КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

В. В. Бандеров, Л. Н. Кашапов,
Н. Ф. Кашапов, В. Ю. Чебакова

МЕТОДЫ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ

медико-биологических данных

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ
В 2 ЧАСТЯХ

1
часть

2023

КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

В. В. Бандеров, Л. Н. Кашапов,
Н. Ф. Кашапов, В. Ю. Чебакова

МЕТОДЫ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ МЕДИКО-БИОЛОГИЧЕСКИХ ДАННЫХ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

В 2 частях

Часть I

Москва
АНО **«Издательский Дом «Научное обозрение»**
2023

УДК 311.1
ББК 60.6
М54

Печатается по рекомендации Редакционно-издательского совета
Института ВМиИТ КФУ (протокол № 3 от 24 ноября 2022 года)

Под редакцией кандидата физико-математических наук, доцента
кафедры анализа данных и технологий программирования
Института ВМиИТ КФУ **В. Т. Дубровина**

Рецензенты:
кандидат физико-математических наук, доцент кафедры
прикладной математики Института транспортных сооружений
Казанского государственного архитектурно-строительного
университета (КГАСУ) **Ф. Г. Габбасов**;
старший научный сотрудник, кандидат технических наук,
доцент кафедры динамики процессов и управления Казанского
национального исследовательского технического университета
им. А. Н. Туполева (КНИТУ-КАИ) **Р. А. Сабитов**

М54 **Методы** математической обработки медико-биологических данных : учеб.-метод. пособие : в 2 ч. - Ч. I / В. В. Бандеров, Л. Н. Кашапов, Н. Ф. Кашапов, В. Ю. Чебакова ; Каз. федер. ун-т ; под ред. В. Т. Дубровина. - Москва : АНО **«Изд. Дом «Науч. обозрение»**, 2023. - 65 с., рис., табл., формулы. - Библиогр.: с. 61-62. – Прил.: с. 63.
ISBN 978-5-6049342-1-0 (ч. 1)
ISBN 978-5-6049342-0-3

В учебно-методическом пособии приводится необходимый базовый материал по прикладной статистике в привязке к конкретным задачам, связанным со статистической обработкой медико-биологических данных.

Предназначено для студентов, обучающихся по направлению «Биотехнические системы и технологии», а также может быть полезно для самостоятельного изучения статистики студентами таких направлений, как «Бизнес-информатика» и «Прикладная математика».

УДК 311.1
ББК 60.6

ISBN 978-5-6049342-1-0 (ч. 1)
ISBN 978-5-6049342-0-3

Предисловие

Данное учебно-методическое пособие, предложенное Вашему вниманию, разработано по курсу «Методы математической обработки медико-биологических данных» и состоит из двух частей.

Часть I настоящего пособия предоставляет студентам возможность улучшить навыки решения задач по прикладной статистике. При этом основной акцент делается на изучение методов решения практических задач, а необходимый теоретический материал поясняется на примерах по обработке медико-биологических данных. Рассмотрение вопросов прикладной статистики сопровождается демонстрацией достаточного количества решений прикладных примеров, ориентирующих обучаемых на будущую профессиональную деятельность.

В первой части приводятся основные понятия прикладной статистики, объясняются принципы проверки статистических критериев на наглядных примерах по обработке статистического материала с целью проверки выдвинутых предположений. В конце данной части предлагаются задачи для самостоятельного решения.

Материал, изложенный в Части I, представляет собой основу, освоение которой необходимо для понимания материала из Части II, в частности различных видов статистического анализа: корреляционного и регрессионного, факторного и дисперсионного, а также кластерного анализа.

Книга рассчитана на обучающихся по направлению: «Биотехнические системы и технологии», но может быть полезна студентам таких направлений, как «Бизнес-информатика», «Прикладная математика», «Прикладная математика и информатика», а также всем, кто интересуется анализом данных.

В пособии содержатся материалы, полученные при поддержке гранта РФФИ № 20-08-01005.

Авторы выражают благодарность редактору В.Т. Дубровину и рецензентам Ф.Г. Габбасову и Р.А. Сабитову.

1. ВЫБОРКА И ЕЁ ПРЕДСТАВЛЕНИЕ

Генеральной совокупностью *называют множество всех мысленно возможных результатов наблюдений изучаемых признаков, сформированных задачами исследования, проводимых в одинаковых условиях.*

Она может быть *конечной*: в качестве примера конечной выборки можно привести пациентов, прикреплённых к поликлинике, – или *бесконечной*, например измерение температуры больного в течение конкретного интервала времени.

Также генеральная совокупность может быть *достижимой* или *недостижимой*. Например, необходимо исследование, по результатам которого следует ответить на вопрос: происходит ли при сахарном диабете повышение уровня фибриногена в крови? При такой постановке вопроса, генеральной совокупностью можно считать уровень фибриногена в крови у всех больных сахарным диабетом на земле в какой-то определённый момент времени. Однако если будет исследоваться годовой показатель летальности среди больных, поступивших в какое-либо эндокринологическое отделение, то генеральной совокупностью будет являться показатель исхода заболевания среди пациентов, поступивших в указанное отделение в течение изучаемого года.

Как правило, невозможно получить полное описание генеральной совокупности, а именно значение всех её элементов с соответствующим числом вхождений, поэтому исследователи вынуждены работать с её частью, полученной при проведении конечного числа серий экспериментов.

Итак, *любое подмножество объектов генеральной совокупности, полученное исследователем и призванное с большей или меньшей точностью охарактеризовать генеральную совокупность, называется* ***выборкой****, или* ***выборочной совокупностью***.

Например, группа больных пневмонией, отобранная для участия в клиническом испытании, будет отражать генеральную совокупность настолько, насколько больные в данном отделении соответствуют всем пациентам с данным диагнозом. Фактически выборка является математической моделью последовательности одинаковых опытов со случайными исходами, проводимых в идентичных условиях, при этом результаты опытов статистически независимы.

Для корректного формирования представления о генеральной совокупности выборка должна отражать её основные характеристики, то есть быть *репрезентативной*.

Одной из задач статистического анализа данных является *исследование генеральной совокупности и принятие решений на основе изучения выборочных данных*. Д. Шпигельхалтер в своей книге приводит следующий пример: он описывает получение статистического обоснования обвинительного заключения бристольского расследования о ненадлежащем исполнении профессиональных обязанностей двумя хирургами и руководителем отделения между 1984–1995 годами, основанного на сравнении количества летальных исходов в Королевской больнице Бристоля с данными в других больницах Соединённого Королевства [16].

Введём следующие обозначения:

• Пусть X – произвольная случайная величина с функцией распределения $F(x) = P(X < x), x \in R$ (R – множество действительных чисел), реализация которой наблюдается в ходе эксперимента. Распределение этой случайной величины неизвестно или известно, но с точностью до параметров этого распределения.

• Совокупность $\{X_k, k = 1, \dots., n\}$ независимых случайных величин, имеющих одинаковые функции распределения: $F_{X_k}(x) = F(x)$, – называется *однородной выборкой объёма n, соответствующей функции распределения $F(x)$*.

• Соответственно, выборка $\{X_k, k = 1, \dots., n\}$ будет называться *неоднородной*, если законы распределения $F_{X_k}(x)$ её элементов неодинаковы. Далее, по умолчанию, будем предполагать, что выборка X_k является *однородной*.

• Выборка $\{X_k, k = 1, \dots, n\}$ из генеральной совокупности X называется *репрезентативной* (*представительной*), если:

1) случайные величины $X_1, \dots, X_n$ имеют то же распределение, что и случайная величина X;

2) случайные величины $X_1, \dots, X_n$ независимы.

• *Реализацией выборки* X_k называется вектор $X_k = \{x_1, \dots, x_n\}$, компоненты которого являются реализациями соответствующих элементов выборки.

• Значение $x_i, i = 1, \dots, n$, называется *вариантой*, или *вариацией*.

• Количество элементов n в выборке называется *объёмом выборки*.

• Разница между максимальным $x_{\max}$ и минимальным $x_{\min}$ значениями выборки называют *размахом выборки*.

• Выборка, записанная в порядке возрастания её отдельных значений: $x_1 \le x_2 \le x_3 \cdots \le x_n$, – называется *вариационным рядом*.

• Далее, пусть дана выборка объёма n, всего k различных значений: $x_1 \ne x_2 \ne x_3 \ne \cdots \ne x_k$. При этом значение x_1 встречается m_1 раз, значение $x_2 - m_2$ раз и т. д., и $m_1 + m_2 + \cdots + m_k = n$. Числа m_1, $m_2, \cdots, m_k$ называют *частотами*. Частоты, нормированные на объём выборки: $\frac{m_i}{n} = \mu_i$, – называют *относительными частотами*.

• Последовательность всех различных значений выборки x_i вместе с их частотами m_i называют *статистическим рядом*.

Значения варианта	x_1	x_2	$\cdots$	x_k
Частота	m_1	m_2		m_k

• Варианты вместе со своими относительными частотами называют *статистическим распределением*.

Значения варианта	x_1	x_2	$\cdots$	x_k
Относительная частота	μ_1	μ_2		μ_k

• Для удобства хранения большого количества данных, а также при работе с выборками из непрерывных генеральных совокупностей используют *интервальное представление данных*. В этом случае диапазон из n значений выборки $[x_{\min}, x_{\max}]$ разбиваем на непересекающиеся интервалы. При этом желательно, чтобы данные не попадали на границы интервалов. Это можно сделать, задав границы интервала с большей точностью, чем значения выборки. Затем подсчитывается число вхождений вариантов выборки в каждый из интервалов. Если значение выборки всё же попало на границу, то будем считать его в равной мере принадлежащим обоим интервалам, тогда к частотам вхождения в эти соседние интервалы будем добавлять по 0,5. При сравнительно равномерном распределении данных в покрывающем диапазон интервале длина интервалов может быть одинаковой – в этом случае их рекомендуется брать как $k = 2\ln n$.

Эмпирическая функция распределения

Определение. *Эмпирической функцией распределения* называется функция вида: $\overline{F_n}(x) = \frac{m_x}{n}$, где m_x – накопленная частота на интервале $(-\infty; x)$. Поскольку m_x – количество выборочных данных меньше x, её удобнее строить по вариационному ряду.

• *Свойства эмпирической функции распределения:*

1. *Значение эмпирической функции распределения* принадлежат отрезку [0,1].

2. $\overline{F_n}(x)$ – неубывающая функция.

3. Если x_1 является наименьшей вариантой, то $\overline{F_n}(x) = 0$, при всех $x \le x_1$. Если же x_n является наибольшей вариантой, то $\overline{F_n}(x) = 1$, при $x > x_n$.

Сама *эмпирическая функция* запишется следующим образом:

$$\overline{F_n}(x) = \begin{cases} 0, \text{ если } x \le x_1 \\ \sum_{i<j} \mu_i, \text{ если } x_j < x \le x_{j+1}. \\ 1, \text{ если } x > x_k \end{cases}$$

Здесь k – количество неповторяющихся значений выборки, $j = 1, \dots, k-1$.

• *Графиком эмпирической функции распределения* $\overline{F_n}(x)$ *является ступенчатая линия, скачкообразно изменяющаяся в точках* $x_1, \dots, x_k$, *с соответствующими величинами скачков, равными* $\mu_1, \mu_2, \dots, \mu_k$.

Пример 1. В течение 10 дней фиксировалось количество обратившихся для вакцинации от гриппа людей. В результате получилась выборка объёма $n = 10$: 6, 8, 5, 15, 8, 6, 1, 5, 3, 8. Построим эмпирическую функцию распределения.

1. Запишем вариационный ряд: 1, 3, 5, 5, 6, 6, 8, 8, 8, 15.
2. Построим статистическое распределение.

Значение варианта	1	3	5	6	8	15
Относительная частота	0,1	0,1	0,2	0,2	0,3	0,1

3. Запишем функцию распределения:

$$\begin{cases} \overline{F}_n(x) = 0, \quad \text{если } x \le 1 \\ \overline{F}_n(x) = 0.1, \ \text{если } 1 < x \le 3 \\ \overline{F}_n(x) = 0.2, \ \text{если } 3 < x \le 5 \\ \overline{F}_n(x) = 0.4, \ \text{если } 5 < x \le 6 \\ \overline{F}_n(x) = 0.6, \ \text{если } 6 < x \le 8 \\ \overline{F}_n(x) = 0.9, \ \text{если } 8 < x \le 15 \\ \overline{F}_n(x) = 1, \ \text{если } x > 15 \end{cases}.$$

4. Построим эмпирическую функцию распределения (рис. 1).

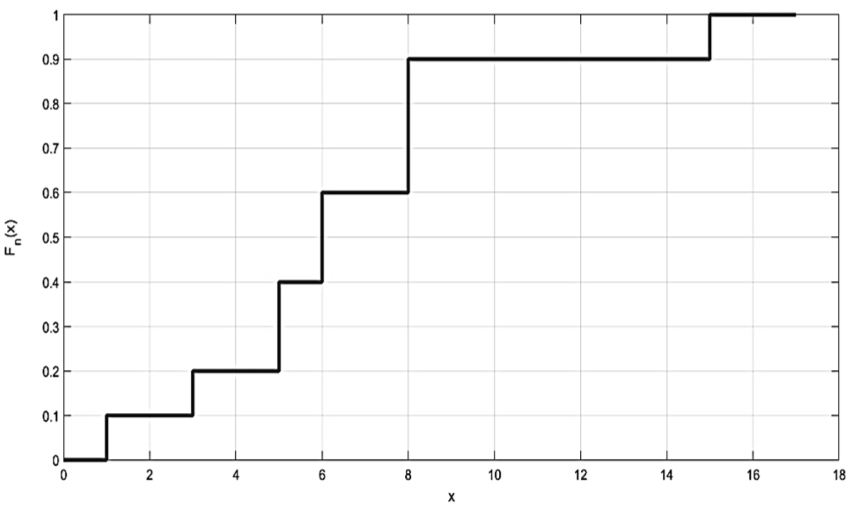

Рисунок 1 – Эмпирическая функция распределения

Графическое представление данных

Графическое представление является наиболее наглядным отображением данных, позволяющим произвести первичный анализ и предположить возможные закономерности.

В статистике, помимо привычных видов диаграмм, используют такие графические объекты, как *полигон частот* и *гистограмма*.

• *Полигон частот* или *относительных частот* представляет собой ломаную линию, соединяющую точки на декартовой координатной плоскости (x_i, m_i) или (x_i, ω_i), $i = 1, \dots, k$ (рис. 2).

В случае интервального представления данных под $\boldsymbol{x_i}$ подразумевают середину интервала как типичного (среднего) представителя.

Для примера 1 представим полигон относительных частот.

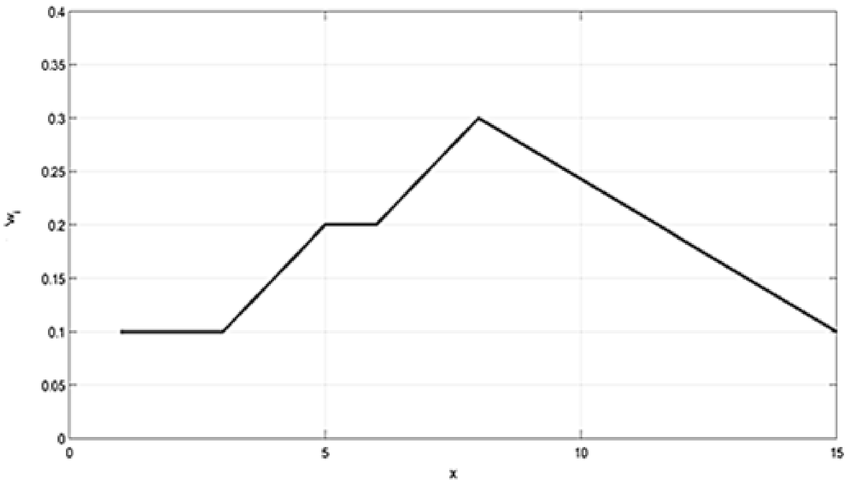

Рисунок 2 – Полигон относительных частот

• *Гистограмма* строится по интервальному представлению данных. Интервалы откладываются по оси x декартовой системы координат, далее на каждом интервале восстанавливается прямоугольник так, чтобы его площадь была равна: m_i – в случае гистограммы частот и μ_i – для гистограммы относительных частот. Таким образом, высота i-го прямоугольника откладывается по оси y и будет равна $\frac{m_i}{h_i}$ или же $\frac{\mu_i}{h_i}$ соответственно, здесь h_i – длина i-го интервала.

Пример 2. Интервальное представление данных.

(2,6)	(6,10)	(10,14)	(14,18)
6	10	4	5

1. Рассчитаем высоту. Объём выборки равен: 6 + 10 + 4 + 5 = 25, длина каждого интервала равна 4. Высота первого столбика: 6 / (4 × 25) = 0.06; аналогично для остальных столбиков: 0.1, 0.04, 0.05. Гистограмма относительных частот примет нижеследующий вид (рис. 3).

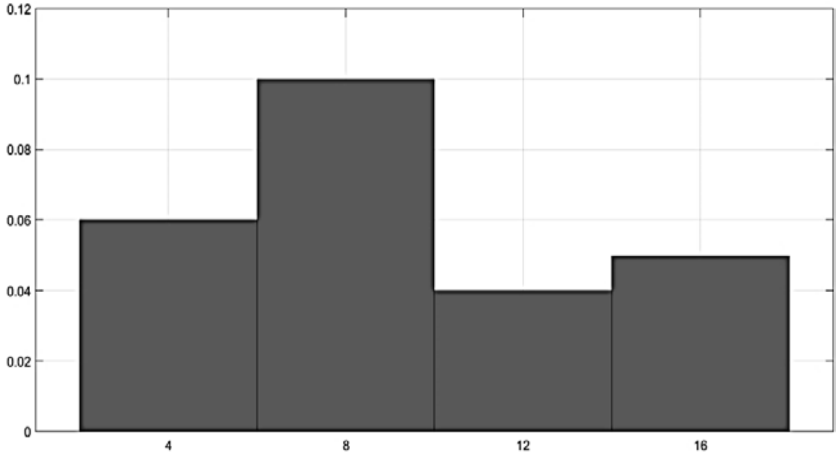

Рисунок 3 – Гистограмма относительных частот

Замечание! Эмпирическая функция распределения представляет собой *оценку функции распределения генеральной совокупности*, гистограмма относительных частот – *оценку функции плотности*.

2. ОЦЕНКИ, СДЕЛАННЫЕ ПО ВЫБОРКЕ, И ИХ СВОЙСТВА

Выборка *– это подмножество (часть) генеральной совокупности*, поэтому параметры распределения или числовые характеристики, подсчитанные по выборке, являются «истинными» и соответствуют генеральной совокупности всего лишь с некоторой точностью. В результате в математической статистике речь идёт об оценках как параметров распределения, так и числовых характеристик.

При оценке параметров используются два подхода: *точечное* и *интервальное оценивание*.

При *точечном оценивании* по результатам испытаний находят число, которое принимают в качестве приближенного значения оцениваемого параметра. Его называют *оценкой параметра*. При этом *оценка случайного параметра* является функцией от результатов испытаний: $\hat{\theta} = S(X_1, X_2, \cdots, X_n)$. Следовательно, ку $\hat{\theta}$ можно считать случайной величиной, с присущими ей законом распределения и числовыми характеристиками.

Пусть $\theta \in \Theta \subseteq \mathcal{R}^1$ – некоторая детерминированная или случайная величина (параметр), а $z_n = \{X_k, k = 1, \dots, n\}$ – выборка.

Определение. *Точечной оценкой параметра θ по выборке z_n* называется любая статистика $\hat{\theta}_n = \varphi_n(z_n)$, принимающая значения из множества θ.

• Выделяют такие свойства оценок параметров:

1. *Несмещённость оценки* – означает, что математическое ожидание ошибок (разницы между оценкой параметра и его истинным значением) равно нулю.

Определение. Величина $\Delta\hat{\theta}_n = \hat{\theta}_n - \theta$ называется *ошибкой оценки $\hat{\theta}_n$.*

Определение. Оценка $\hat{\theta}_n$ называется *несмещённой*, если $M\{\Delta\hat{\theta}_n\} = 0$. Если же $M\{\Delta\hat{\theta}_n\} \neq 0$, но $M\{\Delta\hat{\theta}_n\} \to 0$, при $n \to \infty$, то $\hat{\theta}_n$ называется *асимптотически несмещённой*.

2. Оценки считаются *эффективными*, если они обладают наименьшими дисперсиями.

3. *Состоятельность оценок* характеризует увеличение их точности с увеличением объёма выборки.

Определение. Оценка $\hat{\theta}(X_1, X_2, \cdots, X_n)$ параметра θ называется *состоятельной*, если для всех $\theta \in \Theta$ последовательность $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \cdots, X_n) \xrightarrow{P} \theta$, при $n \to \infty$ ($\xrightarrow{P}$ означает сходимость по вероятности, то есть для $\forall\varepsilon > 0$, $P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \to 0$, при $n \to \infty$).

Понятие «состоятельность оценок» связано только с предельными свойствами последовательности случайной величины. Состоятельность напрямую не связана со свойством оценки при фиксированном объёме выборки и на практике применяется с осторожностью.

Определение. Выборочное математическое ожидание: $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ – называют также *выборочным средним*.

Определение. Оценка $\bar{X}$, составленная по репрезентативной выборке $X_1, X_2, \cdots, X_n$ генеральной совокупности случайной величины X, является *несмещённой оценкой* $M(X)$ (*математического ожидания генеральной совокупности*).

• Найдём математическое ожидание данной оценки:

$$M(\bar{X}) = M\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}M\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\left(\sum_{k=1}^{n} M(X_k)\right).$$

Все $X_1, X_2, \cdots, X_n$ распределены одинаково с генеральной совокупностью X, следовательно:

$$M(\bar{X}) = \frac{1}{n}\left(\sum_{k=1}^{n} M(X_k)\right) = \frac{1}{n}\left(\sum_{k=1}^{n} M(X)\right) = \frac{1}{n}nM(X) = M(X).$$

Согласно определению, $\bar{X}$ будет *несмещённой оценкой*.

• *Проверка на состоятельность*. Для этого вычислим дисперсию данной оценки:

$$D(\bar{X}) = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2} D\left(\sum_{k=1}^{n} X_k\right).$$

Так как $X_1,\ X_2, \cdots, X_n$ – независимые случайные величины, то:

$$D(\bar{X}) = \frac{1}{n^2}\left(\sum_{k=1}^{n} D(X_k)\right) = \frac{1}{n^2}\left(\sum_{k=1}^{n} D(X)\right) = \frac{1}{n^2} nD(X) = \frac{D(X)}{n} \to 0,$$

при $n \to 0$, следовательно, $\bar{X}$ – состоятельная оценка.

• *Проверка на эффективность* при произвольном распределении не выполняется, так как эффективность оценок зависит от закона распределения генеральной совокупности.

В качестве примера представляется интересным проверить на смещённость оценку дисперсии вида:

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

Для дисперсии справедливы следующие равенства:

$$D(X) = M(X - M(X))^2$$

и

$$D(X) = M(X^2) - M(X)^2.$$

Следовательно,

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

Теперь найдём математическое ожидание данной оценки дисперсии генеральной совокупности:

$$M(\overline{S^2}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}({X_i}^2) - \frac{1}{n^2} M\left(\sum_{i=1}^{n} X_i\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}({X_i}^2) - \frac{1}{n^2} M\left(\sum_{i=1}^{n} {X_i}^2 + \sum_{i \neq j} X_i X_j\right),$$

здесь i и j – от 1 до n.

Согласно свойству линейности математического ожидания и определению репрезентативной выборки, получим:

$$M\left({X_i}^2\right) = M(X^2), M(X_i) = M(X) \text{ и } M\left(X_i X_j\right) = M(X_i)M\left(X_j\right).$$

Тогда:

$$M\left(\overline{S^2}\right) = \frac{1}{n}\sum_{i=1}^{n} M(X^2) - \frac{1}{n^2}\sum_{i=1}^{n} M(X^2) - \frac{1}{n^2}\sum_{i\neq j} M(X)M(X) = \frac{1}{n} n\, M(X^2) - \frac{1}{n^2} nM(X^2) - \frac{1}{n^2} n(n-1)M(X)^2 = \frac{n-1}{n} M(X^2) -$$

$$-\frac{n-1}{n} M(X)^2 = \frac{n-1}{n}\left(M(X^2) - M(X)^2\right).$$

Таким образом, $M\left(\overline{S^2}\right) = \frac{n-1}{n} D(X)$, и $\overline{S^2}$ – будет смещённой.

• За *исправленную выборочную дисперсию* принимают:

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

Также при анализе выборок рассматривают:

• *моду* – значение признака, который чаще всего встречается в исследуемом наборе данных.

• Для интервального представления данных за *моду* берут:

$$M_0 = x_0 + h_i \frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)},$$

где x_0 – нижняя граница модального интервала (интервала, имеющего наибольшую частоту); h_i – величина модального интервала; f_1 – частота предмодального интервала; f_2 – частота модального интервала; f_3 – частота послемодального интервала.

Пример 3. Приведены результаты измерения роста (см) случайно отобранных 100 студентов.

Рост	154,1–158,1	158,1–162,1	162,1–166,1	166,1–170,1
Число студентов	10	14	26	28

Рост	170,1–174,1	174,1–178,1	178,1–182,1
Число студентов	12	8	2

Мода будет равна:

$$M_0 = 166{,}1 + (170{,}1 - 166{,}1)\frac{28-26}{(28-26)+(28-12)} \approx 166{,}54.$$

Также моду можно нанести графически (рис. 4).

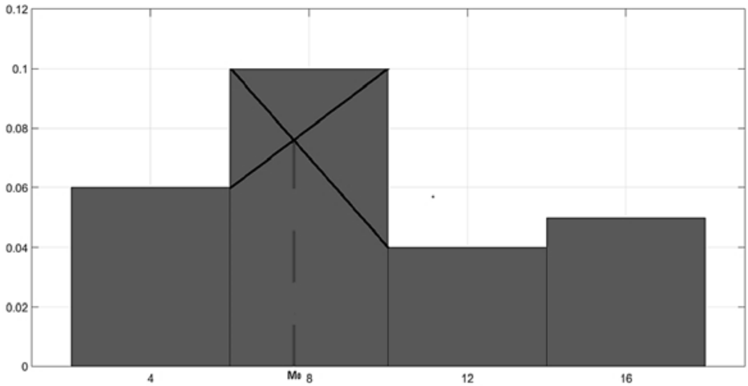

Рисунок 4 – Графическое определение моды

Стоит учитывать, что распределение может не иметь моды. Также мода может не быть типичной характеристикой, если в распределении есть значения, близкие к модальному, но противоположные ему по смыслу.

Пример 4. Представлена успеваемость студентов в группе по предмету.

Оценка	Количество студентов
«Неудовлетворительно»	8
«Удовлетворительно»	5
«Хорошо»	6
«Отлично»	7

Здесь некорректно утверждать, что большинство студентов неуспевающие.

Определение. Пусть F – непрерывное распределение на $\mathcal{R}$ и $\alpha\in(0{,}1)$. Число x_α, являющееся решением уравнения: $F(x) = P\{\xi < x_\alpha\} = \alpha$, – называется α -*квантилем распределения F*.

Определение. *Медианой* называется квантиль уровня ½. Для медианы характерно*:*

$$P\{\xi < x_{½}\} = P\{\xi \geq x_{½}\}.$$

При нахождении медианы по выборке:

1) строят вариационный ряд:

$$x_1 \leq x_2 \leq x_3 \cdots \leq x_n.$$

2) за медиану принимают:

$$x_{\text{med}} = \begin{cases} x_{n+½}, \text{ если } n - \text{нечётное} \\ \frac{x_{n+½}+x_n}{2}, \text{ если } n - \text{чётное} \end{cases}.$$

Так, при построении прогнозов по развитию болезни пользуются таким определением, как ***медиана общей выживаемости***: это *тот период времени от момента начала исследования, который переживает только половина пациентов с определённым диагнозом.*

• При *интервальном оценивании* определяется интервал, который с заданной вероятностью накрывает истинное значение оцениваемого параметра. *Границы интервалов также являются функцией от результатов испытаний:*

$$S_1(X_1, X_2, \cdots, X_n) \leq \theta \leq S_2(X_1, X_2, \cdots, X_n),$$

здесь θ – истинное значение параметра.

В математической статистике рассматриваемые интервалы принято называть *доверительными*, а вероятность, с которой они покрывают истинное значение параметра, называют *доверительной вероятностью*.

Определение. Под *доверительным интервалом* понимается случайный интервал, который с некоторой вероятностью α накрывает истинное значение искомого параметра:

$$\alpha = P(|\theta_x - \theta| < \gamma).$$

Здесь: α – доверительная вероятность; γ – точность определения неизвестного параметра θ с помощью оценки θ_x.

$$\alpha = P(\theta_x - \gamma < \theta < \theta_x + \gamma).$$

Пусть $\alpha = 0{,}5$. Это означает, что для половины повторений эксперимента доверительный интервал покроет истинный параметр θ, а половину построенных интервалов не захватит (рис. 5).

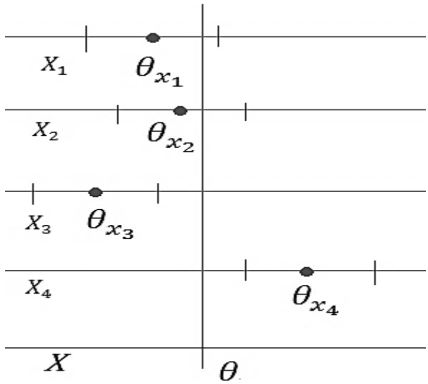

Рисунок 5 – Доверительное оценивание параметра

В целом *доверительный интервал* может быть как *симметричным* относительного оцениваемого параметра, так и *несимметричным*.

Определение. Доверительный интервал вида:

$$\alpha = \mathrm{P}(\theta_x - \gamma_1 < \theta < \theta_x + \gamma_2),$$

при котором $\mathrm{P}(x < \gamma_1) = \mathrm{P}(x \geq \gamma_2) = \frac{1-\alpha}{2}$, или через функцию плотности:

$$\int_{-\infty}^{\gamma_1} f(x)dx = \int_{\gamma_2}^{+\infty} f(x)dx,$$

называют *центральным*.

В случае симметричного распределения при фиксированной вероятности центральный и симметричный доверительные интервалы совпадают.

В качестве примера доверительного оценивания рассмотрим доверительный интервал (с надёжностью γ) математического ожидания $M(X)$ нормально распределённого количественного признака X по выборочной средней $\bar{X}$:

1) при известном среднем квадратическом отклонении σ генеральной совокупности:

$$\bar{X} - t\frac{\sigma}{\sqrt{n}} < M(X) < \bar{X} + t\frac{\sigma}{\sqrt{n}}..$$

Здесь: $t\frac{\sigma}{\sqrt{n}} = \delta$ – точность оценки; t – значение аргумента функции Лапласа, при котором $\Phi_0(t) = \frac{\gamma}{2}$.

Функция Лапласа связана с функцией нормального распределения математическим ожиданием, равным нулю, и дисперсией, равной единице, следующим соотношением:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{t^2}{2}}dt + \frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-\frac{t^2}{2}}dt =$$

$$= \frac{1}{\sqrt{2\pi}} \times \sqrt{\frac{\pi}{2}} + \Phi_0(x) = 0{,}5 + \Phi_0(x);$$

2) если среднеквадратическое отклонение σ генеральной совокупности неизвестно и объём выборки достаточно большой.

В этом случае вместо характеристики σ подставляют выборочную оценку S, равную квадратному корню из выборочной дисперсии, полученную по результатам n измерений, что при большом объёме выборки вполне допустимо, однако границы интервала получаются приближенными:

$$\bar{X} - t\frac{S}{\sqrt{n}} < M(X) < \bar{X} + t\frac{S}{\sqrt{n}};$$

3) при неизвестном среднем квадратическом отклонении σ генеральной совокупности и малом объёме выборки ($n < 30$).

В случае малой выборки используется *нормированная разность*:

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

Данная статистика имеет распределение Стьюдента, которое определяется *функцией плотности*:

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}.$$

Здесь: $\Gamma(\blacksquare)$ – гамма-функция; $k = n - 1$ – называется *степенью свободы*.

В силу чётности данной функции плотности:

$$\alpha = P(|T| < t) = 2\int_0^{t_{\alpha,k}} f(t)dt = P\left(-t_{\alpha,k} < T < t_{\alpha,k}\right),$$

доверительный интервал будет равным:

$$\bar{X} - t_{\alpha,k}\frac{S}{\sqrt{n}} < M(X) < \bar{X} + t_{\alpha,k}\frac{S}{\sqrt{n}}.$$

Пример 5. Исследуется состояние дыхательных путей курильщиков. В качестве характеристики используется показатель функции внешнего дыхания – «максимальная объёмная скорость середины выдоха (л/с)». Требуется найти интервальную оценку среднего значения этого показателя с надёжностью $\gamma = 0{,}95$. В обследуемой группе из 20 случайным образом отобранных курильщиков среднее значение этого показателя осталось равным 2,2 л/с. Выборку из 20 курильщиков считаем представительной.

Доверительный интервал примет вид:

$$2{,}2 - 2{,}09 \times \frac{0{,}73}{\sqrt{20}} < M(X) < 2{,}2 + 2{,}09 \times \frac{0{,}73}{\sqrt{20}}.$$

Статистика Стьюдента $t_{\alpha,k}$ при $\alpha = 0{,}95$ и k = 19 будет приблизительно равна 2,09 и находиться либо с помощью таблиц, либо с помощью обратной функции распределения Стьюдента в статистических пакетах.

Пример 6. На ферме испытывалось влияние витаминов на прибавку в весе телят. Для этой цели было осмотрено 20 телят одного возраста. Средний вес их оказался равным 340 кг, а среднеквадратическое отклонение – 15 кг. Найти доверительный интервал для генеральной средней с надёжностью 0,95.

Как и в предыдущем примере, статистика Стьюдента $t_{\alpha,k}$ при $\alpha = 0{,}95$ и k = 19 будет приблизительно равна 2,09. В данном случае доверительный интервал примет вид:

$$340 - 2.09\frac{15}{\sqrt{20}} < M(X) < 340 + 2.09\frac{15}{\sqrt{20}}.$$

3. ПРОВЕРКА СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

Задача практической статистики состоит в проверке предположений, выдвинутых на основании эмпирических данных. При этом сформулированное предположение будет является *основной гипотезой* H_0. Данную гипотезу H_0 можно принять или отвергнуть, но тогда будет принято обратное утверждение, называемое *альтернативной гипотезой* H_1.

Пример 7. Медицинский препарат перед выпуском в продажу исследуют на токсичность биологическими методами: определённую дозу препарата вводят подопытным животным (мышам, кроликам) и в зависимости от количества летальных исходов (ξ) делают выводы о токсичности препарата.

Итак, пусть за H_0 принимают утверждение «препарат токсичен», а за H_1 – «препарат не токсичен».

Возможны следующие варианты:

1. Гипотеза H_0 верна, но, согласно критерию, гипотезу H_0 отклонили – и, таким образом, опасный для здоровья препарат отправлен в реализацию. Последствием ошибки в данном случае является смерть пациентов.

2. Гипотеза H_0 неверна, но, согласно критерию, её приняли. Последствия такой ошибки – финансовые потери поставщика, которому вернули препарат.

Как правило, в качестве проверяемой гипотезы выбирают ту, для которой важнее избежать ошибки, заключающейся в отклонении верной гипотезы H_0.

Определения.

• Ошибка, состоящая в отклонении верной гипотезы H_0, называется *ошибкой первого рода*; ошибка, состоящая в неотклонении неверной гипотезы H_0, называется *ошибкой второго рода* (см. таблицу ниже).

Результат проверки H_0	Возможные состояния проверки	
	Верна H_0	Верна H_1
Гипотеза H_0 отклоняется	Ошибка первого рода	Правильное решение
Гипотеза H_0 не отклоняется	Правильное решение	Ошибка второго рода

• Вероятность β отклонения основной (нулевой) гипотезы, когда верна альтернативная гипотеза H_1, называют *мощностью критерия*.

• Свойство критерия отклонять H_0, когда H_0 неверна, называют *чувствительностью критерия*.

• Число α, ограничивающее сверху вероятность ошибки первого рода, называют *уровнем значимости*.

Замечание. При этом, уменьшая уровень значимости α, то есть вероятность ошибки первого рода, уменьшается чувствительность критерия.

Наглядным примером может служить задача о диагностике туберкулёза [14].

Пример 8. Если пациент страдает туберкулёзом, то вероятность того, что отдельный рентгеновский анализ положителен, равна: $\rho_0 = 0{,}8$. Если пациент не несёт никаких признаков туберкулёза, то вероятность положительного анализа равна: $\rho_1 = 0{,}01$.

Введём следующие предположения, упрощающие решение:

1. Для всех пациентов, страдающих туберкулёзом, вероятность ρ_0 одинакова.

2. Результаты анализа снимков одного пациента независимы друг от друга.

Предположим, что по n, равном трём независимым рентгеновским снимкам, необходимо вынести заключение: или «человек болен туберкулёзом», или «человек здоров».

При выборе основной гипотезы считается, что ошибка признать больного человека здоровым приводит к большим потерям как для самого больного (потерей возможности диагностировать и вылечить болезнь на ранней стадии), так и для окружения (большой вероятностью заражения от больного). Поэтому H_0 – «человек болен» и H_1 – «человек здоров». Гипотезы относительного здоровья пациента – это гипотезы о распределении случайной величины ξ.

При верной гипотезе H_0 вероятность обнаружения можно выразить следующим распределением ξ:

$$P_0(k) = C_3^k p_0^k (1 - p_0)^{3-k};$$

а при верной гипотезе H_1:

$$P_1(k) = C_3^k p_1^k (1 - p_1)^{3-k}.$$

Здесь $k = 0,\ 1,\ 2,\ 3.$

Интуитивно при малых значениях $\xi \leq l$ распределение будет $P_1(k)$. Если $\xi > l$, то $P_0(k)$. Здесь l – некоторое положительное число.

Вероятность ошибки первого рода в данной интерпретации будет:

$$P(\xi \leq l | H_0) < \alpha.$$

Идеально, чтобы вероятность ошибки была равна нулю, но в результате пациент объявляется больным в любом случае ($p_0 = 1$), так как $\sum_{k=0}^{l} C_3^k p_0^k (1 - p_0)^{3-k} \leq \alpha = 0$. Поэтому выбираем $\alpha \neq 0$.

Пусть α = 0.01, тогда $l = 0$ – наибольшее натуральное число, для которого $\sum_{k=0}^{l} C_3^k p_0^k (1 - p_0)^{3-k} \leq 0.01$. Действительно, при $l = 0$ будет:

$$\sum_{k=0}^{l} C_3^k p_0^k (1 - p_0)^{3-k} = 0.008,$$

а при $l = 1$:

$$\sum_{k=0}^{l} C_3^k p_0^k (1 - p_0)^{3-k} = 0.008 + 3 \times 0.8 \times 0.04 \approx 0.1.$$

Таким образом, если имеется хоть один положительный исход, то принимаем H_0; если все отрицательные, то H_0 отвергаем. Вероятность ошибки первого рода $P(\xi \leq l|H_0)$ будет:

$$C_3^0 p_0^0 (1 - p_0)^3 = 0.008 \approx 0.01.$$

Это означает, что в среднем 99 из 100 больных туберкулёзом будет ставиться положительный диагноз.

Вероятность ошибки второго рода:

$$1 - P(\xi \leq 0|H_1) = 1 - C_3^0 p_1^0 (1 - p_1)^3 = 0.0297.$$

Это означает, что из 100 здоровых людей в среднем трое будут классифицированы как больные.

Мощность критерия:

$$P(\xi \leq 0|H_1) = C_3^0 p_1^0 (1 - p_1)^3 \approx 0.97.$$

Следовательно, из 100 здоровых людей в среднем 97 будут классифицироваться как здоровые.

• Итак, пусть V_0 – это множество всех возможных результатов выборок Z_n в предположении, что H_0 верно, тогда $Sp \in V_0$ будет называться *критической областью*, если из того, что $Z_n \in Sp$, следует, что верную гипотезу H_0 отвергаем, и $P_o(Sp) = P(Z_n \in Sp|H_0 - \text{верна}) = \alpha_p \leq \alpha$.

• Соответственно, множество $\overline{Sp} = V_0 \backslash Sp$ называется *областью принятия гипотезы* H_0 (рис. 6).

• *Статистическим критерием* называется алгоритм проверки гипотезы H_0 по выборке $Z^{(k)} = (Z_1, \dots, Z_k)$, где $Z_i = (z_1, \dots, z_{n_i})$, n_i – объём i-й выборки, k = 1, 2, …

• *Статистикой критерия* будем называть некоторую числовую функцию $f(Z^{(k)})$ по выборке $Z^{(k)}$, обладающую тем свойством, что её закон распределения полностью известен, если гипотеза H_0

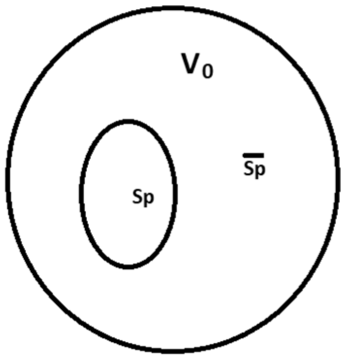

Рисунок 6 – Критическая и доверительная области в предположении, что гипотеза H_0 верна

верна. В этом случае обозначим функцию распределения статистики как $\boldsymbol{F}$.

Пусть статистика переводит критическую область Sp в область на числовой прямой $\Delta_p \in \mathrm{R}^1$ таким образом:

$$Z_n \in Sp <=> f(Z^{(k)}) \in \Delta_p,$$

$$P_o(Sp) = P\big(f(Z^{(k)}) \in \Delta_p \big| H_0 - \text{верна}\big) = \alpha_p \leq \alpha.$$

• Точки, отвечающие за границы отрезка Δ_p, называются *критическими точками*.

Возможны следующие расположения Δ_p на числовой прямой:

1. Левосторонняя критическая область (левосторонняя альтернатива; см. рис. 7).

В этом случае критическая область располагается слева от области принятия гипотезы. Для принятия альтернативы достаточно рассмотреть один из двух случаев:

– статистика, полученная по эмпирическим данным $C_p = f(Z^{(k)})$, меньше статистики C, соответствующей принятому уровню значимости и являющейся решением уравнения $F(C) = \alpha$;

– α_p, рассчитанная как функция распределения от эмпирической статистики C_p ($\alpha_p = F(C_p)$), меньше принятого уровня значимости α. Здесь $F(\blacksquare)$ – функция распределения статистики критерия.

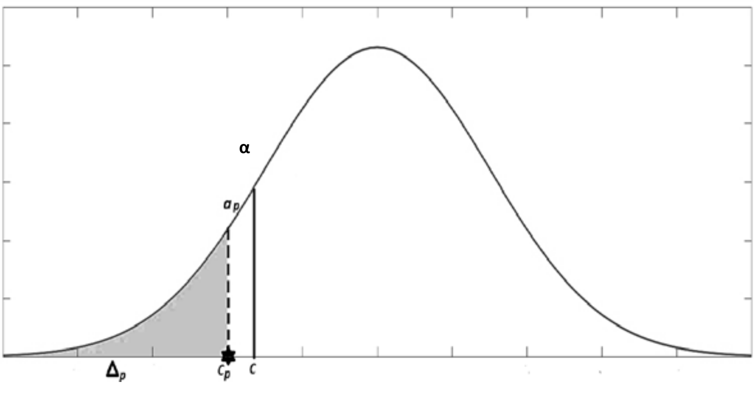

Рисунок 7 – Левосторонняя критическая область

2. Правосторонняя критическая область (правосторонняя альтернатива; см. рис. 8).

В этом случае критическая область располагается справа от области принятия гипотезы, и альтернатива будет приниматься, если $C_p > C$, $\alpha_p = 1 - F\left(C_p\right) < \alpha$.

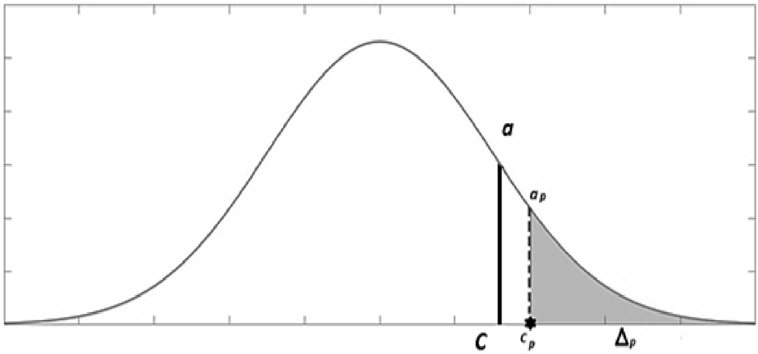

Рисунок 8 – Правосторонняя критическая область

3. Двухсторонняя критическая область (двухсторонняя альтернатива; см. рис. 9).

В этом случае критическая область располагается слева и справа от области принятия гипотезы таким образом:

$$P\left(C_{p1} < C_1\right) = \alpha_{p1}, P\left(C_{p2} < C_2\right) = \alpha_{p2},$$

и

$$\alpha_{p1} + \alpha_{p2} \leq \alpha_1 + \alpha_2 = \alpha.$$

В случае симметричного распределения альтернатива принимается, если $\left|C_p\right| < |C|$, где $F(C) = {}^{\alpha}/_{2}$.

В случае несимметричного распределения желательно рассмотреть отдельно попадание в каждую часть области. Так, для распределения Фишера со степенями свободы, равными 4, при ${}^{\alpha}/_{2} = 0.1$, попадание в левую часть критической области будет означать, что $|C| < C_{p1} \approx 0.2434719365631875$, а в правую – $|C| > C_{p2} \approx 4.107249542250520$ (см. рис. 10).

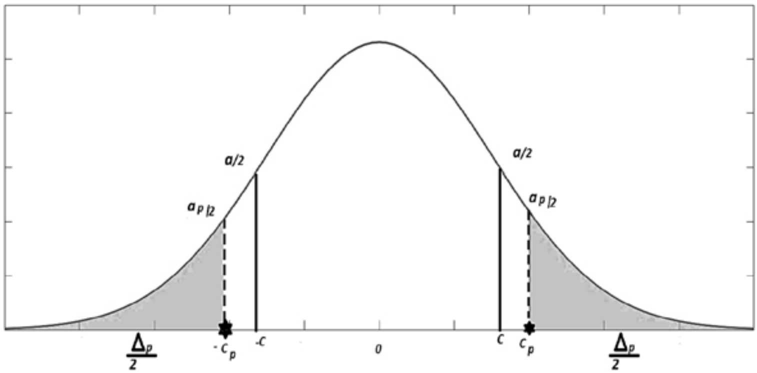

Рисунок 9 – Двухсторонняя критическая область (симметричное распределение статистики)

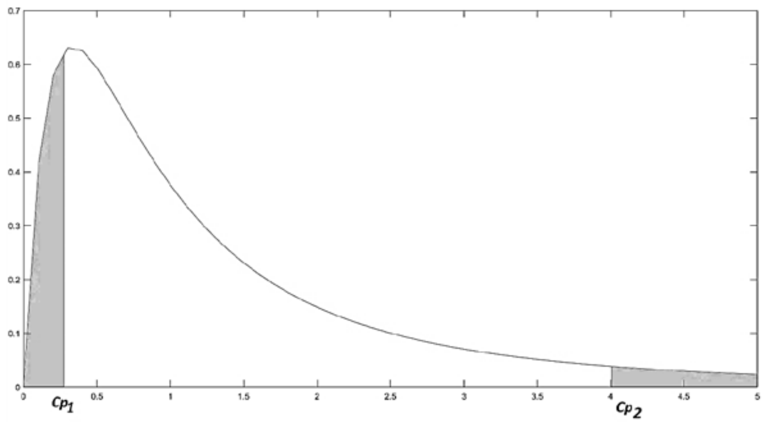

Рисунок 10 – Двухсторонняя критическая область при несимметричном распределении

Чаще всего в медицинских или биологических исследованиях проверяют гипотезы о статистической значимости различий между эмпирическими выборками. Например: повышается ли точность измерительного прибора после приведённой отладки? эффективен ли новый препарат? Обе описанные задачи сводятся к проверке однородности двух выборок. При этом статистические гипотезы, не использующие допущений о конкретном законе распределения, называются *непараметрическими*, в противном случае – *параметрическими*.

В примере 8 экспериментальные выборки могут быть организованы двумя способами:

1. У контрольной группы людей измеряют исследуемый показатель до принятия лекарства и после. В этом случае формируются две связанные выборки, показатели в этих выборках связаны через пациента, у которого они снимались. Такие выборки называются *зависимыми*, или *связанными*. При такой организации эксперимента используются *одновыборочные критерии*.

2. Наличие контрольной (пациенты, не получавшие препарат) и опытной (пациенты, получившие препарат) выборок. Такие выборки называются *несвязанными*, а критерии – *двухвыборочными*.

В качестве примера принятия решения об эффективности препарата рассмотрим ***параметрический критерий Стьюдента о равенстве средних***. Это параметрический метод, используемый для проверки гипотез о достоверности разности средних при анализе количественных данных, позволяет проверить предположение об однородности двух выборок из нормального распределения с одинаковой дисперсией (в примерах, приведённых для иллюстрации критерия Стьюдента, данные условия будем считать выполненными – *их проверку покажем ниже*).

Основная гипотеза формулируются в следующем виде: H_0 – «выборки однородны и их средние равны».

Альтернативная H_1 может быть выбрана как один из трёх возможных вариантов:

1) «среднее первой выборки больше средней второй выборки» (правосторонняя альтернатива);

2) «среднее первой выборки меньше средней второй выборки» (левосторонняя альтернатива);

3) «средние двух выборок не равны» (двухсторонняя альтернатива).

В зависимости от организации выборок *критерий Стьюдента* бывает как *одновыборочным*, так и *двухвыборочным*.

Одновыборочный критерий Стьюдента

Рассмотрим процедуру проведения одновыборочного критерия Стьюдента на конкретном примере, сформулированном нижеследующим образом.

Пример 9. У группы пациентов, страдающих гипертонией, проверили давление до и после приёма исследуемого препарата. Рассмотреть на уровне значимости 0.05 возможность принять или отвергнуть нулевую гипотезу при различных альтернативах.

Данные для этого примера получим следующим образом:

1) сгенерируем одну выборку из нормального распределения со средним 160 и дисперсией, равной 3, а вторую – из нормального распределения со средним 140 и дисперсией, равной 3 (выборка после приёма лекарств);

2) данные в обеих выборках округлены до целого числа;

3) объём нашей группы пациентов равен 40.

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
1	167	135
2	163	142
3	163	141
4	160	139
5	156	143
6	158	137
7	157	141
8	158	141
9	161	141
10	158	137
11	157	140
12	158	138
13	159	137
14	161	136
15	157	143
16	160	137
17	160	139
18	163	136

Окончание таблицы

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
19	159	139
20	159	130
21	156	136
22	155	135
23	160	136
24	159	139
25	161	139
26	157	145
27	165	138
28	160	139
29	163	135
30	159	138
31	159	135
32	162	140
33	161	142
34	156	141
35	155	137
36	156	135
37	155	144
38	160	146
39	158	140
40	158	137

Несмотря на то что вычисление статистики критерия Стьюдента не зависит от рассматриваемой альтернативы, всё же алгоритмически представляется верным сначала обозначить и гипотезу, и альтернативу.

Итак:

1. Случай первый: H_0 = «$\mu_x = \mu_y$» и H_1 = «$\mu_x > \mu_y$».

Формально это означает, что график функции плотности случайной величины Y сдвинут относительно графика функции плотности случайной величины X влево, как, например, на рисунке 11.

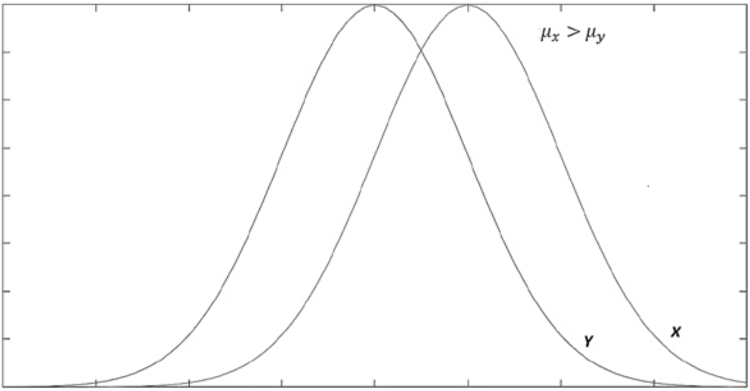

Рисунок 11 – Графики функции плотности нормального распределения с одинаковой дисперсией, но разным математическим ожиданием

Далее рассчитаем *статистику критерия Стьюдента* по формуле:

$$T = \frac{\overline{D}}{S_D}\sqrt{n-1},$$

где $\overline{D}$ – выборка, представляющая собой разность значений вариантов первой и второй выборки:

$$D = \{d_i = x_i - y_i\};$$

${S_D}^2$ – выборочная (смещённая) дисперсия; n – объём выборки.

В примере имеем: $n = 40$; $\overline{D} = 20.5$; ${S_D}^2 \approx 20.05$; $T \approx 28.59$. Статистика данного критерия имеет распределение Стьюдента со степенью свободы, равной $(n-1)$.

При уровне значимости 0.05 альтернатива H_1 = «$\mu_x > \mu_y$» будет означать правостороннюю критическую область, и $\alpha_T \approx 1 - 0.99 < 0.05$.

Вывод будет звучать так: «Данный препарат понижает давление при уровне значимости 0.05».

2. Случай второй: H_0 = «$\mu_x = \mu_y$» и H_1 = «$\mu_x < \mu_y$».

В случае левосторонней альтернативы (которая будет звучать как «давление повысилось») $\alpha_T \approx 0.99 > 0.05 = \alpha$ – и мы вынуждены отклонить альтернативную гипотезу о том, что «препарат повышает давление».

3. Рассмотрим последний вариант: H_0 = «$\mu_x = \mu_y$» и H_1 = «$\mu_x \neq \mu_y$».

Это двусторонняя альтернатива, которая звучит так: «В среднем давление у людей изменилось после приёма лекарств».

В познавательных целях найдём значение статистики Стьюдента, соответствующее уровню значимости 0.025 (распределение Стьюдента является симметричным) и степени свободы, равной 39.

Получим: $t_{\alpha=0.025,n=39} \approx -2.022$ и $|T| > \left|t_{\alpha=0.025,n=39}\right|$.

Делаем *вывод*, что «выборки неоднородны с уровнем значимости 0.05».

Рассмотрим менее идеальный пример (***пример 10***).

У группы пациентов, страдающих гипертонией, проверили давление до и после приёма исследуемого препарата (результат измерения представлен в нижеследующей таблице). Рассмотреть на уровне значимости 0.05 возможность принять или отвергнуть гипотезу при альтернативе, что давление после приёма лекарства понизилось.

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
1	162	140
2	157	160
3	160	158
4	158	163
5	160	125
6	158	156
7	161	166
8	162	160
9	165	165
10	159	162
11	153	130
12	157	134

Продолжение таблицы

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
13	164	156
14	156	132
15	162	128
16	160	140
17	164	132
18	154	125
19	159	140
20	156	135
21	168	129
22	162	162
23	164	138
24	156	147
25	158	150
26	159	155
27	163	126
28	159	159
29	162	128
30	153	135
31	158	130
32	157	140
33	155	125
34	161	145
35	160	149
36	160	158
37	155	155
38	163	160
39	161	126
40	159	142
41	160	134
42	159	159

Продолжение таблицы

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
43	154	120
44	159	135
45	157	125
46	157	157
47	156	166
48	158	126
49	153	132
50	162	162
51	161	128
52	159	141
53	159	168
54	157	132
55	163	122
56	159	159
57	157	161
58	164	135
59	159	145
60	158	126
61	159	132
62	157	125
63	156	156
64	167	128
65	164	168
66	160	137
67	156	165
68	157	128
69	159	159
70	162	135
71	156	164
72	153	124

Окончание таблицы

Номер пациента	До приёма препарата, выборка X	После приёма препарата, выборка Y
73	155	137
74	161	132
75	161	161
76	161	131
77	159	132
78	160	136
79	158	129
80	162	162

Предварительно мы имеем, что у 12 человек давление не изменилось, у 10 человек оно повысилось, и только у 53 человек – стало намного меньше, чем это можно было бы объяснить погрешностью измерения прибора (то есть дисперсией первой выборки).

Итак, наша нулевая гипотеза и альтернатива формулируются как H_0 = «$\mu_x = \mu_y$» и H_1 = «$\mu_x > \mu_y$».

В данном примере имеем: $n = 80$, $\overline{D}$ =16.175, ${S_D}^2 \approx 216.544$ и $T \approx 9.77$.

Статистика данного критерия имеет распределение Стьюдента со степенью свободы $n - 1$.

При уровне значимости 0.05 альтернатива H_1 = «$\mu_x > \mu_y$» будет означать правостороннюю критическую область и $\alpha_T \approx 1 - 0.99 < 0.05$ – и мы вынуждены отклонить гипотезу о равенстве средних.

Одновыборочный критерий Стьюдента применяется также для *сравнения равенства среднего конкретному числу* [9].

Пример 11. Поставщик удобрений утверждает, что применение новой партии удобрений обеспечивает урожайность пшеницы в 60 ц/га. Удобрения внесли на площади в 37 га и получили урожай 55 ц/га при «исправленном» среднем квадратическом отклонении 3 ц/га. При 5%-м уровне значимости оценить справедливость утверждения поставщика.

В данном случае нулевая гипотеза гласит о согласии с поставщиком, тогда как в качестве альтернативы рассмотрим двухстороннюю: «поставщик исказил информацию».

При расчёте статистики учтём:

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

Таким образом, при расчёте статистики критерия Стьюдента имеем:

$$\bar{D} = 60 - 55 = 5, {S_D}^2 = 3^2 \times \frac{37-1}{37} = 8.77,$$

и

$$T = \frac{\bar{D}}{S_D}\sqrt{n-1} = 10.1.$$

В данном случае нас интересует двухсторонняя гипотеза, поэтому найдём значение статистики Стьюдента, соответствующее уровню значимости 0.025 и степени свободы, равной 36. Получим: $t_{\alpha=0.025,n=36} \approx\ -2.028$ и $|T| > \left|t_{\alpha=0.025,n=36}\right|$.

Делаем *вывод:* «с уровнем значимости 0.05 утверждение, сделанное поставщиком, не подтвердилось».

Двухвыборочный критерий Стьюдента

В данном случае имеются две выборки, имеющие нормальное распределение и одинаковую дисперсию и относящиеся к независимым группам наблюдений одной и той же характеристики.

Нулевой гипотезой будет предположение, что выборки однородны, то есть их средние можно считать одинаковыми.

В качестве альтернативы при определении критерия можно выбрать любую из трёх: «средние двух групп не равны», «среднее первой выборки больше» и «среднее второй выборки больше».

Статистика для двухвыборочного критерия Стьюдента считается по формуле:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{n_x {S_x}^2 + n_y {S_y}^2}}\sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}.$$

Здесь n_x, n_y – объёмы соответствующих выборок; $\bar{X}, \bar{Y}$ – их выборочные средние; ${S_x}^2, {S_y}^2$ – выборочные (смещённые) дисперсии.

Статистика данного критерия подчиняется закону распределения Стьюдента с $(n_x + n_y - 2\ )$ степенями свободы.

Пример-шутка с чёрным юмором (пример 12). Средняя продолжительность госпитализации 36 больных пиелонефритом, получивших правильное (соответствующее официальным рекомендациям) лечение, составила 4,51 сут, а 36 больных, получавших неправильное лечение, – 6,28 сут. Стандартные отклонения для этих двух групп – соответственно 1,98 сут и 2,54 сут. Можно ли утверждать на уровне значимости в 0.1, что правильность лечения не влияет на госпитализацию?

Решение. Нулевая гипотеза ставится как предположение, что «правильность лечения не влияет на срок госпитализации». В качестве альтернативы возьмём предположение о том, что «правильность лечения влияет на срок госпитализации, без уточнения в какую сторону».

Статистика критерия Стьюдента будет равна $T = -3.22$, а $t_{0.05,70} \approx\ -1.67$. Так как $|T| > |t_{0.05,70}|$, то нулевая гипотеза отвергается. Влияние правильности лечения на срок госпитализации установлено.

Аналогичный результат будет, если найти: $\alpha_T \approx 0.00097$, что заведомо меньше, чем 0.05.

Вспомним, что есть условия применимости критерия Стьюдента, которые ограничивают его применение и могут влиять на достоверность результата или его объективность.

Так, проверим равенство дисперсий двух генеральных совокупностей по данным примера 10 (***пример 13***). Сделаем это с помощью *критерия Фишера*.

Критерий Фишера проверяет равенство дисперсий двух выборок из нормального распределения.

Нулевая гипотеза формулируется как «равенство дисперсий»: H_0 = «$\sigma^2{}_x = \sigma^2{}_y$», а интересующая нас альтернатива – как «дисперсии двух генеральных совокупностей не равны»: H_1 = «$\sigma^2{}_x \neq \sigma^2{}_y$».

Статистика критерия считается по формуле:

$$f = \frac{S_x}{S_y},$$

где S_x и S_y – поправленные выборочные дисперсии (несмещённые оценки).

Статистика критерия Фишера имеет распределение Фишера со степенями свободы $(n_x - 1)$ и $(n_y - 1)$. У нас они равны 39. Таким образом, в нашем примере: $S_x \approx 10.35$ и $S_y \approx 221.44$, статистика $f \approx 0.04675$. Так как распределение Фишера несимметрично, то рассмотрим $f_{\alpha=0.025,39,39} = 0.053$ и $f_{1-\alpha,39,39} = 1.890$.

Таким образом, мы попадаем в левую критическую область и делаем вывод о том, что «генеральные совокупности двух выборок не одинаковы».

Возвращаясь назад, напомним, что равенство дисперсий проверяемых выборок является одним из условий применимости критерия Стьюдента, и собственно, полученный нелогичный результат по критерию Стьюдента как раз и объясняется невыполнением этого условия.

Замечание. При описании критерия Фишера обычно рекомендуется при расчёте статистики располагать выборки таким образом, чтобы при расчёте делить большую поправленную выборочную дисперсию на меньшую.

Кроме того, существуют рекомендации проверять однородность по критерию Фишера, а потом применять критерий Стьюдента.

Пример 14. По приведённым ниже данным сравнить средние удои коров, получавших различные рационы. Для проверки нулевой гипотезы о равенстве средних принять уровень значимости равным 0,05.

Рацион	Поголовье коров, получавших рацион, гол.	Среднесуточный удой молока в пересчёте на базисную жирность, кг/гол.	Среднеквадратическое отклонение в молочной продуктивности коров, кг/гол.
1	30	16,2	3,8
2	40	17,7	4,2

Решение. Так, проверим на равенство дисперсии двух генеральных совокупностей, отвечающих за эти выборки в примере. Сделаем это с помощью критерия Фишера.

Нулевая гипотеза формулируется в виде равенства дисперсий: H_0 = «$\sigma^2{}_x = \sigma^2{}_y$», а интересующая нас альтернатива – как «дисперсии двух генеральных совокупностей не равны»: H_1 = «$\sigma^2{}_x \neq \sigma^2{}_y$».

Статистика критерия считается по формуле:

$$f = \frac{S_x{}^2}{S_y{}^2},$$

где $S_x{}^2$ и $S_y{}^2$ – поправленные выборочные дисперсии (несмещённые оценки).

Статистика критерия Фишера имеет распределение Фишера со степенями свободы $(n_x - 1)$ и $(n_y - 1)$. В нашем примере они равны 29 и 39. Таким образом, $S_x{}^2 \approx 14.94$ и $S_y{}^2 \approx 18.09$, а статистика $f \approx 0.826$. Так как распределение Фишера несимметрично, то рассмотрим $f_{\alpha=0.025,29,39} = 0.49$ и $f_{1-\alpha,29,39} = 1.961$. Таким образом, мы делаем вывод о том, что «генеральные совокупности двух выборок одинаковы».

Теперь рассмотрим применение критерия Стьюдента. В качестве нулевой гипотезы примем предположение об однородности этих двух групп, а в качестве альтернативной – предположение, что у второй группы удой выше. Статистика критерия Стьюдента будет равна: $T = -0.66$, а $t_{0.05,68} \approx -1.667$. Так как $T > t_{0.05,68}$, то нулевая гипотеза подтверждается. Аналогичный результат будет, если найти $\alpha_T \approx 0.256$, что заведомо больше, чем 0.05.

С помощью критерия Стьюдента можно делать и множественные сравнения с контрольной группой.

Рассмотрим следующий ***практический пример (пример 15)***. В описании патента на изобретение № SU-1753974 A1 показан метод предпосевной обработки семян кормовой свёклы [11]. Так, отобранные семена 3.5–4.5 и 4.5–5.5 мм заливают водой с pH 3,0–5,5 и температурой 20–40 °С, полученной с помощью электролиза, и замачивают в ней на 15–24 ч при комнатной температуре 18–25 °С. Затем семена высушивают на воздухе или при активной вентиляции и температуре 18–50 °С до сыпучего состояния, то есть до влажности 14–18%. Для сравнения использовали сухие семена и семена, замоченные в нейтральной (pH 7,0) и щелочной (pH 8,0–9,0) водах, полученных путём электролиза. Данные для сравнения приведены в нижеследующей таблице.

Вода	pH воды, применяемой для обработки	Энергия прорастания, %	Всхожесть семян, %	Урожайность корнеплодов, ц/га
Кислая	4.5	80.0	89.25	839.0
Щелочная	8.0	80.5	84.0	645.1
Контроль		60.25	70.25	432.5

Можно ли считать, что данная предпосевная обработка влияет на всхожесть семян?

Для принятия решения воспользуемся критерием для определения достоверности различий между частотами определённого количественного признака с ***поправкой Бонферрони***.

Неравенство Бонферрони говорит: *Если k раз применить критерии с уровнем значимости α, то вероятность хотя бы в одном случае найти различие там, где его нет, не превышает произведения k на α.*

Данная поправка позволяет провести сравнение при небольшом количестве парных сравнений (менее 8). В нашем случае количество парных сравнений равно: $3 = C_3^2$. Таким образом, для общего уровня значимости 0.05 каждое сравнение надо проводить при уровне значимости: $0.05/3 \approx 0.017$. Так как количество семян в каждой группе неизвестно, то для удобства возьмём объёмы выборок, равные 100.

Определим основную гипотезу как «выборки однородны», а в качестве альтернативы возьмём двухстороннюю: «выборки неоднородны».

Средняя арифметическая качественных признаков отражает долю или процент особей, имеющих данный признак. В нашем случае: $p_{\text{кис}}$ =0.8925, $p_{\text{щ}} = 0.84$, $p_{\text{конт}}$ =0.7025. Доли случаев, когда семена не взошли, обозначим как $q_{\text{кис}} = 1 - p_{\text{кис}}$ = 0.1075, $q_{\text{щ}} = 1 - p_{\text{щ}}$ =0.16, $q_{\text{конт}} = 1 - p_{\text{конт}}$=0.2975.

Средние квадратические отклонения при обработке подобных данных вычисляют как $\sigma_{\text{кис}} = \sqrt{p_{\text{кис}} q_{\text{кис}}}$, $\sigma_{\text{щ}} = \sqrt{p_{\text{щ}} q_{\text{щ}}}$, $\sigma_{\text{конт}} = \sqrt{p_{\text{контр}} q_{\text{контр}}}$. Для расчета статистики при попарных сравнениях воспользуемся формулами:

$$t_{\text{кис-щ}} = \frac{p_{\text{кисл}} - p_{\text{щ}}}{\sqrt{m_{\text{кисл}}^2 + m_{\text{щ}}^2}} \approx 1.102,$$

$$t_{\text{кис-конт}} = \frac{p_{\text{кисл}} - p_{\text{конт}}}{\sqrt{m_{\text{кисл}}^2 + m_{\text{конт}}^2}} \approx 3.46,$$

$$t_{\text{конт-щ}} = \frac{p_{\text{конт}} - p_{\text{щ}}}{\sqrt{m_{\text{конт}}^2 + m_{\text{щ}}^2}} \approx -2.3291.$$

Здесь m – средняя ошибка частоты качественного признака, выраженная в долях единицы или в процентах:

$$m_{\text{кис}} = \sqrt{\frac{p_{\text{кисл}} q_{\text{кисл}}}{n_{\text{кисл}}}} \approx 0.03,$$

$$m_{\text{щ}} = \sqrt{\frac{p_{\text{щ}} q_{\text{щ}}}{n_{\text{щ}}}} \approx 0.037,$$

$$m_{\text{конт}} = \sqrt{\frac{p_{\text{конт}} q_{\text{конт}}}{n_{\text{конт}}}} \approx 0.046.$$

Статистика данного критерия имеет распределение Стьюдента. С учётом того, что наши выборки имеют одинаковую размерность, имеем: $t_{\alpha=0.016,\nu=198} = -2.159$. В случае двухсторонней альтернативы с уровнем значимости 0.05 нет оснований полагать, что все три выборки различны.

Если же нет необходимости накладывать информацию о типе распределения и его параметрах, то для определения однородности можно воспользоваться непараметрическими критериями. В качестве примера непараметрического критерия для несвязанных выборок рассмотрим применение ***критерия Вилкоксона***, приведённом в [13], для решения нижеследующего примера.

Пример 16. Было изучено общее содержание азота в плазме крови крыс-альбиносов в возрасте 37 и 180 дней. Результаты выражены в граммах на 100 куб. см плазмы. В возрасте 37 дней девять крыс имели: 0.99, 0.83, 0.99, 0.86, 0.9, 0.81, 0.94, 0.87 и 0.1.

В возрасте 180 дней восемь крыс имели: 0.93, 1.18, 0.9, 1.21, 1.2, 0.83, 1.13, 1.12. На уровне значимости 0.05 определите, увеличилось ли содержание азота в крови.

В качестве нулевой гипотезы возьмём предположение, что «выборки однородны». В качестве альтернативы предположим, что «значения второй выборки больше» (левосторонняя альтернатива).

Далее совместим обе выборки в один ряд, упорядоченный по возрастанию, с сохранением информации о принадлежности данных выборкам.

Значение	Выборка	Ранг
0.81	1	1
0.83	1–2	2,5
0.83	1–2	2,5
0.86	1	4
0.87	1	5
0.9	1–2	6,5
0.9	1–2	6,5
0.93	2	8
0.94	1	9
0.99	1	10.5
0.99	1	10.5
0.1	1	12
1.12	2	13
1.13	2	14
1.18	2	15
1.2	2	16
1.21	2	17

Во втором столбце таблицы запись «1–2» означает, что это значение присутствует в обеих выборках и оно включено в общий ряд столько раз, сколько оно встретилось суммарно в обеих выборках. В третьем столбике таблицы приведены ранги соответствующих значений. Фактически это номер данного значения в ряду, кроме случаев, когда имеются одинаковые значения данных. Для одинаковых значений присваивается общий ранг, равный среднему арифметическому числу мест, на которых эти данные находятся.

Затем для подсчёта статистики критерия Вилкоксона сложим ранги, соответствующие значениям 1-й выборки: $U = 1 + 2.5 + 4 + 5 + 6.5 + 9 + 10.5 + 10.5 + 12 = 61$.

Для нахождения α_U воспользуемся предложенным в [13] способом, построенным на том, что статистика критерия Вилкоксона асимптотически нормальна со средним $\mu_U = \frac{n_1(n_1+n_2+1)}{2} + 0.5 = 81$ и дисперсией $\sigma_U^2 = n_1 n_2 (n_1 + n_2 + 1)/12 = 108$. Таким образом, $\alpha_U \approx 0.027 < 0.05$, и на уровне значимости 0.05 принимается альтернатива, что показатели второй выборки больше.

Пример 17. В случае связанных выборок рассмотрим ***непараметрический критерий знаков*** [4]. Разберём его на данных примера 10.

Сформулируем нулевую гипотезу о «случайности уменьшения показателя давления после приёма лекарства». За альтернативную гипотезу возьмём «неслучайный характер уменьшения данных показателя давления после приёма лекарства».

Посчитаем количество «типичных», «нетипичных» и «нулевых» сдвигов.

Под *сдвигом* понимаем разность между первой и второй выборками; под *типичными сдвигами* – преобладающее направление (в нашем случае – положительные сдвиги); под *нетипичными* – сдвиги более редкого, противоположного «типичным сдвигам», знака.

На основе сгенерированных в примере 10 данных получаем:

– 58 типичных сдвигов;

– 10 нетипичных сдвигов;

– 12 нулевых сдвигов.

Далее по таблицам критических точек для *G*-критерия знаков (см. приложение) для уровня значимости 0.05 и количества типичных сдвигов находим нижеследующее.

<table>
<tr><td rowspan="2">Количество типичных сдвигов</td><td colspan="2">Уровень статистической значимости</td></tr>
<tr><td>0.05</td><td>0.01</td></tr>
<tr><td>58</td><td>22</td><td>19</td></tr>
</table>

Это означает, что для уровня значимости 0.05 нетипичных знаков должно быть не более 22, а на уровне 0.01 – не более 19.

Таким образом, мы можем сказать, что нулевая гипотеза о случайности отрицательных сдвигов отвергается на уровне значимости 0.05 и подтверждается гипотеза об уменьшении давления после приёма лекарства.

Помимо описанного выше способа принятия критерия, часто используется способ, при котором задаются не один, а два уровня значимости: $\alpha_1 > \alpha_2$. Далее эмпирическая статистика сравнивается со статистиками, соответствующими уровням значимости α_1, α_2 следующим образом:

– если $t_{эм} < t_{\alpha_1}$, то принимается нулевая гипотеза с уровнем значимости α_1;

– если $t_{эм} > t_{\alpha_2}$, то принимается альтернатива с уровнем значимости α_2;

– если $t_{\alpha_1} < t_{эм} < t_{\alpha_2}$, то говорят о недостаточности данных для принятия решения (рис. 12).

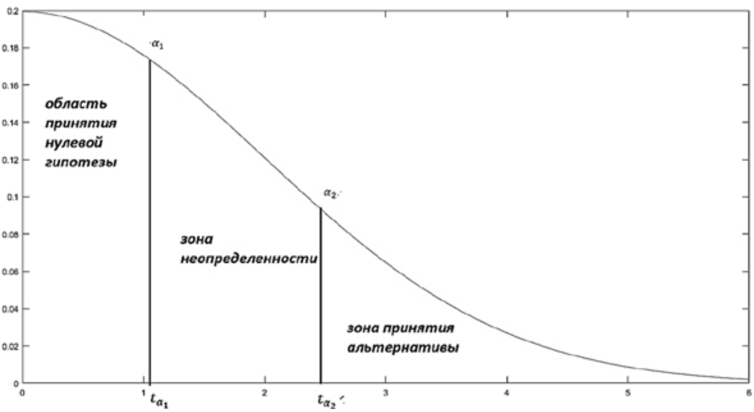

Рисунок 12 – Схема расположения критической области в случае сравнения с двумя уровнями значимости

Далее применим данный метод принятия гипотез при разборе ***критерия*** χ^2. Этот критерий позволяет сделать заключение о распределении выборки, то есть сравнить эмпирическое распределение выборки с предполагаемым теоретическим, как в случае дискретного распределения, так и непрерывного. Кроме того, он позволяет сравнивать распределение двух эмпирических выборок.

На практике закон распределения выборок известен достаточно редко. В случае проверки гипотезы о сравнении выборочного и теоретического распределений критериев называются ***критериями согласия***.

Пример (таблица сопряжённости χ^2***) на сравнение двух эмпирических выборок (пример 18).*** При тромбозе шунта у больных на гемодиализе отслеживаются два признака: «препарат (аспирин – плацебо)» и «тромбоз (есть – нет)». В нижеследующей таблице указаны все их возможные сочетания, поэтому такая таблица называется *таблицей сопряжённости* [2].

Препарат	«Тромбоз есть»	«Тромбоза нет»
Плацебо	18	7
Аспирин	6	13

Примем нулевую гипотезу о том, что «аспирин не влияет на риск тромбоза». Тогда альтернативой будет предположение о том, что «влияние аспирина значимо».

Посчитаем долю признака, отвечающего за развитие тромбоза: всего у нас 44 пациента, из них тромбоз развился у 24, таким образом, доля этого признака от общего количества пациентов составит: (18 + 6) / (18 + 6 + 7 + 13) = 0,5455. Доля признака «тромбоза нет» составит 0.4545.

Если верна нулевая гипотеза, то доля признаков должна выполняться вне зависимости от того, что получал пациент.

Ожидаемое количество пациентов, принимающих плацебо, у которых должен обнаружиться тромбоз, равно: (18 + 7) × 0,5455 = 13,64. Количество пациентов, у которых он не должен обнаружиться, равно 11,36. В группе принимающих аспирин обнаружение тромбоза должно произойти: (6 + 13) × 0,5455 = 10,36 человека –

и не произойти у 8.64 человека. Таблица сопряжённости ожидаемых появлений будет следующей.

Препарат	«Тромбоз есть»	«Тромбоза нет»
Плацебо	13,64	11,36
Аспирин	10,36	8.64

Далее найдём статистику критерия сопряжённости χ^2. Она представляет собой *сумму отношений квадрата разности между эмпирическим появлением и ожидаемым к ожидаемому появлению, проведённую по всем клеткам таблиц сопряжённости:*

$$\chi^2 = \frac{(18-13.64)^2}{13.64} + \frac{(7-11.36)^2}{11.36} + \frac{(6-10.36)^2}{10.36} + \frac{(13-8.64)^2}{8.64} = 7.1.$$

Данная статистика подчиняется распределению χ^2 со степенью свободы, равной 1. Степень свободы считается равной количеству столбцов без 1, умноженному на количество строк, также без 1.

Возьмём за α_1 уровень значимости 0.05, а за α_2 – уровень значимости 0.01. Таким образом, статистика:

$$\chi^2{}_{1-\alpha_1,1} = 3.841\,, \chi^2{}_{1-\alpha_2,1} = 6.635 \text{ и } \chi^2 > \chi^2{}_{1-\alpha_2,1}.$$

В результате мы принимаем альтернативу на уровне значимости 0.01. Таблицы сопряжённости могут быть разной размерности, но общий алгоритм расчёта критерия будет такой же.

Пример (таблица сопряжённости χ^2) сравнения эмпирического распределения выборки с предполагаемым теоретическим [5] (дискретные распределения) (пример 19). Необходимо установить, соответствует ли теоретически ожидаемому расщеплению соотношение частот фенотипов в F_2 кроликов, полученных от скрещивания самок, чёрных с нормальной шерстью (AABB), с пуховыми самцами-альбиносами (aabb). Доминантными признаками является чёрная шерсть (A) и нормальная её длина (B). Все гибриды первого поколения F_1 обладали чёрной короткой шерстью (дигетерозиготными, AaBb). При их скрещивании (AaBb × AaBb) получили 120 потомков: 45 чёрных короткошёрстных, 30 чёрных пуховых, 25 белых короткошёрстных и 20 белых пуховых кроликов. Теоретическое распределение по фенотипу особей, полученных во втором поколении при дигибридном скрещивании кроликов, по закону Менделя составляет: 9 : 3 : 3 : 1.

Нулевая гипотеза в данном случае утверждает, что «полученная популяция кроликов F_2 подчиняется распределению фенотипа по закону Менделя особей, полученных во втором поколении при дигибридном скрещивании». За альтернативу принимается предположение о том, что «наблюдается различие между ожидаемым распределением гибридов второго поколения и распределением по фенотипу полученных кроликов». Составим таблицу.

	Частота, гол.			
	Фенотип кроликов			
	Чёрные, коротко-шёрстные	Чёрные, длинно-шёрстные	Белые, коротко-шёрстные	Белые, длинно-шёрстные
Эмпирическое	45	30	25	20
Теоретическое	67,5	22,5	22,5	7,5

Рассчитаем статистику критерия:

$$\chi^2 = \frac{(45-67{,}5)^2}{67{,}5} + \frac{(30-22{,}5)^2}{22{,}5} + \frac{(25-22{,}5)^2}{22{,}5} + \frac{(20-7{,}5)^2}{7{,}5} = 31{,}1.$$

Степень свободы будет равна: (2 – 1) × (4 – 1) = 3.

Для уровня значимости 0.01 и степени свободы, равной 3, ${\chi^2}_{0.99,3} = 11.34$.

Вывод. Так как $\chi^2 > {\chi^2}_{0.99,3}$, то достоверность отклонения полученной группы кроликов от расщепления фенотипов по закону Менделя подтверждается.

С помощью критерия χ^2 можно также с некой вероятностью установить предпочитаемый ответ [4].

Пример 20. Так, психологу предстоит проверить удовлетворённость работой на данном предприятии. Для этого он проводит опрос случайной выборки из 65 респондентов по выбору ответа из вариантов:

«1» – «Работой вполне доволен».

«2» – «Скорее доволен, чем недоволен».

«3» – «Трудно сказать».

«4» – «Скорее недоволен, чем доволен».

«5» – «Совершенно недоволен».

В качестве нулевой гипотезы предположим, что «ни один из ответов не встречается чаще других». В этом случае можно заключить, что распределение носит равновероятностный характер и ожидаемые частоты появления ответов равны между собой: 65 / 5 = 13. Альтернативой же будет предположение, что «существует наиболее часто встречающийся ответ».

Внесём данные в таблицу.

№ ответа	Эмпирическая частота	Теоретическая частота
1	8	13
2	22	13
3	14	13
4	9	13
5	12	13

Рассчитаем статистику критерия:

$$\chi^2 = \frac{(8-13)^2}{13} + \frac{(22-13)^2}{13} + \frac{(14-13)^2}{13} + \frac{(9-13)^2}{13} + \frac{(12-13)^2}{13} = 9{,}54.$$

Степень свободы в данном случае будет равна 4, $\chi^2{}_{0.99,4} = 13.27$, $\chi^2{}_{0.95,4} = 9.487$, – *в этой ситуации психологу не хватило данных для вывода*, так как $\chi^2{}_{0.95,4} < \chi^2 < \chi^2{}_{0.99,4}$.

Пример 21. В качестве последнего примера проверим вторую выборку из примера 10 на соответствие нормальному распределению. Нулевая гипотеза будет считать «соответствие эмпирического распределения нормальному», альтернативная – «несоответствие эмпирического распределения нормальному». Среднее значение данной выборки равно 143, выборочная дисперсия – 215.9. Эти оценки и примем за параметры теоретического распределения.

Далее представим выборку в интервальном виде.

Номер интервала	Границы	Количество попаданий	Теоретическая вероятность попадания в интервал
1	(-∞, 119.9)	0	4.64
2	(119.9, 124.72)	3	3.902
3	(124.72, 129.54)	16	5.8469
4	(129.54, 134.36)	12	7.875
5	(134.36, 139.18)	9	9.534
6	(139.18, 144)	6	10.375
7	(144,148.82)	3	10.148
8	(148.82,153.64)	2	8.92
9	(153.64,158.46)	8	7.05
10	(158.46,163,28)	14	5.0
11	(163.28,168.1)	7	3.197
12	(168.1,+∞)	0	3.5

Рассчитаем теоретическое попадание в интервал. Для средних интервалов она равна объёму выборки, умноженному на разность между функцией распределения правого и левого концов интервала. Для первого интервала частота попадания будет равна объёму выборки, умноженному на значение функции нормального распределения в правом конце интервала, а для последнего – как объём выборки, умноженный на единицу минус значение функции распределения в левом конце данного интервала.

Далее считаем статистику χ^2 так же, как и в предыдущих случаях. Получим: $\chi^2 = 61.26$ и $\chi^2 > \chi^2{}_{0.99,11}$.

Таким образом, гипотеза о соответствии нормальному распределению не подтвердилась.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1. Средний процент жира в молоке за лактацию коров холмогорских помесей был следующим: 3.4, 3.6, 3.2, 3.1, 2.9, 3.7, 3.2, 3.6, 4.0, 3.4, 4.1, 3.8, 3.4, 4.0, 3.3, 3.7, 3.5, 3.6, 3.4, 3.8. Определите оценки для математического ожидания и дисперсии генеральной совокупности по данной выборке. Определите доверительный интервал для математического ожидания при вероятности в 95% и 99%.

2. Оценивается средний месячный расход на медицинскую помощь в семьях работников данного предприятия, причём средняя ошибка полученной оценки должна быть не более 15 рублей. В пробной выборке из 100 семей выборочная дисперсия составила 350 рублей. Сколько семей должно быть отобрано для проведения обследования?

3. Известна продолжительность (в секундах) физической нагрузки до развития приступа стенокардии у 12 человек с ишемической болезнью сердца: 289, 203, 359, 243, 232, 210, 251, 246, 224, 239, 220, 211. Найдите среднее, стандартное отклонение, медиану, 25-й и 75-й процентили для этих данных.

4. Приведены результаты оценки проницаемости сосудов сетчатки: 1,2; 1,4; 1,6; 1,7; 1,7; 1,8; 2,2; 2,3; 2,4; 6,4; 19,0; 23,6. Найдите среднее, стандартное отклонение, медиану, 25-й и 75-й процентили.

5. При исследовании некоторого показателя, связанного с остротой зрения, были взяты две группы учеников: мальчики и девочки до 12 лет. В первой группе мальчиков выборочное среднее составило 70 $с^2$ и выборочная дисперсия была равна 25 $с^2$ при объёме выборки в 20 мальчиков, тогда как во второй группе девочек выборочное среднее было 75 $с^2$ и выборочная дисперсия составила 36 $с^2$ при объёме выборки в 30 девочек. Определить, существенна ли разность средних в двух выборках при уровне значимости в 0.05.

6. Двум лаборантам поручено определить число эритроцитов в крови, взятой у одних и тех же лиц (для точности эксперимента лаборантам данный факт неизвестен). Каждый из лаборантов проводит по 21 исследованию. Исходя из полученных результатов вычислено стандартное отклонение: 100 000 – у первого лаборанта и 60 000 – у второго. Имеются ли основания предполагать, что второй лаборант работает точнее первого на уровне значимости в 0.01?

7. Двенадцать чёрно-пёстрых коров были покрыты джейсейским быком. Получены нижеследующие данные о количестве молока за лактацию и о жирности молока у матерей и дочерей.

Матери		Дочери	
Кол-во молока, кг	Жирность, %	Кол-во молока, кг	Жирность, %
1 983	3.25	3 509	5.29
3 674	3.81	3 110	6.04
3 976	2.96	3 181	5.24
3 391	3.24	2 997	5.25
4 344	2.82	2 991	5.14
3 784	2.83	3 720	4.72
3 628	2.79	3 268	4.54
3 957	3.08	3 595	4.97
2 185	3.01	2 939	5.13
4 980	3.23	3 213	4.98
2 709	3.68	3 240	5.58
2 807	2.96	3 388	4.81

Можно ли сделать вывод об увеличении показателей у дочерей на уровне значимости в 0.03?

8. При обследовании состояния гланд у студентов первого курса ветеринарного факультета в группе из 84 студентов 39 человек оказались больными. При аналогичном обследовании второкурсников зоофака в группе из 44 студентов больными оказались

28 человек. Определите, есть ли достоверные различия в состоянии здоровья между группами на уровне значимости в 0.01.

9. Были получены нижеследующие данные о весе тушканчиков (в граммах).

Самцы	186	190	165	182	182	182	180
	173	157	179	164	146	173	144
	156	156	156	160	169	161	144
	153	152	151	173			
Самки	162	163	190	188	147	146	145
	157	162	186	175	147	145	145
	155	174	180	148	175	145	144
	153	165	141	146			

Отличаются ли по весу самцы тушканчиков от самок на уровне значимости в 0.04?

10. Пробы по 15 зёрен кукурузы разных стадий зрелости проверяли на устойчивость к раздавливанию. Пробы показали следующие цифры (в единицах давления):

Первая проба: 42, 50, 36, 34, 45, 56, 42, 53, 25, 65, 33, 40, 39, 43, 42.

Вторая проба: 43, 44, 51, 40, 29, 49, 39, 59, 43, 48, 67, 44, 46, 54, 64.

Проверьте, достоверно ли различие между двумя средними двух групп на уровне значимости в 0.01.

11. На 10 парах крыс определяли биологическую ценность белков земляного ореха: сырого P и жареного R. Пары данных (в условных единицах) были следующие: 1) 61 – 55; 2) 60 – 54; 3) 56 – 47; 4) 63 – 59; 5) 56 – 51; 6) 63 – 61; 7) 59 – 57; 8) 56 – 54; 9) 44 – 63; 10) 61 – 58. Достоверна ли разница на уровне значимости в 0.01?

12. В одном классе из 20 детей на удачу были выбраны 10, которым ежедневно начали выдавать апельсиновый сок. Остальные 10 учеников ежедневно получали молоко. Через некоторое время зафиксировали увеличение массы детей в фунтах (1 фунт = 453,6 г).

Ученик	1-й	2-й	3-й	4-й	5-й
Сок	4.0	2.5	3.5	4.0	1.5
Молоко	1.5	3.5	2.5	3.0	2.5

Ученик	6-й	7-й	8-й	9-й	10-й
Сок	1.0	3.5	3.0	2.5	3.5
Молоко	2.0	2.0	2.5	1.5	3.0

Среднее увеличение веса одного ученика в группе, где выдавали апельсиновый сок, составило 2.9 фунта, а в группе, где выдавали молоко, – 2.4. Существенно ли отличается увеличение детей в группах?

13. Сравнивается действие обезболивающих препаратов А и В. (В некоторых случаях одним из «лекарств» может быть инертное плацебо, которое используется для контроля при исследовании действия другого препарата).

В группе больных, которые изъявили желание принять участие в эксперименте, насчитывалось 8 человек. Вполне возможно, что у этих больных возраст, пол, общее состояние и т. д. далеко не одинаковые. Поэтому оба препарата дают каждому больному и фиксируют продолжительность обезболивающего действия каждого из них. С целью обеспечения чистоты эксперимента приняты все разумные меры предосторожности: между приёмами обоих препаратов проходит время, исключающее «перекрывание» их действия; четверо больных получают сначала препарат А, а другие четверо – сначала препарат В, при этом ни один из них не знает, какой именно препарат он принимает; и т. д. Результаты эксперимента приведены в нижеследующей таблице.

Достоверно ли предположение, что длительность действия препарата В больше, чем препарата А?

Больной	Длительность действия препарата, часы	
	A	B
1	3.2	3.8
2	1.6	1.0
3	5.7	8.4
4	2.8	3.6
5	5.5	5.0
6	1.2	3.5
7	6.1	7.3
8	2.9	4.8

14. В нижеследующей таблице приведены данные о 818 наблюдениях, классифицированных по двум признакам: «наличие прививки против холеры» и «отсутствие заболевания». Можно ли на основании этих данных прийти к выводу о зависимости между отсутствием заболевания и наличием прививки?

Наличие прививки	Наличие заболевания	
	Не заболели	Заболели
Привитые	276	3
Непривитые	473	66

15. Среди 2 020 семей с двумя детьми 527 имеют двух мальчиков, 476 – двух девочек, у остальных 1 017 семей дети разного пола. Можно ли считать, что количество мальчиков в семье, имеющей двух детей, является биномиально распределённой случайной величиной?

16. В экспериментах по селекции гороха Г. Мендель наблюдал частоты появления различных видов семян, получаемых в результате скрещивания растений с круглыми жёлтыми и морщинистыми зелёными семенами (в этой задаче частота – количество семян определённого вида). Эти данные и значения теоретических вероятностей, которые определяются согласно теории наследственности Менделя, приведены в таблице.

Вид семян	Частота	Вероятность
Круглые жёлтые	315	$^9/_{16}$
Морщинистые жёлтые	101	$^3/_{16}$
Круглые зелёные	108	$^3/_{16}$
Морщинистые зелёные	32	$^1/_{16}$

Согласуются ли вероятности, полученные согласно теории Менделя, с приведёнными экспериментальными данными?

17. Чётное число мышей было рассажено по одной в клетки, объединённые случайным образом в две группы с одинаковым количеством клеток: группа А – контрольная, группа В – испытуемая. Мышам из группы В был введён препарат, который, предположительно, угнетает палочку Коха (возбудитель туберкулёза). После этого всех животных в случайном порядке заражали туберкулёзом. В таблице приведены дни смерти мышей после инфицирования, при этом данные об одной из мышей были утеряны.

	День				
Группа А	5-й	6-й	7-й	7-й	8-й
Группа В	7-й	8-й	8-й	8-й	9-й

	День				
Группа А	8-й	8-й	9-й	12-й	
Группа В	9-й	12-й	13-й	14-й	17-й

Из предварительных экспериментов известно, что применяемый препарат не является токсичным, поэтому можно считать, что продолжительность жизни в испытуемой группе не меньше, чем в контрольной.

Решить поставленную задачу, сформировав её как задачу проверки статистических гипотез.

18. В таблице приведены результаты эксперимента, в котором изучалось влияние работы метронома на речь людей, страдающих заиканием.

Номер участника	Число заиканий при условии			Номер участника	Число заиканий при условии		
	N	*R*	*A*		*N*	*R*	*A*
1	15	3	5	7	10	0	2
2	11	3	3	8	8	0	3
3	18	1	3	9	13	0	2
4	21	5	4	10	4	1	0
5	6	2	2	11	11	2	4
6	17	0	2	12	17	2	1

Обследовалось 12 человек с тяжёлой формой заболевания. Каждый из них импровизировал трёхминутную речь при условиях *N*, *R*, *A*: *N* – говорить без метронома; *R* – говорить при регулярной (ритмичной) работе метронома (120 ударов за минуту), причём человек был предварительно проинструктирован о необходимости произносить один слог слова на каждый удар метронома; *A* – говорить при неритмичной работе метронома, работающего со случайными интервалами между ударами (от 0.3 до 0.7 с), совершая при этом в среднем те же 120 ударов в минуту (при условии *A*, как и при условии *R*, человек должен произносить один слог на каждый удар метронома).

Приведённые данные, безусловно, свидетельствуют о том, что работа метронома уменьшает количество заиканий. Но существуют ли отличия во влиянии на заикание ритмично и неритмично работающих метрономов?

19. Среди курящих беременных женщин 10 человек имели высокий уровень невротизации, 13 человек – неопределённый результат и 52 человека – низкий; среди некурящих – 5 человек, 14 человек, 102 человека соответственно. Можно ли утверждать, что есть взаимосвязь между табачным анамнезом беременных женщин и уровнем невротизации?

20. Во время эпидемии гепатита А под наблюдением оказалось 47 больных. Из них для лечения 30 больных был применён новый лекарственный препарат, а 17 – лечились прежними средствами. *Итог:* из 30 больных, лечившихся с помощью нового препарата, умерло 4 и выздоровело 26 человек; из 17 больных, лечившихся прежними средствами, умерли 9 и выздоровели 8 человек. Используя критерий χ^2, сделайте вывод об эффективности нового препарата.

21. Исследуется эффективность прививки против сыпного тифа. Под наблюдением находится 17 685 человек. Используя критерий χ^2, сделайте вывод об эффективности прививок на уровне значимости в 0.05.

	Количество заболевших	Количество не заболевших
С прививкой	72	7 988
Без прививки	303	9 322

22. Изучается некоторый признак, предположительно передаваемый по наследству. Для исследования сформированы две попарно связанные выборки количественного выражения признака: отцов и их взрослых сыновей. Проверить однородность этих выборок по указанному признаку при уровне значимости 0.05 и односторонней критической области.

Отцы	Сыновья
117.4	121.3
89.3	103.5
142.0	131.3
103.9	114.1
121.6	135.4
148.9	135.8
107.2	117.9
98.0	110.4
129.5	117.4

Окончание таблицы

Отцы	Сыновья
147.4	152.3
93.0	104.6
98.4	113.0
102.0	113.1
116.2	120.7
133.4	119.0
124.1	135.8
119.9	113.0
114.8	120.4
105.6	116.9
145.6	134.2
121.0	134.7
135.2	121.1
136.1	124.3
97.6	105.2

23. По районам области собраны данные числа лабораторных исследований на одного больного в стационаре за 2011 и 2017 годы. Исследовать значимость различия при уровне значимости 0.01 и односторонней критической области.

Район	Число исследований	
	2011 год	2017 год
1	30.52	47.09
2	27.91	35.07
3	18.39	18.51
4	20.54	22.98
5	27.43	28.08

Окончание таблицы

Район	Число исследований	
	2011 год	2017 год
6	28.38	35.02
7	28.06	23.43
8	20.46	26.19
9	29.27	24.59
10	19.35	25.55
11	17.92	25.91
12	25.37	18.10
13	29.11	29.16
14	25.59	20.61
15	24.11	23.56
16	10.09	12.99
17	21.86	29.75
18	38.16	40.16
19	17.93	22.89
20	23.66	23.82
21	32.41	29.73
22	48.61	48.70
23	28.08	33.14

24. Одной из причин смерти детей в возрасте от 1 недели до 1 года является синдром детской смертности. Обычно смерть наступает на фоне полного здоровья, внезапно, во сне. Изучаются факторы риска для этого показателя, связанные с матерью ребёнка. Всего изучалось 19 047 детей в возрасте до 1 года, из них умерли 44 ребёнка. Применяя критерий χ^2, установите влияние указанных факторов, связанных с риском внезапной детской смертности.

Факторы (мать)		Синдром внезапной детской смертности (количество детей)	
		Умерли	Живы
Возраст матери	до 25 лет	29	7 301
	25 и более	15	11 421
Время окончания предыдущей беременности	менее 1 года	23	4 694
	более 1 года	11	7 339
Курение во время беременности	да	24	5 228
	нет	10	9 595
Самый низкий уровень гемоглобина	менее 12 мг %	26	12 613
	12 мг % и более	7	2 678

25. В описания патента [12] данные для исследования в нижеприведённой таблице были получены во время засушливого 2010 года. Выживаемость растений, как и полевая всхожесть, во многом зависела от погодных условий периода вегетации, поэтому авторами отмечаются пониженные показатели. Целью изобретения является активация функций растений нута за счёт использования инокуляции (нитрагинизации) штаммами клубеньковых бактерий семян нута ризоторфином влажным способом из расчёта 200–250 г на гектарную норму высева в растворе активированной воды с pH 4,2 (1 л воды на 1 ц семян), полученной путём электролиза, и обработанных затем в вакуумной среде при давлении 650–680 мм рт. ст. с одновременным перемешиванием.

Подтвердить или опровергнуть результативность данной обработки семян (см. нижеследующую таблицу).

Полевая всхожесть, сохранность и выживаемость растений нута при различной обработке семян перед посевом

Опытная группа	Варианты опыта	Количество растений, шт./м^2		Полевая всхожесть, %	Сохранность, %	Выживаемость, %
		При полных всходах	Перед уборкой			
1	Ризоторфин + вода (контроль)	41.0	33.5	45.6	81.7	37.2
2	Ризоторфин + сыворотка + вакуум	44.5	36.5	49.4	82.0	40.5

Список рекомендованной литературы

1. Балдин, К. В. Основы теории вероятностей и математической статистики : учебник/ К. В. Балдин, В. Н. Башлыков, А. В. Рокосуев ; общ. ред. К. В. Балдина ; Рос. акад. образования (РАО), Моск. психолого-соц. ин-т (МПСИ). – Москва : Флинта : Изд-во МПСИ, 2010. – 488 с. – Текст : непосредственный.

2. Гланц, С. Медико-биологическая статистика : пер. с англ. / С. Гланц ; пер. с англ. д-ра физ.-мат. наук Ю. А. Данилова ; под ред. Н. Е. Бузикашвили и Д. В. Самойлова. – Москва : Практика, 1998. – 459 с. – Primer of Biostatistics / Stanton A. Glantz, Ph. D. – Fourth edition. – Текст : непосредственный.

3. Горяинова, Е. Р. Прикладные методы анализа статистических данных / Е. Р. Горяинова, А. Р. Панков, Е. Н. Платонов ; Нац. исслед. ун-т « Высшая школа экономики». – Москва : Изд. дом Высшей школы экономики, 2012. – 310 с. – Текст : непосредственный.

4. Ермолаев, О. Ю. Математическая статистика для психологов : учебник / О. Ю. Ермолаев ; Рос. акад. образования. Моск. психолого-соц. ин-т. – 2-е изд., испр. – Москва : Моск. психолого-соц. ин-т : Флинта, 2003. – 335 с. – (Б-ка психолога). – Текст : непосредственный.

5. Крюков, В. И. Генетика. Часть 5. Статистические методы изучения изменчивости : учеб. пособие для вузов / В. И. Крюков. – Орёл : Изд-во ОрёлГАУ, 2006. – 208 с. – Текст : непосредственный.

6. Мамаев, А. Н. Статистические методы в медицине / А. Н. Мамаев, Д. А. Кудлай. – Москва : Практическая медицина, 2021. – 136 с. – Текст : непосредственный.

7. Медик, В. А. Математическая статистика в медицине : учеб. пособие для бакалавриата, специалитета и магистратуры : в 2 т. / В. А. Медик, М. С. Токмачев. – 2-е изд., перераб. и доп. – Т. 1. – Москва : Юрайт, 2018. – 471 с. – Текст : непосредственный.

8. Медик, В. А. Математическая статистика в медицине : учеб. пособие для бакалавриата, специалитета и магистратуры : в 2 т. / В. А. Медик, М. С. Токмачёв. – 2-е изд., перераб. и доп. – Т. 2. – Москва : Юрайт, 2019. – 347 с. – (Бакалавр. Специалист. Магистр). – Текст : непосредственный.

9. Основы математической биостатистики : учебное пособие для обучающихся специальности 36.05.01 «Ветеринария» и направлениям подготовки: 36.03.02 «Зоотехния»; 36.03.01 «Ветеринарно-санитарная

экспертиза» / ФГБОУ ВО «Приморская ГСХА» ; сост. Е. В. Савельева. – Уссурийск, 2016. – 210 с. – Текст : непосредственный.

10. Оуэн, Д. Б. Сборник статистических таблиц / Д. Б. Оуэн ; обраб. табл. Л. С. Барк ; пер. с англ. Л. Н. Большева и В. Ф. Котельниковой ; АН СССР. Мат. ин-т им. В. А. Стеклова. Вычислит. центр. – Москва : [Б. и.], 1966. – IX, 586 с. – Текст : непосредственный.

11. **Патент № 1753974 А1 СССР, МПК A01C 1/00.** Способ предпосевной обработки семян кормовой свёклы : № 4800534/15 : заявл. 15.12.1989 : опубл. 15.08.1992. Бюл. № 30 / Измайлов Ю. С., Фролов И. Н., Михеев В. В., Юферев В. П. ; заявитель Всесоюз. науч.-исслед. ин-т механизации сел. хоз-ва. – 2 с. – Текст : непосредственный.

12. **Патент № 2477942 С2 Российская Федерация, МПК A01C 1/06 (2006.01) A01C 1/08 (2006.01)**. Способ предпосевной обработки семян нута : № 2011123627/13 : заявл. 09.06.2011 : опубл. 27.03.2013 / Мирошников С. А., Малышева А. В., Дерябина Т. Д., Павлов Л. Н., Рогачёв Б. Г., Сидоров Ю. Н. ; Гос. науч. учреждение Всерос. науч.-исслед. ин-т мясного скотоводства Рос. академии с.-х. наук. – Текст : непосредственный.

13. Симушкин, С. В. Теоретические аспекты заданий курсового проекта по математической статистике : [метод. разработка] / С. В. Симушкин.— Казань : Каз. гос. ун-т, 2004. – 68 с. – Текст : непосредственный.

14. Турчин, В. Н. Теория вероятностей и математическая статистика : учеб. для студентов высш. учеб. заведений / В. Н. Турчин. – 2-е изд., перераб. и доп. – Днепр : Лира, 2018. – 752 с. – Текст : непосредственный.

15. Харькова, О. А. Статистические методы и математическое моделирование : учеб. пособие / О. А. Харькова, А. Г. Соловьёв. – Архангельск : Изд-во Сев. гос. мед. ун-та, 2017. – 164 с. – Текст : непосредственный.

16. Шпигельхалтер, Д. Искусство статистики. Как находить ответы в данных / Дэвид Шпигельхалтер ; пер. с англ. Е. Поникарова. – Москва : ООО «Манн, Иванов и Фербер», 2019. – 448 с. – Текст : непосредственный.

Приложение

(справочное)

Критические значения критерия знаков G
для уровней статистической значимости ρ ≤ 0,05 и ρ ≤ 0,01 [10] [1]

n	Уровни стат. значимости (ρ)		*n*	Уровни стат. значимости (ρ)		*n*	Уровни стат. значимости (ρ)	
	0,05	0,01		0,05	0,01		0,05	0,01
5	0	–	27	8	7	49	18	15
6	0	–	28	8	7	50	18	16
7	0	0	29	9	7	52	19	17
8	1	0	30	10	8	54	20	18
9	1	0	31	10	8	56	21	18
10	1	0	32	10	8	58	22	19
11	2	1	33	11	9	60	23	20
12	2	1	34	11	9	62	24	21
13	3	1	35	12	10	64	24	22
14	3	2	36	12	10	66	25	23
15	3	2	37	13	10	68	26	23
16	4	2	38	13	11	70	27	24
17	4	3	39	13	11	72	28	25
18	5	3	40	14	12	74	29	26
19	5	4	41	14	12	76	30	27
20	5	4	42	15	13	78	31	28
21	6	4	43	15	13	80	32	29
22	6	5	44	16	13	82	33	30
23	7	5	45	16	14	84	33	30
24	7	5	46	16	14	86	34	31
25	7	6	47	17	15	88	35	32
26	8	6	48	17	15	90	36	33

[1] Преобладание «типичного» сдвига является достоверным, если $G'_{эмп}$ ниже или равно $G'_{0,05}$, и тем более достоверным, если $G'_{эмп}$ ниже или равен $G'_{0,01}$.

Содержание

Учебное издание

Бандеров Виктор Викторович, кандидат физико-математических наук;
Кашапов Ленар Наилевич, ассистент кафедры биомедицинской инженерии и управления инновациями Инженерного института КФУ;
Кашапов Наиль Фаикович, доктор технических наук, профессор;
Чебакова Виолетта Юрьевна, кандидат физико-математических наук

МЕТОДЫ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ МЕДИКО-БИОЛОГИЧЕСКИХ ДАННЫХ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

В 2 частях

Часть I

Под редакцией кандидата физико-математических наук,
доцента кафедры анализа данных и технологий
программирования Института ВМиИТ КФУ
В. Т. Дубровина

Подписано в печать 24.11.2022. Формат 60×84/16.
Усл. печ. л. 3,72. Тираж 50 экз. Заказ .

Автономная некоммерческая организация
«Издательский Дом «Научное обозрение».

127051, г. Москва, переулок Сухаревский М., д. 9, стр. 1.
Метро «Трубная» или «Цветной бульвар».