DS-поток, 3 курс, осень 2023

Статистика

Лекция 10

6. Проверка статистических гипотез

6.6. Множественная проверка гипотез

Поиск экстрасенсов

Этап 1: Угадайте цвета (**синий** и **оранжевый**) с учетом порядка.

Поиск экстрасенсов

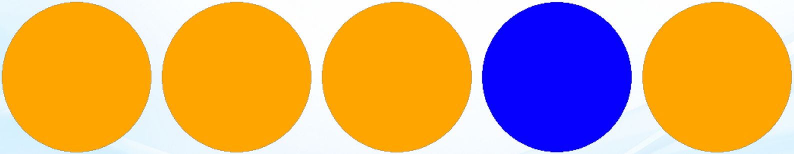
Этап 1: ответы.

Поиск экстрасенсов

Этап 2: Угадайте цвета (**синий** и **оранжевый**) с учетом порядка.

Поиск экстрасенсов

Этап 2: ответы.



Поиск экстрасенсов

В 1950 г. проводились испытания

возможности экстрасенсорного восприятия.

Этап 1: поиск экстрасенсов — испытуемому нужно угадать цвет 10 карт.

$X_1, ..., X_{10} \sim Bern(\theta)$ — результаты (правильно / нет).

$\mathsf{H}_0 \colon \theta = 1/2$ *vs.* $\mathsf{H}_1 \colon \theta > 1/2$ — наугад *vs.* не наугад

Критерий $S = \{T(x) \geqslant c_\alpha\}$, где $T(X) = \sum\limits_{i=1}^{n} X_i \sim Bin(n, \theta)$.

c	7	8	9	10
$\mathsf{P}_{1/2}(T(X) \geqslant c)$	0.172	0.055	0.010	0.001

Берем $c_\alpha = 9$, т.е. H_0 отклоняется если $\sum X_i \geqslant 9$.

Поиск экстрасенсов

Вывод: если человек верно отгадывает хотя бы 9 карт из 10, то он становится предполагаемым экстрасенсом.

В эксперименте приняли участие 1000 человек, при этом

- 9 карт верно отгадали 9 человек;
- 10 карт верно отгадали 2 человека.

В дальнейшем ни один из них не подтвердил свои способности...

$$\mathrm{P}_{1/2}\,(\text{хотя бы один из 1000 угадает 9 или 10 карт верно}) =$$

$$= 1 - \left(1 - C_{10}^{9}/2^{10} - C_{10}^{10}/2^{10}\right)^{1000} = 1 - \left(1 - 11/2^{10}\right)^{1000} \approx 0.99997$$

Гипотезы и критерии

Проверка гипотез:

$X = (X_1, \ldots, X_n)$ — выборка из неизвестного распр. P.

$\mathsf{H}_0\colon\ \mathsf{P} \in \mathscr{P}_0$ — проверяемая гипотеза.

S — критерий для проверки H_0, если H_0 отвергается $\Leftrightarrow X \in S$.

$\mathsf{P}(I_S) \leqslant \alpha$ — уровень значимости.

Размножим по $j \in \{1, \ldots, m\}$:

$X_j = (X_{j,1}, \ldots, X_{j,n_j})$ — j-ая выборка из неизвестного распр. P_j.

$\mathsf{H}_j\colon\ \mathsf{P}_j \in \mathscr{P}_j$ — проверяемая гипотеза.

S_j — критерий для проверки H_j, если H_j отвергается $\Leftrightarrow X_j \in S_j$.

$\mathsf{P}(I_{S_j}) \leqslant \alpha$ — уровень значимости.

Обобщение ошибки

Групповая ошибка первого рода (*familywise error rate*)

Вероятность отвергнуть хотя бы одну верную гипотезу.

$$FWER = \mathrm{P}(V_{\mathrm{P}\,S} > 0),$$

$V_{\mathrm{P}\,S}$ — количество верных гипотез, которые были отвергнуты критерием S для распределения P (задает верные гипотезы).

Что мы знаем?

Пусть $\mathrm{H}_1, ..., \mathrm{H}_{m_0}$ — верные гипотезы (m_0 не знаем).

$$FWER = \mathrm{P}(\text{произошла хотя бы одна ошибка I рода}) =$$

$$= \mathrm{P}\left(\bigcup_{j=1}^{m_0}\{X_j \in S_j\}\right) \leqslant \sum_{j=1}^{m_0} \mathrm{P}(X_j \in S_j) \leqslant \alpha m_0$$

А нам нужно $FWER \leqslant \alpha$.

Методы контроля FWER

Метод Бонферрони

Каждый критерий имеет уровень значимости α/m.

Метод Холма

Пересортируем гипотезы и критерии в порядке возрастания p-value:
$p_1 \leqslant \ldots \leqslant p_m$ — p-value.
$\mathrm{H}_1, \ldots, \mathrm{H}_m$ — соответствующие гипотезы.

Установим $\alpha_j = \frac{\alpha}{m-j+1}$ — уровень значимости критерия S_j.

Пусть $j = \min\{j \mid p_j > \alpha_j\}$ — номер первой неотвергнутой гипотезы.
Отвергаем гипотезы $\mathrm{H}_1, \ldots, \mathrm{H}_{j-1}$.

Всегда мощнее метода Бонферрони.

Обобщение ошибки

Ожидаемая доля ложных отклонений (*false discovery rate*)

$$FDR = \mathsf{E}_{\mathsf{P}} \frac{V_{\mathsf{P}\,S}}{\max(R_S, 1)},$$

$V_{\mathsf{P}\,S}$ — количество верных гипотез, которые были отвергнуты,

R_S — количество отвергнутых гипотез

Утверждение: $FDR \leqslant FWER$

Методы Бенджамини-Хохберга и Бенджамини-Иекутиели

Пересортируем гипотезы и критерии в порядке возрастания p-value.

Пусть $j = \max\{j \mid p_j \leqslant \alpha_j\}$ — номер последней отвергнутой гипотезы.

Отвергаем гипотезы $\mathsf{H}_1, ..., \mathsf{H}_j$.

$\alpha_j = \frac{\alpha j}{m}$ — метод Б.-Х. (*если статистики критериев независимы*)

$\alpha_j = \frac{\alpha j}{m} \Big/ \sum_{j=1}^{m} \frac{1}{j}$ — метод Б.-И. (*работает всегда*)

Замечания

Зависимости:

1. $V_{\mathsf{P}\,S}$ зависит от совместн. распр. P, общего критерия S и выборок X;
2. R_S зависит от общего критерия S и выборок X;
3. *FWER* и *FDR* зависят от совместного распр. P и общего критерия S.

Контроль *FWER* **на уровне** α означает:

for $\forall\,\mathsf{P}_1$ — распределение выборки X_1 (в т.ч. не из H_1):
...
for $\forall\,\mathsf{P}_m$ — распределение выборки X_m (в т.ч. не из H_m):
проверить $FWER \leqslant \alpha$.

Максимум FWER не обязательно достигается при справедливости всех $\mathsf{H}_1, ..., \mathsf{H}_m$.

Итерационные процедуры

Нисходящая процедура

В методе Холма можно выполнять следующие итерации.

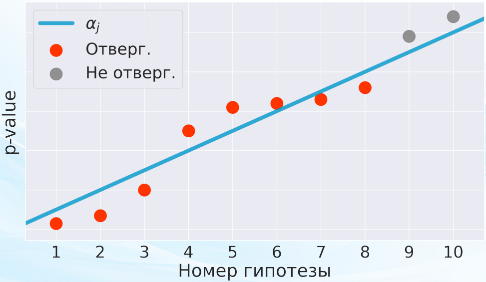
1. Если $p_1 \leqslant \alpha_1$, то отвергнуть H_1 и продолжить, иначе не отвергнуть $\mathrm{H}_1, \ldots, \mathrm{H}_m$ и остановиться;
2. Если $p_2 \leqslant \alpha_2$, то отвергнуть H_2 и продолжить, иначе не отвергнуть $\mathrm{H}_2, \ldots, \mathrm{H}_m$ и остановиться;
3. и т.д.

Итерационные процедуры

Восходящая процедура

В методах Б.-Х. и Б.-И. можно выполнять следующие итерации.

1. Если $p_m > \alpha_m$, то не отвергать H_m и продолжить, иначе отвергнуть $\mathrm{H}_m, \ldots, \mathrm{H}_1$ и остановиться;
2. Если $p_{m-1} > \alpha_{m-1}$, то не отвергать H_{m-1} и продолжить, иначе отвергнуть $\mathrm{H}_{m-1}, \ldots, \mathrm{H}_1$ и остановиться;
3. и т.д.



Скорректированные p-value

p_j — p-value, сравниваем с α_j [+процедура]

$\widetilde{p}_j$ — скорректированные p-value, хотим сравнивать с α

Метод Бонферрони

Запишем пограничные случаи:

$$\begin{cases} p_j = \alpha_j \ (= \alpha/m) \\ \widetilde{p}_j = \alpha \end{cases} \implies \widetilde{p}_j = mp_j$$

Чтобы $\widetilde{p}_j \in [0, 1]$ поправим их: $\widetilde{p}_j = \min(1, mp_j)$

Метод Холма

Аналогично получаем $\widetilde{p}_j = (m - j + 1)p_j$

Если H_{j-1} не отверглась, то дальше не отвергаем
$\Rightarrow \widetilde{p}_j = \max(\widetilde{p}_{j-1}, (m - j + 1)p_j)$

Чтобы $\widetilde{p}_j \in [0, 1]$ поправим их: $\widetilde{p}_j = \min\left(1, \max\left[\widetilde{p}_{j-1}, (m - j + 1)p_j\right]\right)$

Численный пример

Гипотезы, верность и полученные результаты:

Гипотеза	H_1	H_2	H_3	H_4	H_5	H_6	H_7
Верность	Нет	Да	Да	Нет	Да	Нет	Да
p-value	0.015	0.005	0.014	0.009	0.013	0.001	0.8

Верность известна тем, кто сгенерировал выборку, а не аналитикам :)

Перегруппируем:

Гипотеза	$H_{(1)}$	$H_{(2)}$	$H_{(3)}$	$H_{(4)}$	$H_{(5)}$	$H_{(6)}$	$H_{(7)}$
Верность	Нет	Да	Нет	Да	Да	Нет	Да
p-value	0.001	0.005	0.009	0.013	0.014	0.015	0.8

Гипотеза	$H_{(1)}$	$H_{(2)}$	$H_{(3)}$	$H_{(4)}$	$H_{(5)}$	$H_{(6)}$	$H_{(7)}$
Верность	Нет	Да	Нет	Да	Да	Нет	Да
p-value	0.001	0.005	0.009	0.013	0.014	0.015	0.8

Метод Бонферрони:

Гипотеза	p-value	α_j	p-value adj	Отвергаем?
$H_{(1)}$	0.001	0.0071	0.007	True
$H_{(2)}$	0.005	0.0071	0.035	True
$H_{(3)}$	0.009	0.0071	0.063	False
$H_{(4)}$	0.013	0.0071	0.091	False
$H_{(5)}$	0.014	0.0071	0.098	False
$H_{(6)}$	0.015	0.0071	0.105	False
$H_{(7)}$	0.8	0.0071	1.0	False

Ошибок I рода: 1
Верных отвержений: 1

Гипотеза	$H_{(1)}$	$H_{(2)}$	$H_{(3)}$	$H_{(4)}$	$H_{(5)}$	$H_{(6)}$	$H_{(7)}$
Верность	Нет	Да	Нет	Да	Да	Нет	Да
p-value	0.001	0.005	0.009	0.013	0.014	0.015	0.8

Метод Холма:

Гипотеза	p-value	α_j	p-value adj	Отвергаем?
$H_{(1)}$	0.001	0.0071	0.007	True
$H_{(2)}$	0.005	0.0083	0.03	True
$H_{(3)}$	0.009	0.0100	0.045	True
$H_{(4)}$	0.013	0.0125	0.052	False
$H_{(5)}$	0.014	0.0167	0.052	False
$H_{(6)}$	0.015	0.0250	0.052	False
$H_{(7)}$	0.8	0.0500	0.8	False

Ошибок I рода: 1
Верных отвержений: 2

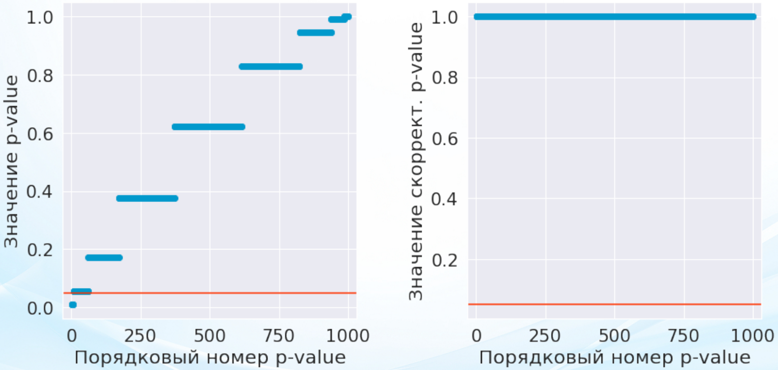
Гипотеза	$H_{(1)}$	$H_{(2)}$	$H_{(3)}$	$H_{(4)}$	$H_{(5)}$	$H_{(6)}$	$H_{(7)}$
Верность	Нет	Да	Нет	Да	Да	Нет	Да
p-value	0.001	0.005	0.009	0.013	0.014	0.015	0.8

Метод Бенджамини-Иекутиелли:

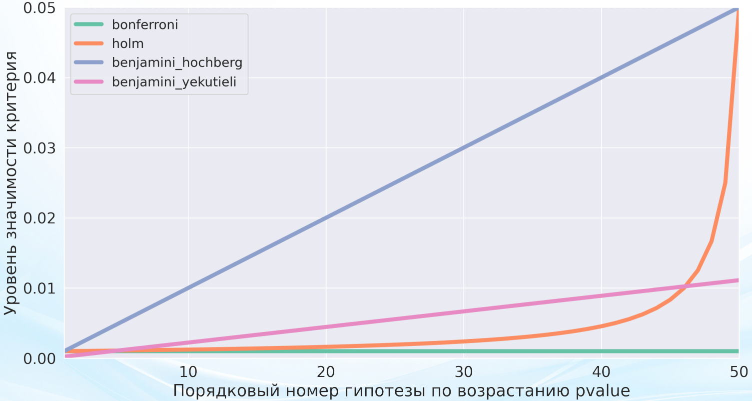
Гипотеза	p-value	α_j	p-value adj	Отвергаем?
$H_{(7)}$	0.8	0.0193	1.0	False
$H_{(6)}$	0.015	0.0165	0.045375	True
$H_{(5)}$	0.014	0.0138	0.045375	True
$H_{(4)}$	0.013	0.0110	0.045375	True
$H_{(3)}$	0.009	0.0083	0.045375	True
$H_{(2)}$	0.005	0.0055	0.045375	True
$H_{(1)}$	0.001	0.0028	0.01815	True

Ошибок I рода: 3
Верных отвержений: 3

Поиск экстрасенсов



Сравнение методов МПГ



Реализация МПГ

```
statsmodels.stats.multitest.multipletests
(pvals, alpha=0.05, method='hs',
is_sorted=False, returnsorted=False)
```

- `pvals` — значения p-value по всем критериям
- `alpha` — желаемый уровень значимости
- `method`:
  - `bonferroni`
  - `holm`
  - `sidak`
  - `holm-sidak`
  - `fdr_bh`
  - `fdr_by`

Возвращает:

- `reject` — для отвергаемых гипотез `True`
- `pvals_corrected` — скорректированые p-value

Простой пример

Знакомая задача:

$X_1, ..., X_n \sim \mathcal{N}(\theta, 1)$

$\mathrm{H}_0: \theta \geqslant 0$ *vs* $\mathrm{H}_1: \theta < 0$

РНМК: $S = \{x \in \mathbb{R} \mid \overline{x} \leqslant c_\alpha\}$

Пусть теперь две одинаковые задачи с независимыми выборками:

$X \sim \mathcal{N}(\theta_1, 1)$

$Y \sim \mathcal{N}(\theta_2, 1)$

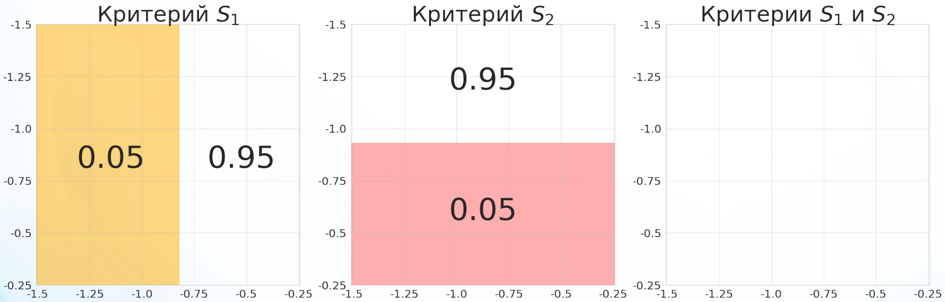
$\mathrm{H}_1: \theta_1 \geqslant 0$ *vs* $\mathrm{H}_1': \theta_1 < 0$

$\mathrm{H}_2: \theta_2 \geqslant 0$ *vs* $\mathrm{H}_2': \theta_2 < 0$

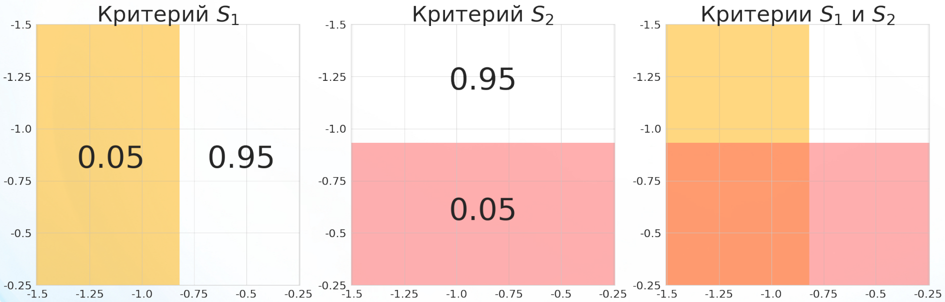
Критерии: $S_1 = \left\{(x, y) \in \mathbb{R}^2 \mid x \leqslant c_\alpha\right\}$

$S_2 = \left\{(x, y) \in \mathbb{R}^2 \mid y \leqslant c_\alpha\right\}$

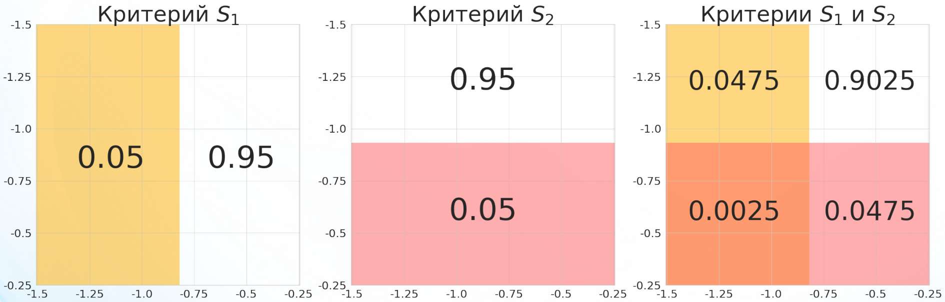
Частая ошибка: Выборки независимы $\Longrightarrow$ МПГ не нужна.



Критерий S_1
0.05
0.95
Критерий S_2
0.95
0.05
Критерии S_1 и S_2
-1.5
-1.25
-1.0
-0.75
-0.5
-0.25



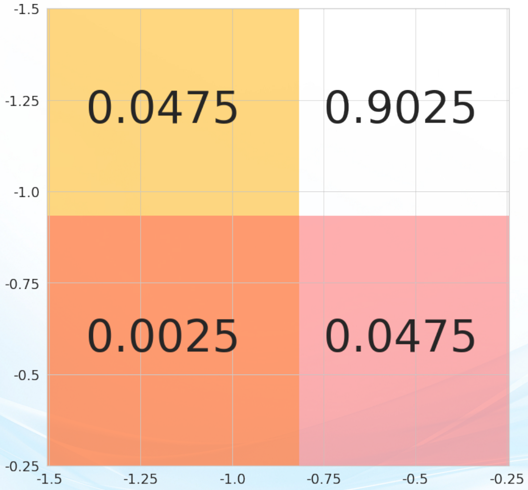
Критерий S_1
0.05
0.95
Критерий S_2
0.95
0.05
Критерии S_1 и S_2
-1.5
-1.25
-1.0
-0.75
-0.5
-0.25



Вывод:

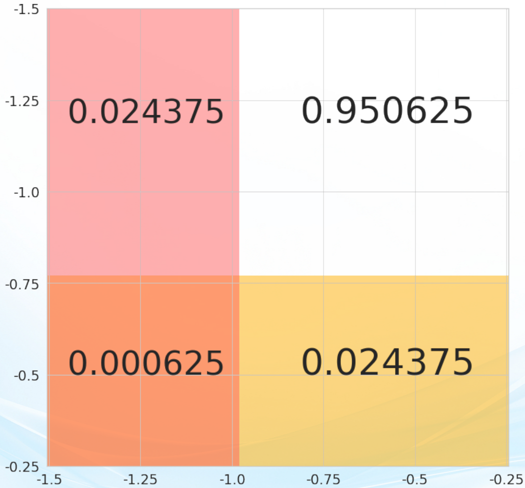
вероятность допустить хотя бы одну ошибку равна 0.0975, если обе основные гипотезы верны.

Сравнение: без корректировки



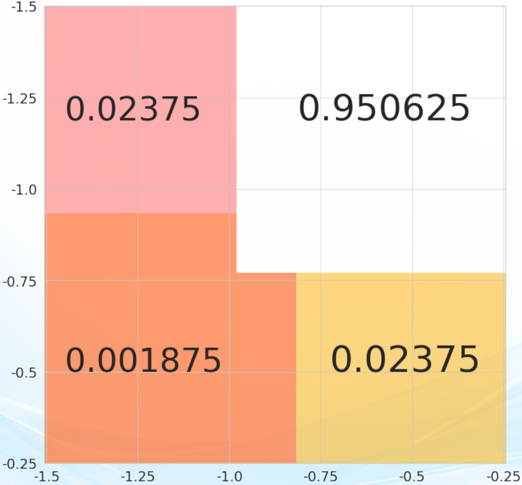
Вероятности указаны при справедливости H_1 и H_2.

Сравнение: метод Бонферрони



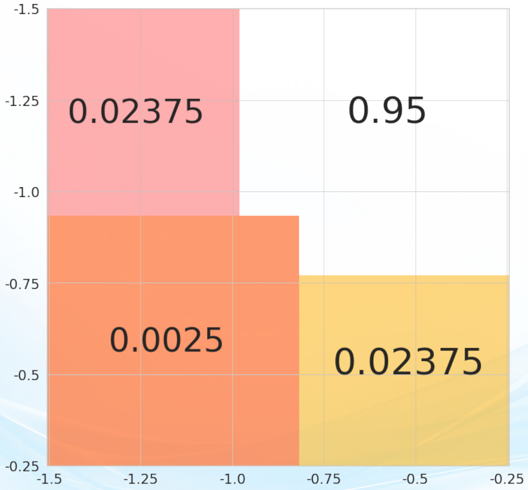
Вероятности указаны при справедливости H_1 и H_2.

Сравнение: метод Холма



Вероятности указаны при справедливости H_1 и H_2.

Сравнение: метод Бенджамини-Хохберга (не контр. FWER)



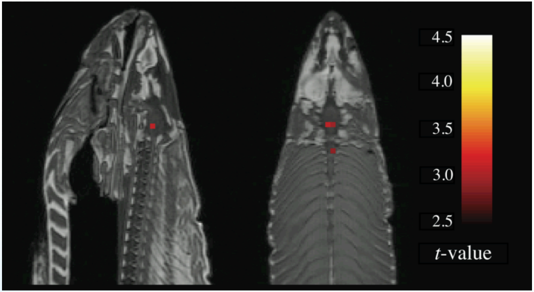
Вероятности указаны при справедливости H_1 и H_2.

Какой подход использовать?

- При первичном анализе данных, при котором только происходит формулировка интересных гипотез, можно вообще не делать поправки на МПГ. При этом **всегда** нужно приводить информацию об общем количестве гипотез и количестве отвергнутых.
- При проведении исследований и отбора признаков для дальнейшего анализа, который обычно является более сложным и дорогим, следует применять методы, контролирующие **FDR**. Обычно берут $FDR \leqslant 0.1$.
- На этапе подтверждения выводов следует проводить строгий контроль за вероятностью ошибок первого рода, контролируя **FWER**. Обычно берут $FWER \leqslant 0.05$.

Удивительные открытия

2009 год. МРТ мозга мертвого самца лосося:



МРТ дает 3D-изображение на 130 000 вокселей.

Эксперимент: Лососю показывали фото и просили его пояснить, какие эмоции испытывают люди с картинки.

Обработка: Для каждого вокселя тестируется гипотеза о наличии активации этого участка мозга.

Удивительные открытия

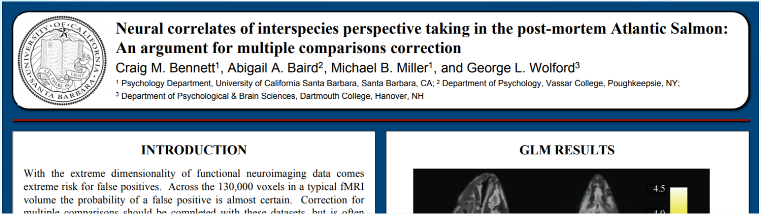
Результат: Для каждой картинки для нескольких вокселей мозга p-value оказывалось меньше 0.001.

Вывод: мертвый лосось реагирует на фотки!!!

Авторы удостоились Шнобелевской премии (2012 год) за открытие в области неврологии.

При применении МПГ лосось переставал на что-либо реагировать...

http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf




ВСЁ!