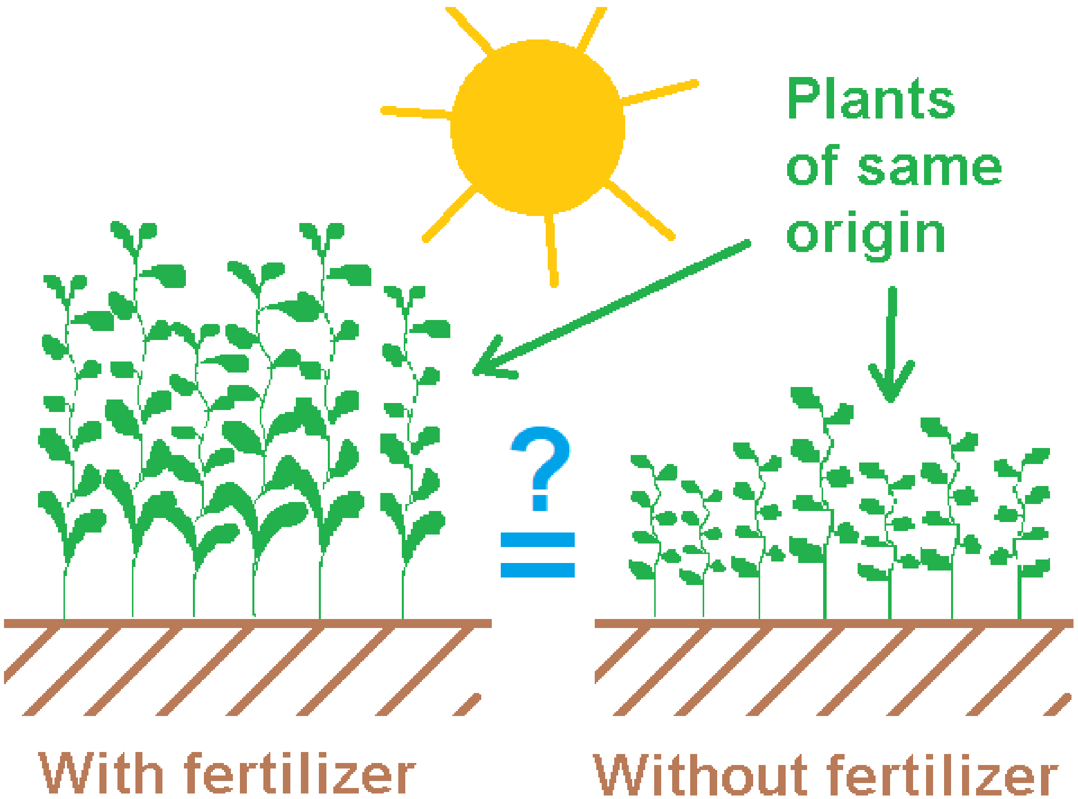
Experimental Data Analysis
in ©MATLAB

Lecture 3:
Hypothesis testing, group differences, paired vs. independent test, effect size

Jan Rusz
Czech Technical University in Prague

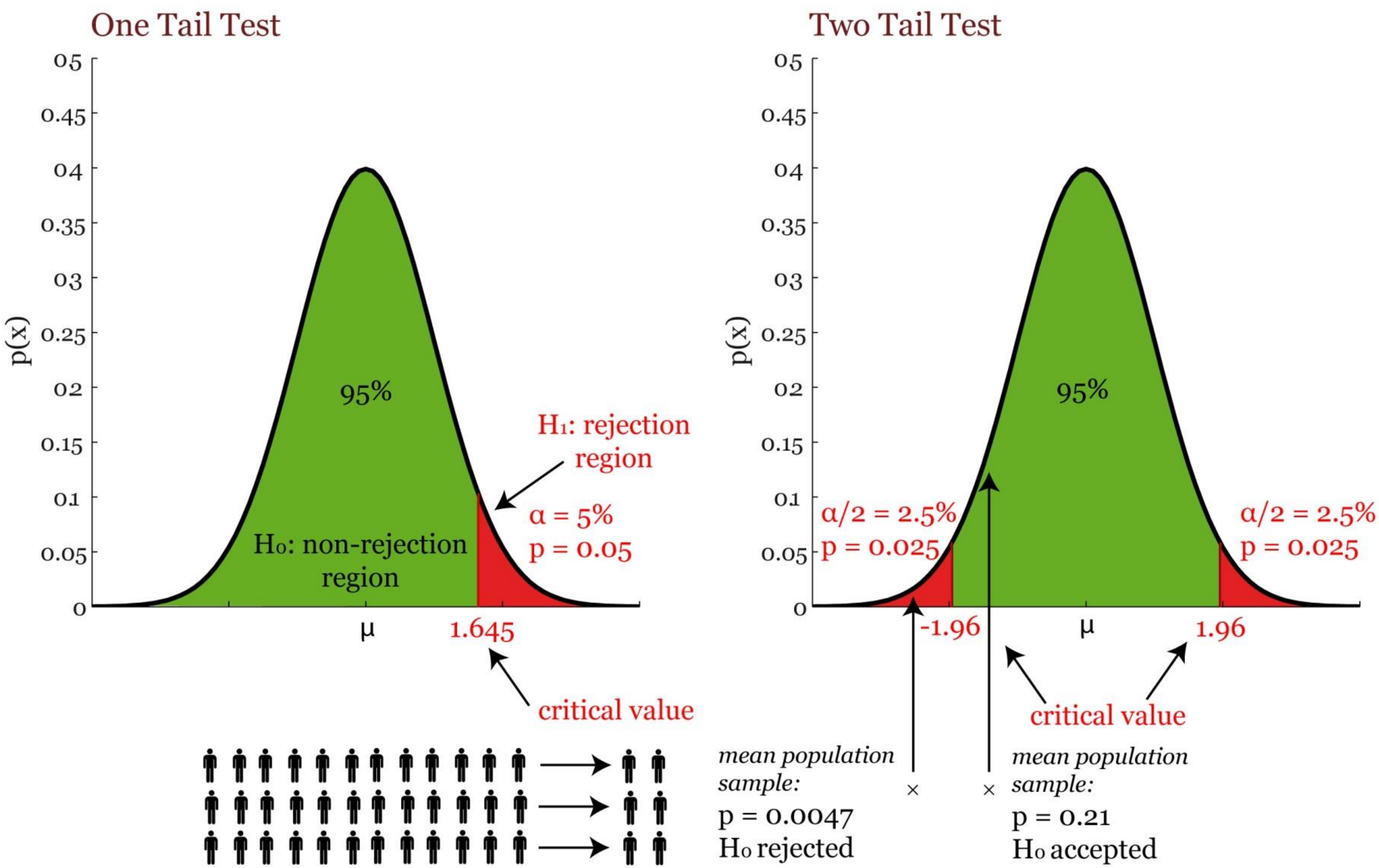
Motivation

- Do size of plants differ in dependence of fertilizer application?
- Obviously yes, but can you quantify the difference?
- Each plant has different size and there is different number of plants in groups.
- What is the probability of co-incidence?



Hypothesis testing

distribution: $\mu = 0, \sigma = 1$



Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

1. step: State null (H_0) and alternative (H_1) hypothesis

$H_0: \mu = 100$

Two Tailed Test: $H_1: \mu \neq 100$

One Tailed Test: $H_1: \mu < 100$

$H_1: \mu > 100$

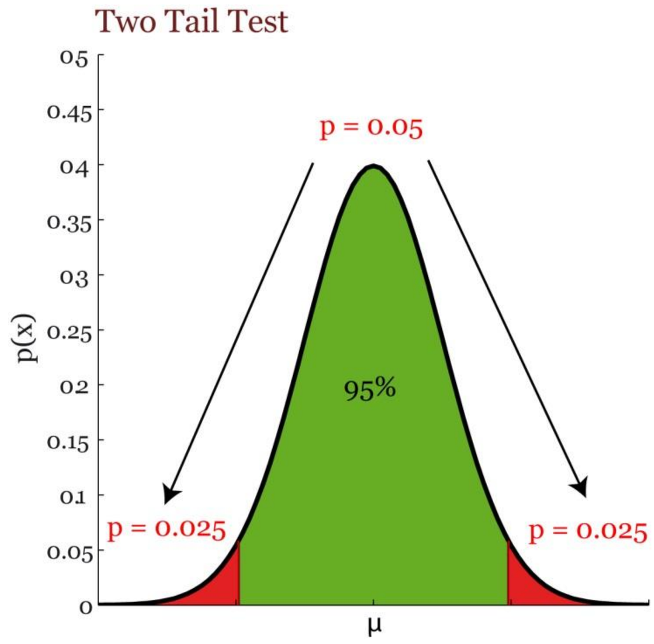
Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

2. step: Choose level of significance (p)

H_0: $\mu = 100$

H_1: $\mu \neq 100$



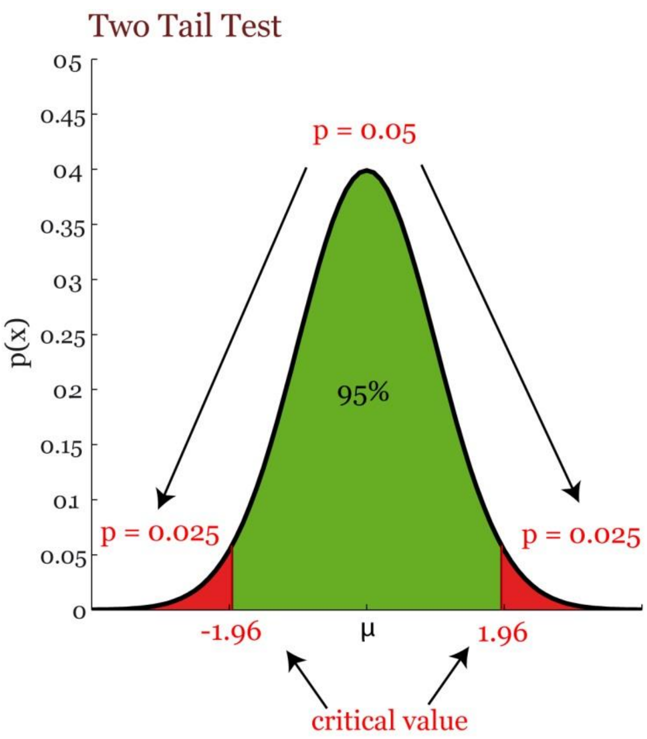
Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

3. step: Find critical values

H_0: $\mu = 100$

H_1: $\mu \neq 100$



Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

4. step: Find test statistic

H_0: $\mu = 100$, $\sigma = 15$

H_1: $\mu \neq 100$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{105-100}{\frac{15}{\sqrt{75}}}$$

z-test: We know standard deviation of the population

$$t = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}}$$

t-test: We calculate actual standard deviation (*S*) of the sample

Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

4. step: Find test statistic

H_0: $\mu = 100$, $\sigma = 15$

H_1: $\mu \neq 100$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{105-100}{\frac{15}{\sqrt{75}}} = \mathbf{2.89}$$

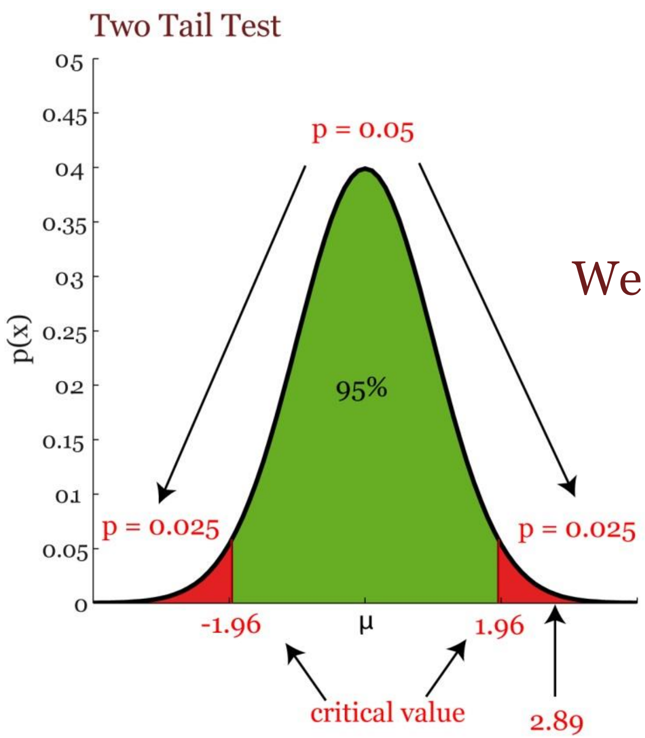
Hypothesis testing: example

The average IQ for the adult population is defined to be 100 with standard deviation of 15. Researcher decided to reproduce this result and tested the IQ in 75 random adults. He obtained average IQ of 105 using this sample. Is there enough evidence to suggest that the average IQ has changed?

5. step: Draw your conclusion

H_0: $\mu = 100, \sigma = 15$

H_1: $\mu \neq 100$



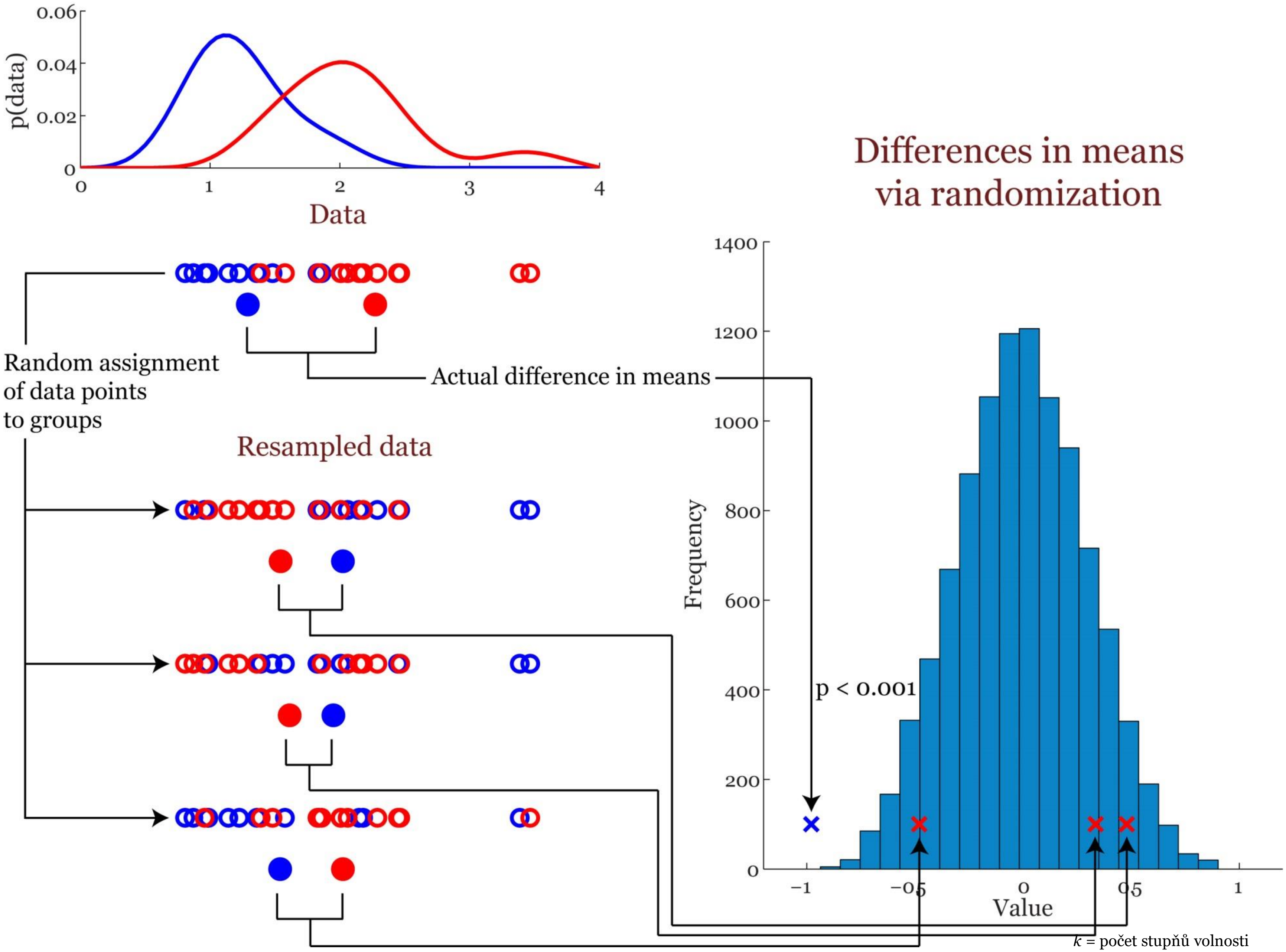
We can REJECT the H_0!

How to report p values?

The most common levels of significance reported with respect to research questions:
$p < 0.05$ * (minimal level of significance)
$p < 0.01$ **
$p < 0.001$ *** (statistically highly significant – less than one in a thousand chance of being wrong)

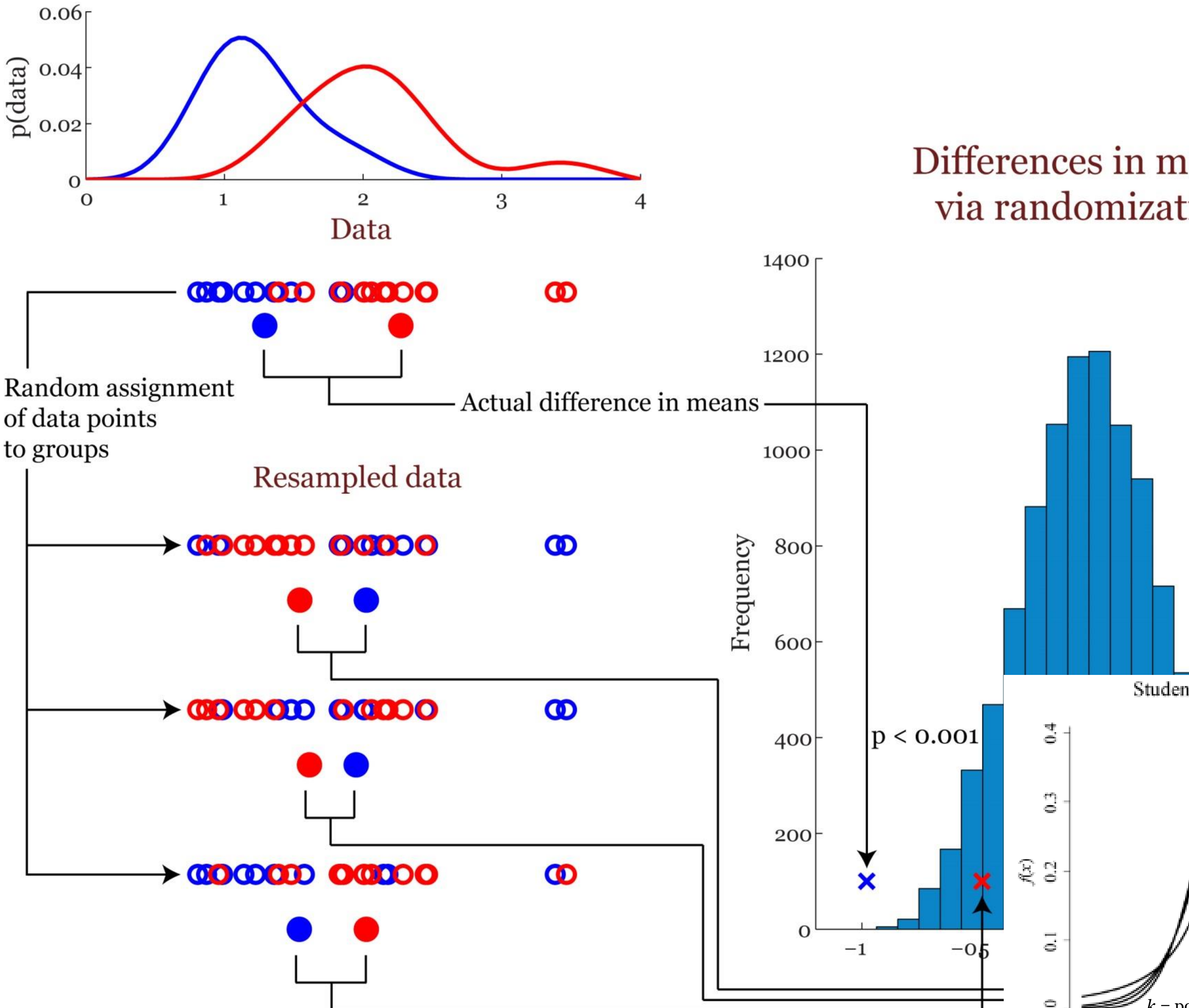
How to report them?
$p <1 , 0.01>$: 2 digits ($p = 0.02$, $p = 0.51$; 3 digits in special cases $p = 0.049$)
$p (0.01 , 0.001>$: 3 digits ($p = 0.009$, $p = 0.001$)
$p (0.001 , 0>$: always $p < 0.001$

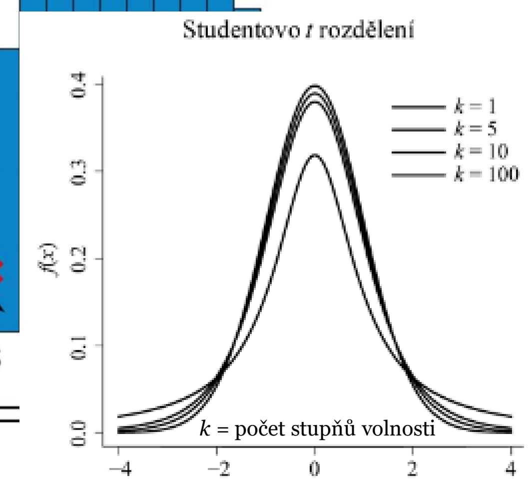
Differences in means via randomization



k = počet stupňů volnosti

Differences in means via randomization





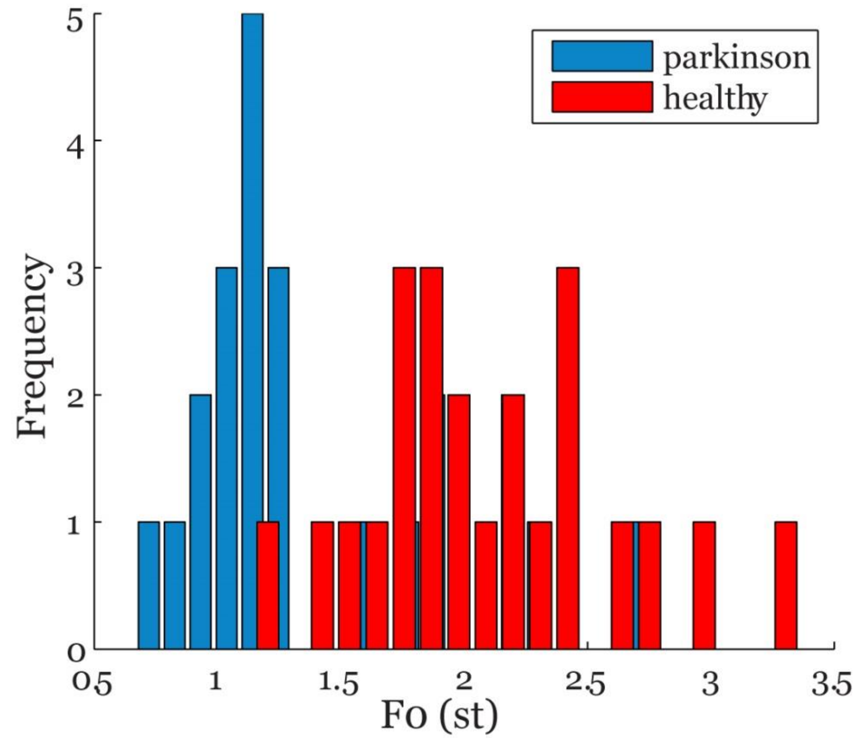
Group differences
for normally distributed data (t-test)

Matlab example 1

Researcher wants to verify if the intonation patterns differ between speakers with Parkinson's disease (PD) and healthy control speakers. He collected short reading texts from 23 speakers with PD and age- and 23 sex-matched controls, extracted fundamental frequency contour and converted it into semitone scale (st).

How to report the results?

$t(45) = -4.40$, $p < 0.001$



Matlab example 2

Group differences
for non-normally distributed data (Wilcoxon rank sum test)

Normal speaker is able to perform sustained vowel phonation without voice breaks that represent impaired function of vocal folds. To verify if vocal fold function differs in patients with Huntington's disease (HD), researcher collected sustained phonations from 34 speakers with HD and 34 age- and sex-matched controls and extracted analyzed number of voice breaks (nvb).

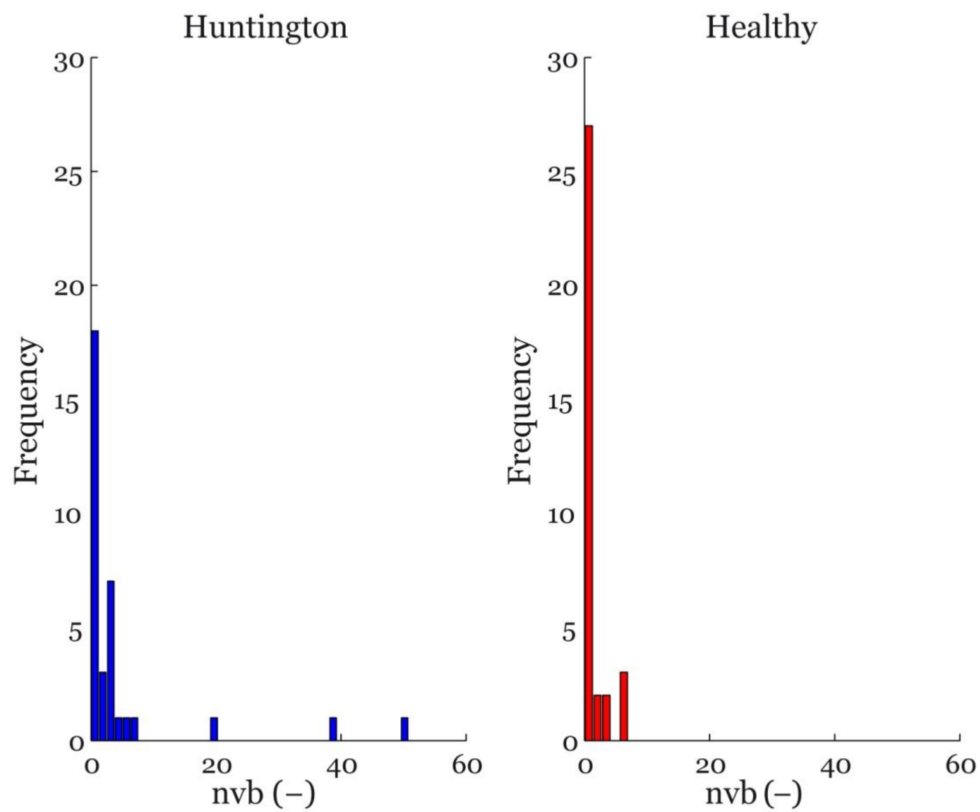
How to report the results?

$z(67) = 3{,}81$, $p < 0.001$

What about t-test?

$t(67) = 1.98$, $p = 0.051$

Failed to found significant group differences



Independent samples

Scores are separate (e.g. testing the blood pressure of group of people on active drugs against group of people taking placebo)

Paired samples

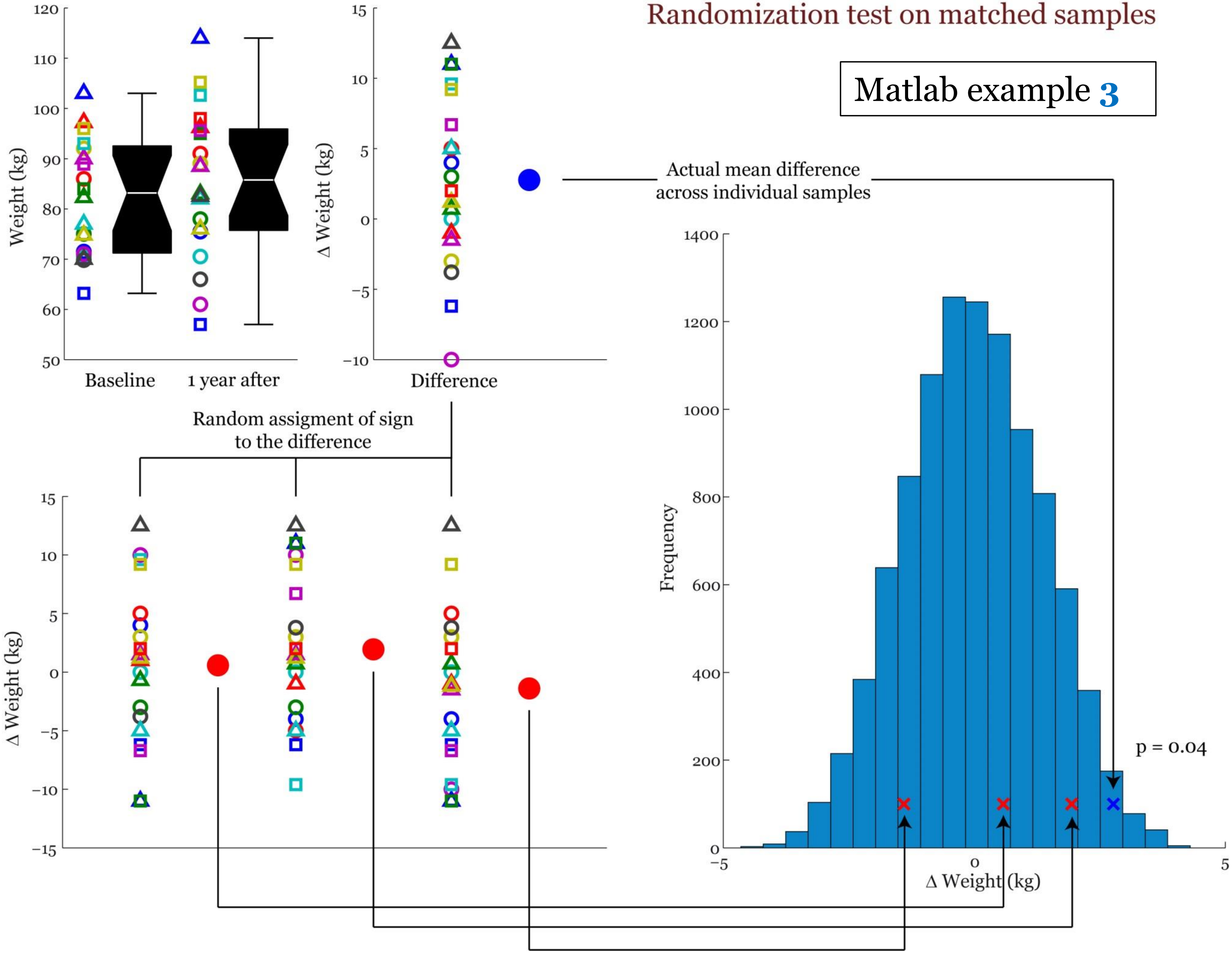
Scores are linked (e.g. measuring the blood pressures of the same people before and after they receive a dose)

How to use paired test?

Is there any change in weight of 20 patients with Parkinson's disease one year after introduction of dopaminergic drugs?

Randomization test on matched samples

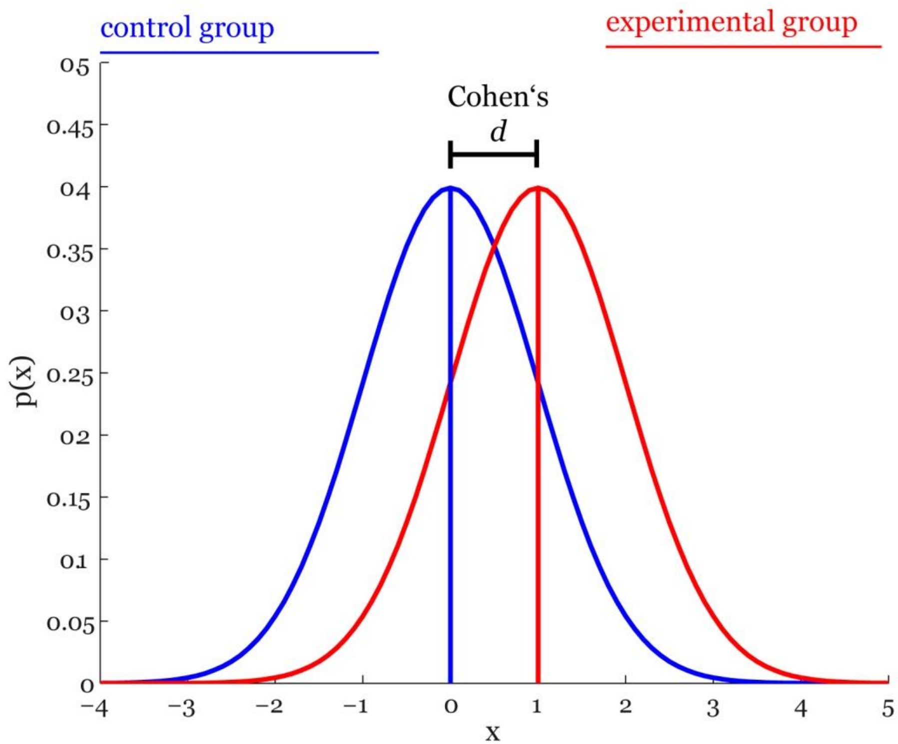
Matlab example 3

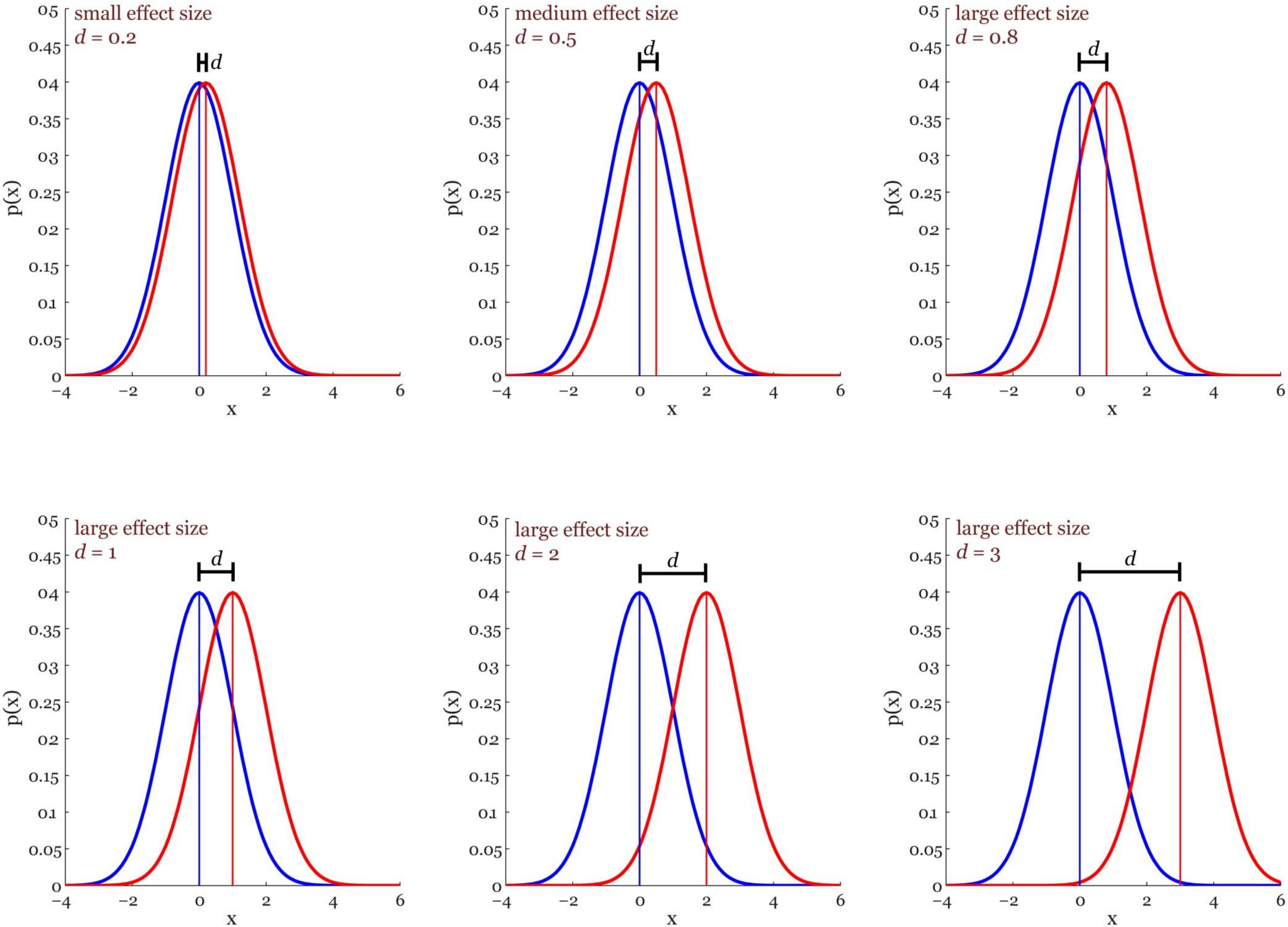


Cohen‘s effect size:

a quantitative measure of the strength of a phenomenon, independent of the measured units.

$$d = \frac{\mu_c - \mu_e}{\sigma} \qquad \sigma = \sqrt{\frac{\sigma_c^2 + \sigma_e^2}{2}}$$



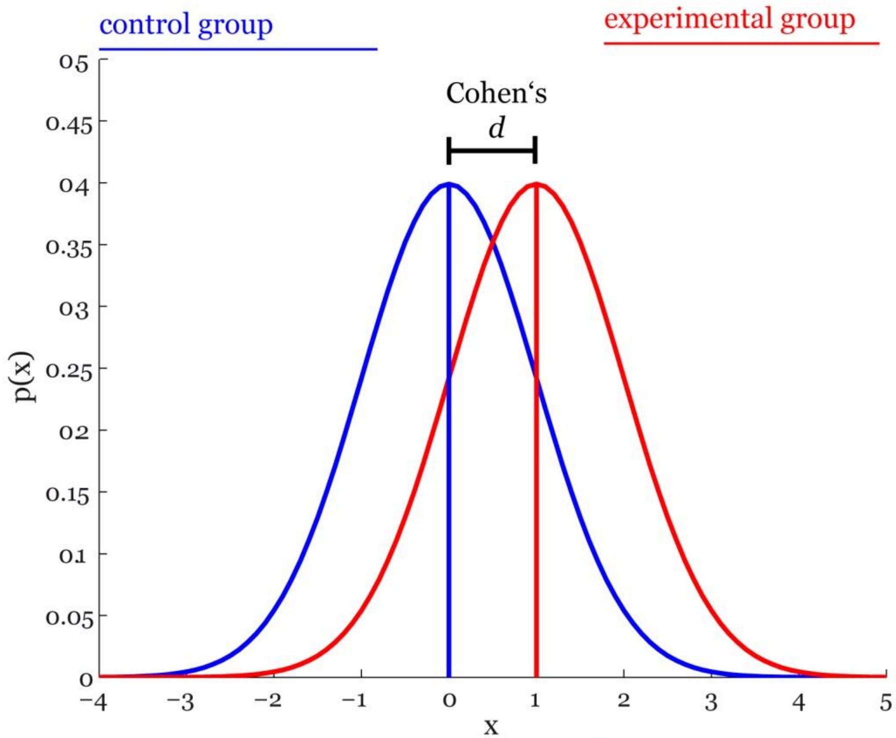

small effect size
d = 0.2
d
medium effect size
d = 0.5
d
large effect size
d = 0.8
d
large effect size
d = 1
d
large effect size
d = 2
d
large effect size
d = 3
d
p(x)
x
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
−4
−2
0
2
4
6

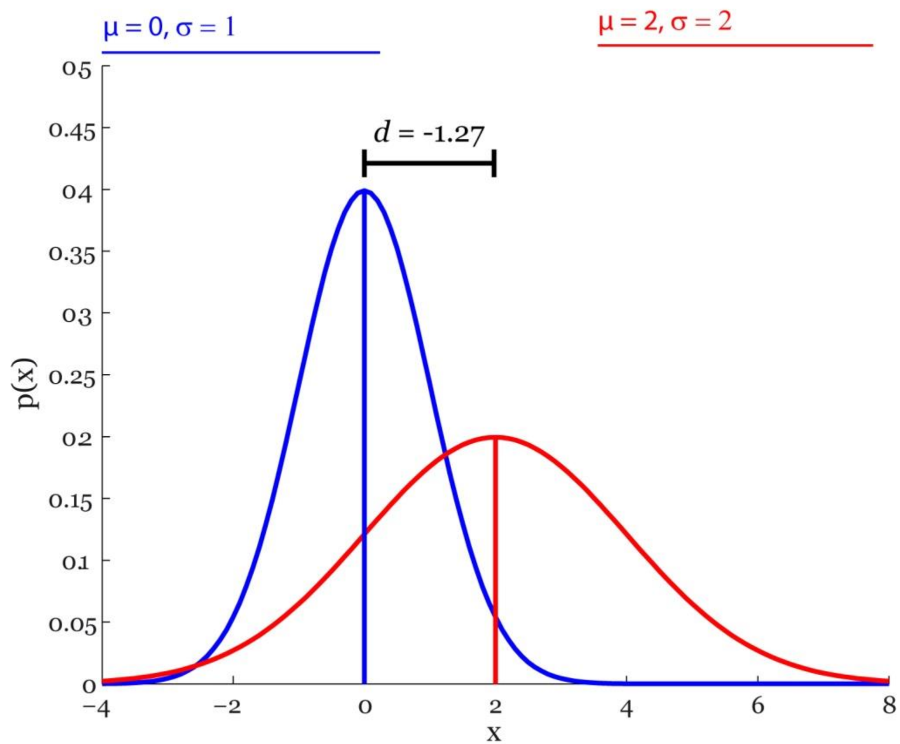
Cohen‘s effect size:

a quantitative measure of the strength of a phenomenon, independent of the measured units.

$$d = \frac{\mu_c - \mu_e}{\sigma} \qquad \sigma = \sqrt{\frac{\sigma_c^2 + \sigma_e^2}{2}}$$

$$\sigma = \sqrt{\frac{1^2 + 2^2}{2}} = 1.58 \qquad d = (0-2)/1.58 = -1.27$$





Effect size estimate for unequal sample size: $d = \frac{\mu_c - \mu_e}{\sigma} \qquad \sigma = \sqrt{\frac{(n_c - 1)\sigma_c^2 + (n_e - 1)\sigma_e^2}{n_c + n_e + 2}}$

Non-parametric effect size estimate for Mann-Whitney U test: $r = \frac{z}{\sqrt{n_c + n_e}}$