

# Хи-квадрат критерий (Chi-square test)

Синонимы: Критерий согласия Пирсона

Разделы: [Метрики](#)

LogInom: [Отчет по регрессии](#)

Критерий согласия для проверки гипотезы о законе распределения исследуемой случайной величины. Во многих практических задачах точный закон распределения неизвестен. Поэтому выдвигается гипотеза о соответствии имеющегося эмпирического закона, построенного по наблюдениям, некоторому теоретическому. Данная гипотеза требует статистической проверки, по результатам которой будет либо подтверждена, либо опровергнута.

Пусть  $X$  — исследуемая случайная величина. Требуется проверить гипотезу  $H_0$  о том, что данная случайная величина подчиняется закону распределения  $F(x)$ . Для этого необходимо произвести выборку из  $n$  независимых наблюдений над случайной величиной  $X$ :  $X^n = (x_1, \dots, x_n)$ ,  $x_i \in [a, b]$ ,  $\forall i = 1 \dots n$ , и по ней построить эмпирический закон распределения  $F'(x)$  случайной величины  $X$ .

Гипотеза  $H'_0$ :  $X^n$  порождается функцией  $F'(x)$ .

Разделим  $[a, b]$  на  $k$  непересекающихся интервалов  $[a_i, b_i]$ ,  $i = 1 \dots k$ .

Пусть  $n_j$  — количество наблюдений в  $j$ -м интервале;

$p_j = F(b_j) - F(a_j)$  — вероятность попадания наблюдения в  $j$ -й интервал при выполнении гипотезы  $H'_0$ ;

$E_j = np_j$  — ожидаемое число попаданий в  $j$ -й интервал;

тогда распределение Хи-квадрат с числом степеней свободы  $k - 1$  будет иметь следующую статистику:

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - E_j)^2}{E_j} \sim \chi_{k-1}^2$$

В зависимости от значения критерия  $\chi^2$  гипотеза  $H_0$  может приниматься либо отвергаться:

- $\chi_1^2 < \chi^2 < \chi_2^2$  — гипотеза  $H_0$  выполняется.
- $\chi^2 \leq \chi_1^2$  — попадает в левый «хвост» распределения. Следовательно, теоретические и практические значения очень близки. Если, к примеру, происходит проверка генератора случайных чисел, который сгенерировал  $n$  чисел из отрезка  $[0,1]$  и выборка  $X^n$  распределена равномерно на  $[0,1]$ , то генератор нельзя называть

случайным (гипотеза случайности не выполняется), т.к. выборка распределена слишком равномерно, но гипотеза  $H_0$  выполняется.

- $\chi^2 \geq \chi_2^2$  — попадает в правый «хвост» распределения, гипотеза  $H_0$  отвергается.