

КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И  
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Кафедра математической статистики

Информационный дайджест:

Математическая статистика

27 мая – 2 июня 2019 г.

30 мая 2015 г., источник: N+1

Для КМС ИВМиИТ подготовил Казанцев А.В.



Илья Щуров

математик, доцент НИУ ВШЭ

## Ложь, наглая ложь и (математическая) статистика

**Стоит ли доверять математической статистике? Если взять наугад сотню научных статей, использующих статистические методы, сколько в них будет ошибочных выводов?**

**Как работает математическая статистика? Часто ли она ошибается? Что означает «различие оказалось статистически значимым ( $p < 0.05$ )»? Сегодня мы попробуем во всём этом разобраться.**

### Играть или не играть?

Было лето. Аркадий Петров с женой Олей и сыном Серёжей гуляли по набережной одного небольшого курортного города. Первый день отпуска прошёл довольно неплохо, море оказалось в меру тёплым, но не таким прозрачным, как обещала реклама — что, впрочем, не помешало всему семейству вдоволь накупаться и наплескаться. А ужин в уютном ресторанчике и прогулка по набережной стали прекрасным завершением дня.

— Па-ап, ма-ам, пойдём туда? — Серёжа показал пальцем на красочный павильон «Испытай удачу».

— Это что? Азартные игры? Не надо нам туда, давай лучше купим тебе сладкой ваты, — забеспокоилась Оля.

— Да мы только посмотрим, — сказал Аркадий. Он был не слишком азартным человеком, но вывеска почему-то привлекла его внимание.

Центральное место в павильоне занимало большое «Колесо Фортуны», разделенное ровно на две половины — красную и зелёную. Рядом висела табличка с инструкцией.

## САМАЯ ЧЕСТНАЯ ИГРА

### ПРАВИЛА

1. ОПУСТИТЕ МОНЕТУ В 10 РУБ.
2. РАСКРУТИТЕ КОЛЕСО.
3. КРАСНАЯ ЗОНА — ПРОИГРЫШ, ЗЕЛЁНАЯ ЗОНА — ВЫИГРЫШ
4. РАЗМЕР ВЫИГРЫША 21 РУБ.
5. КОЛИЧЕСТВО ИГР НЕ ОГРАНИЧЕНО.

Серёжа прочитал инструкцию и стал что-то считать. Через минуту он сказал:

— Игра выгодная. Зелёная и красная зоны одинаковые, так что вероятность выиграть равна вероятности проиграть. Если мы сыграем сто раз, то примерно 50 раз выиграем и 50 раз проиграем. Но выигрыш составит  $21 \times 50 = 1050$  рублей, тогда как за 100 игр мы заплатим 1000 рублей. Значит, мы останемся в плюсе на 50 рублей! Правильно, пап?

Аркадий только недавно рассказывал сыну про теорию вероятностей и закон больших чисел. Рассуждение казалось правильным, но по опыту он знал, что так в жизни не бывает.

— Это всё обман и мошенничество! — прервала их размышления Оля. — Пойдём отсюда скорее.

— Как же так? — удивился Серёжа. — Теория вероятностей нас обманывает?

— Про теорию вероятностей не скажу, но люди обманывают часто. Особенно те, кто организуют лотереи и азартные игры.

— Наверное, колесо на самом деле не симметрично — там где-нибудь прилеплен груз и оно чаще останавливается на красной половине, чем на зелёной. — попробовал примирить теорию с жизненным опытом Аркадий. — Нам, наверное, действительно лучше уйти.

— Но разве мы не можем это проверить? — спросил любознательный Серёжа — и тут же, не дав родителям опомниться, достал из кармана десятирублевую монетку, кинул её в щель и раскрутил колесо. Через полминуты оно остановилось. Стрелка указывала на красную зону. Раздался неприятный звук и засветилось табло «К сожалению, вы проиграли. Попробуйте ещё раз».

— Я же говорила! Ты выкинул 10 рублей на ветер! Кстати, откуда они у тебя? — Оля была недовольна и расстроена. Серёжа тоже расстроился, но по другой причине.

— Если колесо действительно всегда останавливается на красном, нужно жаловаться в полицию — пусть они запретят обманывать людей! — воскликнул он.

— Строго говоря, мы не доказали, что колесо «нечестное». — отметил Аркадий. — Даже если предположить, что колесо симметричное, а игра действительно честная, вероятность проигрыша составляет  $1/2$ . Тот факт, что ты проиграл в этот раз, никак этому не противоречит. Вероятность — такая штука...

— Я думаю, что оно всегда будет останавливаться на красном, чтобы забрать больше денег доверчивых людей. Давайте уже уйдём отсюда! — Ольге совсем не нравилось происходящее.

— Подожди. Давай проверим — интересно ведь. У меня есть ещё одна монетка. — сказал Аркадий. Колесо снова раскрутилось — и на этот раз ко всеобщему удивлению остановилось в зелёной зоне. Раздались фанфары. «ПОБЕДА!» — сообщил автомат. В лоточек упали две 10-рублёвые и одна рублёвая монетка.

— Значит, выиграть всё-таки можно! — воскликнул Серёжа. — Колесо честное!

— Ну уж нет. Этого вы точно не доказали! — ответила Ольга. — Из того факта, что можно выиграть один раз, точно ничего не следует! Просто проиграть гораздо легче, чем выиграть. Вот, смотрите! — она взяла монетки, которые вернул автомат, и снова опустила их в щель. «Количество попыток: 2» — высветилось на табло. Оля раскрутила колесо. Проигрыш. И ещё раз. Снова проигрыш.

— Вот видите! — победно заявила она. — Три проигрыша из четырёх!

— Мам, но ведь это тоже могло получиться в результате случайности! Как это было в первый раз...

— Если рассуждать таким образом, то тебя никакой результат не убедит в том, что колесо нечестное! — Оля стала сердиться. — Даже тысяча проигрышей из тысячи — ты всегда сможешь сказать «это была случайность, на самом деле колесо честное, нам просто не повезло». А раз так, то и дальнейшие эксперименты бессмысленны. Идём. — Она взяла сына за руку и уверенным шагом пошла к выходу. Аркадий пошёл следом, пытаясь вспомнить, что им рассказывали в институте на курсе по математической статистике...

## **Проверка гипотез**

Кто был прав в этой истории? Действительно ли три проигрыша из четырёх является убедительным аргументом в пользу того, что колесо несимметричное? И как вообще можно делать хоть какие-то выводы, если любой результат можно объяснить случайностью? Давайте разбираться.

Во-первых, общая схема рассуждений наших героев выглядит разумной. Вслед за ними, мы будем рассуждать как бы «от противного» — предположим, что колесо на самом деле симметричное, а вероятность выиграть равна вероятности проиграть. Это называется «нулевой гипотезой». Затем мы посмотрим на имеющиеся у нас результаты эксперимента и попробуем понять, насколько они согласуются с нулевой гипотезой.

Для начала, рассмотрим гипотетическую ситуацию, которую обрисовала Ольга: тысяча проигрышей из тысячи. Конечно, если бы это случилось, мы бы вряд ли смогли бы кого-нибудь убедить, что колесо симметричное. Теория вероятностей действительно не запрещает таким событиям происходить, но говорит, что их вероятность чрезвычайно низка, то есть происходят они очень редко. Мы можем её посчитать аккуратно: если вероятность проигрыша в одном раунде равна  $1/2$ , а все игры независимы, то вероятность проиграть в двух раундах равна  $1/2 \times 1/2 = 1/4$ , в трёх —  $1/2 \times 1/2 \times 1/2 = 1/8 = 1/2^3$  и так далее. Вероятность проиграть в тысяче раундов из тысячи равна  $1/2^{1000}$  — это число с 300 нулями после запятой! Если бы мы увидели такой результат, то у нас было бы два объяснения: либо колесо симметричное, но нам просто как-то невероятно, фантастически повезло (вернее, не повезло), либо колесо всё-таки асимметричное и вероятность проиграть больше, чем вероятность выиграть. Второе объяснение кажется более разумным и именно его мы и принимаем — а значит отвергаем наше исходное предположение о симметричности колеса.

Вероятность в  $1/2^{1000}$ , которую мы посчитали, показывает, насколько имеющиеся у нас данные согласуются с нулевой гипотезой. Она называется p-value и обозначается буквой p. Если p мало, то это означает, что нулевая гипотеза кажется нам малореалистичной. Логика такая: если бы мы считали, что нулевая гипотеза верна, то нам пришлось бы прийти к выводу, что произошло событие с очень маленькой вероятностью — иными словами, что нам фантастически повезло. Но выбирая между фантастическим везением и необходимостью отвергнуть нулевую гипотезу мы отдаём предпочтение второму варианту. Именно так проходят все рассуждения в математической статистике.

## Уровень значимости

Кажется очевидным, что тысячи проигрышей из тысячи более, чем достаточно, чтобы убедиться в том, что колесо «нечестное». И даже 999 из 1000 нам бы скорее всего хватило. А если бы мы проиграли сто раз из ста? А если 99 из ста? Или 98? Или пять из пяти? Или два из двух? Где проходит граница между «это можно объяснить случайностью» и «нам должно было фантастически повезти, так не бывает».

Давайте посчитаем. Вероятность двух проигрышей из двух (при условии, что колесо симметричное), равна  $1/2 \times 1/2 = 1/4$ . Она не кажется очень маленькой. Вероятность трёх проигрышей из трёх равна  $1/2^3 = 1/8 = 0,125$ , то есть чуть больше 10%. Для четырёх проигрышей из четырёх это число составит  $1/2^4 = 0,0625$ . Для пяти уже  $1/2^5 = 0,03125$ . Для десяти проигрышей эта вероятность будет близка к одной тысячной.

Чтобы принять решение о том, какую вероятность мы считаем «очень маленькой», нужно провести где-то границу. Эта граница называется «уровнем значимости» и обозначается греческой буквой  $\alpha$ . Чему равняется уровень значимости? Обычно его выбирают равным 5% или 1% (то есть  $\alpha = 0,05$  или  $\alpha = 0,01$ ). Конкретное значение уровня значимости зависит от области деятельности, для которой проводится исследование, и связано с «ценой ошибки» в этой области.

Например, если мы положим, что уровень значимости равен  $\alpha = 0,05$ , то четырёх проигрышей из четырёх будет ещё недостаточно, чтобы отвергнуть нулевую гипотезу о симметричности колеса (то есть мы считаем, что это могло произойти в результате случайности), а пяти проигрышей (из пяти попыток) уже достаточно (не верим в столь низкую вероятность). А если бы мы приняли уровень значимости равным  $\alpha = 0,01$ , то даже

шести проигрышей (из шести попыток) было бы недостаточно, а семи уже достаточно. (Проверьте, что это так.)

Проиллюстрируем на нашем примере, из каких соображений выбирается уровень значимости. Допустим, мы решили, что если придём к выводу о несимметричности колеса, то вызовем полицию и потребуем закрыть игровую точку. Полиция проведёт расследование и установит, действительно ли колесо несимметричное. Если это так, то мы получим премию (и чувство глубокого морального удовлетворения), а если не так, то штраф за ложный вызов. Понятно, что выбор уровня значимости зависит от того, насколько велика премия (и моральное удовлетворение) и насколько велик штраф. Если штраф за ложный вызов маленький, а премия большая, то мы можем выбрать в качестве уровня значимости относительно большое число. А если наоборот, штраф за ложный вызов очень большой, то мы будем действовать более осторожно и заявим в полицию только в том случае, когда получим очень убедительные свидетельства «нечестности» колеса, то есть очень маленькое  $p$ -value. В этом случае уровень значимости должен выбираться маленьким.

## Вычисление $p$ -value

Чему равняется  $p$ -value для случая из нашей истории — три проигрыша из четырёх? Давайте считать.

У нас было четыре попытки, в каждой мы могли выиграть или проиграть. Давайте будем обозначать выигрыш буквой «В», а проигрыш буквой «П» и запишем результат этих четырёх попыток в виде последовательности из четырёх букв. Результат, полученный в истории, записывается как «ПВПП» (он называется «элементарным исходом»). Всего таких разных результатов может быть  $2^4=16$  штук (попробуйте выписать их все). Значит, вероятность каждого элементарного исхода — при условии, что нулевая гипотеза верна — равна  $1/16$ .

Но ситуация «три проигрыша из четырёх» могла реализоваться разными способами. Мы могли проиграть в первый раз и выиграть три других («ВППП») или проиграть в последний раз («ПППВ») или, наконец, в предпоследний («ППВП»). Всего таких комбинаций, дающих «три проигрыша из четырёх», четыре штуки. Вероятность каждой равна  $1/16$ . Значит, вероятность получить «три проигрыша из четырёх» равна  $4/16=1/4$ .

Нашли ли мы  $p$ -value? На самом деле, нет.

Чтобы убедиться в этом, рассмотрим другой пример. Допустим, мы крутили колесо 1000 раз и ровно 500 из них выиграла и ровно 500 проиграла. С одной стороны, это именно такой результат, который мы должны были «в среднем» получить, если колесо «честное». Он точно не свидетельствует против гипотезы о симметричности колеса. (Пожалуй, он может свидетельствовать в пользу того, что игры не являются независимыми, то есть ставит под сомнение всю модель — но это уже немножко другой вопрос.) С другой — вероятность получить ровно 500/500 (а не 499/501 или 502/498) довольно мала (она равна  $C_{1000}^{500}/2^{1000} \approx 0,025$ ). Если бы мы считали, что нашли таким образом  $p$ -value, то должны были бы отвергнуть гипотезу о симметричности (при уровне значимости 5%)! Хотя очевидно, что этого делать не следует.

Чтобы найти  $p$ -value, мало вычислить, какова вероятность получения тех результатов, которые мы на самом деле получили. Нужно учесть, что могли получиться другие

результаты, свидетельствующие (ещё сильнее) в пользу того, что отвергнуть нулевую гипотезу необходимо.

Вернёмся к нашей истории. Вероятность получить три проигрыша из четырёх равна  $1/4$ . Но если бы мы проиграли все четыре раза, это было бы не меньшим аргументом в пользу асимметричности колеса. Значит, нам нужно учесть и этот случай тоже. Вероятность получить его равна  $1/16$  и общая p-value составляет  $5/16$ .

### **Некоторые выводы**

Теперь можно попытаться ответить на вопросы, вынесенные в преамбулу. Статистика — наука точная, но вероятностная природа процессов, которые она описывает, даёт о себе знать: если перед вами есть сотня утверждений, подтверждающихся на уровне значимости 5%, в среднем 5 из них могут оказаться неверными. Но не стоит расстраиваться. Предупрежден — значит, вооружен.

### **Задачи**

- Следует ли отвергнуть нулевую гипотезу (о симметричности и честности колеса) в том случае, который описан в истории, на уровне значимости  $\alpha=0,05$ ?
- Следует ли отвергнуть нулевую гипотезу на уровне значимости  $\alpha=0,1$ , если колесо покрутили 6 раз и 5 раз проиграли? А на уровне значимости 5%?

<https://nplus1.ru/blog/2015/05/30/p-value>