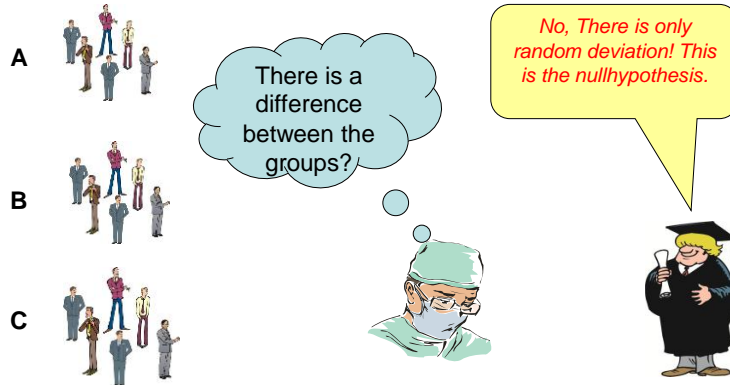
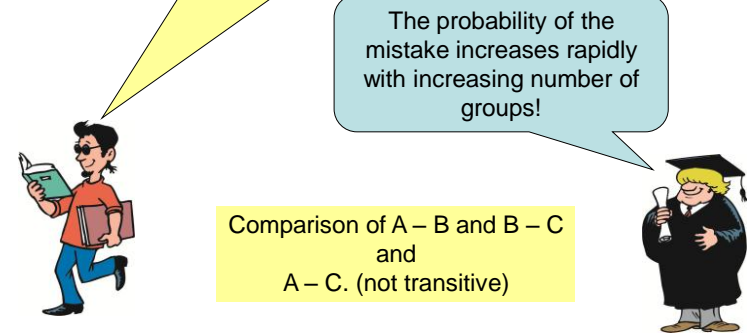


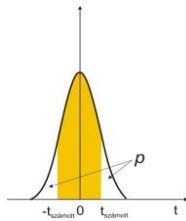
## Analysis of variancia (ANOVA)



## Why don't we do 2-sample t-tests?



## How much is the chance of the mistake?



Suppose, that we reject the nullhypothesis in all cases.

$p$  – probability being outside  
( $1-p$ ) – probability being inside randomly.



Question: How much is the probability to have mistake at least one case?

## How much is the chance that at least one is outside randomly?



1 test:  $p$  (let it be 5%).  
In the case of more than 1 test, the binomial distribution may be used to calculate.

$$1 - (1 - p)^3$$

In the case of 3 groups is about 15%!!!



## More than 2 groups

1. group



We can handle elements separated in groups or together.

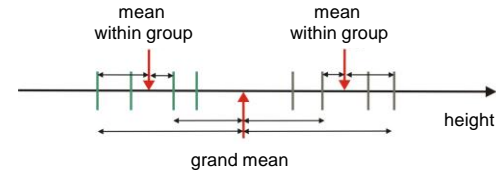


2. group



**mean within group:** calculated from the elements of the given group.  
**grand mean:** calculated from all elements.

## Components of the variance



Remember:

The variance is proportional to the squared sum of the differences from the average!

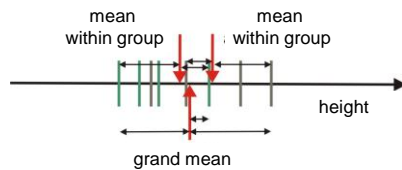


If the groups differ from each other significantly, the average squared differences from the grand mean are significantly higher than the differences in the groups!

## The base of ANOVA

In this case, the difference not so significant!

The total variance is the sum of the variance within groups and the variance between the groups!



## Components of the variance

	1. group	2. group	3. group
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172
mean	171.25	166	172.8

grand mean = 170.58

$$(170 - 170.58) = (170 - 166) + (166 - 170.58)$$

$$(175 - 170.58) = (175 - 172.8) + (172.8 - 170.58)$$

$$(x_{i,j} - \bar{x}) = (x_{i,j} - \bar{x}_j) + (\bar{x}_j - \bar{x})$$

between groups

within group (e.g. random) deviation

$\bar{x}$  - grand mean

$\bar{x}_j$  - mean in the jth group

Take the square!

$$(x - x_{ij})^2 = (x - x_i)^2 + (x_i - x_{ij})^2 + 2(x - x_i)(x_i - x_{ij})$$

$$\sum (x - x_{ij})^2 = \sum (x - x_i)^2 + \sum (x_i - x_{ij})^2 + \sum 2(x - x_i)(x_i - x_{ij})$$

covariance = 0 (independent events)

$$\sum 2(x - x_i)(x_i - x_{ij}) = 0$$

$$\sum (x - x_{ij})^2 = \sum (x - x_i)^2 + \sum (x_i - x_{ij})^2$$

between groups      within groups

(not necessary to know!)

We can decompose the variance!

## Calculation of variances

Variance	Sum of squares	d.f.		F value
between groups	$SS_A = \sum_j n_j (\bar{x}_j - \bar{x})^2$	k-1	$MS_A = \frac{SS_A}{k-1}$	$F = \frac{MS_A}{MS_E}$
Within groups	$SS_E = SS_T - SS_A$	N-k	$MS_E = \frac{SS_E}{N-k}$	
Total	$SS_T = \sum_{i,j} (x_{i,j} - \bar{x})^2$	N-1		

$\bar{x}$  grand mean

$\bar{x}_j$  mean in the jth group

N – no. of all elements  
k – no. of groups

## Nullhypothesis

There is no difference between the groups.

The difference between the averages of the groups due to the random deviation.



Decision: On the base of the comparison of the variance between groups and within groups!



## How can we compare them?

Comparison of variances?  
It was already discussed!

Really,  
in the case of 2-sample t-test.

$$F = \frac{MS_A}{MS_E}$$



## 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 large ( $p > \alpha$ ) – we **accept** the nullhypothesis.

(After decision, if it is necessary,  
we can do t-tests.)

## Example

Groups	n. of elements	sum	mean	Variance		
1	4	685	171.25	10.92		
2	3	498	166	13		
3	5	864	172.8	2.7		
ANOVA						
Factors	SS	df	MS	F	p-value	F crit.
Between groups	89.367	2	44.68	5.782	0.02427	4.257
Within group	69.55	9	7.728			
Total	158.92	11				

$\alpha = 0.05$   
 $p = 0.024$

**Decision:**  
we reject the nullhypothesis, there is a  
significant difference.

## Conditions for ANOVA

- Task: comparison 3 or more independent groups.
- The variable has **normal distribution**.
- sd values are the same in the groups.

## Kruskal-Wallis test

Rank data without separating into groups, than sum  
ranks in each group!



If the variable has no  
normal distribution!

## Ranking

	1. group	2. group	3. group
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172

value	163	165	168	169	170	171	172	172	173	174	175	175
rank	1	2	3	4	5	6	7.5	7.5	9	10	11.5	11.5

group	n. of elements	sum of the ranks
1	4	27.5
2	3	8
3	5	42.5

## Nullhypothesis

There is no difference between the groups.

The difference between the averages of the groups due to the random deviation.



## What kind of the distribution is suitable?

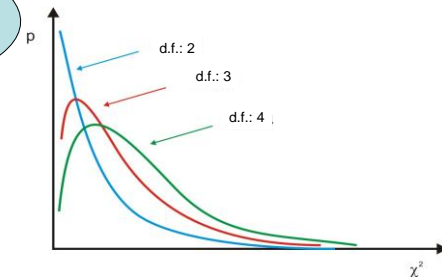
$$H = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

$N$  – no of elements  
 $R_i$  – the sum of the ranks in the  $i$ -th group  
 $n_i$  – no. of elements in the  $i$ -th group

## $\chi^2$ -distribution

The value of  $H \geq 0$ !

$H$  variable has  $\chi^2$ -distribution!



## 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 large ( $p > \alpha$ ) – we **accept** the nullhypothesis.

## Example

group	$n_i$	Sum of the ranks ( $R_i$ )
1	4	27.5
2	3	8
3	5	42.5

$$N = 12$$

$$H = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1)$$

$$4.97 = \frac{12}{12(12+1)} \left( \frac{27.5^2}{4} + \frac{8^2}{3} + \frac{42.5^2}{5} \right) - 3(12+1)$$

$$df. = 3 - 1 = 2$$

$$\alpha = 0.05$$

$$p = 0.083$$

**Decision:**

we accept the nullhypothesis, there is no significant difference.

## Compare them!

	1. group	2. group	3. group
1	173	170	175
2	175	163	174
3	169	165	171
4	168		172
5			172

ANOVA

$$\alpha = 0.05$$

$$p = 0.024$$

Kruskal-Wallis test

$$\alpha = 0.05$$

$$p = 0.083$$

**Decision:**

we reject the nullhypothesis.



**Decision:**

we accept the nullhypothesis.

Hypothesis test?



- Set up the **nullhypothesis!**
- Look for a **variable with known distribution.**
- Calculate the **probability of the random deviation** on the base of the distribution.
- If it is smaller than the significance level **reject**, in opposite case **accept the nullhypothesis.**
- That's all!

