Прикладная статистика. Лабораторная 1.

Алексей Григорьев, 674 гр. ФУПМ МФТИ

Март 2019

№4.1 (a) Сформулируем задачу о проверке гипотезы независимости связанных выборок в терминах корреляции. Из независимости следует некоррелируемость признаков, учитывая это получим:

$$\begin{aligned} \text{нулевая гипотеза} \quad & \rho_{X_1X_2} = 0, \\ \text{альтернатива} \quad & \rho_{X_1X_2} \neq 0, \end{aligned}$$

где ρ – некоторый выборочный коэффициент корреляции.

(b) Допустим, что рассматривается коэффициент корреляции Спирмена, не требующий при дальнейшем рассмотрении нормальности данных. Тогда статистику возьмем согласно критерию Стьюдента:

$$T = \frac{\rho_{X_1X_2}\sqrt{n-2}}{\sqrt{1-\rho^2_{X_1X_2}}}$$

(c) Нулевое распределение данной статистики при $n \to \infty$ имеет вид $St(n-2)$. Таким образом, статистика приближенно подчиняется распредлению Стьюдента с $n-2$ степенями свободы.

(d) Нулевая гипотеза откланяется при попадании реализации статистики в критическую область. Критическая область соответствует наименее вероятным значениям в условиях истинности гипотезы и имеет следующий вид:

$$\mathcal{U} = (-\infty, t_{\alpha/2}] \cup [-t_{\alpha/2}, +\infty),$$

где $t_{\alpha/2}$ – $\frac{\alpha}{2}$-квантиль распределения $St(n-2)$.

(e) На этапе предобработки данных исключаем из выборки выбросы, которые могут повлиять на вычисление коэффициента корреляции. Исключаются объекты, хотя бы один признак которых является ввыбросом. С помощью "ящика с усами"были исключены из выборки следующие страны: Китай, Италия, Корея, Иран. Для остальной части выборки вычисляется предложенная статистика:

$$T(X_1^n, X_2^n) \approx -0.142, \text{ p-val} \approx 0.889,$$

для уровня значимости $\alpha = 0.05$ гипотеза не отвергается (формально в терминах критической области получили, что реализация статистики в нее не попала).