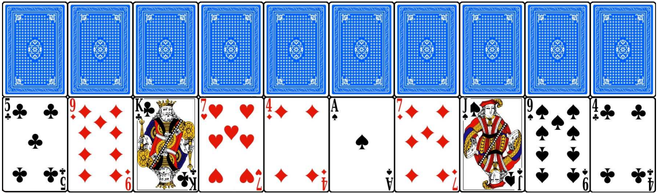
Прикладная статистика 6. Множественная проверка гипотез.

18 марта 2013 г.

Поиск экстрасенсов

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



H_0: испытуемый выбирает цвет карт наугад.
H_1: испытуемый может предсказывать цвет карт.
Статистика t — число угаданных цветов.

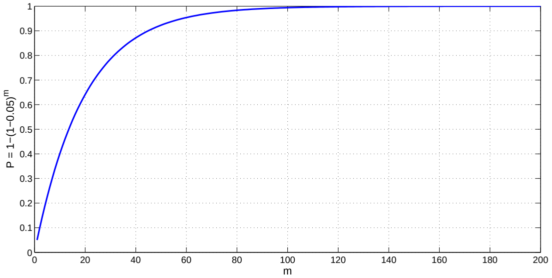
$$P\left(t \geq 9 \,|\, H_0\right) = 11 \cdot \frac{1}{2}^{10} = 0.0107421875,$$

т.е. при $t = 9$ достигаемый уровень значимости $p \approx 0.01$, событие достаточно редкое, можно отклонять H_0 и признавать испытуемого экстрасенсом.

Поиск экстрасенсов

Процедуру отбора прошли 1000 человек.
Девять из них угадали цвета 9 из 10 карт, двое — цвета всех 10 карт.
Ни один из них 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

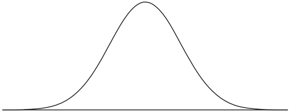
выборка: $\mathbf{X} = \{X_1, \dots, X_n\} \sim P \in \Omega$;

нулевая гипотеза: $H_0 \colon P \in \omega, \quad \omega \in \Omega$;

альтернатива: $H_1 \colon P \notin \omega$;

статистика: $T(\mathbf{X}), \quad T(\mathbf{X}) \sim F(x)$ при $P \in \omega$;

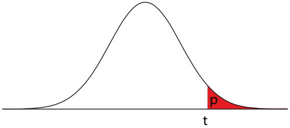
$T(\mathbf{X}) \not\sim F(x)$ при $P \notin \omega$;



реализация выборки: $\mathbf{x} = \{x_1, \dots, x_n\}$;

реализация статистики: $t = T(\mathbf{x})$;

достигаемый уровень значимости: $p(\mathbf{x})$— вероятность при H_0 получить $T(\mathbf{X}) = t$ или ещё более экстремальное;



Гипотеза отвергается при $p(\mathbf{x}) \le \alpha, \quad \alpha$— уровень значимости.

Правило проверки гипотезы

Несимметричность задачи проверки гипотез

	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 достаточно малой наперёд заданной величиной — $P\left(p\left(\mathbf{x}\right) \leq \alpha \,|H_0\,\right) \leq \alpha$.

Вероятность ошибки второго рода минимизируется путём выбора достаточно мощного критерия.

данные: $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_m\} \sim P \in \Omega$;
нулевые гипотезы: $H_i\colon P \in \omega_i, \quad \omega_i \in \Omega$;
альтернативы: $H_i'\colon P \notin \omega_i$;
статистики: $T_i = T(\mathbf{X}_i)$ проверяет H_i против H_i';
реализации статистики: $t_i = t(\mathbf{x}_i)$;
достигаемые уровни значимости: $p_i = p(\mathbf{x}_i), \quad i = 1, \ldots, m$;

$\mathbf{M} = \{1, 2, \ldots, m\}$;
$H_0 = \bigcap\limits_{i \in \mathbf{M}} H_i$ — полная нулевая гипотеза;
$\mathbf{M_0} = \mathbf{M_0}(P) = \{i\colon H_i \text{ верна}\}$ — индексы верных гипотез, $|\mathbf{M_0}| = m_0$;
$\mathbf{R} = \mathbf{R}(P, \alpha) = \{i\colon H_i \text{ отвергнута}\}$ — индексы отвергаемых гипотез,
$|\mathbf{R}| = R$;
$V = |\mathbf{M_0} \cap \mathbf{R}|$ — число ошибок первого рода.

	Число верных H_0	Число неверных H_0	Всего
Число принятых H_0	U	T	$m - R$
Число отвергнутых H_0	V	S	R
Всего	m_0	$m - m_0$	m

Многомерные обобщения ошибки первого рода

Групповая вероятность ошибки (первого рода):

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

Контроль над групповой вероятностью ошибки на уровне α означает

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

Ожидаемая доля ложных отклонений гипотез (среди всех отклонений):

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

Контроль над ожидаемой долей ложных отклонений на уровне α означает

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

Виды контроля

Мера числа ошибок первого рода зависит от неизвестного множества верных нулевых гипотез $\mathbf{M_0}$:

$$FWER = P\left(V \geq 1 \middle| \bigcap_{j \in \mathbf{M_0}} H_j\right) =$$

$$= P\left(\text{отвергнуть хотя бы одну } H_j \text{ из } \bigcap_{j \in \mathbf{M_0}} H_j \middle| \bigcap_{j \in \mathbf{M_0}} H_j\right)$$

Точный контроль:

$$P\left(V \geq 1 \middle| \bigcap_{j \in \mathbf{M_0}} H_j\right) \leq \alpha.$$

Слабый контроль:

$$P\left(V \geq 1 \middle| \bigcap_{j \in \mathbf{M}} H_j\right) \leq \alpha.$$

Сильный контроль:

$$\max_{M_0^* \subseteq M} P\left(V \geq 1 \middle| \bigcap_{j \in \mathbf{M_0^*}} H_j\right) \leq \alpha.$$

Поправка Бонферрони

Теорема

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

Доказательство.

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

$$\leq \sum_{i=1}^{m_0} \alpha/m = m_0\alpha/m \leq \alpha.$$

□

Сильный контроль над FWER на уровне α обеспечивается при любых p_i.

Поправка Бонферрони

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

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

Поправка Бонферрони

При увеличении m в результате применения поправки Бонферрони мощность статистической процедуры резко уменьшается — шансов отклонить хотя бы одну неверную гипотезу практически не остаётся.

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

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

Эксперимент:
$n = 15, m = 200, m_0 = 150; X_{ij} \sim N(0,1), i = 1, \dots, m_0,$
$X_{ij} \sim N(1,1), i = m_0 + 1, \dots, m, j = 1, \dots, n.$

Без поправок:

	Верные H_0	Неверные H_0	Всего
Принятые H_0	141	2	143
Отвергнутые H_0	9	48	57
Всего	150	50	200

Бонферрони:

	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Нисходящая процедура

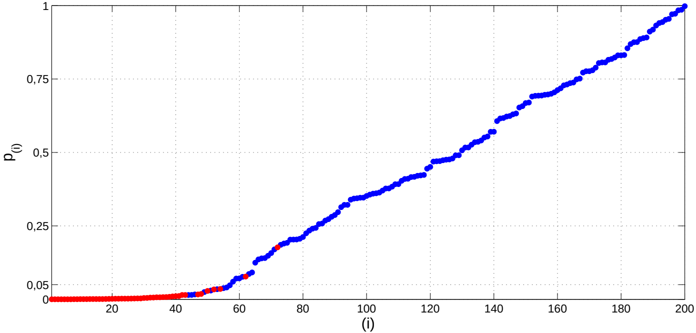
Составим вариационный ряд достигаемых уровней значимости:
$p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$,
$H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ — соответствующие нулевые гипотезы.

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

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

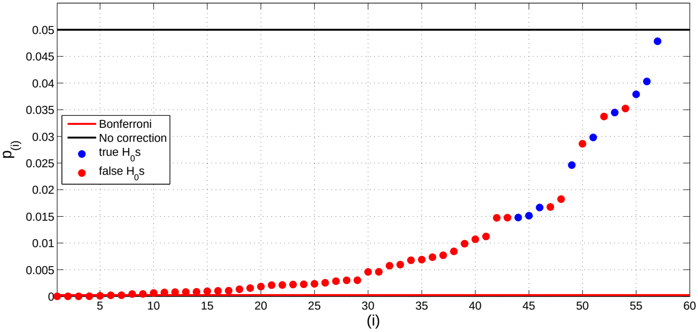
Модельный эксперимент

Отсортированные достигаемые уровни значимости:



Модельный эксперимент

Отсортированные достигаемые уровни значимости:



Метод Холма

Метод Холма — нисходящая процедура со следующими уровнями значимости:

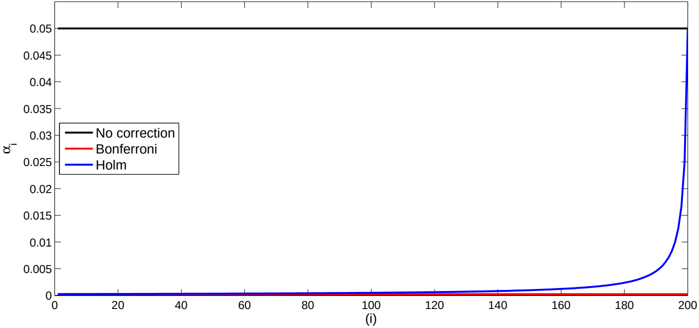
$$\alpha_1 = \frac{\alpha}{m}, \alpha_2 = \frac{\alpha}{m-1}, \dots, \alpha_i = \frac{\alpha}{m-i+1}, \dots, \alpha_m = \alpha.$$

Сильный контроль над $FWER$ на уровне α обеспечивается при любых p_i, без ограничений на характер зависимости между ними.

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

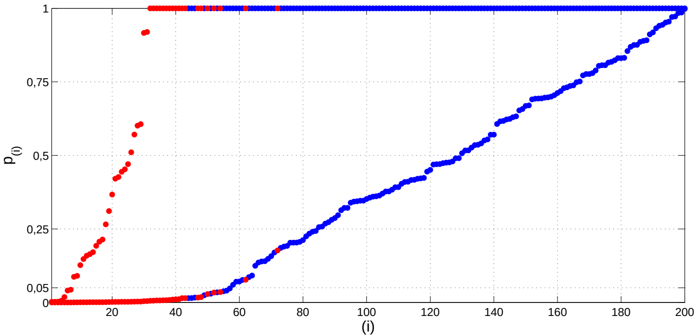
$$\tilde{p}_{(i)} = \min\left((m-i+1)\, p_{(i)}, 1\right).$$

Метод Холма


0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
α_i
No correction
Bonferroni
Holm
0
20
40
60
80
100
120
140
160
180
200
(i)

Модельный эксперимент

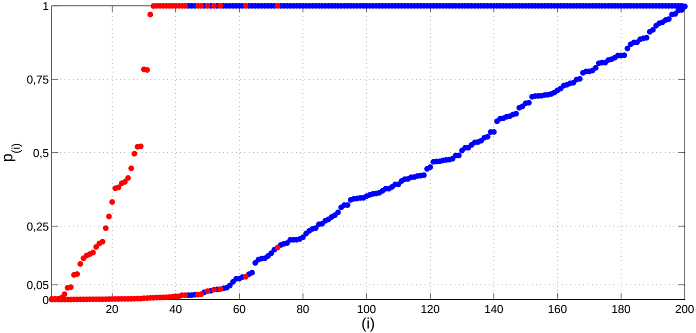
Модифицированные достигаемые уровни значимости: метод Бонферрони



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Модельный эксперимент

Модифицированные достигаемые уровни значимости: метод Холма



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Улучшения

Если совместное распределение статистик $T_1, \ldots, T_m$ известно, константы α_i могут быть найдены точно:

$$P_{0,\ldots,0}\left(\max\left(T_1, \ldots, T_i\right) \geq c_{m-i+1}\right) = \alpha.$$

Если полной информации о распределении статистик нет, можно использовать дополнительные предположения.

Метод Шидака

Метод Шидака — нисходящая процедура со следующими уровнями значимости:

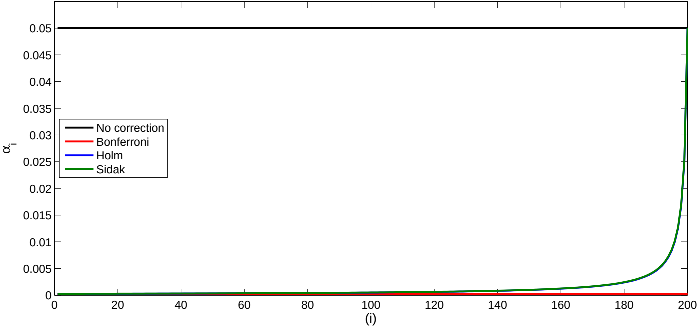
$$\alpha_1 = 1 - (1-\alpha)^{\frac{1}{m}}, \dots, \alpha_i = 1 - (1-\alpha)^{\frac{1}{m-i+1}}, \dots, \alpha_m = \alpha.$$

Сильный контроль над $FWER$ на уровне α обеспечивается при независимых p_i.

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

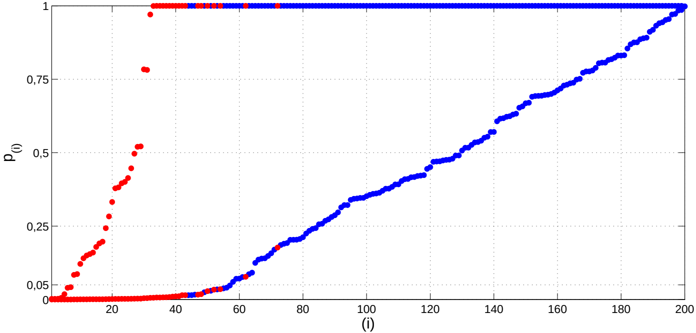
$$\tilde{p}_{(i)} = \min\left(\frac{p_{(i)}\alpha}{\left(1-(1-\alpha)^{\frac{1}{m-i+1}}\right)}, 1\right).$$

Метод Шидака



Модельный эксперимент

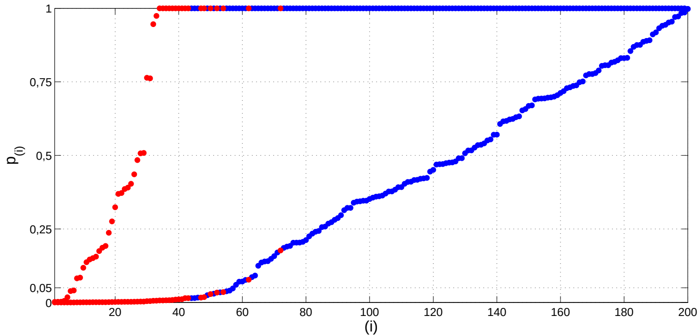
Модифицированные достигаемые уровни значимости: метод Холма



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Модельный эксперимент

Модифицированные достигаемые уровни значимости: метод Шидака



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Положительная зависимость

Будем говорить, что между достигаемыми уровнями значимости $p_1, \ldots, p_m$ положительная зависимость, если

$$P\left(p_1 \leq x_1, \ldots, p_m \leq x_m\right) \geq \prod_{i=1}^{m} P\left(p_i \leq x_i\right).$$

Теорема

Если между статистиками положительная регрессионная зависимость, то наилучшие α_i лежат в диапазоне

$$1-(1-\alpha)^{\frac{1}{m-i+1}} \leq \alpha_i \leq \alpha.$$

Subset pivotality

Subset pivotality: нулевое распределение любого подмножества статистик T_i не зависит от того, верны или неверны соответствующие оставшимся статистикам гипотезы.

$$P\left(\bigcap_{i\in\mathbf{M}^*}\{T_i\geq t^*\}\middle|\bigcap_{i\in\mathbf{M}^*}H_i\right)=P\left(\bigcap_{i\in\mathbf{M}^*}\{T_i\geq t^*\}\middle|H_0\right)$$
$$\leq P\left(\bigcap_{i\in\mathbf{M}}\{T_i\geq t^*\}\middle|H_0\right)$$

Если выполняется это свойство, то сильный контроль над FWER можно обеспечить с использованием максимальной статистики.

Subset pivotality

$M_T = \max_i T_i$ — максимальная статистика.

$$\bigcap_i \{T_i \geq t^*\} = \{M_T \geq t^*\}.$$

Для обеспечения $FWER \leq \alpha$ при выполнении свойства subset pivotality достаточно знать распределение максимальной статистики при справедливости полной нулевой гипотезы $F_{M_T|H_0}(x)$:

$$t_\alpha \equiv F^{-1}_{M_T|H_0}(1-\alpha),$$

$$\begin{aligned} P\left(\bigcap_{i\in\mathbf{M}} \{T_i \geq t^*\} |H_0\right) &= P\left(M_T \geq t_\alpha |H_0\right) \\ &= 1 - F_{M_T|H_0}(t_\alpha) \\ &= \alpha. \end{aligned}$$

Метод Бенджамини-Хохберга

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

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

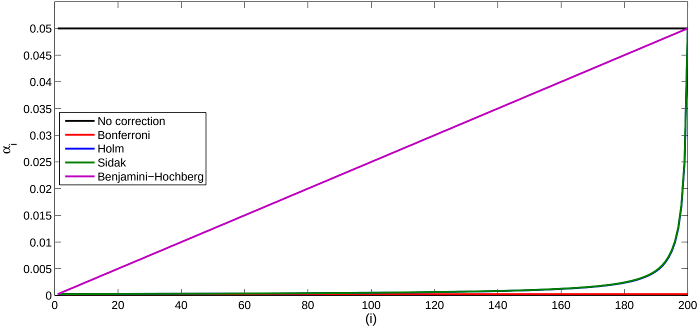
Если p_i независимы или положительно зависимы, то

$$FDR \leq \frac{m_0}{m}\alpha \leq \alpha.$$

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

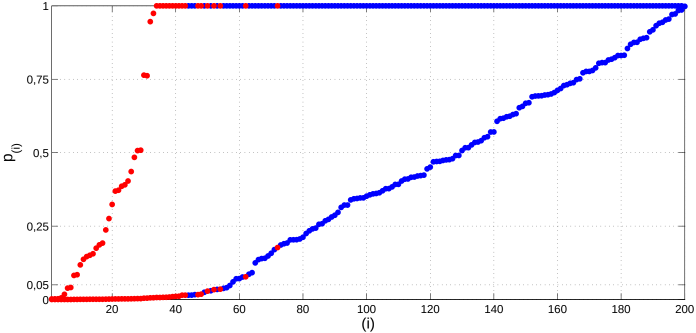
$$\tilde{p}_{(i)} = \min\left(\frac{mp_{(i)}}{i}, 1\right).$$

Метод Бенджамини-Хохберга



Модельный эксперимент

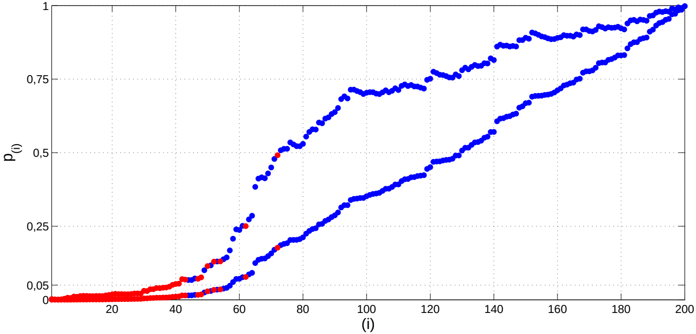
Модифицированные достигаемые уровни значимости: метод Шидака



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Модельный эксперимент

Модифицированные достигаемые уровни значимости: метод Бенджамини-Хохберга



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	12	162
Отвергнутые H_0	0	38	38
Всего	150	50	200

Метод Бенджамини-Иекутиели

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

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

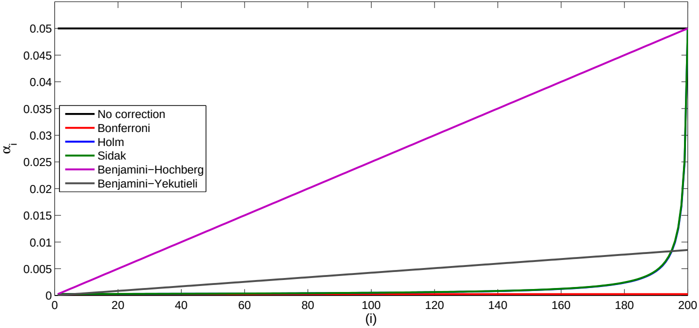
Для любых p_i выполняется

$$FDR \le \frac{m_0}{m}\alpha \le \alpha.$$

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

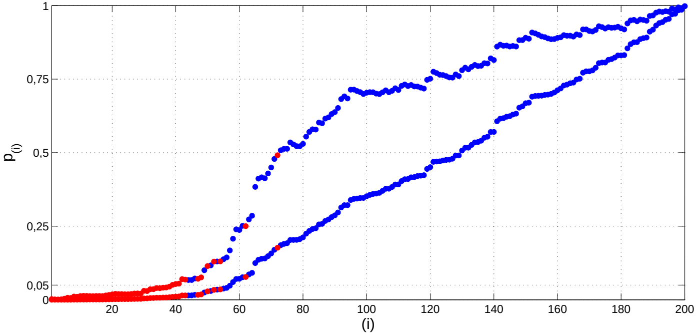
$$\tilde{p}_{(i)} = \min\left(\frac{mp_{(i)}\sum\limits_{i=1}^{m}\frac{1}{i}}{i}, 1\right).$$

Метод Бенджамини-Иекутиели



Модельный эксперимент

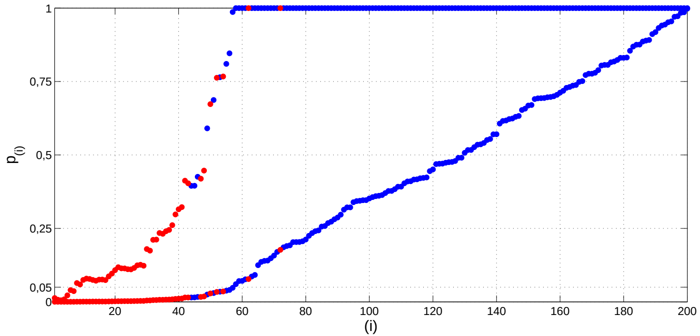
Модифицированные достигаемые уровни значимости: метод Бенджамини-Хохберга



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	24	174
Отвергнутые H_0	0	26	26
Всего	150	50	200

Модельный эксперимент

Модифицированные достигаемые уровни значимости: метод Бенджамини-Иекутиели



	Верные H_0	Неверные H_0	Всего
Принятые H_0	150	43	193
Отвергнутые H_0	0	7	7
Всего	150	50	200

Возможные улучшения

Оценка $m_0 \Rightarrow$ более точная оценка FDR.

Метод Стори:

$$m_0 = 2 \sum_{i=1}^{m} [p_i \geq 0.5] \, .$$

Двухэтапный метод Бенджамини-Хохберга: m_0 — число гипотез, не отвергаемых методом Бенджамини-Хохберга при контроле FDR на уровне $\alpha' = \frac{\alpha}{\alpha+1}$.

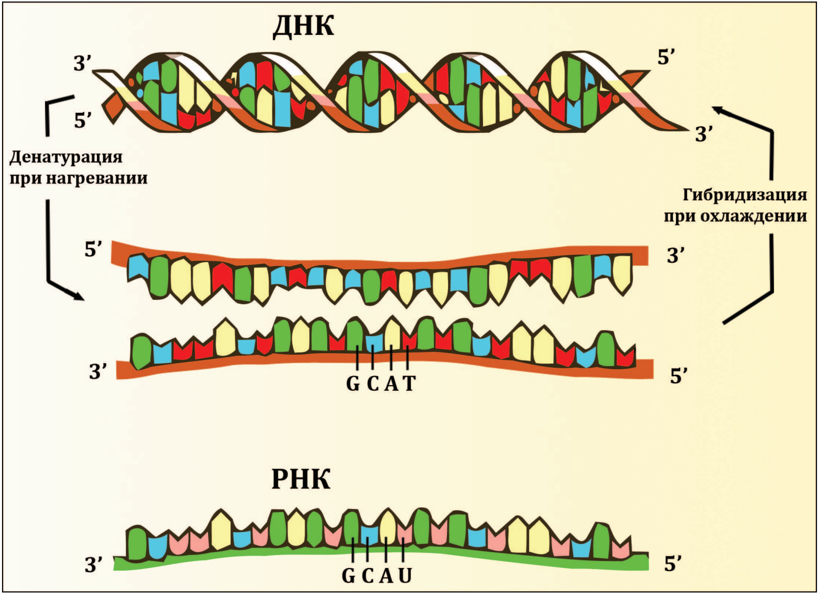
Мутации

	Контроль (100)	Больные (100)	p
Мутация	1 из 100	7 из 100	0.0349
Фамилия начинается с гласной	36 из 100	40 из 100	0.646

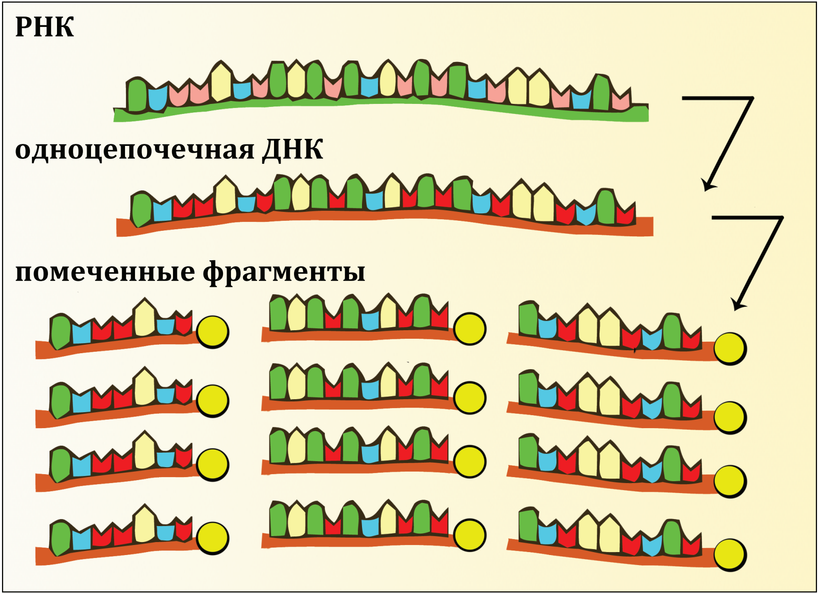
Бонферрони, Холм: p_1 сравнивается с $\frac{0.05}{2} = 0.025$

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

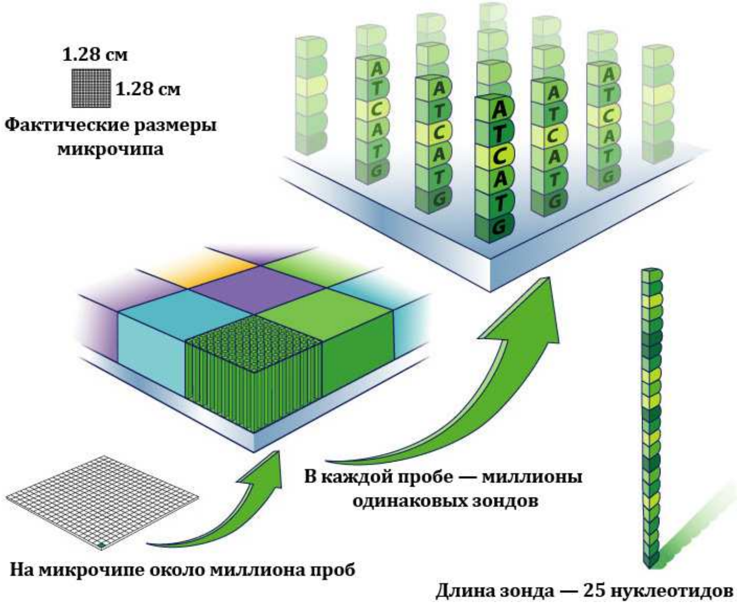
ДНК и РНК


ДНК
3’
5’
5’
3’
Денатурация
при нагревании
Гибридизация
при охлаждении
5’
3’
3’
5’
G C A T
РНК
3’
5’
G C A U

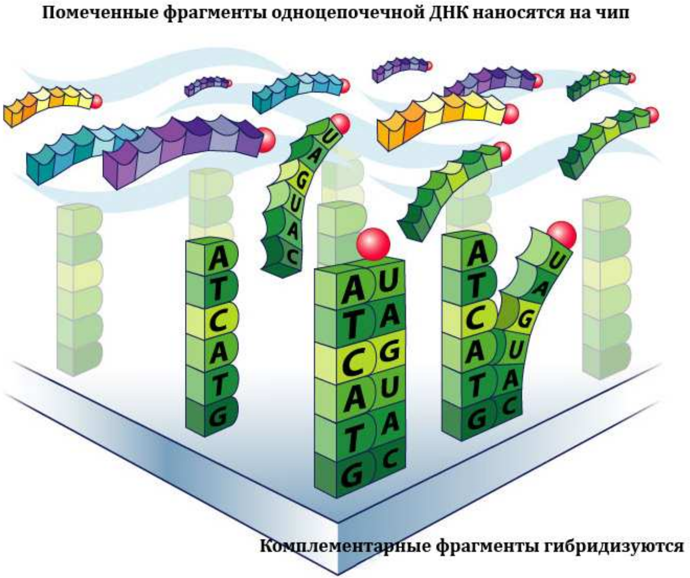
Ход эксперимента



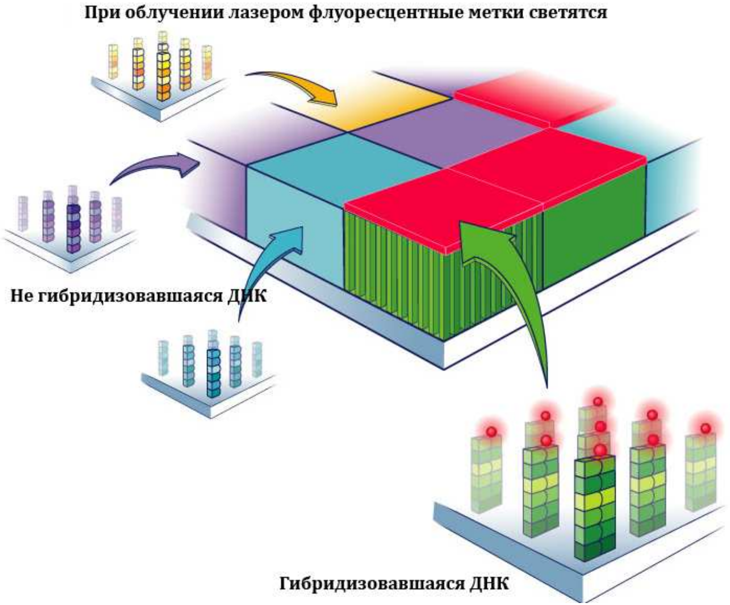
Вид ДНК-микрочипа



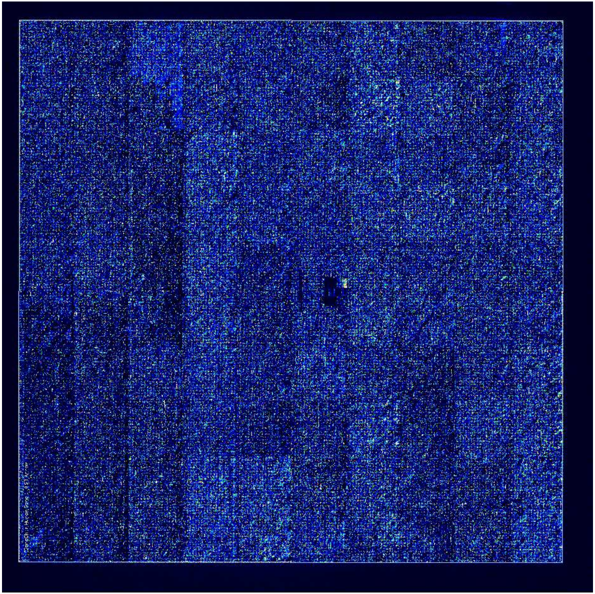
Гибридизация



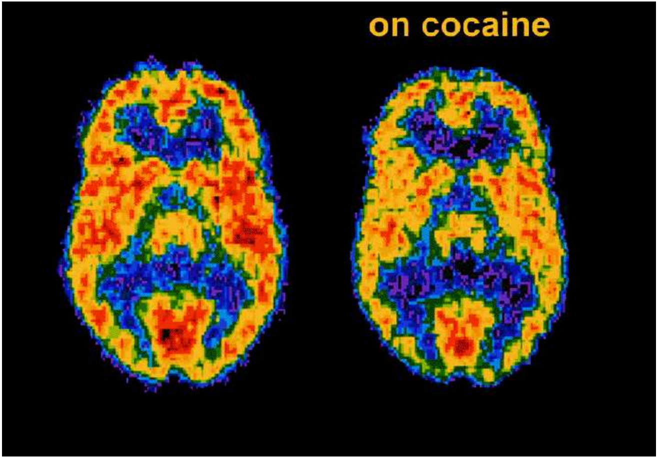
Сканирование



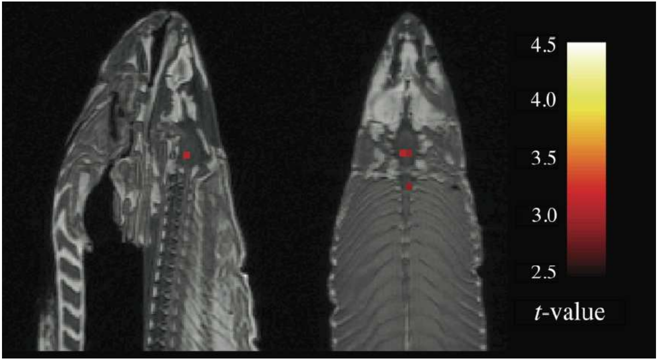
Результат сканирования



SPM



SPM



Прикладная статистика
6. Множественная проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com