

Пример 1.7. На сайте В. Леонова* приводится интерактивный пример анализа сопряженности двух качественных признаков 2×2 . А именно, рассматривается вопрос: "Можно ли утверждать, исходя из данных конкретной выборки, что два исследуемых дискретных качественных признака независимы друг от друга в генеральной совокупности?" Иными словами, определяется то, что *между этими признаками отсутствует взаимосвязь*. Если эта гипотеза будет отвергнута, то с высокой долей вероятности можно утверждать, что такая зависимость существует. Реальным содержанием примера является исследование наличия взаимосвязи между приемом контрацептивных таблеток матерями, и желтухой у детей получающих грудное вскармливание (данные в таблице).

Данные исследования

Прием матерью таблеток	b1. Есть желтуха	b2. Нет желтухи	Всего
a1. Принимала таблетки	33	24	57
a2. Не принимала таблетки	14	45	59
Всего	47	69	116

* <http://www.biometrika.tomsk.ru/freq1.htm>

В этом примере у 33 матерей принимавших таблетки дети болели желтухой, а у 24 матерей также принимавших таблетки дети не болели желтухой. Далее, у 14 матерей, которые не принимали таблетки, дети болели желтухой и у 45 матерей, не принимавших таблетки, дети не болели желтухой. Объем выборки для данного примера равен 116. Комбинации этих категорий между собой можно обозначить как $a1b1 (Z_{11}=33)$, $a1b2 (Z_{12}=24)$, $a2b1 (Z_{21}=14)$ и $a2b2 (Z_{22}=45)$.

По приведенной (без поправки Yates'a) в п.3 параграфа 1.2.4) формуле

$$\chi^2_{\text{расч}} = \frac{1}{n_1 n_2} \sum_{j=1}^M \frac{(n_1 Z_{2j} - n_2 Z_{1j})^2}{Z_{1j} + Z_{2j}}$$

для случая $M=2$ значение критерия можно рассчитать (см. также [рис. А](#)) так

$$\chi^2_{\text{расч}} = \frac{1}{n_1 n_2} \left[\frac{(n_1 Z_{21} - n_2 Z_{11})^2}{Z_{11} + Z_{21}} + \frac{(n_1 Z_{22} - n_2 Z_{12})^2}{Z_{12} + Z_{22}} \right] \approx 14,042.$$

Представленный на [рис. А](#) скриншот решения данной задачи содержит следующие этапы вычислений.

1. В ячейки A6:A8, B5:C6, A10:A14 заносятся пояснительные надписи, в ячейку D10 заносится величина уровня значимости, диапазон B7:C8 – исходные данные.
2. В ячейки B10 и B11 вводятся формулы подсчета сумм $n_1 = a1b1 + a1b2 (=B7+C7)$ – числа женщин, принимающих таблетки, и $n_2 = a2b1 + a2b2 (=B8+C8)$ – женщин, обходящихся без таблеток.

	A	B	C	D	E
1	Гипотеза: взаимосвязи между приемом				
2	контрацептивных таблеток матерями и забо-				
3	леванием желтухой желтухой у детей НЕТ				
4					
5		ж е л т у х а			
6	таблетки	есть	нет		
7	принимала	33	24		
8	не приним	14	45		
9					
10	n1=	57	$\alpha =$	0,005	
11	n2=	59			
12					
13	$\chi^2 =$	14,042	$= ((B10*B8-B11*B7)^2 / ((B7+B8)+(B10*C8-B11*C7)^2 / ((C7+C8)) / B10 / B11$		
14	$\chi^2_{кр} =$	7,879	$= \text{ХИ2ОБР}(D10;1)$		
15					
16					
17	Поскольку $\chi^2 > \chi^2_{кр}$, то гипотеза о независимости				
18	между заболеванием желтухой и приемом				
19	контрацептивных таблеток отвергается				
20					
21	$\alpha_{реальн} =$	1,79E-04	$= \text{ХИ2РАСП}(B13;1)$		
22					

Рис. А. Скриншот расчетного листа MS Excel к примеру 1.7

3. В ячейку B13 вводится формула $=((B10*B8-B11*B7)^2 / ((B7+B8)+(B10*C8-B11*C7)^2 / ((C7+C8)) / B10 / B11$ вычисления величины $\chi^2_{расч}$.

4. В ячейке B14 формулой $=\text{ХИ2ОБР}(D10;1)$ { $\text{CHIINV}(D10; 1)$ } определяется критическое значение $\chi^2_{кр}$.

5. Из сравнения $\chi^2_{расч}$ с $\chi^2_{кр}$ для $\alpha = 0,005$ видно, что вычисленное χ^2 превосходит критическое. Другими словами – выборки статистически различны и поэтому гипотеза о независимости между заболеванием желтухой и приемом контрацептивных таблеток отвергается при уровне значимости $\alpha < 0,005$, то есть

какая-то зависимость между заболеванием и приемом таблеток существует.

Можно отметить, что полученному значению $\chi_{\text{расч}}^2 = 14,042$ при числе степеней свободы равном $df = 1$ будет отвечать достигнутый уровень значимости $\alpha = 0,000179$ ("реальное" значение α в ячейке B21 определяется функцией =ХИ2РАСП (B13;1) {=CHIDIST(B13;1)} – см. [рис. А](#)).

Отметим также, что данную таблицу можно использовать в первом приближении и для проверки нулевой гипотезы о равенстве частот в двух генеральных совокупностях. Применительно к содержанию задачи можно говорить о *проверке гипотезы о равенстве частоты заболевания желтухой в популяциях, принимающих пероральные контрацептивы, и не принимающих таковых.*

