

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Локтионова Оксана Геннадьевна  
Должность: проректор по учебной работе  
Дата подписания: 04.05.2022 14:05:17  
Уникальный программный ключ:  
0b817ca911e6668abb13a5040c3d910ca4d19ca311e

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

« 10 » 02

2022 г.



**МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Методические указания к выполнению практических заданий  
по дисциплине «Математическая статистика»  
для направления подготовки 45.03.03 – Фундаментальная  
и прикладная лингвистика, направленность  
(профиль) «Теоретическая и прикладная лингвистика»

Курск 2022

УДК 51

Составитель: Н. А. Моргунова

Рецензент

Доктор физ.-мат. наук,  
профессор кафедры высшей математики

*Хохлов Н. А.*

**Математическая статистика:** методические указания к выполнению практических заданий по дисциплине «Математическая статистика» для направления подготовки 45.03.03 – Фундаментальная и прикладная лингвистика, направленность (профиль) «Теоретическая и прикладная лингвистика» / Юго-Зап. гос. ун-т; сост.: Н. А. Моргунова – Курск, 2022. – 14 с.

Излагаются методические рекомендации по выполнению практических заданий. Содержатся краткие описания методов, применяемых при решении задач по математической статистике, задания и вопросы для контроля знаний.

Методические указания соответствуют требованиям Федерального государственного образовательного стандарта высшего образования для направления подготовки 45.03.03 - Фундаментальная и прикладная лингвистика, направленность (профиль) «Теоретическая и прикладная лингвистика». Материал предназначен для студентов очной и заочной форм обучения по направлению подготовки 45.03.03 - Фундаментальная и прикладная лингвистика, направленность (профиль) «Теоретическая и прикладная лингвистика», а также будет полезен студентам всех других направлений подготовки, изучающих раздел высшей математики «Математическая статистика».

Текст печатается в авторской редакции

Подписано в печать \_\_\_\_\_ . Формат 60x84 1/16.

Усл. печ. л. . Уч.-изд. л. 0,4. Тираж \_\_\_\_\_ экз. Заказ 804. Бесплатно.

Юго-Западный государственный университет.

305040 Курск, ул. 50 лет Октября, 94

**Цель работ:** освоить необходимый математический аппарат, помогающий анализировать, моделировать и решать прикладные задачи.

### Задания по работам

1. Тема «Выборочный метод».

Выборка задана в виде распределения частот:

$x_i$	2	4	6	7
$m_i$	3	2	4	1

Построить полигон относительных частот.

2. Тема «Точечные оценки параметров распределения».

Выборочная совокупность задана таблицей распределения:

$x_i$	4	7	8	12
$m_i$	5	2	3	10

Найти выборочную среднюю и выборочную дисперсию.

3. Тема «Интервальные оценки параметров распределения».

Признак  $X$  распределен в генеральной совокупности по нормальному закону с известным средним квадратическим отклонением  $\sigma=3,44$ . Найти по данным выборки доверительный интервал для математического ожидания  $\mu$  с надежностью  $\gamma=0,95$ , если объем выборки  $n=20$ , выборочная средняя  $\bar{x}_B = 8,9$ .

4. Тема «Метод наименьших квадратов».

Себестоимость  $Y$  одного экземпляра книги в зависимости от тиража  $X$  (в тысячах экземпляров) в течение ряда лет характеризуется таблицей:

$X$ , тыс. экз.	2	4	6	8	10
$Y$ , тыс. руб.	3	5	6,5	7,5	9

Найти линейную зависимость себестоимости экземпляра книги от тиража.

5. Тема «Корреляционный анализ».

Распределение 100 заводов по производственным средствам в миллионах рублей  $X$  и по суточной выработке  $Y$  характеризуется корреляционной таблицей:

$Y / X$	10	15	20	25	30	35	$m_x$
50	2	2					4
60	2	4	5	6	4		21
70		2	7	12	10	4	35
80				10	10	6	26
90				8		6	14
$m_y$	4	8	12	36	24	16	100

По таблице вычислены выборочные средние признаков  $X$  и  $Y$ , выборочные средние квадратические отклонения и выборочный коэффициент корреляции.

$$\bar{x} = 72,5; \bar{y} = 25,8;$$

$$\sigma_x = 10,62; \sigma_y = 6,43;$$

$$r_b = 0,56.$$

Найти выборочные уравнения прямых линий регрессии  $Y$  на  $X$  и  $X$  на  $Y$ .

6. Тема «Проверка статистических гипотез».

Две независимые выборкам, объемы которых  $n_1 = 30$  и  $n_2 = 25$ , извлечены из нормальных генеральных совокупностей  $X$  и  $Y$ .

$x_i$	1	2	3	4	5
$m_i$	3	5	10	6	6

$y_i$	1	2	3	4	5
$m_i$	1	5	10	6	3

При уровне значимости  $\alpha=0,05$  проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий при конкурирующей гипотезе  $H_1: D(X) > D(Y)$ .

## 7. Тема «Математическая статистика в лингвистике».

Прочитать следующий текст:

*Ликуй, ликуй, наш курский соловей,  
Людские души свято возвышая  
До красоты мелодии своей,  
Сравнимой, может, с музыкою рая!*

Выбрав в качестве исследуемого признака длину слова (число букв в каждом слове текста), составить статистический и вариационный ряды, закон распределения частот; найти эмпирическую функцию распределения.

**Примеры выполнения заданий с кратким описанием  
применяемых методов**

**1. Тема «Выборочный метод»**

Выборка задана в виде распределения частот:

$x_i$	2	4	6	7
$m_i$	3	2	4	1

Построить полигон относительных частот.

*Решение.*

Найдем объем выборки:  $n = 3 + 2 + 4 + 1 = 10$ .

Найдем относительные частоты:

$$w_1 = \frac{3}{10} = 0,3; w_2 = \frac{2}{10} = 0,2; w_3 = \frac{4}{10} = 0,4; w_4 = \frac{1}{10} = 0,1.$$

Распределение относительных частот следующее:

$x_i$	2	4	6	7
$w_i$	0,3	0,2	0,4	0,1

Контроль:  $0,3+0,2+0,4+0,1=1$ .

Далее отложим на оси абсцисс варианты  $x_i$ , а на оси ординат – соответствующие относительные частоты  $w_i$ . Соединив точки  $(x_i, w_i)$  отрезками прямых, получим искомый полигон относительных частот (рис. 1).

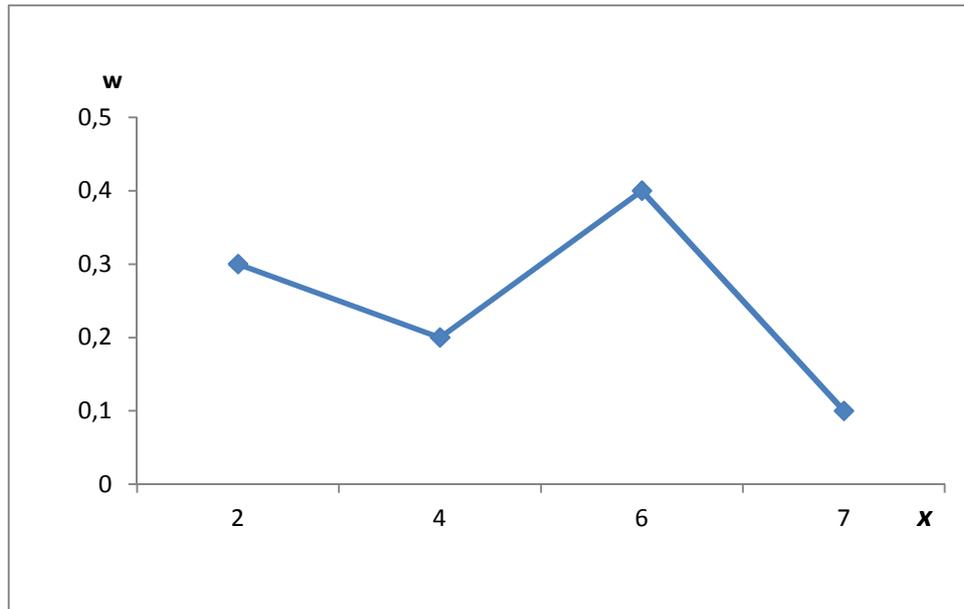


Рис. 1

## 2. Тема «Точечные оценки параметров распределения»

Выборочная совокупность задана таблицей распределения:

$x_i$	4	7	8	12
$m_i$	5	2	3	10

Найти выборочную среднюю и выборочную дисперсию.

*Решение.*

Используя формулу для вычисления выборочной средней

$$\bar{x}_B = \frac{1}{n} \sum_{i=1}^k x_i n_i,$$

получим:

$$\bar{x}_B = \frac{4 \cdot 5 + 7 \cdot 2 + 8 \cdot 3 + 12 \cdot 10}{5 + 2 + 3 + 10} = 8,9.$$

Найдем выборочную дисперсию по формуле:

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

$$\text{Имеем: } D_B = \frac{(4-8,9)^2 \cdot 5 + (7-8,9)^2 \cdot 2 + (8-8,9)^2 \cdot 3 + (12-8,9)^2 \cdot 10}{5+2+3+10} = 11,29.$$

### 3. Тема «Интервальные оценки параметров распределения»

Признак  $X$  распределен в генеральной совокупности по нормальному закону с известным средним квадратическим отклонением  $\sigma=3,44$ . Найти по данным выборки доверительный интервал для математического ожидания  $a$  с надежностью  $\gamma=0,95$ , если объем выборки  $n=20$ , выборочная средняя  $\bar{x}_B = 8,9$ .

*Решение.*

Доверительный интервал для математического ожидания, покрывающий с надежностью  $\gamma$  неизвестный параметр  $a$ , следующий:

$$\left( \bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}} \right).$$

Найдем параметр  $t$  из условия:  $\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475$ .

По таблице приложений значений функции Лапласа находим  $t=1,96$ .

Концы доверительного интервала равны:

$$\bar{x}_B - \frac{t\sigma}{\sqrt{n}} = 8,9 - \frac{1,96 \cdot 3,44}{\sqrt{20}} \approx 7,39,$$

$$\bar{x}_B + \frac{t\sigma}{\sqrt{n}} = 8,9 + \frac{1,96 \cdot 3,44}{\sqrt{20}} \approx 10,41.$$

Итак, доверительный интервал  $(7,39; 10,41)$  покрывает  $a$  с надежностью  $0,95$ .

### 4. Тема «Метод наименьших квадратов»

Себестоимость  $Y$  одного экземпляра книги в зависимости от тиража  $X$  (в тысячах экземпляров) в течение ряда лет характеризуется таблицей:

$X$ , тыс. экз.	2	4	6	8	10
$Y$ , тыс. руб.	3	5	6,5	7,5	9

Найти линейную зависимость себестоимости экземпляра книги от тиража.

*Решение.*

Экспериментальные точки нанесем на координатную плоскость (рис. 2):

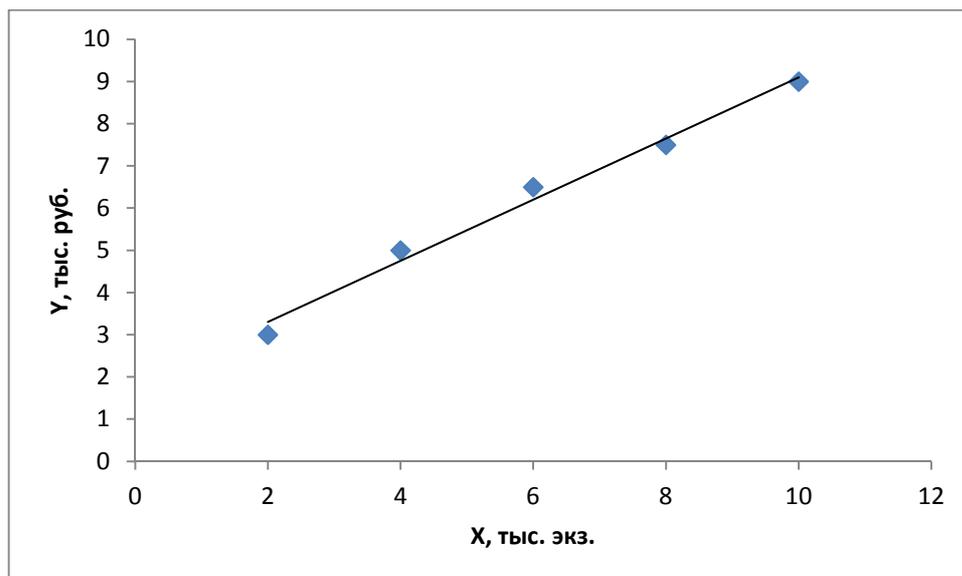


Рис. 2

По характеру расположения на координатной плоскости экспериментальных точек можно предположить, что между  $X$  и  $Y$  существует линейная зависимость  $y = ax + b$ .

Найдем аппроксимирующую (приближаемую) функцию методом наименьших квадратов.

Для определения коэффициентов  $a$  и  $b$  воспользуемся фундаментальной (нормальной) системой уравнений:

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

где  $n$  – количество экспериментальных точек.

Произведем необходимые вычисления и запишем их в таблицу:

$n$	$x_i$	$y_i$	$x_i^2$	$x_i y_i$
1	2	3	4	6
2	4	5	16	20
3	6	6,5	36	39
4	8	7,5	64	60
5	10	9	100	90
$\sum_{i=1}^n$	30	31	220	215

Система линейных уравнений для определения величин  $a$  и  $b$  примет вид:

$$\begin{cases} 220a + 30b = 215, \\ 30a + 5b = 31. \end{cases}$$

Решив систему, получим  $a=0,725$ ,  $b=1,85$ .

Функция себестоимости экземпляра книги в зависимости от тиража имеет вид  $y = 0,725x + 1,85$ .

### 5. Тема «Корреляционный анализ»

Распределение 100 заводов по производственным средствам в миллионах рублей  $X$  и по суточной выработке  $Y$  характеризуется корреляционной таблицей:

$Y / X$	10	15	20	25	30	35	$m_x$
50	2	2					4
60	2	4	5	6	4		21
70		2	7	12	10	4	35
80				10	10	6	26
90				8		6	14
$m_y$	4	8	12	36	24	16	100

По таблице вычислены выборочные средние признаков  $X$  и  $Y$ , выборочные средние квадратические отклонения и выборочный коэффициент корреляции.

$$\bar{x} = 72,5; \bar{y} = 25,8;$$

$$\sigma_x = 10,62; \sigma_y = 6,43;$$

$$r_b = 0,56.$$

Найти выборочные уравнения прямых линий регрессии  $Y$  на  $X$  и  $X$  на  $Y$ .

*Решение.*

Выборочные уравнения прямых линий регрессии  $Y$  на  $X$  и  $X$  на  $Y$  имеют вид:

$$\overline{y_x} - \bar{y} = r_b \frac{\sigma_y}{\sigma_x} (x - \bar{x}),$$

$$\overline{x_y} - \bar{x} = r_b \frac{\sigma_x}{\sigma_y} (y - \bar{y}),$$

где  $\overline{y_x}$ ,  $\overline{x_y}$  - условные средние.

Подставив найденные величины в указанные уравнения, получим искомые уравнения прямых линий регрессии:

$$\overline{y_x} - 25,8 = 0,56 \frac{6,43}{10,62} (x - 72,5) \text{ или } \overline{y_x} = 0,34x + 1,15;$$

$$\overline{x_y} - 72,5 = 0,56 \frac{10,62}{6,43} (y - 25,8) \text{ или } \overline{x_y} = 0,92y + 48,76.$$

### **6. Тема «Проверка статистических гипотез»**

Две независимые выборкам, объемы которых  $n_1 = 30$  и  $n_2 = 25$ , извлечены из нормальных генеральных совокупностей  $X$  и  $Y$ .

$x_i$	1	2	3	4	5
$m_i$	3	5	10	6	6

$y_i$	1	2	3	4	5
$m_i$	1	5	10	6	3

При уровне значимости  $\alpha=0,05$  проверить нулевую гипотезу  $H_0: D(X) = D(Y)$  о равенстве генеральных дисперсий при конкурирующей гипотезе  $H_1: D(X) > D(Y)$ .

*Решение.*

Найдем точечные оценки указанных выборок.

Выборочные средние равны

$$\bar{x}_B = \frac{1 \cdot 3 + 2 \cdot 5 + 3 \cdot 10 + 4 \cdot 6 + 5 \cdot 6}{3 + 5 + 10 + 6 + 6} = 3,23,$$

$$\bar{y}_B = \frac{1 \cdot 1 + 2 \cdot 5 + 3 \cdot 10 + 4 \cdot 6 + 5 \cdot 3}{1 + 5 + 10 + 6 + 3} = 3,2.$$

Исправленные дисперсии вычислим по формулам:

$$S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^k (x_i - \bar{x}_B)^2 m_i,$$

$$S_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^k (y_i - \bar{y}_B)^2 m_i.$$

В результате получим: при  $n_1 = 30$   $S_x^2 = 1,57$ ; при  $n_2 = 25$   $S_y^2 = 1,083$ .

Найдем наблюдаемое значение критерия – отношение большей исправленной дисперсии к меньшей:  $F_{\text{набл}} = \frac{1,57}{1,083} \approx 1,45$ .

Так как конкурирующая гипотеза имеет вид  $D(X) > D(Y)$ , то критическая область – правосторонняя.

По таблице распределения Фишера – Снедекора, по уровню значимости  $\alpha=0,05$  и числам степеней свободы  $k_1 = n_1 - 1 = 30 - 1 = 29$  и  $k_2 = n_2 - 1 = 25 - 1 = 24$  находим критическую точку  $F_{\text{кр}}(0,05; 29; 24) = 1,94$ .

Так как  $F_{\text{набл}} < F_{\text{кр}}$ , то нет оснований отвергать гипотезу о равенстве генеральной дисперсии.

### 7. Тема «Математическая статистика в лингвистике»

Прочитать следующий текст:

*Ликуй, ликуй, наш курский соловей,  
Людские души свято возвышая  
До красоты мелодии своей,  
Сравнимой, может, с музыкою рая!*

Выбрав в качестве исследуемого признака длину слова (число букв в каждом слове текста), составить статистический и вариационный ряды, закон распределения частот; найти эмпирическую функцию распределения.

*Решение.*

Статистический ряд: 5, 5, 3, 7, 7, 7, 4, 5, 8, 2, 7, 7, 5, 9, 5, 1, 7, 3.

Вариационный ряд распределения частот:

$x_i$	1	2	3	4	5	7	8	9
$m_i$	1	1	2	1	5	6	1	1

Вариационный ряд распределения относительных частот:

$x_i$	1	2	3	4	5	7	8	9
$w_i$	1/18	1/18	2/18	1/18	5/18	6/18	1/18	1/18

Эмпирическая функция распределения частот имеет вид:

$x$	$F_{\text{ЭМП}}(x)$
$x \leq 1$	0
$1 < x \leq 2$	1/18
$2 < x \leq 3$	1/18+1/18=2/18=1/9
$3 < x \leq 4$	1/18+1/18+2/18=4/18=2/9
$4 < x \leq 5$	1/18+1/18+2/18+1/18=5/18
$5 < x \leq 7$	1/18+1/18+2/18+1/18+5/18=10/18=5/9
$7 < x \leq 8$	1/18+1/18+2/18+1/18+5/18+6/18=16/18=8/9
$8 < x \leq 9$	1/18+1/18+2/18+1/18+5/18+6/18+1/18=17/18
$x > 9$	1/18+1/18+2/18+1/18+5/18+6/18+1/18+1/18=18/18=1

## Контрольные вопросы

1. Основные понятия математической статистики: понятия генеральной и выборочной совокупностей; определение выборки из генеральной совокупности с функцией распределения  $F(x)$ ; дискретный вариационный ряд, статистическое распределение выборки и полигон; эмпирическая функция распределения; интервальный вариационный ряд и гистограмма.
2. Понятие точечной статистической оценки параметра распределения. Требования предъявляемые к статистическим оценкам (несмещенность, состоятельность, эффективность).
3. Точечная статистическая оценка неизвестного математического ожидания и её свойства.
4. Точечные статистические оценки неизвестной дисперсии (среднего квадратического отклонения): выборочная дисперсия; исправленная выборочная дисперсия. Свойства этих оценок.
5. Интервальные статистические оценки параметров распределений. Доверительный интервал. Надежность и точность интервальной оценки.
6. Доверительный интервал для неизвестного математического ожидания нормального распределения (два случая: при известном и неизвестном среднем квадратическом отклонении).
7. Доверительный интервал для неизвестного среднего квадратического отклонения (дисперсии) нормального распределения.
8. Понятие статистической гипотезы. Нулевая и альтернативная гипотезы. Статистический критерий. Критическая область и область принятия гипотезы. Ошибки 1-го и 2-го рода. Общее правило проверки статистической гипотезы.
9. Проверка гипотезы о законе распределения генеральной совокупности по критерию  $\chi^2$ . Проверка гипотезы о нормальности распределения генеральной совокупности.
10. Проверка гипотезы о равенстве дисперсий двух нормально распределенных случайных величин. Критерий Фишера-Снедекора.
11. Проверка гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин (два случая: дисперсии известны; дисперсии неизвестны и одинаковые).
12. Основные задачи теории корреляции.

13. Понятие функциональной, статистической и корреляционной связи между двумя величинами. Их сходство и различие. Построение корреляционной таблицы.
14. Точечные статистические оценки для ковариации и коэффициента корреляции. Их свойства. Эмпирические прямые регрессии.
15. Доверительный интервал для неизвестного коэффициента корреляции двух нормально распределенных случайных величин.
16. Проверка гипотезы о значимости коэффициента корреляции двух нормально распределенных случайных величин.
17. Суть метода наименьших квадратов.

### РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Гмурман В.Е. Теория вероятностей и математическая статистика [Текст]: учебное пособие / В.Е. Гмурман. – М.: Высшая школа, 2012. – 479с.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике [Текст]: учебное пособие / В.Е. Гмурман. – М.: Высшая школа, 2011.–404с.
3. Математика для гуманитариев [Текст]: учебник / К.В. Балдин. – М.: Дашков и К, 2009. – 512с.
4. Сборник задач по математике для вузов [Текст]: учебное пособие. / А.В. Ефимов, А.С. Поспелов. – М.: Физматлит, 2009. Ч.4 – 432с.
5. Калинина В.Н. Математическая статистика [Текст]: учебник / В.Н. Калинина. – М.: Высшая школа, 2001. – 336 с.
6. Колемаев В.А. Теория вероятностей и математическая статистика [Текст]: учебник / В.А. Колемаев, В.Н. Калинина. – М.: ИНФРА-М, 1997. – 302 с.
7. Пиотровский Р.Г. Математическая лингвистика [Текст]: учебное пособие / Р.Г. Пиотровский и др.– М.: Высшая школа, 1977. –383с.
8. Метод наименьших квадратов [Электронный ресурс]: методические указания и индивидуальные задания к ЛР-15 / Юго-Зап. гос. ун-т; сост.: Л.И. Студеникина, Т.В. Шевцова. – Курск: ЮЗГУ, 2011.– 52 с.
9. Элементы математической статистики и корреляционного анализа [Электронный ресурс]: методические указания и индивидуальные задания к М-15 / Юго-Зап. гос. ун-т; сост.: Е.В. Журавлева, Е.А. Панина. – Курск: ЮЗГУ, 2012.– 35 с.