

# Непараметрические методы анализа

- Если вы имеете несколько групп, то можете использовать *Дисперсионный анализ*.

- Его непараметрическими аналогами являются:

Ранговый дисперсионный анализ Краскела-Уоллиса

Медианный тест

Рассмотрим **критерий Краскела-Уоллиса** подробнее:

Критерий Краскела-Уоллиса является расширением критерия Манна-Уитни и предназначен для сравнения распределений в  $k$  выборках.

$$H_0: F_1 = F_2 = \dots = F_k$$

$H_1$ : Распределения каждой из  $k$  выборок различны

Критерий Краскела-Уоллиса используется, когда невозможно сказать что-либо определенное об альтернативах, т.к. он свободен от распределения.

Число элементов в каждой  $i$ -й выборке ( $i=1, \dots, k$ ) равно  $n_i$

Как было показано выше, заменим наблюдения их рангами , упорядочивая всю совокупность в порядке возрастания.

- $i=1,\dots,k$
- $j=1,\dots,n_i$

Затем для каждой выборки необходимо вычислить суммарный и средний ранги:

$$R_i = \sum_{j=1}^{n_i} r_{ij} \quad \text{и} \quad \bar{R}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} r_{ij}$$

Если между выборками нет систематических различий, то средние ранги  $\bar{R}_i$  не должны значительно отличаться от среднего, рассчитанного по всей совокупности  $\|r_{ij}\|$ .

Значение последнего  $\hat{R} = (N + 1)/2$ .

Здесь  $N$  - общее число наблюдений.

Вычислим величины дисперсий  $(\bar{R}_i - \hat{R})^2$  для каждой выборки  $(\bar{R}_1 - \frac{N+1}{2})^2, \dots, (\bar{R}_k - \frac{N+1}{2})^2$ .

Эти значения при  $H_0$  в совокупности должны быть небольшими. Составляя общую характеристику, разумно учесть различия в числе наблюдений для разных выборок и взять в качестве меры отступления от чистой случайности величину

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left( \bar{R}_i - \frac{N+1}{2} \right)^2$$

# Критерий Манна - Уитни

- Критерий Манна - Уитни можно использовать как непараметрический эквивалент  $t$  - критерия для проверки гипотезы о равенстве средних двух выборок.

$$U_i = \sum_{i=1}^n R(X_i) - \frac{n(n+1)}{2}$$

- Все варианты сравниваемых совокупностей ранжируют в одном общем ряду: каждому значению присваивают ранг, порядковый номер.
- При этом одинаковым значениям вариант должен соответствовать один и тот же средний ранг. После этого ранги вариант суммируют отдельно по каждой выборке.

- Если выборка достаточно велика, то величина статистики сравнивается с табличным значением критерия Стьюдента.
- Метод хорошо подходит для выборок объемом больше 10.
- При меньшем объеме нужно пользоваться специальной таблицей Уилсона-Манна-Уитни.

- Спасибо за внимание!

**Спасибо за внимание!**