

ТЕМА 5

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И ИХ ПРОВЕРКА

В прикладных задачах часто требуется по наблюдениям выборки высказать некоторое суждение (*гипотезу*) относительно интересующих экспериментатора характеристик генеральной совокупности, из которой эта выборка извлечена. То есть, речь идет о *проверке статистических гипотез*.

Гипотеза – это некоторое предположение о параметрах известных распределений (параметрическая) или о виде неизвестного закона распределения (непараметрическая) случайных величин, выдвигаемое в качестве предварительного, условного объяснения.

Теория проверки статистических гипотез является основным инструментом доказательной, а не интуитивной медицины.

Задачи медицинских и биологических исследований, для решения которых необходимо сформулировать статистические гипотезы:

- анализ соответствия распределения значений признака в изучаемой группе какому-либо определенному закону (анализ соответствия распределения нормальному закону);
- сравнение групп по параметрам распределений признака (по средним значениям, дисперсиям).

Например, при проверке статистических гипотез можно получить ответ на следующий вопрос. В двух однородных группах больных гриппом была проведена вакцинация: одной лекарственным средством «А», а другой - «В», среднее время выздоровления в группах неодинаково. Указывает ли это обстоятельство на то, что одно противогриппозное средство по эффективности превосходит другое или же выявленное различие случайно?

Для решения любой подобной задачи выдвигаются две статистические гипотезы:

- **нулевая гипотеза H_0** - гипотеза об отсутствии различий между группами, либо об определенных значениях параметров, либо о соответствии распределения нормальному закону;

- **альтернативная гипотеза H_1** - гипотеза о существовании различий между группами, либо об отличающихся от заданных значениях параметров, либо о несоответствии распределения нормальному закону.

Нулевая гипотеза формулируется таким образом, чтобы она была противоположной той исследовательской (медицинской, биологической) гипотезе, которая послужила поводом для проведения исследования.

Для проверки нулевой гипотезы применяют статистические методы (тесты, критерии).

Статистическими критериями называются правила, согласно которым выясняется, соответствует или нет интересующая нас гипотеза опытным данным.

Статистические критерии - это наиболее широко применяемые статистические средства.

Значение критерия, которое рассчитано по выборочной совокупности, подчиняющейся определенному закону распределения, называется *наблюдаемым*.

Множество возможных значений статистического критерия, при которых основная гипотеза принимается, называется *областью принятия*.

Множество возможных значений статистического критерия, при которых основная гипотеза отвергается, называется *критической областью*.

Точки, разграничивающие критическую область и область принятия гипотезы, называются *критическими точками*.

При проверке статистических гипотез возникают следующие виды ошибок:

- *ошибка первого рода* – это вероятность отвергнуть правильную нулевую гипотезу;
- *ошибка второго рода* – это вероятность принять неправильную нулевую гипотезу.

Ошибка первого рода иначе называется *уровнем статистической значимости*. Уровень значимости - это максимально приемлемая для исследователя вероятность ошибочно отклонить нулевую гипотезу, когда она на самом деле верная, т.е. допускаемая исследователем величина ошибки первого рода. Величина уровня значимости устанавливается исследователем произвольно, однако обычно принимается равным 0,05, 0,01 или 0,001.

Вероятность ошибки второго рода не имеет какого-то особого общепринятого названия, на письме обозначается греческой буквой β . Однако с этой величиной тесно связана другая, имеющая большое статистическое значение – **мощность (чувствительность) критерия**. Она вычисляется по формуле $(1 - \beta)$. Таким образом, чем выше мощность, тем меньше вероятность совершить ошибку второго рода.

Таблица 6.1.

Возможные решения при различных соотношениях результатов статистического теста и истинной ситуации в генеральной совокупности

| | | В генеральной совокупности | |
|------------------------|--------------------|---|--|
| | | H_0 неверна | H_0 верна |
| В статистическом тесте | H_0 отклонена | Истинно-положительный результат | Ложно-положительный результат (ошибка первого рода или α -ошибка) |
| | H_0 не отклонена | Ложно-отрицательный результат (ошибка второго рода или β -ошибка) | Истинно-отрицательный вариант |

Слово «отрицательный» в данном случае не имеет отношения к желательности или нежелательности самого события.

Термин широко используется в медицине. Например, тесты, предназначенные для диагностики заболеваний, иногда дают отрицательный результат (т.е. показывают отсутствие заболевания у пациента), когда, на самом деле пациент страдает этим заболеванием. Такой результат называется ложноотрицательным.

В ходе применения статистического метода вычисляется значение тестовой статистики (например, при применении критерия Стьюдента – значение t), а также соответствующее ему и числу степеней свободы значение p – вероятность справедливости нулевой гипотезы.

Схема проверки статистических гипотез:

1) Выдвигаются две гипотезы: основная (нулевая) « H_0 » и альтернативная (конкурирующая) « H_1 ».

2) Задается уровень значимости « p ». Статистический вывод никогда не может быть сделан со стопроцентной уверенностью. Всегда допускается риск принятия неправильного

решения. При проверке статистических гипотез мерой такого риска является уровень значимости « p ».

3) По исходным данным, т.е. по выборке, вычисляется наблюдаемое (эмпирическое, расчетное) значение критерия.

4) По специальным статистическим таблицам определяется табличное (критическое) значение критерия.

5) Путем сравнения наблюдаемых и табличных значений делается вывод о правильности той или иной гипотезы.

- если рассчитанное в статистическом тесте значение p оказывается больше принятого уровня значимости, то нулевую гипотезу H_0 не отклоняют и различия групп называются **статистически незначимыми**.

- если значение p оказывается меньше уровня значимости, то нулевую гипотезу H_0 отклоняют, при этом следует принять альтернативную гипотезу H_1 . В данном случае различия групп называют **статистически значимыми** (при $p < 0,05$) или **статистически высокосignимыми** (при $p < 0,01$).

В процессе медико-биологических исследований часто возникает проблема сравнения результатов обследования (например, в контрольной и экспериментальной группах или до и после эксперимента).

Для решения этой проблемы существует большое количество статистических критериев. Каждый из них имеет свою специфику, отличаясь друг от друга (например, типами данных, объемами выборок, количеством сравниваемых выборок, качеством сравниваемых выборок (зависимая и независимая) и др.).

Наиболее популярным из таких критериев является t -критерий Стьюдента, который применяется примерно в 30-40% научных медицинских работ

Сравнение связанных и несвязанных выборок производится с помощью разных критериев (табл. 6.2).

Таблица 6.2.

Рекомендуемые к использованию статистические критерии в зависимости от задачи исследования и типа данных.

| Задача | Метод | |
|---|---|---|
| | параметрический | Непараметрический |
| Сравнение двух независимых групп по одному признаку | t -критерий Стьюдента для независимых выборок | Критерии Манна-Уитни, Колмогорова-Смирнова, Вальда-Вольфовица, критерий χ^2 , точный критерий Фишера |
| Сравнение двух зависимых групп по одному признаку | t -критерий Стьюдента для зависимых выборок | Критерий Вилкоксона, критерий знаков, критерий Мак-Нимара |
| Сравнение трех и более независимых групп по одному признаку | ANOVA | ANOVA по Краскелу-Уоллису, медианный критерий, критерий χ^2 |
| Сравнение трех и более зависимых групп по одному признаку | Критерий Кокрена | Критерий Кокрена, ANOVA по Фридману |

Параметрические критерии для проверки гипотезы о различии (или сходстве) между средними значениями

t-критерий Стьюдента – метод проверки однородности выборок, позволяет принять или отвергнуть гипотезу о равенстве средних двух выборок ($H_0: \bar{x}_1 = \bar{x}_2$).

t-критерий Стьюдента используется:

- при проверке гипотезы о равенстве средних двух *независимых* выборок (*двухвыборочный t-критерий*). В этом случае анализируются контрольная и экспериментальная выборки разных объемов. Например, группа больных сахарным диабетом и группа здоровых людей;
- при проверке гипотезы о равенстве средних двух *зависимых* выборок (*парный t-критерий*). В этом случае анализируется одна и та же выборка, но до и после эксперимента. Например, средняя частота пульса у одних и тех же пациентов до и после приема антиаритмического препарата.

Применение критерия Стьюдента возможно, если выполняются следующие два условия:

- рассматриваемые выборки имеют нормальное распределение;
- дисперсии рассматриваемых выборок равны.

Непараметрические критерии для проверки гипотезы о различии (или сходстве) между средними значениями

Сравнение двух независимых групп

U-критерий Уилкоксона (Манна-Уитни) используется для проверки гипотезы о принадлежности сравниваемых *независимых выборок* одной и той же генеральной совокупности.

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны.

Ограничения применимости критерия:

1. В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти.
2. В выборочных данных не должно быть совпадающих значений (все числа – разные) или таких совпадений должно быть очень мало.

Критерий серий Вальда-Вольфовица предназначен для проверки нулевой гипотезы о равенстве целого ряда параметров двух выборок, включая медианы и коэффициента асимметрии.

Сравнение двух зависимых групп

T-критерий Уилкоксона используется в случае *парно связанных* выборок. При этом ранжируют попарные разности – положительные и отрицательные (кроме нулевых) в один ряд так, чтобы наименьшая абсолютная разница (без учета знака) получила первый ранг, одинаковым величинам присваивают один ранг. Отдельно вычисляю т сумму рангов положительных (T^+) и отрицательных (T^-) разностей. Меньшую из двух таких сумм без учета знака считают тестовой статистикой данного критерия. Нулевую гипотезу принимают на данном уровне значимости, если вычисленная статистика превзойдет табличное значение.

Сравнение средних значений нескольких выборок (множественные сравнения)

Если план исследования включает сравнение большего числа групп (больше, чем две группы), совершенно недопустимо просто сравнивать их попарно. Для корректного решения

этой задачи можно воспользоваться, например, дисперсионным анализом. Однако дисперсионный анализ позволяет проверить лишь гипотезу о равенстве всех сравниваемых средних. Но, если гипотеза не подтверждается, нельзя узнать, какая именно группа отличалась от других. Это позволяет сделать методы множественного сравнения, которые в свою очередь также бывают *параметрические* и *непараметрические*. Эти методы дают возможность провести множественные сравнения так, чтобы вероятность хотя бы одного неверного заключения оставалась на первоначальном выбранном уровне значимости, например, 5%.

Параметрические критерии

Критерий Стьюдента для множественных сравнений основан на использовании **неравенства Бонферрони**: если k -раз применить критерий с уровнем значимости α , то вероятность хотя бы в одном случае найти различие там, где его нет, не превышает произведения k на α . Этот метод работает, если число сравнений невелико, обычно не больше 8. При большем числе сравнений **критерий Ньюмана-Кейлса** и **Тьюки** дают более точную оценку вероятности α .

Критерий Даннета более чувствительный, чем предыдущий, особенно при большом числе групп. Критерий Даннета является модификацией критерия Ньюмана-Кейлса. Для проверки критерия средние значения упорядочиваются по абсолютной величине их отличия от контрольной группы, сравнения начинают с группы, наиболее отличающейся от контроля. Если различия с очередной группой не найдены, сравнения прекращаются.

Непараметрические критерии

Критерий Краскела-Уоллиса – непараметрический критерий для сравнения средних значений нескольких независимых выборок – основан на построении объединенного вариационного ряда из вариантов рассматриваемых выборок и присвоении рангов всем вариантам в объединенном ряду, предназначен для проверки равенства медиан нескольких выборок.

Критерий Фридмана – это непараметрический аналог дисперсионного анализа повторных измерений, применяется для анализа повторных измерений, связанных с одним и тем же индивидуумом. Порядок проведения множественных сравнений средних значений представлен на рис. 6.2.

ANOVA (дисперсионный анализ)

Дисперсионный анализ был разработан английским математиком Р.Фишером. Его чаще используют в научно-практических исследованиях общественного здоровья и здравоохранения для изучения влияния одного или нескольких факторов на результативный признак

Оценка значимости различия при альтернативной форме учета реакций. Для оценки значимости расхождения частот какого-либо явления в двух группах может быть использован статистический метод, который носит название **критерия χ^2** . Этот критерий может быть применен, например, при сравнении групп, получивших различные сравниваемые по своей активности препараты; групп, получивших различные дозы изучаемого препарата или одну и ту же дозу различными путями введения и т.д. Для описания результатов такого исследования удобно применять **таблицу сопряженности**, в которой для каждой из групп указывается число пациентов с каждым из градаций признака. Таким образом, для 2-х рассматриваемых групп и 2-х возможных исходов получается таблица размерности 2x2 (рис. 6.1). Для ответа на вопрос о значимости различий между группами вычисляется величина статистики χ^2 , которая является показателем максимально возможных при данном уровне значимости отклонений частот.

Критерий χ^2 может применяться и к таблице сопряженности произвольной размерности.

Точный критерий Фишера основан на переборе всех возможных вариантах заполнения таблицы сопряженности при данной численности групп. Позволяет получить точные значения вероятности событий, столь же или еще менее вероятных, чем те, которые наблюдались в действительности.

Критерий Мак-Нимара применяется для анализа связанных измерений в случае измерения реакции для связанной переменной. Является аналогом параметрического **критерия Стьюдента** для зависимых выборок или непараметрического **T-критерий Уилкоксона**.

Порядок проведения парных и множественных сравнений качественных признаков представлен на рис. 6.1, 6.2.

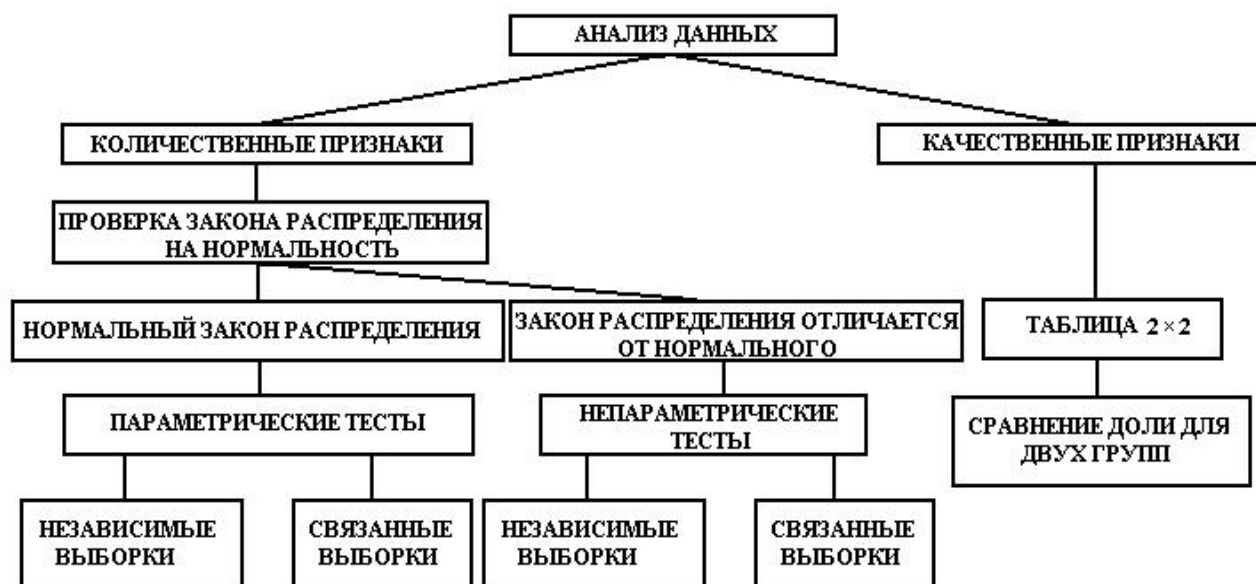


Рис 6.1. Схема проведения парного сравнения средних значений.

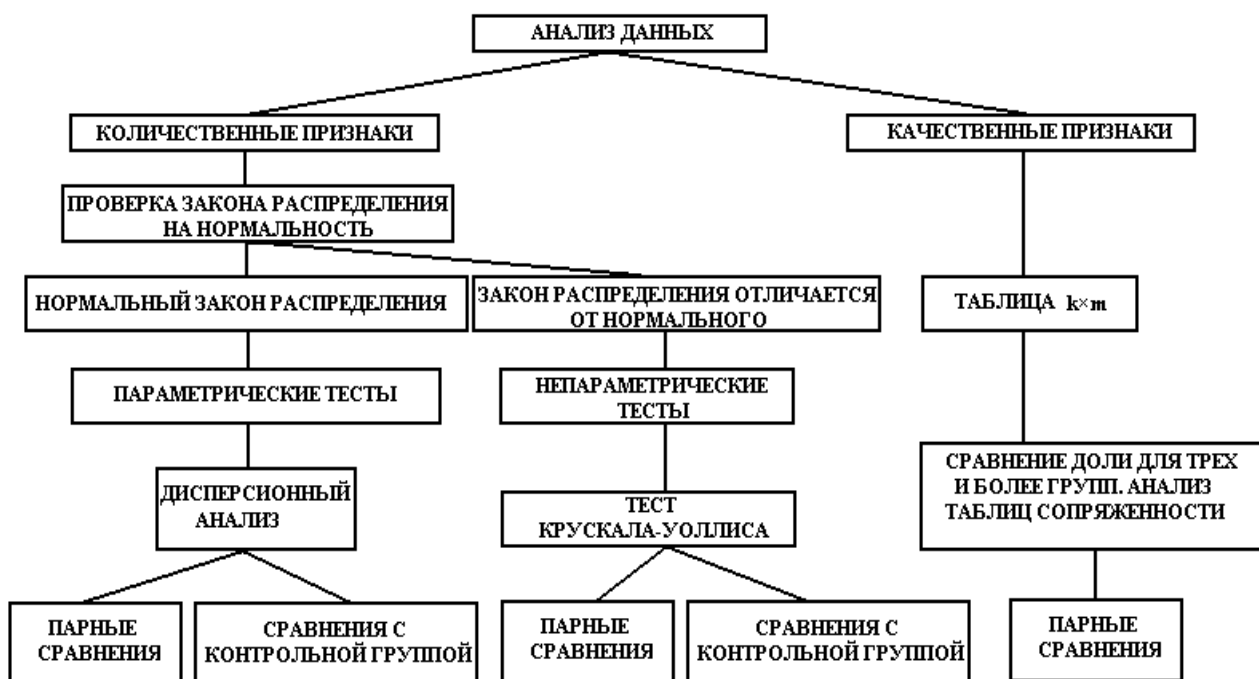


Рис. 6.2. Схема проведения множественного сравнения.

Контрольные вопросы

1. Что такое статистическая гипотеза?
2. Что такое нулевая гипотеза?
3. Что такое альтернативная гипотеза?
4. Что такое ложноположительный и ложноотрицательный результат проверки статистического теста?
5. Что такое ошибка первого рода?
6. Что такое ошибка второго рода?
7. Что такое мощность критерия?

8. Какие параметрические критерии используются для проверки статистических гипотез?
9. Какие непараметрические критерии используются для проверки статистических гипотез?
10. Приведите примеры статистических критериев, используемых для множественных сравнений средних значений нескольких выборок.

Список литературы

1. Гланц С. Медико-биологическая статистика. Пер. с англ. – М.: Практика, 1998. – 459 с.
2. Лях Ю.Е., Гурьянов В.Г., Хоменко В.Н., Панченко О.А. Основы компьютерной биостатистики: анализ информации в биологии, медицине и фармации статистическим пакетом Medstat. – Донецк: 2006. – 214 с.
3. Платонов А.Е. Статистический анализ в медицине и биологии: задача, терминология, логика, компьютерные методы. – М.: Издательство РАМН, 2000. – 52 с.
4. Реброва О.Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. – М.: МедиаСфера, 2002. – 312 с.
5. Сергиенко В.И., Бондарева И.Б. Математическая статистика в клинических исследованиях. - М.: ГЭОТАР-МЕД, 2001. – 256 с.