ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ПРОФЕССИОНАЛЬНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ АРХАНГЕЛЬСКОЙ ОБЛАСТИ
«МИРНИНСКИЙ ПРОМЫШЛЕННО-ЭКОНОМИЧЕСКИЙ ТЕХНИКУМ»

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ВЫПОЛНЕНИЯ ПРАКТИЧЕСКИХ РАБОТ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

ЕН.02 ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

для специальности: 09.02.01 Компьютерные системы и комплексы

2017 г.

Методические рекомендации для ЕН.02 Теория вероятностей и математическая статистика разработаны для выполнения практических работ по разделу – «Математическая статистика» и составлены в соответствии с рабочей программой и учебным планом по специальности 09.02.01 «Компьютерные системы и комплексы».

Организация-разработчик: государственное бюджетное профессиональное образовательное учреждение Архангельской области «Мирнинский промышленно-экономический техникум»

Разработчики:
Пивоварова Т. В., Преподаватель техникума

ОДОБРЕНЫ цикловой комиссией дисциплин специальностей 09.02.01 и 13.02.11	Составлены в соответствии с требованиями ФГОС по специальности среднего профессионального образования 09.02.01 «Компьютерные системы и комплексы» и учебным планом
Председатель цикловой комиссии В.И.Письменник	Заместитель директора техникума по учебной работе М.Н.Венедиктова

СОДЕРЖАНИЕ

Введение	3
Задача № 1 Выборка. Эмпирическая функция распределения ДСВ	3
Задача № 2 Вариационный ряд. Полигон частот	5
Задача № 3 Интервальные оценки параметров распределения	6
Задача № 4 Начальные и центральные моменты	8
Задача № 5 Доверительный интервал. Надёжность	9
Задача № 6 Статистическое распределение выборки	10
Задача № 7 Распределение Пирсона	13
Задача № 8 Нормальное распределение случайной величины	13
Задача № 9 Коэффициент корреляции. Остаточная дисперсия	16
Задача № 10 Корреляционное поле	18
Задача № 11 Эмпирическая функция распределения. Кумулята	21
Задача № 12 Коэффициент корреляции. Остаточная дисперсия	26
Список использованных источников	30

ВВЕДЕНИЕ

Математическая статистика – раздел математики, посвящённый методам сбора, анализа и обработки результатов статистических данных наблюдений для научных и практических целей. Наибольшие трудности при изучении теории вероятностей и математической статистике традиционно вызывают приёмы решения задач, так как сложно алгоритмизировать процесс вычислений. Методические рекомендации предназначены для помощи студентам при выполнении практических работ по учебной дисциплине «Теория вероятностей и математическая статистика».

Задачи с решением

Задача № 1 Выборка. Эмпирическая функция распределения ДСВ

- Из генеральной совокупности извлечена выборка объёма $n = 21$:

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

Требуется:

1)Найти и построить эмпирическую функцию распределения;

2)Найти выборочное среднее, «исправленное» СКО, выборочную моду и медиану.

Решение:

4) Согласно определению эмпирической функции распределения её значение при любом x равно $F^*(x) = \frac{n_x}{n}$, где n_x — количество элементов x_i; выборки, меньших, чем $x; n = \sum n_i$ — объём выборки.

Например, при

$x = 1$ имеем $n_x = 0$, $F^*(x) = 0$; при $x = 4$ $n_x = 3$ $F^*(x) = \frac{3}{21} = \frac{1}{7} \approx 0{,}1$; при $x = 7$ $n_x = 3 + 4 = 7$, $F^*(x) = \frac{7}{21} \approx 0{,}3$; при $x = 9$ $n_x = 3 + 4 + 6 = 13$, $F^*(x) = \frac{13}{21} \approx 0{,}6$; при $x = 11$ $n_x = 3 + 4 + 6 + 8 = 21$; , $F^*(x) = \frac{21}{21} = 1$.

Тогда

$$F^*(x) = \begin{cases} 0 & x \le 2 \\ 0{,}1 & 2 < x \le 6 \\ 0{,}3 & 6 < x \le 8 \\ 0{,}8 & 8 < x \le 10 \\ 1 & x > 10. \end{cases}$$

График эмпирической функции распределения изображён на рисунке 6.

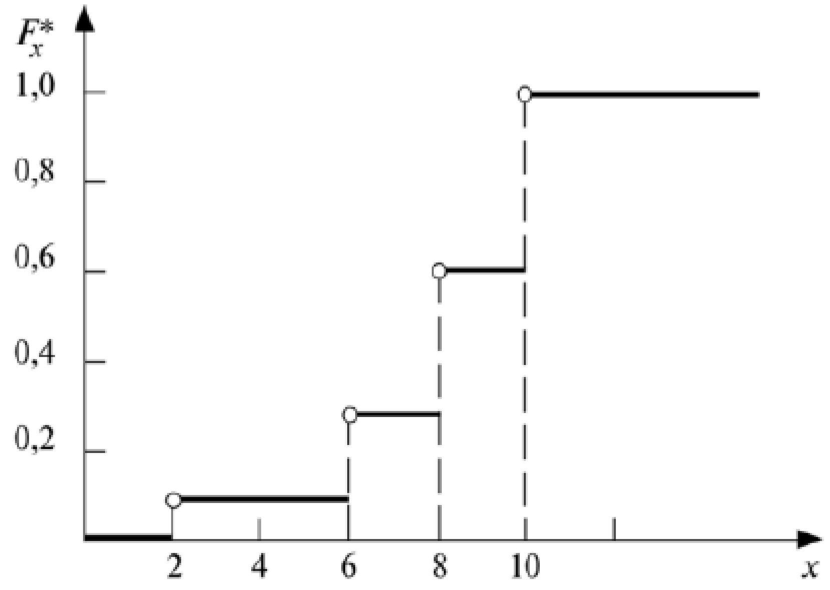


Рисунок 6

2) Определим выборочное среднее выборки по формуле (2):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{21}(2 \cdot 3 + 6 \cdot 4 + 8 \cdot 6 + 10 \cdot 8) = \frac{116}{21} \approx 7{,}52.$$

«Исправленную» дисперсию найдём, используя следующую формулу:

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

$$S^2 = \frac{1}{20} \cdot ((2 - 7{,}52)^2 \cdot 3 + (6 - 7{,}52)^2 \cdot 4 + (8 - 7{,}52)^2 \cdot 6 + (10 - 7{,}52)^2 \cdot 8)$$

$$= \frac{151{,}2384}{20} \approx 7{,}2 \Rightarrow S = \sqrt{7{,}2} \approx 2{,}683.$$

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

.

Не сгруппированные данные образуют дискретный вариационный ряд, содержащий нечётное число вариант $(n = 21 = 2 \cdot 10 + 1)$:

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

Значит, медиана равна

$Me = x_{10} = 8.$

Задача № 2 Вариационный ряд. Полигон частот.

- 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:

Объём 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 следующей таблицы:

x_i	12	13	14	16	17	18	19	20	Сумма
n_i	1	2	3	3	2	1	2	1	15
p_i^*	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	1

Полигон относительных частот изображён на рисунке 7.

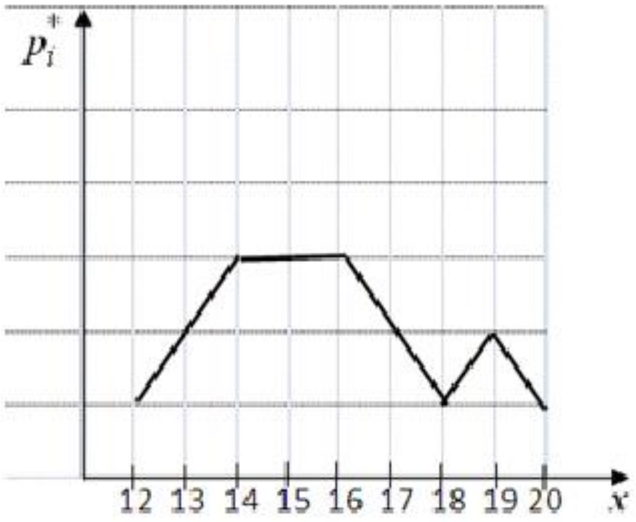


Рисунок 7

Находим выборочное среднее по формуле (2):

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

Для вычисления выборочной дисперсии используем формулу (5):

$$D_B = \frac{1}{15}\cdot((12-15{,}9)^2 + 2\cdot(13-15{,}9)^2 + 3\cdot(14-15{,}9)^2 + \cdots$$
$$+ (20-15{,}9)^2) = \frac{85{,}75}{15} \approx 5{,}72.$$

«Исправленная» дисперсия и СКО:

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

Задача № 3 Интервальные оценки параметров распределения

- Найти выборочное среднее, моду, медиану и выборочное СКО выборки объёмом $n = 70$, распределение которой задано следующей таблицей:

Интервалы	0 – 1,02	1,02 – 2,04	2,04 – 3,06	3,06 – 4,08	4,08 – 5,1	Сумма
Частота	4	28	19	12	7	140

Построить гистограмму и полигон частот.

Решение:

Для построения гистограммы все частоты необходимо разделить на длину интервала, равную 1,02, и откладывать по оси ординат. По оси абсцисс отмечаются границы интервалов (рисунок 8).

Для построения полигона частот найдем середины интервалов и дополним исходную таблицу:

Интервалы $a_i - a_{i+1}$	0 – 1,02	1,02 – 2,04	2,04 – 3,06	3,06 – 4,08	4,08 – 5,1	Сумма
Середины интервалов a_i^*	0,51	1,53	2,55	3,57	4,59	
Частота m_i	4	28	19	12	7	140
Накопленная частота	4	32	51	63	70	

Ломаная линия (рисунок 8) будет соединять точки с координатами $\left(a_i^*; \frac{n_i}{h}\right)$.

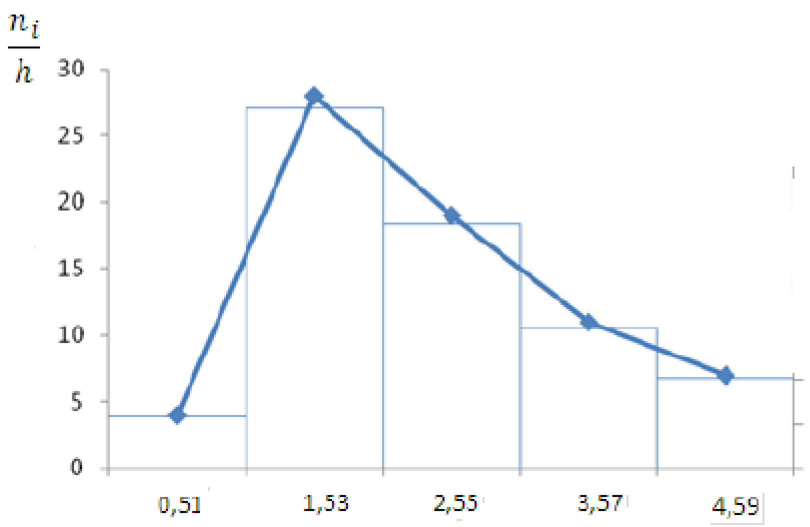


Рисунок 8

Для расчёта выборочного среднего и выборочного СКО составляем вариационный ряд, принимая в качестве вариант середины соответствующих интервалов:

a_i^*	0,51	1,53	2,55	3,57	4,59	**Σ**
m_i	4	28	19	12	7	**70**
$a_i^* \cdot m_i$	2,04	42,84	48,45	42,84	32,13	**168,3**
$(a_i^*)^2 \cdot m_i$	1,0404	65,5452	123,5475	152,9388	147,4767`	**490,5486**

Таким образом:

$$\bar{x} = \frac{168{,}3}{70} = 2{,}404; \;\; \bar{x}^2 = \frac{490{,}5486}{70} = 7{,}008$$

$$D_B = \bar{x}^2 - (\bar{x})^2 = 7{,}008 - 2{,}404^2 = 1{,}229; \quad \sigma_B = \sqrt{D_B} = \sqrt{1{,}229} \approx 1{,}11.$$

Так как наибольшая частота

$$m_i = 28$$

отвечает интервалу 1,02 — 2,04, то

$$x_i = 1{,}02; \; m_{i-1} = 4; \; m_{i+1} = 19; \; h = 1{,}02.$$

Мода (согласно формуле (3)) равна:

$$Mo = 1{,}02 + 1{,}02 \cdot \frac{28-4}{(28-4)+(28-19)} \approx 1{,}76.$$

Определим номер медианного интервала. Так как $32 < \frac{70}{2} < 51$, то номер медианного интервала равен 3, а сам интервал — 2,04 — 3,06. Тогда, по формуле (4), получаем:

$$Me = 2{,}04 + 1{,}02 \cdot \frac{35-32}{19} = 2{,}202.$$

Задача № 4 Начальные и центральные моменты

Дан статистический ряд признака X:

x_i	3,8	3,9	4	4,1	4,2	4,3	4,4	4,5	Σ
n_i	2	2	5	7	6	4	3	1	30

Найти начальные и центральные моменты первых четырёх порядков признака X, а также определить асимметрию и эксцесс.

Решение:
Вычисления проводим по формулам (8) для $\tilde{\nu}_k$ и по формулам (10) для $\tilde{\mu}_k$.

Начальные моменты:

$$\tilde{\nu}_1 = \frac{1}{30}(2\cdot 3{,}9 + 2\cdot 4{,}1 + 5\cdot 4 + 7\cdot 4{,}1 + 6\cdot 4{,}2 + 4\cdot 4{,}3 + 3\cdot 4{,}4 + 1\cdot 4{,}5) = \frac{124{,}2}{30} = 4{,}14.$$

$$\tilde{\nu}_2 = \frac{1}{30}(2\cdot 3{,}9^2 + 2\cdot 4{,}1^2 + 5\cdot 4^2 + 7\cdot 4{,}1^2 + 6\cdot 4{,}2^2 + 4\cdot 4{,}3^2 + 3\cdot 4{,}4^2 + 1\cdot 4{,}5^2) = \frac{515{,}1}{30} = 17{,}17.$$

$$\tilde{\nu}_3 = \frac{1}{30}(2\cdot 3{,}9^3 + 2\cdot 4{,}1^3 + 5\cdot 4^3 + 7\cdot 4{,}1^3 + 6\cdot 4{,}2^3 + 4\cdot 4{,}3^3 + 3\cdot 4{,}4^3 + 1\cdot 4{,}5^3) = \frac{2140{,}1}{30} = 71{,}34.$$

$$\tilde{\nu}_4 = \frac{1}{30}(2\cdot 3{,}9^4 + 2\cdot 4{,}1^4 + 5\cdot 4^4 + 7\cdot 4{,}1^4 + 6\cdot 4{,}2^4 + 4\cdot 4{,}3^4 + 3\cdot 4{,}4^4 + 1\cdot 4{,}5^4) = \frac{8906{,}8}{30} = 296{,}89.$$

Центральные моменты

$$\tilde{\mu}_1 = 0$$
$$\tilde{\mu}_2 = \tilde{\nu}_2 - \tilde{\nu}_1^2 = 17{,}17 - 4{,}14^2 = 0{,}0304;$$
$$\tilde{\mu}_3 = \tilde{\nu}_3 - 3\tilde{\nu}_1\tilde{\nu}_2 + 2\tilde{\nu}_1^3 = 71{,}34 - 3\cdot 4{,}14\cdot 17{,}17 + 2\cdot 4{,}14^3 = -70{,}95;$$
$$\tilde{\mu}_4 = \nu_4 - 4\tilde{\nu}_1\tilde{\nu}_3 + 6\tilde{\nu}_2\tilde{\nu}_1^2 - 3\tilde{\nu}_1^4 = 296{,}89 - 4\cdot 4{,}14\cdot 71{,}34 + 6\cdot 17{,}17\cdot 4{,}14^2 - 3\cdot 4{,}14^4 = -11{,}34$$

Тогда, так как

$$\sigma_B = \sqrt{D_B} = \sqrt{\tilde{\mu}_2} = 0{,}174,$$

то

$$\tilde{A}_s = \frac{\tilde{\mu}_3}{\sigma_B^3} = \frac{-70{,}95}{0{,}174^3} = -13{,}5; \quad E_x = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3 = \frac{-11{,}34}{0{,}174^4} - 3 = -2{,}15.$$

Задача № 5 Доверительный интервал. Надёжность

Предельная нагрузка для выборки из 50 стальных стержней характеризуется следующим рядом:

x_i	11	13	14	16	17
n_i	4	5	30	7	4

Считая распределение предельной нагрузки X нормальным, построить доверительные интервалы для оценки с надёжностью $\gamma = 0{,}99$ средней предельной нагрузки и СКО предельной нагрузки стальных стержней партии, из которой произведена выборка.

Решение:

Вычислим выборочное среднее и исправленное СКО соответственно по формулам

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \frac{11\cdot 4 + 13\cdot 5 + 14\cdot 30 + 16\cdot 7 + 17\cdot 4}{50} = \frac{709}{50} = 14{,}18$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 n_i}{n-1}} = \sqrt{\frac{(11-14{,}18)^2\cdot 3 + \dots + (17-14{,}17)^2\cdot 4}{49}} = \sqrt{\frac{103{,}38}{49}} = 1{,}45$$

По таблице (см. приложение 3) найдём

$t_\gamma = t(0{,}99; 50) = 2{,}679.$

Точность оценки:

$\delta = t_\gamma \cdot \frac{S}{\sqrt{n}} = 2{,}679 \cdot \frac{1{,}45}{7{,}071} = 0{,}55.$

Доверительный интервал для средней предельной нагрузки найдём по формуле (13):

$$\bar{x} - t_\gamma \cdot \frac{S}{\sqrt{n}} = 14{,}18 - 0{,}55 = 13{,}63; \quad \bar{x} + t_\gamma \cdot \frac{S}{\sqrt{n}} = 14{,}18 + 0{,}55 = 14{,}73.$$

$$13{,}63 < a < 14{,}73..$$

Доверительный интервал для СКО предельной нагрузки будем искать по формуле

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

так как

$q = q(0{,}99; 50) = 0{,}3$

$1{,}45\cdot(1-0{,}3) < \sigma < 1{,}45\cdot(1+0{,}3) \Rightarrow 1{,}015 < \sigma < 1{,}885$

Задача № 6 Статистическое распределение выборки

В результате эксперимента получены данные, представленные в виде статистического ряда:

38 60 41 51 33 42 45 21 53 60 60 52 47 46 49 49 14 57 54 59 30 40 50 59 30 61 56 58 42 54 44 42 32 45 60 43 41 58 48 72 48 47 39 28 47 35 65 61 77 67.

Требуется:

1)Записать значения результатов эксперимента в виде вариационного ряда.

2)Представить данную выборку в виде интервального статистического ряда.

3)Найти числовые характеристики выборки:

$\bar{x},\ D_B,\ S^2, S.$

4) Определить доверительные интервалы неизвестного математического ожидания и неизвестного среднего квадратического отклонения. Предполагается, что генеральная совокупность имеет нормальное распределение. Доверительную вероятность принять равной 0,95.

Решение:

4) Расположим значения результатов эксперимента в порядке возрастания, т.е. записываем вариационный ряд:

14 21 28 30 30 32 33 35 38 39 40 41 41 42 42 42 43 44 45 45 46 4747 47 48 48 49 49 50 51 52 53 54 54 56 57 58 58 59 59 60 60 60 60 61 61 65 67 72 77.

2) Объём выборки $n = 50$. Наибольшая варианта — 77, наименьшая — 14. Найдём длину интервала:

$$h = \frac{77 - 14}{1 + 3{,}322 \cdot lg50} = \frac{63}{6{,}644} \approx 9.48.$$

Выбираем длину интервала 9. Интервальный статистический ряд примет вид:

Границы интервалов	$[14; 23[$	$[23; 32[$	$[32; 41[$	$[41; 50[$	$[50; 59[$	$[59; 68[$	$[68; 77]$
m_i	2	3	6	17	10	10	2

3) Для вычисления числовых характеристик составляем вариационный ряд, принимая в качестве вариант середины соответствующих интервалов:

Границы интервалов	[14; 23[	[23; 32[	[32; 41[	[41; 50[	[50; 59[	[59; 68[	[68; 77[	$\sum$
Середины интервалов a_i^*	18,5	27,5	36,5	45,5	54,5	63,5	72,5	
m_i	2	3	6	17	10	10	2	**50**
$a_i^* \cdot m_i$	37	82,5	219	773,5	545	635	145	**2437**
$(a_i^*)^2 \cdot m_i$	684,5	2268,75	7993,5	35194,25	29702,5	40322,5	10512,5	**126678,5**

Таким образом:

$$\bar{x} = \frac{2437}{50} = 48{,}74; \quad \bar{x}^2 = \frac{126678{,}5}{50} = 2533{,}57.$$

$$D_B = \bar{x}^2 - (\bar{x})^2 = 2533{,}57 - 48{,}74^2 = 158{,}9824;$$

$$S^2 = \frac{50}{49} \cdot 158{,}9824 = 161{,}2 \Rightarrow S = \sqrt{161{,}2} = 12{,}7.$$

4) Доверительный интервал для оценки математического ожидания нормально распределённой случайной величины найдём по формуле:

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

Из приложения 3 для $\gamma = 0{,}95$ находим $t_\gamma = 2{,}009$. Далее

$t_\gamma \cdot \frac{s}{\sqrt{n}} =$

$\frac{2{,}009 \cdot 12{,}7}{7{,}071} = 3{,}608$ и $45{,}132 < a < 52{,}348$, т.е. $a \in (45{,}132; 52{,}348)$.

Доверительный интервал для оценки а нормального распределения по несмещённой оценке S определяется из неравенства

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

где величина

$$q = q(\gamma, n)$$

определяется из таблицы (приложение 4).

Имеем

$n = 50;\ S = 12{,}7.$

При

$\gamma = 0{,}95$ и $n = 50$

в таблице приложения находим

$q(0{,}95; 50) = 0{,}21.$

Следовательно,

$S(1-q) = 12{,}7 \cdot 0{,}79 = 10{,}033;\ S(1+q) = 12{,}7 \cdot 1{,}21 = 15{,}367.$

Значит,

$\sigma \in (10{,}033; 15{,}367).$

Задача № 7 Распределение Пирсона

- 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

$$\chi^2 = \sum_{i=1}^{l} \frac{(n_i - n_i')^2}{n_i'} = \frac{(8-5)^2}{5} + \frac{(10-15)^2}{15} + \frac{(18-16)^2}{16} + \cdots + \frac{(9-7)^2}{7} \approx 5$$

В таблице критических точек χ^2 (приложение 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.

Задача № 8 Нормальное распределение случайной величины

- 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 $l = 1 + 3{,}322 \cdot \lg n$.

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

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

Решение:

Подсчитаем количество интервалов разбиения:

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

Из ряда видно, что

$$x_{min} = 0{,}6;\ x_{max} = 4{,}2,$$

поэтому

$$\frac{x_{max} - x_{min}}{8} = \frac{4{,}2 - 0{,}6}{8} \approx 0{,}5.$$

Границы интервалов будут:

$$x_0 = 0{,}6;\ x_1 = 0{,}6 + 0{,}5 = 1{,}1;\ x_2 = 1{,}1 + 0{,}5 = 1{,}6;\ x_3 = 1{,}6 + 0{,}5 = 2{,}1;\ x_4 = 2{,}1 + 0{,}5 = 2{,}6;\ x_5 = 2{,}6 + 0{,}5 = 3{,}1;\ x_6 = 3{,}1 + 0{,}5 = 3{,}6;\ x_7 = 3{,}1 + 0{,}5 = 4{,}1;\ x_8 = 4{,}1 + 0{,}5 = 4{,}6.$$

Частота n_i— интервала $[x_i; x_{i+1}[\ (i = \overline{0{,}7})$ подсчитывается с помощью ряда как число наблюдений, попавших в интервал. Так в первый $(i = 0)$ интервал) 0,6; 1,11 попало 7 значений, во второй [1,1; 1,6[— 14 значений. Сведём полученные данные в таблицу:

$x_i - x_{i+1}$	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:

$$\bar{x} = \frac{\sum x_i^* n_i}{n} = \frac{1}{100}(0{,}85 \cdot 7 + 1{,}35 \cdot 14 + 1{,}85 \cdot 16 + \cdots + 4{,}35 \cdot 4) = \frac{248{,}5}{100} = 2{,}485$$

$$\sigma_B = \sqrt{\frac{\sum (x_i^* - \bar{x})^2 n_i}{n}} = \sqrt{\frac{1}{100}((0{,}85 - 2{,}485)^2 \cdot 7 + (1{,}35 - 2{,}485)^2 \cdot 14 + \cdots + (4{,}35 - 2{,}485)^2 \cdot 4)} = \sqrt{0{,}89} \approx 0{,}95.$$

Найдём теоретические вероятности p_i по формуле

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

где $\Phi(x)$ — функция Лапласа, значения которой даются в приложении 3. Результаты вычислений сведём в таблицу:

i	x_i	x_{i+1}	$x_i - \bar{x}$	$x_{i+1} - \bar{x}$	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	$n_i' = n \cdot p_i$
1	0,6	1,1	–	-1,385	–	-	-0,5	-0,4279	0,0721	7,21
						1,46				
2	1,1	1,6	-1,385	-0,885	-1,46	-0,94	-0,4279	-0,3264	0,1015	10,15
3	1,6	2,1	-0,885	-0,385	-0,94	-0,41	-0,3264	-0,1591	0,1673	16,73
4	2,1	2,6	-0,385	0,115	-0,41	0,12	-0,1591	0,0478	0,2069	20,69
5	2,6	3,1	0,115	0,615	0,12	0,65	0,0478	0,2422	0,1944	19,44
6	3,1	3,6	0,615	1,115	0,65	1,18	0,2422	0,381	0,1388	13,88
7	3,6	4,1	1,115	1,615	1,18	1,71	0,381	0,4564	0,0754	7,54
8	4,1	4,6	1,615	–	1,71	–	0,4564	0,5	0,0436	4,36

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим следующую расчетную таблицу:

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, уровню значимости $\alpha = 0{,}05$ и числу 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.

Задача № 9 Коэффициент корреляции. Остаточная дисперсия

- По заданной таблице зависимости признаков X и Y

X	0	1.7	4.7	7.5	8.5
Y	-3.2	-2.7	-1.0	0.2	1.8

вычислить выборочный коэффициент корреляции и остаточную дисперсию. Записать уравнения прямой регрессии X на Y. Построить корреляционное поле и линию регрессии на корреляционном поле.

Решение:

Вычислим основные выборочные характеристики: Выборочные средние:

$$\bar{x} = \frac{1}{5}(0 + 1{,}7 + 4{,}7 + 7{,}5 + 8{,}5) = \frac{22{,}4}{5} = 4{,}48;$$
$$\bar{y} = \frac{1}{5}(-3{,}2 - 2{,}7 - 1 + 0{,}2 + 1{,}8) = \frac{-4{,}9}{5} = -0{,}98.$$

Найдем оценки для средних квадратичных отклонений и корреляционного момента, для чего составим следующую вспомогательную таблицу:

X	0	1,7	4,7	7,5	8,5	Σ
Y	-3,2	-2,7	-1	0,2	1,8	
$(x_i - \bar{x})^2$	20,07	7,7284	0,0484	9,1204	16,16	**53,128**
$(y_i - \bar{y})^2$	4,9284	2,9584	0,0004	1,3924	7,7284	**17,008**
$(x_i - \bar{x})(y_i - \bar{y})$	9,9456	4,7816	-0,004	3,5636	11,176	**29,462**

$$s_x = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2} = \sqrt{\frac{53{,}128}{4}} \approx 3{,}6444.$$
$$s_y = \sqrt{\frac{1}{n-1}\sum_i (y_i - \bar{y})^2} = \sqrt{\frac{17{,}008}{4}} \approx 2{,}062.$$
$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{29{,}462}{4} = 7{,}3655.$$

Согласно формуле (15):

$$r_B = \frac{s_{xy}}{s_x \cdot s_y} = \frac{7{,}3655}{3{,}6444 \cdot 2{,}062} \approx 0{,}98.$$

Найдем методом наименьших квадратов эмпирическую формулу вида $\bar{y}_x = a + bx.$. Составим систему нормальных уравнений (17) для определения параметров линейной регрессии. Так как

$$\bar{x} = 4{,}48;\ \bar{y} = -0{,}98;\ \overline{x^2} = \frac{1}{5}(0 + 1{,}7^2 + 4{,}7^2 + 7{,}5^2 + 8{,}5^2) = \frac{153{,}48}{5} = 30{,}696;$$
$$\overline{xy} = \frac{1}{5}(0 - 1{,}7 \cdot 2{,}7 - 4{,}7 + 7{,}5 \cdot 0{,}2 + 8{,}5 \cdot 1{,}8) = \frac{7{,}51}{5} = 1{,}502,$$

то:

$$\begin{cases} a + 4{,}48b = -0{,}98 \\ 4{,}48a + 30{,}696b = 1{,}502 \end{cases} \Rightarrow b = \frac{1{,}502 + 4{,}48 \cdot 0{,}98}{30{,}696 - 4{,}48^2} = 0{,}5545;$$
$$a = -0{,}98 - 0{,}5545 \cdot 4{,}48 = -3{,}464.$$

Уравнение регрессии Y на X имеет вид:

$$\bar{y}_x = -3{,}464 + 0{,}5545x.$$

Остаточная дисперсия:

$$D_{\text{ост}} = s_y^2(1 - r_B^2) = 2{,}062^2 \cdot (1 - 0{,}98^2) = 0{,}168.$$

Корреляционное поле и линия регрессии на корреляционном поле изображены на рисунке 9.

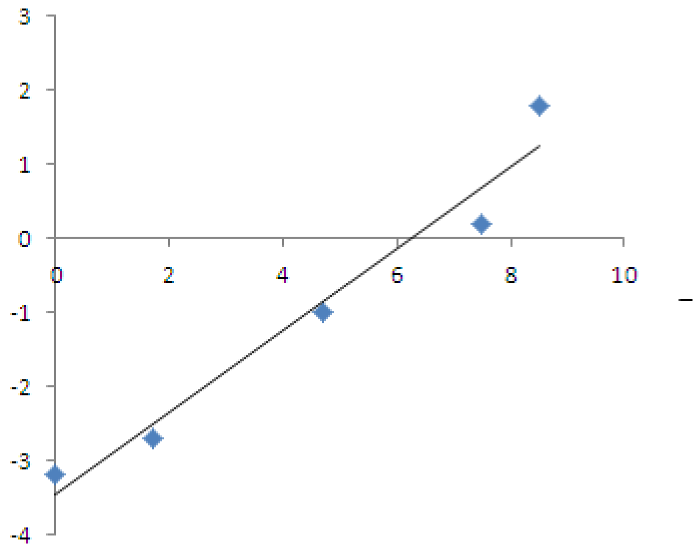


Рисунок 9

Задача № 10 Корреляционное поле

- Таблица значений признака Y при данных значениях признака X имеет вид:

Y	X				
	5	10	15	20	n_y
10	2	-	-	-	2
20	5	4	1	-	10
30	3	8	6	3	20
40	-	3	6	6	15
50	-	-	2	1	3
n_x	10	15	15	10	$n = 50$

Построить корреляционное поле. Найти выборочный коэффициент корреляции, оценить его значимость. Записать уравнения прямой линии регрессии Y на X.

Решение:

Корреляционное поле данной двумерной выборки приведено на рисунке 10.

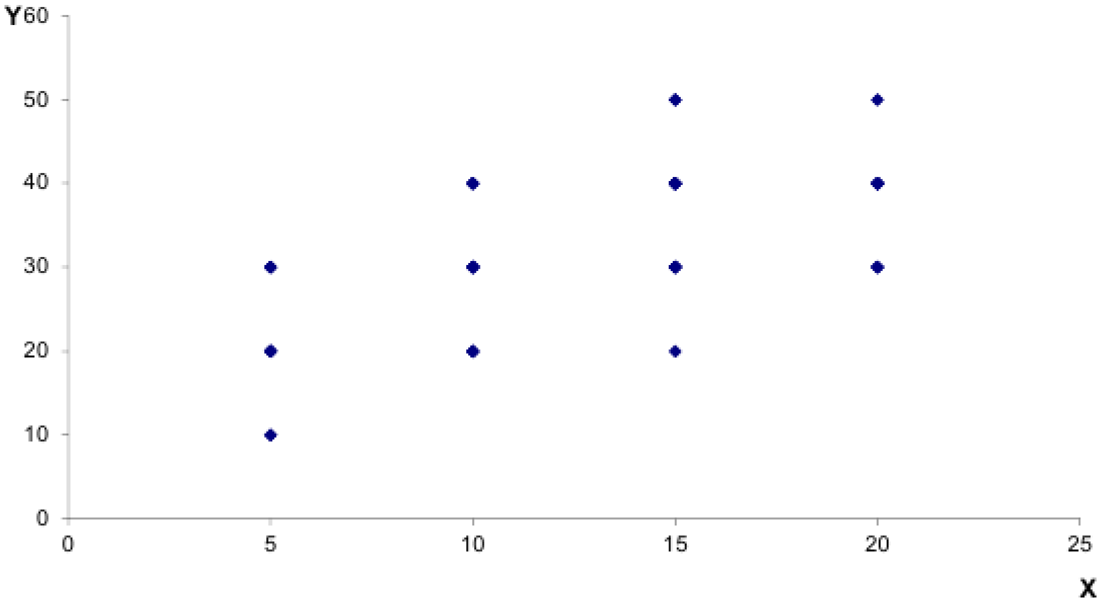


Рисунок 10

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

Для вычисления выборочных числовых характеристик составляем следующую расчётную таблицу:

Y	**X**					$n_{yi}y_j$	$\sum_i n_{ij}x_i$	$y_j^2 n_{yj}$	$y_j \sum_i n_{ij}x_i$
	5	**10**	**15**	**20**	n_y				
10	2	-	-	-	2	20	10	200	100
20	5	4	1	-	10	200	80	4000	1600
30	3	8	6	3	20	600	245	18000	7350
40	-	3	6	6	15	600	240	24000	9600
50	-	-	2	1	3	150	50	7500	2500
n_x	10	15	15	10	$\boldsymbol{n = 50}$	**1570**	**625**	**53700**	**21150**
$n_{xi}x_i$	50	150	225	200	**625**				
$\sum_j n_{ij}y_j$	210	440	540	380	**1570**				
$n_{xi}x_i^2$	250	1500	3375	4000	**9125**				
$x_i \sum_j n_{ij}y_j$	1050	4400	8100	7600	**21150**				

Замечание. Строка

$$\sum_j n_{ij} y_j$$

получается следующим образом:

$210 = 10 \cdot 2 + 20 \cdot 5 + 30 \cdot 3; \quad 440 = 20 \cdot 4 + 30 \cdot 8 + 40 \cdot 3;$ … и т. д;

Столбец

$\sum_i n_{ij} x_i$: $10 = 5 \cdot 2; 30 = 5 \cdot 5 + 10 \cdot 4 + 15 \cdot 1; \ldots$

Вычислим выборочные средние

$\bar{x}$ и $\bar{y}$ $\left(i = \overline{1,4};\ j = \overline{1,5}\right)$:

$$\bar{x} = \frac{1}{n}\sum_i n_{xi} x_i = \frac{625}{50} = 12{,}5;\ \bar{y} = \frac{1}{n}\sum_j n_{xj} x_j = \frac{1570}{50} = 31{,}4.$$

«Исправленные» дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1}\left(\sum_i n_{xi} x_i^2 - \frac{1}{n}\left(\sum_i n_{xi}\, x_i\right)^2\right) = \frac{1}{49}\left(9125 - \frac{1}{50} \cdot 625^2\right) = 26{,}786.$$

$$s_y^2 = \frac{1}{n-1}\left(\sum_j n_{yj} y_j^2 - \frac{1}{n}\left(\sum_j n_{yj}\, y_j\right)^2\right) = \frac{1}{49}\left(53700 - \frac{1}{50} \cdot 1570^2\right) = 89{,}837.$$

Оценку корреляционного момента вычисляем по формуле:

$$s_{xy} = \frac{1}{n-1}\left(\sum_i \sum_j n_{ij} x_i y_j - \frac{1}{n}\left(\sum_i n_{xi}\, x_i\right) \cdot \left(\sum_j n_{yj}\, y_j\right)\right) = \frac{1}{49}\left(21150 - \frac{1}{50} 625 \cdot 1570\right) = 31{,}1.$$

Рассчитав все нужные величины, можно вычислить выборочный коэффициент корреляции:

$$r_B = \frac{31{,}1}{\sqrt{26{,}786} \cdot \sqrt{89{,}837}} = \frac{31{,}1}{5{,}176 \cdot 9{,}478} \approx 0{,}63.$$

Для оценки значимости выборочного коэффициента корреляции вычислим наблюдаемое значение критерия, воспользовавшись формулой (16):

$$T_{\text{набл}} = r_B \cdot \sqrt{\frac{n-2}{1-r_B^2}} = 0{,}63 \cdot \sqrt{\frac{48}{1-0{,}63^2}} = 5{,}62.$$

Затем по таблице критических точек распределения Стьюдента, по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = n - 2 = 48$ найдем критическую точку t_{kp} для двухсторонней критической области:

$t_{kp}(0{,}05; 48) = 2{,}01$

Сравнивая $T_{\text{набл}}$ и t_{kp}, получим, что $|T_{\text{набл}}| > t_{kp}$, следовательно, величины X. Y 21оррелированны.

Оценкой теоретической линии регрессии является эмпирическая линия регрессии, уравнение которой имеет вид

$\bar{y}_x - \bar{y} = b(x - \bar{x})$, где $b = r_B \frac{s_y}{s_x}$.

Тогда

$\bar{y}_x - 31{,}4 = 0{,}63 \cdot \frac{9{,}478}{5{,}176} \cdot (x - 12{,}5)$ или $\bar{y}_x = 1.$

Задача № 11 Эмпирическая функция распределения. Кумулята

Из генеральной совокупности извлечена выборка, представленная в виде статистического ряда:

x_i	4	6	7	8	10	11	12
n_i	3	7	10	11	9	6	4

Требуется:

1)вычислить выборочное среднее $\bar{x}$, выборочную дисперсию D_B, исправленную выборочную дисперсию S^2 и среднее квадратичное отклонение s;

2)найти размах варьирования; моду и медиану;

3)построить полигон частот и эмпирическую функцию распределения;

4)проверить, согласуются ли выборочные данные с гипотезой о нормальном распределении случайной величины X графически и с помощью критерия согласия Пирсона при уровне значимости $\alpha = 0{,}05$, представив данную выборку в виде интервального ряда. Количество интервалов рассчитать по формуле Стерджеса $l = 1 + 3{,}322 \lg n$;

5)найти с доверительной вероятностью $\gamma = 0{,}95\ (\gamma = 0{,}99)$ доверительный интервал для математического ожидания, а также доверительный интервал для $\sigma(x)$.

Решение:

4) Объем выборки равен

$$n = \sum n_i = 3 + 7 + 10 + 11 + 9 + 6 + 4 = 50.$$

Выборочное среднее определим по формуле:

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{50}(4 \cdot 3 + 6 \cdot 7 + 7 \cdot 10 + 8 \cdot 11 + 10 \cdot 9 + 11 \cdot 6 + 12 \cdot 4)$$
$$= \frac{416}{50} = 8{,}32$$

Для нахождения выборочной дисперсии составим следующую вспомогательную таблицу:

x_i	4	6	7	8	1 0	1 1	1 2	$\sum$
n_i	3	7	1 0	1 1	9	6	4	**5 0**
$(x_i - \bar{x})^2 \cdot n_i$	5 5,987	3 7,677	1 7,424	1 ,1264	2 5,402	4 3,094	5 4,17	**2 34,88**

Тогда

$$D_B = \frac{1}{n}\sum(x_i - \bar{x})^2 \cdot n_i = \frac{1}{50} \cdot 234{,}88 = 4{,}697;\ S^2 = \frac{n}{n-1}D_B = \frac{50}{49} \cdot$$

$$4{,}6976 \approx 4{,}8$$

Исправленное среднее квадратичное отклонение будет

$$s = \sqrt{4{,}8} \approx 2{,}19.$$

2) Размах варьирования находится по формуле

$$R = \max x_i - \min x_i = 12 -$$

$4 = 8.$ Так как мода — это варианта, которой соответствует наибольшая частота, то

$4 = 8.$ $Mo = 8.$

Не сгруппированные данные образуют дискретный вариационный ряд, содержащий чётное число вариант $(n = 50 = 2 \cdot 25)$, поэтому $M_e = \frac{x_{25}+x_{26}}{2} = \frac{8+8}{2} = 8$

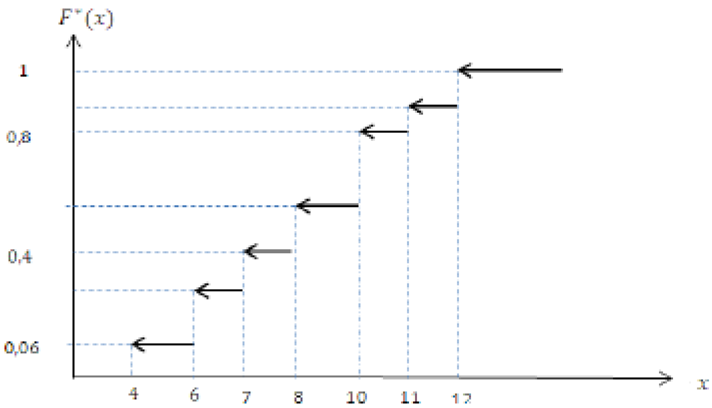
3) Согласно определению эмпирической функции распределения ее значение при любом x равно $F^*(x) = \frac{n_x}{n}$, где n_x — количество элементов x_i выборки, меньших, чем x.

Например, при $x = 3$ имеем $n_x = 0$, $F^*(x) = 0$; при $x = 5$ $n_x = 3$ $F^*(x) = \frac{3}{50} = 0{,}06$; при $x = 6{,}2$ $n_x = 3 + 7 = 10$, $F^*(x) = \frac{10}{50} = 0{,}2$; при $x = 7{,}9$ $n_x = 3 + 7 + 10 = 20$, $F^*(x) = \frac{20}{50} = 0{,}4$; при $x = 9$ $F^*(x) = \frac{31}{50} = 0{,}6$; при $x = 10{,}6$ $F^*(x) = \frac{40}{50} = 0{,}8$; при $x = 11{,}2$ $F^*(x) = \frac{46}{50} = 0{,}92$.

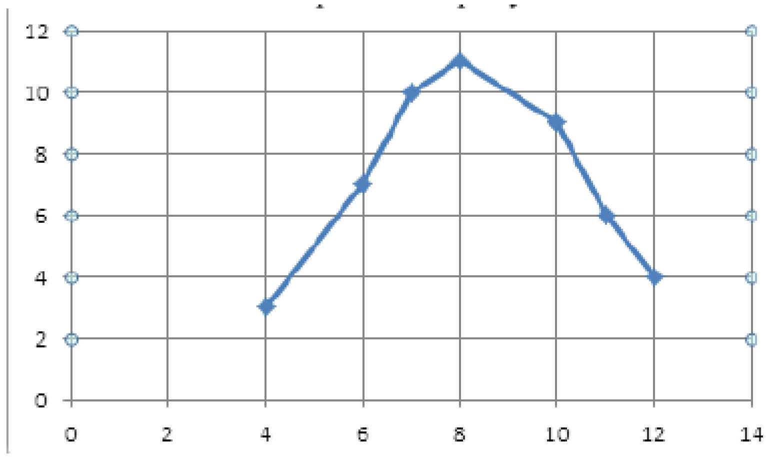
Тогда

$$F^*(x) = \begin{cases} 0 & x \le 4 \\ 0{,}06 & 4 < x \le 6 \\ 0{,}2 & 6 < x \le 7 \\ 0{,}4 & 7 < x \le 8 \\ 0{,}62 & 8 < x \le 10 \\ 0{,}8 & 10 < x \le 11 \\ 0{,}92 & 11 < x \le 12 \\ 1 & x > 12 \end{cases}$$

График эмпирической функции распределения:



Полигон частот изображен на рисунке:



4) Так как полигон частот по форме напоминает кривую Гаусса, то можно сделать предположение о том, что случайная величина X распределена по нормальному закону. Проверим данное утверждение по критерию Пирсона. Вычислим количество интервалов:

$$l = 1 + 3{,}322 \lg n = 1 + 3{,}322 \lg 50 = 6{,}644 \approx 7.$$

Длина интервала

$$h = \frac{R}{h} = \frac{8}{7} \approx 1{,}2.$$

Границы интервалов будут:

$$a_0 = 4,\ a_1 = 4 + 1{,}2 = 5{,}2,\ \ a_2 = 6{,}4,\ \ a_3 = 7{,}6,\ \ a_4 = 8{,}8,$$
$$a_5 = 10,\qquad a_6 = 11{,}2,\qquad a_7 = 12{,}4$$

Посчитаем число выборочных значений, попавших в каждый интервал. Частота m_i интервала $[a_i;\ a_{i+1}]$ $(i = \overline{0,6})$ подсчитывается с помощью ряда, как число наблюдений, попавших в интервал. Так, в первый $(i = 0)$ интервал [4; 5,2] попало 3 значения; во второй $(i = 1)$ — [5,2; 6,4] попало 7 значений. Аналогично получаем частоты 3-7 интервалов.

Полученные данные сведём в следующую таблицу:

$[a_i;\ a_{i+1}]$	4 – 5,2	5,2 – 6,4	6,4 – 7,6	7,6 – 8,8	8,8 – 10	10 – 11,2	11,2 – 12,4
m_i	3	7	10	11	9	6	4

Найдем теоретические вероятности p_i по формуле:

$$p_i = \Phi\left(\frac{a_{i+1} - \bar{x}}{s}\right) - \Phi\left(\frac{a_i - \bar{x}}{s}\right)$$

Результаты вычислений сведем в таблицу:

i	a_i	a_{i+1}	$\frac{a_i-\bar{x}}{s}$	$\frac{a_{i+1}-\bar{x}}{s}$	$\Phi\left(\frac{a_i-\bar{x}}{s}\right)$	$\Phi\left(\frac{a_{i+1}-\bar{x}}{s}\right)$	p_i	np_i
1	4	5,2	–	-1,42	-0,5	-0,4222	0,0778	3,89
2	5,2	6,4	-1,42	-0,88	-0,4222	-0,3106	0,1116	5,58
3	6,4	7,6	-0,88	-0,33	-0,3106	-0,1293	0,1813	9,065
4	7,6	8,8	-0,33	0,22	-0,1293	0,0871	0,2164	10,82
5	8,8	10	0,22	0,77	0,0871	0,2794	0,1923	9,615
6	10	11,2	0,77	1,32	0,2794	0,4066	0,1272	6,36
7	11,2	12,4	1,32	–	0,4066	0,5	0,0934	4,67

Так как ожидаемые (эмпирические) частоты первого и седьмого интервалов группировки не удовлетворяют условию $min(n_i) \geq 5$5, объединим эти интервалы (первый со вторым; а седьмой — с шестым).

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим

i	n_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
1	10	9,47	0,2809	0,029662
2	10	9,065	0,874225	0,09644
3	11	10,82	0,0324	0,002994
4	9	9,615	0,378225	0,039337
5	10	11,03	1,0609	0,096183
Σ	**50**	**50**		**0,264616**

По таблице критических точек распределения χ^2, уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = 5 - 3 = 2$ находим $\chi^2_{kp}(0{,}05; 2) = 6$. Так как

$$\chi^2_{\text{набл}} = 0{,}264616 < 6,$$

то гипотеза о нормальном распределении принимается.

5) Доверительный интервал для математического ожидания найдём по формуле

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

Значение t_γ определим по таблице для доверительной вероятности

$\gamma = 0{,}95$ $(\gamma = 0{,}99)$

и объёму выборки

$n = 50$:

$t_\gamma(0{,}95; 50) = 2{,}009; \quad t_\gamma(0{,}99; 50) = 2{,}679$

Тогда доверительный интервал имеет вид:

для $\gamma = 0{,}95$: $\quad 8{,}32 - 2{,}009 \cdot \frac{2{,}19}{\sqrt{50}} < a < 8{,}32 + 2{,}009 \cdot \frac{2{,}19}{\sqrt{50}} \Rightarrow 7{,}7 < a < 8{,}9$

для $\gamma = 0{,}99$: $\quad 8{,}32 + 2{,}679 \cdot \frac{2{,}19}{\sqrt{50}} < a < 8{,}32 + 2{,}679 \cdot \frac{2{,}19}{\sqrt{50}} \Rightarrow$
$7{,}49 < a < 9{,}15.$

Задача № 12 Коэффициент корреляции. Остаточная дисперсия

По заданной таблице зависимости признаков X и Y:

1}Вычислить выборочный коэффициент корреляции; проверить его на значимость, приняв $\alpha = 0{,}05$.

2)Методом наименьших квадратов выровнять зависимость Y от X по прямой $y = a + bx$.

3)Вычислить остаточную дисперсию, сделать вывод.

4)Построить корреляционное поле и линию регрессии на корреляционном поле.

X	-2	-1	0	1	2	3
Y	5,6	5	4,3	4	3,6	3

Решение:

Найдём выборочные средние х, у, а также оценки для средних квадратичсских отклонений и корреляционного момента, для чего составим следующую вспомогательную таблицу:

i	**1**	**2**	**3**	**4**	**5**	**6**	$\sum$
X	-2	-1	0	1	2	3	**3**
Y	5,6	5	4,3	4	3,6	3	**25,5**
$(x_i - \bar{x})^2$	6,25	2,25	0,25	0,25	2,25	6,25	**17,5**
$(y_i - \bar{y})^2$	1,8225	25	18,49	16	12,96	9	**83,2725**
$(x_i - \bar{x})(y_i - \bar{y})$	-3,375	-1,125	-0,025	-0,125	-0,975	-3,125	**-8,75**

Здесь

$\bar{x} = \frac{3}{6} = 0{,}5;\ \bar{y} = \frac{25{,}5}{6} = 4{,}25.$

Тогда

$$s_x = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2} = \sqrt{\frac{17{,}5}{5}} \approx 1{,}87$$

$$s_y = \sqrt{\frac{1}{n-1}\sum_i (y_i - \bar{y})^2} = \sqrt{\frac{4{,}435}{5}} \approx 0{,}94$$

$$s_{xy} = \frac{1}{n-1}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{-8{,}75}{5} = -1{,}75$$

Выборочное значение коэффициента корреляции:

$$r_B = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-1{,}75}{1{,}87 \cdot 0{,}94} \approx -0{,}996$$

Проверим значимость полученного выборочного коэффициента корреляции. Найдём наблюдаемое значение критерия:

$$T_{\text{набл}} = r_B \cdot \sqrt{\frac{n-2}{1-r_B^2}} = -0{,}996 \cdot \sqrt{\frac{4}{1-0{,}996^2}} = -22{,}4$$

По таблице критических точек распределения Стьюдента, по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = n - 2 = 4$ находим критическую точку двусторонней критической области $t_{kp}(0{,}05; 4) = 2{,}78.$.

Так как $|T_{\text{набл}}| > t_{kp}$, то отвергаем гипотезу о равенстве нулю генерального коэффициента корреляции, значит X и Y-коррелированы.

i	x_i	y_i	x_i^2	$x_i \cdot y_i$
1	-2	5,6	4	-11,2
2	-1	5	1	-5
3	0	4,3	0	0
4	1	4	1	4
5	2	3,6	4	7,2
6	3	3	9	9
$\sum$	**3**	**25,5**	**19**	**4**

Запишем нормальную систему уравнений. Так как

$\bar{x} = 0{,}5;\ \bar{y} = 4{,}25;$

$\overline{x^2} = \frac{19}{6} \approx 3{,}17;\ \overline{xy} = \frac{4}{6} \approx 0{,}67,$

то

$$\begin{cases} a + 0{,}5b = 4{,}25 \\ 0{,}5a + 3{,}17b = 0{,}67 \end{cases}$$

Решая систему по формулам Крамера, получим:

$$a = \frac{\begin{vmatrix} 4{,}25 & 0{,}5 \\ 0{,}67 & 3{,}17 \end{vmatrix}}{\begin{vmatrix} 1 & 0{,}5 \\ 0{,}5 & 3{,}17 \end{vmatrix}} = \frac{4{,}25 \cdot 3{,}17 - 0{,}5 \cdot 0{,}67}{3{,}17 - 0{,}5 \cdot 0{,}5} = \frac{13{,}1375}{2{,}92} \approx 4{,}5;$$

$$b = \frac{\begin{vmatrix} 1 & 4{,}25 \\ 0{,}5 & 0{,}67 \end{vmatrix}}{\begin{vmatrix} 1 & 0{,}5 \\ 0{,}5 & 3{,}17 \end{vmatrix}} = \frac{0{,}67 - 0{,}5 \cdot 4{,}25}{3{,}17 - 0{,}5 \cdot 0{,}5} = \frac{-1{,}455}{2{,}92} \approx -0{,}5$$

Следовательно, зависимость между величинами X и Y выражается приближённой формулой

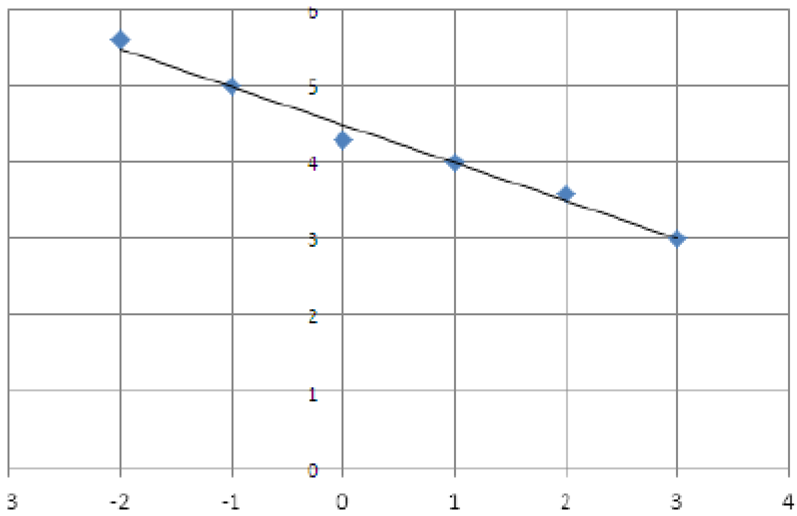
$\bar{y}_x = 4{,}5 - 0{,}5x.$

3) Остаточная дисперсия:

$D_{\text{ост}} = s_v^2(1 - r_B^2) = 4{,}08^2 \cdot (1 - 0{,}996^2) = 0{,}133$

То сеть величина ошибки, которая возникает при замене Y линейной функцией, невелика можно сделать вывод, что между величинами X и Y существует приближённая линейная зависимость.

4) Корреляционное поле и линия регрессии на корреляционном поле представлены на следующем рисунке:



6
5
4
3
2
1
0
3
-2
-1
0
1
2
3
4

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

Основные источники:

1. Теория вероятностей и математическая статистика: учебник для студ. учреждений сред. проф. образования /М. С. Спирина, П. А. Спирин. – 8-е изд., стер. - М.: Издательский центр «Академия», 2017. – 352 с.
2. Теория вероятностей и математическая статистика. Сборник задач: учеб. пособие для студ. учреждений сред. проф. образования / М. С. Спирина, П. А. Спирин. – 2-е изд., стер. – М.: Издательский центр «Академия», 2018. – 192 с.

Дополнительные источники:

1. Белько И.В., Свирид Г.П. Теория вероятностей и математическая статистика. Примеры и задачи. - Минск: Новое знание 2007.

2. Зубков А.М. Севостьянов Б.А. и др. Сборник задач по теории вероятности. – СПб.: Лань, 2009.

3. Кочетков Е.С., Смерчинская С.О. Теория вероятностей в задачах и упражнениях. - М.: Форум, 2008.

4. Юсупов Р.А. Теория вероятностей и математическая статистика. Учебное пособие для студентов вузов – Астрахань: АГТУ, 2000.