

Лабораторная работа 1, задача 4.1

1. Из независимости следует признаков следует их некоррелированность, поэтому можем рассмотреть следующие гипотезы:

$$H_0 : \hat{r}_{X_1 X_2} = 0$$

$$H_1 : \hat{r}_{X_1 X_2} \neq 0,$$

где \hat{r} — корреляция Спирмена

2. Возьмём статистику из критерия Стьюдента:

$$T(X_1^n, X_2^n) = \frac{\hat{r}_{X_1 X_2} \sqrt{n-2}}{\sqrt{1 - \hat{r}_{X_1 X_2}^2}}$$

3. В условиях истинности нулевой гипотезы:

$$T(X_1^n, X_2^n) \rightarrow \text{St}(n-2), n \rightarrow \infty$$

4. Критическая область будет иметь вид:

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

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

При попадании реализации статистики в критическую область нулевая гипотеза отклоняется, иначе данные гипотезе не противоречат.

5. Очевидно, что в выборке есть выбросы в лице Китая, Италии, Кореи, Ирана. Это страны, в которых раньше других распространилась эпидемия

Без них $\hat{r}_{X_1 X_2} \approx -0.032$, $T(X_1^n, X_2^n) \approx -0.159$, данные гипотезе не противоречат.

С ними $\hat{r}_{X_1 X_2} \approx 0.375$, $(X_1^n, X_2^n) \approx 2.622$, гипотеза отвергается.

Отметим также, что даже во втором случае встроенный критерий `scipy.stats` всё равно не отвергает гипотезу. Логично предположить, что в общем случае число выздоровевших не зависит от числа заболевших.