Requirement and related topics

The basics of statistics and especially statistical distributions are advantageous for these descriptions. Further topics are:

www.weibull.de/COM/Data_Analysis.pdf

Introduction

The Analysis Of Variance (ANOVA for short) is about determining the variance of groups (factors) against the unexplained variance (residual variance) and „confirming“ or rejecting a significant influence.
Historically, the ANOVA was the evaluation tool for Design of Experiment (DoE). Alternatively, regression methods can usually do more.

Purpose and usefulness

The ANOVA is used to determine whether the factors in relation to the scatter have a significant effect on the response.

Basics

The known methods of ANOVA are diverse. Only the most important procedures are described in this documentation:
In general, a so-called dispersion decomposition is carried out in an ANOVA in order to differentiate systematic influences of factors from a random dispersion. The general model is:

Total deviation = Factors dev. + Error deviation

$SS_{Total} = SS_{Factors} + SS_{Error}$

$$\sum_{j=1}^{z}\sum_{i=1}^{n}\left(y_{ji}-\bar{y}\right)^2 = n\sum_{j=1}^{z}\left(\bar{y}_j-\bar{y}\right)^2 + \sum_{j=1}^{z}\sum_{i=1}^{n}\left(y_{ji}-\bar{y}_j\right)^2$$

SS	Sum of Squares
y_{ji}	data point from col j and row i
$\bar{y}_j$	mean of col j
$\bar{y}$	mean over all

Measurem.	Factors 1	2	3	...	z
1	y_{11}	y_{21}	y_{31}	⋯	y_{z1}
2	y_{12}	y_{22}	y_{32}	⋯	y_{z2}
3	y_{13}	y_{23}	y_{33}	⋯	y_{z3}
...	...	...	...	...	...
n	y_{1n}	y_{2n}	y_{3n}	⋯	y_{zn}

The variance = Mean Squares MS is the sum of squares based on the degrees of freedom. Here is z = number of factors and n = number of observations:

$$MS_{Total} = \left(\frac{SS_{Total}}{z\,n-1}\right) \quad MS_{Factors} = \left(\frac{SS_{Factors}}{z-1}\right) \quad MS_{Error} = \left(\frac{SS_{Error}}{z\,(n-1)}\right)$$

For a significant test now, the $MS_{Factors}$ are divided through MS_{Error} and it is:

$$F = \frac{MS_{Factors}}{MS_{Error}}$$

The bigger the F-value, the higher the probability of the factor effect. The null hypothesis Ho is: The means of the factors do not differ from one another. Ho is rejected if the probability from the F-distribution with degrees of freedom $f1 = z\text{-}1$ and $f2 = z\,(n\text{-}1)$ is less than the significance level α.

The so-called coefficient of determination R^2 describes how much the effect of the factors is in the model. The maximum is R^2=1. The bigger the scatter the smaller is the R^2.

$$R^2 = 1 - \frac{SS_{Error}}{SS_{Total}}$$

Balanced One-Way ANOVA (μ1 = μ2 = μ3…)

The null hypothesis is to be tested for several data columns with the same size

μ1 = μ2 = μ2 = …

The prerequisite for this test is that the data series are normally distributed. The variances must be the same, which can be checked using the F-test. Alternatively, the t-test is possible, in which different variances are possible. The data series must be independent of each other. For the following example, the null hypothesis is to be tested that all mean values are equal.

For the calculation, the sum of squares SS and the degrees of freedom = Degrees of Freedom = DF are determined as follows:

z →

	A	B	C	
n ↓	1,0	4,0	5,5	
	1,5	5,5	6,5	
	2,5	6,0	8,0	
	4,0	7,0	9,0	
	5,0	9,0	9,5	
$\bar{y}$	2,8	6,3	7,7	$\bar{\bar{y}} = 5{,}6$
$\bar{y} - \bar{\bar{y}}$	-2,8	0,7	2,1	
$(\bar{y} - \bar{\bar{y}})^2$	7,84	0,49	4,41	12,74

$$SS_{Total} = \sum_{j=1}^{z}\sum_{i=1}^{n}\left(y_{j,i} - \bar{\bar{y}}\right)^2 = 100{,}1$$

$$SS_{Factors} = n \cdot (\bar{y} - \bar{\bar{y}})^2 = 5 \cdot 12{,}74 = 63{,}7$$

$$SS_{Error} = SS_{Total} - SS_{Factors} = 36{,}4$$

$$DF_{Total} = n \cdot z - 1 = 14$$

$$DF_{Factors} = z - 1 = 2$$

$$DF_{Error} = DF_{Total} - DF_{Factors} = 12$$

Table of results:

	DF	SS	MS	F	p-val
Factors	2	63,7	31,85	10,50	0,0023
Error	12	36,4	3,03		
Total	14	100,1			

The p-value is calculated using the Fisher distribution

$f1 = DF_{Faktors}$; $f2 = DF_{Error}$

$p\text{-}value = 1 - Fisher(F; f1; f2) = 1 - Fisher(10{,}5; 2; 12) = 0{,}0023$

Because the p-value falls below the specified significance level of α = 0.05, the null hypothesis that the means are equal is rejected.

Two-Way ANOVA balanced

In contrast to the one-way ANOVA, in the two-way there is a target value on which the factors act. The aim here is to determine a relationship between the factors and the target variable. The factors must have the same number of observations (balanced), be independent of one another, have comparable scatter and be normally distributed. The scatter decomposition is here:

$$SS_{abs} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \qquad SS_{Total} = \sum_{i=1}^{n} y_i^2 - SS_{abs}$$

$$SS_A = \frac{1}{bk}\sum_{i=1}^{a} \bar{y}_i^2 - SS_{abs}$$

$$SS_B = \frac{1}{ak}\sum_{j=1}^{b} \bar{y}_j^2 - SS_{abs}$$

$$SS_{AB} = \frac{1}{k}\sum_{i=1}^{a}\sum_{j=1}^{b} \bar{y}_{ji}^2 - SS_A - SS_B - SS_{abs}$$

$$SS_{Error} = SS_{tot} - SS_A - SS_B - SS_{AB}$$

n : total size of data

a : number of levels factor A

b : number of levels factor B

k : number of repetitions

$\bar{y}_i$: mean oft hte i-th factor level from factor A

$\bar{y}_j$: mean oft hte i-th factor level from factor B

The results for a two-factor example with the influences of an additive and the temperature on a process (target variable) are generally output in the tabular form shown. The F value is the ratio of the variance (mean square) of the factors and the interaction to the variance of the spread (error). From this, the probability of error (p-value) is determined via the F-distribution:

	DF	SS	MS	F	p-val
Additive	3	2,608E+02	8,692E+01	4,99	0,008
Temperature	2	8,029E+02	4,014E+02	23,05	0,002
Additive*Temperature	6	3,340E+02	5,567E+01	3,20	0,019
Error	24	4,180E+02	1,742E+01		
Total	35	1,816E+03			

Two-Way ANOVA balanced with random factors

Random factors have levels chosen at random, while fixed factors have levels fixed by e.g. a DoE. The following example resulted in temperatures that were not systematically specified.

Instead of relating the variance *MS* of the additive to the variance of the error *MSError*, here we refer to the variance of the interaction.

	DF	SS	MS	F	p-val	Type
Additive	3	2,61E+02	8,69E+01	1,56	0,294	fix
Temperature	2	8,03E+02	4,01E+02	10,5	0,017	random
Additive*Temperature	6	3,34E+02	5,57E+01	3,2	0,019	
Error	24	4,18E+02	1,74E+01			
Total	35	1,82E+03				

This method is used in the measurement system analysis with ANOVA according to VDA Volume 5. Here, the parts used for the repeat measurement and the testers are random and not the same as, for example, later for the determination of a process capability.

Two-Way ANOVA nested

In a so-called nested ANOVA, there is one factor that cannot be freely combined. All factors in the model must be random factors. In this example, the temperature is generated by different heating processes in an oven. Each temperature level is therefore nested in the additives. Instead of relating the variance *MS* of the additive to *MSError*, here it is related to the second nested factor, temperature.

	DF	SS	MS	F	p-val
Additive	2	8,03E+02	4,01E+02	6,075	0,022
Temperature	9	5,95E+02	6,61E+01	3,794	0,004
Error	24	4,18E+02	1,74E+01		
Total	35	1,82E+03			

Finally, the last factor is related to *MSError*.

A nested ANOVA is used in particular in the measurement system analysis when the parts to be measured repeatedly must always be different due to destructive tests.

Model ANOVA

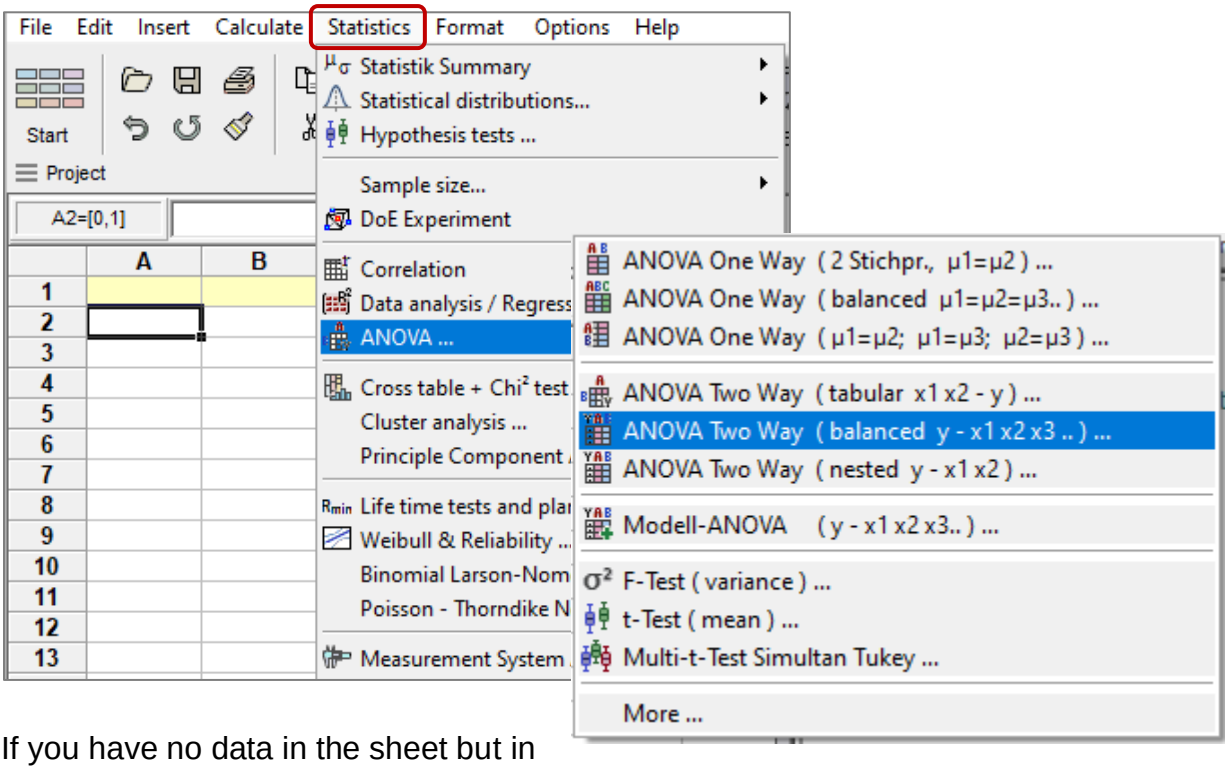
In the so-called model ANOVA, the sum of squares are related to a function from a regression model. Detailed information on this is described under:

www.weibull.de/COM/Data_Analysis.pdf

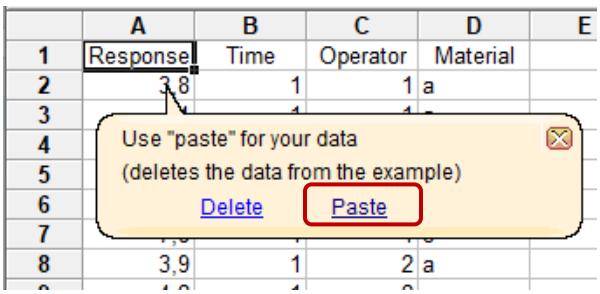
Using Visual-XSel

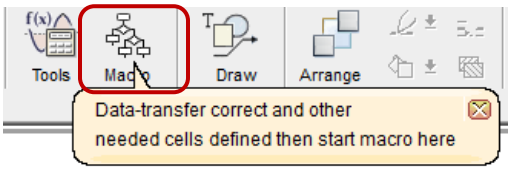
www.crgraph.com

Except for the model ANOVA, all methods are available as templates. These are in the directory ..\Templates\ 03_Data_Analysis\. Direct access is possible via the menu *Statistics*:



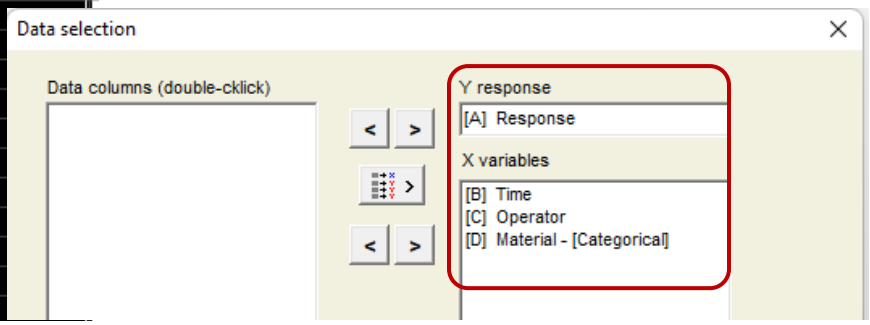
If you have no data in the sheet but in the clipboard, use the speech bubble to paste the data in the template and start the macro





If you have first the data in the sheet, mark the needed columns and use the menu shown above and define the Y response and the X variables

	A	B	C	D
1	Response	Time	Operator	Material
2	3,8	1	1	a
3	4	1	1	a
4	6,3	1	1	b
5	5,9	1	1	b
6	7,6	1	1	c
7	7,8	1	1	c
8	3,9	1	2	a
9	4,2	1	2	a
10	7,2	1	2	b
11	7	1	2	b
12	9,5	1	2	c
13	9,6	1	2	c
14	4,5	1	3	a
15	4	1	3	a



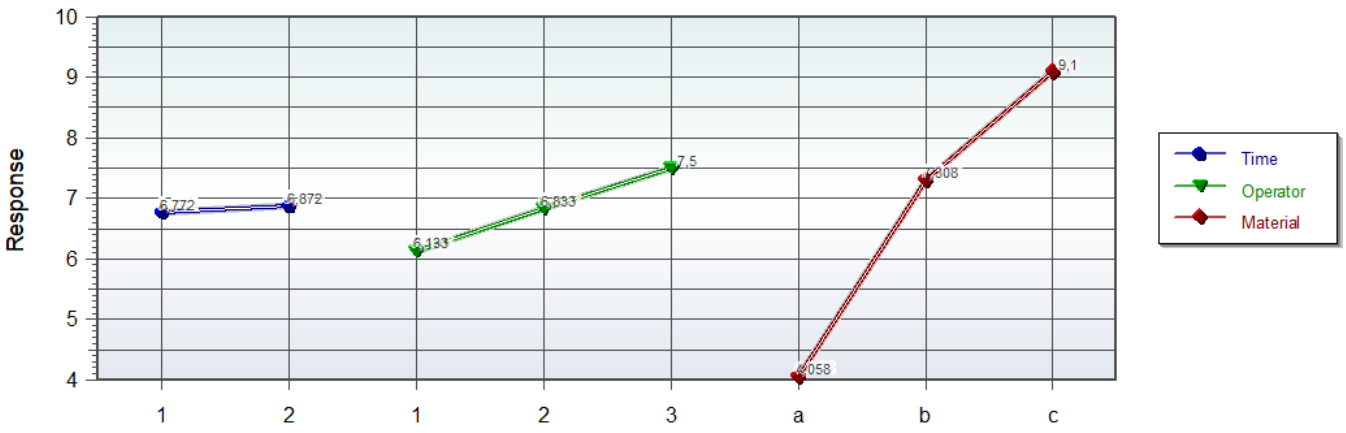
In the following example, the relationship between time, operator and material is to be examined for a strength (target variable) using an ANOVA. Given are the following data series, which are to be marked first. Since there is a target variable and the data is in columns, the ANOVA Two Way balanced should be selected.

After starting the macro, a query follows as to whether to create the model with interaction and what are random factors. Since the operators are not the same people as in the later process, the second parameter is to be declared as random.

The result is displayed in the main window of the embedded template

	DF	SS	MS	F	p-val	Typ	Steps			
Time	1	9,000E-02	9,000E-02	1,26	0,273	fix	2	1	2	
Operator	2	1,121E+01	5,604E+00	78,53	0,000	fix	3	1	2	3
Material	2	1,568E+02	7,838E+01	1098,35	0,000	fix	3	a	b	c
Time*Operator	2	6,200E-01	3,100E-01	4,34	0,026					
Time*Material	2	1,145E+00	5,725E-01	8,02	0,002					
Operator*Material	4	4,284E+00	1,071E+00	15,01	0,001					
Error	22	1,570E+00	7,136E-02							
Total	35	1,757E+02								

S =	2,671E-01
R² =	0,991
R²adj =	0,986



To get back to the original starting table, select the main project under the Project menu item, or close the template on the top right: