

билет 28

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:

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

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

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

- 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  $\chi^2_{расч} \leq \chi^2_{кр}$  – подтверждается.

