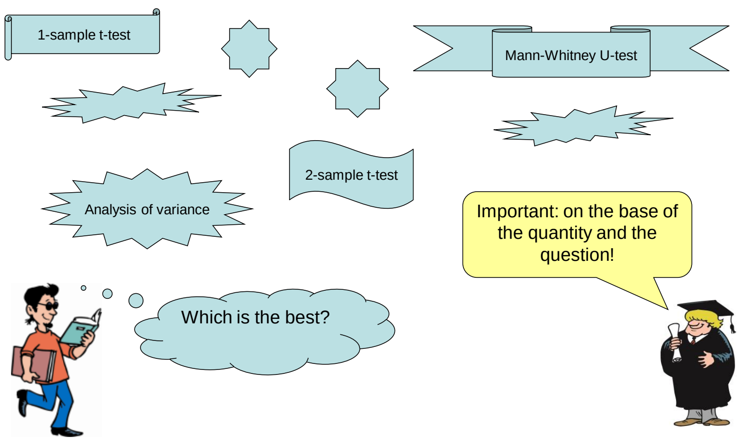
Selection



On the base of variable

numerical

Values of the variable are used in calculations.

categorial

Calculations are based on the frequency data.

On the base of 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 of the distributions.
disadvantage: normally it has less power.

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

Ranks

<u>Rank:</u> numerical or ordinal data belonging to a value in data series sorted according to a certain rule.

<u>Ties</u>:
In the case of same values every value replaced by the average of the ranks.

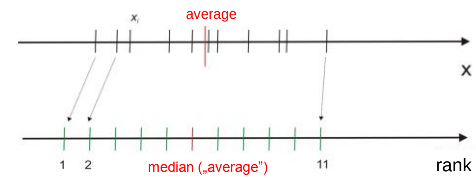
e.g.:
- lieutenant
- major

colonel

etc.

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

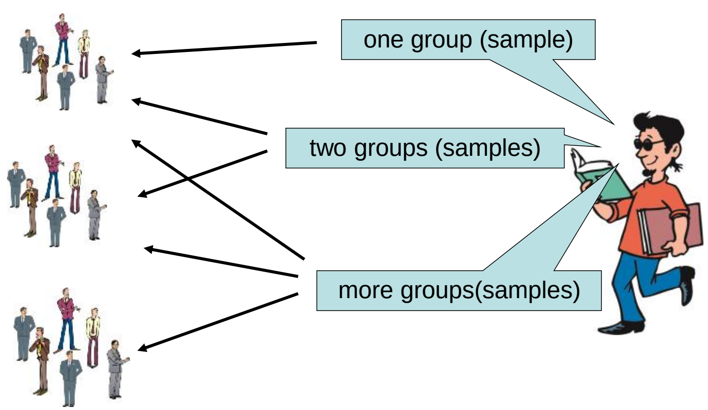
the „average" of the ranks is the median



<u>Rank scale is an equidistance one</u>!

The median plays the same role as the average.

On the base of the question



Summary table of the most frequently used tests

	parametric	**non-parametric**
one group	1-sample t-test,	Wilcoxon's signed-rank test, sign-test
two groups	2-sample t-test	Mann-Whitney U-test
more groups	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	non-parametric
μ = ?	median = ?
Nullhypothesis: $\bar{x} = \mu_0$	Nullhypothesis: the median is a given value
1-sample t-test	Wilcoxon's signed-rank

1-sample t-test

Example: The medicine is effective or not?

Nullhypothesis: not! μ_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 big, it seems to be more probable to be non-random!

What is big difference?

What is the measure of the difference?

Standard error: the average devations of the averages from the μ.

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

So, we must compare the difference to the standard error!

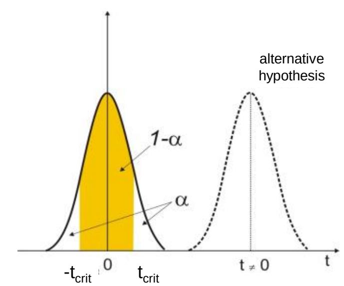
t-value

$$t = \frac{x - \mu_0}{s_x}$$

Compare the difference to the standard error! (frequently μ_0= 0)

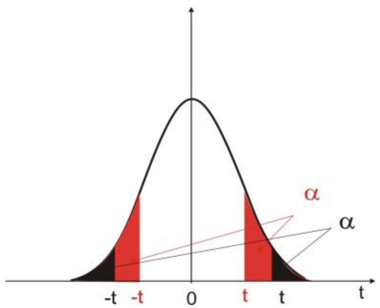
The averages fluctuate around the μ_0 so the *t*-values deviate around the 0.

(providing, that the nullhypothesis is true!)



Why is the t-value better than the original value?

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



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

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

Degree of freedom (d.f.)

I think 3 numbers! (sample)

The average of them: 8! (information!)

they must be!

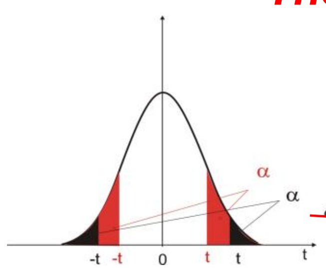
3, 12, 8 or 5, 7, 11 etc.

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
	probability (p)			
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

Different t_{crit} values belong to different sigficance levels.

Decision the base of t-table

t-table

significance level

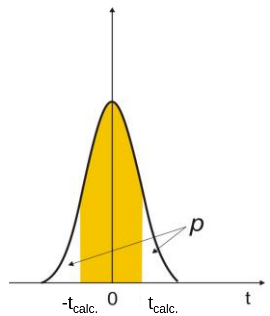
d.f.	0.1	0.05	0.02	0.01
	probability (p)			
1	6.31	12.7	31.8	63.7
2	2.92	4.3	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.6
5	2.02	2.57	3.37	4.03

d.f.: n-1

Select an approriate significance level!

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

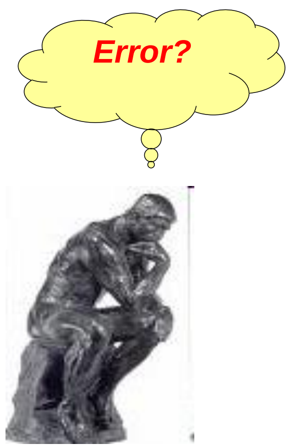
Decision using computer

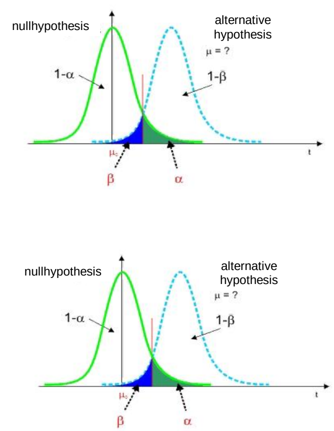


p: probability, that the $t_{calculated}$ is so large randomly.

Decision

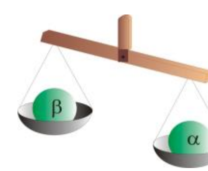
- 1. If the probalitiy of the random deviation is small ($p(|t|\geq t_{crit}) \leq 5\%$) – ***reject*** the nullhypothesis.
- 2. If the probalitiy of the random deviation is large ($p(|t|\geq t_{crit}) > 5\%$) – ***accept*** the nullhypothesis.





The error

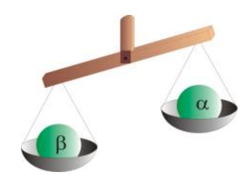
Reject the nullhypothesis



I. type error

α is the measure of the error. Smaller *p* is better.

Accept the nullhypothesis



II. type error

β is the meaure of the error. Larger *p* is better.

Condition for 1-sample t-test

- Task: Decision about the μ 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 any 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!

$\mathbf{H_0}$: $\mu_0 = 0$

H_1: $\mu_0 \neq 0$

known distribution

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

If n is enough large!

(small sample: base distribution of ranks)

$\bar{R}$ -	the average of the signed-ranks
s –	the standard deviation of them

Rememer!
„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

Experimetal design

test from previously collected data?

practical order: Question → experimental design → calculation.

Several problems may arise.

e.g. a few data are suitable only.

„real” 1-sample t-test

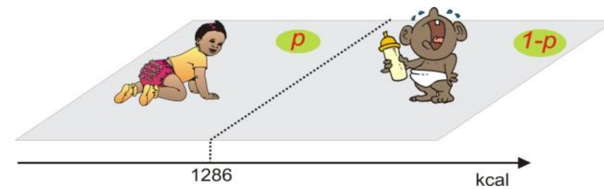
Is it possible that the μ is equal to a value?

$$t = \frac{x - \mu_0}{s_x}$$

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

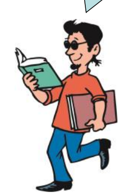
$$z = \frac{|x - np| - 1/2}{\sqrt{np(1-p)}}$$

standard normal distribution

x – no. of children below 1286 kcal.
n – no. of children in test.
p – probabilty, that randomly smaller (look at: binomial distribution)

Decision

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



End of this part!