МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

ГЛАВНОЕ УПРАВЛЕНИЕ ОБРАЗОВАНИЯ, НАУКИ И КАДРОВ

Учреждение образования
«БЕЛОРУССКАЯ ГОСУДАРСТВЕННАЯ
СЕЛЬСКОХОЗЯЙСТВЕННАЯ АКАДЕМИЯ»

ВЫСШАЯ МАТЕМАТИКА

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*Рекомендовано учебно-методическим объединением
по образованию в области сельского хозяйства
в качестве учебно-методического пособия для студентов
учреждений высшего образования, обучающихся по специальности
1-74 01 01 Экономика и организация производства
в отраслях агропромышленного комплекса*

Горки
БГСХА
2019

УДК 51(075)
ББК 22.1я73
В93

Одобрено методической комиссией экономического факультета 25.06.2019 (*протокол № 11*)
и Научно-методическим советом БГСХА 26.06.2019 (*протокол № 10*)

Авторы:
кандидат экономических наук, доцент *Т. Б. Воронкова*;
старший преподаватель *С. Л. Василькова*;
старший преподаватель *Е. Л. Демитриченко*;
кандидат технических наук, доцент *С. В. Курзенков*

Рецензенты:
доктор физико-математических наук, профессор *М. П. Дымков*;
кандидат физико-математических наук, доцент *А. А. Тиунчик*

В93 **Высшая математика. Математическая статистика** : учебно-методическое пособие / Т. Б. Воронкова [и др.]. – Горки : БГСХА, 2019. – 74 с.
ISBN 978-985-467-950-1.

Приведены краткий теоретический материал по теме «Математическая статистика», задания для аудиторных и домашних занятий и список рекомендуемой литературы.

Для студентов учреждений высшего образования, обучающихся по специальности 1-74 01 01 Экономика и организация производства в отраслях агропромышленного комплекса.

УДК 51(075)
ББК 22.1я73

ISBN 978-985-467-950-1

ВВЕДЕНИЕ

Самостоятельная работа студентов играет большую роль в системе высшего образования. Она включает изучение теоретического материала, применение различных подходов и приемов к решению типовых задач по каждой теме, самостоятельное выполнение индивидуальных заданий. Данное методическое пособие содержит необходимый минимум программы курса математической статистики, достаточный для усвоения специальных дисциплин, преподаваемых в сельскохозяйственных вузах.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Дымков, М. П. Теория вероятностей и математическая статистика. (3-й сем). [Электронный ресурс] / Официальный сайт БГЭУ Респ. Беларусь. – Режим доступа: www.bseu.by/hm/учебные материалы/3 семестр/pdf.

2. Лобоцкая, Н. Л. Основы высшей математики / Н. Л. Лобоцкая. – Минск : Выш. шк., 1978.

3. Высшая математика. Общий курс / под ред. проф. А. И. Яблонского. – Минск : Выш. шк., 1993.

4. Гусак, А. А. Высшая математика / А. А. Гусак. – Минск, 2000. – Т. 2.

5. Лихолетов, И. И. Руководство к решению задач по высшей математике, теории вероятностей и математической статистике / И. И. Лихолетов, И. П. Мацкевич. – Минск : Выш. шк., 1976.

6. Булдык, Г. М. Теория вероятностей и математическая статистика / Г. М. Булдык. – Минск : Выш. шк., 1989.

7. Мацкевич, И. П. Теория вероятностей и математическая статистика / И. П. Мацкевич, Г. П. Свирид. – Минск : Выш. шк., 1993.

8. Гурский, Е. И. Сборник задач по теории вероятностей и математической статистике / Е. И. Гурский. – Минск : Выш. шк., 1984.

9. Мацкевич, И. П. Сборник задач и упражнений. Теория вероятностей и математическая статистика / И. П. Мацкевич, Г. П. Свирид, Г. М. Булдык. – Минск : Выш. шк., 1996.

1. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

1.1. Предмет и задачи математической статистики

Математическая статистика изучает методы сбора, систематизации и анализа экспериментальных данных, полученных в результате наблюдения массовых случайных явлений. Исследования математической статистики базируются на понятиях и методах теории вероятностей, но устанавливают свойства математической модели, исходя из наблюдаемых статистических данных.

В зависимости от цели исследования математическая статистика принимает решения в условиях неопределенности. Типичными задачами являются следующие:

- сбор, упорядочивание и представление статистических данных в удобном для анализа виде;
- определение закона распределения изучаемой случайной величины;
- оценивание неизвестных характеристик, параметров распределения изучаемой случайной величины;
- изучение зависимости случайной величины от одной или нескольких величин;
- проверка статистических гипотез, т. е. задача согласования результатов оценивания или других выводов с опытными данными.

1.2. Генеральная и выборочная совокупности

Для изучения количественного признака некоторых объектов или случайных величин проводится статистическое наблюдение. Совокупность результатов изучения всех объектов образует *генеральную совокупность*. Однако провести сплошное наблюдение, как правило, трудно, экономически нецелесообразно или невозможно. В этом случае наилучшим способом исследования является выборочное наблюдение.

Выборочной совокупностью или *выборкой* называется часть объектов, выбранных случайным образом из генеральной совокупности. Под выборкой понимается совокупность независимых случайных величин, одинаково распределенных с изучаемой случайной величиной. Число элементов выборочной совокупности называется *объемом* выборки. Метод статистического исследования объектов на основании выборки называется *выборочным методом.*

Различают выборки *повторные* с возвращением изучаемого объекта в генеральную совокупность и *бесповторные* – без возвращения.

Для получения хороших, подходящих оценок изучаемых характеристик случайной величины выборка должна быть репрезентативной или представительной, т. е. достаточно полно представлять изучаемые признаки генеральной совокупности. Достаточным для этого является соблюдение условия случайного отбора, при котором все элементы генеральной совокупности имеют одинаковую вероятность попасть в выборку. Для обеспечения репрезентативности используют различные методы отбора: простой, типический, механический и серийный отборы.

1.3. Статистические ряды

Для изучения случайной величины X производится выборка объемом n. Допустим значение x_1 встречается в выборке m_1 раз, значение x_2 – m_2 раз и т. д., значение x_k – m_k раз. Значения $x_i, i=\overline{1,\ n}$ называются вариантами. Варианты, записанные по возрастанию, образуют вариационный ряд. Числа m_i, показывающие сколько раз встречаются значения x_i в выборке, называются *частотами*, причем

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

Отношения частот к объему выборки называются *относительными частотами* или *частостями* $w_i = \dfrac{m_i}{n},\ i=\overline{1,\ k},$ причем сумма относительных частот

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

Выборка представляет собой первичный статистический материал, который необходимо представить в виде дискретного или интервального статистических рядов распределения частот или относительных частот. *Дискретным статистическим рядом* называется перечень вариант в виде вариационного ряда и соответствующих им частот или относительных частот. Графическим изображением дискретного ряда служит полигон частот или относительных частот.

Полигоном относительных частот называется ломаная, отрезки которой соединяют точки $(x_i;\ w_i),\ i=\overline{1,\ k}.$ Варианты x_i откладываются на оси $Ox,$ а относительные частоты w_i на оси $Oy.$

Интервальным статистическим рядом называется перечень частичных интервалов и соответствующих им частот или относительных частот. Для построения интервального ряда прежде всего вычисляем количество интервалов, например, по формуле $k = 1 + \log_2 n \approx \approx 1 + 3{,}32 \cdot \lg n.$ Затем размах вариации $R = x_{\max} - x_{\min}$, где $x_{\max}$ – максимальная варианта выборки, а $x_{\min}$ – минимальная. Шаг разбиения вычисляем по формуле $h=\frac{R}{k-1}=\frac{x_{\max}-x_{\min}}{k-1}$. Далее находим границы интервалов $x_0'=x_{\min}-\frac{h}{2}$, $x_1'=x_0'+h$, $x_2'=x_1'+h, \ldots,$ $x_k'=x_{k-1}'+h$ и записываем сами интервалы $\left[x_0';\,x_1'\right)$, $\left[x_1';\,x_2'\right)$, ..., $\left[x_{k-1}';\,x_k'\right]$.

Графическим изображением интервального ряда служит гистограмма и полигон частот или относительных частот. Для построения гистограммы относительных частот откладываем на оси Ox частичные интервалы $\left(x_{i-1}';\,x_i'\right)$, $i=\overline{1,\,k}$ и каждый интервал достраиваем до прямоугольника с высотой, равной относительной частоте w_i или ее плотности $\frac{w_i}{h}$. Полученная ступенчатая фигура называется *гистограммой* относительных частот. Соединив середины верхних оснований прямоугольников, получим полигон относительных частот для интервального статистического ряда.

1.4. Эмпирическая функция распределения

Эмпирической функцией распределения случайной величины X называется функция $F^*(x)$, определяющая для каждого значения x относительную частоту события $\{X<x\}$: $F^*(x)=w(X<x)=\sum\limits_{x_i\,<\,x} w_i$.

Основные свойства эмпирической функции распределения.

1. $F^*(x)\in[0,1]$.

2. $F^*(x)$ является неубывающей функцией.

3. Если $x_{\min}$ – наименьшая варианта, то $F^*(x)=0$ при $x\le x_{\min}$, а $x_{\max}$ – наибольшая, то $F^*(x)=1$ при $x>x_{\max}$.

Эмпирическая функция распределения является оценкой теоретической функции распределения изучаемой случайной величины.

Для приближения к графику теоретической функции распределения нормального закона распределения строят *кумулянту.* Для построения кумулянты на координатной плоскости отмечают точки, абсциссы которых являются границами интервалов или вариантами, а ординаты равны накопленным относительным частотам. Полученные точки соединяют отрезками, полученная ломаная линия является кумулянтой.

Пример. Для заданного дискретного статистического ряда найти и построить график эмпирической функции распределения.

Варианты	–1	–0,5	0	0,5	1	1,5	2	2,5
Частоты	5	8	10	15	22	20	11	9

Решение. Объем выборки $n = 5+8+10+15+22+20+11+9 = 100$. Вычисляем ряд относительных частот $w_i = \dfrac{m_i}{n}$, $i = \overline{1, n}$ и ряд накопленных относительных частот. Получаем таблицу.

№	Варианты	Частоты	Относительные частоты	Накопленные относительные частоты
1	0,5	5	0,05	0,05
2	1	8	0,08	0,05 + 0,08 = 0,13
3	1,5	10	0,1	0,05 + 0,08 + 0,1 = 0,23
4	2	15	0,15	0,05 + 0,08 + 0,1 + 0,15 = 0,38
5	2,5	22	0,23	0,05 + 0,08 + 0,1 + 0,15 + 0,22 = 0,6
6	3	20	0,2	0,05 + 0,08 + 0,1 + 0,15 + 0,22 + 0,2 = 0,8
7	3,5	11	0,11	0,05 + 0,08 + 0,1 + 0,15 + 0,22 + 0,2 + 0,11 = 0,91
8	4	9	0,09	0,05 + 0,08 + 0,1 + 0,15 + 0,22 + 0,2 + 0,11 + 0,09 = 1
∑		100	1	–

Составим эмпирическую функцию распределения:

$$F^*(x) = \sum_{x_i < x} w_i = \begin{cases} 0, \text{ если } x \le 0,5, \\ 0,05, \text{ если } 0,5 < x \le 1, \\ 0,13, \text{ если } 1 < x \le 1,5, \\ 0,23, \text{ если } 1,5 < x \le 2, \\ 0,38, \text{ если } 2 < x \le 2,5, \\ 0,6, \text{ если } 2,5 < x \le 3, \\ 0,8, \text{ если } 3 < x \le 3,5, \\ 0,91, \text{ если } 3,5 < x \le 4, \\ 1, \text{ если } x > 4. \end{cases}$$

График эмпирической функции распределения изображен на рис. 1.

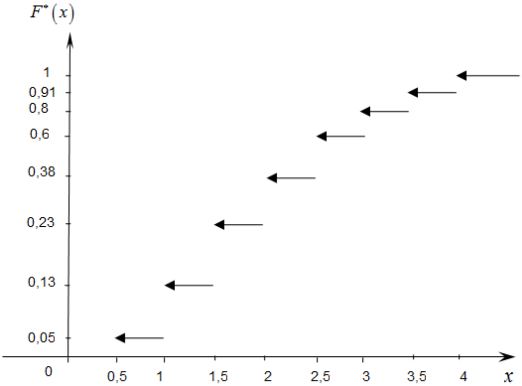


Рис. 1

1.5. Числовые характеристики выборки

К основным числовым характеристикам выборки относятся: среднее значение выборки, мода, медиана, выборочные дисперсия и среднее квадратическое отклонение, коэффициент вариации.

Средним значением выборки или *выборочной средней* называется величина $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ для несгруппированных данных и величина $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i$ для статистических рядов. Причем для интервального статистического ряда $x_i = \frac{x'_{i-1} + x'_i}{2}$, $i = \overline{1, k}$ равно середине интервала.

Модой Mo называется варианта, соответствующая наибольшей частоте. Для интервального статистического ряда мода вычисляется по формуле

$$Mo = x_0 + h\frac{m_0 - m_1}{(m_0 - m_1) + (m_0 - m_2)},$$

где x_0 – начало модального интервала или интервала с наибольшей частотой;

h – ширина интервала;

m_0 – частота модального интервала;

m_1 – частота интервала, предшествующего модальному;

m_2 – частота интервала, следующего за модальным.

Медианой Me называется варианта, разделяющая статистический ряд на две равные по количеству вариант части. Для интервального статистического ряда медиана вычисляется по формуле

$$Me = x_0 + h\frac{n - 2n_1}{2m_e},$$

где x_0 – начало медианного интервала или интервала, накопленная частота которого больше половины объема выборки;

h – ширина интервала;

n_1 – накопленная частота интервала, предшествующего медианному;

m_e – частота медианного интервала.

Выборочной дисперсией называется величина $D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ для несгруппированных данных и величина $D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 m_i$ для статистических рядов. Получим рабочую формулу для вычисления дисперсии:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 m_i = \frac{1}{n}\sum_{i=1}^{k}\left(x_i^2 m_i - 2x_i\overline{x}m_i + \overline{x}^2 m_i\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{k}x_i^2 m_i - 2\overline{x}\frac{1}{n}\sum_{i=1}^{k}x_i m_i + \frac{1}{n}\overline{x}\sum_{i=1}^{k}m_i = \frac{1}{n}\sum_{i=1}^{k}x_i^2 m_i - \left(\overline{x}\right)^2.$$

Выборочным средним квадратическим отклонением называется корень квадратный из выборочной дисперсии $\sigma_{\text{в}} = \sqrt{D_{\text{в}}}$.

Коэффициентом вариации называется величина $v = \frac{\sigma_{\text{в}}}{\overline{x}} \cdot 100\ \%$, определяющая в процентах степень отклонения от выборочной средней для различных статистических рядов.

Пример. Вычислить среднее значение выборки, моду, медиану и выборочные дисперсию и среднее квадратическое отклонение для интервального статистического ряда.

Интер-валы	[0,1; 0,3)	[0,3; 0,5)	[0,5; 0,7)	[0,7; 0,9)	[0,9; 1,1)	[1,1; 1,3)	[1,3; 1,5)	[1,5; 1,7]
Час-тоты	6	10	16	23	19	14	7	5

Решение. Вычислим середины интервалов, накопленные частоты, произведения $x_i m_i$ и $x_i^2 m_i$ для каждого интервала. Получим таблицу с итоговой строкой.

№	Интервалы	Частоты	Накопленные частоты	Середины интервалов	$x_i m_i$	$x_i^2 m_i$
1	[0,1; 0,3)	6	6	0,2	1,2	0,24
2	[0,3; 0,5)	10	16	0,4	4	1,6
3	[0,5; 0,7)	16	32	0,6	9,6	5,76
4	[0,7; 0,9)	23	55	0,8	18,4	14,72
5	[0,9; 1,1)	19	74	1	19	19
6	[1,1; 1,3)	14	88	1,2	16,8	20,16
7	[1,3; 1,5)	7	95	1,4	9,8	13,72
8	[1,5; 1,7]	5	100	1,6	8	12,8
$\sum$	–	100	–	–	86,8	88

Вычислим среднее значение выборки

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \frac{1}{100}\cdot 86,8 = 0,868.$$

Вычислим моду по формуле

$$Mo = x_0 + h\frac{m_0 - m_1}{(m_0 - m_1) + (m_0 - m_2)},$$

где $x_0 = 0,7$ – начало модального интервала или интервала с наибольшей частотой;

$h = 0,2$ – ширина интервала;

$m_0 = 23$ – частота модального интервала;

$m_1 = 16$ – частота интервала, предшествующего модальному;

$m_2 = 19$ – частота интервала, следующего за модальным.

Тогда $Mo = 0,7 + 0,2\cdot\frac{23-16}{(23-16)+(23-19)} \approx 0,827.$

Вычислим медиану по формуле

$$Me = x_0 + h\frac{n - 2n_1}{2m_e},$$

где $x_0 = 0,7$ – начало медианного интервала или интервала, накопленная частота которого больше половины объема выборки;

$h = 0,2$ – ширина интервала;

$n_1 = 32$ – накопленная частота интервала, предшествующего медианному;

$m_e = 23$ – частота медианного интервала.

Тогда $Me = 0,7 + 0,2 \cdot \dfrac{100 - 2 \cdot 32}{2 \cdot 23} \approx 0,857.$

Вычислим выборочные дисперсию и среднее квадратическое отклонение

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 m_i - \overline{x}^2 = \frac{1}{100} \cdot 88 - (0,868)^2 \approx 0,127;$$

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{0,127} \approx 0,356.$$

1.6. Выборочные моменты. Асимметрия и эксцесс нормального распределения

Начальным выборочным моментом порядка r называется величина

$$\nu_r = \frac{1}{n}\sum_{i=1}^{k} x_i^r m_i.$$

Причем, если $r = 1$, получим среднее значение выборки

$$\nu_1 = \overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i;$$

если $r = 2$, получим среднее квадратическое

$$\nu_2 = \overline{x^2} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 m_i.$$

Центральным выборочным моментом порядка r называется величина

$$\mu_r = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^r m_i.$$

Причем, если $r = 1$, получим отклонение

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)m_i = 0;$$

если $r = 2$, получим выборочную дисперсию

$$\mu_2 = D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 m_i.$$

Для оценки отклонения эмпирического распределения изучаемой случайной величины от нормального распределения служат показатели асимметрии и эксцесса.

Асимметрией нормального распределения называется величина

$$A = \frac{\mu_3}{\sqrt{D_{\text{в}}^3}}.$$

Эксцессом нормального распределения называется величина

$$E = \frac{\mu_4}{\sqrt{D_{\text{в}}^4}} - 3,$$

где μ_3, μ_4 – центральные выборочные моменты третьего и четвертого порядков соответственно;

$D_{\text{в}}$ – выборочная дисперсия.

Асимметрия характеризует отклонение эмпирического распределения относительных частот от нормальной кривой по горизонтали: при $A > 0$ асимметрия правосторонняя, при $A < 0$ – левосторонняя. Для симметричного нормального распределения асимметрия $A = 0$ и $\overline{x} = Mo = Me$. Эксцесс характеризует отклонение эмпирического распределения относительных частот от нормальной кривой по вертикали или крутость распределения: при $E > 0$ распределение островершинное, при $E < 0$ – низковершинное. Причем распределение выборки можно принять за нормальное, если показатели асимметрии и эксцесса по абсолютной величине не превосходят своей утроенной ошибки репрезентативности, т. е. $|A| \le 3\sqrt{\frac{6}{n}}$ и $|E| \le 3\sqrt{\frac{24}{n}}$, где n – объем выборки.

Пример. Вычислить асимметрию и эксцесс нормального распределения для интервального статистического ряда, представленного в таблице на с. 10.

Решение. Заполним и рассчитаем следующую таблицу, учитывая, что $\bar{x} = 0{,}868$.

№	Интервалы	Частоты m_i	Середины интервалов x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})m_i$	$(x_i - \bar{x})^3 m_i$	$(x_i - \bar{x})^4 m_i$
1	[0,1; 0,3)	6	0,2	–0,668	–4,008	–1,788	1,195
2	[0,3; 0,5)	10	0,4	–0,468	–4,68	–1,025	0,48
3	[0,5; 0,7)	16	0,6	–0,268	–4,288	–0,308	0,083
4	[0,7; 0,9)	23	0,8	–0,068	–1,564	–0,007	0,0005
5	[0,9; 1,1)	19	1	0,132	2,508	0,044	0,006
6	[1,1; 1,3)	14	1,2	0,332	4,648	0,512	0,17
7	[1,3; 1,5)	7	1,4	0,532	3,724	1,054	0,561
8	[1,5; 1,7]	5	1,6	0,732	3,66	1,961	1,436
$\sum$		100	–	–	0	0,443	3,932

Вычислим центральные моменты третьего и четвертого порядков:

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^3 m_i = \frac{1}{100}\sum_{i=1}^{8}\left(x_i - 0,868\right)^3 = \frac{1}{100}\cdot 0,443 = 0,00443;$$

$$\mu_4 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^4 m_i = \frac{1}{100}\sum_{i=1}^{8}\left(x_i - 0,868\right)^4 = \frac{1}{100}\cdot 3,932 = 0,0393.$$

Вычислим асимметрию и эксцесс нормального распределения, учитывая, что

$$D_{\text{в}} = 0,127 \ : \ A = \frac{\mu_3}{\sqrt{D_{\text{в}}^3}} = \frac{0,00443}{\sqrt{(0,127)^3}} \approx 0,098;$$

$$E = \frac{\mu_4}{\sqrt{D_{\text{в}}^4}} - 3 = \frac{0,0393}{\sqrt{(0,127)^4}} - 3 \approx -0,547.$$

Вычислим утроенные ошибки репрезентативности асимметрии и эксцесса: $3\sqrt{\dfrac{6}{n}} \approx 0,735;$ $3\sqrt{\dfrac{24}{n}} \approx 1,47.$

Анализируя полученные результаты: $A = 0,098;$ $|A| \le 0,735;$ $\overline{x} > Me$ (0,868 > 0,857) и $\overline{x} > Mo$ (0,868 > 0,821), отметим, что данное выборочное распределение имеет незначительную правостороннюю асимметрию. Полученное значение эксцесса также не превышает допустимой ошибки $|E| \le 1,47,$ поэтому выборочное распределение подчиняется нормальному закону распределения.

1.7. Точечные и интервальные оценки параметров распределения

Как правило, изучаемая случайная величина распределена по некоторому закону, зависящему от параметров. Например, нормальный закон распределения имеет два параметра a и σ^2. Для оценки неизвестных параметров распределения из генеральной совокупности произведена выборка, которую можно рассматривать как совокупность n независимых случайных величин, одинаково распределенных с изучаемой случайной величиной. Различают *точечные* и *интервальные* оценки неизвестных параметров распределения.

Точечной оценкой параметра θ называется его приближенное значение $\tilde{\theta},$ зависящее от результатов выборки. Точечная оценка $\tilde{\theta} = f\left(x_1,\, x_2,\, ...,\, x_n\right)$ является функцией выборки или функцией независимых, одинаково распределенных случайных величин и сама представляет собой случайную величину.

Точечная оценка должна быть достаточно близкой к истинному значению параметра, т. е. быть подходящей оценкой. Точечная оценка называется *подходящей*, если она является несмещенной, эффективной и состоятельной оценкой. Оценка называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру: $M\left(\tilde{\theta}\right) = \theta.$ Оценка называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру: $\lim\limits_{n\to\infty}\left(\left|\theta - \tilde{\theta}\right| < \varepsilon\right) = 1.$ Точечная оценка называется *эффективной*, если ее дисперсия наименьшая по сравнению с дисперсиями других оценок.

Допустим, изучаемая случайная величина X распределена по нормальному закону с неизвестными параметрами $a = M(X)$ и $\sigma^2 = D(X)$. В качестве оценки параметра a рассмотрим среднее значение выборки $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, где $x_1, x_2, ..., x_n$ – независимые случайные величины, распределенные по тому же нормальному закону с параметрами $a = M(x_i)$ и $\sigma^2 = D(x_i)$, $i = \overline{1, n}$.

Вычислим числовые характеристики случайной величины $\overline{x}$:

$$M\left(\overline{x}\right) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M\left(x_i\right) = \frac{1}{n} na = a;$$

$$D\left(\overline{x}\right) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D\left(x_i\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Так как математическое ожидание равно оцениваемому параметру $M\left(\overline{x}\right) = a,$ то среднее значение выборки является несмещенной оценкой параметра a.

По теореме Чебышева среднее арифметическое независимых случайных величин сходится по вероятности к их математическому ожиданию, т. е. $\lim\limits_{n\to\infty}\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - a\right| < \varepsilon\right) = 1$ или $\lim\limits_{n\to\infty}\left(\left|\overline{x} - a\right| < \varepsilon\right) = 1.$

Значит, среднее значение выборки является также и состоятельной оценкой параметра a. Также можно доказать, что эта оценка имеет максимальную эффективность $e\left(\overline{x}\right) = 1$ и является эффективной, а значит и подходящей оценкой параметра a.

В качестве оценки параметра σ^2 рассмотрим выборочную дисперсию $D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$. Вычислим математическое ожидание выборочной дисперсии:

$$M\left(D_{\text{в}}\right)=M\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2\right)=M\left(\frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i-a\right)-\left(\overline{x}-a\right)\right)^2\right)=$$

$$=\frac{1}{n}M\sum_{i=1}^{n}\left(\left(x_i-a\right)^2-2\left(x_i-a\right)\left(\overline{x}-a\right)+\left(\overline{x}-a\right)^2\right)=\frac{1}{n}\sum_{i=1}^{n}M\left(x_i-a\right)^2-$$

$$-2M\left(\left(\overline{x}-a\right)\frac{1}{n}\sum_{i=1}^{n}\left(x_i-a\right)\right)+M\left(\overline{x}-a\right)^2=\frac{1}{n}\sum_{i=1}^{n}\sigma^2-M\left(\overline{x}-a\right)^2=\sigma^2-\frac{\sigma^2}{n},$$

или $M\left(D_{\text{в}}\right)=\dfrac{n-1}{n}\sigma^2.$ Тогда $M\left(D_{\text{в}}\right)=\dfrac{n-1}{n}\sigma^2$, а $M\left(\dfrac{n}{n-1}D_{\text{в}}\right)=\sigma^2.$

Величина $S^2=\dfrac{n}{n-1}D_{\text{в}}$ называется исправленной выборочной дисперсией и является несмещенной оценкой параметра σ^2.

Пусть θ – оцениваемый параметр, а $\tilde{\theta}$ – точечная оценка этого параметра. *Надежностью* или *доверительной вероятностью* называется вероятность γ, с которой выполняется неравенство $\left|\theta-\tilde{\theta}\right|<\delta$, или $P\left(\left|\theta-\tilde{\theta}\right|<\delta\right)=\gamma.$ Заменив неравенство $\left|\theta-\tilde{\theta}\right|<\delta$ равносильным $\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta,$ получим $P\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)=\gamma.$

Интервал $\left(\tilde{\theta}-\delta;\,\tilde{\theta}+\delta\right)$, который с заданной надежностью γ включает в себя оцениваемый параметр θ, называется *доверительным интервалом* или *интервальной оценкой* параметра. Величина δ называется *точностью* интервальной оценки. Величина $1-\gamma=\alpha$ называется *уровнем значимости* и задает часть значений параметра θ, которая не входит в доверительный интервал.

Найдем доверительный интервал с заданной надежностью γ для математического ожидания a при известной дисперсии σ^2 нормальной случайной величины X. Точечной оценкой этого параметра является среднее значение выборки $\overline{x}$, которое также можно рассматривать как нормальную случайную величину с математическим ожиданием a и дисперсией $\dfrac{\sigma^2}{n}$. Подставим эти значения в формулу, вычисляющую заданное отклонение нормальной случайной величины:

$$P\left(\left|X-a\right|<\delta\right)=2\Phi\left(\frac{\delta}{\sigma}\right).$$

Получим

$$P\left(\left|\overline{x}-a\right|<\delta\right)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=2\Phi(t)=\gamma,$$

где $t=\dfrac{\delta\sqrt{n}}{\sigma}\Rightarrow\delta=t\dfrac{\sigma}{\sqrt{n}}.$

От неравенства $\left|a-\overline{x}\right|<t\dfrac{\sigma}{\sqrt{n}}$ перейдем к равносильному

$$\overline{x}-t\frac{\sigma}{\sqrt{n}}<a<\overline{x}+t\frac{\sigma}{\sqrt{n}}.$$

Итак, доверительный интервал параметра *a* имеет вид

$$\left(\overline{x}-t\frac{\sigma}{\sqrt{n}};\ \overline{x}+t\frac{\sigma}{\sqrt{n}}\right),$$

где значение *t* находится из уравнения $\Phi(t)=\dfrac{\gamma}{2}$ по таблице значений функции Лапласа из прил. 1.

Пример. Случайная величина *X* имеет нормальное распределение с известным средним квадратическим отклонением σ = 0,36. Найти доверительные интервалы параметра *a* по выборочной средней $\overline{x}=0,868$ и объему выборки *n* = 100 при надежности γ_1 = 0,95 и γ_2 = 0,99.

Решение. 1. Вычислим точность интервальной оценки параметра *a* при надежности $\gamma_1=0,95:\ \delta=t\dfrac{\sigma}{\sqrt{n}}$, где значение *t* находим из уравнения $2\Phi(t)=0,95$ или $\Phi(t)=\dfrac{0,95}{2}=0,475,$ из прил. 1 следует, что $t=1,96.$ Тогда $\delta=1,96\cdot\dfrac{0,36}{\sqrt{100}}\approx0,071,$ границы интервала равны: $\overline{x}-\delta=0,868-0,071=0,797$ и $\overline{x}+\delta=0,868+0,071=0,939.$

Итак, доверительный интервал (0,797; 0,939) включает параметр *a* с надежностью $\gamma_1 = 0,95.$

2. Вычислим точность интервальной оценки параметра *a* при надежности $\gamma_2 = 0,99: \delta = t\frac{\sigma}{\sqrt{n}},$ где значение t находим из уравнения

$2\Phi(t) = 0,99$ или $\Phi(t) = \frac{0,99}{2} = 0,495,$ из прил. 1 находим $t = 2,58.$

Тогда $\delta = 2,58 \cdot \frac{0,36}{\sqrt{100}} \approx 0,093,$ границы интервала равны: $\bar{x} - \delta =$ $= 0,868 - 0,093 = 0,775$ и $\bar{x} + \delta = 0,868 + 0,093 = 0,961.$

Итак, доверительный интервал (0,775; 0,961) включает параметр *a* с надежностью $\gamma_2 = 0,99.$ Длина интервала увеличилась, поскольку он включает уже 99 % значений параметра *a*, но уровень значимости при этом понизился с $\alpha_1 = 5$ % до $\alpha_2 = 1$ %.

1.8. Понятие статистической гипотезы. Статистический критерий. Ошибки первого и второго рода

Статистической гипотезой называется некоторое предположение о генеральной совокупности, проверяемое по выборке. Например, гипотеза о законе распределения случайной величины или о значениях параметров распределения.

Выдвинутое предположение называется *нулевой гипотезой* H_0, *конкурирующей* называют гипотезу H_1, противоречащую нулевой.

Например, если нулевая гипотеза состоит в предположении, что параметр a нормального распределения равен значению a_0, то конкурирующая гипотеза может состоять в предположении, что $a \neq a_0$ или $a \neq a_0$, $a > a_0$. Это записывается так: $H_0 : a = a_0$; $H_1 : a \neq a_0$, или $H_0 : a = a_0$; $H_1 : a \neq a_0$, $a > a_0$. Если нулевая гипотеза отвергается, то принимается конкурирующая. Гипотеза, содержащая только одно предположение, называется *простой*. *Сложная* гипотеза состоит из конечного или бесконечного числа простых гипотез.

Правило, по которому принимают или отвергают нулевую гипотезу, называется *статистическим критерием.* Проверку гипотезы проводят на основании выборки, составляя для этого *статистику* критерия или функцию выборки. Множество значений статистики делится на *критическую* область S и область принятия нулевой гипотезы $\overline{S}$. Основной принцип проверки гипотез заключается в следующем. Если наблюдаемое значение статистики, вычисленное по выборке, попадает в критическую область S, то нулевую гипотезу отвергают и принимают конкурирующую. Если $f_{\text{набл}}\left(x_1, x_2, ..., x_n\right)$ попадает в область принятия гипотезы $\overline{S}$, то нулевую гипотезу принимают.

При проверке гипотез может быть принято неправильное решение и допущены ошибки первого и второго рода. *Ошибка первого рода* состоит в том, что отвергается нулевая гипотеза H_0, хотя на самом деле она верна. *Ошибка второго рода* состоит в том, что принимается нулевая гипотеза H_0, хотя на самом деле она неверна, а верна конкурирующая H_1. Это показано в следующей таблице.

Объективное состояние гипотезы H_0	Субъективные действия	
	Принимаем гипотезу H_0	Отвергаем гипотезу H_0
Гипотеза H_0 верна	Правильное решение	Ошибка первого рода
Гипотеза H_0 неверна	Ошибка второго рода	Правильное решение

Вероятность ошибки первого рода называется уровнем значимости статистического критерия и обозначается $\alpha = P(H_1 / H_0)$. Вероятность ошибки второго рода обозначается $\beta = P(H_0 / H_1)$. Вероятность недопущения ошибки второго рода или вероятность отвергнуть неверную нулевую гипотезу называется мощностью критерия $1 - \beta = = P\left(f_{\text{набл}}\left(x_1, x_2, ..., x_n\right) \in S / H_1\right)$. Ошибки первого и второго рода находятся в обратной зависимости. Одновременное уменьшение этих ошибок возможно только при увеличении объема выборки.

1.9. Критерий согласия Пирсона

Во многих случаях закон распределения изучаемой случайной величины неизвестен, но есть основания предположить, что он имеет вполне определенный вид: нормальный, биномиальный или другой.

Критерием согласия называется критерий о предполагаемом законе распределения изучаемой случайной величины. Применим критерий согласия Пирсона к выдвигаемой нулевой гипотезе H_0 о нормальном законе распределения случайной величины X. Гипотеза H_0 является простой гипотезой и проверяет только одно предположение о законе распределения случайной величины X. Поэтому вместо неизвестных параметров нормального закона распределения a, σ используют их точечные подходящие оценки, среднее значение выборки $\overline{x}$ и корень квадратный из исправленной выборочной дисперсии:

$$\sigma = \sqrt{S^2} = \sqrt{\frac{n}{n-1} D_{\text{в}}}.$$

Критерий χ^2 Пирсона заключается в сравнении эмпирических или наблюдаемых частот и теоретических частот, вычисленных из предположения о нормальном законе распределения. Для проверки выдвинутой гипотезы выборочную совокупность разбивают на k частичных интервалов $[x'_0;\, x'_1),\, [x'_1;\, x'_2),\, ...,\, [x'_{k-1};\, x'_k]$ и получают ряд эмпирических частот m_1, m_2, …, m_k. В соответствии с предполагаемым нормальным законом распределения случайной величины X рассчитываются вероятности ее попадания в интервалы:

$$p_0 = P(-\infty,\, x'_1) = \Phi\left(\frac{x'_1 - \overline{x}}{\sigma}\right) - \Phi(-\infty) = 0,5 + \Phi\left(\frac{x'_1 - \overline{x}}{\sigma}\right);$$

$$p_1 = P(x'_1,\, x'_2) = \Phi\left(\frac{x'_2 - \overline{x}}{\sigma}\right) - \Phi\left(\frac{x'_1 - \overline{x}}{\sigma}\right),\ ...;$$

$$p_k = P\left(x'_{k-1}, +\infty\right) = \Phi(+\infty) - \Phi\left(\frac{x'_{k-1} - \overline{x}}{\sigma}\right) = 0,5 - \Phi\left(\frac{x'_{k-1} - \overline{x}}{\sigma}\right).$$

Тогда теоретические частоты попадания случайной величины в эти интервалы равны $m_i^0 = np_i,\ i = \overline{1,\, k}.$ Если эмпирические частоты m_i сильно отличаются от теоретических $m_i^0,\ i = \overline{1,\, n},$ то выдвинутую гипотезу H_0 о нормальном законе распределения случайной величины X отвергают, в противном случае ее принимают.

В качестве меры расхождения между эмпирическими и теоретическими частотами Пирсоном была предложена статистика

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(m_i - m_i^0\right)^2}{m_i^0} = \sum_{i=1}^{k} \frac{\left(m_i - np_i\right)^2}{np_i}.$$

Согласно теореме Пирсона эта величина имеет распределение χ^2 с числом степеней свободы $s = k - r - 1$, где k – число частичных интервалов, r – количество параметров предполагаемого закона распределения и для нормального закона $r = 2$, тогда $s = k - 3$.

Правило применения критерия χ^2 заключается в следующем:

1. Вычисляем ряд теоретических частот $m_i^0 = np_i$, $i = \overline{1, n}$, причем каждый частичный интервал должен содержать не менее 5 значений эмпирических частот, т. е. $m_i \geq 5$. В противном случае соседние интервалы объединяем в один.

2. Вычисляем наблюдаемое значение статистики

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{\left(m_i - m_i^0\right)^2}{m_i^0} = \sum_{i=1}^{k} \frac{\left(m_i - np_i\right)^2}{np_i}.$$

3. Задаем уровень значимости α и по таблице распределения χ^2 по уровню значимости α и числу степеней свободы $s = k - 3$ находится граница правосторонней критической области, критическая точка $\chi^2_{\alpha,\, s}$.

4. Если $\chi^2_{\text{набл}} < \chi^2_{\alpha,\, s}$, то выдвинутую нулевую гипотезу о нормальном законе распределения случайной величины принимаем. Если $\chi^2_{\text{набл}} > \chi^2_{\alpha,\, s}$, то нулевую гипотезу отвергаем.

1.10. Теория корреляции. Корреляционная зависимость

Случайные величины X и Y могут быть связаны между собой функциональной зависимостью, статистической зависимостью или быть независимыми. Строгая функциональная зависимость между значениями одной величины и значениями другой реализуется довольно редко из-за влияния случайных факторов на величины.

Статистической зависимостью называется зависимость между значениями одной величины и параметрами распределения другой. Частным случаем этой зависимости является *корреляционная* зависимость между значениями одной величины и средними значениями другой.

Изучение корреляционной зависимости удобно проводить для сгруппированных данных. Данные выборочных наблюдений над случайными величинами X и Y группируют и записывают в виде корреляционной таблицы.

Интервалы и средние значения случайных величин		Y				
		$[y_0'; y_1')$ y_1	$[y_1'; y_2')$ y_2	…	$[y_{k-1}'; y_k')$ y_k	m_x
X	$[x_0'; x_1')$ x_1	m_{11}	m_{12}	…	m_{1k}	m_{x_1}
	$[x_1'; x_2')$ x_2	m_{21}	m_{22}	…	m_{2k}	m_{x_2}
	…	…	…	…	…	…
	$[x_{k-1}'; x_k']$ x_k	m_{k1}	m_{k2}	…	m_{kk}	m_{x_k}
	m_y	m_{y_1}	m_{y_2}	…	m_{y_k}	n

Суммы частот в последних строке и столбце равны объему выборки $\sum_{i=1}^{k} m_{xi} = \sum_{j=1}^{k} m_{yj} = n.$ Первый и последний столбец таблицы соответствуют интервальному ряду распределения X, а первая и последняя строка – Y. Причем, под интервалами записаны середины интервалов:

$$x_i = \frac{x_{i-1}' + x_i'}{2}, \ i = \overline{1, k}; \quad y_j = \frac{y_{j-1}' + y_j'}{2}, \quad j = \overline{1, k}.$$

В таблице можно выделить k условных законов распределения Y при $X = x_1, x_2, ..., x_k$ и k условных законов распределения X при $Y = y_1, y_2, ..., y_k$ соответственно и вычислить условные средние:

$$\overline{y}_{x_i} = \frac{1}{m_{x_i}} \sum_{j=1}^{k} y_j m_{ij}, \ j = \overline{1, k}; \ \overline{x}_{y_j} = \frac{1}{m_{y_j}} \sum_{i=1}^{k} x_i m_{ij}, \ i = \overline{1, k}.$$

В этом случае получаем взаимно однозначное соответствие между значениями одной величины и условными средними другой. Эту корреляционную зависимость можно записать уравнениями регрессии:

$$\overline{y}_x = f(x) \text{ и } \overline{x}_y = \varphi(y).$$

Для определения вида функции регрессии (линейная, квадратическая или другая) на координатную плоскость наносим точки $(x_1; \overline{y}_1)$, $(x_2; \overline{y}_2)$, …, $(x_k; \overline{y}_k)$ или $(\overline{x}_1; y_1)$, $(\overline{x}_2; y_2)$, …, $(\overline{x}_k; y_k)$ и по характеру расположения точек выбираем подходящий вид функции регрессии.

1.11. Коэффициент линейной корреляции

Выборочный коэффициент линейной корреляции служит оценкой меры тесноты линейной корреляционной зависимости между случайными величинами и выражается формулой:

$$r_{\text{в}} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y},$$

где $\overline{x}$, $\overline{y}$ – средние значения выборок;

σ_x, σ_y – средние квадратические отклонения случайных величин X и Y.

При этом средняя величина $\overline{xy}$ вычисляется по корреляционной таблице по формуле:

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

Свойства коэффициента линейной корреляции.

1. Значение коэффициента линейной корреляции по модулю не превышает единицы: $|r_{\text{в}}| \le 1$.

2. Если $|r_{\text{в}}| = 1$, то между случайными величинами существует линейная функциональная зависимость.

3. Если коэффициент линейной корреляции $r_{\text{в}} > 0$, то корреляционная зависимость между величинами X и Y прямая, а если $r_{\text{в}} < 0$, то обратная.

4. Если $|r_{\text{в}}| \ge 0,7$, то между случайными величинами существует тесная линейная корреляционная зависимость.

5. Если случайные величины независимы, то коэффициент линейной корреляции $r_{\text{в}} = 0$.

6. Если коэффициент линейной корреляции $r_{\text{в}} = 0$, то случайные величины не коррелированы.

Квадрат коэффициента корреляции $D = r_{\text{в}}^2$ называется *коэффициентом детерминации* в долях или в процентах $D = r_{\text{в}}^2 \cdot 100$ %. Он показывает, на сколько процентов в среднем изменение зависимой переменной Y зависит от независимой переменной X.

Доверительный интервал для коэффициента линейной корреляции между случайными величинами X и Y с заданной надежностью $\gamma = 0{,}95$ находится по формуле:

$$\left(r_{\text{в}} - t\frac{1-r_{\text{в}}^2}{\sqrt{n}};\ r_{\text{в}} + t\frac{1-r_{\text{в}}^2}{\sqrt{n}} \right),$$

где значение t находим из уравнения $2\Phi\left(t\right) = 0{,}95$ или $\Phi\left(t\right) = \frac{0{,}95}{2} = 0{,}475,$ тогда из прил. 1 следует, что $t = 1{,}96$.

1.12. Метод наименьших квадратов нахождения параметров линейной регрессии

Рассмотрим этот метод для не сгруппированных выборочных данных случайных величин X и Y:

№	1	2	…	n
X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

Уравнение линейной регрессии имеет вид

$$y = ax + b,$$

где a, b – неизвестные параметры.

Запишем теоретические значения y: $y_1^{\text{т}} = ax_1 + b;\ \ldots;\ y_n^{\text{т}} = ax_n + b$ и отклонения теоретических значений от наблюдаемых:

$$\begin{aligned}
\varepsilon_1 &= y_1^{\text{т}} - y_1 = ax_1 + b - y_1;\\
\varepsilon_2 &= y_2^{\text{т}} - y_2 = ax_2 + b - y_2;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots;\\
\varepsilon_n &= y_n^{\text{т}} - y_n = ax_n + b - y_n.
\end{aligned}$$

Подберем такие значения параметров a, b, чтобы сумма квадратов отклонений была наименьшей. Для этого рассмотрим функцию

$$F(a,\, b) = \sum_{i=1}^{n}\left(ax_i + b - y_i\right)^2$$

и исследуем ее на экстремум.

Найдем частные производные этой функции:

$$F_a' = 2\sum_{i=1}^{n}(ax_i + b - y_i)x_i = 2\left(a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}x_i y_i\right);$$

$$F_b' = 2\sum_{i=1}^{n}(ax_i + b - y_i) = 2\left(a\sum_{i=1}^{n}x_i + bn - \sum_{i=1}^{n}y_i\right).$$

Приравняем частные производные к нулю и получим систему уравнений:

$$\begin{cases} a\sum_{i=1}^{n}x_i^2 + b\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}x_i y_i, \\ a\sum_{i=1}^{n}x_i + bn = \sum_{i=1}^{n}y_i. \end{cases}$$

Разделим каждое уравнение на объем выборки n и с учетом выборочных характеристик

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n}x_i^2,\ \overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i,\ \overline{xy} = \frac{1}{n}\sum_{i=1}^{n}x_i y_i,\ \overline{y} = \frac{1}{n}\sum_{i=1}^{n}y_i,$$

получим:

$$\begin{cases} \overline{x^2}a + \overline{x}b = \overline{xy}, \\ \overline{x}a + b = \overline{y}. \end{cases}$$

Откуда $a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x^2} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} \cdot \dfrac{\sigma_y}{\sigma_x} = r_в \dfrac{\sigma_y}{\sigma_x}$, $b = \overline{y} - a\overline{x}.$

Подставим найденные значения параметров, получим уравнения регрессии y на x и x на y соответственно:

$$\overline{y}_x = r_в \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right) + \overline{y} \quad \text{и} \quad \overline{x}_y = r_в \frac{\sigma_x}{\sigma_y}\left(y - \overline{y}\right) + \overline{x}.$$

2. ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Задание 1. Статистические ряды.

По статистическим выборочным данным (прил. 2) случайной величины X или Y требуется:

1) составить интервальные статистические ряды распределения частот и относительных частот;

2) построить гистограмму и полигон относительных частот;

3) составить и построить график эмпирической функции распределения, построить кумулянту;

4) вычислить числовые характеристики выборки: среднее значение выборки, моду, медиану, выборочную дисперсию, асимметрию и эксцесс нормального распределения, выборочный коэффициент вариации;

5) найти подходящие точечные оценки параметров a и σ нормального распределения и построить доверительный интервал параметра a с надежностью 95 %.

Решение типового примера

Выполним типовой пример для выборочных данных случайных величин X – стоимости валовой продукции (тыс. руб/га) и Y – стоимости оборотных фондов (тыс. руб/га), приведенных в таблице.

№	X	Y	№	X	Y	№	X	Y	№	X	Y
1	1,291	1,421	26	1,550	1,707	51	2,852	3,245	76	2,947	3,245
2	0,982	2,092	27	3,690	4,064	52	2,275	1,984	77	1,916	2,110
3	0,619	3,062	28	2,554	2,813	53	1,654	2,293	78	2,072	2,282
4	1,743	1,920	29	2,130	1,061	54	2,016	2,028	79	1,841	2,028
5	2,042	2,249	30	2,071	2,774	55	3,125	1,682	80	1,498	1,650
6	1,373	1,512	31	0,958	2,080	56	3,369	3,364	81	1,316	1,449
7	1,339	1,475	32	1,221	1,655	57	2,947	2,143	82	3,043	3,352
8	1,438	1,583	33	3,391	2,577	58	1,802	1,171	83	1,988	2,189
9	0,661	1,814	34	2,476	2,808	59	2,082	3,107	84	2,547	2,805
10	1,647	2,052	35	0,964	2,736	60	1,841	1,431	85	1,300	1,431
11	3,706	4,082	36	2,519	2,247	61	1,527	2,458	86	1,604	1,767
12	3,570	3,932	37	1,889	3,873	62	3,054	1,474	87	2,233	2,459
13	2,542	2,800	38	1,503	2,684	63	1,946	2,346	88	1,338	1,474
14	1,776	1,956	39	2,340	1,263	64	1,064	2,281	89	1,664	1,833
15	1,436	1,581	40	2,550	1,206	65	2,821	1,055	90	3,825	4,212
16	1,155	1,272	41	2,485	1,679	66	1,300	1,345	91	2,669	2,939
17	1,728	1,903	42	2,040	2,074	67	2,232	3,735	92	1,435	1,580
18	3,623	3,990	43	3,516	1,587	68	1,338	2,727	93	1,955	2,153
19	2,672	2,943	44	2,437	1,737	69	1,659	1,827	94	1,237	1,362
20	2,159	2,378	45	1,147	3,141	70	1,951	2,148	95	2,327	2,563
21	1,815	1,999	46	1,097	2,506	71	3,790	4,174	96	1,654	1,821
22	2,034	2,240	47	1,524	1,821	72	3,308	3,643	97	2,275	2,506
23	3,004	3,308	48	2,883	2,221	73	1,743	1,919	98	3,369	3,710
24	3,139	3,457	49	1,441	3,442	74	1,314	1,448	99	2,947	3,245
25	2,944	3,241	50	1,578	3,710	75	2,643	2,911	100	2,042	2,249

Случайная величина X

1. Составим интервальный статистический ряд распределения частот и относительных частот.

Прежде всего вычислим количество интервалов по формуле $k = 1 + \log_2 n \approx 1 + 3,32 \cdot \lg n.$ Поскольку объем выборки $n = 100$, то количество интервалов $k = 1 + \log_2 100 \approx 1 + 3,32 \cdot \lg 100 = 7,64 \approx 8.$

Затем вычислим размах вариации

$$R = x_{\max} - x_{\min} = 3,825 - 0,619 = 3,206.$$

Шаг разбиения вычислим по формуле

$$h = \frac{R}{k-1} = \frac{x_{\max} - x_{\min}}{k-1} = \frac{3,825 - 0,619}{7} = \frac{3,206}{7} = 0,458.$$

Далее находим границы интервалов:

$$x_0' = x_{\min} - \frac{h}{2} = 0,619 - 0,229 = 0,39; \quad x_1' = x_0' + h = 0,39 + 0,458 = 0,848;$$
$$x_2' = x_1' + h = 0,848 + 0,458 = 1,306; \quad x_3' = x_2' + h = 1,306 + 0,458 = 1,764;$$
$$x_4' = x_3' + h = 1,764 + 0,458 = 2,222; \quad x_5' = x_4' + h = 2,222 + 0,458 = 2,68;$$
$$x_6' = x_5' + h = 2,68 + 0,458 = 3,138; \quad x_7' = x_6' + h = 3,138 + 0,458 = 3,596;$$
$$x_8' = x_7' + h = 3,596 + 0,458 = 4,054$$

и записываем сами интервалы: [0,39; 0,848), [0,848; 1,306), [1,306; 1,764), [1,764; 2,222), [2,222; 2,68), [2,68; 3,138), [3,138; 3,596), [3,596; 4,054].

Просматриваем все выборочные данные случайной величины *X* и распределяем их по интервалам. Причем в каждый интервал включаем те значения, которые больше или равны нижней границе интервала и меньше верхней границы. Подсчет частот представлен в таблице.

№	Интервалы	Подсчет частот	Частоты
1	[0,390; 0,848)	II	2
2	[0,848; 1,306)	IIIIIIIIIIII	12
3	[1,306; 1,764)	IIIIIIIIIIIIIIIIIIIIIIIII	25
4	[1,764; 2,222)	IIIIIIIIIIIIIIIIIIIII	21
5	[2,222; 2,680)	IIIIIIIIIIIIIIIII	17
6	[2,680; 3,138)	IIIIIIIIIII	11
7	[3,138; 3,596)	IIIIIII	7
8	[3,596; 4,054]	IIIII	5
$\sum$	–	–	100

Для получения статистического ряда относительных частот разделим частоты m_i на объем выборки, т. е. $w_i = \frac{m_i}{n}$, $n = \overline{1, 8}$.

В следующей таблице представлен интервальный статистический ряд распределения частот и относительных частот.

№	Интервалы	Середины интервалов	Частоты	Накопленные частоты	Относитель-ные частоты	Накопленные относитель-ные частоты
1	[0,390; 0,848)	0,619	2	2	0,02	0,02
2	[0,848; 1,306)	1,077	12	14	0,12	0,14
3	[1,306; 1,764)	1,535	25	39	0,25	0,39
4	[1,764; 2,222)	1,993	21	60	0,21	0,60
5	[2,222; 2,680)	2,451	17	77	0,17	0,77
6	[2,680; 3,138)	2,909	11	88	0,11	0,88
7	[3,138; 3,596)	3,367	7	95	0,07	0,95
8	[3,596; 4,054]	3,825	5	100	0,05	1
$\sum$	–	–	100	–	1	–

2. Строим гистограмму и полигон относительных частот (рис. 2).

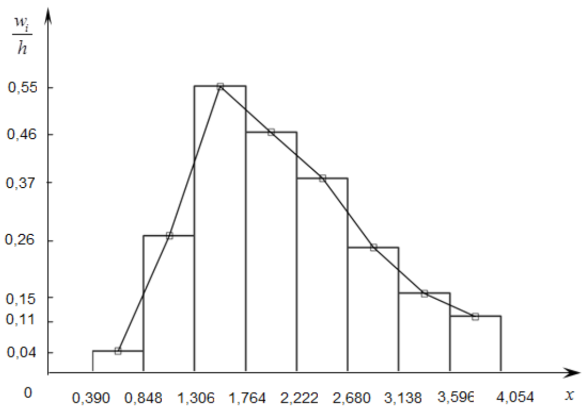


Рис. 2

3. Значения эмпирической функции распределения $F^*(x)$ записаны в строке 7 приведенной выше таблицы. Эмпирическая функция распределения имеет вид

$$F^*(x) = \begin{cases} 0, \text{ если } x \le 0,390, \\ 0,02, \text{ если } 0,390 < x \le 0,848, \\ 0,14, \text{ если } 0,848 < x \le 1,306, \\ 0,39, \text{ если } 1,306 < x \le 1,764, \\ 0,60, \text{ если } 1,764 < x \le 2,222, \\ 0,77, \text{ если } 2,222 < x \le 2,680, \\ 0,88, \text{ если } 2,680 < x \le 3,138, \\ 0,95, \text{ если } 3,138 < x \le 3,596, \\ 1, \text{ если } x > 3,596. \end{cases}$$

График эмпирической функции распределения изображен на рис. 3.

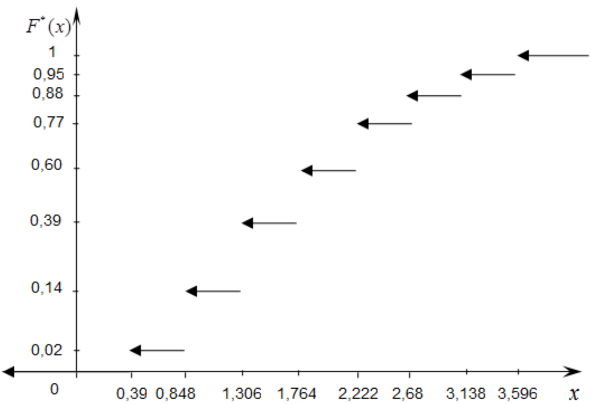


Рис. 3

Для приближения к графику теоретической функции распределения нормального закона распределения построим *кумулянту*. Для построения кумулянты на координатной плоскости отмечаем точки, абсциссы которых являются границами интервалов, а ординаты равны накопленным относительным частотам: (0,39; 0), (0,848; 0,02), (1,306; 0,14), (1,764; 0,39), (2,222; 0,6), (2,680; 0,77), (3,138; 0,88), (3,596; 0,95). Соединим эти точки отрезками, полученная ломаная линия является кумулянтой (рис. 4).

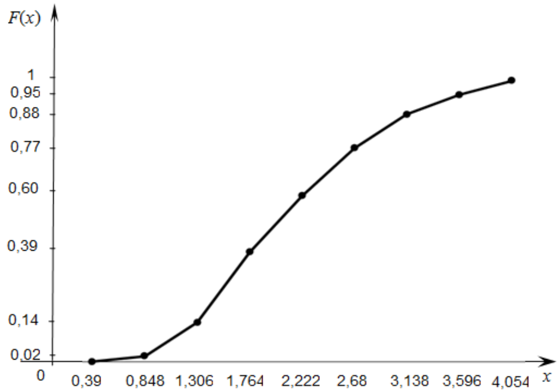


Рис. 4

4. Составим таблицу для вычисления выборочных числовых характеристик случайной величины.

№	Интервалы	Частоты	Середины интервалов	$(x_i - \bar{x})$	$(x_i - \bar{x})m_i$	$(x_i - \bar{x})^2 m_i$	$(x_i - \bar{x})^3 m_i$	$(x_i - \bar{x})^4 m_i$
1	[0,390; 0,848)	2	0,619	–1,489	–2,977	4,431	–6,596	9,818
2	[0,848; 1,306)	12	1,077	–1,031	–12,366	12,743	–13,132	13,532
3	[1,306; 1,764)	25	1,535	–0,573	–14,313	8,194	–4,691	2,686
4	[1,764; 2,222)	21	1,993	–0,115	–2,405	0,275	–0,032	0,004
5	[2,222; 2,680)	17	2,451	0,343	5,840	2,006	0,689	0,237
6	[2,680; 3,138)	11	2,909	0,802	8,817	7,066	5,664	4,539
7	[3,138; 3,596)	7	3,367	1,260	8,817	11,104	13,986	17,615
8	[3,596; 4,054]	5	3,825	1,718	8,588	14,749	25,331	43,507
$\sum$	–	100	–	–	0	60,569	21,220	91,938

Вычислим среднее значение выборки

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \frac{1}{100}\cdot(2\cdot 0,619 + 12\cdot 1,077 + 25\cdot 1,535 + 21\cdot 1,993 + 17\cdot 2,451 +$$

$$+11\cdot 2,909 + 7\cdot 3,367 + 5\cdot 3,825) = \frac{1}{100}\cdot 210,75 \approx 2,108.$$

Вычислим моду по формуле

$$Mo = x_0 + h\frac{m_0 - m_1}{(m_0 - m_1) + (m_0 - m_2)},$$

где $x_0 = 1,306$ – начало модального интервала или интервала с наибольшей частотой;

$h = 0,458$ – ширина интервала;

$m_0 = 25$ – частота модального интервала;

$m_1 = 12$ – частота интервала, предшествующего модальному;

$m_2 = 21$ – частота интервала, следующего за модальным.

Тогда $Mo = 1,306 + 0,458\frac{25-12}{(25-12)+(25-21)} \approx 1,656.$

Вычислим медиану по формуле

$$Me = x_0 + h\frac{n - 2n_1}{2m_e},$$

где $x_0 = 1,764$ – начало медианного интервала или интервала, накопленная частота которого больше половины объема выборки;

h = 0,458 – ширина интервала;

$n_1 = 39$ – накопленная частота интервала, предшествующего медианному;

$m_e = 21$ – частота медианного интервала.

Тогда $Me = 1,764 + 0,458 \cdot \frac{100 - 2\cdot 39}{2\cdot 21} \approx 2,004.$

Вычислим выборочную дисперсию

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 m_i = \frac{1}{100}\cdot 60,569 \approx 0,606$$

и выборочный коэффициент вариации

$$v = \frac{\sqrt{D_{\text{в}}}}{\overline{x}}\cdot 100\ \% = \frac{\sqrt{0,606}}{2,108}\cdot 100\ \% \approx 37\ \%.$$

Вычислим центральные моменты третьего и четвертого порядков:

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^3 m_i, \quad \mu_3 = \frac{1}{100}\cdot 21,220 \approx 0,212;$$

$$\mu_4 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^4 m_i, \ \mu_4 = \frac{1}{100}\cdot 91,938 \approx 0,919.$$

Вычислим асимметрию и эксцесс нормального распределения:

$$A = \frac{\mu_3}{\sqrt{D_{\text{в}}^3}} = \frac{0,212}{\sqrt{(0,606)^3}} \approx 0,45; \ \ E = \frac{\mu_4}{\sqrt{D_{\text{в}}^4}} - 3 = \frac{0,919}{\sqrt{(0,606)^4}} - 3 \approx -0,498.$$

Вычислим утроенные ошибки репрезентативности асимметрии и эксцесса:

$$3\cdot\sqrt{\frac{6}{n}} = 3\cdot\sqrt{\frac{6}{100}} \approx 0,735; \ \ 3\sqrt{\frac{24}{n}} = \sqrt{\frac{24}{100}} \approx 1,47.$$

Анализируя полученные результаты: $A = 0{,}45 > 0$, $\overline{x} > Me$ (2,108 > 2,004) и $\overline{x} > Mo$ (2,108 > 1,656), отметим, что данное выборочное распределение имеет незначительную правостороннюю асимметрию. Полученное значение эксцесса E = –0,498 отрицательное, поэтому выборочное распределение низковершинное. Значения асимметрии и эксцесса не превышают допустимой ошибки $|A| \le 0,735$ и $|E| \le 1,47,$ поэтому выборочное распределение подчиняется нормальному закону распределения.

5. Подходящей точечной оценкой параметра a нормального распределения служит среднее значение выборки $\overline{x} = 2{,}108,$ а параметра σ^2 – исправленная выборочная дисперсия

$$S^2 = \frac{n}{n-1}D_{\text{в}} = \frac{100}{99}\cdot 0,606 \approx 0,612.$$

Тогда с некоторым приближением параметр

$$\sigma_x = \sqrt{S^2} = \sqrt{0,612} \approx 0,782.$$

Найдем доверительный интервал параметра a с надежностью $\gamma = 0,95.$ Вычислим точность интервальной оценки параметра a

$$\delta = t\frac{\sigma}{\sqrt{n}},$$

где значение t находим из уравнения $2\Phi(t) = 0,95$ или $\Phi(t) = \frac{0,95}{2} = 0,475,$ по прил. 1 $t = 1,96.$

Тогда

$$\delta = 1,96 \cdot \frac{0,782}{\sqrt{100}} \approx 0,153$$

и границы интервала равны: $\overline{x} - \delta = 2,108 - 0,153 = 1,955$ и $\overline{x} + \delta =$ $= 2,108 + 0,153 = 2,261$. Итак, доверительный интервал (1,955; 2,261) включает параметр *a* нормального распределения случайной величины *X* с надежностью $\gamma = 0,95$.

Случайная величина Y

1. Составим интервальный статистический ряд распределения частот и относительных частот. Прежде всего вычислим количество интервалов по формуле $k = 1 + \log_2 n \approx 1 + 3,32 \cdot \lg n.$ Поскольку объем выборки $n = 100$, то количество интервалов $k = 1 + \log_2 100 \approx$ $\approx 1 + 3,32 \cdot \lg 100 = 7,64 \approx 8.$

Затем вычислим размах вариации

$$R = y_{\max} - y_{\min} = 4,212 - 1,055 = 3,157.$$

Шаг разбиения вычислим по формуле

$$h = \frac{R}{k-1} = \frac{y_{\max} - y_{\min}}{k-1} = \frac{4,212 - 1,055}{7} = \frac{3,157}{7} = 0,451.$$

Далее находим границы интервалов:

$$y'_0 = y_{\min} - \frac{h}{2} = 1,055 - 0,225 = 0,83;$$
$$y'_1 = y'_0 + h = 0,83 + 0,451 = 1,281;$$

$$y_2' = y_1' + h = 1,281 + 0,451 = 1,732;$$
$$y_3' = y_2' + h = 1,732 + 0,451 = 2,183;$$
$$y_4' = y_3' + h = 2,183 + 0,451 = 2,634;$$
$$y_5' = y_4' + h = 2,634 + 0,451 = 3,085;$$
$$y_6' = y_5' + h = 3,085 + 0,451 = 3,536;$$
$$y_7' = y_6' + h = 3,536 + 0,451 = 3,987;$$
$$y_8' = y_7' + h = 3,987 + 0,451 = 4,438$$

и записываем сами интервалы [0,83; 1,281), [1,281; 1,732), [1,732; 2,183), [2,183; 2,634), [2,634; 3,085), [3,085; 3,536), [3,536; 3,987), [3,987; 4,438].

Просматриваем все выборочные данные случайной величины Y и распределяем их по интервалам. Причем в каждый интервал включаем те значения, которые больше или равны нижней границе интервала и меньше верхней границы. Подсчет частот представлен в следующей таблице.

№	Интервалы	Подсчет частот	Частоты
1	[0,830; 1,281)	IIIIII	6
2	[1,281; 1,732)	IIIIIIIIIIIIIIIIIIII	20
3	[1,732; 2,183)	IIIIIIIIIIIIIIIIIIIIIII	23
4	[2,183; 2,634)	IIIIIIIIIIIIIIIII	17
5	[2,634; 3,085)	IIIIIIIIIIII	12
6	[3,085; 3,536)	IIIIIIIIIII	11
7	[3,536; 3,987)	IIIIII	6
8	[3,987; 4,438]	IIIII	5
$\sum$	–	–	100

Ниже в таблице представим интервальный статистический ряд распределения частот и относительных частот.

№	Интервалы	Середины интервалов	Частоты	Накопленные частоты	Относитель-ные частоты	Накопленные относитель-ные частоты
1	[0,830; 1,281)	1,056	6	6	0,06	0,06
2	[1,281; 1,732)	1,507	20	26	0,20	0,26
3	[1,732; 2,183)	1,958	23	49	0,23	0,49
4	[2,183; 2,634)	2,409	17	66	0,17	0,66
5	[2,634; 3,085)	2,860	12	78	0,12	0,78
6	[3,085; 3,536)	3,311	11	89	0,11	0,89
7	[3,536; 3,987)	3,762	6	95	0,06	0,95
8	[3,987; 4,438]	4,213	5	100	0,05	1
$\sum$	–	–	100	–	1	–

2. Строим гистограмму и полигон относительных частот (рис. 5).

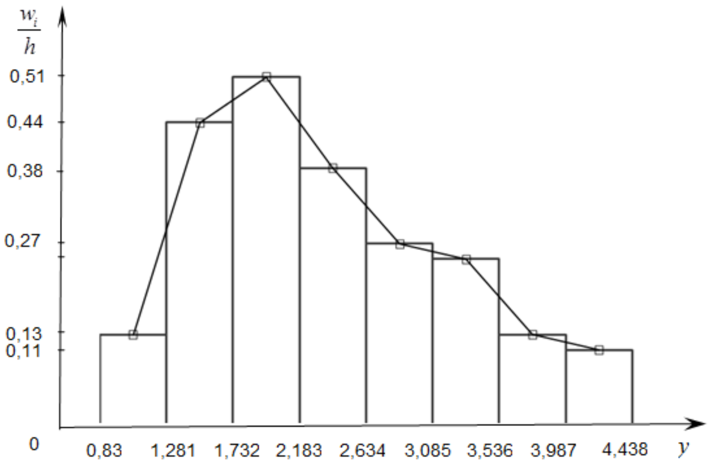


Рис. 5

3. Значения эмпирической функции распределения записаны в строке 7 таблицы, приведенной выше.

Эмпирическая функция распределения имеет вид

$$F^*(y) = \begin{cases} 0, & \text{если } y \le 0,830, \\ 0,06, & \text{если } 0,830 < y \le 1,281, \\ 0,26, & \text{если } 1,281 < y \le 1,732, \\ 0,49, & \text{если } 1,732 < y \le 2,183, \\ 0,66, & \text{если } 2,183 < y \le 2,634, \\ 0,78, & \text{если } 2,634 < y \le 3,085, \\ 0,89, & \text{если } 3,085 < y \le 3,536, \\ 0,95, & \text{если } 3,536 < y \le 3,987, \\ 1, & \text{если } y > 3,987. \end{cases}$$

График эмпирической функции распределения изображен на рис. 6.

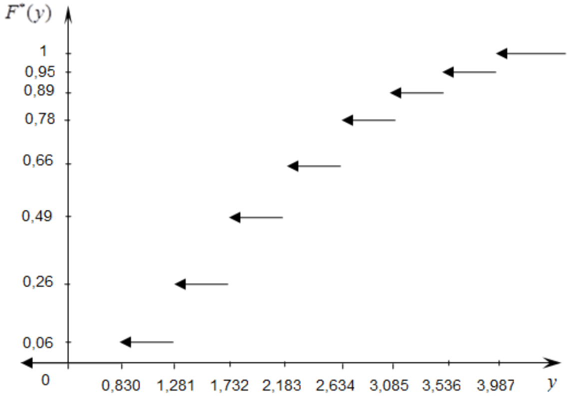


Рис. 6

Для приближения к графику теоретической функции распределения нормального закона распределения построим *кумулянту*. Для построения кумулянты на координатной плоскости отмечаем точки, абсциссы которых являются границами интервалов, а ординаты равны накопленным относительным частотам: (0,830; 0), (1,281; 0,06), (1,732; 0,26), (2,183; 0,49), (2,634; 0,66), (3,085; 0,78), (3,536; 0,89), (3,987; 0,95). Соединим эти точки отрезками, полученная ломаная линия является кумулянтой (рис. 7).

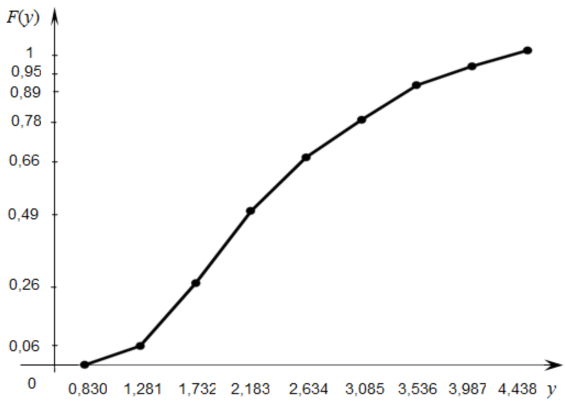


Рис. 7

4. Составим таблицу для вычисления выборочных числовых характеристик случайной величины *Y*.

№	Интервалы	Часто-ты	Середины интерва-лов	$(y_i-\overline{y})$	$(y_i-\overline{y})m_i$	$(y_i-\overline{y})^2 m_i$	$(y_i-\overline{y})^3 m_i$	$(y_i-\overline{y})^4 m_i$
1	[0,830; 1,281)	6	1,056	–1,312	–7,874	10,335	–13,563	17,800
2	[1,281; 1,732)	20	1,507	–0,861	–17,228	14,841	–12,784	11,012
3	[1,732; 2,183)	23	1,958	–0,410	–9,439	3,874	–1,590	0,653
4	[2,183; 2,634)	17	2,409	0,041	0,690	0,028	0,001	0,000
5	[2,634; 3,085)	12	2,860	0,492	5,899	2,900	1,426	0,701
6	[3,085; 3,536)	11	3,311	0,943	10,368	9,773	9,212	8,683
7	[3,536; 3,987)	6	3,762	1,394	8,362	11,653	16,239	22,630
8	[3,987; 4,438]	5	4,213	1,845	9,223	17,013	31,381	57,885
$\sum$	–	100	–	–	0	70,415	30,322	119,365

Вычислим среднее значение выборки

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{k}y_i m_i ==\frac{1}{100}\cdot(6\cdot 1,056+20\cdot 1,507+23\cdot 1,958+17\cdot 2,409+$$

$$+12\cdot 2,860+11\cdot 3,311+6\cdot 3,762+5\cdot 4,213)=\frac{1}{100}\cdot 236,791\approx 2,368.$$

Вычислим моду по формуле

$$Mo=y_0+h\frac{m_0-m_1}{(m_0-m_1)+(m_0-m_2)},$$

где $y_0=1,732$ – начало модального интервала или интервала с наибольшей частотой;

$h=0,451$ – ширина интервала;

$m_0=23$ – частота модального интервала;

$m_1=20$ – частота интервала, предшествующего модальному;

$m_2=17$ – частота интервала, следующего за модальным.

Тогда $Mo=1,732+0,451\cdot\frac{23-20}{(23-20)+(23-17)}\approx 1,882.$

Вычислим медиану по формуле

$$Me=y_0+h\frac{n-2n_1}{2m_e},$$

где $y_0 = 2{,}183$ – начало медианного интервала или интервала, накопленная частота которого больше половины объема выборки;

$h = 0{,}451$ – ширина интервала;

$n_1 = 23$ – накопленная частота интервала, предшествующего медианному;

$m_e = 17$ – частота медианного интервала.

Тогда $Me = 2{,}183 + 0{,}451 \cdot \dfrac{100 - 2 \cdot 23}{2 \cdot 17} \approx 2{,}899.$

Вычислим выборочную дисперсию и выборочный коэффициент вариации:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k}\left(y_i - \overline{y}\right)^2 m_i = \frac{1}{100} \cdot 70{,}415 \approx 0{,}704$$

и

$$v = \frac{\sqrt{D_{\text{в}}}}{\overline{y}} \cdot 100\ \% = \frac{\sqrt{0{,}704}}{2{,}368} \cdot 100\ \% \approx 35\ \%.$$

Вычислим центральные моменты третьего и четвертого порядков:

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{k}\left(y_i - \overline{y}\right)^3 m_i, \quad \mu_3 = \frac{1}{100} \cdot 30{,}322 \approx 0{,}303;$$

$$\mu_4 = \frac{1}{n}\sum_{i=1}^{k}\left(y_i - \overline{y}\right)^4 m_i, \quad \mu_4 = \frac{1}{100} \cdot 119{,}365 \approx 1{,}194.$$

Вычислим асимметрию и эксцесс нормального распределения:

$$A = \frac{\mu_3}{\sqrt{D_{\text{в}}^3}} = \frac{0{,}303}{\sqrt{\left(0{,}704\right)^3}} \approx 0{,}513;$$

$$E = \frac{\mu_4}{\sqrt{D_{\text{в}}^4}} - 3 = \frac{1{,}194}{\sqrt{\left(0{,}704\right)^4}} - 3 \approx -0{,}591.$$

Вычислим утроенные ошибки репрезентативности асимметрии и эксцесса: $3 \cdot \sqrt{\dfrac{6}{n}} = 3 \cdot \sqrt{\dfrac{6}{100}} \approx 0{,}735;$ $3\sqrt{\dfrac{24}{n}} = \sqrt{\dfrac{24}{100}} \approx 1{,}47.$

Анализируя полученные результаты:

$$A = 0{,}513 > 0,\ \ \bar{y} < Me\ (2{,}368 < 2{,}899) \text{ и } \bar{y} > Mo\ (2{,}368 > 1{,}882),$$

отметим, что данное выборочное распределение имеет незначительную правостороннюю асимметрию. Полученное значение эксцесса $E = -0{,}591$ отрицательное, поэтому выборочное распределение низковершинное. Значения асимметрии и эксцесса не превышают допустимой ошибки $|A| \le 0{,}735$ и $|E| \le 1{,}47,$ поэтому выборочное распределение подчиняется нормальному закону распределения.

5. Подходящей точечной оценкой параметра a нормального распределения служит среднее значение выборки $\bar{y} = 2{,}368,$ а параметра σ^2 – исправленная выборочная дисперсия

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{100}{99} \cdot 0{,}704 \approx 0{,}711.$$

Тогда с некоторым приближением параметр

$$\sigma_y = \sqrt{S^2} = \sqrt{0{,}711} \approx 0{,}843.$$

Найдем доверительный интервал параметра a с надежностью $\gamma = 0{,}95.$

Вычислим точность интервальной оценки параметра a:

$$\delta = t \frac{\sigma}{\sqrt{n}},$$

где значение t находим из уравнения $2\Phi(t) = 0{,}95$ или $\Phi(t) = \dfrac{0{,}95}{2} = 0{,}475,$ из прил. 1 находим $t = 1{,}96.$

Тогда $\delta = 1{,}96 \cdot \dfrac{0{,}843}{\sqrt{100}} \approx 0{,}165,$ а границы доверительного интервала вычисляются как $\bar{y} - \delta = 2{,}368 - 0{,}165 = 2{,}203$ и $\bar{y} + \delta = 2{,}368 + 0{,}165 =$ $= 2{,}533.$ Итак, доверительный интервал (2,203; 2,533) включает параметр a случайной величины Y с надежностью $\gamma = 0{,}95$.

Задание 2. Проверка статистических гипотез.

По статистическим выборочным данным случайной величины X или Y требуется:

1. Выдвинуть нулевую гипотезу H_0 о нормальном законе распределения случайных величин. Записать функцию плотности вероятностей и точечные оценки параметров нормального распределения.

2. Рассчитать теоретические частоты нормального распределения и вычислить наблюдаемое значение статистики критерия согласия Пирсона

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(m_i - m_i^0)^2}{m_i^0} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

3. Для уровня значимости $\alpha = 0{,}05$ по таблице критических точек распределения Пирсона (прил. 3) найти границу правосторонней критической области $\chi^2_{\alpha, s}$.

4. Если $\chi^2_{\text{набл}} < \chi^2_{\alpha, s}$, то выдвинутую нулевую гипотезу H_0 о нормальном законе распределения случайной величины принять. Если $\chi^2_{\text{набл}} > \chi^2_{\alpha, s}$, то нулевую гипотезу отвергнуть. Построить графики эмпирических и теоретических кривых нормального распределения.

Решение типового примера

Выполним типовой пример для выборочных данных задания 1, случайных величин X – стоимости валовой продукции (тыс. руб/га) и Y – стоимость оборотных фондов (тыс. руб/га).

Случайная величина X

1. Выдвигаем нулевую гипотезу H_0 о нормальном законе распределения случайной величины X. Функция плотности распределения вероятностей нормального закона имеет вид

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

Точечными оценками параметров нормального распределения a и σ являются вычисленные в задании 1 среднее значение выборки

$\overline{x} = 2,108$ и среднее квадратическое отклонение $\sigma_x = 0,782$. Тогда функция плотности примет вид

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot 0,782} e^{-\frac{(x-2,108)^2}{2\cdot(0,782)^2}} = \frac{1}{0,782\cdot\sqrt{2\pi}} e^{-\frac{(x-2,108)^2}{1,223}}.$$

2. В соответствии с предполагаемым нормальным законом распределения и на основании интервального ряда случайной величины X запишем интервалы: $(-\infty;\ 1,306)$, $[1,306;\ 1,764)$, $[1,764;\ 2,222)$, $[2,222;\ 2,68)$, $[2,680;\ 3,138)$, $[3,138;\ 3,596)$, $[3,596;\ +\infty)$.

Первый и второй интервалы объединены в один, поскольку $m_1 = 2 < 5$. Рассчитываем вероятности попадания случайной величины X в эти интервалы по формулам:

$$p_i = P(x'_{i-1} < X < x'_i) = \Phi\left(\frac{x'_i - \overline{x}}{\sigma_x}\right) - \Phi\left(\frac{x'_{i-1} - \overline{x}}{\sigma_x}\right), \quad i = \overline{1,\ 7}.$$

$$p_1 = P(-\infty < X < 1,306) = \Phi\left(\frac{1,306-2,108}{0,782}\right) - \Phi(-\infty) = -\Phi(1,03) + 0,5 =$$

$$= -0,3485 + 0,5 = 0,1515;$$

$$p_2 = P(1,306 < X < 1,764) = \Phi\left(\frac{1,764-2,108}{0,782}\right) - \Phi\left(\frac{1,306-2,108}{0,782}\right) =$$

$$= -\Phi(0,44) + \Phi(1,03) = -0,1700 + 0,3485 = 0,1785;$$

$$p_3 = P(1,764 < X < 2,222) = \Phi\left(\frac{2,222-2,108}{0,782}\right) - \Phi\left(\frac{1,764-2,108}{0,782}\right) =$$

$$= \Phi(0,15) + \Phi(0,44) = 0,0596 + 0,1700 = 0,2296;$$

$$p_4 = P(2,222 < X < 2,680) = \Phi\left(\frac{2,680-2,108}{0,782}\right) - \Phi\left(\frac{2,222-2,108}{0,782}\right) =$$

$$= \Phi(0,73) - \Phi(0,15) = 0,2673 - 0,0596 = 0,2077;$$

$$p_5 = P(2,680 < X < 3,138) = \Phi\left(\frac{3,138-2,108}{0,782}\right) - \Phi\left(\frac{2,680-2,108}{0,782}\right) =$$

$$= \Phi(1,32) - \Phi(0,73) = 0,4066 - 0,2673 = 0,1393;$$

$$p_6 = P(3{,}138 < X < 3{,}596) = \Phi\left(\frac{3{,}596-2{,}108}{0{,}782}\right) - \Phi\left(\frac{3{,}138-2{,}108}{0{,}782}\right) =$$

$$= \Phi(1{,}90) - \Phi(1{,}32) = 0{,}4713 - 0{,}4066 = 0{,}0647;$$

$$p_7 = P(3{,}596 < X < +\infty) = \Phi(+\infty) - \Phi\left(\frac{3{,}596-2{,}108}{0{,}782}\right) =$$

$$= 0{,}5 - \Phi(1{,}90) = 0{,}5 - 0{,}4713 = 0{,}0287.$$

Тогда теоретические частоты попадания случайной величины в записанные интервалы равны $m_i^0 = np_i$, $i = \overline{1,\ 7}$.

№	Интервалы	m_i	x_i'	x_{i+1}'	$u_i = \dfrac{x_i' - \overline{x}}{\sigma_x}$	$u_{i+1} = -\dfrac{x_{i+1}' - \overline{x}}{\sigma_x}$	$\Phi(u_i)$	$\Phi(u_{i+1})$	p_i	m_i^0
1	$(-\infty;\ 1{,}306)$	14	$-\infty$	1,306	$-\infty$	–1,03	–0,5	–0,3485	0,1515	15,15
2	$[1{,}306;\ 1{,}764)$	25	1,306	1,764	–1,03	–0,44	–0,3485	–0,1700	0,1785	17,85
3	$[1{,}764;\ 2{,}222)$	20	1,764	2,222	–0,44	0,15	–0,1700	0,0596	0,2296	22,96
4	$[2{,}222;\ 2{,}680)$	17	2,222	2,680	0,15	0,73	0,0596	0,2673	0,2077	20,77
5	$[2{,}680;\ 3{,}138)$	12	2,680	3,138	0,73	1,32	0,2673	0,4066	0,1393	13,93
6	$[3{,}138;\ 3{,}596)$	7	3,138	3,596	1,32	1,90	0,4066	0,4713	0,0647	6,47
7	$[3{,}596;\ +\infty)$	5	3,596	$+\infty$	1,90	$+\infty$	0,4713	0,5	0,0287	2,87
$\sum$	–	100	–	–	–	–	–	–	1	100

Вычисляем наблюдаемое значение статистики согласно следующей таблице

$$\chi^2_{\text{набл}} = \sum_{i=1}^{7} \frac{(m_i - m_i^0)^2}{m_i^0} = \sum_{i=1}^{7} \frac{(m_i - np_i)^2}{np_i}.$$

№	Эмпирические частоты m_i	Теоретические частоты m_i^0	$m_i - m_i^0$	$(m_i - m_i^0)^2$	$\dfrac{(m_i - m_i^0)^2}{m_i^0}$
1	14	15,15	–1,15	1,3225	0,0873
2	25	17,85	7,15	51,1225	2,8640
3	21	22,96	–1,96	3,8416	0,1673
4	17	20,77	–3,77	14,2129	0,6843
5	11	13,93	–2,93	8,5849	0,6163
6	7	6,47	0,53	0,2809	0,0434
7	5	2,87	2,13	4,5369	1,5808
Σ	100	100	–	–	6,0434

3. Задаем уровень значимости $\alpha = 0{,}05$ и вычисляем число степеней свободы $s = k - 3 = 7 - 3 = 4$. По таблице критических точек распределения χ^2 (прил. 3) находим границу правосторонней критической области, критическую точку $\chi^2_{\alpha,\, s} = 9{,}5.$

4. Так как $\chi^2_{\text{набл}} = 6{,}0434 < 9{,}5,$ то выдвинутую нулевую гипотезу о нормальном законе распределения случайной величины принимаем. Другими словами, расхождения между эмпирическими и теоретическими частотами являются незначительными, т. е. случайными, и предположение о распределении случайной величины X по нормальному закону вполне согласуется с эмпирическим распределением выборки.

Строим графики эмпирических и теоретических кривых нормального распределения (рис. 8).

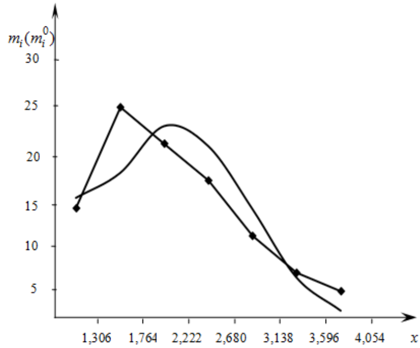


Рис. 8

Случайная величина Y

1. Выдвигаем нулевую гипотезу H_0 о нормальном законе распределения случайной величины Y.

Функция плотности распределения вероятностей нормального закона имеет вид

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

Точечными оценками параметров нормального распределения a и σ являются вычисленные в задании 1 среднее значение выборки $\overline{y} = 2{,}368$ и среднее квадратическое отклонение $\sigma_x = 0{,}843$. Тогда функция плотности примет вид

$$f(y) = \frac{1}{\sqrt{2\pi} \cdot 0{,}843} e^{-\frac{(y-2{,}368)^2}{2\cdot(0{,}843)^2}} = \frac{1}{0{,}843 \cdot \sqrt{2\pi}} e^{-\frac{(y-2{,}368)^2}{1{,}421}}.$$

2. В соответствии с предполагаемым нормальным законом распределения и на основании интервального ряда случайной величины Y

запишем интервалы: (–∞; 1,281), [1,281;1,732), [1,732; 2,183), [2,183; 2,634); [2,634; 3,085); [3,085; 3,536), [3,536; 3,987), [3,987; +∞).

Рассчитываем вероятности попадания случайной величины Y в эти интервалы по формулам:

$$p_i = P\left(y'_{i-1} < Y < y'_i\right) = \Phi\left(\frac{y'_i - \overline{y}}{\sigma_y}\right) - \Phi\left(\frac{y'_{i-1} - \overline{y}}{\sigma_y}\right),\ i = \overline{1,8}.$$

$$p_1 = P\left(-\infty < Y < 1,281\right) = \Phi\left(\frac{1,281 - 2,368}{0,843}\right) - \Phi\left(-\infty\right) = -\Phi\left(1,29\right) + 0,5 =$$
$$= -0,4015 + 0,5 = 0,0985;$$

$$p_2 = P\left(1,281 < Y < 1,732\right) = \Phi\left(\frac{1,732 - 2,368}{0,843}\right) - \Phi\left(\frac{1,281 - 2,368}{0,843}\right) =$$
$$= -\Phi\left(0,75\right) + \Phi\left(1,29\right) = -0,2734 + 0,4015 = 0,1281;$$

$$p_3 = P\left(1,732 < Y < 2,183\right) = \Phi\left(\frac{2,183 - 2,368}{0,843}\right) - \Phi\left(\frac{1,732 - 2,368}{0,843}\right) =$$
$$= -\Phi\left(0,22\right) + \Phi\left(0,75\right) = -0,0871 + 0,2734 = 0,1863;$$

$$p_4 = P\left(2,183 < Y < 2,634\right) = \Phi\left(\frac{2,634 - 2,368}{0,843}\right) - \Phi\left(\frac{2,183 - 2,368}{0,843}\right) =$$
$$= \Phi\left(0,32\right) + \Phi\left(0,22\right) = 0,1255 + 0,0871 = 0,2126;$$

$$p_5 = P\left(2,634 < Y < 3,085\right) = \Phi\left(\frac{3,085 - 2,368}{0,843}\right) - \Phi\left(\frac{2,634 - 2,368}{0,843}\right) =$$
$$= \Phi\left(0,85\right) - \Phi\left(0,32\right) = 0,3023 - 0,1255 = 0,1768;$$

$$p_6 = P\left(3,085 < Y < 3,536\right) = \Phi\left(\frac{3,536 - 2,368}{0,843}\right) - \Phi\left(\frac{3,085 - 2,368}{0,843}\right) =$$
$$= \Phi\left(1,39\right) - \Phi\left(0,85\right) = 0,4177 - 0,3023 = 0,1154;$$

$$p_7 = P\left(3,536 < Y < 3,987\right) = \Phi\left(\frac{3,987 - 2,368}{0,843}\right) - \Phi\left(\frac{3,536 - 2,368}{0,843}\right) =$$
$$= \Phi\left(1,92\right) - \Phi\left(1,39\right) = 0,4726 - 0,4177 = 0,0549;$$

$$p_8 = P\left(3,987 < Y < +\infty\right) = \Phi\left(+\infty\right) - \Phi\left(\frac{3,987 - 2,368}{0,843}\right) =$$
$$= 0,5 - \Phi\left(1,92\right) = 0,5 - 0,4726 = 0,0274.$$

Тогда теоретические частоты попадания случайной величины в записанные интервалы равны $m_i^0 = np_i$, $i = \overline{1,\,8}$.

№	Интервалы	m_i	y'_i	y'_{i+1}	$u_i = \dfrac{y'_i - \overline{y}}{\sigma_y}$	$u_{i+1} = \dfrac{y'_{i+1} - \overline{y}}{\sigma_y}$	$\Phi(u_i)$	$\Phi(u_{i+1})$	p_i	m_i^0
1	$(-\infty;\ 1{,}281)$	6	$-\infty$	1,281	$-\infty$	–1,29	–0,5	–0,4015	0,0985	9,85
2	$[1{,}281;\ 1{,}732)$	20	1,281	1,732	–1,29	–0,75	–0,4015	–0,2734	0,1281	12,81
3	$[1{,}732;\ 2{,}183)$	23	1,732	2,183	–0,75	–0,22	–0,2734	–0,0871	0,1863	18,63
4	$[2{,}183;\ 2{,}634)$	17	2,183	2,634	–0,22	0,32	–0,0871	0,1255	0,2126	21,26
5	$[2{,}634;\ 3{,}085)$	12	2,634	3,085	0,32	0,85	0,1255	0,3023	0,1768	17,68
6	$[3{,}085;\ 3{,}536)$	11	3,085	3,536	0,85	1,39	0,3023	0,4177	0,1154	11,54
7	$[3{,}536;\ 3{,}987)$	6	3,536	3,987	1,39	1,92	0,4177	0,4726	0,0549	5,49
8	$[3{,}987;\ +\infty)$	5	3,987	$+\infty$	1,92	$+\infty$	0,4726	0,5	0,0274	2,74
Σ	–	100	–	–	–	–	–	–	1	100

Вычисляем наблюдаемое значение статистики

$$\chi^2_{\text{набл}} = \sum_{i=1}^{8} \frac{(m_i - m_i^0)^2}{m_i^0} = \sum_{i=1}^{8} \frac{(m_i - np_i)^2}{np_i}.$$

№	Эмпирические частоты m_i	Теоретические частоты m_i^0	$m_i - m_i^0$	$(m_i - m_i^0)^2$	$\dfrac{(m_i - m_i^0)^2}{m_i^0}$
1	6	9,85	–3,85	14,8225	1,5048
2	20	12,81	7,19	51,6961	4,0356
3	23	18,63	4,37	19,0969	1,0251
4	17	21,26	–4,26	18,1476	0,8536
5	12	17,68	–5,68	32,2624	1,8248
6	11	11,54	–0,54	0,2916	0,0253
7	6	5,49	0,51	0,2601	0,0474
8	5	2,74	2,26	5,1076	1,8641
Σ	100	100	–	–	11,1806

3. Задаем уровень значимости $\alpha = 0{,}05$ и вычисляем число степеней свободы $s = k - 3 = 8 - 3 = 5$. По таблице критических точек распределения χ^2 (прил. 3) находим границу правосторонней критической области, критическую точку $\chi^2_{\alpha,\,s} = 11{,}1$.

4. Так как $\chi^2_{\text{набл}} = 11{,}1806 > 11{,}1,$ то выдвинутую нулевую гипотезу о нормальном законе распределения случайной величины отвергаем. Строим графики эмпирических и теоретических кривых нормального распределения (рис. 9).

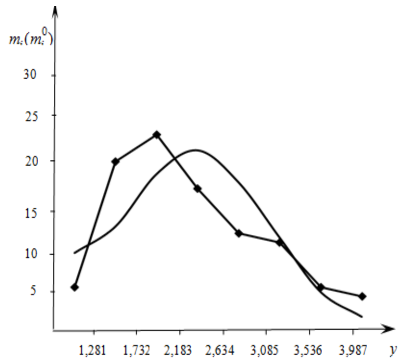


Рис. 9

Задание 3. Корреляция.

1. По результатам задания 1 составить корреляционную таблицу.

2. Найти условные средние $\overline{y}_x$ и $\overline{x}_y$. На координатную плоскость нанести точки $\left(x_i;\ \overline{y}_i\right)$, $i = \overline{1,\ k}$ и $\left(\overline{x}_j;\ y_j\right)$, $j = \overline{1,\ k}$. По характеру расположения точек подобрать подходящий вид функции регрессии.

3. Найти выборочный коэффициент линейной корреляции $r_{\text{в}}$, построить доверительный интервал для коэффициента линейной корреляции между случайными величинами X и Y с надежностью $\gamma = 0{,}95$ и оценить силу корреляционной зависимости между случайными величинами.

4. Найти параметры и составить уравнения линейной регрессии y на x и x на y, построить линии регрессии.

Решение типового примера

1. Данные выборочных наблюдений над случайными величинами X и Y группируем и записываем в виде корреляционной таблицы. Для этого просматриваем все 100 пар значений исходной выборки и проводим подсчет частот.

Суммы частот в последних строке и столбце равны объему выборки: $\sum_{i=1}^{8} m_{x_i} = \sum_{j=1}^{8} m_{y_j} = 100$. Первый и последний столбец таблицы соответствуют интервальному ряду распределения X, а первая и последняя строка – Y.

Случайные величины		Y								
		[0,83; 1,281)	[1,281; 1,732)	[1,281; 1,732)	[1,732; 2,183)	[2,183; 3,085)	[3,085; 3,536)	[3,536; 3,987)	[3,987; 4,438]	m_x
X	[0,39; 0,848)	II								2
	[0,848; 1,306)	IIII	IIIIIIII							12
	[1,306; 1,764)		IIIIIIIIIIII	IIIIIIIIIIIII						25
	[1,764; 2,222)			IIIIIIIIII	IIIIIIIIIII					21
	[2,222; 2,68)				IIIIII	IIIIIIIIIII				17
	[2,68; 3,138)					I	IIIIIIIIII			11
	[3,138; 3,596)						I	IIIIII		7
	[3,596; 4,054]								IIIII	5
	m_y	6	20	23	17	12	11	6	5	100

Записываем корреляционную таблицу, причем вместо интервалов запишем середины интервалов, округлив значения до сотых долей.

Случайные величины		Y								
		1,06	1,51	1,96	2,41	2,86	3,31	3,76	4,21	m_x
X	0,62	2								2
	1,08	4	8							12
	1,54		12	13						25
	1,99			10	11					21
	2,45				6	11				17
	2,91					1	10			11
	3,37						1	6		7
	3,83								5	5
	m_y	6	20	23	17	12	11	6	5	100

2. В данной таблице можно выделить 8 условных законов распределения Y при $X = 0{,}62$; $X = 1{,}08$; $X = 1{,}54$; $X = 1{,}99$; $X = 2{,}45$; $X = 2{,}91$; $X = 3{,}37$; $X = 3{,}83$ и вычислить условные средние по формулам:

$$\overline{y}_{x_1} = \frac{1}{m_{x_1}} \sum_{j=1}^{8} y_j m_{1j} = \frac{1{,}06 \cdot 2}{2} = 1{,}06;$$

$$\overline{y}_{x_2} = \frac{1}{m_{x_2}} \sum_{j=1}^{8} y_j m_{2j} = \frac{1}{12} \cdot \left(1{,}06 \cdot 4 + 1{,}51 \cdot 8\right) = 1{,}36;$$

$$\overline{y}_{x_3} = \frac{1}{m_{x_3}} \sum_{j=1}^{8} y_j m_{3j} = \frac{1}{25} \cdot \left(1{,}51 \cdot 12 + 1{,}96 \cdot 13\right) \approx 1{,}74;$$

$$\overline{y}_{x_4} = \frac{1}{m_{x_4}} \sum_{j=1}^{8} y_j m_{4j} = \frac{1}{21} \cdot \left(1{,}96 \cdot 10 + 2{,}41 \cdot 11\right) \approx 2{,}20;$$

$$\overline{y}_{x_5} = \frac{1}{m_{x_5}} \sum_{j=1}^{8} y_j m_{5j} = \frac{1}{17} \cdot \left(2{,}41 \cdot 6 + 2{,}86 \cdot 11\right) \approx 2{,}70;$$

$$\overline{y}_{x_6} = \frac{1}{m_{x_6}} \sum_{j=1}^{8} y_j m_{6j} = \frac{1}{11} \cdot \left(2{,}86 \cdot 1 + 3{,}31 \cdot 10\right) \approx 3{,}27;$$

$$\overline{y}_{x_7} = \frac{1}{m_{x_7}} \sum_{j=1}^{8} y_j m_{7j} = \frac{1}{7} \cdot \left(3{,}31 \cdot 1 + 3{,}76 \cdot 6\right) \approx 3{,}70;$$

$$\overline{y}_{x_8} = \frac{1}{m_{x_8}} \sum_{j=1}^{8} y_j m_{8j} = \frac{4{,}21 \cdot 5}{5} = 4{,}21.$$

Результаты заносим в таблицу.

x	0,62	1,08	1,54	1,99	2,45	2,91	3,37	3,83
$\overline{y}_x$	1,06	1,36	1,74	2,20	2,7	3,27	3,7	4,21

В этом случае получим взаимно однозначное соответствие между значениями случайной величины X и условными средними величины Y. Эту корреляционную зависимость можно записать уравнением регрессии: $\overline{y}_x = f(x)$. Для определения вида функции регрессии на координатную плоскость наносим точки $(x_1;\, \overline{y}_1)$, $(x_2;\, \overline{y}_2)$, $\ldots, (x_k;\, \overline{y}_k)$ и по характеру расположения точек вдоль прямой линии выбираем прямую регрессию (рис. 10).

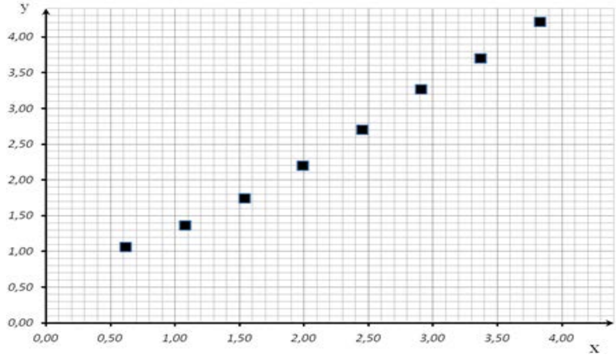


Рис. 10

Затем в корреляционной таблице выделяем 8 условных законов распределения X при Y =1,06; Y = 1,51; Y = 1,96; Y = 2,41; Y = 2,86; Y = 3,31; Y = 3,76; Y = 4,21 соответственно и вычисляем условные средние:

$$\bar{x}_{y_1} = \frac{1}{m_{y_1}} \sum_{i=1}^{8} x_i m_{i1} = \frac{1}{6} \cdot (0,62 \cdot 2 + 1,08 \cdot 4) \approx 0,93;$$

$$\bar{x}_{y_2} = \frac{1}{m_{y_2}} \sum_{i=1}^{8} x_i m_{i2} = \frac{1}{20} \cdot (1,08 \cdot 8 + 1,54 \cdot 12) \approx 1,36;$$

$$\bar{x}_{y_3} = \frac{1}{m_{y_3}} \sum_{i=1}^{8} x_i m_{i3} = \frac{1}{23} \cdot (1,54 \cdot 13 + 1,99 \cdot 10) \approx 1,74;$$

$$\bar{x}_{y_4} = \frac{1}{m_{y_4}} \sum_{i=1}^{8} x_i m_{i4} = \frac{1}{17} \cdot (1,99 \cdot 11 + 2,45 \cdot 6) \approx 2,15;$$

$$\bar{x}_{y_5} = \frac{1}{m_{y_5}} \sum_{i=1}^{8} x_i m_{i5} = \frac{1}{12} \cdot (2,45 \cdot 11 + 2,91 \cdot 1) \approx 2,49;$$

$$\bar{x}_{y_6} = \frac{1}{m_{y_6}} \sum_{i=1}^{8} x_i m_{i6} = \frac{1}{11} \cdot (2,91 \cdot 10 + 3,37 \cdot 1) \approx 2,95;$$

$$\bar{x}_{y_7} = \frac{1}{m_{y_7}} \sum_{i=1}^{8} x_i m_{i7} = \frac{3,37 \cdot 6}{6} = 3,37;$$

$$\bar{x}_{y_8} = \frac{1}{m_{y_8}} \sum_{i=1}^{8} x_i m_{i8} = \frac{3,83 \cdot 5}{5} = 3,83.$$

Результаты заносим в таблицу.

y	1,06	1,51	1,96	2,41	2,86	3,31	3,76	4,21
$\overline{x}_y$	0,93	1,36	1,74	2,15	2,49	2,95	3,37	3,83

В этом случае получим взаимно однозначное соответствие между значениями случайной величины Y и условными средними величины X. Эту корреляционную зависимость можно записать уравнением регрессии: $\overline{x}_y = \varphi(y)$. Для определения вида функции регрессии на координатную плоскость наносим точки $(\overline{x}_1;\ y_1)$, $(\overline{x}_2;\ y_2)$, …, $(\overline{x}_k;\ y_k)$ и по характеру расположения точек вдоль прямой линии выбираем прямую регрессию (рис. 11).

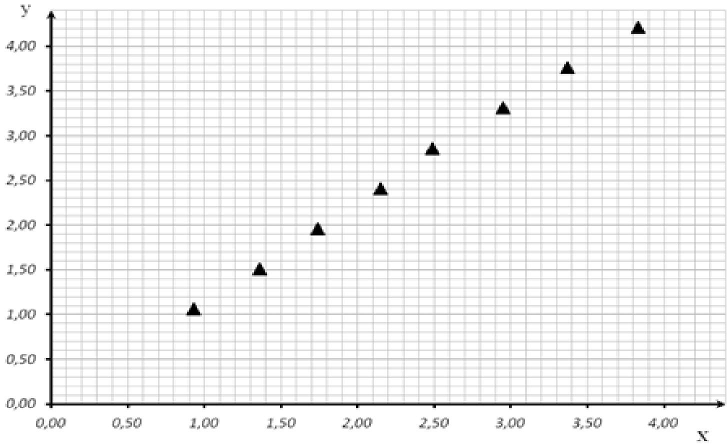


Рис. 11

3. Выборочный коэффициент линейной корреляции служит оценкой меры тесноты линейной корреляционной зависимости между случайными величинами и выражается формулой:

$$r_{\text{в}} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}.$$

Из расчетов задания 1 выбираем средние значения выборок $\overline{x} = 2{,}108$, $\overline{y} = 2{,}368$ и средние квадратические отклонения случайных величин X и Y $\sigma_x = 0{,}782$, $\sigma_y = 0{,}843$.

Величина $\overline{xy}$ вычисляется по корреляционной таблице:

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{k} x_i y_j m_{ij} = \frac{1}{100}\Big(\ 0{,}62\cdot 1{,}06\cdot 2 + 1{,}08\cdot(\ 1{,}06\cdot 4 + 1{,}51\cdot 8\ ) + 1{,}54\times$$
$$\times(1{,}51\cdot 12 + 1{,}96\cdot 13) + 1{,}99\cdot(1{,}96\cdot 10 + 2{,}41\cdot 11) + 2{,}45\cdot(2{,}41\cdot 6 + 2{,}86\cdot 11) +$$
$$+2{,}91\cdot(2{,}86\cdot 1 + 3{,}31\cdot 10) + 3{,}37\cdot(3{,}31\cdot 1 + 3{,}76\cdot 6) + 3{,}83\cdot 4{,}21\cdot 5\Big) = 5{,}627939.$$

Тогда коэффициент корреляции будет равен

$$r_{\text{в}} = \frac{5{,}627939 - 2{,}108\cdot 2{,}368}{0{,}782\cdot 0{,}843} = \frac{0{,}636195}{0{,}659226} \approx 0{,}965.$$

Так как $|r_{\text{в}}| \geq 0{,}7$, то между случайными величинами существует тесная, прямая линейная корреляционная зависимость.

Доверительный интервал для коэффициента линейной корреляции между случайными величинами X и Y с заданной надежностью $\gamma = 0{,}95$ находим по формуле

$$\left(r_{\text{в}} - t\frac{1 - r_{\text{в}}^2}{\sqrt{n}};\ r_{\text{в}} + t\frac{1 - r_{\text{в}}^2}{\sqrt{n}}\right),$$

где значение t находим из уравнения

$$2\Phi(t) = 0{,}95,\ \text{или}\ \Phi(t) = \frac{0{,}95}{2} = 0{,}475.$$

Тогда из прил. 1 следует, что $t = 1{,}96$.

Получаем $\left(0{,}965 - 1{,}96\cdot\dfrac{1 - 0{,}965^2}{\sqrt{100}};\ 0{,}965 + 1{,}96\cdot\dfrac{1 - 0{,}965^2}{\sqrt{100}}\right)$, или (0,952; 0,978).

4. Найдем линейные уравнения функций регрессии. Уравнения регрессии y на x и x на y соответственно имеют вид:

$$\overline{y}_x = r_в \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right) + \overline{y} \text{ и } \overline{x}_y = r_в \frac{\sigma_x}{\sigma_y}\left(y - \overline{y}\right) + \overline{x}.$$

Подставим имеющиеся данные: $\overline{x} = 2{,}108$, $\overline{y} = 2{,}368$, $\sigma_x = 0{,}782$, $\sigma_y = 0{,}843$, $r_в = 0{,}965$, имеем: $\overline{y}_x = 0{,}965 \cdot \frac{0{,}843}{0{,}782} \cdot \left(x - 2{,}108\right) + 2{,}368$, преобразуя, получим: $\overline{y}_x = 1{,}04x + 0{,}18$.

Аналогично $\overline{x}_y = 0{,}965 \cdot \frac{0{,}782}{0{,}843} \cdot \left(y - 2{,}368\right) + 2{,}108$, или $\overline{x}_y = 0{,}9y - 0{,}02$.

Построим эмпирические и теоретические зависимости на одном чертеже, квадратные метки отражают зависимость $\overline{y}_x = f\left(x\right)$, треугольные – $\overline{x}_y = \varphi\left(y\right)$. Прямые регрессии проходят через точку $C\left(\overline{x}; \overline{y}\right) = \left(2{,}108; 2{,}368\right)$ (рис. 12).

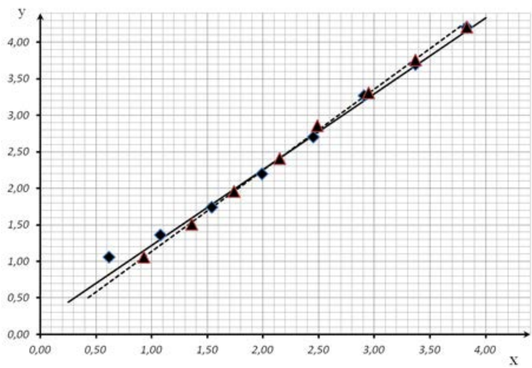


Рис. 12

Задание 4. Вычисление числовых характеристик статистических данных бонитировки почв сельскохозяйственных угодий.

На основании выборочных статистических данных бонитировки почв сельскохозяйственных угодий районов Республики Беларусь, приведенных в прил. 4 требуется:

1. Составить интервальные статистические ряды распределения частот и относительных частот.

2. Построить гистограмму и полигон частот.

3. Вычислить числовые характеристики выборки: среднее значение выборки, моду, медиану, выборочную дисперсию, выборочное среднее квадратическое отклонение.

Решение типового примера

Приведены 50 данных по бонитировке почв сельскохозяйственных угодий районов Республики Беларусь: 31, 40, 36, 33, 35, 36, 40, 37, 30, 45, 34, 34, 39, 40, 37, 35, 48, 38, 41, 33, 41, 37, 37, 40, 34, 44, 44, 34, 43, 35, 44, 39, 44, 36, 34, 35, 41, 41, 30, 36, 34, 32, 34, 35, 35, 39, 38, 35, 36, 36.

1. Построение интервального вариационного ряда начинается с разбиения интервала изменения случайной величины на k частичных интервалов одинаковой ширины и подсчета частот попадания случайной величины в каждый из этих интервалов.

Для определения числа интервалов k можно пользоваться формулой

$$k = 1 + \log_2 n \approx 1 + 3{,}32 \cdot \lg n,$$

где n – объем выборки.

При n = 50 получаем $k \approx 1 + 3{,}32 \cdot \lg 50 \approx 6{,}64 \approx 7$.

Определим границы интервалов. Для этого находим размах выборки $R = x_{\max} - x_{\min}$. В нашем случае $x_{\max} = 48$, $x_{\min} = 30$, $R = 48 - 30 = 18$. Ширину интервала определяем по формуле

$$h = \frac{x_{\max} - x_{\min}}{k-1} = \frac{48-30}{7-1} = \frac{18}{6} = 3.$$

Начало интервала выбираем на 0,5 ширины интервала левее $x_{\min}$, а конец последнего на 0,5 ширины интервала правее $x_{\max}$.

Находим границы интервалов:

$$x'_0 = x_{\min} - \frac{h}{2} = 30 - 1{,}5 = 28{,}5;\ x'_1 = x'_0 + h = 28{,}5 + 3 = 31{,}5;$$

$$x'_2 = x'_1 + h = 31{,}5 + 3 = 34{,}5;\ x'_3 = x'_2 + h = 34{,}5 + 3 = 37{,}5;$$

$$x'_4 = x'_3 + h = 37{,}5 + 3 = 40{,}5;\ x'_5 = x'_4 + h = 40{,}5 + 3 = 43{,}5;$$

$$x'_6 = x'_5 + h = 43{,}5 + 3 = 46{,}5;\ x'_7 = x'_6 + h = 46{,}5 + 3 = 49{,}5.$$

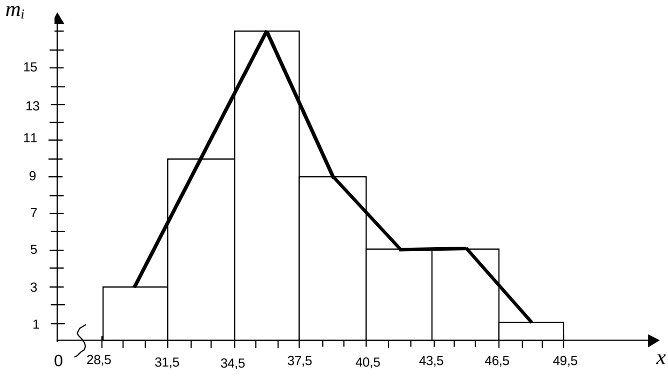
Записываем сами интервалы: [28,5; 31,5), [31,5; 34,5), [34,5; 37,5), [37,5; 40,5), [40,5; 43,5), [43,5; 46,5), [46,5; 49,5].

Просматриваем все выборочные данные случайной величины X и распределяем их по интервалам. Причем в каждый интервал включаем те значения, которые больше или равны нижней границе интервала и меньше верхней границы. Подсчет частот и статистические ряды распределения частот и относительных частот представлены в таблице.

№	Интервалы	Подсчет частот	Частоты m_i	Относительные частоты $w_i = \frac{m_i}{n}$
1	[28,5; 31,5)	III	3	0,06
2	[31,5; 34,5)	IIIIIIIIII	10	0,2
3	[34,5; 37,5)	IIIIIIIIIIIIIIIII	17	0,34
4	[37,5; 40,5)	IIIIIIIII	9	0,18
5	[40,5; 43,5)	IIIII	5	0,1
6	[43,5; 46,5)	IIIII	5	0,1
7	[46,5; 49,5]	I	1	0,02
$\sum$	–	–	50	1

2. Для построения гистограммы частот по оси Ox откладываем частичные интервалы и на каждом из них строим прямоугольник, высота которого равна частоте соответствующего интервала. Полученная при этом фигура называется гистограммой частот.

Для построения полигона частот проводим ломаную линию, соединяющую середины верхних оснований прямоугольников (рис. 13).



Интервалы

Рис. 13

Аналогичным образом строится гистограмма и полигон относительных частот.

3. Среднее значение выборки $\bar{x}$ вычисляем по формуле

$$\bar{x}_{в} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i,$$

где x_i и m_i – соответственно середина и частота i-го интервала $i = \overline{1,7}$.

Середины интервалов определяем по формуле

$$x_i = \frac{x'_{i-1} + x'_i}{2}, \; i = \overline{1,7}.$$

В нашем случае $x_1 = \frac{28,5 + 31,5}{2} = 30$; $x_2 = \frac{31,5 + 34,5}{2} = 33$ и т. д.

Выборочную дисперсию найдем по формуле

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

Расчеты сводим в таблицу.

№	Интервалы	Середины интервалов	Частоты m_i	$x_i m_i$	$x_i^2 m_i$
1	[28,5; 31,5)	30	3	90	2700
2	[31,5; 34,5)	33	10	330	10890
3	[34,5; 37,5)	36	17	612	22032
4	[37,5; 40,5)	39	9	351	13689
5	[40,5; 43,5)	42	5	210	8820
6	[43,5; 46,5)	45	5	225	10125
7	[46,5; 49,5]	48	1	48	2304
$\sum$	–	–	50	1866	70560

Тогда среднее значение выборки $\bar{x} = \frac{1866}{50} = 37,32$.

Выборочная дисперсия $D_{в} = \frac{70560}{50} - (37,32)^2 = 18,4176 \approx 18,42$, а среднее квадратическое отклонение $\sigma_{в} = \sqrt{D_{в}} = \sqrt{18,42} \approx 4,29$.

Моду вычисляем по формуле

$$Mo = x_0 + h\frac{m_0 - m_1}{(m_0 - m_1) + (m_0 - m_2)},$$

где $x_0 = 34{,}5$ – начало модального интервала (интервала с наибольшей частотой);

$h = 3$ – ширина интервала;

$m_0 = 17$ – частота модального интервала;

$m_1 = 10$ – частота интервала, предшествующего модальному;

$m_2 = 9$ – частота интервала, следующего за модальным.

В нашем случае $Mo = 34{,}5 + 3\cdot\frac{17-10}{(17-10)+(17-9)} = 35{,}9.$

Для определения медианы воспользуемся формулой

$$Me = x_0 + h\frac{n - 2n_1}{2m_e},$$

где x_0 – начало медианного интервала (первого из интервалов, у которого накопленная частота больше половины объема выборки);

h – ширина интервала;

n – объем выборки;

n_1 – накопленная частота интервала, предшествующего медианному;

m_e – частота медианного интервала.

Подсчитаем накопленные частоты интервалов и сведем полученные результаты в таблицу.

№	Интервалы	Частоты	Накопленные частоты
1	[28,5; 31,5)	3	3
2	[31,5; 34,5)	10	13
3	[34,5; 37,5)	17	30
4	[37,5; 40,5)	9	39
5	[40,5; 43,5)	5	44
6	[43,5; 46,5)	5	49
7	[46,5; 49,5]	1	50
Σ	–	50	–

Имеем $x_0 = 34{,}5$; $h = 3$; $n = 50$; $n_1 = 13$; $m_e = 17$.

Тогда $Me = 34{,}5 + 3\cdot\frac{25-13}{17} = 36{,}6.$

Задание 5. Уравнения линейной регрессии между двумя рядами статистических данных.

На основании выборочных статистических данных площадей под дорогами Y (%) и распаханности территорий X (%), представленных в прил. 5 корреляционными таблицами, требуется:

1. Найти условные средние $\overline{y}_x$ и $\overline{x}_y$, построить точки $(x_i; \overline{y}_{xi})$, $(\overline{x}_{yj}; y_j)$ и по характеру их расположения подобрать вид функций регрессии.

2. Вычислить коэффициент линейной корреляции и сделать вывод о силе корреляционной связи.

3. Составить уравнения линейной регрессии y на x и x на y, построить линии регрессии.

Решение типового примера

В корреляционной таблице представлены выборочные данные случайных величин Y – площадь под дорогами (%) и X – распаханность территорий (%).

Случайные величины		X						
		39	45	51	57	63	69	m_y
Y	0,6	2	3					5
	0,9		4	5				9
	1,2			2	17	3		22
	1,5			2	6	3		11
	1,8				1	1	1	3
	m_x	2	7	9	24	7	1	$n = 50$

1. Определим условные средние $\overline{y}_{xi}$ по формуле

$$\overline{y}_{xi} = \frac{m_{i1}y_1 + m_{i2}y_2 + \ldots + m_{in}y_n}{m_{i1} + m_{i2} + \ldots + m_{in}} = \frac{1}{m_{xi}} \sum_{j=1}^{n} m_{ij} y_j .$$

$$\overline{y}_{39} = 0,6; \qquad \overline{y}_{45} = \frac{0,6 \cdot 3 + 0,9 \cdot 4}{7} \approx 0,8;$$

$$\overline{y}_{51} = \frac{0,9 \cdot 5 + 1,2 \cdot 2 + 1,5 \cdot 2}{9} = 1,1; \qquad \overline{y}_{57} = \frac{1,2 \cdot 17 + 1,5 \cdot 6 + 1,8}{24} = 1,3;$$

$$\overline{y}_{63} = \frac{1,2 \cdot 3 + 1,5 \cdot 3 + 1,8}{7} \approx 1,4; \qquad \overline{y}_{69} = 1,8.$$

Результаты вычислений занесем в таблицу.

X	39	45	51	57	63	69
Y_x	0,6	0,8	1,1	1,3	1,4	1,8

Аналогичным образом определим условные средние $\bar{x}_{yj}$:

$$\bar{x}_{yj} = \frac{m_{1j}x_1 + m_{2j}x_2 + ... + m_{kj}x_k}{m_{1j} + m_{2j} + ... + m_{kj}} = \frac{1}{m_{yj}} \sum_{i=1}^{k} m_{ij}x_i.$$

Получим:

$$\bar{x}_{0,6} = \frac{39 \cdot 2 + 45 \cdot 3}{5} = 42,6; \qquad \bar{x}_{0,9} = \frac{45 \cdot 4 + 51 \cdot 5}{9} = 48,3;$$

$$\bar{x}_{1,2} = \frac{51 \cdot 2 + 57 \cdot 17 + 63 \cdot 3}{22} \approx 57,3; \quad \bar{x}_{1,5} = \frac{51 \cdot 2 + 57 \cdot 6 + 63 \cdot 3}{11} \approx 57,5;$$

$$\bar{x}_{1,8} = \frac{57 + 63 + 69}{3} = 63.$$

Y	0,6	0,9	1,2	1,5	1,8
X_y	42,6	48,3	57,3	57,5	63

Принимая каждую пару значений $(x_i;\ \bar{y}_{xi})$ и $(\bar{x}_{yj};\ y_j)$ за координаты точек, строим эти точки в прямоугольной системе координат. По характеру расположения построенных точек (рис. 14) сделаем вывод о виде функции регрессии $y(x)$ и $x(y)$.

При этом кружочками покажем точечный график зависимости $y(x)$, треугольниками – $x(y)$.

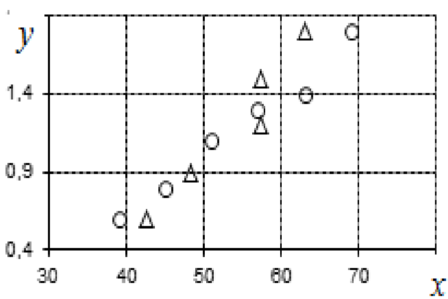


Рис. 14

Характер расположения точек указывает на линейную зависимость между x и $\overline{y}_x$, y и $\overline{x}_y$, поэтому можно предположить, что связь между данными осуществляется по формулам:

$$\overline{y}_x = k_1 x + b_1 \text{ и } \overline{x}_y = k_2 y + b_2.$$

Установим их параметры.

2. Для определения коэффициента корреляции воспользуемся формулой

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y},$$

предварительно вычислив все входящие в нее величины:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_{xi} = \frac{1}{50}\cdot(39\cdot 2 + 45\cdot 7 + 51\cdot 9 + 57\cdot 24 + 63\cdot 7 + 69\cdot 1) \approx 54,6;$$

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{\rho} y_j m_{yj} = \frac{1}{50}\cdot(0,6\cdot 5 + 0,9\cdot 9 + 1,2\cdot 22 + 1,5\cdot 11 + 1,8\cdot 3) \approx 1,2;$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{k} x_i y_i m_{xi} = \frac{1}{50}\cdot(39\cdot 0,6\cdot 2 + 45\cdot(0,6\cdot 3 + 0,9\cdot 4) +$$

$$+51\cdot(0,9\cdot 5 + 1,2\cdot 2 + 1,5\cdot 2) + 57\cdot(1,2\cdot 17 + 1,5\cdot 6 + 1,8\cdot 1) +$$

$$+63\cdot(1,2\cdot 3 + 1,5\cdot 3 + 1,8\cdot 1) + 69\cdot 1,8\cdot 1) = \frac{3321}{50} \approx 66,4;$$

$$D(X) = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 m_{xi} = \frac{1}{50}\cdot((39-54,6)^2\cdot 2 + (45-54,6)^2\cdot 7 +$$

$$+ (51-54,6)^2\cdot 9 + (57-54,6)^2\cdot 24 + (63-54,6)^2\cdot 7 + (69-54,6)^2\cdot 1) =$$

$$= \frac{2088}{50} = 41,76;$$

$$D(Y) = \frac{1}{n}\sum_{j=1}^{k}(y_j - \overline{y})^2 m_{yj} = \frac{1}{50}\cdot((0,6-1,2)^2\cdot 5 + (0,9-1,2)^2\cdot 9 +$$

$$+(1,2-1,2)^2\cdot 22 + (1,5-1,2)^2\cdot 11 + (1,8-1,2)^2\cdot 3) =$$

$$= \frac{1}{50}\cdot(1,8 + 0,81 + 0,99 + 1,08) = \frac{4,68}{50} \approx 0,09;$$

$$\sigma_x = \sqrt{D(X)} \approx 6,5;\ \ \sigma_y = \sqrt{D(Y)} = \sqrt{0,09} = 0,3.$$

Тогда коэффициент корреляции будет равен

$$r = \frac{66,4 - 54,6 \cdot 1,2}{6,5 \cdot 0,3} = \frac{66,4 - 65,52}{1,95} = \frac{0,88}{1,95} = 0,46.$$

Так как коэффициент корреляции r = 0,46, то между площадью дорог и распаханностью территории существует прямая корреляционная зависимость. По величине коэффициента корреляции определяем, что эта связь средняя.

3. Уравнение прямой регрессии $y(x)$ имеет следующий вид:

$$\overline{y}_x - \overline{y} = r\frac{\sigma_y}{\sigma_x}(x - \overline{x}).$$

В нашем случае получим

$$\overline{y}_x - 1,2 = 0,46 \cdot \frac{0,3}{6,46} \cdot (x - 54,6).$$

Упростив, получим зависимость $y(x)$:

$$\overline{y}_x = 0,02x + 0,11.$$

Аналогично получим зависимость $x(y)$:

$$\overline{x}_y - \overline{x} = r\frac{\sigma_x}{\sigma_y}(y - \overline{y}), \quad \overline{x}_y - 54,6 = 0,46 \cdot \frac{6,46}{0,3}(y - 1,2),$$

$$\overline{x}_y = 9,97y + 42,64.$$

Построим эмпирические и теоретические зависимости (рис. 15).

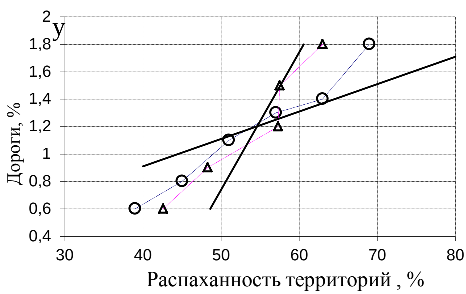


Рис. 15

Нетрудно заметить, что обе прямые проходят через точку ($\overline{x}$; $\overline{y}$), т. е. точку (54,6; 1,2).

ПРИЛОЖЕНИЯ

Приложение 1

Значения функции $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt.$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,39	0,1517	0,78	0,2823	1,17	0,3790
0,01	0,0040	0,40	0,1554	0,79	0,2852	1,18	0,3810
0,02	0,0080	0,41	0,1591	0,80	0,2881	1,19	0,3830
0,03	0,0120	0,42	0,1628	0,81	0,2910	1,20	0,3849
0,04	0,0160	0,43	0,1664	0,82	0,2939	1,21	0,3869
0,05	0,0199	0,44	0,1700	0,83	0,2967	1,22	0,3883
0,06	0,0239	0,45	0,1736	0,84	0,2995	1,23	0,3907
0,07	0,0279	0,46	0,1772	0,85	0,3023	1,24	0,3925
0,08	0,0319	0,47	0,1808	0,86	0,3051	1,25	0,3944
0,09	0,0359	0,48	0,1884	0,87	0,3078	1,26	0,3962
0,10	0,0398	0,49	0,1879	0,88	0,3106	1,27	0,3980
0,11	0,0438	0,50	0,1915	0,89	0,3133	1,28	0,3839
0,12	0,0478	0,51	0,1950	0,90	0,3159	1,29	0,4015
0,13	0,0517	0,52	0,1985	0,91	0,3186	1,30	0,4032
0,14	0,0557	0,53	0,2019	0,92	0,3212	1,31	0,4049
0,15	0,0596	0,54	0,2954	0,93	0,3238	1,32	0,4066
0,16	0,0636	0,55	0,2088	0,94	0,3264	1,33	0,4082
0,17	0,0675	0,56	0,2123	0,95	0,3289	1,34	0,4099
0,18	0,0714	0,57	0,2157	0,96	0,3315	1,35	0,4115
0,19	0,0753	0,58	0,2190	0,97	0,3340	1,36	0,4131
0,20	0,0793	0,59	0,2224	0,98	0,3365	1,37	0,4147
0,21	0,0832	0,60	0,2257	0,99	0,3389	1,38	0,4162
0,22	0,0871	0,61	0,2291	1,00	0,3413	1,39	0,4177
0,23	0,0910	0,62	0,2324	1,01	0,3438	1,40	0,4192
0,24	0,0948	0,63	0,2357	1,02	0,3461	1,41	0,4207
0,25	0,0987	0,64	0,2389	1,03	0,3485	1,42	0,4222
0,26	0,1026	0,65	0,2422	1,04	0,3508	1,43	0,4236
0,27	0,1064	0,66	0,2454	1,05	0,3531	1,44	0,4251
0,28	0,1103	0,67	0,2486	1,06	0,3554	1,45	0,4265
0,29	0,1141	0,68	0,2517	1,07	0,3577	1,46	0,4279
0,30	0,1179	0,69	0,2549	1,08	0,3599	1,47	0,4292
0,31	0,1217	0,70	0,2580	1,09	0,3621	1,48	0,4306
0,32	0,1255	0,71	0,2611	1,10	0,3643	1,49	0,4313
0,33	0,1293	0,72	0,2642	1,11	0,3665	1,50	0,4332
0,34	0,1331	0,73	0,2673	1,12	0,3686	1,51	0,4335
0,35	0,1368	0,74	0,2703	1,13	0,3708	1,52	0,4357
0,36	0,1406	0,75	0,2734	1,14	0,3729	1,53	0,4370
0,37	0,1443	0,76	0,2764	1,15	0,3749	1,54	0,4382
0,38	0,1480	0,77	0,2794	1,16	0,3770	1,55	0,4394
1,56	0,4406	1,82	0,4656	2,16	0,4846	2,68	0,4963

x	***Φ***(***x***)	*x*	***Φ***(***x***)	*x*	***Φ***(***x***)	*x*	***Φ***(***x***)
1,57	0,4418	1,83	0,4664	2,18	0,4854	2,70	0,4965
1,58	0,4429	1,84	0,4671	2,20	0,4861	2,72	0,4967
1,59	0,4441	1,85	0,4678	2,22	0,4868	2,74	0,4969
1,60	0,4452	1,86	0,4686	2,24	0,4875	2,76	0,4971
1,61	0,4463	1,87	0,4693	2,26	0,4881	2,78	0,4973
1,62	0,4474	1,88	0,4699	2,28	0,4887	2,80	0,4974
1,63	0,4484	1,89	0,4706	2,30	0,4893	2,82	0,4976
1,64	0,4495	1,90	0,4713	2,32	0,4898	2,84	0,4977
1,65	0,4505	1,91	0,4719	2,34	0,4904	2,86	0,4979
1,66	0,4515	1,92	0,4726	2,36	0,4909	2,88	0,4980
1,67	0,4525	1,93	0,4732	2,38	0,4913	2,90	0,4981
1,68	0,4535	1,94	0,4738	2,40	0,4918	2,92	0,4982
1,69	0,4545	1,95	0,4744	2,42	0,4922	2,94	0,4984
1,70	0,4554	1,96	0,4750	2,44	0,4927	2,96	0,4985
1,71	0,4564	1,97	0,4756	2,46	0,4931	2,98	0,4986
1,72	0,4573	1,98	0,4761	2,48	0,4934	3,00	0,49865
1,73	0,4582	1,99	0,4767	2,50	0,4938	3,20	0,49931
1,74	0,4591	2,00	0,4772	2,52	0,4941	3,40	0,49966
1,75	0,4599	2,02	0,4783	2,54	0,4945	3,60	0,499841
1,76	0,4608	2,04	0,4793	2,56	0,4948	3,80	0,499928
1,77	0,4616	2,06	0,4803	2,58	0,4951	4,00	0,499968
1,78	0,4625	2,08	0,4812	2,60	0,4953	4,50	0,499997
1,79	0,4633	2,10	0,4821	2,62	0,4956	5,00	0,499999
1,80	0,4641	2,12	0,4830	2,64	0,4959		
1,81	0,4649	2,14	0,4838	2,66	0,4961		

Приложение 2

Производственные показатели сельскохозяйственных предприятий

№ п/п	**Площадь пашни, га**	**Балл пашни**	**Стоимость валовой продукции, тыс. руб/га**	**Стоимость основных производственных фондов, тыс. руб/га**	**Стоимость оборотных фондов тыс. руб/га**	**Среднегодовое число работников, чел./га**
1	2	3	4	5	6	7
1	3243	39	1,291	21,631	1,421	0,064
2	2203	44	0,982	31,031	2,092	0,094
3	1505	42	0,619	46,595	3,062	0,138
4	2400	45	1,743	29,219	1,920	0,086
5	2049	47	2,042	34,224	2,249	0,101
6	3048	46	1,373	23,007	1,512	0,068
7	3124	45	1,339	22,447	1,475	0,066
8	2910	41	1,438	24,098	1,583	0,071

1	2	3	4	5	6	7
9	2540	42	0,661	27,608	1,814	0,081
10	2246	44	1,647	31,222	2,052	0,092
11	1129	45	3,706	62,113	4,082	0,184
12	1172	45	3,570	59,834	3,932	0,177
13	1646	42	2,542	42,603	2,800	0,126
14	2356	43	1,776	29,764	1,956	0,088
15	2914	41	1,436	24,065	1,581	0,071
16	3621	38	1,155	19,366	1,272	0,057
17	2421	42	1,728	28,965	1,903	0,085
18	1155	50	3,623	60,715	3,990	0,180
19	1566	42	2,672	44,780	2,943	0,132
20	1938	42	2,159	36,184	2,378	0,107
21	2305	46	1,815	30,423	1,999	0,090
22	2057	51	2,034	34,091	2,240	0,101
23	1393	44	3,004	50,341	3,308	0,149
24	1333	46	3,139	52,607	3,457	0,156
25	1422	52	2,944	49,315	3,241	0,146
26	1240	44	3,375	56,553	3,716	0,167
27	2232	50	1,875	31,418	2,064	0,093
28	2110	48	1,983	33,235	2,184	0,098
29	2722	46	1,537	25,762	1,693	0,076
30	2379	41	1,759	29,477	1,937	0,087
31	3393	51	1,233	20,667	1,358	0,061
32	1730	49	2,419	40,535	2,664	0,120
33	2051	47	2,040	34,191	2,513	0,101
34	1834	44	2,281	38,236	2,259	0,113
35	3129	48	1,337	22,411	1,472	0,066
36	2597	52	1,611	27,002	1,774	0,080
37	1785	46	2,344	39,286	2,582	0,116
38	3216	46	1,301	21,805	1,433	0,064
39	2252	46	1,858	31,139	2,046	0,092
40	2415	49	1,732	29,037	1,908	0,086
41	1014	47	4,127	69,157	4,545	0,205
42	1625	48	2,575	43,154	2,836	0,128
43	2691	43	1,555	26,059	1,712	0,077
44	2041	51	2,050	34,358	2,258	0,101
45	3283	47	1,274	21,360	1,403	0,063
46	1835	46	2,280	38,215	2,511	0,113
47	1740	40	2,405	40,302	2,648	0,119
48	1018	31	4,111	68,886	4,527	0,204
49	2073	45	2,018	33,828	2,223	0,100
50	2533	47	1,652	27,684	1,819	0,082
51	2699	54	1,550	25,982	1,707	0,077
52	1134	49	3,690	61,839	4,064	0,183
53	1638	46	2,554	42,811	2,813	0,126
54	1964	46	2,130	35,705	2,346	0,105

1	2	3	4	5	6	7
55	2020	42	2,071	34,715	2,281	0,102
56	4367	53	0,958	16,058	1,055	0,047
57	3425	52	1,221	20,474	1,345	0,060
58	1234	46	3,391	56,828	3,735	0,168
59	1690	45	2,476	41,494	2,727	0,123
60	4341	46	0,964	16,154	1,061	0,047
61	1661	44	2,519	42,219	2,774	0,125
62	2215	46	1,889	31,659	2,080	0,093
63	2784	48	1,503	25,188	1,655	0,074
64	1788	47	2,340	39,220	2,577	0,116
65	1641	46	2,550	42,733	2,808	0,126
66	1684	40	2,485	41,642	2,736	0,123
67	2051	42	2,040	34,191	2,247	0,101
68	1190	46	3,516	58,929	3,873	0,174
69	1717	44	2,437	40,842	2,684	0,121
70	3648	46	1,147	19,223	1,263	0,057
71	3819	44	1,097	18,362	1,206	0,054
72	2745	48	1,524	25,546	1,679	0,075
73	2222	39	12,883	31,559	2,074	0,093
74	904	47	1,441	24,148	1,587	0,071
75	2652	39	1,578	26,442	1,737	0,078
76	3758	46	1,113	18,660	1,226	0,055
77	2127	45	1,967	32,969	2,166	0,097
78	2805	49	1,491	25,001	1,643	0,074
79	2695	44	1,552	26,020	1,710	0,077
80	2330	50	1,796	30,096	1,978	0,089
81	3142	39	1,331	22,318	1,466	0,066
82	2103	43	1,990	33,345	2,191	0,098
83	1515	49	2,762	46,287	3,042	0,137
84	3453	55	1,211	20,308	1,334	0,060
85	2409	44	1,737	29,110	1,913	0,086
86	2149	53	1,947	32,631	2,144	0,096
87	3124	48	1,339	22,447	1,475	0,066
88	3048	50	1,373	23,007	1,512	0,068
89	2850	44	1,468	24,605	1,617	0,072
90	2504	45	1,671	28,005	1,840	0,083
91	2108	40	1,985	33,266	2,186	0,098
92	1131	45	3,700	62,003	4,075	0,183
93	1270	54	3,295	55,217	3,629	0,163
94	1184	48	3,534	59,228	3,892	0,175
95	1687	47	2,480	41,568	2,732	0,123
96	2358	52	1,774	29,739	1,954	0,088
97	2815	46	1,486	24,911	1,637	0,073
98	3609	50	1,159	19,430	1,277	0,057
99	2621	55	1,596	26,755	1,758	0,079
100	1189	41	3,519	58,978	3,876	0,174

1	2	3	4	5	6	7
101	1467	32	2,852	47,802	3,141	0,141
102	1839	37	2,275	38,132	2,506	0,113
103	2530	51	1,654	27,717	1,821	0,082
104	2075	43	2,016	33,795	2,221	0,100
105	1339	45	3,125	52,371	3,442	0,155
106	1242	49	3,369	56,462	3,710	0,167
107	1420	47	2,947	49,384	3,245	0,146
108	2322	54	1,802	30,200	1,984	0,089
109	2010	48	2,082	34,888	2,293	0,103
110	2272	50	1,841	30,865	2,028	0,091
111	2739	34	1,527	25,602	1,682	0,075
112	1370	33	3,054	51,186	3,364	0,151
113	2150	42	1,946	32,616	2,143	0,096
114	3933	55	1,064	17,830	1,171	0,052
115	1483	51	2,821	47,286	3,107	0,140
116	3219	38	1,300	21,785	1,431	0,064
117	1875	46	2,232	37,400	2,458	0,110
118	3126	49	1,338	22,433	1,474	0,066
119	2522	49	1,659	27,805	1,827	0,082
120	2145	42	1,951	32,692	2,148	0,096
121	1104	50	3,790	63,519	4,174	0,188
122	1265	49	3,308	55,435	3,643	0,164
123	2401	43	1,743	29,206	1,919	0,086
124	3183	51	1,314	22,031	1,448	0,065
125	1583	48	2,643	44,299	2,911	0,131
126	3142	40	1,331	22,318	1,466	0,066
127	2302	43	1,817	30,463	2,002	0,090
128	1505	43	2,780	46,595	3,062	0,138
129	2752	44	1,520	25,481	1,674	0,075
130	2094	48	1,998	33,489	2,201	0,099
131	3084	47	1,357	22,738	1,494	0,067
132	3142	46	1,331	22,318	1,466	0,066
133	2019	46	2,072	34,733	2,282	0,103
134	2405	43	1,740	29,158	1,916	0,086
135	2264	46	1,848	30,974	2,035	0,091
136	1291	47	3,241	54,319	3,570	0,161
137	1127	43	3,713	62,223	4,089	0,184
138	1464	44	2,858	47,900	3,148	0,142
139	2536	45	1,650	27,652	1,817	0,082
140	2419	40	1,730	28,989	1,905	0,085
141	3261	42	1,283	21,504	1,413	0,063
142	2241	41	1,867	31,292	2,056	0,092
143	1515	52	2,762	46,287	3,042	0,137
144	1615	40	2,591	43,421	2,853	0,128
145	1839	44	2,275	38,132	2,506	0,113
146	2509	44	1,667	27,949	1,836	0,082

1	2	3	4	5	6	7
147	2075	53	2,016	33,795	2,221	0,100
148	1402	42	2,985	50,018	3,287	0,148
149	1731	48	2,417	40,511	2,662	0,120
150	1248	44	3,353	56,190	3,693	0,166
151	1420	47	2,947	49,384	3,245	0,146
152	2184	48	1,916	32,108	2,110	0,095
153	2019	49	2,072	34,733	2,282	0,103
154	2272	46	1,841	30,865	2,028	0,091
155	2793	45	1,498	25,107	1,650	0,074
156	3180	51	1,316	22,052	1,449	0,065
157	1375	47	3,043	51,001	3,352	0,151
158	2105	46	1,988	33,314	2,189	0,098
159	1643	40	2,547	42,681	2,805	0,126
160	3219	32	1,300	21,785	1,431	0,064
161	2608	47	1,604	26,888	1,767	0,079
162	1874	45	2,233	37,420	2,459	0,110
163	3126	54	1,338	22,433	1,474	0,066
164	2514	49	1,664	27,894	1,833	0,082
165	1094	47	3,825	64,102	4,212	0,190
166	1568	45	2,669	44,723	2,939	0,132
167	2916	53	1,435	24,048	1,580	0,071
168	2140	52	1,955	32,769	2,153	0,097
169	3382	46	1,237	20,735	1,362	0,061
170	1798	39	2,327	39,002	2,563	0,115
171	2530	51	1,654	27,717	1,821	0,082
172	1839	44	2,275	38,132	2,506	0,113
173	1242	49	3,369	56,462	3,710	0,167
174	1420	47	2,947	49,384	3,245	0,146
175	2049	47	2,042	34,224	2,249	0,101
176	3048	46	1,373	23,007	1,512	0,068
177	3124	45	1,339	22,447	1,475	0,066
178	2533	47	1,652	27,684	1,819	0,082
179	2699	54	1,550	25,982	1,707	0,077
180	1134	49	3,690	61,839	4,064	0,183
181	1687	47	2,480	41,568	2,732	0,123
182	2358	52	1,774	29,739	1,954	0,088
183	2815	46	1,486	24,911	1,637	0,073
184	3609	50	1,159	19,430	1,277	0,057
185	2401	43	1,743	29,206	1,919	0,086
186	3183	51	1,314	22,031	1,448	0,065
187	1583	48	2,643	44,299	2,911	0,131
188	3142	40	1,331	22,318	1,466	0,066
189	2793	45	1,498	25,107	1,650	0,074
190	3180	51	1,316	22,052	1,449	0,065
191	2533	47	1,652	27,684	1,819	0,082
192	1402	42	2,985	50,018	3,287	0,148

Окончание прил. 2

1	2	3	4	5	6	7
193	1731	48	2,417	40,511	2,662	0,120
194	1248	44	3,353	56,190	3,693	0,166
195	3648	46	1,147	19,223	1,263	0,057
196	3819	44	1,097	18,362	1,206	0,054
197	2745	48	1,524	25,546	1,679	0,075
198	2203	44	0,982	31,031	2,092	0,094
199	1505	42	0,619	46,595	3,062	0,138
200	2400	45	1,743	29,219	1,920	0,086

Приложение 3

Критические точки распределения Пирсона χ^2

Число степеней свободы *k*	**Уровень значимости α**					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,89**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,7	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Приложение 4

Баллы бонитета почв сельскохозяйственных угодий

№ п/п	Варианты									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	36	38	30	33	40	34	31	45	42	42
2	44	46	43	38	43	42	44	45	36	55
3	37	32	30	24	32	33	32	38	43	42
4	27	40	32	31	35	38	36	40	49	45
5	36	29	33	37	37	39	28	44	47	44
6	40	33	38	42	46	48	38	49	44	52
7	24	28	40	31	31	31	37	41	45	40
8	28	33	50	42	36	37	42	43	40	26
9	26	34	31	30	30	34	37	43	40	43
10	31	38	39	40	33	47	49	42	42	47
11	24	30	31	30	31	36	39	43	44	49
12	35	33	48	33	34	47	42	42	56	53
13	37	34	25	34	40	37	35	40	42	42
14	49	38	27	38	31	46	37	44	38	35
15	36	37	39	27	36	38	28	42	44	44
16	42	40	34	35	32	45	26	45	42	44
17	41	34	27	36	36	44	26	39	42	41
18	49	40	32	42	31	46	35	41	45	39
19	46	34	27	36	32	32	35	34	44	48
20	51	37	31	41	39	39	40	40	47	58
21	33	34	27	37	46	36	38	30	44	40
22	46	39	49	40	35	44	43	35	49	40
23	35	36	33	38	41	35	34	36	42	35
24	47	45	41	45	35	48	46	41	26	37
25	30	31	29	37	40	34	34	34	38	37
26	35	39	32	42	38	42	41	37	47	45
27	39	36	26	29	37	37	36	32	51	48
28	44	39	30	34	41	46	44	37	47	38
29	31	43	33	32	36	33	41	30	49	42
30	37	48	41	35	42	32	42	34	45	40
31	30	40	36	28	34	47	31	39	45	47
32	37	43	43	46	40	36	36	41	44	37
33	29	30	31	36	29	27	29	40	47	49
34	24	45	40	42	41	37	40	45	53	34
35	37	39	32	32	43	52	26	37	51	50
36	38	47	40	42	49	30	31	47	44	36
37	37	22	34	32	30	36	32	41	43	41
38	40	38	42	36	32	30	38	41	45	43
39	34	57	26	32	26	35	37	32	41	45
40	43	37	38	38	35	44	40	34	41	37
41	29	47	32	32	34	38	32	35	44	37
42	40	35	42	38	39	41	37	45	46	47
43	32	42	34	33	34	37	31	37	45	46
44	42	35	39	39	40	48	41	42	48	50
45	31	41	31	38	34	32	33	46	42	41
46	41	36	38	44	33	36	39	48	43	47
47	26	40	33	40	32	35	31	38	43	50
48	34	43	32	48	34	44	39	44	41	53
49	32	50	31	29	31	41	27	39	48	49
50	39	30	34	45	34	53	32	42	42	52

№ п/п	Варианты									
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	40	40	47	58	40	40	36	41	44	37
2	38	30	44	40	38	30	29	40	47	49
3	43	35	49	40	43	35	40	45	53	34
4	34	36	42	35	34	36	26	37	51	50
5	46	41	26	37	46	41	31	47	44	36
6	34	34	38	37	34	34	32	41	43	41
7	41	37	47	45	41	37	38	41	45	43
8	36	32	51	48	36	32	37	32	41	45
9	44	37	47	38	44	37	40	34	41	37
10	41	30	49	42	41	30	32	35	44	37
11	42	34	45	40	42	34	37	45	46	47
12	31	39	45	47	31	39	31	37	45	46
13	36	41	44	37	36	41	36	41	44	37
14	29	40	47	49	29	40	29	40	47	49
15	40	45	53	34	40	45	40	45	53	34
16	36	41	44	37	36	41	26	37	51	50
17	29	40	47	49	29	40	31	47	44	36
18	40	32	30	24	32	33	32	38	27	36
19	29	40	32	31	35	38	36	40	31	41
20	41	29	33	37	37	39	28	44	27	37
21	43	33	38	42	46	48	38	49	49	40
22	49	28	40	31	31	31	37	41	33	38
23	30	33	50	42	36	37	42	43	41	45
24	32	34	31	30	30	34	37	43	29	37
25	26	38	39	40	33	47	49	42	32	42
26	35	30	31	30	31	36	39	43	26	29
27	34	33	48	33	34	47	42	42	30	34
28	39	34	25	34	40	37	35	40	33	32
29	34	38	27	38	31	46	37	44	41	35
30	40	34	31	45	42	42	38	43	36	28
31	43	42	44	45	36	55	24	32	43	46
32	32	33	32	38	43	42	31	35	31	36
33	35	38	36	40	49	45	37	37	40	42
34	37	39	28	44	47	44	42	46	32	32
35	46	48	38	49	44	52	31	31	40	42
36	31	31	37	41	45	40	42	36	34	32
37	36	37	42	43	40	26	30	30	42	36
38	30	34	37	43	40	43	40	33	26	32
39	33	47	49	42	42	47	30	31	31	47
40	31	36	39	43	44	49	33	34	32	41
41	40	34	31	45	42	42	34	40	38	41
42	43	42	44	45	36	55	38	31	31	47
43	32	33	32	38	43	42	27	36	32	41
44	35	38	36	40	49	45	35	32	38	41
45	47	45	41	45	35	48	46	41	53	34
46	30	31	29	37	40	34	34	34	44	37
47	35	39	32	42	38	42	41	37	47	49
48	39	36	26	29	37	37	36	32	30	24
49	44	39	30	34	41	46	44	37	32	31
50	47	45	41	45	35	48	46	41	33	37

Приложение 5

Корреляционные таблицы выборочных данных случайных величин *Y* – площадь под дорогами (%) и *X* – распаханность территорий (%)

1.

Случайные величины		X						
		5	10	15	20	25	30	m_y
Y	15	1	2					3
	25		4	2				6
	35			1	23	1		25
	45			1	5	3		9
	55				2	3	2	7
	m_x	1	6	4	30	7	2	$n = 50$

2.

Случайные величины		X						
		20	25	30	35	40	45	m_y
Y	25	2	1					3
	35		3	2				5
	45			3	22	1		26
	55			1	4	3		8
	65				2	4	2	8
	m_x	2	4	6	28	8	2	$n = 50$

3.

Случайные величины		X						
		10	15	20	25	30	35	m_y
Y	20	1	3					4
	30		2	2				4
	40			3	18	4		25
	50			1	5	4		10
	60				2	3	2	7
	m_x	1	5	6	25	11	2	$n = 50$

4.

Случайные величины		X						
		15	20	25	30	35	40	m_y
Y	30	1	3					4
	40		2	2				4
	50			5	20	1		26
	60			2	5	3		10
	70				2	3	1	6
	m_x	1	5	9	27	7	1	$n = 50$

5.

Случайные величины		X						
		25	30	35	40	45	50	m_y
Y	35	3	1					4
	45		3	1				4
	55			3	20	2		25
	65			1	4	3		8
	75				2	3	4	9
	m_x	3	4	5	26	8	4	$n = 50$

6.

Случайные величины		X						
		10	15	20	40	30	35	m_y
Y	20	1	2					3
	30		2	3				5
	40			2	15	5		22
	50			4	5	4		13
	60				2	3	2	7
	m_x	1	4	9	22	12	2	$n = 50$

7.

Случайные величины		X						
		1	6	11	16	21	26	m_y
Y	110	1	2					3
	120		2	2				4
	130			2	24	1		27
	140			1	5	3		9
	150				1	4	2	7
	m_x	1	4	5	30	8	2	$n = 50$

8.

Случайные величины		X						
		2	7	12	17	22	27	m_y
Y	25	1	2					3
	35		3	2				5
	45			3	22	2		27
	55			1	4	3		8
	65				2	3	2	7
	m_x	1	5	6	28	8	2	$n = 50$

9.

Случайные величины		X						
		3	8	13	18	23	28	m_y
Y	30	2	2					4
	40		3	2				5
	50			20	1	4		25
	60			2	5	3		10
	70				2	3	1	6
	m_x	2	5	24	8	10	1	$n = 50$

10.

Случайные величины		X						
		4	9	14	19	24	29	m_y
Y	32	2	2					4
	42		2	3				5
	52			6	21	2		29
	62			1	4	2		7
	72				1	1	3	5
	m_x	2	4	10	26	5	3	$n = 50$

11.

Случайные величины		X						
		6	11	16	21	26	31	m_y
Y	16	2	2					4
	26		3	3				6
	36			1	22	1		24
	46			1	6	6		13
	56				1	1	1	3
	m_x	2	5	5	29	8	1	$n = 50$

12.

Случайные величины		X						
		11	16	21	26	31	36	m_y
Y	21	2	2					4
	31		2	3				5
	41			3	19	3		25
	51			2	5	3		10
	61				3	1	2	6
	m_x	2	4	8	27	7	2	$n = 50$

13.

Случайные величины		X						
		21	26	31	36	41	46	m_y
Y	24	2	2					4
	34		3	2				5
	44			3	20	1		24
	54			2	4	3		9
	61				2	4	2	8
	m_x	2	5	7	26	8	2	$n = 50$

14.

Случайные величины		X						
		13	18	23	28	33	38	m_y
Y	32	2	2					4
	42		2	3				5
	52			6	21	2		29
	62			1	4	2		7
	72				1	1	3	5
	m_x	2	4	10	26	5	3	$n = 50$

15.

Случайные величины		X						
		20	25	30	35	40	45	m_y
Y	37	2	1					3
	47		2	2				4
	57			3	20	1		24
	67			2	2	4		8
	77				3	3	5	11
	m_x	2	3	7	25	8	5	$n = 50$

16.

Случайные величины		X						
		11	16	21	31	36	41	m_y
Y	21	2	1					3
	31		2	4				6
	41			2	14	4		20
	51			5	1	4		10
	61				5	2	4	11
	m_x	2	3	11	20	10	4	$n = 50$

17.

Случайные величины		X						
		11	16	21	26	31	36	m_y
Y	11	2	1					3
	12		2	3				4
	13			2	23	2		27
	14			2	4	1		9
	15				1	5	2	7
	m_x	2	3	7	28	8	2	$n = 50$

18.

Случайные величины		X						
		3	8	13	18	23	28	m_y
Y	20	2	2					3
	25		2	1				3
	30			2	21	3		26
	35			2	3	3		8
	40				3	3	4	10
	m_x	2	4	5	27	9	4	$n = 50$

19.

Случайные величины		X						
		4	9	14	19	24	27	m_y
Y	31	3	1					4
	41		4	2				6
	51			21	1	3		25
	61			2	4	2		8
	71				3	2	2	7
	m_x	3	5	25	8	7	2	$n = 50$

20.

Случайные величины		X						
		1	6	11	16	21	26	m_y
Y	31	1	2					3
	41		3	2				5
	51			1	20	1		22
	61			1	5	2		8
	71				3	6	3	12
	m_x	1	5	4	28	9	3	$n = 50$

СОДЕРЖАНИЕ

Учебное издание

Воронкова Татьяна Борисовна
Василькова Светлана Львовна
Демитриченко Елена Леонидовна и др.

ВЫСШАЯ МАТЕМАТИКА

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие

Редактор *О. Н. Минакова*
Технический редактор *Н. Л. Якубовская*
Корректор *А. С. Зайцева*

Подписано в печать 12.11.2019. Формат 60×84 $^{1}/_{16}$. Бумага офсетная.
Ризография. Гарнитура «Таймс». Усл. печ. л. 4,42. Уч.-изд. л. 3,51.
Тираж 60 экз. Заказ .

УО «Белорусская государственная сельскохозяйственная академия».
Свидетельство о ГРИИРПИ № 1/52 от 09.10.2013.
Ул. Мичурина, 13, 213407, г. Горки.

Отпечатано в УО «Белорусская государственная сельскохозяйственная академия».
Ул. Мичурина, 5, 213407, г. Горки.