## Тема 6.3: «СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ»

Исследователь изучает выборку, но его целью является получение информации о свойствах всей генеральной совокупности, всей популяции. Исследователь должен определить, могут ли результаты выборочного исследования быть распространены на всю генеральную совокупность. Например:

- Эксперимент выявил эффективность лекарства «А» на выборке из 200 пациентов. Можно ли утверждать, что лекарство «А» будет эффективно для любой другой группы аналогичных больных (т.е. для любой другой выборки из той же генеральной совокупности)?
- Эксперимент выявил улучшение показателей образа жизни у испытуемых под влиянием мероприятий по их информированию (в выборке 250 человек). Значит ли это, что результат информирования других аналогичных испытуемых (другой выборки из той же генеральной совокупности) будет таким же?
- Выполнено когортное исследование. В основной группе (лица, подвергающиеся воздействию изучаемого фактора риска) из 300 человек заболели изучаемым заболеванием 100 человек (инцидентность составила 33,3 на 100 испытуемых). В контрольной группе из 250 человек заболели 50 человек (инцидентность равна 20,0 на 100 испытуемых). Можно ли утверждать, что данный фактор действительно является фактором риска развития изучаемого заболевания?

Методологической основой любого научного исследования является формулирование и проверка ***гипотез***. Философия дает следующее толкование этого понятия:

***Гипотеза*** (от греч. hypothesis – основание, предположение) – это положение, выдвигаемое в качестве предварительного, условного объяснения некоторого явления или группы явлений. Это научное предположение, которое требует проверки (в отличие от аксиом и постулатов, не требующих доказательства). Гипотеза – это предположительное, вероятное знание, которое пока не является ни истиной, ни ложью. Гипотезу проверяют в ходе научного исследования. Если будут найдены неопровержимые доказательства гипотезы, то она становится истиной и прекращает свое существование. Если будут обнаружены факты, не согласующиеся с гипотезой, то она станет ложным положением и опять-таки перестанет быть гипотезой.

Таким образом, **алгоритм работы с гипотезой** выглядит так**:**

*1-й этап*. Априорное выдвижение гипотезы. Слово «априорное» (от лат. «a priori» – буквально «от предшествующего») означает «заранее», «первоначально», «бездоказательно», «предварительно».

*2-й этап*. Получение в процессе исследования данных, позволяющих проверить эту гипотезу.

*3-й этап*. Формулирование вывода по итогам проверки гипотезы:

- гипотезу принимают (т.е. признают истинной),
- гипотезу отвергают, отклоняют (т.е. признают ложной).

Существуют понятия «***рабочая гипотеза***» и «***статистическая гипотеза***».

***Рабочая гипотеза*** – это главная, основная гипотеза всего исследования. Она формулируется на подготовительном этапе. Примерами рабочих гипотез могут служить следующие:

- Новое лекарственное средство «….» более эффективно и безопасно, чем лекарственные средства, применяющиеся в настоящее время при данном виде патологии
- Способ реабилитации «….» при …. заболевании позволяет получить лучший результат в более короткие сроки
- Здоровье детей, родившихся в зарегистрированном браке, лучше здоровья детей, родившихся вне брака
- Отсутствие активной интеллектуальной деятельности является фактором риска развития нервно-психических заболеваний в пожилом возрасте
- За период обучения в медицинском вузе образ жизни студентов ухудшается и пр.

***Статистическая гипотеза,*** в отличие от рабочей гипотезы, по своему содержанию является очень узкой. Обычно это предположение относительно вида распределения и его свойств. Например:

- Эмпирическое распределение признака соответствует нормальному распределению
- Средние арифметические двух совокупностей равны между собой
- Дисперсии двух совокупностей различаются

Таким образом, на подготовительном этапе исследователь формулирует *рабочую гипотезу*. На организационном этапе он продумывает построение своего исследования, т.е. технологию проверки своей гипотезы. Потом собирает необходимые данные (методами наблюдения, опроса, либо получая сведения из документов), выполняет их первичную обработку, а затем статистический анализ. В процессе статистического анализа исследователь выдвигает и проверяет десятки, сотни, а иногда и тысячи *статистических гипотез*. Можно сказать, что процесс проверки *рабочей гипотезы* состоит из проверок огромного множества *статистических гипотез* так же, как дом состоит из множества кирпичиков.

Теперь рассмотрим, почему необходимы статистические гипотезы, и что такое стандартные ошибки.

## СТАНДАРТНЫЕ ОШИБКИ

Итак, целью исследования является получение сведений о свойствах генеральной совокупности. Однако изучить генеральную совокупность обычно не представляется возможным, т.к. она слишком велика (например, невозможно изучить всех больных данным заболеванием или всех людей, на кого воздействует данный фактор риска). Поэтому, как правило, изучают не генеральную совокупность, а выборку из нее. Значит, первым же вопросом исследователя оказывается такой: ***«А можно ли утверждать, что полученные выборочным путем статистические характеристики (среднее арифметическое, медиана, дисперсия, относительные показатели и пр.) действительно равны истинным характеристикам генеральной совокупности?»*** Ведь если мы изучим несколько выборок из одной генеральной совокупности, то, скорее всего, они дадут нам несколько ***различающиеся*** сведения. Мы получим разные средние, разные дисперсии, разные доли – просто в силу того, что единицы наблюдения попадают в нашу выборку совершенно случайно.

Для большей наглядности приведем гипотетический пример. Существует генеральная совокупность, состоящая всего из 10 единиц (N=10). Предположим, что изучить ее всю нет возможности. Попытаемся составить представление о свойствах этой совокупности (например, о среднем значении интересующего признака) по результатам нескольких выборочных исследований. Последовательно изучим три случайные выборки разного объема. Все наши выборки являются ***повторными***, т.е. единицы наблюдения после изучения никуда не исчезают, они так и остаются в генеральной совокупности, а следовательно – могут попасть в другую выборку. Полученные результаты представлены в таблице 1.

Таблица 1

**Результаты выборочного изучения количественного признака**

| Изучаемые совокупности | Генеральная совокупность | Выборка 1 | Выборка 2 | Выборка 3 |
|---|---|---|---|---|
| Состав совокупностей (единицы с различными значениями количественного признака) | 16 | 16 | 16 | |
| | 18 | 18 | | 18 |
| | 23 | | 23 | |
| | 24 | | 24 | 24 |
| | 25 | | | 25 |
| | 25 | 25 | | |
| | 26 | | | 26 |
| | 27 | | 27 | |
| | 32 | | | |
| | 34 | | | 34 |
| Число наблюдений | N=10 | $n_1$=3 | $n_2$=4 | $n_3$=5 |
| Среднее арифметическое | $M_{ген}$=25,0 (не известно исследователю) | $M_1$=19,7 | $M_2$=22,5 | $M_3$=25,4 |
| Отклонение выборочных средних ($M_{выб}$) от генерального среднего ($M_{ген}$), т.е. случайная ошибка | – | –5,3 | –2,5 | +0,4 |
| | | (их размер не известен исследователю) | | |

Необходимо усвоить, что статистические оценки никогда не претендуют на идеальную точность. Даже само слово «оценка» в статистике понимается как «указание приближенного значения». Выборка позволяет исследователю получить ***точечную оценку*** истинных свойств генеральной совокупности в форме одного числа. Три выборочные средние (***19,7***, ***22,5*** и ***25,4***) являются точечными оценками генерального среднего, которое исследователю не известно (и никогда не будет известно). Однако логично предположить, что неизвестные истинные свойства генеральной совокупности будут несколько отличаться от полученных по выборке точечных оценок. Причем это отличие может быть как в большую, так и в меньшую сторону. Поэтому точечную оценку принято дополнять ***интервальной оценкой***. Истинные свойства генеральной совокупности представляют не одним числом, а числовым интервалом, в середине которого располагается точечная оценка, полученная по выборке:

$$\mathbf{M_{ген} = M_{выб} \pm \text{предельная ошибка } (\Delta)}$$

Величина предельной ошибки в формулах обычно обозначается прописной греческой буквой «Δ» (читается: «дельта»). Представляя такой интервал, исследователь утверждает (с некоторой уверенностью, которая измеряется в процентах или в долях единицы), что неизвестное истинное значение интересующего его свойства генеральной совокупности находится в указанных границах. Этот интервал называется ***доверительным***. Понятно, что чем меньше предельная ошибка, тем ỳже будет доверительный интервал, а значит – точнее исследование, надежнее выводы.

Для расчета предельной ошибки используется ***стандартная ошибка***. Эта величина в формулах обычно обозначается маленькой латинской буквой ***«m»***. Специалисты в сфере математической статистики могут рассчитывать стандартные ошибки любых выборочных величин. Мы рассмотрим две – стандартную ошибку среднего арифметического и стандартную ошибку экстенсивного показателя.

**А) Стандартная ошибка среднего арифметического** рассчитывается по формуле (формулу необходимо запомнить)***:***

$$\boldsymbol{m = \frac{\sigma}{\sqrt{n}},} \text{ где}$$

σ – среднеквадратическое (стандартное) отклонение,

n – объем выборки.

Стандартная ошибка измеряется в тех же единицах, что и количественный признак.

В зарубежной литературе эту ошибку обозначают **SEM**, т.е. «Standard Error Means», что переводится как «стандартная ошибка среднего». Как видно из формулы, стандартная ошибка

- имеет прямую связь с величиной «σ», т.е. с разнообразием количественного признака.
  - Чем больше разнообразие (σ), тем больше стандартная ошибка (m). Следовательно, больше будет и предельная ошибка (Δ), доверительный интервал будет шире.
  - Чем меньше разнообразие (σ), тем меньше стандартная ошибка (m). Следовательно, предельная ошибка (Δ) тоже будет меньше, доверительный интервал – более узкий.
- имеет обратную связь с величиной «n», т.е. с объемом выборки.
  - Чем больше выборка (n), тем меньше стандартная ошибка (m). Меньше будет и предельная ошибка, а значит будет более узким доверительный интервал.
  - Чем меньше выборка (n), тем больше стандартная ошибка (m). Величина предельной ошибки (Δ) будет больше, доверительный интервал – шире.

Таким образом, широкий доверительный интервал получится при изучении очень вариабельного признака на малой выборке (исследование будет неточным). И наоборот, исследование будет точным (т.е. результат будет близок к точке), доверительный интервал будет узким, если изучается мало разнообразный признак на большой выборке.

Разнообразие признака от исследователя никак не зависит. А вот объем выборки – полностью «в его руках». Поэтому до основного исследования иногда проводят пилотажное, на малом числе единиц наблюдения, чтобы сориентироваться в отношении разнообразия. Если подобные исследования уже выполнялись раньше, и результаты их были опубликованы, то значение среднеквадратического (стандартного) отклонения берут из литературных данных. И в том, и в другом случае, если выяснится, что вариабельность признака высокая, то потребуется большее количество единиц наблюдения. Если же разнообразие окажется небольшим, то и выборку можно сформировать небольшую.

**Б) Стандартная ошибка относительного экстенсивного показателя** рассчитывается по формуле (формулу необходимо запомнить)***:***

$$\boldsymbol{m} = \sqrt{\frac{P_{выб} * \boldsymbol{q}}{\boldsymbol{n}}}\text{, где}$$

$P_{выб}$ - относительный экстенсивный показатель, доля,

q – альтернатива,

n – объем выборки.

Стандартная ошибка относительного показателя измеряется в тех же величинах, что и сам относительный показатель, т.е. в долях единицы или в процентах.

В зарубежной литературе эту ошибку обозначают **SE(p)**, т.е. «Standard Error (proportion)», что переводится как «стандартная ошибка пропорции». Как видим из формулы, эта ошибка также имеет обратную связь с объемом выборки (n): чем больше выборка, тем меньше стандартная ошибка (m), меньше предельная ошибка (Δ), уже доверительный интервал, т.е. лучше, точнее исследование.

В числителе находится произведение относительного показателя ($P_{выб}$) и его альтернативы (q). Величина «q» дополняет экстенсивный показатель до целого размера (до 100% или 1,0) и рассчитывается по формуле:

**$q=1–P_{выб}$** (если $P_{выб}$ представлено в долях единицы)

**$q=100\%–P_{выб}$** (если $P_{выб}$ представлено в процентах)

Другими словами в числителе долю наблюдений, у которых встречается изучаемый качественный признак, надо умножить на долю наблюдений, у которых изучаемый признак не обнаружен. Поэтому в литературе вы можете увидеть следующий вариант этой формулы:

$$\boldsymbol{m} = \sqrt{\frac{\boldsymbol{P}(\boldsymbol{1}-\boldsymbol{P})}{\boldsymbol{n}}}$$

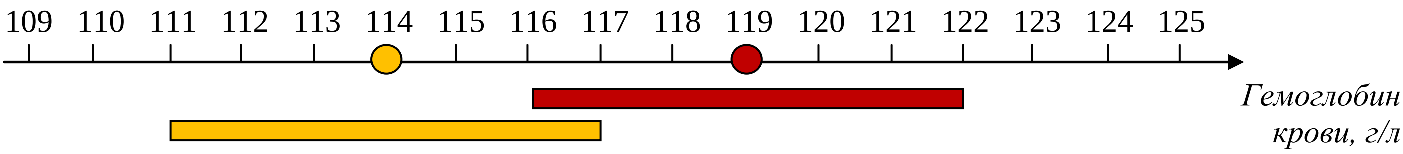
После рассмотрения стандартных ошибок разберем следующий пример. Выполнено экспериментальное исследование с положительным параллельным контролем с целью изучить эффективность и безопасность нового лекарственного средства для лечения анемии. После выполнения вмешательства средний уровень гемоглобина крови у пациентов составил (результаты представлены по схеме M ± Δ):

- в основной группе (получавших новый препарат) **119,0 ± 3,0 г/л**
- в контрольной группе (получавших стандартную терапию) **114,0 ± 3,0 г/л**

Можно ли на основании этой информации утверждать, что выдвинутая рабочая гипотеза истинна – что новое лекарственное средство более эффективно, чем стандартная терапия? Вроде кажется, что можно – ведь 119,0 г/л больше, чем 114,0 г/л. На самом деле такого вывода пока еще сделать нельзя (рисунок 1).

Рисунок 1.

**Точечные и интервальные оценки уровня гемоглобина крови в двух группах пациентов**




Как видим из рисунка 1, можно утверждать, что при лечении анемии **у любой другой группы больных** средний уровень гемоглобина крови

- после применения нового лекарственного средства будет от 116,0 г/л до 122,0 г/л,
- после стандартной терапии будет от 111,0 г/л до 117,0 г/л

То есть вполне возможно, что в других группах пациентов с анемией после применения нового лекарственного средства средний уровень гемоглобина увеличится до 116,1 г/л, а после применения стандартной терапии – до 116,9 г/л. Можно ли делать вывод о том, что новое лекарство более эффективно? Нет, нельзя; такой вывод будет некорректен.

Другими словами, выявленные в исследовании различия (например, различие среднего уровня гемоглобина крови) могут на самом деле оказаться **случайными**, ничего не значащими, ни о чем не говорящими. Они обусловлены всего лишь особенностями единиц наблюдения, тех испытуемых, что **случайно** оказались в нашей выборке. Если бы мы набрали других испытуемых, то могли бы получить совсем другие результаты и сделать другие выводы. …Или же все-таки нет? Каковы бы ни были выборки, из каких бы испытуемых они ни состояли, исследователь снова, и снова, и снова будет обнаруживать, что новое лекарственное средство более эффективно для лечения анемии, чем стандартное? Так, да или нет? Вот так и появляется статистическая гипотеза, т.е. статистическое предположение, нуждающееся в статистической проверке.

## АЛГОРИТМ РАБОТЫ СО СТАТИСТИЧЕСКОЙ ГИПОТЕЗОЙ

Алгоритм работы со статистической гипотезой в целом такой же, как с рабочей гипотезой (выдвижение гипотезы → проверка гипотезы → выводы). Однако, чтобы пояснить ход рассуждений исследователя, нам придется разделить этот процесс на более мелкие составляющие:

*1-й шаг*. Выдвижение нулевой ($H_0$) и альтернативной ($H_1$) гипотез
*2-й шаг*. Выбор статистического критерия
*3-й шаг*. Расчет значения статистического критерия по эмпирическим данным
*4-й шаг*. Определение критического значения статистического критерия
*5-й шаг*. Сравнение рассчитанного значения статистического критерия с его критическим значением. Интерпретация результата и выводы

Рассмотрим каждый из этих шагов более внимательно.

### *1-й шаг*. ВЫДВИЖЕНИЕ НУЛЕВОЙ ($H_0$) И АЛЬТЕРНАТИВНОЙ ($H_1$) ГИПОТЕЗ

Выдвинутую гипотезу называют ***нулевой*** или ***основной*** **(**обозначают $\mathbf{H_0}$**).** Гипотезу, которая противоречит нулевой и является ее логическим отрицанием, называют ***альтернативной*** или ***конкурирующей*** **(**обозначают $\mathbf{H_1}$**)**. Например:

- *Нулевая гипотеза ($H_0$)*: средние арифметические двух совокупностей равны между собой. Коротко это записывают так: $H_0$: $M_1=M_2$
- *Альтернативная гипотеза ($H_1$)*: средние арифметические двух совокупностей не равны между собой ($H_1$: $M_1 \neq M_2$)

Как видите, у исследователя существует выбор только одного из двух вариантов – истинной будет либо нулевая гипотеза ($H_0$), либо альтернативная ($H_1$). Третьего не дано. Перед исследователем два алгоритма рассуждений (рисунок 2).

Рисунок 2.

**Схематичное представление работы с гипотезой**

**1-й вариант работы с гипотезой** | **2-й вариант работы с гипотезой (ПРАВИЛЬНЫЙ)**

Почему же именно второй подход к работе будет правильным? Почему следует не подтверждать гипотезу, а ее опровергать, отклонять? Дело в том, что **подтвердить гипотезу невозможно**: найденные доказательства могут свидетельствовать не только в пользу данной гипотезы, но и в пользу других гипотез. Причем, возможно, некоторые из них в настоящее время еще не высказаны, еще никому «не пришли в голову». Положительный результат исследования не подтверждает именно эту гипотезу. Он ее просто не опровергает. Классическим примером является следующий. Изменения положения Солнца по отношению к Земле были известны человечеству с древнейших времен. Многие тысячелетия эти доказательства рассматривались как надежное подтверждение гипотезы о вращении Солнца вокруг Земли (впоследствии эта гипотеза оказалась ложной). Что же касается опровержения, отклонения выдвинутой гипотезы, то здесь все обстоит гораздо проще. Если будет зафиксирован, хотя бы один факт, не согласующийся с выдвинутой гипотезой, то, следовательно, эта гипотеза не верна. Замечательной иллюстрацией здесь могут служить слова гениального физика-теоретика Альберта Эйнштейна (1879-1955): *«Ни один эксперимент не говорит теории "Да". Он может сказать только "Нет" или "Может быть".»*

Таким образом, в ходе статистического анализа ищут аргументы не в пользу выдвинутой гипотезы, а против нее, т.е. пытаются гипотезу не подтвердить, а отклонить. ***В качестве нулевой (основной) гипотезы ($H_0$) всегда выдвигают гипотезу об отсутствии статистически значимых различий. Альтернативной (конкурирующей) гипотезой ($H_1$) будет являться гипотеза о существовании различий.*** Если удастся найти сильные аргументы против нулевой гипотезы ($H_0$), то она будет отклонена, т.е. признана ложной. Следовательно, истинной будет считаться противоречащая ей альтернативная гипотеза ($H_1$). Другими словами, ***проверка статистической гипотезы – это оценка силы аргументов против нулевой гипотезы.***

Рассмотрим несколько примеров статистических гипотез.

**1-й пример**.

Выполнено экспериментальное исследование с положительным параллельным контролем с целью изучить эффективность и безопасность нового лекарственного средства, применяемого для лечения хронического заболевания. Доля пациентов, у кого наступила ремиссия, в основной группе составила 78,3% ($P_1$), в контрольной группе – 65,4% ($P_2$).

- *Нулевая гипотеза*: различие долей пациентов с ремиссией в группах сравнения случайно, статистически незначимо, т.е. новое лекарственное средство не лучше стандартной терапии ($H_0$: $P_1$=$P_2$).
- *Альтернативная гипотеза*: различие долей пациентов с ремиссией в группах сравнения неслучайно, статистически значимо, т.е. новое лекарственное средство действительно лучше, чем стандартная терапия ($H_1$: $P_1 \neq P_2$).

**2-й пример**.

Выполнено повторное поперечное исследование по изучению физического развития детей в регионе «К». Средний рост 10-летних мальчиков составил 146,3 см ($M_2$). Средний рост 10-летних мальчиков по данным аналогичного исследования, выполненного 15 лет назад, составлял 142,1 см ($M_1$).

- *Нулевая гипотеза*: выявленное различие среднего роста 10-летних мальчиков региона «К» является случайным, статистически незначимым, т.е. в данном году в выборку случайно попали высокие мальчики, а 15 лет назад – мальчики невысокого роста ($H_0$: $M_1$=$M_2$).
- *Альтернативная гипотеза*: выявленное различие среднего роста 10-летних мальчиков региона «К» не является случайным, оно статистически значимо, т.е. 10-летние мальчики региона «К» за 15 прошедших лет действительно стали выше ($H_0$: $M_1 \neq M_2$)

### *2-й шаг.* ВЫБОР СТАТИСТИЧЕСКОГО КРИТЕРИЯ

Проверка выдвинутой нулевой гипотезы осуществляется путем расчета и оценки величины определенного статистического критерия. ***Статистический критерий*** – это строгое математическое правило, формула, куда следует подставить строго определенные эмпирические значения. Например:

- **t-критерий Стьюдента** позволяет проверить нулевую гипотезу о равенстве средних арифметических двух несвязанных групп наблюдений ($H_0$: $M_1=M_2$) и рассчитывается по формуле:

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

где

$M_1$ и $M_2$ – средние арифметические двух групп наблюдений,

$m_1$ и $m_2$ – стандартные ошибки этих средних арифметических

- **критерий Пирсона $\chi^2$** (читается «хи-квадрат») является непараметрическим критерием и позволяет сравнить несколько групп по описательному признаку; рассчитывается по формуле:

$$\chi^2 = \sum \frac{(p - p')^2}{p'}$$

где

p – фактическая абсолютная частота,

p′ – ожидаемая частота.

Различных статистических критериев много, и у каждого из них есть свои условия применения. Очень важно выбрать правильный статистический критерий, тот, который подходит именно в данном случае, для проверки именно этой нулевой гипотезы. В противном случае сделанные исследователем выводы могут оказаться неверными, некорректными. Исследователь допустит систематическую ошибку (смещение). Наиболее часто применяемые статистические критерии и условия их использования представлены в таблице 2.

Таблица 2.

**Статистические критерии и условия их использования**

<table>
<tr><th colspan="2" rowspan="2">Признак</th><th colspan="4">Дизайн исследования</th></tr>
<tr><th>две несвязанные группы</th><th>две связанные группы</th><th>три и более несвязанных групп</th><th>изучение связи признаков</th></tr>
<tr><td colspan="2">1. Количественный, распределение нормальное (параметрические критерии)</td><td>t-критерий Стьюдента</td><td>парный t-критерий Стьюдента</td><td>критерий Фишера (F) (дисперсионный анализ)</td><td>• регрессионный анализ<br>• коэффициент корреляции Пирсона (r)</td></tr>
<tr><td rowspan="2">непараметрические критерии</td><td>2. Количественный, распределение не соответствует нормальному, или описательный порядковый</td><td>критерий Манна-Уитни (U)</td><td>критерий Вилкоксона (W)</td><td>критерий Краскела-Уоллиса (H)</td><td>коэффициент ранговой корреляции Спирмена (ρ)</td></tr>
<tr><td>3. Описательный номинативный</td><td>критерий Пирсона ($\chi^2$)</td><td>критерий Мак-Немара</td><td>критерий Пирсона ($\chi^2$)</td><td>коэффициент сопряженности</td></tr>
</table>

Статистические критерии объединяются в две группы – параметрические и непараметрические критерии.

1) **Параметрические критерии** (первая строка в таблице 1) используют для своего расчета параметры нормального распределения, т.е. среднее арифметическое (M) и дисперсию (D, $\sigma^2$). Параметрические критерии можно использовать только в случаях доказанной нормальности распределения количественного признака.

2) **Непараметрические критерии** (вторая и третья строки в таблице 1) оперируют рангами и частотами. Для их расчета параметры нормального распределения (среднее арифметическое и дисперсия) не нужны. Поэтому такие критерии используют для сравнения групп наблюдений по описательному признаку, а также по количественному признаку, распределение которого не соответствует нормальному. Кроме того непараметрические критерии рекомендуется применять при сравнении групп небольшой численности, т.к. на малых группах сложно доказать нормальность распределения признака.

Покажем на нескольких примерах, как осуществляется выбор статистического критерия для проверки нулевой гипотезы.

1) Выполнено экспериментальное исследование с параллельным положительным контролем. Средний уровень гемоглобина крови после медицинского вмешательства у пациентов основной группы составил 119 г/л ($M_1$), у пациентов контрольной группы 114 г/л ($M_2$). Объем основной группы ($n_1$) 215 человек, объем контрольной группы ($n_2$) 220 человек. Распределение пациентов по уровню гемоглобина крови после вмешательства в обеих группах подчиняется законам нормального распределения. Группы формировались независимо друг от друга с применением рандомизации. Для проверки нулевой гипотезы о равенстве средних ($H_0$: $M_1$=$M_2$) необходимо выбрать **t-критерий Стьюдента для несвязанных групп**.
2) Выполнено экспериментальное исследование с параллельным положительным контролем для сравнительной оценки гиполипидемического действия четырех препаратов. Средний уровень холестерина крови после медицинского вмешательства в четырех группах пациентов составил 6,8 ммоль/л ($M_1$), 4,7 ммоль/л ($M_2$), 5,9 ммоль/л ($M_3$) и 5,1 ммоль/л ($M_4$). Распределение пациентов по уровню холестерина крови после вмешательства во всех четырех группах подчиняется законам нормального распределения. Группы формировались независимо друг от друга с применением рандомизации. Для проверки нулевой гипотезы о равенстве средних ($H_0$: $M_1$=$M_2$=$M_3$=$M_4$) необходимо выполнить **дисперсионный анализ с расчетом F-критерия Фишера**.
3) Выполнено экспериментальное исследование с параллельным контролем на двух малых группах испытуемых ($n_1$=8, $n_2$=7). Одним из показателей эффекта медицинского вмешательства являлось снижение систолического артериального давления. Малая численность групп не позволила убедительно продемонстрировать нормальность распределения этого признака. Группы формировались независимо друг от друга. Для проверки нулевой гипотезы об отсутствии влияния изучаемого медицинского вмешательства на уровень систолического давления следует выбрать непараметрический **критерий Манна-Уитни (U)**.
4) Выполнено неконтролируемое экспериментальное исследование с целью изучения эффективности и безопасности нового гипогликемического лекарственного препарата. Численность группы составила 185 человек. У каждого испытуемого был дважды замерен уровень сахара крови (до лечения и после лечения), а затем по каждому испытуемому было определено изменение этого показателя (путем расчета разности). Группа разностей подчинялась законам нормального распределения. Для проверки нулевой гипотезы об отсутствии эффекта у нового лекарственного препарата следует выбрать **парный t-критерий Стьюдента**.
5) Выполнено неконтролируемое экспериментальное исследование нового психотерапевтического вмешательства. У малой группы пациентов с невротическими расстройствами (n=23) методом психологической диагностики был оценен в баллах уровень тревожности до и после вмешательства. Для проверки нулевой гипотезы об отсутствии психотерапевтического эффекта изучаемого вмешательства следует использовать непараметрический **критерий Вилкоксона (W)**.
6) Выполнено наблюдательное поперечное исследование с целью изучить здоровье работников крупного текстильного предприятия. Проанализированы четыре профессиональные группы работниц – ленточницы, мотальщицы, прядильщицы и ткачихи. Для каждой работницы определена группа здоровья (первая, вторая или третья). Для проверки нулевой гипотезы об отсутствии различий в распределении работниц этих четырех профессий по группам здоровья следует выбрать непараметрический **критерий Пирсона ($\chi^2$)**.

Выбор статистического критерия (параметрический или непараметрический) во многом зависит от «пристрастий» исследователя. Но следует помнить, что непараметрические критерии нерационально используют имеющуюся информацию (например, заменяют фактические числа их ранговыми местами). ***Параметрический критерий обладает большей мощностью в обнаружении реально существующего эффекта, чем эквивалентный ему непараметрический.*** Поэтому по возможности следует использовать параметрический критерий, и только в случае отсутствия обязательных условий для его применения – непараметрический аналог.

Любой статистический критерий может быть двусторонним или односторонним.

1) **Двусторонний вариант критерия** применяется в случаях, когда нельзя заранее, априори, до начала исследования предсказать направление различия сравниваемых групп (т.е. в какой группе значение показателя будет больше, а в какой – меньше). Например, сравнивается образ жизни мужчин и женщин. По результатам исследования доля лиц с гиподинамией среди мужчин составила 28,5%, а среди женщин 21,3%. Разумеется, что не представлялось возможным заранее предвидеть, среди лиц какого пола будет больше доля людей с гиподинамией. В данном случае для проверки нулевой гипотезы об отсутствии различий следует использовать двусторонний вариант критерия. ***В большинстве случаев исследователи встречаются именно с такой ситуацией, поэтому двусторонние критерии применяются гораздо чаще, чем односторонние.***
2) **Односторонний вариант критерия** используется в том случае, когда можно заранее предсказать направление различий. Например, изучается влияние запыленности воздуха в цехах на величину жизненной емкости легких (ЖЕЛ). Сравниваются по этому показателю две группы рабочих – со стажем работы на предприятии менее 3 лет и со стажем более 10 лет. Вряд ли труд в условиях запыленности будет способствовать увеличению ЖЕЛ. Следует ожидать, что в группе рабочих с большим стажем работы ЖЕЛ будет меньше. В этом случае исследователь может применить односторонний вариант статистического критерия.

### *3-й шаг.* Расчет значения статистического критерия по эмпирическим данным

В формулу статистического критерия следует подставить эмпирические данные и выполнить расчеты. Существуют следующие способы выполнения этой работы:

1) Выполнить расчеты вручную, используя калькулятор. Если сведений немного, именно такой путь может оказаться наиболее рациональным.
2) Выполнить расчеты, используя возможности электронной таблицы, например, Microsoft Excel. На первом листе книги Microsoft Excel можно создать базу данных, второй лист использовать для статистического анализа этих данных. В ячейки таблицы вводят формулы. Эти формулы надо аккуратно «собирать» из чисел, арифметических действий, статистических функций (функции в свою очередь требуют указания аргументов – ссылок на массивы и ячейки, указания числа степеней свободы, указания числа «хвостов распределения» и пр.). После ввода формулы (нажатия клавиши Enter) в ячейке появится результат вычислений.
3) Выполнить расчеты, используя возможности специальных программных продуктов, например, пакета Statistica. В эту программу загружается база данных, созданная в электронной таблице, и щелчками мыши в открывающихся диалоговых окнах указываются статистические критерии, которые нужно рассчитать, а также названия полей базы данных, по которым требуется провести эти расчеты. После нажатия ОК на экране появится отчет системы с результатами расчетов.

### *4-й шаг.* Определение критического значения статистического критерия

Статистический критерий позволяет выявить меру расхождения между эмпирическими данными и гипотетическими. Каким образом? Не вдаваясь в математические нюансы, объясним просто. Статистический критерий может принимать бесконечное множество разных значений: 0,86 … 1,25 … 2,17 … 3,31 … 8,59 … 34,42 и т.д. Каждому значению соответствуют две дополняющие друг друга вероятности:

- вероятность того, что истинна нулевая гипотеза ($H_0$),
- вероятность того, что истинна альтернативная гипотеза ($H_1$).

Величина этих вероятностей будет связана с числом единиц наблюдения. Традиционно эти вероятности представляют не в процентах, а в долях единицы. Поскольку гипотез всего две, то сумма этих взаимосвязанных вероятностей равна 1,0. Чем больше одна вероятность, тем меньше другая. Большинство статистических критериев таковы, что по мере роста значения критерия снижается вероятность истинности нулевой гипотезы ($H_0$) и соответственно повышается вероятность истинности альтернативной гипотезы ($H_1$). В качестве примера приведем t-критерий Стьюдента (таблица 3).

Таблица 3.

**Некоторые значения t-критерия Стьюдента и соответствующие им вероятности (для группы из 30 единиц)**

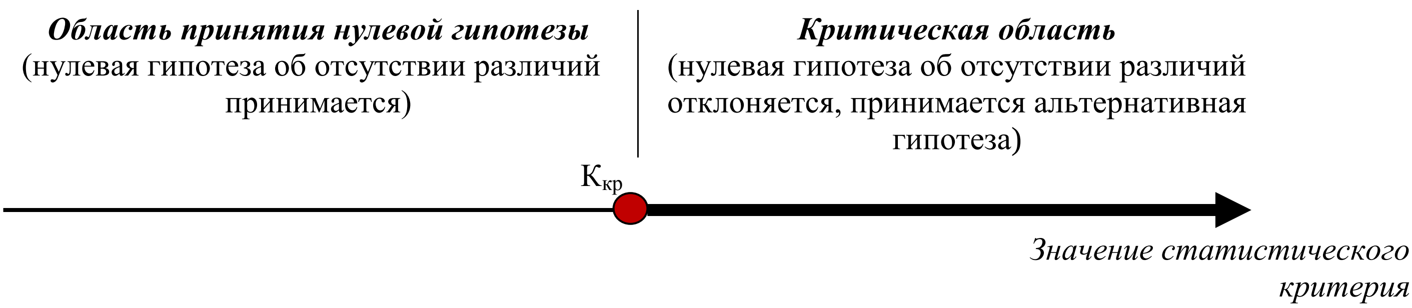
| **Значения t-критерия Стьюдента** | **t=0,01** | **t=0,1** | **t=0,3** | **t=0,5** | **t=0,8** | **t=1,0** | **t=1,5** | **t=2,0** | **t=3,0** |
|---|---|---|---|---|---|---|---|---|---|
| Вероятность истинности нулевой гипотезы ($H_0$) | 0,992 | 0,921 | 0,766 | 0,621 | 0,430 | 0,326 | 0,144 | 0,055 | 0,006 |
| Вероятность истинности альтернативной гипотезы ($H_1$) | 0,008 | 0,079 | 0,234 | 0,379 | 0,570 | 0,674 | 0,856 | 0,945 | 0,994 |

Вероятность истинности нулевой гипотезы называется ***уровень значимости*** и обозначается маленькой латинской буквой ***«р»***. Уровень значимости это одна из важнейших величин в математической статистике.

Как же значение статистического критерия поможет исследователю сделать выбор в пользу одной из гипотез – нулевой или альтернативной? Одно из возможных значений статистического критерия принимается в качестве ***критического значения*** ($K_{кр}$). Оно разделит все возможные значения данного статистического критерия на два подмножества. Те значения, которые меньше критического ($K < K_{кр}$), образуют так называемую «область принятия нулевой гипотезы». Те значения критерия, которые больше критического ($K > K_{кр}$) образуют так называемую «критическую область» (рисунок 3).

Рисунок 3

**Критическое значение статистического критерия и формируемые им области**




***В качестве критического в медицинской науке принимают то значение статистического критерия, которому соответствует уровень значимости (р) равный 0,05. Это значит, что исследователь на 95% уверен в истинности альтернативной гипотезы, в том, что сравниваемые группы действительно различаются. Вероятность того, что эти его выводы ошибочны, и истинной является всё же нулевая гипотеза, не превышает 5% или 0,05.*** Для научных исследований в фармакологии или токсикологии рекомендуют использовать даже меньший уровень значимости (р) равный 0,01, т.е. вероятность ошибочно увидеть различия групп не должна превышать 1%.

Может показаться, что вероятность ошибки величиной 0,05 (или 1/20) – это смехотворно малая величина. На самом же деле – это величина просто гигантская. В технических науках для статистического анализа применяют много меньшие значения «р». Как гласит старая шутка, если бы инженеры и конструкторы выполняли свои расчеты с уровнем значимости р=0,05, то разрушался бы каждый 20-й кирпич в стене, падал бы каждый 20-й самолет, и железнодорожные мосты разваливались бы под каждым 20-м поездом. Такой высокий уровень значимости р=0,05 применяется в медицинской науке только по причине высокой вариабельности изучаемых признаков. Если бы требовалось выполнять статистический анализ с меньшим уровнем значимости «р», то объем выборок должен был бы составлять десятки и сотни тысяч единиц наблюдения.

Определить критическое значение статистического критерия можно следующими способами:

1) Использовать специальные компьютерные программы (например, Microsoft Excel или Statistica). В современных версиях есть статистические функции, которые предложат пользователю в качестве результата расчетов даже не само критическое значение статистического критерия (с которым потребуется сравнить эмпирическое значение критерия), а уже готовый уровень значимости (р) для эмпирического значения. Пользователю остается только сопоставить его с 0,05 и сделать соответствующие выводы. Такими функциями являются, например, СТЬЮДЕНТ.ТЕСТ, F.ТЕСТ, ХИ2.ТЕСТ.
2) Использовать приближенные значения критических точек. Например, для больших выборок в качестве критического значения t-критерия Стьюдента применяют t=2. Именно этой величине соответствует уровень значимости р=0,05, что обеспечивают 95%-ную надежность выводов о существовании различий между группами.
3) Использовать таблицы критических значений различных теоретических распределений.

Именно этот последний способ мы и будем применять на практических занятиях, поэтому познакомимся с алгоритмом его выполнения.

В таблице 4 представлены критические значения t-критерия Стьюдента – они занимают основную массу ячеек. Прежде всего, необходимо определиться с нужной строкой таблицы. В левом столбце указано ***число степеней свободы***. В литературе эта величина может быть обозначена буквой k, либо буквой υ, либо буквами df (от английского «degree of freedom»). Чтобы определить число степеней свободы следует из объема наблюдения вычесть единицу. Например, объем выборки составляет 26 единиц, тогда число степеней свободы будет равно 25.

Далее следует выбрать вариант статистического критерия – двусторонний или односторонний. В верхней части таблицы отмечены уровни значимости для двусторонней критической области, в нижней части – для односторонней области. Как уже было упомянуто выше, в подавляющем большинстве случаев следует выбирать двусторонний вариант статистического критерия.

Затем следует выбрать интересующий исследователя уровень значимости «р». Как уже было сказано, в медицинских научных исследованиях уровень значимости не может быть больше 0,05 (т.е. уверенность в истинности альтернативной гипотезы должна быть не менее 95%). Уровень значимости отражает так называемую α-ошибку (ошибку первого рода), поэтому часто обозначается не буквой «р», а греческой буквой «α» (альфа).

Искомое критическое значение t-критерия Стьюдента будет располагаться на пересечении строки с 25 степенями свободы и столбца с уровнем значимости 0,05 для двустороннего варианта критерия. Это будет число **2,060**.

Таблица 4

## Критические значения t-критерия Стьюдента

| Число степеней свободы | Уровень значимости α (двусторонняя критическая область) | | | | | |
|---|---|---|---|---|---|---|
| | 0,10 | 0,05 | 0,02 | 0,01 | 0,002 | 0,001 |
| 1 | 6,314 | 12,706 | 31,821 | 63,657 | 318,309 | 636,619 |
| 2 | 2,920 | 4,303 | 6,965 | 9,925 | 22,327 | 31,599 |
| 3 | 2,353 | 3,182 | 4,541 | 5,841 | 10,215 | 12,924 |
| 4 | 2,132 | 2,776 | 3,747 | 4,604 | 7,173 | 8,610 |
| 5 | 2,015 | 2,571 | 3,365 | 4,032 | 5,893 | 6,869 |
| 6 | 1,943 | 2,447 | 3,143 | 3,707 | 5,208 | 5,959 |
| 7 | 1,895 | 2,365 | 2,998 | 3,499 | 4,785 | 5,408 |
| 8 | 1,860 | 2,306 | 2,896 | 3,355 | 4,501 | 5,041 |
| 9 | 1,833 | 2,262 | 2,821 | 3,250 | 4,297 | 4,781 |
| 10 | 1,812 | 2,228 | 2,764 | 3,169 | 4,144 | 4,587 |
| 11 | 1,796 | 2,201 | 2,718 | 3,106 | 4,025 | 4,437 |
| 12 | 1,782 | 2,179 | 2,681 | 3,055 | 3,930 | 4,318 |
| 13 | 1,771 | 2,160 | 2,650 | 3,012 | 3,852 | 4,221 |
| 14 | 1,761 | 2,145 | 2,624 | 2,977 | 3,787 | 4,140 |
| 15 | 1,753 | 2,131 | 2,602 | 2,947 | 3,733 | 4,073 |
| 16 | 1,746 | 2,120 | 2,583 | 2,921 | 3,686 | 4,015 |
| 17 | 1,740 | 2,110 | 2,567 | 2,898 | 3,646 | 3,965 |
| 18 | 1,734 | 2,101 | 2,552 | 2,878 | 3,610 | 3,922 |
| 19 | 1,729 | 2,093 | 2,539 | 2,861 | 3,579 | 3,883 |
| 20 | 1,725 | 2,086 | 2,528 | 2,845 | 3,552 | 3,850 |
| 21 | 1,721 | 2,080 | 2,518 | 2,831 | 3,527 | 3,819 |
| 22 | 1,717 | 2,074 | 2,508 | 2,819 | 3,505 | 3,792 |
| 23 | 1,714 | 2,069 | 2,500 | 2,807 | 3,485 | 3,768 |
| 24 | 1,711 | 2,064 | 2,492 | 2,797 | 3,467 | 3,745 |
| 25 | 1,708 | 2,060 | 2,485 | 2,787 | 3,450 | 3,725 |
| 26 | 1,706 | 2,056 | 2,479 | 2,779 | 3,435 | 3,707 |
| 27 | 1,703 | 2,052 | 2,473 | 2,771 | 3,421 | 3,690 |
| 28 | 1,701 | 2,048 | 2,467 | 2,763 | 3,408 | 3,674 |
| 29 | 1,699 | 2,045 | 2,462 | 2,756 | 3,396 | 3,659 |
| 30 | 1,697 | 2,042 | 2,457 | 2,750 | 3,385 | 3,646 |
| 40 | 1,684 | 2,021 | 2,423 | 2,704 | 3,307 | 3,551 |
| 50 | 1,676 | 2,009 | 2,403 | 2,678 | 3,261 | 3,496 |
| 60 | 1,671 | 2,000 | 2,390 | 2,660 | 3,232 | 3,460 |
| 70 | 1,667 | 1,994 | 2,381 | 2,648 | 3,211 | 3,435 |
| 80 | 1,664 | 1,990 | 2,374 | 2,639 | 3,195 | 3,416 |
| 90 | 1,662 | 1,987 | 2,368 | 2,632 | 3,183 | 3,402 |
| 100 | 1,660 | 1,984 | 2,364 | 2,626 | 3,174 | 3,390 |
| 120 | 1,658 | 1,980 | 2,358 | 2,617 | 3,160 | 3,373 |
| 140 | 1,656 | 1,977 | 2,353 | 2,611 | 3,149 | 3,361 |
| 160 | 1,654 | 1,975 | 2,350 | 2,607 | 3,142 | 3,352 |
| 180 | 1,653 | 1,973 | 2,347 | 2,603 | 3,136 | 3,345 |
| 200 | 1,653 | 1,972 | 2,345 | 2,601 | 3,131 | 3,340 |
| 300 | 1,650 | 1,968 | 2,339 | 2,592 | 3,118 | 3,323 |
| 400 | 1,649 | 1,966 | 2,336 | 2,588 | 3,111 | 3,315 |
| 500 | 1,648 | 1,965 | 2,334 | 2,586 | 3,107 | 3,310 |
| 1000 | 1,646 | 1,962 | 2,330 | 2,581 | 3,098 | 3,300 |
| Число степеней свободы | 0,05 | 0,025 | 0,01 | 0,005 | 0,001 | 0,0005 |
| | Уровень значимости α (односторонняя критическая область) | | | | | |

*5-й шаг*. **Сравнение рассчитанного значения статистического критерия с его критическим значением. Интерпретация результата и выводы.**

Значение статистического критерия, рассчитанного по эмпирическим данным, количественно отражает аргументы против нулевой гипотезы ($H_0$). Это рассчитанное эмпирическое значение ($К_{факт}$) нужно сравнить с критической точкой ($К_{кр}$).

- Если рассчитанное значение статистического критерия **меньше критического** ($К_{факт}<К_{кр}$), то соответствующий ему уровень значимости будет больше 0,05 **(p>0,05)**. **Аргументов для отклонения нулевой гипотезы у исследователя недостаточно**. Нулевую гипотезу о том, что группы наблюдений не имеют статистически значимых различий ($H_0$), отклонить нельзя. Исследователь вынужден с ней согласиться. Что это значит в контексте конкретного исследования? То, что новый лекарственный препарат не имеет преимуществ перед стандартной терапией (для экспериментального исследования). То, что данный фактор, не является фактором риска развития изучаемой патологии (для продольного эпидемиологического исследования). То, что за время, прошедшее с первого исследования, изучаемый объект никак не изменился (для повторного поперечного исследования).
- Если рассчитанное значение статистического критерия больше критического ($К_{факт}>К_{кр}$), то соответствующий ему уровень значимости будет меньше 0,05 **(p<0,05)**. **Аргументов для отклонения нулевой гипотезы достаточно**. Причем, чем больше эмпирическое значение статистического критерия, тем меньше его уровень значимости, а, следовательно, больше аргументов против нулевой гипотезы. А поскольку нулевая гипотеза признана ложной, то **принимается, т.е. признается истинной, альтернативная гипотеза ($H_1$)**, являющаяся ее логическим отрицанием. Что это значит в контексте конкретного исследования? То, что новый лекарственный препарат действительно более эффективен, чем стандартная терапия (для экспериментального исследования). То, что данный фактор, действительно является фактором риска развития изучаемой патологии (для продольного эпидемиологического исследования). То, что за время, прошедшее с первого исследования, изучаемый объект действительно изменился (для повторного поперечного исследования).

Теперь перейдем от слов к делу. Научимся работать со статистическими гипотезами на примере параметрических критериев.

## t-КРИТЕРИЙ СТЬЮДЕНТА ДЛЯ НЕСВЯЗАННЫХ (НЕЗАВИСИМЫХ) ГРУПП

Величина t-критерия является одним из значений теоретического распределения вероятности, которое известно как «t-распределение Стьюдента». Оно было получено в 1908 г. английским математиком и химиком Уильямом Госсетом (1876-1937), который публиковал свои работы под псевдонимом «Student», т.е. «студент». t-критерий Стьюдента является, пожалуй, самым часто используемым параметрическим критерием. Основными условиями его применения являются:

- Сравнение выполняют ***только по двум группам наблюдений***. Этот критерий не предназначен для сравнения нескольких групп парами (такая ошибка встречается очень часто). Другими словами, с помощью этого критерия можно сравнить характеристики мужчин и женщин (двух групп сравнения), но нельзя сравнить характеристики студентов лечебного, педиатрического и стоматологического факультетов (т.е. трех группы сравнения).
- Зарегистрированный у единиц наблюдения количественный признак должен соответствовать ***нормальному распределению в обеих сравниваемых группах***. Зачастую проверка этого условия исследователями не выполняется, что может привести к неправомерному использованию данного статистического критерия, а значит к получению неверных выводов. Если распределение признака отличается от нормального, то следует использовать непараметрические критерии.
- Сравниваемые группы наблюдений должны быть ***достаточно велики***, чтобы можно было судить о характере распределения и его параметрах. Если группы малы, то следует использовать для их сравнения непараметрические критерии.

Методика расчета t-критерия Стьюдента зависит от способа формирования сравниваемых групп наблюдений. Группы бывают:

1. **Независимые (несвязанные).** Они формируются отдельно друг от друга случайным образом, т.е. состоят из разных испытуемых.
2. **Связанные.** Связанные группы наблюдений существуют в следующих случаях:
   - Группы сформированы **методом копи-пара (парно-сопряженным методом)**, т.е. в группах есть парные друг другу единицы наблюдения. Напомним, что представляет собой этот метод формирования групп. Предположим, исследователь считает, что важными факторными признаками для данного исследования будут являться пол, возраст и стаж заболевания. Тогда из генеральной совокупности он выбирает две одинаковые по этим признакам единицы наблюдения. Одну он отправляет в основную группу, другую – в контрольную. Тогда в основной группе будет, например, мужчина в возрасте от 40 лет до 49 лет, у которого заболевание было выявлено 3 года назад, и в контрольной группе будет такая же единица наблюдения – мужчина в возрасте от 40 до 49 лет, у которого заболевание было выявлено 3 года назад.
   - В исследовании участвует одна группа испытуемых, которая помещается в разные условия (например, признак измерен до вмешательства и после вмешательства)

t-критерий Стьюдента рассчитывается по формуле:

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

где

$M_1$ и $M_2$ – средние арифметические двух несвязанных групп наблюдений,
$m_1$ и $m_2$ – стандартные ошибки этих средних арифметических,
$\sqrt{m_1^2 + m_2^2}$ – стандартная ошибка разности средних арифметических.

Если расчеты выполняются вручную, то в числителе располагают абсолютное различие средних (т.е. из большего вычитают меньшее), поэтому t-критерий всегда получается величиной положительной.

Рассмотрим, технологию проверки нулевой гипотезы на примере следующей задачи.

### Задача 1

Необходимо оценить статистическую значимость различия длительности госпитализации в больницах №1 и №2 за отчетный год. Исходные данные представлены в таблице 4.

Таблица 4

**Длительность госпитализации пациентов в больницах №1 и №2 (в днях)**

| Больница №1 ($n_1$=8) | Больница №2 ($n_2$=8) |
|---|---|
| 14 | 15 |
| 16 | 17 |
| 15 | 16 |
| 17 | 17 |
| 18 | 19 |
| 15 | 19 |
| 14 | 16 |
| 15 | 16 |
| $M_1$=124/8=**15,5 дня** | $M_2$=135/8=**16,9 дня** |

Как видим, средняя длительность госпитализации пациентов в больнице №1 меньше, чем в больнице №2 (15,5 дня против 16,9 дня). Но нами изучены всего лишь случайные выборки из множества случаев госпитализации. Может быть, случайно в выборку по больнице №1 попали пациенты с малыми сроками госпитализации, а в выборку по больнице №2 – пациенты с длительной госпитализацией? Может быть, если бы были изучены другие выборки, то вывод о длительности госпитализации оказался бы иным – в больнице №1 пациенты лечатся дольше, чем в больнице №2?

**Решение:**

Прежде чем использовать t-критерий Стьюдента для несвязанных групп, необходимо проверить соблюдение условия о нормальности распределения. И сразу же оговоримся, что ***в данном случае применять t-критерий Стьюдента НЕЛЬЗЯ, т.к. выборки очень малы***. ***Однако в учебных целях, чтобы не работать с большими массивами исходных данных, предположим, что применение t-критерия Стьюдента будет корректным***.

1) Выдвигаем нулевую и альтернативную гипотезы.
   - *Нулевая гипотеза*: средняя длительность госпитализации в больницах №1 и №2 не имеет статистически значимых различий ($H_0$: $M_1=M_2$)
   - *Альтернативная гипотеза*: средняя длительность госпитализации в больницах №1 и №2 различается ($H_1$: $M_1 \neq M_2$)
2) Выбираем t-критерий Стьюдента для несвязанных групп.
3) Рассчитаем стандартные ошибки двух выборочных средних арифметических. Поскольку выборки очень малы, в знаменатель внесем поправку – из объема выборки вычтем единицу.
   - Больница №1:

$$\sigma_1 = \sqrt{D_1} = \sqrt{2{,}00} = 1{,}41 \text{ (дня)}$$

$$m_1 = \frac{\sigma_1}{\sqrt{n_1 - 1}} = \frac{1{,}41}{\sqrt{8 - 1}} = 0{,}53 \text{ (дня)}$$

   - Больница №2:

$$\sigma_2 = \sqrt{D_2} = \sqrt{2{,}13} = 1{,}46 \text{ (дня)}$$

$$m_2 = \frac{\sigma_2}{\sqrt{n_2 - 1}} = \frac{1{,}46}{\sqrt{8 - 1}} = 0{,}55 \text{ (дня)}$$

4) Рассчитаем t-критерий Стьюдента по эмпирическим данным, воспользовавшись формулой:

$$t_{\text{эмп}} = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{16{,}9 - 15{,}5}{\sqrt{0{,}53^2 + 0{,}55^2}} = \mathbf{1{,}84}$$

5) Определим критическое значение t-критерия Стьюдента ($t_{кр}$). Для этого:
   - Определим **число степеней свободы (df)**. Для независимых выборок число степеней свободы рассчитывается следующим образом – численность каждой из сравниваемых групп уменьшается на единицу:

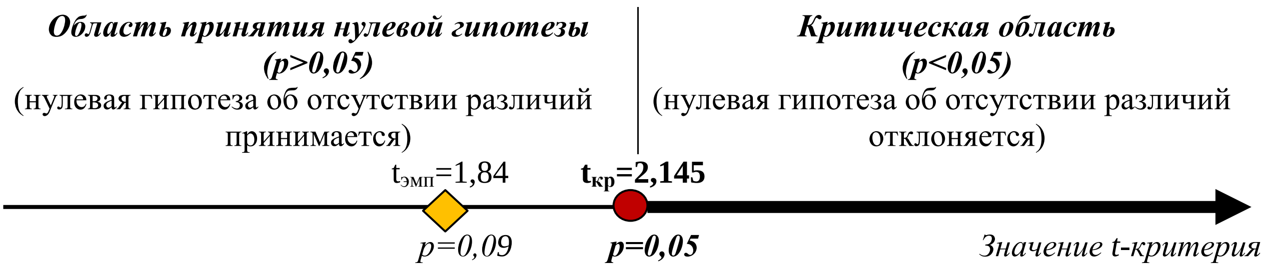
$$df = (n_1 - 1) + (n_2 - 1) = (8 - 1) + (8 - 1) = \mathbf{14}$$

   - Выбираем **двусторонний вариант t-критерия** (двустороннюю критическую область), т.к. заранее, изначально невозможно предсказать направление различия (в какой больнице средняя длительность госпитализации будет больше, а в какой – меньше).
   - Выбираем **уровень значимости p=0,05**, так как это максимально допустимая для медицинских исследований вероятность ошибочно увидеть несуществующие различия.
   - По таблице критических значений t-распределения Стьюдента (таблица 4) найдем критическую точку, соответствующую условиям нашего исследования: $\mathbf{t_{кр}=2{,}145}$

6) Сравним значение t-критерия Стьюдента, рассчитанное по эмпирическим данным, с его критическим значением (рисунок 4).

Рисунок 4

**Критическое значение t-критерия Стьюдента и формируемые им области**




**Вывод:** Для отклонения нулевой гипотезы о равенстве средней длительности госпитализации в больницах №1 и №2 достаточных аргументов не получено. Статистически значимых различий в длительности лечения пациентов в этих больницах не выявлено ($p > 0{,}05$).

## ПАРНЫЙ t-КРИТЕРИЙ СТЬЮДЕНТА ДЛЯ СВЯЗАННЫХ ГРУПП

Если сравниваемые группы наблюдений являются связанными (т.е. либо сформированы парно-сопряженным отбором, либо это одна группа испытуемых в разных состояниях), то методика расчета t-критерия Стьюдента будет уже другой:

$$t = \frac{\bar{A}}{m_{\bar{A}}}$$

где

$\bar{A}$ – средняя разность значений по каждой паре наблюдений (среднее арифметическое разностей),

$m_{\bar{A}}$ – стандартная ошибка средней разности.

Сущность метода заключается в следующем. По каждой паре наблюдений нужно определить разность значений признака (А). Таким образом, две группы значений превращаются в одну группу разностей. Именно с этим распределением и идет дальнейшая работа в соответствии с алгоритмом:

- Доказать нормальность распределения разностей
- Определить центр распределения, т.е. рассчитать среднее арифметическое (среднюю разность $\bar{A}$)
- Определить среднеквадратическое отклонение этого распределения ($\sigma_A$)
- Рассчитать стандартную ошибку средней разности ($m_{\bar{A}}$)
- Рассчитать t-критерий Стьюдента по указанной выше формуле, с помощью которого проверить нулевую гипотезу об отсутствии статистически значимых различий в двух группах наблюдений.

Обратите внимание, что в последующей задаче 2 (по связанным выборкам) цифровые данные точно такие же, как и в задаче 1 (по независимым выборкам). Но теперь представим, что эти числовые данные характеризуют изменение какого-либо количественного признака (физиологического, биохимического, морфологического и пр.) у группы из 8 испытуемых под влиянием изучаемого медицинского вмешательства. Какой это признак, и в каких единицах он измеряется – не суть важно.

***Как и в предыдущей задаче 1, сразу же оговоримся, что в данном случае применение t-критерия Стьюдента будет НЕКОРРЕКТНЫМ – слишком мала выборка (всего 8 единиц наблюдения). Однако в учебных целях, чтобы не тратить время на работу с большим массивом числовых данных, представим, что применение t-критерия будет допустимым.***

### Задача 2

Оценить статистическую значимость изменения количественного признака «К» в группе испытуемых под влиянием изучаемого медицинского вмешательства (таблица 5).

Таблица 5

**Изменение количественного признака «К» в группе испытуемых под влиянием изучаемого медицинского вмешательства**

| Результаты обследования до вмешательства | Результаты обследования после вмешательства | Изменение количественного признака |
|---|---|---|
| 14 | 15 | +1 |
| 16 | 17 | +1 |
| 15 | 16 | +1 |
| 17 | 17 | 0 |
| 18 | 19 | +1 |
| 15 | 19 | +4 |
| 14 | 16 | +2 |
| 15 | 16 | +1 |

**Решение:**

В учебных целях будем считать, что группа разностей подчиняется законам нормального распределения, и применение парного t-критерия Стьюдента в данном случае является правомерным.

1) Выдвигаем нулевую и альтернативную гипотезы.
   - *Нулевая гипотеза ($H_0$)*: статистически значимых изменений в количественной характеристике «К» под влиянием изучаемого медицинского вмешательства не произошло.
   - *Альтернативная гипотеза ($H_1$)*: под влиянием изучаемого медицинского вмешательства количественная характеристика «К» у пациентов изменяется.
2) Выполняем расчеты величин, необходимых для проверки нулевой гипотезы. Методика выполнения расчетов отражена в таблице 6.

Таблица 6

**Промежуточные величины, необходимые для расчета парного t-критерия Стьюдента для связанных групп**

| Результаты обследования до вмешательства | Результаты обследования после вмешательства | Изменение количественного признака | Отклонение изменения от среднего ($d_{\bar{A}}$) | Квадрат отклонений ($d_{\bar{A}}^2$) |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 14 | 15 | +1 | -0,38 | 0,14 |
| 16 | 17 | +1 | -0,38 | 0,14 |
| 15 | 16 | +1 | -0,38 | 0,14 |
| 17 | 17 | 0 | -1,38 | 1,90 |
| 18 | 19 | +1 | -0,38 | 0,14 |
| 15 | 19 | +4 | 2,62 | 6,86 |
| 14 | 16 | +2 | 0,62 | 0,38 |
| 15 | 16 | +1 | -0,38 | 0,14 |

- Определим среднее изменение изучаемого количественного признака ($\bar{A}$).

  $\bar{A}$ = (0 + 1*5 + 2 + 4) / 8 = 11 / 8 = **1,38**

  В среднем у каждого из 8 испытуемых изучаемая количественная характеристика увеличилась на 1,38.

  Дальнейшие расчеты позволят нам рассчитать среднеквадратическое отклонение в этом ряду, а затем и стандартную ошибку среднего.
- Определим отклонение изменения признака у каждого испытуемого от его среднего значения ($d_{\bar{A}}$) (столбец 4).
- Возведем в квадрат каждое отклонение ($d_{\bar{A}}^2$) (столбец 5).
- Рассчитаем сумму квадратов отклонений

$$\sum d_{\bar{A}}^2 = 0{,}14 \times 5 + 1{,}90 + 6{,}86 + 0{,}38 = \mathbf{9{,}84}$$

- Рассчитаем дисперсию изучаемого ряда изменения признака ($D_A$). Поскольку выборка очень мала в знаменатель внесем поправку – из объема наблюдения вычтем единицу.

$$D_A = \frac{\sum d_{\bar{A}}^2}{n-1} = \frac{9{,}84}{8-1} = \mathbf{1{,}41}$$

- Рассчитаем среднеквадратическое отклонение изучаемого ряда изменения признака ($\sigma_A$).

$$\sigma_A = \sqrt{D_A} = \sqrt{1{,}41} = \mathbf{1{,}19}$$

- Рассчитаем стандартную ошибку среднего изменения признака ($m_{\bar{A}}$). Поскольку выборка очень мала, в знаменатель еще раз внесем поправку – из объема наблюдения вычтем единицу.

$$m_{\bar{A}} = \frac{\sigma_A}{\sqrt{n_A - 1}} = \frac{1{,}19}{\sqrt{8-1}} = \frac{1{,}19}{2{,}65} = \mathbf{0{,}45}$$

3) Рассчитаем значение парного t-критерия Стьюдента для связанных групп по эмпирическим данным.

$$t_{\text{эмп}} = \frac{\bar{A}}{m_{\bar{A}}} = \frac{1{,}38}{0{,}45} = \mathbf{3{,}07}$$

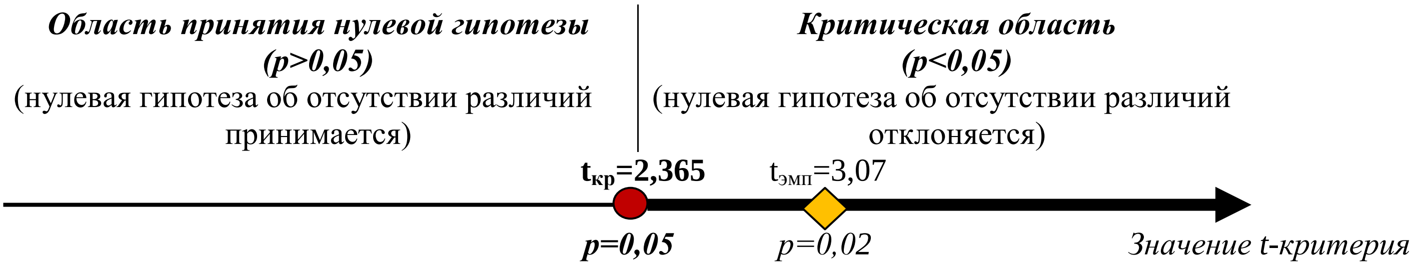
4) Найдем критическое значение t-критерия Стьюдента ($t_{кр}$).
   - Найдем **число степеней свободы (df).** Теперь у нас только одна группа значений, поэтому число степеней свободы определяется по обычной формуле:

     df = n – 1 = 8 – 1 = **7**
   - Выбираем **двусторонний вариант t-критерия** (двустороннюю критическую область), т.к. заранее, изначально невозможно уверенно предсказать направление изменений. Может быть, изучаемый количественный признак «К» будет увеличиваться, а может – уменьшаться.
   - Выбираем **уровень значимости p=0,05**, так как это максимально допустимая для медицинских научных исследований вероятность ошибочно выявить различия там, где их нет.
   - По таблице критических значений t-распределения Стьюдента найдем критическую точку, соответствующую условиям нашего исследования (таблица 4): **$t_{кр}$=2,365**

5) Сравним значение t-критерия Стьюдента, рассчитанное по эмпирическим данным, с его критическим значением для условий нашего исследования (рисунок 5) и сделаем вывод.

Рисунок 5

**Критическое значение t-критерия Стьюдента и формируемые им области**




**Вывод**: Нулевая гипотеза об отсутствии различий в характеристиках групп может быть отклонена с уровнем значимости p<0,05. Под влиянием изучаемого медицинского вмешательства количественный признак «К» действительно увеличивается.

Специалисты рекомендуют в описании полученных результатов использовать не приблизительную величину уровня значимости (как это в выводе сделали мы), а его точное значение.

То есть следует писать не «p<0,05», а «p=0,02». Точная величина уровня значимости может быть легко получена в электронной таблице Microsoft Excel с помощью функции СТЬЮДРАСП. Аргументами этой функции (т.е. то, что нужно указать в диалоговом окне) являются:

- значение t-критерия для которого требуется выполнить вычисления;
- число степеней свободы;
- хвосты (1 – для одностороннего варианта критерия, 2 – для двустороннего варианта критерия).

Итак, нам удалось отклонить нулевую гипотезу об отсутствии влияния изучаемого медицинского вмешательства на организм человека с уровнем значимости p=0,02. Другими словами, мы на 98% можем быть уверены, что данное медицинское вмешательство действительно эффективно, действительно увеличивает интересующий нас количественный признак «К» у пациентов. Группы наблюдений имеют статистически значимые различия. Эти группы, как бы, извлечены из разных генеральных совокупностей, имеющих принципиально разные характеристики.

А если нас по какой-либо причине не устраивает такая надежность выводов? Например, нам хотелось бы быть уверенными в эффективности изучаемого вмешательства на 99%. Увы, для отклонения нулевой гипотезы с такой уверенностью (т.е. с уровнем значимости p=0,01) достаточно сильных аргументов у нас нет. По таблице t-распределения легко найти критическое значение двустороннего t-критерия при уровне значимости p=0,01 для числа степеней свободы df=7. Оно равно 3,499. Наше эмпирическое значение 3,07 попадает в область принятия нулевой гипотезы.

Обратите внимание, что если бы мы рассчитывали t-критерий как в примере 1 (с независимыми группами), то получили бы число **1,84** (а не **3,07**) и не смогли бы отклонить нулевую гипотезу, т.е. не увидели бы различий там, где они есть (совершили бы ошибку второго типа, ошибку β).

## t-КРИТЕРИЙ ДЛЯ ОТНОСИТЕЛЬНЫХ ПОКАЗАТЕЛЕЙ

В рамках этой же темы рассмотрим еще один статистический критерий. На самом деле он не относится к группе параметрических, т.к. применяется для сравнения совокупностей по качественному, а не по количественному нормально распределенному признаку. Однако формула его расчета очень похожа на формулу t-критерия Стьюдента для несвязанных групп. В российской литературе его обычно называют «t-критерий для относительных показателей». А вот зарубежные авторы часто для его обозначения применяют другие символы, например, греческую букву $\zeta$ (читается «дзета»).

Формула для оценки различия двух групп по относительному показателю следующая:

$$t(\text{или } \zeta) = \frac{P_1 - P_2}{\sqrt{m_1^2 + m_2^2}}$$

где

$P_1$ и $P_2$ – относительные показатели,
$m_1$ и $m_2$ – стандартные ошибки этих относительных показателей.

Этот критерий можно применять при соблюдении ряда условий:

- Относительные показатели являются ***экстенсивными***, т.е. отражают долю единиц наблюдения, обладающих некоторым свойством, в общем объеме группы.
- Оба сравниваемых показателя находятся ***в интервале от 20% до 80%***. Если же один или оба показателя менее 20% либо более 80%, то применение данного критерия будет ошибочным – он может привести к некорректным выводам. В таких случаях используются другие статистические критерии, например, $\chi^2$ Пирсона.
- Сравниваемые группы должны быть ***достаточно велики***. Это оценивают путем проверки одновременного выполнения двух условий (для каждой группы):
    - $n * P > 5$ (если P представлено в процентах, то больше 500)

- $n * q > 5$ (если q представлено в процентах, то больше 500; напомним, что величина «q» называется альтернатива и дополняет Р до целого).

### Задача 3

Выполнен медицинский осмотр 164 детей трехлетнего возраста и 262 детей четырехлетнего возраста. Выявлены нарушения осанки функционального характера у 22% трехлетних детей и у 29% четырехлетних детей. Является ли статистически значимым различие доли числа детей с нарушениями осанки в двух возрастных группах?

**Решение:**

1) Выдвигаем нулевую и альтернативную гипотезы
   - *Нулевая гипотеза*: в сравниваемых возрастных группах статистически значимых различий в долях числа детей с нарушениями осанки нет ($H_0$: $P_1=P_2$)
   - *Альтернативная гипотеза*: в сравниваемых возрастных группах доли числа детей с нарушениями осанки имеют статистически значимые различия ($H_0$: $P_1 \neq P_2$).
2) Проверим выполнение условий, позволяющих использовать t-критерий для относительных показателей:
   - $P_1 \times n_1 = 22 \times 164 = 3\,608\ (> 500)$
   - $q_1 \times n_1 = 78 \times 164 = 12\,792\ (> 500)$
   - $P_2 \times n_2 = 29 \times 262 = 7\,598\ (> 500)$
   - $q_2 \times n_2 = 71 \times 262 = 18\,602\ (> 500)$
3) Рассчитаем t-критерий для относительных показателей
   - Рассчитаем стандартные ошибки сравниваемых относительных показателей

$$m_1 = \sqrt{\frac{P_1 * q_1}{n_1}} = \sqrt{\frac{22 * 78}{164}} = \sqrt{10{,}46} = 3{,}23\ (\%)$$

$$m_2 = \sqrt{\frac{P_2 * q_2}{n_2}} = \sqrt{\frac{29 * 71}{262}} = \sqrt{7{,}86} = 2{,}80\ (\%)$$

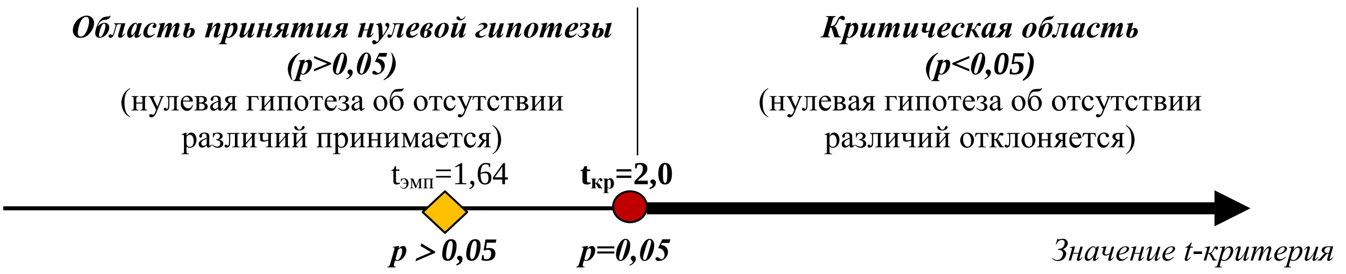
   - Рассчитаем t-критерий для относительных показателей по эмпирическим данным

$$t_{\text{эмп}} = \frac{\mathrm{P}_1 - \mathrm{P}_2}{\sqrt{m_1^2 + m_2^2}} = \frac{29 - 22}{\sqrt{3{,}23^2 + 2{,}80^2}} = \frac{7}{4{,}27} = \mathbf{1{,}64}$$

4) Критическое значение t-критерия для относительных показателей при уровне значимости p=0,05 равно 2,0.
5) Сравним значение t-критерия для относительных показателей, рассчитанное по эмпирическим данным, с его критическим значением (рисунок 6).

Рисунок 6

**Критическое значение t-критерия для относительных показателей и формируемые им области**




**Вывод**: Нулевая гипотеза об отсутствии статистически значимых различий в долях числа детей с нарушениями осанки в сравниваемых возрастных группах не может быть отклонена. С трехлетнего до четырехлетнего возраста осанка у детей не ухудшается ($p > 0{,}05$).