



## Проверка гипотез о равенстве средних, о равенстве дисперсий, о равенстве вероятностей, о значимости выборочного коэффициента корреляции, о независимости событий.

### Равенство средних при известных дисперсиях

В случае, когда дисперсии известны, для проверки гипотезы о равенстве разности средних некоторому значению применяется статистика:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

где

$\bar{x}_1$  и  $\bar{x}_2$  - выборочные средние,

$\mu_1$  и  $\mu_2$  - гипотетические генеральные средние,

$n_1$  и  $n_2$  - объемы выборок,

$\sigma_1^2$  и  $\sigma_2^2$  - известные генеральные дисперсии.

### Дисперсии неизвестны, но равны

#### Теорема

В случае, когда дисперсии неизвестны, но равны, для проверки гипотезы применяется статистика:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

где

$\bar{x}_1$  и  $\bar{x}_2$  - выборочные средние,

$\mu_1$  и  $\mu_2$  - гипотетические генеральные средние,

$n_1$  и  $n_2$  - объемы выборок,

$s_p^2$  - объединённая оценка дисперсии.

Вычисляется объединённая оценка дисперсии по формуле:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

где  $s_1^2$  и  $s_2^2$  - выборочные дисперсии.

### Дисперсии неизвестны и не предполагаются равными

- В самом общем случае, когда дисперсии неизвестны и не равны, точный критерий для проверки гипотезы о равенстве средних указать трудно. В этом случае пользуются приближительными формулами.
- Как и следовало ожидать для проверки гипотезы применяется  $t$ -статистика, в которой вместо теоретических значений дисперсий стоят выборочные оценки, то есть статистика критерия имеет вид

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

**Теорема**

Для проверки гипотезы о равенстве дисперсий  $H_0: \sigma_1^2 = \sigma_2^2$  используется статистика  $F = \frac{s_1^2}{s_2^2}$ , которая имеет распределение Фишера с числом степеней свободы числителя  $n_1 - 1$  и знаменателя  $n_2 - 1$ .

Здесь  $\sigma_1^2$  и  $\sigma_2^2$  - дисперсии генеральных совокупностей,  $s_1^2$  и  $s_2^2$  - выборочные дисперсии,  $n_1$  и  $n_2$  - объемы выборок. Без ограничения общности считаем, что  $s_1^2 > s_2^2$ .

**Проверка гипотез о равенстве вероятностей**

В качестве критерия для сравнения двух вероятностей биномиальных распределений используется случайная величина  $U$ , наблюдаемое значение найдем по формуле:

$$u_{\text{набл}} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{p \cdot q}{n_1} + \frac{p \cdot q}{n_2}}}$$

где  $m_1 / n_1$  - относительная частота (частость) появления события в первой выборке;

$m_2 / n_2$  - относительная частота (частость) появления события во второй

$q = 1 - p$  - средняя частость непоявления события;

$n_1$  - объем первой выборки;

$n_2$  - объем второй выборки.

Так как конкурирующая гипотеза - двусторонняя, критическое значение (укр.) следует находить по таблице функции Лапласа из равенства:

$$\Phi(u_{кр}) = \frac{1 - \alpha}{2}$$

Если  $|U_{\text{набл}}| < U_{\text{кр}}$  то, точка не попадает в критическую область, можно принять нулевую гипотезу о равенстве параметра в первой выборке и второй.

**Проверка гипотез о значимости выборочного коэффициента корреляции**

Пусть двумерная генеральная совокупность  $(X, Y)$  распределена нормально. Из этой совокупности извлечена выборка объема  $n$  и по ней найден выборочный коэффициент корреляции  $r_B$ , который оказался отличным от нуля. Так как выборка отобрана случайно, то еще нельзя заключить, что коэффициент корреляции генеральной совокупности  $R_r$  также отличен от нуля. В конечном счете нас интересует именно этот коэффициент, поэтому возникает необходимость при заданном уровне значимости проверить нулевую гипотезу  $H_0: R_r = 0$  о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе  $H_1: R_r \neq 0$ .

Если нулевая гипотеза отвергается, то это означает, что выборочный коэффициент корреляции значимо отличается от нуля (кратко говоря, значим), а  $X$  и  $Y$  коррелированы, т. е. связаны линейной зависимостью.

Если нулевая гипотеза будет принята, то выборочный коэффициент корреляции незначим, а  $X$  и  $Y$  некоррелированы, т. е. не связаны линейной зависимостью.

В качестве критерия проверки нулевой гипотезы примем случайную величину

$$T = r_B \frac{\sqrt{n-2}}{\sqrt{1-r_B^2}}.$$

Величина при справедливости нулевой гипотезы имеет распределение Стюдента с  $k = n - 2$  степенями свободы. Поскольку конкурирующая гипотеза имеет вид критическая область - двусторонняя.

Обозначим значение критерия, вычисленное по данным наблюдений, через  $T_{\text{набл}}$  и сформулируем правил проверки нулевой гипотезы.

**Правило.** Для того чтобы при заданном уровне значимости проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе, надо вычислить наблюдаемое значение критерия:

$$T_{\text{набл}} = r_B \frac{\sqrt{n-2}}{\sqrt{1-r_B^2}}$$

и по таблице критических точек распределения Стюдента по заданному уровню значимости и числу степеней свободы найти критическую точку  $t_{кр}(a, k)$  для двусторонней критической области.

Если  $|T_{набл}| < t_{кр}$  – нет оснований отвергнуть нулевую гипотезу.

Если  $|T_{набл}| > t_{кр}$  – нулевую гипотезу отвергают.

### Проверка гипотезы о независимости двух выборок

Для проверки гипотезы о независимости двух выборок  $X$  и  $Y$  используется таблица частот, которые можно было бы ожидать в том случае, если переменные оказались бы независимыми. В общем случае критерий  $\chi^2$  независимости принято применять следующим образом:

1. Множество значений выборки  $X$  разбивается на  $s$  непересекающихся классов (интервалов, разрядов), а множество значений  $Y$  на  $r$  непересекающихся интервалов.
2. Составляется таблица исходных данных в виде списка экспериментальных частот  $Z_{ij}$  всех комбинаций категорий двух качественных переменных. Таблица дополняется строкой и столбцом сумм "попаданий" компонент выборки в конкретный класс другой выборки:

$$Z_{Yk} = \sum_i Z_{ik}, k = 1, \dots, r \quad Z_{Xk} = \sum_j Z_{kj} k = 1, \dots, s$$

Переменные (выборки), разделенные по классам (экспериментальные частоты)					
интервалы $i=1,2,\dots,s$ выборки $X$	интервалы $j=1,2,\dots,r$ выборки $Y$				
	элементы $Y_1$	элементы $Y_2$	***	элементы $Y_r$	
элементы $X_1$	$Z_{11}$	$Z_{12}$	***	$Z_{1r}$	$Z_{X1} = \sum_j Z_{1j}$
элементы $X_2$	$Z_{21}$	$Z_{22}$	***	$Z_{2r}$	$Z_{X2} = \sum_j Z_{2j}$
***	***	***	***	***	***
элементы $X_s$	$Z_{s1}$	$Z_{s2}$	***	$Z_{sr}$	$Z_{Xs} = \sum_j Z_{sj}$
	$Z_{Y1} = \sum_i Z_{i1}$	$Z_{Y2} = \sum_i Z_{i2}$	***	$Z_{Yr} = \sum_i Z_{ir}$	$n = \sum_i \sum_j Z_{ij}$

3. Составляется таблица "относительных частот" по соотношению

$$T_{ij} = \frac{Z_{ij}}{Z_{Yj} Z_{Xi}}, i = 1, \dots, s, j = 1, \dots, r.$$

Переменные (выборки), разделенные по классам				
	элементы $Y_1$	элементы $Y_2$	***	элементы $Y_r$
элементы $X_1$	$T_{11}$	$T_{12}$	***	$T_{1r}$
элементы $X_2$	$T_{21}$	$T_{22}$	***	$T_{2r}$
***	***	***	***	***
элементы $X_s$	$T_{s1}$	$T_{s2}$	***	$T_{sr}$

4. Определяется расчетное значение  $\chi^2$  по формуле

$$\chi^2_{расч} = n \left( \sum_i \sum_j T_{ij} - 1 \right).$$

5. Критическое значение  $\chi^2_{кр}$  ( $s$  категорий  $X$  и  $r$  категорий  $Y$ ) определяется для числа степеней свободы  $df = (s-1)(r-1)$
6. Проводится сравнение расчетного значения  $\chi^2_{расч}$  с критическим  $\chi^2_{кр}$ , определенному по "обычным" уровням значимости, равному 0,05 или 0,01 или другому выбранному значению.
7. Строится заключение: при  $\chi^2_{расч} > \chi^2_{кр}$  гипотеза об отсутствии связи между признаками и параметрами отвергается, при  $\chi^2_{расч} < \chi^2_{кр}$  – подтверждается.

<https://mse.msu.ru/wp-content/uploads/2020/03/Лекция-7-Две-выборки-дополнительный-материал.pdf?ysclid=l4sbatmb7r182706342>

<http://math-info.hse.ru/fj/2017-18/ps-ms/hypo-test.pdf>

[https://www.matburo.ru/Examples/Files/ms\\_pg\\_14.pdf?ysclid=l4sbtrp1te402341399](https://www.matburo.ru/Examples/Files/ms_pg_14.pdf?ysclid=l4sbtrp1te402341399)

<https://helpiks.org/7-18188.html>

<https://studfile.net/preview/9099246/page:20/>

[http://koi.tspu.ru/biostat/Appendix\\_5.pdf](http://koi.tspu.ru/biostat/Appendix_5.pdf)

