22 марта 2020 г. Андреев Павел, 626 группа

Задача 2.4

1

Выборки:	$X_1^{15} = (X_{1,1}, .., X_{1,15})$
	$X_2^{15} = (X_{2,1}, .., X_{2,15})$
	выборки связанны
Гипотеза:	H_0 : Средние выборок равны
	$\mathrm{med}(X_1) = \mathrm{med}(X_2)$
	($\mu_1 = \mu_2$, зависит от критерия)
Альтернатива:	H_1 : Средние выборок отличаются
	$\mathrm{med}(X_1) \neq \mathrm{med}(X_2)$
	$(\mu_1 \neq \mu_2)$

2

Рассмотрим два критерия: t-критерий Стъюдента для связанных выборок и критерий знаковых рангов Уилкоксона для связанных выборок.

t-критерий является параметрическим, он предполагает нормальность исследуемых выборок, следовательно для его применения необходимо убедиться в том, что данные не слишком отличаются от нормальных. t-критерий устойчив к небольшим отклонениям от нормальности, следовательно, если бы ку-ку график получился не очень скошеным, то можно было бы его применять. Альтернативно, можно было бы проверить гипотезу о нормальности выборки, например с помощью критерия Шапиро-Уилка.

Критерий знаковых рангов Уилкоксона, напротив, является непараметричексим и не предпологает нормальности данных, он предполагает лишь симметричность функции распределения относительно медианы.

t-критерий проверяет гипотезу о равенстве матожиданий, критерий же знаковых рангов Уилкоксона проверяет гипотезу о равенстве медиан. Хотя для использования t-критерия необходимо дополнительно проверять нормальность, в итоге он может оказаться мощнее чем непараметрические критерии, однако в случае значительных отклонений от нормальности лучше использовать один из непараметричекси критериев, например Уилкоксона.

3

Статистика t-критерия:

$$T(X_1^n, X_2^n) = \frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}} \sim St(n-1)$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2}, \quad D_i = X_{1,i} - X_{2,i}$$

Статистика критерия знаковых рангов Уилкоксона:

$$W(X_1^n, X_2^n) = \sum_{i=1}^{n} \mathrm{rank}(|X_{1,i} - X_{2,i}|) \cdot \mathrm{sign}(X_{1,i} - X_{2,i}) \sim \text{табличное}$$

Нулевые распределения для обоех статистик является известными. При заданной функции распредления значение достигаемого уровня значимости для двухсторонней альтернативы для обоих критериев вычисляется как (при статистике равной K):

$$p_{val} = 2 \cdot (1 - \mathrm{CDF}(|K|))$$