

## Семинар 11: Введение в АБ-тесты

«Если выпадет решка, то отпустим конечно. И в небо летит монетка с одинаковыми сторонами.»

Олег ЛСП о неудачном дизайне АБ-теста

На этом семинаре разговор пойдёт про АБ-тесты и проверку гипотез. Задачки из этого семинара призваны просто продемонстрировать откуда растут ноги всей замысловатой процедуры с проверкой гипотезы. Из-за этого понятия в задачках иногда формулируются не очень чётко. Если вы хотите уметь профессионально проверять гипотезы, а не просто иметь об этом какое-то представление, лучше почитайте учебник.

### Упражнение 1 (которое сеет в наших головах раздор и сомнение)

В Селе АБтестово проживает 4 человека. У каждого из них свой рост:

Маша	150
Паша	160
Саша	180
Даша	190

Дедя Фёдор, Шарик и Матроскин проезжают через АБтестово в Простоквашино транзитом. Каждый из них заинтересовался ростом местных жителей и решил по небольшой подвыборке из двух человек посчитать средний рост всех жителей АБтестово.

- а) Посчитайте настоящий средний рост в АБтестово по всей генеральной совокупности.
- б) Шарик посчитал средние по Саше и Даше и сказал, что это оценка среднего роста в АБтестово. Сколько у него получилось? Насколько сильно эта оценка отличается от настоящего среднего?
- в) К Матроскину в выборку затесались Маша и Паша. Какую оценку он получил? Далека ли она от реального среднего?
- г) К дяде Фёдору в выборку попали Маша и Саша. Как дела обстоят с его оценкой?
- д) Подерутся ли между собой Шарик, Матроскин и дядя Фёдор? Почему результаты получились именно такими? Может ли так происходить в реальности?

### Упражнение 2 (в котором происходит исследование)

Жизнь в Простоквашино изрядно испортилась. Почтальону Печкину надоела вся эта ругань. Чтобы раз и навсегда покончить с раздорами, он сел на велосипед и поехал в АБтестово. Там он опросил всех четверых жителей села, а после стал фантазировать что могло бы получиться в качестве среднего, если бы он опросил только двух каких-то жителей.

- а) Является ли средний рост случайной величиной? Сколько значений принимает эта случайная величина (сколько вариантов опросить местных жителей есть у Печкина)?

- б) Найдите все возможные значения среднего роста в АБтестово. Постройте гистограмму для этого среднего значения. Как и в прошлый раз, столбики стройте с шагом 5, верхнюю границу включайте в столбик. Отметьте на картинке рост, который получил Шарик, дядя Фёдор и Матроскин. Какая из оценок ближе всего к центру распределения?
- в) Какова вероятность оказаться в хвостах распределения? Какова вероятность оказаться в его центре?

### Упражнение 3 (в котором вскрывается правда)

Построив распределение для среднего значения роста в АБтестово, Печкин очень сильно удивился. Оказалось, что это случайная величина. Печкин решил узнать у своего друга по переписке, Роналда Фишера, как правильно делать выводы, когда ты видишь только **часть генеральной совокупности, то есть выборку**.

Фишер объяснил Печкину, что  $\bar{x}$  при большом числе наблюдений, вошедших в него, имеет нормальное распределение. Когда мы хотим сделать выводы о среднем, нам нужно работать сразу со всем распределением. Например, с помощью правила трёх сигм для него можно построить доверительный интервал, то есть интервал, в котором с вероятностью 99% лежит истинное значение среднего.

- а) Найдите стандартное отклонение для Шарика, Матроскина и Фёдора по формуле

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}.$$

- б) Постройте для каждого из парней доверительный интервал по правилу трёх сигм. Обратите внимание, что стандартное отклонение, которое мы посчитали в первом пункте — стандартное отклонение для роста. Нам нужно скорректировать его на число наблюдений, чтобы получить стандартное отклонение для среднего, то есть надо построить интервал

$$\left( \bar{x} - 3 \cdot \frac{\hat{\sigma}}{\sqrt{n}}; \quad \bar{x} + 3 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right)$$

- в) Лежит ли настоящий средний рост во всех трёх доверительных интервалах? Что это означает? Насколько широкими вышли интервалы?
- г) Кто такой Роналд Фишер? Хороших ли друзей заводит себе Печкин?

### Упражнение 4 (в котором в Простоквашино наступает мир)

Печкин приехал на велосипеде из АБтестово в Простоквашино и принёс его жителям новое знание. Матроскин, Шарик и дядя Фёдор были поражены этим знанием. Все склоки и ссоры закончились. Жители Простоквашино помирились. Прошла неделя. Как-то вечером ребята пили чай да призадумались: а можно ли по собранным наблюдениям как-то проверить гипотезу о том, что средний рост в АБтестово равен 160?<sup>1</sup>

---

<sup>1</sup>ЧАЙ? ОНИ ТОЧНО ТАМ ЧАЙ ПИЛИ?

Посреди ночи простоквашинская братва завалилась к Печкину и стала мучить его вопросами. Мудрый почтальон набросал следующие мысли:

1. Мы знаем, что  $\bar{x}$  — случайная величина, которая имеет нормальное распределение.
2. Значит расстояние  $\bar{x} - 160$  — это тоже случайная величина с нормальным распределением.
3. Если наша гипотеза верна,  $\bar{x} - 160 = 0$  и распределение концентрируется вокруг нуля.
4. Значит мы можем построить для расстояния  $\bar{x} - 160$  доверительный интервал. Если окажется, что ноль оказался внутри доверительного интервала, мы не можем отвергнуть гипотезу. Если он оказалось за пределами интервала, мы отвергаем гипотезу.
5. При этом, если мы будем пользоваться правилом 3-х сигм, при отвержении гипотезы, мы ошибёмся с вероятностью 1%, так как наш доверительный интервал будет покрывать истинное значение с вероятностью 99% (есть ещё другая ошибка, **ошибка 2 рода**, зря согласиться с гипотезой, но про неё мы на этом семинаре говорить не будем).

Проверьте гипотезу о том, что  $\mu$ , так обычно обозначают то среднее значение, которое задумала природа, равно 160. Будем использовать выборку дяди Фёдора. Используя её же, проверим гипотезу о том, что  $\mu = 100$ .

### Упражнение 5 (в котором дядя Фёдор помогает людям)

Дядя Фёдор настолько был в восторге от проведённого исследования, что написал статью об этом на [habr.ru](https://habr.ru). Теперь ему пишут со всех концов мира. Например, вчера дяде Фёдору пришло три письма:

- Аристарх, Пантелей и Иван исследуют рост людей. Они сделали три выборки. Аристарх занимается баскетболом, поэтому он опросил своих друзей по команде. Пантелей измеряет рост людей у остановки, где люди ждут автобус. Иван залезает в дома к молодым девушкам и измеряет их рост, пока они спят. Что такое репрезентативность выборки? Чья выборка будет репрезентативной? Почему?
- Хипстер Сергей пишет, что он опросил в Москве и Питере по 100 человек. Каждому он задавал вопрос: "Кофе любишь?" В Москве "Да" сказали 50 человек, в Питере 55 человек. Можно ли исходя из этого сделать вывод, что в Питере кофе любят больше? Как правильно узнать, где кофе любят больше?
- Знахарка Акулина пишет, что смешала в тазике "доктор Мом" с соком редьки. Этот настой она дала простудившейся внучке. Внучка выздоровела. Означает ли это, что лекарство работает? Как правильно проверить работоспособность лекарства?
- Фермер Андрей хочет проверить насколько хорошо работает его особый навоз с секретным ингредиентом. Для этого он разделил поле с помидорами на два куска: западный и восточный. На западном куске он использует свой новый навоз. На восточном — старый. Когда кусты начнут плодоносить он сможет посмотреть на какой части поля урожай оказался в среднем с куста больше и сделает выводы. Правильно ли поступает Андрей?
- Рузвельт сражается на выборах в президенты со своим оппонентом Альфом Лэнданом<sup>2</sup>.

---

<sup>2</sup>пахнет вмешательством дяди Фёдора в выборы

Происходит это аж в 1936 году. Журнал «Литерари Дайджест» опрашивает накануне выборов аж 10 миллионов человек насчёт выборов президента. На основе такого огромного числа респондентов журнал предсказывает победу республиканцу Лэндану с перевесом (60 на 40). В выборах побеждает демократ Рузвель — как раз с таким же перевесом, но в обратную сторону. Как думаете, почему так произошло и что журнал сделал не так?

Помогите дяде Фёдору ответить на эти вопросы.

## Ещё задачи

В этом разделе находится ещё пара задач, которые можно порешать руками. Обязательно попробуйте решить их дома.

### Упражнение 6 (экзамены)

Ежегодно более 200000 людей по всему миру сдают стандартизированный экзамен GMAT при поступлении на программы MBA. В повседневной ситуации средний результат составляет 525 баллов, стандартное отклонение — 100 баллов.

Сто студентов закончили специальные подготовительные курсы и сдали экзамен. Средний полученный ими балл — 541.4. Проверьте гипотезу о неэффективности специальных подготовительных курсов. Есть ли в них смысл?

### Упражнение 7 (монета Олега)

Олег подбрасывает монетку и орёт: "ОРЁЛ-РЕШКА-ОРЁЛ-РЕШКА!". Ещё он недавно посмотрел фильм Кристофера Нолана "Тёмный рыцарь". Там ему очень понравился Харви Дент. Потому что у него тоже была монетка, которую тот подбрасывал. Олегу стало интересно: а правильная ли у него монетка. Действительно ли она выпадает орлом с вероятностью  $\frac{1}{2}$ ?

- а) Олег подбросил монетку трижды и получил комбинацию: ОРР. Найдите долю выпадения орла. Дальше будем обозначать эту долю как  $\hat{p}$ .
- б) В семинаре по статистике мы выяснили, что для выборки из нулей и единиц дисперсию можно найти как  $\hat{p} \cdot (1 - \hat{p})$ . Для того, чтобы найти её для доли, по аналогии со средним, нужно поделить на  $n$ .

В конечном итоге стандартное отклонение для доли считается по формуле

$$\sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}.$$

На теории вероятностей вы докажете это более строго. Найдите стандартное отклонение доли.

- в) Можно показать, что  $\hat{p}$  имеет нормальное распределение<sup>3</sup>. Постройте для вашей оценки доли 95% доверительный интервал по формулам:

---

<sup>3</sup>на самом деле асимптотически нормальное (при большом числе наблюдений)

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Найдите его ширину. Лежит ли  $\frac{1}{2}$  в этом интервале?

- г) Олег подбросил монетку ещё два раза и получил ОРРОР. Найдите доверительный интервал для этой ситуации. Найдите его ширину. Стал ли он уже? Почему это произошло?

## Упражнение 8 (сборник жизненных историй)

- Во время Второй Мировой войны американские военные собрали статистику попаданий пуль в фюзеляж самолёта. По самолётам, вернувшимся из полёта на базу, была составлена карта повреждений среднестатистического самолёта. С этими данными военные обратились к статистике Абрахаму Вальду с вопросом, в каких местах следует увеличить броню самолёта. Что посоветовал Абрахам Вальд и почему? Как это связано с репрезентативностью?
- Компания, выпускающая колу решила протестировать как повлияет на продажи увеличение количества сахара в напитке. Для этого была собрана фокус-группа, которой предложили на выбор два напитка: со стандартным содержанием сахара и увеличенным. Людям нужно было попробовать оба и выбрать тот, который им больше понравился. В результате этого исследования выяснилось, что большее количество людей предпочитает Кока-Колу с увеличенным количеством сахара. Было решено увеличить количество сахара в напитке для более широкой аудитории. Неожиданно продажи упали. Вопрос: что в этом АБ-тесте пошло не так?
- Радомир работает аналитиком в интернет-магазине. Его главная задача на следующие полгода — увеличить число покупок. Для этого он решил поменять цвет кнопки «купить» с красного на зелёный. На его взгляд это должно привести к росту покупок. Как это можно проверить, не уничтожив при этом бизнес?
- Часто в социальных сетях можно увидеть вот такие посты:

Помогите нашему студенту в исследовании. Пожалуйста, пройдите небольшой опрос, он займет менее 5 минут.  
[https://docs.google.com/forms/d/e/1FAIpQLSfnp-8-jzV\\_Y..](https://docs.google.com/forms/d/e/1FAIpQLSfnp-8-jzV_Y..)

Влияние дополнительного образования на заработную плату.

Опрос проводится студентом экономического отделения РАНХиГС для дипломной работы.

\* Required

Заполните, пожалуйста, форму.

Пол \*

☐ М

☐ Ж

Возраст (лет) \*

Влияние дополнительного образования на заработную плату.

docs.google.com

Как считаете, какие проблемы возникнут у исследователя с выборкой?

- Радомир ушёл заниматься машинным обучением в крупную ритейл-компанию. Его главная задача в этом полугодии — оптимизация издержек. Радомиру нужно научиться предсказывать продажи помидоров в разных частях города, чтобы более оптимально заполнять ими торговые точки.

Очень плохо, когда в магазин привозят слишком много помидоров и они протухают, либо наоборот слишком мало помидоров, и их не хватает людям.

Радомир с ходу понял, что перед ним задача регрессии и обучил случайный лес. Теперь он умеет предсказывать спрос на помидоры в каждом магазине города в любой день года. Судя по тестовой выборке модель Радомира помогает сэкономить. Как можно было бы убедиться в этом на практике и не поломать продажи? На что надо обратить внимание при дизайне АБ-теста?