

Voraussetzung und verwandte Themen

Für diese Beschreibungen sind Grundlagen der Statistik vorteilhaft. Weiterführende und verwandte Themen sind:

www.crgraph.de/literatur

www.versuchsmethoden.de/Hypothesentests.pdf

www.versuchsmethoden.de/Multiple_Regression.pdf

Stichworte: ANOVA - Varianzanalyse - Mittelwertvergleich - One-Way - Two Way - Balanciert

Einführung

In der Varianzanalyse (**A**nalysis of **V**ariance oder kurz ANOVA) geht es darum, die Varianzen von Gruppen (Faktoren) gegenüber der unerklärten Varianz (Reststreuung) zu bestimmen und einen signifikanten Einfluss zu bestätigen oder abzulehnen.

Historisch war die ANOVA das Auswertetool für Versuchspläne (DoE). Alternativ können Regressionsverfahren in der Regel mehr leisten.

Ziel und Nutzen

Mit der ANOVA soll die Hypothese auf gleiche Mittelwerte von Datenreihen, oder der Zusammenhang von Einflüssen (Faktoren) auf eine Zielgröße (z.B. Messwerte) bestimmt werden.

Grundlagen

Die bekannten Verfahren der ANOVA sind vielfältig. In diese Dokumentation werden nur die wichtigsten Verfahren beschrieben:

Allgemein wird in einer ANOVA eine sogenannte Streuungszerlegung durchgeführt, um systematische Einflüsse von Faktoren von einer zufälligen Streuung zu unterscheiden. Das allgemeine Modell ist:

Gesamtabweichung = erklärte Abweichung + unerklärte Abweichung

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

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

Summe der quadrierten Gesamtabweichung	Summe der quadrierten Abweichungen zwischen den Faktorstufen	Summe der quadrierten Abweichungen innerhalb der Faktorstufen
--	--	---

SS	Sum of Squares
y_{ji}	Datenpunkt von Spalte j und Zeile i
\bar{y}_j	Mittelwert der Spalte j
\bar{y}	Mittelwert über alle Daten

		Faktoren				
		1	2	3	...	z
Messung	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}

Die Varianzen = Mean Square MS ergeben sich mit Hilfe der Freiheitsgrade $z = \text{Anzahl Faktoren}$ und $n = \text{Anzahl Beobachtungen (Versuche)}$:

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

Die Werte $MS_{Factors}$ und MS_{Error} werden ins Verhältnis zueinander gesetzt. Diesen Ausdruck bildet die Teststatistik für die ANOVA, den Test auf gleiche Mittelwerte.

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

Je größer der F-Wert ist, desto höher ist die Wahrscheinlichkeit für den Einfluss des Faktors. Die Nullhypothese H_0 lautet: Die Mittelwerte der Faktoren unterscheiden sich nicht voneinander. H_0 wird abgelehnt, wenn die Wahrscheinlichkeit aus der F-Verteilung mit den Freiheitsgraden $f_1 = z-1$ und $f_2 = z (n-1)$ kleiner als das Signifikanzniveau α ist, siehe www.versuchsmethoden.de/Hypothesentests.

Das sogenannte Bestimmtheitsmaß R^2 beschreibt wie viel die Faktoren die Zusammenhänge beschreiben. Der maximal Wert ist $R^2=1$. Je größer die Streuung ist, desto geringer ist R^2 .

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

One-Way ANOVA balanciert ($\mu_1 = \mu_2 = \mu_3 \dots$)

Für mehrere Datenspalten mit gleichem Umfang ist die Nullhypothese

$$\mu_1 = \mu_2 = \mu_3 = \dots$$

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	
	5,0	9,0	9,5	
n				
\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

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

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

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

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

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

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

Ergebnistabelle:

	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			

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 $\alpha = 0,05$ unterschreitet, wird die Nullhypothese, dass die Mittelwerte gleich sind, verworfen.

Two-Way ANOVA balanciert

Im Gegensatz zur One-Way ANOVA gibt es in dem 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 Streuungszersetzung ist hier:

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

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

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

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

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

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 Zusatzstoffes 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.

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](http://www.versuchsmethoden.de/Multiple%20Regression.pdf)

Weiterführende Beschreibungen

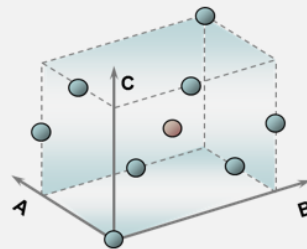
Ausführliche softwareunabhängige Beschreibungen zum Thema DoE und der dazugehörigen Auswertungen gibt es im

Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden

Definitive Screening Designs DSD

Sogenannte Definitive Screening Designs sind sehr neu von Jones und Nachtsheim entwickelte Versuchspläne mit sehr geringem Versuchsumfang.

Sie ermöglichen die Auswertung von quadratischen Modellen und basieren deshalb auf 3 Stufen. Zwischen den Hauptfaktoren untereinander und den quadratischen Termen gibt es keine Vermengung (orthogonal). Die Wechselwirkungen sind nicht zu 100% vermengt.



Nr	A	B	C	D
1	0	1	-1	-1
2	0	-1	1	1
3	-1	0	-1	1
4	1	0	1	-1
5	-1	-1	0	-1
6	1	1	0	1
7	-1	1	1	0
8	1	-1	-1	0
9	0	0	0	0

In der generischen Erzeugung dieser Versuchspläne (iterativ mit Hilfe der Determinante) ergibt sich regulär die Anzahl Versuche mit $n = 2^p + 2$. Manche Pläne, z.B. für $p=5$ sind dann allerdings teilweise zwischen den Hauptfaktoren vermengt. Hier müssen bis zu 3 Versuchszeilen ergänzt werden. Der Gesamtumfang ergibt sich somit zu:

$$n = 2^p + 2 + (1..3)$$

Alle Faktoren müssen durchgehend auf 3 Stufen sein und es lassen sich keine kategorialen Faktoren darstellen. Nachteilig ist auch, dass keine Auswertung aller möglichen



Weitere Informationen und Leseproben:

crgraph.de/Literatur

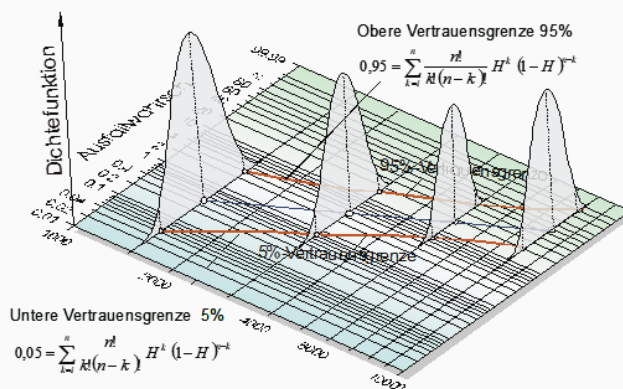
Speziell das Buch

Weibull & Zuverlässigkeitsmethoden

vertieft anwendungsbezogen die Statistiken und Methoden rund um Weibull und aller weitere Verteilungen. Die Versuchsplanung behandelt hier spezielle Lebensdauerfragen aufgrund unterschiedlicher Belastungen, Temperaturen, etc.

2.5.1 Vertrauensbereich der Weibull-Gerade

Bei der Weibull-Auswertung handelt es sich praktisch immer um eine Stichprobe. Die Gerade im Weibull-Diagramm entspricht also nur der Stichprobe. Je mehr Teile geprüft oder ausgewertet werden, desto mehr streuen die „Punkte“ um die Weibull-Gerade. Man kann statistisch eine Abschätzung über den Bereich der Grundgesamtheit machen. Hierfür wird ein sogenannter „Vertrauensbereich“ eingeführt. In der Regel gibt man diesen mit 90% an. Die obere Vertrauensgrenze entspricht dann einer Aussagewahrscheinlichkeit von $P_A=95\%$.



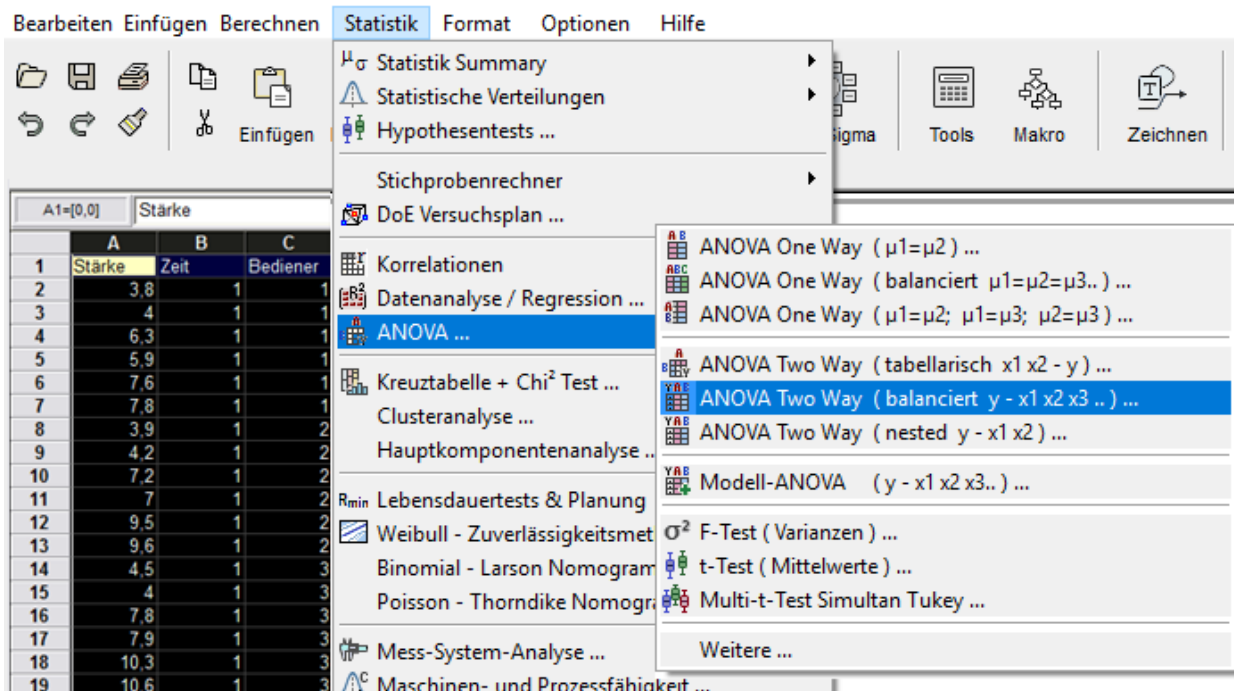
Weitere Informationen und Leseproben:

crgraph.de/Literatur

Anwendung in Visual-XSel

www.crgraph.de

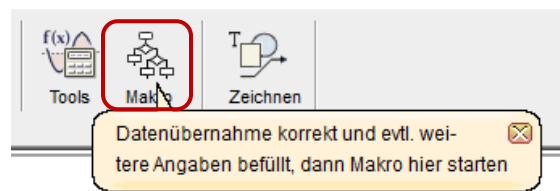
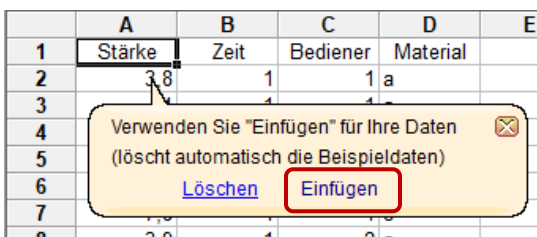
Bis auf die Modell-ANOVA sind alle Verfahren als Templates verfügbar. Diese befinden sich im Verzeichnis `..\Templates\03_Datenauswertung\`. Ein direkter Zugriff ist über den Menüpunkt Statistik möglich:



Wenn die Daten bereits in der Tabelle vorliegen, sind die gewünschten Spalten zu markieren und die Menüpunkte, wie gezeigt, aufzurufen. Es folgt noch eine Abfrage, ob die Datenspalten in der richtigen Zuordnung vorliegen, was durch die Pfeiltasten noch geändert werden kann.

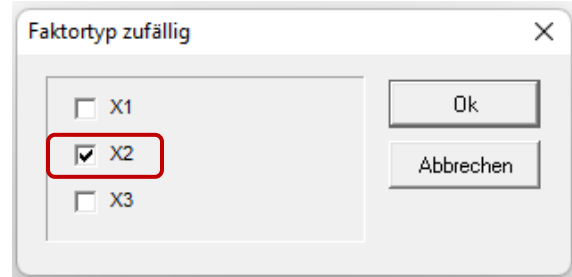


Wenn man noch keine Daten in der Tabelle hat aber in der Zwischenablage, ist nach Öffnen des Templates „Einfügen“ in der Sprechblase zu drücken und danach das Makro zu starten.



In dem Beispiel soll über eine ANOVA der Zusammenhang von Zeit, Bediener und Material auf eine Stärke (Zielgröße) untersucht werden.

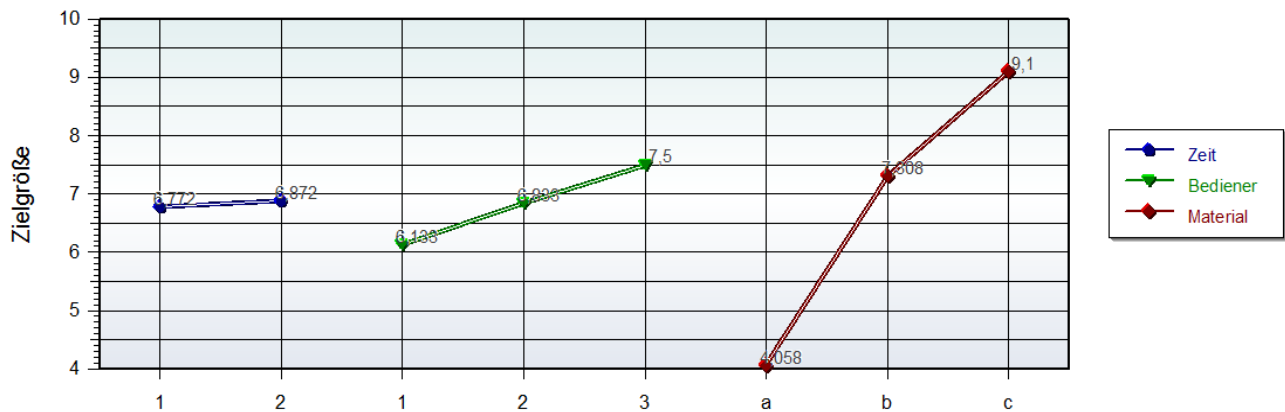
Nach Start des Makros folgt eine Abfrage, ob das Modell mit Wechselwirkung erstellt werden soll und welche zufällige Faktoren sind. Da die Bediener nicht die selben Personen sind, wie bei dem späteren Prozess, ist hier der zweite Parameter als zufällig zu deklarieren.



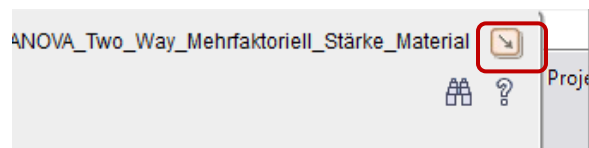
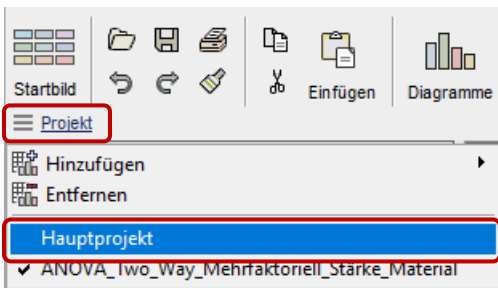
Das Ergebnisse wird im Hauptfenster des eingebetteten Templates ausgegeben:

	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			
Error	22	1,570E+00	7,136E-02					
Total	35	1,757E+02						

S =	2,671E-01
R ² =	0,991
R ² adj =	0,986



Um wieder in die ursprüngliche Ausgangstabelle zu gelangen, ist unter dem Menüpunkt *Projekt* das *Hauptprojekt* auszuwählen, oder das Template rechts außen zu schließen.





Software – Literatur – Consulting – Schulungen



Software

Unsere Software **Visual-XSel** ist ein leistungsfähiges Tool für alle wichtigen statistischen Qualitäts- und Zuverlässigkeitsmethoden. Nicht umsonst ist diese Software in vielen großen Firmen im Einsatz – crgraph.de/Referenzen.



Weitere Informationen zum aktuellen Thema finden Sie auf den nächsten Seiten oder unter www.versuchsmethoden.de/Versionen.pdf



Eigene Literatur

Unsere **Taschenbücher der statistischen Qualitäts- und Zuverlässigkeitsmethoden** beinhalten aktuelle und weiterführende Themen, z.B. zu Systemanalysen, Versuchsplanung und Datenauswertung, sowie zur Mess-System-Analyse und Prozessfähigkeit, Weibull- und Zuverlässigkeitsmethoden (neutral und softwareunabhängig).



Weitere Informationen finden Sie unter crgraph.de/Literatur



Consulting & Schulungen & Six Sigma

Bei unseren Inhouse- oder Online-Schulungen wird die praxisnahe Anwendung von statistischen Methoden vermittelt. Wir haben über 20 Jahre Erfahrung, insbesondere in der Automobilindustrie und unterstützen Sie bei Ihren Problemstellungen, führen Auswertungen für Sie durch, oder erstellen firmenspezifische Auswertevorlagen.



Weitere Informationen finden Sie unter crgraph.de/Schulungen



Hotline

Haben Sie noch Fragen, oder Anregungen? Wir stehen Ihnen gerne zur Verfügung:

Tel. +49 (0)8151-9193638

e-mail: info@crgraph.de

Besuchen Sie uns auf unserer Home-Page: www.crgraph.de