Статистические тесты

Тестирование гипотез

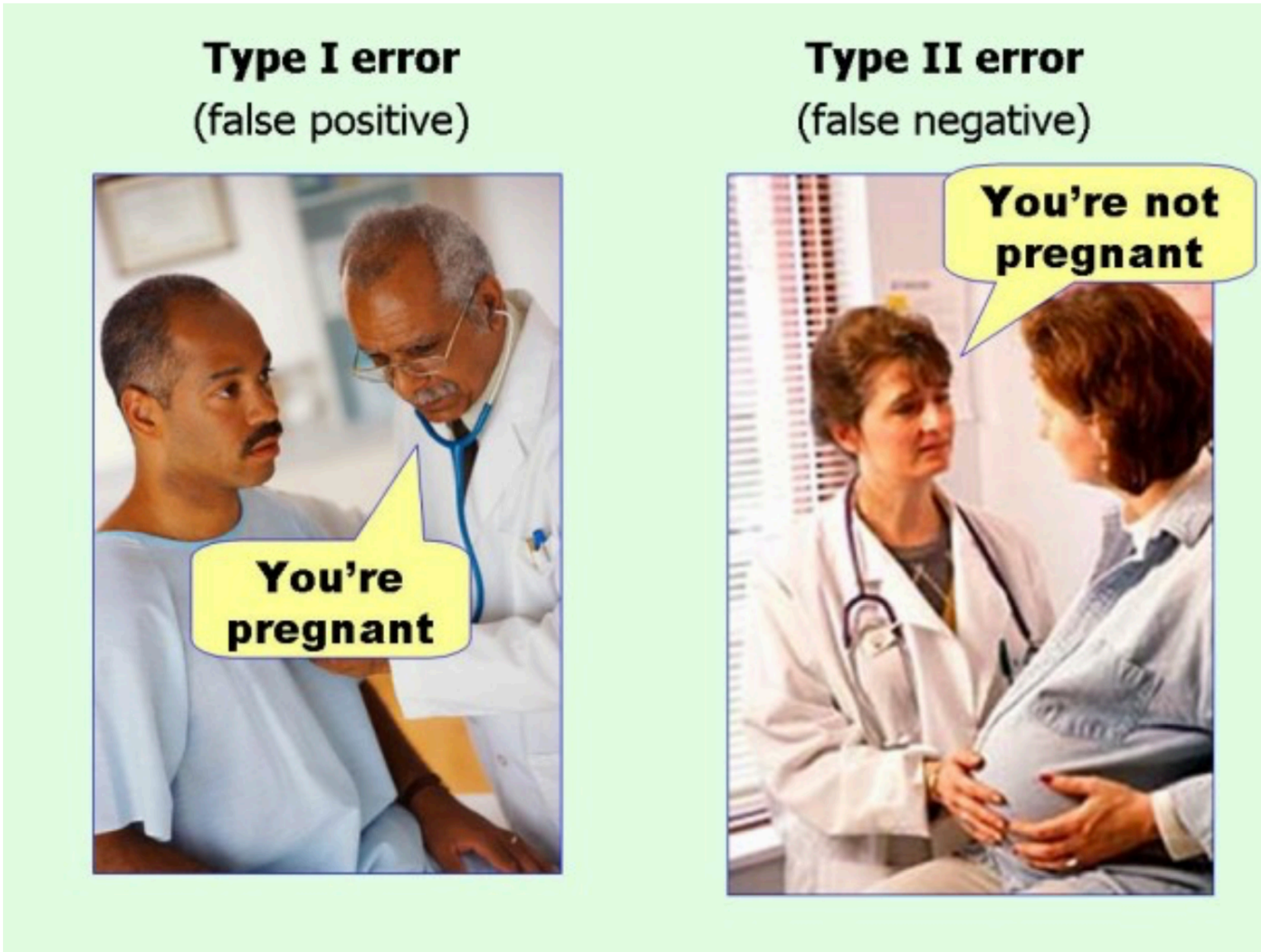
Нулевая гипотеза (H0) – это основное проверяемое предположение, которое обычно формулируется как отсутствие различий, отсутствие влияние фактора, отсутствие эффекта, равенство нулю значений выборочных характеристик и т.п.
Примером нулевой гипотезы в педагогике является утверждение о том, что различие в результатах выполнения двумя группами учащихся одной и той же контрольной работы вызвано лишь случайными причинами.

Другое проверяемое предположение (не всегда строго противоположное или обратное первому) называется конкурирующей или альтернативной гипотезой (H1). Обычно она соответствует предположению, что мы нашли значимое воздействие какого-то фактора

Тренировка

Сформулируйте нулевую и альтернативную гипотезу в данных случаях:
1) Есть две группы студентов, одна приходит всегда раньше первой и слушает лекцию раньше. Вторая слушает ту же лекцию, но позже. Известны оценки первой и второй группы студентов;
2) Есть две группы людей, одни в числе близких друзей имеют людей с повышенным весом, вторые нет. Известны веса людей в первой и второй группах;
3) Есть две группы студентов. Одна работает весь семестр, вторая учит предмет в последнюю неделю. Известны оценки по этому предмету для обеих групп;
4) Есть две группы людей. Одни слушают смотрят телеканал “Домашний” в течении не менее одного часа, вторые - нет. Известны результаты теста (в баллах) на критическое мышление для обеих групп;
5) Есть два чата в Телеграм - флудящий и нефлудящий. Известно количество годноты, кинутой в данный чат в течении месяца (расписано по дням);
6) Есть две группы программистов - пишущие только на Python, и пишущие только на C/C++ . Известна статистика самоубийств среди первых и вторых в течении года;
6) Есть три группы программистов - пишущие только на Python, пишущие только на C/C++ и пишущие только на JavaScript. Известна статистика самоубийств среди первых и вторых в течении года;

Ошибки первого и второго рода



Ошибки первого и второго рода

H0	верная	ложная
Отклоняется	Ошибка первого рода (alpha, FP)	Решение верное
Не отклоняется	Решение верно	Ошибка второго рода (beta, FN)

Задача

Представим, что мы хотим проверить, насколько хорошо витамин С помогает в лечении простуды. Для этого мы делим пациентов на пары (на основе пола, возраста, здоровья и т.д.). Далее считаем сколько, в скольких парах люди, принимавшие витамин С, выздоровели от простуды раньше. Гипотезы:

H0: P (витамин С лучше) = ½
H1: P (витамин С лучше) != ½

Допустим, что наш эксперимент состоял в том, что мы собрали 17 пар наблюдений и наблюдали в 13 парах, что раньше выздоровел принимавший витамин С.

Как оценить насколько подтвердилась наша гипотеза?

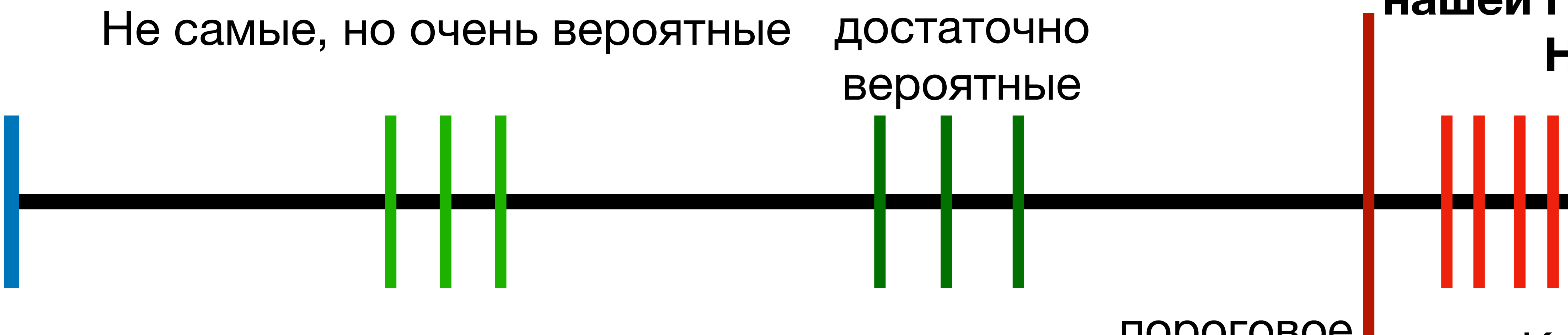
Задача

У нас есть 17 испытаний с вероятностью успеха p. По определению - биномиальное распределение.

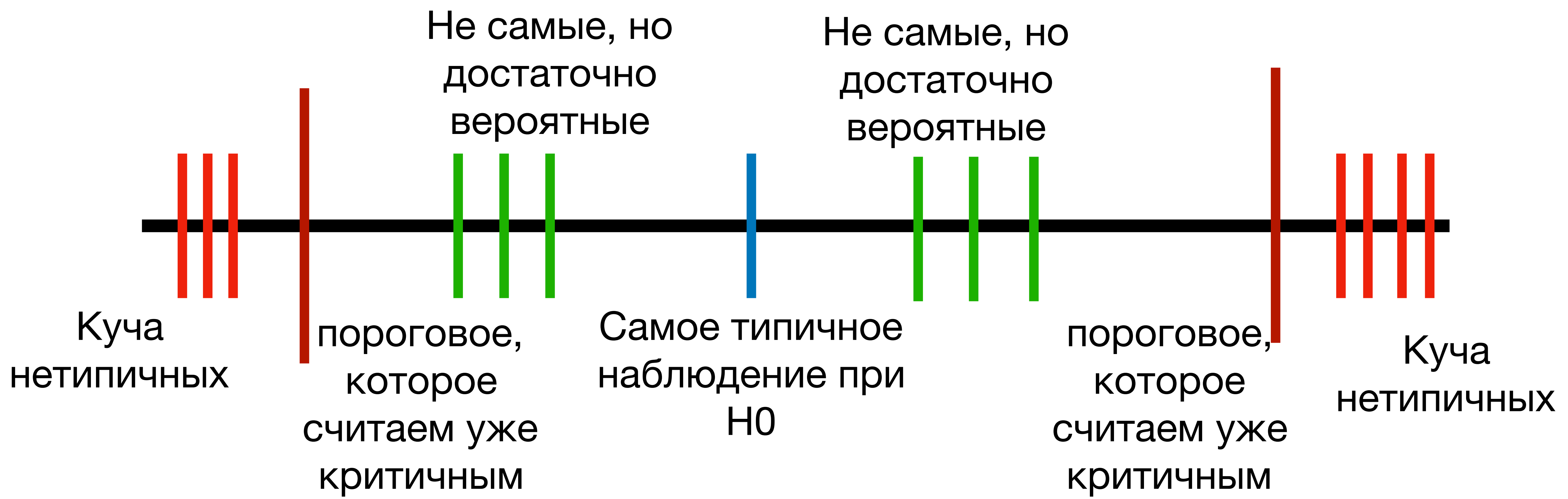
Можно посчитать вероятность нашего наблюдения при условии H0:

$$C_{17}^{13} p^{13} (1-p)^4 = 0.018$$

Но это не говорит нам о вероятности нашей гипотезы H0..
Можно рассуждать иначе - пусть мы будем считать это наблюдение значимым. Тогда логично считать и все менее вероятные при условии нашей гипотезы наблюдения значимыми.



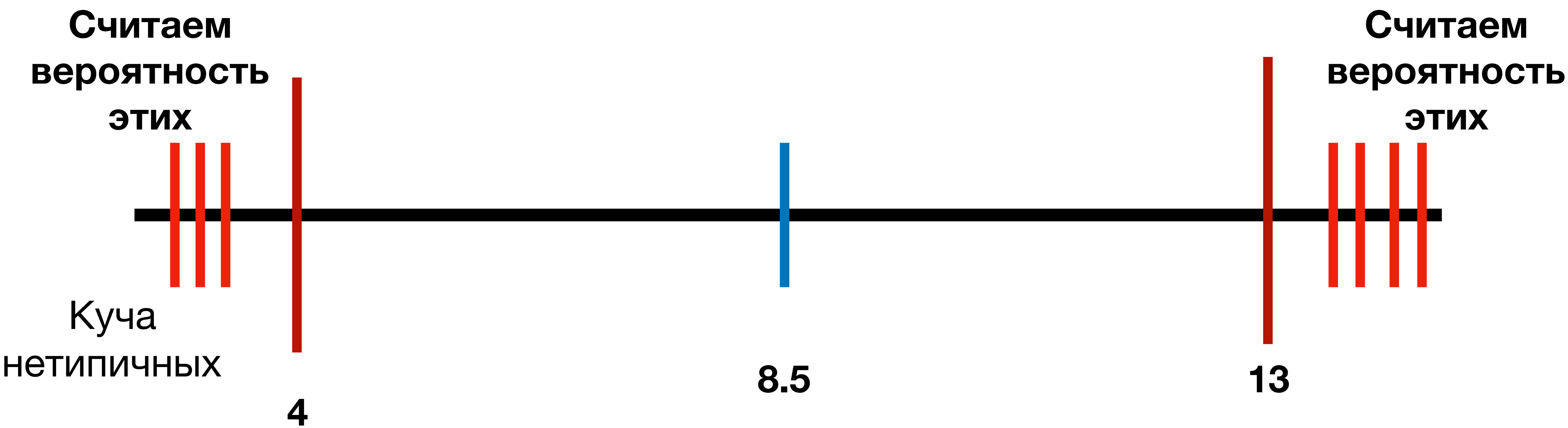
Задача



Часто у распределения нетипичные наблюдения с обеих сторон смотрим, потому иногда удобнее представлять что-то такое

Задача

Такие значения для нас - от 13 до 17, и от 4 до 0.



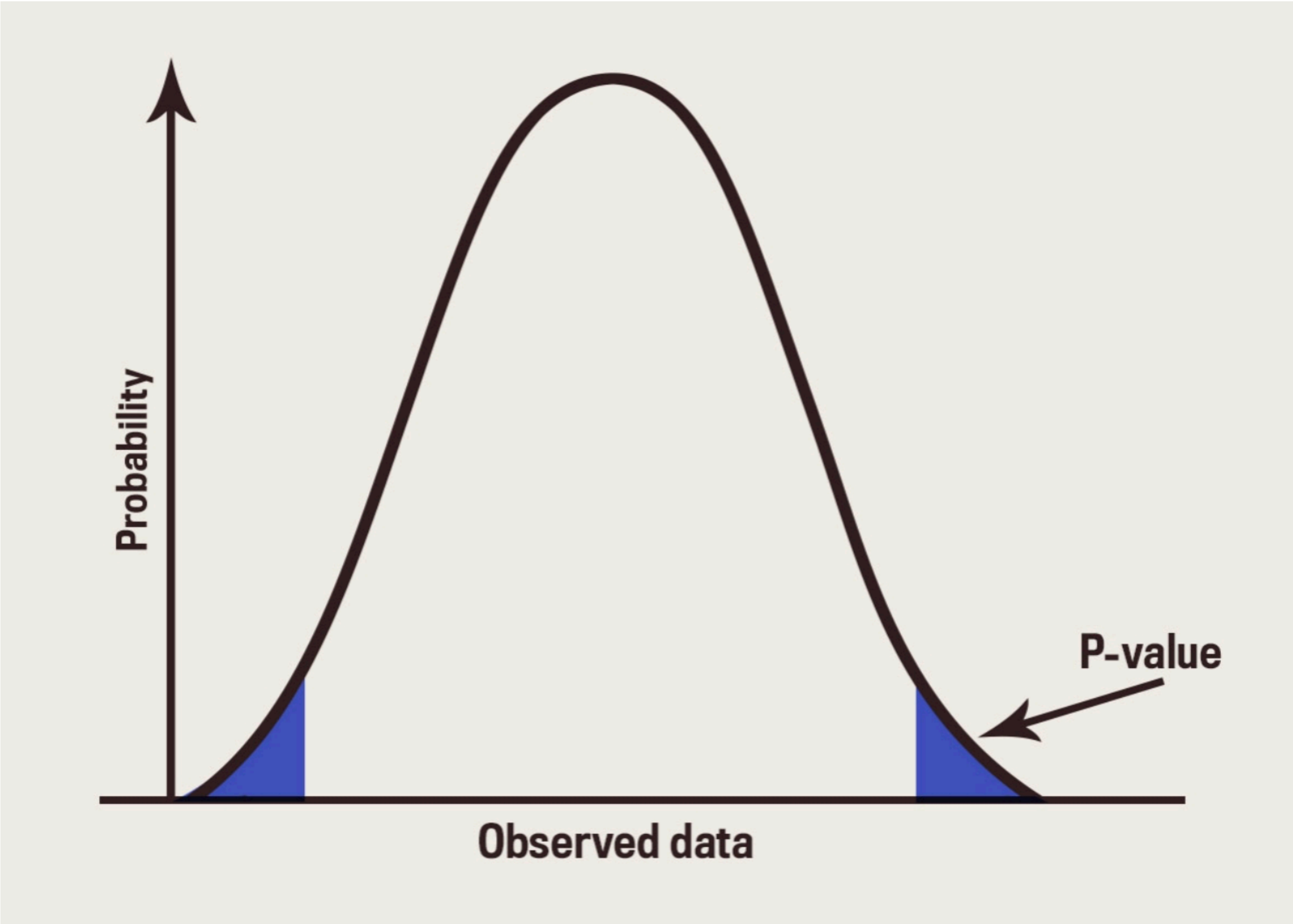
$$\sum_{i=13}^{17} C_{17}^{i} p^{i} (1-p)^{17-i} + \sum_{i=0}^{4} C_{17}^{i} p^{i} (1-p)^{17-i} = 0.049$$

Обычно берут порог 0.05, потому наше наблюдение значимо

P-value

P-value - вероятность получить результат как минимум такой же критический как тот, что мы наблюдаем, считая, что нулевая гипотеза является правильной

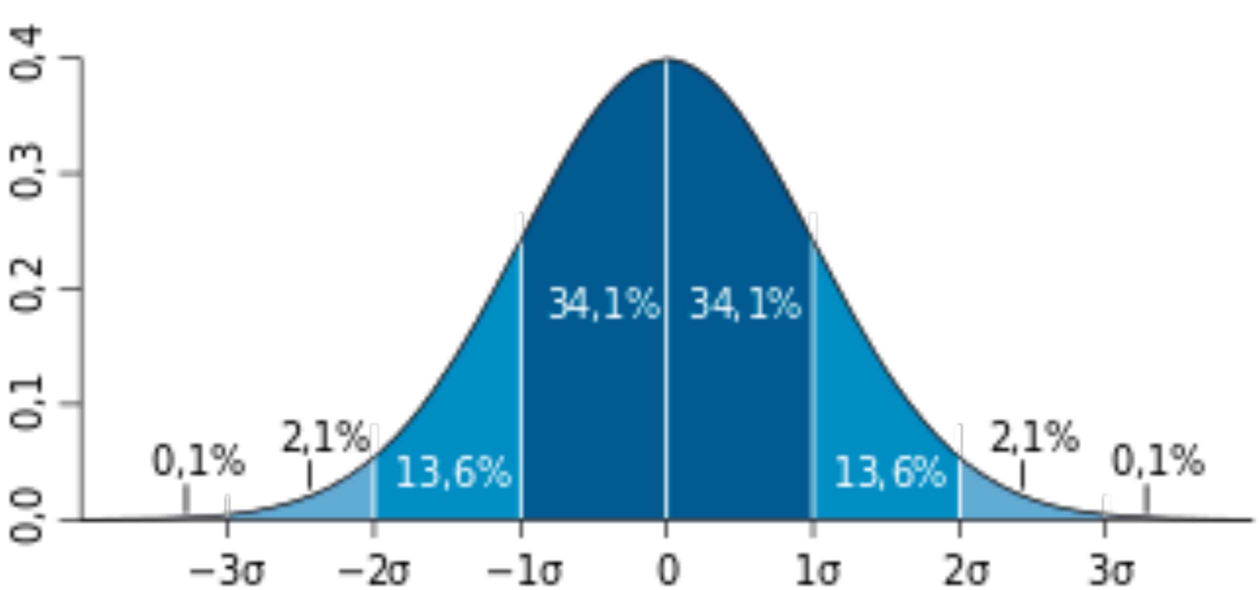
Другими словами - если нулевая гипотеза верна, то насколько вероятно получить ту выборку, которую мы получили, или более критичную



Тесты

Параметрические

Предполагают, что генеральная совокупность распределена по какому-то закону, использует параметры этой совокупности



Непараметрические

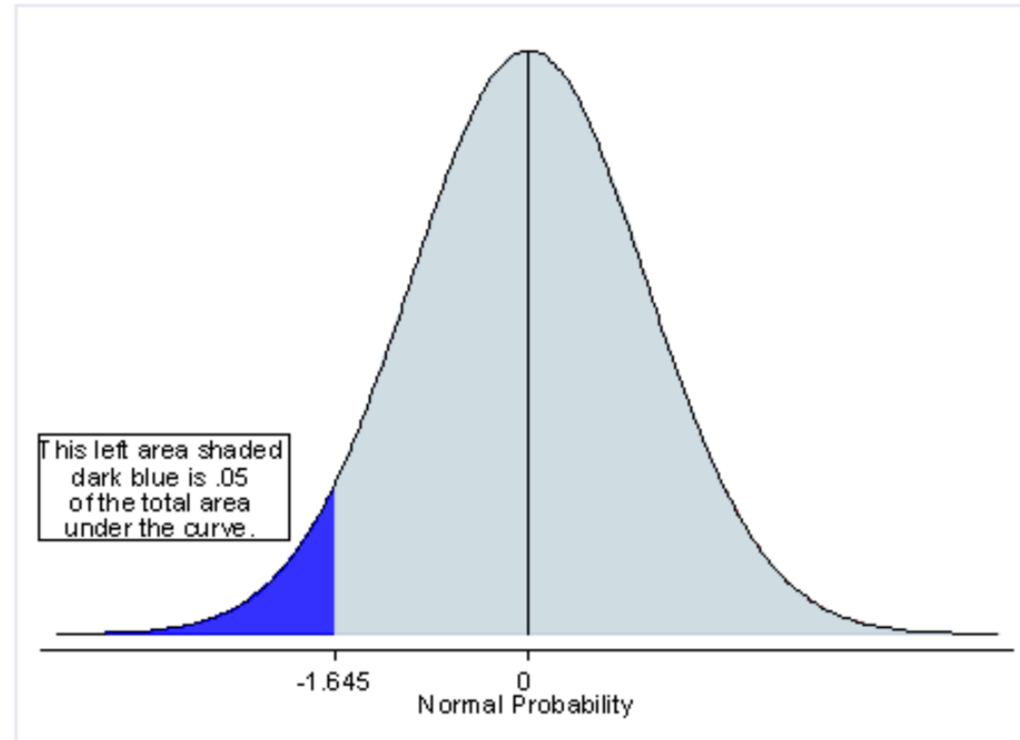
Не делает предположений о генеральной совокупности, критерий “свободен от распределения”.

Вычисление p-value

А) Для односторонних тестов

1) Левосторонний тест

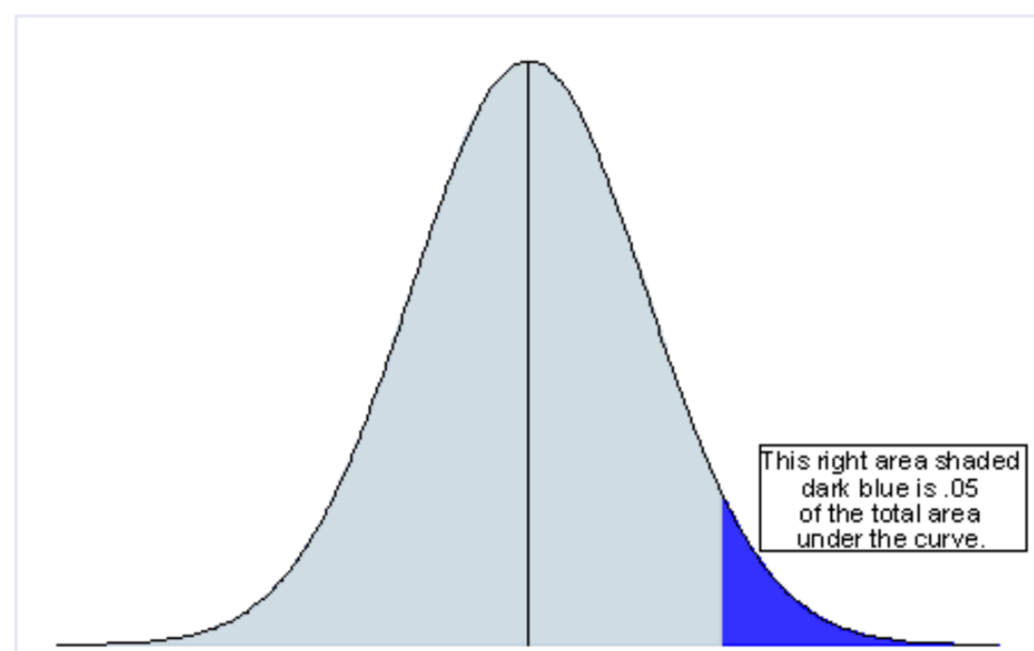
$$P(x < test\_value) = p\_value$$



2) Правосторонний тест

$$P(x > test\_value) = p\_value$$

$$1 - P(x < test\_value) = p\_value$$

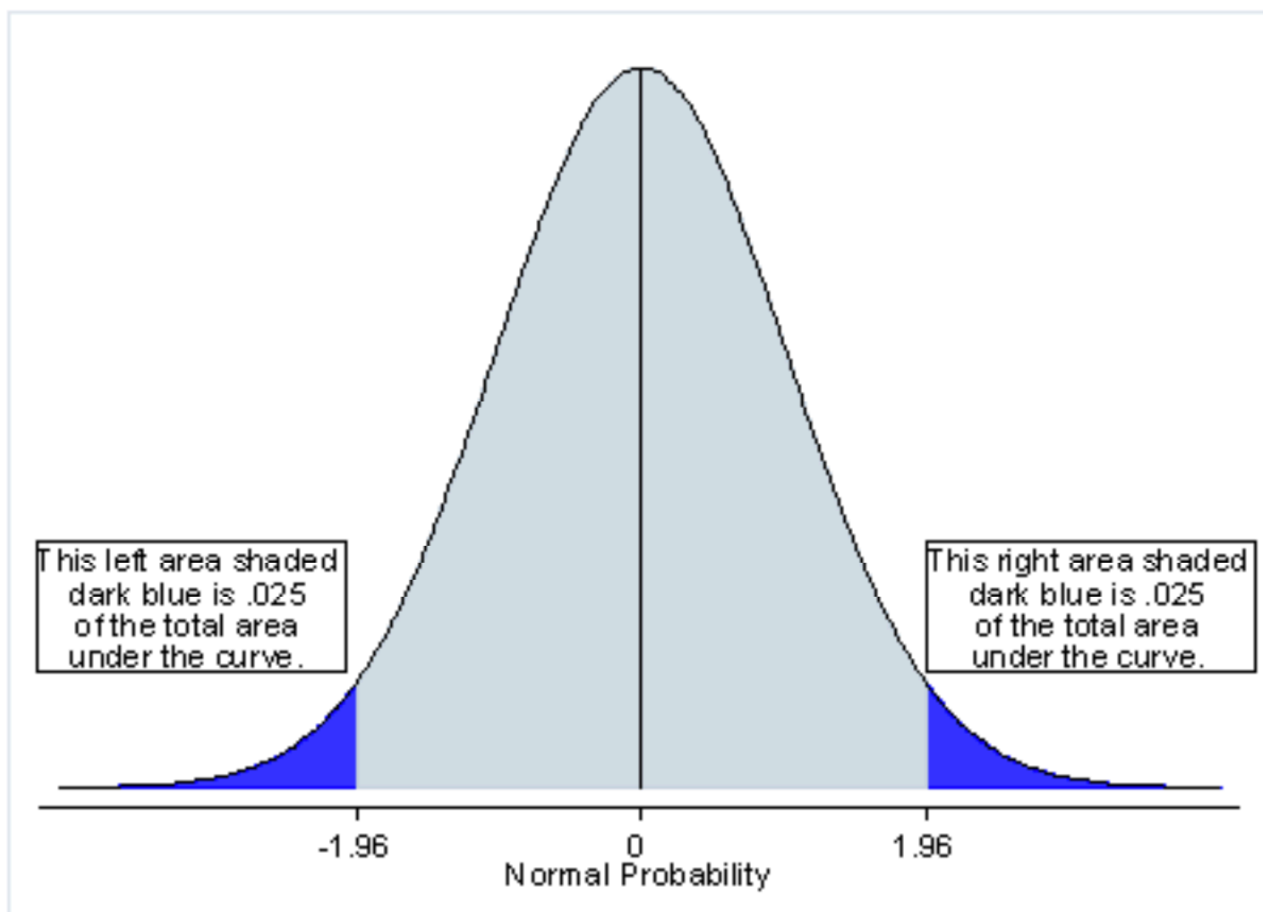


Вычисление p-value

Б) Для двусторонних тестов

$$P(x < -abs(test\_value)) + P(x > abs(test\_value)) = p\_value$$

$$P(x < -abs(test\_value)) + 1 - P(x < abs(test\_value)) = p\_value$$



Тест на равенство среднего

Какие знаете?

Тест на равенство среднего

Пусть N < 40.

Если X1, … XN распределены одинаково нормально и независимо, то их сумма тоже будет распределена нормально, а, значит и среднее будет распределено нормально

$$X_1, \ldots X_N \sim N(\mu, \sigma^2)$$

$$\sum X_i^N \sim N(\mu * N, \sigma^2 * N)$$

$$\overline{X} \sim N(?, ?)$$

Тест на равенство среднего

Пусть N < 40.
Если X1, … XN распределены одинаково нормально и независимо, то их сумма тоже будет распределена нормально, а, значит и среднее будет распределено нормально

$$X_1, \ldots X_N \sim N(\mu, \sigma^2)$$

$$\sum X_i^N \sim N(\mu * N, \sigma^2 * N)$$

$$\overline{X} \sim N(\mu, \sigma^2/N)$$

Тест на равенство среднего

$$X_1, \ldots X_N \sim N(\mu, \sigma^2)$$

$$\sum X_i^N \sim N(\mu * N, \sigma^2 * N)$$

$$\overline{X} \sim N(\mu, \sigma^2/N)$$

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

Тест на равенство среднего

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{N}} \sim N(0,1)$$

Получается, мы можем, зная распределение выборочного среднего, давать оценку среднему генеральной совокупности

В чем разница между средним генеральной совокупности и выборочным средним?

Тест на равенство среднего

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

Получается, мы можем, зная распределение выборочного среднего, давать оценку среднему генеральной совокупности

Среднее генеральной совокупности - вполне конкретное число.

Выборочное среднее - случайная величина

Тест на равенство среднего

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

Какие проблемы с этой формулой?

Тест на равенство среднего

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{N}} \sim N(0,1)$$

Какие проблемы с этой формулой?

А известна ли нам дисперсия?

Тест на равенство среднего

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{N}} \sim N(0,1)$$

Какие проблемы с этой формулой?

А известна ли нам дисперсия? Если да - то хорошо, нормальный тест

Одновыборочный тест Стьюдента

Если нет, то мы ее оцениваем выборочной дисперсией, и тогда тест Стьюдента

$$s^2 = \frac{\sum_i^N (X_i - \overline{X})^2}{N - 1}$$

$$\frac{\overline{X} - \mu}{s/\sqrt{N}} \sim t(df)$$

df - число степеней свободы.
В данном случае равно n - 1, так как для подсчета выборочного отклонения используется выборочное же среднее

Равенство средних

$$\overline{X} \sim N(\mu_x, \sigma_x^2/N_x)$$

$$\overline{Y} \sim N(\mu_y, \sigma_y^2/N_y)$$

$$\overline{Y} - \overline{X} \sim ?$$

Равенство средних

Двухвыборочный тест (равенство средних)

$$\overline{X} \sim N(\mu_x, \sigma_x^2/N_x)$$

$$\overline{Y} \sim N(\mu_y, \sigma_y^2/N_y)$$

$$\overline{Y} - \overline{X} \sim N(\mu_y - \mu_x, \sigma_y^2/N_y + \sigma_x^2/N_x)$$

Если знаем дисперсии - то все хорошо..

$$\frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{\sqrt{\dfrac{\sigma_y^2}{N_y} + \dfrac{\sigma_x^2}{N_x}}} \sim N(0,1)$$

Двувыборочный тест Стьюдента при неравенстве дисперсий (тест Уэлча)

Если не знаем и дисперсии не равны, то (и то, это не совсем точно):

$$\frac{\overline{Y} - \overline{X} - (\mu_y - my_x)}{\sqrt{\frac{s_y^2}{N_y} + \frac{s_x^2}{N_x}}} \sim t(df) \quad df = min(N_x - 1, N_y - 1)$$

Двувыборочный тест Стьюдента при равенстве дисперсий

Если не знаем и дисперсии равны, то (и то, это не совсем точно):

$$\frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{s_{xy}\sqrt{\frac{1}{N_y} + \frac{1}{N_x}}} \sim t(df)$$

$$s_{xy} = \sqrt{\frac{(N_x - 1)s_x^2 + (N_y - 1)s_y^2}{N_x + N_y - 2}}$$

$$df = N_x + N_y - 2$$

pooled variance

Парный тест

У нас есть парные наблюдения, например, пациент до и после терапии. Достаточно предполагать, что разницы распределены нормально

$$\overline{X - Y} \sim N(\mu_{x-y}, \sigma^2_{x-y}/N_{x-y})$$

μ_{x-y}, σ_{x-y} **Именно для разниц**

Если знаем дисперсию (хотя откуда) - то нормальный тест

Если знаем дисперсию (хотя откуда) - то тест Стьюдента (одновыборочный), просто в качестве элемента выборки - разница в паре

N >= 40

Если N >= 40, то даже не предполагая нормальность генеральной совокупности, мы получаем, что

$$\overline{X} \sim N(\mu, \sigma^2/N)$$

Почему?

N >= 40

Если N >= 40, то даже не предполагая нормальность генеральной совокупности, мы получаем, что

$$\overline{X} \sim N(\mu, \sigma^2/N)$$

По ЦПТ. Ограничения на распределение есть, но они чаще всего удолетворяются.

Что тоже важно, это то, что стандартное отклонение при таком числе наблюдений начинает давать хорошую оценку дисперсии, потому мы можем считать, что мы ее знаем и всегда прибегать к нормальному тесту, а не тесту Стьюдента

N >= 40, двувыборочный тест при равенстве дисперсий

$$\frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{s_{xy}\sqrt{\frac{1}{N_y} + \frac{1}{N_x}}} \sim N(0,1)$$

$$s_{xy} = \sqrt{\frac{(N_x - 1)s_x^2 + (N_y - 1)s_y^2}{N_x + N_y - 2}}$$