и b отвечают соответственно за сжатие и сдвиг вдоль оси t

базисного вейвлета, позволяя полностью сконструировать пространство $L^2(\mathbf{R})$.

Результатом преобразования (113) является поверхность W — дву-

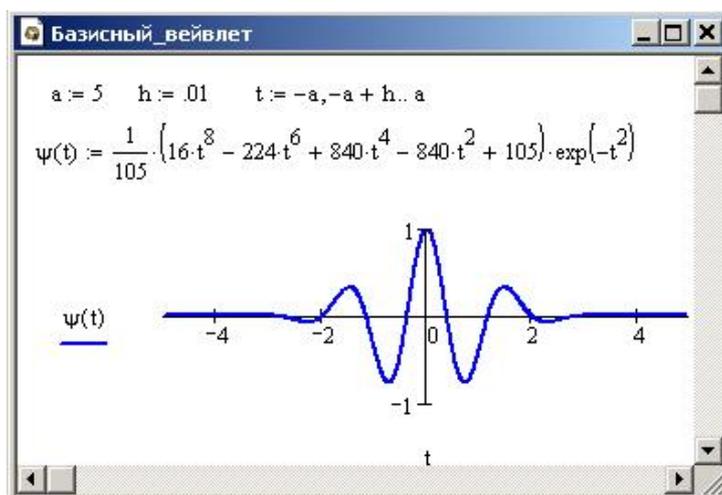


Рис. 33. График базисного вейвлета ψ .

мерный набор коэффициентов (его визуализируют путем проектирования коэффициентов на плоскость ab и сопоставления величине коэффициентов некой цветовой гаммы), в котором, как и в коэффициентах преобразования Фурье, заключена информация о частотных характеристиках сигнала (благодаря параметру a). Более того, сохраняется информация и о поведении сигнала во времени (благодаря параметру b), что выгодно отличает вейвлет-разложение от преобразования Фурье. Платой за это является некоторое ухудшение разрешения по частоте, так как любой вейвлет, в отличие от синусоиды, в частотном пространстве не представляет собой δ -функции. Однако этот недостаток во многом компенсируется гладкостью набора вейвлет-коэффициентов, которая в Фурье-анализе достигается лишь применением искусственных операций. Полученные коэффициенты легко сопоставить с привычным спектром мощности. Для этого достаточно найти следующий интеграл [1]:

$$E(a) = \int W^2(a, b) db. \quad (114)$$

Зависимость $E(a)$ иногда называют *скалограммой* (рис. 34). Смысл *энергетического спектра* ей придает тот факт, что для вейвлет-преобразования существует аналог теоремы Парсеваля. Для того чтобы сопоставить спектр мощности и скалограмму, необходимо перейти от параметра a к частоте ω , воспользовавшись соотношением

$$a\omega = \omega_\psi, \quad (115)$$

где ω_ψ — это константа, зависящая от выбора базисного вейвлета. Проще

всего ее можно оценить, применив формулу (115) к скалограмме модельного сигнала с известными частотами.

Известно, что если функция $f(t)$ самоаффинна, то есть остается инвариантной при масштабировании вида

$$t \text{ а } \lambda t, \quad f \text{ а } \lambda^H f,$$

то ее *спектр мощности* имеет степенной характер:

$$P(\omega) \sim \omega^{-2H-1}. \quad (116)$$

Такое же соотношение, очевидно, справедливо и для скалограммы (114) [14, 15, 16], однако она представляет собой уже гораздо более гладкую

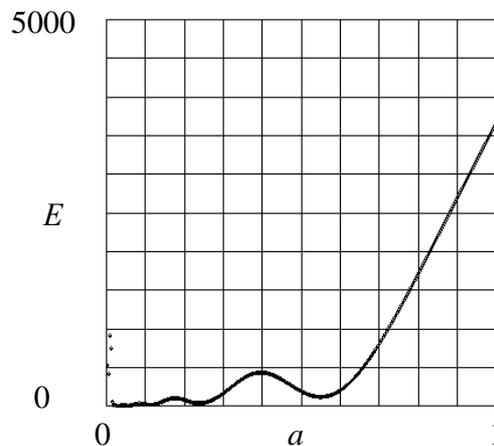


Рис. 34. Скалограмма.

функцию. Это, впрочем, является не главным достоинством вейвлет-преобразования. Естественность использования вейвлетов при анализе фрактальных сигналов обусловлена иерархичностью вейвлетного базиса. Фрактальность подразумевает воспроизведение каких-либо характеристик сигнала на разных масштабах, а вейвлет-преобразование как бы производит иерархизацию особенностей сигнала в зависимости от их масштаба (частоты), с сохранением их пространственной (временной) локализации. В результате, самоподобие сигнала приводит к самоподобию картины его вейвлет-разложения. При этом разработаны различные (помимо скалограммы) способы извлечения численных характеристик из такой самоподобной картины. Так как набор вейвлет-коэффициентов несет в себе избыточную информацию, то часто используют только самую информативную часть данных, а именно, локальные максимумы (вдоль параметра b , при заданном значении a). Картины локальных максимумов иногда называют *скелетонам* (рис. 32 в). Для определения фрактальной размерности D по скелетону пользуются тем фактом, что

$$\lim_{a \rightarrow 0} \frac{\log N(a)}{\log a^{-1}} = D, \quad (117)$$

где $N(a)$ — это количество локальных максимумов на масштабе a . Для кривых Ван дер Вардена, как доказал профессор из Костромы С.Б. Козырев [17], $D = 1$.

В качестве еще одного преимущества вейвлет-преобразования перед преобразованием Фурье можно привести предсказуемость влияния граничных эффектов, а именно, если базисный вейвлет имеет компактный носитель $[-S; S]$, то на масштабе a влияние границы распространяется на расстояние aS . Область, внутри которой не происходит влияния граничных эффектов на величину вейвлет-коэффициентов, называют *треугольником достоверности*. Наличие такой области делает формулу (114) не совсем удобной для практических расчетов. Для устранения влияния граничных эффектов бывает удобно заменить интегрирование усреднением, которое позволяет учитывать только часть коэффициентов.

Формула (5) позволяет определять размерности монофракталов, однако из вейвлет-разложения относительно легко можно извлечь любые мультифрактальные характеристики [14]. Кроме того, вейвлет-преобразование (113) обобщается на многомерный случай [5], что позволяет анализировать двумерные наборы данных напрямую, не прибегая к анализу одномерных срезов.

В качестве примера рассмотрим модельный временной ряд из 65 значений (табл. 60) для дискретного аргумента $t \in [3, 4]$, взятого с шагом $h = 1/64$. [Функцию (118) К. Вейерштрасс предложил в качестве примера непрерывной и всюду недифференцируемой функции.] Значения времени определяются по формуле: $t_n = 3 + n/64, n = 0 \dots 64$.

Т а б л и ц а 60

n	X								
0	0	13	3,0505	26	3,3155	39	3,369	52	2,5433
1	0,7733	14	2,6724	27	3,4755	40	2,7386	53	3,0879
2	1,0156	15	2,58	28	2,7213	41	3,4879	54	2,704
3	1,5902	16	1,9065	29	3,0406	42	3,5459	55	2,5266
4	1,3209	17	2,7284	30	2,5023	43	3,8025	56	1,6669
5	2,1386	18	2,968	31	2,2819	44	3,1234	57	2,1972
6	2,0552	19	3,4905	32	1,5157	45	3,4905	58	2,0552
7	2,1973	20	3,1234	33	2,2817	46	2,9681	59	2,1388
8	1,6668	21	3,8024	34	2,5022	47	2,7286	60	1,321
9	2,5263	22	3,546	35	3,0406	48	1,9066	61	1,5902
10	2,7038	23	3,4881	36	2,7212	49	2,5798	62	1,0157
11	3,088	24	2,7387	37	3,4753	50	2,6723	63	0,7735
12	2,5433	25	3,3689	38	3,3155	51	3,0506	64	0

$$X(t) = \sum_{i=1}^6 \frac{1 - \cos(b^i \pi t)}{b^{(2-D)i}} \quad (118)$$

Отметив точки $(t_n, X(t_n))$ на числовой плоскости, мы, рассматривая полученную картину (рис. 34 а), могли бы поставить привычный в математической статистике вопрос: какая здесь имеется регрессия? Можно даже выдвинуть предположение, что здесь имеется либо полиномиальная зависимость, либо гармоническая. Вообще говоря, исходя только из табл. 60 и из расположения точек в корреляционном поле, трудно судить о виде зависимости величины X от времени t . Для выявления свойств данного ряда составим программу в какой-либо среде, например, в Vision Basic.

Программа вейвлет-преобразования в VB:

```
Private Sub Form_Click()
pi = 3.14159
ix = 1000: iy = 8000: ix3 = ix: iy3 = iy + 200           'Начало отсчета
m1 = 500: m2 = 10 * m1: m3 = m2: m4 = m3               'Масштабы
bb = 2: d1 = 1.6                                       'Параметры функции X(t) Вейерштрасса
alf = 3: bet = 4:                                       'Концы отрезка
fara = 0.1                                             'Регулировка оттенков серого цвета
.....

For i = 0 To 9                                         'Рисуем бумагу в клетку
  For j = 0 To 8
    X1 = ix + m1 * i: Y1 = iy - m1 * j
    X2 = ix + m1 * (i + 1): Y2 = iy - m1 * (j + 1)
    Line (X1, Y1)-(X2, Y2), 0, B
  Next
Next
.....

n = 10000                                             'Рисуем график функции X(t)
raz = bet - alf: hag = raz / (n - 1): step1 = 1 / n   'Шаг
sdwigB = 1.1
ReDim x(n)                                           'Определение массива
For i = 1 To n                                       'Вычисление значений функции X(t)
  summa = 0
  t = alf + (i - 1) * hag
  For j = 1 To 25
    slog = 1 / bb ^ ((2 - d1) * j) * (1 - Cos(bb ^ j * pi * t))
    summa = summa + slog
  Next
  x(i) = summa
Next
For j = 1 To n
  X1 = ix + m2 * ((j - 1) / (n - 1) + sdwigB): X2 = iy - m1 * (x(j))
  PSet (X1, X2), 0
  For rq = 0 To 5
    Circle (X1, X2), rq, vbBlue
  Next
Next
```

Next

```

.....
n = 65
raz = bet - alf: hag = raz / (n - 1): step1 = 1 / (n - 1)
fara = 0.1 'Регулировка оттенков серого цвета
sdwigB = 1.1
ReDim x(n+1)
For i = 1 To n
  summa = 0
  t = alf + (i - 1) * hag
  For j = 1 To 25
    slog = 1 / bb ^ ((2 - d1) * j) * (1 - Cos(bb ^ j * pi * t))
    summa = summa + slog
  Next
  x(i) = summa
Next
Next
For j = 1 To n 'Построение корреляционного поля
  X1 = ix + m2 * (j - 1) / (n - 1): X2 = iy - m1 * x(j)
  PSet (X1, X2), 0
  For rq = 0 To 50
    Circle (X1, X2), rq, vbBlue
  Next
Next
e = 2.718281
alfa = 0: beta = 0.02
X1 = ix3 + m3 * (-0.025): X2 = ix3 + m3 * 0: Y1 = iy3 + m4 * 0
Line (X1, Y1)-(X2, Y1), 0
For k = 0 To 10
  X1 = ix3 + m3 * (-0.025): X2 = ix3 + m3 * 0: Y1 = iy3 + m4 * 1 / 2 ^ k
  Line (X1, Y1)-(X2, Y1), 0
Next
For a = 0.000001 To 0.6 Step 15 / m2 'Построение поверхности W
  Ka = a ^ 0.5
  pic = 0
  For B = 0 To 1 Step 10 / m2
    Wab = 0
    ii = 0
    For t = 0 To 1 Step step1
      ii = ii + 1
      s = ((t - B) / a) ^ 2
      z2 = 16 * e ^ (-s) * (105 - 840 * s + 840 * s ^ 2 - 224 * s ^ 3 + 16 * s ^ 4)
      Wab = Wab + z2 * x(ii) * step1 'Интегрирование
    Next
    Wab = Wab / Ka
    If beta >= alfa And beta >= Wab Then X15 = ix3 + m3 * (B + sdwigB)
    If beta >= alfa And beta >= Wab Then X25 = iy3 + m4 * a
    If beta >= alfa And beta >= Wab Then PSet (X15, X25), 0
    If beta >= alfa And beta >= Wab Then pic = pic + 1
    alfa = beta
    beta = Wab
  Next
  cwet = 255 * (1 + Sin(fara * Wab)) / 2

```

```

cw = RGB(cwet, cwet, cwet)
X1 = ix3 + m3 * B: X2 = iy3 + m4 * a: PSet (X1, X2), cw
Next
Next
End Sub

```

По корреляционному полю на рис. 35 *а* построена поверхность W , представленная на рис. 35 *б*, из которого хорошо видно, что исследуемая нелинейная регрессия обладает периодичностями. Об этом свидетельствуют

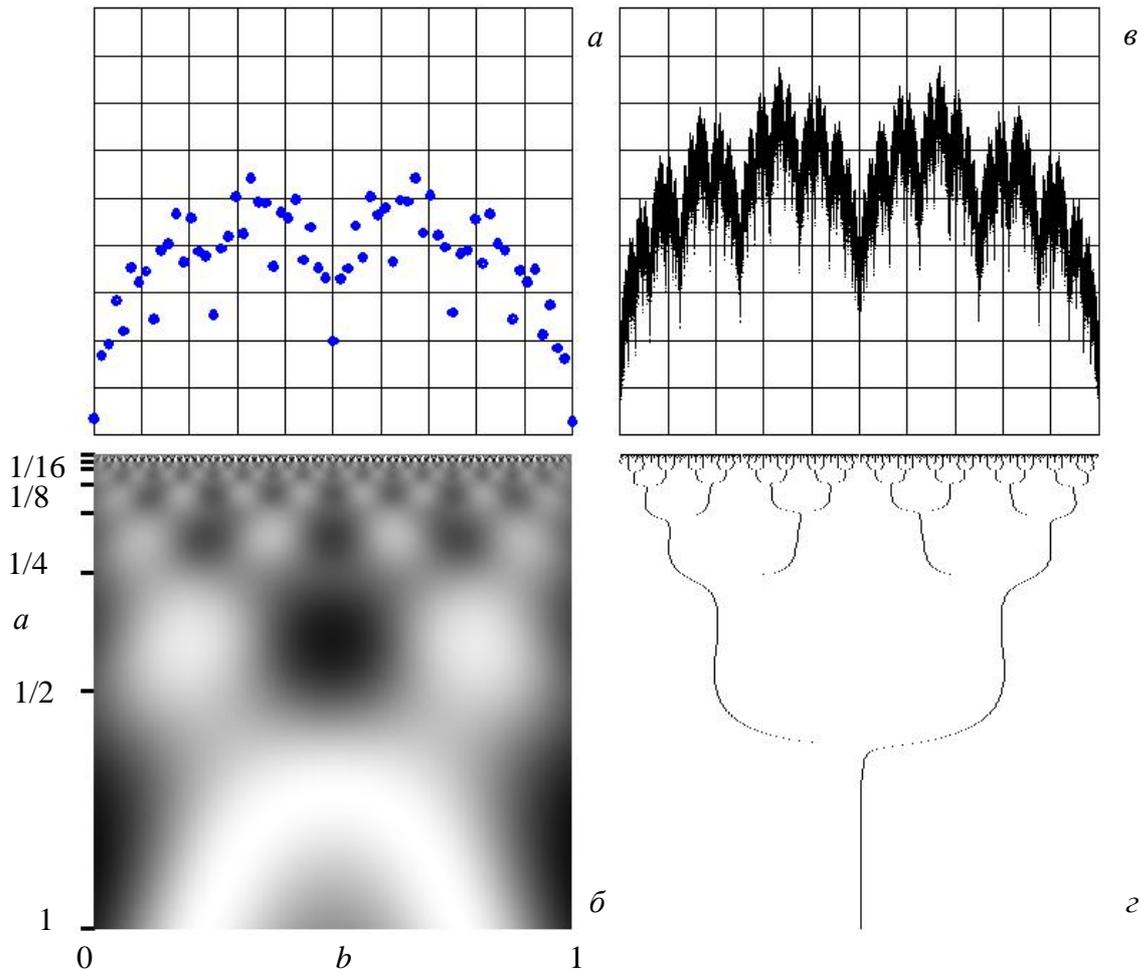


Рис. 35. Анализ нелинейной регрессии посредством вейвлет-анализа: *а* — корреляционное поле, построенное на основании табл. 60; *б* — поверхность W — вейвлет-образ 65 данных точек; *в* — график функции (118); *г* — скелетон поверхности W .

повторяющиеся 3 острова, заключенные в полосе от $a = 1/4$ до $a = 1/2$. В этой полосе два холма и одна впадина. В полосе $[1/8, 1/4]$ имеется 4 холма, 3 впадины; в полосе $[1/16, 1/8]$ — 8 холмов, 7 впадин; в полосе $[1/32, 1/16]$ — 16 холмов, 15 впадин. На более мелких значениях параметра a картина начинает искажаться, так как в рассматриваемом массиве (табл. 60) мало данных. На рис. 35 *в* представлен график функции $X(t)$, из которого хорошо видно, что исследуемая регрессия на самом деле фрактальная, она обладает свойством аффинного самоподобия. Свойство периодичности хорошо демонстрирует скелетон, приведенный на рис. 35 *г*.

$$\text{Значения функции } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$$

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	0,4854
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	0,49999997
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		

Приложение 3

Таблица значений $t_g = t(g, n)$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	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	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Приложение 4

Таблица значений $q_g = q(g, n)$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	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

Приложение 5

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	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

Приложение 6

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,3138	12,7062	31,82	63,7	318,3	637,0
2	2,9200	4,3037	6,97	9,92	22,33	31,6
3	2,3534	3,1824	4,54	5,84	10,22	12,9
4	2,1318	2,7764	3,75	4,60	7,17	8,61
5	2,0150	2,5706	3,37	4,03	5,89	6,86
6	1,9432	2,4469	3,14	3,71	5,21	5,96
7	1,8946	2,3646	3,00	3,50	4,79	5,40
8	1,8595	2,3060	2,90	3,36	4,50	5,04
9	1,8331	2,2622	2,82	3,25	4,30	4,78
10	1,8125	2,2281	2,76	3,17	4,14	4,59
11	1,7959	2,2010	2,72	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
∞	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 α (односторонняя критическая область)						

Приложение 7

Критические точки распределения Фишера — Снедекора $(k_1$ — число степеней свободы большей дисперсии, k_2 — число степеней свободы меньшей дисперсии)

Уровень значимости $\alpha = 0,01$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45
Уровень значимости $\alpha = 0,05$												
$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	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

ЛИТЕРАТУРА

1. Астафьева Н.М. *Вейвлет-анализ: основы теории и примеры применения* // УФН, т.166, №11, 1996.
2. Венецкий И.Г. Кильдышев Г.С. *Основы теории вероятностей и математической статистики*.— М.: Статистика, 1968.— 320 с.
3. Гмурман В.Е. *Теория вероятностей и математическая статистика: Учебное пособие для вузов*.— М.: Высшая школа, 1997.
4. Губин В.И. *Лекции по высшей математике. Ч.2: Учебное пособие для студентов дневной и заочной форм обучения*.— Тюмень: Изд-во ТГНГУ, 1996.
5. Добеши И. *Десять лекций по вейвлетам*.— Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001 (пер. с англ. Daubechies I., Ten Lectures on Wavelets, 1992).
6. Калинина В.Н., Панкин В.Ф. *Математическая статистика: Учебник для студ. сред. спец. учеб. заведений*.— М.: Высшая школа, 2001.
7. Карасев А.И., Аксютин З.М., Савельева Т.И. *Курс высшей математики для экономических вузов. Ч.2: учеб. пособие для студентов вузов*.— М.: Высшая школа, 1982.
8. Климова Е.З., Миллионщикова А.П., Фастовец Н.О. *Задачи по математической статистике*.— М.: МИНХ И ГП им. И.М. Губкина, 1976.
9. Колемаев В.А., Калинина В.Н. *Теория вероятностей и математическая статистика*.— М.: «Инфра — М», 1997.
10. Осташков В.Н., Смирнов Е.И. *Формирование нелинейного мышления студентов посредством визуализации самоподобных множеств* // Труды вторых колмогоровских чтений.— Ярославль: Изд-во ЯГПУ, 2004.— С. 173 – 189.
11. Степнов М.Н. *Статистические методы обработки результатов механических испытаний: Справочник*.— М.: Машиностроение, 1985.
12. Ферстер Э., Рёнц Б. *Методы корреляционного и регрессивного анализа*.— М.: «Статистика», 1987.
13. Янпольский А.Р. *Лабораторный вычислительный практикум*.— М.: Изд-во ВИА, 1968.
14. Hansen A., Scmittbuhl J., Batrouni G.G. *Distinguishing fractional and white noise in one and two dimensions* // Phys. Rev. E, **63**, 2001.
15. Hwang W-L., Mallat S. *Characterization of Self-Similar Multifractals with Wavelet Maxima* // Applied and Computational Harmonic Analysis, **1**, p. 316-328, 1994.
16. Simonsen I., Hansen A. *Fast algorithm for generating long self-affine profiles* // Phys. Rev. E, **65**, 2002.
17. Козырев С.Б. *О фрактальной размерности кривой Ван дер Вардена*. // Препринт.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Адекватность модели 52
– уравнения регрессии 65
Алгоритм 17, 124
Аппроксимация 118
Аппроксимирующая прямая 35
Асимметрия 13, 144

Б

Базисный вейвлет 156
Бумага вероятностная 34

В

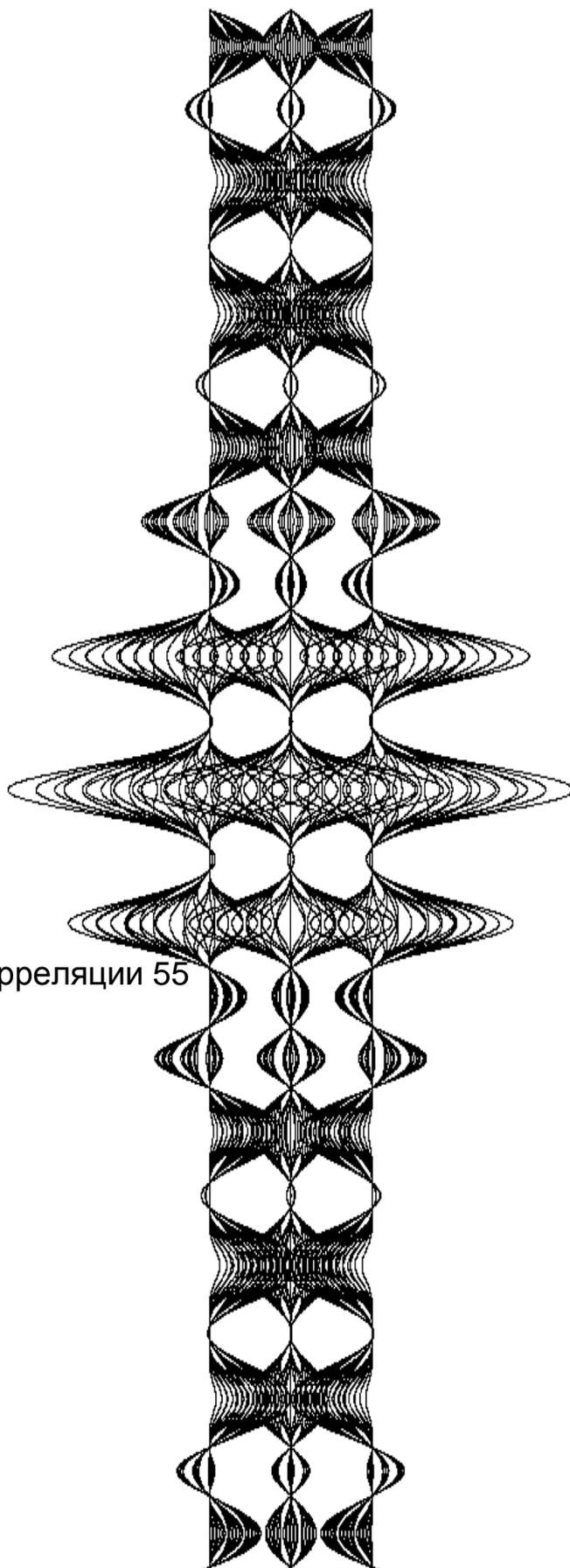
Варианта 6
– условная 13, 19, 24, 56
Варьирование признака 7
Вейвлет-преобразование 156
Временной ряд 154
Выборка 6
– взвешенная 12
– генеральная 6
– простая 12
– средняя 13
Выборочный метод 6
Выборочный коэффициент линейной корреляции 55

Г

Гарантийный максимум 59
– минимум 59
Генеральная совокупность 6
Гипотеза нулевая 37
– альтернативная 37
Гистограмма относительных частот 8
– частот 8, 144
Графический метод 52

Д

Данные сгруппированные 50, 77, 80
Двусторонняя критическая область 58
Дисперсия 12, 144



– исправленная 12
 Длина интервала 13
 Доверительная вероятность 67
 Доверительный интервал 59, 67

З

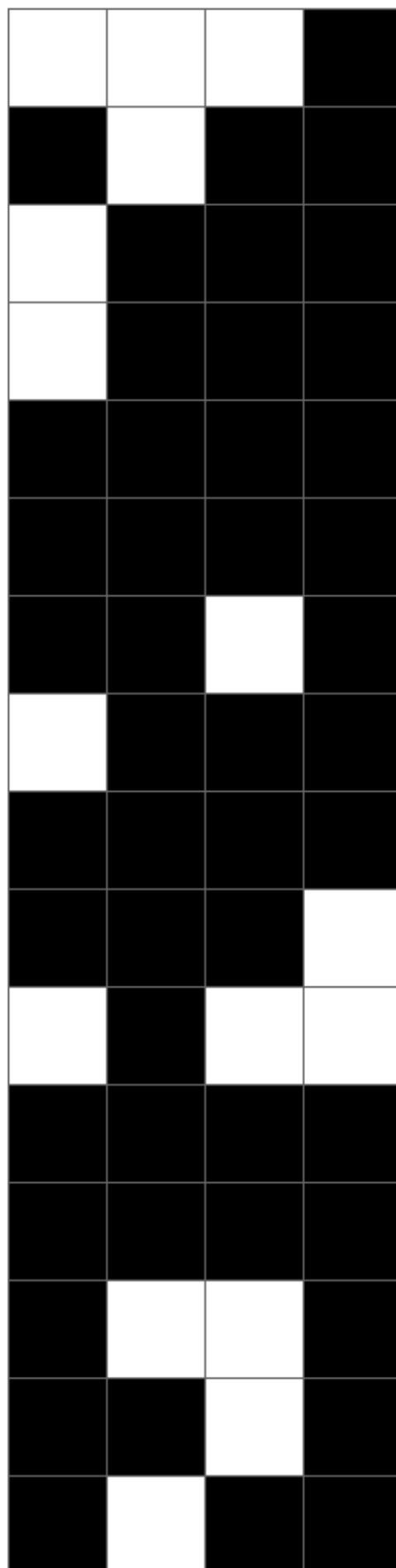
Зависимость корреляционная 47
 – корреляционная линейная 48
 Значение скорректированное 116
 Значения сгруппированные 11
 Значимость коэффициента корреляции 61, 63
 – коэффициента линейной корреляции 60
 – уравнения регрессии 99

И

Индекс корреляции 96, 97
 Интервал 6
 – доверительный 15
 – частичный 18
 Интерполирование 93
 Интерпретация 119

К

Качественные признаки 48
 Квантиль 142
 Ковариация 140
 Колеблемость признака 14
 Конечные разности второго порядка 94
 Корректировка 125,
 Корреляционная зависимость линейная 53
 – связь криволинейная 52, 95
 – таблица 54
 Корреляционное отношение 100
 – поле 52
 Корреляционный анализ 51
 Корреляция линейная 52
 Корреляция парная 53
 – парная линейная 53
 – частная 115
 Коэффициент вариации 15, 20
 – детерминации 60, 118
 – корреляции 55, 144
 – – множественный 120



- линейной корреляции 55
- парной корреляции 115
- частной корреляции 115
- эластичности 54, 55

Критерий Колмогорова 42

– Пирсона 39

– приближенный 45, 46

– Романовского 44

– Стьюдента 58

– Фишера — Снедекора 57

– Ястремского 44

Критическая область критерия 38

Кумулята 8

Л

Линия регрессии 52

– тренда 66

Ложный нуль 13

М

Математическое ожидание 11, 144

Медиана 13, 144

Метод наименьших квадратов 53

– перебора 115

Мода 13, 144

Модель 52

Модельное уравнение регрессии 63

Моменты условные начальные 14

– условные центральные 14

Мощность критерия 38

Н

Надежность 17, 59

Несмещенные оценки 23

Неучтенные факторные признаки 54

О

Объем выборки 6, 144

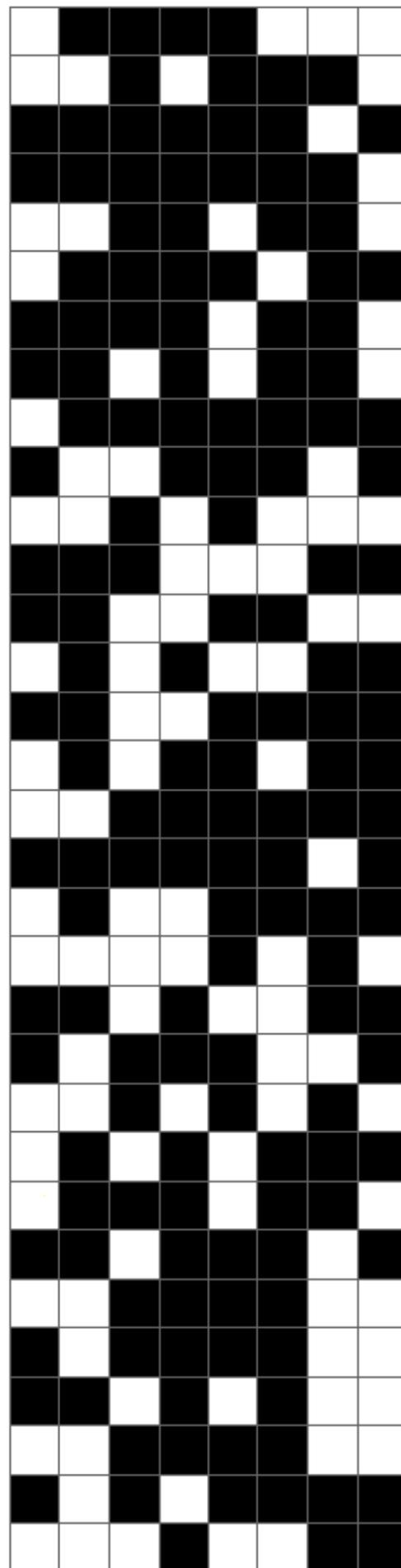
– – генеральной 6

– – нечетный 13

– – четный 13

Область допустимых значений критерия 38

Остаточная дисперсия 63



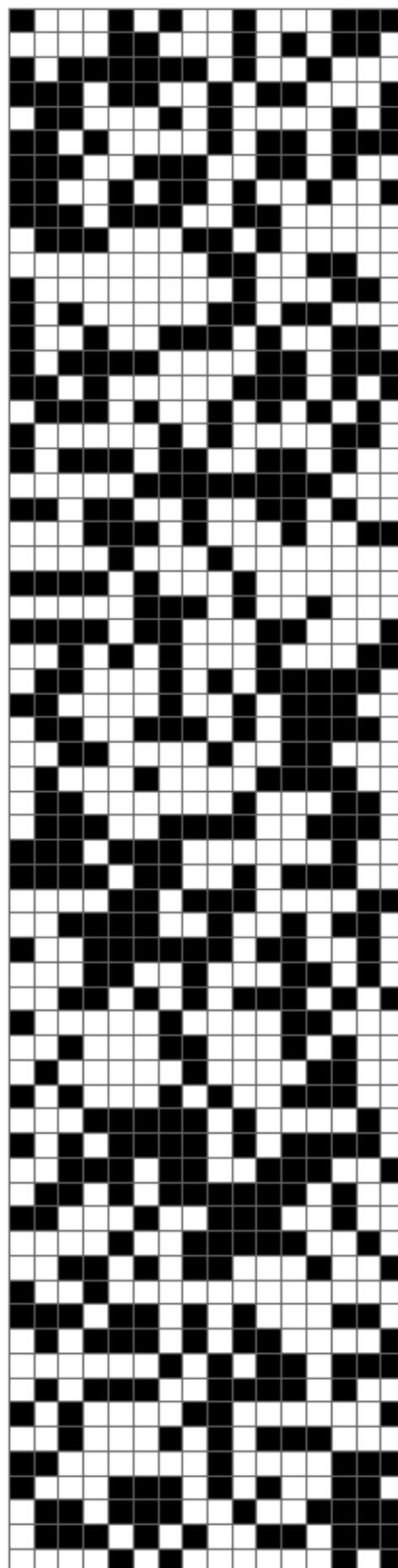
- сумма квадратов 63
- Отклонение ср. кв. выборочное 12
- Относительная погрешность уравнения 63
- Оценка генеральной средней 26
 - надежности 125
 - несмещенная 23
 - статистическая смещенная 36
 - – состоятельная 36
 - – эффективная 36
 - точечная 36

П

- Плотность вероятности 147
- Погрешность 59
- Показатель тесноты связи 116
- Полигон частот 8
 - частот относительных 8
- Предикатор 115
- Признак 6
 - варьирования 8
 - результативный 51, 115
- Признаки факторные 51, 115

Р

- Размах варьирования 6
- Размерность фрактальная 158
- Распределение статистическое 9
 - биномиальное 41, 148
 - – отрицательное 148
 - Вейбулла 148
 - геометрическое 147
 - гипергеометрическое 147
 - Коши 147
 - нормальное 41, 148
 - – логарифмически 147
 - логистическое 148
 - Пуассона 41, 148
 - равномерное 148
 - Стьюдента 58, 148
 - Фишера 147
 - экспоненциальное 41, 147
 - β 147



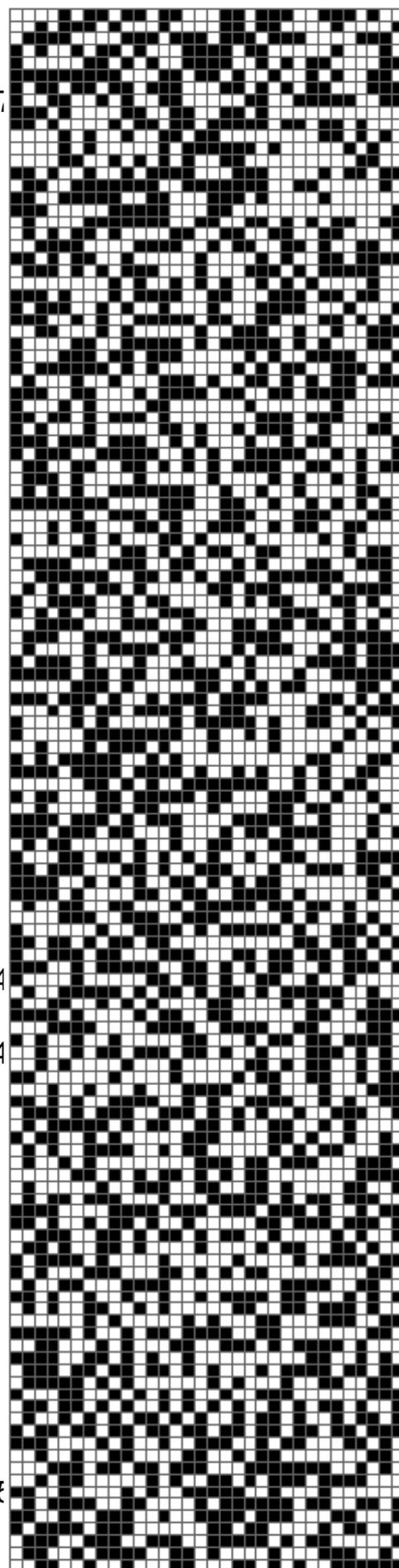
- γ 147
- χ^2 147
- Рассеяние относительно линии регрессии 57
- результативного признака 59
- Регрессии уравнение 48
- Регрессионная модель 60
- Регрессия гиперболическая 96, 97
- линейная 52
- логарифмическая 98
- множественная 115
- параболическая 92
- экспоненциальная 93
- Регрессор 115
- Ряд вариационный дискретный 8
- – интервальный 6
- – непрерывный 17

С

- Связь криволинейная 95
- Скалограмма 157
- Скелетон 158
- СЛАУ 53
- Согласованность распределений 49
- Содержательный анализ взаимосвязей 60
- Спектр мощности 158
- Способ выравнивания 93
- Среднее квадратическое отклонение 13, 144
- Средняя ошибка аппроксимации 118
- Стандартная ошибка при лин. регрессии 144
- – уравнения регрессии 63, 144
- Статистики 8
- Статистическая гипотеза 37
- – нулевая 37
- Статистический критерий 37, 39
- Степень влияния фактора 116
- Схема обработки данных 39

Т

- Теория корреляций 52
- Теснота связи 52
- Тип переменной количественного признака 5
- Треугольник достоверности 159



У

Уровень доверия 17

– значимости 38

Условная средняя распределения признака 52

Усредненное влияние 54

Ф

Форма корреляционной зависимости 57

Фрактальность 159

Функция Ван дер Вардена 155

– Вейерштрасса 159

– распределения 9, 147

– – интегральная 9

– – эмпирическая 9

– самоаффинная 158

– самоподобная 155

– эмпирическая 9, 17

Х

Характер корреляционной зависимости 52

Характеристики выборочные числовые 11

хи-квадрат 39

Ч

Частость 7

Частота 6

– значения признака 100

– непрерывного вариационного ряда 16

– появления признака 52

Частоты накопленные 9

– относительные 8

– теоретические 39

– эмпирические 6

Число степеней свободы 39

Ш

Шаг 13

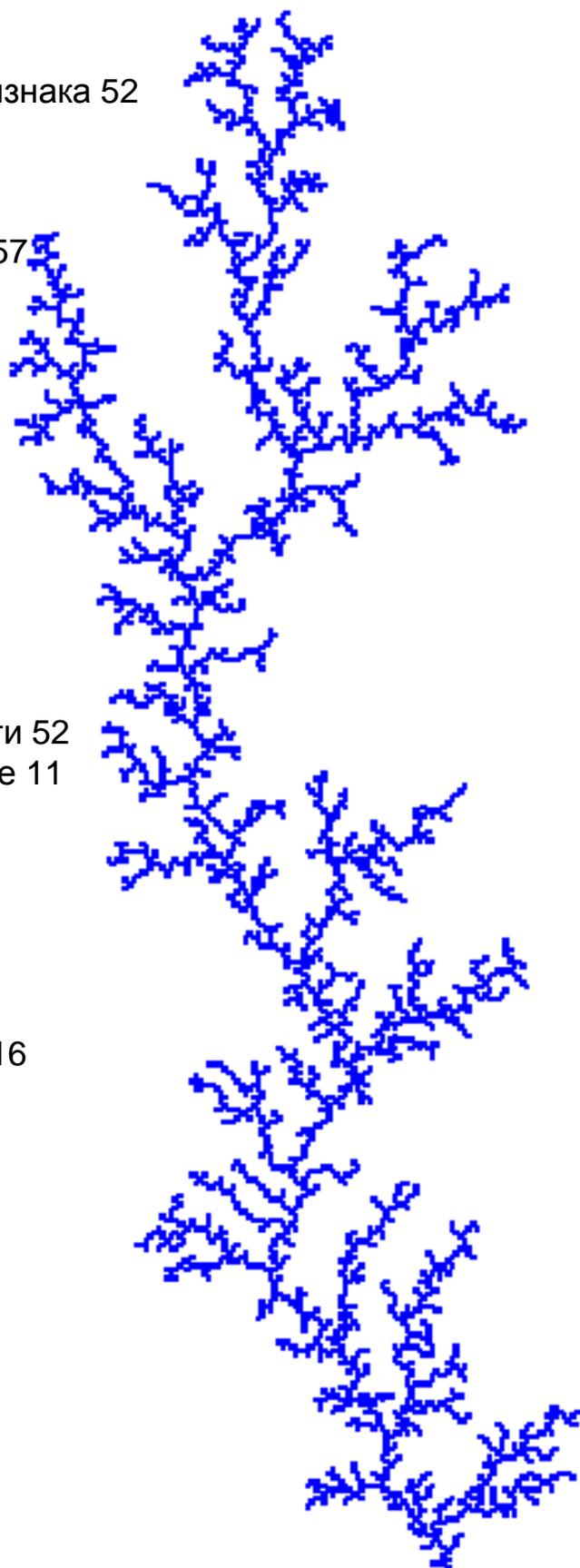
Э

Экспериментальные данные 115

Эксцесс 14, 144

Эмпирические данные 115

Энергетический спектр 157



ОБОЗНАЧЕНИЯ

- a — математическое ожидание 10
 a — сжатие базисного вейвлета вдоль оси времени 152
 b — сдвиг базисного вейвлета вдоль оси времени 152
 c — статистика критерия Ястремского 40
 $\det C$ — детерминант, определитель матрицы C 113
 h — шаг, длина частичного интервала 6
 i — индекс корреляции 97
 k — количество частичных интервалов 6
 k — число степеней свободы (параметр распределения Пирсона) 37
 l — число вариантов дискретного вариационного ряда 41
 n_i — частота 6
 n'_i — округленные частоты 34
 m_3, m_4 — условные центральные моменты 3-го, 4-го порядка 14
 n — объем выборки 6
 n_{ij} — частота совместного появления пары (x_i, y_j) 48
 n_{x_i} — частота появления признака X 48
 p — число факторных признаков 118
 q — табличное значение 16
 r — сумма числа параметров теоретического закона распределения 37
 r — коэффициент линейной корреляции 51
 \hat{r} — оценка для коэффициента r линейной корреляции 55
 r_{xy} — коэффициент парной корреляции 111
 $r_{xy(z)}$ — коэффициент частной корреляции 111
 s — число интервалов вариационного ряда 37
 t_p — расчетная статистика 54
 $t_{кр}$ — табличная статистика распределения Стьюдента 54
 t_γ — табличное значение 16
 t_γ — табличная статистика функции Лапласа 55
 u_i — условная варианта 12
 \bar{x} — выборочная средняя 10
 x_i — значение признака, варианта 6
 x_{\max} — наибольшая варианта 6
 x_{\min} — наименьшая варианта 6
 \bar{x}_y — средняя произведения значений признаков X, Y 51
 y_i — выровненные частоты 34
 \bar{y}_{x_i} — условное среднее признака Y 48
- A_s — асимметрия 13
 C — ложный нуль 12
 D — фрактальная размерность 154

- D_T — генеральная дисперсия
 D_u — остаточная дисперсия 59
 D_n^+, D_n^- — статистики 40
 $E(a)$ — скалограмма 152
 $F_B(x)$ — функция распределения выборочная 8
 F_H — вычисленная статистика критерия Фишера — Снедекора 57
 F_T — табличная статистика критерия Фишера — Снедекора 58
 E_x — эксцесс 13
 $F(x)$ — интегральная функция распределения 9
 H_0 — нулевая гипотеза 37
 J — статистика критерия Ястремского 40
 $K(\lambda)$ — функция вероятности 39
 K_3 — коэффициент эластичности 50
 $L^2 \textcircled{\mathbb{R}}$ — линейное пространство функций 151
 M — накопленные эмпирические частоты 39
 M' — накопленные теоретические частоты 39
 $M_e X$ — медиана 11
 $M_o X$ — мода 11
 M_i^* — условный начальный i -й момент 13
 $M(X)$ — генеральная средняя 15
 $P(\omega)$ — спектр мощности 153
 Q — сумма квадратов отклонений 58
 Q_e — остаточная сумма квадратов 58
 Q_R — сумма квадратов отклонений условных средних 58
 R — размах варьирования признака 6
 \mathbb{R} — множество всех действительных чисел
 R^2 — коэффициент детерминации 56
 $R_{1.2L.k}$ — множественный (совокупный) коэффициент корреляции 115
 $\hat{R}_{1.2L.k}$ — скорректированный совокупный коэффициент корреляции 116
 S, \hat{S} — выборочные средние квадратические отклонения 11
 S_{A_s} — среднее квадратическое асимметрии 41
 S_{E_x} — среднее квадратическое эксцесса 41
 S_{a_r} — стандартные ошибки коэффициентов a_r уравнения линейной регрессии 59, 113
 S_R — среднеквадратическая ошибка R 116
 S_u — среднее квадратическое отклонение условных вариантов u_i 52
 S_x — выборочное среднее квадратическое отклонение признака X 51
 \hat{S}^2 — исправленная выборочная дисперсия
 S^2 — выборочная дисперсия 10

S_{yx}^2 — средний квадрат отклонений фактических значений y от вычисленных по уравнению регрессии 97

S_y^2 — средний квадрат отклонений опытных значений y от их средней 97

\hat{S}^2 — исправленная дисперсия 11

\hat{S}_Y^2 — общая дисперсия результативного признака 115

$\hat{S}_{1,2L.k}^2$ — остаточная дисперсия 115

$\hat{S}_{Y.1.2L.k}^2$ — средний квадрат отклонений фактических значений признака Y от значений, вычисленных с учетом всех факторных признаков 117

$\hat{S}_{Y.2.3L.k}^2$ — средний квадрат отклонений фактических значений признака Y от значений, вычисленных с учетом всех факторных признаков, кроме первого 117

V — коэффициент вариации 14

W_i — накопленная частота 7

$[W^\Psi f](a, b)$ — одномерное непрерывное вейвлет-преобразование 151

X — признак, производственный показатель 6

α — уровень значимости 35

γ — уровень доверия, надежность 15

δ — относительная погрешность 59

ε — средняя ошибка аппроксимации 118

η_{xy} — корреляционное отношение 96

λ — параметр распределения Пуассона 38

λ — статистика критерия Колмогорова 39

λ_α — критическое значение статистики критерия Колмогорова 40

σ — генеральная средняя квадратическая 15

σ^2 — дисперсия 10

σ_r — средняя квадратическая ошибка 55

σ_y — стандартная ошибка уравнения регрессии 59

$\varphi(u_i)$ — табличное значение функции 34

χ^2 — статистика критерия Пирсона 36

$\chi_{кр}^2$ — табличное значение 37

χ_0^2 — вычисленное значение 37

ψ — базисный вейвлет 151

ω — частота 152

ω_i — относительная частота 7

$\Delta^2 y$ — конечные разности второго порядка 95

$\Phi(x)$ — функция Лапласа 55

БИБЛИОГРАФИЧЕСКИЙ СЛОВАРЬ

БЕРНУЛЛИ Якоб (Jacob Bernoulli) (27.12.1654, Базель,— 16.8.1705, там же) — проф. математики Базельского ун-та (с 1687). Ознакомившись с первым мемуаром Г. Лейбница по дифференциальному исчислению (1684) в том же году, Бернулли вскоре блестяще применил новые идеи к изучению свойств ряда кривых. Совместно с братом Иоганном положил начало вариационному исчислению. При этом особое значение имели выдвигнутая и частью решённая Якобом Бернулли изопериметрическая задача и найденное им решение поставленной Иоганном Бернулли задачи о брахистохроне. В «Искусстве предположений» («Ars conjectandi», Basileae, 1713) доказал так называемую теорему Бернулли — важный частный случай закона больших чисел. В связи с вычислением суммы одинаковых степеней натуральных чисел открыл числа Бернулли. Работал также в области физики (определение центра качения тел и сопротивления тел различной формы, движущихся в жидкости).



БЕРНУЛЛИ Иоганн (Johann Bernoulli) (27.7.1667, Базель,— 1.1. 1748, там же) — младший брат Якоба Б., иностранный почётный чл. Петербургской АН (1725), проф. математики Гронингенского (с 1695) и Базельского (с 1705) университетов. Был деятельным сотрудником Г. Лейбница в разработке дифференциального и интегрального исчислений, в области которых им был сделан ряд открытий. Дал первое систематическое изложение дифференциального и интегрального исчислений, продвинул разработку методов решения обыкновенных дифференциальных уравнений, поставил классическую задачу о



геодезических линиях и нашёл характерное геометрическое свойство этих линий, а позднее вывел их дифференциальное уравнение. Ожесточённый спор о решении вариационных задач, разгоревшийся между Иоганном и Якобом Бернулли, в некоторой мере способствовал постановке новых проблем в этой области. Иоганну Бернулли принадлежат также исследования по механике: теория удара, движение тел в сопротивляющейся среде, учение о живой силе и др.

ГАУСС Карл Фридрих (Gauss Carl Friedrich) (30.4.1777, Брауншвейг,— 23.2.1855, Гёттинген) — немецкий математик, внёсший фундаментальный вклад также в астрономию и геодезию, иностранный чл.-корр. (1802) и

иностраннный почётный чл. (1824) Петербургской АН. Родился в семье водопроводчика. Учился в Гёттингенском ун-те (1795—98). В 1799 получил доцентуру в Брауншвейге, в 1807 — кафедру математики и астрономии в Гёттингенском ун-те, с которой была также связана должность директора Гёттингенской астрономии, обсерватории. На этом посту Гаусс оставался до конца жизни.

Отличительными чертами творчества Гаусса являются глубокая органичная связь в его исследованиях между теоретической и прикладной математикой, необычайная широта проблематики.

Работы Гаусса оказали большое влияние на развитие алгебры, теории чисел, дифференциальной геометрии, теории тяготения, классической теории электричества и магнетизма, геодезии, целых отраслей теоретической астрономии. Во многих областях математики труды Гаусса содействовали повышению требований к логической отчётливости доказательств, однако сам Гаусс оставался в стороне от работ по строгому обоснованию математического анализа, которые проводил в его время О. Коши.

Первое крупное сочинение Гаусса по теории чисел и высшей алгебре («*Арифметические исследования*», 1801) во многом предопределило дальнейшее развитие этих дисциплин. Гаусс даёт здесь обстоятельную теорию квадратичных вычетов, первое доказательство квадратичного закона взаимности — одной из центральных теорем теории чисел. Гаусс даёт также новое подробное изложение арифметической теории квадратичных форм, до того построенной Ж. Лагранжем, в частности тщательную разработку теории композиции классов таких форм. В конце книги излагается теория уравнений деления круга (т. е. уравнений $x^n - 1 = 0$), которая во многом была прообразом теории Галуа. Помимо общих методов решения этих уравнений Гаусс установил связь между ними и построением правильных многоугольников. Он, впервые после древнегреческих учёных, сделал значительный шаг вперёд в этом вопросе, а именно: Гаусс нашёл все те значения n , для которых правильный n -угольник можно построить циркулем и линейкой; в частности, решив уравнение $x^{17} - 1 = 0$, он дал построение правильного 17-угольника при помощи циркуля и линейки. Гаусс придавал этому открытию очень большое значение и завещал выгравировать правильный 17-угольник, вписанный в круг, на своём надгробном памятнике, что и было исполнено. Алгебраические интересы Гаусса связаны с основной теоремой алгебры; он дал несколько её доказательств — первое из них в 1799.

Астрономические работы Гаусса (1800—20) в основном связаны с решением проблемы определения орбит малых планет и исследованием их возмущений. Гаусс как астроном получил широкую известность после разработки метода вычисления эллиптических орбит планет по трём наблю-



дениям, успешно применённого им к первым открытым малым планетам Церера (1801) и Паллада (1802). Результаты исследований по вычислению орбит Гаусс опубликовал в сочинении «*Теория движения небесных тел*» (1809). В 1794—95 открыл и в 1821—23 разработал основной математический метод обработки неравноценных наблюдательных данных (метод наименьших квадратов). В связи с астрономическими вычислениями, основанными на разложении интегралов соответствующих дифференциальных уравнений в бесконечные ряды, Гаусс занялся исследованием вопроса о сходимости бесконечных рядов [в работе, посвященной изучению гипергеометрического ряда (1812)].

Работы Гаусса по геодезии (1820—30) связаны с поручением провести геодезическую съёмку и составить детальную карту Ганноверского королевства; Гаусс организовал измерение дуги меридиана Гёттинген — Альтона, в результате теоретической разработки проблемы создал основы высшей геодезии («*Исследования о предметах высшей геодезии*», 1842—47). Для оптической сигнализации Гаусс изобрёл специальный прибор — гелиотроп. Изучение формы земной поверхности потребовало углублённого общего геометрического метода для исследования поверхностей. Выдвинутые Гауссом в этой области идеи получили выражение в соч. «*Общие исследования о кривых поверхностях*» (1827). Руководящая мысль этого сочинения заключается в том, что при изучении поверхности как бесконечно тонкой гибкой плёнки основное значение имеет не уравнение поверхности в декартовых координатах, а дифференциальная квадратичная форма, через которую выражается квадрат элемента длины и инвариантами которой являются все собственные свойства поверхности — прежде всего её кривизна в каждой точке. Другими словами, Гаусс предложил рассматривать те свойства поверхности (т. е. *внутренние*), которые не зависят от изгибаний поверхности, не изменяющих длин линий на ней. Созданная таким образом внутренняя геометрия поверхностей послужила образцом для создания n -мерной римановой геометрии.

Исследования Гаусса по теоретической физике (1830—40) являются в значительной мере результатом тесного общения и совместной научной работы с В. Вебером. Вместе с В. Вебером Гаусс создал абсолютную систему электромагнитных единиц и сконструировал в 1833 первый в Германии электромагнитный телеграф. В 1835 Гаусс основал Магнитную обсерваторию при Гёттингенской астрономической обсерватории. В 1838 он издал труд «*Общая теория земного магнетизма*». Небольшое сочинение «*О силах, действующих обратно пропорционально квадрату расстояния*» (1834—40) содержит основы теории потенциала. К теоретической физике примыкают также разработка (1829) Гауссом принципа наименьшего принуждения (принцип Гаусса) и работы по теории капиллярности (1830). К числу физических исследований Гаусса относятся и его «*Диоптрические исследования*» (1840), в которых он заложил основы теории построения изображения в системах линз.

Очень многие исследования Гаусса остались неопубликованными и в виде очерков, незаконченных работ, переписки с друзьями входят в его научное наследие. Вплоть до 2-й мировой войны 1939—45 оно тщательно разрабатывалось Гёттингенским учёным обществом, которое издало 12 томов сочинений Гаусса. Наиболее интересными в этом наследии являются дневник Гаусса и материалы по неевклидовой геометрии и теории эллиптических функций. Дневник содержит 146 записей, относящихся к периоду от 30 марта 1796, когда 19-летний Гаусс отметил открытие построения правильного 17-угольника, по 9 июля 1814. Эти записи дают отчётливую картину творчества Гаусса в первой половине его научной деятельности; они очень кратки, написаны на латинском языке и излагают обычно сущность открытых теорем. Материалы, относящиеся к неевклидовой геометрии, обнаруживают, что Гаусс пришёл к мысли о возможности построения наряду с евклидовой геометрией и геометрии неевклидовой в 1818, но опасение, что эти идеи не будут поняты, было причиной того, что Гаусс их не разрабатывал далее и не опубликовывал. Более того, он категорически запрещал опубликовывать их тем, кого посвящал в свои взгляды. Когда вне всякого отношения к этим попыткам Гаусса неевклидова геометрия была построена и опубликована Н. И. Лобачевским, Гаусс отнёсся к публикациям Н. И. Лобачевского с большим вниманием, был инициатором избрания его чл.-корр. Гёттингенского учёного общества, но своей оценки великого открытия Н. И. Лобачевского по существу не дал. Архивы Гаусса содержат также обильные материалы по теории эллиптических функций и своеобразную их теорию; однако заслуга самостоятельной разработки и публикации теории эллиптических функций принадлежит Карлу Якоби и Нильсу Абелю.

ГНЕДЕНКО Борис Владимирович [р. 19.12.1911 (1.1.1912), Симбирск] — советский математик, акад. АН УССР (1948; чл.-корр. 1945). Окончил Саратовский ун-т (1930), в 1934—45 работал в Московском ун-те (проф. с 1942), в 1945—50 проф. Львовского, в 1950—58 — Киевского и с 1960 — Московского ун-тов. В 1945—60 в Ин-те математики АН УССР (в 1955—58 директор). Основные труды по теории вероятностей, математической статистике, истории математики. Автор известного университетского курса теории вероятностей. Чл. Международного статистического ин-та (1957), Международной ассоциации по применению статистики в физических науках (1959), Американского ин-та математической статистики (1960), почётный чл. Лондонского статистического общества. Награждён орденами Трудового Красного Знамени, Дружбы народов и орденом ГДР «За заслуги перед отечеством».



КОЛМОГОРОВ Андрей Николаевич [12(25).4.1903, Тамбов,— 20.10.1987, Москва] — советский математик, акад. АН СССР (1939) и Академии педагогических наук СССР (1966). Герой Социалистич. Труда (1963). Окончил Московский ун-т (1925), с 1931 проф. там же. Почётный чл. Московского математического общества (1953), в 1964—66 и 1974—83 его президент, редактор журнала «Успехи математических наук» (1946—54 и с 1983), редактор математического отдела 1-го издания Большой советской энциклопедии и член Главной редакции 2-го издания Большой советской энциклопедии и 3-го издания



Малой советской энциклопедии. Научную деятельность начал в области теории функций действительного переменного, где ему принадлежат фундаментальные работы по тригонометрическим рядам, теории меры, теории множеств, теории интеграла, теории приближения функции. В дальнейшем Колмогоров внёс существенный вклад в разработку конструктивной логики, топологии (где им создана теория верхних гомологий), механики (теория турбулентности), теории дифференциальных уравнений, функционального анализа. Основополагающее значение имеют работы Колмогорова в области теории вероятностей, где он совместно с А. Я. Хинчиным начал применять методы теории функций действительного переменного (с 1925). Это позволило Колмогорову решить ряд трудных проблем и построить широко известную систему аксиоматического обоснования теории вероятностей (1933), заложить основы теории марковских случайных процессов с непрерывным временем. Позднее Колмогоров развил (примыкая к исследованиям А. Я. Хинчина) теорию стационарных случайных процессов, процессов со стационарными приращениями, ветвящихся процессов. Колмогоров внёс важный вклад в теорию информации. Ему принадлежат исследования по теории стрельбы, статистическим методам контроля массовой продукции, применениям математических методов в биологии, математической лингвистике. Принимал деятельное участие в разработке вопросов математического образования в средней школе и ун-тах.

Колмогоров — почётный доктор наук Парижского ун-та (1955), иностранный чл. Польской АН (1956), почётный член Королевского статистического общества Великобритании (1956), чл. Международного статистического ин-та (1957), почётный чл. Американской академии искусств и наук в Бостоне (1959), чл. Германской академии естествоиспытателей «Леопольдина» (1959), почётный доктор наук Стокгольмского ун-та (1960), иностранный член Американского философского общества в Филадельфии (1961), почётный доктор наук Индийского статистического ин-та в Калькутте (1962), почётный чл. Американского метеорологического общества (1962), Индийского математического общества (1962), Лондонского мате-

матического общества (1962), иностранный чл. Нидерландской королевской АН (1963), Лондонского королевского общества (1964), почётный чл. Румынской академии (1965; чл.-корр. 1957), Венгерской АН (1965), иностранный чл. Национальной АН США (1967), Парижской АН (1968), АН ГДР (1977), почётный чл. Международной академии истории науки (1977). Ленинская премия (1965), Государственная премия СССР (1941), премия имени П. Л. Чебышева АН СССР (1951), Международная премия фонда Э. Бальзана (1963), Золотая медаль Американского метеорологического общества, медаль имени Гельмгольца АН ГДР (1976), Международная премия Вольфа (1981), Международная премия имени Н. И. Лобачевского АН СССР (1986). Награждён 7 орденами Ленина (1944, 1945, 1953, 1961, 1963, 1973, 1975), орденом Октябрьской Революции (1983), орденом Трудового Красного Знамени (1940), а также венгерским орденом Знамени (1975).

КОШИ Огюстен Луи (Cauchy Augustin Louis) (21.8.1789, Париж,—23.5. 1857, Со) — французский математик, иностранный почётный чл. Петербургской АН (1831), чл. Парижской АН (1816). Окончил Политехническую школу (1807) и Школу мостов и дорог (1810) в Париже. В 1810—13 работал инженером в Шербуре. В 1816—30 преподавал в Политехнической школе и в Коллеж де Франс, с 1848—в Парижском ун-те и в Коллеж де Франс.



Труды Коши относятся к различным областям математики (преимущественно к математическому анализу) и математической физики. Его курсы анализа («Курс анализа», 1821, «Резюме лекций по исчислению бесконечно малых», 1823, «Лекции по приложениям анализа к геометрии», т. 1—2, 1826—28), основанные на систематическом использовании понятия предела, послужили образцом для большинства курсов позднейшего времени. В них он дал определение понятия непрерывности функции, чёткое построение теории сходящихся рядов (признак Коши, критерий Коши), определение интеграла как предела сумм и др. Коши систематически развивал основы теории аналитических функций комплексного переменного (уравнения Коши — Римана); дал выражение аналитической функции в виде интеграла Коши, разложение функции в степенной ряд; разработал теорию вычетов. В области теории дифференциальных уравнений Коши принадлежат: постановка задачи Коши, основные теоремы существования решений и метод интегрирования уравнений с частными производными 1-го порядка. В работах по теории упругости он рассматривал тело как сплошную среду и оперировал напряжением и деформацией, относимой к каждой точке. В работах по оптике Коши дал математическую разработку теории Френеля и теории дисперсии. Коши принадлежат также исследования по теории чисел, алгебре, геометрии.

По политическим убеждениям Коши — ультрароялист, сторонник Бурбонов (после Революции 1830 был в эмиграции до 1838), клерикал.

КРАМЕР Карл Харальд (Cramer Karl Harald) (р. 25.9.1893, Стокгольм) — шведский математик. Проф. математики и математической статистики (с 1929), ректор (1950—58) Стокгольмского ун-та, канцлер шведских унтов (1958—61). Основные труды по теории вероятностей, математической статистике, математической теории страхования.



ЛАПЛАС Пьер Симон (Laplace Pierre Simon) (23.3.1749, Вомон-ан-Ож, Нормандия,— 5.3.1827, Париж) — французский астроном, математик и физик, иностранный почётный член Петербургской АН (1802), член Парижской АН (1785; адъюнкт с 1773), чл. Французской академии (1816). Учился в школе бенедиктинцев, из которой вышел, однако, убеждённым атеистом. В 1766 Л. приехал в Париж, где Ж. Д'Аламбер через пять лет помог ему получить место проф. Военной школы. Деятельно участвовал в реорганизации системы высшего образования во Франции, в создании Нормальной и Политехнической школ. В 1790 Лаплас был назначен председателем Палаты мер и весов, руководил введением в жизнь новой метрической системы мер. С 1795 в составе руководства Бюро долгот.



Научное наследие Лапласа относится к области небесной механики, математики и математической физики. Фундаментальными являются работы Лапласа по дифференциальным уравнениям, в частности по исследованию так называемого уравнения Лапласа, интегрированию методом «каскадов» уравнений с частными производными. Введённые Лапласом сферической функции имеют разнообразные применения. В алгебре Лапласу принадлежит важная теорема о представлении определителей суммой произведений дополнительных миноров (теорема Лапласа).

Для разработки созданной им математической теории вероятностей Лаплас ввёл так называемые производящие функции и широко применял преобразование, носящее его имя (преобразование Лапласа). Теория вероятностей явилась основой для изучения всевозможных статистических закономерностей, в особенности в области естествознания. До него первые шаги в этой области были сделаны Б. Паскалем, П. Ферма, Я. Бернулли и другими. Лаплас привёл их выводы в систему, усовершенствовал методы доказательств, сделав их менее громоздкими; доказал (1812) теорему, носящую его имя (теорема Лапласа), развил теорию ошибок и способ наименьших квадратов, позволяющие находить наиболее вероятнейшие значения

измеренных величин и степень достоверности этих подсчётов. Классический труд Лапласа «*Аналитическая теория вероятностей*» («*Theorie analytique des probabilites*») издавался трижды при его жизни — в 1812, 1814 и 1820; в качестве введения к последним изданиям была помещена работа «*Опыт философии теории вероятностей*» (1814), в которой в популярной форме разъясняются основные положения и значение теории вероятностей.

Вместе с А. Лавуазье в 1779—84 Лаплас занимался физикой, в частности вопросом о скрытой теплоте плавления тел и работами с созданным ими ледяным калориметром. Для измерения линейного расширения тел они впервые применили зрительную трубу; изучали горение водорода в кислороде. Лаплас активно выступал против ошибочной гипотезы о флогистоне. Позднее снова вернулся к физике и математике. Он опубликовал ряд работ по теории капиллярности и установил закон, носящий его имя. В 1809 Лаплас занялся вопросами акустики; вывел формулу для скорости распространения звука в воздухе. Лапласу принадлежит барометрическая формула для вычисления изменения плотности воздуха с высотой над поверхностью Земли, учитывающая влияние влажности воздуха и изменение ускорения свободного падения. Занимался также геодезией.

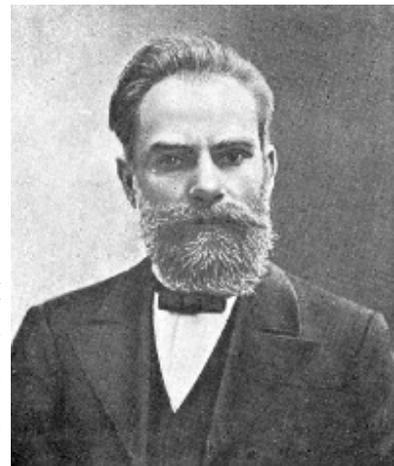
Лаплас развил методы небесной механики и завершил почти всё то, что не удалось его предшественникам в объяснении движения тел Солнечной системы на основе закона всемирного тяготения Ньютона; ему удалось доказать, что закон всемирного тяготения полностью объясняет движение планет, если представить их взаимные возмущения в виде рядов. Он установил также, что эти возмущения носят периодический характер. В 1780 Лаплас предложил новый способ вычисления орбит небесных тел. Исследования Лапласа доказали устойчивость Солнечной системы в течение очень длительного времени. Далее Лаплас пришёл к заключению, что кольцо Сатурна не может быть сплошным, так как в этом случае оно было бы неустойчиво, и предсказал открытие сильного сжатия Сатурна у полюсов. В 1789 Лаплас рассмотрел теорию движения спутников Юпитера под действием взаимных возмущений и притяжения к Солнцу. Он получил полное согласие теории с наблюдениями и установил ряд законов этих движений. Одной из главных заслуг Лапласа было открытие причины ускорения в движении Луны. В 1787 он показал, что средняя скорость движения Луны зависит от эксцентриситета земной орбиты, а последний меняется под действием притяжения планет. Лаплас доказал, что это возмущение не вековое, а долгопериодическое и что впоследствии Луна станет двигаться замедленно. По неравенствам в движении Луны Лаплас определил величину сжатия Земли у полюсов. Ему принадлежит также разработка динамической теории приливов. Небесная механика во многом обязана трудам Лапласа, которые подытожены им в классическом сочинении «Трактат о небесной механике» (т. 1—5, 1798—1825). Космогоническая

гипотеза Лапласа имела огромное философское значение. Она изложена им в приложении к его книге «Изложение системы мира» (т. 1—2, 1796).

По философским взглядам Лаплас примыкал к французским материалистам; известен ответ Лапласа Наполеону I, что в своей теории о происхождении Солнечной системы он не нуждался в гипотезе о существовании бога. Ограниченность механистического материализма Лапласа проявилась в попытке объяснить весь мир, в том числе физиологические, психические и социальные явления, с точки зрения механистического детерминизма. Своё понимание детерминизма Лаплас рассматривал как методологический принцип построения всякой науки. Образец окончательной формы научного познания Лаплас видел в небесной механике. Лапласовский детерминизм стал нарицательным обозначением механистической методологии классической физики. Материалистическое мировоззрение Лапласа, ярко сказавшееся в научных трудах, контрастирует с его политической неустойчивостью. При всяком политическом перевороте Лаплас переходил на сторону победивших; сначала был республиканцем, после прихода к власти Наполеона — министром внутренних дел; затем был назначен членом и вице-председателем сената, при Наполеоне получил титул графа империи, а в 1814 подал свой голос за низложение Наполеона; после реставрации Бурбонов получил пэрство и титул маркиза.

ЛЯПУНОВ Александр Михайлович [25.5 (6.6). 1857, Ярославль,— 3.11 1918, Одесса] — русский математик и механик, академик Петербургской АН (1901; чл.-корр. 1900). Ученик П. Л. Чебышева. Окончил Петербургский университет (1880), с 1885 доцент и 1892 проф. Харьковского ун-та. С 1902 работал в Петербурге.

Родился в Ярославле в семье директора Демидовского лицея известного астронома М.В. Ляпунова. Незадолго до этого его отец оставил ученую деятельность в обсерватории Казанского университета после столкновений с реакционной администрацией университета, установившей полицейский казарменный режим после освобождения Н.И. Лобачевского от должности ректора. В 1864 году отец совсем оставил службу и поселился с семьей в имении жены в Симбирской губернии, где посвятил себя обучению двоих старших сыновей — Александра и Сергея. В кабинете отца братья гусиными перьями выводили между линейками крупные буквы. Отец обладал удивительно быстрой способностью счета, чему обучал детей. Долгие зимние вечера просиживали за черчением географических карт, устраивали игры, состоящие в путешествиях по странам света. В доме была библиотека, богатая сочинениями на русском, немецком и французском языках по математике, астрономии, естественным наукам, философии, истории, этнографии, политической экономии, литературе. После внезапной смерти отца обуче-



ние Александра Михайловича продолжалось в семье дяди Р.М. Сеченова (брат знаменитого физиолога Ивана Михайловича Сеченова), где Александр и его двоюродная сестра, будущая жена Наталья Рафаиловна, готовились по предметам гимназического курса и языкам. В 1870 г. А.М. Ляпунов вместе с матерью и братьями переехал в Нижний Новгород, где был принят сразу в третий класс гимназии.

В старших классах у А.М. Ляпунова появилось стремление к точным наукам, которым он уделял много внимания. Закончив гимназию в 1876 г. с золотой медалью, он поступил на естественное отделение физико-математического факультета Петербургского университета, где слушал лекции по химии профессора Д.И. Менделеева. Вскоре он понял, что имеет большую склонность к математическим наукам, и уже через месяц перешел на математическое отделение университета, где профессорами в то время были П.Л. Чебышев и его ученики А.А. Коркин и Е.И. Золотарев, но продолжал посещать лекции Менделеева.

Первые самостоятельные научные работы А.М. Ляпунов провел под руководством Д.К. Бобылева. Еще будучи студентом четвертого курса, Ляпунов получил золотую медаль за написанную им работу на предложенную факультетом тему по гидростатике. На основании этой работы он написал и опубликовал свои две первые научные статьи «*О равновесии тяжелых тел в тяжелых жидкостях, содержащихся в сосуде определенной формы*» и «*О потенциале гидростатических давлений*», в которых уточнялись условия и приводились новые строгие доказательства ранее неточно обоснованных теорем гидростатики. Первую из этих работ А.М. Ляпунов впоследствии защитил как диссертацию на звание кандидата математических наук. По заявлению профессора механики Д.К. Бобылева А.М. Ляпунова оставляют при университете для подготовки к профессорскому званию.

В 1882 г. А.М. Ляпунов успешно сдал магистерские экзамены и начал работать над другой диссертацией, тему которой ему предложил П.Л. Чебышев — исследование эллипсоидальных форм равновесия вращающейся жидкости. Хотя он ранее предлагал этот вопрос другим ученым (Е.И. Золотареву, С.В. Ковалевской) и знал все трудности, связанные с этим исследованием, но, очевидно, как говорил В.А. Стеклов, «Чебышев уже тогда усматривал из ряда вон выходящие силы в молодом человеке, если рискнул возложить на его плечи такой непосильный труд». Ляпунов работал над поставленной задачей в течение двух лет, но, как он писал: «После нескольких неудачных попыток я должен был отложить решение вопроса на неопределенное время. Но вопрос этот навел меня на другой, именно на вопрос об устойчивости эллипсоидальных форм равновесия, который и составил предмет моей магистерской диссертации».

Эта первая большая работа «*Об устойчивости эллипсоидальных форм равновесия вращающейся жидкости*» сразу обратила на себя внимание математиков, механиков, физиков и астрономов во всем мире. В 1885

г. он блестяще защитил в Петербургском университете эту работу в качестве диссертации на степень магистра прикладной математики. В том же году он был утвержден в звании приват-доцента и получил предложение занять кафедру механики в Харьковском университете, освободившуюся после избрания В.Г. Имшенецкого в члены Академии Наук. До 1890 г. он один вел все преподавание по кафедре механики, что отнимало у него очень много времени. Про прекрасные лекции А.М. Ляпунова в Харьковском университете можно судить по словам его ученика академика В.А. Стеклова: «... в аудиторию вместе со старым деканом профессором Леваковским, которого уважали все студенты, вошел красавец-мужчина, почти ровесник некоторых наших товарищей, и, после того как ушел декан, начал дрожащим от волнения голосом читать вместо курса динамики систем курс динамики точки, который мы уже прослушали у профессора Деларю. Шел уже 4-й год моего студенчества; в Москве в течение года я слушал таких лекторов, как Давыдов, Цингер, Столетов, Орлов; два года был студентом Харьковского университета; курс механики мне уже был знаком. Но с самого начала лекции я услышал то, чего раньше не слышал и не встречал ни в одном из известных мне пособий. И вся недружелюбность курса разлетелась прахом. Силой своего таланта, волшебству которого в большинстве случаев неосознанно поддается молодежь, Александр Михайлович, сам не зная того, покорила за один час враждебно настроенную аудиторию. С этого же дня Александр Михайлович занял особое место в глазах студентов: к нему стали относиться с исключительным уважением. Большинство, которым не были чужими интересы науки, стали напрягать все силы, чтобы хоть немного приблизиться к той высоте, к которой вел Александр Михайлович своих слушателей. Появился особый стыд перед ним за свое незнание, большинство не решались даже начать говорить с ним только из боязни показать перед ним свое незнание».

В Харьковском университете А.М. Ляпунов читал различные общие и специальные курсы теоретической механики, интегрирования дифференциальных уравнений, теории вероятностей. Эти курсы не были опубликованы и остались в литографированном виде. По механике А.М. Ляпунов прочитал шесть курсов: кинематику, динамику материальной точки, динамику систем материальных точек, теорию притяжения, теорию деформированных тел и гидростатику. Кроме Харьковского университета, А.М. Ляпунов читал аналитическую механику в Харьковском технологическом институте (1887—1893 гг.).

Начиная с 1888 г. А.М. Ляпунов опубликовал ряд работ, посвященных устойчивости движения механических систем с конечным числом степеней свободы. В 1892 г. он защитил на эту тему докторскую диссертацию, одним из его оппонентов был Н.Е. Жуковский, защитивший за десять лет до этого докторскую диссертацию по этому же вопросу. К харьковскому периоду жизни А.М. Ляпунова относятся его исследования по теории потенциала и по движению твердого тела в жидкости, которые тесно пере-

плетаются с исследованиями его ученика, а в дальнейшем соратника В.А. Стеклова. С 1899 по 1902 г. Ляпунов состоял председателем Харьковского математического общества и редактором его «Сообщений». Все свои работы этого периода, посвященные главным образом теории потенциала и теории вероятностей, он докладывал на заседаниях Общества. Занятие теорией вероятностей было лишь кратким эпизодом в научной работе Ляпунова, тем не менее, и в этой области он добился результатов фундаментального значения. Им дано доказательство в весьма широких условиях центральной предельной теоремы, высказанной, но не доказанной полностью П.Л. Чебышевым; примененный в этом доказательстве метод является ныне одним из основных в теории вероятностей.

В 1900 г. А.М. Ляпунов был избран членом-корреспондентом Академии наук, а в 1901 г. — ординарным академиком по кафедре прикладной математики, остававшейся вакантной в течение семи лет после смерти П.Л. Чебышева. В 1902 г. Ляпунов переехал в Петербург и целиком отдался научной работе — исследованию фигур небесных тел, т.е. исследованию форм равновесия равномерно вращающейся жидкости. Этими исследованиями он вписал славную страницу в историю мировой науки и показал ошибочность ряда результатов первоклассных зарубежных ученых. В 1908 г. Ляпунов был участником IV Международного математического конгресса в Риме. Он принимал участие в издании полного собрания сочинений Л. Эйлера и был одним из редакторов восемнадцатого и девятнадцатого математических томов этого собрания.

Ляпунов создал современную строгую теорию устойчивости равновесия и движения механических систем, определяемых конечным числом параметров. С математической стороны этот вопрос сводится к исследованию предельного поведения решений систем обыкновенных дифференциальных уравнений при стремлении независимого переменного к бесконечности. Устойчивость определялась Ляпуновым по отношению к возмущениям начальных данных движения. До работ Ляпунова вопросы об устойчивости обычно решались по первому приближению, т. е. путём отбрасывания всех нелинейных членов уравнений, причём не выяснялась законность такой линеаризации уравнений движения. Выдающейся заслугой Ляпунова является построение общего метода для решения задач об устойчивости. Основным трудом в этой области до сих пор является докторская диссертация Ляпунов *«Общая задача об устойчивости движения»* (1892). В этой работе определяются основные понятия теории устойчивости, указываются случаи, когда рассмотрение линейных уравнений первого приближения дает решение вопроса об устойчивости, и проводится подробное исследование некоторых важных случаев, когда первое приближение не даёт решения вопроса об устойчивости. Диссертация и последующие работы Ляпунова в рассматриваемой области содержат целый ряд фундаментальных результатов в теории обыкновенных дифференциальных уравнений как линейных, так и нелинейных. Все работы по теории

устойчивости движения отечественных и зарубежных учёных, выполненные после Ляпунова, основаны на его идеях и методах.

Большой цикл исследований Ляпунова посвящен теории фигур равновесия равномерно вращающейся жидкости, частицы которой взаимно притягиваются по закону всемирного тяготения. До Ляпунова были установлены для однородной жидкости эллипсоидальные фигуры равновесия. Ляпунов впервые доказал существование фигур равновесия однородной и слабо неоднородной жидкости, близких к эллипсоидальным. Он установил, что от некоторых эллипсоидальных фигур равновесия ответвляются близкие к ним неэллипсоидальные фигуры равновесия однородной жидкости, а от других эллипсоидальных фигур равновесия ответвляются фигуры равновесия слабо неоднородной жидкости. Ляпунов разрешил также задачу, предложенную ему ещё в начале его научной деятельности П. Л. Чебышёвым, о возможности ответвления от эллипсоидальной фигуры равновесия с наибольшей (возможной для эллипсоидов) угловой скоростью неэллипсоидальных фигур равновесия. Ответ получился отрицательным. Ляпунов впервые строго доказал существование близких к сфере фигур равновесия медленно вращающейся неоднородной жидкости при весьма общих предположениях об изменении плотности с глубиной. Ляпунов занимался также исследованием устойчивости как эллипсоидальных фигур, так и открытых им новых фигур для случая однородной жидкости. Сама постановка вопроса об устойчивости для сплошной среды (жидкость) до работ Ляпунова была неясной. Он впервые строго поставил вопрос и с помощью тонкого математического анализа провел исследование устойчивости фигур равновесия. В частности, он доказал неустойчивость грушевидных фигур равновесия. Цикл работ Ляпунова по фигурам равновесия вращающейся жидкости и устойчивости этих фигур занимает центральное место во всей теории фигур равновесия.

Небольшим по объёму, но весьма важным для дальнейшего развития науки был цикл работ Ляпунова по некоторым вопросам математической физики. Среди работ этого цикла основное значение имеет его труд «*О некоторых вопросах, связанных с задачей Дирихле*» (1898). Эта работа основана на исследовании свойств потенциала от зарядов и диполей, непрерывно распределённых по некоторой поверхности. Наиболее существенным является исследование потенциала двойного слоя (случай диполей). Далее Ляпунов получил важные результаты, касающиеся поведения производных решения задачи Дирихле при приближении к поверхности, на которой задано граничное условие. На этой основе им впервые были доказаны симметрия функции Грина для задачи Дирихле и формула, дающая решение этой задачи в виде интеграла по поверхности от произведения функции, входящей в граничное условие, на нормальную производную функции Грина. При всех этих условиях Ляпунов налагает на граничную поверхность некоторые ограничения; поверхности, удовлетворяющие им, называются теперь поверхностями Ляпунова.

В теории вероятностей Ляпунов предложил новый метод исследования (метод «характеристических функций»), замечательный по своей общности и плодотворности; обобщая исследования П. Л. Чебышева и А. А. Маркова, Ляпунов доказал центральную предельную теорему (теорему Ляпунова) теории вероятностей при значительно более общих условиях, чем его предшественники. В целом ряде работ Ляпунова содержится большое число принципиально новых понятий математического анализа. Язык и рассуждения Ляпунова отличаются большой точностью, и высказываемое мнение о трудности чтения его работ очень часто объясняется только особой сложностью рассматриваемых проблем. Все исследования Ляпунова являются источником новых работ во многих направлениях математики. Ляпунов — иностранный чл.-корр. Парижской АН (1916).

В конце июня 1917 г. Ляпунов вместе с женой, у которой обострился туберкулез легких, уехал в Одессу к брату Борису Михайловичу. К концу лета состояние здоровья Наталии Рафаиловны приняло угрожающий характер. Сознание близости смерти жены, которая угасала с каждым днем, надвигающаяся слепота, трудные материальные условия жизни — все это явилось причиной чрезвычайно мрачного состояния духа А.М. Ляпунова. Несмотря на это, он нашел в себе силы в сентябре 1918 г. по приглашению физико-математического факультета Одесского (Новороссийского) университета начать чтение курса «*О форме небесных тел*», посвященного изложению его последних работ. 31 октября 1918 г. Н.Р. Ляпунова скончалась. В тот же день А.М. Ляпунов выстрелил в себя и, не приходя в сознание, скончался 3 ноября в 5 часов дня.

Научные заслуги А.М. Ляпунова были признаны всем миром: он состоял почетным членом Петербургского, Харьковского и Казанского университетов, почетным членом Харьковского математического общества, иностранным членом Академии в Риме, членом-корреспондентом Парижской академии наук.

В 1969 в АН СССР учреждена Золотая медаль имени А. М. Ляпунова.

НЕЙМАН Ежи (Юрий) (Neuman Jerzy) (16.4.1894, Бендеры,—5.8.1981, Беркли) — американский математик и статистик, чл. Национальной АН США (1963), с 1938 проф. Калифорнийского ун-та в Беркли. Основные труды по математической статистике (в частности, по теории статистических выводов) и теории вероятностей.



ПИРСОН Карл (Чарлз) [Pearson Karl Charles] (27.3.1857, Лондон,— 27.4.1936, там же) — английский математик, биолог, философ, член Лондонского королевского общества (189R), с 1884 профессор Лондонского университета. Основные труды по математической статистике (кривые Пирсона, распределение Пирсона). Разработал теорию корреляции, тесты математической статистики и критерии согласия.



ПАСКАЛЬ Блез (Pascal Blaise) (19.6.1623, Клермон-Ферран,— 19.8.1662, Париж) — французский философ, писатель, математик и физик. Родился в семье высокообразованного юриста, занимавшегося математикой. Рано проявил выдающиеся математические способности, войдя в историю науки как классический пример отроческой гениальности.

Первый математический трактат Паскаля «*Опыт о конических сечениях*» («*Essay pour les coniques*», Р., 1640) являлся развитием трудов Ж. Дезарга, содержал одну из основных теорем проективной геометрии — теорему Паскаля. В 1641 (по другим сведениям, в 1642) Паскаль сконструировал суммирующую машину. К 1654 закончил ряд работ по арифметике, теории чисел, алгебре и теории вероятностей (опубликовано в 1665). Круг математических интересов Паскаля был весьма разнообразен. Паскаль нашёл общий алгоритм для нахождения признаков делимости



любого целого числа на любое другое целое число [трактат «*О характере делимости чисел*» («*Caracteres de divisibilite des nombres,...*»)], способ вычисления биномиальных коэффициентов (треугольник Паскаля), сформулировал ряд основных положений элементарной теории вероятностей [«*Трактат об арифметическом треугольнике*» («*Traite du triangle arithmetique*», Р., 1665) и переписка с П. Ферма]. В этих работах Паскаль впервые точно определил и применил для доказательства метод математической индукции. Труды Паскаля, содержащие изложенный в геометрической форме интегральный метод решения ряда задач на вычисление площадей фигур, объёмов и площадей поверхностей тел, а также других задач, связанных с циклоидой, явились существенным шагом в развитии анализа бесконечно малых.

Вместе с Г. Галилеем и С. Стевином Паскаль является основоположником классической гидростатики: он установил её основной закон — так называемый закон Паскаля, принцип действия гидравлического пресса,

указал на общность основных законов равновесия жидкостей и газов. Опыт, проведённый под руководством Паскаля (1648), подтвердил предположение Э. Торричелли о существовании атмосферного давления.

Работа Паскаля над проблематикой точных наук в основном относится к 1640—50. Разочаровавшись в «отвлеченности» этих наук, Паскаль обращается к религиозным вопросам и философской антропологии. Сблизившись с представителями янсенизма, он с 1655 ведёт полумонашеский образ жизни в янсенистской обители Пор-Руайяль-де-Шан, вступив в энергичную полемику по вопросам религиозной этики с иезуитами; плодом этой полемики стали *«Письма к провинциалу»* (1657) — шедевр французской сатирической прозы.

Место Паскаля в истории философии определяется тем, что это первый мыслитель, который прошёл через опыт механистического рационализма 17 века и со всей остротой поставил вопрос о границах «научности», указывая при этом на «доводы сердца», отличные от «доводов разума», и тем предвосхищая последующую иррационалистическую тенденцию в философии.

Паскаль сыграл значительную роль в формировании французской классической прозы.

ПУАССОН Симеон Дени (Poisson Simeon Denis) (21.6.1781, Питивье, деп. Луара,— 25.4.1840, Париж) — французский механик, физик, математик, иностранный почётный чл. Петербургской АН (1826), чл. Парижской АН (1812). По окончании в 1800 Политехнич. школы в Париже работал там же (с 1806 проф.), с 1809 проф. Парижского ун-та. Основные труды по теоретической и небесной механике, математике и математической физике. Пуассон впервые записал уравнения аналитической механики в составляющих импульса. В гидромеханике Пуассон обобщил уравнение Навье — Стокса на случай движения сжимаемой вязкой жидкости с учётом теплопередачи. Решил ряд задач теории упругости и обобщил уравнения теории упругости на анизотропные тела. В области небесной механики исследовал устойчивость движения планет Солнечной системы, занимался решением задач о возмущениях планетных орбит и о движении Земли вокруг её центра тяжести. В теории потенциала ввёл так называемое уравнение Пуассона и применил его к решению задач по гравитации и электростатике. Пуассону принадлежат работы по интегральному исчислению (интеграл Пуассона), исчислению конечных разностей (формула суммирования Пуассона), теории дифференциальных уравнений с частными производными, теории вероятностей, где он доказал частный случай закона больших чисел и одну из предельных теорем (теорема Пуассона, рас-



пределение Пуассона). Исследовал вопросы теплопроводности, магнетизма, капиллярности, распространения звуковых волн и баллистики.

РОМАНОВСКИЙ Всеволод Иванович [22.11 (4.12).1879, Алма-Ата,— 6.10. 1954, Ташкент] — советский математик, акад. АН Узбекской ССР (1943). Окончил Петербургский ун-т (1906). В 1911—15 доцент, затем проф. Варшавского ун-та, в 1915—18 проф. Донского (Ростов-на-Дону), с 1918 Среднеазиатского (Ташкент) ун-тов. Основные труды по математической статистике и теории вероятностей. Автор работ в области математического анализа, в частности по интегрированию дифференциальных уравнений с частными производными. Романовский имеет большие заслуги в воспитании национальных научных кадров в Узбекской ССР



Именем В. И. Романовского назван Ин-т математики АН Узбекской ССР

СТЬЮДЕНТ (Student) [псевдоним Уильяма Сили Госсета (William Sealy Gosset)] (13.6.1876, Кантербери,— 16.10.1937, Биконсфилд) — английский математик и статистик. Один из основоположников теории статистических оценок и проверки гипотез. Установил статистическое правило проверки гипотез (критерий Стьюдента), распределение отношения двух независимых случайных величин (распределение Стьюдента).



ФЕРМА Пьер (Fermat Pierre) (17.8. 1601, Бомонде-Ломань,— 12.1.1665, Кастр) — французский математик. По профессии юрист, с 1631 был советником парламента в Тулузе. Автор ряда выдающихся работ, большинство из которых было издано после смерти Ферма его сыном под назв. «*Разные математические сочинения*» («*Varia opera mathematica*», Tolosae, 1679); при жизни Ферма полученные им результаты становились известны учёным благодаря переписке и личному общению.



Ферма является одним из создателей теории чисел, где с его именем связаны две знаменитые теоремы: великая теорема Ферма и малая теорема Ферма. В области геометрии Ферма в более систематической форме, чем Р. Декарт, развил метод координат, дав уравнения

прямой и кривых 2-го порядка и наметив доказательство положения о том, что все кривые 2-го порядка — конические сечения. В области метода бесконечно малых дал общее правило дифференцирования степенной функции, которое распространил на любые рациональные показатели. В подготовке современных методов дифференциального исчисления большое значение имело данное им правило нахождения экстремумов. Доказал в общем виде правило интегрирования степенной функции, полученное в частных случаях уже ранее, также распространил его на случай дробных и отрицательных степеней. В трудах Ферма, таким образом, получили систематическое развитие оба основных процесса метода бесконечно малых, однако он, как и его современники, прошёл мимо связи между операциями дифференцирования и интегрирования. Эта связь была чётко установлена несколько позднее И. Ньютоном и Г. Лейбницем. Работы Ферма оказали большое влияние на дальнейшее развитие математики. В области физики с именем Ферма связано установление вариационного принципа геометрической оптики.

ФИШЕР Роналд Эйлмер (Fisher Ronald Aylmer) (17.2.1890, Лондон,— 29.7.1962, Аделаида, Австралия) — английский статистик и генетик, один из основателей математич. статистики, чл. Лондонского королевского общества (1929). Основные труды по статистике и генетич. теории эволюции. Ввёл понятие «достаточной статистики», построил теорию точечных и интервальных статистич. оценок, разработал методику планирования экспериментов и внёс существенный вклад в создание теории статистич. проверки гипотез.



ЧЕБЫШЁВ Пафнутий Львович [4(16).5.1821, с. Окатово, ныне Калужской обл.,— 26.11(8.12). 1894, Петербург] — русский математик и механик, акад. Петербургской АН (1856; адъюнкт с 1853). Родился в дворянской семье. Первоначальное образование получил дома, шестнадцати лет поступил в Московский ун-т. В 1841 за соч. «*Вычисление корней уравнений*» (тема была предложена факультетом) награждён серебряной медалью. В том же году окончил Московский ун-т. В 1846 при Московском ун-те защитил магистерскую диссертацию «*Опыт элементарного анализа теории вероятностей*» (М., 1845). В 1847 переехал в Петербург, где в том же году при ун-те защитил диссертацию «*Об интегрировании с помощью логарифмов*» на право чтения лекций, был утверждён в звании доцента и начал чтение лекций по алгебре и теории чисел. В 1849 защитил при Петербургском ун-те докторскую диссертацию «*Теория сравнений*»



(СПБ, 1849), удостоенную в том же году Демидовской премии Петербургской АН, и в 1850 стал проф. Петербургского ун-та. Длительное время принимал живое участие в работах артиллерийского отделения военно-учёного комитета и учёного комитета Министерства народного просвещения. В 1882 Чебышёв прекратил чтение лекций в Петербургском ун-те и, выйдя в отставку, целиком отдался научной работе, продолжавшейся до последних дней его жизни. Чебышёв является основателем Петербургской математической школы, наиболее крупными представителями которой были А. Н. Коркин, Е. И. Золотарёв, А. А. Марков, Г.Ф. Вороной, А. М. Ляпунов, В. А. Стеклов, Д. А. Граве.

Характерными чертами творчества Чебышёва являются разнообразие областей исследования, умение находить посредством элементарных средств большие научные результаты и неизменный интерес к вопросам практики. Исследования Чебышёва относятся к теории приближения функций многочленами, интегральному исчислению, теории чисел, теории вероятностей, теории механизмов и многим другим областям математики и смежных областей знания. В каждой из упомянутых областей Чебышёв сумел создать ряд основных, общих методов и выдвинул идеи, наметившие ведущие направления в дальнейшем развитии этих областей. Стремление увязать проблемы математики с принципиальными вопросами естествознания и техники в значительной мере определяют его своеобразие как учёного. Многие его открытия навеяны прикладными интересами. Это неоднократно подчёркивал и сам Чебышёв, говоря, что в создании новых методов исследования «науки находят себе верного руководителя в практике» и что «сами науки развиваются под влиянием ее: она открывает им новые предметы для исследования...» (Полн. собр. соч., т. 5, 1951, с. 150).

К теории вероятностей Чебышёв обращался несколько раз — в начале, середине и конце научного пути («Опыт элементарного анализа теории вероятностей», 1845; «Элементарное доказательство одного общего положения теории вероятностей», 1846; «О средних величинах», 1867; «О двух теоремах относительно вероятностей», 1887). В идейном отношении ему принадлежит заслуга систематичного введения в рассмотрение случайных величин и создание нового приёма доказательства предельных теорем теории вероятностей — *метода моментов*. Им был доказан *закон больших чисел* в весьма общей форме; при этом его доказательство поражает своей простотой и элементарностью. Исследование условий сходимости функций распределения сумм независимых случайных величин к нормальному закону Чебышёв не довел до полного завершения. Однако посредством некоторого дополнения методов Чебышёва это удалось сделать А. А. Маркову. Без строгих выводов Чебышёв наметил также возможность уточнений этой предельной теоремы в форме асимптотических разложений функции распределения суммы независимых слагаемых по степеням $n^{-1/2}$, где n — число слагаемых. Работы Чебышёва по теории вероятно-

стей составляют важный этап в ее развитии; кроме того, они явились базой, на которой выросла русская школа теории вероятностей, вначале состоявшая из непосредственных учеников Чебышёва.

В теории чисел Чебышёв, впервые после Евклида, существенно продвинул изучение вопроса о распределении простых чисел («*Об определении числа простых чисел, не превосходящих данной величины*», 1849; «*О простых числах*», 1852). Чебышёв впервые доказал, что функция $\pi(x)$ — число простых чисел, не превосходящих x , удовлетворяет неравенствам

$$a \frac{x}{\ln x} < \pi(x) < b \frac{x}{\ln x},$$

где $a < 1$ и $b > 1$ — вычисленные Чебышёвым постоянные ($a = 0,921$, $b = 1,06$). Исследование расположения простых чисел в ряду всех целых привело Чебышёв также к исследованию квадратичных форм с положительными определителями. Позднее теория квадратичных форм была предметом исследований ряда его учеников — А. Н. Коркина, Е. И. Золотарёва, А. А. Маркова, Г. Ф. Вороного. Работа Чебышёва «*Об одном арифметическом вопросе*» (1866), посвященная приближению чисел рациональными числами, сыграла важную роль в развитии теории диофантовых приближений. Чебышёв явился создателем новых направлений исследований в теории чисел и новых методов исследований, а также организатором русской школы теории чисел.

Наиболее многочисленны работы Чебышёва в области математического анализа; ему была, в частности, посвящена диссертация на право чтения лекций (1847), в которой Чебышёв исследовал интегрируемость некоторых иррациональных выражений в алгебраических функциях и логарифмах. Интегрированию алгебраических функций Чебышёв посвятил также ряд других своих мемуаров. В одном из них («*Об интегрировании иррациональных дифференциалов*», 1853) была в качестве следствия общих результатов получена известная его теорема об условиях интегрируемости в элементарных функциях дифференциального бинома. Вторым большим направлением исследований Чебышёва по математическому анализу явились его работы по построению общей теории ортогональных многочленов. Толчком к созданию этой теории послужило параболическое интерполирование по способу наименьших квадратов. К этому же кругу идей примыкают исследования Чебышёва по проблеме моментов и по квадратурным формулам. Имея в виду сокращение вычислений, Чебышёв предложил рассматривать квадратурные формулы с равными коэффициентами («*О квадратурах*», 1873). Исследования по квадратурным формулам и по теории интерполирования были тесно связаны с задачами, которые ставились перед Чебышёвым в артиллерийском комитете.

Чебышёв является основоположником *конструктивной теории функций*, основным составляющим элементом которой является теория наилучшего приближения функций. Простейшая постановка задачи Чебы-

шёва такова («Теория механизмов, известных под названием параллелограммов», 1854): дана непрерывная функция $f(x)$; среди всех многочленов степени n найти такой $P(x)$, чтобы в данном промежутке $[a, b]$ выражение $\max |f(x) - P(x)|$ было возможно меньшим. Помимо указанного равномерного наилучшего приближения, Чебышёв рассматривал также квадратичное приближение.

Теория машин и механизмов была одной из тех дисциплин, которыми Чебышёв систематически интересовался всю жизнь. Особенно многочисленны его работы, посвященные синтезу шарнирных механизмов, в частности параллелограмму Уатта («О некотором видоизменении коленчатого параллелограмма Уатта», 1861; «О параллелограммах», 1869; «О параллелограммах, состоящих из трёх каких-либо элементов», 1879, и др.). Большое внимание он уделял конструированию и изготовлению конкретных механизмов. Интересны, в частности, его механизмы с остановками, а также *парадоксальный механизм*, в котором передаточное отношение между ведущим и ведомым валами меняется в зависимости от направления движения. Следует отметить также его стопоходящую машину, имитирующую движение животного при ходьбе, а также автоматический арифмометр; изучение параллелограмма Уатта и стремление усовершенствовать его натолкнуло Чебышёва на постановку задачи о наилучшем приближении функций. Оттолкнувшись же от этой прикладной задачи, он заложил основы большой математической теории, значение которой оказалось несравненно более широким, чем решение первичной практической задачи. К прикладным работам Чебышёва относится также оригинальное исследование «О построении географических карт» (1856), где он поставил задачу найти такую картографическую проекцию данной страны, сохраняющую подобие в малых частях, чтобы наибольшее различие масштабов в разных точках карты было наименьшим. Чебышёв высказал без доказательства мнение, что для этого отображение должно сохранять на границе постоянство масштаба, что впоследствии и было доказано Д. А. Граве.

Чебышёв оставил глубокий и яркий след в развитии математики, дал толчок созданию и развитию многих её разделов как собственными исследованиями, так и постановкой соответствующих вопросов перед молодыми учёными. Так, по его совету А. М. Ляпунов начал цикл исследований по теории фигур равновесия вращающейся жидкости, частицы которой притягиваются по закону всемирного тяготения. Труды Чебышёва ещё при его жизни нашли широкое признание как в России, так и за границей; он был избран иностранным чл. Берлинской АН (1871), Болонской АН (1873), Парижской АН (1874; чл.-корр. 1860), Лондонского королевского общества (1877), Шведской королевской АН (1893) и почётным членом многих других русских и иностранных научных обществ, академий и университетов.

В 1944 в АН СССР учреждена премия имени П. Л. Чебышёва.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	3
ГЛАВА 1. ВАРИАЦИОННЫЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ.....	5
§ 1. Первичная обработка результатов наблюдений	5
§ 2. Расчет выборочных характеристик статистического распределения.....	11
§ 3. Интервальные (доверительные) оценки параметров распределения	16
§ 4. Лабораторная работа № 1. Построение вариационных рядов. Расчет числовых характеристик.....	18
ГЛАВА 2. ПРОВЕРКА СООТВЕТСТВИЯ ВЫБОРКИ НОРМАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ	34
§ 5. Построение кривой нормального распределения по опытным данным.....	34
§ 6. Статистические оценки параметров распределения	36
§ 7. Проверка статистических гипотез.....	37
§ 8. Лабораторная работа № 2. Построение кривой нормального распределения по опытным данным. Проверка гипотезы о нормальном распределении выборки	47
ГЛАВА 3. ПАРНАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ	51
§ 9. Понятие корреляционной зависимости. Задачи теории корреляции.....	51
§ 10. Парная линейная корреляция	53
§ 11. Коэффициент корреляции, его свойства и значимость.....	57
§ 12. Определение надежности (доверительного интервала) коэффициента корреляции	59
§ 13. Коэффициент детерминации	60
§ 14. Проверка адекватности модели	61
§ 15. Оценка величины погрешности.....	63
§ 16. Лабораторная работа № 3. Построение модели линейной корреляции по несгруппированным данным.....	65
§ 17. Лабораторная работа № 4. Построение выборочного уравнения линии регрессии по сгруппированным данным	77
ГЛАВА 4. ПАРНАЯ НЕЛИНЕЙНАЯ КОРРЕЛЯЦИЯ	95
§ 18. Нелинейная корреляционная зависимость.....	95
§ 19. Определение силы криволинейной связи.....	100
§ 20. Проверка адекватности модели	102
§ 21. Лабораторная работа № 5. Построение модельного уравнения нелинейной корреляции.....	104
ГЛАВА 5. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ КОРРЕЛЯЦИЯ.....	114
§ 22. Понятие множественной корреляции	114
§ 23. Измерение тесноты множественной линейной корреляционной связи	119
§ 24. Проверка адекватности модели множественной линейной корреляции.....	122
§ 25. Экономическая интерпретация уравнения регрессии.....	123
§ 26. Лабораторная работа № 6. Построение модели множественной линейной корреляции.....	126
ГЛАВА 6. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ	142
§ 27. Пакет MathCAD и математическая статистика.....	142
§ 28. Исследование временных рядов средствами вейвлет-анализа.....	154
ПРИЛОЖЕНИЯ.....	163
ЛИТЕРАТУРА.....	169
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ.....	170
ОБОЗНАЧЕНИЯ.....	176
БИБЛИОГРАФИЧЕСКИЙ СЛОВАРЬ	179