Тест Манна-Уитни

$X_1, \ldots, X_{n_x} \sim iid$ tretment group
$Y_1, \ldots, Y_{n_y} \sim iid$ control group $\Bigg]$ нез. друг от друга

$H_0 : \mu_x = \mu_y$
$H_A : \mu_x > \mu_y$

Z-Тест (ЦПТ):

→ iid
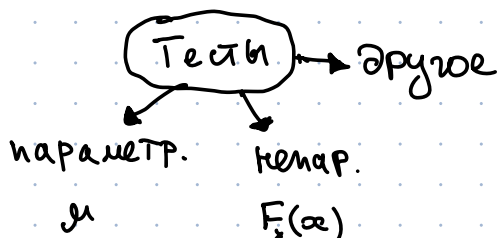→ нет выбросов
→ $n \to \infty$ Big data

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\hat{\sigma}_x^2}{n_x} + \dfrac{\hat{\sigma}_y^2}{n_y}}} \underset{H_0}{\overset{asз}{\sim}} N(0,1)$$

Z_{obs} vs $Z_{1-\alpha}$

Pvalue vs α

Что делать если предпосылки сломались?

- Набл. зав.
- Выбросы
- med вместо $\bar{x}$
- не знаем ас. р-е.
  . . . .

$\Rightarrow$

① Бутстрэп

② Непарамстр. Тесты

Эти Техники решат не все проблемы, но хоть какие-то

Тесты → другое

параметр.   непар.
    μ $F_\xi(x)$

Критерий знаков

МЭР: больше 1/2 кв. дешевле 1 млн. ₽

$x_1, \ldots, x_n \sim iid$

$z_i = [x_i < 1 \text{ млн.}] = 1 \quad \oplus$
$ \geq \phantom{1 \text{ млн.}]} = 0 \quad \ominus$

$H_0: \mathbb{P}(X < 1) \geq \frac{1}{2} \quad = \frac{1}{2}$
$H_A: \mathbb{P}(X < 1) < \frac{1}{2}$

$T = \sum\limits_{i=1}^{n} z_i \underset{H_0}{\sim} Bin\left(\frac{1}{2}; n\right)$

$\overline{T}_{obs} \qquad vs \qquad \overline{T}_{cr}$

~~80% < 1 млн.~~
~~20% > 1 млн~~
~~1 ЛУХХРИ 1 млрд.~~

~~$H_0: \mu_x = 1$ млн.~~
~~$H_A: \mu_x < 1$ млн.~~

выбросы

$T = \sum\limits_{i=1}^{n} [X_i < 1 \text{ млн.}]$

$[\text{истина}] = 1 \quad [\text{ложь}] = 0$

Тест Манна-Уитни

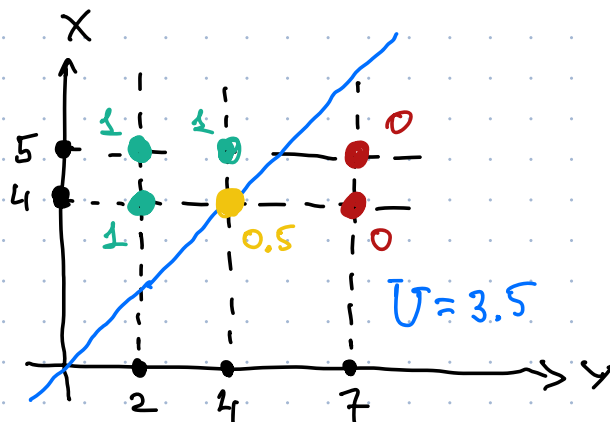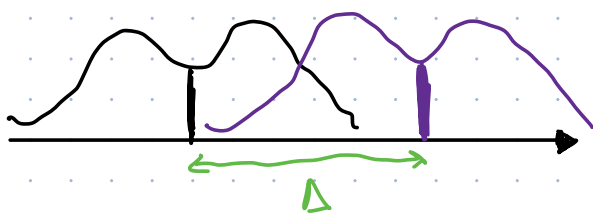Если у нас две незав. выборки $\Rightarrow$ Тест знаков не выйдет

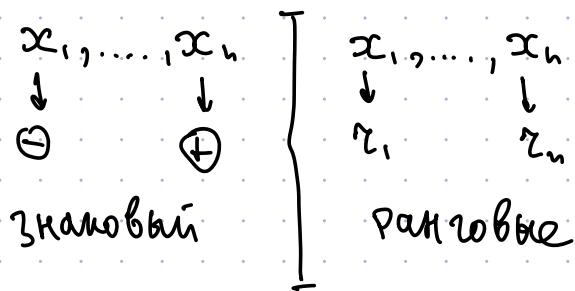$X_1, \ldots, X_n \sim iid$
$Y_1, \ldots, Y_m \sim iid$

$U = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} [X_i > Y_j] \underset{H_0}{\sim} \boxed{?}$

$H_0: F_X(t) = F_Y(t)$
$H_A: F_X(t) = F_Y(t + \Delta), \ \Delta \neq 0$

$U = 3.5$

можно ввести тест ст.з. рангов.

$x_1, \ldots, x_n$ ↓ ⊖   ↓ ⊕   знаковый

$x_1, \ldots, x_n$ ↓ z_1 ↓ z_n ранговые

3   1   -2   4   3

-2   1   3   3   4   набл.
1   2   3.5   3.5   5   ранги

$\dfrac{3+4}{2}$

R_y — сумма рангов набл. из y
в выборке (X, Y)

набл.	откуда	ранг
2	y	1
4	z	2.5
4	x	2.5
5	x	4
7	y	5

$R_x = 2.5 + 4 = 6.5$

$R_y = 8.5$

$n_x \cdot n_y = 6$

$U_x = R_x - \dfrac{n_x \cdot (n_x + 1)}{2}$

$6.5 - \dfrac{2 \cdot 3}{2} = 3.5$

$U_y = R_y - \dfrac{n_y \cdot (n_y + 1)}{2}$

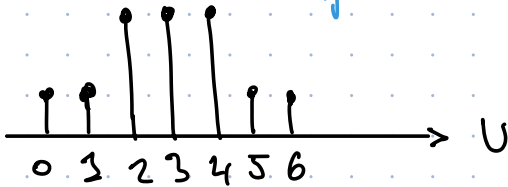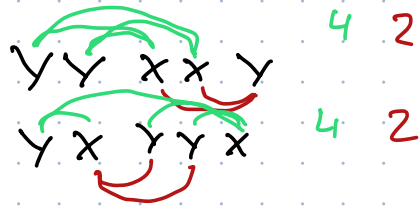$8.5 - \dfrac{3 \cdot 4}{2} = 2.5$

$U_x + U_y = n_x \cdot n_y$

$P(U = 6) = \dfrac{3! \cdot 2!}{5!}$

$\overbrace{YYY}^{3!} \ \overbrace{XX}^{2!}$
$\underbrace{}_{5!}$

$$\mathbb{P}(U=5) = \frac{3! \cdot 2!}{5!}$$

$$\mathbb{P}(U=4) = 2 \cdot \frac{3! \cdot 2!}{5!}$$

$$\ldots \quad U = \sum_{ij} [x_i > y_j]$$

Табличное р.-е.

$$\Rightarrow U_{cr}$$

Огружкое св-во:

$$\frac{U - \mathbb{E}(U)}{\sqrt{Var(U)}} \xrightarrow[\substack{n_x \to \infty \\ n_y \to \infty}]{d} \mathcal{N}(0,1)$$

Big data

$$\alpha \quad \beta$$
$$MDE = \Delta$$
$$\Rightarrow n$$

$$\mathbb{E}(U) = n_y \cdot n_x \cdot \frac{1}{2}$$

$$Var(U) = \ldots$$