

## 2.4

Усманова Анастасия, 674 группа

29 марта 2020 г.

### 1 Формальное описание задачи

Хотим проверить гипотезу равенства средних для классификаторов. Если мы не отклоним эту гипотезу, то исследователь не может утверждать, что какая-то модель лучше.

**Выборка:**  $X^i = \{X_j^i\}_{j=1}^{14}$ ,  $i \in \{b_0, b_1\}$ ,

$X_j^i$  - время восстановления j-го пациента при использовании одеяла вида i

Выборки связанные, n = 14

**Нулевая гипотеза:**  $H_0 : medX^0 = medX^1 // \mu_0 = \mu_1 // P(X^0 > X^1) = 1/2 //$  одеяла не отличаются

**Альтернатива:**  $H_1 : medX^0 \neq medX^1 // \mu_0 \neq \mu_1 // P(X^0 > X^1) \neq 1/2 //$  одеяла отличаются

### 2 Методы решения

Предложим несколько методов решения.

- **Критерий знаковых рангов Уилкоксона**

**Статистика:**  $T(X^0, X^1) = \sum_{k=1}^{14} rank(|X_k^0 - X_k^1|) * sign(X_k^0 - X_k^1)$

Нулевое распределение: табличное (F). Критерий непараметрический. Проверяет гипотезу о равенстве медиан.

**Условия применимости:** Требуется симметричность функции распределения относительно медианы.

**Преимущества и недостатки:** Не требует нормальности данных (+). При нормальных данных может быть менее мощным, чем, например, t-критерий Стьюдента (-).

**Достижимый уровень значимости:**  $p = 2(1 - F(|T|))$

- **t-критерий Стьюдента**

**Статистика:**  $T(X^0, X^1) = \frac{\bar{X}^0 - \bar{X}^1}{S/\sqrt{14}}$

$S = \sqrt{\frac{1}{13} \sum_{i=1}^{14} (D_i - \bar{D})^2}$ ,  $D_i = X_i^0 - X_i^1$

Нулевое распределение: St(13). Критерий параметрический. Проверяет гипотезу о равенстве математических ожиданий выборок.

**Условия применимости:** Требуется нормальность выборки (проверка - критерий Шапиро-Уилка, qq-plot).

**Преимущества и недостатки:** Требуется нормальность данных (-). Может оказаться мощнее критерия знаковых рангов(+).

**Достижимый уровень значимости:**  $p = 2(1 - F_{St(13)}(|T|))$

• *Критерий знаков*

**Статистика:**  $T(X^0, X^1) = \sum_{i=1}^{14} [X_i^0 > X_i^1]$

Нулевое распределение:  $\text{Bin}(14, 1/2)$ . Критерий непараметрический. Проверяет гипотезу о равенстве медиан распределений для связанных выборок.

**Условия применимости:** Необходимо выполнение условия  $(X_i^1 \neq X_i^0 \ \forall i)$ . Простота выборок.

**Преимущества и недостатки:** Не требует нормальности данных (+). Выбрасывает много информации о данных (-).

**Достижимый уровень значимости:**  $p = 2(1 - F_{Bi(14, 1/2)}(|T|))$