42. Критерий Колмогорова-Смирнова однородности двух выборок.

Status	Completed

Даны две выборки $\vec{X}$ и $\vec{Y}$ размерами n и m соответственно из неизвестных распределений F и G соответственно.

$H_0 = \{F = G\}$

$H_a = \{F \neq G\}$

F и G **должны** иметь **непрерывные** функции распределения.

Теорема. Если H_0, то

$$\Psi = \sqrt{\frac{nm}{n+m}} \sup_y |F_n^*(y) - G_m^*(y)| \Rightarrow \eta \sim \mathcal{K}.$$

$\mathcal{K}(t)$ — все та же функция распределения Колмогорова, непрерывная и табулированая.

Критерий Колмогорова-Смирнова

$$\delta = \begin{cases} 0, & \Psi < c \\ 1, & \Psi \geq c \end{cases}$$

где $\mathcal{K}(c) = 1 - \epsilon$.

Свойства

1. Критерий состоятельный.
    1. **"Доказательство".** Если распределения разные, то у них будут разные функции распределения, следовательно точная верхняя грань модуля их разности не будет стремиться к нулю, а корень перед $\sup$ будет стремиться к бесконечности.
2. Критерий имеет асимптотический размер $1 - \epsilon$.

3. Реально допустимый уровень значимости: $\epsilon^* = 1 - \mathcal{K}(\Psi)$.

👍 Молодец, выучил все вопросы, теперь точно сдашь экзамен!