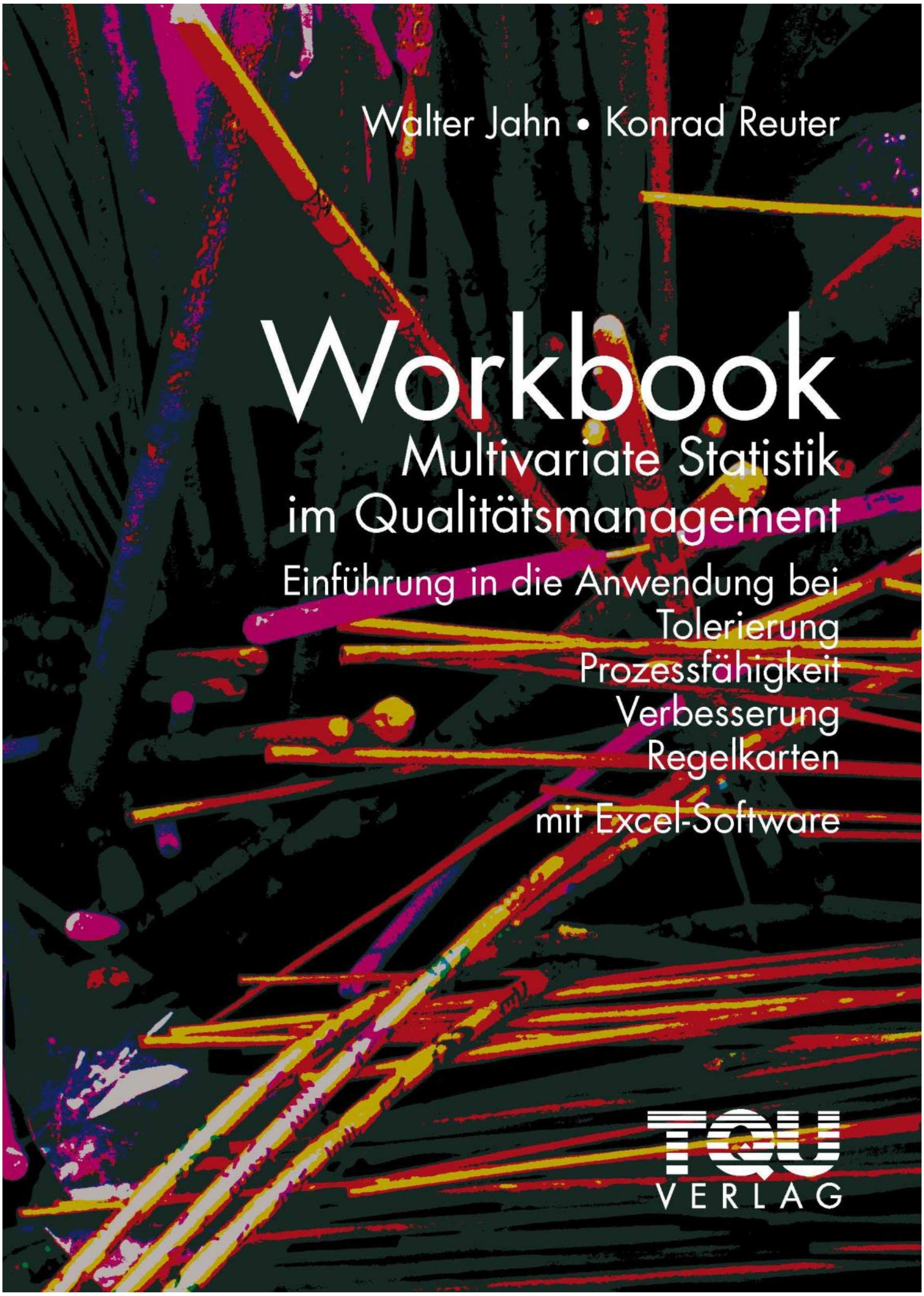
Walter Jahn • Konrad Reuter
Workbook
Multivariate Statistik
im Qualitätsmanagement
Einführung in die Anwendung bei
Tolerierung
Prozessfähigkeit
Verbesserung
Regelkarten
mit Excel-Software
TQU
VERLAG

Walter Jahn† • Konrad Reuter

Workbook

Multivariate Statistik im Qualitätsmanagement

Einführung in die Anwendung bei

Tolerierung
Prozessfähigkeit
Verbesserung
Regelkarten

mit Excel-Software

Die Autoren

Dr. Walter Jahn† hat Mathematik studiert und war als Leiter des Fachgebietes mathematische Statistik an der Universität Leipzig in Lehre, Forschung und Anwendung tätig. Die Ergebnisse seiner Forschung liegen in zahlreichen Publikationen vor. Er war freiberuflicher Berater und vorrangig auf dem Gebiet der Anwendung multivariater statistischer Methoden in der Industrie tätig. Seine Spezialgebiete waren das moderne Produktaudit mit der multivariaten statistischen Tolerierung, den multivariaten Prozessfähigkeitsuntersuchungen und der Auswahl wesentlicher Regressoren.

Dr. Konrad Reuter hat Maschinenbau studiert und war langjährig in der Textilforschung tätig. Er ist freiberuflicher Berater und Trainer mit großer praktischer Erfahrung im modernen Qualitätsmanagement. Als Wissenspartner im TQU Verbund hat er sich auf die praktische Anwendung der Methoden und Verfahren der technischen Statistik spezialisiert. beratung@konrad-reuter.de

Workbook
Multivariate Statistische Methoden für das Qualitätsmanagement

Die Abbildungen und Tabellen aus EXCEL-Dateien können von den beigestellten Quelldateien leicht abweichen, da diese laufenden Verbesserungen unterliegen können, die nach dem Druck erfolgt sind.

Eigenverlag und Eigenvertrieb
Herausgeber: Prof. Dr. Jürgen P. Bläsing
TQU Verlag Ulm, Magirus-Deutz-Straße 18, 89077 Ulm Deutschland
Fon: +49 731 14660-200, verlag@tqu-group.com, www.tqu-group.com
Erste Auflage 2012, überarbeitet 2018

Inhaltsverzeichnis

Motivation

Probleme

Ein berechtigter Anspruch der Kunden besteht in der simultanen Erfüllung aller relevanten Kundenanforderungen. Damit wird nach Crosby die Qualität eines Produktes oder einer Dienstleistung verständlich definiert.
Verluste durch Qualitätsschwankungen sind keine außergewöhnlichen Ereignisse in einem Unternehmen. Schwankungen können neben den üblichen Ursachen auch technische Weiterentwicklung, laufende Anpassungen, fehlerhafte Spezifizierungen der Produkte und Inputs, auch die zunehmende Kompliziertheit der Produkte und Prozesse sein.
Zum Nachweis der simultanen Erfüllung **aller** Kundenanforderungen werden Prozessfähigkeitsindizes verwendet. Diese bieten die ideale Möglichkeit, das Problem zu definieren und zu entscheiden, wie das Problem gelöst werden muss.

Ziele von Verbesserung

Die multivariate statistische Analyse von Produkten und Prozessen zeigt widerspruchsfrei, wie stark die Streuungen der Produktvariablen durch die zugeordneten Input- und Prozessvariablen verursacht werden. Erst mit diesem Wissen können wir die Streuungen der Produktvariablen durch Steuerung der Input- und Prozessvariablen reduzieren.
Zur Verbesserung der Produkte und Prozesse sind verstärkt statistische Verfahren eingesetzt worden (neben den klassischen SPC-Methoden u. a. „Six Sigma“-Methoden).
CAQ Software mit statistischen Verfahren ist weit verbreitet.

Dies besagt aber noch nichts über den Grad der effektiven Verwendung dieser Methoden. Neue Software (z. B. erweiterte Anwendung von Excel, Statgraphics, S-plus for Windows …) ist zur Realisierung multivariaten statistischen Methoden erforderlich.
Das grundlegende Ziel der Prozesssteuerung ist es, dafür zu sorgen, dass die Messwerte der Produktvariablen im Toleranzgebiet liegen.

Die Prozesse sind so zu steuern und regeln, dass

- sie auf den Sollwert zentriert sind und
- eine Reduzierung der Streuung erreicht wird.

Begrenztheit der konventionellen univariaten Methoden

Nicht abzustreiten ist, dass ein gehöriges Maß an Erfahrung und eine ganze Portion an Heuristik zum gegenwärtig anerkannt hohen Stand der Produktqualität beitragen.

Ein Blick in Unternehmenspraktiken und auf Schulungsprogramme sowie qualitätsrelevante Veröffentlichungen zeigt, dass fast ausschließlich mit statistischen Methoden gearbeitet wird, die der univariaten Statistik zuzuordnen sind.

Dies bedeutet, dass auch bei komplexen Produkten und Prozessen die interessierenden Variablen als unabhängig angesehen werden und der Versuch unternommen wird, diese einzeln zu überwachen, zu steuern, bzw. zu optimieren (Regelkarten, Prozessfähigkeit, Maßketten, Messunsicherheit).

Damit sind unzweifelhaft Erfolge erreicht worden. Trotzdem stößt diese Vorgehensweise an grundsätzliche Grenzen. In Unternehmen wird dies immer dann erkennbar, wenn bei Nichterfüllung der Anforderungen die Fragen nach den eigentlichen Ursachen trotz der Ursache-Wirkungs-Diagramme nur mangelhaft beantwortet werden können, da die Quantifizierung der Auswirkungen möglicher, wahrscheinlicher und tatsächlicher Ursachen auf die Produktvariablen nicht zuordenbar gegeben ist.

Eine Regelkarte als statistisches Verfahren zur Merkmalsüberwachung erfordert die Festlegung einer (kleinen) Irrtumswahrscheinlichkeit α. Häufig wird dafür der Wert von 1 % eingesetzt. Ist ein Produkt durch mehrere Merkmale charakterisiert (m > 20 ist nicht selten) und werden diese einzeln überwacht, so erhöht sich die Irrtumswahrscheinlichkeit nach der Beziehung $\alpha_{ges}=1-(1-\alpha)^m$ mit m als der Anzahl der Produktmerkmale (Variablen). Die gleiche Logik betrifft die Prozessfähigkeit. Es wird erkennbar, dass die Wahrscheinlichkeit von Fehlentscheidungen mit möglichen finanziellen Konsequenzen exponentiell mit der Anzahl der Variablen ansteigt.

Überwachungen des Prozesses mit mehreren Produktvariablen und daraus folgende Entscheidungen für jede Produktvariable separat sind folglich nicht sinnvoll . Stellen Sie sich bitte wieder das Platonsche Höhlengleichnis mit zwei Produktvariablen vor, von denen eine die Kundenanforderung erfüllt und die andere nicht. Das Problem ist mit den univariaten Methoden nicht lösbar.

Die Multivariate Sicht

Es ist ein Axiom, dass Produkte durch mehrere, nicht unabhängige Produktvariable $Y_1, Y_2, \ldots, Y_m$ charakterisiert werden. Versuchen Sie ein Produkt zu finden, das nur durch eine Variable charakterisiert wird! Selbst der Messwert als Ergebnis des Messprozesses wird durch mehrere Variable beschrieben. Hinsichtlich aller relevanten Produktvariablen müssen Kundenanforderungen spezifiziert und simultan erfüllt werden.

Jedes Produkt ist stets das Ergebnis eines Prozesses. Das zeigt uns die Abbildung 4. Die Inputvariablen eines Prozesses sind <u>immer</u> Produktvariablen eines Vorläuferprozesses. Über die Input- und Prozessvariablen können die Produktvariablen beeinflusst werden. Damit muss man akzeptieren, dass sowohl die Input- als auch die Produktvariablen stochastischer Natur sind. Bei den Prozessvariablen kann es ohne weiteres auch fixe Variable geben.

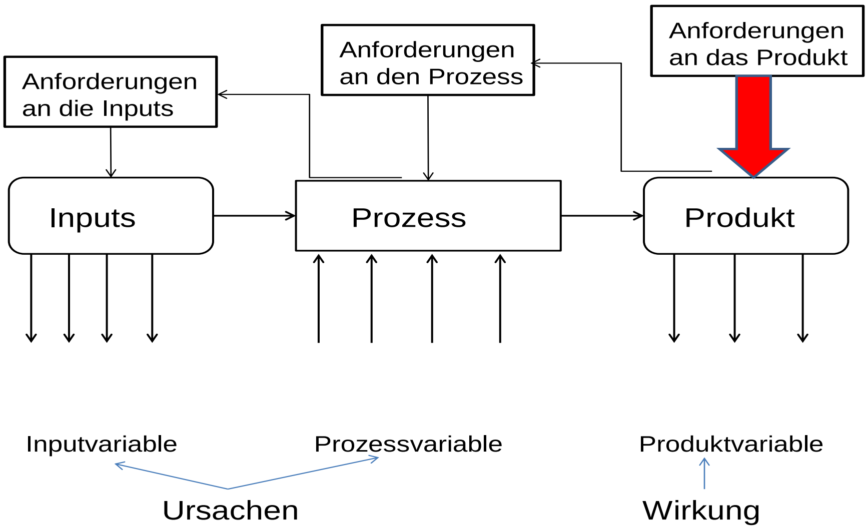

Abb. 4: Prozess mit den Anforderungen

Die Abhängigkeiten zwischen den Input-, Prozess- und Produktvariablen sind durch die Gesetze der Naturwissenschaften, der Ökonomie usw. gegeben. Wir fassen die paarweisen Abhängigkeiten zwischen allen möglichen Variablen zur Abhängigkeitsstruktur zusammen. Diese sind sehr stabil, durch den Menschen kaum beeinflussbar und damit quasi der „Fingerabdruck“ des Prozesses.

Jegliche anzuwendende Methodik zur Prozessanalyse und Steuerung muss auf der Abhängigkeitsstruktur zwischen den Variablen aufbauen. Die folgende Abbildung beschreibt diese Abhängigkeitsstruktur ganz allgemein.

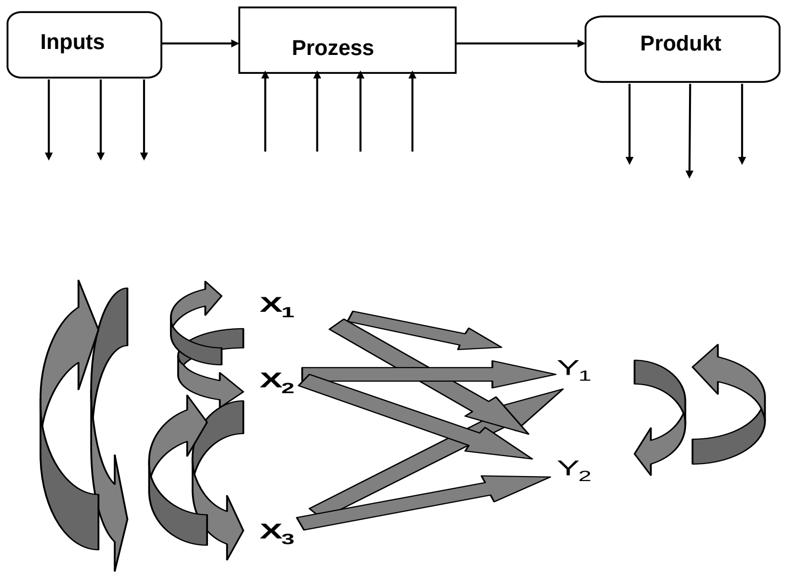

Abb. 5: Abhängigkeitsstruktur der Variablen eines Prozesses

Partielle Korrelationskoeffizienten

Die Betrachtung der Korrelationskoeffizienten für alle möglichen Paare von Variablen z. B. r_{12} für X_1 und X_2 lässt die Wirkungen der nicht in den Korrelationskoeffizienten einfließenden Variablen $X_3, \ldots, X_n$ außer acht.
Diese Variablen wirken aber nach Abbildung 11 auf r_{12}.
Das legt die Frage nahe, wie stark r_{12} durch diese „restlichen“ Variablen beeinflusst wird?
Zur Beantwortung nehmen wir an, dass die Wirkungen der Variablen $X_3, \ldots, X_n$ auf X_1 und X_2 konstant sind und berechnen den daraus resultierenden „bedingten“ Korrelationskoeffizienten, den man partiellen Korrelationskoeffizienten nennt.

Um das noch klarer zu erklären betrachten wir nur drei Zufallsvariable Y, X_1 und X_2 in der Abbildung 13.

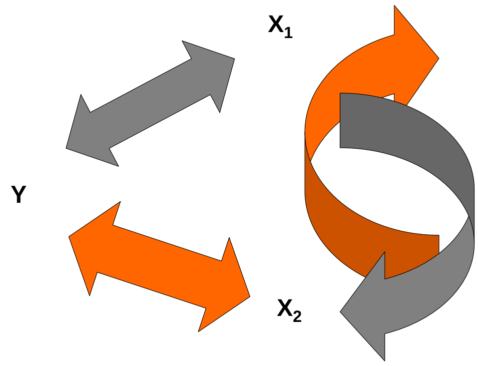

Abb. 13: Abhängigkeitsstruktur zwischen drei Zufallsvariablen

In dieser Abbildung setzen wir die Wirkung von X_1 auf X_2 konstant. Das entspricht dem Kappen der Abhängigkeit zwischen X_1 und X_2.

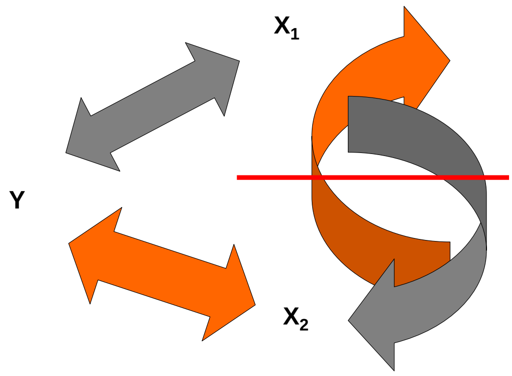

Abb. 14: Kappen der Abhängigkeit zwischen X_1 und X_2

In diesem Falle erhalten wir die unverfälschten - direkten - Einflüsse von X_1 und X_2 auf Y. In Formeln kann das z. B. wie folgt ausgedrückt werden

$$r_{YX1/X2} = \frac{r_{YX1} - r_{YX2} r_{X1X2}}{\sqrt{\left(1 - r_{YX2}^2\right)}\sqrt{\left(1 - r_{X1X2}^2\right)}}$$

d. h. die Korrelation zwischen Y und X_1 wird vom Einfluss der Korrelationen zwischen Y und X_2 und X_1 und X_2 bereinigt.

Das Ergebnis ist der partielle Korrelationskoeffizient zwischen Y und X_1 unter der Bedingung, dass die Abhängigkeit zwischen X_1 und X_2 weder die Abhängigkeit zwischen Y und X_1 noch die zwischen Y und X_2 beeinflusst.

partielle Korrelationskoeffizienten		bremsweg	geschw	profil	reaktion	temperatur
		Y	X_1	X_2	X_3	X_4
bremsweg	Y		0,9677	-0,9528	0,9136	0,9168
geschw	X_1			0,8834	-0,8771	-0,8379
profil	X_2				0,8679	0,9550
reaktion	X_3					-0,8077
temperatur	X_4					

Tabelle 6: Partielle Korrelationskoeffizienten

Interpretation:

- Die partiellen Korrelationskoeffizienten von Y mit den anderen Zufallsvariablen werden größer.
- Die negative Korrelation r_{12} = -0,2520 zwischen Geschwindigkeit und Profiltiefe wird zum positiven partiellen Korrelationskoeffizienten $r_{X1,\ X2/Y,X3,\ X4}$ = 0,8834.
- Die sehr kleine Korrelation zwischen der Profiltiefe und der Reaktionszeit wird viel größer.

Diese partiellen Korrelationskoeffizienten können in der Praxis nicht beobachtet werden; es sind notwendige Konstrukte zum Erkennen der totalen (durch andere Variablen unverfälschten) Abhängigkeiten, die für die Steuerung des Prozesses zwingend erforderlich sind.

Regressionsanalyse

Einfache lineare Regression

Wir beginnen mit der Betrachtung zweier unkorrelierter Variabler z.B. von Y_1 und Y_2. In einer Abbildung stellen wir die Verteilungsdichten von Y_1 und Y_2 und zusätzlich die bedingte Verteilung $Y_2 = f(Y_2/Y_1)$ dar. Die bedingte Verteilung ist eine Funktion von Y_1.

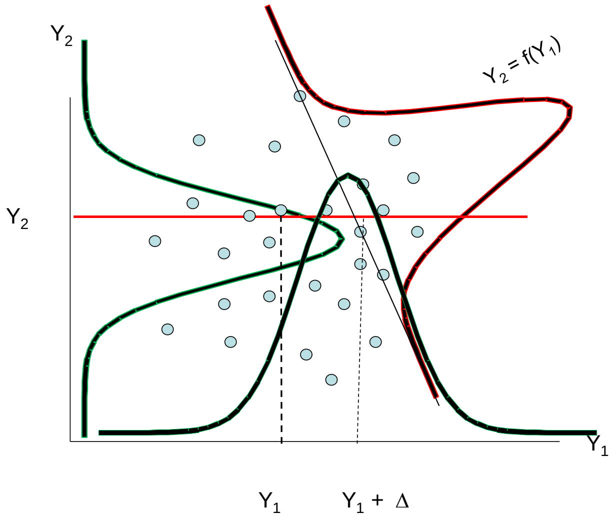

Abb. 15: Unabhängige Produktvariable Y_1 und Y_2, kreisförmig umrissenes Streugebiet

Im unabhängigen Fall gewinnt man aus der Messung von z. B. Y_2 keine Information über Y_1. Jede Richtung im Kreis ist gleichberechtigt. Hieraus folgt, die Veränderung eines Messwertes Y_1 um den Betrag Δ führt zu keiner Veränderung von Y_2. Daher ist die bedingte Verteilung (rote Verteilung) von Y_2 unter der Bedingung von Y_1 ebenso breit wie die ursprüngliche Verteilung von Y_2.

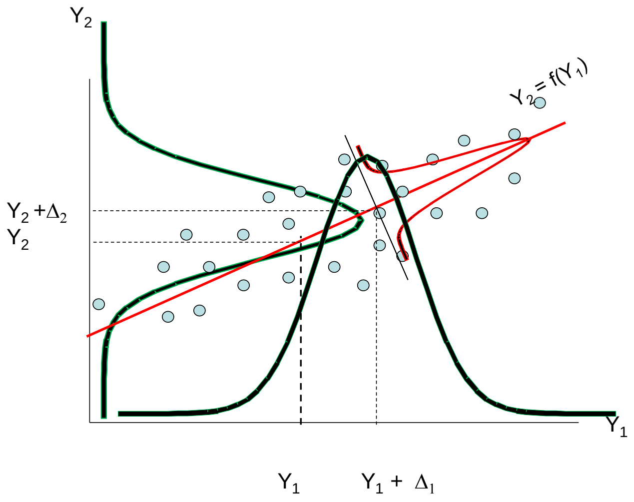

Abb. 16: Korrelierte Produktvariable Y_1 und Y_2, elliptisch umrissenes Streugebiet

Multivariate Prozessfähigkeit

Die Fähigkeit eines Prozesses ist sein inhärentes Potenzial, Produkte mit vorgegebenen Eigenschaften zu produzieren. Die vorgegebenen Eigenschaften werden durch die Sollwerte und Toleranzgrenzen definiert. Das inhärente Potenzial ist die Fähigkeit des Prozesses, Werte für die Produktvariablen zu produzieren, die mit vorgegebener Wahrscheinlichkeit innerhalb des Toleranzbereiches liegen. Das bedeutet, ein Prozess ist fähig, wenn er simultan alle relevanten Kundenanforderungen erfüllt.

Für eine Produktvariable Y wird ein univariater Prozessfähigkeitsindex C_p nach

$$C_p = \frac{\text{Toleranzbreite}}{\text{Breite der Häufigkeitsverteilung}} = \frac{\text{was der Kunde fordert}}{\text{was der Kunde erhält}} = \frac{T_o - T_u}{6 \cdot s}$$

berechnet.

Wenn $C_p < 1$, dann ist die 6s Breite der Häufigkeitsverteilung größer als die Toleranzbreite. Daraus folgt, es werden Produkte gefertigt, die außerhalb des Toleranzintervalls liegen. Wenn $C_p > 1$, dann ist die 6s Breite der Häufigkeitsverteilung kleiner als die Toleranzbreite. Trotzdem kann Ausschuss produziert werden. Das liegt daran, dass bei C_p nur das Streuverhalten der Produktvariablen Y beachtet wurde. Aber die Verteilung von Y wird durch Mittelwert und Streuung charakterisiert. Folglich müssen wir noch die Abweichung des Mittelwertes vom Sollwert in die Fähigkeitsbetrachtung einbeziehen. Das geschieht durch einen Korrekturterm der Dezentralität.

$$k = \frac{1/2T}{\left(1 - \left|\overline{Y} - M\right|\right)}$$

Mit dem Korrekturterm k kann C_p in der folgenden Weise berechnet werden.

$$C_{pk} = \frac{C_p}{k}$$

Diese Darstellung mit dem Korrekturterm k ist informativer als die übliche Berechnung von C_{po} und C_{pu}.

Die multivariate Erweiterung der Prozessfähigkeitsberechnungen basiert auf der Verallgemeinerung der Standardabweichung s zur Kovarianzmatrix S und des Toleranzintervalls T zum mehrdimensionalen Toleranzbereich über die Berechnung mit der Kotoleranzmatrix Σ^*_{YY}. Die Interpretation der multivariaten PFI muss dieselbe bleiben wie bei den univariaten. Die multivariaten PFI werden nach [Jahn] mit den Formeln

$$\boldsymbol{MC}_p = \sqrt{\frac{\boldsymbol{T}^T \boldsymbol{S}_{YY}^{-1} \boldsymbol{T}}{\boldsymbol{T}^T \left(\Sigma^*_{YY}\right)^{-1} \boldsymbol{T}}}$$

multipler Prozessfähigkeitsindex,

$$K = \sqrt{1 + \left|\overline{Y} - Soll\right|^T \left(\Sigma^*_{YY}\right)^{-1} \left|\overline{Y} - Soll\right|}$$

Prozesssteuerung

Weiter vorn haben wir bereits geschrieben, dass die Steuerung eines Prozesses in den folgenden Stufen erfolgen muss.

- Berechnung der Prozessgleichung in den wesentlichen Input- und Prozessvariablen,
- Berechnung des Toleranzintervalls für eine Produktvariable bzw. des Toleranzbereiches für mehrere Produktvariable
- Berechnung des Toleranzbereiches für die Inputvariablen, die ja Produktvariablen von Vorläuferprozessen sind - Quantifizierung des ex- und/oder internen Kunden-Lieferanten-Verhältnisses als Ausdruck der Kommunikation zwischen Prozessen
- Berechnung der Steuerintervalle in Abhängigkeit von dem Toleranzbereich für die Inputvariablen für die Prozessvariablen
- Messung, Fähigkeitsnachweise für die Inputvariablen
- Berechnung der Einstellung für die Prozessvariablen in Abhängigkeit von einem oder mehreren Wertesätzen für die Inputvariablen
- Einsetzen dieser Werte in die Prozessgleichung

Dieses Vorgehen gewährleistet, dass die Werte für die Produktvariablen im Toleranzgebiet liegen.
Am besten lässt sich das Vorgehen anhand eines Beispiels demonstrieren.

Beispiel Bremsweg:
Produktvariable Y = Bremsweg in Abhängigkeit von

den Prozessvariablen X_1 = Geschwindigkeit,
X_2 = Reaktionszeit des Fahrers und

den Inputvariablen Z_1 = Tiefe des Reifenprofils,
Z_2 = Umgebungstemperatur.

Die Korrelationsmatrix für dieses Beispiel besitzt die Werte der Tabelle:

R_{YXZ}	Y	X_1	Z_1	X_2	Z_2
Y	1	0,862	-0,431	0,428	0,324
X_1	0,862	1	-0,252	0,152	0,264
Z_1	-0,431	-0,252	1	0,097	0,568
X_2	0,428	0,152	0,097	1	0,259
Z_2	0,324	0,264	0,568	0,259	1

Tabelle 21: Korrelationsmatrix für das (erweiterte) Bremswegbeispiel

Berechnung der Prozessgleichung mit den wesentlichen Input- und Prozessvariablen

$$y = b_0 + b_{Yx1/z1,x2,z2} \cdot x_1 + b_{Y.z1/x1,x2,z2} \cdot z_1 + b_{Y.x2/x1,z1,z2} x_2 + b_{Y.z2/x1,z1,x2} z_2$$

Für die umgerechneten Toleranzgrenzen gilt dann

$$T_{\hat{Y}o} = \frac{T_o - Soll}{\sqrt{n}} \quad \text{und} \quad T_{\hat{Y}u} = \frac{Soll - T_u}{\sqrt{n}}$$

Die Einzelwertkarte für Toleranzgrenzen:

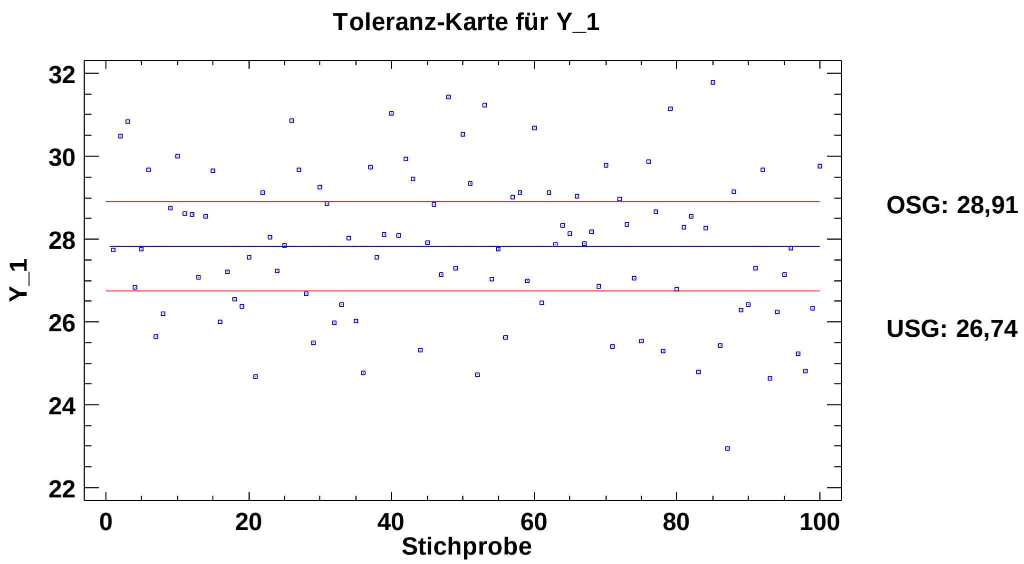

Abb. 26: Einzelwert Annahme Regelkarte für Y_1

Diese Karte spiegelt die Realität des Prozesses wider mit einer Fähigkeit von C_{pk} = 0,18.

Der Vergleich der Karten bestätigt eindrucksvoll, dass:

- Die Karten mit den Toleranzgrenzen für die Kunden sinnvoller sind, als die Regelkarten mit den statistischen Grenzen - siehe noch einmal unsere Qualitätsdefinition: „simultane Erfüllung aller relevanten Kundenanforderungen“.
- Die Karten mit den Toleranzgrenzen bestätigen die Prozessfähigkeitsindizes, sowohl die univariaten als auch die multivariaten.
- Die multivariate statistische Grenze nach Hotelling ist hauptsächlich vom Stichprobenumfang abhängig, daher haben sehr verschiedene Verläufe für das aggregierte Produkt dieselbe Grenze. Das sollte aber nicht sein.

Ausblick

Diesen kurzen Leitfaden haben wir in der Überzeugung und aufgrund unserer praktischen Erfahrungen bei der Lösung betrieblicher Probleme geschrieben, um Ihnen zu zeigen, dass

- es in jedem Unternehmen zur Lösung anstehende Probleme gibt.
- jedes Problem auf dem generellen Ursache-Wirkungs-Prinzip basiert.
- wir Ihnen daher die multivariaten statistischen Methoden vorstellen müssen, die Sie bei der Lösung der Probleme unbedingt benötigen.

Die Multivariaten statistischen Methoden unterscheiden sich wesentlich von den Versuchsplanmethoden, die auf der grundsätzlichen Voraussetzung fixer Input- und Prozessvariablen basieren.

Außerdem wollen wir Ihnen zeigen, dass die anzuwendenden Methoden in einem „organischen“ Zusammenhang stehen. Soll z. B. ein Prozess so gesteuert und geregelt werden, dass dieser simultan alle relevanten Kundenanforderungen erfüllt, dann müssen die Anforderungen vorher durch Sollwerte und Toleranzgrenzen für einen oder alle in der Regel korrelierten Produktvariablen spezifiziert vorliegen oder spezifiziert werden. Das Toleranzintervall oder der Toleranzbereich ist das Zielgebiet für die Steuerung.

Der Nachweis der simultanen Erfüllung aller relevanten Kundenanforderungen liefert Ihnen den quantifizierten Ausdruck für die Qualität und damit eine sinnvolle Grenze für die multivariate Regelkarte.

Wir hoffen, Ihnen den Einstieg in diese Methodik zu erleichtern, wünschen Ihnen Erfolgserlebnisse, würden uns riesig freuen von Ihnen etwas über Ihre Erfolge zu hören, nehmen Ihre Anregungen zu Verbesserungen und Korrekturen gerne entgegen und stehen Ihnen mit Rat und Tat zur Seite.