Министерство сельского хозяйства Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Казанский государственный аграрный университет»
Институт механизации и технического сервиса

Кафедра физики и математики

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Методические указания
для практических и самостоятельных работ

Казань, 2016

УДК 51 (07)
ББК 22.01Р

Составитель: Киселева Н.Г.

Под редакцией зав. кафедры физики и математики, д.т.н., профессора Р.И. Ибятова.

Рецензенты:
Доцент кафедры «Технический сервис» Казанского ГАУ, к.т.н. Г.Р. Муртазин.
Доцент кафедры «Анализ данных и исследование операций» Института вычислительной математики и информационных технологий КФУ, к.ф.-м.н., Е.П. Шустова.

Печатается по решению методической комиссии ИМ и ТС (протокол № 8 от 14.03.2016 г.), кафедры физики и математики (протокол № 6 от 11.03.2016 г.).

Киселева Н.Г. Математические методы обработки данных: метод. указания для практ. и самост. работ**.** – Казань: Изд-во Казанского ГАУ, 2016. – 54с.

Методические указания предназначены для студентов всех направлений подготовки Института механизации и технического сервиса, Института экономики, Агрономического факультета и Факультета лесного хозяйства и экологии, обучающихся по программе ФГОС ВО (уровень бакалавриата и магистратуры). Содержат краткие теоретические сведения и разобранные примеры с подробными пояснениями. Приведены задания для самостоятельной работы и вопросы для самоконтроля.

Методические указания также можно использовать для самостоятельной работы аспирантов при изучении дисциплин «Математическое моделирование», «Прикладная математика».

УДК 51 (07)
ББК 22.01Р

СОДЕРЖАНИЕ

Введение

Данные методические указания способствуют формированию общепрофессиональных и профессиональных компетенций обучающихся по направлениям подготовки: 35.03.04 – Агрономия, 35.03.03 – Агрохимия и агропочвоведение, 35.03.05 – Садоводство, 35.03.07 – Технология производства и переработки сельскохозяйственной продукции, 21.03.02 – Землеустройство и кадастры, 35.03.06 – Агроинженерия, 23.03.03 – Эксплуатация транспортно-технологических машин и комплексов, 20.03.01 – Техносферная безопасность, 23.05.01 – Наземные транспортно-технологические средства, 35.03.01 – Лесное дело, 38.03.01 – Экономика, 38.04.02 – Менеджмет, 38.03.04 – Государственное и муниципальное управление.

Методические указания содержат необходимые теоретические сведения и основные формулы по темам «Первичная обработка статистических данных», «Метод произведений и метод сумм для вычислений $\overline{x}_B$, D_B», «Элементы корреляционного анализа», «Регрессионный анализ», «Критерий согласия Пирсона», «Однофакторный дисперсионный анализ». В данных методических указаниях представлены задачи с подробными решениями, что позволяет использовать его не только для аудиторных занятий, но и для самостоятельной работы студентов.

Тема №1. Первичная обработка статистических данных

Цель работы: научиться обрабатывать экспериментальные данные, строить вариационный ряд, эмпирическую функцию распределения, полигон и гистограмму ряда, определять его характеристики.

Краткие теоретические сведения.

Генеральной совокупностью называется вся исследуемая совокупность объектов.

Выборочной совокупностью или просто **выборкой** называют совокупность случайно отобранных объектов.

Объемом совокупности (выборочной или генеральной) называют число объектов этой совокупности.

Пусть из генеральной совокупности извлечена выборка, причем значение x_1 наблюдалось n_1 раз, x_2 наблюдалось n_2 раз, …, $x_к$ наблюдалось $n_к$ раз. Общий объем выборки можно определить как:

$$n = \sum_{i=1}^{k} n_i \ .$$

Наблюдаемые значения x_i называют **вариантами**, а последовательность вариант, записанных в возрастающем порядке, – **вариационным рядом**.

Числа наблюдений называют **частотами**, а их отношения к объему выборки $\frac{n_i}{n} = W_i$ – **относительными частотами**.

Модой M_0 называется варианта, имеющая наибольшую частоту.

Медианой M_e называется варианта, которая делит вариационный ряд на две части с одинаковым числом вариант в каждой. Для вариационного ряда с нечетным числом членов медиана равна серединной варианте, а для ряда с четным числом членов – полусумме двух серединных вариант.

Размахом выборки называется разность между максимальным и минимальным элементами выборки.

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот.

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

или

x_i	x_1	x_2	…	x_k
w_i	w_1	w_2	…	w_k

Статистическое распределение можно задать также в виде последовательности интервалов одинаковой длины и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал).

x	$x_0 - x_1$	$x_1 - x_2$	…	$x_{к-1} - x_к$
n_i	n_1	n_2	…	$n_к$

Алгоритм группировки выборочных данных при построении интервального ряда.

1) Найти наименьшее и наибольшее значения признака в совокупности и определить размах варьирования: $R = x_{max} - x_{min}$.

2) Определить число интервалов k. Для этого используют формулу Стерджесса:

$$k = 1 + 3{,}322 \cdot lgn$$

3) Найти постоянную величину интервала: $h = \frac{R}{k}$.

4) Определить границы интервалов. За начало первого интервала следует взять $x_0 = x_{min}$. Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала h. Конец последнего интервала должен удовлетворять условию: $x_{max} \le x_k$.

5) Подсчитать число выборочных данных, которые попадут в каждый из полученных интервалов: n_1, n_2, …, n_k. При этом только один из промежутков будет замкнут с двух сторон: $[x_0 ; x_1]$, а остальные промежутки будут замкнуты только справа: $(x_i; x_{i+1}]$, $i = 1, …, k$-1.

6) Результаты вычислений занести в таблицу, которая называется «Статистическое распределение интервального ряда».

Эмпирической функцией распределения случайной величины (функцией распределения выборки) называют функцию F_x^* относительной частоты числа наблюдений n_x: $F_x^* = \frac{n_x}{n}$, т.е. относительной частоты события $X < x$.

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$, …, $(x_k; n_k)$. Для построения полигона частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им частоты n_i. Точки $(x_i; n_i)$ соединяют отрезками прямых и получают полигон частот.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1; W_1)$, $(x_2; W_2)$, …, $(x_k; W_k)$. Для построения полигона относительных частот на оси абсцисс откладывают варианты x_i, а на оси ординат – соответствующие им относительные частоты W_i. Точки $(x_i; W_i)$ соединяют отрезками прямых и получают полигон относительных частот.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению n_i / h (плотность частоты).

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению W_i / h (плотность относительной частоты).

Выборочной средней $\bar{x}_в$ называют среднее арифметическое значение признака выборочной совокупности.

Если значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$, причем $n_1 + n_2 + \ldots + n_k = n$, то

$$\bar{x}_в = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n}.$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{x}_в$.

Если значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$, причем $n_1 + n_2 + \ldots + n_k = n$, то

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

Выборочным средним квадратическим отклонением называют квадратный корень из выборочной дисперсии:

$$\sigma_в = \sqrt{D_в}.$$

Пример: Имеется выборка, содержащая 30 числовых значений некоторого признака случайной величины X:

19	25	22	16	22	14	17	19	18	20
22	26	24	18	16	19	22	14	18	14
25	17	18	14	20	18	24	25	16	18

Построить:
1) статистическое распределение выборки;
2) полигон частот;
3) эмпирическую функцию распределения;
4) интервальный ряд;
5) гистограмму частот;

вычислить:
6) выборочную среднюю;
7) выборочную дисперсию;
8) выборочное среднее квадратическое отклонение;
9) моду;
10) медиану.

Решение.

1) Статистическое распределение выборки представляет собой таблицу, в которой первая строка содержит варианты (значения случайной величины, расположенные в порядке возрастания), а вторая – соответствующие частоты (сколько раз эти значения встречались).

x_i	14	16	17	18	19	20	22	24	25	26
n_i	4	3	2	6	3	2	4	2	3	1

2) Полигон частот – это ломаная, соединяющая точки с координатами (x_i; n_i) (рис.1.1).

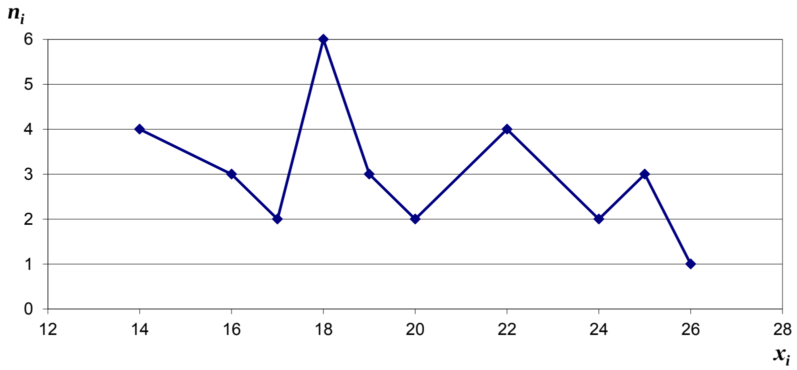

Рисунок 1.1. – Полигон частот.

3) Проведем вычисления для построения эмпирической функции распределения:

$x \le 14$	$F_x^* = 0$
$14 < x \le 16$	$F_x^* = \dfrac{4}{30} = \dfrac{2}{15} \approx 0{,}13$
$16 < x \le 17$	$F_x^* = \dfrac{4+3}{30} = \dfrac{7}{30} \approx 0{,}23$
$17 < x \le 18$	$F_x^* = \dfrac{4+3+2}{30} = \dfrac{9}{30} = 0{,}3$
$18 < x \le 19$	$F_x^* = \dfrac{4+3+2+6}{30} = \dfrac{15}{30} = 0{,}5$
$19 < x \le 20$	$F_x^* = \dfrac{4+3+2+6+3}{30} = \dfrac{18}{30} = 0{,}6$
$20 < x \le 22$	$F_x^* = \dfrac{4+3+2+6+3+2}{30} = \dfrac{20}{30} \approx 0{,}67$

$22 < x \le 24$	$F_x^* = \dfrac{4+3+2+6+3+2+4}{30} = \dfrac{24}{30} = 0{,}8$
$24 < x \le 25$	$F_x^* = \dfrac{4+3+2+6+3+2+4+2}{30} = \dfrac{26}{30} \approx 0{,}87$
$25 < x \le 26$	$F_x^* = \dfrac{4+3+2+6+3+2+4+2+3}{30} = \dfrac{29}{30} \approx 0{,}97$
$x > 26$	$F_x^* = \dfrac{4+3+2+6+3+2+4+2+3+1}{30} = \dfrac{30}{30} = 1$

Получили эмпирическую функцию распределения:

$$F_x^* = \begin{cases} 0, & \text{при } x \le 14 \\ 0{,}13, & \text{при } 14 < x \le 16 \\ 0{,}23, & \text{при } 16 < x \le 17 \\ 0{,}3, & \text{при } 17 < x \le 18 \\ 0{,}5, & \text{при } 18 < x \le 19 \\ 0{,}6, & \text{при } 19 < x \le 20 \\ 0{,}67, & \text{при } 20 < x \le 22 \\ 0{,}8, & \text{при } 22 < x \le 24 \\ 0{,}87, & \text{при } 24 < x \le 25 \\ 0{,}97, & \text{при } 25 < x \le 26 \\ 1, & \text{при } x > 26 \end{cases}$$

4) Построим интервальный ряд.

Найдем наименьшее и наибольшее значения признака в совокупности и определим размах варьирования: $R = x_{max} - x_{min}$. = 26 – 14 = 12.

Определим число интервалов k. Для этого воспользуемся формулой Стерджесса:

$$k = 1 + 3{,}322 \cdot lgn = 1 + 3{,}322 \cdot lg30 \approx 6.$$

Найдем постоянную величину интервала: $h = \dfrac{R}{k} = \dfrac{12}{6} = 2$.

Определим границы интервалов. За начало первого интервала следует взять $x_0 = x_{min}$.= 14. Промежуточные интервалы получаем, прибавляя к концу предыдущего интервала длину частичного интервала $h = 2$:
$x_1 = 14 + 2 = 16$, $x_2 = 16 + 2 = 18$, $x_3 = 18 + 2 = 20$, $x_4 = 20 + 2 = 22$, $x_5 = 22 + 2 = 24$, $x_6 = 24 + 2 = 26$. Будем рассматривать следующие промежутки: [14; 16], (16; 18], (18; 20], (20; 22], (22; 24], (24; 26] и посчитаем количество вариант для каждого промежутка.

Получим статистическое распределение интервального ряда:

x_i	14 –16	16 – 18	18 – 20	20 – 22	22 – 24	24 – 26
n_i	7	8	5	4	2	4

5) Построим гистограмму частот:

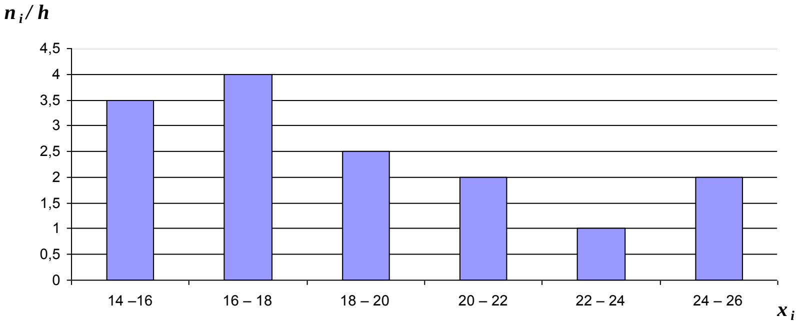

Рисунок 1.2. – Гистограмма частот.

6) Вычислим характеристики ряда. Статистическое распределение ряда, вычисленное выше (пункт 1), имеет вид:

x_i	14	16	17	18	19	20	22	24	25	26
n_i	4	3	2	6	3	2	4	2	3	1

Выборочная средняя:

$$\bar{x}_{\text{в}} = \frac{\sum n_i x_i}{n} = \frac{14\cdot 4 + 16\cdot 3 + 17\cdot 2 + 18\cdot 6 + 19\cdot 3 + 20\cdot 2 + 22\cdot 4 + 24\cdot 2 + 25\cdot 3 + 26\cdot 1}{30} = \frac{580}{30} \approx 19{,}3$$

7) Выборочная дисперсия:

$$D_{\text{в}} = \frac{\sum n_i (x_i - \bar{x}_B)^2}{n} = \frac{4(14-19{,}3)^2 + 3(16-19{,}3)^2 + 2(17-19{,}3)^2 + 6(18-19{,}3)^2 + 3(19-19{,}3)^2 + 2(20-19{,}3)^2 +}{30}$$

$$\frac{+4(22-19{,}3)^2 + 2(24-19{,}3)^2 + 3(25-19{,}3)^2 + 1(26-19{,}3)^2}{} = \frac{382{,}7}{30} \approx 12{,}76\cdot$$

8) Среднее квадратическое отклонение: $\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{12{,}76} \approx 3{,}57$.

9) Запишем вариационный ряд (все значения в порядке возрастания):

14, 14, 14, 14, 16, 16, 16, 17, 17, **18, 18**, **18**, **18**, **18**, **18**, 19, 19, 19, 20, 20, 22, 22, 22, 22, 24, 24, 25, 25, 25, 26

Мода – наиболее часто встречаемое значение, $M_0 = 18$.

10) Медиана – варианта, которая делит вариационный ряд на две части с одинаковым числом вариант в каждой. Для нашего ряда с четным числом членов – медиана равна полусумме двух серединных вариант.

14, 14, 14, 14, 16, 16, 16, 17, 17, 18**,** 18, 18, 18, 18, **18**, **19**, 19, 19, 20, 20, 22, 22, 22, 22, 24, 24, 25, 25, 25, 26

Медиана $M_e = \dfrac{18+19}{2} = 18{,}5$.

Тема №2. «Метод произведений» и «Метод сумм» для вычисления выборочной средней $\overline{x}_B$, выборочной дисперсии D_B.

Цель работы: научиться «методом произведений» и «методом сумм» вычислять выборочную среднюю $\overline{X}_B$ и выборочную дисперсию D_B в случае распределения равностоящих вариант.

Краткие теоретические сведения.

«Метод произведений»

Пусть выборка задана в виде распределения равноотстоящих вариант и соответствующих им частот. В этом случае удобно находить $\overline{X}_B, D_B$ методом произведений по формулам:

$$\overline{X}_B = M_1^* \cdot h + C\ ; \qquad D_B = \left[M_2^* - (M_1^*)^2\right] \cdot h^2,$$

где h – шаг (разность между двумя соседними вариантами);

C – ложный ноль (варианта, которая имеет наибольшую частоту);

$$M_1^* = \frac{\sum n_i u_i}{n}$$ – условный момент I порядка;

$$M_2^* = \frac{\sum n_i u_i^{\,2}}{n}$$ – условный момент II порядка.

«Метод сумм»

Пусть выборка задана в виде распределения равноотстоящих вариант и соответствующих им частот. Для вычисления $\overline{X}_B, D_B$ воспользуемся методом сумм по формулам:

$$\overline{X}_B = M_1^* \cdot h + C\ ; \qquad D_B = \left[M_2^* - (M_1^*)^2\right] \cdot h^2,$$

где h – шаг (разность между двумя соседними вариантами);

C – ложный ноль (варианта, которая имеет наибольшую частоту);

$$M_1^* = \frac{d_1}{n}$$ – условный момент I порядка;

$$M_2^* = \frac{(S_1 + 2S_2)}{n}$$ – условный момент II порядка;

$d_1 = a_1 - b_1$, $S_1 = a_1 + b_1$, $S_2 = a_2 + b_2$. Таким образом, в конечном счете надо вычислить числа a_1, a_2, b_1, b_2.

Пример 1. По данной выборке вычислить «методом произведений» $\overline{X}_B$, D_B:

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

Составим расчётную таблицу:

1) запишем варианты в первый столбец;
2) запишем частоты во второй столбец, сумму частот (100) поместим в нижнюю клетку столбца;
3) в качестве ложного нуля *C* выберем варианту, которая имеет наибольшую частоту (*C* = 16). В клетке третьего столбца, которая принадлежит строке, содержащей ложный нуль, пишем 0; над нулем последовательно запишем: -1, -2; под нулём 1, 2, 3;
4) в четвёртый столбец записываем произведение частот n_i на условные варианты u_i , сумму произведений $u_i \cdot n_i$ (23) записываем в нижнюю клетку столбца;
5) произведение частот на квадраты условных вариант, т.е. $u_i^2 \cdot n_i$ запишем в пятый столбец. Сумму чисел $(n_i \cdot u_i)^2$ (127) записываем в нижнюю клетку столбца;
6) произведения частот на квадраты условных вариант, увеличенных на единицу, т.е. $n_i \cdot (u_i + 1)^2$, запишем в шестой контрольный столбец; сумму чисел столбца (273) помещаем в нижнюю клетку шестого столбца.

Получим следующую таблицу:

x_i	n_i	u_i	$u_i n_i$	$n_i u_i^2$	$n_i(u_i+1)^2$
12	5	-2	-10	20	5
14	15	-1	-15	15	0
16	50	0	0	0	50
18	16	1	16	16	64
20	10	2	20	40	90
22	4	3	12	36	64
	n = 100		$\sum n_i u_i$ = 23	$\sum n_i u_i^2 = 127$	$\sum n_i (u_i + 1)^2 = 273$

Для контроля вычисления пользуются тождеством:

$$\sum n_i (u_i + 1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + n \,.$$

Контроль: $273 = \sum n_i \cdot (u_i + 1)^2$,

$$\sum n_i u_i^2 + 2\sum n_i u_i + n = 127 + 2 \cdot 23 + 100 = 273 \,.$$

Для вычисления $\overline{X}_B$ и D_B воспользуемся формулами:

$$\overline{X}_B = M_1^* \cdot h + C \,; \quad D_B = [M_2^* - (M_1^*)^2] \cdot h^2 \,,$$

где h – шаг (разность между двумя соседними вариантами); C – ложный ноль,

$M_1^* = \dfrac{\sum n_i u_i}{n}$ – условный момент I порядка;

$M_2^* = \dfrac{\sum n_i u_i^2}{n}$ – условный момент II порядка;

$$M_1^* = \frac{23}{100} = 0{,}23; \qquad M_2^* = \frac{127}{100} = 1{,}27\cdot$$

Вычислим $\overline{X}_B$ и D_B (учитывая, что h = 14 – 12 = 2; C = 16):

$\overline{X}_B = M_1^* \cdot h + C = 0{,}23\cdot 2 + 16 = 16{,}46;$

$D_B = \left[M_2^* - (M_1^*)^2\right]\cdot h^2 = \left[1{,}27 - 0{,}23^2\right]\cdot 2^2 = 4{,}8684\cdot$

Пример 2: По данной выборке вычислить «методом сумм» $\overline{X}_B$, D_B:

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

Составим расчётную таблицу:

1) запишем варианты в первый столбец;
2) запишем частоты во второй столбец, сумму частот (100) поместим в нижнюю клетку столбца;
3) в качестве ложного нуля C выберем варианту, которая имеет наибольшую частоту (C = 16). В клетках строки, содержащей ложный нуль, запишем нули;
4) в четвёртом столбце над и под уже помещенным нулем запишем еще по одному нулю;
5) в оставшихся незаполненными над нулем клетках третьего столбца (исключая самую верхнюю) запишем последовательно накопленные частоты: 5; 5 + 15 = 20; сложив все накопленные частоты, получим число b_1 = 25, которое поместим в верхнюю клетку третьего столбца. В оставшихся незаполненными под нулем клетках третьего столбца (исключая самую нижнюю) запишем последовательно накопленные частоты: 4; 4 + 10 = 14; 14 + 16 = 30; сложив все накопленные частоты, получим число a_1 = 48, которое поместим в нижнюю клетку третьего столбца;
6) аналогично заполняется четвертый столбец, причем суммируют частоты третьего столбца; сложив все накопленные частоты, расположенные над нулем, получим число b_2 = 5, которое поместим в верхнюю клетку четвертого столбца; сумма накопленных частот, расположенных под нулем, равна числу a_2 = 22, которое поместим в нижнюю клетку четвертого столбца.

Получим следующую таблицу:

x_i	n_i	$b_1 = 25$	$b_2 = 5$
12	5	5	5
14	15	20	0
16	50	0	0
18	16	30	0
20	10	14	18
22	4	4	4
	$n = 100$	$a_1 = 48$	$a_2 = 22$

Найдем $d_1 = a_1 - b_1 = 48 - 25 = 23$;

$S_1 = a_1 + b_1 = 48 + 25 = 73$;

$S_2 = a_2 + b_2 = 22 + 5 = 27$.

Найдем условные моменты первого и второго порядков:

$M_1^* = d_1 / n = 23/100 = 0{,}23$; $\quad M_2^* = (S_1 + 2S_2)/n = (73 + 2\cdot 27) / 100 = 1{,}27$.

Вычислим $\overline{X}_B$ и D_B (учитывая, что $h = 14 - 12 = 2$; $C = 16$):

$\overline{X}_B = M_1^* \cdot h + C = 0{,}23\cdot 2 + 16 = 16{,}46$;

$D_B = \left[M_2^* - (M_1^*)^2\right] \cdot h^2 = \left[1{,}27 - 0{,}23^2\right] \cdot 2^2 = 4{,}8684$.

Тема №3. Элементы корреляционного анализа

Цель: определить тесноту связи между признаками *X* и *Y*.

Краткие теоретические сведения.

Все явления и процессы, которые существуют в природе и обществе, взаимосвязанные. Изучение взаимосвязей – важная задача статистики, которую она решает с помощью особых методов. Существуют *функциональные* и *корреляционные* связи.

При *функциональной* связи каждому значению *x* соответствует ровно одно возможное значение *y*, которое может быть вычислено по точной формуле $y = f(x)$. *Например*, в функции $y = 2x$ каждому значению *x* соответствует в два раза большее значение *y*.

Но такого рода функциональные (однозначные) связи между переменными величинами встречаются не всегда. *Например*, при одном и том же росте масса различных индивидуумов может быть различна. В данном случае, можно и целесообразно вести речь о *среднем значении* массы индивидуумов.

Корреляционная связь проявляется в том, что при изменении одной из величин изменяется среднее значение другой.

Виды корреляционных связей:

1) по форме может быть прямолинейной или криволинейной;

2) по направлению может быть положительной («прямой») или отрицательной («обратной»).

Линейной называется связь, которая графически изображается прямой линией.

Нелинейной (криволинейной) называется связь, которая графически изображается любой другой линией, кроме прямой (гиперболой, параболой и т.д.).

Связь называется *положительной* («*прямой*»), если при увеличении одного параметра второй тоже увеличивается. Связь называется *отрицательной* («*обратной*»), если при увеличении одного параметра второй уменьшается.

Математической мерой корреляции двух случайных величин служит коэффициент корреляции. Обозначается символом ***r***. Коэффициент корреляции ***r*** характеризует величину, отражающую степень взаимосвязи двух переменных между собой. Он может варьировать в пределах от -1 (отрицательная корреляция) до +1 (положительная корреляция). В зависимости от того, насколько ***r*** приближается к 1, различают связи:

$0{,}91 \leq |\boldsymbol{r}| < 1$ очень сильная связь;

$0{,}71 \leq |\boldsymbol{r}| < 0{,}9$ сильная связь;

$0{,}51 \leq |\boldsymbol{r}| < 0{,}7$ значительная связь;

$0{,}31 \leq |\boldsymbol{r}| < 0{,}5$ умеренная связь;

$|\boldsymbol{r}| < 0{,}3$ связь очень слабая.

Если коэффициент корреляции равен 0, то это говорит об отсутствии корреляционных связей между переменными.

При линейной зависимости между признаками коэффициент корреляции определяется методом квадратов (метод Пирсона).

Метод квадратов (*метод Пирсона*).

1) определить для каждого вариационного ряда средние значения $\bar{x}$ и $\bar{y}$;

2) найти отклонения $(x-\bar{x})$ и $(y-\bar{y})$ каждого числового значения от среднего значения своего вариационного ряда;

3) полученные отклонения перемножить и суммировать;

4) каждое отклонение возвести в квадрат и суммировать по каждому ряду.

Подставить полученные значения в формулу расчета коэффициента корреляции:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}.$$

ОШИБКА КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}},$$

где n – число парных вариант.

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Оценка значимости коэффициента корреляции проводится по критерию Стьюдента путем сравнения его значения с величиной случайной ошибки:

$$t_r = \frac{r}{m_r}.$$

Статистика t_r подчиняется распределению Стьюдента с числом степени свободы $\nu = n-2$.

Из таблицы распределения критерия Стьюдента (приложение 1) находят критическое значение $t_{кр} = t(\alpha, \nu)$.

Если $|t_r| > t_{кр}$, то параметр r статистически значим. В противном случае ($|t_r| < t_{кр}$) параметр r статистически незначим.

Пример. Найти коэффициент корреляции по данным n = 10 наблюдений, которые получены при изучении зависимости между ростом (X, см) и массой (Y, кг) некоторых животных:

x	31	32	33	34	35	36	40	41	42	46
y	7,8	8,3	7,6	9,1	9,6	9,8	11,8	12,1	14,7	13,0

Решение.

Составим расчетную вспомогательную таблицу:

№	x_i	y_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
1	31	7,8	-6	36	-2,58	6,6564	15,48
2	32	8,3	-5	25	-2,08	4,3264	10,4
3	33	7,6	-4	16	-2,78	7,7284	11,12
4	34	9,1	-3	9	-1,28	1,6384	3,84
5	35	9,6	-2	4	-0,78	0,6084	1,56
6	36	9,8	-1	1	-0,58	0,3364	0,58
7	40	11,8	3	9	1,42	2,0164	4,26
8	41	12,1	4	16	1,72	2,9584	6,88
9	42	14,7	5	25	4,32	18,6624	21,6
10	46	13,0	9	81	2,62	6,8644	23,58
Σ	**370**	**103,8**	**-**	**222**	**-**	**51,796**	**99,3**

Вычислим $\bar{x} = \frac{370}{10} = = 37$, $\bar{y} = \frac{103{,}8}{10} = = 10{,}38$.

Подставляем в формулу:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{99{,}3}{\sqrt{222 \cdot 51{,}796}} = \frac{99{,}3}{107{,}232} = 0{,}926.$$

Вывод: связь положительная (прямая, т.е. чем больше рост животных, тем больше у них масса), очень тесная.

ОШИБКА КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r:

$$m_r = \sqrt{\frac{1-0{,}926^2}{10-2}} = 0{,}1335.$$

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Оценка значимости коэффициента корреляции проводится по критерию Стьюдента путем сравнения его значения с величиной случайной ошибки:

$$t_r = \frac{r}{m_r} = \frac{0{,}926}{0{,}1335} = 6{,}94.$$

Статистика t_r подчиняется распределению Стьюдента с числом степени свободы $\nu = n-2$.

Из таблицы распределения критерия Стьюдента (приложение 1) находят критическое значение $t_{кр} = t(\alpha, \nu)$ =2,31.

Так как $|t_r| > t_{кр}$, то параметр *r* статистически значим.

Тема №4. Регрессионный анализ

Цель: разработать регрессионные модели: 1) линейную $\hat{y}_i = b_0 + b_1 x_i$,

2) логарифмическую $\hat{y}_i = b_0 + b_1 \cdot \ln x_i$,

3) гиперболическую $\hat{y}_i = b_0 + \frac{b_1}{x_i}$.

Краткие теоретические сведения.

1) Для построения модели линейной регрессии воспользуемся методом наименьших квадратов. Составим систему уравнений:

$$\begin{cases} b_0 n + b_1 \sum x_i = \sum y_i, \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i. \end{cases}$$

Из системы уравнений находим коэффициенты b_0 и b_1 по формулам:

$$b_0 = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i y_i \cdot \sum x_i}{n\sum x_i^2 - (\sum x_i)^2}, \; b_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

Получим регрессионную линейную модель $\hat{y}_i = b_0 + b_1 x_i$.

2) Для построения логарифмической модели регрессии воспользуемся методом наименьших квадратов. Составим систему уравнений:

$$\begin{cases} b_1 \sum \ln x_i + b_0 n = \sum y_i, \\ b_1 \sum \ln x_i^2 + b_0 \sum \ln x_i = \sum y_i \ln x_i. \end{cases}$$

Из системы уравнений находим коэффициенты b_0 и b_1:

$$b_0 = \frac{\sum y_i \cdot \sum \ln x_i - n \cdot \sum y_i \ln x_i}{D},$$

$$b_1 = \frac{\sum \ln x_i \cdot \sum y_i \ln x_i - \sum y_i \cdot \sum (\ln x_i)^2}{D},$$

где $D = \left(\sum \ln x_i\right)^2 - n \sum \left(\ln x_i\right)^2$.

Получим уравнение $\hat{y}_i = b_0 + b_1 \cdot \ln x_i$.

3) Для построения гиперболической модели регрессии воспользуемся методом наименьших квадратов. Составим систему уравнений:

$$\begin{cases} b_1 \sum \frac{1}{x_i} + b_0 n = \sum y_i, \\ b_1 \sum \frac{1}{x_i^2} + b_0 \sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i}. \end{cases}$$

Из системы уравнений находим коэффициенты b_0 и b_1:

$$b_0 = \frac{1}{D} \cdot \left(\sum \frac{1}{x_i} \cdot \sum \frac{y_i}{x_i} - \sum y_i \cdot \sum \frac{1}{x_i^2} \right),$$

$$b_1 = \frac{1}{D} \cdot \left(\sum y_i \cdot \sum \frac{1}{x_i} - n \cdot \sum \frac{y_i}{x_i} \right),$$

где $D = \left(\sum \frac{1}{x_i} \right)^2 - n \cdot \sum \frac{1}{x_i^2}$. Получим уравнение $\hat{y}_i = b_0 + \frac{b_1}{x_i}$.

ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Для оценки качества модели определяют среднюю ошибку аппроксимации – это среднее отклонение расчетных данных от фактических. Чем меньше эти отклонения, тем ближе теоретические значения подходят к эмпирическим данным, это лучшее качество модели.

Среднюю ошибку аппроксимации рассчитывают по формуле:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\%.$$

Допустимый предел средней ошибки аппроксимации 8-10%. На практике для грубого приближения регрессии к реальной зависимости полагают, что значение средней ошибки аппроксимации не должно превышать 12-15%. Маленькая ошибка аппроксимации указывает на хорошие перспективы при использовании модели для прогнозных расчетов.

ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛИ

Оценка адекватности уравнения регрессии в целом производится на основе *F*-критерия Фишера:

$$F = \frac{\overline{S}_y^2}{\overline{S}_{ост}^2}$$

(отношение общей и остаточной дисперсий показывает, во сколько раз уравнение регрессии предсказывает результаты опыта лучше, чем среднее $\bar{y}$),

где $\overline{S}_y^2 = \dfrac{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}{k_1}$ – общая дисперсия, $\overline{S}_{ост}^2 = \dfrac{\sum (y_i - \hat{y}_i)^2}{k_2}$ – остаточная дисперсия.

Вычисленное значение *F*-критерия Фишера сравнивается с табличным значением (приложение 2): $F_{табл}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = n-1, k_2 = n-m-1$, где n – число наблюдений, m – число параметров при переменной x (в нашем случае $m = 1$).

Если вычисленное значение $F > F_{табл}$, то полученное уравнение регрессии адекватно (статистически значимо). В противном случае ($F < F_{табл}$) уравнение регрессии неадекватно (статистически незначимо).

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Оценка значимости коэффициентов регрессии проводится по критерию Стьюдента путем сравнения их значений с величиной случайной ошибки:

$$t_{b0} = \frac{b_0}{m_{b_0}},\ t_{b1} = \frac{b_1}{m_{b_1}}.$$

Случайные ошибки коэффициентов регрессии определяются по формулам:

$$m_{b0} = \frac{\overline{S}^{2}_{\hat{\imath}\tilde{n}\grave{o}} \cdot \sqrt{\sum x_i^2}}{n\sigma_x},\ m_{b1} = \frac{\overline{S}^{2}_{\hat{\imath}\tilde{n}\grave{o}}}{\sigma_{\tilde{o}} \cdot \sqrt{n}},$$

где $\overline{S}^{2}_{\hat{\imath}\tilde{n}\grave{o}} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$ – остаточная дисперсия, $\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2}$ – среднее квадратическое отклонение.

Статистики t_{b_0} и t_{b_1} подчиняются распределению Стьюдента с числом степени свободы $\nu = n - 2$. Из таблицы распределения критерия Стьюдента (приложение 1) находят критическое значение $t_{кр} = t(\alpha, \nu)$.

Если $\left|t_{b_0}\right| > t_{\hat{e}\eth}$, то параметр b_0 статистически значим. В противном случае ($\left|t_{b0}\right| < t_{\hat{e}\eth}$) параметр b_0 статистически незначим.

Если $\left|t_{b1}\right| > t_{\hat{e}\eth}$, то параметр b_1 статистически значим. В противном случае ($\left|t_{b1}\right| < t_{\hat{e}\eth}$) параметр b_1 статистически незначим.

***Пример* 1.** Найти уравнение прямой линии регрессии по данным n = 8 наблюдений, которые получены при изучении зависимости количества продаж товара y от затрат на рекламу этого товара x:

x	1,5	4,0	5,0	7,0	8,5	10,0	11,0	12,5
y	5,0	4,5	7,0	6,5	9,5	9,0	11,0	9,0

***Решение*.**

Составим расчетную вспомогательную таблицу:

№ п/п	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$\hat{y}_i = a + bx_i$	$(y_i - \hat{y}_i)^2$	A_i, %
1	1,5	5,0	2,25	25	7,5	4,525	0,2256	9,50
2	4,0	4,5	16	20,25	18	5,85	1,8225	30,00
3	5,0	7,0	25	49	35	6,38	0,3844	8,86
4	7,0	6,5	49	42,25	45,5	7,44	0,8836	14,46
5	8,5	9,5	72,25	90,25	80,75	8,235	1,6002	13,32
6	10,0	9,0	100	81	90	9,03	0,0009	0,33
7	11,0	11,0	121	121	121	9,56	2,0736	13,09
8	12,5	9,0	156,25	81	112,5	10,355	1,8360	15,06
$\sum$	**59,5**	**61,5**	**541,75**	**509,75**	**510,25**	**-**	**8,8269**	**104,61**

$$\bar{x}=\frac{\sum x_i}{n}=\frac{59{,}5}{8}=7{,}44;\quad \bar{\grave{A}}=\frac{\sum \grave{A}_i}{n}=\frac{104{,}61}{8}=13{,}08.$$

Вычислим коэффициенты b_0 и b_1 по формулам:

$$b_0=\frac{\sum y_i\cdot\sum x_i^2-\sum x_i y_i\cdot\sum x_i}{n\sum x_i^2-(\sum x_i)^2}=\frac{61{,}5\cdot 541{,}75-510{,}75\cdot 59{,}5}{8\cdot 541{,}75-(59{,}5)^2}=3{,}73,$$

$$b_1=\frac{n\sum x_i y_i-\sum x_i\cdot\sum y_i}{n\sum x_i^2-(\sum x_i)^2}=\frac{8\cdot 510{,}25-59{,}5\cdot 61{,}5}{8\cdot 541{,}75-(59{,}5)^2}=0{,}53.$$

Получим уравнение вида: $\hat{y}_i=3{,}73+0{,}53x_i$.

ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Качество построенной модели оценим по средней ошибке аппроксимации:

$$\bar{\grave{A}}=\frac{1}{n}\sum\left|\frac{y_i-\hat{y}_i}{y_i}\right|\cdot 100\%=\frac{\sum A_i}{n}=\frac{104{,}61}{8}=13{,}08\%.$$

Вывод: <u>в среднем расчетные значения отклоняются от фактических на 13,08%.</u>

Прямая, построенная по этому уравнению, показана на рисунке вместе с исходными данными (рис. 4.1).

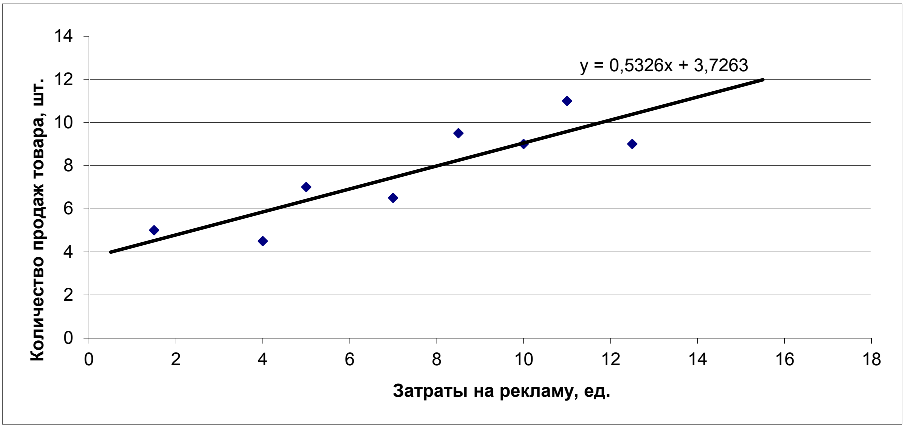

Рисунок 4.1. – Зависимость количества продаж товара от затрат на рекламу.

ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛИ

Число степеней свободы $k_1 = n - 1 = 8 - 1 = 7$

$$k_2 = n - 2 = 8 - 2 = 6$$

Общая дисперсия $\overline{S}_y^2 = \dfrac{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}{k_1} = \dfrac{509{,}75 - \frac{1}{8}\cdot 61{,}5^2}{7} = 5{,}28.$

Остаточная дисперсия $\overline{S}_{\text{îñò}}^2 = \dfrac{\sum \left(y_i - \hat{y}_i\right)^2}{k_2} = \dfrac{8{,}8269}{6} = 1{,}47.$

F – критерий Фишера $F = \dfrac{\overline{S}_y^2}{\overline{S}_{ост}^2} = \dfrac{5{,}28}{1{,}47} = 3{,}59.$

Из таблицы значений F-критерия Фишера (приложение 2) находим $F_{табл}(\alpha; k_1; k_2) = 4{,}21$.

Вывод: $F < F_{табл} (3{,}59 < 4{,}21)$ линейное уравнение регрессии адекватно описывает фактические значения количества продаж товара от затрат на рекламу этого товара.

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Случайные ошибки коэффициентов линейной регрессии: $m_{b_0} = \dfrac{\overline{S}_{\text{îñò}}^2 \cdot \sqrt{\sum x_i^2}}{n\sigma_x}$, $m_{b1} = \dfrac{\overline{S}_{\text{îñò}}^2}{\sigma_x \cdot \sqrt{n}}$.

Вычислим среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - (\overline{x})^2} = \sqrt{\frac{1}{8}\cdot 541{,}75 - 7{,}44^2} = 3{,}52.$$

Тогда $m_{b_0} = \dfrac{\overline{S}_{\text{îñò}}^2 \cdot \sqrt{\sum x_i^2}}{n\sigma_x} = \dfrac{1{,}47\cdot\sqrt{541{,}75}}{8\cdot 3{,}52} = 1{,}22$,

$$m_{b1} = \frac{\overline{S}_{\text{îñò}}^2}{\sigma_{\text{õ}}\cdot\sqrt{n}} = \frac{1{,}47}{3{,}52\cdot\sqrt{8}} = 0{,}15.$$

Оценим значимости коэффициентов регрессии по критерию Стьюдента путем сравнения их значений с величиной случайной ошибки:

$$t_{b_0} = \frac{b_0}{m_{b_0}} = \frac{3{,}73}{1{,}22} = 3{,}06,\ t_{b1} = \frac{b_1}{m_{b1}} = \frac{0{,}53}{0{,}15} = 3{,}53.$$

Из таблицы распределения критерия Стьюдента (приложение 1) находим критическое значение: $t_{кр} = t(\alpha, \nu) = t(\alpha, n-2) = t(0{,}05; 6) = 2{,}45$.

Так как $\left|t_{b_0}\right| > t_{\hat{e}\eth}$ и $\left|t_{b_1}\right| > t_{\hat{e}\eth}$, то параметры a и b статистически значимы.

Вывод: модель по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений к осуществлению прогнозов.

***Пример* 2.** Найти логарифмическое уравнение линии регрессии по данным $n = 8$ наблюдений, которые получены при изучении зависимости количества продаж товара y от затрат на рекламу этого товара x:

x	1,5	4,0	5,0	7,0	8,5	10,0	11,0	12,5
y	5,0	4,5	7,0	6,5	9,5	9,0	11,0	9,0

***Решение*.**

Составим расчетную вспомогательную таблицу:

№ п/п	x_i	y_i	$\ln x_i$	$(\ln x_i)^2$	$y_i \cdot \ln x_i$	$\hat{y}_i = b_0 + b_1 \ln x_i$	$(y_i - \hat{y}_i)^2$	A_i, %
1	1,5	5,0	0,4055	0,1644	2,0273	3,8040	1,4304	36,40
2	4,0	4,5	1,3863	1,9218	6,2383	6,4650	3,8613	3,75
3	5,0	7,0	1,6094	2,5903	11,2661	7,0704	0,0050	34,38
4	7,0	6,5	1,9459	3,7866	12,6484	7,9833	2,2000	23,26
5	8,5	9,5	2,1401	4,5799	20,3306	8,5100	0,9801	45,09
6	10,0	9,0	2,3026	5,3019	20,7233	8,9509	0,0024	39,92
7	11,0	11,0	2,3979	5,7499	26,3768	9,2095	3,2059	49,83
8	12,5	9,0	2,5257	6,3793	22,7316	9,5563	0,3095	37,01
$\sum$	**59,5**	**61,5**	**14,7134**	**30,4741**	**122,3424**	**-**	**11,9946**	**269,64**

$$b_0 = \frac{\sum y_i \cdot \sum \ln x_i - n \cdot \sum y_i \ln x_i}{D} = \frac{61{,}5 \cdot 14{,}7134 - 8 \cdot 0122{,}3424}{14{,}7134^2 - 8 \cdot 30{,}4741} = 2{,}704,$$

$$b_1 = \frac{\sum \ln x_i \cdot \sum y_i \ln x_i - \sum y_i \cdot \sum (\ln x_i)^2}{D} = \frac{14{,}7134 \cdot 122{,}3424 - 61{,}5 \cdot 30{,}4741}{14{,}7134^2 - 8 \cdot 30{,}4741} = 2{,}713.$$

Получим уравнение вида $\hat{y}_i = 2{,}704 + 2{,}703 \ln x_i$.

ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Качество построенной модели оценим по средней ошибке аппроксимации:

$$\overline{\hat{A}} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_i}{y} \right| \cdot 100\% = \frac{\sum A_i}{n} = \frac{269{,}64}{8} = 33{,}71\%.$$

Вывод: в среднем расчетные значения отклоняются от фактических на 33,71%.

ОЦЕНКА АДЕКВАТНОСТИ УРАВНЕНИЯ

Число степеней свободы $k_1 = n - 1 = 8 - 1 = 7$

$k_2 = n - 2 = 8 - 2 = 6$

Общая дисперсия $\overline{S}_y^2 = \dfrac{\sum y_i^2 - \dfrac{1}{n}\left(\sum y_i\right)^2}{k_1} = \dfrac{509{,}75 - \dfrac{1}{8}\cdot 61{,},5^2}{7} = 5{,}28.$

Остаточная дисперсия $\overline{S}_{ост}^2 = \dfrac{\sum \left(y_i - \hat{y}_i\right)^2}{k_2} = \dfrac{11{,}9946}{6} = 2{,}00.$

F–критерий Фишера $F_{табл} = \dfrac{\overline{S}_y^2}{\overline{S}_{ост}^2} = \dfrac{5{,}28}{2{,}00} = 2{,}64.$

Из таблицы значений F-критерия Фишера (приложение 2) находим $F_{табл}(\alpha;\, k_1;\, k_2) = 4{,}21$.

Вывод: $F_{табл} < F_{k_1;k_2}$ $(2{,}64 < 4{,}21)$ логарифмическое уравнение регрессии адекватно описывает фактические значения количества продаж товара от затрат на рекламу этого товара.

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Случайные ошибки коэффициентов регрессии:

$$m_{b_0} = \frac{\overline{S}_{ост}^2 \cdot \sqrt{\sum x_i^2}}{n\sigma_x},\quad m_{b_1} = \frac{\overline{S}_{ост}^2}{\sigma_x \cdot \sqrt{n}}.$$

Вычислим среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\frac{1}{n}\sum x_i^2 - (\overline{x})^2} = \sqrt{\frac{1}{8}\cdot 541{,}75 - 7{,}44^2} = 3{,}52.$$

Тогда

$$m_{b0} = \frac{\overline{S}_{ост}^2 \cdot \sqrt{\sum x_i^2}}{n\sigma_x} = \frac{2{,}00 \cdot \sqrt{541{,}75}}{8 \cdot 3{,}52} = 1{,}65,$$

$$m_{b1} = \frac{\overline{S}_{ост}^2}{\sigma_x \cdot \sqrt{n}} = \frac{2{,}00}{3{,}52 \cdot \sqrt{8}} = 0{,}06 \cdot$$

Оценим значимости коэффициентов регрессии по критерию Стьюдента путем сравнения их значений с величиной случайной ошибки:

$$t_{b_0} = \frac{b_0}{m_{b_0}} = \frac{2{,}704}{1{,}65} = 1{,}64,\; t_{b_1} = \frac{b_1}{m_{b1}} = \frac{2{,}703}{0{,}06} = 45{,}05.$$

Из таблицы распределения критерия Стьюдента находим критическое значение: $t_{кр} = t(\alpha,\nu) = t(\alpha, n-2) = t(0{,}05;6) = 2{,}45$.

Так как $\left|t_{b_0}\right| < t_{\text{êð}}$, то параметр b_0 статистически незначим.

Так как $\left|t_{b_1}\right| > t_{\text{êð}}$, то параметр b_1 статистически значим.

Вывод: <u>модель по *F*-критерию Фишера в целом адекватна, но часть коэффициентов регрессии незначимы. В этом случае модель пригодна для принятия некоторых решений, но не для осуществления прогнозов.</u>

***Пример* 3.** Найти гиперболическое уравнение линии регрессии по данным $n = 8$ наблюдений, которые получены при изучении зависимости количества продаж товара y от затрат на рекламу этого товара x:

x	1,5	4,0	5,0	7,0	8,5	10,0	11,0	12,5
y	5,0	4,5	7,0	6,5	9,5	9,0	11,0	9,0

***Решение*.**

Составим расчетную вспомогательную таблицу:

№ п/п	x_i	y_i	x_i^2	$\frac{y_i}{x_i}$	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\hat{y}_i = b_0 + \frac{b_1}{x_i}$	$(y_i - \hat{y}_i)^2$	A_i, %
1	1,5	5,0	2,25	3,33	0,667	0,4444	16,957	142,962	239,14
2	4,0	4,5	16	1,13	0,250	0,0625	12,203	59,329	171,18
3	5,0	7,0	25	1,40	0,200	0,0400	11,632	21,455	66,17
4	7,0	6,5	49	0,93	0,143	0,0204	10,980	20,070	68,92
5	8,5	9,5	72,25	1,12	0,118	0,0138	10,692	1,422	12,55
6	10,0	9,0	100	0,90	0,100	0,0100	10,491	2,223	16,57
7	11,0	11,0	121	1,00	0,091	0,0083	10,387	0,375	5,57
8	12,5	9,0	156,25	0,72	0,080	0,0064	10,263	1,595	14,03
$\sum$	**59,5**	**61,5**	**541,75**	**10,52**	**1,648**	**0,6059**	-	**249,431**	**594,13**

$$b_0 = \frac{1}{D}\cdot\left(\sum\frac{1}{x_i}\cdot\sum\frac{y_i}{x_i} - \sum y_i\cdot\sum\frac{1}{x_i^2}\right) = \frac{1{,}648\cdot 10{,}52 - 61{,}5\cdot 0{,}6059}{1{,}648^2 - 8\cdot 0{,}6059} = 9{,}35,$$

$$b_1 = \frac{1}{D} \cdot \left(\sum y_i \cdot \sum \frac{1}{x_i} - n \cdot \sum \frac{y_i}{x_i} \right) = \frac{61{,}5 \cdot 1{,}648 - 8 \cdot 10{,}52}{1{,}648^2 - 8 \cdot 0{,}6059} = 11{,}41.$$

Получим уравнение вида $\hat{y}_i = 9{,}35 + \frac{11{,}41}{x_i}$.

ОЦЕНКА КАЧЕСТВА МОДЕЛИ

Качество построенной модели оценим по средней ошибке аппроксимации:

$$\overline{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_i}{y} \right| \cdot 100\% = \frac{\sum A_i}{n} = \frac{594{,}13}{8} = 74{,}27\%.$$

Вывод: <u>в среднем расчетные значения отклоняются от фактических на 74,27%.</u>

ОЦЕНКА АДЕКВАТНОСТИ УРАВНЕНИЯ

Число степеней свободы $k_1 = n - 1 = 8 - 1 = 7$

$k_2 = n - 2 = 8 - 2 = 6$

Общая дисперсия $\overline{S}_y^2 = \frac{\sum y_i^2 - \frac{1}{n} \cdot \left(\sum y_i\right)^2}{k_1} = \frac{509{,}75 - \frac{1}{8} \cdot 61{,},5^2}{7} = 5{,}28.$

Остаточная дисперсия $\overline{S}_{\hat{i}\tilde{n}\grave{o}}^2 = \frac{\sum \left(y_i - \hat{y}_i\right)^2}{k_2} = \frac{249{,}431}{6} = 41{,}57.$

F – критерий Фишера $F_{табл} = \frac{\overline{S}_y^2}{\overline{S}_{ост}^2} = \frac{5{,}28}{41{,}57} = 0{,}13.$

Из таблицы значений F-критерия Фишера (приложение 2) находим $F_{табл}(\alpha;\ k_1;\ k_2) = 4{,}21$.

Вывод: $F_{табл} < F_{k_1;k_2}$ $(0{,}13 < 4{,}21)$ <u>логарифмическое уравнение регрессии адекватно описывает фактические значения количества продаж товара от затрат на рекламу этого товара.</u>

ОЦЕНКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Случайные ошибки коэффициентов регрессии:

$$m_{b_0} = \frac{\overline{S}_{\hat{i}\tilde{n}\grave{o}}^2 \cdot \sqrt{\sum x_i^2}}{n\sigma_x},\ m_{b_1} = \frac{\overline{S}_{\hat{i}\tilde{n}\grave{o}}^2}{\sigma_x \cdot \sqrt{n}}.$$

Вычислим среднее квадратическое отклонение:

$$\sigma_x = \sqrt{\frac{1}{n} \sum x_i^2 - (\overline{x})^2} = \sqrt{\frac{1}{8} \cdot 541{,}75 - 7{,}44^2} = 3{,}52.$$

Тогда

$$m_{b0} = \frac{\overline{S}^2_{ост} \cdot \sqrt{\sum x_i^2}}{n\sigma_x} = \frac{41{,}57 \cdot \sqrt{541{,}75}}{8 \cdot 3{,}52} = 34{,}36;$$

$$m_{b1} = \frac{\overline{S}^2_{ост}}{\sigma_x \cdot \sqrt{n}} = \frac{41{,}57}{3{,}52 \cdot \sqrt{8}} = 1{,}19.$$

Оценим значимости коэффициентов регрессии по критерию Стьюдента путем сравнения их значений с величиной случайной ошибки:

$$t_{b_0} = \frac{b_0}{m_{b_0}} = \frac{9{,}35}{34{,}36} = 0{,}27\ ,\ t_{b_1} = \frac{b_1}{m_{b1}} = \frac{11{,}41}{1{,}19} = 9{,}59\,.$$

Из таблицы распределения критерия Стьюдента находим критическое значение: $t_{кр} = t(\alpha,\nu) = t(\alpha, n-2) = t(0{,}05;6) = 2{,}45$.

Так как $\left|t_{b_0}\right| < t_{кр}$, то параметр b_0 статистически незначим.

Так как $\left|t_{b_1}\right| > t_{кр}$, то параметр b_1 статистически значим.

Вывод: модель по F-критерию Фишера в целом адекватна, но часть коэффициентов регрессии незначимы. В этом случае модель пригодна для принятия некоторых решений, но не для осуществления прогнозов.

ВЫБОР ОПТИМАЛЬНОГО УРАВНЕНИЯ

Метод наименьших квадратов позволяет построить уравнение практически любой формы парной связи и выбрать оптимальное уравнение. Его сущность заключается в минимизации отклонений фактических значений y_i от предсказанных $\hat{y}_i$ по одному из уравнений регрессии. Из множества уравнений исследователю необходимо выбрать одно, которое бы хорошо описывало изучаемое явление и логически было обосновано.

Наилучшие результаты опыта дает **уравнение линейной регрессии**, так как оно характеризуется меньшей остаточной дисперсией $(\overline{S}^2_{ост} = 1{,}47)$ и наибольшим фактическим F–критерием ($F_{табл}$= 3,59), а также имеет самую маленькую ошибку аппроксимации – среднее отклонение расчетных данных от фактических (13,08%). По критерию Стьюдента оба коэффициента в уравнении регрессии значимы ($t_1 > t_{кр}$ и $t_2 > t_{кр}$).

Вид уравнения регрессии	Степень свободы		Дисперсия		Критерий Фишера		Средняя Ошибка аппр.	Критерий Стьюдента		
	k_1	k_2	общ	ост.	$F_{табл}$	F_{k_1,k_2}	$\overline{A}_i$, %	t_1	t_2	$t_{кр}$
$\hat{y}_i = 3{,}73 + 0{,}53x_i$	7	6	5,28	1,47	3,59	4,21	13,08	3,06	3,53	2,45
$\hat{y}_i = 2{,}704 + 2{,}703 \ln x$	7	6	5,28	2,00	2,64	4,21	33,71	1,64	45,05	2,45
$\hat{y}_i = 9{,}35 + \dfrac{11{,}41}{x_i}$	7	6	5,28	41,57	0,13	4,21	74,27	0,27	9,59	2,45

Тема №5. Критерий согласия Пирсона

Цель работы: научиться проверять гипотезу о нормальном распределении с помощью критерия Пирсона.

Краткие теоретические сведения.

Критерием согласия Пирсона называется критерий проверки гипотезы о нормальном распределении генеральной совокупности.

Для того, чтобы при заданном уровне значимости α проверить гипотезу о нормальном распределении генеральной совокупности необходимо:

а) вычислить выборочную среднюю $\overline{x}_B$ и выборочное среднее квадратическое отклонение σ_B;

б) вычислить теоретические частоты

$$n_i^{'} = \frac{nh}{\sigma_B}\varphi(u_i),$$

где n – объем выборки (сумма всех частот), h – шаг (разность между двумя соседними вариантами), $u_i = \dfrac{x_i - \overline{x}_B}{\sigma_B}$, $\varphi(u_i) = \dfrac{1}{\sqrt{2\pi}} e^{-u^2/2}$ – определяется по таблице приложения 4.

в) вычислить наблюдаемое значение критерия

$$\chi^2_{набл} = \sum_{i=1}^{S} \frac{(n_i - n_i^{'})^2}{n_i^{'}},$$

где n_i – эмпирические (опытные) частоты, $n_i^{'}$ – теоретические частоты, s – число групп выборки;

г) найти $\chi^2_{кр}(\alpha, k)$, где $k = s - 3$, по таблице критических точек распределения χ^2 (см. приложение 3);

д) если $\chi^2_{набл} < \chi^2_{кр}$ – нет оснований отвергнуть нулевую гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо. Если $\chi^2_{набл} > \chi^2_{кр}$ – гипотезу отвергаем, то есть, эмпирические и теоретические частоты различаются значимо.

Замечание. Малочисленные частоты ($n_i < 5$) следует объединить; в этом случае и соответствующие им теоретические частоты также надо сложить. Если производилось объединение частот, то при определении числа степеней свободы по формуле $k = s - 3$ следует в качестве s принять число групп выборки, оставшихся после объединения частот.

Задание. По данному распределению выборки проверить гипотезу о нормальном распределении генеральной совокупности при уровне значимости $\alpha = 0{,}01$.

x_i	10,6	10,8	11,0	11,2	11,4	11,6	11,8
n_i	5	10	17	30	20	12	6

Решение.

а) Выборочная средняя

$$\bar{x}_в = \frac{\sum n_i x_i}{n} = \frac{10{,}6\cdot 5 + 10{,}8\cdot 10 + 11{,}0\cdot 17 + 11{,}2\cdot 30 + 11{,}4\cdot 20 + 11{,}6\cdot 12 + 11{,}8\cdot 6}{5 + 10 + 17 + 30 + 20 + 12 + 6} = 11{,}22.$$

и выборочная дисперсия

$$D_в = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n} = \frac{5(10{,}6 - 11{,}22)^2 + 10(10{,}8 - 11{,}22)^2 + 17(11 - 11{,}22)^2 + 30(11{,}2 - 11{,}22)^2 +}{100}$$

$$\frac{+\ 20(11{,}4 - 11{,}22)^2 + 12(11{,}6 - 11{,}22)^2 + 6(11{,}8 + 11{,}22)^2}{} = 0{,}0892,$$

тогда выборочное среднее квадратическое отклонение

$\sigma_B = \sqrt{D_B} \approx 0{,}2987.$

б) Вычислим теоретические частоты, учитывая, что n = 100, h = 0,2, $\sigma_B = 0{,}2987$, по формуле $n_i' = \dfrac{nh}{\sigma_B}\varphi(u_i) = \dfrac{100\cdot 0{,}2}{0{,}2987}\varphi(u_i) \approx 67\cdot\varphi(u_i)$.

Составим расчетную таблицу:

x_i	$u_i = \dfrac{x_i - \overline{x}_B}{\sigma_B}$	$\varphi(u_i)$	$n_i' = 67 \cdot \varphi(u_i)$
10,6	-2,08	0,0459	3
10,8	-1,41	0,1476	10
11,0	-0,74	0,3034	21
11,2	-0,07	0,3980	27
11,4	0,60	0,3332	23
11,6	1,27	0,1781	2
11,8	1,94	0,0608	4

в) Для вычисления $\chi^2_{набл} = \sum\limits_{i=1}^{7} \dfrac{(n_i - n_i')^2}{n_i'}$, составим расчетную таблицу:

i	n_i	n_i'	$(n_i - n_i')$	$(n_i - n_i')^2$	$\dfrac{(n_i - n_i')^2}{n_i'}$
1	5	3	2	4	1,333
2	10	10	0	0	0
3	17	21	-4	16	0,762
4	30	27	3	9	0,333
5	20	23	-3	9	0,391
6	12	12	0	0	0
7	6	4	2	4	1,000
	$\sum$=100				$\sum$=3,819

Таким образом, $\chi^2_{набл} = 3{,}819$.

г) По таблице критических точек распределения χ^2 (приложение 3) найдем:

$$\chi^2_{кр}(\alpha, k) = \chi^2_{кр}(0{,}01; 7-3=4) = 13{,}3$$

д) Так как $\chi^2_{набл} < \chi^2_{кр}$ – нет оснований отвергнуть нулевую гипотезу о нормальном распределении генеральной совокупности. Другими словами, эмпирические и теоретические частоты различаются незначимо (случайно).

Тема №6. Однофакторный дисперсионный анализ

Цель работы: научиться анализировать влияние различных факторов на результат эксперимента методами однофакторного анализа для одинакового числа испытаний на всех уровнях.

Краткие теоретические сведения.

Дисперсионный анализ – это статистический метод, позволяющий анализировать влияние различных факторов на результат эксперимента.

Под **фактором** понимают различные, независимые, качественные показатели, влияющие на изучаемые признаки. Факторы обозначаются прописными начальными буквами латинского алфавита: A, B, C, … .

Факторы, контролируемые и измеряемые в процессе исследования, называются **регулируемыми.**

Признаки, изменяющиеся под воздействием тех или иных факторов, называют **результативными**. Для их обозначения используются конечные буквы латинского алфавита X, Y, Z.

Основная идея дисперсионного анализа состоит в сравнении факторной дисперсии, определяемой влиянием регулируемого фактора и остаточной дисперсии, обусловленной действием неконтролируемых и случайных причин.

С помощью F – критерия устанавливается влияние фактора на признак.

Условия применения дисперсионного анализа к выборочным данным:

1) выборочные данные должны быть взяты из совокупностей с нормальным законом распределения;

2) дисперсии всех совокупностей должны быть равны.

Однофакторный дисперсионный анализ решается по следующей схеме:

1) Вычислить общую среднюю: $\overline{x}_B = \sum_{j=1}^{p} \frac{\overline{x}_{гр\, j}}{p}$.

2) Вычислить сумму квадратов отклонений групповых средних от общей средней:

$$S_{общ} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \overline{x}_B)^2 .$$

Имеет место формула: $S_{общ} = \sum_{j=1}^{p} \sum_{i=1}^{q} x_{ij}^2 - pq\overline{x}_B^2$.

3) Вычислить факторную сумму квадратов отклонений групповых средних от общей средней (рассеяние между группами):

$$S_{факт} = q\sum_{j=1}^{p}(x_{гр\,j} - \bar{x}_B)^2 \ .$$

Имеет место формула: $S_{факт} = q\sum_{j=1}^{p}\bar{x}^2_{гр\,j} - pq\bar{x}^2_B$

4) Вычислить остаточную сумму квадратов отклонений значений группы от своей групповой средней (рассеяние внутри групп):

$$S_{ост} = S_{общ} - S_{факт} \ .$$

5) Вычислить факторную, остаточную дисперсии:

$$S^2_{факт} = \frac{S_{факт}}{p-1}\ ,\ S^2_{ост} = \frac{S_{ост}}{p(q-1)}$$

и величину:

$$F_{набл} = \frac{S^2_{факт}}{S^2_{ост}}\ .$$

6) Находим $F_{кр}(\alpha, p-1, p(q-1))$ используя таблицу значений (приложение 2).

Если $F_{набл} < F_{кр}$, то различие групповых средних **незначимое** (фактор оказывает незначительное воздействие на результаты испытаний и его не следует учитывать).

Если $F_{набл} > F_{кр}$, то различие групповых средних **значимое** (фактор оказывает существенное влияние на результаты испытаний и его следует учитывать).

Замечание: Если x_{ij} – десятичные дроби с ***m*** знаками после запятой, то целесообразно перейти к нормированным и центрированным вариантам:

$$y_{ij} = 10^m \cdot (x_{ij} - c)\ .$$

При этом факторная и остаточная дисперсии увеличатся в 10^{2m} раз каждая.

***Пример* 1.** В трех филиалах одного банка были организованы три уровня различных услуг для клиентов. После этого в течение шести месяцев измерялись объемы вкладов (тыс. руб.). Методом дисперсионного анализа проверить нулевую гипотезу о существенном влиянии организации услуг на объем вкладов при уровне значимости 0,05. Данные измерений приведены в таблице:

Номер измерения	Уровни фактора		
	F_1	F_2	F_3
1	16	17	18
2	15	25	27
3	10	15	30
4	20	28	14
5	18	30	40
6	14	22	34

Решение.

1) Имеем $p = 3$, $q = 6$, значения вариант – целые числа. Находим групповые и общую средние:

$$\bar{x}_{гр1} = \frac{16+15+10+20+18+14}{6} = \frac{93}{6} = 15{,}5\,,$$

$$\bar{x}_{гр2} = \frac{17+25+15+28+30+22}{6} = \frac{137}{6} = 22{,}8\,,$$

$$\bar{x}_{гр3} = \frac{18+27+30+30+40+34}{6} = \frac{163}{6} = 27{,}2\,,$$

$$\bar{x}_{B} = \frac{\bar{x}_{гр1}+\bar{x}_{гр2}+\bar{x}_{гр3}}{3} = \frac{15{,}5+22{,}8+27{,}2}{3} = 21{,}8\,.$$

2) Для расчета $S_{общ}$ составим таблицу квадратов:

Номер измерения	Уровни фактора		
	F_1	F_2	F_3
1	256	289	324
2	225	625	729
3	100	225	900
4	400	784	496
5	324	900	1600
6	196	484	1156
Сумма	**1501**	**3307**	**4905**

Получаем

$$S_{общ} = \sum_{j=1}^{p}\sum_{i=1}^{q} x_{ij}^2 - pq\bar{x}_B^2 = 1501+3307+4905-3\cdot 6\cdot 21{,}8^2 = 1158{,}68.$$

3) Вычислим

$$S_{факт} = q\sum_{j=1}^{p} \bar{x}_{гр\,j}^2 - pq\bar{x}_B^2 = 6\cdot(15{,}5^2+22{,}8^2+27{,}2^2-3\cdot 21{,}8^2) = 445{,}26.$$

4) Находим $S_{ост} = S_{общ} - S_{факт} = 1158{,}65 - 445{,}26 = 713{,}42.$

5) Факторная, остаточная дисперсии равны:

$$S_{факт}^2 = \frac{S_{факт}}{p-1} = \frac{445{,}23}{2} = 222{,}63, \quad S_{ост}^2 = \frac{S_{ост}}{p(q-1)} = \frac{713{,}42}{15} = 47{,}56,$$

тогда $F_{набл} = \frac{S_{факт}^2}{S_{ост}^2} = \frac{222{,}63}{47{,}56} = 4{,}68.$

6) По таблице значений (приложение 4) находим $F_{кр}(\alpha, p-1, p(q-1)) = F_{кр}(0{,}05; 2; 15) = 3{,}6$.

Так как $F_{набл} > F_{кр}$, то нулевая гипотеза о существенном влиянии организации услуг на объем вкладов принимается.

***Пример* 2.** Произведено по 4 испытания на каждом из трех уровней фактора F. Методом дисперсионного анализа при уровне значимости 0,05 проверить гипотезу о равенстве групповых средних.

Номер измерения	Уровни фактора		
	F_1	F_2	F_3
1	0,38	0,20	0,21
2	0,36	0,24	0,22
3	0,35	0,26	0,31
4	0,31	0,30	0,34
Групповая средняя	**0,35**	**0,25**	**0,27**

Решение.

Поскольку данные варианты являются дробными числами, то для упрощения вычислений переходим к новым, целым и центрированным вариантам, полагая

$$y_{ij} = 10^m \cdot (x_{ij} - c) = 10^2 \cdot (x_{ij} - 0{,}29) = 100x_{ij} - 29$$

(общее среднее арифметическое значений x_{ij} равно $c = \frac{0{,}35 + 0{,}25 + 0{,}27}{3} = 0{,}29$).

Новые варианты заносим в расчетную таблицу:

Номер измерения	F_1		F_2		F_3	
	U_{i1}	U_{i1}^2	U_{i2}	U_{i2}^2	U_{i3}	U_{i3}^2
1	9	81	-9	81	-8	64
2	7	49	-5	25	-7	49
3	6	36	-3	9	2	4
4	2	4	1	1	5	25
Суммы	**24**	**170**	**-16**	**116**	**-8**	**142**

1) Имеем $p = 3$, $q = 4$.

Находим групповые и общие средние:

$$\overline{U}_{гр1} = \frac{24}{4} = 6\,,\ \overline{U}_{гр2} = -\frac{16}{4} = -4\,,\ \overline{U}_{гр3} = -\frac{8}{4} = -2\,,\ \overline{U}_{B} = \frac{6+(-4)+(-2)}{4} = 0\,.$$

2) Находим

$$S_{U\,общ} = \sum_{j=1}^{p}\sum_{i=1}^{q}(u_{ij} - \overline{u}_B)^2 = (170-0)^2 + (116-0)^2 + (142-0)^2 = 428$$

3) Находим

$$S_{U\,факт} = q\sum_{j=1}^{p}(U_{гр\,j} - \overline{u}_B)^2 = 4\cdot (6-0)^2 + (-4-0)^2 + (-2-0)^2 = 224.$$

4) Вычислим

$$S_{U\,ост} = S_{U\,общ} - S_{U\,факт} = 428 - 224 = 204.$$

5) Факторная, остаточная дисперсии и $F_{набл}$ равны:

$$S^2_{факт} = \frac{S_{U\,факт}}{p-1} = \frac{224}{2} = 112, \qquad S^2_{ост} = \frac{S_{U\,ост}}{p(q-1)} = \frac{204}{6} = 22{,}67,$$

$$F_{набл} = \frac{S^2_{факт}}{S^2_{ост}} = \frac{112}{22{,}67} = 4{,}48.$$

6) По таблице значений (приложение 2) находим

$$F_{кр}(\alpha, p-1, p(q-1)) = F_{кр}(0{,}05; 2; 9) = 4{,}26\,.$$

Так как $F_{набл} > F_{кр}$, то гипотезу о равенстве групповых средних данных совокупностей отвергаем.

Задания для самостоятельной работы

Тема №1. Первичная обработка статистических данных

Задание: Имеется выборка, содержащая 15 числовых значений некоторого признака случайной величины X.

Построить:
1) статистическое распределение выборки;
2) полигон частот;
3) эмпирическую функцию распределения;
4) интервальный ряд;
5) гистограмму частот;

вычислить:
6) выборочную среднюю;
7) выборочную дисперсию;
8) выборочное среднее квадратическое отклонение;
9) моду;
10) медиану.

1.	***X***	17	10	26	20	4	17	20	26	20	4	10	29	20	17	10
2.	***X***	13	8	3	17	8	20	13	8	3	23	8	13	17	8	13
3.	***X***	15	22	27	6	15	36	27	22	15	27	22	27	6	27	31
4.	***X***	14	12	7	14	12	17	14	12	17	22	20	12	14	20	12
5.	***X***	11	5	14	18	23	11	14	5	18	25	11	18	14	18	23
6.	***X***	12	16	23	30	12	5	28	12	16	23	5	12	16	12	16
7.	***X***	13	18	8	21	28	13	21	26	21	8	13	18	21	18	21
8.	***X***	10	6	13	10	17	10	13	20	17	10	20	13	10	21	13
9.	***X***	18	13	24	7	32	24	18	30	7	24	13	18	30	13	24
10.	***X***	15	10	4	15	17	22	10	15	24	10	15	4	10	17	10
11.	***X***	13	7	10	13	3	13	7	17	13	18	3	13	7	10	13
12.	***X***	12	19	12	5	19	24	12	19	35	24	19	30	12	30	12
13.	***X***	14	9	20	14	23	20	26	9	23	14	26	23	20	29	23
14.	***X***	15	10	3	19	15	10	28	15	3	10	15	19	10	24	10
15.	***X***	17	7	14	12	17	20	22	17	14	7	12	17	12	17	14

16.	***X***	16	11	6	25	16	21	11	26	11	16	11	25	16	21	11
17.	***X***	13	16	20	7	13	27	20	24	13	7	20	16	24	20	16
18.	***X***	12	25	8	12	8	25	8	12	3	12	3	8	33	8	30
19.	***X***	19	8	14	24	19	24	33	24	19	8	14	24	30	24	14
20.	***X***	10	14	8	10	14	8	3	10	8	16	8	10	8	18	16
21.	***X***	14	18	22	14	7	18	22	25	14	22	27	25	22	7	18
22.	***X***	15	11	4	11	15	22	15	11	4	15	25	11	29	22	11
23.	***X***	17	24	12	24	17	12	24	4	17	34	24	4	30	24	12
24.	***X***	16	13	8	16	13	18	21	16	13	21	18	13	16	23	13
25.	***X***	10	5	16	10	20	5	22	20	10	22	16	20	25	20	16

Тема №2. «Метод произведений» и «Метод сумм» для вычисления выборочной средней $\overline{X}_B$, выборочной дисперсии D_B

Задание: По данной выборке вычислить «методом произведений» и «методом сумм» $\overline{X}_B$, D_B :

1.

x_i	12	14	16	18	20	22	24
n_i	6	12	18	24	17	13	10

2.

x_i	11	13	15	17	19	21	23
n_i	8	14	21	25	13	10	9

3.

x_i	25	30	35	40	45	50	55
n_i	9	10	16	22	18	15	10

4.

x_i	16	20	24	28	32	36	40
n_i	10	15	16	26	13	11	9

5.

x_i	20	22	24	26	28	30	32
n_i	11	12	20	24	14	12	7

6.

x_i	23	25	27	29	31	33	35
n_i	8	15	18	23	15	12	9

7.

x_i	18	20	22	24	26	28	30
n_i	10	13	16	21	19	13	8

8.

x_i	10	13	16	19	22	25	28
n_i	7	13	19	22	16	13	10

9.

x_i	22	24	26	28	30	32	34
n_i	9	16	19	23	14	11	8

10.

x_i	15	20	25	30	35	40	45
n_i	12	14	19	20	14	12	9

11.

x_i	11	15	19	23	27	31	35
n_i	8	11	18	26	15	12	10

12.

x_i	24	26	28	30	32	34	36
n_i	7	12	19	24	17	12	9

13.

x_i	17	20	23	26	29	32	35
n_i	10	12	17	25	14	12	10

14.

x_i	13	15	17	19	21	23	25
n_i	9	13	17	24	14	13	10

15.

x_i	18	22	26	30	34	38	42
n_i	11	12	16	23	15	14	9

16.

x_i	14	18	22	26	30	34	38
n_i	10	11	19	21	16	12	11

17.

x_i	21	23	25	27	29	31	33
n_i	8	13	19	21	16	13	10

18.

x_i	16	18	20	22	24	26	28
n_i	7	12	14	24	20	12	11

19.

x_i	17	21	25	29	33	37	41
n_i	9	12	15	23	18	15	8

20.

x_i	12	15	18	21	24	27	30
n_i	8	11	19	24	17	12	9

21.

x_i	10	14	18	22	26	30	34
n_i	6	13	18	23	17	14	9

22.

x_i	19	21	23	25	27	29	31
n_i	9	10	14	26	16	15	10

23.

x_i	22	25	28	31	34	37	40
n_i	10	12	16	22	18	13	9

24.

x_i	18	20	22	24	26	28	30
n_i	9	11	13	26	16	15	10

25.

x_i	19	23	27	31	35	39	43
n_i	7	13	20	23	14	12	11

Тема №3. Элементы корреляционного анализа

Задание. Вычислить коэффициент корреляции ***r*** Пирсона:

1.

X	41	40	39	38	36	37	39	36	36	38
Y	47	48	44	45	42	45	50	47	46	46

2.

X	25	30	25	30	35	35	40	40	45	45
Y	36	38	42	36	24	28	24	20	22	20

3.

X	21	20	18	14	20	21	17	18	16	15
Y	19	18	16	13	19	17	16	15	14	13

4.

X	28	23	24	25	26	27	29	23	29	26
Y	30	25	26	28	27	31	30	24	31	28

5.

X	20	20	35	20	30	25	25	27	30	30
Y	18	19	25	20	25	21	23	22	23	24

6.

X	56	58	61	60	59	58	56	57	59	56
Y	46	47	46	48	44	45	42	46	50	46

7.

X	17	14	18	15	16	17	16	19	13	15
Y	30	40	35	40	40	35	35	25	45	38

8.

X	56	56	58	57	57	54	55	52	55	60
Y	38	36	40	41	36	39	38	36	37	39

9.

X	17	18	21	25	29	33	37	41	45	44
Y	36	32	36	46	40	48	50	45	50	46

10.

X	20	19	18	21	23	22	23	24	25	25
Y	40	38	36	45	42	50	49	45	36	40

11.

X	40	39	38	37	35	36	38	35	36	37
Y	46	47	42	44	41	44	48	43	42	45

12.

X	26	32	26	31	34	33	38	40	42	44
Y	35	37	41	36	25	27	26	22	21	24

13.

X	20	22	19	15	20	19	18	17	16	14
Y	18	19	16	13	18	17	16	15	13	15

14.

X	27	24	23	25	24	26	29	24	29	25
Y	30	26	25	27	28	30	28	23	31	27

15.

X	19	20	24	23	28	25	24	26	29	30
Y	18	19	25	21	24	23	22	23	25	26

16.

X	46	48	51	50	49	48	46	47	49	46
Y	35	37	36	38	34	35	32	36	40	36

17.

X	16	15	18	16	15	17	16	18	13	14
Y	26	30	28	29	30	24	25	22	25	28

18.

X	36	36	38	37	37	34	35	32	35	40
Y	28	26	30	31	26	29	28	26	27	29

19.

X	27	28	21	25	29	23	27	21	25	24
Y	36	32	36	33	40	28	40	35	38	36

20.

X	20	19	18	21	23	22	23	24	25	22
Y	30	38	36	35	32	40	39	35	36	40

Тема №4. Регрессионный анализ

Задание. Разработать регрессионные модели: 1) линейную $\hat{y}_i = b_0 + b_1 x_i$, 2) логарифмическую $\hat{y}_i = b_0 + b_1 \cdot \ln x_i$, 3) гиперболическую $\hat{y}_i = b_0 + \dfrac{b_1}{x_i}$. Оценить качество и адекватность моделей, значимость коэффициентов регрессионных моделей, выбрать оптимальное уравнение.

1.

X	18	19	17	18	16	19	18	22	17	18	15	20	18	17	21
Y	14	19	16	14	14	16	14	21	18	16	12	19	18	18	18

2.

X	22	23	20	22	21	20	21	18	16	22	18	25	13	23	17
Y	24	21	21	19	23	23	25	21	20	19	24	23	18	23	18

3.

X	17	20	19	22	18	16	23	25	14	16	22	18	24	22	23
Y	19	22	18	24	22	17	20	22	19	21	19	21	19	19	22

4.

X	16	13	16	14	19	17	21	17	21	10	20	19	11	13	20
Y	15	16	17	14	18	16	18	16	21	11	17	21	12	12	17

5.

X	22	15	19	21	23	18	23	19	18	21	17	20	16	22	20
Y	18	13	15	16	19	17	19	18	15	18	13	17	15	20	19

6.

X	18	13	16	19	20	20	17	17	18	15	16	15	15	19	16
Y	17	14	17	18	22	19	18	21	20	17	19	17	14	21	15

7.

X	14	17	17	15	19	16	18	17	16	18	21	20	15	19	21
Y	14	18	20	14	20	15	18	15	17	16	19	17	17	18	17

8.

X	19	20	15	16	15	14	16	18	19	13	14	18	17	18	17
Y	17	19	17	15	16	16	18	19	20	14	15	19	20	17	16

9.

X	19	22	16	14	21	20	18	17	15	22	15	18	16	20	17
Y	19	21	18	15	20	21	19	16	17	19	18	19	15	17	20

10

X	15	21	16	14	19	20	18	17	20	22	19	22	23	20	17
Y	16	18	18	15	16	21	19	15	18	20	20	21	21	21	18

11

X	16	22	18	18	14	20	19	18	20	17	16	15	19	17	20
Y	17	20	16	17	16	22	19	20	20	21	19	17	17	18	18

12	X	18	19	21	18	16	19	18	21	17	18	16	22	16	20	21
	Y	14	17	18	14	14	15	18	20	15	16	17	19	17	19	17

13	X	22	18	19	20	21	20	21	19	17	22	18	16	15	21	17
	Y	18	17	17	19	20	17	17	20	16	19	16	17	16	19	18

14	X	20	21	17	19	19	16	23	23	15	16	20	18	14	18	22
	Y	22	20	18	20	22	19	20	21	19	21	19	21	19	19	21

15	X	16	14	15	18	16	17	21	17	21	18	20	19	17	18	20
	Y	15	15	17	18	18	16	18	18	20	16	19	18	15	20	17

16	X	14	15	19	21	23	21	23	17	20	18	20	17	16	24	16
	Y	15	18	20	18	20	22	19	18	21	17	18	20	15	21	19

17	X	22	21	16	18	20	17	16	19	18	15	16	16	17	20	15
	Y	21	20	20	19	21	19	17	21	20	17	19	17	18	19	19

18	X	15	22	18	18	20	14	16	18	16	17	21	20	15	19	17
	Y	14	20	19	17	21	15	15	16	17	18	19	18	17	18	15

19	X	19	20	16	19	15	14	16	16	18	13	17	18	17	18	15
	Y	18	19	17	20	17	18	19	19	20	15	16	19	19	17	18

20	X	16	19	15	14	16	20	17	16	18	17	16	15	13	19	17
	Y	19	20	18	16	17	21	19	17	19	18	18	17	15	21	21

21	X	17	22	16	14	21	20	18	16	15	23	19	25	22	20	16
	Y	19	20	18	17	18	21	19	18	18	19	20	22	21	19	18

22	X	17	21	19	18	15	23	24	22	20	18	17	16	19	20	21
	Y	15	20	18	17	16	21	21	20	18	20	18	17	17	21	18

23

X	19	18	20	18	16	19	21	16	17	19	15	22	15	17	21
Y	16	17	18	20	17	20	18	19	18	18	18	19	18	19	20

24

X	22	20	19	21	21	20	21	18	16	22	18	19	15	23	17
Y	20	21	17	19	22	19	18	19	16	19	16	20	16	21	18

25

X	17	22	19	20	19	17	21	20	20	16	22	18	20	16	16
Y	17	21	18	18	20	19	20	19	19	18	20	19	20	19	17

Тема №5. Критерий согласия Пирсона

Задание. Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности *X* с эмпирическим распределением выборки:

1.

x_i	12	14	16	18	20	22	24
n_i	6	8	12	24	22	17	11

2.

x_i	26	28	30	32	34	36	38
n_i	10	15	16	18	20	13	8

3.

x_i	25	30	35	40	45	50	55
n_i	13	14	16	25	12	11	9

4.

x_i	11	13	15	17	19	21	23
n_i	9	10	17	21	16	15	12

5.

x_i	14	17	20	23	26	29	32
n_i	10	12	15	20	18	16	9

6.

x_i	17	19	21	23	25	27	29
n_i	9	12	16	24	15	13	11

7.

x_i	10	14	18	22	26	30	34
n_i	6	8	12	18	23	19	14

8.

x_i	23	25	27	29	31	33	35
n_i	8	14	17	23	15	14	9

9.

x_i	6	10	14	18	22	26	30
n_i	12	15	17	21	14	13	8

10.

x_i	12	15	18	21	24	27	30
n_i	13	15	17	20	14	13	8

11.

x_i	11	15	19	23	27	31	35
n_i	12	14	18	20	16	11	9

12.

x_i	8	12	16	20	24	28	32
n_i	8	13	18	22	15	14	10

13.

x_i	13	16	19	22	25	28	31
n_i	11	15	20	23	12	10	9

14.

x_i	10	12	14	16	18	20	22
n_i	12	13	18	20	14	13	10

15.

x_i	14	18	22	26	30	34	38
n_i	8	12	17	19	20	13	11

16.

x_i	10	13	16	19	22	25	28
n_i	12	17	20	16	14	13	8

17.

x_i	15	20	25	30	35	40	45
n_i	7	9	11	19	23	18	13

18.

x_i	20	22	24	26	28	30	32
n_i	11	14	17	19	16	13	10

19.

x_i	17	20	23	26	29	32	35
n_i	10	13	19	22	13	12	11

20.

x_i	20	25	30	35	40	45	50
n_i	12	17	18	21	13	10	9

21.

x_i	14	17	20	23	26	29	32
n_i	7	9	13	24	20	16	11

22.

x_i	10	15	20	25	30	35	40
n_i	11	13	15	18	17	14	12

23.

x_i	18	22	26	30	34	38	42
n_i	9	13	17	19	16	15	11

24.

x_i	13	18	23	28	33	38	43
n_i	10	12	15	26	14	12	11

25.

x_i	17	19	21	23	25	27	29
n_i	8	10	13	25	18	14	12

Тема №6. Однофакторный дисперсионный анализ

Задание. В течение шести лет использовались три различных технологии по выращиванию сельскохозяйственной культуры. Данные по эксперименту (в ц/га) приведены в таблице. При уровне значимости $\alpha = 0{,}05$ методом дисперсионного анализа установить влияние различных технологий на урожайность культуры.

-1-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,31	0,32	0,36
2	0,29	0,36	0,38
3	0,30	0,29	0,31
4	0,35	0,33	0,34
5	0,28	0,32	0,35
6	0,30	0,30	0,33

-2-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,38	0,32	0,36
2	0,44	0,41	0,39
3	0,32	0,33	0,40
4	0,31	0,35	0,38
5	0,44	0,40	0,37
6	0,36	0,38	0,35

-3-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,18	0,24	0,36
2	0,28	0,36	0,12
3	0,12	0,28	0,22
4	0,14	0,40	0,45
5	0,32	0,16	0,40
6	0,25	0,30	0,28

-4-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,38	0,20	0,20
2	0,36	0,24	0,22
3	0,35	0,26	0,30
4	0,30	0,30	0,34
5	0,29	0,28	0,30
6	0,30	0,31	0,29

-5-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,10	0,18	0,14
2	0,15	0,16	0,18
3	0,14	0,25	0,30
4	0,18	0,22	0,27
5	0,20	0,32	0,33
6	0,16	0,28	0,40

-6-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,20	0,16	0,18
2	0,25	0,15	0,20
3	0,26	0,25	0,28
4	0,28	0,22	0,29
5	0,20	0,23	0,32
6	0,19	0,25	0,38

-7-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,32	0,26	0,29
2	0,31	0,30	0,26
3	0,30	0,27	0,30
4	0,29	0,28	0,31
5	0,34	0,27	0,30
6	0,30	0,30	0,28

-8-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,36	0,30	0,29
2	0,34	0,34	0,33
3	0,32	0,35	0,30
4	0,34	0,32	0,28
5	0,35	0,28	0,31
6	0,30	0,30	0,35

-9-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,36	0,34	0,30
2	0,38	0,35	0,32
3	0,34	0,32	0,36
4	0,38	0,37	0,29
5	0,34	0,29	0,31
6	0,36	0,31	0,34

-10-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,28	0,30	0,35
2	0,33	0,35	0,30
3	0,31	0,28	0,31
4	0,36	0,34	0,33
5	0,29	0,32	0,28
6	0,35	0,30	0,32

-11-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,32	0,34	0,36
2	0,41	0,36	0,44
3	0,30	0,29	0,38
4	0,32	0,31	0,36
5	0,31	0,32	0,32
6	0,32	0,30	0,33

-12-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,17	0,20	0,26
2	0,24	0,26	0,27
3	0,19	0,28	0,30
4	0,17	0,21	0,24
5	0,32	0,31	0,35
6	0,26	0,27	0,26

-13-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,36	0,26	0,24
2	0,38	0,28	0,22
3	0,30	0,29	0,31
4	0,35	0,30	0,34
5	0,32	0,27	0,29
6	0,27	0,31	0,28

-14-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,18	0,12	0,16
2	0,14	0,16	0,18
3	0,20	0,29	0,21
4	0,16	0,15	0,14
5	0,24	0,22	0,25
6	0,18	0,20	0,23

-15-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,24	0,22	0,26
2	0,31	0,27	0,32
3	0,22	0,29	0,31
4	0,26	0,25	0,24
5	0,34	0,20	0,31
6	0,31	0,24	0,30

-16-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,18	0,16	0,20
2	0,24	0,26	0,28
3	0,30	0,29	0,31
4	0,31	0,30	0,29
5	0,24	0,22	0,24
6	0,32	0,30	0,33

-17-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,26	0,28	0,26
2	0,30	0,32	0,34
3	0,31	0,28	0,31
4	0,26	0,26	0,28
5	0,31	0,29	0,31
6	0,30	0,31	0,30

-18-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,18	0,22	0,18
2	0,14	0,24	0,25
3	0,22	0,23	0,20
4	0,26	0,24	0,26
5	0,23	0,22	0,24
6	0,26	0,29	0,22

-19-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,22	0,24	0,31
2	0,27	0,26	0,29
3	0,24	0,28	0,22
4	0,30	0,31	0,35
5	0,23	0,26	0,30
6	0,24	0,30	0,24

-20-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,20	0,15	0,20
2	0,24	0,24	0,23
3	0,29	0,26	0,25
4	0,26	0,25	0,24
5	0,20	0,28	0,18
6	0,31	0,29	0,22

-21-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,20	0,27	0,24
2	0,25	0,23	0,24
3	0,24	0,21	0,22
4	0,19	0,20	0,19
5	0,20	0,28	0,20
6	0,18	0,22	0,26

-22-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,24	0,16	0,18
2	0,25	0,18	0,20
3	0,26	0,25	0,22
4	0,28	0,24	0,21
5	0,22	0,21	0,24
6	0,19	0,25	0,27

-23-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,32	0,25	0,28
2	0,31	0,31	0,26
3	0,30	0,27	0,30
4	0,28	0,26	0,22
5	0,31	0,27	0,33
6	0,31	0,29	0,29

-24-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,31	0,30	0,28
2	0,24	0,28	0,24
3	0,32	0,35	0,30
4	0,30	0,32	0,28
5	0,25	0,30	0,27
6	0,32	0,31	0,25

-25-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,22	0,25	0,24
2	0,21	0,20	0,18
3	0,24	0,22	0,26
4	0,28	0,27	0,22
5	0,32	0,26	0,30
6	0,26	0,21	0,24

-26-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,22	0,22	0,24
2	0,20	0,25	0,34
3	0,31	0,29	0,30
4	0,26	0,27	0,22
5	0,31	0,32	0,30
6	0,32	0,30	0,28

-27-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,22	0,24	0,26
2	0,21	0,20	0,24
3	0,24	0,29	0,28
4	0,22	0,28	0,26
5	0,30	0,32	0,30
6	0,28	0,26	0,28

-28-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,25	0,22	0,24
2	0,22	0,20	0,26
3	0,24	0,26	0,25
4	0,30	0,24	0,28
5	0,22	0,21	0,29
6	0,30	0,31	0,30

-29-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,23	0,25	0,25
2	0,28	0,26	0,23
3	0,30	0,29	0,31
4	0,24	0,28	0,24
5	0,31	0,26	0,28
6	0,29	0,31	0,25

-30-

Год	Технология (фактор)		
	F_1	F_2	F_3
1	0,28	0,22	0,26
2	0,24	0,26	0,28
3	0,20	0,29	0,21
4	0,26	0,25	0,24
5	0,24	0,22	0,25
6	0,25	0,20	0,23

Вопросы для самоконтроля

1. Виды совокупностей. Статистические показатели и их классификация. Статистическое распределение выборки, основные правила построения.

2. Понятие о статистическом графике, его основные элементы и правила построения. Виды статистических графиков. Эмпирическая функция распределения.

3. Полигон и гистограмма частот, построение графиков. Мода и медиана.

4. Показатели вариации и методы их расчета. Дисперсия, ее математические свойства и методы расчета.

5. Элементы корреляции. Функциональные и корреляционные связи. Виды корреляционных связей.

6. Понятие линейной корреляции. Нахождение параметров уравнения регрессии, линейный коэффициент корреляции.

7. Регрессионный анализ. Понятие криволинейной зависимости, оценка тесноты связи при криволинейной зависимости.

8. Оценка значимости коэффициентов регрессии.

9. Критерий согласия Пирсона.

10. Суть однофакторного дисперсионного анализа. Условия применения дисперсионного анализа к выборочным данным.

ПРИЛОЖЕНИЯ

Приложение 1

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
∞	1,64	1,96	2,33	2,58	3,09	3,29

Приложение 2

Значения *F*–критерия Фишера при уровне надежности 95% (α = 0,05)

k_1 / k_2	ЧИСЛО СТЕПЕНЕЙ СВОБОДЫ											
	1	2	3	4	5	6	7	8	9	10	11	14	∞
1	161	200	216	225	230	234	237	239	241	242	243	245	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,71	8,5
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,87	5,6
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,64	4,3
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	3,96	3,6
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,53	3,2
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,53	3,31	3,24	2,9
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,03	2,7
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,86	2,5
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60	2,3
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53	2,1
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,41	2,38
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	1,6
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,2
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,69	1,0

Приложение 3

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Приложение 4

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

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0264	0258	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Литература

1. Киселева Н.Г., Зиннатуллина А.Н., Еникеева С.Р. Теория вероятностей и математическая статистика: Учебно-методические пособие / Н.Г. Киселева, А.Н. Зиннатуллина, С.Р. Еникеева. – Казань: Изд-во Казанского ГАУ, 2014. – 133 с.

2. Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учеб. пособие для магистров / Н. И. Сидняев. – М. : Юрайт, 2012. – 399 с.

3. Туганбаев А.А. Теория вероятностей и математическая статистика: учеб. пособие / А.А. Туганбаев, В.Г. Крупин. – Санкт-Петербург: Изд-во «Лань», 2011. – 224 с.

4. Попов А.М. Теория вероятностей и математическая статистика: учебник для бакалавров / А.М. Попов, В.Н. Сотников; под ред. проф. А.М. Попова. – М.: Издательство Юрайт, 2011. – 440 с.

5. Чашкин, Ю.Р. Математическая статистика. Анализ и обработка данных: Учебное пособие / Ю.Р. Чашкин; Под ред. С.Н. Смоленский. - Рн/Д: Феникс, 2010. - 236 с.