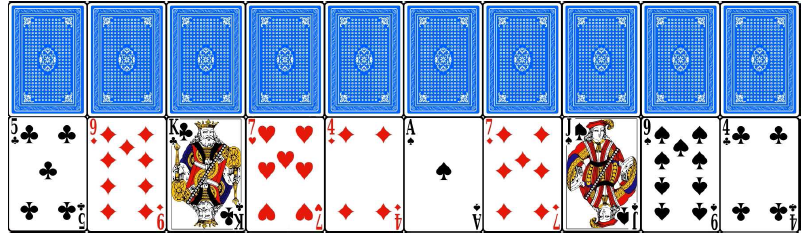
* 1. Проблема множественной проверки статистических гипотез

Работа посвящена проблеме множественной проверки статистических гипотез. Для того, чтобы понять в чем заключается проблема, давайте рассмотрим несколько примеров.

Пример 1.1.1. Поиск экстрасенсов

В 1950 году парапсихолог Джозеф Райн проводил эксперименты по выявлению людей с экстрасенсорными способностями. Каждый испытуемый должен был угадать цвет каждой карты в последовательности из 10 карт.



В эксперименте приняли участие 1000 человек. Из них 12 человек угадали 9 из 10 карт, двое — все 10 карт. В последующих экспериментах ни один из них не подтвердил свои способности. Почему? В чем причина?

Рассмотрим данный эксперимент в терминах проверки статистических гипотез.

1. *Выбор нулевой гипотезы*

H0 – испытуемый не обладает экстрасенсорными способностями

H1 – испытуемый обладает экстрасенсорными способностями

1. *Определение случайной величины и ее распределения*

Стохастический эксперимент состоит в наблюдении случайной величины ξ – числа правильно угаданных карт, ξ=0,1,2,…,10.

Случайная величина ξ имеет биномиальное распределение с параметрами (10, p), где p – вероятность «успеха» («успех» – цвет карты определен правильно).

Если гипотеза H0 справедлива, то ξ имеет распределение



где  – вероятность «успеха» при условии, что человек не имеет экстрасенсорных способностей. Оба исхода эксперимента (цвет карты определен правильно или не правильно) равновероятны для обычного человека.

Если гипотеза H1 справедлива, то ξ имеет распределение



где  – вероятность «успеха» при условии, что человек имеет экстрасенсорные способности.

1. *Выбор критерия (выбор критического множества)*

Естественно отклонять H0, когда ξ принимает большие значения и не отклонять H0, когда ξ принимает малые значения. Итак,



Если *ξ∈S*, т.е. *ξ≥l*, то H0 отклоняется.

Если *ξ∉S*, т.е. *ξ<l*, то H0 не отклоняется.

1. *Выбор уровня значимости, определение l*

Выбираем α=0,05. Вероятность ошибки I рода не превышает α:





Итак, *l=9* и критерий принимает вид 

Если *ξ≥9*, то H0 отклоняется – испытуемый обладает экстрасенсор-ными способностями.

Если *ξ<9*, то H0 не отклоняется – испытуемый не обладает экстрасен-сорными способностями

**Правильное определение цвета 9 или 10 карт из 10 – свидетельство того, что испытуемый – экстрасенс.**

1. *Вероятности ошибок I и II рода*

Вероятность ошибки I рода



можно проинтерпретировать так: в одном случае из 100 экстрасенсом будем считать не экстрасенса.

Запишем вероятности ошибок I и II рода в таблицу:

Справедливая гипотеза

|  |  |  |
| --- | --- | --- |
| Принимается гипотеза | H0 – испытуемый не обладает экстрасенсорными способностями | H1 – испытуемый обладает экстрасенсорными способностями |
| H0 – испытуемый не обладает экстрасенсорными способностями |  |  |
| H1 – испытуемый обладает экстрасенсорными способностями |  |  |

1. *Проблема множественной проверки гипотез*

В эксперименте участвовали 1000 человек. Вероятность того, что ни один из них случайно не угадает 9 или 10 карт из 10 составляет



Тогда вероятность того, что из 1000 человек хотя бы один человек случайно угадает 9 или 10 карт из 10 равна



Если мы посмотрим как ведет себя эта вероятность в зависимости от количества испытуемых, то увидим, что она растет очень быстро.

При 100 испытуемых мы найдем хотя бы одного экстрасенса с вероятностью больше ½:



а при 500 испытуемых вероятность найти хотя бы одного экстрасенса:



уже близка к единице.

То, что с помощью этой статистической процедуры мы находим экстрасенсов является прекрасным примером эффекта множественной проверки гипотез. **Если мы проверяем одновременно большое количество статистических гипотез, вероятность совершения хотя бы одной ошибки I рода становится очень большой.**

Вероятность совершения хотя бы одной ошибки I рода для разных уровней значимости и числа гипотез m приведена на рисунке 1[1].

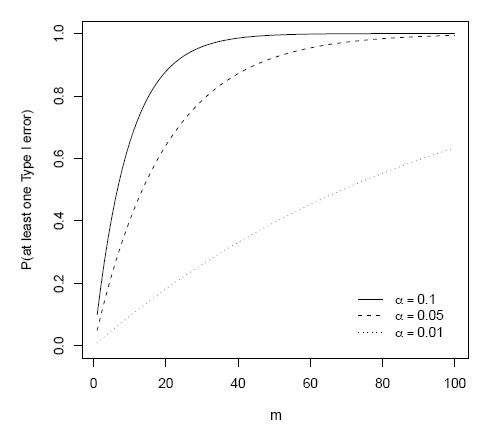


Рис. 1. Вероятность совершения хотя бы одной ошибки I рода для разных уровней значимости α и числа статистических гипотез m

[1] Frank Bretz Multiple comparisons using R / Frank Bretz, Torsten Hothorn, Peter Westfall, 2011. – 182 p.

Пример 1.1.2. Нейронаука

Еще один яркий пример действия эффекта множественной проверки гипотез можно найти в нейронауке. Представьте, что мы анализируем данные функциональной магнитно-резонансной томографии (фМРТ). Типичный дизайн эксперимента с такими данными следующий: мы берем контрольную группу испытуемых, с которыми ничего не происходит, и измеряем активность их мозга. Потом мы берем группу испытуемых, состояние которых мы каким-то образом поменяли, измеряем также их активность мозга. Дальше сравниваем эти две выборки данных, пытаясь выяснить, на какие области мозга подействовало различие между нашими двумя экспериментальными условиями.

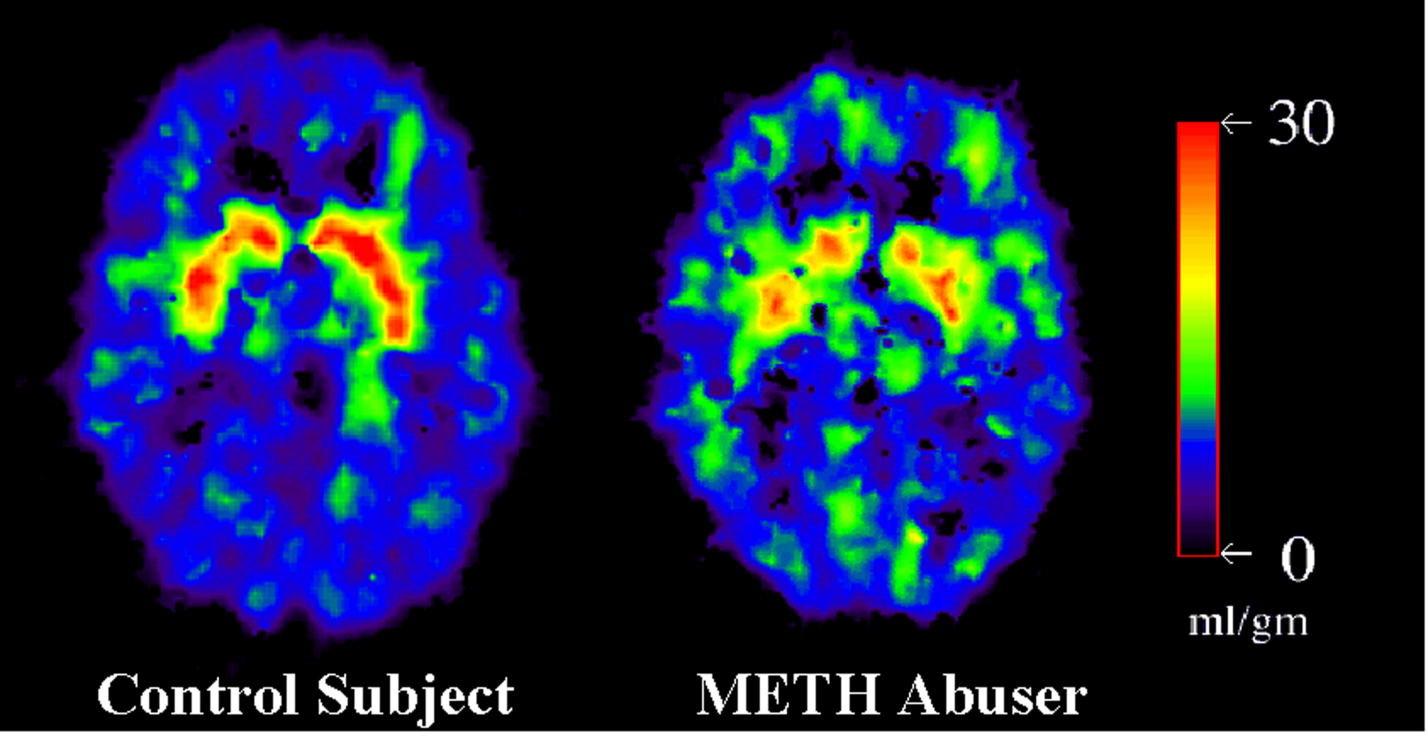


Рис. 2. Источник https://www.bnl.gov/bnlweb/pubaf/pr/2001/bnlpr030101.htm

Решение такой задачи связано с проверкой очень большого количества статистических гипотез. Фактически для двумерного изображения мозга мы проверяем гипотезу в каждой точке, для трехмерного изображения мозга, которая возникает при магнитно-резонансной томографии, мы проверяем гипотезу в каждом вокселе, то есть в каждом трехмерном пикселе трехмерного изображения мозга. Пикселей могут быть тысячи, вокселей могут быть миллионы. Таким образом, нам нужно проверить очень много гипотез. И если ничего не делать, эффект множественной проверки гипотез будет проявляться очень ярко.

Пример 1.1.3. Исследование мертвого лосося.

Это исследование стало важной вехой в области функциональной магнитно-резонансной томографии (фМРТ), с помощью которой измеряется уровень оксигенации крови в мозге (как правило, человека) при выполнении им заданий. Такие исследования широко используются, цитируются в прессе, они многое открыли нам о мозге, наших умственных способностях, а также о ряде болезненных состояний. Но до появления работы [2] не во всех подобных исследованиях делались поправки на множественные сравнения. Теперь ситуация начала улучшаться. Благодаря мёртвому лососю.

*Объект.* В фМРТ-исследовании принял участие один взрослый атлантический лосось (Salmo salar). Длина лосося – 46 см, вес – 1,7 кг. На момент сканирования лосось был мёртв.

*Задача.* Лосось должен был выполнять неограниченные по времени задания на вынесение суждений о поведении людей как результате их психического и эмоционального состояния. Лососю демонстрировали серию фотографий, на которых изображались индивидуумы в различных социальных ситуациях с ярко выраженной эмоциональной валентностью. Лосось должен был определить, какую эмоцию испытывает человек на фотографии.

*Модель.* Стимулы были организованы в блочную схему, где каждая фотография показывалась в течение 10 секунд, после чего следовали 12 секунд перерыва. Всего было показано 15 фотографий. Общее время сканирования составило 5,5 минут.

*Предварительная обработка данных.* Обработка снимков осуществлялась в статистической среде SPM, версия 2. Предварительная обработка функциональных данных сканирования включила аффинное преобразование временных рядов фМРТ по 6 параметрам, совместную регистрацию данных в Т1-взвешенное анатомическое изображение, а также сглаживание гауссовой функцией с шириной на половине высоты в 8 мм.

*Анализ.* Статистические показатели по вокселям данных о лососе рассчитывались через обычную оценку по методу наименьших квадратов в общей линейной модели. Предикторы гемодинамической реакции были смоделированы функцией в виде единичного прямоугольного импульса с канонической гемодинамической реакцией. Был использован 128-секундный временной высокочастотный фильтр, чтобы убрать шумы на низких частотах. Поправку на автокорреляцию не вводили.

*Результаты.* t-критерий использовали для выявления районов со статистически значимым изменением сигнала при предъявлении фотографии по сравнению с отсутствием таковой. t-значение для 131 степени свободы должно было превысить 3,15, что соответствовало нескорректированному p-значению = 0,001. 3 вокселя превысили порог.

Обнаружены несколько активных вокселей в кластере, расположенном в мозговой полости лосося. Размер этого кластера составил 81 мм3, а уровень значимости кластера р = 0,001. Из-за плохого разрешения получаемых эхо-планарных изображений и относительно маленького размера мозга лосося невозможно провести границы между различными отделами мозга.

Из 8064 вокселей, составивших общий объём исследования, **16 вокселей были статистически значимы.**

При проведении аналогичных сравнений по t-критерию с применением FDR и FWER поправок на множественные сравнения не было обнаружено активных вокселей даже при мягком уровне значимости (p = 0,25).

Итак, когда статистический анализ проводили с поправками на множественную проверку статистических гипотез, мертвый лосось оставался мёртвым. Когда же часть поправок опустили, он внезапно ожил: статистически значимая активность обнаружилась в его головном и спинном мозге.

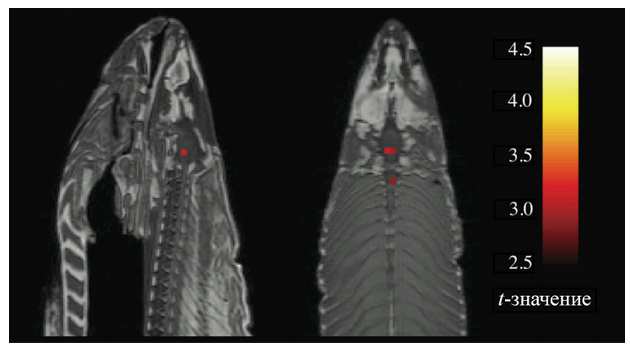


Рис. 3. Результаты фМРТ лосося. Активные области выделены красным. Источник: <http://brights-russia.org/article/scientific-method/it-died-for-science.html>

[2] Bennett et al. [Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction](http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf) // Journal of Serendipitous and Unexpected Results, 2010.