

Лекция

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ РАЗЛИЧИЙ

План

- 1. Параметрические и непараметрические критерии**
 - 2. Рекомендации к выбору критерия различия**
 - 3. Непараметрические критерии для *зависимых выборок***
 - Критерии знаков (G-критерий)**
 - Парный критерий T – Вилкоксона**
 - 4. Непараметрические критерии для *независимых выборок***
 - U - критерий Манна-Уитни**
 - H - критерий Крускала-Уоллиса**
-

задача сравнения результатов обследования,
какого либо *признака в разных условиях*
измерения (например, до и после определенного
воздействия) или обследования контрольной и
экспериментальной групп, оценка *характера*
изменения того или иного *психологического*
показателя в одной или нескольких группах в
разные периоды времени или *выявить динамику*
изменения этого показателя под влиянием
экспериментальных воздействий.

Для решения подобных задач используется достаточно большой **набор статистических способов** называемых в наиболее общем виде ***критериями различий***.

Эти критерии позволяют оценить ***степень статистической достоверности различии*** между ***разнообразными показателями***, измеренными согласно плану проведения исследования.

Важно : уровень достоверности различии включается в план проведения эксперимента т.е. исследователь при постановке экспериментальной задачи априори выбирает **уровень достоверности различии** (как правило, от 5% и выше в зависимости от особенностей решаемой задачи) который будет считаться приемлемым.

Существует достаточно большое количество критериев различий, которые имеет свою специфику различаясь между собой *по различным основаниям.*



Основания: *тип измерительной шкалы*, для которой предназначен тот или иной критерий.

Например, с помощью некоторых критериев можно обрабатывать данные, полученные только в номинальных шкалах. Ряд критериев дает возможность обрабатывать данные, полученные в порядковой, интервальной и шкале отношений.

Критерии различаются

- **по максимальному объему выборки**, который они могут охватить,
- **по количеству выборок**, которые можно сравнивать между собой с их помощью. Так, существуют критерии, позволяющие оценить различия сразу в трех и большем числе выборок.
- Некоторые критерии позволяют сопоставлять **неравные по численности выборки**.
- Дифференцирующим критерии признаком, служит **само качество выборки**, она может быть **связной (зависимой)** или **несвязной (независимой)**.
- Выборки **взяты из одной или нескольких генеральных совокупностей**. Именно эта характеристика выборки служит наиболее важным основанием, по которому, прежде всего, выбираются критерии. Кроме того, критерии различаются **по мощности**.

Мощность критерия — это его способность выявлять различия или отклонять нулевую гипотезу, если она неверна. Напомним, что ошибке первого рода соответствует отказ от нулевой гипотезы. Можно сказать также, что мощность критерия характеризует его способность избегать ошибки второго рода.

Задачи	Условия	Методы
1. Выявление различий в уровне исследуемого признака	а) 2 выборки испытуемых	Q критерий Розенбаума U критерий Манна - Уитни Ф критерий (угловое преобразование Фишера)
	б) 3 и больше выборок испытуемых	S критерий Джонкира H критерий Крускала - Уоллиса
2. Оценка сдвига значений исследуемого признака	а) 2 замера на одной и той же выборке испытуемых	T критерий Вилкоксона G критерий знаков Ф критерий (угловое преобразование Фишера) t - критерий Стьюдента
	б) 3 и более замеров на одной и той же выборке испытуемых	x критерий Фридмана L критерий тенденций Пейджа t - критерий Стьюдента
3. Выявление различий в распределении признака	а) при сопоставлении эмпирического распределения с теоретическим	χ^2 критерий Пирсона X критерий Колмогорова - Смирнова t - критерий Стьюдента
	б) при сопоставлении двух эмпирических распределений	X критерий Пирсона X критерий Колмогорова - Смирнова Ф критерий (угловое преобразование Фишера)
4. Выявление степени согласованности изменений	а) двух признаков	Ф коэффициент корреляции Пирсона t коэффициент корреляции Кен- далла R бисериальный коэффициент корреляции П корреляционное отношение Пирсона r_s коэффициент ранговой корреляции Спирмена r коэффициент корреляции Пирсона Линейная и криволинейная регрессии
	б) трех или большего числа признаков	r_s коэффициент ранговой корреляции Спирмена r коэффициент корреляции Пирсона Множественная и частная корреляции Линейная, криволинейная и множественная регрессия Факторный и кластерный анализы
5. Анализ изменений признака под влиянием контролируемых условий	а) под влияние одного фактора	S критерий Джонкира L критерий тенденций Пейджа Однофакторный дисперсионный анализ
	б) под влиянием двух факторов одновременно	Двухфакторный дисперсионный анализ

Большое разнообразие критериев различия **предоставляет следующие возможности**

- выбирать критерий адекватный типу шкалы, в которой получены экспериментальные данные;
 - работать со связными (зависимыми) и несвязными (независимыми) выборками;
 - работать с неравными по объему выборками;
 - выбирать из критериев разные по мощности (в зависимости от целей исследования);
-

Параметрические и непараметрические критерии

Все критерии различий условно подразделены на две группы параметрические и непараметрические критерии

- Критерии различия называют **параметрическим**, если он основан *на конкретном типе распределения генеральной совокупности* (как правило, нормальном) или использует параметры этой совокупности (средние, дисперсии и т.д.).
 - Критерий различия называют **непараметрическим**, если он *не базируется на предположении о типе распределения генеральной совокупности* и не использует параметры этой совокупности. Для непараметрических критериев предлагается также использовать такой термин как *критерии свободный от распределения*.
-

при оценке различий в *распределениях далеких от нормального* непараметрические критерии могут выявить значимые различия, в то время как параметрические критерии таких различий не обнаружат. Важно отметить, что,

- 1.** во-первых, непараметрические критерии выявляют значимые различия и в том случае если распределение близко к нормальному,
 - 2.** во-вторых, при вычислениях вручную непараметрические критерии являются значительно менее трудоемкими, чем параметрические.
-

2. Рекомендации к выбору критерия различия

При подготовке экспериментального исследования необходимо заранее запланировать характеристики сопоставляемых выборок (прежде всего *связность* — *несвязность и однородность*), их величину (объем), тип измерительной шкалы и вид используемого критерия различия.

Последовательно это можно представить в виде поэтапного алгоритма

1. Прежде всего, следует определить является ли **выборка связанной** (зависимой) или несвязанной (независимой).
 2. Следует определить **однородность** (неоднородность) **выборки**.
 3. Затем следует оценить **объем выборки** и, зная ограничения каждого критерия по объему выбрать соответствующий критерий.
 4. При этом целесообразнее всего начинать работу с выбора наименее **трудоемкого критерия**.
 5. Если используемый критерий не выявил различия — следует применить **более мощный**, (но одновременно и более трудоемкий) критерий.
 6. Если в распоряжении психолога имеется несколько критериев, то следует выбирать те из них, которые **наиболее полно используют информацию**, содержащуюся в экспериментальных данных.
 7. При малом объеме выборки следует увеличивать **величину уровня значимости** (не менее 1%) так как небольшая выборка и низкий уровень значимости приводят к увеличению вероятности принятия ошибочных решений.
-

3. Непараметрические критерии для связанных выборок

3.1. Критерии знаков G

Назначение критерия G

Критерий знаков G предназначен для установления *общего направления сдвига* исследуемого признака.

Сравнивая результаты до и «после какого либо воздействия (например, тренинга) исследователь видит тенденции повторного измерения — большинство показателей может увеличиваться или уменьшаться.

Он позволяет установить, в какую сторону в выборке в целом изменяются значения признака при переходе от первого измерения ко второму: изменяются ли показатели в сторону улучшения, повышения или усиления или, наоборот, в сторону ухудшения, понижения или ослабления.

Критерий знаков G относится к непараметрическим и применяется только для *связанных (зависимых) выборок*.

Он дает возможность установить, *насколько однонаправленно изменяются значения признака при повторном измерении связанной, однородной выборки*.

Критерий знаков применяется к данным, полученным в ранговой, интервальной и шкале отношений.

Ограничения критерия знаков

Количество наблюдений в обоих замерах - не менее 5 и не более 300

- **Суть критерия знаков G** состоит в том, что он определяет, не слишком ли много наблюдается "нетипичных сдвигов", чтобы сдвиг в "типичном" направлении считать преобладающим? Ясно, что чем меньше "нетипичных сдвигов", тем более вероятно, что преобладание "типичного" сдвига является преобладающим.

$G_{ЭМП}$ - это количество "нетипичных" сдвигов. Чем меньше $G_{ЭМП}$, тем более вероятно, что сдвиг в "типичном" направлении статистически достоверен.

Гипотезы

- H_0 : Преобладание типичного направления сдвига является случайным.
 - H_1 : Преобладание типичного направления сдвига не является случайным.
-

АЛГОРИТМ Расчет критерия знаков G

1. Подсчитать количество нулевых реакций и исключить их из рассмотрения. В результате n уменьшится на количество нулевых реакций.
 2. Определить преобладающее направление изменений. Считать сдвиги в преобладающем направлении "типичными".
 3. Определить количество "нетипичных" сдвигов. Считать это число эмпирическим значением $G_{\text{эмп}}$.
 4. По Таблице Приложения определить критические значения $G_{\text{кр}}$ для данного n .
 5. Сопоставить $G_{\text{эмп}}$ с $G_{\text{кр}}$. Если $G_{\text{эмп}}$ меньше $G_{\text{кр}}$ или по крайней мере равен ему, сдвиг в типичную сторону может считаться достоверным.
-

критерии знаков G является одним из самых простых по способу вычисления. Традиционно он считается одним из *наименее мощных*. Однако можно утверждать, что если критерии знаков *показал значимые различия* на 1% уровне, то другие более мощные критерии подтвердят эти различия. В то же время, если критерий знаков не выявил значимых различий, возможно, что более мощные критерии, напротив, такое различие выявят.

Для применения **критерия G** необходимо соблюдать следующие условия:

- Измерение может быть проведено в шкале порядка, интервалов и отношений
 - Выборка должна быть однородной и связной
 - Число элементов в сравниваемых выборках должно быть равным
 - **критерий знаков G** может применяться при величине типичного сдвига от 5 до 300 (на большую величину не рассчитана таблица достоверности)
 - При большом числе сравниваемых парных значений критерий знаков достаточно эффективен
 - При равенстве типичных и нетипичных сдвигов критерий знаков неприменим, следует использовать другие критерии.
-

3.2. Парный критерий T – Вилкоксона

Для решения задач, в которых осуществляется сравнение двух рядов чисел (выборок), кроме критерия знаков G можно использовать парный критерий T - Вилкоксона.

Этот критерий является более мощным, чем критерий знаков G , и применяется для *оценки различий экспериментальных данных, полученных в двух разных условиях на одной и той же выборке испытуемых*. Он позволяет выявить *не только направленность изменений, но и их выраженность*, т е он позволяет установить, насколько сдвиг показателей в каком-то одном направлении является более интенсивным, чем в другом.

Критерий T основан на ранжировании абсолютных величин разности между двумя рядами выборочных значений в первом и втором эксперименте (например «до» и «после» какого-либо воздействия).

Ранжирование абсолютных величин означает, что **знаки разностей не учитываются**, однако в дальнейшем наряду с общей суммой рангов находится отдельно сумма рангов как для положительных так и для отрицательных сдвигов.

Если интенсивность сдвига в одном из направлении оказывается большей то и соответствующая сумма рангов также оказывается больше.

Этот сдвиг, как и в случае критерия знаков, называется **типичным**, а противоположный меньший по сумме рангов сдвиг — **нетипичным**. Как и для критерия знаков эти два сдвига оказываются дополнительными друг к другу.

Критерии T— Вилкоксона базируется на величине нетипичного сдвига который называется в дальнейшем $T_{эм.}$

Гипотезы

H_0 : Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

H_1 : Интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении.

Ограничения в применении критерия Вилкоксона

1. Минимальное количество испытуемых, прошедших измерения в двух условиях - 5 человек. Максимальное количество испытуемых - 50 человек, что диктуется верхней границей имеющихся таблиц. Критические значения T приведены в Табл. Приложения
 2. Нулевые сдвиги из рассмотрения исключаются, и количество наблюдений n уменьшается на количество этих нулевых сдвигов. Можно обойти это ограничение, сформулировав гипотезы, включающие отсутствие изменений, например: "Сдвиг в сторону увеличения значений превышает сдвиг в сторону уменьшения значений и тенденцию сохранения их на прежнем уровне".
-

АЛГОРИТМ критерия Вилкоксона

1. Составить список испытуемых в любом порядке, например, алфавитном.
2. Вычислить разность между индивидуальными значениями во втором и первом замерах ("после" - "до"). Определить, что будет считаться "типичным" сдвигом и сформулировать соответствующие гипотезы.
3. Перевести разности в абсолютные величины и записать их отдельным столбцом (иначе трудно отвлечься от знака разности).
4. Проранжировать абсолютные величины разностей, начисляя меньшему значению меньший ранг. Проверить совпадение полученной суммы рангов с расчетной.
5. Отметить кружками или другими знаками ранги, соответствующие сдвигам в "нетипичном" направлении.
6. Подсчитать сумму этих рангов по формуле:

$$T = \sum R_r$$

где R_r - ранговые значения сдвигов с соответствующим знаком.

7. Определить критические значения T для данного n по Табл. VI Приложения 1. Если $T_{эмп}$ меньше или равен $T_{кр}$, сдвиг в "типичную" сторону по интенсивности достоверно преобладает.

Для применения критерия T - Вилкоксона необходимо соблюдать следующие условия.

- Измерение может быть проведено во всех шкалах, кроме номинальной.
 - Выборка должна быть связной.
 - Число элементов в сравниваемых выборках должно быть равным.
 - Критерий T — Вилкоксона может применяться при численности выборки от 5 до 50 (на большую величину не рассчитана таблица достоверности).
-

4.1. U - критерий Манна-Уитни

Назначение критерия

Критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного. Он позволяет выявлять различия между *малыми* выборками, когда $n_1 \cdot n_2 \geq 3$ или $n_1 = 2, n_2 \geq 5$.

Описание критерия

Существует несколько способов использования критерия и несколько вариантов таблиц критических значений, соответствующих этому способу).

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами значений выборки. Мы 1-й выборкой называем тот ряд значений, в котором значения, по предварительной оценке, выше, а 2-й выборкой - тот, где они предположительно ниже.

Чем меньше область перекрещивающихся значений, тем более вероятно, что различия достоверны. Иногда эти различия называют различиями в *расположении* двух выборок .

Эмпирическое значение критерия $U_{\text{эмп}}$ отражает то, *насколько велика зона совпадения между рядами.*

Поэтому *чем меньше $U_{\text{эмп}}$, тем более вероятно, что различия достоверны.*

Гипотезы

- H_0 : Уровень признака в выборке 2 не ниже уровня признака в выборке 1.
 - H_1 : Уровень признака в выборке 2 ниже уровня признака в выборке 1.
-

Ограничения критерия U

1. В каждой выборке должно быть не менее 3 наблюдений: $n_1, n_2 \geq 3$; допускается, чтобы в одной выборке было 2 наблюдения, но тогда во второй их должно быть не менее 5.

2. В каждой выборке должно быть не более 60 наблюдений; $n_1, n_2 \leq 60$. Однако уже при $n_1, n_2 > 20$ ранжирование становится достаточно трудоемким.

(в случае, если $n_1, n_2 > 20$, лучше использовать другой критерий, а именно угловое преобразование Фишера)

АЛГОРИТМ критерия U Манна-Уитни

1. Перенести все данные испытуемых на индивидуальные карточки.
 2. Пометить карточки испытуемых выборки 1 одним цветом, скажем красным, а все карточки из выборки 2 - другим, например синим.
 3. Разложить все карточки в единый ряд по степени нарастания признака, не считаясь с тем, к какой выборке они относятся, как если бы мы работали с одной большой выборкой.
 4. Проранжировать значения на карточках, приписывая меньшему значению меньший ранг. Всего рангов получится столько, сколько у нас $(n_1 + n_2)$.
 5. Вновь разложить карточки на две группы, ориентируясь на цветные обозначения: красные карточки в один ряд, синие - в другой.
-

6. Подсчитать сумму рангов отдельно на красных карточках (выборка 1) и на синих карточках (выборка 2). Проверить, совпадает ли общая сумма рангов с расчетной.

7. Определить большую из двух ранговых сумм.

8. Определить значение **U** по формуле:

$$U = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x$$

где n_1 - количество испытуемых в выборке 1;

n_2 - количество испытуемых в выборке 2;

T_x - большая из двух ранговых сумм;

n_x - количество испытуемых в группе с большей суммой рангов.

9. Определить критические значения **U** по Табл. Приложения .

Если $U_{\text{эмп}} > U_{\text{кр } 005}$, то гипотеза H_0 принимается.

Если $U_{\text{эмп}} \leq U_{\text{кр } 005}$, то гипотеза H_0 отвергается.

Чем меньше значения **U**, тем достоверность различий выше.

Правила ранжирования

1. Меньшему значению начисляется меньший ранг. Наименьшему значению начисляется ранг 1.

Наибольшему значению начисляется ранг, соответствующий количеству ранжируемых значений. Например, если $n=7$, то наибольшее значение получит ранг 7, за возможным исключением для тех случаев, которые предусмотрены правилом 2.

2. В случае, если несколько значений равны, им начисляется ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны.

Например, 3 наименьших значения равны 10 секундам. Если бы мы измеряли время более точно, то эти значения могли бы различаться и составляли бы, скажем, 10,2 сек; 10,5 сек; 10,7 сек. В этом случае они получили бы ранги, соответственно, 1, 2 и 3. Но поскольку полученные нами значения равны, каждое из

$$\frac{1+2+3}{3} = \frac{6}{3} = 2$$

Допустим, следующие 2 значения равны 12 сек. Они должны были бы получить ранги 4 и 5, но, поскольку они равны, то получают средний

$$\frac{4+5}{2} = 4.5 \text{ и т. д.}$$

3. Общая сумма рангов должна совпадать с расчетной, которая определяется по формуле:

$$\sum(R_i) = \frac{N \cdot (N + 1)}{2}$$

где N - общее количество ранжируемых наблюдений (значений). Несовпадение реальной и расчетной сумм рангов будет свидетельствовать об ошибке, допущенной при начислении рангов или их суммировании. Прежде чем продолжить работу, необходимо найти ошибку и устранить ее.

Н - критерий Крускала-Уоллиса

Назначение критерия

Критерий предназначен для оценки различий одновременно между *тремя, четырьмя* и т.д. выборками по *уровню* какого-либо признака.

Он позволяет установить, что *уровень* признака *изменяется* при переходе от группы к группе, но не указывает на направление этих изменений.

Описание критерия

Данный критерий является продолжением *критерия U* на большее (чем 2) количество сопоставляемых выборок. Все индивидуальные значения ранжируются так, как если бы это была одна большая выборка. Затем все индивидуальные значения возвращаются в свои первоначальные выборки, и подсчитывается суммы полученных ими рангов отдельно по каждой выборке.

Если различия между выборками случайны, суммы рангов не будут различаться сколько-нибудь существенно, так как высокие и низкие ранги равномерно распределяются между выборками.

Если в одной из выборок будут преобладать низкие значения рангов, в другой - высокие, а в третьей - средние, то **критерий H** позволит установить эти различия.

Гипотезы

H_0 : Между выборками 1, 2, 3 и т. д. существуют лишь случайные различия по уровню исследуемого признака.

H_1 : Между выборками 1, 2, 3 и т. д. существуют неслучайные различия по уровню исследуемого признака.

Ограничения критерия Н

1. При сопоставлении **3-х выборок** допускается, чтобы в одной из них $n=3$, а двух других $n=2$. Но при таких численных составах выборок мы сможем установить различия лишь на низшем уровне значимости ($p \leq 0,05$).

Для того, чтобы оказалось возможным диагностировать различия на более высоком уровне значимости ($p \leq 0,01$), необходимо, чтобы в каждой выборке было не менее 3 наблюдений, или чтобы по крайней мере в одной из них было 4 наблюдения, а в двух других - по 2; при этом неважно, в какой именно выборке сколько испытуемых, а важно соотношение 4:2:2.

2. Критические значения критерия Н и соответствующие им уровни значимости приведены в Табл. Приложения . Таблица предусмотрена только для трех выборок и $\{n_1, n_2, n_3\} \leq 5$.

При большем количестве выборок и испытуемых в каждой выборке необходимо пользоваться Таблицей критических значений критерия χ^2 , поскольку критерий Крускала-Уоллиса асимптотически приближается к распределению χ^2 .

Количество степеней свободы при этом определяется по формуле: $V=c-1$ где c - количество сопоставляемых выборок.

АЛГОРИТМ критерия Н Крускала-Уоллиса

1. Перенести все показатели испытуемых на индивидуальные карточки.
 2. Пометить карточки испытуемых группы 1 определенным цветом, например красным, карточки испытуемых группы 2 - синим, карточки испытуемых групп 3 и 4 - соответственно, зеленым и желтым цветом и т. д. (Можно использовать, естественно, и любые другие обозначения.)
 3. Разложить все карточки в единый ряд по степени нарастания признака, не считаясь с тем, к какой группе относятся карточки, как если бы мы работали с одной объединенной выборкой.
 4. Проранжировать значения на карточках, приписывая меньшему значению меньший ранг. Надписать на каждой карточке ее ранг. Общее количество рангов будет равняться количеству испытуемых в объединенной выборке.
 5. Вновь разложить карточки по группам, ориентируясь на цветные или другие принятые обозначения.
 6. Подсчитать суммы рангов отдельно по каждой группе. Проверить совпадение общей суммы рангов с расчетной.
-