Statistical methods in biology and medicine

Statistics

- Group of mathematical methods concerning the collection, analysis and interpretation of the data
- A complete description of the world is both impossible and impractical (statistics represent a tool for reducing the variability of the data)
- Statistics creates mathematical models of the reality that can be helpful in making decisions
- It works correctly only when the assumptions of its methods are met

Descriptive statistics

- Population-wide – works with the data related to whole surveyed population (e.g. census, medical registry)
- Inductive – conclusions based on sample data (obtained from a part of the target population) are extrapolated to whole population (assumption: random selection of the sample)

Statistics as a data processing tool

- „raw data" – often difficult to grasp
- Descriptive statistics can make the data (of given sample) understandable

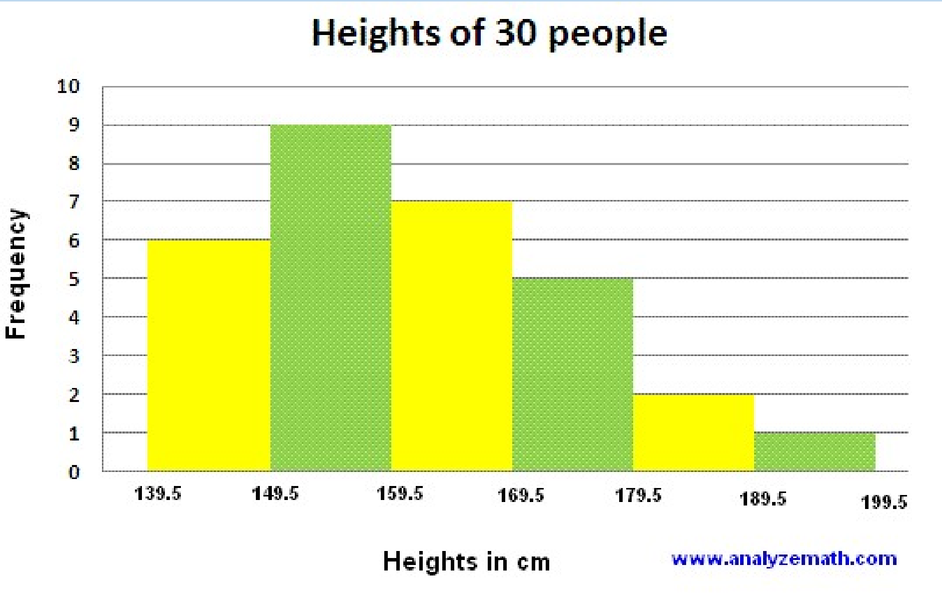
kod	cislo	adrenalin	noradrenalin	hypokineza	ERα 397/Pvull	ERα 351/Xbal
TTCBI13-2013	1	354	3643	baze	CT	AG
TTCKE14-2013	2	307	2955	apex	TT	AA
TTCKH15-2013	3	473	6076	apex	CT	AG
TTCAJ16-2013	4	341	2108	apex	CT	AG
TTCCHM17-2013	5	321	2031	apex	CC	GG
TTCCHS18-2013	6	426	1931	apex	TT	AA
TTCRK19-2013	7	508	1753	difuzni	TT	AA
TTCPD20-2013	8	374	1088	difuzni	CT	AA
TTCMJ21-2013	9	597	1798	apex	CC	GG
TTCPO22-2013	10	420	2856	apex	CT	AG
TTVVA23-2013	11	367	2657	apex	CT	AA
TTCNL24-2013	12	327	2467	apex	CT	AG
TTCJF25-2013	13	395	3929	apex	CC	GG
TTCZM26-2013	14	344	3706	apex	CT	AG
TTCHJ27-2013	15	426	4225	apex	TT	AA
TTCGT28-2013	16	265	2406	apex	CT	AG
TTCSB29-2013	17	295	3186	apex	CT	AG

Kinds of data

- Continuous (always quantitative) – the parameter can theoretically be of any value in a given interval (e.g. glucose concentration: 0-∞; ejection fraction: 0-100%)
  - Ratio vs. interval data – only differences, but not ratios of two values can be determined (e.g. IQ score)
- Categorical (usually qualitative) – the parameter can only be of some specified values (e.g. blood group: 0, A, B, AB; sex: male, female; a disease is present/absent)
  - Ordinal data – are categorical, but quantitative (they can be ordered – e.g. heart failure classification NYHA I-IV)
  - Count data – can be ordered and form a linearly increasing row (e.g. number of children in a family: 0,1,2…) - they are often treated as continuous data
  - Binary data – only two possibilities (patients / healthy controls)

The distribution of continuous data - histograms

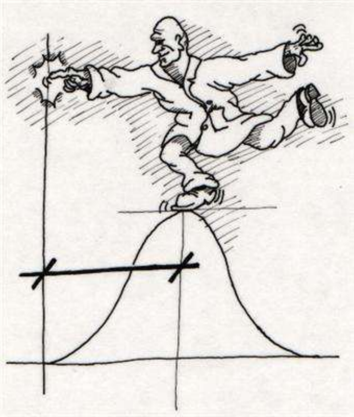
- The distribution of a continuous parameter can be visualized graphically (e.g. using histograms)
- The values usually cluster around some numbers

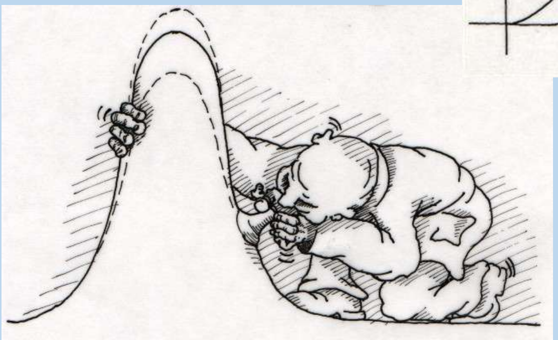


Description of continuous data

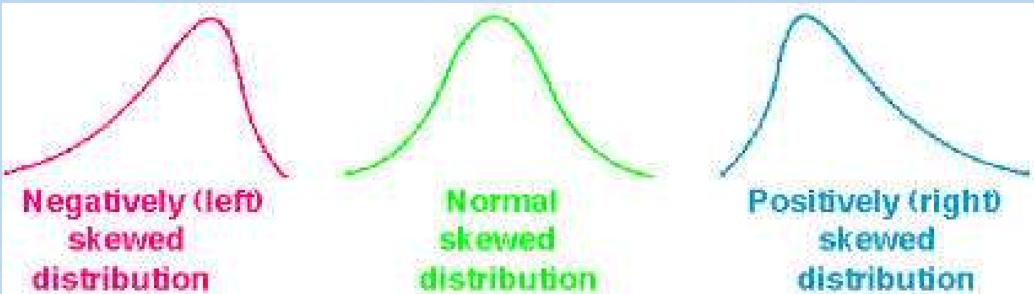
- **Measures of central tendency**
  - The arithmetic mean (μ)
    - sum of values divided by their number (n)
  - The median (= 50% quantile)
    - cuts the order of values in half
  - The mode
    - most frequent value
- **Measures of variability**
  - variance (σ^2)
  - standard deviation (SD, σ)
  - coefficient of variance (CV)
    - CV = σ/μ
  - standard error of mean (SE, SEM = $\sigma/\sqrt{n}$)
  - min-max (= range)
  - quartiles
    - upper 25%
    - median
    - lower 75%
  - skewness
  - kurtosis

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x} - x_i)^2}{n-1}}$$



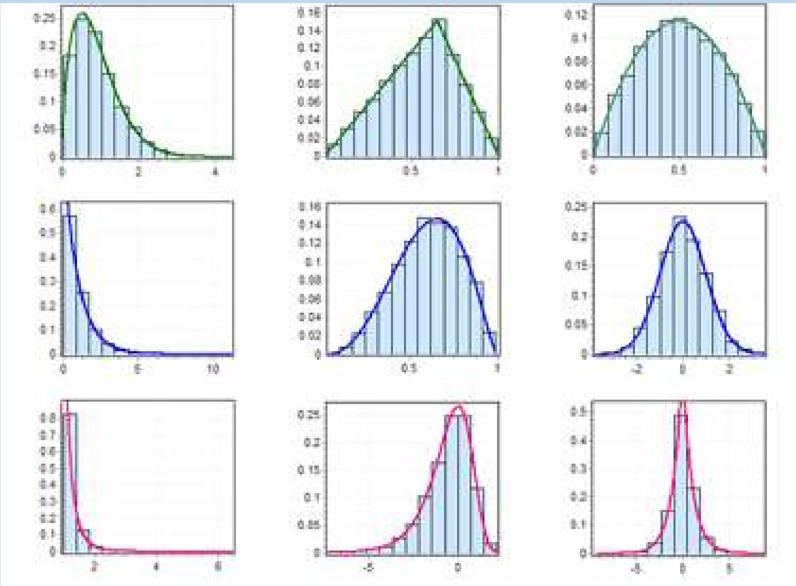


The probability distribution of continuous random variable



- Probability density function
- In graphs each (continuously) quantifiable variable (x axis) is linked to its probability (y axis)

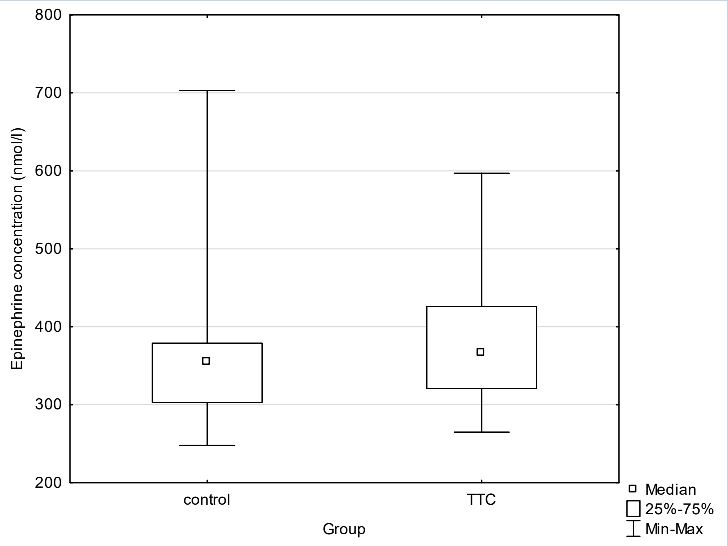
Examples of continuous data distribution



Histograms + corresponding probability density functions

Other ways of graphical visualisation

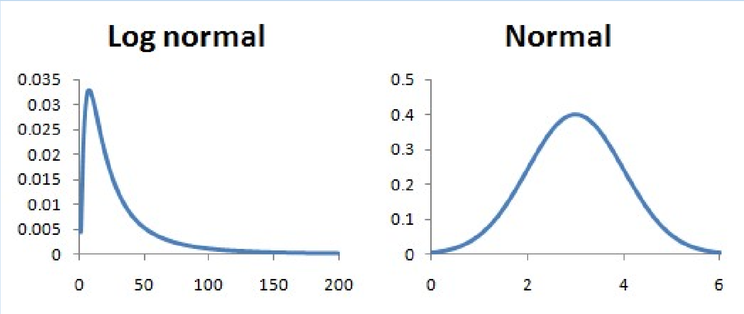
- Box and whisker plots



- Instead of median e.g. mean can be used, instead of quartiles („box“) ±σ, instead of range („whiskers“) e.g. non-outlying values… etc.

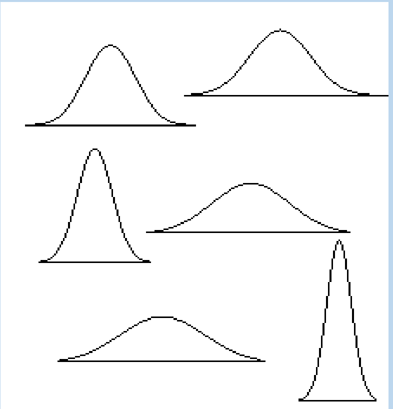
Normal distribution of the data

- Defined by the Gaussian function $y = a^{-(x-b)^2/2c^2} + d$, where a, b, c, d are real numbers
- Graphical representation is the **Gaussian „bell“ curve**
- mean = median = mode
- A random variable **x** is normally distributed when its value can be interpreted as the sum of an infinite number of independent effects with equal absolute value
- E.g.: throwing a coin, we assign the value of +1 to a head and -1 to a tail. When throwing many times ($n \rightarrow \infty$), the probability distribution of the resulting value will be normal

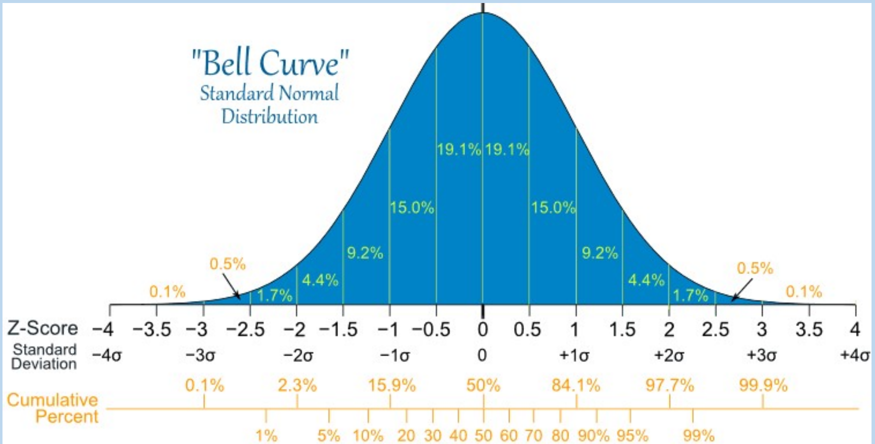


- Log-normal distribution: after logarithmic transformation of the data, we obtain the Gaussian curve (with **the geometric mean** at its peak) – an example of data transformation

Normal (Gaussian) vs. symmetric distribution



- Not each symmetric distribution is normal
  - Meeting of several assumptions is necessary
    - interval frequency distribution
    - distribution function
    - skewness = 0, kurtosis = 0
  - Data transformation („normalization“)
    - Creating normal distribution by applying a formula
- Student distribution is an approximation of normal distribution in small datasets

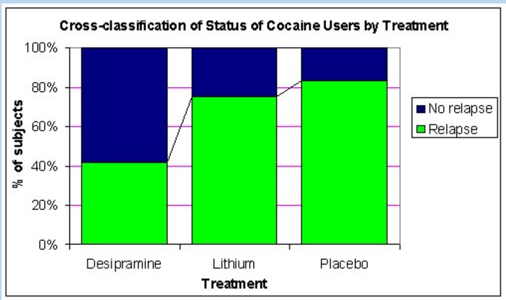


Description of categorical data

- Sumarization of given categories of the dataset (frequency table)
- When more than one categorical parameter is available, we can create **contingency tables** (and, based on them, we can eventually create graphs)

		Body Image			
		About Right	Overweight	Underweight	**Total**
Gender	Female	560	163	37	760
	Male	295	72	73	440
	Total	855	235	110	1200

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Totals	87	13	100

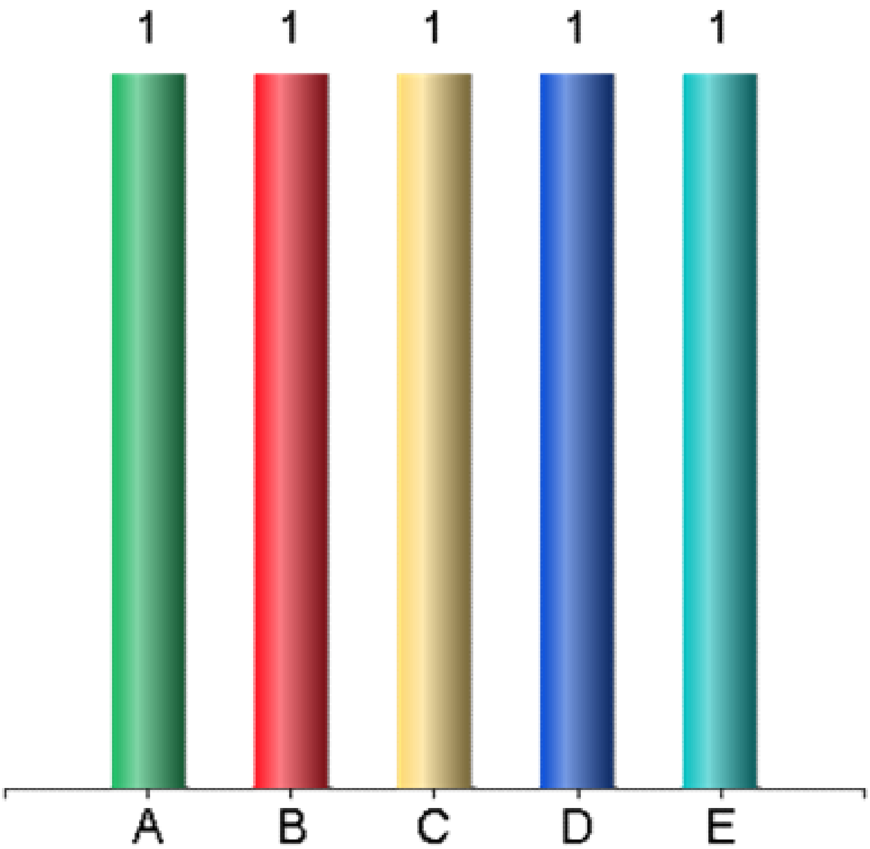


Expression of categorical data variability - examples

- Variation ratio
  - $v = 1 – (f_m/N)$, where f_m = number of cases in mode (most frequent) category, N = total number of cases
  - Ratio of all cases **except** the mode category to the total number of cases
- Shannon-Wiener diversity index
  - Expresses the uncertainty in prediction, to which category will a case belong
  - $H' = -\sum p_i * \ln(p_i)$, where p_i is a proportion of category i from total sample
  - If p_i = 100% then H' = 0; Higher value corresponds with higher diversity
  - Widely used in ecology, common range between: 1,5 – 3,5

The level of education (basic, high school, university) is an example of...

A. Ordinal variable ✓
B. Interval variable
C. Binary variable
D. Continuous variable
E. Qualitative variable



formulation of statistical hypotheses

- Research hypothesis (e.g. drug A has better effect than drug B, blood pressure decreases during the treatment, there is a correlation between sex and body height etc…) – can be formulated both for an experiment or for an observation
- Testing of research hypothesis uses a proof by contradiction
- For statistical hypothesis testing, a **null hypothesis H_0** must be defined (e.g. between two groups, there is no difference in means, there is no difference in variances, there is no correlation between two parameters, a parameter does not change in time…resp. any observed difference is only due to a chance)
- During the testing of null hypothesis, we try to refute it (or, more exactly, to show that it is highly improbable)
- If the null hypothesis, is not true, then its negation must be true – **alternative hypothesis H_A** (there is a difference, there is a correlation…)
- The result of hypothesis testing can thus be:
  - A) non-refutal of the null hypothesis (at certain level of statistical significance α)
  - B) refutal of the null hypothesis favouring the alternative hypothesis

Repetitions - errors in hypothesis testing

	Real nature of the null hypothesis	
Statistical decision	H_0 true	H_0 false
H_0 refuted	**type I error (α) = false pos.**	**Correctl y pos. (1-β)**
H_0 confirmed	**Correctly neg. (1-α)**	**type II error (β) = false neg.**

- Type I error rate (α) – also **significance level**
- α must be defined before the statistical testing – 0.05 is usual in biomedicine (i.e. when H_0 is refuted, there is 95% certainty, that it is really false and the observed difference/correlation is real)
- 1-α = specificity of a statistical test
- 1-β – also **power of a test (sensitivity of a test)**
- P-value – probability that the observed result was obtained under the assumption that H_0 is true
- **When p < α, we refute the null hypothesis at a given significance level and the alternative hypothesis is valid**
- We say that the difference (effect) is **statistically significant** (that, of course, does not mean that it has to be significant practically)

P-value

- “A p-value doesn‘t “prove“ anything. It's simply a way to use surprise as a basis for making a reasonable decision.“

Cassie Kozyrkov

Statistical tests

- For different statistical hypotheses, different tests are used
- The selection of the right test depends on:
    - the number of compared groups
    - the character of the data (categorical vs. continuous)
    - the distribution of the data
    - mutual dependence of the data

Power of a test...

A. Express its practical (not statistical) significance

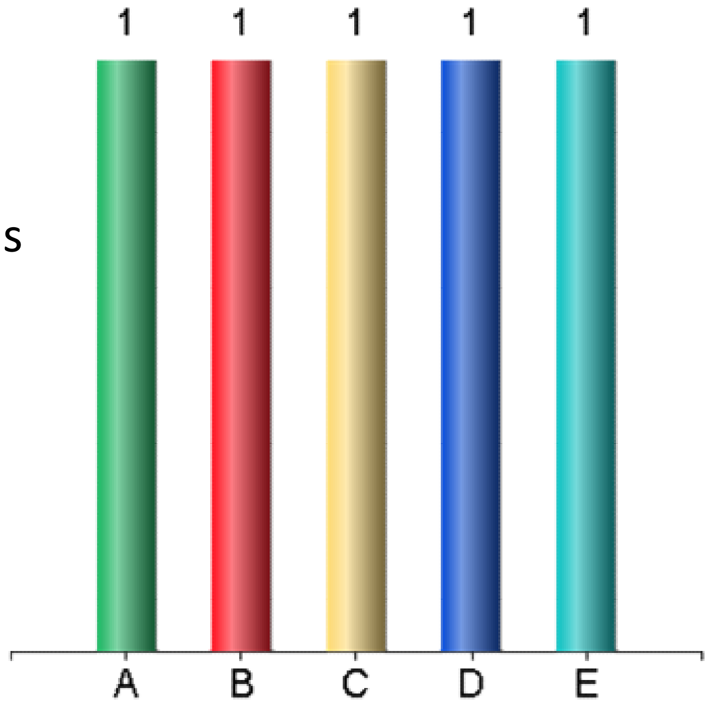
B. Increases with increasing variability of the data

✔ C. Express the ability of a test to correctly refute the null hypothesis

D. Is expressed by a letter p

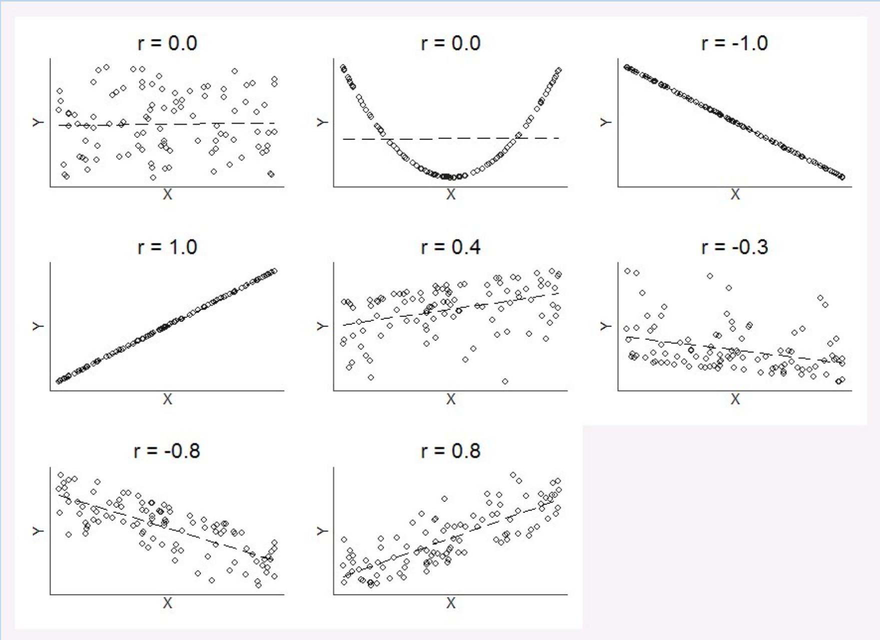
E. Is a probability that the alternative hypothesis is true in a case

when the null hypothesis is refuted



Correlation of two continuous parameters

- Mutual dependence of two parameters - correlation
- Expressed by correlation coefficient (r)
- r generally express the size of the effect
- r can achieve values in the interval from -1 to 1, where 0 corresponds to no correlation, 1 corresponds to 100% positive correlation (when one factor increases, the other does the same) and -1 corresponds to total negative correlation



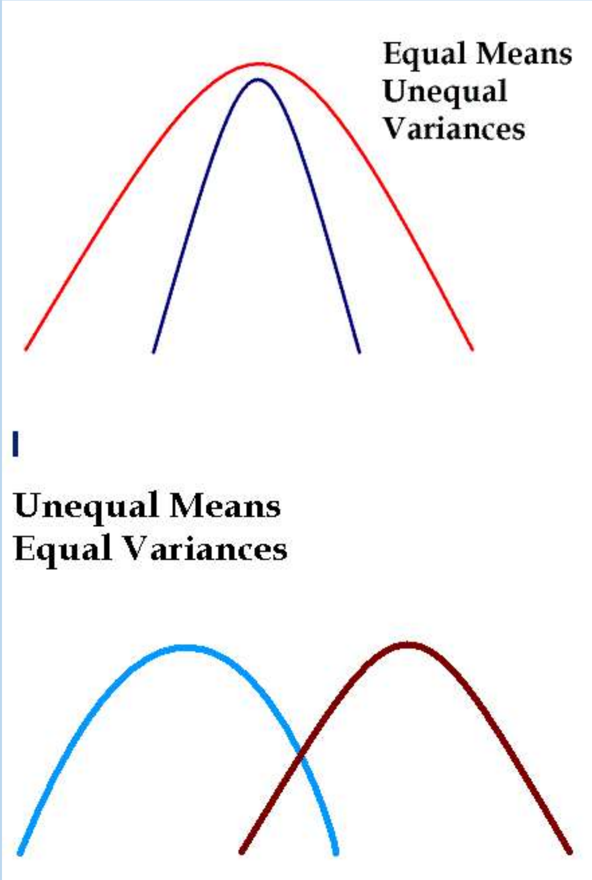
- Besides r, the corresponding p-value can be determined (H_0 – the variables are independent)
- Correlation of a categorical vs. continuous variable – see the „tests for continuous data“ (categorical variable define the groups that are compared by the tests)

Examples of correlation coefficients

- Pearson coefficient (parametric) – measures linear correlation between variables
  - The main assumption is approximately normal distribution of the data
- Spearman coefficient (non-parametric) – measures the rank correlation of the variables
- None of the coefficients can reveal e.g. U-shaped dependence

Comparing the continuous variable in two and more samples

- H_0 – there is no difference in the value of the variable between the samples (or is due to chance – e.g. the concentration of glycated hemoglobin in treated and untreated diabetics is equal)
- Generally, central tendency (more often; see further) or variability (e.g. F-test, Levene test) can be tested



Parametric vs. non-parametric tests for continuous data

Parametric

- Use the values
- Have higher power, but only when their assumptions are met (esp. normal distribution of the data in each sample)
- If the distribution is not normal, we can try to transform (normalize) them

Non-parametric

- Use ranks of values
- Power is generally lower (but the difference is small in big samples)
- They are more „robust“ – their use is not that dependent on data distribution

The normality can be tested by normality tests (e.g. Kolmogorov-Smirnov, Shapiro-Wilks – they compare the real distribution with the normal distribution) and „by eye“ evaluation of whether the histograms correspond to Gaussian curve (in small samples, the normal probability plot is a better choice)

Tests for continuous data - paired vs. unpaired tests

Paired (matched samples)

- Used when to each value from sample A, we can match one value from sample B that differs only by its membership in the sample (e.g. comparing salaries in two hospitals: director A – director B; head physician A – head physician B… up to charwoman A – charwoman B)
- **Most often, this design is used to assess the change in time** (e.g. patients‘ weight now vs. after 5 years: patient XY – and other patients – is the same person now as well as after 5 years and differs only by the time difference)
- They assess differences between the samples (or their ranks)

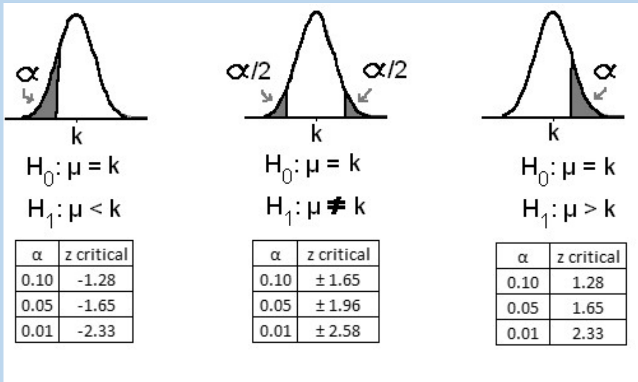
Unpaired (unmatched samples)

- Used in independent samples (they can differ in size)
- They compare the actual values of the variable between the samples (or their ranks)
- It is necessary to decide between the paired or unpaired design before the start of the study (pairing is technically challenging, but paired tests have higher power)

One-tailed vs. two-tailed tests

One-tailed

- H_0 is asymmetric: e.g. drug A is not better than drug B – but we are not interested whether it is or is not worse
- They have higher power



α	z critical
0.10	-1.28
0.05	-1.65
0.01	-2.33

α	z critical
0.10	± 1.65
0.05	± 1.96
0.01	± 2.58

α	z critical
0.10	1.28
0.05	1.65
0.01	2.33

Two-tailed

- H_0 is symmetric: there is no difference between drug A and drug B (i.e. A is neither better nor worse than B)
- They can reveal the differences in both ways
- They are usually more suitable – we don‘t know the result a priori, and we are interested in both possible effects

Tests for continuous data, 2 samples – examples

Test	Parametric	Non-parametric
Paired	Paired (dependent) Student‘s t-test	Wilcoxon paired test Sign test
Unpaired	Unpaired (independent) Student‘s t-test	Mann-Whitney U-test * Kolmogorov-Smirnov test

- * has almost the same power as unpaired t-test, but it has an assumption of similar variability in both samples (as well as t-test)

Tests for continuous data, more than 2 samples – examples

Test	Parametric	Non-parametric
Paired	Repeated measures ANOVA (Analysis Of VAriance) – RMANOVA	Friedman test („ANOVA“)
Unpaired	One-way ANOVA (and its variants)	Kruskal-Wallis test („ANOVA“)

- When ANOVA rejects H_0, it is necessary to find out which specific samples differ from each other – post hoc tests

Choose the best test

In a clinical trial, patients take either a new drug to treat epilepsy or a placebo. The study is randomized (the study group is randomly drawn). Only patients, which have at least one and at most ten seizures in three months are included. The study evaluates a number of seizures during the first year of treatment

A. Paired t-test

B. Unpaired t-test

C. Mann-Whitney U-test

D. Sign test

E. Repeated measures ANOVA