

Лабораторная работа 7.2.

Корреляционный анализ категоризованных данных.

Выполняется в R.

Задание.

Для двух признаков A и B , каждый из которых имеет два возможных уровня, создать самостоятельно две таблицы сопряженности вида

Уровни признака	Уровни признака B		
	B	\bar{B}	Σ
A	n_{11}	n_{12}	$n_{1\bullet}$
\bar{A}	n_{21}	n_{22}	$n_{2\bullet}$
Σ	$n_{\bullet 1}$	$n_{\bullet 2}$	n

определив соответствующие вероятности и частоты так, чтобы в одном случае признаки были зависимы, во втором – независимы.

Для этого необходимо задать количество наблюдений n и вероятности $P(A) = p$, $P(B|A) = q_1$, $P(B|\bar{A}) = q_2$.

Если признаки A и B независимы, $P(B|A) = P(B|\bar{A})$, то есть $q_1 = q_2 = q$.

Сначала разбиваем всю выборку пропорционально вероятностям p и $(1 - p)$.

Для это используем функцию

$$n_1 = \text{rbinom}(1, n, p); n_2 = n - n_1$$

Получаем значения n_1 и n_2 . Записываем их в таблицу.

Далее, если признаки независимы и $q_1 = q_2 = q$, разбиваем значения n_1 и n_2 пропорционально вероятностям q и $(1 - q)$. Получаем соответственно значения n_{11} , n_{12} и n_{21} , n_{22} .

Если признаки зависимы, разбиваем значение n_1 пропорционально вероятностям q_1 и $(1 - q_1)$, получаем соответственно значения n_{11} , n_{12} .

Затем значение n_2 – пропорционально вероятностям q_2 и $(1 - q_2)$, получаем соответственно значения n_{21} , n_{22} .

Заполняем таблицу.

Для полученных таблиц проверяем сопряженность признаков по критерию квадратичной сопряженности и критерию Крамера.

Для нахождения коэффициента Крамера в R использовать функцию `cramerV` пакета `rcompanion`.

Построить доверительный интервал для данного коэффициента, задав в функции `cramerV` аргумент `ci=TRUE`.