Сравнительная статистика в медицинских исследованиях

Врач-исследователь постоянно сопоставляет группы (например, новый препарат vs. стандартная терапия). Главная цель — не просто описать численные расхождения, а ответить, являются ли они **статистически значимыми**.

Статистическая значимость означает, что различия вряд ли случайны и будут воспроизведены при повторных исследованиях. Это превращает наблюдение в научно обоснованный вывод.

Для корректной оценки значимости различий необходимо выбрать подходящий статистический метод. Этот выбор определяется тремя фундаментальными факторами:

1

Число сравниваемых групп

Базовые методы для двух групп, дисперсионный анализ для трех и более.

2

Тип выборок

Независимые (разные, не связанные группы) или зависимые/парные (одни и те же объекты в разных условиях).

3

Вид и тип распределения признака

Количественный (непрерывный) или качественный (категориальный) признак, а также его соответствие нормальному распределению.

Ответы на эти вопросы формируют логическую схему для выбора критерия (t-критерий, Манна-Уитни, χ^2).

Гипотезы и уровень иначимости

Нулевая гипотеза (H_0)

Гипотеза «об отсутствии эффекта». Предполагает, что наблюдаемые различия — результат случайной вариации, а не систематического эффекта. **Пример:** Эффективность препарата А не отличается от Б.

Альтернативная гипотеза (H_1)

Логическое отрицание H_0. Постулирует, что различия не случайны и отражают реальный эффект. **Задача:** Собрать свидетельства, чтобы отклонить H_0 в пользу H_1.

P-Value

Решение об отклонении H_0 принимается на основе **p-value** (уровень значимости, α), который quantifies вероятность получить наблюдаемые результаты, если H_0 верна.

0.05

Пороговое Значение (α)

Общепринятый допустимый риск ошибки первого рода (ошибочно отвергнуть верную H_0).

$p < \alpha$

Значимые Различия

Отвергаем H_0. Вероятность случайности мала. Вывод в пользу реальности эффекта.

$p \geq \alpha$

Незначимые Различия

Нет оснований для отклонения H_0. Наблюдаемые различия признаются статистически незначимыми.

Классификация выборок: зависимые vs. независимые

Правильная классификация критически важна для выбора корректного математического аппарата.

Независимые выборки

- Формируются из разных, не связанных групп испытуемых.
- Отбор в одну группу не зависит от отбора в другую (часто через рандомизацию).
- **Пример:** Сравнение холестерина у группы на статине и группы на плацебо.

Зависимые (парные) выборки

- Данные получены от одних и тех же объектов в разных условиях или в разное время.
- Между измерениями существует постоянная связь.
- **Пример:** Измерение давления у пациентов до и после приема лекарства.

> Зависимые выборки исключают межиндивидуальную вариабельность, делая анализ более чувствительным к эффекту воздействия.

Параметрические vs. непараметрические критерии

Параметрические критерии

- Основаны на параметрах распределения (среднее, дисперсия).
- Требуют **нормального распределения** признака (проверка Шапиро-Уилка/Колмогорова-Смирнова).
- Высокая статистическая мощность.
- **Пример:** t-критерий Стьюдента.

Непараметрические критерии

- Основаны на анализе **рангов** (порядковых номеров).
- Не требуют нормального распределения (универсальны).
- Используются для малых выборок ($n < 30$), порядковых данных, или при наличии выбросов.
- **Пример:** Критерии Манна-Уитни, Вилкоксона.

Критерий Стьюдента (t-критерий)

1

Независимые выборки

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

M_1 и M_2 — выборочные средние значения сравниваемых групп, а m_1 и m_2 — их стандартные ошибки, вычисляемые как отношение стандартного отклонения (σ) к квадратному корню из объема выборки ($m = \sigma/\sqrt{n}$). Количество степеней свободы для этого варианта расчета определяется по формуле $df = n_1 + n_2 - 2$.

2

Зависимые выборки

$$t = d / m\_d.$$

Здесь d представляет собой среднюю арифметическую разность между парными измерениями, а m_d — стандартную ошибку этой средней разности. Число степеней свободы в данном случае равно $df = n - 1$, где n — количество пар наблюдений

Критерий Стьюдента (t-критерий)

1

Независимые выборки

Предположим, мы проводим исследование, в котором сравнивается эффективность двух гипотензивных препаратов (А и Б). Группа пациентов была случайным образом разделена на две независимые группы:

Группа 1 (n_1=15): получала препарат А.

Группа 2 (n_2=12): получала препарат Б.

После курса лечения было зафиксировано среднее диастолическое артериальное давление (ДАД) в каждой группе. Наша цель — определить, существует ли статистически значимое различие в эффективности этих препаратов, то есть различаются ли средние значения ДАД между группами.

Критерий Стьюдента (t-критерий)

1

Независимые выборки

Исходные данные:
Группа 1 (Препарат А):

Объем выборки: $n_1 = 15$

Среднее значение ДАД: $M_1 = 82.4$ мм рт. ст.

Стандартное отклонение: $\sigma_1 = 5.6$

Группа 2 (Препарат Б):

Объем выборки: $n_2 = 12$

Среднее значение ДАД: $M_2 = 87.9$ мм рт. ст.

Стандартное отклонение: $\sigma_2 = 6.1$

Предварительное условие: Предполагается, что распределение ДАД в обеих группах является нормальным (проверено, например, с помощью критерия Шапиро-Уилка).

Критерий Стьюдента (t-критерий)

1

Независимые выборки

Шаг 1: Формулировка гипотез
Нулевая гипотеза (H_0): Средние значения ДАД в группах, получавших препарат А и препарат Б, не различаются.

$M_1 = M_2$ или $M_1 - M_2 = 0$

Альтернативная гипотеза (H_1): Средние значения ДАД в группах, получавших препарат А и препарат Б, различаются.

$M_1 \neq M_2$ (используется двусторонний критерий, так как мы не утверждаем заранее, какой препарат лучше).

Уровень значимости принимаем стандартным: $\alpha = 0.05$.

Шаг 2: Расчет стандартных ошибок средних (m)
Стандартная ошибка среднего показывает, насколько точно выборочное среднее приближает генеральное.

Для Группы 1:
$m_1 = \sigma_1 / \sqrt{n_1} = 5.6 / \sqrt{15} \approx 5.6 / 3.873 \approx 1.446$

Для Группы 2:
$m_2 = \sigma_2 / \sqrt{n_2} = 6.1 / \sqrt{12} \approx 6.1 / 3.464 \approx 1.761$

Критерий Стьюдента (t-критерий)

1

Независимые выборки

Шаг 3: Расчет наблюдаемого значения t-критерия
Используем формулу для независимых выборок:

$t = |M_1 - M_2| / \sqrt{(m_1^2 + m_2^2)}$

Находим разность средних:
$|M_1 - M_2| = |82.4 - 87.9| = 5.5$

Возводим стандартные ошибки в квадрат и складываем:
$m_1^2 + m_2^2 = (1.446)^2 + (1.761)^2 \approx 2.091 + 3.101 \approx 5.192$

Извлекаем квадратный корень из суммы:
$\sqrt{(m_1^2 + m_2^2)} = \sqrt{5.192} \approx 2.279$

Рассчитываем t-критерий:
$t = 5.5 / 2.279 \approx 2.413$

Наблюдаемое значение t-критерия: $t \approx 2.413$

Шаг 4: Определение числа степеней свободы (df)
Для критерия Стьюдента с независимыми выборками и неравными дисперсиями (в данном случае мы используем общий подход) число степеней свободы рассчитывается по формуле:

$df = n_1 + n_2 - 2 = 15 + 12 - 2 = 25$

Критерий Стьюдента (t-критерий)

1

Независимые выборки

Шаг 5: Сравнение с критическим значением и интерпретация
Используем таблицу критических значений t-критерия Стьюдента.

Для уровня значимости $\alpha = 0.05$ и числа степеней свободы $df = 25$ двустороннее критическое значение $t_{кр} \approx 2.060$.

Правило принятия решения:

Если $|t_{набл}| \geq t_{кр}$, то различия статистически значимы (отвергаем H_0).

Если $|t_{набл}| < t_{кр}$, то различия статистически незначимы (не отвергаем H_0).

В нашем случае:
$t_{набл} = 2.413 > t_{кр} = 2.060$

Так как наблюдаемое значение t-критерия (2.413) превышает критическое значение (2.060), мы отвергаем нулевую гипотезу (H_0).

Заключение
На основании проведенного анализа с использованием t-критерия Стьюдента для независимых выборок было установлено, что разница в средних значениях диастолического артериального давления между группой, получавшей препарат А (82.4 ± 5.6 мм рт. ст.), и группой, получавшей препарат Б (87.9 ± 6.1 мм рт. ст.), является статистически значимой на уровне значимости $p < 0.05$.

Это дает нам статистические основания утверждать, что препарат А продемонстрировал значимо лучшую эффективность в снижении диастолического давления по сравнению с препаратом Б.

Критерий Стьюдента (t-критерий)

Зависимые выборки

Постановка задачи

Предположим, мы исследуем эффективность новой диеты для снижения веса. Группа из 10 добровольцев следует этой диете в течение 8 недель. Их вес (в кг) измеряется до начала диеты и после ее окончания. Наша цель — определить, привела ли диета к статистически значимому изменению веса.

Предполагается, что разности весов распределены нормально.

Исходные данные:

Испытуемый	Вес до (X)	Вес после (Y)
1	85.2	81.4
2	92.7	89.1
3	78.9	76.3
4	101.5	98.2
5	88.3	85.7
6	94.0	90.5
7	79.6	78.1
8	96.2	92.8
9	83.4	81.0
10	90.1	87.5

Критерий Стьюдента (t-критерий)

Зависимые выборки

2

Пошаговое решение
Шаг 1: Формулировка гипотез
Нулевая гипотеза (H_0): Диета не оказывает эффекта на вес. Средняя разность весов (до - после) в генеральной совокупности равна нулю.

$\mu_d = 0$

Альтернативная гипотеза (H_1): Диета оказывает эффект на вес. Средняя разность весов (до - после) в генеральной совокупности не равна нулю.

$\mu_d \neq 0$ (используется двусторонний критерий).

Уровень значимости: $\alpha = 0.05$.

Шаг 2: Расчет разностей (d) для каждой пары
Рассчитаем разницу "Вес до" - "Вес после" для каждого испытуемого.

	Вес до (X)	Вес после (Y)	Разность (d = X - Y)
1	85.2	81.4	3.8
2	92.7	89.1	3.6
3	78.9	76.3	2.6
4	101.5	98.2	3.3
5	88.3	85.7	2.6
6	94.0	90.5	3.5
7	79.6	78.1	1.5
8	96.2	92.8	3.4
9	83.4	81.0	2.4
10	90.1	87.5	2.6

Критерий Стьюдента (t-критерий)

Зависимые выборки

2

Шаг 3: Расчет средней разности ($\bar{d}$) и стандартного отклонения разностей (σ_d)
Средняя разность ($\bar{d}$):
$\bar{d} = (\Sigma d) / n = (3.8 + 3.6 + 2.6 + 3.3 + 2.6 + 3.5 + 1.5 + 3.4 + 2.4 + 2.6) / 10 = 29.3 / 10 = 2.93$

Стандартное отклонение разностей (σ_d):

Сначала рассчитаем сумму квадратов отклонений от средней разности.
$\Sigma(d - \bar{d})^2 = (3.8-2.93)^2 + (3.6-2.93)^2 + (2.6-2.93)^2 + (3.3-2.93)^2 + (2.6-2.93)^2 + (3.5-2.93)^2 + (1.5-2.93)^2 + (3.4-2.93)^2 + (2.4-2.93)^2 + (2.6-2.93)^2$
$= (0.87)^2 + (0.67)^2 + (-0.33)^2 + (0.37)^2 + (-0.33)^2 + (0.57)^2 + (-1.43)^2 + (0.47)^2 + (-0.53)^2 + (-0.33)^2$
$= 0.7569 + 0.4489 + 0.1089 + 0.1369 + 0.1089 + 0.3249 + 2.0449 + 0.2209 + 0.2809 + 0.1089 = 4.541$

Далее рассчитаем дисперсию:
$s^2 = \Sigma(d - \bar{d})^2 / (n - 1) = 4.541 / 9 \approx 0.5046$

Стандартное отклонение – это квадратный корень из дисперсии:
$\sigma_d = \sqrt{s^2} = \sqrt{0.5046} \approx 0.710$

Шаг 4: Расчет стандартной ошибки средней разности (m_d)
$m_d = \sigma_d / \sqrt{n} = 0.710 / \sqrt{10} \approx 0.710 / 3.162 \approx 0.225$

Критерий Стьюдента (t-критерий)

Зависимые выборки

2

Шаг 5: Расчет наблюдаемого значения t-критерия
t = đ / m_d = 2.93 / 0.225 ≈ 13.02

Наблюдаемое значение t-критерия: t ≈ 13.02

Шаг 6: Определение числа степеней свободы (df)
Для парного критерия число степеней свободы равно:
df = n - 1 = 10 - 1 = 9

Шаг 7: Сравнение с критическим значением и интерпретация
Используем таблицу критических значений t-критерия Стьюдента.

Для уровня значимости α = 0.05 и числа степеней свободы df = 9 двустороннее критическое значение tкр ≈ 2.262.

Правило принятия решения:

Если |tнабл| ≥ tкр, то отвергаем H_0.

Если |tнабл| < tкр, то не отвергаем H_0.

В нашем случае:
tнабл = 13.02 > tкр = 2.262

Так как наблюдаемое значение t-критерия (13.02) значительно превышает критическое значение (2.262), мы отвергаем нулевую гипотезу (H_0).

Заключение
На основании проведенного парного t-теста было установлено, что соблюдение диеты привело к статистически значимому снижению веса (t(9) = 13.02, p < 0.0

Непараметрические критерии для количественных признаков

Применяются при нарушении нормальности, малых выборках или порядковых данных.

1

Критерий Манна-Уитни

Аналог t-критерия для **независимых выборок**. Основан на ранжировании объединенного ряда данных и подсчете сумм рангов (ΣR). Если **Uэмп ≤ Uкр**, различия значимы.

2

Критерий Вилкоксона

Аналог парного t-критерия для **зависимых выборок**. Основан на ранжировании абсолютных разностей между парными измерениями. Если **Тэмп ≤ Ткр**, наблюдается значимый сдвиг.

Критерий x^2 Пирсона

Цель

Оценка статистической значимости различий между частотами исходов в разных группах.

Применимость

Анализ таблиц сопряженности, содержащих данные о частоте исходов (номинальная или порядковая шкала).

Сфера

Эпидемиология, клинические исследования, биостатистика и социология.

Структура четырехпольной таблицы сопряженности

Критерий χ^2 наиболее часто используется для анализа таблиц 2x2, где и фактор риска, и исход являются бинарными переменными (имеют два состояния).

	Исход есть (1)	Исхода нет (0)	Всего
Фактор риска есть (1)	A (наблюдения с фактором и исходом)	B (наблюдения с фактором без исхода)	**A + B**
Фактор риска отсутствует (0)	C (наблюдения без фактора с исходом)	D (наблюдения без фактора без исхода)	**C + D**
Всего	**A + C**	**B + D**	**N = A + B + C + D**

Каждая ячейка (A, B, C, D) содержит фактическое количество наблюдений.

Пример исследования: курение и артериальная гипертония

Рассмотрим конкретный кейс: проверка гипотезы о влиянии курения на частоту развития артериальной гипертонии (АГ).

Группа 1 (Курящие)

- **Объем:** 70 человек (не менее 1 пачки сигарет ежедневно).
- **Исход (АГ):** 40 человек.
- **Без исхода:** 30 человек (70 – 40).

Группа 2 (Некурящие)

- **Объем:** 80 человек (аналогичный возраст).
- **Исход (АГ):** 32 человека.
- **Без исхода:** 48 человек (80 – 32).

Задача: определить, является ли разница в частоте АГ между группами статистически значимой.

Заполнение таблицы фактическими данными

Вносим полученные данные в четырехпольную таблицу сопряженности (фактические частоты *O_{ij}*).

	АГ есть (1)	АГ нет (0)	Всего
Курящие (1)	40	30	**70**
Некурящие (0)	32	48	**80**
Всего	**72**	**78**	**150**

Основной вопрос

Существуют ли статистически значимые различия между долей лиц с артериальной гипертонией среди курящих и некурящих?

Условия и ограничения применения критерия x^2

Шкала Измерения

Сопоставляемые переменные должны быть измерены в **номинальной** (например, пол, наличие/отсутствие) или **порядковой** шкале.

Независимость Выборок

Группы должны быть **независимыми** (не "до/после"). Для связанных выборок применяются тест Мак-Немара или Q-критерий Кохрена.

Ожидаемые Частоты (2x2)

Ожидаемое значение (E_{ij}) в каждой ячейке должно быть **не менее 10**. Иначе — использовать Точный критерий Фишера.

Ожидаемые Частоты (Многопольные)

В многопольных таблицах: E_{ij} **менее 5** не более чем в 20% ячеек.

Расчет статистики χ^2

Значение критерия χ^2 находится путем сравнения фактических ($O\_\{ij\}$) и ожидаемых ($E\_\{ij\}$) частот по всем ячейкам.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Подставляем значения из примера:

(40-33.6)2/33.6 + (30-36.4)2/36.4 + (32-38.4)2/38.4 + (48-41.6)2/41.6

Полученное значение χ^2: 4.396

Определение степеней свободы

Число степеней свободы (**f**) определяет форму распределения χ^2 и необходимо для поиска критического значения.

Формула для f:

$$f = (r - 1) \times (c - 1)$$

Где **r** — число строк, **c** — число столбцов.

Для нашего примера (таблица 2x2):

f = (2 - 1) * (2 - 1) = **1**

Для четырехпольной таблицы (2 ряда, 2 столбца) всегда **одна степень свободы**.

Интерпретация результатов и выводы

Сравниваем рассчитанное значение χ^2 с критическим значением, найденным по таблице для заданного уровня значимости (*p*) и числа степеней свободы (**f**).

4.396

Расчетное χ^2

Получено в нашем исследовании

3.841

Критическое χ^2

Для **p=0.05** и **f=1**

Критерий принятия решения:

Если Расчетное $\boldsymbol{\chi^2}$ > Критического $\boldsymbol{\chi^2}$, то нулевая гипотеза об отсутствии связи отвергается.

Окончательный вывод

Сравнение значений критерия χ^2 в нашем примере:

$$4.396 > 3.841$$

Поскольку расчетное значение χ^2 (4.396) превышает критическое значение (3.841) при уровне значимости **p=0.05** и одной степени свободы, мы делаем следующий вывод:

Связь Статистически Значима

Зависимость частоты случаев артериальной гипертонии от наличия курения является статистически значимой (**$p < 0.05$**).

Результат

Мы можем отвергнуть нулевую гипотезу и утверждать, что курение статистически связано с повышенным риском развития АГ в исследуемой выборке.