

Seminararbeit:

**Breakthroughs in Statistical
Methodology**

**G.E.P. Box and K.B. Wilson - On the
Experimental Attainment of Optimum
Conditions**

Autorin: Natalie Grimm

Matrikelnummer: 10567152

Ludwig- Maximilians- Universität München

Betreuer: Paul Fink, M.Sc.

28.05.2014

Inhaltsverzeichnis

1	Geschichtlicher Abriss bis 1951	2
1.1	Allgemeines	2
1.2	Faktorenanalyse	2
1.2.1	Idee	2
1.2.2	Studium	2
1.2.3	Vorgehensweise und Methodik	3
1.2.4	Problematik	4
2	On the Experimental Attainment of Optimum Conditions	5
2.1	Idee	5
2.2	Neuerung im Vergleich zur Faktorenanalyse	6
2.3	Vorteile	6
2.4	Vorgehensweise und Methodik	6
2.4.1	Allgemein	7
2.4.2	Rechenbeispiel	8
3	Auswirkungen dieser neuen Herangehensweise	11
3.1	Statistik im Allgemeinen	11
3.2	Fazit	11

1 Geschichtlicher Abriss bis 1951

1.1 Allgemeines

Bis zur Veröffentlichung der Arbeit von George E.P. Box und K.B. Wilson im Jahr 1951, war in dem Gebiet der Statistik und insbesondere der Versuchsplanung die wissenschaftlichen Erkenntnisse von W.S. Gossett mit der „Entdeckung“ der Student t-Verteilung und dem t-Test im Jahr 1908 sowie R.A. Fisher mit seiner Einführung des Maximum-Likelihood-Prinzip und der Varianzanalyse (ANOVA) voran getreten. Weitere Erkenntnisse für die Versuchsplanung und die Statistik im Allgemeinen lieferten zudem die Faktorenanalyse, dessen Grundidee von C. Spearman stammt. In seinen Versuchen Intelligenz besser zu verstehen und beschreiben zu können, entwarf er die Faktorenanalyse. Die Testergebnisse sollten erklären, ob Intelligenz als latente Größe auf einen allgemeinen Faktor oder auf mehrere spezifische Bereiche eingeschränkte Faktoren zurückzuführen ist. (Kuß, 2012, vgl. S. 266) Versuchsplanung ist in dem Maße nicht nur für Unternehmen wichtig effizient zu arbeiten, um eine geringere Anzahl an Versuchen durchführen und damit Kosten sparen zu können, sondern auch die Möglichkeit einer Betrachtung von Interaktionen zwischen den Einflüssen. (Kiesling, Kiesling, vgl.)

1.2 Faktorenanalyse

Im Folgenden Abschnitt wird ein verstärkter Blick auf die Faktorenanalyse geworfen, da diese im direkten Bezug zu Box und Wilsons Artikel steht und eine Einführung für das weitere Verständnis von Nöten ist.

1.2.1 Idee

Die Grundidee der Faktorenanalyse, die als ein Überbegriff für verschiedene Untermethoden verstanden werden kann, beschreibt den Prozess Strukturen zwischen den einzelnen Variablen in Form der Korrelation zu erkennen und die Daten zu reduzieren. Faktoren werden dabei als die Gruppe der hoch-korrelierten Variablen bezeichnet. (Backhaus et al., 2011, vgl. S. 330)

1.2.2 Studium

Im Studienfach Statistik der Ludwig- Maximilians- Universität München wird die Faktorenanalyse vor allem in den Bereichen der Veranstaltungen der Versuchsplanung und Multivariaten Verfahren behandelt sowie in der Veranstaltung Generalisierte Lineare

Modelle angeschnitten. (Ludwigs-Maximilians-Universität, 2013, vgl.) Zudem wird in weiteren Studienfächern wie zum Beispiel in den Bereichen der Betriebswissenschaften, Soziologie und Psychologie dieses Thema behandelt.

1.2.3 Vorgehensweise und Methodik

Nun wird im Folgenden das Grundmodell der Faktorenanalyse sowie die Vorgehensweise und Methodik im Kurzen beschrieben und allgemein erläutert:

Am Ausgangspunkt werden viele, häufig miteinander korrelierte Variablen x_1, \dots, x_p betrachtet. Zwischen diesen und den latenten Variablen z_1, \dots, z_k besteht ein Zusammenhang und eine Korrelation. Die Vorstellung hinter dieser Analyse besagt, dass die Korrelation zwischen x_i, x_j durch die latenten Größen beziehungsweise Faktoren bewirkt werden. Ein Faktor ist hierbei eine latente Variable, die sich oftmals als eine kategoriale Variable darstellen kann.

Das Grundmodell besagt:

$$X - \mu = \Gamma F + U \quad (1.1)$$

wobei

X : $(p \times 1)$ - Vektor der beobachtbaren korrelierten Größen der Untersuchung

μ : Erwartungswert von X , als Vektor mit $(p \times 1)$ Eintragungen

F : $(k \times 1)$ - Vektor der gemeinsamen Faktoren mit $k \ll p$

Γ : $(p \times k)$ - Matrix der Faktorladungen $= (\gamma_{ij})$

γ_{ij} : Gewicht des j -ten Faktors für z.B. *Item_i* oder *Aufgabe_i* mit $Cov(X_i, F_j) = \gamma_{ij}$

U : $(p \times 1)$ - Vektor der spezifischen Faktoren

darstellen.

Faktorladungen beschreiben hierbei die Maßgröße für den Zusammenhang zwischen der Variable und den Faktoren. Diese ist ebenfalls der Korrelationskoeffizient zwischen Variablen und den Faktoren.

Die Faktorenanalyse spezifiziert eine lineare Grundstruktur, die sich im Folgenden ausführen lässt:

$$X_1 = \mu_1 + \gamma_{11}F_1 + \dots + \gamma_{1k}F_k + U_1 \quad (1.2)$$

⋮

$$X_p = \mu_p + \gamma_{p1}F_1 + \dots + \gamma_{pk}F_k + U_p . \quad (1.3)$$

Gleichfalls wurden folgende weitere Annahmen getroffen:

(1) $E(F) = 0$

(2) $Cov(U) = diag(\psi_1^2, \dots, \psi_p^2) = \psi$

(3) $Cov(F, U) = 0$

(4) $Cov(F) = I$.

Die Fehlerterme $\rho(U_i, U_j)$ korrelieren nicht miteinander, ist gleich null, da die partielle Korrelation der X -Werte nach den latenten Faktoren verschwinden (vgl. (2)). In (3) wird ersichtlich, dass die Kovarianz $Cov(F, U)$ durch die Zerlegung in unabhängige Komponenten null ergibt. Die $Cov(F)$ wird als orthogonales Faktorenmodell benannt, da es die

Einheitsmatrix I ergibt.

Des Weiteren ist die Definition des Fundamentalsatzes für $s = Cov(X)$ für spätere Annahmen von Nöten:

$$s = \Gamma\Gamma^T + \psi . \quad (1.4)$$

Die Faktorenanalyse zerlegt die Kovarianzmatrix, in

$$\Sigma = \text{Faktorenladungen} \cdot \text{Faktorenladungen}^T + \text{Diagonalelemente} . \quad (1.5)$$

(Koch and Lindenlaub, 2012, vgl S. 98-100)

In der Praxis, der explorativen Datenanalyse, wird die Faktorenanalyse nach folgenden sechs Schritten durchgeführt: Während in Schritt 1 die irrelevanten Variablen im Vorfeld aussortieren werden und die Korrelationsmatrix gebildet wird, folgt in Schritt 2 die Extraktion der Faktoren. Im dritten Schritt werden die Kommunalitäten bestimmt, die sich auch als Kommunalitätenprobleme darstellen können. Dies bedeutet, dass die gemeinsamen Faktoren nicht die Gesamtvarianz erklären können und somit in der Regel kleiner eins sind. Die Zahl der Faktoren wird im vierten Schritt bestimmt, meist vom Anwender. Danach werden in den letzten beiden Schritten die Faktoren interpretiert und die Faktorwerte bestimmt. (Backhaus et al., 2011, vgl. S. 335-369) Genauere Einzelheiten werden bei der Gegenüberstellung in Kapitel 2.2 erläutert.

1.2.4 Problematik

Jedoch können Probleme bei der Faktorenanalyse auftreten. Das Rotationsproblem zu einem beschreibt die Problematik, dass die Kovarianzmatrix von X nach dem Fundamentalsatz zerlegt werden soll, um die Daten zufriedenstellend interpretieren zu können. Da sich jedoch eine orthogonale Matrix finden lassen kann, die jene Bedingungen erfüllt, führt das zu einem Problem der Uneindeutigkeit und lässt viele Möglichkeiten der Interpretation offen. Als weitere Probleme werden noch das Kommunalitätenproblem, in denen ψ und k nicht eindeutig bestimmt werden können und die Schätzung der Faktorenanalyse behandelt. (Koch and Lindenlaub, 2012, vgl. S. 101-103) Diese werden jedoch im Folgenden nicht weiter behandelt.

2 On the Experimental Attainment of Optimum Conditions

Im Nachfolgendem Kapitel wird nun auf den Artikel von G.E.P. Box und K.B. Wilson eingegangen und ihre bedeutende Entdeckung im Bezug auf die Statistik erläutert.

2.1 Idee

Imperial Chemical Industries (ICI), aus Manchester, erkannte als eine der ersten Firmen die Bedeutsamkeit einer Zusammenarbeit mit Statistikern. In diesem Fall genauer gesagt, G. Box als Statistiker und K. Wilson als Chemiker. Diese Gemeinschaftsarbeit in der Studie „On the Experimental Attainment of Optimum Conditions“ (Johnson and Kotz, 1991, S. 267) legte die Grundlage für das ganze Feld der Response Surface Methodology, kurz RSM. Diese war in ihren Anfängen besonders wichtig für Chemiebranche um die Anzahl an Versuchen bei Einstellungen von zum Beispiel Temperatur, Konzentration sowie Druck etc. zu reduzieren. Zu Beginn dieser Versuche wurde mit ersten Ableitungen der Response Surface bis hin zur Surface mit unbekanntem Schätzungen gearbeitet. Nach einigen ersten Erfolgen wie offensichtlichen Interaktionen zwischen den Inputs wurde sich nicht nur auf faktorielle Designs konzentriert. Das Ziel dieser neuen Analyseverfahren ist, einen optimalen Response zu finden. Dieser optimale Punkt soll entweder das Maximum oder Minimum darstellen, jedoch in einer mehrdimensionalen Betrachtung. Bei einer chemischen Betrachtung dieses Problems betrachtet, sollen die Anzahl der Versuche minimiert werden um so unter anderem Kosten zu sparen. Allgemein zusammengefasst soll eine Optimierung der Versuchsanzahl stattfinden um den besten Response zu finden. Das Problem stellt sich hier gleich zu Beginn dieser Idee: Der Response η ist vermutlich abhängig von k quantitativen Faktoren oder Variablen $x_1, \dots, x_t, \dots, x_k$. Dies wird folgend ausgedrückt für die u -te Kombination der Faktorebenen ($u = 1, \dots, N$):

$$\eta = \varphi(x_{1u}, \dots, x_{ku}) . \quad (2.1)$$

Die experimentelle Umgebung hierbei ist in R mit k -dimensionalen Faktorraum. Deutlicher ausgedrückt, beschreibt das Problem, dass die kleinste Anzahl an Experimenten zu finden ist, an dem der Punkt $(x_1^0, \dots, x_t^0, \dots, x_k^0)$ in dem Raum R , wenn η ein Maximum oder Minimum ist. Für ein Beispiel im chemischen Kontext bedeutet es, dass die kleinstmögliche Anzahl an Experimenten mit der Ausbeute, der Reinheit und den Kosten eines Produkts die Responsevariablen darstellen, die maximiert oder minimiert werden sollen. Die Faktoren, die den Response beeinträchtigen sind Variablen wie Tem-

peratur, Druck, Reaktionszeit, Proportionen der reagierenden Substanzen etc. (Johnson and Kotz, 1991, vgl. S. 267-271)

2.2 Neuerung im Vergleich zur Faktorenanalyse

Die Neuerungen im Vergleich zu Faktorenanalyse sind hierbei, dass bei der RSM nicht nur höhergeordnete Ableitungen betrachtet werden, sondern auch die Verwendung der Taylor Reihenapproximation. Des Weiteren werden alle Ableitungen bis zu einer höheren Ordnung untersucht. Die Idee des steilsten Anstiegs wird hierbei ebenso verwendet um eine Optimierung des Response zu ermitteln. Ein weiterer Unterschied zwischen der Faktorenanalyse und RSM beträgt die Zielsetzung dieser beiden Methoden: Strukturbeschreibung gegenüber Optimierung der Versuchsplanung. Während in der Faktorenanalyse erst die x_2 - Achse im dimensionalen Raum und danach die x_1 - Achse angepasst werden, wird in diesem Fall beide Richtungen im Raum gleichzeitig angepasst, was wiederum die Anzahl der Versuche verringert. Ebenso werden bei der faktorielle Analyse nur lineare Interaktionen zum Beispiel vom Umfang x_1x_2 sowie x_2x_3 etc. angewendet. Im Gegensatz dazu werden bei der RSM die Interaktionen nicht nur im linearen Maße sondern auch mehrdimensionale Interaktion beispielsweise $x_1^5x_2^3$ gebraucht. (Johnson and Kotz, 1991, vgl. S- 267-298)

2.3 Vorteile

Die Vorteile dieser neuen Herangehensweise sind kurz und knapp zusammengefasst: bessere Prognosefähigkeit unter bestimmten Bedingungen, anwendbar nicht nur auf die Chemiebranche sowie die Optimierung. Dieser Weg ist nicht nur schneller in der mehrdimensionalen Betrachtung durch die Möglichkeit der gleichzeitigen Veränderung von beispielsweise x_1 und x_2 , die bei der faktoriellen Betrachtung nicht möglich ist. (Johnson and Kotz, 1991, vgl. S. 267-308)

2.4 Vorgehensweise und Methodik

Im folgenden Abschnitt wird kurz auf die Vorgehensweise und Methodik bei der Faktorenanalyse eingegangen. Zuerst wird auf die experimentelle Strategie, die Größe des experimentellen Fehlers sowie die Komplexität der Response Surface und die sequentielle ableiten dieser, da die Ableitungen aufeinander aufbauen, abgehandelt. Des weiteren ist es wichtig die Regionen nicht zu sehr zu begrenzen, da bei einer Begrenzung vielleicht nur ein lokales Minimum oder Maximum zu finden ist, jedoch kein absolutes. Danach werden die sequentielle Prozesse mit dem Ziel auf höhere Responsevariablen anwenden zu könne. Die „One factor at a time“ (Johnson and Kotz, 1991, S. 271) - Methode von Friedmann und Savage wird hierbei vernachlässigt, da eine effizientere Methode benötigt wird. Diese neue Methode evaluiert von Box und Wilson führt zuerst die „Maximum Gains Method“ (Johnson and Kotz, 1991, S. 271) sowie die „Steepest Ascent“ (Johnson and Kotz, 1991,

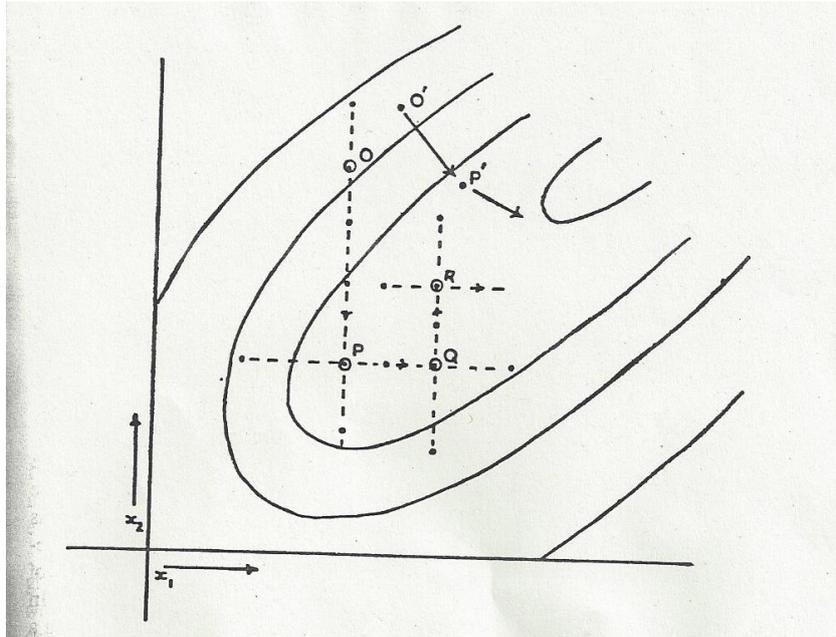


Abbildung 2.1:

Abbildung 2.2:

S. 271) Idee. Die „Maximum Gains Method“ (Johnson and Kotz, 1991, S. 271) wird auf der Grafik 1. dargestellt. Punkt O ist der Ursprung in einem k -dimensionalen Raum mit Distanz r von Punkt O zu Punkt P . Der Gewinn für den Response ist das Maximum. O wird als $\varphi(O)$ und $\varphi(P) = \varphi(x_1, \dots, x_k)$ von Punkt P . Die Distanz wird wie folgt beschrieben: $r^2 = \sum x_t^2 \rightarrow \varphi(P) - \varphi(O)$. Zudem ist die Skala nicht invariant, das heißt, dass die Distanz unter den Punkten jegliche Bedeutung haben kann. Praktisch gesehen wird um den Punkt O herum die Skala gelegt mit der Wahl der Faktorenlevels, die somit die Einheit der Skala festlegen. Als weitere Methode bzw. Idee wird die „Steepest Ascent Formula“, (Johnson and Kotz, 1991, S. 271) die den experimentellen Fehler und die Anzahl der Ableitungen reduzieren soll. (Johnson and Kotz, 1991, vgl. S. 272-273)

2.4.1 Allgemein

Allgemein betrachtet werden aufbauend auf die Methoden, die in Kapitel 2.4 erläutert wurden, die Response Surfaces gebildet. Mithilfe der Taylorreihenentwicklung werden bs basierend auf den geschätzten β s, die die Ableitung in die jeweilige Richtung, angegeben. η gibt dabei die Interaktionsterme an und ist ein Schätzer für die Steigung, die abgeleitet f ergeben. Dabei ist f die Schätzung für φ , die die wahren Ableitungen darstellen. Der Response η in dem Faktorraum an jedem Punkt $(x_1, \dots, x_t, \dots, x_k)$ wird durch die Regressionsgleichung angegeben:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{11} x_1^2 + \dots + \beta_{12} x_1 x_2 + \dots + \beta_{111} x_1^3 + \dots \quad (2.2)$$

(Johnson and Kotz, 1991, vgl. S. 270-308)

2.4.2 Rechenbeispiel

Nun wird abschließend noch ein Anwendungsbeispiel vorgeführt. Zur Veranschaulichung dieser Methode bei Anwendung chemischer Prozesse wird ein Versuch durchgeführt, wobei eine Substanz A mit einer Substanz B miteinander verbunden, so dass das gewünschte Produkt C entsteht sowie weitere Nebenprodukte.



Dieses Experiment ist beliebig oft reproduzierbar sowie die Ausbeute von C feststellbar. Das Ziel ist es an Minimum an Versuchen zu bestimmen, um eine für den Anwender optimale Ausbeute von Substanz C zu erhalten. Die Faktoren hierbei sind zum einem die Temperatur x_1 , die Konzentration x_2 und das molare Verhältnis x_3 von Substanz B zu Substanz A . Den Response stellt hierbei die errechneten Kosten pro Pfund von dem finalen Produkt C dar. Die Herstellungskosten y hierbei sollen minimiert werden. Als ein konkretes Beispiel werden für die

Temperatur in Celsius:	145
Konzentration in Prozent:	20
Molares Verhältnis von B/A:	6

vorgegeben. Es wird weiter festgelegt, dass die Konzentration nicht unter 18 Prozent fallen darf. Hierbei wird ein 2^3 faktorielles Design erstellt.

Nachfolgend wurde jede Variable in zwei Level mit 1 und -1 unterteilt. Auf unser Beispiel bezogen bedeutet das folgendes:

Temperatur in Celsius:	140 -1
	145 1
Konzentration in Prozent:	20 -1
	24 1
Molares Verhältnis von B/A:	6 -1
	7 1 .

Die Koordinaten der in diesem Design verschiedenen Möglichkeiten und den damit zusammenhängenden Kosten werden wie folgt bestimmt: (1,1,1) mit 17, (1,1,-1) mit 9, (1,-1,1) mit 12, (1,-1,-1) mit 15, (-1,1,-1) mit 24, (-1,1,-1) mit 11, (-1,-1,1) mit 7 und (-1,-1,-1) mit 5.

Die Effekte und ihre erwartenden Werte mit der Bedingung bis zur dritten Ordnung werden nachfolgend dargestellt. Zum Beispiel:

$$b_0 \rightarrow \varphi_0 + \frac{1}{2}\varphi_{11} + \frac{1}{2}\varphi_{22} + \frac{1}{2}\varphi_{33} = 12.50 \pm 0.4 \quad (2.4)$$

$$b_1 \rightarrow \varphi_1 \left(+\frac{1}{6}\varphi_{111} + \frac{1}{2}\varphi_{122} + \frac{1}{2}\varphi_{133} \right) = 0.75 \pm 0.4 \quad (2.5)$$

$$b_2 \rightarrow \varphi_2 \left(+\frac{1}{6}\varphi_{222} + \frac{1}{2}\varphi_{112} + \frac{1}{2}\varphi_{233} \right) = 2.75 \pm 0.4 \quad (2.6)$$

$$b_3 \rightarrow \varphi_3 \left(+\frac{1}{6}\varphi_{333} + \frac{1}{2}\varphi_{113} + \frac{1}{2}\varphi_{223} \right) = 2.50 \pm 0.4 \quad (2.7)$$

$$b_{12} \rightarrow \varphi_{12} = -3.00 \pm 0.4 \quad (2.8)$$

$$b_{13} \rightarrow \varphi_{13} = -1.25 \pm 0.4 \quad (2.9)$$

$$b_{23} \rightarrow \varphi_{23} = 2.75 \pm 0.4 \quad (2.10)$$

$$b_{123} \rightarrow \varphi_{123} = 0.00 \pm 0.4 \quad (2.11)$$

Hierbei beträgt die Streuungsvarianz 0.4. Nun wird die Methode des steilsten Abstiegs, da ein Minimum gesucht wird, verwendet und erweitert um alle Effekte der zweiten Ordnung. In diesem Beispiel sind die Gleichungen nicht zufriedenstellend, da die Effekte der ersten Ordnung für die Faktoren positiv sind, wobei ein Minimum gesucht wird. Folgend werden noch weitere zwei Experimente durchgeführt, die jedoch ui hohe Kosten aufweisen und nachfolgend noch weitere drei Versuche. Es wird die Matrix T' für ein zusammengesetztes Design mit (13×10) Eintragungen erstellt. Dreizehn Eintragungen, da dreizehn Experimente durchgeführt wurden sowie 10 Eintragungen, da $b_0, b_1, b_2, b_3, b_{11}, b_{22}, b_{23}, b_{12}, b_{13}$ und b_{23} . Nun werden die Gleichungen für b_0 und die quadratischen Effekte auf die Schätzungen angegeben:

$$b_0 \rightarrow \varphi_0 + \frac{1}{2}(\varphi_{111} + \varphi_{222} + \varphi_{333}) + \frac{1}{2}(\varphi_{122} + \varphi_{133} + \varphi_{112} + \varphi_{233} + \varphi_{113} + \varphi_{223}) + \frac{9}{8}\varphi = 6.750 \pm 0.8 \quad (2.12)$$

$$b_{11} \rightarrow \frac{1}{2}\varphi_{11} - \frac{1}{2}(\varphi_{111} + \varphi_{112} + \varphi_{113}) - \frac{3}{8}\varphi_{123} = 3.000 \pm 0.3 \quad (2.13)$$

$$b_{22} \rightarrow \frac{1}{2}\varphi_{22} - \frac{1}{2}(\varphi_{222} + \varphi_{122} + \varphi_{223}) - \frac{3}{8}\varphi_{123} = 1.625 \pm 0.3 \quad (2.14)$$

$$b_{33} \rightarrow \frac{1}{2}\varphi_{33} - \frac{1}{2}(\varphi_{333} + \varphi_{113} + \varphi_{233}) - \frac{3}{8}\varphi_{123} = 1.125 \pm 0.3 \quad (2.15)$$

. Unter der Annahme, dass die dritte und höhere Ordnung vernachlässigt werden, wird folgende Gleichung für den besten Fit, in der Umgebung des Designs, geschätzt:

$$y = 6.750 + 0.750x_1 + 2.750x_2 + 2.500x_3 + 3.000x_1^2 + 1.625x_2^2 + 1.125x_3^2 - 3.000x_1x_2 - 1.250x_1x_3 + 2.750x_2x_3 \quad (2.16)$$

Matrix C^{-1} mit den Effekten der Variablen von 0,1,2,3,11,22,33,12,13 und 23 als (10×10) - Matrix gibt die Korrelationen zwischen den Koeffizienten an. Nach einigen weiteren Versuchen, die Varianzen der dreizehn Experimente so umzustellen und umzuwandeln, dass es die Streuung am besten erklärt, siehe (Johnson and Kotz, 1991, S. 302-308), folgt eine neue Matrix C^{-1} . Mit dieser neuen Matrix folgt die Residuenquadratsumme mit drei Freiheitsgraden: $S_0 = 0.00 + 0.66 = 0.66$. Wenn der Fit perfekt ist, beträgt die Standardabweichung: $s = (0.66/3)^{\frac{1}{2}} = 0.47$. Falls die Art der gefitteten Surface geklärt werden soll, wird es auf die Standardform zurückgeführt und folgende Gleichung gelöst:

$\lambda^3 - 5.5141\lambda^2 + 4.9114\lambda + 0.9290 = 0$ mit folgenden Wurzeln $\lambda_1 = -0.1597$, $\lambda_2 = 1.3434$ und $\lambda_3 = 4.3305$. Danach werden die Koordinaten für eine Bestimmung des stationären Punktes S bestimmt. (Johnson and Kotz, 1991, vgl. S. 305-308) Nachdem die Variablen nach den Ergebnissen der vorherigen Analyse angepasst wurden, werden nun basierend auf fünf Freiheitsgraden die geschätzte Standardabweichung $s = (19.93/5)^{\frac{1}{2}} = 2.00$ geschätzt. Aus den neu berechneten $\lambda_1 = 0.9461$, $\lambda_2 = 1.4999$ und $\lambda_3 = 4.9113$ ergibt sich ein neues Zentrum $x_1^0 = -0.7601$, $x_2^0 = -0.1093$ und $x_3^0 = -0.4014$ mit den geschätzten Kosten $y^0 = 2.6719$. Das Endergebnis für das Minimum an Kosten von 3.2 ergab bei 11 df und Varianz von 1.7: für x_1 141°C, für x_2 20% und für die Konzentration und molares Verhältnis A/B mit x_3 6.2. (Johnson and Kotz, 1991, vgl. S. 298-308)

3 Auswirkungen dieser neuen Herangehensweise

Im Folgenden Kapitel werden im Kurzen die Auswirkungen dieser neuen Herangehensweise erläutert.

3.1 Statistik im Allgemeinen

Diese Neuerung im Bereich der Versuchsplanung brachte nicht nur eine weitere Methode sondern auch eine Alternative zu dem faktoriellen Design. Diese Optimierung und Weiterführung der Ideen (vgl. Kapitel 2) legte einen Grundstein für weitergehende Entwicklungen im Bereich der Statistik.

3.2 Fazit

Abschließend ist zu sagen, dass Box und Wilsons „On the Experimental Attainment of Optimum Conditions“ (Johnson and Kotz, 1991, S. 267) eine langanhaltende Wirkung auf den Bereich der Statistik hatte. Sie beeinflussten nicht nur die Art und Weise wie die Industrie und auch die Universitäten Experimente plant und durchführt sondern auch eine Weiterführung- und Entwicklung zu weiteren praktischen Designinnovationen sowie theoretischen Aspekten im Bereich der RSM.(Johnson and Kotz, 1991, vgl. S. 268)

Literaturverzeichnis

- Backhaus, K., B. Erichson, W. Plinke, and R. Weiber (2011). *Multivariate Analysemethoden - Eine anwendungsorientierte Einführung*, Volume 13. Springer-Verlag Berlin Heidelberg.
- Johnson, N. L. and S. Kotz (1991). *Breakthroughs in Statistics 1890-1989*, Volume 1. Springer New York.
- Kiesling, D. T. Versuchsplanung - Teil1 - Einführung und Grundlagen.
<http://www.statistik.lmu.de/~monecke/teaching/vp1112/vp-folien/Versuchsplanung-Teil1.pdf>.
- Koch, S. and C. Lindenlaub (2012, 04). Multivariate Verfahren - Mitschrift zur Vorlesung von Prof. Dr. Tutz.
<http://www.statistik.lmu.de/~fachsch/vorlesungsdateien/Multivariate12.pdf>.
- Kuß, A. (2012). *Marktforschung - Grundlagen der Datenerhebung und Datenanalyse*, Volume 4. Auflage. Springer Gabler Verlag.
- Ludwigs-Maximilians-Universität, M. (2013, 06). Modulhandbuch Statistik.
<http://www.uni-muenchen.de/studium/studienangebot/studiengaenge/studienfaecher/statistik/bachelor/modulhandbuch.pdf>.

Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Seminararbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Kraiburg, den 28.05.2014

(Natalie Grimm)