

СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОВЕРКИ ГИПОТЕЗ

Основные определения

Генеральной совокупностью называется совокупность однородных (относительно некоторого признака) объектов, на которой **ставится** статистическая задача

Выборочной совокупностью (выборкой) называется совокупность случайно отобранных предметов из генеральной совокупности. Совокупность объектов, на которой **решается** статистическая задача

Объемом совокупности (выборочной или генеральной) называется число объектов этой совокупности ($n \ll N$)

Виды выборок

Определение. Выборки называются **независимыми (несвязными)**, если процедура эксперимента и полученные результаты измерения некоторого свойства у испытуемых одной выборки не оказывают влияния на особенности протекания этого же эксперимента и результаты измерения этого же свойства у испытуемых (респондентов) другой выборки.



Виды выборок

Определение. Выборки называется **зависимыми (связными)** если процедура эксперимента и полученные результаты измерения некоторого свойства, проведенные на одной выборке, оказывают влияние на другую.

Одна и та же группа испытуемых, на которой дважды проводилось обследование (пусть даже разных качеств, признаков, особенностей), по определению оказывается зависимой, или связной выборкой.

ШКАЛЫ

1. Номинальная шкала (шкала наименований)

- делит все объекты на группы по какому-либо признаку (различию).

Примеры шкалы наименований: пол, национальность, семейное положение, образование, здоровый – больной, левша – правша, тип темперамента, тип личности и т. п.



2. Шкала порядка (порядковая, ранговая, ординальная)

- предназначена для измерения (обозначения) степени различия какого-либо признака или свойства у разных объектов.

Пример.



Разновидности порядкового шкалирования (измерения):

- ранжирование (в ряд),
- группировка (ранжирование по группам),
- парное сравнение,
- метод рейтинга,
- метод полярных профилей.



3. Интервальная шкала (интервальное измерение)

– это такое присвоение чисел объектам, когда определено расстояние между объектами и предусмотрена общая для всех объектов постоянная единица измерения. Иначе говоря, в интервальной шкале вводится единица и масштаб измерения. Нулевая точка шкалы выбирается произвольно.

Пример.



4. Шкала отношений (пропорциональная шкала)

- отличается от интервальной только тем, что ее нулевая точка не произвольна, а указывает на полное отсутствие измеряемого свойства.

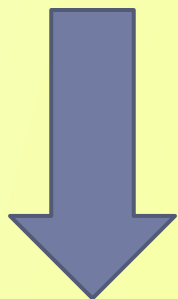
Примеры пропорциональных шкал:

расстояние, длина отрезков или физических объектов, время, температура по Кельвину (абсолютный нуль).

Возможны все арифметические действия.



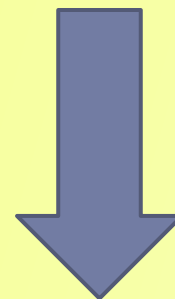
▣ **Определение . Статистическая гипотеза** – это любое предположение о свойствах случайных величин или событий.



H_0 (подлежит проверке): отсутствие связи в генеральной совокупности

нулева

я



H_1 : наличие связи в генеральной совокупности

альтернативн

ая

Проверка статистических

гипотез! Какие выборки, какие измерения?

1. Уровень статистической значимости

□ *Определение.* **Статистическая значимость (p-уровень значимости)** – количественная оценка надежности установленной связи.



Чем меньше значение p-уровня, тем выше статистическая значимость результата исследования, который подтверждает научную гипотезу

2. Статистический критерий и число степеней свободы

Определение . **Статистический критерий** – это инструмент для определения уровня статистической значимости гипотезы

Определение. **Число степеней свободы** – это количество возможных направлений изменчивости признака.

Зависит от:

- объема выборки,
- от числа изучаемых признаков

Определение. **Назначение критерия** – проверка статистической гипотезы путем определения p -уровня значимости



Структура статистического критерия

- формула расчета наблюдаемого (эмпирического) значения критерия по выборочным статистикам;
- правило определения числа степеней свободы;
- критическое значение критерия для данного числа степеней свободы;
- правило соотнесения наблюдаемого значения критерия с критическим для определения вероятности того, что H_0 верна.

Алгоритм проверки гипотезы

- Выбор критерия в зависимости от вида исходных данных и статистической гипотезы
- Расчет по исходным данным (или по имеющимся статистикам) наблюдаемого (эмпирического) значения критерия и числа степеней свободы
- Применение «Таблицы критических значений критерия» позволяет определить значение p -уровня для данного числа степеней свободы

Выбор метода статистической проверки гипотезы

Определяется:

- типом используемой шкалы
- количеством сравниваемых групп
- зависимостью или независимостью сравниваемых выборок

Классификация критериев

- ▣ **Непараметрические критерии статистики:** закон распределения выборок не известен , наблюдения независимые
- ▣ **Параметрические критерии :** вид распределения или функция распределения выборки нам заданы

Выбор критерия для проверки

Зависимость выборок		Независимые	Зависимые
Признак	метрический	Параметрические методы сравнения	
		t-Стьюдента, для независимых выборок	t-Стьюдента, для зависимых выборок
		Непараметрические методы сравнения	
	ранговый	Медианный критерий	Критерий знаков
	номинальный		Критерий Макнамара



Параметрические критерии

Критерий Стьюдента

! Распределение выборок – нормальное

Случай независимых выборок

n_1 - число элементов в первой выборке

n_2 - число элементов во второй выборке

Статистика критерия:

$$t_{\text{эмп}} = \frac{\bar{x} - \bar{y}}{\sigma_{x-y}}$$

□ $\bar{x}$ $\bar{y}$ - средние арифметические в экспериментальной и контрольной группах

$$\sigma_{x-y} = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

число степеней свободы: $k = n_1 + n_2 - 2$

Принятие решения:

1. Сравнить полученное значение $t_{\text{эмп}}$ с критическим значением t -распределения Стьюдента (приложение I).
2. Если $t_{\text{эмп}} < t_{\text{крит}}$, то гипотеза H_0 принимается, в противном случае нулевая гипотеза отвергается и принимается альтернативная гипотеза.

Пример:

Экспериментальная группа	Контрольная группа
12,13,9,8,12,10,11,9,9,6	11,8,10,12,15,12,10,5,4

Сформулировать нулевую и альтернативную гипотезы и проверить, какая верна

Случай связанных (парных) выборок

1. Вычислим: $d_i = y_i - x_i$

2. Вычислим: $d = \frac{\sum d_i}{n}$

3. Вычислим: $S_d = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n(n-1)}}$

Статистика критерия: $t_{\text{эмп.}} = \frac{d}{S_d}$

Число степеней свободы $k = n - 1$

Принятие решения. Если $t_{\text{эмп}} < t_{\text{крит}}$, то нулевая гипотеза принимается, в противном случае принимается альтернативная.

Пример.

До начала работы	11, 9, 8, 7, 8, 9, 12, 10, 11
После окончания	10, 10, 9, 8, 9, 8, 11, 12, 12

Данные тестирования до и после начала эксперимента. Сформулировать нулевую и альтернативную гипотезы и проверить, какая из них принимается



Непараметрические критерии

□ Критерий Макнамары

!!! Условия:

- выборки зависимые;
- для получения данных использовалась шкала наименований;
- выборки случайные;
- результаты измерений интересующего свойства не влияют друг на друга.

X - состояние некоторого свойства в рассматриваемой совокупности объектов при первичном измерении свойства.

Y — при вторичном измерении.

- $x_1, x_2, \dots, x_M$ — до применения средства;
- $y_1, y_2, \dots, y_N$ — после применения средства.

Классификация x_i	Классификация y_i		
		$y_i = 0$	$y_i = 1$
	$x_i = 0$	a	b
	$x_i = 1$	c	d

Гипотезы: $H_0: P(x_i = 0, y_i = 1) = P(x_i = 1, y_i = 0).$

$H_1: P(x_i = 0, y_i = 1) \neq P(x_i = 1, y_i = 0).$



Статистика критерия и принятие решения:

□ Если $n=b+c > 20$, то $t_{\text{эмп}} = \frac{(b-c)^2}{b+c}$

Если $t_{\text{эмп}} < 3.84$, то H_0 – принимается и альтернативная гипотеза отклоняется.

□ Если $n=b+c \leq 20$, то вычисляем $T2 = \min(b, c)$. По n и $T2$ находим $t_{\text{эмп}}$ по таблице на стр. III и сопоставляем это значение с $t_{\text{кр.}} = 0.025$.

Если $t_{\text{эмп}} < T_{\text{кр.}}$, то нулевая гипотеза отклоняется.

а) Если $b < c$, то принимается гипотеза $H1: P(x_i=0, y_i=1) < P(x_i=1, y_i=0)$.

б) Если $b > c$, то принимается гипотеза $H1: P(x_i=0, y_i=1) > P(x_i=1, y_i=0)$.

Пример 1.

- Проверка эффективности форм адапционной работы среди учащихся первых классов. Эффективность каждого цикла работы выяснялась по отношению учащихся к обучению (ответы: да, нравится/ нет, не нравится). Из 150 человек отобрана выборка – 20 учащихся. Используется критерий Макнамары.

<i>1 опрос</i>	<i>2 опрос</i>	
	<i>Нрав.</i>	<i>Не нрав.</i>
<i>Нрав.</i>	$a=3$	$b=8$
<i>Не нрав.</i>	$c=3$	$d=6$

Задание: сформулировать гипотезы и проверить, какая принимается

Пример 2.

- В эксперименте 100 испытуемых. Проверяется наличие (отсутствие) качества в ходе двух последовательных срезов, между которыми - работа по формированию этого качества.

<i>1 срез</i>	<i>2 срез</i>		
	<i>наличие</i>	<i>отсутствие</i>	
<i>наличие</i>	$a=20$	$b=15$	$\Sigma=35$
<i>отсутствие</i>	$c=35$	$d=30$	$\Sigma=65$
	$\Sigma=55$	$\Sigma=45$	

Задание: сформулировать гипотезы и проверить, какая принимается

Критерий знаков (G-критерий)

□ **Условия:**

- зависимые выборки
- шкала не ниже ранговой

Наблюдения над случайными переменными X и Y , полученные при рассмотрении двух зависимых выборок

Н₀: в состоянии изучаемого свойства нет значимых различий при первичном и вторичном измерениях

Н₁: законы распределения величин X и Y различны

▣ **Статистика критерия:** $t_{\text{эмп}}$ – количество пар со знаком плюс, n – количество пар без нулей

! Нулевая гипотеза принимается на уровне значимости 0,025, если $t_{\text{эмп}} < n - t_a$ (см. приложение на стр. 112)

Пример: новый метод, два среза

Учащиеся (№)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Первое выполнение	2	2	2	2	2	3	3	3	3	3	2	2	3	3	3
Второе выполнение	2	3	3	4	3	2	3	4	4	3	4	3	2	4	4
Знак разности отметок	0	+	+	+	+	-	0	+	+	0	+	+	-	+	+

Задание: сформулировать гипотезы и проверить, какая принимается



Медианный критерий

УСЛОВИЯ:

- выборки независимые;
 - выборки случайные;
 - шкала измерений не менее порядковой;
 - Совокупный объем выборок не менее 40.
 - $x_1, x_2 \dots x_{N1}$ — состояния изучаемого свойства в первой совокупности; $y_1, y_2 \dots y_{N2}$ — состояния изучаемого свойства во второй совокупности.
- $N1$ - объем первой выборки; $N2$ - объем второй выборки;
 MeX – медиана совместной выборки.

Систематизация:

	Выборка 1	Выборка 2
$> \text{Me}X$	A	B
$\leq \text{Me}X$	C	D

Важно! Критерий можно применять, если $A, B, C, D \geq 5$.

H_0 : обе совокупности имеют одинаковую медиану (нет значимых различий между выборками);

H_1 : совокупности имеют различные медианы (существуют значимые различия между выборками).

Статистика критерия:

$$t_{\text{эмп}} = \frac{N(|AD - BC|)}{(A+B)(C+D)(A+C)(B+D)}$$

Принятие решения: H_0 принимается, если $t_{\text{эмп}} < 3.84$.

Важно!

Если наблюдается достаточно большое количество значений выборок вблизи медианы, то следует использовать уточненное значение медианы.

Пример.

- Для сопоставления уровня развития визуальной памяти в двух классах были проведены замеры по шкале рангов. Результаты представлены в таблице:

Контр. класс	3	4	5	4	4	6	6	4	5	7	4	4	5	5	3
Экспер. класс	6	4	5	4	6	3	5	5	6	7	7	4	7	5	4
Контр. класс	4	2	7	6	5	5	5	4	6	3	4	5	4	4	6
Экспер. класс	7	3	5	7	5	5	6	4	6	7	4	4	3	5	6

Являются ли различия в уровне развития визуальной памяти у учащихся двух этих классов статистически значимыми?

