

Математика и статистика, часть 2**Взаимосвязь между качественными признаками. Таблицы сопряженности. Критерий хи-квадрат. (08.06.2020)***А. А. Макаров, А. А. Тамбовцева, Н. А. Василёнок**Задача 1 рассматривалась в видео к семинару.*

Задача 1. В таблице приведены результаты исследования взаимосвязи между гендером и тем, является ли человек правой или левой рукой. Используя имеющиеся данные, проверьте гипотезу об отсутствии связи между двумя признаками на 5% уровне значимости, используя критерий хи-квадрат.

Гендер / Рука	Правша	Левша	
Мужчина	28	12	40
Женщина	47	13	60
	75	25	100

Задача 2. Для распределения χ^2 с одной степенью свободы рассчитайте следующие вероятности:

- (a) $P(\chi^2 > 1.69)$;
- (b) $P(\chi^2 > 3.13)$;
- (c) $P(\chi^2 > 5)$.

Задача 3. Ниже приведены данные исследования, посвященного маркетинговой эффективности рекламной акции «экономичная упаковка». В ходе исследования были опрошены 100 респондентов. Фиксировался пол респондента и наличие или отсутствие эффекта на него рекламной акции.

Гендер / Эффект	Нет	Есть	
Мужчины		15	50
Женщины			50
	40	60	100

- (a) Сформулируйте нулевую гипотезу для проверки взаимосвязанности двух качественных признаков.
- (b) Заполните пропущенные ячейки таблицы. Вычислите ожидаемые значения частот таблицы в предположении верной нулевой гипотезы.
- (c) Рассчитайте значение статистики критерия хи-квадрат.
- (d) Рассчитайте значение p -value для статистики критерия. Сделайте вывод относительно нулевой гипотезы.

Задача 4. Откройте файл `titanic.csv` в R. В файле содержится информация о пассажирах «Титаника». Мы знаем имя, пол, возраст, класс путешествия, порт посадки, а также выжил ли пассажир в результате крушения корабля. Давайте попробуем выяснить, существует ли связь между классом путешествия (`Pclass`) и шансами на спасение (`Survived`).

- (a) Сформулируйте нулевую и альтернативную гипотезы.
- (b) Постройте таблицу сопряженности для двух признаков с помощью команды `table()`. Сколько степеней свободы будет иметь статистика хи-квадрат?
- (c) Рассчитайте статистику критерия с помощью команды `chisq.test()`. Какой вывод можно сделать о нулевой гипотезе?
- (d) Выведите на экран ожидаемые частоты и стандартизированные остатки. Что можно предположить о характере связи между признаками?