

Анализ мощности

Математические методы в зоологии с использованием R

Марина Варфоломеева

Экономим силы с помощью анализа мощности

- Тестирование гипотез (двухвыборочный t-критерий)
- Статистические ошибки при проверке гипотез
- Мощность статистического теста
- *A priori* анализ мощности, оценка величины эффекта
- Как влиять на мощность тестов

Вы сможете

- сравнивать средние значения при помощи t-критерия, интерпретировать и описывать результаты
- дать определение ошибок I и II рода, и графически изобразить их отношение к мощности теста
- оценивать величину эффекта и необходимый объем выборки по данным пилотного исследования
- загружать данные из .xls в R
- строить боксплоты с помощью ggplot2

Тестирование гипотез

Тест Стьюдента (t-критерий)

Двухвыборочный тест Стьюдента (Student, 1908) используется для проверки значимости различий между средними значениями двух нормально распределенных величин.

$$t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{SE}$$

Гипотезы: $H_0 : \mu_1 = \mu_2$, $H_A : \mu_1 \neq \mu_2$

Условия применимости:

- Выборки случайны и независимы друг от друга
- Величины нормально распределены
- Дисперсии в группах одинаковы

$$SE = \sqrt{\frac{sd_1^2(n_1 - 1) + sd_2^2(n_2 - 1)}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$df = n_1 + n_2 - 2$$

Тест Уэлча (модификация теста Стьюдента для случая разных дисперсий)

$$t = \frac{\bar{\mu}_1 - \bar{\mu}_2}{SE}$$

Условия применимости:

- Выборки случайны и независимы друг от друга
- Величины нормально распределены

$$SE = \sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}$$

Приблизительное число степеней свободы рассчитывается по уравнению Уэлча-Саттеруэйта (Welch–Satterthwaite equation)

$$df \approx \frac{\left(\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2} \right)^2}{\frac{\frac{sd_1^4}{n_1^2 \cdot df_1}}{+} \frac{\frac{sd_2^4}{n_2^2 \cdot df_2}}$$

t-распределение

t-статистика подчиняется t-распределению.

Иными словами, если много раз взять выборки **из одной** совокупности (т.е. при условии, что H_0 верна) и посчитать между ними разницу, то она будет подчиняться t-распределению.

Форма t-распределения зависит только от одного параметра — числа степеней свободы df



В хвостах этого распределения находятся редкие значения



Обычно используется уровень значимости α 0.05 или 0.01.

Уровень значимости α — это вероятность ошибочно отвергнуть справедливую нулевую гипотезу. Т.е. это вероятность найти различия там, где их нет (вероятность ошибки I рода).

Для t-теста α — это вероятность ошибочно сделать вывод о том, что средние выборок различаются **при условии, что эти выборки получены из одной генеральной совокупности.**

Тестирование гипотезы о равенстве двух средних при помощи t-критерия

- Для конкретных данных считаем значение t-критерия
- Сравниваем его с теоретическим распределением t (распределением при условии, что H_0 верна)
- Принимаем решение, отвергнуть ли H_0



Пример: Снотворное

В датасете `sleep` содержатся данные об увеличении продолжительности сна по сравнению с контролем после применения двух снотворных препаратов (Cushny, Peebles, 1905, Student, 1908)

```
data(sleep)  
View(sleep)
```

Двухвыборочный t-критерий

Сравним увеличение продолжительности сна при помощи двухвыборочного t-критерия.

```
tt <- t.test(extra ~ group, sleep)
tt
```

```
##
## Welch Two Sample t-test
##
## data:  extra by group
## t = -1.8608, df = 17.776, p-value = 0.07939
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.3654832  0.2054832
## sample estimates:
## mean in group 1 mean in group 2
##           0.75           2.33
```

Что спрятано в результатах?

Как называются отдельные элементы результатов можно узнать посмотрев их структуру при помощи функции `str()`

```
str(tt)
```

```
## List of 9
## $ statistic : Named num -1.86
## ..- attr(*, "names")= chr "t"
## $ parameter : Named num 17.8
## ..- attr(*, "names")= chr "df"
## $ p.value : num 0.0794
## $ conf.int : atomic [1:2] -3.365 0.205
## ..- attr(*, "conf.level")= num 0.95
## $ estimate : Named num [1:2] 0.75 2.33
## ..- attr(*, "names")= chr [1:2] "mean in group 1" "mean in group 2"
## $ null.value : Named num 0
## ..- attr(*, "names")= chr "difference in means"
## $ alternative: chr "two.sided"
## $ method : chr "Welch Two Sample t-test"
## $ data.name : chr "extra by group"
## - attr(*, "class")= chr "htest"
```

Можно получить элементы результатов в виде отдельных цифр

```
tt$parameter # степени свободы
```

```
##          df  
## 17.77647
```

```
tt$p.value # доверительная вероятность
```

```
## [1] 0.07939414
```

```
tt$statistic # значение t-критерия
```

```
##          t  
## -1.860813
```

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

Ошибка I рода



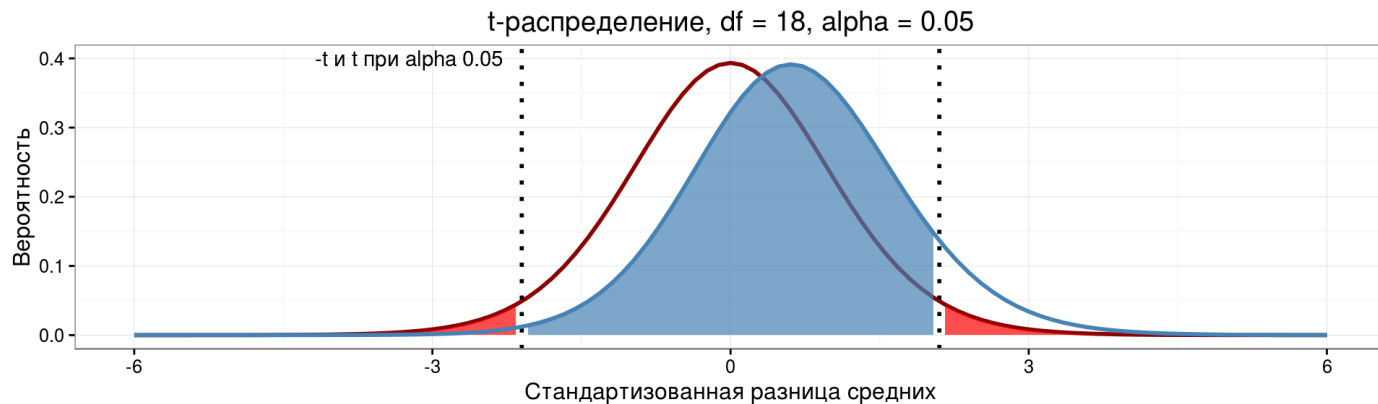
Ошибка I рода — вероятность отвергнуть H_0 , когда верна H_0

Мы этого не знаем, но может быть верна H_A ...



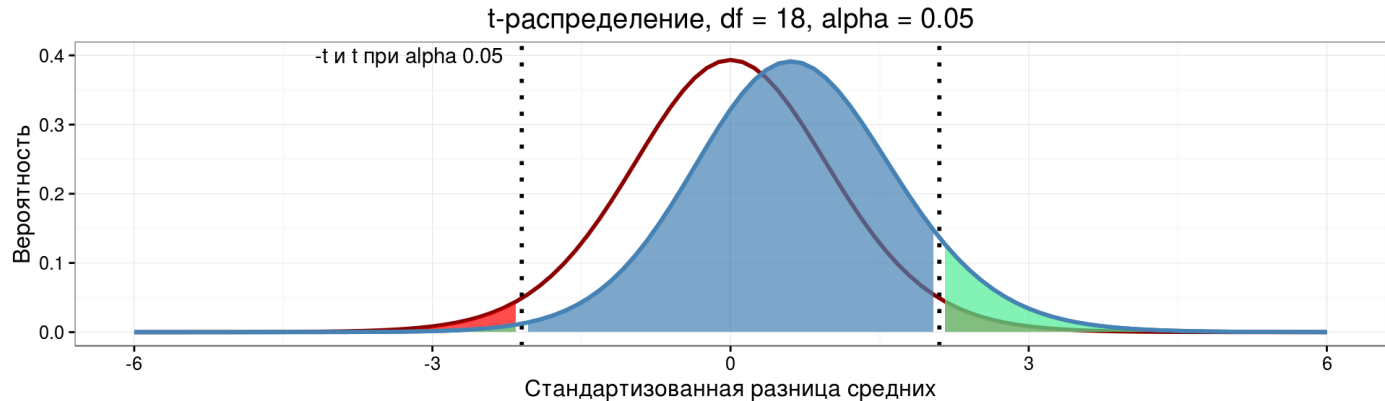
Можно построить еще одно распределение статистики — распределение, при условии того, что верна H_A

Ошибка II рода



Ошибка II рода — вероятность принять H_0 , когда верна H_A

Мощность теста — способность выявлять различия



Мощность теста - вероятность отвергнуть H_0 , когда верна H_A

$$Power = 1 - \beta$$

Обычно считается, что хорошо, когда мощность не меньше 0.8

Т.е. что в 80% случаев мы можем найти различия заданной величины, если они есть.

Анализ мощности

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

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

A priori анализ мощности

Что нужно

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

Величина эффекта

d Коэна (Cohen's d)

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

Где сигма может оцениваться одним из способов:

- среднеквадратичное отклонение (d Коэна)

$$\sigma = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

- обобщенное стандартное отклонение (g Хеджа)

$$\sigma = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Величина эффекта

Яков Коэн предложил делить эффекты на сильные, умеренные и слабые (Cohen, 1982)

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

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


Задание

Рассчитайте величину умеренных и слабых эффектов для t-критерия

Решение

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

```
##  
##      Conventional effect size from Cohen (1982)  
##  
##          test = t  
##          size = medium  
##      effect.size = 0.5
```

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

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

Величина эффекта из пилотных данных

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

σ - стандартное отклонение плотности халиотисов:

- Плотность крупных халиотисов на 50м² 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
 - Предположили, что коммерческий лов и любительский лов равны
 - Коммерческий лов = 11.6 экз. м⁻²
 - Коммерческий + любительский лов = 23.2 экз. м⁻²

Данные для анализа мощности собраны

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

```
## [1] 0.8375451
```

Функции для анализа мощности t-критерия:

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

Считаем объем выборки

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

```
pwr_half
```

```
##  
##      Two-sample t test power calculation  
##  
##              n = 23.37829  
##              d = 0.8375451  
##      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 пробы каждого типа**, если мы верно оценили величину эффекта.

Задание

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

Вам понадобятся функции `cohen.ES()` и `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 = 585.6093  
##          d = 0.2  
##      sig.level = 0.01  
##          power = 0.8  
##      alternative = two.sided  
##  
## NOTE: n is number in *each* group
```

Пример: Улитки на устрицах в мангровых зарослях*

В каких зонах мангровых зарослей на устрицах предпочитают обитать улитки (Minchinton, Ross, 1999)?

- Факторы:
 - Сайт (А и В)
 - Зона зарослей (LZ - ближе к земле, MZ - средняя часть, SZ(+TR)с деревьями, SZ(-TR) - ближе к морю, без деревьев)
- Собрали по 5 проб - число улиток на раковинах устриц

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

Скачиваем данные с сайта

Не забудьте войти в вашу директорию для матметодов при помощи `setwd()`

```
library(downloader)
```

```
# в рабочем каталоге создаем суб-директорию для данных
```

```
if(!dir.exists("data")) dir.create("data")
```

```
# скачиваем файл
```

```
download(
```

```
  url = "https://varmara.github.io/mathmethr/data/minch.xls",
```

```
  destfile = "data/minch.xls")
```

Читаем данные из файла

```
library(readxl)
minch <- read_excel(path = "data/minch.xls", sheet = 1)
```

```
str(minch) # Структура данных
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    40 obs. of  6 variables:
## $ NA      : chr  "1" "2" "3" "4" ...
## $ site     : chr  "A" "A" "A" "A" ...
## $ zone     : chr  "SZ(-TR)" "SZ(-TR)" "SZ(-TR)" "SZ(-TR)" ...
## $ limpt    : num  0.16 0.11 0.1 0.16 0.15 0.12 0 0.03 0.05 0.43 ...
## $ limpt100: num  16 11 10 16 15 12 0 3 5 43 ...
## $ sqlim100: num  4 3.32 3.16 4 3.87 ...
```

Просмотреть, что получилось можно так:

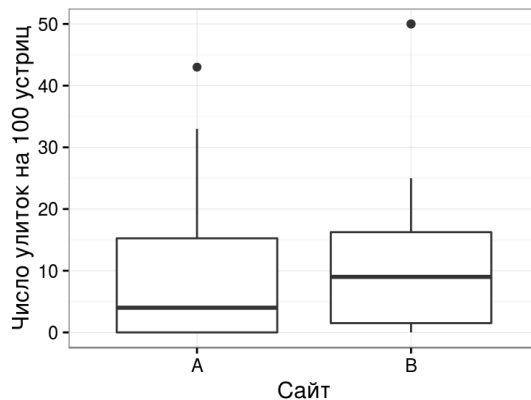
```
head(minch)      # Первые несколько строк файла
```

```
## # A tibble: 6 × 6
##   NA    site    zone limpt limpt100 sqlim100
##   <chr> <chr>   <chr> <dbl>   <dbl>   <dbl>
## 1     1     A SZ(-TR)  0.16     16     4.000
## 2     2     A SZ(-TR)  0.11     11     3.317
## 3     3     A SZ(-TR)  0.10     10     3.162
## 4     4     A SZ(-TR)  0.16     16     4.000
## 5     5     A SZ(-TR)  0.15     15     3.873
## 6     6     A SZ(+TR)  0.12     12     3.464
```

Боксплоты числа улиток

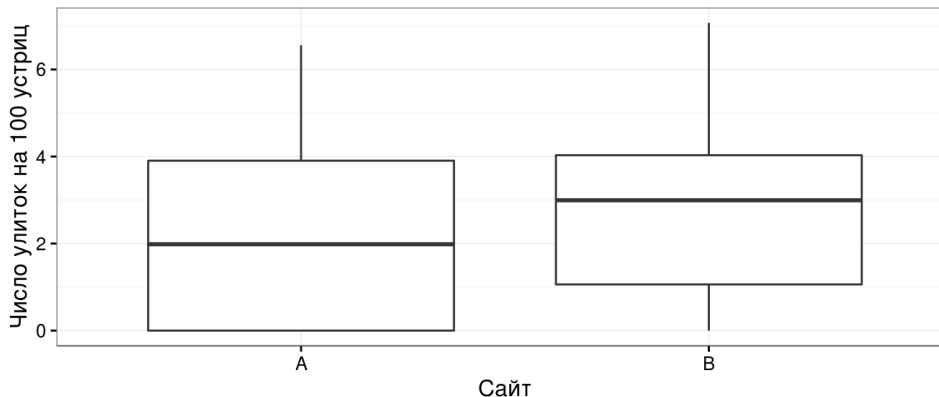
Нормальное ли тут распределение?

```
ggplot(data = minch, aes(x = site, y = limpt100)) +  
  geom_boxplot() +  
  labs(y = "Число улиток на 100 устриц",  
       x = "Сайт")
```



Боксплот корня из численности улиток (**sqlim100**) для двух сайтов

```
ggplot(data = minch, aes(x = site, y = sqlim100)) +  
  geom_boxplot() +  
  labs(y = "Число улиток на 100 устриц",  
       x = "Сайт")
```



- Распределение стало больше походить на нормальное. Можно пользоваться t-критерием для сравнения значений корня из численности улиток (**sqlim100**) для двух сайтов.

**A priori анализ мощности по
данным пилотного исследования**

Анализ мощности по данным пилотного исследования

Представим, что эти данные — это данные пилотного исследования.

Мы хотим выяснить по этим данным, сколько нужно собрать проб, чтобы показать, что плотность улиток различается на двух сайтах.

Величина эффекта по исходным данным

```
library(effsize)
eff_snail <- cohen.d(minch$sqlim100, minch$site)
eff_snail
```

```
##
## Cohen's d
##
## d estimate: -0.3650246 (small)
## 95 percent confidence interval:
##      inf      sup
## -1.027557  0.297508
```

```
# Вычислим модуль, поскольку для `pwr.t.test()` эффект должен быть положительным
effect_snail <- abs(eff_snail$estimate)
```

Задание

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

Используйте функцию `pwr.t.test()`

Решение

```
pwr_snail <- pwr.t.test(n = NULL, d = effect_snail, power = 0.8, sig.level = 0.05)
pwr_snail
```

```
##
##      Two-sample t test power calculation
##
##              n = 118.7808
##              d = 0.3650246
##      sig.level = 0.05
##              power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

- Нужна выборка **119 площадок с каждого сайта**, чтобы с вероятностью 0.8 обнаружить различия плотности улиток между сайтами.

Задание

Представьте, что в датасете **sleep** содержатся данные пилотного исследования.

Оцените, какой объем выборки нужно взять, чтобы показать, что число часов дополнительного сна после применения двух препаратов различается?

Решение

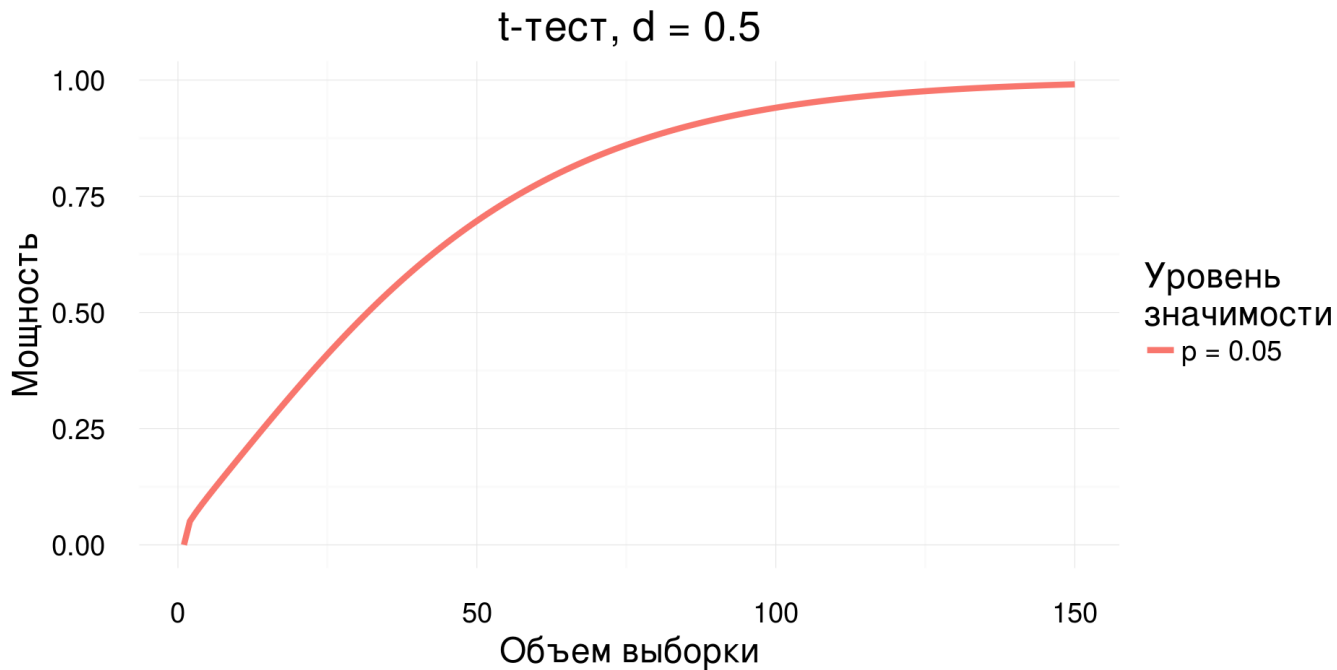
```
eff_sleep <- cohen.d(sleep$extra, sleep$group)
effect_sleep <- abs(eff_sleep$estimate)
pwr_sleep <- pwr.t.test(n = NULL, d = effect_sleep, power = 0.8, sig.level = 0.05)
pwr_sleep
```

```
##
##      Two-sample t test power calculation
##
##              n = 23.6672
##              d = 0.8321811
##      sig.level = 0.05
##              power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

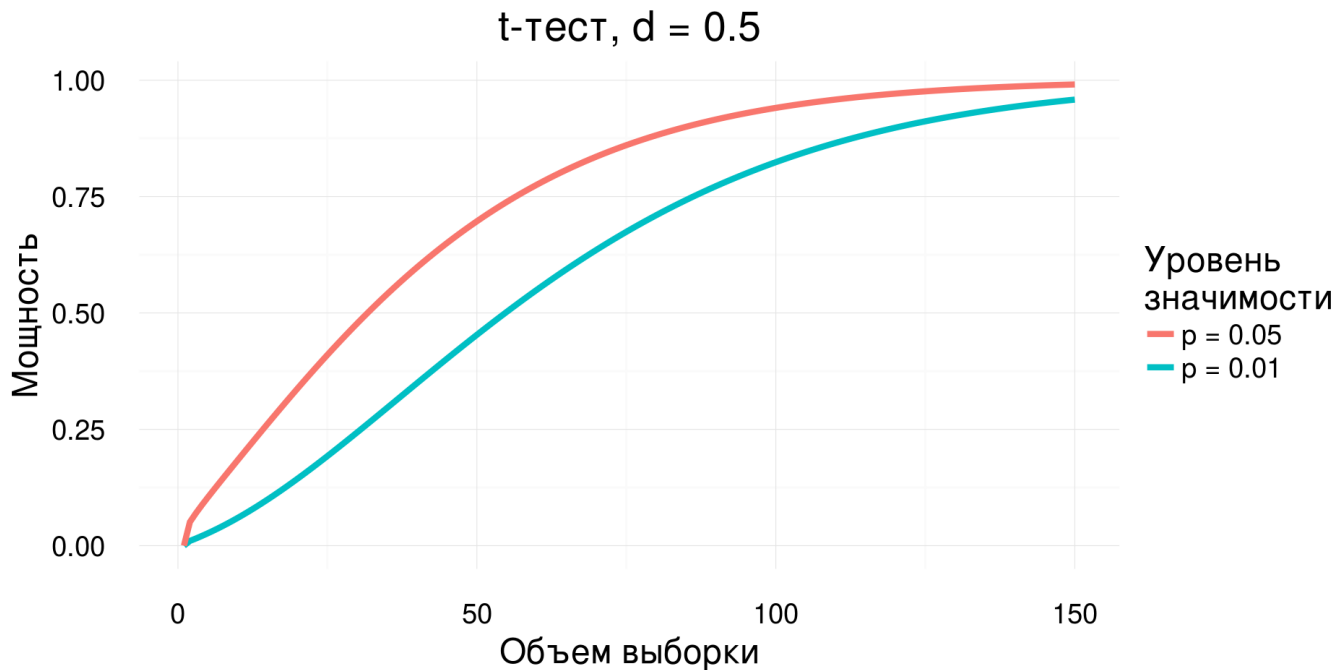
Нужна выборка **24 человека в каждой из групп**, чтобы с вероятностью 0.8 обнаружить различия числа часов дополнительного сна после применения двух препаратов.

Как влиять на мощность теста?

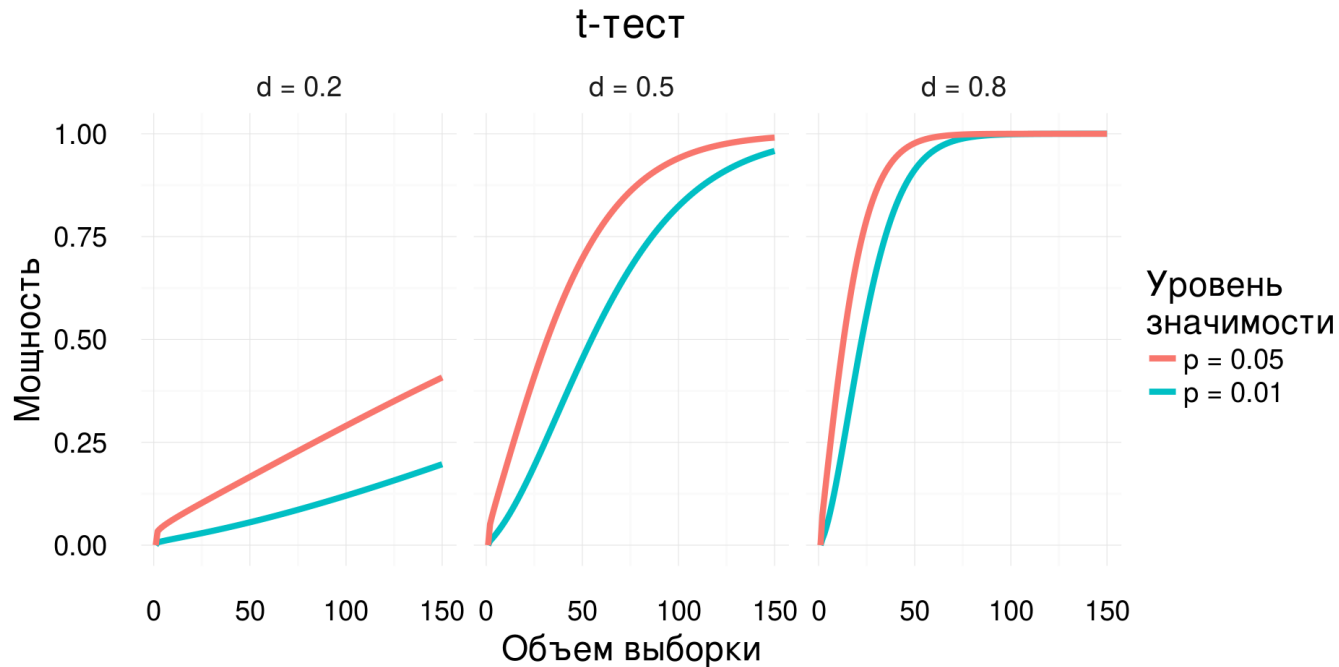
Чем больше объем выборки — тем больше мощность



Чем больше уровень значимости — тем больше мощность



Чем больше величина различий — тем больше мощность



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

- Чтобы не находить несуществующих эффектов, фиксируем уровень значимости
- Чтобы не пропустить значимое, рассчитываем величину эффекта, объем выборки 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.](#)