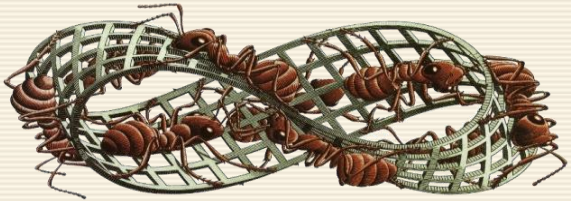
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ

Бондарчук С.С. Бондарчук И.С.

Статобработка экспериментальных данных в MS Excel

учебное пособие



Томск, 2018

УДК 519.254
ББК 22.172я73

С.С. Бондарчук, И.С. Бондарчук.
Б81 Статобработка экспериментальных данных в MS Excel: учебное пособие.
– Томск: Издательство Томского государственного педагогического университета, 2018. – 433 с.

Издание охватывает базовые направления математической статистики, используемые при анализе экспериментальных данных в областях биологии, медицины, химии, при анализе физических и педагогических измерений.

Каждый раздел издания посвящен конкретному аспекту анализа и обработки данных эксперимента. Теоретические положения сопровождается и дополняется методическими указаниями и примерами, реализованными исключительно в среде MS Excel. Критические значения статистик, не поддерживаемых функциями электронных таблиц, представлены соответствующими аппроксимациями.

Для студентов (бакалавров, магистрантов и аспирантов), преподавателей и научных работников, занимающихся анализом и статистической обработкой экспериментальных данных.

УДК 519.254
ББК 22.172я73

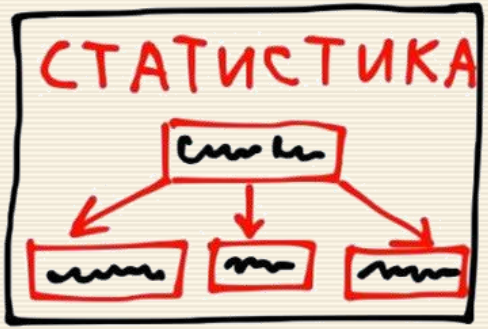
Рецензенты: В.А. Архипов, доктор физико-математических наук, профессор
А.С. Жуков, доктор физико-математических наук

Печатается в рамках Программы повышения конкурентоспособности ТГУ
Публикуется в авторской редакции
ISBN 978–5–89428–861–1

О г л а в л е н и е

Предисловие

При разработке настоящего пособия решались две главные задачи: дать читателю краткие сведения по основополагающим понятиям и методам статистики (сложившуюся терминологию, традиционные обозначения, основные принятые определения и формулы); а также представить практическую реализацию этих методов к решению различных прикладных вопросов.



Каждый раздел издания посвящен отдельному аспекту анализа данных только с помощью MS Excel; макросы VBA для решения задач не используются. В достаточной мере прокомментированы и пояснены алгоритмы, реализующие методы статистического анализа. Все примеры и скриншоты листов электронных таблиц, сопровождающих изложение, содержат точное указание всех используемых формул и делают, по мнению авторов, процесс освоения представленного материала максимально легким, приятным и ненавязчивым. В этом отношении издание можно рассматривать как "сборник рецептов" методов статистики, наиболее часто применяемых в практике анализа экспериментальных данных при их относительно простой реализации на электронных таблицах MS Excel.

Существенное внимание уделено примерам по темам, традиционно вызывающим затруднение при освоении материала: "при изучении наук примеры полезнее правил" (сэр Исаак Ньютон). Иллюстрированный примерами материал в своем изложении строился по принципу полноты и воспроизводимости: читатель имеет возможность для приведенных исходных данных проверять и контролировать результаты вычислений. Умение решать задачи означает не только знание теории, но и способность использовать имеющееся знание для получения нового. Исходные данные примеров взяты или соответствуют таковым для наиболее популярных сборников задач, учебных пособий и монографий.

Изложенные положения базовых элементов теории статистики ориентированы на наглядным объяснение используемых методов с определенной математической строгостью, поскольку "в любой науке столько истины, сколько в ней математики" (Иммануил Кант), а целью математики, да и, в конечном итоге, статистики, является осмысление действительности.

Microsoft Excel входит в состав Microsoft Office и на сегодняшний день является одним из наиболее популярных приложений в мире. Благодаря простоте инструментария и наличия большого количества статистических и других полезных встроенных функций в рамках пакета может решаться широкий спектр прикладных задач, связанные с анализом экспериментальных данных. MS Excel доступен на большей части учебных, лабораторных и домашних компьютеров. "Прозрачность" реализации алгоритмов вычислений для представленных примеров решения задач обработки данных инструментарием электронных таблиц позволит читателю "прочувствовать" статистические методы, порядки и зависимости числовых величин, а освоение материала обеспечит накопление шаг за шагом необходимого опыта восприятия и теории статистики, и квалифицированного использование специализированного программного обеспечения. "Adde parvum parvo magnus acertus erit" ("Добавляя малое к малому, получишь большую кучу", – Овидий).

Предполагается, что читатель знаком с основами работы в электронных таблицах – структурой данных в ячейках, абсолютной и относительной адресацией, арифметическими операциями, автозаполнением диапазонов ячеек и пр. Не применяющийся в обыденной практике инструментарий (именование массивов, формул и работа с массивами) представлен в приложении П5, где дано его соответствующее описание. В этом же приложении приведен перечень сообщений MS Excel об ошибках в представлении данных и при выполнении операций, способах их устранения.

Статистическая обработка данных – задача достаточно непростая, требующая базовых знаний теории и наработанного практического опыта. Авторы данного пособия придерживаются мнения о том, что наличие практических навыков в сочетании со знанием основных понятий и определений создадут тот базис, который далее позволит продвигаться читателю в направлении профессионализма выполнения как статистических исследований, так и в практике решения иных расчетных задач.

Приведенные примеры и нижеследующий расширенный перечень тем могут сориентировать и помочь читателю разобраться с порядком и выполнением практических статистических исследований.

Описательная статистика, представление данных

Оценивается нормальность распределения признака (см. ниже рекомендации по теме "Оценка нормальности распределения…"). Если распределение нормальное, то при анализе данных указываются среднее значение, стандартное отклонение, доверительный интервал (раздел 1.4 пример 1.2) и стандартная ошибка. Анализ погрешностей прямых и косвенных измерений дан в пораздела П4.3.

При описании количественного признака, распределение которого отличается от нормального, указываются медиана, значения нижнего и верхнего квартилей (или 25% и 75% перцентилей).

При описании качественного признака для каждого его значения указываются абсолютные величины, их процентные доли в структуре совокупности.

При наличии достаточного количества данных строится вариационный ряд (приложение П2, пример П2.1), дифференциальная и интегральная функции распределения (раздел 1.2, пример 2.1), полигон частот (пример 1.3).

Для выборки, заданной в том числе и в форме интервального ранжированного частотного ряда, в примере П2.2 приведены расчетные формулы описательной статистики: определения среднего, медианы, моды, параметров вариации и некоторых гистограмм.

В приложение П4 даны правила представления числового материалов исследований:

подраздел П4.1 – округления чисел;

подраздел П4.2 – инструментария округления чисел в MS Excel;

подраздел П4.3 – обработки данных прямых и косвенных измерений.

Исключение грубых погрешностей

При экспериментальных измерениях могут появляться ошибки (выбросы, промахи), допущенные исследователем при фиксации измерений, при действиях с приборами и т.д. Выбросом (промахом) считается наблюдение, которое лежит аномально далеко от остальных из серии параллельных наблюдений. То есть выбросы – это значения количественного признака, располагающиеся на краях интервала допустимых значений. Выброс, промах – резко отклоняющееся значение наблюдаемой величины. В англоязычной литературе в качестве синонимического понятия используются термины maverick – резко выделяющийся результат; straggler – оторвавшийся результат.

В разделе 8 представлен ряд критериев и алгоритмы, позволяющие исключить из данных грубые промахи: критерий Романовского, Шарлье, Райта, Диксона и др. Эти критерии основаны на статистических выборочных оценках, поскольку в большинстве случаев действительные значения параметров распределения неизвестны.

Параметрические и непараметрические критерии исключения промахов даны в подразделах

- 8.1 Критерии Райта и правило "трех сигм";
- 8.2 Критерий Романовского (пример 8.1);
- 8.3 Критерий Шарлье (пример 8.2);
- 8.4 Правило "ящик с усами" (пример 8.3);
- 8.5 Правило Томпсона (критерий Рошера) (пример 8.4);
- 8.6 Критерий Диксона (Q-критерий) (пример 8.5).

Оценка нормальности распределения, равенства распределений

Результаты наблюдений можно оценить наиболее полно, если их распределение является нормальным. Поэтому существенную роль при обработке результатов наблюдений играет проверка принадлежности выборки нормальному (Гауссовому) закону распределению с помощью соответствующих методов и критериев.

Критерии нормальности являются частным случаем критериев согласия и составляют группу статистических критериев, предназначенных для проверки соответствия данных распределению Гаусса.

В разделе 3 представлены алгоритмы и примеры проверки данных "на нормальность", когда используются

- 3.1 Критерий согласия Пирсона хи-квадрат (примеры 3.1 и 3.2);
- 3.2 Критерий Колмагорова (пример 3.4);
- 3.4 Критерий Крамера-Мизеса-Смирнова (пример 3.6);
- 3.5 Метод моментов (пример 3.7).

Тестирование данных на нормальность практически всегда является первым этапом анализа экспериментальных результатов, так как большое количество статистических методов исходит из предположения нормальности распределения изучаемых данных.

Если требуется сравнить два произвольных распределения, то используется критерий Колмогорова-Смирнова (пример 3.5).

Сравнение двух и более групп по количественному признаку

Алгоритмы сравнения групп по количественному признаку разделяются на два класса – параметрические и непараметрические.

непараметрические тесты

Для функций распределения в выборках, неподчиняющихся нормальному закону распределения, при сравнении средних двух выборок используются следующие критерии

- для независимых – критерий Крамера-Уэлча (раздел 5.1, пример 5.1);
- для независимых выборок при оценке различий между двумя выборками по признаку, измеренному в количественной или порядковой шкале – критерий Вилкоксона-Манна-Уитни (раздел 5.3, пример 5.3). Критерий оперирует не с абсолютными значениями элементов двух выборок, а с результатами их парных сравнений.
- для связанных – критерий Вилкоксона (раздел 5.4, пример 5.4), применяемый для сопоставления показателей, измеренных в двух разных условиях на одной и той же выборке испытуемых. Критерий позволяет установить не только направленность изменений, но и их выраженность. С его помощью определяется, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Для сравнения трех и более независимых выборок используется дисперсионный анализ по Краскелу-Уоллису (Kruskal-Wallis test) (раздел 5.2, пример 5.2), использующий ранги исходные значения и их суммы в сравниваемых группах. Критерий проверяет нулевые гипотезы о принадлежности выборок одному и тому же распределению или распределениям, характеризующимися одинаковыми медианами.

параметрические тесты

Если функции распределения в выборках соответствуют распределению Гаусса (нормальному), то для сравнения средних двух выборок используются следующие критерии

- для связанных – парный тест Стьюдента (раздел 4.1, пример 4.1);
- для независимых с равными дисперсиями – гомоскедастический тест Стьюдента (раздел 4.1, пример 4.2);
- для независимых с разными дисперсиями – гетероскедастический тест Стьюдента (раздел 4.1, пример 4.3).
- для зависимых и независимых выборок при проверке гипотезы о равенстве долей в двух совокупностях – так называемый z-критерий (раздел 4.2, пример 4.5)). Критерий обобщается до проверки гипотезы о равенстве долей в трех и более совокупностях (пример 4.6).

Критерий Стьюдента равенства среднего значения (раздел 4.1, пример 4.4) используется для сравнения результатов анализа с каким-либо значением, которое можно считать точной (паспортной) величиной.

Для сравнения трех и более независимых выборок используется однофакторный дисперсионный анализ ANOVA (раздел 4.3, пример 4.7), в результате которого находится значение F-критерия Фишера. Оценка статистической значимости различий показателей выборок проводится сравнением рассчитанного значения F-критерия с его критическим значением либо через P-значение.

В случае, если нулевая гипотеза при применении ANOVA отвергается, то оценку разности средних можно провести с помощью теста множественных сравнений Шеффе (раздел 4.4, пример 4.8).

Для выбора той или иной группы тестов (параметрические/непараметрические) предварительно необходимо оценить нормальность распределения в исследуемых выборках (см. указания по анализу на нормальность "Оценка нормальности распределения").

Дисперсионный анализ. Однородность

Дисперсионный анализ включает в себя проверку гипотез о значимости различий двух и более выборок, на основании которых формируется заключение об их однородности. В приложении П1 достаточно подробно описаны понятия однородности дисперсии и ошибки среднего.

В разделе 2.1 представлен алгоритм и пример сравнения дисперсий двух выборок для "классического" критерия Фишера (пример 2.1). Если выборок более двух, когда проведено, например, несколько серий экспериментов, то при одинаковом количестве парных опытов в разделе 2.2 дано краткое описание критерия Кохрена (пример 2.2). Для разного числа парных опытов в серии в разделе 2.3 дан алгоритм и пример 2.3 анализа однородности в рамках критерия Бартлета.

Для сравнения двух связанных (парных) выборок в разделе 2.4 описан критерий знаков. Критерий эффективно используется (примеры 2.4 и 2.5) для решения задач, когда сравниваются измерения, сделанные в разные событийные или временные моменты (например, до и после обработки образцов).

В разделе 2.5 описан тест Левена (Levene's test), используемый для оценки равенства дисперсий нескольких выборок, когда проверяется нулевая гипотеза о равенстве (гомогенности, гомоскедастичности) дисперсий выборок (пример 2.6).

Критерий однородности χ^2, применяемый к данным, измеренным в порядковой шкале по нескольким градациям (например, низкий, средний, высокий) при анализе однородности двух выборок, дан в разделе 2.6 (пример 2.7).

Для данных, измеренных в порядковой дихотомической шкале для двух градаций (да/нет и т.д.), при анализе однородности двух выборок (например, экспериментальной и контрольной групп) в разделе 2.7 (примеры 2.8, 2.9) описан алгоритм применения критерия (углового преобразования) Фишера.

Исследование статистических связей

Раздел 6 руководства представляет описание элементов теории и алгоритмов, относящихся к исследованию статистических связей двух (зависимой и независимой) переменных.

В подразделе 6.1 дано краткое описание основных теоретических положений, которые легли в понятие линейной регрессии (Linear regression) – широко используемой в практических исследованиях модели функциональной зависимости одной величины от другой. Дан алгоритм построения уравнения регрессии (пример 6.1), проводится анализ значимости и точности коэффициентов; строится доверительный интервал для коэффициентов и определяются показатели качества уравнения регрессии (коэффициенты детерминации, эластичности и средняя ошибка аппроксимации).

Для относительно сложных функциональных зависимостей, которые можно свести к линейным, в подразделе 6.2 рассматриваются задачи линеаризации нелинейных зависимостей с помощью замены переменных (примеры 6.2 и 6.3).

Для задачи идентификация параметров математической модели, описывающей изменение во времени концентрации для простой химической реакции в примере 6.4 представлен алгоритм расчета порядка и константы реакции на базе исходных таблицы экспериментальных данных.

В подразделе 6.3 представлены краткое описание расчета коэффициентов корреляции Пирсона и Спирмена, приводится таблица силы (тесноты) связи по шкале Чеддока (scale Cheddok). Определяются значимость (примере 6.7) и доверительный интервал линейного коэффициента корреляции.

В примерах 6.5 и 6.6 даны применения коэффициентов бисериальной корреляции для анализа дискриминативности заданий теста, а также при определении валидности посредством коррелирования значений тестовых оценок с независимыми характеристиками критерия, выраженными в дихотомической шкале.

В подразделе 6.4 (пример 6.8) описана задача градуировки (калибровки), заключающаяся в создании шкалы (градуировочной функции типа уравнения линейной регрессии, таблицы), где связываются показания средства измерения с искомым значением физической величины. Исходя из уравнения градуировочной линии и статистических характеристик фонового сигнала определяются наименьшее содержание вещества $C_{\min}$, которое может быть обнаружено, а также предел "надёжного обнаружения" аналита $C_{\text{над}}$.

Сравнение относительных показателей (частот, долей…) между группами. Таблицы сопряженности

Таблицы сопряженности применяются для исследования вопроса о наличии статистической зависимости между распределениями двух переменных при проверке гипотезы о существовании связи между двумя признаками с использованием (чаще всего) точного теста Фишера или критерия согласия Пирсона).

Наиболее распространенной, в частности в медикобиологических и педагогических исследованиях, является проблема установления статистической значимости различий в распределении частот в одной и той же группе объектов до и после применения контролируемых воздействий.

Типовые задачи, решаемые с помощью таблиц сопряженности, приведены в примерах с 7.1 по 7.11 раздела 7**.**

Выбор метод анализа зависит размерности таблицы и уровня числовых значений в ячейках:

- если в таблице 2×2 наименьшее значение ожидаемых частот менее 5, то для сравнения используется точный критерий Фишера (раздел 7.1, примеры 7.5 , 7.6 , 7.7 и 7.8),
- если в таблице 2×2 наименьшее значение ожидаемых частот находится в интервале от 5÷10, то для сравнения используется критерий хи-квадрат Пирсона с поправкой на непрерывность Йейтса (раздел 7.1, пример 7.1).

В остальных случаях таблиц большей размерности используется критерий хи-квадрат Пирсона (раздел 7.2, примеры с 7.9 , 7.10 и 7.11).

При анализе таблиц 2×2 в сомнительных случаях, когда, например, анализ данных с учетом и без учета поправки Йейтса дают противоположные результаты, для расчетов можно использовать критерий χ^2 с поправкой на максимального правдоподобия (раздел 7.1, пример 7.2).

Если при проверке статистических гипотез требуется определить не только достигнутые уровни значимости связи, но и оценить величину эффекта (effect size), то есть нормированную от 0 до 1 силу связи между признаками, то можно применить критерий V Крамера (Cramer's V test), методика которого дана в разделе 7.1, пример 7.3.

Когда (например, в анализе медицинских данных) кроме исследования связей в таблицах 2×2 для количественной оценки зависимости вероятности исхода от факторных воздействий необходимо рассчитать на заданном уровне значимости показатели отношения шансов или фактора рисков, можно использовать схемы вычислений, данные разделе 7.3, пример 7.12.

Знакомясь с этим изданием, читатель может пользовать его как учебное и справочное руководство. Представленный материал подходит для каждого, кто совершенствует или еще только приобретает навыки по анализу экспериментальных данных.

Издание подготовлено при поддержке Программы повышения конкурентоспособности ТГУ.

Путь к мудрости

К мудрости путь – по ухабам ошибок;
Иди же и носа не вешай:
Ушибы, ушибы и снова ушибы,
Но реже, и реже, и реже

The road to wisdom

The road to wisdom? – Well, it's plain
and simple to express:
Err and err and err again
but less

Piet Hein

Все вероятности равны 50%. Либо случится, либо нет.
Мерфология, логические предложения Кольварда

1. Базовые термины

> С т а т и с т и к а (от лат.: *status* – состояние дел) – отрасль знаний, наука, в которой излагаются общие вопросы сбора, измерения и анализа массовых статистических (количественных или качественных) данных; изучение количественной стороны массовых явлений в числовой форме.

Математической статистикой называется раздел математики, занимающийся разработкой методов получения, описания и обработки экспериментальных данных с целью выявления и изучения закономерностей случайных массовых явлений для научных и практических выводов.

1.1. Измерения, шкалы и величины

В основе любого наблюдения и анализа лежат измерения – процедура, которая данному наблюдаемому состоянию объекта ставит в соответствие определенное обозначение: число, номер или символ. Измерением является процесс присвоения значений характеристик изучаемых объектов согласно определенных правил. В процессе подготовки данных (зарегистрированной информации) измеряется не сам объект, а его параметры, характеристики, которые и подвергаются статистическому исследованию.

Статистические исследования имеют дело со случайными величинами.

Случайная величина (random variable) – некоторая функция, принимающая одно из своих возможных значений в результате эксперимента (опыта, испытания) такая, что для любой совокупности её значений можно указать вероятность* того, что полученное в результате эксперимента конкретное значение будет принадлежать этой совокупности. В результате определяется распределение вероятностей случайной величины.

Случайная величина полностью определяется своим распределением вероятностей.

Шкала – правило, в соответствии с которым объектам присваиваются значения и формируются данные исследования различного типа (символьные, порядковые и числовые). Числовые данные могут быть дискретными и непрерывными.

Дискретные данные (discrete data) являются значениями признака, общее число которых конечно или бесконечно и может быть представлено (подсчитано) при помощи натуральных или целых чисел.

Пример дискретных данных: суточный городской пассажиропоток, количество листьев у растения.

Непрерывные данные (continues data) – данные, принимающие любые значения в некотором интервале. Непрерывные данные предполагают некоторую точность измерения.

Пример непрерывных данных: температура, вес, длина.

Вопросы сплошного и выборочного наблюдений, колеблемость, динамика, тенденции развития, цикличность и сезонность данных обсуждаются в приложениях П1.1 и П1.2.

---

* Вероятность события A, обозначаемая $P(A)$ есть число в диапазоне от нуля до единицы, указывающее, насколько правдоподобно, что событие A произойдет.

Ш к а л ы . Кроме специальных имеется несколько обычно используемых типов шкал измерений: номинальная, порядковая, интервальная, относительная и дихотомическая.

Н о м и н а л ь н а я ш к а л а (nominal scale) – шкала, содержащая только категории; данные в ней не могут упорядочиваться, с ними не могут быть произведены никакие арифметические действия.

Номинальная шкала состоит из названий, категорий, имен для классификации и сортировки объектов или наблюдений по некоторому признаку. Для этой шкалы применимы две операции: равно (=) и не равно (≠). Пример шкалы: профессия, город проживания, семейное положение.

П о р я д к о в а я ш к а л а (ordinal scale) – шкала, в которой числа присваивают объектам для обозначения относительной позиции объектов, но не величины различий между ними. Шкала измерений дает возможность ранжировать значения, но не позволяет определить "насколько одна величина больше (или меньше) другой. Для этой шкалы применимы 4 операции: равно (=), не равно (≠), больше (>) и меньше (<). Пример такой шкалы: номер в рейтинге популярности (1-й, 3-й и т.д.), место в соревнованиях по бегу (1, 2, 3-е), хотя при этом неизвестно, насколько один бегун быстрее другого.

И н т е р в а л ь н а я ш к а л а (interval scale) – шкала, разности между значениями которой могут быть выражены числами, однако отношения их значений не имеют особого смысла. Измерения в этой шкале упорядочены по рангам и разделены определенными интервалами. В шкале используются размерные единицы измерения (градус, секунда и т.д.); измеряемому объекту присваивается число, равное количеству единиц измерения, которое он содержит. Нулевая точка шкалы выбирается произвольно – например, летоисчисления в разных календарях, нулевая температура и т.д. Результаты измерения по шкале интервалов можно обрабатывать всеми математико-статистическими методами, кроме вычисления отношений.

Относительная шкала, шкала отношений (ratio scale) – шкала, в которой есть определенная точка отсчета и возможны отношения между значениями шкалы. Эта шкала имеет строго определенную нулевую точку и не накладывает никаких ограничений на использование математического аппарата. При оценке результатов измерений по этой шкале можно определить, "во сколько раз" один объект больше другого.

При использовании шкалы отношений измерение какой-либо величины сводится к экспериментальному определению соотношению этой величины с эталоном. При измерении размера можно определить, во сколько раз длина объекта больше длины другого тела, принятого за единицу – например, метровой линейки.

Дихотомическая шкала (dichotomous scale) – шкала, содержащая только две категории. В литературе наряду с этим используется термин "бинарные данные". Пример шкалы: пол (мужской и женский).

Признак (characteristic, variable) – характеристика объекта исследования (наблюдения). Различают качественные (attribute, qualitative) и количественные (quantitative) признаки.

Номинальные (nominal data) и порядковые (ordered data) данные, которые отражают условные коды неизмеряемых категорий или условную степень выраженности признака, называют качественными данными (qualitative data).

Количественные данные (quantitative), измеряемые числами, имеющих содержательный смысл. Количественные данные могут быть непрерывными (continues data) или дискретными (discrete data).

1.2. Генеральная совокупность. Выборка

> Генеральная совокупность (анг.: parent population, general population, лат.: generalis – общий, всеобщий) – совокупность объектов, из которых производится выборка.

Это совокупность всех подлежащих изучению объектов или мыслимых результатов наблюдений, которые могут быть получены в данных (неизменных) условиях. Различают конечные, содержащие конечное число элементов, и бесконечные, содержащие бесконечное число элементов, генеральные совокупности.

Генеральная совокупность часто содержит конечное число объектов. Однако если это число достаточно велико, то иногда, в целях упрощения вычислений или для облегчения теоретических выводов, допускают, что генеральная совокупность состоит из бесчисленного множества объектов. Такое допущение оправдывается тем, что увеличение объёма генеральной совокупности практически не сказывается на результатах обработки данных выборки.

Генеральная совокупность характеризуется генеральным распределением с генеральными параметрами, например, генеральным математическим ожиданием и генеральной дисперсией. Генеральные параметры могут оцениваться по выборочным данным.

В реальности параметры генеральной совокупности неизвестны и ее характеристики носят характер предположения, которое называется статистической гипотезой. Если предположение противоречит наблюдаемым данным, то гипотезу отклоняют, как ложную; если не противоречит, то принимают. Степень противоречия определяется вероятностью, которая обычно задается и, в свою очередь, зависит от степени различия фактической и генеральной совокупностей.

Статистическая гипотеза – предположение о виде и свойствах распределения величины, которое можно подтвердить или опровергнуть применением статистических методов к данным выборки.

> **Статистическая гипотеза** – предположение о виде и свойствах распределения величины, которое можно подтвердить или опровергнуть применением статистических методов к данным выборки.

> **Выборка (выборочная совокупность)** – часть объектов из генеральной совокупности, отобранных для изучения, с целью получения информации о всей генеральной совокупности.

Характеристики выборки:

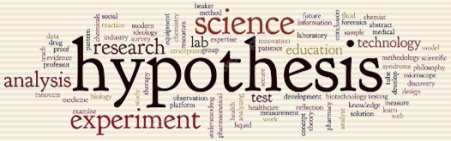
- Качественная характеристика выборки – природа выбираемых объектов и используемые способы построения выборки;
- Количественная характеристика выборки — количество выбранных случаев; число объектов, составляющих выборочную совокупность, часто называемая объемом выборки.

Относительно времени статистические исследования могут быть

- **проспективными**, когда выборки выделяются на основе исходного фактора, а в выборках анализируется некоторый результирующий фактор;
- **ретроспективными**, когда выборки выделяются на основе результирующего фактора, а в выборках анализируется некоторый исходный фактор.

При сравнении двух (и более) выборок важным параметром является их зависимость. Если можно установить гомоморфную пару (то есть, когда одному случаю из выборки X соответствует один и только один случай из выборки Y и наоборот) для каждого случая в двух выборках (и это основание взаимосвязи является важным для измеряемого на выборках признака), такие выборки называются зависимыми (парными). Примером зависимых выборок являются два измерения какого-либо признака до и после экспериментального воздействия. В случае, если такая взаимосвязь между выборками отсутствует, то эти выборки считаются независимыми.

Соответственно, зависимые выборки всегда имеют одинаковый объём, а объём независимых может отличаться.



Сравнение выборок производится с помощью различных статистических критериев (например, Пирсона, Стьюдента, критерия знаков и пр.). Критерий для проверки гипотезы (hypothesis test; синонимы statistical test – статистический критерий; significance test – критерий значимости; test – критерий) – это решающее правило (метод), отвергающее или принимающее нулевую гипотезу на основе выборочных наблюдений.

Статистический критерий – строгое математическое правило, по которому принимается или отвергается та или иная статистическая гипотеза с известным уровнем значимости.

> Гипотеза (hypothesis) научная – утверждение, которое можно подтвердить или опровергнуть на основании результатов исследования.
>
> Статистическая гипотеза – представление научной гипотезы в форме, приемлемой для проверки методами статистического анализа данных.

Методы математической статистики позволяют оценить случайную ошибку изучаемых признаков выборки определенного объема. Также можно решить обратную задачу – определить объем выборки, удовлетворяющий заданным требованиям точности.

Кроме объема выборки, существенную роль играет способ формирования выборки. Не вдаваясь в детали, можно отметить, что выборка, которая сохраняет все свойства генеральной совокупности, называется репрезентативной выборкой.

Свойство репрезентативности – необходимое условие для того, чтобы выводы, сделанные для выборочной совокупности, можно было распространить на генеральную совокупность.

> Выборка является репрезентативной (или представительной), если она достаточно полно представляет изучаемые признаки генеральной совокупности.

Условием обеспечения репрезентативности выборки является, согласно закону больших чисел, соблюдение случайности отбора, т.е. все объекты генеральной совокупности должны иметь равные вероятности попасть в выборку.

Элемент выборки – случайная величина – переменная, которая в результате испытания в зависимости от случая принимает одно из множества возможных значений (заранее неизвестное). Более строго, случайная величина определяется как функция, заданная на множестве элементарных исходов (или в пространстве элементарных событий). Различают дискретные и непрерывная случайные величины.

Непрерывная случайная величина – случайная величина, функция распределения которой непрерывна в любой точке и дифференцируема всюду, кроме отдельных точек; множество возможных значений случайной величины бесконечно или несчетно.

Анализ репрезентативности выборки особенно важен на начальном этапе исследований, когда численность генеральной совокупности неизвестна, но известны некоторые параметры опыта, позволяющие оценить репрезентативность.

В простейшем случае достаточную численность выборки n^* можно оценить по формуле

$$n^* = \left(\frac{t_\infty S}{\Delta}\right)^2,$$

t_∞ – значение обратной функции распределения Стьюдента с "числом степеней свободы" для заданной стандартной доверительной вероятности,

S – стандартное отклонение, рассчитанное по выборке,

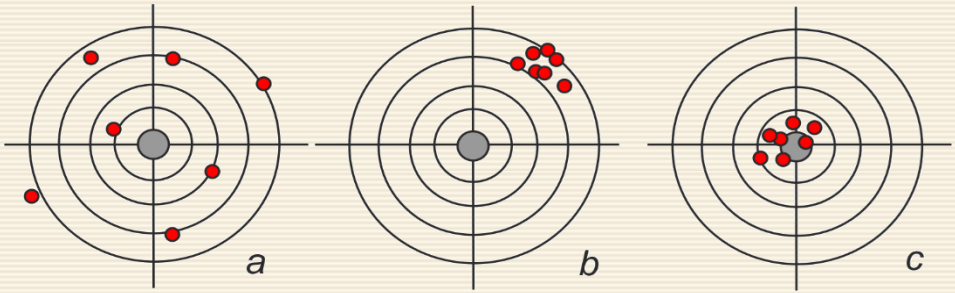
Δ – заданная абсолютная погрешность определения среднего арифметического значения, введенная в именованных числах, т.е. в тех же единицах измерения, что и варианты выборки.

В случае подсчета количества неделимых объектов исследования (например, количество листьев) абсолютная погрешность может быть установлена равной единице.

С увеличением числа параллельных опытов (экспериментов в одинаковых условиях) связаны понятия воспроизводимости и точности результатов.

> Воспроизводимость (reproducibility) характеризует то, насколько близко находятся друг от друга независимые повторные измерения.

> Точность (precision) результатов измерений определяет степень их близости оценок к истинному (стандартному, паспортному) значению параметра; близость к нулю погрешности результата измерения.



a) невысокая точность и воспроизводимость;
b) хорошая воспроизводимость;
c) хорошая точность и воспроизводимость.

1.3. Функции распределения

Предел функции "3 в квадрате", когда 3 стремится к 4, равен 16.
Математический анекдот

Эмпирическая функция распределения (sample distribution function). Пусть $X_1, \ldots, X_n$ случайная выборка из совокупности $\mathbb{R}$. Эмпирическая функция распределения для любого α определяется по формуле

$$F_n(\alpha) = \frac{\text{число наблюдений } X_i \text{ из выборки, для которых } X_i \leq \alpha}{n}.$$

Эмпирическую функцию распределения можно рассматривать как приблизительную оценку истинной функции распределения $F(x)$ случайной величины X. Очевидно, что точность такого приближения обычно улучшается с ростом объёма выборки.

Интегральная функция распределения (cumulative distribution function). Функция распределения $F(x)$ случайной величины X, определенная для всех x, определяет вероятность исхода $P\{X \leq x\}$ или, другими словами, события вида "X меньше x".

Обе функции играют важнейшую роль в широком спектре научных и технических приложений. Например, на рис. 1.1 приведены дифференциальная и интегральная функции распределения, характеризующие размеры микрочастиц алюминия, получающиеся при распылении жидкого металла форсуночным устройством.

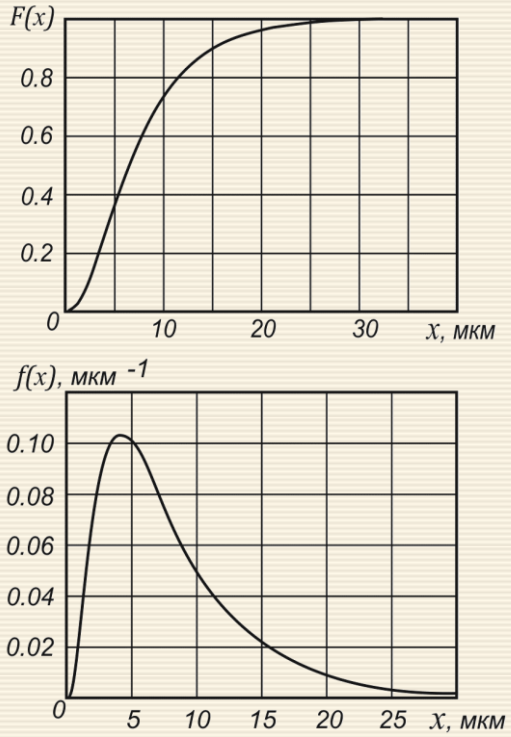


Рис. 1.1. Интегральная и дифференциальная функции распределения размеров микрочастиц алюминия

Как правило, для интегральной функции распределения используется обозначение $F(x)$. То есть, согласно определению, $P\{X \le x\}$. Если X – случайная величина из совокупности $\mathbb{R}$, то задание $F(x)$ для всех x полностью характеризует совокупность $\mathbb{R}$.

$F(x)$ — неубывающая функция, которая стремится к 0 при $x \to -\infty$ и стремится к 1 при $x \to +\infty$.

С помощью $F(x)$ можно найти вероятность попадания случайной величины X в любой промежуток $(a, b]$ на прямой $P(a < X \le b) = F(b) - F(a)$.

Производная от интегральной функции распределения $F(x)$ непрерывной случайной величины называется дифференциальной функцией распределения этой случайной величины или дифференциальным законом распределения

$$f(x) = \frac{dF(x)}{dx}.$$

Дифференциальная функция распределения иначе называется плотностью распределения вероятности. График дифференциальной функции распределения $f(x)$ называется кривой распределения вероятностей случайной величины X.

Дифференциальная функция f в точке x имеют следующий вероятностный смысл: $f(x)$ есть плотность вероятности в том смысле, что

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}.$$

Предел при $\Delta x \to 0$ средней плотности вероятности случайной величины – это п л о т н о с т ь распределения вероятностей.

Можно говорить, что $P(x < X \leq x + \Delta x)$ есть вероятность того, что случайная величина X примет значение в полуоткрытом интервале $(x;\ x + \Delta x]$ малой длины Δx. Данное свойство может быть сформулировано иначе: вероятность $P(x < X \leq x + \Delta x)$ с точностью до бесконечно малой более высокого порядка малости, чем Δx, равна $f(x)\Delta x$.

Дифференциальная функция обладает следующими свойствами:

- $f(x) \geq 0$ всюду в области определения $f(x)$.
- $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Обратите внимание, что дифференциальная функция распределения есть величина р а з м е р н а я .

По аналогии с измерением длины объекта: отношение вероятности того, что случайная величина X примет значение из интервала $(x;\ x + \Delta x)$, к длине этого интервала Δx – это средняя плотность вероятности случайной величины на этом интервале.

Пример 1.1 Для выборки, заданной в виде вариационного ряда (исходные данные представлены в диапазоне B5:D11 рис. 1.2, построить функцию распределения в дифференциальной и интегральной формах.

Эмпирическую функцию распределения $f(y)$ через непрерывную переменную y обычно представляют в виде кусочно-постоянной зависимости; для вариационного ряда $\{(X_i;\ n_i)\}\ i = 1,2,\dots,n$ определяют как

$$f_j(y) = \begin{cases} 0, & \text{если } y \le X_1, \\ \dfrac{n_i}{\sum n_i}, & X_i < y \le X_{i+1}, \\ 0 & \text{при } y > X_n. \end{cases}$$

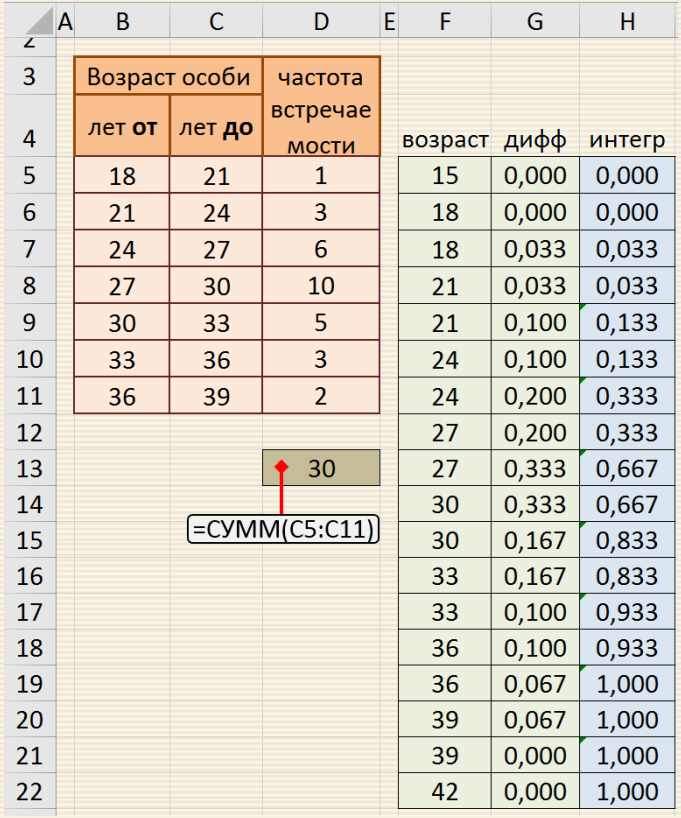


Рис. 1.2. Подготовка данных для построения функций распределения

Исходя из этого рассчитываются (рис. 1.2) необходимые значения интервального "дифференциального" ряда $f_j(y)$ (диапазон ячеек G5:G22). Кумулятивная функция $F_k(y)$ формируется (диапазон ячеек H5:H22) суммированием в соответствии с правилами численного интегрирования по всем "предыдущим" отрезкам $i = 0,1,2, \ldots$ (рис. 1.2). Используя команды MS Excel ВСТАВКА-ДИАГРАММЫ-ТОЧЕЧНАЯ ломаной линией строится графики дифференциальной (рис. 1.3) и интегральной (рис. 1.4) функций распределения.

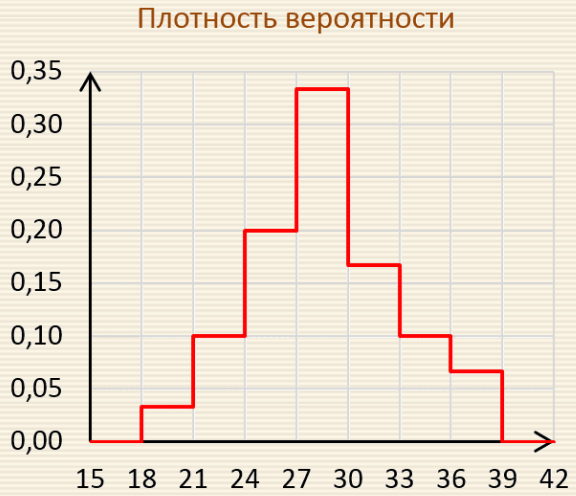


Рис. 1.3. Дифференциальная функция распределения

Рис. 1.4. Интегральная функция распределения

Другой характеристикой распределения является таблица значений (для дискретных распределений) или плотность (для абсолютных непрерывных). Эмпирическим (выборочным) аналогом таблицы или плотности является так называемая гистограмма (рис. 1.5).

Гистограмма – это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

Гистограмма – это столбчатая диаграмма, которая показывает частоту повторяемости значений.

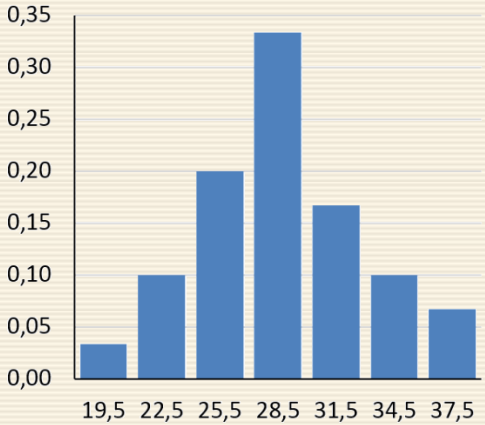


Рис. 1.5. Гистограмма

1.4. Статистические гипотезы и критерии

Типовой задачей анализа данных в исследованиях является установление совпадений или различий характеристик выборок. Статистическая гипотеза – некоторое предположение о законе распределения случайной величины или о параметрах этого закона в рамках данной выборки, которые можно проверить, опираясь на результаты наблюдений в этой (случайной) выборке. Для этого формулируются гипотезы:

– гипотеза об отсутствии различий (так называемая нулевая гипотеза);

– гипотеза о значимости различий (так называемая альтернативная гипотеза).

Пример статистической гипотезы: "генеральная совокупность распределена по нормальному закону", "различие между дисперсиями двух выборок незначимо", "различие между среднеарифметическими значениями двух выборок незначимо" и т.д.

> Нулевая гипотеза H_0 предполагает, что между рассматриваемыми показателями достоверного различия нет, т.е., в частности, обе анализируемые группы составляют однородный материал, одну совокупность (различий "нуль").

Для принятия решений о том, какую из гипотез (нулевую или альтернативную) следует принять, используют решающие правила – статистические критерии.

При этом на основании информации о результатах наблюдений вычисляется число, называемое эмпирическим значением критерия. Это число сравнивается с известным (например, заданным таблично) эталонным числом, называемым критическим значением критерия.

Критические значения приводятся, в общем случае, для нескольких уровней значимости α. Уровнем значимости называется вероятность ошибки, заключающейся в отклонении (не принятии) нулевой гипотезы, то есть вероятность того, что различия сочтены существенными, а они на самом деле случайны.

> Уровень значимости – это вероятность ошибочного отклонения (отвержения) гипотезы, в то время как она на самом деле верна (обычно речь идет об отклонении нулевой гипотезы).
> Уровень значимости – это вероятность допустимой ошибки в формулируемом утверждении или выводе.

Решение о том, можно ли считать высказывание H_0 справедливым для генеральной совокупности, принимается по выборочным данным, т.е. по ограниченному объему информации. Следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов:

> • ошибка первого рода совершается при отклонении гипотезы H_0 (т.е. принимается альтернативная H_1), тогда как на самом деле гипотеза H_0 верна; вероятность такой ошибки обозначается $P(H_1/H_0)$;
>
> • ошибка второго рода совершается при принятии гипотезы H_0, тогда как на самом деле высказывание H_0 неверно и следовало бы принять гипотезу H_1; вероятность ошибки второго рода обычно обозначается как $\beta = P(H_1/H_0)$.

Уровень значимости α определяет ошибку первого рода, т.е. $\alpha = P(H_1/H_0)$. Поэтому вероятность α задается малым числом, поскольку это вероятность ошибочного высказывания.

При этом обычно используются стандартные значения: 0.05; 0.01; 0.005. Например, α = 0.05 означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объема, то в среднем в 5 случаях из 100 совершим ошибку первого рода. В прикладных исследованиях обычно ограничиваются значением 0.05, то есть, грубо говоря, допускается не более чем 5% возможность ошибки.

> Таким образом, если полученное исследователем эмпирическое значение критерия оказывается меньше или равно критическому, то принимается нулевая гипотеза – считается, что на заданном уровне значимости (то есть при том значении α, для которого рассчитано критическое значение критерия) исследуемые характеристики выборок совпадают.

В противном случае, если эмпирическое значение критерия оказывается строго больше критического, то нулевая гипотеза отвергается и принимается альтернативная гипотеза – исследуемые характеристики выборок считаются различными с достоверностью различий 1-α. Например, если α = 0.05 и принята альтернативная гипотеза, то достоверность различий равна 0.95 или 95%.

Мощность критерия — вероятность отклонения основной (или нулевой) гипотезы при проверке статистических гипотез в случае, когда конкурирующая (или альтернативная) гипотеза верна. Чем выше мощность статистического теста, тем меньше вероятность совершить ошибку второго рода.

Другими словами, чем меньше эмпирическое значение критерия (чем левее оно находится от критического значения), тем больше степень совпадения характеристик сравниваемых объектов. И наоборот, чем больше эмпирическое значение критерия (чем правее оно находится от критического значения), тем сильнее различаются характеристики сравниваемых объектов.

Критерий однородности – критерий проверки гипотез о том, что две (или более) выборки взяты из одного распределения вероятностей (из одной и той же генеральной совокупности).

Понятие "однородность", то есть "отсутствие различия", может быть формализовано в терминах вероятностной модели различными способами. Наивысшая степень однородности (**абсолютная однородность**) достигается, если обе выборки, характеризующиеся распределениями $F(x)$ и $G(x)$, взяты из одной и той же генеральной совокупности, т.е. справедлива следующая нулевая гипотеза

$$H_0\colon\ F(x) = G(x) \text{ при любом } x.$$

Отсутствие абсолютной однородности означает, что верна альтернативная гипотеза, согласно которой

$$H_1\colon F(x_0) \neq G(x_0)$$

хотя бы при одном значении аргумента x_0. Если гипотеза H_0 принята, то, в частности, выборки можно объединить в одну, если нет – то нельзя.

В некоторых случаях целесообразно проверять не совпадение функций распределения, а лишь совпадение некоторых характеристик случайных величин – математических ожиданий, медиан, дисперсий, коэффициентов вариации и т.д. Дополнительный материал о понятии однородности дан в приложении П1.

Дополнение. *P*-значение, *P*-уровень (анг. *P-value*) – величина, используемая при тестировании статистических гипотез. Фактически это вероятность ошибки при отклонении нулевой гипотезы (ошибки первого рода). Проверка гипотез с помощью *P*-значения является альтернативой классической процедуре проверки через критическое значение распределения.

Обычно *P*-значение равно вероятности того, что случайная величина с данным распределением (распределением тестовой статистики при нулевой гипотезе) примет значение, не меньшее, чем фактическое значение тестовой статистики.

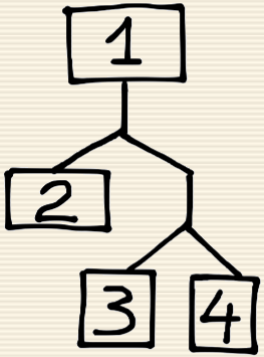
> *P*-уровень: рассчитанная в ходе статистического теста вероятность ошибочного отклонения нулевой гипотезы.

Для принятия решения о том, необходимо ли отклонить нулевую гипотезу по результатам статистического теста, значение *P* сравнивают с принятым исследователем критическим (пороговым) уровнем значимости (α-level).

> Правило принятия нулевой гипотезы: Если *P*-уровень меньше уровня значимости (α-level), то нулевая гипотеза отклоняется. В обратном случае говорят, что нулевая гипотеза не противоречит данным.

Чем меньше *P*-уровень, тем более тестовая статистика называется значимой. Чем меньше *P*-уровень, тем сильнее основания отвергнуть нулевую гипотезу. Таким образом, *P*-уровень находится в убывающей зависимости от надёжности результата. Необходимо отметить, что использование *p*-значений для проверки нулевых гипотез подвергается критике со стороны многих специалистов: их использование нередко приводят к ошибкам первого рода (false positive). В частности, журнал BASP в 2015 году запретил публикацию статей, в которых используются *P*-значения. Редакторы журнала объяснили это тем, что сделать исследование, в котором получено $P < 0.05$ не очень сложно, и такие низкие значения *p* слишком часто становятся оправданием для низкопробных исследований.

Проверка статистической гипотезы осуществляется с помощью статистического критерия в соответствии со следующим алгоритмом:

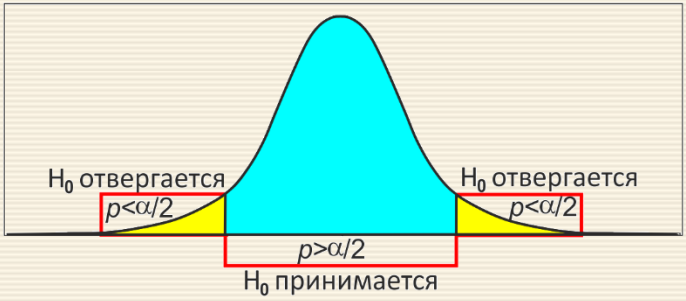


1. Формулировка гипотезы. Гипотеза формулируется в терминах различия величин. Например, есть случайная величина x и константа a. Они не равны (арифметически), но нужно установить значимость статистического различия между ними.

Существует два типа критериев:

- двухсторонний критерий вида: $x \neq a$;
- односторонний критерий вида: $x < a$ или $x > a$.

Необходимо отметить, что знаки больше, меньше и равно здесь используются не в арифметическом, а в "статистическом" смысле. Их необходимо понимать как"значимо больше", "значимо меньше", "различие незначимо".



2. Задаются уровнем значимости α. Вероятность α есть вероятность ошибочного заключения; обычно используют некоторые "стандартные" значения: 0.05; 0.01; 0.005; 0.001.

Уровень значимости α – это вероятность ошибки первого рода, т.е. вероятность того, что будет принята гипотеза H_1, если на самом деле для генеральной совокупности верна гипотеза H_0.

При проверке любой статистической гипотезы возможны следующие основные варианты:

- гипотеза H_0 подтверждается;
- гипотеза H_0 отвергается.

3. Установка закона распределения. На данном этапе устанавливается или постулируется закон распределения для так называемых параметрических критериев. Соответствие нормальному распределению устанавливается, например, приведенными в разделе 3 способами. В случаях невозможности использования параметрических применяются критерии, которые не зависят от вида распределения – так называемые непараметрические критерии.

4. Вычисление тестовой статистики. Определяется величину специально составленной выборочной характеристики – статистического критерия (например, некого критерия K). В общем случае статистическим критерием называют однозначно определенное правило, устанавливающее условия, в рамках которых проверяемую гипотезу H_0 следует либо отвергнуть, либо принять.

По элементам конкретной выборки (выборок), полученных в результате наблюдений (эксперимента), подсчитывается числовое значение критерия $K_{\text{эмп}}$.

Например, при сравнении двух дисперсий D по критерию Фишера тестовая статистика вычисляется по формуле $F = D_{\max}/D_{\min}$.

5. Расчётное значения критерия позволяют судить о расхождении выборки с нулевой гипотезой: анализируется попадание критерия в область существенного расхождении данных с гипотезой H_0 (так называемую критическую область, характеризуемую значением $K_{\text{крит}}$).

Критическая область выбирается так, чтобы вероятность попадания в нее была минимальной (равной α), если верна нулевая гипотеза H_0, и максимальной в противоположном случае.

Наиболее часто подтверждение или отрицание нулевой гипотезы проводится сравнением с табличным (критическим) значением: $K_{\text{эмп}} \leq K_{\text{крит}}$ или $K_{\text{эмп}} \geq K_{\text{крит}}$. Так, например, при сравнении двух дисперсий по критерию Фишера тестовая статистика F сравнивается с так называемым табличным "критическим" значением $F_{\text{крит}}$, которое зависит от уровня значимости α и числа степеней свободы выборок df_1 и df_2.

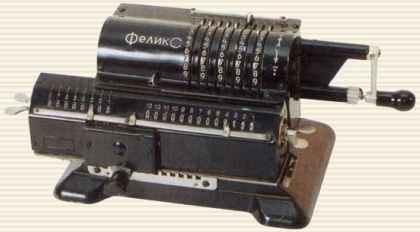
6. Вывод. На основании сравнения делается вывод о том, принимается ли гипотеза (например, значимо ли различие между средними, дисперсиями и т.д.).

Необходимо отметить, что даже если гипотеза H_0 принимается, это вовсе не означает, что H_0 является единственно подходящей гипотезой. Подтверждение гипотезы означает, что расхождение между выборочными данными и нулевой гипотезой невелико, то есть H_0 не противоречит результатам наблюдений.

1.5. Описательная статистика

Эмпирические (опытные, экспериментальные) выборки (совокупности) состоят из отдельных вариант (элементов), которые объединены общностью некоторых свойств (признаков, переменных). Выборки могут быть получены в результате медико-биологического или технического эксперимента, научного опыта, социологического опроса и т.п. Источник появления выборок для статистического анализа значения не имеет.

Единственное требование к анализируемым данным определяется используемыми методами расчета. Они применимы только к количественным выборкам, т.е. к таким эмпирическим выборкам, варианты которых измерены в количественной шкале. Если варианты выборок измерены в порядковой или номинальной шкале, следует применять иные методы расчета описательной статистики.



Количество вариант совокупности в источниках называют по-разному. Так, если речь идет об эмпирической выборке, количество ее элементов может называться численностью, объемом, величиной или размером. Термин "размерность" употреблять в значении "численность" не следует, т.к. он зарезервирован для описания так называемых многомерных совокупностей. Традиционными в отечественной статистической литературе являются термины "выборка", "варианта" и "численность", поэтому по возможности следует придерживаться их.

Чтобы по выборочным данным можно было судить о свойствах генеральной совокупности, выборка должна быть отобрана случайно. Используется два способа формирования выборки:

- повторный отбор, когда каждый элемент, случайно отобранный и обследованный, возвращается в генеральную совокупность и, теоретически, может быть отобран повторно. Например, если в течение некоторого времени на занятии каждый день опрашивают студента, пришедшего последним.
- бесповторный отбор, когда отобранный элемент не возвращается в общую совокупность. При этом если объем генеральной совокупности велик, любой случайный выбор считают бесповторным.

> Точечная оценка какого-либо параметра генеральной совокупности определяется его значением (числом).

Основные параметры распределения признака в генеральной совокупности есть

- выборочное среднее,
- исправленная выборочная дисперсия,
- доля признака в выборке, среднее отклонение,
- мода и медиана.

Как правило, рассчитываются следующие выборочные точечные и интервальные статистические показатели описательной статистики:

- показатели положения: среднее значение и его стандартная ошибка, доверительный интервал, медиана;
- показатели разброса (рассеяния): стандартное отклонение, среднее отклонение, размах, коэффициент вариации, межквартильный размах;
- показатели формы распределения: коэффициент асимметрии, эксцесс.

Кроме перечисленных показателей, рассчитывается достаточная численность выборки из анализа заданных и рассчитанных выборочных показателей.

Выборочное среднее значение, среднеарифметическое выборочное (mean) – наиболее часто применяемый статистический показатель, характеризующий середину эмпирической совокупности. Выборочная средняя – это несмещенная оценка математического ожидания, так как средняя из выборочных средних стремится к своему теоретическому значению по генеральной совокупности.

Вычисление среднего значения выборки производится по формуле:

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

где n – численность выборки,

x_i – значения вариант выборки.

Функция MS Excel: СРЗНАЧ(данные).

Если все варианты разнесены по M классам, каждый из которых характеризуется определенным значением вариант и частотой f_j, то среднее выборки можно вычислить по следующей формуле

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{M} f_j x_j$$

где f_j – частота класса,

x_j – (среднее) значение класса.

Расчет среднего значения, как и многих других статистических показателей, имеет смысл только для качественно однородных групп: "мухи отдельно, котлеты отдельно". Строго говоря, вычисление среднего значения законно только для таких эмпирических выборок, которые не противоречат гипотезе о нормальности статистического распределения.

Тесно с выборочным средним значением связано понятие математического ожидания, которое можно рассматривать в некотором роде как предельное значение выборочного среднего.

Математическим ожиданием непрерывной случайной величины X с плотностью распределения $p(x)$ принята величина (интеграл Лебега-Стилтьеса)

$$M[X] = \int_{-\infty}^{+\infty} xp(x)dx$$

В англоязычной литературе математическое ожидание часто обозначается через $E[X]$(скорее всего от англ. Expected value или нем. Erwartungswert), в русской – $M[X]$ (возможно, от англ. Mean value или нем. Mittelwert, а возможно и от "Математическое ожидание"). В статистике достаточно часто используют обозначение(читается "мю").

Основные свойства математического ожидания следующие.

Свойство 1. Если имеются переменные X, Y, Z, то математическое ожидание их суммы равно сумме математических ожиданий. Работает такое равенство $M(X+Y) = M(X) + M(Y)$.

Свойство 2. Если переменную (т.е. каждое значение переменной) умножить на постоянную величину (a), то математическое ожидание такой величины будет равно произведению матожидания переменной и этой константы. Формально выражаясь, имеет место равенство $M(aX) = aM(X)$.

Свойство 3. Математическое ожидание постоянной величины a есть сама эта величина: $M(a) = a$.

Свойство 4. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий $M(X \cdot Y) = M(X) \cdot M(Y)$.

Основным статистическим показателем, характеризующим разброс выборки, является дисперсия, вычисляемая по формуле:

$$D = \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$$

В иностранной литературе дисперсию иногда называют вариансой.

Представленная формула вычисляет так называемую несмещенную (исправленную) выборочную оценку дисперсии. Формула смещенной оценки отличается от показанной формулы делителем не (n-1), а n. Предполагается, что формула смещенной оценки должна использоваться, если известна вся генеральная совокупность, что на самом деле, конечно, встречается редко. Более подробно данные понятия обсуждаются в приложении П1.

Дисперсия для генеральной совокупности определяется смещенной оценкой $D = \frac{1}{n}\sum_i (x_i - \overline{x})^2$

В MS Excel для вычисления выборочной и генеральной дисперсий используются функции ДИСП.В (данные) и ДИСП.Г (данные).

Если все варианты разнесены по M классам, каждый из которых характеризуется определенным значением вариант и частотой f_j, то дисперсию можно вычислить по формуле $D = \frac{1}{n-1}\sum_{j=1}^{M} f_j\left(x_j - \overline{x}\right)^2$

Вычисление точечных оценок для вариационного ряда приведено в приложении П2.2.

Стандартным выборочным отклонением S (средним квадратическим отклонением, средним квадратичным отклонением) называют корень квадратный из дисперсии $S = \sqrt{D}$.

Стандартное отклонение – это мера того, насколько широко разбросаны точки данных относительно их среднего.

Функции MS Excel: для выборочных данных: D: ДИСП.В (данные); для S: СТАНДОТКЛОН.В (данные)
для генеральной выборки: D: ДИСП.Г (данные); для S: СТАНДОТКЛОН.Г (данные)

Доля какого-либо признака, характеризующего определенные элементы выборки, определяется отношением числа m этих элементов к объему выборки n $\omega = \frac{m}{n}$.

В генеральной совокупности доля признака определяется вероятностью появления указанного свойства. Для вычисления доли элементов, описываемых каким-либо признаков (условием), в MS Excel можно использовать функции СЧЁТЕСЛИ(…) и СЧЁТ(…).

Мода (mode), модальный класс – класс, обладающий наибольшей частотой. Значение, которое принимает данный класс, называют модой. Функция MS Excel: МОДА(данные).

Медиана (median) – это число, которое является серединой множества чисел выборки, то есть половина чисел имеют значения большие, чем медиана, а половина чисел имеют значения меньшие, чем медиана.

Для вычисления медианы количественной выборки численностью n сначала строится интервальный вариационный ряд, т.е. исходная выборка упорядочивается по возрастанию.

Для нечетного n медианой является варианта полученного интервального вариационного ряда, имеющая порядковый номер (n+1)/2. Для четного n медиана равна среднему значению двух средних вариант. Некоторые исследователи предпочитают медиану среднему значению, считая ее более точной оценкой меры положения выборки. Функция MS Excel: МЕДИАНА(данные).

Выборочное среднее отклонение (выборочная оценка среднего отклонения), подобно стандартному отклонению, характеризует разброс эмпирической выборки относительно среднего значения и вычисляется по формуле

$$\overline{d} = \frac{1}{n}\sum_i |x_i - \overline{x}|,$$

n – численность выборки,
x_i – значения вариант выборки,
$\overline{x}$ – выборочное среднее значение.

Среднее отклонение отражает так называемый модульный подход к вычислению меры отклонения между величинами в противоположность тому, что стандартное отклонение отражает квадратический подход. Функция MS Excel: СРОТКЛ(данные).

Между признаками выборочной совокупности и соответствующими признаками генеральной совокупности, как правило, существует некоторое расхождение, которое называется ошибкой статистического наблюдения:

- Ошибки регистрации, или технические ошибки, связаны с недостаточной квалификацией наблюдателей, неточностью подсчетов, несовершенством приборов и т.п.
- Под ошибкой репрезентативности понимают расхождение между выборочной характеристикой и разыскиваемой (истинной) характеристикой генеральной совокупности. Здесь возможны систематические ошибки, связанные с нарушением установленных правил отбора, и случайные ошибки, которые объясняются недостаточно равномерным представлением в выборочной совокупности различных категорий единиц генеральной совокупности.

Случайная ошибка с увеличением объема выборки уменьшается. В случае большой выборки (n>100) определение предельной ошибки для среднего и доли основано на центральной предельной теореме, вследствие которой среднее и доля при большом числе измерений имеют распределения близкие к нормальному.

При описании результатов экспериментального исследования в медико-биологических науках стандартную ошибку принято обозначать символом m. Обычно используется понятная большинству исследователей традиционная запись, характеризующая среднее значение и его стандартную ошибку, в виде $\overline{x} \pm m$.

Стандартная ошибка большой выборки для генеральной средней m и генеральной доли m_ω рассчитывается в зависимости от условий отбора в соответствии с нижеследующими соотношениями.

	Повторный отбор (или $n \ll N$, или $N = \infty$)	Бесповторный отбор
Стандартная ошибка среднего standard error of the mean, SEM	$m = \frac{S}{\sqrt{n}}$,	$m = \frac{S}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}$,
Стандартная ошибка доли (вероятности проявления признака)	$m_\omega = \sqrt{\frac{\omega(1-\omega)}{n}}$,	$m_\omega = \sqrt{\frac{\omega(1-\omega)}{n}\left(1 - \frac{n}{N}\right)}$,

N – объем генеральной совокупности.

Ошибка стандартного отклонения вычисляется по формуле: $m_s = \frac{S}{\sqrt{2n}}$.

Замечание. Стандартная ошибка не есть ошибка точности. Смысл m – это, насколько исследователь ошибается, принимая данные выборочной совокупности за генеральную (см. радел П1.1).

В случае малой выборки ($n < 30$) при отсутствии данных о нормальности распределения признака предельная ошибка $m_{\text{пред}}$ для генеральной средней определяется по формуле $m_{\text{пред}} = t_{\alpha,n-1}\frac{S}{\sqrt{n-1}}$,

где $t_{\alpha,n-1}$ – табличное значение критерия Стьюдента для уровня значимости α при числе степеней свободы n-1.

В Excel коэффициент доверия для малой выборки рассчитывается при помощи функции СТЬЮДЕНТ.ОБР.2Х ($\alpha; n-1$).

Необходимо отметить, что причисление выборки к категории "большой" или "малой" достаточно произвольно и зависит от дисперсии выборки; четкой границы между большой и малой выборками нет. Выборка с малыми разбросами может считаться большой, тогда как выборка такого же объема из существенно разнородной совокупности является малой.

При помощи формулы предельной ошибки выборки решают следующие задачи:

- определение доверительного интервала на заданном уровне значимости;
- расчет доверительного интервала с заданной доверительной вероятностью для генеральной доли: $\omega \pm \Delta$.
- определение необходимого объема выборки n для определения доверительного интервала с заданной точностью Δ на заданном уровне значимости. Формулы объема выборки приведены в таблице:

	Повторный отбор (или $n \ll N$, или $N = \infty$)	Бесповторный отбор
Для генеральной средней	$n = \left(\frac{tS}{\Delta}\right)^2$,	$n = \left[\frac{1}{N} + \left(\frac{\Delta}{tS}\right)^2\right]^{-1}$,
Для генеральной доли	$n = \omega(1-\omega)\left(\frac{t}{\Delta}\right)^2$,	$n = \left[\frac{1}{N} + \frac{1}{\omega(1-\omega)}\left(\frac{\Delta}{t}\right)^2\right]^{-1}$.

Следует отметить, что предельную ошибку для среднего большой выборки можно рассчитать в Excel также при помощи функции ДОВЕРИТ(α; S; n), где

α – допустимая вероятность ошибки, т.е. уровень значимости α;

S – генеральное среднеквадратическое отклонение, предполагающееся известным, либо его оценка.

Доверительным (confidence interval, CI) называют интервал, который покрывает неизвестный параметр с заданной надёжностью.

Доверительной вероятностью или надежностью P измерений называется вероятность попадания истинного значения измеряемой величины в данный интервал (выражается в долях или процентах).

Одним из условий построения доверительного интервала является его максимальная узость, т.е. он должен быть насколько это возможно коротким. Отсюда следует, что доверительный интервал должен накрывать максимальные вероятности распределения, а сама оценка должна быть в центре.

Для нормального распределения вероятность отклонения (истинного показателя от оценки) в большую сторону равна вероятности отклонения в меньшую сторону. Следует также отметить, что для несимметричных распределений интервал справа не равен интервалу слева.

Чтобы рассчитать нижнюю и верхнюю границу интервала, требуется знать точный вид распределения, определяющего связь степени отклонения и вероятности накрытия оцениваемого показателя.

В частности, используя известные свойства нормального распределения (см. приложение П3), можно рассчитать доверительные интервалы для среднего арифметического, которые с заданной вероятностью накрывают истинное среднее или математическое ожидание.

Для установления верхней и нижней границы требуется знать параметры нормального распределения. Как правило, они не известны, поэтому используют выборочные оценки среднего арифметического и дисперсии, хотя хорошее приближение проявляется только при больших выборках. Когда выборки малые, рекомендуется использование распределение Стьюдента, хотя это распределение для среднего имеет место только тогда, когда исходные данные имеют достаточно редко встречающееся на практике нормальное распределение. Минимально рекомендуемая достаточность выборки – 30 наблюдений.

Обычно при построении доверительных интервалов для оценки среднего используют только верхний $\alpha/2$ квантиль и не используют нижний $\alpha/2$ квантиль, поскольку стандартное нормальное распределение симметрично относительно среднего, то бишь нуля. Поэтому нужды вычисления нижнего $\alpha/2$ квантиля нет, поскольку он равен значению верхнего со знаком минус. Отметим также, что, несмотря на форму распределения x соответствующая случайная величина $\overline{x}$ распределена приблизительно нормально $\mathbb{N}(\mu, \sigma^2/n)$ в соответствии с центральной предельной теоремой (см. приложение П3). Следовательно, вышеуказанное выражение для доверительного интервала является приближенным и только в случае нормального распределения доверительный интервал становится точным.

При заданном уровне значимости α рассчитать предельное отклонение $\Delta\overline{x}$ от среднеарифметического $\overline{x}$ можно по следующей формуле:

$$\Delta\overline{x} = t\frac{S}{\sqrt{n}},$$

S – стандартное отклонение по выборке (возвращается Excel-функцией СТАНДОТКЛОН.В (данные);

n – объем выборки (возвращается Excel-функцией СЧЕТ (данные).

К о э ф ф и ц и е н т д о в е р и я t, связанный с доверительной вероятностью – вероятностью того, что случайная ошибка репрезентативности на самом деле не превосходит вычисленную предельную ошибку; рассчитывается как обратное значение функции стандартного нормального распределения (для задаваемого уровня значимости α возвращается Excel-функцией НОРМ.СТ.ОБР(1-$\alpha/2$)).

Для расчета доверительного интервала в MS Excel имеется функция ДОВЕРИТ.НОРМ (α, S, n), которая возвращает полуинтервал $\pm\Delta\overline{x}$. Соответственно, нижняя и верхняя граница математического ожидания для генеральной выборки μ – это среднее $\mp$ полученное значение $\mu = \overline{x} \pm \Delta\overline{x}$.

Платой за простоту изложенного подхода является его асимптотичность, т.е. необходимость использования относительно больших выборок. Для "малых" выборок в MS Excel служит функция ДОВЕРИТ.СТЬЮДЕНТ(α, S, n).

Пример 1.2 Для выборки (диапазон B2:H14 на рис. 1.6) на уровне значимости 0.1 рассчитать величину доверительного интервала, считая выборку и большой и малой.

Видно (рис. 1.6), что предельная ошибка, вычисленная по формуле для малой выборки несколько больше, чем по формуле для большой выборки, но в данном случае различие невелико. Таким образом, доверительный интервал для среднего равен 9.0385 ± 1.5696.

Когда элементы выборки расположены упорядоченно от самой маленькой величины переменной x до самой большой величины, то значение $x^{(1)}$, до которого расположен 1% наблюдений (и выше которого, соответственно, расположены 99% наблюдений), называется первым п р о ц е н т и л е м (percentiles).

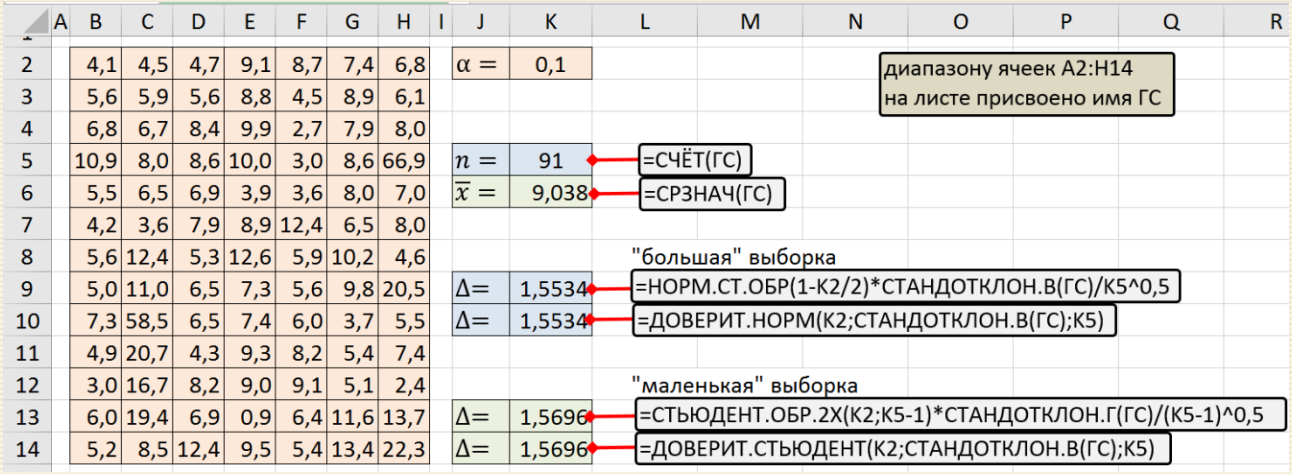


Рис. 1.6. Доверительный интервал Δ для среднего

Величина $x^{(2)}$, до которой находится 2% наблюдений, называется 2-м процентилем, и т.д. Величины $x^{(10)}$, $x^{(20)}$, … которые делят упорядоченный набор значений на 10 равных групп, т.е. 10-й, 20-й, 30-й,..., 90 и процентили, называются д е ц и л я м и . Величины $x^{(25)}$, $x^{(50)}$, $x^{(75)}$, которые делят упорядоченный набор значений на 4 равные группы, т.е. 25-й, 50-й и 75-й процентили, называются к в а р т и л я м и . 50-й процентиль — это медиана.

Часто процентили применяются для описания рассеяния данных, на которые не влияет выброс (аномальное значение), "сглаживая" тем самым экстремальные величины в наблюдениях.

Межквартильный размах (interquartile range), интерквартильный размах. Квартили, а также медиана, обеспечивают разбиение упорядоченной количественной выборки (в виде вариационного ряда) на 4 подмножества равной численности. Вычисление данных показателей производится по правилам, принятым для вычисления медианы.

Межквартильный размах выборки (интерквартильный размах) характеризует степень разброса данных в абсолютных числах. Выборочный межквартильный размах – это разность между верхней и нижней квартилями выборки, иначе 75% и 25% процентилями выборки. Вычисление межквартильного размаха упорядоченной по возрастанию количественной выборки производится по формуле:

$$R_{\mu} = \mu_{3/4} - \mu_{1/4},$$

где $\mu_{3/4}$ – значение верхней квартили выборки,
$\mu_{1/4}$ – значение нижней квартили выборки.

Межквартильный размах является более репрезентативной оценкой разброса значений выборки по сравнению с точечной оценкой стандартного отклонения. Межквартильный размах находит применения в качестве основы одного из методов выявления аномальных наблюдений (выбросов). Критерии исключения промахов (аномальностей, выбросов) приведены в разделе 8. Величина $0.5R_{\mu}$ также используется как характеристика рассеяния и называется семиинтеркваритильной широтой.
Вычисление R_{μ} функциями MS Excel: КВАРТИЛЬ.ВКЛ (данные;3) - КВАРТИЛЬ.ВКЛ (данные;1).

Числа n_i, показывающие, сколько раз встречаются варианты x_i в ряде наблюдений, называются частотами, а отношение их к объему выборки – частостями или относительными частотами (relative frequency) (p_i^*).

Рассчитываются частости по очевидным соотношениям

$$p_i^* = \frac{n_i}{n}, \text{ где } n = \sum n_i .$$

Полигоном частот называют ломаную, отрезки которой соединяют точки с координатами

$$(x_1, n_1), (x_2, n_2), \ \ldots, (x_M, n_M);$$

полигоном частостей – с координатами $(x_1, p_1^*), (x_2, p_2^*), \ldots, (x_M, p_M^*)$.

Варианты *(x_i)* откладываются на оси абсцисс, а частоты и частости – на оси ординат. Как правило, для случайных величин дискретного типа употребляются полигон и ступенчатая кумулятивная кривая, а для непрерывных – гистограмма и ломаная кумулятивная кривая.

Гистограммой (histogram) частот (частостей) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению n_i/h – плотность частоты (p_i^*/h или $n_i/(n \cdot h)$ – плотности частости).

Очевидно, что площадь гистограммы частот равна объему выборки, а площадь гистограммы частостей равна единице.

Размах выборки (размах вариации, амплитуда ряда) характеризует степень разброса данных в абсолютных числах. Выборочный размах – это разность между максимумом и минимумом вариант выборки. Вычисление размаха количественной выборки производится по формуле:

$$R = x_{max} - x_{min}$$

где x_{max} – значение максимальной варианты,
x_{min} – значение минимальной варианты выборки.

Функции MS Excel: МИН(данные); МАКС(данные).

Коэффициент вариации (variation coefficient), относительное среднеквадратичное отклонение. Все показатели, рассмотренные выше, привязаны к масштабу исходных данных и не позволяют получить сопоставительное представление о вариации анализируемой совокупности. Для получения относительной (безразмерной) меры разброса данных используется коэффициент вариации, который является отношением среднеквадратичного отклонения к среднему арифметическому:

$$V = \frac{S}{\overline{x}}.$$

Вычисление V производится делением стандартного отклонения на среднее значение =СТАНДОТКЛОН.Г (данные) / СРЗНАЧ (данные) либо, при необходимости =СТАНДОТКЛОН.В (данные) / СРЗНАЧ (данные) .

Коэффициент вариации чаще всего выражается в процентах, для чего можно установить через "Формат ячеек" (правая кнопка мыши в MS Excel) выбором процентного представления величины.

Коэффициент вариации, в отличие от других показателей разброса значений, используется как весьма информативный индикатор вариации данных. В статистике принято считать, что если коэффициент вариации менее 33%, то совокупность данных является однородной, если более 33%, то – неоднородной. Эта информация может быть полезна для предварительного описания данных и определения возможностей проведения дальнейшего анализа. Кроме того, коэффициент вариации позволяет сравнивать степень разброса различных данных независимо от их масштаба и единиц измерений.

Дисперсия, среднее квадратическое отклонение, среднее линейное отклонение являются мерилом разброса и однородности данных. Однако все они связаны с размерностью и масштабом исходных данных и не дают "независимой" (относительной) характеристики меры разброса. Коэффициент вариации лишен данной особенности.

Отметим, что однородность – понятие относительное и растяжимое. Она не имеет точных границ и критериев. В рассматриваемом отношении под однородными данными следует понимать некоторый уровень их рассеяния, при котором рассчитываемые статистические показатели (например, среднее) будут давать надёжную и качественную характеристику анализируемой совокупности. Более подробное описание затронутых понятий приведено в приложении П1.

В статистических исследованиях широко используется понятие вариационного ряда.

> **Вариационный ряд** (set of order statistic) – последовательность значений заданной выборки $x_1, x_2, \dots, x_m$, расположенных в порядке неубывания $x_1 \le x_2 \le \dots \le x_m$.

k-той порядковой статистикой называется k-е значение в вариационном ряду x_k.

Рангом r_i наблюдения x_i называется его порядковый номер в вариационном ряду.

> **Вариационный ряд** (set of order statistic) – ряд, в котором сопоставлены (по степени возрастания или убывания) варианты и соответствующие им **частоты**.

Напомним, что варианты – отдельные количественные выражения признака. Классическое понимание термина "варианта" предполагает, что вариантой называется каждое уникальное значение признака, без учета количества повторов.

Частота – число, показывающее, сколько раз повторяется варианта. Сумма всех частот (которая, разумеется, равна числу всех исследуемых) обычно обозначается как n.

Например, в вариационном ряду показателей систолического артериального давления, измеренного у десяти пациентов: 110, 120, 120, 130, 130, 130, 140, 140, 160, 160, 170;
вариантами являются только 6 значений: 110, 120, 130, 140, 160, 170.

В приведенном примере вариационный ряд будет выглядеть в виде следующей таблицы

Давление	Частота
110	1
120	2
130	3
140	2
160	2
170	1

Анализ статистических характеристик для вариационного ряда дан в разделе П2.2.

Различают следующие виды вариационных рядов:

- простой – это ряд, в котором каждая варианта встречается единожды (все частоты равны 1);
- взвешенный – ряд, в котором одна или несколько вариант встречаются неоднократно.

Взвешенный вариационный ряд служит для описания больших массивов чисел, именно в этой форме изначально представляются собранные данные большинства биомедицинских исследований.

Для того, чтобы охарактеризовать вариационный ряд, специальным образом рассчитываются показатели – средние величины, вариабельность, репрезентативность выборочных данных и пр.

Пример 1.3 Для выборки, заданной вариационным рядом (столбец J – значения; K – частоты) построить полигон частот (рис. 1.7).

Для построения полигона выделяются данные и выполняются команды ВСТАВКА→ Диаграммы→Точечная.

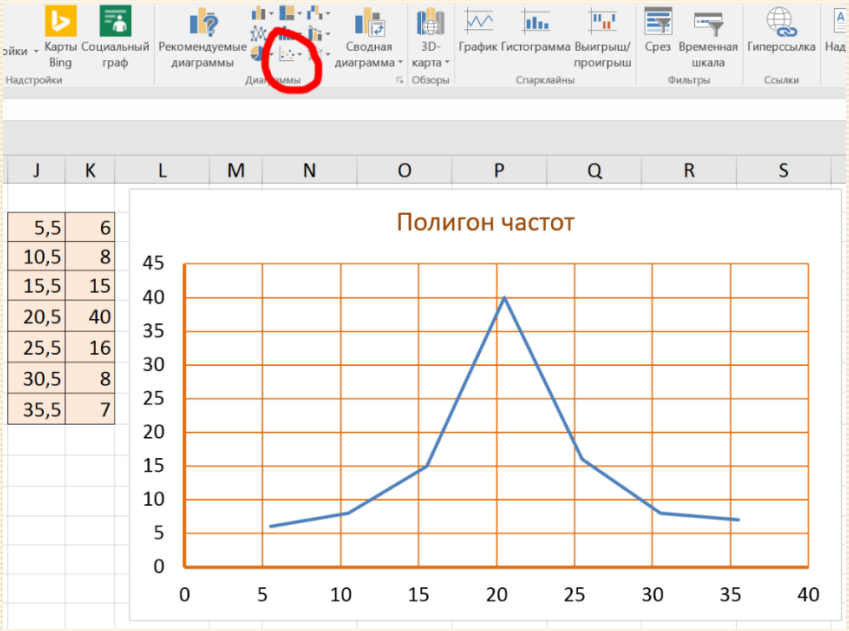


Рис. 1.7. Скриншот экрана полигона частот

2. Дисперсионный анализ. Однородность

Дисперсионный анализ (Analysis of Variance, ANOVA) включает в себя проверку гипотез, связанных с оценкой выборочной дисперсии. Можно выделить три основных вида гипотез:

1. $D_1 = D_2$ (значимо ли различие между двумя дисперсиями?)
2. $D_1 > D_2$ (одна дисперсия значимо больше другой?)
3. $D_1 = D_2 = \cdots = D_K$ (значимо ли различие между несколькими (K) дисперсиями?)

Дисперсия вычисляется из случайных величин, и поэтому сама также является случайной величиной. Напомним, что дисперсии, в отличие от средних, может подчиняться подчиняются распределению χ^2.

2.1. Критерий Фишера

Первые две гипотезы дисперсионного анализа проверяются с помощью критерия Фишера $F_{\text{эмп}}$. Причем первая гипотеза – с помощью двустороннего критерия, а вторая – с помощью одностороннего. Строго говоря, эти критерии не равны, но в общем случае разницей можно пренебречь. Проверка производится по следующей формуле:

$$F_{\text{эмп}} = \frac{D_{\max}}{D_{\min}} \leqslant F_{\text{крит}}\left(\alpha,\ df(\text{для } D_{\max}),\ df(\text{для } D_{\min})\right).$$

Критическое значение (правостороннее) $F_{\text{крит}}$ в MS Excel определяется функцией F.ОБР.ПХ(...) от трех параметров: α – уровень значимости;

$df\,(\text{для } D_{\max})$ – число степеней свободы для выборки с максимальной дисперсией;

$df\,(\text{для } D_{\min})$ – число степеней свободы для выборки с меньшей дисперсией.

Пример 2.1 Задача: сравнить дисперсии двух выборок, данные которых даны определяются диапазонами ячеек B3:D9 и F3:G9 (рис. 2.1) на уровне значимости 0.01.

Скриншот решения дан на этом же рисунке.

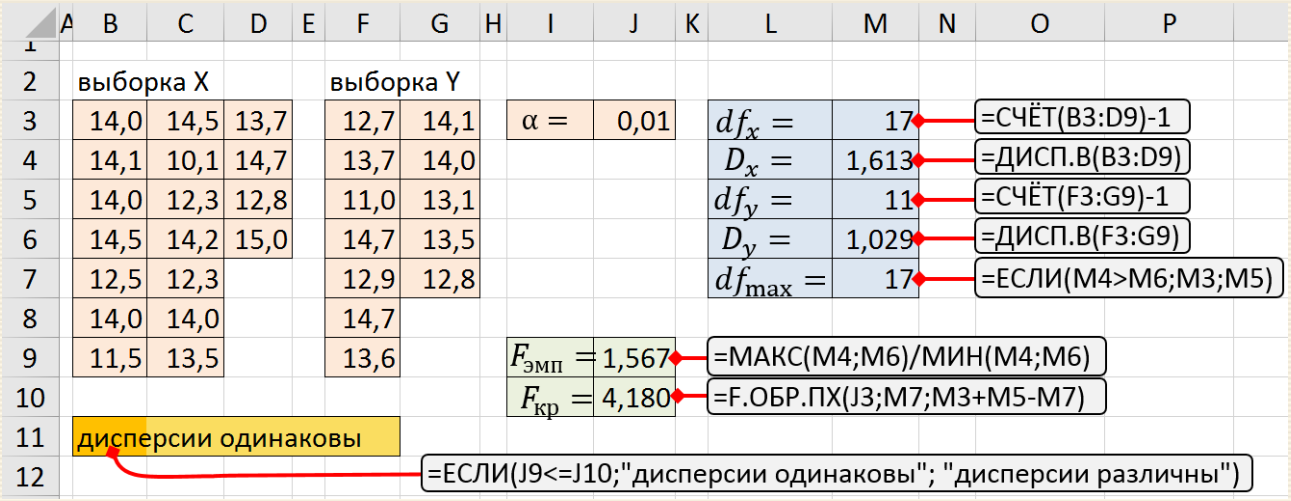


Рис. 2.1 Критерий Фишера равенства дисперсий двух выборок

2.2. Критерий Кохрена

С третьей гипотезой (сравнение нескольких дисперсий или проверка однородности дисперсий) дело обстоит сложнее. Попарно сравнивать дисперсии некорректно (например, какой вывод нужно сделать, если из трех дисперсий дисперсии 1, 2 и 1, 3 различаются незначимо, а дисперсии 2, 3 – значимо?). В дисперсионном анализе для сравнения нескольких дисперсий существует ряд специальных критериев, наиболее простым из которых является критерий Кохрена, предполагающий независимость распределенных по нормальному закону данных. Количество наблюдений m одинаково в каждой выборке.

Проверка однородности дисперсий по критерию Кохрена (Cochran's test) для одинакового числа m параллельных опытов включает вычисление доли максимальной дисперсии среди всех дисперсий:

$$G = \frac{D_{\max}}{\sum_i D_i},$$

которая затем сравнивается с критическим значением $G_{\text{крит}}(\alpha, k, df)$, где df – число степеней свободы каждой дисперсии (должно быть одинаковым у всех дисперсий и равно m-1), k – число дисперсий (серий опытов), α – уровень значимости.

William Gemmell Cochran

При использовании критерия Кохрена строится правосторонняя критическая область, критическое значение $G_{\text{крит}}(\alpha, k, df)$ уровня значимости α статистики G вычисляется как квантиль порядка $1-\alpha/k$ бета-распределения с параметрами формы $df/2$ и $df(k-1)/2$. Для этой цели в Excel можно использовать функцию БЕТА.ОБР $(1-\alpha/k;\ df/2;\ df(k-1)/2)$.

Критерий Кохрена чувствителен к случаю, когда ожидается равенство дисперсий за исключением возможно только одной, каковая может быть больше остальных. Достоинства критерия: простота вычислений. Ограничения критерия касаются числа степеней свободы дисперсий; критерий выявляет отклонения только в большую сторону.

Пример 2.2 Задача: сравнить дисперсии шести серий опытов; для каждой серии проведено по пять опытов по уровню значимости α=0.05.

	опыт 1	опыт 2	опыт 3	опыт 4	опыт 5
серия 1	14.0	14.5	13.7	12.7	14.1
серия 2	14.1	10.1	14.7	13.7	14.0
серия 3	14.0	12.3	12.8	11.0	13.1
серия 4	14.5	14.2	15.0	14.7	13.5
серия 5	12.5	12.3	11.5	12.9	12.8
серия 6	14.0	14.0	13.5	14.7	13.6

На рис. 2.2 даны результаты проверки гипотезы H_0: $D_1 = D_2 = D_3 = D_4 = D_5 = D_6$ равенства шести выборок, выполненной с помощью критерия Кохрена.

Сравнивая расчетное значение G =0.570 статистики G с ее критическим значением $G_{(0.05)} = 0.480$, приходим к выводу, что гипотеза о равенстве дисперсий противоречит реальным данным наблюдений и поэтому отвергается на уровне значимости $\alpha = 0.05$.

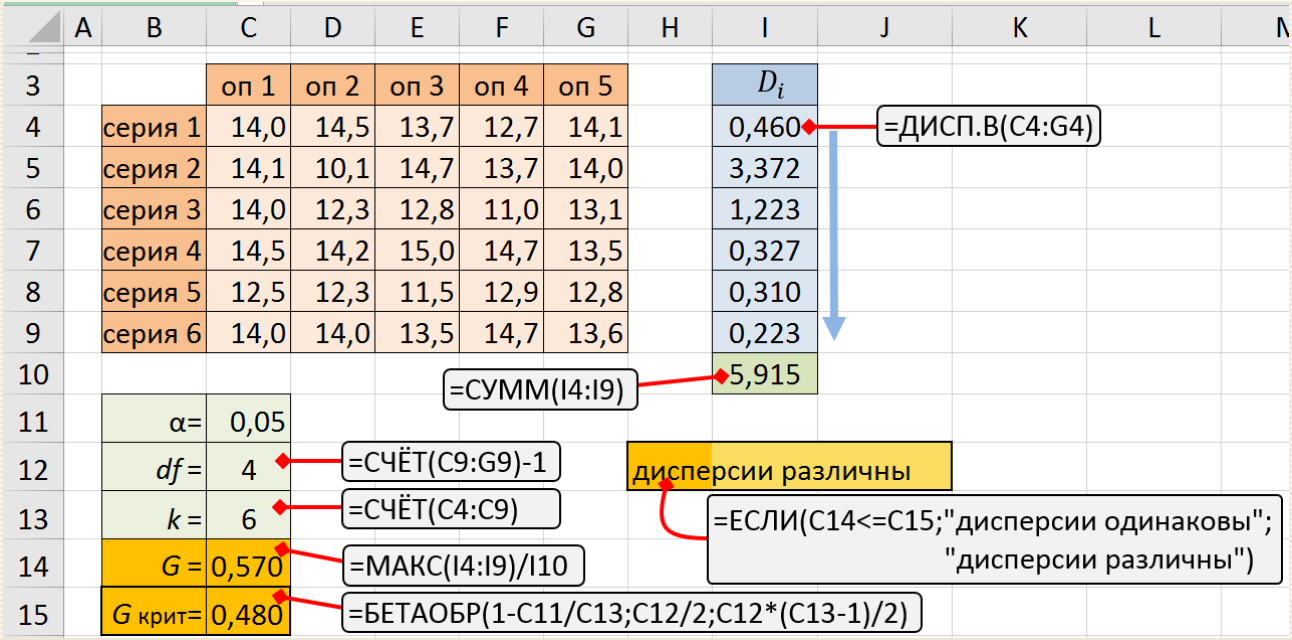


Рис. 2.2 Проверка гипотезы равенства дисперсий шести выборок (серий опытов)

2.3. Критерий Бартлетта

Критерий Бартлетта (Bartlett's test) оценки равенства дисперсий для разного числа параллельных экспериментов. В основе критерия лежит статистика

$$T = \frac{(M-k)\cdot \ln D - \sum_{i=1}^{k} \nu_i \ln D_i}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{M-k}\right)},$$

степени своды i-й выборки, объем которой равен m_i;	$\nu_i = m_i - 1$,
суммарный объем всех k выборок	$M = \sum_{i=1}^{k} m_i$,
выборочная дисперсия i-й выборки x_{ij} – j-й элемент i-й выборки	$D_i = \frac{1}{\nu_i}\sum_{i=1}^{m_i}\left(x_{ij} - \overline{x_i^*}\right)^2$,
выборочное среднее i-й выборки	$\overline{x_i^*} = \frac{1}{m_i}\sum_{i=1}^{m_i} x_{ij}$,
взвешенное среднее k выборочных дисперсий	$D = \frac{1}{M-k}\sum_{i=1}^{m_i} \nu_i D_i$.

где k – число сравниваемых дисперсий (число серий, выборок);

Если проверяемая гипотеза верна (то есть если дисперсии исследуемых случайных величин равны) и все m_i больше 3, то статистика T имеет распределение, близкое к χ^2-распределению с $(k-1)$ степенями свободы.

При проверке гипотезы о равенстве дисперсий с помощью критерия Бартлетта используется правосторонняя критическая область $[\chi^2(\alpha; k-1), \infty)$, где $\chi^2(\alpha; k-1)$ – критическое значение порядка α хи-квадрат распределения с $(k-1)$ степенями свободы. При этом значимость $\alpha^* = P\left(\chi^2_{k-1} \geq T\right)$.

Критерий Бартлетта очень чувствителен к отклонениям распределений исследуемых случайных величин от нормального распределения. Значимость статистики T может указывать не на отсутствие однородности дисперсий, а просто на отклонение от нормальности.

Если гипотеза о равенстве дисперсий принята, то в качестве оценки дисперсии следует использовать взвешенное среднее D выборочных дисперсий.

Maurice Stevenson Bartlett

Пример 2.3 В диапазоне D4:J9 рис. 2.3 приведены результаты измерений шести опытов (по каждому из которых сделано от 5 до 7 измерений). Используя эти данные, необходимо проверить на уровне значимости α = 0.05 гипотезу о равенстве дисперсий шести серий опытов.

На рис. 2.3 даны результаты проверки гипотезы H_0: $D_1 = D_2 = D_3 = D_4 = D_5 = D_6$ равенства шести выборок, выполненной с помощью критерия Бартлетта.

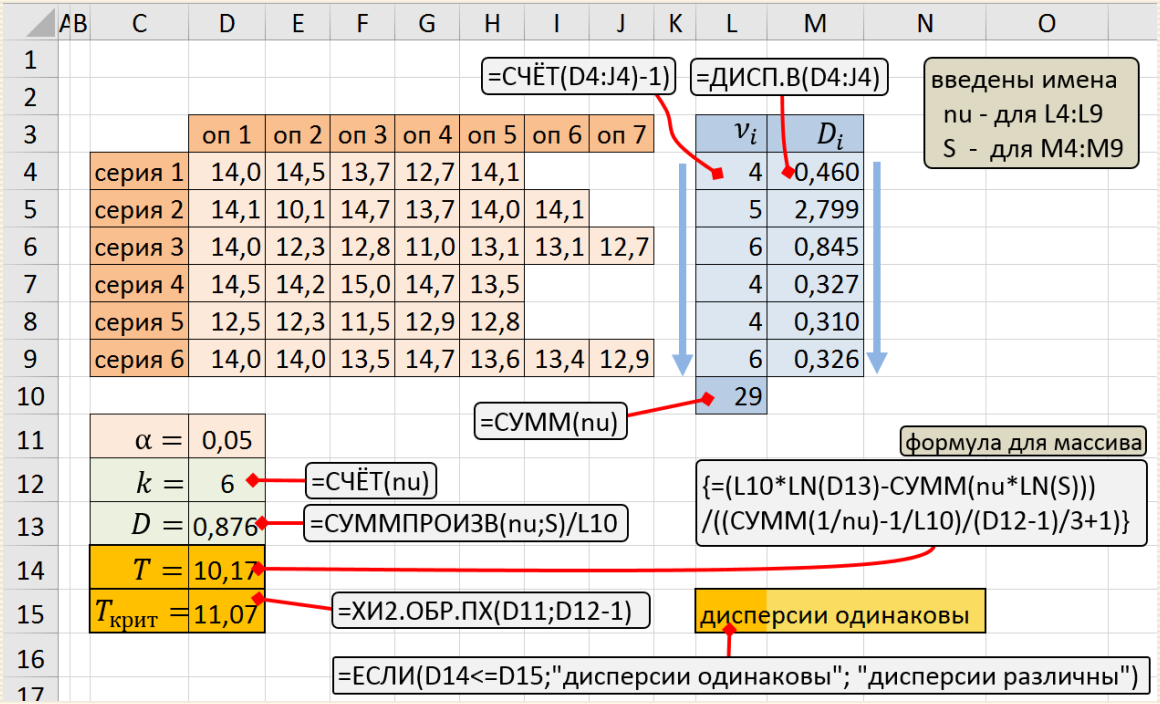


Рис. 2.3 Проверка гипотезы равенства дисперсий серий опытов по критерию Бартлетта

Итоговые результаты вычислений находятся в ячейках D14:D15. Сравнивая расчетное значение критерия T = 10.17 статистики T с ее критическим значением $T_{крит} = \chi^2(0.05;\ 5) = 11.07$, можно сделать вывод, что гипотеза о равенстве дисперсий не противоречит реальным данным наблюдений и посему она принимается на уровне значимости α = 0.05.

2.4. Критерий знаков

При сравнении двух парных выборок достаточно удобен критерий знаков (sign test), используемый именно для связанных выборок, то есть таких, когда элементы выборок соответствуют одному и тому же объекту, но измерения сделаны в разные моменты (например, до и после воздействия). Чаще всего сравниваются параметры до и после эксперимента у одного и того же животного.

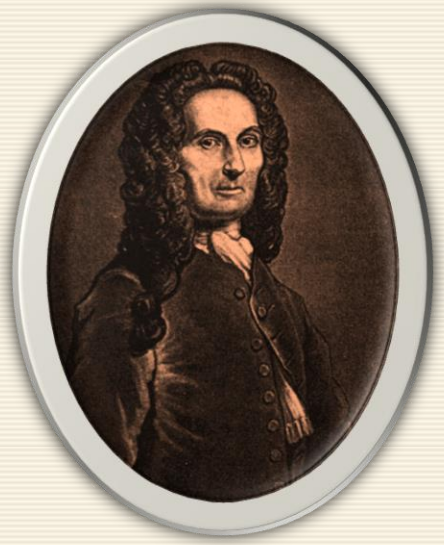
В практических приложениях наряду с задачей о соответствии выборочных наблюдений предполагаемому закону распределения может возникнуть задача о проверке соответствия распределений двух генеральных совокупностей по результатам выборочных наблюдений.

Пусть $x_1, \dots, x_n$ – выборка объёма n наблюдений случайной величины X, имеющей неизвестное распределение $F_X(x)$, $y_1, \dots, y_n$ – выборка наблюдений случайной величины Y, имеющей неизвестное распределение $F_Y(y)$.

> Выборки называются однородными, если $F_X(\xi) = F_Y(\xi)$. Иными словами, выборки однородные, если они получены из одной и той же генеральной совокупности, или являются наблюдениями одной и той же случайной величины;
>
> Основная гипотеза однородности: две (или более) выборки взяты из одного распределения вероятностей H_0: $F_X(\xi) = F_Y(\xi)$.

Одним из наиболее простых критериев проверки гипотезы об однородности распределения случайных величин X и Y является критерий знаков, используемый для проверки двух связанных выборок. Критерий является примером непараметрического критерия математической статистики, т.е. критерия, использующего не сами численные значения элементов выборки, а структурные свойства выборки (например, отношения порядка между её элементами и пр.).

Мощность непараметрических критериев, как правило, меньше, чем мощность их параметрических аналогов. Причина этого связана с неизбежной потерей части информации, содержащейся в выборке. Однако непараметрические методы могут применяться при менее строгих предположениях о свойствах наблюдаемых случайных величин; они, как правило, достаточно просты с вычислительной точки зрения.



Abraham de Moivre

Если выборки получены из одной и той же генеральной совокупности, то значения x_i и y_i $(i = 1, \dots, n)$ взаимозаменяемы, и, следовательно, вероятности появления положительных и отрицательных разностей x_i и y_i равны,
т.е. $P(x_i - y_i > 0) = P(x_i - y_i < 0) = 1/2$.

Пусть K – число знаков "+" в последовательности знаков разностей $x_1 - y_1, \dots, x_n - y_n$. Если в этой последовательности разностей содержатся нулевые элементы, то они исключаются из рассмотрения. Далее считается, что в последовательности разностей нулевых элементов нет. При условии, что основная гипотеза H_O верна, а пары наблюдений (x_i, y_i), $(i = 1, \dots, n)$, и, следовательно, знаки разностей $x_i - y_i$ независимы, число K знаков "+" имеет биномиальное распределение $\mathbb{B}(n, 1/2)$.

Таким образом, проверка гипотезы однородности сводится к проверке гипотезы о параметре p биномиального распределения: H_0: $p = 1/2$. Можно показать, что эта гипотеза эквивалентна гипотезе о равенстве медиан распределений $F_X(x)$ и $F_Y(y)$.

Математическое ожидание $\mu_K = np$ и дисперсия $D = np(1-p)$. В соответствии с предельной теоремой Муавра-Лапласа при большом числе испытаний n статистика K имеет закон распределения, близкий к нормальному: $K \sim \mathbb{N}\left(np, \sqrt{np(1-p)}\right)$.

Частота "успеха" $H = K/n$ также имеет нормальное распределение $H \sim \mathbb{N}\left(p, \sqrt{p(1-p)/n}\right)$.
В качестве статистики критерия используется так называемая стандартизованная частота:

$$Z = \frac{H - 1/2}{\sqrt{1/4\pi}} = 2\sqrt{n}\left(H - \frac{1}{2}\right),$$

которая при условии истинности H_0 имеет стандартное нормальное распределение $\mathbb{N}(0,1)$.

Основная гипотеза H_0 должна отклоняться при больших отличиях частоты знаков "+" от значения $1/2$ как в меньшую, так и в большую сторону, т.е. в области больших абсолютных значений статистики критерия Z. Таким образом, критическая область для статистики Z должна выбираться двусторонней.

Условие равновероятности отклонений является необходимым, но не достаточным условием однородности выборок $x_1, \dots, x_n$ и $y_1, \dots, y_n$. Это означает, что из принятия основной гипотезы критерия знаков не следует однородность выборок, а следует лишь возможность однородности. Если же основная гипотеза критерия знаков отклоняется, то отклоняется и гипотеза однородности выборок.

Если в выборке имеются случаи $x_i = y_i$, то их следует исключить из выборки, уменьшив число наблюдений. Статистика критерия – это число элементов в выборке, при которых $x_i > y_i$.

Пример 2.4 Исследуется действие некой новой методики преподавания на учащихся, измеряемое с помощью некоторого теста. В результате выборочного тестирования испытуемых получены следующие результаты.

Уровень до тренинга	30	39	35	34	40	35	22	22	32	23	16	34	33	34
Уровень после тренинга	34	39	26	33	34	40	25	21	30	24	15	27	35	30

Определить, является ли действие методики преподавания на уровень усвоения материала статистически значимым при уровне значимости α = 0.1, используя критерий знаков.

На рис. 2.4. дан скриншот проводимых вычислений.

Поскольку $|Z| < Z_{\text{крит}}$, то нулевая гипотеза подтверждается – статистически значимой разницы между выборками нет, положительный эффект новой методики отсутствует. Утверждать, что новой методики оказывает значимое действие на уровень усвоения материала учащимися, нельзя.

В ряде областей (педагогика, психология, социология) критерий знаков используется в несколько иной форме. В частности, при решении разобранной выше задачи сравниваются результаты "до" и "после" воздействия на учащихся можно наблюдать тенденции повторного измерения – большинство показателей могут увеличиваться или, напротив, уменьшаться. Для того чтобы доказать эффективность воздействия, необходимо выявить статистически значимую тенденцию в смещении (обычно употребляемый термин "сдвиг") показателей. Типичным называют сдвиг, численность которого больше; нетипичным, соответственно, сдвиг меньшего объема.

Необходимо заметить, что количество измерений должно быть не меньше 5; при равенстве типичных и нетипичных сдвигов критерий знаков неприменим.

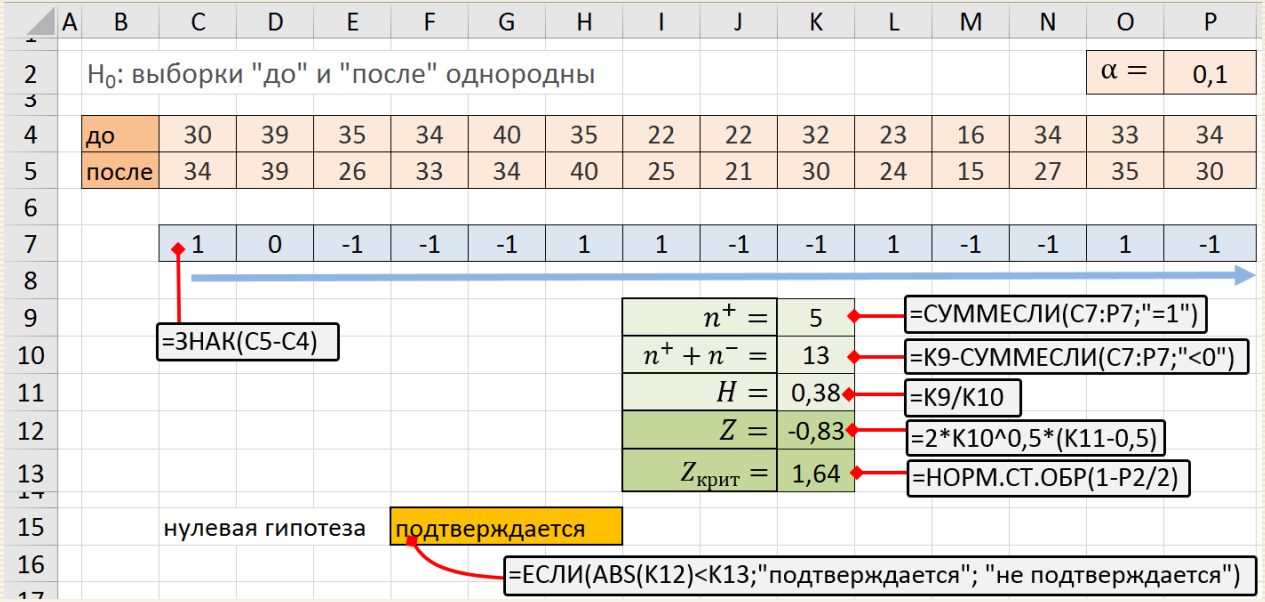


Рис. 2.4. Скриншот решения задачи по анализу эффективности воздействие методики преподавания

Пример 2.5 Для условий примера 2.4 на рис. 2.5 дан скриншот решения по следующему алгоритму.

1. Подсчитывается количество n^- "отрицательных" и "положительных" n^+ сдвигов. эмпирическим значением $G_{\text{эмп}}$ считается то количество сдвигов, которых меньше:

$$G_{\text{эмп}} = \min\{n^-; n^+\}.$$

2. Для определения $G_{\text{крит}}$ используется функция MS Excel БИНОМ.ОБР, которая возвращает наименьшее значение, для которого интегральное биномиальное распределение больше заданного значения критерия или равно ему

БИНОМ.ОБР(<число испытаний>; <вероятность успеха>; α)

Jakob Bernoulli

Обязательный аргумент <число испытаний> задает число испытаний, соответствующее формуле Бернулли, которая определяет вероятность реализации какого-либо события при независимых испытаниях.

Аргумент <вероятность успеха> в данном случае из очевидных допущений равен 0.5.

3. На последнем этапе сопоставляются между собой $G_{\text{крит}}$ и $G_{\text{эмп}}$. Если $G_{\text{эмп}}$ меньше $G_{\text{крит}}$, то сдвиг и "типичную" сторону считается достоверным.

В примере число положительных сдвигов превосходит количество сдвигов в отрицательном направлении. Поэтому в данной задаче типичным является положительный сдвиг.

Из таблицы видно, что число таких сдвигов равно 8. Эмпирическое значение критерия определяется, как число нетипичных сдвигов. В рассматриваемом случае $G_{\text{эмп}}$=5.

Рис. 2.5. Скриншот решения задачи по анализу эффективности воздействие методики преподавания

Нулевая гипотеза принимается, если $G_{\text{эмп}} \geq G_{\text{крит}}(\alpha, n = n^- + n^+)$). Поскольку в данном примере $G_{\text{эмп}}$=5 > 1 = $G_{\text{крит}}$(0.01; 13=5+8), то нулевая гипотеза принимается, и типичный сдвиг является случайным на заданном уровне значимости.

2.5. Тест Левина

Тест Левена (Levene's test) – это статистика, используемая для оценки равенства дисперсий нескольких выборок, когда проверяется нулевая гипотеза о равенстве дисперсий выборок (гомогенность (homogeneity), гомоскедастичность (homoscedasticity) дисперсий в отличии от их гетероскедастичности (heteroscedasticity).

Для теста Левена однородности групповых дисперсий значения групповых средств рассчитываются по одному из следующих вариантов (остатков от среднего):

$$z_{ij} = \begin{cases} \left|x_{ij} - \overline{x}_i\right| \\ \left|x_{ij} - \tilde{x}_i\right| \\ \left|x_{ij} - \bar{\bar{x}}_i\right| \end{cases},$$

где $\overline{x}_i = \sum_{j=1}^{n_i} x_{ij}$ – среднеарифметическое выборки;

$\tilde{x}_i$ – медиана для x_{ij},

$\bar{\bar{x}}_i$ – 10% обрезанное среднее (когда отбрасываются 5% точек в нижнем и верхнем хвостах распределения, т.е. вычисляется среднее значение для данных между 5-ми и 95-ми процентилями.

Howard Levene

После подобной трансформации данных "в остатки" выполняется тест ANOVA по абсолютной величине этих остатков. Если групповые отклонения равны, то средний размер остатка должен быть одинаковым для всех групп.

Статистика испытаний W определяется следующим образом:

$$W = \frac{n-k}{k-1} \cdot \frac{\sum_{i=1}^{k} n_i (\overline{z}_i - \overline{z}_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (z_{ij} - \overline{z}_{i.})^2}, \quad \overline{z}_{i.} = \sum_j z_{ij}, \quad n = \sum n_i .$$

где k – количество различных групп данных с числом элементов n_i в каждой i-той группе,

После пересчета исходной выборки техника анализа критерия Левена совпадает с таковой применяемой в ANOVA (см. раздел 4.3).

Основным показателем для принятия решения о подтверждении нулевой гипотезы является сравнение F-критерия Фишера с его критическим значением на заданном уровне значимости α.

Анализ применимости критерия показал, что использование обрезанного среднего лучше всего применять, когда распределение выборочных данных асимметричное и "растянутое". Использование медианы показывает хорошие результаты, когда базовые данные существенно асимметричны. Использование пересчета по среднему обеспечивает большую мощность критерия для симметричных и достаточно крутых распределений.

Хотя оптимальный выбор зависит от распределения данных в исходной выборке, преобразование, основанное на медиане, рекомендуется для применения, обеспечивающего достаточную устойчивость к типам "ненормальных" данных при сохранении большой мощности. В любом случае выбор преобразования данных должен основываться на знаниях о распределении исходных величин.

Необходимо отметить, что тест Левена менее чувствителен к отклонениям данных от нормального закона распределения, чем тест Бартлетта. Но в случае, когда исследуемые выборки соответствуют закону распределения Гаусса, то тест Бартлетта является более мощным.

Пример 2.6 В разных классах используется четыре метода обучения. Требуется определить наличие существенной разницы между показателями усвоения материала на уровне значимости 0.05; результаты тестирования* даны на рис. 2.6 в области B4:E11.

Алгоритм решения следующий.

1. Рассчитываются характеристики выборки: в ячейке H6 – количество групп; в ячейке H7 – общее число элементов выборки.

2. В диапазоне B14:E21 создается матрица "осколков по среднеарифметическому" z_{ij}, для которой вычисляются: в ячейке H9 – среднеарифметическое "осколков"; в диапазоне B23:E23 рассчитываются функции групповых дисперсий $n_i(\overline{z}_i - \overline{z}_{..})^2$.

3. Межгрупповая, внутригрупповая и общая составляющие изменчивости, обусловленная различием средних значений под влиянием фактора, вычисляются в ячейках K5:K7. Через соотношения этих величин рассчитывается основной показатель, который определяет статистическую значимость влияния фактора. Детально используемые соотношения приведены в формулах вычислений примера 4.7.

Основным показателем для принятия решения о подтверждении нулевой гипотезы является сравнение F-критерия Фишера с его критическим значением на заданном уровне значимости α.

При анализе с уровнем значимости α если уровень ошибки выше или равен α ($P_{\text{value}} \geq \alpha$), подтверждается нулевая гипотеза.

Из результатов (рис. 2.6) видно, что $F = 0.19 < F_{\text{крит}} = 2.95$ и $P_{\text{value}} = 0.90 > \alpha = 0.05$; посему отклонить нулевую гипотезу нельзя и можно заключить, что существенной разницы между методами нет.

*http://www.real-statistics.com/one-way-analysis-of-variance-anova/basic-concepts-anova

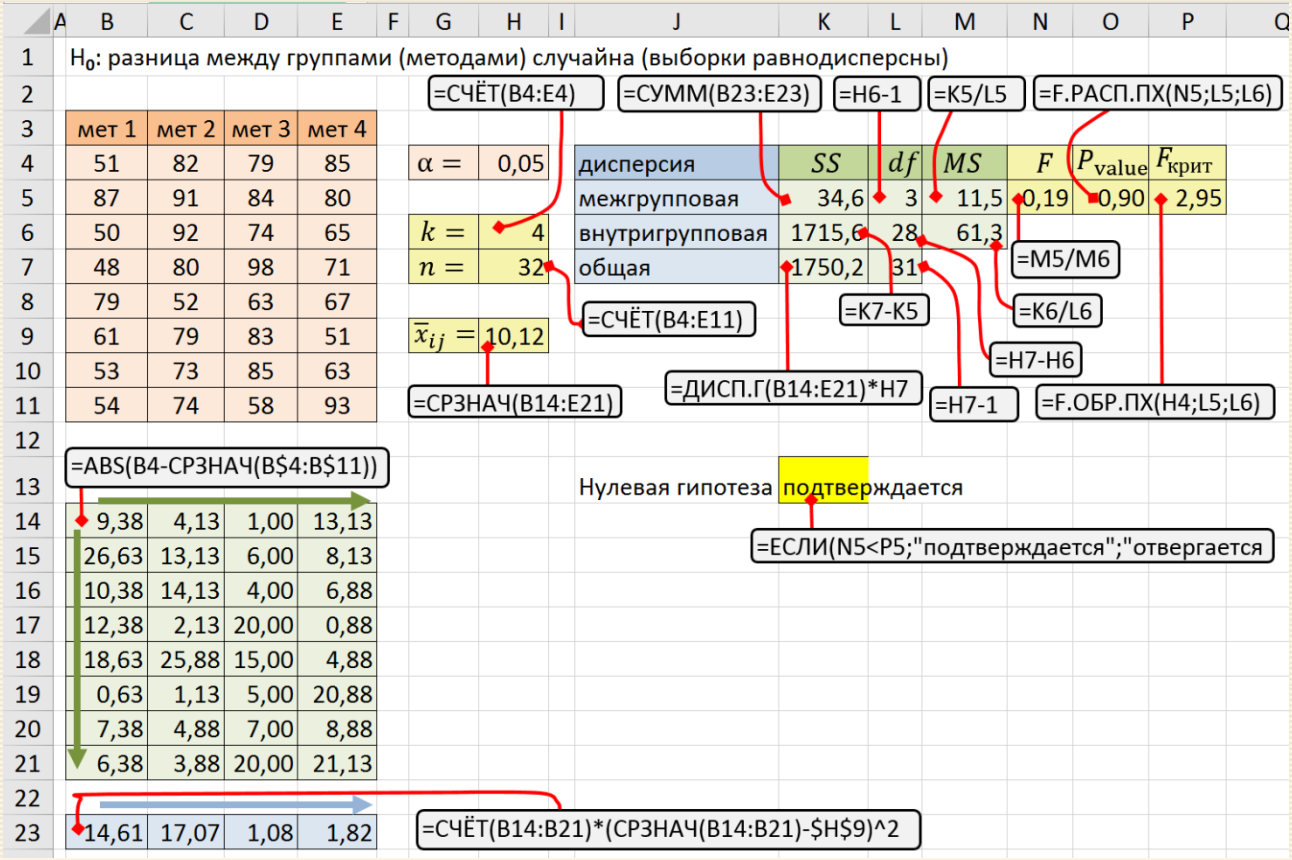


Рис. 2.6. Анализ дисперсионной однородности по тесту Левена

2.6. Достоверности совпадений и различий для порядковой шкалы

Критерий однородности χ^2 может использоваться для данных, измеренных в порядковой шкале по нескольким градациям (например, низкий, средний, высокий) при анализе однородности двух выборок. Пример подобных данных дан в табл.2.1.

Уровень знаний	Число правильно решенных задач
Низкий	0-10
Средний	11-15
Высокий	16 и более

Таблица 2.1 – Соответствие (перевод) шкалы отношений и порядковой шкалы

Задачи данного типа предполагают известными начальные (до начала эксперимента) состояния экспериментальной (ЭГ) и контрольной групп (КГ) и конечные (после окончания эксперимента) состояния. Анализ совпадений и различий позволяет, в частности, делать выводы об эффективности экспериментальной методики обучения.

Для данных, измеренных в порядковой шкале из L градаций, можно использование критерий однородности χ^2, эмпирическое значение $\chi^2_{\text{эмп}}$ которого вычисляется по следующей формуле:

$$\chi^2_{\text{эмп}} = N_{\text{КГ}} \cdot N_{\text{ЭГ}} \cdot \sum_{i=1}^{L} \frac{\left(\frac{q_i^{\text{КГ}}}{N_{\text{КГ}}} - \frac{q_i^{\text{ЭГ}}}{N_{\text{ЭГ}}}\right)^2}{q_i^{\text{КГ}} + q_i^{\text{ЭГ}}} .$$

Характеристикой каждой группы является число ее членов (частота), набравших тот или иной балл: $\{q_i^{\text{КГ}}\}, i = 1,2 \dots, L$ и $\{q_i^{\text{ЭГ}}\}, i = 1,2 \dots, L$.

Например, $q_2^{\text{ЭГ}}$ – число учащихся экспериментальной группы, правильно решивших от 11 до 15 задач (таб. 2.1). Естественно, что численности групп равны $N_{\text{КГ}} = \sum_{i=1}^{\text{КГ}} q_i^{\text{КГ}}$ и $N_{\text{ЭГ}} = \sum_{i=1}^{\text{ЭГ}} q_i^{\text{ЭГ}}$.

В качестве критического значения статистики используется правостороннее распределение χ^2.

Условия применимости критерия в исследованиях данного типа следующие.

1. Число градаций L должно быть не менее трех. Если требуется использование дихотомической $(L = 2)$ шкалы (да/нет, решил/не решил и т.д.), то целесообразнее применять критерий Фишера.
2. Для любого значения балла (градации) в любой из сравниваемых выборок число членов, получивших данный балл (попавших в данную градацию) должно быть не менее пяти, то есть в данном случае $q_i^{\text{КГ}} \geq 5$, $q_i^{\text{ЭГ}} \geq 5$ для любого значения i.

Алгоритм определения достоверности совпадений и различий для экспериментальных данных предусматривает следующие шаги.

1. Переход от шкалы отношений (если данные заданы в таковой) к порядковой шкале.
2. Подсчет необходимых для анализа межгрупповых значений $\chi^2_{\text{эмп}}$ для порядковых данных.
3. Сравнительный анализ полученных значений $\chi^2_{\text{эмп}}$ с критическим $\chi^2_{\text{крит}}$; формулирование заключения.

Пример 2.7. Проведены измерения в контрольной (КГ) и экспериментальной группах (ЭГ) в начале и конце периода, когда в ЭГ использовалась некоторая новая методика обучения (данные по исследованию представлены в таблице ниже). Требуется выполнить сравнительный анализ уровней знаний со значимостью 0.05 характеристик выборок контрольной и экспериментальных групп в соответствии с порядковой трехбалльной шкалой таблицы 2.1.

КГ до	15	13	11	18	10	8	20	7	8	12	15	16	13	14	14	10	8
	19	7	8	11	12	15	16	13	5	11	19	18	9	6	15	12	15
КГ после	16	12	14	17	11	9	15	8	6	13	17	19	15	11	9	14	17
	19	8	6	9	12	11	17	10	8	8	20	19	6	14	10	11	9
ЭГ до	12	11	15	17	18	6	8	10	16	12	15	14	19	13	19	9	12
	12	11	16	12	8	13	7	15	8	17	18	6	8	10	16	8	
ЭГ после	15	18	12	20	16	11	13	7	14	17	19	16	12	15	19	13	12
	18	14	13	18	13	13	15	18	9	14	12	20	16	11	13	15	

Если данные были исходно получены в шкале отношений, то предварительно необходимо выполнить их перевод (в рассматриваемом случае являющимся корректным) в шкалу отношений.

В данном примере рассматривается порядковая шкала с $L = 3$ уровнями (градациями, показателями).

Первый этап вычислений: переход от шкалы отношений к порядковой шкале. Уровень знаний определялся количеством правильных решений каких-либо задач в соответствии с установленными правилами агрегации.

Можно убедиться, что для данных рассматриваемого примера имеют место следующие значения (таб. 2.2)

Таблица 2.2 – Исходные данные в порядковой шкале

Уровень знаний	КГ до	ЭГ до	КГ после	ЭГ после
Низкий	11	11	13	2
Средний	16	13	12	19
Высокий	7	9	9	12

При этом, естественно, $\sum_{i=1}^{\text{КГ}} q_i^{\text{КГ}} = N_{\text{КГ}}$ и $\sum_{i=1}^{\text{ЭГ}} q_i^{\text{ЭГ}} = N_{\text{ЭГ}}$.

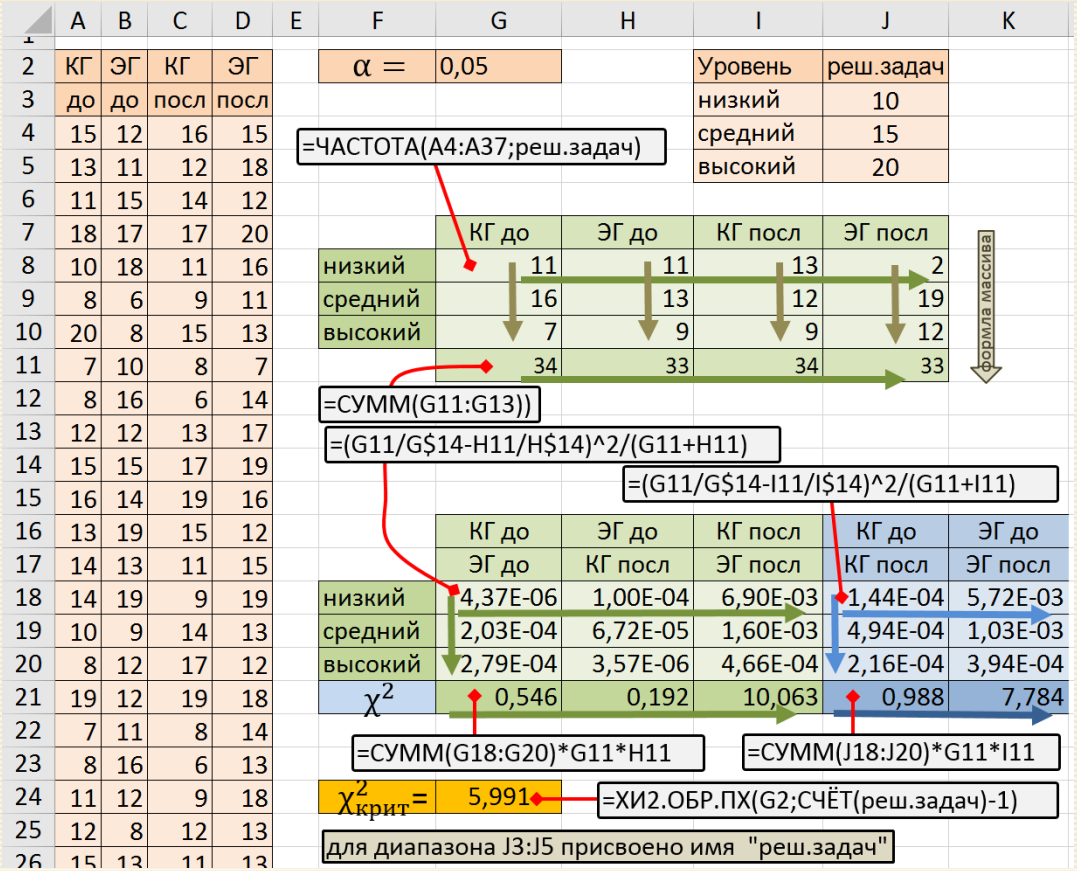
Переход от шкалы отношений к порядковой шкале может быть проведен в соответствие со схемой вычислений, представленной на рис. 2.7 скриншота листа Excel.

Последовательно выполняются следующие действия (см. рис.2.7).

A1. Заносятся исходные данные (столбцы A-D) – результаты тестирования контрольной и экспериментальной групп до и после начала исследования, а также градаций порядковой шкалы (ячейки J3:J5, которым присваивается имя "реш.задач").

A2. В диапазоне ячеек G8:J10 формируются частоты по уровням градации групп:

- в ячейку G8 вводится формула =ЧАСТОТА(A4:A37; реш.задач) и с помощью автозаполнения (тиражирования) растягивается до ячейки J8;
- каждая из ячеек G8:J8 растягивается на две последующих строки; к каждой тройке ячеек применяются правила работы с массивами (F2 и комбинация Ctrl + Shift + Enter).



	A	B	C	D
2	КГ	ЭГ	КГ	ЭГ
3	до	до	посл	посл
4	15	12	16	15
5	13	11	12	18
6	11	15	14	12
7	18	17	17	20
8	10	18	11	16
9	8	6	9	11
10	20	8	15	13
11	7	10	8	7
12	8	16	6	14
13	12	12	13	17
14	15	15	17	19
15	16	14	19	16
16	13	19	15	12
17	14	13	11	15
18	14	19	9	19
19	10	9	14	13
20	8	12	17	12
21	19	12	19	18
22	7	11	8	14
23	8	16	6	13
24	11	12	9	18
25	12	8	12	13
26	15	13	11	13

Рис. 2.7. Скриншот листа Excel сравнения выборок в порядковой шкале

A3. В ячейку G18 вводится формула =(G8/G$11-H8/H$11)^2/(G8+H8) подсчета составляющих для значения эмпирического критерия однородности; формула растягивается на диапазон G18:I20. Полученные составляющие суммируются по соответствующим ячейкам G21:I21 (значения χ^2 для соответствующих подгрупп сравнений). Для остальных сравниваемых групп аналогично описанной процедуре в ячейках J21:K21 определятся соответствующие значения χ^2.

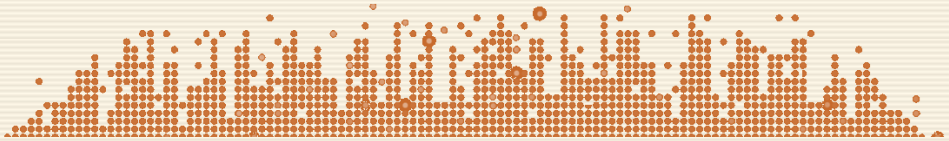
A4. В ячейке J24 рассчитывается значение $\chi^2_{\text{крит}}$.

На заключительном этапе выполняется собственно анализ данных; основной вывод следующий.

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают ($\chi^2_{\text{эмп}} = 0.546 < \chi^2_{\text{крит}} = 5{,}991$), а конечные (после окончания эксперимента) – различаются ($\chi^2_{\text{эмп}} = 10.063 > \chi^2_{\text{крит}} = 5{,}991$). Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментальной методики обучения.

Могут быть сформулированы и другие результаты вычислений по уровню значимости α = 0.05:

- однородность контрольной группы до начала и после окончания эксперимента сохранилась;
- характеристики экспериментальной группы до начала и после окончания эксперимента изменились.



2.7. Достоверности совпадений и различий для дихотомической шкалы

Для данных, измеренных в порядковой дихотомической шкале* для двух градаций (да/нет, решил/не решил и т.д.), для анализа однородности двух выборок (например, экспериментальной ЭГ и контрольной КГ)

	градация показателя	
	положительная	отрицательная
КГ	n_1^+	n_1^-
ЭГ	n_2^+	n_2^-

$n_1 = n_1^+ + n_1^-$

$n_2 = n_2^+ + n_2^-$

используется критерий (угловое преобразование) Фишера, который определяется соотношением

$$\varphi_{\text{эмп}} = \frac{2\left|\arcsin\left(\sqrt{p_1}\right) - \arcsin\left(\sqrt{p_2}\right)\right|}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\cdot,$$

где доля $p_1 = n_1^+/n_1$ и $p_2 = n_2^+/n_2$ – доли числа членов контрольной (КГ) и экспериментальной (ЭГ) групп, отвечающие положительной градации показателя ("да", "решил" и т.д.); n –групповые численности.

Критерий Фишера сопоставляет две выборки по частоте встречаемости интересующего исследователя эффекта (показателя).

*Дихотомическая шкала (dichotomous scale) – шкала, содержащая только две категории

Г.Ф. Лакин* отмечает, что долевые критерии пригодны при не слишком больших и не слишком маленьких значениях p $(0.25 < p < 0.75)$, что чаше всего относится к небольшим выборкам. Универсальность критерия достигается использование поправки Йейтса на непрерывность, когда частоты рассчитываются по соотношениям

$$p_1 = \frac{n_1^+}{n_1} + \frac{1}{2n_1}, \quad p_2 = \frac{n_2^+}{n_2} + \frac{1}{2n_2}.$$

Решение достигается сравнением $\varphi_{эмп}$ с критическим значением $\varphi_{крит}$: если $\varphi_{эмп} \leq \varphi_{крит}$, то характеристики сравниваемых выборок совпадают с уровнем значимости α; если $\varphi_{эмп} > \varphi_{крит}$, то достоверность различий характеристик сравниваемых выборок составляет $1 - \alpha$.

Для ограниченных выборок в качестве $\varphi_{крит}$ используется критическое значение Стьюдента для числа степеней свободы $df = n_1 + n_2 - 2$, который для заданного уровня значимости вычисляется через MS Excel функцию СТЬЮДЕНТ.ОБР.2Х(α; df). Для относительно больших выборок можно использовать обратное значение стандартного нормального распределения НОРМ.СТ.ОБР(1–α/2).

Условия применимости критерия следующие.

1. Ни одна из сопоставляемых долей не должна быть равной нулю. Формально нет препятствий для применения метода в случаях, когда доля наблюдений в одной из выборок равна нулю, однако в этих случаях результат может оказаться неоправданно завышенным.
2. Верхний предел в критерии φ отсутствует – выборки могут быть сколь угодно большими.

---

*Лакин Г.Ф. Биометрия: Учеб. пособие для биол. спец. вузов. М.: Высш. шк., 1990. 352 с.

Нижний предел составляют два наблюдения в одной из выборок. Однако должны соблюдаться следующие соотношения в численности этих двух выборок:

- если в одной выборке всего 2 наблюдения, то во второй должно быть не менее 30;
- если в одной из выборок всего 3 наблюдения, то во второй должно быть не менее 7;
- если в одной из выборок всего 4 наблюдения, то во второй должно быть не менее 5;
- если объемы выборок более пяти, то возможны любые сопоставления.

Пример 2.8. Изучалось влияние некого лекарства на выздоровление пациентов. Результаты исследования приведены в следующей таблице

	выздоровели	не выздоровели
КГ	3	11
ЭГ	23	13

На уровне значимости 0.05 проверить эффективность (положительность воздействия) препарата.

На рис. 2.8 представлены скриншоты двух решений задачи, когда доли рассчитываются без (левый рисунок) и с поправкой Йейтса (рисунок справа). Различие алгоритмов заключается лишь соотношениях для вычислений долей успешных результатов исследования. Можно отметить, что в первом случае разность долей "успеха" (и соответственно, $\varphi_{эмп}$) при расчете с поправкой Йейтса меньше.

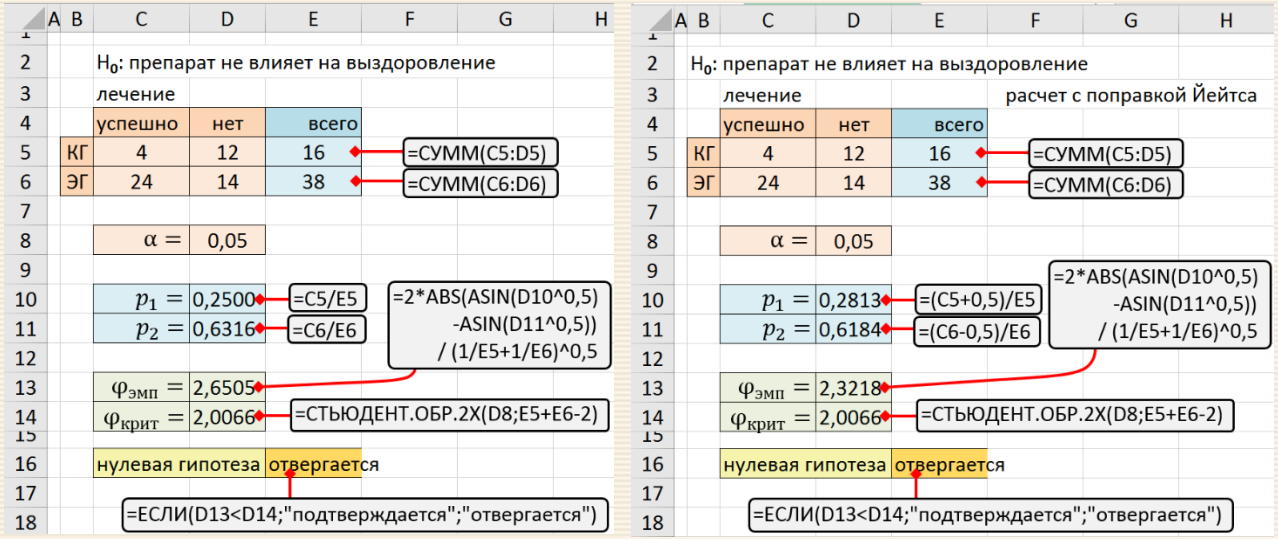


Рис 2.8 Расчет эффективности препарата без (слева) и с поправкой Йейтса (справа)

Пример 2.9. Выполнить сравнительный анализ различий знаний учащихся контрольной и экспериментальной групп (до и после применения некой методики) при оценке по двум уровням – "не усвоили материал" (число правильно решенных задач меньше либо равно 10) и "успешно усвоили материал" (число правильно решенных задач строго больше 10). Требуется определить достоверность различий состояний экспериментальной и контрольной групп после окончания эксперимента.

Таблица исходных данные в шкале отношений

Контрольная группа до эксперимента (КГ до)

КГ до	15	13	11	18	10	8	20	7	8	12	14	14	19	7
	8	11	12	15	16	13	5	11	19	18	9	6	15	

Экспериментальная группа до эксперимента (ЭГ до)

ЭГ до	12	11	15	17	18	6	8	10	16	12	13	19	12	11
	16	12	8	13	7	15	8	9						

Контрольная группа после эксперимента (КГ после)

КГ после	16	12	14	17	11	9	15	8	6	13	11	9	19	8
	6	9	12	11	17	10	8	8	20	19	6	14	10	

Экспериментальная группа после эксперимента (ЭГ после)

ЭГ после	15	18	12	20	16	11	13	7	14	17	15	19	18	14
	10	18	13	13	15	18	9	14						

Требуется определить состояния экспериментальной и контрольной групп (совпадают или различаются средние значения) до и после эксперимента, сделать вывод об эффекте изменений состояния групп вследствие применения экспериментальной методики обучения.

Таким образом, есть четыре выборки $\{x_i^{\text{КГ}}\}, i = 1,2,\dots,n_{\text{КГ}}$ и $\{x_i^{\text{ЭГ}}\}, i = 1,2,\dots,n_{\text{ЭГ}}$ данных до и после испытаний. Алгоритм сопоставления выборок для экспериментальных данных, исходно заданных в шкале отношений, включает два этапа – переход от шкалы отношений к дихотомической шкале и собственно анализ этих данных (скриншот на рис. 2.9).

Этап первый – переход от шкалы отношений к дихотомической шкале. Последовательно выполняются следующие действия.

A1. Заносятся исходные данные (диапазон данных A6:D32) – результаты тестирования контрольной и экспериментальной групп до и после начала исследования, а также контрольного числа решенных задач (ячейка L4).

A2. В диапазоне G8:J8 определяются объемы исследуемых выборок контрольной и экспериментальной групп, которые потребуются в дальнейших вычислениях.

A3. В ячейки G13:J13 (строчка "n^+") вводятся формулы, по которым подсчитывается число учащихся, решивших в контрольных и экспериментальных группах больше указанного (в ячейке L4) числа задач.

A4. В ячейки G14:J14 (строчка "p") вводятся формулы, по которым подсчитывается доля учащихся, прошедшая через контрольный барьер (отношения значений G13:J13 к соответствующим численностям групп G8:J8) и корни из этих величин в G15:J15.

A5. В ячейке K6 рассчитывается вспомогательное значение c (неизменяемая часть критерия $\varphi_{\text{эмп}}$)

$$c = \frac{2}{\sqrt{1/n_{\text{КГ}} + 1/n_{\text{ЭГ}}}}.$$

A6. В ячейках G24:J24 рассчитываются значения $\varphi_{\text{эмп}}$ для межгрупповых отношений; в G19 заносится величина критического значения $\varphi_{\text{крит}}$.

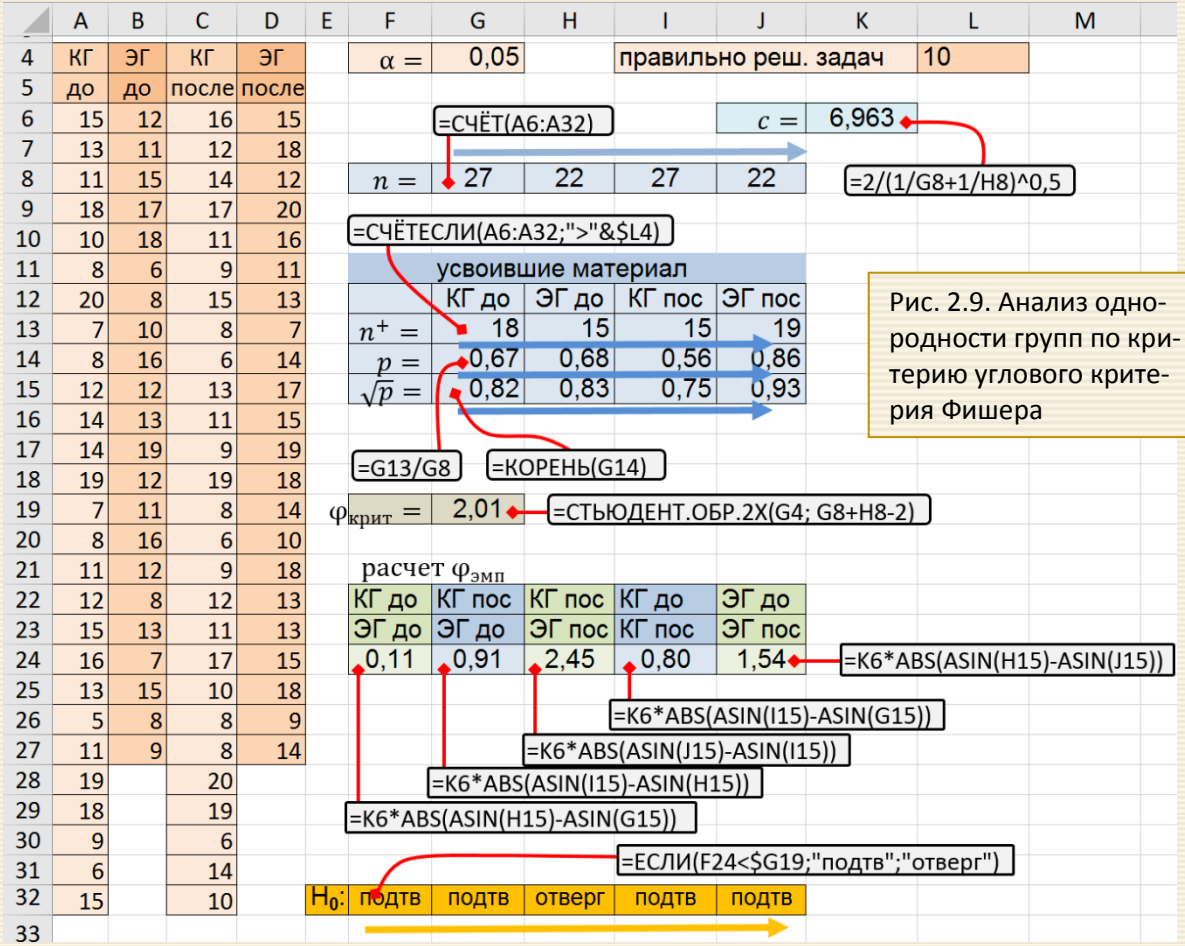


Рис. 2.9. Анализ однородности групп по критерию углового критерия Фишера

На заключительном этапе выполняется сравнение эмпирических значений с критическим (ячейки G32:J32); по величинам КГ и ЭГ до и после (эксперимента) формулируется основной вывод.

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают ($\varphi_{эмп} = 0.11 < \varphi_{крит} = 2.01$), а конечные (после окончания эксперимента) – различаются ($\varphi_{эмп} = 2.45 > \varphi_{крит} = 2.01$). Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментальной методики обучения.

Могут быть сформулированы и другие результаты вычислений по уровню значимости $\alpha = 0.05$:

- характеристики контрольной группы до начала и после окончания эксперимента совпадают;
- контрольная группа после эксперимента осталась на уровне экспериментальной группы до эксперимента;
- характеристики экспериментальной группы до начала и после окончания эксперимента совпадают.

3. Критерии согласия

В практических задачах статистики модель закона распределения заранее не известна и возникает задача выбора модели, согласующейся с результатами наблюдений случайной величины.

Анализируется соответствие неизвестной (выборочной) функция распределения $F_{\text{эмп}}(x)$ исследуемой случайной величины X с известной теоретической $F_{\text{теор}}(x)$, т.е. высказывается гипотеза: H_0: $F_{\text{эмп}}(x) = F_{\text{теор}}(x)$. В качестве теоретических могут рассматриваться нормальное, равномерное либо иное распределения вероятности. Это определяется сущностью изучаемого явления, а также результатами предварительной обработки полученных экспериментальных данных (видом гистограммы, полигона частот, соотношением основных числовых характеристик).

Критерии, с помощью которых проверяется гипотеза о законе распределения, называются критериями согласия.

> Критерием согласия называют критерий, который позволяет установить, является ли расхождение эмпирического и теоретического распределений случайным или значимым, т. е. согласуются ли данные наблюдений с выдвинутой статистической гипотезой или не согласуются.

Распределение генеральной совокупности, имеющее место в силу выдвинутой гипотезы, называют теоретическим. Обычно эмпирические и теоретические частоты различаются в силу того, что:

- расхождение случайно и связано с ограниченным количеством наблюдений;
- расхождение неслучайно и объясняется тем, что статистическая гипотеза о теоретическом законе распределения ошибочна.

3.1. Критерий согласия Пирсона

Критерий χ^2 (хи-квадрат), проверяющий значимость расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот, был предложенв1900 году и широко используется в настоящее время. Более того, его применяют для решения широкого круга задач, связанных с анализом номинальных данных, т.е. данных, выражаемым не количеством, а качеством объектов (вид растения, встречаемость (частоты) элементов и т.д.). В качестве статистического показателя Пирсон рассматривал квадраты отклонений наблюдаемых частот, исходно обозначаемых O (Observed), и ожидаемых – E (Expected)

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

Формальное определение критерия χ^2 следующее – распределение χ^2 (хи-квадрат) с df степенями свободы – это распределение суммы квадратов df независимых стандартных нормальных случайных величин. Распределение χ^2 – это семейство распределений, каждое из которых зависит от числа степеней свободы.

С увеличением степеней свободы распределение хи-квадрат стремится к нормальному. Это объясняется действием центральной предельной теоремы, согласно которой сумма большого количества независимых случайных величин имеет нормальное распределение.

Критерий χ^2 применяется для сравнения распределений объектов двух совокупностей на основе измерений по шкале наименований в двух н е з а в и с и м ы х выборках.

Критерий χ^2 отвечает на вопрос, с одинаковой ли частотой встречаются разные значения признака в эмпирическом и теоретическом распределениях или в двух и более эмпирических распределениях. Это один из наиболее часто применяемых критериев при анализе данных как для количественных дискретных вариаций, так и для непрерывных. Преимущество метода состоит в том, что он позволяет сопоставлять распределения признаков, представленных в любой шкале, начиная от шкалы наименований. В самом простом случае это оценка двух альтернативных (дихотомических) показателей, когда можно обоснованно применять критерий χ^2: "да-нет", "живой-мертвый", "решил задачу – не решил задачу" и т.д.

В случае, если исследователя интересует проблема совпадения (или несовпадения) его данных с результатами других исследователей, то с помощью метода χ^2 можно сопоставить два эмпирических распределения – скажем, ваше и приведенное в других работах.

Аналогично можно сопоставлять распределения выборок из трех и более альтернатив. Например, если в выборке из 50 человек 30 выбрали ответ (а), 16 человек – ответ (б) и четверо – ответ (в), то с помощью метода χ^2 можно проверить, отличается ли это распределение от равномерного распределения или от распределения ответов в другой выборке, где ответ (а) выбрали 12 человек, ответ (б) – 25 человек, ответ (в) – 13 человек.

В тех случаях, если признак измеряется количественно (в килограммах, секундах, миллиметрах и т.д.), необходимо представить всю совокупность значений признака по нескольким классам и затем с помощью метода χ^2 сопоставить частоты встречаемости признака в каждом классе. В остальном принципиальная схема применения метода не меняется.

При сопоставлении эмпирического распределения с теоретическим определяется степень расхождения между эмпирическими и теоретическими частотами. При сопоставлении двух эмпирических распределений определяется степень расхождения между эмпирическими частотами и теоретическими частотами, которые наблюдались бы в случае совпадения двух этих эмпирических распределений.

Чем больше расхождение между двумя сопоставляемыми распределениями, тем больше эмпирическое значение $\chi^2_{\text{эмп}}$.

Для нахождения $\chi^2_{\text{эмп}}$ обычно используется следующая формула:

$$\chi^2_{\text{эмп}} = \sum_{i=1}^{M} \frac{\left(x_{i\ \text{экс}} - x_{i\ \text{теор}}\right)^2}{x_{i\ \text{теор}}},$$

где $x_{i\ \text{теор}}$ – теоретически ожидаемое число или показатель для данного класса (диапазона, интервала, группы);

$x_{i\ \text{экс}}$ – фактически наблюдаемое; M – количество классов (интервалов).

Ограничения критерия следующие.

1. Объем выборки должен быть достаточно большим: n > 30. При n < 30 критерий χ^2 дает весьма приближенные значения. Точность критерия повышается при больших значениях n.
2. Теоретическая частота для каждой ячейки таблицы не должна быть меньше 5. Это означает, что если число классов M задано заранее и не может быть изменено, то применять метод χ^2, не накопив определенного минимального числа наблюдений, нельзя.

Если, например, проверяются предположения о том, что частота заболеваний гриппом неравномерно распределяются по 7 дням недели, то потребуется исследование 5·7=35 случаев для анализа. Таким образом, если количество классов M задано заранее (M=7), как в данном случае, минимальное число наблюдений ($n_{\min}$) определяется по формуле: $n_{\min} = 5 \cdot M = 35$.

3. Выбранные классы должны включать всё распределение, то есть охватывать весь диапазон вариативности признаков. При этом группировка на классы должна быть одинаковой во всех сопоставляемых распределениях.

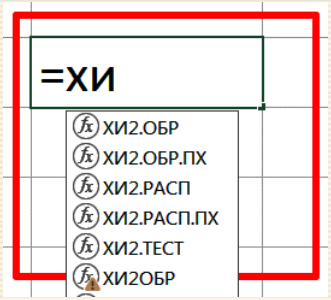
4. Необходимо (скажем, желательно) вносить так называемую "поправку на непрерывность" Yates'a при сопоставлении распределений признаков, которые принимают всего два значения (таблиц сопряженности 2×2). Уточнение обусловлено тем, что теоретическое распределение $\chi^2_{\text{эмп}}$ непрерывно, тогда как набор вычисленных значений χ^2 дискретен.

$$\chi^2_{\text{эмп}} = \sum \frac{\left(\left|x_{i\ \text{экс}} - x_{i\ \text{ожид}}\right| - 1/2\right)^2}{x_{i\ \text{ожид}}}.$$

Здесь $x_{i\ \text{экс}}$ – фактически наблюдаемое, а $x_{i\ \text{ожид}}$ – ожидаемое (расчетное или теоретическое) число или показатель для данной группы.

5. Классы должны быть неперекрещивающимися: если наблюдение отнесено к одному классу, то оно уже не может быть отнесено ни к какому другому классу. И, очевидно, что сумма наблюдений по классам всегда должна быть равна общему количеству наблюдений.

В MS Excel есть несколько функций, связанных с критерием χ^2-квадрат.



ХИ2.ОБР – возвращает критическое значение критерия при заданной вероятности слева (как в статистических таблицах);

ХИ2.ОБР.ПХ – возвращает критическое значение критерия при заданной вероятности справа. Функция по сути дублирует предыдущую. Но здесь можно сразу указывать уровень α, а не вычитать его из 1. Это более удобно, т.к. в большинстве случаев нужен именно правый хвост распределения.

ХИ2.РАСП – возвращает *P*-уровень (*P*-level) слева (можно рассчитать плотность распределения).

ХИ2.РАСП.ПХ – *P*-уровень справа.

ХИ2.ТЕСТ – по двум заданным диапазонам частот сразу проводит тест хи-квадрат. Количество степеней свободы берется на одну меньше, чем количество частот в столбце (как и должно быть), возвращая значение *P*-уровень.

Критерий согласия хи-квадрат для непрерывной вариации

Во многих практических задачах точный закон распределения неизвестен, поэтому выдвигается гипотеза о соответствии имеющегося в распоряжении исследователя эмпирического закона, построенного по наблюдениям, которая (гипотеза) требует статистической проверки.

Пусть X – исследуемая случайная величина. Требуется проверить гипотезу H_0 о том, что данная случайная величина подчиняется закону распределения $f(x)$. Для этого необходимо произвести выборку из n независимых наблюдений и по ней построить эмпирический закон распределения $f_0(x)$.

Сравниваются экспериментальная и теоретическая функции распределений. Общий алгоритм проверки "нулевой" гипотезы следующий.

1. Выдвинуть гипотезу H_0: $f(x) = f_0(x)$ равенства эмпирического и теоретического законов распределения. Здесь $f_0(x)$ – плотность вероятности гипотетического закона распределения: нормального, равномерного или какого-либо другого.

2. Вычислить значение критерия по формуле

$$\chi^2_{\text{эмп}} = \sum_{i=1}^{M} \frac{\left(n_i - n_i^{\text{теор}}\right)^2}{n_i^{\text{теор}}},$$

где n_i – число данных в i-том интервале;

$n_i^{\text{теор}}$ – число данных, которое должно содержаться в i-том интервале согласно выбранному закону (например, нормального) распределения, т.е. при условии, что гипотеза H_0 верна.

3. По имеющейся в MS Excel функции (или из справочных таблиц) выбирается "критическое" значение $\chi^2_{\text{крит}} = \chi^2_{\text{крит}}(\alpha, df)$, где α – заданный уровень значимости, а df – число степеней свободы, определяемое по формуле $df = M - 1 - s$, s – число параметров, от которых зависит выбранный гипотезой H_0 закон распределения. Значение s для равномерного и нормального закона равно 2.

4. Если $\chi^2_{\text{эмп}} > \chi^2_{\text{крит}}$, то гипотеза H_0 отклоняется. В противном случае оснований ее отклонить нет и H_0 принимается.

Пример 3.1 Проверка соответствия выборки (задана таблицей) нормальному закону распределения при уровне значимости $\alpha = 0.01$.

n_i	3	8	25	40	46	31	6	2
$n_i^{\text{т}}$	2.12	8.93	25.39	43.03	43.52	26.26	9.45	2.30

На рис. 3.1 приведен скриншот проверки соответствия выборки нормальному закону распределения.

Сопоставление данной выборки и соответствующих значений нормального распределения показывает, что $\chi^2_{\text{крит}} \approx 15.09$ и $\chi^2_{\text{эмп}} \approx 2.98$ на доверительном уровне α=0.01.

Поскольку $\chi^2_{\text{эмп}} < \chi^2_{\text{крит}}$, то нулевая гипотеза (соответствия выборки нормальному закону распределения) подтверждается, т.е. выборка соответствует нормальному закону распределения.

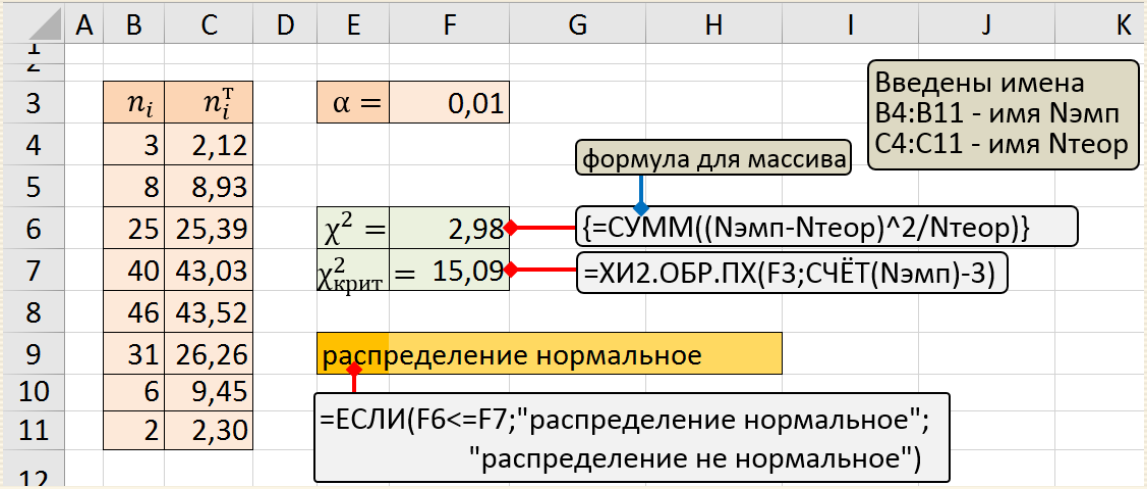


Рис. 3.1 Скриншот проверки соответствия выборки нормальному закону распределения

Пример 3.2 Задача: оценить соответствие выборки нормальному распределению (для выборки, заданной выше в форме таблицы классов).

Для анализа используется критерий χ^2, с помощью которого на заданном уровне значимости сравниваются эмпирические частоты n_i для классов значений $x \in [x_i^{\text{нач}}; x_i^{\text{кон}})$ и соответствующие нормальному распределению теоретические частоты $n_i^{\text{т}}$.

Необходимые для идентификации параметры нормального распределения среднее арифметическое и стандартное отклонение определяются формулами

$$\overline{x} = \frac{1}{n}\sum\left[\frac{x_i^{\text{нач}} + x_i^{\text{кон}}}{2} \cdot n_i\right] = \frac{1}{2n}\sum(x_i^{\text{нач}} + x_i^{\text{кон}}) \cdot n_i \,,$$

$$S = \sqrt{\frac{1}{n-1}\sum\left[n_i \cdot \left(\frac{x_i^{\text{нач}} + x_i^{\text{кон}}}{2} - \overline{x}\right)^2\right]} \,,$$

где число элементов выборки $n = \sum n_i$.

Определить необходимое для вычислений теоретическое для нормального закона значение $n_i^{\text{т}}$ можно по соотношению

$$n_i^{\text{т}} = n \cdot \int_{x_i^{\text{нач}}}^{x_i^{\text{кон}}} p(x, \overline{x}, S)\, dx \,.$$

Вычислить данный интеграла можно точно следующим образом

$$n_i^{\text{т}} = n \cdot \int_{x_i^{\text{нач}}}^{x_i^{\text{кон}}} p(x, \overline{x}, S)\, dx = n \cdot \int_{-\infty}^{x_i^{\text{кон}}} p(x, \overline{x}, S)\, dx - n \cdot \int_{-\infty}^{x_i^{\text{нач}}} p(x, \overline{x}, S)\, dx \ ,$$

используя Excel-функцию НОРМ.РАСП(x; $\overline{x}$; S; ИСТИНА).

После того как определены n – объем выборки (ячейка H5);

$\overline{x}$ – среднеарифметическое для выборки (ячейка H6);

S – стандартное выборочное отклонение (ячейка H7)

- для первого класса диапазона $x \in (-\infty; x_i^{\text{кон}})$ в ячейку E4 заносится формула =H$5*НОРМ.РАСП(C4; H6; H7; ИСТИНА);
- в ячейку E5 формула), тиражируемая на диапазон E6:E10;

  =H$5*НОРМ.РАСП(C5; H$6; H$7; ИСТИНА) -H$5*НОРМ.РАСП(B5; H$6; H$7; ИСТИНА) ;
- в ячейку E5 формула =H$5*(1-НОРМ.РАСП(B11; H$6; H$7; ИСТИНА)) для последнего классового диапазона $x \in [x_i^{\text{нач}}; +\infty)$.

Скриншот проверки малой выборки "на нормальность" для изложенного подхода дан на рис. 3.2.

Можно использовать другое осреднение оценки интеграла по отрезку $[x_i^{\text{нач}}; x_i^{\text{кон}})$

$$n_i^{\text{т}} = n \cdot \int_{x_i^{\text{нач}}}^{x_i^{\text{кон}}} p(x, \overline{x}, S)\, dx \approx n \cdot (x_i^{\text{кон}} - x_i^{\text{нач}}) \cdot p\left(\frac{x_i^{\text{кон}} + x_i^{\text{нач}}}{2}, \overline{x}, S\right).$$

Для такого подхода (скриншот приведен на рис. 3.3.) в ячейку E4 заносится формула =H$5*(C4-B4)*НОРМ.РАСП((C4+B4)/2;H$6;H$7;) и тиражируется на диапазон E5:E11.

Замечание. После вычисления всех "теоретических" вероятностей $n_i^{\text{т}}$ полезно проверить, выполняется ли контрольное соотношение $\sum n_i^{\text{теор}} \approx n = \sum n_i$.

Рис. 3.2. Скриншот проверки выборки "на нормальность"

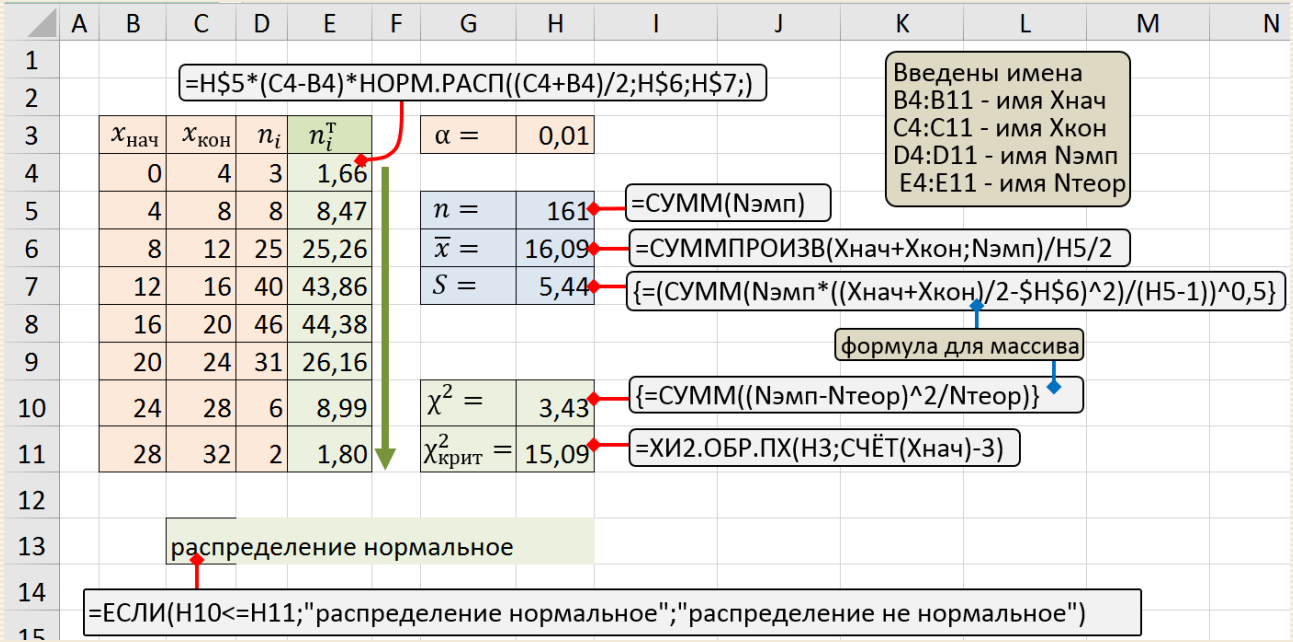


Рис. 3.3. Скриншот проверки выборки "на нормальность"

В обоих случаях степень свободы, использующейся в параметрах функции для критического значения критерия χ^2, равна числу классов минус 1 и еще минус 2 (число параметров нормального распределения).

Критерий согласия хи-квадрат для дискретной вариации

В данном разделе на примере сравнения экспериментальной и теоретической функций распределения "взгляда" лабораторного животного рассматривается общий алгоритм проверки нулевой гипотезы для качественной дискретной вариации.

Пример 3.3 Распределение "взгляда" лабораторного животного.

Чтобы определить цветовые предпочтения у лабораторного животного, перед ним в клетке были повешены 4 цветных круга (красный, синий, зеленый и желтый). К голове животного были прикреплены сенсоры фиксации глаз и были получены результаты, представленные в нижерасположенной таблице.

Цвет	Красный $n_1^{\text{эмп}}$	Синий $n_2^{\text{эмп}}$	Зеленый $n_3^{\text{эмп}}$	Желтый $n_4^{\text{эмп}}$	Всего "взглядов"
Количество "взглядов"	14	5	8	5	32

Полученные эмпирические частоты необходимо сопоставить с теоретическими. Если животное не отдает предпочтения ни одному цвету, то данное распределение показателя направленности "взгляда" не будет отличаться от равномерного распределения: на всех цветовых кругах задержка взгляда будет примерно одинаковой частоты.

Нулевая гипотеза: распределение "взглядов" лабораторного животного между цветами не отличается от равномерного распределения.

Теоретическая частота "взгляда" при равномерном распределении определяется просто. Если бы все "взгляды" животного распределялись равномерно между 4-мя кругами, то, очевидно, каждый из них получил бы по 1/4 всех ее "взглядов".

Теоретическая частота при сопоставлении эмпирического распределения с равномерным определяется формулой: $n_i^{\text{т}} = n/M$, где $n = \sum n_i^{\text{эмп}}$ – количество наблюдений; M – количество классов признака (M = 4 – рассматриваются четыре цветовых круга).

В данном случае признак – "взгляд" животного, направленный на какой-либо из цветов; количество классов признака равно 4 – это направления "взгляда" по количеству цветовых кругов; количество наблюдений – 32. Таким образом, в рассматриваемом случае для всех частот $n_i^{\text{т}} = n^{\text{т}} = 36/4 = 9$.

Далее с этой теоретической частотой сравниваются все эмпирические частоты. Заметим, что по заданной исходной таблице данных на первый взгляд отмечается явное предпочтение красного цвета.

Обсчет данного примера (рис .3.4) показывает, что $\chi^2_{\text{крит}} \approx 7.815$ и $\chi^2_{\text{эмп}} \approx 6.000$ на доверительном уровне α =0.05.

Поскольку $\chi^2_{\text{эмп}} < \chi^2_{\text{крит}}$, то нулевая гипотеза (соответствия выборки равномерному закону распределения) подтверждается. Распределение "взгляда" лабораторного животного между цветными кругами не отличается от равномерного распределения.

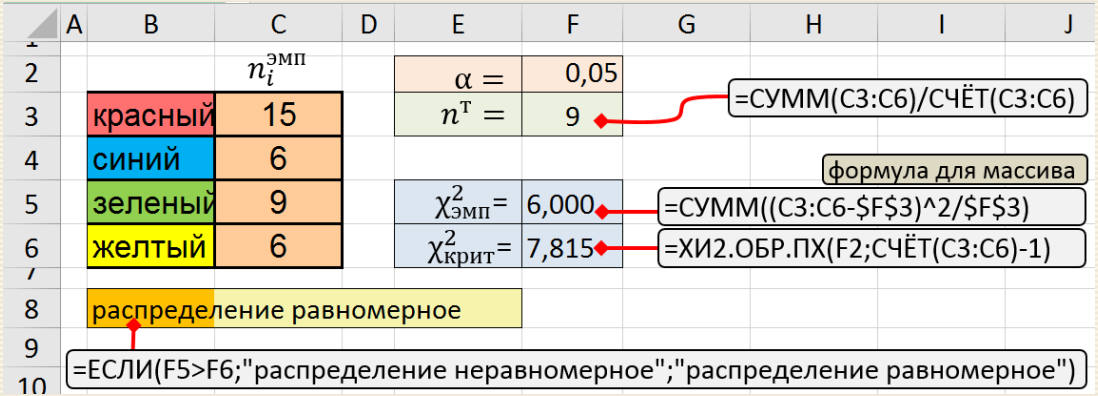


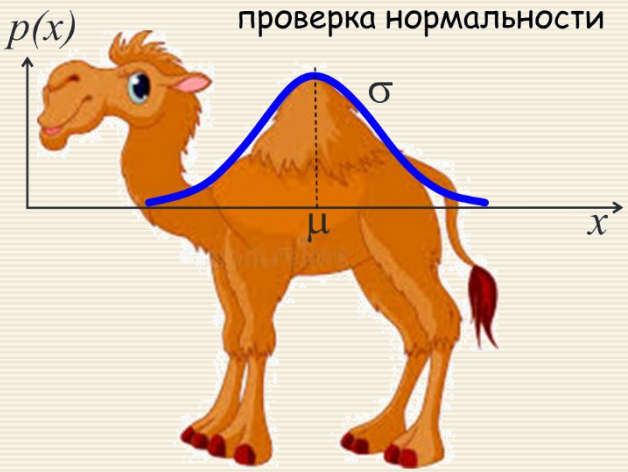
Рис. 3.4 Скриншот проверки соответствия выборки равномерному распределению

Критерий χ^2 является своего рода базой для исследования таблиц сопряженности – средства представления совместного распределения двух переменных, предназначенное для анализа связи между ними.

Подробный разбор примеров анализа таблиц сопряженности (контингентности (contingency) факторов) приведен в разделе 7.

3.2. Критерий Колмагорова

Классический критерий Колмогорова предназначен для проверки простых гипотез о принадлежности анализируемой выборки некоторому полностью известному закону распределения.



Пусть $x_1, x_2, \dots, x_n$ – выборка независимых одинаково распределённых случайных величин, $F_{\text{эмп}}(x)$ эмпирическая функция распределения, $F_{\text{теор}}(x_i)$ – некоторая "истинная" (теоретическая) функция распределения с известными параметрами.

Критерий можно применять и тогда, когда сравниваются между собой два экспериментальных распределения, но в данном случае ответ будет состоять только в одном – принадлежат ли эти два распределения к какому-то одному типу без, естественно, определения самого типа распределения.

Структура критерия Колмогорова-Смирнова базируется на принципе сравнения распределений на сравнении процентильных кривых этих распределений, а характер распределения не учитывается (процентильных кривые – частотное распределение данных, построенных по принципу суммирования накопленной частоты встречаемости всех значений ниже данного).

Андрей Николаевич Колмогоров

Критерий Колмогорова в качестве меры расхождения между теоретическим и эмпирическим распределениями рассматривает максимальное значение абсолютной величины разности между эмпирической и теоретической функцией распределения $\left|F_{\text{эмп}}(x_i) - F_{\text{теор}}(x_i)\right|$. Первоначально по выборочным данным определяют точечные оценки параметров предполагаемого распределения (например, $\overline{x}, S$). Затем строится эмпирическая функция распределения $F_{\text{эмп}}(x_i)$ по данным выборки. Теоретическая функция распределения строится по предполагаемому закону распределения (например, нормальному).

Наблюдаемое значение критерия

$$\lambda = \sqrt{n}\, D = \sqrt{n}\ \times \max\left|F_{\text{эмп}}(x_i) - F_{\text{теор}}(x_i)\right| .$$

сравнивают с критическим значением (таблица ниже) и делают соответствующий вывод.

α	0.40	0.30	0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
$\lambda_{\text{крит}}(\alpha)$	0.89	0.97	1.07	1.22	1.36	1.48	1.63	1.73	1.95	2.03

Если статистика $\lambda = \sqrt{n}\, D$ превышает процентную точку распределения Колмогорова $\lambda_{\text{крит}}(\alpha)$ заданного уровня значимости α, то нулевая гипотеза о соответствии закону отвергается. Иначе гипотеза принимается на уровне α.

При справедливости гипотезы H_0 статистика $\lambda < \lambda_{\text{крит}}$ имеет асимптотическое распределение Колмогорова, а $\lambda_{\text{крит}}$ определяется из $P\{\lambda > \lambda_\alpha\} = \alpha$, где α – уровень значимости (табл. выше).

Для наиболее часто используемой области $\alpha \in [0.005; 0.100]$ критическое значение $\lambda_{\text{крит}}(\alpha)$ можно с ошибкой менее одного процента аппроксимировать зависимостью

$$\lambda_{\text{крит}}(\alpha) = 0{,}835 - 0{,}171 \cdot \ln(\alpha)\,.$$

Пример 3.4 В таблице представлена выборка по изменению размера некоторого объекта (например, длины взрослого животного). На уровне значимости 0.05 проверить гипотезу, что размеры распределены по нормальному закону.

Размер, см	94-100	100-106	106-112	112-118	118-124	124-130	130-136	136-142
Число животных	3	7	11	20	28	19	10	2

Алгоритм выполнения анализа нормальности предусматривает следующее (рис. 3.5).

1. Для известных диапазонов ($x_i^{\text{нач}} \div x_i^{\text{кон}}$) размеров рассчитываются (столбец F) средних значений диапазонов классов $x_i^{\text{средн}} = (x_i^{\text{нач}} + x_i^{\text{кон}})/2$.

2. Вычисляется (столбец G) эмпирическая кумулятивная функция относительных частот $F_i^{\text{эмп}}$.

3. В ячейках L8 и L9 рассчитываются среднее арифметическое и выборочное стандартное отклонение по выборке по соотношениям частотного ряда

$$\overline{x} = \frac{1}{n}\sum n_i x_i^{\text{средн}}, \quad S = \left[\frac{1}{n-1}\sum n_i \left(x_i^{\text{средн}} - \overline{x}\right)^2\right]^{0.5}, \quad n = \sum n_i .$$

4. В столбце H рассчитываются теоретические частоты для нормального закона распределения

$$n_i^{\text{теор}} = n \cdot p_i ,$$

где вероятность "попадания" в i–й отрезок определяется интегральной Excel-функцией вероятностей нормального распределения НОРМ.РАСП($x; \overline{x}; S$):

- для первого участка (все размеры от $-\infty$ до x_0=94) через $= n$ ·НОРМ.РАСП($x_0; \overline{x}; S$) ;
- для внутренних участков размеры от $x_i^{\text{нач}}$ до $x_i^{\text{кон}}$) посредством
$$= n \cdot (\text{НОРМ.РАСП}(x_i^{\text{кон}}; \overline{x}; S)\text{-НОРМ.РАСП}(x_i^{\text{нач}}; \overline{x}; S)) ;$$
- для последнего участка (все размеры от x_{k+1}=142 до +∞): $= n$ ·(1-НОРМ.РАСП($x_{k+1}; \overline{x}; S$)) .

5. В столбцах I и J рассчитываются относительные частоты $f_i^{\text{теор}}$ и "теоретическая" кумулятивная функция относительных частот $F_i^{\text{теор}}$.

6. Вычисляются λ (ячейка N17), $\lambda_{\text{крит}}(\alpha)$ (ячейка N3) и через их сравнение (ячейка M19) формулируется ответ по достоверности нулевой гипотезы.

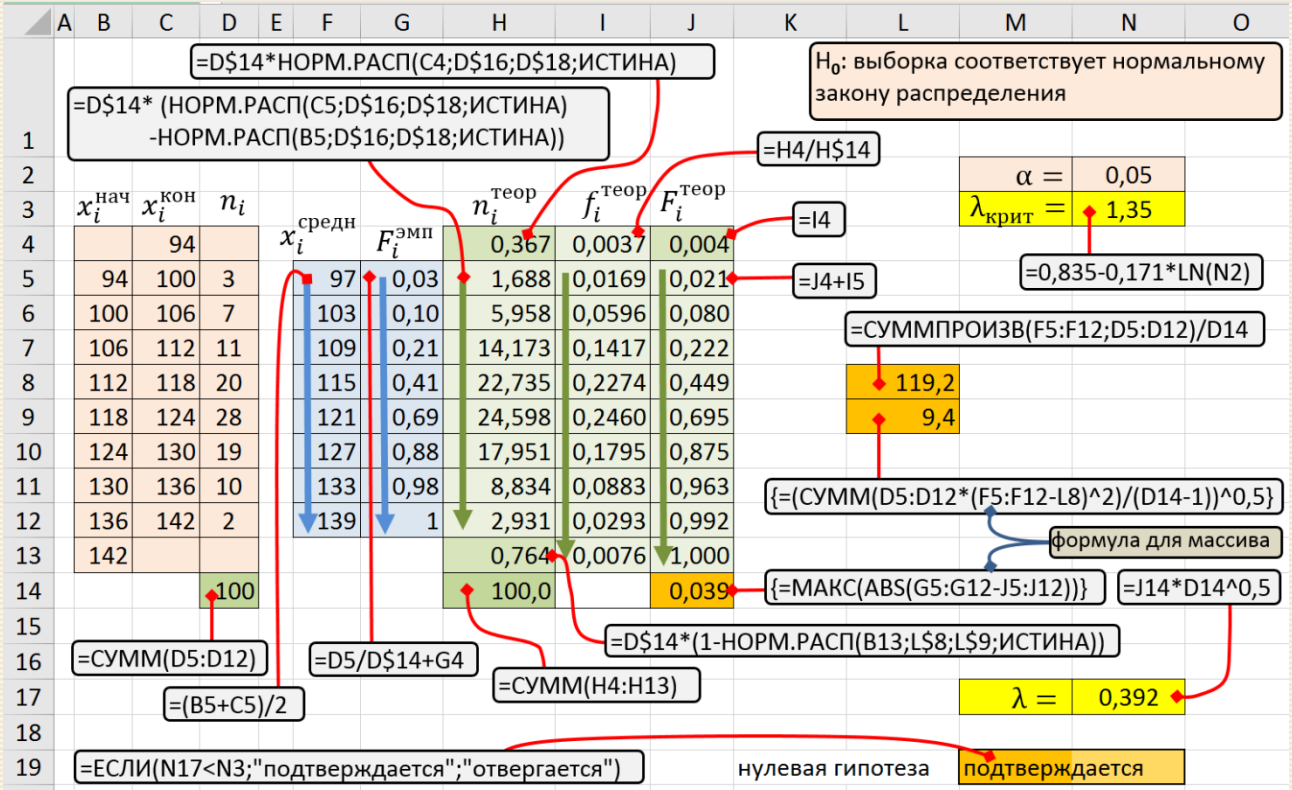


Рис. 3.5. Скриншот проверки выборки "на нормальность" по критерию Колмагорова

3.3. Критерий Колмагорова-Смирнова

Критерий Колмогорова-Смирнова использует ту же самую идею, что и критерий Колмогорова, но если в критерии Колмогорова сравнивается эмпирическая функция распределения с теоретической, то в критерии Колмогорова-Смирнова сравниваются две эмпирические функции распределения.

При проверке гипотезы о совпадении распределений в двух независимых выборках статистика критерия Смирнова $D_{m,n}$ определяется как максимум модуля разности между одной эмпирической функцией $F_1(x)$, построенной по выборке $x_1, x_2, \dots, x_n$, и другой эмпирической функцией $F_2(x)$, построенной по выборке $y_1, y_2, \dots, y_m$

$$\lambda = \sqrt{\frac{mn}{m+n}} \times \max_x |F_1(z) - F_2(z)|,$$

где z – разряды, по которым рассчитываются разницы частотных распределений $F_1(z)$ и $F_2(z)$.

Николай Васильевич Смирнов

При справедливости гипотезы H_0 статистика $\lambda < \lambda_{крит}$ имеет асимптотическое распределение Колмогорова, а $\lambda_{крит}$ определяется из $P\{\lambda > \lambda_\alpha\} = \alpha$, где α – уровень значимости.

Пример 3.5 Проанализировать на уровне значимости 0.05 равенство двух эмпирических выборок X и Y, заданных вариационными частотными рядами (диапазон ячеек B6:E11 рис. 3.6). Алгоритм вычислений аналогичен использованному для решения примера 3.4.

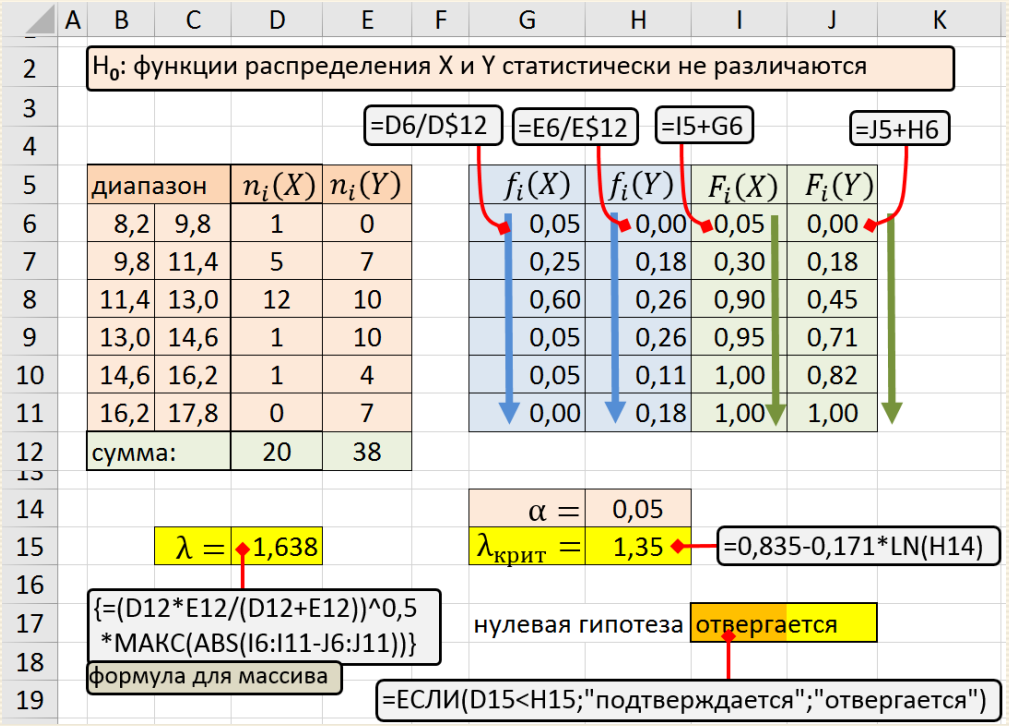


Рис. 3.6. Анализ равенства выборок критерием Колмагорова-Смирнова

3.4. Критерий Крамера-Мизеса-Смирнова

Пример 3.6 При испытаниях деталей из алюминиевого сплава получены значения относительного сужения имеющихся в них отверстий:

0.320	0.327	0.390	0.409	0.285	0.292	0.305	0.308	0.252	0.420
0.430	0.261	0.310	0.360	0.298	0.299	0.313	0.315	0.290	0.340

Проверить гипотезу о нормальном распределении относительного сужения уровне значимости 0.1.

Richard von Mises

Для анализа нормальности распределения используется критерий омега-квадрат.

Критерий омега-квадрат (ω^2, иначе – критерий Крамера-Мизеса-Смирнова) достаточно надёжен при объеме выборки $n \geq 15$ для проверки гипотезы, подчиняется ли случайная величина некоторому закону распределения, если известны (предполагаются известными) его параметры (так называемая простая гипотеза). Критерий основан на расчёте суммы квадратов разностей между накопленной частостью (эмпирической функцией распределения) и теоретической функцией распределения.

Статистику критерия для нормального распределения, расположив результаты (отсортировав их по возрастанию) в вариационном ряду, можно рассчитать по зависимости:

$$\omega^2 = \left(1 + \frac{1}{2n}\right)\left[\frac{1}{12n} + \sum_{i=1}^{n}(F(x_i) - W)^2\right].$$

Здесь $F(x_i)$ – значения функции нормального распределения, которое рассчитывается интегральной Excel-функцией

=НОРМ.РАСП(<x_i>; <$\overline{x}$>; <S >; ИСТИНА)

с параметрами $\overline{x}$ и S соответствующих оценок генеральной совокупности; $W = (2i - 1)/2n$ – величина накопленной частости.

Расчётную величину статистики сравнивают с табличным* $\omega^2_{\text{крит}}$. Сама величина $\omega^2_{\text{крит}}$ зависит от вида распределения; для нормального (гауссовского) распределения они имеют значения

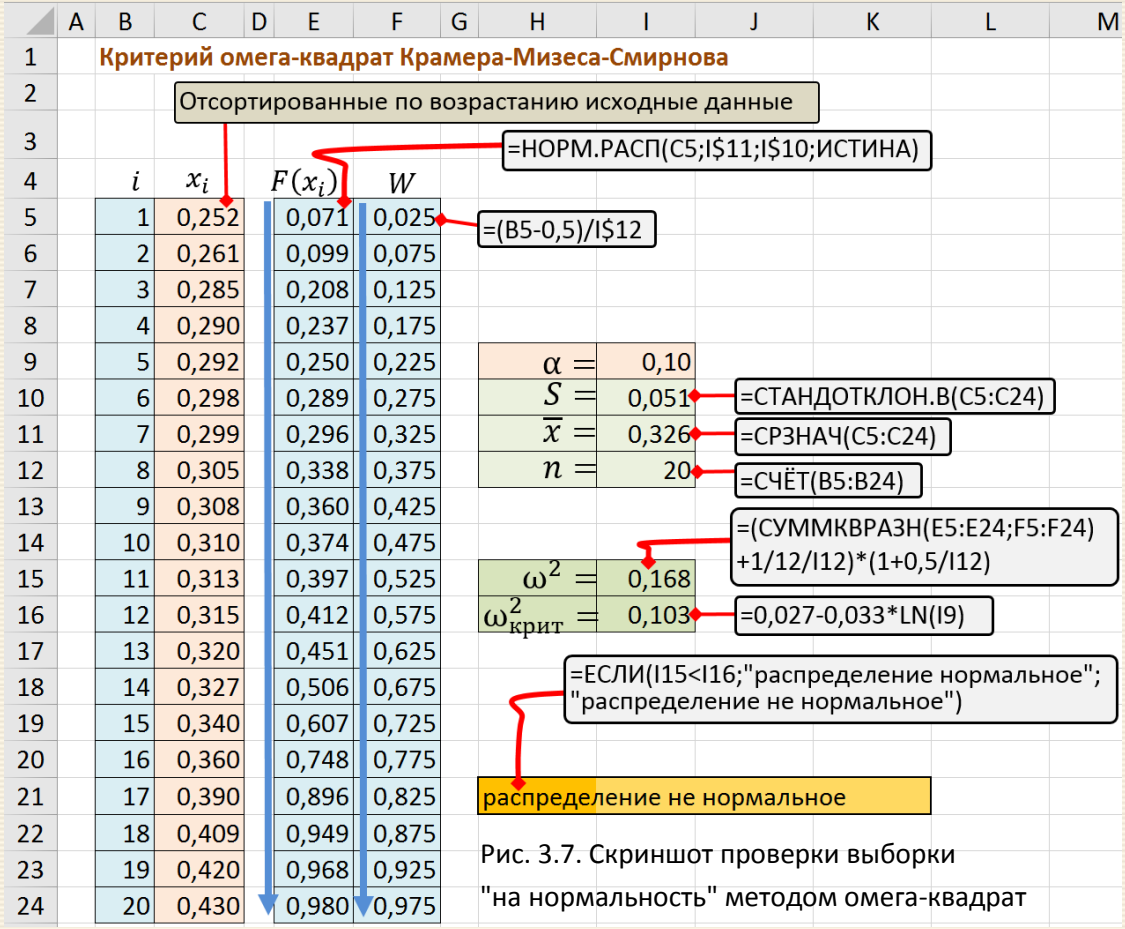
α	0.1	0.05	0.01	0.005	0.001
$\omega^2_{\text{крит}}$	0.1035	0.126	0.1788	0.2018	0.2559

которая с ошибкой не выше 0.5% аппроксимируется зависимостью $\omega^2_{\text{крит}} = 0.027 - 0.033 \cdot \ln(\alpha)$.

Если $\omega^2 < \omega^2_{\text{крит}}$, то нулевая гипотеза подтверждается, т.е. распределение считается соответствующим нормальному на выбранным уровнем значимости α.

Скриншот проверки выборки "на нормальность" для метода омега-квадрат (критерия Крамера-Мизеса-Смирнова) дан на рис. 3.7.

*Мартынов Г. В. Критерии омега-квадрат. М.: Наука, 1978.



i	x_i	$F(x_i)$	W
1	0,252	0,071	0,025
2	0,261	0,099	0,075
3	0,285	0,208	0,125
4	0,290	0,237	0,175
5	0,292	0,250	0,225
6	0,298	0,289	0,275
7	0,299	0,296	0,325
8	0,305	0,338	0,375
9	0,308	0,360	0,425
10	0,310	0,374	0,475
11	0,313	0,397	0,525
12	0,315	0,412	0,575
13	0,320	0,451	0,625
14	0,327	0,506	0,675
15	0,340	0,607	0,725
16	0,360	0,748	0,775
17	0,390	0,896	0,825
18	0,409	0,949	0,875
19	0,420	0,968	0,925
20	0,430	0,980	0,975

$\alpha =$	0,10
$S =$	0,051
$\overline{x} =$	0,326
$n =$	20
$\omega^2 =$	0,168
$\omega^2_{крит} =$	0,103

Рис. 3.7. Скриншот проверки выборки "на нормальность" методом омега-квадрат

3.5. Метод моментов

Люди делятся на три категории:
умеющие считать и не умеющие считать.
Мерфология, Закон Уинкорна

В случае выборок небольшого объема для проверки гипотезы о соответствии выборочного распределения нормальному закону можно использовать простые критерии, основанные на сравнении генеральных параметров распределения и их оценок, полученных по выборке. В качестве оцениваемых параметров удобнее всего брать моменты распределения – асимметрию и эксцесс.

Асимметрия (A) – это мера несимметричности графика плотности реального распределения в сравнении с нормальным распределением. Оценивается соотношением

$$A = \frac{1}{nS^3}\sum_{i=1}^{n}(x_i - \overline{x})^3,$$

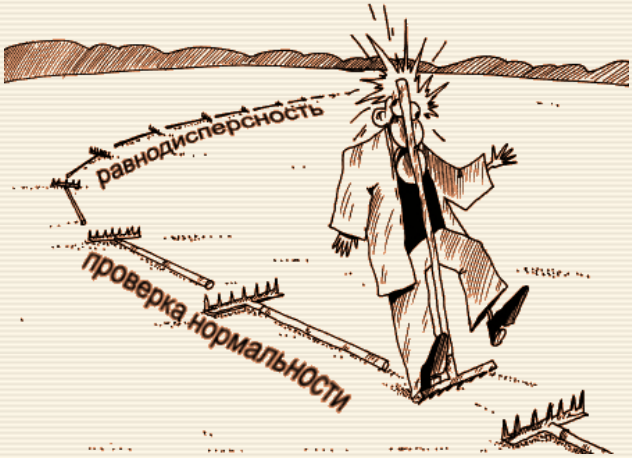
Эксцесс (E) показывает меру вытянутости графика плотности реального распределения в сравнении с нормальным распределением. Оценивается соотношением

$$E = \frac{1}{nS^4}\sum_{i=1}^{n}[(x_i - \overline{x})^4 - 3],$$

Выше использованы следующие обозначения: n – численность выборки, x_i, i=1,2…,n – значения вариант выборки, $\overline{x}$ – выборочное среднее значение, S – стандартное отклонение.

По численным значениям асимметрии A и эксцесса E и их дисперсии D_A и D_E можно приближенно оценить нормальность распределения результатов испытаний.

Если $|E| \le 5\sqrt{D_E}$, а $|A| \le 3\sqrt{D_A}$, то наблюдаемое распределение можно считать нормальным.



В Excel асимметрию A и эксцесс E можно вычислить при помощи статистических функций СКОС (для A) и ЭКСЦЕСС (для E).

Дисперсию эксцесса D_E можно определить формулой

$$D_E = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)},$$

дисперсия D_A коэффициента асимметрии A выборочной совокупности вычисляется по соотношению:

$$D_A = \frac{6(n-1)}{(n+1)(n+3)}.$$

Пример 3.7 Задача: установить – соответствует или нет выборка нормальному распределению

32.30	32.61	32.63	32.28	31.60	32.48	32.68	32.26	31.70	32.47
31.74	32.29	32.36	32.46	32.17	32.28	32.92	32.74	32.25	31.73

На рис. 3.8 приведен скриншот выполненного анализа соответствия выборки нормальному распределению. Оценка нормальности проводится методом моментов.

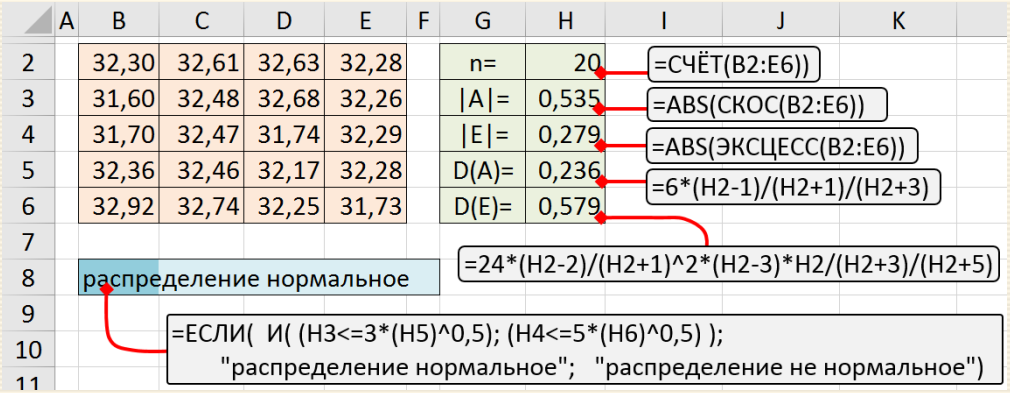
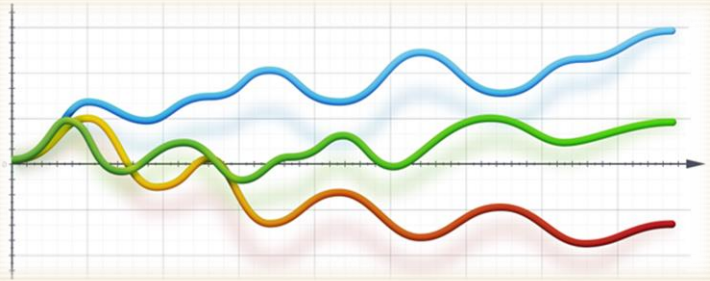


Рис. 3.8. Скриншот проверки малой выборки "на нормальность"



4. Параметрические критерии

Нельзя заранее правильно определить,
какую сторону бутерброда мазать маслом.
Мерфология, Закон своенравия природы

4.1. Критерий Стьюдента

Традиционный метод проверки однородности (критерий Стьюдента) позволяет найти вероятность того, что средние значения в выборках относятся к одной и той же совокупности.

William Sealy Gosset

Стьюдента t-критерий был разработан Уильямом Госсетом для оценки качества пива в компании Гиннесс. В связи с коммерческой тайной статья Госсета вышла в 1908 году в журнале "Биометрика" под псевдонимом "Student" (студент).

Стьюдента t-критерий – общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стьюдента. Наиболее частые случаи применения t-критерия связаны с проверкой равенства средних значений $\overline{x} = \overline{y}$ в двух выборках $X = \{x_1, x_2, \dots, x_{n_x}\}$ и $Y = \{y_1, y_2, \dots, y_{n_y}\}$.

Обычно t-статистика строится по следующему общему принципу: в числителе случайная величина с нулевым математическим ожиданием (при выполнении нулевой гипотезы), а в знаменателе – выборочное стандартное отклонение этой случайной величины, получаемое как квадратный корень из несмещенной оценки дисперсии.

К л а с с и ч е с к и е у с л о в и я п р и м е н и м о с т и к р и т е р и я С т ь ю д е н т а . Согласно математической теории статистики должны быть выполнены два условия применимости критерия Стьюдента, основанного на использовании статистики t:

а) результаты наблюдений имеют н о р м а л ь н ы е р а с п р е д е л е н и я с математическими ожиданиями μ_x и μ_y и выборочными дисперсиями D_x и D_y в первой и во второй выборках соответственно;

б) дисперсии результатов наблюдений в первой и второй выборках совпадают: $D_x = D_y$.

Если условия *а)* и *б)* выполнены, то нормальные распределения выборок отличаются только математическими ожиданиями* и гипотеза H_0 формулируется как $H_0 : \mu_x = \mu_y$, а альтернативная как $H_1 : \mu_x \neq \mu_y$.

Если хотя бы одно из условий *а)* и *б)* не выполнено, то нет оснований считать, что статистика t имеет распределение Стьюдента, поэтому применение классического метода, строго говоря, не обосновано.

*Математическое ожидание – среднее (в пределе) значение случайной величины

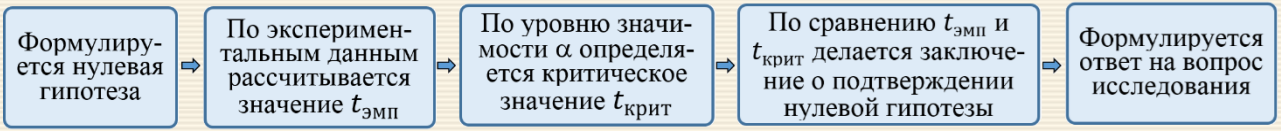
Специальным случаем является выполнение только условия *а)* – нормальности распределения, когда можно использовать "модификацию" распределения Стьюдента – критерий Беренса-Фишера.

При использовании критерия Стьюдента выделяется случай его применения для проверки гипотезы о равенстве средних для одной и той же группы объектов – так называемый парный *t*-критерий. Выборки при этом называют зависимыми, связанными.

В случае сравнения генеральных средних двух независимых, несвязанных выборок (так называемый двухвыборочный *t*-критерий) анализируются, например, средние значения контрольной и экспериментальной (опытной) групп (выборок) разных объемов. Специальным случаем здесь является выполнение только условия *а)* – нормальности распределения, когда можно использовать "модификацию" распределения Стьюдента – критерий Беренса-Фишера.

В различного рода исследованиях возникают задачи сравнения результатов анализа с каким-либо значением, которое можно считать точной (паспортной) величиной. С математической точки зрения в таких случаях сравнение данных сводится к проверке значимости отличия случайной величины $\overline{x}$ от константы μ. Тест, предназначенный для сравнения среднего значения и константы, обычно называется простым тестом Стьюдента или критерием Стьюдента равенства среднего значения.

Общая схема исследования имеет вид:



Конкретная зависимость для расчета выбирается в соответствии с нижеследующей таблицей.

1	парный t-критерий одновыборочный критерий зависимые выборки $n_x = n_y = n$	$D_x = D_y$	$t = \lvert\overline{d}\rvert \sqrt{\frac{n \cdot df}{\sum_i d_i^2 - n\overline{d}^2}}$, $df = n - 1.$	$d_i = x_i - y_i$ разности между парами, $\overline{d}$ – среднее разностей,
2	двухвыборочный t-критерий – гомоскедастический тест two-group unpaired-test, независимые выборки	$D_x = D_y$	$t = \frac{\lvert\overline{x} - \overline{y}\rvert}{\overline{S}} \sqrt{\frac{n_x n_y}{n_x + n_y}}$, $\overline{S} = \sqrt{\frac{df_x \cdot D_x + df_y \cdot D_y}{df}}$, $df = df_x + df_y = (n_x + n_y - 2)$.	
3	двухвыборочный t-критерий – гетероскедастический тест, независимые выборки	$D_x \neq D_y$	$t = \frac{\lvert\overline{x} - \overline{y}\rvert}{\sqrt{\Omega_x + \Omega_y}}$, где $\Omega = \frac{D}{n}$, $df = \frac{(\Omega_x + \Omega_y)^2}{\Omega_x^2/df_x + \Omega_y^2/df_y}$.	
4	критерий Стьюдента равенства среднего значения, μ – заданное математическое ожидание		$t = \frac{\lvert\overline{x} - \mu\rvert\sqrt{n}}{S}$, $df = n - 1$.	$S = \sqrt{D}$ – стандартное отклонение по выборке

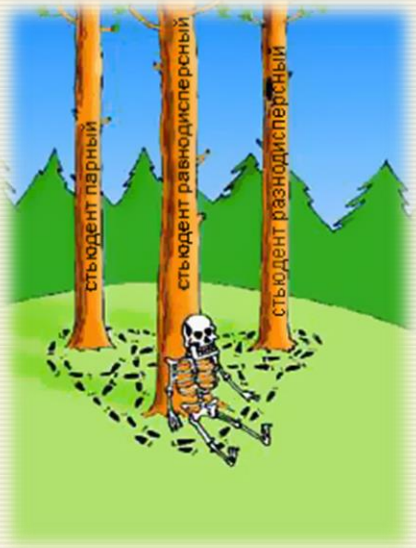
Значимое различие между $\overline{x}$ и $\overline{y}$ (и выборками *X* и *Y*) имеет место тогда, когда $t > t_{\text{крит}}(\alpha, df)$.

Пример 4.1 П а р н ы й критерий Стьюдента для зависимых выборок.

Задача: для оценки эффективности нового гипогликемического средства были проведены измерения уровня глюкозы в крови пациентов, страдающих сахарным диабетом, до и после приема препарата. В результате были получены следующие данные:

ДО	9,6	8,1	8,8	7,9	9,2	8,0	8,4	10,1	7,8	8,1
ПОСЛЕ	5,7	5,4	6,4	5,5	5,3	5,2	5,1	6,9	7,5	5,0

Имеются ли статистически значимые (на уровне значимости α=0.05) различия содержания глюкозы в крови до и после приема нового препарата?



Эмпирическое значение критерия Стьюдента вычисляется по соотношению

$$t = |\overline{d}| \sqrt{\frac{n \cdot df}{\sum_i d_i^2 - n\overline{d}^2}}$$

где $d_i = x_i - y_i$ разности между парами переменных,

$\overline{d}$ – среднее этих разностей,

$df = n - 1$.

Скриншот решения в MS Excel приведен на рис. 4.1.

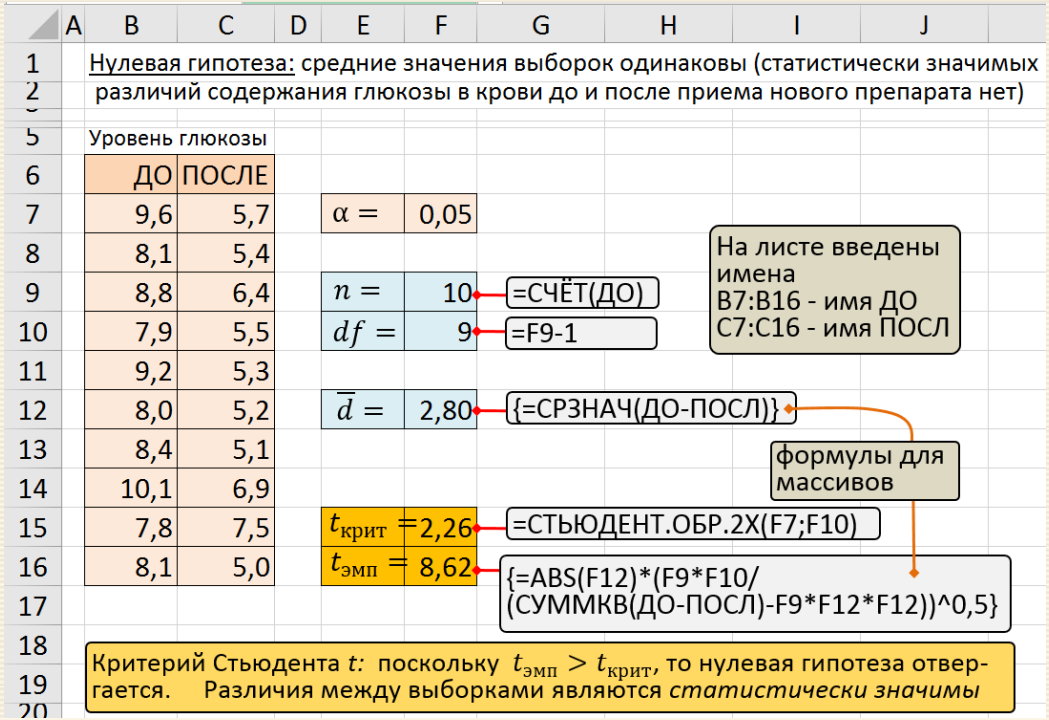


Рис. 4.1. Скриншот использования парного критерия Стьюдента для зависимых выборок

Пример 4.2 Выполнить сравнение средних значений двух независимых выборок на уровне значимости α=0.05. Используется критерий Стьюдента для независимых выборок равных дисперсий – гомоскедастический тест.

выборка *X*			
15	19	12	11
13	7	15	19
11	8	16	18
18	11	13	9
10	12	14	6
8	15	14	15
20	16	10	12
7	13	8	15
8	5		

выборка *Y*			
15	18	17	14
18	14	19	12
12	13	16	20
20	18	12	16
16	13	15	11
11	13	19	13
13	15	13	
7	18	12	
14	9	15	

Эмпирическое значение критерия Стьюдента вычисляется по соотношению

$$t = \frac{|\overline{x} - \overline{y}|}{\overline{S}} \sqrt{\frac{n_x n_y}{n_x + n_y}}, \quad \overline{S} = \sqrt{\frac{df_x \cdot D_x + df_y \cdot D_y}{df}},$$

$$df = df_x + df_y = (n_x + n_y - 2).$$

Скриншот решения в MS Excel приведен на рис. 4.2.

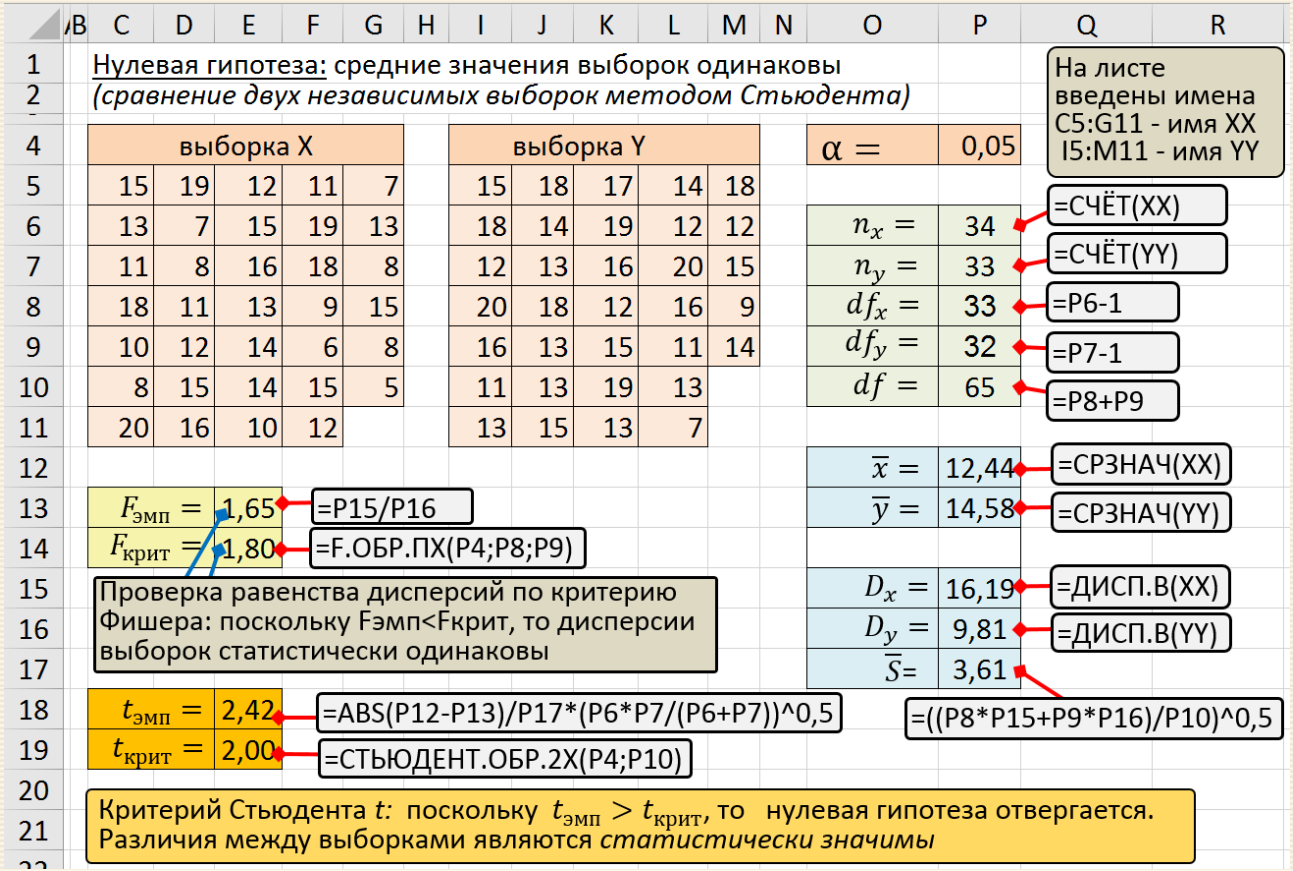


Рис. 4.2. Скриншот гомоскедастического теста критерия Стьюдента

Пример 4.3 (критерий Стьюдента для независимых выборок разных дисперсий – гетероскедастический тест). Задача: выполнить сравнение средних значений двух независимых выборок на уровне значимости α=0.05.

выборка X				11.1	14.1
11.4	11.9	11.5	11.6	12.0	11.5
12.4	12.1	12.6	12.1	12.5	12.2
8.2	10.1	10.7	10.4	11.3	14.8

выборка Y				17.7	11.4
14.3	14.4	14.9	14.3	17.5	17.5
10.8	11.4	16.3	16.1	11.4	11.9
12.5	12.2	17.0	16.6	12.3	17.3
13.0	14.4	14.1	13.9	12.0	13.5
10.0	10.0	10.0	10.0	13.2	13.9
15.8	12.1	10.0	10.0	12.0	15.5

Эмпирическое значение критерия Стьюдента вычисляется по соотношению $t = \frac{|\overline{x} - \overline{y}|}{\sqrt{\Omega_x + \Omega_y}}$, $df = \frac{\left(\Omega_x + \Omega_y\right)^2}{\Omega_x^2/df_x + \Omega_y^2/df_y}$, где $\Omega = \frac{D}{n}$.

Скриншот решения в MS Excel приведен на рис. 4.3.

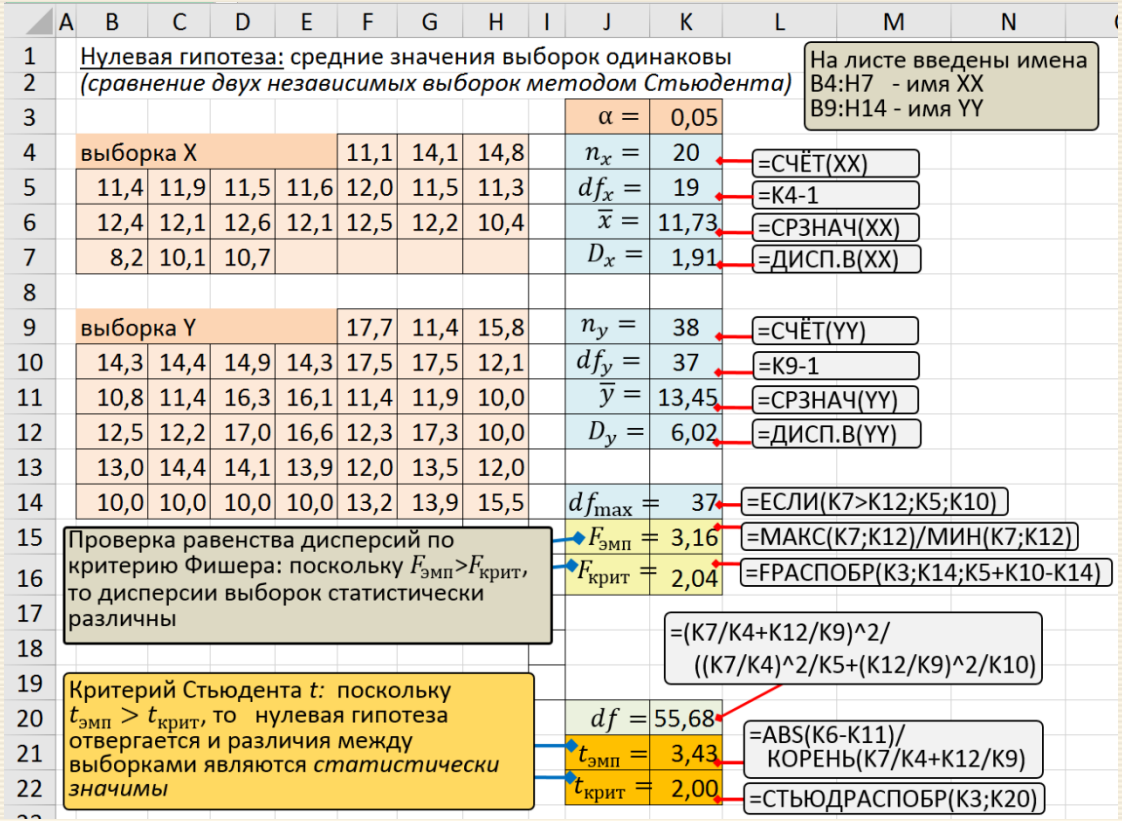


Рис. 4.3. Скриншот гетероскедастического теста критерия Стьюдента

Пример 4.4 Критерий Стьюдента равенства среднего значения. Задача: при определении никеля в стандартном образце сплава получена серия значений (% масс.) 12.11; 12.44; 12.32; 12.28; 12.42. Содержание никеля согласно паспорту образца – 12.38%. Содержит ли использованная методика систематическую погрешность? Проверку провести на уровне значимости 0.05.

При решении паспортное содержание никеля (12.38) считаем действительным (точным) значением и применяем простой тест Стьюдента. Вычисляем все необходимые для расчета $n, \overline{x}, S$ параметры (рис. 4.4) и сравниваем значения t-статистики. Поскольку отличие результата анализа от действительного значения на уровне α=0.05 незначимо $t_{\text{эмп}} = 1.116 < t_{\text{крит}} = 2.776$, методика не содержит систематической погрешности.

Эмпирическое значение критерия Стьюдента вычисляется по соотношению

$$t = \frac{|\overline{x} - \mu|\sqrt{n}}{S}, \qquad S = \sqrt{D}\,, \quad df = n - 1\,.$$

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Нулевая гипотеза: среднее значение проб никеля								На листе введено имя C5:C9 - имя NIK				
2	совпадает с паспортным значением												
4			% Ni		$\alpha =$	0,05			Критерий Стьюдента t: поскольку $t_{эмп} < t_{крит}$, то нулевая гипотеза подтверждается, то есть нет статистически значимых различий среднего содержания проб и паспортного значения: методика не содержит систематической погрешности				
5			12,11										
6			12,44		$n =$	5	=СЧЁТ(NIK)						
7			12,32		$\overline{x} =$	12,314	=СРЗНАЧ(NIK)						
8			12,28		$S =$	0,132	=СТАНДОТКЛОН.В(NIK)						
9			12,42										
10					$t_{эмп} =$	1,116	=ABS(F7-C11)/F8*КОРЕНЬ(F6)						
11		$\mu =$	12,38		$t_{крит} =$	2,776	=СТЬЮДЕНТ.ОБР.2Х(F4;F6-1)						

Рис. 4.4. Скриншот решения по сопоставлению среднего значения паспортному

4.2. *Z*-критерий

При проверке гипотезы вместо t–критерия можно использовать z–критерий, основанный на нормированном нормальном распределении статистики критерия. В этом случае вычисляется статистика

$$z = \frac{|\overline{x} - \mu|}{S/\sqrt{n}}$$

и сравнивается с ее критическим значением для нормированного нормального распределения при заданном уровне значимости ($\overline{x}$ – случайная величина выборочного среднего, μ — значение математического ожидания, $S/\sqrt{n}$ – стандартная ошибка этой величины).

Проверка гипотезы о равенстве долей в двух совокупностях

Для проверки гипотезы о равенстве долей в двух совокупностях используют параметрический критерий, который носит название в ряде источников как z-критерий. Критерий может быть использован для зависимых и независимых выборок.

Суть: имеются две совокупности, генеральные доли признака в которых равны соответственно p_1 и p_2. Необходимо проверить нулевую гипотезу (о равенстве генеральных долей): $p_1 = p_2$, то есть различие долей исследуемых признаков в генеральных совокупностях статистически незначимо. Альтернативная гипотеза – различие долей исследуемых признаков в генеральных совокупностях статистически значимо.

Для проверки нулевой гипотезы из этих совокупностей взяты две независимые выборки достаточно большого объема n_1 и n_2.

Выборочные доли признака равны соответственно

$$\omega_1 = \frac{m_1}{n_1} \quad \text{и} \quad \omega_2 = \frac{m_2}{n_2},$$

где m_1 и m_2 – соответственно число элементов первой и второй выборок, обладающих данным признаком.

Ограничения критерия: должно выполняться неравенство

$$[n_1\omega_1,\ n_1(1-\omega_1),\ n_2\omega_2,\ n_2(1-\omega_2)] > 5\,.$$

Уровень значимости получаемого результата равен α (достоверность, или доверительная вероятность результата равна $1-\alpha$).

Для проверки гипотезы используется интеграл Лапласа, то есть функция стандартного нормального распределения

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt\,.$$

В MS Excel ее значение можно вычислить функцией НОРМ.СТ.ОБР $(1-2\alpha)$.

Pierre-Simon de Laplace

Статистика критерия рассчитывается по соотношению

$$u_{\text{эмп}} = \frac{|\omega_1 - \omega_2| - \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \qquad \text{где} \qquad p = \frac{m_1 + m_2}{n_1 + n_2}.$$

Если $u_{\text{эмп}} < u_{\text{крит}}$, то нулевая гипотеза принимается. Значение $u_{\text{крит}}$ определяется из соотношения

$$\Phi\left(u_{\text{крит}}\right) = 1 - 2\alpha$$

через Excel-функцию НОРМ.СТ.ОБР$(1 - 2\alpha)$.

Доверительный интервал I, содержащий разность между долями в двух независимых группах, определяется соотношением

$$I = (p_1 - p_2) \pm Z\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}.$$

Пример 4.5 Сравнить на уровне значимости α=0.05 эффективность двух различных методик лечения, когда с использованием первой методики из 61 больного не излечилось 8 больных, а при применении второй методики из 67 больных остались больными 10.

Нулевая гипотеза: различие двух методик лечения на уровне значимости α = 0.05 статистически незначимо. Скриншот вычислений по проверке гипотезы о равенстве долей двух выборок дан на рис. 4.5.

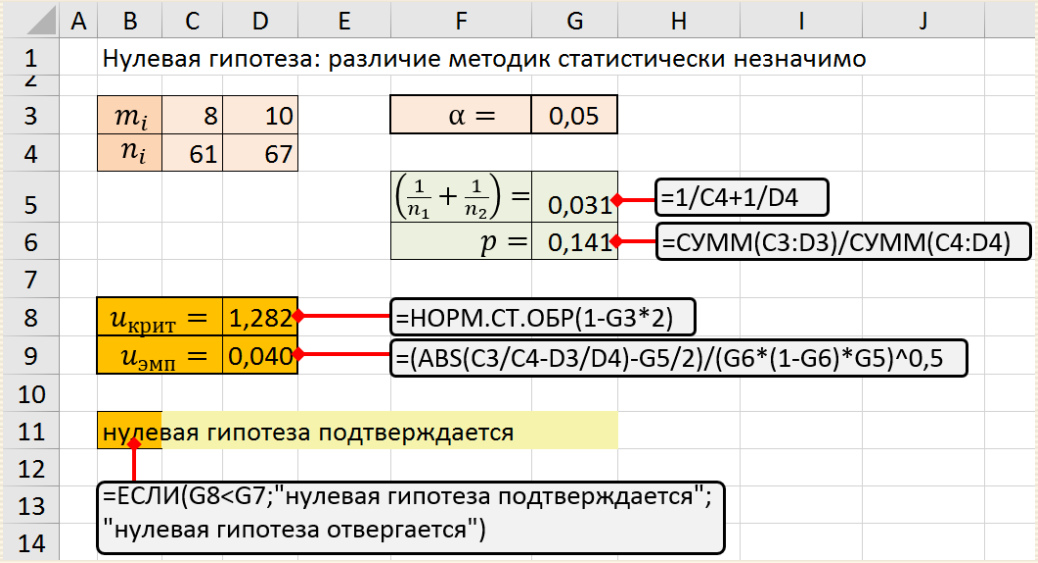


Рис. 4.5. Скриншот решения по проверке гипотезы о равенстве долей двух выборок

Проверка гипотезы о равенстве долей в двух и более совокупностях

Рассматриваются k совокупностей, генеральные доли признака в которых равны соответственно $p_i.(i = 1,2, \dots, k)$. Необходимо проверить нулевую гипотезу о равенстве генеральных долей, то равенства $p_1 = p_2 = \cdots = p_k$. Для проверки нулевой гипотезы из данных совокупностей берутся независимые выборки достаточно большого объема $n_i.(i = 1,2, \dots, k)$.

Выборочные доли признака равны $\omega_i = m_i/n_i$ $(i = 1,2,\dots,k)$, где m_i – число элементов в i-й выборке, обладающих данным признаком.

Уровень значимости получаемого результата равен α, тогда достоверность, или доверительная вероятность результата равна $1 - \alpha$.

Для проверки гипотезы о равенстве долей в совокупностях используют критерий χ^2, то есть функцию распределения для уровня значимости α и числа степеней свободы $df = k - 1$.

Статистика критерия рассчитывается по соотношению

$$\chi^2_{\text{эмп}} = \frac{1}{p(1-p)} \sum_{i=1}^{k} n_i(\omega_i - p)^2 \qquad \text{где} \qquad p = \frac{\sum m_i}{\sum n_i}.$$

Нулевая гипотеза формулируется следующим образом: $p_1 = p_2 = \cdots = p_k$. Если $\chi^2_{\text{эмп}} < \chi^2_{\text{крит}}$, то нулевая гипотеза принимается. Значение $\chi^2_{\text{крит}}$ определяется для числа степеней свободы $df = k - 1$ и определяется Excel-функцией =ХИ2.ОБР.ПХ(α; df).

Пример 4.6 Контрольную работу по высшей математике по индивидуальным вариантам выполняли студенты четырёх групп. В первой группе было предложено 105 задач, из которых было правильно решено 60, во второй группе из 140 предложенных задач было верно решено 69 задач, в третьей группе – 125 задач, из них правильно было решено 63 задачи, в четвертой группе – 160 и 105 соответственно. На уровне значимости α = 0.05 необходимо выяснить, можно ли считать, что различия в усвоении учебного материала студентами четырех групп первого курса существенны.

Нулевая гипотеза: уровень усвоения материала студентами различных групп статистически не различается.

Скриншот вычислений по проверке гипотезы о равенстве долей двух выборок дан на рис. 4.6.

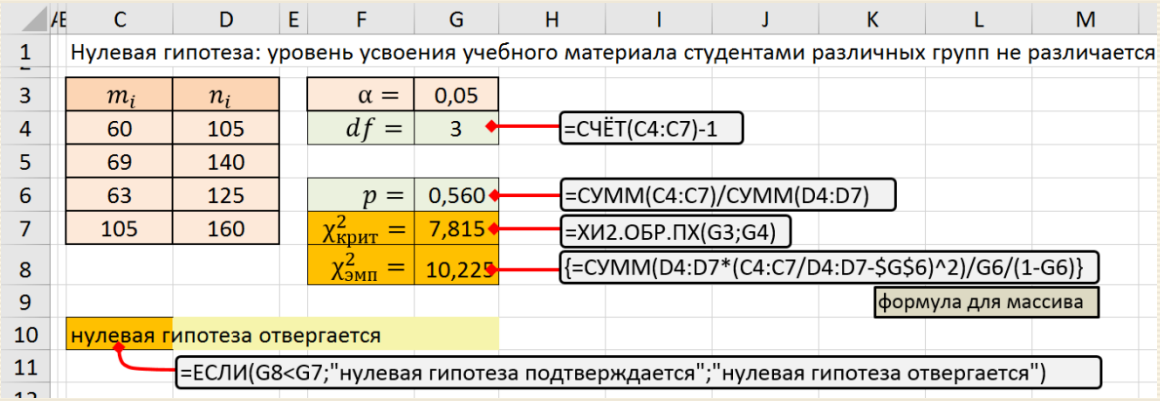


Рис. 4.6. Скриншот решения по проверке гипотезы о равенстве долей двух выборок

Поскольку $\chi^2_{эмп} > \chi^2_{крит}$ (10.225 >7.815), то нулевая гипотеза отвергается.

Таким образом, полученные в опыте данные противоречат гипотезе об одинаковом уровне усвоения учебного материала студентами всех четырех групп. Другими словами – с достоверностью 95% уровень усвоения учебного материала студентами различных групп различается.

4.3. Однофакторный анализ ANOVA

Анализ дисперсии (ANOVA, **AN**alysis **O**f **VA**riance, one-way ANOVA) определяет наличие различий между групповыми средними; подход был разработан сэром Рональдом Фишером. В простой форме ANOVA обеспечивает статистическое испытание равновеликости средних нескольких групп и некоторым образом обобщает t-критерий на три и более группы данных; позволяет проверить гипотезу о существовании влияния изучаемого фактора на зависимую переменную.

Математическая модель однофакторного ANOVA предполагает выделение в общей изменчивости зависимой переменной двух ее составляющих: м е ж г р у п п о в а я (sum of squares between groups) составляющая изменчивости обусловлена различием средних значений под влиянием фактора; в н у т р и г р у п п о в а я (sum of squares within groups) составляющая изменчивости обусловлена влиянием неучтенных причин. Соотношение этих двух составляющих изменчивости и есть основной показатель, определяющий статистическую значимость влияния фактора.

Sir Ronald Aylmer Fisher

Пусть имеется k выборок, которые будут называться группами или сериями (столбцы анализа данных, соответствующие 4 методам в рассматриваемом ниже примере). Для этих столбцов используется индекс j. Каждая группа состоит из данных по группу (серии) размера n_j.

Элементами рассматриваемой выборки являются строки столбца, для которых используется индекс i. Таким образом, образец j-я группа есть множество $\{x_{1j}, x_{2j}, \dots, x_{n_j j}\}$, а общая выборка состоит из всех элементов x_{ij} $(i = 1, \dots, n_j;\ \ j = 1, \dots, k)$, количество которых n определяется суммой $n = \sum n_j$. Параметры анализируемых данных можно свести в следующую таблицу

дисперсия	SS	df	MS
B Between Groups межгрупповая	$\sum_{j=1}^{k} n_j(\overline{x}_j - \overline{x})^2$	$k-1$	$\frac{SS_B}{df_B}$
W Within Groups внутригрупповая	$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$	$n-k$	$\frac{SS_W}{df_W}$
T Total общая	$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x})^2$	$n-1$	$\frac{SS_T}{df_T}$

Здесь SS – сумма квадратов; $MS = SS/df$ – средний квадрат;

$\overline{x} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}$ – общее среднее; $\overline{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$ – групповое среднее.

Можно отметить, что MS_B является мерой вариабельности группового средства вокруг общего среднего. MS_W (mean squared error) является мерой вариабельности каждой группы вокруг ее среднего значения, а также мерой общей изменчивости. Если нулевая гипотеза истинна, то MS_W и MS_B являются мерами одной и той же ошибки, поэтому можно ожидать, что $F = MS_B/MS_W$ будет около единицы. Если нулевая гипотеза ложна, то можно ожидать $F > 1$.

Нулевая гипотеза заключается в том, что любая разница между группами (сериями) обусловлена случайностью. Если нулевая гипотеза верна, это означает, что математические ожидания всех k групп равны.

Основным показателем для принятия решения о подтверждении нулевой гипотезы является сравнение F-критерия Фишера с его критическим значением на заданном уровне значимости α.

$$F = \frac{MS_B}{MS_W}, \quad F_{\text{крит}} = F(\alpha, df_B, df_W).$$

При анализе с уровнем значимости α если уровень ошибки выше или равен α ($P_{\text{value}} \geq \alpha$), подтверждается гипотеза о равенстве средних значений. При уровне ошибки меньшей α (т.е. $P_{\text{value}} < \alpha$) подтверждается гипотеза о различие по крайней мере двух средних значений.

Ограничения метода*:* 1) дисперсии выборок должны быть однородны; что проверяется каким-либо соответствующим критерием; 2) формально численность любой группы не должна быть меньше двух объектов.

Альтернативный подход – сравнение независимых выборок по критерию H Краскела-Уоллеса.

Пример 4.7 В разных классах используется четыре метода обучения. Требуется определить наличие существенной разницы между показателями усвоения материала на уровне значимости 0.05; результаты тестирования* даны на рис. 4.7 в области B4:E11.

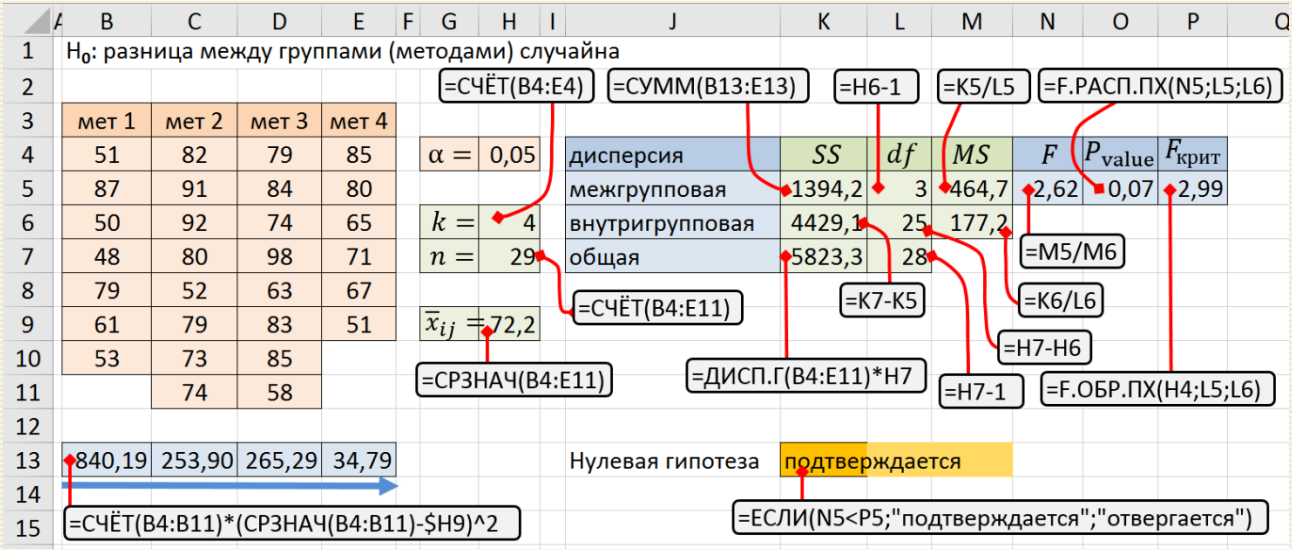


Рис. 4.7. Однофакторный анализ ANOVA проверки равенства показателей усвоения материала

Из результатов видно, что $F = 2.62 < F_{\text{крит}} = 2.99$ и $P_{\text{value}} = 0.07 > \alpha = 0.05$; посему отклонить нулевую гипотезу нельзя и можно заключить, что существенной разницы между методами нет.

*http://www.real-statistics.com/one-way-analysis-of-variance-anova/basic-concepts-anova

В MS Excel через надстройку "Анализ данных" (скриншоты на рис. 4.8) в соответствии с алгоритмами ANOVA можно получить результаты однофакторного дисперсионного анализа (рис. 4.9).

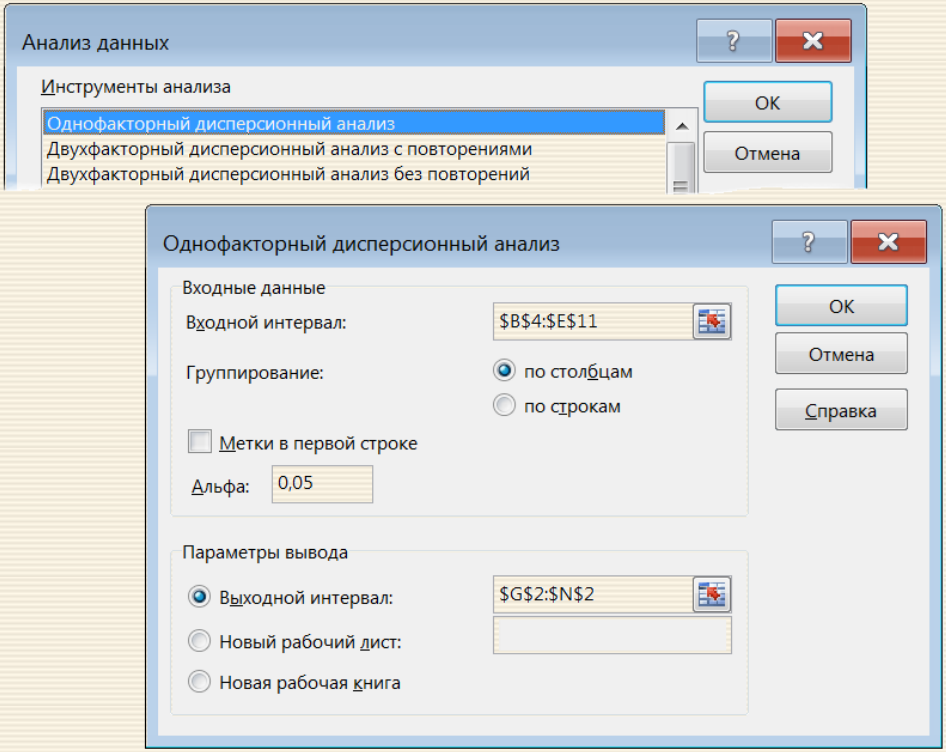
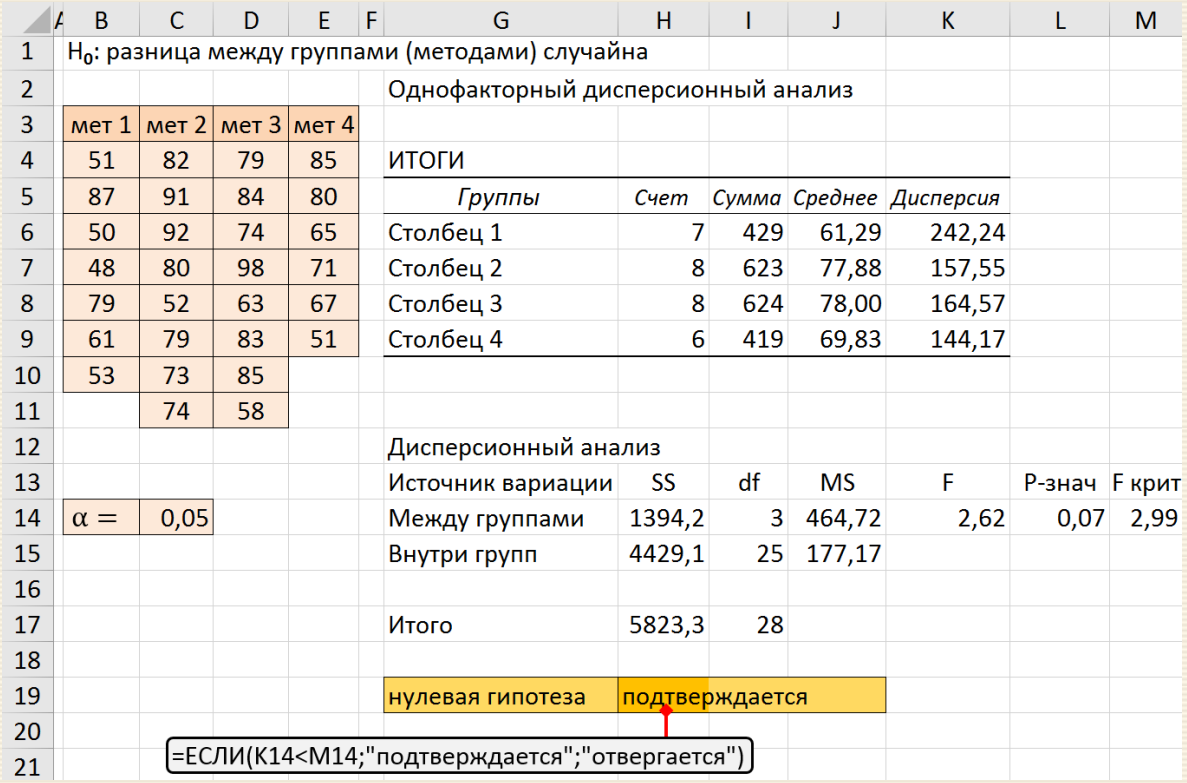


Рис. 4.8. Скриншот выбора инструментария однофакторного дисперсионного анализа



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		H_0: разница между группами (методами) случайна											
2							Однофакторный дисперсионный анализ						
3		мет 1	мет 2	мет 3	мет 4								
4		51	82	79	85		ИТОГИ						
5		87	91	84	80		*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
6		50	92	74	65		Столбец 1	7	429	61,29	242,24		
7		48	80	98	71		Столбец 2	8	623	77,88	157,55		
8		79	52	63	67		Столбец 3	8	624	78,00	164,57		
9		61	79	83	51		Столбец 4	6	419	69,83	144,17		
10		53	73	85									
11			74	58									
12							Дисперсионный анализ						
13							Источник вариации	SS	df	MS	F	P-знач	F крит
14		α =	0,05				Между группами	1394,2	3	464,72	2,62	0,07	2,99
15							Внутри групп	4429,1	25	177,17			
16													
17							Итого	5823,3	28				
18													
19							нулевая гипотеза	подтверждается					
20													
21			=ЕСЛИ(K14<M14;"подтверждается";"отвергается")										

Рис. 4.9. Скриншот результатов применения инструментария дисперсионного анализа

4.4. Критерий множественных сравнений

Результатом использования инструмента ANOVA является вывод, что среднее по крайней мере для одной генеральной совокупности отличается от остальных, в то время как в исследованиях интересно знать, какие из групп значимо отличаются от других. Для корректного заключения о том, какое групповое среднее значимо отличаются от остальных, применяются так называемые статистические критерии множественных сравнений.

Можно отметить, что решения задачи путем выполнения попарного сравнения пар данных критерием Стьюдента не вполне корректно, поскольку данный критерий предполагает сравнение именно двух выборок.

Критерии, применяемые п о с л е использования ANOVA, отклонившего нулевую гипотезу, называются апостериорными критериями (post hoc tests — от лат. "после того, как"). Для справки – критерии, применяемые до использования ANOVA, именуются априорными критериями (a priori tests).

Апостериорный тест Шеффе (Scheffe's test) используется после отклонения ANOVA нулевой гипотезы и дает оценку значимости различий посредством модифицированного t-критерия Стьюдента. Статистика критерия Шеффе t_s вычисляется по соотношению

$$t_s = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{MS_W\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

где $\overline{x}_1, \overline{x}_2$ – сравниваемые средние значения анализируемых выборок объемами n_1 и n_2,

MS_W – внутригрупповой средний квадрат (см. таблицу и пример раздела 4.3).

Решение о значимости различий средних принимается либо по обычному условию $t_s > t_s^{\text{крит}}$, либо по малости P_{value} относительно уровня значимости α.

Критическое значение теста Шеффе $t_s^{\text{крит}}$ определяется через таковое для критерия Фишера $F_{\text{крит}}$ по соотношению

$$t_s^{\text{крит}} = \sqrt{df_B \cdot F_{\text{крит}}(\alpha, df_B, df_W)}\ ,$$

рассчитываемое для заданного уровня значимости α и соответствующих степеней свободы df_B и df_W.

Для оценки P-значения P_{value} можно использовать соотношение

$$P_{\text{value}} = F\left(\frac{t_s^2}{df_B}, df_B, df_W\right).$$

Тест Шеффе имеет те же допущения по использованию, что и ANOVA.

Пример 4.8 Партия расфасованных по 30 пакетам удобрений была распределена по трем различным методам хранения (A, B и C). После некоторого срока хранения определялось содержание влаги в каждом пакете, данные сведены в таблицу.

A	10.1	7.3	5.6	6.2	8.4	8.1	8.0	7.6	5.3	7.2			
B	11.7	12.2	11.8	7.8	8.9	9.9	12.4	11.0	10.3	13.8	10.5	9.8	9.1
C	10.2	12.0	8.8	8.7	10.5	11.0	9.1						

На уровне значимости 0.05 проверить гипотезу о том, что условия хранения продукта не оказывают влияния на содержание влаги.

На первом этапе решения проводится анализ ANOVA (см. раздел 4.3, пример 4.7), согласно которому нулевая гипотеза равенства средних отвергается (рис. 4.10а).

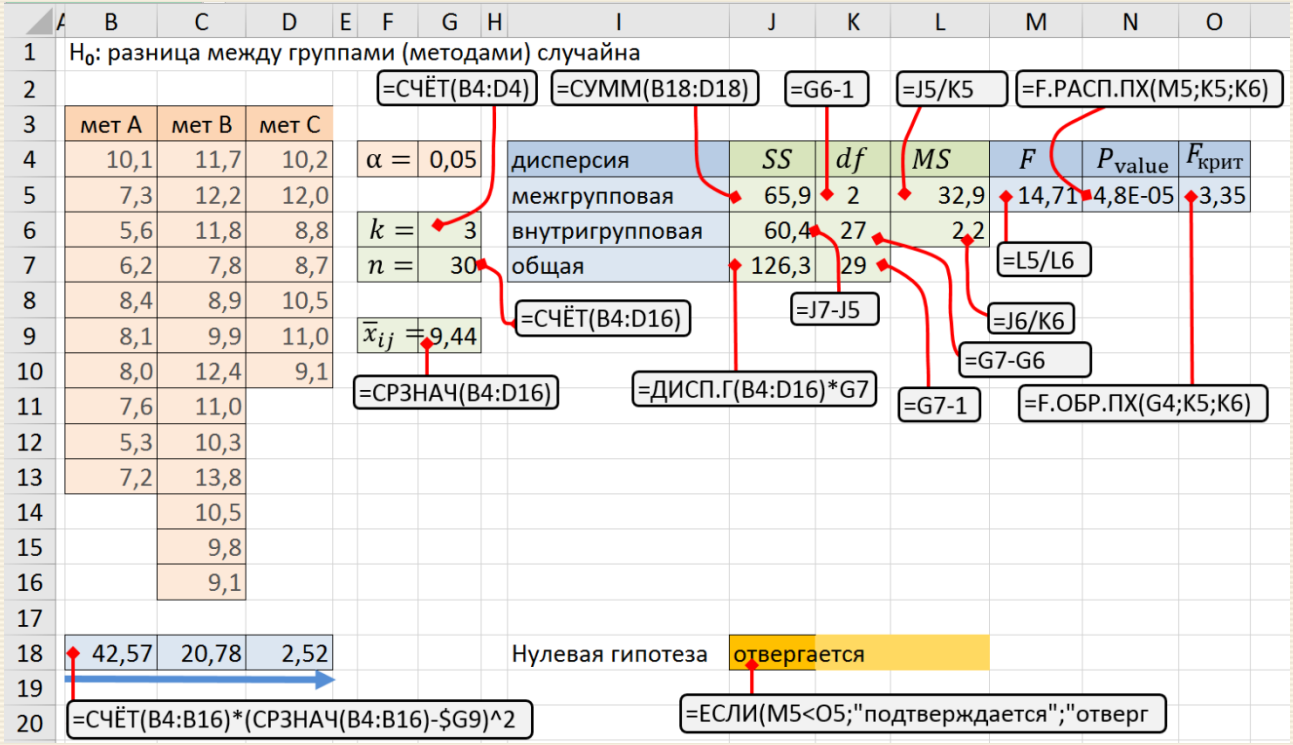


Рис. 4.10а. Однофакторный анализ ANOVA проверки равенства методов хранения A, B и C

На рис. 4.10b дан скриншот выполнения апостериорного (post hoc) теста Шеффе по оценке разницы методов хранения A, B и C, являющего "продолжением" анализа ANOVA и использующим ряд уже полученных (рис. 4.10a) результатов.

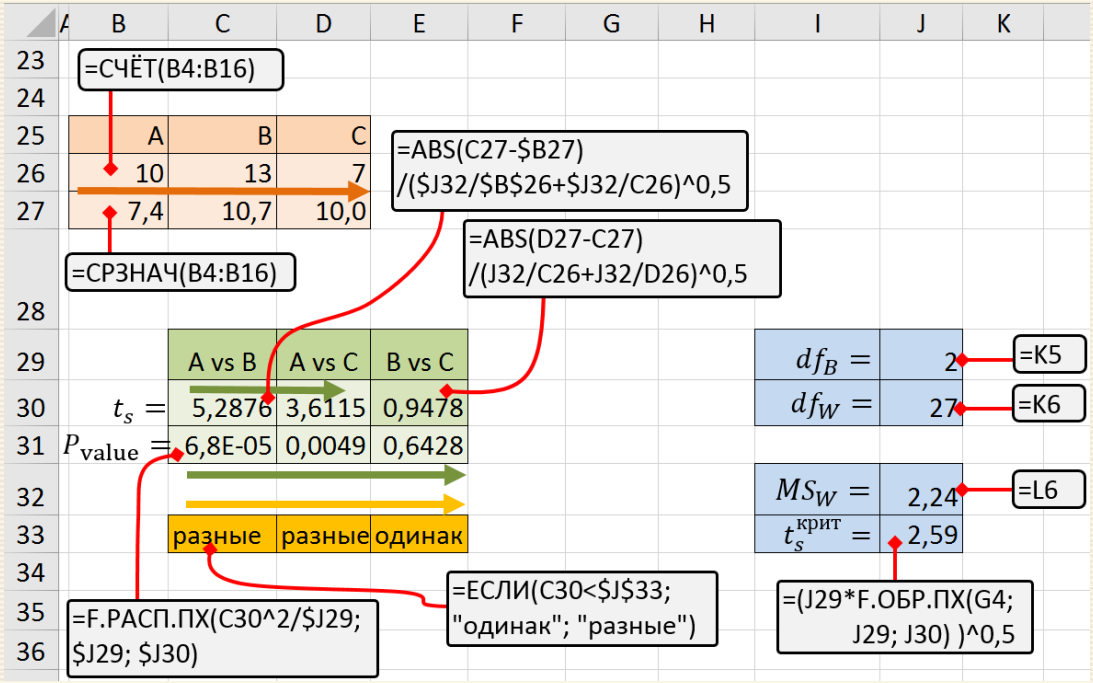


Рис. 4.10b. Применение теста Шеффе для оценки разницы методов хранения A, B и C

Далее рассчитываются количества данных (B26:D26) и средние значения (B27:D27) для трех методов; для наглядности переносятся значения df_B, df_W и MS_W. В ячейке J33 вычисляется критическое значение для теста и статистика парных сравнений A vs B , A vs C и B vs D*; в диапазоне ячеек C26:E26 рассчитываются значения P_{value} для каждой пары.

Результаты сравнения статистики теста с критическим значения выведены в ячейках C33:E33.

Исходя из проведенного анализа можно сделать вывод, что влияние метода хранения на отклик фактора влажности для методов B и C одинаково, а оба эти метода различаются со способом хранения A на уровне значимости 0.05.

Henry Scheffe

*vs (от лат. versus) – против

5. Непараметрические критерии

5.1. Критерий Крамера-Уэлча

Классические условия применимости критерия Стьюдента в подавляющем большинстве технических, экономических, медицинских и иных задач не выполняются. Тем не менее, при больших и примерно равных объемах выборок его можно применять. При конечных объемах выборок традиционный метод носит неустранимо приближенный характер. Вместо критерия Стьюдента целесообразно для проверки гипотезы H_0 равенства средних в двух выборках использовать критерий Крамера-Уэлча, основанный на статистике

$$T_{\text{эмп}} = \frac{|\overline{x} - \overline{y}|}{\sqrt{\frac{D_x}{n_x} + \frac{D_y}{n_y}}}.$$

Критерий Крамера-Уэлча имеет прозрачный смысл – разность выборочных средних арифметических выборок *X* и *Y* делится на естественную оценку среднего квадратического отклонения этой разности. Естественность указанной оценки состоит в том, что неизвестные (в исследовании) дисперсии заменены их выборочными оценками. Из многомерной центральной предельной теоремы (приложение П3) вытекает то, что при росте объемов выборок распределение статистики T Крамера-Уэлча сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1.

Итак, при справедливости H_0 и больших объемах выборок распределение статистики T приближается к стандартному нормальному распределению $\Phi(x)$, из которого и можно выбирать критические значения.

Из асимптотической нормальности статистики T следует, что правило принятия решения для критерия Крамера-Уэлча выглядит следующим образом:

- если $T_{\text{эмп}} \le T_{\text{кр}}$ (т. е. $T_{\text{эмп}} \le \Phi(1-\alpha/2)$), то гипотеза однородности (равенства) математических ожиданий принимается на уровне значимости α;
- если $T_{\text{эмп}} > T_{\text{кр}}$ (т. е. $T_{\text{эмп}} > \Phi(1-\alpha/2)$), то гипотеза однородности (равенства) математических ожиданий отклоняется на уровне значимости α.

В прикладной статистике наиболее часто применяется уровень значимости α=0.05. Тогда значение модуля статистики T Крамера-Уэлча надо сравнивать с критическим $T_{\text{кр}}$ значением

$$T_{\text{кр}} = \Phi(1-\alpha/2) = 1.96\,.$$

Из сказанного выше следует, что применение критерия Крамера-Уэлча не менее обосновано, чем применение критерия Стьюдента. Дополнительное преимущество – не требуется равенства дисперсий $D_x = D_y$. Распределение статистики T не является распределением Стьюдента, однако и распределение статистики t не является таковым в реальных ситуациях.

Таким образом, критерий Крамера-Уэлча является достаточно эффективным "заменителем" такого известного в различных предметных приложениях t-критерия (Стьюдента).

Пример 5.1 Выполнить сравнение средних значений двух независимых выборок* на уровне значимости α=0.05. Используется критерий Крамера-Уэлча.

выборка *X*				
15	19	12	11	8
13	7	15	19	5
11	8	16	18	7
18	11	13	9	13
10	12	14	6	8
8	15	14	15	15
20	16	10	12	

выборка *Y*				
15	18	17	14	18
18	14	19	12	12
12	13	16	20	15
20	18	12	16	9
16	13	15	11	14
11	13	19	13	
13	15	13	7	

Эмпирическое значение данного критерия (ячейка P15) рассчитывается на основании информации об объемах n_x и n_y выборок (ячейки P6 и P7), выборочных средних $\overline{x}$ и $\overline{y}$ (ячейки P9 и P10) и выборочных дисперсиях D_x и D_y (ячейки P12 и P13) сравниваемых выборок *X* и *Y*.

Поскольку при росте объемов выборок распределение статистики T Крамера-Уэлча сходится к стандартному нормальному распределению, то критическое значение $T_{\text{кр}}$ можно определить через функцию НОРМ.СТ.ОБР(P), которая возвращает обратное значение стандартного нормального распределения (со средним, равным нулю, и стандартным отклонением, равное единице).

*Для иллюстрации метода взяты выборки сравнительно небольшого объема

Параметр функции P есть вероятность, соответствующая нормальному распределению, которую через уровень значимости α можно определить значением $P = 1 - \alpha/2$.

Рис. 5.1. Скриншот сравнения средних значений выборок для критерия Крамера-Уэлча

5.2. Тест Краскела-Уоллиса

Дисперсионный анализ по Краскелу-Уоллису (Kruskal-Wallis test) относится к группе непараметрических методов статистики – при выполнении соответствующих расчетов принадлежности выборок тому или иному вероятностному распределению (например, нормальному) не требуется. Вместо этого используются ранги исходных значений и их суммы в сравниваемых группах. Критерий используется для сравнения трех или более выборок и проверяет нулевые гипотезы, согласно которым различные выборки были взяты из одного и того же распределения или из распределений с одинаковыми медианами. Статистика критерия Краскела-Уоллиса основана на вычислении значения

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1), \quad n = \sum n_i \ ,$$

где n_i – число наблюдений и R_i – сумма рангов наблюдений в i-той группе; k – количество серий экспериментов.

> Ранг – номер наблюдения, присвоенный ему при процедуре ранжирования.
> Ранжирование – процедура присвоения рангов элементам выборки.

Ранг представляет собой порядковый номер конкретного наблюдения в ряду упорядоченных по возрастанию (или убыванию) наблюдений. Чем больше значение H-критерия, тем больше оснований отклонить нулевую гипотезу об отсутствии разницы между сравниваемыми группами. Если рассчитанное по выборочным данным значение H превышает определенное критическое значение $H_{\text{крит}}$, то нулевая гипотеза отклоняется.

Критическое значение определяется с учетом принятого уровня значимости и числа степеней свободы; в частности, при k>5 H-критерий можно определить как критическое значение критерия хи-квадрат для числа степеней свободы k−1. При меньшем числе сравниваемых групп вносятся определенные поправки (например, Iman-Davenport modification – поправка Имана-Давенпорта).

Необходимо отметить, что при выполнении обычного дисперсионного анализа по ранговым номерам исходных значений анализируемой переменной результат совпадает с тестом Краскела-Уоллиса. Видимо отсюда и следует использование в названии метода Краскела-Уоллиса слов "дисперсионный анализ". Кроме того, при наличии двух сравниваемых групп, тест Краскела-Уоллиса будет идентичен тесту Манна-Уитни.

Для анализируемых данных, удовлетворяющих условиям нормальности и однородности групповых дисперсий, статистическая мощность теста Краскела-Уоллиса составляет около 95% от результатов обычного параметрического дисперсионного анализа. Однако при нарушении этих условий мощность тест Краскела-Уоллиса может оказаться даже выше, чем у обычного дисперсионного анализа.

Для расчета несмещенных оценок H-критерия Краскела-Уоллиса значения анализируемой переменной теоретически должны иметь одинаковый разброс и форму распределения во всех сравниваемых группах. Однако ряд авторов отмечает, что на практике нарушение этих условий мало сказывается на качестве получаемых при помощи критерия выводов.

В процессе исследования наблюдения ранжируют (им присваиваются ранги), упорядочивая их по величине и назначая им номера (называемые рангами), соответствующие их месту в упорядочении. Обычно наблюдения ранжируются от меньшего к большему.

Средний ранг. Пусть имеется выборка из n упорядоченных по возрастанию наблюдений $x_1 \le x_2 \le \cdots \le x_n$. Предполагается, что наблюдение x_i имеет ту же величину, что и (совпадающие с ним) некоторые из остальных x наблюдений.

Средний ранг x_i в ранжировании наблюдений $x_1, x_2, \ldots, x_n$ есть среднее арифметическое из рангов, которые были бы назначены x_i и остальным значениям x, таким же, что и x_i, если бы равные наблюдения оказались различными.

Пример. Ранжируется выборка из пяти наблюдений {20, 22, 25, 25, 25}. Значение "25" встречается в ней трижды. Если бы равные наблюдения считались различными, то набор рангов для этой выборки был бы (1, 2, 3, 4, 5). Поскольку все значения "25" равноправны, присваиваем им усреднённый ранг (3+4+5)/3=4 и получаем набор рангов {1, 2, 4, 4, 4}.

Нулевая гипотеза заключается в том, что любая разница между группами (сериями) обусловлена случайностью. Если нулевая гипотеза верна, это означает, что математические ожидания всех k групп равны. Принимается нулевая гипотеза в случае, когда

$$H < H_{\text{крит}} = \chi^2(\alpha, k-1)\,.$$

Пример 5.2 Три группы испытуемых принимали некий препарат, последствия воздействия которого оценивалось неким индексом И; данные представлены в таблице.

группа 1	1.22	1.24	1.31	1.31	1.45	1.52	1.84	2.52	
группа 2	1.47	1.52	1.55	1.70	1.93	2.00	3.00		
группа 3	1.56	1.58	1.81	1.89	2.00	2.00	2.55	2.58	4.00

Требуется определить наличие существенной разницы между оказанным на группы воздействием по уровню значимости 0.05.

На рис. 5.2 дан скриншот решения поставленной задачи.

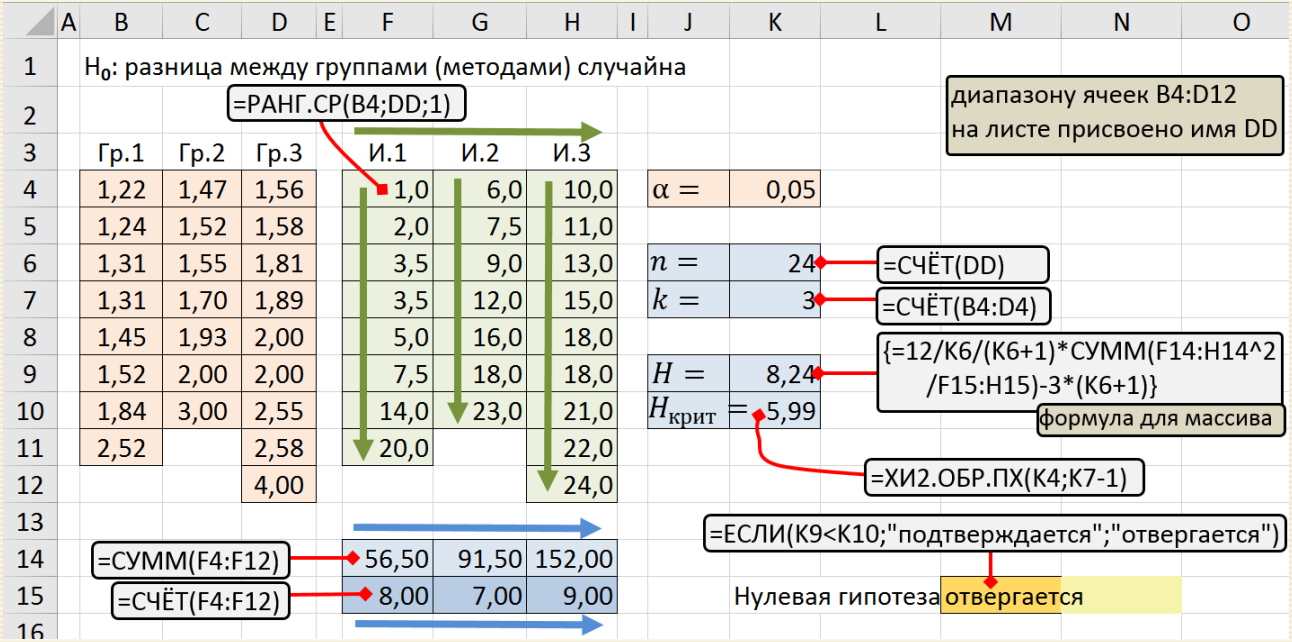


Рис. 5.2. Скриншот дисперсионного анализа по Краскелу-Уоллису

Из результатов видно, что $H = 8.24 > H_{\text{крит}} = 5.99$ и поэтому нулевая гипотеза отклоняется и можно заключить, что имеется существенная разница между по воздействию препарата на различные группы испытуемых.

5.3. Критерий Вилкоксона-Манна-Уитни

Критерий Вилкоксона-Манна-Уитни является непараметрическим и используется для оценки различий между двумя выборками по признаку, измеренному в количественной или порядковой шкале. Данный критерий оперирует не с абсолютными значениями элементов двух выборок, а с результатами их парных сравнений. Например, существенно, что учащийся Долгин решил больше задач, чем учащийся Перевозкин, а на сколько больше – не важно.

Имеются две выборки $\{x_i^{\text{КГ}}\}, i = 1,2,\dots,N_{\text{КГ}}$ и $\{x_i^{\text{ЭГ}}\}, i = 1,2,\dots,N_{\text{ЭГ}}$ до и после эксперимента. Для каждого элемента первой выборки $\{x_i^{\text{КГ}}\}$, определяется число a_i элементов второй выборки $\{x_i^{\text{ЭГ}}\}$, которые превосходят его $(x_i^{\text{КГ}})$ по своему значению (то есть число таких $x_j^{\text{ЭГ}}$, что $x_j^{\text{ЭГ}} > x_i^{\text{КГ}}$).

Сумма $U = a_1 + a_2 + \dots + a_{N_{\text{КГ}}} = \sum a_i$ этих чисел по всем $N_{\text{КГ}}$ членам первой выборки называется эмпирическим значением критерия Манна-Уитни.

С использованием значения U определяется эмпирическое значение критерия Вилкоксона:

$$W_{\text{эмп}} = \frac{\left|\frac{N_{\text{КГ}}N_{\text{ЭГ}}}{2} - U\right|}{\sqrt{\frac{N_{\text{КГ}}N_{\text{ЭГ}}}{12}(N_{\text{КГ}} + N_{\text{ЭГ}} + 1)}}.$$

Распределение нормированной и центрированной статистики Вилкоксона W при росте объемов выборок приближается к стандартному нормальному распределению (с нулевым математическим ожиданием и единичной дисперсией).

Из асимптотической нормальности статистики W следует, что правило принятия решения для критерия Вилкоксона следующее:

- если $|W| \le \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$, то гипотеза однородности (равенства) функций распределений принимается на уровне значимости α,
- если же $|W| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$, то гипотеза однородности функций распределений отклоняется на уровне значимости α.

Здесь $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ – квантиль порядка $\left(1-\frac{\alpha}{2}\right)$ стандартного нормального распределения (с математическим ожиданием 0 и дисперсией 1). В прикладной статистике наиболее часто применяется уровень значимости α =0.05. Тогда значение модуля статистики $|W|$ Вилкоксона сравнивается с критическим значением $W_{\text{кр}} = \Phi^{-1}(1-\alpha/2) = 1.96$.

В MS Excel $W_{\text{кр}}$ вычисляется посредством функции НОРМСТОБР(1- α/2).

Особенности использования критерия Вилкоксона-Манна-Уитни следующие:

1. Каждая выборка должна содержать не менее трех элементов; если же в одной из выборок всего два элемента, то во второй их должно быть не менее пяти.
2. Не имеет значения какую выборку считать первой, а какую второй.
3. Критерий Вилкоксона-Манна-Уитни плохо применим в условиях, когда число отличающихся друг от друга значений в выборках мало.

Пример 5.3 Для экспериментальной и контрольной группам были получены данные по уровню знаний в шкале отношений. Измерение уровня знаний заключалось по результатам проведения теста из 20 задач. Считалось, что характеристикой учащегося (признаком) является число правильно решенных им задач. Результаты измерений уровня знаний в контрольной и экспериментальной группах до и после эксперимента приведены в таблице (исходные данные в шкале отношений).

Другими словами, имеются выборки $\{x_i^{\text{КГ}}\}, i = 1,2, \dots, N_{\text{КГ}}$ и $\{x_i^{\text{ЭГ}}\}, i = 1,2, \dots, N_{\text{ЭГ}}$ до и после эксперимента.

Таблица исходных данных в шкале отношений

Контрольная группа до эксперимента (КГ до)

КГ до	15	13	11	18	10	8	20	7	8	12	15	16	13	14	14
	19	7	8	11	12	15	16	13	5	11	19	18	9	6	15

Экспериментальная группа до эксперимента (ЭГ до)

ЭГ до	12	11	15	17	18	6	8	10	16	12	15	14	19	13	19
	12	11	16	12	8	13	7	15	8	9					

Контрольная группа после эксперимента (КГ после)

КГ после	16	12	14	17	11	9	15	8	6	13	17	19	15	11	9
	19	8	6	9	12	11	17	10	8	8	20	19	6	14	10

Экспериментальная группа после эксперимента (ЭГ после)

ЭГ после	15	18	12	20	16	11	13	7	14	17	19	16	12	15	19
	18	14	13	18	13	13	15	18	9	14					

Требуется определить состояния экспериментальной и контрольной групп (совпадают или различаются средние значения) до и после эксперимента, сделать вывод об эффекте изменений состояния групп вследствие применения экспериментальной методики обучения.

Алгоритм определения соответствия характеристик состояния групп следующий.

A1. Заносятся исходные данные (A6:B20, D6:E20, G6:H20, J6:K20) – результаты тестирования контрольной и экспериментальной групп до и после начала исследования, а также величина уровня значимости α. Указанным данным присваиваются имена в соответствии с пояснениями, приведенными на нижней части рис. 5.3.

A2. В диапазоне ячеек M6:N20 вводятся (используя автозаполнение (растяжку)) формула

=СЧЁТЕСЛИ(КГ_до;">"&ЭГ_до) подсчета количества элементов контрольной группы, превосходящих по величине конкретный элемент экспериментальной группы для данных до начала эксперимента (область M6:N20, озаглавленная "до экспер-та"). Применяются правила работы с массивами: диапазон M6:N20 выделяется, нажимается клавиша F2 и комбинация (т.е. одновременно три) клавиши Ctrl + Shift + Enter.

A3. Аналогично п. A2 формируются частоты для диапазонов ячеек P6:Q20 подсчитываются соответствующие частоты контрольная и экспериментальной группа после эксперимента.

A4. Подсчитываются объемы выборок N (ячейки T10 и T11), значения критерия U Манна-Уитни (ячейки T13 и T14) и эмпирические значения критерия $W_{\text{эмп}}$ Вилкоксона-Манна-Уитни (ячейки U18 и U19) для данных до и после эксперимента.

A5. Данный критерий основан на нормированной и центрированной статистики Манна-Уитни и асимптотически стремится к стандартному нормальному распределению. Так что критическое значение $W_{\text{кр}}$ (ячейка T16) можно определить через параметры нормального стандартного распределения функцией =НОРМСТОБР(1-T4/2).

Henry Berthold Mann

На заключительном этапе выполняется анализ данных; вывод следующий.

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают ($W_{\text{эмп}} \cong 0.52 < W_{\text{кр}} \cong 1.96$), а конечные (после окончания эксперимента) – различаются ($W_{\text{эмп}} \cong 2.52 > W_{\text{кр}} \cong 1.96$).

Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментальной методики обучения.

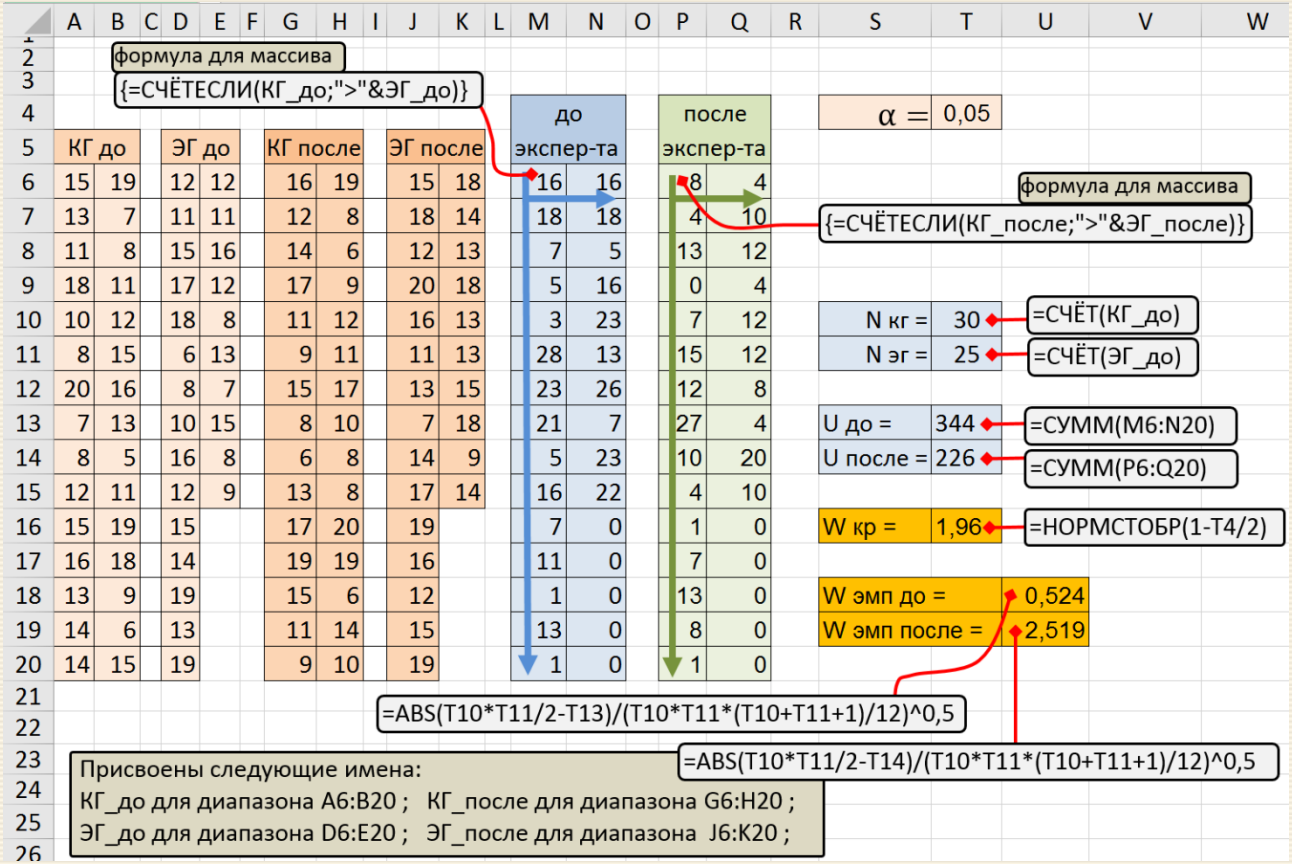


Рис. 5.3. Скриншот листа Excel для схемы вычислений критерия Вилкоксона

5.4. Критерий Вилкоксона связанных выборок

Критерий применяется для сопоставления показателей, измеренных в двух разных условиях на одной и той же выборке испытуемых. Он позволяет установить не только направленность изменений, но и их выраженность. С его помощью определяется, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Frank Wilcoxon

Критерий применим в тех случаях, когда признаки измерены по шкале порядка и сдвиги между вторым и первым замерами также могут упорядочиваться. В принципе, можно применять *T*-критерий Вилкоксона и в тех случаях, когда сдвиги принимают только три, как в критерии знаков, значения: -1, 0 и +1.

Суть метода состоит в том, что сопоставляются выраженности сдвигов в том и ином направлениях по абсолютной величине. Для этого сначала ранжируются абсолютные величины сдвигов и суммируются ранги. Если сдвиги в положительную и в отрицательную сторону происходят случайно, то суммы рангов абсолютных значений их будут примерно равны. Если же интенсивность (типичного) сдвига в одном из направлений перевешивает, то сумма рангов абсолютных значений сдвигов в противоположную (нетипичную) сторону будет значительно ниже, чем это могло бы быть при случайных изменениях.

Алгоритм анализа строится на предположения о том, что типичным сдвигом является более часто встречающееся направление, а нетипичный сдвиг принадлежит редко встречающемуся направлению.

Нулевая гипотеза *T*-критерия Вилкоксона формулируется следующим образом: интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении. Численно *T*-критерия равен сумме рангов нетипичного сдвига, а достоверные различия имеют место в том случае, если $T \le T_{\text{крит}}$.

Ограничения в применения *T*-критерия Вилкоксона связаны с объемом выборки: минимальное количество элементов равно 5; максимальное количество – 50, что определяется верхней границей имеющихся таблиц.

Пример 5.4 На уровне значимости 0.1 сравнить две выборки ("до" и "после", диапазоны данных B3:B12 и C3:C12 на рис. 5.4)

В ходе вычислений рассчитываются парные разности выборок (E3:E12), модули разностей (F3:F12) и средние ранги этих модулей (G3:G12). Для удобства дальнейшего анализа определяется (ячейка K5) параметр

$$\delta = \begin{cases} 1, & \text{если положительных разностей меньше отрицательных;} \\ -1, & \text{если положительных разностей больше отрицательных.} \end{cases}$$

Финальная для рангов формула =ЕСЛИ (E3*K$5<0;""; РАНГ.СР(F3; F$3:F$12;1)) в ячейки диапазона H3:H12 заносит пустые значения для рангов типичных сдвигов и действительные величины для нетипичных. В ячейке H14 рассчитывается сумма последних, которая и является статистикой критерия.

Критическое значение $T_{\text{крит}}$ определяется из соответствующих аппроксимаций табличных данных для уровней значимости 0.05 и 0.1.

$$T_{\text{крит}} = \begin{cases} (0.2149\alpha - 1.5236)\alpha + 3.8668 & \text{для } \alpha = 0.05; \\ (0.2013\alpha - 2.2913)\alpha + 7.5402 & \text{для } \alpha = 0.1 \ . \end{cases}$$

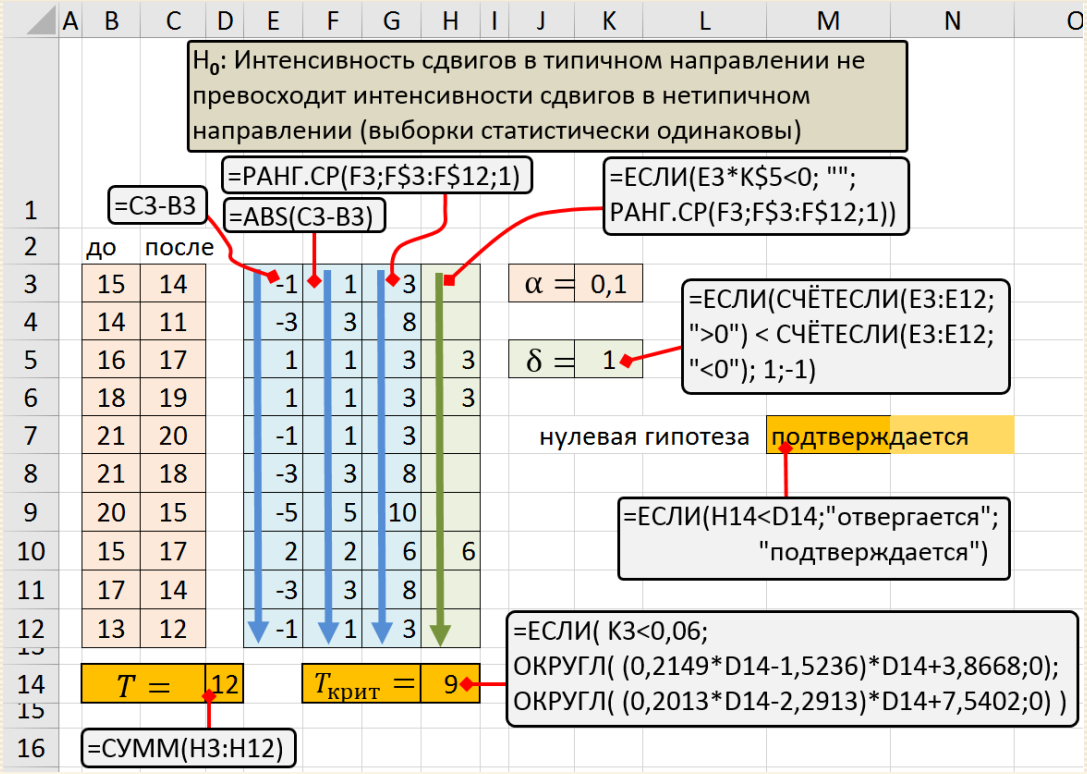


Рис. 5.4. Скриншот вычислений для парного критерия Вилкокса

Поскольку $T = 12 > T_{\text{крит}} = 9$, то можно заключить следующее: H_0 подтверждается; выборки статистически одинаковы.

6. Линейные и нелинейные зависимости. Регрессия. Коэффициент корреляции

6.1. Линейная регрессия

Линейная регрессия (linear regression) – используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной y от другой или нескольких других переменных (факторов, регрессоров, независимых переменных) x с функцией линейной зависимости.

В частном случае, когда фактор единственный (без учёта константы), говорят о парной линейной регрессии (a, b – константы): $y = a + bx$.

Когда количество факторов (без учёта константы) больше одного, то говорят о множественной регрессии.

Метод наименьших квадратов (МНК, Ordinary Least Squares, OLS) – математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. В данном случае используется для аппроксимации точечных значений линейной функцией. МНК является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным.

Пусть имеется n значений некоторой переменной y_i (это могут быть результаты наблюдений) и соответствующих переменных x_i. Задача заключается в том, чтобы взаимосвязь между y и x аппроксимировать некоторой функцией $y' = a + bx$ с некоторыми неизвестными параметрами a, b.

Фактически необходимо найти наилучшие значения параметров a и b, максимально приближающие значения y_i' к конкретным значениям y_i. Штрих в выражении y' добавлен для идентификации переменной, а отнюдь не обозначает производную. Соответственно, имеем $y_i' = a + bx_i$.

Сущность применения МНК (в данном случае) заключается в том, чтобы найти такие параметры a и b, при которых сумма квадратов отклонений (ошибок, для регрессионных моделей их часто называют остатками регрессии) будет минимальной:

$$\sum_{i=1}^{n}(y_i - y_i')^2 \to \min$$

Параметры a и b в MS Excel определяются функциями ОТРЕЗОК и НАКЛОН или ЛИНЕЙН.

Оценка параметров уравнения регрессии

Определяются выборочные дисперсии $D(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $D(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$,

среднеквадратические отклонения $S(x) = \sqrt{D(x)}$, $S(y) = \sqrt{D(y)}$.

средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических: $\mathrm{Err} = \frac{100}{n}\sum\frac{|y_i - y_i'|}{y_i}\,\%$.

Коэффициент эластичности показывает, на сколько процентов изменится величина результативной переменной *y*, если величина факторной переменной изменится на 1 %.

В общем случае коэффициент эластичности рассчитывается по формуле:

$$E = \frac{\partial y}{\partial x} : \frac{y}{x},$$

где $\partial y / \partial x$ – производная результативной переменной y по факторной переменной x.

Зачастую средний коэффициент эластичности определяется по формуле: $\overline{E} = b\,\overline{x}/\overline{y}$.

Анализ точности определения оценок коэффициентов регрессии

Несмещенной оценкой дисперсии возмущений является величина

$$D_y = \frac{1}{n-m-1}\sum_{i=1}^{n}(y_i - y_i')^2 .$$

То есть D_y – необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии). Соответственно, $S_y = \sqrt{D_y}$ – стандартная ошибка оценки (стандартная ошибка регрессии).

Стандартные отклонения параметров a и b определяются формулами

Стандартное отклонение S_a случайной величины a определется формулой

$$S_a = S_y \frac{\sqrt{\sum x_i^2}}{n \cdot S(x)}$$

Стандартное отклонение S_b случайной величины b

$$S_b = \frac{S_y}{\sqrt{n} \cdot S(x)}$$

Оценка значимости показателей уравнения регрессии

С помощью МНК получают лишь оценки параметров уравнения регрессии, которые характерны для конкретного статистического наблюдения (конкретного набора значений x и y).

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитывается t-критерий Стьюдента и доверительные интервалы каждого показателя. Выдвигается гипотеза H_0 о случайной природе показателей, то есть о незначимом их отличии от нуля.

Для проверки значимости параметров (значимо ли они отличаются от нуля для генеральной совокупности) используются статистические методы проверки гипотез.

Проводится проверка гипотезы H_0 о равенстве отдельных коэффициентов регрессии нулю на уровне значимости α. В случае, если основная гипотеза окажется неверной, возможно принятие альтернативной. Для проверки гипотезы используется t-критерий Стьюдента. Найденное по данным наблюдений значение t-критерия t_a (t_b) сравнивается с критическим значением распределения Стьюдента $t_{\text{крит}}$. Критическое значение определяется в зависимости от уровня значимости α и числа степеней свободы df, которое в случае линейной парной регрессии равно $df = n - 2$, n – число наблюдений. Если фактическое значение t_a (t_b) больше $t_{\text{крит}}$ (по модулю), то основную гипотезу отвергают и считают, что с вероятностью $(1 - \alpha)$ параметр или статистическая характеристика в генеральной совокупности значимо отличается от нуля.

Если фактическое значение t_a (t_b) $< t_{\text{крит}}$ (по модулю), то нет оснований отвергать основную гипотезу, т.е. параметр или статистическая характеристика в генеральной совокупности незначимо отличается от нуля при уровне значимости α.

$$t_a = \frac{a}{S_a}, \quad t_b = \frac{b}{S_b},$$

Расчет критического значение критерия Стьюдента $t_{\text{крит}}$ в MS Excel реализуется функцией СТЬЮДЕНТ.ОБР.2Х (α; df).

Полученные коэффициенты регрессии a и b статистически значимы и их можно использовать в уравнении линейной регрессии для анализов и прогнозов, если $t_a > t_{\text{крит}}$ и $t_b > t_{\text{крит}}$.

Доверительный интервал для коэффициентов уравнения регрессии

Доверительные интервалы коэффициентов регрессии, которые с надёжностью $(1-\alpha)$ определяются следующими соотношениями: $\left(a - S_a t_{\text{крит}};\ a + S_a t_{\text{крит}}\right)$, $\left(b - S_b t_{\text{крит}};\ b + S_b t_{\text{крит}}\right)$.

С вероятностью $(1-\alpha)$ можно утверждать, что значение данных параметров будут лежать в найденных интервалах.

Надежность уравнения регрессии

Целью анализа является получение некоторой оценки, с помощью которой можно было бы утверждать, что при некотором уровне α полученное уравнение регрессии – статистически надёжно и его можно привлекать для анализа и прогнозирования. Для этого используется коэффициент детерминации R^2.

$$R^2 = 1 - \frac{\sum(y_i - y_i')^2}{\sum(y_i - \overline{y})^2}.$$

Проверка значимости модели регрессии проводится с использованием *F*-критерия Фишера, расчетное значение которого находится как отношение дисперсии исходного ряда наблюдений изучаемого показателя и несмещенной оценки дисперсии остаточной последовательности для данной модели.

Если расчетное значение с $df_1 = (m)$ и $df_2 = (n - m - 1)$ степенями свободы больше табличного при заданном уровне значимости (m – число факторов в модели), то модель считается значимой.

Оценка статистической значимости парной линейной регрессии производится по следующему алгоритму:

1. Выдвигается нулевая гипотеза о том, что уравнение в целом статистически незначимо:
   $H_0: R^2 = 0$ на уровне значимости α.
2. Далее определяют фактическое значение *F*-критерия: $F = \frac{R^2}{1 - R^2} \frac{n - m - 1}{m}$.
3. Критическое значение $F_{\text{крит}}$ для заданного уровня значимости α, принимая во внимание, что число степеней свободы для общей суммы квадратов (большей дисперсии) равно 1 и число степеней свободы остаточной суммы квадратов (меньшей дисперсии) при линейной регрессии равно n-2 определяется через функцию MS Excel FРАСПОБР $(\alpha; 1; n - 2)$.
   $F_{\text{крит}}$ – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α. Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна.
4. Если фактическое значение F-критерия меньше табличного, то считается, что нет основания отклонять нулевую гипотезу. В противном случае нулевая гипотеза отклоняется и с вероятностью $(1 - \alpha)$ принимается альтернативная гипотеза о статистической значимости уравнения в целом.

Если фактическое значение $F > F_{\text{крит}}$, то коэффициент детерминации статистически значим (найденная оценка уравнения регрессии статистически надёжна).

Показатели качества уравнения регрессии

В качестве таковых обычно приводят следующие показатели:

- коэффициент детерминации;
- средний коэффициент эластичности;
- средняя ошибка аппроксимации.

Пример расчета показателей уравнения регрессии

Пример 6.1 Алгоритм построения линейной регрессии предполагает сортировку исходных данных по переменной x; (диапазон B5:C22 – исходные данные, диапазон E5:F22 – отсортированные (рис. 6.1). Для удобства вычислений вводятся имена массивов переменных (диапазоны и соответствующие имена приведены на рис. 6.1).

В ячейку K4 заносится величина принятого уровня значимости. Далее через функции ОТРЕЗОК и НАКЛОН определяются параметры a (ячейка N5)и b (ячейка N6) уравнения регрессии, а с использованием автозаполнения в диапазоне H5:H22 формируется массив $\{y_i' = a + bx_i\}$.

Для оценки значимости коэффициентов регрессии a и b в диапазоне R6:R8 определяются параметры $t_{\text{крит}}$, t_a, t_b. В ячейках K16 и K14 рассчитываются средние значения коэффициента эластичности $\overline{E}$ и ошибки аппроксимации Err. Для оценки статистической надёжности в ячейке K15 вычисляется коэффициент детерминации R^2.

В ячейках Q17 и Q18 рассчитываются эмпирическое и критическое значения критерия Фишера F и $F_{\text{крит}}$, сравнением которых определяется статистическая значимость уравнения в целом (рис. 6.1).

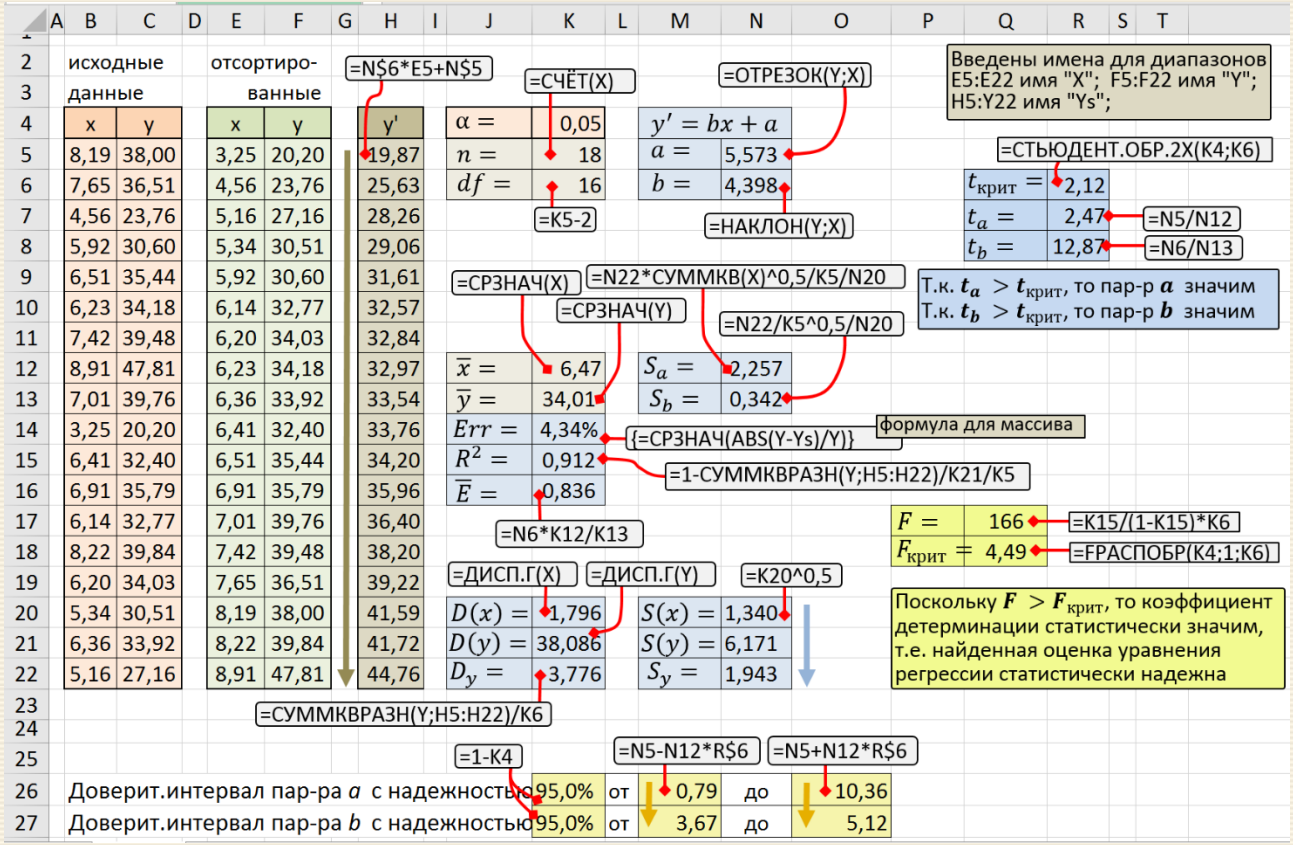


Рис. 6.1. Скриншот листа Excel расчета параметров линейной регрессии

Доверительные интервалы коэффициентов регрессии с надёжностью $(1-\alpha)\cdot 100\%$ рассчитываются в ячейках M26 (нижний предел), O26 (верхний предел) – для коэффициента a и M27, O27 – для коэффициента b.

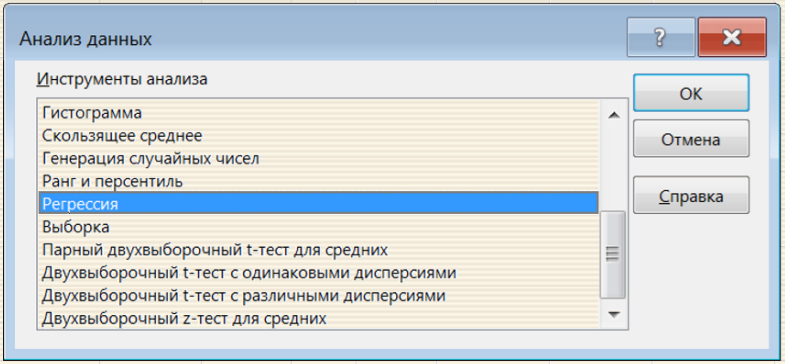
На рис. 6.2 дан график построенной линейной регрессии.

Рис. 6.2. Скриншот листа Excel для графика линейной регрессии

Искомые статистические характеристики могут быть определены с помощью надстройки Excel "Анализ данных" панели "Регрессия" (меню ДАННЫЕ панель АНАЛИЗ ДАННЫХ). На рис. 6.3. представлен скриншот использования надстройки, цветом выделены искомые данные.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,954924
R-квадрат	0,911879
Нормированный R-квадрат	0,906372
Стандартная ошибка	1,943103
Наблюдения	18



Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	625,1316	625,13	165,6	0
Остаток	16	60,4104	3,7756		
Итого	17	685,542			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	5,572603	2,256912	2,469	0,025	0,79	10,36	0,79	10,36
Переменная X 1	4,397742	0,341775	12,867	7,E-10	3,67	5,12	3,67	5,12

Рис. 6.3. Скриншот листа Excel надстройки "Анализ данных" – "Регрессия"

6.2. Нелинейные зависимости. Аппроксимация и идентификация параметров

Статистический анализ показывает, что вероятность
крупного выигрыша в лотерею всегда одинакова
и не зависит от того, купили вы лотерейный билет или нет.
Народная мудрость

Некоторые сложные функциональные зависимости можно свести к линейным и использовать аппарат регрессионного анализа, описанный в предыдущем разделе. При этом задача нахождения регрессионной кривой по обработке и анализу погрешностей экспериментальных данных сводится к решению следующих задач:

1. Линеаризация нелинейных зависимостей, которая осуществляется соответствующей заменой переменных. Примеры такой замены приведены в таблице.

   В некоторых случаях различные замены переменных могут приводить одну и ту же функцию к линейному виду несколькими способами. Например, эта ситуация возможна для зависимости $y = Ax^n$, соответствующие замены переменных приведены в первых двух строках таблицы.
2. Нахождение наилучших значений коэффициентов a и b в линейной зависимости $y = ax + b$ или коэффициента a в зависимости $y = ax$ согласно методу наименьших квадратов.
3. Нахождение случайных и приборных погрешностей этих коэффициентов.
4. Определение по найденным значениям коэффициентов a и b их "физического" содержания. Последняя задача решается стандартным приемом метода переноса погрешностей при косвенных измерениях.

Линеаризация через замену переменных

	Исходная функция	Замена переменных	Новая функция
1	$y = Ax^n$	$X = x^n,\ a = A$	$Y = aX$
2	$y = Ax^n$	$Y = \ln y,\quad X = \ln x,$ $a = n,\quad b = \ln A$	$Y = aX + b$
3	$y = Ax^{an}$	$Y = \ln y,\quad b = \ln A$	$Y = ax + b$
4	$y = ax^n + b$	$X = x^n$	$Y = aX + b$
5	$y = \frac{1}{ax^n + b}$	$Y = \frac{1}{y},\ X = x^n$	$Y = aX + b$
6	$y = \frac{1}{a + bx}$	$Y = \frac{1}{y},\ X = \frac{1}{x}$	$Y = aX + b$
7	$y = ax^n + bx^m$	$Y = yx^{-m},$ $X = x^{n-m}$	$Y = aX + b$
8	$y = a \cdot \sin x + b \cdot \cos x$	$Y = \frac{y}{\cos x}.\quad X = \mathrm{tg}x$	$Y = aX + b$

> Л и н е а р и з а ц и я – (лат. linearis – линейный), один из методов приближённого представления нелинейных систем, при котором исследование нелинейной системы заменяется анализом линейной системы, в некотором смысле эквивалентной исходной.

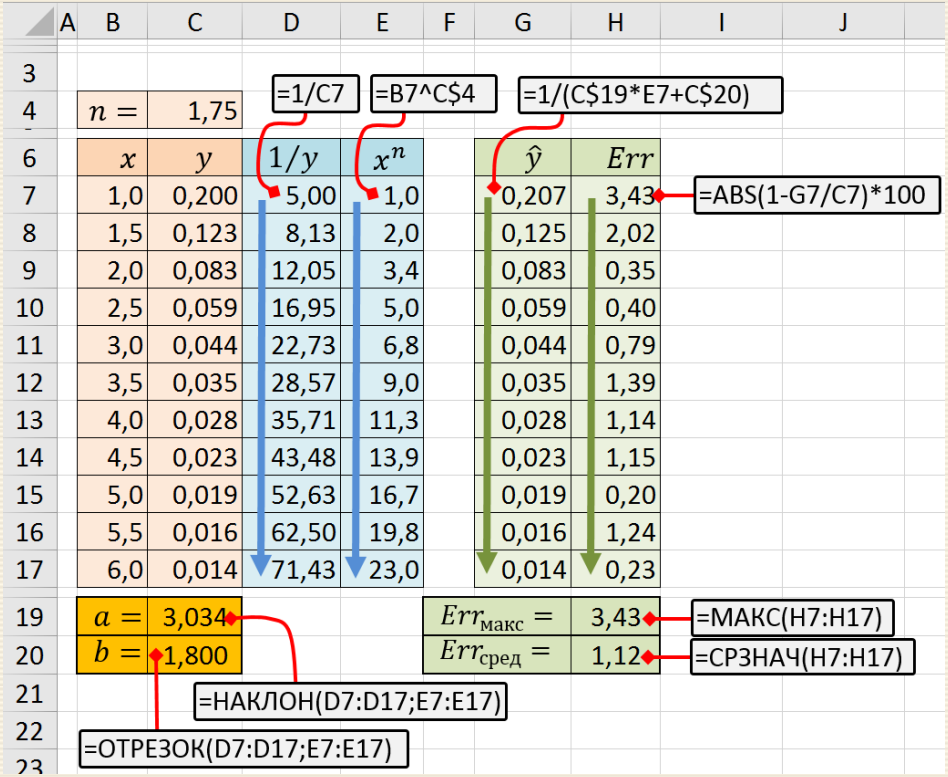
Простейшие методы линеаризации

- метод логарифмирования — применяется к степенным функциям;
- метод обратного преобразования — для дробных функций;
- комплексный метод — для дробных и степенных функций.

Пример 6.2 Определить константы a и b для аппроксимирующего уравнения $y = 1/(ax^n + b)$ по известной таблице данных B7:C17 для значения n=1.75 (рис. 6.4).

В соответствие с соотношениями линеаризации переходя к новым переменным $1/y$ (D7:D17) и x^n (E7:E17) константы a и b для уже "нового" линейного уравнения определяются (рис. 6.4) через Excel-функции НАКЛОН(…) и ОТРЕЗОК(…).

Для контроля и оценки качества аппроксимации в диапазоне G7:G17 рассчитываются "теоретические" значения $\hat{y} = 1/(ax_i^n + b)$ и их процентные расхождения Err с исходными данными.



	A	B	C	D	E	F	G	H	I	J
3										
4		$n =$	1,75							
6		x	y	$1/y$	x^n		$\hat{y}$	Err		
7		1,0	0,200	5,00	1,0		0,207	3,43		
8		1,5	0,123	8,13	2,0		0,125	2,02		
9		2,0	0,083	12,05	3,4		0,083	0,35		
10		2,5	0,059	16,95	5,0		0,059	0,40		
11		3,0	0,044	22,73	6,8		0,044	0,79		
12		3,5	0,035	28,57	9,0		0,035	1,39		
13		4,0	0,028	35,71	11,3		0,028	1,14		
14		4,5	0,023	43,48	13,9		0,023	1,15		
15		5,0	0,019	52,63	16,7		0,019	0,20		
16		5,5	0,016	62,50	19,8		0,016	1,24		
17		6,0	0,014	71,43	23,0		0,014	0,23		
19		$a =$	3,034				$Err_{макс} =$	3,43		
20		$b =$	1,800				$Err_{сред} =$	1,12		
21										
22										
23										

Рис. 6.4. Расчет констант a и b нелинейного уравнения

Пример 6.3 Определить константы a, b и n для аппроксимирующего уравнения $y = 1/(ax^n + b)$ по известной таблице данных B7:C17 рис. 6.5.

В случае, если требуется выполнить аппроксимацию данных (предыдущий пример 6.2), когда неизвестны все три (a, b и n) параметра уравнения $y = 1/(ax^n + b)$, то более простым и эффективным может быть подход подбора этих параметров с использованием надстройки MS Excel "Поиск решения" (Solver) по минимизации коэффициента вариации. Технология построения решения следующая.

1. В ячейки C2:C4 варьируемых переменных заносятся (рис. 6.5a) некоторые предварительные значения искомых параметров a, b и n, для которых в диапазоне G7:G17 рассчитываются "теоретические" значения $\hat{y} = 1/(ax_i^n + b)$ и их процентные расхождения Err (F7:F17) с исходными данными.

2. Меню ДАННЫЕ→ Поиск решения открывает окно (рис. 6.6), где указываются адреса целевой функции, варьируемых переменных, направления и используемый алгоритм оптимизации.

3. Исполнив Найти решение в случае успеха поиска в адресах варьируемых ячеек будут находится значения искомых параметров (рис. 6.5b).

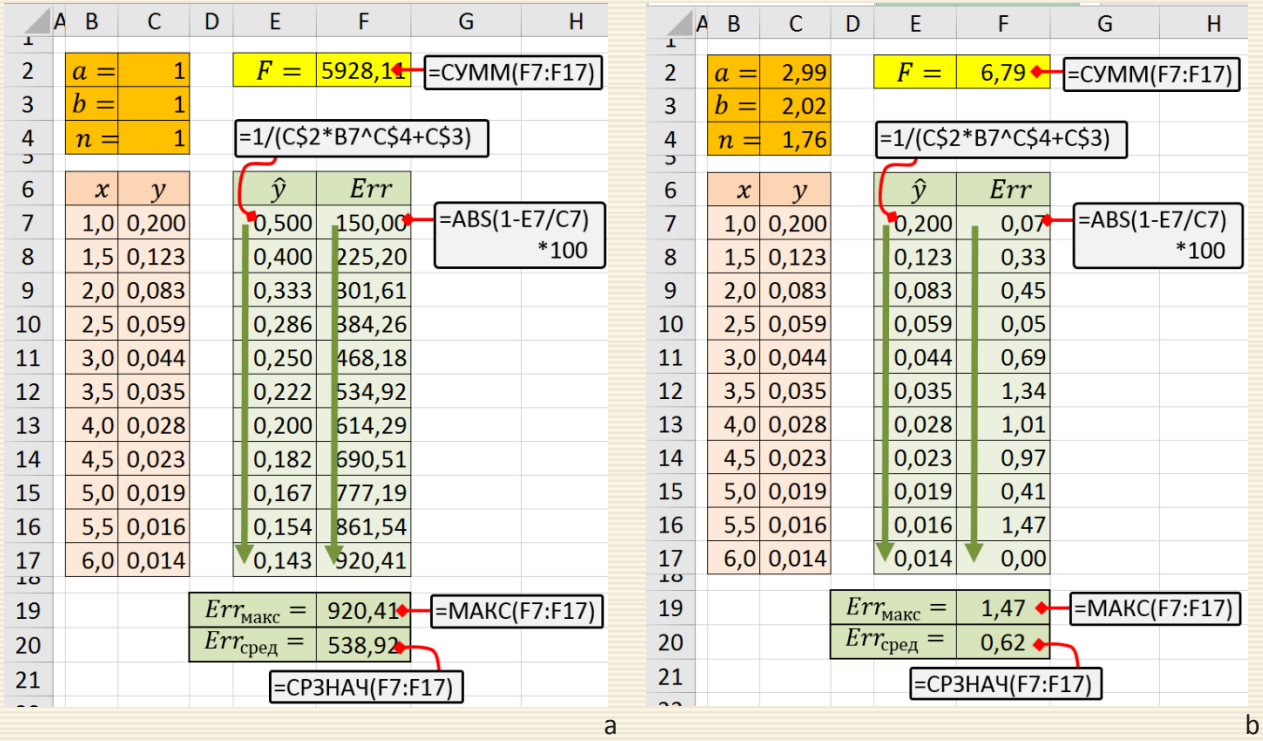


Рис. 6.5. Расчет констант a, b и n для нелинейного уравнения

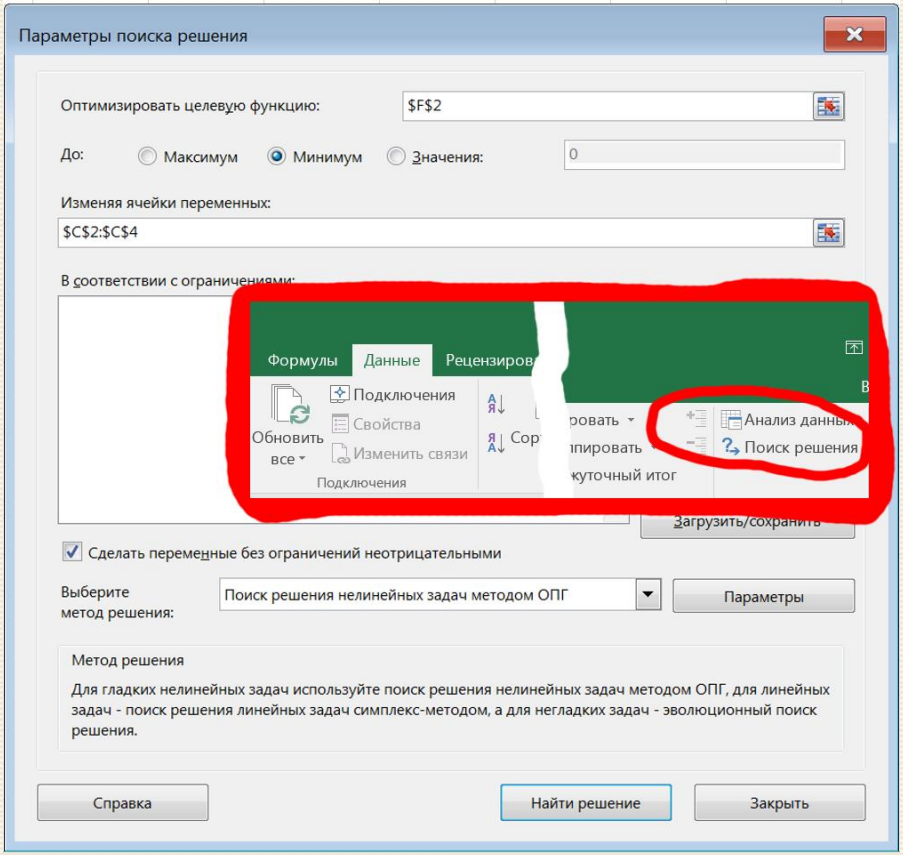


Рис. 6.6. Окно надстройки "Поиск решения" (Solver)

Идентификация параметров математической модели

По сути все естественные науки, использующие математический аппарат, опираются на математическое моделирование, когда объект исследования заменяется его математической моделью и изучается именно она. В простейших случаях математическая модель, представляющая собой уравнение (или систему уравнений), содержит собственно говоря зависимые и независимые переменные, определяющие их связь в рассматриваемом процессе (уравнении), а также параметры, которые характеризуют форму и количественное соотношения имеющейся связи.

Например, математическая модель, описывающая изменение во времени t (независимая переменная) концентрации C (зависимая переменная) в простой химической реакции имеет вид

$$\frac{dC}{dt} = -kC^p, \quad C(t=0) = C_0\,, \quad t \geq 0\,, \tag{6.1}$$

где параметры (коэффициенты, константы) модели есть k – константа реакции и p – порядок реакции.

> А п п р о к с и м а ц и я (от лат. proxima – ближайшая) – в смысле процесса это научный метод, состоящий в замене одного объекта другим (аппроксимации как результата процесса аппроксимации), в каком-то смысле близкому к исходному, но более простому.

Например, в результате химического эксперимента было получена таблица $\{C_i, t_i\}$, $i = 1,2,\dots,n$ изменения концентрации реагента в различные моменты времени. Уравнение (6.1) есть, в принципе, и математическая модель процесса, и аппроксимация исходной экспериментальной таблицы.

И д е н т и ф и к а ц и е й (от лат. identifico – отождествлять) параметров математической модели называется определение (или уточнение) входящих в модель коэффициентов при минимизации отклонений расчетных показателей от фактических.

Задача идентификации параметров математической модели формулируется как задача минимизации, в которой целевая функция – оценка степени совпадения исходных характеристик и выходных параметров, получаемых с помощью математической модели, а варьируемые (идентифицируемые) параметры – параметры анализируемой математической модели.

Пример 6.4 Определить константу и порядок простой химической реакции для экспериментально полученной зависимости концентрация бромистого нитрозила в реакции его разложения

t, мин	0	18	36	54	72	90	108	126	144
C, мМ	25.4	18.7	14.8	11.6	9.7	8.0	6.8	5.8	5.0

Теоретической основой задачи идентификации порядка реакции p по экспериментальным данным зависимости концентрации реагента C от времени t протекания этой реакции является уравнение (6.1), аналитическое решение которого имеет вид

$$C = \begin{cases} C_0 \exp(-kt), & \text{если } p = 1, \\ \left[C_0^{1-p} - (1-p)kt\right]^{1/(1-p)}, & \text{если } p \neq 1. \end{cases}$$

Явно выраженная из этого уравнения константа реакции k_i, вычисленная по двум значениям концентрации при известном значении порядка реакции p подчиняется уравнению

$$k_i = \frac{1}{t}\begin{cases}\ln(C_0/C_i), & \text{если } p = 1,\\ \left[C_0^{1-p} - C_i^{1-p}\right]^{1/(1-p)}, & \text{если } p \neq 1.\end{cases}$$

Решение задачи строится по алгоритму предыдущего примера, когда в качестве варьируемого параметра используется порядок реакции p (ячейка C8 рис. 6.7), а целевой функцией $\mathcal{F}$ (ячейка F9) является коэффициент вариации для рассчитанных значений $\{k_i\}$.

Выражения для целевой ячейки, отражающие "постоянство" константы k модели реакции, выбрано из-за простоты его реализации через статистические функции Excel:

СТАНДОТКЛОН.Г(<данные>)/СРЗНАЧ(<данные>).

	A	B	C	D	E	F	G	H	I	J	K	L
2		$t =$	0	18	36	54	72	90	108	126	144	
3		$C =$	0,0254	0,0187	0,0148	0,0116	0,0097	0,0080	0,0068	0,0058	0,0050	
4												
5			$k =$	0,139	0,131	0,136	0,132	0,135	0,134	0,136	0,137	
6												
7		переменная порядок реакции										
8		$p =$	1,55		$\overline{k} =$	0,135						
9		$1 - p =$	-0,55		$\mathcal{F} =$	0,018						
10												

=ЕСЛИ($C8=1;LN($C3/D3)/D2;($C3^$C9-D3^$C9)/$C9/D2)

=СРЗНАЧ(D5:K5)

=СТАНДОТКЛОН.Г(D5:K5)/F8

целевая ячейка

Рис. 6.7. Расчет кинетических параметров реакции разложения бромистого нитрозила

6.3. Коэффициенты корреляции Пирсона и Спирмена

Джон Лайтерут, он же архиепископ Ушер Ирландский,
подсчитал со всем усердием в Кембридже, в 1654 году,
что создатель сотворил человека из глины точно
в 9 часов утра 23 октября 4004г. до рождения Христа.
Справка

К о р р е л я ц и я (от лат. correlatio "соотношение, взаимосвязь") или корреляционная зависимость – статистическая взаимосвязь двух или более случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

К о р р е л я ц и о н н ы й а н а л и з – метод, позволяющий обнаружить зависимость между несколькими случайными величинами.

Для графического представления корреляционной связи можно использовать прямоугольную систему координат с осями, которые соответствуют обеим переменным. Каждая пара значений маркируется при помощи определённого символа. Такой график называется диаграммой рассеяния.

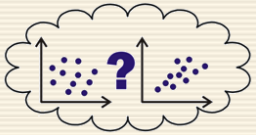
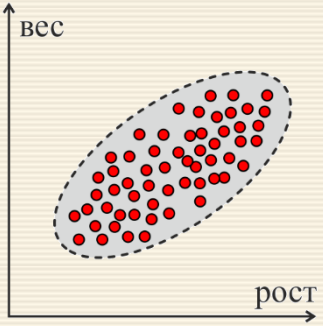


Диаграмма рассеяния (также точечная диаграмма, scatter plot) – математическая диаграмма, изображающая значения двух переменных в виде точек на декартовой плоскости.

На диаграмме рассеяния каждому наблюдению (или элементарной единице набора данных) соответствует точка, координаты которой (в декартовой системе координат) равны значениям двух каких-то параметров этого наблюдения. Если предполагается, что один из параметров зависит от другого, то обычно значения независимого параметра откладывается по горизонтальной оси, а значения зависимого – по вертикальной.

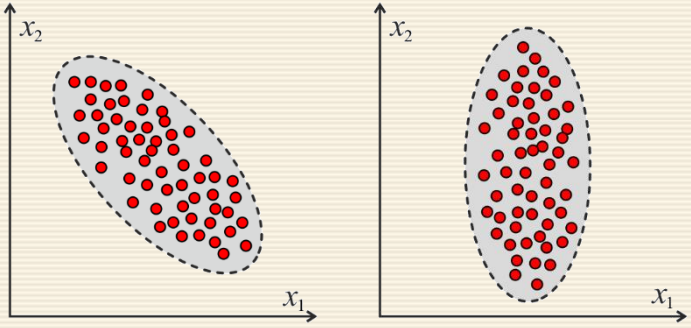
Допустим, проводится независимое измерение различных параметров у одного типа объектов. Из этих данных можно получить качественно новую информацию – о взаимосвязи этих параметров.

Например, измеряем рост и вес человека, каждое измерение представлено точкой в двумерном пространстве (рисунок справа).



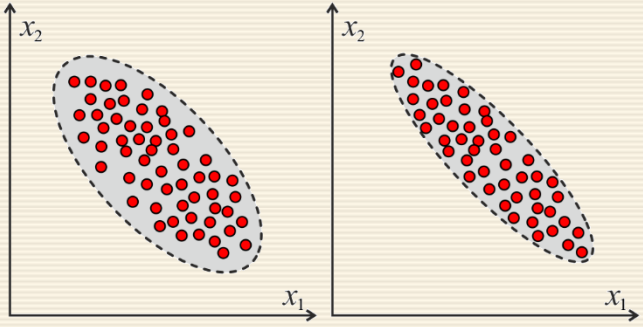
Несмотря на то, что величины носят случайный характер, в общем наблюдается некоторая зависимость – величины коррелируют. В данном случае это положительная корреляция (при увеличении одного параметра второй тоже увеличивается).

Возможны и другие случаи, представленные на диаграммах справа.



Отрицательная корреляция Отсутствие корреляции

Взаимосвязь между переменными необходимо охарактеризовать численно, чтобы, например, различать такие "похожие" случаи, приведенные на диаграммах справа.



Для этого вводится коэффициент корреляции**.** Из ряда используемых для этого соотношений ниже рассматриваются коэффициенты корреляции Пирсона и Спирмена.

Виды связи между признаками, характеризующими явление, можно определить как

- ф у н к ц и о н а л ь н а я – присущая физической природе, когда изменение величины одного признака строго вызывает изменение другого признака. Например, зависимость расстояния от времени и скорости.
- к о р р е л я ц и о н н а я – величине одного признака соответствует ряд варьирующих значений другого признака (например, зависимость частоты пульса от температуры тела).

Пусть $(x_1, \dots x_n), (y_1, \dots y_n)$ – набор значений двух факторов на выборке объёма n. Значимые характеристики корреляционной связи определяется значением коэффициента корреляции, обычно обозначаемого как r. Свойства коэффициента корреляции r

- r изменяется в интервале от –1 до +1.
- Знак r означает, увеличивается ли одна переменная по мере того, как увеличивается другая (положительный r), или уменьшается ли одна переменная по мере того, как увеличивается другая (отрицательный r).
- Величина r указывает, как близко расположены точки к прямой линии. В частности, если r= +1 или r= –1, то имеется абсолютная (функциональная) корреляция по всем точкам, лежащим на линии (практически это маловероятно); если $r \cong 0$, то линейной корреляции нет (хотя может быть нелинейное соотношение). Чем ближе r к крайним точкам (±1), тем больше степень линейной связи.
- Величина r обоснована только в диапазоне значений x и y в выборке. Нельзя заключить, что он будет иметь ту же величину при рассмотрении значений x и y, которые значительно больше, чем их значения в выборке.

- Коэффициент корреляции r безразмерен, т.е. не имеет единиц измерения.
- x и y могут взаимозаменяться, не влияя на величину r $(r_{xy} = r_{yx})$.
- Корреляция между x и y не обязательно означает соотношение причины и следствия.
- Величина r^2 представляет собой долю вариабельности y, которая обусловлена линейным соотношением с x.

При расчете коэффициента корреляции r могут возникать ошибки в случаях, если:

- соотношение между двумя переменными нелинейное, например, квадратичное;
- данные включают более одного наблюдения по каждому случаю;
- есть аномальные значения (выбросы);
- данные содержат ярко выраженные подгруппы наблюдений.

Сила (теснота) связи определяется шкалой Чеддока (scale Cheddok):

Количественная мера тесноты связи	Качественная характеристика силы связи
0.1 – 0.3	Слабая
0.3 – 0.5	Умеренная
0.5 – 0.7	Заметная
0.7 – 0.9	Высокая
0.9 – 0.99	Весьма высокая

Значимость линейного коэффициента корреляции для уровня значимости α определяется через критерий Стьюдента.

Критическое для оценки значимости коэффициента корреляции $r_{\text{крит}}$ вычисляется по формуле: $$r_{\text{крит}} = t(\alpha, df)\frac{\sqrt{1-r^2}}{\sqrt{df}},$$

где число степеней свободы $df = n - m - 1$, $m = 1$ – количество объясняющих переменных.

Если $r > r_{\text{крит}}$, то полученное значение коэффициента корреляции r признается значимым (нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается).

Соответственно, если $r < r_{\text{крит}}$, то допускается гипотеза о равенстве нулю коэффициента корреляции. Другими словами, в этом случае коэффициент корреляции статистически считается не значимым.

Расчет критического значение критерия Стьюдента $t_{\text{крит}}$ в MS Excel реализуется функцией СТЬЮДЕНТ.ОБР.2Х (α;*df*).

В парной линейной регрессии проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

Доверительный интервал для коэффициента корреляции – интервальная оценка для линейного коэффициента корреляции определяется диапазоном значений

$$\left(r - t_{\text{крит}}\frac{1-r^2}{\sqrt{n}}\ ;\quad r + t_{\text{крит}}\frac{1-r^2}{\sqrt{n}}\right).$$

Значимость отличия между коэффициентами корреляции r_1 и r_2 оценивается тестовой статистикой

$$\Delta t_{\text{набл}} = 0.5 \cdot \ln\left(\frac{(1+r_1)(1-r_2)}{(1-r_1)(1+r_2)}\right)\left[\sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}\right]^{-1},$$

которая сравнивается с критическим значением $t_{\text{крит}}$.

Методами корреляционного анализа решаются следующие задачи:

1) Взаимосвязь. Есть ли взаимосвязь между параметрами?
2) Прогнозирование. Если известно поведение одного параметра, то можно предсказать поведение другого параметра, коррелирующего с первым.
3) Классификация и идентификация объектов. Корреляционный анализ помогает подобрать набор независимых признаков для классификации.

Расчет коэффициента корреляции Пирсона

Выборочный коэффициент корреляции Пирсона (т.е. коэффициент корреляции, определяемый по выборке) равен:

$$r = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{j=1}^{n}(y_i-\overline{y})^2}}, \qquad \overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ \ \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i\ .$$

Коэффициент корреляции, подсчитанный таким образом, называется коэффициентом корреляции Пирсона; в MS Excel реализуется функцией КОРРЕЛ.

Бисериальный коэффициент корреляции Пирсона

Бисериальная корреляция (лат. bis series – два ряда, две серии) — метод корреляционного анализа отношения переменных, одна из которых измерена в дихотомической шкале наименований, а другая — в интервальной шкале отношении или порядковой шкале. Название метода связано с тем, что сравниваются две альтернативные серии объектов: интервальные значения X и дихотомические величины Y, имеющие значения 0 или 1. Для описания связи между перечисленными видами переменных используется точечный бисериальный коэффициент корреляции Пирсона.

Наиболее характерно применение коэффициентов бисериальной корреляции при анализе дискриминативности заданий теста, а также при определении валидности путем коррелирования значений тестовых оценок с независимыми характеристиками критерия, выраженными в дихотомической шкале.

Под дискриминативностью теста понимается способность теста дифференцировать испытуемых в диапазоне от "максимального" до "минимального" результата, набранного по данному теста.

Валидность (или обоснованность) процедуры измерения состоит в однозначности (устойчивости) получаемых результатов относительно измеряемых свойств объектов, т.е. относительно предмета измерения.

Для описания связи между рассматриваемыми видами переменных X и Y используется точечный бисериальный коэффициент корреляции Пирсона

$$r_{pb} = \frac{\overline{x}_1 - \overline{x}_0}{S_x} \sqrt{\frac{n_1 n_0}{n(n-1)}},$$

где $\overline{x}_1$ – среднее по X объектов со значением единицы по Y;
$\overline{x}_0$ – среднее по X объектов со значением нуль по Y;
S_x – стандартное выборочное отклонение всех значений по X;
n_1 – число объектов с единицей по Y; n_0 – число объектов с нулем по Y.

Очевидно, что размерность (объем, длина) n значений X и Y равна $n = n_1 + n_0$.

Значение r_{pb} изменяется от -1 до +1. Можно отметить, что когда переменные с $Y = 1$ имеют среднее по X, равное среднему переменных с нулем по Y, то $r_{pb} = 0$.

Для тестов значение r_{pb} показывает диагностическую значимость проверяемого пункта задания и степень корреляции с общими результатами. Коэффициент корреляции характеризует валидность отдельных заданий. Структура тестов должна быть такой, чтобы корреляция результатов по заданию и индивидуальными баллами была достаточно высокой; рекомендуемое значение $r_{pb} \geq 0.5$.

Пример 6.5 Оценить диагностическую значимость личного опросника – корреляции между типичным ответом на отдельный пункт (утверждение – отрицание, значения Y заданы в диапазоне ячеек B9:B26 на рис. 6.8) с общим результатом теста (значения X заданы в диапазоне ячеек C9:C26).

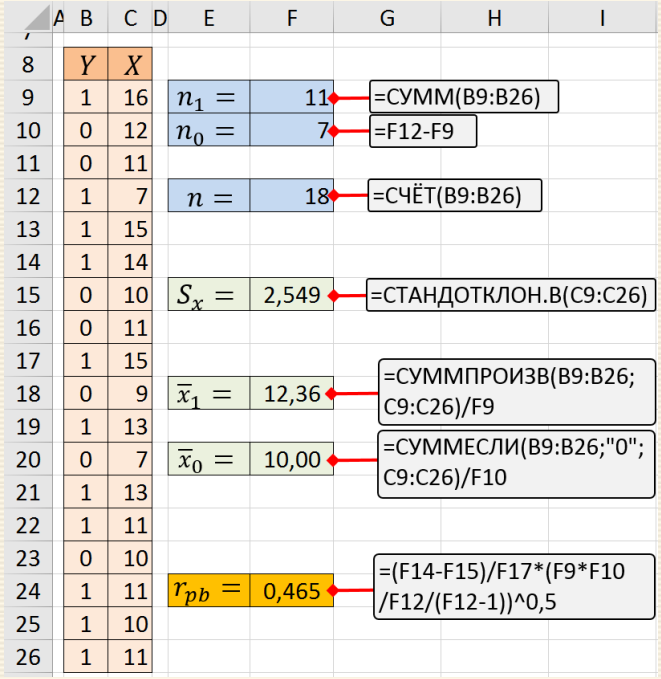


Рис. 6.8. Расчет точечного бисериального коэффициента корреляции

Исходя из структуры массива Y на рис. 6.8 даны Excel-формулы расчета параметров $n_1 = \sum y_i$, $\bar{x}_1 = \sum x_i y_i / n_1$; $\bar{x}_0$ определяется через функцию условного суммирования

СУММЕСЛИ(…) значений x_i, соответствующие значение y_i которых равно нулю.

Вычисленное значение r_{pb}= 0.465<0.5 показывает, что проверяемый пункт опросника имеет среднюю диагностическую значимость и слабую корреляцию с общим результатом теста.

Помимо того, что статистическая обработка данных тестирования позволяет оценивать испытуемых, необходимо оценивать и качество непосредственно самого теста; результаты которого чаще всего представляемого в дихотомической шкале (единица – верный ответ, нуль – неверный).

Пример 6.6 В таблице результатов тестирования (дана ниже) представлены данные 10 испытуемых (ФИО 1 – ФИО 10) по 8 заданиям в виде упорядоченной бинарной матрицы после предварительной обработки, когда удалены не позволяющие дифференцировать уровень знаний столбцы с заданиями, выполненные всеми участниками или невыполненными никем. Определить диагностические значимости заданий и уровень корреляции между заданиями.

Исходные данные заданы в виде двумерного массива $y_{i,j}$, $i = 1, \dots n,\ j = 1, \dots, m$, где n – число испытуемых, m – количество тестовых заданий, по которой формируется вектор x_j числа индивидуально набранных испытуемыми баллов и число правильно данных ответов n_{1j} на конкретное задание

$$x_i = \sum_{j=1}^{m} y_{i,j}\ ,\ \ j = 1, \dots, n, \quad n_{1j} = \sum_{i=1}^{n} y_{i,j}\ ,\ \ j = 1, \dots, m\ .$$

Результаты тестирования

	Задание 1	Задание 2	Задание 3	Задание 4	Задание 5	Задание 6	Задание 7	Задание 8
ФИО 1	1	1	1	1	1	1	1	0
ФИО 2	1	1	0	1	1	1	0	1
ФИО 3	1	1	1	1	1	1	0	0
ФИО 4	1	1	0	1	1	0	1	0
ФИО 5	1	1	1	0	1	0	0	0
ФИО 6	1	1	1	0	0	0	0	0
ФИО 7	0	1	1	0	0	0	0	0
ФИО 8	0	0	1	1	0	0	0	0
ФИО 9	0	0	0	0	0	1	0	0
ФИО 10	1	0	0	0	0	0	0	0

Рассчитываются $p_j = n_{1j}/n$ – доля верных ответов на j-тое задание; $q_j = 1 - p_j$ – доля неверных ответов на j-тое задание и доля "одновременно" верных ответов в заданиях с номерами l и k

$$p_{lk} = \left(\sum_{i=1}^{n} y_{i,l} \cdot y_{i,k} \right) / n \, .$$

На рис. 6.9*a* дан скриншот части листа MS Excel, где реализованы вычисления вышеприведенных параметров. На рис. 6.9*b* представлены результаты искомых корреляционных показателей. Кроме бисериального коэффициента r_{pb} для анализа наличия связи между заданиями теста для каждой пары заданий рассчитываются коэффициенты корреляции Пирсона (так называемые коэффициенты φ_{lk}), которые определяют связь между заданиями с номерами l и k

$$\varphi_{lk} = \frac{p_{lk} - p_l p_k}{\sqrt{p_l q_l p_k q_k}} \, .$$

Можно отметить, что показательным параметром тестового задания является вариация (дисперсия) тестовых долей p_i и q_i. Чем больше вариация, тем качественнее задание дифференцирует испытуемых.

При анализе коэффициентов корреляция заданий друг с другом необходимо учитывать то, что если корреляция между двумя заданиями близка к единице, то одно из них лишнее.

Отрицательная корреляция какого-либо задания с другими указывает на следующее. Если задание отрицательно коррелирует с несколькими другими – это значит, что "правильные" ответы не соответствуют таковым для других заданий. Скорее всего в этом задании имеются ошибки либо задание не соответствует проверяемой предметной области.

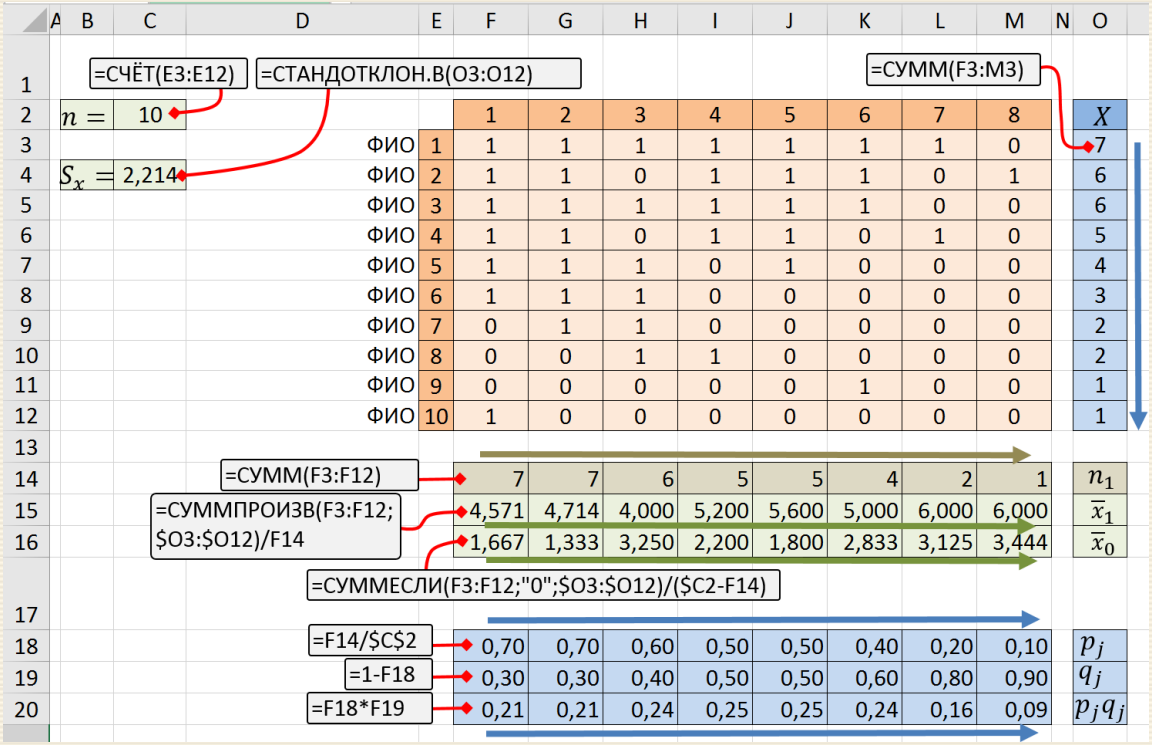


Рис. 6.9*а*. Вычисления "долевых" и средних параметров

Полученные результаты указывают на то, что отрицательная корреляция занятия 3 с заданиями 1, 6-8 и для него $r_{pb} = 0.175$ – его следует удалить. Подозрительным является и задание 8, поскольку для него $r_{pb} = 0.365$.

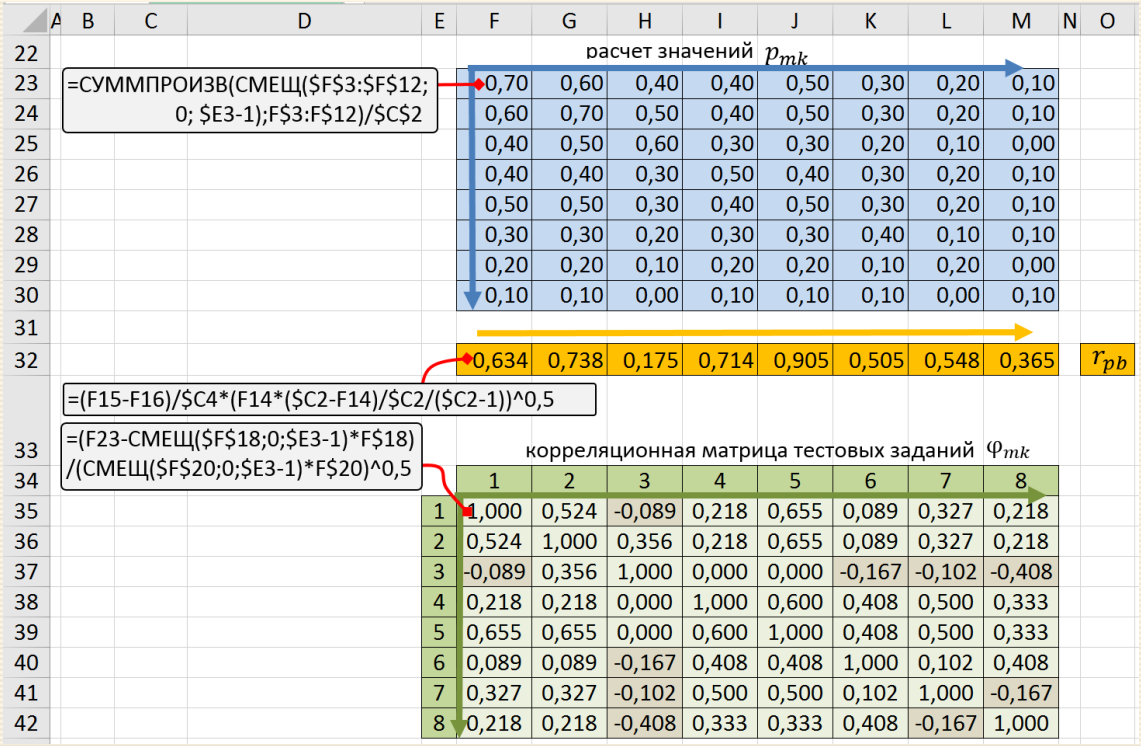


Рис. 6.9*b*. Расчет точечных бисериальных коэффициентов корреляции

Расчет коэффициента корреляции Спирмена

Помимо коэффициента корреляции r для определения тесноты связи используются и другие, менее точные показатели, например, коэффициент корреляции рангов Спирмэна, когда показатель рассчитывается на основе метода выстраивания параллельных рядов и ранжирования (присваивания порядковых номеров) значений x и y.

Karl (Charles) Pearson

Коэффициент ранговой корреляции Спирмена используется для выявления и оценки тесноты связи между двумя рядами сопоставляемых количественных показателей. В том случае, если ранги показателей, упорядоченных по степени возрастания или убывания, в большинстве случаев совпадают (большему значению одного показателя соответствует большее значение другого показателя – например, при сопоставлении роста человека и массы его тела), делается вывод о наличии прямой корреляционной связи.

Соответственно, если ранги показателей имеют противоположную направленность (большему значению одного показателя соответствует меньшее значение другого – например, при сопоставлении возраста и частоты сердечных сокращений), то говорят об обратной связи между показателями.

В связи с тем, что коэффициент является методом непараметрического анализа, проверка на нормальность распределения не требуется.

Charles Edward Spearman

Сопоставляемые показатели могут быть измерены как в непрерывной шкале (например, рост человека), так и в порядковой (например, баллы экспертной оценки от 1 до 100).

Эффективность и качество оценки методом Спирмена снижается, если разница между различными значениями какой-либо из измеряемых величин достаточно велика. Не рекомендуется использовать коэффициент Спирмена, если имеет место неравномерное распределение значений измеряемой величины.

Коэффициент Спирмена определяется по формуле:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)},$$

где d – разность рангов (порядковых номеров) признаков x и y;

n – количество пар значений x и y.

Пример 6.7 Знания 10 студентов проверены по двум тестам: A и B. Оценки по стобалльной системе оказались следующими (в первой строке по тесту A, во второй - по тесту B):

A	95	90	86	84	75	70	62	60	57	50
B	68	68	62	64	55	55	49	46	46	40

Найти коэффициент корреляции и проверить при α = 0.01 его значимость. Указать доверительный интервал для коэффициента корреляции.

Скриншот расчета коэффициента Спирмена приведен на рис. 6.10.

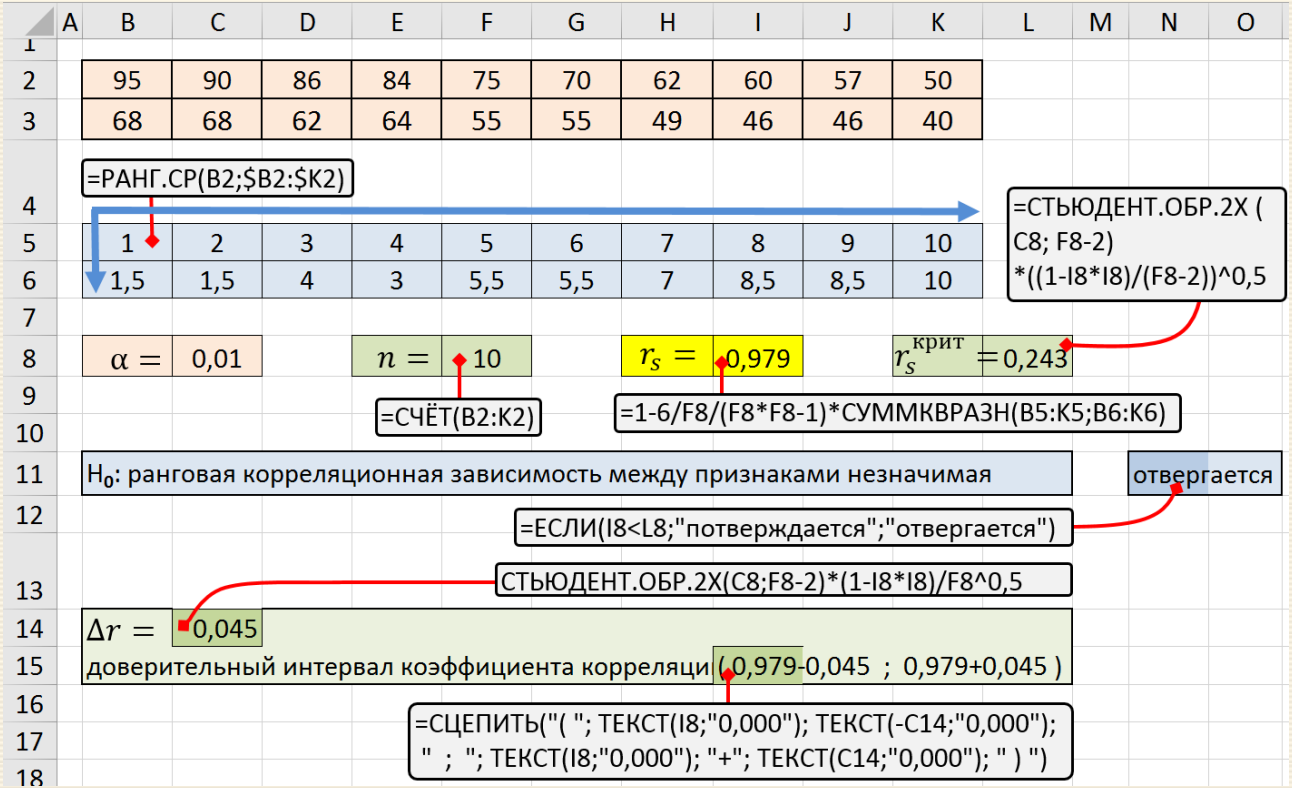


Рис. 6.10. Скриншот листа Excel расчета коэффициента корреляции Спирмена

6.4. Градуировка. Пределы обнаружения аналита

Суть задачи градуировки (калибровки) заключается в создании шкалы (градуировочной функции, таблицы), связывающей показания средства измерения с искомым значением физической величины.

Градуировка (от лат. gradus – шаг, ступень, степень) – процесс построения градуировочной функции опытным путем. Очевидно, что абсолютные методы измерений градуировки не требуют.

Неявно градуировка присутствует даже в таком методе, как титриметрия. Концентрация титранта устанавливается через стандартизацию, а затем уже эта величина используется для вычисления концентрации в неизвестном образце.

В большинстве своем градуировочная функция строится как линейная и представляется уравнением линейной регрессии

$$y = kx + y_0 ,$$

где y – уровень аналитического сигнала, x – измеряемая величина (например, концентрация какого-либо вещества – аналита).

Константы k и y_0 определяются в соответствии с алгоритмом, изложенным в подразделе 6.1.

На примере анализа концентрации C вещества-аналита ниже рассматриваются характеристики, определяющие параметры точности измерений, использующие градуировочную функцию

$$y = kC + y_0.$$

1. $C_{\text{обн}}$ (limit of blank, LoB) – предел обнаружения, основанный на результатах измерения *холостого опыта* (blank), не содержащего определяемого компонента. Представляют собой то наименьшее содержание аналита, при котором по данной методике можно обнаружить статистически значимое присутствие определяемого компонента в анализируемом объекте (рис. 6.11).

2. $C_{\text{мин}}$ (limit of detection, LoD) – наименьшее содержание вещества, которое может быть обнаружено по данной методике с заданной степенью достоверности и величиной стандартного отклонения S_0. C_{min} является самой низкой концентрацией аналита, которая может быть надёжно отличной от $C_{\text{обн}}$ и при которой обнаружение возможно.

3. $C_{\text{над}}$ (limit of quantitation, LoQ) – пределу "надёжного обнаружения" введенное для практического применения понятие (нижней границей определяемых содержаний). $C_{\text{над}}$ есть наименьшая концентрация, при которой аналит может быть не только надёжно обнаружен, но и где допущения "остаются за бортом". Понятно, что $C_{\text{над}} \geq C_{\text{мин}}$.

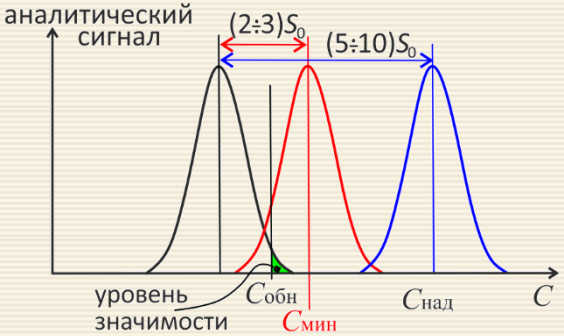


Рис. 6.11. Пределы обнаружения

Стандартное отклонение фонового сигнала S_0, (уровень шума*)* рассчитано из нескольких десятков параллельных измерений y_0), то критерий

$$\frac{y - y_0}{S_0} > 3\,,$$

обеспечивает приблизительно 90% доверительную вероятность для нормального и не нормального распределений аналитического сигнала (см. правило "трех сигм"). Таким образом, $y_{\text{мин}} = 3S_0 + y_0$. Если градуировочная функция линейна, то, подставив это значение в уравнение градуировочной функции, получается выражение для предела обнаружения наименьшего содержания вещества

$$C_{\text{мин}} = \frac{3S_0}{k}.$$

Проводя аналогичные оценки можно получить, что минимальное значение надёжного обнаружения

$$C_{\text{над}} = \frac{6S_0}{k}.$$

На практике выбирают значения $C_{\text{над}}$ в 5, 6 или 10 раз превышающее величину S_0/k.

Пример 6.8. Для экспериментальных данных (концентрация/сигнал, заданы в диапазоне B4:C13 рис. 6.12) определить значения наименьшей концентрации обнаружения вещества $C_{\text{мин}}$ и $C_{\text{над}}$ – предел "надёжного обнаружения".

На рис. 6.12 представлен скриншот вычислений, согласно которому на первом этапе рассчитываются угловой коэффициент градуировочной линии (ячейка G9), стандартное отклонение (ячейка G4) и затем вычисляются искомые значения $C_{\text{мин}}$ и $C_{\text{над}}$.

Для наглядности в левой части рис. 6.12 дан график рассчитанных зависимостей – градуировочная прямая (ячейки диапазона D4:D13) и предельная линия $C_{\min}$ (ячейки B15:C16).

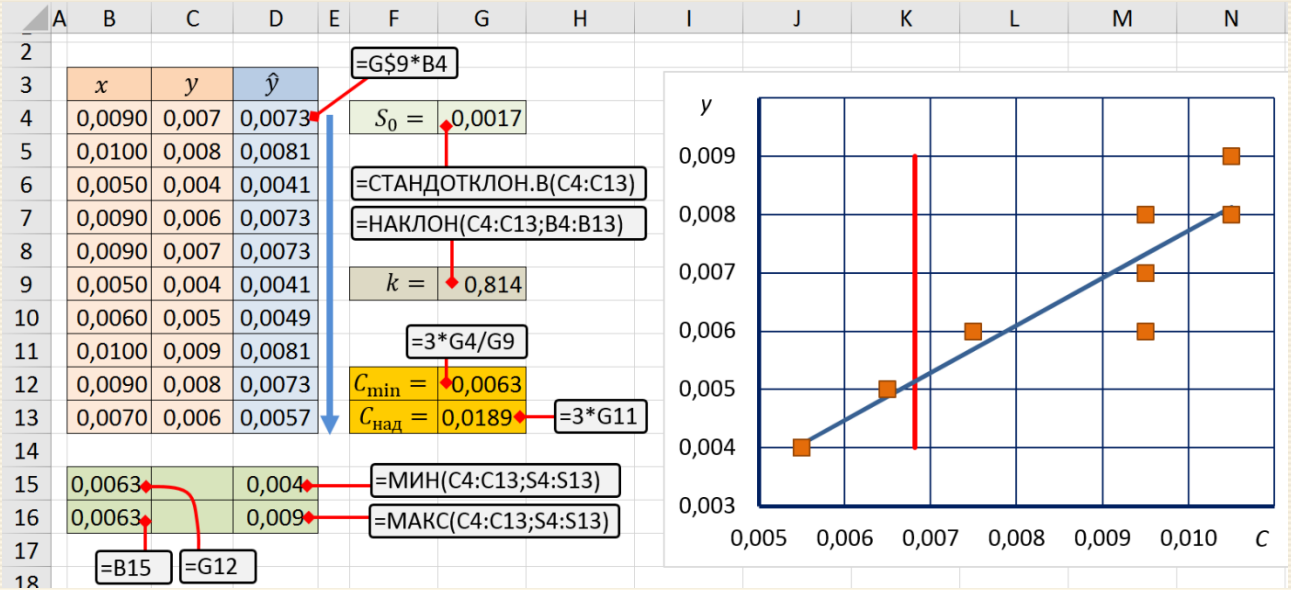


Рис. 6.12. Расчет пределов обнаружения вещества

7. Таблицы сопряженности. Корреляции качественных признаков

Сложные проблемы всегда имеют простые,
легкие для понимания неправильные решения.
Мерфология

Таблица сопряжённости, или **таблица контингентности**, или факторная таблица в статистике – средство представления совместного распределения двух переменных, предназначенное для исследования связи между ними.

Строки таблицы сопряжённости соответствуют значениям одной переменной, столбцы – значениям другой переменной.

Несмотря на кажущуюся примитивность используемых в таблице шкал данных, измерения могут быть использованы для проверки некоторых статистических гипотез и для вычисления показателей корреляции качественных признаков. В "свернутом" виде результаты наблюдений представляются таблицей сопряженности, состоящей из r строк и c столбцов, в ячейках которых проставлены частоты событий. На пересечении строки и столбца указывается частота совместного появления f_{ij} соответствующих значений двух признаков x_i и y_j. Сумма частот по строке $f_{\Sigma i}$ называется маргинальной частотой строки; сумма частот по столбцу $f_{\Sigma j}$ маргинальной частотой столбца. В таблице сопряжённости могут быть представлены как абсолютные, так и относительные частоты (в долях или процентах). Относительные частоты могут рассчитываться по отношению: *а)* к маргинальной частоте по строке; *б)* к маргинальной частоте по столбцу; *в)* к объёму выборки.

Типовой вид таблицы сопряженности приведен в таблице 7.1.

Таблица 7.1. – Переменные (выборки), разделенные по классам

	элементы y_1	элементы y_2	* * *	элементы y_c
элементы x_1	f_{11}	f_{12}	* * *	f_{1c}
элементы x_2	f_{21}	f_{22}	* * *	f_{2c}
* * *	* * *	* * *	* * *	* * *
элементы x_r	f_{r1}	f_{r2}	* * *	f_{rc}

Алгоритм анализа данных таблиц сопряженности включает следующие этапы.

1. Для ответа на вопрос о наличии статистической взаимосвязи с помощью критерия χ^2 следует сначала рассчитать ожидаемое количество наблюдений E_{ij} в каждой из ячеек

$$E_{ij} = \frac{\left(\sum_{j=1}^{c} f_{ij}\right) \cdot \left(\sum_{i=1}^{r} f_{ij}\right)}{\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij}} \qquad (7.1)$$

или, если использовать так называемые маргинальные частоты формулу расчета

$$f_{\Sigma j} = \sum_{i=1}^{r} f_{ij}\ , \quad f_{\Sigma i} = \sum_{j=1}^{c} f_{ij}\ ,$$

теоретических частот можно представить в виде

$$E_{ij} = \frac{f_{\Sigma j} \cdot f_{\Sigma i}}{f_{\Sigma\Sigma}}, \qquad \text{где} \qquad f_{\Sigma\Sigma} = \sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij} \qquad (7.2)$$

Ожидаемые значения E_{ij} не есть обязательно целые числа.

2. Рассчитывается значение критерия $\chi^2_{\text{эмп}}$. Общая формула имеет вид:

$$\chi^2_{\text{эмп}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(f_{ij} - E_{ij}\right)^2}{E_{ij}}, \qquad (7.3)$$

где i – номер строки ($i = 1, \dots, r$),

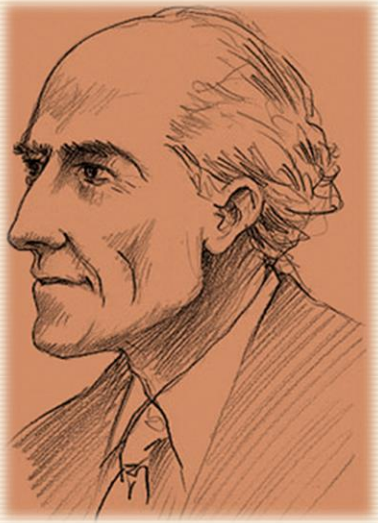
j – номер столбца ($j = 1, \dots, c$);

r, c – количество строк и столбцов таблицы;

в случае таблиц 2×2 r=2 и c=2;

f_{ij} – фактическое наблюдение в ячейке ij.

3. Для заданного уровня значимости α по числу степеней свободы $df = (r-1)\cdot(c-1)$ определяется критическое значение $\chi^2_{\text{крит}}$. Сравнением значений $\chi^2_{\text{эмп}}$ и $\chi^2_{\text{крит}}$ делается заключение о верности нулевой гипотезы.



Karl (Charles) Pearson

7.1. Критерий χ^2 для таблиц сопряженности. Таблицы 2×2

Критерий χ^2 для таблиц 2×2. Поправка Yates'a

Простейшая таблица сопряженности 2x2 (табл. 7.2) содержит распределение частот f_{ij} двух качественных признаков, измеренных в номинальной шкале с двумя "делениями" (дихотомическая шкала).

Таблица 7.2 – Общий вид таблицы сопряженности 2×2

		признак *B*	
		да, присутствует	нет, отсутствует
признак *A*	да, присутствует	f_{11} число элементов выборки, обладающих признаками *A* и *B* одновременно	f_{12} число элементов выборки, обладающих признаком *A*, но не обладающих признаком *B*
	нет, отсутствует	f_{21} число элементов выборки, обладающих признаком *B*, но не обладающих *A*	f_{22} число элементов выборки, не обладающих ни одним из признаков *A* и *B*

Нулевая гипотеза: статистически значимая связь между признаками *A* и *B* отсутствует.

Наиболее распространенной, в частности в биологических и педагогических исследованиях, является проблема установления статистической значимости различий в распределении частот в одной и той же группе объектов до и после применения контролируемых воздействий (табл. 7.3).

Таблица 7.3 – Вид таблицы сопряженности 2×2 для одной и той же группы объектов

	контролируемое воздействие	
	положительный результат	отрицательный результат
до воздействия	f_{11} число объектов выборки с положительным эффектом до и после воздействия	f_{12} число объектов выборки с положительным эффектом до воздействия и отрицательным после
после воздействия	f_{21} число элементов выборки, с положительным результатом после и отрицательным до воздействия	f_{22} число объектов выборки с отрицательным эффектом до и после воздействия

Критерий χ^2 для таблиц сопряженности был предложен Карлом Пирсоном в 1900 году*. С помощью данного критерия оценивается значимость различий между фактическим (выявленным в результате исследования) количеством исходов или качественных характеристик выборки, попадающих в каждую категорию, и теоретическим количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы.

*Pearson E. S. The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table // Biometrika, 1947. Vol. 34. P. 139–167.

Для применения критерия χ^2 Пирсона требуется, в общем случае, соблюдение следующих условий:

1. Номинальные или порядковые данные (возможно создание категорий из непрерывных данных).
2. Независимость наблюдений (отбор участников исследования из генеральной совокупности производится независимо друг от друга).
3. Независимость групп (метод нельзя применять для исследований типа "до – после").
4. Ожидаемое (не фактическое) число наблюдений в любой из ячеек должно быть не менее 5 (или 10) для четырехпольных таблиц.
5. Доля ячеек с ожидаемым числом наблюдений менее 5 не должна превышать 20 % для многопольных таблиц.
6. Для расчета критерия χ^2 используются только абсолютные фактические и ожидаемые числа (проценты и доли для расчетов не должны использоваться).

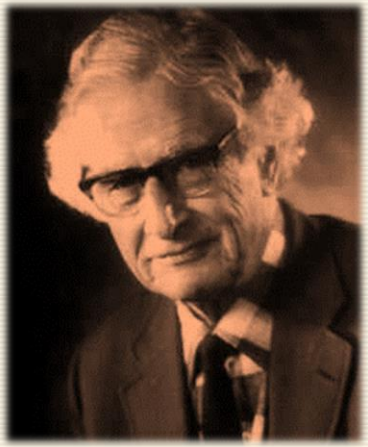
Вычисленное значение критерия χ^2 изменяется скачкообразно, так как основывается на частотах, которые являются целыми числами. В то же время табличные значения для распределения χ^2 составлены для непрерывной шкалы, поэтому в 1934 году английский статистик Фрэнк Йейтс (Frank Yates) предложил поправку* на непрерывность, которая сейчас известна под названием поправки Йейтса (Yates's correction).

*Yates F. Contingency tables involving small numbers and the chi-square test // Supplement to the Journal of the Royal Statistical Society. – 1934. – Vol. 1. – P. 222.

Поправка заключается в вычитании 0.5 из абсолютного значения разности между фактическим и ожидаемым количеством наблюдений в каждой ячейке, что ведет к уменьшению величины критерия:

$$\chi^2_{\text{эмп}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(\left|f_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}}. \qquad (7.4)$$

В части литературных источников отмечается, что применение поправки Yates'a целесообразно. В некоторых учебниках оговаривается, что ее применение необходимо при небольших объемах выборки и/или при количестве ожидаемых наблюдений в любой из ячеек менее 5 или менее 10. В ряде работ предлагается поправку на непрерывность применять всегда. Тем не менее, не все статистики согласны с необходимостью применением поправки, так как было показано, что она может приводить к получению заниженных значений критерия, а значит, увеличивает вероятность ошибки второго типа, то есть вероятности необнаружения различий там, где они есть. При наличии больших выборок различия в значениях критерия $\chi^2_{\text{эмп}}$, получаемых с использованием поправки Йейтса и без нее незначительны, однако при малых выборках различия могут быть существенными.



Frank Yates

Следует помнить, что поправка Yates'a применяется только для таблиц 2×2, то есть при анализе двух дихотомических переменных.

Критическое значение $\chi^2_{\text{кр}}$ определяется для числа степеней свободы $df = (r-1)(c-1)$; значение критерия $\chi^2_{\text{эмп}}$ сравнивается с критическим значением. Если $\chi^2_{\text{эмп}} \leq \chi^2_{\text{кр}}$, то подтверждается нулевая гипотеза об отсутствии статистически значимой связи между изучаемыми признаками и исходами. Если $\chi^2_{\text{эмп}} > \chi^2_{\text{кр}}$, то на основании применения критерия χ^2 Пирсона нулевая гипотеза может быть отвергнута.

Вероятность того, что, отвергая нулевую гипотезу, совершается ошибка (первого рода), численно равна уровню значимости α, задаваемого при проверке гипотезы.

Интерпретация χ^2 теста зачастую усложняется, когда в таблице сопряженности имеются ячейки с нулевыми значениями наблюдаемых частот (см. критерий Фишера точный). Дело в том, что если пара (x_i, y_j) значений переменных не наблюдалась в выборке, то это может означать, что объем выборки не столь велик, чтобы зафиксировать такую редкую комбинацию, либо что данная комбинация невозможна по каким-то объективным причинам. В последнем случае действительное число степеней свободы анализируемой системы меньше числа степеней свободы таблицы сопряженности, на основании которого произведена оценка уровня значимости χ^2 теста.

Корректировка применения χ^2 теста возможна лишь в том случае, если эмпирические данные, наполняющие таблицу сопряженности, есть результаты независимой случайной выборки относительно большого объема n. Последнее требование вызвано тем, что выборочное распределение χ^2 аппроксимирует табличное распределение статистики χ^2 только при больших n. Естественно, что возникает вопрос о том, насколько велико должно быть n, чтобы иметь возможность использовать данный тест. Ответ на этот вопрос зависит от числа ячеек и величин маргинальных сумм.

Вообще говоря, чем меньше число ячеек и чем более близки между собой по величине маргиналы, тем меньше может быть n.

Существует, однако, практическое число, позволяющее оценить снизу по n диапазон возможного применения критерия χ^2: если в данной таблице сопряженности любая из теоретических ожидаемых частот E_{ij} в ячейке (i, j) не больше 5, то рекомендуется произвести, если это возможно, модификацию таблицы либо воспользоваться другим критерием.

В общем случае корректировка таблицы размера $r \times c$ затруднительна. Практика показала, что если число ячеек велико, а ожидаемые частоты, равные или меньше пяти, встречаются лишь в одной-двух ячейках, то проведение корректировки нецелесообразно; во всех иных случаях разумной альтернативой является объединение категорий (градаций) с тем, чтобы элиминировать (удалить) подобные ячейки. Естественно, такое объединение должно быть таким, чтобы получаемая в результате комбинация не была содержательно бессмысленной.

Далее даются реализации изложенных принципов на примерах.

Пример 7.1 Определить, зависит ли успешность прохождения студентами тестирования по математике от прохождения ими адаптивного курса в начале обучения в вузе. В группе из 20 человек, посещавших адаптивный курс, успешно справились с тестированием 12 человек, а в группе из 25 человек, не посещавших адаптивный курс, – 10 человек. В данном случае признак *A* – посещение адаптивного курса, *B* – успешность прохождения тестирования.

Последовательно выполняются следующие действия (рис. 7.1).

1. Заносятся исходные данные (диапазон E5:F6) – результаты тестирования, а также величину уровня значимости (C3).

2. Маргинальные и теоретические частоты, а также значения для вычисления критерия $\chi^2_{\text{эмп}}$ определяются в соответствии с формулами, приведенными на рис. 7.1. Формула ячейки G5 служит для автозаполнения ячейки G6, по E7 заполняются F7 и G7.

3. Теоретические частоты определяются формулой ячейки I5, которая "растягивается" на диапазон I5:J6.

4. В ячейке F9 подсчитывается значение $\chi^2_{\text{эмп}}$, а в F10 величина критического значения $\chi^2_{\text{крит}}$. Поскольку $\chi^2_{\text{эмп}} = 1.78 < \chi^2_{\text{крит}} = 3.84$, то нулевая гипотеза о независимости успешности выполнения студентами заданий от прохождения ими адаптивного курса подтверждается.

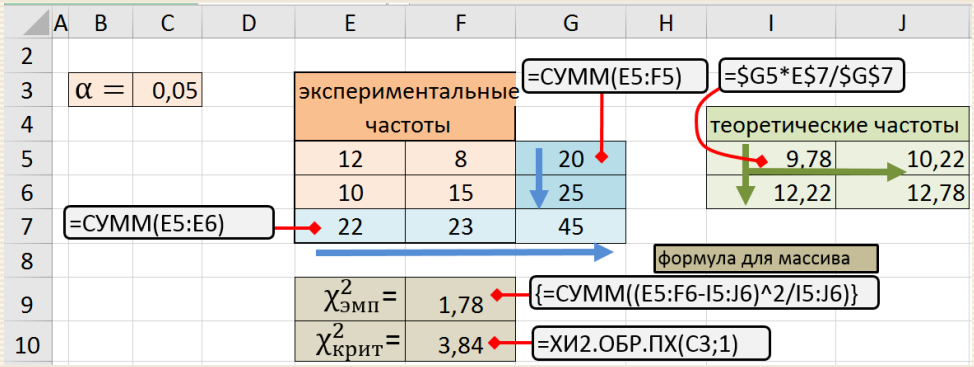


Рис. 7.1. Скриншот для схемы вычислений критерия χ^2 для таблиц 2×2

Таким образом, успешность выполнения студентами тестовых заданий по математике не зависит от прохождения ими адаптивного курса в начале обучения в вузе.

Скриншот решения этой же задачи с учетом поправки Yates'a (расчета критерия по соотношению (7.4)) приведен на рис. 7.2.

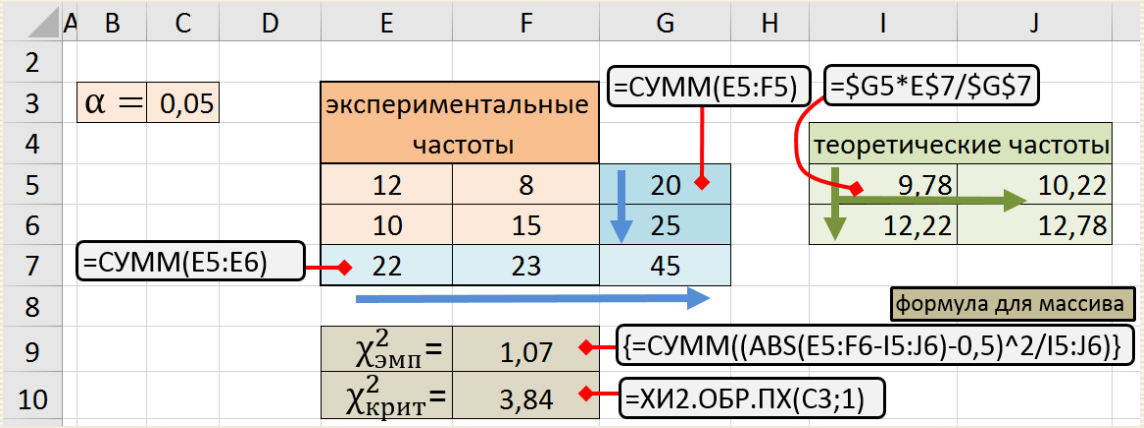


Рис. 7.2. Скриншот листа Excel расчета χ^2 с поправкой Yates'a

Критерий χ^2 с поправкой на правдоподобие

В сомнительных случаях, когда, например, анализ данных с учетом и без учета поправки Йейтса дают противоположные результаты, то в расчетах можно использовать вычисление критерия χ^2 с поправкой на метод максимального правдоподобия*, при котором оценка неизвестного параметра производится путем максимизации функции правдоподобия. Расчет $\Lambda\chi^2$ производится по формуле

$$\Lambda\chi^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij} \cdot \ln\left(\frac{f_{ij}}{E_{ij}}\right), \qquad (7.5)$$

после чего полученные значения критерия $\Lambda\chi^2$ сравниваются с критическим $\chi^2_{\text{крит}}$ прежним образом. При больших выборках значения $\Lambda\chi^2$ и χ^2 приблизительно равны. При малых выборках значение $\Lambda\chi^2$ обычно несколько меньше, а потому считается некоторыми специалистами предпочтительнее.

Пример 7.2 Для данных примера 7.1 на рис. 7.3 дан скриншот листа завершенных вычислений с поправкой на правдоподобие. Критерий $\Lambda\chi^2 = 1.79 < \chi^2_{\text{крит}} = 3.84$, что опять-таки подтверждает нулевую гипотезу о независимости успешности выполнения студентами заданий от прохождения ими адаптивного курса.

*Field A. Discovering statistics using SPSS / SAGE Publications, 2005. – 779 p.

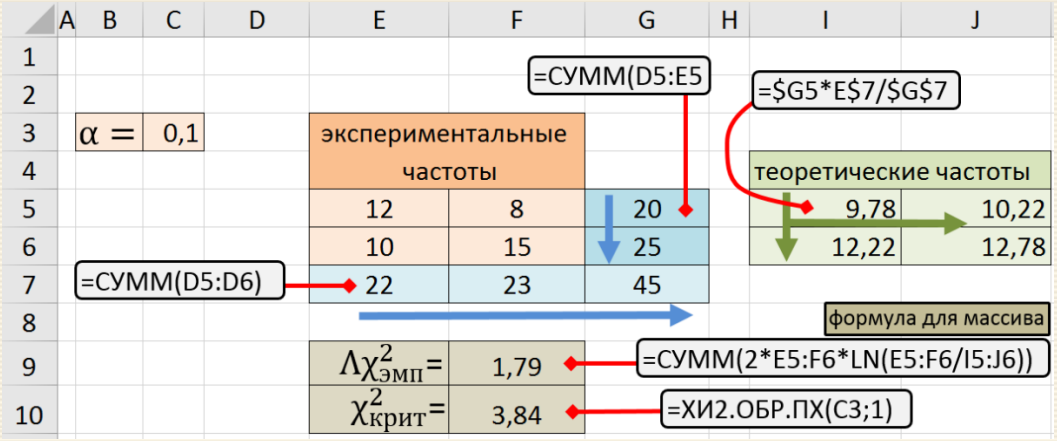


Рис. 7.3. Схема вычислений для критерия с поправкой на правдоподобие

Критерий V Крамера для таблиц сопряженности

Не исключено, что при анализе одних и тех же данных различными критериями можно попасть в ситуацию, когда какие-то критерии говорят о том, что нулевую гипотезу можно отвергнуть, а другие – наоборот. В частности, даже сильную статистическую связь сложно выявить при малом числе наблюдений, в то время как при больших выборках даже слабая и маловажная связь становится статистически значимой. Поэтому ошибочно было бы делать вывод о силе взаимосвязи между переменными только на основании достигнутого уровня значимости P, а также сравнивать по значениям P силу взаимосвязи между признаками в совокупностях с разным числом наблюдений.

В некоторых исследованиях необходимо не только представлять достигнутые уровни значимости при проверке статистических гипотез, но и оценивать величину эффекта (effect size), то есть силу связи между признаками*.

Критерии, оценивающие силу связи между номинальными переменными, могут принимать значения от 0 до 1. Они не могут иметь отрицательных значений, так как данные, измеряемые по номинальной шкале, не имеют порядкового отношения, что не позволяет изучать направление зависимости.

Для оценки силы связи между признаками таблиц сопряженности можно применять критерий V Крамера (Cramer's V), основанного на критерии χ^2 и значение которого варьируются от 0 до 1:

$$V = \sqrt{\frac{\chi^2_{\text{эмп}}}{n \cdot (\min(r, c) - 1)}}, \quad n = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij}.$$

*Wilkinson L. Statistical methods in psychology journals: guidelines and explanations // American Psychologist. – 1999. – Vol. 54. – P. 594–604.

Пример 7.3 Для данных примера 7.1 на рис. 7.4. дан скриншот вычисления критерия V Крамера. Основная часть схемы повторяет таковую для вышеприведенных примеров.

Если интерпретировать полученное значение V согласно рекомендациям Rea & Parker* (табл. 7.4, рис. 7.4), то можно заключить, что для рассматриваемы данных имеется слабой силы связь между факторами адаптивного обучения и положительным исходом тестирования.

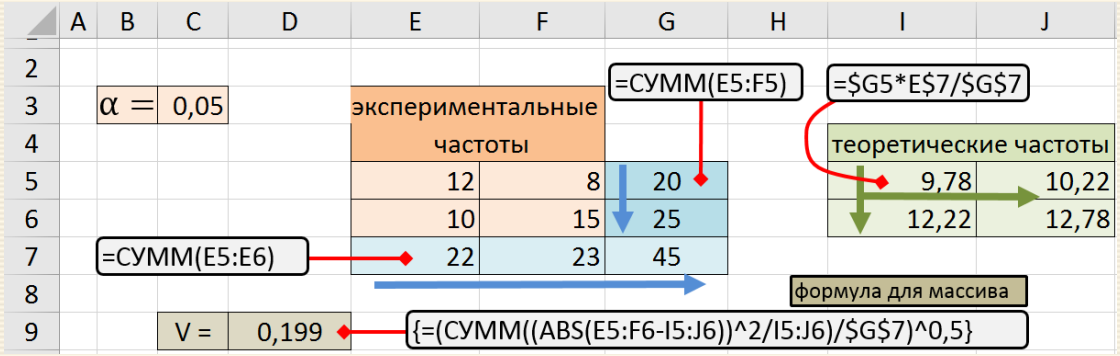


Рис. 7.4. Схема вычислений для критерия V Крамера

*Richard A. Parker, & Louis M. Rea. Designing and Conducting Survey Research: A Comprehensive Guide / 4 edition, Jossey-Bass, 2014. – 360 p.

Таблица 7.4. – Интерпретация значений критерия V Крамера

Значение критерия V	Сила взаимосвязи
[0.0 – 0.1)	Несущественная
[0.1 – 0.2)	Слабая
[0.2 – 0.4)	Средняя
[0.4 – 0.6)	Относительно сильная
[0.6 – 0.8)	Сильная
[0.8 –1.0]	Очень сильная

Пример 7.4 На сайте В. Леонова* приводится интерактивный пример анализа сопряженности двух качественных признаков 2×2. А именно, рассматривается вопрос: "Можно ли утверждать, исходя из данных конкретной выборки, что два исследуемых дискретных качественных признака независимы друг от друга в генеральной совокупности?" Иными словами, определяется то, что между этими признаками отсутствует взаимосвязь. Если эта гипотеза будет отвергнута, то с высокой долей вероятности можно утверждать, что такая зависимость существует. Реальным содержанием примера является исследование наличия взаимосвязи между приемом контрацептивных таблеток матерями, и желтухой у детей, получающих грудное вскармливание.

*http://www.biometrica.tomsk.ru/freq1.htm

В этом примере (рис. 7.5) у 33 матерей, принимавших таблетки, дети болели желтухой, а у 24 матерей, также принимавших таблетки, дети не болели. Далее, у 14 матерей, которые не принимали таблетки, дети болели желтухой; у 45 матерей, не принимавших таблетки, дети не болели желтухой.

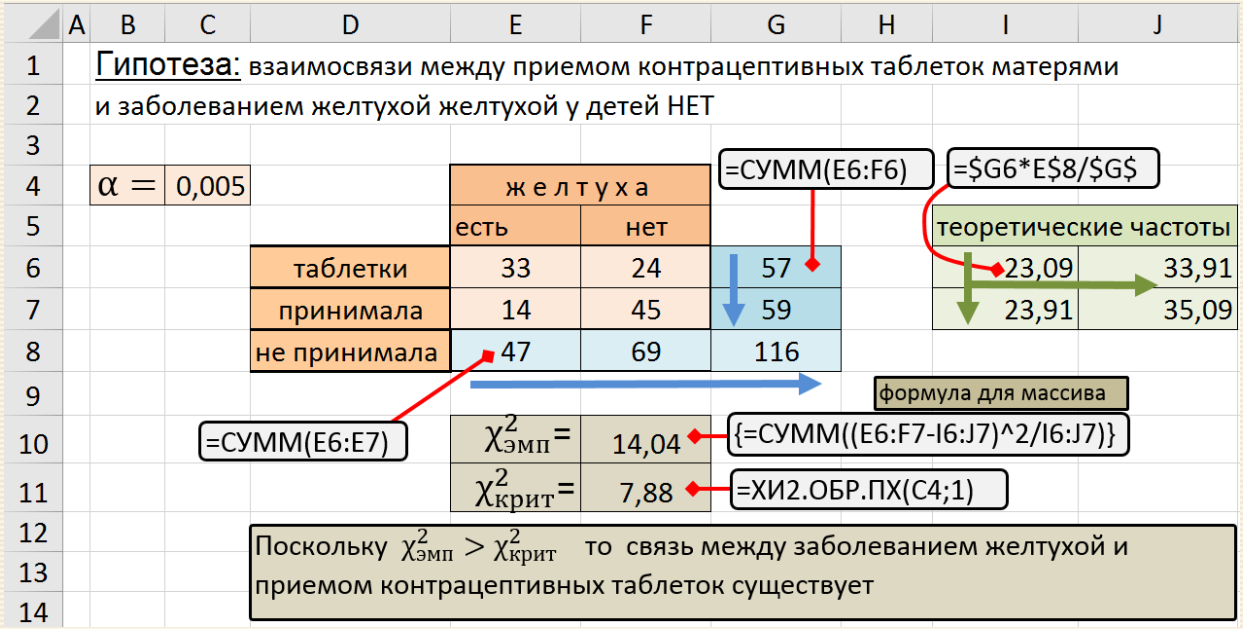


Рис. 7.5. Решение о наличии взаимосвязи приема таблеток и болезнью желтухой

Из сравнения $\chi^2_{эмп}$ с $\chi^2_{крит}$ для α=0.005 видно, что вычисленное χ^2 превосходит критическое. Другими словами – выборки статистически различны и поэтому гипотеза о независимости между заболеванием желтухой и приемом контрацептивных таблеток отвергается при уровне значимости α < 0.005, т.е. какая-то зависимость между заболеванием и приемом таблеток существует.

Точный критерий Фишера (Fisher's exact test)

Если вам непонятно какое-то слово в техническом тексте,
не обращайте на него внимания.
Текст полностью сохраняет смысл и без него.
Мерфология, Закон Купера

Критерий χ^2 применим для анализа таблиц сопряженности 2×2, если ожидаемые значения в любой из ее клеток не меньше 5. Когда число наблюдений невелико, это условие не выполняется и критерий χ^2 неприменим.

В этом случае используют точный критерий Фишера. Он основан на переборе всех возможных вариантов заполнения таблицы сопряженности при данной численности групп.

Общий вид таблицы сопряженности 2×2

		признак *B*	
		да, присутствует	нет, отсутствует
признак *A*	да, присутствует	f_{11}	f_{12}
	нет, отсутствует	f_{21}	f_{22}

$$P_i = \frac{C_{f_{\Sigma i1}}^{f_{11}^i} C_{f_{\Sigma i2}}^{f_{21}^i}}{C_{f_{\Sigma i+\Sigma j}}^{f_{\Sigma j1}}} = \frac{(f_{11}+f_{12})! \cdot (f_{21}+f_{22})! \cdot (f_{11}+f_{21})! \cdot (f_{12}+f_{22})!}{(f_{11}+f_{12}+f_{21}+f_{22})! \cdot f_{11}^i! \cdot f_{12}^i! \cdot f_{21}^i! \cdot f_{22}^i!}$$

Если f_{11} есть минимальное значение в исходной таблице, то вероятность можно определить через сумму

$$P = \sum_{i=0}^{f_{11}} P_i = \frac{(f_{11}+f_{12})! \cdot (f_{21}+f_{22})! \cdot (f_{11}+f_{21})! \cdot (f_{12}+f_{22})!}{(f_{11}+f_{12}+f_{21}+f_{22})! \cdot f_{11}^i! \cdot f_{12}^i! \cdot f_{21}^i! \cdot f_{22}^i!} + \\ + \sum_{i=0}^{f_{11}} \frac{1}{(f_{11}-i)! \cdot (f_{21}+i)! \cdot (f_{22}-i)! \cdot (f_{12}+i)!}. \quad (7.6)$$

Полученная сумма представляет собой значение P для одностороннего варианта точного критерия Фишера.

Нулевая гипотеза состоит в том, что выборки однородны, то есть, например, между курением и кашлем нет никакой связи.

"Классические" правила использования точного критерия Фишера следующие:

1. Вычисляется вероятность для исходной таблицу;
2. Строятся все возможные варианта заполнения таблицы при неизменных суммах по строкам и столбцам. Для этого в одной из клеток проставляются все целые числа от 0 до максимально возможного, пересчитывая числа в остальных клетках так, чтобы суммы по столбцам и строкам оставались неизменными;

3. Вычисляется вероятности для полученных таблиц;

4. Суммируются вероятности получить исходной таблицы (для одностороннего критерия), и не превышающие оную (для двухстороннего критерия);

5. По заданному уровню значимости принимается статистическое решение.

Если вероятность P не превосходит уровень значимости α, тогда нулевую гипотезу о независимости признаков отклоняют. В противном случае нет оснований отклонять нулевую гипотезу.

Sir Ronald Aylmer Fischer

Пример 7.5 Определить, зависит ли успешность прохождения студентами тестирования по математике от прохождения ими адаптивного курса в начале обучения в вузе для задачи типа примера 7.1, когда исходная таблица имеет вид

12	8
10	15

Алгоритм о д н о с т о р о н н е г о к р и т е р и я реализуется следующим образом.

1. Вводится исходная таблица (ячейки B2:C3 на рис. 7.6), рассчитываются маргинальные частоты (ячейки D2:D3; B4:B5), суммарная частота (ячейка D4) и уровень вероятности в ячейке E5.

Заметим, что при больших значениях частот могут возникнуть ошибки типа "переполнению" при вычислении факториалов или существенная потеря точности при неправильной организации последовательности выполнения операций умножения и деления. Преодоления этих трудностей возможно методом предварительного логарифмирования соотношения (7.6), использования приближенных соотношений расчета факториала и т.д. В частности, при $n > 20$ можно использовать формулу

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n.$$

Устанавливается положение элемента таблицы, имеющего минимальное значение.

2. Оформленный в п.1 блок формул и данных копируется в рядом стоящее место, значение в позиции минимального элемента уменьшается на 1, а остальные элементы таблицы корректируются таким образом, чтобы сохранились маргинальные частоты. Рассчитывается соответствующий этой таблице уровень вероятности.

3. Если в ячейке минимального значения не достигнуто нулевое значение, то п.2 повторяется.

4. Суммируются все значения полученных вероятностей, выполняется сравнение с заданным уровнем значности и формулируется вывод о справедливости нулевой гипотезы.

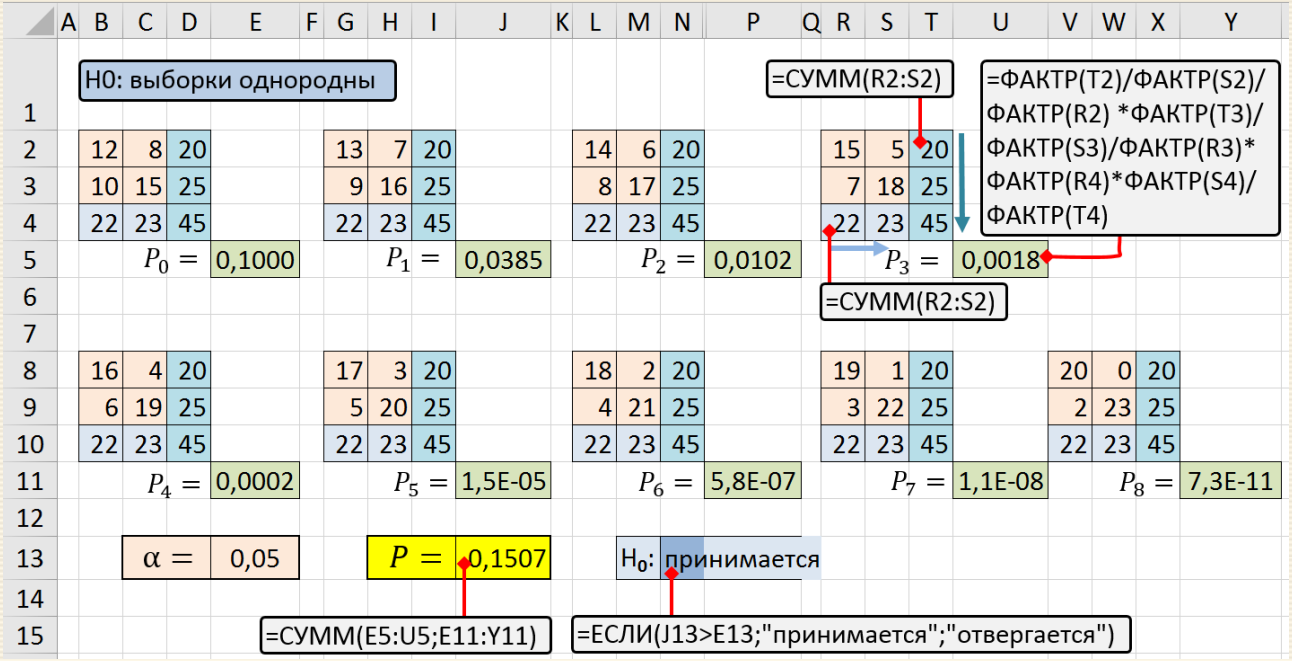


Рис. 7.6. "Классический" алгоритм точного теста Фишера

Пример 7.6 Определить, зависит ли успешность прохождения студентами тестирования по математике от прохождения ими адаптивного курса в начале обучения в вузе для задачи типа примеров 7.1 и 7.5, когда исходная таблица имеет вид

9	10
13	3

В MS Excel имеется функция, возвращающая плотность гипергеометрического распределения, лежащего в основе точного теста Фишера ГИПЕРГЕОМ.РАСП (

<число успехов в выборке>; <размер выборки>;
<число успехов в совокупности>; <размер совокупности>;
<интегральная>),

которая может существенно упростить вычисления.

Поскольку в точном критерии анализ связан с отношение частот (вероятностей), то их сравнение не зависит от последовательности строк или столбцов. Алгоритм анализа можно и в этом отношении упростить (рис. 7.7), если исходную таблицу перестановками строк или столбцов привести к виду, когда минимальный элемент таблицы будет в левом верхнем углу. В рассматриваемом примере после перестановки столбцов таблица будет выглядеть так

3	13
10	9

В этом случае искомое значение вероятности для о д н о с т о р о н н е г о к р и т е р и я можно вычислить формулой (для работы с массивами) =СУММ(ГИПЕРГЕОМ.РАСП(

<массив целых чисел от 0 до минимального в таблице, т.е. от 0 до B2>;
<размер выборки успехов, что равно сумме B2+B3>;
<число успехов в совокупности, определяется как B2+C2>;
<размер совокупности, рассчитываемы функцией СУММ(B2:C3)>;
<используется дифференциальное распределение, т.е. интегральная =ЛОЖЬ))

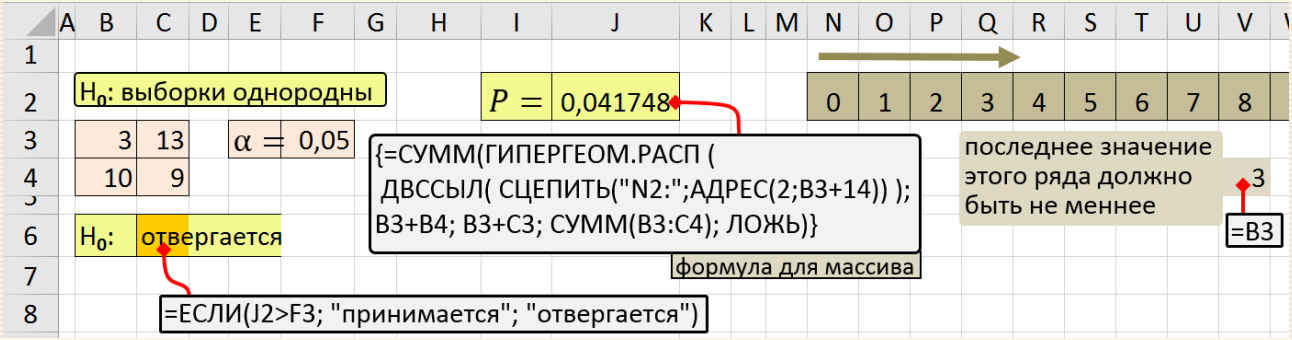


Рис. 7.7. Односторонний точный тест Фишера

Пример 7.7 Для тех же данных примера 7.6 скриншот расчета двухстороннего критерия точного теста Фишера дано на рис. 7.8.

Точный тест Фишера, в отличие от критерия χ^2, имеет одно- и двусторонний варианты. Большинство исследователей используют односторонний вариант скорее всего потому, что он дает меньшие значения вероятности P, что выглядит более привлекательно в представлении результатов.

Алгоритм получения двустороннего варианта теста состоит в переборе и суммировании всех возможных вариантов заполнения таблицы сопряженности при неизменных маргинальных суммах.

Реально вычисления могут производится по следующей схеме:

- на первое место ставится столбец с меньшей маргинальной суммой;
- рассчитывается вероятность P_0 для исходной таблицы =ГИПЕРГЕОМ.РАСП(B2; D6; D7; D8; ЛОЖЬ);
- суммируются все вероятности P_Σ (для значений от нуля до маргинального значения первого столбца) для тех величин, не превосходящих P_0;
- P-значение двустороннего точного варианта критерия Фишера определяется суммой $P = P_0 + P_\Sigma$.

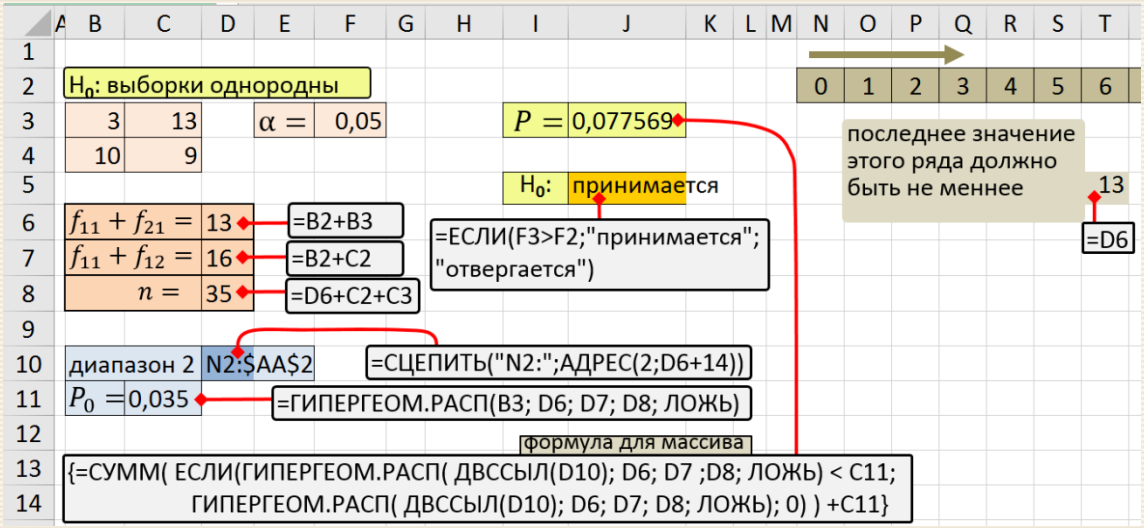


Рис. 7.8. Двухсторонний точный тест Фишера

Пример 7.8 Определить, зависит ли успешность прохождения студентами интернет-тестирования по математике от прохождения ими адаптивного курса в начале обучения в вузе для задачи типа примера 7.1 с исходной таблицей

0	5
10	9

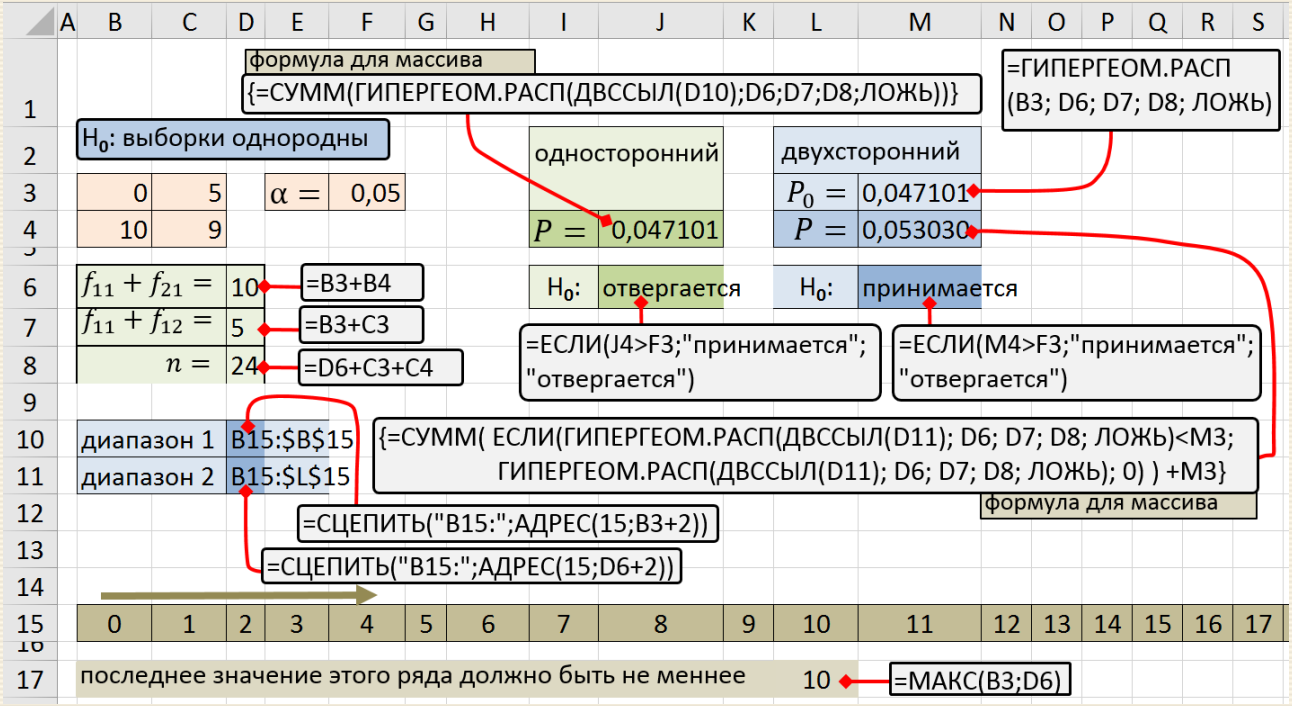


Рис. 7.9. Точный тест Фишера (одностороннее и двухстороннее значения)

7.2. Критерий χ^2 для таблиц сопряженности $r \times c$

Если вариацию качественного признака изучаемого явления требуется разбить не на две группы (как в случае дихотомического признака), а на несколько групп, то соответствующий числовой материал располагают в виде таблицы с несколькими строками (r) и столбцами (c).

Схема вычислений аналогична тем, что изложена для таблиц размером 2×2 и ниже рассматривается на нескольких примерах.

Пример 7.9 По данным диагностики темперамента у 93 подростков с помощью опросника Г. Айзенка и диагностики профессиональных предпочтений с помощью опросника Климова определяли, влияет ли тип темперамента на профессиональные предпочтения*. Результаты исследования представлены в табл. 7.5.

Таблица 7.5. – Профессиональные предпочтения

Тип темперамента	Профессиональные предпочтения		
	Техника	Знаковая техника	Человек
Холерик	1	5	15
Сангвиник	13	5	9
Меланхолик	2	17	3
Флегматик	16	3	4

*Т. А. Юрьева, А. П. Филимонова, Н. А. Чалкина. Статистическая оценка связи между качественными признаками в педагогических исследованиях // Вестник Амурского государственного университета. 2014. Вып. 65: Сер. Естеств. и экон. науки. – С. 11-16.

Последовательно выполняются следующие действия (см. рис. 7.10).

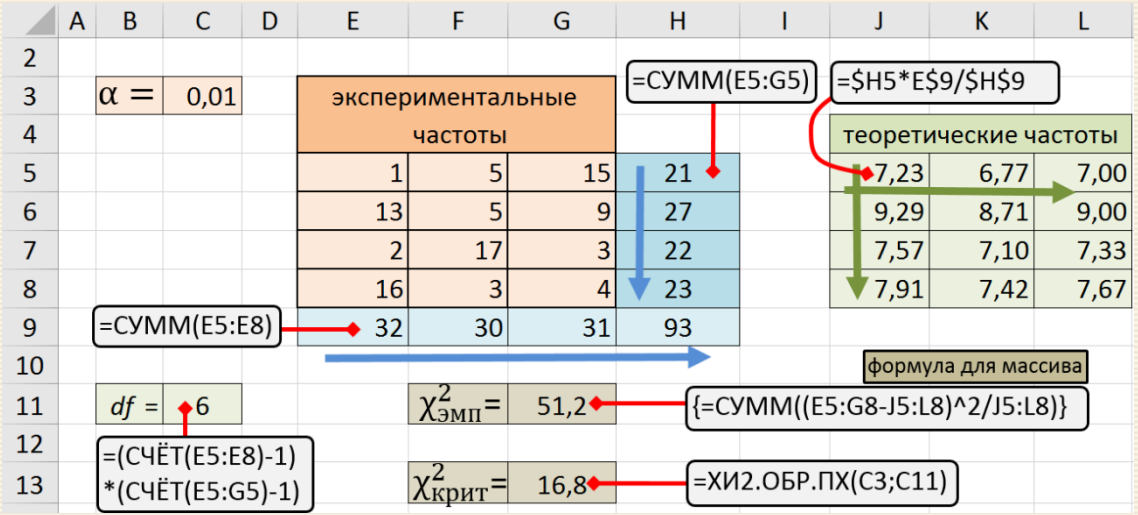


Рис. 7.10. Скриншота листа Excel для схемы вычислений критерия χ^2 для таблиц 4×3

1. Заносятся исходные данные (ячейки E5:G8) – результаты тестирования, а также величину уровня значимости (ячейка C3).

2. Маргинальные и теоретические частоты, а также значения для вычисления критерия $\chi^2_{эмп}$ определяются в соответствии с формулами, приведенными на рис. 7.10. Формула ячейки H5 служит для автозаполнения до G8, по E9 заполняются – заполнение до H9.

3. Теоретические частоты определяются формулой ячейки J5, которая "растягивается" на диапазон J5:L8.

4. В ячейке G11 подсчитывается значение $\chi^2_{\text{эмп}}$, а в G12 величина критического значения $\chi^2_{\text{крит}}$. Количество степеней свободы $df = (r-1)(c-1)$ определяется числом строк и столбцов таблицы в ячейке C11. Поскольку $\chi^2_{\text{эмп}} = 51.2 > \chi^2_{\text{крит}} = 16.8$, то нулевая гипотеза об отсутствии влияния типа темперамента на профессиональные предпочтения не подтверждается.

Таким образом, тип темперамента влияет на профессиональные предпочтения.

Пример 7.10 В Тегульдетском районе Томской области в 1999 и 2002 годах В. П. Перевозкин выполнил сбор материала и получил распределение сочетаний хромосомных инверсий комаров *Anopheles messeae* (данные приведены в ячейках C6:G7 на рис. 7.11).

Вопрос: изменилось или нет распределение по сочетанию инверсий *Anopheles messeae* за три года? Иными словами – выборки 1999 и 2002 года статистически одинаковы или различны?

Как обычно сравниваются расчетное (эмпирическое) и критическое значения критерия χ^2. Поскольку в данном случае $\chi^2_{\text{эмп}} < \chi^2_{\text{крит}}$ (5.14<9.49), то можно сделать вывод: гипотеза о неизменности распределения *Anopheles messeae* подтверждается.

На рис. 7.11. представлен скриншот решения данной задачи. Теоретические частоты инверсий рассчитываются по соотношениям (7.1) автозаполнением ячеек C12 и C13 до столбца G.

Нулевая гипотеза: Распределение *Anopheles* по кариотипам в Тегульдете с 1999 по 2002 год не изменилось

год выборки	кариотипы, фактические данные				
	XL11 2R11	XL22 2R11	XL11 2R01	XL00 2R00	XL11 2R00
1999	38	26	16	62	69
2002	37	18	14	78	93

$\alpha =$ 0,05

=ХИ2.ОБР.ПХ(J4; (СЧЁТ(C6:G6)-1) *(СЧЁТ(C6:C7)-1))

=СУММ(C$6:C$7)/СУММ($C$6:$G$7)*СУММ($C6:$G6)

год выборки	кариотипы, теоретические частоты				
	XL11 2R11	XL22 2R11	XL11 2R01	XL00 2R00	XL11 2R00
1999	35,1	20,6	14,0	65,5	75,8
2002	39,9	23,4	16,0	74,5	86,2

$\chi^2_{крит} =$ 9,49

$\chi^2_{эмп} =$ 5,14

формула для массива {=СУММ((C6:G7-C12:G13)^2/C12:G13)}

Поскольку $\chi^2_{эмп} < \chi^2_{крит}$ то гипотеза о неизменности распределения *Anopheles* подтверждается

Рис. 7.11. Расчет соответствия распределений сочетаний хромосомных инверсий

Пример 7.11 Дигибридное скрещивание; анализ гибридов второго поколения. Доказать, что у мух, у которых учитываются две пары альтернативных признаков, в гибридах второго поколения наблюдается расщепление по фенотипическим классам 9 : 3 : 3 : 1. Отношение, полученное экспериментально, следующее: 135 : 51 : 54 : 18.

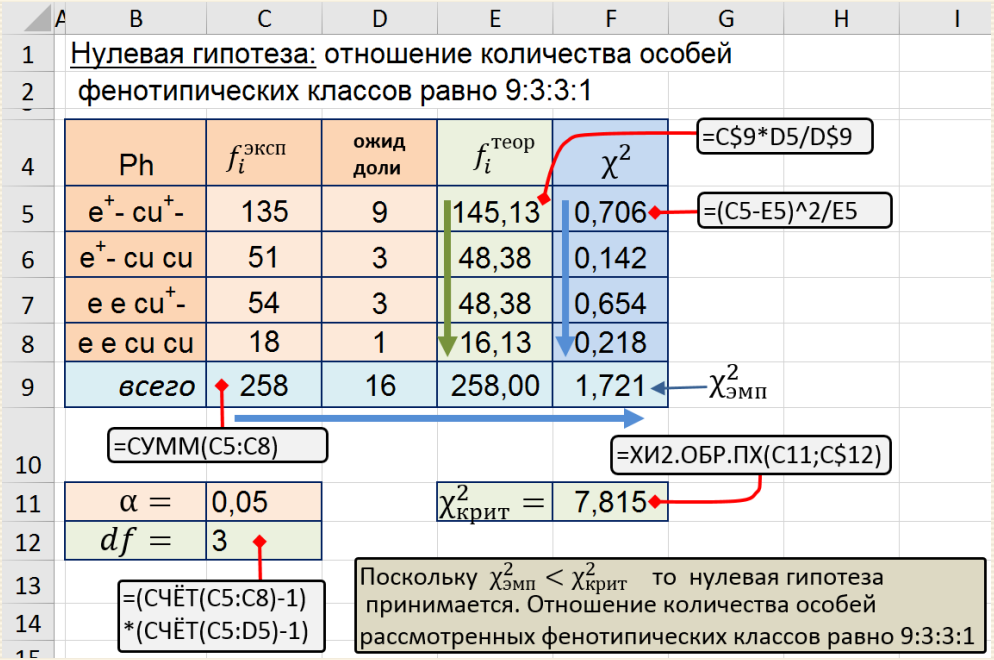
Алгоритм анализа заключается в построении ряда теоретических частот

$$f_i^{\text{теор}} = \frac{\sum f_i^{\text{эксп}}}{\sum f_i^{\text{ожид}}} \cdot f_i^{\text{ожид}}$$

для данного количества экспериментальных значений $f_i^{\text{эксп}}$, вычисление $\chi^2_{\text{эмп}}$ и $\chi^2_{\text{крит}}$ с их последующим сравнением (рис. 7.12).

Поскольку $\chi^2_{\text{эмп}} = 1.72 < \chi^2_{\text{кр}} = 7.82$, то формулируется вывод: нулевая гипотеза принимается по уровню значимости α = 0.05**.** Отношение количества особей рассмотренных фенотипических классов равно 9 : 3 : 3 : 1.

Практически все промежуточные вычисления, используемые для реализации "классического" алгоритма расчета $f_i^{\text{теор}}, f_i^{\text{ожид}}$ (рис. 7.12), могут быть опущены, если использовать возможности MS Excel в части использования инструментария работы с формулами для массивов (рис. 7.13).



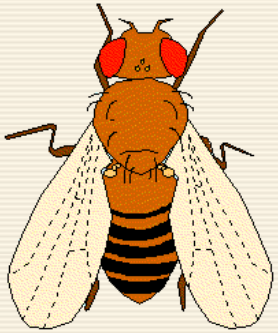


Рис. 7.12. Анализ справедливости соотношения классов 9 : 3 : 3 : 1 для гибридов второго поколения

Нулевая гипотеза: отношение количества особей фенотипических классов равно 9:3:3:1

Ph	$f_i^{эксп}$	$f_i^{ожид}$
e^+- cu^+-	135	9
e^+- cu cu	51	3
e e cu^+-	54	3
e e cu cu	18	1

$\alpha =$ 0,05

Введено имя Fэксп
для диапазона C6:C9
имя Fожид для диапазона D6:D9

формула для массива

{=СУММ((Fэксп-СУММ(Fэксп)/СУММ(Fожид)
*Fожид)^2*СУММ(Fожид)/СУММ(Fэксп)/Fожид)}

$\chi^2_{эмп} =$ 1,721

$\chi^2_{крит} =$ 7,815

=ХИ2.ОБР.ПХ(H5;(СЧЁТ(Fэксп)-1))

Поскольку $\chi^2_{эмп} < \chi^2_{крит}$ то нулевая гипотеза принимается.
Отношение количества особей рассмотренных фенотипических классов равно 9:3:3:1

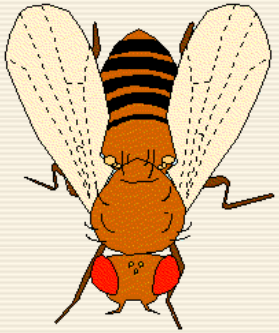


Рис. 7.13. Анализ отношения классов с использованием формул для массивов

7.3. Риски и шансы

> Относительный риск – это отношение частоты исходов среди исследуемых, на которых оказывал влияние изучаемый фактор, к частоте исходов среди исследуемых, не подвергавшихся влиянию этого фактора. Обычно используемое обозначение данного показателя – RR (relative risk).

В медицинской статистике и эпидемиологии RR – отношение риска наступления определенного события у лиц подвергшихся воздействию фактора риска, по отношению к контрольной группе.

$$RR = \frac{p_{\text{экспериментальной}}}{p_{\text{контрольной}}} = \frac{p_{\text{exposed}}}{p_{\text{non exposed}}}.$$

Типичный пример: исследование некого заболевания на базе 300 респондентов:

	заболевание есть	заболевания нет
женщины	f_{11} (40)	f_{12} (150)
мужчины	f_{21} (10)	f_{22} (100)

Обе переменные, входящие в таблицу, являются дихотомическими. Наличие заболевания (да-нет) – является переменной риска, а пол с двумя категориями (женщины-мужчины) — независимой (причинной) переменной.

Относительный риск используется для сравнения вероятности исхода в зависимости от наличия фактора риска. Например, при оценке влияния курения на частоту сердечных заболеваний.

Условия и ограничения применения относительного риска:

- показатели фактора и исхода должны быть измерены в номинальной шкале (например, курит - не курит, живой-мертвый, есть-нет и т.д.).
- метод позволяет проводить анализ только таблиц 2×2, когда и фактор, и исход являются дихотомическими переменными, то есть когда имеются только два возможных значения.

Относительный риск применяется при проспективных* исследованиях, когда исследуемые группы формируются по признаку наличия или отсутствия фактора риска (например, исследование типа до/после для зависимых выборок, когда группу составляют в настоящем и наблюдают в будущем).

	исход есть (1)	исхода нет (0)
фактор риска есть (1)	f_{11}	f_{12}
фактора риска нет (0)	f_{21}	f_{22}

Значение относительного риска определяется по следующей формуле:

$$RR = \frac{\frac{f_{11}}{f_{11}+f_{12}}}{\frac{f_{21}}{f_{21}+f_{22}}} = \frac{f_{11}}{f_{21}} \cdot \frac{(f_{21}+f_{22})}{(f_{11}+f_{12})}$$

Границы доверительного 95-процентного интервала (confidence interval) – верхняя $B_{\text{верх}}$ и нижняя $B_{\text{нижн}}$ определятся по следующим соотношениям.

*prospectivus: "открывающий вид"; prospicere: "глядеть вдаль, смотреть вперёд"

$$B_{\text{нижн}} = \exp\left[\ln(RR) - 1.96\sqrt{\frac{1}{f_{11}+f_{12}}\cdot\frac{f_{12}}{f_{11}} + \frac{1}{f_{21}+f_{22}}\cdot\frac{f_{22}}{f_{21}}}\right],$$

$$B_{\text{верх}} = \exp\left[\ln(RR) + 1.96\sqrt{\frac{1}{f_{11}+f_{12}}\cdot\frac{f_{12}}{f_{11}} + \frac{1}{f_{21}+f_{22}}\cdot\frac{f_{22}}{f_{21}}}\right].$$

Показателя относительного риска является значимым, когда $B_{\text{нижн}} \geq 1$ или $B_{\text{верх}} \leq 1$, то есть если единица находится внутри доверительного интервала $B_{\text{нижн}} \leq 1 \leq B_{\text{верх}}$, то статистическая значимость влияния фактора на частоту исхода отсутствует, а **п**оказатель относительного риска не значим.

Характер связи фактора и исхода в зависимости от показателя относительного риска определяется через сравнение с единицей:

- Если $RR = 1$, можно сделать вывод, что исследуемый фактор не влияет на вероятность исхода (отсутствие связи между фактором и исходом).
- При значениях $RR > 1$ делается вывод о том, что фактор повышает частоту исходов (прямая связь) – в экспериментальной группе событие развивается чаще чем в контрольной.
- При значениях $RR < 1$ – фактор снижает вероятность исхода при воздействии (обратная связь) – в экспериментальной группе событие развивается реже, чем в контрольной.

Ш а н с (chance) – это отношение вероятности того, что события произойдёт, к вероятности того, что событие не произойдёт.

Например, шанс выпадения тройки на игральной кости равен $1/5$ (один к пяти).

Отношение шансов – статистический показатель OR (odds ratio), один из основных способов описать в численном выражении то, насколько отсутствие или наличие определённого исхода связано с присутствием или отсутствием определённого фактора в конкретной статистической группе.

> Отношение шансов – характеристика, применяемая в математической статистике для количественного описания тесноты связи признака двух признаков в некоторой статистической популяции.

Отношение шансов – это дробь, в числителе которой стоят шансы некоторого события для одной группы, а в знаменателе – шансы того же события, но для другой группы. Данное выражение применяется также для расчета выборочных оценок отношения.

Часто рассматривается отношение шансов изучаемого события в основной группе к шансам в контрольной:

$$OR = \frac{\text{Ш}_{\text{экспериментальной}}}{\text{Ш}_{\text{контрольной}}} = \frac{S_{\text{exposed}}}{S_{\text{non exposed}}}.$$

Для представленного выше типового примера отношение шансов "заболеть" рассчитывается по следующей формуле:

$$OR = \frac{f_{11}}{f_{12}} : \frac{f_{21}}{f_{22}} = \frac{f_{11} \cdot f_{22}}{f_{12} \cdot f_{21}}.$$

Границы доверительного 95-процентного интервала (confidence interval) – верхняя $B_{\text{верх}}$ и нижняя $B_{\text{нижн}}$ определятся по соотношениям

$$B_{\text{нижн}} = \exp\left[\ln(OR) - 1.96\sqrt{\frac{1}{f_{11}}+\frac{1}{f_{12}}+\frac{1}{f_{21}}+\frac{1}{f_{22}}}\right],$$

$$B_{\text{верх}} = \exp\left[\ln(OR) + 1.96\sqrt{\frac{1}{f_{11}}+\frac{1}{f_{12}}+\frac{1}{f_{21}}+\frac{1}{f_{22}}}\right].$$

Показателя относительного риска является значимым, когда $B_{\text{нижн}} \geq 1$ или $B_{\text{верх}} \leq 1$,
то есть если единица находится внутри доверительного интервала $B_{\text{нижн}} \leq 1 \leq B_{\text{верх}}$, то статистическая значимость влияния фактора на частоту исхода отсутствует, а показатель отношение шансов не значим.

Как и в случае анализа рисков характер связи фактора и исхода в зависимости от показателя отношение шансов определяется через сравнение с единицей.

Пример 7.12 Было проведено исследование некого заболевания на базе 300 респондентов мужчин и женщин. из 110 мужчин подвержены заболеванию оказалось 10 человек; из 190 женщин болело 40. Рассчитать риски и шансы заболеваемости для женщин.

На рис. 7.14 дан скриншот вычислений, из которого видна значимость рассчитанных показателей. Вывод: риск заболеть у женщин в 2.67 раза выше, чем у мужчин. Отношение шансов заболеваемости в зависимости от пола (женщины / мужчины) определяется как 2.32.

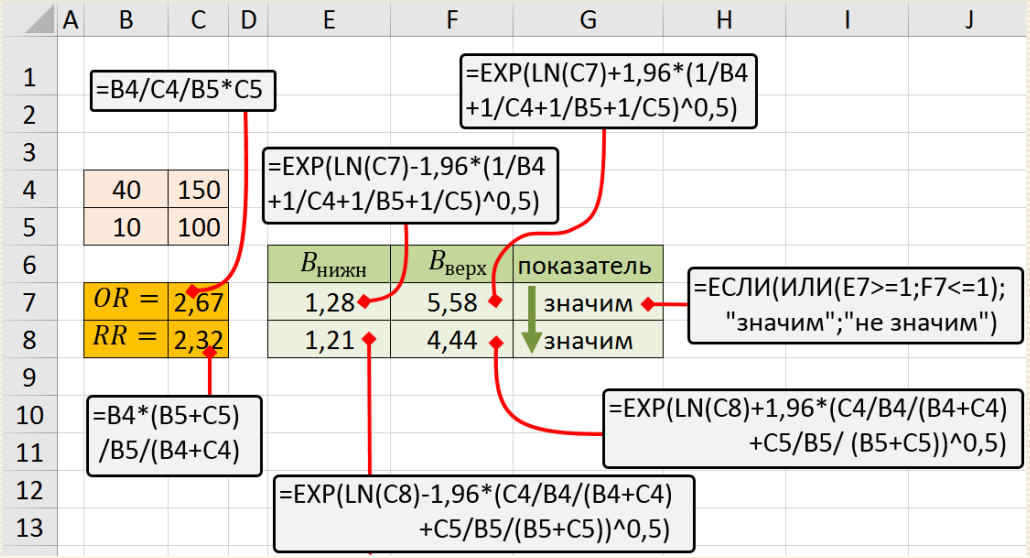


Рис. 7.14. Скриншот расчета рисков и отношения шансов

8. Исключение грубых погрешностей

> Выброс (outlier; синонимы maverick – резко выделяющийся результат, straggler – оторвавшийся результат) – резко отклоняющееся значение наблюдаемой величины.

Выбросом считается наблюдение, которое лежит аномально далеко от остальных из серии параллельных наблюдений. То есть выбросы – это значения количественного признака, располагающиеся на краях интервала допустимых значений.

Источниками выбросов (промахов) нередко бывают ошибки, допущенные исследователем при измерении. Наиболее характерными из них являются: неправильный отсчет по шкале измерительного устройства, неправильная запись результата наблюдения (описка), неправильная запись значений отдельных мер использованного набора и т.п., ошибки при действиях с приборами, если они повторяются при измерениях.

Причинами грубых погрешностей могут быть внезапные или кратковременные изменения условий измерения или незамеченные неисправности в аппаратуре.

Оценка наличия грубых погрешностей решается методами математической статистики – статистической проверкой гипотез. Суть метода сводится к следующему. Выдвигается нулевая гипотеза относительно результата измерения, который вызывает некоторое сомнение и рассматривается как грубый промах в связи с большим отклонением от других результатов измерения. При этом нулевая гипотеза заключается в утверждении, что "сомнительный" результат в действительности принадлежит к возможной совокупности полученных в данных условиях результатов измерений, и получение такого результата вероятно.

Пользуясь определенными статистическими критериями проверяется нулевая гипотеза; в случае ее подтверждения промах из исходных данных исключают, если нет – то результат эксперимента (измерения) оставляют. Выбор того или иного критерия основан на принципе практической уверенности. Для этого задаются достаточно малой вероятностью того, что сомнительный результат действительно мог бы иметь место. Вероятность определяется уровнем значимости $\alpha = 0.001 \div 0.1$.

Для выбранного критерия определяют критическую область значений проверки нулевой гипотезы. Если значение критерия попадает в эту область, то гипотеза отвергается.

Известен ряд критериев, которые позволяют исключить грубые промахи. К ним, в частности, можно отнести критерий Романовского, Шарлье, Райта и др. Эти критерии основаны на статистических оценках (по выборке) параметров распределения, поскольку в большинстве случаев действительные значения параметров распределения неизвестны.

8.1. Критерии Райта и правило "трех сигм"

Критерий "правило трех сигм" является одним из простейших для проверки результатов, подчиняющихся нормальному закону распределения. Одним из распространенных обозначений среднего квадратического является греческая буква (сигма). В данном подразделе (и только здесь) для указанного параметра будет использоваться σ. Сущность правила трех сигм:

> если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического (стандартного) отклонения.

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основания предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально. С этой целью для выборки (включая подозрительный результат) вычисляется центр распределения и оценка стандартного отклонения результата наблюдений.

Сомнительный результат $x_i^{\text{сом}}$, который удовлетворяет условию

$$|x_i^{\text{сом}} - \overline{x}| \geq 3\,\sigma\,,$$

считается имеющим грубую погрешность и удаляется, а ранее вычисленные характеристики распределения уточняются.

Этому критерию аналогичен критерий Райта, основанный на том, что если остаточная погрешность больше четырех сигм, то этот результат измерения является грубой погрешностью и должен быть исключен при дальнейшей обработке. Оба критерия надёжны при числе измерений $n > (20 \div 50)$. Их правомочно применять, когда известна величина генерального среднеквадратического отклонения σ.

Может оказаться, что при новых значениях $\overline{x}$ и σ другие результаты попадут в категорию аномальных. В работе* указано, что дважды использовать критерии грубой погрешности не рекомендуется.

*Никипорец Э.Н., Парамонова Л.А., Черновский Н.М. Сборник задач по взаимозаменяемости и метрологическому обеспечению в авиационной технике: Учебное пособие – М.: Изд-во МАИ, 1990. – 108 с.

Правило $|x_i^{\text{сом}} - \overline{x}| \geq 3\,\sigma$ обычно считается слишком жестким, поэтому рекомендуется назначать границу цензурирования в зависимости от объема выборки (критерий Райта):

$6 < n < 100$	$\left\|x_i^{\text{под}} - \overline{x}\right\| \geq 4\,\sigma$
$100 < n < 1000$	$\left\|x_i^{\text{под}} - \overline{x}\right\| \geq 4.5\,\sigma$
$1000 < n < 10000$	$\left\|x_i^{\text{под}} - \overline{x}\right\| \geq 5\,\sigma$

Данное правило также применимо только для нормального закона.

В общем случае границы цензурирования выборки зависят не только от объема n, но и от вида распределения. Назначая ту или иную границу, необходимо оценить уровень значимости α, то есть вероятность исключения какой-либо части отсчетов, принадлежащих обрабатываемой выборке.

8.2. Критерий Романовского

Критерий Романовского применяется, если число измерений объемом $n \leq 20$.
Значение критерия для сомнительного значения $x^{\text{сом}}$ вычисляется по соотношению

$$\beta = \frac{|x^{\text{сом}} - \overline{x}|}{S},$$

где $\overline{x}$ – среднее значение выборки (центр распределения) и S – среднее квадратическое отклонение рассчитываются без учета сомнительного значения $x_{\text{сом}}$.

Расчетную величину β сравнивают с критическим (табличным) $\beta_{крит}$ на требуемом (заданном) уровне значимости α. Если $\beta \geq \beta_{крит}$, то значение $x^{сом}$ считается промахом (грубой погрешностью) и отбрасывается.

По таблице справа критические значения критерия Романовского определяются через n б е з у ч е т а сомнительного значения.

Значения критерия Романовского

α	n = 4	n = 6	n = 8	n = 10	n = 12	n = 15	n = 20
0.01	1.73	2.16	2.43	2.62	2.75	2.90	3.08
0.02	1.72	2.13	2.37	2.54	2.66	2.80	2.96
0.05	1.71	2.10	2.27	2.41	2.52	2.64	2.78
0.10	1.69	2.00	2.17	2.29	2.39	2.49	2.62

Значения $\beta_{крит}$ выбирать из таблицы не всегда удобно; при автоматизированной обработке более эффективно использовать аппроксимации соответствующих данных, сведенные в следующей таблице.

α	соотношение	максимальная ошибка, %
0.01	$\beta_{крит} = 3.53\ e^{-2.9/n}$	1.18
0.02	$\beta_{крит} = 3.35\ e^{-2.7/n}$	1.12
0.05	$\beta_{крит} = 3.10\ e^{-2.4/n}$	1.18
0.10	$\beta_{крит} = 2.89\ e^{-2.2/n}$	1.34

Пример 8.1 Провести на уровне значимости 0.01 проверку на наличие грубых ошибок по отсортированным по убыванию данным

65.0	51.0	48.0	46.5	44.0	42.5	41.5	40.0	38.0	36.0	22.0

Поскольку грубой ошибкой может быть одно из крайних значений вариационного ряда, рассчитываем значения $\overline{x}, S$ и количество вариант для диапазона B4:B13, а в ячейке B3 будет находиться сомнительное значение $x^{\text{сом}}$, минимальное либо максимальное в выборке в зависимости от направления сортировки (рис. 8.1 и 8.2).

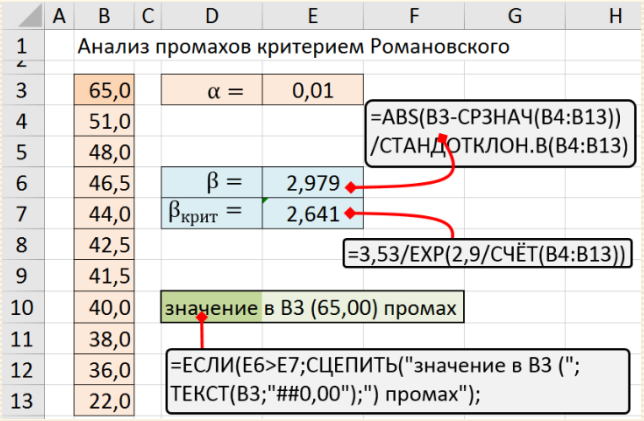


Рис. 8.1 Анализ на промах максимального значения в выборке

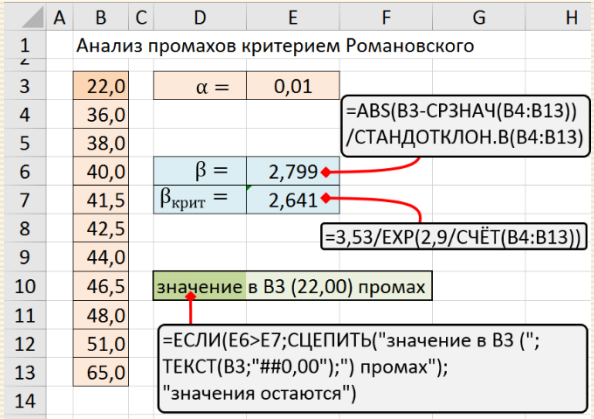


Рис. 8.2 Анализ на промах минимального значения в выборке

Так, на рис. 8.1 значения расположены по убыванию, и поэтому проверяется максимальное значение. Чтобы проверить минимальное значение следует расположить значения по возрастанию – пересортировать таблицу (рис. 8.2).

8.3. Критерий Шарлье

Критерий Шарлье используется, если число измерений велико ($n > 20$). Тогда по теореме Бернулли число результатов, превышающих по абсолютному значению среднее арифметическое значение на величину $K_S S$, будет $n[1 - \Phi(K_S)]$, где $\Phi(K_S)$ — значение нормированной функции Лапласа для аргумента $x = K_S$.

В случае если значение $(x_i - \overline{x})$ превосходит по модулю значение $K_S S$, то данный результат (элемент выборки) отбрасывается. Значение K_S в зависимости от числа измерений n приводятся в следующей таблице

Критические значения критерия Шарлье

n	5	10	20	30	40	50	100
K_S	1.29	1.65	1.96	2.13	2.24	2.32	2.58

Приведенные данные можно определить через Excel-функцию, возвращающей обратное значение стандартного нормального распределения (распределения с нулевым средним и единичным стандартным отклонением) от аргумента $(n - 0.5)/n$

$$=\text{НОРМ.СТ.ОБР}\left(\frac{n-0.5}{n}\right).$$

Пользуясь данным критерием, отбрасывается результат, для которого выполняется неравенство

$$|x_i - \overline{x}| > K_S\, S\,. \qquad (*)$$

Пример 8.2 При измерении размеров листа акации были получены следующие результаты

11.8	11.5	16.8	16.1	16.8	16.8	16.3	16.3
12.6	12.0	12.4	5.6	18.9	18.5	18.6	16.6
13.1	13.9	13.0	17.7	17.7	17.1	17.1	16.7
14.8	14.4	14.3	14.6	14.5	14.4	14.4	16.3
15.1	15.3	15.5	15.5	15.4	15.4	15.4	16.2
19.8	19.1	19.8	17.2	17.4	17.7	20.8	20.6

Необходимо проанализировать данные и выполнить отсев грубых погрешностей данных.

Для решения выполняется перенос исходных данных (область B3:I8 рис. 8.3) с исключением (отбросом) сомнительных элементов на область "очищенных данных" (область B11:I16). "Плохие" значения идентифицируется и отсеиваются неравенством (∗). Для этого

- в ячейку B11 вносится формула =ЕСЛИ(ABS(B3-$L$4)>$L$5;"";B3), которая заносит в текущую ячейку значение из таблицы исходных данных, если выполняется соотношение $|x_i - \overline{x}| \leq K_S\, S$, либо "пусто" в ином случае;

- формула ячейки B11 тиражируется автозаполнением на область выходных данных B11:I16 (рис. 8.3).

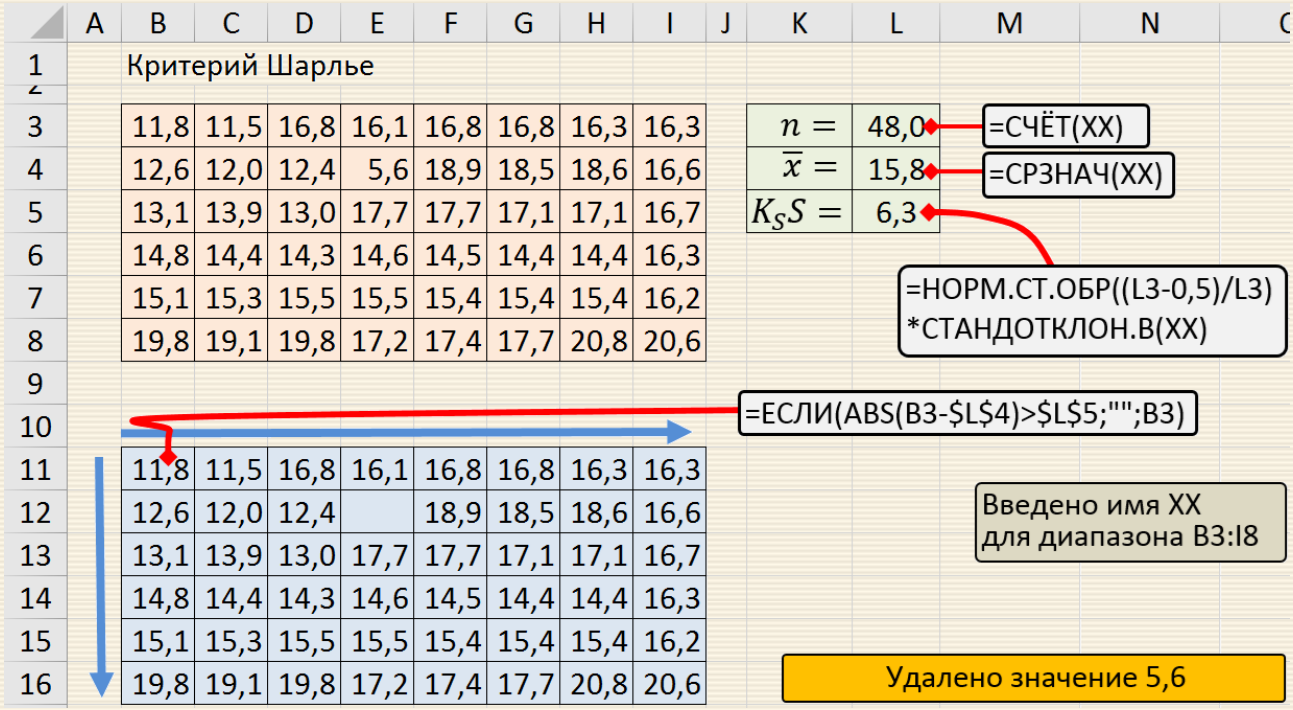


Рис. 8.3. Скриншот реализации вычислений для критерия Шарлье

8.4. Правило "ящик с усами"

Правило "ящик с усами" (или "коробчатая диаграмма" (box-and-whiskers plot)) получило название от типа соответствующего графика, используемого для наглядного представления разброса эмпирических данных с нанесенными значениями медианы и квартилей, как это показано на рис. 8.4.

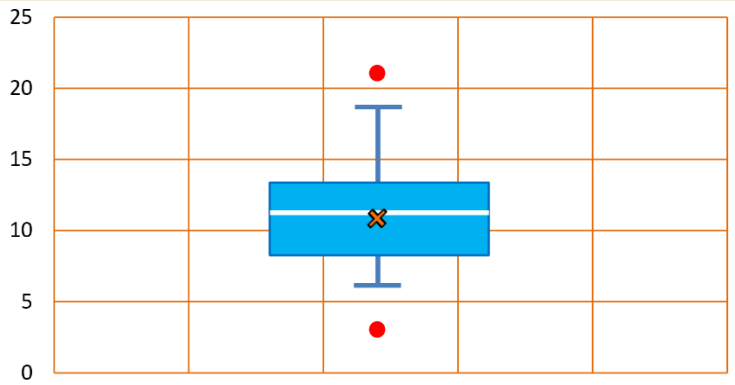


Рис. 8.4. Типовой вид диаграммы "ящик с усами"

Крестик посередине – это среднее арифметическое по выборке.

Линия чуть выше или ниже крестика – медиана.

Нижняя и верхняя грань прямоугольника (типа ящика) соответствует первому и третьему квартилю (значениям, отделяющим ¼ и ¾ выборки). Расстояние между 1-м и 3-м квартилем – это межквартильный размах (или расстояние).

Горизонтальные черточки на конце "усов" – максимальное и минимальное значение (выбросы – красные точки – игнорируются).

Порядок вычислений при анализе выбросов по правилам метода "ящик с усами" следующий:

а) Определяются выборочные значения межквартильного размаха R_μ и медианы μ.

б) Выборочные значения, меньшие (μ-1.5R_μ) и большие (μ+1.5R_μ), называются м я г к и м и (подозрительными) выбросами.

в) Выборочные значения, меньшие (μ-3R_μ) и большие (μ+3R_μ), называются э к с т р е м а л ь н ы м и выбросами и должны быть исключены.

Вычисление межквартильного размаха R_μ упорядоченной по возрастанию количественной выборки производится по формуле:

$$R_\mu = \mu_{3/4} - \mu_{1/4},$$

где $\mu_{1/4}$, $\mu_{3/4}$ – значение нижней и верхней квартилей выборки.

Критерий удобен для автоматической идентификации любого числа экстремально малых и больших значений выборки. Критерий очень популярен, однако его рекомендуется применять только в случае, если численность выборки достаточно велика.

Пример 8.3 Выполнить отсев грубых погрешностей для данных примера 8.2 на базе критерия "ящик с усами".

Выделение цветом ячеек на листе решения (рис. 8.4) выполнено с помощью настроек инструментария "Условное форматирование" (рис. 8.6-8.9).

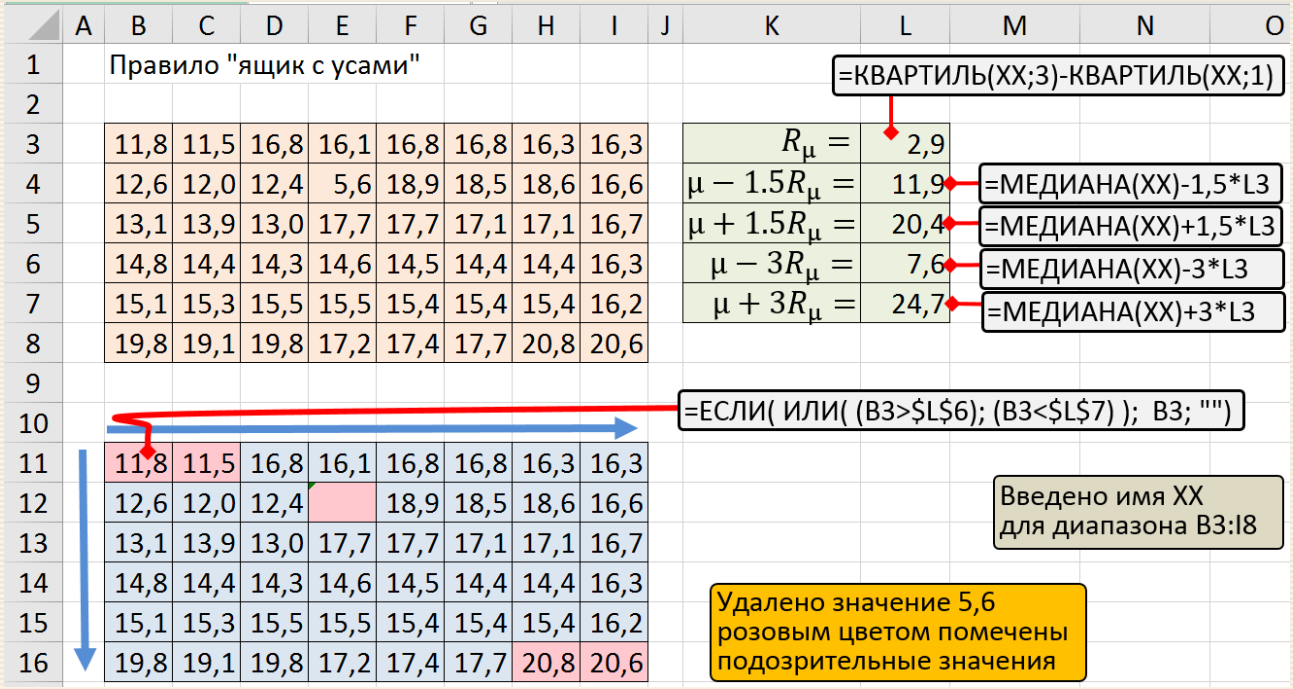


Рис. 8.5. Скриншот реализации вычислений для правила "ящик с усами"

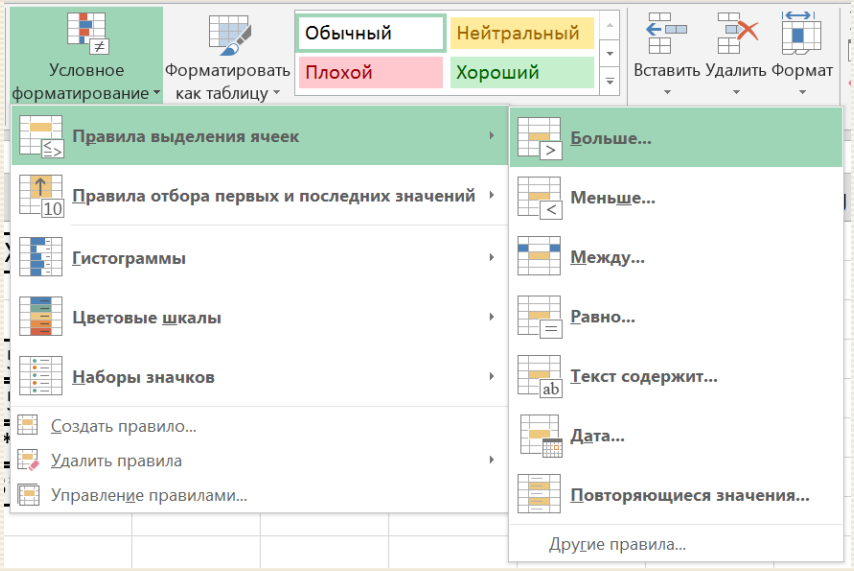


Рис. 8.6. Инструментарий "Условное форматирование"

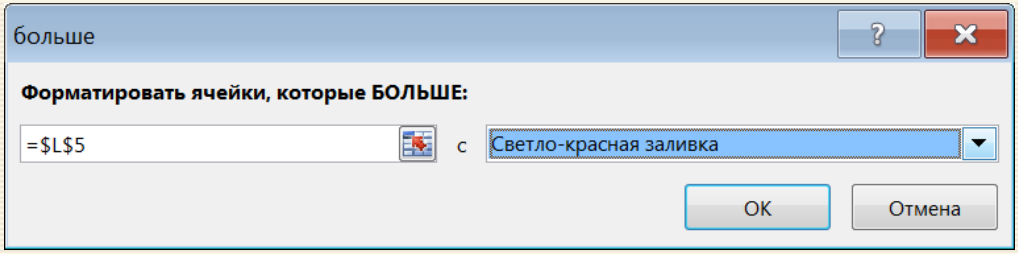


Рис. 8.7. Выбор цвета ячеек по условию

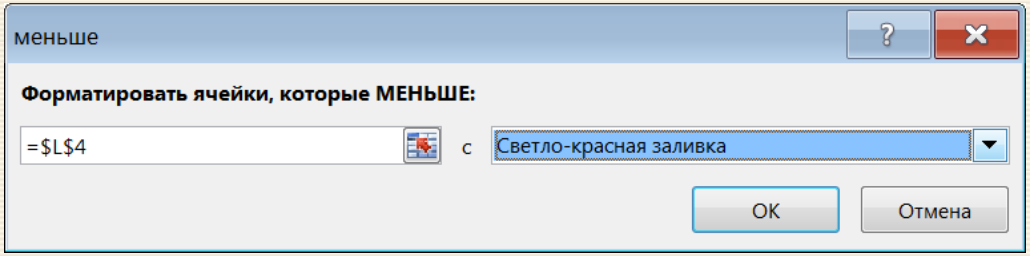


Рис. 8.8. Выбор цвета ячеек по условию

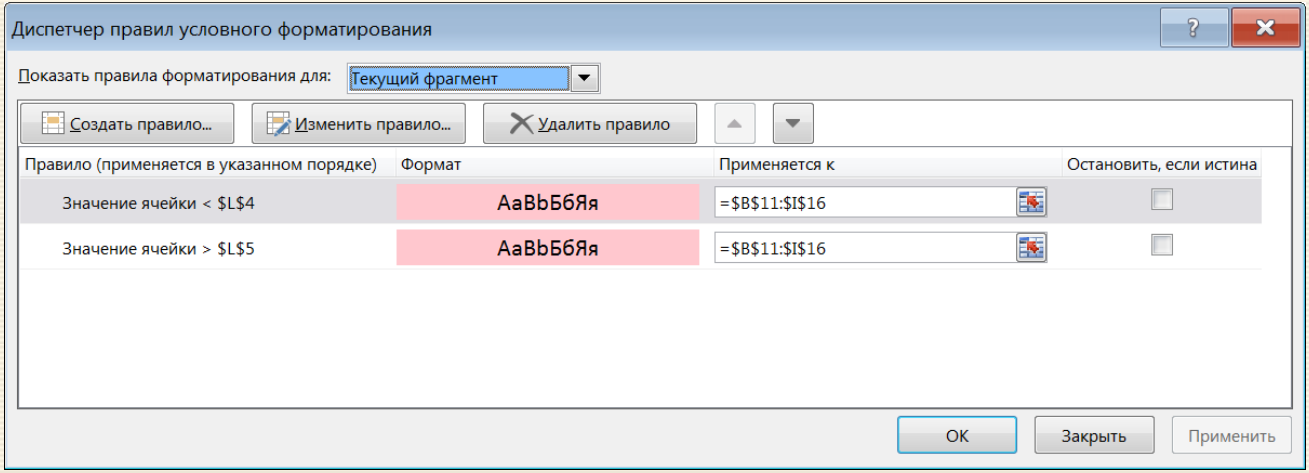


Рис. 8.9. Панель диспетчера условного форматирования

В MS Excel 2016 в состав типовых диаграмм включен "Ящик с усами" (рис. 8.10). Для данных примера 8.3 на рис. 8.11 приведен вид данной диаграммы, где на оси абсцисс указано количество выбросов и при наведении курсора на графическое изображение точки отображается ее значение.

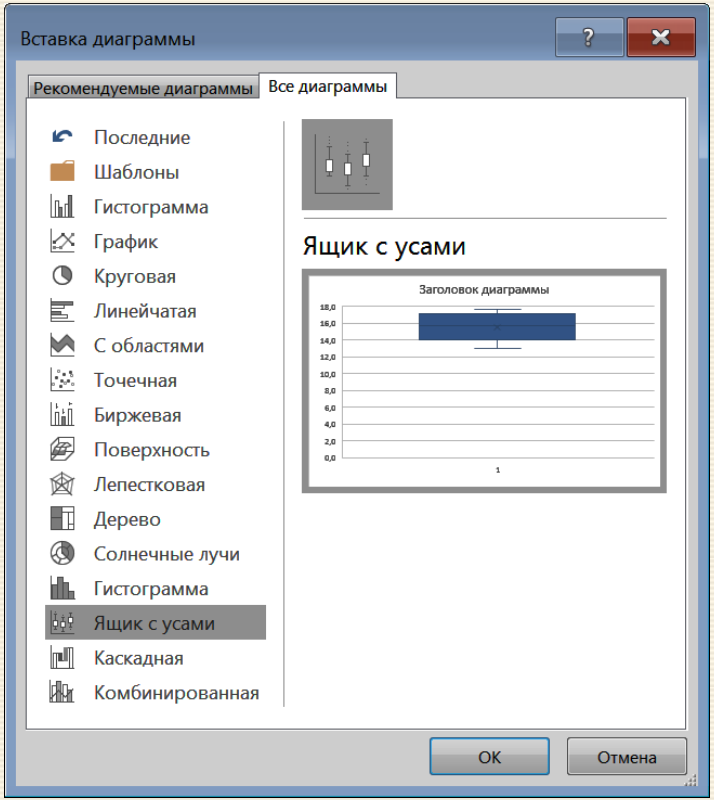


Рис. 8.10. Панель диспетчера условного форматирования

Рис. 8.11. Диаграмма "ящик с усами"

8.5. Правило Томпсона (критерий Рошера)

В правиле Томпсона (критерии Рошера) для исключения выбросов используется статистика

$$T = \frac{|x^{\text{сом}} - \overline{x}|}{S},$$

где x_i – результаты наблюдений,
$\overline{x}$ – среднее значение,
S –стандартное отклонение, корень квадратный из дисперсии.

Величина статистики критерия сравнивается со значением для двусторонней критической области

$$T_{\text{крит}} = \sqrt{\frac{(n-1)\, t^2_{\alpha,(n-2)}}{n-2+t^2_{\alpha,(n-2)}}}$$

где $t_{\alpha,(n-2)}$ – значение обратной функции t-распределения с параметрами α и $(n\text{-}2)$,
α – заданный уровень значимости.

При величине статистики, большей критического значения, измерение (наблюдение) исключается.

Критерий является неполным, поскольку заранее предполагается исключение только одного аномального значения. В работе* Пирсона указано, что критерий Томпсона обладает маскирующим эффектом, который заключается в том, что при наличии в выборке более одного аномального значения (выброса) критерий их не обнаруживает. В наибольшей степени это касается для малых (n<30) выборок.

Пример 8.4 Условия задачи те же, что и в предыдущих примерах. Расчетную величину T сравнивают с критическим (табличным) $T_{\text{крит}}$ на требуемом (заданном) уровне значимости α. Если $T \geq T_{\text{крит}}$, (либо, что тоже самое, $|x^{\text{сом}} - \overline{x}| \geq S \cdot T_{\text{крит}}$), то значение $x^{\text{сом}}$ считается промахом (грубой погрешностью) и отбрасывается (рис. 8.12).

*Pearson E.S., Chandra S.C. Biometrika // Annals of mathematical statistics, 1936. P. 30-49.

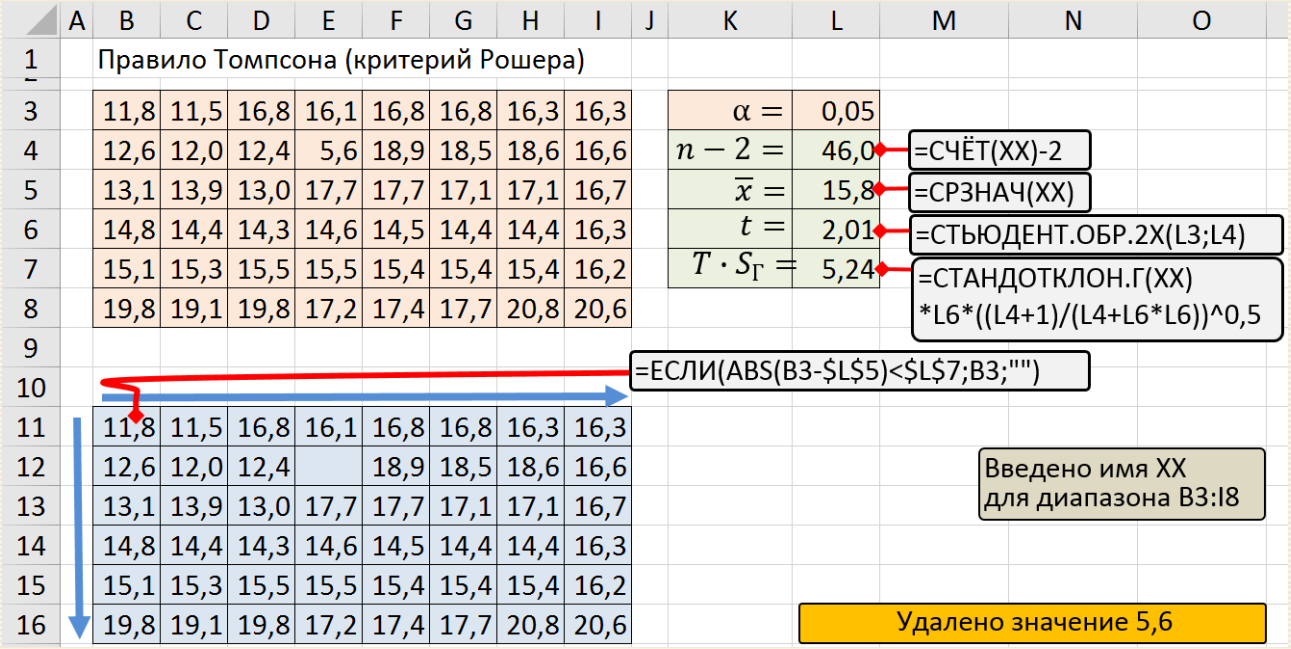


Рис. 8.12. Скриншот анализа данных по правилу Томпсона (критерию Рошера)

8.6. Критерий Диксона (Q-критерий)

При нормальном распределении контролируемого параметра для исключения грубых погрешностей (грубых ошибок) распространен критерий Диксона (другое название Q-критерий). В зависимости от объёма выборки критерий (или коэффициент) Диксона для одностороннего выброса (набольшего или наименьшего) обозначают так, как представлено в таблице слева.

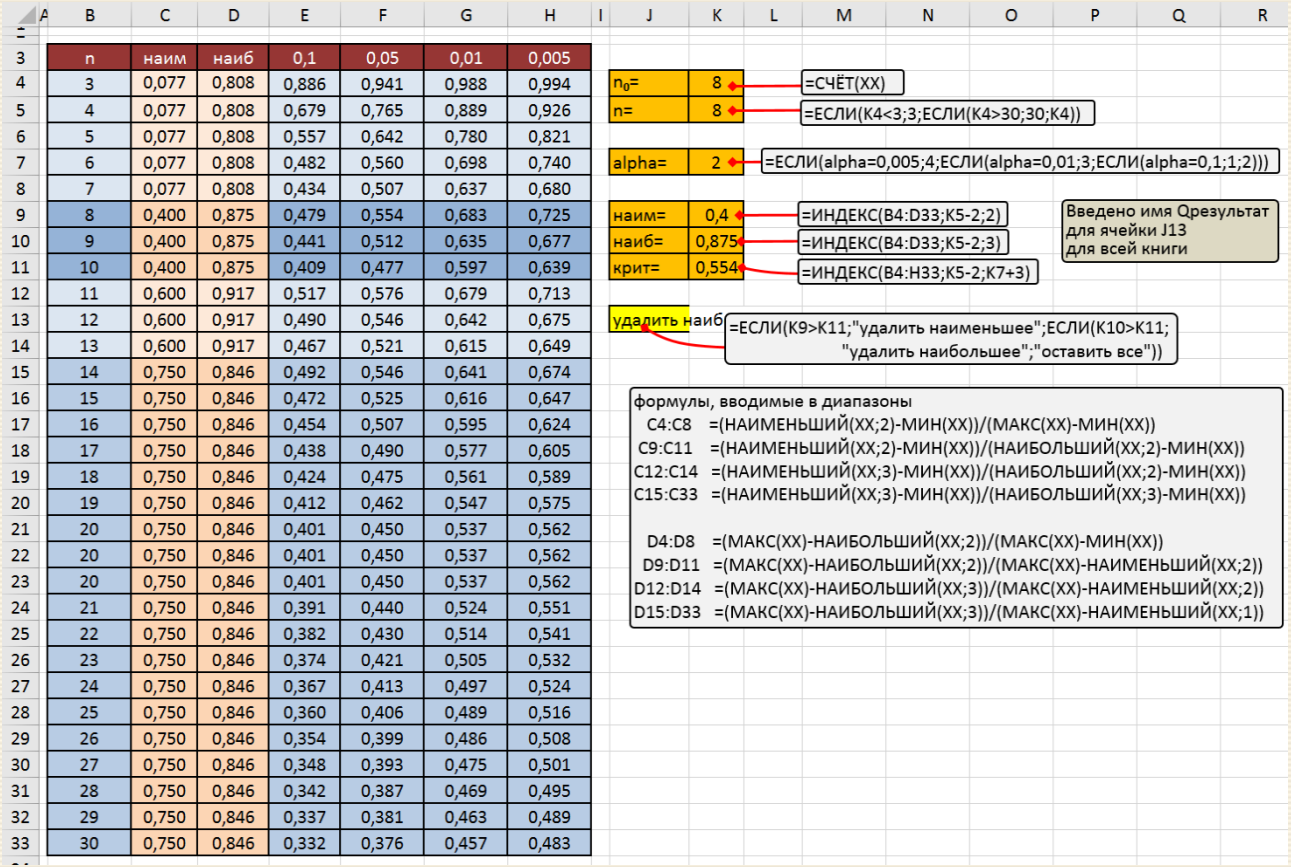
n	критерий Диксона одностороннего выброса	
	наименьшего	наибольшего
3…7	$Q_{10} = \frac{x_2 - x_1}{x_n - x_1}$	$Q_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$
8…10	$Q_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$	$Q_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$
11…13	$Q_{21} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	$Q_{21} = \frac{x_n - x_{n-2}}{x_n - x_2}$
14…30	$Q_{22} = \frac{x_3 - x_1}{x_{n-2} - x_1}$	$Q_{22} = \frac{x_n - x_{n-2}}{x_n - x_3}$

Здесь $x_1, x_1, \dots, x_n$ – результаты испытаний в вариационном ряду.

Рассчитанный критерий Диксона $Q_{\text{эмп}}$ сравнивают с его табличным значением $Q_{\text{крит}}$, приведённым в таблице* рис. 8.13 (данные, подкрашенные голубым цветом).

Сомнительное значение считают грубой ошибкой и отбрасывают, если $Q_{\text{эмп}} > Q_{\text{крит}}$.

*Данные сайта http://arhiuch.ru/lab5.html



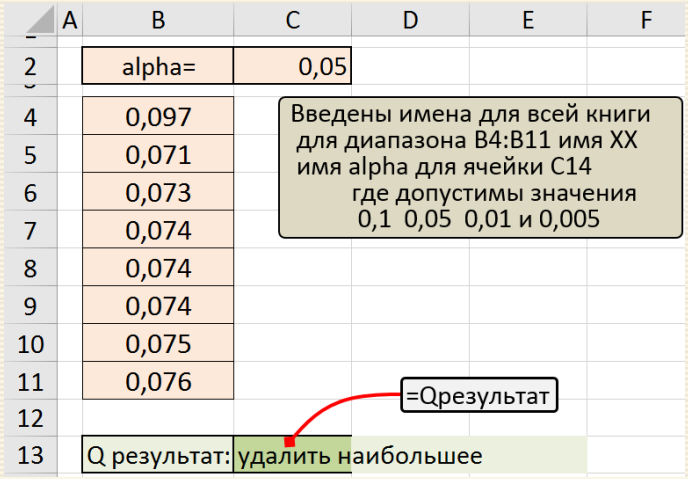
n	наим	наиб	0,1	0,05	0,01	0,005
3	0,077	0,808	0,886	0,941	0,988	0,994
4	0,077	0,808	0,679	0,765	0,889	0,926
5	0,077	0,808	0,557	0,642	0,780	0,821
6	0,077	0,808	0,482	0,560	0,698	0,740
7	0,077	0,808	0,434	0,507	0,637	0,680
8	0,400	0,875	0,479	0,554	0,683	0,725
9	0,400	0,875	0,441	0,512	0,635	0,677
10	0,400	0,875	0,409	0,477	0,597	0,639
11	0,600	0,917	0,517	0,576	0,679	0,713
12	0,600	0,917	0,490	0,546	0,642	0,675
13	0,600	0,917	0,467	0,521	0,615	0,649
14	0,750	0,846	0,492	0,546	0,641	0,674
15	0,750	0,846	0,472	0,525	0,616	0,647
16	0,750	0,846	0,454	0,507	0,595	0,624
17	0,750	0,846	0,438	0,490	0,577	0,605
18	0,750	0,846	0,424	0,475	0,561	0,589
19	0,750	0,846	0,412	0,462	0,547	0,575
20	0,750	0,846	0,401	0,450	0,537	0,562
20	0,750	0,846	0,401	0,450	0,537	0,562
20	0,750	0,846	0,401	0,450	0,537	0,562
21	0,750	0,846	0,391	0,440	0,524	0,551
22	0,750	0,846	0,382	0,430	0,514	0,541
23	0,750	0,846	0,374	0,421	0,505	0,532
24	0,750	0,846	0,367	0,413	0,497	0,524
25	0,750	0,846	0,360	0,406	0,489	0,516
26	0,750	0,846	0,354	0,399	0,486	0,508
27	0,750	0,846	0,348	0,393	0,475	0,501
28	0,750	0,846	0,342	0,387	0,469	0,495
29	0,750	0,846	0,337	0,381	0,463	0,489
30	0,750	0,846	0,332	0,376	0,457	0,483

Рис. 8.13. Скриншот "обслуживающего" листа

Если первичная обработка данных с помощью Q-критерия повторяется достаточно часто, то целесообразно создать в книге электронных таблиц "обслуживающего" листа (рис. 8.13), который через общекнижное имя Qрезультат возвращает результат анализа критерия Диксона.

Алгоритм анализа одностороннего выброса с помощью "обслуживающего" листа достаточно прост:

1) на отдельном (рабочем) листе исследуемых данных им присваивается общекнижное имя XX;
2) ячейке (в данном случае C14 рис. 8.14) устанавливается уровень значимости; ячейке присваивается общекнижное имя alpha;
3) ячейке C16 присваивается значение Qрезультат, где и высветится заключение по анализу наличия выброса исходных данных.



Пример 8.5 Для выборки (данные в таблице слева) определить наличие односторонних выбросов.

На рис. 8.14 дан скриншот вычислительного анализа задачи с использованием "обслуживающего" листа.

Рис. 8.14. Скриншот анализа одностороннего выброса

9. Основы планирования эксперимента

Сомнение вызывает не сама возможность планирования...
но возможность успешного планирования.
Фридрих фон Хайек

Планирование эксперимента – это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Целью планирования эксперимента является стремление к минимизации общего числа опытов посредством одновременного варьирования всеми переменными, определяющими процесс, по специальным правилам (алгоритмам) на базе математического аппарата, формализующего действия экспериментатора использованием стратегии, позволяющей принимать обоснованные решения после каждой серии экспериментов.

Поиск оптимальных условий, построение интерполяционных формул, выбор существенных факторов и так далее относится к задачам планирования эксперимента.

> Оптимальный режим – набор значений факторов, позволяющий получить наиболее выгодное значение параметра оптимизации (функции отклика).

Необходимо отметить, что большое количество исследовательских задач формулируется как экстремальные: определение оптимальных условий процесса, оптимального состава композиции и т.д.

Благодаря оптимальному расположению точек в факторном пространстве и линейному преобразованию координат удается преодолеть недостатки классического регрессионного анализа, в частности корреляцию между коэффициентами уравнения регрессии.

Выбор плана эксперимента определяется постановкой задачи исследования и особенностями объекта. Процесс исследования обычно разбивается на отдельные этапы, определяющие стратегию эксперимента и возможность его оптимального управления. Планирование эксперимента позволяет варьировать одновременно все факторы и получать количественные оценки основных и взаимодействующих эффектов. Исследуемые эффекты определяются с меньшей ошибкой, чем при традиционных методах исследования. В конечном счете применение методов планирования значительно повышает эффективность эксперимента.

Фактором называется измеряемая переменная величина, принимающая в некоторый момент времени определенное значение и влияющая на объект исследования. Факторы имеют область определения, внутри которой задаются его конкретные значения. Область определения может быть непрерывной или дискретной. Обычно при планировании эксперимента значения факторов принимаются дискретными – выбираются некоторые конкретные значения – уровни факторов. По своей природе факторы разделяются на количественные и качественные.

К количественным относятся те факторы, которые можно измерять; качественные факторы – это технологические способы, приборы и пр. Хотя к качественным факторам не соответствует числовая шкала, но при планировании эксперимента для них вводят условную порядковую шкалу в соответствии с уровнями, т.е. производится их кодирование; порядок уровней фиксируется.

Фактор считается заданным, если указаны его название и область определения. В выбранной области определения он может иметь несколько значений, которые соответствуют числу его различных состояний. Выбранные для эксперимента количественные или качественные состояния фактора носят название у р о в н е й ф а к т о р а .

При выборе области определения факторов важен выбор нулевого (основного) уровня: исходного состояния объекта исследования. Оптимизация связана с улучшением состояния объекта по сравнению с состоянием в нулевой точке (исходными, рабочими значениями параметров).

> У р о в е н ь ф а к т о р о в – значения факторов, которыми задаются при изучении их влияния на параметр оптимизации.
>
> Ф а к т о р н о е п р о с т р а н с т в о – пространство, координаты которого соответствуют рассматриваемым факторам.
>
> Ф а к т о р ы – независимые переменные, варьируемые экспериментатором при изучении объекта исследования.

Предположим, в некоторой задаче фактор (температура) может изменяться от 100 до 200 °С. Естественно, что за нулевой уровень можно принять среднее значение фактора, соответствующее 150 °С.

После установления нулевой точки выбирают интервалы варьирования факторов, которые в кодированных величинах соответствуют +1 и –1. Интервалы варьирования выбирают с учетом того, что значения факторов, соответствующие уровням +1 и –1, должны отличаться от значения нулевого уровня. Поэтому во всех случаях величина интервала варьирования должна быть больше удвоенной квадратичной ошибки фиксирования данного фактора.

Чрезмерное увеличение величины интервалов варьирования нежелательно, так как это может привести в дальнейших исследованиях к снижению эффективности поиска оптимума; малый интервал варьирования уменьшает область эксперимента и замедляет поиск оптимума. Минимальное число уровней, обычно применяемое на первой стадии работы, равно двум (верхний и нижний уровни), обозначаемые в кодированных координатах через значения +1 и –1.

> Интервал варьирования – разность между двумя натуральными значениями фактора, соответствующая единице кодированного значения фактора.

При решении задачи планирования используется математическая модель исследования, под которой понимается уравнение, связывающее параметр оптимизации с факторами z_i:

$$y = \varphi(z_1, z_2, \dots, z_k)$$

где выражение $\varphi(z_1, z_2, \dots, z_k)$, как это принято в математике, трактуется как "функция от…" и называется функцией отклика.

> Функция отклика (параметр оптимизации) – величина, характеризующая результаты эксперимента.

> Уравнение регрессии – математическая модель процесса, полученная посредством математикостатистической обработки экспериментальных данных и представленная в полиномиальной форме.

При планировании эксперимента достаточно простим видом функции отклика является алгебраический полином первой степени:

$$y = c_0 + c_1 z_1 + c_2 z_2 + \cdots + c_{12} z_1 z_2 . + c_{13} z_1 z_3 + \cdots + c_{123} z_1 z_2 z_3 \ldots ,$$

где c_0, c_1, c_2 – линейные коэффициенты, c_{12}, c_{13}, c_{23} – коэффициенты парного взаимодействия, c_{123} – коэффициенты тройного взаимодействия.

Для анализа функции отклика удобно использовать кодированное представление факторов

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_{12} x_1 x_2 . + b_{13} x_1 x_3 + \cdots + b_{123} x_1 x_2 x_3 \ldots ,$$

где коэффициенты b имеют прежний смысл.

Анализ математической модели при планировании эксперимента нужен для определения численных значений коэффициентов в целях сокращения их количества – чем больше коэффициентов, тем больше нужно опытов. Очевидно, что полиномы первой степени имеют наименьшее число коэффициентов и позволяют предсказывать направление наискорейшего улучшения параметра оптимизации.

При планировании по схеме полного факторного эксперимента (ПФЭ) реализуются все возможные комбинации факторов на всех выбранных для исследования уровнях. При двух значениях факторов необходимое количество опытов N при ПФЭ определяется по формуле

$$N = 2^n,$$

где N – количество опытов; n – количество исследуемых факторов.

При двух значениях уровни факторов представляют собой границы исследуемой области по данному технологическому параметру. Например, изучается влияние на выход продукта y трех факторов: температуры (z_1) в диапазоне 100–200 °С, давления (z_2) (2-6)·10^5 Па и времени пребывания (z_3) 20 мин.

Верхний уровень по температуре $z_1^{\max}$ равен 200 °С, нижний $z_1^{\min}$ равен 100°С. Тогда для z_1 имеем

$$z_1^0 = \frac{z_1^{\max} + z_1^{\min}}{2} = 150\ ^\circ\text{C}\,;\ \ \Delta z_1 = \frac{z_1^{\max} - z_1^{\min}}{2} = 50\ ^\circ\text{C}\,.$$

Вообще для любого фактора z_j

$$z_j^0 = \frac{z_j^{\max} + z_j^{\min}}{2}\ ;\ \ \Delta z_j = \frac{z_j^{\max} - z_j^{\min}}{2}.$$

Точка с координатами $(z_1^0, z_2^0, \ldots)$ называется центром плана (основным уровнем); Δz_j – интервал варьирования по фактору z_j. От естественных (натуральных) переменных $z_1, z_2, \ldots$ линейным преобразованием можно перейти к новым (кодированным) $x_1, x_2, \ldots$

$$x_j = \frac{z_j - z_j^0}{\Delta z_j}. \qquad (9.1)$$

Для переменных $x_1, x_2, \ldots$ верхний уровень равен +1, нижний уровень -1, координаты центра плана равны нулю и совпадают с началом координат. Число возможных комбинаций N для трех факторов на двух уровнях равно $N = 2^3 = 8$.

Кодирование фактора – линейное преобразование факторного пространства с переносом начала координат в центр эксперимента и выбором масштаба по осям координат в единицах варьирования факторов.

Кодированная система координат – безразмерная система координат, в которой на осях откладывают кодированные значения факторов.

Центр плана (эксперимента) – условия проведения эксперимента, при которых значения всех факторов соответствуют серединам интервалов, определяемых граничными значениями области вариации каждого фактора.

Условия эксперимента удобно записывать в виде табл. 9.1, которую называют матрицей планирования эксперимента.

Таблица 9.1. – Матрица планирования эксперимента 2^2

номер опыта	x_1	x_2	y
1	+1	+1	y_1
2	-1	+1	y_2
3	+1	-1	y_3
4	-1	-1	y_4

Таким образом, для двух факторов построение матрицы планирования элементарно. Для большего числа факторов необходимо использовать какие-либо правила построения таких матриц. Например, при появлении фактора x_3 в предыдущей таблице произойдут следующие изменения: при появлении нового столбца каждая комбинация уровней исходной таблицы проявится дважды (табл. 9.2).

Таблица 9.2. – Матрица планирования эксперимента 2^3

номер опыта	x_1	x_2	x_3	y
1	+1	+1	+1	y_1
2	-1	+1	+1	y_2
3	+1	-1	+1	y_3
4	-1	-1	+1	y_4
5	+1	+1	-1	y_5
6	-1	+1	-1	y_6
7	+1	-1	-1	y_7
8	-1	-1	-1	y_8

Для расширения матрицы планирования данный способ не единственный – также можно использовать перемножение столбцов, правило чередования знаков и другие алгоритмы.

Для составления планов-таблиц регулярных дробных реплик часто используют так называемое правило двоичного кода. Оно гласит, что для модели в виде гиперболоида знаки "+" и "–" в столбцах плана должны чередоваться по степеням двойки: в столбце x_1 – через 1 (т.е. 2^0), в столбце x_2 – через 2 (т.е. 2^1), в столбце x_3 – через 4 (т.е. 2^2), в столбце x_n – через 2^{n-1} .

В матрице планирования эксперимента в значениях уровней факторов цифру 1 часто опускают; с учетом взаимодействия факторов x_1 и x_2 табл. 9.1 можно переписать в виде табл. 9.3.

номер опыта	x_1	x_2	x_1x_2	y
1	+	+	+	y_1
2	–	+	–	y_2
3	+	–	–	y_3
4	–	–	+	y_4

Таблица 9.3. – Матрица планирования эксперимента

План проведения экспериментов (матрица планирования 2^3) записывается в виде нижеследующей таблицы 9.4. Представленный в таблице план в безразмерном масштабе геометрически может быть пространственно интерпретирован в виде восьми вершин куба, где в качестве начала координат служит столбец с так называемой фиктивной переменной $x_0 = 1$.

Таблица 9.4. – Расширенная матрица планирования полного факторного эксперимента 2^3

Номер опыта	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	+1	-1	-1	-1	+1	+1	+1	-1	y_1
2	+1	+1	-1	-1	-1	-1	+1	+1	y_2
3	+1	-1	+1	-1	-1	+1	-1	+1	y_3
4	+1	+1	+1	-1	+1	-1	-1	-1	y_4
5	+1	-1	-1	+1	+1	-1	-1	+1	y_5
6	+1	+1	-1	+1	-1	+1	-1	-1	y_6
7	+1	-1	+1	+1	-1	-1	+1	-1	y_7
8	+1	+1	+1	+1	+1	+1	+1	+1	y_8

Эффекты взаимодействия определяются аналогично линейным эффектам.

Пользуясь планом, представленным в табл. 9.4, можно записать полное уравнение регрессии с линейными коэффициентами и коэффициентами взаимодействия

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2. + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3\ . \qquad (9.2)$$

Значение свободного члена (b_0) берут как среднее арифметическое всех значений параметра (допустим, оптимизации) в матрице:

$$b_0 = \frac{1}{N}\sum_{i=1}^{N} y_i\ . \qquad (9.3a)$$

где y_i – значения параметра оптимизации в i-м опыте; N – число серий опытов в матрице.

Коэффициенты уравнения регрессии b_j определяются скалярным произведением столбца y на соответствующий столбец x_j, деленным на число опытов в матрице планирования N:

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}y_i\ . \qquad (9.3b)$$

где x_{ji} – кодированное значение фактора x_j в i-м опыте.

Коэффициенты регрессии, характеризующие парное (и тройное) взаимодействие факторов, находят по формулам

$$b_{ij} = \frac{1}{N}\sum_{k=1}^{N} x_{ik}x_{jk}y_k\ , \quad i \neq j, \quad i,j = 1,2,\dots,N\ ,$$
$$b_{ijk} = \frac{1}{N}\sum_{q=1}^{N} x_{iq}x_{jq}x_{kq}y_q\ , \quad i \neq j,\ \ i \neq k,\ \ j \neq k, \quad i,j,k = 1,2,\dots,N\ . \qquad (9.3c)$$

Можно показать, что коэффициенты уравнения регрессии не коррелированы между собой. Значимость коэффициентов уравнения регрессии можно (и должно) проверять для каждого коэффициента в отдельности по, например, критерию Стьюдента. Исключение из уравнения регрессии незначимого коэффициента не скажется на остальных. При этом выборочные коэффициенты b_j, оказываются так называемыми несмешанными оценками для соответствующих теоретических коэффициентов и характеризуют вклад соответствующего фактора в величину y. Все коэффициенты b_j уравнений для $\hat{y}$ определяются с одинаковой точностью.

Пример 9.1 Построить план дробного факторного эксперимента 2^{4-1} и определить коэффициенты уравнения регрессии для *Y*={10; 8; 8; 7; 9; 8; 8; 6.5}

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2. + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 .$$

Число факторов равно 4; нужно найти 8 коэффициентов полинома. Выбираются 8 из 16 опытов плана ПФЭ 2^4 таким образом, чтобы были определены независимые коэффициенты при самих факторах, смешанные коэффициенты при парных сочетаниях факторов; тройными и четверным сочетаниями пренебрегается.

Из табл. 9.4 берутся столбцы x_1, x_2, x_3 и достраивается столбец $x_4 = x_1x_2x_3$. Столбцы x_{ij} строятся перемножением соответствующих значений x_ix_j (рис. 9.1).

По соотношениям (9.3) рассчитываются коэффициенты уравнения регрессии, в столбце $\hat{y}$ вычисляются значения полинома. Можно отметить, что в данном случае точность аппроксимации достаточно высокая.

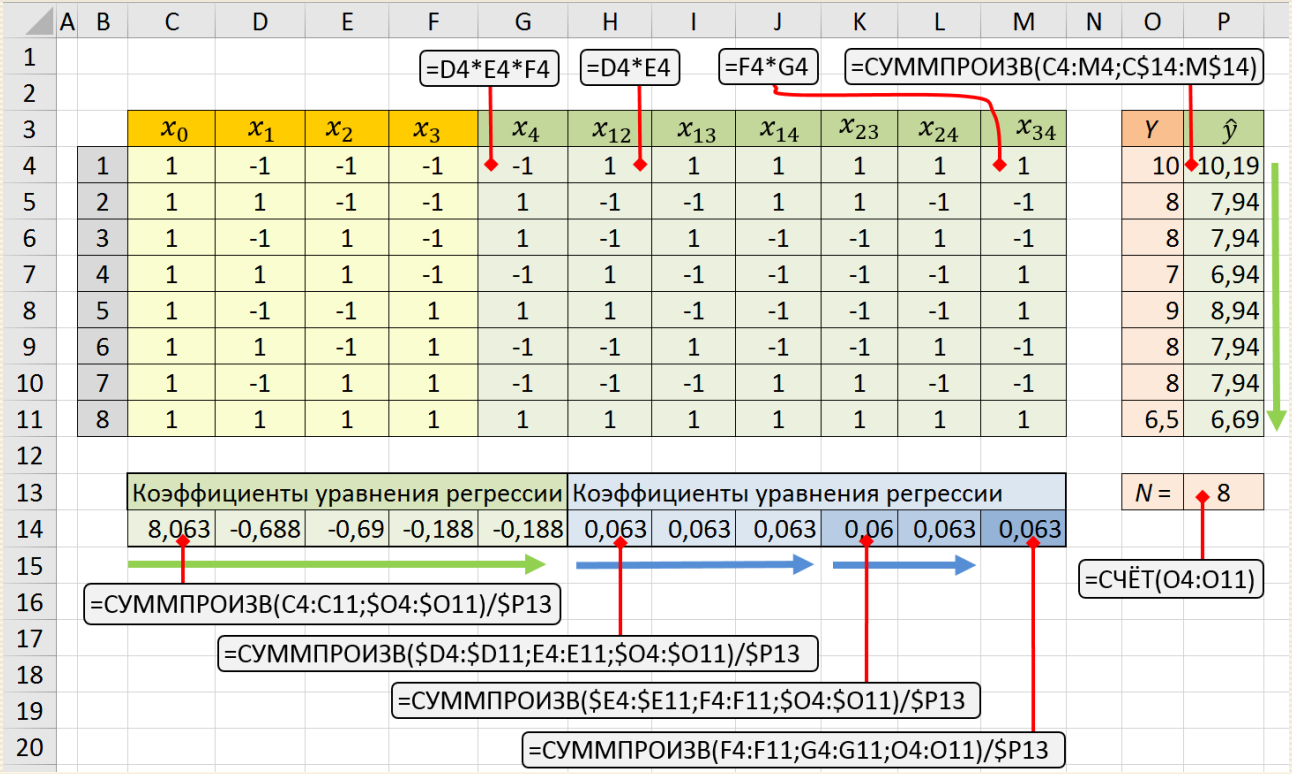


Рис. 9.1. План дробного факторного эксперимента и коэффициенты уравнения регрессии

При большом числе факторов (n>3) проведение полного факторного эксперимента связано с большим числом, экспериментов, значительно превосходящим число коэффициентов линейной модели. Если при получении модели можно ограничиться линейным приближением, то есть получить адекватную модель в виде полинома $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$, то число экспериментов можно резко сократить в результате использования так называемого дробного факторного эксперимента. Так, например, в полном факторном эксперименте типа 2^2 при линейном приближении коэффициент регрессии b_{12} можно принять равным нулю, а соответствующий столбец x_1x_2 матрицы (табл. 9.5) использовать для "нового" (третьего) фактора x_3.

Таблица 9.5. – Матрица планирования эксперимента

номер опыта	x_0	x_1	x_2	$x_3(x_1x_2)$	y
1	+1	+1	+1	+1	y_1
2	+1	-1	+1	-1	y_2
3	+1	+1	-1	-1	y_3
4	+1	-1	-1	+1	y_4

В этом случае линейная модель будет определяться уравнением $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$. Для определения коэффициентов этого уравнения достаточно провести четыре эксперимента вместо восьми в полном факторном эксперименте. Такой вид исследований называется полурепликой. При увеличении числа факторов (n>3) возможно применение реплик большей дробности.

Генерирующее соотношение – соотношение, показывающее, какие взаимодействия заменены новыми факторами при построении дробной реплики.

Дробная реплика – план эксперимента, являющийся частью полного факторного эксперимента.

Дробные реплики обозначают зависимостью 2^{n-p}, где p – число линейных эффектов, приравненных к эффектам взаимодействия.

При p = 1 получают полуреплику; при p = 2 получают 1/4 – реплику; при p = 3 получают 1/8 – реплику и т.д. по степеням числа 2. Так, например, если в полном факторном эксперименте 2^3 (табл. 9.4) один из эффектов взаимодействия (x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$) заменим четвертым фактором x_4, то получим полуреплику 2^{4-1} от полного факторного эксперимента 2^4. Если два эффекта взаимодействия заменить факторами x_4 и x_5, то получим ¼-реплику 2^{5-2} от полного факторного эксперимента 2^5.

Если для каждого конкретного набора факторов проводится несколько опытов, то значения y_i определяются, как правило, их средним значением (параллельных наблюдений) для значений данного набора факторов.

Параллельные измерения – измерения, выполненные при одинаковых значениях факторов. Используются для оценки дисперсии воспроизводимости.

Схема дисперсионного и регрессионного анализов планированного эксперимента

Ниже в общем виде приводится схема дисперсионного и регрессионного анализов планированного эксперимента для K факторов, когда каждый опыт в матрице планирования повторялся m раз (табл. 9.6).

Таблица 9.6. – Матрица планирования и результаты измерений

Номер опыта	x_0	x_1	x_2	...	x_K	y_{i1}	y_{i2}	...	y_{im}	$\overline{y}$
1	+1	+1	-1	...	+1	y_{11}	y_{12}	...	y_{1m}	$\overline{y}_1$
2	+1	-1	-1	...	+1	y_{21}	y_{22}	...	y_{2m}	$\overline{y}_2$
3	+1	+1	+1	...	+1	y_{31}	y_{32}	...	y_{3m}	$\overline{y}_3$
...	...	...	...	...	...	...	...	...	...	...
N	+1	-1	+1	...	-1	y_{N1}	y_{N2}	...	y_{Nm}	$\overline{y}_N$

При осуществлении регрессионного анализа предполагается следующее:

- в ходе проведения опыта значение каждого из факторов фиксируется на определенном неизменном уровне, а если вариация фактора и имеет место, то ее влияние на величину функции отклика незначительно;
- для каждого значения фактора (сочетания факторов) опыт повторяется m_i раз (выполняется серия из параллельных опытов).

В каждой строчке матрицы планирования определяется среднее значение измеряемой величины по m параллельным опытам:

$$\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij} \ , \quad i = 1,2,\dots,N \tag{9.4}$$

и выборочная дисперсия

$$D_i = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}\left(y_{ij} - \overline{y}_i\right)^2, \quad i = 1,2,\dots,N \ . \tag{9.5}$$

Далее проверяется однородность выборочных дисперсий по критерию Кохрена, если имеет место быть одинаковое количество параллельных опытов в каждой серии ($m_i = m$). Для этого составляется отношение максимальной дисперсии к сумме всех дисперсий:

$$G = \frac{D_{\max}}{\sum_{i=1}^{N} D_i}, \quad D_{\max} = \max_i\{D_i\} \ . \tag{9.6}$$

Полученное отношение сравнивается с табличным $G_{\text{крит}}(\alpha, df_1, df_2)$, где уровень значимости обычно принимается $\alpha = 0.05$, $df_1 = m - 1$, $df_2 = N$. Если $G < G_{\text{крит}}$, то считается, что дисперсии серии опытов о д н о р о д н ы .

В случае, если каждая серия определяется различным числом экспериментов m_i, то для анализа однородности используется критерий Бартлетта.

Величина y есть случайная величина с нормальным законом распределения. Одной из характеристик этого распределения является дисперсия воспроизводимости, которая получается в результате усреднения дисперсий в каждой строке матрицы, если они однородны. Если же гипотеза об однородности дисперсий не подтверждается, то (один из путей) следует искать такое преобразование критерия оптимизации, которое делает дисперсии однородными (например, логарифмированием данных).

Построчная дисперсия характеризует рассеивание результатов относительного среднего арифметического, заданных условиями одной строки матрицы.

В качестве оценки дисперсии воспроизводимости $D_{\text{воспр}}$ берется среднее арифметическое дисперсий опытов D_i; число степеней свободы $df_{\text{воспр}}$ этой дисперсии равно сумме чисел степеней свободы дисперсий опытов

$$D_{\text{воспр}} = \frac{1}{df_{\text{воспр}}} \sum_{i=1}^{N} \sum_{j=1}^{m_i} \left(y_{ij} - \overline{y}_i\right)^2 , \quad df_{\text{воспр}} = \sum_{i=1}^{N} m_i - N . \tag{9.7a}$$

Для одинакового количества параллельных опытов в каждой серии ($m_i = m$)

$$D_{\text{воспр}} = \frac{1}{df_{\text{воспр}}} \sum_{i=1}^{N} D_i , \quad df_{\text{воспр}} = N(m-1) . \tag{9.7b}$$

> Дисперсия воспроизводимости – дисперсия, характеризующая воспроизводимость эксперимента; вычисляется по данным параллельных измерений.

Число степеней свободы – понятие, учитывающее в статистических ситуациях связи, ограничивающие свободу изменения случайных величин; вычисляют как разность между числом экспериментальных точек и числом связей.

Коэффициенты уравнения регрессии определяются по формулам (9.3), где в качестве y_i используются средние значения, рассчитанные по параллельным экспериментам по соотношению (9.4).

Учитывая, что дисперсия $\overline{y}$ полученного по выборке объема m во столько же раз меньше дисперсии единичного измерения $D_{\overline{y}} = D_{\text{воспр}}/m$ в рассматриваемом примере (табл. 9.4) дисперсия D_{b_j} и стандартное отклонение S_{b_j} коэффициентов определяется следующим образом:

$$D_{b_j} = \frac{D_{\text{воспр}}}{\sum_{i=1}^{N} m_i}\,, \quad S_{b_j} = \sqrt{D_{b_j}}\,, \tag{9.8a}$$

В случае одинакового количества параллельных опытов в каждой серии ($m_i = m$)

$$D_{b_j} = \frac{D_{\text{воспр}}}{Nm}\,, \quad S_{b_j} = \sqrt{D_{b_j}}\,. \tag{9.8b}$$

Значимость коэффициентов проверяется по критерию Стьюдента. Для всех коэффициентов уравнения регрессии составляется t-отношение

$$t_j = \frac{|b_j|}{S_{b_j}}\,, \quad j = 0, \ldots, K\,, \tag{9.9}$$

которое сравнивается с критическим $t_{\text{крит}}$ для уровня значимости α (например, 0.05) и числа степеней свободы $df = df_{\text{воспр}}$.

Если $t_j < t_{\text{крит}}(\alpha, df)$, то соответствующий выборочный коэффициент b_j как незначимый отсеивается из уравнения регрессии. При этом остальные коэффициенты пересчитывать нет необходимости.

При практических вычислениях (см. пример 9.1) значимость коэффициентов удобнее определить единичной функцией δ_j (0 – коэффициент незначим, 1 – значим):

$$\delta_j = \begin{cases} 0, & \text{если } t_j/t_{\text{крит}} < 1, \\ 1, & \text{если } t_j/t_{\text{крит}} \geq 1. \end{cases}$$

Адекватность – соответствие.

Адекватность уравнения регрессии – соответствие уравнения регрессии опытным данным. Обычно соответствие оценивают в пределах ошибки воспроизводимости.

Дисперсия адекватности $D_{\text{ад}}$ определяется формулой

$$D_{\text{ад}} = \frac{1}{df_{\text{ад}}}\left[\sum_{i=1}^{N}\sum_{j=1}^{m_i}(y_{ij} - \hat{y}_i)^2 - \sum_{i=1}^{N}\sum_{j=1}^{m_i}(y_{ij} - \overline{y}_i)^2\right], \qquad (9.10a)$$

$$df_{\text{ад}} = \left(\sum_{i=1}^{N} m_i - k\right) - \left(\sum_{i=1}^{N} m_i - N\right) = N - k,$$

где k — число значимых коэффициентов в уравнении регрессии $k = \sum \delta_i$.

В случае одинакового количества параллельных опытов в каждой серии ($m_i = m$)

$$D_{\text{ад}} = \frac{m}{N-k}\sum_{i=1}^{N}\left(\hat{y}_i - \overline{y}_i\right)^2 \ , \quad df_{\text{ад}} = N - k \ , \tag{9.10b}$$

Проверка значимости различия между дисперсией адекватности и дисперсией воспроизводимости осуществляется по критерию Фишера, согласно которому стохастическая модель считается адекватной, если на принятом уровне значимости α выполняется неравенство

$$F = \frac{\max(D_{\text{ад}}, D_{\text{воспр}})}{\min(D_{\text{ад}}, D_{\text{воспр}})} < F_{\text{крит}}(\alpha, df_1, df_2), \tag{9.11}$$

где α — уровень значимости; df_1 – степень свободы большей дисперсии из $D_{\text{ад}}$ и $D_{\text{воспр}}$,

df_2 – меньшей.

Если $F > F_{\text{крит}}$, то для адекватного описания эксперимента необходимо увеличить порядок аппроксимирующего полинома.

После расчета значимых коэффициентов регрессии и оценки адекватности модели можно вернуться к натуральному масштабу факторов и получить выражение регрессионной зависимости в окончательном виде:

$$\hat{y} = b_0 + \sum_{j=1}^{K} b_j x_j = b_0 + \sum_{j=1}^{K} b_j \left(\frac{z_j - z_j^0}{\Delta z_j}\right)$$

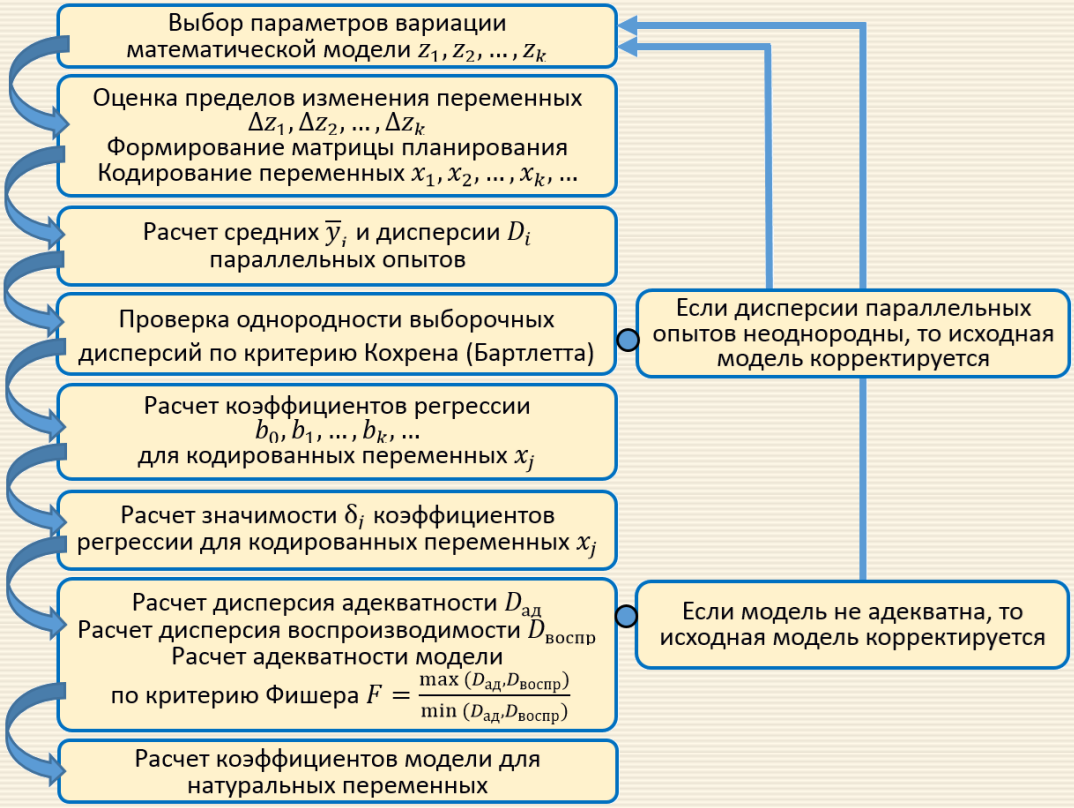
или $$\hat{y} = c_0 + \sum_{j=1}^{K} c_j z_j \ , \qquad (9.12)$$

где $$c_0 = b_0 - \sum_{j=1}^{K} c_j z_j^0 \ , \quad c_j = \delta_j \frac{b_j}{\Delta z_j} \ , \quad j = 1, \dots, K \ .$$

Полученная математическая модель (9.12) определяет степень влияния различных факторов на целевую функцию и позволяет предсказывать значения целевой функции при определенных значениях факторов. Модель может использоваться для поиска оптимальных условий (набора конкретных значений факторов), при которых целевая функция достигает экстремума (минимума или максимума)*. Например, при поиске условий измерения с минимальной погрешностью, при оптимизации условий достижения максимального выхода производимого продукта и т.д.

*Изложение представленного материала и возможный дальнейший анализ данных соответствует учебному пособию С.Л. Ахназарова, В. В. Кафаров. Методы оптимизации эксперимента в химической технологии // М.: Высш. шк., 1985. 327 с.

Схема построения и анализа математической модели для задачи планирования эксперимента



Для различных результатов исследования ниже приведена таблица по возможным заключениям выполненного предварительного анализа модели эксперимента.

Результат анализа	Возможные причины	Возможное решение
Модель адекватна и все коэффициенты линейной модели значимы		Завершение предварительного анализа модели, переход к более детальному исследованию.
Модель адекватна, часть коэффициентов линейной модели незначима	Неудачный выбор интервалов варьирования	Расширить интервал варьирования
	Фактор не влияет на параметр оптимизации	Стабилизировать значение фактора на определенном уровне или оставить неконтролируемым
	Плохая воспроизводимость опытов	Увеличить число параллельных опытов
Модель адекватна, но все коэффициенты незначимы	Большая ошибка эксперимента	Уменьшить дисперсии за счет увеличения числа опытов
	Выбраны узкие интервалы варьирования	Расширить интервал варьирования
Модель неадекватна	Неудачный выбор области факторного пространства	Изменить интервалы варьирования либо перенести центр плана

Пример 9.2 Определяется оптимальный состав фотохромного стекла в системе Li_2O – Al_2O_3 – SiO_2. В качестве параметров оптимизации (y) рассматривалась оптическая плотность в облученном состоянии. Для обеспечения получения вещества с необходимыми свойствами необходимо построить математическую модель для зависимости оптической плотности от следующих технологических факторов:

z_1, z_2 — исходные концентрации хлора и брома, г-атом/100 г стекла;
z_3 — соотношение $Ag : Cl$;
z_4 — температура варки, °С;
z_5 — время выдержки, ч;
z_6 — содержание Al_2O_3, мол. доли;
z_7 — соотношение Li_2O / SiO_2.

Условия эксперимента приведены в таблицах 9.7 и 9.8.

Таблица 9.7. – Матрица планирования и результаты измерений

	z_1	z_2	z_3	z_4	z_5	z_6	z_7
Основной уровень z^0	0.0425	0.0187	0.0675	1325	1.75	0.1395	0.4165
Интервал варьирования Δz	0.0205	0.0093	0.0325	25	0.25	0.0125	0.0835
+1	0.0630	0.0280	0.1	1350	2	0.1570	0.5
–1	0.0220	0.0094	0.035	1300	1.50	0.1240	0.333

Таблица 9.8. – Матрица планирования и результаты измерений

Номер серии	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_{i1}	y_{i2}
1	+1	–1	+1	–1	–1	+1	–1	0.000	0.000
2	+1	+1	+1	–1	+1	–1	–1	0.108	0.150
3	+1	–1	–1	–1	+1	+1	+1	0.000	0.000
4	+1	+	–1	–1	–1	–1	+1	0.194	0.160
5	+1	–1	+1	+1	–1	–1	+1	0.298	0.292
6	+1	+	+1	+1	+1	+1	+1	0.400	0.408
7	+1	–1	–1	+1	+1	–1	–1	0.255	0.278
8	+1	+	–1	+1	–1	+1	–1	0.453	0.408

Каждый опыт в матрице планирования (табл. 9.7) повторен два раза. Для определения коэффициентов линейного уравнения регрессии

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7$$

использована ПФЭ со следующими генерирующими соотношениями

$$x_4 = x_1x_2x_3, \quad x_5 = x_1x_2, \quad x_6 = x_1x_3 \quad x_7 = x_2x_3\,.$$

Вычисления выполняются в следующей последовательности.

1. В ячейки C4:J11, L16:M23, D30 заносятся исходные данные в соответствии с табл. 9.8 (рис. 9.2, 9.3).

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	+	—	+	—	+	—	+	—
2	+	+	+	—	—	+	—	—
3	+	—	—	—	—	+	+	+
4	+	+	—	—	+	—	—	+
5	+	—	+	+	—	—	—	+
6	+	+	+	+	+	+	+	+
7	+	—	—	+	+	+	—	—
8	+	+	—	+	—	—	+	—

=ЕСЛИ(C4="+";1;-1)

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	1	-1	1	-1	1	-1	1	-1
2	1	1	1	-1	-1	1	-1	-1
3	1	-1	-1	-1	-1	1	1	1
4	1	1	-1	-1	1	-1	-1	1
5	1	-1	1	1	-1	-1	-1	1
6	1	1	1	1	1	1	1	1
7	1	-1	-1	1	1	1	-1	-1
8	1	1	-1	1	-1	-1	1	-1

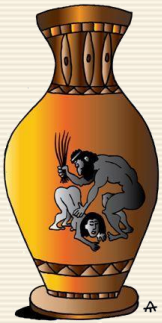


Рис. 9.2 Трансформация исходных знаковых" данных C4:J11 в числовые значения C16:J23

2. В диапазонах O16:O23 и P16:P23 по соотношениям (9.4) и (9.5) рассчитываются средние значения $\overline{y}_i$ и дисперсии D_i параллельных опытов.

3. В ячейках I29 и I30 определяются число N серий экспериментов и количество m опытов в серии (рис. 9.3).

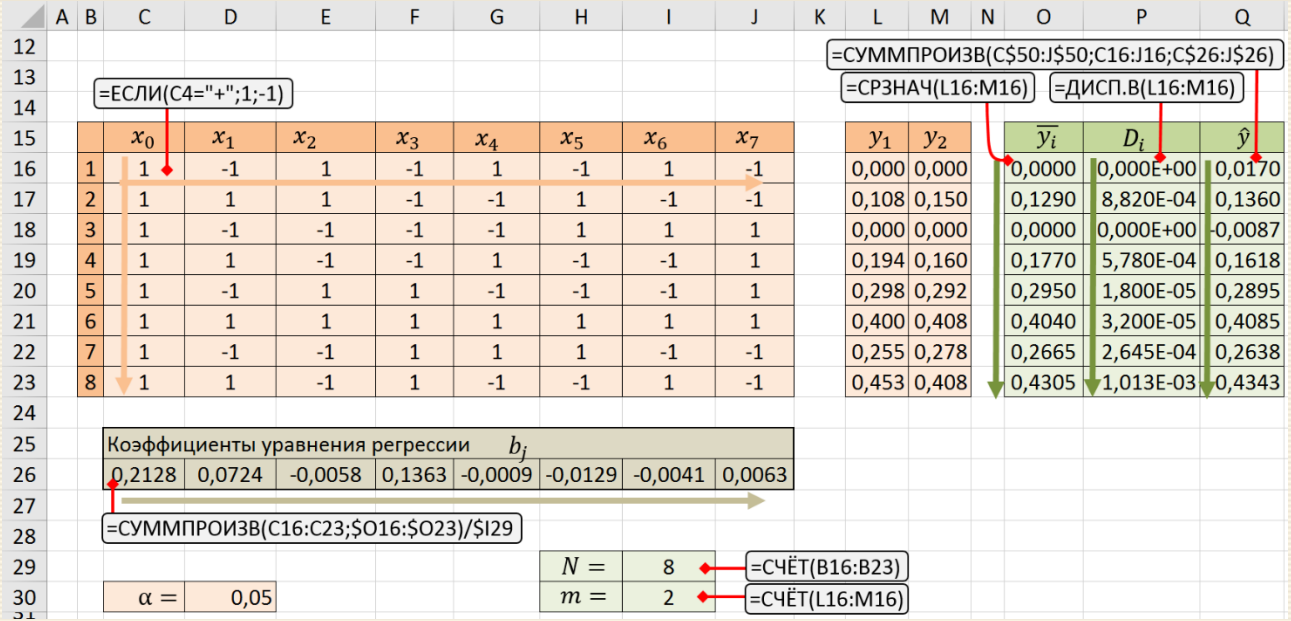


Рис. 9.3 Расчет коэффициентов уравнения регрессии

4. В диапазоне C26:J26 по соотношениям (9.3b) рассчитываются коэффициенты уравнения регрессии.

5. Проводится проверка однородности результатов экспериментов (рис. 9.4) по критерию Кохрена:

- в ячейках I34 и I35 для критерия Кохрена G рассчитываются степени свободы $df_1 = m - 1$, и $df_2 = N$;

- в ячейках D34, D35 определяется величина экспериментального G и $G_{\text{крит}}(\alpha, df_1, df_2)$ через Excel-функцию БЕТА.ОБР $(1-\alpha/N\,;\ df_1/2\,;\ df_1(df_2-1)/2)$;
- через формулу сравнения эмпирического и критического значений критерия в ячейке C36 формулируется заключение об однородности.

6. Рассчитывается $df_{\text{воспр}} = N(m-1)$ (ячейка E40) и дисперсия $D_{\text{воспр}}$ (формула (9.7), ячейка E39), дисперсия D_{b_j} и стандартное отклонение S_{b_j} для коэффициентов регрессии (формулы (9.8), ячейки E44 и E45).

7. Вычисляется критическое значение критерия Стьюдента (ячейка D49) и определяется значимость коэффициентов уравнения регрессии $b_i (i = 0,1,\dots,7)$, указываемая в диапазоне C50:J50 в виде единичной функции δ_j (1 – коэффициент значим; 0 – не значим).

8. Подсчитывается число значимых коэффициентов в уравнении регрессии (ячейка D53) и проводится проверка адекватности уравнения экспериментальным данным, для чего (рис. 9.4)

- в ячейке D57 по $df_{\text{ад}}$ (ячейка E58) определяется величина дисперсии адекватности $D_{\text{ад}}$ – расчёт проводится по формуле (9.11);
- в ячейках H58 и H59 рассчитываются степени свободы для критерия Фишера;
- в ячейках H57 и H60 определяются величины экспериментального и критического значений критерия Фишера; по их соотношению в ячейке C61 формулируется заключение об адекватности (рис. 9.5) уравнения регрессии экспериментальным данным.

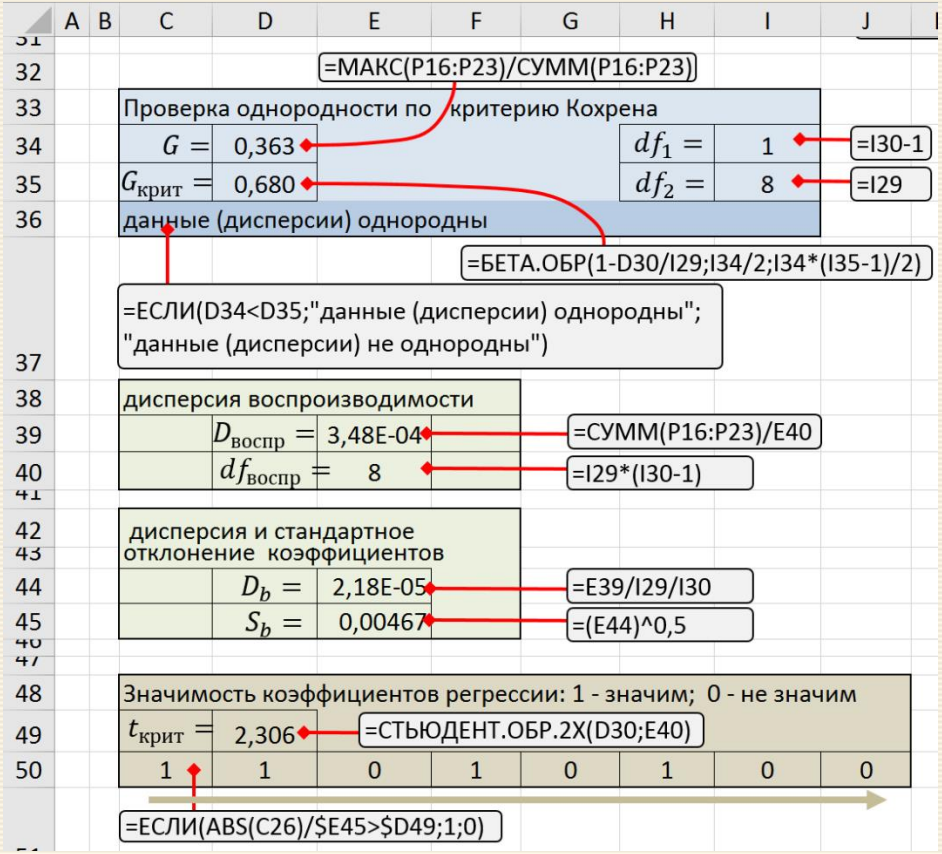


Рис. 9.4 Расчет значимости коэффициентов уравнения регрессии

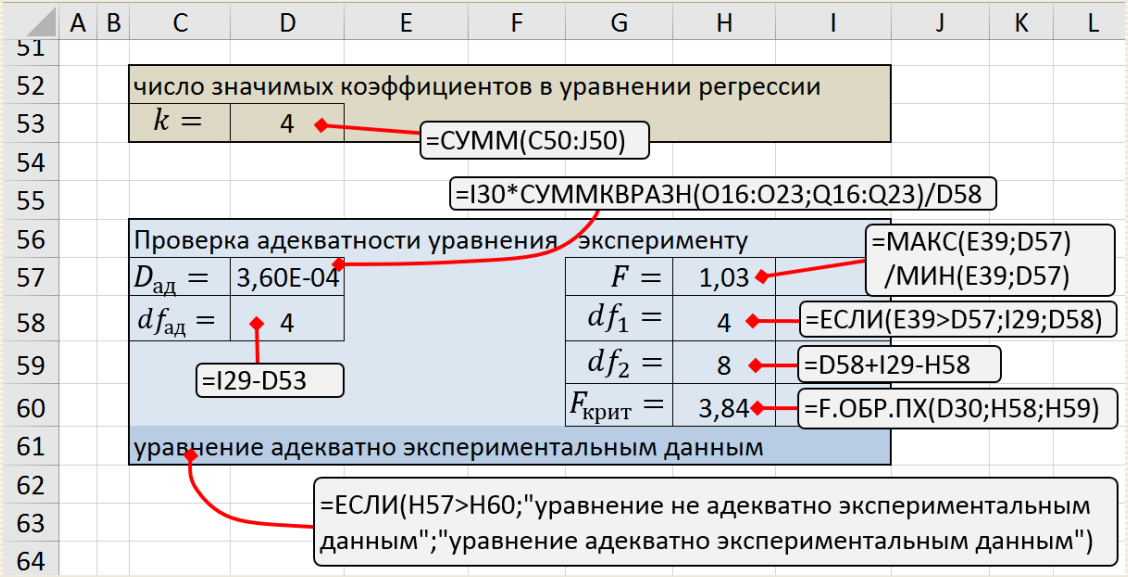


Рис. 9.5 Проверка адекватности математической модели

Таким образом, адекватное экспериментальным данным уравнения регрессии для кодированных переменных имеет вид

$$y = 0.2128 + 0.0724x_1 + 0.1363x_3 - 0.0129x_5 .$$

Данное уравнение для дальнейших исследований на основе формул (9.1) и данных табл. 9.7 можно записать в естественных переменных зависимость оптической плотности τ от параметров процесса

$$\tau = -0.1302 + 3.5305 \cdot [\mathrm{Cl}]z_1 + 4.1923\left(\frac{\mathrm{Ag}}{\mathrm{Cl}}\right) - 0.0515\, t_{\text{выд}} ,$$

где $[\mathrm{Cl}]$ — концентрация хлора; $\left(\frac{\mathrm{Ag}}{\mathrm{Cl}}\right)$ — соотношение Ag : Cl; $t_{\text{выд}}$ — время выдержки, часов.

На рис. 9.6 дан скриншот заключительной части рабочего листа MS Excel, где производится пересчет кодированных факторов к натуральным.

	A	B	C	D	E	F	G	H	I	J
66										
67			z_0	0,04	0,02	0,07	1325	1,75	0,14	0,42
68			Δz_0	0,02	0,01	0,03	25	0,25	0,01	0,08
70			c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7
71			-0,1302	3,5305	0	4,1923	0	-0,0515	0	0
72										
73										
74										

=D26/D68*D50

=C26-СУММПРОИЗВ(D71:J71;D67:J67)

Рис. 9.6 Пересчет кодированных факторов к натуральным значениям

Приложение П1. Однородность, дисперсия и ошибка среднего

П1.1. Однородность и коэффициент вариации

При описании статистических данных используется такое понятие как "однородность", существенно влияющая на точность рассчитываемых показателей и качество аналитических выводов*. Чем однороднее данные, тем надёжнее и ближе к реальности результаты статистического анализа. Тем не менее однородность – понятие относительное и достаточно растяжимое, не имеющая точных границ и критериев. Под однородными данными следует понимать некоторый уровень их рассеяния, при котором рассчитываемые статистические показатели (например, среднее) будут давать надёжную и качественную характеристику анализируемой совокупности. Граница, отделяющая однородные данные от неоднородных, весьма плавная и размытая.

Основным мерилом разброса (и, соответственно, однородности) данных являются показатели вариации: дисперсия, среднее квадратическое отклонение, среднее линейное отклонение, напрямую связанные с масштабом исходных данных и не дающие "независимой", (безразмерной, относительной) характеристики меры разброса. Безразмерной величиной, характеризующей относительный разброс данных, является коэффициент вариации, который рассчитывается как отношение среднего квадратического отклонения к среднему для выборки.

Данный показатель вариации, учитывающий меру рассеяния учитывает, и единиц измерения не имеет и не связан с масштабом анализируемых данных. Исходя из этого факта, коэффициенты вариации можно сравнивать между собой и тем самым сопоставлять относительную меру рассеяния данных, независимо от их масштаба.

*Излагается по материалам и текстам Дмитрия Езепова https://statanaliz.info/

Таким образом, основным показателем, характеризующим однородность данных, является коэффициент вариации. В статистике принято считать, что, если значение коэффициента менее 33%, то совокупность данных является однородной, если более 33%, то – неоднородной.

На последующих примерах коэффициенты вариации иллюстрируется на данных некоторых выборок. На рис. П1.1 дано графическое изображение данных для выборки со средним значением 100 и генерируемых вокруг этого функцией случайных чисел значений самой выборки с максимальным отклонением ± 40.

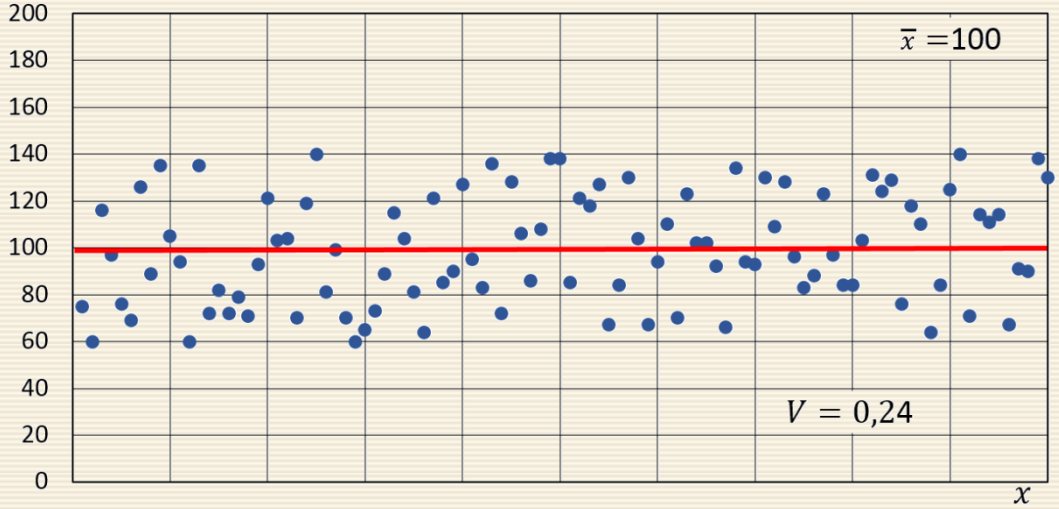


Рис. П1.1. Выборка с максимальным отклонением ± 40 единиц

Для указанных на рис. П1.1 данных коэффициент вариации составил 0.24, то есть совокупность можно считать однородной.

На рис. П1.2 дано графическое изображение данных для выборки с тем же средним значением 100 и выборки с максимальным отклонением ± 80.

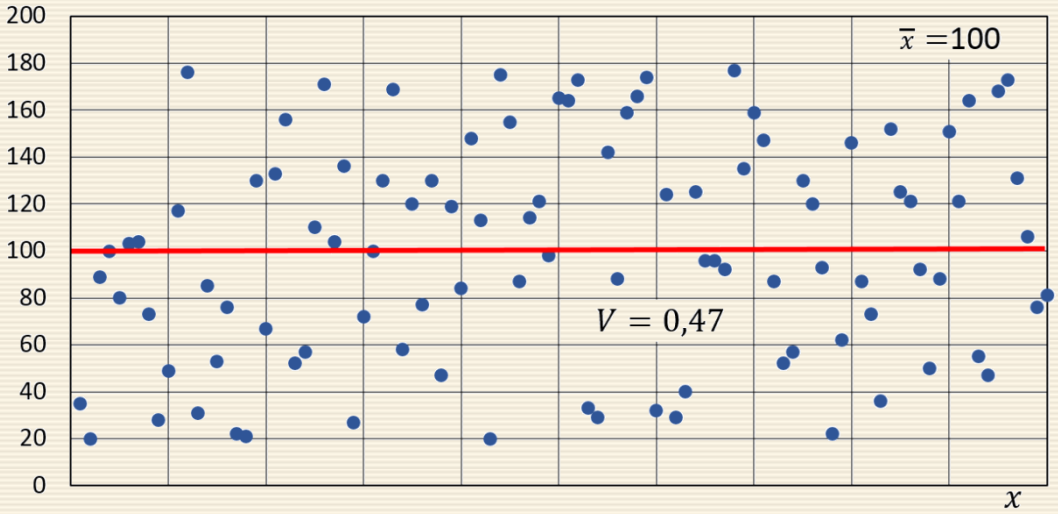


Рис. П1.2. Выборка с максимальным отклонением ± 80 единиц

Отчетливо видно, что данные здесь более рассеяны. Коэффициент вариации в этом случае составляет 45%, следовательно, выборку не следует считать однородной.

Из приведенных рисунков можно получить первичное представление об отличии однородных данных от неоднородных. На первом – данные однородны со значением коэффициента вариации 24%, на втором – неоднородны со значением вариации 0.47. Чем более однородны данные, тем ближе они находятся к среднему значению. Чем менее однородны, тем больше рассеяны и находятся дальше друг от друга и от своей средней. Неоднородность может иметь не только случайный характер в некотором диапазоне, но может быть вызвана совершенно различными обстоятельствами и иметь самую замысловатую конфигурацию.

Далее представлены некоторые типы рассеяния значений, которые встречаются в реальности как и отдельно, так и в комбинациях. Если уровень коэффициента вариации ниже 33%, то данные принято считать однородными, независимо от характера разброса. Но если вариация выше 33%, то данные должны быть проанализированы на данный предмет.

На рис. П1.3 представлен тип неоднородности, который объясняется присутствием в выборке аномальных значений. Данные выборки соответствуют значениям рис. П1.1, одно из которых заменено на значение 500, существенно отличающееся от других элементов выборки.

Коэффициент вариации с 0.24 сразу вырос до 0.45. Всего одно аномальное значение из сотни увеличило показатель вариации почти в 2 раза!

Вывод: удаление из выборки аномальных наблюдений существенно снижает показатель вариации. Анализ исключения аномальностей – грубых промахов и выбросов – приведен в разделе 8.

Некоторые разбросы связаны с качественным различием в данных. Например, если анализируются данные по предприятиям из различных отраслей, то, отличие будет объясняться, прежде всего, их различной производственной направленностью.

Рис. П1.3. Выборка с аномальным значением

Чаще всего встречаются разбросы данных без каких-либо видимых на то причин. Даже после удаления аномальных наблюдений уровень разброса практически сохраняется и требуются специальные приемы анализа ("принудительная" группировка и пр.).

Выше перечисленные типы разброса относятся к условно говоря пространственному представлению данных, когда характеризуется явление или процесс в разных местах пространства, но в одно и то же время.

В статистике различают и другой вид данных – динамические. Таковые характеризуют развитие одного и того же объекта во времени. Как правило, это один и тот же показатель, который фиксируется через некоторые промежутки времени. Полученные данные называются д и н а м и к о й . Рассматривая такие данные с точки зрения однородности можно обнаружить некоторые важные особенности.

Приведенные ранее на рис. П1.1 можно интерпретировать как динамические, а горизонтальную ось считать осью времени (рис. П1.4.)

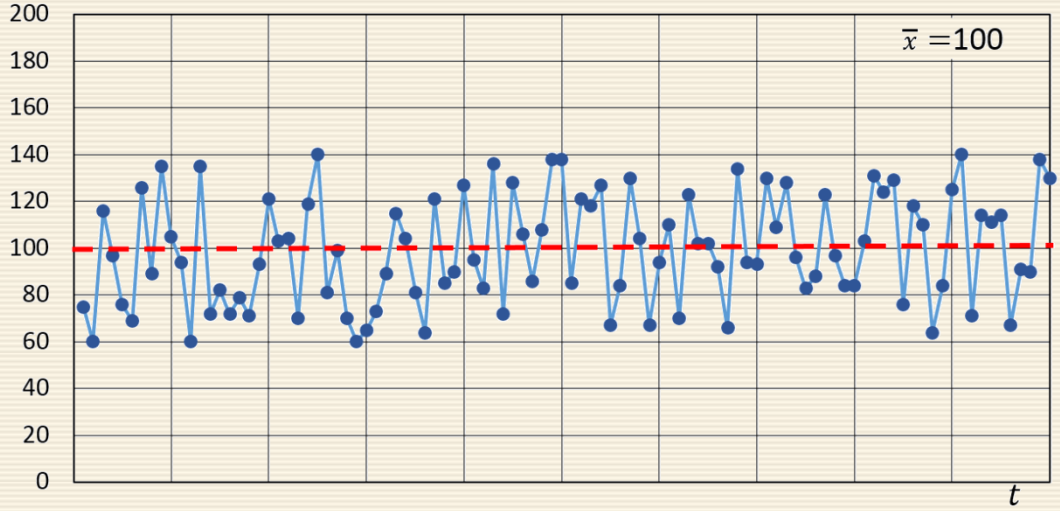


Рис. П1.4. Ряд динамики

Уровень колеблемости (рассеяние данных в динамике) в динамике измеряется так же, как и уровень рассеяния в пространстве – с помощью коэффициента вариации. Если колеблемость носит случайный, то есть необъяснимый характер, то что процесс считается нестабильным, т.е. стабильность оценивается коэффициентом вариации. Чем больше необъяснимая вариация, тем менее стабилен процесс.

Можно отметить следующее: стабильность (или колеблемость) тесно связана с понятием прогнозируемости. Чем стабильнее процесс, тем его легче прогнозировать, и наоборот. Другими словами, чем меньше случайные отклонения показателя от ожидаемого уровня, тем точнее можно сделать прогноз. Таким образом, с помощью коэффициента вариации измеряют уровень колеблемости (или стабильности) и вместе с тем уровень прогнозируемости динамических данных.

Еще один тип разброса данных связан с проявлением какого-либо явления в динамике. Многие явления в природе имеют некоторую т е н д е н ц и ю развития. Если она нулевая (уровень не растет и не уменьшается в долгосрочной перспективе), то данные будут похожи на колебания вокруг постоянного уровня (среднего значения), как на вышеприведенных рисунках выше. Но чаще значения показателей под воздействием некоторых факторов со временем растут или уменьшаются, то есть имеют некоторую тенденцию (рис. П1.5). Например, изменение размеров листьев при росте растения.

Понятное дело, что наличие тенденции отдаляет многие значения от средней, что автоматически отражается на коэффициенте вариации, завышая его.

Коэффициент вариации не определяет и не учитывает наличие тенденций и зависит только от отклонений от средней в рассматриваемой совокупности. В итоге многие значения под действием вполне конкретных факторов находятся далеко от среднего уровня, увеличивая тем самым коэффициент вариации с 0.24 до 0.28. Уменьшить коэффициент вариации в ряде динамики, в котором присутствует выраженная тенденция, можно путем ее специального учета.

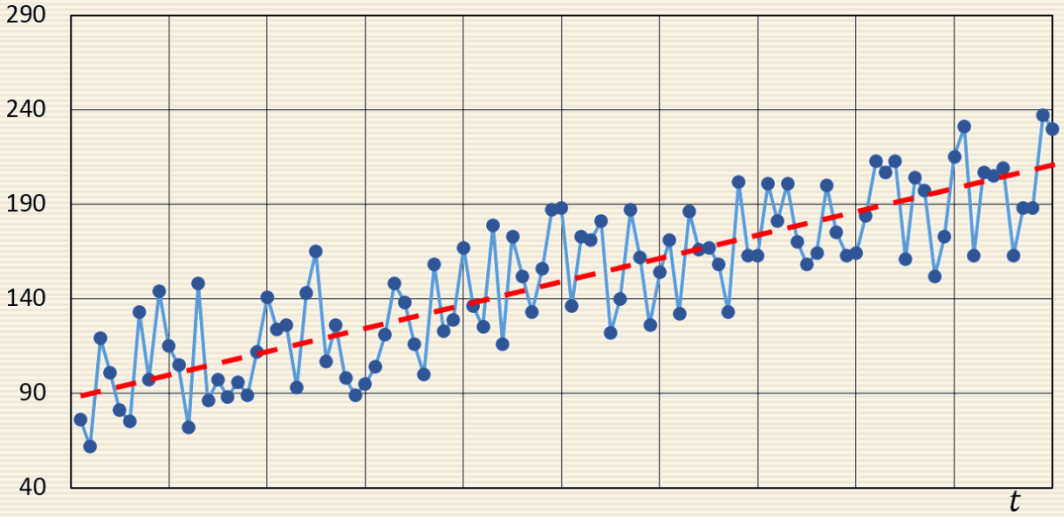


Рис. П1.5. Ряд динамики с тенденцией развития (роста)

Еще один тип разбросов связан с наличием в данных динамики цикличности или сезонности (рис. П1.6), когда в наблюдениях присутствует какой-то достаточно контролируемый и прогнозируемый фактор. Коэффициент вариации при этом существенно увеличивается. Например, наблюдается цикличность всплеска продаж для каких-либо товаров в выходные дни.

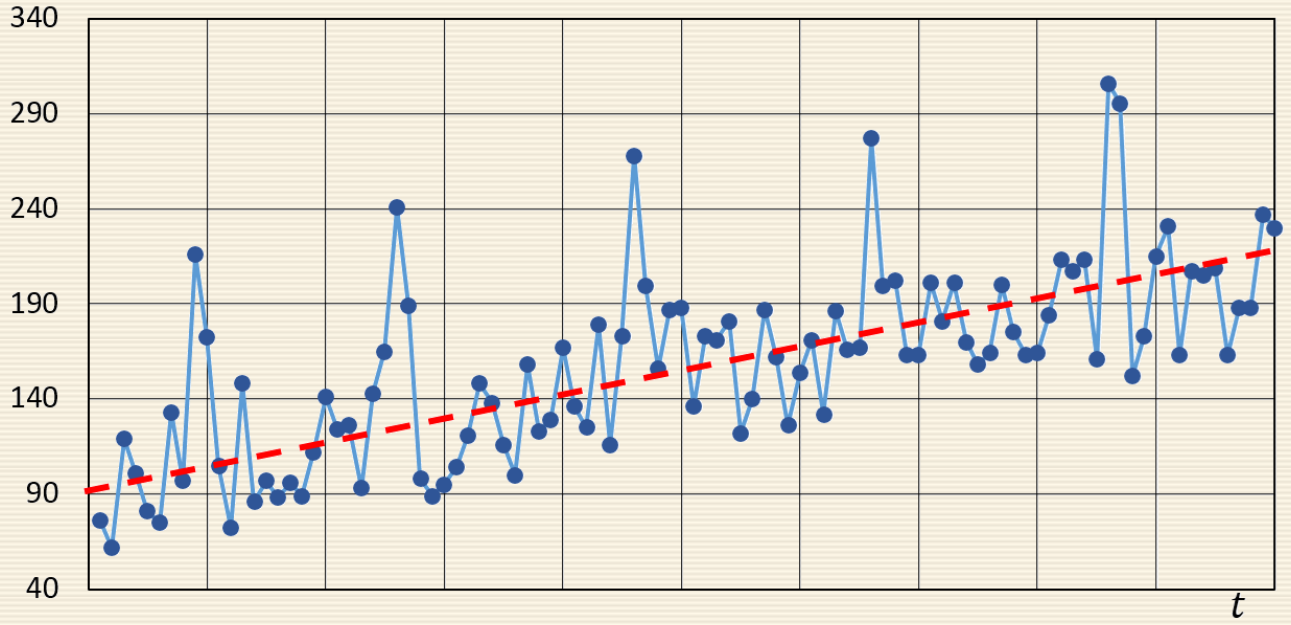


Рис. П1.6. Ряд динамики с тенденцией развития при наличии цикличности

Цикличность и тенденция увеличивают общую вариацию, что может привести к неверным выводам относительно прогнозируемости. Для решения проблемы обе компоненты убираются из ряда динамики с помощью специальных методов. Уровень прогнозируемости правильно считать по оставшейся части динамического ряда (только случайной компоненте).

Таким образом, выше представлены основные типы причин, которые оказывают непосредственное влияние на уровень коэффициента вариации. Часто в данных присутствуют не один вид, а несколько. Например, динамика может иметь и тенденцию, и сезонность, и аномальные всплески с провалами.

Основные способы, с помощью которых можно устранить или снизить неоднородность данных, следующие:

- устранение аномальных наблюдений;
- группировка данных по резкому качественному отличию;
- "принудительная" группировка;
- устранение тенденции в динамических данных;
- устранение сезонности в динамических данных.

В реальности однородные данные в готовом виде встречаются редко, поэтому перед проведением анализа их необходимо обработать таким образом, чтобы качество данных не пострадало и не исказились закономерности, а неоднородность исчезла.

П1.2. Дисперсия смещенная и несмещенная. Стандартная ошибка среднего

С точки зрения охвата объекта исследования, статистический анализ разделяется на два вида: сплошной и выборочный. Сплошное наблюдение, сплошной статистический анализ предполагает изучение генеральной совокупности данных, то есть всего явления во всем его многообразии без распространения выводов на другие элементы, не входящие в анализируемую совокупность. Из названия данного типа явствует, что наблюдению подвергаются абсолютно все элементы. Результат анализа распространяется на всю генеральную совокупность без каких-либо допущений и поправок на ошибку. Сплошное статистическое исследование является наиболее полным и точным, так как дополнительные знания почерпнуть уже неоткуда – информация собрана со всех элементов объекта исследования.

Хорошим примером сплошного наблюдения является перепись населения. Но, тем не менее, у данного наблюдения есть отрицательное качество: на организацию и проведение исследования могут потребоваться значительные ресурсы.

Название метода "выборочное наблюдение" точно отражает его суть: из генеральной совокупности отбирается и анализируется только часть данных, а выводы распространяют на всю генеральную совокупность. Отбор данных происходит таким образом, чтобы выборка была репрезентативной, то есть, сохранила внутреннюю структуру и закономерности генеральной совокупности. Если это условие не соблюдено, то дальнейший анализ во многом теряет смысл.

Сам анализ выборочных данных происходит так же, как и при сплошном наблюдении (рассчитываются различные показатели, делаются прогнозы и т.д.), только с поправкой на ошибку. Понятно, что при повторной выборке его значение всегда будет иным.

При малом объеме данных средняя величина постоянно меняется и для решения проблемы необходимо увеличивать выборку. Большая выборка, очевидно, дает более надёжные результаты, чем маленькая, но даже в этом случае ошибка имеется, хотя и становится меньше. Но достаточно часто увеличение объема выборки весьма затруднено (стоимость, сроки исследования и пр.) и тогда, volens-nolens необходимо работать с малыми выборками.

У выборочного наблюдения есть один существенный плюс и один минус, однако по сравнению со сплошным наблюдением крайности меняются местами. Плюс заключается в том, что для проведения выборочного обследования требуется гораздо меньше ресурсов. Минус – в том, что выборочное наблюдение всегда ошибочно. Поэтому основная задача проведения выборочного наблюдения – добиться максимальной точности при приемлемых затратах на его проведение.

Выборочная несмещенная дисперсия

Дисперсия, как и среднее арифметическое, меняет свое значение от выборки к выборке, причем имеет место быть интересная особенность вычислений. Дисперсия рассчитывается от средней величины, которая, в свою очередь тоже рассчитывается по выборке, то есть является ошибочной. Как же это обстоятельство влияет на саму дисперсию?

Если бы была известна истинная средняя величина (по генеральной совокупности), то ошибка дисперсии была бы связана только с нерепрезентативностью, то есть с тем, что данные в выборке оказались бы ближе или дальше от средней, чем в целом по генеральной совокупности. Соответственно, что при многократном повторении данные стремились бы к своему реальному расположению относительно средней.

Выборочный показатель, который при многократном повторении выборки стремится к своему теоретическому значению, называется несмещенной оценкой. Термин "оценка" употребляется в связи с незнанием реального значения (в генеральной совокупности) показателя, а посредством выборочного наблюдения показатель именно и только оценивается. Оценка показателя – это есть его характеристика, рассчитанная по выборке.

Примером из жизни могут служить оценки в школе. Школьнику задаются вопросы, на основе чего оцениваются его знания (производится как бы выборочное наблюдение). Оценка знаний школьника может быть ошибочна, что многие знают по себе, хотя почему-то каждый считает, что его оценки занижают. А преподователи зачастую считают, что выставляют завышенные оценки.

Выборочное среднее – это несмещенная оценка математического ожидания, так как среднее из выборочных средних стремится к своему теоретическому значению по генеральной совокупности. Оно расположено в центре выборки; среднее всегда находится в центре значений, по которым оно рассчитано – на то оно и среднее. А раз выборочное среднее находится в центре выборки, то из этого следует, что сумма квадратов расстояний от каждого значения выборки до выборочного среднего всегда меньше, чем до любой другой точки, в том числе и до среднего генеральной совокупности. Следовательно, дисперсия в каждой выборке будет занижена; средняя из заниженных дисперсий также даст заниженное значение. То есть при многократном повторении эксперимента выборочная дисперсия не будет стремиться к своему истинному значению (как выборочное среднее), а будет смещена относительно истинного значения по генеральной совокупности.

Отклонение выборочного среднего от генеральной показано на рис. П1.7.

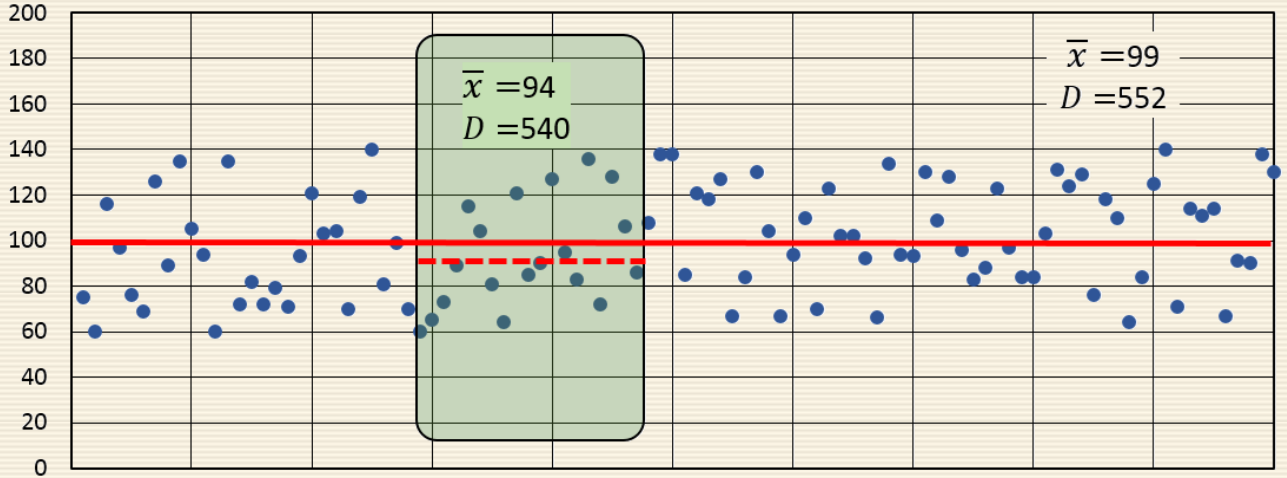


Рис. П1.7. Среднее и дисперсия для выборочной и генеральной совокупностей

Несмещенность оценки – одна из важных характеристик статистического показателя. Смещенная оценка показателя заранее говорит о тенденции к ошибке. Поэтому показатели стараются оценивать таким образом, чтобы их оценки были несмещенными (как у средней арифметической). Для того, чтобы решить проблему смещенности оценки выборочной дисперсии в ее расчет вносят корректировку – домножают на $n/(n-1)$, либо сразу при расчете в знаменатель ставят не n, а $(n-1)$.

Отметим лишний раз, что соотношение между выборочной и генеральной дисперсией составляет $n/(n-1)$.

	Выборочная смещенная дисперсия (генеральная совокупность)	Выборочная несмещенная дисперсия ("малые" выборки)
формула	$D = \frac{\sum(x_i - \overline{x})^2}{n}$	$D = \frac{\sum(x_i - \overline{x})^2}{n-1}$
функция Excel вычисления дисперсии	ДИСП.Г (данные)	ДИСП.В (данные)
функция Excel вычисления стандартного отклонения	СТАНДОТКЛОН.Г (данные)	СТАНДОТКЛОН.В (данные)

Под выборочной дисперсией понимают, как правило, именно несмещенный вариант.

Легко увидеть, что с ростом n (объема выборки) данное выражение стремится к единице, то есть разница между значениями выборочной и генеральной дисперсиями уменьшается. Таким образом, эффект смещенной дисперсии проявляется в небольших выборках. В больших выборках можно использовать генеральную дисперсию. При большом объеме выборки (более 100 наблюдений) разница между смещенной и несмещенной дисперсиями практически исчезает.

Свойства дисперсии

Свойство 1. Дисперсия постоянной величины A равна 0 (нулю), т.е. у постоянной величины нет отклонений: $D(A) = 0$.

Свойство 2. Если случайную величину умножить на постоянную A, то дисперсия этой случайной величины увеличится в A^2 раз. Другими словами, постоянный множитель можно вынести за знак дисперсии, возведя его в квадрат: $D(AX) = A^2 \cdot D(X)$. Данное свойство вполне очевидно, если вспомнить, что при расчете дисперсии отклонения от средней возводятся в квадрат.

Свойство 3. Если к случайной величине добавить (или отнять) постоянную A, то дисперсия останется неизменной: $D(A + X) = D(X)$. Это свойство также вполне понятно, т.к. все значения и их среднее увеличиваются на одну и ту же величину, и при взятии их разностей, величина A просто сокращается.

Свойство 4. Если случайные величины X и Y независимы, то дисперсия их суммы равна сумме их дисперсий: $D(X + Y) = D(X) + D(Y)$.

Свойство 5. Если случайные величины X и Y независимы, то дисперсия их разницы также равна сумме дисперсий. следует из того, что дисперсия всегда положительна (все отклонения от средней возводятся в квадрат): $D(X - Y) = D(X) + D(Y)$.

Расчет дисперсии и стандартной ошибки среднего арифметического

Чтобы получить дисперсию среднего арифметического достаточно иметь только одну выборку. Это легко доказать: среднее арифметическое рассчитывается по формуле:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n},$$

где x_i – значения переменной,

n – количество значений.

Используя свойства дисперсии, согласно которым, 1) – постоянный множитель можно вынести за знак дисперсии, возведя его в квадрат и 2) – дисперсия суммы независимых случайных величин равняется сумме соответствующих дисперсий. Предполагая, что каждое случайное значение x_i обладает одинаковым разбросом, можно получить формулу дисперсии средней арифметического:

$$D(\overline{x}) = D\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n^2} D(x_1 + x_2 + \cdots + x_n) = \frac{1}{n^2} nD(x) = \frac{D(x)}{n}.$$

На практике генеральная дисперсия $D(x)$ известна далеко не всегда, а в качестве таковой используют выборочную дисперсию.

Стандартное отклонение средней арифметического (или стандартная ошибка среднего) есть не что иное как корень квадратный из дисперсии. Поэтому формула для вычисления стандартной ошибки среднего m при использовании генеральной дисперсии имеет вид

$$m = S(\overline{x}) = \frac{S(x)}{\sqrt{n}}.$$

Данная формула на практике используется чаще всего, поскольку генеральная дисперсия, как правило, не известна.

П2.1. Степень(и) свободы. Классы и группировка

Для того чтобы свести к минимуму ошибки, в таблицах критических значений статистических критериев в общем количестве данных не учитывают те, которые можно вывести методом дедукции. Оставшиеся данные составляют так называемое число степеней свободы.

> Число степеней свободы df – число данных из выборки, значения которых могут быть случайными.

Так, если среднее трех данных равно 2, то первые два из них могут принимать любые значения, и если они определены, то третье значение фактически становится известным. Если, например, значение первого данного равно 2, а второго – 1, то третье может быть равным только 3. Таким образом, в такой выборке имеются только две степени свободы.

В общем случае для выборки в n данных существует $df = n$-1 степень свободы. Если имеются две независимые выборки, то число степеней свободы для первой из них составляет $(n_1 - 1)$, а для второй совокупности – $(n_2 - 1)$. А поскольку при определении достоверности разницы между ними опираются на анализ каждой выборки, то число степеней свободы, по которому определяется критическое значение (например, критерия Стьюдента t), будет равно $(n_1 - n_2) - 2$.

Если же речь идет о двух зависимых выборках, то в основе расчета лежит вычисление суммы разностей, полученных для каждой из n пар данных (т.е., например, разностей между результатами до и после воздействия на одного и того же испытуемого). Поскольку одну (любую) из этих разностей можно вычислить, зная остальные разности и их сумму, число степеней свободы для определения того же (например, парного критерия Стьюдента t) будет равно $n - 1$.

> Группировка – разбиение совокупности на группы (классы), однородные по какому-либо признаку или объединение отдельных единиц совокупности в группы, однородные по каким-либо признакам.

Обычно метод группировки рассматривается в отношении группировочного признака к интервалу значений. Ниже приводятся основные используемые термины.

> Класс (анг.: class, лат.: classis – разряд, класс, группа). Совокупность однородных предметов, обладающих каким-то определенным качеством, свойством или отношением.

> Кластерный анализ (анг.: cluster analysis) есть статистический метод выделения групп (классов, групп, блоков), схожих между собой элементов в их множестве.

> Варианты – отдельные количественные выражения признака. Классическое понимание термина "варианта" предполагает, что вариантой называется каждое уникальное значение признака без учета количества повторов.

Имеется два пути практической группировки и получения вариационного ряда: задавшись границами классовых интервалов (классов) или задавшись их количеством.

В "Пакете анализа" электронных таблиц Microsoft Excel имеются удобные средства группировки для случая, когда пользователем заданы границы классовых интервалов (в терминологии Microsoft Excel – к а р м а н о в), либо они устанавливаются автоматически по специальному встроенному в "Пакет анализа" алгоритму. Поэтому выполнить группировку не представляет труда. Во втором случае для вариационного ряда число классов равно числу градаций переменной, выбранном исследователем. При этом число классов дискретного вариационного ряда обычно равно числу градаций вариант выборки, измеренной в порядковой шкале. Для интервального вариационного ряда число классов задается на основе какого-либо правила. Возможность задать число классов Microsoft Excel не предоставляет.

Критерием правильности выбора количества классов считается верная передача типа распределения эмпирических частот данной выборочной совокупности. Если выбрано слишком мало классов, можно потерять характерную картину эмпирического распределения. При слишком подробном делении можно затушевать реальную картину распределения частот случайными отклонениями.

Выделяются несколько способов вычисления числа классов для выборок умеренной численности. Повсеместно распространенным является правило Sturgess'a.

По правилу Sturgess'a число классов M для выборки объемом n оценивается формулой

$$M = 1 + \log_2 n \ ,$$

$$\text{или} \quad M \approx 1 + 1.44 \cdot \ln n \ ,$$

$$\text{или} \quad M \approx 1 + 3.32 \cdot \lg n \ .$$

После решения вопроса о числе классов производится вычисление границ классовых интервалов и разнесение вариант исходной количественной выборки по классовым интервалам.

Интервал группировки – это интервал значений варьирующего признака, лежащих в пределах определенной группы. Каждый интервал имеет свою длину (ширину), верхнюю и нижнюю границы.

Нижняя граница интервала – это наименьшее значение признака в интервале, а верхняя граница интервала – его наибольшее значение. За нижнюю границу первого интервала принимают наименьшее значение признака в совокупности единиц наблюдения. Верхняя граница последнего интервала не может быть меньше наибольшего значения признака в совокупности единиц наблюдения.

Ширина интервала – это разность между верхней и нижней границами. Интервалы группировки в зависимости от их ширины бывают равными и неравными. Неравные делятся на прогрессивно возрастающие, прогрессивно убывающие, произвольные и специализированные.

Если вариация признака проявляется в сравнительно узких границах и распределение носит равномерный характер, то строят группировку с равными интервалами величины h

$$h = \frac{x_{\text{макс}} - x_{\text{мин}}}{M},$$

где $x_{\text{макс}}$, $x_{\text{мин}}$ – максимальное и минимальное значение признака в совокупности (выборке);

M – число групп (классов).

Данная формула определяет так называемый шаг интервала.

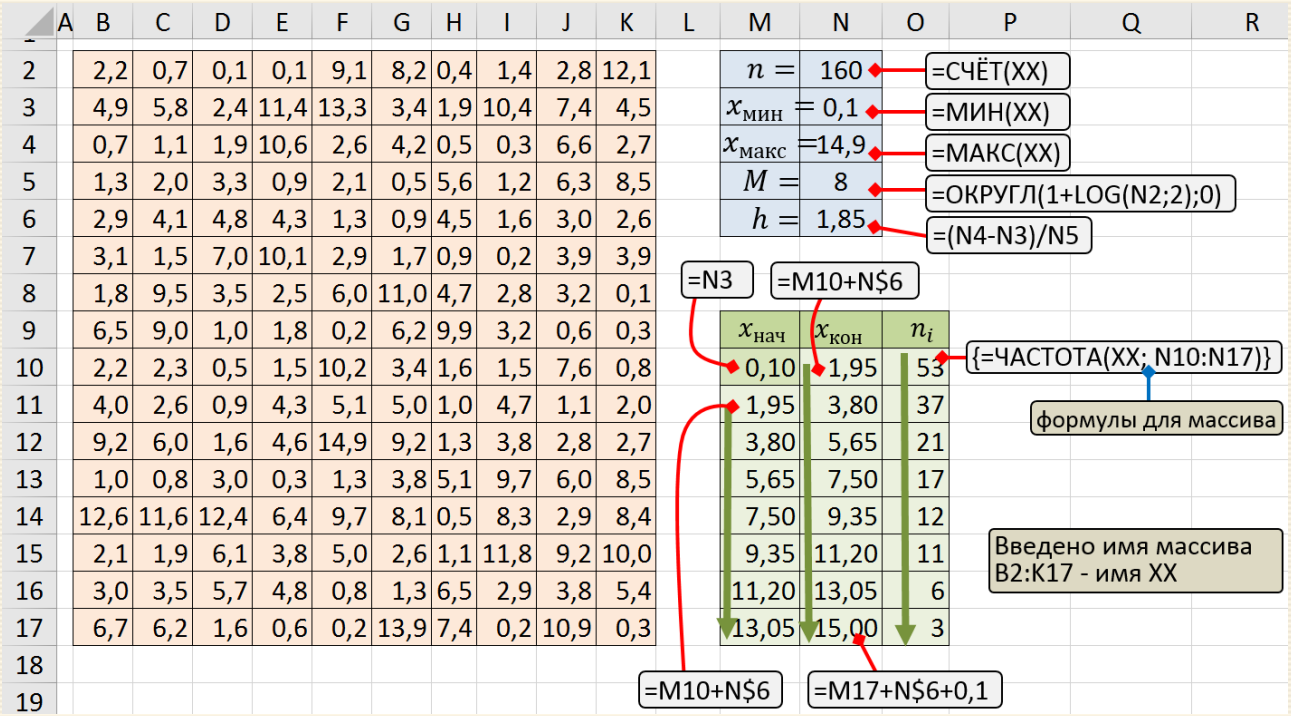
Если размах вариации признака в совокупности велик и значения признака варьируются неравномерно, то используют группировку с неравными интервалами. Неравные интервалы могут быть получены, если построенная группировка с равными интервалами содержит группы, не отражающие определенные типы изучаемого явления (процесса) или не содержащие ни одной единицы совокупности. В этом случае возникает необходимость увеличения размера интервала через объединение двух или более последовательных "пустых".

Выбор равных или неравных интервалов зависит от степени заполнения интервалов. Интервалы группировок могут быть закрытыми и открытыми. Закрытыми интервалами являются интервалы, в которых указаны верхняя и нижняя границы. Открытые интервалы имеют только одну границу (верхнюю – у первого, нижнюю – у последнего). Если в основании группировки лежит дискретный признак, то нижняя граница последующего интервала равна верхней границе предыдущего, увеличенную на 1.

Пример П2.1 Задача: привести выборку к форме таблицы классов.

2.2	0.7	0.1	0.1	9.1	8.2	0.4	1.4	2.8	12.1	2.2	2.3	0.5	1.5	10.2	3.4	1.6	1.5	7.6	0.8
4.9	5.8	2.4	11.4	13.3	3.4	1.9	10.4	7.4	4.5	4.0	2.6	0.9	4.3	5.1	5.0	1.0	4.7	1.1	2.0
0.7	1.1	1.9	10.6	2.6	4.2	0.5	0.3	6.6	2.7	9.2	6.0	1.6	4.6	14.9	9.2	1.3	3.8	2.8	2.7
1.3	2.0	3.3	0.9	2.1	0.5	5.6	1.2	6.3	8.5	1.0	0.8	3.0	0.3	1.3	3.8	5.1	9.7	6.0	8.5
2.9	4.1	4.8	4.3	1.3	0.9	4.5	1.6	3.0	2.6	12.6	11.6	12.4	6.4	9.7	8.1	0.5	8.3	2.9	8.4
3.1	1.5	7.0	10.1	2.9	1.7	0.9	0.2	3.9	3.9	2.1	1.9	6.1	3.8	5.0	2.6	1.1	11.8	9.2	10.0
1.8	9.5	3.5	2.5	6.0	11.0	4.7	2.8	3.2	0.1	3.0	3.5	5.7	4.8	0.8	1.3	6.5	2.9	3.8	5.4
6.5	9.0	1.0	1.8	0.2	6.2	9.9	3.2	0.6	0.3	6.7	6.2	1.6	0.6	0.2	13.9	7.4	0.2	10.9	0.3

На первом этапе определяются параметры выборки $n, x_{макс}, x_{мин}$, число классов (интервалов) по формуле Sturgess'a и диапазон размера класса h (ячейки N2-N6); формируются массивы границ классов ($x_{нач}$ – M10:M17) и ($x_{кон}$ – N10:N17). Скриншот решения дан на рис. П2.1.



2	2,2	0,7	0,1	0,1	9,1	8,2	0,4	1,4	2,8	12,1	
3	4,9	5,8	2,4	11,4	13,3	3,4	1,9	10,4	7,4	4,5	
4	0,7	1,1	1,9	10,6	2,6	4,2	0,5	0,3	6,6	2,7	
5	1,3	2,0	3,3	0,9	2,1	0,5	5,6	1,2	6,3	8,5	
6	2,9	4,1	4,8	4,3	1,3	0,9	4,5	1,6	3,0	2,6	
7	3,1	1,5	7,0	10,1	2,9	1,7	0,9	0,2	3,9	3,9	
8	1,8	9,5	3,5	2,5	6,0	11,0	4,7	2,8	3,2	0,1	
9	6,5	9,0	1,0	1,8	0,2	6,2	9,9	3,2	0,6	0,3	
10	2,2	2,3	0,5	1,5	10,2	3,4	1,6	1,5	7,6	0,8	
11	4,0	2,6	0,9	4,3	5,1	5,0	1,0	4,7	1,1	2,0	
12	9,2	6,0	1,6	4,6	14,9	9,2	1,3	3,8	2,8	2,7	
13	1,0	0,8	3,0	0,3	1,3	3,8	5,1	9,7	6,0	8,5	
14	12,6	11,6	12,4	6,4	9,7	8,1	0,5	8,3	2,9	8,4	
15	2,1	1,9	6,1	3,8	5,0	2,6	1,1	11,8	9,2	10,0	
16	3,0	3,5	5,7	4,8	0,8	1,3	6,5	2,9	3,8	5,4	
17	6,7	6,2	1,6	0,6	0,2	13,9	7,4	0,2	10,9	0,3	
18											
19											

$x_{нач}$	$x_{кон}$	n_i
0,10	1,95	53
1,95	3,80	37
3,80	5,65	21
5,65	7,50	17
7,50	9,35	12
9,35	11,20	11
11,20	13,05	6
13,05	15,00	3

Рис. П2.1. Скриншот задачи построения вариационного ряда

При выполнении Excel-функции ЧАСТОТА(...) количество элементов в возвращаемом массиве на единицу больше числа элементов в массиве интервалов для учета величин, превышающих верхнюю границу интервала (т.е. содержащую наибольшее значение).

Чтобы специально не анализировать максимальные значения выборки можно просто увеличить верхнюю границу последнего интервала (ячейка N17) на положительное малое (в данном случае на величину 0.1).

Для вычисления частот формула ячейки O10 тиражируется на диапазон O11: O17, после чего выполняется операция работы с массивами F2 и затем Ctrl + Shift + Enter.

Отметим, что график вариационного ряда (рис. П2.2) самостоятельно дает качественную информацию по распределению данных в выборке.

	$x_{нач}$	$x_{кон}$	$x_{средн}$	n_i
3	0,10	1,95	1,03	53
4	1,95	3,80	2,88	37
5	3,80	5,65	4,73	21
6	5,65	7,50	6,58	17
7	7,50	9,35	8,43	12
8	9,35	11,20	10,28	11
9	11,20	13,05	12,13	6
10	13,05	15,00	14,03	3

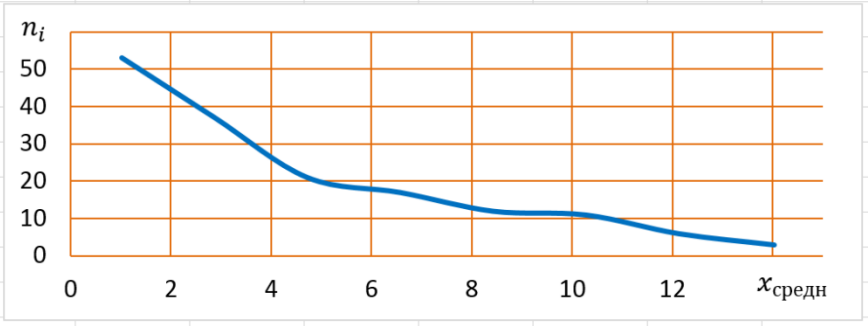


Рис. П2.2. График вариационного ряда

П2.2. Описательная статистика. Интервальный ранжированный частотный ряд

Описательная (дескриптивная) статистика (descriptive statistics) имеет дело с обработкой эмпирических данных и их наглядным представлением; количественной оценкой основных статистических показателей. Для выборки, заданной совокупностью элементов, целесообразно использовать инструментарий Пакет анализа MS Excel. В случае, если выборка задана в виде частотного ряда, для расчета показателей описательной статистики вычисления организуются с помощью встроенных функций электронных таблиц.

В состав Microsoft Excel входит надстройка Пакет анализа (вход через главное меню Данные), которая содержит статистические процедуры описательной статистики. Это средство является, по-видимому, наиболее часто используемым из всего пакета анализа, поскольку позволяет быстро и просто вычислять основные статистические характеристики одномерных выборок. Загрузка инструмента производится в соответствии с диалоговыми окнами рис. П2.3

На рис. П2.3 в качестве примера показано диалоговое окно средства Описательная статистика, где в области Входные данные и входной интервал указан диапазон ячеек, содержащий данные по выборке. При необходимости указывается, как сгруппированы выборочные данные (по столбцам или по строкам). Если входной диапазон данных задается с заголовками, то устанавливается флажок опции Метки в первой строке. Если заголовки отсутствует, то данным автоматически присваиваются заголовки Столбец1, Столбец2, … (или Строка1, Строка2, …).

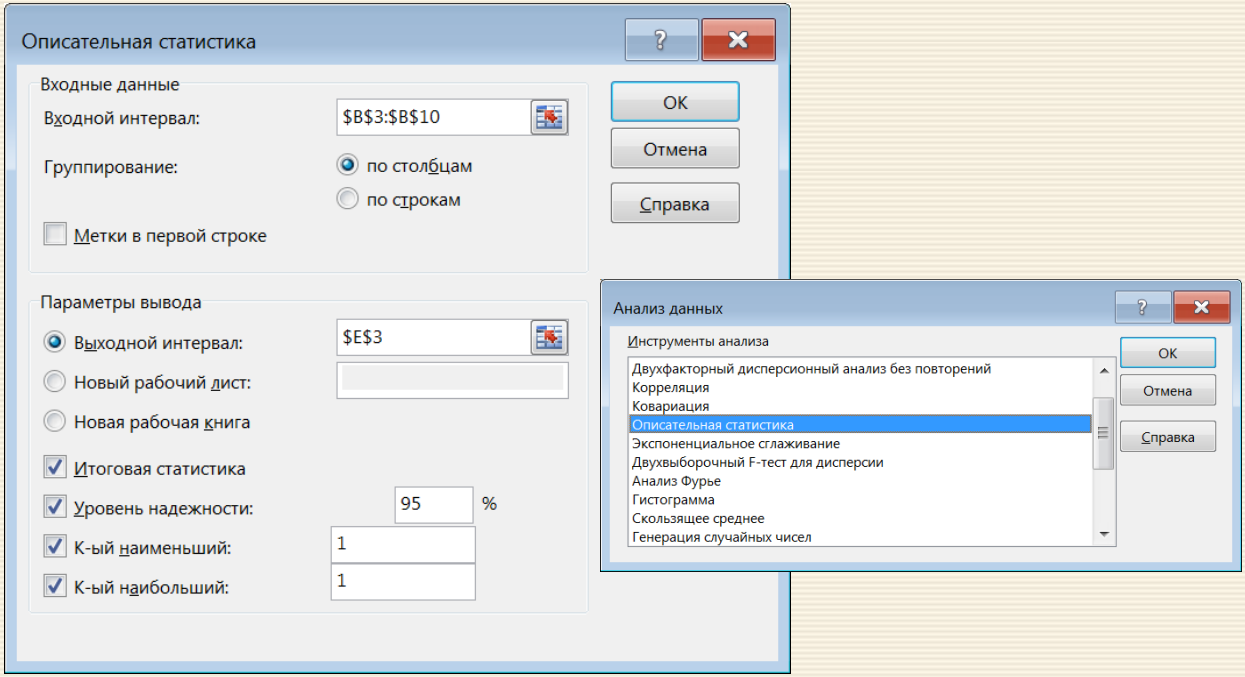


Рис. П2.3. Исходные данные и диалоговое окно Описательная статистика

В области Параметры вывода указывается область, где будут размещены результаты анализа. Предусмотрены возможности вывода на текущий рабочий лист (переключатель Выходной интервал), при этом необходимо указать выходной интервал (достаточно указать адрес одной ячейки, которая определяет верхний левый угол выходного диапазона).

Вывод также может выполнен на новый рабочий лист текущей рабочей книги начиная с ячейки A1 (переключатель Новый рабочий лист).

В соответствующие окна вводятся дополнительные параметры для проведения выбранной статистической процедуры. В частности, опция Уровень надёжности указывает, надо ли вычислять границу доверительного интервала для среднего. В поле ввода рядом с этой опцией задается доверительный уровень в процентах. Граница вычисляется с помощью распределения Стьюдента, т.е. здесь неявно используется предположение о нормальности распределения генеральной совокупности. Поэтому к данному показателю при малых выборках следует относиться с осторожностью.

В полях ввода рядом с опциями К-ый наибольший и К-ый наименьший указываются порядки выводимых наибольшего и наименьшего значений. Если эти порядки равны 1, то выводятся соответственно максимальное и минимальное выборочные значения.

На рис. П2.4 показан рабочий лист с результатами вычисления параметров описательной статистики инструментом Пакет анализа, а рядом справа приведены функции MS Excel, через которые можно получить соответствующие аналогичные значения.

Замечание. Поскольку в выборке отсутствуют повторяющиеся (модные) значения, то в соответствующих строках высвечивается ошибка #Н/Д.

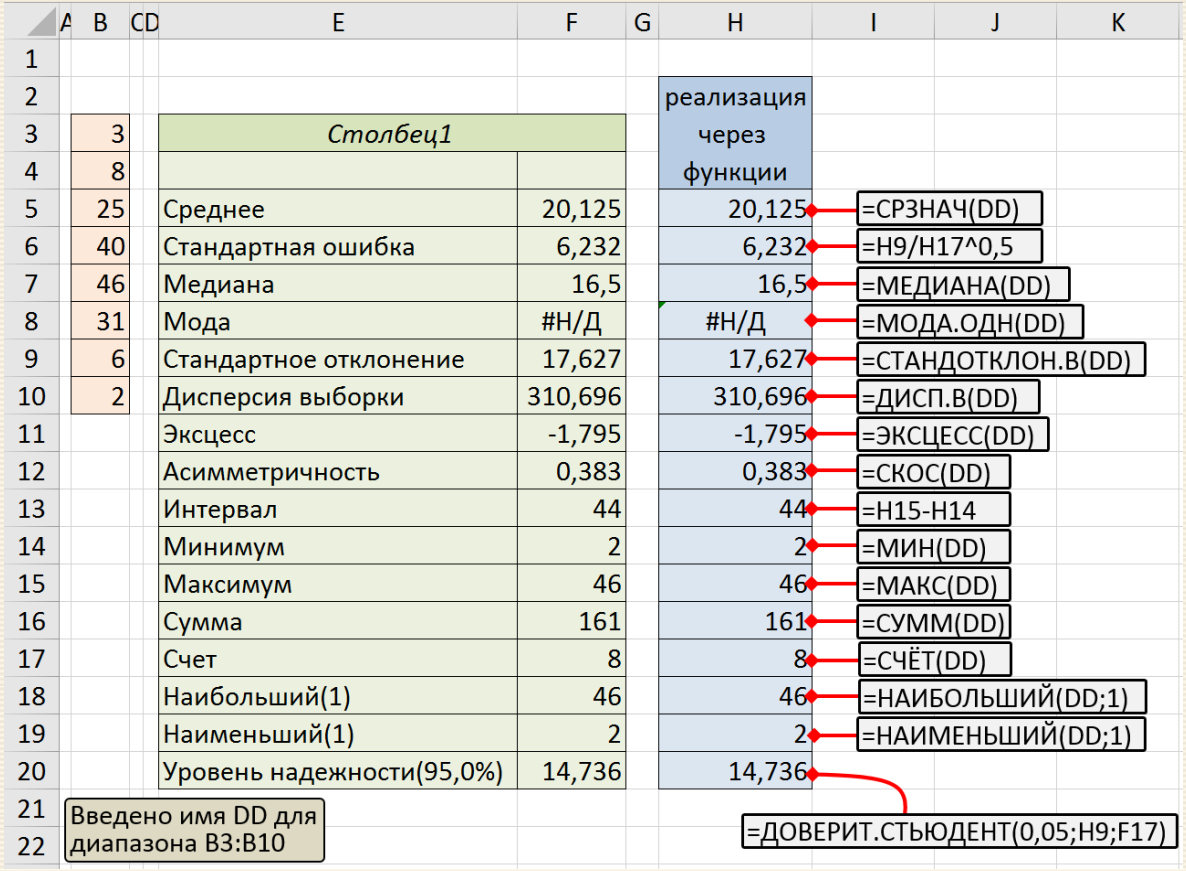


Рис. П2.4. Результаты вычисления параметров описательной статистики

Инструмент Гистограмма из Пакета анализа предназначен для предварительной оценки распределения выборки и построения столбиковых диаграмм эмпирических плотностей вероятностей. Исходными данными является входной диапазон выборочных значений и интервал карманов, определяющий границы столбцов гистограммы. Инструментарий Гистограмма подсчитывает число выборочных значений, попавших в каждый карман (Частота). Частоты могут суммироваться (подсчитываются кумулятивные суммы), которые в результатах выдаются под наименованием Интегральный процент, представляющий эмпирическую функцию распределения. Предусмотрена построение гистограммы по этим отсортированным частотам.

Диалоговое окно Гистограмма показано на рис. П2.5. В области Входные данные задаются адрес диапазона ячеек с выборочными значениями (поле ввода Входной интервал) и адрес диапазона, содержащего границы карманов (поле ввода Интервал карманов) в порядке возрастания. При подсчете количества попаданий выборочных значений в карманы в число попавших в данный карман включаются значения, равные нижней границе кармана и меньшие верхней границы кармана.

Если не указывать интервал границ карманов, будут автоматически созданы равновеликие интервалы, количество которых определяется по формуле Sturgess'a. Более подробно построение интервалов (классов) дано в примере П2.1. Пример вывода данных в этом случае приведен на рис. П2.6. Для заданного массива интервалов (карманов) на рис. П2.7 приведен скриншот результатов работы рассматриваемого средства.

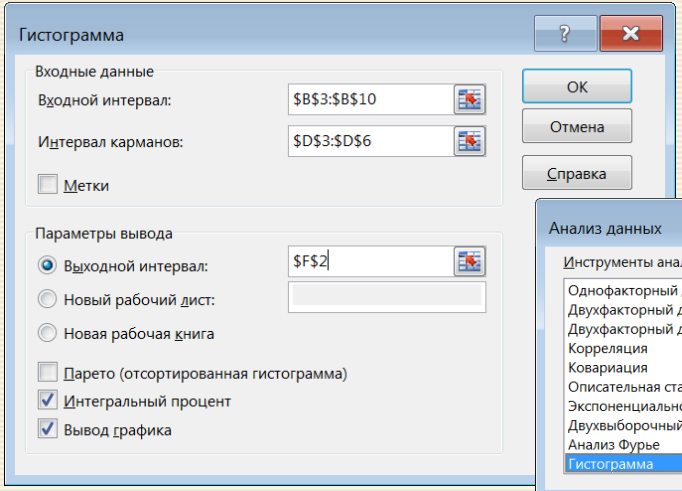


Рис. П2.5. Исходные данные и диалоговое окно Гистограмма

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3		3		*Карман*	*Частота*	*Интегр-й %*						
4		8		2	1	12,50%						
5		25		24	3	50,00%						
6		40		Еще	4	100,00%						
7		46										
8		31										
9		6										
10		2										

Рис. П2.6. Гистограмма с карманами, определенными формулой Sturgess'a

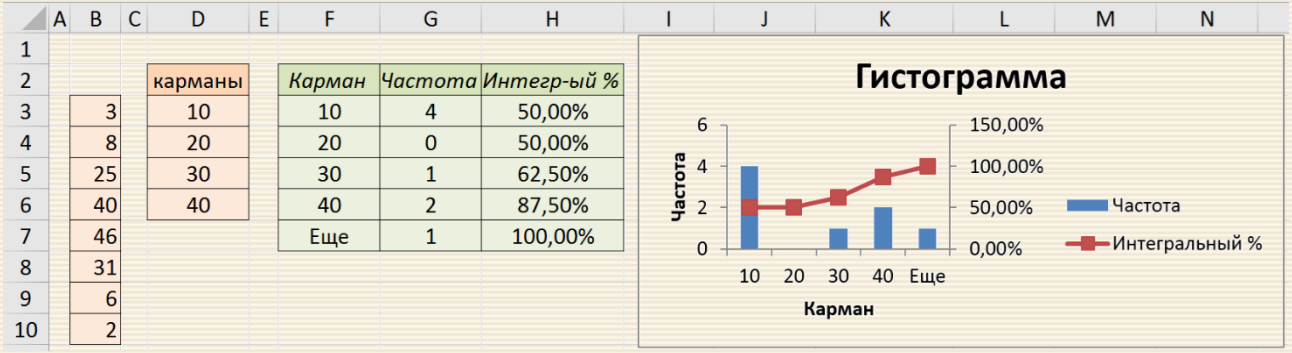


Рис. П2.7. Построение гистограммы с заданным массивом карманов

Если установлен флажок опции Парето (отсортированная гистограмма), то выводятся таблица частот и таблица отсортированных в порядке убывания частот. Если также установлен флажок опции Вывод графика, выводится гистограмма отсортированных частот, как показано на рис. П2.8. Если установлена опция Интегральный процент, то выводятся две таблицы: одна содержит неотсортированные частоты и интегральные проценты, вторая — отсортированные частоты и соответствующие интегральные проценты.

Диаграмма (кривая) Парето (Pareto) – графическое отражение закона Парето, кумулятивная зависимость распределения определённых ресурсов или результатов от большой совокупности (выборки) причин.

Закон Парето: 20% усилий дают 80% результата, а остальные 80% усилий – лишь 20% результата.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		3		карманы		*Карман*	*Частота*	*Интегр-ый %*	*Карман*	*Частота*	*Интегр-ый %*
3		8		10		10	4	50,00%	10	4	50,00%
4		25		20		20	0	50,00%	40	2	75,00%
5		40		30		30	1	62,50%	30	1	87,50%
6		46		40		40	2	87,50%	Еще	1	100,00%
7		31				Еще	1	100,00%	20	0	100,00%
8		6									
9		2									
10											
11											
12											
13											
14											
15											
16											
17											

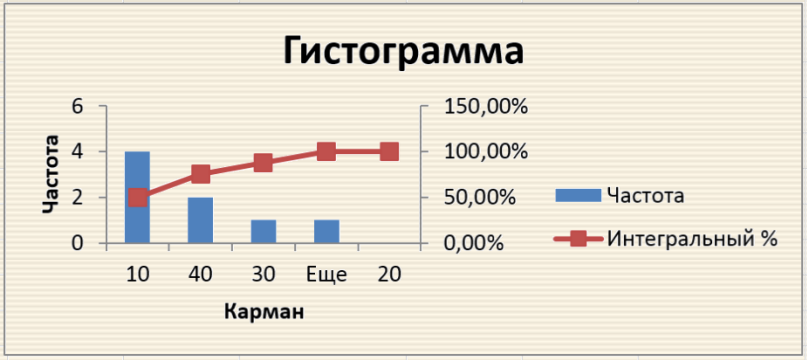


Рис. П2.8. Построение гистограммы Парето

Для достаточно широкого спектра исследовательских задач исходные данные для статистической оценки задаются интервальным ранжированным частотным рядом.

> Р а н ж и р о в а н н ы й р я д – это распределение отдельных единиц совокупности в порядке возрастания или убывания исследуемого признака.

Например, данные наблюдения могут быть представлены в виде ранжированного распределения особей по классам возрастов с указанием частот встречаемости (в виде вариационного ряда)

Возраст особи		частота встречаемости
лет от	лет до	
18	21	1
21	24	3
24	27	6
27	30	10
30	33	5
33	36	3
36	39	2

Требуется вычислить статистические оценки данных (в списке стандартных функций электронных таблиц необходимые средства для данной формы исходных данных отсутствуют). А именно – найти среднее, медиану, моду и параметры вариации.

Медиана (*Me*) – это величина, которая соответствует варианту, находящемуся в середине ранжированного ряда.

Для ранжированного ряда с нечетным числом индивидуальных величин (например, 1, 2, 4, 4, 6, 7, 8, 8, 10) медианой будет величина, которая расположена в центре ряда, т.е. пятая величина.

Для ранжированного ряда с четным числом индивидуальных величин (например, 1, 5, 7, 10, 11, 14) медианой будет средняя арифметическая величина, которая рассчитывается из двух смежных величин. В данном случае медиана равна (7+10) : 2= 8.5.

Таким образом, для нахождения медианы сначала необходимо определить ее порядковый номер (ее положение в ранжированном ряду) по формуле

$$N_{Me} = \frac{n+1}{2},$$ где n – объем совокупности.

Численное значение медианы определяют по накопленным частотам дискретного вариационного ряда, для чего требуется указать интервал нахождения медианы в интервальном ряду распределения. Медианным называют первый интервал, где сумма накопленных частот превышает половину наблюдений от общего числа всех наблюдений.

Численное значение медианы определятся по формуле

$$Me = x_{Me} + \Delta \frac{\frac{n+1}{2} - S_{-1}}{f_{Me}},$$

где x_{Me} – нижняя граница медианного интервала;
Δ – величина интервала;
S_{-1} – накопленная частота интервала, которая предшествует медианному;
f_{Me} – частота медианного интервала.

Модой (*Mo*) называют значение признака, которое встречается наиболее часто у единиц совокупности. Для дискретного ряда модой будет являться вариант с наибольшей частотой. Для определения моды интервального ряда сначала определяют модальный интервал (интервал, имеющий наибольшую частоту). Затем в пределах этого интервала находят то значение признака, которое может являться модой.

Чтобы найти конкретное значение моды, необходимо использовать формулу

$$Mo = x_{Mo} + \Delta \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}$$

где x_{Mo} – нижняя граница модального интервала;

f_{Mo} – частота модального интервала;

f_{Mo-1} – частота интервала, предшествующего модальному;

f_{Mo+1} – частота интервала, следующего за модальным.

Размах вариации R. Это самый доступный по простоте расчета абсолютный показатель, который определяется как разность между самым большим и самым малым значениями признака у единиц данной совокупности:

Вычисление размаха количественной вариации (выборки) производится по формуле:

$$R = x_{max} - x_{min}$$

где x_{max} – значение максимальной варианты,

x_{min} – значение минимальной варианты выборки.

Функции MS Excel: МИН(данные); МАКС(данные).

Размах вариации (размах разброса данных) – важный показатель, но только крайних отклонений. Для более точной характеристики рассеяния вариации признака используются другие показатели.

Среднее отклонение (выборочная оценка среднего отклонения), подобно стандартному отклонению, характеризует разброс эмпирической выборки относительно среднего значения и вычисляется по формуле

$$\overline{d} = \frac{1}{n}\sum_i |x_i - \overline{x}|,$$

n – численность выборки,

x_i – значения вариант выборки.

Выборочное среднее значение $\overline{x}$ определяется очевидным соотношением

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i}.$$

Среднее отклонение отражает так называемый модульный подход к вычислению меры отклонения между величинами в противоположность тому, что стандартное отклонение отражает квадратический подход.

Функция MS Excel: СРОТКЛ(данные);

Частотно-взвешенная оценка среднего отклонения вычисляется соотношением

$$\overline{d} = \frac{\sum(|x_i - \overline{x}|f_i)}{\sum f_i}.$$

При использовании показателя среднего линейного отклонения возникают определенные неудобства, связанные с расчетом модуля разности. Широкое распространение получили обобщающие показатели, найденные с использованием вторых степеней отклонений.

В предположении принадлежности выборки генеральной совокупности к таким показателям относятся несмещенная оценка дисперсии D. Частотно-взвешенная оценка дисперсии вариационного ряда вычисляется соотношением

$$D = \frac{\sum[(x_i - \overline{x})^2 f_i]}{\sum f_i} = \frac{\sum x_i^2 f_i}{\sum f_i} - \overline{x}^2, \quad S = \sqrt{D}.$$

Кроме показателей вариации, выраженных абсолютными величинами, в статистических исследованиях используются относительные показатели вариации V_*, в частности для сравнения разброса признаков нескольких совокупностей.

Данные показатели рассчитываются как отношение размаха вариации к средней величине признака (коэффициент осцилляции), отношение среднего линейного отклонения к средней величине признака (линейный коэффициент вариации), отношение среднего квадратического отклонения к средней величине признака (коэффициент вариации) и, как правило, выражаются в процентах.

Формулы расчета относительных показателей вариации (в процентах):

– коэффициент осцилляции V_R $$V_R = \frac{R}{\overline{x}} \cdot 100\%,$$

– линейный коэффициент вариации V_A $$V_A = \frac{\overline{d}}{\overline{x}} \cdot 100\%,$$

– коэффициент вариации V $$V = \frac{S}{\overline{x}} \cdot 100\%.$$

Отметим следующее: из приведенных формул видно, что чем больше коэффициент V_* ближе к нулю, тем меньше вариация значений признака.

В статистической практике наиболее часто применяется коэффициент вариации V. Он используется не только для сравнительной оценки вариации, но и для характеристики однородности совокупности. Напоминание: для распределений, близких к нормальному, совокупность считается однородной, если коэффициент вариации не превышает 33%.

Пример П2.2 Определить среднее, медиану, моду и параметры вариации по исходным данным, приведенным на рис. П2.3 в диапазоне C3:D11.

Расчет описанных статистических параметров для интервального ранжированного частотного ряда можно выполнить по следующей схеме.

1. В ячейке F5 формулой =(B5+C5)/2 определяется средний по диапазону возраст. Далее автозаполнение на диапазон F6:F11.

2. Формулой =СУММ(D$5:D5) в ячейке G5 находится накопленные частоты. Ячейка протягивается до ячейки G11.

3. В соответствии с нижеприведенной таблицей и формулами рис. П2.3 заполняются ячейки, рассчитывающие необходимые величины.

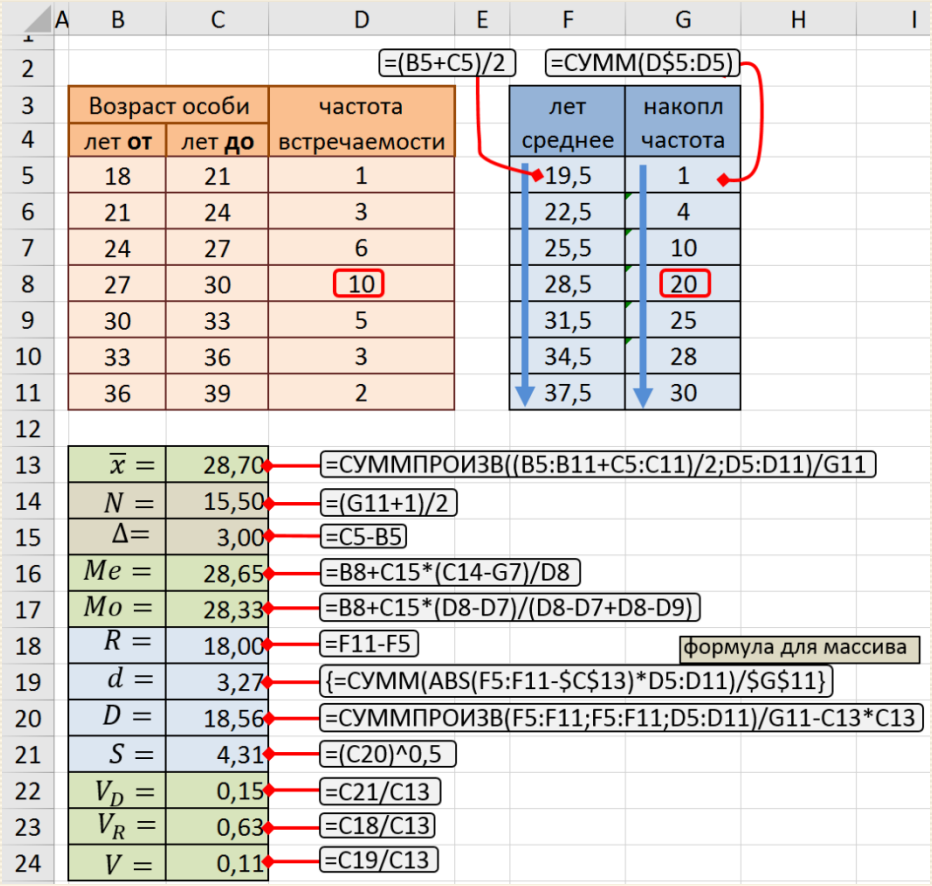


Рис. П2.3 Скриншот схемы вычислений

адрес ячейки	Описание	Формула для вычислений
B13	Подсчитывается среднее по формуле	$\overline{x} = \frac{\sum x_i f_i}{\sum f_i}$
B14	Определяется порядковый номер медианы N_{Me}	$N_{Me} = \frac{n+1}{2}$ где $n = \sum x_i$
B16	Рассчитывается значение медианы. По величине N_{Me} определяется номер строки, где значение накопленной частоты "содержит" N_{Me}. В данном примере это строка 8, соответствующее накопленная частота отмечена красной рамкой.	$Me = x_{Me} + \Delta \frac{\frac{n+1}{2} - S_{-1}}{f_{Me}}$
B17	Рассчитывается значение моды. По максимальному значению частоты в столбце D определяется номер строки, по положению которой в ряду распределения рассчитывается мода. Ячейка отмечается красной рамкой.	$Mo = x_{Mo} + \Delta \frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}}$
B18	Рассчитывается размах вариации. Для ранжированного ряда это весьма просто	$R = x_{max} - x_{min}$
B19	Вычисляется частотно-взвешенная оценка среднего отклонения	$\overline{d} = \frac{\sum(\lvert x_i - \overline{x}\rvert f_i)}{\sum f_i}$

B20	Считается дисперсия выборки	$D = \frac{\sum x_i^2 f_i}{\sum f_i} - \overline{x}^2$
B21	Среднеквадратическое отклонение	$S = \sqrt{D}$
B22	Коэффициент вариации	$V = \frac{S}{\overline{x}} \cdot 100\%$
B23	Коэффициент осцилляции	$V_R = \frac{R}{\overline{x}} \cdot 100\%$
B24	Линейный коэффициент вариации	$V_A = \frac{\overline{d}}{\overline{x}} \cdot 100\%$

Приложение П3. Нормальное распределение

Для проведения статистических расчетов часто необходимо располагать информацией о виде функции распределения. Наиболее важным из них является нормальное распределение. С ним связаны распределения хи-квадрат, Стьюдента, Фишера, а также интеграл вероятностей. Для указанных законов функции распределения аналитически в простой форме не представимы и определяются с использованием стандартных процедур прикладных программ. Нормальное распределение занимает важнейшее место в статистике, поскольку очень многие эмпирические распределения признаков приближаются к нему.

Р а с п р е д е л е н и е в е р о я т н о с т е й – это закон, описывающий область значений случайной величины и вероятности их исхода (появления). Функция распределения в теории вероятностей – функция, характеризующая распределение случайной величины. Другими словами – вероятность того, что случайная величина примет значение, меньшее или равное x, где x – произвольное действительное число.

Н о р м а л ь н о е р а с п р е д е л е н и е, также называемое распределением Гаусса – распределение вероятностей, которое в одномерном случае задаётся функцией плотности вероятности, совпадающей с функцией Гаусса:

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

В формуле распределения параметр μ – математическое ожидание (среднее в пределе значение), медиана и мода распределения, а параметр σ – среднеквадратическое (стандартное) отклонение (σ^2 – дисперсия) распределения (здесь использованы "классические" обозначения). Отметим, что одномерное нормальное распределение является двухпараметрическим семейством распределений.

Любая случайная величина имеет функцию распределения – зависимость плотности вероятности от значения случайной величины. Для нормального распределения (распределения Гаусса) функция распределения имеет следующий вид (рис. П3.1):

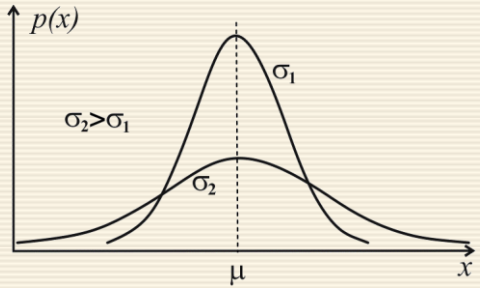


Рис. П3.1. Плотность вероятности нормального распределения

Обозначение	$N(\mu, \sigma^2)$
Математическое ожидание	μ
Медиана	μ
Мода	μ
Дисперсия	σ^2
Коэффициент асимметрии	0
Коэффициент эксцесса	0

Johann Carl Friedrich Gauß

Размещение вариант в вариационном ряду при нормальном распределении характеризуется определенными закономерностями. В частности, в нормальной кривой отклонения от средней арифметической охватывают приблизительно 6σ: три сигмы справа от средней и столько же слева.

Зная вариационную кривую распределения вариант по тому или иному признаку и предполагая, что распределение является нормальным, можно заранее предсказать, какой процент изученных особей (или вариант) укладывается в пределах $\pm 1\sigma$; в пределах $\pm 2\sigma$; в пределах $\pm 3\sigma$ (рис. П3.2). А именно, в пределах $\pm 1\sigma$ располагается 68.3% всех вариант данного ряда; $\pm 2\sigma$ – 95.5% и в пределах $\pm 3\sigma$ находится 99.7% всех вариант.

Например, при исследовании размеров раковин моллюсков рода *Benedictia* были получены следующие данные (мм): 32; 35; 37; 45; 41; 35; 39; 39; 45; 41. "Перенос" данных на генеральную совокупность позволяет предположить (рис. П3.3), что ≈ 68% популяции будут иметь средний размер раковины плюс-минус одно стандартное отклонение $\overline{x} \pm 1\sigma$, то есть размеры раковин будут лежать в интервале от 34.62 до 43.18 мм.

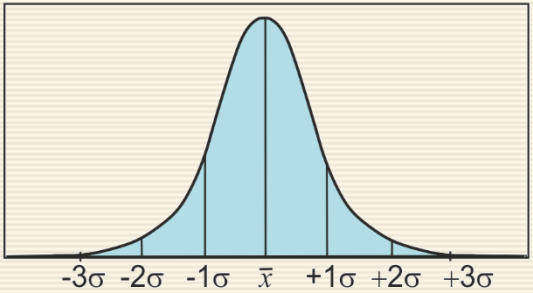


Рис. П3.2. Нормальная вариационная кривая

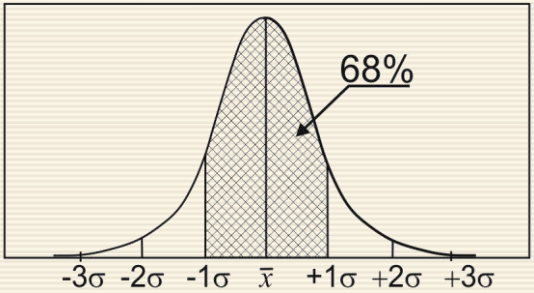


Рис. П3.3. Нормальное распределение значений размеров раковин у моллюсков *Benedictia* вокруг среднего

Около 98% популяции будет иметь размер раковины 38.9 плюс-минус два стандартных отклонения (8.56), то есть размер раковины будет лежать в интервале (30.34÷47.46) мм, и практически 100% будут лежать в интервале плюс-минус три стандартных отклонения от 38.9.

Нормальное распределение является наиболее важным в связи с центральной предельной теоремой теории вероятностей: распределение суммы независимых случайных величин стремится к нормальному с увеличением их количества при произвольном законе распределения отдельных слагаемых, если слагаемые обладают конечной дисперсией.

Как правило, для аналитических измерений условие теоремы выполняется, поэтому для результатов физического и химического анализа обычно постулируется нормальное распределение.

Центральная предельная теорема Чебышева: Если случайная величина подвержена воздействию бесконечного числа бесконечно малых случайных факторов, то она имеет нормальное распределение.

Так как реальные физические, химические, биологические явления часто представляют собой результат суммарного воздействия многих факторов, то в таких случаях нормальное распределение является хорошим приближением наблюдаемых значений. Для известных выборочных значений среднего $\overline{x}$ и стандартного отклонения S функция плотности нормального распределения

$$p(x) = \frac{1}{S\sqrt{2\pi}} \exp\left(-\frac{(x-\overline{x})^2}{2S^2}\right)$$

– унимодальная и симметричная, а аргумент x может принимать любые действительные значения.

Чебышев Пафнутий Львович

Здесь же необходимо упомянуть закон больших чисел, гарантирующий устойчивость средних значений некоторых случайных событий для достаточно длинной серии экспериментов.

Предсказать результат испытания конкретного испытания, когда реализуется конкретно-определенное значение случайной величины, невозможно в силу именно в случайности события. Тем не менее, при многократных испытаниях проявляются обладающие устойчивостью закономерности массовых событий и явлений, когда средние характеристики групповых испытаний стремятся к некоторому значению при увеличении (в принципе до бесконечности) числа испытаний. Другими словами – для большого числа параллельных (выполненных в одинаковых условиях) опытов исход каждого из них случаен и не определён, а средний результат серии уже далеко не случаен и закономерен.

Данный факт рассматривается в ряде теорем (Якоба Бернулли, С.Д. Пуассона, П.Л. Чебышёва, А.А. Маркова, А.Н. Колмогорова и др.), которые носят общее название закона больших чисел.

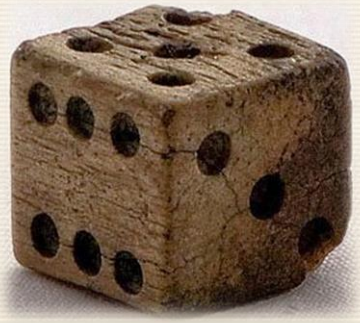
> Закон больших чисел – это обобщенное название ряда теорем, из которых следует, что при неограниченном увеличении числа испытаний средние величины стремятся к некоторым постоянным.

Достаточно удобной формулировкой является Борелевский закон больших чисел, названный в честь Эмиля Бореля. Закон гласит, что если эксперимент повторяется много раз независимо при одинаковых условиях, то число благоприятных реализаций любого конкретного события происходит соответственно (приблизительного равному) вероятности проявления события. При этом, чем больше число повторений, тем лучше приближение.

То есть, если E обозначает конкретное событие, p вероятность его появления, а N_n – число реализаций E в первых n испытаниях, тогда достоверно, что

$$\frac{N_n(E)}{n} \to p, \quad n \to \infty .$$

"Почувствовать" закон больших чисел можно на моделировании бросков шестигранной игральной кости, на которой равновероятно может выпасть одно из чисел 1, 2, 3, 4, 5 или 6. Согласно закону больших чисел, при большом количестве бросков среднее значение выпадающих очков будет стремиться к величине $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$, при этом точность (близость среднего арифметического к значению 3.5) будет возрастать по мере увеличения числа бросков.



На рис. П3.4 дан скриншот листа электронных таблиц, где моделируются результаты бросания кости с помощью функции MS Excel =1+ОКРУГЛ(5*СЛЧИС();0). Отметим, что в если функции ОКРУГЛ() число разрядов равно нулю, то значение округляется до ближайшего целого.

Выпадающее количество очков при "бросках" заносится в диапазон данных B2:AK15 (рис. П3.4); в столбцах AM и AN рассчитывается количество данных для области и среднее значение выпадающего числа очков для рассматриваемого количества данных.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1																																							
2		3	4	6	5	2	6																															6	4,33
3		1	2	5	2	4	1	3	1	5	2	2	5																									18	3,28
4		2	1	2	6	5	4	5	6	3	4	3	5	1	2	5	4	5	5	4	2	6	1	3	6													42	3,55
5		5	2	1	3	3	5	2	5	3	6	3	4	2	2	3	4	5	2	4	4	5	3	3	5	4	1	5	3	4	4							72	3,53
6		1	2	3	4	6	1	1	3	5	2	4	3	5	5	2	5	2	2	5	4	6	5	6	4	1	3	4	6	3	4	4	5	2	3	2	3	108	3,52
7		3	4	1	2	3	3	1	5	2	6	5	1	5	2	3	1	1	2	5	5	2	2	4	3	2	5	2	1	4	1	2	6	1	1	3	6	144	3,37
8		4	2	5	1	5	2	6	1	5	1	2	5	4	4	5	2	1	2	5	6	2	5	3	2	4	3	4	3	4	4	3	6	6	4	3	3	180	3,40
9		5	2	3	3	4	5	1	1	5	3	4	4	3	6	3	4	1	4	3	5	1	3	2	2	1	3	2	4	5	5	1	2	2	4	5	5	216	3,37
10		5	5	5	4	3	2	5	1	5	5	3	3	2	3	4	1	6	3	2	3	6	2	1	2	2	5	4	6	1	3	4	2	1	3	4	4	252	3,37
11		2	1	5	3	5	3	3	2	1	2	1	6	3	5	2	2	2	1	3	3	1	1	5	2	5	2	1	5	2	2	3	1	5	2	6	4	288	3,30
12		4	3	3	3	3	5	4	1	3	6	6	4	3	4	1	2	6	6	1	3	4	4	1	5	3	4	4	3	2	6	4	4	5	3	3	2	324	3,33
13		2	6	2	5	4	6	2	3	4	3	6	3	2	5	3	3	2	4	5	5	3	3	2	1	5	6	5	6	3	1	2	4	5	4	3	6	360	3,37
14		1	6	2	4	3	3	5	3	3	4	6	5	4	2	3	4	4	5	1	5	5	3	5	3	2	6	2	5	2	6	2	2	5	5	4	3	396	3,40
15		6	1	5	5	5	5	2	4	6	1	6	5	3	4	5	5	6	3	6	3	4	5	6	1	5	6	4	6	3	4	2	5	5	5	2	6	432	3,47

Рис. П3.4. Средние значения выпадающего числа очков в зависимости от количества бросков

На рис. П3.5 приведен график зависимости среднего значения выпадающего числа очков от количества бросков.

На рис. П3.6 дана зависимость частости (относительной частоты) выпадения 4 очков от общего количества бросков кости.

Рис. П3.5. Среднее значение выпадающего числа очков от количества бросков кости

Рис. П3.6. Зависимость частости выпадения 4 очков от общего количества бросков кости

Последствия нарушения нормального закона распределения.

1) Нарушаются условия теоремы (выделяется более весомая группа факторов). Например, при анализе высокочистых веществ наблюдается неравномерное распределение примесей.

2) Произвольное объединение нескольких выборок даже если каждая из них подчинялась распределению Гаусса (рис. П3.7).

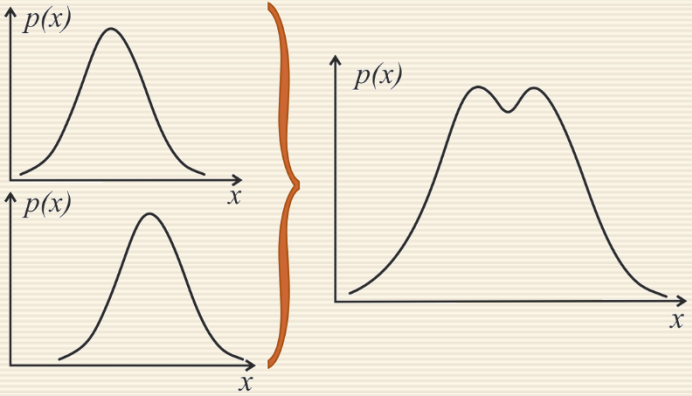


Рис. П3.7. Нормальная вариационная кривая

3) Косвенные измерения. Линейная комбинация нормально распределенных случайных величин также будет подчиняться нормальному распределению. А нелинейная комбинация (например, произведение двух величин) не сохранит нормального распределения. Однако, чем меньше погрешность, тем меньше отличие от нормального распределения, поэтому даже для нелинейных преобразований в некоторых случаях можно принять нормальное распределение.

4) Результат измерения является дискретной величиной (например, некоторые радиоактивационные методы анализа, ряд биохимических и рентгеноспектральных методов – так называемые счетные методы).

В этом случае результат измерения подчиняется распределению Пуассона*. При больших λ распределение Пуассона переходит в нормальное распределение.

С т а н д а р т и з и р у ю щ е е п р е о б р а з о в а н и е . Если случайная величина x имеет известное математическое ожидание μ и дисперсию σ^2 то случайная величина $y = x - \mu$ называется центрированной, величина $y = x/\sigma$ – нормированной, а $y = (x - \mu)/\sigma$ – стандартизированной. Последнее преобразование называется стандартизирующим и используется на практике для получения стандартизированных выборок (в качестве значений μ и σ обычно берутся выборочные среднее и стандартное отклонение). Для выполнения этого преобразования в Excel есть специальная функция НОРМАЛИЗАЦИЯ(x; $\overline{x}$; σ).

Можно и уместно отметить, что стандартизирующее преобразование не изменяет тип распределения, а изменяет только значения математического ожидания и дисперсии.

Функция плотности нормального распределения стандартизованной величины u имеет вид

$$P(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

На рисунке П3.8. приведен скриншот графика плотности вероятности нормального распределения.

* Для фиксированного числа $\lambda > 0$ дискретное распределение, задаваемое функцией вероятности $p(k) = \lambda^k e^{-\lambda}/k!$, называется распределением Пуассона.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
3	σ	x	-3,00	-2,75	-2,50	-2,25	-2,00	-1,75	-1,50	-1,25	-1,00	-0,75	-0,50	-0,25	0,00	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75	3,00
4	1	$p(x)$	0,00	0,01	0,02	0,03	0,05	0,09	0,13	0,18	0,24	0,30	0,35	0,39	0,40	0,39	0,35	0,30	0,24	0,18	0,13	0,09	0,05	0,03	0,02	0,01	0,00
5	2	$p(x)$	0,04	0,05	0,07	0,09	0,11	0,13	0,16	0,19	0,21	0,23	0,25	0,26	0,27	0,26	0,25	0,23	0,21	0,19	0,16	0,13	0,11	0,09	0,07	0,05	0,04

=НОРМ.РАСП(C3; 0; 1;)

=НОРМ.РАСП(C3; 0; 1,5;)

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

Рис. П3.8. Плотность вероятности нормального распределения

П4.1. Округление чисел

При обработке результатов расчеты проводят как на калькуляторах, так на компьютерах, часто делая при этом ошибки, связанные с "магией цифр". Например, результат измерения записывается в виде: x = 1.7469432$\pm$0.025. При ошибке 0.025 последние цифры числа ничего не означают и фраза "так точнее" свидетельствует о математической безграмотности.

Если измерения проводились прибором, дающим значимые цифры до сотых долей, то грамотная запись результата должна быть 1.75$\pm$0.03. Таким образом, всегда нужно производить необходимые округления, чтобы не было ложного впечатления о большей, чем это есть на самом деле, точности результатов эксперимента.

При обработке результатов руководствуются следующими правилами округления:

- Погрешность измерения округляют до первой значащей цифры, всегда увеличивая ее на единицу.

  Примеры: 8.27≈8; 0.237≈0.3; 0.0862≈0.09; 0.00035≈0.0004; 857.3≈900; 43.5≈50.

- Результаты измерения округляют с точностью "до погрешности", то есть последняя значащая цифра в результате должна находиться в том же разряде, что и погрешности.

  Примеры: 243.871$\pm$0.026≈243.87$\pm$0.03; 243.871$\pm$2.6≈244$\pm$3; 1053$\pm$ 47≈1050$\pm$50.

- Округление результата измерения достигается простым отбрасыванием цифр, если первая из отбрасываемых цифр меньше 5.

  Примеры: 8.337 (округлить до десятых) ≈ 8.3; 833.438 (округлить до целых) ≈ 833; 0.27375 (округлить до сотых) ≈ 0.27.

■ Если первая из отбрасываемых цифр больше или равна 5 (а за ней одна или несколько цифр отличны от нуля), то последняя из остающихся цифр увеличивается на единицу.

Примеры: 8.3351 (округлить до сотых) ≈ 8.34; 0.2510 (округлить до десятых) ≈ 0.3; 271.515 (округлить до целых) ≈ 272.

■ Если отбрасываемая цифра равна 5, а за ней нет значащих цифр (или стоят одни нули), то последнюю оставляемую цифру увеличивают на единицу, когда она нечетная, и оставляют неизменной, когда она четная.

Примеры: 0.875 (округлить до сотых) ≈ 0.88; 0.5450 (округлить до сотых) ≈ 0.54; 275.500 (округлить до целых) ≈ 276.

Примечание. Значащими называют верные цифры числа, кроме нулей, стоящих впереди числа. Например, 0.00402 – в этом числе имеется три значащих цифры; 4, 0, 2, а первые три нуля незначащие.

Важно помнить, что точность результата определяется точностью исходных данных. Если одна из составляющих исходных данных имеет большую погрешность, чем другие, то и результат будет иметь такую же погрешность, несмотря на более точные другие составляющие. А для всех величин, входящих в промежуточные расчеты, число цифр после запятой должно быть на одну больше числа значащих цифр исходных данных.

Запись и округление результата измерения

Доверительный интервал значения или погрешность результата при прямых измерениях рассчитывается по случайной выборке, что можно трактовать как погрешность. Результат измерения также содержит лишь ограниченное число значащих цифр, несущих информацию о величине этого результата. В связи с этим числовые значения результата (погрешности) должны быть округлены. При округлении используют следующие правила:

1. Предварительно результат и погрешность записывают в нормальном виде: общий показатель степени выносят за скобку или заменяют соответствующей приставкой: микро, милли, кило, мега и др. Например, x = 0.22 ± 0.03 м = (22 ± 3)·10^{-2} м = 22 ± 3 см.

Запрещены записи вида x = 22·10^{-2} ± 30·10^{-3} м или x = 0.22 ± 3·10^{-2} м.

Показатель 10^1 не выносится.

2. Среднее значение x округляют до того разряда, которым оканчивается результат или округленная погрешность Δx:

Неокругленный результат	Округленный результат
1237.2 ± 32	(12.4 ± 0.3)·10^2
(7.854 ± 0.0476) ·10–3	(7.85 ± 0.05)·10^{-3}
83.2637 ± 0.0126	83.264 ± 0.013
2.48 ± 0.931	2.5 ± 0.9
2.48 ± 0.96	2.5 ± 1.0

Если погрешность округляется до двух значащих цифр, но вторая из них равна нулю, то этот нуль сохраняется, а в соответствующем ему разряде результата записывается получающаяся там значащая цифра: $x = 3.48 \pm 0.10$.

Наиболее употребительные масштабирующие приставки приведены в таблице.

десятичный множитель	Приставка		Обозначение		Пример
	русская	междуна-родная	русское	междуна-родное	
10^{-9}	нано	nano	н	n	нм — нанометр
10^{-6}	микро	micro	мк	μ	мкм — микрометр, микрон
10^{-3}	милли	milli	м	m	мН — миллиньютон
10^{-2}	санти	centi	с	c	см — сантиметр
10^{-1}	деци	deci	д	d	дм — дециметр
10^{1}	дека	deca	да	da	дал — декалитр
10^{2}	гекто	hecto	г	h	гПа — гектопаскаль
10^{3}	кило	kilo	к	k	кН — килоньютон
10^{6}	мега	mega	М	M	МПа — мегапаскаль
10^{9}	гига	giga	Г	G	ГГц — гигагерц
10^{12}	тера	tera	Т	T	ТВ — теравольт

П4.2. Округления чисел в MS Excel

Простое округление по математическим правилами не всегда удовлетворяет в части представления результатов вычислений – отображения желаемого количества знаков после запятой, для чего в Excel имеются специальных функции. Более того, встречаются совсем малограмотные пользователи, которые формат чисел (видимое количество знаков после запятой) принимают за округление. Данные команды вызывается либо с ленты *Главная* → *Число*, либо через правую кнопку *Формат ячеек* (или нажатие на клавиатуре <Ctrl+1>); вид панели приведен на рис. П4.1.

Первая вкладка Число (открыта на рис. П4.1) задает числовой формат. То есть информацию можно представить, как обычное число, проценты и т.д. По умолчанию всем ячейкам придается Общий формат, то есть ячейка никак не отформатирована. В данном окне можно перейти к другому формату (рис. П4.1.а).

Следующий пункт Числовой. Здесь задается количество видимых знаков после запятой (по умолчанию их два), а также можно отделить группы разрядов (тысячи, миллионы и т.д.) друг от друга. Здесь все интуитивно понятно: указывается количество знаков после запятой и наслаждаемся внешним видом. Однако вид никак не влияет на точность числа в ячейке. Поэтому не стоит надеяться на формат, когда нужно реальное округление.

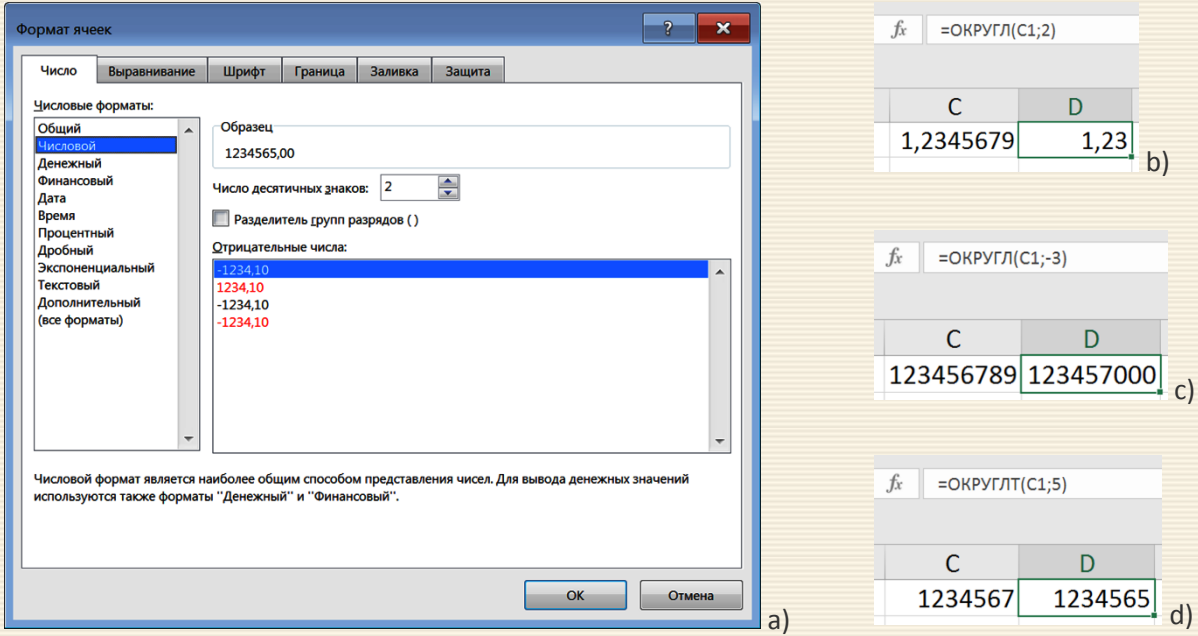


Рис. П4.1. Панель формата и функции округления чисел

ОКРУГЛ Для "настоящего" округления числа по математическим правилам существует функция ОКРУГЛ. С помощью функции число округляется до заданного количества знаков (рис. П4.1.b). Функция имеет следующий синтаксис (набор заполняемых параметров):

ОКРУГЛ (*число или ссылка на округляемое число; количество оставляемых знаков*).

Другие варианты использования данной функции рассматриваются на примере задачи, когда число нужно округлить до, например, тысяч (три последние цифры всегда нули). Другими словами – необходимо провести округление не десятичной дроби, а разрядов, то есть сделать так, чтобы некоторое количество знаков в конце числа всегда были нулями. Например, число 123 456 789 нужно округлить до тысяч, т.е. до 123 457 000. Это делается с помощью той же формулы ОКРУГЛ, только количество округляемых разрядов пишется с минусом. Как это выглядит для числа 123 456 789 видно из рис. П4.1.c.

Ниже представлены функции с некоторыми особенностями округления, которые встречаются в практической обработке данных.

ОКРУГЛВВЕРХ и ОКРУГЛВНИЗ Иногда требуется произвести округление в большую или меньшую сторону независимо от близости к числу с требуемым количеством разрядов (знаков после запятой или нулей в конце числа). Например, расчетные цены округляют вверх, чтобы не уменьшить доход, возраст человека округляют вниз до целого, чтобы узнать полное количество лет. Для этих целей используются функции ОКРУГЛВВЕРХ и ОКРУГЛВНИЗ. Данные функции имеют такие же параметры, как и ОКРУГЛ (ссылки на число и количество знаков до или после запятой).

С помощью функции ОКРУГЛТ можно добиться округления не только до нужного разряда (слева или справа от запятой), но и вообще до нужной точности (кратности). К примеру, нужно рассчитать заказ изделий в штуках, но так, чтобы он был равен целому количеству ящиков (рис. П4.1.d). Если в ящике 5 штук, то заказываемое у поставщика количество должно быть кратным пяти.

П4.3. Погрешности прямых и косвенных измерений

Измеряй микрометром. Отмечай мелом. Отрубай топором.
Мерфология, правило точности Рэя

Основной целью большинства экспериментов, в том числе и лабораторных, является измерение физических, в основном, величин.

Измерения не могут быть абсолютно точными в принципе – не существует способов определения истинного значения измеряемой величины как вследствие природы самих измеряемых объектов, так из-за ограниченной точности приборов.

> Измерение – совокупность операций для определения отношения одной (измеряемой) величины к другой однородной величине, принятой за единицу.

Точность результата измерений (precision) – одна из характеристик качества измерения, отражающая близость к нулю погрешности результата измерения. Следует отметить, что о повышении качества измерений всегда говорят термином "увеличить точность" – притом, что величина, характеризующая точность, при этом должна уменьшиться.

Точность средства измерений (accuracy) – степень совпадения показаний измерительного прибора с истинным значением измеряемой величины. Чем меньше разница, тем больше точность прибора.

Абсолютная погрешность приближенного значения – это модуль разности точного (истинного или действительного $x_{\text{ист}}$) и приближенного x значения: $\Delta x = |x - x_{\text{ист}}|$.

Относительная погрешность измерения – отношение абсолютной погрешности измерения к опорному (истинному или действительному) значению измеряемой величины:

$$\delta_x = \frac{\Delta x}{x_{\text{ист}}}, \quad \delta_x = \frac{\Delta x}{\overline{x}}.$$

Относительная погрешность является безразмерной величиной; её численное значение может указываться в процентах.

Обработка результатов измерений предполагает определение наилучшей оценки измеряемой величины $x_{\text{наил}}$ и в определении точности полученного результата $x = x_{\text{наил}} \pm \Delta x$, где Δx есть доверительная погрешность для рассчитанного по результатам наблюдений доверительного интервала $(x_{\text{наил}} - \Delta x;\ x_{\text{наил}} + \Delta x)$.

По природе влияния на результат измерений погрешности делятся на случайные, систематические и грубые (промахи). Систематические погрешности определяются факторами (учитываемыми или не учитываемыми), действующими одинаковым образом при повторе однотипных измерений. Систематические погрешности не уменьшаются с увеличением числа измерений. Случайные погрешности обусловливаются большим количеством трудноучитываемых (неисключаемых) факторов, влияющих на измерительную аппаратуру, на исследуемый объект или процесс и пр. Величина таких погрешностей определяется посредством многократных измерений. Грубые погрешности (промахи) обычно связаны с ошибками регистрации результатов измерений. Способы их отсева приведены в разделе 8.

По способу получения результата измерения делятся на прямые и косвенные.

> Прямые измерения – это измерения, при которых искомое значение физической величины определяется непосредственно путём сравнения с мерой этой величины.

> Косвенные измерения – измерения, при которых значение величины находится на основании известной зависимости между этой величиной и величинами, подвергаемыми прямым измерениям.

Например, скорость автомобиля может быть определена по спидометру (прямое измерение) или найдена из отношения пройденного пути и времени движения (косвенное измерение).

В общем случае погрешность Δx_{Σ} прямых измерений вычисляется по значениям "случайной" $\Delta x_{\text{случ}}$ и приборной $\Delta x_{\text{приб}}$ составляющих

$$\Delta x_{\Sigma} = \sqrt{\left(\Delta x_{\text{случ}}\right)^2 + \left(\Delta x_{\text{приб}}\right)^2}\ . \qquad \text{(П4.1)}$$

Погрешность $\Delta x_{\text{приб}}$ результата прямого однократного измерения зависит от используемых приборов; на лицевой панели некоторых из них указывается так называемый класс точности К. Значение К выражает в процентах от нижнего $x_{\text{мин}}$ и верхнего предела измерения $x_{\text{макс}}$ через предельно допустимую погрешность $\Delta x_{\text{приб}}^{\text{макс}}$ $\Delta x_{\text{приб}}$, а именно

$$\text{К} = \frac{\Delta x_{\text{приб}}^{\text{макс}}}{x_{\text{макс}} - x_{\text{мин}}} \cdot 100\%\,,$$

Из этого соотношения можно определить предельно допустимую погрешность

$$\Delta x_{\text{приб}}^{\text{макс}} = 10^{-2}\text{К} \cdot \left(x_{\text{макс}} - x_{\text{мин}}\right). \qquad \text{(П4.2)}$$

При отсутствии указания класса точности за допустимую погрешность прибора (например, миллиметровой линейки, барометра, термометра, часов с секундной стрелкой) можно принять цену деления прибора. В нижеприведенной таблице приведены достаточно часто используемые в лабораторном эксперименте предельные (т.е. отвечающих вероятности p=1) погрешности $\Delta x_{\text{приб}}^{\text{макс}}$ инструментов.

Прибор или инструмент	Цена деления прибора	Предельная погрешность $\Delta x_{\text{приб}}^{\text{макс}}$
Измерительная линейка	1 см/дел	0.5 см
Весы технические до 2 кг	–	1 г
Весы аналитические до 0,2 кг	0.1 мг/дел	1 мг
Секундомер механический	0.1 с/дел	0.3 с
	0.2 с/дел	0.4 с
Секундомер электрический	0.001 с/дел	0.03 с
Табличная величина	$\pm$5 единиц последнего приведенного в ее записи разряда	

При p <1 приборную погрешность принято рассчитывать по формуле

$$\Delta x_{\text{приб}} = t(\alpha, \infty)\, \Delta x_{\text{приб}}^{\text{макс}}/3\,,$$

где $t(\alpha, \infty)$ – квантиль распределения Стьюдента, а значение предельной погрешности $\Delta x_{\text{приб}}^{\text{макс}}$ определяется либо по классу точности (П4.2), либо берется в зависимости от цены деления прибора.

Увеличение точности измерений можно достигнуть двумя способами – за счет соответствующего улучшения метода измерений (аппаратуры) либо за счет увеличения числа измерений до определенного значения. При выборе числа измерений необходимо учитывать систематическую погрешность измерений $\Delta x_{\text{приб}}$, определяемую классом точности используемого прибора или другими аналогичными обстоятельствами. Снижения случайной погрешности через увеличение числа измерений оправдывается до тех пор, пока суммарная (общая) погрешность измерений Δx_{Σ} не будет полностью определяться ее систематической составляющей. Для этого необходимо, чтобы доверительный интервал, определенный с задаваемой надёжностью, был существенно меньше величины систематической погрешности $\Delta x \ll \Delta x_{\text{приб}}$. Практически можно удовлетвориться гораздо менее жестким требованием $\Delta x \leq \Delta x_{\text{приб}}/3$.

Случайные погрешности возникают из-за одновременного действия многих независимых причин, каждая из которых в отдельности в какой-то мере влияет на результат измерения. Часто распределение случайной погрешности предполагается "нормальным" (см. Приложение П3), что дает возможность использования при обработке данных соответствующий аппарат параметрической математической статистики. В частности, в этом случае погрешность оценивается через произведение стандартной ошибки среднего m бесповторного отбора на квантиль распределения Стьюдента $t(\alpha, n-1)$ для числа степеней свободы nи уровня значимости α

$$\Delta x_{\text{случ}} = t(\alpha, n-1) \cdot m = t(\alpha, n-1) \sqrt{\frac{1}{n(n-1)} \sum_i (x_i - \overline{x})^2} \ , \qquad \text{(П4.3)}$$

где n – количество измерений.

Таким образом, в случае приблизительного равенства случайной погрешности, полученной из разброса отдельных измерений, с погрешностью прибора $\Delta x_{\text{приб}}$, результирующее значение погрешности прямых измерений Δx_{Σ} (границы доверительного интервала) определяются соотношением

$$\Delta x_{\Sigma} = \sqrt{[t(\alpha, n-1)\cdot m]^2 + \left[t(\alpha,\infty)\frac{\Delta x_{\text{приб}}}{3}\right]^2}\,, \qquad \text{(П4.4)}$$

В MS Excel квантиль $t(\alpha, n-1)$ рассчитывается через функцию СТЬЮДЕНТ.ОБР.2Х (α; $n-1$), $t(\alpha,\infty)$ посредством НОРМ.СТ.ОБР(1- α/2).

Стандартная ошибка среднего бесповторного отбора для заданных значений <данные> измерений

$$m = \sqrt{\frac{1}{n(n-1)}\sum_i (x_i - \overline{x})^2}$$

вычисляется в Excel выражением =СТАНДОТКЛОН.В(<данные>) / СЧЁТ(<данные>)^0,5 .

Пример П4.1 Неким прибором, класс точности которого К = 2,5 с диапазоном шкалы 0-100, измерено 10 отсчетов величины X:

№	1	2	3	4	5	6	7	8	9	10
X	145	140	145	105	130	150	150	155	175	160

Необходимо дать результат измерения для уровня значимости α=0.05 .

Алгоритм обработки данных следующий.

1. Проверка наличия промахов. Исходные данные из диапазона B4:B13 копируются в ячейки D4:D13 и сортируются по возрастанию. По критерию Романовского проверяются "на выброс" наименьшее и наибольшие (сомнительные $x^{\text{сом}}$) значения D4 и D13, для чего в ячейках G10 и D11 рассчитывается соответствующие статистики (рис. П4.2)

$$\beta = \frac{|x^{\text{сом}} - \overline{x}|}{S}.$$

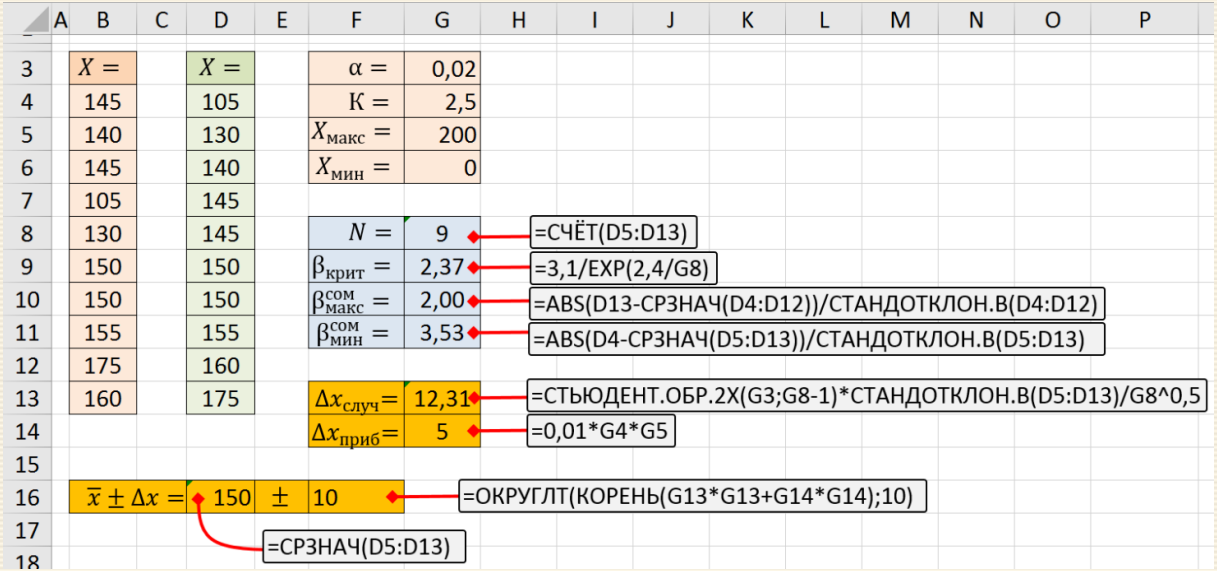


Рис. П4.2. Обработка результатов прямых измерений

Сравнение этих значений с критическим уровнем (ячейка G9) требует отбросить наименьшее значение и далее работать только с данными измерений диапазона D5:D13.

2. По соотношению (П4.2) и известным параметрам прибора в ячейке G14 рассчитывается аппаратная погрешность, а по соотношению (П4.3) в ячейке G13 – случайная. По этим значениям в соответствии с соотношением (П4.1) в ячейке F16 определяется общая округляемая погрешность измерений.

Иногда необходимо объединить результаты нескольких серий прямых измерений одной и той же физической величины. Эту задачу можно решить следующим образом. Пусть результаты измерений представлены в виде $x_i = \overline{x}_i \pm \Delta x_i$. Наилучшее значение $\overline{x}$ и его погрешность $\Delta_{\overline{x}}$ вычисляются по формулам:

$$\overline{x} = \frac{\sum_i \omega_i x_i}{\sum_i \omega_i}, \quad \Delta_{\overline{x}} = \frac{1}{\sqrt{\sum_i \omega_i}}, \qquad \text{где} \qquad \omega_i = \frac{1}{(\Delta x_i)^2}$$

– статистический вес каждой серии измерений.

Пример П4.2 В трех различных условиях измерена длина некого объекта L. Результаты измерений представлены в виде: $L_1 = 10 \pm 3$ см; $L_2 = 11 \pm 2$ см; $L_3 = 10 \pm 2$ см .

Чтобы объединить эти измерения определяются статистические вклады (веса) каждого измерения $\omega_1 = 1/3^2 \approx 0.11$; $\omega_2 = \omega_3 = 1/2^2 = 0.25$; суммарный вес $\sum_i \omega_i = \omega_1 + \omega_2 + \omega_3 \approx 0.61$.

Весовая оценка длины объекта равна $\overline{L} = (10 \cdot 0.11 + 11 \cdot 0.25 + 10 \cdot 0.25)/0.61 = 10.4$; погрешность оценки $\Delta_{\overline{L}}$ составит $\Delta_{\overline{L}} = 1/0.61^{0.5} = 1.28$. После округления окончательно: $L = 10 \pm 1$ см.

Обработка данных косвенных измерений

В большинстве экспериментов искомая величина непосредственно не измеряется, а определяется через другие k величин (аргументов) $x_1, x_2, \dots, x_k$. Измеряемое значение F вычисляется на основе заданной функциональной зависимости

$$F = F(x_1, x_2, \dots, x_k)\,. \qquad (П4.5)$$

Если для каждого аргумента в выражении (П4.5) экспериментально найдены средние значения $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k$ и вычислены погрешности $\Delta x_1, \Delta x_2, \dots, \Delta x_k$, то за наилучшее приближение для величины F принимается значение $\overline{F} = F(\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k)$, получающееся при подстановке в выражение (П4.5) вместо истинных значений аргументов их средних экспериментальных значений.

Доверительная погрешность ΔF косвенных измерений величины F определяется погрешностями $\Delta x_1, \Delta x_2, \dots, \Delta x_k$ (однократных или многократных) прямых измерений.

По определению, полный дифференциал функции – это приращение функции, вызванное малыми приращениями аргументов:

$$dF = \sum \frac{\partial F}{\partial x_i} dx_i\,.$$

Используя замену дифференциалов абсолютными погрешностями (при условии, что абсолютные погрешности достаточно малы) $dF \approx \Delta F$, $dx_i \approx \Delta dx_i$ можно оценить полное приращение функции ΔF, обусловленное изменением ее аргументов x_i на малые величины Δx_i.

Для практических расчётов погрешности ΔF косвенно измеренной величины F, выраженной функцией (П4.5), в предположении, что x_i – статистически независимые величины, применяется статистическое суммирование

$$\Delta F = \sqrt{\sum \left(\frac{\partial F}{\partial x_i}\Delta x_i\right)^2} \ . \qquad \text{(П4.6)}$$

Таким образом, для того чтобы определить абсолютную погрешность результата косвенного измерения, следует найти частные производные функции F по всем аргументам

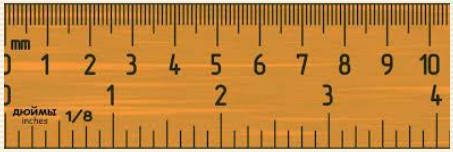
$$\frac{\partial F}{\partial x_i}, \quad i = 1, \dots, k \ ,$$

подставить в них найденные на предыдущем этапе измерений средние значения аргументов $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k$ и рассчитать ΔF по формуле (П4.6).

Выражение (П4.6) удобно при вычислениях погрешности косвенных измерений величин, определяемых формулой, в которую входят сумма или разность, а также формулой, содержащей степенную функцию. Значения ΔF и δF для некоторых функций приведены в нижеследующей таблице.

Замечание. При непосредственных расчетах погрешности Δx_i должны определяться на одном и том же уровне значимости. Погрешность косвенного измерения ΔF будет соответствовать данному значению уровня значимости.

Связь погрешностей прямых и косвенных измерений



	$F = F(x, y, \dots)$	ΔF	$\delta F = \Delta F / \overline{F}$
1	$F = ax$	$a\,\Delta x$	δx
2	$F = x^a$	$a \cdot (\overline{x})^{a-1} \Delta x$	$a\,\delta x$
3	$F = x/(1 \pm x)$	$\Delta x/(1 \pm \overline{x})^2$	$\delta x/(1 \pm \overline{x})^2$
4	$F = \exp(x/a)$	$(\Delta x/a) \cdot \exp(\overline{x}/a)$	$\Delta x/a$
5	$F = \ln x$	δx	$\delta x / \ln \overline{x}$
6	$F = \sin(x/a)$	$(\Delta x/a) \cdot \cos(\overline{x}/a)$	$(\Delta x/a) \cdot \mathrm{ctg}(\overline{x}/a)$
7	$F = ax \pm by$	$\sqrt{(a\,\Delta x)^2 + (b\,\Delta y)^2}$	$\Delta F/(a\,\overline{x} \pm b\,\overline{y})$
8	$F = xy$	$\sqrt{(\overline{y}\,\Delta x)^2 + (\overline{x}\,\Delta y)^2}$	$\sqrt{(\delta x^2) + (\delta y^2)}$
9	$F = x/y$	$\sqrt{(\Delta x/\overline{y})^2 + (\Delta y/\overline{x})^2}$	$\sqrt{(\delta x^2) + (\delta y^2)}$
10	$F = ax \pm by \pm cz$	$\sqrt{(a\,\Delta x)^2 + (b\,\Delta y)^2 + (c\,\Delta z)^2}$	$\Delta F/F$
11	$F = ax^{\pm\alpha} y^{\pm\beta} z^{\pm\gamma}$	$F \cdot\ \delta F$	$\sqrt{(\alpha\,\delta x)^2 + (\beta\,\delta y)^2 + (\gamma\,\delta z)^2}$

Пример П4.3 Прямыми измерениями найдены значения радиуса r и линейной скорости v равномерного вращения по окружности материальной точки массы m. Необходимо оценить значение центробежной силы F, действующей на этот материальный объект, характеризуемый следующими результатами эксперимента: m = 400 ± 10 г; r = 100 ± 5 мм; v = 30 ± 1 м/с. Величина силы определяется соотношением $F = mv^2/r$.

Алгоритм, использующий вычисление производных измеряемой величины по её аргументам, следующий.

1. В ячейки C2:C4 и E2:E4 заносятся исходные данные согласно системе СИ единиц физических величин (рис. П4.3), полученными и обработанными в результате прямых измерений.

В ячейке C9 вычисляется среднее значение силы $\overline{F} = mv^2/r$.

2. Находятся частные производные и вычисляются их значения при средних значениях аргументов

$$\Delta F_m = \frac{\partial F}{\partial m} = \frac{v^2}{r}\Delta m, \quad \Delta F_r = \frac{\partial F}{\partial r}\Delta r = \frac{mv^2}{r^2}\Delta r, \quad \Delta F_v = \frac{\partial F}{\partial v}\Delta v = 2\frac{mv}{r}\Delta v$$

и вычисляются в ячейках C6:C8.

3. Вычисляется полная абсолютная ΔF (ячейка C10) и относительная δF (C11) погрешности по соотношениям:

$$\Delta F = \sqrt{\Delta F_m^2 + \Delta F_r^2 + \Delta F_v^2} \ . \quad \delta F = \frac{\Delta F}{\overline{F}}.$$

4. Переводя значения $\overline{F}$ и ΔF (C9 и C10) в более крупную единицу (ньютоны в килоньютоны) и округляя их результаты обработки данных записывается в виде $F = 3.6 \pm 0.3$ кН; $\delta F = 7\%$.

	A	B	C	D	E	F	G	H
2		$m =$	0,4	$\pm$	0,01			
3		$r =$	0,1	$\pm$	0,01			
4		$v =$	30	$\pm$	1			
5								
6		$\Delta F_m =$	90		=C4*C4/C3*E2			
7		$\Delta F_r =$	180		=C2*C4*C4/C3/C3*E3			
8		$\Delta F_v =$	240		=2*C2*C4/C3*E4			
9		$\overline{F} =$	3600		=C2*C4*C4/C3			
10		$\Delta F =$	313		=(C6^2+C7^2+C8^2)^0,5			
11		$\delta F =$	0,07		=C8/C9			
12								
14		$\overline{F} \pm \Delta F =$	3,6	$\pm$	0,3	kN	$\delta F =$	7%
15								
16								

=ОКРУГЛТ(0,001*C10;0,1)

=ОКРУГЛТ(0,001*C2*C4*C4/C3;0,1)

формат ячейки "процентный"

Рис. П4.3. Обработка результатов косвенных измерений

Массивы и имена в MS Excel

Под м а с с и в о м понимается набор данных, объединенных в группу.

Массивы бывают одномерные (элементы массива образуют строку или столбец) или двумерные (матрица). Практически в любой таблице MS Excel при желании можно найти один или несколько таких массивов.

Формулы массива в MS Excel – это специальные формулы для обработки данных массивов. Формулы массива делятся на две категории – те, которые возвращают одно значение и те, что дают на выходе целый набор (массив) значений.

На рис. П5.1 приведен простой пример вычислений с массивами значения

$$\sum M_i \, (N_i - 3^2) \, ,$$

где массив N – значения в диапазоне B2:B9

и M – в диапазоне F2:F9.

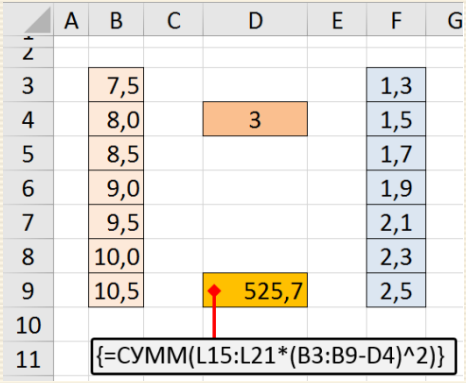


Рис. П5.1. Пример вычислений с массивами

После ввода в ячейку D9 формулы =СУММ(F3:F9*(B3:B9-D4)^2) в ней появится какое-то "неправильное" значение или даже сообщение об ошибке вида "ЗНАЧ!", но после выполнения F2 и Ctrl + Shift + Enter высветится искомый правильный результат.

Фигурные скобки, появившиеся в формуле – отличительный признак формулы массива. Вводить их вручную с клавиатуры бессмысленно и бесполезно – они автоматически появляются при нажатии комбинации клавиш активации формулы массива.

Замечание. В некоторых версиях MS Excel массивы, участвующие в операциях, должны быть однотипного расположения – либо все столбиковые, либо строчные.

Для упрощения адресации диапазонам данных (например, вследствие того, что в списке может быть много и даже очень много элементов) могут быть даны (присвоены) и м е н а . Имя может представлять ячейку, диапазон ячеек, формулу или значение константы.

Область действия имени. Все имена имеют область действия: это либо конкретный лист (локальный уровень листа), либо вся книга (глобальный уровень книги, общекнижный). Область действия имени – это область, в которой имена распознаются без уточнений. Примеры:

- Если для Листа1 определено имя "Данные", то это имя без уточнения распознается только на листе "Лист1" и нигде более. Чтобы использовать локальное имя листа в другом листе, его можно уточнить, предварив именем листа: Лист1!Данные.
- Если имя "Данные" определено для книги, то это имя распознается на всех листах этой книги, но не в какой-либо другой книге.

Имя должно быть уникальным в своей области определения. В случае одинаковых имен в разных областях определения (например, имя "Данные" определено на Лист1 и как глобальное для книги), Excel по умолчанию использует значение имени на листе: локальный уровень листа выше глобального уровня книги.

Чтобы на Лист1 использовать глобальное имя необходимо его адресовать полностью:

ИмяФайлаКниги!Данные.

Синтаксические правила для имен следующие.

- Ограничение наименований. В качестве определенного имени нельзя использовать буквы "C", "c", "R" и "r", поскольку эти буквы используются как сокращенное имя строки и столбца.
- Имена в виде ссылок на ячейки запрещены. Имена не могут быть такими же, как ссылки на ячейки. Нельзя, например, использовать X1, X$1 или R1C1.
- Ограничение символов. Первым символом имени должна быть буква, знак подчеркивания (_) или обратная косая черта (\). Остальные символы имени могут быть буквами, цифрами, точками и знаками подчеркивания.
- Пробелы не допускаются. В качестве разделителей слов можно использовать символы подчеркивания или точки.
- Длина имени не должно быть длиннее 255 символов.
- Регистр: имя может состоять из строчных и прописных букв. Excel не различает строчные и прописные буквы в именах. Например, если создать имя "Данные", то создание нового имени "ДАННЫЕ" Excel заблокирует.

Имена присваиваются (в новых версиях MS Excel) на вкладке Формулы → Диспетчер имен (Formulas → Name manager) или в старых версиях – через меню Вставка → Имя → Присвоить (Insert → Name → Define). В появившейся вкладке указывается собственно имя, диапазон данных и область распространения (работы) имени в книге или конкретном листе MS Excel (рис. П5.2, левый). Имя диапазону (или ячейке) для всей книги) можно для выделенных ячеек можно присвоить непосредственно в адресной строке (рис. П5.2, в центре). Второй способ создания имени ячейки или диапазона ячеек для книги.

Выделяется ячейка, диапазон ячеек или несмежный диапазон, которому требуется присвоить имя.

В поле "Имя" у левого края строки формул вводится собственно имя, которое будет использоваться для ссылки на указанную ячейку или диапазон (рис. П5.2, справа).

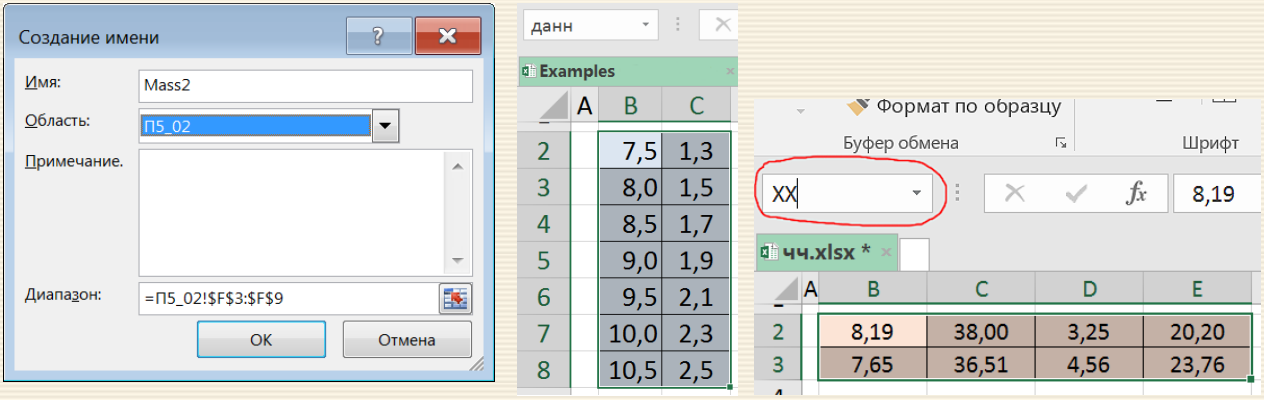


Рис. П5.2. Создание имени массива

Проверка имен.

Кроме управления каталогом через меню "Диспетчер имен" можно создать их список. Перечень будет состоять из двух столбцов: в первом имена, а во втором их краткие описания (как и если они были введены при создании). Указывается ячейка, с которой список будет начинаться: на вкладке Формулы нажимаются пункты Использовать в формуле и Вставить имена. В открывшемся диалоговом окне Вставка имен выбираются Все имена (рис. П5.3).

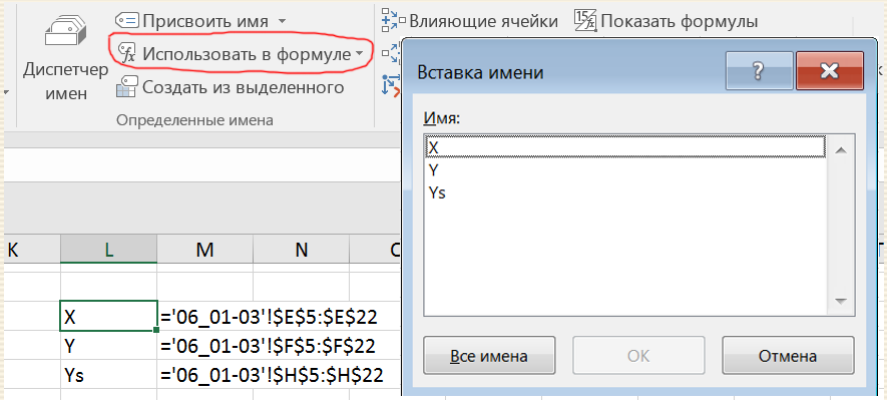


Рис. П5.3. Составление списка имен

На рис. П5.4 приведены вычисления предыдущего примера (рис. П5.1.) с использованием имен переменных и массивов.

По приведенным выше правилам можно присваивать имена формулам MS Excel. На рис. П5.5 решение упоминаемого выше примера приведено с наименованиями значений, массивов и формул.

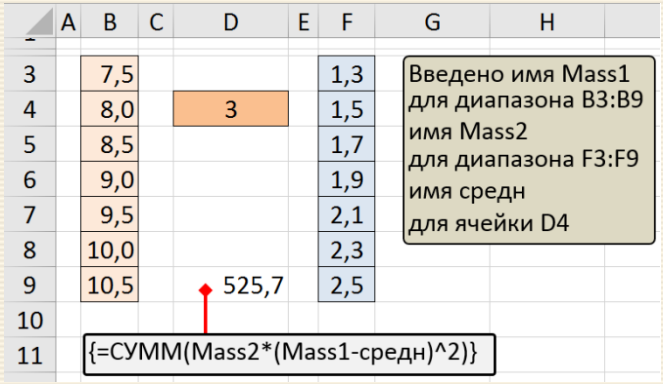


Рис. П5.4. Пример вычислений с именованными значениями и массивами

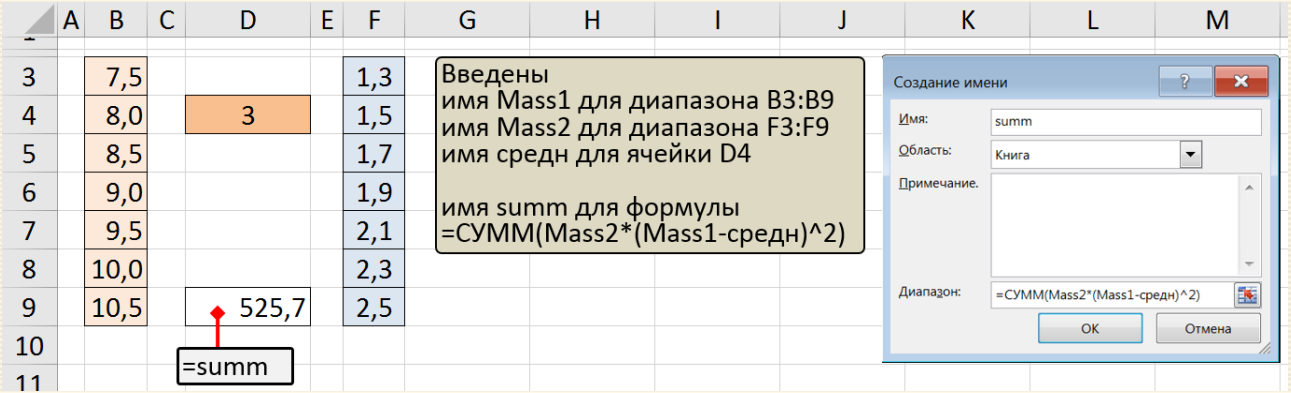


Рис. П5.5. Пример вычислений с именованными значениями, массивами и формулами

Перечень использованных функций MS Excel

СЧЁТ (<диапазон> или <имя>)	Подсчитывается количество ячеек, содержащих числа
СЧЁТЕСЛИ (<диапазон>; <условие>)	Подсчитывается количество в диапазоне, если выполняется некоторое условие, заданное в текстовом виде. Примеры условий: "Вася", ">7".
СРЗНАЧ (<диапазон> или <имя>)	Возвращается среднее(арифметическое) значение аргументов
МАКС (<диапазон> или <имя>)	Возвращается наибольшее значение из набора данных
МИН (<диапазон> или <имя>)	Возвращается наименьшее значение в списке аргументов
ДИСП.В (<диапазон> или <имя>) ДИСП.Г (<диапазон> или <имя>)	Оценивается дисперсия по выборке Дисперсия генеральной совокупности
ABS (<величина>)	Возвращает модуль (абсолютную величину) числа
СТАНДОТКЛОН.В (<диапазон>) СТАНДОТКЛОН.В (<имя>) СТАНДОТКЛОН (<диапазон>) СТАНДОТКЛОН (<имя>)	Оценивает стандартное отклонение по выборке. Предполагается, что аргументы являются только выборкой из генеральной совокупности

СТАНДОТКЛОН.Г (<диапазон>) СТАНДОТКЛОН.Г (<имя>)	Стандартное отклонение для генеральной совокупности	
НОРМ.СТ.ОБР (P)	Возвращает обратное значение стандартного нормального распределения, имеющего нулевое среднее и единичное стандартное отклонение. При заданном α вероятность $P = 1 - \alpha/2$.	
СТЬЮДЕНТ.ОБР.2Х (α;df)	Возвращает двустороннее обратное t-распределение Стьюдента	α – уровень значимости; df – число степеней свободы, характеризующее распределение
FОБР.ПХ (α; df_1, df_2) FРАСПОБР (α; df_1, df_2)	Возвращается значение, обратное (правостороннему) F-распределению вероятностей	α – уровень значимости; df_1 – (числитель степеней свободы); df_2 – (знаменатель степеней свободы)
БЕТА.ОБР (вероятность; альфа; бета; [A]; [B]) БЕТАОБР (вероятность; альфа; бета; [A]; [B]) альфа, бета – параметры распределения	Возвращает обратную функцию к интегральной функции плотности бета-распределения вероятности. p – вероятность, связанная с бета-распределением; [A] – нижняя граница интервала изменения x (необязательный аргумент); [B] –Верхняя граница интервала изменения x (необязательный аргумент).	

ЕСЛИ (<ЛУ>; <значение при ЛУ=истина>; <значение при ЛУ=ложь>)	ЛУ – логическое условие например A7<5; A6=F4*Y9
И (ЛУ$_1$; ЛУ$_2$)	возвращает значение "истина", если ЛУ$_1$ = истина **и** ЛУ$_2$ = истина. В остальных случаях И(ЛУ$_1$; ЛУ$_2$) возвращает "ложь"
СКОС (<диапазон> или <имя>)	Возвращает асимметрию распределения относительно его среднего. Положительная асимметрия указывает на отклонение распределения в сторону положительных значений. Отрицательная асимметрия указывает на отклонение распределения в сторону отрицательных значений.
ЭКСЦЕСС (<диапазон> или <имя>)	Возвращает эксцесс распределения, характеризующий его "остроконечность" или сглаженность по сравнению с нормальным распределением. Положительный эксцесс указывает на относительную остроконечность, отрицательный – на относительно сглаженное распределение.
ДОВЕРИТ.НОРМ $(\alpha; S; n)$	Возвращает доверительный интервал для среднего генеральной совокупности с нормальным распределением; α – уровень значимости; S – стандартное отклонение; n – размер выборки.

ДОВЕРИТ.СТЬЮДЕНТ (α; S; n)	Возвращает доверительный интервал для среднего генеральной совокупности, используя распределение Стьюдента.
НАКЛОН (значения_y; значения_x)	Возвращает наклон линии линейной регрессии значения_y – массив или диапазон ячеек, содержащих зависимые числовые данные, значения_x – множество независимых данных.
ОТРЕЗОК (значения_y; значения_x)	Вычисляет точку пересечения оптимальной линии регрессии с осью y
СРОТКЛ(данные);	Возвращает среднее абсолютных значений отклонений точек данных от среднего.
ЧАСТОТА (массив_данных; массив_интервалов)	Вычисляет частоту появления значений в интервале значений и возвращает массив чисел. массив_данных – массив или ссылка на множество значений, для которых вычисляются частоты; массив_интервалов – массив или ссылка на множество интервалов, в которые группируются значения аргумента "массив_данных".
ОКРУГЛ (…); ОКРВВЕРХ.МАТ (…); ОКРВНИЗ.МАТ (…); ОКРУГЛВВЕРХ (…); ОКРУГЛВНИЗ (…); ОКРУГЛТ (…);	Функции округления чисел. См. раздел П4.2.

Виды ошибок при задании формул

Формула в Microsoft Excel представляет собой синтаксическую конструкцию, начинающуюся со знака равенства (=) и предназначенную для обработки данных с последующим помещением результатов обработки в ячейку, где записана сама формула. Формула может содержать одну или несколько функций, связанных между собой арифметическими операторами или вложенных друг в друга. Если при задании формулы были допущены ошибки, результатом ее вычисления будет так называемое значение ошибки, которое появится в ячейке. В зависимости от вида ошибки в ячейке, содержащей формулу, записываются различные значения.

Первым символом значения ошибки является символ диез (#), за которым следует текст. Текст значения ошибки может завершаться восклицательным знаком или знаком вопроса.

Ниже приводится список значений ошибок с пояснением наиболее распространенных причин их возникновения и указанием мер по их устранению.

Ошибка ##### Причины возникновения ошибки:

1. Вводимое числовое значение не умещается в ячейке.

Устранение ошибки – увеличьте ширину столбца перемещением его границы, расположенной между заголовками столбцов.

2. Используется формула, возвращаемый результат которой не умещается в ячейке.

Устранение ошибки – увеличьте ширину столбца перемещением его границы, расположенной между заголовками столбцов. Кроме того, можно изменить формат числа ячейки, для чего следует выбрать команду Ячейки в меню Формат, затем вкладку Число и указать подходящий формат.

3. При определении числа дней между двумя датами, а также количества часов между двумя временными промежутками получается отрицательное значение.

Устранение ошибки – введите правильно формулу, чтобы число дней или часов было положительным числом.

Ошибка #ЗНАЧ! Причины возникновения ошибки:

1. Вместо числового или логического значения введен текст, и Microsoft Excel не может преобразовать его к нужному типу данных.

Устранение ошибки - проверьте в формуле правильность задания типов аргументов. Например, если в ячейке A1 содержится число 8, в ячейке B1 – текстовое значение "Вася", а в ячейке C1 – формула =A1+B1, то в ячейке C1 будет выведена ошибка #ЗНАЧ!. Если все же необходимо сложить два таких значения, то следует использовать функцию СУММ, игнорирующую текстовые значения. Для рассматриваемого примера функция = СУММ(A1:B1) вернет значение 8.

2. После ввода или редактирования формулы массива нажимается клавиша Enter.

Устранение ошибки — для редактирования формулы укажите ячейку или диапазон ячеек, содержащих формулу массива, нажмите клавишу F2, а затем – комбинацию клавиш Ctrl+Shift +Enter.

3. Использована неправильная размерность матрицы данных в соответствующей функции.

Устранение ошибки — укажите правильную размерность матрицы данных.

Ошибка #ДЕЛ/0! Причины возникновения ошибки:

1. В качестве делителя используется ссылка на ячейку, содержащую нулевое или пустое значение (если аргумент является пустой ячейкой, то ее содержимое интерпретируется как нуль). Такая ситуация чаще всего возникает случайно: например, если ячейка содержит формулу =A1/B1, а содержимое ячейки B1 по какой-либо причине было удалено.

Устранение ошибки — измените ссылку или введите ненулевое значение в ячейку, используемую в качестве делителя.

Кроме того, в качестве делителя можно ввести значение #Н/Д. В этом случае ошибка #ДЕЛ/0! сменится на #Н/Д, что указывает на неопределенность значения делителя.

2. В формуле содержится явное деление на нуль, например, = 8/0.

Устранение ошибки — исправьте формулу.

Ошибка #ИМЯ? Причины возникновения ошибки:

1. Используемое в формуле имя было удалено или не было определено. *Устранение ошибки* – определите имя. Для этого выберите команду Имя в меню Вставка, а затем – команду Создать. Кроме того, команда Создать используется для добавления имени, отсутствующего в списке.
2. Имеется ошибка в написании имени. *Устранение ошибки* – исправьте написание имени.

Чтобы вставить правильное имя в формулу, выделите имя в строке формул, выберите команду Имя в меню Вставка, а затем – команду Вставить. На экране появится диалоговое окно Вставка имени. Выделите нужное имя и щелкните по кнопке ОК.

3. Имеется ошибка в написании имени функции. *Устранение ошибки* — исправьте написание имени функции вручную или вставьте функцию с помощью мастера функций.
4. В формулу введен текст, не заключенный в двойные кавычки. Microsoft Excel пытается распознать такой текст как имя. *Устранение ошибки* – заключите текст формулы в двойные кавычки. Например, если в ячейке A1 содержится значение 100, а в ячейке B1 формула ="Итого:"&A1, то в ячейке B1 будет выведен результат Итого:100.
5. В ссылке на диапазон ячеек пропущен знак двоеточия (:). *Устранение ошибки* – исправьте формулу в части ссылок на диапазон(ы) ячеек. Например =СУММ(A1:C10).

Ошибка #Н/Д Значение ошибки #Н/Д (неопределенные данные) предотвращает использование ссылки на пустую ячейку. Введите в ячейки листа значение #Н/Д, если они должны содержать данные, но в текущий момент эти данные отсутствуют. Формулы, ссылающиеся на эти ячейки, тоже будут иметь значение #Н/Д. Причины возникновения ошибки:

1. Для функций ГОР, ПРОСМОТР, ПОИСКПОЗ или ВПР (функции ссылок и автоподстановок) задан недопустимый аргумент искомое значение. *Устранение ошибки* – задайте правильный аргумент искомое значение (значение или ссылку не диапазон ссылок).

2. Функции ВПР или ГПР используются для обработки неотсортированной таблицы.
Устранение ошибки – по умолчанию для функций просмотра таблиц сведения должны располагаться в возрастающем порядке (аргумент интервальный просмотр опущен или имеет значение ИСТИНА). Чтобы найти искомое значение в неотсортированной таблице, установите для аргумента интервальный просмотр значение ЛОЖЬ.

3. В формуле массива используется аргумент, не соответствующий размеру диапазона, определяющегося числом строк и столбцов. *Устранение ошибки* – если формула массива введена в несколько ячеек, откорректируйте диапазон ссылок формулы на соответствие числу строк и столбцов или введите формулу массива в недостающие ячейки. Например, если формула массива введена в первые 15 ячеек столбца C (C1:C15), а сама формула ссылается на первые 10 ячеек столбца A (A1:A10), то в ячейках C1:C15 будет отображаться ошибка #Н/Д. Чтобы исправить эту ошибку, уменьшите диапазон в формуле (например, C1:C10) или увеличьте диапазон, на который ссылается формула (например, A1:A15).

4. Не заданы один или несколько аргументов стандартной или пользовательской функции листа.
Устранение ошибки – задайте все необходимые аргументы функции.

5. Используется пользовательская функция, обращение к которой приводит к ошибке.
Устранение ошибки – проверьте, что книга, использующая функцию листа, открыта, и убедитесь в правильности работы функции (проведите отладку в редакторе VBA).

Ошибка #ССЫЛКА! Причина возникновения ошибки:
Ячейки, на которые ссылаются формулы, были удалены или в эти ячейки было помещено содержимое других скопированных ячеек. *Устранение ошибки* – измените формулы или сразу же после удаления или вставки скопированного восстановите прежнее содержимое ячеек с помощью кнопки Отменить.

Ошибка #ЧИСЛО! Причины возникновения ошибки:

1. В функции с числовым аргументом используется неприемлемый аргумент. *Устранение ошибки* – проверьте правильность использования в функции аргументов.

2. Задана функция (например, статистическая функция СТЬЮДРАСПОБР), при вычислении которой используется итерационный процесс. При этом итерационный процесс не сходится и результат не может быть получен. *Устранение ошибки* — используйте другое начальное приближение для этой функции.

3. Введена формула, рассчитывающая числовое значение, которое слишком велико или слишком мало, чтобы его можно было представить в Microsoft Excel. *Устранение ошибки* – проверьте и откорректируйте формулу так, чтобы в результате ее вычисления получалось число, попадающее в рабочий диапазон Microsoft Excel. Например, число 222 является слишком большим, чтобы быть использованным в качестве аргумента функции ФАКТР (функция вычисления факториала числа), поэтому формула =ФАКТР(222) помещает в ячейку значение ошибки #ЧИСЛО!.

Ошибка #ПУСТО! Причина возникновения ошибки: использован оператор, задающий пересечение диапазонов, не имеющих общих ячеек. *Устранение ошибки* — задайте правильно размерность пересекающихся диапазонов или не используйте оператор пересечения, если диапазоны не являются таковыми. В Microsoft Excel оператором пересечения диапазонов является пробел ( ). Например, диапазоны A1:A5 и B1:B5 содержат массивы единиц. В этом случае формула =СУММ(A1:A5; B1:B5) будет выдавать значение ошибки #ПУСТО!, а формула =СУММ(A1:A5;A1:B3) рассчитает значение 3. Для суммирования непересекающихся диапазонов A1:A5 и B1:B5 необходимо воспользоваться стандартной синтаксической конструкцией функции СУММ, т. е. =СУММ(A1A5;B1:B5).

Пакет "Анализ данных" Microsoft Excel. Инструментарий

В состав Microsoft Excel входит набор средств анализа данных (Анализ данных), предназначенный для решения достаточно сложных статистических задач. Для проведения анализа данных с помощью инструментов пакета указываются входные данные и параметры; анализ проводится с помощью подходящей статистической макрофункции, а результат помещается в выходной диапазон.

Для того чтобы отыскать команду вызова надстройки Пакет анализа, необходимо воспользоваться меню Данные.

Если в меню Данные отсутствует команда Анализ данных, то в этом случае необходимо в меню Файл-Параметры => Надстройки выполнить Управление-Надстройки Excel-Перейти и поставить "галочку" на "Пакет анализа". После этого в меню Данные появится требуемая команда.

Эта ситуация наиболее типична, так как надстройка Пакет анализа инсталлируется при стандартной установке.

Если в меню Данные отсутствует команда Анализ данных, а в Надстройках нет элемента Пакет анализа, то без установочного файла или компакт-диска в этом случае не обойтись. Для "доустановки" Excel с дистрибутива Microsoft Office нужно перейти в папку Панель управления и через Установка и удаление программ выполнить требуемое.

Доступные средства. Чтобы просмотреть список доступных инструментов анализа, выберите команду Анализ данных в меню Данные (рис. П5.4).

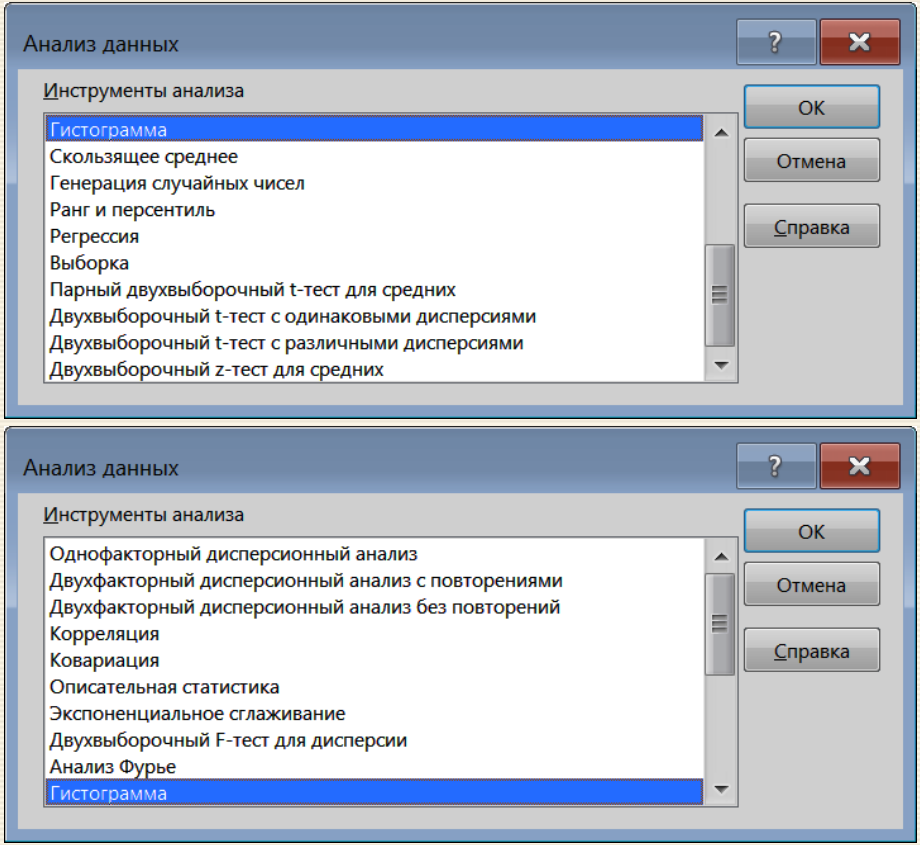


Рис. П5.4. Панель доступных инструментов анализа данных

Дисперсионный анализ

Пакет анализа включает в себя три средства дисперсионного анализа. Выбор конкретного инструмента определяется числом факторов и числом выборок в исследуемой совокупности данных.

Однофакторный дисперсионный анализ используется для проверки гипотезы о сходстве средних значений двух или более выборок, принадлежащих одной и той же генеральной совокупности. Этот метод распространяется также на тесты для двух средних (к которым относится, например, *t*-критерий).

Двухфакторный дисперсионный анализ с повторениями представляет собой более сложный вариант однофакторного анализа, включающее более чем одну выборку для каждой группы данных.

Двухфакторный дисперсионный анализ без повторения представляет собой двухфакторный анализ дисперсии, не включающий более одной выборки на группу. Используется для проверки гипотезы о том, что средние значения двух или нескольких выборок одинаковы (выборки принадлежат одной и той же генеральной совокупности). Данный метод распространяется также на тесты для двух средних, такие как *t*-критерий.

Корреляция

Используется для количественной оценки взаимосвязи двух наборов данных, представленных в безразмерном виде. Коэффициент корреляции выборки представляет собой ковариацию двух наборов данных, деленную на произведение их стандартных отклонений.

Корреляционный анализ дает возможность установить, пропорциональны ли величины в наборах данных. А именно, большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция), или, наоборот, малые значения одного набора связаны с большими значениями другого (отрицательная корреляция), или данные двух диапазонов никак не связаны (околонулевая корреляция).

Ковариация

Используется для вычисления среднего произведения отклонений точек данных от относительных средних. Ковариация является мерой связи между двумя диапазонами данных. Ковариационный (как и корреляционный) анализ дает возможность установить ассоциированность наборов данных по величине.

Гистограмма

Используется для вычисления выборочных и интегральных частот попадания данных в указанные интервалы значений, при этом, генерируются числа попаданий для заданного диапазона ячеек. Например, необходимо выявить тип распределения успеваемости в группе студентов. Таблица гистограммы состоит из границ шкалы оценок и численности студентов, уровень успеваемости которых находится между нижней границей и текущей границей. Наиболее часто повторяемый уровень является модой интервала данных.

Экспоненциальное сглаживание*

Предназначается для предсказания значения на основе прогноза для предыдущего периода, скорректированного с учетом погрешностей в данном прогнозе. Использует константу сглаживания, по величине которой определяет, насколько сильно влияют на прогнозы погрешности предыдущего прогноза. Более старым наблюдениям приписываются экспоненциально убывающие веса, при этом в отличие от скользящего среднего учитываются все предшествующие наблюдения ряда, а не те, что попали в определенное окно. Формула метода простого экспоненциального сглаживания имеет следующий вид: $\hat{y}_t = (1 - \alpha)\hat{y}_{t-1} + \alpha\hat{y}_t$, где 0 < α < 1 – коэффициент экспоненциального сглаживания.

Двухвыборочный *F*-тест для дисперсий

Двухвыборочный *F*-тест применяется для сравнения дисперсий двух генеральных совокупностей. Например, *F*-тест можно использовать для выявления различия в дисперсиях временных характеристик, вычисленных по двум выборкам.

Описательная статистика

Данное средство анализа служит для создания одномерного статистического отчета, содержащего информацию о центральной тенденции и изменчивости входных данных.

*Исторически метод экспоненциального сглаживания был независимо открыт Броуном и Холтом для решения задач прогнозирования спроса на запасные части вооружения и военной техники ВМС США

Скользящее среднее

Используется для расчета значений в прогнозируемом периоде на основе среднего значения переменной для указанного числа предшествующих периодов. Метод скользящей средней состоит в том, что исходный эмпирический временной ряд y_i преобразуется в ряд сглаженных значений (оценок) по формуле

$$\hat{y}_t = \frac{1}{p} \sum_{j=t-m}^{t+m} y_j ,$$

где p – размер окна; j – порядковый номер уровня в окне сглаживания; $m = (p-1)/2$.

Скользящее среднее, в отличие от простого среднего для всей выборки, содержит сведения о тенденциях изменения данных. Процедура может использоваться для прогноза сбыта, инвентаризации и других процессов.

Генерация случайных чисел

Используется для заполнения диапазона случайными числами, извлеченными из одного или нескольких распределений. С помощью данной процедуры можно моделировать объекты, имеющие случайную природу, по известному распределению вероятностей. Например, можно использовать нормальное распределение для моделирования совокупности данных по росту индивидуумов, или использовать распределение Бернулли для двух вероятных исходов, чтобы описать совокупность результатов бросания монетки.

Проведение *t*-теста

Пакет анализа включает в себя три средства анализа среднего для совокупностей различных типов:

Двухвыборочный *t*-тест Стьюдента с одинаковыми дисперсиями служит для проверки гипотезы о равенстве средних для двух выборок. Эта форма *t*-теста предполагает совпадение дисперсий генеральных совокупностей и обычно называется гомоскедастическим *t*-тестом.

Двухвыборочный *t*-тест Стьюдента с разными дисперсиями используется для проверки гипотезы о равенстве средних для двух выборок данных из разных генеральных совокупностей. Эта форма *t*-теста предполагает несовпадение дисперсий генеральных совокупностей и обычно называется гетероскедастическим *t*-тестом. Если тестируется одна и та же генеральная совокупность, используйте парный тест.

Парный двухвыборочный *t*-тест для средних используется для проверки гипотезы о различии средних для двух выборок данных. В нем не предполагается равенство дисперсий генеральных совокупностей, из которых выбраны данные. Парный тест используется, когда имеется естественная парность наблюдений в выборках, например, когда генеральная совокупность тестируется дважды.

Ранг и персентиль

Используется для вывода таблицы, содержащей порядковый и процентный ранги для каждого значения в наборе данных. Данная процедура может быть применена для анализа относительного взаиморасположения данных в наборе.

Регрессия

Линейный регрессионный анализ заключается в подборе графика для набора наблюдений с помощью метода наименьших квадратов. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных. Например, на спортивные качества спортсмена влияют несколько факторов (возраст, рост и вес). Регрессия "распределяет" меру качества по этим факторам. Результаты регрессии впоследствии могут быть использованы для предсказания качеств другого спортсмена.

Выборка

Создает выборку из генеральной совокупности, рассматривая входной диапазон как генеральную совокупность. Если совокупность слишком велика для обработки или построения диаграммы, можно использовать представительную выборку. Кроме того, если предполагается периодичность входных данных, то можно создать выборку, содержащую значения только из отдельной части цикла.

Например, если входной диапазон содержит данные для квартальных привесов, создание выборки с периодом 4 разместит в выходном диапазоне значения привесов из одного и того же квартала.

Двухвыборочный *z*-тест для средних"

Выполняется двухвыборочный z-тест для средних с известными дисперсиями, который используется для проверки основной гипотезы об отсутствии различий между средними двух генеральных совокупностей относительно односторонней и двусторонней альтернативных гипотез.

Вопросы для самопроверки

Т.1 Определение генеральной совокупности. Количество элементов в генеральной совокупности. Два примера генеральных совокупностей. Материалы для контроля

Т.2 Разбивка случайной величины на интервалы (классы); выбор числа интервалов. Эвристическая формула Sturgess'a для определения "оптимального" числа интервалов. Материалы для контроля

Т.3 Свойство репрезентативности как необходимое условие. Закон больших чисел. Материалы для контроля Материалы для контроля

Т.4 Нулевая гипотеза. Критерии, эмпирическое и критическое значения критерия. Уровень значимости, его смысл. Материалы для контроля Материалы для контроля

Т.5 Правило принятия нулевой гипотезы. Критерии однородности. Материалы для контроля Материалы для контроля Материалы для контроля

Т.6 Алгоритм проверки статистической гипотезы с помощью (статистического) критерия. Материалы для контроля

Т.7 Выборочное среднее значение, среднеарифметическое выборочное. Формулы расчета среднеарифметического выборочного таблицы данных и данных виде интервального частотного ряда. Материалы для контроля

Т.8 Выборочное среднее отклонение (выборочная оценка среднего отклонения). Модульный и квадратический подходы и подход к вычислению меры отклонения между величинами. Материалы для контроля

Т.9 Доверительный интервал. Условие построения доверительного интервала. Предельное отклонение при заданном уровне значимости α. Материалы для контроля Материалы для контроля

Т.10 Дисперсия как основной статистический показатель. Формулы расчета дисперсии для таблицы данных и данных в виде интервального частотного ряда. Материалы для контроля Материалы для контроля

Т.11 Дисперсионный анализ проверки гипотез, связанных с оценкой выборочной дисперсии. Основные виды гипотез. Материалы для контроля

Т.12 Классические условия применимости критерия Стьюдента. Случаи равенства/неравенства дисперсий выборок. Материалы для контроля Материалы для контроля

Т.13 Общая схема (4 случая) исследования равенства средних по критерию Стьюдента. Материалы для контроля

Т.14 Показатели разброса данных: дисперсия, среднее квадратическое отклонение, среднее линейное отклонение. Понятие однородности. Материалы для контроля Материалы для контроля Материалы для контроля Материалы для контроля

Т.15 Планирование эксперимента. Процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Материалы для контроля

Т.16 Линейное преобразование (кодирование) факторного пространства. Кодированная система координат, центр плана.

Материалы для контроля

Материалы для контроля

Материалы для контроля

Т.17 Проверка однородности выборочных дисперсий по критериям Кохрена и Бартлетта.

Материалы для контроля

Материалы для контроля

Т.18 Уравнение регрессии, значимость и качество показателей уравнения. Надёжность уравнения регрессии.

Материалы для контроля

Материалы для контроля

Материалы для контроля

Р.1 В результате обследования 8596 мужчин по росту были получены результаты, приведенные в таблице, построенной на основании вариационного ряда, разбитого на 11 равных интервалов. Проверить гипотезу: полученные результаты имеют нормальный закон распределения при уровне значимости $\alpha = 0.1$, приняв за математическое ожидание и дисперсию гипотетического закона их оценки, вычисленные по экспериментальным данным.

Интервал	Частота
[142; 147)	7
[147; 152)	56
[152; 157)	253
[157; 162)	1064
[162; 167)	2214
[167; 172)	2560
[172; 177)	1708
[177; 182)	550
[182; 187)	112
[187; 192)	22
[192; 197)	3

Р.2 В двух водоемах были взяты пробы рыбы (щуки). Было подсчитано число лучей в хвостовых плавниках щуки. По уровню значимости 0.1 определить: одинаковы или нет средние размеры в пробах.

Первая выборка

53	51	52	55	56	49
54	53	52	53	51	55
51	51	56	54	54	53
52	55	53	53	56	53
56	55	50	54	49	54
52	51	55	52		

Вторая выборка

56	49	51	52	54	56
51	55	53	55	53	54
54	53	54	54	55	53
56	53	52	56	52	52
49	54	54	55	54	55
55	54	51			

Р.3 Сравнить среднее двух независимых выборок методом Стьюдента по уровню значимости $\alpha=0.075$.

выборка X

12.6	13.4	12.4	13.3	13.1	12.0	11.9	11.3	15.0	16.4	14.4	14.9	14.4
12.8	12.6	12.5	12.0	12.4	12.4	12.4	11.9	14.3	15.2	16.1	16.6	14.7
13.2	12.9	8.5	9.8	10.7	10.4	10.6	13.9	14.2	13.1	15.3	13.1	13.5

выборка Y

9.6	10.2	14.3	14.3	15.3	14.5	17.6	17.9	17.8	11.3	14.4	13.6	11.6
15.4	16.8	11.0	11.2	16.6	15.9	11.1	11.7	16.1	11.6	10.0	12.6	14.8
12.4	14.5	12.1	11.9	17.5	16.4	12.3	17.8	13.5	14.1	13.4		

Р.4 Используя критерий Крамера-Уэлча выполнить сравнение средних значений двух независимых выборок на уровне значимости α=0.05.

выборка X						
18	14	15	19	12	11	8
14	12	13	7	15	19	5
13	20	11	8	16	18	7
18	16	18	11	13	9	13
13	11	10	12	14	6	8
13	13	8	15	14	15	15
15	7	20	16	10	12	

выборка Y						
12	19	15	18	17	14	18
15	7	18	14	19	12	12
16	8	12	13	16	20	15
13	11	20	18	12	16	9
14	12	16	13	15	11	14
14	15	11	13	19	13	
10	16	13	15	13	7	

Р.5 Четверо студентов-химиков провели одни и те же опыты. Можно или нет объединить результаты этих опытов? Для анализа используйте уровень значимости α=0.05.

	опыт 1	опыт 2	опыт 3	опыт 4	опыт 5
студент 1	14.0	14.5	13.7	12.7	14.1
студент 2	14.1	10.1	14.7	13.7	14.0
студент 3	14.0	12.3	12.8	11.0	13.1
студент 4	14.5	14.2	15.0	14.7	13.5

Р.6 Для проверки эффективности новой развивающей программы были созданы две группы детей шестилетнего возраста. Одна группа (экспериментальная) занималась по новой программе, вторая (контрольная) – по старой.

После эксперимента дети обеих групп были протестированы по методике Керна-Йерасика (школьная зрелость). Результаты тестирования по вербальной шкале занесены в таблицу. Можно ли сделать заключение об эффективности новой программы и ее преимуществе перед старой.

Результаты тестирования по вербальной шкале (сырые баллы)		
№ исп.	эксп.	контр.
1	14	13
2	13	13
3	11	14
4	8	12
5	12	14
6	13	14
7	13	12
8	13	13
9	11	15
10	12	13
11	14	11
12	13	12
13	12	14
14	14	9
15	10	14

Р.7 Была исследована группа детей с заболеванием крови до лечения препаратами и после лечения. В таблицу занесены показатели er крови по результатам медицинского обследования.

Сделать анализ результативности лечения данным препаратом, используя статистические критерии.

Лабораторные данные (результаты обследования детей с ОЛЛ)		
пациент	до лечения	после лечения
1.	4.20	2.65
2.	2.00	2.38
3.	2.33	3.50
4.	2.40	3.50
5.	1.80	4.80
6.	0.80	4.00
7.	2.00	3.70
8.	2.10	4.12
9.	2.80	4.20
10.	1.46	2.89
11.	3.85	3.70
12.	1.64	2.58
13.	2.60	3.20
14.	1.70	2.10

Р.8 Разбейте выборку на классы, рассчитайте частоты признака для каждого класса

11.6	11.3	12.0	10.7	11.3	11.4	11.1	11.2	14.9	14.4	15.3	10.2	13.5
13.1	12.9	13.1	12.3	12.6	12.9	13.8	14.7	16.1	16.6	14.7	14.4	15.4
8.5	9.9	9.8	10.4	11.7	13.4	14.7	13.1	12.8	13.2	12.6	13.1	12.4
14.7	16.3	15.3	15.5	14.4	15.6	15.5	14.8	14.3	14.3	15.3	14.5	13.4
15.0	15.2	15.4	9.8	13.2	12.6	13.1	9.8	11.0	11.2	16.6	15.9	11.6
14.2	13.8	13.9	14.5	17.0	17.4	17.3	11.1	12.1	11.9	17.5	16.4	13.1
11.2	11.1	8.5	17.0	12.2	11.1	14.9	12.7	12.6	14.8	14.4	13.6	14.2
12.7	12.1	17.8	17.5	11.3	16.5	13.2	13.9	9.6	10.0	14.5	16.8	13.9
12.6	15.4	14.1	13.9	12.9	13.6	12.2	15.2	12.9	12.6	13.4	9.8	10.6
9.0	10.6	10.3	10.0	10.4	13.2	12.6	13.1	17.6	17.9	17.8	11.3	10.4
12.4	13.3	13.1	12.0	11.9	11.3	15.0	16.4	11.1	11.7	16.1	11.6	10.7
12.5	12.0	12.4	12.4	12.4	11.9	14.3	15.2	12.3	17.8	13.5	14.1	9.8

Р.9 Чтобы определить пищевые предпочтения кошек перед ними ставились чашки с различными кормами. В таблице представлено количество подходов к чашкам.

Корм	Вискас рыбный	Вискас куриный	корм "Прима"	корм "Джеба"	Китти-кэт
Количество подходов	14	5	8	6	12

Имеются ли предпочтения кошки к какому-либо корму? Использовать уровень значимости $\alpha = 0.05$.

Р.10 Разбейте выборку на классы, рассчитайте частоты признака для каждого класса

10.5	10.9	10.5	11.4	11.9	10.7	10.9	11.5	10.2	9.7	10.6	16.8
12.8	11.2	12.3	12.7	11.6	12.9	14.9	14.3	14.3	14.3	15.3	14.5
7.7	10.3	9.7	11.2	5.7	11.2	11.8	14.5	11.0	11.2	16.6	15.9
14.2	13.8	15.9	14.6	18.1	18.5	16.8	12.2	12.1	11.9	17.5	16.4
11.7	10.6	16.9	15.7	11.7	12.6	15.2	12.4	12.6	14.8	14.4	13.6
12.4	12.2	16.7	15.8	11.3	17.2	12.3	13.5	9.6	10.0	14.5	11.3
12.1	14.7	13.8	14.7	12.2	12.6	12.4	15.5	11.6	13.4	12.4	15.4
12.3	17.8	13.5	14.1	10.2	16.8	10.2	14.5	11.1	11.7	16.1	11.6
17.6	17.9	17.8									

Р.11 Измерялись размеры зерен. Выборка, содержащая 9570 элементов, разбитая на 10 одинаковых по величине интервалов, представлена таблицей.

Проверить гипотезу о нормальности распределения полученных измерений при уровне значимости $\alpha = 0.05$, приняв за параметры гипотетического распределения: математическое ожидание и дисперсию, их оценки, вычисленные по экспериментальным данным.

Интервал	Частота
[9.0; 8.8)	52
[8.8; 8.6)	148
[8.6; 8.4)	400
[8.4; 8.2)	1483
[8.2; 8.0)	2742
[8.0; 7.8)	2579
[7.8; 7.6)	1397
[7.6; 7.4)	530
[7.4; 7.2)	170
[7.2; 7.0)	69

Р.12 Сравнить среднее двух независимых выборок методом Стьюдента по уровню значимости $\alpha=0.05$.

выборка X

10.5	10.9	10.5	11.4	11.9	10.7	10.9	11.5	11.2	5.7	11.2	11.8
12.8	11.2	12.3	12.7	11.6	12.9	14.9	14.3	7.7	10.3	9.7	

выборка Y

14.2	13.8	15.9	14.6	18.1	18.5	16.8	12.2	12.2	12.6	12.4	15.5
11.7	10.6	16.9	15.7	11.7	12.6	15.2	12.4	12.1	14.7	13.8	14.7
12.4	12.2	16.7	15.8	11.3	17.2	12.3	13.5	10.2	9.7	10.6	

Р.13 Проводились исследования с целью выяснить, влияет ли прием нифедипина (препарат, обладающий способностью расширять сосуды) на среднее артериальное давление (мм.рт.ст.) после приема кокаина. После введения кокаина, собакам вводили физиологический раствор либо нифедипин. Были получены следующие данные*:

физ.раствор – 156, 171, 133, 102, 129, 150, 120,110,112, 130, 105;
нифедипин – 73, 81, 103, 88, 130, 106, 106, 111, 122, 108, 99.

Влияет ли нифедипин на среднее артериальное давление?

Использовать уровень значимости $\alpha=0.05$.

* Гланц С. Медико-биологическая статистика / Пер. с англ. М.: Практика, 1998.

Р.14 Определялся оптимальный состав фотохромного стекла в системе $Li_2O–Al_2O_3–SiO_2$. В качестве параметра оптимизации (Y) рассматривалась оптическая плотность в облученном состоянии. Надо определить состав стекла и условия его варки, обеспечивающие максимальную плотность. В качестве независимых факторов были выбраны: Z_1 – температура варки (1300÷1350°C); Z_2 – время выдержки (1.5÷2 час); Z_3 – содержание A_2O_3 (0.124÷0.157 мол. доли). Результаты эксперимента представлены в таблице.

Матрица планирования эксперимента

	x_1	x_2	x_3	y_1	y_2	y_3
1	+1	+1	+1	12.22	12.22	13.08
2	-1	+1	+1	14.27	14.43	13.67
3	+1	-1	+1	17.65	18.92	18.28
4	-1	-1	+1	18.51	17.71	19.10
5	+1	+1	-1	10.48	9.02	10.58
6	-1	+1	-1	10.24	10.74	10.34
7	+1	-1	-1	14.85	14.56	14.40
8	-1	-1	-1	15.91	15.50	16.39

1. Построить линейное уравнение регрессии.
2. Провести проверку однородности параллельных опытов.
3. Выполнить оценку дисперсии воспроизводимости (экспериментов).
4. Сделать проверку адекватности уравнения регрессии экспериментальным данным.
5. Уравнение представить (записать) в естественных переменных.

Р.15 Исследовалось влияние на выход полимера *Y*(кг) – температуры Z_1 (160÷170°С), концентрации инициатора Z_2 (8÷15 %), продолжительности реакции Z_3 (1÷5 ч). Результаты эксперимента представлены в таблице.

Матрица планирования эксперимента

	x_1	x_2	x_3	y_1	y_2	y_3
1	+1	+1	+1	25.01	27.86	25.19
2	-1	+1	+1	26.90	28.88	30.09
3	+1	-1	+1	22.34	23.65	21.14
4	-1	-1	+1	24.18	25.89	24.65
5	+1	+1	-1	18.16	18.10	19.68
6	-1	+1	-1	19.83	19.58	20.30
7	+1	-1	-1	13.99	14.86	14.64
8	-1	-1	-1	17.40	16.76	17.15

1. Построить линейное уравнение регрессии.
2. Провести проверку однородности параллельных опытов.
3. Выполнить оценку дисперсии воспроизводимости (экспериментов).
4. Сделать проверку адекватности уравнения регрессии экспериментальным данным.
5. Уравнение представить (записать) в естественных переменных.

Р.16 Построить (сформулировать) математическую модель процесса образования окрашенного никеля с диметилглиоксимом в присутствии железа (II).

В качестве изучаемых факторов выбраны:

Z_1 – длина волны, нм,

Z_2 – pH раствора,

Z_3 – концентрация $(NH_4)_2S_2O_8$, %.

Параметр оптимизации – оптическая плотность раствора.

Указание: при статистическом анализе использовать уровень значимости α=0.05; считать, что кодированные фактора взаимодействия следующие

$x_4 = x_1x_2$; $x_5 = x_1x_3$; $x_6 = x_2x_3$; $x_7 = x_1x_2x_3$.

Условия планирования

Характеристики плана	Факторы		
	Z_1	Z_2	Z_3
Нулевой уровень	530	9	5
Интервал варьирования	10	1	2
Верхний уровень	540	10	7
Нижний уровень	520	8	3

Матрица планирования и результаты эксперимента

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2
1	+	—	—	+	+	—	—	+	0.2860	0.1260
2	+	—	+	+	—	—	+	—	0.4450	0.5050
3	+	+	—	+	—	+	—	—	0.2210	0.0610
4	+	+	+	+	+	+	+	+	0.1330	0.2930
5	+	—	—	—	+	+	+	—	0.0720	0.1520
6	+	—	+	—	—	+	—	+	0.3820	0.2620
7	+	+	—	—	—	—	+	+	0.0366	0.1140
8	+	+	+	—	+	—	—	—	0.1770	0.3170

1. Построить линейное уравнение регрессии.
2. Провести проверку однородности параллельных опытов.
3. Выполнить оценку дисперсии воспроизводимости (экспериментов).
4. Сделать проверку адекватности уравнения регрессии экспериментальным данным.
5. Уравнение представить (записать) в естественных переменных.

Р.17 Результаты измерений размера детенышей в выборке 1000 самок представлены в таблице. В первой строке указаны размеры детенышей, во второй – частоты появления детенышей соответствующих размеров.

Размеры	98.0	98.5	99.0	99.5	100	100.5	101.0	101.5	102.0	102.5
Частоты	21	47	87	158	181	201	142	97	41	25

Проверить при помощи критериев согласия гипотезу: выборочное распределение имеет нормальный закон распределения, при уровне значимости α = 0.05. За параметры нормального закона распределения принять их оценки, вычисленные по экспериментальным данным.

Р.18 Было определено количество детей n_i и суточное время t_i, посвященное чтению:

t_i, мин	15	20	25	30	35	40	45	50	55
n_i	6	13	38	74	106	85	30	10	4

При уровне значимости 0.05 проверить гипотезу о нормальном распределении времени чтения по критерию согласия Пирсона.

Р.19 Была проанализирована сменная выработка рабочих; эмпирическое распределение выборки приведено в таблице:

$x_i - x_{i+1}$	3-8	8-13	13-18	18-23	23-28	28-33	33-38
n_i	6	8	15	40	16	8	7

Используя критерий Колмогорова, при уровне значимости 0.01 проверить, согласуется ли гипотеза о нормальном распределении выработки с эмпирическим распределением выборки.

Р.20 В течение месяца на двух участках сборочного цеха выборочно проводились моментные наблюдения с целью выявления потерь времени, данные которых представлены в таблицах:

x_i, мин	14	17	20	23	26	29	32	35	38	41
n_i	2	4	10	15	20	27	18	16	8	5

y_i, мин	16	20	24	28	32	36	40	44
n_i	3	9	12	17	16	13	7	3

Можно ли считать, что при уровне значимости 0.05 по результатам проверок на двух участках потери времени описываются одной и той же функцией распределения, т.е. являются устойчивым и закономерным процессом в данном цехе? Использовать критерий Колмогорова-Смирнова.

Р.21 Имеются две выборки (в условных единицах) усвоения некого препарата, апробируемого на двух группах A и B животных. На уровне значимости 0.01 проверить нулевую гипотезу H_0: $F_1(x) = F_2(x)$ об однородности двух выборок

A	3	4	6	10	13	17	
B	1	2	5	7	16	20	22

Р.22 Контрольную работу по высшей математике выполняли студенты двух групп. В первой группе из 100 задач студенты правильно решили 56 задач, а во второй группе из 120 задач ошиблись в 56 задачах. При уровне значимости 0.01 проверить гипотезу о том, что материал одинаково усвоен студентами обеих групп.

Каталог примеров

Пример 1.1 Для выборки, заданной в виде вариационного ряда, строится функция распределения в дифференциальной и интегральной формах.

Пример 1.2 Для выборки на заданном уровне значимости рассчитывается величина доверительного интервала (выборка считается и большой и малой).

Пример 1.3 Для выборки (заданы данные) строится полигон частот.

Пример 2.1 Сравниваются дисперсий двух выборок по критерию Фишера.

Пример 2.2 Проверятся однородность (дисперсий) результатов измерений шести серий (по пять опытов в серии) по критерию Кохрена.

Пример 2.3 Проверятся однородность (дисперсий) результатов измерений шести серий опытов (в каждой серии от 5 до 7 измерений) по критерию Бартлетта.

Пример 2.4 Определяется, является ли действие методики преподавания на уровень усвоения материала статистически значимым (на заданном уровне); используется критерий знаков.

Пример 2.5 Для условий примера 2.4 проводится решение по модифицированному алгоритму вычислений для критерия знаков.

Пример 2.6 В разных классах используется четыре метода обучения. Требуется определить наличие существенной разницы между показателями усвоения материала на уровне значимости 0.05 для заданной таблицы результатов тестирования. Используется тест Левена.

Пример 2.7. Проведены измерения в контрольной (КГ) и экспериментальной группах (ЭГ) в начале и конце периода, когда в ЭГ использовалась некоторая новая методика обучения. Выполняется сравнительный анализ уровней знаний со значимостью 0.05 характеристик выборок контрольной и экспериментальных групп в соответствии с заданной порядковой трехбалльной шкалой.

Пример 2.8. Изучалось влияние некого лекарства на выздоровление пациентов для данных, представленных в дихотомической шкале. Проверяется достоверность совпадений и различий выборок на уровне значимости 0.05 и определяется эффективность (положительность воздействия) препарата.

Пример 2.9. Выполняется сравнительный анализ различий знаний учащихся контрольной и экспериментальной групп (до и после применения некой методики) для дихотомической шкалы при оценке по двум уровням – "усвоили или не усвоили материал" (по число правильно решенных задач, большему или меньшему 10). Определяется достоверность различий состояний экспериментальной и контрольной групп после окончания эксперимента.

Пример 3.1 Проверка соответствия выборки (задана таблицей) нормальному закону распределения (используется критерий χ^2 дискретной вариации).

Пример 3.2 Оценка соответствия нормальному распределению для выборки, заданной в форме таблицы классов.

Пример 3.3 Сравниваются частоты (цветовые предпочтения) животного равномерному распределению (используется критерий χ^2 дискретной вариации).

Пример 3.4 Для выборки (измерений размера некоторого объекта) на заданном уровне значимости проверяется гипотеза о соответствии размеров распределению нормального закона. Используется критерий Колмогорова.

Пример 3.5 На заданном уровне значимости анализируется равенство двух эмпирических выборок, заданных вариационными частотными рядами. Используется критерий Колмогорова-Смирнова.

Пример 3.6 Оценивается соответствия ограниченной выборки (20 значений) нормальному распределению по критерию омега-квадрат (критерию Крамера-Мизеса-Смирнова).

Пример 3.7 Оценка соответствия ограниченной выборки (20 значений) нормальному распределению методом моментов.

Пример 4.1 Проверка равенства средних значений двух выборок парным (зависимых выборок) критерием Стьюдента.

Пример 4.2 Проверка равенства средних значений двух выборок (гомоскедастический тест) критерием Стьюдента.

Пример 4.3 Проверка равенства средних значений двух выборок (гетероскедастический тест) критерием Стьюдента.

Пример 4.4 Критерий Стьюдента равенства среднего значения математическому ожиданию. Задача по определении соответствия образца никеля паспортной величине. Анализируется присутствия в методике систематической погрешности.

Пример 4.5 Использование z-критерия для сравнения эффективности двух различных методик лечения.

Пример 4.6 По результатам контрольной работу выясняется, можно ли считать, что различия в усвоении учебного материала студентами четырех групп первого курса существенны. Используется *z*-критерий.

Пример 4.7 Определяется наличие существенной разницы между показателями усвоения материала на заданном уровне значимости для четырех методов обучения в разных классах. Используется однофакторный анализ ANOVA.

Пример 4.8 Апостериорный тест Шеффе используется после отклонения ANOVA нулевой гипотезы и оценивается значимость различий средних по содержание влаги после хранения пакетов удобрений для трех различных методов хранения.

Пример 5.1 Непараметрический критерий Крамера-Уэлча: сравнение средних значений двух независимых выборок.

Пример 5.2 По некому индексу оценивались последствия воздействия некого препарата на три группы испытуемых. Используется дисперсионный анализ по непараметрическому критерию Краскела-Уоллиса (Kruskal-Wallis test).

Пример 5.3 Используется непараметрический критерий Вилкоксона-Манна-Уитни для оценки различий между экспериментальной и контрольной группами по уровню знаний в шкале отношений по результатам проведения тестов.

Пример 5.4 На заданном уровне значимости сравниваются две связанные выборки (эксперимент типа "до" и "после"). Определяется, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Пример 6.1 На конкретных данных дан алгоритм построения и исследования показателей уравнения линейной регрессии.

Пример 6.2 Определяются константы a и b для аппроксимирующего уравнения $y = 1/(ax^n + b)$ по известной таблице данных для значения n=1.75.

Пример 6.3 Определяются константы a, b и n для аппроксимирующего уравнения $y = 1/(ax^n + b)$ по известной таблице данных.

Пример 6.4 Определяются константа и порядок простой химической реакции для экспериментально полученной зависимости концентрация бромистого нитрозила в реакции его разложения.

Пример 6.5 Оценивается диагностическая значимость личного опросника – корреляция между типичным ответом на отдельный пункт с общим результатом теста.

Пример 6.6 Определяются диагностическая значимость заданий и уровень корреляции между заданиями для данных 10 испытуемых по 8 заданиям в виде упорядоченной бинарной матрицы.

Пример 6.7 По заданным данным рассчитывается коэффициент корреляции Спирмена и проверяется его значимость. Указывается доверительный интервал для коэффициента корреляции.

Пример 6.8. Для экспериментальных данных (концентрация/сигнал) определяются значения наименьшей концентрации обнаружения вещества $C_{\min}$ и $C_{\text{над}}$ – предел "надёжного обнаружения".

Пример 7.1 Средствами таблицы сопряженности 2×2 определяется зависимость успешности прохождения студентами интернет-тестирования от прохождения ими адаптивного курса. Используется критерия χ^2. Приводится решение этой же задачи с учетом поправки Yates'a.

Пример 7.2 Для данных примера 7.1 дано решение задачи с поправкой на правдоподобие.

Пример 7.3 Для данных примера 7.1 дано решение задачи с использованием критерия V Крамера. Основная часть схемы вычислений повторяет таковую для вышеприведенных примеров.

Пример 7.4 Приводится пример анализа сопряженности двух качественных признаков 2×2. Определяется наличие взаимосвязи между признаками (приемом медикамента и болезнью).

Пример 7.5 Определяется зависимость успешности прохождения студентами интернет-тестирования от прохождения ими адаптивного курса в начале обучения в вузе. Используется "классический" алгоритм точного теста Фишера.

Пример 7.6 Определяется зависимость успешности прохождения студентами интернет-тестирования по математике от прохождения ими адаптивного курса в начале обучения в вузе. Используется упрощенный алгоритм точного одностороннего теста Фишера.

Пример 7.7 Для данных примера 7.6 рассчитывается двухсторонний критерий точного теста Фишера.

Пример 7.8 Определяется зависимость успешности прохождения студентами интернет-тестирования от прохождения ими адаптивного курса в начале обучения в вузе для задачи типа примера 7.1. Рассчитываются одно- и двухсторонний критерии точного теста Фишера.

Пример 7.9 Средствами таблицы сопряженности 4×3 определяется влияние темперамента на профессиональные предпочтения. Используется критерия χ^2.

Пример 7.10 Средствами таблицы сопряженности 5×2 определяется наличие изменений распределения сочетаний инверсий *Anopheles messeae* за три года. Используется критерия χ^2.

Пример 7.11 Критерием χ^2 сравнивается полученное экспериментально соотношение 135:51:54:18 гибридов с фенотипическими классами 9:3:3:1 при учете двух пар альтернативных признаков.

Пример 7.12 Рассчитываются риски и шансы заболеваемости для женщин по материалам исследования некого заболевания на базе 300 респондентов мужчин и женщин.

Пример 8.1 По критерию Романовского на заданном уровне значимости проводится проверка на наличие грубых ошибок по отсортированным по убыванию данным.

Пример 8.2 По критерию Шарлье проводится проверка на наличие грубых ошибок данных, полученных при измерении размеров листа акации.

Пример 8.3 По критерию "ящик с усами" проводится проверка на наличие грубых ошибок данных, полученных при измерении размеров листа акации.

Пример 8.4 По правилу Томпсона (критерию Рошера) проводится проверка на наличие грубых ошибок данных, полученных при измерении размеров листа акации.

Пример 8.5 По критерию Диксона (Q-критерию) проводится анализа наличия одностороннего выброса в данных с использованием "обслуживающего" листа.

Пример 9.1 Строится план дробного факторного эксперимента 2^{4-1} и определяются коэффициенты уравнения регрессии для заданных значений эксперимента.

Пример 9.2 Планируется полный факторный эксперимент: строится математическая модель для зависимости оптической плотности от технологических факторов.

Пример П2.1 Дан алгоритм решения задачи по приведению выборки к форме таблицы классов.

Пример П2.2 Приведены расчетные формулы определения среднего, медианы, моды и параметров вариации по данным в форме интервального ранжированного частотного ряда.

Пример П4.1 Неким прибором, класс точности которого К = 2,5 с диапазоном шкалы 0-100, измерено 10 отсчетов величины X. Необходимо дать результат измерения для уровня значимости α=0.05 .

Пример П4.2 В трех различных условиях измерена длина некого объекта L. Результаты измерений представлены в виде: $L_1 = 10 \pm 3$ см; $L_2 = 11 \pm 2$ см; $L_3 = 10 \pm 2$ см . Необходимо объединить эти измерения и определить общую погрешность.

Пример П4.3 Прямыми измерениями найдены значения радиуса r и линейной скорости v равномерного вращения материальной точки массы m. Необходимо оценить по заданному соотношению значение центробежной силы, действующей на этот материальный объект, по результатам прямых измерений.

Критерии (кр.), методы и формулы

Терминология

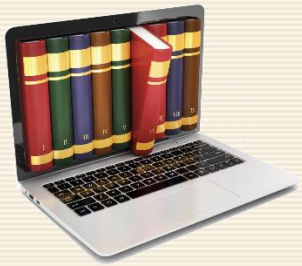
Абсолютная погрешность
Адекватность
Амплитуда ряда
Аппроксимация
Асимметрия

Валидность процедуры измерения
Варианса
Варианта
Варианты: классическое понимание
Вариации размах
Вариационный ряд
Вариационный ряд: выборочное среднее
Вариационный ряд: медиана
Вариационный ряд: мода
Вариационный ряд: оценка дисперсии
Вариационный ряд: оценка среднего отклонения
Вариационных рядов виды
Величина случайная
Вероятности *P*-значение, *P*-уровень
Выборка
Выборка репрезентативная
Выборки зависимые
Выборки интерквартильный размах
Выборки межквартильный размах
Выборки однородные
Выборки парные
Выборки размах
Выборки характеристики
Выборки элемент
Выборочное среднее значение
Выброс

Генеральная совокупность
Генерирующее соотношение
Гипотеза альтернативная
Гипотеза научная
Гипотеза научная статистическая
Гипотеза однородности основная
Гипотеза статистическая
Гипотезы статистической алгоритм проверки
Гистограмма
Гистограмма частостей
Гистограмма частот
Градуировка
Группировка

Данные динамические
Данные динамические: цикличность и сезонность
Данные дискретные
Данные количественные
Данные непрерывные
Данные номинальные и порядковые

Перечень использованных источников



1. Агапова Е.Г., Битехтина Е.А. Обработка экспериментальных данных в MS Excel: методические указания к выполнению лабораторных работ для студентов дневной формы обучения / Хабаровск: Изд-во Тихоокеан. гос. ун-та, 2012. 32 с.
2. Архипов В.А., Березиков А.П. Основы теории инженерно-физического эксперимента: учебное пособие / Федеральное агентство по образованию, Гос. образовательное учреждение высш. проф. образования "Томский политехнический ун-т". Томск, 2008. 205 с.
3. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента в химической технологии / Москва, Высш. шк., 1985. 327 с.
4. Бондарчук С.С., Годованная И.Г., Перевозкин В.П. Основы практической биостатистики / М-во образования Российской Федерации, Гос. образовательное учреждение высшего проф. образования "Томский государственный педагогический университет" (ТГПУ). Томск, 2009. 130 с.
5. Бондарчук И.С., Курзина И.А., Бондарчук С.С. Методология решения задач физической химии инструментом Solver MS Excel // Высшее образование сегодня, 2014. № 9. С. 22-24.
6. Бондарчук И.С., Федорова В.А. Алгоритмы идентификации кинетических параметров простых реакций // Перспективы развития фундаментальных наук: сборник на учных трудов XI Международной конференции студентов и молодых ученых / ред. Е.А. Вайтулевич. Национальный исследовательский Томский политехнический университет, 2014. С. 555-557.
7. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя / Санкт-Петербург, Питер, 2008. 608 с.

8. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / Москва, Юрайт, 2016. 479 с.

9. Горбунова А.А., Лемешко Б.Ю., Лемешко С.Б. Критерии проверки гипотез об однородности дисперсий при наблюдаемых законах, отличных от нормального // Материалы X международной конференции "Актуальные проблемы электронного приборостроения" АПЭП-2010. Т.6, Новосибирск, 2010. С.36-41.

10. Езепов Д.А. Видеокурс "Продвинутый уровень MS Excel" URL: https://statanaliz.info/, Электронная книга "Трюки Excel". URL: https://statanaliz.info/kniga-tryuki-excel (дата обращения: 08.05.2018).

11. Жигунов В.В., Ростовцев Р.Н., Бурцева Ю.В., Жигунов К.В., Якунова Е.В. Методы обработки экспериментальных данных: учебное пособие. Тула, Изд-во ТулГУ, 2016. 78 с.

12. Заляжных В.В. Критерий Диксона. URL: http://arhiuch.ru/lab5.html (дата обращения: 08.05.2018).

13. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. Статистика знает всё / Изд-во URSS, 2017. 608 с.

14. Лакин Г.Ф. Биометрия: Учеб. пособие для биол. спец. вузов. Москва, Высш. шк., 1990. 352 с.

15. Леонов В.П. Биометрика – журнал для медиков и биологов, сторонников доказательной биомедицины. URL: http://www.biometrica.tomsk.ru/freq1.htm (дата обращения: 08.05.2018).

16. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учеб. пособие / Москва, Финансы и статистика, 2002. 368 с.

17. Минько А.А. Статистический анализ в MS Excel / Москва, Издательский дом "Вильямс", 2004. 448 с.

18. Никипорец Э.Н., Парамонова Л.А., Черновский Н.М. Сборник задач по взаимозаменяемости и метрологическому обеспечению в авиационной технике: Учебное пособие / Москва, Изд-во МАИ, 1990. 108 с.

19. Новиков Д.А. Статистические методы в педагогических исследованиях (типовые случаи) / Москва, МЗ-Пресс, 2004. 67 с.
20. Практическое руководство по статистическому анализу в Excel Real Statistics Using Excel. URL: http://www.real-statistics.com/one-way-analysis-of-variance-anova/basic-concepts-anova (дата обращения: 08.05.2018).
21. Халафян А.А., Боровиков В.П., Калайдина Г.В. Теория вероятностей, математическая статистика и анализ данных: Основы теории и практика на компьютере / Изд-во URSS, 2017. 320 с.
22. Ходасевич Г.Б. Обработка экспериментальных данных на ЭВМ. Ч. 1. Обработка одномерных данных / Санкт-Петербург, Санкт-Петербургский гос. ун-т телекоммуникаций им. проф. М.А. Бонч-Бруевича, 2000. 100 с.
23. Шкляр В.Н. Планирование эксперимента и обработка результатов / Изд-во Томского политехнического университета, 2010. 89 с.
24. Юрьева Т.А., Филимонова А.П., Чалкина Н.А. Статистическая оценка связи между качественными признаками в педагогических исследованиях // Вестник Амурского государственного университета. 2014. Вып. 65: Сер. Естеств. и экон. науки. С. 11-16.
25. Parker R.A., Rea L.M. Designing and Conducting Survey Research: A Comprehensive Guide **/** 4 edition, Jossey-Bass, 2014. 360 p.
26. Pearson E. S. The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table // Biometrika, 1947. Vol. 34. P. 139–167.
27. Wilkinson L. Statistical methods in psychology journals: guidelines and explanations // American Psychologist, 1999. Vol. 54. P. 594–604.
28. Yates F. Contingency tables involving small numbers and the chi-square test // Supplement to the Journal of the Royal Statistical Society, 1934. Vol. 1. P. 222.

Учебное издание

Сергей Сергеевич Бондарчук
Иван Сергеевич Бондарчук

СТАТОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В MS EXCEL

Учебное пособие

Редактор: Г.В. Белозерова
Ответственный за выпуск: Л.В. Домбраускайте

Печать: трафаретная
Бумага: офсетная
Усл. печ. л.: 25.2
Уч. изд. л.: 10.2

Сдано в печать: 24.06.2018
Формат: 60×84/16
Заказ: 1381/у
Тираж: 500 экз.

Издательство Томского государственного
педагогического университета
634041, г. Томск, пр. Комсомольский, 75
Отпечатано в типографии Издательства ТГПУ,
г. Томск, ул. Герцена, 49. Тел. (3822) 52-12-93
e-mail: publish@tspu.edu.ru

9 785894 288611