

Анализ мощности

Линейные модели, осень 2014

Марина Варфоломеева

Каф. Зоологии беспозвоночных, СПбГУ

Экономим силы с помощью анализа мощности

- Статистические ошибки при проверке гипотез
- Мощность статистического теста
- *A priori* анализ мощности, оценка величины эффекта
- *Post hoc* анализ мощности
- Как влиять на мощность тестов

Вы сможете

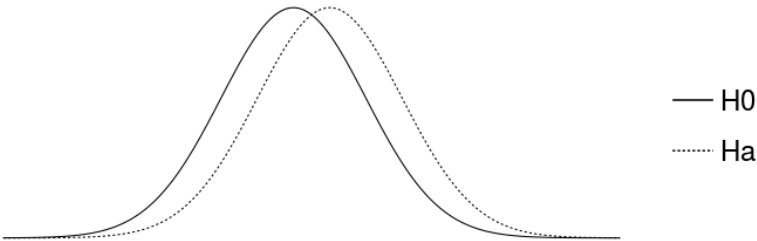
- дать определение ошибок I и II рода, мощности теста, и изобразить их вероятности на графике
- оценивать величину эффекта и необходимый объем выборки по данным пилотного исследования (*a priori* анализ мощности)
- загружать данные из .csv в R
- строить боксплоты с помощью **ggplot2**, раскрашивать их, менять темы оформления
- сравнивать средние значения при помощи *t*-критерия, интерпретировать и описывать результаты
- рассчитывать фактическую мощность теста (*post hoc* анализ мощности)

Статистические ошибки при проверке гипотез

Типы ошибок при проверке гипотез

	$H_0 == \text{TRUE}$	$H_0 == \text{FALSE}$
Отклонить H_0	Ошибка I рода Ложно-положительный результат	Верно Положительный результат
Сохранить H_0	Верно Отрицательный результат	Ошибка II рода Ложно-отрицательный результат

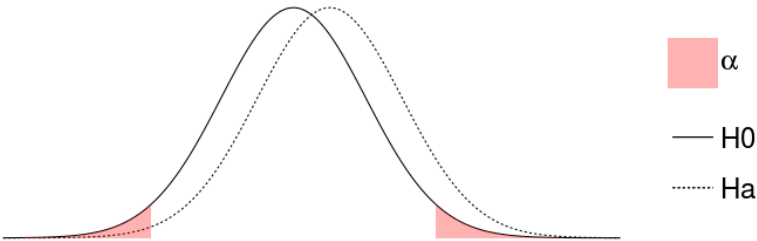
Вероятности гипотез



Типы ошибок при проверке гипотез

	$H_0 == \text{TRUE}$	$H_0 == \text{FALSE}$
Отклонить H_0	Ошибка I рода Ложно-положительный результат	Верно Положительный результат
Сохранить H_0	Верно Отрицательный результат	Ошибка II рода Ложно-отрицательный результат

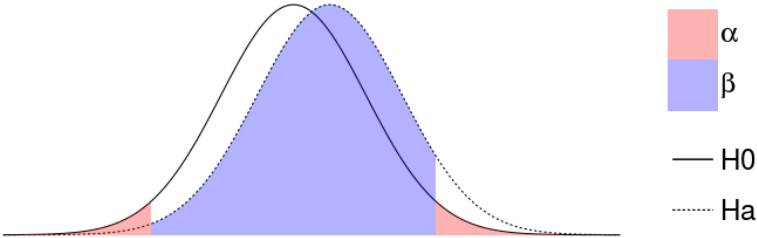
Ошибки I рода



Типы ошибок при проверке гипотез

	$H_0 == \text{TRUE}$	$H_0 == \text{FALSE}$
Отклонить H_0	Ошибка I рода Ложно-положительный результат	Верно Положительный результат
Сохранить H_0	Верно Отрицательный результат	Ошибка II рода Ложно-отрицательный результат

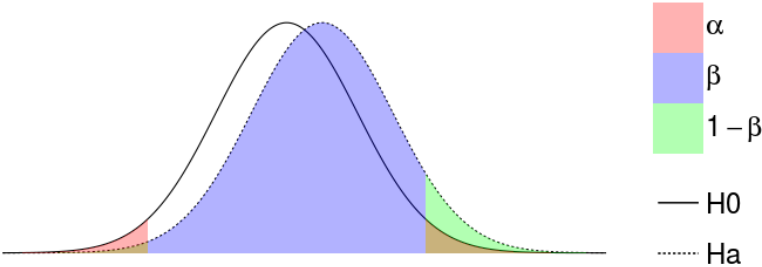
Ошибки II рода



Типы ошибок при проверке гипотез

	H0 == TRUE	H0 == FALSE
Отклонить H0	Ошибка I рода Ложно-положительный результат	Верно Положительный результат
Сохранить H0	Верно Отрицательный результат	Ошибка II рода Ложно-отрицательный результат

Мощность теста - способность выявлять различия $Power = 1 - \beta$



Анализ мощности

A priori

- какой нужен объем выборки, чтобы найти различия с разумной долей уверенности?
- различия какой величины мы можем найти, если известен объем выборки?

Post hoc

- смогли бы мы найти различия при помощи нашего эксперимента (α , n), если бы величина эффекта была X ?

A priori анализ мощности

Пример: Заповедник спасает халиотисов *

Лов халиотисов (коммерческий и любительский) запретили, организовав заповедник.

Стало ли больше моллюсков через несколько лет? (Keough, King, 1991)

Для *a priori* анализа нужно знать

- тест - *t*-критерий
- уровень значимости - $\alpha = 0.05$
- желаемая мощность теста - 80%
- ожидаемая величина эффекта - ?

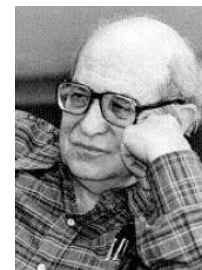
* - Данные из Quinn, Keough, 2002, Box 9-5, Fig 9-7

Как оценить ожидаемую величину эффекта?

d Коэна (Cohen's d)

$$d = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sigma}$$

- Пилотные исследования
- Литература
- Общебиологические знания
- Технические требования



Яков Коэн

Как оценить стандартное отклонение для расчета величины эффекта?

- как среднеквадратичное стандартное отклонение (d Коэна)
- как обобщенное стандартное отклонение (g Хеджа)

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

$$g = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Величина эффекта из общих соображений

Яков Коэн (1982)

сильные, умеренные и слабые эффекты

```
library(pwr)  
cohen.ES(test = "t", size = "large")
```

```
#  
#   Conventional effect size from Cohen (1982)  
#  
#       test = t  
#       size = large  
#   effect.size = 0.8
```

Задача:

Рассчитайте величину умеренных и слабых эффектов для t-критерия

```
library()  
cohen.ES()
```

Подсказка: обозначения можно посмотреть в файлах справки

```
help(cohen.ES)  
?cohen.ES  
cohen.ES # курсор на слове, нажать F1
```

Величина эффекта из пилотных данных

$$d = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sigma}$$

σ - стандартное отклонение плотности халиотисов:

- Плотность крупных халиотисов на 50м^2 была $\bar{x} = 47.5$, $SD = 27.7$

$\bar{\mu}_1 - \bar{\mu}_2$ - средний вылов халиотисов в год:

- Масса выловленных коммерческим способом + данные о размерах -> численность -> плотность -> коммерческий лов = 11.6 экз. м^{-2}
- -> общий вылов = 23.2 экз. м^{-2} (если любительский и коммерческий лов равны)

Величина эффекта из пилотных данных

$$d = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sigma}$$

σ - стандартное отклонение плотности халиотисов:

- Плотность крупных халиотисов на 50м^2 была $\bar{x} = 47.5$, $SD = 27.7$

$\bar{\mu}_1 - \bar{\mu}_2$ - средний вылов халиотисов в год:

- Масса выловленных коммерческим способом + данные о размерах -> численность -> плотность -> коммерческий лов = 11.6 экз. м^{-2}
- -> общий вылов = 23.2 экз. м^{-2} (если любительский и коммерческий лов равны)

```
alpha <- 0.05
power <- 0.80
sigma <- 27.7 # варьирование плотности халиотисов
diff <- 23.2 # ожидаемые различия плотности халиотисов
(effect <- diff/sigma) # величина эффекта
```

```
# [1] 0.838
```

Считаем объем выборки

Функции для анализа мощности t-критерия

- при одинаковых объемах групп `pwr.t.test()`
- при разных объемах групп `pwr.t2n.test()`

```
pwr.t.test(n = NULL, d = effect, power = power, sig.level = alpha,  
           type = "two.sample", alternative = "two.sided")
```

```
#  
#       Two-sample t test power calculation  
#  
#               n = 23.4  
#               d = 0.838  
#       sig.level = 0.05  
#               power = 0.8  
#       alternative = two.sided  
#  
# NOTE: n is number in *each* group
```

- Чтобы с вероятностью 0.8 выявить различия плотности халиотисов в местах, где лов разрешен и запрещен, нужно обследовать **по 24 места каждого типа**, если мы верно оценили величину эффекта.

Задача:

Рассчитайте сколько нужно обследовать мест, чтобы обнаружить слабый эффект с вероятностью 0.8, при уровне значимости 0.01

```
cohen.ES()  
pwr.t.test()
```

Решение

```
cohen.ES(test = "t", size = "small")
```

```
#  
#      Conventional effect size from Cohen (1982)  
#  
#      test = t  
#      size = small  
#      effect.size = 0.2
```

```
pwr.t.test(n = NULL, d = 0.2, power = 0.8, sig.level = 0.01,  
           type = "two.sample", alternative = "two.sided")
```

```
#  
#      Two-sample t test power calculation  
#  
#      n = 586  
#      d = 0.2  
#      sig.level = 0.01  
#      power = 0.8  
#      alternative = two.sided  
#  
# NOTE: n is number in *each* group
```

Пример: Улитки*

Улитки *Patelloida mimula* на устрицах *Saccostrea glomerata*? Minchinton, Ross, 1999

Сколько нужно проб, чтобы показать, что плотность улиток различается между сайтами?

Читаем данные из файла

Не забудьте войти в вашу директорию для матметодов

Читаем из **.csv**

```
minch <- read.table("./data/minch.csv",  
                    header = TRUE, sep = "\t")
```

Или читаем из **.xls**

```
library(XLConnect)  
minch <- readWorksheetFromFile("./data/minch.xls",  
                               sheet = 1)
```

- Кто помнит, как посмотреть, что внутри переменной **minch**?

* - Данные из Quinn, Keough, 2002, Box 9-5, Fig 9-7.



Фото: <http://users.monash.edu.au/~murray/stats/BIO4200/Eworksheets/images/oysters.jpg>

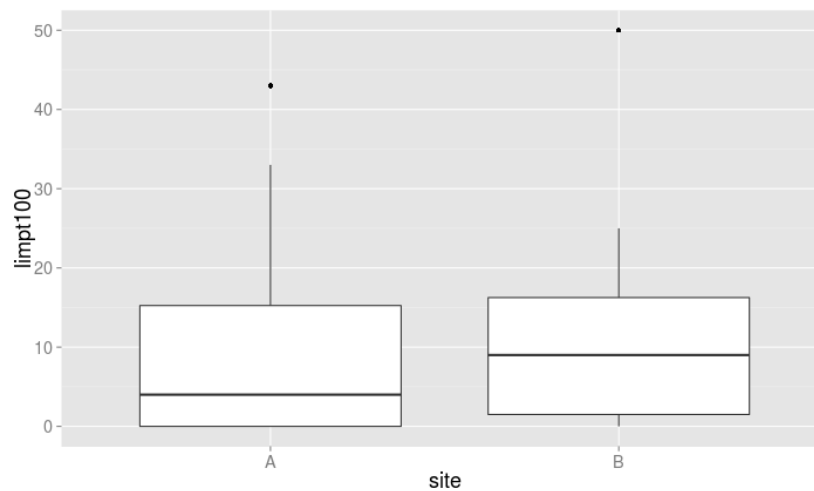
Просмотреть, что внутри переменной `minch` можно так:

```
# Структура данных
str(minch)
# Первые несколько строк
head(minch, 2)
# Первые три значения переменной zone
minch$zone[1:3]
# 2-3 строки и 3, 5, 7 столбцы
minch[2:3, c(1, 3, 5)]
# Полностью столбцы site и zone
minch[, c("site", "zone")]
```

Боксплот числа улиток в двух сайтах

Геом `geom_boxplot`

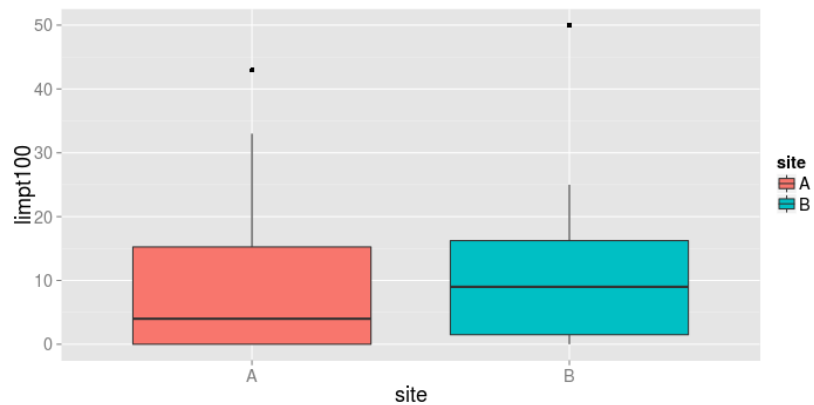
```
library(ggplot2)
gglimp <- ggplot(data = minch, aes(x = site, y = limpt100))
gglimp + geom_boxplot()
```



Раскрашиваем график

эстетика `fill`

```
gglimp + geom_boxplot(aes(fill = site))
```



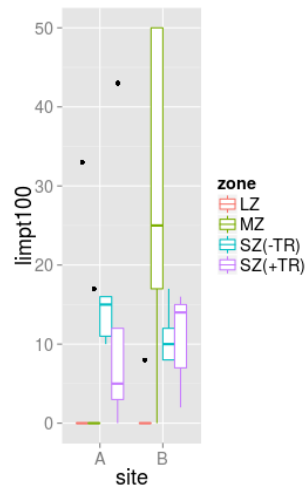
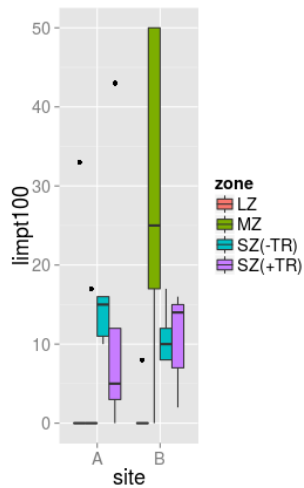
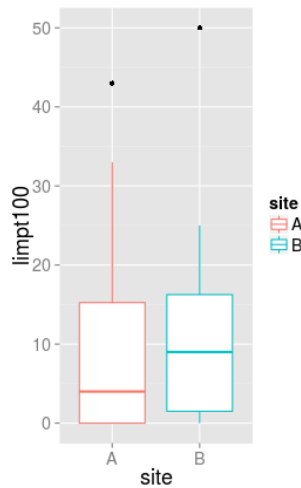
Задание: Поэкспериментируйте с эстетиками

Чем отличаются результаты применения эстетик `fill` и `colour`?

```
ggplot()  
aes()  
geom_boxplot()
```

Решение:

```
gglimp + geom_boxplot(aes(colour = site))  
gglimp + geom_boxplot(aes(fill = zone))  
gglimp + geom_boxplot(aes(colour = zone))
```



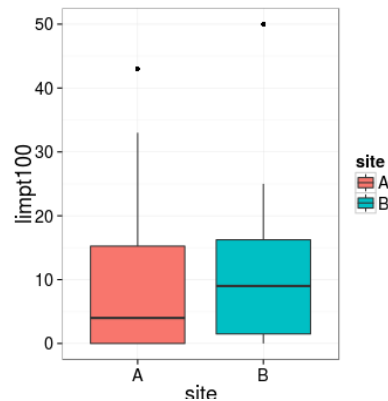
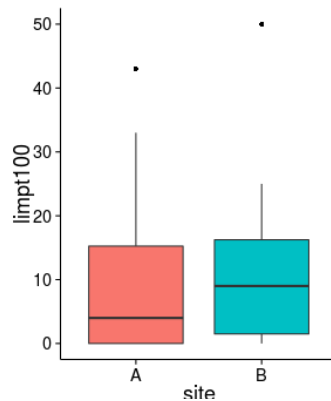
Не нравится тема? Можно привинтить другую!

Можно прибавить к графику `theme_bw()`, `theme_classic()`, `theme_grey()`, `theme_minimal()`, `theme_light()`

```
gglimp + geom_boxplot(aes(fill = site)) + theme_classic()
```

Можно установить для всех последующих графиков `theme_set()`

```
theme_set(theme_bw()) # тема до конца сеанса  
gglimp + geom_boxplot(aes(fill = site))
```



A priori анализ мощности по данным пилотного исследования

Пилотное исследование

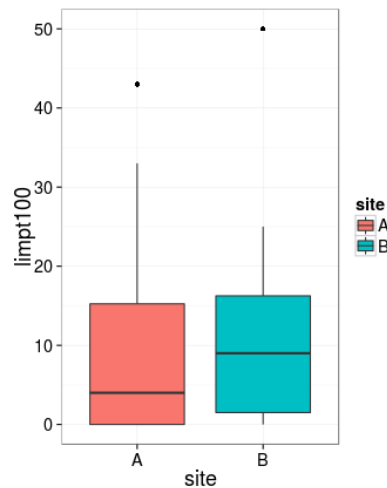
Какой объем выборки нужен, чтобы доказать что численность улиток различается между сайтами?

Величина эффекта по исходным данным

```
library(effsize)
effect <- cohen.d(minch$limpt100, minch$site)
effect
```

```
#
# Cohen's d
#
# d estimate: -0.293 (small)
# 95 percent confidence interval:
#   inf      sup
# -0.953    0.368
```

- как добыть из переменной effect значение величины эффекта?



Обращения к переменным по имени - \$

Как называется в структуре объекта элемент, где записана величина эффекта?

```
str(effect) # effect$estimate
```

```
# List of 7
# $ method      : chr "Cohen's d"
# $ name         : chr "d"
# $ estimate     : Named num -0.293
# .. attr(*, "names")= chr "A"
# $ conf.int     : Named num [1:2] -0.953 0.368
# .. attr(*, "names")= chr [1:2] "inf" "sup"
# $ var          : Named num 0.326
# .. attr(*, "names")= chr "A"
# $ conf.level   : num 0.95
# $ magnitude    : chr "small"
# - attr(*, "class")= chr "effsize"
```

```
# Для pwr.t.test() эффект должен быть положительным, поэтому вычислим модуль
effect <- abs(effect$estimate)
```

- Очень слабый эффект 0.293

Задача:

Рассчитайте объем выборки, чтобы показать различия плотности улиток между сайтами с вероятностью 0.8?

```
pwr.t.test()
```

Решение

```
pwr.t.test(n = NULL, d = effect, power = 0.8, sig.level = 0.05,  
           type = "two.sample", alternative = "two.sided")
```

```
#  
#       Two-sample t test power calculation  
#  
#           n = 184  
#           d = 0.293  
#       sig.level = 0.05  
#           power = 0.8  
#       alternative = two.sided  
#  
# NOTE: n is number in *each* group
```

- Площадок должно быть **185 с каждого сайта**, чтобы с вероятностью 0.8 обнаружить различия плотности улиток между сайтами.

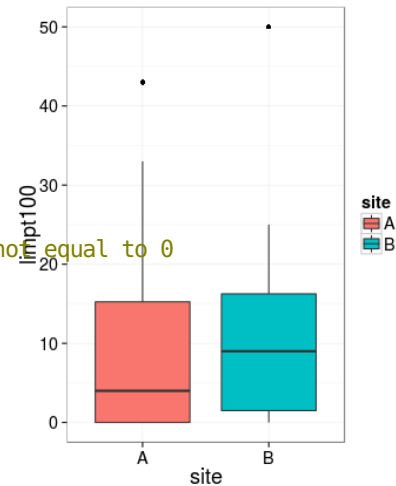
Post hoc анализ мощности

На самом деле различия действительно не были найдены

```
t.test(limpt100 ~ site, data = minch,  
       var.equal = FALSE)
```

```
#  
#  Welch Two Sample t-test  
#  
# data:  limpt100 by site  
# t = -0.926, df = 36.7, p-value = 0.3605  
# alternative hypothesis: true difference in means is not equal to 0  
# 95 percent confidence interval:  
#  -12.44   4.64  
# sample estimates:  
# mean in group A mean in group B  
#           9.05           12.95
```

- Достоверных различий плотности улиток между локациями не обнаружено (t-критерий)



Post hoc анализ - когда различий не нашли

Какова была реальная величина эффекта?

Хватило ли нам мощности, чтобы выявлять такие незначительные различия?

Для post hoc анализа нужно знать

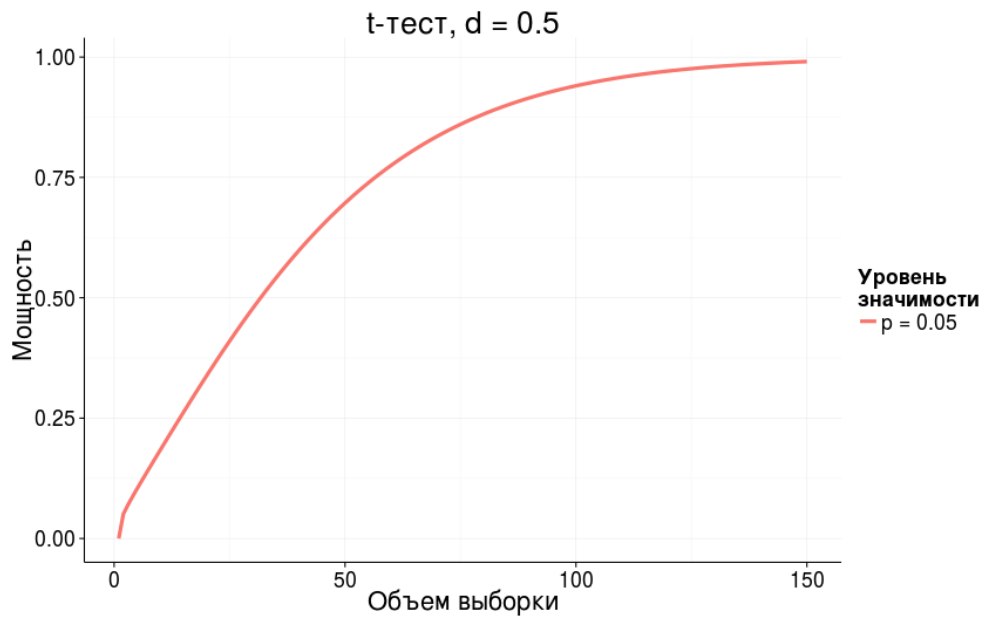
- тест (H_0 отвергнута!) — t-критерий
- уровень значимости — $\alpha = 0.05$
- фактический объем выборки — 20
- фактическая величина эффекта — ?
- реальная мощность теста - ?

```
effect_real <- cohen.d(minch$limpt100, minch$site)
effect_real <- abs(effect_real$estimate)
pwr.t.test(n = 20, d = effect_real,
           power = NULL, sig.level = 0.05,
           type = "two.sample",
           alternative = "two.sided")
```

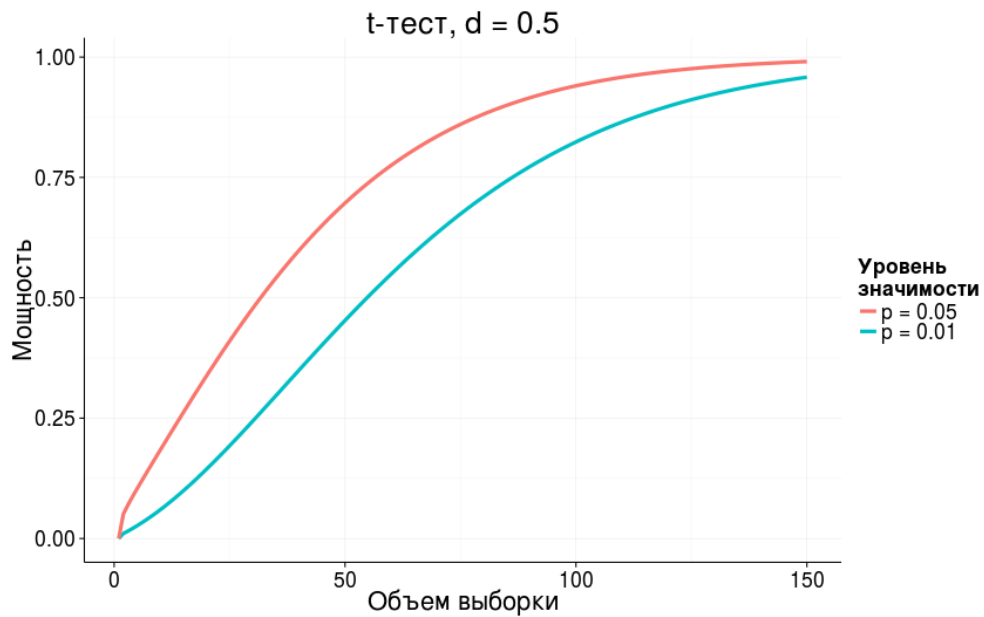
```
#
#       Two-sample t test power calculation
#
#               n = 20
#               d = 0.293
#       sig.level = 0.05
#       power = 0.147
#       alternative = two.sided
#
# NOTE: n is number in *each* group
```

Как влиять на мощность теста?

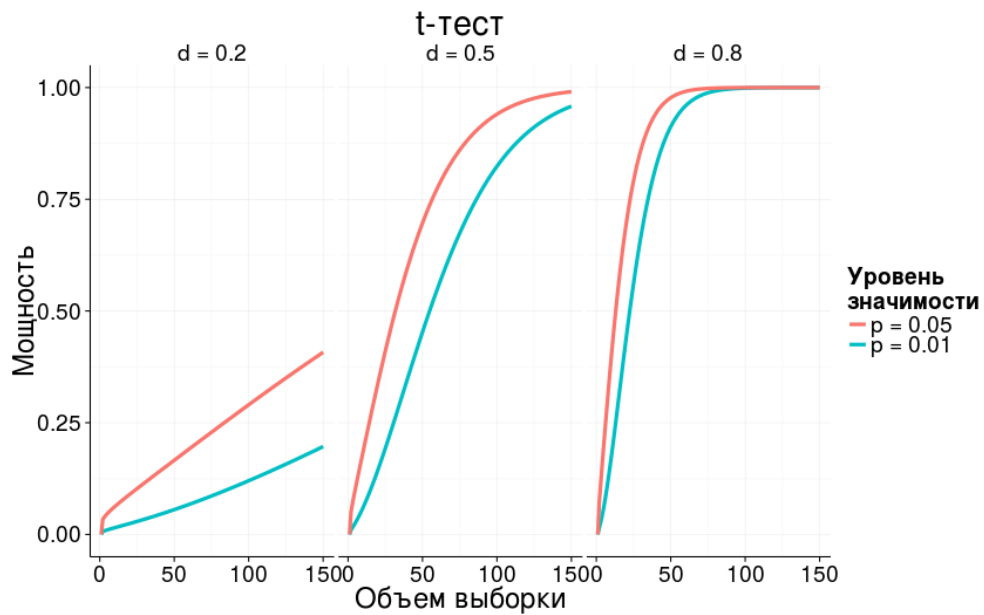
Чем больше объем выборки—тем больше мощность



Чем больше уровень значимости—тем больше мощность



Чем больше величина различий—тем больше мощность



Задание:

Какие из факторов, влияющих на мощность теста, мы **не можем** контролировать?

- Мы не можем контролировать внешние факторы
 - величину эффекта (ES)
 - фоновую изменчивость (σ^2)

Каким образом можно повлиять на мощность теста?

- Мощность теста можно регулировать, если
 - изменить число повторностей
 - выбрать другой уровень значимости (α)
 - определиться, какие эффекты действительно важны (ES)

Take home messages

- Контролируем статистические ошибки:
 - чтобы не находить несуществующих эффектов, фиксируем уровень значимости
 - чтобы не пропустить значимое, рассчитываем величину эффекта, объем выборки и мощность теста
 - когда не обнаружили достоверных эффектов, оцениваем величину эффекта и мощность теста
- Способность выявлять различия зависит
 - от объема выборки,
 - от уровня значимости
 - от величины эффекта

Дополнительные ресурсы

- Quinn, Keough, 2002, pp. 164-170
- Open Intro to Statistics: [4.6 Sample Size and Power](#), pp. 193-197
- Sokal, Rohlf, 1995, pp. 167-169.
- Zar, 1999, p. 83.
- [R Data Analysis Examples - Power Analysis for Two-group Independent sample t-test. UCLA: Statistical Consulting Group.](#)
- [R Data Analysis Examples - Power Analysis for One-sample t-test. UCLA: Statistical Consulting Group.](#)
- [FAQ - How is effect size used in power analysis? UCLA: Statistical Consulting Group.](#)