ANOVA

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 devided 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 smaler is the R^2.

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

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

Für mehrere Datenspalten mit gleichem Umfang ist die Nullhypothese

µ1 = µ2 = µ2 = …

zu testen. Die Voraussetzung für diesen Test ist, dass die Datenreihen normalverteilt sind. Die Varianzen müssen gleich sein, was über den F-Test geprüft werden kann. Alternativ ist der t-Test möglich, bei dem unterschiedliche Varianzen möglich sind. Die Datenreihen müssen voneinander unabhängig sein. Für folgendes Beispiel soll die Nullhypothese geprüft werden, dass alle Mittelwerte gleich sind.

Für die Berechnung werden die *Sum of Squares* kurz SS und die Freiheitsgrade = Degrees of Freedom = DF wie folgt bestimmt:

z →

↓	A	B	C	
	1,0	4,0	5,5	
	1,5	5,5	6,5	
	2,5	6,0	8,0	
	4,0	7,0	9,0	
n	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	$\boxed{31{,}85}$	$\boxed{10{,}50}$	0,0023
Error	12	36,4	$\boxed{3{,}03}$		
Total	14	100,1			

Der *p-value* errechnet sich über die Fisher-Verteilung mit $f1 = DF_{Faktors}$; $f2 = DF_{Error}$

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

Da der *p-value* das festgelegte Signifikanzniveau von α = 0,05 unterschreitet, wird die Nullhypothese, dass die Mittelwerte gleich sind, verworfen.

Two-Way ANOVA balanciert

Im Gegensatz zur One-Way ANOVA gibt es in der Two-Way eine Zielgröße auf die Faktoren wirken. Das Ziel ist es hier einen Zusammenhang zwischen den Faktoren und der Zielgröße zu bestimmen. Die Faktoren müssen gleich viele Beobachtungen aufweisen (balanciert) unabhängig voneinander sein, vergleichbare Streuungen haben, sowie normalverteilt sein.

Die Streuungszerlegung ist hier:

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

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

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

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

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

n : Gesamtanzahl

a : Anzahl Stufen Faktor A

b : Anzahl Stufen Faktor B

k : Anzahl Wiederholungen

$\bar{y}_i$: Mittelwert der i-ten Faktorstufe von Faktor A

$\bar{y}_j$: Mittelwert der j-ten Faktorstufe von Faktor B

Die Ergebnisse für ein zweifaktorielles Beispiel mit den Einflüssen eines Zusatzstoffe und der Temperatur auf einen Prozess (Zielgröße) werden generell in der gezeigten tabellarischen Form ausgegeben. Der F-Wert ist das Verhältnis zwischen der Varianz (Mean Square) der Faktoren und der Wechselwirkung zur Varianz der Streuung (Error). Hieraus wird die Irrtumswahrscheinlichkeit (p-value) über die F-Verteilung bestimmt:

	DF	SS	MS	F	p-val
Zusatzstoff	3	2,608E+02	8,692E+01	4,99	0,008
Temperatur	2	8,029E+02	4,014E+02	23,05	0,002
Zusatzstoff*Temperatur	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 balanciert mit Zufallsfaktoren (Random)

Zufallsfaktoren haben zufällig ausgewählte Stufen, während die Stufen von festen Faktoren z.B. durch eine DoE festgelegt wurden. Im folgenden Beispiel ergaben sich Temperaturen, die nicht systematisch vorgegeben wurden.

ANOVA

Anstelle die Varianz MS des Zusatzstoffes auf die Varianz des Fehlers MS_{Error} zu beziehen, wird hier auf die Varianz der Wechselwirkung bezogen.

	DF	SS	MS	F	p-val	Typ
Zusatzstoff	3	2,61E+02	8,69E+01	1,56	0,294	fest
Temperatur	2	8,03E+02	4,01E+02	10,5	0,017	zufällig
Zusatzstoff*Temperatur	6	3,34E+02	5,57E+01	3,2	0,019	
Error	24	4,18E+02	1,74E+01			
Total	35	1,82E+03				

Dieses Verfahren wird bei der Mess-System-Analyse mit ANOVA nach VDA Band 5 verwendet. Hier sind verwendete Teile für die Wiederholmessung und die Prüfer zufällig und nicht dieselben, wie z.B. später für die Bestimmung einer Prozessfähigkeit.

Two-Way ANOVA geschachtelt (nested)

In einer sogenannten geschachtelten ANOVA gibt es einen nicht frei kombinierbaren Faktor. Alle Faktoren im Modell müssen Zufallsfaktoren sein. In diesem Beispiel wird die Temperatur durch unterschiedliche Aufheizvorgänge in einem Ofen erzeugt. Jede Temperaturstufe ist also in den Zusatzstoffen geschachtelt. Anstelle die Varianz MS des Zusatzstoffes auf MS_{Error} zu beziehen, wird hier auf den zweiten geschachtelten Faktor Temperatur bezogen.

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

Der letzte Faktor wird schließlich auf MS_{Error} bezogen.

Eine geschachtelte ANOVA wird insbesondere bei der Mess-System-Analyse angewendet, wenn die eigentlich wiederholend zu messenden Teile aufgrund von zerstörenden Prüfungen immer andere sein müssen.

Auf weitere Beschreibungen sei auf das Buch der Versuchsplanung von Prof. Kleppmann verwiesen.

Modell ANOVA

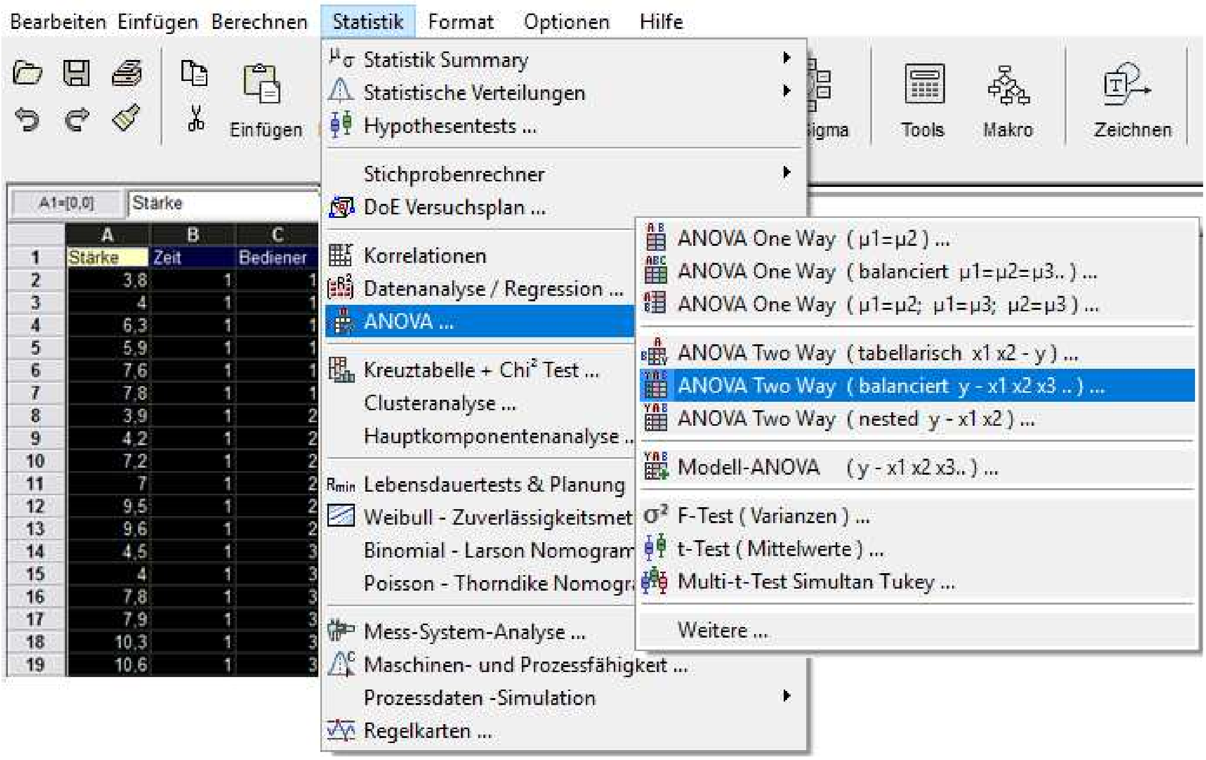
Bei der sogenannten Modell-ANOVA werden die Sum of Squares auf eine Funktion aus einem Regressionsmodell bezogen. Ausführliche Informationen sind hierzu beschrieben unter: www.versuchsmethoden.de/Multiple_Regression.pdf

ANOVA

Anwendung in Visual-XSel 15.0/16.0

www.crgraph.de

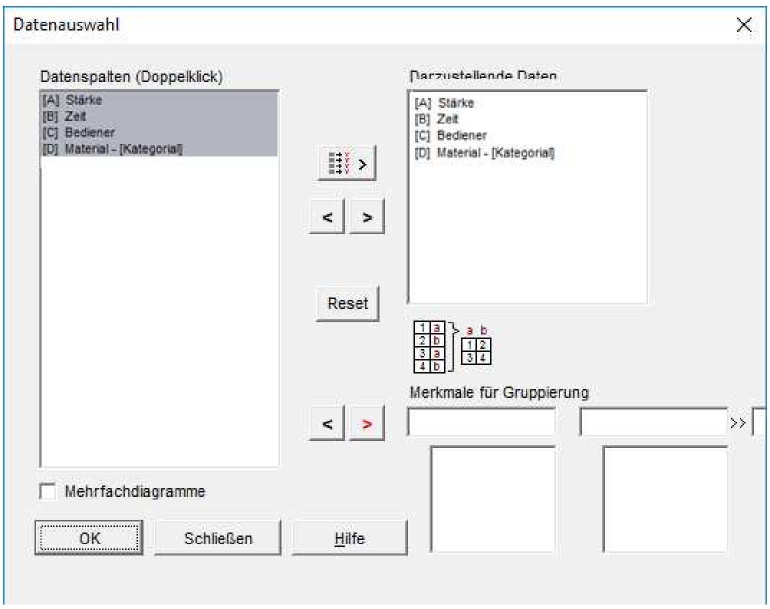
Bis auf die Modell-ANOVA sind alle Verfahren als Templates verfügbar. Diese befinden sich im Verzeichnis ..\Templates\03_Datenauswertung\..



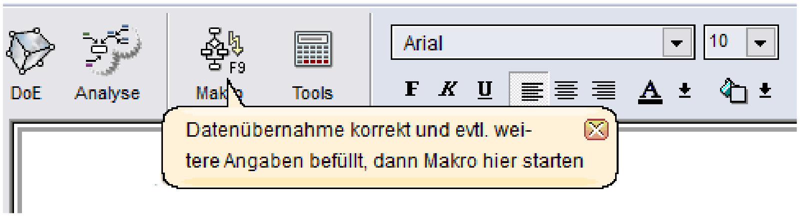
Im folgenden Beispiel soll über eine ANOVA der Zusammenhang von Zeit, Bediener und Material auf eine Stärke (Zielgröße) untersucht werden. Gegebene sind folgende Datenreihen, die zunächst zu markieren sind. Da es eine Zielgröße gibt und die Daten in Spalten vorliegen, ist die ANOVA Two Way balanciert auszuwählen:

Im darauffolgenden Dialog kann nochmal überprüft werden, ob die richtigen Datenspalten verwendet wurden (*Darzustellende Daten*).

Nachdem das Template geladen wurde, ist das Makro über die entsprechende Ikone, oder über F9 zu starten:



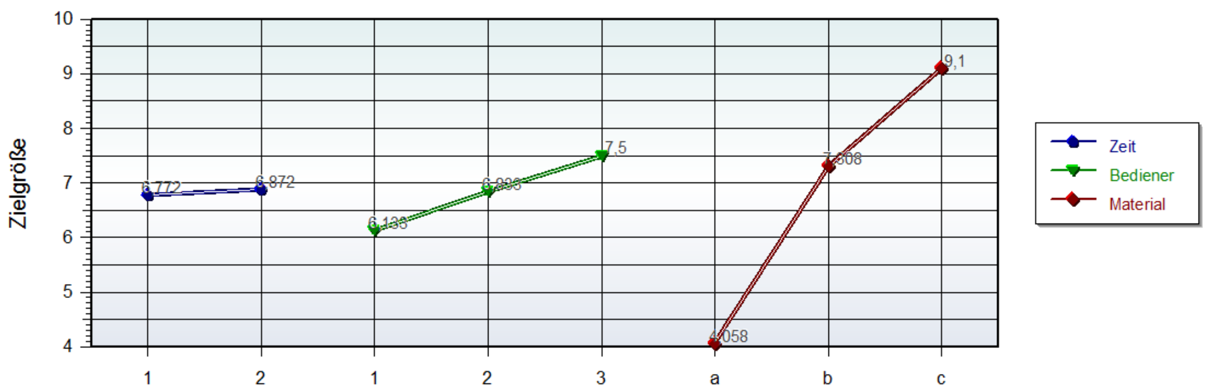
ANOVA



Es folgt eine Abfrage, ob das Modell mit Wechselwirkung erstellt werden soll und welche zufällige Faktoren sind. Da die Bediener nicht die selben Personen sind, wie bei dem späteren Prozess, ist hier der zweite Parameter als zufällig zu deklarieren.

Das Ergebnise wird im Hauptfenster des eingebetteten Templates ausgegeben:

	DF	SS	MS	F	p-val	Typ	Stufen					
Zeit	1	9,000E-02	9,000E-02	0,29	0,643	fest	2	1	2			
Bediener	2	1,121E+01	5,604E+00	4,28	0,083	zufällig	3	1	2	3		
Material	2	1,568E+02	7,838E+01	73,18	0,001	fest	3	a	b	c		
Zeit*Bediener	2	6,200E-01	3,100E-01	4,34	0,026							
Zeit*Material	2	1,145E+00	5,725E-01	8,02	0,002							
Bediener*Material	4	4,284E+00	1,071E+00	15,01	0,001						S =	2,671E-01
Error	22	1,570E+00	7,136E-02								R² =	0,991
Total	35	1,757E+02									R²adj =	0,986



Um wieder in die ursprüngliche Ausgangstabelle zu gelangen, ist unter dem Menüpunkt *Projekt* das *Hauptprojekt* auszuwählen (V15), oder das Template rechts außen zu schließen. Ab V16 ist der Menüpunkt Projekt links außen.