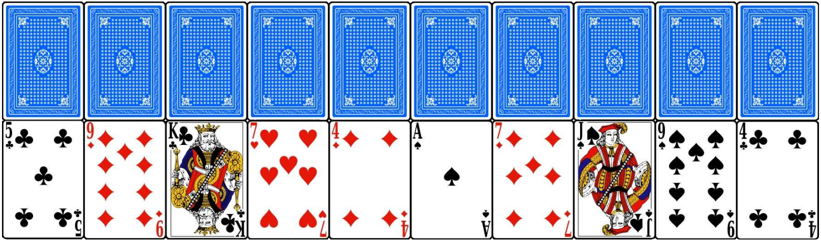
Прикладная статистика. Занятие 10. Множественная проверка гипотез.

3 мая 2011 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.

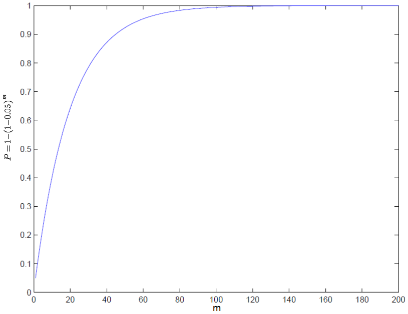
$$P\left(t \geq 9 \,|\, H_0\right) = 11 \cdot \frac{1}{2}^{10} = 0.0107421875,$$

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: $1-\left(1-11\cdot\frac{1}{2}^{10}\right)^{1000}\approx 0.9999796.$



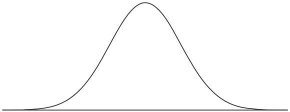
Математическая формулировка

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

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

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

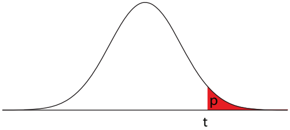
статистика: $T(\mathbf{X}), \quad T(\mathbf{X}) \sim F(x)$ при $P \in \omega;$
$T(\mathbf{X}) \nsim F(x)$ при $P \notin \omega;$



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 $p(\mathbf{x}) \leq \alpha, \quad \alpha$— 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 — $P\left(p\left(\mathbf{x}\right) \le \alpha \,|\, H_0\right) \le \alpha$.

Вероятность ошибки второго рода минимизируется путём выбора достаточно мощного критерия.

данные:	$\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_m\} \sim P \in \Omega$;
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:	$p_i = p(\mathbf{x}_i), \quad i = 1, \dots, m$;

$\mathbf{M} = \{1, 2, \dots, m\}$;
$H_0 = \bigcup_{i \in \mathbf{M}} H_i$ — 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 q означает

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

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

Theorem

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

Доказательство.

$$FWER = P\left(V \ge 1\right) \le P\left(\bigcup_{i=1}^{m_0}\left\{\tilde{P}_i \le \alpha\right\}\right) \le \sum_{i=1}^{m_0} P\left(\tilde{P}_i \le \alpha\right) \le$$

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

$\square$

Сильный контроль обеспечивается при любых p_i, без ограничений на характер зависимости между ними.

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 с одинаковой дисперсией $\sigma^2 = 1$ и разностью средних $\mu_1 - \mu_2 = 1$ при проверке гипотезы однородности на уровне значимости 0.05 имеет мощность 0.9 уже при длине выборок $n = 23$.
Если проверяется одновременно 10 гипотез, для достижения той же мощности необходима длина выборок $n = 36$; при $n = 23$ мощность равна 0.667.
Если проверяется одновременно 100 гипотез, для достижения той же мощности необходима длина выборок $n = 49$; при $n = 23$ мощность равна 0.374.
Если проверяется одновременно 1000 гипотез, для достижения той же мощности необходима длина выборок $n = 62$; при $n = 23$ мощность равна 0.165.

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

Метод Холма

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.

Нисходящая процедура Холма:

1. Если $p_{(1)} \geq \alpha/m$, принять все нулевые гипотезы $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ и остановиться; иначе отвергнуть $H_{(1)}$ и продолжать.
2. Если $p_{(2)} \geq \alpha/\left(m-1\right)$, принять нулевые гипотезы $H_{(2)}, H_{(3)}, \ldots, H_{(m)}$ и остановиться; иначе отвергнуть $H_{(2)}$ и продолжать.
3. ...

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

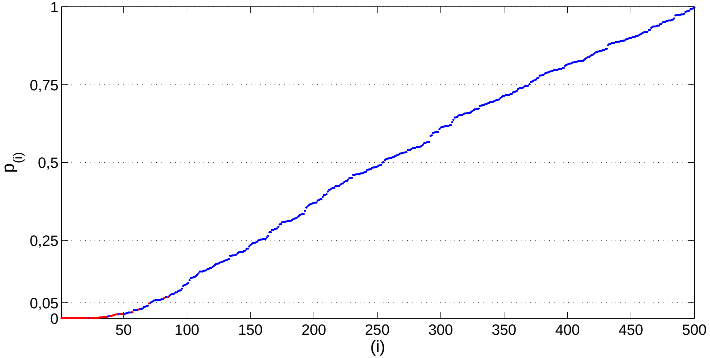
Метод Холма

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

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

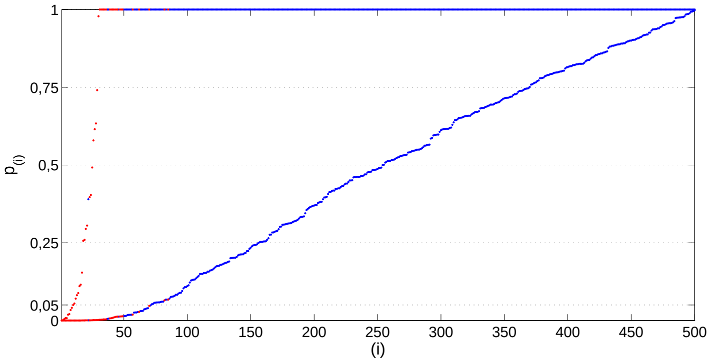
Метод Холма

Отсортированные достигаемые уровни значимости



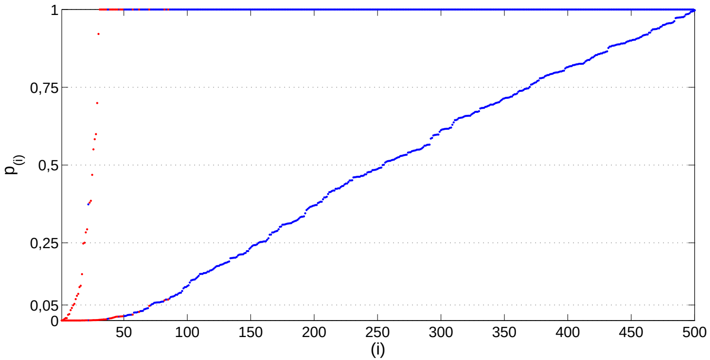
Метод Холма

Модифицированные по Бонферрони достигаемые уровни значимости



Метод Холма

Модифицированные по Холму достигаемые уровни значимости



Метод Хохберга

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.

Восходящая процедура Хохберга:

1. Если $p_{(m)} \leq \alpha$, отклонить все нулевые гипотезы $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$ и остановиться; иначе принять $H_{(m)}$ и продолжать.
2. Если $p_{(m-1)} \geq \alpha/2$, отклонить нулевые гипотезы $H_{(1)}, H_{(2)}, \ldots, H_{(m-1)}$ и остановиться; иначе принять $H_{(m-1)}$ и продолжать.
3. ...

Достигаемые уровни значимости модифицируются аналогичным образом, но рассматриваются в обратном порядке.

Метод обеспечивает сильный контроль над $FWER$ при независимости или положительной регрессионной зависимости статистик.

Нисходящая процедура: общий вид

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

Theorem

Уровни значимости нисходящей процедуры, обеспечивающие строгий контроль над $FWER$, лежат в диапазоне

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

Возможные 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 статистики, при помощи которых проверяются гипотезы, независимы, можно брать
$$\alpha_i = 1 - (1-\alpha)^{\frac{1}{m-i+1}}.$$
- Если между статистиками положительная регрессионная зависимость, то α_i лежат в диапазоне
$$1 - (1-\alpha)^{\frac{1}{m-i+1}} \leq \alpha_i \leq \alpha.$$
- Subset pivotality: нулевое распределение любого подмножества статистик T_i не зависит от того, верны или неверны соответствующие оставшимся статистикам гипотезы.
$$\begin{aligned} P\left(\bigcup_{i\in\mathbf{M}^*}\{T_i \geq t^*\} \middle| \bigcup_{i\in\mathbf{M}^*} H_i\right) &= P\left(\bigcup_{i\in\mathbf{M}^*}\{T_i \geq t^*\} | H_0\right) \\ &\leq P\left(\bigcup_{i\in\mathbf{M}}\{T_i \geq t^*\} | H_0\right) \end{aligned}$$

Максимальная статистика

$M_T = \max_i T_i$ — максимальная статистика.

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

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

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

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

Метод Беньямини-Хохберга

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

Метод Беньямини-Хохберга:
Отвергаются гипотезы $H_{(1)}, H_{(2)}, \ldots, H_{(k)}$, где

$$k = max\left\{i \colon p_{(i)} \leq \frac{i}{m}\alpha\right\}.$$

Если статистики, при помощи которых проверяются гипотезы, независимы, или если их между ними и статистиками, соответствующими верным нулевым гипотезам, положительная регрессионная зависимость, выполняется

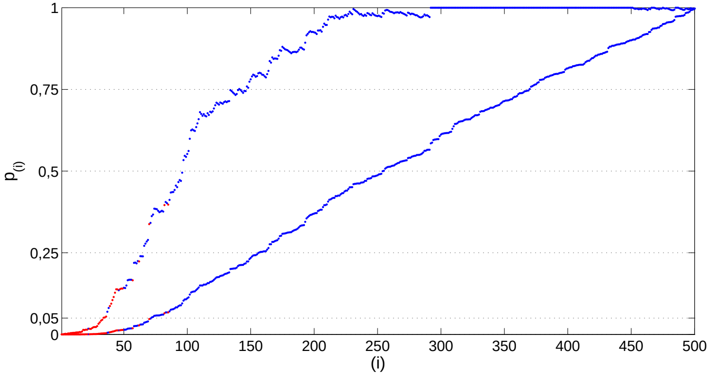
$$FDR \leq \alpha \cdot m_0/m \leq \alpha.$$

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

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

Метод Беньямини-Хохберга

Модифицированные по Беньямини-Хохбергу достигаемые уровни значимости



Метод Беньямини-Иекутиели

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

Метод Беньямини-Иекутиели:
Отвергаются гипотезы $H_{(1)}, H_{(2)}, \ldots, H_{(k)}$, где

$$k = max\left\{i \colon p_{(i)} \leq \frac{i}{m}\alpha\frac{1}{\sum_{i=1}^{m}\frac{1}{i}}\right\}.$$

При любом характере зависимости между гипотезами выполняется

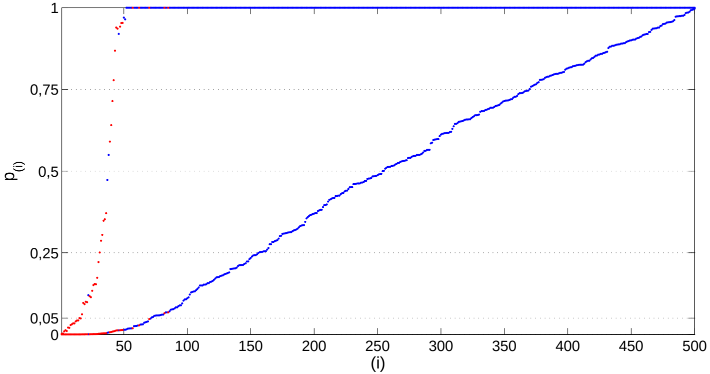
$$FDR \leq \alpha \cdot m_0/m \leq \alpha.$$

Альтернативный вид: переход к модифицированным достигаемым уровням значимости.

$$\tilde{p}_{(i)} = \min\left(m/i\sum_{i=1}^{m}\frac{1}{i}\cdot p_{(i)}, 1\right).$$

Метод Беньямини-Иекутиели

Модифицированные по Беньямини-Иекутиели достигаемые уровни значимости



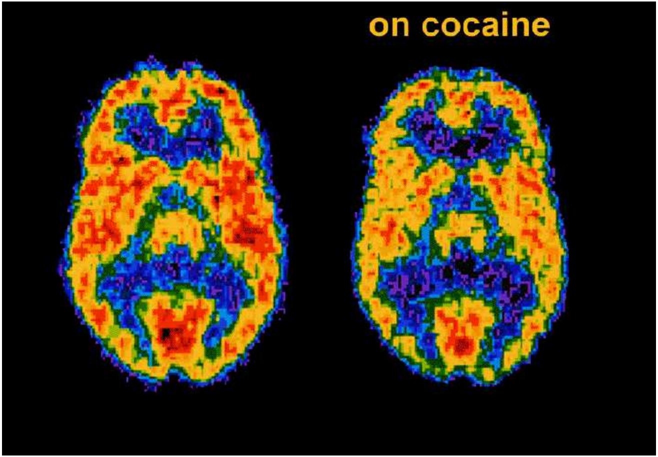
Возможные улучшения

Оценка $m_0 \Rightarrow$ более точная оценка FDR.

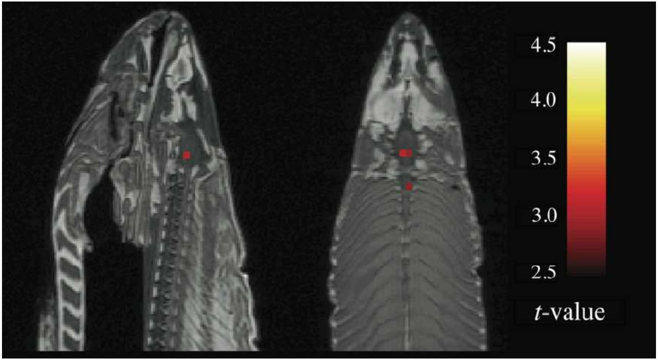
Распределение p_i-х: $g(x) = m_0 f_0(x) + (1 - m_0) f_1(x)$

Используются оценки по гистограмме p_i-х.

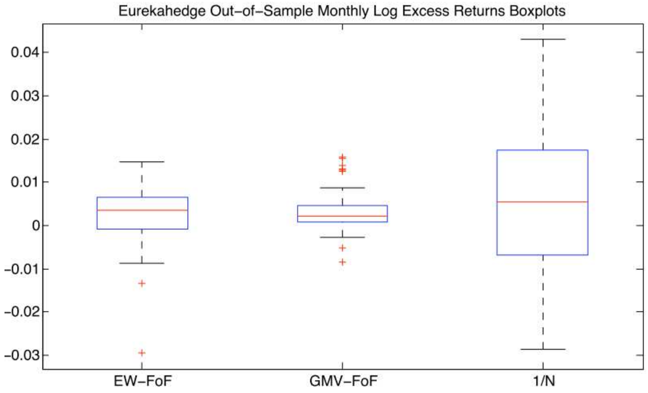
SPM



SPM



FoF


Eurekahedge Out–of–Sample Monthly Log Excess Returns Boxplots
0.04
0.03
0.02
0.01
0
–0.01
–0.02
–0.03
EW–FoF
GMV–FoF
1/N

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 $1 - (1 - 0.05)^{\frac{1}{2-1+1}} \approx 0.02532$

Прикладная статистика
Семинар 10. Множественная проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com