0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра высшей математики

УТВЕРЖДАЮ:
Проректор по учебной работе
_________ О.Г. Локтионова
« 30 » _______________ 2021г.

# МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Методические указания к выполнению практических заданий для студентов всех направлений подготовки бакалавров по дисциплине «Математические методы обработки данных»

Курск 2021

УДК 51

Составители: Н.А. Хохлов

Рецензент

Кандидат физико-математических наук,
доцент кафедры высшей математики
*Дмитриев В.И.*

**Математические методы обработки данных**: методические указания к выполнению практических заданий для студентов всех направлений подготовки бакалавров по дисциплине «Математические методы обработки данных»/ Юго-Зап. гос. ун-т; сост.: Н.А. Хохлов – Курск, 2021. – 20 с.

Методические рекомендации по выполнению практических заданий содержат описание методов, применяемых при решении задач Математических методов обработки данных, задания и вопросы для контроля знаний. Методические указания соответствуют требованиям Федерального государственного образовательного стандарта высшего образования по дисциплине «Математические методы обработки данных». Материал предназначен для студентов всех направлений подготовки бакалавров по дисциплине «Математические методы обработки данных».

Текст печатается в авторской редакции

Подписано в печать _______ . Формат 60x84 1/16.
Усл. печ. л. . Уч.-изд. л. .Тираж ____экз. Заказ_____. Бесплатно.
Юго-Западный государственный университет.
305040 Курск, ул. 50 лет Октября, 94

## Введение

Основной формой обучения студентов является самостоятельная работа с учебником и учебными пособиями. Поэтому каждый студент с самого начала занятий должен выработать для себя рациональную систему работы над курсом, постоянно практикуясь при этом в решении задач. В противном случае усвоение и практическое использование материала затруднены. Чрезвычайно важны систематические занятия.

Рассмотрение решения типовых примеров и задач в параграфе, как правило, расположено по возрастающей трудности. Здесь же представлены индивидуальные задания. Для подготовки к защите представлен список контрольных вопросов.

Для выполнения заданий достаточно аккуратно записанных лекций и внимательного изучения методических рекомендаций, предложенных в данном учебном пособии. Кроме того, весь теоретический материал по данным темам хорошо представлен в учебных пособиях, указанных в списке литературы.

Структура заданий соответствует практическим занятиям курса

### Практические занятия

| № | Наименование практического занятия | Объем, час. |
|---|---|---|
| 1 | 2 | 3 |
| 1 | Выборочные методы математической статистики, статистические методы, используемые в психологии | 4 |
| 2 | Нахождение оценок параметров распределений; доверительный интервал, квантили распределения | 4 |
| 3 | Проверка гипотез о параметрах распределений, Проверка статистических гипотез | 4 |
| 4 | Корреляционный анализ, Построение и анализ простой линейной регрессии | 6 |
| Итого: | | 18 |

## Практическая работа 1. Выборочные методы математической статистики, статистические методы, используемые в психологии

Пример**.** Имеются данные о стаже рабочих цеха: 6, 6, 10, 10, 7, 2, 2, 5, 8, 8, 12, 9, 10, 10, 7, 7, 6, 7, 2, 3. Построить
а) дискретный вариационный ряд;

б) интервальный вариационный ряд.

**Решение.**

**а)** Построение **дискретного** вариационного ряда.

Находится объём выборки, то есть $n$ – общее количество чисел (у нас $n$ = 20).

Составляется таблица, где $x_i$ – варианты (числа из условия), расположенные в порядке возрастания, $n_i$ – сколько раз встречается каждое число.

б) Построение интервального вариационного ряда.

1) Находится величина интервала:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3,322 \cdot \lg n} = \frac{12 - 2}{1 + 3,322 \cdot \lg 20} = \frac{10}{5,3} = 1,9$$

2) Шкала интервалов формируется так:

$$a_1 = x_{i\min} = 2; a_2 = a_1 + h = 2 + 1,9 = 3,9; a_3 = a_2 + h = 3,9 + 1,9 = 5,8;$$
$$a_4 = 7,7; a_5 = 9,6; a_6 = 11,5; a_7 = 13,4 > x_{\max} = 12.$$

| $x_i$ | 2 | 3 | 5 | 6 | 7 | 8 | 9 | 10 | 12 |
|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 3 | 1 | 1 | 3 | 4 | 2 | 1 | 4 | 1 |

3) Составляется таблица, в которой в первой строке формируются границы интервалов, а число во второй строке – это общая сумма частоты встреч всех чисел дискретного ряда, попадающих в соответствующий интервал.

| Интервал | [2; 3,9) | [3,9; 5,8) | [5,8; 7,7) | [7,7; 9,6) | [9,6; 11,5) | [11,5;13,4] |
|---|---|---|---|---|---|---|
| $n_{i\ нов}$ | 3+1=4 | 1 | 3+4=7 | 2+1=3 | 4 | 1 |

**Задания**

1. По имеющимся статистическим данным построить дискретный и интервальный вариационные ряды. Изобразить их графически: построить полигон, гистограмму (деление провести на 4 равных интервала) и кумулятивную кривую.

1. Имеются данные о стаже рабочих цеха: 6, 6, N +1, 10, 11, 2, 2, 5, 8, 8, 12, 9, N +2, 10, 7, 7, 6, 7, 2, 3, 4, 3, 8, 6, 5, 7, 9, N + 1, 9, 5.

2. Имеются следующие данные о среднегодовых вкладах в банках (тыс. руб.): 100, 100, 50, 50, 100, 10□N, 100, 200, 150, 80, 10□N, 150 , 80, 60, 80, 80, 150, 130, 120, 10N, 100, 500, 800, 600, 60, 80, 700, 400, 150, 10N.

3. Имеются данные о дневной выручке денег от продажи товаров в торговых киосках города (тыс. руб.): 2, 2, 5, 7, 2, N + 1, 6, 3, 3, 7, 8, 2, N +2, 4, 9, 4, 3, 5, 5, 7, 8, N+1, 8, 9, N + 2, 8, 6, 3, 3, 4.

4. Имеются данные о средней месячной заработной плате рабочих – сдельщиков (тыс. руб.): 1,0; 1,2;1,2;1,25; 1,5; 1,5; 1+0,1□N; 1,35; 1,5; 1,5; 1+0,1□N; 1,3;1,45; 1,85; 1,8; 1,85; 1,8; 1,9; 1,7; 1,8; 1,6; 1,7; 1,4; 1,5; 1+0,1N; 1,4; 1,4; 1,3; 1,2; 1,5.

5. Имеются данные о выработке продукции рабочими бригадами за смену (в штуках): 14; 7; 8; 9; N + 5; 12; 3; 6; 7; 8; 6; 9; 8; 6; 13; 11; 9; 11; N + 6; 10; 11; 12; 9; 8; 12; 13; 11; N + 5; 6; 7.

6. Имеются следующие данные о количестве произведенной продукции рабочими цеха за смену (в штуках): 16; 22; 15 + N; 25; 15; 19; 16; 17; 18; 13; N + 16; 19; 14; 16; 11; 15; 12; 22; 14; 10; N + 15; 22; 17; 18; 16; 14; 17; 22; 13; 15.

7. Имеются следующие данные о среднем сроке службы деталей не-которых отобранных механизмов (в месяцах): 7; 8,2; 8,6; 7; 7,5 + 0,2N; 8; 8+0,1N; 8,8; 7,2; 7,2; 6,1; 6; 6; 10; 8,2; 7,5; 6; 6,1; 7,2; 8,8; 7,7; 6,1; 7,5+0,2N; 8; 8; 8,8; 6,1; 7,7; 7,2; 6.

8. Имеются следующие данные о выплавке чугуна за отчетный период на заводе (тыс. т): 5,6; 5,2; 5,3; 5,5; 5+0,1N; 5,5; 5,3; 5,6; 5+0,1N; 5,6; 5,4; 5,8; 5,3; 5,8; 5,5; 5,2; 5,7; 5,8; 5,6; 5,4; 5+0,1N; 5,4; 5,3; 5,5; 5,6; 5,8; 5,3; 5,2; 5,8; 5,7.

9. Имеются следующие данные о производстве часов по годам (млн. шт.): 20; 21; 25 +N; 30 – N; 27; 20; 20; 30; 33; 22; 23; 35; 33; 32; 32; 29; 22; 25; 33; 30 – N; 25 + N; 22; 25; 24; 33; 32; 25; 33; 24; 30 – N.

10. Имеются следующие данные об уровне энерговооруженности труда (кВт): 50; 52; 50; 52; 52; 50 + N; 60 – N; 60; 63; 60; 50 + N; 55; 55; 54; 53; 50 + N; 59; 57; 55; 52; 54; 50 + N; 60 – N; 63; 50; 53; 54; 55; 57; 59.

11. Имеются следующие данные о себестоимости одной единицы продукции (тыс. руб.): 13; 13; 12; 11; 12; 12; 10; 9; 9; 8 + N; 10; 10; 8; 12; 9 + N; 12; 11; 10; 15; 9 + N; 15; 13; 11; 12; 15; 9; 8; 9 + N; 15; 13.

12. Имеются данные по заводам за отчетный период о среднегодовой стоимости основных промышленно – производственных фондов (млн. руб.): 100; 130; 150; 140; 100 + 10N; 100; 100 + 10N; 100; 120; 110; 120; 100 +10N; 160; 160, 150; 140; 160; 100 + 10N; 120; 150; 110; 130; 140; 150; 100 + 10N; 150; 120; 120; 140; 100 + 10N.

13. Имеются следующие данные по заводам за отчетный период о фактическом выпуске продукции (млн. руб.): 140; 140; 150; 180; 200 – 10N; 170; 130; 170; 150; 150; 120; 110; 120; 100; 200 – 10N; 160; 180; 170; 120; 130; 130; 120; 150; 200 – 10N; 140; 120; 130; 130; 170; 160.

14. Имеются данные по группе предприятий об основных производ-ственных фондах (млн. руб.): 3; 4; 5; 8; N + 5; 10; 7; 6; 5; 4; N + 5; 10; N + 5; 11; N + 5; 4; 6; 7; 8; 3; 8; 4; 5; 3; 8; N + 5; 7; 6; 5; 3.

15. Имеются данные по группе предприятий о валовой продукции (млн. руб.): 3; 5; 10; N + 6; 6; 4; 7; 8; 8; 3; 5; 10; 6; 6; 9; 4; 5; N + 6; 8; 3; 6; 5; 3; 8; 4; 7; 5; N + 6; 9; 9.

**Контрольные вопросы**

1. Дайте определения: перестановок, сочетаний, размещений.
2. Сформулируйте классическое определение вероятностей. Укажите недостатки этого определения.
3. Какое событие называется достоверным, невозможным?

4. Дайте определение полной группы событий.
5. Какие события называются несовместными, совместными, противоположными, независимыми?

## Практическая работа 2. Нахождение оценок параметров распределений; доверительный интервал, квантили распределения

**Пример.** Пусть средняя продолжительность обучения равна $M^{*}X=500$ часов. Предлагается новая технология обучения. Экспериментальное обучение по новой технологии показало среднее время обучения 560 часов. Пусть среднеквадратическое отклонение

$$\sigma_{\overline{X}} = \frac{\sigma_x}{\sqrt{n}} = 45$$

Вопрос: достаточно ли это основание для перехода на новую технологию. Так как $M^{*}X$ приближенно распределена по нормальному закону, то выберем уровень значимости $\alpha=0,05$ и третий вид критической области. Находим $t_{\alpha}=1,96$. Для принятия гипотезы о том что отклонение математического ожидания значимо нужно, чтобы выполнялось неравенство

$$\left|\overline{X} - m\right| > 1{,}96\sigma = 88.$$

В нашем примере отклонение составило 60 часов. Следовательно, оно могло получиться за счет действия только случайных факторов. И для перехода к новой технологии оснований недостаточно.

Пусть у нас есть две независимые выборки случайных величин X и Y. Часто нужно проверить гипотезу о том, что математические ожидания двух выборок из нормальных случайных величин равны, т. е. MX=MY. Если дисперсии $\sigma^2_x$ и $\sigma^2_y$ известны и гипотеза выполняется, то величина $Z= M^{*}X - M^{*}Y$ распределена по нормальному закону с параметрами $a=0$ и

$$\sigma^2 = \frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}.$$

Выберем критерий значимости и критическую область вида III. Если величина Z попадает в критическую область, то гипотеза о равенстве математических ожиданий противоречит данным наблюдений и ее следует отклонить. Если нет, то гипотеза согласуется с данными наблюдений и ее следует принять.

Пример. Имеются данные об испытании на разрыв двух партий проволоки, произведенных двумя разными заводами. Получены результаты. Для первой партии $n_1$=50, $M^*X$=120,8 кг/мм$^2$, $\sigma_x$=8,0 кг/мм$^2$ и Для второй партии $n_2$=50, $M^*Y$=128,2 кг/мм$^2$, $\sigma_y$=9,4 кг/мм$^2$. Найдем

$$\sigma_z = \sqrt{\frac{8^2}{50} + \frac{9{,}4^2}{50}} = 1{,}75$$

Для $\alpha$=5% находим $t_\alpha$=1,96. Тогда критическое отклонение равно 1,75·1,96=3,43 кг/мм$^2$. В нашем примере отклонение равно 7,4. Это значит что гипотезу о равенстве математических ожиданий следует отклонить.

**Задания**

Построить доверительный интервал для оценки математического ожидания а нормального распределения с надежностью P, зная выборочное среднее $\bar{x}$, объём выборки n, выборочную дисперсию $S^{*2}$.

Таблица 2.2

Индивидуальные данные к заданию 4

| **m** | $\bar{x}$ | **P** | **n** | $S^{*2}$ |
|---|---|---|---|---|
| 1 | 1,9 | 0,9 | 70 | 1,96 |
| 2 | 2,7 | 0,95 | 75 | 2,25 |
| 3 | 3,5 | 0,99 | 55 | 1,21 |
| 4 | 1,8 | 0,9 | 60 | 1,44 |
| 5 | 4,6 | 0,95 | 65 | 1,69 |
| 6 | 2,5 | 0,99 | 80 | 2,56 |
| 7 | 15,3 | 0,9 | 85 | 2,89 |
| 8 | 12,2 | 0,95 | 90 | 3,24 |
| 9 | 14,7 | 0,99 | 95 | 3,61 |
| 10 | 5,8 | 0,9 | 105 | 4,41 |
| 11 | 6,2 | 0,95 | 110 | 4,84 |
| 12 | 7,5 | 0,99 | 115 | 5,29 |
| 13 | 8,3 | 0,9 | 120 | 5,76 |

| | | | | |
|---|---|---|---|---|
| 14 | 9,7 | 0,95 | 125 | 6,25 |
| 15 | 10,4 | 0,99 | 130 | 6,76 |
| 16 | 1,3 | 0,9 | 135 | 7,29 |
| 17 | 2,3 | 0,95 | 140 | 7,84 |
| 18 | 17,2 | 0,9 | 155 | 9,61 |
| 20 | 16,6 | 0,95 | 45 | 0,81 |

## Контрольные вопросы

1. Какие виды случайных величин вы знаете?
2. Перечислите важнейшие характеристики случайных величин.
3. Какие важнейшие распределения случайных величин вы знаете?
4. Если p – вероятность наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность того, что событие A в n независимых испытаниях наступит ровно m раз Pn(m), можно найти используя локальную теорему Лапласа или интегральную теорему Лапласа?
5. Какое название носят величины, значения которых нельзя заранее указать и которые зависят от случайных причин?
6. Если случайная величина может принимать отдельные, изолированные значения, причем их количество конечно или бесконечно, но счетно, то такая величина носит название дискретной, непрерывной или смешанной?
7. Как называется перечень всех значений дискретной случайной величины и их вероятностей?
8. Как находят математическое ожидание дискретной случайной величины: как среднее арифметическое, среднее геометрическое, среднее квадратическое?

## Практическая работа 3. Проверка гипотез о параметрах распределений. Проверка статистических гипотез

**Пример.**

Выдвинутая гипотеза $H_0$: предполагаемый закон распределения f(x) – нормальный закон распределения с параметрами $M(X) = \bar{x}$, $D(X) = S^{*2}$.

Конкурирующая гипотеза $H_1$: предполагаемый закон распределения нормальным не является.

Таблица 3.1

Исходные данные

| m | Интервалы | | | | | | |
|---|---|---|---|---|---|---|---|
| | Частоты | | | | | | |
| 1 | (1; 3) | (3; 5) | (5; 7) | (7; 9) | (9; 11) | (11; 13) | (13; 15) |
| | 15 | 25 | 30 | 50 | 45 | 30 | 5 |

Объем выборки:

$$n = 15 + 25 + 30 + 50 + 45 + 30 + 5 = 200$$

Прежде чем искать числовые характеристики выборки, заменим интервалы соответствующими им серединами.

Таблица 3.2

Построение дискретного вариационного ряда

| Интервалы | (1; 3) | (3; 5) | (5; 7) | (7; 9) | (9; 11) | (11; 13) | (13; 15) |
|---|---|---|---|---|---|---|---|
| Середины | 2 | 4 | 6 | 8 | 10 | 12 | 14 |
| Частоты | 15 | 25 | 30 | 50 | 45 | 30 | 5 |

Числовые характеристики:

– выборочное среднее:

$$\bar{x} = \frac{1}{n}\sum_i x_i n_i = \frac{1}{200}(2 \cdot 15 + 4 \cdot 25 + 6 \cdot 30 + 8 \cdot 50 + 10 \cdot 45 + 12 \cdot 30 + 14 \cdot 5) =$$

$$= 7{,}95$$

– выборочная дисперсия:

$$S^2 = \overline{x^2} - (\bar{x})^2$$

$$\overline{x^2} = \frac{1}{200}(2^2 \cdot 15 + 4^2 \cdot 25 + 6^2 \cdot 30 + 8^2 \cdot 50 + 10^2 \cdot 45 + 12^2 \cdot 30 +$$

$$+ 14^2 \cdot 5) = 72{,}7$$

$$S^2 = 72{,}7 - 7{,}95^2 = 9{,}4975.$$

– выборочное среднее квадратическое отклонение:

$$S = \sqrt{S^2} = 3{,}08$$

Для проверки выдвинутой гипотезы используем критерий согласия Пирсона.

$$\tau_i = \frac{x_i - \bar{x}}{S},$$

где $x_i$ – границы интервалов.

$\Phi(x_i)$ – значения функции Лапласа.

$p_i = \Phi(x_i) - \Phi(x_{i-1})$

Теоретические частоты: $n_i^{/} = n \cdot p_i$

Необходимые расчеты представим в таблице.

Таблица 3.3

Расчет значения критерия Пирсона

| Интервалы | $n_i$ | $\tau_i$ | $\Phi(\tau_i)$ | $p_i$ | $n_i^{/}$ | $\frac{(n_i - n_i^{/})^2}{n_i^{/}}$ |
|---|---|---|---|---|---|---|
| | – | –2,26 | –0,4881 | – | – | – |
| (1; 3) | 15 | –1,61 | –0,4463 | 0,0418 | 8,36 | 5,27 |
| (3: 5) | 25 | –0,96 | –0,3315 | 0,1148 | 22,96 | 0,18 |

| | | | | | | |
|---|---|---|---|---|---|---|
| (5; 7) | 30 | –0,31 | –0,1217 | 0,2098 | 41,96 | 3,41 |
| (7; 9) | 50 | 0,34 | 0,1331 | 0,2548 | 50,96 | 0,02 |
| (9; 11) | 45 | 0,99 | 0,3389 | 0,2058 | 41,16 | 0,36 |
| (11; 13) | 30 | 1,64 | 0,4495 | 0,1106 | 22,12 | 2,81 |
| (13; 15) | 5 | 2,29 | 0,4890 | 0,0395 | 7,9 | 1,06 |
| Сумма | 200 | – | – | 0,9771 | 195,42 | 13,11 |

Расчетное значение критерия Пирсона: $\chi^2_{\text{расч}} = 13{,}11$.

Табличное значение критерия Пирсона находим при доверительной вероятности Р = 0,95: $\chi^2_{\text{т}}(0{,}05; 7-2-1) = 9{,}5$.

Так как расчетное значение превышает табличное значение, то выдвинутая гипотеза отвергается.

Числа M,N,P,K в заданиях задаются преподавателем индивидуально.

Задания

1. По имеющимся статистическим данным построить дискретный и интервальный вариационные ряды. Изобразить их графически: построить полигон, гистограмму (деление провести на 4 равных интервала) и кумулятивную кривую.

Имеются данные о стаже рабочих цеха: 6, 6, N +1, 10, 11, 2, 2, 5, 8, 8, 12, 9, N +2, 10, 7, 7, 6, 7, 2, 3, 4, 3, 8, 6, 5, 7, 9, N + 1, 9, 5.

2. По заданному в нижеследующих задачах статистическому ряду выборки найти числовые характеристики:

а) выборочную дисперсию;
б) выборочное среднее квадратическое отклонение;
в) размах выборки;
г) асимметрию;
д) эксцесс.

Имеются следующие данные об уровне энерговооруженности труда (кВт): 50; 52; 50; 52; 52; 50 + N; 60 – N; 60; 63; 60; 50 + N; 55; 55; 54; 54; 54; 60 – N; 63; 63; 55; 60 – N; 60 – N; 50; 50 + N; 55; 55; 50; 54; 52; 52. Найти также среднюю энерговооруженность труда.

3. Построить доверительный интервал для оценки математического ожидания a нормального распределения с надежностью P, зная выборочное среднее $\overline{x}$, объём выборки $n = N^2 \cdot 25$ и генеральное среднее квадратичное отклонение $\sigma = \dfrac{N}{4}$. Индивидуальные задания смотри в табл. 1

Таблица 1

**Индивидуальные задачи к заданию 3**

| **n** | $\overline{x}$ | **P** | N | $\overline{x}$ | **P** |
|---|---|---|---|---|---|
| 1 | 75,09 | 0,9 | 2 | 85,13 | 0,99 |
| 3 | 25,17 | 0,91 | 4 | 65,10 | 0,98 |
| 5 | 55,14 | 0,92 | 6 | 75,11 | 0,97 |
| 7 | 15,15 | 0,93 | 8 | 45,08 | 0,96 |
| 9 | 65,12 | 0,94 | 10 | 25,16 | 0,95 |
| 11 | 75,07 | 0,9 | 12 | 35,17 | 0,99 |
| 13 | 95,06 | 0,91 | 14 | 25,18 | 0,98 |
| 15 | 45,05 | 0,92 | 16 | 65,19 | 0,97 |
| 17 | 85,04 | 0,93 | 18 | 75,20 | 0,96 |
| 19 | 75,03 | 0,94 | 20 | 15,25 | 0,95 |
| 21 | 35,21 | 0,9 | 22 | 85,24 | 0,99 |
| 23 | 45,22 | 0,91 | 24 | 35,23 | 0,98 |
| 25 | 95,02 | 0,92 | 26 | 25,26 | 0,97 |
| 27 | 15,28 | 0,93 | 28 | 75,27 | 0,96 |
| 29 | 25,29 | 0,94 | 30 | 65,35 | 0,95 |
| 31 | 45,30 | 0,9 | 32 | 95,32 | 0,99 |
| 33 | 85,34 | 0,91 | 34 | 115,33 | 0,98 |
| 35 | 65,31 | 0,92 | | | |

## Контрольные вопросы

1. Дайте понятие вариационного ряда.
2. Какие виды вариационных рядов вы знаете?
3. Какие графики используются для изображения дискретных вариационных рядов?
4. Перечислите важнейшие точечные характеристики выборки.

5. Дайте понятие доверительного интервала.
6. Дайте определение основной и конкурирующей гипотез.

**Практическая работа 4. Корреляционный анализ, построение и анализ простой линейной регрессии.**

**Пример.**

Пусть N = 9, n = 50. Тогда A = 12, B = 3, C = 2, D = 3.

Для заданных значений параметров A, B, C, D корреляционная таблица имеет вид:

Таблица 3.4

Корреляционная таблица для заданных A, B, C, D.

| Y \ X | 0 | 1 | 2 | 3 | 4 | 5 | $n_y$ |
|---|---|---|---|---|---|---|---|
| 1 | – | 3 | – | – | – | – | 3 |
| 2 | 3 | – | 2 | – | – | – | 5 |
| 3 | – | – | 1 | 2 | 12 | 1 | 16 |
| 4 | – | – | – | – | 1 | – | 1 |
| $n_x$ | 3 | 3 | 3 | 2 | 13 | 1 | 25 |

Объём выборки равен 2 +3 + 3 +1 + 2 + 12 + 1 + 1 = 25.

1) Вычислим выборочные характеристики:

$$\overline{x} = \frac{1}{25}(0\cdot 3 + 1\cdot 3 + 2\cdot 3 + 3\cdot 2 + 4\cdot 13 + 6\cdot 1) = \frac{72}{25};$$

$$\overline{y} = \frac{1}{25}(1\cdot 3 + 2\cdot 5 + 3\cdot 16 + 4\cdot 1) = \frac{13}{5};$$

$$\overline{x^2} = \frac{1}{25}(0^2\cdot 3 + 1^2\cdot 3 + 2^2\cdot 3 + 3^2\cdot 2 + 4^2\cdot 13 + 6^2\cdot 1) = \frac{266}{25};$$

$$\overline{y^2} = \frac{1}{25}(1^2\cdot 3 + 2^2\cdot 5 + 3^2\cdot 16 + 4^2\cdot 1) = \frac{183}{25};$$

$$S_X^2 = \overline{x^2} - (\overline{x})^2 = \frac{266}{25} - \left(\frac{72}{25}\right)^2 = 2{,}346;$$

$$S_Y^2 = \overline{y^2} - (\overline{y})^2 = \frac{183}{25} - \left(\frac{13}{5}\right)^2 = 0{,}56;$$

$S_X = 1{,}532;$ $\qquad S_Y = 0{,}748;$

$$\overline{x \cdot y} = \frac{1}{25}(1 \cdot 1 \cdot 3 + 0 \cdot 2 \cdot 3 + 2 \cdot 2 \cdot 2 + 2 \cdot 3 \cdot 1 + 3 \cdot 3 \cdot 2 + 4 \cdot 3 \cdot 12 + 5 \cdot 3 \cdot 1 +$$

$$+ 4 \cdot 4 \cdot 1) = \frac{210}{25}$$

$$\operatorname{cov}(X, Y) = \overline{x \cdot y} - \overline{x} \cdot \overline{y} = \frac{210}{25} - \frac{72}{25} \cdot \frac{13}{5} = 0{,}912;$$

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{S_X \cdot S_Y} = \frac{0{,}912}{1{,}532 \cdot 0{,}748} = 0{,}796.$$

2) Проверим значимость коэффициента корреляции по критерию Стьюдента.

Выдвинутая гипотеза $H_0$: коэффициент корреляции значим.

Конкурирующая гипотеза $H_1$: коэффициент корреляции незначим.

Расчетное значение критерия Стьюдента:

$$t_{расч} = \frac{\rho_{XY}}{\sqrt{1 - \rho_{XY}^2}} \cdot \sqrt{n - 2} = \frac{0{,}796}{\sqrt{1 - 0{,}796^2}} \sqrt{25 - 2} = 6{,}31.$$

При доверительной вероятности Р = 0,95 табличное значение критерия Стьюдента

$t_{табл}$ (0,05; 25 – 2) = 2,07.

Так как расчетное значение превышает табличное, то выдвинутая гипотеза отвергается, т.е. коэффициент корреляции значим. Между величинами X и Y существует тесная прямая линейная связь, т.е. с увеличением X увеличивается Y и наоборот.

3) В общем случае уравнения прямых регрессий имеют вид:

$$\text{Y на X: } y - \bar{y} = \rho_{XY} \cdot \frac{S_Y}{S_X}(x - \bar{x}),$$

$$\text{X на Y: } x - \bar{x} = \rho_{XY} \cdot \frac{S_X}{S_Y}(y - \bar{y}).$$

При решении данной задачи уравнения прямых регрессий примут вид:

Y на X:

$y - 2{,}6 = 0{,}388(x - 2{,}88)$;

X на Y:

$x - 2{,}88 = 1{,}63(y - 2{,}6)$.

Графически прямые регрессии изображены на рис. 3.1.

Рис. 4.1. Прямые регрессии

Числа A,B,C,D в заданиях задаются преподавателем индивидуально.

Задания

1. Для двух случайных величин X и Y проведена серия испытаний. Результаты испытаний записаны в следующую корреляционную таблицу. Четные варианты индивидуальные задания берут из таблицы 2.1, а нечетные – из таблицы 2.2.

Таблица 2.1

Индивидуальные данные к заданию 6

| Y \ X | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | D | C | – | – | – | – |
| 2 | – | C | B | – | – | – |
| 3 | – | – | A | B | A | 1 |
| 4 | – | – | – | – | D | C |

Таблица 2.2

Индивидуальные данные к заданию 6

| Y \ X | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 1 | | | | | | A |
| 2 | | | | C | B | A |
| 3 | | D | 3 | B | | |
| 4 | C | 2 | | | | |

Для этих случайных величин:

Вычислить числовые характеристики выборочные средние, выборочные дисперсии, ковариацию и выборочный коэффициент корреляции $\rho_{XY}$.

Проверить для доверительной вероятности P = 0,95 значимость коэффициента корреляции $\rho_{XY}$. Сделать вывод о тесноте взаимосвязи.

Написать уравнения прямых регрессий Y на X и X на Y.

В подходящем масштабе изобразить на графике точки (x, y) из корреляционной таблицы и прямые регрессии.

2. Над случайными величинами X, Y, Z проведена серия из 8 наблюдений. Результаты записаны в таблицу

Таблица 2.3

Индивидуальные данные к заданию 2

| | X | Y | Z |
|---|---|---|---|
| 1 | 1 | A | 0 |
| 2 | 0 | 1 | A |
| 3 | 2 | B | 3 |
| 4 | C | 2 | 3 |
| 5 | 3 | 1 | 1 |
| 6 | 2 | 0 | -1 |
| 7 | A | 3 | B |
| 8 | 1 | C | D |

Составить матрицы моментов и корреляционную. Вычислить коэффициент множественной корреляции между переменной Z (как функции от X, Y) и переменными X, Y.

## Контрольные вопросы

1. Дайте определение прямой регрессии Y на X, X на Y.
2. Как вычисляется линейный коэффициент парной корреляции?
3. Как вычисляется индекс корреляции R?
4. Как осуществляется оценка статистической значимости линейного коэффициента парной корреляции?
5. Как осуществляется оценка статистической значимости индекса корреляции?

# РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие. -М.: ЮРАЙТ, 2012.–479с. Текс: непосредственный
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учебное пособие. -М.: ЮРАЙТ, 2011.-404с. Текс: непосредственный
3. Расчёт вероятностей случайных событий: индивидуальные задания и методические указания по выполнению модуля 13 / Юго-Зап. гос. ун-т; сост.: Е.В.Журавлёва, Е.А.Панина. –Курск: ЮЗГУ, 2011. -50 с. Текст: электронный.
4. Повторные испытания. Случайные величины: индивидуальные задания к модулю / Юго-Зап. гос. ун-т ; сост.: Е. В. Журавлева, Е. А. Панина. - Курск : ЮЗГУ, 2019. - 54 с. - Текст: электронный.
5. Элементы математической статистики и корреляционного анализа: методические указания и индивидуальные задания к модулю / Юго-Зап. гос. ун-т ; сост.: Е.В. Журавлева [и др.]. - Курск: ЮЗГУ, 2020. - 35 с. : табл. - Текст: электронный.