Министерство образования Республики Беларусь
Учреждение образования «Витебский государственный
университет имени П.М. Машерова»
Кафедра геометрии и математического анализа

С.М. Бородич, Т.В. Кавитова

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

Витебск
ВГУ имени П.М. Машерова
2020

УДК 519.22(075.8)
ББК 22.172я73
Б83

Печатается по решению научно-методического совета учреждения образования «Витебский государственный университет имени П.М. Машерова». Протокол № 6 от 18.06.2020.

Авторы: старшие преподаватели кафедры геометрии и математического анализа ВГУ имени П.М. Машерова **С.М. Бородич, Т.В. Кавитова**

Р е ц е н з е н т :
заведующий кафедрой математики
и информационных технологий УО «ВГТУ»,
кандидат физико-математических наук *Т.В. Никонова*

Научный редактор:
доцент кафедры геометрии и математического анализа ВГУ
имени П.М. Машерова, кандидат физико-математических наук *Т.Л. Сурин*

Бородич, С.М.

Б83 Математическая статистика : методические рекомендации / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2020. – 48 с.

Настоящее учебное издание содержит теоретический материал, примеры решения задач и задачи для самостоятельного решения. Оно может использоваться для организации самостоятельной работы студентов, а также при проведении практических занятий.

Предназначается для студентов, обучающихся по специальностям «Программное обеспечение информационных технологий», «Прикладная математика», «Прикладная информатика», «Математика и информатика», «Компьютерная безопасность (радиофизические методы и программно-технические средства)», «Управление информационными ресурсами».

УДК 519.22(075.8)
ББК 22.172я73

Содержание

Введение ………………………………………………………………… 4

1. Выборка. Статистическое распределение выборки и ее графическое представление ……………………………………………………… 5

2. Эмпирическая функция распределения ……………………………… 8

3. Числовые характеристики выборки ………………………………… 9

4. Точечные оценки неизвестных параметров распределения ……… 10

5. Методы нахождения оценок неизвестных параметров распределения …………………………………………………………………… 14

6. Распределения χ^2 (хи-квадрат), Стьюдента и Фишера ………… 17

7. Интервальное оценивание. Доверительные интервалы для параметров нормального распределения ………………………………… 20

8. Проверка параметрических гипотез ………………………………… 21

9. Проверка гипотез о законе распределения ………………………… 26

10. Статистическое исследование зависимости между случайными величинами. Линейная регрессия ……………………………………… 29

11. Выборочное корреляционное отношение ………………………… 33

Задачи для самостоятельного решения ……………………………… 37

Литература ……………………………………………………………… 47

Введение

Математическая статистика, являющаяся частью прикладной математической дисциплины «Теория вероятностей и математическая статистика», изучает случайные явления, используя одинаковые с теорией вероятностей методы и понятия. Математическая статистика отличается от теории вероятностей специфичностью решаемых ею задач. Если в теории вероятностей мы считаем заданной модель явления и производим расчёт возможного реального течения этого явления, то в математической статистике на основе реализаций каких-либо случайных событий (т. е. на основе экспериментальных данных) подбирается подходящая теоретико-вероятностная модель.

Учебное издание организовано следующим образом: в пунктах 1–3 рассматриваются основные понятия описательной статистики, применяемые для первичного анализа экспериментальных данных (выборка, способы её представления, эмпирическая функция распределения, основные числовые характеристики выборки); в пунктах 4–7 излагаются основы теории точечных и интервальных оценок неизвестных параметров распределений; в пунктах 8, 9 рассмотрены некоторые важные задачи проверки статистических гипотез; пункты 10, 11 посвящены вопросам статистического исследования зависимости между случайными величинами.

По каждой теме приводится краткий теоретический материал. Кроме того, издание содержит разобранные примеры, позволяющие закрепить, углубить теоретические знания и получить навыки практического использования методов математической статистики. В конце издания приведен список задач, который полностью охватывает все изучаемые темы и вполне обеспечивает необходимую самостоятельную работу студентов.

Настоящее учебное издание адресуется прежде всего студентам факультета математики и информационных технологий ВГУ имени П.М. Машерова, обучающимся по специальностям «Программное обеспечение информационных технологий», «Прикладная математика», «Прикладная информатика», «Математика и информатика», «Компьютерная безопасность (радиофизические методы и программно-технические средства)», «Управление информационными ресурсами».

1. Выборка. Статистическое распределение выборки и ее графическое представление

Будем предполагать, что каждый исход некоторого случайного эксперимента характеризуется одним из возможных значений случайной величины *X*. Повторив *n* раз эксперимент в одинаковых условиях, получим последовательность из *n* наблюдённых значений этой случайной величины: $x_1,\ x_2,\ \ldots,\ x_n$. Набор чисел $x_1,\ x_2,\ \ldots,\ x_n$ называется *выборкой*.

Числа $x_i\ (i=1,\ 2,\ \ldots,\ n)$ называются *элементами выборки*, а их количество *n* — *объёмом выборки*.

Выборку можно упорядочить, расположив её элементы в неубывающем порядке:

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}.$$

Полученная таким образом последовательность чисел

$$x_{(1)},\ x_{(2)},\ \ldots,\ x_{(n)}$$

называется *вариационным рядом.*

Различные значения случайной величины *X*, представленные в выборке, называются *вариантами*.

Если в выборке объема *n* варианта x_i встретилась n_i раз, то число n_i называется *частотой*, а число $w_i = \dfrac{n_i}{n}$ – *относительной частотой (частостью)* этой варианты. Очевидно, что

$$\sum_{i=1}^{n} n_i = n, \qquad \sum_{i=1}^{n} w_i = 1.$$

Статистическим рядом распределения выборки называют перечень вариант и соответствующих им частот или относительных частот. Обычно статистический ряд записывается в виде таблицы, первая строка которой содержит варианты x_i, а вторая — их частоты n_i (или относительные частоты w_i). При этом варианты располагаются в порядке возрастания.

Как правило, статистический ряд распределения составляется в случае, когда наблюдаемая случайная величина *X* является дискретной.

Графически статистический ряд распределения выборки изображается следующим образом. В прямоугольной декартовой системе координат строят точки с координатами $(x_i,\ n_i)$. Последовательно соединив их отрезками, получают ломаную линию, называемую *полигоном частот*. Аналогично строится *полигон относительных частот*.

Пример 1. Имеется выборка значений случайной величины *X* объёма $n = 15$:

$$0,\ \ 3,\ \ -5,\ \ -3,\ \ 1,\ \ 0,\ \ 1,\ \ 3,\ \ 0,\ \ 0,\ \ -3,\ \ 1,\ \ -1,\ \ 0,\ \ -1.$$

Вариационным рядом для неё будет последовательность

$$-5,\ \ -3,\ \ -3,\ \ -1,\ \ -1,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 0,\ \ 1,\ \ 1,\ \ 1,\ \ 3,\ \ 3.$$

Статистический ряд распределения данной выборки имеет вид

x_i	-5	-3	-1	0	1	3
n_i	1	2	2	5	3	2

или

x_i	-5	-3	-1	0	1	3
w_i	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{5}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Полигон частот изображён на рис. 1.

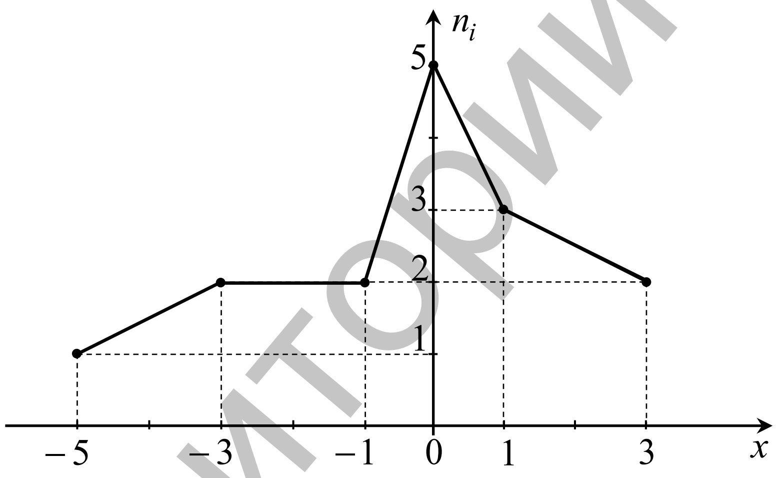

Рис.1

В случае, когда X является непрерывной случайной величиной или же вариационный ряд имеет большое количество вариант, элементы выборки объединяют в группы (разряды), представляя результаты опытов в виде *интервального (группированного) статистического ряда распределения*. Для этого интервал, содержащий все значения выборки, разбивают на k непересекающихся интервалов

$$[a_0,\, a_1),\ \ [a_1,\, a_2),\ \ \ldots,\ \ [a_{k-1},\, a_k].$$

Затем для каждого интервала разбиения определяют *частоту* – количество элементов выборки, попавших в этот интервал: n_i – частота, соответствующая i-му интервалу разбиения $(i = 1,\ 2,\ \ldots,\ k)$. Наряду с частотами находят также *относительные частоты*

$$w_i = \frac{n_i}{n}$$

(n – объём выборки). Интервальный статистический ряд записывают в виде таблицы, в первой строке которой содержатся интервалы группировки, а во второй – соответствующие частоты n_i или относительные частоты w_i.

Интервальный статистический ряд распределения изображают с помощью *гистограммы частот*. Для этого на оси абсцисс откладывают интервалы группировки и на них, как на основании, строят прямоугольники с высотами

$$h_i = \frac{n_i}{a_i - a_{i-1}} \quad (i = 1,\ 2,\ \ldots,\ k).$$

Аналогично строится *гистограмма относительных частот*.

Пример 2. Выборка 100 значений наблюдаемой непрерывной случайной величины X представлена в виде следующего интервального статистического ряда распределения:

Интервал	[22, 24)	[24, 26)	[26, 28)	[28, 30)	[30, 32)	[32, 34]
Частота	2	14	34	40	8	2

Построить гистограмму частот.

Решение. В данном случае $a_i - a_{i-1} = 2$ $(i = 1,\ 2,\ 3,\ 4,\ 5,\ 6)$. Находим высоты прямоугольников:

$$h_1 = \frac{2}{2} = 1, \quad h_2 = \frac{14}{2} = 7, \quad h_3 = \frac{34}{2} = 17,$$

$$h_4 = \frac{40}{2} = 20, \quad h_5 = \frac{8}{2} = 4, \quad h_6 = \frac{2}{2} = 1.$$

Гистограмма частот изображена на рис. 2.

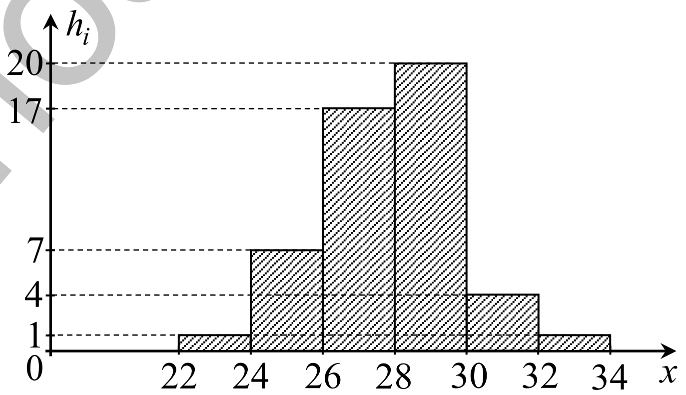

Рис. 2

•

2. Эмпирическая функция распределения

Эмпирической (выборочной) функцией распределения случайной величины *X*, построенной по выборке $x_1, x_2, \dots, x_n$, называют функцию $F_n^*(x)$, которая при каждом $x \in \mathbf{R}$ равна относительной частоте события $X < x$, т.е.

$$F_n^*(x) = \frac{n_x}{n},$$

где n_x – число элементов выборки, меньших x.

Эмпирическая функция распределения обладает следующими свойствами:

1. $0 \le F_n^*(x) \le 1$ для любого $x \in \mathbf{R}$;
2. $F_n^*(x)$ не убывает на $\mathbf{R}$;
3. $F_n^*(x) = 0$ при $x \le x_{\min}$, $F_n^*(x) = 1$ при $x > x_{\max}$, где $x_{\min}$ и $x_{\max}$ – соответственно наименьший и наибольший элементы выборки;
4. $F_n^*(x)$ непрерывна слева в любой точке $x_0 \in \mathbf{R}$.

График эмпирической функции распределения имеет ступенчатый вид. В промежутках между соседними вариантами выборки $F_n^*(x)$ сохраняет постоянное значение. В точках оси Ox, равных вариантам выборки, $F_n^*(x)$ претерпевает скачки; величина скачка в точке $x = x_i$ равна относительной частоте варианты x_i.

Пример 3. В условиях примера 1 найти эмпирическую функцию распределения случайной величины *X* и построить её график.

Решение. В данном случае $n = 15$. По статистическому ряду распределения находим

$$F_{15}^*(x) = \begin{cases} 0 & \text{при} \quad x \le -5, \\ \dfrac{1}{15} & \text{при} \quad -5 < x \le -3, \\ \dfrac{3}{15} & \text{при} \quad -3 < x \le -1, \\ \dfrac{5}{15} & \text{при} \quad -1 < x \le 0, \\ \dfrac{10}{15} & \text{при} \quad 0 < x \le 1, \\ \dfrac{13}{15} & \text{при} \quad 1 < x \le 3, \\ 1 & \text{при} \quad x > 3. \end{cases}$$

График $F_{15}^*(x)$ изображён на рис. 3.

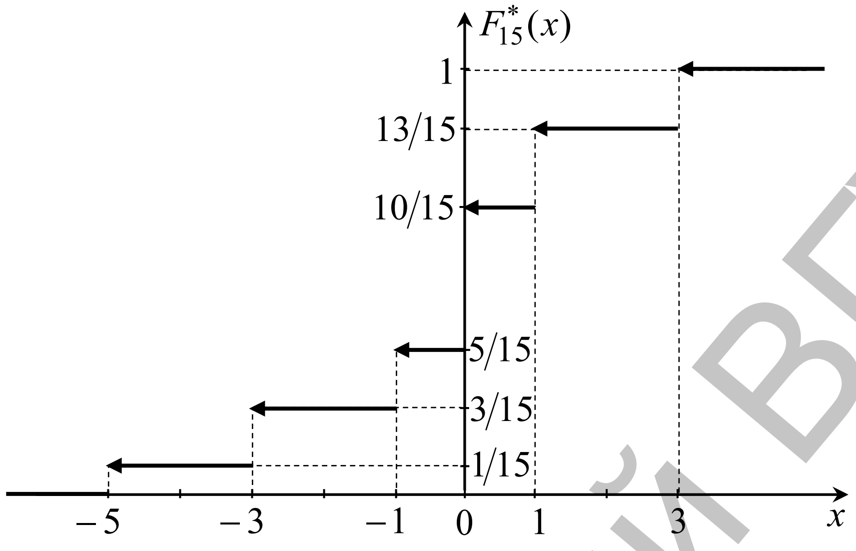

Рис. 3 ●

3. Числовые характеристики выборки

Пусть x_1, x_2, …, x_n – выборка объёма *n*.

Выборочным средним $\overline{x}$ называется среднее арифметическое всех значений выборки:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Выборочной дисперсией s^2 называют среднее арифметическое квадратов отклонений значений выборки от выборочного среднего $\overline{x}$:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 .$$

Выборочную дисперсию можно вычислять также по формуле

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 .$$

Замечание 1. Если выборка представлена статистическим рядом распределения, то значения $\overline{x}$ и s^2 находят по формулам

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i , \qquad (1)$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\overline{x})^2 , \qquad (2)$$

где x_i – варианты, n_i – соответствующие им частоты. ●

Замечание 2. Для интервального статистического ряда выборочное среднее и выборочную дисперсию также вычисляют по формулам (1) и (2).

В качестве x_i берут середины интервалов ряда, а в качестве n_i – частоты соответствующих интервалов. ●

Выборочное среднее квадратическое отклонение s определяется формулой

$$s = \sqrt{s^2}\,.$$

Модой M_o^* выборки называют варианту, имеющую наибольшую частоту.

Медианой M_e^* выборки называется число, которое делит вариационный ряд

$$x_{(1)}, x_{(2)}, \ldots, x_{(n)}$$

на две части, содержащие равное число элементов, а именно: если $n = 2k$ $(k \in \mathbf{N})$, то $M_e^* = \dfrac{x_{(k)} + x_{(k+1)}}{2}$; если $n = 2k+1$, то $M_e^* = x_{(k+1)}$.

4. Точечные оценки неизвестных параметров распределения

Пусть $F(x, \theta)$ – функция распределения случайной величины X, θ – неизвестный параметр. Предполагаем, что общий вид функции $F(x, \theta)$ задан.

Пусть в результате n наблюдений за случайной величиной X получена выборка

$$x_1, x_2, \ldots, x_n\,. \tag{3}$$

По выборке требуется найти приближённое значение параметра θ.

Элементы выборки (3) можно рассматривать последовательно как частные значения n независимых случайных величин

$$X_1,\ X_2,\ \ldots,\ X_n,$$

каждая из которых имеет тот же закон распределения, что и случайная величина X.

Любая функция случайных величин $X_1, X_2, \ldots, X_n$ называется *статистикой*.

Пусть $\tilde{\theta} = \tilde{\theta}(X_1, X_2, \ldots, X_n)$ – некоторая статистика. Значение $\tilde{\theta}(x_1, x_2, \ldots, x_n)$, принятое статистикой $\tilde{\theta}$ на выборке (3), называется её *выборочным значением*.

Важными примерами статистик являются *выборочное среднее*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и *выборочная (статистическая) дисперсия*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

Определённые в пункте 3 числовые характеристики выборки $\overline{x}$ (выборочное среднее) и s^2 (выборочная дисперсия) являются выборочными значениями одноимённых статистик $\overline{X}$ и S^2.

Статистика $\widetilde{\theta}$, выборочное значение которой принимается за приближённое значение неизвестного параметра θ, называется его *точечной оценкой* или просто *оценкой*.

Основными свойствами оценок, характеризующими их качество, являются *несмещённость*, *состоятельность* и *эффективность*.

Оценку $\widetilde{\theta}$ параметра θ называют *несмещённой*, если

$$M(\widetilde{\theta}) = \theta .$$

В противном случае оценка называется *смещённой*.

Разность $M(\widetilde{\theta}) - \theta$ называется *смещением*.

Несмещённость оценки гарантирует, что при оценивании неизвестного параметра не будет систематических ошибок в сторону завышения или занижения.

Оценку $\widetilde{\theta}_n = \widetilde{\theta}_n(X_1, X_2, \ldots, X_n)$ параметра θ называют *состоятельной*, если она сходится по вероятности к θ при $n \to \infty$, т. е. для любого $\varepsilon > 0$ выполняется соотношение

$$\lim_{n\to\infty} P(|\widetilde{\theta}_n - \theta| \geq \varepsilon) = 0 .$$

Свойство состоятельности означает, что с ростом объёма выборки выборочные значения оценки приближаются к неизвестному значению параметра.

Теорема 1. Выборочное среднее $\overline{X}$ является несмещённой и состоятельной оценкой математического ожидания наблюдаемой случайной величины *X*.

Доказательство. Поскольку случайные величины X_i распределены так же, как и сама наблюдаемая случайная величина *X*, то $M(X_i) = M(X)$, $i = 1, 2, \ldots, n$. В силу свойств математического ожидания имеем

$$M(\overline{X}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n} nM(X) = M(X),$$

и значит, $\overline{X}$ – несмещённая оценка $M(X)$.

Так как последовательность $X_1, X_2, \ldots, X_n$ состоит из независимых одинаково распределённых случайных величин с конечным математическим ожиданием и дисперсией, то в силу следствия теоремы Чебышева (см. [1], [2])

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} M(X).$$

Следовательно, $\overline{X}$ – состоятельная оценка $M(X)$. ◄

Теорема 2. Выборочная дисперсия S^2 является смещённой и состоятельной оценкой дисперсии случайной величины *X*.

Ограничимся лишь доказательством смещённости оценки S^2.

Доказательство. Преобразуем формулу выборочной дисперсии:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}[(X_i - M(X)) - (\overline{X} - M(X))]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - 2\cdot\frac{1}{n}(\overline{X} - M(X))\sum_{i=1}^{n}(X_i - M(X)) +$$

$$+\frac{1}{n}\cdot n(\overline{X} - M(X))^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - 2(\overline{X} - M(X))^2 + (\overline{X} - M(X))^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - M(X))^2 - (\overline{X} - M(X))^2.$$

Таким образом,

$$M(S^2) = \frac{1}{n}\sum_{i=1}^{n} M[(X_i - M(X))^2] - M[(\overline{X} - M(X))^2] =$$

$$= \frac{1}{n}\cdot nD(X) - D(\overline{X}) = D(X) - \frac{D(X)}{n} = \frac{n-1}{n}D(X).$$

Поскольку $M(S^2) \neq D(X)$, то S^2 – смещённая оценка дисперсии случайной величины *X*. ◄

Замечание 3. Из доказательства теоремы 2 следует, что статистика

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (4)$$

является несмещённой оценкой дисперсии σ^2. Действительно,

$$S_0^2 = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}S^2,$$

и значит,

$$M(S_0^2) = \frac{n}{n-1}M(S^2) = \frac{n}{n-1}\cdot\frac{n-1}{n}D(X) = D(X).$$

Можно доказать также, что S_0^2 является состоятельной оценкой $D(X)$. ●

Статистика S_0^2, определённая формулой (4), называется *исправленной выборочной дисперсией*.

Отметим, что при больших значениях *n* выборочные значения статистик S^2 и S_0^2 отличаются мало. Поэтому на практике оценку S_0^2 используют для оценки дисперсии в основном лишь при малых объёмах выборки (обычно при $n \le 30$).

Несмещённая оценка $\tilde{\theta}$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещённых оценок параметра θ.

Эффективность оценки означает, что статистика, используемая в качестве точечной оценки неизвестного параметра, обеспечивает минимальный среди всех несмещённых оценок разброс приближённых значений параметра около его истинного значения.

Пусть $p(x, \theta)$ – плотность распределения непрерывной случайной величины X или же вероятность $p(x, \theta) = P(X = x)$ – в случае, если случайная величина X дискретная. В предположении, что $p(x, \theta)$ и статистика $\tilde{\theta}$ удовлетворяют некоторым условиям регулярности (см. [11]), для дисперсии несмещенной оценки $\tilde{\theta}$ параметра θ выполняется *неравенство Рао – Крамера*:

$$D(\tilde{\theta}) \ge \frac{1}{J_n(\theta)},$$

где

$$J_n(\theta) = nM\left[\left(\frac{\partial \ln p(X,\theta)}{\partial \theta}\right)^2\right]$$

– *информация Фишера* относительно неизвестного параметра θ, содержащаяся в выборке объема n.

Таким образом, для доказательства эффективности несмещённой оценки $\tilde{\theta}$ (в случае выполнения упомянутых выше условий регулярности) достаточно установить равенство

$$D(\tilde{\theta}) = \frac{1}{J_n(\theta)}.$$

Пример 4. Пусть $x_1, x_2, \ldots, x_n$ – выборка из нормально распределенной генеральной совокупности с параметрами m и σ. Показать, что выборочное среднее $\overline{X}$ является эффективной оценкой параметра m.

Решение. Нам известно, что $\overline{X}$ – несмещенная оценка $M(X)$, и так как $M(X) = m$, то $\overline{X}$ – несмещенная оценка параметра m.

Находим

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Вычислим информацию Фишера $J_n(m)$. Имеем:

$$p(x,m)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}};$$

$$\ln p(x,m) = -\frac{(x-m)^2}{2\sigma^2}-\ln\sqrt{2\pi}\sigma,$$

$$\frac{\partial \ln p(x,m)}{\partial m} = \frac{x-m}{\sigma^2},$$

$$M\left[\left(\frac{\partial \ln p(X,m)}{\partial m}\right)^2\right] = M\left[\frac{(X-m)^2}{\sigma^4}\right] = \frac{1}{\sigma^4}M(X-M(X))^2 =$$

$$= \frac{D(X)}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2},$$

$$J_n(m) = nM\left[\left(\frac{\partial \ln p(X,m)}{\partial m}\right)^2\right] = \frac{n}{\sigma^2}.$$

Таким образом,

$$D(\overline{X}) = \frac{1}{J_n(m)},$$

и следовательно, $\overline{X}$ – эффективная оценка параметра m. ●

Замечание 4. Оценки S^2 и S_0^2 не являются эффективными оценками дисперсии. В случае, когда известно математическое ожидание m наблюдаемой случайной величины X, несмещённой, состоятельной и эффективной оценкой $D(X)$ является оценка

$$S_1^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2 . \text{ ●}$$

5. Методы нахождения оценок неизвестных параметров распределения

Рассмотрим два наиболее распространённых метода получения точечных оценок неизвестных параметров распределения: метод максимального правдоподобия и метод моментов.

Метод максимального правдоподобия. Предположим, что известен вид закона распределения случайной величины X, но неизвестен параметр θ, которым определяется этот закон (параметр θ может быть и векторным: $\theta=(\theta_1, \theta_2, \ldots, \theta_r)$). Пусть $x_1, x_2, \ldots, x_n$ – выборка наблюдений случайной величины X, по которой требуется оценить параметр θ.

Функцией правдоподобия для оценки параметра θ называется функция

$$L(\theta)=p(x_1,\theta)p(x_2,\theta)\ldots p(x_n,\theta),$$

где $p(x, \theta)$ – плотность распределения случайной величины X, если эта величина непрерывная, и $p(x, \theta) = P(X = x)$, если случайная величина X дискретная.

Метод максимального правдоподобия состоит в том, что в качестве оценки параметра θ принимается статистика $\tilde{\theta}$ (векторная статистика $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_r)$, если $\theta = (\theta_1, \ldots, \theta_r)$), выборочное значение которой $\tilde{\theta}(x_1, x_2, \ldots, x_n)$ доставляет максимум функции правдоподобия. Такую оценку называют *оценкой максимального правдоподобия* (коротко: *МП-оценкой*).

Для упрощения вычислений, связанных с получением МП-оценки, в некоторых случаях удобно использовать логарифмическую функцию правдоподобия $\ln L(\theta)$.

Метод моментов. Пусть X — непрерывная случайная величина, плотность распределения вероятностей которой $p(x, \theta_1, \theta_2, \ldots, \theta_r)$ содержит неизвестные параметры $\theta_1, \theta_2, \ldots, \theta_r$. Определим с помощью этой плотности r каких-либо теоретических моментов случайной величины X, например, первые r начальных моментов:

$$\alpha_k(\theta_1, \theta_2, \ldots, \theta_r) = M\left[X^k\right] = \int_{-\infty}^{+\infty} x^k p(x, \theta_1, \theta_2, \ldots, \theta_r)\,dx, \quad k = 1, 2, \ldots, r.$$

По выборке наблюдений случайной величины X найдем значения соответствующих выборочных моментов α_k^*:

$$\alpha_k^* = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \qquad k = 1, 2, \ldots, r.$$

Приравнивая теоретические моменты α_k их выборочным аналогам α_k^*, получаем систему r уравнений с неизвестными $\theta_1, \theta_2, \ldots, \theta_r$:

$$\alpha_k(\theta_1, \theta_2, \ldots, \theta_r) = \alpha_k^*, \qquad k = 1, 2, \ldots, r.$$

Решая полученную систему относительно неизвестных $\theta_1, \theta_2, \ldots, \theta_r$, находим оценки $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_r$ неизвестных параметров.

Аналогично находятся оценки неизвестных параметров по выборке наблюдений дискретной случайной величины.

Пример 5. Найти МП-оценки математического ожидания m и дисперсии σ^2 нормально распределенной генеральной совокупности.

Решение. Пусть $x_1, x_2, \ldots, x_n$ – выборка наблюдений случайной величины X с плотностью распределения

$$p(x, m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Найдем функцию правдоподобия $L(m, \sigma^2)$:

$$L(m,\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-m)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{\sum(x_i-m)^2}{2\sigma^2}}.$$

Логарифмическая функция правдоподобия имеет следующий вид:

$$\ln L(m,\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{\sum(x_i-m)^2}{2\sigma^2}.$$

Используя необходимые условия максимума $L(m,\sigma^2)$, получаем систему уравнений для нахождения оценок максимального правдоподобия:

$$\begin{cases} \dfrac{\partial \ln L(m,\sigma^2)}{\partial m} = \dfrac{1}{\sigma^2}\sum(x_i-m)=0, \\ \dfrac{\partial \ln L(m,\sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum(x_i-m)^2=0. \end{cases}$$

Из первого уравнения системы находим

$$m=\frac{1}{n}\sum x_i=\bar{x}.$$

Подставляя это значение во второе уравнение, получаем

$$\sigma^2=\frac{1}{n}\sum(x_i-\bar{x})^2=s^2.$$

Таким образом, $\tilde{m}=\overline{X}$, $\tilde{\sigma}^2 = S^2$ – искомые оценки. ●

Пример 6. Пусть случайная величина X распределена по показательному закону с плотностью распределения вероятностей

$$p(x,\lambda)=\begin{cases}\lambda e^{-\lambda x} & \text{при } x\geq 0, \\ 0 & \text{при } x<0,\end{cases}$$

содержащей неизвестный параметр λ. Используя метод моментов, по выборке $x_1, x_2,\ldots, x_n$ найти оценку параметра λ.

Решение. Вычисляем теоретический начальный момент первого порядка:

$$\alpha_1(\lambda) = \int_{-\infty}^{+\infty} xp(x,\lambda)dx = \lambda\int_{0}^{+\infty} xe^{-\lambda x}dx = \frac{1}{\lambda}.$$

Начальный выборочный момент первого порядка равен выборочному среднему:

$$\alpha_1^*=\bar{x}.$$

Приравниваем найденные теоретический и выборочный моменты:

$$\frac{1}{\lambda}=\bar{x}.$$

Отсюда получаем $\lambda=\frac{1}{\bar{x}}$. Следовательно, $\tilde{\lambda}=\frac{1}{\overline{X}}$ – искомая оценка. ●

6. Распределения χ^2 (хи-квадрат), Стьюдента и Фишера

Рассмотрим три широко распространённых в математической статистике распределения – χ^2, Стьюдента и Фишера. Приведём также используемые в дальнейшем теоремы о статистиках, имеющих эти распределения.

Распределение χ^2. Пусть случайные величины $X_1, X_2, \dots, X_n$ независимы и распределены нормально с параметрами $m=0$ и $\sigma=1$. *Распределением χ^2 (или χ^2-распределением) с n степенями свободы* называется закон распределения случайной величины

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2 .$$

Плотность распределения χ^2 с *n* степенями свободы имеет вид

$$p_{\chi^2}(x) = \begin{cases} 0 & \text{при } x \le 0, \\ \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{при } x > 0, \end{cases}$$

где $\Gamma(a) = \int\limits_0^{\infty} t^{a-1} e^{-t} dt$ – гамма-функция Эйлера.

С возрастанием числа степеней свободы *n* распределение χ^2 приближается к нормальному (при $n > 30$ распределение χ^2 практически не отличается от нормального).

Для $n \le 30$ составлена таблица, содержащая такие значения $\chi_n^2 = \chi_{\alpha,n}^2$, для которых

$$P(\chi_n^2 > \chi_{\alpha,n}^2) = \int\limits_{\chi_{\alpha,n}^2}^{+\infty} p_{\chi^2}(x) dx = \alpha$$

(α – заданный уровень вероятности). Значение $\chi_{\alpha,n}^2$ называется *квантилью распределения χ^2*, отвечающей заданному уровню вероятности α и заданному уровню степеней свободы *n*.

Для $n > 30$ и $\alpha \le 0{,}5$ применяют формулу

$$\chi_{\alpha,n}^2 = 0{,}5(u_\alpha + \sqrt{2n-1})^2,$$

где u_α (*квантиль уровня α нормального распределения с параметрами* $m=0$, $\sigma=1$) находится по таблице функции Лапласа из условия

$$\Phi(u_\alpha) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u_\alpha} e^{-\frac{t^2}{2}} dt = 1-\alpha .$$

Теорема 3. Пусть наблюдаемая случайная величина *X* распределена нормально с параметрами m и σ. Тогда статистика

$$\frac{(n-1)S_0^2}{\sigma^2}$$

имеет распределение χ^2 с $n-1$ степенями свободы. Здесь n – объём выборки, S_0^2 – исправленная выборочная дисперсия.

Распределение Стьюдента. Пусть $X_0, X_1, \ldots, X_n$ – независимые случайные величины, распределённые нормально с параметрами $m=0$ и $\sigma=1$. *Распределением Стьюдента с n степенями свободы* называется закон распределения случайной величины

$$T_n = \frac{X_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}.$$

Замечание 5. Очевидно, что случайную величину T_n можно записать в следующем виде:

$$T_n = \frac{X_0\sqrt{n}}{\sqrt{\chi_n^2}}. \bullet$$

Плотность распределения Стьюдента с n степенями свободы имеет вид

$$p_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbf{R}.$$

При $n \to \infty$ распределение Стьюдента приближается к нормальному с параметрами $m=0$, $\sigma=1$. Практически уже при $n>30$ распределение Стьюдента можно считать приближённо нормальным.

Составлена таблица *квантилей $t_{\alpha/2,n}$ распределения Стьюдента*, отвечающих числу степеней свободы n и заданному уровню вероятности α. Значение квантили $t_{\alpha/2,n}$ определяется из условия

$$P(|T_n| > t_{\alpha/2,n}) = 2\int_{t_{\alpha/2,n}}^{+\infty} p_T(x)dx = \alpha.$$

Теорема 4. Пусть наблюдаемая случайная величина X распределена нормально с параметрами m и σ. Тогда статистика

$$\frac{(\overline{X}-m)\sqrt{n}}{S_0}$$

имеет распределение Стьюдента с $n-1$ степенями свободы (n – объём выборки, $\overline{X}$ – выборочное среднее, $S_0 = \sqrt{S_0^2}$ – *исправленное выборочное среднее квадратическое отклонение*).

Теорема 5. Пусть $\overline{X}_1$, S_{01}^2 и $\overline{X}_2$, S_{02}^2 — выборочные средние и исправленные выборочные дисперсии в двух независимых выборках объёмов n_1 и n_2 из нормально распределённых генеральных совокупностей, математические ожидания которых известны и равны m_1 и m_2 соответственно, а неизвестные дисперсии одинаковы. Тогда случайная величина

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (m_1 - m_2)}{\sqrt{(n_1-1)S_{01}^2 + (n_2-1)S_{02}^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1+n_2-2)}{n_1+n_2}}$$

имеет распределение Стьюдента с $\nu = n_1 + n_2 - 2$ степенями свободы.

Распределение Фишера. Пусть случайные величины U^2 и V^2 имеют χ^2- распределения с n_1 и n_2 степенями свободы соответственно. *Распределением Фишера с* n_1 *и* n_2 *степенями свободы* называется закон распределения случайной величины

$$F_{n_1, n_2} = \frac{U^2}{V^2} \cdot \frac{n_2}{n_1}.$$

Плотность распределения Фишера с n_1 и n_2 степенями свободы имеет вид

$$p_F(x) = \begin{cases} 0 & \text{при } x \le 0, \\ \dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)} x^{\frac{n_1}{2}-1}\left(1 + \dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & \text{при } x > 0. \end{cases}$$

Для распределения Фишера составлена таблица, в которой для различных значений n_1, n_2 и α указаны значения квантилей F_{α, n_1, n_2}, определяемых из условия

$$P(F_{n_1, n_2} > F_{\alpha, n_1, n_2}) = \int_{F_{\alpha, n_1, n_2}}^{+\infty} p_F(x)dx = \alpha.$$

Теорема 6. Пусть S_{01}^2 и S_{02}^2 – исправленные выборочные дисперсии в двух независимых выборках объёмов n_1 и n_2, взятых из нормально распределенных генеральных совокупностей с одинаковыми дисперсиями. Тогда случайная величина

$$F = \frac{S_{01}^2}{S_{02}^2}$$

имеет распределение Фишера *с* $n_1 - 1$ *и* $n_2 - 1$ степенями свободы.

7. Интервальное оценивание. Доверительные интервалы для параметров нормального распределения

Пусть θ – неизвестный параметр распределения случайной величины X.

Пусть статистики $\tilde{\theta}_1 = \tilde{\theta}_1(X_1, \ldots, X_n)$ и $\tilde{\theta}_2 = \tilde{\theta}_2(X_1, \ldots, X_n)$ таковы, что интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ содержит (накрывает) истинное значение параметра θ с заданной вероятностью $p = 1 - \alpha$:

$$P(\tilde{\theta}_1 < \theta < \tilde{\theta}_2) = 1 - \alpha .$$

Тогда интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$ называется *доверительным интервалом* для параметра θ, вероятность $p = 1 - \alpha$ – *доверительной вероятностью (надёжностью)*, а число α – *уровнем значимости*.

Пусть X – нормально распределённая случайная величина. Тогда:

1. Если среднее квадратическое отклонение σ известно, то доверительный интервал для математического ожидания m имеет вид

$$\left(\overline{X} - u_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \overline{X} + u_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right),$$

где n – объём выборки, $\overline{X}$ – выборочное среднее, $u_{\alpha/2}$ находится по таблице значений функции Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

из условия

$$\Phi_0(u_{\alpha/2}) = \frac{1-\alpha}{2}.$$

2. Если среднее квадратическое отклонение σ неизвестно, то доверительный интервал для математического ожидания m имеет вид

$$\left(\overline{X} - t_{\alpha/2, n-1} \frac{S_0}{\sqrt{n}},\ \overline{X} + t_{\alpha/2, n-1} \frac{S_0}{\sqrt{n}} \right),$$

где

$$S_0 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

(исправленное выборочное среднее квадратическое отклонение), $t_{\alpha/2, n-1}$ находится по таблице квантилей распределения Стьюдента из условия

$$P\left(T_{n-1} > t_{\alpha/2, n-1}\right) = \alpha / 2$$

(здесь T_{n-1} – случайная величина, распределённая по закону Стьюдента с $n-1$ степенью свободы).

Пример 7. По результатам 6 независимых наблюдений над нормально распределённой случайной величиной X найдены выборочное среднее $\bar{x} = 5{,}63$ и исправленная выборочная дисперсия $s_0^2 = 0{,}0625$. Требуется найти доверительный интервал для математического ожидания случайной величины X, соответствующий доверительной вероятности $1-\alpha = 0{,}99$.

Решение. Находим $s_0 = \sqrt{s_0^2} = \sqrt{0{,}0625} = 0{,}25$. В нашем случае $n-1=5$, $\alpha/2 = 0{,}005$. По таблице квантилей распределения Стьюдента находим значение $t_{\alpha/2,\,n-1} = t_{0{,}005;5} = 4{,}03$. Вычисляем

$$t_{\alpha/2,\,n-1}\frac{s_0}{\sqrt{n}} = 4{,}03 \cdot \frac{0{,}25}{\sqrt{6}} \approx 0{,}41.$$

Искомый доверительный интервал таков: $(5{,}63-0{,}41;\ 5{,}63+0{,}41)$, т. е. $(5{,}22;\ 6{,}04)$. ●

Доверительный интервал для среднего квадратического отклонения σ нормально распределённой случайной величины X имеет вид

$$\left(S_0\sqrt{\frac{n-1}{\chi^2_{\alpha/2,\,n-1}}},\ S_0\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2,\,n-1}}} \right),$$

где $\chi^2_{1-\alpha/2,\,n-1}$ и $\chi^2_{\alpha/2,\,n-1}$ находятся по таблице квантилей χ^2-распределения из условий

$$P(\chi^2_{n-1} < \chi^2_{1-\alpha/2,\,n-1}) = \alpha/2, \quad P(\chi^2_{n-1} > \chi^2_{\alpha/2,\,n-1}) = \alpha/2$$

(здесь χ^2_{n-1} – случайная величина, имеющая χ^2-распределение с $n-1$ степенью свободы).

8. Проверка параметрических гипотез

Статистической гипотезой или просто *гипотезой* называется любое предположение о распределениях наблюдаемых в эксперименте случайных величин, проверяемое по результатам наблюдений.

Статистическая гипотеза называется *параметрической*, если в ней сформулировано предположение относительно значений параметров распределения известного вида.

Параметрическая гипотеза называется *простой*, если она однозначно определяет параметры функции распределения; в противном случае гипотеза называется *сложной*.

Относительно параметра (параметров) распределения формулируется некоторая *основная (нулевая)* гипотеза H_0. Наряду с гипотезой H_0 рассматривается одна из *альтернативных (конкурирующих)* гипотез H_1. Если основная гипотеза отклоняется, то принимается альтернативная.

Правило, которое позволяет по результатам соответствующих наблюдений принять или отклонить гипотезу H_0, называется *статистическим критерием* (или просто *критерием*) для проверки гипотезы H_0.

Проверка соответствия результатов эксперимента (выборочных данных) выдвинутой гипотезе H_0 осуществляется с помощью специально подобранной статистики Z, точное или приближённое распределение которой в случае справедливости гипотезы H_0 известно. Статистика Z называется *статистикой критерия*. По выборочному значению $z_в$ статистики критерия судят о справедливости гипотезы H_0.

Множество всех возможных значений статистики критерия разбивают на два непересекающихся подмножества: одно из них содержит значения статистики Z, при которых гипотеза H_0 отклоняется, другое – при которых она принимается.

Критической областью называется совокупность всех значений статистики критерия, при которых принимается решение отклонить гипотезу H_0.

Совокупность всех значений статистики критерия, при которых гипотезу H_0 принимают, называется *областью принятия гипотезы*.

Таким образом, если выборочное значение $z_в$ статистики критерия принадлежит критической области, то гипотезу H_0 отклоняют; если же $z_в$ принадлежит области принятия гипотезы, то гипотезу H_0 принимают.

Поскольку проверка статистической гипотезы осуществляется на основе выборочных данных, носящих случайный характер, то всегда присутствует возможность принятия неправильного решения. С каждым статистическим критерием связаны ошибки двух родов:

ошибка первого рода — отклонение гипотезы H_0, когда на самом деле она верна;

ошибка второго рода — принятие гипотезы H_0, когда эта гипотеза ошибочна.

Вероятность ошибки первого рода называется *уровнем значимости* критерия и обозначается α. Как правило, уровень значимости (вероятность отклонения верной гипотезы H_0) задаётся заранее; при этом для α обычно используются стандартные значения: $\alpha = 0{,}001;\ 0{,}005;\ 0{,}01;\ 0{,}05;\ 0{,}1$.

При заданном уровне значимости α критическую область выбирают такой, чтобы вероятность ошибки второго рода была минимальной.

Различают левостороннюю, правостороннюю и двустороннюю критические области. *Левосторонняя* критическая область определяется неравенством $Z < z_1$, *правосторонняя* — неравенством $Z > z_2$. *Двусторонняя* критическая область определяется совокупностью неравенств $Z < z_1$ и $Z > z_2$, где $z_1 < z_2$.

Точки z_1 и z_2, отделяющие критическую область от области принятия гипотезы, называются *критическими*.

Критические точки ищут исходя из требования, чтобы при условии справедливости гипотезы H_0 вероятность попадания значения статистики критерия в критическую область была равна принятому уровню значимости α:

$P(Z < z_1) = \alpha$ – для критерия с левосторонней критической областью;

$P(Z > z_2) = \alpha$ – для критерия с правосторонней критической областью;

$P(Z < z_1) + P(Z > z_1) = \alpha$ – для критерия с двусторонней критической областью.

При нахождении критических точек используется таблица распределения, которому подчиняется (точно или приближённо) статистика Z в предположении, что верна гипотеза H_0.

Замечание 6. Если распределение статистики критерия (при условии справедливости гипотезы H_0) симметрично относительно нуля, то для критерия с двусторонней критической областью критические точки z_1 и z_2 также выбираются симметричными относительно нуля: $z_1 = -z_2$. Таким образом, в данном случае критическая область задаётся неравенством $|Z| > z_2$, и значит, точка z_2 должна удовлетворять условию $P(|Z| > z_2) = \alpha$, которое (в силу симметричности относительно нуля распределения статистики Z) равносильно условию $P(Z > z_2) = \dfrac{\alpha}{2}$. ●

Рассмотрим следующий пример построения статистического критерия для проверки параметрической гипотезы о математическом ожидании нормально распределённой с.в.

Пусть наблюдаемая в эксперименте случайная величина X имеет нормальный закон распределения с неизвестными параметрами m и σ. По результатам выборки объёма n требуется при заданном уровне значимости α проверить гипотезу $H_0 : m = m_0$. Предполагается, что в качестве альтернативной гипотезы выбрана гипотеза $H_1 : m \neq m_0$.

Статистикой критерия выбирается случайная величина

$$T = \frac{(\overline{X} - m_0)\sqrt{n}}{S_0},$$

где S_0 – исправленное выборочное среднее квадратическое отклонение. Если гипотеза H_0 справедлива, то в силу теоремы 4 статистика T имеет распределение Стьюдента с $n-1$ степенями свободы. При данном выборе альтернативной гипотезы ($m \neq m_0$) строят двустороннюю критическую область. По таблице квантилей распределения Стьюдента по уровню значимости α и числу степеней свободы $n-1$ находят критическую точку $t_{\alpha/2,\,n-1}$, для которой

$$P\left(|T| > t_{\alpha/2,\,n-1}\right) = \alpha .$$

Неравенство $|T| > t_{\alpha/2,\,n-1}$ определяет критическую область критерия.

Далее вычисляется выборочное значение $T_в$ статистики критерия:

$$T_в = \frac{(\bar{x} - m_0)\sqrt{n}}{s_0},$$

где

$$s_0 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} .$$

Если $|T_в| > t_{\alpha/2,\,n-1}$, то гипотеза H_0 отклоняется и принимается гипотеза H_1. Если $|T_в| \leq t_{\alpha/2,\,n-1}$, то считается, что оснований для отклонения гипотезы H_0 нет.

Если в рассмотренном нами примере альтернативную гипотезу «$m \neq m_0$» заменить на гипотезу «$m < m_0$», то критическую область нужно строить левосторонней, а если на гипотезу «$m > m_0$», то правосторонней.

В таблицах 1–4 приводятся критерии проверки некоторых параметрических гипотез о параметрах нормально распределенных генеральных совокупностей.

Таблица 1. **Критерии для проверки гипотезы о математическом ожидании случайной величины, имеющей нормальное распределение**

Гипотеза H_0	Условие	Статистика Z критерия	Распределение Z, если H_0 верна	Гипотеза H_1	Критич. область
$m = m_0$	σ известно	$\frac{(\bar{X}-m_0)\sqrt{n}}{\sigma}$	Нормальное с параметрами (0, 1)	$m > m_0$	$Z > u_\alpha$
				$m < m_0$	$Z < -u_\alpha$
				$m \neq m_0$	$\lvert Z \rvert > u_{\alpha/2}$
	σ неизвестно	$\frac{(\bar{X}-m_0)\sqrt{n}}{S_0}$	Стьюдента с числом степеней свободы $\nu = n-1$	$m > m_0$	$Z > t_{\alpha,\,\nu}$
				$m < m_0$	$Z < -t_{\alpha,\,\nu}$
				$m \neq m_0$	$\lvert Z \rvert > t_{\alpha/2,\,\nu}$

Таблица 2. **Критерии для проверки гипотезы о равенстве математических ожиданий двух случайных величин, имеющих нормальное распределение**

Гипотеза H_0	Условия	Статистика Z критерия	Распределение Z, если H_0 верна	Гипотеза H_1	Критич. область
$m_1 = m_2$	σ_1 и σ_2 известны	$\dfrac{\bar{X}_1-\bar{X}_2}{\sqrt{\sigma_1{}^2/n_1+\sigma_2{}^2/n_2}}$	Нормальное с параметрами (0, 1)	$m_1 > m_2$	$Z > u_\alpha$
				$m_1 < m_2$	$Z < -u_\alpha$
				$m_1 \neq m_2$	$\lvert Z \rvert > u_{\alpha/2}$
	σ_1 и σ_2 неизвестны, $\sigma_1 = \sigma_2$	$\dfrac{\bar{X}_1-\bar{X}_2}{\sqrt{(n_1-1)S_{01}^2+(n_2-1)S_{02}^2}}\times \sqrt{\dfrac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}$	Стьюдента с числом степ. свободы $\nu = n_1 + n_2 - 2$	$m_1 > m_2$	$Z > t_{\alpha,\nu}$
				$m_1 < m_2$	$Z < -t_{\alpha,\nu}$
				$m_1 \neq m_2$	$\lvert Z \rvert > t_{\alpha/2,\nu}$
	σ_1 и σ_2 неизвестны, $\sigma_1 \neq \sigma_2$	$\dfrac{\bar{X}_1-\bar{X}_2}{\sqrt{S_{01}^2/n_1+S_{02}^2/n_2}}$	Стьюдента с числом степ. свободы $\nu \approx \dfrac{(S_{01}^2/n_1+S_{02}^2/n_2)^2}{\dfrac{(S_{01}^2/n_1)^2}{n_1-1}+\dfrac{(S_{02}^2/n_2)^2}{n_2-1}}$	$m_1 > m_2$	$Z > t_{\alpha,\nu}$
				$m_1 < m_2$	$Z < -t_{\alpha,\nu}$
				$m_1 \neq m_2$	$\lvert Z \rvert > t_{\alpha/2,\nu}$

(S_{01}^2 и S_{02}^2 – исправленные выборочные дисперсии)

Таблица 3. **Критерий для проверки гипотезы о дисперсии случайной величины, имеющей нормальное распределение**

Гипотеза H_0	Условие	Статистика Z критерия	Распределение Z, если H_0 верна	Гипотеза H_1	Критич. область
$\sigma^2 = \sigma_0{}^2$	m неизвестно	$\dfrac{(n-1)S_0^2}{\sigma_0^2}$	χ^2 с числом степ. свободы $\nu = n - 1$	$\sigma^2 > \sigma_0{}^2$	$Z > \chi^2_{\alpha,\nu}$
				$\sigma^2 < \sigma_0{}^2$	$Z < \chi^2_{1-\alpha,\nu}$
				$\sigma^2 \neq \sigma_0{}^2$	$Z > \chi^2_{\alpha/2,\nu}$ или $Z < \chi^2_{1-\alpha/2,\nu}$

Таблица 4. **Критерий для проверки гипотезы о равенстве дисперсий двух случайных величин, имеющих нормальное распределение**

Гипотеза H_0	Условие	Статистика Z критерия	Распределение Z, если H_0 верна	Гипотеза H_1	Критич. область
$\sigma_1{}^2 = \sigma_2{}^2$	m_1 и m_2 неизвестны	$\dfrac{S_{01}^2}{S_{02}^2}$	Фишера с $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$ степ. свободы	$\sigma_1{}^2 > \sigma_2{}^2$	$Z > F_{\alpha,\nu_1,\nu_2}$

Пример 8. Из нормально распределённой генеральной совокупности извлечена выборка объема n = 13 и по ней найдена исправленная выборочная дисперсия s_0^2=10,3. Требуется при уровне значимости 0,02, проверить нулевую гипотезу H_0: $\sigma^2 = 12$, приняв в качестве конкурирующей гипотезы H_1: $\sigma^2 \neq 12$.

Решение. Для проверки гипотезы H_0 используем статистику

$$Z = \frac{(n-1)S_0^2}{\sigma_0^2},$$

имеющую, при условии справедливости нулевой гипотезы, χ^2-распределение с $\nu = n - 1$ степенями свободы (см. таблицу 3).

Найдем выборочное значение статистики критерия

$$z_в = \frac{(13-1)\cdot 10{,}3}{12} = 10{,}3.$$

Так как конкурирующая гипотеза имеет вид $\sigma^2 \neq 12$, то критическая область — двусторонняя.

По таблице квантилей χ^2-распределения находим критические точки:

$$z_1 = \chi^2_{1-\alpha/2,\,\nu} = \chi^2_{0{,}99,\,12} = 3{,}6, \qquad z_2 = \chi^2_{\alpha/2,\,\nu} = \chi^2_{0{,}01,\,12} = 26{,}2.$$

Следовательно, областью принятия гипотезы H_0 является отрезок [3,6; 26,2]. Так как $z_в$ принадлежит этому отрезку, то гипотеза H_0 принимается. ●

9. Проверка гипотез о законе распределения

Пусть необходимо проверить гипотезу H_0, состоящую в том, что наблюдаемая в эксперименте случайная величина X распределена по некоторому известному закону (нормальному, биномиальному, Пуассона и т.д.). Статистический критерий, применяемый для проверки такой гипотезы, называется *критерием согласия*.

Критериев согласия существует много. Рассмотрим наиболее часто используемый на практике критерий согласия – *критерий* χ^2 *Пирсона*.

Предположим, что сформулирована гипотеза H_0, состоящая в том, что случайная величина X имеет закон распределения известного вида, зависящий от параметров $\theta_1, \ldots, \theta_r$ (например, нормальный закон с параметрами m и σ или закон Пуассона с параметром λ).

Пусть в результате наблюдений за случайной величиной X получена выборка объёма n: $x_1, x_2, \ldots, x_n$.

Для проверки гипотезы H_0 с помощью критерия χ^2 Пирсона поступают следующим образом.

1. Выбирают уровень значимости α (вероятность отвергнуть гипотезу H_0 в случае, если она верна).

2. По выборке методом максимального правдоподобия (см. п. 5) оценивают параметры предполагаемого закона распределения.

3. Область возможных значений случайной величины X разбивают на k непересекающихся множеств S_1, S_2, $\ldots S_k$, которые представляют собой интервалы в случае, когда X – непрерывная случайная величина, либо группы отдельных значений, если эта величина дискретная.

4. Для каждого множества S_i, $i = 1, 2, \ldots, k$, подсчитывают число n_i элементов выборки, попавших в это множество (т.е. находят *эмпирические частоты*).

5. Используя предполагаемый закон распределения, вычисляют гипотетические вероятности p_i попадания случайной величины X в множества S_i:

$$p_i = P(X \in S_i), \quad i = 1,\ 2,\ \ldots,\ k.$$

6. Находят *теоретические частоты* n_i' попадания значений случайной величины X в множества S_i:

$$n_i' = np_i.$$

7. Вычисляют выборочное значение $\chi_в^2$ статистики критерия χ^2 Пирсона – случайной величины

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'}. \tag{5}$$

Применение статистики (5) для проверки гипотезы H_0 основано на следующей теореме.

Теорема 7. Если гипотеза H_0 верна, то статистика (5) критерия χ^2 асимптотически при $n \to \infty$ распределена по закону χ^2 с $\nu = k - r - 1$ степенями свободы, где k – число множеств S_i, r – число параметров предполагаемого распределения, оцененных по выборке.

8. По таблице квантилей χ^2-распределения по заданному уровню значимости α и числу степеней свободы $\nu = k - r - 1$ находят значение $\chi_{\alpha,\nu}^2$, удовлетворяющее условию

$$P(\chi_\nu^2 > \chi_{\alpha,\nu}^2) = \alpha.$$

9. Сравнивают значения $\chi_в^2$ и $\chi_{\alpha,\nu}^2$. В соответствии с критерием χ^2 Пирсона гипотеза H_0 принимается, если $\chi_в^2 \le \chi_{\alpha,\nu}^2$ (в этом случае говорят также, что гипотеза H_0 согласуется с результатами наблюдений). Если $\chi_в^2 > \chi_{\alpha,\nu}^2$, то гипотеза H_0 отклоняется.

Замечание 7. Необходимым условием применения критерия χ^2 Пирсона является выполнение неравенства $n_i' \ge 5$ для всех множеств S_i. Если для некоторых множеств S_i это условие не выполняется, то их следует объединить с соседними.

Пример 9. Результаты исследования прочности на сжатие 200 образцов бетона представлены в виде сгруппированного статистического ряда:

Интервалы прочности, МПа	19–20	20–21	22–23	23–24	24–25	25–26
Частоты n_i	10	26	56	64	30	14

С помощью критерия χ^2 Пирсона требуется проверить гипотезу H_0 о нормальном распределении прочности на сжатие (случайной величины X). Уровень значимости принять равным 0,05.

Решение. Определяем значения x_i^* середин интервалов и находим точечные оценки математического ожидания m и среднего квадратического отклонения σ гипотетического нормального распределения:

$$\widetilde{m} = \overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i^* n_i =$$

$$= \frac{1}{200}(19{,}5\cdot 10 + 20{,}5\cdot 26 + 21{,}5\cdot 56 + 22{,}5\cdot 64 + 23{,}5\cdot 30 + 24{,}5\cdot 14) =$$

$$= 22{,}1 \text{ МПа};$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} (x_i^* - \overline{x})^2 n_i =$$

$$= \frac{1}{200}((-2{,}6)^2\cdot 10 + (-1{,}6)^2\cdot 26 + (-0{,}6)^2\cdot 56 +$$

$$+ 0{,}4^2\cdot 64 + 1{,}4^2\cdot 30 + 2{,}4^2\cdot 14) = 1{,}52 \text{ МПа}^2;$$

$$\widetilde{\sigma} = \sqrt{s^2} = \sqrt{1{,}52} \approx 1{,}233 \text{ МПа}.$$

По формуле

$$u_i = (x_i - \widetilde{m})/\widetilde{\sigma},$$

вычисляем концы нормированных интервалов, при этом наименьшее значение u_i полагаем равным $-\infty$, а наибольшее — $+\infty$:

$$u_0 = -\infty; \quad u_1 \approx -1{,}70; \quad u_2 \approx -0{,}89; \quad u_3 \approx -0{,}08;$$

$$u_4 \approx 0{,}73; \quad u_5 \approx 1{,}54; \quad u_6 = +\infty.$$

Находим вероятности p_i попадания случайной величины X, имеющей нормальное распределение с параметрами $m = 22{,}1$, $\sigma = 1{,}233$, в частичные интервалы $[x_{i-1}, x_i)$ по формуле

$$p_i = P\{ x_{i-1} \le X < x_i \} = \Phi_0(u_i) - \Phi_0(u_{i-1}),$$

где

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}\,dt.$$

Используем при этом таблицу значений функции Лапласа. Полученные результаты, а также дальнейшие вычисления, необходимые для определения выборочного значения статистики критерия χ^2, приведем в таблице:

Интервалы наблюдаемых значений с.в. X $[x_{i-1}, x_i)$	Частоты n_i	Нормированные интервалы $[u_{i-1}, u_i)$	p_i	$n_i' = np_i$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
19 – 20	10	$(-\infty; -1{,}70)$	0,045	9	1	0,11
20 – 21	26	$[-1{,}70; -0{,}89)$	0,142	28,4	5,76	0,20
21 – 22	56	$[-0{,}89; -0{,}08)$	0,281	56,2	0,04	0,00
22 – 23	64	$[-0{,}08; 0{,}73)$	0,299	59,8	17,64	0,29
23 – 24	30	$[0{,}73; 1{,}54)$	0,171	34,2	17,64	0,52
24 – 25	14	$[1{,}54; +\infty)$	0,062	12,4	2,56	0,23
$\sum$	$n = 200$		1,000	200,0		$\chi_в^2 =$ $= 1{,}35$

В результате вычислений получили $\chi_в^2 = 1{,}35$. Находим по таблице квантилей χ^2-распределения по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $\nu = k - r - 1 = 6 - 2 - 1 = 3$ критическое значение $\chi^2_{0{,}5,\,3} = 7{,}815$. Так как $\chi_в^2 < \chi^2_{0{,}5,\,3}$, то гипотеза о нормальном распределении предела прочности на сжатие принимается. ●

10. Статистическое исследование зависимости между случайными величинами. Линейная регрессия

Пусть эксперимент описывается двумерной случайной величиной (X, Y). Изучается связь между случайными величинами X и Y.

Зависимость математического ожидания одной случайной величины от значений, принимаемых другой, называется *регрессионной зависимостью (регрессией)*.

Условное математическое ожидание $M(Y \mid X = x)$ случайной величины Y, рассматриваемое как функция от x, т. е. $M(Y \mid X = x) = \varphi(x)$, называется *функцией регрессии Y на X*. Аналогично определяется *функция регрессии X на Y*.

Вид функции регрессии (линейная, квадратичная, показательная и т. д.) характеризует форму связи между случайными величинами.

Для нахождения функций регрессий надо знать закон распределения случайного вектора (X, Y). Однако на практике этот закон, как правило, неизвестен. Поэтому, исходя из экспериментальных данных, строят оценки функций регрессий Y на X и X на Y, называемые *выборочными (эмпириче-*

скими) функциями этих регрессий и обозначаемые соответственно $\overline{y}_x = f(x)$ и $\overline{x}_y = g(y)$.

Пусть (x_1, y_1), (x_2, y_2),…, (x_n, y_n) — выборка значений двумерной случайной величины (X, Y). Если наблюденные данные изобразить в виде точек в декартовой системе координат, то получится точечная диаграмма, называемая *корреляционным полем*.

Общий вид выборочной функции регрессии $\overline{y}_x = f(x;\, a_0, a_1, \ldots, a_m)$ (a_i – параметры) часто выбирают исходя из анализа расположения точек (x_i, y_i) на корреляционном поле. Неизвестные значения параметров находятся из условия минимизации функции

$$S(a_0, a_1, \ldots, a_m) = \sum_{i=1}^{n}(y_i - f(x_i; a_0, a_1, \ldots, a_m))^2.$$

При этом применяется метод наименьших квадратов.

Для параметров линейной выборочной функции регрессии $\overline{y}_x = ax + b$ получается следующая система уравнений:

$$\begin{cases} b\sum x_i + a\sum x_i^2 = \sum x_i y_i, \\ nb + a\sum x_i = \sum y_i. \end{cases}$$

Параметры линейной выборочной функции регрессии $\overline{x}_y = cy + d$ находятся из аналогичной системы.

Величина

$$r_в = \frac{\frac{1}{n}\sum x_i y_i - \overline{x}\overline{y}}{s_X s_Y},$$

где s_X, s_Y – выборочные средние квадратические отклонения случайных величин X и Y, а $\overline{x}$, $\overline{y}$ – их выборочные средние, называется *выборочным коэффициентом корреляции* этих величин.

Заметим, что при замене (x_i, y_i) на $\left(\dfrac{x_i - c_1}{h_1}, \dfrac{y_i - c_2}{h_2}\right)$, где $h_1 > 0$, $h_2 > 0$, а постоянные c_1 и c_2 произвольны, выборочный коэффициент корреляции не меняет своего значения (т. е. переход к условным вариантам не изменяет величины $r_в$).

Имеет место оценка $|\, r_в \,| \le 1$.

Выборочный коэффициент корреляции характеризует тесноту линейной связи между случайными величинами X и Y (оценивает величину рассеяния наблюденных данных (x_i, y_i) относительно прямых линий регрессий $\overline{y}_x = ax + b$ и $\overline{x}_y = cy + d$). Чем ближе $|\, r_в \,|$ к 1, тем линейная связь между X и Y сильнее; чем ближе $|\, r_в \,|$ к 0, тем эта связь слабее. Если $|\, r_в \,| = 1$, то между

X и Y существует линейная функциональная связь. Если $r_в = 0$, то линейная регрессионная зависимость между X и Y отсутствует.

Выборочные уравнения прямых линий регрессий Y на X и X на Y можно записать в следующем виде:

$$\overline{y}_x - \overline{y} = r_в \frac{s_Y}{s_X}(x - \overline{x}), \qquad \overline{x}_y - \overline{x} = r_в \frac{s_X}{s_Y}(y - \overline{y}).$$

Величины $\rho_{YX} = r_в \frac{s_Y}{s_X}$ и $\rho_{XY} = r_в \frac{s_X}{s_Y}$ называются *линейными коэффициентами регрессий*.

Выборочный коэффициент корреляции является точечной оценкой коэффициента корреляции генеральной совокупности (т. е. теоретического коэффициента корреляции).

Пусть $r_в$ – выборочный коэффициент корреляции, вычисленный по выборке объема n из генеральной совокупности (X, Y), имеющей двумерное нормальное распределение. Если $r_в \neq 0$, то это еще не значит, что коэффициент корреляции генеральной совокупности $r_Г$ также отличен от нуля. Проверка нулевой гипотезы H_0: $r_Г = 0$ против альтернативной гипотезы H_1: $r_Г \neq 0$ осуществляется с помощью статистики

$$T = \frac{r_в \sqrt{n-2}}{\sqrt{1 - r_в^2}},$$

которая, при условии справедливости гипотезы H_0, имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы. Критическая область критерия при уровне значимости α определяется неравенством $|T| > t_{\alpha/2,\, n-2}$.

Если гипотеза H_0 отклоняется, то это означает, что выборочный коэффициент корреляции $r_в$ значимо отличается от нуля (коротко: значим), а X и Y коррелированы, т. е. связаны линейной регрессионной зависимостью.

Если гипотеза H_0 принимается, то выборочный коэффициент корреляции незначим, а X и Y некоррелированы, т.е. не связаны линейной регрессионной зависимостью.

Пример 10. Найти уравнения линейных регрессий Y на X и X на Y по 10 парам наблюдаемых значений двумерной случайной величины (X, Y):

x_i	6	11	11	7	8	10	12	6	10	9
y_i	27	32	33	30	30	33	34	29	31	32

Решение. Найдем коэффициенты (параметры) искомых уравнений. Для этого составим расчетную таблицу:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
6	27	162	36	729
11	32	352	121	1024
11	33	363	121	1089
7	30	210	49	900
8	30	240	64	900
10	33	330	100	1089
12	34	408	144	1156
6	28	168	36	784
10	31	310	100	961
9	32	288	81	1024
$\sum x_i = 90$	$\sum y_i = 310$	$\sum x_i y_i = 2831$	$\sum x_i^2 = 852$	$\sum y_i^2 = 9656$

Записываем систему уравнений для нахождения коэффициентов a и b уравнения линейной регрессии Y на X:

$$\begin{cases} 90b + 852a = 2831, \\ 10b + 90a = 310. \end{cases}$$

Решая ее, находим: $a \approx 0{,}98$, $b \approx 22{,}18$. Следовательно, уравнение линейной регрессии Y на X имеет вид

$$\bar{y}_x = 0{,}98x + 22{,}18.$$

Коэффициенты c и d уравнения линейной регрессии X на Y удовлетворяют системе уравнений

$$\begin{cases} 310d + 9656c = 2831, \\ 10d + 310c = 90. \end{cases}$$

Решая систему, находим $c \approx 0{,}89$, $d \approx -18{,}59$. Следовательно, уравнение линейной регрессии X на Y имеет вид

$$\bar{x}_y = 0{,}89y - 18{,}59. \bullet$$

Пример 11. По выборке объема $n = 122$, извлеченной из нормальной двумерной совокупности (X, Y), найден выборочный коэффициент корреляции $r_в = 0{,}4$. При уровне значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции $r_Г$ при конкурирующей гипотезе H_1: $r_Г \neq 0$.

Решение. Найдем выборочное значение статистики критерия:

$$T_в = \frac{0{,}4\sqrt{122-2}}{\sqrt{1-0{,}4^2}} \approx 4{,}78.$$

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

По уровню значимости α = 0,05 и числу степеней свободы $\nu = n-2 = 122-2 = 120$ находим по таблице квантилей распределения Стьюдента критическую точку $t_{\alpha/2,n-2} = t_{0,025;120} = 1,98$.

Поскольку $T_в$ = 4,78 > 1,98, нулевую гипотезу отвергаем. Другими словами, выборочный коэффициент корреляции значимо отличается от нуля. ●

11. Выборочное корреляционное отношение

Будем исследовать тесноту любой, вообще говоря, нелинейной корреляционной связи между случайными величинами X и Y.

Наблюдённые данные для случайного вектора (X,Y) представим в виде *корреляционной таблицы*.

$X \backslash Y$	y_1	y_2	…	y_l	n_i
x_1	m_{11}	m_{12}	…	m_{1l}	n_1
x_2	m_{21}	m_{22}	…	m_{2l}	n_2
…	…	…	…	…	…
x_k	m_{k1}	m_{k2}	…	m_{kl}	n_k
m_j	m_1	m_2	…	m_l	n

Эта таблица включает в себя следующие элементы:

$x_1, x_2, \ldots, x_k$ и $y_1, y_2, \ldots, y_l$ – различные наблюдённые значения (варианты) случайных величин X и Y соответственно;

m_{ij} – число появлений пары значений (x_i, y_j) в выборке (частота этой пары);

n_i – частота значения x_i $(n_i = \sum_j m_{ij})$;

m_j – частота значения y_j $(m_j = \sum_i m_{ij})$;

n – объём выборки $(n = \sum_i n_i = \sum_j m_j = \sum_{i,j} m_{ij})$.

Для каждого значения x_i найдём среднее значение случайной величины Y:

$$\bar{y}_i = \frac{1}{n_i}\sum_j y_j m_{ij} .$$

Величины $\bar{y}_i$ называются *групповыми (условными) средними*.

Рассеяние условных средних относительно общего выборочного среднего характеризуется величиной

$$s_{\bar{y}}^{\ 2}=\frac{1}{n}\sum_i(\bar{y}_i-\bar{y})^2 n_i\,,$$

которая называется *межгрупповой дисперсией*.

Величина

$$s_{\bar{y}}=\sqrt{s_{\bar{y}}^{\ 2}}$$

называется *межгрупповым средним квадратическим отклонением*.

Выборочным корреляционным отношением Y на X называют величину

$$\eta_{YX}=\frac{s_{\bar{y}}}{s_Y}\,,$$

где s_Y – выборочное среднее квадратическое отклонение случайной величины Y. Аналогично определяется η_{XY} – выборочное корреляционное отношение X на Y.

Выборочное корреляционное отношение обладает следующими свойствами:

1. $0\le\eta_{YX}\le 1$.
2. Если $\eta_{YX}=0$, то корреляционная зависимость Y на X отсутствует.
3. Если $\eta_{YX}=1$, то существует функциональная зависимость Y от X.
4. $\eta_{YX}\ge|r_в|$.
5. $\eta_{YX}=|r_в|$ тогда и только тогда, когда имеет место точная линейная корреляционная зависимость Y от X.

Выборочное корреляционное отношение η_{YX} является мерой тесноты корреляционной зависимости Y от X:

чем ближе η_{YX} к 1, тем выше степень корреляционной зависимости Y от X, которая при $\eta_{YX}=1$ переходит в функциональную зависимость;

чем ближе η_{YX} к 0, тем меньше степень корреляционной зависимости Y от X, причём при $\eta_{YX}=0$ корреляционная зависимость Y на X отсутствует.

Проверка гипотезы H_0 об отсутствии корреляционной зависимости Y от X производится с помощью статистики

$$W=\frac{(n-k)\eta_{YX}^2}{(k-1)(1-\eta_{YX}^2)}, \tag{6}$$

где n – объём выборки, k – число различных наблюдённых значений случайной величины X в выборке. Если гипотеза H_0 верна, то статистика (6) имеет распределение Фишера с числом степеней свободы $\nu_1=k-1$ и $\nu_2=n-k$. Границу критического множества определяет квантиль F_{α,ν_1,ν_2} (α – уровень значимости). Гипотеза H_0 отвергается, если выборочное значение статистики W принадлежит критическому множеству, т. е. если $W_в>F_{\alpha,\nu_1,\nu_2}$.

При проверке гипотезы H_0 о наличии *линейной* корреляционной связи между случайными величинами Y и X используется статистика

$$W^* = \frac{(n-k)(\eta_{YX}^2 - r_в^2)}{(k-2)(1-\eta_{YX}^2)}, \qquad (7)$$

имеющая при справедливости гипотезы H_0 распределение Фишера с $\nu_1 = k-2$ и $\nu_2 = n-k$ степенями свободы.

Пример 12. Данные 15 наблюдений над двумерной случайной величиной (X,Y) представлены в таблице:

x_i	2	4	9	13	15
y_{ij}	1, 3, 4	7, 8, 12	14, 19, 21	11, 9, 6	8, 7, 3

Проверить гипотезу об отсутствии корреляционной зависимости Y от X при уровне значимости $\alpha = 0{,}05$.

Решение. Имеем $k=5$, $n=15$, $n_i = 3$ $(i=1,2,3,4,5)$. Находим

$$\bar{y}_1 = \frac{1+3+4}{3} \approx 2{,}67, \quad \bar{y}_2 = \frac{7+8+12}{3} = 9, \quad \bar{y}_3 = \frac{14+19+21}{3} = 18,$$

$$\bar{y}_4 = \frac{11+9+6}{3} \approx 8{,}67, \quad \bar{y}_5 = \frac{8+7+3}{3} = 6;$$

$$\bar{y} = \frac{1}{15}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \frac{133}{15} \approx 8{,}87;$$

$$s_{\bar{y}}^{\,2} = \frac{1}{n}\sum_{i=1}^{k}(\bar{y}_i - \bar{y})^2 n_i =$$

$$= \frac{(2{,}67-8{,}87)^2 + (9-8{,}87)^2 + (18-8{,}87)^2 + (8{,}67-8{,}87)^2 + (6-8{,}87)^2}{5} =$$

$$= \frac{(-6{,}2)^2 + 0{,}13^2 + 9{,}13^2 + (-0{,}2)^2 + (-2{,}87)^2}{5} =$$

$$= \frac{38{,}44 + 0{,}0169 + 83{,}3569 + 0{,}04 + 8{,}2369}{5} = \frac{130{,}0907}{5} = 26{,}01814;$$

$$s_Y^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - (\bar{y})^2 = \frac{1641}{15} - 8{,}87^2 = 30{,}7231.$$

Тогда

$$\eta_{YX}^2 = \frac{26{,}01814}{30{,}7231} \approx 0{,}8469.$$

Вычисляем далее

$$W_в = \frac{(n-k)\eta_{YX}^2}{(k-1)(1-\eta_{YX}^2)} = \frac{10\cdot 0{,}8469}{4\cdot(1-0{,}8469)} \approx 13{,}83$$

По таблице квантилей распределения Фишера находим $F_{0,05;4;10} = 3{,}48$.

Как видим, $W_в > F_{0,05;4;10}$. Следовательно, нулевую гипотезу об отсутствии корреляционной зависимости Y от X отвергаем. Таким образом, корреляционная зависимость Y от X существует. ●

Пример 13. В условиях примера 12 проверить гипотезу о наличии линейной корреляционной зависимости Y от X при уровне значимости α = 0,05.

Решение. Находим

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{43}{5} = 8{,}6; \qquad \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_i y_{ij} = \frac{1218}{15} = 81{,}2;$$

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^{\,2} n_i - (\bar{x})^2 = 99 - 8{,}6^2 = 25{,}04.$$

Тогда

$$r_в^2 = \frac{\left(\dfrac{1}{n}\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{n_i} x_i y_{ij} - \bar{x}\cdot\bar{y}\right)^2}{s_X^2 s_Y^2} = \frac{(81{,}2 - 8{,}6\cdot 8{,}87)^2}{25{,}04\cdot 30{,}7231} \approx 0{,}0314.$$

Находим выборочное значение статистики (7):

$$W_в^* = \frac{(n-k)(\eta_{YX}^2 - r_в^2)}{(k-2)(1-\eta_{YX}^2)} = \frac{10(0{,}8469 - 0{,}0314)}{3(1-0{,}8469)} \approx 17{,}76.$$

По таблице квантилей распределения Фишера находим критическое значение $F_{0,05;3;10} = 3{,}71$.

Так как $W_в^* > F_{0,05;3;10}$, то гипотезу о наличии линейной корреляционной зависимости Y от X следует отклонить. ●

Задачи для самостоятельного решения

Задачи к пунктам 1-3

1. Имеются результаты измерения длины (в мм) 30 случайно отобранных заготовок:

39, 41, 40, 40, 43, 41, 44, 42, 41, 41, 43, 42, 39, 40, 42,
43, 41, 42, 41, 39, 42, 41, 42, 40, 41, 43, 41, 39, 40, 41.

Составить вариационный и статистический ряды данной выборки. Построить полигон частот.

2. Измерения ёмкости у 80 транзисторов дали следующие результаты:

1,9	3,1	1,3	0,7	3,2	1,1	2,9	2,7	2,7	4,0
1,7	3,2	0,9	0,8	3,1	1,2	2,6	1,9	2,3	3,2
4,1	1,3	2,4	4,5	2,5	0,9	1,4	1,6	2,2	3,1
1,5	1,1	2,3	4,3	2,1	0,7	1,2	1,5	1,8	2,9
0,8	0,9	1,7	4,1	4,3	2,6	0,9	0,8	1,2	2,1
3,2	2,9	1,1	3,2	4,5	2,1	3,1	5,1	1,1	1,9
0,9	3,1	0,9	3,1	3,3	2,8	2,5	4,0	4,3	1,1
2,1	3,8	4,6	3,8	2,3	3,9	2,4	4,1	4,2	0,9

Построить гистограмму относительных частот по этой выборке, предварительно проведя группировку. В качестве длины интервала группировки взять следующие значения:

а) $\Delta = 0{,}3$; б) $\Delta = 1{,}2$.

3. Найти эмпирическую функцию распределения и начертить ее график для выборки, представленной следующим статистическим рядом:

x_i	39	40	41	42	43	44
n_i	4	5	9	7	4	1

4. Построить гистограмму частот для выборки, представленной следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
n_i	1	2	7	18	12	8	2

5. Найти среднее и дисперсию выборки, представленной следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13
n_i	1	2	4	2	1	1

6. Доказать следующие свойства выборочного среднего:

а) $\sum\limits_{i=1}^{n}(x_i - \bar{x}) = 0$;

б) $\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 < \sum\limits_{i=1}^{n}(x_i - a)^2$, где $a \in \mathbf{R}$, $a \neq \bar{x}$.

Задачи к пункту 4

7. Пусть $\tilde{\theta}$ – несмещенная оценка параметра θ, $D(\tilde{\theta}) < \infty$. Показать, что $(\tilde{\theta})^2$ является смещенной оценкой θ^2, и вычислить смещение.

8. Доказать, что выборочная дисперсия S^2 является состоятельной оценкой дисперсии наблюдаемой случайной величины *X*.

9. Показать, что значение эмпирической функции распределения $F_n^*(x)$ в точке *x* является несмещенной и состоятельной оценкой значения теоретической функции распределения *F*(*x*) в той же точке.

10. Показать, что выборочное среднее $\bar{x}$ является эффективной оценкой параметра λ распределения Пуассона.

11. Показать, что относительная частота $\frac{k}{n}$ появления события *A* в *n* независимых испытаниях является эффективной оценкой вероятности *p* появления события *A* в одном испытании.

12. По выборке объема *n* из нормально распределенной генеральной совокупности с параметрами *m* и σ (*m* известно) оценивается дисперсия σ^2. Показать, что статистика

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

является эффективной оценкой σ^2.

13. Пусть $x_1, x_2, \ldots, x_n$ – выборка из генеральной совокупности с показательной плотностью распределения

$$p(x,\theta) = \begin{cases} (1/\theta)e^{-x/\theta} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0, \end{cases}$$

где $\theta > 0$. Показать, что $\bar{x}$ является эффективной оценкой параметра θ.

14. Пусть $x_1, x_2, \ldots, x_n$ – выборка из нормально распределенной генеральной совокупности с параметрами *m* и σ. Найти информацию Фишера $J_n(\sigma^2)$.

Задачи к пункту 5

15. Найти оценку максимального правдоподобия параметра λ распределения Пуассона.

16. Найти оценку максимального правдоподобия параметра $\theta > 0$ показательного распределения с плотностью

$$p(x,\theta)=\begin{cases}(1/\theta)e^{-x/\theta} & \text{при } x\geq 0,\\ 0 & \text{при } x<0.\end{cases}$$

17. Случайная величина X имеет плотность распределения

$$p(x)=\begin{cases}kx & \text{при } x\in[0,\sqrt{2/k}],\\ 0 & \text{при } x\notin[0,\sqrt{2/k}].\end{cases}$$

Найти оценку максимального правдоподобия математического ожидания случайной величины X по выборке объема n.

18. Найти оценку максимального правдоподобия параметра θ равномерного на отрезке $[0, \theta]$ распределения.

19. Случайная величина X имеет равномерное на отрезке $[a, b]$ распределение. Методом моментов найти оценки неизвестных параметров a и b.

20. Используя метод моментов, найти оценки неизвестных параметров $a>0$ и $b>0$ для гамма-распределения с плотностью

$$p(x)=\begin{cases}\dfrac{b^a}{\Gamma(a)}x^{a-1}e^{-bx} & \text{при } x>0,\\ 0 & \text{при } x\leq 0.\end{cases}$$

Задачи к пункту 7

21. По выборке объема $n=64$ найдена средняя длина детали $\bar{x}=50$ мм. Считая, что длина детали X – нормально распределенная случайная величина, найти доверительный интервал, который с доверительной вероятностью $1-\alpha=0{,}95$ накрывает неизвестное математическое ожидание m длины детали, если генеральное среднее квадратическое отклонение $\sigma=0{,}5$ мм.

22. Имеется выборка объема 12 из нормально распределённой генеральной совокупности:

№ элемента выборки	1	2	3	4	5	6	7	8	9	10	11	12
Наблюдённое значение	–0,5	1,2	0	0,8	1,2	–0,4	0,2	–0,2	1,5	0,6	–0,4	1,0

Найти доверительный интервал с вероятностью $1-\alpha=0{,}95$ для ее математического ожидания.

23. Из нормально распределенной генеральной совокупности извлечена выборка объема n = 10. Результаты приведены в таблице:

Наблюденное значение (x_i)	–2	1	2	3	4	5
Частота (n_i)	2	1	2	2	2	1

Найти доверительный интервал для математического ожидания m генеральной совокупности. Доверительную вероятность $1 - \alpha$ принять равной 0,95.

24. Найти минимальный объем выборки, на основании которой можно было бы оценить математическое ожидание времени исполнения некоторой технической операции с ошибкой, не превышающей 10 с, и надежностью $1-\alpha=0{,}95$, если предположить, что время исполнения этой технической операции X является нормально распределенной случайной величиной, имеющей среднее квадратическое отклонение $\sigma = 50$ с.

25. Систематические ошибки измерительного прибора практически равны нулю, а случайные распределены нормально со средним квадратическим отклонением $\sigma = 20$ м. Необходимо, чтобы абсолютное значение разности между оценкой измеряемой величины и истинным ее значением не превосходило 10 м. Определить, с какой вероятностью будет выполнено это условие, если число наблюдений равно 25.

26. По независимым выборкам $x_1, x_2,\dots, x_n$ и $y_1, y_2,\dots, y_n$ из двух нормально распределённых генеральных совокупностей с параметрами (m_1, σ_1) и (m_2, σ_2) соответственно построить доверительный интервал с доверительной вероятностью $1-\alpha$ для разности $m_1 - m_2$, если σ_1 и σ_2 известны.

27. Построить доверительный интервал с вероятностью $1-\alpha=0{,}95$ для дисперсии σ^2 нормально распределенной случайной величины X, если исправленная выборочная дисперсия $s_0^2=10$, а объем выборки $n = 20$.

28. Из партии однотипных сопротивлений отобрано для контроля 10 штук. Измерения дали следующие отклонения от номинала (в килоомах):

№ сопротивления	1	2	3	4	5	6	7	8	9	10
Отклонение	+1	+2	–2	+2	+4	+2	+5	+3	–2	+10

Считая, что отклонение сопротивления от номинала X – нормально распределенная случайная величина, найти доверительные интервалы для математического ожидания m и дисперсии σ^2 этого отклонения. Доверительную вероятность $1-\alpha$ принять равной 0,95.

Задачи к пункту 8

29. 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, причем выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило $\bar{x} = 9{,}3$ л. Предполагается, что выборка расходов топлива получена из нормально распределенной генеральной совокупности с дисперсией $\sigma^2 = 4$ л^2. Проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива. Уровень значимости принять равным 0,05.

30. Автомат производит болты с номинальным значением контролируемого размера m_0 = 40 мм. Результаты предыдущих измерений дают основание предполагать, что действительные размеры болтов образуют нормально распределенную совокупность с дисперсией $\sigma^2 = 1$ мм^2. Партия болтов бракуется, если среднее выборочное контролируемого размера будет больше 40,1 мм. Найти вероятность ошибок первого и второго рода при альтернативной гипотезе H_1: m = 40,3 мм, если решение принимается по выборке объема n = 36.

31. Техническая норма предусматривает в среднем 40 с на выполнение определенной технической операции. От работников поступили сигналы, что они в действительности затрачивают на эту операцию больше времени. Для проверки произведены измерения времени выполнения операции у 16 работников и получены следующие результаты: $\bar{x} = 42$ с (среднее время выполнения операции), $s_0 = 3{,}5$ с (исправленное среднее квадратическое отклонение). Предполагается, что распределение контролируемого промежутка времени является нормальным. Можно ли по имеющимся хронометрическим данным на уровне значимости α = 0,01 отклонить гипотезу о том, что действительное среднее время исполнения данной технической операции соответствует норме?

32. Для нормально распределенной случайной величины проверяется нулевая гипотеза о математическом ожидании H_0: m = 10 при альтернативной гипотезе H_1: m = 9. Известна дисперсия распределения: $\sigma^2 = 4$. Какой минимальный объем выборки n следует взять, чтобы ошибка первого рода была равна α = 0,01, а ошибка второго рода не превышала 0,1?

33. Выборка 50 электроламп завода $А$ показала среднюю продолжительность работы $\bar{x}_А = 1282$ ч, а такая же по объему выборка того же типа ламп завода $Б$ — $\bar{x}_Б = 1208$ ч. Предполагается, что продолжительности работы электроламп, выпускаемых заводами $А$ и $Б$, являются нормально распределенными случайными величинами со средними квадратическими отклонениями $\sigma_A = 80$ ч и $\sigma_Б = 94$ ч соответственно. Проверить гипотезу о том, что эти заводы выпускают лампы одинакового качества (средний срок службы ламп обоих заводов одинаков). Уровень значимости принять равным 0,05.

34. При обработке втулок на станке-автомате было отобрано две пробы по 10 штук деталей в каждой. Результаты измерения диаметров этих втулок в порядке обработки указаны в следующей таблице:

№ детали	1	2	3	4	5	6	7	8	9	10
Проба 1	2,066	2,063	2,068	2,060	2,067	2,063	2,059	2,062	2,062	2,060
Проба 2	2,063	2,060	2,057	2,056	2,059	2,058	2,062	2,059	2,059	2,057

Предполагается, что выборки диаметров втулок получены из нормально распределенных генеральных совокупностей с одинаковыми средними квадратическими отклонениями. Проверить гипотезу о том, что средние (математические ожидания) генеральных совокупностей в моменты выбора обеих проб равны, т. е. что режим работы станка от пробы к пробе не изменился. Уровень значимости принять равным 0,05.

35. Выдвинута гипотеза, что применение нового типа резца сокращает время обработки некоторой детали. Проведено 10 измерений времени, затрачиваемого на обработку этой детали старым и новым резцами. Получены следующие результаты (в минутах): старый тип резца — 58, 58, 56, 38, 70, 38, 42, 75, 68, 67; новый тип резца — 57, 55, 63, 24, 67, 43, 33, 68, 56, 54. Предполагается, что время обработки детали старым и новым резцами — случайные величины, имеющие нормальные распределения с равными средними квадратическими отклонениями. Проверить гипотезу равенства среднего времени, затрачиваемого на изготовление детали с помощью двух типов резцов. Уровень значимости принять равным 0,05.

36. Точность работы станка-автомата проверяется по дисперсии контролируемого размера деталей, которая не должна превышать $\sigma_0^2 = 0{,}04$. Взята проба из 11 случайно отобранных деталей, и получены следующие результаты (в мм): 100,6; 99,6; 100,0; 100,1; 100,3; 100,0; 99,9; 100,2; 100,4; 100,6; 100,5. Предполагается, что распределение контролируемого размера является нормальным. На основании имеющихся данных проверить, обеспечивает ли станок заданную точность. Уровень значимости принять равным 0,05.

37. До наладки станка была проверена точность изготовления 10 втулок и найдено значение исправленной выборочной дисперсии $s_{01}^2 = 9{,}6$ мкм^2. После наладки подверглись контролю еще 15 втулок и получено новое значение исправленной выборочной дисперсии $s_{02}^2 = 5{,}7$ мкм^2. Предполагается, что контролируемый размер втулки имеет нормальное распределение. Можно ли считать, что в результате наладки станка точность изготовления деталей увеличилась? Принять $\alpha = 0{,}15$.

38. При измерении производительности двух агрегатов получены следующие результаты (в кг вещества за час работы):

№ замера	1	2	3	4	5
Агрегат *A*	14,1	10,1	14,7	13,7	14,0
Агрегат *B*	14,0	14,5	13,7	12,7	14,1

Можно ли считать, что производительности агрегатов *A* и *B* одинаковы, в предположении, что обе выборки получены из нормально распределённых генеральных совокупностей? Принять $\alpha = 0{,}10$.

39. Из нормально распределённых генеральных совокупностей с параметрами (m_1, σ) и (m_2, σ) получены две выборки объемов n_1 и n_2 соответственно. Предлагается отклонить гипотезу H_1: $m_1 = m_2$, если доверительные интервалы для m_1 и m_2 не пересекаются. Показать, что при доверительной вероятности $1-\alpha$ уровень значимости этого критерия меньше α.

Задачи к пункту 9

40. При 50 подбрасываниях монеты герб появился 20 раз. Можно ли считать монету симметричной? Принять $\alpha = 0{,}10$.

41. При 120 бросаниях игральной кости шестерка выпала 40 раз. Согласуется ли этот результат с утверждением, что кость правильная? Принять $\alpha = 0{,}05$.

42. На экзамене студент отвечает только на один вопрос по одной из трех частей курса. Анализ вопросов, заданных 60 студентам, показал, что 23 студента получили вопросы из первой, 15 — из второй и 22 — из третьей части курса. Можно ли считать, что студент, идущий на экзамен, с равной вероятностью получит вопрос по любой из трех частей курса? Принять $\alpha = 0{,}10$.

43. Приведены данные о фактических объёмах сбыта (в условных единицах) в пяти районах:

Район	1	2	3	4	5
Фактический объём сбыта	110	130	70	90	100

Согласуются ли эти результаты с предположением о том, что сбыт продукции в этих районах должен быть одинаковым? Принять $\alpha = 0{,}01$.

44. Приведены данные об отказах аппаратуры за 10 000 часов работы (всего обследовано $n = 757$ экземпляров):

Число отказов, k	0	1	2	3	4	5	≥ 6
Количество случаев, в которых наблюдалось k отказов, n_k	427	235	72	21	1	1	0

Проверить гипотезу H_0 о том, что число отказов имеет распределение Пуассона, при $\alpha = 0{,}10$.

45. Произведены 500 измерений боковой ошибки наводки при стрельбе с самолета по наземной цели. Результаты измерений (в тысячных долях радиана) сведены в статистический ряд:

$x_{i-1} \div x_i$	$-4 \div -3$	$-3 \div -2$	$-2 \div -1$	$-1 \div 0$	$0 \div 1$	$1 \div 2$	$2 \div 3$	$3 \div 4$
n_i	6	25	72	133	120	88	46	10

С помощью критерия χ^2 Пирсона проверить гипотезу H_0 о том, что измерения подчиняются нормальному распределению при $\alpha = 0{,}01$.

46. Выборка представлена следующим интервальным статистическим рядом:

$x_{i-1} - x_i$	3,0–3,6	3,6–4,2	4,2–4,8	4,8–5,4	5,4–6,0	6,0–6,6	6,6–7,2
n_i	2	8	35	43	22	15	5

При уровне значимости $\alpha = 0{,}01$ проверить гипотезу о нормальном распределении генеральной совокупности.

47. Выборка представлена интервальным статистическим рядом:

$x_{i-1} - x_i$	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
n_i	133	45	15	4	2	1

При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о показательном распределении генеральной совокупности.

Задачи к пунктам 10, 11

48. По данным выборки вычислить коэффициент корреляции и найти уравнения прямых регрессий Y на X и X на Y. Изобразить эти прямые на корреляционном поле.

x_i	8	10	5	8	9
y_i	1	3	1	2	3

49. Вычислить коэффициент корреляции и найти уравнения прямых регрессий Y на X и X на Y по данным, приведенным в следующей корреляционной таблице:

X \ Y	5 – 15	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65
10 – 20	5	7				
20 – 30		20	23			
30 – 40			30	47	2	
40 – 50			10	11	20	6
50 – 60				9	7	3

50. По выборке объема n = 122, извлеченной из нормальной двумерной совокупности (X, Y), найден выборочный коэффициент корреляции $r_в = 0{,}4$. При уровне значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции $r_Г$ при конкурирующей гипотезе H_1: $r_Г \neq 0$.

51. Какое наименьшее значение выборочного коэффициента корреляции следует считать значимым на 5%-ном уровне значимости (т. е. при α = 0,05), если объем выборки n = 38? Предполагается, что выборка извлечена из нормально распределенной двумерной генеральной совокупности

52. Найти выборочное уравнение регрессии $\overline{y}_x = ax^2 + bx + c$ по данным корреляционной таблицы

X \ Y	25	45	110
2	20		
3		30	1
5		1	48

53. Найти выборочное уравнение регрессии Y на X вида $\overline{y}_x = b + \dfrac{a}{x}$ по данным корреляционной таблицы

X \ Y	3,4–3,8	3,8–4,2	4,2–4,6	4,6–5,0	5,0–5,4	5,4–5,8	5,8–6,2	6,2–6,6
4 – 6						2	3	3
6 – 8				3	3	2		1
8 – 10			1	1	2	1	1	
10 – 12		1	4	3				
12 – 14		1	2	3				
14 – 16		3	2					
16 – 18	2	2						
18 – 20	3	1						

54. Для совокупности 50 однотипных предприятий исследуется зависимость между величиной основных производственных фондов X (млн. руб.) и суточной выработкой продукции Y (т).

Величина ОПФ, млн. руб.(X)	Середины интервалов	Суточная выработка продукции, т (Y)					Всего n_i
		7 – 11	11 – 15	15 – 19	19 – 23	23 – 27	
	x_i \ y_j	9	13	17	21	25	
20 – 25	22,5	2	1	–	–	–	3
25 – 30	27,5	3	6	4	–	–	13
30 – 35	32,5	–	3	11	7	–	21
35 – 40	37,5	–	1	2	6	2	13
40 – 45	42,5	–	–	–	1	1	2
Всего n_i		5	11	17	14	3	

Вычислить выборочный коэффициент корреляции $r_в$ и статистическое корреляционное отношение η_{YX}. При уровне значимости $\alpha = 0{,}05$ проверить гипотезу о наличии линейной корреляционной связи между случайными величинами X и Y.

Литература

1. Бородич, С.М. Теория вероятностей и математическая статистика : методические рекомендации / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2015. – 51 с.

2. Бородич, С.М. Теория вероятностей и математическая статистика : методические рекомендации / С.М. Бородич, Т.В. Кавитова. – Витебск : ВГУ имени П.М. Машерова, 2017. – 52 с.

3. Бочаров, П.П. Теория вероятностей. Математическая статистика / П.П. Бочаров, А.В. Печинкин. – 2-е изд. – М.: ФИЗМАТЛИТ, 2005. – 296 с.

4. Герасимович, А.И. Математическая статистика / А.И. Герасимович, Я.И. Матвеева. – Минск: Выш. шк., 1978. – 200 с.

5. Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2003. – 479 с.

6. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов / В.Е. Гмурман. – 9-е изд., стер. – М.: Высш. шк., 2004. – 404 с.

7. Емельянов, Г.В. Задачник по теории вероятностей и математической статистике: учеб. пособие / Г.В. Емельянов, В.П. Скитович. – 2-е изд., стер. – СПб. [и др.]: Лань, 2007. – 332 с.

8. Кремер, Н.Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н.Ш. Кремер. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2004. – 573 с.

9. Кремлёв, А.Г. Математика. Раздел “Статистика” : учеб. пособие / А.Г. Кремлёв. – Екатеринбург: Изд-во УрГЮА, 2001. – 140 с.

10. Максимов, Ю.Д. Математика. Выпуск 8. Математическая статистика: опорный конспект / Ю.Д. Максимов. – СПб.: Изд-во СПбГПУ, 2002. – 96 с.

11. Севастьянов, Б.А. Курс теории вероятностей и математической статистики / Б.А. Севастьянов. – М.: Наука. Главная редакция физико-математической литературы, 1982. – 256 с.

Учебное издание

БОРОДИЧ Сергей Митрофанович
КАВИТОВА Татьяна Валерьевна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

Технический редактор *Г.В. Разбоева*
Компьютерный дизайн *Л.Р. Жигунова*

Подписано в печать 2020. Формат 60x84 $^{1}/_{16}$. Бумага офсетная.
Усл. печ. л. 2,79. Уч.-изд. л. 1,56. Тираж экз. Заказ .

Издатель и полиграфическое исполнение – учреждение образования
«Витебский государственный университет имени П.М. Машерова».

Свидетельство о государственной регистрации в качестве издателя,
изготовителя, распространителя печатных изданий
№ 1/255 от 31.03.2014.

Отпечатано на ризографе учреждения образования
«Витебский государственный университет имени П.М. Машерова».
210038, г. Витебск, Московский проспект, 33.