И.В. Марченко

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Практикум

Могилев 2009

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
«МОГИЛЕВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. А.А. КУЛЕШОВА»

И.В. Марченко

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Практикум

им.А.А.Кулешова

Могилев 2009

УДК 519.22(075.8)
ББК 22.172
М30

*Печатается по решению редакционно-издательского совета
УО «МГУ им. А.А. Кулешова»*

Рецензент
доцент кафедры математического анализа,
информатики и вычислительной техники
УО «МГУ им. А.А. Кулешова»
Л.А. Мазаник

Марченко, И.В.

М30 Математическая статистика: практикум / И.В. Марченко. – Могилев: УО «МГУ им. А.А. Кулешова», 2009. – 28 с.: ил.

ISBN 978-985-480-514-6.

Данное издание включает лабораторные работы и индивидуальные задания по основным разделам курса «Математическая статистика». Задания снабжены примерами и методическими рекомендациями по их решению с использованием табличного процессора MS Excel.

Предназначается для студентов физико-математических и экономических факультетов всех видов обучения.

УДК 519.22(075.8)
ББК 22.172

ISBN 978-985-480-514-6

ЛАБОРАТОРНАЯ РАБОТА № 1

Тема: Генеральная совокупность и выборка. Вариационные ряды. Полигон и гистограмма. Числовые характеристики выборки

Пример. Пятьдесят абитуриентов получили на вступительных экзаменах следующие баллы:

12,	14,	19,	15,	14,	18,	13,	16,	17,	12,
20,	17,	15,	13,	17,	16,	20,	14,	14,	13,
17,	16,	15,	19,	16,	15,	18,	17,	15,	14,
16,	15,	15,	18,	15,	15,	19,	14,	16,	18,
18,	15,	15,	17,	15,	16,	16,	14,	14,	17.

Требуется:

1. Составить вариационный ряд.
2. Записать таблицу частот и относительных частот.
3. Построить полигон частот и полигон относительных частот.
4. Найти эмпирическую функцию распределения $F^*(x)$. Построить ее график.
5. Найти числовые характеристики данной выборки ($\overline{X}_в$, $D_в$, $\sigma_в$, $D_{испр.в.}$, эмпирические моменты второго порядка ν_2^*, μ_2^*, коэффициент вариации V), воспользовавшись их статистическими определениями. Полученные результаты проверить с помощью встроенных функций Excel.

Решение.

1) Объем данной выборки $n = 50$. Вариационный ряд (варианты, расположенные в порядке возрастания): 12, 13, 14, 15, 16, 17, 18, 19, 20. Для построения вариационного ряда удобно воспользоваться следующими встроенными функциями Excel:

МИН(число1;число2; ...) – возвращает наименьшее значение в списке аргументов.

МАКС(число1;число2; ...) – возвращает наибольшее значение из набора значений.

Здесь число1, число2, ... – числа, среди которых требуется найти наименьшее (соответственно, наибольшее).

2) Составим таблицу частот и относительных частот. ***Частота*** n_k ***варианты*** – это количество повторов варианты в выборке. ***Относительная частота (частость)*** μ_k ***варианты*** – это отношение ее частоты к объему выборки

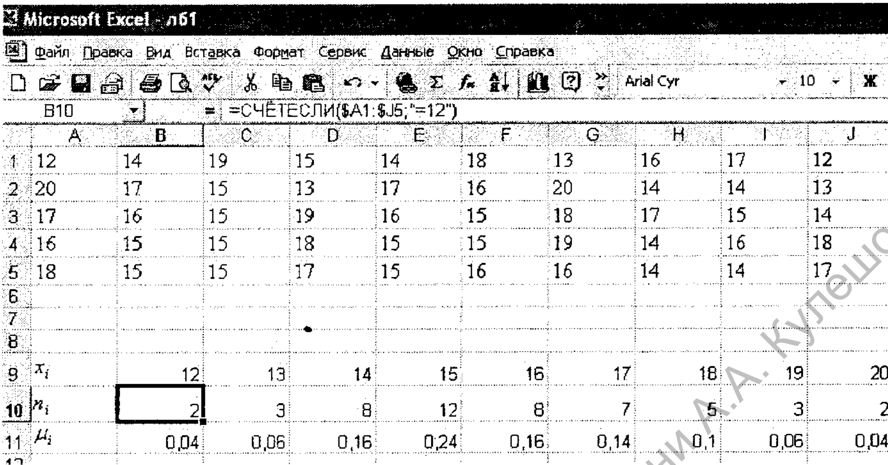
$$\mu_i = \frac{n_i}{n}. \tag{1}$$

Скопируем данную выборку в таблицу Excel и построим для нее таблицу частот и относительных частот. Для этого выполним следующие действия:

1. Заголовки строк (значения первого столбца) этой таблицы создадим в редакторе Microsoft Equation 3.0, для вызова которого выполним команды **Вставка / Объект ... / Microsoft Equation 3.0.**
2. Подсчет частот вариант осуществим статистической функцией **СЧЕТЕСЛИ**. Для этого выполним команды **Вставка / Функция / Статистические / СЧЕТЕСЛИ**. В появившемся диалоговом окне в качестве диапазона введем диапазон ячеек, в котором располагаются значения вариант данной выборки, а в качестве условия введем $=x_i$, где x_i – это фиксированное значение варианты. Если выборка располагается в диапазоне **A1:J5**, то в ячейку **B10** со значением частоты варианты $x_1 = 12$ следует ввести формулу СЧЕТЕСЛИ(**$A1:$J5; "=12"**) (см. рис. 1, строка формул отображает формулу в выделенной ячейке).
3. Относительные частоты определяются формулой (1), которая для ячейки **B11** примет вид = **B10**/50.
4. Итоговая таблица находится в диапазоне **A9 : J11** (см. рис. 1).

Замечание. *При заполнении таблицы используйте операцию копирования и не забывайте об относительных и абсолютных ссылках.*

3) По полученной в п. 2) таблице построим полигон частот и полигон относительных частот. ***Полигоном частот (относительных частот)*** называется ломаная линия, соединяющая точки с координатами (x_i, n_i) (соответственно, (x_i, μ_i)). Для построения этих ломаных воспользуемся графическими средствами Excel.



	A	B	C	D	E	F	G	H	I	J
1	12	14	19	15	14	18	13	16	17	12
2	20	17	15	13	17	16	20	14	14	13
3	17	16	15	19	16	15	18	17	15	14
4	16	15	15	18	15	15	19	14	16	18
5	18	15	15	17	15	16	16	14	14	17
6										
7										
8										
9	x_i	12	13	14	15	16	17	18	19	20
10	n_i	2	3	8	12	8	7	5	3	2
11	μ_i	0,04	0,06	0,16	0,24	0,16	0,14	0,1	0,06	0,04

Рис. 1

1. Выполним команды **Вставка / Диаграмма**. В появившемся диалоговом окне выберем тип диаграммы – график с маркерами, помечающими точки данных. Нажмем кнопку **Далее**.
2. В следующем окне **Мастера диаграмм** укажем диапазон – диапазон частот **B10:J10**, расположив ряды в строках. На вкладке **Ряд** выберем в качестве подписей по оси X диапазон значений вариант **B9:J9**. Нажмем кнопку **Далее**.
3. На третьем шаге **Мастера** на вкладке **Линии сетки** добавим по оси X основные линии. Нажмем кнопку **Далее**.
4. На последнем шаге разместим диаграмму на имеющемся листе и нажмем кнопку **Готово**.

Полученная диаграмма представляет собой полигон частот (см. рис. 2). Аналогично тому, как это делалось выше, построим полигон относительных частот (см. рис. 3).

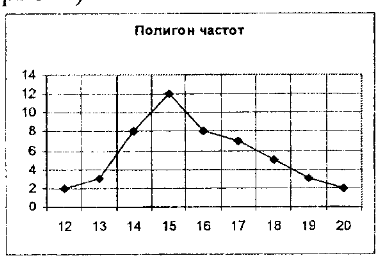


Рис. 2

Рис. 3

4) Если закон распределения выборки задан таблицей относительных частот

x_i	x_1	x_2	x_3	...	x_k
μ_i	μ_1	μ_2	μ_3	...	μ_k

то эмпирическая функция распределения имеет вид

$$F^*(x) = \begin{cases} 0, & -\infty < x \le x_1, \\ \mu_1, & x_1 < x \le x_2, \\ \mu_1 + \mu_2, & x_2 < x \le x_3, \\ \ldots \\ \sum\limits_{i=1}^{k-1} \mu_i, & x_{k-1} < x \le x_k, \\ 1, & x > x_k. \end{cases} \qquad (2)$$

Для вычисления значений этой функции и построения ее графика в Excel создадим таблицу (см. рис. 4, диапазон ячеек **A32 : K33**). В первую строку введем значения правых концов промежутков $|x_i, x_{i+1}|$ задания функции $F^*(x)$. Поскольку далее требуется построить график этой функции, то в ячейку **K32** вместо значения " $+\infty$ " поместим любое значение, большее $x_k = 20$, например, "21". Во второй строке таблицы вычислим значения функции на соответствующем промежутке. Для этого в ячейку **B32** введем значение "0", в ячейку **C32** поместим формулу

"=B33+B11" (в ячейке **B11** находится значение относительной частоты μ_1) и распространим ее путем копирования на последующие ячейки этой строки. По полученной таблице, воспользовавшись редактором формул Microsoft Equation 3.0, запишем функцию $F^*(x)$ в виде (2). Например, значение "0,04" находится в ячейке **C33**, соответствующей ячейке **C32**, содержимое которой есть число "13" – правый конец промежутка $(12;\ 13]$. Следовательно, при $x \in (12;\ 13]$ функция $F^*(x)$ принимает значение "0,04". В результате получим

$$F^*(x) = \begin{cases} 0, & -\infty < x \le 12, \\ 0{,}04, & 12 < x \le 13, \\ 0{,}1, & 13 < x \le 14, \\ 0{,}26, & 14 < x \le 15, \\ 0{,}5, & 15 < x \le 16, \\ 0{,}66, & 16 < x \le 17, \\ 0{,}8, & 17 < x \le 18, \\ 0{,}9, & 18 < x \le 19, \\ 0{,}96, & 19 < x \le 20, \\ 1, & 20 < x < +\infty. \end{cases}$$

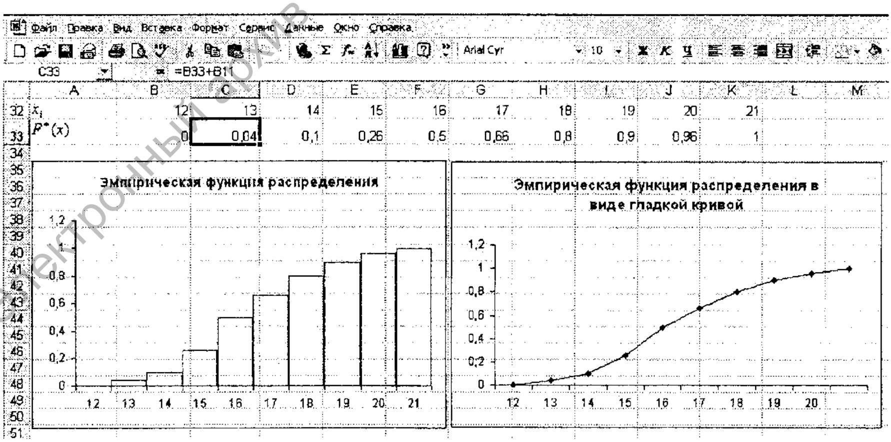


Рис. 4

Замечание.

График этой функции построим в виде гистограммы и в виде графика с маркерами (см. п. 3). Результаты построения приведены на рис. 4.

5) Для нахождения числовых характеристик выборки выполним необходимые промежуточные вычисления, результаты которых занесем в таблицу (см. рис. 5, диапазон ячеек **A53 : J55**).

B54 ▾ fx =(B9-B60)*(B9-B60)

	A	B	C	D	E	F	G	H	I	J
53	$x_i \mu_i$	0,48	0,78	2,24	3,6	2,56	2,38	1,8	1,14	0,8
54	$(x_i - \overline{X}_в)^2$	14,2884	7,7284	3,1684	0,6084	0,0484	1,4884	4,9284	10,3684	17,8084
55	$(x_i - \overline{X}_в)^2 \mu_i$	0,571536	0,463704	0,506944	0,146016	0,007744	0,208376	0,49284	0,622104	0,712336
56										
57										
58										
59		по формуле	проверка							
60	$\overline{X}_в$	15,78	15,78							
61	$D_в$	3,7316	3,7316							
62	$D_{испр.}$	3,8077551	3,8077551							
63	$\sigma_в$	1,93173497	1,931735							

Рис. 5

Выборочное среднее определяется формулой

$$\overline{X}_в = \sum_{i=1}^{k} x_i \mu_i,$$

поэтому предварительно найдем произведения значений вариант на соответствующие относительные частоты, для чего в ячейку **B53** введем формулу

= B9*B11,

при этом ячейка **B9** содержит значение варианты x_1, а ячейка **B11** – относительную частоту этой варианты. На остальные ячейки в строке **53** распространим данную формулу путем копирования ее из ячейки **B11**. После этого занесем в ячейку **B60** сумму значений в строке **53** с помощью функции **СУММ**. Эта сумма и будет искомым выборочным средним. Для проверки полученного результата воспользуемся статистической функцией **СРЗНАЧ**, которая возвращает среднее арифметическое своих аргументов. Чтобы найти с ее помощью выборочное среднее, введем в ячейку **C60** формулу

=СРЗНАЧ(A1:J5),

указав в качестве диапазона диапазон значений данной выборки.

Выборочной дисперсией называется сумма произведений квадратов отклонений вариант на соответствующие относительные частоты, т.е.

$$D_в = \sum_{i=1}^{k}\left(x_i - \overline{X}_в\right)^2 \mu_i .$$

Для вычисления дисперсии произведем вспомогательные вычисления в строках **54** и **55**. В ячейку **B54** введем формулу

=(B9-B60)*(B9-B60),

а в ячейку **B55**

=B54*B11,

после чего скопируем введенные формулы на остальные ячейки в этих строках. Результатом суммирования значений из строки 55 является выборочная дисперсия (ячейка **B61**). Проверку проведем с помощью статистической функции **ДИСПРА**, выбрав в качестве диапазона ее значений диапазон значений выборки.

Исправленной выборочной дисперсией называется величина

$$D_{испр.в} = \frac{n}{n-1}\cdot D_в .$$

Для ее нахождения в ячейку **B62** введем формулу =50/49*B61. Результат проверим функцией **ДИСПА**.

Поскольку ***выборочное среднее квадратическое отклонение*** $\sigma_в = \sqrt{D_в}$, для его вычисления поместим в ячейку **B63** формулу =КОРЕНЬ(B61). Результат проверим функцией **СТАНДОТКЛОН**.

Коэффициентом вариации называется процентное отношение выборочного среднего квадратического отклонения к выборочному среднему

$$V = \sigma_в / \overline{X}_в \cdot 100\% . \tag{3}$$

Эмпирическим начальным моментом порядка т называют выборочное среднее случайной величины X^m, которое определяется формулой

$$v_m^* = \overline{X}_в\left(X^m\right) = \sum_{i=1}^{k} x_i^m \mu_i . \tag{4}$$

Эмпирическим центральным моментом порядка т называют выборочное среднее отклонения в m-й степени

$$\mu_m^* = \overline{X}_в\left(X - \overline{X}_в\right)^m = \sum_{i=1}^{k}\left(x_i - \overline{X}_в\right)^m \mu_i. \qquad (5)$$

Эмпирические моменты и коэффициент вариации находятся в соответствии с формулами (3) – (5) аналогично тому, как это делалось выше.

Задача. В результате 50 независимых измерений некоторой величины получены данные

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

Требуется:

1. Составить вариационный ряд.
2. Записать таблицу частот и относительных частот.
3. Построить полигон частот и полигон относительных частот.
4. Найти эмпирическую функцию распределения $F^*(x)$. Построить ее график.
5. Найти числовые характеристики данной выборки ($\overline{X}_в$, $D_в$, $\sigma_в$, $D_{испр.в.}$, эмпирические моменты второго порядка ν_2^*, μ_2^*, коэффициент вариации V), воспользовавшись их статистическими определениями. Полученные результаты проверить с помощью встроенных функций Excel.

Указание. Методика выполнения этого задания та же, что и в примере.

ЛАБОРАТОРНАЯ РАБОТА № 2

Тема: Точечные оценки параметров распределения. Доверительные интервалы для математического ожидания и среднего квадратического отклонения. Критерий согласия χ^2 Пирсона.

Пример. Исходные данные см. в примере к лабораторной работе 1.

Требуется:

1) Исходя из графика эмпирической функции распределения $F^*(x)$, выдвинуть гипотезу о законе распределения генеральной совокупности.

2) Найти точечные оценки математического ожидания и среднего квадратического отклонения. Записать с их учетом плотность распределения вероятности $f(x)$.

3) С помощью критерия согласия Пирсона проверить предположение о нормальном законе распределения данной выборки.

4) Если случайная величина X распределена нормально, то найти доверительные интервалы для математического ожидания и среднего квадратического отклонения, взяв доверительную вероятность $\gamma = 0{,}95$.

5) Вычислить вероятность $P(14{,}5 < X < 18{,}5)$.

Решение.

1) Вид эмпирической функции распределения (см. рис. 4) позволяет выдвинуть гипотезу о нормальном законе распределения.

Замечание. Если возникают сомнения, то следует рассмотреть полигон или гистограмму относительных частот, которые дают представление о плотности распределения вероятности. В нашем случае кривая напоминает график плотности распределения вероятности нормального распределения.

2) Функция плотности распределения вероятности нормального распределения имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (6)$$

Точечными оценками параметров a и σ нормального распределения являются выборочное среднее $\overline{X}$ и выборочное среднее квадратическое отклонение $S = \sqrt{D_в}$, т.е. $a \approx \overline{X}$, $\sigma \approx S$. Для рассматриваемого примера после подстановки найденных в п. 5) лабораторной работы 1 значений $\overline{X} = 15,78$ и $S = 1,932$ в формулу (6) получаем

$$f(x) = 0,21e^{-\frac{(x-15,78)^2}{7,45}}.$$

3) Для проверки предположения о нормальном законе распределения воспользуемся критерием согласия Пирсона. Чтобы построить интервальную таблицу относительных частот, определим шаг таблицы. Для этого найдем размах варьирования

$$R = |x_{\max} - x_{\min}| = 20 - 12 = 8$$

и подберем такое число $n \geq 5$, на которое его удобно разделить. Возьмем число частичных интервалов $n = 8$. Тогда шаг таблицы $h = 1$.

$\|x_i, x_{i+1}\|$	[12, 13)	[13, 14)	[14, 15)	[15, 16)	[16, 17)	[17, 18)	[18, 19)	[19, 20]
m_i	2	3	8	12	8	7	5	5

Поскольку выбранный критерий согласия требует, чтобы частота варианты в каждом частичном интервале была не меньше 5, и первые два интервала не удовлетворяют этому условию, объединим их. В результате получим новую интервальную таблицу относительных частот

$\|x_i, x_{i+1}\|$	[12, 14)	[14, 15)	[15, 16)	[16, 17)	[17, 18)	[18, 19)	[19, 20]
m_i	5	8	12	8	7	5	5

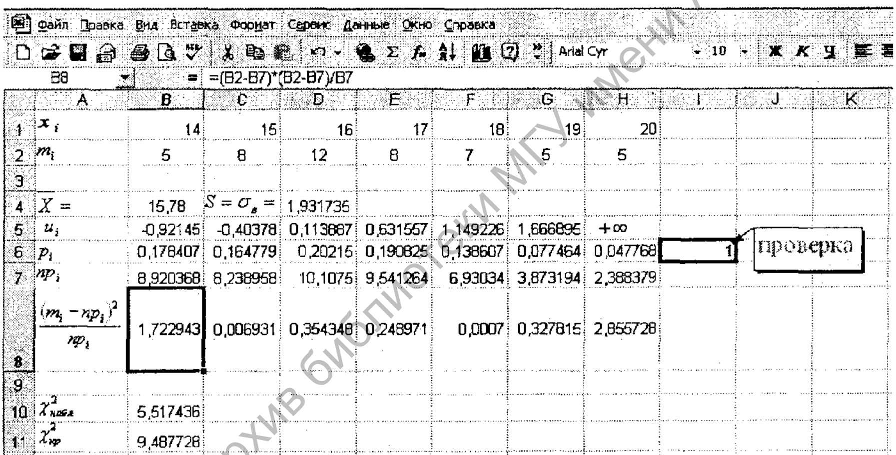
Внесем эти данные в таблицу Excel (см. рис. 6, ячейки **A1:H2**). В первую строку введем значения правых[1] концов промежутков $|x_i, x_{i+1}|$, а во вторую строку соответствующие им частоты m_i.
Найдем вероятности p_i по формуле

[1] Значение x_1 не участвует в дальнейших вычислениях, поэтому нет необходимости вносить его в таблицу Excel.

$$p_i = P\left(X \in \left[x_i; x_{i+1}\right)\right) = F\left(x_{i+1}\right) - F\left(x_i\right) = \Phi\left(\frac{x_{i+1} - \overline{X}}{S}\right) - \Phi\left(\frac{x_i - \overline{X}}{S}\right). \quad (7)$$

Поместим в ячейки **B4** и **D4** соответственно $\overline{X}$ и S, которые были вычислены в лабораторной работе 1. Для удобства вычислений составим таблицу (рис. 6, диапазон **A5 : H8**). Значения $u_i = \dfrac{x_i - \overline{X}}{S}$ в пятой строке вычислим с помощью статистической функции **НОРМАЛИЗАЦИЯ**. Для этого поместим в ячейку **B5** формулу

=НОРМАЛИЗАЦИЯ(B1;B4;D4)



	A	B	C	D	E	F	G	H	I	J	K
1	x_i	14	15	16	17	18	19	20			
2	m_i	5	8	12	8	7	5	5			
3											
4	$\overline{X}$ =	15,78	$S=\sigma_в=$	1,931735							
5	u_i	-0,92145	-0,40378	0,113887	0,631557	1,149226	1,666895	$+\infty$			
6	p_i	0,178407	0,164779	0,20215	0,190825	0,138607	0,077464	0,047768	1	проверка	
7	np_i	8,920368	8,238958	10,1075	9,541264	6,93034	3,873194	2,388379			
8	$\frac{(m_i - np_i)^2}{np_i}$	1,722943	0,006931	0,354348	0,248971	0,0007	0,327815	2,855728			
9											
10	$\chi^2_{набл}$	5,517436									
11	$\chi^2_{кр}$	9,487728									

Рис. 6

При этом ячейка **B1** содержит значение "14" правого конца первого частичного интервала [12;14). Распространим эту формулу путем копирования на ячейки **C5:G5**. Поскольку возможные значения нормально распределенной случайной величины лежат в промежутке (-∞,+∞), то значение в ячейке **H5** не вычисляем (оно соответствует $x_i = 20$), а вводим "+∞" с помощью Microsoft Equation 3.0.

Значения вероятности p_i вычислим по формуле (7) с помощью статистической функции НОРМСТРАСП. Для этого в ячейку **C6** введем формулу

=НОРМСТРАСП(C5)–НОРМСТРАСП(B5)

и скопируем ее в ячейки **D6:G6**. В первую **B6** и последнюю **H6** ячейки данной строки введем соответственно формулы[2]

=НОРМСТРАСП(B5) и =1 – НОРМСТРАСП(G5)

Обязательно ***выполним проверку***: сумма значений p_i должна быть равной единице.

Вычислим наблюдаемое значение критерия $\chi^2_{набл.} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}$, просуммировав с помощью функции **СУММ** значения **B8 : H8**. Получим $\chi^2_{набл.} \approx 5{,}52$.

Найдем по уровню значимости $\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05$ и числу степеней свободы $\nu = k - r - 1 = 7 - 2 - 1 = 4$, где $k = 7$ – число интервалов, а r – число параметров распределения (для нормального распределения $r = 2$), критическое значение критерия $\chi^2_{кр} = \chi^2_{\alpha;\nu} = \chi^2_{0,05;4}$. Для этого введем в ячейку **B11** формулу

=ХИ2ОБР(0,05;4).

Имеем $\chi^2_{кр} \approx 9{,}49$.

Сравним наблюдаемое и критическое значения критерия. Так как $\chi^2_{набл.} < \chi^2_{кр}$, то нет оснований отклонить нулевую гипотезу о нормальном распределении генеральной совокупности, из которой извлечена данная выборка.

4) Концы доверительного интервала для математического ожидания определяются соотношениями $\overline{X} \pm t_{\alpha,n} \frac{S}{\sqrt{n}}$. Для вычисления значения $t_{\alpha;n} = t_{0,05;50}$ воспользуемся следующей стандартной функцией Excel

=СТЬЮДРАСПОБР(0,05;50).

[2] Это связано с тем, что для функции НОРМСТРАСП имеет место равенство НОРМСТРАСП(x)= $\frac{1}{2} + \Phi(x)$.

Доверительный интервал для среднего квадратического отклонения σ имеет вид

$$S \cdot (1-q) < \sigma < S \cdot (1+q).$$

Значение $q = q_{\gamma;n} = q_{0,95;50} = 0,21$ найдем по таблице значений этой функции (см. приложение).

В результате получим

$$15,23 < a < 16,33,\ 1,53 < \sigma < 2,34.$$

5) Вероятность попадания случайной величины X в заданный интервал $(\alpha;\beta)$ определяется формулой

$$P(\alpha < X < \beta) = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta - \overline{X}}{S}\right) - \Phi\left(\frac{\alpha - \overline{X}}{S}\right). \qquad (8)$$

Для нахождения вероятности $P(14,5 < X < 18,5)$ вычислим значения $\Phi\left(\frac{18,5 - \overline{X}}{S}\right)$ и $\Phi\left(\frac{14,5 - \overline{X}}{S}\right)$ аналогично тому, как это делалось в п. 3.

Задача. Для выборки из задачи лабораторной работы 1 выполните задания рассмотренного выше примера.

ЛАБОРАТОРНАЯ РАБОТА № 3

Тема: Метод наименьших квадратов. Линейная регрессия. Выборочный коэффициент корреляции.

Задача 1. Дана таблица экспериментальных данных

x_i	0,25	0,37	0,44	0,55	0,60	0,62	0,68	0,70	0,73	0,75	0,82	0,84	0,87	0,88	0,90	0,95	1
y_i	2,57	2,31	2,12	1,92	1,75	1,71	1,60	1,51	1,50	1,41	1,33	1,31	1,25	1,20	1,19	1,15	1

Требуется:

1) Построить корреляционное поле и высказать предположение о виде функции регрессии Y на X.

2) Методом наименьших квадратов найти коэффициенты уравнения регрессии Y на X. Построить полученную линию на координатной плоскости.

3) Найти интервальные оценки для коэффициентов модельного уравнения регрессии Y на X, взяв уровень значимости $\alpha = 0{,}05$.

4) Найти эмпирический коэффициент корреляции и проверить гипотезу о его значимости при уровне значимости $\alpha = 0{,}05$.

Решение.

1) Для построения корреляционного поля (см. рис. 7) командами **Вставка / Диаграмма** вызовем **Мастер диаграмм** и выполним следующие действия:

1. В появившемся диалоговом окне выберем тип диаграммы – точечная. Нажмем кнопку **Далее**.

2. На следующем шаге на вкладке **Диапазон данных** в строке **Диапазон** укажем числовые данные исходной таблицы, выбрав ряды в строках.

3. В третьем окне укажем заголовки осей: **ось X** (категорий) – X, **ось Y** (значений) – Y.

4. На последнем шаге поместим диаграмму на имеющемся листе. Нажмем кнопку **Готово**.

По расположению точек на координатной плоскости можно высказать предположение о линейной регрессионной зависимости Y на X.

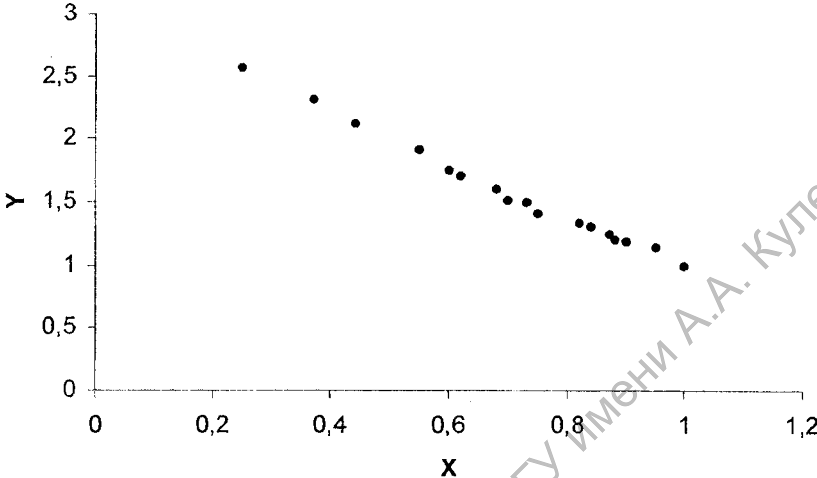


Рис. 7. Корреляционное поле для регрессионной зависимости Y на X

2) Согласно выдвинутой в пункте 1) гипотезе эмпирическое уравнение регрессии Y на X имеет вид $\bar{Y}_x = b_0 + b_1 x$. Коэффициенты этого уравнения b_0, b_1 найдем методом наименьших квадратов. Для этого решим систему

$$\begin{cases} b_0 + b_1 \overline{X} = \overline{Y}, \\ b_0 \overline{X} + b_1 \overline{X^2} = \overline{XY}, \end{cases} \tag{9}$$

где

$$\frac{\sum x_i}{n} = \overline{X}, \quad \frac{\sum y_i}{n} = \overline{Y}, \quad \frac{\sum x_i^{\,2}}{n} = \overline{X^2}, \quad \frac{\sum x_i y_i}{n} = \overline{XY}. \tag{10}$$

Числовые характеристики (10) вычислим средствами Excel аналогично тому, как это делалось в лабораторной работе 1. Для данной задачи получаем

$\overline{X} = 0{,}702941$, $\overline{Y} = 1{,}578235$, $\overline{X^2} = 0{,}535853$, $\overline{XY} = 1{,}023041$.

Договоримся в дальнейшем при выполнении промежуточных вычислений оставлять 6 знаков после запятой, а в конечных результатах – 4 знака после запятой.

Система (9) принимает вид

$$\begin{cases} b_0 & + & 0{,}702941 \cdot b_1 & = & 1{,}578235, \\ 0{,}702941 \cdot b_0 & + & 0{,}535853 \cdot b_1 & = & 1{,}023041. \end{cases}$$

Решим ее методом Крамера. Вычислим основной Δ и вспомогательные Δ_0, Δ_1 определители этой системы с помощью встроенной функции **МОПРЕД** из категории **Математические**.

$$\Delta = \begin{vmatrix} 1 & 0{,}702941 \\ 0{,}702941 & 0{,}535853 \end{vmatrix} = 1 \cdot 0{,}535853 - 0{,}702941 \cdot 0{,}702941 = 0{,}041727,$$

$$\Delta_0 = \begin{vmatrix} 1{,}578235 & 0{,}702941 \\ 1{,}023041 & 0{,}535853 \end{vmatrix} = 0{,}126564, \quad \Delta_1 = \begin{vmatrix} 1 & 1{,}578235 \\ 0{,}702941 & 1{,}023041 \end{vmatrix} = -0{,}086365.$$

Тогда по формулам Крамера имеем

$$b_0 = \frac{\Delta_0}{\Delta} = \frac{0{,}126564}{0{,}041727} \approx 3{,}0332, \quad b_1 = \frac{\Delta_1}{\Delta} = \frac{-0{,}086365}{0{,}041727} \approx -2{,}0698.$$

Подставив найденные коэффициенты в эмпирическое уравнение регрессии Y на X, получим $\overline{Y}_x = -2{,}0698x + 3{,}0332$.

Проверим найденные значения коэффициентов с помощью статистической функции Excel **ЛИНЕЙН**. Введем в ячейку **F33** формулу

=ЛИНЕЙН(B2:R2;B1:R1),

для которой в диапазонах **B2:R2** и **B1:R1** содержатся соответственно значения y_i и x_i. Чтобы эта формула была формулой массива, выделим ячейки **F33** и **G33**, нажмем клавишу F2, а затем – клавиши CTRL+ +SHIFT+ENTER.

Изобразим эту прямую на плоскости. Для этого выберем любые две точки, через которые проходит данная прямая, например, (0; 3,0332) и (1,4655; 0). Введем координаты этих точек в таблицу Excel и построим аналогично тому, как это делалось в пункте 1), проходящую через эти точки прямую с помощью **Мастера диаграмм**, выбрав вид диаграммы "**точечная диаграмма, на которой значения соединены отрезками**". Результат на рис. 8.

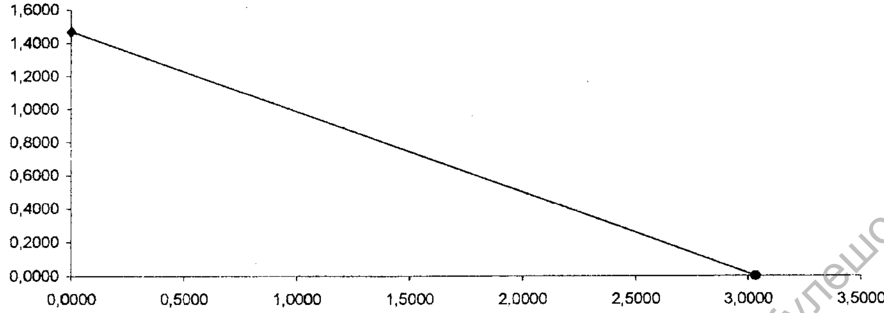


Рис. 8. График линии регрессии Y на X

3) В силу выдвинутого выше предположения о линейной корреляционной зависимости между величинами X и Y модельное уравнение регрессии имеет вид

$$M(Y \mid X) = \beta_0 + \beta_1 X .$$

Построим интервальные оценки для коэффициентов β_0 и β_1 этого уравнения по формулам

$$b_0 - t_{\frac{\alpha}{2},n-2} S_{b_0} < \beta_0 < b_0 + t_{\frac{\alpha}{2},n-2} S_{b_0},$$

$$b_1 - t_{\frac{\alpha}{2},n-2} S_{b_1} < \beta_1 < b_1 + t_{\frac{\alpha}{2},n-2} S_{b_1},$$

где $S_{b_0} = S_e \cdot \sqrt{\dfrac{\overline{X^2}}{n\left(\overline{X^2} - \left(\overline{X}\right)^2\right)}}$, $S_{b_1} = S_e \cdot \sqrt{\dfrac{1}{n\left(\overline{X^2} - \left(\overline{X}\right)^2\right)}}$

есть средние квадратические отклонения коэффициентов регрессии, а

$S_e = \sqrt{\dfrac{\sum (y_i - b_0 - b_1 x_i)^2}{n-2}}$ – среднее квадратическое отклонение остатков $e_i = y_i - b_0 - b_1 x_i$.

Значения функции Стьюдента находим аналогично тому, как это делалось в лабораторной работе 2. Для нахождения значения S_e составим таблицу (см. рис. 9), в первой строке которой вычислим остатки e_i, а во второй их квадраты. Формула Excel для вычисления, например, остатка e_1 имеет следующий вид

=B2-G32-F32*B1.

Здесь в ячейках **B1** и **B2** находятся значения x_1 и y_1, а в ячейках **G32** и **F32** находятся, соответственно, значения коэффициентов b_0 и b_1, поэтому ссылки на эти ячейки должны быть абсолютными.

H59 fx =H2-G32-F32*H1

	A	B	C	D	E	F	G	H	I	J	K	L	M
21		$\overline{X}$	0,702941										
22		$\overline{Y}$	1,578235										
23													
24	x_i^2	0,062500	0,136900	0,193600	0,302500	0,360000	0,384400	0,462400	0,490000	0,532900	0,562500	0,672400	0,705600
25	y_i^2	6,604900	5,336100	4,494400	3,686400	3,062500	2,924100	2,560000	2,280100	2,250000	1,988100	1,768900	1,716100
26	$x_i y_i$	0,642500	0,854700	0,932800	1,056000	1,050000	1,060200	1,088000	1,057000	1,095000	1,057500	1,090600	1,100400
27		$\overline{X^2}$	0,535853										
28		$\overline{Y^2}$	2,671335										
29		$\overline{XY}$	1,023041										
30													
31		Δ	0,041727			b_1	b_0						
32		Δ_0	0,126564			-2,0698	3,033176						
33		Δ_1	-0,086365		проверка	-2,0698	3,033176						
34													
59	e_i	0,054271	0,042646	-0,002468	0,025209	-0,041302	-0,039906	-0,025719	-0,074323	-0,022229	-0,070833	-0,005948	0,015448
60	e_i^2	0,002945	0,001819	0,000006	0,000635	0,001706	0,001592	0,000661	0,005524	0,000494	0,005017	0,000035	0,000239
61													
62		S_e	0,044540		S_{b_0}	0,038711		S_{b_1}	0,052883				
63		$t_{0,025;15}$	2,489878										
64													
65		2,9368	$< \beta_0 <$	3,1296		-2,2015	$< \beta_1 <$	-1,9381					

Рис. 9

После выполнения необходимых действий получаем следующие результаты

$$S_e = \sqrt{\frac{0{,}029757}{15}} \approx 0{,}044540,$$

$$S_{b_0} = \sqrt{\frac{0{,}535853}{17\left(0{,}535853 - 0{,}702941^2\right)}} \cdot 0{,}044540 \approx 0{,}038711,$$

$$S_{b_1} = \sqrt{\frac{1}{17\left(0{,}53583 - 0{,}702941^2\right)}} \cdot 0{,}044540 \approx 0{,}052883.$$

Таким образом, доверительные интервалы для коэффициентов модельного уравнения регрессии имеют вид

$$2{,}9368 < \beta_0 < 3{,}1296,\ -2{,}2015 < \beta_1 < -1{,}9381.$$

4) Эмпирический коэффициент корреляции найдем по формуле

$$r_{xy}^* = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sqrt{\overline{X^2} - \left(\overline{X}\right)^2} \cdot \sqrt{\overline{Y^2} - \left(\overline{Y}\right)^2}}. \qquad (11)$$

Подставив в соотношение (11) найденные значения

$\overline{Y^2} = 2{,}671335$, $\overline{X} = 0{,}702941$, $\overline{Y} = 1{,}578235$, $\overline{X^2} = 0{,}535853$, $\overline{XY} = 1{,}023041$, получим

$$r_{xy}^* = \frac{1{,}023041 - 0{,}702941 \cdot 1{,}578235}{\sqrt{0{,}535853 - (0{,}702941)^2} \cdot \sqrt{2{,}671335 - (2{,}671335)^2}} \approx -0{,}995140.$$

Проверить найденное значение можно с помощью статистической функции Excel **КОРРЕЛ**.

Решение об адекватности линейного уравнения регрессии экспериментальным данным примем на основании критерия Стьюдента. Для этого сравним наблюдаемое значение критерия $t_{набл.} = \frac{r_{xy}^* \sqrt{n-2}}{\sqrt{1-\left(r_{xy}^*\right)^2}}$ с критическим $t_{кр} = t_{\alpha;n-2}$, которое найдем по уровню значимости α и числу степеней свободы $n-2$, где n – количество экспериментальных точек.

В данном случае

$$t_{набл.} = \frac{-0{,}995140\sqrt{15}}{\sqrt{1-(-0{,}995140)^2}} = -39{,}1391.$$

Используя функцию **СТЬЮДРАСПОБР**, получаем $t_{кр} = t_{0,05;15} = 2{,}13\ 1451$. Так как $\left|t_{набл.}\right| > t_{кр}$, то выборочный коэффициент корреляции значимо отличается от нуля, т.е. величины X и Y коррелированы. Следовательно, линейная регрессия модельной функции $M(Y \mid X) = \beta_0 + \beta_1 X$ выбрана удачно (согласуется с экспериментальными данными).

ЗАДАНИЯ ДЛЯ ИНДИВИДУАЛЬНОЙ РАБОТЫ

Требования к оформлению ИДЗ

1. Обязательно указать формулы, теоремы, определения математической статистики, которые используются при решении задачи. Указывать формулы Excel и способ решения в Excel не следует.

2. Работа должна содержать номер задания и его условие.

3. На проверку предоставлять печатный вариант (листы не больше формата А4) и электронный вариант ИДЗ.

Задача 1. По данному статистическому материалу опыта требуется:

1. Составить вариационный ряд.

2. Записать таблицу частот и относительных частот.

3. Построить полигон частот и относительных частот.

4. Найти эмпирическую функцию распределения и построить ее график.

5. Вычислить числовые характеристики данной выборки ($\overline{X}_в$, $D_в$, $\sigma_в$, $D_{испр.в.}$, коэффициент вариации V), воспользовавшись их статистическими определениями. Полученные результаты проверить с помощью встроенных функций Excel.

6. Исходя из графика эмпирической функции распределения $F^*(x)$, выдвинуть гипотезу о законе распределения генеральной совокупности.

7. Найти точечные оценки математического ожидания и среднего квадратического отклонения. Записать с их учетом плотности распределения вероятности $f(x)$.

8. С помощью критерия согласия Пирсона проверить предположение о нормальном законе распределения данной выборки.

9. Если случайная величина X распределена нормально, то найти доверительные интервалы для математического ожидания и среднего квадратического отклонения, взяв доверительную вероятность $\gamma = 0{,}95$.

10. Вычислить вероятность $P(36 < X < 42)$.

Статистические данные для задачи находятся в следующей таблице, из которой берутся 5 строк, начиная с i-й строки, где i – ваш номер в журнале. Например, если ваш номер в журнале 26, то следует взять 26 – 28 и 1 – 2 строки.

№ строки	***Данные задачи***									
1	43	32	44	25	43	40	31	28	41	35
2	38	41	32	38	24	43	25	37	46	38
3	46	49	32	34	31	24	41	50	38	29
4	40	31	28	41	35	31	26	34	49	32
5	43	25	37	46	38	46	26	38	37	49
6	24	41	50	38	29	43	37	46	38	25
7	41	32	34	49	44	43	32	44	25	43
8	37	31	47	50	34	38	41	32	38	24
9	25	37	40	32	35	46	49	32	34	31
10	28	44	43	31	44	46	26	38	37	49
11	38	35	29	43	38	43	37	46	38	25
12	31	26	34	49	32	41	32	34	49	44
13	46	26	38	37	49	37	31	47	50	34
14	43	37	46	38	25	25	37	40	32	35
15	40	31	28	41	35	31	26	34	49	32
16	43	25	37	46	38	46	26	38	37	49
17	24	41	50	38	29	43	37	46	38	25
18	28	44	43	31	44	46	26	38	37	49
19	38	35	29	43	38	43	37	46	38	25
20	31	26	34	49	32	41	32	34	49	44
21	43	32	44	25	43	40	31	28	41	35
22	38	41	32	38	24	43	25	37	46	38
23	46	49	32	34	31	24	41	50	38	29
24	46	26	38	37	49	37	31	47	50	34
25	43	37	46	38	25	25	37	40	32	35
26	24	41	50	38	29	43	37	46	38	25
27	41	32	34	49	44	43	32	44	25	43
28	37	31	47	50	34	38	41	32	38	24

Задача 2. По данным следующей таблицы:

1) Построить корреляционное поле и высказать предположение о виде функции регрессии Y на X.

2) Методом наименьших квадратов найти коэффициенты уравнения регрессии Y на X. Построить полученную линию на координатной плоскости.

3. Найти интервальные оценки для коэффициентов модельного уравнения регрессии Y на X, взяв уровень значимости $\alpha = 0{,}05$.

4. Найти эмпирический коэффициент корреляции и проверить гипотезу о его значимости при уровне значимости $\alpha = 0{,}05$.

Данные для таблицы определяются следующим образом. Номер варианта i – последняя цифра номера в журнале, число экспериментальных точек из таблицы 10 + j, где j – остаток от деления номера в журнале на 3. Например, если ваш номер в журнале 25, то требуется выполнить вариант 5 с числом экспериментальных точек 11.

Вариант 0.

x_i	0	10	20	30	40	50	60	70	80	90	100	110
y_i	0,533	0,552	0,574	0,596	0,619	0,645	0,667	0,690	0,710	0,732	0,756	0,764

Вариант 1.

x_i	75	76	77	80	82	85	88	90	91	92	94	95
y_i	2,1	2,0	2,5	2,4	3,6	4,0	4,1	5,0	5,4	5,1	5,5	6,2

Вариант 2.

x_i	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,10	0,11
y_i	0,533	0,552	0,574	0,596	0,619	0,645	0,667	0,690	0,710	0,734	0,765	0,762

Вариант 3.

x_i	10	15	20	25	30	35	40	45	50	55	60	65
y_i	1,8	2,7	2,5	4,5	4,4	6,3	6,5	6,5	9,5	9,5	19,4	10,2

Вариант 4.

x_i	31	30	35	42	40	55	48	64	59	70	75	80
y_i	2,0	2,6	3,0	3,9	5,2	7,0	6,2	7,5	8,6	12,2	13,2	14,5

Вариант 5.

x_i	3,0	3,6	4,0	4,5	5,2	5,6	6,0	6,4	7,0	7,5	8,0	9,0
y_i	1,98	1,92	1,93	1,81	1,83	1,70	1,73	1,68	1,60	1,66	1,43	1,41

Вариант 6.

x_i	75	76	77	80	82	85	88	90	91	92	93	94
y_i	2,1	2,0	2,4	2,4	3,6	4,0	4,1	5,0	5,4	5,1	5,1	5,4

Вариант 7.

x_i	0,050	0,070	0,100	0,125	0,150	0,175	0,200	0,225	0,250	0,275	0,300	0,325
y_i	0,005	0,052	0,012	0,015	0,017	0,025	0,026	0,033	0,034	0,043	0,046	0,048

Вариант 8.

x_i	0	1	2	3	4	5	6	7	8	9	10	11
y_i	0,537	0,552	0,567	0,598	0,619	0,625	0,667	0,690	0,710	0,735	0,756	0,766

Вариант 9.

x_i	0	10	20	30	40	50	60	70	80	90	100	110
y_i	0,73	0,75	0,74	0,76	0,79	0,80	0,82	0,85	0,86	0,88	0,90	0,91

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. **Айвазян С.А.** и др. Прикладная статистика: Исследование зависимостей. – М.: Финансы и статистика, 1985.
2. **Гмурман В.Е.** Теория вероятностей и математическая статистика. – М.: Высш. шк., 1999.
3. **Гмурман В.Е.** Руководство к решению задач по теории вероятностей и математической статистики. – М.: Высш. шк., 1999.
4. **Гихман И.И., Скороход А.В.** Введение в теорию случайных процессов. – М.: Наука, 1977.
5. **Калинина В.Н., Колемаев В.А.** Теория вероятностей и математическая статистика. – М.: ИНФРА-М, 2001.
6. **Коваленко И.Н., Филиппова А.А.** Теория вероятностей и математическая статистика. – М.: Высш. шк., 1973.
7. **Колемаев В.А.** и др. Теория вероятностей и математическая статистика. – М.: Высш. шк., 1991.
8. **Лихолетов И.И.** Высшая математика, теория вероятностей и математическая статистика. – Мн.: Выш. шк., 1976.
9. **Микулик Н.А., Рейзина Г.Н.** Руководство к решению технических задач по теории вероятностей и математической статистике. – Мн.: Выш. шк., 1977.
10. **Нейман Ю.** Вводный курс теории вероятностей и математической статистики. – М.: Наука, 1968.
11. **Пугачев В.С.** Теория вероятностей и математическая статистика. – М.: Наука, 1979.
12. **Савич Л.К., Смольская Н.А.** Теория вероятностей и математическая статистика. – Мн.: Адукацыя і выхаванне, 2006.
13. **Себер Дж.** Линейный регрессионный анализ. – М.: Мир, 1980.
14. **Феллер В.** Введение в теорию вероятностей и ее приложения. – Т. 2. – М.: Мир, 1967.
15. **Худсон Д.** Статистика для физиков. – М.: Мир, 1967.
16. **Щиголев Б.М.** Математическая обработка наблюдений. – М.: Наука, 1969.

ПРИЛОЖЕНИЕ

Таблица значений $q = q(\gamma, n)$

n	γ		
	0,95	0,99	0,999
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

СОДЕРЖАНИЕ

Лабораторная работа № 1 3

Лабораторная работа № 2 11

Лабораторная работа № 3 16

Задания для индивидуальной работы 22

Рекомендуемая литература 26

Приложение 27

Учебное издание

Марченко Ирина Васильевна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Практикум

Технический редактор *А.Н. Гладун*
Компьютерная верстка *А.Л. Позняков*
Корректор *Г.В. Тетерукова*

Подписано в печать 3.02.2009.
Формат 60x84/16. Гарнитура Times New Roman Cyr.
Усл.-печ. л. 1,4. Уч.-изд. л. 1,35. Тираж 80 экз. Заказ № 42

Учреждение образования “Могилевский государственный университет им. А.А. Кулешова”, 212022, Могилев, Космонавтов, 1
ЛИ № 02330/278 от 30.04.2004 г.

Отпечатано на ризографе отдела оперативной полиграфии МГУ им. А.А. Кулешова. 212022, Могилев, Космонавтов, 1