

Семинар 9: Введение в АБ-тесты

Задача 1 (которая сеет в наших головах раздор и сомнение)

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

Маша	140
Паша	150
Саша	200
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?

Задача 2 (в которой 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?

Задача 3 (в которой Печкин находит решение проблемы)

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.

Фишер объяснил Печкину, что \bar{x} имеет нормальное распределение. Когда мы хотим сделать выводы о среднем, нам нужно работать сразу со всем распределением. Например, с помощью правила двух сигм для него можно построить доверительный интервал, то есть интервал, в котором с вероятностью 95% 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 двух 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

$$\left(\bar{x} - 2 \cdot \frac{\hat{\sigma}}{\sqrt{n}}; \quad \bar{x} + 2 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \right)$$

- 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?

Задача 4 (в которой 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 соры 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?

Посреди ночи простоквашинская братва завалилась к Печкину и стала мучать 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.
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 наблюдаемое нами расстояние оказалось внутри доверительного интервала, мы

не можем отвергнуть гипотезу. Если оно оказалось за пределами интервала, мы отвергаем гипотезу.

5. При этом, если мы будем пользоваться правилом 3-х сигм, мы ошибёмся с вероятностью 5%, так как наш доверительный интервал будет покрывать истинное значение с вероятностью 95%.

Проверьте гипотезу о том, что $\mu = 160$ по этому алгоритму, используя выборку дяди Фёдора. Используя её же проверьте гипотезу о том, что $\mu = 120$. В данном случае буквой μ мы обозначили настоящее среднее.

Задача 5 (в которой дядя Фёдор помогает людям)

Дядя Фёдор настолько был в восторге от проведённого исследования, что написал статью об этом на 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. 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?

Помогите дяде Фёдору ответить на эти вопросы.

Ещё задачи

В этом разделе находится ещё пара задач, которые можно порешать руками. Попробуйте решить их дома. На следующем семинаре мы их все разберём.

Задача 6 (про экзамены)

Ежегодно более 200000 людей по всему миру сдают стандартизированный экзамен GMAT при поступлении на программы MBA. 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 программы.

Задача 7 (проблемы с монеткой)

Олег подбрасывает монетку и орёт: "ОРЕЛ-РЕШКА-ОРЕЛ-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. ЛСП стало интересно: а правильная ли у него монетка. Действительно ли она выпадает орлом с вероятностью $\frac{1}{2}$?

1. Олег подбросил монетку трижды и получил комбинацию: ОРР. Найдите долю выпадения орла. Далее будем обозначать эту долю как \hat{p} .
2. На теории вероятностей вы докажете, что стандартное отклонение для доли считается по формуле $\sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$. Найдите среднее отклонение доли.
3. Можно показать, что \hat{p} имеет нормальное распределение. Постройте для вашей оценки доли 95% доверительный интервал по формулам:

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

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

4. 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?

Задача 8 (про Вальда)

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 статистику Абрахаму Вальду с вопросом, в каких местах следует увеличить броню самолёта. Что посоветовал Абрахам Вальд и почему? Как это связано с репрезентативностью?