Семинар 17

→ множественное тестирование гипотез.

Упр (ручка)

5 гипотез: (нулевых)

$$\overset{H_1}{H_{01}} \quad \overset{H_2}{H_{02}} \quad \dots \quad \overset{H_5}{H_{05}}$$

$$H_{A1} \quad - - - - \quad H_{A5}$$

p-value:

p_1	p_2	p_3	p_4	p_5
0.08	0.01	0.07	0.02	0.10

а) примените метод Бонферрони, чтобы гарантировать $P(FD>0) = FWER \le 0.05$

б) — // — Холма-Бонферрони, чтобы гарантировать $FWER \le 0.05$

в) — // — Бенджамини-Хохберга, чтобы гарантировать $FDR \overset{*}{=} E\left(\frac{FD}{D}\right) \le 0.05$

* доопр-ем $\frac{FD}{D} = 0$ если $D=0$, $FD=0$

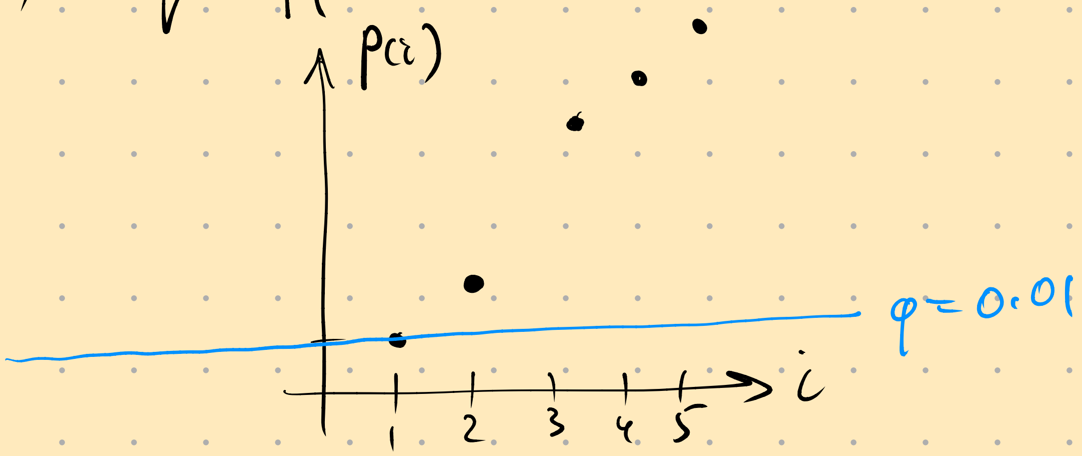
упорядочим p-значения

$$p_{(1)} \le p_{(2)} \le p_{(3)} \le p_{(4)} \le p_{(5)}$$

$$0.01 \le 0.02 \le 0.07 \le 0.08 \le 0.10$$

а) $\varphi = \frac{\alpha}{M} = \frac{0.05}{5} = 0.01$

Хотим $FWER = P(FD>0) \le \alpha$



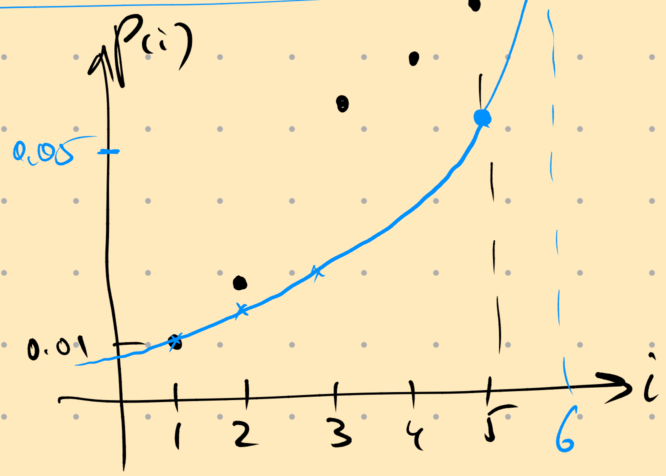
Отвергните $H_{(i)}$ у которых $p_{(i)} \le \varphi$

Вывод: $H_{(1)}$ $p_{(1)}=0.01 \Rightarrow$ отверг.

$\left.\begin{matrix} H_{(2)} \\ H_{(3)} \\ H_{(4)} \\ H_{(5)} \end{matrix}\right\}$ не отвергается.

гарантия $P(FD>0) \leq \alpha = 0{,}05$

б)



$q(i) = \dfrac{\alpha}{M+1-i} = \dfrac{0.05}{6-i}$

$\alpha = 0.05$

$M = 5$

Алгоритм Холма-Бонферрони:

Шаг 1: $\underset{0.01}{p_{(1)}} \overset{?}{\leq} q(1) = \dfrac{\alpha}{M} = \dfrac{\alpha}{5} = 0{,}01$

да → $H_{(1)}$ отверг. к шагу 2

нет → EXIT

Шаг 2: $\underset{0.02}{p_{(2)}} \overset{?}{\leq} q(2) = \dfrac{\alpha}{M+1-2} = \dfrac{0.05}{6-2} = \dfrac{0.05}{4} = 0.0125$

да → $H_{(2)}$ отверг к шагу 3

нет → EXIT.

Итог: $H_{(1)}$ отверг.

$\left.\begin{matrix} H_{(2)} \\ H_{(3)} \\ H_{(4)} \\ H_{(5)} \end{matrix}\right\}$ не отвергнуты.

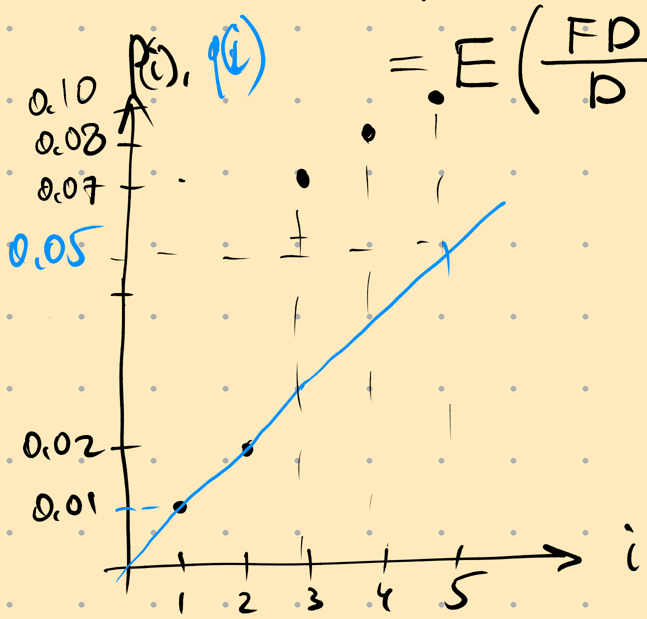
$P(FD>0) \leq \alpha$

FD = false discoveries = кол-во ложно отвергнутых нулевых гипотез.

б) алгоритм Бенжамини – Хохберга.

(!) цель $FDR \le \alpha$

false discovery rate $= E\left(\frac{FD}{D}\right)$



FD – число ложно отвергнутых нулевых гипотез

D – число отвергнутых нулевых гипотез.

$$q(i) = \frac{\alpha i}{M} = \frac{0.05 \cdot i}{5} = 0.01 i$$

Шаг 1 $0.01 = p_{(1)} \overset{?}{\le} q(1) = 0.01 \cdot 1 = 0.01$

да → $H_{(1)}$ отверг. перех к шаг 2

нет → EXIT

Шаг 2 $0.02 = p_{(2)} \overset{?}{\le} q(2) = 0.01 \cdot 2 = 0.02$

да → $H_{(2)}$ отверг. перех к шаг 3.

нет → EXIT

Шаг 3 $0.07 = p_{(3)} \overset{?}{\le} q(3) = 0.01 \cdot 3 = 0.03$

да → $H_{(3)}$ отверг. перех. к шагу 4

нет → EXIT

Вывод:

$H_{(1)}$ отверг.
$H_{(2)}$ отверг.

$\left.\begin{array}{l} H_{(3)} \\ H_{(4)} \\ H_{(5)} \end{array}\right\}$ не отвергн.

Упр. Что больше?

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

$$FDR = E\left(\frac{FD}{D}\right)$$

* доопр. $FD/D = 0$ при $D=0$

$$E\left(\begin{cases} FD/D, & D>0 \\ 0, & D=0 \end{cases}\right)$$

таблица сопр-сти (количеств)

	верные $H_{0,i}$	неверные $H_{0,i}$	Σ
отвергнутые $H_{0,i}$	FD	TD	D
не отвергн. $H_{0,i}$	TN	FN	N
Σ	M_0	M_1	M

две „типичных" таблички

FD=0	TD=3	D=3
TN=5	FN=6	N=11

FD=2	TD=1	D=3
TN=5	FN=6	N=11

$FD=0$ $\quad I[FD>0]=0$

$$\frac{FD}{D}=0$$

$FD>0$ $\quad I[FD>0]=1$

$$\frac{FD}{D}=\frac{2}{3}$$

$$FWER = P(FD>0) = E(I[FD>0])$$

$$FDR = E\left(\frac{FD}{D}\right) = E\left(\frac{FD}{D}\right)$$

$$\frac{FD}{D} \overset{*}{\leq} I[FD>0]$$

* при согласишении $\frac{FD}{D}=0$ при $D=0$.

$$FDR = E\left(\frac{FD}{D}\right) \leq P(FD>0) = FWER$$

замечание: из $FWER \leq 0.05$ следует, что $FDR \leq 0.05$

Упр. предположим, что все $H_1, H_2, H_3, \ldots, H_m$ верны. Как связаны FWER и FDR?

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

	H_0 верные	H_0 неверные	Σ
H_0 отвергн.	FD	0	D=FD
H_0 не отверг.	TN	0	N=TN
Σ	$M=M_0$	0	M

$$FDR = E\left(\frac{FD}{D}\right) \overset{⊛}{=}$$

договр. ⊛ $\frac{FD}{D} = 0$ если $D=0$

по согл.

$$\frac{FD}{D} = \begin{cases} 1, & \text{если } D=FD>0 \\ 0, & \text{если } D=FD=0. \end{cases}$$

$$= E(\text{индек-р}) = P(FD>0)$$

$$\boxed{FWER = FDR}\ !!$$

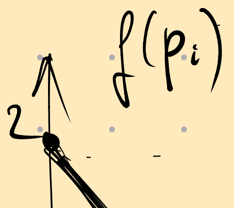
Упр. → провер-ляем гипотезы $H_1, H_2, \ldots$, незав.

Истинные вер-сти

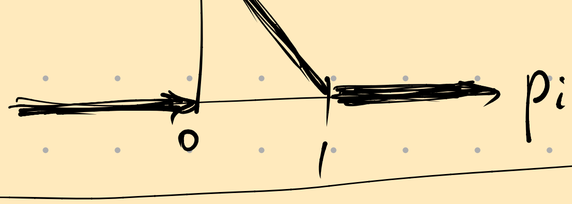
$$P(H_i \text{ верна}) = 0.3 \quad \forall i$$

если H_i не верна, то p_i = p-value для i-ой нулевой гипотезы.

p_i имеет плотность.



$$f(p_i) = \begin{cases} 2-2p_i & p_i \in [0;1] \end{cases}$$



$\{0, \quad p_i \notin [0;1]$

Если нулевая гипотеза верна, то p-value $\sim U[0;1]$

если H_i верна, то $p_i \sim U[0;1]$

$M=3$ FDR? используем алгоритм Бенджамини-Хохберга, таргетирую $FDR \leq 0.05$

$$FDR \overset{*}{=} E\left(\frac{FD}{D}\right) = E\left(\frac{FD}{D}\Big|M=0\right)\cdot P(M=0) + \ldots + E\left(\frac{FD}{D}\Big|M=3\right)\cdot P(M=3).$$

* доопр. $\frac{FD}{D}=0$ при $D=0$

$M=0$ ← кол-во верных нулевых гипотез.

$\hookrightarrow FD=0$. $\quad E\left(\frac{0}{D}\Big|M=0\right)=0$

$$FDR = E\left(\frac{FD}{D}\Big|M=1\right)\cdot P(M=1) + \ldots + E\left(\frac{FD}{D}\Big|M=3\right)\cdot P(M=3)$$

$$M \sim Bin(3, 0.3)$$

$$P(M=1) = C_3^1 \cdot 0{,}3^1 \cdot 0{,}7^2$$
$$P(M=2) = C_3^2 \cdot 0{,}3^2 \cdot 0{,}7^1$$
$$P(M=3) = C_3^3 \cdot 0{,}3^3 \cdot 0{,}7^0 = 0{,}3^3$$

$$E\left(\frac{FD}{D}\Big|M=1\right) \overset{?}{=}$$

H_1 верна $\quad H_2$ не верна $\quad H_3$ не верна

$FD \nearrow 0, \ \searrow 1$

$$= E\left(\frac{0}{D}\Big|M=1, FD=0\right)\cdot P(FD=0|M=1) + E\left(\frac{1}{D}\Big|M=1, FD=1\right)\cdot P(FD=1|M=1)$$

$\downarrow$ 0

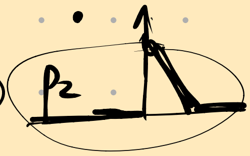
одна верная H_0

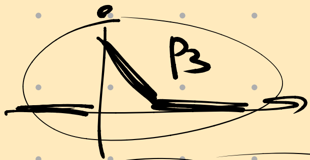
$$P(FD=1 \mid M=1) \overset{?}{=}$$

[Алгоритм Бенджамини-Хохберга]

H_1 верна, H_2 не верна, H_3 не верна.

$p_1 \sim U[0;1]$





сит. 1:	$p_1 < p_2 < p_3$, $p_1 < \frac{0.05}{3}$	$p_1 < p_3 < p_2$
сит 2:	$p_2 < p_1 < p_3$, $p_2 < \frac{0.05}{3}$, $p_1 < \frac{2\cdot 0.05}{3}$	$p_3 < p_1 < p_2$, $p_3 < \frac{0.05}{3}$, $p_1 < \frac{2\cdot 0.05}{3}$
сит 3:	$p_2 < p_3 < p_1$, $p_2 < \frac{0.05}{3}$, $p_3 < \frac{2\cdot 0.05}{3}$, $p_1 < \frac{3\cdot 0.05}{3}$	$p_3 < p_2 < p_1$, —//—

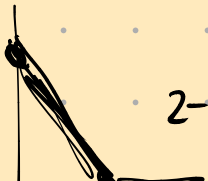
$$P\begin{pmatrix} p_1 < p_2 < p_3 \\ p_1 < 0.05 \end{pmatrix} \overset{?}{=} E\left[\underline{P(p_1 < p_2 < p_3 \mid p_1)}\right]$$

$p_1 \sim U[0;1]$

$$P(p_3 > p_2 > p_1 \mid p_1) =$$

$$= \frac{1}{2}\left(1 - F_A(p_1)\right)^2 = \frac{1}{2}\left(1 - 2p_i + p_i^2\right)^2$$

$f_A(x) =$



$2-2p_i$

$$F_A(p_i) = \int_0^{p_i} 2-2t \, dt = 2t - t^2 \Big|_{t=0}^{p_i} = \begin{cases} 0 \\ 2p_i - p_i^2, & p_i \in [0,1] \\ 1 \end{cases}$$

$$P(p_1 < p_2 < p_3) = E\left(\frac{1}{2}\left(1 - 2p_1 + p_1^2\right)^2\right) =$$

$$= \frac{1}{2} \cdot E\left(1 + 4p_1^2 + p_1^4 - 4p_1 - 4p_1^3 + 2p_1^2\right) =$$

$p_1 \sim U[0,1]$

$$E(p_1^k) = \frac{1}{k+1}$$

$$= \frac{1}{2}\left(1 + 4 \cdot \frac{1}{3} + \frac{1}{5} - 4 \cdot \frac{1}{2} - 4 \cdot \frac{1}{4} + 2 \cdot \frac{1}{3}\right) \;!!$$