Varianzanalysen

ANOVA zwischen zwei Datenreihen

In der Varianz-Analyse (Analysis Of Variance) zwischen zwei Versuchsreihen, soll ein signifikanter Unterschied ermittelt werden. In einem Beispiel soll bestimmt werden, ob die Köpergrößen zwischen Europäern und Afrikanern unterschiedlich sind. Folgende Daten liegen vor:

Europäer	Afrikaner
159	187
163	173
156	177
173	181
161	
169	

Zunächst bildet man die Quadratsumme der Abweichungen für den Mittelwert, der den sogenannten Korrekturfaktor entspricht.

$$SQA_m = \frac{\left(\sum \mathrm{Re}\,ihe1 + \sum \mathrm{Re}\,ihe2\right)^2}{n_{ges}} = CF$$

mit n_{ges} = Anzahl der Messungen Reihe1 und Reihe 2
Danach wird die Gesamtsumme der quadrierten Abweichungen ermittelt:

$$GSQ = \sum \mathrm{Re}\,ihe1_i^2 + \sum \mathrm{Re}\,ihe2_i^2 - CF$$

mit dem dazugehörigen Freiheitsgrad $f_G = n_{ges}$-1
Weiterhin ist die Summe der quadrierten Abweichungen der einzelnen Datenreihen zu bilden

$$SQA = \frac{\left(\sum \mathrm{Re}\,ihe1\right)^2}{n_1} + \frac{\left(\sum \mathrm{Re}\,ihe2\right)^2}{n_2} - CF$$

mit dem Freiheitsgrad $f_A = 1$
Die Quadratsumme des Fehlers errechnet sich durch:

$$SQF = GSQ - SQA$$

mit dem Freiheitsgrad $f_F = n_{ges}$ -2
Die Varianzen bestimmen sich entsprechend:

$$V_A = \frac{SQA}{f_A} \qquad V_F = \frac{SQF}{f_F}$$

Der sogenannte F-Wert ist der Quotient dieser beiden Varianzen

$$F = \frac{V_A}{V_F}$$

der mit einem kritischen F-Wert F_{krit} auf einem festgelegten Signifikanzniveau, z.B. 95%, verglichen wird. Ist $F > F_{krit,\, fA,\, fF}$, so kann gesagt werden, dass die beiden Datenreihen signifikant unterschiedlich sind.

Der prozentuale Anteil am Gesamteffekt berechnet sich durch

$$SQA' = SQA - f_A \cdot V_F$$

$$A = \frac{SQA'}{GSQ} 100\%$$

und beschreibt die Mittelwerteffekte. Die Differenz zu 100% entspricht den Fehleranteil

Das Verfahren kann mit der Vorlagendatei *ANOVA_Zwei_Datenreihen.vxg* im Verzeichnis */Statistik* angewendet werden.

ANOVA & ANOM mit mehreren Faktoren

In der Varianz-Analyse mit mehreren Faktoren werden die Einflüsse von Versuchsparametern auf eine Zielgröße untersucht.
Es soll herausgefunden werden, welchen Einfluss die Parameter im Verhältnis der Streuungen auf das Versuchsergebnis haben.
Nach Varianzanalyse wird durch den statistischen F-Test ausgegeben, ob der Parameter einen signifikanten Einfluss hat und wie groß sein prozentualer Anteil gegenüber der Reststreuung ist. Es wird vorausgesetzt, dass die Abweichungen normalverteilt sind. Andernfalls ist das Ergebnis nicht eindeutig.

Im folgendem Beispiel sind die dargestellten Versuche durchgeführt worden.

	Zielgröße	Temperatur	Druck	Einstellzeit	Reinigung
1	-20	1	1	1	1
2	-10	1	2	2	2
3	-30	1	3	3	3
4	-25	2	1	2	3
5	-45	2	2	3	1
6	-65	2	3	1	2
7	-45	3	1	3	2
8	-65	3	2	1	3
9	-70	3	3	2	1

Die einzelnen Schritte der ANOVA:

1) Bildung der Quadratsumme der Abweichungen für den Mittelwert, der auch Korrekturfaktor bezeichnet wird

$$SQM = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2 = CF$$

2) Bildung der Quadratsumme der Abweichungen des Gesamtmittelwertes

$$GSQ = \sum_{i=1}^{n} Y_i^2 - CF$$

3) Bildung der Summe der quadrierten Abweichungen bezüglich der Faktoren

$$SQA_A = \frac{1}{n_{A1}}\left(\sum Y_{A1}\right)^2 + \frac{1}{n_{A2}}\left(\sum Y_{A2}\right)^2 + \frac{1}{n_{A2}}\left(\sum Y_{A2}\right)^2 - CF$$

wobei n_{A1}, n_{A2} und n_{A3} jeweils die Anzahl der Punkte gleicher Einstellungen für A ist und A steht in unserem Beispiel für die Temperatur. Für B (Druck) gilt analog:

$$SQA_B = \frac{1}{n_{B1}}\left(\sum Y_{B1}\right)^2 + \frac{1}{n_{B2}}\left(\sum Y_{B2}\right)^2 + \frac{1}{n_{B2}}\left(\sum Y_{B2}\right)^2 - CF$$

$$SQA_C = \frac{1}{n_{C1}}\left(\sum Y_{C1}\right)^2 + \ldots\ldots..$$

usw.

4) Abschätzung der Varianzen der einzelnen Faktoren als Quotient aus der quadrierten Abweichung zu dem Freiheitsgrad

$$V_A = \frac{SQA_A}{FHG_A} \qquad V_B = \frac{SQA_B}{FHG_B} \qquad V_C = \ldots. \qquad \ldots. \qquad \ldots..$$

wobei FHG = Anzahl Stufen –1 ist (Anzahl unabhängiger Einstellungen, die man von einer Stufe ausgehend noch verändern kann, im Beispiel ist FHG_A=2).

5) Bestimmung der Fehlervarianz
Bei der Auswertung von Versuchsplänen können im allgemeinen zwei Arten von Fehlern auftreten:

F1 = Fehler innerhalb einer Merkmalskombination, wobei dieser 0 sein sollte bei entsprechender Sorgfalt der Durchführung.
F2 = Fehler in der Wiederholung von Messungen

Die Varianz des Fehlers F2 kann abgeschätzt werden nach folgender Regel: Man zieht die Quadratsumme der Faktoren mit den geringsten quadrierten Abweichungen zusammen, in unserem Beispiel SQA_{C+D} = 400. Es sollte etwa die halbe Anzahl FHG's verwendet werden. Somit ergibt sich die Fehlervarianz durch

$$V_{F2} = \frac{SQA_{C+D}}{FHG_{C+D}}$$

6) Berechnung der Verhältnisse der Faktorvarianzen zur Fehlervarianz

$$F_A = \frac{V_A}{V_{F2}} \qquad F_B = \frac{V_B}{V_{F2}} \qquad F_C = ...$$

7) Bestimmung der Signifikanz der entsprechenden Faktoren

Der vorher ermittelte F-Wert kann mit einem kritischen F-Wert verglichen werden. Es wird die Nullhypothese aufgestellt $F_A > F_{krit}$, es gibt mit x % einen signifikanten Unterschied. Den kritischen F-Wert, z.B. für A erhält man aus F-Tabellen mit dem Freiheitsgrad $f_1 = FHG_A = 2$ und $f_2 = FHG_{C+D} = 4$ und einem Signifikanzniveau von 95%.

8) Prozentuale Bedeutung eines Faktors

Ein wichtiges Ergebnis der Varianzanalyse ist der prozentuale Anteil eines Faktors auf die Zielgröße. Dieser bestimmt sich z.B. für A durch:

$$SQA'_A = SQA_A - FHG_A \; V_{F2}$$

$$A_A = \frac{SQA'_A}{GSQ} 100\%$$

Der prozentuale Anteil von F2 bestimmt sich durch:

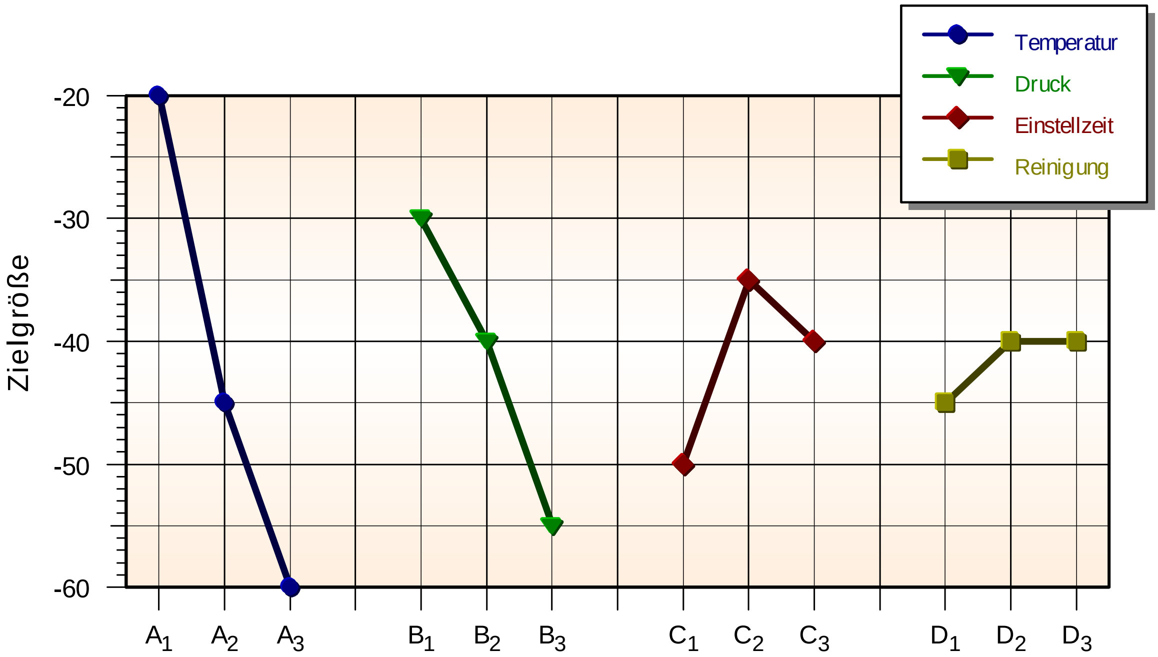
$$A_{F2} = \frac{SQA_{F2} - V_{F2}}{GSQ} 100\%$$

Für das Beispiel ergeben sich insgesamt folgende Ergebnisse:

	FHG	SQA	V	F	SQ'	Anteil %	F kritisch
Temperatur	2	2450	1225	12,25	2250	59,2	6,94
Druck	2	950	475	4,75	750	19,7	6,94
Einstellzeit	2	350	175	1,75	150	3,9	6,94
Reinigung	2	50	25	0,25			
Fehler 2	4	400	100			7,9	

Der entscheidende Einfluss ist somit die Temperatur.

In der sogenannten **ANOM** (Analysis Of Means) werden die mittleren Werte der Zielgrößen jeder Einstellung eines jeden Faktors dargestellt. Für das in der ANOVA beschriebene Beispiel ergibt sich folgende Darstellung:



Für die dargestellten Verfahren ist die Vorlage *ANOVA_Mehrfaktoriell.vxg* im Verzeichnis \Statistik anzuwenden.
Werden mehrere Messungen für jeweils eine Faktoreinstellung verwendet, so muss der Fehleranteil nicht aus den kleinsten Faktoranteilen geschätzt werden, sondern kann direkt bestimmt werden. Zunächst wird die Quadratsumme für den Fehler errechnet:

$$SQA_{F2} = GSQ - \sum_{i=1}^{p} SQA_i \quad \text{mit} \quad p = \text{Anzahl Faktoren}$$

und die Varianz ist:

$$V_{F2} = \frac{SQA_{F2}}{f_{F2}} \qquad f_{F2} = (ny \cdot n - 1) - (n - 1)$$

mit ny = Anzahl Wiederholungen, n = Anzahl Versuche

Die relativen Anteile bestimmen sich analog zur bisherigen Vorgehensweise über:

$$SQA_x^{'} = SQA_x - f_x \cdot V_{F2}$$

$$A_x = \frac{SQA'_x}{GSQ} 100\%$$

Für die ANOVA mit Wiederholungen ist die Vorlage
ANOVA_Mehrfaktoriell_Wiederholungen.vxg im Verzeichnis \Statistik zu verwenden.