Статистические тесты

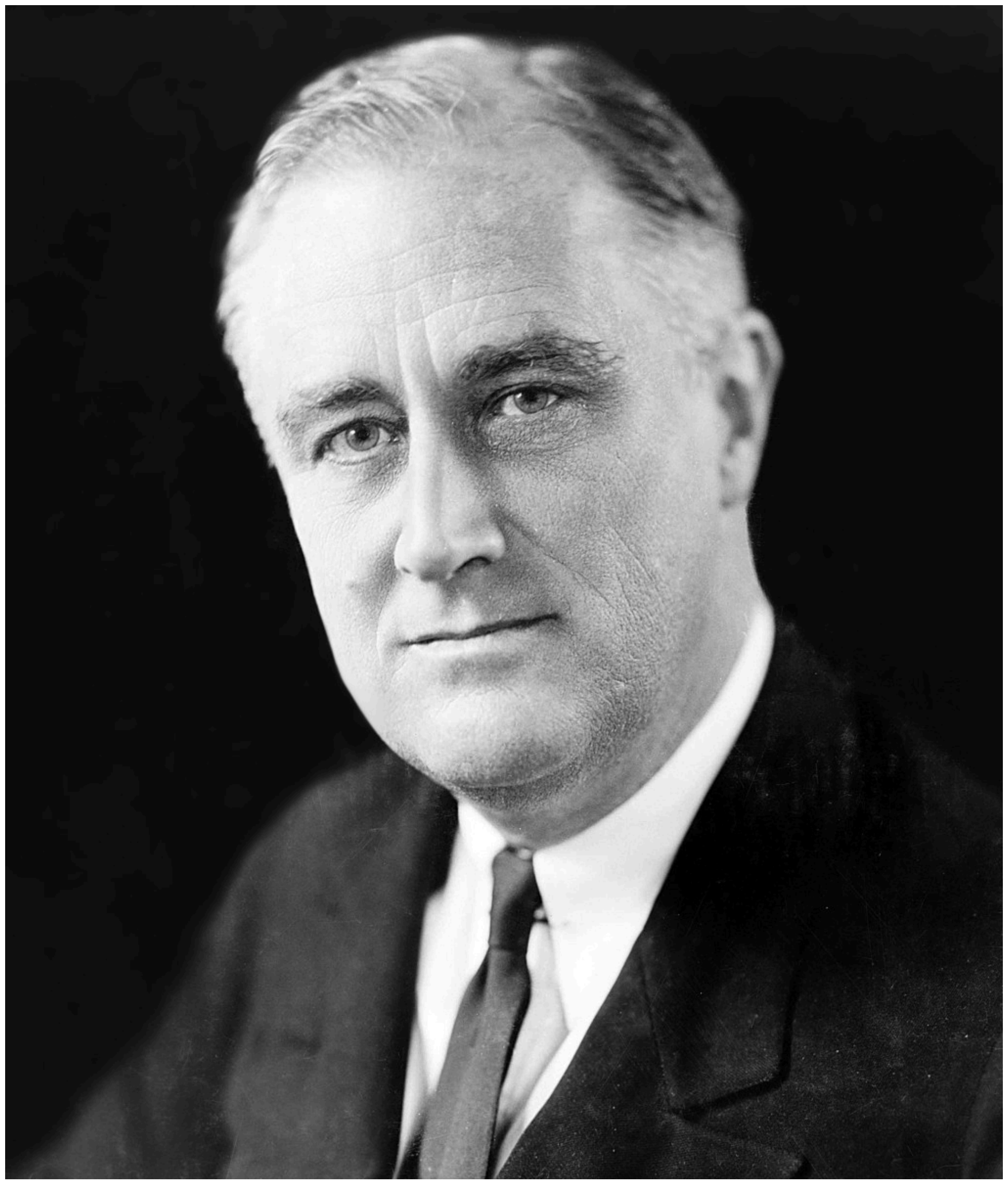
Понятие выборки

Генеральная совокупность (в англ. — population) — совокупность всех объектов (единиц), относительно которых учёный намерен делать выводы при изучении конкретной проблемы.

Выборка или выборочная совокупность — множество случаев (испытуемых, объектов, событий, образцов), с помощью определённой процедуры выбранных из генеральной совокупности для участия в исследовании.

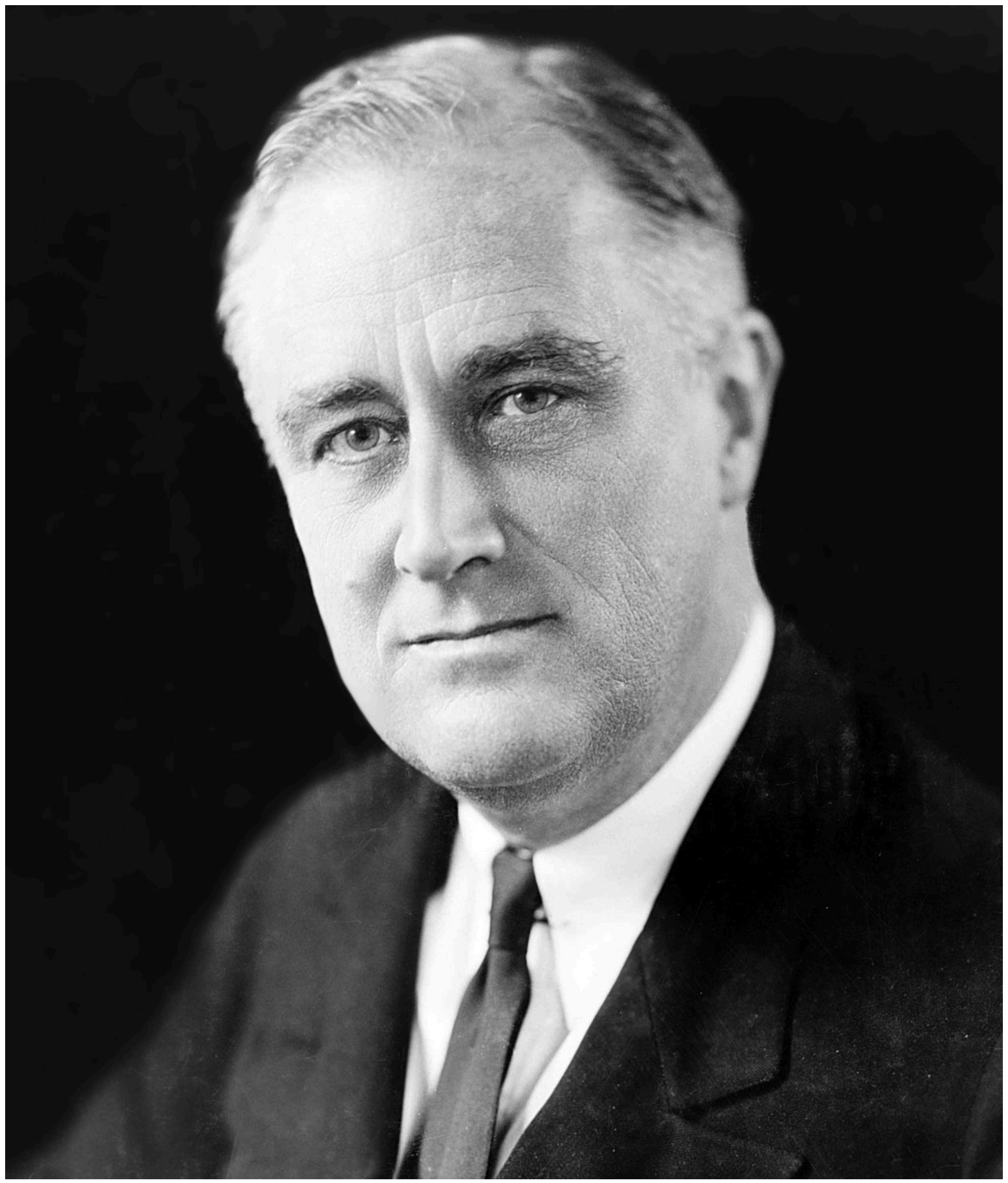
Репрезентативность - выборка может рассматриваться в качестве репрезентативной или нерепрезентативной

Пример важности репрезентативности



Журнал “Литерари Дайджес” - 2 млн собранных подписей, победа Альфреда Лондона
Гэллап - 3000 подписей - победа второго кандидата. Кому поверить?

Пример важности репрезентативности



Журнал "Литерари Дайджес" - 2 млн собранных подписей у владельцев телефонов (роскошь на тот момент), победа Альфреда Лондона

Гэллап - 3000 подписей - победа Рузвельта. Предсказал и то, что "Литерари Дайджес" ошибочно предскажет победу Лондона.

Тренировка

1. Необходимо собрать репрезентативную выборку людей по весу и по размеру используемой посуды;
2. Необходимо собрать выборку людей по политическому спектру (левые-правые);
3. Необходимо собрать выборку птиц по месту гнездования;
4. Необходимо собрать выборку людей по их отношению к курению;
5. Необходимо собрать выборку людей по употреблению ими легких наркотиков.

Тестирование гипотез

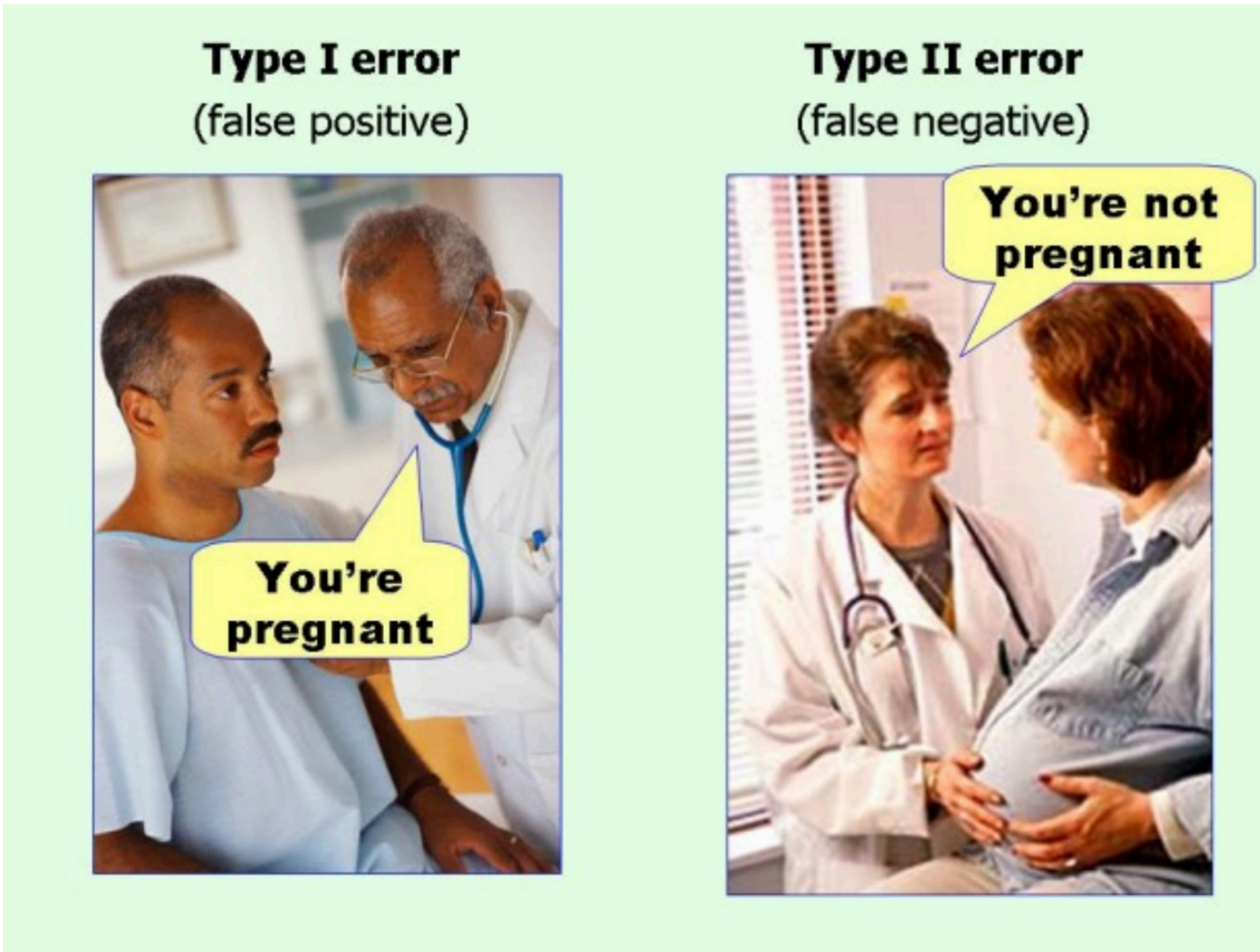
Нулевая гипотеза (H0) – это основное проверяемое предположение, которое обычно формулируется как отсутствие различий, отсутствие влияние фактора, отсутствие эффекта, равенство нулю значений выборочных характеристик и т.п.
Примером нулевой гипотезы в педагогике является утверждение о том, что различие 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;
7) Есть две группы людей - одна ест всю еду палочками, вторая есть наиболее удобным способом. В течении месяца с помощью тестов (дающих число в заданном диапазоне) измерялся уровень агрессивности данных людей.

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 C 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 C, выздоровели от простуды раньше. Гипотезы:

H0: P (витамин C лучше) = ½
H1: P (витамин C 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 C.

Как оценить насколько подтвердилась наша гипотеза?

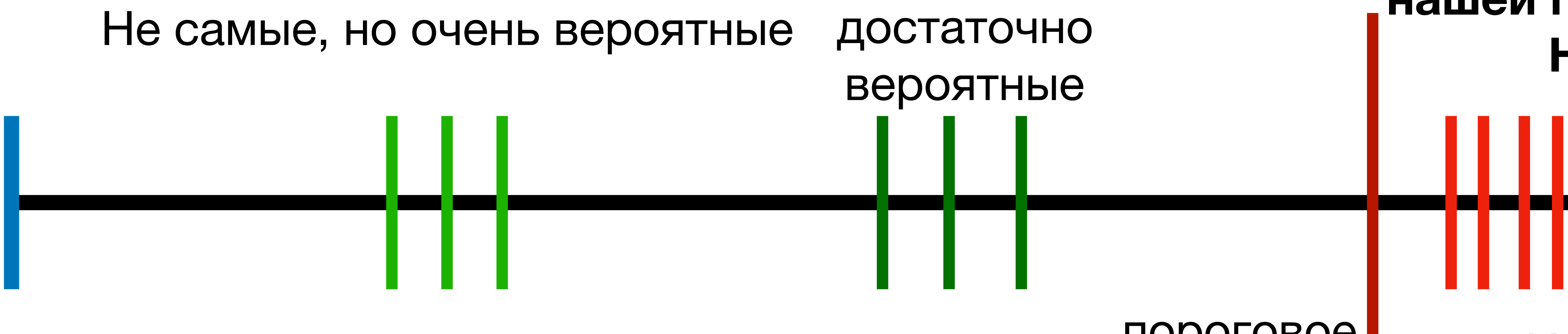
Задача

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

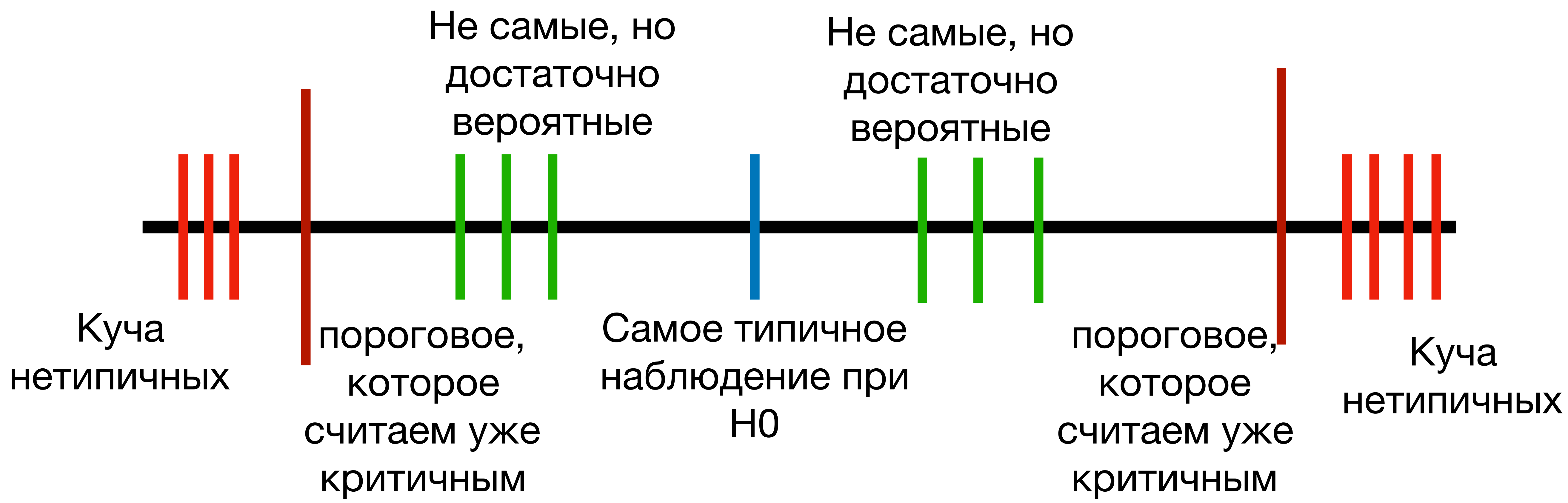
Можно посчитать вероятность нашего наблюдения при условии H0:

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

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.



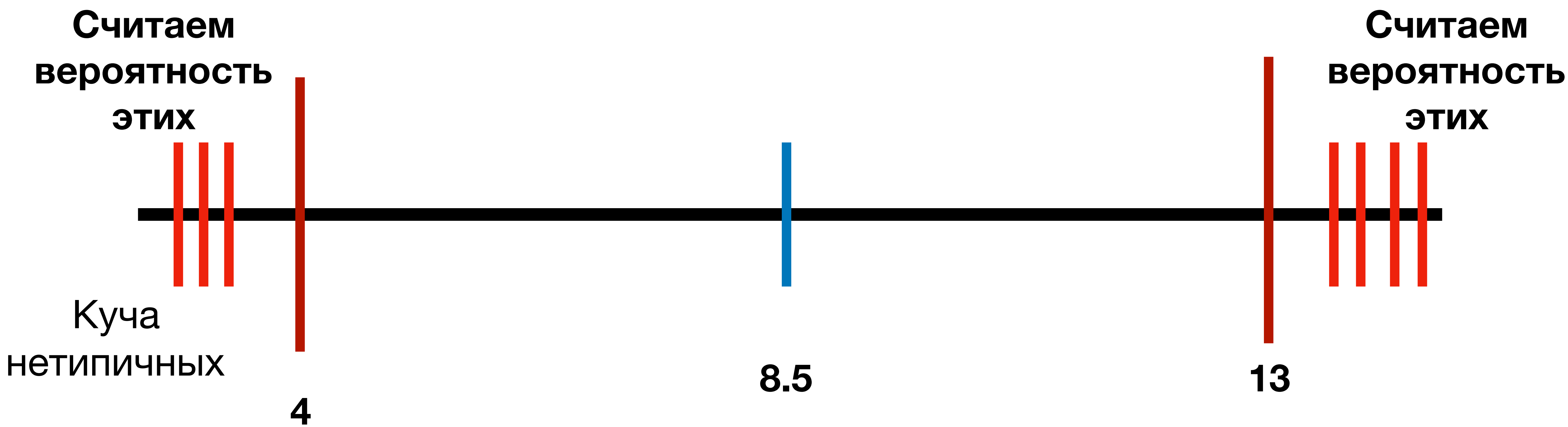
Задача



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.



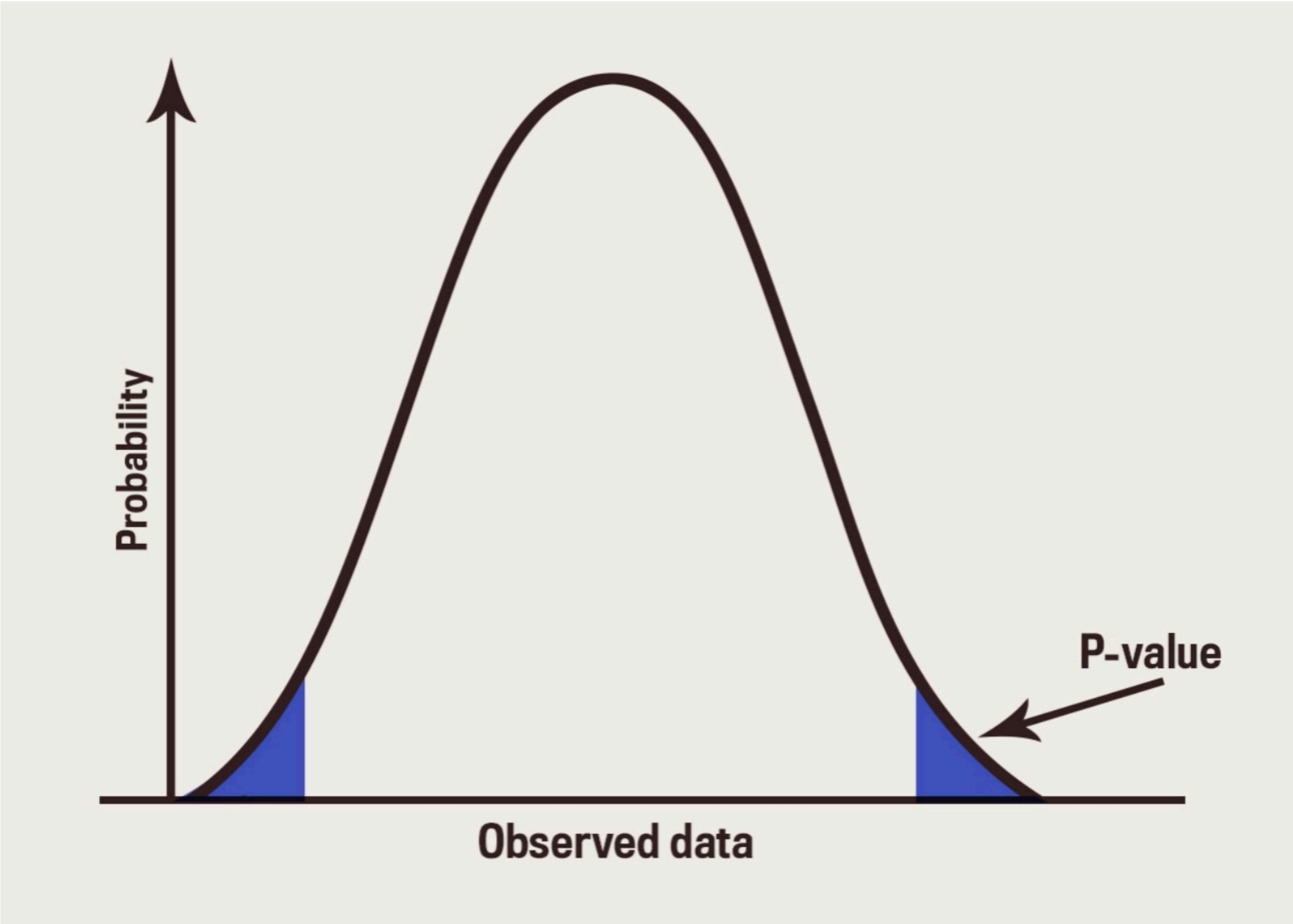
$$\sum_{i=13}^{17} C_{17}^{i} p^{i}(1-p)^{17-i} + \sum_{i=0}^{4} C_{17}^{i} p^{i}(1-p)^{17-i} = 0.049$$

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

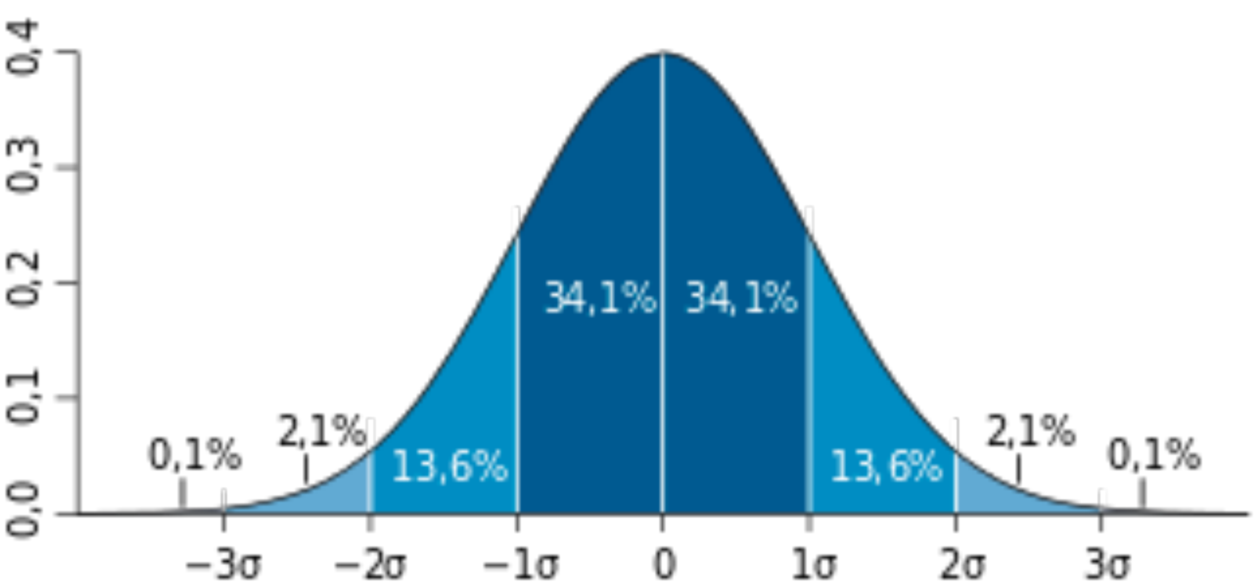
Другими словами - если нулевая гипотеза верна, то насколько вероятно получить ту выборку, которую мы получили, или более критичную



Тесты

Параметрические

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



Непараметрические

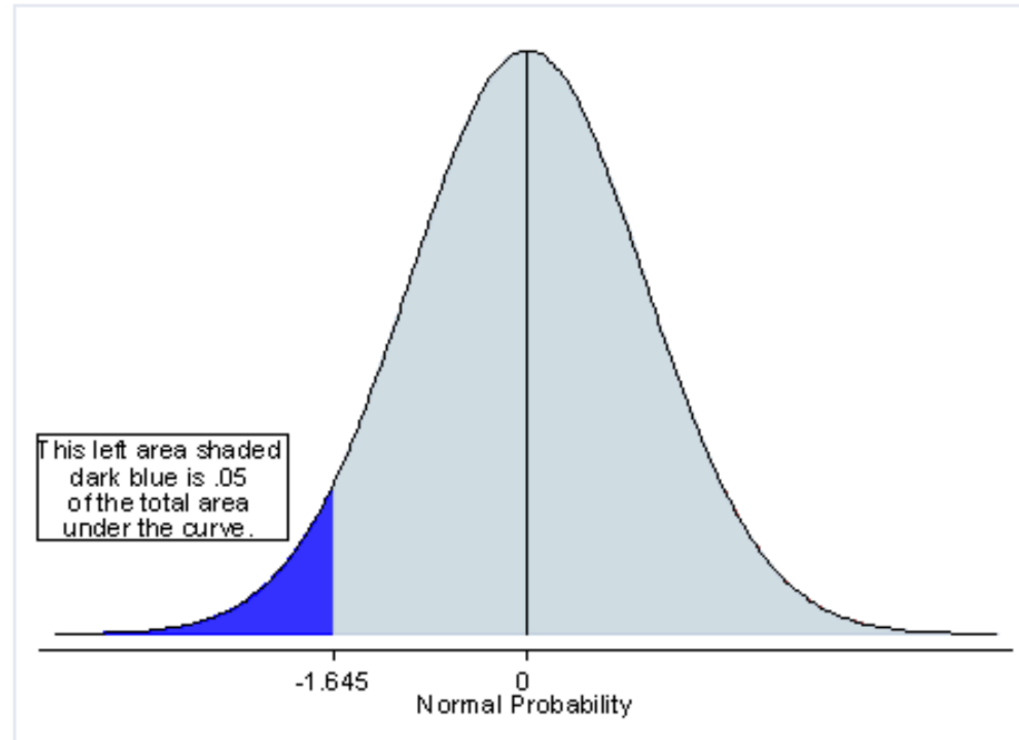
Не делает предположений о генеральной совокупности, критерий “свободен от распределения”.

Вычисление p-value

А) Для односторонних тестов

1) Левосторонний тест

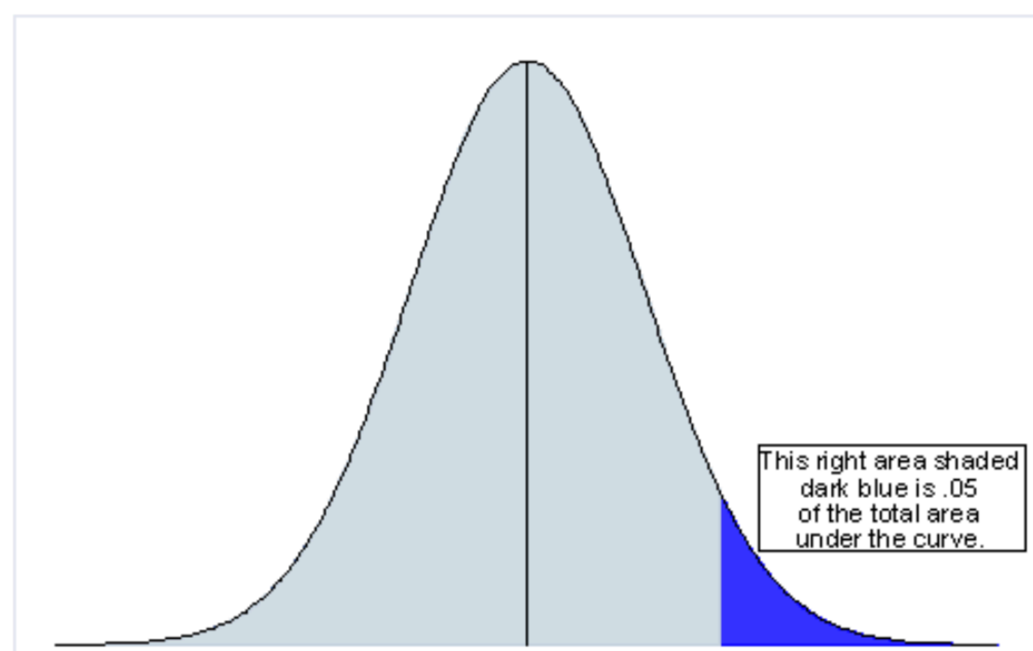
$$P(x < test\_value) = p\_value$$



2) Правосторонний тест

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

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

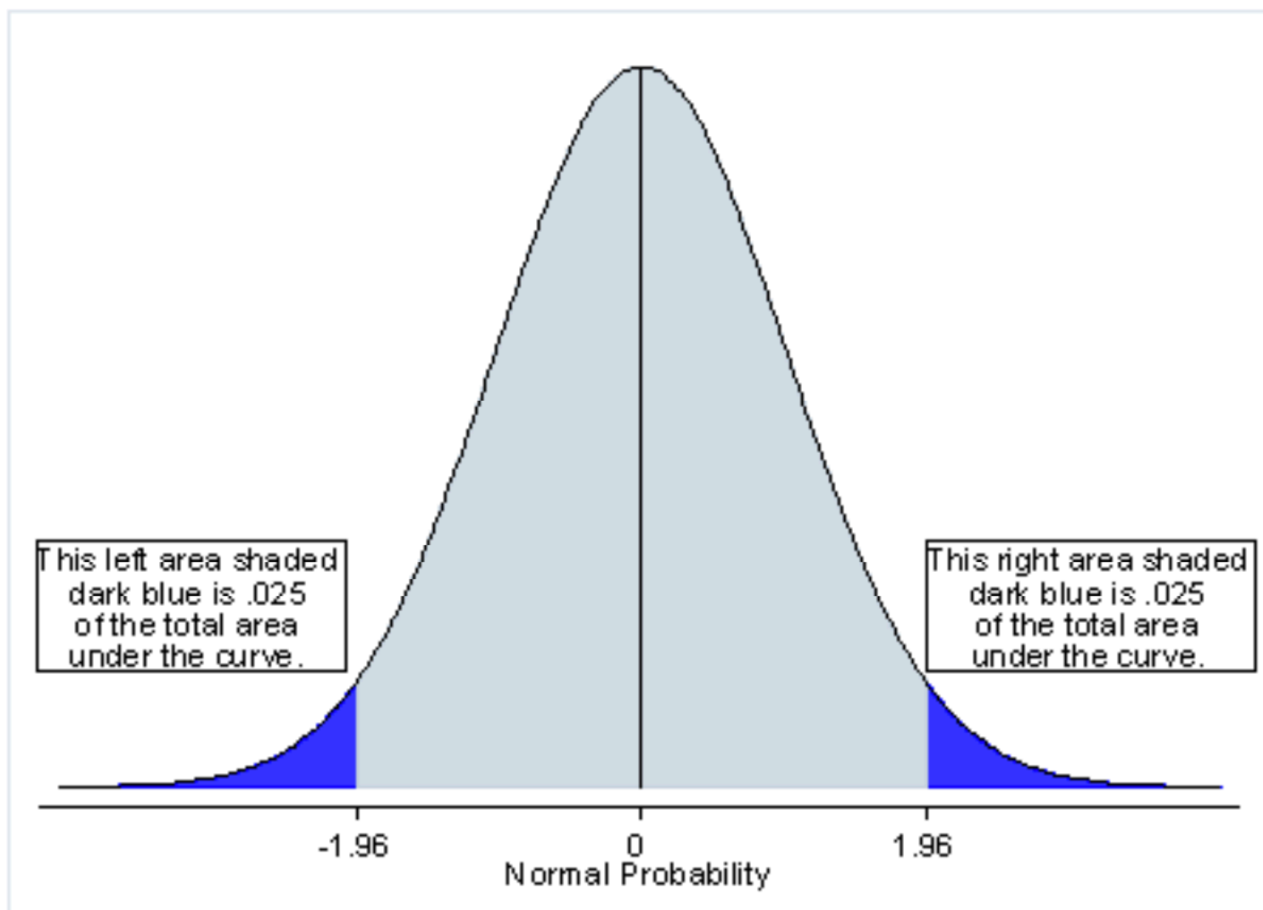


Вычисление p-value

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

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

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

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

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

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

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

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

R. t-test

Student's t-Test

Description

Performs one and two sample t-tests on vectors of data.

Usage

```
t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

Все тесты Стьюдента в одном флаконе

Одновыборочный тест

```
human_height <- starwars[starwars$species == "Human", ]$height
t.test(human_height, mu=180, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  human_height
## t = -1.4899, df = 30, p-value = 0.1467
## alternative hypothesis: true mean is not equal to 180
## 95 percent confidence interval:
##  172.0466 181.2437
## sample estimates:
## mean of x
##  176.6452
```

Двухвыборочный тест

```
human_height <- starwars[starwars$species == "Human", ]$height
not_human_height <- starwars[starwars$species != "Human", ]$height
t.test(x=human_height, y=not_human_height, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  human_height and not_human_height
## t = 0.37878, df = 55.468, p-value = 0.7063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.06723  16.22712
## sample estimates:
## mean of x mean of y
##  176.6452  174.0652
```

Важно!

В R из x вычитается y

Двухвыборочный тест

```
human_height_eyeblue <- starwars[starwars$species == "Human" &
                                  starwars$eye_color == "blue", ]$height
human_height_eyenotblue <- starwars[starwars$species == "Human" &
                                     starwars$eye_color != "blue", ]$height
t.test(x=human_height_eyeblue, y=human_height_eyenotblue,
       alternative = "two.sided", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  human_height_eyeblue and human_height_eyenotblue
## t = -0.54474, df = 29, p-value = 0.5901
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.115584   7.019092
## sample estimates:
## mean of x mean of y
##  175.0833  177.6316
```

Двухвыборочный тест, левосторонняя альтернатива

```r
human_height_eyeblue <- starwars[starwars$species == "Human" &
                                     starwars$eye_color == "blue", ]$height
human_height_eyenotblue <- starwars[starwars$species == "Human" &
                                        starwars$eye_color != "blue", ]$height
t.test(x=human_height_eyeblue, y=human_height_eyenotblue,
       alternative = "less", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  human_height_eyeblue and human_height_eyenotblue
## t = -0.54474, df = 29, p-value = 0.295
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 5.400066
## sample estimates:
## mean of x mean of y
##  175.0833  177.6316
```

Двухвыборочный тест, правосторонняя альтернатива

```
human_height_eyeblue <- starwars[starwars$species == "Human" &
                                     starwars$eye_color == "blue", ]$height
human_height_eyenotblue <- starwars[starwars$species == "Human" &
                                     starwars$eye_color != "blue", ]$height
t.test(x=human_height_eyeblue, y=human_height_eyenotblue,
       alternative = "greater", var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  human_height_eyeblue and human_height_eyenotblue
## t = -0.54474, df = 29, p-value = 0.705
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -10.49656       Inf
## sample estimates:
## mean of x mean of y
##  175.0833  177.6316
```

Двухвыборочный тест, разница равна А

```
human_height_eyeblue <- starwars[starwars$species == "Human" &
                                     starwars$eye_color == "blue", ]$height
human_height_eyenotblue <- starwars[starwars$species == "Human" &
                                        starwars$eye_color != "blue", ]$height
t.test(x=human_height_eyeblue, y=human_height_eyenotblue, alternative = "two.sided", var.equal = T,
       mu = 10)
```

```
##
##  Two Sample t-test
##
## data:  human_height_eyeblue and human_height_eyenotblue
## t = -2.6825, df = 29, p-value = 0.01194
## alternative hypothesis: true difference in means is not equal to 10
## 95 percent confidence interval:
##  -12.115584   7.019092
## sample estimates:
## mean of x mean of y
##  175.0833  177.6316
```

Парный тест

```r
# Weight of the mice before treatment
before <-c(200.1, 190.9, 192.7, 213, 241.4, 196.9, 172.2, 185.5, 205.2, 193.7)
# Weight of the mice after treatment
after <-c(392.9, 393.2, 345.1, 393, 434, 427.9, 422, 383.9, 392.3, 352.2)

t.test(x=before, y=after, alternative = "two.sided", paired = T)
```

```
##
##  Paired t-test
##
## data:  before and after
## t = -20.883, df = 9, p-value = 6.2e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -215.5581 -173.4219
## sample estimates:
## mean of the differences
##                 -194.49
```

R. z-test

А нет его в стандартной библиотеке

install.packages(“BSDA")
librarry(“BSDA”)

Функционал почти такой-же, не умеет обсчитывать Равенство дисперсий.

Возьмите датасет airquality.
Выберите наблюдения за 5й и 6й месяцы. Какой тест вы примените для того, чтобы проверить гипотезу, что среднее значение Ozone в этих двух группах отличаются?
Напишите p-value, которое вам выдал тест и напишите ваш вывод на уровне значимости 0.05.