

Исследование взаимосвязей при помощи критерия хи-квадрат

Поиск взаимосвязей в данных

Статистический критерий — правило, по которому можно сделать выбор между гипотезами.

Статистические гипотезы:

- Основная (нулевая). Признаки не связаны.
- Конкурирующая (альтернативная). Признаки связаны.

Последовательность действий при проверке гипотез

1. Формулируем исследовательский вопрос.
2. Формулируем статистические гипотезы для его проверки: нулевую (основную) и альтернативную (конкурирующую).
3. Выбираем статистический критерий для проверки статистических гипотез с учётом:
 - типа гипотезы;
 - требований к критерию;
 - типа данных/шкалы.
- 4.1. Определяем критическое значение статистики критерия.
- 5.1. Рассчитываем выборочное значение статистики критерия и сравниваем с критическим.

или

- 4.2. Устанавливаем критический уровень значимости / вероятность ошибки первого рода.
 - 5.2. Рассчитываем по выборке статистику критерия и достигнутый уровень значимости.
 - Сравниваем достигнутый уровень значимости с установленной вероятностью.

Критерий хи-квадрат

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(E_{ij} - O_{ij})^2}{E_{ij}}$$

E_{ij} – ожидаемая частота (в условии независимости признаков)

O_{ij} – наблюдаемая частота (полученная на выборке)

Количество строк (R) и столбцов (C) таблицы

Ожидаемая частота

$$E_{ij} = \frac{(M_i \times M_j)}{N}$$

Количество степеней свободы (df)

$$df = (r - 1) \times (c - 1)$$

r — количество строк (rows)

c — количество столбцов (columns)

Критическое значение критерия

Число степеней свободы	Уровни значимости		
	$p = 0,05$	$p = 0,01$	$p = 0,001$
1	3,84	6,63	10,83
2	5,99	9,21	13,82
3	7,81	11,07	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,51
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88