

## Test in two groups

Question: May the samples derive from the same population? May the parameters of the two populations be the same?

**parametric**

$$\mu_1 = \mu_2 ?$$

Null hypothesis:  $\mu_1 = \mu_2$

2-sample t-test

**non-parametric**

Null hypothesis: same.

Mann-Whitney U-test

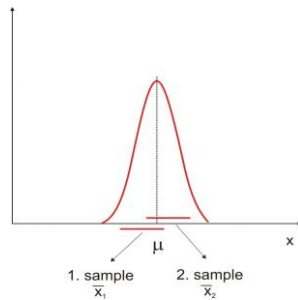
## Why is not a paired test?

	A			B			C		
	172	184	12	162	170	8	162	184	22
	180	172	-8	165	172	7	165	180	15
	165	180	15	172	175	3	172	175	3
	184	175	-9	180	180	0	180	172	-8
	162	170	8	184	184	0	184	170	-14
	Avg	3.6		Avg	3.6		Avg	3.6	
	Sd	11.33		Sd	3.782		Sd	15.11	
different s <sub>d</sub> !	Se	5.066		Se	1.691		Se	6.757	
	t	0.711		t	2.129		t	0.533	
	p	0.517		p	0.1		p	0.622	

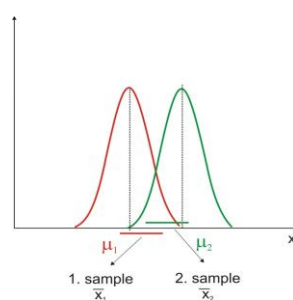
arbitrary paired values!

## Two-sample t-test

one population  
(the deviation of the averages is random)



two populations  
(the deviation of the averages is not random.)



## Standard error

$$s_1 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n_1 - 1}}$$

$$s_2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n_2 - 1}}$$

$$s_{\bar{x},1} = \frac{s_1}{\sqrt{n_1}}$$

$$s_{\bar{x},2} = \frac{s_2}{\sqrt{n_2}}$$

**Common standard error:** the weighted average of the two standard errors.

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

## 2-sample t-test

$$\bar{x}_1 \neq \bar{x}_2$$



It may be random (null hypothesis) or non-random (alternative hypothesis). Known distribution is necessary!

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^*} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

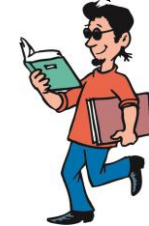
$$s^* = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2}}$$

## Test

The t-value is same!



How much is the d.f.?



$$d.f. = n_1 + n_2 - 2 \\ ((n_1 - 1) + (n_2 - 1))$$

## Conditions for the test

- Task: comparison of two **independent** samples.
- The quantity has **normal distribution**.
- The sd-s are **same** in the groups.



This is new!  
How is it proved?

## Test for standard deviations

How can I do?



Null hypothesis: the two standard deviations are the same and the difference is random (sampling error).

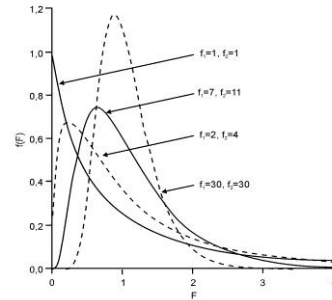
It is similar to the hypothesis testing!



## F-test

A so-called F-distribution belongs to the nullhypothesis.

$$F = \frac{s_1^2}{s_2^2}$$

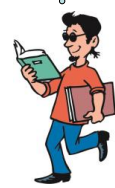


Degree of freedom:  
nominator:  $n_1-1$   
denominator:  $n_2-1$

## Degree of freedom

Using a computer it is not so important.  
Using F-table always the higher value is in the nominator.  
( $F \geq 0$  and d.f. depends on the situation.)

Which variance is in the nominator?



## Decision

- 1. If the probability of the random deviation is small ( $p \leq \alpha$ ) – we **reject** the nullhypothesis.
- 2. If the probability of the random deviation is high ( $p > \alpha$ ) – we **accept** the nullhypothesis.

## If the two standard deviations are not the same!

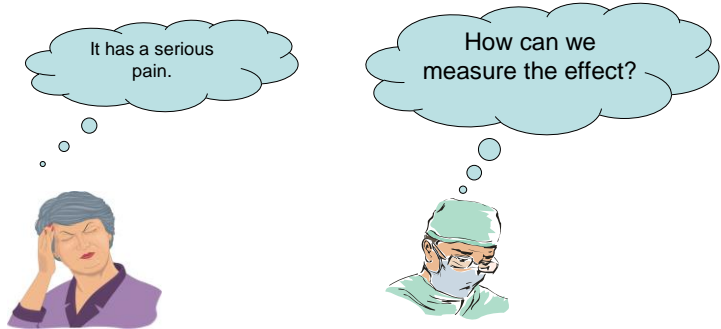
Correction:

Correction of the degree of freedoms.

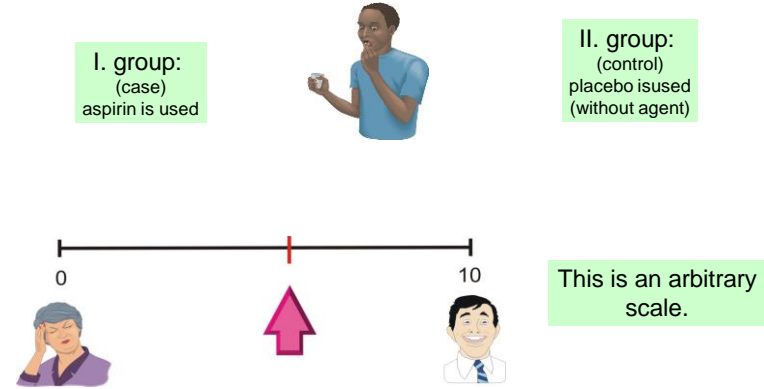
Correction of the t-values.

# Mann-Whitney U-test

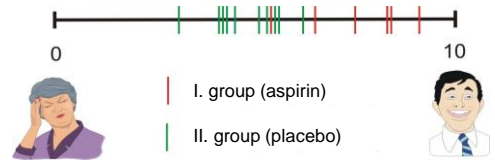
Example: Is the painkiller effective?



# Experiment

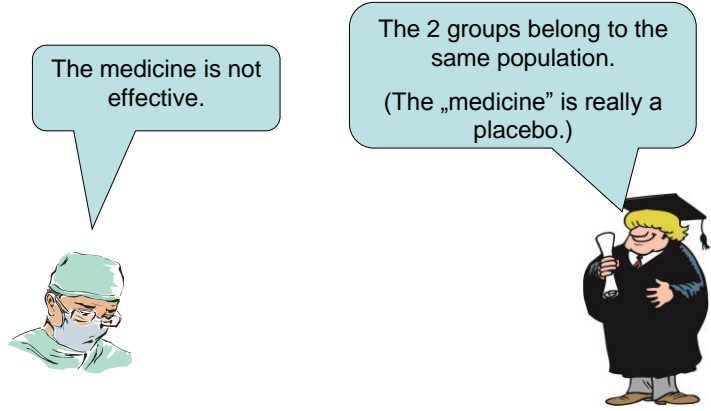


# Results



Value	3.1	4.1	4.2	4.3	4.5	5.1	5.3	5.4	5.5
Rank	1	2	3	4	5	6	7	8	9
Value	5.6	6.2	6.2	6.5	7.5	8.3	8.3	8.4	9.1
Rank	10	11.5	11.5	13	14	15.5	15.5	17	18

# The nullhypothesis



## The sum of the ranks (Gauss story)

Add the numbers from 1 to 100!

It is easy to calculate!

$$1 + 100 = 101$$

$$2 + 99 = 101 \dots$$

$$\sum_{i=1}^n i = \frac{n}{2} \cdot (n + 1)$$



## Sum of the ranks

$T$  – the sum of the ranks in the I. group, in the case of random deviation. The expected value is:

$$n_1 \cdot \frac{n_1 + n_2 + 1}{2}$$

( $n_1$  element, their average =  $(n_1 + n_2 + 1)/2$ )

Nullhypothesis: the deviation from this is random.

Small  $n$ : U-distribution describes the probability of the random deviation.

## The transformation (if $n$ is enough large)

$T$  – the sum of the ranks in the I. group. The expected value in the case of random distribution is:

$z$  has standard normal distribution.

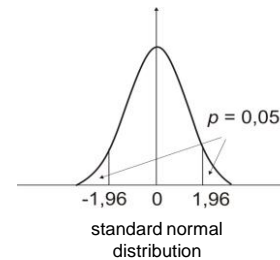
$$z = \frac{T - n_1(n_1 + n_2 + 1)/2}{s}$$

$$n_1 \cdot \frac{n_1 + n_2 + 1}{2}$$

$$s = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$



## Decision



The calculated z-value: 3.24.

Higher than 1.96.

**Conclusion:** we reject the nullhypothesis.

Calculated p-value < 0.1%.

The conclusion is same.

