МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Высшая математика»

# МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

***Методические рекомендации к лабораторным работам для студентов направления подготовки 01.03.04 «Прикладная математика» очной формы обучения***

Могилев 2021

УДК 519.22
ББК 22.172
М34

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Высшая математика» «24» декабря 2020 г., протокол № 4

Составители: канд. физ.-мат. наук, доц. В. Г. Замураев;
канд. пед. наук, доц. Е. Л. Старовойтова

Рецензент канд. физ.-мат. наук, доц. В. А. Ливинская

Методические рекомендации содержат лабораторные работы по дисциплине «Математическая статистика».

Учебно-методическое издание

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

| | |
|---|---|
| Ответственный за выпуск | В. Г. Замураев |
| Корректор | И. В. Голубцова |
| Компьютерная верстка | Н. П. Полевничая |

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 56 экз. Заказ № .

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.
Пр-т Мира, 43, 212022, г. Могилев.

# Содержание

# 1 Лабораторная работа № 1. Первичная обработка выборочных данных

***Цель работы*:** закрепление теоретических сведений по способам представления, группировки и графической интерпретации эмпирических данных.

***Теоретические сведения***

*Генеральной совокупностью* называется совокупность всех подлежащих изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом. Генеральная совокупность – это случайная величина (СВ) $X(\omega)$, заданная на пространстве элементарных событий $\Omega$ с выделенным в нем классом подмножеств событий, для которых указаны их вероятности.

*Выборочной совокупностью* (*выборкой*) называется совокупность объектов, отобранных случайным образом из генеральной совокупности.

Число объектов (наблюдений) в совокупности, генеральной или выборочной, называется ее *объемом*; обозначается соответственно через $N$ и $n$.

Пусть изучается некоторая СВ $X$. С этой целью над СВ $X$ производится ряд независимых опытов (наблюдений). В каждом из этих опытов величина $X$ принимает то или иное значение. Пусть она приняла $n_1$ раз значение $x_1$, $n_2$ раз – значение $x_2$, ..., $n_k$ раз – значение $x_k$. При этом $n_1 + n_2 + ... + n_k = n$ – объем выборки. Изменение изучаемого признака $x_i$ данной статистической совокупности называется его *вариацией*. Наблюдаемые значения объекта (признака) $x_i$, извлеченного при выборке из генеральной совокупности, называют *вариантами*. Значения $x_1$, $x_2$, …, $x_k$ называются вариантами СВ $X$. Вся совокупность значений СВ $X$ представляет собой первичный статистический материал, который подлежит дальнейшей обработке, прежде всего – упорядочению.

Операция расположения значений случайной величины (признака) по неубыванию называется *ранжированием* статистических данных. Полученная таким образом последовательность $x_{(1)}$, $x_{(2)}$, …, $x_{(n)}$ значений СВ $X$, где $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$ и $x_{(1)} = \min\limits_{1 \le i \le n} X_i$, …, $x_{(n)} = \max\limits_{1 \le i \le n} X_i$, называется *вариационным рядом*.

Числа $n_i$, показывающие, сколько раз встречаются варианты $x_i$ в ряде наблюдений, называются *частотами*, а отношение их к объему выборки – *частостями* или *относительными частотами* $W_i$, т. е. $W_i = \dfrac{n_i}{n}$, где $n = \sum\limits_{i=1}^{k} n_i$ и $\sum\limits_{i=1}^{k} W_i = 1$.

Перечень вариантов и соответствующих им частот или частостей называется *статистическим распределением* выборки или дискретным *статистическим рядом*. Записывается статистическое распределение в виде таблицы, первая строка которой содержит варианты $x_i$, а вторая – их частоты $n_i$ (или относительные частоты $W_i$).

В случае, когда число значений признака СВ $X$ велико или признак является непрерывным (т. е. когда СВ $X$ может принять любое значение в некотором интервале), составляют *интервальный статистический ряд*: в первую строку таблицы статистического распределения вписывают частичные промежутки $[x_0, x_1)$, $[x_1, x_2)$, …, $[x_{k-1}, x_k)$, которые берут обычно одинаковыми по длине: $h = x_1 - x_0 = x_2 - x_1 = \ldots$. Во второй строке статистического ряда вписывают количество наблюдений $n_i$ $\left(i = \overline{1,k}\right)$, попавших в каждый интервал.

Для определения длины интервала $h$ можно использовать формулу $h = \dfrac{R}{k}$, где $R = x_{\max} - x_{\min}$ – размах варьирования – разность между наибольшим и наименьшим значениями признака; $k = 1 + \log_2 n \approx 1 + 3{,}322\lg n$ – количество частичных интервалов.

Одним из способов обработки вариационного ряда является построение эмпирической функции распределения. *Эмпирической* (*статистической*) *функцией распределения* называется функция $F^*(x)$, определяющая для каждого значения $x$ частость события $\{X < x\}$, т. е. $F_n^*(x) = W\{X < x\} = \dfrac{n_x}{n}$, где $n$ – объем выборки; $n_x$ – число наблюдений, меньших $x$. Эмпирическая функция распределения $F_n^*(x)$ удовлетворяет тем же условиям, что и теоретическая функция распределения $F(x)$, и является оценкой вероятности события $\{X < x\}$, т. е. оценкой теоретической функции распределения $F(x)$ СВ $X$.

Статистическое распределение изображается графически (для наглядности) в виде *полигона* и *гистограммы*. Полигон служит для изображения дискретного статистического ряда. Полигоном абсолютных частот называют ломаную, отрезки которой соединяют точки с координатами $(x_1;n_1)$, $(x_2;n_2)$, …, $(x_k;n_k)$. Полигоном относительных частот (полигоном частостей) называют ломаную, отрезки которой соединяют точки с координатами $(x_1;W_1)$, $(x_2;W_2)$, …, $(x_k;W_k)$. Варианты $x_i$ откладываются на оси абсцисс, а частоты $n_i$ и относительные частоты $W_i$ – на оси ординат. Полигон относительных частот является статистическим аналогом многоугольника распределения.

Для непрерывно распределенного признака полигон частот строят, взяв в качестве значений $x_1$, $x_2$, …, $x_k$ середины интервалов, т. е. значения $x_i^* = \dfrac{x_{i+1} + x_i}{2}$.

Только для непрерывно распределенной случайной величины строят гистограмму частот – ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины $h$, а высоты этих прямоугольников равны отношению $\dfrac{n_i}{h}$ – плотности частоты. Для построения гистограммы относительных частот строят ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины $h$, а высоты этих прямоугольников равны отношению $\frac{W_i}{h} = \frac{n_i}{n \cdot h}$ – плотности частости. Гистограмма частот является статистическим аналогом дифференциала функции распределения (плотности) $f(x)$ СВ $X$.

Площадь гистограммы относительных частот равна единице, а площадь гистограммы абсолютных частот равна объему выборки.

Накопленной частотой в точке $x_i$ называется суммарная частота элементов статистической совокупности со значениями признака, меньшими чем $x_i$.

Пусть статистическое распределение выборки объема $n$ имеет следующий вид (таблица 1).

Таблица 1

| | | | | | |
|---|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | $x_3$ | … | $x_k$ |
| $n_i$ | $n_1$ | $n_2$ | $n_3$ | … | $n_k$ |

*Выборочным средним* $\overline{x}_B$ (другие обозначения: $\overline{x}$, $M^*(X)$, $m_x^*$) называется среднее арифметическое всех значений выборки, которое вычисляется по одной из формул:

$$\overline{x}_B = \frac{1}{n}\sum_{i-1}^{k} x_i \cdot n_i \text{ или } \overline{x}_B = \sum_{i-1}^{k} x_i \cdot W_i,$$

где $W_i$ – относительная частота, $W_i = \frac{n_i}{n}$.

В случае интервального статистического ряда в качестве $x_i$ берут середины его интервалов, а в качестве $n_i$ – соответствующие им частоты.

*Выборочной дисперсией* $D_B$ называется среднее арифметическое квадратов отклонений значений выборки от выборочной средней $\overline{x}_B$, т. е.

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

*Выборочное среднее квадратическое отклонение* выборки определяется формулой $\sigma_B = \sqrt{D_B}$.

*Исправленной выборочной дисперсией* называется величина $S^2$, которая находится по одной из формул:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_B)^2 \cdot n_i \text{ или } S^2 = \frac{n}{n-1}D_B.$$

Исправленным выборочным средним квадратическим отклонением называется величина $S = \sqrt{S^2}$.

Для непрерывно распределенного признака формулы для выборочных средних будут такими же, но в качестве значений $x_1$, $x_2$, $x_3$, …, $x_k$ надо брать не концы промежутков $[x_0; x_1)$, $[x_1; x_2)$, …, а их середины $\frac{x_0 + x_1}{2}$, $\frac{x_1 + x_2}{2}$, …, а в качестве $n_i$ – соответствующие им частоты.

*Размахом вариации* называется число $R = x_{(n)} - x_{(1)}$, где $x_{(1)} = \min\limits_{1 \le k \le n} x_k$ и $x_{(n)} = \max\limits_{1 \le k \le n} x_k$, или $R = x_{\max} - x_{\min}$, где $x_{\max}$ – наибольшая и $x_{\min}$ – наименьшая варианта ряда. *Модой* $M_0^*$ вариационного ряда называется варианта, имеющая наибольшую частоту. *Медианой* $M_e^*$ вариационного ряда называется значение признака (СВ $X$), приходящееся на середину ряда, причем:

– если $n = 2k$ (т. е. ряд $(x_{(1)}, x_{(2)}, \ldots, x_{(k)}, x_{(k+1)}, x_{(2k)})$ имеет четное число членов), то $M_e^* = \frac{x_{(k)} + x_{(k+1)}}{2}$;

– если $n = 2k + 1$ (т. е. ряд $(x_{(1)}, x_{(2)}, \ldots, x_{(k)}, x_{(k+1)}, x_{(2k+1)})$ имеет нечетное число членов), то $M_e^* = x_{(k+1)}$.

## ***Часть 1. Первичная обработка выборочных данных: генеральная совокупность и выборка (дискретный ряд)***

### ***Постановка задачи***

***Задача***. Дан статистический материал: случайная величина $X$ – результат $n$ измерений некоторой величины. Требуется:

1) построить вариационный и статистический ряды;
2) построить полигон абсолютных и относительных частот;
3) найти эмпирическую функцию распределения $F_n^*(x)$ и построить ее график;
4) найти выборочные значения числовых характеристик случайной величины $X$: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднее квадратическое отклонение, размах вариации, моду, медиану.

### ***Пример выполнения***

Случайная величина $X$ – размеры мужской обуви, проданной магазином за день.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 44 | 39 | 43 | 42 | 41 | 38 | 43 | 42 | 44 | 42 |
| 43 | 41 | 40 | 40 | 42 | 39 | 40 | 42 | 41 | 45 |
| 43 | 42 | 43 | 38 | 39 | 43 | 41 | 40 | 43 | 41 |
| 44 | 45 | 43 | 42 | 45 | 43 | 38 | 43 | 42 | 43 |
| 39 | 42 | 43 | 44 | 42 | 41 | 43 | 43 | 44 | 45 |

1 Проранжировав (упорядочив по неубыванию) статистические данные (т. е. исходный ряд), получим вариационный ряд

$$\left( x_{(1)}, x_{(2)}, ..., x_{(8)} \right) = \left( 38, 39, 40, 41, 42, 43, 44, 45 \right).$$

Варианты (значения, которые может принимать СВ $X$ ) в данном случае следующие: $x_1 = 38$; $x_2 = 39$; $x_3 = 40$; $x_4 = 41$; $x_5 = 42$; $x_6 = 43$; $x_7 = 44$; $x_8 = 45$.

Подсчитав частоту $n_i$ (т. е. сколько раз в вариационном ряду встречается данная варианта) и относительные частоты $W_i = \dfrac{n_i}{n}$ вариант, получим статистическое распределение выборки (дискретный статистический ряд) (таблица 2).

Таблица 2

| $x_i$ | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 3 | 4 | 4 | 6 | 10 | 14 | 5 | 4 |
| $W_i$ | 0,06 | 0,08 | 0,08 | 0,12 | 0,2 | 0,28 | 0,1 | 0,08 |

Контроль: $3 + 4 + 4 + 6 + 10 + 14 + 5 + 4 = 50$;

$$0,06 + 0,08 + 0,08 + 0,12 + 0,2 + 0,28 + 0,1 + 0,08 = 1.$$

2 Для построения полигона абсолютных частот (рисунок 1) на координатной плоскости изобразим ломаную, соединяющую точки с координатами $(x_1; n_1)$, $(x_2; n_2)$, …, $(x_8; n_8)$.

Для построения полигона относительных частот (рисунок 2) на координатной плоскости изобразим ломаную, соединяющую точки с координатами $(x_1; W_1)$, $(x_2; W_2)$, …, $(x_8; W_8)$.

3 Найдем эмпирическую функцию распределения, воспользовавшись формулой $F_n^*(x) = W\{X < x\} = \dfrac{n_x}{n}$, где $n$ – объем выборки; $n_x$ – число наблюдений, меньших $x$.

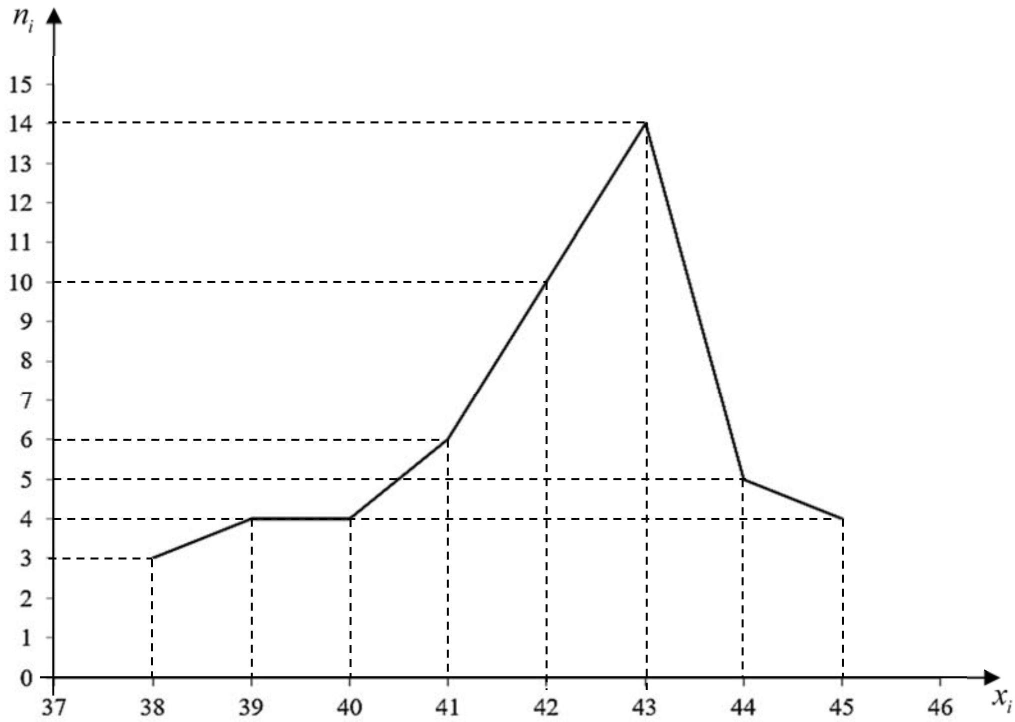


Рисунок 1

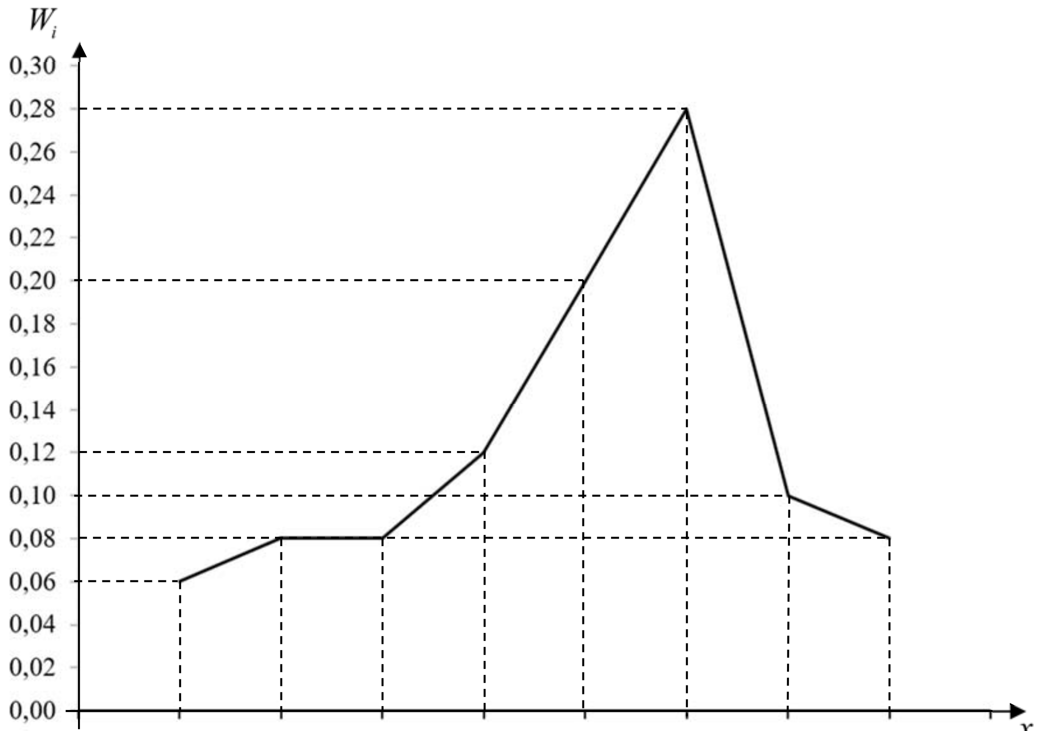


Рисунок 2

$$F_8^*(x)=\begin{cases}0 & \text{при} \quad x\le 38,\\ 0,06 & \text{при} \quad 38<x\le 39,\\ 0,06+0,08=0,14 & \text{при} \quad 39<x\le 40,\\ 0,06+0,08+0,08=0,22 & \text{при} \quad 40<x\le 41,\\ 0,06+0,08+0,08+0,12=0,34 & \text{при} \quad 41<x\le 42,\\ 0,06+0,08+0,08+0,12+0,2=0,54 & \text{при} \quad 42<x\le 43,\\ 0,06+0,08+0,08+0,12+0,2+0,28=0,82 & \text{при} \quad 43<x\le 44,\\ 0,06+0,08+0,08+0,12+0,2+0,28+0,1=0,92 & \text{при} \quad 44<x\le 45,\\ 0,06+0,08+0,08+0,12+0,2+0,28+0,1+0,08=1 & \text{при} \quad x>45.\end{cases}$$

График эмпирической функции распределения имеет вид (рисунок 3).

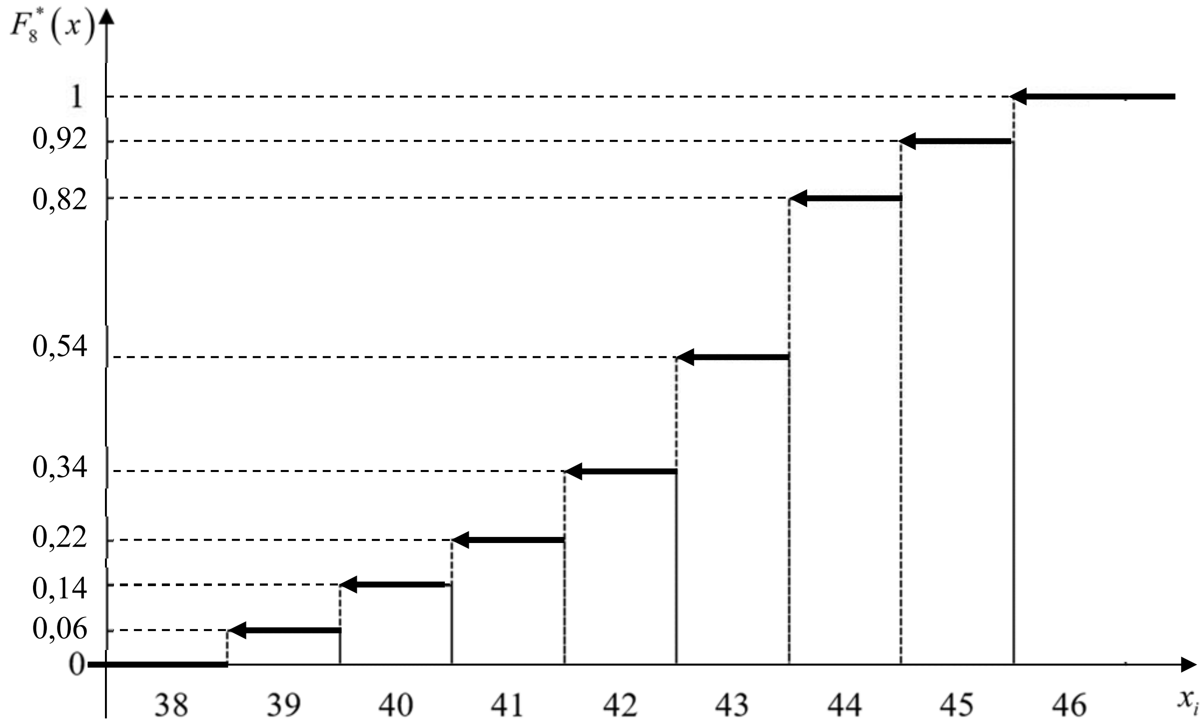


Рисунок 3

4 Найдём характеристики выборки.
Найдем выборочное среднее:

$$\overline{x}_B=\frac{1}{50}\sum_{i=1}^{8}x_i\cdot n_i=\frac{1}{50}(38\cdot 3+39\cdot 4+40\cdot 4+41\cdot 6+42\cdot 10+$$

$$+43\cdot 14+44\cdot 5+45\cdot 4)=\frac{1}{50}\cdot 2098=41,96.$$

Найдем выборочную дисперсию:

$$D_B = \frac{1}{50}\sum_{i=1}^{8}(x_i - \overline{x}_B)^2 \cdot n_i = \frac{1}{50}\Big((38-41{,}96)^2 \cdot 3 + (39-41{,}96)^2 \cdot 4 +$$

$$+(40-41{,}96)^2 \cdot 4 + (41-41{,}96)^2 \cdot 6 + (42-41{,}96)^2 \cdot 10 +$$

$$+(43-41{,}96)^2 \cdot 14 + (44-41{,}96)^2 \cdot 5 + (45-41{,}96)^2 \cdot 4\Big) = \frac{1}{50} \cdot 175{,}92 = 3{,}52.$$

Найдем выборочное среднее квадратическое отклонение:

$$\sigma_B = \sqrt{D_B} = \sqrt{3{,}52} \approx 1{,}88.$$

Найдем исправленную выборочную дисперсию:

$$S^2 = \frac{50}{50-1} \cdot 3{,}52 \approx 3{,}59.$$

Найдем исправленное выборочное среднее квадратическое отклонение:

$$S = \sqrt{3{,}59} \approx 1{,}89.$$

Найдем размах вариации:

$$R = x_{(8)} - x_{(1)} = 45 - 38 = 7.$$

Найдем моду (варианту, имеющую наибольшую частоту):

$$M_0^* = 43.$$

Так как вариационный ряд $\left(x_{(1)}, x_{(2)}, ..., x_{(8)}\right) = (38, 39, 40, 41, 42, 43, 44, 45)$ имеет четное число членов $(8 = 2k \Rightarrow k = 4)$, то найдем медиану:

$$M_e^* = \frac{x_{(k)} + x_{(k+1)}}{2} = \frac{x_{(4)} + x_{(5)}}{2} = \frac{41+42}{2} = 41{,}5.$$

***Контрольные вопросы***

1 Дайте определение понятий: «частота», «относительная частота», «вариационный ряд».

2 Какой вариационный ряд называется дискретным, а какой – интервальным?

3 Каким образом строятся полигоны частот и относительных частот?

4 Дайте определение эмпирической функции распределения.

5 По каким формулам рассчитываются выборочное среднее, выборочная дисперсия, размах вариации, мода и медиана?

***Варианты заданий к лабораторной работе № 1 (часть 1)***

***Задача***. Дан статистический материал: случайная величина $X$ – результат $n$ измерений некоторой величины. Требуется:

1) построить вариационный и статистический ряды;

2) построить полигон абсолютных и относительных частот;

3) найти эмпирическую функцию распределения $F_n^*(x)$ и построить ее график;

4) найти выборочные значения числовых характеристик случайной величины $X$: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднее квадратическое отклонение, размах вариации, моду, медиану.

Указание: $i$ – порядковый номер элемента в выборке (таблица 3), начиная с которого следует отсчитать $n$ чисел, двигаясь по строкам слева направо, для формирования статистического материала для своего варианта.

Таблица 3

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 11 | 10 | 15 | 19 | 22 | 30 | 24 | 26 | 28 | 18 | 27 | 23 | 9 | 14 | 7 |
| $n$ | 100 | 120 | 100 | 110 | 110 | 110 | 120 | 110 | 120 | 115 | 120 | 115 | 110 | 115 | 110 |

Продолжение таблицы 3

| Номер варианта | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 2 | 1 | 29 | 4 | 21 | 6 | 13 | 25 | 8 | 20 | 3 | 12 | 5 | 17 | 16 |
| $n$ | 110 | 120 | 110 | 115 | 110 | 115 | 100 | 110 | 120 | 115 | 110 | 110 | 100 | 100 | 110 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 38 | 43 | 49 | 32 | 42 | 42 | 47 | 33 | 39 | 38 | 29 | 51 | 40 | 28 | 29 |
| 47 | 43 | 42 | 40 | 28 | 27 | 43 | 26 | 38 | 56 | 26 | 53 | 54 | 53 | 37 |
| 48 | 42 | 44 | 46 | 49 | 53 | 27 | 30 | 51 | 35 | 35 | 37 | 30 | 41 | 29 |
| 40 | 48 | 49 | 45 | 55 | 26 | 46 | 49 | 35 | 30 | 52 | 35 | 37 | 36 | 44 |
| 30 | 32 | 44 | 26 | 49 | 27 | 56 | 51 | 50 | 40 | 51 | 55 | 35 | 42 | 28 |
| 25 | 54 | 45 | 44 | 45 | 31 | 47 | 38 | 29 | 29 | 38 | 41 | 42 | 44 | 55 |
| 32 | 39 | 42 | 42 | 32 | 46 | 55 | 42 | 40 | 56 | 25 | 40 | 31 | 57 | 34 |
| 28 | 57 | 54 | 29 | 32 | 35 | 42 | 52 | 35 | 48 | 52 | 40 | 52 | 30 | 52 |
| 32 | 44 | 31 | 40 | 47 | 42 | 32 | 45 | 54 | 54 | 53 | 47 | 39 | 53 | 30 |
| 37 | 26 | 49 | 36 | 43 | 40 | 48 | 50 | 54 | 36 | 29 | 50 | 45 | 38 | 44 |
| 37 | 51 | 27 | 47 | 45 | 38 | 50 | 39 | 55 | 28 | 40 | 34 | 47 | 26 | 30 |

## ***Часть 2. Первичная обработка выборочных данных: генеральная совокупность и выборка (интервальный ряд)***

***Задача***. По данному статистическому материалу требуется:

1) составить статистический ряд распределения случайной величины $X$;

2) составить интервальный статистический ряд распределения случайной величины $X$;

3) построить полигон и гистограмму относительных частот;

4) найти эмпирическую функцию распределения и построить ее график;

5) найти характеристики выборки: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднее квадратическое отклонение.

### ***Пример выполнения***

Случайная величина $X$ – рост (с точностью до сантиметра) 30 наудачу отобранных студентов.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 178 | 160 | 154 | 183 | 155 | 153 | 167 | 186 | 163 | 155 | 157 | 175 | 170 | 166 | 159 |
| 173 | 182 | 167 | 171 | 169 | 179 | 165 | 156 | 179 | 158 | 171 | 175 | 173 | 164 | 172 |

1 Проранжировав (упорядочив по неубыванию) статистические данные и подсчитав частоту $n_i$, получим статистический ряд распределения (таблица 4).

2 Так как СВ $X$ – рост студента – непрерывная случайная величина, то для исследования данной выборки составим интервальный статистический ряд, для чего найдем количество и длину частичных интервалов. Для нахождения числа интервалов воспользуемся формулой $k = 1 + \log_2 n$. В данном случае объем выборки $n = 30$, поэтому $k = 1 + \log_2 30 \approx 1 + 4{,}91 = 5{,}91$. Так как число интервалов – целое число, то разобьем исходные данные на $k = 6$ частичных интервалов. Для нахождения длины частичных интервалов воспользуемся формулой

$h = \frac{R}{k}$. В данном случае $x_{\max} = 186$, $x_{\min} = 153$. Значит, размах варьирования $R = x_{\max} - x_{\min} = 186 - 153 = 33$. Тогда $h = \frac{33}{6} \approx 5{,}5$.

Таблица 4

| $x_i$ | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 163 | 164 | 165 | 166 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Продолжение таблицы 4

| $x_i$ | 167 | 169 | 170 | 171 | 172 | 173 | 175 | 178 | 179 | 182 | 183 | 186 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 |

Подсчитаем число студентов $n_i$, рост которых попадает в каждый из полученных частичных интервалов, и составим интервальный статистический ряд (добавив столбцы с данными, которые понадобятся впоследствии) (таблица 5).

Таблица 5

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| $[x_i; x_{i+1})$ – частичные интервалы | $n_i$ – частота в интервале | $W_i = \frac{n_i}{n}$ – относительная частота | $n_x$ – накопленная частота | $\frac{n_x}{n}$ – относительная накопленная частота | $\frac{W_i}{h}$ – плотность относительной частоты | $x_i^* = \frac{x_{i+1} + x_i}{2}$ – середины интервалов |
| $[153;158{,}5)$ | 7 | 7/30 | 7 | 7/30 | 0,042 | 155,75 |
| $[158{,}5;164)$ | 3 | 3/30 | 10 | 10/30 | 0,018 | 161,25 |
| $[164;169{,}5)$ | 6 | 6/30 | 16 | 16/30 | 0,036 | 166,75 |
| $[169{,}5;175)$ | 6 | 6/30 | 22 | 22/30 | 0,036 | 172,25 |
| $[175;180{,}5)$ | 5 | 5/30 | 27 | 27/30 | 0,030 | 177,75 |
| $[180{,}5;186]$ | 3 | 3/30 | 30 | 1 | 0,018 | 183,25 |

Контроль: $7 + 3 + 6 + 6 + 5 + 3 = 30 = n$; $\frac{7}{30} + \frac{3}{30} + \frac{6}{30} + \frac{6}{30} + \frac{5}{30} + \frac{3}{30} = 1$.

3 Для построения полигона относительных частот (рисунок 4) на координатной плоскости по данным столбцов 3 и 7 таблицы 5 изобразим ломаную, соединяющую точки с координатами $(x_1^*; W_1)$, $(x_2^*; W_2)$, …, $(x_6^*; W_6)$.

Для построения гистограммы относительных частот (рисунок 5) на координатной плоскости по данным столбцов 1 и 6 таблицы 5 построим ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы $(x_i; x_{i+1})$, а высоты равны $\dfrac{W_i}{h}$ – плотности относительной частоты.

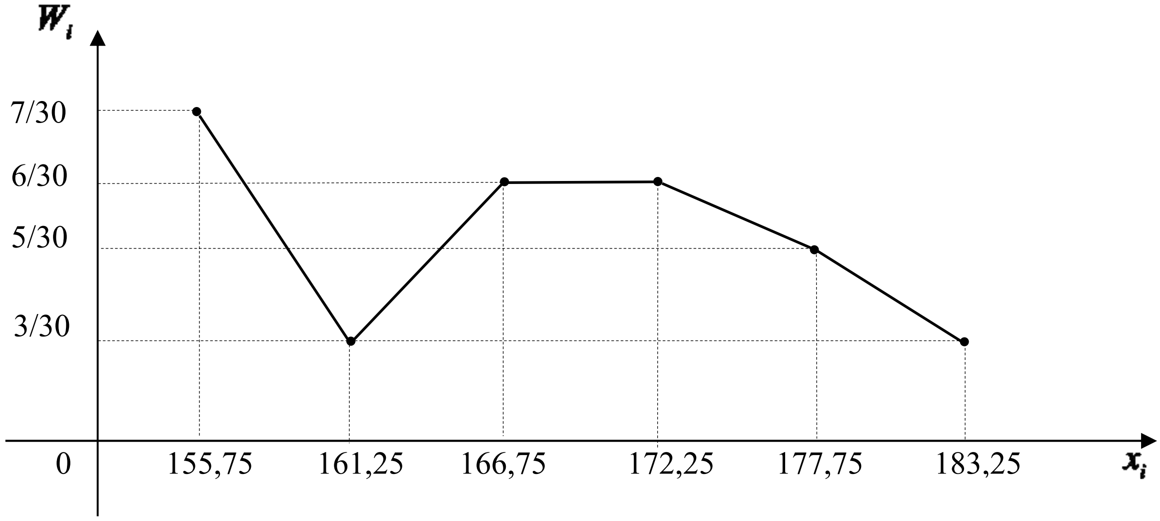


Рисунок 4

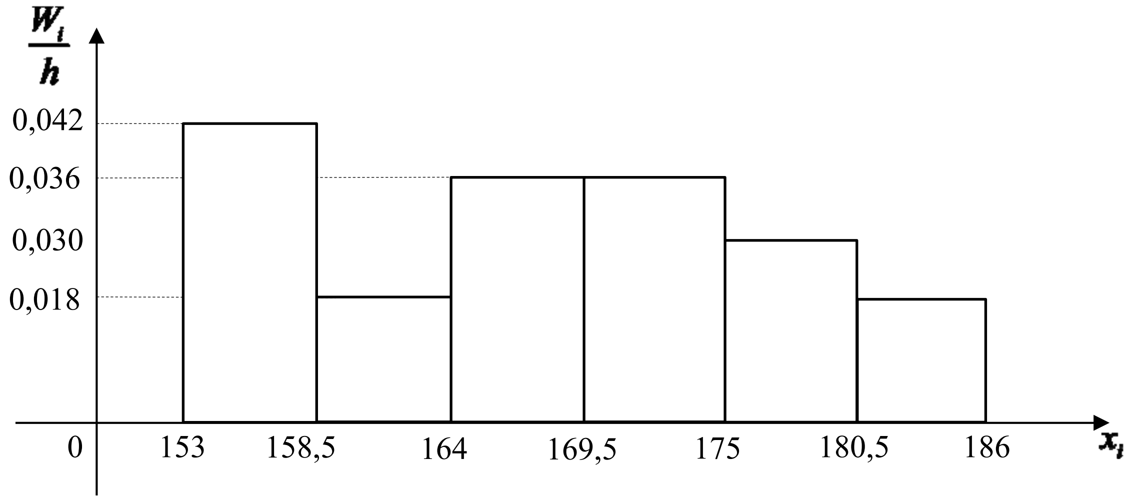


Рисунок 5

4 Эмпирическую функцию распределения найдем по данным столбца 5 таблицы 5 по формуле $F^*(x) = W(X < x) = \dfrac{n_x}{n}$.

Тогда эмпирическая функция распределения будет иметь вид:

$$F^*(x) = \begin{cases} 0, & x \le 153, \\ 7/30, & x = 158{,}5, \\ 10/30, & x = 164, \\ 16/30, & x = 169{,}5, \\ 22/30, & x = 175, \\ 27/30, & x = 180{,}5, \\ 1, & x \ge 186. \end{cases}$$

График эмпирической функции распределения имеет следующий вид (рисунок 6).

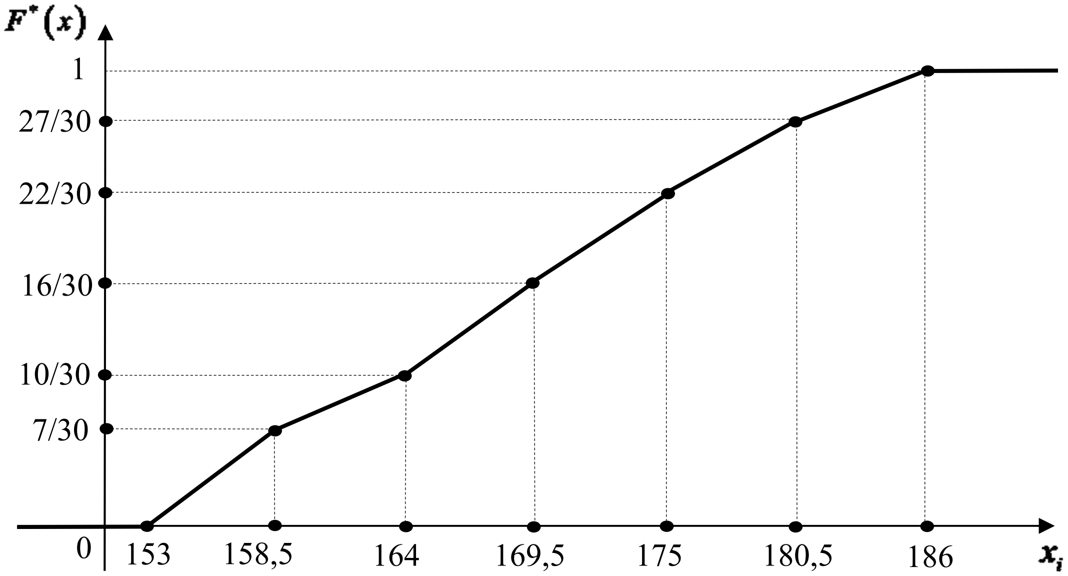


Рисунок 6

5 Найдем характеристики выборки.

Найдем выборочное среднее (в качестве $x_i$ берем середины интервалов, а в качестве $n_i$ – соответствующие им частоты):

$$\overline{x}_B = \frac{1}{30}\sum_{i=1}^{k} x_i^* \cdot n_i = \frac{1}{30}\big(155{,}75 \cdot 7 + 161{,}25 \cdot 3 + 166{,}75 \cdot 6 + 172{,}25 \cdot 6 + $$
$$+ 177{,}75 \cdot 5 + 183{,}25 \cdot 3\big) = 168{,}22.$$

Найдем выборочную дисперсию (в качестве $x_i$ берем середины интервалов, а в качестве $n_i$ – соответствующие им частоты):

$$D_B = \frac{1}{30}\sum_{i=1}^{k}\left( x_i^* \right)^2 \cdot n_i - \left( \overline{x}_B \right)^2 = \frac{1}{30}\big(155,75^2 \cdot 7 + 161,25^2 \cdot 3 + 166,75^2 \cdot 6 +$$

$$+ 172,25^2 \cdot 6 + 177,75^2 \cdot 5 + 183,25^2 \cdot 3\big) - 168,22^2 = 81,43.$$

Найдем выборочное среднее квадратическое отклонение:

$$\sigma_B = \sqrt{81,43} \approx 9,02.$$

Найдем исправленную выборочную дисперсию:

$$S^2 = \frac{30}{30-1} \cdot 81,43 \approx 84,24\,.$$

Найдем исправленное выборочное среднее квадратическое отклонение:

$$S = \sqrt{84,24} \approx 9,18\,.$$

***Варианты заданий к лабораторной работе № 1 (часть 2)***

***Задача***. По данному статистическому материалу требуется:

1) составить статистический ряд распределения случайной величины $X$;

2) составить интервальный статистический ряд распределения случайной величины $X$;

3) построить полигон и гистограмму относительных частот;

4) найти эмпирическую функцию распределения и построить ее график;

5) найти характеристики выборки: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение, исправленную выборочную дисперсию, исправленное выборочное среднее квадратическое отклонение.

Указание: $i$ – порядковый номер элемента в выборке (таблица 6), начиная с которого следует отсчитать $n$ чисел, двигаясь по строкам слева направо, для формирования статистического материала для своего варианта.

Таблица 6

| Номер варианта | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 11 | 10 | 15 | 19 | 22 | 30 | 24 | 26 | 28 | 18 | 27 | 23 | 9 | 14 | 7 |
| $n$ | 100 | 120 | 100 | 110 | 110 | 110 | 120 | 110 | 120 | 115 | 120 | 115 | 110 | 115 | 110 |

Продолжение таблицы 6

| Номер варианта | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 2 | 1 | 29 | 4 | 21 | 6 | 13 | 25 | 8 | 20 | 3 | 12 | 5 | 17 | 16 |
| $n$ | 110 | 120 | 110 | 115 | 110 | 115 | 100 | 110 | 120 | 115 | 110 | 110 | 100 | 100 | 110 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 40 | 25 | 29 | 27 | 66 | 45 | 62 | 41 | 39 | 22 | 59 | 23 | 42 | 66 |
| 78 | 68 | 36 | 23 | 77 | 55 | 39 | 51 | 57 | 62 | 24 | 45 | 36 | 71 | 40 |
| 64 | 60 | 26 | 56 | 30 | 21 | 28 | 21 | 53 | 60 | 77 | 74 | 31 | 31 | 78 |
| 28 | 34 | 45 | 68 | 23 | 68 | 31 | 46 | 68 | 66 | 44 | 73 | 39 | 27 | 26 |
| 77 | 42 | 70 | 35 | 80 | 32 | 51 | 47 | 71 | 79 | 20 | 73 | 35 | 21 | 55 |
| 48 | 46 | 28 | 63 | 68 | 46 | 49 | 56 | 74 | 50 | 32 | 22 | 60 | 77 | 28 |
| 44 | 36 | 33 | 38 | 71 | 60 | 47 | 34 | 54 | 37 | 47 | 74 | 33 | 73 | 51 |
| 40 | 33 | 50 | 59 | 77 | 43 | 57 | 35 | 52 | 29 | 75 | 33 | 58 | 50 | 56 |
| 37 | 26 | 20 | 36 | 58 | 59 | 43 | 79 | 42 | 57 | 23 | 59 | 79 | 61 | 45 |
| 41 | 73 | 23 | 53 | 25 | 52 | 46 | 49 | 74 | 23 | 43 | 48 | 69 | 52 | 64 |
| 62 | 21 | 79 | 44 | 34 | 72 | 21 | 72 | 71 | 65 | 20 | 51 | 21 | 57 | 55 |

## 2 Лабораторная работа № 2. Точечное оценивание

***Цель работы***: практическое ознакомление с методами статистического оценивания параметров генеральной совокупности по выборочным данным (метод моментов, метод наибольшего правдоподобия).

***Теоретические сведения***

*Статистической оценкой* $\Theta^*$ неизвестного параметра $\Theta$ теоретического распределения называется функция $f = f\left(X_1, X_2, ..., X_n\right)$ от наблюдаемых случайных величин $X_1$, $X_2$, …, $X_n$.

*Точечной* называется статистическая оценка, которая определяется одним числом $\Theta^* = f\left(x_1, x_2, ..., x_n\right)$, где $x_1, x_2, ..., x_n$ – результаты $n$ наблюдений над количественным признаком $X$ (выборка).

*Несмещенной* называется точечная оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.

*Смещенной* называется точечная оценка, математическое ожидание которой не равно оцениваемому параметру.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя: $\overline{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i n_i$, где $x_i$ – варианта выборки; $p_i$ – частота варианты $x_i$; $n = \sum_{i=1}^{k} n_i$ – объем выборки.

Если первоначальные варианты $x_i$ – большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число $C$, т. е. перейти к *условным вариантам* $u_i = x_i - C$ (в качестве $C$ выгодно принять число, близкое к выборочной средней; поскольку выборочная средняя неизвестна, число $C$ выбирают «на глаз»). Тогда $\overline{x}_B = C + \frac{1}{n}\sum_{i=1}^{k} u_i n_i$.

Смещенной оценкой генеральной дисперсии служит выборочная дисперсия, которую вычисляют по одной из формул:

$$D_B = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_B)^2 n_i \text{ или } D_B = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2.$$

Если первоначальные варианты $x_i$ – большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число $C$, равное выборочной средней или близкое к ней, т. е. перейти к условным вариантам $u_i = x_i - C$ (дисперсия при этом не изменится). Тогда

$$D_B(X) = D_B(u) = \overline{u^2} - \left(\overline{u}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} u_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} u_i n_i\right)^2.$$

Если первоначальные варианты являются десятичными дробями с $k$ десятичными знаками после запятой, то, чтобы избежать действий с дробями, умножают первоначальные варианты на постоянное число $C = 10^k$, т. е. переходят к условным вариантам $u_i = Cx_i$. При этом дисперсия увеличится в $C^2$ раз. Поэтому, найдя дисперсию условных вариант, надо разделить ее на $C^2$:

$$D_B(X) = \frac{D_B(u)}{C^2}.$$

Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия, которую находят по одной из формул:

$$s^2 = \frac{n}{n-1}D_B = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_B)^2 n_i \text{ или } s_X^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2\right).$$

В условных вариантах формула для вычисления исправленной выборочной дисперсии имеет вид:

$$s_u^2 = \frac{1}{n-1}\left( \sum_{i=1}^{k} u_i^2 n_i - \left( \frac{1}{n} \sum_{i=1}^{k} u_i n_i \right)^2 \right),$$

причем если $u_i = x_i - C$, то $s_X^2 = s_u^2$, а если $u_i = Cx_i$, то $s_X^2 = \frac{s_u^2}{C^2}$.

*Метод моментов* точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов соответствующим эмпирическим моментам того же порядка.

Если распределение определяется одним параметром, то для его отыскания приравнивают один теоретический момент одному эмпирическому моменту того же порядка. Если распределение определяется двумя параметрами, то приравнивают два теоретических момента двум соответствующим эмпирическим моментам того же порядка.

*Метод наибольшего правдоподобия* точечной оценки неизвестных параметров заданного распределения сводится к отысканию максимума функции одного или нескольких оцениваемых параметров.

1 Дискретные случайные величины. Пусть $X$ – дискретная случайная величина, которая в результате $n$ опытов приняла возможные значения $x_1$, $x_2$, …, $x_n$. Допустим, что вид закона распределения величины $X$ задан, но неизвестен параметр $\Theta$, которым определяется этот закон. Требуется найти его точечную оценку $\Theta^* = \Theta^*(x_1, x_2, ..., x_n)$.

Обозначим вероятность того, что в результате испытания величина $X$ примет значение $x_i$ через $p(x_i;\Theta)$. *Функцией правдоподобия* дискретной случайной величины $X$ называется функция аргумента $\Theta$ следующего вида:

$$L(x_1, x_2, ..., x_n;\Theta) = p(x_1;\Theta) \cdot p(x_2;\Theta) \cdot ... \cdot p(x_n;\Theta).$$

Оценкой *наибольшего правдоподобия* параметра $\Theta$ называется такое его значение $\Theta^*$, при котором функция правдоподобия достигает максимума.

Так как функции $L$ и $\ln L$ достигают максимума при одном и том же значении $\Theta$, то вместо отыскания максимума функции $L$ удобнее находить максимум функции $\ln L$.

Логарифмической функцией правдоподобия называют функцию $\ln L$.

Точку максимума функции $\ln L$ аргумента $\Theta$ ищут по следующей схеме:

1) находят производную $\frac{d \ln L}{d\Theta}$;

2) приравнивают производную к нулю и находят критическую точку $\Theta^*$ – корень полученного уравнения (его называют уравнением правдоподобия);

3) находят вторую производную $\frac{d^2 \ln L}{d\Theta^2}$; если вторая производная при $\Theta = \Theta^*$ отрицательна, то $\Theta^*$ – точка максимума. Найденную точку максимума $\Theta^*$ принимают в качестве оценки наибольшего правдоподобия параметра $\Theta$;

2 Непрерывные случайные величины. Пусть $X$ – непрерывная случайная величина, которая в результате $n$ испытаний приняла возможные значения $x_1$, $x_2$, …, $x_n$. Допустим, что вид плотности распределения – функции $f(x)$ – задан, но неизвестен параметр $\Theta$, которым определяется эта функция.

Функцией правдоподобия непрерывной случайной величины $X$ называется функция аргумента $\Theta$ следующего вида:

$$L(x_1, x_2, ..., x_n; \Theta) = f(x_1; \Theta) \cdot f(x_2; \Theta) \cdot ... \cdot f(x_n; \Theta).$$

Оценку наибольшего правдоподобия неизвестного параметра распределения непрерывной случайной величины ищут так же, как в случае дискретной случайной величины.

Если плотность распределения $f(x)$ непрерывной случайной величины определяется двумя неизвестными параметрами $\Theta_1$ и $\Theta_2$, то функция правдоподобия есть функция двух независимых аргументов $\Theta_1$ и $\Theta_2$, которая имеет вид:

$$L = f(x_1; \Theta_1; \Theta_2) \cdot f(x_2; \Theta_1; \Theta_2) \cdot ... \cdot f(x_n; \Theta_1; \Theta_2).$$

Далее находят логарифмическую функцию правдоподобия и для отыскания ее максимума составляют и решают систему

$$\begin{cases} \dfrac{d \ln L}{d\Theta} = 0, \\ \dfrac{d^2 \ln L}{d\Theta^2} = 0. \end{cases}$$

***Задачи для самостоятельного решения***

1 Из генеральной совокупности извлечена выборка объема $n = 50$ (таблица 7).

Таблица 7

| $x_i$ | 2 | 5 | 7 | 10 |
|---|---|---|---|---|
| $n_i$ | 16 | 12 | 8 | 14 |

Найти несмещенную оценку генеральной средней.

2 Найти выборочную среднюю и выборочную дисперсию по данному распределению выборки объема $n = 10$ (таблица 8).

Таблица 8

| $x_i$ | 1250 | 1270 | 1280 |
|---|---|---|---|
| $n_i$ | 2 | 5 | 3 |

3 По выборке объема $n = 41$ найдена смещенная оценка $D_B = 3$ генеральной дисперсии. Найти несмещенную оценку генеральной совокупности.

4 Найти выборочную дисперсию по данному распределению выборки объема $n = 10$ (таблица 9).

Таблица 9

| $x_i$ | 0,01 | 0,04 | 0,08 |
|---|---|---|---|
| $n_i$ | 5 | 3 | 2 |

5 Даны результаты измерения роста (в сантиметрах) случайно отобранных 100 студентов (таблица 10).

Таблица 10

| Рост | 154–158 | 158–162 | 162–166 | 166–170 | 170–174 | 174–178 | 178–182 |
|---|---|---|---|---|---|---|---|
| Число студентов | 10 | 14 | 26 | 28 | 12 | 8 | 2 |

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

6 Найти исправленную выборочную дисперсию по данному распределению выборки $n = 10$ (таблица 11).

Таблица 11

| $x_i$ | 102 | 104 | 108 |
|---|---|---|---|
| $n_i$ | 2 | 3 | 5 |

7 Найти исправленную выборочную дисперсию по данному распределению выборки $n = 100$ (таблица 12).

Таблица 12

| $x_i$ | 1250 | 1275 | 1280 | 1300 |
|---|---|---|---|---|
| $n_i$ | 20 | 25 | 50 | 5 |

8 Найти исправленную выборочную дисперсию по данному распределению выборки $n = 20$ (таблица 13).

Таблица 13

| $x_i$ | 0,1 | 0,5 | 0,7 | 0,9 |
|---|---|---|---|---|
| $n_i$ | 6 | 12 | 1 | 1 |

9 Найти исправленную выборочную дисперсию по данному распределению выборки $n = 10$ (таблица 14).

Таблица 14

| $x_i$ | 23,5 | 26,1 | 28,2 | 30,4 |
|---|---|---|---|---|
| $n_i$ | 2 | 3 | 4 | 1 |

10 Случайная величина $X$ (число семян сорняков в пробе зерна) распределена по закону Пуассона. В таблице 15 приведено распределение семян сорняков в $n = 1000$ пробах зерна (в первой строке указано количество $x_i$ сорняков в одной пробе; во второй строке указана частота $n_i$ – число проб, содержащих $x_i$ семян сорняков).

Таблица 15

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 405 | 366 | 175 | 50 | 8 | 4 | 2 |

Найти методом моментов точечную оценку неизвестного параметра $\lambda$ распределения Пуассона.

11 Случайная величина $X$ (число нестандартных изделий в партии изделий) распределена по закону Пуассона. В таблице 16 приведено распределение нестандартных изделий в $n = 200$ партиях (в первой строке указано количество $x_i$ нестандартных изделий в одной партии; во второй строке указана частота $n_i$ – число партий, содержащих $x_i$ нестандартных изделий).

Таблица 16

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

Найти методом моментов точечную оценку неизвестного параметра $\lambda$ распределения Пуассона.

12 Случайная величина $X$ (число появлений события $A$ в $m$ независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром $p$. В таблице 17 приведено эмпирическое распределение числа появлений события в 10 опытах по 5 испытаний в каждом (в первой строке указано число $x_i$ появлений события $A$ в одном опыте; во второй строке указана частота

$n_i$ – количество опытов, в которых наблюдалось $x_i$ появлений события $A$).

Таблица 17

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $n_i$ | 5 | 2 | 1 | 1 | 1 |

Найти методом моментов точечную оценку неизвестного параметра $p$ биномиального распределения.

13 Случайная величина $X$ (время работы элемента) имеет показательное распределение $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$. В таблице 18 приведено эмпирическое распределение среднего времени работы $n = 200$ элементов (в первой строке приведено среднее время $x_i$ работы элемента в часах; во второй строке указана частота $n_i$ – количество элементов, проработавших в среднем $x_i$ часов).

Таблица 18

| $x_i$ | 2,5 | 7,5 | 12,5 | 17,5 | 22,5 | 27,5 |
|---|---|---|---|---|---|---|
| $n_i$ | 133 | 45 | 15 | 4 | 2 | 1 |

Найти методом моментов точечную оценку неизвестного параметра показательного распределения.

14 Случайная величина $X$ (отклонение контролируемого размера изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами $a$ и $\sigma$. В таблице 19 приведено эмпирическое распределение отклонения от номинала $n = 200$ изделий (в первой строке указано отклонение $x_i$ (мм); во второй строке приведена частота $n_i$ – количество изделий, имеющих отклонение $x_i$).

Таблица 19

| $x_i$ | 0,3 | 0,5 | 0,7 | 0,9 | 1,1 | 1,3 | 1,5 | 1,7 | 1,9 | 2,2 | 2,3 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 6 | 9 | 26 | 25 | 30 | 26 | 21 | 24 | 20 | 8 | 5 |

Найти методом моментов точечные оценки неизвестных параметров $a$ и $\sigma$ нормального распределения.

15 Случайная величина $X$ (ошибка измерения дальности радиодальномером) подчинена равномерному закону распределения с неизвестными параметрами $a$ и $b$. В таблице 20 приведено эмпирическое распределение средней ошибки $n = 200$ измерений дальности (в первой строке указана средняя ошибка $x_i$; во второй строке указана частота $n_i$ – количество измерений, имеющих среднюю ошибку $x_i$).

Таблица 20

| $x_i$ | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n_i$ | 21 | 16 | 15 | 26 | 22 | 14 | 21 | 22 | 18 | 25 |

Найти методом моментов точечные оценки неизвестных параметров $a$ и $b$ равномерного распределения.

16 Случайная величина $X$ (число появления события $A$ в $m$ независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром $p$. В таблице 21 приведено эмпирическое распределение числа появлений события $A$ в 1000 испытаний (в первой строке указано число $x_i$ появлений события в одном опыте из $m = 10$ испытаний; во второй строке приведена частота $n_i$ – число опытов, в которых наблюдалось $x_i$ появлений события $A$) (см. таблицу 13).

Таблица 21

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 2 | 3 | 10 | 22 | 26 | 20 | 12 | 5 |

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра $p$ биномиального распределения.

17 Случайная величина $X$ (число поврежденных стеклянных изделий в одном контейнере) распределена по закону Пуассона с неизвестным параметром $\lambda$. В таблице 22 приведено эмпирическое распределение числа поврежденных изделий в 500 контейнерах (в первой строке указано количество $x_i$ поврежденных изделий в одном контейнере; во второй строке приведена частота $n_i$ – число контейнеров, содержащих $x_i$ поврежденных изделий).

Таблица 22

| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 199 | 169 | 87 | 31 | 9 | 3 | 1 | 1 |

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра $\lambda$ распределения Пуассона.

18 Случайная величина $X$ (время безотказной работы элемента) имеет показательное распределение $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$. В таблице 23 приведено эмпирическое распределение среднего времени работы 1000 элементов (в первой строке указано среднее время $x_i$ безотказной работы одного элемента в часах; во второй строке указана частота $n_i$ – количество элементов, проработавших в среднем $x_i$ часов).

Таблица 23

| $x_i$ | 5 | 15 | 25 | 35 | 45 | 55 | 65 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 365 | 245 | 150 | 100 | 70 | 45 | 25 |

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра $\lambda$ показательного распределения.

***Ответы***: 1) 5,75; 2) 1269; 109; 3) 3,075; 4) 0,0007; 5) 166; 33,44; 6) 6,93; 7) 168,88; 8) 0,0525; 9) 4,89; 10) 0,9; 11) 0,5; 12) 0,22; 13) 0,2; 14) 1,26; 0,5; 15) 2,24; 22,38; 16) 0,4; 17) 1; 18) 0,05.

***Контрольные вопросы***

1 Дайте определение понятия «точечная оценка».

2 Какая оценка называется несмещенной? Какая оценка называется смещенной?

3 Что является несмещенной оценкой генеральной средней?

4 Что является смещенной оценкой генеральной дисперсии?

5 Что является несмещенной оценкой генеральной дисперсии?

6 Что такое «условная варианта»?

7 В чем состоит метод моментов?

8 В чем состоит метод наибольшего правдоподобия?

9 Дайте определение понятия «функция правдоподобия», «логарифмическая функция правдоподобия».

## 3 Лабораторная работа № 3. Интервальное оценивание

***Цель работы***: практическое ознакомление с методами интервального оценивания параметров генеральной совокупности по выборочным данным.

***Теоретические сведения***

При статистической обработке результатов наблюдений часто необходимо не только найти оценку $\Theta^*$ неизвестного параметра $\Theta$, но и охарактеризовать точность этой оценки. С этой целью вводится понятие доверительного интервала.

*Доверительным интервалом* для параметра $\Theta$ называется интервал $(\Theta_1, \Theta_2)$, содержащий (накрывающий) истинное значение $\Theta$ с заданной вероятностью $p = 1 - \alpha$, т. е.

$$P(\Theta_1 < \Theta < \Theta_2) = 1 - \alpha.$$

Число $1 - \alpha$ называется *доверительной вероятностью*, а значение $\alpha$ –

*уровнем значимости*. Статистики $\Theta_1 = \Theta_1(x_1,...,x_n)$ и $\Theta_2 = \Theta_2(x_1,...,x_n)$, определяемые по выборке $x_1$, …, $x_n$ из генеральной совокупности с неизвестным параметром $\Theta$, называются соответственно *нижней и верхней границами доверительного интервала*. Условие $P(\Theta_1 < \Theta < \Theta_2) = 1-\alpha$ означает, что в большой серии независимых экспериментов, в каждом из которых получена выборка объёма $n$, в среднем $(1-\alpha)100$ % из общего числа построенных доверительных интервалов содержат истинное значение параметра $\Theta$.

Длина доверительного интервала, характеризующая точность интервального оценивания, зависит от объёма выборки $n$ и доверительной вероятности $1-\alpha$: при увеличении объёма выборки длина доверительного интервала уменьшается, а с приближением доверительной вероятности к единице – увеличивается. Выбор доверительной вероятности определяется конкретными условиями. Обычно используются значения $1-\alpha$, равные 0,90; 0,95; 0,99.

При решении некоторых задач применяются односторонние доверительные интервалы, границы которых определяются из условий

$$P(\Theta < \Theta_2) = 1-\alpha,\ P(\Theta_1 < \Theta) = 1-\alpha.$$

Эти интервалы называются соответственно *левосторонними и правосторонними доверительными интервалами*.

Чтобы найти доверительный интервал для параметра $\Theta$, необходимо знать закон распределения статистики $\Theta^* = \Theta^*(x_1,...,x_n)$, значение которой является оценкой параметра $\Theta$. При этом для получения доверительного интервала наименьшей длины при данном объёме выборки $n$ и заданной доверительной вероятности $1-\alpha$ в качестве оценки $\Theta^*$ параметра $\Theta$ следует брать эффективную либо асимптотически эффективную оценку.

Один из методов построения доверительных интервалов состоит в следующем. Предположим, что существует статистика $Y = Y(\Theta^*,\Theta)$ такая, что:

а) закон распределения $Y$ известен и не зависит от $\Theta$;

б) функция $Y(\Theta^*,\Theta)$ непрерывна и строго монотонна по $\Theta$.

Пусть, далее, $(1-\alpha)$ – заданная доверительная вероятность, а $y_{\frac{\alpha}{2}}$ и $y_{1-\frac{\alpha}{2}}$ – квантили распределения статистики $Y$ порядков $\dfrac{\alpha}{2}$ и $1-\dfrac{\alpha}{2}$ соответственно. Тогда с вероятностью $1-\alpha$ выполняется неравенство

$$y_{\frac{\alpha}{2}} < Y(\Theta^*,\Theta) < y_{1-\frac{\alpha}{2}}.$$

Решая данное неравенство относительно $\Theta$, найдём границы $\Theta_1$ и $\Theta_2$ доверительного интервала для $\Theta$. Если плотность распределения статистики $Y$ симметрична относительно оси $Oy$, то доверительный интервал имеет наименьшую длину, а если это распределение несимметрично, то длину, близкую к наименьшей.

***Пример 1*** – Рассмотрим выборку $x_1$, …, $x_n$ из нормально распределённой генеральной совокупности и найдём доверительный интервал для математического ожидания $m$ при условии, что дисперсия генеральной совокупности известна и равна $\sigma^2$, а доверительная вероятность равна $1-\alpha$.

*Решение*

В качестве оценки математического ожидания $m$ возьмём выборочное среднее $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Для нормально распределённой генеральной совокупности выборочное среднее является эффективной оценкой $m$. Выборочное среднее $\overline{X}$ в данном случае имеет нормальное распределение $N\left(m, \sigma/\sqrt{n}\right)$.

Рассмотрим статистику $U = \frac{\overline{X}-m}{\sigma/\sqrt{n}}$, имеющую нормальное распределение $N(0,1)$ независимо от значения параметра $m$. Кроме того, $U$ как функция $m$ непрерывна и строго монотонна. Следовательно,

$$P\left(u_{\frac{\alpha}{2}} < U < u_{1-\frac{\alpha}{2}}\right) = 1-\alpha,$$

где $u_{\frac{\alpha}{2}}$ и $u_{1-\frac{\alpha}{2}}$ – квантили нормального распределения $N(0,1)$.

Решая неравенство

$$u_{\frac{\alpha}{2}} < \frac{\overline{X}-m}{\sigma/\sqrt{n}} < u_{1-\frac{\alpha}{2}}$$

относительно $m$, получим, что с вероятностью $1-\alpha$ выполняется следующее условие:

$$\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} < m < \overline{X} - \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}.$$

Так как квантили нормального распределения связаны соотношением $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$, полученный доверительный интервал для $m$ можно записать следующим образом:

$$\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} < m < \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}.$$

Если распределение генеральной совокупности не является нормальным, то в отдельных случаях по выборкам большого объёма можно построить доверительные интервалы для неизвестных параметров приближённо, используя при этом предельные теоремы теории вероятностей и вытекающие из них асимптотические распределения и оценки.

***Пример 2*** – Пусть в $n$ независимых испытаниях схемы Бернулли успех наступил $x$ раз. Найти доверительный интервал для вероятности успеха $p$ в одном испытании.

*Решение*

Эффективной оценкой вероятности успеха $p$ в одном испытании является относительная частота $p^* = h = \dfrac{x}{n}$. По теореме Муавра–Лапласа относительная частота $p^*$ имеет асимптотически нормальное распределение $N\left(p, \sqrt{\dfrac{pq}{n}}\right)$, где $q = 1 - p$.

Рассмотрим статистику $U = \dfrac{h-p}{\sqrt{pq/n}}$, которая, следовательно, имеет асимптотически нормальное распределение $N(0,1)$ независимо от значения $p$. При больших $n$ тогда имеем

$$P\left(\left|\frac{h-p}{\sqrt{pq/n}}\right| < u_{1-\frac{\alpha}{2}}\right) \approx 1-\alpha .$$

Отсюда получаем, что с вероятностью $\approx 1-\alpha$ выполняется неравенство

$$h - u_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < p < h + u_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} .$$

Заменяя значения $p$ и $q$ в обеих частях последнего неравенства их оценками $p^* = h$, $q^* = 1-h$, получаем, что доверительный интервал для вероятности успеха в схеме Бернулли приближённо имеет вид:

$$h - u_{1-\frac{\alpha}{2}}\sqrt{\frac{h(1-h)}{n}} < p < h + u_{1-\frac{\alpha}{2}}\sqrt{\frac{h(1-h)}{n}} .$$

***Пример 3*** – При проверке 100 деталей из большой партии обнаружено 10 бракованных деталей.

Найти 95-процентный приближённый доверительный интервал для доли бракованных деталей во всей партии.

Какой минимальный объём выборки следует взять для того, чтобы с вероятностью 0,95 можно было утверждать, что доля бракованных деталей во всей партии отличается от частоты появления бракованных деталей в выборке не более чем на 1 %?

*Решение*

Оценка доли бракованных деталей в партии по выборке равна $p^* = h = 0{,}1$. По таблице находим квантиль $u_{1-\frac{\alpha}{2}} = u_{0{,}975} = 1{,}96$. Тогда 95-процентный доверительный интервал для доли бракованных деталей в партии приближённо имеет вид $0{,}041 < p < 0{,}159$.

Представим доверительный интервал в виде неравенства

$$|h - p| < u_{1-\frac{\alpha}{2}}\sqrt{\frac{h(1-h)}{n}},$$

которое выполняется с вероятностью $\approx 1-\alpha = 0{,}95$.

Так как по условию задачи $|h - p| \le 0{,}01$, то для определения $n$ получим неравенство

$$u_{0{,}975}\sqrt{\frac{h(1-h)}{n}} \le 0{,}01.$$

Отсюда следует, что

$$1{,}96\sqrt{\frac{0{,}1(1-0{,}1)}{n}} \le 0{,}01$$

и $n \ge (0{,}3\cdot 196)^2 = 3457{,}44$. Значит, минимальный объём выборки $n = 3458$.

***Задачи для самостоятельного решения***

1 Показать, что если дисперсия генеральной совокупности $\sigma^2$ неизвестна, а в качестве оценки дисперсии используется статистика $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, то при доверительной вероятности $1-\alpha$ доверительный интервал для математического ожидания имеет вид:

$$\overline{x} - \frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1) < m < \overline{x} + \frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1),$$

где $t_{1-\frac{\alpha}{2}}(n-1)$ – квантиль распределения Стюдента с $n-1$ степенью свободы.

2 Выборка объёма $n$ получена из генеральной совокупности, имеющей нормальное распределение $N(m,\sigma)$. Выборочные оценки определялись по результатам $n$ наблюдений. Используя эти данные, а также результаты примера 1 и задачи 1, найти 90-процентный и 99-процентный доверительные интервалы для математического ожидания (среднего) следующих характеристик:

а) ёмкость конденсатора, если $\overline{x} = 20$ мкФ, $n = 16$, среднеквадратичное отклонение известно и равно 4 мкФ;

б) время безотказной работы электронной лампы, если $\overline{x} = 500$ ч, $n = 100$, среднеквадратичное отклонение известно и равно 10 ч;

в) диаметр вала, если $\overline{x} = 30$ мм, $n = 9$, $s^2 = 9$ мм$^2$;

г) содержание углерода в единице продукта, если $\overline{x} = 18$ г, $n = 25$, $s^2 = 16$ г$^2$.

3 Методом моделирования получить 10 выборок объёма 25 из генеральной совокупности, имеющей нормальное распределение $N(5,1)$. Для каждой выборки найти доверительный интервал для математического ожидания $m$, считая, что дисперсия генеральной совокупности известна и равна единице. Доверительную вероятность принять равной 0,9. Какая часть из полученных интервалов накроет параметр $m = 5$?

4 Решить предыдущую задачу, считая, что дисперсия $\sigma^2$ генеральной совокупности неизвестна.

5 Решить задачи 3 и 4 при доверительной вероятности 0,99.

6 Показать, что если $m$ известно, а оценка дисперсии

$$\sigma^{*2} = s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2,$$

то доверительный интервал для дисперсии при доверительной вероятности $1-\alpha$ имеет вид:

$$\frac{ns_0^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} < \sigma^2 < \frac{ns_0^2}{\chi^2_{\frac{\alpha}{2}}(n)},$$

где $\chi_p^2(n)$ – квантиль распределения $\chi^2$ с $n$ степенями свободы.

7 Показать, что если $m$ неизвестно, $m^* = \overline{x}$, а $\sigma^{*2} = s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$,

то доверительный интервал для дисперсии при доверительной вероятности $1-\alpha$ имеет вид:

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}.$$

8 Выборка объёма $n$ получена из генеральной совокупности, имеющей нормальное распределение $N(m,\sigma)$. Выборочные оценки определялись по результатам $n$ наблюдений. Используя эти данные, а также результаты задач 6 и 7, найти доверительные интервалы для дисперсии следующих характеристик:

а) диаметр вала, если $\bar{x}$ = 29 мм, $n=16$, $s^2=4{,}5$ мм$^2$; найти 90-процентный и 95-процентный доверительные интервалы;

б) содержание углерода в единице продукта, если $\bar{x}$ = 18,8 г, $n=9$, $s^2$ = 20 г$^2$; найти 90-процентный и 99-процентный доверительные интервалы.

9 С автоматической линии, производящей подшипники, было отобрано 400 шт., причём 10 оказалось бракованными. Найти 90-процентный доверительный интервал для вероятности появления бракованного подшипника. Сколько подшипников надо проверить, чтобы с вероятностью 0,9973 можно было утверждать, что вероятность появления бракованного подшипника не отличается от частоты более чем на 5 %?

10 Для проверки утверждения о том, что вероятность $p$ отказа прибора равна 0,01, было проведено испытание 100 приборов, при этом один из приборов отказал. Построить 95-процентную верхнюю границу одностороннего доверительного интервала для $p$ по этим данным.

## 4 Лабораторная работа № 4. Проверка статистических гипотез

***Цель работы***: практическое ознакомление с методами проверки статистических гипотез.

***Теоретические сведения***

*Статистической гипотезой* называется любое предположение о генеральной совокупности, проверяемое по выборке.

Статистические гипотезы делятся на параметрические – гипотезы о параметрах распределения известного вида – и непараметрические – гипотезы о виде неизвестного распределения.

*Нулевой* (*основной*) называют выдвинутую гипотезу $H_0$. *Конкурирующей* (*альтернативной*) называют гипотезу $H_1$, которая противоречит нулевой.

*Простой* называют гипотезу, содержащую только одно предположение. *Сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

При проверке гипотезы могут быть допущены ошибки первого и второго рода. *Ошибка первого рода* состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность допустить ошибку первого рода называется *уровнем значимости* и обозначается $\alpha$, т. е. $\alpha = P(H_1 | H_0)$. Часто уровень значимости принимается равным 0,05, 0,01, 0,001. *Ошибка второго рода* состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность допустить ошибку второго рода равна $\beta$, т. е. $\beta = P(H_0 | H_1)$, тогда $1-\beta = P(H_1 | H_1)$. Величина $1-\beta$ называется *мощностью критерия.*

Проверка статистических гипотез осуществляется на основе данных выборки. Для этого используют специальным образом подобранную случайную величину (выборочную статистику). Её обозначают $U$ или $Z$ – если случайная величина распределена по нормальному закону; $F$ – если она имеет распределение Фишера–Снедекора; $T$ – если она имеет распределение Стьюдента; $\chi^2$ – если она имеет распределение «хи-квадрат» Пирсона. В общем случае будем обозначать эту случайную величину $K$.

*Статистическим критерием* называют случайную величину $K$, которая служит для проверки нулевой гипотезы. *Наблюдаемым значением* $K_{\text{набл}}$ называют значение критерия, вычисленное по выборкам.

После выбора определённого критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другое – при которых она принимается. *Критической областью* называют совокупность значений критерия, при которых нулевую гипотезу отвергают. *Областью принятия гипотезы* называют совокупность значений критерия, при которых гипотезу принимают. Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

*Критическими точками* $k_{\text{кр}}$ называют точки, отделяющие критическую область от области принятия гипотезы. *Правосторонней* называют критическую область, определяемую неравенством $K > k_{\text{кр}}$, где $k_{\text{кр}} > 0$. *Левосторонней* называют критическую область, определяемую неравенством $K < k_{\text{кр}}$, где $k_{\text{кр}} < 0$. Это *односторонние* области. *Двусторонней* называют критическую область, определяемую неравенствами $K < k_1$, $K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенством $|K| > k_{\text{кр}}$, где $k_{\text{кр}} > 0$.

*Мощность критерия* – вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Таким образом, если уровень значимости выбран, то критическую область следует строить так, чтобы мощность критерия была максимальной. Выполнение этого требования должно обеспечить минимальную ошибку второго рода.

Этапы проверки гипотезы.

1 Формулируют нулевую $H_0$ и альтернативную $H_1$ гипотезы.

2 Выбирают критерий $K = K(X_1, X_2, ..., X_n)$.

3 По выборке $x_1, x_2, ..., x_n$ вычисляют наблюдаемое значение критерия $K_{\text{набл}} = K(x_1, x_2, ..., x_n)$.

4 Выделяют критическую область и область принятия гипотезы.

Для отыскания критической области задают уровень значимости $\alpha$ и ищут критические точки исходя из следующих соотношений:

– для правосторонней критической области $P(K > k_{\text{кр}}) = \alpha$, где $k_{\text{кр}} > 0$;

– для левосторонней критической области $P(K < k_{\text{кр}}) = \alpha$, где $k_{\text{кр}} < 0$;

– для двусторонней области $P(K < k_1) + P(K > k_2) = \alpha$, где $k_1 < k_2$;

– для двусторонней симметричной области

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

5 Если наблюдаемое значение критерия принадлежит области принятия гипотезы, то нет оснований отвергнуть нулевую гипотезу; если же оно принадлежит критической области, то нулевую гипотезу отвергают в пользу альтернативной.

Рассмотрим вначале проверку параметрических гипотез.

1 *Проверка гипотезы о математическом ожидании нормально распределённой случайной величины.* Рассмотрим генеральную совокупность $X$ нормального распределения с параметрами $m$ и $\sigma$, т. е. $X \in N(m;\sigma)$. Нужно проверить нулевую гипотезу о том, что $m = m_0$ ($H_0 : m = m_0$). Можно сформулировать две основные модели и построить для них соответствующие критерии значимости.

*Модель 1. Среднее квадратическое отклонение* $\sigma$ *известно.* Пусть генеральная совокупность имеет нормальное распределение: $X \in N(m;\sigma)$, причём $\sigma$ известно. На основании случайной выборки $(x_1, \ldots, x_n)$ из этой генеральной совокупности требуется проверить нулевую гипотезу $H_0 : m = m_0$ при заданном уровне значимости $\alpha$.

Для проверки значимости отличия выборочной средней $\bar{x}_B$ в случайной выборке объёмом $n$ от математического ожидания $m_0$ генеральной совокупности выбирается критерий

$$U = \frac{(\bar{X} - m_0)\sqrt{n}}{\sigma}.$$

$U \in N(0;1)$ – нормированная нормально распределённая случайная величина.

Критическую точку $u_{\text{кр}}$ находят по таблице значений функции Лапласа.

По выборке вычисляют наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{(\overline{x}_B - m_0)\sqrt{n}}{\sigma}.$$

Критическая область строится в зависимости от вида альтернативной гипотезы.

*Правило 1.* Альтернативная гипотеза $H_1 : m > m_0$.

Критическая область правосторонняя $(u_{\text{кр}};+\infty)$.

Определяем $u_{кр}$ из условия $\Phi(u_{\text{кр}}) = \dfrac{1-2\alpha}{2}$.

Если $U_{\text{набл}} < u_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $U_{\text{набл}} > u_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 2.* Альтернативная гипотеза $H_1 : m < m_0$.

Критическая область левосторонняя $(-\infty;-u_{\text{кр}})$.

Определяем $u_{\text{кр}}$ из условия $\Phi(u_{\text{кр}}) = \dfrac{1-2\alpha}{2}$.

Если $U_{\text{набл}} > -u_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $U_{\text{набл}} < -u_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 3.* Альтернативная гипотеза $H_1 : m \neq m_0$.

Критическая область двусторонняя $(-\infty;-u_{\text{кр}}) \cup (u_{\text{кр}};+\infty)$.

Определяем $u_{\text{кр}}$ из условия $\Phi(u_{\text{кр}}) = \dfrac{1-\alpha}{2}$.

Если $|U_{\text{набл}}| < u_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $|U_{\text{набл}}| > u_{\text{кр}}$, то нулевую гипотезу отвергают.

*Модель 2. Среднее квадратическое отклонение $\sigma$ неизвестно.* Пусть генеральная совокупность имеет нормальное распределение: $X \in N(m;\sigma)$, причём $\sigma$ неизвестно. На основании случайной выборки $(x_1,\ldots,x_n)$ из этой генеральной совокупности требуется проверить нулевую гипотезу $H_0 : m = m_0$ при заданном уровне значимости $\alpha$.

Для проверки нулевой гипотезы применяется случайная величина

$$T = \frac{(\bar{X} - m_0)\sqrt{n}}{S}.$$

Величина $T$ имеет распределение Стьюдента с $k = n-1$ степенями свободы.

Вычисляется наблюдаемое значение критерия $T_{\text{набл}} = \dfrac{(\overline{x} - m_0)\sqrt{n}}{s}$.

Критическая область строится в зависимости от конкурирующей гипотезы.

*Правило 1.* Альтернативная гипотеза $H_1 : m > m_0$.

Критическая область правосторонняя $(t_{\text{кр}}; +\infty)$.

По таблице критических точек распределения Стьюдента находят критическую точку $t_{\text{кр}} = t_{\alpha,k}$, где $k = n-1$; $\alpha$ – заданный уровень значимости.

Если $T_{\text{набл}} < t_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $T_{\text{набл}} > t_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 2.* Альтернативная гипотеза $H_1 : m < m_0$.

Критическая область левосторонняя $(-\infty; -t_{\text{кр}})$.

$t_{\text{кр}}$ определяется, как и в случае правосторонней критической области.

Если $T_{\text{набл}} > -t_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $T_{\text{набл}} < -t_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 3.* Альтернативная гипотеза $H_1 : m \neq m_0$.

Критическая область двусторонняя $(-\infty; -t_{\text{кр}}) \cup (t_{\text{кр}}; +\infty)$.

По таблице критических точек распределения Стьюдента находят критическую точку $t_{\text{кр}} = t_{\frac{\alpha}{2},k}$.

Если $|T_{\text{набл}}| < t_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $|T_{\text{набл}}| > t_{\text{кр}}$, то нулевую гипотезу отвергают.

*2 Проверка гипотезы о дисперсии случайной величины, распределённой по нормальному закону.* Дисперсия характеризует такие важные технологические и конструкторские показатели, как точность машин, погрешность показаний контрольно-измерительных приборов, ритмичность производства, устойчивость работы автоматических линий и др.

Пусть случайная величина $X$ распределена по нормальному закону. Генеральная дисперсия неизвестна, но есть основания по теоретическим предположениям или по предыдущим опытам считать её равной $\sigma_0^2$. Из генеральной совокупности производится выборка объёмом $n$ и вычисляется исправленная выборочная дисперсия $s^2$. Требуется по заданному уровню значимости $\alpha$ проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$.

В качестве критерия оценки нулевой гипотезы используют случайную величину

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

которая имеет распределение «хи-квадрат» Пирсона с $k = n-1$ степенями свободы.

По выборке вычисляют наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{(n-1)s^2}{\sigma_0^2}.$$

Критическая область строится в зависимости от альтернативной гипотезы.

*Правило 1.* Альтернативная гипотеза $H_1 : \sigma^2 > \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{\text{кр}}\right) = \alpha$. По таблице критических точек распределения $\chi^2$, по заданному уровню значимости $\alpha$ и числу степеней свободы $k = n-1$ находят критическую точку $\chi^2_{\text{кр}} = \chi^2(\alpha, k)$.

Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 2.* Альтернативная гипотеза $H_1 : \sigma^2 < \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{\text{кр}}\right) = 1-\alpha$. По таблице критических точек распределения $\chi^2$, по заданному уровню значимости $\alpha$ и числу степеней свободы $k = n-1$ находят критическую точку $\chi^2_{\text{кр}} = \chi^2(1-\alpha, k)$.

Если $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, то нулевую гипотезу отвергают.

*Правило 3.* Альтернативная гипотеза $H_1 : \sigma^2 \neq \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{\text{пр. кр}}\right) = \dfrac{\alpha}{2}$. Левостороннюю критическую область строят из требования $P\left(\chi^2 > \chi^2_{\text{лев. кр}}\right) = 1 - \dfrac{\alpha}{2}$. По таблице критических точек распределения $\chi^2$ находят правостороннюю и левостороннюю критические точки $\chi^2_{\text{пр. кр}} = \chi^2\left(\dfrac{\alpha}{2}, k\right)$, $\chi^2_{\text{лев. кр}} = \chi^2\left(1 - \dfrac{\alpha}{2}, k\right)$.

Если $\chi^2_{\text{лев. кр}} < \chi^2_{\text{набл}} < \chi^2_{\text{пр. кр}}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

*Замечание.* Если число степеней свободы $k > 30$, то критическую точку $\chi^2(\alpha, k)$ можно найти из *равенства Уилсона–Гильферти*

$$\chi^2(\alpha, k) = k\left(1 - \frac{2}{9k} + z_\alpha \cdot \sqrt{\frac{2}{9k}}\right)^3,$$

где $z_\alpha$ находят, используя функцию Лапласа, из условия $\Phi(z_\alpha) = \dfrac{1-2\alpha}{2}$.

*3 Проверка гипотез о равенстве математических ожиданий двух случайных величин, распределённых по нормальному закону.* Пусть исследуются две случайные величины $X$ и $Y$, которые имеют нормальное распределение, т. е. $X \in N(m_X, \sigma_X)$, $Y \in N(m_Y, \sigma_Y)$, причём $\sigma_X$ и $\sigma_Y$ известны, а $m_X$ и $m_Y$ неизвестны. Производятся две независимые выборки объёмами $n_1$ и $n_2$ $(n_1 > 30, n_2 > 30)$, взятые из двух независимых генеральных совокупностей, и по ним находятся выборочные средние $\bar{x}_B$ и $\bar{y}_B$. Выдвигается нулевая гипотеза $m_X = m_Y$.

Для проверки нулевой гипотезы используется нормированная нормальная случайная величина

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{D(X)}{n_1} + \dfrac{D(Y)}{n_2}}}.$$

$Z \in N(0;1)$ – нормированная нормально распределённая случайная величина.

Критическую точку $z_{кр}$ находят по таблице функции Лапласа.

По выборке вычисляют наблюдаемое значение критерия

$$Z_{\text{набл}} = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\dfrac{D(X)}{n_1} + \dfrac{D(Y)}{n_2}}}.$$

Критическая область строится в зависимости от вида альтернативной гипотезы.

*Правило 1.* Альтернативная гипотеза $H_1 : m_X > m_Y$.

Критическая область правосторонняя.

Определяем $z_{\text{пр. кр}}$ из условия $\Phi(z_{\text{кр}}) = \dfrac{1 - 2\alpha}{2}$.

*Правило 2.* Альтернативная гипотеза $H_1 : m_X < m_Y$.

Критическая область левосторонняя, $z_{\text{лев. кр}} = -z_{\text{пр. кр}}$.

*Правило 3.* Альтернативная гипотеза $H_1 : m_X \neq m_Y$.

Критическая область двусторонняя.

Определяем $z_{\text{пр.кр}}$ из условия $\Phi(z_{\text{кр}}) = \dfrac{1 - \alpha}{2}$, $z_{\text{лев. кр}} = -z_{\text{пр. кр}}$.

Если средние квадратические отклонения неизвестны, но можно предположить, что они равны, т. е. $\sigma_1 = \sigma_2$, то наблюдаемое значение критерия находят по формуле

$$T_{\text{набл}} = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}},$$

где $s_X^2$ и $s_Y^2$ – исправленные выборочные дисперсии. По таблице критических точек распределения Стьюдента, по заданному уровню $\alpha$ находят критическую точку в зависимости от вида альтернативной гипотезы: $t_{\text{пр.кр}} = t(\alpha, k)$ – для правосторонней критической области; $t_{\text{лев.кр}} = -t(\alpha, k)$ – для левосторонней критической области; $t_{\text{пр.кр}} = t\left(\dfrac{\alpha}{2}, k\right)$ – для двусторонней критической области, где число степеней свободы $k = n_1 + n_2 - 2$.

*Замечание.* Если равенство $\sigma_1 = \sigma_2$ не может быть обосновано из общих соображений, то оно подлежит предварительной проверке с помощью критерия Фишера–Снедекора.

*4 Проверка гипотез о равенстве дисперсий двух случайных величин, распределённых по нормальному закону.* На практике задача сравнения дисперсий возникает, если требуется сравнить точность приборов, инструментов, самих методов измерений и т. д. Очевидно, предпочтительнее тот прибор, инструмент и метод, который обеспечивает наименьшее рассеяние результатов измерений, т. е. наименьшую дисперсию.

Пусть генеральные совокупности $X$ и $Y$ распределены нормально. По независимым выборкам с объёмами, соответственно равными $n_1$ и $n_2$, извлечённым из этих совокупностей, найдены исправленные выборочные дисперсии $s_X^2$ и $s_Y^2$. Требуется по исправленным дисперсиям при заданном уровне значимости $\alpha$ проверить нулевую гипотезу $H_0 : D(X) = D(Y)$.

В качестве критерия проверки нулевой гипотезы принимают случайную величину $F = \dfrac{S_{\text{б}}^2}{S_{\text{м}}^2}$. Величина $F$ имеет *распределение Фишера–Снедекора* со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где $n_1$ – объём выборки, по которой вычислена большая дисперсия $S_{\text{б}}^2$; $n_2$ – объём выборки, по которой вычислена меньшая дисперсия $S_{\text{м}}^2$.

Наблюдаемое значение критерия находят по формуле

$$F_{\text{набл}} = \frac{s_{\text{б}}^2}{s_{\text{м}}^2}.$$

Критическая область строится в зависимости от вида конкурирующей гипотезы.

*Правило 1.* Альтернативная гипотеза $H_1 : \mathrm{D}(X) > D(Y)$.

Критическая область правосторонняя, строится исходя из требования $P\left(F > F_{\text{кр}}(\alpha, k_1, k_2)\right) = \alpha$.

Критическую точку $F_{\text{кр}} = F(\alpha, k_1, k_2)$ находят по таблице критических точек распределения Фишера–Снедекора.

Если $F_{\text{набл}} < F_{\text{кр}}$, нет оснований отвергать нулевую гипотезу.

*Правило 2.* Альтернативная гипотеза $H_1 : \mathrm{D}(X) \neq D(Y)$.

Критическая область двусторонняя. Правостороннюю критическую точку $F_{\text{кр}} = F\left(\dfrac{\alpha}{2}, k_1, k_2\right)$ находят по таблице критических точек распределения Фишера–Снедекора.

Если $F_{\text{набл}} < F_{\text{кр}}$, нет оснований отвергать нулевую гипотезу.

Рассмотрим теперь проверку непараметрических гипотез и критерии согласия. Пусть выборка $X_1, \ldots, X_n$ произведена из генеральной совокупности с неизвестной теоретической функцией распределения, относительно которой имеются две непараметрические гипотезы: простая основная $H_0$: $F(x) = F_0(x)$ и сложная конкурирующая $H_1$: $F(x) \neq F_0(x)$, где $F_0(x)$ – известная функция распределения. Иными словами, мы хотим проверить, согласуются эмпирические данные с нашим гипотетическим предположением относительно теоретической функции распределения или нет. Поэтому критерии для проверки гипотез $H_0$ и $H_1$ носят название *критериев согласия*. Приведём три наиболее часто употребляемых критерия согласия.

*Критерий Колмогорова.* Эмпирическая функция распределения $F^*(x)$ представляет собой состоятельную оценку теоретической функции распределения $F(x)$. Поэтому можно сравнить эмпирическую функцию распределения $F^*(x)$ с гипотетической $F_0(x)$ и если мера расхождения между ними мала, то считать справедливой гипотезу $H_0$. Наиболее естественной и простой из таких мер (будем предполагать, что $F_0(x)$ – непрерывная функция) является равномерное расстояние $\rho_u = \sup\limits_{-\infty < x < +\infty} \left| F^*(x) - F_0(x) \right|$. Однако при построении критерия Колмогорова более удобно пользоваться нормированнным расстоянием $\rho = \sqrt{n}\rho_u$.

Итак, рассмотрим статистику

$$\rho = \rho(X_1, \ldots, X_n) = \sqrt{n} \sup_{-\infty < x < +\infty} \left| F^*(x) - F_0(x) \right|.$$

Критерий Колмогорова предписывает принять гипотезу $H_0$, если $\rho < C$, и отвергнуть – в противном случае, где $C$ – критическое значение критерия.

Если гипотеза $H_0$ справедлива, то распределение статистики $\rho$ не зависит от гипотетической функции распределения $F_0(x)$. Поэтому можно рассчитать таблицы, которые по заданному объёму выборки $n$ и критическому значению $C$ позволяют определить уровень значимости критерия $\alpha$.

При $n \to \infty$ распределение статистики $\rho$ сходится к распределению Колмогорова, и критическое значение $C$ при большом объёме выборки практически совпадает с $(1-\alpha)$-квантилью $k_{1-\alpha}$ распределения Колмогорова.

При практической реализации критерия Колмогорова сначала по выборке $X_1, \ldots, X_n$ составляют вариационный ряд $X_1^*, \ldots, X_n^*$. Затем находят $F_0(X_i^*)$ и определяют значения статистики $\rho$ по формуле

$$\rho = \sqrt{n} \max_{1 \le i \le n} \left( \left| F_0(X_i^*) - \frac{2i-1}{2n} \right| + \frac{1}{2n} \right).$$

Наконец, сравнивают полученное значение $\rho$ с критическим значением $C$ для заданного уровня значимости $\alpha$ и принимают или отвергают гипотезу $H_0$.

*Критерий* $\omega^2$. Пусть $F_1(x)$ – некоторая функция распределения, не совпадающая с $F_0(x)$. Критерий Колмогорова хорошо разделяет выборки (имеет большую мощность) из генеральных совокупностей с теоретическими функциями распределения $F_0(x)$ и $F_1(x)$, если $|F_0(x)-F_1(x)|$ достаточно велико хотя бы на малом интервале изменения $x$. Встречается и обратная ситуация, когда $|F_0(x)-F_1(x)|$ мало, но постоянно на достаточно большом интервале изменения $x$. В этом случае для разделения гипотез $H_0$ и $H_1$ естественно пользоваться каким-либо интегральным расстоянием.

Статистика $\omega^2$ критерия $\omega^2$ задаётся выражением

$$\omega^2 = \omega^2(X_1,...,X_n) = n\int_{-\infty}^{+\infty}\left(F^*(x)-F_0(x)\right)^2 f_0(x)dx,$$

где $f_0(x)$ – плотность гипотетического распределения вероятностей. Предполагается, что критическая область $W_k$ состоит из всех тех точек $(x_1,...,x_n)$, для которых $\omega^2 > C$, где $C$ – критическое значение критерия. Используя вариационный ряд $X_1^*$, …, $X_n^*$, статистику $\omega^2$ можно записать в более удобном для практических расчётов виде:

$$\omega^2 = \sum_{i=1}^{n}\left(F_0(X_i^*)-\frac{2i-1}{2n}\right)^2 + \frac{1}{12n}.$$

Распределение статистики $\omega^2$ при условии справедливости гипотезы $H_0$ также не зависит от гипотетической функции распределения $F_0(x)$ и при увеличении объёма выборки сходится к $\omega^2$-распределению. Если задан уровень значимости $\alpha$ критерия, то критическое значение $C$ практически совпадает с $(1-\alpha)$-квантилью $a_{1-\alpha}$ $\omega^2$-распределения.

Практическая реализация критерия $\omega^2$ происходит в той же последовательности, что и критерия Колмогорова: сначала по выборке $X_1$, …, $X_n$ определяется вариационный ряд $X_1^*$, …, $X_n^*$, затем находятся $F_0(X_i^*)$ и вычисляется значение статистики $\omega^2$ и, наконец, полученное значение $\omega^2$ сравнивается с критическим значением $C$ и либо принимается, либо отвергается гипотеза $H_0$.

В литературе иногда критериями $\omega^2$ называют целое семейство критериев, основанных на интегральных расстояниях с различными весовыми функциями.

*Критерий* $\chi^2$. Критерий $\chi^2$ является аналогом критерия $\omega^2$ для дискретной наблюдаемой величины $X$, хотя и применяется как в дискретном, так и в непрерывном случае.

Начнём с дискретного случая. Пусть наблюдаемая случайная величина $X$ может принимать только $k$ значений с неизвестными вероятностями $\theta_1, \ldots, \theta_k$. Основная гипотеза $H_0$ выделяет среди всех распределений случайных величин, принимающих данные $k$ значений, одно фиксированное распределение, для которого значения вероятностей $\theta_i$ известны и равны $p_i$. Обозначим через $n_i$ число тех элементов выборки $X_1, \ldots, X_n$, которые приняли значение $x_i$. Поскольку, в силу закона больших чисел надлюденная частота $\theta_i^* = n_i/n$ с ростом объёма выборки $n$ стремится к вероятности $\theta_i$, мы должны признать гипотезу $H_0$ справедливой, если все $\theta_i^*$ мало отличаются от $p_i$. Введём статистику

$$\chi^2 = \chi^2(X_1, \ldots, X_n) = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}.$$

С одной стороны, эта статистика является мерой равномерной близости всех $\theta_i^*$ к $p_i$, с другой стороны, она асимптотически при $n \to \infty$ независимо от гипотетических вероятностей $p_i$ имеет $\chi^2$-распределение с $k-1$ степенями свободы. Таким образом, критерий $\chi^2$ предписывает принять гипотезу $H_0$, если $\chi^2 < C$, и отвергнуть, если $\chi^2 \geq C$, где $C$ – критическое значение критерия.

Если задан уровень значимости $\alpha$, то критическое значение $C$ примерно совпадает с $(1-\alpha)$-квантилью $h_{1-\alpha}$ $\chi^2$-распределения.

При практической реализации критерия $\chi^2$ нужно следить за тем, чтобы объём выборки был велик, иначе неправомочна аппроксимация $\chi^2$-распределением распределения статистики $\chi^2$. Обычно считается, что достаточным условием для этого является выполнение неравенства $n_i \geq 5$ для всех $i$; в противном случае маловероятные значения объединяются в одно или присоединяются к другим значениям, причём объединённому значению приписывается суммарная вероятность и, разумеется, уменьшается число степеней свободы при определении критического значения $C$.

В общем случае, не обязательно дискретной наблюдаемой величины $X$, поступают следующим образом. Сначала всю прямую разбивают на $k$ непересекающихся интервалов. Затем определяют гипотетические вероятности $p_i$ попадания значений величины в каждый из интервалов и числа $n_i$ элементов выборки, попавших в эти интервалы. Наконец, вычисляют значение статистики

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$$

и сравнивают его с критическим значением $C$. Как и в дискретном случае, маловероятные интервалы объединяют. Для того чтобы увеличить мощность критерия $\chi^2$, необходимо уменьшать интервалы разбиения, однако этому препятствует ограничение на числа попавших в каждый интервал наблюдений.

Часто требуется проверить не совпадение теоретической функции распределения $F(x)$ с известной функцией распределения $F_0(x)$, а принадлежность $F(x)$ заданному параметрическому семейству $F(x;\theta_1,...,\theta_m)$ функций распределения, зависящему от $m$ неизвестных параметров. Для того чтобы воспользоваться вышеописанными критериями, нужно из семейства $F(x;\theta_1,...,\theta_m)$ выделить ту функцию распределения $F_0(x)=F(x;\theta_{10},...,\theta_{m0})$, с которой уже и будет производиться сравнение эмпирической функции распределения выборки. Поэтому сначала, предполагая, что верна основная гипотеза $H_0$, находят оценки $\theta_1^*$, …, $\theta_m^*$ неизвестных параметров, а затем, полагая $\theta_{10}=\theta_1^*$, …, $\theta_{m0}=\theta_m^*$, с помощью выбранного критерия согласия проверяют простую основную гипотезу $H_0$: $F(x)=F_0(x)=F\left(x;\theta_1^*,...,\theta_m^*\right)$ против сложной конкурирующей гипотезы $H_1$: $F(x)\neq F_0(x)=F\left(x;\theta_1^*,...,\theta_m^*\right)$.

В общем случае, уровень значимости критерия согласия будет зависеть от выбранной оценки. Обычно на практике для критериев Колмогорова и $\omega^2$ считают уровень значимости таким же, как и в случае простой гипотезы $H_0$. Что касается критерия $\chi^2$, то для него при определении уровня значимости уменьшают число степеней свободы $\chi^2$-распределения на число $m$ неизвестных параметров.

***Задачи для самостоятельного решения***

1 По техническим условиям средняя прочность троса составляет 2000 кг. В результате испытаний 20 кусков троса было установлено, что средняя прочность на разрыв равна 1955 кг при средней ошибке 25 кг. Удовлетворяет ли образец троса техническим условиям?

2 По утверждению руководства фирмы, средний размер дебиторского счета равен 187,5 тыс. р. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборки равна 175 тыс. р. при среднем квадратическом отклонении 35 тыс. р. Может ли оказаться в действительности правильным объявленный размер дебиторского счета? Принять уровень значимости равным 0,05.

3 Хронометраж затрат времени на сборку узла машины 25 слесарями показал, что среднее время сборки узла $\overline{x}_B=16$ мин, а исправленное среднее квадратическое отклонение $s=4,5$. В предположении о нормальном распределении решить вопрос о том, можно ли на уровне значимости $\alpha=0,05$ считать 15 мин нормативом.

4 По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем, причём выборочное среднее расхода топлива на 100 км пробега по результатам испытаний составило 9,3 л. Предполагая, что выборка расходов топлива

получена из нормально распределённой генеральной совокупности со средним $m$ и дисперсией 4 л$^2$, используя критерий значимости, проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива.

5 Большая партия изделий может содержать некоторую долю дефектных. Поставщик утверждает, что эта доля составляет 5 %; покупатель предполагает, что доля дефектных изделий равна 10 %. Условия поставки: из партии случайным образом выбирается и проверяется 10 изделий; партия принимается на условиях поставщика, если при проверке обнаружено не более одного дефектного изделия; в противном случае партия принимается на условиях покупателя. Сформулировать эту задачу в терминах теории проверки статистических гипотез и ответить на следующие вопросы:

а) Каковы статистика критерия, область её значений, критическая область?

б) Какое распределение имеет статистика критерия?

в) В чём состоят проверяемая и альтернативная гипотезы?

г) В чём состоят ошибки первого и второго рода и каковы их вероятности?

6 Методом моделирования получить выборку объёма 100 из генеральной совокупности, имеющей нормальное распределение (параметры выбрать произвольно). Проверить гипотезу $H_0$ о том, что генеральная совокупность распределена по нормальному закону с помощью:

а) критерия Колмогорова;

б) критерия $\omega^2$;

в) критерия $\chi^2$.

Проверку производить для уровня значимости $\alpha = 0,05$.

7 Методом моделирования получить выборку объёма 100 из генеральной совокупности, имеющей распределение Пуассона. С помощью критерия $\chi^2$ проверить гипотезу о том, что генеральная совокупность распределена по закону Пуассона. Проверку производить для уровня значимости $\alpha = 0,05$.

***Ответы***: 1) $U_{\text{набл}} = -8,05$, $u_{\text{кр}} = 1,96$; образец троса не удовлетворяет техническим условиям; 2) $T_{\text{набл}} = -1,129$, $t_{\text{кр}}(0,05;9) = 2,26$; нет оснований считать неправильным объявленный размер дебиторского счета; 3) $T_{\text{набл}} = 1,11$, $t_{\text{кр}} = t(0,25;24) = 2,064$; нет оснований не считать 15 мин нормативом; 4) при $\alpha = 0,05$ критическая область определяется неравенством $\overline{X} < 9,342$; изменение конструкции двигателя привело к уменьшению расхода топлива; 5) а) число дефектных изделий; $\{0,1,...,10\}$; $\{2,3,...,10\}$; б) биномиальное $B(10,p)$; в) $H_0$: $p = 0,05$, верно утверждение поставщика; $H_1$: $p = 0,10$, верно утверждение покупателя; г) ошибка первого рода: партия принята на условиях покупателя, в то время как верно утверждение поставщика; ошибка второго рода: партия принята на условиях поставщика, в то время как верно утверждение покупателя; $\alpha \approx 0,086$, $\beta \approx 0,736$.

# Список литературы

1 **Бочаров, П. П.** Теория вероятностей. Математическая статистика : учебное пособие / П. П. Бочаров, А. В. Печинкин. – Москва: Гардарика, 1998. – 328 с.

2 **Гмурман, В. Е.** Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие для вузов / В. Е. Гмурман. – 11-е изд., перераб. и доп. – Москва : Юрайт, 2020. – 406 с.

3 **Гмурман, В. Е.** Теория вероятностей и математическая статистика : учебник для вузов / В. Е. Гмурман. – 12-е изд. – Москва : Юрайт, 2020. – 479 с.

4 Математическая статистика: статистическая проверка статистических гипотез, элементы теории корреляции : методические рекомендации к практическим занятиям для студентов всех специальностей и направлений подготовки дневной и заочной форм обучения / Сост. Т. Ю. Орлова, С. Ф. Плешкунова, Д. В. Роголев. – Могилев : Белорус.-Рос. ун-т, 2018. – 46 с.

5 **Письменный, Д. Т.** Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д. Т. Письменный. – 7-е изд. – Москва : Айрис-пресс, 2015. – 287 с.

6 Сборник задач по математике для втузов: учебное пособие: в 4 ч. / Под общ. ред. А. В. Ефимова, А. С. Поспелова. – 3-е изд., перераб. и доп. – Москва : Физ.-мат. лит., 2003. – Ч. 4. – 432 с.