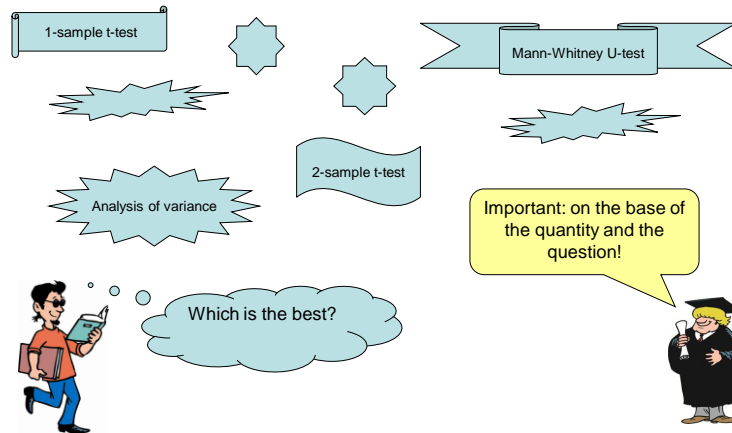


## Selection



## According to the variable

parametric

Testing the hypothesis on the parameter(s) of a known distribution.  
(most frequently the normal distribution).

non-parametric

Testing the hypothesis on the parameter(s) or types of an unknown distribution.

## Non-parametric tests

Distribution free methods.

advantage : independent from the distribution.  
disadvantage: normally it's power is less.

**Ranking tests:**  
Instead of original values we use the so-called **rank**s.

## Ranks



**Rank:** numerical or ordinal data belonging to a value in data series sorted according to a certain rule.

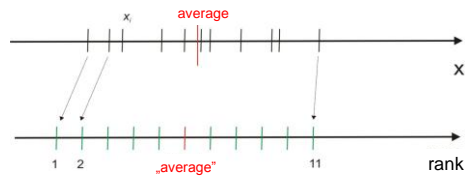
**„tie“:**

In the case of same values every value replaced by the average of the ranks.

e.g.:  
•lieutenant  
•major  
colonel  
etc.

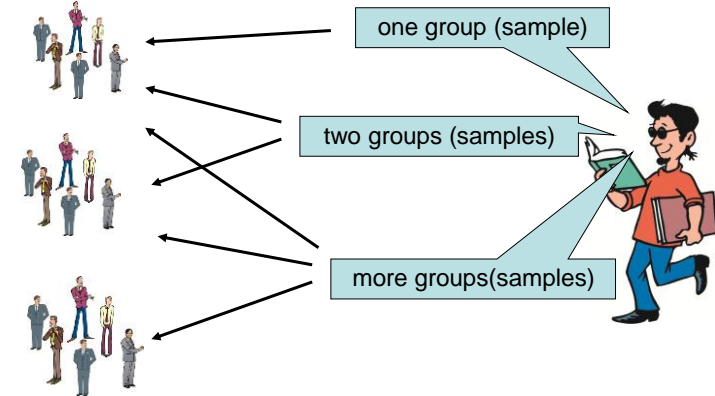
value:	1.2	2	2	3.5	4
rank:	1	2,5	2,5	4	5

**the „average” of the ranks is the median**



The median plays the role of the average.

**according to the question**



## **Summary table**

	parametric	non-parametric
<b>one group</b>	1-sample t-test,	Wilcoxon's signed-rank test, sign-test
<b>two groups</b>	2-sample t-test	Mann-Whitney U-test
<b>more groups</b>	ANOVA	Kruskall-Wallis test

## **Examination in one group**

Question: On the base of the sample the parameter of the population may be a given value?

**parametric**

$\mu = ?$

Null hypothesis:  $\bar{x} = \mu_0$

1-sample t-test

**non-parametric**

median = ?

Null hypothesis: the median is a given value

Wilcoxon's signed-rank

## 1-sample t-test

**example:** The medicine effective or not?



**Nullhypothesis:** not!  $\mu_0 = 0$ . But the average is not 0!

sample	Average
1.	-0.2 °C
2.	-1 °C
3.	-1.5 °C



If the difference is bigger, it seems to be more probable being non random.

## What does it mean big difference?

What is the measure of the difference?

**Standard error:** the average deviations of the averages from the  $\mu$ .

$(\bar{x} \pm s_{\bar{x}})$  ~ 68% - confidence interval.

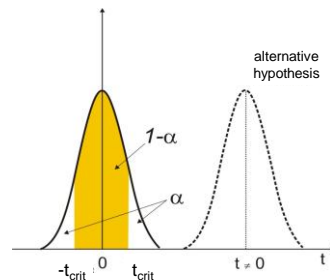
## t-value

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

Compare the difference to the standard error!  
( $\mu_0$  very frequently = 0)

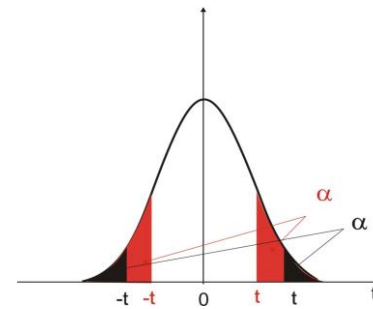
The averages fluctuate around the  $\mu_0$  so the  $t$ -values deviate around the 0.

(providing, that the nullhypothesis is true!)



## Why the t-value is better?

We are able to calculate the probabilities on the base of this distribution!!! (Student- or  $t$ -distribution)



It describes only the **random deviations** of the  $t$ -values!

The shape of the distribution depends on the no. of elements.

## Why has it t-distribution?

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} \longrightarrow \bar{x}$$

The fluctuation of the average has normal distribution. The numerator is variable having normal distribution!

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \longrightarrow s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}$$

variance is the sum of squared probability variables.

remember!

t-distribution

$$\xi_n = \frac{\sqrt{n} \cdot \xi}{\sqrt{\sum_i \xi_i^2}}$$

(Quad erat demonstrandum)

the **t** variable has t-distribution.



## Degree of freedom (d.f.)

I think 3 numbers! (sample)

The average of them: 8! (information!)



they must be!

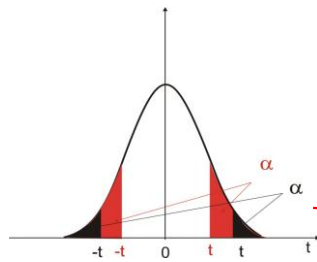
3, 12, 8 or 5, 7, 11 etc.

d.f. = n

3, 12, **9** or 5, 7, **12** etc.

d.f. = n-1

## The t-table



t-table

significance level

d.f.	0.1	0.05	0.02	0.01
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

Different  $t_{crit}$  values belong to different significance level.

d.f.: n-1

## Decision the base of t-table

t-table

significance level

d.f.	0.1	0.05	0.02	0.01
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

d.f.: n-1

Select an appropriate significance level!

If  $\geq 2.78$  **reject**, if smaller **accept** the nullhypothesis.

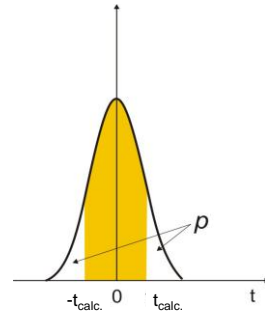


## Decision using computer

I am able to integrate!!!



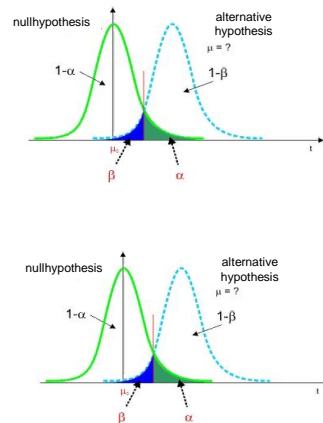
$p$ : probability, that the  $t_{\text{calculated}}$  is so large randomly.



## Decision

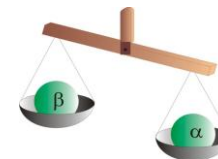
- 1. If the probability of the random deviation is small ( $p(|t| \geq t_{\text{crit}}) \leq 5\%$ ) – **reject** the null hypothesis.
- 2. If the probability of the random deviation is large ( $p(|t| \geq t_{\text{crit}}) > 5\%$ ) – **accept** the null hypothesis.

error?



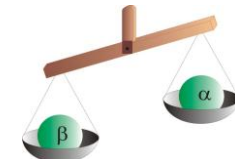
## The error

Reject the null hypothesis



$\alpha$  is the measure of the error. Smaller  $p$  is better.

Accept the null hypothesis



$\beta$  is the measure of the error. Larger  $p$  is better.

## Condition for 1-sample t-test

- Task: Decision about the  $\mu$  on the base of one sample.
- The variable must have **normal distribution**.



What can we do if it isn't true?

## Wilcoxon's signed-rank test

Example: Is there an effect of an entertaining movie on the patients? (The numbers are scores)

n	before	after	Diff.
1	2	2	0
2	0	1	1
3	3	2	-1
4	2	4	2
5	1	3	2
6	3	3	0
7	1	4	3
8	1	5	4
9	5	2	-3
10	4	4	0

Normal distribution?



## Ranking

Sort the absolute values of the differences (without 0-s)! Let the sign of ranks be same then the differences!  
Calculate the averages and sd of signed-ranks!



Diff.	absolute value	rank	Signed-rank
0	0		
1	1	1.5	1.5
-1	1	1.5	-1.5
2	2	3.5	3.5
2	2	3.5	3.5
0	0		
3	3	5.5	5.5
4	4	7	7
-3	3	5.5	-5.5
0	0		

## The nullhypothesis

There is no effect of the movie!



The median = 0!  
The deviation is random!

$$H_0: \mu_0 = 0$$

$$H_1: \mu_0 \neq 0$$



## known distribution



$$t = \frac{\bar{R} - 0}{s / \sqrt{n}}$$

If n is enough large!

$\bar{R}$  - the average of the signed-ranks

s - the standard deviation

Remember!  
„average” of the ranks = median



## Decision

This is known!!!

Of course! This is similar to the 1-sample t-test!!!



## Paired t-test

If the data may be paired according to a rule!

Observation on the same person, paired organ (e.g. kidney).

Rare, on the base of viewpoints (age, profession, etc.).

Look at:  
decreasing  
the fewer.



Experimental design

## Experimental design

test from previously collected data?

**practical order:**  
Question →  
experimental design  
→ calculation.

Many problems: e.g. a few data are suitable only.



## „real” 1-sample t-test

Is it possible that the  $\mu$  is equal to a value?

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

Rare case.



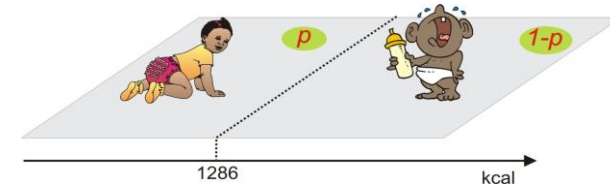
## sign-test

Example: Energy uptake in the population of 2-year old children.

**Question:** May be the median

(This derives from an another test) is 1286 kcal?

**Nullhypothesis:** median = 1286 kcal (deviation is random).



## Test

Small sample

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

binomial distribution

Large sample

$$z = \frac{|x - np| - 1/2}{\sqrt{np(1-p)}}$$

standard normal distribution

$x$  – no. of children below 1286 kcal.

$n$  – no. of children in test.

$p$  – probability, that randomly smaller (look at: binomial distribution)

## Decision

Calculate the probability of the random deviation. (binomial, or standard normal distribution)

End of this part!

