

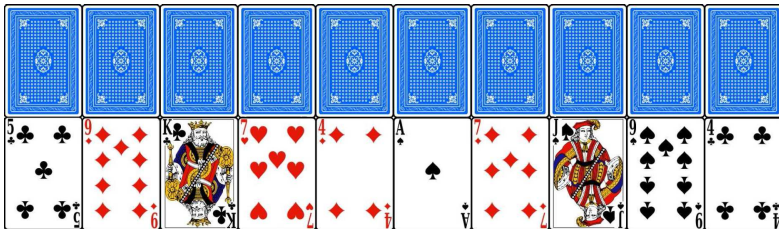
## Прикладная статистика 4. Множественная проверка гипотез.

Рябенко Евгений  
riabenko.e@gmail.com

3 марта 2014 г.

## Поиск экстрасенсов

Joseph Rhine, 1950: исследования возможности экстрасенсорного восприятия. Первый этап — поиск экстрасенсов.  
Испытуемому предлагается угадать цвет 10 карт.



$H_0$ : испытуемый выбирает ответ наугад.

$H_1$ : испытуемый может предсказывать цвета карт.

Статистика  $t$  — число карт, цвета которых угаданы.

$$P(t \geq 9 | H_0) = 11 \cdot \frac{1}{2}^{10} = 0.0107421875,$$

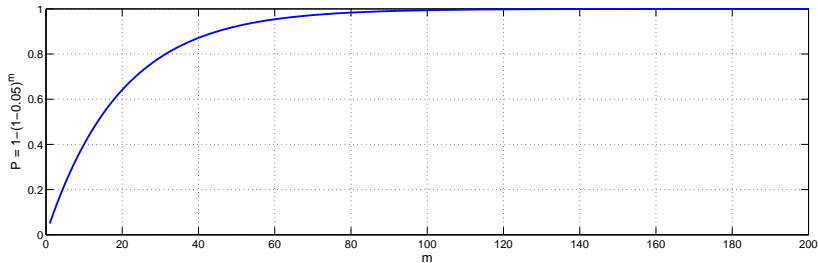
т. е. при  $t = 9$  получаем достигаемый уровень значимости  $p \approx 0.01$  — можно отклонять  $H_0$ .

## Поиск экстрасенсов

Процедуру отбора прошли 1000 человек.

Девять из них угадали цвета 9 из 10 карт, двое — цвета всех 10 карт. Ни один в последующих экспериментах не подтвердил своих способностей.

Вероятность того, что из 1000 человек хотя бы один случайно угадает цвета 9 или 10 из 10 карт:  $1 - \left(1 - 11 \cdot \frac{1}{2}^{10}\right)^{1000} \approx 0.9999796$ .

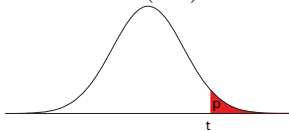


# Математическая формулировка

выборка:  $X^n = (X_1, \dots, X_n) \sim P \in \Omega$ ;  
 нулевая гипотеза:  $H_0: P \in \omega, \omega \in \Omega$ ;  
 альтернатива:  $H_1: P \notin \omega$ ;  
 статистика:  $T(X^n), T(X^n) \sim F(x)$  при  $P \in \omega$ ;  
 $T(X^n) \not\sim F(x)$  при  $P \notin \omega$ ;



реализация выборки:  $x^n = (x_1, \dots, x_n)$ ;  
 реализация статистики:  $t = T(x^n)$ ;  
 достигаемый уровень значимости:  $p(x^n)$  — вероятность при  $H_0$  получить  $T(X^n) = t$  или ещё более экстремальное;



$$p(x^n) = P(T \geq t | H_0)$$

Гипотеза отвергается при  $p(x^n) \leq \alpha$ ,  $\alpha$  — уровень значимости.

# Правило проверки гипотезы



## Несимметричность задачи проверки гипотезы

	$H_0$ верна	$H_0$ неверна
$H_0$ принимается	$H_0$ верно принята	Ошибка второго рода
$H_0$ отвергается	Ошибка первого рода	$H_0$ верно отвергнута

Вероятность ошибки первого рода жёстко ограничивается малой величиной:

$$p = P(T(X^n) \geq t | H_0) = P(p \leq \alpha | H_0) \leq \alpha.$$

Вероятность ошибки второго рода минимизируется путём выбора достаточно мощного критерия.

## Математическая постановка

данные:  $\mathbf{X} = \{X_1^{n_1}, \dots, X_m^{n_m}\}$ ,  $X_i^{n_i} \sim P_i \in \Omega_i$ ;  
 нулевые гипотезы:  $H_i: P_i \in \omega_i$ ,  $\omega_i \in \Omega_i$ ;  
 альтернативы:  $H'_i: P_i \notin \omega_i$ ;  
 статистики:  $T_i = T(X_i^{n_i})$  проверяет  $H_i$  против  $H'_i$ ;  
 реализации статистик:  $t_i = T(x_i^{n_i})$ ;  
 достигаемые уровни значимости:  $p_i = p(x_i^{n_i})$ ,  $i = 1, \dots, m$ ;

$$\mathbf{M} = \{1, 2, \dots, m\};$$

$\mathbf{M}_0 = \mathbf{M}_0(P) = \{i: H_i \text{ верна}\}$  — индексы верных гипотез,  $|\mathbf{M}_0| = m_0$ ;

$\mathbf{R} = \mathbf{R}(P, \alpha) = \{i: H_i \text{ отвергнута}\}$  — индексы отвергаемых гипотез,  
 $|\mathbf{R}| = R$ ;

$V = |\mathbf{M}_0 \cap \mathbf{R}|$  — число ошибок первого рода.

	Число верных $H_i$	Число неверных $H_i$	Всего
Число принятых $H_i$	$U$	$T$	$m - R$
Число отвергнутых $H_i$	$V$	$S$	$R$
Всего	$m_0$	$m - m_0$	$m$

## Многомерные обобщения ошибки первого рода

**Групповая вероятность ошибки** первого рода (familywise error rate):

$$FWER = P(V \geq 1).$$

Контроль над групповой вероятностью ошибки на уровне  $\alpha$  означает

$$FWER = P(V \geq 1) \leq \alpha \quad \forall P.$$

Как этого добиться?

Параметры  $\alpha_1, \dots, \alpha_m$  — уровни значимости, на которых необходимо проверять гипотезы  $H_1, \dots, H_m$ ; задача — выбрать их так, чтобы обеспечить  $FWER \leq \alpha$ .



# Поправка Бонферрони

Метод Бонферрони:

$$\alpha_1 = \dots = \alpha_m = \alpha/m.$$

## Теорема

Если гипотезы  $H_i$ ,  $i = 1, \dots, m$ , отвергаются при  $p_i \leq \alpha/m$ , то  $FWER \leq \alpha$ .

## Доказательство.

$$\begin{aligned} FWER = P(V \geq 1) &\leq P\left(\bigcap_{i=1}^{m_0} \{p_i \leq \alpha/m\}\right) \leq \sum_{i=1}^{m_0} P(p_i \leq \alpha/m) \leq \\ &\leq \sum_{i=1}^{m_0} \alpha/m = \frac{m_0}{m} \alpha \leq \alpha. \end{aligned}$$



# Поправка Бонферрони

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

$$\tilde{p}_i = \min(1, mp_i).$$

# Поправка Бонферрони

При увеличении  $m$  в результате применения поправки Бонферрони мощность статистической процедуры резко уменьшается — шансы отклонить неверные гипотезы падают.

Пример: критерий Стьюдента для независимых выборок,  $X_1^n \sim N(\mu_1, 1)$ ,  $X_2^n \sim N(\mu_2, 1)$ ,  $H_0: \mathbb{E}X_{1i} = \mathbb{E}X_{2j}$ ,  $H_1: \mathbb{E}X_{1i} \neq \mathbb{E}X_{2j}$ ,  $\mu_1 - \mu_2 = 1$ :

$m$	$n$	Мощность
1	23	0.9
10	23	0.67
100	23	0.37
1000	23	0.16
1000	62	0.9

Если проверяется одновременно 1000000 гипотез, при размере выборок  $n = 10$  мощность 0.9 достигается при расстоянии между средними выборок в пять стандартных отклонений.

## Модельный эксперимент

$$n = 20, \quad m = 200, \quad m_0 = 150;$$

$$X_{ij} \sim N(0, 1), \quad i = 1, \dots, m_0, \quad j = 1, \dots, n;$$

$$X_{ij} \sim N(1, 1), \quad i = m_0 + 1, \dots, m, \quad j = 1, \dots, n;$$

$$H_i: \mathbb{E}X_{ij} = 0, \quad H'_i: \mathbb{E}X_{ij} \neq 0;$$

для проверки используем одновыборочный критерий Стьюдента.

Без поправок:

	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	142	0	142
Отвергнутых $H_i$	8	50	58
Всего	150	50	200

Бонферрони:

	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	27	177
Отвергнутых $H_i$	0	23	23
Всего	150	50	200

# Нисходящие методы множественной проверки гипотез

Составим вариационный ряд достигаемых уровней значимости:

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)},$$

$H_{(1)}, H_{(2)}, \dots, H_{(m)}$  — соответствующие гипотезы.

- 1 Если  $p_{(1)} \geq \alpha_1$ , принять все нулевые гипотезы  $H_{(1)}, H_{(2)}, \dots, H_{(m)}$  и остановиться; иначе отвергнуть  $H_{(1)}$  и продолжить.
- 2 Если  $p_{(2)} \geq \alpha_2$ , принять все нулевые гипотезы  $H_{(2)}, H_{(3)}, \dots, H_{(m)}$  и остановиться; иначе отвергнуть  $H_{(2)}$  и продолжить.
- 3 ...

Каждый достигаемый уровень значимости  $p_{(i)}$  сравнивается со своим уровнем значимости  $\alpha_i$ .

# Метод Холма

**Метод Холма** — нисходящая процедура со следующими уровнями значимости:

$$\alpha_1 = \frac{\alpha}{m}, \alpha_2 = \frac{\alpha}{m-1}, \dots, \alpha_i = \frac{\alpha}{m-i+1}, \dots, \alpha_m = \alpha.$$

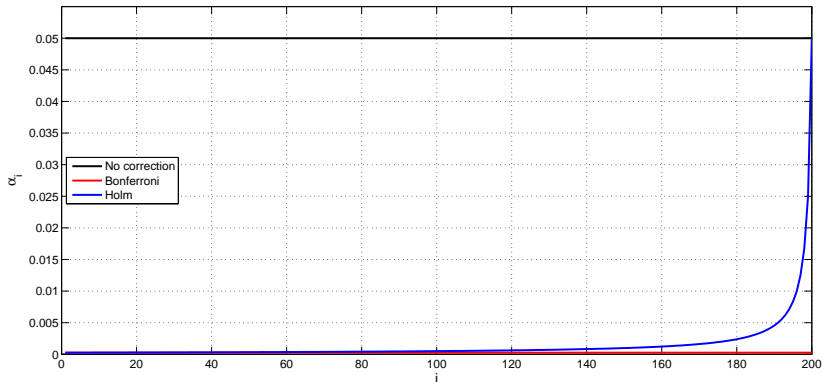
Метод обеспечивает контроль над  $FWER$  на уровне  $\alpha$  при любых  $p_i$  и  $T_i$ .

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

$$\tilde{p}_{(i)} = \min \left( 1, \max \left( (m-i+1) p_{(i)}, \tilde{p}_{(i-1)} \right) \right).$$

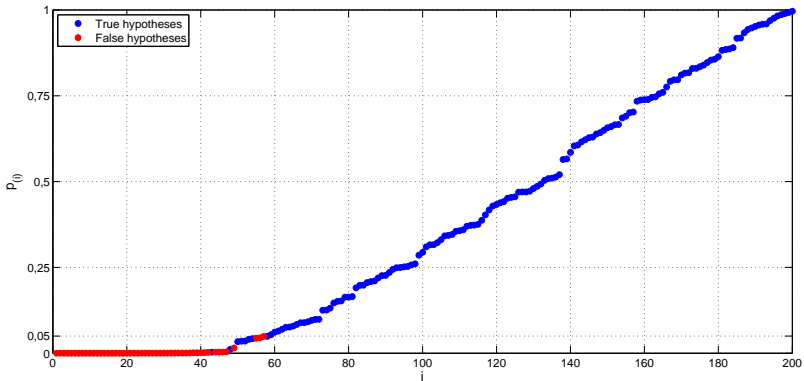
## Метод Холма

Метод Холма равномерно мощнее поправки Бонферрони, поскольку все его уровни значимости  $\alpha_i$  не меньше:



## Модельный эксперимент

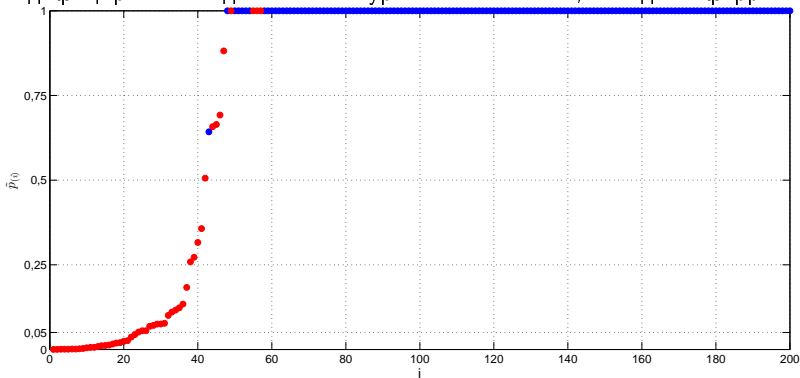
Отсортированные достигаемые уровни значимости:





# Модельный эксперимент

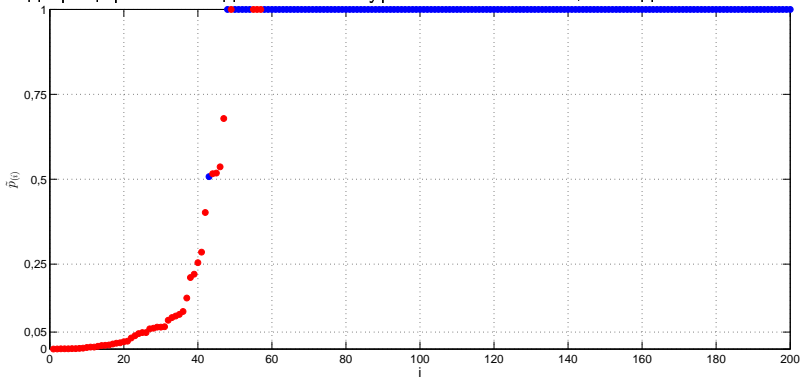
Модифицированные достигаемые уровни значимости, метод Бонферрони:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	27	177
Отвергнутых $H_i$	0	23	23
Всего	150	50	200

# Модельный эксперимент

Модифицированные достигаемые уровни значимости, метод Холма:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	24	174
Отвергнутых $H_i$	0	26	26
Всего	150	50	200

## Идеи для дальнейших улучшений

- Дополнительно оценить  $m_0$ .
- Сделать дополнительные предположения:
  - о характере зависимости между статистиками;
  - о совместном распределении статистик.
- Учесть зависимость между статистиками с помощью перестановочных методов.

## Предварительное оценивание $m_0$

Из доказательства теоремы 1 следует, что метод Бонферрони контролирует  $FWER$  на уровне  $\frac{m_0}{m}\alpha$ .

Примеры методов оценки  $m_0$ :

- метод Стори:

$$\hat{m}_0(\lambda) = \frac{1}{1-\lambda} \left( 1 + \sum_{i=1}^m [p_i > \lambda] \right), \quad \lambda \in [0, 1);$$

- метод наименьшего наклона Бенджамини-Хохберга:

$$\hat{m}_0 = \min \left( m, \frac{1}{S_{i_0}} + 1 \right),$$
$$S_i = \frac{1 - p_{(i)}}{m - i + 1}, \quad i = 1, \dots, m,$$
$$i_0 = \min \{i : S_i < S_{i-1}\}.$$

Как правило, доказательств контроля  $FWER$  для процедур с предварительной оценкой  $m_0$  нет, но на практике они часто работают хорошо.

## Одношаговый метод Шидака

Метод Шидака:

$$\alpha_1 = \alpha_2 = \dots = \alpha_m = 1 - (1 - \alpha)^{\frac{1}{m}}.$$

Метод обеспечивает контроль над  $FWER$  на уровне  $\alpha$  при условии, что статистики  $T_i$  **независимы** или выполняется следующее свойство:

$$P(T_1 \leq t_1, \dots, T_m \leq t_m) \geq \prod_{i=1}^m P(T_i \leq t_i) \quad \forall t_1, \dots, t_m$$

(**positive orthant dependence**).

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

$$\tilde{p}_i = 1 - (1 - p_i)^m.$$

## Нисходящая модификация

Нисходящий метод Шидака (метод Шидака-Холма) — нисходящая процедура со следующими уровнями значимости:

$$\alpha_1 = 1 - (1 - \alpha)^{\frac{1}{m}}, \dots, \alpha_i = 1 - (1 - \alpha)^{\frac{1}{m-i+1}}, \dots, \alpha_m = \alpha.$$

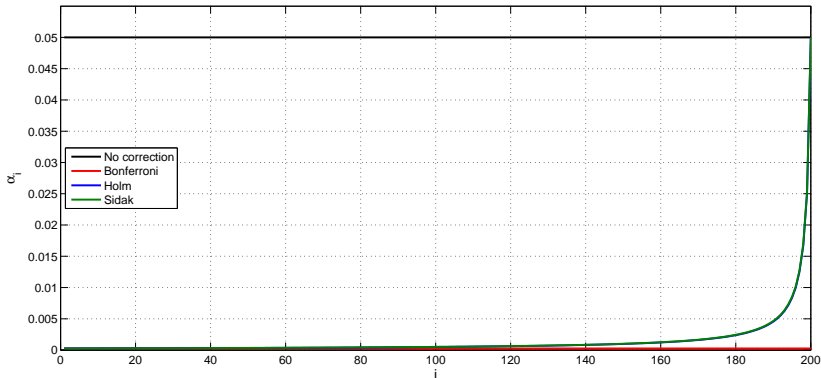
Метод обеспечивает контроль над  $FWER$  на уровне  $\alpha$  при условии, что статистики  $T_i$  **независимы** или для них выполняется свойство **positive orthant dependence**.

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

$$\tilde{p}_{(i)} = \max \left( 1 - (1 - p_{(i)})^{(m-i+1)}, \tilde{p}_{(i-1)} \right).$$

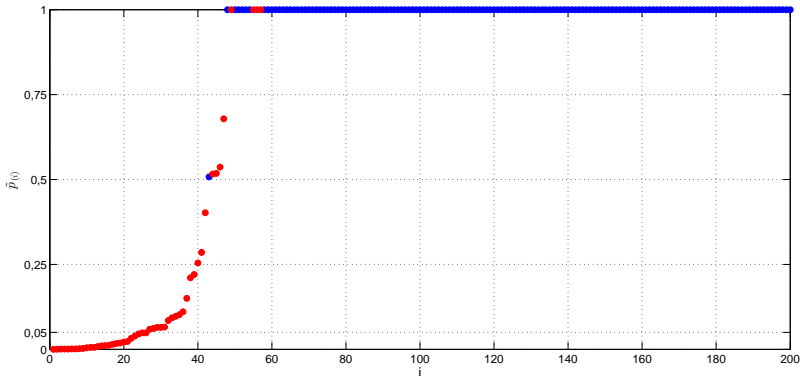
# Нисходящая модификация

На практике при достаточно больших  $m$  не слишком отличается от метода Холма:



# Модельный эксперимент

Модифицированные достигаемые уровни значимости, метод Холма:

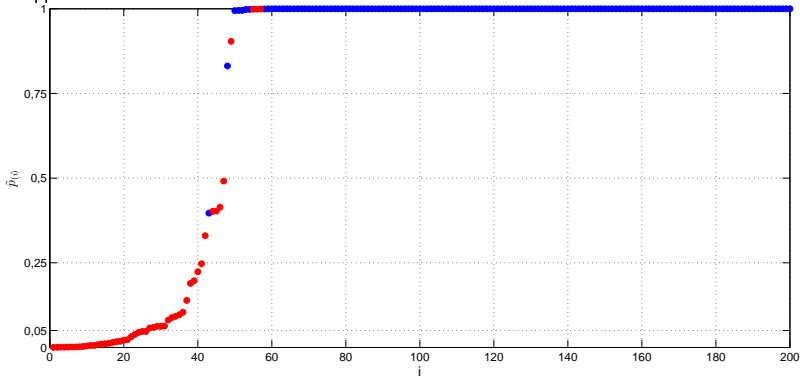


	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	24	174
Отвергнутых $H_i$	0	26	26
Всего	150	50	200



# Модельный эксперимент

Модифицированные достигаемые уровни значимости, нисходящий метод Шидака:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	24	174
Отвергнутых $H_i$	0	26	26
Всего	150	50	200

## Зависимость между статистиками

- Не учитывая характер зависимости между статистиками, нельзя построить контролирующую *FWER* процедуру мощнее, чем метод Холма.
- Если статистики независимы, нельзя построить контролирующую *FWER* процедуру мощнее, чем метод Шидака-Холма.
- Чем сильнее связь между статистиками, тем меньше нужно модифицировать уровни значимости.

Для построения мощной процедуры множественной проверки гипотез необходимо учесть структуру зависимости статистик.

# Параметрические методы

Если совместное нулевое распределение статистик  $T_1, \dots, T_m$  известно, константы  $\alpha_i$  могут быть так, что контроль над  $FWER$  будет точным ( $FWER = \alpha$ ).

## Примеры:

- метод HSD Тьюки для попарных сравнений нормально распределённых выборок друг с другом;
- критерий Даннета для сравнения средних  $m$  нормально распределённых выборок со средним контрольной выборки.

## Перестановочные методы

Неявно учесть зависимости между статистиками можно при помощи перестановочных методов. Подробнее: Bretz, раздел 5.1.

Методы обеспечивает контроль над  $FWER$  на уровне  $\alpha$  при условии выполнения свойства **subset pivotality**:

$$P\left(\bigcap_{i \in M^*} \{T_i \geq t^*\} \mid \bigcap_{i \in M^*} H_i\right) = P\left(\bigcap_{i \in M^*} \{T_i \geq t^*\} \mid \bigcap_{i \in M} H_i\right) \quad \forall M^* \in M$$

(нулевое распределение любого подмножества статистик  $T_i$  не зависит от того, верны или неверны соответствующие оставшимся статистикам гипотезы).

## Многомерные обобщения ошибки первого рода

Ожидаемая доля ложных отклонений гипотез (false discovery rate):

$$FDR = \mathbb{E} \left( \frac{V}{R} [R > 0] \right).$$

Контроль над ожидаемой долей ложных отклонений на уровне  $\alpha$  означает

$$FDR = \mathbb{E} \left( \frac{V}{R} [R > 0] \right) \leq \alpha \quad \forall P.$$

Для любой процедуры множественной проверки гипотез  $FDR \leq FWER$ .

## Восходящие методы множественной проверки гипотез

Составим вариационный ряд достигаемых уровней значимости:

$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)},$$

$H_{(1)}, H_{(2)}, \dots, H_{(m)}$  — соответствующие гипотезы.

- 1 Если  $p_{(m)} \leq \alpha_m$ , отвергнуть все нулевые гипотезы  $H_{(1)}, H_{(2)}, \dots, H_{(m)}$  и остановиться; иначе принять  $H_{(m)}$  и продолжить.
- 2 Если  $p_{(m-1)} \leq \alpha_{m-1}$ , отвергнуть все нулевые гипотезы  $H_{(1)}, H_{(2)}, \dots, H_{(m-1)}$  и остановиться; иначе принять  $H_{(m-1)}$  и продолжить.
- 3 ...

Каждый достигаемый уровень значимости  $p_{(i)}$  сравнивается со своим уровнем значимости  $\alpha_i$ .

## Метод Бенджамини-Хохберга

Метод Бенджамини-Хохберга — восходящая процедура со следующими уровнями значимости:

$$\alpha_1 = \frac{\alpha}{m}, \dots, \alpha_i = \frac{\alpha i}{m}, \dots, \alpha_m = \alpha.$$

Метод обеспечивает контроль над  $FDR$  на уровне  $\alpha$  при условии, что статистики  $T_i$  **независимы** или выполняется следующее свойство:

$$P(X \in D | T_i = x) \text{ неубывает по } x \quad \forall i \in M_0,$$

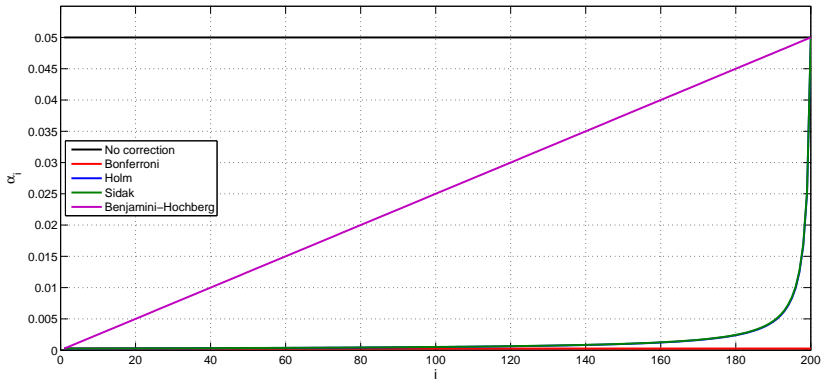
где  $D$  — произвольное возрастающее множество, то есть, такое, что из  $x \in D$  и  $y \geq x$  следует  $y \in D$ .

(**PRDS on**  $T_i, i \in M_0$  (positive regression dependency on each one from a subset)).

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

$$\tilde{p}_{(i)} = \min \left( 1, \frac{mp_{(i)}}{i}, \tilde{p}_{(i+1)} \right).$$

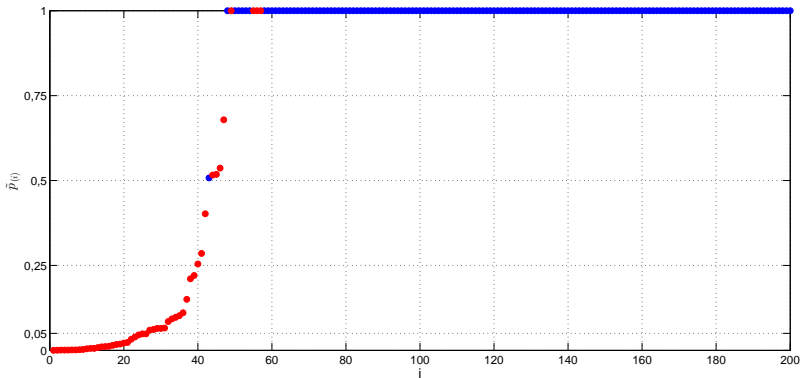
# Метод Бенджамини-Хохберга





# Модельный эксперимент

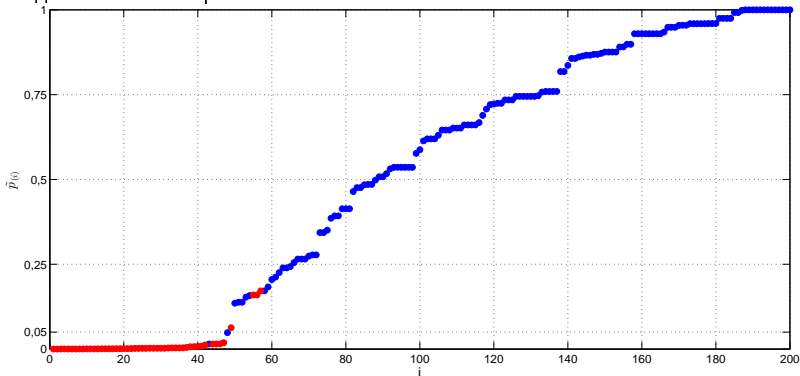
Модифицированные достигаемые уровни значимости, метод Холма:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	24	174
Отвергнутых $H_i$	0	26	26
Всего	150	50	200

# Модельный эксперимент

Модифицированные достигаемые уровни значимости, нисходящий метод Бенджамини-Хохберга:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	148	4	152
Отвергнутых $H_i$	2	46	48
Всего	150	50	200

# Метод Бенджамини-Иекутиели

Метод Бенджамини-Иекутиели — восходящая процедура со следующими уровнями значимости:

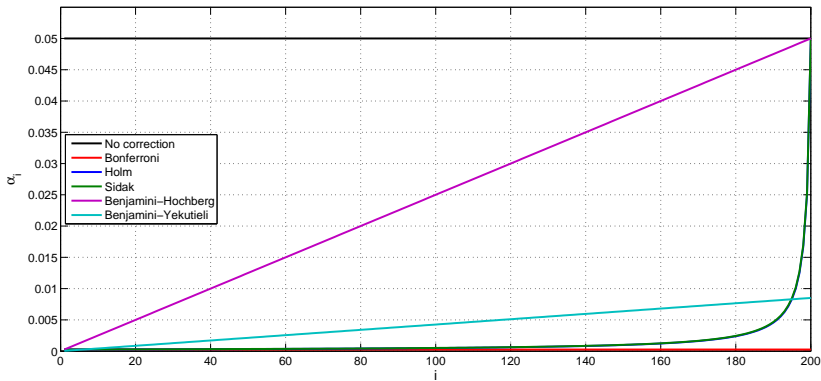
$$\alpha_1 = \frac{\alpha}{m \sum_{j=1}^m \frac{1}{j}}, \dots, \alpha_i = \frac{\alpha i}{m \sum_{j=1}^m \frac{1}{j}}, \dots, \alpha_m = \frac{\alpha}{\sum_{j=1}^m \frac{1}{j}}.$$

Метод обеспечивает контроль над  $FDR$  на уровне  $\frac{m_0}{m} \alpha \leq \alpha$  при любых  $p_i$  и  $T_i$ .

Альтернативный вид — переход к модифицированным достигаемым уровням значимости:

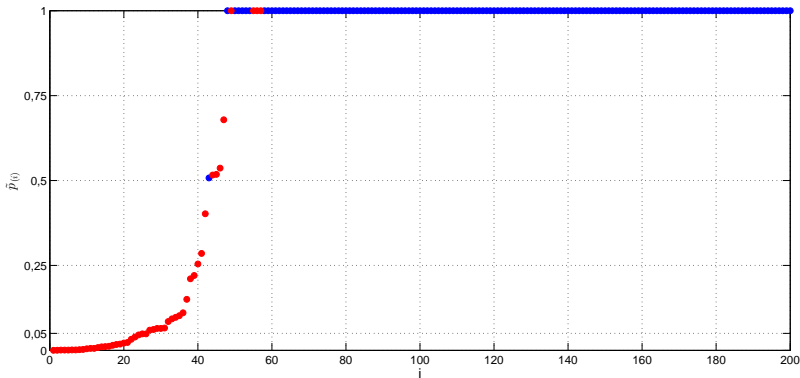
$$\tilde{p}_{(i)} = \min \left( 1, \frac{mp_{(i)} \sum_{j=1}^m \frac{1}{j}}{i}, \tilde{p}_{(i+1)} \right).$$

# Метод Бенджамини-Иекутиели



# Модельный эксперимент

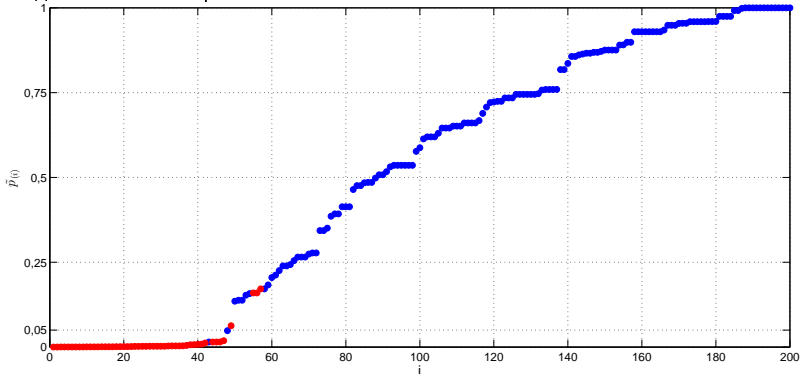
Модифицированные достигаемые уровни значимости, метод Холма:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	24	174
Отвергнутых $H_i$	0	26	26
Всего	150	50	200

# Модельный эксперимент

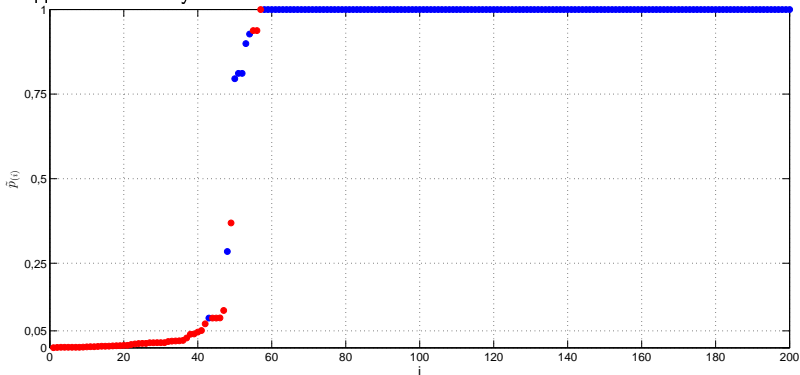
Модифицированные достигаемые уровни значимости, нисходящий метод Бенджамини-Хохберга:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	148	4	152
Отвергнутых $H_i$	2	46	48
Всего	150	50	200

# Модельный эксперимент

Модифицированные достигаемые уровни значимости, нисходящий метод Бенджамини-Иекутиели:



	Верных $H_i$	Неверных $H_i$	Всего
Принятых $H_i$	150	10	160
Отвергнутых $H_i$	0	40	40
Всего	150	50	200

## Мутации

	Контроль (100)	Больные (100)	$p$
Мутация	1 из 100	8 из 100	0.0349
Фамилия начинается с гласной	36 из 100	40 из 100	0.6622

Бонферрони, Холм:  $p_1$  сравнивается с  $\frac{0.05}{2} = 0.025$

Шидак:  $p_1$  сравнивается с  $1 - (1 - 0.05)^{\frac{1}{2}} \approx 0.02532$



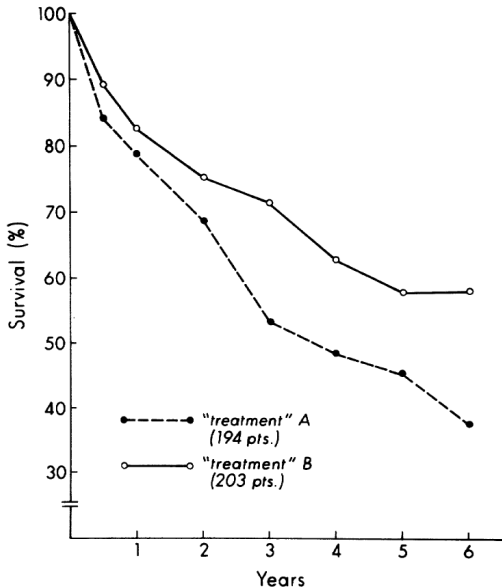
## Подгруппы

Lee et al., Clinical judgment and statistics. Lessons from a simulated randomized trial in coronary artery disease, 1980: 1073 пациента с ишемической болезнью сердца были искусственно разделены на две случайные группы, лечение в двух группах проходило одинаково. Исследовалась выживаемость пациентов.

Важными факторами, влияющими на выживаемость, являются число поражённых артерий (одна, две, три) и тип сокращений левого желудочка (нормальный, абнормальный).

Для одной из шести подгрупп по этим уровням фактора были обнаружены значимые различия между в выживаемости пациентов первого и второго типов.

# Подгруппы



## Подгруппы

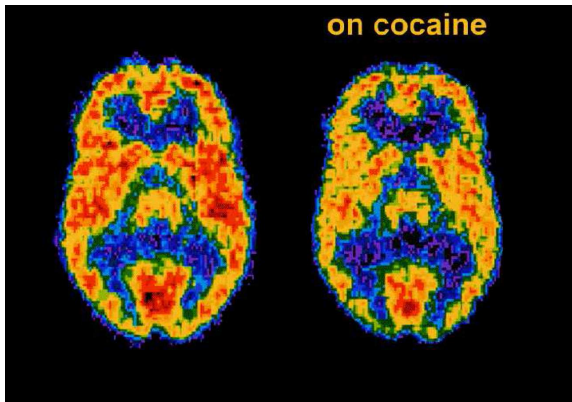
Ishitani, Lin, Caffeine consumption and the risk of breast cancer in a large prospective cohort of women, 2008: анализировалась связь потребления кофеина, кофе и чая и риска возникновения рака груди, всего около 50 подгрупп. Показано, что:

- употребление четырёх и более чашек кофе в день связано с увеличением риска злокачественного рака груди ( $p = 0.08$ );
- потребление кофеина связано с увеличением риска возникновения эстроген- и прогестерон-независимых опухолей и опухолей больше 2 см ( $p = 0.02$  и  $p = 0.02$ );
- потребление кофе без кофеина связано со снижением риска рака груди у женщин в постменопаузе, принимающих гормоны ( $p = 0.02$ ).

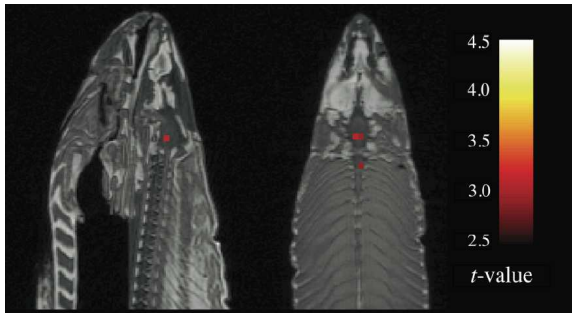
См. также:

- <http://youtu.be/QysrgLXMPwA>
- <http://xkcd.com/882/>
- <http://wmbriggs.com/blog/?p=9308>

# SPM



# SPM



## Литература

- попроще — Bretz;
- посложнее — Dickhaus;
- перестановочные методы (permutation methods) — Westfall, 2008, и другие работы этого автора;
- positive orthant dependence — Holland, 1987;
- subset pivotality — Westfall, 2008;
- PRDS — Benjamini, 2001;
- SPM — Nichols, 2003.

Bretz F., Hothorn T., Westfall P. *Multiple Comparisons Using R*. — Boca Raton: Chapman and Hall/CRC, 2010.

Dickhaus T. *Simultaneous Statistical Inference With Applications in the Life Sciences*. — Heidelberg: Springer, 2014.

Westfall P., Troendle J. (2008). *Multiple testing with minimal assumptions*. Biometrical Journal, 50(5), 745–755.

Holland B.S., Copenhaver M.D. (1987). *An Improved Sequentially Rejective Bonferroni Test Procedure*. Biometrics, 43(2), 417–423.

Benjamini Y., Yekutieli D. (2001). *The control of the false discovery rate in multiple testing under dependency*. Annals of Statistics, 29(4), 1165–1188.

Nichols T.E., Hayasaka, S. (2003). *Controlling the familywise error rate in functional neuroimaging: a comparative review*. Statistical Methods in Medical Research, 12(5), 419–46.