

Статистические гипотезы и критерии. Этапы принятия статистического решения

Статистические гипотезы

Формулирование гипотез систематизирует предположения исследователя и представляет их в лаконичном виде.

Различают два вида гипотез: *научные и статистические*.

Научная гипотеза - это предполагаемое решение проблемы (формулируется как теорема).

Статистическая гипотеза – это предположение о распределении вероятностей, которое мы хотим проверить по имеющимся данным.

Статистические гипотезы

Статистические гипотезы подразделяются на:

- *нулевые,*
- *альтернативные,*
- *ненаправленные,*
- *направленные.*

Статистические гипотезы

- **Нулевая гипотеза** (обозначается как H_0) - это гипотеза об отсутствии значимых различий между какими-либо показателями или о случайности связи между ними.

Гипотеза называется нулевой, потому что содержит 0 ($X_1 - X_2 = 0$, X_1 и X_2 – сопоставляемые значения признаков).

- **Альтернативная гипотеза** - это гипотеза о значимости различий между этими показателями или о достоверности связи между ними и обозначается как H_1 .

Альтернативная гипотеза – это то, что мы хотим доказать, поэтому её часто называют **экспериментальной гипотезой**.

Статистические гипотезы

Например, необходимо выяснить, одинаков ли умственный уровень у учеников школ № 1 и № 2. Перед началом проведения исследования необходимо сформулировать соответствующие гипотезы. В нашем случае нулевая и альтернативная гипотезы будут выглядеть следующим образом.

H_0 : уровни умственного развития учеников школ № 1 и № 2 не отличаются друг от друга.

H_1 : уровни умственного уровня учеников школ № 1 и № 2 значимо различны.

Направленная и ненаправленная гипотезы

Гипотеза называется *ненаправленной*, если ее формулировка предполагает определение лишь отличий или не отличий X_1 и X_2 без указания направления этих отличий (в большую или меньшую сторону).

Если гипотеза содержит указание на направление отличий, например, X_1 не превышает X_2 (для гипотезы H_0) или X_1 превышает X_2 (для гипотезы H_1), то гипотеза рассматривается как *направленная*.

Примеры формулирования гипотез.

Направленные гипотезы

H_0 : x_1 не превышает x_2

H_1 : x_1 превышает x_2

Ненаправленные гипотезы

H_0 : x_1 не отличается от x_2

H_1 : x_1 отличается от x_2

Статистические критерии

Решение о наличии или отсутствии связей между выборками по тем или иным измеряемым признакам принимается в соответствии с определенным, так называемым, *статистическим критерием*.

Статистический критерий - это правило, обеспечивающее принятие выдвигаемой гипотезы как истинной и отклонение как ложной с высоким уровнем достоверности.

Статистические критерии

Статистические критерии делятся на:

- *параметрические,*
- *непараметрические,*
- *односторонние,*
- *двусторонние.*

К *параметрическим* критериям относятся критерии, в вычислении которых используются параметры распределения (средние значения и средне квадратичные отклонения).

К *непараметрическим* критериям относятся критерии, не требующие знание параметров распределения. Для их вычисления достаточно иметь значения частот или рангов.

Параметрические критерии

Преимуществом параметрических критериев является то, что они позволяют прямо оценить различия в средних или в дисперсиях сравниваемых выборок.

Однако, их применение является корректным лишь по отношению к статистическим совокупностям, имеющим нормальные или близкие к нормальным распределения.

Кроме того, процедура их вычисления в сравнении с непараметрическими критериями несколько более сложная.

Непараметрические критерии

Непараметрические критерии вычисляются проще, но они позволяют оценить лишь средние тенденции, например, ответить на вопрос, чаще ли в выборке А встречаются более высокие значения показателя, а выборке Б - более низкие.

Тем не менее, если непараметрические критерии показывают наличие различий - значит, эти различия действительно есть.

Однако если различия с помощью непараметрических методов не выявляются, то это еще не говорит о том, что их нет. Возможно, они есть, но для их выявления необходимо воспользоваться процедурой нормализации исходных распределений и соответствующими **более мощными параметрическими критериями.**

Понятие *одностороннего* либо *двустороннего* критерия связано с формулировкой гипотез.

Если "**нулевая**" гипотеза формулируется о равенстве ($X1 = X2$), то для проверки используется двусторонний критерий.

Если "**нулевая**" гипотеза формулируется о неравенстве, то возможны следующие варианты:

- если $X1 \neq X2$, то используется двусторонний критерий;
- если $X1 > X2$ или $X1 < X2$, то односторонний критерий.

Проверка гипотезы

Метод, который используется для принятия решения относительно справедливости статистической гипотезы, называется *проверкой гипотезы*.

Основной принцип проверки гипотезы состоит в том, что выдвигается нулевая гипотеза H_0 , с тем чтобы попытаться опровергнуть ее и тем самым подтвердить альтернативную гипотезу H_1 .

При проверке любой статистической гипотезы решение исследователя никогда не принимается с уверенностью, поскольку всегда остается риск принятия неправильного решения. Поэтому в психолого-педагогических исследованиях при формулировке выводов и принятии решений всегда указывается степень их достоверности (уровень значимости).

Принятие и отвержение гипотез

Для принятия или отвержения гипотез в процессе формулирования статистического вывода необходимо знать: 1) эмпирическое значение критерия, 2) его критические значения, 3) сам принцип принятия-отвержения гипотез.

Принцип принятия-отвержения гипотезы по А. Д. Наследову. Согласно ему, если эмпирическое значение критерия (коэффициента) превышает его критическое значение или равно ему, то отвергается H_0 и принимается H_1 . В виде символов реализацию этого принципа можно записать следующим образом:

$ЭМП \geq КР \Rightarrow H_1, !$ на соответствующем уровне значимости

$ЭМП < КР \Rightarrow H_0!$

Для данных правил есть **исключения**. В случае реализации любого из этих принципов при работе с Т-критерием Вилкоксона, U-критерием Манна-Уитни и G-критерием знаков устанавливается обратное соотношение знаков $<$ и $>$.

Статистическая значимость

Уровень значимости (уровень достоверности) результата исследования (так называемый ***p*-уровень**) – это вероятность того, что мы сочли различия существенные, а они на самом деле случайны.

Значение *p*-уровня находится в обратной зависимости от надежности результата. Уровень значимости показывает вероятность ошибки при принятии гипотезы.

Уровень значимости – это вероятность отклонения нулевой гипотезы, в то время как она верна.

Например, *p* -уровень = 0,01 показывает, что имеется 1%-ная вероятность, что найденная в выборке связь между переменными является случайной и характерна только для данной выборки.

Соотношение значимости и p - уровня

Традиционная интерпретация уровней значимости при $\alpha = 0,05$

Уровень значимости	Решение	Возможный статистический вывод
$p > 0,1$	Принимается H_0	«Статистически достоверные различия не обнаружены»
$p \leq 0,1$	сомнения в истинности H_0 , неопределенность	«Различия обнаружены на уровне статистической тенденции»
$p \leq 0,05$	значимость, отклонение H_0	«Обнаружены статистически достоверные (значимые) различия»
$p \leq 0,01$	высокая значимость, отклонение H_0	«Различия обнаружены на высоком уровне статистической значимости»



$p > 0,1$



$p \leq 0,1$



$p \leq 0,05$



$p \leq 0,01$



$p \leq 0,001$

Для каждого статистического критерия критические значения p - уровня можно взять из соответствующих специальных таблиц.

Знание значения p - уровня необходимо для подтверждения или опровержения статистических гипотез.

В психологии принято считать *низшим уровнем статистической значимости 5%-ый уровень* ($p \leq 0,05$), *достаточным - 1%-ый уровень* ($p \leq 0,01$) и *высшим 0,1%-ый уровень* значимости ($p \leq 0,001$).

Результат $p \leq 0,05$ большинством специалистов считается приемлемой границей статистической значимости.

Число степеней свободы

В большинстве случаев одно и тоже эмпирическое значение критерия может оказаться значимым или незначимым в зависимости от количества наблюдений в исследуемой выборке (n) или от так называемого количества степеней свободы.

Число степеней свободы (обозначается как df или ν) равно числу классов вариационного ряда минус число условий, при которых он был сформирован.

К числу таких условий относят, прежде всего, *объем выборки* (n).

Например, выборка состоит из 50 человек. Если в первом классе – 20 человек, во втором классе – 20 человек, то в третьем должны оказаться 10 человек.

Мы ограничены только одним условием – объемом выборки. Мы не свободны в определении количества испытуемых в третьем классе, "свобода" простирается только на первые два класса $\nu = n - 1 = 50 - 1 = 49$.

Число степеней свободы

Зная n и/или число степеней свободы, по специальным таблицам можно определить критические значения критерия и сопоставить с ними полученное эмпирическое значение.

В случае зависимых выборок число степеней свободы равно $n-1$.

Число степеней свободы для этих независимых выборок будет составлять $(n_1 + n_2) - 2$.

Схематическое представление психолого-педагогического исследования

Перед началом исследования требуется сравнить выборки 1 и 2, чтобы удостовериться, что испытуемые имеют одинаковый исходный уровень. Эта процедура называется **«оценка достоверности различий»**. Указанные группы 1 и 2 являются *независимыми выборками*.

На фазе заключительных срезов сравниваются показатели выборок 1 и 3, чтобы удостовериться, что был сдвиг каких-либо психолого-педагогических параметров под влиянием стимульного воздействия. Эта процедура называется **«оценка достоверности сдвига»**. Необходимо также убедиться в том, что сдвиг был вызван именно стимульным воздействием, а не влиянием другого неконтролируемого фактора. Для этого следует снова оценить достоверность сдвига, но уже в выборках 2 и 4.

Экспериментальная группа	Контрольная группа
1. Начальный срез	2. Начальный срез
<i>Стимульное воздействие есть</i>	<i>Стимульного воздействия нет</i>
3. Конечный срез	4. Конечный срез

Классификация критериев

	Параметрические критерии	Непараметрические критерии
Определение согласованности изменений (корреляция)	r_{xy} (коэффициент корреляции Пирсона)	r_s (коэффициент корреляции Спирмена) Ф(коэффициент сопряженности) Q (коэффициент ассоциации)
Сравнение распределений		χ^2 - критерий Пирсона
Оценка достоверности различий	t-критерий Стьюдента для независимых выборок	U-критерий Манна-Уитни (2 выборки испытуемых) H-критерий Краскела-Уоллеса (3 и более выборок испытуемых)
Оценка достоверности различий при повторных измерениях	t-критерий Стьюдента для зависимых выборок	T-критерий Вилкоксона G – критерий знаков χ^2 - критерий Фридмана
Анализ изменений признака Сравнение дисперсий	F-критерий Фишера 21	