Лекция 17

• множественное тестирование гипотез.

$$\begin{array}{ccccc} H_{0,1} & H_{0,2} & \ldots & \ldots & H_{0,M} \\ \updownarrow & \updownarrow & & & \uparrow \\ H_{A,1} & H_{A,2} & & & H_{A,M}. \end{array}$$

всего:
M решений   каждая $H_{0,i}$ → отвергн. / → не отвергнута.

для краткости: $H_i \equiv H_{0,i}$

количества

	нулевая верна	нулевая ложная	Σ
нулевая отвергн.	FD	TD	D
нулевая не отвергн.	TN	FN	N
Σ	M_0	M_1	M

общее кол-во „обнаружений"

число не отвергнутых нулевых гипотез.

$M_0 + M_1 = M$

традиция: [F – false ложно, T – верно true] [D – discovery, N – non-discovery]

FD = false discoveries =
= число „открытий" =
= число ложно отвергнутых нулевых гипотез.

TN = true non-discoveries =
= число вер не гнутых нулевых

непорд.

! классич. трактовка: M_0, M_1, M — фиксир. конст!

индив-ая вер-сть ошибки 1-го рода
$P(H_i \text{ отвергнута} \mid H_i \text{ верна})$

$M=3$
$I_0=\{1,3\}$

Обознач: I_0 = индексы верных H_i

Опр. групповая ошибка 1-го рода.
family wise error rate

$$FWER = P(\text{хотя бы одну } H_i \text{ отвергнуть} \mid I_0) =$$

$$= P(FD > 0)$$

Цель: ограничить FWER

метод 1 [Šidák] метод Шидака

Предположения:

- все $H_1, H_2, \dots H_M$ верны. (все нулевые)
- p_i независимы

индив. критерий для H_i

H_i отв $\quad$ H_i не отверг. $\quad p_i$

$0 \qquad q \qquad 1$

Цель: $FWER \leq \alpha$

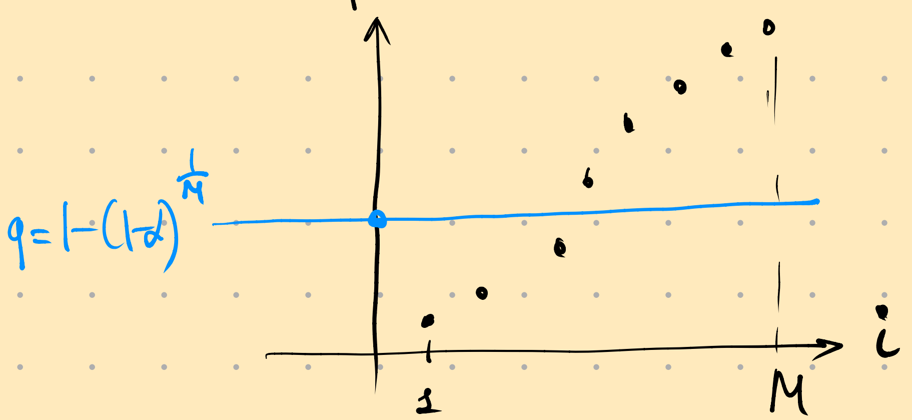
$$FWER = P(FD>0) = 1 - P(FD=0) =$$
$$= 1-(1-q)^M = \alpha$$

$$(1-q)^M = 1-\alpha$$

$$1-q = (1-\alpha)^{\frac{1}{M}}$$

$$q = 1-(1-\alpha)^{\frac{1}{M}}$$

$p_{(i)}$



отсортированные p-значения:

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

критерий (Šidák): отвергаем те H_i у к-рых $p_i \le q = 1-(1-\alpha)^{\frac{1}{M}}$

Гарантирует: если все $H_1, \ldots, H_M$ верны и p_i незав-но: $FWER \le \alpha$.

Метод Bonferroni (Бонферрони)

- некоторые $H_1, \ldots, H_M$ верны, нек-рые — нет.
- p_i м. быть зависимы.

Цель: $FWER = P(FD > 0) \le \alpha$

индив. критерий для H_i: H_i отв при $0 \le p_i \le q$, H_i не отв. при $q < p_i \le 1$

$P(FD > 0) =$

$= P($хотя бы одна H_i при $i \in I_0$ отвергнута$) =$

$$= P\Big(\bigcup_{i \in I_0} p_i \le q\Big) \le \sum_{i \in I_0} P(p_i \le q) = \sum_{i \in I_0} q = q \cdot I_0$$

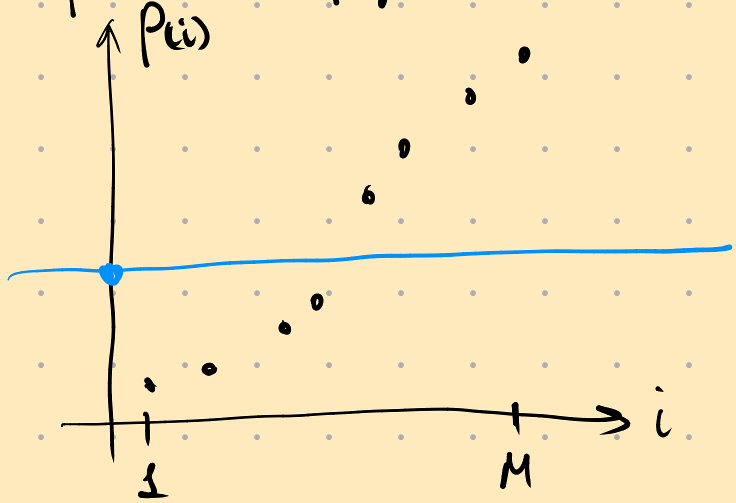
при верной нулевой H_i: $p_i \sim U[0;1]$

цель: $q \cdot M_0 \le \alpha$

хочется взять: $q = \frac{\alpha}{M_0}$ поэтому возьмем $q = \frac{\alpha}{M}$

$M_0 \le M$

Критерий Бонферрони:



I_0 не знаем!!
M_0

$q = \frac{\alpha}{M}.$

Отвергаем все H_i, у которых $p_i \le q = \frac{\alpha}{M}.$

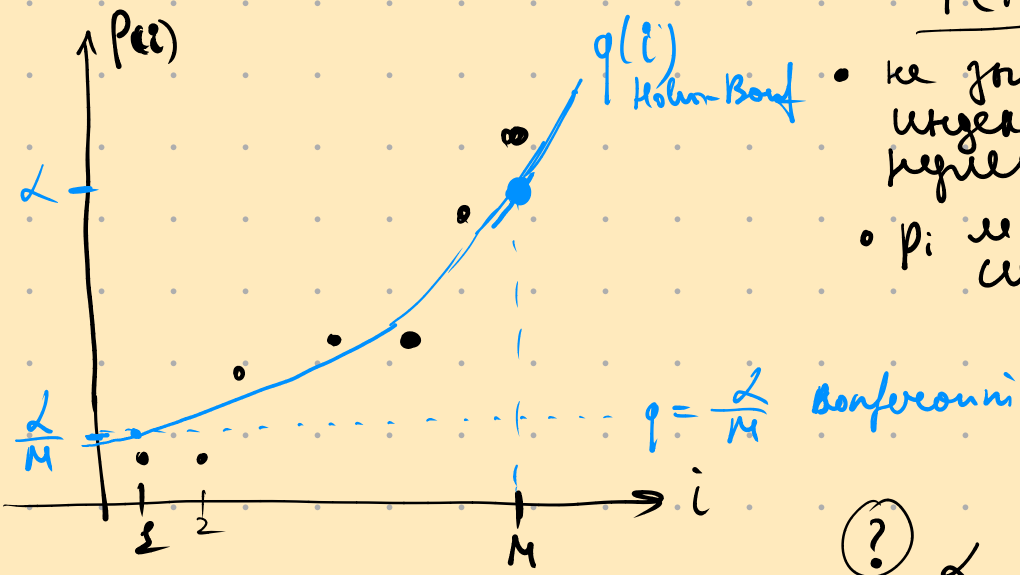
Гарантии: $\forall I_0$
зависимы/не зависимы p_i

FWER $\le \alpha$.

Метод Холма - Бонферрони (Holm – Bonferroni)

Цель: FWER = $P(FD > 0) \le \alpha$

- не знаем I_0 индексы верных нулевых
- p_i м. быть зависимы.



$$q(i) = \frac{\alpha}{M+1-i}$$

Шаг 1

$p_{(1)} \overset{?}{\le} \frac{\alpha}{M}$

да: $H_{(1)}$ отвергаем, перех к шаг 2.

нет: EXIT

Шаг 2

$p_{(2)} \overset{?}{\le} \frac{\alpha}{M-1}$

нет

$H_{(2)}$ отверг.
перех к шаг 3

EXIT

Шаг 3

$$p_{(3)} \overset{?}{\le} \frac{\alpha}{M-2}$$

да: $H_{(3)}$ отверг, перех к шаг 4

нет: EXIT

...

Шаг i

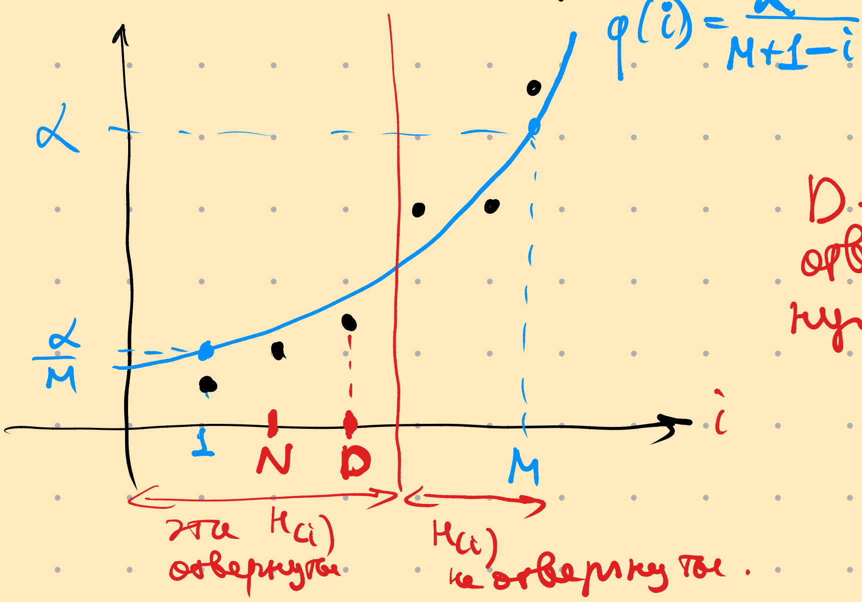
$$p_{(i)} \overset{?}{\le} \frac{\alpha}{M+1-i}$$

да: $H_{(i)}$ отверг, к шаг $(i+1)$

нет: EXIT.

док-во:

Лемма. N – наименьший номер верной отвергнутой гипотезы.



D = число отвергнутых нулевых гипотез.

$N \le D$!

Лемма: $p_{(N)} \le \frac{\alpha}{M_0}$

док-во: по определению Холма–Бонферрони

$$p_{(N)} \le \frac{\alpha}{M+1-N}$$

(ниже синей гиперболы)

верная гип. с наим. но...

i : $\underbrace{(1)\ (2) \ldots (N-1)}_{N-1 \text{ гипотеза неверные}}\ (N)\ (N+1) \ldots\ldots (M)$

$$N-1 \le M - M_0$$

$$M_0 \le M+1-N$$

$$p_{(N)} \le \frac{\alpha}{M+1-N} \le \frac{\alpha}{M_0}$$

завершение док-ва:

$$FWER = P(FD>0) \le P\left(p_{(N)} \le \frac{\alpha}{M_0}\right) \le P\left(\bigcup_{i \in T_0} p_i \le \frac{\alpha}{M_0}\right) \le$$

есть какая-то верная гипотеза с $p_i \le \frac{\alpha}{M_0}$

$$\le \sum_{i \in T_0} P\left(p_i \le \frac{\alpha}{M_0}\right) = M_0 \cdot \frac{\alpha}{M_0} = \alpha$$

$$p_i \sim U[0;1]$$

Другая популярная цель.

контроль: FDR

$FWER = P(FD>0)$

FDR - false discovery rate

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

(*) соглашение
если $D=0$, то $FD=0$
и доопр-м $\frac{FD}{D}=0$.

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

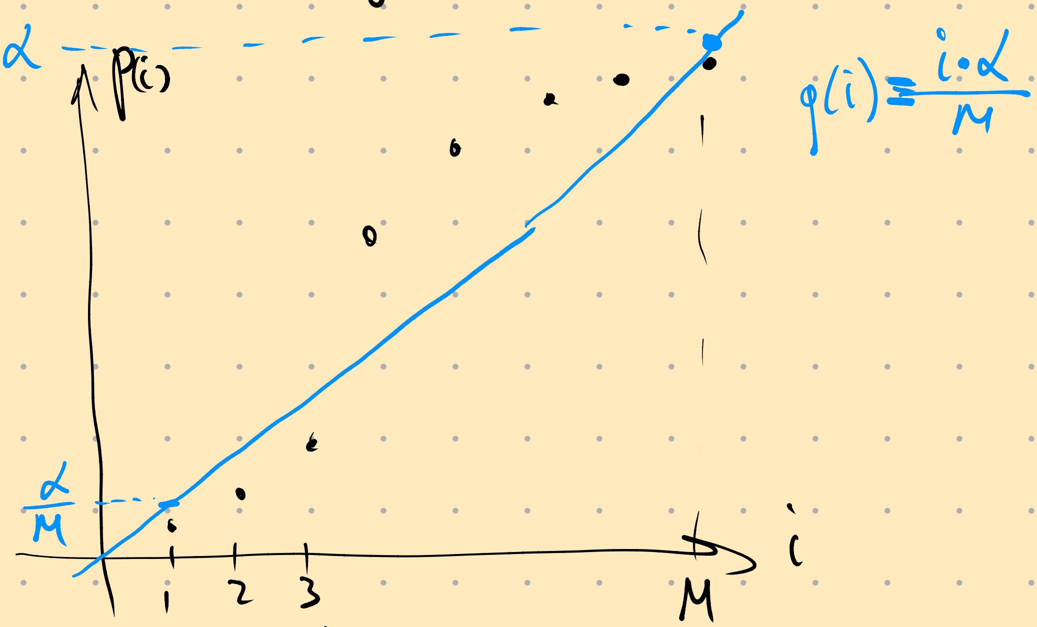
цель: контроль за FDR

$FDR \le \alpha$

процедура Бенджамини – Хохберга.
[Benjamini-Hochberg]

- p_i независимы (можно ослабить)
- индексы I_0 (верных нулевых гипотез) не знаем.

алгоритм



шаг 1 $\quad p_{(1)} \overset{?}{\leq} q(1) = \frac{\alpha}{M}$

да ↙ отвергаем $H_{(1)}$ и к шагу 2

нет ↘ EXIT

шаг 2 $\quad p_{(2)} \overset{?}{\leq} q(2) = \frac{2\alpha}{M}$

да ↙ отвергаем $H_{(2)}$ и к шагу 3

нет ↘ EXIT.

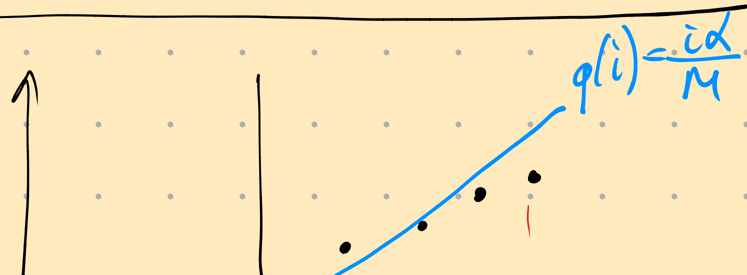
⋮

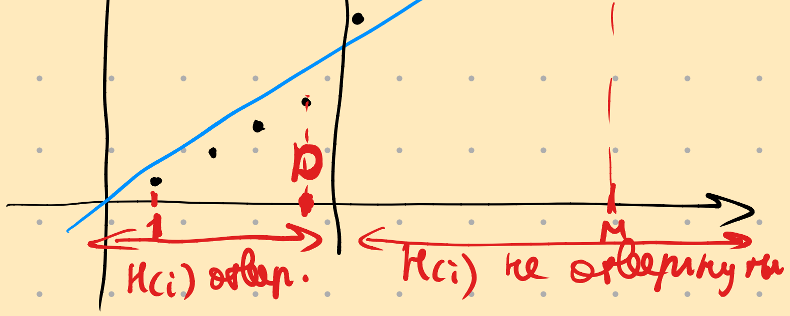
шаг (i) $\quad p_{(i)} \overset{?}{\leq} q(i) = \frac{i\alpha}{M}$

да ↙ отвергаем $H_{(i)}$ и к шагу $(i+1)$

нет ↘ EXIT.

док-во: цель: $\boxed{FDR \leq \alpha}$





трюк.

D_i = число отвергнутых гипотез, если $\boxed{\text{мысленно}}$ подменить p_i на 0.

$$D_i \to D, \quad D_i \to D+1$$

лемма $\quad P(p_i \le q(d), D=d) = P(p_i \le q(d), D_i = d)$

зачем такая очевидная лемма?

→ p_i и D зависимые с.в.

→ p_i и D_i независимые с.в.

* $\frac{FD}{D} = 0$ [опр при $D=0$]

док-во:

$$E\left(\frac{FD}{D}\right) = \sum_{d=1}^{M} E\left(\frac{FD}{D} \Big| D=d\right)\cdot P(D=d) =$$

$$= \sum_{d=1}^{M} \frac{1}{d} E\left(FD \cdot I(D=d)\right) =$$

$$= \sum_{d=1}^{M} \sum_{i\in I_0} \frac{1}{d} \cdot E\left(I[p_i \le q(d)] \cdot I[D=d]\right) =$$

$$= \sum_{d=1}^{M} \sum_{i\in I_0} \frac{1}{d} \cdot P(p_i \le q(d), D=d) \overset{\text{лемма}}{=}$$

$$= \sum \sum \frac{1}{d} \cdot P(p_i \le q(d), D_i = d) \overset{\text{незав.}}{=}$$

$$= \sum \sum_{i\in I_0} \frac{1}{d} \cdot \underbrace{P(p_i \le q(d))} \cdot P(D_i = d) =$$

для $i \in I_0 \quad p_i \sim U[0;1]$

$$P\big(p_i \le \rho(d)\big) = \rho(d) = \frac{d\cdot\alpha}{M}$$

$$= \sum_{d=1}^{M} \sum_{i\in I_0} \frac{1}{\cancel{d}} \cdot \frac{\cancel{d}\cdot\alpha}{M} \cdot P\big(D_i = d\big) =$$

$$= \sum_{i\in I_0} \frac{\alpha}{M} \cdot \underbrace{\sum_{d=1}^{M} P\big(D_i = d\big)}_{=1} = M_0 \cdot \frac{\alpha}{M} = \frac{M_0}{M}\cdot\alpha \le \alpha.$$