# L8.

## Множественное тестирование.

Вгадно! Сильнико? ☺

если одна гипотеза $H_0$ (1 к)
Цель: $\alpha = P(\text{отв } H_0 \mid H_0 \text{ верна})$

Упр. 1000 (независимых) наборов данных
$H_0^1 \ldots\ldots H_0^{1000}$ $\boxed{\alpha = 0.01}$

$H_0^i$ (об отсут. эффекта) все верны.

а) P(хотя бы $H_0^i$ будет отвергнута) = ?
б) E(кол-во отвергнутых гипотез $H_0^i$) = ?

p-value hacking.

а) $P = 1 - (1-\alpha)^{1000} =$
$= 1 - \left(1 - \frac{1}{100}\right)^{1000} \approx$
$\left(1 - \frac{1}{100}\right)^{100} \approx e^{-1}$
$\approx 1 - e^{-10} > 0{,}999$

б) $N$ - кол-во отвергн-х $H_0^i$
$N \sim Bin(n=1000, p=0.01)$
$E(N) = 10$

| | $H_0$ не отв | $H_0$ отвергнут |
|---|---|---|
| $H_0$ верна | $U$ | $V$ |
| $H_0$ не верна | $T$ | $S$ |

$U + V = M_0$

$U + V + T + S = M$

общее кол-во проверяемых гипотез.

family wise error rate

$$FWER = P(V > 0)$$

false discovery rate

$$FDR = E\left(\frac{V}{\max(V+S, 1)}\right)$$

часто пишут $E\left(\frac{V}{V+S}\right)$

утв $\quad FWER \overset{?}{\geq} FDR$

док-во:

$$E(I(V>0)) \text{ vs } E\left(\frac{V}{\max(V+S,1)}\right)$$

если $V=0 \quad I=0 \quad \frac{V}{\max(V+S,1)} = 0$

если $V>0 \quad I=1 \quad \frac{V}{\max(V+S,1)} \leq 1$

жёсткая цель $FWER \leq \alpha \Rightarrow$ мягкая цель : $FDR \leq \alpha$

Цель

$FWER = P(V>0) \leq \alpha$.

| | $H_0$ не отв | $H_0$ отвергнут |
|---|---|---|
| $H_0$ верна | $U$ | $V$ |
| $H_0$ не верна | $T$ | $S$ |

поправка Бонферонни:

уменьшить ур-нь знач-мости для каждой $H_0^i$.

$p_1, p_2 \; \ldots \; p_M$ – p-значения

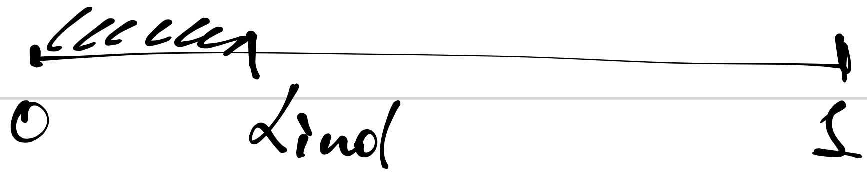
$I_0$ – множество верных $H_0$ $\quad \{p_i < \alpha_{ind}\} = \{H_0^i \text{ отвер}\}$

$$P(V>0) = P\left(\bigcup_{i \in I_0} \{p_i < \alpha_{ind}\}\right) \leq \sum_{i \in I_0} P(p_i < \alpha_{ind}) =$$

факт: При верной $H_0$ p-value $\sim U[0;1]$

$$P(V>0) = \sum_{i \in I_0} P(p_i < \alpha_{ind}) = \alpha_{ind} \cdot M_0$$

$M_0$ - кол-во верных $H_0^i$





худший случай $P(V>0) \leq \alpha_{ind} \cdot M \leq \alpha$

цель

поправка Бонферрони $\alpha_{ind} = \frac{\alpha}{M}$

!!

---

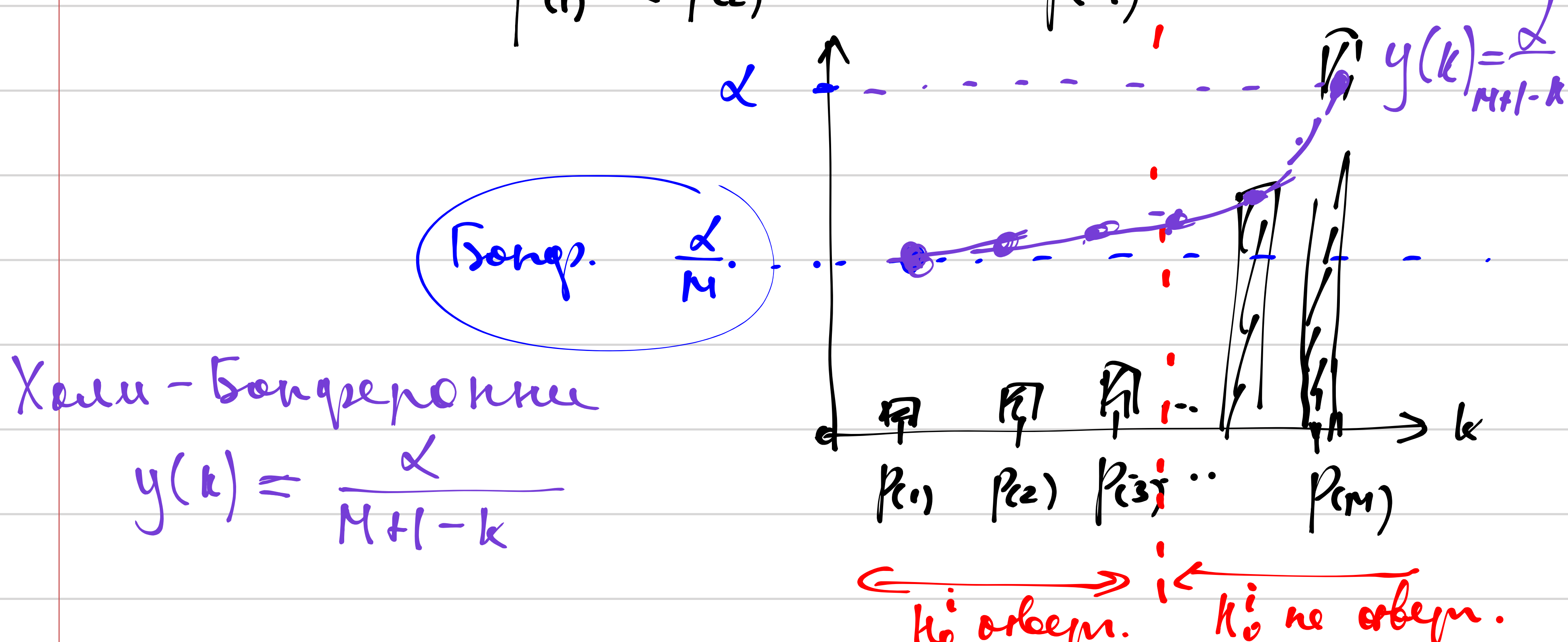
Holm-Bonferroni (поправка)

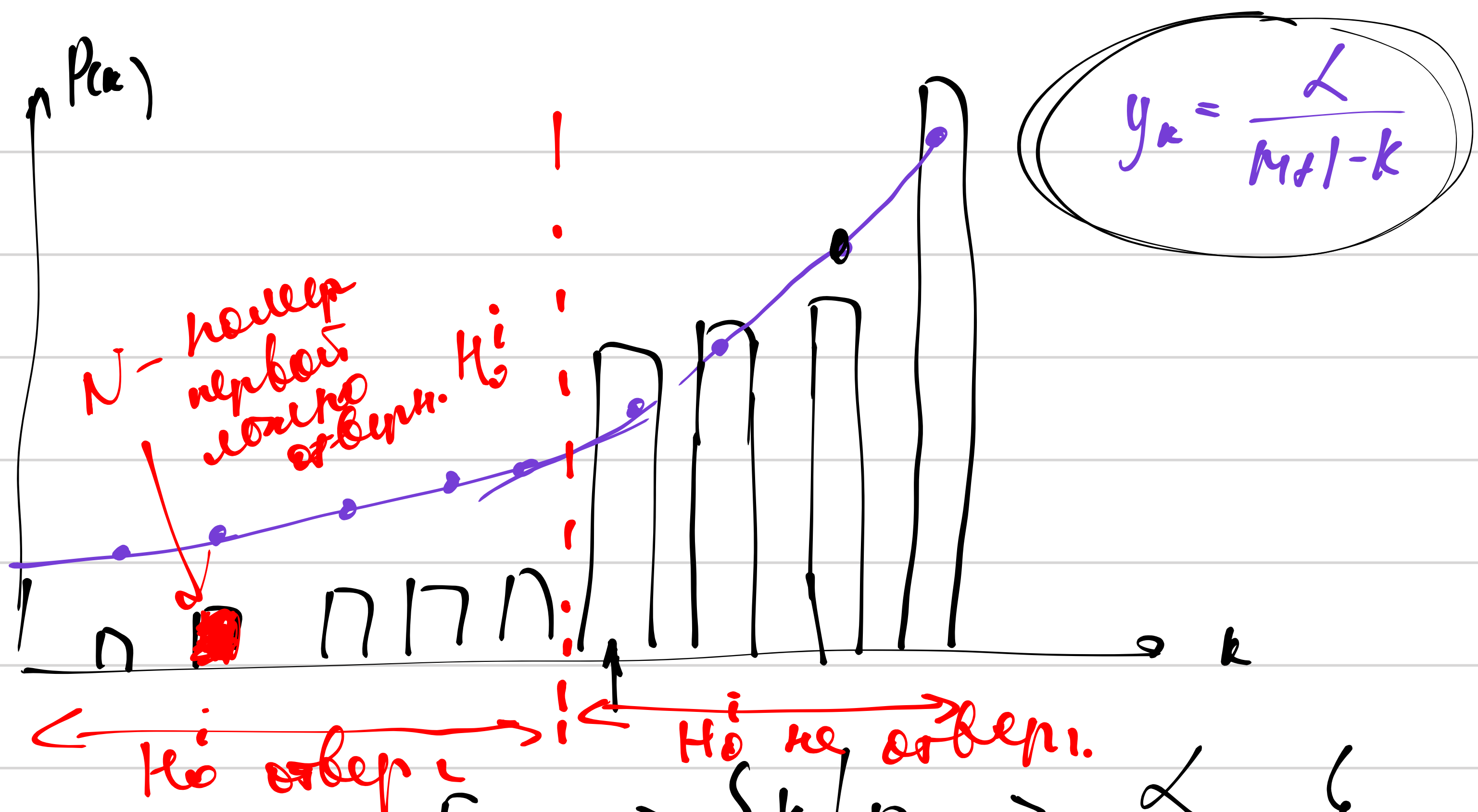
Цель: $P(V>0) \leq \alpha$

Алгоритм.

Шаг 1) считаем p-значение для каждой $H_0^i$ и сортируем.

$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(M)}$






$$F = \min\left\{k \,\middle|\, p(k) \geq \frac{\alpha}{M+1-k}\right\}$$

Алгоритм: гипотезы $F, F+1 \ldots M$ не отверг.
гипотезы $1, 2, 3, \ldots (F-1)$ отверг.

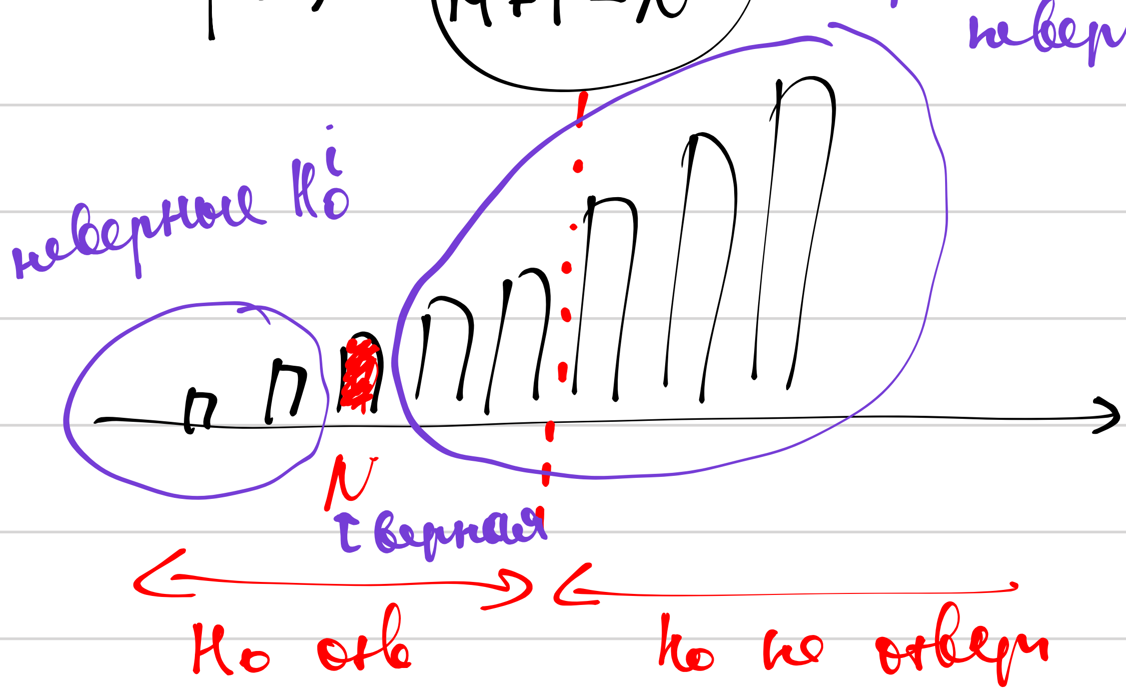
Теорема: алгоритм гарантирует $P(V>0) \leq \alpha$

док-во: найдём ложно отвергнутую $H_0^i$ с наим-м номером.

[p(i) сортир и задают сортировку $H_0^i$]

Какое у неё могло быть p-значение?

$$p(N) < \frac{\alpha}{M+1-N}$$




$M_0 =$ кол-во верных $H_0^i$

$M =$ кол-во всего $H_0^i$

$\{V>0\} \Rightarrow$

$M_0 \leq 1 + (M-N)$

$$\left\{ p(N) < \frac{\alpha}{M+1-N} \leq \frac{\alpha}{M_0} \right\}$$

Шаг 2.

$$FWER = P(V>0) \le P\left(p_{(N)} \le \boxed{\frac{\alpha}{M_0}}\right) \le$$

$$\le P\left(\bigcup_{i\in I_0}\left\{p_i \le \frac{\alpha}{M_0}\right\}\right) \le \sum_{i\in I_0} P\left(p_i \le \frac{\alpha}{M_0}\right) =$$

$$= M_0 \cdot \frac{\alpha}{M_0} = \alpha.$$

---

Цель $FDR = E\left(\frac{V}{\max(V+S, 1)}\right) \le \alpha$

| | $H_0$ не отв | $H_0$ отвергнут |
|---|---|---|
| $H_0$ верна | $U$ | $V$ |
| $H_0$ не верна | $T$ | $S$ |

---

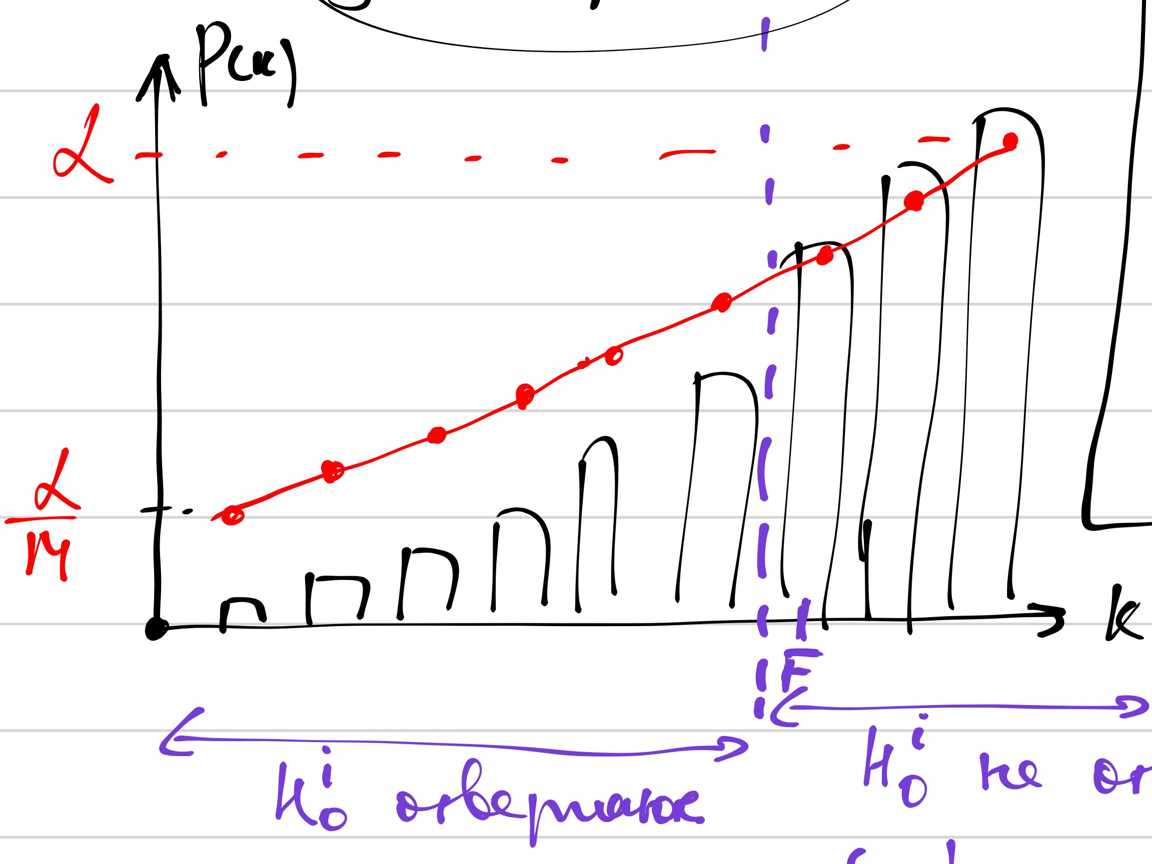
Benjamini-Hochberg алгоритм.

Цель: $FDR \le \alpha$

Шаг 1. Сортируем p-значения

$$p_{(1)} \le p_{(2)} \le \ldots \le p_{(m)}$$

Шаг 2. $y(k) = \frac{\alpha}{M} \cdot k.$




$$F = \min_k \{k \mid p_{(k)} \ge y(k)\}$$

Теорема: Если $p_i$ независимы, то алгоритм BH гарантирует $FDR \le \alpha$