БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А. А. Жукова М. Л. Минец

БИОМЕТРИЯ

В ТРЕХ ЧАСТЯХ

ЧАСТЬ 2 ОСНОВНЫЕ ТЕХНИКИ АНАЛИЗА ДАННЫХ

Рекомендовано
Учебно-методическим объединением
по естественно-научному образованию в качестве пособия
для студентов учреждений высшего образования,
обучающихся по специальностям
1-31 01 01 «Биология (по направлениям)»,
1-31 01 02 «Биохимия», 1-31 01 03 «Микробиология»

МИНСК
БГУ
2020

УДК 57.087.1(075.8)
ББК 28с.я73-1
Ж86

Р е ц е н з е н т ы:
кафедра общей биологии и ботаники Белорусского государственного педагогического университета имени Максима Танка
(заведующий кафедрой кандидат сельскохозяйственных наук *А. В. Деревинский*);
доктор биологических наук, член-корреспондент Национальной академии наук Беларуси *В. П. Семенченко*

Жукова, А. А.
Ж86 Биометрия : пособие. В 3 ч. Ч. 2. Основные техники анализа данных / А. А. Жукова, М. Л. Минец. – Минск : БГУ, 2020. – 151 с.
ISBN 978-985-566-955-6.

Рассматриваются общие подходы и методы сравнительного анализа данных, развиваются положения, изложенные в первой части пособия, которое вышло в БГУ в 2019 г. Материал иллюстрируется примерами из биологии и пошаговой инструкцией применения рассмотренных подходов в пакетах программ *Excel* и *Statistica*.

УДК 57.087.1(075.8)
ББК 28с.я73-1

ISBN 978-985-566-955-6 (ч. 2)
ISBN 978-985-566-728-6

ВВЕДЕНИЕ

Биология давно перестала быть лишь описательной наукой. Сегодня ее развитие невозможно без экспериментальных исследований, а решение любой задачи не обходится без сравнений. Так, эколог сопоставляет плотности разных популяций, биохимик – концентрации фермента у контрольных и опытных организмов, ботаник – урожайности различных культур и т. д. Именно сравнение позволяет установить, насколько наблюдаемые различия существенны, т. е. доказать, что определенный фактор оказывает влияние на изучаемый признак.

Вторая часть курса дисциплины «Биометрия» – «Основные техники анализа данных» – знакомит с методами, позволяющими оценивать статистическую значимость различий между выборками (тестами). В ходе реализации того или иного теста обычно рассчитывается некая величина, называемая статистическим критерием, на основе которого судят о размере различий между сравниваемыми группами.

Выбор теста определяется характером анализируемых данных. Когда данные подчиняются закону нормального распределения и объемы анализируемых выборок достаточно велики, применяют параметрические тесты. Они (в сравнении с непараметрическими) обладают большой мощностью, т. е. способностью выявлять различия между сопоставляемыми выборками, поэтому по возможности им следует отдавать предпочтение. Однако если данные распределены ненормально и/или объемы выборок невелики, корректный статистический анализ можно выполнить лишь применяя непараметрические критерии.

Методов оценки статистической значимости различий существует множество, но все они построены по одному принципу. Сначала нужно сформулировать нулевую гипотезу, т. е. предположить, что исследуемые факторы не оказывают влияния на исследуемую величину признака и полученные в ходе эксперимента различия случайны. Затем следует определить, какова вероятность получить наблюдаемые различия при условии справедливости нулевой гипотезы. Если эта вероятность мала, то отвергаем нулевую гипотезу и заключаем, что результаты эксперимента статистически значимы. Это, разумеется, еще не означает, что действие изучаемых факторов доказано, но, во всяком случае, маловероятно, что результат обусловлен случайностью.

Желаем успехов в освоении сравнительных техник анализа данных и их практического применения!

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

α – уровень значимости (вероятность ошибки 1-го рода)
α' – уровень значимости при множественном сравнении
β – вероятность ошибки 2-го рода
F – критерий Фишера, F-критерий
H – критерий Краскела – Уоллиса
H_0 – нулевая гипотеза (нуль-гипотеза)
$H_\text{А}$ ($H_\text{а}$ или H_1) – альтернативная гипотеза
k – число сравнений
l – интервал сравнения
m – число групп
μ – среднее генеральной совокупности
N – объем (число членов) генеральной совокупности
n – объем выборки (численность группы)
P – вероятность справедливости нулевой гипотезы
p – доля
$\hat{p}$ – выборочная оценка доли
Σ – суммирование
σ – стандартное отклонение генеральной совокупности
σ^2 – дисперсия генеральной совокупности
s – выборочная оценка стандартного отклонения
s^2 – выборочная оценка дисперсии
$s_{\hat{p}}$ – стандартная ошибка доли
$s_{\bar{x}}$ – стандартная ошибка среднего
T – критерий Манна – Уитни, критерий Вилкоксона
t – критерий Стьюдента, t-критерий
t_α – критическое значение t при уровне значимости α
df (либо ν) – число степеней свободы
$df_\text{вну}$ ($\nu_\text{вну}$) – внутригрупповое число степеней свободы
$df_\text{меж}$ ($\nu_\text{меж}$) – межгрупповое число степеней свободы
U – критерий Манна – Уитни
χ^2 – критерий хи-квадрат
χ^2_r – критерий Фридмана
$\bar{x}$ – выборочное среднее
z – критерий z (величина со стандартным нормальным распределением)
$z_\text{Т}$ – критерий Вилкоксона для независимых групп

1. РАСПРЕДЕЛЕНИЕ ВЕРОЯТНОСТЕЙ

1.1. КРАТКИЕ ИСТОРИЧЕСКИЕ СВЕДЕНИЯ О ТЕОРИИ ВЕРОЯТНОСТЕЙ

Возникновение понятия вероятности связано как с потребностями страхования, получившего значительное распространение в XVII в., когда заметно укрепились торговые связи благодаря морским путешествим, так и с запросами азартных игр, в которых выигрыш зависит не от умений игрока, а от случайности.

Азартные игры в то время практиковались главным образом среди знати. Особенно была распространена игра в кости. Подсчет возможных и благоприятствующих выигрышу случаев представлял большие трудности, поэтому для решения таких задач некоторые игроки обращались к ученым. Возникновение теории вероятностей относится к середине XVII в. и связано с исследованиями *Блеза Паскаля, Пьера Ферма* и *Христиана Гюйгенса* именно в области теории азартных игр. В их работах постепенно сформировались такие важные понятия, как вероятность и математическое ожидание, были установлены их основные свойства и приемы вычисления.

Блез Паскаль
(*Blaise Pascal*)
1623–1662

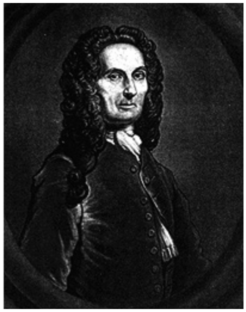
Пьер Ферма
(*Pierre de Fermat*)
1601–1665

Христиан Гюйгенс
(*Christiaan Huygens*)
1629–1695

Крупный шаг вперед в развитии теории вероятностей связан с работами *Якоба Бернулли*. Ему принадлежит первое доказательство одного из важнейших положений теории – закона больших чисел.

Другой важный этап в развитии теории вероятностей соотносится с именем *Абрахама де Муавра*. Этот ученый впервые рассмотрел и обосновал своеобразный закон, часто наблюдаемый в случайных явлениях, – нормальный закон (закон Гаусса).

Якоб Бернулли
(*Jakob Bernoulli*)
1654–1705



Абрахам де Муавр
(*Abraham de Moivre*)
1667–1754

Пьер-Симон де Лаплас
(*Pierre-Simon de Laplas*)
1749–1827

Впервые стройное и систематическое изложение основ теории вероятностей дал математик *Пьер-Симон де Лаплас*. Кроме того, он доказал одну из форм центральной предельной теоремы (теоремы Муавра – Лапласа) и применил ряд приложений теории вероятностей на практике, в частности к анализу ошибок наблюдений и измерений.

Карл Гаусс также внес существенный вклад в развитие теории, дав еще более общее обоснование нормальному закону и разработав метод обработки экспериментальных данных – метод наименьших квадратов.

Следует отметить и работы *Симеона-Дени Пуассона*, доказавшего более общую, чем у Я. Бернулли, форму закона больших чисел, впервые применившего теорию вероятностей к задачам стрельбы. Помимо этого, с именем С.-Д. Пуассона связан один из законов распределения, играющий большую роль в теории вероятностей и ее приложениях.

Карл Фридрих Гаусс
(*Johann Carl Friedrich Gauß*)
1777–1855

Симеон-Дени Пуассон
(*Simeon-Denis Poisson*)
1781–1840

Таким образом, для XVIII – начала XIX в. характерны бурное развитие теории вероятностей и повсеместное увлечение ею. Теория вероятностей становится модной наукой, ее начинают применять не только в случаях, где это правомерно, но и там, где это ничем не оправдано. Данному периоду присущи многочисленные попытки применить теорию вероятностей к изучению общественных явлений, к «моральным» или «нравственным» наукам. Появились работы по судопроизводству, истории, политике, даже богословию, в которых применялся аппарат теории вероятностей. Естественно, подобные попытки были обречены на неудачу и не могли сыграть положительной роли в развитии науки. Их косвенным результатом оказалось то, что в 20–30-х гг. XIX в. в Западной Европе повсеместное увлечение теорией вероятностей сменилось разочарованием и скептицизмом. На теорию стали смотреть как на науку сомнительную, второсортную, как на вид математических развлечений, вряд ли достойный серьезного изучения.

Однако именно в это время в России создается знаменитая Петербургская математическая школа. Среди ее ученых следует назвать *В. Я. Буняковского*, автора первого курса теории вероятностей на русском языке, создателя современной русской терминологии в теории вероятностей, автора оригинальных исследований в области статистики и демографии. Его учеником был русский математик *П. Л. Чебышев*, который расширил и обобщил закон больших чисел.

А. А. Марков, в свою очередь, существенно увеличил область применения закона больших чисел и центральной предельной теоремы, распространив их не только на независимые, но и зависимые опыты. Именно он заложил основы новой ветви теории вероятностей – теории случайных, или стохастических, процессов.

Виктор Яковлевич
Буняковский
1804–1889

Пафнутий Львович
Чебышев
1821–1894

Андрей Андреевич
Марков
1856–1922

Характерными особенностями работ Петербургской математической школы были исключительная четкость постановки задач, математическая строгость применяемых методов и тесная связь теории с требованиями практики. Трудами российских ученых теория вероятностей была выведена с задворков науки и поставлена наравне с точными математическими науками.

1.2. ВЕРОЯТНОСТЬ СОБЫТИЯ И ЕЕ СВОЙСТВА

Что такое вероятность (вероятность события)? Предположим, мы держим в руках игральную кость с шестью сторонами. При подбрасывании она упадет на стол любой из шести возможных сторон. А если число подбрасываний увеличить до 100, возможно, мы получим распределение, представленное в табл. 1.

Таблица 1

Опыт 1. Серия из 100 подбрасываний

Сторона (количество точек)	1	2	3	4	5	6
Количество выпадений	20	15	18	12	16	19
Относительная частота выпадений	0,20	0,15	0,18	0,12	0,16	0,19

После 1000 подбрасываний, вероятно, будет распределение выпавших граней игральной кости, указанное в табл. 2.

Таблица 2

Опыт 2. Серия из 1000 подбрасываний

Сторона (количество точек)	1	2	3	4	5	6
Количество выпадений	170	160	173	171	159	167
Относительная частота выпадений	0,170	0,160	0,173	0,171	0,159	0,167

Ситуацию, как в табл. 3, можно наблюдать, увеличив число подбрасываний до 5000.

Таблица 3

Опыт 3. Серия из 5000 подбрасываний

Сторона (количество точек)	1	2	3	4	5	6
Количество выпадений	815	849	830	850	810	846
Относительная частота выпадений	0,163	0,170	0,166	0,170	0,162	0,169

Вспомним, вначале мы определили возможное выпадение какой-либо стороны игральной кости как одной из шести возможных. В каждой серии опыта относительная (вероятностная) частота выпадений каждой стороны тоже приближалась к 1/6. Отсюда получаем определение понятия *вероятность* как отношения числа случаев, содействующих наступлению ожидаемого события A, к числу возможных исходов:

$$P(A) = m/n,$$

где m – число исходов, благоприятствующих событию A; n – число всех возможных исходов.

Предположим, в ящике находятся 5 черных и 10 белых шаров. Какова вероятность того, что наугад вытащенный шар окажется белым? Так как из 15 шаров в ящике 10 белых, то из 15 ($n = 15$) возможных исходов только 10 ($m = 10$) способствуют появлению именно белого шара, поэтому вероятность этого события равна $P(A) = 10/15 \approx 0{,}67$.

Теория вероятностей содержит ряд основных понятий, на которых базируется. С понятием «вероятность» мы уже познакомились, но что понимают под термином «событие» в теории вероятностей? Это всякий факт, который в результате опыта может либо произойти, либо нет. Приведем несколько примеров событий:

- A – появление орла при бросании монеты;
- B – появление трех орлов при трехкратном бросании монеты;
- C – попадание в цель при выстреле;
- D – появление туза при вынимании карты из колоды.

Рассматривая вышеперечисленные события, видно, что каждое из них обладает какой-то степенью возможности: одни – больше, другие – меньше, причем в некоторых случаях сразу можно решить, какое из них более, а какое менее возможно. К примеру, очевидно, что событие A более возможно, чем B и D. Относительно события C аналогичного вывода сделать нельзя, для этого следует уточнить условия опыта.

Существует подразделение событий, исходя из вероятности их наступления, представленное в табл. 4.

Таблица 4

Подразделение событий

Случайные	Достоверные	Невозможные
В результате опыта событие может произойти или не произойти	В результате опыта событие происходит всегда	В результате опыта событие никогда не произойдет

Если подбросить игральную кость, то заранее нельзя предсказать, какой стороной она упадет. Исход испытания зависит от случая. События, исход которых точно предсказать нельзя, называют *случайными*. Случайные события обозначаются прописными буквами латинского алфавита: A, B, C и т. д.

Случайные события называются *несовместимыми* (несовместными), если появление одного из них исключает возникновение другого в одном и том же испытании.

Так, при подбрасывании монеты нельзя одновременно зафиксировать два события: «монета упала вверх орлом» и «монета упала вверх решкой». Она может упасть вверх либо орлом, либо решкой.

Два случайных события A и B называются *независимыми*, если наступление одного из них не влияет на наступление другого, в противном случае события A и B являются *зависимыми*. К примеру, при одновременном бросании двух монет выпадение на одной из них орла, а на другой решки – независимые события. При вытягивании экзаменационных билетов вероятность вытащить самый простой (например, № 15) восьмым студентом зависит от результатов вытягивания билетов предыдущими студентами.

Для сравнения различных событий по степени их возможности давайте установим единицу их измерения. В качестве такой единицы естественно принять вероятность *достоверного события*, т. е. события, которое в результате опыта обязательно должно произойти. Пример – выпадение не более шести очков при броске одной игральной кости. Если приписать достоверному событию вероятность равную единице, то все другие события – возможные, но не достоверные – будут характеризоваться вероятностями меньше, чем единица. Противоположностью по отношению к достоверному событию является *невозможное событие*, т. е. такое, которое в данном опыте произойти не может, например, появление 12 очков при бросании одной игральной кости. Вероятность невозможного события равна нулю. Итак, *диапазон изменения вероятности случайного события – от 0 до 1* ($0 \le P(A) \le 1$).

Подытожим вышесказанное в виде свойств вероятности.

Свойство 1. Вероятность достоверного события равна 1

Если событие достоверно, то каждый исход испытания благоприятствует событию. В этом случае $m = n$, следовательно, $P(A) = m/n = n/n = 1$.

Свойство 2. Вероятность невозможного события равна 0

Действительно, если событие невозможно, то ни один из исходов испытания не содействует событию. В этом случае $m = 0$, следовательно, $P(A) = m/n = 0/n = 0$.

Свойство 3. Вероятность случайного события есть положительное число, заключенное между 0 и 1

Если случайному событию благоприятствует лишь часть из общего числа всех исходов испытания, то в этом случае $0 < m < n$, а значит, $0 < m/n < 1$, следовательно

$$0 < P(A) < 1.$$

Таким образом, вероятность любого события удовлетворяет неравенству

$$0 \le P(A) \le 1.$$

1.3. ТЕОРЕМЫ СЛОЖЕНИЯ И УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ

Событие, противоположное событию A, принято обозначать $\overline{A}$.

Примеры противоположных событий:

• A – попадание при выстреле;
$\overline{A}$ – промах при выстреле;

• B – выпадение орла при бросании монеты;
$\overline{B}$ – выпадение решки при бросании монеты.

Очевидно, если вероятность достоверного события A равна единице, то вероятность противоположного невозможного события $\overline{A}$ – ноль. Из этих аксиоматических свойств вероятности следует, что вероятности события A и противоположного события $\overline{A}$ в сумме равны единице, т. е. $P(A) + P(\overline{A}) = 1$. В обобщенном виде имеем следующее.

Правило сложения. Вероятность наступления одного из двух (все равно, какого именно) или нескольких независимых и несовместимых событий A, B, C … K равна сумме их вероятностей:

$$P(A + B + C + \ldots + K) = P(A) + P(B) + P(C) + \ldots + P(K).$$

Пример 1. На клумбе растут 20 красных, 30 синих и 40 белых астр. Какова вероятность сорвать в темноте не белую астру?

Решение. $P(A + B + C) = P(A) + P(B) + P(C)$,

$20/90 + 30/90 = 50/90$.

Пример 2. В лотерее участвует 1000 билетов. Из них на один билет выпадает выигрыш в 500 долларов, на 10 билетов – по 100 долларов, на 50 – по 20 долларов, на 100 – по 5 долларов, остальные билеты невыигрышные. Майкл купил один билет. Какова вероятность выиграть не менее 20 долларов?

Решение. Рассмотрим события:

• A – выиграть не менее 20 долларов;

• A_1 – выиграть 20 долларов;

• A_2 – выиграть 100 долларов;

• A_3 – выиграть 500 долларов.

Очевидно, $A = A_1 + A_2 + A_3$.

По теореме сложения вероятностей

$$P(A) = P(A_1) + P(A_2) + P(A_3) = 0{,}050 + 0{,}010 + 0{,}001 = 0{,}061.$$

Пример 3. Вероятности присутствия всех зубов, отсутствия некоторых или отсутствия всех у взрослого равны 0,67; 0,24 и 0,09. Сколько составляет вероятность того, что у пациента есть несколько зубов?

Решение. Рассмотрим события:

A – присутствуют все зубы;

B – отсутствуют некоторые зубы;

C – отсутствуют все зубы.

По теореме сложения вероятностей

$$P\ (A + B) = 0{,}67 + 0{,}\ 24 = 0{,}91.$$

Правило умножения. Вероятность совместного появления двух или нескольких независимых событий равна произведению вероятностей этих событий:

$$P(A, B, C, \ldots, K) = P(A)\ P(B)\ P(C) \ldots P(K).$$

Пример 4. Вероятности присутствия всех зубов, отсутствия некоторых или отсутствия всех у взрослого равны 0,67; 0,24 и 0,09. Найдите вероятность того, что у двух пациентов присутствуют все зубы.

Решение. Рассмотрим события:

A – присутствуют все зубы;

B – отсутствуют некоторые зубы;

C – отсутствуют все зубы.

По теореме умножения вероятностей

$$P\ (A) \cdot P\ (A) = 0{,}67 \cdot 0{,}67 = 0{,}45.$$

1.4. ЭМПИРИЧЕСКИЕ И ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН

Начиная любой эксперимент или приступая к наблюдению, невозможно точно сказать, каков будет результат. Однако практически всегда можно предсказать интервал, в пределах которого будет находиться интересующая исследователя величина.

Так, проводя на уроке биологии в девятом классе демонстрационный опыт по измерению кровяного давления в покое и после физической нагрузки, можно ожидать, что в покое как систолическое, так и диастолическое давление у подростка будет находиться в диапазоне 100 – 140/70 – 90 мм рт. ст. соответственно. В случае плановой медицинской проверки уровня гемоглобина в крови предполагается, что он составит в норме от 135 до 160 г/л у взрослого мужчины и от 120 до 140 г/л у взрослой женщины. А максимальный балл контрольной работы по теме «Химические компоненты живых организмов» не превысит десять.

Обычно вначале, анализируя интервалы результатов, исследователь отображает их как вариационный ряд и изображает графически в виде вариационной кривой или гистограммы для первоначального заключения о вариации изучаемого признака. Такие заключения он делает на основе данных, полученных эмпирическим путем. Соответственно, распределение, построенное на базе этих данных, является *эмпирическим*. Затем полученное эмпирическое распределение он соотносит с определенным *теоретическим распределением вероятности*. В каждом случае выбирается свое теоретическое распределение вероятностей. Зная тип распределения, можно воспользоваться разработанными специально для него приемами математической обработки, получить достоверную информацию о явлении и правильно оценить различия между параметрами разных выборок.

1.5. ОСНОВНЫЕ ТИПЫ РАСПРЕДЕЛЕНИЯ БИОЛОГИЧЕСКИХ ПРИЗНАКОВ

Биометрия, изучая случайные события и поведение таких же величин, может дать более точный прогноз, исходя из теоретических законов поведения случайных величин, относящихся к разным типам распределений. Под *случайной величиной* в теории вероятностей понимают любую количественную характеристику, которая в результате случайного эксперимента может принять одно из некоторого множества значений. В зависимости от типа данных изучаемой переменной (дискретная или непрерывная) ее эмпирическое распределение может соответствовать либо теоретическому дискретному, либо теоретическому непрерывному (рис. 1).

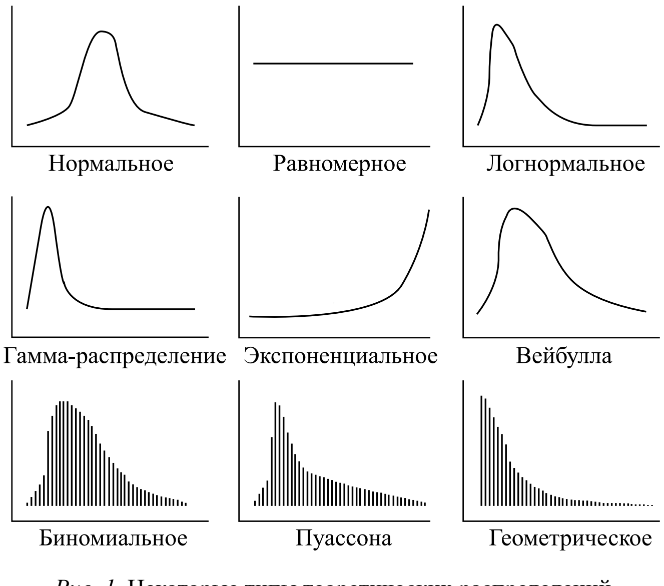


Рис. 1. Некоторые типы теоретических распределений случайной величины:
непрерывные – нормальное, равномерное, логнормальное, гамма-распределение, экспоненциальное, Вейбулла; дискретные – биномиальное, Пуассона, геометрическое (по Шитикову и др., 2003)

1.5.1. Нормальное распределение

Нормальное распределение (нормальный закон) играет большую роль в математической статистике, поскольку многие статистические критерии предполагают, что анализируемые с их помощью экспериментальные данные распределены нормально. График нормального распределения имеет вид колоколообразной кривой (рис. 2).

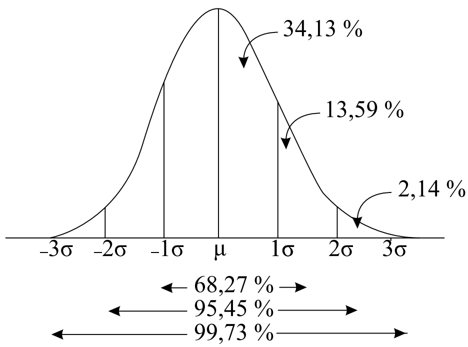


Рис. 2. Нормальное распределение и иллюстрация правила трех сигм

Важной особенностью является то, что форма и положение графика функции нормального распределения определяются только двумя параметрами: средней μ и стандартным отклонением σ. В зависимости от μ распределение будет сдвигаться в ту или иную строну вдоль координатной прямой (рис. 3). Величина стандартного отклонения определяет ширину распределения. При небольших значениях σ кривая будет узкой, тогда как при большом стандартном отклонении колокол приобретет широкое основание (рис. 4). С математической точки зрения нормальная кривая охватывает значения от $-\infty$ до $+\infty$.

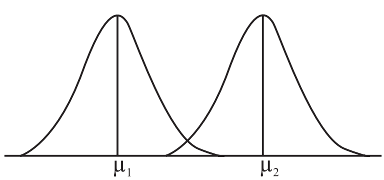


Рис. 3. Результат изменения среднего ($\mu_2 > \mu_1$)

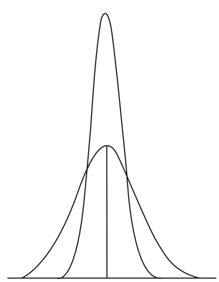


Рис. 4. Результат изменения стандартного отклонения

Нормальное (Гауссовское) распределение используется для приближенного описания явлений, которые носят вероятностный, случайный характер.

По нормальному закону распределено огромное число *непрерывных признаков*, в том числе биологических, поэтому он занимает центральное место не только в биометрии, но и статистике в целом. Учитывая важность нормального распределения, необходимо знать некоторые основные свойства, присущие ему:

• имеет колоколообразную форму;

• симметричен относительно среднего;

• описывается уравнением $P(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2[(x_i - \mu)/\sigma]^2}$;

• полностью определяется двумя параметрами – средним (μ) и стандартным отклонением (σ);

• сдвигается вправо, если среднее увеличивается, и влево, если среднее уменьшается (при постоянном стандартном отклонении);

• становится остроконечным, если стандартное отклонение уменьшается, и «сплющивается» вдоль оси признака, если стандартное отклонение увеличивается (при постоянном среднем);

• при строго нормальном распределении среднее, мода и медиана совпадают по абсолютной величине;

• в случае строго нормального распределения коэффициенты асимметрии (g_1) и эксцесса (g_2) равны нулю;

• при строго нормальном распределении 68,3 % всех вариант находится в диапазоне ±1σ относительно среднего значения (μ). С вероятностью 0,683 (68,3 %) значение случайной величины попадет в интервал от –1σ до +1σ, а с вероятностью 0,317 (1 – 0,683; 31,7 %) это значение может попасть за пределы данного интервала. Соответственно, в пределах от –2σ до +2σ находится 95,5 % вариант, в пределах от –3σ до +3σ – 99,7 % вариант. Это свойство известно как *правило трех сигм* (см. рис. 2).

Последние три свойства часто используют для экспресс-оценки нормальности распределения. Программа *Statistica* (см. прил. 1) позволяет более точно проверить соответствие эмпирического распределения теоретически нормальному с использованием модуля *Distribution Fitting* (Подгонка распределения) либо при помощи «вероятностной бумаги» *Normal probability plot* (График нормальных вероятностей) модуля *Descriptive Statistics* (Описательные статистики). Однако более надежный результат получается при помощи специальных тестов (критериев) этого же модуля: Колмогорова – Смирнова, Лилльефорса, Шапиро – Уилка. Все эти тесты оценивают вероятность соответствия фактических частот вариант теоретически ожидаемым.

Итак, известно что кривая нормального распределения имеет колоколообразный вид. Общая площадь под кривой равна 1 или 100 % и отражает вероятность возможных событий. Вероятность того, что некая величина *x* находится между двумя заданными значениями x_0 и x_1, равна площади под кривой между этими значениями (рис. 5). А вероятность того, что величина x будет больше значения x_2, определяется площадью заштрихованной области при $x > x_2$. Данный подход лежит в основе определения вероятности соответствия расчетных величин эмпирического распределения теоретическому (например, распределения хи-квадрат (χ^2), распределения Стьюдента (t-распределение), распределения Фишера (F-распределение)).

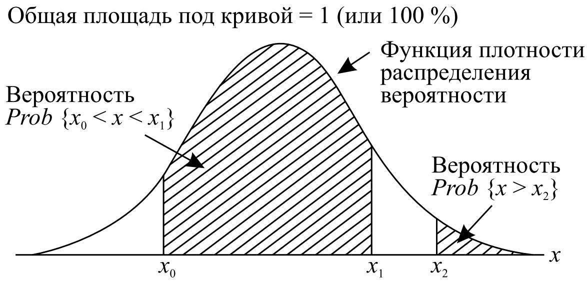


Рис. 5. Определение вероятности случайного события (по Петри и др., 2009)

В настоящее время считается, что наиболее чувствительным (мощным) является тест Шапиро – Уилка, и именно его рекомендуют использовать для проверки нормальности распределения (см. прил. 1, шаги 1–3).

1.5.2. Стандартное нормальное распределение

Кривая нормального распределения имеет вид колоколообразной кривой, точная форма которой определяется двумя параметрами – средним (μ) и стандартным отклонением (σ), поэтому существует бесконечное множество кривых нормального распределения. Они имеют сходную колоколообразную форму. Понятно, что все их описать невозможно, однако, если известны параметры конкретного нормального распределения, его можно преобразовать к стандартному нормальному распределению со средним значением равным нулю и стандартным отклонением равным единице. Такое нормальное распределение со средним значением $\mu = 0$ и стандартным отклонением $\sigma = 1$ также называют *стандартным нормальным* или *z-распределением*.

При анализе переменных, измеряемых в разных единицах измерения, иногда необходимо перейти от первоначальных единиц измерения (градусы, кг, см, экз./м^2 и т. п.) к стандартизованным, т. е. преобразовать данные переменной, имеющей нормальное распределение со средним μ и стандартным отклонением σ в стандартное с параметрами 0 ($\mu = 0$) и 1 ($\sigma = 1$). Стандартизация выполняется по формуле

$$z = \frac{x - \mu}{\sigma}.$$

На практике значения среднего и стандартного отклонения для генеральной совокупности точно определить нельзя, поэтому их заменяют значением выборочной средней арифметической $\overline{x}$ и стандартным отклонением s, вычисляя для каждой варианты соответствующее значение z:

$$z = \frac{x_i - \overline{x}}{s}.$$

1.5.3. Биномиальное распределение

Биномиальное распределение во многом близко к нормальному. Отличие в том, что оно характеризует поведение дискретных признаков, выраженных целыми числами. При биномиальном распределении проявляется та же закономерность, что и при нормальном: чем ближе значения дискретного признака к центру распределения, тем выше вероятность их появления.

Математически распределение называется биномиальным, если вероятности появления отдельных значений признака выражаются величинами, соответствующими коэффициентам разложения бинома Ньютона:

$$(p + q)^k,$$

где p – вероятность появления признака; q – вероятность непоявления признака; k – число классов, отличающихся по появлению признака.

Коэффициенты при отдельных членах разложения бинома Ньютона в случае возведения его в разные степени будут следующими:

$$(p + q)^1 = p + q,$$
$$(p + q)^2 = p^2 + 2pq + q^2,$$
$$(p + q)^3 = p^3 + 3p^2q + 3pq^2 + q^3,$$
$$(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4.$$

Примерами описания признаков с помощью биномиального распределения могут служить количество лучей в плавниках рыб, число хвостовых щитков у рептилий, плодовитость (размер выводка), число поврежденных участков на листьях, поражения глистными инвазиями млекопитающих, число погибших мышей в группах, подвергавшихся экспериментальному воздействию. В основе биномиального распределения лежит альтернативное проявление качественного признака: он может присутствовать у единичного объекта или отсутствовать, проявиться или нет.

Так, клубень картофеля может быть больным или здоровым (качественный признак), тогда проба из нескольких клубней будет содержать некоторое число здоровых (количественный признак). Если отобрать несколько равнообъемных проб, то образуется выборка чисел, для которой можно построить гистограмму распределения. Вероятность события (клубень больной) составляет p, а вероятность альтернативного события (клубень здоровый) равна $q = 1 - p$. При равенстве вероятностей событий $p = q = 0{,}5$ большинство проб (вариант) будут иметь около половины возможных событий (поровну больных и здоровых корнеплодов), и распределение примет симметричную форму (рис. 6). В случае неравенства вероятностей наблюдается та или иная степень асимметрии распределения (рис. 7).

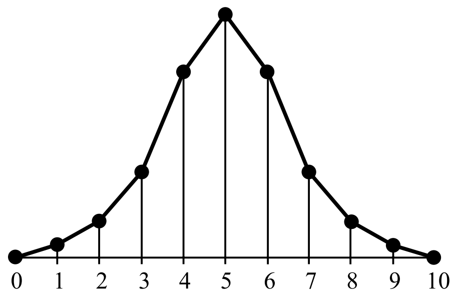


Рис. 6. График биномиального распределения (при $p = q = 0{,}5$)

Биномиальное распределение характеризуется двумя параметрами: средней арифметической и стандартным отклонением. Рассмотрим их расчет.

Пример. Предположим, что получены данные по числу щенков в помете у 76 самок серебристо-черных лисиц.

5	6	6	5	5	6	3
5	5	5	6	6	6	4
4	4	3	5	6	5	3
6	7	5	4	5	4	5
4	1	3	4	3	4	4
5	4	4	6	6	6	3
5	4	6	4	6	8	3
6	6	6	6	8	5	7
6	5	8	4	4	5	5
5	6	5	5	5	4	5
7	5	3	6	6	2	

Вначале распределим данные в вариационный ряд (табл. 5).

Таблица 5

Вариационный ряд по числу щенков в помете

Число потомков, x_i	1	2	3	4	5	6	7	8
Число вариант, f_i (количество самок)	1	1	8	16	23	21	3	3

Изобразим полученный вариационный ряд графически (см. рис. 7).

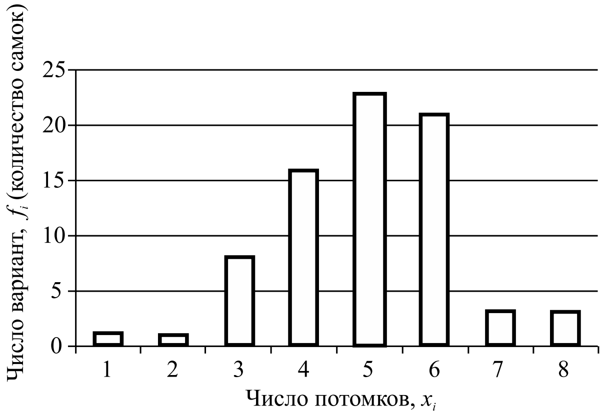


Рис. 7. Эмпирическое биномиальное распределение
(n = 76; $\bar{x}$ = 4,96; s = 1,33)

Все основные параметры распределения можно вычислить по знакомым формулам

$$\bar{x} = \frac{\sum x_i}{n} = 4{,}96 \text{ экз./самку},$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} = 1{,}33 \text{ экз./самку}.$$

Однако обычно для расчета параметров биномиального распределения используют более простые формулы:

$$\bar{x} = kp \text{ и } s = \sqrt{kpq},$$

где k – число изучаемых независимых событий, распределение которых изучается (в нашем примере k = 8); p – вероятность рождения отдельного детеныша; q – вероятность его нерождения.

Для определения вероятности рождения отдельного детеныша p необходимо соотнести общее число появившихся потомков у 76 самок (377 щенков) к возможному максимальному числу потомков (608 щенков; величина определяется, исходя из максимально возможной плодовитости одной самки: $8 \cdot 76 = 608$)

$$p = 377 / 608 = 0,62.$$

Тогда вероятность нерождения детеныша q будет 0,38 ($q = 1 - p$). Подставим полученные величины вероятностей в формулы и вычислим среднюю арифметическую и стандартное отклонение:

$$\bar{x} = kp = 8 \cdot 0,62 = 4,96 \text{ экз./самку,}$$

$$s = \sqrt{kpq} = 8 \cdot 0,62 \cdot 0,38 = 1,37 \text{ экз./самку.}$$

Результаты оказываются практически идентичными.

1.5.4. Распределение Пуассона

Распределение Пуассона также называется распределением редких событий. Как и биномиальное распределение, оно характеризует поведение дискретных признаков, выраженных целыми числами, но описывает редкие события, происходящие 1, 2, 3 и т. д. раз на сотни и тысячи обычных событий. Поэтому в отличие от биномиального распределения, где величины p и q могут быть близки друг к другу, при распределении Пуассона величина p мала, а q – приближается к единице.

В качестве примера можно привести частоту нарушений хромосомного аппарата на каждую тысячу митозов, встречаемость семян сорняка в большой серии навесок семян культурного растения или животных на отрезках длинных маршрутов (на пробных площадках обширной территории), число повторных попаданий животных в ловушки, частоту рождения троен и четверен у женщин.

Распределение Пуассона резко асимметрично, дисперсия равна средней арифметической, что и является критерием для оценки характера распределения изучаемого признака (рис. 8). Рассмотрим это на следующем примере.

Пример. В течение одного года (1946) пометили кольцами и выпустили на волю 32 буревестника. В последующие пять лет часть из них отлавливали повторно: 7 экз. по одному разу, 7 – по два, 2 – по три, 1 – четыре раза, 15 экз. окольцованных птиц повторно не попадались.

Обобщим результаты эксперимента (табл. 6).

Таблица 6

Итог эксперимента

Число повторных отловов, *x*	Число отловленных птиц, *a*	Число случаев повторного отлова, $x \cdot a$
0	15	0
1	7	7
2	7	14
3	2	6
4	1	4
n	32	31

Представим вариационный ряд графически (см. рис. 8). Асимметрия в частотах встречаемости птиц позволяет предполагать распределение Пуассона.

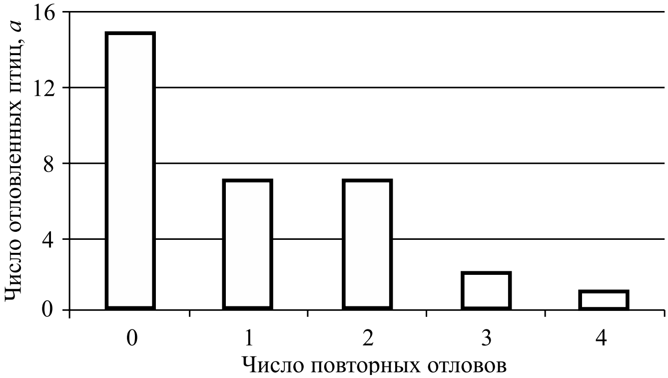


Рис. 8. Эмпирическое распределение Пуассона с параметрами $n = 32$, $\overline{x} \approx s^2 = 0{,}968$

$$\overline{x} = \frac{31}{32} = 0{,}968 \text{ экз.},$$

$$s = \sqrt{\frac{15(0-0{,}968)^2 + 7(1-0{,}968)^2 + 7(2-0{,}968)^2 + 2(3-0{,}968)^2 + (4-0{,}968)^2}{32-1}} =$$
$$= 1{,}121 \text{ экз.},$$

$$s^2 = 1{,}257.$$

Расчеты показали: выборочные значения средней арифметической и дисперсии примерно равны, поэтому предположение о соответствии эмпирического распределения данных теоретическому распределению Пуассона верно.

1.5.5. Логнормальное распределение

Логнормальное распределение характеризуется тем, что логарифмы исходных значений выборки образуют правильное нормальное распределение. Распределение же исходных значений, как правило, умеренно смещено в правую сторону вариационной кривой. Эта модель подходит для описания таких показателей, как концентрация веществ в различных средах, гидрохимические, физиологические и биохимические характеристики. Кроме того, данной моделью удачно может быть описано распределение атмосферного и почвенного загрязнений, а также распределение численности и биомассы бентосных организмов.

* * *

Знание законов распределения изучаемых признаков может дать исследователю информацию о состоянии экосистемы или популяции, об экологии вида, указать на определенную тенденцию в направлении естественного отбора. К примеру, сильный положительный эксцесс в распределении жизненно важных признаков популяции может означать ужесточение стабилизирующего отбора, а асимметричное отклонение от нормального распределения – смену стабилизирующего отбора на движущий.

Однако очень важно правильно определять тип распределения для корректного использования методов математической статистики. Так, применение более мощных параметрических критериев возможно только при нормальном распределении экспериментальных данных. В случае ненормального распределения необходимо использовать непараметрические методы анализа.

2. ОБЩИЕ ПРИНЦИПЫ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

2.1. НУЛЕВАЯ И АЛЬТЕРНАТИВНАЯ ГИПОТЕЗЫ

В любых биологических экспериментах и наблюдениях особое значение имеют расхождения, на основе которых судят об эффективности тех или иных исследований, например по сходству или отличию изучаемого признака между опытной и контрольной группами делают заключение о результатах опыта. При этом особенно важно определить, можно ли данное различие считать закономерным, характерным для всей генеральной совокупности, рассматривать его как результат действия изучаемого фактора, или же оно случайно и является следствием недостаточного количества данных и в следующем опыте может не проявиться.

Данные выводы осуществляются с помощью определенных статистик и называются статистическими. Теория статистического вывода занимает центральное место в статистике. Основным способом, с помощью которого делаются статистические выводы, является проверка статистических гипотез. *Статистическая гипотеза* – предположение о свойстве генеральной совокупности, которое можно проверить, опираясь на данные выборки.

При проверке статистических гипотез используются два понятия: нулевая и альтернативная гипотезы. Обозначается гипотеза буквой H (лат. *hypothesis*). Гипотеза о том, что совокупности, сравниваемые по одному или нескольким признакам, не отличаются, называется *нулевой* (нуль-гипотезой) – H_0. Соответственно, предположение, что сравниваемые совокупности все же отличаются друг от друга – *альтернативная гипотеза*, которая обозначается H_A (в разных источниках могут встречаться H_a или H_1). Принято считать, что нулевая гипотеза H_0 – гипотеза о сходстве, а альтернативная H_A – о различии. Таким образом, принятие нулевой гипотезы H_0 свидетельствует об отсутствии различий, а гипотезы H_A – об их наличии.

Ниже приведены примеры нулевых и альтернативных гипотез.

Ситуация 1. Директору американской школы предъявлен иск в дискриминации учителей по полу. Юрист проводит проверку представленных директором документов о заработной плате учителей-мужчин и учителей-женщин.

Нулевая гипотеза (H_0). Размеры заработной платы учителей-мужчин и учителей-женщин равны, если не принимать во внимание случайные отклонения. Иными словами, реальная разница в размерах заработной платы учителей-мужчин и учителей-женщин не изменилась бы, даже если бы мы сложили все заработные платы в шляпу, хорошо их перемешали и раздали сотрудникам без учета их пола.

Альтернативная гипотеза (H_A). Размеры заработной платы учителей-мужчин и учителей-женщин не равны. Иными словами, различие в размерах заработной платы учителей-мужчин и учителей-женщин превышает простую случайность.

Ситуация 2. Высказано предположение, что средняя масса плодов яблони сорта *A* не отличается от средней массы плодов яблони сорта *B*.

Нулевая гипотеза (H_0). Средняя масса плодов яблони сорта *A* не отличается от средней массы плодов яблони сорта *B*.

Альтернативная гипотеза (H_A). Средняя масса плодов яблони сорта *A* отличается от средней массы плодов яблони сорта *B*.

Ситуация 3. Проведено исследование о влиянии диеты на массу тела. Эксперимент проходил в четырех группах. Испытуемые первой группы (контрольной) питались обычной пищей, члены второй ели только фрукты, третьей – овощи, четвертой – мясо. Через месяц у всех участников измерили массу тела. Высказано предположение, что диета оказала влияние на массу тела испытуемых.

Нулевая гипотеза (H_0). Диета не оказывает влияние на массу тела испытуемых, т. е. средняя масса тела в экспериментальных группах не отличается от средней массы тела в контрольной.

Альтернативная гипотеза (H_A). Диета оказывает влияние на массу тела испытуемых, т. е. средняя масса тела в экспериментальных группах отличается от средней массы тела в контрольной.

Обе гипотезы не могут быть верными одновременно. Чтобы правильно определить одну из двух сформулированных гипотез, необходимо воспользоваться специальными методами, позволяющими оценить статистическую значимость различий между выборками. Эти методы также называют тестами или критериями. В ходе применения того или иного теста рассчитывается некая величина, называемая статистическим критерием. *Статистический критерий* – определенное правило, устанавливающее условия, при которых проверяемую нулевую гипотезу следует либо отклонить, либо нет. На основе рассчитанного критерия определяют, противоречит ли выдвинутая гипотеза фактическим данным.

2.2. ТЕСТИРОВАНИЕ ГИПОТЕЗ В СТАТИСТИКЕ

Тестирование гипотезы – процедура, в которой решается, принять гипотезу или нет. Проверка выдвинутой гипотезы осуществляется по собранным экспериментальным данным, которые подставляются в соответствующую формулу для расчета статистического критерия. В конечном итоге весь имеющийся массив выборочных данных сводится к одному числу.

К примеру, при сравнении массы плодов яблони сортов A, B и C следует сформировать три выборки одинакового объема, взвесить каждое яблоко и на основе полученных данных определить величину статистического критерия. Допустим, нужно вычислить величину критерия Фишера (F-критерия) по формуле

$$F = s^2_{\text{меж}} \Big/ s^2_{\text{меж}}.$$

В ходе вычислений три массива данных будут сведены к одному числу F. Так как анализировали экспериментальные (эмпирические) данные, полученную величину назовем *эмпирическим значением* и обозначим символом $F_{\text{эмп}}$. Ее мы и будем интерпретировать для принятия или отклонения нулевой гипотезы.

Все статистики любого критерия (Фишера, Стьюдента, хи-квадрат и т. п.) подчиняются известным теоретическим распределениям вероятности. Так, график плотности вероятности F-распределения (F-критерия) может иметь вид, представленный на рис. 9.

Полученное эмпирическое значение статистического критерия нужно связать с уже известным теоретическим распределением, которому оно подчиняется, чтобы оценить вероятность справедливости нулевой гипотезы. Более просто – сравниваем полученное эмпирическое значение ($F_{\text{эмп}}$) с *критическим* ($F_{\text{кр}}$). Если значение попало в критическую область ($F_{\text{эмп}} \geq F_{\text{кр}}$), отвергаем нулевую гипотезу, и это значит, что статистически значимые различия найдены. Если эмпирическое значение не попадает в критическую область ($F_{\text{эмп}} < F_{\text{кр}}$), принимаем нулевую гипотезу, отличий нет.

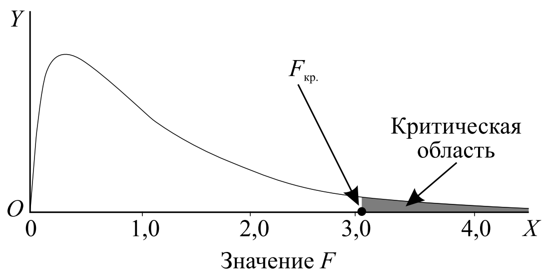


Рис. 9. Распределение вероятности для статистики критерия Фишера

Выбранную границу отсечки критической области называют *критическим уровнем* значимости критерия. По площади данная область чаще всего равна величине 0,05, которую обозначают символом α и называют *уровнем значимости* (или *вероятностью ошибки*).

Большинство компьютерных пакетов обеспечивают автоматическое вычисление эмпирического значения используемого статистического критерия, сравнение его с критическим значением, а также расчет величины P – вероятности справедливости нулевой гипотезы. В конечном итоге именно эту величину P и будем сравнивать с уровнем значимости α. Если P оказывается больше, чем α, то нулевая гипотеза сохраняется, если меньше – отклоняется и принимается альтернативная.

$P > \alpha$, сохраняем H_0 $P \leq \alpha$, принимаем H_A

В биологических исследованиях достаточным считают уровень значимости 0,05 (5 %). Также в качестве критического уровня значимости принимают α = 0,01 (1 %) или даже α = 0,001 (0,1 %).

Статистические компьютерные программы обычно рассчитывают величину P в долях единицы (вспоминаем, что вероятность варьирует в пределах от 0 до 1 или от 0 % до 100 %). Поэтому, получив, например, $P = 0{,}87$, можно заключить, что верна нулевая гипотеза, вероятность ее справедливости составляет 87 %. В случае $P = 0{,}007$ отклоняем нулевую гипотезу и принимаем альтернативную, ошибка в отклонении верной нулевой гипотезы небольшая – порядка 0,7 %.

Форма кривой любого теоретического распределения меняется в зависимости от разных чисел степени свободы (рис. 10).

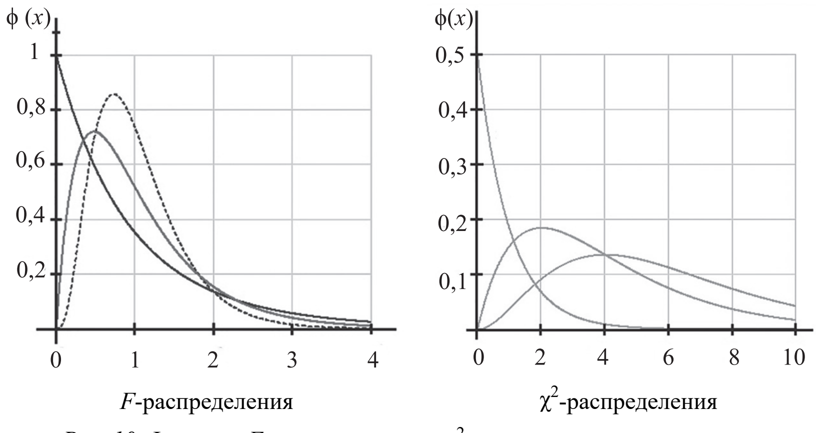


Рис. 10. Функции F-распределения и χ^2-распределения в зависимости от разных чисел степеней свободы

Из-за разной формы теоретического распределения величина критического значения, с которым сравнивают полученную по экспериментальным данным эмпирическую величину, бывает разной. На нее влияет принятый исследователем уровень значимости α (0,05; 0,01 или 0,001), поэтому, выбирая уровень значимости α, мы также воздействуем на величину критического значения.

2.3. ДВА РОДА ОШИБОК

Несмотря на то, что исследователь надеется сделать правильный вывод относительно нулевой гипотезы, нужно признать факт, что он может принять и неправильное решение, когда отвергает / не отвергает нулевую гипотезу, потому что это решение принимается исходя из данных анализируемой выборки. Возможные ошибки при тестировании гипотез показаны в табл. 7.

Таблица 7

Возможные ошибки при проверке гипотез

Решение по критерию	В действительности	
	H_0 верна	H_0 не верна
Отклонить H_0	Ошибка 1-го рода, α	Нет ошибки (правильное решение), 1 – β
Принять H_0	Нет ошибки (правильное решение), 1 – α	Ошибка 2-го рода, β

Ошибка 1-го рода. Отвергаем нулевую гипотезу, когда она истинна, и делаем вывод, что различия есть, когда в действительности их нет. Если, например, в ходе опыта установлено, что новое лекарство эффективнее, хотя на самом деле его действие не отличается от старого, то это ошибка 1-го рода. Максимальная приемлемая вероятность ошибки 1-го рода называется уровнем значимости и обозначается α. Обычно α принимают равной 0,05 (5 %), однако можно взять и другой уровень значимости, например 0,01 (1 %) или 0,001 (0,1 %).

Ошибка 2-го рода. Мы не отвергаем нулевую гипотезу, когда она ложна и делаем вывод, что различий нет, когда в действительности они есть. Если, например, установлено, что оба лекарства одинаково эффективны, хотя на самом деле их действие различается, то допущена ошибка 2-го рода. Ее вероятность обозначается β. Вероятность 1 – β называется мощностью критерия: чем она больше, тем меньше вероятность ошибки 2-го рода.

2.4. ПАРАМЕТРИЧЕСКИЕ И НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

Существуют две большие группы статистических критериев (тестов): параметрические и непараметрические. Применение первых сопряжено с вычислением таких выборочных характеристик, как средняя арифметическая и стандартное отклонение (дисперсия). Так, например, критерий Фишера (F-критерий) относится к параметрическим тестам, которые применимы лишь в случаях, когда анализируемые выборки и, следовательно, генеральная совокупность, откуда они взяты, распределяются по нормальному закону.

Непараметрические критерии таких ограничений не имеют, они не требуют, чтобы анализируемые данные подчинялись нормальному закону распределения. При вычислении непараметрических критериев нет необходимости находить средние величины и показатели вариации. При вычислении таких критериев выборочные данные обычно заменяют рангами (т. е. числами 1, 2, 3 и т. д., описывающими положение исходных значений в упорядоченном наборе данных) и не делают никаких предположений относительно типа распределения этих данных. Поэтому для непараметрических критериев используют термин «критерий свободный от распределения» или «ранговый критерий». Непараметрический критерий применяют, когда объем выборки небольшой (настолько, что невозможно правильно определить, имеют данные нормальное распределение или нет) и когда изучаемые признаки не количественные, а качественные, и выражаются порядковыми номерами, индексами и другими условными знаками.

Параметрические критерии обладают более высокой мощностью по сравнению с непараметрическими. Иными словами, они способны с большой достоверностью отвергать нулевую гипотезу, если она неверна. Однако если данные распределены ненормально и/или объемы выборок невелики, корректный статистический анализ можно выполнить лишь применяя непараметрические критерии.

* * *

Резюмируя вышесказанное, еще раз перечислим этапы проверки статистических гипотез:

1) сформулировать нулевую и альтернативную гипотезы;

2) отобрать необходимые данные для анализа;

3) выбрать статистический критерий;

4) вычислить эмпирическое значение статистики критерия;

5) сравнить расчетное эмпирическое значение статистики критерия с критическим значением соответствующего распределения;

6) интерпретировать результаты относительно принятого уровня значимости:

• если $P > \alpha$, то сохраняем H_0;

• если $P \leq \alpha$, то принимаем H_A.

3. ЧИСЛОВЫЕ ДАННЫЕ. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

В первой части пособия обучение анализу данных мы начали с изучения графических и числовых инструментов для описания распределения одной переменной. Это служит основой и для сравнения нескольких распределений. Изучение статистических результатов начнется схожим образом: с выводов об одиночном распределении, затем сравним два распределения.

Вспомним, что важными свойствами любого распределения являются его центр и разброс. Если распределение нормально, описываем его центр средним μ, а разброс – стандартным отклонением σ ($\bar{x}$ и s для выборки соответственно).

Одними из наиболее часто используемых статистических методов для сравнения двух групп являются t-тесты. К примеру, перед исследователем может стоять задача вычислить, влияет ли содержание белка в диете на прибавку массы белых крыс (табл. 8). Нужно ответить на вопрос: является ли статистически значимой полученная в опыте разница между средними прибавками массы?

Таблица 8

Прибавка массы белых крыс в возрасте между 28 и 84 днями жизни при разном содержании белка в диете (по Дж. У. Снедекору, 1961)

Содержание белка в диете	Число животных	Прибавка массы, г							Средняя прибавка, г
Высокое	7	146	119	161	113	129	83	123	124,9
Низкое	7	70	118	101	85	107	132	94	101,0

Начнем с обсуждения организации эксперимента. Целью исследования является изучение действия одного фактора – содержания белка в диете, поэтому экспериментатор стремится организовать опыт так, чтобы сравниваемые группы животных отличались лишь по исследуемому фактору. По всем другим условиям опыта группы должны быть если не идентичны (это идеальная ситуация), то практически сходны.

При планировании эксперимента специалист проводит огромную предварительную работу. Исходя из целей исследования и реальных возможностей, решается вопрос о выборе материала: или животных определенной генетической линии, или гибридов первого поколения между линиями, или просто беспородных животных. Необходимо так организовать разведение животных, чтобы к началу эксперимента было нужное число крыс определенного возраста. В примере не оговорено, велось ли исследование на самках или самцах, ограничивалась ли какими-то пределами масса животных в начале эксперимента (28 дней жизни) и т. п. Не описаны и условия эксперимента в интервале 28-го и 84-го дней жизни животных, которые также должны быть одинаковыми для обеих групп: нелепостью было бы, например, содержать животных одной группы в индивидуальных клетках, а животных другой группы – по три в клетке. Подразумевается, наконец, что взвешивание животных в начале и конце эксперимента проводится согласно принятым методическим требованиям.

Обсуждая биометрические задачи, мы всегда предполагаем наличие профессиональной культуры при проведении соответствующего эксперимента. Однако даже если опыт организован профессионально грамотно, возникает еще один принципиально важный вопрос. Так, в нашем распоряжении 14 животных, любое из них с равными основаниями может быть взято в ту или иную экспериментальную группу (вариант опыта). Каким образом из 14 животных выбрать 7, которые будут получать диету с высоким содержанием белка, и 7 – с низким содержанием белка?

Это можно решить следующим образом: животные распределяются по вариантам опыта случайно (такая процедура носит название *рандомизации*). Современные представления о биологической изменчивости позволяют утверждать, что 14 крыс просто не могут быть идентичны по всем признакам и свойствам. Рандомизация, естественно, не приведет к тому, что животные, отнесенные к разным вариантам опыта, станут идентичными по всем показателям. Рандомизация представляет собой в некотором смысле механическую процедуру, снимающую субъективизм экспериментатора, определенные неконтролируемые предпочтения (см. дискуссионный вопрос в конце раздела). При проведении рандомизации во всех экспериментах обеспечивается выравнивание в среднем, происходит устранение системных ошибок в формировании групп. Таким образом, распределение животных по вариантам опыта будет осуществлено случайно, вне зависимости от желаний и склонностей экспериментатора.

Вернемся к примеру. Предположим, что продуманы все детали техники эксперимента, 14 животных разбиты на две группы. Какие статистические задачи возникнут после того, как эксперимент будет завершен? Обратите внимание: мы обсуждаем статистические задачи до начала эксперимента, поскольку это последний этап его планирования. Работа ведется с количественным признаком «приращение массы между 28-м и 84-м днями жизни». Объемы выборок невелики: $n_1 = n_2 = 7$, поэтому вопрос о том, является ли распределение генеральной совокупности нормальным, вряд ли удастся решить, используя данные этого эксперимента. Допустим, это предположение правомочно и полученные средние прибавки (средние арифметические) $m_1 = 124{,}9$ и $m_2 = 101{,}0$ можно использовать в качестве оценок средних значений μ_1 и μ_2 генеральных совокупностей. Тогда, чтобы ответить на поставленный вопрос, нужно проверить нулевую гипотезу H_0: $\mu_1 = \mu_2$. Альтернативная гипотеза H_A в данном случае $\mu_1 \neq \mu_2$ (поскольку нет оснований полагать, например, что $\mu_1 > \mu_2$, но не $\mu_1 < \mu_2$).

Обратите внимание, что биолога в рассматриваемом примере интересует разность $\mu_1 - \mu_2$, вопрос же о сравнении σ_1^2 и σ_2^2 не ставится. Быть может, в данном случае это и оправдано, но еще Р. Фишер как-то заметил, что биолог склонен делать акцент на сопоставлении средних, часто пренебрегая сравнением изменчивости, т. е. дисперсий.

Тесты значимости для среднего в нормально распределенной генеральной совокупности основаны на среднем выборочном $\bar{x}$, которое прогнозирует неизвестное μ. Выборочное распределение $\bar{x}$ зависит от σ. Это не проблема, когда σ известно, однако обычно необходимо прогнозировать его значение, несмотря на то, что мы прежде всего заинтересованы в μ. Для прогнозирования стандартного отклонения σ в генеральной совокупности используется выборочное стандартное отклонение s.

Если на уровне значимости α нулевая гипотеза будет принята, то для рассмотренного примера говорят, что разница между средними привесами не существенна. Если нулевая гипотеза будет отклонена и, следовательно, принята альтернативная, говорят, что разница статистически значима.

3.1. СРАВНЕНИЕ ДВУХ ГРУПП

3.1.1. Тест Стьюдента для двух независимых групп

В задачах рассмотрим ситуации, когда нужно сравнить только две группы. В этом случае обычно применяют критерий Стьюдента или t-тест, который часто используется в научных публикациях. Следует сразу запомнить, что данный критерий предназначен для сравнения именно двух групп, а не нескольких попарно.

Подробнее о принципе метода. Предположим, нужно испытать диуретическое действие нового препарата. Для этого набираем десять добровольцев, случайным образом разделяем их на две группы – контрольную, которая будет получать плацебо, и экспериментальную, что будет получать препарат, а затем определяем суточный диурез (результаты представлены на рис. 11, *а*). Средний диурез в экспериментальной группе на 240 мл больше, чем в контрольной. Впрочем, сомнительно подобными данными убедить кого-нибудь, что препарат – диуретик, так как группы слишком малы.

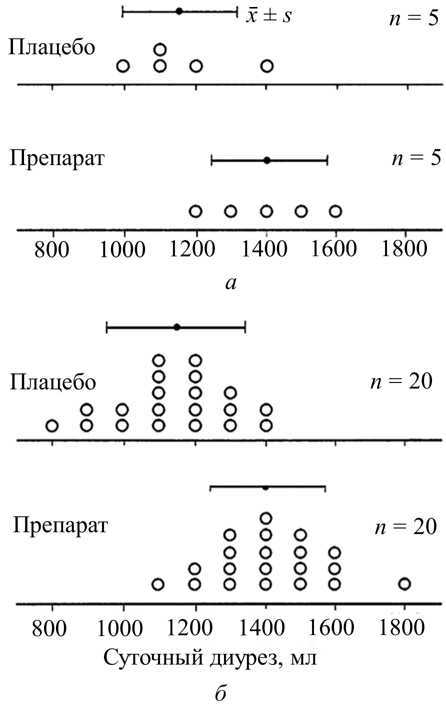


Рис. 11. Результаты испытаний предполагаемого диуретика: *а* – диурез после приема плацебо и препарата (в обеих группах по 5 человек); *б* – средние и стандартные отклонения остались прежними, однако доверие к результату повысилось (в обеих группах по 20 человек)

Повторим эксперимент, увеличив число участников. Теперь в обеих группах по 20 человек. Результаты представлены на рис. 11, *б*. Средние и стандартные отклонения примерно те же, что и в эксперименте, где меньше участников. Тем не менее кажется, что результаты второго эксперимента заслуживают больше доверия. Почему?

Вспомним еще раз, что точность выборочной оценки среднего характеризуется стандартной ошибкой среднего

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

где n – объем выборки; σ – стандартное отклонение в генеральной совокупности, из которой извлечена выборка.

С увеличением объема выборки стандартная ошибка среднего уменьшается, следовательно, понижается и неопределенность в оценке выборочных средних, а вместе с тем и неопределенность в оценке их разности. Применительно к нашему эксперименту, мы будем более уверены в диуретическом действии препарата. Точнее сказать, мы менее уверены в справедливости гипотезы об отсутствии диуретического действия (будь такая гипотеза верна, обе группы можно было бы считать двумя случайными выборками из одной нормально распределенной совокупности).

Чтобы формализовать приведенные рассуждения, рассмотрим отношение

$$t = \frac{\text{Разность выборочных средних}}{\text{Стандартная ошибка разности выборочных средних}}.$$

Для двух случайных выборок, извлеченных из одной нормально распределенной совокупности, это отношение, как правило, будет близко к нулю. Чем меньше (по абсолютной величине) t, тем больше вероятность нулевой гипотезы. Соответственно, чем больше t, тем больше оснований отвергнуть нулевую гипотезу и считать, что различия статистически значимы.

Для нахождения величины t нужно знать разность выборочных средних и ее ошибку. Вычислить разность выборочных средних нетрудно: нужно вычесть из одного среднего другое – сложнее найти ошибку разности. Для этого обратимся к более общей задаче нахождения стандартного отклонения разности двух чисел, случайным образом извлеченных из одной совокупности.

Для нахождения стандартного отклонения разности рассмотрим представленную на рис. 12, *а* гипотетическую генеральную совокупность из 200 членов, где среднее равно 0, стандартное отклонение 1. Выберем наугад два члена совокупности и вычислим разность. Выбранные элементы помечены черными кружками, полученная разность представлена таким же кружком на рис. 12, *б*.

Извлечем еще пять пар (на рис. 12 они различаются штриховкой), вычислим разность для каждой пары. Похоже, что разброс разностей будет больше разброса исходных данных.

Из исходной совокупности выберем наугад еще 100 пар, для каждой из них вычислим разность. Теперь все разности, включая вычисленные ранее, изображены на рис. 12, *в*. Стандартное отклонение для полученной совокупности разностей – примерно 1,4, т. е. на 40 % больше, чем в исходной совокупности.

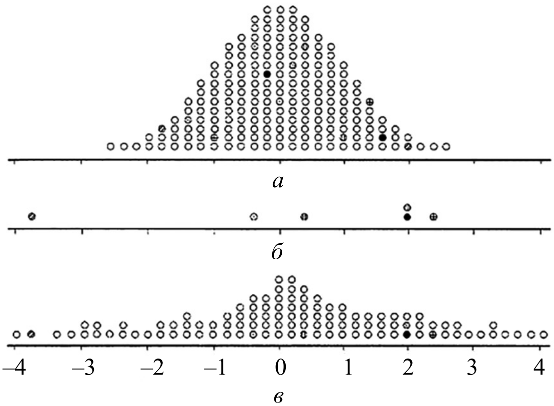


Рис. 12. Нормально распределенная совокупность
и нахождение разности ее элементов:
а – выбор случайных пар объектов в совокупности;
б – разности первых шести пар; *в* – разности еще ста пар.
(Разброс разностей больше, чем разброс исходных значений)

Углубляясь в математическую сторону вопроса, можно доказать, что *дисперсия разности двух случайно извлеченных значений равна сумме дисперсии совокупностей, из которых они извлечены* (что интересно, дисперсия суммы двух случайно извлеченных значений тоже равна сумме дисперсий совокупностей, из которых они извлечены).

Если значение X извлечено из совокупности, имеющей дисперсию σ_X^2, а значение Y из совокупности, имеющей дисперсию σ_Y^2, то распределение всех возможных значений $X - Y$ имеет дисперсию

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2.$$

Почему дисперсия разностей больше дисперсии каждой совокупности, легко понять на нашем примере (см. рис. 12): в половине случаев составляющие пары лежат по разные стороны от среднего, поэтому их разность еще больше отклоняется от среднего, чем они.

Все пары чисел были извлечены из одной совокупности. Ее дисперсия равна 1. В таком случае дисперсия разностей будет

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 = 1+1 = 2.$$

Стандартное отклонение есть квадратный корень из дисперсии, поэтому оно равно квадратному корню из двух, т. е. больше стандартного отклонения исходной совокупности примерно на 40 %, как и получилось в нашем примере.

Чтобы оценить дисперсию разности членов двух совокупностей по выборочным данным, нужно в приведенной выше формуле заменить дисперсии их выборочными оценками:

$$s^2_{X-Y} = s^2_X + s^2_Y.$$

Данной формулой можно воспользоваться и для оценки стандартной ошибки разности выборочных средних. В самом деле, стандартная ошибка выборочного среднего – стандартное отклонение совокупности средних значений всех выборок объемом n из этой генеральной совокупности.

$$s^2_{\bar{X}-\bar{Y}} = s^2_{\bar{X}} + s^2_{\bar{Y}}.$$

Тем самым искомая стандартная ошибка разности средних

$$s_{\bar{X}-\bar{Y}} = \sqrt{s^2_{\bar{X}} + s^2_{\bar{Y}}}.$$

Сейчас можно вычислить отношение t.

Напомним, что для сравнения двух групп мы рассматриваем отношение

$$t = \frac{\text{Разность выборочных средних}}{\text{Стандартная ошибка разности выборочных средних}}.$$

Воспользовавшись результатом предыдущих вычислений, имеем

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2_{\bar{X}_1} + s^2_{\bar{X}_2}}}.$$

Если ошибку среднего выразить через выборочное стандартное отклонение, получим другую запись этой формулы:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}},$$

где n – объем выборки.

Если обе выборки извлечены из одной совокупности, то выборочные дисперсии s_1^2 и s_2^2 – оценки одной и той же дисперсии σ^2, поэтому их можно заменить на *объединенную оценку дисперсии*. Для выборок равного объема объединенная оценка дисперсии вычисляется как

$$s^2 = \frac{s_1^2 + s_2^2}{2}.$$

Значение t, полученное на основе объединенной оценки

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s^2}{n} + \dfrac{s^2}{n}}}.$$

Если объем выборок одинаков ($n_1 = n_2 = n$), оба способа вычисления t дадут один и тот же результат. Однако если объем выборок разный, то это не так.

Посмотрим, какие значения t можно получить, извлекая случайные пары выборок из одной и той же нормально распределенной совокупности.

Так как выборочные средние обычно близки к среднему по совокупности, значение t чаще будет близко к нулю. Тем не менее иногда мы все же будем получать большие по абсолютной величине значения t. Чтобы понять, какую величину t следует считать достаточно большой, чтобы отвергнуть нулевую гипотезу, проведем мысленный эксперимент.

Вернемся к испытаниям предполагаемого диуретика. Допустим, в действительности препарат не оказывает диуретического действия. Тогда и контрольную группу, которая получает плацебо, и экспериментальную, которая принимает препарат, можно считать случайными выборками из одной совокупности. Пусть это будет совокупность из 200 человек, представленная на рис. 13, *а*. Члены контрольной и экспериментальной групп различаются штриховкой. В нижней части рисунка данные по этим двум выборкам показаны так, как их видит исследователь. Взглянув на них, трудно подумать, что препарат – диуретик. Полученное по этим выборкам значение t равно –0,2.

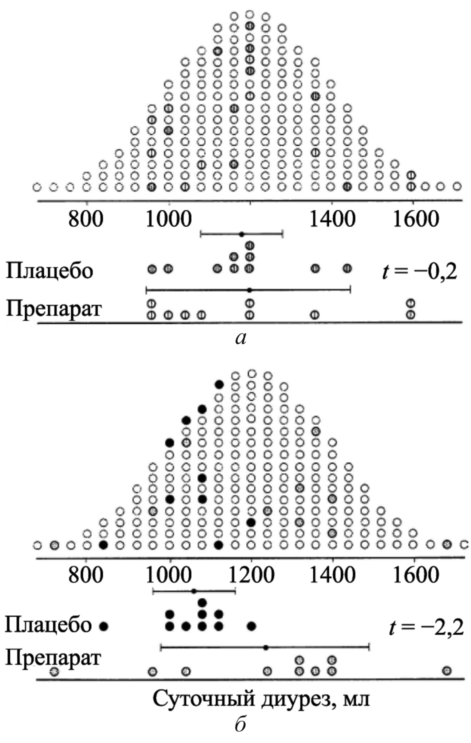


Рис. 13. Испытания предполагаемого диуретика: *а* – сравнение первой пары выборок; *б* – сравнение второй пары выборок

Разумеется, с не меньшим успехом можно было бы извлечь любую другую пару выборок, что и сделано на рис. 13, *б*. Как и следовало ожидать, две новые выборки отличаются не только друг от друга, но и от извлеченных ранее (см. рис. 13, *а*). Интересно, что на этот раз нам «повезло» – средний диурез различается сильно. Соответствующее значение t равно –2,1. В этом случае мы бы сочли препарат диуретиком.

Мы могли бы продолжить извлекать пары выборок и убедились бы, что для большинства из них значение t близко к 0. Разных пар выборок можно извлечь более 10^{27}. На рис. 14, *а* приведено распределение значений t, вычисленных по 200 парам выборок. По нему уже можно судить о распределении t. Оно симметрично относительно нуля, поскольку любую из пары выборок можно счесть первой. Как мы и предполагали, чаще всего значения t близки к нулю, значения меньше –2 и больше +2 встречаются редко.

На рис. 14, *б* видно, что в 10 случаях из 200 (в 5 % всех случаев) t меньше –2,1 или больше +2,1. Иначе говоря, если обе выборки извлечены из одной совокупности, вероятность того, что значение t лежит вне интервала от – 2,1 до +2,1, составляет 5 %. Продолжая извлекать пары выборок, будет видно, что распределение принимает форму гладкой кривой, показанной на рис. 14, *в*.

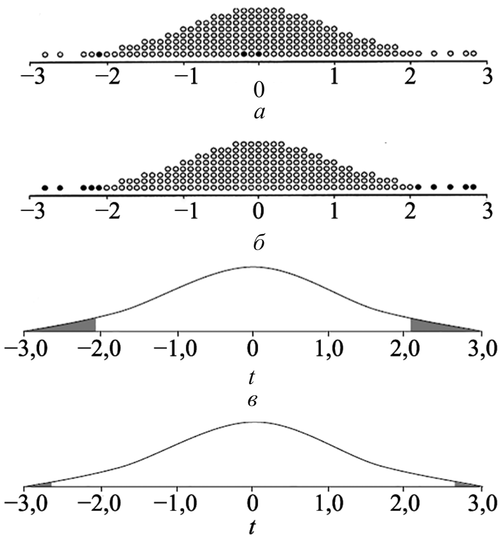


Рис. 14. Распределение Стьюдента:
а – рассчитанное по 200 парам случайных выборок по 10 членов в каждой (см. рис. 13); *б* – 5 % крайних значений t; *в* – кривая t-распределения (две заштрихованные области отсекают 5 % всей площади под кривой); *г* – площадь заштрихованной области составляет 1 % обшей площади под кривой

Так, 5 % наибольших (по абсолютной величине) отклонений от центра распределения образуют две заштрихованные области (см. рис. 14, *в*) (сумма их площадей как раз и составляет 5 % всей площади под кривой). Следовательно, большие значения t начинаются там, где начинается заштрихованная область, т. е. с $|t| \geq 2{,}1$. Вероятность получить столь высокое значение t, извлекая случайные выборки из одной совокупности, не превышает 5 %.

Описанный способ выбора критического значения t предопределяет возможность ошибки: в 5 % случаев будем находить различия там, где их нет. Чтобы снизить вероятность ошибочного заключения, можем выбрать более высокое критическое значение. К примеру, чтобы площадь заштрихованной области составляла 1 % от обшей площади под кривой, критическое значение должно составлять $|t| \geq 2{,}878$ (см. рис. 14, *г*).

Итак, мы нашли, что если две выборки извлечены из одной и той же совокупности, вероятность получить значение t больше +2,1 или меньше –2,1 составляет всего 5 %. Следовательно, если значение t находится вне интервала от –2,1 до +2,1, нулевую гипотезу следует отклонить, а наблюдаемые различия признать статистически значимыми.

Обратите внимание, что таким образом мы выявляем отличия экспериментальной группы от контрольной как в меньшую, так и большую сторону. Именно поэтому мы отвергаем нулевую гипотезу как при $t < -2{,}1$, так и $t > +2{,}1$. Данный вариант критерия Стьюдента называется двусторонним, обычно именно его и используют. Существует и односторонний вариант критерия, который применяется гораздо реже. К слову, говоря о критерии Стьюдента, мы далее будем иметь ввиду именно его двусторонний вариант.

Вернемся к рис. 13, *б*. На нем показаны две случайные выборки из одной и той же совокупности, при этом $t = -2{,}2$. Как мы только что выяснили, нам следует отвергнуть нулевую гипотезу и признать исследуемый препарат диуретиком, что неверно. Хотя все расчеты были выполнены правильно, вывод ошибочен. Такие случаи возможны.

Разберемся подробнее. Если значение t меньше –2,1 или больше +2,1, то при уровне значимости 0,05 мы сочтем различия статистически значимыми. Это означает, что если бы наши группы представляли собой две случайные выборки из одной и той же совокупности, то вероятность получить наблюдаемые различия (или более сильные) равна 0,05. Следовательно, ошибочный вывод о существовании различий будем делать в 5 % случаев. Один из таких случаев и показан на рис. 13, *б*.

Чтобы застраховаться от подобных ошибок, можно принять уровень значимости не 0,05, а, скажем, 0,01. Тогда, как видно из рис. 14, *г*, мы должны отвергать нулевую гипотезу при $t < -2{,}88$ или $t > +2{,}88$. Тогда уже мы не признаем подобные различия статистически значимыми. Однако, во-первых, ошибочные выводы о существовании различий не исключены, просто их вероятность снизилась до 1 %. Во-вторых, вероятность не найти различиий там, где они есть, теперь повысилась.

Критические значения t зависят не только от уровня значимости, но и числа степеней свободы df. Если объемы обеих выборок одинаковы, то число степеней свободы для критерия Стьюдента будет $2(n - 1)$. Таким образом, чем больше объем выборок, тем меньше критическое значение t. Это и понятно, ведь чем больше выборка, тем менее выборочные оценки зависят от случайных отклонений и тем точнее представляют исходную совокупность.

Так, выборки разного размера будут иметь разные t-распределения даже при одинаковых значениях μ и σ. Для обозначения t-распределения с k степеней свободы используется обозначение $t(k)$. Кривые плотности для $t(k)$-распределений похожи по форме на стандартные нормальные кривые, они симметричны и имеют колоколообразную форму. Однако разброс t-распределений несколько больше, чем у стандартного нормального распределения. Причиной является дополнительная изменчивость, вызванная замещением фиксированного параметра σ случайной переменной s. По мере увеличения числа степеней свободы k кривая плотности $t(k)$ становится все более похожа на нормальную кривую $N(0, 1)$. Это отражает факт, что s приближается по значению к σ по мере увеличения размера выборки.

На рис. 15 сравниваются кривые плотности нормального стандартного распределения и t-распределение с пятью степенями свободы. Сходство по форме очевидно, так же как и факт, что t-распределение имеет бо́льшие значения в хвостах и меньшие в центре по сравнению со стандартным нормальным распределением.

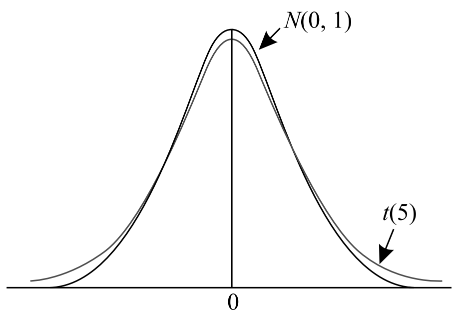


Рис. 15. Кривые плотности для стандартного нормального и $t(5)$ распределений

Во многих пособиях по статистике в приложениях приведены таблицы критических величин для разных критериев, в том числе и для t-распределения. Как и в случае с проверкой нормальности распределений, программное обеспечение (в частности, программа *Statistica*) позволяет не использовать такие таблицы, выводя на экран итоговое значение P (справедливость вероятности нулевой гипотезы), которое нужно сравнить с требуемым уровнем значимости α.

Пошаговое рассмотрение процедуры сравнения двух независимых групп в программе *Statistica* представлено в прил. 2.

3.1.2. Тест Стьюдента для одиночной выборки (сравнение с константой)

С помощью t-распределений мы также можем анализировать выборки из нормально распределенных генеральных совокупностей, сравнивая выборочную среднюю не с другой выборочной средней, а с определенной константой.

Предположим, одиночная случайная выборка размера n взята из генеральной совокупности с неизвестным средним μ. Чтобы проверить гипотезу H_0: $\mu = \mu_0$, основанную на одиночной случайной выборке размера n, рассчитаем t-статистику одиночной выборки:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}.$$

При условии, что случайная переменная T имеет распределение t для $df = n - 1$, для проверки гипотезы H_0 имеем следующие варианты альтернативной гипотезы, отображенные на рис. 16.

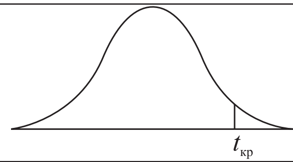
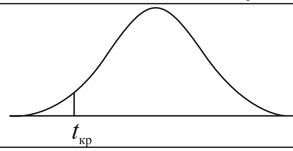
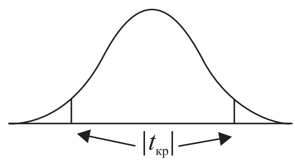
Если H_a: $\mu > \mu_{0,}$ то $T_{эмп} \geq t_{кр}$	
Если H_a: $\mu < \mu_{0,}$ то $T_{эмп} \leq t_{кр}$	
Объединяя два приведенных неравенства: если H_a: $\mu \neq \mu_{0,}$ то $T_{эмп} \geq \lvert t_{кр} \rvert$	

Рис. 16. Варианты альтернативной гипотезы

Первые два варианта на рис. 16 отражают односторонние варианты теста, если мы знаем, в какую сторону возможны отклонения, в ином случае следует применять двусторонний вариант. Напомним, что программы статистического анализа также приводят величину P (вероятности справедливости нулевой гипотезы), тогда H_A: $P \le \alpha$. Рассчитанные таким образом значения P абсолютно точны в генеральных совокупностях с нормальным распределением и приблизительно точны для больших выборок в других случаях.

Пример. В соответствии со спецификацией для приготовления препарата витамин C в микстуре должен содержаться в готовом продукте в количестве 40 мг на 100 г. Эти нормы разработаны для среднего μ содержания витамина C. Мы можем проверить нулевую гипотезу о том, что среднее содержание витамина C в партии соответствует данным спецификации, иначе говоря, мы проверяем, что

$$H_0: \mu = 40,$$
$$H_A: \mu \ne 40.$$

При этом в исследованной нами выборке $n = 8$; $\bar{x} = 22{,}5$; s = 7,2.

Расчетная величина теста Стьюдента вычисляется следующим образом:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{22{,}5 - 40}{7{,}2/\sqrt{8}} = -6{,}88.$$

Поскольку число степеней свободы $n - 1 = 7$, то в данном случае мы имеем дело с $t(7)$-распределением. На рис. 17 показано значение P в случае анализа нашей выборки ($t \ge |6{,}88|$).

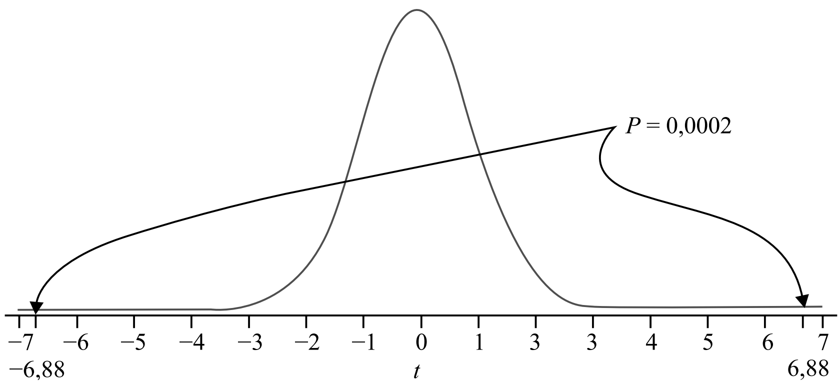


Рис. 17. t(7)-распределение и значения P для примера с витамином C

Из таблицы критических значений (которые часто приводятся в пособиях) можно увидеть, что уровню значимости 0,05 соответствует $t_{кр} = |5{,}408|$. Так как мы получили t меньше –5,408, следовательно, при двустороннем варианте теста значение P будет меньше, чем $2 \cdot 0{,}0005$. Программное обеспечение высчитывает точное значение $P = 0{,}0002$. Очевидно, что данные теста не подтверждают соответствие со средним μ 40 мг на 100 г. Таким образом, мы отвергаем H_0 и с очень большой долей вероятности заключаем, что содержание витамина *C* в данной партии не соответствует указанному в спецификации.

В примере мы проверили соответствие экспериментальных данных нормам спецификации ($\mu = 40$ мг на 100 г) относительно двусторонней альтернативы ($\mu \neq 40$ мг на 100 г), поскольку у нас не было подозрений, что в партии может быть слишком много или слишком мало витамина *C*. Если бы мы заранее знали, что в смесь было внесено необходимое количество витамина *C*, но подозревали, что некоторая его часть могла быть утеряна или уничтожена в процессе производства, то могли бы использовать односторонний тест, однако, если такой уверенности нет, лучше применять двусторонний.

Если бы для рассмотренного примера мы хотели проверить, был ли витамин *C* потерян или уничтожен в процессе производства, мы бы проверяли H_0: $\mu = 40$ против H_A: $\mu < 40$. Результат t-теста не изменяется, $t = -6{,}88$. Однако так как мы теперь рассматриваем отклонения только в сторону меньших значений, то значение P будет равно половине значения из предыдущего примера (программное обеспечение дает $P = 0{,}0001$). Таким образом, можно сделать вывод, что в процессе производства некоторое количество витамина *C* было утеряно или разрушено. Напомним, что выводы, сделанные в данном примере, основаны на предположении, что в смесь добавлено строго необходимое количество витамина *C*.

Пошаговое рассмотрение процедуры сравнения выборки с константой в программе *Statistica* представлено в прил. 4.

3.1.3. Сравнение двух зависимых групп (парный тест Стьюдента)

С зависимыми выборками исследователь имеет дело каждый раз, когда измерения значений изучаемого признака выполняются дважды на одних и тех же объектах. При анализе таких данных получаются парные результаты: наблюдение одинаковых объектов до и после воздействия.

Пример. Проведен цикл занятий по повышению навыков владения иностранным языком преподавателями. В начале мероприятия педагоги прошли тест на понимание разговорного английского. После завершения цикла занятий тест был проведен снова. В табл. 9 показаны результаты первого и второго тестов. Максимально возможная оценка в тесте – 36 баллов.

Таблица 9

Результаты тестирования уровня владения английским языком преподавателей

Преподаватель	1-й тест	2-й тест	Изменение	Преподаватель	1-й тест	2-й тест	Изменение
1	32	34	2	11	30	36	6
2	31	31	0	12	20	26	6
3	29	35	6	13	24	27	3
4	10	16	6	14	24	24	0
5	30	33	3	15	31	32	1
6	33	36	3	16	30	31	1
7	22	24	2	17	15	15	0
8	25	28	3	18	32	34	2
9	32	26	–6	19	23	26	3
10	20	26	6	20	23	26	3

Для анализа данных отнимем от результатов второго теста результаты первого, чтобы увидеть прогресс каждого участника. Чтобы определить, действительно ли занятия улучшили понимание разговорного английского языка у преподавателей, сделаем проверку:

$$H_0\text{: } \mu = 0 \text{ против } H_A\text{: } \mu > 0.$$

Здесь μ – это среднее изменение, которое могло бы быть достигнуто, если бы все преподаватели в мире прошли данные занятия английского. Нулевая гипотеза говорит, что изменений не произойдет, а альтернативная – что результаты второго теста в среднем будут выше.

Увеличение числа набранных баллов (разность между результатами второго и первого тестов) для 20 тестируемых имеют следующие характеристики: $\bar{x} = 2{,}5$ и $s = 2{,}9$.

Применим к полученным данным расчет значения t, аналогично тому, как мы делали в предыдущем разделе для одиночной выборки:

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{2{,}5}{2{,}893/\sqrt{20}} = 3{,}88.$$

Значение P находим на этот раз из t(19)-распределения (помним, что количество степеней свободы меньше размера выборки на 1). Если смотреть по таблицам критических значений, то что для t(19)-распределения число 3,86 находится между верхним 0,001 и нижним 0,0005 пределами для величины P. Компьютерные программы статистической обработки данных дают точное значение $P = 0{,}00053$. Очевидно, что возрастание оценок с большой долей вероятности нельзя объяснить только случайностью, и что занятия были эффективны в повышении числа набранных в тесте баллов. В публикациях рутинные детали статистических процедур обычно опускают, поэтому рассмотренный тест был бы представлен в форме «Улучшение оценок статистически значимо (t = 3,86; df = 19; P = 0,00053)».

Результаты теста позволяют сделать вывод, что в среднем оценки возросли. Однако статистически значимое, но небольшое увеличение будет недостаточным основанием утверждать, что занятия значительно улучшили навыки участников. Для количественной оценки того, насколько улучшились результаты, суждение лучше дополнить доверительным интервалом (см. разд. 6).

Пошаговое рассмотрение процедуры сравнения двух зависимых групп в программе *Statistica* представлено в прил. 3.

Дискуссионный вопрос: надежность *t*-тестов

Использование t-процедуры в рассмотренном примере (см. табл. 9) имеет некоторые сложности. Во-первых, преподаватели не являются случайной выборкой из генеральной совокупности, поскольку имел место некоторый отбор в пользу мотивированных и энергичных людей, готовых потратить время на занятия. Следовательно, не совсем понятно, к какой генеральной совокупности применимы результаты. Во-вторых, внимательно посмотрев на данные, видно, что некоторые учителя по первому тесту получили оценки близкие к максимальным, а значит, они не могли сильно увеличить их, даже если личные навыки преподавателей значительно улучшились после занятий. В этом слабость системы оценивания, которая использовалась в качестве измерительного инструмента в данном исследовании. Разница в оценках может неправильно показывать эффективность занятий, а это могло быть причиной того, что среднее увеличение оказалось небольшим.

Последняя проблема заключается в том, что предварительный анализ данных демонстрирует их отклонение от нормальности. В ходе проведения теста мы предполагаем, что набор изменений имеет нормальное распределение, поскольку применяем к ним t-процедуру. Однако в рассмотренном примере один человек потерял 6 баллов между первым и вторым тестами и, таким образом, занизил среднее увеличение для группы с 2,95 до 2,5. С другой стороны, это резко отличающееся значение (−6) можно рассматривать как выброс.

Очевидно, что t-тесты не надежны в присутствии выбросов, поскольку последние сильно влияют на $\bar{x}$ и s. Если мы удалим выброс в рассмотренном примере, значение t изменится от 3,86 до 5,98, и величина P станет гораздо меньше. Другими словами, в данном случае выброс делает результаты теста существенно менее значимыми.

* * *

Итак, нулевая гипотеза, проверяемая в ходе t-теста, заключается в том, что *обе сравниваемые группы происходят из одной генеральной совокупно-*

сти, т. е. наблюдаемые различия между средними значениями сравниваемых выборок случайны и не вызваны действием изучаемого фактора. Рассмотренные варианты *t*-теста относятся к группе параметрических методов анализа. Их корректное применение требует, чтобы выборки подчинялись нормальному закону распределения. Если значения признака хотя бы в одной из сравниваемых групп распределены ненормально, применение параметрического *t*-теста для их сравнения будет приводить к искаженным результатам.

Однако *t*-тесты достаточно устойчивы к умеренной ненормальности распределения генеральной совокупности за исключением выбросов или сильных смещений. При этом большие выборки улучшают точность *P* и критических величин *t*-распределений даже в ситуациях, когда генеральная совокупность не подчиняется закону нормального распределения. Следовательно, мы можем быть менее обеспокоены нормальностью, когда выборка большая. По мере увеличения выборки ее стандартное отклонение s приближается к стандартному отклонению σ для генеральной совокупности (связано с законом больших чисел), поэтому в больших выборках s будет точно предсказывать σ вне зависимости от формы распределения.

В малых выборках для проверки на наличие смещения распределения или выбросов наряду с рассмотренными в подразд. 1.5.1 и прил. 1 методами можно также использовать стеблевые или точечные диаграммы. Для большинства задач *t*-тест может быть применен при достаточно большом объеме выборки $n \geq 30$, за исключением случаев, когда присутствуют сильные выбросы или ярко выраженные смещения.

В этом разделе мы не обсуждали, как проводить сравнительный анализ при очевидно нормальном распределении, основываясь на небольшой выборке. Наиболее простой способ работать с такими данными – использовать специальные непараметрические тесты. Однако они имеют два недостатка. Во-первых, у них меньше мощности, чем у тестов, разработанных для конкретных типов распределений (как, например, *t*-тесты). Во-вторых, нам часто необходимо модифицировать положения гипотезы для использования тестов, не зависящих от распределения. В таких тестах, например, центр распределения обычно определяется относительно медианы, нежели среднего (это важно, когда распределение имеет смещение). Поэтому непараметрический тест не задает вопрос «Изменилось ли среднее?» как его задает *t*-тест.

3.2. СРАВНЕНИЕ ДВУХ И БОЛЕЕ ГРУПП

Мы рассмотрели, как оценить различия между двумя выборками с помощью разнообразных методов попарного сравнения. Однако в исследовательской работе часто встречаются задачи, когда приходится сравнивать не одну группу с другой, а одновременно несколько групп. К примеру, стоит задача оценить различия в развитии эмбрионов (в днях) у кроликов разных пород: шиншилл, голландских и польских. Проводя попарные сравнения значений групп, возрастает риск ошибки 1-го рода и, как следствие, имеет место высокая вероятность некорректного заключения.

Английский математик и генетик Р. Фишер в 20-х гг. XX в. предложил метод комплексной оценки сравниваемых средних, получивший название дисперсионного анализа. Дисперсионный анализ (*ANOVA* (*Analysis Of Variance*)) устанавливает роль отдельных факторов в изменчивости того или иного признака. Влияние одного фактора на изучаемый признак сложно выделить в чистом виде. И хотя при проведении опытов стараются сохранить условия максимально однородными, различные опыты обычно дают неодинаковые результаты из-за многих других факторов, меняющихся от опыта к опыту и не поддающихся контролю. В исследовании, предполагающем изучение влияния каких-либо воздействий на тот или иной признак, практически невозможно, а нередко и не нужно, учитывать влияние одновременно всех действующих факторов. Обычно исследователь изучает ведущие факторы, наиболее влияющие на признак. Неучитываемые факторы называют случайными или остаточными. Именно общая изменчивость признака в выборках определяется суммарным воздействием *учитываемых* и *остаточных факторов*. Поэтому и возникает задача разложения общей изменчивости признака на составные части, с одной стороны, определяемые изучаемыми конкретными факторами, с другой – вызванные случайными, неконтролируемыми причинами.

Рассмотрим общий принцип дисперсионного анализа на примере данных эксперимента, проведенного на 20 поросятах четырехмесячного возраста. Поросята случайным образом были разбиты на четыре группы по пять животных. Каждую группу кормили определенным кормом (табл. 10). Экспериментатора интересует, отражают ли эти данные различия в рационах и можно ли какой-либо из них рекомендовать в качестве предпочтительного.

Таблица 10

Результаты эксперимента

	Корм 1	Корм 2	Корм 3	Корм 4
	$m = 4$ (число групп)			
$n = 5$ (число наблюдений в каждой группе)	60,8	68,7	102,6	87,9
	57	67,7	102,1	84,2
	65	74	100,2	83,1
	58,6	66,3	96,5	85,7
	61,7	69,8	96,5	90,3

Если изучают воздействие на признак одного фактора, дисперсионный анализ будет *однофакторным*, если одновременно исследуют на признак действие двух, трех или большего числа регулируемых факторов, то дисперсионный анализ называют *двух-*, *трех-* и *многофакторным*. Таким образом, задачей дисперсионного анализа является оценка влияния одного или нескольких факторов на исследуемый признак. В нашем примере мы изучаем воздействие одного фактора – рациона питания, поэтому будем выполнять однофакторный дисперсионный анализ.

Анализ данных мы начинаем с формулировки нулевой гипотезы. В данном случае она заключается в том, что ни одна из диет не влияет на набор веса у поросят, т. е. все групповые средние равны.

Вспомним, что на основе выборочной средней мы характеризуем среднюю генеральной совокупности μ, поэтому нулевую гипотезу в данном случае можно записать H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$ (в общем виде H_0: $\mu_1 = \mu_2 \ldots = \mu_k$, где k – число сравниваемых групп). Если все наши выборки принадлежат одной и той же генеральной совокупности, значит, каждая выборочная средняя характеризует одну и ту же среднюю μ.

Отклоняя нулевую гипотезу и принимая альтернативную о влиянии изучаемого фактора, мы не можем сказать, между какими группами есть различия. Мы всего лишь делаем вывод, что, по крайней мере, существует хотя бы одна пара сравниваемых групп, принадлежащих к разным генеральным совокупностям. Поэтому альтернативную гипотезу при дисперсионном анализе записывают H_1: $\mu_i \neq \mu_j$.

Таким образом, нулевая гипотеза сводится к предположению, что и генеральные межгрупповые средние, и дисперсии равны между собой, а это значит, что различия, наблюдаемые между выборочными показателями, вызваны случайными факторами, а не влиянием изучаемого фактора – корма животного. Это предположение проверяется при помощи F-критерия, который соотносит две дисперсии – дисперсию изучаемой совокупности, оцененную по выборочным средним (межгрупповая дисперсия, $s^2_{\text{меж}}$), и дисперсию изучаемой совокупности, оцененную по выборочным дисперсиям (внутригрупповая дисперсия, $s^2_{\text{вну}}$):

$$F = s^2_{\text{меж}} \,/\, s^2_{\text{вну}}.$$

При нулевой гипотезе о равенстве групповых средних межгрупповая дисперсия будет подобна внутригрупповой. Если же между группами есть различия, то межгрупповая дисперсия будет больше, чем внутригрупповая.

К слову, перед проведением дисперсионного анализа необходимо проверить несколько предположений. Так, F-критерий относится к параметрическим тестам, которые применимы лишь в случаях, когда анализируемые выборки (каждая) распределены нормально, и дисперсия во всех группах одинакова. Объем выборки должен быть достаточный, чтобы проверить эти допущения (см. прил. 5).

Положим, для рассматриваемого примера все предположения выполняются. Проверка выдвинутой нулевой гипотезы будет состоять из следующих этапов:

- вычисление эмпирического значения критерия $F_{\text{эмп}}$;
- вычисление числа степеней свободы. Определение уровня значимости. Нахождение критического значения $F_{\text{кр}}$ (по специальным таблицам или в программе *MS Excel*);
- сравнение расчетного эмпирического значения статистики критерия $F_{\text{эмп}}$ с критическим значением $F_{\text{кр}}$. Принятие или опровержение нулевой гипотезы.

Шаг 1. Вычисление эмпирического значения критерия $F_{\text{эмп}}$. Для вычисления эмпирического значения критерия необходимо воспользоваться формулой

$$F = s_{\text{меж}}^2 / s_{\text{вну}}^2.$$

где $s_{\text{меж}}^2$ – межгрупповая дисперсия, характеризующая вариативность групповых средних; $s_{\text{вну}}^2$ – внутригрупповая дисперсия, характеризующая вариативность признака в каждой группе.

Если ни одна из диет не влияет на набор веса у поросят, то все четыре выборки относятся к одной генеральной совокупности, а значит, любая из четырех дисперсий дает одинаково хорошую оценку вариации. Поэтому в качестве оценки дисперсии совокупности возьмем среднее выборочных дисперсий. Эта оценка называется внутригрупповой дисперсией:

$$s_{\text{вну}}^2 = \frac{1}{4}\left(s_{\text{корм1}}^2 + s_{\text{корм2}}^2 + s_{\text{корм3}}^2 + s_{\text{корм4}}^2\right).$$

Вспомним, что выборочная дисперсия вычисляется по формуле

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

Произведем расчет внутригрупповой дисперсии, подставив в приведенную выше формулу рассчитанные величины дисперсии каждой группы (табл. 11):

$$s_{\text{вну}}^2 = \frac{1}{4}(9,39 + 8,57 + 8,71 + 8,39) = 8,76.$$

Таблица 11

Величины дисперсии каждой группы

Тип корма	Корм 1	Корм 2	Корм 3	Корм 4
x_{i}	60,8	68,7	102,6	87,9
	57,0	67,7	102,1	84,2
	65,0	74,0	100,2	83,1
	58,6	66,3	96,5	85,7
	61,7	69,8	96,5	90,3
$\bar{x}$	60,62	69,3	99,58	86,24
$\sum\left(\bar{x}_i - \bar{x}\right)^2$	37,57	34,3	34,8	33,5
n	5	5	5	5
s^2	9,39	8,57	8,71	8,39

Дисперсия для каждой группы вычислялась относительно среднего этой группы. Поэтому внутригрупповая дисперсия не зависит от того, насколько различаются средние изучаемых групп.

Оценим теперь дисперсию совокупности по выборочным средним, т. е. *межгрупповую*. Так как мы предположили, что все четыре выборки относятся к одной генеральной совокупности, то стандартное отклонение четырех выборочных средних является оценкой стандартной ошибки средней. Стандартная ошибка средней следующим образом связана со стандартным отклонением и объемом совокупности

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}.$$

Поэтому дисперсию совокупности можно рассчитать как

$$\sigma^2 = n\sigma_{\bar{x}}^2.$$

Воспользуемся этим, чтобы оценить дисперсию совокупности по разбросу значений групповых средних – межгрупповую дисперсию (табл. 12):

$$s_{\text{меж}}^2 = ns_{\bar{x}}^2,$$

где $s_{\bar{x}}^2$ – квадрат стандартного отклонения (дисперсия) выборки из четырех выборочных средних; n – число наблюдений в группе.

Таблица 12

Величины межгрупповой дисперсии

<table>
<tr><th>Тип корма</th><th>Корм 1</th><th>Корм 2</th><th>Корм 3</th><th>Корм 4</th><th>Квадрат стандартного отклонения выборки из четырех выборочных средних –</th></tr>
<tr><td rowspan="5">x_i</td><td>60,8</td><td>68,7</td><td>102,6</td><td>87,9</td><td rowspan="7">302,62</td></tr>
<tr><td>57,0</td><td>67,7</td><td>102,1</td><td>84,2</td></tr>
<tr><td>65,0</td><td>74,0</td><td>100,2</td><td>83,1</td></tr>
<tr><td>58,6</td><td>66,3</td><td>96,5</td><td>85,7</td></tr>
<tr><td>61,7</td><td>69,8</td><td>96,5</td><td>90,3</td></tr>
<tr><td>$\bar{x}$</td><td>60,62</td><td>69,3</td><td>99,58</td><td>86,24</td></tr>
<tr><td>n</td><td>5</td><td>5</td><td>5</td><td>5</td></tr>
</table>

$$s^2_{\text{меж}} = 5 \cdot 302{,}62 = 1513{,}1.$$

Вычислим значение $F_{\text{эмп}}$:

$$F_{\text{эмп}} = s^2_{\text{меж}} \,/\, s^2_{\text{вну}} = 1513{,}1/8{,}76 = 172{,}67.$$

Шаг 2. Определение уровня значимости. Вычисление числа степеней свободы. Нахождение критического значения $F_{\text{кр}}$. По данным эксперимента мы определили эмпирическое значение критерия F – $F_{\text{эмп}}$. Для принятия или опровержения выдвинутой нулевой гипотезы необходимо сравнить полученную величину с критическим значением – $F_{\text{кр}}$.

Критическое значение $F_{\text{кр}}$, начиная с которого нулевую гипотезу отклоняют, определяется уровнем значимости (обычно $\alpha = 0{,}05$; задается исследователем) и числом степеней свободы для каждой из сравниваемых дисперсий.

Для межгрупповой дисперсии рассчитывают *межгрупповое число степеней свободы* по формуле

$$v_{\text{меж}} = m - 1,$$

где m – число сравниваемых групп. В нашем примере четыре такие группы, значит, межгрупповое число степеней свободы равно трем.

Для внутригрупповой дисперсии также вычисляют *внутригрупповое число степеней свободы* по формуле

$$v_{\text{вну}} = m(n - 1),$$

где m – число сравниваемых групп; n – объем каждой группы.

В данном примере $v_{\text{вну}} = 4(5 - 1) = 16$.

Критическое значение $F_{\text{кр}}$ можно найти в специальных таблицах или рассчитать в программе *MS Excel*. Рассмотрим, как это можно сделать при помощи соответствующих таблиц.

Используя рассчитанные величины числа степеней свободы сравниваемых дисперсий ($\nu_{\text{меж}} = 3$ и $\nu_{\text{вну}} = 16$) по табл. 13, определяем критическое значение $F_{\text{кр}}$: для $\alpha = 0{,}05$ оно составило 3,24.

Таблица 13

Критические значения F для $\alpha = 0{,}05$

$v_{\text{вну}}$	$v_{\text{меж}}$									
	1	2	3	4	5	6	7	8	9	10
1	161	199	216	225	230	234	237	239	241	242
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96

Окончание табл. 13

$v_{вну}$	$v_{меж}$									
	1	2	3	4	5	6	7	8	9	10
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	$\boxed{3,24}$	3,01	2,85	2,74	2,66	2,59	2,54	2,49

В большинстве пакетов для прикладного статистического анализа (например, *Statistica*) программа самостоятельно находит величину критического значения используемого теста, но не всегда выводит ее на экран в итоговой статистике рабочей книги.

Шаг 3. Сравнение расчетного эмпирического значения статистики критерия $F_{эмп}$ с критическим значением $F_{кр}$. Принятие или опровержение нулевой гипотезы. Расчетная величина эмпирического значения $F_{эмп}$ (172,67) больше критического значения $F_{кр}$ (3,24). Мы отвергаем нулевую гипотезу и делаем вывод, что диета влияет на набор веса у поросят.

Варианты проведения дисперсионного анализа с использованием программы *Excel* представлен в разд. 3.5, программы *Statistica* – в прил. 5 и 6.

Таким образом, аппарат дисперсионного анализа вполне применим, если распределение куполообразно и не резко асимметрично, а групповые дисперсии различаются не более чем в 1,5–2,5 раза. По-видимому, допустимы и отклонения больше, но при условии равенства числа наблюдений в группах в однофакторном анализе.

3.3. ОШИБКИ В ИСПОЛЬЗОВАНИИ КРИТЕРИЯ СТЬЮДЕНТА

Критерий Стьюдента предназначен для сравнения двух групп. Однако на практике он широко (что неправильно) используется для оценки различий большого числа групп посредством попарного их сравнения. При этом вступает в силу *эффект множественных сравнений*.

Пример. Анализируют влияние препаратов А и Б на уровень глюкозы плазмы. Исследование проводят на трех группах: получавших препарат (А или Б) и получавших плацебо (В). С помощью критерия Стьюдента проводят три парных сравнения: группу А сравнивают с группой В, группу Б – с группой В и, наконец, А с Б. Получив достаточно высокое значение t в каком-либо из трех сравнений, сообщают, что $P \leq 0{,}05$. Это означает, что вероятность ошибочного заключения о существовании различий не превышает 5 %, что неверно: вероятность ошибки значительно превышает 5 %.

Разберемся подробнее. В исследовании был принят пятипроцентный уровень значимости. Значит, вероятность ошибиться при сравнении групп А и В – 5 %. Кажется, все правильно, но точно так же мы ошибемся в 5 % случаев при сопоставлении групп Б и В. И, наконец, при сравнении групп А и Б ошибка также возможна в 5 % случаев. Следовательно, вероятность ошибиться хотя бы в одном из трех сравнений составит не 5 %, а значительно больше. В общем случае эта вероятность

$$P' = 1 - (1 - 0{,}05)^k,$$

где k – число сравнений.

При небольшом числе сравнений можно использовать приближенную формулу

$$P' = 0{,}05k,$$

т. е. вероятность ошибиться хотя бы в одном из сравнений примерно равна вероятности ошибиться в одном сравнении, умноженная на число сравнений.

Итак, в нашем исследовании вероятность ошибиться хотя бы в одном из сравнений составляет примерно 15 %. При сравнении четырех групп число пар и, соответственно, возможных попарных сравнений равно 6, поэтому при уровне значимости 0,05 в каждом из сравнений вероятность неверно обнаружить различие хотя бы в одном равна уже не 0,05, а примерно $6 \cdot 0{,}05 = 0{,}30$. И когда исследователь, выявив таким способом эффективный препарат, будет говорить про вероятность ошибки 5 %, в действительности она будет равна 30 %.

Предположим, мы имеем результаты сравнения трех выборок растений (по 10 измерений в каждой) из одной генеральной совокупности. На рис. 18 результаты представлены в виде, принятом в научных публикациях. Напомним, что столбиками изображены выборочные средние, вертикальные черточки «усов» задают интервалы в плюс-минус одну стандартную ошибку среднего. Нередко (и неправильно) исследователи приступают к попарному сравнению таких групп с помощью критерия Стьюдента и получают следующие значения: 2,39 (при сравнении 1-й и 2-й группы), 0,93 (2-й и 3-й группы) и 1,34 (1-й и 3-й группы). Так как в каждом сравнении участвуют 2 группы по 10 растений, число степеней свободы равно 2(10 – 1) = 18. По таблице критических значений t находим, что при таком значении числа степеней свободы и уровне значимости 5 % критическое значение t равно 2,101. Таким образом, нам пришлось бы заключить, что в первой паре сравнений одна из групп отличалась бы по росту, в то время как остальные сравнения не показывали бы различий.

Задумайтесь над этим результатом. Что в нем не так?

Если растения 3-й группы дали результаты, не отличающиеся от 1-й и 2-й, то как две последние оказались различны? К сожалению, столь странный вывод часто не смущает исследователей, а лишь вдохновляет на создание изощренного раздела «Обсуждение результатов».

Дисперсионный анализ приведенных данных, который нужно применять в такой ситуации, дает значение $F = 2{,}74$ (число степеней свободы $\nu_{меж} = m - 1 = 3 - 1 = 2$ и $\nu_{вну} = m\,(n - 1) = 3\,(10 - 1) = 27$). Критическое значение F для пятипроцентного уровня значимости равно 3,35, т. е. превышает полученное. Таким образом, дисперсионный анализ говорит об отсутствии различий между группами.

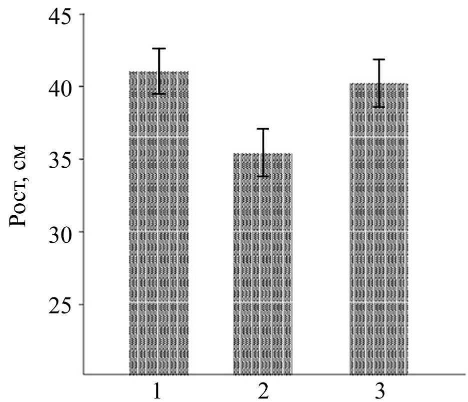


Рис. 18. Рост растений в трех выборках из одной генеральной совокупности

В заключение приведем три правила:

• критерий Стьюдента может быть использован для проверки гипотезы о различии средних только для двух групп;

• если схема эксперимента предполагает большое число групп, необходимо применить дисперсионный анализ;

• если критерий Стьюдента был применен исследователем для проверки различий между несколькими группами без введения поправки Бонферрони (см. разд. 3.4), то истинный уровень значимости можно получить, умножив приведенный уровень значимости, на число возможных сравнений.

3.4. МНОЖЕСТВЕННЫЕ СРАВНЕНИЯ

В предыдущем разделе мы познакомились с эффектом множественных сравнений. Он состоит в том, что при многократном применении критерия Стьюдента вероятность ошибочно найти различия там, где их нет, возрастает. Если исследуемых групп больше двух, то следует воспользоваться дисперсионным анализом. Однако он позволяет проверить лишь гипотезу о равенстве всех средних. Если она не подтверждается, нельзя узнать, какая именно группа отличается от других.

Это позволяют сделать специальные *методы множественного сравнения*, которые основаны на критерии Стьюдента, но учитывают, что сравнивается более одной пары средних. Сразу поясним, когда следует использовать данные методы. Подход состоит в том, чтобы в первую очередь с помощью дисперсионного анализа проверить нулевую гипотезу о равенстве всех средних, а уже затем, если нулевая гипотеза отвергнута, выделить среди них отличные от остальных, используя методы множественного сравнения. Заметим, что некоторые авторы считают этап дисперсионного анализа излишним и предлагают сразу применять методы множественного сравнения, так как они показывают наличие различий между группами.

Простейший из методов множественного сравнения – введение *поправки Бонферрони*. Как уже было показано, при трехкратном применении критерия Стьюдента с пятипроцентным уровнем значимости, вероятность обнаружить различия там, где их нет, составляет не 5 %, а $3 \cdot 5 = 15$ %. Из чего следует, что если мы хотим обеспечить вероятность ошибки α, то в каждом из сравнений нужно принять уровень значимости α/k, что и есть поправка Бонферрони. К примеру, при трехкратном сравнении уровень значимости должен быть $0{,}05/3 = 1{,}7$ %.

Поправка Бонферрони хорошо работает, если число попарных сравнений невелико. Если оно превышает восемь, метод становится слишком «строгим», и даже весьма большие различия приходится признавать статистически незначимыми.

Конечно, существуют и не такие жесткие методы множественного сравнения, которые схожи с поправкой Бонферрони в том, что, будучи модификацией критерия Стьюдента, учитывают многократность сравнений.

При большом числе сравнений часто используют такие критерии, как *Ньюмена – Кейлса*, *Тьюки*, *Даннета* (вариант критерия Ньюмена – Кейлса для сравнения нескольких групп с одной контрольной). Более подробно с ними можно ознакомиться, например, в книге С. Гланца (1999).

3.5. СРАВНЕНИЕ ДВУХ И БОЛЕЕ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *MICROSOFT EXCEL*

Так как все рассматриваемые в этом разделе тесты параметрические, то они применимы лишь в случаях, когда анализируемые выборки (каждая) распределены нормально, и дисперсия во всех группах одинакова. Объем выборки должен быть достаточный, чтобы проверить эти предположения. Некоторые исследователи не рекомендуют применять параметрические методы, если объем каждой из групп составляет менее 30 наблюдений, даже если выборочные данные имеют нормальное распределение.

Допустим, что в учебных целях для рассматриваемых примеров все предположения выполняются.

3.5.1. Сравнение двух независимых групп: тесты Фишера и Стьюдента

Задача. Изучали влияние различных кормов на прирост 10 поросят. Получены следующие данные о весе изученных животных каждой группы (табл. 14).

Таблица 14

Вес животных каждой группы, кг

Корм 1	60,8	57	65	58,6	61,7
Корм 2	68,7	67,7	74	66,3	69,8

Отражают ли полученные данные различия в рационах и можно ли какой-либо из них рекомендовать в качестве предпочтительного?

Пакет анализа в *Excel* содержит два способа сравнения двух независимых групп:

• двухвыборочный t-тест с одинаковыми дисперсиями;

• двухвыборочный t-тест с различными дисперсиями.

Вначале необходимо выполнить тест Фишера, чтобы определить, какой тест из двух можно использовать.

Шаг 1. Ввод данных. Введите данные по предложенному образцу (рис. 19).

	A	B	C	D
1		Корм 1	Корм 2	
2		60,8	68,7	
3		57	67,7	
4		65	74	
5		58,6	66,3	
6		61,7	69,8	
7	дисперсия			
8				

Рис. 19. Ввод данных

Шаг 2. Вычисление дисперсий сравниваемых выборок. Формулировка нулевой гипотезы. Тест Фишера. Выполним тест Фишера (F-тест), чтобы определить, какой вариант из двух (двухвыборочный t-тест с одинаковыми дисперсиями или двухвыборочный t-тест с различными дисперсиями) можно использовать.

В данном случае F-тест представляет собой отношение дисперсий сравниваемых выборок:

$$F = \frac{s_1^2}{s_2^2}.$$

Проверяется нулевая гипотеза H_0: $\sigma_1^2 = \sigma_2^2$. При расчете теста Фишера в числитель помещают дисперсию с большим значением.

Рассчитаем дисперсию для каждой группы. Для этого введем функцию *ДИСП* (*MS Excel* 2003) или *ДИСП.В* (*MS Excel* 2010) в ячейку *B7*. Ввод функции можно осуществить несколькими способами, например нажав кнопку f_x (вставка функции) или сочетание клавиш *Shift* + *F3*. В появившимся диалоговом окне *Мастер функций – шаг 1 из 2* в поле *Категория* выбираем *Статистические* и, найдя функцию *ДИСП.В*, нажимаем на нее (рис. 20), затем нажимаем кнопку *ОК*.

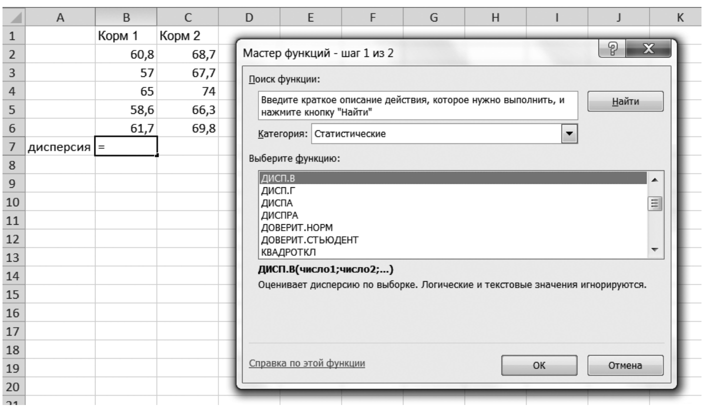


Рис. 20. Вставка функции

Возникнет новое диалоговое окно, в которое вводим диапазон данных выборки «Корм 1» (рис. 21), затем нажимаем кнопку *ОК*. В текущей ячейке появится число 9,4. Аналогично вычисляем дисперсию второй выборки в ячейке *C7*, величина которой будет равна 8,6.

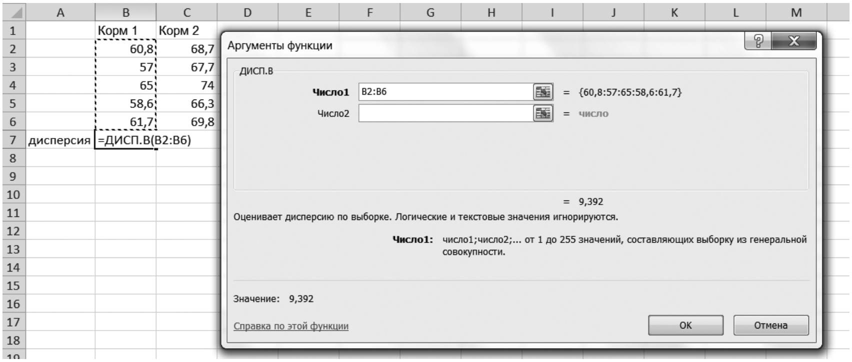


Рис. 21. Вычисление дисперсии

Оценим, есть ли достоверные отличия между дисперсиями двух групп. Нулевая гипотеза: дисперсии двух групп не отличаются. Краткая запись нулевой и альтернативной гипотез следующая:

- H_0: $\sigma_1^2 = \sigma_2^2$, $F_{\text{эмп}} < F_{\text{кр}}$, $P > 0{,}05$ – дисперсии не отличаются;
- H_1: $\sigma_1^2 \neq \sigma_2^2$, $F_{\text{эмп}} \geq F_{\text{кр}}$, $P \leq 0{,}05$ – дисперсии отличаются.

Для проверки нулевой гипотезы выполним тест Фишера. Переходим на вкладку *Данные*. Находим пакет *Анализ данных* и нажимаем на него (рис. 22). В *Инструменты анализа* выбираем *Двухвыборочный F-тест для дисперсии*, нажимаем *ОК*.

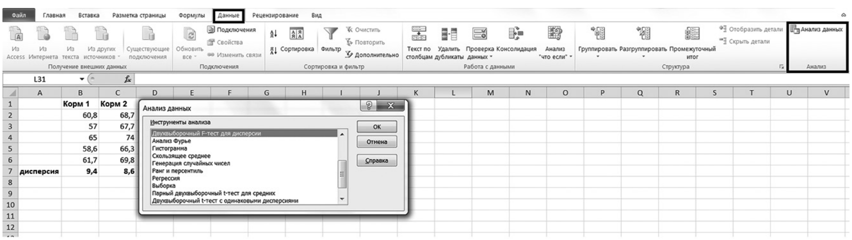


Рис. 22. Инструменты анализа

Заполняем диалоговое окно *Двухвыборочный F-тест для дисперсии* в соответствии с образцом (рис. 23). В *Интервал переменной 1* всегда вводят массив данных с большей дисперсией. Если введен массив с метками – ставят флажок в поле *Метки*. В поле *Выходной интервал* указывают место вывода результатов. Нажмите *ОК*.

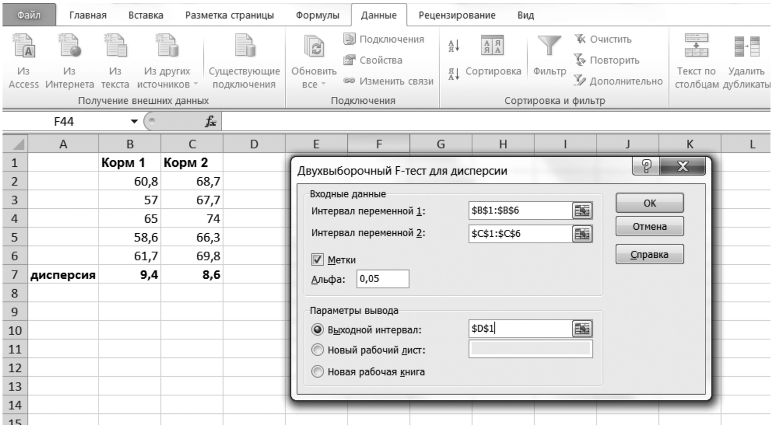


Рис. 23. Окончательный вид диалогового окна «Двухвыборочный F-тест для дисперсии»

Рассмотрим таблицу с полученными результатами. В ней приведены средние для обеих выборок, дисперсии, количество наблюдений, число степеней свободы, вычисленное значение F ($F_{эмп}$), вероятность справедливости гипотезы о равенстве дисперсий и F-критическое ($F_{кр}$) (рис. 24).

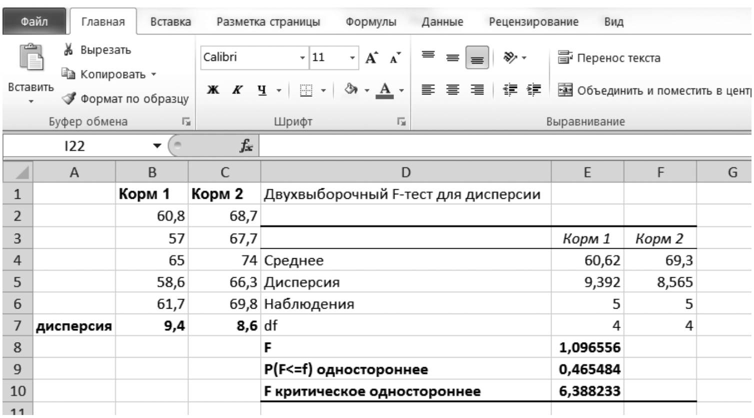


Рис. 24. Результат теста Фишера

Расчетная величина F (1,1) меньше F-критического (6,4) с вероятностью 0,46 ($P > 0{,}05$), значит, верна нулевая гипотеза и дисперсии двух групп не различаются. Поэтому для сравнения средних нужно применять двухвыборочный t-тест с одинаковыми дисперсиями.

Шаг 3. Формулировка гипотезы. Сравнение двух независимых групп. Тест Стьюдента. Нулевая гипотеза следующая: ни одна из диет не влияет на набор веса у поросят, т. е. групповые средние равны. В общем виде нулевую и альтернативную гипотезы запишем так:

- H_0: $\mu_1 = \mu_2$, $|t_{эмп}| < |t_{кр}|$, $P > 0{,}05$ – средние не отличаются;
- H_1: $\mu_1 \neq \mu_2$, $|t_{эмп}| \geq |t_{кр}|$, $P \leq 0{,}05$ – средние отличаются.

Выполним тест Стьюдента и узнаем, какая из гипотез верна. Для этого на вкладке *Данные* нужно нажать на пакет *Анализ данных* (см. рис. 22). В *Инструменты анализа* выбрать *Двухвыборочный t-тест с одинаковыми дисперсиями*, нажать *ОК*. Настроить диалоговое окно в соответствии с образцом (рис. 25), нажать *ОК*.

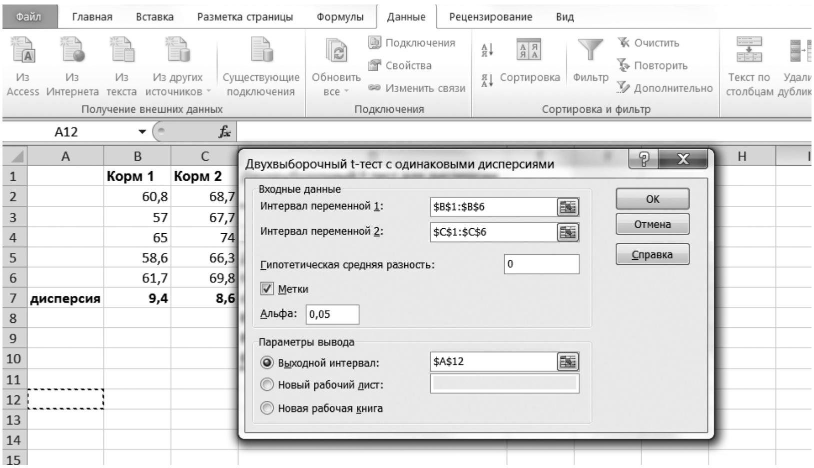
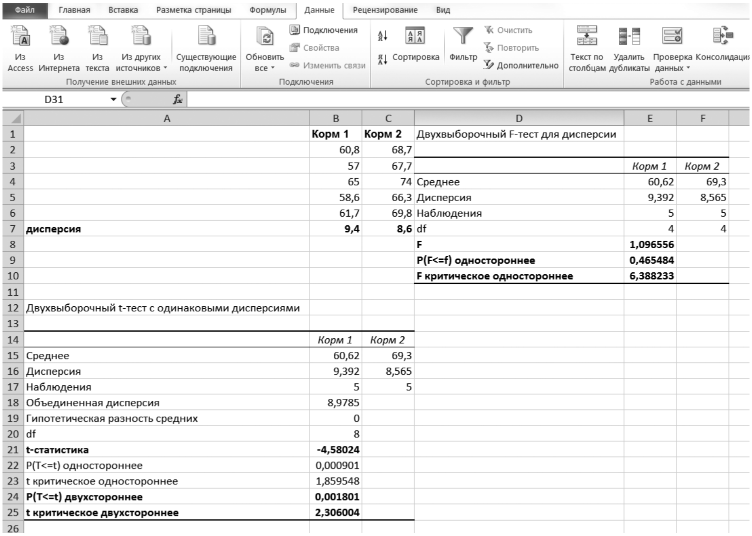


Рис. 25. Окончательный вид диалогового окна «Двухвыборочный *t*-тест с одинаковыми дисперсиями»

Рассмотрим таблицу с полученными результатами. В ней приведены средние для обеих выборок, дисперсии, количество наблюдений, число степеней свободы, вычисленное значение t ($t_{эмп}$), t-критическое ($t_{кр}$) для одно- и двустороннего сравнения и соответствующие вероятности справедливости гипотезы о равенстве средних (рис. 26).

Какую нужно использовать итоговую статистику: одно- или двустороннюю? При формулировке нулевой гипотезы о равенстве средних (наш вариант; гипотетическая средняя разность равна 0; H_0: $\mu_1 = \mu_2$) применяется двустороннее значение. Односторонние критерии используются при проверке, какая из средних арифметических больше или меньше. Вариант записи нулевой гипотезы в этом случае H_0: $\mu_1 > \mu_2$.



	A	B	C	D	E	F
1		**Корм 1**	**Корм 2**	Двухвыборочный F-тест для дисперсии		
2		60,8	68,7			
3		57	67,7		*Корм 1*	*Корм 2*
4		65	74	Среднее	60,62	69,3
5		58,6	66,3	Дисперсия	9,392	8,565
6		61,7	69,8	Наблюдения	5	5
7	**дисперсия**	**9,4**	**8,6**	df	4	4
8				**F**	**1,096556**	
9				**P(F<=f) одностороннее**	**0,465484**	
10				**F критическое одностороннее**	**6,388233**	
11						
12	Двухвыборочный t-тест с одинаковыми дисперсиями					
13						
14		*Корм 1*	*Корм 2*			
15	Среднее	60,62	69,3			
16	Дисперсия	9,392	8,565			
17	Наблюдения	5	5			
18	Объединенная дисперсия	8,9785				
19	Гипотетическая разность средних	0				
20	df	8				
21	**t-статистика**	**-4,58024**				
22	P(T<=t) одностороннее	0,000901				
23	t критическое одностороннее	1,859548				
24	**P(T<=t) двухстороннее**	**0,001801**				
25	**t критическое двухстороннее**	**2,306004**				
26						

Рис. 26. Результаты анализа

В рассматриваемом примере расчетная величина t по модулю (4,6) больше t-критического двустороннего (1,9), P равно 0,002 ($P < 0,05$), а значит, верна альтернативная гипотеза, нулевая же отклонена. Таким образом, рацион питания влияет на набор веса у поросят.

3.5.2. Сравнение двух зависимых групп тестом Стьюдента

Задача. Для выяснения эффективности нового удобрения его внесли в одинаковом количестве на девять одинаковых по площади участков, а в конце года измерили урожайность выращиваемой культуры. На следующий год по аналогичной схеме выполнили еще один эксперимент, однако со старым удобрением. Данные урожайности представлены в табл. 15.

Таблица 15

Информация об урожайности, ц

Тип удобрения	Участок								
	1	2	3	4	5	6	7	8	9
Новое	2250	2410	2260	2200	2360	2320	2240	2300	2090
Старое	1920	2020	2060	1960	1960	2140	1980	1940	1790

Различается ли средняя урожайность культуры в зависимости от используемого удобрения?

Шаг 1. Ввод данных. Введите данные по предложенному образцу (рис. 27).

	A	B	C
1	**Новое**	**Старое**	
2	2250	1920	
3	2410	2020	
4	2260	2060	
5	2200	1960	
6	2360	1960	
7	2320	2140	
8	2240	1980	
9	2300	1940	
10	2090	1790	
11			

Рис. 27. Ввод данных

Шаг 2. Формулировка гипотезы. Нулевая гипотеза следующая: тип удобрения не влияет на урожайность культуры, т. е. групповые средние равны. Общий вид нулевой и альтернативной гипотезы:

- H_0: $\mu_1 = \mu_2$, $|t_{\text{эмп}}| < |t_{\text{кр}}|$, $P > 0{,}05$ – средние не отличаются;
- H_1: $\mu_1 \neq \mu_2$, $|t_{\text{эмп}}| \geq |t_{\text{кр}}|$, $P \leq 0{,}05$ – средние отличаются.

Шаг 3. Сравнение двух зависимых групп. Тест Стьюдента. Выполним тест Стьюдента и узнаем, какая из гипотез верна. На вкладке *Данные* нажать на пакет *Анализ данных*. В *Инструменты анализа* выбрать *Парный двухвыборочный t-тест для средних*, нажать *ОК*. Настроить диалоговое окно в соответствии с образцом (рис. 28). Нажать *ОК*.

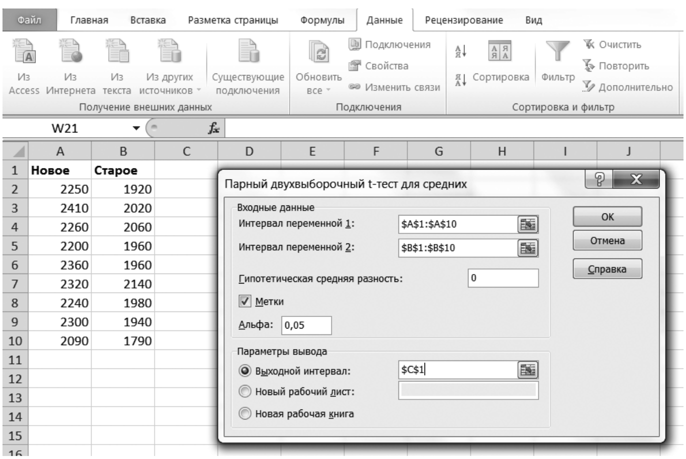
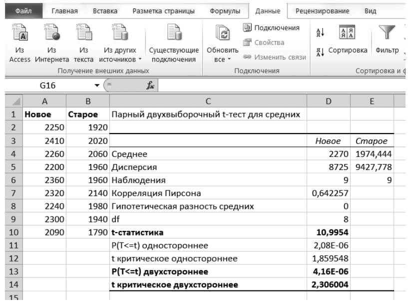


Рис. 28. Окончательный вид диалогового окна «Парный двухвыборочный *t*-тест для средних»



	A	B	C	D	E
1	**Новое**	**Старое**	Парный двухвыборочный t-тест для средних		
2	2250	1920			
3	2410	2020		*Новое*	*Старое*
4	2260	2060	Среднее	2270	1974,444
5	2200	1960	Дисперсия	8725	9427,778
6	2360	1960	Наблюдения	9	9
7	2320	2140	Корреляция Пирсона	0,642257	
8	2240	1980	Гипотетическая разность средних	0	
9	2300	1940	df	8	
10	2090	1790	**t-статистика**	**10,9954**	
11			P(T<=t) одностороннее	2,08E-06	
12			t критическое одностороннее	1,859548	
13			**P(T<=t) двухстороннее**	**4,16E-06**	
14			**t критическое двухстороннее**	**2,306004**	
15					

Рис. 29. Результат теста Стьюдента

В рассматриваемом примере расчетная величина t (10,99) больше t критического двухстороннего (2,3), P равно $4{,}16 \cdot 10^{-6}$ ($P \ll 0{,}001$), а значит, верна альтернативная гипотеза, нулевая отклонена (рис. 29). Средняя урожайность культуры в зависимости от типа удобрения различается.

3.5.3. Однофакторный дисперсионный анализ для независимых групп

При однофакторном дисперсионном анализе изучается влияние какого-либо одного фактора на изменчивость одного признака. Одновременно сравниваться могут две и более выборок.

Задача. Изучали влияние различных кормов на прирост 20 поросят. Данные о весе изученных животных каждой группы см. в табл. 16.

Таблица 16

Вес животных каждой группы, кг

Корм 1	60,8	57,0	65,0	58,6	61,7
Корм 2	68,7	67,7	74,0	66,3	69,8
Корм 3	102,6	102,1	100,2	96,5	96,5
Корм 4	87,9	84,2	83,1	85,7	90,3

Отражают ли полученные данные различия в рационах?

Шаг 1. Ввод данных. Введите данные по предложенному образцу (рис. 30). Данные сгруппированы по столбцам (один столбец – одна выборка).

	A	B	C	D	E
1	**Корм1**	**Корм2**	**Корм3**	**Корм4**	
2	60,8	68,7	102,6	87,9	
3	57	67,7	102,1	84,2	
4	65	74	100,2	83,1	
5	58,6	66,3	96,5	85,7	
6	61,7	69,8	96,5	90,3	

Рис. 30. Ввод данных

Шаг 2. Формулировка гипотезы. Нулевая гипотеза следующая: ни одна из диет не влияет на набор веса у поросят, т. е. все групповые средние равны. Общий вид нулевой и альтернативной гипотезы:

- H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$, $F_{эмп} < F_{кр}$, $P > 0{,}05$ – средние не отличаются;
- H_1: $\mu_i \neq \mu_j$, $F_{эмп} \geq F_{кр}$, $P \leq 0{,}05$ – средние отличаются.

Принимая нулевую гипотезу, мы утверждаем, что все наши выборки принадлежат одной и той же генеральной совокупности со средним μ. Принимая альтернативную гипотезу, нельзя сказать, между какими именно группами есть различия. Поэтому альтернативную гипотезу при дисперсионном анализе записывают: H_1: $\mu_i \neq \mu_j$.

Шаг 3. Сравнение четырех независимых групп. Однофакторный дисперсионный анализ. В задаче изучается воздействие одного фактора – рациона питания, поэтому будем выполнять однофакторный дисперсионный анализ. На вкладке *Данные* нажать на пакет *Анализ данных*. В *Инструменты анализа* выбрать *Однофакторный дисперсионный анализ*, нажать *ОК*. Настроить диалоговое окно в соответствии с образцом (рис. 31), нажать *ОК*.

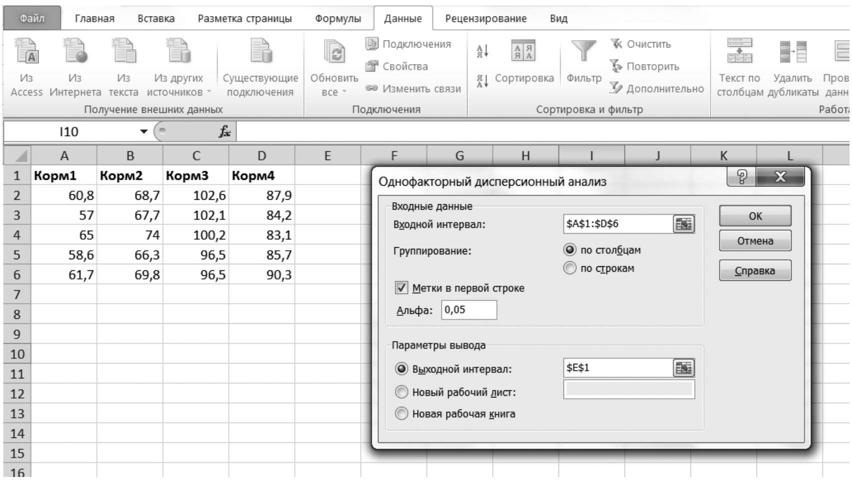
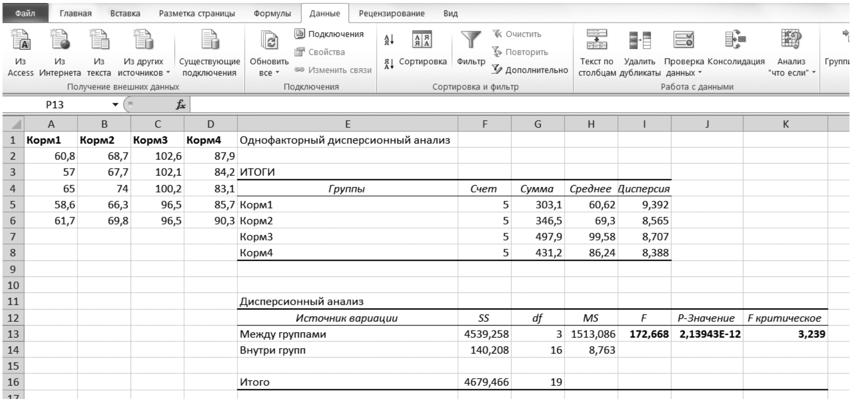


Рис. 31. Окончательный вид диалогового окна «Однофакторный дисперсионный анализ»

Получены следующие результаты, представленные на рис. 32. В первой таблице приведены некоторые показатели описательной статистики по каждой из исследуемых групп (объем выборки, сумма вариант, среднее и дисперсия), во второй – результаты вычислений внутри- и межгрупповых дисперсий (столбец *MS*), вычисленное значение F ($F_{эмп}$), вероятность справедливости нулевой гипотезы P и F-критическое ($F_{кр}$).



	A	B	C	D	E	F	G	H	I	J	K
1	**Корм1**	**Корм2**	**Корм3**	**Корм4**	Однофакторный дисперсионный анализ						
2	60,8	68,7	102,6	87,9							
3	57	67,7	102,1	84,2	ИТОГИ						
4	65	74	100,2	83,1	*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
5	58,6	66,3	96,5	85,7	Корм1	5	303,1	60,62	9,392		
6	61,7	69,8	96,5	90,3	Корм2	5	346,5	69,3	8,565		
7					Корм3	5	497,9	99,58	8,707		
8					Корм4	5	431,2	86,24	8,388		
9											
10											
11					Дисперсионный анализ						
12					*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
13					Между группами	4539,258	3	1513,086	**172,668**	**2,13943E-12**	**3,239**
14					Внутри групп	140,208	16	8,763			
15											
16					Итого	4679,466	19				

Рис. 32. Результат однофакторного дисперсионного анализа

Расчетная величина F (172,7) больше F-критического (3,2), P равно $2{,}14 \cdot 10^{-12}$ ($P \ll 0{,}001$), значит, верна альтернативная гипотеза, нулевая отклонена. Рацион питания влияет на набор веса у поросят.

3.5.4. Двухфакторный дисперсионный анализ без повторений

Программой *Excel* предусмотрено два инструмента для выполнения двухфакторного дисперсионного анализа:

• без повторений – для анализа данных без повторов комбинаций факторов;

• с повторениями – для анализа данных с повторами комбинаций факторов.

Задача. Изучали урожайность пшеницы четырех разных сортов, полученной при использовании пяти типов удобрений (табл. 17). Опыт проведен на 20 одинаковых участках.

Таблица 17

Урожайность пшеницы, ц/га

Тип удобрения	Сорт 1	Сорт 2	Сорт 3	Сорт 4
Удобрение *A*	19	25	17	21
Удобрение *B*	22	19	19	18
Удобрение *C*	26	23	22	25
Удобрение *D*	18	26	20	23
Удобрение *E*	21	22	21	24

Влияет ли сорт и тип удобрения на урожайность пшеницы?

В задаче изучают влияние на урожайность двух факторов: сорт и удобрение. Для каждого участка приведена одна величина урожайности для каждого сорта и типа удобрения, поэтому для анализа влияния факторов «сорт» и «удобрение» будем применять двухфакторный дисперсионный анализ без повторений. Если бы урожайность изучалась для трех участков каждого сорта и типа удобрения (с повторами комбинаций факторов), для анализа использовался бы двухфакторный дисперсионный анализ с повторениями.

Шаг 1. Ввод данных. Введите данные по предложенному образцу. (рис. 33).

	A	B	C	D	E	F
1		**Сорт 1**	**Сорт 2**	**Сорт 3**	**Сорт 4**	
2	**Удобрение A**	19	25	17	21	
3	**Удобрение B**	22	19	19	18	
4	**Удобрение C**	26	23	22	25	
5	**Удобрение D**	18	26	20	23	
6	**Удобрение E**	21	22	21	24	
7						

Рис. 33. Ввод данных

Шаг 2. Формулировка гипотез. Так как изучают влияние на урожайность двух факторов – сорт и удобрение, формулируем две нулевые гипотезы.

1. Нулевая гипотеза 1: тип удобрения не влияет на урожайность пшеницы. Общий вид нулевой и альтернативной гипотезы:

• H_0: $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu$, $F_{\text{эмп}} < F_{\text{кр}}$, $P > 0{,}05$ – урожайность одинаковая;

• H_1: $\mu_i \neq \mu_j$, $F_{\text{эмп}} \geq F_{\text{кр}}$, $P \leq 0{,}05$ – урожайность разная.

2. Нулевая гипотеза 2: сорт не влияет на урожайность пшеницы. Общий вид нулевой и альтернативной гипотезы:

• H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$, $F_{\text{эмп}} < F_{\text{кр}}$, $P > 0{,}05$ – урожайность одинаковая;

• H_1: $\mu_i \neq \mu_j$, $F_{\text{эмп}} \geq F_{\text{кр}}$, $P \leq 0{,}05$ – урожайность разная.

Шаг 3. Проверка гипотез. Двухфакторный дисперсионный анализ без повторений. На вкладке *Данные* нажать на пакет *Анализ данных*. В *Инструменты анализа* выбрать *Двухфакторный дисперсионный анализ без повторений*, нажать *ОК*. Настроить диалоговое окно в соответствии с образцом (рис. 34), нажать *ОК*.

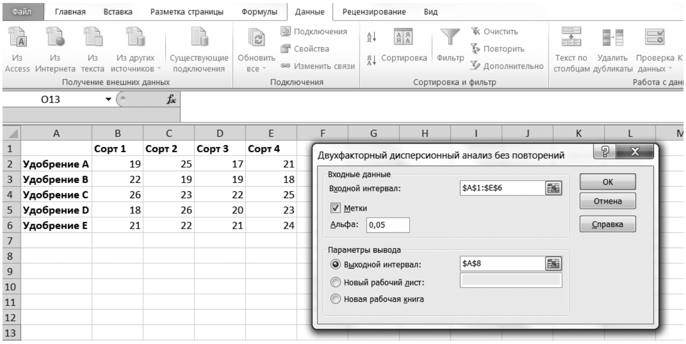
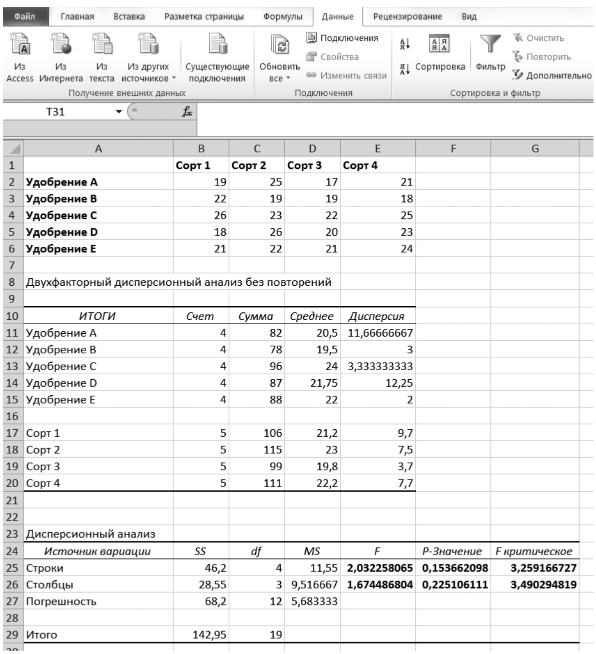


Рис. 34. Окончательный вид диалогового окна «Двухфакторный дисперсионный анализ повторения»

Получены результаты (рис. 35). В первой таблице приведены некоторые показатели описательной статистики по каждой из исследуемых групп (объем выборки, сумма вариант, среднее и дисперсия), во второй – результаты двух тестов Фишера.



	Сорт 1	Сорт 2	Сорт 3	Сорт 4
Удобрение A	19	25	17	21
Удобрение B	22	19	19	18
Удобрение C	26	23	22	25
Удобрение D	18	26	20	23
Удобрение E	21	22	21	24

Двухфакторный дисперсионный анализ без повторений

ИТОГИ	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
Удобрение A	4	82	20,5	11,66666667
Удобрение B	4	78	19,5	3
Удобрение C	4	96	24	3,333333333
Удобрение D	4	87	21,75	12,25
Удобрение E	4	88	22	2
Сорт 1	5	106	21,2	9,7
Сорт 2	5	115	23	7,5
Сорт 3	5	99	19,8	3,7
Сорт 4	5	111	22,2	7,7

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Строки	46,2	4	11,55	**2,032258065**	**0,153662098**	**3,259166727**
Столбцы	28,55	3	9,516667	**1,674486804**	**0,225106111**	**3,490294819**
Погрешность	68,2	12	5,683333			
Итого	142,95	19				

Рис. 35. Результаты анализа

Фактор «удобрение» сгруппирован в строках. Расчетная величина F (2,03) меньше F-критического (3,26), P равно 0,15 ($P > 0{,}05$), значит, верна нулевая гипотеза. Тип удобрения не влияет на урожайность пшеницы.

Фактор «сорт» сгруппирован в столбцах. Расчетная величина F (1,67) меньше F-критического (3,49), P равно 0,22 ($P > 0{,}05$), значит, верна нулевая гипотеза. Урожайность разных сортов пшеницы одинакова.

3.5.5. Двухфакторный дисперсионный анализ с повторениями

Двухфакторный дисперсионный анализ с повторениями позволяет проверить не только воздействие на изучаемый признак двух факторов, но и их возможное совместное влияние. В задаче в подразд. 3.5.4 предполагалось, что каждому сочетанию факторов соответствует одно наблюдение. Если вместо одного измерения провести несколько (два, три), то можно оценить возможный эффект их взаимодействия.

Задача. Изучали урожайность пшеницы четырех разных сортов, полученной при использовании пяти типов удобрений (табл. 18). Опыт проведен на 60 одинаковых участках.

Таблица 18

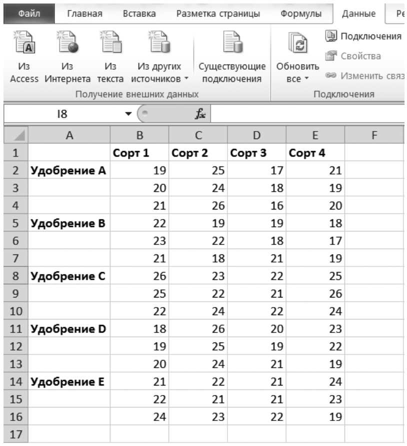
Урожай пшеницы разных сортов, ц/га

Тип удобрения	Сорт 1	Сорт 2	Сорт 3	Сорт 4
Удобрение *A*	19	25	17	21
	20	24	18	19
	21	26	16	20
Удобрение *B*	22	19	19	18
	23	22	18	17
	21	18	21	19
Удобрение C	26	23	22	25
	25	22	21	26
	22	24	22	24
Удобрение D	18	26	20	23
	19	25	19	22
	20	24	21	19
Удобрение *E*	21	22	21	24
	22	21	21	23
	24	23	22	19

Влияет ли сорт и тип удобрения на урожайность пшеницы?

Обратите внимание, что число наблюдений во всех ячейках должно быть одинаковым (в данном случае $n = 3$). Если количество повторов отличается для ячеек с разными комбинациями сорта и типа удобрения, то для таких данных инструмент двухфакторного дисперсионного анализа использовать нельзя.

Шаг 1. Ввод данных. Введите данные по предложенному образцу (рис. 36).



	A	B	C	D	E	F
1		Сорт 1	Сорт 2	Сорт 3	Сорт 4	
2	Удобрение A	19	25	17	21	
3		20	24	18	19	
4		21	26	16	20	
5	Удобрение B	22	19	19	18	
6		23	22	18	17	
7		21	18	21	19	
8	Удобрение C	26	23	22	25	
9		25	22	21	26	
10		22	24	22	24	
11	Удобрение D	18	26	20	23	
12		19	25	19	22	
13		20	24	21	19	
14	Удобрение E	21	22	21	24	
15		22	21	21	23	
16		24	23	22	19	
17						

Рис. 36. Ввод данных

Шаг 2. Формулировка гипотез. Так как изучают влияние на урожайность двух факторов – сорт и удобрение – и их возможного совместного взаимодействия, формулируем три нулевые гипотезы.

1. Нулевая гипотеза 1: тип удобрения не влияет на урожайность пшеницы. Общий вид нулевой и альтернативной гипотезы следующий:

• H_0: $\mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu$, $F_{\text{эмп}} < F_{\text{кр}}$, $P > 0{,}05$ – урожайность одинаковая;

• H_1: $\mu_i \neq \mu_j$, $F_{\text{эмп}} \geq F_{\text{кр}}$, $P \leq 0{,}05$ – урожайность разная.

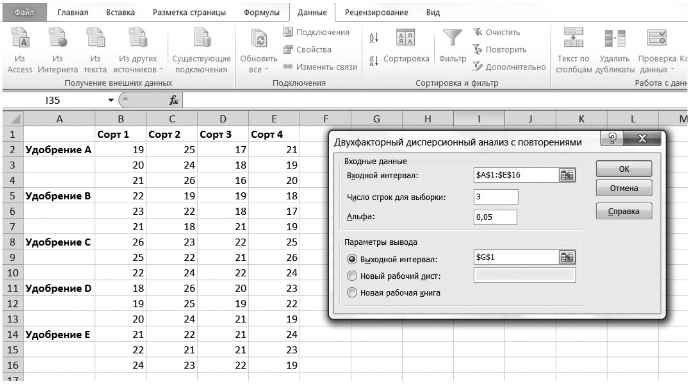
2. Нулевая гипотеза 2: сорт не влияет на урожайность пшеницы. Общий вид нулевой и альтернативной гипотезы:

• H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$, $F_{\text{эмп}} < F_{\text{кр}}$, $P > 0{,}05$ – урожайность одинаковая;

• H_1: $\mu_i \neq \mu_j$, $F_{\text{эмп}} \geq F_{\text{кр}}$, $P \leq 0{,}05$ – урожайность разная.

3. Нулевая гипотеза 3: взаимное воздействие типа удобрения и сорта не влияет на урожайность пшеницы.

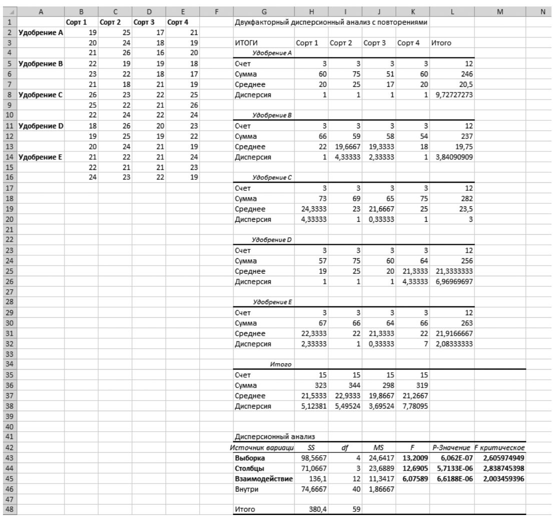
Шаг 3. Проверка гипотез. Двухфакторный дисперсионный анализ без повторений. На вкладке *Данные* нажать на пакет *Анализ данных*. В *Инструменты анализа* выбрать *Двухфакторный дисперсионный анализ с повторениями*, нажать *ОК*. Настроить диалоговое окно в соответствии с образцом (рис. 37). Нажать *ОК*.



	Сорт 1	Сорт 2	Сорт 3	Сорт 4
Удобрение A	19	25	17	21
	20	24	18	19
	21	26	16	20
Удобрение B	22	19	19	18
	23	22	18	17
	21	18	21	19
Удобрение C	26	23	22	25
	25	22	21	26
	22	24	22	24
Удобрение D	18	26	20	23
	19	25	19	22
	20	24	21	19
Удобрение E	21	22	21	24
	22	21	21	23
	24	23	22	19

Рис. 37. Окончательный вид диалогового окна «Двухфакторный дисперсионный анализ с повторениями»

Получены результаты (рис. 38). Посмотрите последнюю таблицу *Дисперсионный анализ*. В трех последних столбцах отображены результаты трех тестов Фишера, соответствующее значение P и F-критическое.



Двухфакторный дисперсионный анализ с повторениями

ИТОГИ	Сорт 1	Сорт 2	Сорт 3	Сорт 4	Итого
Удобрение A					
Счет	3	3	3	3	12
Сумма	60	75	51	60	246
Среднее	20	25	17	20	20,5
Дисперсия	1	1	1	1	9,72727273
Удобрение B					
Счет	3	3	3	3	12
Сумма	66	59	58	54	237
Среднее	22	19,6667	19,3333	18	19,75
Дисперсия	1	4,33333	2,33333	1	3,84090909
Удобрение C					
Счет	3	3	3	3	12
Сумма	73	69	65	75	282
Среднее	24,3333	23	21,6667	25	23,5
Дисперсия	4,33333	1	0,33333	1	3
Удобрение D					
Счет	3	3	3	3	12
Сумма	57	75	60	64	256
Среднее	19	25	20	21,3333	21,3333333
Дисперсия	1	1	1	4,33333	6,96969697
Удобрение E					
Счет	3	3	3	3	12
Сумма	67	66	64	66	263
Среднее	22,3333	22	21,3333	22	21,9166667
Дисперсия	2,33333	1	0,33333	7	2,08333333
Итого					
Счет	15	15	15	15	
Сумма	323	344	298	319	
Среднее	21,5333	22,9333	19,8667	21,2667	
Дисперсия	5,12381	5,49524	3,69524	7,78095	

Дисперсионный анализ

Источник вариаци	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Выборка	98,5667	4	24,6417	**13,2009**	**6,062E-07**	**2,605974949**
Столбцы	71,0667	3	23,6889	**12,6905**	**5,7133E-06**	**2,838745398**
Взаимодействие	136,1	12	11,3417	**6,07589**	**6,6188E-06**	**2,003459396**
Внутри	74,6667	40	1,86667			
Итого	380,4	59				

Рис. 38. Результаты анализа

Источник вариации *Выборка* отражает воздействие фактора «удобрение». Расчетная величина F (13,2) больше F-критического 2,6), P равно $6{,}062 \cdot 10^{-7}$ ($P << 0{,}001$), значит, верна альтернативная гипотеза, а нулевая отклонена. Тип удобрения влияет на урожайность пшеницы.

Источник вариации *Столбцы* есть воздействие фактора «сорт». Расчетная величина F (12,7) больше F-критического (2,8), P равно $5{,}7133 \cdot 10^{-6}$ ($P \ll 0{,}001$), значит, верна альтернативная гипотеза, а нулевая – отклонена. Сорт пшеницы влияет на урожайность.

Источник вариации *Взаимодействие* отражает взаимное воздействие типа удобрения и сорта. Расчетная величина F (6,1) больше F-критического (2,0), P равно $6{,}6188 \cdot 10^{-6}$ ($P \ll 0{,}001$), значит, верна альтернативная гипотеза, нулевая отклонена. Взаимное влияние удобрения и сорта на урожайность пшеницы есть.

Дискуссионный вопрос: теория малых выборок

Теория малых выборок появилась в начале XX в. и заполнила существенный пробел математической статистики. Все существовавшие до этого таблицы интегралов вероятности основывались на предположении, что число наблюдений, послуживших для определения эмпирических констант (средней арифметической, стандартного отклонения, средней ошибки и т. д.), очень велико. При справедливости такого предположения (в выводе число наблюдений принималось даже бесконечно большим) справедливы и соотношения между величиной и своей средней ошибкой. Так, например, согласно хорошо известному правилу трех сигм, вероятность того, что случайно взятый элемент совокупности будет отставать от арифметического среднего (в ту или другую сторону) не менее чем на три средних квадратичных (стандартных) отклонения, равна приблизительно 0,003.

Всякая математическая формула справедлива только при соблюдении условий, которые были положены в ее выводе. Тем не менее, так как они строго никогда не соблюдаются, формулы применяются и при отклонении от принятых условий, но только если оно не будет очень велико. В практической работе исследуемая выборка из общей совокупности никогда не бывает бесконечной, а иногда состоит всего из нескольких экземпляров. Естественно, применимость формул классической математической статистики к подобным случаям вызывает большие сомнения, и, например, уже в руководстве Е. Е. Слуцкого (1912) по теории корреляции не рекомендовалось вычислять коэффициенты корреляции менее чем для 30 наблюдений, причем и это число устанавливалось сугубо эмпирическим путем. Есть и работы, прямо доказывавшие, что применимость критериев ошибки, выработанных для выборок большого объема, к малым выборкам приводит к грубо ошибочным суждениям.

Начало статистике малых выборок было положено в первом десятилетии XX в. публикацией работы У. Госсета, где он под псевдонимом Стьюдент постулировал t-распределение. В отличие от теории нормального распределения, теория t-распределения для малых выборок не требует априорного знания или точных оценок математического ожидания (среднего арифметического) и дисперсии генеральной совокупности, а также допущений относительно параметров. В t-распределении одно из отклонений от выборочного среднего всегда фиксировано, так как сумма всех отклонений должна равняться нулю. Это сказывается на сумме квадратов при вычислении выборочной дисперсии как несмещенной оценки дисперсии генеральной совокупности и ведет к тому, что число степеней свободы df получается равным числу измерений минус единица для каждой выборки. Известны также классические работы английского статистика Р. Фишера (в честь которого получило свое название F-распределение) по дисперсионному анализу – статистическому методу, ориентированному на анализ малых выборок.

Работы Стьюдента и Фишера открыли новую эпоху в развитии математической статистики, так как при соблюдении определенных ограничений (наличие закона нормального распределения в генеральной совокупности) они дали возможность определить вероятность отклонения определенного размера для выборок любого объема. На основе этой теории были составлены таблицы, приводимые во всех руководствах по вариационной статистике. Из них ясно, что для получения минимальной значимости (доверительной вероятности, равной 0,05) отношение величины к своей средней ошибке должно равняться 1,96 при бесконечном числе испытаний, 2,00 – при 60 степенях свободы, 2,042 – при 30 степенях свободы (разница, как видим, очень небольшая, позволяющая даже число 30 признавать достаточно большим). А вот для пяти степеней свободы мы имеем t, равное уже 2,57, для трех – 3,18, для двух (при наличии только трех исследованных индивидов совокупности) t возрастает до 4,3.

Из многочисленных статистик, которые можно обоснованно применять к малым выборкам, кроме t-теста и дисперсионного анализа, можно упомянуть точный критерий Фишера, непараметрический дисперсионный анализ Фридмана, H-критерий Краскела – Уоллеса, U-критерий Манна – Уитни, медианный критерий, критерий знаков, T-критерий Вилкоксона. При этом для каждой исходной статистики обычно вычисляется нормальная аппроксимация (z-статистика) и уровень значимости P нулевой гипотезы об отсутствии различий в разбросе значений двух выборок. Если $P > 0,05$, нулевая гипотеза может быть принята.

Определенного ответа на вопрос «Какой объем должна иметь выборка, чтобы ее можно было считать малой?» не существует. Однако условной границей между малой и большой выборкой принято считать $df = 30$. Основанием для этого в какой-то мере произвольного решения служит результат сравнения t-распределения (для малых выборок) с нормальным распределением (z). Расхождение значений t и z имеет тенденцию возрастать с уменьшением и снижаться с увеличением *df*. Фактически, *t* начинает тесно приближаться к *z* задолго до предельного случая, когда $t = z$. Простое визуальное изучение табличных значений *t* позволяет увидеть, что это приближение становиться довольно быстрым, начиная с $df = 30$ и выше. Сравнительные величины *t* (при $df = 30$) и *z* равны соответственно: 2,04 и 1,96 для $P = 0,05$; 2,75 и 2,58 для $P = 0,01$; 3,65 и 3,29 для $P = 0,001$.

Таким образом, уже на этапе планирования исследования важно учитывать мощность применяемых статистических критериев, которые определяются вариабельностью выборки и заданным уровнем значимости. Для очень малых выборок совершенно неприложимо правило, используемое для больших выборок, и точность вывода возрастает пропорционально корню квадратному числа исследованных: выигрыш точности при переходе от одной степени свободы к двум оказывается неизмеримо большим. Именно поэтому при простых сопоставлениях работать с одной повторностью категорически не рекомендуется.

Бывает и так, что внешне очевидные результаты не имеют под собой настоящих оснований. Как, например, одно из исследований насекомых-вредителей. Агрономы определяли, насколько сильно вредят кукурузе гусеницы кукурузного мотылька. Получились вполне приемлемые результаты: разница в урожае между пораженными и непораженными растениями – почти вдвое. Казалось, что и обрабатывать статистически ничего не нужно ведь и так все ясно. Однако нашелся вдумчивый исследователь, который заметил, что пораженные растения, различающиеся по степени поражения, не различаются по урожайности. Другими словами, здесь очевидно что-то не так: если гусеницы вредят растению, то чем сильнее они вредят, тем меньше должен быть урожай. Стало быть, на какой-то стадии исследования произошла ошибка. И, скорее всего, дело было так: для того, чтобы измерять урожайность, среди здоровых растений отбирали самые здоровые (во всех смыслах этого слова), а среди больных старались подобрать наиболее хилые. Вот эта ошибка репрезентативности и привела к тому, что возникли такие результаты. Обратите внимание, что только вдумчивый анализ взаимосвязи «поражение – урожай» привел к выяснению истинной причины. А кукурузный мотылек, оказывается, почти и не вредит кукурузе...

А какие еще ошибки планирования этого исследования можете предположить вы?

4. ЧИСЛОВЫЕ ДАННЫЕ. НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

Рассматриваемые в предыдущем разделе параметрические тесты Стьюдента и дисперсионный анализ применимы, когда наблюдаемый признак подчиняется нормальному распределению, дисперсии сравниваемых групп статистически равны и количества наблюдений достаточно для проверки данных предположений. Однако некоторые исследователи не рекомендуют применять параметрические методы, если объем каждой из групп составляет менее 30 наблюдений, даже если выборочные данные имеют нормальное распределение. Применяя тот или иной метод сравнения групп без проверки верности допущений, есть риск того, что, выполнив правильную последовательность действий, ученый придет к ошибочным выводам.

В этом разделе будут рассмотрены непараметрические методы, которые используют при невыполнении хотя бы одного допущения. При применении непараметрических методов исходные данные заменяют рангами. При этом сохраняется большая часть информации о распределении исходных данных, но нет необходимости знать, что это за распределение. Соответственно, отпадает и нужда в равенстве дисперсий.

Рассмотрение принципа применения каждого критерия будет состоять из следующих шагов:

1) формулировка нулевой гипотезы;
2) замена исходных данных рангами;
3) вычисление эмпирического значения критерия;
4) принятие нулевой либо альтернативной гипотезы.

4.1. СРАВНЕНИЕ ДВУХ ГРУПП

4.1.1. Две независимые группы

Для сравнения двух независимых (несвязанных) выборок из непараметрических методов наиболее часто используется *U-критерий Манна – Уитни* – непараметрическая альтернатива *t*-критерия для независимых выборок. Этот статистический метод был предложен Ф. Вилкоксоном в 1945 г., однако в 1947 г. метод был улучшен и расширен Х. Б. Манном и Д. Р. Уитни, поэтому *U*-критерий чаще называют их именами.

В пособии отражены три варианта упрощенных вычислений критерия Манна – Уитни, предложенных С. Гланцом (1999). Ознакомиться с более детальными расчетами можно в книге «*Biostatistical analysis*» *J. H. Zar* (2010).

Вариант 1. Вычисление *T*-критерия Манна – Уитни

Задача. Двум группам крыс (самцам и самкам) были сделаны инъекции экспериментального лекарства. Время реакции на него (в секундах) преведено ниже.

Самки:	9,4	7,4	9,6			
Самцы:	8,6	10,5	11,4	9,4	7,9	9,0

Зависит ли значение времени реакции на лекарство от пола крыс?

Шаг 1. Формулировка нулевой гипотезы. Время реакции на лекарство одинаково у самок и самцов.

Шаг 2. Замена исходных данных рангами. Против каждого значения времени реакции указан его ранг. Ранг 1 присвоен наименьшей варианте – 7,4. Заменяя исходные значения их рангами, работают с объединенными данными обеих выборок, поэтому ранг 2 присвоен значению 7,9 из выборки самцов. Варианта 9,4 регистрировалась дважды, поэтому ранги 5 и 6 суммируют и делят на два – получен ранг 5,5, применяемый для обоих значений одинаковой варианты. Ранг 9 присвоен максимальному значению 11,4 (рис. 39).

Самки		*Самцы*	
x_i	Ранг	x_i	Ранг
7,4	1	7,9	2
9,4	5,5	8,6	3
9,6	7	9	4
		9,4	5,5
		10,5	8
		11,4	9
$n_м = 3$	$T = 13,5$	$n_б = 6$	

Рис. 39. Значения и их ранги

После ранжирования общее количество полученных рангов должно совпадать с количеством наблюдений в объединенной выборке.

Шаг 3. Вычисление эмпирического значения критерия. Сумма рангов членов меньшей группы дает величину T – эмпирическое значение рассчитываемого критерия (см. рис. 39). При одинаковых объемах групп значение T можно вычислить для любой из них.

Шаг 4. Принятие гипотезы. Критические значения критерия Манна – Уитни приведены в табл. 19. Столбец критических значений содержит пары чисел. Если T эмпирическое не больше первого из них или не меньше второго, то различия статистически значимы и нулевая гипотеза отвергается. В рассматриваемой задаче объем группы меньшей ($n_{м}$) – 3, большей ($n_{б}$) – 6. Различия статистически значимы, если $T \leq 7$ или $T \geq 23$. Расчетное значение T – 13,5. Нулевая гипотеза верна.

Таблица 19

Критические значения критерия (двусторонний вариант) Манна – Уитни (фрагмент)

Численность группы		Уровень значимости α = 0,05	
маленькой	большой	Критические значения	
3	4	6	18
	5	6	21
	5	7	20
	6	7	23
	7	7	26
	7	8	25
	8	8	28
4	4	11	25
	5	11	29
	5	12	28
	6	12	32
	7	13	35
	8	14	38
	8	12	40

Вариант 2. Вычисление z_T-статистики (критерия Вилкоксона для независимых групп)

Задача. Двум группам крыс (самцам и самкам) были сделаны инъекции экспериментального лекарства. Время реакции на него (в секундах) преведено ниже.

Самки:	9,4	7,4	9,6	8,4	10,1	10,2	7,6	11,6	8,5	
Самцы:	8,6	10,5	11,4	9,4	7,9	9,0	10,8	10,4	8,3	11,7

Зависит ли значение времени реакции на лекарство от пола крыс?

Шаг 1. Формулировка нулевой гипотезы. Время реакции на лекарство одинаково у самок и самцов.

Шаг 2. Замена исходных данных рангами. Проводится как и в варианте 1 (с. 74). Величина T – сумма рангов каждой выборки (рис. 40).

Самки		*Самцы*	
x_i	Ранг	x_i	Ранг
7,4	1	7,9	3
7,6	2	8,3	4
8,4	5	8,6	7
8,5	6	9	8
9,4	9,5	9,4	9,5
9,6	11	10,4	14
10,1	12	10,5	15
10,2	13	10,8	16
11,6	18	11,4	17
		11,7	19
$n_м = 9$	$T = 77,5$	$n_б = 10$	$T = 112,5$

Рис. 40. Значения и их ранги

Шаг 3. Вычисление эмпирического значения z_T. При численности групп больше восьми распределение T приближается к нормальному со средним μ_T и стандартным отклонением σ_T

$$\mu_T = \frac{n_м\left(n_м + n_б + 1\right)}{2}, \qquad \sigma_T = \sqrt{\frac{n_м n_б\left(n_м + n_б + 1\right)}{12}}.$$

В таком случае величина z_T имеет стандартное нормальное распределение. Это позволяет сравнить z_T с критическими значениями нормального распределения (табл. 20, последняя строка).

Переход от T к z_T осуществляется по формуле

$$z_T = \frac{T - \mu_T}{\sigma_T}.$$

Вычисляем среднее μ_T всех возможных значений T и стандартное отклонение σ_T, затем z_T (используем сумму рангов наименьшей группы)

$$\mu_T = 9\left(9 + 10 + 1\right)/2 = 90, \qquad \sigma_T = \sqrt{9 \cdot 10\left(9 + 10 + 1\right)/12} = 12,25,$$

$$z_T = \left(77,5 - 90\right)/12,25 = -1,021.$$

Шаг 4. Принятие гипотезы. Распределение t с $\nu = \infty$ идентично нормальному распределению, критическое значение z_α равно критическому значению $t_{\alpha,\infty}$. Модуль расчетной величина z_T меньше критического значения z (см. табл. 20, последняя строка). Нулевая гипотеза верна.

Таблица 20

Критические значения *t* (двусторонний вариант) (фрагмент)

Уровень значимости α			
v	0,05	0,01	0,001
1	12,706	63,656	636,578
2	4,303	9,925	31,600
3	3,182	5,841	12,924
4	2,776	4,604	8,610
5	2,571	4,032	6,869
...	...	...	...
90	1,987	2,632	3,402
100	1,984	2,626	3,390
120	1,980	2,617	3,373
140	1,977	2,611	3,361
160	1,975	2,607	3,352
...	...	...	...
180	1,973	2,603	3,345
200	1,972	2,601	3,340
...	...	...	...
∞	1,960	2,576	3,291

Вариант 3. Вычисление U-критерия Манна – Уитни

В программе *Statistica* приведен итог расчета двух величин – z_T и U (прил. 7). Один из вариантов расчета величины U-критерия Манна – Уитни следующий:

$$U = T - n_{\text{м}}\left(n_{\text{м}} + 1\right)/2,$$

где T – сумма рангов меньшей группы; $n_{\text{м}}$ – объем меньшей группы.

Для ее расчета воспользуемся данными задачи, представленными в варианте 2 (с. 75):

$$U = 77{,}5 - 9(9 + 1) = 32{,}5.$$

* * *

Критерий суммы рангов Манна – Уитни применяется для проверки однородности двух независимых совокупностей одинаковой или разной численности. Выборки могут принадлежать порядковой или количественной шкале. Проверка на нормальность распределения не требуется. Используют как непараметрический аналог критерия Стьюдента.

4.1.2. Две зависимые группы

В разд. 3 было описано использование парного критерия Стьюдента для сравнения двух зависимых (связанных) групп. Парный критерий Вилкоксона – непараметрический аналог критерия Стьюдента для сравнения связанных групп. Как и критерий Манна – Уитни, он работает не с исходными данными, а с их рангами. По сравнению с парным критерием Стьюдента, является менее чувствительным, но позволяет проводить сравнение зависимых групп при несоблюдении какого-либо допущения.

Парный критерий Вилкоксона основан на ранжировании абсолютных величин разности исходных данных без учета знака. Рассмотрим принцип работы этого критерия на примере.

Задача. Проведено изучение экспериментального лекарства на ответную реакцию на раздражитель у девяти самок крыс. Время реакции на раздражитель до и после инъекции препарата приведен в табл. 21.

Таблица 21

Время реакции на раздражитель, с

Время	№ самки								
	1	2	3	4	5	6	7	8	9
До	10,5	11,4	9,4	9,0	10,8	11,2	8,4	10,3	9,9
После	10,8	9,9	9,8	10,8	10,8	8,9	12,0	11,2	11,4

Повлиял ли препарат на время реакции крыс?

Шаг 1. Формулировка нулевой гипотезы. Время реакции у самок крыс на раздражитель одинаково до и после применения препарата.

Шаг 2. Замена исходных данных рангами.

Шаг 3. Вычисление эмпирического значения парного критерия Вилкоксона T. Поэтапный образец расчета парного критерия Вилкоксона отражен в табл. 22 и состоит из следующих действий:

- определение разности между временем реакции после применения препарата и до его введения. Полученные значения вносим в столбец «Разность»;
- поиск модуля полученной разности (столбец «Модуль разности»);
- присваивание модулю разности ранг. Переменная с нулевыми изменениями (самка под № 5) в анализе не участвует, поэтому будет восемь рангов, а не девять. Если есть одинаковые значения модулей разности, им присваивается средний ранг. Так, дважды встречался модуль разности 1,5. Его ранг 4,5 получен усреднением рангов 4 и 5;
- в столбец $T+$ внесение рангов для положительных разностей, для отрицательных – в столбец $T-$;
- подсчет суммы рангов для положительных и отрицательных разностей;
- меньшая из двух сумм рангов – эмпирическое значение парного критерия Вилкоксона T. В рассматриваемом примере $T = 11,5$.

Таблица 22

Пример расчета парного критерия Вилкоксона

№ самки	До (x_i)	После (x_i')	Разность ($x_i' - x_i$)	Модуль разности $\lvert x_i' - x_i\rvert$	Ранг	T+	T–
1	10,5	10,8	0,3	0,3	1	1	
2	11,4	9,9	−1,5	1,5	**4,5**		4,5
3	9,4	9,8	0,4	0,4	2	2	
4	9	10,8	1,8	1,8	6	6	
5	10,8	10,8	0	0	0	0	
6	11,2	8,9	−2,3	2,3	**7**		**7**
7	8,4	12	3,6	3,6	8	8	
8	10,3	11,2	0,9	0,9	3	3	
9	9,9	11,4	1,5	1,5	4,5	4,5	
Сумма рангов						24,5	11,5

Шаг 4. Принятие гипотезы. При сравнении двух связанных групп вручную, есть два варианта анализа в зависимости от числа сравниваемых пар этих групп – меньше или больше 100.

Для проверки выдвинутой нулевой гипотезы при числе пар меньше 100, полученную величину парного критерия Вилкоксона $T_{\text{эмп}}$ сравнивают с табличным критическим значением распределения Вилкоксона $T_{\text{кр}}$ (таблицы есть в книге «*Biostatistical analysis*» *J. H. Zar* (2010)). При числе пар большем 100, рассчитывают величину z.

В программе *Statistica* (прил. 8) вне зависимости от числа сравниваемых пар, приведен итог расчета двух величин – T и Z (z). Формулы, связывающие z-статистику с парным критерием Вилкоксона, следующие:

$$z = \frac{T - \overline{T}}{SE_T}, \qquad \overline{T} = \frac{n(n+1)}{4}, \qquad SE_T = \sqrt{\frac{n(n+1)(2n+1)}{24}},$$

где T – значение парного критерия Вилкоксона; $\overline{T}$ – среднее значение парного критерия Вилкоксона; SE_T – стандартная ошибка парного критерия Вилкоксона; n – количество наблюдений, не имеющих нулевых изменений.

Определим величину z-статистики:

$$\overline{T} = 8(8+1)/4 = 18,$$

$$SE_T = \sqrt{\frac{8(8+1)(2\cdot 8+1)}{24}} = 7{,}14,$$

$$z = (11{,}5 - 18)/7{,}14 = -0{,}91.$$

В случае двустороннего распределения критическое значение z эквивалентно $t_{\alpha} = 0{,}05, \infty$ и равно 1,96. Если величина модуля z-статистики меньше, чем 1,96, принимаем нулевую гипотезу, если больше – альтернативную. В рассматриваемом примере абсолютное значение расчетной величины z-статистики меньше критического значения. Оставляем нулевую гипотезу – время реакции у самок крыс на раздражитель одинаково до и после использования препарата.

* * *

Парный критерий Вилкоксона применяется для сравнения парных выборок. Изучаемый признак порядковый – количественный или качественный. Проверка на нормальность распределения количественного признака не требуется. Используют как непараметрический аналог критерия Стьюдента.

4.2. СРАВНЕНИЕ БОЛЕЕ ДВУХ ГРУПП

4.2.1. Более двух независимых групп

Критерий Краскела – Уоллиса, который иногда называют непараметрическим дисперсионным анализом, позволяет провести сравнение двух и более групп. Этот критерий работает не с исходными значениями, а с их рангами. Так же как и дисперсионный анализ, критерий Краскела – Уоллиса поможет выяснить, имеются ли различия между группами, однако не сможет показать, между какими именно.

Задача. Изучали влияние разных типов удобрения на урожайность озимой ржи. Исследования проведены на 18 одинаковых по площади участках. В конце года измерили урожайность выращиваемой культуры. Результаты испытаний приведены в табл. 23.

Таблица 23

Влияние удобрений на урожайность озимой ржи

Вид удобрения	Урожайность, ц/га						
Удобрение 1	8	8,4	9	8,6	10,1	–	–
Удобрение 2	8,3	10,2	9	10	10	9,3	–
Удобрение 3	8,8	10	9,7	12	9,2	9,1	8,2

Повлиял ли тип удобрения на урожайность озимой ржи?

Шаг 1. Формулировка нулевой гипотезы. Тип удобрения не влияет на урожайность озимой ржи.

Шаг 2. Замена исходных данных рангами. В табл. 24 напротив каждого значения урожайности указан его ранг. При замене исходных величин рангами работают с упорядоченными по возрастанию объединенными данными трех выборок. Каждому значению в этом упорядоченном по возрастанию ряду присваивается свой ранг, а по сути – номер его места в ранжированном ряду значений признака. Так, ранг 1 характеризует наименьшую варианту 8, ранг 18 – максимальное значение 12. Варианта 10 регистрировалась трижды, поэтому ранги 13, 14 и 15 суммируют и делят на три – получен ранг 14, применяемый для трех исходных значений (см. табл. 24).

Таблица 24

Ранжирование значений урожайности и сумма рангов

Удобрение 1 ($n = 5$)	Ранг	Удобрение 2 ($n = 6$)	Ранг	Удобрение 3 ($n = 7$)	Ранг
8	1	8,3	3	8,8	6
8,4	4	10,2	17	10	14
9	7,5	9	7,5	9,7	12
8,6	5	10	14	12	18
10,1	16	10	14	9,2	10
		9,3	11	9,1	9
				8,2	2
Сумма рангов (R_i)	33,5		66,5		71

Общее количество полученных рангов должно быть равно числу наблюдений в объединенной выборке ($N = 18$). Суммировав значения рангов отдельно для каждого столбца, получаем величину суммы рангов (R_i) каждой группы.

Шаг 3. Вычисление эмпирического значения критерия. Для определения величины критерия воспользуемся формулой

$$H = \frac{12}{N(N+1)} \cdot \sum \frac{R_i^2}{n_i} - 3\left(N+1\right),$$

где R_i – сумма рангов для каждой группы; n_{i} – количество наблюдений в каждой группе; N – общее количество наблюдений в объединенной выборке.

$$H_{\text{эмп}} = \frac{12}{18\left(18+1\right)} \cdot \left(\frac{33{,}5^2}{5} + \frac{66{,}5^2}{6} + \frac{71^2}{7} \right) - 3\left(18+1\right) = 2{,}01.$$

Шаг 4. Принятие гипотезы. Для подтверждения или опровержения нулевой гипотезы необходимо найти критическое значение $H_{\text{кр}}$.

Если число сравниваемых групп три, при этом количество наблюдений в каждой группе не менее пяти (для четырех групп – общее число наблюдений не менее 10), то полученное значение критерия Краскела – Уоллиса $H_{эмп}$ сравнивают с критическим значением хи-квадрат Пирсона χ^2 (табл. 25). Это возможно, так как распределение H близко распределению χ^2 с числом степеней свободы $df = k - 1$, где k – число групп. Если расчетное значение $H_{эмп}$ равно или превышает критическое значение χ^2, то H_0 отвергается.

При числе наблюдений в группах менее пяти в качестве критического значения используют готовые табличные значения распределения Краскела – Уоллиса.

Таблица 25

Критические значения χ^2 (фрагмент)

	Уровень значимости							
df	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
1	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828
2	1,386	2,773	4,605	5,991	7,378	9,210	10,597	13,816
3	2,366	4,108	6,251	7,815	9,348	11,345	12,838	16,266
4	3,357	5,385	7,779	9,488	11,143	13,277	14,860	18,467
5	4,351	6,626	9,236	11,070	12,833	15,086	16,750	20,515

Из таблицы критических значений критерия χ^2 для числа степеней свободы $df = k - 1 = 3 - 1 = 2$ и уровня значимости 0,05 критическое значение χ^2 составляет 5,99 (см. табл. 25). Так как расчетное значение меньше критического ($H_{эмп} = 2,01$), то нулевая гипотеза не отклоняется – урожайность озимой ржи в экспериментальных группах одинакова.

В прил. 9 приведена последовательность решения этой задачи с помощью программы *Statistica*.

* * *

Критерий Краскела – Уоллиса является непараметрическим аналогом однофакторного дисперсионного анализа и применяется для сравнения двух и более независимых групп одинакового или разного объема, в том числе для групп с малым числом наблюдений. Данные, полученные для каждой из групп, могут не подчиняться нормальному распределению и иметь разные дисперсии.

4.2.2. Более двух зависимых групп

Критерий Фридмана применяют для сравнения двух и более связанных (зависимых) групп. Как и критерии, рассмотренные в предыдущих подразделах данного раздела, его расчет основан на ранжированных значениях изучаемого признака. Рассмотрим, как проводится ранжирование исходных данных и вычисление величины критерия Фридмана на примере задачи.

Задача. Преподаватель решил узнать, различалась или нет продолжительность подготовки его студентов к семинару в пятницу. Результаты представлены в табл. 26.

Таблица 26

Продолжительность подготовки студентов к семинару, мин

Студент	Понедельник	Вторник	Среда	Четверг
Олег	9	4	10	11
Семен	4	2	4	8
Юрий	8	7	10	15
Петр	4	9	7	10
Иван	7	8	10	10
Игнат	9	12	14	17

Обнаружил ли преподаватель разницу во времени подготовки к семинару?

Шаг 1. Формулировка нулевой гипотезы. Время подготовки к семинару одинаково на протяжении всех дней.

Шаг 2. Замена исходных данных рангами. Осуществляется путем упорядочивания значений каждого объекта наблюдения (студента). В результате получаем число рядов, равное числу студентов, участвующих в опросе. Обратите внимание, что ранее мы упорядочивали группы, сейчас – наблюдения каждого отдельно взятого объекта. В случае, если значения совпадают (например, у Семена (см. табл. 26) время подготовки одинаково в понедельник и среду), совпадающим значениям присваивается средний ранг: $(2 + 3)/2 = 2{,}5$. Полученные значения рангов суммируют отдельно для каждого столбца, получая величину суммы рангов ($R_м$) каждой группы (табл. 27).

Таблица 27

Ранжирование значений времени подготовки и сумма рангов

Студент	Понедельник	Вторник	Среда	Четверг
Олег	2	1	3	4
Семен	2,5	1	2,5	4
Юрий	2	1	3	4
Петр	1	3	2	4
Иван	1	2	3,5	3,5
Игнат	1	2	3	4
Сумма рангов $R_м$	9,5	10	17	23,5

Шаг 3. Вычисление эмпирического значения критерия. Для определения величины критерия Фридмана воспользуемся формулой

$$\chi_r^2 = \frac{12}{nk(k+1)} \sum \left(R_м - \frac{n(k+1)}{2} \right)^2,$$

где $R_м$ – суммы рангов для сравниваемых групп; k – количество сравниваемых групп; n – количество наблюдений в группе.

$$\chi_r^2 = 12 \Big/ \left(6 \cdot 4(4+1)\right) \left((9{,}5-15)^2 + (10-15)^2 + (17-15)^2 + (23{,}5-15)^2 \right) = 13.$$

Для упрощения процедуры вычислений отдельно рассчитали отношение

$$n(k + 1) / 2 = 15.$$

Шаг 4. Принятие нулевой/альтернативной гипотезы. Для подтверждения или опровержения нулевой гипотезы необходимо найти критическое значение $\chi^2_{кр}$.

При большой численности группы величина критерия Фридмана приблизительно соответствует распределению χ^2 с числом степеней свободы $df = k - 1$ (табл. 28), но при $k = 3$ и $n \leq 9$ и при $k = 4$ и $n \leq 4$ это приближение оказывается слишком грубым. В таком случае следует воспользоваться точными значениями $\chi^2_{кр}$.

Таблица 28

Фрагмент таблицы критических значений χ^2 (по Гланц, 1999)

Уровень значимости								
df	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
1	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828
2	1,386	2,773	4,605	5,991	7,378	9,210	10,597	13,816
3	2,366	4,108	6,251	7,815	9,348	11,345	12,838	16,266
4	3,357	5,385	7,779	9,488	11,143	13,277	14,860	18,467
5	4,351	6,626	9,236	11,070	12,833	15,086	16,750	20,515

Из таблицы критических значений критерия χ^2 для числа степеней свободы $df = k - 1 = 4 - 1 = 3$ и уровня значимости 0,05 критическое значение χ^2 составляет 7,82 (см. табл. 28). Так как расчетное значение больше критического ($\chi_r^2 = 13$), то нулевая гипотеза отклоняется – время подготовки к семинару на протяжении всех дней разное.

В прил. 10 приведена последовательность решения этой задачи с помощью программы *Statistica*.

* * *

Критерий Фридмана является непараметрическим аналогом дисперсионного анализа повторных измерений двух и более групп одинакового объема. Использование этого критерия не требует нормального распределения и основано на ранжировании значений изучаемого признака.

5. КАЧЕСТВЕННАЯ ИЗМЕНЧИВОСТЬ. АНАЛИЗ ЧАСТОТ

Статистические процедуры, с которыми мы познакомились в предыдущих разделах, предназначены для анализа *количественных признаков*. Однако очень многие признаки невозможно измерить числом, например, можно быть мужчиной либо женщиной, мертвым либо живым, врачом, юристом, рабочим и т. д. Тогда мы имеем дело с *качественными признаками*, которые не связаны между собой никакими арифметическими соотношениями, упорядочить их также нельзя. Единственный способ описания качественных признаков состоит в том, чтобы подсчитать число объектов, имеющих одно и то же значение. Кроме того, можно подсчитать, какая доля от общего числа объектов приходится на то или иное значение.

Таким образом, многие статистические исследования производят подсчеты, а не измерения. Опрос общественного мнения спрашивает у выборки людей ответы на вопросы, данные которых представляют собой категории «Да», «Нет» и «Не знаю». Либо в эксперименте сравнивается четыре вида лечения заболевания; данные представляют собой число субъектов, которым назначена каждая процедура и эффективность лечения.

Данные в таких экспериментах – количество подсчетов для объектов каждой категории. Параметры, о которых мы хотим сделать вывод, – это пропорции (доли) в генеральной совокупности. Схема сравнительного анализа и получаемые выводы о пропорциях в исследуемых выборках очень похожи на таковые для средних. На основе выборки мы с определенной долей вероятности делаем выводы о генеральной совокупности, при этом можем иметь дело с одной выборочной совокупностью, со сравнением двух или более выборок.

5.1. СРАВНЕНИЕ ДВУХ ДОЛЕЙ

Поскольку сравнительные исследования широко распространены, часто возникает задача сравнить доли двух групп (например, мужчины и женщины), которые имеют некоторую характеристику. Сравниваемые данные состоят из двух независимых случайных выборок, размером n_1 из генеральной совокупности 1 и размером n_2 из генеральной совокупности 2. Доля носителей признака в каждой выборке оценивает соответствующую долю в генеральной совокупности. Приведем подробнее обозначения, которые мы будем использовать в этом разделе (табл. 29).

Таблица 29

Используемые обозначения

Генеральная совокупность	Доля носителей признака в генеральной совокупности	Размер (объем) выборки	Количество носителей признака в выборке	Доля носителей признака в выборке
1	p_1	n_1	x_1	$\hat{p} = x_1 / n_1$
2	p_2	n_2	x_2	$\hat{p} = x_2 / n_2$

Выборочная доля оценивает неизвестную исследователю долю генеральной совокупности. Обычно последняя намного больше, чем выборка (скажем, по крайней мере, в 10 раз), отдельные наблюдения почти независимы и приблизительно описываются биномиальным распределением. Однако мы знаем, что когда выборка велика, доля выборки $\hat{p}$ имеет примерно нормальное распределение, поэтому в таком случае можно рассматривать процедуры вывода, основанные на нормальном приближении. Такие процедуры будут аналогичны методам определения среднего значения нормального распределения (а сравнение двух групп будет похоже на тест Стьюдента). Если же размер выборки мал, нужно базировать тесты для определения p на биномиальных распределениях, работать с которыми менее удобно из-за их дискретности.

В статистических терминах при сравнении долей (как и в случае сопоставления количественных признаков) мы имеем дело с выводом о вероятности справедливости нулевой гипотезы об отсутствии различий, т. е. проверяем H_0: $p_1 = p_2$.

Для сравнения долей используется критерий *z,* аналогичный *t*-критерию Стьюдента:

$$z = \frac{\text{Разность выборочных долей}}{\text{Стандартная ошибка разности выборочных долей}}.$$

Пусть $\hat{p}_1$ и $\hat{p}_2$ – выборочные доли в исследуемых выборках. Поскольку стандартная ошибка – это стандартное отклонение всех возможных значений $\hat{p}$, полученных по выборкам заданного объема, а также дисперсия разности равна сумме дисперсии, стандартная ошибка разности долей

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{s^2_{\hat{p}_1} + s^2_{\hat{p}_2}}.$$

Следовательно, *z*-критерий для сравнения двух выборочных долей имеет вид

$$z = \frac{\hat{p}_1 - \hat{p}_2}{s_{\hat{p}_1 - \hat{p}_2}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{s^2_{\hat{p}_1} + s^2_{\hat{p}_2}}}.$$

Если сравниваемые выборки имеют объемы n_1 и n_2, то дисперсии этих выборок

$$s^2_{\hat{p}_1} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \qquad s^2_{\hat{p}_2} = \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}$$

Другими словами, развернутую формулу для вычисления z-критерия для разности долей можно записать следующим образом:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}.$$

Напомним, что буквой z обозначаются величины со стандартным нормальным распределением, т. е. нормальным распределением со средним равным 0 и стандартным отклонением равным 1. В случае сравнения долей нормальное распределение имеет место только при достаточно больших объемах выборок: когда значения np и $n(1-p)$ больше пяти. Если хотя бы для одной выборки это условие не выполняется, то критерий z не применим, и нужно воспользоваться точным критерием Фишера (разд. 5.4).

Как и в случае t-критерия, о статистически значимом различии долей можно говорить, если значение z окажется достаточно большим. Отличие состоит в том, что величина t подчиняется распределению Стьюдента, а z – стандартному нормальному распределению. Соответственно, для нахождения больших значении z нужно воспользоваться стандартным нормальным распределением (см. рис. 14).

Поскольку при увеличении числа степеней свободы распределение Стьюдента стремится к нормальному, критические значения z можно найти в последней строке ($n = \infty$) таблиц критических значений t (для пятипроцентного уровня значимости оно составляет 1,96, для однопроцентного – 2,58).

Нормальное распределение служит лишь приближением для распределения z. При этом оценка P оказывается заниженной, и нулевая гипотеза будет отвергаться слишком часто. Причина состоит в том, что z принимает только дискретные значения, тогда как приближающее его нормальное распределение непрерывно. Для компенсации излишнего «оптимизма» критерия z введена *поправка Йейтса*, называемая также *поправкой на непрерывность*. С учетом данной поправки выражение для z имеет следующий вид:

$$z = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Поправка Йейтса слегка уменьшает значение z, уменьшая тем самым расхождение с нормальным распределением.

Пример. Теперь можно сравнить две группы по какому-либо качественному признаку, например доля курящих среди молодых людей в возрасте 20–25 лет. Результаты опроса приведены в табл. 30.

Таблица 30

Доля курящих среди молодых людей в возрасте 20–25 лет

Пол	n	x	$\hat{p}$
Юноши	7180	1630	0,227
Девушки	9916	1684	0,170
Всего	**17 096**	**3314**	**0,194**

Итак, среди опрошенных юношей курят (22,7 %), а среди девушек – 17 %. Объединенная оценка доли курящих составила 19,4 %.

Величина np для каждой из выборок равна соответственно $n_1 p_1 = 7180 \cdot 0{,}227$ и $n_2 p_2 = 9916 \cdot 0{,}170$. Оба значения значительно больше пяти, поэтому можно воспользоваться приведенным выше критерием с учетом поправки Йейтса. Ввиду большой выборки получили величину $z = 9{,}34$, которая оказалась гораздо выше 1,96 – критического значения для пятипроцентного уровня значимости. Рассчитанная величина $P < 0{,}0005$, следовательно, различия статистически значимы.

5.2. СООТВЕТСТВИЕ ДОЛИ В ОДНОЙ ВЫБОРКЕ ТЕОРЕТИЧЕСКОМУ ЗНАЧЕНИЮ

Предположим, нужно оценить долю p некоторой качественной характеристики, такой как количество мужчин среди большой группы населения, например страны. Выбираем простую случайную выборку размера n из генеральной совокупности и записываем количество утвердительных ответов.

Напомним, что доля выборки $\hat{p}$ имеет приблизительно нормальное распределение со средним $\mu_p = p$ и стандартным отклонением σ:

$$\sigma = \sqrt{p(1-p)/n}.$$

При выполнении теста значимости нулевая гипотеза указывает значение для $\hat{p}$, и мы предполагаем, что оно соответствует истинному значению в генеральной совокупности при расчете значения P (проверяем H_0: $p = p_0$).

Для теста проводим стандартизацию $\hat{p}$ и вычисляем z-статистику одиночной выборки, на основе которой делаем заключение о значимости различий:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0\left(1 - p_0\right)}{n}}}.$$

Исходя из распределения переменной Z и формулировки альтернативной гипотезы, находим величину P (рис. 41).

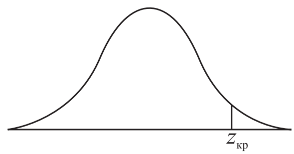
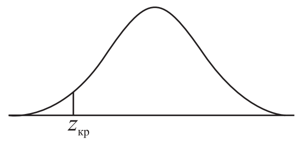
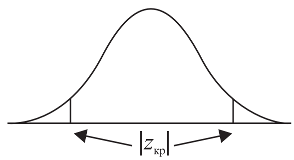
Если H_A: $p > p_0$, то P $(Z_{\text{эмп}} \geq z_{\text{кр}})$	
Если H_A: $p < p_0$, то P $(Z_{\text{эмп}} \leq z_{\text{кр}})$	
Объединяя два приведенных неравенста: если H_A: $p \neq p_0$, то P $(Z_{\text{эмп}} \geq \lvert z_{\text{кр}} \rvert)$	

Рис. 41. Определение величины P

Мы рекомендуем применение такого теста значимости z для большой выборки, если ожидаемое количество в каждой из групп – np_0 и $n(1-p_0)$ больше 10. Если это условие не соблюдается, следует использовать другие процедуры, например точный критерий Фишера (см. разд. 5.4).

Пример. Французский натуралист Ж.-Л. Бюффон однажды бросил монету 4040 раз и подсчитал число выпавших орлов. Доля их составила

$$p = 2048 / 4040 = 0{,}5069.$$

Если бы монета Ж.-Л. Бюффона была точно сбалансирована, то вероятность выпадения орла при любом подбрасывании была бы равна 0,5. Чтобы оценить, являются ли полученные данные свидетельством того, что монета не была сбалансирована, тестируем

$$H_0\colon p = 0{,}5,$$
$$H_{\text{A}}\colon p \neq 0{,}5.$$

Статистика теста считается следующим образом:

$$z = \frac{0,5069 - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{4040}}} = 0,88.$$

Рассчитанное значение P (рис. 42) равно 0,38. Значит, полученные в эксперименте данные совместимы с гипотезой сбалансированной монеты. Поскольку вычисления такого типа очень просты, в некоторые программы статистической обработки они не входят.

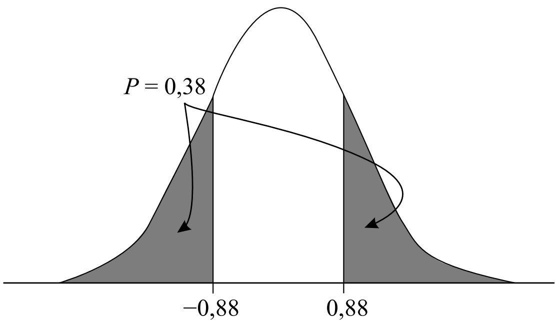


Рис. 42. Расчет двустороннего теста и значения P для примера с монетой

Если мы захотим для этой же задачи посчитать количество решек, просто поменяем при вычислениях знак итоговой тестовой статистики z ($z = -0,88$). Значение P останется прежним.

Наш вывод базируется на предположении, что монета должна быть идеально сбалансирована, тогда истинная вероятность выпадения орла будет точно 0,5. Тест значимости в рассмотренном примере демонстрирует то, что монета Ж.-Л. Бюффона статистически неотличима от идеальной монеты на основе 4040 бросков.

5.3. СРАВНЕНИЕ ДВУХ И БОЛЕЕ ГРУПП

Рассмотренный в разд. 5.1 подход хорошо работает, если качественный признак, который нас интересует, принимает два значения (черный – белый, самец – самка, да – нет). Более того, поскольку метод является прямым аналогом критерия Стьюдента, число сравниваемых выборок также должно быть равно двум.

Понятно, что при исследованиях и число значений признака, и число выборок может оказаться больше двух. Для анализа таких случаев нужен иной метод, аналогичный дисперсионному анализу. Так, с виду подход, который мы изложим, сильно отличается от критерия z, но в действительности между ними много общего.

Рассмотрим принцип метода на следующем примере. Совокупность из 44 мышей разделили на две группы по 19 и 25 мышей. Первой группе сделали инъекцию, содержащую патогенные бактерии, с последующим введением сыворотки. Животные из второй группы служили контролем: им сделали только бактериальные инъекции. После некоторого времени инкубации оказалось, что 24 мыши погибли, а 20 выжили. Из погибших 6 принадлежали к экспериментальной группе (вводили сыворотку), а 25 – к контрольной. Вопрос: повышает ли введение сыворотки выживаемость мышей? Иными словами, имеются ли статистически значимые различия между долями, отражающими выживаемость мышей в обеих группах?

Отметим, что при построении критерия будем рассматривать не долю, а число мышей. Занесем результаты испытания в таблицу (табл. 31). Для каждой из групп укажем число мышей. У нас два признака: введение сыворотки (да/нет) и выживаемость (да/нет). В таблице указаны все их возможные сочетания, поэтому она называется таблицей сопряженности. В данном случае размер таблицы будет 2×2 (две строки, два столбца), ее еще называют четырехпольной таблицей.

Таблица 31

Выживаемость мышей при введении сыворотки

Вариант опыта	Погибло	Выжило
Контроль	18	7
Сыворотка	6	13

Посмотрим на клетки, расположенные по диагонали, идущей из верхнего левого в нижний правый угол таблицы. Числа в них заметно больше чисел в других клетках. Это наводит на мысль о связи между введением сыворотки и гибелью мышей.

Теперь обратите внимание на табл. 32. Это таблица ожидаемых чисел, которые мы получили бы, если бы введение сыворотки не влияло на выживаемость мышей. Как рассчитать ожидаемые числа разберем позже, а пока сфокусируемся на особенностях таблицы. Кроме дробных чисел в клетках, можно заметить еще одно отличие от табл. 31 – суммарные данные по строкам в правом столбце и столбцам в нижней строке. В правом нижнем углу мы получили цифру, отражающую общее число мышей, участвующих в испытании. Обратите внимание, что, хотя числа в клетках табл. 31 и 32 разные, суммы по строкам и столбцам будут одинаковы.

Таблица 32

Выживаемость мышей при введении сыворотки: ожидаемые числа, %

Вариант опыта	Погибло	Выжило	Всего
Контроль	13,64	11,36	25,00
Сыворотка	10,36	8,64	19,00
Всего	**24**	**20**	**44**

Как рассчитать ожидаемые числа? Инъекции только патогенных бактерий (контроль) получали 25 мышей, дополнительные инъекции сыворотки – 19. Погибли всего 24 особи из 44, принимавших участие в эксперименте, что составило 54,55 % случаев, выжили 20 из 44, т. е. в 45,45 % случаев.

Примем нулевую гипотезу о том, что введение сыворотки не влияет на выживаемость мышей. Тогда их гибель должна с равной частотой 54,55 % наблюдаться в контрольной и экспериментальной группах. Рассчитав, сколько составляет 54,55 % от 25 и 19, получим соответственно 13,64 и 10,36 – это и есть ожидаемые числа погибших мышей в контрольной и экспериментальной группах. Таким же образом можно получить ожидаемые числа выживших – 45,45 % от 25, т. е. 11,36 в контрольной группе, и 45,45 % от 19 – это 8,64 в экспериментальной. Обратите внимание, что ожидаемые числа рассчитываются до второго знака после запятой. Такая точность понадобится при дальнейших вычислениях.

Сравним табл. 31 и 32. Числа в клетках довольно сильно различаются, следовательно, реальная картина отличается от той, которая наблюдалась бы, если бы сыворотка не оказывала влияния на выживаемость мышей. Осталось сформулировать критерий, который характеризовал бы данные различия одним числом, а затем найти его критическое значение, т. е. поступить так, как в случае критериев F, t или z.

Критерий χ^2 не требует никаких предположений относительно параметров совокупности, из которой извлечены выборки, – это один из непараметрических критериев.

Критерий χ^2 вычисляется следующим образом:

$$\chi^2 = \sum \frac{(F - T)^2}{T},$$

где F – фактическое (наблюдаемое) число в клетке исходной таблицы сопряженности; T – теоретически ожидаемое число в той же клетке таблицы ожидаемых чисел.

Суммирование проводится по всем клеткам таблицы. Как видно из формулы, чем больше разница наблюдаемого и ожидаемого чисел, тем больше вклад вносит клетка в величину χ^2. При этом клетки с малым ожидаемым числом вносят большой вклад. Таким образом, критерий представляет собой одно число, которое, во-первых, измеряет различия, во-вторых, учитывает их величину относительно ожидаемых чисел.

Применение χ^2 правомерно, если *ожидаемое число в любой из ячеек таблицы больше или равно пяти* (иначе мы должны прибегнуть к точному критерию Фишера).

Критическое значение χ^2 зависит от размеров таблицы сопряженности, т. е. от количества строк и столбцов. Размер таблицы выражается числом степеней свободы: $v = (R - 1)(C - 1)$, где R – количество строк; C – количество столбцов. Легко подсчитать, что для 2×2 таблицы имеем $v = 1$.

Применим критерии χ^2 к данным по введению сыворотки мышам (см. табл. 31 и 32):

$$\chi^2 = \sum \frac{(F-T)^2}{T} = \frac{(18-13{,}64)^2}{13{,}64} + \frac{(7-11{,}36)^2}{11{,}36} +$$
$$+ \frac{(6-10{,}36)^2}{10{,}36} + \frac{(13-8{,}64)^2}{8{,}64} = 7{,}10.$$

Это много или мало? Для ответа найдем критическое значение χ^2.

На рис. 43 показано распределение возможных значений χ^2 для таблиц сопряженности размером 2×2 для случая, когда между изучаемыми признаками нет никакой связи. Величина χ^2 превышает 3,84 только в 5 % случаев. Таким образом, 3,84 – критическое значение χ^2 для пятипроцентного уровня значимости.

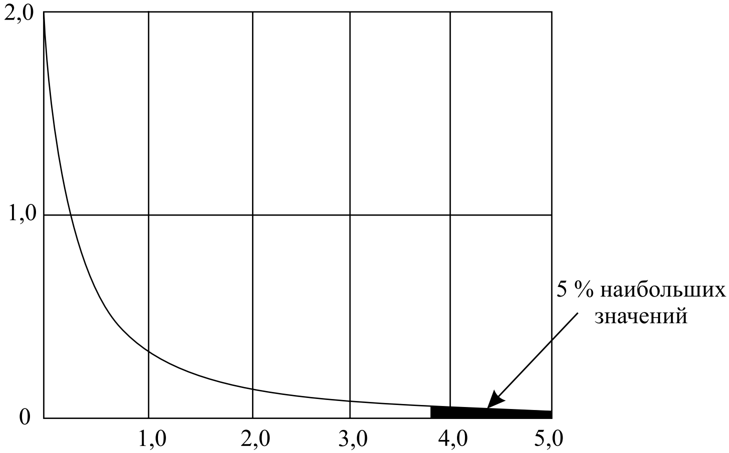


Рис. 43. Распределение χ^2 с одной степенью свободы (v = 1)

Критическое значение χ^2 зависит от размеров таблицы сопряженности. Так, в нашем примере строк таблицы, т. е. сравниваемых вариантов опыта, могло быть больше (например, можно добавить группу, которой вводили только сыворотку), или в другом примере у нас было бы больше число возможных исходов (столбцов таблицы).

Размер таблицы определяет число степеней свободы v. Критические значения χ^2 для разных v можно найти в специальных таблицах (см., например, Гланц, 1999), программы статистической обработки рассчитывают его самостоятельно, выводя на выходе величину P.

В примере с мышами мы получили значение 7,10, поэтому отклоняем нулевую гипотезу об отсутствии связи между введением сыворотки и выживаемостью мышей.

Разумеется, как и все критерии значимости, χ^2 дает вероятностную оценку истинности той или иной гипотезы. Другими словами, в действительности сыворотка может и не оказывать влияния на выживаемость мышей. Однако, как показал критерий, это маловероятно.

Применение критерия χ^2 правомерно, если ожидаемое число в любой из клеток больше или равно пяти. Это условие аналогично условию применимости критерия z. В противном случае мы вынуждены использовать точный критерий Фишера.

Формула для χ^2 в случае таблицы 2×2 (т. е. при одной степени свободы) дает несколько завышенные значения (сходная ситуация была с критерием z). Это вызвано тем, что теоретическое распределение χ^2 непрерывно, тогда как набор вычисленных значений χ^2 дискретен. На практике это приведет к тому, что нулевая гипотеза будет отвергаться слишком часто. Чтобы компенсировать данный эффект, в формулу, так же как и в случае z-критерия, вводят поправку Йейтса. Заметим, что такая поправка применяется только при $df = 1$, т. е. для таблиц 2×2.

Применим поправку Йейтса к изучению связи между введением сыворотки и гибелью мышей (см. табл. 31, 32):

$$\chi^2 = \frac{\left(|18 - 13,64| - \frac{1}{2}\right)^2}{13,64} + \frac{\left(|7 - 11,36| - \frac{1}{2}\right)^2}{11,36} +$$

$$+ \frac{\left(|6 - 10,36| - \frac{1}{2}\right)^2}{10,36} + \frac{\left(|13 - 8,64| - \frac{1}{2}\right)^2}{8,64} = 5,57.$$

Если без поправки значение χ^2 равнялось 7,10 и было больше 6,635 – критического значения для однопроцентного уровня значимости, то скорректированное значение χ^2 оказалось меньше (т. е. для однопроцентного уровня значимости мы уже не признали бы различия значимыми). Однако даже скорректированное по Йейтсу значение χ^2 по-прежнему превосходит 5,024 – критическое значение для 2,5 % уровня значимости.

Рассмотренный критерий χ^2 может применяться для таблицы сопряженности произвольной формы, в том числе, когда число строк или столбцов больше двух. Обратите внимание, что критерий z в таких случаях неприменим.

Итак, пошагово пропишем *порядок применения критерия* χ^2:

1) постройте по имеющимся данным таблицу сопряженности;

2) подсчитайте число объектов в каждых строке и столбце и найдите, какую долю от общего числа объектов составляют эти величины;

3) зная доли, подсчитайте (с точностью до двух знаков после запятой) ожидаемые числа – количество объектов, которое попало бы в каждую клетку таблицы, если бы связь между строками и столбцами отсутствовала;

4) найдите величину, характеризующую различия наблюдаемых и ожидаемых значений. Если таблица сопряженности имеет размер 2×2, примените поправку Йейтса;

5) вычислите число степеней свободы, выберите уровень значимости и по таблицам определите критическое значение χ^2. Сравните его с полученным для вашей таблицы.

Как вы помните, для таблиц сопряженности размером 2×2 критерий χ^2 применим только тогда, когда все ожидаемые числа больше пяти. В случае с большими таблицами критерий применим, если все ожидаемые числа не меньше еденицы и доля клеток с ожидаемыми числами меньше пяти не превышает 20 %. При невыполнении перечисленных условий критерии χ^2 может дать ложные результаты. В таком случае можно собрать дополнительные данные, однако это не всегда осуществимо. Тем не менее есть и более простой путь – объединить несколько строк или столбцов.

Отметим, что χ^2 позволяет установить существование различий между группами по качественному признаку. Однако по его результату нельзя определить, какие именно группы отличаются друг от друга. С похожей ситуацией мы сталкивались в дисперсионном анализе: при сравнении нескольких групп дисперсионный анализ позволяет обнаружить непосредственно факт существования различий, но не указывает выделяющиеся группы. Последнее позволяют сделать процедуры множественного сравнения.

Так, если бы в примере с мышами тестировали два вида сыворотки, то могли бы сравнить эти две группы с контрольной. Вначале проверяем гипотезу о том, что введение сыворотки двух типов одинаково влияет на выживаемость мышей. Для этого выделим из исходной таблицы подтаблицу, содержащую данные по этим двум группам. Если наблюдаемые и ожидаемые числа в новой таблице окажутся довольно близки, вычисленный χ^2 с поправкой Йейтса даст величину значительно меньше критического значения, поэтому гипотеза об отсутствии межгрупповых различий остается в силе. Следовательно, данные группы можно объединить в одну. Полученную группу мышей, получавших сыворотку, сравним с контрольной. Если сыворотка влияет на их выживаемость, значение χ^2 окажется больше критического для соответствующего уровня значимости. Заметьте, мы выполнили два сравнения, используя одни и те же данные, поэтому в этой ситуации нужно применить поправку Бонферрони (см. разд. 3.4).

5.4. ТОЧНЫЙ КРИТЕРИЙ ФИШЕРА

Когда число наблюдений небольшое, используют *точный критерий Фишера*, который основан на переборе всех возможных вариантов заполнения таблицы сопряженности при данной численности групп, поэтому чем они меньше, тем проще его применить.

Значение P рассчитывается из данных 2×2 таблицы по формуле

$$P = \frac{\dfrac{R_1!R_2!C_1!C_2!}{N!}}{O_{11}!O_{12}!O_{21}!O_{22}!},$$

где R_1 и R_2 – маргинальные строковые суммы; C_1 и C_2 – маргинальные суммы по столбцам; O_{11}, O_{12}, O_{21} и O_{22} – числа в клетках 2×2 таблицы; N – общее число наблюдений.

Восклицательный знак, как и всегда в математике, обозначает факториал (факториал числа – произведение целых чисел от этого числа до единицы; факториал нуля равен единице).

Обозначения, используемые в точном критерии Фишера, показаны на рис. 44.

			Суммы по строкам
	O_{11}	O_{12}	$\boldsymbol{R_1}$
	O_{21}	O_{22}	$\boldsymbol{R_1}$
Суммы по столбцам	$\boldsymbol{C_1}$	$\boldsymbol{C_2}$	$\boldsymbol{N}$

Рис. 44. Обозначения, используемые в точном критерии Фишера

Построив все варианты заполнения таблицы, возможные при данных суммах по строкам и столбцам, по этой же формуле рассчитывают и их вероятность. *Вероятности, которые не превосходят вероятность исходной таблицы (включая и эту вероятность), суммируют.* Полученная сумма есть величина P для двустороннего варианта точного критерия Фишера.

Существуют одно- и двусторонний варианты точного критерия Фишера. Во многих пособиях описан именно односторонний вариант, он же обычно используется в компьютерных программах и приводится в статьях. Оно и неудивительно, ведь односторонний вариант дает меньшую величину *P*.

Ниже показаны данные, которые получили Мак-Кинни и соавторы (цит. по Гланцу, 1999), решив выяснить, насколько часто в статьях из двух известных медицинских журналов указан вариант критерия. Выборка, как видно из табл. 33, невелика, потому критерий χ^2 применить нельзя. Таким образом, для анализа использования в журналах точного критерия Фишера воспользуемся им же.

Таблица 33

Частота указания варианта точного критерия Фишера в двух медицинских журналах

Журнал	Критерий		**Всего**
	указан	не указан	
New England Journal of Medicine	1	8	**9**
Lancet	10	4	**14**
Всего	**11**	**12**	**23**

Из приведенной выше формулы для *P* следует, что вероятность при тех же значениях сумм по строкам и столбцам таблицы получить такой же набор чисел в клетках, как в табл. 33:

$$P = \frac{\frac{9!14!11!12!}{23!}}{1!8!10!4!} = 0,00666.$$

Это небольшая вероятность. Теперь возьмем наименьшее из чисел в клетках (это единица в ячейке O_{11}, что на пересечении первой строки и первого столбца) и уменьшим его на 1. Числа в остальных клетках изменим так, чтобы суммы по строкам и столбцам остались прежними (табл. 34).

Таблица 34

Изменение данных табл. 33 для расчета точного критерия Фишера

Журнал	Критерий		**Всего**
	указан	не указан	
New England Journal of Medicine	0	9	**9**
Lancet	11	3	**14**
Всего	**11**	**12**	**23**

Рассчитаем вероятность получить такую измененную таблицу:

$$P = \frac{\frac{9!14!11!12!}{23!}}{1!9!11!3!} = 0,00027.$$

Обратите внимание, что числитель можно заново не вычислять, так как его значение зависит только от сумм по строкам и столбцам, которые не изменились.

Поскольку наименьшее число в ячейке O_{11} теперь равно нулю, дальше уменьшать его невозможно.

Таким образом, односторонний вариант точного критерия Фишера дает $P = 0{,}00666 + 0{,}00027 = 0{,}00695$.

Чтобы рассчитать значение двустороннего варианта точного критерия Фишера, нужно перебрать и все остальные возможные варианты заполнения таблицы при условии неизменности сумм по строкам и столбцам. Получить данные варианты несложно, нужно только заметить, что при постоянных суммах по строкам и столбцам значения во всех четырех клетках полностью определяются значением в любой из них. К примеру, возьмем число в левой верхней клетке и будем увеличивать его на 1, пересчитывая каждый раз числа в остальных клетках. В результате получим еще восемь вариантов заполнения табл. 33 (рис. 45).

			Всего
	2	7	9
	9	5	14
Всего	11	12	23

$P = 0{,}05330$

			Всего
	6	3	9
	5	9	14
Всего	11	12	23

$P = 0{,}12438$

			Всего
	3	6	9
	8	6	14
Всего	11	12	23

$P = 0{,}18657$

			Всего
	7	2	9
	4	10	14
Всего	11	12	23

$P = 0{,}02665$

			Всего
	4	5	9
	7	7	14
Всего	11	12	23

$P = 0{,}31983$

			Всего
	8	1	9
	3	11	14
Всего	11	12	23

$P = 0{,}00242$

			Всего
	5	4	9
	6	8	14
Всего	11	12	23

$P = 0{,}27985$

			Всего
	9	0	9
	2	12	14
Всего	11	12	23

$P = 0{,}00007$

Рис. 45. Возможные варианты заполнения табл. 33

Для двух последних вариантов вероятность не превышает вероятности исходного варианта заполнения (0,00666), составляя соответственно 0,00242 и 0,00007. Таким образом, для всех возможных вариантов заполнения таблицы, кроме исходного, у нас есть еще три варианта маловероятного заполнения. Просуммировав соответствующие вероятности и прибавив к ним вероятность исходного варианта, получим

$$P = 0{,}00666 + 0{,}00027 + 0{,}00242 + 0{,}00007 = 0{,}00944.$$

Это и есть значение двустороннего варианта точного критерия Фишера.

Итак, различие частоты правильного использования точного критерия Фишера в журналах *New England Journal of Medicine* и *Lancet* статистически значимо ($P = 0{,}009$). В данном случае общий вывод при переходе от одностороннего к двустороннему варианту не изменился, однако так бывает далеко не всегда. Еще более грубая ошибка происходит, когда автор рассчитывает только вероятность получения исходной таблицы, пренебрегая построением остальных вариантов заполнения. Естественно, это приводит к сильному занижению P, т. е. к выявлению различий там, где их нет.

В заключение изложим *алгоритм вычисления точного критерия Фишера*:

1) определите вероятность получить исходную таблицу;

2) постройте остальные возможные варианты заполнения таблицы при неизменных суммах по строкам и столбцам. Для этого в одной из клеток проставьте все целые числа от нуля до максимально возможного, пересчитывая числа в остальных клетках так, чтобы суммы по строкам и столбцам оставались неизменными;

3) вычислите вероятности для всех полученных таблиц;

4) просуммируйте вероятность получить исходную таблицу и все вероятности, которые ее не превышают.

6. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

До настоящего момента мы в основном занимались нахождением различий между группами, не слишком интересуясь величиной этих различий, формулировали нулевую гипотезу, например, предполагали, что экспериментальные группы – это просто две случайные выборки из одной и той же совокупности. Затем оценивали вероятность получить наблюдаемые различия при условии, что нулевая гипотеза верна. Если эта вероятность была мала, отвергали нулевую гипотезу и делали вывод, что различия статистически значимы.

При таком подходе мы всегда получаем только качественный результат: либо отклоняем нулевую гипотезу, либо нет, либо признаем различия статистически значимыми, либо нет. Однако количественная оценка различий от нас ускользает. Между тем вероятность выявления различий зависит не только от их величины, но и от численности групп. Даже небольшие различия при достаточно большой численности групп могут оказаться статистически значимыми. При этом речь может идти о совсем небольшой разнице в несколько миллиметров ртутного столба.

Характеристика, которая дополняет и даже заменяет качественное суждение «значимо – незначимо», – *доверительный интервал*.

Вспомним свойства нормального распределения и применим его к рассуждениям о стандартной ошибке среднего. Используя правило трех сигм можно утверждать, что истинное среднее (среднее генеральной совокупности) в 95 % случаев лежит на расстоянии не больше двух ошибок среднего от выборочного среднего. Промежуток длиной в четыре ошибки среднего – это и есть 95-процентный доверительный интервал. Смысл доверительного интервала из этого примера достаточно ясен: неизвестно, чему равна некоторая величина, но с заданной вероятностью можно указать интервал, в котором она находится.

Доверительные интервалы часто используют для определения границ нормы лабораторного показателя (в биологии и медицине).

В разд. 3 мы определили критерий Стьюдента как

$$t = \frac{\text{Разность выборочных средних}}{\text{Стандартная ошибка разности выборочных средних}}.$$

Вычислив t, его сравнивают с критическим значением t_α для заданного уровня значимости α. Для двух случайных выборок из одной совокупности вероятность получить значение t, по абсолютной величине превышающее t_α, весьма мала, а именно, не превышает α (напомним, что уровень значимости α – максимально приемлемая вероятность ошибочно признать существование различий там, где их нет). Поэтому, получив достаточно большое значение t, мы делаем вывод о статистической значимости различий.

Для случайных выборок, извлеченных из одной совокупности, распределение всех возможных значений t (распределение Стьюдента) симметрично относительно среднего, равного нулю (см. рис. 14, 15). Если же выборки извлечены из двух совокупностей с разными средними, то распределение всех возможных значений t будет иметь среднее, отличное от нуля.

Формулу для t можно видоизменить так, чтобы распределение t всегда было симметрично относительно нуля. Рассмотрим математическую запись новой формулы

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\bar{\mu}_1 - \bar{\mu}_2\right)}{s_{\bar{x}_1 - \bar{x}_2}}.$$

Заметим, что если обе выборки извлечены из одной совокупности, то разность истинных средних равна нулю и в этом случае новая формула совпадает с предыдущей.

Поскольку истинных средних (средних генеральной совокупности) мы не знаем, то и вычислить значение t по этой формуле нельзя, однако она позволяет сделать другое – оценить разность $\mu_1 - \mu_2$, т. е. истинную величину различий. Для этого вместо вычисления t выберем его подходящее значение и, подставив в формулу, определим величину $\mu_1 - \mu_2$. Как выбрать подходящее значение?

По определению 100α % всех возможных значений t расположены левее $-t_\alpha$ или правее $+t_\alpha$. Остальные 100(1 – α) % значений t попадают в интервал от $-t_\alpha$ до $+t_\alpha$. К примеру, 95 % значений t находятся в интервале от $-t_{0,05}$ до $+t_{0,05}$ (критические значения t, в частности $t_{0,05}$, можно найти в таблицах критических значений (см. Гланц, 1999)).

Значит, в 100(1 – α) % всех случаев

$$-t_\alpha < \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\bar{\mu}_1 - \bar{\mu}_2\right)}{s_{\bar{x}_1 - \bar{x}_2}} < +t_\alpha,$$

где t_α – критическое значение t для уровня значимости α и числа степеней свободы $v = n_1 + n_2 - 2$.

Преобразуя это неравенство, получаем

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_\alpha s_{\bar{x}_1 - \bar{x}_2} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_\alpha s_{\bar{x}_1 - \bar{x}_2}.$$

Таким образом, разность истинных средних отличается от разности выборочных средних менее чем на произведение t_α и стандартной ошибки разности выборочных средних. Это неравенство задает доверительный интервал для разности средних $\mu_1 - \mu_2$. К примеру, 95-процентный доверительный интервал для разности средних определяется неравенством

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{0,05} s_{\bar{x}_1 - \bar{x}_2} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{0,05} s_{\bar{x}_1 - \bar{x}_2}.$$

В этот интервал разность истинных средних попадает в 95 % случаев. Значения левой и правой частей приведенного неравенства называют соответственно *нижним* и *верхним доверительным пределами* и обозначают L_1 и L_2.

Важно помнить, что данный способ определения доверительного интервала, как и критерий Стьюдента, на котором он базируется, можно применять только тогда, когда совокупность имеет хотя бы приближенно нормальное распределение. Доверительные интервалы (как и тесты значимости) для среднего в нормально распределенной генеральной совокупности основаны на среднем выборочном, которое прогнозирует неизвестное μ. Выборочное распределение зависит от σ. Для прогнозирования стандартного отклонения σ в генеральной совокупности мы также используем выборочное стандартное отклонение.

Формула для $100(1 - \alpha)$-процентного доверительного интервала для средней арифметической похожа на уже выведенную ранее:

$$\bar{x} - t_\alpha s_x < \mu < \bar{x} + t_\alpha s_x.$$

Отметим, что если объем выборки достаточно велик (от 20 и выше), то можно воспользоваться правилом двух стандартных ошибок и в качестве t_α использовать $t_{0,05}$, приблизительно равное двум.

Аналогичную формулу можно записать и для нахождения доверительного интервала доли:

$$\hat{p} - t_\alpha s_{\hat{p}} < p < \hat{p} + t_\alpha s_{\hat{p}}.$$

Смысл доверительного интервала для среднего (доли) совершенно аналогичен смыслу доверительного интервала для разности средних (долей).

Напомним, что приведенный доверительный интервал для доли можно использовать, когда объем выборки достаточно велик (и данные можно приблизительно описать, используя нормальное распределение). Когда выборка мала (в исследованиях так бывает нередко) и распределение отличается от нормального, рассчитывать доверительные интервалы на основе стандартных ошибок среднего недопустимо. В таких случаях приходится вычислять точные значения доверительных интервалов, используя биномиальное распределение. Заметим, что в случае соответствия данных закону нормального распределения, при оценке долей по выборкам небольшого объема расчет доверительного интервала особенно желателен. Причина в том, что, если выборка мала, изменение признака даже у одного из ее членов приведет к резкому изменению долей.

Обычно рассчитывают 95-процентный или 99-процентный доверительные интервалы, однако в зависимости от задач исследования можно рассчитать и другие величины доверительного интервала.

К примеру, вернемся к задаче оценки улучшения разговорного английского языка у преподавателей после цикла занятий (см. табл. 9) по результатам тестирования до начала и по окончании тренинга.

Рассчитаем 90-процентный доверительный интервал для среднего увеличения во всей генеральной совокупности, для этого из таблицы критических значений найдем критическое значение $t_{0,1} = 1,729$.

Доверительный интервал в таком случае рассчитаем как

$$\overline{x} \pm t\frac{s}{\sqrt{n}} = 2,5 \pm 1,729\frac{2,893}{\sqrt{20}} = 2,5 \pm 1,12 = \left(1,38; 3,62\right).$$

Итак, предполагаемое среднее увеличение составляет 2,5 балла с пределами погрешности в $\pm 1,12$ балла для 90-процентного доверительного интервала. Доверительный интервал показывает, что несмотря на статистическую значимость результата эффект занятий был относительно небольшой.

Ранее мы видели, что разные выборки дают разные оценки среднего и стандартного отклонения. Точно так же по разным выборкам будут получаться разные доверительные интервалы (и это не удивительно, ведь доверительный интервал рассчитывают по среднему и стандартному отклонениям), поэтому для других выборок и доверительный интервал будет иным.

Если бы мы построили по всем возможным выборкам из генеральной совокупности 90-процентные доверительные интервалы, то доля не содержащих истинного значения составила бы ровно 10 %, что соответствует чувствительности критерия.

Заметьте, что доверительный интервал может включать и отрицательные значения, тем самым подтверждая, что выборочные данные не противоречат гипотезе, что в действительности результат может не измениться и даже ухудшиться.

Итак, мы только что убедились, что 90-процентный доверительный интервал может и не содержать истинного значения, однако, как правило, содержит его (в 90 % случаев). Вообще, *истинное значение содержат k процентов k-процентных доверительных интервалов*. Иными словами, в этом случае k – вероятность того, что интервал содержит истинное значение показателя. От этой вероятности k зависит ширина интервала. Чем больше k, тем шире k-процентный доверительный интервал. Посмотрите на рис. 46, где для примера приведены рассчитанные по тем же данным в дополнение к 95-процентному еще 90-процентный и 99-процентный доверительные интервалы для разности двух выборок.

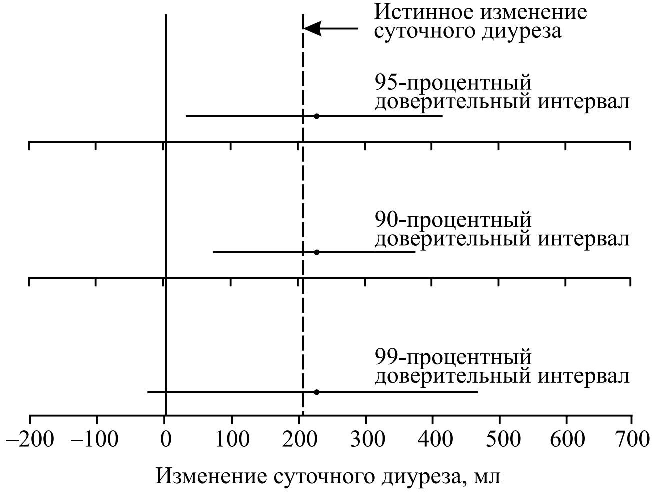


Рис. 46. Три доверительных интервала, рассчитанные для одной и той же разности средних.

На рис. 46 видно, что 90-процентный доверительный интервал более узкий по сравнению с 95-процентным. Однако еще раз подчеркнем, что сужение доверительного интервала произошло в результате снижения вероятности того, что он действительно содержит истинное значение.

Подведем итоги. Приводя *k*-процентный доверительный интервал, мы сообщаем, во-первых, в каких пределах находится истинное значение неизвестной величины, во-вторых, с какой вероятностью *k*. Так, в приведеном ранее примере с английским языком, говоря «90-процентный доверительный интервал составил 1,38–3,62 балла», имеют в виду следующее: «Вероятность того, что истинное значение улучшения разговорного английского языка у преподавателей лежит в пределах 1,38–3,62 балла, составляет 90 %». Не исключено, что истинное значение может оказаться и вне доверительного интервала. С 90-процентными доверительными интервалами такое происходит в 10 % случаев, с 95-процентными – в 5 %.

Доверительные интервалы можно также использовать для оценки статистической значимости различий, поскольку нахождение доверительного интервала имеет общую базу с традиционными методами проверки гипотез. В обоих случаях встречаем разность выборочных средних, ее стандартную ошибку и распределение Стьюдента.

Истинная разность средних может находиться в любой точке доверительного интервала, поэтому *если доверительный интервал содержит ноль, нельзя отвергнуть возможность того, что различий нет*, т. е. нулевую гипотезу. С другой стороны, нахождение истинной разности средних вне доверительного интервала маловероятно, поэтому, если доверительный интервал не содержит нуля, справедливость нулевой гипотезы о равенстве средних маловероятна. Отсюда можно сформулировать правило.

Если $100(1-\alpha)$-процентный доверительный интервал разности средних не содержит нуля, то различия статистически значимы ($P \leq \alpha$), если же интервал содержит ноль, то различия статистически не значимы ($P > \alpha$).

Таким образом, говоря о статистически значимых различиях, всегда полезно привести доверительный интервал, что даст возможность судить о величине различий. Если статистическая значимость обнаружена благодаря большому объему выборки, а не существенной величине различий, доверительный интервал укажет на это. Другими словами, использование доверительных интервалов позволяет среди статистически значимых различий выделить те, которые слишком малы.

Если различия выражены слабо, статистическая значимость будет обусловлена исключительно большой численностью групп. Поэтому, знакомясь с исследованием, важно знать не только уровень значимости, но и величину различий. Авторы публикаций не всегда указывают доверительные интервалы, но обычно все же приводят численность групп, средние величины и их стандартные ошибки либо стандартные отклонения. А значит, в таких случаях можно самостоятельно рассчитать и построить доверительный интервал нужной точности. Этого часто достаточно, чтобы понять, имеет ли исследование сугубо академическую или еще и практическую ценность.

Анализируя данные путем построения доверительного интервала, необходимо не забывать о выполнении двух обязательных условий:

• выборка должна быть случайной;

• распределение должно быть нормальным.

Оба этих требования следует обязательно соблюдать, чтобы утверждение о доверительном интервале было корректным.

Еще раз подчеркнем, что в основе построения доверительного интервала лежит предположение, что в генеральной совокупности изучаемые данные имеют нормальное распределение. Работая с выборкой, никогда нельзя точно сказать, является совокупность нормально распределенной или нет, поскольку имеется только случайная выборка. Поэтому на практике для экспресс-оценки нормальности распределения можно обратиться к гистограмме данных (как вариант, стеблевым или точечным диаграммам), чтобы удостовериться, что распределение является приблизительно нормальным, т. е. не слишком асимметричным и без особо отличающихся экстремальных значений. Если распределение данных на гистограмме выглядит приблизительно нормально, доверительный интервал примерно корректен. Если гистограмма немного асимметрична, необходимо иметь не очень малую выборку ($n \geq 20$). Если гистограмма умеренно асимметрична или имеет несколько умеренно отличающихся значений, необходимо иметь выборку большого размера ($n \geq 30$). Если гистограмма сильно асимметрична или имеет большие экстремальные значения, то это должно поставить под сомнение оценку предполагаемых различий выборок с помощью доверительных интервалов. Для более уверенного понимания распределения анализируемых данных рекомендуем воспользоваться одним из способов, изложенных в прил. 1.

Если условие нормальности распределения не выполняется совсем, например распределение сильно асимметрично, можно попробовать преобразование данных (к примеру, путем логарифмирования), которое может привести к нормальному распределению уже трансформированных данных. Однако в этом случае необходимо помнить, что в результате будет получен доверительный интервал для среднего логарифмов значений генеральной совокупности.

КРАТКАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ УМЕНИЙ: АЛГОРИТМ ДЕЙСТВИЙ

Еще раз напомним, что этап планирования эксперимента (включая схему статистического анализа) важен до начала получения данных и их обработки.

Данные анализируемых выборок служат для выводов о генеральной совокупности, при этом мы исходим из предположения, что работа ведется со случайными выборками, поэтому очень важно избегать ошибок сбора данных. Если есть подозрение в качестве исходных данных, необходимо понимать, что, даже применяя верные подходы анализа, можно получить неправильный результат: работает принцип *garbage in, garbage out**, хорошо известный информатикам и специалистам по компьютерному анализу данных.

До проведения сравнительного анализа желательно предварительно провести разведочный анализ данных и получить их графическое и/или числовое описание (см. Биометрия, ч. 1, с. 78).

Общая схема проведения сравнительного анализа

1. Сформулируйте нулевую гипотезу H_0 (постулирует отсутствие различий между сравниваемыми группами) и альтернативную гипотезу H_A (различия статистически значимы).

2. Проанализируйте имеющиеся данные. Определите, к какому типу данных относятся изучаемые признаки (количественные, качественные, ранговые), определите число групп, объемы выборок в каждой из них. Выявите выбросы и определите, стоит ли их удалять (может повлиять как на выбор теста, так и его результаты).

3. Выберите тест (рис. 47).

Для количественных данных проведите предварительную проверку выборок на нормальность распределения. Если все анализируемые группы имеют нормальное распределение, используем параметрические тесты (хорошо работают при достаточном объеме каждой из анализируемых выборок ($n \geq 20$)). Если хотя бы одна из анализируемых групп не подчиняется закону нормального распределения, используем непараметрические тесты.

Если данные получены путем повторных измерений на одних и тех же объектах – это связанные (зависимые) выборки.

При сравнении двух и более независимых групп необходимым условием применения параметрических методов является однородность дисперсий, если это условие не выполняется, используют непараметрические аналоги.

---

*Дословно с англ. ‘мусор на входе – мусор на выходе’. Фраза подразумевает, что при неверных исходных данных будут получены неверные результаты, даже если алгоритм правилен.

4. Оцените полученную в тесте на основе анализа данных вероятность справедливости нулевой гипотезы P относительно уровня значимости α. Примите или отвергните H_0 при заданном α:

$$P > \alpha \rightarrow \text{сохраняем } H_{0,}$$
$$P \leq \alpha \rightarrow \text{принимаем } H_{A.}$$

5. Для дополнительной информации о величине различий между сравниваемыми выборками рассчитайте доверительный интервал.

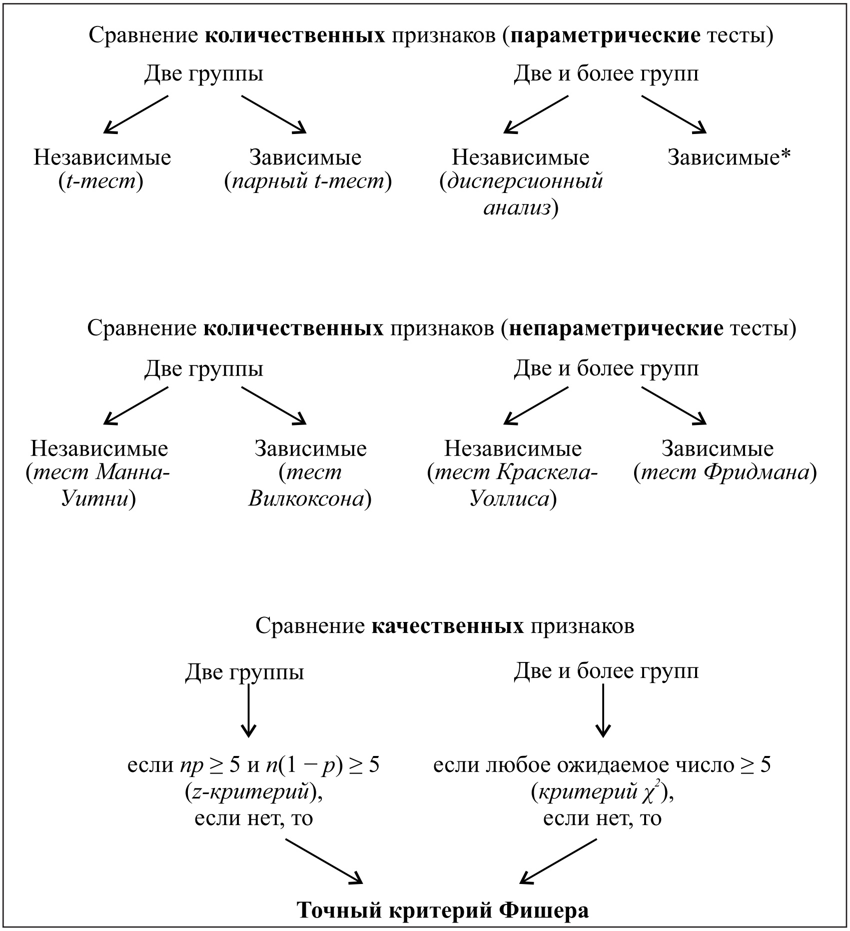


Рис. 47. Схема выбора теста

---

* В пособии тесты данного типа не рассматривались.

Итак, основная идея статистических тестов – моделирование ситуации: как выглядела бы картина, если бы эксперимент повторяли большое количество раз (при условии справедливости одной и той же H_0).

Помните, что многие тесты требуют выполнения определенных условий (чаще всего примерно нормальное распределение данных).

Рассчитанная в ходе статистического теста вероятность (P) оценивает ситуацию с позиции справедливости H_0. Малые значения P (меньше уровня значимости α) говорят в пользу H_A. Тем не менее еще раз подчеркнем, что малые значения P не обозначают, что H_0 неверна, они лишь говорят о том, что это маловероятно.

И основное: статистическая значимость не то же, что значимость практическая. Помните, что большие выборки могут сделать даже совсем малые различия статистически значимыми и, наоборот, очень малые выборки могут не показать статистически значимых различий даже в случае их реального существования. Поэтому очень важно дополнять результат теста о наличии различий количественной характеристикой, т. е. приводить доверительные интервалы.

Доверительные интервалы характеризуют показатели генеральной совокупности на основе выборочных. Так, 95-процентный доверительный интервал показывает границы, в которые показатель попадает в 95 % случаев при многократном повторении эксперимента. Дополняя анализируемые результаты доверительным интервалом, не забывайте, что в сути его построения лежит предположение о нормальном распределении данных изучаемой генеральной совокупности.

ПРИЛОЖЕНИЯ

1. ПРОВЕРКА НОРМАЛЬНОСТИ РАСПРЕДЕЛЕНИЯ ЗНАЧЕНИЙ ПРИЗНАКА С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*

В качестве примера использованы данные о массе (г) 63 самцов взрослых бурозубок (см. Биометрия, ч. 1, с. 22).

Шаг 1. Ввод данных. Вводим весь массив данных ($n = 63$) в один столбец, начиная с первой ячейки (рис. П1).

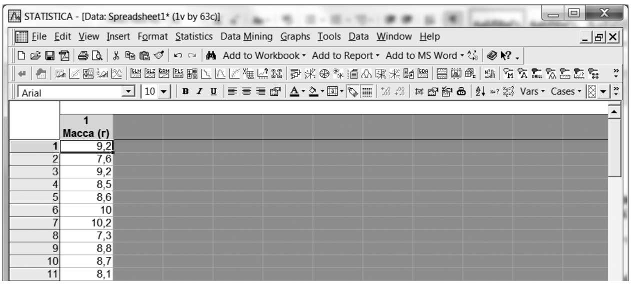


Рис. П1

Шаг 2. Запуск модуля *Descriptive statistics*. Используем команды *Statistics* (Статистические процедуры) → *Basic Statistics/Tables* (Основные статистики /Таблицы) → *Descriptive statistics* (Описательные статистики) → *OK* (рис. П2). Откроется диалоговое окно *Descriptive statistics*.

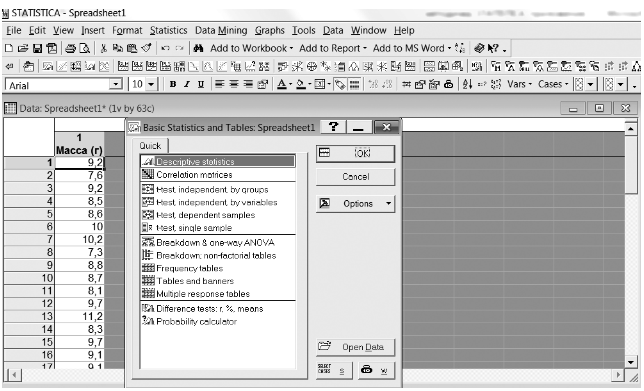


Рис. П2

Шаг 3. Настройка модуля *Descriptive statistics*. Переходим на вкладку *Normality*. Первоначально кликаем по кнопке *Variables* (Переменные). Открывается диалоговое окно выбора списка переменных для анализа. Для выбора переменной, кликаем по пункту *Масса (г)* и нажимаем кнопку *OK*. В поле *Distribution* (Распределение) установите флажок рядом с названием *Shapiro – Wilk's W-test* (*W*-тест Шапиро – Уилка). Можно указать количество интервалов (столбиков) гистограммы, на которые будет разбита выборка *Number of intervals*, которое по умолчанию задано равным десяти (рис. П3).

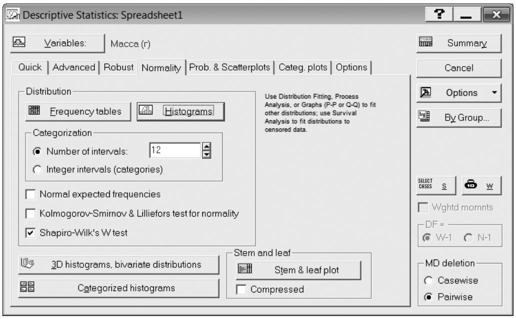


Рис. П3

После нажатия кнопки *Histograms* появляется график с гистограммой распределения значений признака и ожидаемая кривая нормального распределения (рис. П4). По оси *OX* – масса самцов взрослых бурозубок. По оси *OY* – табличные значения для нормального распределения.

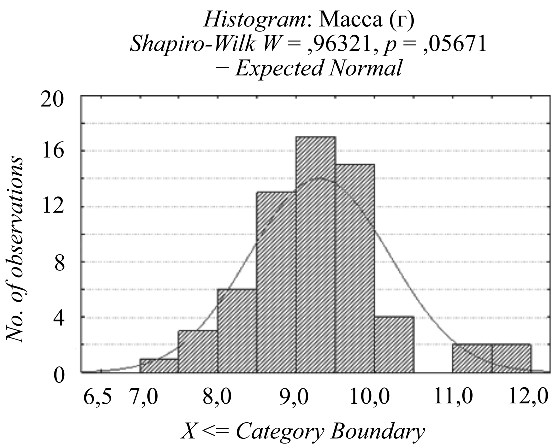


Рис. П4

Выше гистограммы размещены результаты расчета *W*-теста Шапиро – Уилка (*W* = 0,96321), проверяющего нулевую гипотезу о том, что наблюдаемое распределение признака не отличается от теоретически ожидаемого нормального распределения, и расчетная величина вероятности справедливости этой гипотезы *p* (*P*). Поскольку вероятность справедливости данной гипотезы *P* оказалась больше 0,05 (0,05671), принимаем, что она действительно верна: анализируемое распределение не отличается от нормального.

Шаг 4. *Normal probability plot* (График нормальных вероятностей) как способ проверки данных на нормальность распределения. Переходим на вкладку *Prob. & Scatterplots* (Вероятностные графики и диаграммы рассеяния) и жмем на кнопку *Normal probability plot* **(**График нормальных вероятностей) (рис. П5).

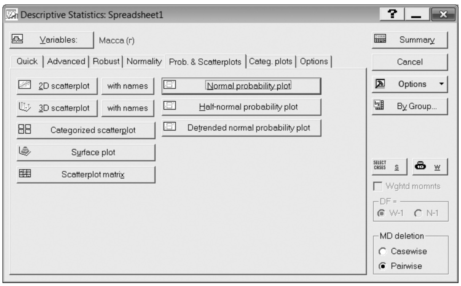


Рис. П5

Появляется график *Normal P-plot* (рис. П6). Точки полученного графика расположены примерно вдоль диагональной (теоретически ожидаемой) прямой, что указывает на нормальность распределения анализируемых данных.

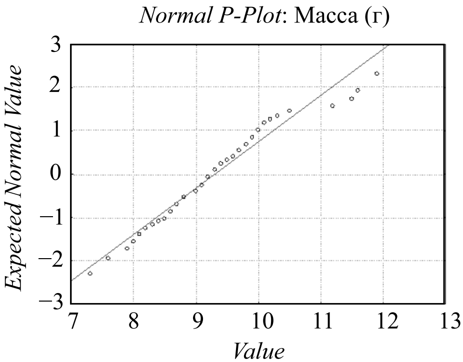


Рис. П6

Шаг 5. Проверка на нормальность распределения при помощи критерия хи-квадрат (χ^2). При анализе данных полезно определить соответствие эмпирического распределения тому или иному закону распределения случайных величин. Воспользуемся специальным модулем программы *Statistica Distribution Fitting* (Подгонка распределения) для проверки соответствия наших данных нормальному распределению. Для изучения возможностей этого модуля воспользуемся данными, отраженными в шаге 1.

Запускаем модуль *Distribution Fitting: Statistics* → *Distribution Fitting*. В открывшемся окне необходимо указать природу случайной величины – *Continuous Distributions* (Непрерывная) или *Discret Distributions* (Дискретная) и предполагаемый закон распределения, которому подчиняется изучаемый признак (рис. П7). Поскольку мы хотим проверить, подчиняются ли данные о массе бурозубок нормальному распределению, в списке непрерывных распределений *Continuous Distributions* выбираем *Normal* и жмем *OK*.

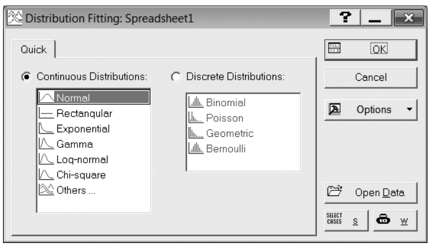


Рис. П7

В открывшемся окне *Fitting Continuous Distributions* указываем анализируемую переменную путем нажатия кнопки *Variable*. После выбора переменной слева вверху становится активным выпадающее меню *Distribution*, где можно изменить предполагаемый закон распределения (рис. П8). По умолчанию активна вкладка *Quick*. На ней есть две кнопки: *Summary: Observed and expected distribution* (Результат: Наблюдаемое и ожидаемое распределение) и *Plot of observed and expected distribution* (График наблюдаемого и ожидаемого распределения).

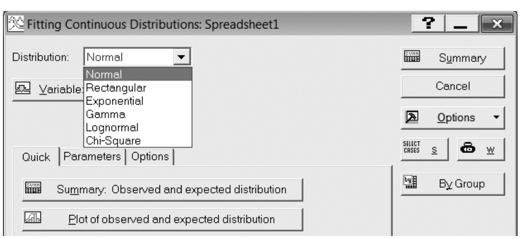


Рис. П8

При нажатии на кнопку *Summary: Observed and expected distribution* будут отображены численные характеристики изучаемой переменной в виде таблицы. При нажатии на кнопку *Plot of observed and expected distribution* будет построена кривая нормального распределения (у него те же средняя арифметическая и стандартное отклонение, что и в анализируемой совокупности) и гистограмма распределения данных по массе бурозубок (рис. П9).

Variable: Масса (г), *Distribution: Normal*
Chi-Square test = 4,31777, *df* = 3 (*adjusted*), *p* = 0,22913

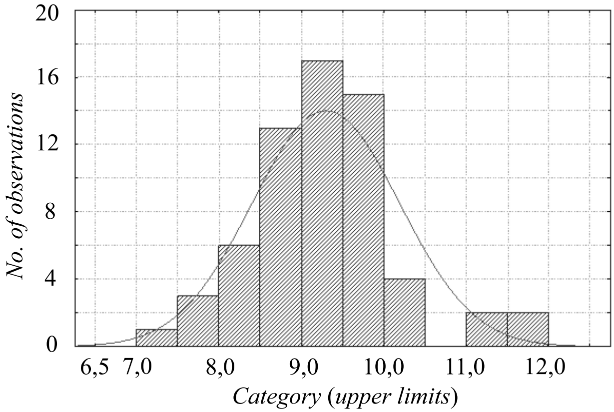


Рис. П9

В заголовке гистограммы указаны: анализируемая переменная, предполагаемый закон распределения, значение критерия χ^2 , число степеней свободы *df* и *p* – уровень значимости критерия χ^2. Вероятность справедливости нулевой гипотезы оказалась больше 0,05 (0,22913), и это значит, что она действительно верна, так как анализируемое распределение не отличается от нормального.

На вкладке *Parameters* окна *Fitting Continuous Distributions* приведено следующее (рис. П10):

1) *Number of categories* (Количество категорий) – задается по умолчанию;

2) *Lower Limit, Uper Limit* (Нижний предел, верхний пределы) – по умолчанию берутся выборочные минимальное и максимальное значения. Изменив эти параметры, можно исключить из рассмотрения значения, не попадающие в интересующий интервал;

3) *Mean* (Средняя) – определяется программой автоматически;

4) *Variance* (Дисперсия) – определяется программой автоматически. На основе определенной программой величин *Mean* и *Variance* на графике строится кривая нормального распределения. Изменив данные параме-

тры, меняется форма и расположение кривой нормального распределения на графике;

5) *Observed mean* (Выборочная средняя);

6) *Observed variance* (Выборочная дисперсия).

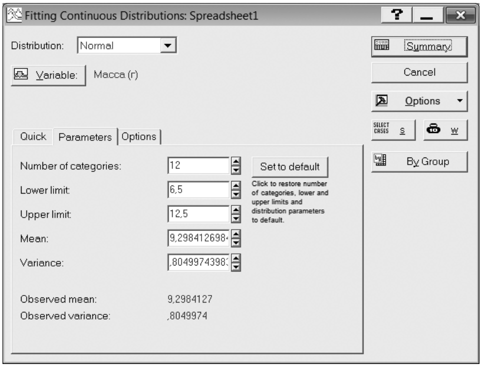


Рис. П10

На вкладке *Options* окна *Fitting Continuous Distributions* отображены четыре настройки, по которым по умолчанию анализируется переменная (рис. П11).

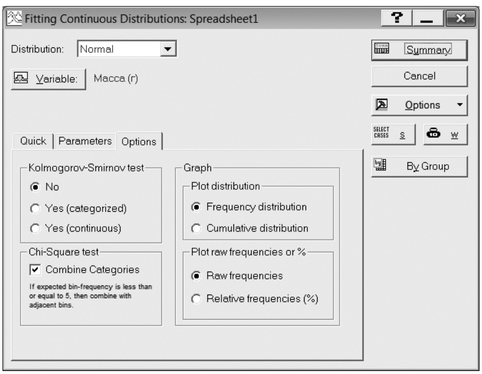


Рис. П11

1) *Kolmogorov – Smirnov test* – проверяет соответствие данных тому или иному закону распределения. По умолчанию выбрана опция *No* – тест не вычисляется, при выборе *Yes* (*categorized*) – вычисляется по группированным (интервальным) данным, при выборе *Yes* (*continuous*) – по негруппированным данным;

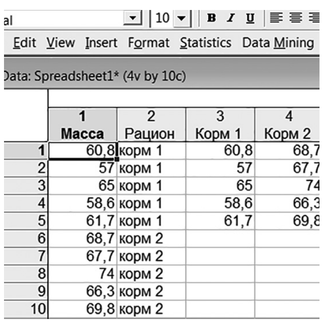
2) в поле *Plot distribution* рамки *Graph* при установке позиции *Cumulative distribution* будет построен график функции распределения. Соответственно, при выборе позиции *Relative frequencies* (*%*) в поле *Plot raw frequencies or %* будет выведено процентное соотношение относительных частот. Оставив предлагаемые по умолчанию настройки, будет выведен полученный ранее график плотности распределения со значениями относительных частот (см. рис. П9);

3) в поле *Chi-Square test* по умолчанию выставлен флажок на *Combine Categories*, поэтому при проверке соответствия выборочных данных тому или иному закону распределения, если в интервал попадает менее пяти значений, то он объединяется с соседним и т. д. пока количество значений в интервалах будет не менее пяти, иначе они не объединяются.

2. СРАВНЕНИЯ ДВУХ НЕЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. ТЕСТ СТЬЮДЕНТА ДЛЯ НЕЗАВИСИМЫХ ВЫБОРОК

В качестве примера использованы данные эксперимента о влиянии рациона питания на набор веса (см. подразд. 3.5.1). Этот тест применим, если анализируемые выборки (каждая) распределены нормально, и дисперсии в группах одинаковы. Объем выборки должен быть достаточный, чтобы проверить данные предположения ($n \geq 30$). Допустим в учебных целях, что все предположения выполняются.

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из двух столбцов и десяти строк. Первый столбец – «Масса», второй – «Рацион». Закодируйте каждую величину массы животного его соответствующим рационом (рис. П12). Обратите внимание, что для этого теста в программе *Statistica* возможен и другой вариант оформления: данные для каждой из групп («Корм 1» и «Корм 2») можно ввести в отдельные столбцы.



	1 Масса	2 Рацион	3 Корм 1	4 Корм 2
1	60,8	корм 1	60,8	68,7
2	57	корм 1	57	67,7
3	65	корм 1	65	74
4	58,6	корм 1	58,6	66,3
5	61,7	корм 1	61,7	69,8
6	68,7	корм 2		
7	67,7	корм 2		
8	74	корм 2		
9	66,3	корм 2		
10	69,8	корм 2		

Рис. П12

Шаг 2. Запуск модуля *T-Test for independent Samples by Groups.* Используйте команды *Statistics* → *Basic Statistics/Tables* или нажмите кнопку  на дополнительной панели инструментов → *t-test, independent, by groups* (если в таблице с данными есть группирующая переменная) или *t-test, independent*, *by variables* (если данные внесены в самостоятельные столбцы). Нажмите *OK* (рис. П13). В этом приложении описывается вариант теста, при котором в таблице с данными имеется группирующая переменная.

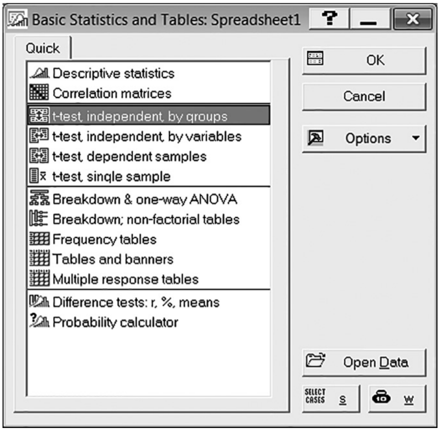


Рис. П13

Шаг 3. Выбор переменных. В открывшемся диалоговом окне модуля *T-Test for independent Samples by Groups* нажмите кнопку *Variables*. Выберите зависимую («Масса») и группирующую переменные («Рацион»). Нажмите *OK* (рис. П14).

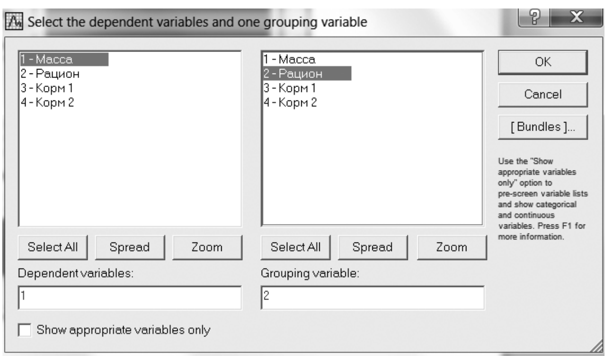


Рис. П14

После проведенных настроек модуль *T-Test for independent Samples by Groups* имеет следующий вид (рис. П15).

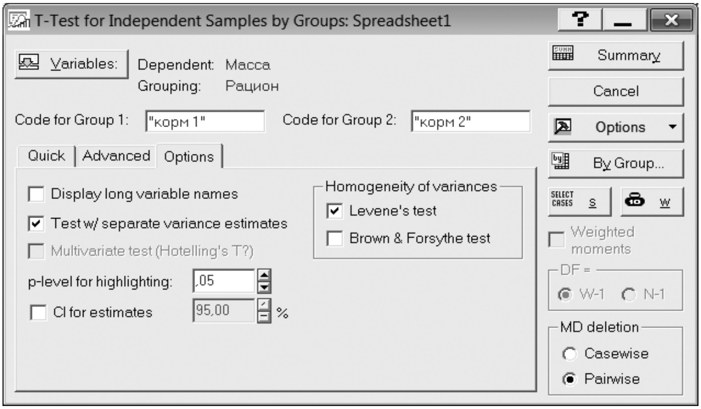


Рис. П15

Шаг 4. Настройка проверки предположения о равенстве дисперсий. Перейдите на вкладку *Options*. Отметьте галочками позиции *Levene's test* и *Test w/separate variance estimates* (рис. П16). Первая позиция нужна для проведения оценки равенства дисперсий с помощью критерия Левена, а вторая позволит вывести результаты расчета критерия Стьюдента для случая, когда дисперсии в сравниваемых группах не равны. Нажмите на кнопку *Summary* для получения результатов анализа.

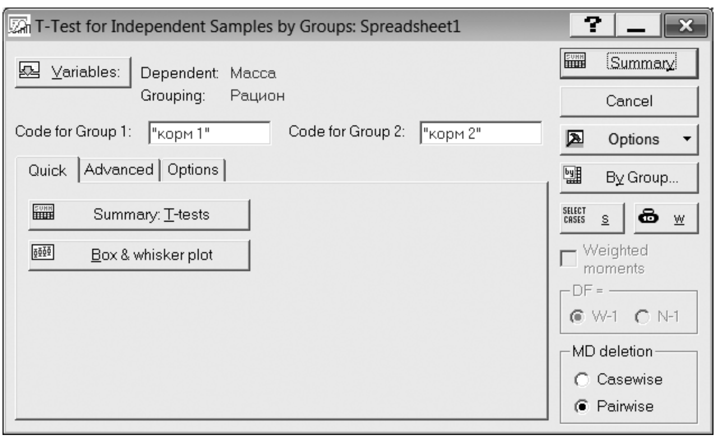


Рис. П16

Шаг 5. Интерпретация результатов анализа. Результаты анализа данных представлены на рис. П17.

T-tests; Grouping: Рацион (Spreadsheet1)
Group 1: корм 1
Group 2: корм 2

Variable	Mean корм 1	Mean корм 2	t-value	df	p	t separ. var.est.	df	p 2-sided	Valid N корм 1	Valid N корм 2	Std.Dev. корм 1	Std.Dev. корм 2	F-ratio Variances	p Variances	Levene F(1,df)	df Levene	p Levene
Масса	60,62000	69,30000	-4,58024	8	0,001801	-4,58024	7,983068	0,001811	5	5	3,064637	2,926602	1,096556	0,930967	0,025033	8	0,878206

Рис. П17

Характеристики столбцов:

• *Mean* (корм 1) – среднее значение веса в группе «корм 1»;

• *Mean* (корм 2) – среднее значение веса в группе «корм 2»;

• *t-value* – значение рассчитанного программой t-критерия Стьюдента;

• *df* – число степеней свободы;

• p – вероятность справедливости гипотезы о том, что сравниваемые средние значения не различаются. В нашем примере $P < 0,05$ ($p = 0,0018$), из чего можно сделать вывод о том, что статистически значимые различия между средними значениями веса в двух сравниваемых группах имеются;

• *Valid N* (корм 1) – объем выборки «корм 1»;

• *Valid N* (корм 2) – объем выборки «корм 2»;

• *Std. Dev.* (корм 1) – стандартное отклонение выборки «корм 1»;

• *Std. Dev.* (корм 2) – стандартное отклонение выборки «корм 2»;

• *F-ratio, Variances* – значение F-критерия Фишера, с помощью которого проверяется гипотеза о равенстве дисперсий в сравниваемых выборках;

• *p, Variances* – вероятность справедливости гипотезы о том, что дисперсии сравниваемых выборок не различаются.

Так как в шаге 4 мы ввели дополнительную настройку проверки однородности дисперсий при помощи теста Левена, результат данной проверки отражен в трех последних столбцах рис. П17. Вероятность справедливости гипотезы о том, что дисперсии сравниваемых выборок не различаются, подтверждена и тестом Левена ($p = 0,0878$). На практике рекомендуется использовать и тест Фишера, и тест Левена. Если оба теста подтверждают гипотезу об однородности дисперсий, то результаты сравнения средних нужно трактовать по графам *t-value* и p (как в нашем случае), в противном случае результаты сравнения средних трактуют по данным, приведенным в столбцах *t separ. var. est. – df – p 2-sided* (выделено рамкой на рис. П17).

Шаг 6. Визуализация описательных статистик сравниваемых групп. Модуль *T-Test for independent Samples by Groups* (см. рис. П16) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами (рис. П18) – достаточно нажать кнопку *Box & whisker plot* на вкладке *Quick*.

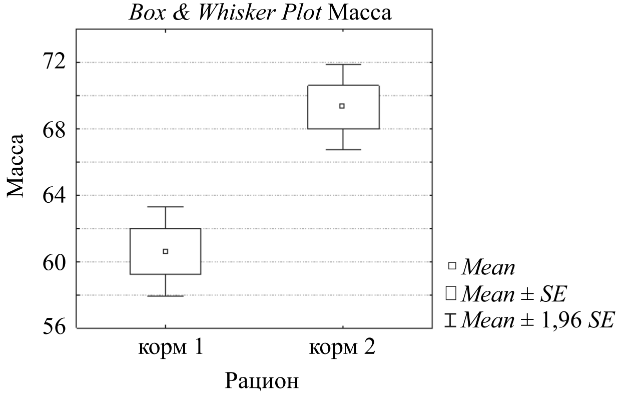


Рис. П18

3. СРАВНЕНИЯ ДВУХ ЗАВИСИМЫХ (СВЯЗАННЫХ) ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. ТЕСТ СТЬЮДЕНТА ДЛЯ ЗАВИСИМЫХ ВЫБОРОК

В качестве примера использованы данные эксперимента о возможном влиянии удобрения на урожайность культуры (см. подразд. 3.5.2). Этот тест применим, если анализируемые выборки (каждая) распределены нормально. Объем выборки должен быть достаточный, чтобы проверить данные предположения ($n \geq 30$). Допустим в учебных целях, что все предположения выполняются.

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из двух столбцов и девяти строк. Первый столбец – «Новое», второй – «Старое» (рис. П19).

File Edit View Insert Format Statistics

	1 Новое	2 Старое
1	2250	1920
2	2410	2020
3	2260	2060
4	2200	1960
5	2360	1960
6	2320	2140
7	2240	1980
8	2300	1940
9	2090	1790

Рис. П19

Шаг 2. Запуск модуля *T-Test for Dependent Samples.* Используйте команды *Statistics* → *Basic Statistics/Tables* или нажмите кнопку на дополнительной панели инструментов → *t-test, dependent samples* (рис. П20). Нажмите *OK*.

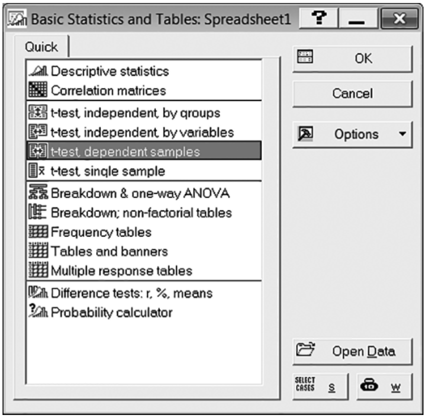


Рис. П20

Шаг 3. Выбор переменных. В открывшемся диалоговом окне модуля *T-Test for Dependent Samples* нажмите кнопку *Variables*. Выберите первую (*First variable list*) и вторую *Second variable list* (*optional*) переменные, участвующие в анализе. Нажмите *OK* (рис. П21).

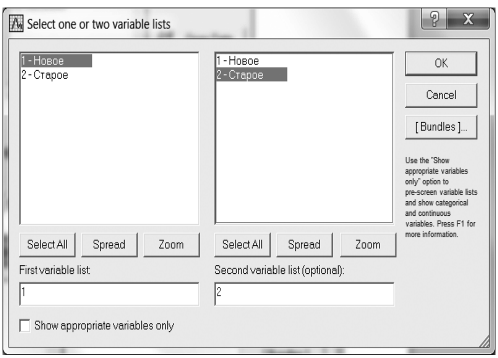


Рис. П21

После проведенных настроек модуль *T-Test for Dependent Samples* имеет вид, представленный на рис. П22. Нажмите на кнопку *Summary: T-tests*.

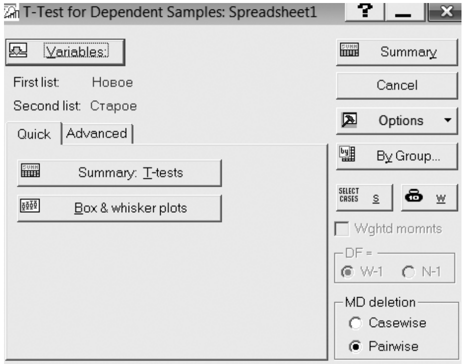


Рис. П22

Шаг 3. Интерпретация результатов анализа. Результаты анализа данных представлены на рис. П23.

T-test for Dependent Samples (Spreadsheet1)
Marked differences are significant at p < ,05000

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Новое	2270,000	93,40771						
Старое	1974,444	97,09674	9	295,5556	80,63980	10,99540	8	0,000004

Рис. П23

В таблице с результатами расчетов приведено следующее (см. рис. П23):

• *Mean* – средние значения урожайности для каждой из сравниваемых групп;

• *Std. Dv.* – стандартные отклонения для каждой из групп;

• *N* – число наблюдений;

• *Diff.* – разница между средними;

• *Std. Dv. Diff.* – стандартное отклонение для средней разницы;

• *t* – значение *t*-критерия;

• *df* – число степеней свободы;

• *p* – вероятность справедливости гипотезы о том, что средние величины урожайности в сравниваемых группах не различаются. Как видно из таблицы на рис. П23, $P \ll 0{,}001$. Это говорит о том, что средняя урожайность при использовании нового и старого удобрений значительно различается.

Шаг 4. Визуализация описательных статистик сравниваемых групп. Модуль *T-Test for Dependent Samples* (см. рис. П22) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами (рис. П24) – достаточно нажать кнопку *Box & whisker plots*.

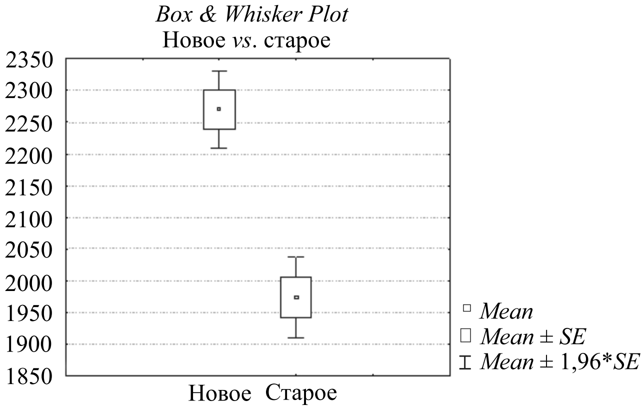


Рис. П24

4. СРАВНЕНИЯ ВЫБОРОЧНОЙ СРЕДНЕЙ С КОНСТАНТОЙ С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. ТЕСТ СТЬЮДЕНТА ДЛЯ СРАВНЕНИЯ ВЫБОРОЧНОЙ СРЕДНЕЙ С КОНСТАНТОЙ

Тест Стьюдента применим, если анализируемая выборка распределена нормально. Ее объем должен быть достаточный, чтобы проверить это предположение ($n \geq 30$).

Задача. Согласно государственному стандарту, предельно допустимая концентрация (ПДК) некоторого загрязнителя составляет 100 единиц. При замерах содержания этого вещества в 10 пробах городской почвы получены следующие данные: 108, 99, 112, 100, 101, 98, 95, 105, 90, 102 единиц. Превышает ли среднее содержание поллютанта в исследованных образцах почвы ПДК?

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из одного столбца и десяти строк (рис. П25).

	1 Загрязнитель
1	108
2	99
3	112
4	100
5	101
6	98
7	95
8	105
9	90
10	102

Рис. П25

Шаг 2. Запуск модуля T-Test for Single Means. Используйте команды *Statistics → Basic Statistics/Tables* или кликните на кнопку на дополнительной панели инструментов → *t-test, single sample* (рис. П26). Нажмите *OK*.

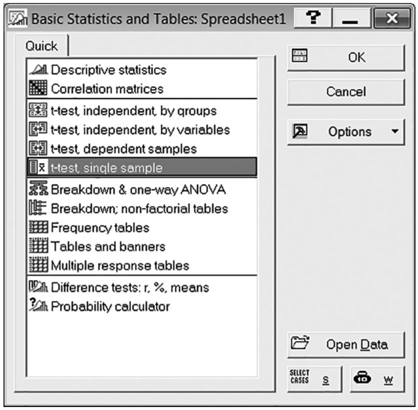


Рис. П26

Шаг 3. Выбор переменных. В открывшемся диалоговом окне модуля *T-Test for Single Means* нажмите кнопку *Variables*. Выберите *Загрязнитель* и нажмите *OK* (рис. П27).

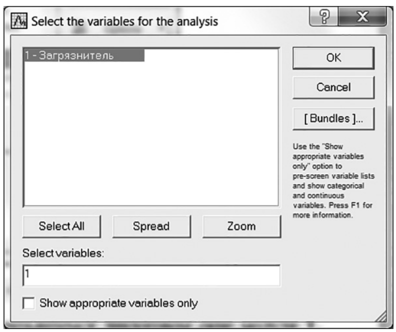


Рис. П27

Таких переменных можно ввести несколько (например, если бы аналогичные отборы проб почвы производились в разные месяцы, то в анализ можно было бы включить и эти данные). При таком варианте программа сравнит все указанные переменные с одним контрольным значением. Это значение задается в поле *Reference values* (Контрольные значения). В нашей задаче данная величина ПДК равна 100, ее мы вносим напротив опции *Test all means against*

(Сравнить все средние с). При необходимости включенные в анализ переменные можно сопоставить с разными контрольными значениями. Для этого нужно активировать опцию *Test means against different user-defined constants* (Сравнить средние с разными константами, заданными пользователем).

После проведенных настроек вкладка модуль *T-Test for Single Means* имеет вид как на рис. П28. Нажмите на кнопку *Summary: T-tests*.

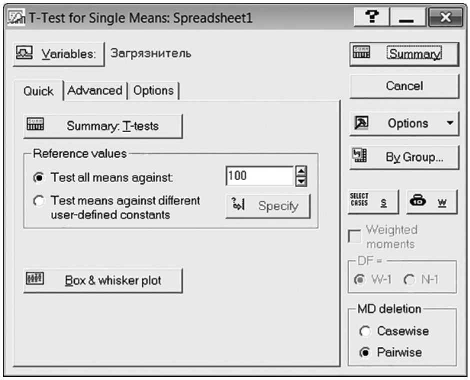


Рис. П28

Шаг 4. Интерпретация результатов анализа. Принятие гипотезы. Результаты анализа данных представлены на рис. П29.

Variable	Test of means against reference constant (value) (Spreadsheet1)							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
Загрязнитель	101,0000	6,306963	10	1,994437	100,0000	0,501395	9	0,628128

Рис. П29

В таблице с результатами расчетов приведено следующее:

- *Mean* – среднее значение, рассчитанное на основе выборочных данных;
- *Std Dv.* – стандартное отклонение;
- *N* – объем выборки;
- *Std. Err.* – стандартная ошибка;
- *Reference Constant* – контрольное значение;
- *t-value* – значение *t*-критерия;
- *df* – число степеней свободы;
- *p* – вероятность справедливости гипотезы о том, что выборочная средняя не отличается от контрольной величины. В нашем случае $p = 0{,}62$ ($P > 0{,}05$), поэтому, несмотря на некоторое превышение средней концентра-

ции поллютанта в пробах, оно является незначительным и статистически не отличается от нормы.

Шаг 5. Визуализация описательных статистик сравниваемых групп. Модуль *T-Test for Single Means* (см. рис. П28) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки вариации признака. Нажмите кнопку *Box & whisker plot*. В окне *Box-Whisker Type* (рис. П30) подсветите один из предложенных вариантов построения «Коробочки с усами», нажмите *OK* – график готов (рис. П31).

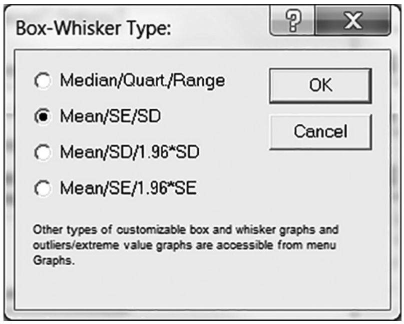


Рис. П30

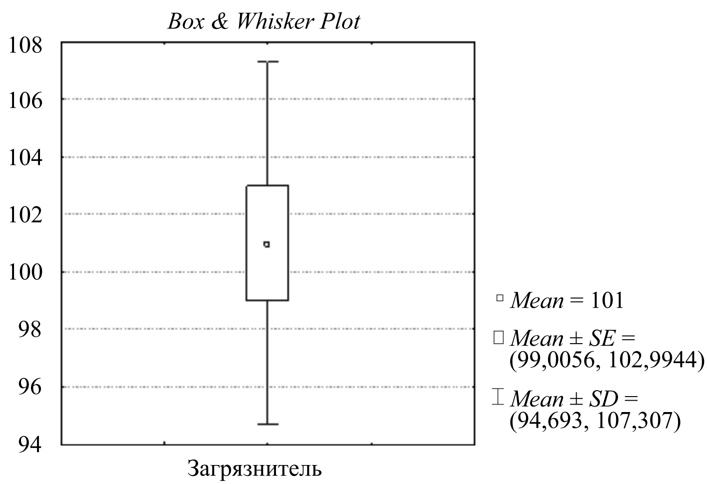


Рис. П31

5. СРАВНЕНИЯ ДВУХ И БОЛЕЕ НЕЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ *ONE-WAY ANOVA*

В качестве примера использованы данные эксперимента о влиянии рациона питания на набор веса (см. подразд. 3.5.3).

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из двух столбцов и двадцати строк. Первый столбец – «Масса», второй – «Рацион». Закодируйте каждую величину массы животного его соответствующим рационом (рис. П32).

	1 Масса	2 Рацион
1	60,8	корм 1
2	57	корм 1
3	65	корм 1
4	58,6	корм 1
5	61,7	корм 1
6	68,7	корм 2
7	67,7	корм 2
8	74	корм 2
9	66,3	корм 2
10	69,8	корм 2
11	102,6	корм 3
12	102,1	корм 3
13	100,2	корм 3
14	96,5	корм 3
15	96,5	корм 3
16	87,9	корм 4
17	84,2	корм 4
18	83,1	корм 4
19	85,7	корм 4
20	90,3	корм 4

Рис. П32

Шаг 2. Запуск модуля *One-way ANOVA*. Используйте команды *Statistics → ANOVA* или нажмите кнопку на дополнительной панели инструментов. Откроется диалоговое окно модуля *General ANOVA/MANOVA*. Оставьте предлагаемый по умолчанию вариант диалогового окна: в списке *Type of analysis* (Вид анализа) – *One-way ANOVA* (Однофакторный дисперсионный анализ), в списке *Specification method* (Задание метода) – *Quick specs dialog* (Диалог быстрых спецификаций). Нажмите *OK* (рис. П33).

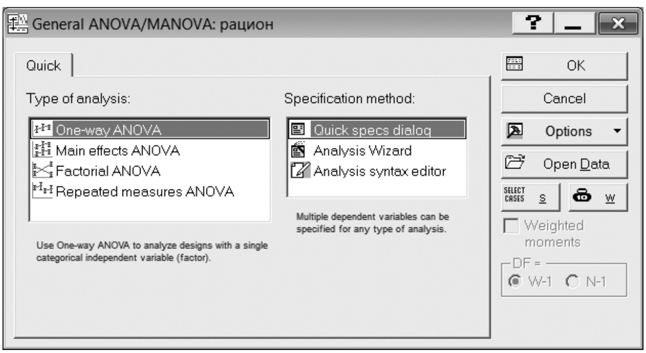


Рис. П33

Шаг 3. Выбор переменных. В открывшемся диалоговом окне *ANOVA/MANOVA One-way ANOVA* на вкладке *Quick* нажмите кнопку *Variables*. Выберите зависимую («Масса») и группирующую переменные («Рацион»). Нажмите *OK* (рис. П34).

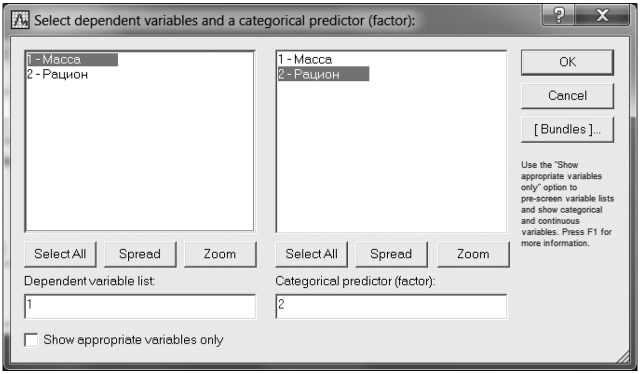


Рис. П34

Нажмите кнопку *Factor codes* (коды факторов), затем кнопку *All* – указали программе, что в анализе будут участвовать все экспериментальные группы (рис. П35).

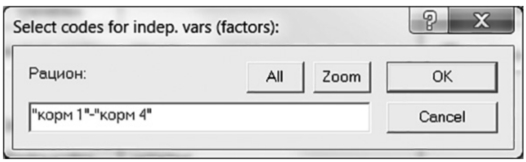


Рис. П35

После проведенных настроек вкладка *Quick* диалогового окна *ANOVA/MANOVA One-way ANOVA* имеет вид как на рис. П36. Нажмите *OK*.

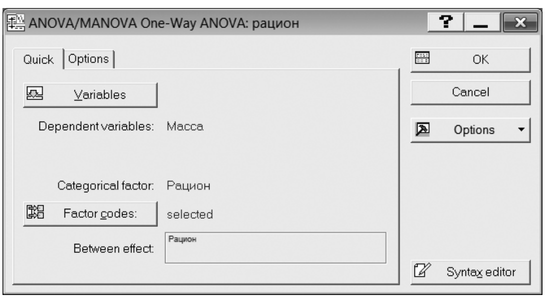


Рис. П36

Шаг 4. Проверка предположений. Рассмотрите открывшееся диалоговое окно *ANOVA Results* (Результаты анализа). Автоматически программа откроет его на вкладке *Quick* (рис. П37). Для проверки условий, при которых допустимо применение дисперсионного анализа, нажмите на кнопку *More results* (Дополнительные результаты), расположенную в нижней части окна *ANOVA Results*.

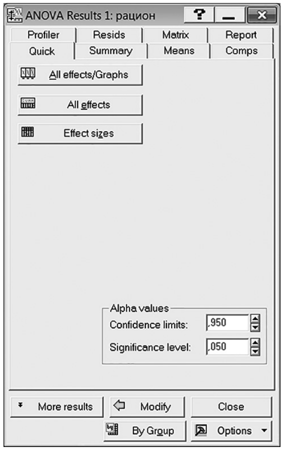


Рис. П37

Перейдите на вкладку *Assumptions* (Предположения) (рис. П38). На вкладке представлены различные критерии проверки гипотезы однородности дисперсий (критерий Кохрана, Хартли, Бартлетта, критерий Левена) и графические средства проверки соответствия закона распределения переменной нормальному закону (гистограммы, диаграммы рассеяния, нормальные вероятностные графики).

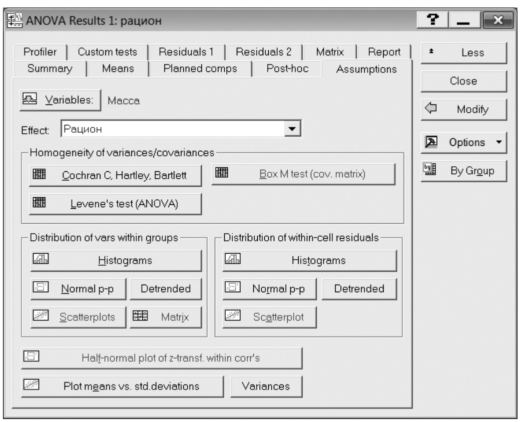


Рис. П38

Для проверки однородности групповых дисперсий в поле *Homogeneity of variances/covariances* нажмите на кнопку *Levene's test* (*ANOVA*) (тест Левена). Результат теста (рис. П39) указывает на отсутствие статистически значимой разницы между дисперсиями всех групп (P = 0,982825; при P > 0.05 нулевая гипотеза о равенстве дисперсий подтверждается).

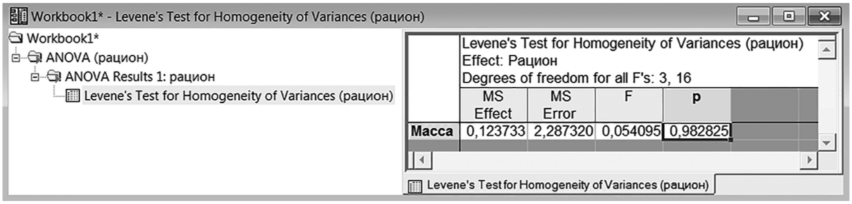


Рис. П39

Для проверки нормальности распределения нажмите на кнопку *Normal p-p* в поле *Distribution of vars within groups* (Распределение переменных внутри групп) (см. рис. П38). Если в появившемся окне *Select groups for* (рис. П40) выбрать *All Groups* и нажать *OK*, то программа построит график «Вероятностная бумага» для всех групп. Нужно проанализировать распределение внутри каждой экспериментальной группы, поэтому необходимо выбрать все четыре группы и нажать *OK*.

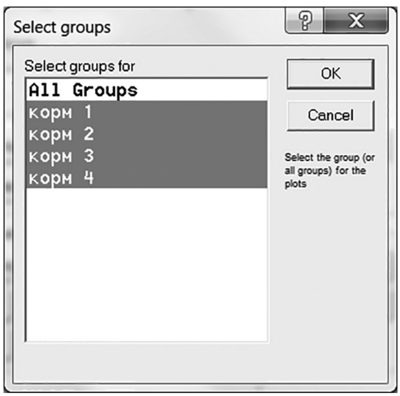


Рис. П40

Ниже приведен пример графика, построенного на графике «Вероятностная бумага» для группы «корм 1» (рис. П41). Точки наблюдения примерно укладываются вдоль теоретически ожидаемой прямой, что указывает на нормальность распределения анализируемых данных. Аналогичная ситуация характерна и для остальных трех групп. При большом числе наблюдений можно оценить характер распределений анализируемых групп с помощью гистограмм (кнопка *Histograms* в поле *Distribution of vars within groups*) (см. рис. П38).

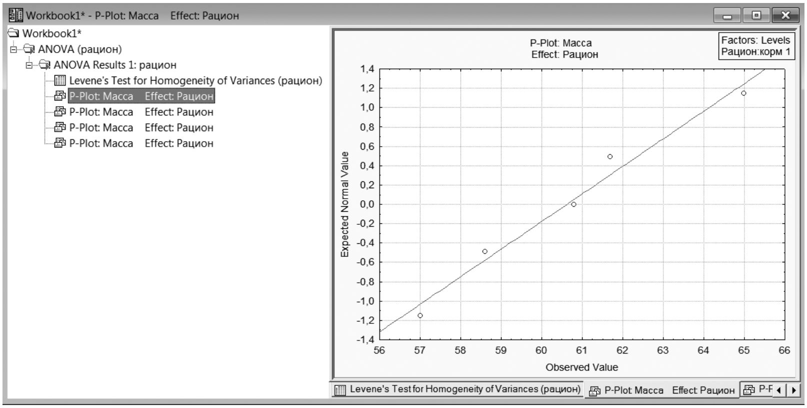


Рис. П41

Еще одним дополнительным условием применимости дисперсионного анализа является отсутствие корреляции между средними и стандартными отклонениями. Нажмите кнопку *Plot means vs. std. deviations*, расположенную в нижней части вкладки *Assumptions* (см. рис. П38). Из диаграммы рассеяния видно, что средние и стандартные отклонения не коррелируют (рис. П42).

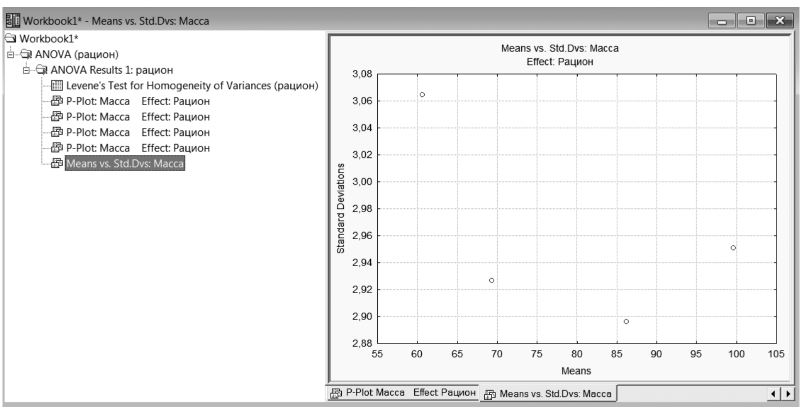


Рис. П42

Таким образом, основные условия применимости дисперсионного анализа выполнены.

Шаг 5. Выполнение дисперсионного анализа. Перейдите на вкладку *Summary* (см. рис. П38) и нажмите кнопку *Test all effects* (Проверить все эффекты). Появится рабочая книга с результатами дисперсионного анализа (рис. П43).

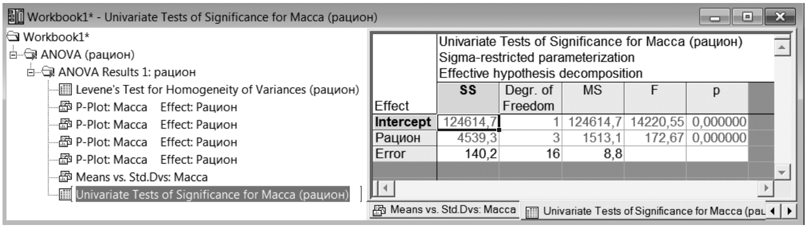


Рис. П43

Расчетная величина *F*-критерия Фишера и достигнутый в ходе анализа уровень значимости *P* находятся на пересечении второй строки и предпоследнего и последнего столбца таблицы. При $P \leq 0{,}05$ нулевую гипотезу об отсутствии разницы между средними значениями сравниваемых групп следует отвергнуть. Следовательно, в нашем примере $P \ll 0{,}001$ рацион питания влияет на набор веса ($F_{эмп} = 172{,}7$; $P \ll 0{,}001$).

Шаг 6. Множественные сравнения. Выясним, между какими группами есть различия. Воспользуемся методами множественных сравнений вкладки *Post-hoc* (рис. П44).

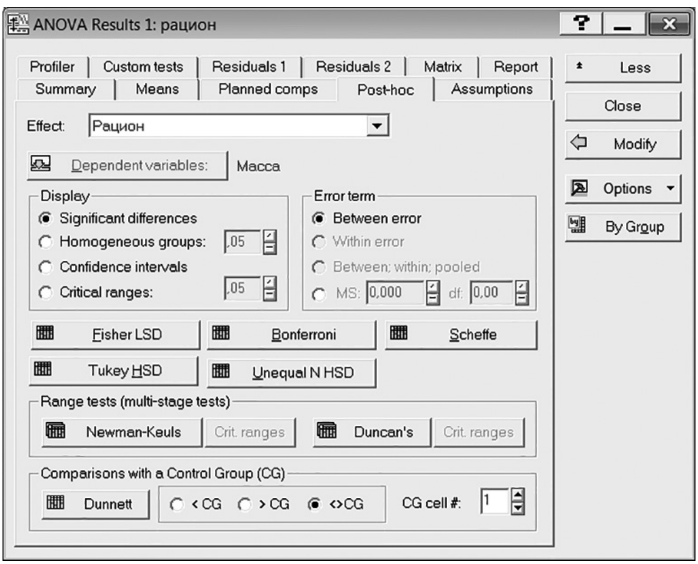


Рис. П44

Здесь предлагается несколько вариантов тестов, различающихся по мощности (*Fisher LSD*, *Bonferroni*, *Scheffe*, *Tukey HSD*, *Newman – Keuls*, *Duncan's*, *Dunnet*). Наиболее часто используемыми являются тесты Тьюки (*Tukey HSD*) и Ньюмана – Кейлса (*Newman – Keuls*). Нажмем на кнопку *Tukey HSD* (тест Тьюки для выборок с одинаковыми объемами; при сравнении групп разных объемов – использовать кнопку *Unequal N HSD*).

Появилась рабочая книга с таблицей вероятностей *P* (рис. П45). Если вероятность, стоящая на пересечении строки и столбца с соответствующими номерами групп больше 0,05, то гипотезу о равенстве средних данных групп принимаем, в противном случае отвергаем. Из таблицы видно, что между всеми сравниваемыми группами есть отличия.

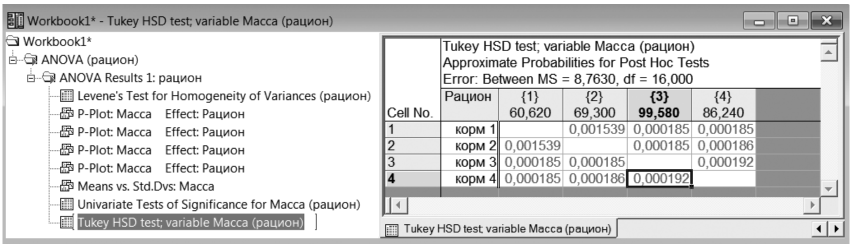


Рис. П45

Шаг 7. Графическое представление средних. Чтобы посмотреть средние и их графики, перейдите на вкладку *Means* (рис. П46). Нажмите кнопку *Plot* рядом с кнопкой *Observed*, *unweighted*.

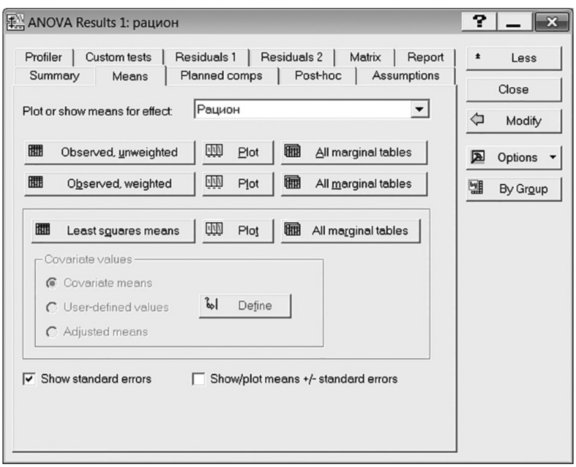


Рис. П46

Видно, что средние величины всех четырех групп отличаются друг от друга (рис. П47).

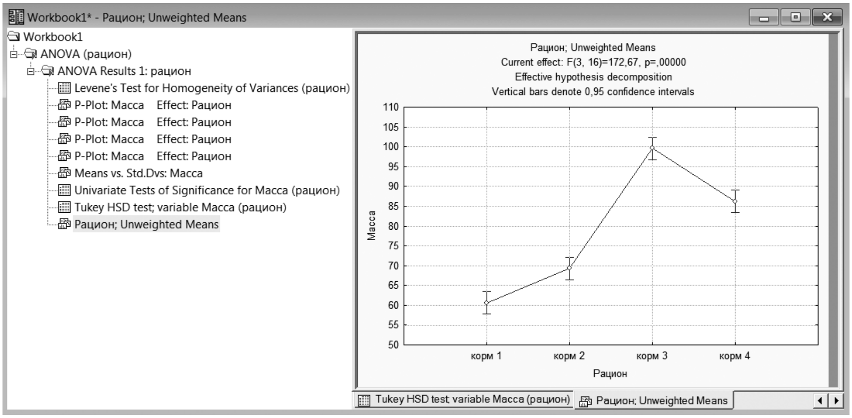


Рис. П47

6. МНОГОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ *FACTORIAL ANOVA* ДЛЯ СРАВНЕНИЯ НЕЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*

В качестве примера использованы данные эксперимента об урожайности пшеницы четырех разных сортов, полученной при использовании пяти типов удобрений. (см. подразд. 3.5.5).

Перед проведением дисперсионного анализа формулируем три нулевые гипотезы (см. подразд. 3.5.5, шаг 2).

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из 3 столбцов и 60 строк. Первый столбец – «Удобрение», второй – «Сорт», третий – «Урожайность». Закодируйте каждую величину урожайности кодами соответствующих факторов (рис. П48).

ta: DA_тест (3v by 60c)

	1 Удобрение	2 Сорт	3 Урожайность
1	Удобрение A	Сорт 1	19
2	Удобрение A	Сорт 1	20
3	Удобрение A	Сорт 1	21
4	Удобрение A	Сорт 2	25
5	Удобрение A	Сорт 2	24
6	Удобрение A	Сорт 2	26
7	Удобрение A	Сорт 3	17
8	Удобрение A	Сорт 3	18
9	Удобрение A	Сорт 3	16
10	Удобрение A	Сорт 4	21
11	Удобрение A	Сорт 4	19
12	Удобрение A	Сорт 4	20
13	Удобрение B	Сорт 1	22
14	Удобрение B	Сорт 1	23
15	Удобрение B	Сорт 1	21
16	Удобрение B	Сорт 2	19
17	Удобрение B	Сорт 2	22
18	Удобрение B	Сорт 2	18
19	Удобрение B	Сорт 3	19
20	Удобрение B	Сорт 3	18

Рис. П48

Шаг 2. Запуск модуля *Factorial ANOVA*. Используйте команды *Statistics* → *ANOVA* или нажмите кнопку на дополнительной панели инструментов. Откроется диалоговое окно модуля *General ANOVA/MANOVA*. Из списка *Type of* *analysis* (Вид анализа) выберите *Factorial ANOVA* (Многофакторный дисперсионный анализ), а в списке *Specification method* (Задание метода) – *Quick specs dialog* (Диалог быстрых спецификаций). Нажмите *OK* (рис. П49).

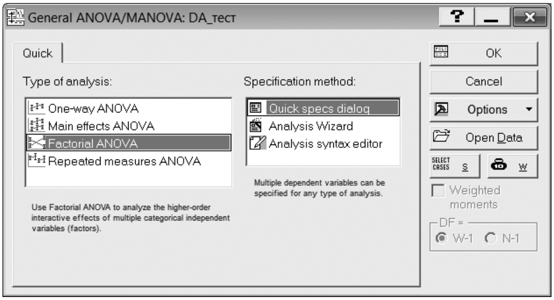


Рис. П49

Шаг 3. Выбор переменных. В открывшемся диалоговом окне *ANOVA/ MANOVA Factorial ANOVA* на вкладке *Quick* нажмите кнопку *Variables*. Выберите зависимую («Урожайность») и группирующие переменные («Удобрение» и «Сорт»). Нажмите *OK* (рис. П50).

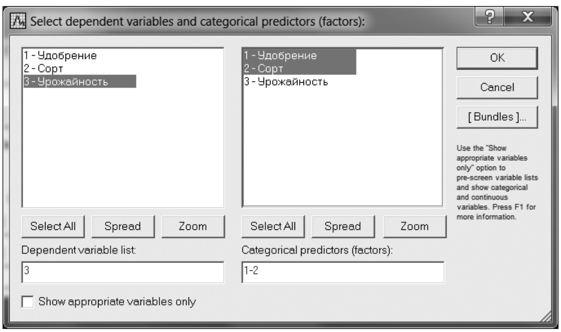


Рис. П50

Нажмите кнопку *Factor codes* (коды факторов), затем кнопку *All* – указали программе, что в анализе будут участвовать все экспериментальные группы (рис. П51).

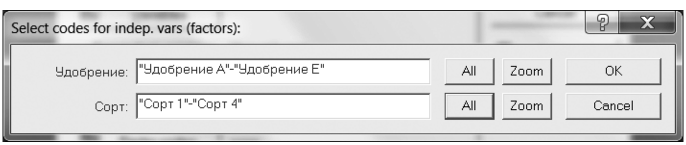


Рис. П51

После проведенных настроек вкладка *Quick* диалогового окна *ANOVA/ MANOVA One-way ANOVA* имеет вид как на рис. П52. Нажмите *OK*.

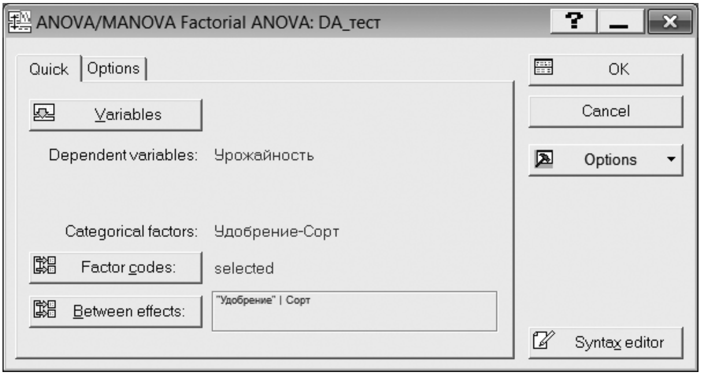


Рис. П52

Шаг 4. Проверка предположений. Проводится так же, как и при однофакторном дисперсионном анализе на вкладке *Assumptions* (см. прил. 5, шаг 4) за исключением следующего момента. При проверке предположений о возможности проведения *Factorial ANOVA*, ее необходимо дополнить проверкой всех возможных группировок данных, получаемой исходя из предполагаемого воздействия одного из факторов: «Удобрение», «Сорт» или их взаимного действия (см. по стрелке в поле *Effect* (эффект) (рис. П53).

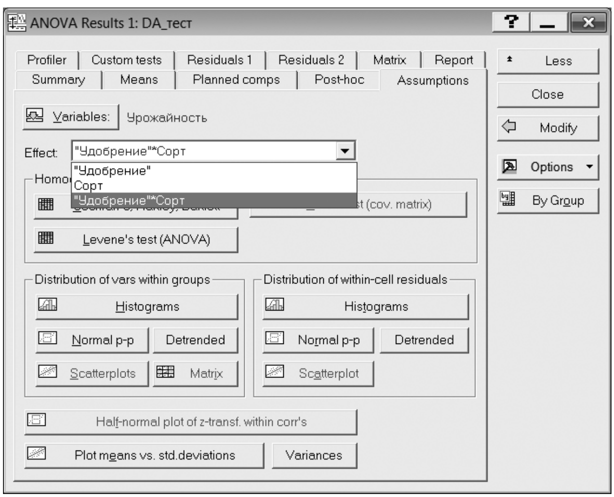


Рис. П53

Шаг 5. Выполнение дисперсионного анализа. Перейдите на вкладку *Summary* (рис. П54) и нажмите кнопку *Test all effects* (Проверить все эффекты).

Effect	Univariate Tests of Significance for Урожайность Sigma-restricted parameterization Effective hypothesis decomposition SS	Degr. of Freedom	MS	F	p
Intercept	27477,60	1	27477,60	14720,14	0,000000
Удобрение	98,57	4	24,64	13,20	0,000001
Сорт	71,07	3	23,69	12,69	0,000006
Удобрение*Сорт	136,10	12	11,34	6,08	0,000007
Error	74,67	40	1,87		

Рис. П54

Появится таблица с результатами дисперсионного анализа. Самое главное в ней – строки со второй по четвертую, расчетные величины F-критерия Фишера и достигнутый в ходе анализа уровень значимости P для проверки выдвинутых гипотез (находятся в предпоследнем и последнем столбцах соответственно). Исходя из полученных результатов, делаем заключение о влиянии на урожайность пшеницы факторов «Удобрение», «Сорт» и их взаимного действия.

Шаг 6. Множественные сравнения проводится так же, как и при однофакторном дисперсионном анализе на вкладке *Post-hoc* (см. прил. 5, шаг 6). При проверке каждой из трех нулевых гипотез, исходя из фактора воздействия, для анализа задавалась своя компоновка групп, поэтому возможны разные варианты проведения *Post-hoc* анализа исходя из выбранного фактора в поле *Effect*. В качестве примера ниже представлен вариант *Post-hoc* анализа для фактора «Сорт» (рис. П55).

Cell No.	Tukey HSD test; variable Урожайность (DA_тест) Approximate Probabilities for Post Hoc Tests Error: Between MS = 1,8667, df = 40,000 Сорт	{1} 21,533	{2} 22,933	{3} 19,867	{4} 21,267
1	Сорт 1		0,037219	0,009557	0,950191
2	Сорт 2	0,037219		0,000164	0,009557
3	Сорт 3	0,009557	0,000164		0,037219
4	Сорт 4	0,950191	0,009557	0,037219	

Рис. П55

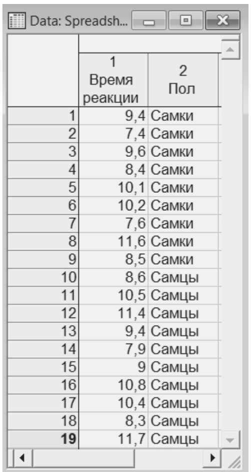
Исходя из теста Тьюки заключаем о различиях в урожайности пшеницы всех сортов, за исключением сортов 1 и 4.

В рассматриваемой задаче возможны три варианта *Post-hoc* анализа, поскольку все выдвинутые нулевые гипотезы были отклонены (см. рис. П54).

7. СРАВНЕНИЯ ДВУХ НЕЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. КРИТЕРИЙ *MANN – WHITNEY U-TEST*

В качестве примера использованы данные эксперимента о времени реакции крыс на инъекцию экспериментального лекарства. (см. подразд. 4.1.1).

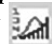
Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из 2 столбцов и 19 строк. Первый столбец – «Время реакции», второй – «Пол». Закодируйте каждую величину данных соответствующими кодами (рис. П56).



Data: Spreadsh...

	1 Время реакции	2 Пол
1	9,4	Самки
2	7,4	Самки
3	9,6	Самки
4	8,4	Самки
5	10,1	Самки
6	10,2	Самки
7	7,6	Самки
8	11,6	Самки
9	8,5	Самки
10	8,6	Самцы
11	10,5	Самцы
12	11,4	Самцы
13	9,4	Самцы
14	7,9	Самцы
15	9	Самцы
16	10,8	Самцы
17	10,4	Самцы
18	8,3	Самцы
19	11,7	Самцы

Рис. П56

Шаг 2. Запуск модуля *Comparing Two Groups*. Используйте команды *Statistics → Nonparametrics* или нажмите кнопку  на дополнительной панели инструментов. В диалоговом окне *Nonparametric Statistics* выберите процедуру *Comparing two independent samples* (*groups*). Нажмите *OK*.

Шаг 3. Выбор переменных. В открывшемся модуле *Comparing Two Groups* нажмите кнопку *Variables*. Выберите зависимую («Время реакции») и группирующую переменные («Пол»). Нажмите *OK*. После данных действий нажмите на кнопку *Mann – Whitney U test* либо *M-W U test* (рис. П57).

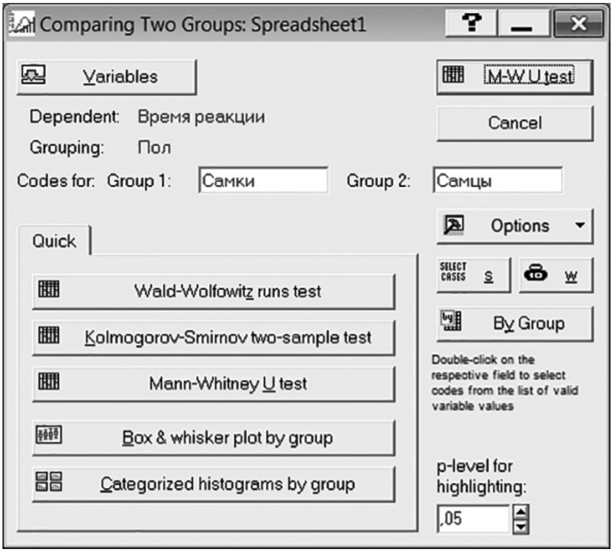
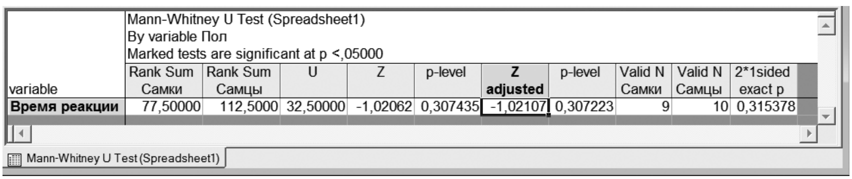


Рис. П57

Шаг 4. Интерпретация полученных результатов (рис. П58).



Mann-Whitney U Test (Spreadsheet1)
By variable Пол
Marked tests are significant at p <,05000

variable	Rank Sum Самки	Rank Sum Самцы	U	Z	p-level	Z adjusted	p-level	Valid N Самки	Valid N Самцы	2*1sided exact p
Время реакции	77,50000	112,5000	32,50000	-1,02062	0,307435	-1,02107	0,307223	9	10	0,315378

Mann-Whitney U Test (Spreadsheet1)

Рис. П58

В таблице с результатами расчетов приведено следующее:

• *Rank Sum* **–** сумма рангов выборок «Самки» и «Самцы»;

• *U* – расчетное значение критерия Манна – Уитни

• *Z* – нормальная аппроксимация статистики Манна – Уитни для больших выборок;

• *p-level* **–** вероятность справедливости нулевой гипотезы;

• *Z-adjusted* – скорректированная нормальная аппроксимация статистики Манна – Уитни;

• *p-level* – скорректированная вероятность справедливости гипотезы;

• *Valid N* – объем выборки;

• *2*1 sided exact p* – здесь вероятность *p* равна 1 минус кумулятивная односторонняя вероятность соответствующей статистики Манна – Уитни.

Для принятия или опровержения нулевой гипотезы интерпретируют величину *p-level*. При большом числе наблюдений в выборках (20 и более) ее значение смотрят в пятом столбце таблицы (вслед за *Z*), иначе – в седьмом (вслед за *Z-adjusted*). Так как $P > 0{,}05$, можно утверждать, что нулевая гипотеза верна и время реакции у самцов и самок на инъекцию экспериментального лекарства статистически одинакова.

Шаг 5. Визуализация описательных статистик сравниваемых групп. Модуль *Comparing Two Groups* (см. рис. П57) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами (рис. П59). Нажмите кнопку *Box & whisker plots by group* и выберите столбец с данными «Время реакции», нажмите *OK*.

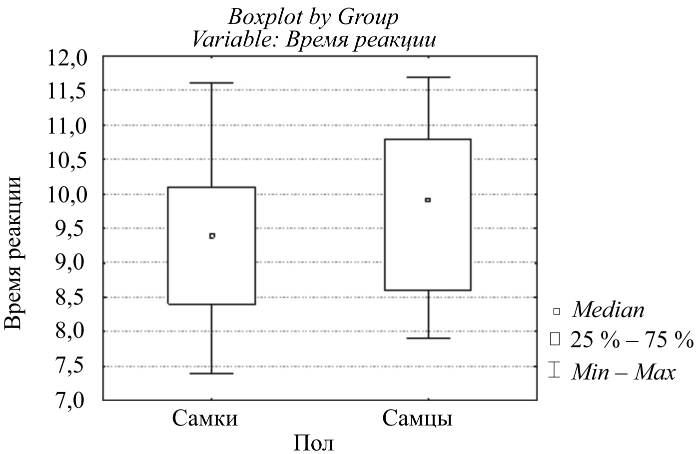
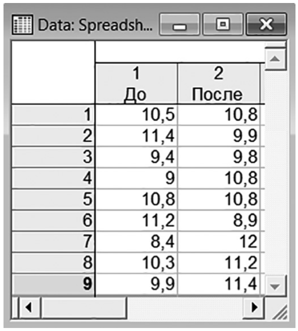


Рис. П59

8. СРАВНЕНИЯ ДВУХ ЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. КРИТЕРИЙ *WILCOXON MATCHED PAIRS TEST*

В качестве примера использованы данные эксперимента о сравнении времени реакции самок крыс до и после инъекции экспериментального лекарства (см. подразд. 4.1.2).

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из двух столбцов и девяти строк. Первый столбец – «До», второй – «После» (рис. П60).



	1 До	2 После
1	10,5	10,8
2	11,4	9,9
3	9,4	9,8
4	9	10,8
5	10,8	10,8
6	11,2	8,9
7	8,4	12
8	10,3	11,2
9	9,9	11,4

Рис. П60

Шаг 2. Запуск модуля *Comparing two variables*. Используйте команды *Statistics → Nonparametrics* или нажмите кнопку на дополнительной панели инструментов. В диалоговом окне *Nonparametric Statistics* выберите процедуру *Comparing two dependent samples* (*variables*). Нажмите *OK*.

Шаг 3. Выбор переменных. В открывшемся модуле *Comparing two variables* нажмите кнопку *Variables*. Выберите переменные. Нажмите *OK*. После нажмите на кнопку *Wilcoxon matched pairs test* (рис. П61).

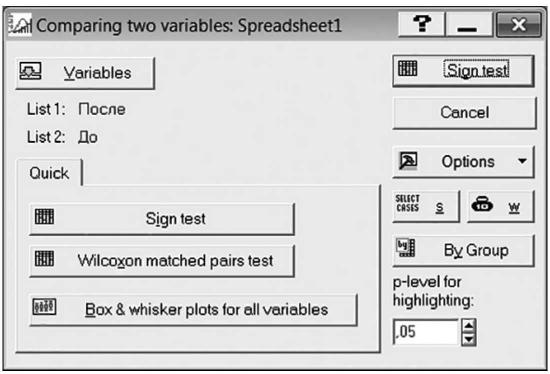


Рис. П61

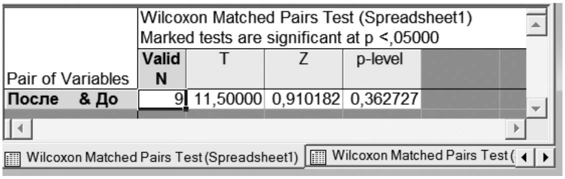
Шаг 4. Интерпретация полученных результатов. В таблице с результатами расчетов приведено следующее (рис. П62):

• *Valid N* – число сравниваемых пар;

• *T* – меньшая из двух сумм рангов модулей разности;

• *Z* – нормальная аппроксимация статистики Вилкоксона для больших выборок;

• *p-level* – вероятность справедливости нулевой гипотезы.



Pair of Variables	Valid N	T	Z	p-level
После & До	9	11,50000	0,910182	0,362727

Рис. П62

Для принятия или опровержения нулевой гипотезы интерпретируют величину *p-level*. Так как $P > 0{,}05$ ($P = 0{,}36$), можно утверждать, что нулевая гипотеза верна и время реакции у самок крыс на раздражитель одинаково до и после применения препарата.

Шаг 5. Визуализация описательных статистик сравниваемых групп. Модуль *Comparing two variables* (см. рис. П61) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами. Нажмите кнопку *Box & whisker plots for all variables* и выберите столбцы с данными «До» и «После», нажмите *OK*. В окне *Box-Whisker Type* (рис. П63) подсветите один из предложенных вариантов построения графика «Коробочка с усами», нажмите *OK*. График получен (рис. П64).

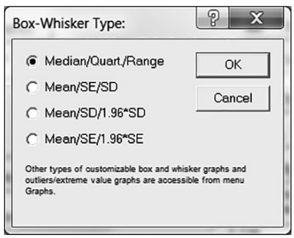


Рис. П63

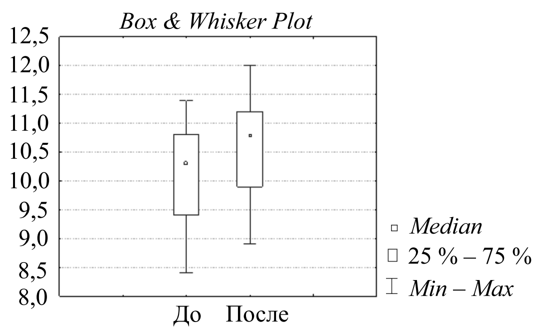
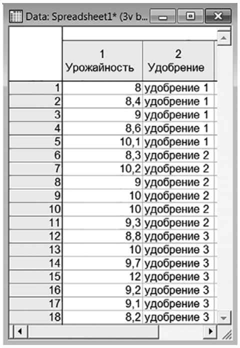


Рис. П64

9. СРАВНЕНИЯ ДВУХ И БОЛЕЕ НЕЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. КРИТЕРИЙ *KRUSKAL – WALLIS H-TEST*

В качестве примера использованы данные эксперимента о сравнении урожайности озимой ржи в зависимости от типа используемого удобрения. (см. подразд. 4.2.1).

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из 2 столбцов и 18 строк. Первый столбец – «Урожайность», второй – «Удобрение» (рис. П65).



Data: Spreadsheet1* (3v b...

	1 Урожайность	2 Удобрение
1	8	удобрение 1
2	8,4	удобрение 1
3	9	удобрение 1
4	8,6	удобрение 1
5	10,1	удобрение 1
6	8,3	удобрение 2
7	10,2	удобрение 2
8	9	удобрение 2
9	10	удобрение 2
10	10	удобрение 2
11	9,3	удобрение 2
12	8,8	удобрение 3
13	10	удобрение 3
14	9,7	удобрение 3
15	12	удобрение 3
16	9,2	удобрение 3
17	9,1	удобрение 3
18	8,2	удобрение 3

Рис. П65

Шаг 2. Запуск модуля *Kruskal – Wallis ANOVA and Median Test*. Используйте команды *Statistics* → *Nonparametrics* или нажмите кнопку на дополнительной панели инструментов. В диалоговом окне *Nonparametric Statistics* выберите процедуру *Comparing multiple indep. samples* (*groups*). Нажмите *OK*.

Шаг 3. Выбор переменных. В открывшемся модуле *Kruskal – Wallis ANOVA and Median Test* нажмите кнопку *Variables*. Выберите переменные, нажмите *OK***.** Нажмите кнопку *Codes* (Коды), затем *All* – указали программе, что в анализе будут участвовать все экспериментальные группы (рис. П66).

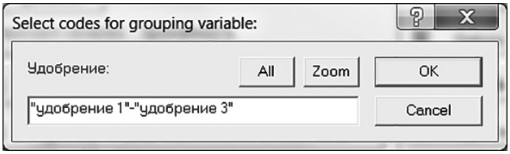


Рис. П66

После данных действий нажмите *Summary: Kruskal – Wallis ANOVA & Median test* или кнопку *Summary* (рис. П67).

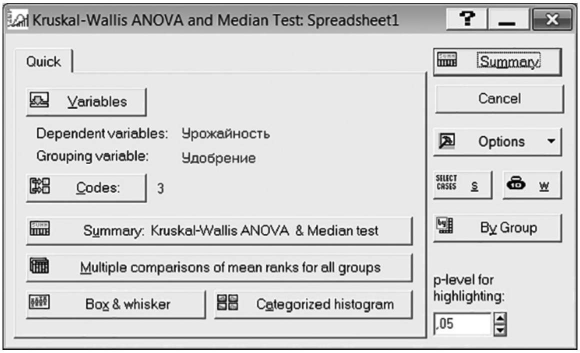


Рис. П67

Шаг 4. Интерпретация полученных результатов. Принятие гипотезы. Слева, в поле открывшейся рабочей книги, выберите *Kruskal – Wallis ANOVA by Ranks*. Полный результат теста Краскела – Уоллиса представлен на рис. П68.

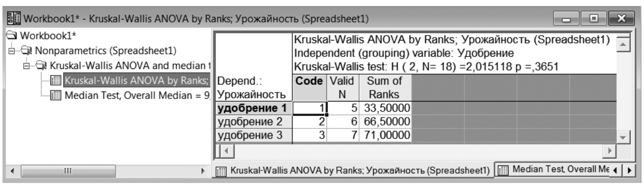


Рис. П68

В таблице с результатами расчетов приведены:

• *Code* – уникальный код группы (число);

• *Valid N* – число значений в группе;

• *Sum of Ranks* **–** сумма рангов;

• *Kruskal – Wallis test:* H (2, N = 18) = 2,015118 p = ,3651 (2 – число степеней свободы; N – общее количество наблюдений в объединенной выборке); 2,015118 – эмпирическое значение H-теста; p – вероятность справедливости нулевой гипотезы.

Для принятия или опровержения нулевой гипотезы интерпретируют величину p. Так как $P > 0,05$ ($p = 0,3651$), можно утверждать, что нулевая гипотеза верна и урожайность озимой ржи в экспериментальных группах одинакова.

Шаг 5. Визуализация описательных статистик сравниваемых групп. Модуль *Kruskal – Wallis ANOVA and Median Test* предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами. Нажмите кнопку *Box & whisker* и выберите столбец с данными «Урожайность», нажмите *OK*. В окне *Box-Whisker Type* (рис. П69) подсветите один из предложенных вариантов построения графика «Коробочка с усами», нажмите *OK*. График получен (рис. П70).

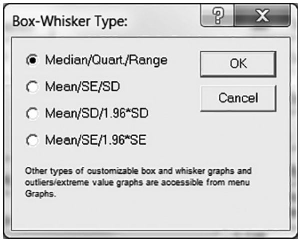


Рис. П69

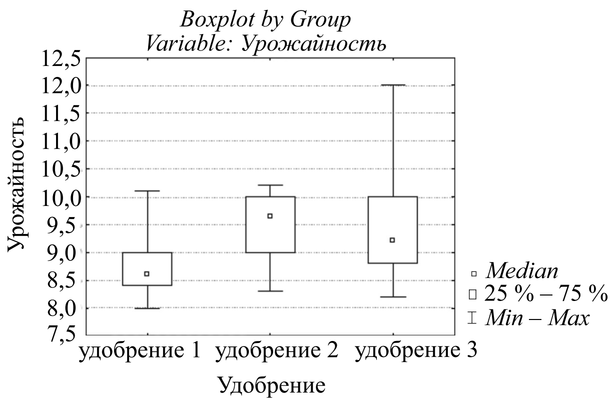


Рис. П70

10. СРАВНЕНИЯ ДВУХ И БОЛЕЕ ЗАВИСИМЫХ ГРУПП С ИСПОЛЬЗОВАНИЕМ ПРОГРАММЫ *STATISTICA*. КРИТЕРИЙ *FRIEDMAN ANOVA*

В качестве примера использованы данные опроса о времени подготовки студентов к семинару (см. подразд. 4.2.2).

Шаг 1. Ввод данных. Создайте электронную таблицу, состоящую из четырех столбцов и шести строк. Каждый столбец отражает время, затраченное каждым студентом для подготовки к семинару в определенный день недели (рис. П71).

Data: Spreadsheet1* (5v by 6c)

	1 Понедельник	2 Вторник	3 Среда	4 Четверг
Олег	9	4	10	11
Семен	4	2	4	8
Юрий	8	7	10	15
Петр	4	9	7	10
Иван	7	8	10	10
Игнат	9	12	14	17

Рис. П71

Шаг 2. Запуск модуля *Friedman ANOVA by Ranks*. Используйте команды *Statistics → Nonparametrics* или нажмите кнопку на дополнительной панели инструментов. В диалоговом окне *Nonparametric Statistics* выберите процедуру *Comparing multiple dep. samples* (*variables*), нажмите *OK*.

Шаг 3. Выбор переменных. В открывшемся модуле *Friedman ANOVA by Ranks* нажмите кнопку *Variables*, выберите переменные. Нажмите на кнопку *Summary: Friedman ANOVA &Kendall' s concordance* или кнопку *Summary* (рис. П72).

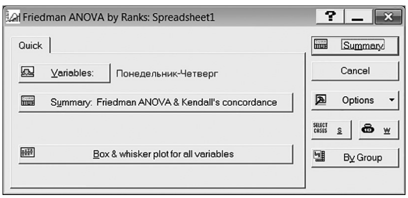


Рис. П72

Шаг 4. Интерпретация полученных результатов. Принятие гипотезы. В таблице с результатами расчетов приведены (рис. П73):

- *ANOVA Chi Sqr.* ($N = 6$, $df = 3$) = 13,60345 p = ,00350 (N – число наблюдений в группе; $df = 3$ – число степеней свободы), 13,60345 – эмпирическое значение теста Фридмана; p – вероятность справедливости нулевой гипотезы;
- *Average Rank* – средний ранг;
- *Sum of Ranks* – сумма рангов по столбцу;
- *Mean* – среднее значение по столбцу;
- *Std. Dev* – стандартное отклонение;
- *Coeff. of Concordance* – коэффициент ранговой корреляции Кендалла.

Friedman ANOVA and Kendall Coeff. of Concordance (S
ANOVA Chi Sqr. (N = 6, df = 3) = 13,60345 p = ,00350
Coeff. of Concordance = ,75575 Aver. rank r = ,70690

Variable	Average Rank	Sum of Ranks	Mean	Std.Dev.
Понедельник	1,583333	9,50000	6,83333	2,316607
Вторник	1,666667	10,00000	7,00000	3,577709
Среда	2,833333	17,00000	9,16667	3,371449
Четверг	3,916667	23,50000	11,83333	3,430258

Рис. П73

Анализируя *Sum of Ranks*, можно говорить о качестве влияния фактора. Очевидно, что продолжительность подготовки увеличивается по мере приближения даты семинара.

Coeff. of Concordance рассчитывается путем усреднения коэффициентов корреляции Спирмена для каждой пары, участвующей в анализе групп. Чем больше различия между группами, тем ближе коэффициент корреляции Кендалла к 1.

Шаг 5. Визуализация описательных статистик сравниваемых групп. Модуль *Friedman ANOVA by Ranks* (см. рис. П72) предоставляет возможность построить график «Коробочка с усами» для визуальной оценки разницы между сравниваемыми группами.

Нажмите кнопку *Box & whisker plot for all variables*. В появившемся окне *Box-Whisker Type* (рис. П74) подсветите один из предложенных вариантов построения графика «Коробочка с усами», нажмите *OK*. График получен (рис. П75).

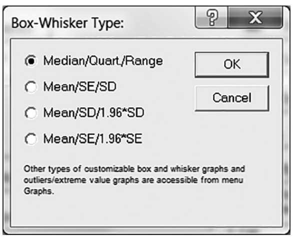


Рис. П74

Рис. П75

СПИСОК ЛИТЕРАТУРЫ

Baldi, B. The practice of Statistics in the Life Sciences / B. Baldi, D. S. Moore. – 2nd ed. – New York, 2012. – 721 p.

Sokal, R. R. Introduction to Biostatistics: second edition / R. R. Sokal, F. J. Rohlf. – 2nd ed. – Dover edition, 2009. – 366 p.

Zar, J. H. Biostatistical analysis / J. H. Zar. – 5th ed. – Pearson, 2010. – 946 p.

Берк, К. Н. Анализ данных с помощью Microsoft Excel / К. Н. Берк, П. Кэйри. – М., 2005. – 560 с.

Боровиков, В. STATISTICA. Искусство анализа данных на компьютере / В. Боровиков. – СПб., 2003. – 688 с.

Волкова, П. А. Статистическая обработка данных в учебно-исследовательских работах / П. А. Волкова, А. Б. Шипунов. – М., 2008. – 60 с.

Вуколов, Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учеб. пособие / Э. А. Вуколов. – 2-е изд., испр. и доп. – М., 2008. – 464 с.

Гланц, С. Медико-биологическая статистика / С. Гланц. – М., 1999. – 459 с.

Гржибовский, А. М. Сравнение количественных данных трех и более парных выборок с использованием программного обеспечения STATISTICA И SPSS: параметрические и непараметрические критерии / А. М. Гржибовский, С. В. Иванов, М. А. Горбатова // Наука и здравоохранение. – 2016. – № 2. – С. 5–29.

Жукова, А. А. Биометрия : пособие. В 3 ч. Ч. 1. Описательная статистика / А. А. Жукова, М. Л. Минец. – Минск : БГУ, 2019. – 100 с.

Ивантер, Э. В. Элементарная биометрия : учеб. пособие / Э. В. Ивантер, А. В. Коросов. – Петрозаводск, 2010. – 104 с.

Ивантер, Э. В. Введение в количественную биологию : учеб. пособие / Э. В. Ивантер, А. В. Коросов. – Петрозаводск, 2011. – 302 с.

Куприенко, Н. В. Статистика. Методы анализа распределений. Выборочное наблюдение / Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – СПб., 2009. – 138 с.

Лакин, Г. Ф. Биометрия : учеб. пособие / Г. Ф. Лакин. – 4-е изд., перераб. и доп. – М., 1990. – 352 с.

Мастицкий, С. Э. Методическое пособие по использованию программы STATISTICA при обработке данных биологических исследований / С. Э. Мастицкий. – Минск, 2009. – 76 с.

Новиков, Д. А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д. А. Новиков, В. В. Новочадов. – Волгоград, 2005. – 84 с.

Петри, А. Наглядная статистика в медицине : пер. с англ. / А. Петри, К. Сэбин. – 2-е изд., перераб. и доп. – М., 2009. – 168 с.

Рокицкий, П. Ф. Биологическая статистика / П. Ф. Рокицкий. – Минск, 1973. – 320 с.

Сигел, Э. Практическая бизнес-статистика / Э. Сигел. – М., 2004. – 1056 с.

Тюрин, Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. – М., 1998. – 528 с.

Унгуряну, Т. Н. Сравнение трех и более независимых групп с использованием непараметрического критерия Краскела-Уоллиса в программе STATA / Т. Н. Унгуряну, А. М. Гржибовский // Экология человека. – 2014. – № 6 – С. 55–58.

Халафян, А. А. STATISTICA 6.0. Статистический анализ данных / А. А. Халафян.– 3-е изд. – М., 2007. – 512 с.

Харькова, О. А. Сравнение двух несвязанных выборок с использованием пакета статистических программ STATA: непараметрические критерии / О. А. Харькова, А. М. Гржибовский // Экология человека. – 2014. – № 4. – С. 60–64.

Харькова, О. А. Сравнение двух парных выборок с помощью пакета статистических программ STATA: непараметрические критерии / О. А. Харькова, А. М. Гржибовский // Экология человека. – 2014. – № 12. – С. 55–60.

Шитиков, В. К. Количественная гидроэкология: методы системной идентификации / В. К. Шитиков, Г. С. Розенберг, Т. Д. Зинченко. – Тольятти, 2003. – 463 с.

Содержание

Учебное издание

Жукова Анна Анатольевна
Минец Маргарита Леонидовна

БИОМЕТРИЯ

Пособие

В трех частях

Часть 2

Основные техники анализа данных

Редактор *Д. В. Мацур*
Художник обложки *Т. Ю. Таран*
Компьютерная верстка *Д. О. Бабенко*

Подписано в печать 10.11.2020. Формат 60×84/16. Бумага офсетная.
Цифровая печать. Усл. печ. л. 8,83. Уч.-изд. л. 10,4.
Тираж 135 экз. Заказ 8300.

Белорусский государственный университет.
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/270 от 03.04.2014.
Пр. Независимости, 4, 220030, Минск.

Издательско-полиграфическое частное унитарное предприятие «Донарит».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/289 от 17.04.2014.
Ул. Октябрьская, 25, офис 2, 220030, г. Минск, Республика Беларусь.