МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский национальный технический университет

Кафедра «Инженерная математика»

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие
для студентов специальностей
1-38 01 01 «Механические и электромеханические приборы и аппараты», 1-38 01 02 «Оптико-электронные и лазерные приборы и системы», 1-52 02 01 «Технология и оборудование ювелирного производства», 1-38 02 02 «Биотехнические и медицинские аппараты и системы», 1-54 01 01 «Метрология, стандартизация и сертификация (машиностроение и приборостроение)»

В 2 частях

Часть 2

Рекомендовано учебно-методическим объединением высших учебных заведений Республики Беларусь по образованию в области приборостроения

Под редакцией *М. А. Князева*

Минск
БНТУ
2020

УДК 519.21(075.8)
ББК 22.17я7
Т33

Составители:
Н. К. Прихач, И. В. Прусова, Л. В. Бокуть, Н. А. Кондратьева

Рецензенты:
первый проректор Белорусского государственного университета,
д-р пед. наук, *Д. Г. Медведев*;
канд. техн. наук, доцент, профессор БИП *В. А. Бахмат*

Т33 **Теория** вероятностей и математическая статистика : учебно-методическое пособие для студентов специальностей 1-38 01 01 «Механические и электромеханические приборы и аппараты», 1-38 01 02 «Оптико-электронные и лазерные приборы и системы», 1-52 02 01 «Технология и оборудование ювелирного производства», 1-38 02 02 «Биотехнические и медицинские аппараты и системы», 1-54 01 01 «Метрология, стандартизация и сертификация (машиностроение и приборостроение)» : в 2 ч. / сост.: Н. К. Прихач [и др.]; под ред. М. А. Князева. – Минск : БНТУ, 2020. – Ч. 2. – 72 с.
ISBN 978-985-583-557-9 (Ч. 2).

Издание содержит материалы для организации системы непрерывного освоения знаний по дисциплине «Математика» раздела «Математическая статистика» для студентов технических специальностей. Задания для самостоятельного обучения разработаны с учетом рекомендации кафедры «Инженерная математика» приборостроительного факультета Белорусского национального технического университета и согласованы с требованиями к уровню подготовки, определенному базовым стандартом по математике.

Часть 1 вышла в 2020 г.

УДК 519.21(075.8)
ББК 22.17я7

ISBN 978-985-583-557-9 (Ч. 2)
ISBN 978-985-583-373-5

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Вторая часть учебно-методического пособия «Теория вероятностей и математическая статистика» посвящена математической статистике.

Математическая статистика – наука о математических методах систематизации и использования статистических данных для научных и практических выводов. Во многих своих разделах математическая статистика опирается на теорию вероятностей, позволяющую оценить надежность и точность выводов, полученных на основании ограниченного статистического материала.

Пособие включает в себя следующие разделы:

1. Вариационные ряды и их графическое изображение.
2. Числовые характеристики законов распределения эмпирических величин.
3. Статистическая проверка гипотез.
4. Элементы теории регрессионного и корреляционного анализа.

Пособие имеет следующую структуру: в начале каждого раздела даются необходимые теоретические сведения, далее приводится краткая сводка рабочих формул, применение которых иллюстрируется решением типовых задач. Затем приведены задачи для самостоятельного решения, к которым имеются ответы. Преподаватель может использовать задачи для практических занятий и индивидуальных домашних заданий. Пособие содержит статистические таблицы, необходимые для решения задач, а также проверочный тест.

Вторая часть учебно-методического пособия «Теория вероятностей и математическая статистика» предназначена, в первую очередь, для студентов второго курса инженерных специальностей приборостроительного факультета БНТУ. Однако приведенный материал может быть также полезен при проведении практических занятий со студентами механико-технологического и спортивно-технического факультетов БНТУ.

§ 1. ВАРИАЦИОННЫЕ РЯДЫ И ИХ ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ

1.1. Краткие теоретические сведения

Результаты наблюдений массовых явлений, случайных величин составляют статистические данные или *статистический материал*.

Генеральной совокупностью называется совокупность объектов произвольной природы, обладающих признаками, доступными для наблюдения и количественного измерения.

Объекты, входящие в генеральную совокупность, называются ее *элементами*, а их общее число N – ее *объемом*.

Но получение экспериментальных данных – достаточно сложный, трудоемкий процесс, а в некоторых случаях и просто невозможный. Поэтому из всей генеральной совокупности приходится выбирать только определенную часть объектов, которую называют *выборочной совокупностью* или *выборкой* объема n.

Предположим, что над случайной величиной (признаком) X производится ряд независимых опытов (наблюдений). В каждом из этих опытов признак X принимает определенное значение $x_1, x_2, ..., x_n$. Совокупность этих значений рассматривается как простая выборка.

Наблюдаемое значение x_i называют *вариантой*, а их последовательность, записанную в возрастающем порядке – *вариационным рядом*.

Дискретным вариационным рядом или *статистическим распределением выборки* называют таблицу, которая в первой строке содержит значения $x_1, x_2, ... x_k \left(x_1 < x_2 < ... < x_k\right)$, а во второй – числа их повторений $n_1, n_2, ..., n_k$; $n_1 + n_2 + ... + n_k = n$ (см. табл. 1.1).

Таблица 1.1

x_i	x_1	x_2		x_k
n_i	n_1	n_2		n_k

Числа n_i называют *частотой*, а отношение $p_i^* = \dfrac{n_i}{n}$ – *относительной частотой (частостью)* варианты $x_i \left(i = \overline{1, k}\right)$. Дискретный вариационный ряд представляет собой выборку значений дискретной случайной величины.

При большом числе наблюдений статистический ряд перестает быть удобной формой записи статистического материала, он становится громоздким и мало наглядным. Для придания ему большей компактности и наглядности строится так называемый *интервальный* статистический ряд. В этом случае весь диапазон наблюдаемых значений X разделяется на интервалы и подсчитывается количество значений m_i, p_i^*, приходящееся на каждый интервал (табл. 1.2).

Таблица 1.2

Интервал $[x_{i-1}, x_i)$	$[x_0, x_1)$	$[x_1, x_2)$	…	$[x_{l-1}, x_l]$	Сумма
Середины x_i^*	x_1^*	x_2^*	…	x_l^*	
Частоты m_i	m_1	m_2	…	m_l	n
Относительные частоты p_i^*	p_1^*	p_2^*	…	p_l^*	1

Длину интервала – h – проще выбирать одинаковой. Практика показывает, что число интервалов рационально выбирать порядка 7–20. Для нахождения длины интервала можно воспользоваться формулой

$$h = \frac{x_{\max} - x_{\min}}{l}, \qquad (1.1)$$

где $l = 1 + 3{,}322 \lg n$ – количество интервалов, рассчитываемое по *формуле Стерджеса*.

Если в результате вычисления по формуле (1.1) длина интервала получится дробным числом, то выбирают либо близкое целое число, либо близкую простую дробь.

Графически статистический ряд можно представить в виде полигона частот или относительных частот. *Полигоном частот (относительных частот)* называют ломаную линию, отрезки которой соединяют точки $(x_i; n_i)$ (соответственно $(x_i; p_i^*)$). Полигоны обычно служат для изображения выборки в случае дискретных случайных величин (рис. 1.1):

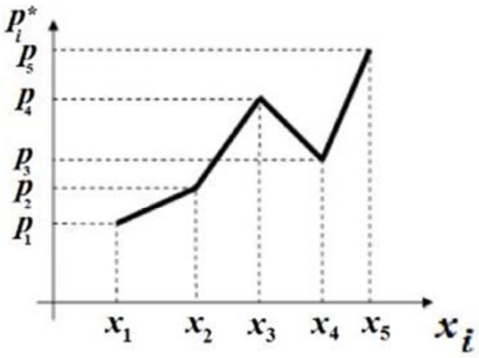

Рис. 1.1. Полигон относительных частот

Интервальный статистический ряд часто оформляется графически в виде гистограммы. *Гистограммой* называется ступенчатая фигура (рис. 1.2), состоящая из прямоугольников, основаниями которых служат отрезки, равные длине интервала, а высотами являются относительные частоты, разделенные на длину интервала.

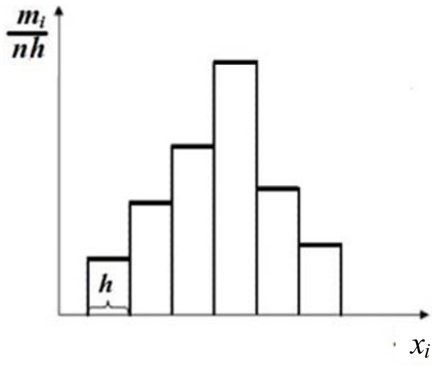

Рис. 1.2. Гистограмма относительных частот

Гистограмма обычно служит для изображения выборки в случае непрерывных случайных величин. Площадь гистограммы равна единице. Если на гистограмме соединить прямолинейными отрезками середины верхних оснований прямоугольников, получим полигон распределения.

В теории вероятностей для характеристики распределения случайной величины X служит функция распределения

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

которая определяет для каждого значения x вероятность того, что случайная величина X примет значение, меньшее x, т. е. равное вероятности события $A = \{X < x\}$, где x – любое действительное число.

Одной из основных характеристик выборки является выборочная (эмпирическая) функция распределения

$$F^*(x) = \frac{n_x}{n},$$

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

Другими словами, $F^*(x)$ – есть относительная частота появления события $A = \{X < x\}$ в n независимых испытаниях. Главное различие между $F(x)$ и $F^*(x)$ состоит в том, что $F(x)$ определяет вероятность события A, а выборочная функция распределения $F^*(x)$ – относительную частоту этого события.

Свойства функции $F^*(x)$:

1. $0 \le F^*(x) \le 1$.

2. $F^*(x)$ – неубывающая функция.

3. $F^*(-\infty) = 0$; $F^*(+\infty) = 1$.

Функция $F^*(x)$ является «ступенчатой», имеются разрывы в точках, которым соответствуют наблюдаемые значения вариантов (рис. 1.3). Величина скачка равна относительной частоте варианта.

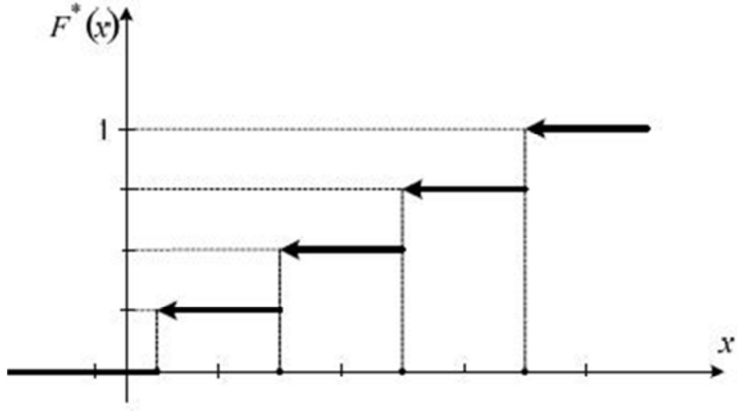

Рис. 1.3. График эмпирической функции распределения

Аналитически $F^*(x)$ задается следующим соотношением:

$$F^*(x) = \begin{cases} 0, & x < x_1 \\ \sum\limits_{j=1}^{i-1} p_j^*, & x_{i-1} < x \le x_i \quad i = \overline{2, n}, \\ 1, & x > x_n \end{cases} \tag{1.2}$$

где p_i^* – относительные частоты;

x_i – элементы вариационного ряда (варианты).

З*амечание*. В случае интервального вариационного ряда под x_i понимается середина i-го частичного интервала. Эмпирическую функцию распределения непрерывной случайной величины так же называют «*накопленной частотой*».

1.2. Примеры решения задач

1. Дана статистическая совокупность чисел:

32 17 22 15 22 17 20 26 27 32 17 32 17 22 15 26 17 22 15 20 26 32 22 32 37 22 15 20 27 26 32 37 22 20 27 32 37 26 32 17 32 22 15 20 26 22 32 22 32 37

Написать дискретный вариационный ряд, построить интервальную обработку. Построить полигон частот, гистограмму относительных частот, эмпирическую функцию распределения и ее график.

Решение. Составим вариационный ряд – запишем числа в порядке возрастания:

15	15	15	15	15	17	17	17	17	17
17	20	20	20	20	20	22	22	22	22
22	22	22	22	22	22	26	26	26	26
26	26	27	27	27	32	32	32	32	32
32	32	32	32	32	32	37	37	37	37

1) Дискретный вариационный ряд имеет вид (табл. 1.3):

Таблица 1.3

x_i	15	17	20	22	26	27	32	37	Сумма
n_i	5	6	5	10	6	3	11	4	50
p_i^*	0,1	0,12	0,1	0,2	0,12	0,06	0,22	0,8	1

2) Полигон частот данного дискретного ряда изображен на рис. 1.4 (построен с помощью пакета MS Excel).

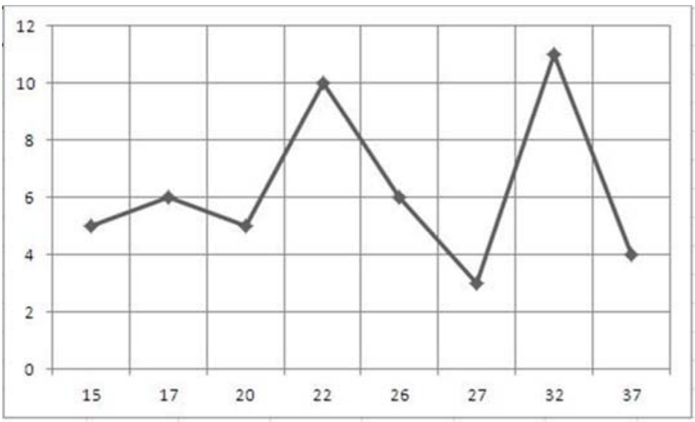

Рис. 1.4. Полигон частот

3) Для построения эмпирической функции распределения нужно найти накопленные относительные частоты w_i, которые вычисляются по формулам

$$w_1 = p_1^*;\ w_2 = p_1^* + p_2^*;\ w_3 = p_1^* + p_2^* + p_3^*;\ \dots;\ w_k = \sum_{i=1}^{k} p_i^*.$$

Для нахождения накопленных относительных частот составим следующую вспомогательную табл. 1.4.

Таблица 1.4

x_i	15	17	20	22	26	27	32	37	Сумма
p_i^*	0,1	0,12	0,1	0,2	0,12	0,06	0,22	0,8	1
w_i	0,1	0,22	0,32	0,52	0,64	0,7	0,92	1	

Тогда, согласно формуле (1.1).

$$F^*(x) = \begin{cases} 0, & x \le 15 \\ 0{,}1, & 15 < x \le 17 \\ 0{,}22, & 17 < x \le 20 \\ 0{,}32, & 20 < x \le 22 \\ 0{,}52, & 22 < x \le 26 \\ 0{,}64, & 26 < x \le 27 \\ 0{,}7, & 27 < x \le 32 \\ 0{,}92, & 32 < x \le 37 \\ 1, & x > 37 \end{cases}.$$

4) Для удобства интервальной обработки расширим интервал, на котором расположены варианты x. Пусть $x_{\min} = 14$; $x_{\max} = 38$. Тогда, согласно формуле (1.1)

$$h = \frac{38 - 14}{1 + 3{,}322 \lg 50} \approx \frac{24}{6} = 4.$$

Границы интервалов будут равны

$$x_0 = 14;\ x_1 = 14 + 4 = 18;\ x_2 = 18 + 4 = 22;\ x_3 = 26;$$
$$x_4 = 30;\ x_5 = 34;\ x_6 = 38.$$

Частота m_i интервала $[x_i;\ x_{i+1})$ $\left(i = \overline{0,5}\right)$ подсчитывается с помощью ряда как число наблюдений, попавших в интервал. Так, в первый $(i = 0)$ интервал $[14; 18)$ попало $5 + 6 = 11$ значений; во второй $[18; 22)$ – 5 значений. Аналогично $m_2 = 10$; $m_3 = 9$; $m_4 = 11$; $m_5 = 4$. Сведем полученные данные в табл. 1.5.

Таблица 1.5

$[x_{i-1}, x_i)$	$[14;18)$	$[18;22)$	$[22;26)$	$[26;30)$	$[30;34)$	$[34;38)$
x_i^*	16	20	24	28	32	36
m_i	11	5	10	9	11	4
p_i^*	0,22	0,1	0,2	0,18	0,22	0,08
$\dfrac{m_i}{h \cdot n}$	0,055	0,025	0,05	0,045	0,055	0,02

Гистограмма относительных частот представлена на рис. 1.5:

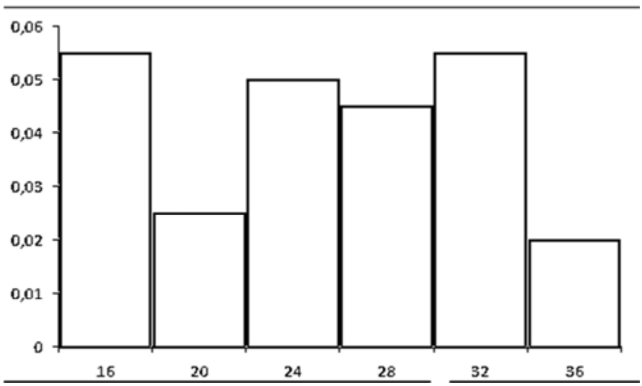

Рис. 1.5. Гистограмма относительных частот

1.3. Задачи для самостоятельного решения

Дана статистическая совокупность чисел. Написать вариационный ряд, построить интервальную обработку. Построить полигон распределения, гистограмму относительных частот, эмпирическую функцию распределения и ее график.

1. 1 6 12 7 1 12 1 2 8 4 3 13 1 5 5 10 2 2 2 4 3 11 2 11 3 4 5 7 6 6 9 10 1 3 1 2.

2. 12 16 15 14 18 10 20 16 14 14 15 17 23 16 11 12 21 17 16 15 10 13 14 19 18 16 17 20 14 18.

3. 181 169 178 171 178 179 172 181 179 168 174 167 169 171 179 181 181 183 172 176 165 178 176 176 190 183 169 192 185 173 179 187 181 168 179 184 177 184 185 186.

4. 14,3 17,1 15,4 16,6 13,9 11,6 16,2 19,9 21,9 16,5 14,7 12,1 15,7 14,7 12,1 15,7 14,7 9,4 18,8 15,1 17,9 14,1 16,2 17,7 14,2 17,3 14,1 17,6 15,6 14,1 13,5 8,8 16,7 15,1 13,7 20,7 12,3 15,2 13 16,4 16,1 8,8 16,3 12,7 17,3 15,2 11,7 14,5 15,8 16,7 18,2 9,9 10,9 18,3 17,5 14,6 12,5 14,6 15,6 12,4 15,8 13,4 14,2.

Проверочный тест 1

Условие задачи	Варианты ответов
1. Совокупность объектов, из которых производится выборка, называется….	1) Генеральной 2) Средней 3) Вероятной 4) Выборочной 5) Массовой
2. Часть отобранных объектов из генеральной совокупности называется….	1) Генеральной выборкой 2) Выборочной совокупностью 3) Репрезентативной совокупностью 4) Вариантами
3. Различные значения признака (случайной величины X) называются ….	1) Относительной частотой 2) Выборкой 3) Частотами 4) Вариантами
4. Числа, показывающие, сколько раз встречаются варианты из данного интервала, называются…	1) Относительной частотой 2) Выборкой 3) Вариантами 4) Частотами
5. Отношение частоты данного варианта к общей сумме частот называется…	1) Относительной частотой 2) Выборкой 3) Вариантами 4) Частотами
6. 3, 3, 1, 2, 5, 4, 2, 2, 4, 0, 2, 3, 2, 0, 2 – выборка. Относительная частота варианты 2 составляет…	1) 5 2) 2/3 3) 2/5 4) 3
7. Полигон служит для изображения…	1) Дискретного ряда 2) Гистограммы 3) Интервального ряда 4) Выборочной функции
8. Эмпирической функцией распределения называется относительная частота того, что признак (случайная величина X) примет значение…	1) Больше заданного x 2) Меньше заданного x 3) Равное заданному x

Условие задачи	Варианты ответов
9. Гистограмма служит для изображения …	1) Дискретного ряда 2) Кумулянты 3) Интервального ряда 4) Выборочной функции
10. Выберите номер неправильного ответа. Следующие выражения являются свойствами эмпирической функции распределения	1) $0 \le F^*(x) \le 1$ 2) $F^*(x)$ – неубывающая функция 3) $F^*(x)$ – невозрастающая функция 4) $F^*(-\infty)=0;\ F^*(+\infty)=1$

Ответы к проверочному тесту 1

№	1	2	3	4	5	6	7	8	9	10
	1	2	4	4	1	3	1	2	3	3

§ 2. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ЭМПИРИЧЕСКИХ ВЕЛИЧИН

2.1. Краткие теоретические сведения

Одними из задач математической статистики являются установление закона распределения случайной величины X (генеральной совокупности) и оценка параметров этого закона.

Вид закона выбирается из каких-либо теоретических или практических соображений, а параметры следует вычислять исходя из параметров этого закона.

Важнейшим этапом обработки статистических данных является вычисление оценок числовых характеристик исследуемой случайной величины.

Полученные оценки позволяют в числовой форме описать характерные черты статистического распределения и являются базой для построения математической модели изучаемого случайного явления.

Любая величина $\tilde{\theta}$, определяемая как функция выборочных значений $\tilde{\theta} = \varphi(x_1, x_2, ..., x_k)$, называется *выборочной статистикой* или просто *статистикой*. Статистика $\tilde{\theta}$, используемая в качестве приближенного значения неизвестного параметра θ, называется *статистической оценкой* параметра θ.

Существует два вида оценок параметров: точечные и интервальные.

Оценка называется *точечной,* если она определяется одним числом, и *интервальной*, если она определяется двумя числами – концами интервала.

К точечным статистическим оценкам предъявляется ряд требований.

Если $\tilde{\theta}$ – статистическая оценка параметра θ, то она должна удовлетворять следующим условиям:

1) быть *несмещенной*, это означает, что $M\left(\tilde{\theta}\right) = \theta$;

2) быть *состоятельной*, т. е. при $n \to +\infty$ предел по вероятности последовательности таких оценок должен быть равен искомому параметру, т. е. вероятность того, что $\left|\theta - \tilde{\theta}\right| > \delta > 0$, стремится к нулю при $n \to +\infty$;

3) быть эффективной, т. е. дисперсия $D(\tilde{\theta})$ – наименьшая, или быть *асимптотически эффективной*, что означает, что $\lim\limits_{n\to\infty} D(\tilde{\theta}) = 0.$

Число $\delta > 0$ называется *точностью оценки*, если имеет место равенство $|\theta - \tilde{\theta}| < \delta.$ Если это неравенство имеет место с некоторой вероятностью γ, то число γ называется *надежностью оценки* или *уровнем надежности*.

Среди выборочных характеристик случайной величины выделяют показатели, относящиеся к центру распределения, показатели рассеяния вариант и меры формы распределения.

К показателям, характеризующим центр распределения, относят выборочную среднюю, а также выборочную моду и медиану.

Выборочной средней $\overline{x}$ называют среднее арифметическое значение случайной величины X по выборочной совокупности объема n:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i, \tag{2.1}$$

где x_i – варианты дискретного ряда (или середины интервалов непрерывного вариационного ряда);

n_i – соответствующие им частоты;

$n = \sum\limits_{i=1}^{k} n_i$ – объем выборки.

Выборочная средняя служит *несмещенной оценкой математического ожидания* признака X или генеральной совокупности.

Выборочной модой Mo называют варианту, которая имеет наибольшую частоту.

Медианой Me называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если $n = 2k + 1$, то $Me = x_{k+1}$, а если $n = 2k$, то $Me = \frac{x_k + x_{k+1}}{2}$.

Средние величины не отражают изменчивости (вариации) значений признака. Чтобы охарактеризовать рассеяние наблюдаемых значений количественного признака выборки вокруг своего среднего значения, вводят свободную характеристику – *выборочную дисперсию*.

Выборочной дисперсией $D_{\text{В}}$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{x}$:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \bar{x}^2. \tag{2.2}$$

Выборочным средним квадратическим отклонением (стандартом) называют квадратный корень из выборочной дисперсии:

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}}. \tag{2.3}$$

Выборочная дисперсия является смещенной оценкой генеральной дисперсии. В качестве *несмещенной оценки генеральной дисперсии* служит «исправленная» выборочная дисперсия

$$s^2 = \frac{n}{n-1} D_{\text{В}} = \frac{1}{n-1}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 n_i. \tag{2.4}$$

При достаточно больших n выборочная и исправленная дисперсии мало отличаются, поэтому на практике исправленной дисперсией пользуются, если $n < 100$.

Выборочная средняя и дисперсия вариационного ряда являются частными случаями более общего понятия – *выборочных (эмпирических) моментов*.

Начальный момент $\tilde{\nu}_k$ k-го порядка вариационного ряда определяется по формуле

$$\tilde{\nu}_k = \frac{1}{n}\sum_i x_i^k n_i. \tag{2.5}$$

Центральный момент $\tilde{\mu}_k$ k-го порядка вариационного ряда определяется по формуле

$$\tilde{\mu}_k = \frac{1}{n}\sum_i \left(x_i - \bar{x}\right)^2 n_i. \tag{2.6}$$

В частности $\tilde{\nu}_1 = \bar{x}$, $\tilde{\mu}_1 = 0$, $\tilde{\mu}_2 = D_{\text{В}}$.

Центральные моменты первых четырех порядков выборки $\{x_i\}$ выражаются через начальные моменты $\tilde{\nu}_k$ по формулам

$$\begin{gathered}\tilde{\mu}_2 = \tilde{\nu}_2 - \tilde{\nu}_1^2, \;\; \tilde{\mu}_3 = \tilde{\nu}_3 - 3\tilde{\nu}_1\tilde{\nu}_2 + 2\tilde{\nu}_1^3, \\ \tilde{\mu}_4 = \tilde{\nu}_4 - 4\tilde{\nu}_1\tilde{\nu}_3 + 6\tilde{\nu}_2\tilde{\nu}_1^2 - 3\tilde{\nu}_1^4.\end{gathered} \tag{2.7}$$

Выборочным коэффициентом асимметрии вариационного ряда называется число

$$\tilde{A}_s = \frac{\tilde{\mu}_3}{\sigma_{\text{В}}^3}. \tag{2.8}$$

Выборочным эксцессом вариационного ряда называется число

$$\tilde{E}_x = \frac{\tilde{\mu}_4}{\sigma_{\text{В}}^4} - 3. \tag{2.9}$$

Асимметрия называется также нормированным третьим центральным моментом, а эксцесс – нормированным четвертым центральным моментом признака X. Знаки асимметрии и эксцесса указывают на отклонения графика закона распределения X от нормального распределения, для которого $\tilde{A}_s = 0$; $\tilde{E}_x = 0$. При $\tilde{A}_s > 0$ большая часть вариант будет расположена слева от $\bar{x}$ – имеет место левосторонняя асимметрия распределения, при $\tilde{A}_s < 0$ – правосторонняя. Если $\tilde{A}_s = 0$ – то распределение имеет симметричную форму.

Положительное значение эксцесса указывает на то, что полигон распределения около моды имеет более высокую острую вершину, чем нормальная кривая, с тем же центром и той же дисперсией.

Отрицательное значение эксцесса имеет место для кривых с более низким и более плоским характером вершины по сравнению с нормальной кривой.

Точечные оценки не указывают величину ошибки, которая совершается при замене $M(X)$ и $\sigma(X)$ их приближенными значениями (оценками). Поэтому иногда выгодно пользоваться *интервальной оценкой*, которая определяется двумя числами $\tilde{\theta}_1$ и $\tilde{\theta}_2$ – концами интервала, накрывающего оцениваемый параметр θ с заданной вероятностью (*надежностью*).

Пусть $\tilde{\theta}$ – точечная оценка параметра θ. Она тем лучше, чем меньше разность $|\theta - \tilde{\theta}|$. Тогда в качестве характеристики *точности* оценки можно взять некоторое $\delta > 0$, такое, что $|\theta - \tilde{\theta}| < \delta$. Но в статистике можно говорить лишь о вероятности (надежности) $\gamma = 1 - \alpha$, с которой выполняется это неравенство. Число α называется *уровнем значимости*.

Доверительной вероятностью оценки называется вероятность $\gamma = 1 - \alpha$ выполнения неравенства $|\theta - \tilde{\theta}| < \delta$. Обычно γ задается заранее и наиболее часто полагают $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99$ и пр. Таким образом,

$$P\left(\left|\theta - \tilde{\theta}\right| < \delta\right) \Leftrightarrow P\left(\tilde{\theta} - \delta < \theta < \tilde{\theta} + \delta\right).$$

Доверительный интервал – это интервал $\left(\tilde{\theta} - \delta,\ \tilde{\theta} + \delta\right)$, который накрывает неизвестный параметр θ с заданной надежностью $\gamma = 1 - \alpha$.

Границы интервала и его величина находятся по выборочным данным и поэтому являются случайными величинами в отличие от оцениваемого параметра θ, поэтому говорят, что $\left(\tilde{\theta} - \delta,\ \tilde{\theta} + \delta\right)$ накрывает, а не содержит истинное значение θ.

Величина доверительного интервала существенно зависит от объема выборки n (уменьшается с ростом n) и значения доверительной вероятности γ (увеличивается с приближением γ к единице).

Интервальной оценкой с надежностью γ математического ожидания a нормально распределенной случайной величины (признака) X по выборочной средней $\bar{x}$ при известном среднем квадратическом отклонении σ генеральной совокупности служит доверительный интервал

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

где $\delta = t \cdot \dfrac{\sigma}{\sqrt{n}}$ – точность оценки;

n – объем выборки;

t – значение аргумента функции Лапласа $\Phi(t)$, при котором $\Phi(t) = \dfrac{\gamma}{2}$.

При неизвестном σ (в условиях эксперимента σ обычно неизвестно) доверительный интервал для математического ожидания a. нормально распределенной случайной величины X имеет вид

$$\overline{x} - t_\gamma \cdot \frac{s}{\sqrt{n}} < a < \overline{x} + t_\gamma \cdot \frac{s}{\sqrt{n}}, \tag{2.10}$$

где t_γ – квантиль распределения Стьюдента, определяемый по таблицам (прил. 1), а параметры $\overline{x}$, s находятся по данным выборки.

Доверительный интервал для σ задается неравенствами

$$\begin{gathered} s \cdot (1 - q) < \sigma < s \cdot (1 + q), \text{ если } q < 1, \\ 0 < \sigma < s \cdot (1 + q), \text{ если } q > 1. \end{gathered} \tag{2.11}$$

Величина q находится по таблице $q = q(\gamma, n)$ и зависит от надежности и объема выборки.

2.2. Примеры решения задач

1. Из генеральной совокупности извлечена выборка объемом $n = 21$:

x_i	2	6	8	10	$\sum$
n_i	3	4	6	8	21

Требуется найти выборочное среднее, «исправленное» среднее квадратическое отклонение, выборочную моду и медиану.

Решение. Определим выборочное среднее выборки по формуле (2.1):

$$\overline{x} = \frac{1}{21}\left(2 \cdot 3 + 6 \cdot 4 + 8 \cdot 6 + 10 \cdot 8\right) = \frac{116}{21} \approx 7,52.$$

«Исправленную» дисперсию найдем, используя формулу (2.6):

$$s^2 = \frac{\left(2-7,52\right)^2 \cdot 3 + \left(6-7,52\right)^2 \cdot 4 + \left(8-7,52\right)^2 \cdot 6 + \left(10-7,52\right)^2 \cdot 8}{20} \approx 7,2,$$

$$s = \sqrt{7,2} \approx 2,683.$$

Так как мода – это варианта, которой соответствует наибольшая частота, то $Mo = 10$.

Несгруппированные данные образуют дискретный вариационный ряд, содержащий нечетное число вариант $\left(n = 21 = 2 \cdot 10 + 1\right)$.

i	1	2	3	4	5	6	7	8	9	10	11
x_i	2	2	2	6	6	6	6	8	8	8	8
i	12	13	14	15	16	17	18	19	20	21	
x_i	8	8	10	10	10	10	10	10	10	10	

Значит, медиана равна $Me = x_{10} = 8$.

2. Записать в виде вариационного ряда выборку 20, 19, 12, 13, 16, 17, 17, 14, 16, 14, 13, 19, 18, 16, 14. Представить статистическое распределение выборки. Построить полигон относительных частот для статистического ряда. Вычислить числовые характеристики выборки: выборочное среднее, «исправленную» и выборочную дисперсии, «исправленное» среднеквадратическое отклонение (СКО).

Решение. Объем выборки $n = 15$. Упорядочив элементы выборки по возрастанию, получим вариационный ряд

12, 13, 13, 14, 14, 14, 16, 16, 16, 17, 17, 18, 19, 19, 20.

Статистическое распределение исходной выборки можно записать в виде табл. 2.1.

Таблица 2.1

x_i	12	13	14	16	17	18	19	20	Сумма
n_i	1	2	3	3	2	1	2	1	15
p_i^*	1/15	2/15	3/15	3/15	2/15	1/15	2/15	1/15	1

Полигон относительных частот изображен на рис. 2.1.

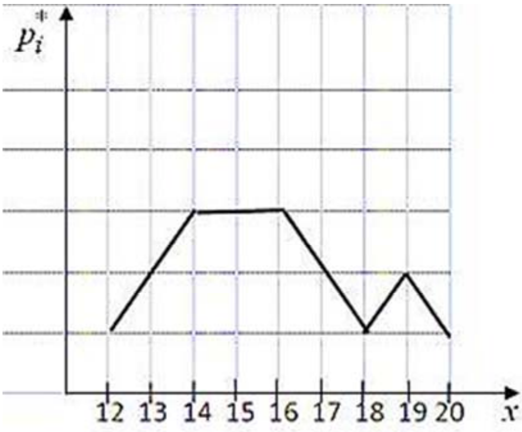

Рис. 2.1. Полигон относительных частот

Находим выборочное среднее по формуле (2.1):

$$x = \frac{12 + 13 \cdot 2 + 14 \cdot 3 + 16 \cdot 3 + 17 \cdot 2 + 18 + 19 \cdot 2 + 20}{15} = \frac{238}{15} \approx 15,9.$$

Для вычисления выборочной дисперсии используем формулу (2.2):

$$D_B = \frac{(12 - 15,9)^2 + 2 \cdot (13 - 15,9)^2 + 3 \cdot (14 - 15,9)^2 + \ldots + (20 - 15,9)^2}{15} =$$

$$= \frac{85,75}{15} \approx 5,72.$$

«Исправленная» дисперсия и СКО:

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{15}{14} \cdot 5,72 \approx 6,13; \quad s = \sqrt{6,13} \approx 2,48.$$

3. Найти выборочное среднее, выборочное и «исправленное» СКО выборки объемом $n = 70,$ распределение которой задано табл. 2.2.

Таблица 2.2

Интервалы	0–1,02	1,02–2,04	2,04–3,06
Частота	4	28	19
Интервалы	3,06–4,08	4,08–5,1	Сумма
Частота	12	7	70

Решение.

Для расчета выборочного среднего и выборочного СКО составляем вариационный ряд, принимая в качестве вариант середины соответствующих интервалов (табл. 2.3).

Таблица 2.3

$x_i - x_{i+1}$	0–1,02	1,02–2,04	2,04–3,06	3,06–4,08	4,08–5,1	$\sum$
x_i^*	0,51	1,53	2,55	3,57	4,59	
m_i	4	28	19	12	7	70
$x_i^* \cdot m_i$	2,04	42,84	48,45	42,84	32,13	168,3
$\left(x_i^*\right)^2 \cdot m_i$	1,04	65,545	123,548	152,939	147,477`	490,549

Таким образом, $\overline{x} = \frac{168,3}{70} = 2,404;$ $\overline{x^2} = \frac{490,549}{70} = 7,008;$

$$D_{\text{В}} = \overline{x^2} - \left(\overline{x}\right)^2 - 7,008 - 2,404^2 = 1,229; \; s^2 = \frac{70}{69} \cdot 1,229 = 1,434.$$

Тогда $\sigma_{\text{В}} = \sqrt{1,229} \approx 1,11;$ $s = \sqrt{1,434} \approx 1,197.$

4. Из генеральной совокупности извлечена выборка, представленная в виде дискретного ряда.

x_i	6	8	10	12	15	18	20
n_i	5	7	10	14	10	8	6

Требуется:

1) вычислить выборочное среднее $\overline{x}$, выборочную дисперсию $D_{\text{В}}$, исправленную выборочную дисперсию s^2 и среднее квадратичное отклонение s;

2) найти с доверительной вероятностью $\gamma = 0,95$ доверительный интервал для математического ожидания, а также доверительный интервал для $\sigma(x)$, предполагая, что случайная величина распределена по нормальному закону;

3) вычислить моду и медиану;

4) построить эмпирическую функцию распределения.

Решение. 1) Объем выборки равен

$$n = \sum_{i=1}^{n} n_i = 5 + 7 + 10 + 14 + 10 + 8 + 6 = 60.$$

Для расчета выборочного среднего и дисперсии составим следующую вспомогательную табл. 2.4.

Таблица 2.4

x_i	6	8	10	12	15	18	20	$\sum$
n_i	5	7	10	14	10	8	6	60
$x_i \cdot n_i$	30	56	100	168	150	144	120	768
$x_i^2 \cdot n_i$	180	448	1000	2016	2250	2592	2400	10886

Тогда, согласно формулам (2.1), (2.2), (2.4)

$$\overline{x} = \frac{768}{60} = 12,8; \quad D_{\text{В}} = \frac{10886}{60} - 12,8^2 = 17,6; \quad s^2 = \frac{60}{59} \cdot 17,6 \approx 17,9;$$

$$s = \sqrt{17,9} \approx 4,23.$$

2) Доверительный интервал для математического ожидания найдем по формуле (2.10). Значение t_γ определим из таблицы (прил. 1) по доверительной вероятности $\gamma = 0,95$ и объему выборки $n = 60$: $t_\gamma = 2,001$. Тогда

$$\delta = t_\gamma \cdot \frac{s}{\sqrt{n}} = 2,001 \cdot \frac{4,23}{\sqrt{60}} \approx 1,093 \Rightarrow 12,8 - 1,093 = 11,707;$$

$$12,8 + 1,093 = 13,893.$$

Доверительный интервал для математического ожидания имеет вид:

$$11,707 < a < 13,893.$$

Доверительный интервал для дисперсии определим по формуле (2.11). Так как (см. прил. 2) $q = q(0,95;60) = 0,188 < 1,$ то границы интервала принимают вид:

$$s(1-q) = 4,23 \cdot (1 - 0,188) = 3,435;$$

$$s(1+q) = 4,23 \cdot (1 + 0,188) = 5,025.$$

Т. е. $3,435 < \sigma < 5,025.$

3) Вычислим медиану и моду. Так как $n = 60 = 2 \cdot 30,$ то

$$Me = \frac{1}{2}(x_{30} + x_{31}) = \frac{12+12}{2} = 12.$$

Мода тоже равна 12, так как именно варианта $x = 12$ имеет наибольшую частоту, равную 14.

4) Для построения эмпирической функции распределения нужно найти *относительные частоты* $p_i^* = \dfrac{n_i}{n},$ а также *накопленные относительные частоты* w_i.

Для нахождения относительных и накопленных относительных частот составим табл. 2.5.

Таблица 2.5

x_i	6	8	10	12	15	18	20	Сумма
n_i	5	7	10	14	10	8	6	60
p_i^*	0,08	0,12	0,17	0,23	0,17	0,13	0,01	1
w_i	0,08	0,2	0,37	0,6	0,77	0,9	1	

Аналитическое выражение $F^*(x)$ через накопленные относительные частоты (см. формулу (1.2)):

$$F^*(x) = \begin{cases} 0 & x \le 6 \\ 0{,}08 & 6 < x \le 8 \\ 0{,}2 & 8 < x \le 10 \\ 0{,}37 & 10 < x \le 12 \\ 0{,}6 & 12 < x \le 15 \\ 0{,}77 & 15 < x \le 18 \\ 0{,}9 & 18 < x \le 20 \\ 1 & x > 20 \end{cases}.$$

График эмпирической функции изображен на рис. 2.2.

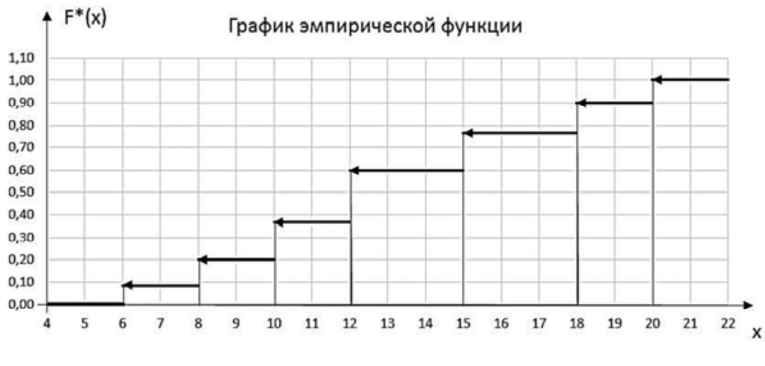

Рис. 2.2. График эмпирической функции распределения

5. Дан вариационный ряд признака X:

x_i	1	3	5	7	9	$\sum$
n_i	10	40	20	20	10	100

Найти начальные и центральные моменты первых четырех порядков, а также асимметрию и эксцесс.

Решение. Вычисления проводим по формулам (2.5) для $\tilde{\nu}_k$ и (2.7) для $\tilde{\mu}_k$.

Начальные моменты:

$$\tilde{\nu}_1 = \frac{1}{100}\left(1\cdot 10 + 3\cdot 40 + 5\cdot 20 + 7\cdot 20 + 9\cdot 10\right) = \frac{460}{100} = 4,6;$$

$$\tilde{\nu}_2 = \frac{1}{100}\left(1\cdot 10 + 9\cdot 40 + 25\cdot 20 + 49\cdot 20 + 81\cdot 10\right) = \frac{2660}{100} = 26,6;$$

$$\tilde{\nu}_3 = \frac{1\cdot 10 + 27\cdot 40 + 125\cdot 20 + 343\cdot 20 + 729\cdot 10}{100} = \frac{17740}{100} = 177,4;$$

$$\tilde{\nu}_4 = \frac{1\cdot 10 + 81\cdot 40 + 625\cdot 20 + 2401\cdot 20 + 6561\cdot 10}{100} = 1293,8.$$

Центральные моменты:

$$\tilde{\mu}_1 = 0;$$

$$\tilde{\mu}_2 = 26,6 - 4,6^2 = 5,44; \quad D_{\text{В}} = \tilde{\mu}_2 = 5,44; \quad \sigma_{\text{В}} = \sqrt{5,44} = 2,33;$$

$$\tilde{\mu}_3 = 177,4 - 3\cdot 4,6\cdot 26,6 + 2\cdot 4,6^3 = 4,992;$$

$$\tilde{\mu}_4 = 1293,8 - 4\cdot 4,6\cdot 177,4 + 6\cdot 26,6\cdot 4,6^2 - 3\cdot 4,6^4 = 63,5392.$$

Вычислим значения асимметрии и эксцесса, используя формулы (2.8) и (2.9):

$$\tilde{A}s = \frac{4,992}{2,33^3} \approx 0,3946; \quad \tilde{E}x = \frac{63,5392}{2,33^4} - 3 \approx -0,8442.$$

2.3. Задачи для самостоятельного решения

1. Инженер по контролю качества продукции обнаружил в 10 партиях электроламп, произведенных заводом, следующее число бракованных изделий: 5, 3, 7, 1, 0, 6, 3, 4, 5, 2 . Найти среднее число (выборочное среднее) и стандартное отклонение («исправленое» среднее квадратическое отклонение) бракованных ламп. Вычислить медиану.

2. Найти несмещенную оценку дисперсии случайной величины X по заданному вариационному ряду

x_i	3	6	8	10
n_i	8	14	10	18

3. Найти выборочное среднее, выборочную дисперсию, а также «исправленное» среднее квадратическое отклонение по заданному распределению выборки. Построить полигон частот и эмпирическую функцию распределения.

а)	x_i	10,6	10,8	11	11,2	11,4	11,6
	n_i	5	10	17	30	20	12

б)	x_i	2	3	5	6	8	9	10
	n_i	4	10	21	30	20	10	5

4. Найти точечные оценки параметров a и σ нормального распределения, исходя из данной выборки:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5
n_i	6	9	26	25	30	26	21
x_i	1,7	1,9	2,1	2,3			
n_i	24	20	8	5			

5. Найти выборочное среднее и «исправленную» дисперсию для интервального статистического ряда. Построить гистограмму относительных частот.

Границы интервалов	$[0;\ 0{,}02)$	$[0{,}02;\ 0{,}04)$	$[0{,}04;\ 0{,}06)$	$[0{,}06;\ 0{,}08]$
m_i	4	8	5	14

6. Ниже приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Рост	154–158	158–162	162–166	166–170	170–174	174–178	178–182
Число студентов	10	14	26	28	12	8	2

Найти выборочное среднее, «исправленное» среднее квадратическое отклонение, выборочную моду и медиану.

7. Среднее время сборки изделия составляло 90 минут. Инженер изобрел новый метод сборки этого изделия, и продолжительность сборки 10 изделий новым способом составила 79; 74; 112; 95; 83; 96; 77; 84; 70; 90 (мин.). Построить доверительный интервал для нового среднего времени сборки с надежностью 95 %.

8. Из генеральной совокупности извлечена выборка:

x_i	0,2	0,4	0,8	0,9	1	1,1	1,2	1,3	1,5	1,8
n_i	5	7	12	15	20	18	13	5	3	2

Оценить с надежностью 0,99 математическое ожидание a нормально распределенного признака генеральной совокупности по выборочному среднему при помощи доверительного интервала. Построить доверительный интервал для дисперсии.

9. Дан статистический ряд признака X. Найти начальные и центральные моменты первых четырех порядков признака, а также определить асимметрию и эксцесс.

x_i	2,6	3	3,4	3,8	4,2
n_i	8	20	45	15	12

10. В итоге многократных измерений некоторой физической величины одним прибором получены результаты, представленные в таблице. Построить по этим данным:

1) интервальный вариационный ряд с равными интервалами, полигон частот и гистограмму относительных частот;

2) вычислить основные числовые характеристики выборки: выборочное среднее, дисперсию, среднее квадратическое отклонение, моду и медиану;

3) найти с доверительной вероятностью $\gamma = 0{,}95$ доверительный интервал для математического ожидания, а также доверительный интервал для среднего квадратического отклонения.

19,2	20,0	18,2	21,6	20,3	18,9	16,5	17,2	18,7	20,5
21,3	12,9	19,5	21,3	16,7	21,2	17,5	17,5	15,8	16,7
17,6	19,6	15,8	18,5	15,4	12,2	17,1	12,6	23,6	17,8
15,5	16,9	16,1	15,5	16,6	19,6	19,9	16,9	16,0	18,2
19,8	15,0	19,0	19,2	17,0	16,3	14,9	16,8	18,8	16,5

Ответы

1	$\bar{x} = 3{,}6;\ s = 2{,}221;\ Me = 3{,}5$	**2**	$s^2 = 6{,}28$
3	а) $\bar{x} = 11{,}18;\ D_B = 0{,}0714;\ s = 0{,}27$ б) $\bar{x} = 6{,}23;\ D_B = 4{,}197;\ s = 2{,}06$	**4**	$\bar{x} = 1{,}26;\ \sigma_B = 0{,}49$
5	$\bar{x} = 0{,}049;\ s^2 = 0{,}0005$	**6**	$\bar{x} = Me = 166;\ s = 5{,}81;$ $Mo = 166{,}4$
7	$(77{,}038;\ 94{,}962)$	**8**	$0.91 < a < 1{,}03;$ $0{,}074 < \sigma^2 < 0{,}129$
9	$\tilde{A}s = 0{,}1352;$ $\tilde{E}x = -0{,}337$	**10**	$\bar{x} = 17{,}71;\ D_B = 5{,}042;$ $s^2 = 5{,}145;\ s = 2{,}268;$ $Mo = 19{,}2;\ Me = 17{,}6;$ $17{,}07 < a < 18{,}35;$ $1{,}79 < \sigma < 2{,}74$

Проверочный тест 2

Условие задачи	Варианты ответов
1. Статистическая оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, называется…	1) Несмещенной 2) Состоятельной 3) Эффективной 4) Прямой 5) Обратной
2. Статистическая оценка, которая (при заданном объеме выборки) имеет наименьшую возможную дисперсию, называется	1) Несмещенной 2) Состоятельной 3) Эффективной 4) Прямой 5) Обратной
3. Проведено 5 измерений (без систематических ошибок) некоторой случайной величины: 5, 7, 8, 9, 10. Тогда несмещенная оценка математического ожидания равна…	1) 7 2) 7,2 3) 7,5 4) 7,8 5) Другой ответ
4. Мода вариационного ряда 21; 21; 23; 23; 23; 23; 24; 25; 27; 29; 29; 30 равна...	1) 27 2) 21 3) 29 4) 23 5) Другой ответ
5. Медиана вариационного ряда 22; 23; 24; 25; 26; 28; 28; 28; 31; 32 равна	1) 22 2) 27 3) 24 4) 28 5) Другой ответ
6. В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты (в мм): 14, 16, 17. Тогда несмещенная оценка дисперсии измерений равна…	1) 4 2) 3 3) 5 4) 2 5) Другой ответ

Условие задачи	Варианты ответов
7. Чему равна сумма доверительной вероятности и уровня значимости $\gamma + \alpha$?	1) 0 2) неотрицательному числу 3) какому-то числу от 0 до 1 4) 1
8. Дан доверительный интервал (18,5; 25,9) для оценки математического ожидания нормально распределенного количественного признака. Тогда точечная оценка математического ожидания равна…	1) 22,2 2) 22,5 3) 22 4) 22,6 5) Другой ответ
9. Точечная оценка параметра распределения равна 21. Тогда его интервальная оценка может иметь вид…	1) (20;21) 2) (20;22) 3) (0;21) 4) (21;22) 5) Другой ответ
10. Ширина доверительного интервала зависит от …	1) уровня значимости и числа наблюдений 2) уровня значимости 3) числа наблюдений

Ответы к проверочному тесту 2

№	1	2	3	4	5	6	7	8	9	10
	1	3	4	4	2	5	4	1	2	1

§ 3. СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

3.1. Краткие теоретические сведения

Статистической гипотезой называется предположение относительно параметров или вида распределения изучаемой случайной величины. Гипотеза называется *параметрической,* если в ней содержится некоторое утверждение о параметрах распределения случайной величины (когда сам закон распределения считается известным), и *непараметрической* – в иных случаях.

Нулевой H_0 *(основной)* гипотезой называется предположение, которого мы придерживаемся изначально, пока наблюдения не заставят нас признать обратное.

Альтернативной (*конкурирующей*) гипотезой H_1 называется гипотеза, которая противоречит H_0, и которую мы принимаем, если отвергаем основную гипотезу.

Случайная величина K, построенная по наблюдениям для проверки нулевой гипотезы, называется *статистикой критерия*.

Схема построения критерия такова: все выборочное пространство делится на две взаимодополняющие области – область отклонения основной гипотезы H_0 и область принятия этой гипотезы. Область $K_{\text{кр}}$, при попадании в которую выборочной точки отвергается основная гипотеза, называется *критической.* При проверке гипотез могут быть ошибки двух родов.

Ошибка первого рода: H_0 отвергается (принимается H_1) в то время, как в действительности верна гипотеза H_0. Вероятность ошибки первого рода называют *уровнем значимости* и обозначают α

$$P\left(H_1 / H_0\right) = \alpha.$$

Величину $\gamma = 1 - \alpha$, то есть вероятность принять верную гипотезу, называют *уровнем доверия* (*доверительным уровнем*).

Ошибка второго рода: H_0 принимается в то время, как верна гипотеза H_1. Вероятность ошибки второго рода обозначается β.

Вероятность принять гипотезу H_1, если она верна, называют *мощностью критерия.*

Вычисленное по выборке значение критерия называют *наблюдаемым значением* $K_{\text{набл}}$.

Критическими точками (*границами*) называют точки $k_{\text{кр}}$, отделяющие критическую область от области принятия гипотезы. Критические области разделяются на правосторонние и левосторонние. *Правосторонняя* область определяется неравенством $K > k_{\text{кр}}$, *левосторонняя* – $K < k_{\text{кр}}$. Это односторонние области. Существуют также и двусторонние области, определяемые неравенствами $K < k_{\text{кр}1}$, $K > k_{\text{кр}2}$, где $k_{\text{кр}2} > k_{\text{кр}1}$ ($k_{\text{кр}1}$ и $k_{\text{кр}2}$ – критические точки).

Для каждого критерия, т. е. соответствующего распределения, обычно составлены таблицы, по которым находят $k_{\text{кр}}$.

После того как критическая точка найдена, по данным выборки вычисляют наблюдаемое значение критерия. Если $K_{\text{набл}} > k_{\text{кр}}$ (для правосторонней области) нулевую гипотезу отвергают, если наоборот, то принимают.

Проверка гипотезы о нормальном распределении генеральной совокупности. Критерий согласия Пирсона.

Одной из задач математической статистики является установление истинного закона распределения случайной величины на основании экспериментальных данных. Критерии, устанавливающие закон распределения, называются *критериями согласия*.

Алгоритм применения критерия Пирсона.

1) Из генеральной совокупности образовывается случайная выборка, и на ее основе делается предположение о нормальном законе распределения. Выдвигается гипотеза H_0: «*генеральная совокупность распределена нормально*».

2) Вычисляются выборочные числовые характеристики $\overline{x}$, σ_B.

3) Вычисляются теоретические частоты.

а) Для дискретного ряда

$$n_i' = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i),$$

где n – объем выборки;

h – шаг (разность между двумя соседними вариантами).

$$u_i = \frac{x_i - \overline{x}}{\sigma_B};\ \varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}.$$

Значения $\varphi(u)$ определяются из таблицы прил. 3.

б) Для интервального ряда $n_i' = n \cdot p_i$, где n – объем выборки, $p_i = \Phi(z_{i+1}) - \Phi(z_i)$ – теоретические вероятности попадания в интервалы $x_i - x_{i+1}$, $z_i = \dfrac{x_i - \overline{x}}{\sigma_B}$, $z_{i+1} = \dfrac{x_{i+1} - \overline{x}}{\sigma_B}0$, $\Phi(z)$ – функция Лапласа, значения которой определяются по таблице прил. 4.

4) Находится наблюдаемое значение критерия Пирсона по формуле

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

5) По таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $k = l - 3$ (l – число групп для дискретного ряда или число интервалов для интервального ряда) находят критическую точку $\chi^2_{kp}(\alpha; k)$ правосторонней критической области.

6) Если $\chi^2 < \chi^2_{kp}(\alpha;\ k)$ – нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо. Если $\chi^2 > \chi^2_{kp}(\alpha;\ k)$ – гипотезу отвергают.

Замечание. Малочисленные варианты и интервалы (содержащие малочисленные частоты $n_i < 5$) следует объединить, а соответствующие им частоты – сложить. Если производилось объединение частот, то в формуле $k = l - 3$ следует в качестве l принять число групп или интервалов выборки, оставшихся после объединения частот.

3.2. Примеры решения задач

1. Требуется при уровне значимости $\alpha = 0,05$ проверить по критерию согласия Пирсона гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические частоты n_i и теоретические частоты n_i':

n_i	8	10	18	27	17	11	9
n_i'	5	15	16	25	20	12	7

Решение. Определим наблюдаемое значение критерия Пирсона по формуле

$$\chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - n_i'\right)^2}{n_i'} = \frac{(8-5)^2}{5} + \frac{(10-15)^2}{15} +$$

$$+ \frac{(18-16)^2}{16} + \ldots + \frac{(9-7)^2}{7} \approx 5.$$

В таблице критических точек χ^2 (прил. 5) находим при уровне значимости $\alpha = 0,05$ значение $\chi^2_{kp}(0,05;\, 4) = 9,488$ (имеем $k = l - 3 =$ $= 7 - 3 = 4$ степени свободы). Значение $\chi^2 = 5 < \chi^2_{kp}$. Следовательно, выдвинутая гипотеза о нормальном распределении генеральной совокупности не отвергается.

2. Дан вариационный ряд случайной величины X с $n = 100$ вариантами.

x_i	2	5	8	11	14	17	20	$\sum$
n_i	9	13	20	25	15	10	8	100

Построить теоретический закон распределения случайной величины X. Используя критерий Пирсона при уровне значимости $\alpha = 0,01$, установить, согласуется ли полученный закон с данной выборкой. Найти также асимметрию и эксцесс эмпирического распределения и дать им надлежащее толкование.

Решение. Первую рабочую таблицу (табл. 3.1) используем для вычисления эмпирических начальных и центральных моментов данного вариационного ряда.

Таблица 3.1

x_i	n_i	$x_i n_i$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$
2	9	18	36	72	144
5	13	65	325	1625	8125
8	20	160	1280	10240	81920
11	25	275	3025	33275	366025
14	15	210	2940	41160	576240
17	10	170	2890	49130	835210
20	8	160	3200	6400	1280000
$\sum$	100	1058	13696	199502	3147664

При помощи таблицы, применяя формулы (2.5)–(2.9), находим последовательно

$$\tilde{\nu}_1 = \overline{x} = \frac{1058}{100} = 10,58;\ \tilde{\nu}_2 = \frac{13696}{100} = 136,96;$$

$$\tilde{\nu}_3 = \frac{199502}{100} = 1995,02;\ \tilde{\nu}_4 = \frac{3147664}{100} = 31476,64.$$

$$\tilde{\mu}_1 = 0;\ \tilde{\mu}_2 = \sigma_B^2 = 136,96 - 10,58^2 = 25,0236;\ \sigma_B = \sqrt{25,0236} \approx 5,002;$$

$$\tilde{\mu}_3 = 1995,02 - 3 \cdot 10,58 \cdot 136,96 - 2 \cdot 10,58^3 = 16,48382;$$

$$\tilde{\mu}_4 = 31476,64 - 4 \cdot 10,58 \cdot 1995,02 + 6 \cdot 136,96 \cdot 10,58^2 -$$
$$-3 \cdot 10,58^4 = 1442,977.$$

$$\tilde{A}s = \frac{16,48382}{5,002^3} \approx 0,132;\ \tilde{E}x = \frac{1442,977}{5,002^4} - 3 \approx -0,7.$$

Построим (в пакете MS Excel) эмпирический полигон (рис. 3.1) и, исходя из его внешнего вида, выдвинем гипотезу: случайная величина X распределена по нормальному закону с параметрами $a = \overline{x} = 10{,}58$; $\sigma = \sigma_B = 5{,}002$.

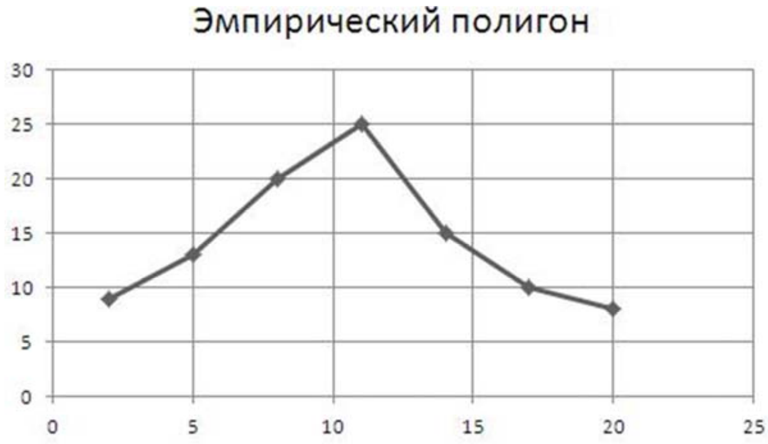

Рис. 3.1. Эмпирический полигон

Тем самым теоретический закон распределения имеет вид (принимаем $\sigma = 5$):

$$f(x) = \frac{1}{5 \cdot \sqrt{2\pi}} \cdot e^{\frac{-(x-10{,}58)^2}{50}}.$$

Для проверки того, насколько полученный закон согласуется с эмпирическими данными, построим новую табл. 3.2, в которую впишем теоретические частоты, отклонения теоретических и эмпирических частот и определим наблюдаемое значение критерия:

$$\chi_H^2 = \sum_i \frac{(n_i - n_i')^2}{n_i'}.$$

Предположим $u_i = \dfrac{x_i - \overline{x}}{\sigma_B}$ (принимаем $\sigma = 1$). Тогда

$$n_i' = \frac{h \cdot n}{\sigma_B} \cdot \varphi(u_i) \approx 60 \cdot \varphi(u_i),$$

где $\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}}$.

Таблица 3.2

x_i	n_i	u_i	$\varphi(u_i)$	n_i'	$\frac{(n_i - n_i')^2}{n_i'}$
2	9	–1,716	0,0915	5,49	2,243
5	13	–1,116	0,2140	12,84	0,002
8	20	–0,516	0,3492	20,95	0,043
11	25	0,084	0,3975	23,85	0,055
14	15	0,684	0,3157	18,94	0,821
17	10	1,284	0,1749	10,50	0,024
20	8	1,884	0,0676	4,06	3,829
$\sum$	100				7,017

Из таблицы находим $\chi_H^2 = 7,017$, а при $\alpha = 0,01$ и числе степеней свободы $k = 7 - 2 - 1 = 4$ ($l = 7$ – число групп выборки, $m = 2$ – число параметров задачи) из таблицы критических точек χ^2 находим $\chi_{kp}^2 = 13,3$. Следовательно, гипотеза о нормальном распределении генеральной совокупности не отвергается.

Эмпирический полигон, изображенный на рис. 3.1, скошен вправо (что соответствует $\tilde{A}s = 0,132$) и «слегка низковершинный» (что согласуется с $\tilde{E}x = -0,7$).

3. Из генеральной совокупности извлечена выборка, представленная в виде ряда.

Требуется проверить, согласуются ли выборочные данные с гипотезой о нормальном распределении случайной величины X с помощью критерия согласия Пирсона при уровне значимости

$\alpha = 0,05,$ разбив отрезок $[x_{\min};\ x_{\max}]$ на l интервалов одинаковой длины. Величину l рассчитать по формуле Стерджеса.

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

Решение. Подсчитаем количество интервалов разбиения:

$$l = 1 + 3,322 \cdot \lg n = 1 + 3,322 \cdot \lg 100 = 7,644 \approx 8.$$

Так как $x_{\min} = 0,6;\ x_{\max} = 4,2,$ то $h = \dfrac{4,2 - 0,6}{8} \approx 0,5.$

Границы интервалов:

$$x_0 = 0,6;\ x_1 = 0,6 + 0,5 = 1,1;\ x_2 = 1,1 + 0,5 = 1,6;\ x_3 = 1,6 + 0,5 = 2,1;$$

$$x_4 = 2,1 + 0,5 = 2,6; x_5 = 2,6 + 0,5 = 3,1;\ x_6 = 3,1 + 0,5 = 3,6;$$

$$x_7 = 3,1 + 0,5 = 4,1;\ x_8 = 4,1 + 0,5 = 4,6.$$

Частота n_i интервала $[x_i;\ x_{i+1})$ $\left(i = \overline{0,7}\right)$ подсчитывается с помощью ряда как число наблюдений, попавших в интервал. Так, в первый $(i = 0)$ интервал $[0,6;\ 1,1)$ попало 7 значений, во второй $[1,1;\ 1,6)$ – 14 значений. Сведем полученные данные в табл. 3.3.

Таблица 3.3

$x_i - x_{i+1}$	0,6–1,1	1,1–1,6	1,6–2,1	2,1–2,6	2,6–3,1	3,1–3,6	3,6–4,1	4,1–4,6
x_i^*	0,85	1,35	1,85	2,35	2,85	3,35	3,85	4,35
n_i	7	14	16	18	16	15	10	4

Объем выборки равен $n=\sum n_i=100.$ Выборочное среднее и дисперсия определяются по формулам

$$\overline{x}=\frac{\sum x_i n_i}{n}=\frac{1}{100}\left(0{,}85\cdot 7+1{,}35\cdot 14+1{,}85\cdot 16+\ldots+4{,}35\cdot 4\right)=$$

$$=\frac{248{,}5}{100}=2{,}485;$$

$$\sigma_B=\sqrt{\frac{\sum\left(x_i^*-\overline{x}\right)\cdot n_i}{n}}=\sqrt{\frac{\left(0{,}85-2{,}485\right)^2\cdot 7+\ldots+\left(4{,}35-2{,}485\right)^2\cdot 4}{100}}=$$

$$=\sqrt{0{,}89}\approx 0{,}95.$$

Найдем теоретические вероятности p_i по формуле

$$p_i=P\left(z_i<z<z_{i+1}\right)=\Phi\left(\frac{x_{i+1}-\overline{x}}{\sigma_B}\right)-\Phi\left(\frac{x_i-\overline{x}}{\sigma_B}\right),$$

где $\Phi(x)$ – функция Лапласа, значения которой даются в прил. 2. Результаты вычислений сведем в табл. 3.4.

Таблица 3.4

x_i	x_{i+1}	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	$n_i'=n\cdot p_i$
0,6	1,1	–	–1,46	–0,5	–0,4279	0,0721	7,21
1,1	1,6	–1,46	–0,94	–0,4279	–0,3264	0,1015	10,15
1,6	2,1	–0,94	–0,41	–0,3264	–0,1591	0,1673	16,73
2,1	2,6	–0,41	0,12	–0,1591	0,0478	0,2069	20,69
2,6	3,1	0,12	0,65	0,0478	0,2422	0,1944	19,44
3,1	3,6	0,65	1,18	0,2422	0,381	0,1388	13,88
3,6	4,1	1,18	1,71	0,381	0,4564	0,0754	7,54
4,1	4,6	1,71	–	0,4564	0,5	0,0436	4,36

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим следующую расчетную табл. 3.5.

Таблица 3.5

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	7	7,21	–0,210	0,044	0,006
2	14	10,15	3,850	14,823	1,460
3	16	16,73	–0,730	0,533	0,032
4	18	20,69	–2,690	7,236	0,350
5	16	19,44	–3,440	11,834	0,609
6	15	13,88	1,120	1,254	0,090
7	10	7,54	2,460	6,052	0,803
8	4	4,36	–0,360	0,130	0,030
$\sum$	**100**	**100**			**3,379**

По таблице критических точек распределения χ^2, для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $k = 8 - 2 - 1 = 5$ находим $\chi^2_{kp} = 11{,}1$. Так как $\chi^2 = 3{,}379 < \chi^2_{kp}(0{,}05; 5)$, то нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности.

3.3. Задачи для самостоятельного решения

1. Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}01$ установить, согласуется ли гипотеза о выборе закона распределения случайной величины X со случайной выборкой, если ее эмпирические и теоретические частоты заданы таблицей:

а)

n_i	14	18	32	70	20	36	10
n_i'	10	24	34	80	18	22	12

б)

n_i	6	8	13	15	20	16	10	7	5
n_i'	5	9	14	16	18	16	9	6	7

2. Используя критерий Пирсона, при уровне значимости $\alpha = 0,01$ установить, согласуется ли гипотеза о выборе нормального распределения признака X со случайной выборкой, заданной вариационным рядом:

а)	x_i	2	3	4	5	6	7	8
	n_i	4	11	25	30	15	10	5
б)	x_i	–10	–5	0	5	10	15	20
	n_i	3	17	24	32	20	5	1
в)	x_i	80	90	100	110	120	130	140
	n_i	4	8	14	40	16	12	6

Найти асимметрию и эксцесс эмпирического распределения и дать им надлежащее толкование.

3. В итоге многократных измерений некоторой физической величины одним прибором получены результаты, представленные в таблице. Проверить, согласуются ли выборочные данные с гипотезой о нормальном распределении случайной величины X графически и с помощью критерия согласия Пирсона при уровне значимости $\alpha = 0,05$.

а)

19,2	20,0	18,2	21,6	20,3	18,9	16,5	17,2	18,7	20,5
21,3	12,9	19,5	21,3	16,7	21,2	17,5	17,5	15,8	16,7
17,6	19,6	15,8	18,5	15,4	12,2	17,1	12,6	23,6	17,8
15,5	16,9	16,1	15,5	16,6	19,6	19,9	16,9	16,0	18,2
19,8	15,0	19,0	19,2	17,0	16,3	14,9	16,8	18,8	16,5

б)

1,03	1,1	0,95	0,75	0,58	0,59	0,56	0,69	0,71	0,65
1,02	0,66	0,68	0,57	0,54	0,7	0,72	0,82	0,83	0,91
0,65	0,95	0,85	0,55	0,74	0,76	0,77	0,82	0,66	0,69
0,99	0,88	1,02	1,05	0,81	0,79	0,72	0,61	0,58	0,59
0,91	0,81	0,71	0,73	0,84	0,69	0,82	0,9	0,8	0,77

Ответы

1	а) Не согласуется $\chi^2 = 13,9; \quad \chi^2_{kp} = 13,3$ б) Согласуется $\chi^2 = 1,5; \quad \chi^2_{kp} = 15,1$	**2**	а) Согласуется $\chi^2 = 4,1; \ \chi^2_{kp} = 13,3$ б) Согласуется $\chi^2 = 1,9; \ \chi^2_{kp} = 9,2$ в) не согласуется $\chi^2 = 9,809; \ \chi^2_{kp} = 5,9$
3	а) Согласуется $\chi^2 = 3,7; \ \chi^2_{kp} = 5,9$ б) Согласуется $\chi^2 = 3,75; \ \chi^2_{kp} = 7,8$		

Проверочный тест 3

Условие задачи	Варианты ответов
1. Статистической гипотезой называют предположение …	1) о равенстве двух параметров 2) о неравенстве двух величин 3) о виде или параметрах неизвестного закона распределения случайной величины
2. Нулевая гипотеза – это …	1) выдвинутая гипотеза, которую нужно проверить 2) альтернативная гипотеза 3) гипотеза, определяющая закон распределения 4) гипотеза о равенстве нулю параметра распределения
3. Альтернативная гипотеза – это …	1) выдвинутая гипотеза, которую нужно проверить 2) гипотеза, определяющая закон распределения 3) гипотеза, противоположная нулевой 4) гипотеза о равенстве нулю параметра распределения

Условие задачи	Варианты ответов
4. Что называют ошибкой первого рода?	1) гипотеза H_0 верна и ее принимают согласно критерию 2) гипотеза H_0 верна, но ее отвергают согласно критерию 3) гипотеза H_0 неверна и ее отвергают согласно критерию 4) гипотеза H_0 неверна, но ее принимают согласно критерию
5. Что называют ошибкой второго рода?	1) гипотеза H_0 верна и ее принимают согласно критерию 2) гипотеза H_0 верна, но ее отвергают согласно критерию 3) гипотеза H_0 неверна и ее отвергают согласно критерию 4) гипотеза H_0 неверна, но ее принимают согласно критерию
6. Если нулевую гипотезу в результате проверки критерия отвергают, какова вероятность при этом совершить ошибку?	1) β 2) α 3) $1-\beta$ 4) γ

Ответы к проверочному тесту 3

№	1	2	3	4	5	6
	3	1	3	2	4	2

§ 4. ЛИНЕЙНАЯ РЕГРЕССИЯ. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА

4.1. Краткие теоретические сведения

Одной из основных задач математической статистики является исследование зависимости между двумя или несколькими переменными. Строгая функциональная зависимость реализуется редко, так как одна или обе величины подвержены еще и случайным факторам.

Статистической называется зависимость, при которой изменение одной из величин влечет за собой изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой. В этом случае статистическую зависимость называют *корреляционной*.

Основная цель изучения зависимостей между случайными величинами заключается в прогнозировании с данной вероятностью области изменения одной случайной величины на основании наблюдаемых значений другой случайной величины. На практике при исследовании зависимости между случайными величинами X и Y часто ограничиваются исследованием зависимости между X и условным математическим ожиданием $M(Y/x) = f(x)$. Функция $M(Y/x) = f(x)$ называется *функцией регрессии первого рода* или модельной функцией регрессии Y на X, а график ее – *линией регрессии.*

Линейной регрессией называется сведение наблюдаемой на опыте зависимости некоторой переменной (*зависимой* или *объясняемой*) от одной или более других переменных (*независимых* или *объясняющих*) к линейной зависимости (в предположении, что строгая линейная зависимость между ними нарушается случайными ошибками). Для проведения линейной регрессии часто используется *метод наименьших квадратов*.

В простейшем случае речь идет о двух переменных. Пусть x – независимая переменная, y – зависимая и между ними существует следующая связь: $y_i = a + bx_i + \varepsilon_i$, где a и b – числовые коэффициенты, ε_i – случайные ошибки. Задача состоит в том, чтобы по имеющимся наблюдениям $(x_1, y_1); (x_2, y_2), ..., (x_n, y_n)$ построить оценки для *a* и *b*. Согласно методу наименьших квадратов необходимо решить следующую математическую задачу:

$$S=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2\to\min. \tag{4.1}$$

Решаем задачу, вычисляя частные производные суммы квадратов по каждому из коэффициентов и приравнивая эти производные к нулю. Получаем систему *нормальных уравнений*, которая позволяет получить оценки параметров *a* и *b*:

$$\begin{cases}2\sum_{i=1}^{n}\left(y_i-a-bx_i\right)=0\\ 2\sum_{i=1}^{n}\left(y_i-a-bx_i\right)x_i=0\end{cases}\sim\begin{cases}n\cdot a+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i\\ a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_iy_i\end{cases}$$

Из последней системы следуют формулы для определения параметров уравнения парной линейной регрессии Y на X:

$$b=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{x^2}-\overline{x}^2};\;\; a=\overline{y}-b\overline{x}, \tag{4.2}$$

где $\overline{xy}=\frac{1}{n}\sum_{i=1}^{n}x_iy_i$; $\overline{x^2}=\frac{1}{n}\sum_{i=1}^{n}x_i^2$.

Уравнение вида $\overline{y}_x=a+b\overline{x}$ называется *уравнением линейной регрессии*, а получаемые из него значения $\overline{y}_i=a+bx_i$ называются *предсказанными* значениями, в отличие от *наблюдаемых* значений y_i. Важным и практически значимым результатом линейной регрессии является то, что она позволяет «предсказывать» значения зависимой переменной даже для таких значений независимых, которые реально не наблюдались.

На практике часто бывает важно знать, существует ли зависимость между некоторыми наблюдаемыми величинами, и если да, то какая она – сильная или слабая, положительная или отрицательная.

Для выяснения этих обстоятельств используется *корреляционный анализ*.

Корреляционный анализ позволяет на основе выборочных данных оценить наличие, силу и направленность статистической взаимосвязи.

По направлению связи бывают *прямыми*, когда зависимая переменная растет с увеличением факторного признака (X) и *обратными*, при которых рост последнего сопровождается уменьшением зависимой переменной. Эти связи также можно назвать положительными и отрицательными.

Пусть для двух показателей X и Y (случайных величин) имеется выборка связанных пар наблюдений $\{(x_1, y_1); (x_2, y_2); ...; (x_n, y_n)\}$, где n – число наблюдений.

Выборочной ковариацией (*корреляционным моментом*) называется величина

$$K_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}), \tag{4.3}$$

где $\overline{x}$, $\overline{y}$ – средние значения выборочных данных для величин X и Y соответственно.

Ковариация является мерой взаимосвязи случайных величин. Если ковариация равна нулю, то взаимосвязь величин отсутствует. Если $K_{xy} > 0,$ то существует прямая взаимосвязь, а если $K_{xy} < 0,$ то обратная взаимосвязь случайных величин. Но эта характеристика обладает рядом существенных недостатков. Во-первых, ее значение зависит от единиц измерения исследуемых случайных величин. Во-вторых, она не позволяет оценить силу зависимости между ними. Для устранения данных недостатков вводится относительная мера взаимосвязи (безразмерная величина) – коэффициент корреляции. Существует несколько типов коэффициентов корреляции. Рассмотрим только один – коэффициент линейной корреляции (Пирсона), который характеризует степень линейной взаимосвязи двух случайных величин.

Выборочным коэффициентом корреляции для выборки вида $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ называется выборочная характеристика

$$r_в = \frac{K_{xy}}{s_x \cdot s_y}, \qquad (4.4)$$

где K_{xy} – оценка корреляционного момента (ковариация);

s_x и s_y – «исправленные» средние квадратические отклонения.

Отметим основные свойства коэффициента корреляции.

1. Выборочный коэффициент корреляции может принимать значения от – 1 до + 1. В зависимости от того, насколько $|r_в|$ приближается к 1, различают связь слабую, умеренную, заметную, достаточно тесную, тесную и весьма тесную, т. е. чем ближе $|r_в|$ к 1, тем теснее связь.

2. Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции не изменится.

3. При $r_в = \pm 1$ корреляционная связь представляет линейную функциональную зависимость.

4. При $r_в = 0$ линейная корреляционная связь отсутствует.

Таким образом, по абсолютной величине и знаку коэффициента можно судить о степени зависимости (сильной или слабой) и о ее характере (положительной или отрицательной).

Степень линейной зависимости можно качественно оценить с помощью шкалы Чеддока (табл. 4.1).

Таблица 4.1

Значение коэффициента корреляции	Теснота связи
0,1–0,3	Слабая
0,3–0,5	Умеренная
0,5–0,7	Заметная
0,7–0,9	Высокая
0,9–0,99	Весьма высокая

При большом числе наблюдений одно и то же значение x может встретиться n_x раз, одно и то же значение y может встретиться n_y раз, одна и та же пара чисел (x, y) может наблюдаться n_{xy} раз. Поэтому данные наблюдений группируют, т. е. подсчитывают частоты n_x, n_y, n_{xy}. Все сгруппированные данные записывают в виде таблицы, которую называют *корреляционной* (табл. 4.2).

Таблица 4.2

X	Y						n_x
	y_1	y_2	…	y_j	…	y_m	
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1m}	n_{x1}
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2m}	n_{x2}
…	…	…	…	…	…	…	…
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{im}	n_{xi}
…	…	…	…	…	…	…	…
x_k	n_{k1}	n_{k2}	…	n_{kj}	…	n_{km}	n_{xk}
n_y	n_{y1}	n_{y2}	…	n_{yj}	…	n_{ym}	n

Здесь $x_1, x_2, ..., x_k$; $y_1, y_2, ..., y_m$ – значения случайных величин X и Y соответственно, а n_{x1}, $n_{x2}, ..., n_{xk}$; $n_{y1}, n_{y2}, ..., n_{ym}$ – соответствующие частоты, n_{ij} – частота, с которой встречается пара $\left(x_i, y_j\right)$; $n = \sum_{i=1}^{k}\sum_{j=1}^{n} n_{ij}$.

Для *коэффициента корреляции* случайных величин X и Y в случае *сгруппированных* данных используется выражение

$$r_{\text{в}} = \frac{\sum n_{xy} xy - n \cdot \overline{x} \cdot \overline{y}}{n \cdot \sigma_x \cdot \sigma_y}. \tag{4.5}$$

После подсчета $\overline{x}, \overline{y}, \sigma_x, \sigma_y$ и $r_{\text{в}}$ получают выборочное уравнение линейной регрессии Y на X в виде

$$\overline{y}_x - \overline{y} = r_{\text{в}} \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right). \tag{4.6}$$

Величину $D_{\text{ост}} = s_y^2\left(1 - r_B^2\right)$ называют *остаточной дисперсией* случайной величины Y относительно случайной величины X; она характеризует величину ошибки, которая возникает при замене Y линейной функцией.

При $r_B = \pm 1$ остаточная дисперсия равна нулю, т. е. при представлении Y в виде линейной функции от X не возникает ошибки, а Y и X связаны *линейной функциональной* зависимостью.

Выборочный коэффициент корреляции обычно используется в предположении нормальности данных. В этом случае из равенства нулю теоретического коэффициента R_{xy} следует независимость случайных величин (в более общем случае это неверно). В случае нормального распределения можно проверить гипотезу $R_{xy} = 0$. Пусть

$$T = r_{\text{в}} \cdot \sqrt{\frac{n-2}{1-r_{\text{в}}^2}}. \qquad (4.7)$$

Если гипотеза $R_{xy} = 0$ верна, то T имеет распределение Стьюдента с $k = n - 2$ степенями свободы. При уровне значимости α выберем критическую точку $t_{kp} = t_{kp}(\alpha, k)$ для двусторонней области (см. прил. 6). Если $|T| < t_{kp}$, то гипотеза $R_{xy} = 0$ принимается, иначе – отвергается.

Наличие корреляции примерно может быть определено с помощью корреляционного поля. Его получим, если нанесем на график в определенном масштабе точки, соответствующие наблюдаемым одновременно значениям двух величин $\left(x_i; y_j\right)$.

4.2. Примеры решения задач

1. При исследовании зависимости между случайными величинами X и Y была получена таблица измерений соответствующих значений этих величин.

X	–2	–1	0	1	2	3
Y	–0,4	0,2	0,7	1,6	2	3,5

Требуется:

1) методом наименьших квадратов аппроксимировать статистическую зависимость между этими величинами линейной функцией;

2) вычислить оценку корреляционного момента и коэффициент корреляции, а также остаточную дисперсию;

3) построить корреляционное поле и линию регрессии на корреляционном поле.

Решение.

1) Найдем методом наименьших квадратов эмпирическую формулу вида $\overline{y}_x = a + bx.$ Для вычисления коэффициентов a и b составим табл. 4.3.

Таблица 4.3

№	1	2	3	4	5	6	Сумма
x_i	–2	–1	0	1	2	3	3
y_i	–0,4	0,2	0,7	1,6	2	3,5	7,6
$x_i y_i$	0,8	–0,2	0	1,6	4	10,5	16,7
x_i^2	4	1	0	1	4	9	19

Тогда $\overline{x} = \dfrac{3}{6} = \dfrac{1}{2};$ $\overline{y} = \dfrac{7,6}{6} = \dfrac{38/5}{6} = \dfrac{19}{15};$ $\overline{x^2} = \dfrac{19}{6};$ $\overline{xy} = \dfrac{16,7}{6} = \dfrac{167}{60}.$

Согласно формуле (4.2)

$$b = \frac{\dfrac{167}{60} - \dfrac{1}{2}\cdot\dfrac{19}{15}}{\dfrac{19}{6} - \dfrac{1}{4}} = \frac{2,15}{2,9167} \approx 0,737; a = \frac{19}{15} - 0,737\cdot\frac{1}{2} \approx 0,898.$$

Следовательно, модель имеет вид $\overline{y}_x = 0,898 + 0,737x.$

Проверим правильность линейной модели. Для этого вычислим значения аппроксимирующей функции $\overline{y}_x = 0,898 + 0,737x$ и внесем полученные значения в табл. 4.4.

Таблица 4.4

№	1	2	3	4	5	6
x_i	–2	–1	0	1	2	3
y_i	–0,4	0,2	0,7	1,6	2	3,5
$\overline{y}_i$	–0,576	0,161	0,898	1,635	2,372	3,109
$\varepsilon_i = \overline{y}_i - y_i$	–0,176	–0,039	0,198	0,035	0,372	–0,391

Из таблицы видно, что значения аппроксимирующей функции приблизительно совпадают с y_i для всех точек x_i. Следовательно, исследуемая функциональная зависимость может быть приближенно описана линейной моделью $\overline{y}_x = 0,898 + 0,737x$.

Определим меру отклонения S:

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = (-0,176)^2 + (-0,039)^2 + 0,198^2 + 0,035^2 +$$

$$+ 0,372^2 + (-0,391)^2 = 0,3642.$$

Вычисленное значение S небольшое $(\to \min)$, что еще раз подтверждает правильность выбора модели.

2) Составим следующую вспомогательную табл. 4.5 для нахождения оценки корреляционного момента и выборочного коэффициента корреляции.

Таблица 4.5

x_i	y_i	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x}) \cdot (y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
–2	–0,4	–2,5	–1,6667	4,16675	6,25	2,7779
–1	0,2	–1,5	–1,067	1,6001	2,25	1,1378
0	0,7	–0,5	–0,567	0,2834	0,25	0,3211
1	1,6	0,5	0,3333	0,1667	0,25	0,1111
2	2	1,5	0,7333	1,1	2,25	0,5377
3	3,5	2,5	2,2333	5,5833	6,25	4,9876
$\sum$	7,6			12,9	17,5	9,8733

Тогда

$$K_{xy} = \frac{1}{5} \cdot 12{,}9 = 2{,}58; \; s_x = \sqrt{\frac{17{,}5}{5}} \approx 1{,}871,$$

$$s_y = \sqrt{\frac{9{,}8732}{5}} = \sqrt{1{,}97464} \approx 1{,}405.$$

Согласно формуле (4.4)

$$r_{\text{в}} = \frac{2{,}58}{1{,}871 \cdot 1{,}405} = \frac{2{,}58}{2{,}628755} \approx 0{,}98.$$

Остаточная дисперсия

$$D_{\text{ост}} = s_y^2 \left(1 - r_B^2\right) = 1{,}97464 \cdot \left(1 - 0{,}98^2\right) \approx 0{,}078.$$

3) Корреляционное поле и линия регрессии на корреляционном поле изображены на рис. 4.1 (построено в пакете Mathcad).

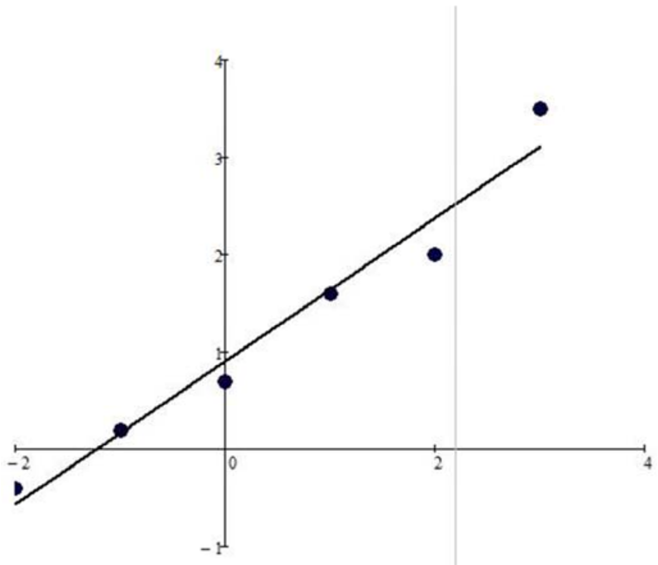

Рис. 4.1. Корреляционное поле и линия регрессии на корреляционном поле

2. Таблица значений признака Y при данных значениях признака X имеет вид (табл. 4.6).

Таблица 4.6

X	Y				n_x
	5	10	15	20	
8	7				7
12	10	12	1		23
16		5	28	4	37
20		5	5	9	19
24				14	14
n_y	17	22	34	27	$n=100$

Построить корреляционное поле. Найти выборочный коэффициент корреляции, оценить его значимость, приняв $\alpha = 0,05$. Записать уравнения прямой линии регрессии Y на X.

Решение. Корреляционное поле данной двумерной выборки (построенное в пакете MS Excel) приведено на рис. 4.2.

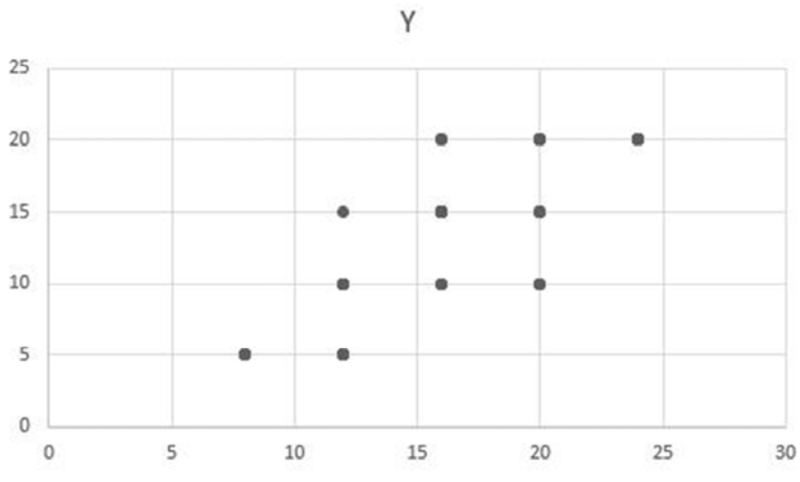

Рис. 4.2. Корреляционное поле исходной двумерной выборки

По виду поля корреляции можно судить о том, что между величинами существует зависимость.

Для вычисления выборочных числовых характеристик составляем расчетную табл. 4.7.

Таблица 4.7

X	Y				n_x	$n_x \cdot x_i$	$n_x x_i^2$
	5	10	15	20			
8	7				7	56	448
12	10	12	1		23	276	3312
16		5	28	4	37	592	9472
20		5	5	9	19	380	7600
24				14	14	336	8064
n_y	17	22	34	27	$n = 100$	1640	28896
$n_y \cdot y_j$	85	220	510	540	1355	Сумма	
$n_y y_j^2$	425	2200	7650	10800	21075		

Находим средние значения

$$\overline{x} = \frac{1640}{100} = 16,4, \quad \overline{y} = \frac{1355}{100} = 13,55, \quad \overline{x^2} = \frac{28896}{100} = 288,96,$$

$$\overline{y^2} = \frac{21075}{100} = 210,75,$$

$$\sigma_x = \sqrt{\overline{x^2} - \overline{x}^2} = \sqrt{288,96 - 16,4^2} = \sqrt{20} \approx 4,472,$$

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{210,75 - 13,55^2} = \sqrt{27,1475} \approx 5,2103.$$

$$\sum n_{xy} xy = 5 \cdot 8 \cdot 7 + 5 \cdot 12 \cdot 10 + 10 \cdot 10 \cdot 12 + 10 \cdot 16 \cdot 5 + 10 \cdot 20 \cdot 5 +$$
$$+ 15 \cdot 1 \cdot 12 + 15 \cdot 16 \cdot 28 + 15 \cdot 5 \cdot 20 + 20 \cdot 4 \cdot 16 +$$
$$+ 20 \cdot 20 \cdot 9 + 20 \cdot 14 \cdot 24 = 24120.$$

Находим коэффициент корреляции по формуле (4.5):

$$r_{\text{в}} = \frac{24120 - 100 \cdot 16,4 \cdot 13,55}{100 \cdot 4,472 \cdot 5,2103} \approx 0,815.$$

Для оценки значимости выборочного коэффициента корреляции вычислим наблюдаемое значение критерия, воспользовавшись формулой (4.7):

$$T = 0,815 \cdot \sqrt{\frac{98}{1 - 0,815^2}} \approx 24,03.$$

Затем по таблице критических точек распределения Стьюдента, по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 100 - 2 = 98$ найдем критическую точку t_{kp} для двухсторонней критической области:

$$t_{kp}\left(0,05;98\right) = 1,98.$$

Сравнивая T и t_{kp}, получим, что $\left|T\right| > t_{kp}$, следовательно, величины X, Y коррелированы.

Оценкой теоретической линии регрессии является эмпирическая линия регрессии, уравнение которой имеет вид $\overline{y}_x - \overline{y} = b\left(x - \overline{x}\right)$, где $b = r_{\text{в}} \dfrac{\sigma_y}{\sigma_x}$. Тогда

$$\overline{y}_x - 13,55 = 0,815 \cdot \frac{5,2103}{4,472} \cdot \left(x - 16,4\right) \text{ или } \overline{y}_x = 0,95x - 2,08.$$

4.3. Задачи для самостоятельного решения

1. При исследовании зависимости между случайными величинами X и Y была получена таблица измерений соответствующих значений этих величин.

Требуется:

а) методом наименьших квадратов аппроксимировать статистическую зависимость между этими величинами линейной функцией;

б) вычислить оценку корреляционного момента и коэффициент корреляции, а также остаточную дисперсию;

в) построить корреляционное поле и линию регрессии на корреляционном поле.

1)

X	1	2	3	4	5
Y	5,2	4,2	3	0,7	1,2

2)

X	0	0,5	1	1,5	2
Y	3	6	8,5	13	15

3)

X	1,6	3,8	0,7	3,4	1,6
Y	–2,2	–1,6	–2,5	–0,9	–2,2

4)

X	9,1	0,2	5,9	5,4	9,9
Y	2	5,9	0,1	0	1,6

5)

X	0	5	10	15	20	25
Y	9	7	6,3	5,1	3,9	2,1

6)

X	0	1	2	3	4	5	6	7	8	9
Y	–4	–3,2	–2	–1	–0,7	0	0,5	1	1,6	2,1

7)

X	0	0,4	0,8	1,2	1,6	2	2,4	2,8	3,2	3,6
Y	5	7,5	11,4	14,5	17,2	19,9	22	24	26,1	28

2. Для исследования зависимости случайных величин X и Y получены статистические данные, представленные в корреляционной таблице.

Требуется:

а) построить корреляционное поле, сделать вывод;

б) определить выборочный коэффициент корреляции;

в) при уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о равенстве генерального коэффициента корреляции нулю при конкурирующей гипотезе $H_1 : R_{xy} \neq 0$;

г) найти уравнение прямой регрессии Y на X;

д) построить линию регрессии на корреляционном поле.

1)

X	Y				
	3	6,5	10	13,5	17
0	15	5			
0,5	7	10	3		
1		4	8	9	2
1,5				6	1

2)

X	Y			
	0,5	0,6	0,7	0,8
0,5		2		8
0,6		4	2	9
0,7	2	12	3	1
0,8	21	14		
0,9	2			

3)

X	Y				
	7	10	13	16	19
10	5				
20	10	11	8		
30	2	2	3		
40		16	9	4	
50			3	11	2
60				8	6

Ответы

1	1) $\overline{y}_x = -1{,}15x + 6{,}31$ $s_{xy} = -2{,}875;\ r_в = -0{,}948;$ $D_{ост} = 0{,}37.$	2) $\overline{y}_x = 6{,}2x + 2{,}9$ $s_{xy} = 3{,}875;\ r_в = 0{,}994;$ $D_{ост} = 0{,}3.$
	3) $\overline{y}_x = 0{,}611x - 3{,}045$ $s_{xy} = 1{,}445;\ r_в = 0{,}978;$ $D_{ост} = 0{,}04.$	4) $\overline{y}_x = 0{,}486x - 2{,}786$ $s_{xy} = 7{,}145;\ r_в = 0{,}988;$ $D_{ост} = 0{,}085.$
	5) $\overline{y}_x = -0{,}257x + 8{,}781$ $s_{xy} = -22{,}5;\ r_в = -0{,}992;$ $D_{ост} = 0{,}094.$	6) $\overline{y}_x = 0{,}656x - 3{,}535$ $s_{xy} = 6{,}04;\ r_в = 0{,}988;$ $D_{ост} = 0{,}097$
	7) $\overline{y}_x = 6{,}453x + 5{,}995$ $s_{xy} = 9{,}464;\ r_в = 0{,}994;$ $D_{ост} = 0{,}74$	
2	1) $r_в = 0{,}823;\ \ \overline{y}_x = 7{,}348x + 3{,}334$ 2) $r_в = -0{,}791;\ \ \overline{y}_x = -0{,}795x + 1{,}18$ 3) $r_в = 0{,}792;\ \ \overline{y}_x = 0{,}191x + 5{,}339$	

Проверочный тест 4

Условие задачи	Варианты ответов
1. Связь считается сильной, если значение выборочного коэффициента корреляции…	1) Равно 0 2) В диапазоне от 0 до 0,3 3) В диапазоне от 0,7 до 1 4) В диапазоне от 1 до 2 5) Принимает положительные значения
2. Если одному значению первого признака соответствует несколько значений второго – это связь…	1) Функциональная 2) Положительная 3) Регрессионная 4) Прямолинейная 5) Корреляционная

Условие задачи	Варианты ответов
3. Если ковариация (K_{xy}) больше нуля, то…	1) Взаимосвязь величин отсутствует 2) Существует прямая взаимосвязь 3) Существует обратная взаимосвязь
4. Коэффициент линейной корреляции может принимать значения…	1) Близкое к нулю 2) Измеряется в процентах 3) Близкое к единице(+1 или –1) 4) От –1 до +1
5. Для изображения корреляционной зависимости используется график:	1) Линейный 2) Радиальный 3) График рассеяния точек 4) Динамический
6. Коэффициент корреляции между уровнем шума и снижением слуха с учетом стажа у рабочих механосборочного цеха равен $r_в =$ $= +0{,}91$. Установленная связь…	1) Прямая и сильная 2) Обратная и слабая 3) Обратная и сильная 4) Прямая и слабая
7. Выборочное уравнение прямой линии регрессии Y на X имеет вид $\overline{y}_x = 1{,}4 - 1{,}8x$, а средние квадратические отклонения равны $\sigma_x =$ $= 0{,}12; \sigma_y = 0{,}54$. Тогда коэффициент корреляции равен….	1) –0,6 2) –0,4 3) –0,02 4) 0,4 5) Другой ответ
8. По результатам выборки, извлеченной из генеральной совокупности (X, Y), вычислены выборочный коэффициент регрессии Y на X (b = 2,5) и выборочные средние $\overline{x} = 30$; $\overline{y} = 41$. Тогда выборочное уравнение прямой линии регрессии Y на X имеет вид …	1) $\overline{y}_x = 2{,}5x + 116$ 2) $\overline{y}_x = 2{,}5x - 75$ 3) $\overline{y}_x = 2{,}5x - 116$ 4) $\overline{y}_x = 2{,}5x + 132$ 5) Другой ответ

<table>
<tr><th>Условие задачи</th><th>Варианты ответов</th></tr>
<tr><td>9. Зависимость средней выборки одного рабочего за смену Y (штук) от квалификации X (разряды) приведена в таблице:

<table>
<tr><td>X</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Y</td><td>12</td><td>19</td><td>23</td><td>30</td></tr>
</table>

Уравнение регрессии $\overline{y}_x = a + bx$. Тогда коэффициент b равен…</td>
<td>1) 2,8
2) 5,3
3) 5,8
4) 7,2
5) Другой ответ</td></tr>
</table>

Ответы к проверочному тесту 4

№	1	2	3	4	5	6	7	8	9
	3	5	2	4	3	1	2	1	3

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Латышева, И. Г. Конспект лекций по математике для студентов инженерно-технических специальностей. В четырех частях. Часть 3. – И. Г. Латышева, И. В. Прусова, Л. А. Барминова, О. Г. Вишневская, Е. А. Глинская, Н. А. Кондратьева, А. Н. Мелешко. – Электрон. дан. – БНТУ, 2007. – Режим доступа: http://rep.bntu.by/handle/data/1214.

2. Лисьев, В. П. Теория вероятностей и математическая статистика: учебное пособие / Московский государственный университет экономики, статистики и информатики. – М., 2006. – 199 с.

4. Маталыцкий, М. А. Теория вероятностей и математическая статистика: пособие / М. А. Маталыцкий, Т. В. Русилко. – 2-е изд., перераб. и доп. // Гродно: ГрГУ, 2009. – 219 с.

3. Савич, Л. К. Теория вероятностей и математическая статистика: учебное пособие для экономических специальностей учреждений, обеспечивающих получение высш. образования / Л. К. Савич, Н. А. Смольская; науч. ред. О. И. Лаврова // Мн.: Адукацыя і выхаванне, 2009. – 208 с.: ил.

4. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман – 9-е изд., стер. – М: Высш. шк., 2003. – 479 с., ил.

5. Математическая статистика в примерах и задачах: учебное пособие / Г. С. Евдокимова, Смол. гос. ун-т. – Смоленск: Изд-во СмолГУ, 2014. – 98 с. – Режим доступа https://studfile.net/preview/3544158/.

6. Высшая математика. Руководство к решению задач. Ч. 2 / К. Н. Лунгу, Е. В. Макаров // М.:ФИЗМАТЛИТ, 2007. – 384 с.

7. Аверьянова, С. Ю. Лабораторный практикум по математической статистике в среде ЭТ MS Excel: учебное пособие / С. Ю. Аверьянова, Н. В. Растеряев // Южный федеральный университет. – Ростов-на-дону: Издательство Южного федерального университета, 2014. – 64 с.

8. Рябушко, А. Индивидуальные задания по высшей математике. В 4 частях. Часть 4. Операционное исчисление. Элементы теории устойчивости. Теория вероятностей. Математическая статистика / А. Рябушко // Мн.: Вышэйшая школа, 2013. – 336 с.

8. Дубровина, О. В. Прикладная математика: метод. пособие по выполнению практических и лабораторных работ для студентов заочного отделения специальности 1-54 01 01 «Метрология, стан-

дартизация и сертификация» / О. В. Дубровина, Н. К. Прихач, В. М. Романчак // Мн.: БНТУ, 2009. – 70 с.

9. Еськова, О. И. Основы статистической обработки информации: пособие / О. И. Еськова, Л. П. Авдашкова, М. А. Грибовская. – Минск: Беларусь, 2011. – 175 с.: ил.

10. Кондратьева, Н. А. Математика. Специальные разделы. Элементы теории функций комплексной переменной и операционного исчисления. Теория вероятностей. Элементы математической статистики / Н. А. Кондратьева, Н. К. Прихач, Н. Н. Буснюк., А. Н. Мелешко. – Электрон. издан. – БНТУ, 2014. – Режим доступа: http://rep.bntu.by/handle/data/9383

Таблица значений $t_\gamma = t(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	2,78	4,6	8,61	20	2,093	2,861	3,833
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,5	5,41	35	2,032	2,72	3,6
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,2	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,64	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,98	2,617	3,374
18	2,11	2,9	3,97	∞	1,96	2,576	3,291
19	2,1	2,88	3,92				

ПРИЛОЖЕНИЕ 2

Таблица значений $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	18	0,37	0,58	0,88
6	1,09	2,01	3,88	19	0,32	0,49	0,73
7	0,92	1,62	2,98	20	0,28	0,43	0,63
8	0,8	1,38	2,42	25	0,26	0,38	0,56
9	0,71	1,2	2,06	30	0,24	0,35	0,5
10	0,65	1,08	1,8	35	0,22	0,32	0,46
11	0,59	0,98	1,6	40	0,21	0,3	0,43
12	0,55	0,9	1,45	45	0,188	0,269	0,38
13	0,52	0,83	1,33	50	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,7	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,16	0,211
18	0,4	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,6	0,92	250	0,089	0,12	0,162

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
0,0	0,3989	0,35	0,3752	0,68	0,3166	1,2	0,1942
0,03	0,3988	0,36	0,3739	0,7	0,3123	1,3	0,1714
0,04	0,3986	0,37	0,3725	0,71	0,3101	1,4	0,1497
0,05	0,3984	0,38	0,3712	0,72	0,3079	1,5	0,1295
0,06	0,3982	0,4	0,3683	0,73	0,3056	1,6	0,1109
0,07	0,3980	0,41	0,3668	0,74	0,3034	1,7	0,0940
0,08	0,3977	0,42	0,3653	0,75	0,3011	1,8	0,0790
0,1	0,3970	0,43	0,3637	0,76	0,2989	1,9	0,0656
0,11	0,3965	0,44	0,3621	0,77	0,2966	2	0,0540
0,12	0,3961	0,45	0,3605	0,78	0,2943	2,1	0,0440
0,13	0,3956	0,46	0,3589	0,8	0,2897	2,15	0,0431
0,14	0,3951	0,47	0,3572	0,81	,2874	2,2	0,0355
0,15	0,3945	0,48	0,3555	0,82	0,2850	2,25	0,0347
0,16	0,3939	0,5	0,3521	0,83	0,2827	2,3	0,0283
0,17	0,3932	0,51	0,3503	0,84	0,2803	2,4	0,0224
0,18	0,3925	0,52	0,3485	0,85	0,2780	2,5	0,0175
0,2	0,3910	0,53	0,3467	0,86	0,2756	2,6	0,0136
0,21	0,3902	0,54	0,3448	0,87	0,2732	2,7	0,0104
0,22	0,3894	0,55	0,3429	0,88	0,2709	2,8	0,0079
0,23	0,3885	0,56	0,3410	0,9	0,2661	2,9	0,0060
0,24	0,3876	0,57	0,3391	0,91	0,2637	3	0,0044
0,25	0,3867	0,58	0,3372	0,92	0,2613	3,1	0,0033
0,26	0,3857	0,6	0,3332	0,93	0,2589	3,2	0,0024
0,27	0,3847	0,61	0,3312	0,94	0,2565	3,3	0,0017
0,28	0,3836	0,62	0,3292	0,95	0,2541	3,4	0,0012
0,3	0,3814	0,63	0,3271	0,96	0,2516	3,5	0,0009
0,31	0,3802	0,64	0,3251	0,97	0,2492	3,6	0,0006
0,32	0,3790	0,65	0,3230	0,98	0,2468	3,7	0,0004
0,33	0,3778	0,66	0,3209	1	0,2420	3,8	0,0003
0,34	0,3765	0,67	0,3187	1,1	0,2179	3,9	0,0002

ПРИЛОЖЕНИЕ 4

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

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,31	0,2117	0,62	0,2324	0,93	0,3238
0,01	0,0040	0,32	0,1255	0,63	0,2357	0,94	0,3264
0,02	0,0080	0,33	0,1293	0,64	0,2389	0,95	0,3289
0,03	0,0120	0,34	0,1331	0,65	0,2422	0,96	0,3315
0,04	0,0160	0,35	0,1368	0,66	0,2454	0,97	0,3340
0,05	0,0199	0,36	0,1406	0,67	0,2486	0,98	0,3365
0,06	0,0239	0,37	0,1443	0,68	0,2517	0,99	0,3389
0,07	0,0279	0,38	0,1480	0,69	0,2549	1,00	0,3413
0,08	0,0319	0,39	0,1517	0,70	0,2589	1,01	0,3438
0,09	0,0359	0,40	0,1554	0,71	0,2611	1,02	0,3461
0,10	0,0398	0,41	0,1591	0,72	0,2642	1,03	0,3485
0,11	0,0438	0,42	0,1628	0,73	0,2673	1,04	0,3508
0,12	0,0478	0,43	0,1664	0,74	0,2703	1,05	0,3531
0,13	0,0517	0,44	0,1700	0,75	0,2734	1,06	0,3554
0,14	0,0557	0,45	0,1736	0,76	0,2764	1,07	0,3577
0,15	0,0596	0,46	0,1772	0,77	0,2794	1,08	0,3599
0,16	0,0636	0,47	0,1808	0,78	0,2823	1,09	0,3621
0,17	0,0675	0,48	0,1844	0,79	0,2852	1,10	0,3643
0,18	0,0714	0,49	0,1879	0,80	0,2881	1,11	0,3665
0,19	0,0753	0,50	0,1915	0,81	0,2910	1,12	0,3686
0,20	0,0793	0,51	0,1950	0,82	0,2939	1,13	0,3708
0,21	0,0832	0,52	0,1985	0,83	0,2967	1,14	0,3729
0,22	0,0871	0,53	0,2019	0,84	0,2995	1,15	0,3749
0,23	0,0910	0,54	0,2054	0,85	0,3023	1,16	0,3770
0,24	0,0948	0,55	0,2088	0,86	0,3051	1,17	0,3790
0,25	0,0987	0,56	0,2123	0,87	0,3078	1,18	0,3810
0,26	0,1026	0,57	0,2157	0,88	0,3106	1,19	0,3830
0,27	0,1064	0,58	0,2190	0,89	0,3133	1,20	0,3849
0,28	0,1103	0,59	0,2224	0,90	0,3159	1,21	0,3869
0,29	0,1141	0,60	0,2257	0,91	0,3186	1,22	0,3883

Окончание прил. 4

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,30	0,1179	0,61	0,2291	0,92	0,3212	1,23	0,3907
1,24	0,3925	1,57	0,4418	1,90	0,4713	2,45	0,4929
1,25	0,3944	1,58	0,4429	1,91	0,4719	2,46	0,4931
1,26	0,3962	1,59	0,4441	1,92	0,4726	2,48	0,4934
1,27	0,3980	1,60	0,4452	1,93	0,4732	2,50	0,4938
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,52	0,4941
1,29	0,4015	1,62	0,4474	1,95	0,4744	2,54	0,4945
1,30	0,4032	1,63	0,4484	1,96	0,4750	2,55	0,4946
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,56	0,4948
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,60	0,4453
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,65	0,4960
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,70	0,4965
1,35	0,4115	1,68	0,4535	2,02	0,4783	2,75	0,4970
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,80	0,4974
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,85	0,4978
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,90	0,4981
1,39	0,4177	1,72	0,4573	2,10	0,4821	2,95	0,4984
1,40	0,4192	1,73	0,4582	2,12	0,4830	3,00	0,4987
1,41	0,4207	1,74	0,4591	2,14	0,4838	3,10	0,4990
1,42	0,4222	1,75	0,4599	2,16	0,4846	3,20	0,4993
1,43	0,4236	1,76	0,4608	2,18	0,4854	3,30	0,4995
1,44	0,4251	1,77	0,4616	2,20	0,4861	3,40	0,4997
1,45	0,4265	1,78	0,4625	2,22	0,4868	3,50	0,4998
1,46	0,4279	1,79	0,4633	2,24	0,4875	3,60	0,4999
1,47	0,4292	1,80	0,4641	2,26	0,4881	3,70	0,4999
1,48	0,4306	1,81	0,4649	2,28	0,4887	3,80	0,4999
1,49	0,4319	1,82	0,4656	2,30	0,4893	3,90	0,49995
1,50	0,4332	1,83	0,4664	2,32	0,4898	4,00	0,49997
1,51	0,4345	1,84	0,4671	2,34	0,4904	4,10	0,49998
1,52	0,4357	1,85	0,4678	2,36	0,4909	4,20	0,49999
1,53	0,4370	1,86	0,4686	2,38	0,4913	4,30	0,49999
1,54	0,4382	1,87	0,4693	2,40	0,4918	4,40	0,49999
1,55	0,4394	1,88	0,4699	2,42	0,4922	4,50	0,499997
1,56	0,4406	1,89	0,4706	2,44	0,4927	5,00	0,4999997

ПРИЛОЖЕНИЕ 5

Критические точки распределения χ^2

k	α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,64	5,02	3,84	0,0039	0,00098	0,00016
2	9,21	7,38	5,99	0,103	0,051	0,020
3	11,35	9,35	7,82	0,352	0,216	0,115
4	13,28	11,14	9,49	0,711	0,484	0,297
5	15,09	12,83	11,07	1,145	0,831	0,554
6	16,81	14,45	12,59	1,635	1,237	0,872
7	18,48	16,01	14,07	2,167	1,690	1,24
8	20,09	17,54	15,51	2,733	2,180	1,65
9	21,67	19,02	16,92	3,325	2,700	2,09
10	23,21	20,48	18,31	3,940	3,247	2,56
11	24,73	21,92	19,68	4,575	3,816	3,05
12	26,22	23,34	21,03	5,226	4,404	3,57
13	27,69	24,74	22,36	5,892	5,009	4,11
14	29,14	26,12	23,69	6,571	5,629	4,66
15	30,58	27,49	24,996	7,261	6,262	5,23
16	32,00	28,85	26,296	7,962	6,908	5,81
17	33,41	30,19	27,59	8,672	7,564	6,41
18	34,81	31,53	28,87	9,390	8,231	7,01
19	36,19	32,85	30,14	10,117	8,907	7,63
20	37,57	34,17	31,41	10,851	9,591	8,26
21	38,93	35,48	32,67	11,591	10,283	8,90
22	40,29	36,78	33,92	12,338	10,982	9,54
23	41,64	38,08	35,17	13,091	11,689	10,2
24	42,98	39,36	36,42	13,848	12,401	10,9
25	44,31	40,65	37,65	14,611	13,12	11,5
26	45,64	41,92	38,89	17,292	15,379	12,2
27	46,96	43,20	40,11	18,114	16,151	12,9
28	48,28	44,46	41,34	18,939	16,928	13,6
29	49,59	45,72	42,56	19,768	17,708	14,3
30	50,89	46,98	43,77	20,599	18,493	15,0

ПРИЛОЖЕНИЕ 6

Критические точки распределения Стьюдента

k	α (односторонняя критическая область)					
	0,001	0,005	0,01	0,025	0,05	0,1
	α (двусторонняя критическая область)					
	0,002	0,01	0,02	0,05	0,1	0,2
1	318,309	63,657	31,821	12,706	6,314	3,078
2	22,327	9,925	6,965	4,303	2,920	1,886
3	10,215	5,841	4,541	3,182	2,353	1,638
4	7,173	4,604	3,747	2,776	2,132	1,533
5	5,893	4,032	3,365	2,571	2,015	1,476
6	5,208	3,707	3,143	2,447	1,943	1,440
7	4,785	3,499	2,998	2,365	1,895	1,415
8	4,501	3,355	2,896	2,306	1,860	1,397
9	4,297	3,250	2,821	2,262	1,833	1,383
10	4,144	3,169	2,764	2,228	1,812	1,372
11	4,025	3,106	2,718	2,201	1,796	1,363
12	3,930	3,055	2,681	2,179	1,782	1,356
13	3,852	3,012	2,650	2,160	1,771	1,350
14	3,787	2,977	2,624	2,145	1,761	1,345
15	3,733	2,947	2,602	2,131	1,753	1,341
16	3,686	2,921	2,583	2,120	1,746	1,337
17	3,646	2,898	2,567	2,110	1,740	1,333
18	3,610	2,878	2,552	2,101	1,734	1,330
19	3,579	2,861	2,539	2,093	1,729	1,328
20	3,552	2,845	2,528	2,086	1,725	1,325
21	3,527	2,831	2,518	2,080	1,721	1,323
22	3,505	2,819	2,508	2,074	1,717	1,321
23	3,485	2,807	2,500	2,069	1,714	1,319
24	3,467	2,797	2,492	2,064	1,711	1,318
25	3,450	2,787	2,485	2,060	1,708	1,316
26	3,435	2,779	2,479	2,056	1,706	1,315
27	3,421	2,771	2,473	2,052	1,703	1,314
28	3,408	2,763	2,467	2,048	1,701	1,313
29	3,396	2,756	2,462	2,045	1,699	1,311
30	3,385	2,750	2,457	2,042	1,697	1,310

Учебное издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие
для студентов специальностей
1-38 01 01 «Механические и электромеханические приборы
и аппараты», 1-38 01 02 «Оптико-электронные и лазерные приборы
и системы», 1-52 02 01 «Технология и оборудование ювелирного
производства», 1-38 02 02 «Биотехнические и медицинские
аппараты и системы», 1-54 01 01 «Метрология, стандартизация
и сертификация (машиностроение и приборостроение)»

В 2 частях

Часть 2

С о с т а в и т е л и:
ПРИХАЧ Наталия Константиновна
ПРУСОВА Ирина Васильевна
БОКУТЬ Людмила Валентиновна
КОНДРАТЬЕВА Наталья Анатольевна

Редактор *Е. О. Германович*
Компьютерная верстка *Е. А. Беспанской*

Подписано в печать 15.10.2020. Формат 60×84 $^{1}/_{16}$. Бумага офсетная. Ризография.
Усл. печ. л. 4,24. Уч.-изд. л. 3,32. Тираж 100. Заказ 593.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет.
Свидетельство о государственной регистрации издателя, изготовителя, распространителя
печатных изданий № 1/173 от 12.02.2014. Пр. Независимости, 65. 220013, г. Минск.