*МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ*
федеральное государственное бюджетное образовательное учреждение

высшего образования

«Курганский государственный университет»

Кафедра фундаментальной математики и методики преподавания математики

# МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

**(Часть 2)**

**Методические указания для практических занятий и самостоятельной работы для студентов факультета «Математика и информационные технологи » направления 01.03.01 «Математика»**

Курган 2018

**Кафедра** фундаментальной математики и методики преподавания математики.

**Дисциплина**: «Математическая статистика» (направление 01.03.01 «Математика»).

**Составил:** ст. преподаватель Е.А. Лукерьянова.

Утверждены на заседании кафедры «13» марта 2017 г.

Рекомендованы методическим советом университета «12» декабря 2016 г.

## Содержание

Введение ........................................ 4

Тема 6. Сглаживание экспериментальной зависимости по методу наименьших квадратов ........................................ 5

Тема 7. Корреляционно-регрессионный анализ ........................................ 9

Тема 8. Проверка статистических гипотез ........................................ 27

Тема 9. Однофакторный дисперсионный анализ ........................................ 39

Примерные задания для контрольной работы ........................................ 44

Приложения ........................................ 47

Список литературы ........................................ 51

## ВВЕДЕНИЕ

Целью изучения дисциплины «Математическая статистика» является получение фундаментального образования, способствующего развитию личности.

Задачи дисциплины: изучить основы математической статистики; овладеть методами и приемами решения задач математической статистики; овладеть методами сбора, обработки и анализа статистических данных; формировать навыки проведения сплошного и выборочного наблюдения, умения выделять конкретное вероятностно-статистическое содержание в прикладных задачах учебной и профессиональной деятельности.

### Краткое содержание дисциплины

Выборочный метод. Статистическая оценка параметров распределения. Методы расчета сводных характеристик выборки. Элементы теории корреляции. Статистическая проверка статистических гипотез. Однофакторный дисперсионный анализ.

Методические указания включают в себя: планы занятий, задания для рубежного контроля, таблицы, список литературы. Планы занятий содержат вопросы для повторения, теоретическую справку, образцы решения задач, задачи различного уровня сложности для решения в аудитории, самостоятельного решения, решения дома, таблицы значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$,$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}}\, dz$, таблица значений $t_y = t(\gamma, n)$, $q_\gamma = q(\gamma, n)$, критические точки распределения $x^2$.

Цель данного пособия – оказать помощь студентам в организации самостоятельной работы по курсу «Математическая статистика», связанной с подготовкой к практическим занятиям, текущему и рубежному контролю, подготовкой к сдаче зачета по дисциплине «Математическая статистика».

## Тема 6. Сглаживание экспериментальной зависимости по методу наименьших квадратов

### Теоретическая справка

Полагая, что *x* и *y* связаны линейной зависимостью $y=ax+b$, определить коэффициенты а и b используя метод наименьших квадратов, если данные опыта представлены следующей таблицей значений x и y (таблица 1).

*Таблица 1 – Данные опыта*

| | | | | | |
|---|---|---|---|---|---|
| *x* | 1 | 2 | 3 | 4 | 5 |
| *y* | 3 | 6,1 | 9,1 | 12 | 16 |

**Решение**

Подберем параметры а и b так, чтобы сумма квадратов отклонений $y_i = f(x_i)$ была минимальной. Так как каждое отклонение зависит от а и b, то сумма квадратов отклонений есть функция F этих параметров:

$$F(a,b) = \sum_{i=1}^{n}(y_1 - ax_1 - b)\cdot(-x_1) = 0$$

или

$$F(a,b) = \sum_{i=1}^{n}(y_1 - (ax_1 + b))^2\,, \qquad F(a,b) = \sum_{i=1}^{n}(y_1 - ax_1 - b)^2.$$

Для описания минимума приравняем к нулю соответствующие частные производные:
$$\begin{cases} \dfrac{dF}{da} = 2\sum_{i=1}^{n}(y_1 - ax_1 - b)\cdot(-x_1) = 0, \\ \dfrac{dF}{db} - 2\sum_{i=1}^{n}(y_1 - ax_1 - b)\cdot(-1) - 0\,. \end{cases}$$

Составим расчетную таблицу (таблица 2).

*Таблица 2 – Расчетная таблица*

| $x_i$ | $y_i$ | $x_i y_i$ | $x_i^2$ |
|---|---|---|---|
| 1 | 3 | 3 | 1 |
| 2 | 6,1 | 12,2 | 4 |
| 3 | 9,1 | 27,3 | 9 |
| 4 | 12 | 48 | 16 |
| 5 | 16 | 80 | 25 |
| $\sum x_i = 15$ | $\sum y_i = 46{,}2$ | $\sum x_i y_i = 170{,}5$ | $\sum x_i^2 = 55$ |

Система примет вид:

$$\begin{cases}170{,}5 - 55a - 15b = 0, \\ 46{,}2 - 15a - 5b = 0.\end{cases} \quad \begin{cases}55a + 15b = 170{,}5\,, \\ 15a + 5b = 46{,}2\,.\end{cases} \quad \begin{cases}11a + 3b = 34{,}1\,, \\ 15a + 5b = 46{,}2.\end{cases}$$

Решим  систему, используя метод Крамера.

$$d = \begin{vmatrix} 11 & 3 \\ 15 & 5 \end{vmatrix} = 55 - 45 = 10\,, d \neq 0;$$

$$d_1 = \begin{vmatrix} 34{,}1 & 3 \\ 46{,}2 & 5 \end{vmatrix} = 34{,}1 \cdot 5 - 46{,}2 \cdot 3 = 170{,}5 - 138{,}6 = 31{,}9;$$

$a = \frac{d_1}{d}$ ,  $a = \frac{31{,}9}{10}$ =3,19 ;

$$d_1 = \begin{vmatrix} 11 & 34{,}1 \\ 15 & 46{,}2 \end{vmatrix} = 11 \cdot 46{,}2 - 15 \cdot 34{,}1 = 508{,}2 - 511{,}5 = -3{,}3;$$

$b = \frac{d_2}{d}$ ,  $b = \frac{-3{,}3}{10}$ $= -0{,}33\ -$;

итак, $y = 3{,}19x - 0{,}33.$

**Вопросы для повторения**

1 Задача сглаживания экспериментальной зависимости.

2 Метод наименьших квадратов.

3 Подготовиться к самостоятельной работе по темам: «Статистическое распределение выборки», «Статистические оценки параметров распределений».

**Задачи для решения в аудитории**

1 Результаты измерения величин *x* и *y* даны в таблице 3. Существует линейная зависимость $y = ax + b$, способом наименьших квадратов определить коэффициенты a и b.

*Таблица 3 – Результаты измерения величин x и y*

| | | | | | |
|---|---|---|---|---|---|
| *x* | -2 | 0 | 1 | 2 | 4 |
| y | 0,5 | 1 | 1,5 | 2 | 3 |

2 В таблице 4 приведены  данные за 5 лет о внесении минеральных удобрений и урожае сахарной свеклы с гектара посева в определенный год.

*Таблица 4 – Данные о внесении минеральных удобрений, внесенные по годам и урожайность*

| Год | Минеральные удобрения, ц | Урожайность с 1 га, т |
|---|---|---|
| 1971 | 4 | 20 |
| 1972 | 5 | 24 |
| 1973 | 6 | 29 |
| 1974 | 8 | 35 |
| 1975 | 9 | 50 |

Предполагая линейную зависимость урожайности от количества внесенных удобрений $y = ax + b$, найти по этим данным параметры $a$ и $b$.

3 Произведен ряд опытов по изучению зависимости физической величины $y$ от физической величины $x$. Результаты опытов приведены в таблице 5. Построить по методу наименьших квадратов квадратичную зависимость вида $y = av^2 + bv + c$, наилучшим образом согласующуюся с экспериментальными данными.

*Таблица 5 – Экспериментальные данные*

| | | | | | | |
|---|---|---|---|---|---|---|
| $x$ | -2 | -1 | $\frac{1}{2}$ | 2 | 2,5 | 3 |
| y | 9 | 0 | -4,5 | 0 | 3,2 | 8 |

**Задачи для самостоятельного решения**

1 В ОТК завода «Фрейзер» была измерена глубина паза 20 плашек. Результаты приведены в таблице 6.

*Таблица 6 – Результаты глубины паза 20 плашек (в мм)*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2,4 | 2,6 | 2,7 | 2,5 | 2,5 | 2,8 | 2,8 | 2,5 | 2,6 | 2,6 |
| 2,7 | 2,4 | 2,4 | 2,5 | 2,1 | 2,6 | 2,5 | 2,5 | 2,3 | 2,2 |

Требуется:

1) построить статистическое распределение частот;
2) построить полигон частот;
3) дать характеристику распределения признака, вычислив для этого:

а) $\bar{x}_B$; б) медиану; в) дисперсию, среднее квадратическое отклонение, размах варьирования, среднее абсолютное отклонение, коэффициент вариации; г) моду.

2 При сверлении отверстий одним и тем же сверлом и последующем измерении диаметров получены следующие данные (в мм): 40,25; 40,37; 40,33; 40,28; 40,29; 40,41; 40,33; 40,28; 40,29; 40,28; 40,35; 40,35; 40,41; 40,35; 40,28; 40,41; 40,29; 40,37; 40,37; 40,33.

Требуется:

4) построить статистическое распределение частот;
5) построить полигон частот;
6) дать характеристику распределения признака, вычислив для этого:
   а) $\bar{x}_B$; б) медиану; в) дисперсию, среднее квадратическое отклонение, размах варьирования, среднее абсолютное отклонение, коэффициент вариации; г) моду.

3 Колхозы и совхозы области, вошедшие в выборку по урожайности овощей, были сгруппированы определенным образом (таблица 7).

*Таблица 7 – Количество хозяйств и их урожайность*

| Урожайность, ц/га | 20 | 70 | 120 | 170 | 220 | 270 | 320 |
|---|---|---|---|---|---|---|---|
| Количество хозяйств | 11 | 23 | 26 | 17 | 20 | 2 | 1 |

**Задание**

Оценить с надежностью 0,97 математическое ожидание *a* нормально распределенного признака генеральной совокупности при помощи доверительного интервала, используя данные таблицы 7.

4 Из генеральной совокупности извлечена выборка объема $n = 12$. Оценить с надежностью 0,95 математическое ожидание *a* нормально распределенного признака генеральной совокупности при помощи доверительного интервала (таблица 8).

*Таблица 8 – Статистический ряд*

| $x_i$ | -0,5 | -0,4 | -0,2 | 0 | 0,2 | 0,6 | 0,8 | 1 | 1,2 | 1,5 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 |

**Задачи для решения дома**

1 В опыте исследована зависимость глубины проникания $h$ тела в преграду от удельной энергии $\varepsilon$ (энергии, приходящейся на квадратный сантиметр площади соударения). Экспериментальные данные приведены в таблице 9. Требуется по методу наименьших квадратов подобрать и построить прямую, изображающую зависимость $h$ от $\varepsilon$.

*Таблица 9 – Зависимость проникновения тела в преграду от удельной энергии*

| $i$ | $\varepsilon_i$(кгм/$см^2$) | $h_i$(мм) |
|---|---|---|
| 1 | 41 | 4 |
| 2 | 50 | 8 |
| 3 | 81 | 10 |
| 4 | 104 | 14 |
| 5 | 120 | 16 |
| 6 | 139 | 20 |
| 7 | 154 | 19 |
| 8 | 180 | 23 |
| 9 | 208 | 26 |
| 10 | 241 | 30 |
| 11 | 250 | 31 |
| 12 | 269 | 36 |
| 13 | 301 | 37 |

## Тема 7. Корреляционно-регрессионный анализ

**Теоретическая справка**

Во многих задачах требуется установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других величин.

Две случайные величины могут быть связаны либо функциональной зависимостью, либо зависимостью другого рода, называемой статистической, либо быть независимыми.

Строгая функциональная зависимость реализуется редко, так как обе величины или одна из них подвержены еще действию случайных факторов, причем среди них могут быть и общие для обеих величин (под общими здесь подразумеваются такие факторы, которые воздействуют и на Y, и на X).
В этом случае возникает статистическая зависимость.

Например, если Y зависит от случайных факторов $Z_1$, $Z_2$, $V_1$, $V_2$, а X зависит от случайных факторов $Z_1$, $Z_2$, $U_1$, то между Y и X имеется статистическая зависимость, так как среди случайных факторов есть общие, а именно: $Z_1$ и $Z_2$.

***Статистической*** называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой; в этом случае статистическую зависимость называют ***корреляционной.***

Приведем пример случайной величины Y, которая не связана с величиной X функционально, а связана корреляционно. Пусть Y – урожай зерна, X – количество удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т. е. Y не является функцией от X. Это объясняется влиянием случайных факторов (осадки, температура воздуха и др.). Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т.е. Y связан с X корреляционной зависимостью.

Определение 1. Две случайные величины $X$ и $Y$ находятся в *корреляционной* зависимости, если каждому значению любой из этих величин соответствует определенное распределение вероятностей другой величины.

Определение 2. *Условным математическим ожиданием* (*УМО)* дискретной случайной величины $X$ при $Y = y$ *(y* – определенное возможное значение *Y)* называется сумма произведений возможных значений величины $X$ на их условные вероятности.

$$M_y(X) = \sum_{i=1}^{n} x_i P_y(X = x_i),$$

где $P_y(X = x_i)$ – условная вероятность равенства $X = x_i$ при условии, что $Y = y$.

Для непрерывных величин $M_y(X) = \int_{-\infty}^{+\infty} x\varphi_y(x)dx,$

где $\varphi_y(x)$ – плотность вероятности случайной непрерывной величины $X$ при условии $Y = y$.

*УМО* $M_y(X)$ есть функция от $y$: $M_y(X) = f(y)$, которую называют функцией *регрессии* величины $X$ на величину $Y$.

Аналогично определяется *УМО* случайной величины $Y$ и функция регрессии $Y$ на $X$:

$$M_x(Y) = g(x).$$

Уравнение $x = f(y)(y = g(x))$ называется *уравнением регрессии* $X$ на $Y$ ($Y$ на $X$), а линия на плоскости, соответствующая этому уравнению, называется *линией регрессии.*

Линия регрессии $Y$ на $X$ ($X$ на $Y$) показывает, как в среднем зависит $Y$ от $X$ ($X$ от $Y$).

Для характеристики корреляционной зависимости между случайными величинами вводится понятие коэффициента корреляции.

Определение 3. Если $X$ и $Y$ – независимые случайные величины, то

$$M(XY) = M(X) \cdot M(Y).$$

Если же $X$ и $Y$ не являются независимыми случайными величинами, то, вообще говоря, $M(XY) \neq M(X) \cdot M(Y)$.

Условимся за меру связи (зависимости) двух случайных величин $X$ и $Y$ принять безразмерную величину $r$, определяемую соотношением:

$$r = \frac{M(XY) - M(X)M(Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Или более кратно соотношением:

$$r = \frac{\mu}{\sigma_1 \sigma_2},$$

где $\mu = M(XY) - M(X) \cdot M(Y), \sigma_1 = \sigma(X), \sigma_2 = \sigma(Y),$

и называемую *коэффициентом корреляции.*

Легко видеть, что

$$\mu = M(XY) - M(X) \cdot M(Y) = M[(X - M(X))(Y - M(Y))].$$

Определение 4. Случайные величины $X$ и $Y$ называются *некоррелированными,* если $r = 0$, и *коррелированными,* если $r \neq 0$.

Отметим некоторые свойства коэффициента корреляции.

1 Если $X$ и $Y$ – независимые случайные величины, то коэффициент корреляции равен нулю. Заметим, что обратное утверждение, вообще говоря, неверно.

2 Укажем без доказательства, что $|r| < 1$. При этом если $|r| = 1$, то между случайными величинами $X$ и $Y$ имеет место функциональная, а именно линейная зависимость.

3 Коэффициент корреляции характеризует относительную величину отклонения математического ожидания произведения $M(XY)$ от произведения математических ожиданий $M(X)M(Y)$ величин $X$ и $Y$. Так как это отклонение имеет место только для зависимых величин, то можно сказать, что коэффициент корреляции характеризует тесноту зависимости между $X$ и $Y$ (таблица 10).

*Таблица 10 – Количественный критерий оценки тесноты зависимости*

| Величина коэффициента корреляции | Характер связи |
|---|---|
| До $\lvert\pm 0{,}3\rvert$ | Практически отсутствует |
| $\lvert\pm 0{,}3\rvert - \lvert\pm 0{,}5\rvert$ | Слабая |
| $\lvert\pm 0{,}5\rvert - \lvert\pm 0{,}7\rvert$ | Умеренная |
| $\lvert\pm 0{,}7\rvert - \lvert\pm 1\rvert$ | Сильная |

Определение 5. Корреляционная зависимость между случайными величинами $X$ и $Y$ называется *линейной корреляцией,* если обе функции регрессии *f(y)* и *g(x)* являются линейными. В этом случае обе линии регрессии являются прямыми; они называются *прямыми регрессии.*

Запишем уравнение прямой регрессии $Y$ на $X$, т. е. найдем коэффициенты линейной функции $g(x) - Ax + B$.

Обозначим $M(X) = a$, $M(Y) = b$, $M[(X-a)^2] = \sigma_1^2$, $M[(Y-b)^2] - \sigma_2^2$, $A = \frac{\mu}{\sigma_1^2}$.

Полученный коэффициент называется *коэффициентом регрессии* $Y$ на $X$ и обозначается через $p(Y/X)$:

$$p(Y/X) = \frac{\mu}{\sigma_1^2}.$$

Таким образом, уравнение прямой регрессии $Y$ на $X$ имеет вид:

$$y = p(Y/X)(x - a) + b.$$

Аналогично можно получить уравнение прямой регрессии $X$ на $Y$:

$$y = p(X/Y)(y - b) + a,$$

где коэффициент регрессии $X$ на $Y$.

$$p(X/Y) = \frac{\mu}{\sigma_2^2}$$

Уравнения прямых регрессии можно записать в более симметричном виде, если воспользоваться коэффициентом корреляции. С учетом этого коэффициента

$$p\left(\frac{Y}{X}\right) = r\frac{\sigma_2}{\sigma_1},\ p\left(\frac{X}{Y}\right) = r\frac{\sigma_1}{\sigma_2},$$

и поэтому уравнения прямых регрессии принимают вид

$$y - b = r\frac{\sigma_2}{\sigma_1}(x - a),\ x - a = r\frac{\sigma_1}{\sigma_2}(y - b).$$

Из уравнений прямых регрессии видно, что обе эти прямые проходят через точку $(a;b)$; угловые коэффициенты прямых регрессии равны соответственно: (обозначения углов смотрите рисунок 1)

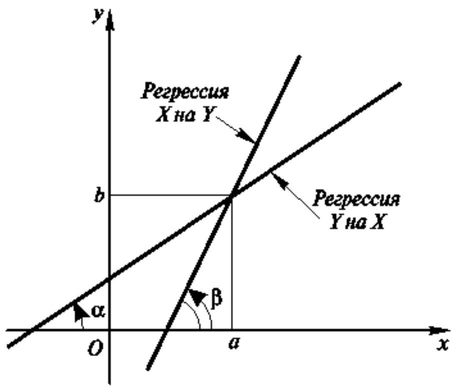


*Рисунок 1 – Прямые регрессии*

$$\mathrm{tg}\,\alpha = r\frac{\sigma_2}{\sigma_1},\ \mathrm{tg}\,\beta = \frac{1}{r}\cdot\frac{\sigma_2}{\sigma_1}.$$

Так как $|r| \le 1$, то $|\mathrm{tg}\,\alpha| \le |\mathrm{tg}\,\beta|$. Это означает, что прямая регрессии $Y$ на $X$ имеет меньший наклон к оси абсцисс, чем прямая регрессии $X$ на $Y$. Чем ближе $|r|$ к $1$, тем меньше угол между прямыми регрессии. Эти прямые сливаются тогда и только тогда, когда $|r| = 1$.

При $r = 0$ прямые регрессии имеют уравнения $y = b;\ x = a$.

В этом случае $M_x(Y) = b = M(Y);\ M_y(X) = a = M(X)$.

Из формул видно, что коэффициенты регрессии имеют тот же знак, что и коэффициент корреляции $r$, и связаны соотношением:

$$p(Y/X)p(X/Y) = r^2.$$

Пусть проведено $n$ опытов, в результате которых получены следующие значения системы величин $(X;Y):(x_i, y_i),\ i = 1,\ 2,\ ...,\ n$. За приближенные значения $M(X)$, $M(Y)$, $D(X)$ и $D(Y)$ принимают их выборочные значения:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y}_в = \frac{1}{n}\sum_{i=1}^{n} y_i,\ s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2,$$

$$s_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_в)^2.$$

Оценкой для $\mu$ служит величина

$$\mu_в = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)(y_i - \bar{y}_в).$$

Заменяя в соотношениях величины $\mu$, $\sigma_1$, $\sigma_2$ их выборочными значениями $\mu_в$, $s_1$, $s_2$, получим приближенные значения коэффициента корреляции и коэффициентов регрессии:

$$r \approx \frac{\mu_в}{s_1 s_2},\ p(Y/X) \approx \frac{\mu_в}{s_1^2},\ p(X/Y) \approx \frac{\mu_в}{s_2^2}$$

($\frac{\mu_в}{s_1 s_2}$ и $\frac{\mu_в}{s_1^2}, \frac{\mu_в}{s_2^2}$ – выборочные коэффициенты соответственно корреляции и регрессии).

Подставляя в уравнения вместо $a$, $b$, $p(Y/X)$ и $p(X/Y)$ их приближенные значения, получим выборочные уравнения прямых регрессии: $y - \bar{y}_в = \frac{\mu_в}{s_1^2}(x - \bar{x}_в),\ x - \bar{x}_в = \frac{\mu_в}{s_2^2}(y - \bar{y}_в)$.

При большом числе наблюдений одно и то же значение $x$ может встретиться $n_x$ раз, одно и то же значение $y - n_y$ раз, одна и та же пара чисел

$(x, y)$ может наблюдаться $n_{xy}$ раз. Поэтому данные наблюдений группируют, подсчитывают частоты $n_x$, $n_y$, $n_{xy}$. Все сгруппированные данные записывают в виде таблицы, которую называют *корреляционной* (таблица 11).

*Таблица 11 – Корреляционная таблица*

| Y | X | | | | |
|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | $n_y$ |
| 0.4 | **5** | **-** | **7** | **14** | 26 |
| 0.6 | - | 2 | 6 | 4 | 12 |
| 0.8 | 3 | 19 | - | - | 22 |
| $n_x$ | 8 | 21 | 13 | 18 | n= 60 |

В первой строке таблицы указаны наблюдаемые значения (10; 20; 30; 40) признака *X*, а в первом столбце – наблюдаемые значения (0,4; 0,6; 0,8) признака *Y*. На пересечении строк и столбцов находятся частоты $n_{xy}$ наблюдаемых пар значений признаков. Например, частота 5 указывает, что пара чисел (10; 0,4) наблюдалась 5 раз. Все частоты помещены в прямоугольнике, стороны которого проведены «жирными» отрезками. «Черточка» в таблице означает, что соответственная пара чисел, например (20; 0,4), не наблюдалась.

В последнем столбце записаны суммы частот строк. Например, сумма частот первой строки «жирного» прямоугольника равна $n_y = 5 + 7 + 14 = 26$; это число указывает, что значение признака *Y*, равное 0,4 (в сочетании с различными значениями признака *X*), наблюдалось 26 раз.

В последней строке записаны суммы частот столбцов. Например, число 8 указывает, что значение признака *X*, равное 10 (в сочетании с различными значениями признака *Y*), наблюдалось 8 раз.

В клетке, расположенной в нижнем правом углу таблицы, помещена сумма всех частот (общее число всех наблюдений $n$). Очевидно, $\sum n_x = \sum n_y = n$. В нашем примере

$$\sum n_x = 8 + 21 + 13 + 18 = 60 \text{ и } \sum n_y = 26 + 12 + 22 = 60.$$

Выборочный коэффициент корреляции определяется равенством:

$$r_в = \frac{\sum n_{xy} xy - n\bar{x}\bar{y}}{n\bar{\sigma}_x\bar{\sigma}_y},$$

где $x$, $y$ – варианты (наблюдавшиеся значения) признаков $X$ и $Y$; $n_{xy}$ – частота пары вариант $(x, y)$; $n$ – объем выборки (сумма всех частот);
$\bar{\sigma}_x$, $\bar{\sigma}_y$ – выборочные средние квадратические отклонения; $\bar{x}$, $\bar{y}$ – выборочные средние.

Выборочный коэффициент корреляции $r_в$ является оценкой коэффициента корреляции $r$ генеральной совокупности и поэтому также служит для измерения линейной связи между величинами – количественными признаками $Y$ и $X$. Допустим, что выборочный коэффициент корреляции, найденный по выборке, оказался отличным от нуля. Так как выборка отобрана случайно, то отсюда еще нельзя заключить, что коэффициент корреляции генеральной совокупности также отличен от нуля. Возникает необходимость проверить гипотезу о значимости (существенности) выборочного коэффициента корреляции (или, что то же, о равенстве нулю коэффициента корреляции генеральной совокупности). Если гипотеза о равенстве нулю генерального коэффициента корреляции будет отвергнута, то выборочный коэффициент корреляции значим, а величины $X$ и $Y$ коррелированы; если гипотеза принята, то выборочный коэффициент корреляции незначим, а величины $X$ и $Y$ не коррелированы.

Если выборка имеет достаточно большой объем и хорошо представляет генеральную совокупность (репрезентативна), то заключение о тесноте линейной зависимости между признаками, полученное по данным *выборки,* в известной степени может быть распространено и на *генеральную совокупность.* Например, для оценки коэффициента корреляции $r_г$ нормально распределенной генеральной совокупности (при $n \geq 50$) можно воспользоваться формулой:

$$r_в - 3\frac{1 - r_в^2}{\sqrt{n}} \leq r_г \leq r_в + 3\frac{1 + r_в^2}{\sqrt{n}}.$$

З а м е ч а н и е 1. Знак выборочного коэффициента корреляции совпадает со знаком выборочных коэффициентов регрессии, что следует из формул:

$$\rho_{yx} = r_в\frac{\bar{\sigma}_y}{\bar{\sigma}_x};\ \rho_{yx} = r_в\frac{\bar{\sigma}_x}{\bar{\sigma}_y}.$$

Требуется по данным корреляционной таблицы вычислить выборочный коэффициент корреляции. Можно значительно упростить расчет, если перейти к условным вариантам (при этом величина $r_в$ не изменится):

$$u_i = (x_i - C_1)/h_1 \text{ и } v_j = (y_j - C_2)/h_2.$$

В этом случае выборочный коэффициент корреляции вычисляют по формуле:

$$r_в = \left(\sum n_{uv} uv - n\bar{u}\,\bar{v}\right)/(n\bar{\sigma}_u \bar{\sigma}_v).$$

Величины $\bar{u}$, $\bar{v}$, $\bar{\sigma}_u$ и $\bar{\sigma}_v$ можно найти методом произведений, а при малом числе данных – непосредственно исходя из определений этих величин. Остается указать способ вычисления $\sum n_{uv} uv$, где $n_{uv}$ – частота пары условных вариант $(u, v)$.

Можно доказать, что справедливы формулы (смотреть пояснение в конце параграфа):

$$\sum n_{uv} uv = \sum vU, \text{ где } U = \sum n_{uv} u,$$

$$\sum n_{uv} uv = \sum uV, \text{ где } V = \sum n_{uv} v.$$

Для контроля целесообразно выполнить расчеты по обеим формулам и сравнить результаты; их совпадение свидетельствует о правильности вычислений.

Покажем на примере, как пользоваться приведенными формулами.

**Пример 1.** Вычислить $\sum n_{uv} uv$ по данным корреляционной таблицы 12.

*Таблица 12 – Корреляционная таблица*

| Y | X | | | | | | $n_x$ |
|---|---|---|---|---|---|---|---|
| | 10 | 20 | 30 | 40 | 50 | 60 | |
| 15 | 5 | 7 | - | -- | -- | -- | 12 |
| 25 | - | 20 | 23 | -- | -- | - | 43 |
| 35 | - | - | 30 | 47 | 2 | - | 79 |
| 45 | | | 10 | 11 | 20 | 6 | 47 |
| 55 | | | | 9 | 7 | 3 | 19 |
| $n_y$ | 5 | 27 | 63 | 67 | 29 | 9 | n = 200 |

Решение. Перейдем к условным вариантам:

$u_i = \frac{x_i - C_1}{h_1} = (x_i - 40)/10$ (в качестве ложного нуля $C_1$ взята варианта $x = 40$, расположенная примерно в середине вариационного ряда; шаг $h_1$ равен разности между двумя соседними вариантами: $20 - 10 = 10$) и $v_j = (y_j - C_2)/h_2 = (y_j - 35)/10$ (в качестве ложного нуля $C_2$ взята варианта

$y = 35$, расположенная в середине вариационного ряда; шаг $h_2$ равен разности между двумя соседними вариантами: $25 - 15 = 10$).

Составим корреляционную таблицу в условных вариантах. Практически это делают так: в первом столбце вместо ложного нуля $C_2$ (варианты 35) пишут 0; над нулем последовательно записывают $-1$, $-2$; под нулем пишут 1, 2. В первой строке вместо ложного нуля $C_1$ (варианты 40) пишут 0; слева от нуля последовательно записывают $-1$, $-2$, $-3$; справа от нуля пишут 1, 2. Все остальные данные переписывают из первоначальной корреляционной таблицы 12.

В итоге получим корреляционную таблицу 13 в условных вариантах.

*Таблица 13 – Таблица условных вариантов*

| v | u | | | | | | $n_v$ |
|---|---|---|---|---|---|---|---|
| | -3 | -2 | -1 | 0 | 1 | 2 | |
| -2 | 5 | 7 | - | -- | -- | -- | 12 |
| -1 | - | 20 | 23 | -- | -- | - | 43 |
| 0 | - | - | 30 | 47 | 2 | - | 79 |
| 1 | | | 10 | 11 | 20 | 6 | 47 |
| 2 | | | | 9 | 7 | 3 | 19 |
| $n_u$ | 5 | 27 | 63 | 67 | 29 | 9 | n = 200 |

Теперь для вычисления искомой суммы $\sum n_{uv}uv$ составим расчетную таблицу (таблица 14).

1 В каждой клетке, в которой частота $n_{uv} \neq 0$, записывают в правом верхнем углу произведение частоты $n_{uv}$ на варианту $u$. Например, в правых верхних углах клеток первой строки записаны произведения: $5 \cdot (-3) = -15$; $7 \cdot (-2) = -14$.

2 Складывают все числа, помещенные в правых верхних углах клеток одной строки и их сумму записывают в клетку этой же строки столбца $U$. Например, для первой строки $U = -15 + (-14) = -29$.

3 Умножают варианту $v$ на $U$ и полученное произведение записывают в последнюю клетку той же строки, т. е, в клетку столбца $vU$. Например, в первой строке таблицы $v = -2$, $U = -29$; следовательно, $vU = (-2) \cdot (-29) = 58$.

*Таблица 14 – Расчетная таблица*

| $v$ | $u$ | | | | | | $U = \sum_u n_{uv}u$ | $vU$ |
|---|---|---|---|---|---|---|---|---|
| | −3 | −2 | −1 | 0 | 1 | 2 | | |
| −2 | −15<br>5<br>−10 | −14<br>7<br>−14 | — | — | — | — | −29 | 58 |
| −1 | — | −40<br>20<br>−20 | −23<br>23<br>−23 | — | — | — | −63 | 63 |
| 0 | — | — | −30<br>30<br>0 | 0<br>47<br>0 | 2<br>2<br>0 | — | −28 | 0 |
| 1 | — | — | −10<br>10<br>10 | 0<br>11<br>11 | 20<br>20<br>20 | 12<br>6<br>6 | 22 | 22 |
| 2 | — | — | — | 0<br>9<br>18 | 7<br>7<br>14 | 6<br>3<br>6 | 13 | 26 |
| $V = \sum_v n_{uv}v$ | −10 | −34 | −13 | 29 | 34 | 12 | | $\sum_v vU = 169$ |
| $uV$ | 30 | 68 | 13 | 0 | 34 | 24 | $\sum_u uV = 169$ | ←Контроль↑ |

4 Наконец, сложив все числа столбца $vU$, получают сумму $\sum_v vU$, которая равна искомой сумме $\sum n_{uv}uv$. Например, для таблицы 14 имеем $\sum_v vU = 169$; следовательно, искомая сумма $\sum n_{uv}uv - 169$.

Для контроля аналогичные вычисления производят по столбцам: произведения $n_{uv}v$ записывают в левый нижний угол клетки, содержащей частоту $n_{uv} \neq 0$; все числа, помещенные в левых нижних углах клеток одного столбца, складывают и их сумму записывают в строку $V$; далее умножают каждую варианту $u$ на $V$ и результат записывают в клетках последней строки.

Наконец, сложив все числа последней строки, получают сумму $\sum_u uV$, которая также равна искомой сумме $\sum n_{uv}uv$. Например, для таблицы 14 имеем $\sum_u uV = 169$; следовательно, $\sum n_{uv}uv = 169$.

Теперь, когда мы научились вычислять $\sum n_{uv}uv$, приведем пример на отыскание выборочного коэффициента корреляции.

**Пример 2.** Вычислить выборочный коэффициент корреляции $r_в = (\sum n_{uv}uv - n\bar{u}\,\bar{v})/(n\bar{\sigma}_u\bar{\sigma}_v)$ по данным корреляционной таблицы 12.

Р е ш е н и е. Перейдя к условным вариантам, получим корреляционную таблицу 15. Величины $\bar{u}$, $\bar{v}$, $\bar{\sigma}_u$ и $\bar{\sigma}_v$ можно вычислить методом произведений;

однако, поскольку числа $u_i$, $v_i$ малы, вычислим $\bar{u}$ и $\bar{v}$, исходя из определения средней, а $\bar{\sigma}_u$ и $\bar{\sigma}_v$ – используя формулы:

$$\bar{\sigma}_u = \sqrt{\overline{u^2} - (\bar{u})^2},\ \bar{\sigma}_v = \sqrt{\overline{v^2} - (\bar{v})^2}.$$

Найдем $\bar{u}$ и $\bar{v}$:

$$\bar{u} = (\sum n_u u)/n = [5\cdot(-3) + 27\cdot(-2) + 63\cdot(-1) + 29\cdot 1 + 9\cdot 2]/200 = $$
$$= -0{,}425.$$
$$\bar{v} = (\sum n_v v)/n = [12\cdot(-2) + 43\cdot(-1) + 47\cdot 1 + 19\cdot 2]/200 = 0{,}09.$$

Вычислим вспомогательную величину $\overline{u^2}$, а затем $\sigma_u$:

$$\overline{u^2} = \frac{\sum n_u u^2}{n} = (5\cdot 9 + 27\cdot 4 + 63\cdot 1 + 29\cdot 1 + 9\cdot 2)/200 = 1{,}405.$$

$$\bar{\sigma}_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1{,}405 - (0{,}425)^2} = 1{,}106.$$

Аналогично получим $\bar{\sigma}_v = 1{,}209$.

Найдем искомый выборочный коэффициент корреляции, учитывая, что ранее уже вычислена $\sum n_{uv} uv = 169$:

$$r_в = (\sum n_{uv} uv - n\bar{u}\,\bar{v})/(n\bar{\sigma}_u\bar{\sigma}_v) = [169 - 200\cdot(-0{,}425)\cdot 0{,}09]/(200\cdot 1{,}106\cdot 1{,}209) = 0{,}603.$$

Итак, $r_в = 0{,}603$.

Задача.

Найти выборочное уравнение прямой линии регрессии *Y* на *X* по данным корреляционной таблицы 15:

*Таблица 15 – Корреляционная таблица*

| *Y* \ *X* | 20 | 25 | 30 | 35 | 40 | $n_y$ |
|---|---|---|---|---|---|---|
| 16 | 4 | 6 | - | - | - | 10 |
| 26 | - | 8 | 10 | - | - | 18 |
| 36 | - | - | 32 | 3 | 9 | 44 |
| 46 | - | - | 4 | 12 | 6 | 22 |
| 56 | - | - | - | 1 | 5 | 6 |
| $n_x$ | 4 | 14 | 46 | 16 | 20 | *n*=100 |

**Решение**

Выборочное уравнение прямой линии регрессии Y на X имеет вид:

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

где $r_B$ — выборочный коэффициент корреляции, причем

$$r_B = \frac{\sum_{i=1}^{k} n_{xy} \cdot x \cdot y - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y},$$

$\sigma_x \cdot \sigma_y$ — выборочные средние квадратические отклонения;

$\bar{x} \cdot \bar{y}$ —выборочные средние признаков X и Y;

$\overline{y_x}$ —условная средняя.

Выборочные средние признаков X и Y найдем методом произведений:

$\bar{x} = M_{1,x}^* \cdot h_1 + C_1$,

$\bar{y} = M_{1,y}^* \cdot h_2 + C_2$.

Пусть $C_1 = 30, C_2 = 36.$ Составим вспомогательные таблицы (таблицы 16,17):

*Таблица 16 — Вспомогательная таблица*

| $x_i$ | $n_i$ | $u_i$ | $n_i \cdot u_i$ | $n_i \cdot u_i^2$ | $n_i \cdot (u_i + 1)^2$ |
|---|---|---|---|---|---|
| *20* | *4* | *-2* | *-8* | *16* | *4* |
| *25* | *14* | *-1* | *-14* | *14* | *0* |
| *30* | *46* | *0* | *0* | *0* | *46* |
| *35* | *16* | *1* | *16* | *16* | *64* |
| *40* | *20* | *2* | *40* | *80* | *180* |
|  | *100* |  | *34* | *126* | *294* |

**Контроль:**

$$\sum_{i=1}^{k} h_i \cdot (u_i + 1)^2 = 294.$$

$$\sum_{i=1}^{k} n_i \cdot u_i^2 + 2\sum_{i=1}^{k} n_i \cdot u_i + n = 126 + 2 \cdot 34 + 100 = 294.$$

$M_{1,x}^* = \frac{\sum_{i=1}^{k} n_1 \cdot u_i}{n}$, $M_{1,x}^* = \frac{34}{100} = 0{,}34$, $h_1 = 5$, $\bar{x} = 0{,}34 \cdot 5 + 30 = 31{,}7$.

*Таблица 17 — Вспомогательная таблица*

| $y_i$ | $n_i$ | $v_i$ | $n_i \cdot v_i$ | $n_i \cdot v_i^2$ | $n_i \cdot (v_i + 1)^2$ |
|---|---|---|---|---|---|
| 16 | 10 | $-2$ | $-20$ | 40 | 10 |
| 26 | 18 | $-1$ | $-18$ | 18 | 0 |
| 36 | 44 | 0 | 0 | 0 | 44 |
| 46 | 22 | 1 | 22 | 22 | 88 |
| 56 | 6 | 2 | 12 | 24 | 54 |
|  | 100 |  | $-4$ | 104 | 196 |

**Контроль:**

$$\sum_{i=1}^{k} n_i \cdot (v_1 + 1)^2 = 196,$$

$$\sum_{i=1}^{k} n_i \cdot v_i^2 + 2\sum_{i=1}^{k} n_i \cdot v_i + n = 104 + 2 \cdot (-4) + 100 = 196.$$

$$M_{1,y}^{*} = \frac{\sum_{i=1}^{k} n_i \cdot v_i}{n}, M_{1,y}^{*} = \frac{-4}{100} = -0{,}04, \bar{y} = -0{,}04 \cdot 10 + 36 = -0{,}4 + 36 = = 35{,}6.$$

Вычислим $\sigma_x, \sigma_y$:

$$\sigma_x = \sqrt{D_{Bx}}, \sigma_y = \sqrt{D_{By}}.$$

Найдем $D_{Bx}$ и $D_{By}$. Воспользуемся методом произведений.

$$D_B = (M_2^{*} - (M_1^{*})^2) \cdot h^2;$$

$$M_{2,x}^{*} = \frac{\sum_{i=1}^{k} n_i \cdot u_i^2}{n}, M_{2,x}^{*} = \frac{126}{100} = 1{,}26.$$

$$D_{Bx} = (1{,}26 - (0{,}34)^2) \cdot 25 = 28{,}61.$$

$$\sigma_x = \sqrt{28{,}61} = 5{,}35.$$

Аналогично найдем $\sigma_y$, $\sigma_y = 10{,}2$.

Вычислим $\sum_{i=1}^{n} n_{x,y} \cdot x \cdot y$.

$$\sum_{i=1}^{n} n_{x,y} = 16\cdot(20\cdot4+25\cdot6)+26\cdot(8\cdot25+10\cdot30)+36\cdot$$
$$\cdot(32\cdot30+3\cdot35+9\cdot40)+46\cdot(4\cdot30+12\cdot35+6\cdot40)+56\cdot$$
$$\cdot(35+5\cdot40)=$$
$$=3680+13000+36\cdot(960+105+300)+35880+13160=$$
$$=117020.$$

Найдем $r_B = \frac{\sum_{i=1}^{n} n_{x,y}\cdot x\cdot y - n\cdot\bar{x}\cdot\bar{y}}{n\cdot\sigma_x\cdot\sigma_y}$,

$$r_B = \frac{117020-100\cdot31{,}7\cdot35{,}6}{100\cdot5{,}35\cdot10{,}2} = \frac{4168}{5457} \approx 0{,}76;$$

$$\bar{y}_x - 35{,}6 = 0{,}76\cdot\frac{10{,}2}{5{,}35}\cdot(x-31{,}7);$$

$$\bar{y}_x = 1{,}45\cdot x - 10{,}365.$$

**Проверка гипотезы о значимости выборочного коэффициента корреляции**

Пусть двумерная генеральная совокупность *(X, Y)* распределена нормально. Из этой совокупности извлечена выборка объема $n$ и по ней найден выборочный коэффициент корреляции $r_в$, который оказался отличным от нуля. Так как выборка отобрана случайно, то еще нельзя заключить, что коэффициент корреляции генеральной совокупности также отличен от нуля. Поэтому возникает необходимость при заданном уровне значимости $\alpha$ проверить нулевую гипотезу $H_0$: $r_г=0$ о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1$: $r_г \neq 0$.

Если нулевая гипотеза отвергается, то это означает, что выборочный коэффициент корреляции значимо отличается от нуля (кратко говоря, значим), а *X* и *У* коррелированны, т. е. связаны линейной зависимостью.

Если нулевая гипотеза будет принята, то выборочный коэффициент корреляции незначим, а *X* и У не коррелированны, т. е. не связаны линейной зависимостью.

В качестве критерия проверки нулевой гипотезы примем случайную величину $T = \frac{r_в\cdot\sqrt{n-2}}{\sqrt{1-r_в^2}}$.

Величина $T$ при справедливости нулевой гипотезы имеет распределение Стьюдента с $k = n - 2$ степенями свободы.

Поскольку конкурирующая гипотеза имеет вид $r_г \neq 0$,то критическая область – двусторонняя.

Обозначим значение критерия, вычисленное по данным наблюдений, через $T_{набл}$ и сформулируем правило проверки нулевой гипотезы.

**Правило**. Для того, чтобы при заданном уровне значимости $\alpha$ проверить нулевую гипотезу $H_0$: $r_г = 0$ *о* равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_1: r_г \neq 0$, надо вычислить наблюдаемое значение критерия: $T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1-r_в^2}}$, и по таблице критических точек распределения Стьюдента, по заданному уровню значимости и числу степеней свободы $k = n - 2$ найти критическую точку $t_{кр}$ ($\alpha$; $k$) для двусторонней критической области.

Если $|T_{набл}| < t_{кр}$ — нет оснований отвергнуть нулевую гипотезу.

Если $|T_{набл}| > t_{кр}$ — нулевую гипотезу отвергают.

**Пример.** По выборке объема n = 122, извлеченной из нормальной двумерной совокупности, найден выборочный коэффициент корреляции $r_в$ = 0,4. При уровне значимости $\alpha =$ 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе $H_1: r_г \neq 0$.

Решение. Найдем наблюдаемое значение критерия:

$$T_{набл} = \frac{r_в \cdot \sqrt{n-2}}{\sqrt{1-r_в^2}} = \frac{0.4 \cdot \sqrt{122-2}}{\sqrt{1-0.4^2}} = 4.78.$$

По условию, конкурирующая гипотеза имеет вид $r_г \neq 0$, поэтому критическая область — двусторонняя.

По уровню значимости 0,05 и числу степеней свободы $k = 122 - 2 = 120$ находим по таблице для двусторонней критической области критическую точку $t_{кр} = t_{кр}$ (0,05; 120)= 1,98.

Поскольку $T_{набл} > t_{кр}$ — нулевую гипотезу отвергаем. Другими словами, выборочный коэффициент корреляции значимо отличается от нуля, т. е. $X$ и $Y$ коррелированны.

**Вопросы для повторения**

1 Функциональная, статистическая и корреляционная зависимость.

2 Условные средняя и корреляционная зависимость.

3 Линейная корреляция. Криволинейная корреляция.

4 Выборочное уравнение прямой линии регрессии *Y* на *X* и *X* и *Y*.

5 Выборочный коэффициент корреляции и его свойства.

6 Выборочное корреляционное отношение.

7 Корреляционное отношение как мера корреляционной связи.

**Задачи для решения в аудитории**

1 Найти выборочное уравнение прямой линии регрессии *Y* на *X* по данным $n = 5$ наблюдений (таблица 18).

*Таблица 18 – Данные 5 наблюдений*

| | | | | | |
|---|---|---|---|---|---|
| *x* | 1 | 1,5 | 3,5 | 4,5 | 5,5 |
| *y* | 1,25 | 1,4 | 1,5 | 1,75 | 2,25 |

2 Найти выборочное уравнение прямой линии регрессии *Y* на *X* по данным, приведенным в корреляционной таблице 19.

*Таблица 19 – Корреляционная таблица*

| y \ x | 20 | 25 | 30 | 35 | 40 | $n_y$ |
|---|---|---|---|---|---|---|
| 16 | 4 | 6 | - | - | - | 10 |
| 26 | - | 8 | 10 | - | - | 18 |
| 36 | - | - | 32 | 3 | 9 | 44 |
| 46 | - | - | 4 | 12 | 6 | 22 |
| 56 | - | - | - | 1 | 5 | 6 |
| $n_x$ | 4 | 14 | 46 | 16 | 20 | $n = 100$ |

3 Найти выборочные уравнения прямых линий регрессии *Y* на *X* и *X* на *Y* по данным, приведенным в корреляционной таблице 20.

*Таблица 20 – Корреляционная таблица*

| y \ x | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | $n_y$ |
|---|---|---|---|---|---|---|---|---|---|
| 100 | 2 | 1 | - | - | - | - | - | - | 3 |
| 120 | 3 | 4 | 3 | - | - | - | - | - | 10 |
| 140 | - | - | 5 | 10 | 8 | - | - | - | 23 |
| 160 | - | - | - | 1 | - | 6 | 1 | 1 | 9 |
| 180 | - | - | - | - | - | - | 4 | 1 | 5 |
| $n_x$ | 5 | 5 | 8 | 11 | 8 | 6 | 5 | 2 | $n = 50$ |

4 Найти выборочное уравнение регрессии $\bar{Y}_x = Ax^2 + Bx + C$ по данным, приведенным в корреляционной таблице 21.

*Таблица 21 – Корреляционная таблица*

| y \ x | 2 | 3 | 5 | $n_y$ |
|---|---|---|---|---|
| 25 | 20 | | | 20 |
| 45 | | 30 | 1 | 31 |
| 110 | | 31 | 48 | 49 |
| $n_x$ | 20 | 31 | 49 | *n = 50* |

5 Вычислить выборочный коэффициент корреляции по данным следующей таблицы (таблица 22).

*Таблица 22 – Данные 12 наблюдений*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 80 | 92 | 83 | 91 | 90 | 86 | 85 | 85 | 85 | 83 | 80 | 78 |
| $y_i$ | 76 | 84 | 75 | 85 | 84 | 81 | 76 | 77 | 75 | 79 | 78 | 78 |

**Задачи для решения дома**

1 Найти выборочное уравнение регрессии *Y* на *X* по данным *n = 10* наблюдений (таблица 23).

*Таблица 23 – Результаты наблюдений*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| $y_i$ | 8,6 | 8,9 | 8,9 | 9 | 9,1 | 9,2 | 9,2 | 9,2 | 9,3 | 9,4 |

2 В таблице приведены данные, характеризующие корреляционную зависимость среднесуточного съема стали с 1 $м^2$ площади подамартеновских печей от простоя (таблица 24).

*Таблица 24 – Корреляционная зависимость среднесуточного съема стали*

| | | | | | |
|---|---|---|---|---|---|
| Простой в процентах к календарному времени работы(x) | 35,4 | 22,8 | 29,6 | 18,6 | 14,5 |
| Среднесуточный съем стали с 1 $м^2$плозади(y) | 2,01 | 4,43 | 3,67 | 5,63 | 6,55 |

Составить по этим данным уравнение прямой регрессии Y на X. Каким должен быть среднесуточный объем стали при: а) $x = 25\%$; б) $x = 10\%$.

3 Найти уравнение параболической регрессии Y на X в форме $y = Ax^2 + Bx + C$ по данным таблицы 25.

*Таблица 25 – Данные наблюдений*

| | | | | | |
|---|---|---|---|---|---|
| X | -2 | -1 | 1 | 3 | 4 |
| У | -0,1 | -2,1 | -2 | 2,8 | 4,4 |

## Тема 8. Проверка статистических гипотез

### Теоретическая справка

*Статистической* называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

*Нулевой (основной)* называют выдвинутую гипотезу $H_0$.

*Конкурирующей (альтернативной)* называют гипотезу $H_1$, которая противоречит нулевой.

Различают гипотезы, которые содержат одно и более одного предположений.

*Простой* называют гипотезу, содержащую только одно предположение.

*Сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

В итоге проверки гипотезы могут быть допущены ошибки двух родов.

*Ошибка первого рода* состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность ошибки первого рода называют *уровнем значимости* и обозначают через $\alpha$.

*Ошибка второго рода* состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность ошибки второго рода обозначают через $\beta$.

*Статическим критерием* (или просто *критерием*) называют случайную величину $K$, которая служит для проверки гипотезы.

*Наблюдаемым* (*эмпирическим*) значением $K_{набл}$ называют то значение критерия, которое вычислено по выборкам.

*Критической областью* называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

*Областью принятия гипотезы* (*областью допустимых значений*) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

*Основной принцип проверки статистических гипотез*: если наблюдаемое значение критерия принадлежит критической области, то нулевую гипотезу отвергают; если наблюдаемое значение критерия принадлежит области принятия гипотезы, то гипотезу принимают.

*Критическими точками* (*границами*) $K_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

*Правосторонней* называют критическую область, определяемую неравенством $K > K_{кр}$, где $K_{кр}$ – положительное число.

*Левосторонней* называют критическую область, определяемую неравенство $K < K_{кр}$, где $K_{кр}$ − отрицательное число.

*Двусторонней* называют критическую область, определяемую неравенством $K < k_1, K > k_2$, где $k_1 > k_2$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенствами (в предложении, что $k_{кр} > 0$):

$K < -k_{кр}, K > k_{кр}$, или равносильным неравенством $|K| > k_{кр}$.

Для отыскания критической области задаются уровнем значимости $\alpha$ и ищут критические точки, исходя из следующих соотношений:

а) для правосторонней критической области:

$$P(K > k_{кр}) = \alpha(k_{кр} > 0);$$

б) для левосторонней критической области:

$$P(K < k_{кр}) = \alpha(k_{кр} < 0);$$

в) для двусторонней симметричной области:

$$P(K > k_{кр}) = (\alpha/2)(k_{кр} > 0), P(K < -k_{кр}) = \alpha/2.$$

*Мощностью критерия* называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Другими словами, мощность критерия есть вероятность того, что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.

**Сравнение дисперсий двух генеральных совокупностей**

По независимым выборкам, объемы которых $n_1, n_2$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии $s_x^2$ и $s_y^2$. Требуется сравнить эти дисперсии.

**Правило 1.** Для того, чтобы при заданном уровне значимости $\alpha$ проверить нулевую гипотезу $H_0: D(x) = D(Y)$ о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе $H_1$: *D(X) > D(Y),* надо вычислить наблюдаемое значение критерия (отношение большей исправленной дисперсии к меньшей):

$$F_{набл} = s_Б^2 / s_М^2 .$$

И по таблице критических точек распределения Фишера – Снедекора, по заданному уровню значимости $\alpha$ и числам степеней свободы $k_1 = n_1 - 1, k_2 = n_2 - 1$ ($k_1 -$ число степеней большей исправленной дисперсии) найти критическую точку $F_{кр}(\alpha; k_1; k_2)$. Если $F_{набл} < F_{кр} -$ нет оснований отвергнуть нулевую гипотезу. Если $F_{набл} > F_{кр} -$ нулевую гипотезу отвергают.

**Правило 2.** При конкурирующей гипотезе $H_1: D(x) \neq D(Y)$ критическую точку $F_{кр}(\alpha/2; k_1; k_2)$ ищут по уровню значимости $\alpha/2$ (вдвое меньше заданного) и числам степеней свободы $k_1$ и $k_2$($k_1$ – число степеней сводобы большей дисперсии).

Если $F_{набл} < F_{кр} -$ нет оснований отвергнуть нулевую гипотезу.

Если $F_{набл} > F_{кр} -$ нулевую гипотезу отвергают.

**Пример.** По двум независимым выборкам, объемы которых $n_1 = 11$ и $n_2 = 14$, извлеченным из нормальных генеральных совокупностей X и У, найдены исправленные выборочные дисперсии $s_X^2 = 0{,}76$ и $s_Y^2 = 0{,}38$.

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 отношение большей исправленной дисперсии к меньшей:
$F_{набл} = 0{,}76/0{,}38 = 2$.

По условию конкурирующая гипотеза имеет вид *D (X) > D(Y),* поэтому критическая область – правосторонняя.

По таблице, по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_1 - 1 = 11 - 1 = 10$ и $k_2 = n_2 - 1 = 14 - 1 = 13$ находим критическую точку

$$F_{кр} = (0{,}05; 10; 13) = 2{,}67 .$$

Так как $F_{набл} < F_{кр}$ – нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначимо.

### Проверка гипотезы о нормальном распределении генеральной совокупности

Пусть эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот:

$x_i$ $x_1$ $x_2$ ... $x_N$;

$n_i$ $n_1$ $n_2$ ... $n_N$.

Требуется, используя критерий Пирсона, проверить гипотезу о том, что генеральная совокупность $X$ распределена нормально.

**Правило 1**. Для того чтобы при заданном уровне значимости $\alpha$ проверить гипотезу о нормальном распределении генеральной совокупности, надо:

1) вычислить непосредственно (при малом числе наблюдений) или упрощенным методом (при большом числе наблюдений), например, методом произведений или сумм, выборочную среднюю $\overline{x_В}$ и выборочное среднее квадратическое отклонение $\sigma_В$.

2) вычислить теоретические частоты:

$$n_i = \frac{nh}{\sigma_В} \cdot \varphi(u_i).$$

где n – объем выборки (сумма всех частот), h – шаг (разность между двумя соседними вариантами), $u_i = \frac{x_i - \overline{x_В}}{\sigma_В}$, $\varphi(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-u^2/2}$.

3) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого:

а) составляют расчетную таблицу, по которой находят наблюдаемое значение критерия:

$$\chi^2_{набл} = \sum \frac{(n_i - n_i')^2}{n_i'};$$

б) по таблице критических точек распределения $\chi^2$, по заданному уровню значимости $\alpha$ и числу степеней свободы k = s – 3 (s – число групп выборки) находят критическую точку $\chi^2_{кр}(\alpha; k)$ правосторонней критической области.

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

$x^2_{\text{набл}} > x^2_{\text{кр}}$ – гипотезу отвергают. Другими славами, эмпирические и теоретические частоты различаются значимо.

Пусть эмпирическое распределение задано в виде последовательности интервалов $(x_i, x_{i+1})$ и соответствующих им частот $n_i$($n_i$ – сумма частот, которые попали в *i-й* интервал):

$(x_1, x_2)$ $(x_2, x_3)$ … $(x_s, x_{s+1})$.

$n_1$ $n_2$ $n_s$

Требуется, используя критерий Пирсона, проверить гипотезу о том, что генеральная совокупность *X* распределена нормально.

**Правило 2.** Для того чтобы, при уровне значимости $\alpha$, проверить гипотезу о нормальном распределении генеральной совокупности, надо:

1) вычислить, например, методом произведений, выборочную среднюю $\bar{x}$ и выборочное среднее квадратическое отклонение $\sigma_B$, причем в качестве вариант $x_i^*$ принимают среднее арифметическое концов интервала:

$x_i^* = (x_i + x_{i+1})/2$;

2) пронормировать X, т. е. перейти к случайной величине Z = (X$-\overline{x^*})/\sigma^*$, и вычислить концы интервалов: $z_i$ = $(x_i - \overline{x^*})/\sigma^*$, $z_{i+1}$ = $(x_i - \overline{x^*})/\sigma^*$, причем наименьшее значение Z, т. е. $z_1$, полагают равным –$\infty$, а наибольшее, т. е. $z_{s+1}$, полагают равным $\infty$;

3) вычислить теоретические частоты: $n_i' = n \cdot P_i$

где n – объем выборка (сумма всех частот);

$P_i = \Phi(z_{i+1}) - \Phi(z_i)$ – вероятности попадания X в интервалы $(x_i, x_{i+1})$;

Ф (Z) – функция Лапласа;

4) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого:

а) составляют расчетную таблицу, по которой находят наблюдаемое значение критерия Пирсона:

$x^2_{\text{набл}}=\sum(n_i - n_i')^2/n_i'$;

б) по таблице критических точек распределения $x^2$, по заданному уровню значимости $\alpha$ и числу степеней свободы k = s – 3 (s – число интервалов выборки) находят критическую точку правосторонней критической области $x^2_{\text{кр}}(\alpha; k)$.

**Пример 1**

Используя критерий Пирсона, при уровне значимости 0,05, проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n=200 (таблица 26).

*Таблица 26 – Эмпирическое распределение выборки объема n=200*

| $x_i$ | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 15 | 20 | 31 | 20 | 30 | 27 | 24 | 20 | 13 |

**Решение**

Используя метод произведений, найдем $\bar{x}_B$ и $\sigma_B$. Пусть C = 12, h=2. Все вычисления будем заносить в таблицу (таблица 27).

*Таблица 27 – Эмпирическое распределение выборки объема n=200*

| $x_i$ | $n_i$ | $u_i$ | $n_i \cdot u_i$ | $n_i \cdot u_i^2$ | $n_i \cdot (u_i + 1)^2$ |
|---|---|---|---|---|---|
| 4 | 15 | $-4$ | $-60$ | 240 | 135 |
| 6 | 20 | $-3$ | $-60$ | 180 | 80 |
| 8 | 31 | $-2$ | $-62$ | 124 | 31 |
| 10 | 20 | $-1$ | $-20$ | 20 | 0 |
| 12 | 30 | 0 | $0$ | 0 | 30 |
| 11 | 27 | 1 | $27$ | 27 | 108 |
| 16 | 24 | 2 | 48 | 96 | 216 |
| 18 | 20 | 3 | 60 | 180 | 320 |
| 20 | 13 | 4 | 52 | 208 | 325 |
| | 200 | | 15 | 1075 | 1245 |

**Контроль**

$$\sum_{i=1}^{9} n_i \cdot (x_i + 1)^2 = 1245;$$

$$\sum_{i=1}^{9} n_i \cdot x_i^2 + 2\sum_{i=1}^{9} n_i \cdot \mathrm{x_i} + \mathrm{n} = 1075 - 30 + 200 = 1245;$$

$$M_1^* = \frac{-15}{100} = -0{,}15,\ M_2^* = \frac{1075}{100} = 10{,}75;$$

$$\bar{x}_B = M_1^* \cdot h + c,\ \bar{x}_B = -0{,}15 \cdot 2 + 12 = -0{,}3 + 12 = 11{,}7;$$

$$D_B = (M_2^* - (M_1^*)^2) \cdot h^2;$$

$$D_B = (10{,}75 - (-0{,}15)^2) \cdot 4 = (10{,}75 - 0{,}0225) \cdot 4 = 42{,}91;$$

$$\sigma_B = 6{,}6, \text{по формуле } n_i^{\prime} = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i) = 60{,}61 \cdot \varphi(u_i).$$

Для вычисления выравнивающих частот, составим расчетную таблицу (таблица 28).

Значение функции $\varphi(u_i)$ найдем в приложении А (таблица А.1).

*Таблица 28 – Эмпирическим распределением выборки объема n=200*

| $i$ | $x_i$ | $u_i = \dfrac{x_i - \bar{x}_B}{\sigma_B}$ | $\varphi(u_i)$ | $n_i^t = 60{,}61 \cdot \varphi(u_i)$ |
|---|---|---|---|---|
| 1 | 4 | $-1{,}17$ | 0,2012 | 12,2 |
| 2 | 6 | $-0{,}87$ | 0,2756 | 16,7 |
| 3 | 8 | $-0{,}56$ | 0,3410 | 20,7 |
| 4 | 10 | $-0{,}26$ | 0,3857 | 23,4 |
| 5 | 12 | $0{,}05$ | 0,3984 | 24,1 |
| 6 | 14 | 0,35 | 0,3867 | 23,4 |
| 7 | 16 | 0,65 | 0,3230 | 19,6 |
| 8 | 18 | 0,95 | 0,2541 | 15,4 |
| 9 | 20 | 1,26 | 0,1804 | 10,9 |

Сравним эмпирические и теоретические частоты:

а) составим таблицу (таблица 29), из которой найдем наблюдаемое значение критерия $x^2_{набл} = \sum_i \frac{(n_i - n_i^t)^2}{n_i^t}$.

*Таблица 29 – Сравнение эмпирических и теоретических частот*

| $i$ | $n_i$ | $n_i^t$ | $n_i - n_i^t$ | $(n_i - n_i^t)^2$ | $\dfrac{(n_i - n_i^t)^2}{n_i^t}$ |
|---|---|---|---|---|---|
| 1 | 15 | 12,2 | 2,8 | 7,84 | 0,64 |
| 2 | 20 | 16,7 | 3,3 | 10,89 | 0,65 |
| 3 | 31 | 20,7 | 10,3 | 106,09 | 5,1 |
| 4 | 20 | 23,4 | $-3{,}4$ | 11,56 | 0,49 |
| 5 | 30 | 24,1 | 5,9 | 34,81 | 1,4 |
| 6 | 27 | 23,4 | 3,6 | 12,96 | 0,6 |
| 7 | 24 | 19,6 | 4,4 | 19,36 | 0,39 |
| 8 | 20 | 15,4 | 4,6 | 21,16 | 1,37 |
| 9 | 13 | 10,9 | 2,1 | 4,41 | 0,4 |
| | | | | | $x^2_{набл} = 11{,}04$ |

б) по таблице критических точек распределения $x^2$ (приложение А, таблица А.5), по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы

$k = s - 3 = 9 - 3 = 6$ находим критическую точку правосторонней критической области $x^2_{кр}(0{,}05/6) = 12{,}6.$ Так $x^2_{набл} < x^2_{кр}$ – гипотезу о нормальном распределении генеральной совокупности нет оснований отвергать.

Найдем по таблице критических точек распределения $x^2$ (приложение А, таблица А.5) по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы k = 7 критическую точку правосторонней критической области: $x^2_{кр.}$(0,05;7)=14,1.

Так как $x^2_{набл.}$>$x^2_{кр.}$, гипотезу о равномерном распределении X отвергаем.

**Вопросы для повторения**

1 Понятие статистической гипотезы.

2 Нулевая и конкурирующая гипотеза.

3 Ошибки первого и второго ряда.

4 Статистический критерий.

5 Критическая область. Область принятия гипотезы.

6 Основной принцип проверки статистических гипотез.

7 Критические точки.

8 Правосторонняя и левосторонняя критическая область. Односторонние и двусторонние критические области.

9 Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона.

10 Сравнение двух дисперсий нормальных генеральных совокупностей.

**Задачи для решения в аудитории**

1 По двум независимым выборкам, объемы которых $n_1$ = 9 и $n_2$= 16, извлеченным из нормальных генеральных совокупностей *X* и *Y,* найдены исправленные выборочные дисперсии $s_X^2 = 34{,}02$ и $s_Y^2 = 12{,}15$. При уровне значимости 0,01, проверить нулевую гипотезу $H_0$: *D(X)* = *D(Y)* о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1$: *D(X)>D(Y).*

2 По двум независимым выборкам, объемы которых $n_1 = 14$ и $n_2 = 10$, извлеченным из нормальных генеральных совокупностей Х и У, найдены исправленные выборочные дисперсии $s_X^2 = 0{,}84$ и $s_Y^2 = 2{,}52$. При уровне значимости $\alpha$ = 0,1, проверить нулевую гипотезу $H_0{:}D(x) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1{:}D(x) \neq D(Y)$.

3 Почему при проверке с помощью критерия Пирсона гипотезы о нормальном распределении генеральной совокупности число степеней свободы находят по формуле k=s-3?

4 Используя критерий Пирсона, при уровне значимости 0,05, проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n = 200 (таблица 30).

*Таблица 30 – Выборка объема n = 200*

| *xi* | *5* | *7* | *9* | *11* | *13* | *15* | *17* | *19* | *21* |
|---|---|---|---|---|---|---|---|---|---|
| *ni* | *15* | *26* | *25* | *30* | *26* | *21* | *24* | *20* | *13* |

5 При уровне значимости 0,05 проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты (таблица 31).

*Таблица 31 – Эмпирические и теоретические частоты*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Эмпирические частоты | 6 | 12 | 16 | 40 | 13 | 8 | 5 |
| Теоретические частоты | 4 | 11 | 15 | 43 | 15 | 6 | 6 |

6 В колхозе выращивали овес сорта «Укосный – 550» на сено.

На 65 случайно выбранных гектарах получен следующий урожай (таблица 32).

*Таблица 32 – Урожайность овса сорта «Укосный – 550»*

| Урожайность ц/га | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 |
|---|---|---|---|---|---|---|---|---|
| Количество гектаров | 1 | 2 | 5 | 14 | 19 | 15 | 6 | 3 |

По критерию Пирсона проверить, что урожайность овса распределена нормально. Найти процент гектаров, для которых урожайность не менее 90 ц/га. Построить полигон относительных частот.

7 Известны следующие выборочные данные о посеве за день яровых (в процентах к плану) хозяйствами области: 7; 1; 3; 4; 3; 3; 1; 5; 4; 2; 6; 5; 4; 3; 2; 2; 4; 3; 5; 6; 4; 3; 2; 5; 4; 3; 2; 5; 4; 6; 4; 5; 2; 4; 5; 3; 3; 2; 4; 6; 5; 4; 3; 2; 3; 1; 4; 5; 4; 3. Составить дискретный вариационный ряд, проверить по критерию Пирсона, что распределение нормальное. Найти процент хозяйств, для которых отклонение $|x - \bar{x}|$ меньше 2,8. Построить полигон частот и кривую теоретических частот.

8 Результаты выборочного обследования коров фермерского стада по годовому надою даны в таблице 33.

*Таблица 33 – Результаты выборочного обследования коров фермерского стада по годовому надою*

| Годовой надой (тыс. кг) | 2,7-2,9 | 2,9-3,1 | 3,1-3,3 | 3.3-3,5 | 3,5-3,7 | 3,7-3,9 | 3,9-41 | 4,1-4,3 |
|---|---|---|---|---|---|---|---|---|
| Кол-во коров | 5 | 16 | 33 | 43 | 37 | 20 | 7 | 2 |

Выяснить, является ли распределение годового надоя нормальным? Найти процент надоев в среднем не менее 3500 кг и процент коров, у которых средний годовой надой отличается от среднего годового надоя меньше, чем на 500 кг. Построить кривую теоретических частот.

9 Почему при проверке по критерию Пирсона гипотеза о показательном распределении генеральной совокупности число степеней свободы определяется равенством $\kappa = s - 2$, где $s$ —число интервалов выборки.

10 В результате испытания $200$ элементов на длительность работы получено эмпирическое распределение, приведенное в таблице 34 (в первом столбце указаны интервалы времени в часах, во втором столбце − частоты, т.е. количество элементов, проработавших время в пределах соответствующего интервала).

*Таблица 34 – Длительность работы элементов*

| $x - x_{i+1}$ | $0-5$ | $5-10$ | $10-15$ | $15-20$ | $20-25$ | $25-30$ |
|---|---|---|---|---|---|---|
| $n_i$ | *133* | *45* | *15* | *4* | *2* | *1* |

Требуется, при уровне значимости 0,05, проверить гипотезу о том, что время работы элементов распределено по показательному закону.

11 Произведено $n = 100$ опытов. Каждый опыт состоял из $N = 10$ испытаний в каждом из которых вероятность $\mathrm{p}$ появления события *А* равна 0,3. В итоге получено следующее эмпирическое распределение: в первой строке указано число $x_i$ появлений событий А в одном опыте; во второй строке − частота $n_i$, т.е. число опытов, в которых наблюдались $x_i$ появлений события А (таблица 35).

Требуется, при уровне значимости 0,05, проверить гипотезу о том, что дискретная случайная величина $X$ (число появлений события *А*) распределена по биномиальному закону.

*Таблица 35 – Эмпирическое распределение*

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $n_i$ | 2 | 10 | 27 | 32 | 23 | 6 |

12 Почему при проверке с помощью критерия Пирсона гипотезы о равномерном распределении генеральной совокупности X число степеней свободно определяется из равенства $\kappa = s - 2$, где $s$ –число интервалов выборки.

13 Почему параметры *a* и *b* равномерно распределенной случайной величины *x* оценивают по формулам:

$$a^* = \bar{x}_\beta - \sqrt{3\sigma_\beta}, \quad b^* = \bar{x}_\beta + \sqrt{3\sigma_\beta}?$$

14 Произведено n = 200 испытаний, в результате каждого из которых событие A появляется в различные моменты времени. В итоге было получено эмпирическое распределение, приведенное в таблице 36. Требуется, при уровне значимости 0,05, проверить гипотезу о том, что время появления событий распределено равномерно.

*Таблица 36 – Эмпирическое распределение*

| Интервал $x_{i-1} - x_i$ | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 | 12-14 | 14-16 | 16-18 | 18-20 | 20-22 |
|---|---|---|---|---|---|---|---|---|---|---|
| Частота $n_i$ | 21 | 16 | 15 | 26 | 22 | 14 | 21 | 22 | 18 | 25 |

15 Отдел технического контроля проверил $n = 200$ партий одинаковых изделий и получил следующее эмпирическое распределение (таблица 37). Требуется, при уровне значимости 0,05, проверить гипотезу о том, что число нестандартных изделий $x$ распределено по закону Пуассона.

*Таблица 37 – Количество нестандартных изделий*

| Количество нестандартных изделий в одной партии $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Количество партий, содержащих $x_i$ нестандартных изделий, $n_i$ | 116 | 56 | 22 | 4 | 2 |

**Задачи для самостоятельного решения**

1 По двум независимым выборкам, объемы которых $n_1 = 9$ и $n_2 = 6$, извлеченным из нормальных генеральных совокупностей $X$ и Y, найдены выборочные дисперсии $D_B(X) = 14{,}4$ и $D_B(Y) = 20{,}5$. При уровне значимости 0,1 проверить нулевую гипотезу $H_0$:$D(X)=D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(x) \neq D(Y)$.

2 Двумя методами проведены измерения одной и той же физической величины. Получены следующие результаты:

а) в первом случае $x_1 = 9{,}6$; $x_2=10{,}0$; $x_3 = 9{,}8$; $x_4 = 10{,}2$; $x_5=10{,}6$;

б) во втором случае $y_1=10{,}4$; $y_2 =9{,}7$; $y_3 = 10{,}0$; $y_4 =10{,}3$.

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости $\alpha = 0{,}1$?

3 Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объемы которых $n_1 = 10$ и $n_2 = 8$. В результаты контролируемого размера отобранных изделий получены следующие результаты:

$x_i$ 1,08 1,10 1,12 1,14 1,15 1,25 1,36 1,38 1,40 1,42;

$y_i$ 1,11 1,12 1,18 1,22 1,33 1,35 1,36 1,38.

Можно ли считать, что станки обладают одинаковой точностью [$H_0: D(x) = D(Y)$], если принять уровень значимости $\alpha = 0{,}1$ и в качестве конкурирующей гипотезы $H_1: D(x) \neq D(Y)$?

4 Н. Хольмбергом наблюдалось распределение красных кровяных шариков по 169 отделениям прибора гемацитомера; число $v_i$ отделений, содержащих по $n_i$ красных кровяных шариков, указаны в следующей таблице (таблица 38).

*Таблица 38 – Распределение красных кровяных шариков по 169 отделениям прибора гемацитомера*

| $v_i$ | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 17 | 13 | 14 | 15 | 15 | 21 | 18 | 17 | 16 | 9 | 14 |

Найти среднее значение числа красных кровяных шариков в одном отделении и, приняв его за параметр $\lambda$ распределения Пуассона, проверить с помощью критерия $z^2$ с 5 % уровнем значимости гипотезу о том, что выборка согласуется с указанным распределением Пуассона.

5 При измерении роста 359 школьников полученые результаты, представлены в таблице 39.

*Таблица 39 – Рост школьников*

| Рост, см | 173 | 172 | 171 | 170 | 169 | 168 | 167 | 166 | 165 |
|---|---|---|---|---|---|---|---|---|---|
| Число школьников | 1 | 3 | 11 | 14 | 17 | 47 | 53 | 50 | 44 |

| Рост, см | 164 | 163 | 162 | 161 | 160 | 159 | 158 | 157 | 156 |
|---|---|---|---|---|---|---|---|---|---|
| Число школьников | 41 | 26 | 19 | 11 | 13 | 4 | 2 | 2 | 1 |

Можно ли предполагать, что распределение школьников по росту мало отличается от нормального?

6 В течении 10 часов регистрировали прибытие автомашин к бензоколонке и получили эмпирическое распределение, приведенное в таблице 40 (в первом столбце указан интервал времени в часах, во втором столбце — количество машин, прибывших в этом интервале). Всего было зарегистрировано 200 машин.

*Таблица 40 – Время прибытия автомашин*

| Время прибытия | 8-9 | 9-10 | 10-11 | 11-12 | 13-14 | 14-15 |
|---|---|---|---|---|---|---|
| $n_i$ | 12 | 40 | 22 | 16 | 28 | 6 |

| Время прибытия | 15-16 | 16-17 | 17-18 |
|---|---|---|---|
| $n_i$ | 33 | 18 | 14 |

Требуется, при уровне значимости 0,01, проверить гипотезу о том, что время прибытия машин распределено равномерно.

## Тема 9. Однофакторный дисперсионный анализ

### Теоретическая справка

На практике часто возникают следующие ситуации. Есть один фактор, принимающий конечное число значений и влияющий на конечный результат. Например, требуется оценить влияние:

- различных плавок на механические свойства металла;
- свойств сырья на показатели качества продукции;

- количества вносимых удобрений на урожайность;
- способа действия на поставленную цель и т. п.

Такие задачи решает однофакторный анализ. При этом используется следующая терминология.

**Фактор** – то, что должно оказывать влияние (способ действия).

**Уровень фактора** – конкретная реализация (значение) фактора.

**Отклик** – значение измеряемого признака (величина результата).

По измеряемому признаку идет сравнение уровней фактора.

Данные (статистический материал) для однофакторного анализа получают следующим образом.

Пусть имеется $m$ уровней фактора. Каждый уровень наблюдают несколько раз (не обязательно одинаковое число). Другими словами, каждый способ действия применяют несколько раз. Пусть для определенности проводится $n$ наблюдений. Результатом таких испытаний являются $n$ выборок значений измеряемого признака (откликов). Данные можно представить в виде таблицы – матрицы:

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} = (x_{ij}), \qquad (i = 1,2,\dots,m;\ j = 1,2,\dots,n),$$

где $x_{ij}$ – отклик на $i$-м уровне фактора при $j$-м наблюдении.

В рамках этого анализа проверяется существенность влияния уровней фактора на конечный результат.

### Однофакторный дисперсионный анализ

Обычно предполагают, что все наблюдения $x_{ij}$ принадлежат некоторому сдвиговому семейству распределений. Если рассматривается семейство нормальных распределений, то для исследования применяется ***однофакторный дисперсионный анализ***.

В этом случае однофакторная дисперсионная модель выглядит так:

$$x_{ij} = \mu + F_i + \varepsilon_{ij},$$

где $\mu$ – средний вклад фактора;

$F_i$ – эффект, обусловленный влиянием $i$-го уровня фактора;

$\varepsilon_{ij}$ – случайная компонента, или возмущение, вызываемая неконтролируемыми факторами.

Условия применения дисперсионного анализа таковы:

1) математическое ожидание возмущения $\varepsilon_{ij}$ равно нулю для любых $i$, т.е. $M[\varepsilon_{ij}]=0$;

2) возмущения $\varepsilon_{ij}$ взаимно независимы;

3) дисперсия переменной $x_{ij}$ (или возмущение $\varepsilon_{ij}$) постоянна для любых *i,j*, т.е. $D[\varepsilon_{ij}]=\sigma^2$;

4) переменная $x_{ij}$ (или возмущение $\varepsilon_{ij}$) имеет нормальный закон распределения $N(0,\sigma^2)$.

Существенность влияния уровней фактора отражает гипотеза о равенстве средних нескольких (а точнее, *m*) совокупностей. Если полагать, что $X_i, i=\overline{1,m}$ – нормальные случайные величины, соответствующие измеряемому признаку на *i*-м уровне фактора, с математическими ожиданиями $a_i, i=\overline{1,m}$ и одинаковыми дисперсиями $\sigma^2$, то задача дисперсионного анализа сводится к проверке нулевой гипотезы $H_0: a_1=a_2=\cdots=a_m$ (таблица 41).

*Таблица 41 – Схема дисперсионного анализа*

| Компоненты дисперсии | Сумма квадратов (смещенная оценка дисперсий) | Число степеней свободы | Средний квадрат (несмещенная оценка дисперсий) |
|---|---|---|---|
| Межгрупповая (факторная) | $Q_1=n\sum_{i=1}^{m}(\bar{x}_{i*}-\bar{x}_{**})^2$ | m-1 | $s_1^2=\dfrac{Q_1}{m-1}$ |
| Внутригрупповая (остаточная) | $Q_2=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i*})^2$ | mn-m | $s_2^2=\dfrac{Q_2}{mn-m}$ |
| Общая (полная) | $Q_2=\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{**})^2$ | mn-1 | |

Здесь $\bar{x}_{i*}=\dfrac{1}{n}\sum_{j=1}^{n}x_{ij}$ – групповая средняя для *i*-го фактора (* – усреднение по соответствующему индексу); $\bar{x}_{**}=\dfrac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}x_{ij}=\dfrac{1}{m}\sum_{i=1}^{m}\bar{x}_{i*}$ – общая средняя.

Для проверки нулевой гипотезы используется *F*-статистика (критерий Фишера – Снедекора), имеющая вид $F=\dfrac{s_1^2}{s_2^2}$. При уровне значимости $\alpha$

нулевая гипотеза отвергается, если $F > F_{\alpha;m-1;mn-m}$, где $F_{\alpha;m-1;mn-m}$ берется из таблицы критических значений критерия.

**Ранговый однофакторный анализ**

Если ничего неизвестно про распределение откликов, то использовать количественные значения $x_{ij}$ при проверке гипотезы трудно. Поэтому для наблюдений определяется не зависящее от распределения отношение «больше-меньше», т.е. каждому $x_{ij}$ во всей совокупности ставится в соответствие **ранги** $r_{ij}$.

В этом случае проверяется гипотеза об однородности выборок (при нулевой гипотезе расположение рангов по местам равновероятно). Для этого используется ранговый критерий Краскела-Уоллеса (для непрерывных распределений наблюдений). Он имеет вид:

$$H = \frac{12n}{N(N+1)}\sum_{i=1}^{m}\left(R_{i*} - \frac{N+1}{2}\right)^2,$$

где $N$ – общее число наблюдений; $R_{i*} = \frac{1}{n}\sum_{j=1}^{n} r_{ij}$ – средний ранг на *i*-м уровне фактора.

При большом числе наблюдений критерий имеет другой вид:

$$H = \frac{12n}{N(N+1)}\sum_{i=1}^{m} R_{i*}{}^2 - 3(N+1).$$

В этом случае критерий Краскела-Уоллеса имеет распределение $\chi^2(m-1)$. При уровне значимости $\alpha$ гипотеза отвергается, если $H > \chi^2_{1-\alpha}(m-1)$, где $\chi^2_{1-\alpha}$ – квантиль уровня $1-\alpha$.

**Вопросы для повторения**

1 Понятие о дисперсионном анализе.

2 Общая, факторная и остаточная суммы квадратов отклонений.

3 Общая, факторная и остаточная дисперсии.

4 Сравнение нескольких средних методом дисперсионного анализа.

**Задачи для решения в аудитории**

1 При уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборка извлечена из нормальной совокупности с одинаковыми дисперсиями.

*Таблица 42 – Факторный анализ признака*

| Номер испытания | Уровни фактора | | | | |
|---|---|---|---|---|---|
| | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ |
| 1 | 42 | 66 | 35 | 64 | 70 |
| 2 | 55 | 91 | 50 | 70 | 79 |
| 3 | 67 | 96 | 60 | 79 | 88 |
| 4 | 67 | 98 | 69 | 81 | 90 |
| Xj | 57,75 | 87,75 | 53,50 | 73,50 | 81,75 |

2 Студентов 1 курса опрашивали с целью выявления занятий, которым они посвящают свое свободное время. Проверьте, различаются ли распределение вербальных и невербальных предпочтений студентов (таблица 43).

*Талица 43 – Распределение вербальных и невербальных предпочтений студентов*

| N | $П_1$ | $П_2$ |
|---|---|---|
| 1 | 12 | 17 |
| 2 | 18 | 19 |
| 3 | 23 | 25 |
| 4 | 10 | 7 |
| 5 | 15 | 17 |
| $x_{ср}$ | 15.6 | 17 |

**Задачи для самостоятельного решения**

1 При уровне значимости 0.05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборка извлечина из нормальной совокупности с одинаковыми дисперсиями (таблица 44).

*Таблица 44 – Факторный анализ признака*

| Номер испытания | Уровни фактора | | | |
|---|---|---|---|---|
| | $F_1$ | $F_2$ | $F_3$ | $F_4$ |
| 1 | 6 | 6 | 9 | 7 |
| 2 | 7 | 7 | 12 | 9 |
| 3 | 8 | 11 | 13 | 10 |
| 4 | 11 | 12 | 14 | 10 |
| Xj | 8 | 9 | 12 | 9 |

2 Провести однофакторный дисперсионный анализ (неслучайное или зависимое формирование групп). Провести попарное сравнение средних по результатам дисперсионного анализа (таблица 45).

*Таблица 45 – Исходные данные для дисперсионного анализа*

| Гербицид (способ внесения) | Повторности | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Харпес (под культивацию) | 31,2 | 28,6 | 32,1 | 32,1 | 34,0 |
| Харпес (Ленточный) | 28,1 | 31,0 | 32,3 | 28,7 | 29,6 |
| Харпес (под борование) | 28,6 | 32,4 | 26,9 | 29,3 | 32,1 |

**Задачи для решения дома**

1 Предположим, что в педагогическом эксперименте участвовали три группы студентов по 10 человек в каждой. В группах применили различные методы обучения: в первой – традиционный ($F_1$), во второй – основанный на компьютерных технологиях ($F_2$), в третьей – метод, широко использующий задания для самостоятельной работы ($F_3$). Знания оценивались по десятибалльной системе.

Требуется обработать полученные данные об экзаменах и сделать заключение о том, значимо ли влияние метода преподавания, приняв за уровень значимости $\alpha$=0,05.

Результаты экзаменов заданы таблицей 46, $F_j$ – уровень фактора, $x_{ij}$ – оценка i-го учащегося, обучающегося по методике $F_j$.

*Таблица 46 – Результаты экзаменов*

| | i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Уровень фактора $F_j$ | $F_1$ | 7 | 5 | 6 | 4 | 6 | 7 | 8 | 6 | 5 | 7 |
| | $F_2$ | 9 | 8 | 10 | 8 | 7 | 10 | 10 | 9 | 7 | 6 |
| | $F_3$ | 6 | 7 | 6 | 6 | 9 | 5 | 7 | 8 | 7 | 8 |

**Примерные задания для контрольной работы**

1 Известно распределение рабочих механического цеха по тарифным разрядам (таблица 47).

*Таблица 47 – Распределение рабочих механического цеха по тарифным разрядам*

| | | | | | | |
|---|---|---|---|---|---|---|
| Тарифные разряды | 1 | 2 | 3 | 4 | 5 | 6 |
| Количество рабочих | 3 | 2 | 9 | 20 | 16 | 3 |

Дать характеристику распределения признака. Построить полигон частот. Найти асимметрию и эксцесс.

2 Районы области получили картофелеуборочные комбайны в следующем количестве: 2, 3, 4, 1, 5, 4, 6, 4, 7, 4, 5, 3, 2, 4, 5, 7, 6, 2, 4, 3, 4, 6, 4, 3, 4, 4, 6, 4, 4, 6, 3, 4, 4, 5, 6, 4. Найти среднее число комбайнов, отправленных в район и доверительные интервалы для среднего значения с надежностью 0,95.

3 Произведено 10 измерений одним прибором (без систематической ошибки) некоторой физической величины, причем исправленное среднее квадратическое отклонение случайных ошибок измерений оказалось равным 0,8. Найти точность прибора с надежностью 0,95.

4 В некоторой местности в течение 300 суток регистрировалось среднесуточная температура воздуха. В итоге наблюдений было получено эмпирическое распределение. Требуется, при уровне значимости 0,05, проверить гипотезу о том, что среднесуточная температура воздуха распределена равномерно (таблица 48).

*Таблица 48 – Среднесуточная температура воздуха в течение 300 суток*

| $x_{i-1}-x_i$ | -40-(-30) | -30-(-20) | -20-(-10) | -10-0 | 0-10 | 10-20 | 20-30 | 30-40 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 25 | 40 | 30 | 45 | 40 | 46 | 48 | 26 |

5 Результаты измерения величины X и Y даны в таблице 49.

*Таблица 49 – Результаты измерения величины X и Y*

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

Предполагая, что между X и Y существует линейная зависимость $y = ax + b$, способом наименьших квадратов определить коэффициенты $a$ и $b$.

6 Дано следующее распределение (таблица 50).

*Таблица 50 – Статистическое распределение*

| $x_{i-1}-x_i$ | *3,45–3,55* | *3,55–3,65* | *3,65–3,75* | *3,75–3,85* | *3,85–3,95* |
|---|---|---|---|---|---|
| $n_i$ | *1* | *5* | *3* | *4* | *7* |

Провести анализ статистических данных. Построить гистограмму и полигон частот.

7 Из генеральной совокупности извлечена выборка объема n=10 (таблица 51).

*Таблица 51 – Выборка объема n=10*

| $x_i$ | -2 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $n_i$ | 2 | 1 | 2 | 2 | 2 | 1 |

Оценить с надежностью 0,95 математическое ожидание $a$ нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала.

8 Выборочным путем установлено, что в партии из 200 деталей 160 стандартных. Найти относительную частоту таких деталей и надежность утверждения, что полученная частота является оценкой вероятности с относительной погрешностью не более 6%.

9 Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ по данным, приведенным в корреляционной таблице 52. Оценить силу корреляционной связи по выборочному корреляционному соотношению.

*Таблица 52 – Корреляционная таблица*

| Y \ X | 2 | 3 | 5 | $n_y$ |
|---|---|---|---|---|
| 25 | 20 | - | - | 20 |
| 45 | - | 30 | 1 | 31 |
| 110 | - | 31 | 48 | 79 |
| $n_x$ | 20 | 31 | 49 | $n = 100$ |

10 В итоге проверки на нестандартность 200 ящиков консервов получено следующее эмпирическое распределение (таблица 53).

*Таблица 53 – Эмпирическое распределение*

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $n_i$ | 132 | 43 | 20 | 3 | 2 |

Требуется при уровне значимости 0,05 проверить гипотезу о том, что случайная величина X (число нестандартных коробочек) распределена по закону Пуассона.

## ПРИЛОЖЕНИЯ

### Приложение А

### Статистические таблицы

*Таблица А.1 – Значения функции* $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

| | 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 | 0270 | 0264 | 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 |

*Таблица А.2 – Значения функции* $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \cdot \int\limits_0^x e^{-\frac{z^2}{2}} dz$

| | Φ(*x*) | *x* | Φ(*x*) | *x* | Φ(*x*) | *x* | Φ(*x*) |
|---|---|---|---|---|---|---|---|
| 0,00 | 0,0000 | 0,32 | 0,1255 | 0,64 | 0,2389 | 0,96 | 0,3315 |
| 0,01 | 0,0040 | 0,33 | 0,1293 | 0,65 | 0,2422 | 0,97 | 0,3340 |
| 0,02 | 0,0080 | 0,34 | 0,1331 | 0,66 | 0,2454 | 0,98 | 0,3365 |
| 0,03 | 0,0120 | 0,35 | 0,1368 | 0,67 | 0,2486 | 0,99 | 0,3389 |
| 0,04 | 0,0160 | 0,36 | 0,1406 | 0,68 | 0,2517 | 1,00 | 0,3413 |
| 0,05 | 0,0199 | 0,37 | 0,1443 | 0,69 | 0,2549 | 1,01 | 0,3438 |
| 0,06 | 0,0239 | 0,38 | 0,1480 | 0,70 | 0,2580 | 1,02 | 0,3461 |
| 0,07 | 0,0279 | 0,39 | 0,1517 | 0,71 | 0,2611 | 1,03 | 0,3485 |
| 0,08 | 0,0319 | 0,40 | 0,1554 | 0,72 | 0,2642 | 1,04 | 0,3508 |
| 0,09 | 0,0359 | 0,41 | 0,1591 | 0,73 | 0,2673 | 1,05 | 0,3531 |
| 0,10 | 0,0398 | 0,42 | 0,1628 | 0,74 | 0,2703 | 1,06 | 0,3554 |
| 0,11 | 0,0438 | 0,43 | 0,1664 | 0,75 | 0,2734 | 1,07 | 0,3577 |
| 0,12 | 0,0478 | 0,44 | 0,1700 | 0,76 | 0,2764 | 1,08 | 0,3599 |
| 0,13 | 0,0517 | 0,45 | 0,1736 | 0,77 | 0,2794 | 1,09 | 0,3621 |
| 0,14 | 0,0557 | 0,46 | 0,1772 | 0,78 | 0,2823 | 1,10 | 0,3643 |
| 0,15 | 0,0596 | 0,47 | 0,1808 | 0,79 | 0,2852 | 1,11 | 0,3665 |
| 0,16 | 0,0636 | 0,48 | 0,1844 | 0,80 | 0,2881 | 1,12 | 0,3686 |
| 0,17 | 0,0675 | 0,49 | 0,1879 | 0,81 | 0,2910 | 1,13 | 0,3708 |
| 0,18 | 0,0714 | 0,50 | 0,1915 | 0,82 | 0,2939 | 1,14 | 0,3729 |
| 0,19 | 0,0753 | 0,51 | 0,1950 | 0,83 | 0,2967 | 1,15 | 0,3749 |
| 0,20 | 0,0793 | 0,52 | 0,1985 | 0,84 | 0,2995 | 1,16 | 0,3770 |
| 0,21 | 0,0832 | 0,53 | 0,2019 | 0,85 | 0,3023 | 1,17 | 0,3790 |
| 0,22 | 0,0871 | 0,54 | 0,2054 | 0,86 | 0,3051 | 1,18 | 0,3810 |
| 0,23 | 0,0910 | 0,55 | 0,2088 | 0,87 | 0,3078 | 1,19 | 0,3830 |
| 0,24 | 0,0948 | 0,56 | 0,2123 | 0,88 | 0,3106 | 1,20 | 0,3849 |
| 0,25 | 0,0987 | 0,57 | 0,2157 | 0,89 | 0,3133 | 1,21 | 0,3869 |
| 0,26 | 0,1026 | 0,58 | 0,2190 | 0,90 | 0,3159 | 1,22 | 0,3883 |
| 0,27 | 0,1064 | 0,59 | 0,2224 | 0,91 | 0,3186 | 1,23 | 0,3907 |
| 0,28 | 0,1103 | 0,60 | 0,2257 | 0,92 | 0,3212 | 1,24 | 0,3925 |
| 0,29 | 0,1141 | 0,61 | 0,2291 | 0,93 | 0,3238 | 1,25 | 0,3944 |
| 0,30 | 0,1179 | 0,62 | 0,2324 | 0,94 | 0,3264 | | |
| 0,31 | 0,1217 | 0,63 | 0,2357 | 0,95 | 0,3289 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1,26 | 0,3962 | 1,59 | *0,4441* | 1,92 | 0,4726 | *2,50* | 0,4938 |
| 1,27 | 0,3980 | 1,60 | *0,4452* | 1,93 | 0,4732 | *2,52* | 0,4941 |
| 1,28 | 0,3997 | 1,61 | *0,4463* | 1,94 | 0,4738 | *2,54* | 0,4945 |
| 1,29 | 0,4015 | 1,62 | *0,4474* | 1,95 | 0,4744 | *2,56* | 0,4948 |
| 1,30 | 0,4032 | 1,63 | *0,4484* | 1,96 | 0,4750 | *2,58* | 0,4951 |
| 1,31 | 0,4049 | 1,64 | *0,4495* | 1,97 | 0,4756 | *2,60* | 0,4953 |
| 1,32 | 0,4066 | 1,65 | *0,4505* | 1,98 | 0,4761 | *2,62* | 0,4956 |
| 1,33 | 0,4082 | 1,66 | *0,4515* | 1,99 | 0,4767 | *2,64* | 0,4959 |
| 1,34 | 0,4099 | 1,67 | *0,4525* | 2,00 | 0,4772 | *2,66* | 0,4961 |
| 1,35 | 0,4115 | 1,68 | *0,4535* | 2,02 | 0,4783 | *2,68* | 0,4963 |
| 1,36 | 0,4131 | 1,69 | *0,4545* | 2,04 | 0,47,93 | *2,70* | 0,4965 |
| 1,37 | 0,4147 | 1,70 | *0,4554* | 2,06 | 0,4803 | *2,72* | 0,4967 |
| 1,38 | 0,4162 | 1,71 | *0,4564* | 2,08 | 0,4812 | *2,74* | 0,4969 |
| 1,39 | 0,4177 | 1,72 | *0,4573* | 2,10 | 0,4821 | *2,76* | 0,4971 |
| 1,40 | 0,4192 | 1,73 | *0,4582* | 2,12 | 0,4830 | *2,78* | 0,4973 |
| 1,41 | 0,4207 | 1,74 | *0,4591* | 2,14 | 0,4838 | *2,80* | 0,4974 |
| 1,42 | 0,4222 | 1,75 | *0,4599* | 2,16 | 0,4846 | *2,82* | 0,4976 |
| 1,43 | 0,4236 | 1,76 | *0,4608* | 2,18 | 0,4854 | *2,84* | 0,4977 |
| 1,44 | 0,4251 | 1,77 | *0,4616* | 2,20 | 0,4861 | *2,86* | 0,4979 |
| 1,45 | 0,4265 | 1,78 | *0,4625* | 2,22 | 0,4868 | *2,88* | 0,4980 |
| 1,46 | 0,4279 | 1,79 | *0,4633* | 2,24 | 0,4875 | *2,90* | 0,4981 |
| 1,47 | 0,4292 | 1,80 | *0,4641* | 2,26 | 0,4881 | *2,92* | 0,4982 |
| 1,48 | 0,4306 | 1,81 | *0,4649* | 2,28 | 0,4887 | *2,94* | 0,4984 |
| 1,49 | 0,4319 | 1,82 | *0,4656* | 2,30 | 0,4893 | *2,96* | 0,4985 |
| 1,50 | 0,4332 | 1,83 | *0,4664* | 2,32 | 0,4898 | *2,98* | 0,4986 |
| 1,51 | 0,4345 | 1,84 | *0,4671* | 2,34 | 0,4904 | *3,00* | 0,49865 |
| 1,52 | 0,4357 | 1,85 | *0,4678* | 2,36 | 0,4909 | *3,20* | 0,49931 |
| 1,53 | 0,4370 | 1,86 | *0,4686* | 2,38 | 0,4913 | *3,40* | 0,49966 |
| 1,54 | 0,4382 | 1,87 | *0,4693* | 2,40 | 0,4918 | *3,60* | 0,499841 |
| 1,55 | 0,4394 | 1,88 | *0,4699* | 2,42 | 0,4922 | *3,80* | 0,499928 |
| 1,56 | 0,4406 | 1,89 | *0,4706* | 2,44 | 0,4927 | *4,00* | 0,499968 |
| 1,57 | 0,4418 | 1,90 | *0,4713* | 2,46 | 0,4931 | *4,50* | 0,499997 |
| 1,58 | 0,4429 | 1,91 | *0,4719* | 2,48 | 0,4934 | *5,00* | 0,499997 |

*Таблица А.3 – Значения* $q = q(\gamma, n)$

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

**Таблица А.4 – Значения** $t_\gamma = t(\gamma, n)$

| γ / n | 0,95 | 0,99 | 0,999 |
|---|---|---|---|
| 5 | 2,78 | 4,60 | 8,61 |
| 6 | 2,57 | 4,03 | 6,86 |
| 7 | 2,45 | 3,71 | 5,96 |
| 8 | 2,37 | 3,50 | 5,41 |
| 9 | 2,31 | 2,36 | 5,04 |
| 10 | 2,26 | 3,25 | 4,78 |
| 11 | 2,23 | 3,17 | 4,59 |
| 12 | 2,20 | 3,11 | 4,44 |
| 13 | 2,18 | 3,06 | 4,32 |
| 14 | 2,16 | 3,01 | 4,22 |
| 15 | 2,15 | 2,98 | 4,14 |
| 16 | 2,13 | 2,95 | 4,07 |
| 17 | 2,12 | 2,92 | 4,02 |
| 18 | 2,11 | 2,90 | 3,97 |
| 19 | 2,10 | 2,88 | 3,92 |
| 20 | 2,093 | 2,861 | 3,883 |
| 25 | 2,064 | 2,797 | 3,745 |
| 30 | 2,045 | 2,756 | 3,659 |
| 35 | 2,032 | 2,720 | 3,600 |
| 40 | 2,023 | 2,708 | 3,558 |
| 45 | 2,016 | 2,692 | 3,527 |
| 50 | 2,009 | 2,679 | 3,502 |
| 60 | 2,001 | 2,662 | 3,464 |
| 70 | 1,996 | 2,649 | 3,439 |
| 80 | 1,001 | 2,640 | 3,418 |
| 90 | 1,987 | 2,633 | 3,403 |
| 100 | 1,984 | 2,627 | 3,392 |
| 120 | 1,980 | 2,617 | 3,374 |
| ∞ | 1,960 | 2,576 | 3,291 |

## Список литературы

**Учебные пособия**

1 Вентцель, Е. С. Теория вероятностей / Е. С Вентцель. – 4-е изд., стереотип. – Москва : Наука, Физматгиз, 1969. – 576 с.

2 Гнеденко, Б. В. Курс теории вероятностей / Б. В. Гнеденко. – 8-е изд., испр. и доп. – Москва : Едиториал УРСС, 2005. – 406 с.

3 Гмурман, В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – 9-е изд., стер. – Москва : Высшая школа, 2003. – 479 с.

4 Крамер, Г. Математические методы статистики / Г. Крамер. – Москва : Наука, 1975.

5 Кремер, Н. М. Теория вероятностей и математическая статистика / Н. М. Кремер. – Москва : Наука, 2000.

6 Севастьянов, Б. А. Курс теории вероятностей и математической статистики / Б. Севастьянов. – Москва : Наука. Гл. ред. физ.-мат. лит., 1982. – 256 с.

7 Павловский, З. Введение в математическую статистику / З. Павловский. – 5-е изд., стер. – Москва : 2011. – 220 с.

8 Уилкс, С. Математическая статистика / С. Уилкс. – Москва : Наука. Главная редакция физико-математической литературы, 1967. – 632 с.

**Задачники**

1 Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – 9-е изд., стер. – Москва : Высшая школа, 2004. – 404 с.

2 Емельянов, Т. В. Задачник по теории вероятностей и математической статистике /сост. Т. В. Емельянов, В. П. Скитович. – Ленинград : Издательство ленинградского университета, 1967. – 329 с.

3 Лозинский, С. И. Сборник задач по теории вероятностей и математической статистике / С. И. Лозинский. – Москва : Статистика, 1967. – 126 с.

Лукерьянова Елена Александровна

**МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

**(часть 2)**

**Методические указания для практических занятий и самостоятельной работы для студентов факультета «Математика и информационные технологии» направления 01.03.01 « Математика»**

Редактор Л.П. Чукомина

| | | |
|---|---|---|
| Подписано к печати 21.05.18 | Формат бумаги 60   84 1/16 | Бумага тип 65г/м$^2$ |
| Печать цифровая | Усл. печ.л. 3.31 | Уч.-изд л. 3.31 |
| Заказ №99 | Тираж 25 | Не для продажи |

БИЦ Курганского государственного университета.
640020, г. Курган, ул. Советская 63/4.
Курганский государственный университет.