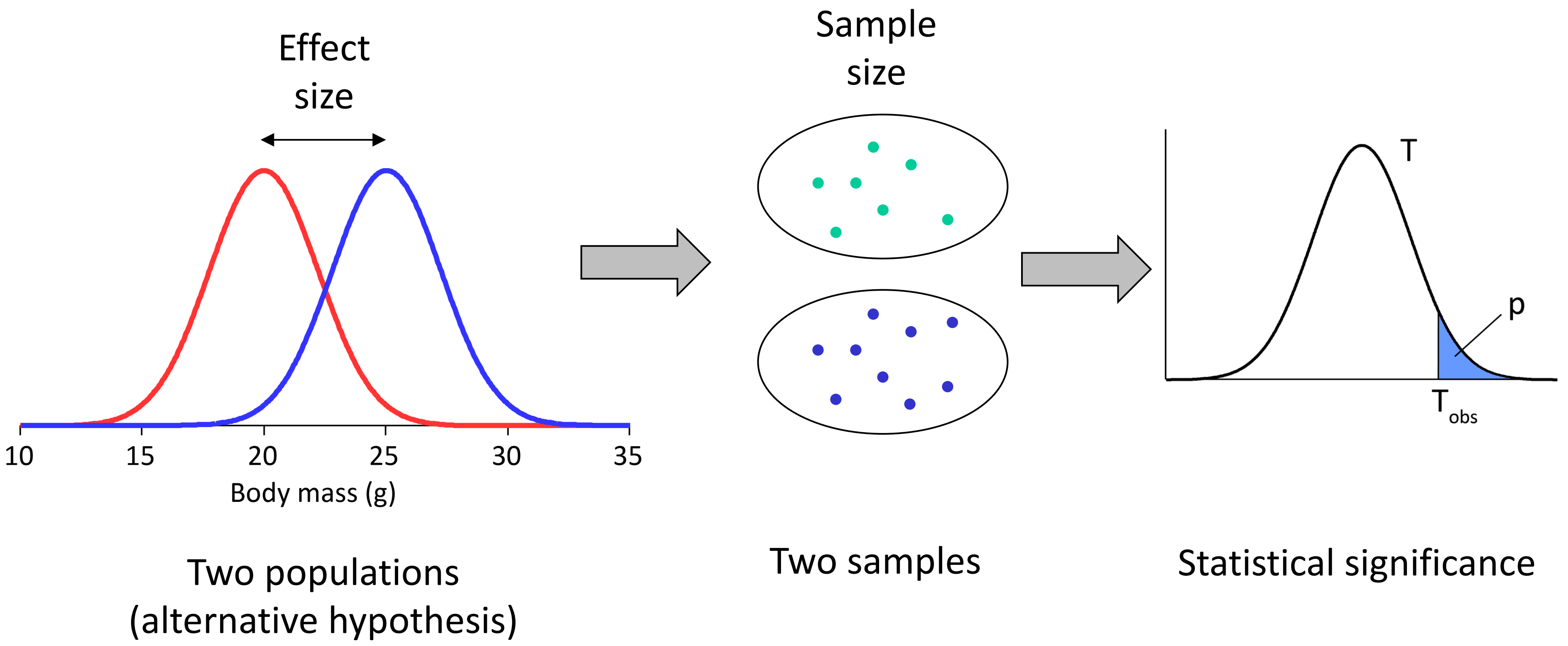
P-values and statistical tests
7. Statistical power

Marek Gierliński
Division of Computational Biology

Hand-outs available at http://is.gd/statlec

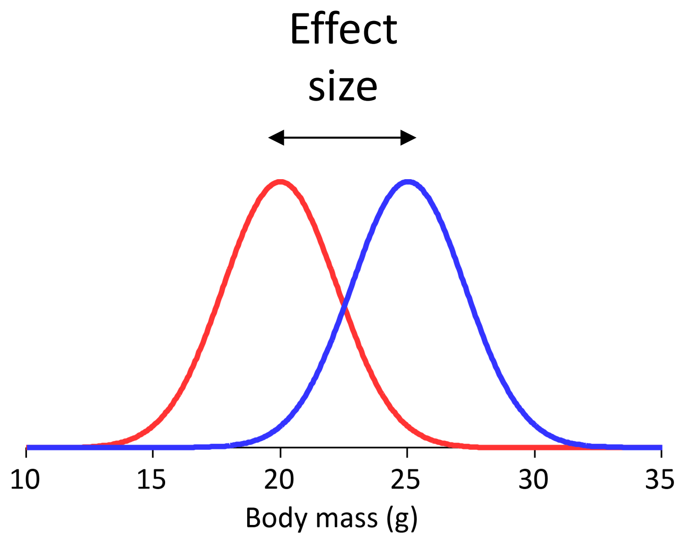
Statistical power: what is it about?



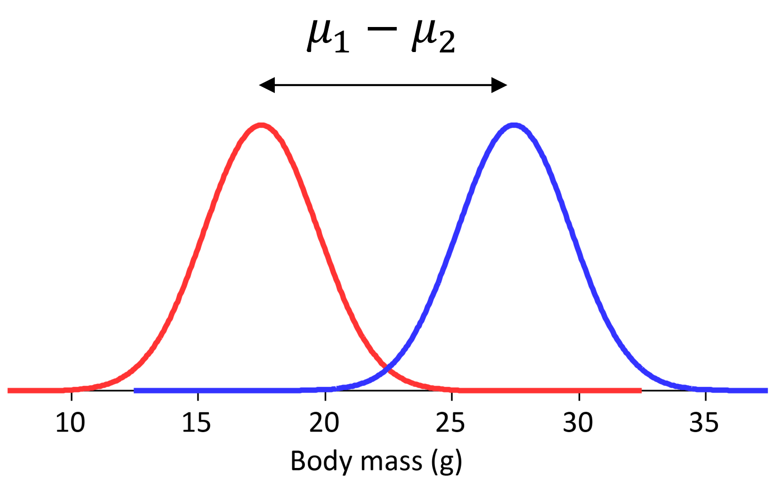
How does our ability to call a change “significant” depend on the effect size and the sample size?

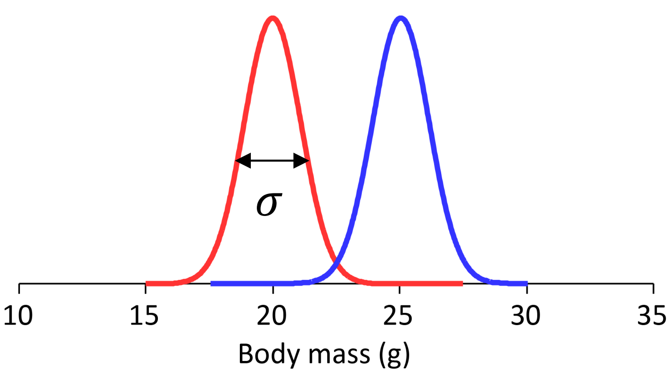
Effect size

Effect size describes the alternative hypothesis

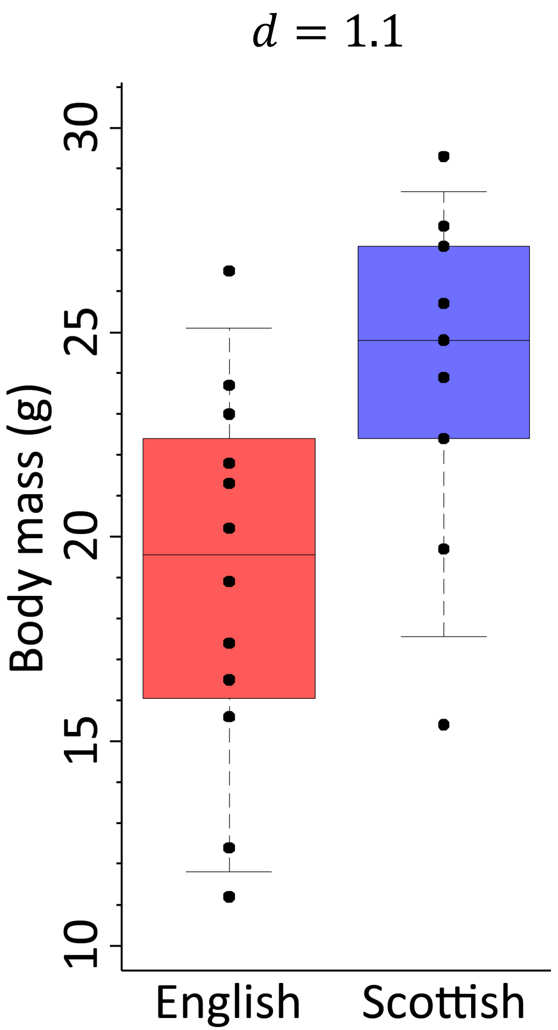


$$\boxed{\frac{\mu_1 - \mu_2}{\sigma}}$$





Effect size for two sample means



$$d = \frac{M_1 - M_2}{SD}$$

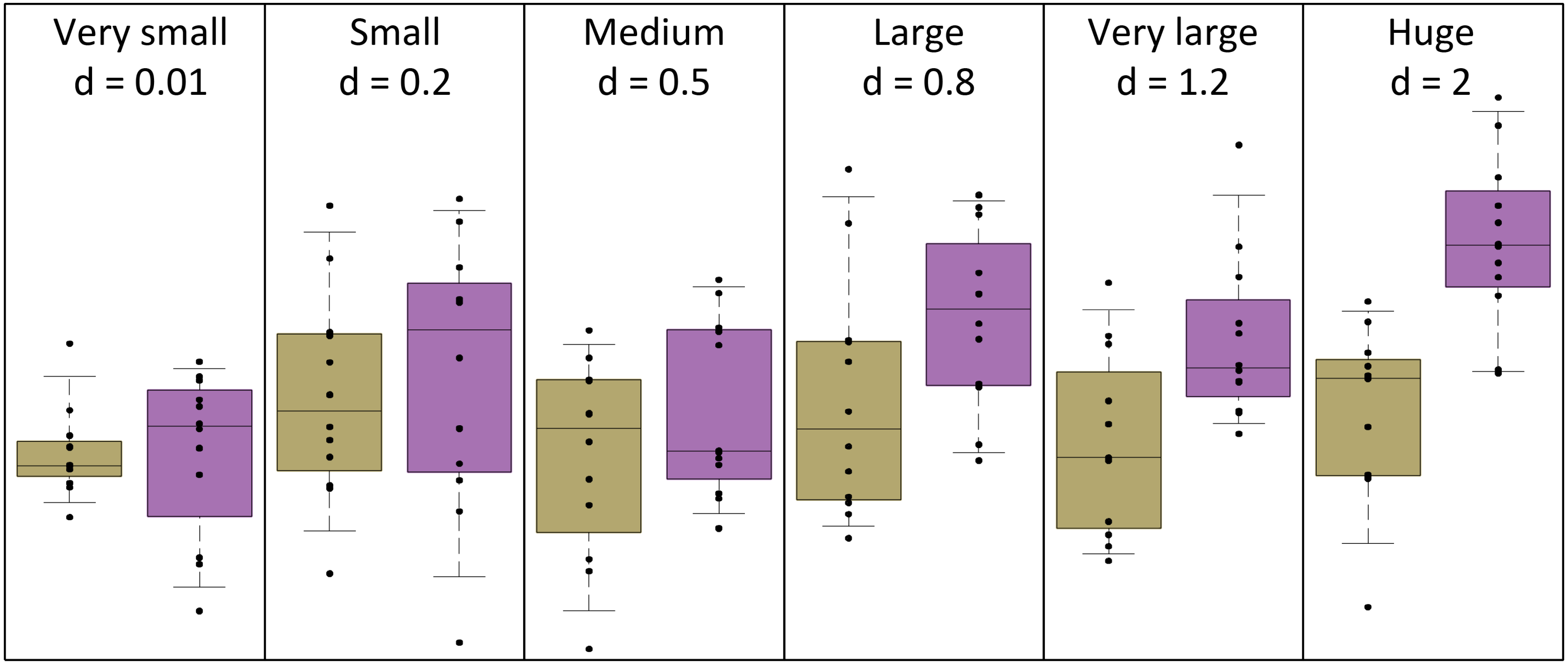
Cohen's d

$$SD = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 + 2}}$$

$$t = \frac{M_1 - M_2}{SE}$$

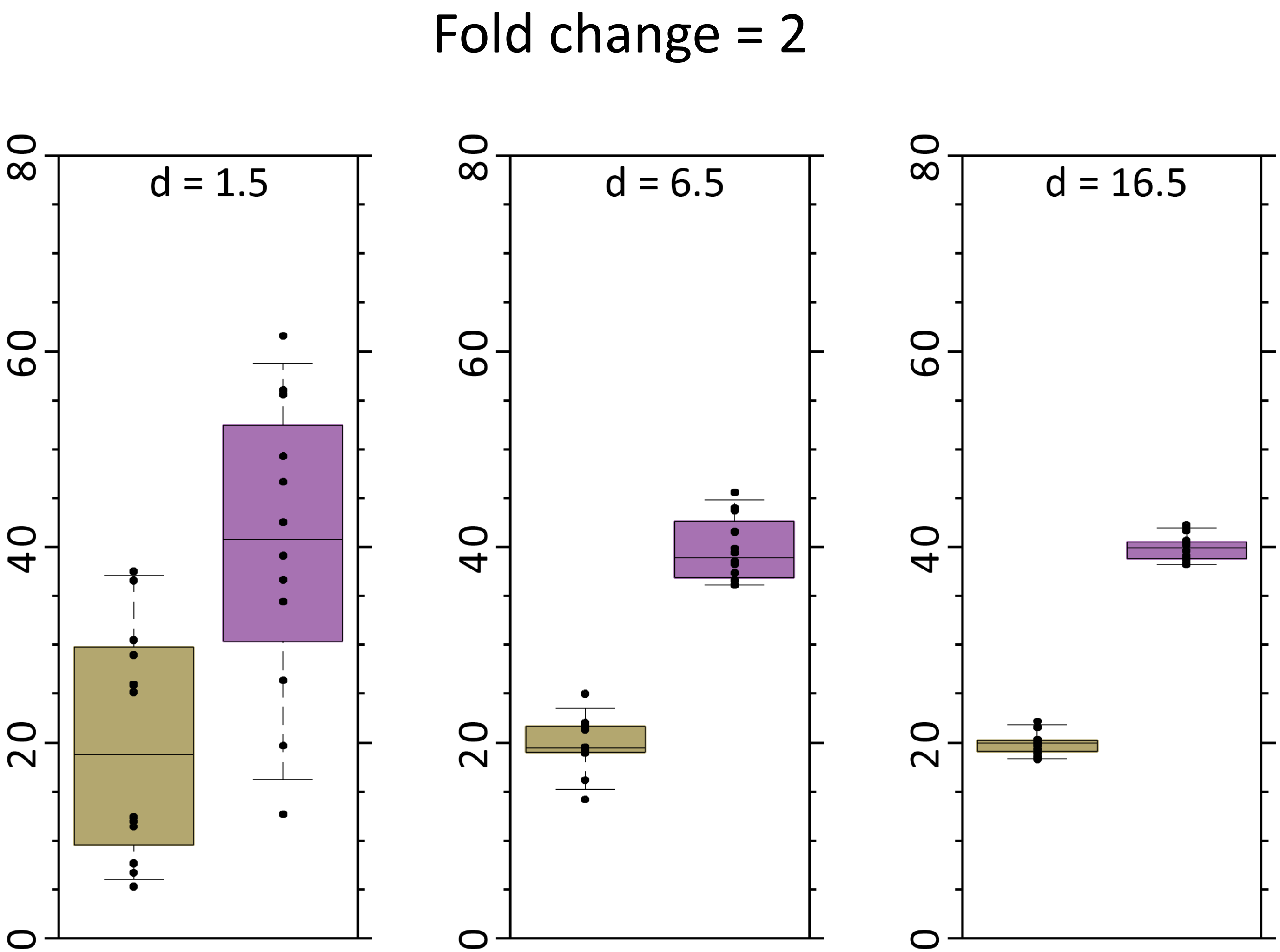
$$d = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Effect size for two sample means



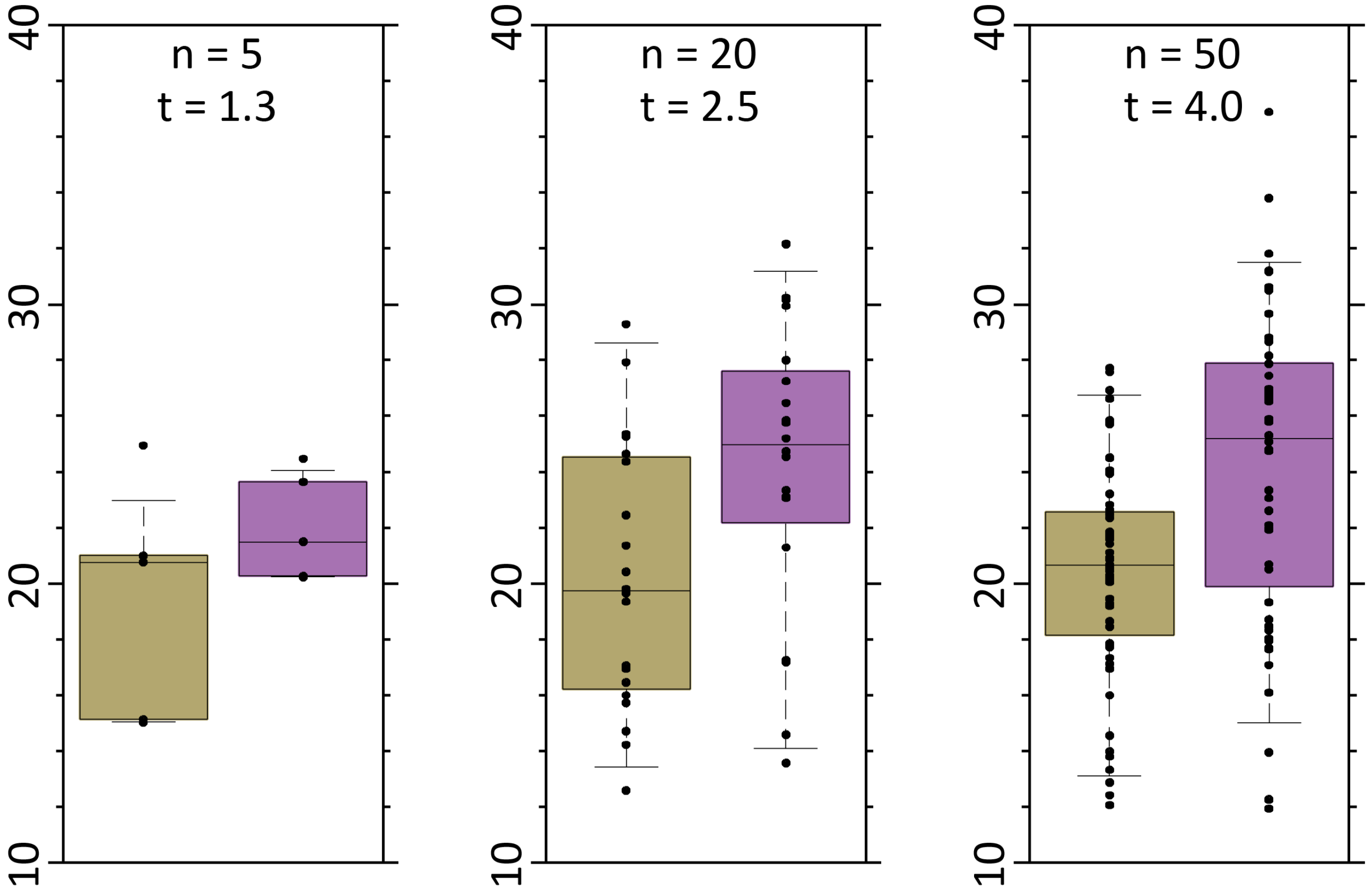
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*

Effect size depends on the standard deviation



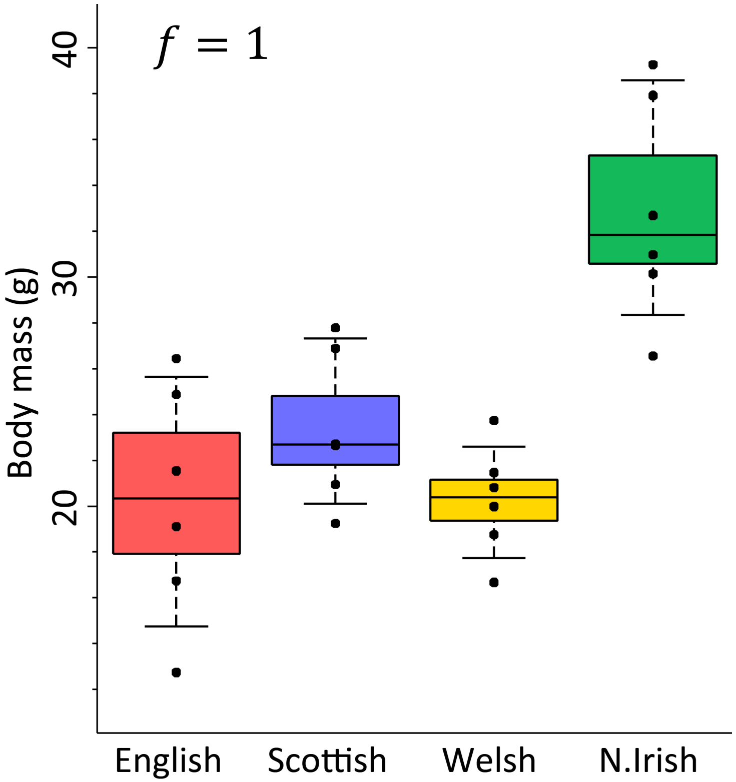
Effect size does not depend on the sample size

Effect size = 0.8



Effect size describes the alternative hypothesis

Effect size in ANOVA



Test statistic

$$F = \frac{MS_B}{MS_W}$$

H_0: $MS_B = MS_W$

H_1: $MS_B = MS_W + nMS_A$

Added variance

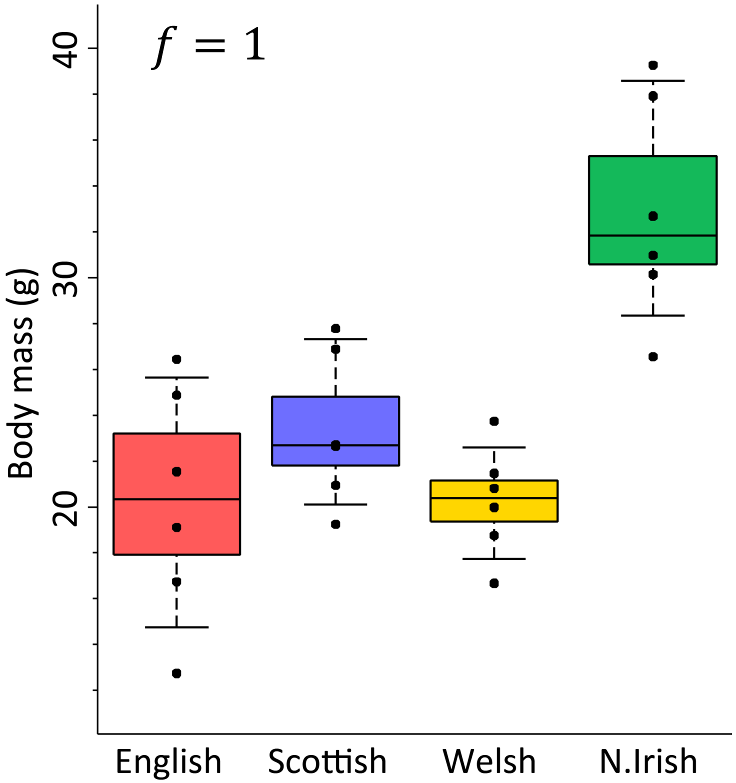
$$f^2 = \frac{MS_A}{MS_W}$$

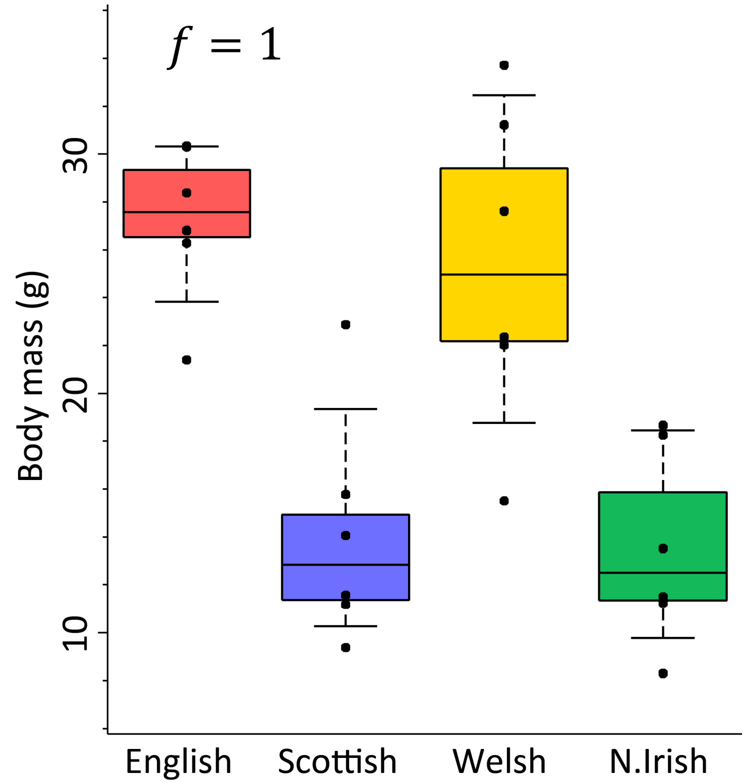
Cohen's f

$$f^2 = \frac{F-1}{n}$$

For the purpose of this calculation we only consider groups of equal sizes, n

Effect size in ANOVA





Effect size in frequency tables: odds ratio

	Dead	Alive	Total
Drug A	68	12	80
Drug B	70	30	100
Total	138	42	180

p = 0.013

$q_B - q_A = 0.30 - 0.15 = 0.15$

Not useful for small proportions

	Dead	Alive	Total
Drug A	$p_A = 0.85$	$q_A = 0.15$	1
Drug B	$p_B = 0.70$	$q_B = 0.30$	1
Total	1	1	

Odds of survival

$$\frac{q_A}{p_A} = \frac{0.15}{0.85} = 0.18 : 1$$

$$\frac{q_B}{p_B} = \frac{0.30}{0.70} = 0.43 : 1$$

Odds ratio

$$\omega = \frac{q_B/p_B}{q_A/p_A} = \frac{0.43}{0.18} = 2.4$$

Effect size

Data	Statistical test	Effect size	Formula
Two sets, size n_1 and n_2	t-test	Cohen's d	$d = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$
k groups of n points each	ANOVA	Cohen's f	$f = \sqrt{\frac{F - 1}{n}}$
2×2 contingency table	Fisher's exact	Odds ratio	$\omega = \frac{q_B/p_B}{q_A/p_A}$
Paired data $x_1, x_2, \dots, x_n$ and $y_1, y_2, \dots, y_n$	Significance of correlation	Pearson's r	$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - M_x}{SD_x}\right)\left(\frac{y_i - M_y}{SD_y}\right)$

How to do it in R?

```
> library(MBESS)
# Mouse body weight data
> English = c(16.5, 21.3, 12.4, 11.2, 23.7, 20.2, 17.4, 23, 15.6, 26.5, 21.8, 18.9)
> Scottish = c(19.7, 29.3, 27.1, 24.8, 22.4, 27.6, 25.7, 23.9, 15.4)
> n1 = length(English)
> n2 = length(Scottish)

# t-test with equal variances, extract test statistic
> test = t.test(English, Scottish, var.equal=TRUE)
> t = test$statistic[['t']]
# confidence limits on the non-centrality parameter (t in this case)
> nct.limits = conf.limits.nct(t, n1 + n2 - 2)
# find Cohen's distance and its limits
> sn = sqrt((n1 + n2) / (n1 * n2))
> d = t * sn
> d.lower = nct.limits$Lower.Limit * sn
> d.upper = nct.limits$Upper.Limit * sn

> d
[1] -1.102067
> d.lower
[1] -2.021337
> d.upper
[1] -0.1579345
```

Statistical power

t-test

Statistical testing

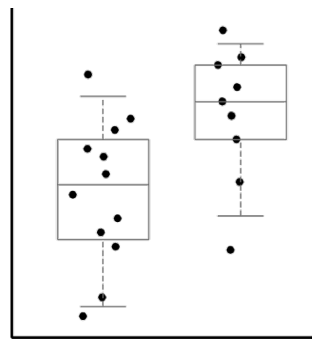
Statistical model

Null hypothesis
H_0: no effect

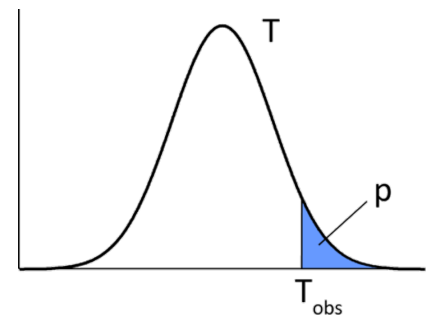
All other assumptions

Significance level
$\alpha = 0.05$

Data



Statistical test against H_0



p-value: probability that the observed effect is random

$p < \alpha$
Reject H_0
(at your own risk)
Effect is real

$p \geq \alpha$
Accept H_0 (!!!)

This table

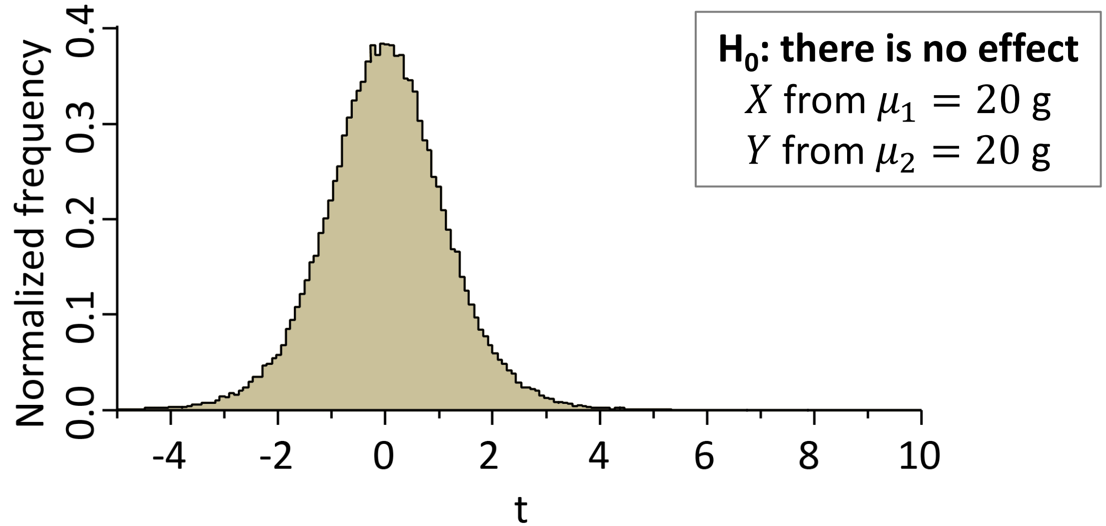
	H_0 is true	H_0 is false	
H_0 rejected	**type I error (α)** false positive	**correct decision** true positive	Positive
H_0 accepted	**correct decision** true negative	**type II error (β)** false negative	Negative
	No effect	Effect	

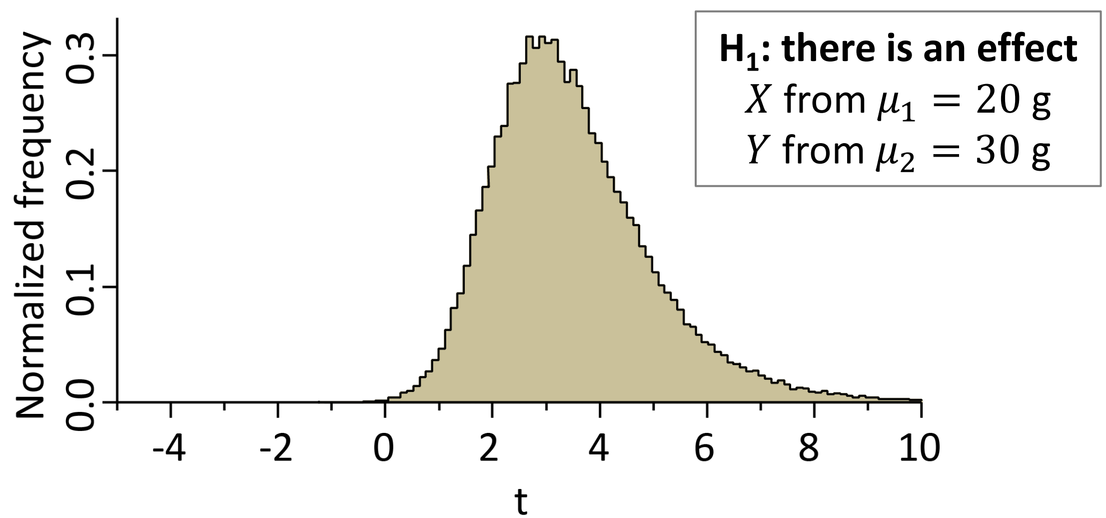
Gedankenexperiment

Draw 100,000 pairs of samples (X, Y) of size $n = 5$

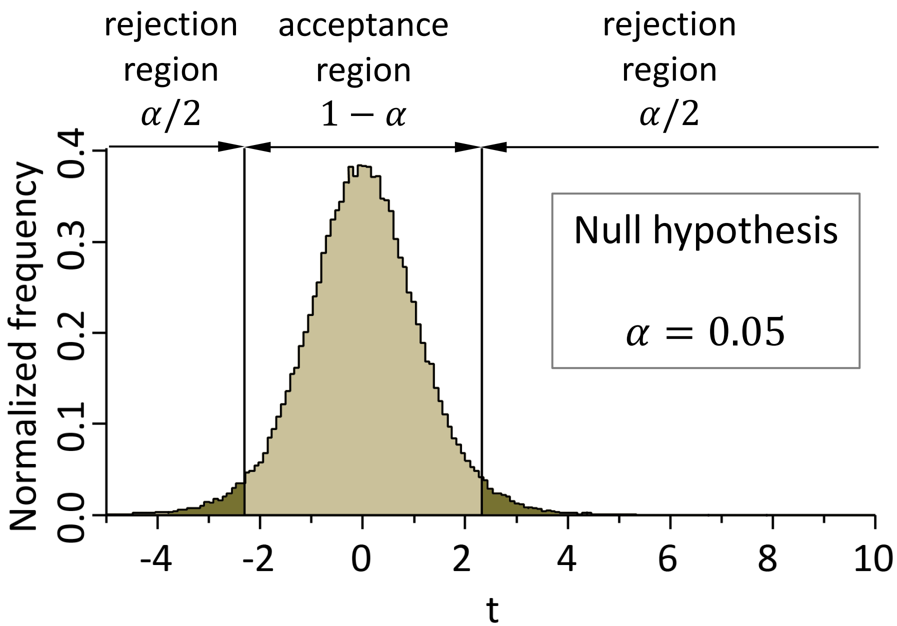
Find $t = (M_1 - M_2)/SE$ for each pair

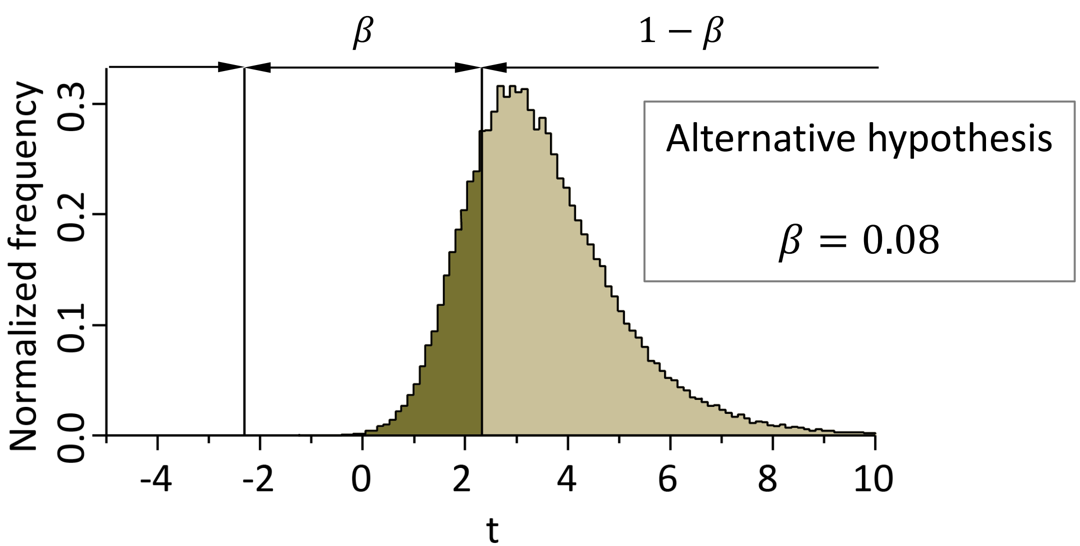
Build sampling distribution of t





One alternative hypothesis





	H_0 true	H_0 false
reject	FP $\boldsymbol{\alpha}$	TP
accept	TN	FN $\boldsymbol{\beta}$

Power of the test

$$P = 1 - \beta$$

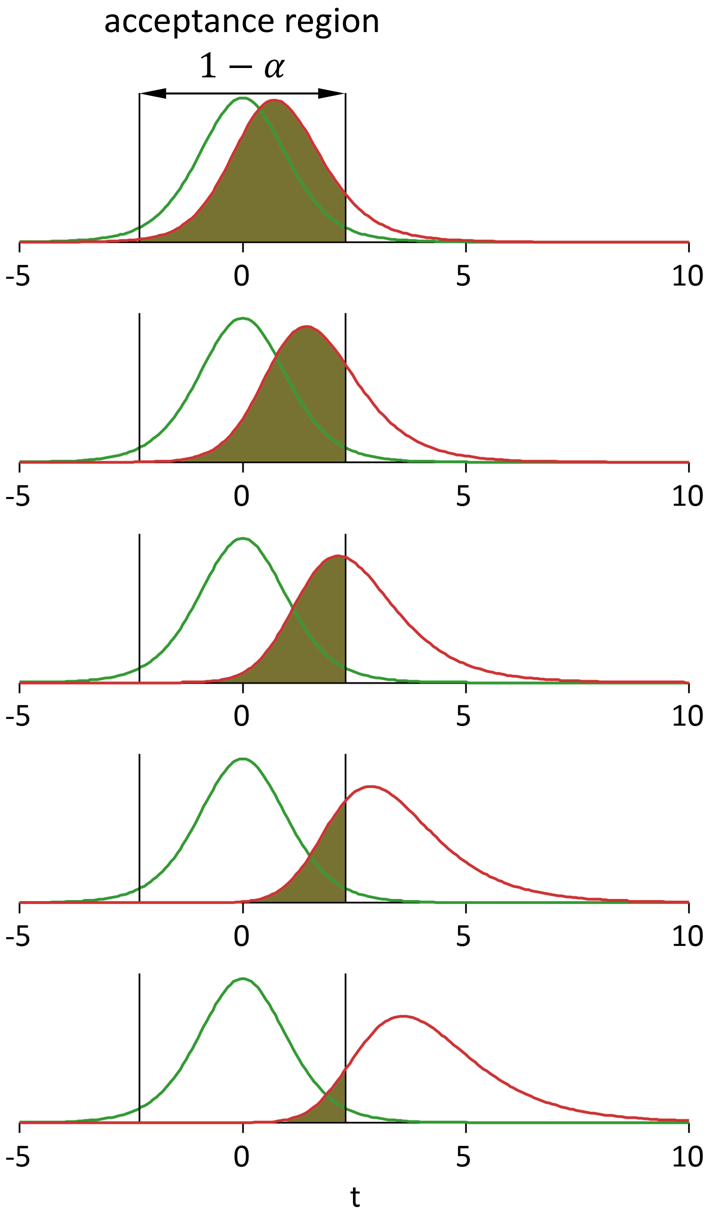
Probability that we correctly reject H_0

Statistical power

The probability of correctly rejecting the null hypothesis

(choosing the alternative, when it is true)

Multiple alternative hypotheses



$\mu_1 = 22$ g
$d = 0.49$
$\beta = 0.90$

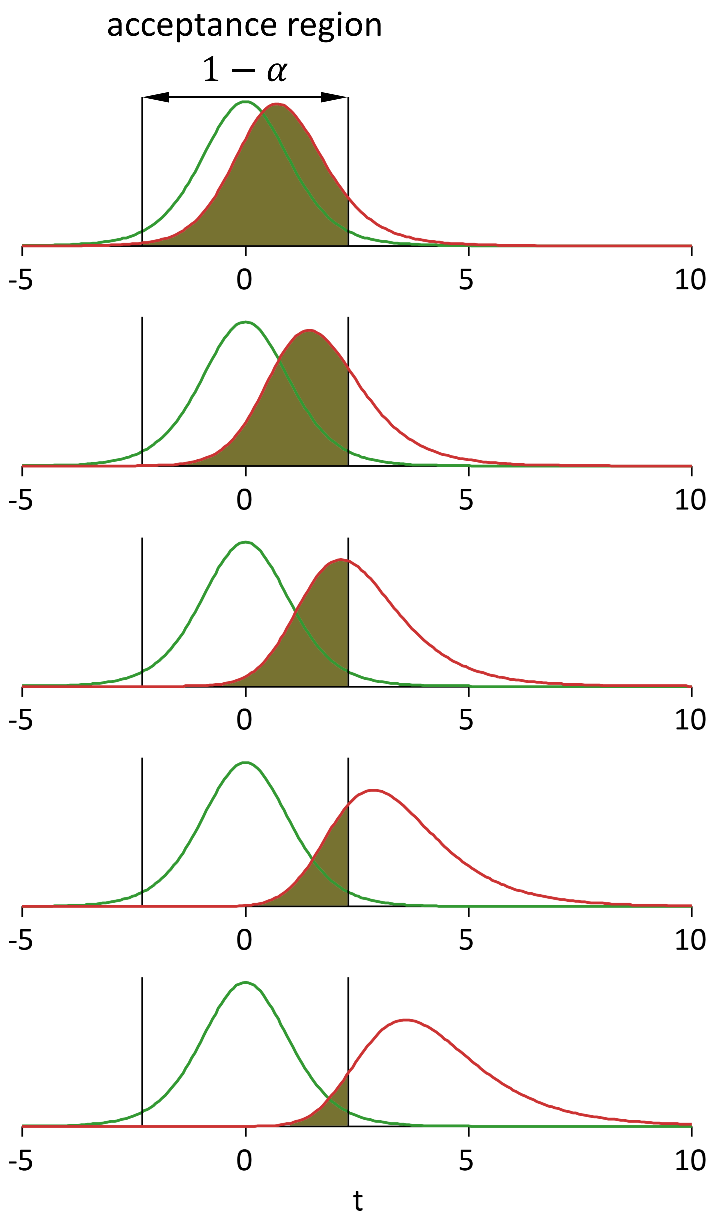
$\mu_1 = 24$ g
$d = 0.98$
$\beta = 0.72$

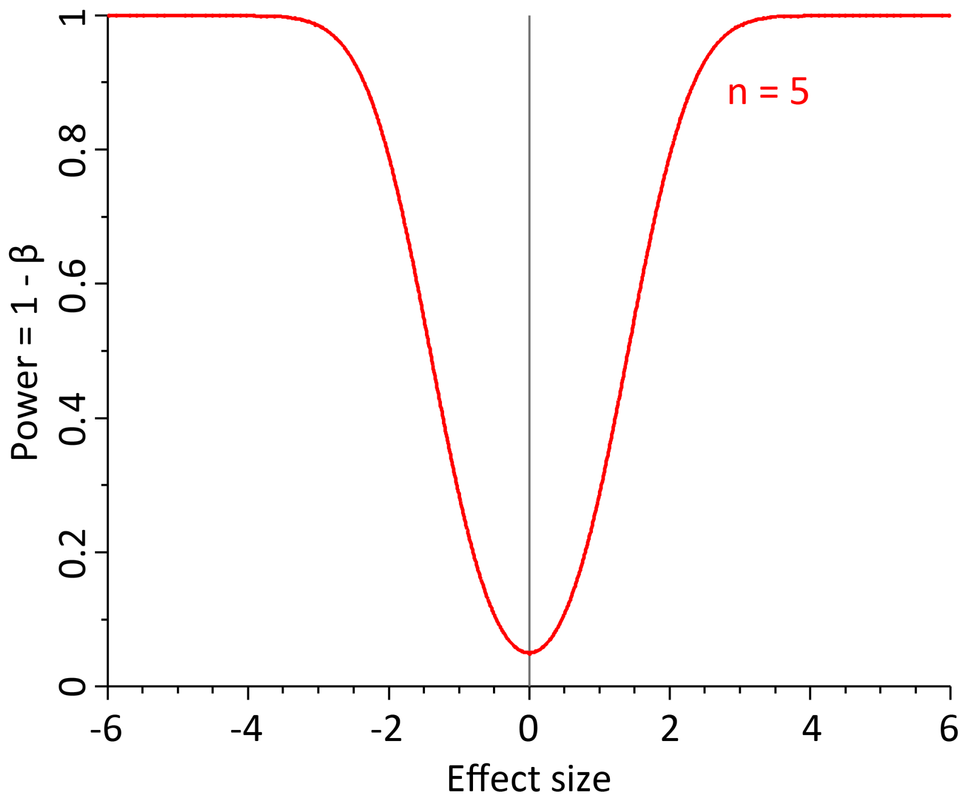
$\mu_1 = 26$ g
$d = 1.47$
$\beta = 0.47$

$\mu_1 = 29$ g
$d = 1.96$
$\beta = 0.23$

$\mu_1 = 30$ g
$d = 2.45$
$\beta = 0.08$

Power curve

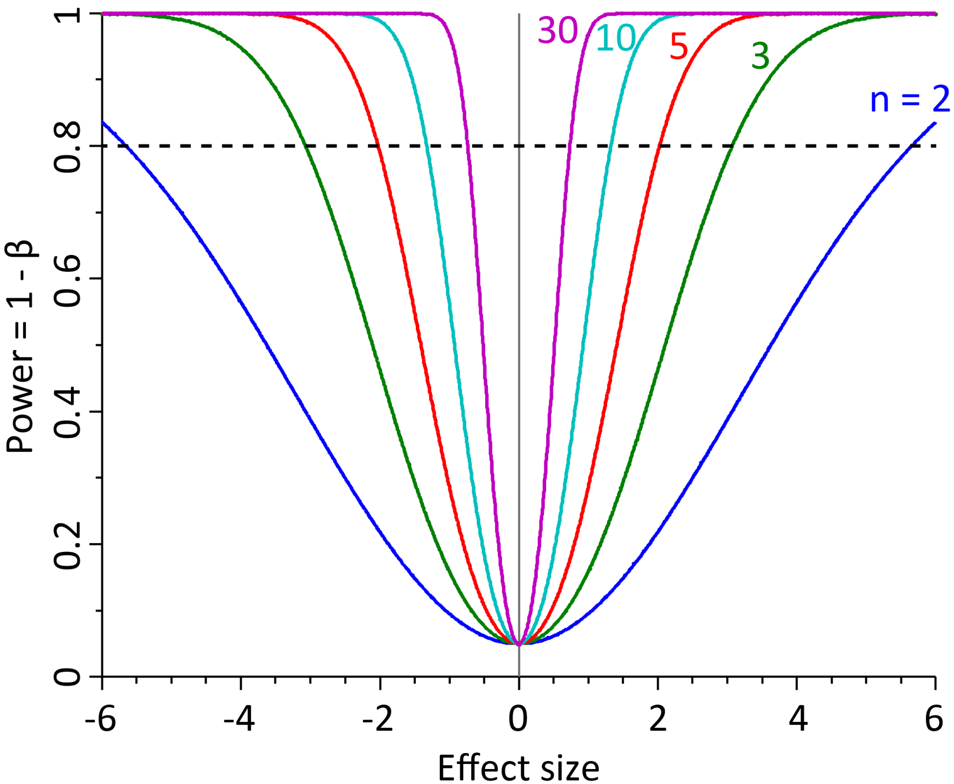


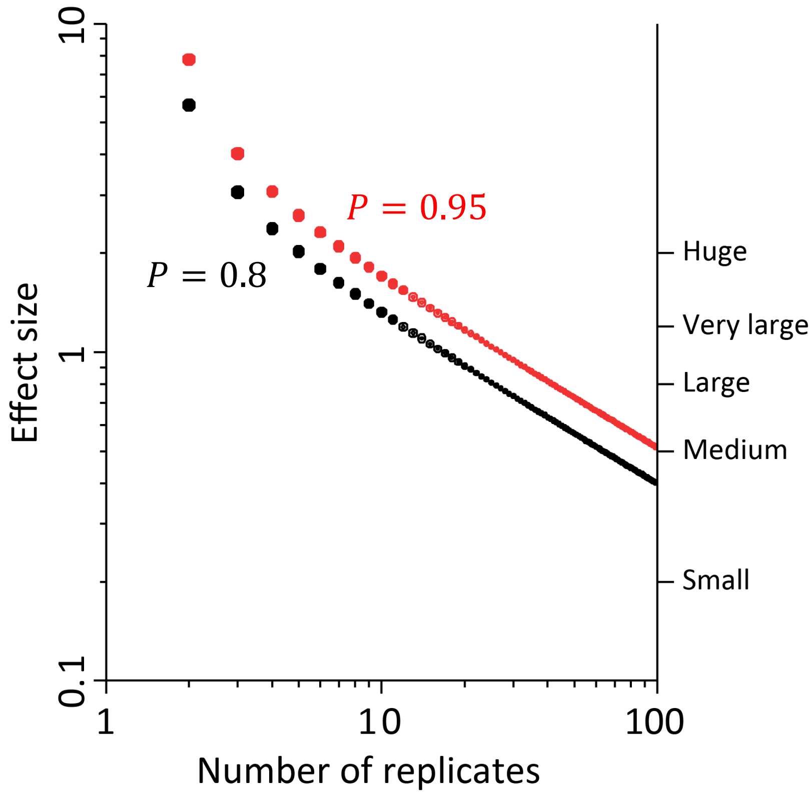


β - type II error (false negative) probability

$$\text{Power} = 1 - \beta$$

Power curves





How to do it in R?

```
# Find sample size required to detect the effect size d = 1
> power.t.test(d=1, sig.level=0.05, power=0.8, type="two.sample",
alternative="two.sided")

     One-sample t test power calculation

              n = 16.71473
              d = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

> power.t.test(d=1, sig.level=0.05, power=0.95, type="two.sample",
alternative="two.sided")

     One-sample t test power calculation

              n = 26.98922
              d = 1
      sig.level = 0.05
          power = 0.95
    alternative = two.sided
```

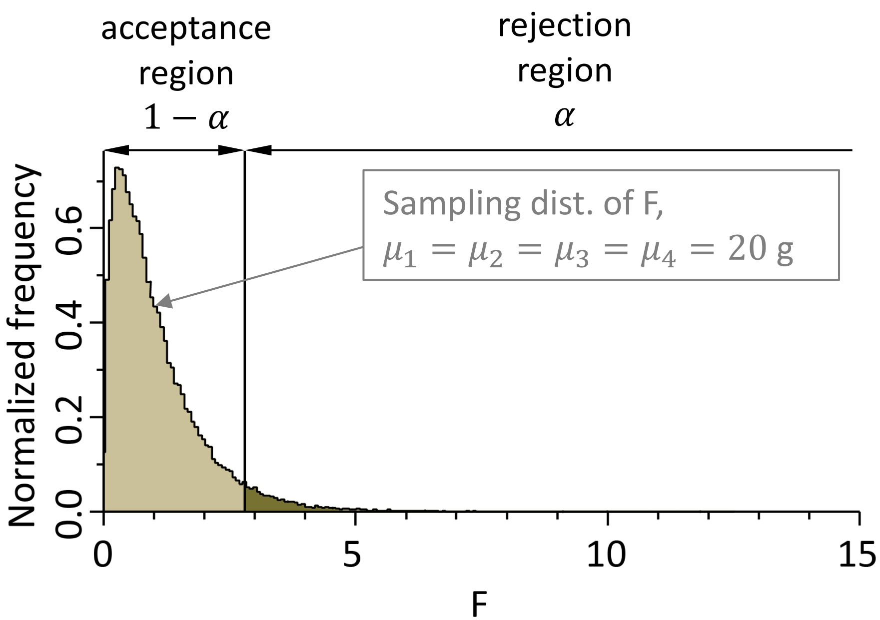
Statistical power

ANOVA

One alternative hypothesis

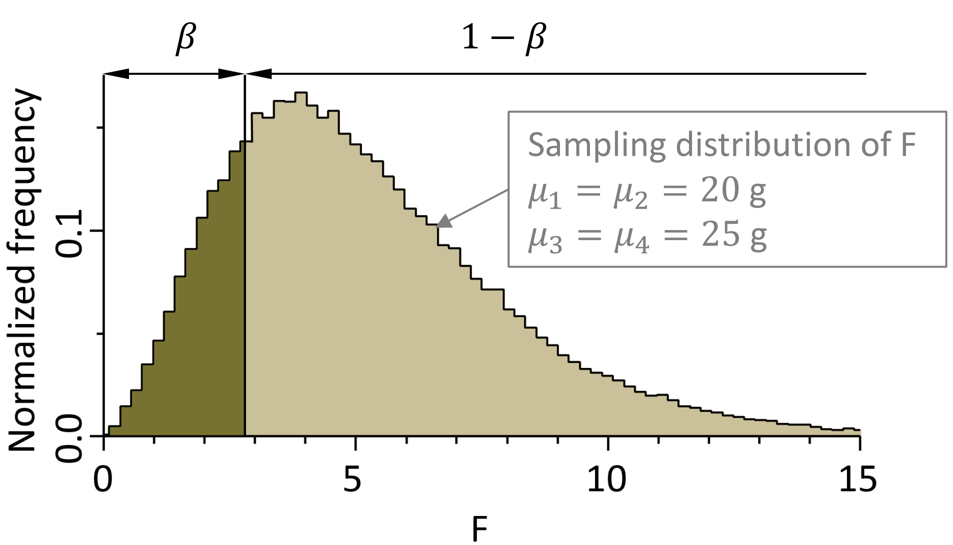
Null hypothesis

$\alpha = 0.05$

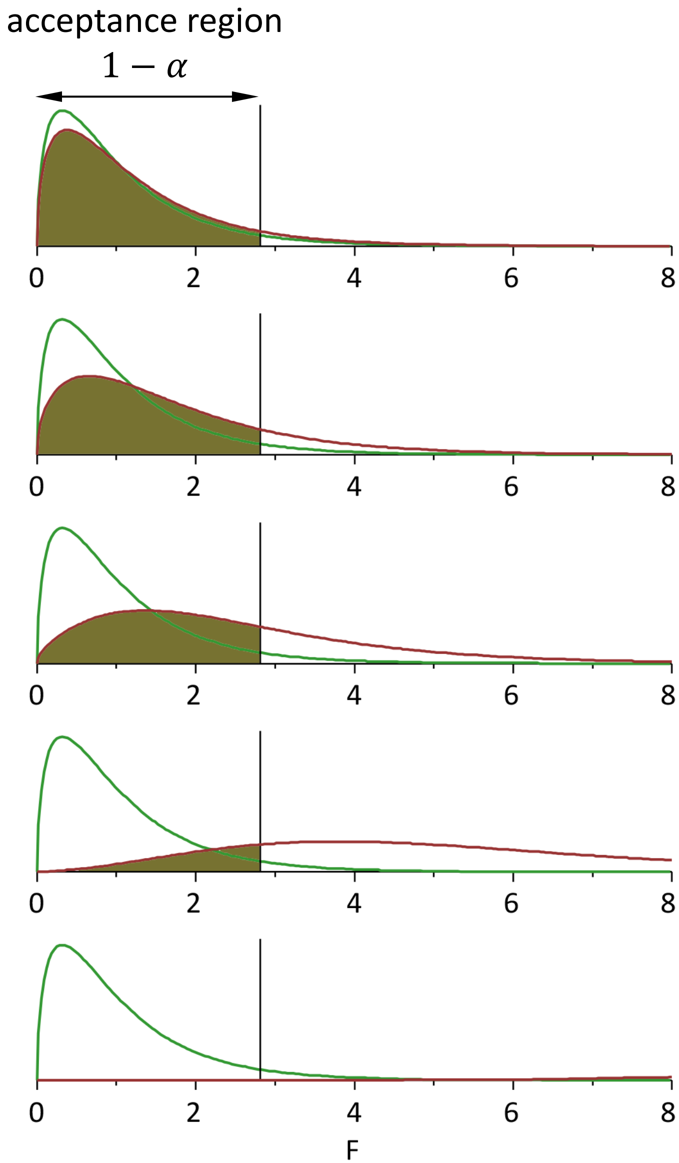


Alternative hypothesis

$\beta = 0.20$



Multiple alternative hypotheses



$f = 0.1$
$\beta = 0.92$

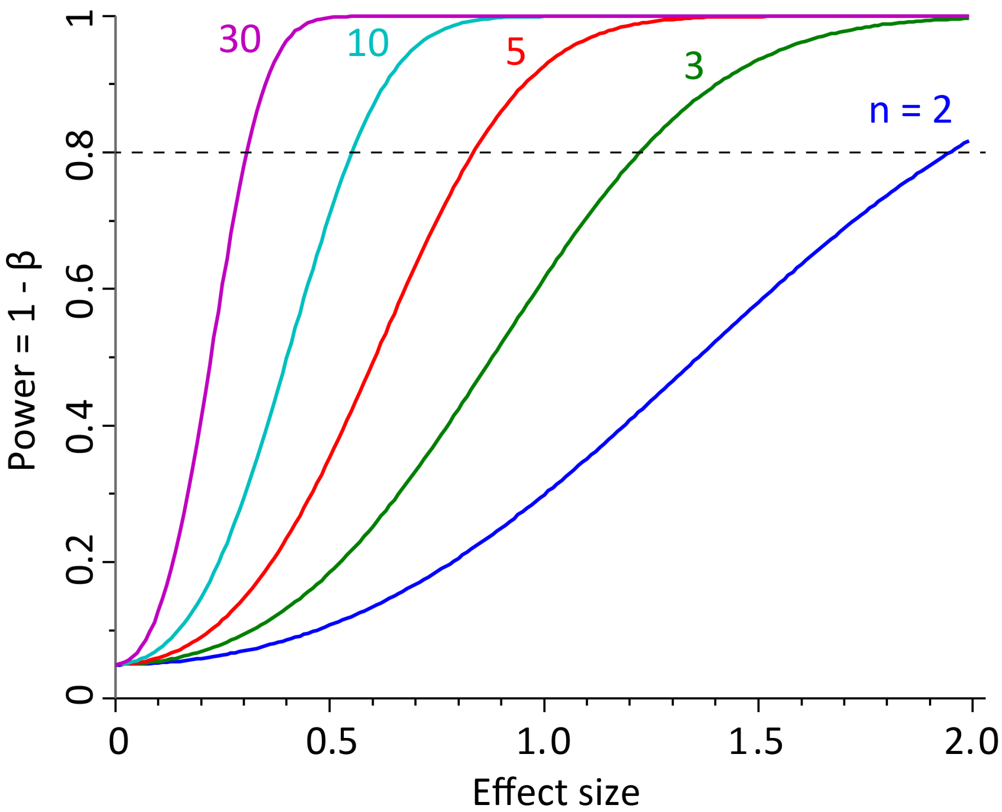
$f = 0.2$
$\beta = 0.83$

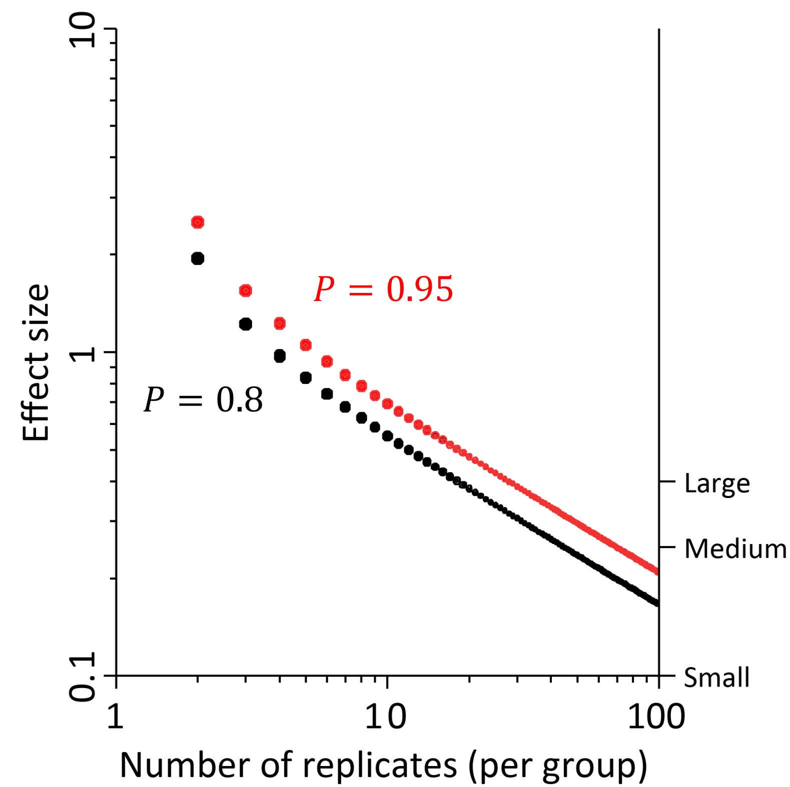
$f = 0.3$
$\beta = 0.65$

$f = 0.5$
$\beta = 0.20$

$f = 1$
$\beta = 3\times10^{-5}$

Power curves





How to do it in R?

```
> library(pwr)

# Find sample size required to detect a “large” effect size f = 0.4
> pwr.anova.test(k=4, f=0.4, sig.level=0.05, power=0.8)

     Balanced one-way analysis of variance power calculation

              k = 4
              n = 18.04262
              f = 0.4
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

Worked example

Example: how toxicity affects rat brains

Pilot experiment

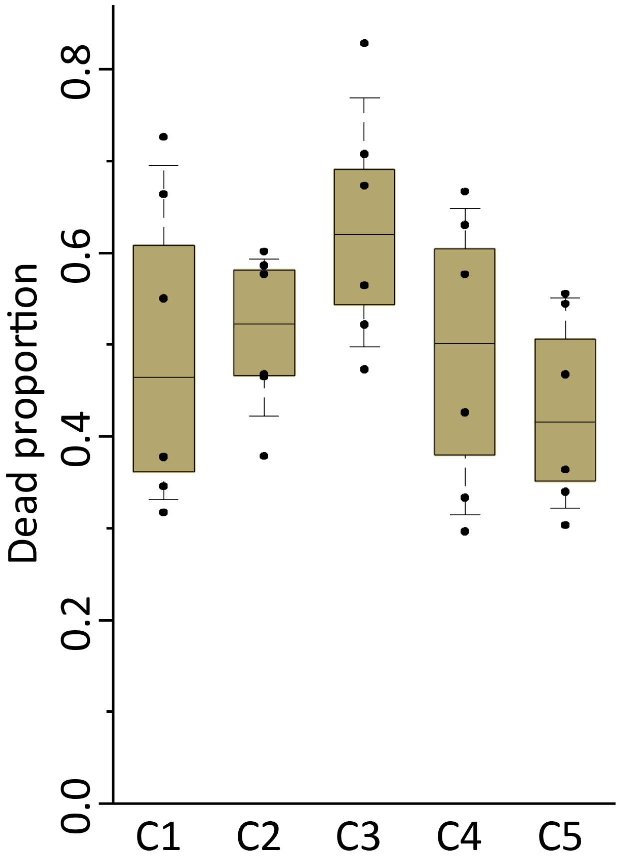
Connected neurons in 5 chambers
Put neurotoxin in C3
Count dead and alive cells
See how it spreads

Power analysis

How many replicates do we need to...

1) detect a 10% difference between chambers? (power in t-test)
2) detect the observed C1-C5 effect in ANOVA? (power in ANOVA)

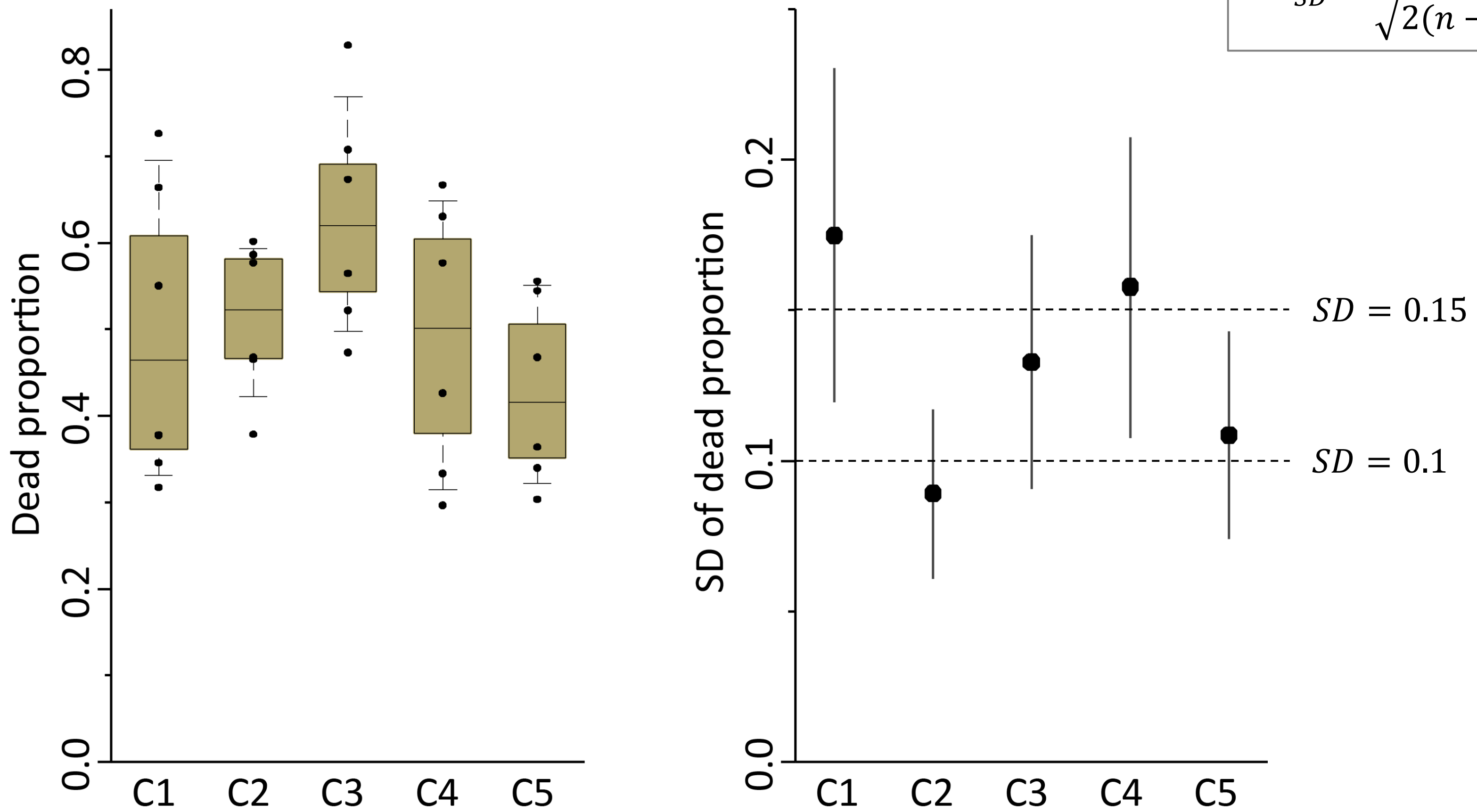
$k = 5$ chambers
$n = 6$ replicates in each



Samson *at al.* (2016)
DOI:10.1038/srep33746

How many replicates to detect a difference of 0.1 between chambers?

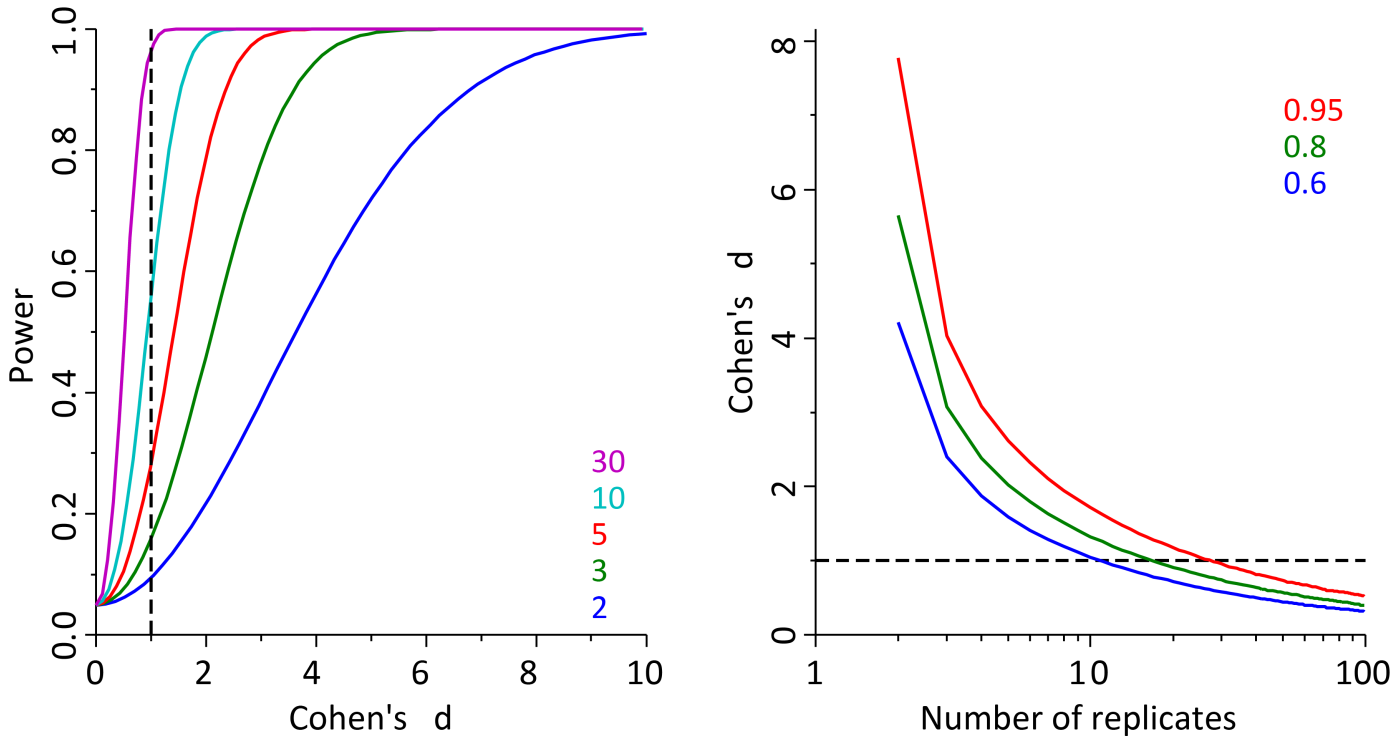
Assess your data variability based on the pilot



Standard error of SD

$$SE_{SD} = \frac{SD}{\sqrt{2(n-1)}}$$

Better scenario: $SD = 0.1$

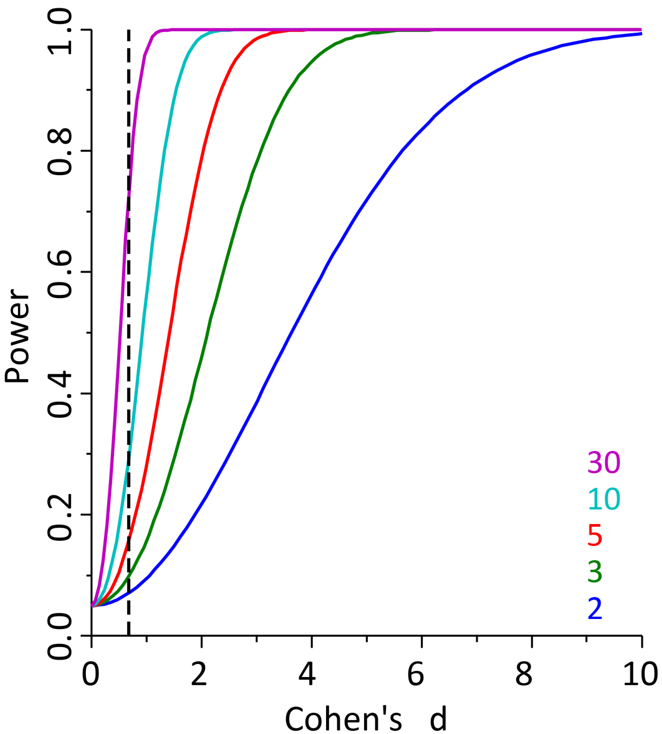


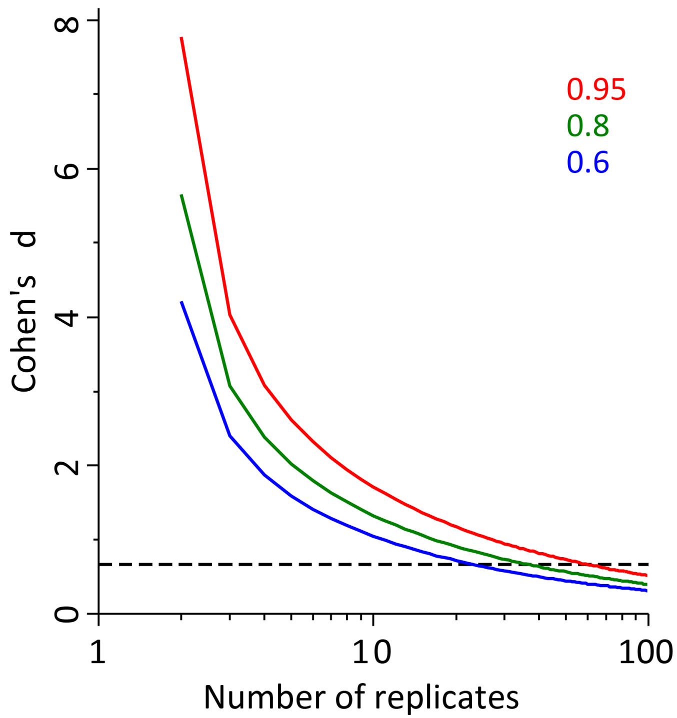
Cohen's d:

$$d = \frac{\Delta M}{SD} = \frac{0.1}{0.1} = 1$$

```
> power.t.test(d=1, sig.level=0.05, power=0.8, type="two.sample", alternative="two.sided")

     Two-sample t test power calculation

              n = 16.71477
          delta = 1
             sd = 1
      sig.level = 0.05
          power = 0.8
```

Worse scenario: $SD = 0.15$





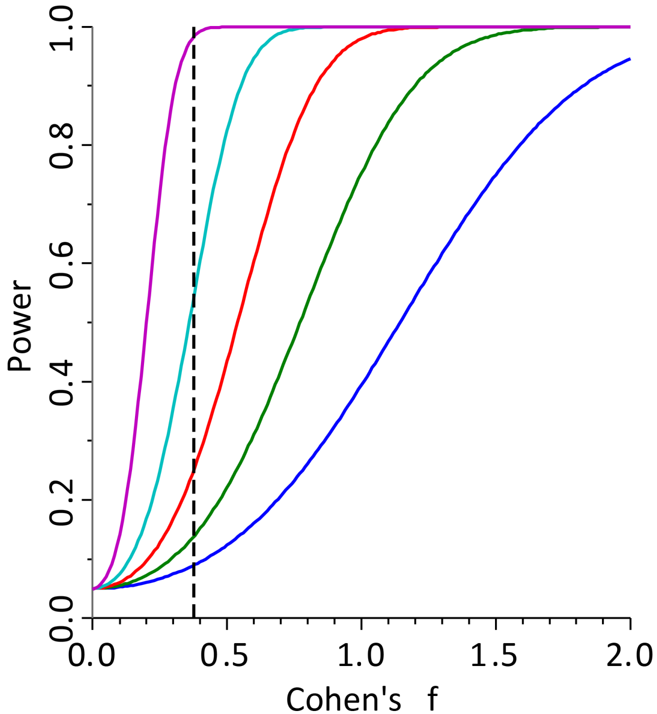
Cohen's d:

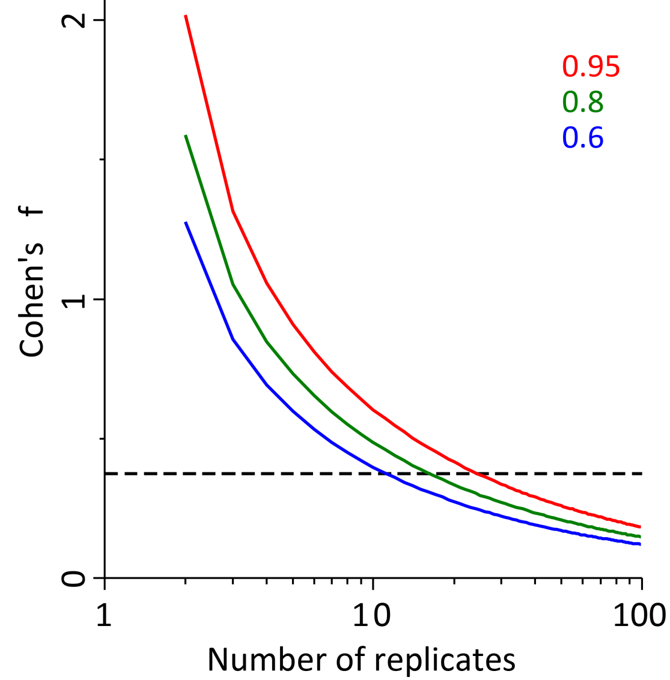
$$d = \frac{\Delta M}{SD} = \frac{0.1}{0.15} \approx 0.67$$

```
> power.t.test(d=0.67, sig.level=0.05, power=0.8, type="two.sample", alternative="two.sided")

     Two-sample t test power calculation

              n = 35.95548
          delta = 0.67
             sd = 1
      sig.level = 0.05
          power = 0.8
```

How many replicates to detect the observed C1-C5 effect in ANOVA?

Power in ANOVA





$$f = \sqrt{\frac{F-1}{n}} = 0.38$$

How many replicates do we need?

```
> library(pwr)
> rat = read.table('http://tiny.cc/rat_toxicity', header=TRUE)
# Here n = 6 and k = 4

> rat.aov = aov(Proportion ~ Chamber, data=rat)
# Extract F value
> F = summary(rat.aov)[[1]]$F[1]
# Effect size: Cohen's f
> f = sqrt((F - 1)/n)

# What is the power of this experiment?
> pwr.anova.test(k=4, n=6, f=f, sig.level=0.05)

              k = 6
              n = 5
              f = 0.3760972
      sig.level = 0.05
          power = 0.2507655

# How many replicates to get power of 0.8?
> pwr.anova.test(k=4, f=f, sig.level=0.05, power=0.8)

              k = 6
              n = 16.06243
              f = 0.3760972
      sig.level = 0.05
          power = 0.8
```

Hand-outs available at
http://tiny.cc/statlec