

Faktorenanalyse

Marcel Noack

12. April 2007

Inhaltsverzeichnis

1	Einleitung	3
1.1	Grundlagen	3
1.2	Vorüberlegungen	8
1.3	Probleme	10
2	Die Grundidee der Faktorenanalyse	14
2.1	Die Korrelation	14
2.2	Das Fundamentaltheorem	15
2.3	Pattern-Matrix und Structure-Matrix	17
2.4	Ein einfaches Beispiel	20
3	Methoden der Faktorextraktion	22
3.1	Principal Component Analysis / Hauptkomponentenanalyse	22
3.1.1	Dekompositionsstrategie	26
3.1.2	Beispiel	28
3.1.3	Extraktion weiterer Faktoren	31
3.2	“Die” Faktorenanalyse (Hauptachsenfaktorenanalyse)	32
3.3	Die Varianzen im Common-Factor Modell	32
3.4	Maximum Likelihood Faktorenanalyse	33
3.4.1	Erläuterung: Maximum Likelihood	35
3.4.2	Canonical Factoring	37
3.4.3	Maximale Determinante der Residual-Korrelations-Matrix	37
3.5	Least Squares Methoden	38
3.5.1	Erläuterung: Least Squares	38
3.5.2	Principal Axis Factoring mit iterativer Schätzung der Kommunalitäten	42
3.5.3	MinRes (Minimum Residuals)	42
3.6	Weitere Ansätze	42
3.6.1	α -Factoring	43
3.6.2	Image Analysis	43

4	Probleme der Faktorenanalyse	44
4.1	Das Kommunalitätenproblem	44
4.2	Anzahl der Faktoren	46
4.2.1	Das Kaiser-Kriterium	46
4.2.2	Der Scree-Test	46
4.2.3	Das kummulierte Varianzkriterium	47
4.3	Das Rotationsproblem	47
4.3.1	Die Einfachstruktur - Die Simple Structure	48
4.3.2	Orthogonale Rotation	51
4.3.3	Oblique Rotation	55
4.3.4	Überlegungen zur Rotation	59
4.4	Bestimmung der Faktorwerte	62
4.5	Interpretation der Faktoren	63
5	SPSS-Beispiele	65
5.1	Erziehung	65
5.2	Wichtigkeit	70
6	Appendix	75
6.1	Appendix A - χ^2	75
6.2	Appendix B - Matrix-Algebra	78
6.2.1	Skalarmultiplikation	78
6.2.2	Multiplikation	78
6.2.3	Addition und Subtraktion	81
6.2.4	Transponieren	82
6.2.5	Diagonalmatrizen	82
6.2.6	Die Spur einer Matrix	83
6.2.7	Determinante	84
6.2.8	Adjunkte	87
6.2.9	Inverse	88
6.2.10	Der Rang einer Matrix	92
6.2.11	Idempotente Matrix	93
6.2.12	Diverses	94
6.3	Appendix C - Determinanten	95
6.4	Appendix D - Eigenwerte	98

Kapitel 1

Einleitung

Faktorenanalyse? Was ist das? Darauf eine kurze, einfache Antwort zu geben ist kaum möglich. Dies liegt an einem einfachen Sachverhalt: es gibt “die” Faktorenanalyse gar nicht. Es handelt sich hierbei vielmehr um einen Oberbegriff verschiedener Verfahren und Ansätze, bei denen man wiederum an einer Vielzahl von Stellen unter mehreren Optionen wählen kann. Unterschieden wird zwischen der *konfirmatorischen Faktorenanalyse*, in der die Anzahl der Faktoren bereits als bekannt angenommen wird und darüber hinaus bestimmte strukturelle Annahmen getroffen werden, sowie der *explorativen Faktorenanalyse*, in der es darum geht, auf der Grundlage empirischer Daten möglichst wenige latente Faktoren zu entdecken, die die Korrelation zwischen den manifesten Variablen erklären. Im Folgenden werden wir uns genauer mit der explorativen Faktorenanalyse beschäftigen. Eines jedoch haben alle Verfahren, die sich unter dem Schlagwort “Faktorenanalyse” vereinigen lassen gemein: es handelt sich bei ihnen um Verfahren der Dimensionsreduktion. Durch ihre Anwendung ist es möglich, umfangreiches Datenmaterial auf wenige, einfachere Faktoren zu reduzieren. Wie dies im Einzelnen geschieht soll Thema dieses Skriptes sein.

1.1 Grundlagen

Das Hauptziel der Faktorenanalyse ist die Ableitung hypothetischer Größen aus einer Menge beobachteter Variablen. Hierbei kann es sich um gemeinsame Faktoren handeln, die mit mehreren Variablen in Beziehung stehen oder um spezifische Faktoren, die nur mit einer Variablen zu tun haben. In Graphik 1.1 sind diese beobachteten, manifesten Variablen mit x_i bezeichnet, wobei i die Werte 1 bis 5 annimmt. Die latenten Größen, die *Faktoren* kürzen wir mit ξ_i oder F_i ab. Die Korrelation zwischen einer Variable j und einem Faktor

k bezeichnen wir als Faktorladung a_{jk} . Die Ausprägungen der gemeinsamen Faktoren k bei den Merkmalsträgern i bezeichnet man als Faktorwerte, abgekürzt f_{ik} .

Nachfolgend einige Begrifflichkeiten und ihre Abkürzungen, die uns im Verlaufe dieses Skriptes begleiten werden:

$$\begin{array}{lcl}
 \text{Gesamte Varianz} & s_j^2 = 1 & = h_j^2 + b_j^2 + e_j^2 = h_j^2 + u_j^2 \\
 \text{Reliabilität} & r_{j\mathbf{J}} & = h_j^2 + b_j^2 = 1 - e_j^2 \\
 \text{Kommunalität} & h_j^2 & = h_j^2 = 1 - u_j^2 \\
 \text{Uniqueness} & u_j^2 & = b_j^2 + e_j^2 = 1 - h_j^2 \\
 \text{Specificity} & b_j^2 & = b_j^2 = u_j^2 - e_j^2 \\
 \text{Fehlervarianz} & e_j^2 & = e_j^2 = 1 - r_{jk}
 \end{array}$$

Die gesamte Varianz einer Variablen kann also in mehrere Elemente zerlegt werden. Wichtig ist das Konzept der Kommunalität. Wie wir später sehen werden, setzt sie sich in einem orthogonalem Modell aus der Summe der quadrierten Faktorladungen zusammen. Sie ist der durch die Faktoren erklärte Anteil an der gesamten Varianz. Die Uniqueness ist der Teil, der nicht durch die gemeinsamen Faktoren erklärt werden kann. Die Uniqueness lässt sich weiter unterteilen. Erstens in die Specificity, wobei es sich um den Teil der Uniqueness handelt, der sich nicht durch die gemeinsamen Faktoren erklären lässt. Hierfür ist die Variable selbst oder ein Faktor zuständig, auf dem nur diese spezifische Variable lädt. Dies kann daran liegen, dass erklärende Variablen fehlen, oder weil dieser Teil nur durch die Variable selbst zu erklären ist. Zweitens durch die Fehlervarianz, die sich auf die immer mehr oder weniger fehlerbehaftete Messung zurückführen lässt. Hier sei schon einmal auf einen wichtigen Unterschied in verschiedenen Verfahren hingewiesen.

- In der Hauptkomponentenanalyse (Principal Components Analysis) gehen wir von dem Modell $\mathbf{R} = \mathbf{ACA}'$ aus. Es wird angenommen, dass sich die Varianz der Variablen komplett durch die Faktoren erklären lassen, Annahmen über die Uniqueness werden nicht gemacht. Die Faktoren reproduzieren 100%.
- In der Hauptachsenanalyse wird das Modell um die Matrix \mathbf{U} erweitert, also $\mathbf{R} = \mathbf{ACA}' + \mathbf{U}$. Hier wird nicht davon ausgegangen, dass die Faktoren 100% erklären. Die Uniqueness kann nicht erklärt werden.

In der Graphik 1.1 sind implizit schon einige Annahmen des faktoranalytischen Modells für die Anfangslösung (die noch rotiert werden muss, dazu später mehr) dargestellt. In Anlehnung an folgende Graphik sind die Formeln rot eingefärbt, sofern sie sich auf eine negative, also verneinende Annahme des Modells beziehen.

$$\text{cov}(F_i, F_j) = 0 \quad \text{cov}(F_i, u_i) = 0 \quad \text{cov}(u_i, u_j) = 0$$

$$\text{cov}(F_i, x_i) = r_{F_i x_i} = b_i = \beta_i,$$

$$\text{cov}(x_i, x_j) = r_{x_i x_j} = b_i b_j = \beta_i \beta_j$$

Es werden also keine Kovarianzen zwischen den Faktoren, zwischen Faktoren und Uniqueness sowie zwischen den einzelnen Uniquenesses angenommen.

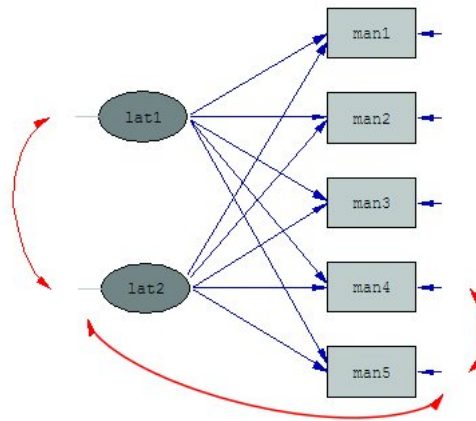


Abbildung 1.2: Annahmen

Diese Annahmen und Restriktionen gelten für die Anfangslösung. Bei der Anfangslösung handelt es sich um die erste Faktorladungsmatrix (die wir beispielsweise von einem Programm wie SPSS erhalten). Die Restriktion der unkorrelierten Faktoren kann in der Rotation durch ein obliques Rotationsverfahren, das uns die finale Lösung liefert, vom Forscher aufgegeben werden. Auch die Restriktion der unkorrelierten Fehlerterme kann beispielsweise in konfirmatorischen Faktorenanalysen vom Forscher aufgegeben werden, um die Anpassungsgüte (fit) des Modells zu verbessern.

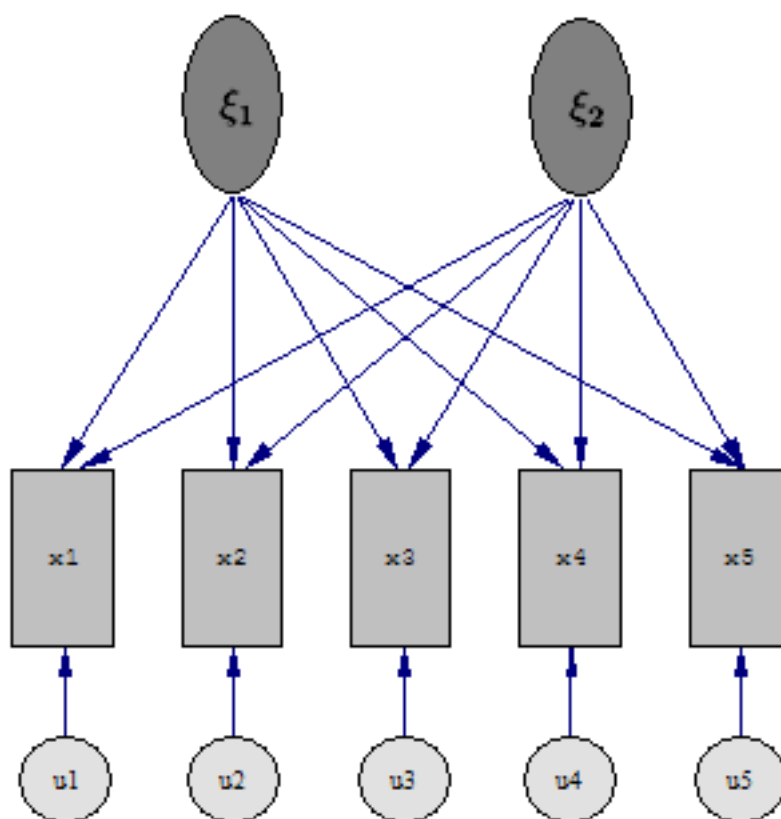


Abbildung 1.1: 2-Faktor 5-Variablen Modell

Faktorwerte und Faktorladung 1 (*Beispiel*)

Stellen wir uns Herrn Mustermann vor. Herr Mustermann soll -in Anlehnung an Graphik 1.1- einen Test bearbeiten, der aus 5 Teilaufgaben besteht. Von Beruf ist Herr Mustermann Übersetzer für 7 Sprachen und schreibt nebenher erfolgreich Romane und Kurzgeschichten. Sein Abitur hat er wegen Mathe und Physik nur mit Ach und Krach bestanden. Nun zu den Teilaufgaben. Es handelt sich hierbei um folgende Aufgaben:

1. Kurvendiskussion
2. Dreisatz
3. Räumliche Vorstellung an Hand der Lage von Würfeln im Raum
4. Aus 4 Worten das unpassende identifizieren
5. Synonyme Wortpaare nach Vorgabe bilden (Hund verhält sich zu bellen wie Auto zu ...)

Wir können uns denken, dass Herr Mustermann bei den ersten 3 Aufgaben bedeutend mehr Schwierigkeiten hat, als bei den letzten beiden. Sein sprachliches Talent hilft ihm nicht bei mathematischen Aufgaben, genauso wenig wie ihn seine geringes mathematisches Verständnis bei den sprachlichen Aufgaben behindert. Bei den **Faktorwerten** handelt es sich um die "Begabungen" von Herrn Mustermann, wenn wir sie so nennen wollen. Er zeichnet sich durch einen sehr hohen Wert auf dem ersten Faktor aus, der sich auf sprachliche Fähigkeiten bezieht. Auf dem zweiten Faktor hingegen, der mathematische Fertigkeiten anzeigt ist Herr Musstermanns Wert bestenfalls bescheiden. Bei den **Faktorladungen** handelt es sich um den Zusammenhang von Test und Faktor. Es ist einsichtig, dass Herr Musstermanns ganzes Sprachvermögen ihm nicht bei einer Kurvendiskussion hilft, weil dieser Test mit dem Faktor "sprachliche Begabung" nichts zu tun hat.

Zwischen den Beobachtungsdaten und den zugrunde liegenden Faktoren wird folgender linearer Zusammenhang angenommen:

$$\mathbf{Z} = \mathbf{F} \cdot \mathbf{A}'$$

wobei:

\mathbf{X} = Matrix der Ursprungsdaten

\mathbf{F} = Matrix der Faktorwerte

\mathbf{A}' = Transponierte ¹ Matrix der Faktorladungen

¹siehe Appendix

Den generellen Ablauf einer Faktorenanalyse sehen wir in Graphik 1.3

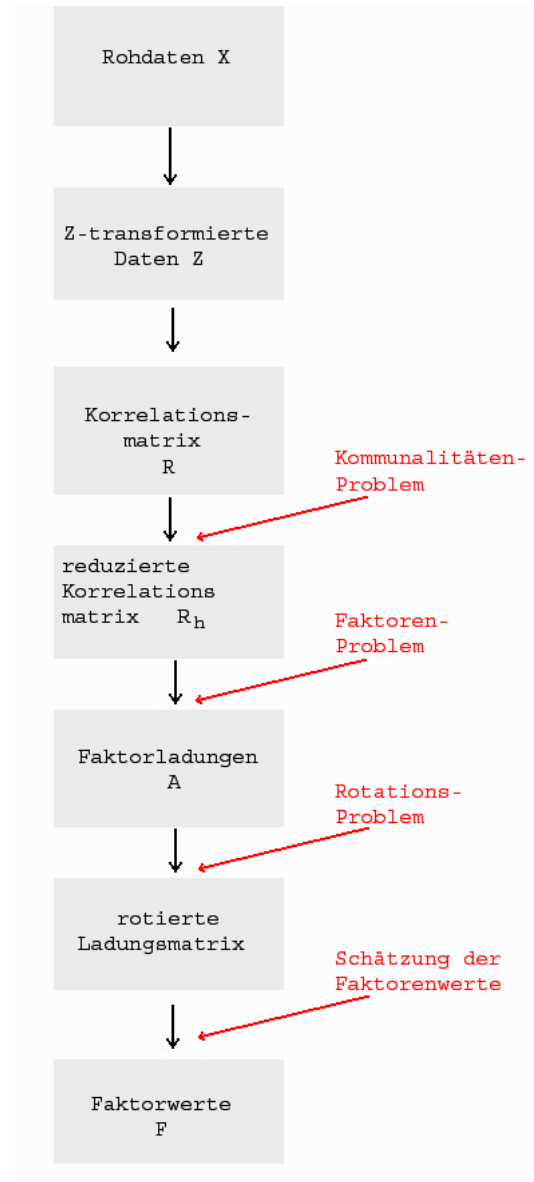


Abbildung 1.3: Ablauf der Faktorenanalyse

1.2 Vorüberlegungen

Bevor wir uns dem eigentlichen Thema dieses Skriptes widmen, ein paar Gedanken, die den Einstieg hoffentlich erleichtern. Wir sehen in Graphik 1.4

3 Vektorpaare, die jeweils 2 Variablen darstellen. Das Paar links bildet fast einen rechten Winkel (90°). Dies entspricht einer Korrelation von nahezu $r=0$. Die mittige und rechte Abbildung zeigen, wie sich die Lage, und damit auch die durch beide Vektoren aufgespannte Fläche verkleinert, wenn die Korrelation zunimmt. Das rechte Paar bildet einen Winkel von 0° , dies entspricht einer Korrelation von $r=1$.

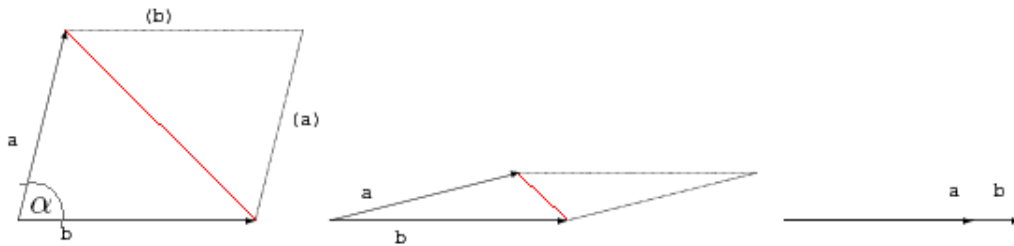
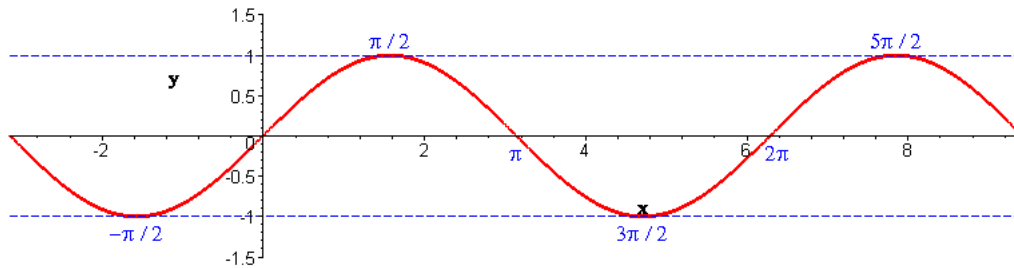


Abbildung 1.4: Flächeninhalt und lineare Abhängigkeit

Wir können an Hand dieser Graphik sehen, dass die Korrelation der Variablen den potentiell informationsbehaltenden Raum bis zu dem Extremfall verkleinert, dass beide Vektoren genau übereinanderliegen, und exakt die gleichen Informationen transportieren. Wie wir uns vielleicht erinnern, können wir den Flächeninhalt eines Vierecks als Flächeninhalt zweier Dreiecke verstehen. Also ist der Flächeninhalt des Vierecks identisch mit dem doppelten Flächeninhalt eines der Dreiecke. Den Flächeninhalt eines Dreiecks berechnen wir über die altbekannte Formel “ $1/2$ Grundseite \cdot Höhe”. Diese Formel kann überführt werden in “ $1/2|a| \cdot |b| \cdot \sin \alpha$ ”, wobei α den von a und b eingeschlossenen Winkel bezeichnet. Wenn wir der Einfachheit halber annehmen, dass a und b jeweils die Länge 1 haben, dann berechnet sich der Flächeninhalt des Vierecks über $2 \cdot 1/2 \cdot |a| \cdot |b| \cdot \sin \alpha$, wobei sich die Formel sehr stark vereinfacht, da $2 \cdot 1/2 = 1$ und a und b ebenfalls 1 sind. Es bleibt also auszurechnen: $2 \cdot 1/2 \cdot 1 \cdot 1 \cdot \sin \alpha$, also nur $\sin \alpha$ bleibt. Für uns ist nur der Bereich zwischen 0 und 2π von Interesse, da uns nur der Bereich zwischen 0° und 360° interessiert. Wie in Graphik 1.5 zu erkennen, wird $\sin \alpha$ maximal bei $\pi/2$.

Dies entspricht einem Winkel von 90° . Wir sehen hier, dass sich mathematisch zeigen lässt, dass der “Informationsraum”, den zwei Variablen aufspannen maximal ist, wenn beide Variablen unkorreliert sind. Wir könnten also eine Variable *weglassen*, ohne Informationen einzubüßen, wenn die Variablen maximal korreliert sind, also eine Korrelation von 1 aufweisen. Ebenso kann man sich jetzt schon überlegen, ob sich nicht eine gemeinsame Dimension mehrerer Variablen finden lässt, wenn sie wechselseitig hoch korreliert sind. Eine weiterführende Überlegung dieser Art wird im Appendix C behandelt.

Abbildung 1.5: $f(x) = \sin x$

1.3 Probleme

Ein Kausalsystem von Faktoren erzeugt immer eine eindeutige Korrelationsmatrix. Dies bedeutet aber nicht, dass man von einer Korrelationsmatrix eindeutig auf einen Faktor oder eine Struktur von Faktoren rückschließen kann. Zur Verdeutlichung dieses Sachverhalts kann man sagen, dass $7 \cdot 7$ ein eindeutiges Ergebnis hat, nämlich 49. $7 \cdot 7$ ist jedoch nicht die einzige Berechnung, die 49 zum Ergebnis hat. $50 - 1$, oder $\sqrt{2401}$ liefern ebenfalls 49 als Ergebnis. Generell können so viele Faktoren extrahiert werden, wie Variablen vorhanden sind. Da dies aber unter der Zielsetzung der Dimensionsreduktion keinen Sinn ergibt, müssen weniger Faktoren extrahiert werden als Variablen vorhanden sind. Aber wieviele? Zwei? Drei? Sechs? Hierfür gibt es Kriterien, die bei der Entscheidung helfen. Dazu später genaueres. Vorher treffen wir hier auf eine mathematische Besonderheit, der wir uns als Sozialwissenschaftler stellen müssen. Es liefern nämlich mehrere Modelle mathematisch äquivalente Ergebnisse, die jedoch theoretisch vollkommen unterschiedlich zu bewerten sind. Hier stellen sich uns Fragen, die nicht aus dem mathematisch-statistischen Modell heraus beantwortet werden können, sondern die soziologischen, psychologischen, politikwissenschaftlichen Theorien - oder in welcher Disziplin die Faktorenanalyse auch immer angewendet wird - (auf)fordern, ein Modell (1 oder 2 Faktoren? Korrelieren die Faktoren oder nicht?) als "das Richtige" zu identifizieren, und so die Anwendung zu legitimieren.

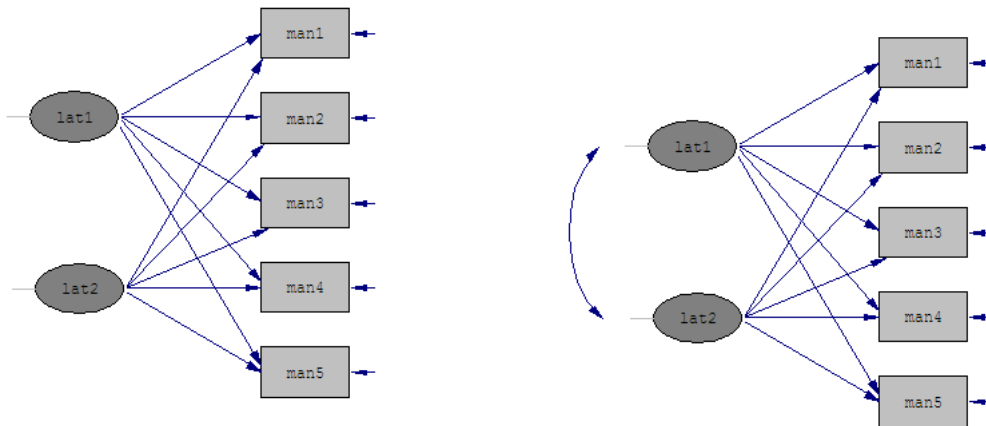


Abbildung 1.6: Problem der Faktorkorrelation

Ein Problem betrifft die Ungewissheit, ob die Faktoren (hier mit lat1 und lat2 als Abkürzung für “latente Variable”) korrelieren oder nicht. Je nachdem, ob die Faktoren korrelieren (hier mittels gebogener Linie dargestellt) ändern sich die Ladungen der Faktoren auf die von ihnen beeinflussten manifesten Variablen (hier mit man1 bis man5 benannt). Die Frage nach der Korrelation der Faktoren kann nicht aus dem Modell abgeleitet beantwortet werden, sondern es muss die Theorie des anwendenden Fachs zur Beantwortung der Frage herangezogen werden. Generell wird davor gewarnt, einfach “eine handvoll Variablen in die Faktorenanalyse zu werfen” und zu hoffen, dass die Faktoren “schon interpretierbar” sein werden. So schreibt Karl Überla:

Die Faktorenanalyse stellt also die Frage, welches die einfachste Struktur ist, die die Daten genügend genau reproduziert und erklärt. Damit ist sie entscheidend von der Art und der Anzahl der vorliegenden Daten abhängig. Eine einseitige Auswahl der Variablen muss notwendigerweise andere Faktoren ergeben oder in den Vordergrund stellen als breiter angelegte Studien. Das Ergebnis der Faktorenanalyse ist von der Anlage der ganzen Untersuchung bestimmt. Es kann nicht genug davor gewarnt werden, planlos gesammelte Daten einer Faktorenanalyse zu unterwerfen und zu hoffen, dass ein vorhandenes Standardprogramm sinnvolle Resultate bringt.

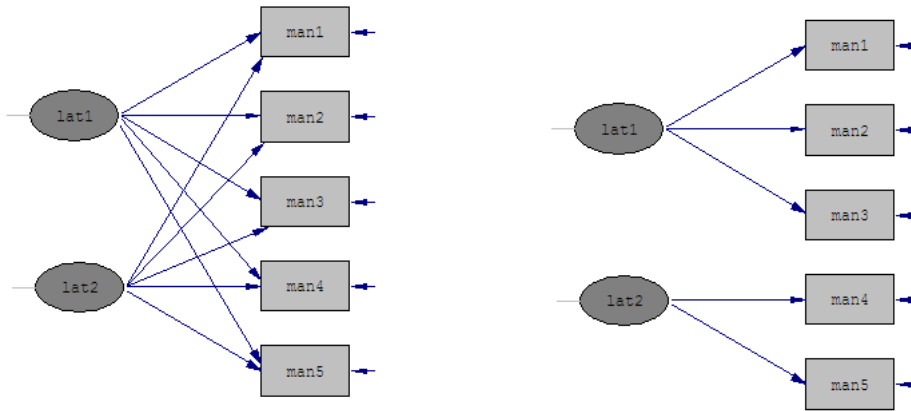


Abbildung 1.7: Problem der Faktorladungen

Ebenfalls stellt sich die Frage nach den einzelnen Faktorladungen. Eine Faktorladung bezeichnet die Korrelation zwischen einem Faktor und einer standardisierten Originalvariablen. So ist bei gleicher Faktorzahl und gleicher Faktorkorrelation mehr als ein Modell denkbar, das den Daten gleich gut angepasst ist, obwohl sich die Faktorladungen unterscheiden. So stellt sich zum Beispiel die Frage, ob es Fremdladungen auf Variablen einer nicht zum Faktor gehörenden Itematterie gibt, oder nicht.

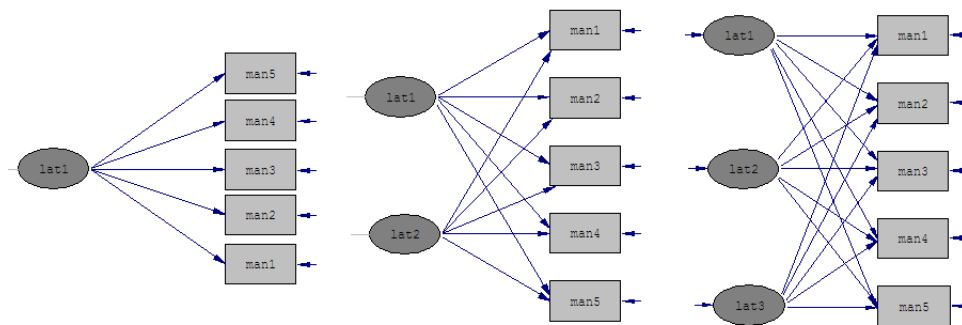


Abbildung 1.8: Problem der Faktorzahl

Auch die Anzahl der Faktoren ist nicht determiniert. Es existieren mehrere -mathematisch- "gleichgute" Modelle, die sich in ihrer Faktorzahl unterscheiden. Vielleicht sieht man hier am deutlichsten, dass es die fachspezifische Theorie ist, die den Ausschlag für die Wahl eines Modells geben muss. So kann -und muss- z.B. der Psychologe aus der Psychologie -und nicht aus der

Mathematik- entscheiden, ob er die Intelligenzleistung durch ein Modell mit zwei, fünf oder X zugrunde liegenden latenten Faktoren erklärt.

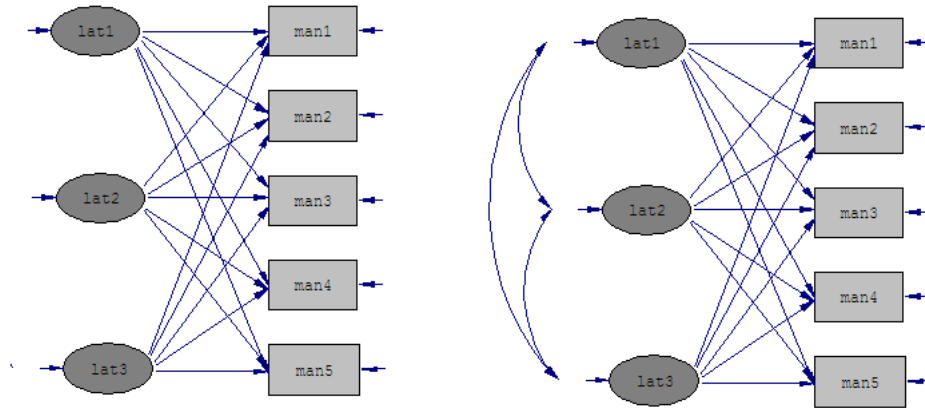


Abbildung 1.9: Kombination der Probleme

Als ob dies alles nicht schon unerfreulich genug wäre muss man sich natürlich vor Augen halten, dass die Probleme nicht unabhängig voneinander existieren, also in Kombination auftreten. Diese ganzen Probleme lassen sich aber leichter lösen, als es nach diesem Abschnitt den Anschein hat. Eine gute Theorie - im Vorfeld recherchiert oder ausgearbeitet - liefert Hinweise und Begründungen, welches Modell als passend angesehen werden kann.

Kapitel 2

Die Grundidee der Faktorenanalyse

Ausgangspunkt einer jeden Faktorenanalyse ist die Korrelationsmatrix \mathbf{R} . In ihr sind alle Korrelationen zwischen den beobachteten, manifesten Variablen festgehalten. Zur Generierung einer Korrelationsmatrix bietet es sich an, die beobachteten Daten zunächst zu standardisieren. Durch eine Z-Transformation erhält die standardisierte Variable einen neuen Mittelwert von 0 und eine Varianz von 1. Dies geschieht über folgende Formel:

2.1 Die Korrelation

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

Die Korrelation zweier Variablen errechnet sich über folgende Formel

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

oder ausführlich geschrieben:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Die Vereinfachung der Berechnung der Korrelation durch Z-transformierte Variablen sieht so aus: Bei so standardisierten Variablen gilt $\bar{x} = 0$ und

$s = s^2 = 1$. Es vereinfacht sich

$$r = \frac{s_{xy}}{s_x \cdot s_y}$$

zu

$$r = \frac{z_{xy}}{1 \cdot 1}$$

also bleibt nur die Z-transformierte Kovarianz übrig, hier mit z_{xy} bezeichnet. Die unstandardisierte Kovarianz

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

vereinfacht sich nach Standardisierung, da in beide Variablen $\bar{y}, \bar{x} = 0$ gilt zu

$$\sum_{i=1}^n (x_i^z - 0)(y_i^z - 0) = \sum_{i=1}^n (x_i^z)(y_i^z)$$

wobei der Exponent z anzeigen soll, dass die Daten Z-transformiert sind. Wir sehen, dass das Produkt zweier Z-transformierter Variablen identisch ist mit der Korrelation der ursprünglichen, unstandardisierten Daten.

2.2 Das Fundamentaltheorem

Der Kern der Faktorenanalyse ist das so genannte Fundamentaltheorem. Es besagt, dass sich die Korrelationsmatrix \mathbf{R} durch die Matrizen der Faktorenwerte und Faktorladungen reproduzieren lässt. Die Grundgleichung des Fundamentaltheorems der Faktorenanalyse lautet:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}'$$

Dabei sind mit den fetten Buchstaben jeweils unterschiedliche Matrizen bezeichnet.

- \mathbf{Z} =Matrix der standardisierten Daten
- \mathbf{F} =Matrix der Faktorenwerte
- \mathbf{A} =Matrix der Faktorenladungen
- \mathbf{R} =Korrelationsmatrix

Folgende Gleichung ist für uns von Interesse. Sie ist der Ansatzpunkt, von dem aus wir Schritt für Schritt fortfahren. Die Gleichung der Korrelationsmatrix.

$$\mathbf{R} = \frac{1}{n-1} \mathbf{Z}' \mathbf{Z}$$

Durch Substitution von $(\mathbf{F} \mathbf{A}') = \mathbf{Z}$, was sich aus der Grundgleichung erklärt, ergibt sich:

$$\mathbf{R} = \frac{1}{n-1} (\mathbf{F} \mathbf{A}')' (\mathbf{F} \mathbf{A}')$$

Aus den Rechenregeln für inverse Matrizen ist bekannt, dass: $(\mathbf{A} \mathbf{B})' = (\mathbf{B}' \mathbf{A}')$, also gilt auch $(\mathbf{A}' \mathbf{F})' = (\mathbf{F}' \mathbf{A})$, was wir umgehend einsetzen.

$$\mathbf{R} = \frac{1}{n-1} (\mathbf{A} \mathbf{F}') (\mathbf{F} \mathbf{A}')$$

Im nächsten Schritt lassen wir einfach die Klammern weg.

$$\mathbf{R} = \frac{1}{n-1} \mathbf{A} \mathbf{F}' \mathbf{F} \mathbf{A}'$$

Da es sich bei $\frac{1}{n-1}$ um einen Skalar handelt, ist es erlaubt, ihn an eine andere Stelle zu setzen.

$$\mathbf{R} = \mathbf{A} \frac{1}{n-1} \mathbf{F}' \mathbf{F} \mathbf{A}'$$

An dieser Stelle erkennen wir eine Struktur wieder: so wie in der zweiten Formel $\mathbf{R} = \frac{1}{n-1} \mathbf{Z}' \mathbf{Z}$ die Korrelationsmatrix der Daten beschreibt, so beschreibt auch $\frac{1}{n-1} \mathbf{F}' \mathbf{F}$ eine Korrelationsmatrix. Allerdings handelt es sich hierbei um die Korrelationsmatrix der Faktoren. Setzen wir $\frac{1}{n-1} \mathbf{F}' \mathbf{F} = \mathbf{C}$ führt uns das zu folgender Formel:

$$\mathbf{R} = \mathbf{A} \mathbf{C} \mathbf{A}'$$

Wenn die Faktoren unkorreliert sind handelt es sich bei der Korrelationsmatrix der Faktoren um die Identitätsmatrix, also

$$\mathbf{C} = \mathbf{I} :$$

In diesem Fall vereinfacht sich die Formel weiter, und zwar zu:

$$\mathbf{R} = \mathbf{ACA}' = \mathbf{AIA}' = \mathbf{AA}'$$

Bei unkorrelierten Faktoren lässt sich also die Korrelationsmatrix als Produkt der Matrix der Faktorladung \mathbf{A} und der transponierten Matrix der Faktorladung \mathbf{A}' errechnen. Wenn die Faktoren nicht orthogonal sind kommt die Korrelationsmatrix der Faktoren \mathbf{C} hinzu. Wie wir \mathbf{A} bestimmen, und somit die Faktorladungen werden wir in späteren Abschnitten sehen, in denen es um die Methoden der Faktorextraktion geht.

2.3 Pattern-Matrix und Structure-Matrix

Generell gilt es zwischen der Pattern-Matrix und der Structure-Matrix zu unterscheiden. In der *Structure-Matrix* befinden sich die Korrelationen zwischen Faktoren und Variablen. In der *Pattern-Matrix* sind die standardisierten linearen Regressionskoeffizienten eingetragen. Wenn die Faktoren orthogonal sind, also unkorreliert, dann sind Pattern und Structure Matrix identisch. Wenn es sich jedoch um oblique, also korrelierende Faktoren handelt, dann sind Pattern und Structure Matrix verschieden. Es ist möglich, dass bei einem Vergleich zwischen einem orthogonalen und einem obliquen Modell beide Pattern-Matrizen identisch sind, dies gilt jedoch nicht für die Structure-Matrizen! In folgender Graphik ist die Berechnung der Korrelation zwischen x_1 und x_2 verdeutlicht. Hier mag man stutzen. Wieso sind in diesem Fall Regressionskoeffizienten und Korrelationskoeffizienten identisch? Wir gehen davon aus, dass die Ursprungsdaten Z-transformiert werden. Die Regressionskoeffizienten sind dementsprechend β_i -Koeffizienten, auch die Korrelation ist standardisiert. Da in diesem Fall $\bar{x} = 0$ und $s = s^2 = 1$ vereinfacht sich,

$$r = \frac{s_{xy}}{s_x s_y}, \quad b = \frac{s_{xy}}{s_x^2}$$

zu

$$r = \frac{z_{xy}}{1 \cdot 1}, \quad b = \frac{z_{xy}}{1^2}$$

$$r = b = s_{xy}$$

Wobei mit z_{xy} die Z-transformierte Kovarianz bezeichnet ist. Was ist nun wieder die Z-transformierte Kovarianz? Man kann nachweisen, dass $r = \beta$. Wir haben gesehen, dass

$$r_{xy} \text{ und } \beta = z_{xy}.$$

Wenn also $r = z_{xy}$ gelten soll, muss $\frac{s_{xy}}{s_x s_y} = z_{xy}$ gelten, wobei der Exponent z anzeigen soll, dass die Variable Z-transformiert ist.

Wir wollen beweisen, dass folgende Gleichung wahr ist:

$$r_{xy} = \beta$$

Dazu schreiben wir beide Formeln ausführlich:

$$\frac{s_{xy}}{s_x s_y} = z_{xy}$$

Wir schreiben die Z-transformierte Kovarianz z_{xy} ausführlich aus, wobei der Exponent z anzeigt, dass die Variablen Z-transformiert sind. Er bezieht sich also nicht wie sonst üblich auf eine Berechnungsvorschrift.

$$\frac{s_{xy}}{s_x s_y} = \sum \frac{[(x_i^z - \bar{x}^z)(y_i^z - \bar{y}^z)]}{n}$$

Da die Mittelwerte Z-transformierter Variablen gleich Null sind vereinfacht sich der Term zu:

$$\frac{s_{xy}}{s_x s_y} = \sum \frac{[(x_i^z - 0)(y_i^z - 0)]}{n}$$

Wir lassen -0 in der nächsten Zeile wegfallen und multiplizieren auf beiden Seiten mit $s_x s_y$, wodurch der Bruch auf der linken Seite verschwindet.

$$s_{xy} = \sum \frac{[(x_i^z)(y_i^z)]}{n} s_x s_y$$

Nun stellen wir $s_x s_y$ vor die Summe, um eine bessere Übersicht zu erhalten.

$$s_{xy} = s_x s_y \frac{\sum [(\frac{x_i - \bar{x}}{s_x})(\frac{y_i - \bar{y}}{s_y})]}{n}$$

Da es sich bei $s_x s_y$ um eine Konstante handelt, sind wir berechtigt, sie in die Summe zu schreiben.

$$s_{xy} = \sum \frac{s_x s_y [(\frac{x_i - \bar{x}}{s_x})(\frac{y_i - \bar{y}}{s_y})]}{n}$$

Nun stellen wir in der Summe die Werte so um, dass wir auf einen Blick sehen, dass wir jeweils $\frac{s_x}{s_x}$ und $\frac{s_y}{s_y}$ kürzen können.

$$s_{xy} = \sum \frac{\left(\frac{s_x(x_i - \bar{x})}{s_x}\right)\left(\frac{s_y(y_i - \bar{y})}{s_y}\right)}{n}$$

Nachdem wir nun gekürzt haben, erkennen wir unsere altbekannte Formel für die Kovarianz. Die Aussage ist also wahr.

$$s_{xy} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n}$$

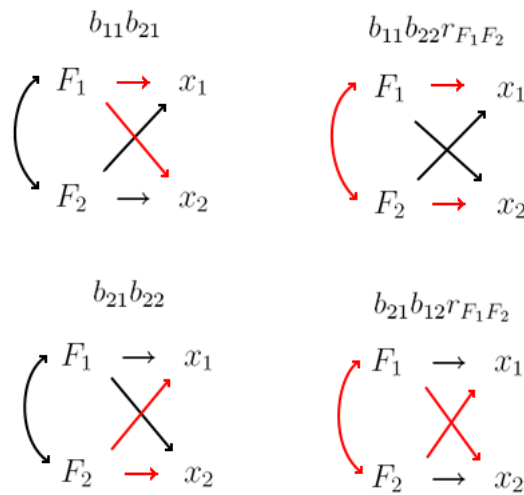


Abbildung 2.1: Korrelation r_{12}

Es gilt: $r_{12} = b_{11}b_{21} + b_{21}b_{22} + b_{11}b_{22}r_{F_1F_2} + b_{21}b_{12}r_{F_1F_2}$. An diesem Beispiel können wir sehen, worin hier der Unterschied zwischen r und β liegt. Im Fall von korrelierenden Faktoren sind nicht nur die “Wege” über einen Faktor möglich, sondern auch “Umwege” über die Korrelation der Faktoren. Diese Umwege sind für den Unterschied zwischen der Pattern Matrix und der Structure Matrix im obliquen Fall verantwortlich. Im orthogonalen Fall wären die beiden rechten Abbildungen mit den rot markierten Korrelationen nicht vorhanden, da die Korrelation schlicht und ergreifend wegfällt.

2.4 Ein einfaches Beispiel

Wir betrachten eine beliebige Korrelationsmatrix \mathbf{R} :

$$\mathbf{R} = \begin{pmatrix} 1 & 0.7225 & 0.085 & 0.08 \\ 0.7225 & 1 & 0.08 & 0.075 \\ 0.085 & 0.08 & 1 & 0.5625 \\ 0.08 & 0.075 & 0.5625 & 1 \end{pmatrix}$$

Wir werden nun die in der Hauptdiagonalen befindlichen Selbstkorrelationen, also alle Einsen durch die zugehörigen Kommunalitätenschätzer ersetzen (siehe S.44). Die reduzierte Korrelationsmatrix bezeichnen wir mit \mathbf{R}_h . Näheres dazu in dem entsprechenden Kapitel.

$$\mathbf{R}_h = \begin{pmatrix} 0.8125 & 0.7225 & 0.085 & 0.08 \\ 0.7225 & 0.6425 & 0.08 & 0.075 \\ 0.085 & 0.08 & 0.6425 & 0.5625 \\ 0.08 & 0.075 & 0.5625 & 0.4925 \end{pmatrix}$$

Wir suchen nun einen Vektor (Faktor), der mit sich selbst multipliziert diese Matrix möglichst genau reproduziert (\mathbf{R}^+). Wie man an solch einen Faktor gelangt, soll auch späteren Kapiteln vorbehalten sein.

$$\mathbf{R}^+ \approx \mathbf{a}_1 \mathbf{a}'_1 = \begin{pmatrix} 0.9 \\ 0.8 \\ 0.05 \\ 0.05 \end{pmatrix} (0.9 \ 0.8 \ 0.05 \ 0.05) = \begin{pmatrix} 0.81 & 0.72 & 0.045 & 0.045 \\ 0.72 & 0.64 & 0.04 & 0.04 \\ 0.045 & 0.04 & 0.0025 & 0.0025 \\ 0.045 & 0.04 & 0.0025 & 0.0025 \end{pmatrix}$$

Die reproduzierte Korrelationsmatrix ist nicht ausreichend gut. Wir benötigen also einen weiteren Faktor. Nun suchen wir einen Faktor a_2 , der mit sich selbst multipliziert den Rest der Korrelationsmatrix, die Residualmatrix \mathbf{R}_1 gut reproduziert.

$$\mathbf{R}_1 = \mathbf{R}_h - \mathbf{R}^+ = \begin{pmatrix} 0.0025 & 0.025 & 0.04 & 0.035 \\ 0.0025 & 0.025 & 0.04 & 0.035 \\ 0.04 & 0.04 & 0.64 & 0.56 \\ 0.035 & 0.035 & 0.56 & 0.49 \end{pmatrix}$$

Die Residualmatrix lässt sich restlos durch den Faktor a_2 aufklären. Es gilt:

$$\mathbf{R}_1 = \mathbf{a}_2 \mathbf{a}'_2 = \begin{pmatrix} 0.05 \\ 0.05 \\ 0.08 \\ 0.07 \end{pmatrix} (0.05 \ 0.05 \ 0.8 \ 0.7) = \begin{pmatrix} 0.0025 & 0.025 & 0.04 & 0.035 \\ 0.0025 & 0.025 & 0.04 & 0.035 \\ 0.04 & 0.04 & 0.64 & 0.56 \\ 0.035 & 0.035 & 0.56 & 0.49 \end{pmatrix}$$

Man kann das gleiche Modell auch in einer Gleichung schreiben:

$$\mathbf{R}_h = \begin{pmatrix} 0.8125 & 0.7225 & 0.085 & 0.08 \\ 0.7225 & 0.6425 & 0.08 & 0.075 \\ 0.085 & 0.08 & 0.6425 & 0.5625 \\ 0.08 & 0.075 & 0.5625 & 0.4925 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.05 \\ 0.8 & 0.05 \\ 0.05 & 0.8 \\ 0.05 & 0.7 \end{pmatrix} \begin{pmatrix} 0.9 & 0.8 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.8 & 0.7 \end{pmatrix}$$

Wir benötigen also zwei Faktoren, um die Korrelationsmatrix zu reproduzieren. Die Matrix \mathbf{A} enthält das Faktormuster (Factor-Pattern), ihre Elemente sind die Faktorladungen.

Es ergibt sich :

$$\mathbf{A} = (\mathbf{a}_1 \mid \mathbf{a}_2) = \begin{pmatrix} 0.9 & 0.05 \\ 0.8 & 0.05 \\ 0.05 & 0.8 \\ 0.05 & 0.7 \end{pmatrix}$$

Hier haben wir an Hand eines kurzen Beispiels gesehen, was die Intention der Faktorenanalyse ist. Die 4×4 -Matrix der ursprünglichen Variablen ist durch 2 Faktoren voll reproduzierbar. Wir haben die Dimension von 4 Variablen auf 2 Faktoren reduziert.

Kapitel 3

Methoden der Faktorextraktion

Hier wird eine Vielzahl unterschiedlicher Ansätze der Faktorextraktion vorgestellt. Beginnen wollen wir mit der Hauptkomponentenanalyse, auch Principal Component Analysis genannt. Auf Grund einiger Besonderheiten, auf die später eingegangen wird, wird sie oft als eigenständiges Verfahren angesehen. So ist beispielsweise im Statistik-Programm SPSS “Principal Components” eine Option in der Faktorextraktion, die als Bestandteil der Faktorenanalyse ausgewählt werden kann. Im Statistik-Programm Stata jedoch ist die Principal Components Analysis eine gleichberechtigte Option neben der Faktorenanalyse und keine Unteroption der Faktorextraktion. Andere Ansätze lassen sich unter Oberbegriffen zusammenfassen, die einem auch in anderen Kontexten über den Weg laufen können. So lassen sich die Ansätze der *Common-Factor-Modelle* erstens unter den *Least-Squares-Methoden* beschreiben, welche aus der linearen Regression bekannt sein dürfte, und den Methoden, die auf dem Gedanken der *Maximum-Likelihood-Schätzung* aufbauen, die unter anderem auch im Zusammenhang mit der logistischen Regression eine wichtige Rolle spielt. Als dritte Kategorie gibt es noch die “anderen Typen”, die als Gemeinsamkeit die Zielrichtung der Dimensionsreduktion haben, aber nicht genug Ähnlichkeiten vereinen, um eine so enge Einteilung zu ermöglichen wie die ersten beiden Kategorien. In dieser dritten Gruppe befassen wir uns mit der Image Analysis, dem α -Factoring und der schon angesprochenen Principal Component Analysis.

3.1 Principal Component Analysis / Hauptkomponentenanalyse

Die Hauptkomponentenanalyse ist eine Methode, ein bestehendes Set von Variablen in ein anderes Set zu transformieren. Es wird versucht, die Originalva-

riablen durch eine kleinere Anzahl “dahinterliegender Variablen” zu ersetzen. Diese dahinterliegenden Variablen -Hauptkomponenten genannt- sind Linearkombinationen der Ausgangsdaten. Das Verfahren der Hauptkomponentenanalyse besteht nun darin, durch eine orthogonale (rechtwinkelige) Transformation der Ursprungsvariablen eine neue Menge unkorrelierter Variablen zu erhalten. Dies geschieht über einen Ansatz, der an die Least-Squares Regression erinnert, die aus der bivariaten Statistik bekannt sein dürfte. Dort wird die Distanz zwischen Wert und vorhergesagtem Wert minimiert.

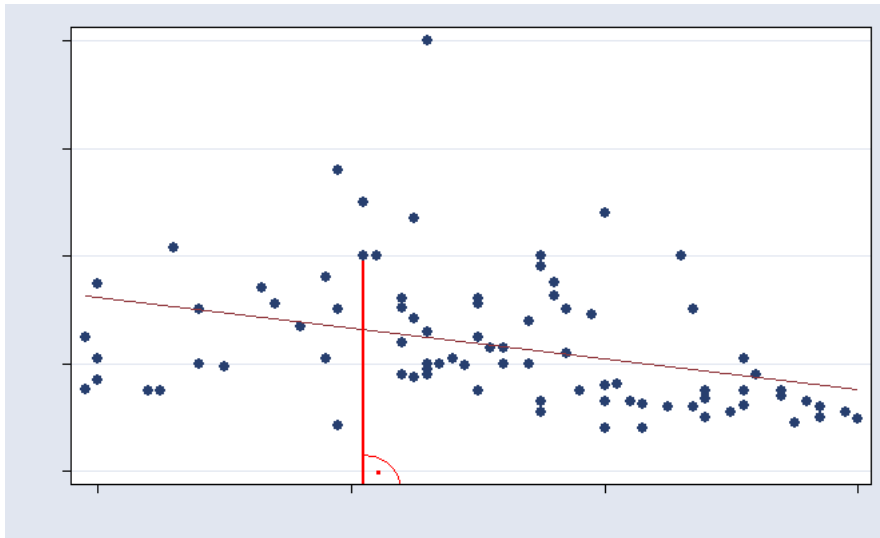


Abbildung 3.1: Regression

Diese Distanz wird rechtwinkelig zur X -Achse, also parallel zur Y -Achse gemessen. Zur Veranschaulichung wird dies an Hand eines Datenpunktes in Graphik 3.1 dargestellt. Es wird also der Wert $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ minimiert (hier wird der quadrierte “Fehler” minimiert, daher der Name des Verfahrens *kleinsten Fehlerquadrate*, oder auf Englisch: *least squares*), wobei y_i der gemessene und \hat{y}_i der durch die Regressionsgerade vorhergesagte Wert ist. Die Principal Axis findet man auf eine ähnliche, jedoch nicht identische Weise. Hier wird auch eine Distanz zwischen einer Geraden und den Datenpunkten minimiert, jedoch ist diesmal die Distanz nicht rechtwinkelig zur X -Achse, sondern rechtwinkelig zur Principal Axis selbst. In Graphik 3.2 ist zur Verdeutlichung eine solche Distanz markiert, auch wenn es sich bei der Geraden eigentlich um die Regressionsgerade handelt. Um den Unterschied zwischen den beiden Ansätzen zu verstehen ist dies jedoch unerheblich. Im multivariaten Fall wird nicht nur eine Principal Axis “gefunden”, sondern mehrere. Diese Achsen sind absteigend in ihrer Bedeutung und wechselseitig

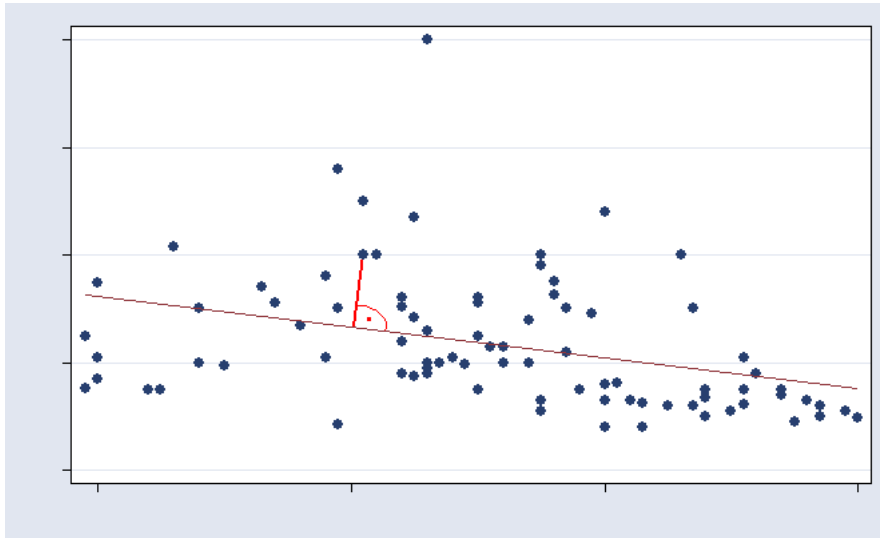


Abbildung 3.2: Principal Axis

unabhängig, also stehen alle rechtwinkelig aufeinander und erklären sukzessiv maximale Varianz. Bekannte Beispiele für solche wechselseitig rechtwinkligen Gebilde sind die karthesischen Koordinatensysteme im \mathbb{R}^2 und \mathbb{R}^3 , welches in Graphik 3.3 abgebildet ist. Diese Eigenschaften Principal Axes bedeuten nun folgendes:

- Die erste Achse erklärt maximale Varianz. Keine andere Achse erklärt mehr!
- Die zweite Achse steht rechtwinkelig auf der ersten Achse und erklärt weniger Varianz als diese, aber mehr als die Dritte.
- Die dritte Achse steht rechtwinkelig auf den ersten beiden Achsen und erklärt weniger Varianz als die erste und zweite Achse.
- Die vierte Achse steht rechtwinkelig auf allen drei Achsen und erklärt weniger Varianz als die dritte Achse, also auch weniger als die erste und die zweite Achse.
- Und so weiter...

Man erhofft sich davon, dass einige wenige der ersten so erhaltenen, dahinterstehenden Variablen für den größten Teil der Variation der Ursprungsvariablen verantwortlich sind. So erhalten wir eine effektive Reduktion der Dimensionalität. Eine graphische Darstellung der ersten beiden Hauptkomponenten in einem Koordinatensystem kann hilfreich sein, um Cluster in den

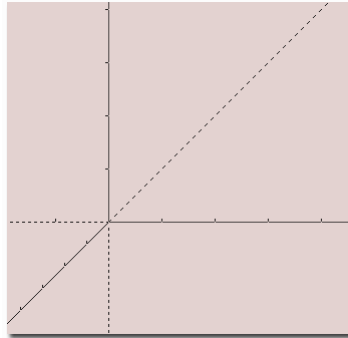


Abbildung 3.3: Dreidimensionales kartesisches Koordinatensystem

Daten zu finden. Dies ist jedoch nur sinnvoll, wenn die ursprünglichen Variablen hoch korreliert sind. Da wir durch unsere Transformation ein unkorreliertes Set neuer Variablen erhalten, ist es natürlich äußerst empfehlenswert, vorher zu überprüfen, ob die Ursprungsvariablen an sich korreliert sind, oder nicht. Liegen keine signifikanten Korrelationen vor, ist eine Hauptkomponentenanalyse überflüssig. Wir würden nur Hauptkomponenten erhalten, die den Ursprungsdaten sehr ähnlich sind, bis zum Extremfall der Einheitsmatrix, in der die Hauptkomponenten und die Ursprungsvariablen identisch sind. Das Einzige was sich ändern würde, wäre die Ordnung der Variablen, da sie nun nicht mehr z.B. alphabetisch geordnet wären, sondern nach absteigender Bedeutung. Die Hauptkomponentenanalyse ist ein mathematisches Verfahren, bei dem kein statistisches Modell zur Erklärung der Fehlerstruktur verlangt wird. Es werden auch keine Annahmen über die Verteilungsform der Ursprungsvariablen gemacht. Trotzdem ist die Interpretation der Hauptkomponenten einfacher, wenn eine multivariate Normalverteilung vorausgesetzt werden kann.

Indem wir die Principal Axes herausarbeiten, müssen wir keine hypothetischen Faktoren annehmen. Die neuen Achsen sind mathematische Funktionen der beobachteten Variablen. Die Hauptkomponentenmethode versucht *nicht* die *Korrelation* bzw. die *Kovarianz* der Variablen zu erklären wie es die Faktorenanalyse, namentlich die *Hauptachsenmethode* tut, sondern vielmehr über so viel *Varianz* wie möglich Rechenschaft abzulegen. Das primäre mathematische Werkzeug, mit dem solch eine hierarchische Dekomposition oder Transformation erreicht werden kann, wird “characteristic equation” oder “eigen equation” genannt.

Die Ähnlichkeit der Principal Component Analysis zur Faktorenanalyse liegt darin, dass beide Verfahren zur Dimensionsreduktion eingesetzt werden. Eben-

so sind beide Verfahren dazu geeignet, Interdependenzen zwischen Variablen zu erforschen. Die Unterschiede zwischen den beiden Verfahren sind folgende:

- Die Faktorenanalyse (Hauptachsenmethode) repräsentiert die Kovarianzstruktur in Form eines hypothetischen Kausalmodells. Die Korrelationen sollen durch eine kleinere Nummer von Faktoren erklärt werden.
- Die Principal Component Analysis fasst die Daten durch Linearkombinationen der beobachteten Daten zusammen. Hierbei ist keine Aussage über ein Kausalmodell notwendig, das als fragwürdig angesehen werden kann. Dafür entdeckt man aber auch keine zugrunde liegende Kausalstruktur, falls solch eine Struktur existiert.

Wir sehen, das Ziel der Principal Component Analysis ist ein anderes als das der Faktorenanalyse.

3.1.1 Dekompositionsstrategie

Wir suchen einen Vektor,

$$a_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{pmatrix}$$

der mit sich selbst multipliziert die Korrelationsmatrix bestmöglich approximiert. Für diesen Vektor a soll also gelten:

$$\mathbf{R} \cong a_1 a_1'$$

da:

$$a_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} \left| \begin{matrix} a_1 \cdot a_1' & (a_{11} \ a_{21} \ a_{31}) = a_1' \\ \hline \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \end{matrix} \right. = \mathbf{R}$$

Abbildung 3.4: Dyadisches Produkt: Reproduzierte Korrelationsmatrix

Wenn wir nun auf beiden Seiten rechtsseitig (in der Matrizenrechnung ist nämlich $\mathbf{AB} \neq \mathbf{BA}$) mit a_1 multiplizieren ergibt sich folgendes:

$$\mathbf{R}a_1 \cong a_1 a'_1 a_1$$

Wir setzen nun $a'_1 a_1 = \lambda_1$, da

$$\begin{array}{c|c} a'_1 \cdot a_1 & \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = a_1 \\ \hline a'_1 = (a_{11} \ a_{21} \ a_{31}) & \sum_{k=1}^m a_{jk}^2 = \lambda_1 \end{array}$$

Abbildung 3.5: Skalarprodukt: Eigenwert

$$\mathbf{R}a_1 \cong a_1 \lambda_1$$

Wir erhalten also einen Skalar, den sogenannten Eigenwert. Desweiteren gilt in der Matrizenrechnung: $\mathbf{A} = \mathbf{AI}$. Wir können also als nächsten Schritt schreiben:

$$\mathbf{R}a_1 \cong \lambda_1 a_1 \mathbf{I}$$

Dies wiederum lässt sich zusammenfassen zu:

$$\mathbf{R}a_1 - \lambda_1 a_1 \mathbf{I} \cong \mathbf{0}$$

Nun klammern wir a_1 aus und erhalten folgendes:

$$(\mathbf{R} - \lambda_1 \mathbf{I})a_1 \cong \mathbf{0}$$

→ $(\mathbf{R} - \lambda_1 \mathbf{I})$ darf nicht invertierbar sein, da sonst nur eine triviale Lösung existiert, also $|\mathbf{R} - \lambda_1 \mathbf{I}| = 0$.

Die Berechnung des Eigenvektors a_1 und des (größten) Eigenwerts λ_1 ist aus der Mathematik wohl bekannt, und wird dort als Eigenwertproblem bezeichnet.

Da, wie bereits oben bemerkt $\lambda_1 = a'_1 a_1$ gilt, sind wir zu folgender Aussage berechtigt:

$$\begin{array}{c|c} a'_1 \cdot a_1 & \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} = a_1 \\ \hline a'_1 = (a_{11} \ a_{21} \ a_{31}) & \sum_{k=1}^m a_{jk}^2 = \lambda_1 \end{array}$$

Abbildung 3.6: Eigenwert

Der erste λ_j ergibt sich als Summe der quadrierten Faktorladungen ($\sum a_{jk}^2$) des ersten Faktors. Aus der einfachen Regressionsanalyse ist r^2 als Determinationskoeffizient bekannt. Dieser drückt den Anteil der von Y durch X erklärten Varianz aus. Da $a = r$ ist $a^2 = r^2$. So ergibt sich der erste λ_j als Summe von m Determinationskoeffizienten, also quadrierten Faktorladungen, die mit dem ersten Faktor verbunden sind. Damit gibt er die Summe der durch den ersten Faktor F_1 erklärten Varianz an.

3.1.2 Beispiel

Das Lösen dieser Gleichung ergibt Eigenwerte (λ_j) und Eigenvektoren (hier: a), die mit der Matrix \mathbf{R} , aus der sie berechnet wurden in Zusammenhang stehen. Die Gleichung sieht wie folgt aus:

An Hand eines Beispiels sehen wir vielleicht deutlicher, wie dies von Statten geht. Als Beispiel wählen wir folgende Matrix \mathbf{A} :

$$\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

Nun subtrahieren wir von dieser Matrix die λ -Diagonalmatrix

$$(\mathbf{A} - \lambda \mathbf{I}) = \begin{pmatrix} 4 - \lambda & 2 \\ 2 & 1 - \lambda \end{pmatrix}$$

Die Determinante dieser Matrix lautet folgendermaßen:

$$(4 - \lambda) \cdot (1 - \lambda) - 2 \cdot 2$$

Um eine nicht-triviale Lösung zu erlauben, muss diese Determinante Null sein. Warum liefert nur dies eine nicht-triviale Lösung? Für den Fall, dass die Determinante nicht 0 ist (die Matrix also nicht singulär ist), kann die Gleichung

$$(\mathbf{A} - \lambda \mathbf{I})a = 0$$

ganz einfach durch Invertieren (was nur für Matrizen mit Determinanten ungleich Null möglich ist) gelöst werden, nämlich folgendermaßen:

$$(\mathbf{A} - \lambda \mathbf{I})^{-1} \cdot (\mathbf{A} - \lambda \mathbf{I}) \cdot a = (\mathbf{A} - \lambda \mathbf{I})^{-1} \cdot 0$$

$$\mathbf{I} \cdot a = (\mathbf{A} - \lambda \mathbf{I})^{-1} \cdot 0$$

$$a = \text{Nullvektor}$$

Für eine nicht-triviale Lösung darf also

$$(\mathbf{A} - \lambda \mathbf{I})^{-1}$$

nicht existieren. Dies ist dann der Fall, wenn

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

da sich die Inverse wie folgt berechnet:

$$(\mathbf{A} - \lambda \mathbf{I})^{-1} = \frac{\text{adj}(\mathbf{A} - \lambda \mathbf{I})}{\det(\mathbf{A} - \lambda \mathbf{I})}$$

Wir sehen also, dass die Berechnung bei einer Determinante von 0 nicht möglich ist, da wir durch 0 dividieren müssten. Nun zurück zu unserem Beispiel: Wir haben folgende Gleichung zu lösen:

$$\begin{aligned} (4 - \lambda) \cdot (1 - \lambda) - 2 \cdot 2 &= 0 \\ \Leftrightarrow (4 - \lambda) \cdot (1 - \lambda) &= 4 \\ \Leftrightarrow \lambda^2 - 5\lambda &= 4 \\ \Leftrightarrow \lambda^2 - 5\lambda &= 0 \\ \Leftrightarrow \lambda(\lambda - 5) &= 0 \\ \lambda_1 = 0 \vee \lambda_2 = 5 & \end{aligned}$$

Die gesuchten Eigenvektoren ergeben sich aus den beiden Gleichungssystemen:

$$(\mathbf{A} - \lambda_1 \mathbf{I})a = 0 \text{ und } (\mathbf{A} - \lambda_2 \mathbf{I})a = 0$$

Für λ_1 ergibt sich folgendes homogenes Gleichungssystem:

$$\begin{pmatrix} 4 - 0 & 2 \\ 2 & 1 - 0 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 4a_{11} + 2a_{21} = 0 \\ 2a_{11} + 1a_{21} = 0 \end{pmatrix} = \begin{pmatrix} 2a_{11} + a_{21} = 0 \\ 2a_{11} + a_{21} = 0 \end{pmatrix}$$

Daraus folgt: $-2a_{11} = a_{21}$. Somit sind die Komponenten des ersten Eigenvektors nur eindeutig bis auf eine multiplikative Konstante. Eindeutigkeit lässt sich nur durch Normierung erreichen. Üblicherweise werden Eigenvektoren auf die Länge 1 normiert, d.h. es gelte:

$$a' a = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} (a_{11} \ a_{21}) = a_{11}^2 + a_{21}^2 = 1$$

also:

$$\begin{aligned} a_{11}^2 + a_{21}^2 &= 1 && \text{durch Substitution, da } -2a_{11} = a_{21} \\ a_{11}^2 + (-2a_{11})^2 &= 1 \\ a_{11}^2 + 4a_{11}^2 &= 1 \\ 5a_{11}^2 &= 1 \\ a_{11}^2 &= \frac{1}{5} \\ a_{11} &= \sqrt{\frac{1}{5}} \end{aligned}$$

Da $-2a_{11} = a_{21}$ ergibt sich: $a_{11} = \sqrt{\frac{1}{5}}$ und $a_{21} = -2 \cdot \sqrt{\frac{1}{5}}$. Also lautet der erste Eigenvektor:

$$a_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{pmatrix}$$

Analog berechnet sich der zweite Eigenvektor. Die Eigenvektoren lauten:

$$a_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{pmatrix}, \quad a_2 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$$

Es gilt:

$$a_1' a_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \end{pmatrix} \cdot \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} = 0$$

d.h. die beiden Eigenvektoren sind orthogonal zueinander. Allgemein gilt: Eine reelle $n \times n$ -Matrix besitzt reelle Eigenwerte λ_n , die nicht voneinander verschieden zu sein brauchen. Diese λ_n entsprechen n Eigenvektoren v_n , die paarweise orthogonal zueinander stehen. Also:

$$a_i' a_j = 0 \text{ für } i \neq j; \quad i, j = 1, 2, 3, \dots, n$$

Dies ist zugegebenermaßen ein sehr einfaches Beispiel; die Berechnung der λ_j für eine größere Matrix als eine 2×2 -Matrix ist um einiges aufwendiger als hier. Wichtig für uns ist aber die Erkenntnis, dass der größte λ_j den Teil der Varianz repräsentiert, der durch die erste Principal Axis erklärt wird. Hier sei schon einmal angemerkt, dass wir in der Principal Component Analysis mit der normalen Korrelationsmatrix arbeiten, in der alle Einträge der Hauptdiagonalen den Wert 1 haben. In Faktoranalytischen Ansätzen wird mit reduzierten Korrelationsmatrizen gearbeitet, in der in der Hauptdiagonalen die geschätzten Kommunalitäten stehen. Der zweitgrößte λ_j repräsentiert den Teil, der durch die zweite Principal Axis erklärt wird, und so weiter.

In einer Korrelationsmatrix besteht die Hauptdiagonale nur aus 1en. Die Spur der Matrix ist identisch mit der Summe der Elemente der Hauptdiagonalen. Dies ist ebenfalls die Summe der λ_j . Also wissen wir, dass in einer $n \times n$ -Korrelationsmatrix die Hauptdiagonale n Elemente besitzt, und dass diese Elemente *alle* 1 sind. Also ist die Summe aller $\lambda_j = n$. Wenn wir also einen bestimmten λ_j durch n teilen, erhalten wir den Anteil an Varianz, den dieser λ_j erklärt. Es gilt:

$$\sum_{j=1}^n \frac{\lambda_j}{n} = 1$$

3.1.3 Extraktion weiterer Faktoren

Nach der Extraktion des ersten Faktors F_1 soll nun der zweite Faktor F_2 die Restkorrelation bestmöglich approximieren. Wir subtrahieren die reproduzierte Korrelationsmatrix von der ursprünglichen Korrelationsmatrix. So erhalten wir eine Matrix, in der die Abweichungen zwischen den beiden Matrizen stehen.

$$\mathbf{R} - a_1 a_1' \cong a_2 a_2'$$

Nun suchen wir einen Faktor, der diese Abweichungsmatrix bestmöglich darstellt und gehen wie gehabt vor.

$$(\mathbf{R} - \lambda_1 \mathbf{I}) a_1 \cong \mathbf{0}$$

nun

$$((\mathbf{R} - a_1 a_1') - \lambda_2 \mathbf{I}) a_2 \cong \mathbf{0}$$

oder, wenn $\mathbf{R} - a_1 a_1' = \mathbf{R}^+$ für die erste Extraktion gesetzt wird:

$$(\mathbf{R}^+ - \lambda_2 \mathbf{I}) a_2 \cong \mathbf{0}$$

Bedingung: $\lambda_2 = a_2' a_2$, Orthogonalität der Faktoren F_1 und F_2 , also $a_1' a_2 = 0$.

Eine andere Möglichkeit der Berechnung der weiteren Faktoren verzichtet auf die intuitiv einleuchtendere Berechnung der Restkorrelationsmatrix. Indem man aus der Ursprungsmatrix nicht nur den größten Eigenwert λ_1 extrahiert, sondern auch $\lambda_2, \lambda_3, \dots, \lambda_m$ mit den zugehörigen Eigenvektoren a_2, a_3, \dots, a_m erhält man ebenso weitere orthogonale Faktoren. Dieses Verfahren ist weniger rechenaufwendig und wird daher öfter genutzt.

3.2 “Die” Faktorenanalyse (Hauptachsenfaktorenanalyse)

Die Faktorenanalyse ist wie die Hauptkomponentenanalyse ein Verfahren, dass auf Dimensionsreduktion abzielt. Wie in der Hauptkomponentenanalyse wird nicht zwischen abhängigen und unabhängigen Variablen unterschieden. Die Idee ist, neue “Supervariablen” -oder Faktoren- zu konstruieren, die hoffentlich zu einem besseren Verständnis der Daten führen. Während dies bei der Hauptkomponentenanalyse durch orthogonale Transformation der Daten erreicht wird, der kein statistisches Modell zu Grunde liegt, basiert die Faktorenanalyse auf einem statistischen Modell. Hier wird versucht, die Kovarianzstruktur der Daten zu erklären, und nicht die Varianzen. Varianzen, die nicht durch die gemeinsamen Faktoren erklärt werden können werden durch Rest- oder Fehlerterme erklärt.

3.3 Die Varianzen im Common-Factor Modell

Historisch handelt es sich beim Common-Factor Modell um die Principal Axis Factoring Prozedur, die die Dekompositions-Strategien der Principal Components Analysis auf die reduzierte Korrelationsmatrix anwendet. Näheres zu dem Begriff der reduzierten Korrelationsmatrix im Kapitel “Das Kommunalitätenproblem” auf Seite 44.

Nachdem die Elemente der Hauptdiagonalen durch die Kommunalitätsschätzer \hat{h}_j ersetzt worden sind können die Faktoren so im Common-Factor Modell so extrahiert werden wie in der Principal Component Analysis. Es gilt:

$$\det(\mathbf{R}_1 - \lambda \mathbf{I}) = 0$$

wobei \mathbf{R}_1 die Korrelationsmatrix mit den geschätzten Kommunalitäten in der Hauptdiagonalen ist.

$$\mathbf{R}_1 = \mathbf{R} - \mathbf{U}^2$$

3.4 Maximum Likelihood Faktorenanalyse

Der Maximum-Likelihood Ansatz verläuft folgendermaßen: Die Maximum-Likelihood Lösung findet die *wahrscheinlichsten* Populationswerte, die die beobachtete Korrelationsmatrix produziert hätten. Dies geschieht unter der Hypothese, dass ein k-Common Factor Modell den Daten perfekt angepasst ist. Desweiteren geht man davon aus, dass die Population multivariat normalverteilt ist. Was dies bedeutet sei in Grafik 3.7 im trivariaten Fall verdeutlicht. Bei dieser Graphik handelt es sich um eine multivariate Normalverteilung mit unkorrelierten Variablen.

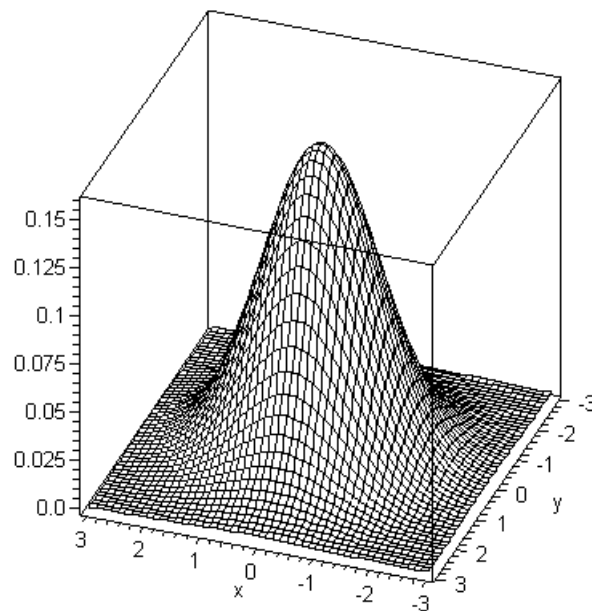


Abbildung 3.7: Multivariate Normalverteilung

Wir nehmen also an, dass die vorhandene Stichprobe aus einer Population stammt, in der ein k-Common Faktor Modell exakt gilt, und wo die Verteilung der Variablen (inklusive der Faktoren) multivariat normal ist. Was unbekannt ist, sind die exakten Konfigurationen der Parameter, z.B. Die exakten Ladungen auf den Variablen. Das Ziel ist nun, die zugrundeliegenden Populationsparameter zu finden, die unter der bestehenden Hypothese die größte Wahrscheinlichkeit (maximum likelihood) haben, die beobachtete Korrelationsmatrix zu “verursachen”. Mit der Maximum Likelihood-Methode

ist ein χ^2 -Signifikanztest assoziiert, der Aussagen über die Güte der Anpassung trifft. Dabei gilt:

$$U_k = N [\ln |\mathbf{C}| - \ln |\mathbf{R}| + \text{tr}(\mathbf{R}\mathbf{C}^{-1}) - m]$$

wobei:

$$\begin{aligned} N &= \text{Stichprobengröße} \\ m &= \text{Anzahl der Variablen} \\ \mathbf{R} &= \text{Korrelationsmatrix} \\ \mathbf{C} &= \mathbf{F}\mathbf{F}' + \mathbf{U}^2, \text{ wobei} \\ \mathbf{F} &= \text{Matrix der Faktorladungen} \\ \mathbf{U}^2 &= \text{Matrix der Uniqueness} \\ \ln &= \text{logarithmus naturalis, } \log_e \\ \text{tr}(\mathbf{R}\mathbf{C}^{-1}) &= \text{Spur der Matrix } \mathbf{R}\mathbf{C}^{-1} \end{aligned}$$

und den Freiheitsgraden

$$df_k = \frac{1}{2}[(m - k)^2 - (m + k)]$$

wobei

$$\begin{aligned} k &= \text{Anzahl der hypothetischen Faktoren} \\ m &= \text{Anzahl der Variablen} \end{aligned}$$

Hierbei ist es nicht so, wie in der bivariaten Statistik, dass ein großes χ^2 angestrebt wird. Ein großer χ^2 , oder besser ein hoher φ oder Cramers V-Wert zeigen dort einen starken Zusammenhang an. Er wird gemessen über die Differenz der Kontingenztafel und der Indifferenztafel. Diese Tabellen können wir als Matrizen auffassen. Nun sehen wir, dass ein großer χ^2 -Wert eine große Differenz der Matrizen anzeigt. Dies ist in der bivariaten Statistik zur Messung des Zusammenhanges zweier nominaler Variablen insofern gut, als dass eine große Differenz äquivalent ist zu einem starken Zusammenhang. Hier liegt der Fall genau andersherum. Es ist für uns gerade eben *nicht* von Vorteil, wenn wir einen großen χ^2 -Wert erhalten. Wir streben ein möglichst *kleines*(!) χ^2 an, da genau dann die Differenz zwischen Korrelationsmatrix und reproduzierter Korrelationsmatrix gering ist, also die Reproduktion der Ursprungskorrelationsmatrix aus den Faktoren gut. Faktorenanalysen nach dem ML-Prinzip bilden auch die Grundlage zur Analyse für Strukturgleichungsmodelle, die mit Programmpaketen wie LISREL von Karl Jöreskog und Dag Sörbom, EQS von Peter Bentler, M-Plus von Bengt Muthén, Amos (SPSS-Erweiterung) oder Sepath möglich ist, um die Bekanntesten zu nennen. Ein gutes Beispiel zur Maximum Likelihood Methode kann im Detail in "Datenanalyse mit Stata" von Ulrich Kohler und Frauke Kreuter nachgelesen werden. Um ein besseres Verständnis für die Idee hinter Maximum

Likelihood zu fördern, da ML auch in anderen Verfahren wie z.B. der logistischen Regression eine Rolle spielt, wird hier ein kurzer Abriss dieses Beispiels gegeben.

3.4.1 Erläuterung: Maximum Likelihood

Bei diesem –leicht makaberen– Beispiel stellen wir uns die Titanic vor. Wir nehmen an, die Wahrscheinlichkeit diese Unglück überlebt zu haben beträgt $P_{(\text{überlebt})} = 0.60$. Logischerweise beträgt dann die Wahrscheinlichkeit, nicht überlebt zu haben $P_{(\text{nicht überlebt})} = 0.40$. Nun ziehen wir 3 der beim Auslaufen auf der Titanic befindlichen Passagiere und betrachten sie daraufhin, ob sie gerettet wurden, oder nicht. Betrachten wir dazu Graphik 3.8.

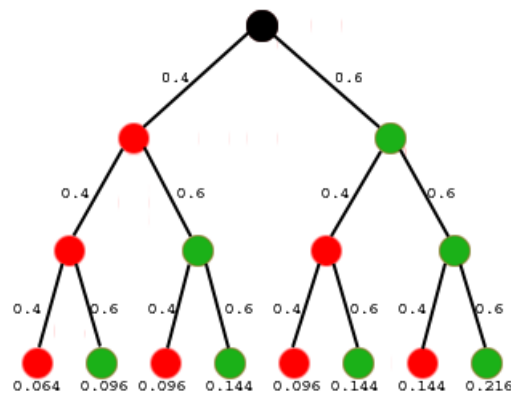


Abbildung 3.8: Titanic Baumdiagramm

Hier sehen wir alle möglichen Ergebnisse für 3 gezogene Passagiere für das dichotome Merkmal “überlebt”, wobei die Wahrscheinlichkeit für jeden Zug identisch ist. Wenn wir die verschiedenen Möglichkeiten sortieren, und nach Überlebenden zusammenrechnen erhalten wir:

überlebt	nicht überlebt	Wahrscheinlichkeit	Ergebnis
3	0	0.216	0.216
2	1	$0.144 + 0.144 + 0.144$	0.432
1	2	$0.096 + 0.096 + 0.096$	0.288
0	3	0.064	0.064
			$\sum = 1$

Allgemein beträgt die Wahrscheinlichkeit eine Stichprobe vom Umfang n zu ziehen, in dem ein dichotomes Merkmal h mal auftritt, und die Wahrscheinlichkeit für jeden Zug identisch ist:

$$P(h|\pi, n) = \binom{n}{h} \pi^h (1 - \pi)^{n-h}$$

Sie ist also binomialverteilt. Dabei steht $\binom{n}{h}$ für $\frac{n!}{h!(n-h)!}$. Mit dieser Formel wird die Anzahl *aller* Möglichen Stichproben errechnet, in denen das Merkmal h mal auftritt. π steht für den Anteil des dichotomen Merkmals in der Grundgesamtheit. π^h ist die Wahrscheinlichkeit für h Elemente, die das Merkmal π tragen, zur Potenz da “und-Verknüpfungen” multipliziert werden. Also: $\pi^h = \pi_1 \cap \pi_2 \cap \pi_3 \dots \pi_h = \pi_1 \cdot \pi_2 \cdot \pi_3 \dots \pi_h$. Analog verhält es sich mit $(1 - \pi)^{n-h}$. $1 - \pi$ steht für den Anteil der Elemente, die das Merkmal π nicht tragen. Es ist logisch, dass der Anteil aller Elemente die das Merkmal tragen + dem Anteil aller Elemente die das Merkmal nicht tragen = 1 sein muss, da $(1 - \pi) + (\pi) = 1$. Der Exponent $n - h$ erklärt sich dadurch, dass wenn von n Elementen h das entsprechende Merkmal tragen sollen $m = n - h$ Elemente übrig bleiben. Also wenn von 3 Elementen 2 das Merkmal tragen bleiben $3 - 2 = 1$ Elemente übrig die das Merkmal *nicht* tragen.

In unserem Titanic-Beispiel oben haben wir den Wert $\pi = 0.6$ gegeben. Was wäre, wenn wir diesen Wert π nicht kennen würden? In der Praxis ist das Ergebnis, was wir durch Anwendung der Maximum Likelihood Methode erhalten wollen nicht eine spezielle Wahrscheinlichkeit, sondern der Wert π aus der Grundgesamtheit. Normalerweise ist dieser nicht bekannt, jedoch kann man überlegen, welcher Wert von π in der Grundgesamtheit gegeben sein muss, damit unsere Stichprobe die Maximale Wahrscheinlichkeit (englisch: Maximum Likelihood) besitzt, gezogen zu werden. Hierzu können wir verschiedene Werte für π in die oben genannte Formel einsetzen und dann den Wert Auswählen, der die größte Wahrscheinlichkeit aufweist. Dies bedeutet, wir suchen nach dem Wert für π , der die *Likelihood*

$$\mathcal{L}(\pi|(h, n) = \binom{n}{h} \pi^h (1 - \pi)^{n-h}$$

maximiert. Dabei ist die Berechnung von $\binom{n}{h}$ uninteressant, da sie für uns eine Konstante darstellt. In unserem Beispiel würden wir jedesmal $\binom{3}{2} = 3$ ausrechnen. In Graphik 3.9 sind die Werte $\frac{1}{100}, \frac{2}{100}, \frac{3}{100}, \dots, \frac{100}{100}$ für π verwendet worden.

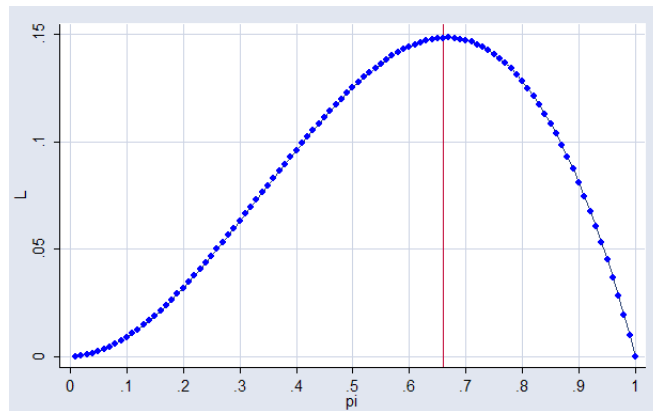


Abbildung 3.9: Maximum Likelihood

Wir sehen, das Maximum der Likelihood liegt etwa bei 0.66. Hierbei handelt es sich um das Ergebnis der Maximum Likelihood Schätzung für den Anteil π der Überlebenden in der Grundgesamtheit, wenn wir eine Stichprobe bestehend aus 3 Personen mit 2 Überlebenden gezogen haben.

3.4.2 Canonical Factoring

Bei der Canonical Factoring Methode handelt es sich um eine Variation des Maximum Likelihood Ansatzes. Das Kriterium hierbei ist, dass die Kanonische Korrelation zwischen den k -Common-Factors und den beobachteten Variablen maximal für das getestete Set aus allen möglichen k -Common-Factor Modellen ist.

3.4.3 Maximale Determinante der Residual-Korrelations-Matrix

Auch hierbei handelt es sich um eine Maximum Likelihood-basierte Methode. Hier wird Jöreskog's Lösung als sehr gut angesehen. Prinzipiell ist sie nicht sehr unterschiedlich von der Lösung durch die Eigenequation. Es gilt:

$$\det(\mathbf{R}_2 - \lambda \mathbf{I}) = 0$$

wobei

$$\mathbf{R}_2 = \mathbf{U}^{-1}(\mathbf{R} - \mathbf{U}^2)\mathbf{U}^{-1}$$

$$\mathbf{R}_2 = \mathbf{U}^{-1}\mathbf{R}_1\mathbf{U}^{-1}$$

3.5 Least Squares Methoden

Der Least-Squares Ansatz in der Common Factor Analysis geht folgendermaßen vor: Nachdem eine gewisse Anzahl an Faktoren extrahiert wurde wird die Residualkorrelation minimiert. Der “degree of fit” zwischen der reproduzierten und der beobachteten Korrelationsmatrix wird betrachtet. Man untersucht die quadrierten Differenzen. An dieser Stelle betrachten wir noch einmal, wie die Least Squares oder kleinste Fehlerquadrat Methode in der einfachen linearen Regression angewendet wird, um noch einmal in Erinnerung zu rufen, wie diese Methode funktioniert:

3.5.1 Erläuterung: Least Squares

Das Verfahren der kleinsten Fehlerquadrate geht zurück auf Carl Friedrich Gauss, uns allen noch in guter Erinnerung, zierte sein Konterfei doch den 10 DM Schein. Das Verfahren selbst ist uns aus der einfachen linearen Regression bekannt.

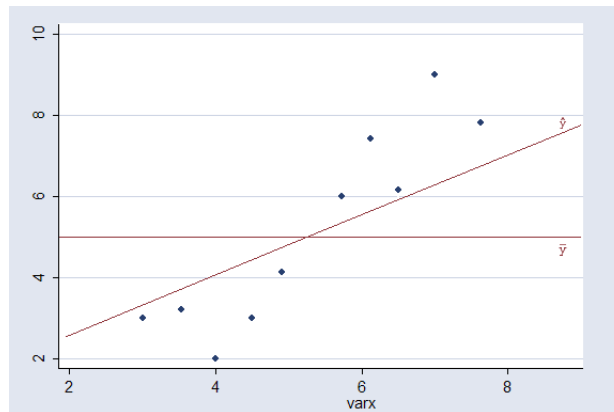


Abbildung 3.10: Bivariate lineare Regression

Wie können wir die Regressionsgerade $\hat{y}_i = a + b \cdot x_i$ bestimmen, die bestmöglich durch die Datenpunkte verläuft? Denn genau das ist es, was das Verfahren der kleinsten Fehlerquadrate bestimmt. Die bestmögliche Regressionsgerade. Verdeutlichen wir uns erst einmal an Hand eines einzelnen Datenpunktes was die folgenden Begrifflichkeiten überhaupt bedeuten. Wir werden von erklärten Fehlern, nicht erklärten Fehlern und totalen Fehlern sprechen.

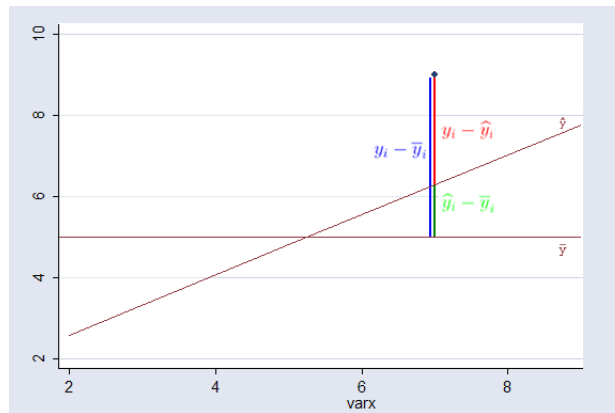


Abbildung 3.11: Varianzzerlegung für einen Datenpunkt

Der grüne Bereich bezeichnet den *erklärten Fehler*. Der rote Bereich bezeichnet die Abweichung der Fälle von ihrem vorhergesagten Wert auf der Regressionsgerade, dies ist der *nicht erklärte Fehler*. Beide Bereiche, also rot und grün addiert ergeben den blauen Bereich, den *totalen Fehler*. Da wir es aber nicht nur mit einem einzelnen Fall zu tun haben, sondern mit n Fällen, müssen wir mehrere Datenpunkte mit jeweils erklärten Fehlern, nicht erklärten Fehlern und totalen Fehlern berücksichtigen.

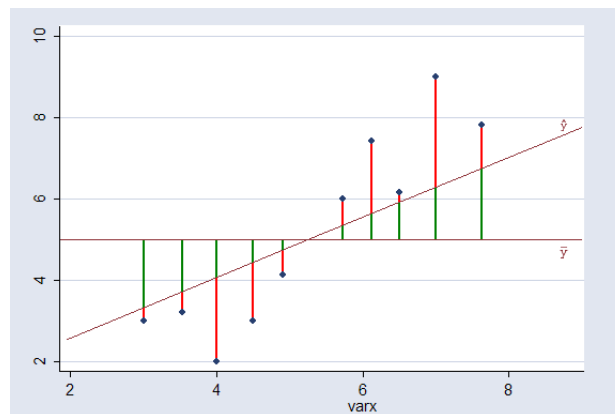


Abbildung 3.12: Varianzzerlegung für alle Datenpunkte

Für alle Datenpunkte werden nun die einzelnen Fehlerterme aufsummiert. Wir haben es also hier nicht mit

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + y_i - \hat{y}_i$$

für einen einzelnen Datenpunkt, sondern mit der Summe mehrerer -nämlich n - Datenpunkte, unterteilt nach den verschiedenen Fehlerarten zu tun:

$$\sum_{i=1}^n y_i - \bar{y} = \sum_{i=1}^n \hat{y}_i - \bar{y} + \sum_{i=1}^n y_i - \hat{y}_i$$

Hierbei stoßen wir auf ein altbekanntes Problem. Der Term $\sum_{i=1}^n y_i - \bar{y}$ ist $= 0$. Dies ist immer so und erklärt sich aus der Definition des arithmetischen Mittels $\bar{y} = \frac{\sum y_i}{n}$

$$\begin{aligned} \sum y_i - \bar{y} &= \sum y_i - \sum \bar{y} \\ \sum y_i - \bar{y} &= \sum y_i - n\bar{y} \\ \sum y_i - \bar{y} &= \sum y_i - n \frac{\sum y_i}{n} \\ \sum y_i - \bar{y} &= \sum y_i - \sum y_i \\ \sum y_i - \bar{y} &= 0 \end{aligned}$$

Dieses Problem ist uns von der Berechnung der Standardabweichung bekannt. Aus diesem Grund ist die Standardabweichung als Wurzel aus der quadrierten Standardabweichung, uns besser bekannt als Varianz definiert: $s = \sqrt{s^2}$, oder ausgeschrieben: $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$ und $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$. Auch hier löschen sich positive Abweichungen nach oben und negative Abweichungen nach unten aus und ergeben, schlussendlich, Null. Dies wird vermieden, wenn der Term $\sum y_i - \bar{y}$ quadriert wird. Es resultiert nun:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Die Idee, die hinter dem Verfahren der kleinsten Fehlerquadrate steckt kann nun schon fast intuitiv erahnt werden. Der Term $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ ist die Fehler-Quadrat Summe. Sie setzt sich wie folgt zusammen: $y_i - \hat{y}_i$ ist der Fehlerterm. Dieser Term ist quadriert. Und da wir mit n quadrierten Fehlertermen zu tun haben, handelt es sich um eine Summe. Die sogenannte Fehler-Quadrat-Summe. Dies sind die ominösen Fehlerquadrate.

Wir haben also jetzt ein Verständnis dafür, woher der Begriff Fehlerquadrate kommt. Aber was ist mit *kleinsten* Fehlerquadraten gemeint? Auch dies ist kein Buch mit 7 Siegeln. Wenn wir uns verschiedene Geraden ansehen, wie in Graphik 3.13 verändert sich der Term $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ für jede einzelne dieser Geraden. Er ist minimal für die grüne Gerade und größer für *alle* anderen möglichen Geraden, einige von den unendlich vielen Möglichen sind in rot eingezeichnet. Wir suchen die Gerade, für die $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ minimal ist.

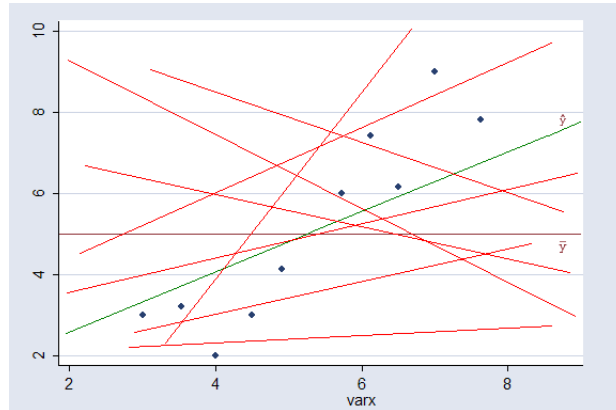


Abbildung 3.13: verschiedene Regressionsgeraden

Wir wissen, dass die Gleichung einer Gerade im \mathbb{R}^2 die Form $f(x) = mx + b$ besitzt. Die Formel $\hat{y}_i = a + b \cdot x_i$ besagt, von den anderen Bezeichnungen abgesehen, nichts anderes. Da wir $\hat{y}_i = a + b \cdot x_i$ in $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ substituieren können, erhalten wir: $f(a, b) = \sum_{i=1}^n (y_i - (a + bx_i))^2$. Durch partielles Ableiten $\frac{\partial f(a,b)}{\partial a}$ nach a , beziehungsweise $\frac{\partial f(a,b)}{\partial b}$ nach b erhalten wir die bekannten Schätzer $a = \bar{y} - \bar{x}b$ für die Regressionskonstante und $b = \frac{s_{xy}}{s_x^2}$ für den Regressionskoeffizienten.

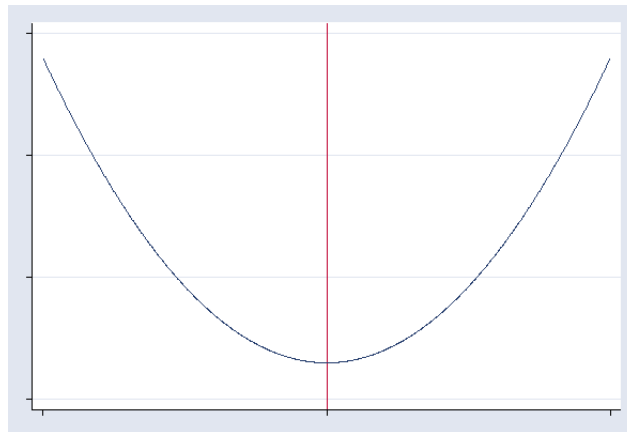


Abbildung 3.14: Bestimmung des kleinsten Fehlerquadrates

Eine optische Verdeutlichung kann man sich in Graphik 3.14 ansehen. Da es sich bei der Funktion um eine Quadratische handelt, kann man durch Ableitung die Extremwerte errechnen. Hier wollen wir ermitteln, wo das Minimum dieser Funktion liegt. Der Extremwert liegt dort, wo die erste Ableitung gleich Null ist. Es könnte sich um eine Minimum oder Maximum handeln. Wenn die

zweite Ableitung in dem Punkt der Nullstelle der ersten Ableitung positiv ist, handelt es sich um ein Minimum. Wie dies beispielsweise aussehen könnte ist in eben genannter Graphik dargestellt. Dort läge das kleinste Fehlerquadrat am gekennzeichneten Scheitelpunkt der Parabel.

3.5.2 Principal Axis Factoring mit iterativer Schätzung der Kommunalitäten

Bei diesem Ansatz, dem der Least-Squares Gedanke zugrunde liegt geht man folgendermaßen vor:

- Es werden k Faktoren angenommen, die für die beobachtete Korrelation verantwortlich sind.
- Einsetzen von Anfangs-Kommunalitäten-Schätzungen.
- Extraktion von k Faktoren, die die beobachtete Korrelationsmatrix nach den Least-Squares Kriterien am besten reproduzieren können.
- Um das Factor Pattern zu erhalten, das die beobachtete Korrelationsmatrix bestmöglich reproduzieren kann, werden die Kommunalitäten neu geschätzt. Dies geschieht auf der Basis des als erste Lösung erhaltenen Factor Patterns.
- Dieser Schrittweise (iterative) Prozess wird solange wiederholt, bis keine Verbesserung in der Kommunalitätenschätzung mehr festgestellt werden kann.

3.5.3 MinRes (Minimum Residuals)

Auch hierbei handelt es sich um eine iterative Schätzung, der der Least-Squares Gedanke zu Grunde liegt. Allerdings wird ein Anderer und auch effizienterer Algorithmus verwendet. Für diese Technik existiert ein χ^2 -Test, der für grosse Stichproben angemessen ist. Nach Harry Harman kann dieser Test, der unabhängig von der speziellen Methode der Faktorextraktion ist, auch bei anderen Methoden der Extraktion angewendet werden, und kann als Mittel genutzt werden, um die "completeness of factorization" zu prüfen.

3.6 Weitere Ansätze

Es existieren noch weitere Ansätze neben der Principal Component Analysis und der Faktorenanalyse mit ML und LS-Verfahren. An dieser Stelle soll nicht

im Detail auf diese Verfahren eingegangen werden. An dieser Stelle werden sie der Vollständigkeit halber kurz erwähnt, sowie knapp dargestellt, da man ihnen z.B. in SPSS als Auswahlmöglichkeit bei der Faktorextraktion über den Weg laufen kann.

3.6.1 α -Factoring

Der Name α -Factoring geht auf den Reliabilitätskoeffizienten α von Cronbach zurück. Mit dem α -Koeffizienten wird die Reliabilität der aus allen Testitems gebildeten Summenscores geschätzt. Hierbei werden alle Testitems als eigenständige Tests für ein und dasselbe Merkmal angesehen. Die Reliabilität des Summenscores (α) ergibt sich als durchschnittliche Paralleltestreliabilität für alle mögliche Paare von Items. Das Anliegen der α -Faktorenanalyse ist es nun, Faktoren mit möglichst hoher Generalisierbarkeit zu bestimmen.

3.6.2 Image Analysis

Die Image-Analyse wählt einen anderen Lösungsansatz für das Kommunalitätenproblem. Es wird von einer Population von Versuchspersonen und einer Population von Variablen ausgegangen. Die gemeinsame Varianz einer Variablen ist als derjenige Varianzanteil definiert, der potentiell durch multiple Regression von allen anderen Variablen der Variablenpopulation vorhergesagt werden kann. Dieser gemeinsame Varianzanteil wird als *Image der Variablen* (im Sinne von "Abbildung") bezeichnet. Der nicht vorhersagbare Varianzanteil wird als *Anti-Image* bezeichnet. Für die konkrete Durchführung einer Image-Analyse stehen nur Stichprobendaten zur Verfügung. Die kompletten Universen der Versuchspersonen sowie der Variablen sind nur teilweise greifbar. Dementsprechend sind das Image und das Anti-Image jeweils Schätzungen auf Grund der Stichprobe. Man spricht deshalb nicht vom *Image*, sondern vom *Partial-Image* der Variablen. Hierfür werden die konkreten Messwerte einer Variablen i durch vorhergesagte Werte \hat{x} bzw. \hat{z} ersetzt, die man durch die multiplen Regression zwischen der entsprechenden Variable i und den übrigen $n-1$ Variablen erhält. Aus der Korrelationsmatrix dieser vorhergesagten Messwerte (in der die Hauptdiagonale aus Einsen besteht) werden nach der Hauptachsenmethode Faktoren extrahiert. Die Korrelationen der Variablen kommen so nur auf Grund der gemeinsamen Varianzen mit allen anderen Variablen zustande. So ist gewährleistet, dass die extrahierten Faktoren nur gemeinsame Varianz aufklären.

Kapitel 4

Probleme der Faktorenanalyse

4.1 Das Kommunalitätenproblem

Wir werden in einem späteren Kapitel sehen, dass wir nicht immer eine Korrelationsmatrix benötigen, sondern teilweise auch eine speziell bearbeitete, nämlich eine *reduzierte Korrelationsmatrix*. Was ist eine reduzierte Korrelationsmatrix? Hier sind in die Elemente der Hauptdiagonalen, in einer normalen Korrelationsmatrix bestehend nur aus Einsen, durch die entsprechenden Kommunalitätenschätzungen ersetzt worden. Also:

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{pmatrix} \longrightarrow \begin{pmatrix} h_1^2 & r_{12} & r_{13} & r_{14} \\ r_{21} & h_2^2 & r_{23} & r_{24} \\ r_{31} & r_{32} & h_3^2 & r_{34} \\ r_{41} & r_{42} & r_{43} & h_4^2 \end{pmatrix}$$

Der Unterschied besteht darin, dass die Elemente r_{11} , r_{22} , r_{33} und r_{44} einer normalen Korrelationsmatrix alle gleich 1 sind. In einer reduzierten Korrelationsmatrix sind diese Elemente der Hauptdiagonalen ≤ 1 . Es handelt sich hierbei um die Kommunalitäten h^2 . In der Principal Component Analysis wird versucht, die Varianzen der manifesten Variablen zu erklären. Hier benötigt man die normale Korrelationsmatrix. In der Faktorenanalyse jedoch wird nicht versucht, die Varianzen der manifesten Variablen zu erklären, sondern deren Korrelationen. Da sich die Varianz einer Variablen durch die Kommunalität (h_j^2) und die Uniqueness (u_j^2) zusammensetzen, die Uniqueness für uns aber nur in der Variable selbst begründet liegt, interessiert uns nur die Kommunalität. Deshalb wird die Uniqueness von der normalen Korrelationsmatrix subtrahiert, und es bleibt die reduzierte Korrelationsmatrix mit Kommunalitäten in der Hauptdiagonale übrig. Also:

$$\begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{pmatrix} - \begin{pmatrix} u_1^2 & 0 & 0 & 0 \\ 0 & u_2^2 & 0 & 0 \\ 0 & 0 & u_3^2 & 0 \\ 0 & 0 & 0 & u_4^2 \end{pmatrix} = \begin{pmatrix} h_1^2 & r_{12} & r_{13} & r_{14} \\ r_{21} & h_2^2 & r_{23} & r_{24} \\ r_{31} & r_{32} & h_3^2 & r_{34} \\ r_{41} & r_{42} & r_{43} & h_4^2 \end{pmatrix}$$

Da $r_{jj} - u_j^2 = h_j^2$ identisch ist mit $1 - u_j^2 = h_j^2$.

Das Problem, das nun besteht ist folgendes: h_j^2 ist a priori unbekannt. Die Kommunalität ergibt sich erst nach der Extraktion. Es werden also im Vorfeld Schätzer für die Kommunalitäten benötigt. Dies wird als *Kommunalitätenproblem* bezeichnet. Gewöhnlich werden folgende Kommunalitätenschätzer verwendet:

- Die maximale Korrelation einer gegebenen Variablen aus ihrem "Korrelationspool" mit allen verbleibenden Variablen. Der Nachteil dieser Methode ist, dass nur die Beziehung zu *einer* anderen Variablen berücksichtigt wird

- Die Triadenmethode

$$h_j^2 = \frac{r_{jk}r_{jl}}{r_{kl}}$$

wobei l und k = Variablen, die maximal mit j korrelieren

- Durchschnitt aller Korrelationen einer gegebenen Variablen mit allen anderen

$$h_j^2 = \frac{\sum_{k=1}^n r_{jk}}{n-1}$$

- Die quadrierte multiple Korrelation (SMC = squared multiple correlation) einer Variablen mit den restlichen Variablen. Die quadrierte multiple Korrelation wird als beste Schätzung angesehen.

$$h_j^2 = r^2$$

Der multiple Determinationskoeffizient r_j^2 kann bestimmt werden über $r_j^2 = 1 - \frac{1}{r^{jj}}$, wobei r^{jj} ein Diagonalelement von \mathbf{R}^{-1} ist. Es kann gezeigt werden, worauf an dieser Stelle jedoch verzichtet wird, dass die wahre Kommunalität stets zwischen r^2 und 1 liegt, also $r_j^2 \leq h_j^2 \leq 1$. Bei r_j^2 handelt es sich demnach um eine Untergrenze.

4.2 Anzahl der Faktoren

Da generell so viele Faktoren extrahiert werden können, wie Variablen vorhanden sind, ist eine wichtige Frage, wie viele Faktoren genutzt werden sollen. Die beiden geläufigsten Kriterien, um zu bestimmen wie viele Faktoren extrahiert werden sollen sind das Kaiser-Kriterium und der Scree-Test, ein Weiteres das kumulierte Varianzkriterium. Oft sind die Ergebnisse unterschiedlich. Es bietet sich also immer an, mehrere Kriterien zu nutzen und danach eine Entscheidung nach gründlicher Überlegung zu treffen. Diese Kriterien werden im Folgenden kurz dargestellt.

4.2.1 Das Kaiser-Kriterium

Das Kaiser-Kriterium (oder Kaiser-Gutman-Kriterium) besagt, dass nur Faktoren mit $\lambda > 1$ extrahiert werden. Da die ursprünglichen Variablen durch Standardisierung eine Streuung von 1 haben, sollen nur Faktoren extrahiert werden, die mehr Streuung aufklären als die Ursprungsvariablen verursachen. Faktoren mit einem Eigenwert < 1 tragen so nicht zu einer Verdichtung bei, da sie weniger Streuung erklären, als Ursprungsvariablen (die sie ja ersetzen sollen) an Streuung verursachen. Dieses Kriterium ist weit verbreitet, aber es überschätzt oft die Anzahl der bedeutsamen Faktoren.

4.2.2 Der Scree-Test

Der Name des Scree-Tests leitet sich vom englischen “scree” für Geröll ab. Hier werden Eigenwerte gegen Faktorenzahl geplottet. Man extrahiert so viele Faktoren, wie *links* vom Knick liegen. Wenn die Eigenwerte *rechts* vom Knick ungefähr auf einer Parallele zur X-Achse liegen, so lassen sie sich in der Regel als das Resultat von “Zufallsfaktoren” erklären. Das Verfahren liefert jedoch nicht immer eindeutige Lösungen. Diese Graphik erinnert stark an den Elbow-Plot aus der Clusteranalyse, mit dem die optimale Clusterzahl ermittelt werden soll. Eine Modifikation ist die sogenannte Parallel Analysis, in der über das Eigenwertdiagramm der Daten ein zweites Diagramm mit Zufallsdaten gelegt wird. Es werden dann nur diejenigen Faktoren extrahiert, deren Eigenwerte höher sind als die Zufallswerte.

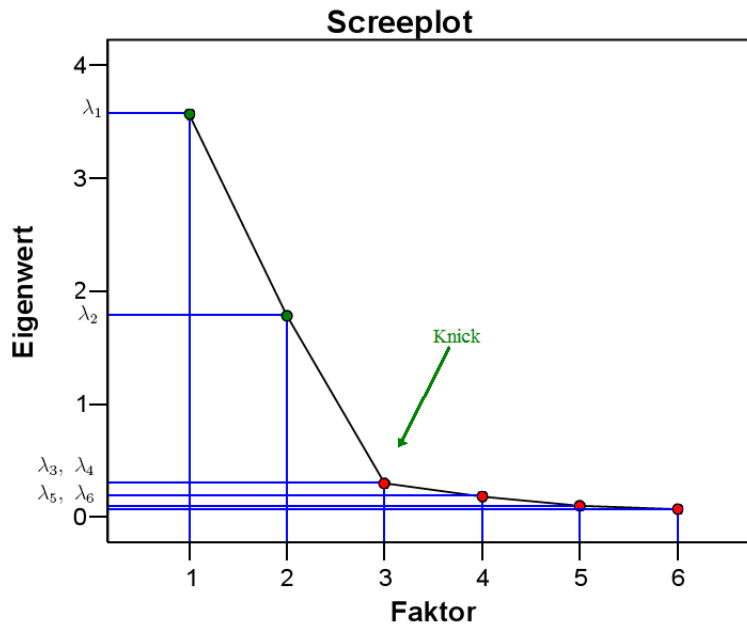


Abbildung 4.1: Screepplot-Beispiel

4.2.3 Das kummulierte Varianzkriterium

Nach diesem Kriterium werden so viele Faktoren extrahiert, bis ein vorgegebenes Maß an Varianz erklärt ist. Sollen beispielsweise 85% der Varianz erklärt werden, so werden so viele Faktoren extrahiert, bis dieses Maß an Erklärung erreicht ist, ganz gleich ob die Faktoren beispielsweise nach dem Kaiserkriterium zu extrahieren wären, weil sie $\lambda \leq 1$ aufweisen, oder nicht. Hiermit soll der Informationsverlust verringert werden, wenn die nach anderen Kriterien zu extrahierenden Faktoren einen zu geringen Teil der Varianz erklären. Ob dieses Verfahren sinnvoll ist, da evtl. zu viele relativ unbedeutende Faktoren extrahiert werden, sei dahingestellt.

4.3 Das Rotationsproblem

Das Verfahren der Faktorenextraktion ist so gestaltet, dass der erste Faktor einen möglichst grossen Teil der Varianz der manifesten Variablen erklärt. Der zweite Faktor soll einen möglichst grossen Teil der vom ersten Faktor noch nicht erklärten Varianz aufklären, und so weiter. Deshalb werden die Faktorladungen des zweiten Faktors im Schnitt absolut kleiner sein als die

Faktorladungen des ersten Faktors. Sie werden im mittleren Bereich von

$$|0,3| \leq a_{jk} \leq |0,7|$$

liegen. Diese Lösung eignet sich also nicht besonders, um sie sinnvoll zu interpretieren. Die Faktorstruktur ist insofern verwischt, als dass alle Variablen irgendwie mit den Faktoren in Beziehung zu stehen scheinen. Durch Rotation kann die Interpretation vereinfacht werden. Doch welchem Kriterium soll solch eine rotierter Lösung genügen? Hier treffen wir auf den Begriff der Einfachstruktur.

4.3.1 Die Einfachstruktur - Die Simple Structure

Ein Rotationsergebnis wird dann ausreichen, wenn die so erreichte Faktorenlösung eine sogenannte Simple Structure oder Einfachstruktur hat, d.h. wenn alle Faktoren möglichst zentral in ihren Variablenclustern liegen. Die 5 Kriterien nach Louis Leon Thurstones lauten wie folgt:

1. Jede Zeile der neuen Faktorladungsmatrix hat mindestens eine Null.
2. Bei empirisch sinnvollen Faktoren soll jede Spalte der Matrix mindestens m Nullen haben.
3. In jedem Spaltenpaar der Matrix sollte es eine Reihe von Variablen geben, die auf einem Faktor kaum, auf einem anderen hoch laden.
4. Bei 4 oder mehr Faktoren sollte in jedem Spaltenpaar eine Reihe der Variablen überhaupt nicht laden.
5. In einem Spaltenpaar sollte nur eine geringe Anzahl von Variablen vorhanden sein, die auf beiden Faktoren laden.

Auch wenn es schwer anzugeben ist, was die Mindestvoraussetzungen für eine Einfachstruktur sind, so kann man doch angeben, was *“die” einfachste Struktur* in einem r -Faktor m -Variablen Modell ist.

Die Faktorstruktur ist am einfachsten, wenn alle Variablen eine Faktorkomplexität von 1 haben, also alle Variablen nur auf einem einzigen Faktor laden. Für $r \geq 2$ Faktoren hat die simpelste Pattern Matrix

1. Nur ein einziges von Null verschiedenes Element pro Reihe
2. Jede Spalte beinhaltet einige Null-Elemente
3. Zwischen keinem Spaltenpaar überschneiden sich die von Null verschiedenen Elemente

Als Beispiel kann man sich eine Matrix vorstellen, die so aufgebaut ist wie die Einheitsmatrix (\mathbf{I} bzw. \mathbf{E}), z.B.:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Dies macht allerdings nicht unbedingt viel Sinn, weil wir dann so viele Faktoren extrahiert hätten (Anzahl der Spalten), wie wir Variablen haben (Anzahl der Zeilen). Unter dem Gesichtspunkt der Dimensionsreduktion ein unbefriedigendes Ergebnis. Wenn wir also weniger Faktoren erhalten, als Variablen, sagen wir 3 Faktoren bei 7 Variablen, dann schieben sich die Spalten sozusagen ineinander, z.B. so:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Dass r kleiner sein muss als m ergibt sich aus der Forderung nach Dimensionsreduktion. Mit "echten" Daten wird man so eine Einfachstruktur kaum jemals erblicken. Es wird also eine Hauptfrage, wie man eine Faktorstruktur so rotiert, dass sie der Einfachstruktur möglichst nahe kommt. Bei zweidimensionalen Lösungen, also zwei Faktoren kann man noch graphisch rotieren. Hierbei dreht man das Koordinatensystem so über die Punktcluster, bis die Achsen möglichst zentral in den Variablenclustern liegen. Bei höherdimensionalen Lösungen ist dies nicht mehr möglich und wir überlassen die aufwendigen Berechnungen dem Computer.

Nun zur Rotation. Warum ist es zulässig, eine Anfangslösung zu drehen? Durch Drehung des Koordinatensystems, in dem die Variablenvektoren (Vektoren, die die Faktorladungen jeweils einer Variablen enthalten: Zeilen der Faktormatrix) liegen, wird versucht eine Einfachstruktur zu erreichen. Diese Drehung ist zulässig, da die erklärten Varianzanteile der Variablen (Kommunalitäten) hierdurch unverändert bleiben. Es wird also nicht die aufgeklärte Varianz durch die Rotation nicht verändert, nur ihre Verteilung auf die Faktoren. Die Rotation der Faktoren erfolgt so, dass die Faktorladungen absolut

möglichst gross oder möglichst gering werden. Zur Verdeutlichung betrachten wir Graphik 4.2:

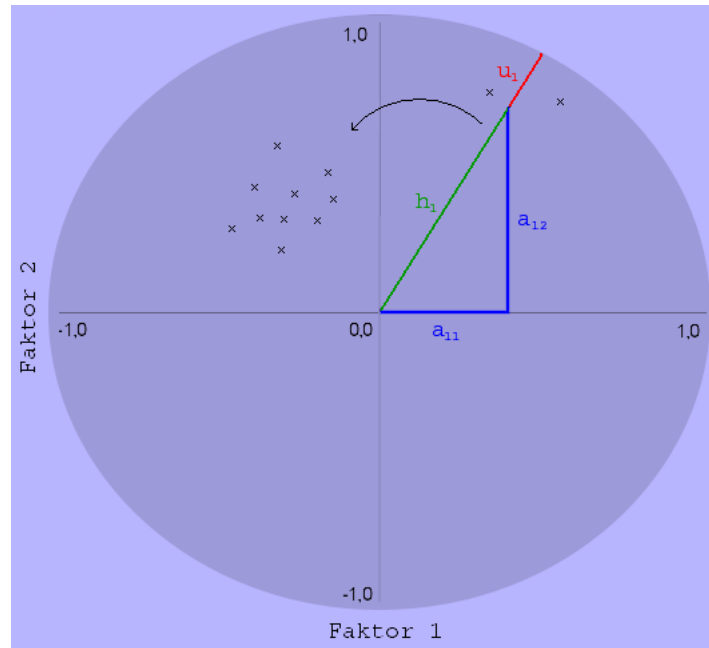


Abbildung 4.2: Geometrische Interpretation von Faktorladungen und Kommunalitäten

Da Kommunalität und Uniqueness addiert 1 ergeben, also $h_1 + u_j = 1$, haben wir in der Graphik einen Kreis mit Radius 1 (Einheitskreis) vor uns. Wir können also die anfängliche Lösung rotieren, ohne dass sich h_1 oder u_j in ihren Werten verändern, sie ergeben addiert immer 1. Die einzige Veränderung stellt sich bei a_{11} und a_{12} ein. Wir erinnern uns an den Satz des Pythagoras: $a^2 + b^2 = c^2$. Dieser gilt hier, da es sich bei dem Dreieck h_1 , a_{11} , a_{12} um ein Rechtwinkeliges handelt ebenso. Es gilt $a_{11}^2 + a_{12}^2 = h_1^2$. Es stehen zwei Arten der Rotation zur Verfügung: die orthogonale Rotation, in der die Faktoren in 90° Winkeln zueinander stehen. Die andere Möglichkeit besteht in der obliquen Rotation. Hier wird nicht an einer Orthogonalität der Faktoren festgehalten. Es ist möglich, dass die Faktoren in einer obliquen Rotation rechtwinkelig zueinander stehen (nicht miteinander korrelieren). Hier kann man sich sicher sein, dass diese Orthogonalität nicht im Verfahren der Rotation, sondern in den Daten begründet liegt. Die folgende Auflistung der

Rotationsverfahren erhebt keinen Anspruch auf Vollständigkeit. So sind einige Verfahren wie Orthoblique, Maxplane, Parsimax, Varisim oder Tandem nicht erwähnt.

4.3.2 Orthogonale Rotation

Bei orthogonalen Rotationsverfahren sind die Faktoren nach der Rotation unkorreliert. Diese Eigenschaft hat Vorteile, so ist zum Beispiel in der Literatur angegeben, dass das Ward's-Linkage Clusterverfahren orthogonale Variablen als Voraussetzung benötigt. Es gibt allerdings auch Nachteile, so zum Beispiel ist fraglich, ob die resultierenden Faktoren nun wirklich unkorreliert, also orthogonal zueinander sind, oder ob diese Eigenschaft ein vom Rotationsverfahren verursachtes Ergebnis ist.

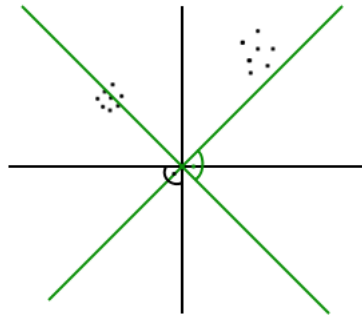


Abbildung 4.3: Orthogonale Rotation

Varimax

Das *Varimax*-Kriterium ist sehr bekannt und oft genutzt. Es maximiert die Varianz der quadrierten Faktorladungen für jede Spalte, also jeden Faktor (somit die Eigenwerte, in unserer Graphik 4.4 beispielhaft λ_1). Dies erleichtert die Interpretation der Faktoren. Die Summe der Faktorladungen pro Spalte kann Werte > 1 annehmen. Wir haben im Kapitel "Anzahl der Faktoren" gesehen, welche Bedeutung ein $\lambda_j > 1$ besitzt.

	F1	F2	F3	F4
Test1	a_{11}^2	a_{12}^2	a_{13}^2	a_{14}^2
Test2	a_{21}^2	a_{22}^2	a_{23}^2	a_{24}^2
Test3	a_{31}^2	a_{32}^2	a_{33}^2	a_{34}^2
Test4	a_{41}^2	a_{42}^2	a_{43}^2	a_{44}^2
Test5	a_{51}^2	a_{52}^2	a_{53}^2	a_{54}^2
	λ_1			

Abbildung 4.4: Varimax

Das Maß für die Einfachheit, auch bekannt als “row-Varimax-criterion” lautet:

$$V = \sum_{i=1}^m v_j = \frac{1}{n^2} \left(\sum_{j=1}^m n_j \sum_{i=1}^n a_{ij}^4 - \left(\sum_{i=1}^n a_{ij}^2 \right)^2 \right)$$

Quartimax

Ein anderes Verfahren der orthogonalen Rotation ist das *Quartimax*-Kriterium. Hier werden nicht die quadrierten Faktorladungen für jede Spalte maximiert, sondern für jede Reihe (also jede Variable). Dies ist gleichbedeutend mit der Maximierung der Kommunalität, in Graphik 4.5 beispielhaft h_1^2 . Dies soll die Erklärung der einzelnen Variablen erleichtern.

	F1	F2	F3	F4	
Test1	a_{11}^2	a_{12}^2	a_{13}^2	a_{14}^2	h_1^2
Test2	a_{21}^2	a_{22}^2	a_{23}^2	a_{24}^2	
Test3	a_{31}^2	a_{32}^2	a_{33}^2	a_{34}^2	
Test4	a_{41}^2	a_{42}^2	a_{43}^2	a_{44}^2	
Test5	a_{51}^2	a_{52}^2	a_{53}^2	a_{54}^2	

Abbildung 4.5: Quartimax

Die Idee, die hinter diesem Verfahren steht, ist diese: Die Variabilität ist am größten, wenn ein Element der quadrierten Faktorladungen gleich gross ist wie die Kommunalität, und alle anderen Ladungen in der Reihe gleich Null. Kurz, die maximale Varianz der quadrierten Faktorladungen für eine Variable i ist äquivalent zur größten faktoriellen Einfachheit für diese Variable. Quantifiziert man die faktorielle Einfachheit dieser Variable i erhält man:

$$q_i = \frac{\sum_{j=1}^m (a_{ij}^2 - \bar{a}_{ij}^2)^2}{m}$$

wobei m =Anzahl der Spalten in der pattern matrix. Es wird so Rotiert, dass

$$q = \sum_{i=1}^n q_i = \sum_{i=1}^n \frac{\sum_{j=1}^m (a_{ij}^4) - \left(\sum_{j=1}^m a_{ij}^2 \right)^2}{m^2}$$

also über alle Variablen maximiert wird. Da $\sum_{j=1}^m a_{ij}^2$ der Kommunalität entspricht und auch m Konstant ist, entspricht die Maximierung des gesamten Terms der Maximierung von

$$Q = \sum_{i=1}^n \sum_{j=1}^m a_{ij}^4$$

In der Anwendung betont dieses Verfahren die Einfachheit der Interpretation der Variablen auf Kosten der Einfachheit der Interpretation der Faktoren.

Equimax, Biquartimax

Wenn man bedenkt, dass das Varimax-Kriterium die Spalten vereinfacht, und das Quartimax-Kriterium die Reihen, so ist es nicht verwunderlich, dass einige gewichtete Mixturen aus beiden Verfahren entwickelt wurden.

$$\alpha \text{ Quartimax} + \beta \text{ Varimax} = \text{Maximum}$$

$$\sum_{j=1}^m \left(\sum_{i=1}^n a_{ij}^2 \right)^2 - \frac{1}{n} \left(\gamma \left(\sum_{i=1}^n a_{ij}^2 \right)^2 \right) = \text{Maximum}$$

Wobei $\gamma = \frac{\beta}{\alpha + \beta}$.

- $\gamma = 0 \rightarrow$ Quartimax
- $m = 1 \rightarrow$ Varimax
- $\gamma = \frac{m}{2} \rightarrow$ Equimax
- $\gamma = 0.5 \rightarrow$ Biquartimax

4.3.3 Oblique Rotation

Oblique Rotationsverfahren besitzen einen Vorteil gegenüber orthogonalen Verfahren: Man kann sich sicher sein, dass wenn eine orthogonale Lösung resultiert dies tatsächlich daran liegt, dass die Faktoren nicht korrelieren, und dies nicht bloss ein vom Verfahren aufgezwungenes Artefakt ist. Die Lösungen sind den Daten besser angepasst, können aber schwerer zu interpretieren sein. Desweiteren ist zu überlegen, ob oblique Verfahren nicht sinnvoller in der Anwendung sind, da so nach einer explorativen Faktorenanalyse mit obliquer Rotation sich die Konstruktion eines Strukturgleichungsmodells anbieten kann, dass auf Korrelation der Faktoren angewiesen ist (Siehe hierzu Graphiken 4.7 und 4.8). Bei unkorrelierten Faktoren würden alle Verbindungen zwischen den runden Elementen wegfallen, es liesse sich kein sogenanntes Strukturmodell aufstellen. Oblique Faktoren werden auch *Faktoren erster Ordnung* oder *Primärfaktoren* genannt. Lässt man über die Korrelationsmatrix der Faktoren eine weitere Faktorenanalyse laufen, und erhält ein interpretierbares Ergebnis, so hat man *Faktoren zweiter Ordnung* oder *Sekundärfaktoren* ermittelt. Die folgenden Ansätze basieren auf der Annahme, dass definierte Cluster von Variablen, die separate Dimensionen repräsentieren, vorhanden sind. Wenn diese Cluster in der Anfangslösung korrekt identifiziert werden, dann wird jedes Cluster von Variablen Ladungen von nahezu Null auf allen bis auf eine Referenz-Achse haben. Zur Verdeutlichung Graphik 4.9

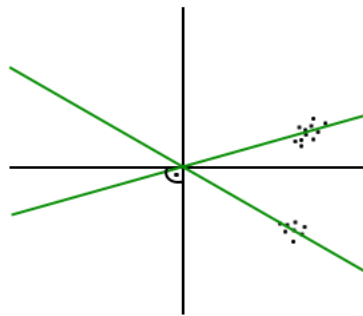


Abbildung 4.6: Oblique Rotation

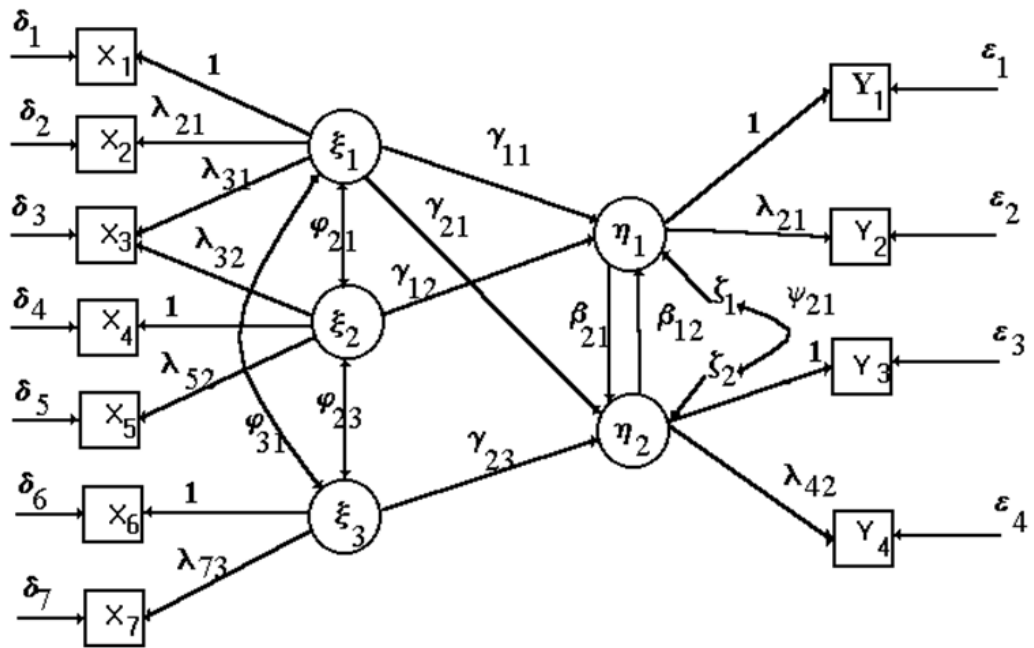


Abbildung 4.7: Beispiel: Strukturgleichungsmodell 1

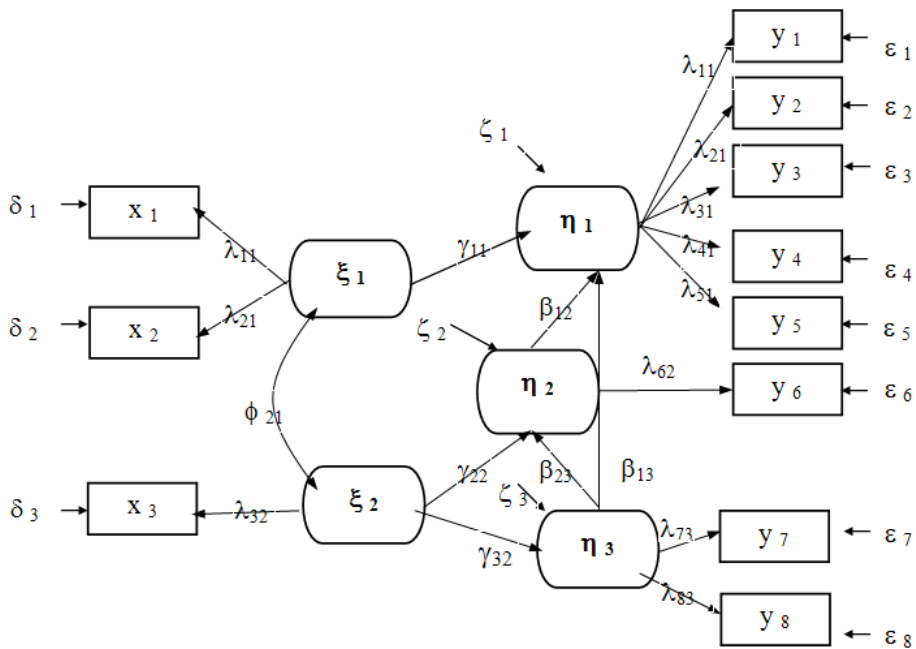


Abbildung 4.8: Beispiel: Strukturgleichungsmodell 2

Quartimin

Das Quartimin-Kriterium, ist das oblique Pendant zum orthogonalen Quartimax-Kriterium. Es ist definiert als:

$$N = \sum_{i=1}^n \sum_{j < k=1}^m w_{ij}^2 w_{ik}^2$$

wobei w_{ij} und w_{ik} Projektionen auf der j -ten und k -ten Referenz-Achse darstellen. Was wir suchen, ist diejenige oblique Rotation, die die Faktorladungen liefert, die N minimieren

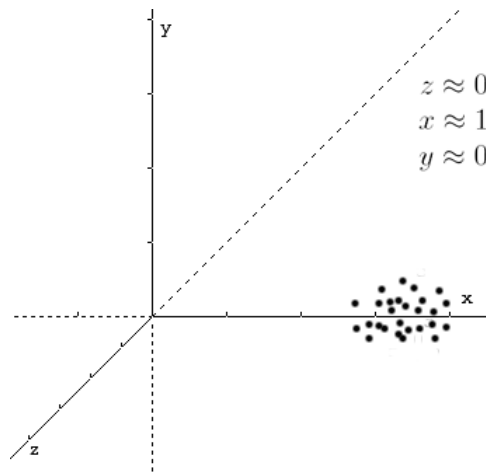


Abbildung 4.9: Idee Referenzachsen

Covarimin und Biquartimin

Parallel zu dem Unterschied zwischen den orthogonalen Rotationen Varimax und Quartimax existieren auch in den obliquen Rotationen Varimax-ähnliche Gegenstücke zu dem Quartimax-ähnlichen Quartimin. Hierbei handelt es sich um die Verfahren Covarimin und Biquartimin. Der Wert der hier minimiert wird ist die Kovarianz der quadrierten Elemente der Projektionen auf die Referenz-Achsen.

$$C = \sum_{j < k=1}^m \left(n \sum_{i=1}^n w_{ij}^2 w_{ik}^2 - \sum_{i=1}^n w_{ij}^2 \sum_{i=1}^n w_{ik}^2 \right)$$

Wenn beide Verfahren auf die selben Daten angewendet werden, so tendiert das Covarimin-Kriterium dazu, Faktoren zu finden, die weniger oblique sind, also eher orthogonal. Das Biquartimin-Kriterium hingegen findet Faktoren, die stärker oblique sind, also weniger orthogonal.

Indirektes Oblimin

Auch hier existiert eine Mixtur aus beiden Verfahren:

$$B = \alpha N + \beta \frac{C}{n} = \text{Minimum}$$

$$B = \sum_{j < k=1}^m \left(n \sum_{i=1}^n w_{ij}^2 w_{ik}^2 - \gamma \sum_{i=1}^n w_{ij}^2 \sum_{i=1}^n w_{ik}^2 \right)$$

wobei $\gamma = \frac{\beta}{\alpha + \beta}$

- $\gamma = 0 \rightarrow$ Quartimin, am stärksten oblique
- $\gamma = 0.5 \rightarrow$ Biquartimin, weniger oblique
- $\gamma = 1 \rightarrow$ Covarimin, am wenigsten oblique

Binormamin

Ein weiteres Kriterium ist Binormamin. Es versucht γ als Korrektur für den zu obliquen Bias des Quartimin-Kriteriums und des zu orthogonalen Bias des Covarimin-Kriteriums zu nutzen. Im Vergleich mit dem Biquartimin-Kriterium soll es bessere Ergebnisse liefern, wenn die Daten entweder sehr komplex oder sehr simpel sind.

Direktes Oblimin

Dieses Kriterium bedient sich nicht der Idee der Referenzachsen. Hier werden die Faktorladungen der Anfangsfaktoren vereinfacht, Referenzachsen spielen keine Rolle.

$$D = \sum_{j=k=1}^m \left(\sum_{i=1}^n a_{ij}^2 a_{ik}^2 - \frac{d}{n} \left(\sum_{i=1}^n a_{ij}^2 \sum_{i=1}^n a_{ik}^2 \right) \right)$$

Auch hier kann man den Grad der Obliqueheit durch Variation von d bestimmen. Je größer d gewählt wird, desto stärker oblique wird die Lösung sein. Wenn nur ein Faktor existiert, dann liefert $d = 0$ die korrekte Lösung.

Oblimax

Das Oblimax-Kriterium basiert auf dem Versuch, die Anzahl der kleinen und grossen Ladungen zu maximieren, auf Kosten der schlecht zu interpretierenden mittleren Faktorladungen. Dieses Kriterium ist äquivalent zur orthogonalen Quartimax-Lösung, aber führt zu anderen Ergebnissen als das Quartimin-Verfahren (dem obliquen Gegenstück zu Quartimax), wenn es ohne Restriktionen zu orthogonalen Achsen verwendet wird.

Promax

Ein anderer Ansatz steckt hinter der Promax-Methode. Hierbei handelt es sich um eine Rotation, die auf den besten fit mit einer vorher definierte Zielmatrix abzielt. Die Matrix, auf die hin rotiert werden soll ist eine vorherigen orthogonalen Lösung. Die obliquen Faktoren, deren "fit" am besten ist sind die gesuchte Lösung der promax-Rotation.

4.3.4 Überlegungen zur Rotation

Bei genauer Betrachtung kann das kundige Auge eine Besonderheit erkennen, die wir schon aus einem anderen Teil der Faktorenanalyse kennen. So wie die Principal Components Analysis versucht, die Varianzen der Variablen zu erklären und die Hauptachsen-Faktorenanalyse die Kovarianzen, so scheinen auch die verschiedenen Rotationsverfahren diesen Weg zu beschreiben. Schauen wir uns die orthogonalen und obliquen Verfahren einmal unter diesem Gesichtspunkt an.

Orthogonale Verfahren

Betrachten wir einmal einen Teil der Formel des Quartimax-Kriteriums genauer:

$$\frac{\sum_{j=1}^m (a_{ij}^4) - \left(\sum_{j=1}^m a_{ij}^2 \right)^2}{m^2}$$

Die Ähnlichkeit zur Formel für die Varianz ist nicht auf den ersten Blick zu sehen. Nach einigen Umformungen ist sie jedoch schon augenfälliger:

Wenn wir $a_{ij}^2 = x_i$ und $m^2 = n$ setzen erhalten wir

$$\frac{\sum x_i^2 - (\sum x_i)^2}{n}$$

Wie hängt nun

$$\frac{\sum (x_i - \bar{x})^2}{n}$$

mit der Form

$$\frac{\sum x_i^2 - (\sum x_i)^2}{n}$$

zusammen?

Durch einige Umformungen sehen wir:

$$\begin{aligned} s_x^2 &= \frac{\sum (x_i - \bar{x})^2}{n} \\ &= \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} \\ &= \frac{\sum x_i^2}{n} - \frac{\sum 2x_i\bar{x}}{n} + \frac{\sum \bar{x}^2}{n} \\ &= \overline{x^2} - \frac{\sum 2x_i}{n} \frac{\sum \bar{x}}{n} + \frac{n\bar{x}^2}{n} \\ &= \overline{x^2} - \frac{\sum 2x_i}{n} \frac{n\bar{x}}{n} + \frac{n\bar{x}^2}{n} \\ &= \overline{x^2} - \frac{2\sum x_i}{n} \bar{x} + \bar{x}^2 \\ &= \overline{x^2} - 2\bar{x} \cdot \bar{x} + \bar{x}^2 \\ &= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 \\ &= \overline{x^2} - \bar{x}^2 \\ &= \frac{\sum x_i^2}{n} - \frac{(\sum x_i)^2}{n} \\ &= \frac{\sum x_i^2 - (\sum x_i)^2}{n} \end{aligned}$$

Beispielhaft am Quartimax-Kriterium haben wir gesehen, dass die Maximierung der Varianz eine Rolle spielt. Nun zu den obliquen Verfahren.

Oblique Verfahren

Hier beschäftigen wir uns exemplarisch mit dem Covarimin-Kriterium. In der Formel befindet sich folgender Term.

$$\sum_{i=1}^n w_{ij}^2 w_{ik}^2 - \sum_{i=1}^n w_{ij}^2 \sum_{i=1}^n w_{ik}^2$$

Setzen wir nun $w_{ij}^2 = x_i$ und $w_{ik}^2 = y_i$ erhalten wir

$$\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

Wie hängt dies nun mit der Kovarianz zusammen? Die Formel für die Kovarianz lautet:

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Formen wir nun genau so um wie im vorherigen Beispiel erhalten wir folgendes:

$$\begin{aligned} s_{xy} &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \\ &= \frac{\sum x_i y_i - \sum x_i \bar{y} - \sum \bar{x} y_i + \sum \bar{x} \bar{y}}{n} \\ &= \frac{\sum x_i y_i}{n} - \frac{\sum x_i \bar{y}}{n} - \frac{\sum \bar{x} y_i}{n} + \frac{\sum \bar{x} \bar{y}}{n} \\ &= \bar{x} \bar{y} - \frac{\sum x_i \sum \bar{y}}{n} - \frac{\sum \bar{x} \sum y_i}{n} + \frac{\sum \bar{x} \sum \bar{y}}{n} \\ &= \bar{x} \bar{y} - \bar{x} \frac{n \bar{y}}{n} - \frac{n \bar{x}}{n} \bar{y} + \bar{x} \bar{y} \\ &= \bar{x} \bar{y} - \bar{x} \bar{y} - \bar{y} \bar{x} + \bar{x} \bar{y} \\ &= \bar{x} \bar{y} - \bar{x} \bar{y} - \bar{x} \bar{y} + \bar{x} \bar{y} \\ &= \bar{x} \bar{y} - 2 \bar{x} \bar{y} + \bar{x} \bar{y} \\ &= \bar{x} \bar{y} - \bar{x} \bar{y} \\ &= \frac{\sum xy}{n} - \frac{\sum x \sum y}{n} \\ &= \frac{\sum xy - \sum x \sum y}{n} \end{aligned}$$

An dieser Stelle erkennen wir die Ähnlichkeit zur Formel der Kovarianz.

4.4 Bestimmung der Faktorwerte

Es ist nicht nur interessant, die Variablen auf eine geringere Anzahl von Faktoren zu reduzieren, auch die Werte, die die Objekte hinsichtlich der extrahierten Faktoren annehmen ist von Interesse. Als *Faktorwerte* bezeichnet man also die Ausprägung der Faktoren bei den Objekten. Wie in Kapitel 1 erläutert ist das Ziel der Faktorenanalyse, die standardisierte Ursprungsmatrix \mathbf{Z} als Linearkombination von Faktoren darzustellen. Es galt:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}'$$

- \mathbf{Z} =Matrix der standardisierten Daten
- \mathbf{F} =Matrix der Faktorwerte
- \mathbf{A} =Matrix der Faktorladungen

Wir haben uns bisher mit der Bestimmung der Faktorladungsmatrix \mathbf{A} beschäftigt. \mathbf{Z} ist als standardisierte Ursprungsdatenmatrix gegeben. Also müssen wir die Gleichung nach \mathbf{F} auflösen. Also:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}'$$

Rechtsseitiges Multiplizieren mit $(\mathbf{A}')^{-1}$ ergibt

$$\mathbf{Z}(\mathbf{A}')^{-1} = \mathbf{F}\mathbf{A}'(\mathbf{A}')^{-1}$$

Da nach Definition $\mathbf{A}'(\mathbf{A}')^{-1} = \mathbf{I}$ folgt

$$\mathbf{Z}(\mathbf{A}')^{-1} = \mathbf{F}\mathbf{I}$$

Da $\mathbf{F}\mathbf{I} = \mathbf{F}$ ergibt sich die folgende Gleichung

$$\mathbf{F} = \mathbf{Z}(\mathbf{A}')^{-1}$$

Ein Problem ist, dass es sich bei der Matrix \mathbf{A} nicht um eine Quadratische handelt, da es ja ein fundamentales Ziel der Faktorenanalyse ist, weniger Faktoren zu extrahieren als Variablen vorhanden sind. Es handelt sich um eine $i \times j$ -Matrix, wobei i die Anzahl der Variablen anzeigt, und j die Anzahl der Faktoren, wobei unter dem Gesichtspunkt der Dimensionsreduktion $j < i$ gilt. Wir gehen also folgendermaßen vor:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}'$$

Multiplikation von rechts mit \mathbf{A} ergibt:

$$\mathbf{Z}\mathbf{A} = \mathbf{F}\mathbf{F}'\mathbf{A}$$

Die Matrix $(\mathbf{A}'\mathbf{A})$ ist definitionsmäßig quadratisch, und somit prinzipiell invertierbar, sofern $\det(\mathbf{A}'\mathbf{A}) \neq 0$. Im nächsten Schritt multiplizieren wir rechtsseitig mit $(\mathbf{A}'\mathbf{A})^{-1}$

$$\mathbf{Z}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{F}(\mathbf{A}'\mathbf{A})(\mathbf{A}'\mathbf{A})^{-1}$$

Da $\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$ gilt auch $(\mathbf{A}'\mathbf{A})(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{I}$ Wir erhalten also

$$\mathbf{Z}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{F}\mathbf{I}$$

$$\mathbf{F} = \mathbf{Z}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}$$

Nicht in allen Fällen ist das Lösen dieser Gleichung unproblematisch. In diesen Fällen werden die Faktorwerte geschätzt. Eine Möglichkeit ist hierbei die Schätzung durch multiple Regression, detailliert nachzulesen im Buch von Karl Überla.

4.5 Interpretation der Faktoren

Ein "Problem" der Faktorenanalyse liegt darin, dass das Verfahren nur Faktoren liefert. Die Interpretation der Faktoren ist Sache des Forschers. Nur wenn die Faktoren inhaltlich sinnvoll sind, ist die Faktorenanalyse erfolgreich gewesen. Wir betrachten ein Beispiel aus dem Sport:

Wir versuchen folgende 10 Sportarten auf einige wenige erklärende Faktoren zu reduzieren.:

- 100m Sprint
- Weitsprung
- Kugelstoßen
- Hochsprung
- 400m Lauf
- 110m Hürden

- Diskuswurf
- Stabhochsprung
- 1500m Lauf

Da die Anfangslösung in der Regel nicht zu interpretieren ist betrachten wir hier die rotierte Lösung:

	Faktor1	Faktor2	Faktor3	Faktor4	h^2
100m	−.03	.94	−.01	−.18	.91
Weit	.07	.84	.21	−.06	.75
Kugel	.88	.20	.01	−.27	.88
Hoch	−.30	.02	.78	−.06	.70
400m	.03	.84	.04	.42	.89
Hürden	.07	.38	.68	.01	.61
Diskus	.90	−.05	.11	−.07	.83
Stab	.43	−.12	.69	.03	.67
Speer	.76	−.01	−.08	.22	.62
1500m	−.05	.00	−.02	.97	.95

Die Variablen, die hoch auf dem entsprechenden Faktor laden sind rot markiert. Nun fassen wir die markierten Variablen pro Faktor zwecks Interpretation zusammen. Wir erkennen: Die Variablen, die auf dem ersten Faktor hoch laden sind alle Wurfdisziplinen. Also scheint es sich bei dem ersten Faktor um einen Wurf-Faktor zu handeln. Der zweite Faktor hängt stark mit Sprintdisziplinen zusammen. Der zweite Faktor ist also ein Sprintfaktor. Der dritte Faktor bezieht sich auf Sprungkraft. Der vierte Faktor hängt sehr stark mit dem 1500m Lauf zusammen. Hierbei handelt es sich um einen Ausdauer-Faktor. Dieses Beispiel ist recht einfach zu interpretieren. Es ist aber nicht schwer einzusehen, dass kompliziertere Lösungen dem Forscher eine größere Anstrengung bei der Benennung der Faktoren abverlangen.

Kapitel 5

SPSS-Beispiele

5.1 Erziehung

Betrachten wir einmal ein handfestes Beispiel. In Graphik 5.1 sehen wir die Korrelationsmatrix von 9 Variablen die sich mit der Beurteilung der Wichtigkeit zu verschiedenen Zielen der Erziehung befassen.

Korrelationen

		ERZIEHUNGSZIEL: SELBSTAENDIGKEIT	ERZIEHUNGSZIEL: SELBSTVERTRAUEN	ERZIEHUNGSZIEL: GUTE UMGANGSFORMEN	ERZIEHUNGSZIEL: DURCHSETZUNGSFAEHIIGKEIT	ERZIEHUNGSZIEL: GUTE SCHULLEISTUNGEN	ERZIEHUNGSZIEL: VERANTWORTUNGSB EW.	ERZIEHUNGSZIEL: KRITIKFAEHIGKEIT	ERZIEHUNGSZIEL: VERSTAE NDNIS F. ANDERE	ERZIEHUNGSZIEL: FLEISS
ERZIEHUNGSZIEL: SELBSTAENDIGKEIT	Korrelation nach Pearson	1	,618**	,305**	,462**	,197**	,476**	,449**	,425**	,231**
	Signifikanz (2-seitig)		,000	,000	,000	,000	,000	,000	,000	,000
	N	4444	4442	4441	4443	4439	4444	4434	4438	4435
ERZIEHUNGSZIEL: SELBSTVERTRAUEN	Korrelation nach Pearson	,618**	1	,334**	,487**	,205**	,474**	,444**	,442**	,226**
	Signifikanz (2-seitig)	,000		,000	,000	,000	,000	,000	,000	,000
	N	4442	4442	4439	4441	4437	4442	4433	4437	4433
ERZIEHUNGSZIEL: GUTE UMGANGSFORMEN	Korrelation nach Pearson	,305**	,334**	1	,421**	,520**	,415**	,259**	,346**	,557**
	Signifikanz (2-seitig)	,000	,000		,000	,000	,000	,000	,000	,000
	N	4441	4439	4441	4440	4437	4441	4431	4435	4433
ERZIEHUNGSZIEL: DURCHSETZUNGSFAEHIIGKEIT	Korrelation nach Pearson	,462**	,487**	,421**	1	,360**	,482**	,478**	,391**	,354**
	Signifikanz (2-seitig)	,000	,000	,000		,000	,000	,000	,000	,000
	N	4443	4441	4440	4443	4438	4443	4433	4437	4434
ERZIEHUNGSZIEL: GUTE SCHULLEISTUNGEN	Korrelation nach Pearson	,197**	,205**	,520**	,360**	1	,325**	,155**	,196**	,609**
	Signifikanz (2-seitig)	,000	,000	,000	,000		,000	,000	,000	,000
	N	4439	4437	4437	4438	4440	4439	4429	4433	4434
ERZIEHUNGSZIEL: VERANTWORTUNGSB EW.	Korrelation nach Pearson	,476**	,474**	,415**	,482**	,325**	1	,540**	,527**	,401**
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000		,000	,000	,000
	N	4444	4442	4441	4443	4439	4444	4434	4438	4435
ERZIEHUNGSZIEL: KRITIKFAEHIGKEIT	Korrelation nach Pearson	,449**	,444**	,259**	,478**	,155**	,540**	1	,565**	,225**
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,000		,000	,000
	N	4434	4433	4431	4433	4429	4434	4434	4430	4428
ERZIEHUNGSZIEL: VERSTAE NDNIS F. ANDERE	Korrelation nach Pearson	,425**	,442**	,346**	,391**	,196**	,527**	,565**	1	,312**
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,000	,000		,000
	N	4438	4437	4435	4437	4433	4438	4430	4438	4431
ERZIEHUNGSZIEL: FLEISS	Korrelation nach Pearson	,231**	,226**	,557**	,354**	,609**	,401**	,225**	,312**	1
	Signifikanz (2-seitig)	,000	,000	,000	,000	,000	,000	,000	,000	
	N	4435	4433	4433	4434	4434	4435	4428	4431	4436

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Abbildung 5.1: Korrelation der Erziehungsziel-Variablen

Diese Matrix wollen wir einer Maximum Likelihood-Faktorenanalyse unterziehen. Durch die Dimensionsreduktion, die uns hoffentlich - auf die zu Grunde liegenden Faktoren führt, erhalten wir eine handhabbare Menge an Faktoren als wir nun als Variablen vor uns haben.

Erklärte Gesamtvarianz

Faktor	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	4,186	46,506	46,506	3,695	41,051	41,051
2	1,458	16,203	62,709	1,034	11,489	52,540
3	,747	8,299	71,008			
4	,561	6,229	77,237			
5	,481	5,343	82,580			
6	,440	4,888	87,468			
7	,390	4,329	91,797			
8	,377	4,192	95,989			
9	,361	4,011	100,000			

Extraktionsmethode: Maximum Likelihood.

Abbildung 5.2: Erklärte Gesamtvarianz

In der Graphik 5.2 sehen wir, dass genau so viele Faktoren existieren, wie Variablen vorhanden sind, nämlich 9. Wir bedienen uns allerdings nicht aller Faktoren, sondern wir benutzen nur einen Teil dieser Faktoren. Nur die ersten zwei Faktoren verfügen über einen Eigenwert größer als eins. Es werden nach dem Kaiser-Kriterium also nur 2 Faktoren extrahiert. Wieviel % der gesamten Varianz die einzelnen Faktoren erklären, wird über $\frac{\lambda_j}{m}$ errechnet, wobei m hier 11 ist, da 11 Variablen vorliegen. Eine andere Möglichkeit, die Anzahl der Faktoren zu bestimmen ist der Screeplot.

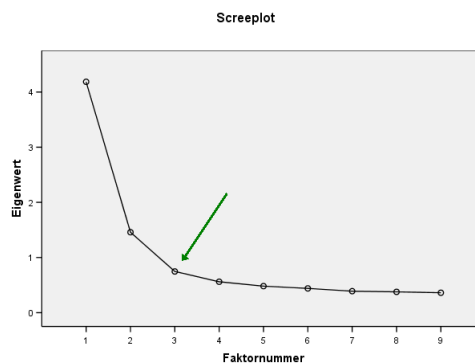


Abbildung 5.3: Screeplot der Erziehungsvariablen

Wir extrahieren die Faktoren, die sich links des Knickes befinden. Es handelt sich hier um eine 2–Faktoren Lösung.

Faktorenmatrix^a

	Faktor	
	1	2
ERZIEHUNGSZIEL: SELBSTAENDIGKEIT	,627	,331
ERZIEHUNGSZIEL: SELBSTVERTRAUEN	,636	,332
ERZIEHUNGSZIEL: GUTE UMGANGSFORMEN	,647	-,301
ERZIEHUNGSZIEL: DURCHSETZUNGSFA EHIGKEIT	,673	,104
ERZIEHUNGSZIEL: GUTE SCHULLEISTUNGEN	,568	-,519
ERZIEHUNGSZIEL: VERANTWORTUNGSB EW.	,724	,146
ERZIEHUNGSZIEL: KRITIKFAEHIGKEIT	,617	,360
ERZIEHUNGSZIEL: VERSTAENDNIS F. ANDERE	,630	,246
ERZIEHUNGSZIEL: FLEISS	,634	-,482

Extraktionsmethode: Maximum-Likelihood.

a. 2 Faktoren extrahiert. Es werden 4 Iterationen benötigt.

Abbildung 5.4: Faktormatrix Erziehungsvariablen

Es werden also 2 Faktoren extrahiert. In Graphik 5.5 und 5.6 können wir uns anschauen, wie die Variablen in diesem von den Faktoren aufgespannten zweidimensionalen Raum liegen. Diese Faktoren können interpretiert werden. Es handelt sich hierbei *nicht* um die Anfangslösung, die erst noch rotieren werden müsste, sie ist bereits rotiert. Aus der Vielzahl von Rotationskriterien, die uns zur Verfügung stehen haben wir das Varimax-Kriterium ausgewählt. Dieses Verfahren vereinfacht insbesondere die Interpretation der Faktoren.

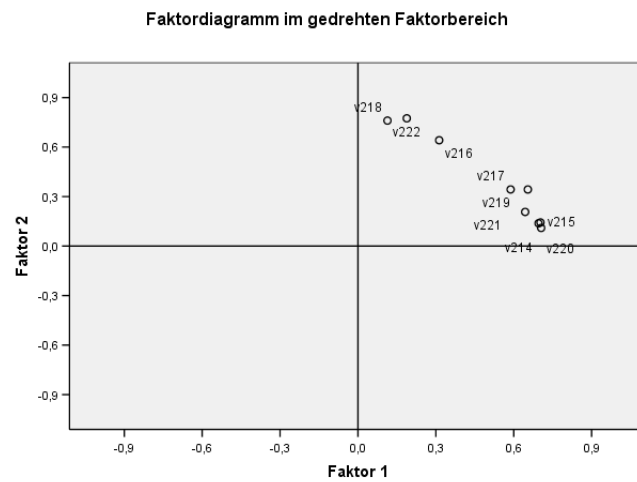


Abbildung 5.5: Rotierte Faktoren in SPSS

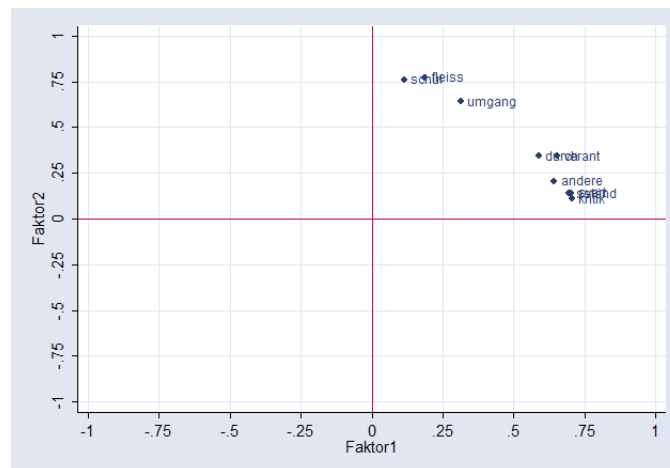


Abbildung 5.6: Rotierte Lösung (nicht SPSS)

Rotierte Faktormatrix^a

	Faktor	
	1	2
ERZIEHUNGSZIEL: SELBSTAENDIGKEIT	,695	,138
ERZIEHUNGSZIEL: SELBSTVERTRAUEN	,703	,143
ERZIEHUNGSZIEL: GUTE UMGANGSFORMEN	,313	,642
ERZIEHUNGSZIEL: DURCHSETZUNGSFA EHIGKEIT	,588	,343
ERZIEHUNGSZIEL: GUTE SCHULLEISTUNGEN	,113	,760
ERZIEHUNGSZIEL: VERANTWORTUNGSB EW.	,654	,343
ERZIEHUNGSZIEL: KRITIKFAEHIGKEIT	,705	,110
ERZIEHUNGSZIEL: VERSTAENDNIS F. ANDERE	,644	,207
ERZIEHUNGSZIEL: FLEISS	,188	,774

Extraktionsmethode: Maximum-Likelihood.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.
a. Die Rotation ist in 3 Iterationen konvergiert.

Abbildung 5.7: Rotierte Faktormatrix Erziehungsvariablen

Die rotierten Faktoren sind besser zu interpretieren als die ursprüngliche Lösung. Die Werte, die in der rotierten Faktormatrix stehen können wir nun verwenden, um die Faktoren inhaltlich zu interpretieren. Die Variablen, die mit dem ersten Faktor in Zusammenhang stehen lassen sich alle mit dem Erziehungsziel der Vermittlung eines *positiven Selbstkonzeptes* beschreiben. Die Variablen, die mit dem zweiten Faktor in Beziehung stehen lassen sich als Variablen subsumieren, die eher in Richtung einer *positiven Aussenwirkung* zielen.

5.2 Wichtigkeit

In unserem zweiten SPSS-Beispiel betrachten wir 8 Variablen die sich auf das Vertrauen zu verschiedenen Bereichen des öffentlichen Lebens beziehen.

		Korrelationen							
		VERTRAUEN: BUNDESVER FASSUNGSG GERICHT	VERTRAUEN: BUNDESTAG	VERTRAUEN: KATHOLISC H KIRCHE	VERTRAUEN: EVANGELISC HE KIRCHE	VERTRAUEN: JUSTIZ	VERTRAUEN: FERNSEHEN	VERTRAUEN: ZEITUNGSW ESEN	VERTRAUEN: BUNDESREG IERUNG
VERTRAUEN: BUNDESVERFASSUNGSG GERICHT	Korrelation nach Pearson Signifikanz (2-seitig) N	1 .000 8357	.547** .000 8328	.285** .000 8111	.274** .000 8137	.493** .000 8330	.167** .000 8329	.186** .000 8322	.417** .000 8332
VERTRAUEN: BUNDESTAG	Korrelation nach Pearson Signifikanz (2-seitig) N	.547** .000 8328	1 .000 8521	.382** .000 8233	.337** .000 8261	.485** .000 8480	.283** .000 8491	.280** .000 8479	.375** .000 8492
VERTRAUEN: KATHOLISCHE KIRCHE	Korrelation nach Pearson Signifikanz (2-seitig) N	.285** .000 8111	.382** .000 8233	1 .000 8307	.743** .000 8181	.342** .000 8259	.245** .000 8277	.219** .000 8269	.387** .000 8268
VERTRAUEN: EVANGELISCHE KIRCHE	Korrelation nach Pearson Signifikanz (2-seitig) N	.274** .000 8137	.337** .000 8251	.743** .000 8181	1 .000 8330	.348** .000 8291	.252** .000 8303	.221** .000 8291	.340** .000 8293
VERTRAUEN: JUSTIZ	Korrelation nach Pearson Signifikanz (2-seitig) N	.493** .000 8330	.485** .000 8480	.342** .000 8259	.348** .000 8291	1 .000 8559	.302** .000 8528	.283** .000 8512	.455** .000 8500
VERTRAUEN: FERNSEHEN	Korrelation nach Pearson Signifikanz (2-seitig) N	.167** .000 8329	.283** .000 8491	.245** .000 8277	.252** .000 8303	.302** .000 8528	1 .000 8589	.639** .000 8550	.312** .000 8536
VERTRAUEN: ZEITUNGSWESSEN	Korrelation nach Pearson Signifikanz (2-seitig) N	.186** .000 8322	.260** .000 8479	.219** .000 8269	.221** .000 8291	.283** .000 8512	.639** .000 8550	1 .000 8572	.283** .000 8524
VERTRAUEN: BUNDESREGIERUNG	Korrelation nach Pearson Signifikanz (2-seitig) N	.417** .000 8332	.375** .000 8492	.387** .000 8266	.340** .000 8293	.455** .000 8500	.312** .000 8536	.283** .000 8524	1 .000 8571

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Abbildung 5.8: Korrelationsmatrix der Vertrauensvariablen

Diese 8 Variablen werden daraufhin untersucht, wieviele Faktoren für ihre Korrelation verantwortlich sind. Wir sehen, dass wir es mit 3 Faktoren mit einem Eigenwert größer als 1 zu tun haben. Wenn wir den Screeplot betrachten kommen wir zu dem selben Ergebnis, nämlich 3 Faktoren zu extrahieren.

Erklärte Gesamtvarianz									
Faktor	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	3,572	44,652	44,652	2,906	36,326	36,326	2,038	25,473	25,473
2	1,259	15,733	60,385	1,155	14,440	50,766	1,538	19,224	44,698
3	1,112	13,902	74,287	.871	10,884	61,649	1,356	16,952	61,649
4	.647	8,085	82,372						
5	.499	6,239	88,610						
6	.362	4,529	93,140						
7	.298	3,719	96,859						
8	.251	3,141	100,000						

Extraktionsmethode: Maximum Likelihood.

Abbildung 5.9: Erklärte Gesamtvarianz der Vertrauensvariablen

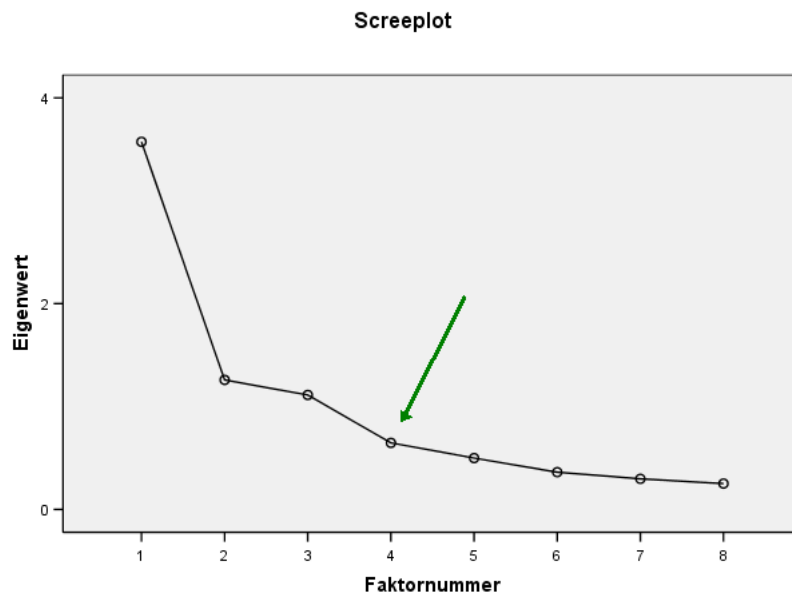


Abbildung 5.10: Screepplot Vertrauensvariablen

Die Anfangslösung der Faktoren wird auch hier nicht interpretiert. Erst nach einer Rotation der Faktoren hin zu einer Einfachstruktur werden sie inhaltlich interpretiert, da sie dann viel einfacher zu handhaben sind.

Faktorenmatrix^a

	Faktor		
	1	2	3
VERTRAUEN: BUNDESVERFASSUNGS GERICHT	,452	,297	,325
VERTRAUEN: BUNDESTAG	,601	,469	,419
VERTRAUEN: KATHOLISCHE KIRCHE	,775	-,188	,018
VERTRAUEN: EVANGELISCHE KIRCHE	,877	-,354	-,053
VERTRAUEN: JUSTIZ	,522	,287	,134
VERTRAUEN: FERNSEHEN	,454	,498	-,544
VERTRAUEN: ZEITUNGSWESEN	,401	,432	-,435
VERTRAUEN: BUNDESREGIERUNG	,578	,412	,288

Extraktionsmethode: Maximum-Likelihood.

a. Es wurde versucht, 3 Faktoren zu extrahieren. Es werden mehr als 25 Iterationen benötigt. (Konvergenz=,004). Die Extraktion wurde abgebrochen.

Abbildung 5.11: Faktormatrix Vertrauensvariablen

Hier sehen wir, dass die rotierte Lösung inhaltlich viel besser “zu fassen” ist als die vorherige unrotierte Lösung. Der erste Faktor lässt sich als *Vertrauen in das politische System* oder *Vertrauen in den Rechtsstaat* beschreiben. Der zweite Faktor bezieht sich ganz klar auf *Vertrauen in die Kirche* oder *Vertrauen in Gott* und der dritte Faktor bezieht sich auf *Vertrauen in die Medien*.

Rotierte Faktorenmatrix^a

	Faktor		
	1	2	3
VERTRAUEN: BUNDESVERFASSUNGS GERICHT	,608	,154	,067
VERTRAUEN: BUNDESTAG	,842	,162	,142
VERTRAUEN: KATHOLISCHE KIRCHE	,294	,729	,134
VERTRAUEN: EVANGELISCHE KIRCHE	,204	,917	,128
VERTRAUEN: JUSTIZ	,514	,234	,230
VERTRAUEN: FERNSEHEN	,166	,120	,841
VERTRAUEN: ZEITUNGSWESEN	,171	,106	,705
VERTRAUEN: BUNDESREGIERUNG	,713	,189	,201

Extraktionsmethode: Maximum-Likelihood.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 4 Iterationen konvergiert.

Abbildung 5.12: Rotierte Faktormatrix Vertrauensvariablen

Wir können uns in folgender Graphik anschauen, wie die Variablen im dreidimensionalen Faktorraum liegen.

Zuerst die unrotierte Lösung:

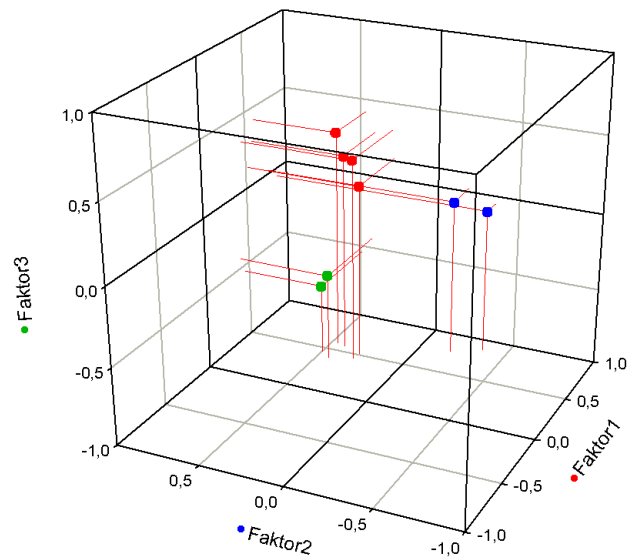


Abbildung 5.13: Dreidimensionaler unrotierter Faktorraum (nicht SPSS)

Nach der rechtwinkligen Varimax-Rotation:

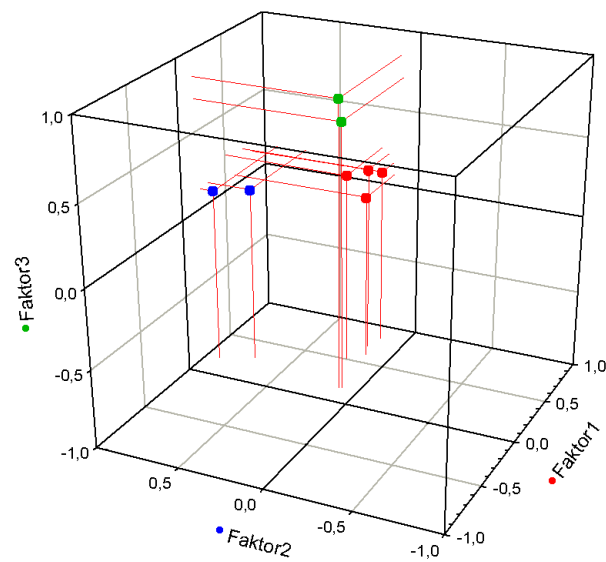


Abbildung 5.14: Dreidimensionaler rotierter Faktorraum (nicht SPSS)

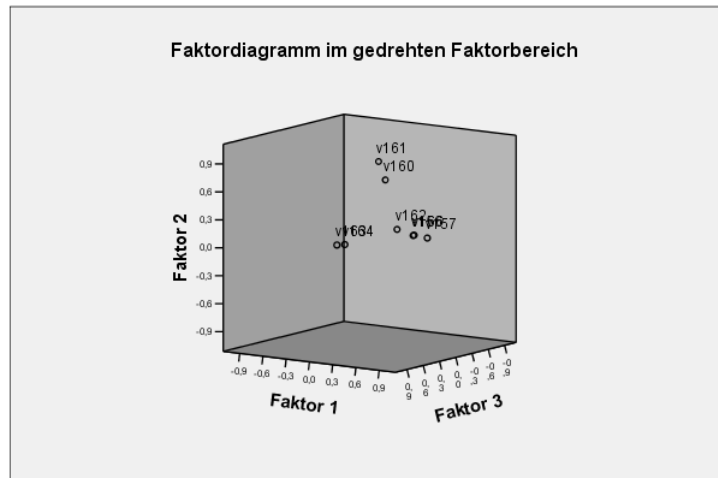


Abbildung 5.15: Dreidimensionaler rotierter Faktorraum

Da alle Variablen positiv auf allen 3 Faktoren laden können wir uns auf den positiven Bereich aller drei Faktoren (Achsen) beschränken, und erhalten ein übersichtlicheres Bild:

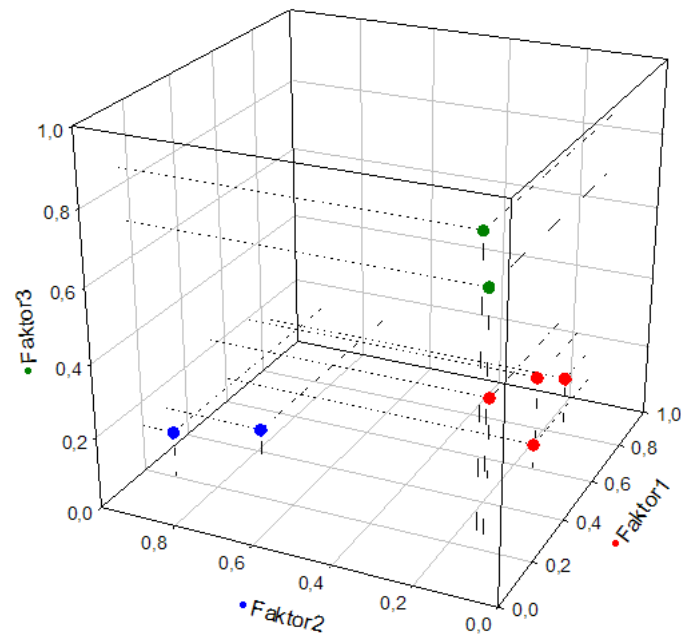


Abbildung 5.16: Dreidimensionaler rotierter Faktorraum, positiver Bereich (nicht SPSS)

Kapitel 6

Appendix

6.1 Appendix A - χ^2

Die Messung der linearen Abhängigkeit zweier Variablen ist möglich über die Betrachtung des n-Dimensionalen Volumens, der durch die Determinante beschrieben wird. Wir kennen einen Koeffizienten zur Berechnung des Zusammenhangs zwischen zwei Variablen, in dem die Determinante explizit berechnet wird. Dabei handelt es sich um den Koeffizienten χ^2 für nominale Variablen. Die Determinante, abgekürzt $\det(\mathbf{A})$, oder $|\mathbf{A}|$ berechnet sich für eine beliebige 2×2 -Tabelle folgendermaßen:

Für die Matrix \mathbf{A}

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

berechnet sich die $\det(\mathbf{A})$ über:

$$\det(\mathbf{A}) = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

Die Formel für χ^2 , die die Kontingenztabelle mit der Indifferenztabelle vergleicht, und für alle $m \times n$ -Tabellen gilt lautet folgendermaßen:

$$\chi^2 = \sum \frac{(f_b - f_e)^2}{f_e}$$

Für uns ist jedoch folgende Formel interessant, die *NUR* für 2×2 -Tabellen gültig ist:

$$\chi^2 = \frac{n \cdot (ad - bc)^2}{(a + c)(a + b)(b + d)(d + c)}$$

Hier erkennen wir, dass die “ χ^2 -Matrix”

$$\mathbf{X} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

den Ausdruck $ad-bc$ als Determinante hat.

In den beiden Ausdrücken von STATA und SPSS sehen wir das χ^2 -Ergebnis für die Kreuztabelle:

$$\begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$$

```
. tab Geschlecht Schwimmen, chi2
```

Geschlecht	Schwimmen		Total
	nein	ja	
weiblich	2	4	6
Männlich	4	8	12
Total	6	12	18

Pearson chi2(1) = 0.0000 Pr = 1.000

Abbildung 6.1: χ^2 in STATA

Hierbei ist die Determinante $2 \cdot 8 - 4 \cdot 4 = 16 - 16 = 0$. Dementsprechend ist auch $\chi^2=0$. Da beide Spalten der Tabelle linear abhängig sind, also die eine Spalte das Vielfache der Anderen ist, sind auch die Proportionen der Schwimmer und Nichtschwimmer bezüglich des Geschlechteranteils identisch (jeweils 1:2).

Geschlecht * Schwimmen Crosstabulation

Count		Schwimmen		Total
		Nichtschwimmer	Schwimmer	
Geschlecht	Weiblich	2	4	6
	Männlich	4	8	12
Total		6	12	18

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,000 ^b	1	1,000		
Continuity Correction ^a	,000	1	1,000		
Likelihood Ratio	,000	1	1,000		
Fisher's Exact Test				1,000	,706
Linear-by-Linear Association	,000	1	1,000		
N of Valid Cases	18				

a. Computed only for a 2x2 table

b. 3 cells (75,0%) have expected count less than 5. The minimum expected count is 2,00.

Abbildung 6.2: χ^2 in SPSS

Hier liegt also kein Zusammenhang zwischen Geschlecht und Schwimmfähigkeit vor. Der Vektor für die Schwimmer

$$\text{Schwimmer} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

ist ein Vielfaches - nämlich das Doppelte - des Vektors der Nichtschwimmer.

$$\text{Nichtschwimmer} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

Die Vektoren sind demnach parallel, sie spannen keinen Raum auf. Die Determinante hat also den Wert 0.

6.2 Appendix B - Matrix-Algebra

Eine Matrix besteht aus m Zeilen und n Spalten. Matrizen werden mit fetten Großbuchstaben bezeichnet. Eine $m \times n$ -Matrix sieht folgendermaßen aus:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}$$

6.2.1 Skalarmultiplikation

Um eine Matrix mit einem Skalar zu multiplizieren muss jedes Element der Matrix mit diesem Skalar multipliziert werden.

$$\mathbf{A} \cdot \varphi = \varphi \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} \varphi a_{11} & \varphi a_{12} & \varphi a_{13} \\ \varphi a_{21} & \varphi a_{22} & \varphi a_{23} \\ \varphi a_{31} & \varphi a_{32} & \varphi a_{33} \end{pmatrix}$$

Rechenregeln für Skalarmultiplikation

$$\begin{aligned} (\alpha + \beta)\mathbf{A} &= \alpha\mathbf{A} + \beta\mathbf{A} \\ \alpha(\mathbf{A} + \mathbf{B}) &= \alpha\mathbf{A} + \alpha\mathbf{B} \end{aligned}$$

$$\begin{aligned} \alpha(\beta\mathbf{A}) &= (\alpha\beta)\mathbf{A} = (\beta\alpha)\mathbf{A} = \beta(\alpha\mathbf{A}) \\ \alpha(\mathbf{A}\mathbf{B}) &= (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B}) = \alpha\mathbf{A}\mathbf{B} \end{aligned}$$

6.2.2 Multiplikation

In der Rechnung mit Matrizen ist Einiges zu beachten, was uns auf den ersten Blick unlogisch oder verwirrend erscheint. So ist jedem von uns aus dem Alltag bekannt, dass 3 mal 5 identisch ist mit 5 mal 3. Dies gilt in der Matrizenrechnung nicht, oder nur für ganz bestimmte Matrizen \mathbf{A} und \mathbf{B} , nämlich für idempotente Matrizen. Im "Normalfall" gilt:

$$\mathbf{A} \cdot \mathbf{B} \neq \mathbf{B} \cdot \mathbf{A}$$

Berechnen wir $\mathbf{A} \cdot \mathbf{B}$ mit

$$\mathbf{A} = \begin{pmatrix} 2 & -3 & 1 \\ -1 & 4 & 0 \end{pmatrix}; \mathbf{B} = \begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -3 \end{pmatrix}$$

erhalten wir folgendes:

$$\begin{pmatrix} 2 & -3 & 1 \\ -1 & 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -3 \end{pmatrix} = \begin{pmatrix} -1 & -7 \\ 13 & 7 \end{pmatrix}$$

Berechnen wir nun $\mathbf{B} \cdot \mathbf{A}$ erhalten wir ein anderes Ergebnis:

$$\begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -3 \end{pmatrix} \cdot \begin{pmatrix} 2 & -3 & 1 \\ -1 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 5 & -5 & 3 \\ 6 & -4 & 4 \\ 13 & -27 & 5 \end{pmatrix}$$

Matrixmultiplikation ist also nicht kommutativ! Wann können wir überhaupt zwei Matrizen miteinander *multiplizieren*? Dies ist nicht immer möglich. Man kann zwei Matrizen nur multiplizieren, wenn die Anzahl der Spalten der ersten Matrix identisch ist mit der Anzahl der Zeilen der zweiten Matrix. Wir können als eine 4×5 -Matrix mit einer 5×6 -Matrix multiplizieren und erhalten so eine 5×5 -Matrix als Ergebnis. Es ist jedoch nicht möglich, eine 5×6 -Matrix mit einer 4×5 -Matrix zu multiplizieren.

$$\begin{array}{ll} \underbrace{3 \times 4}_{\text{grün}} * \underbrace{4 \times 5}_{\text{grün}} = 3 \times 5 & \underbrace{4 \times 3}_{\text{rot}} * \underbrace{5 \times 4}_{\text{rot}} = X \\ \underbrace{5 \times 6}_{\text{grün}} * \underbrace{6 \times 5}_{\text{grün}} = 5 \times 5 & \underbrace{6 \times 5}_{\text{grün}} * \underbrace{5 \times 6}_{\text{grün}} = 6 \times 6 \\ \underbrace{7 \times 5}_{\text{grün}} * \underbrace{5 \times 3}_{\text{grün}} = 7 \times 3 & \underbrace{5 \times 7}_{\text{rot}} * \underbrace{3 \times 5}_{\text{rot}} = X \end{array}$$

Abbildung 6.3: Matrixmultiplikation

Als Beispiel multiplizieren wir nun die schon angesprochenen Matrizen \mathbf{A} und \mathbf{B} , um zu sehen, wie wir an die Werte kommen, die in der resultierenden Matrix \mathbf{C} stehen. Wir berechnen:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C}$$

also

$$\begin{pmatrix} 2 & -3 & 1 \\ -1 & 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -3 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

Welchen Wert hat nun beispielsweise der Eintrag c_{11} ?

$$\begin{pmatrix} 2 & -3 & 1 \\ -1 & 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -3 \end{pmatrix}$$

Der Wert von c_{11} ergibt sich durch

$$c_{11} = 2 \cdot 3 + (-3) \cdot 4 + 1 \cdot 5 = -1$$

Es wird also das erste Element der ersten Zeile der Matrix \mathbf{A} mit dem ersten Element der ersten Spalte der Matrix \mathbf{B} multipliziert. Dann wird das zweite Element der ersten Zeile der Matrix \mathbf{A} mit dem zweiten Element der ersten Spalte der \mathbf{B} multipliziert und dieses Ergebnis zu dem vorherigen Ergebnis addiert. Zu guter Letzt wird das Ergebnis der Multiplikation des dritten Elements der ersten Zeile der Matrix \mathbf{A} mit dem dritten Element der ersten Spalte der Matrix \mathbf{B} zu den beiden vorherigen addiert. Wir erhalten folgende 4 Werte:

$$\begin{aligned} c_{11} &= 2 \cdot 3 + (-3) \cdot 4 + 1 \cdot 5 = -1 \\ c_{12} &= 2 \cdot 1 + (-3) \cdot 2 + 1 \cdot (-3) = -7 \\ c_{21} &= (-1) \cdot 3 + 4 \cdot 4 + 0 \cdot 5 = 13 \\ c_{22} &= (-1) \cdot 1 + 4 \cdot 2 + 0 \cdot (-3) = 7 \end{aligned}$$

Also erhalten wir als Ergebnis der Multiplikation $\mathbf{A} \cdot \mathbf{B}$ die Matrix

$$\mathbf{C} = \begin{pmatrix} -1 & -7 \\ 13 & 7 \end{pmatrix}$$

Rechenregeln zur Matrixmultiplikation

Assoziativgesetz $(AB)C = A(BC)$

$$(AB)C = ABC$$

$$A(BC) = ABC$$

linksseitiges Distributivgesetz $A(B + C) = AB + AC$

rechtsseitiges Distributivgesetz $(A + B)C = AC + BC$

im Allgemeinen $AB \neq BA$

$$AI = IA = A$$

6.2.3 Addition und Subtraktion

Addition und *Subtraktion* von Matizen sind einfacher zu bewerkstelligen als Multiplikation oder Division von Matrizen. Eine Einschränkung ist jedoch, dass nur Matrizen der gleichen Ordnung addiert oder subtrahiert werden können. Schauen wir uns nun die Addition zweier Matrizen an:

$$\begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 2 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 4 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3+5 & 1+4 \\ 5+1 & 2+2 \\ 2+1 & 4+3 \end{pmatrix} = \begin{pmatrix} 8 & 5 \\ 6 & 4 \\ 3 & 7 \end{pmatrix}$$

die Subtraktion zweier Matrizen verläuft analog. Auch hier ein Beispiel:

$$\begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 2 & 4 \end{pmatrix} - \begin{pmatrix} 5 & 4 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3-5 & 1-4 \\ 5-1 & 2-2 \\ 2-1 & 4-3 \end{pmatrix} = \begin{pmatrix} -2 & -3 \\ 4 & 0 \\ 1 & 1 \end{pmatrix}$$

Rechenregeln Addition und Subtraktion

$$(A + B) + C = A + (B + C)$$

$$A + B = B + A$$

$$A + \mathbf{0} = A$$

$$A + (-A) = \mathbf{0}$$

Wobei A , B und $C = n \times m$ -Matrizen und $\mathbf{0}$ die $n \times m$ -Nullmatrix

6.2.4 Transponieren

Als *transponieren* einer Matrix wird der Vorgang bezeichnet, durch den die Zeilen einer Matrix zu Spalten werden und Spalten zu Zeilen. Technischer gesprochen: Aus dem Eintrag a_{ij} wird der a_{ji} . Die transponierte Matrix von \mathbf{A} wird mit \mathbf{A}' oder auch \mathbf{A}^t bezeichnet.

$$\mathbf{A} = \begin{pmatrix} a & b & c \\ x & y & z \end{pmatrix} \xrightarrow{\text{transponieren}} \mathbf{A}' \text{ bzw. } \mathbf{A}^t = \begin{pmatrix} a & x \\ b & y \\ c & z \end{pmatrix}$$

Transponieren und multiplizieren wirken sich kombiniert folgendermaßen aus:

- $\mathbf{A} \times \mathbf{B}$ = Multiplikation Reihe mal Spalte von \mathbf{A} und \mathbf{B}
- $\mathbf{A} \times \mathbf{B}'$ = Multiplikation Reihe mal Reihe von \mathbf{A} und \mathbf{B}
- $\mathbf{A}' \times \mathbf{B}$ = Multiplikation Spalte mal Spalte von \mathbf{A} und \mathbf{B}
- $\mathbf{A}' \times \mathbf{B}'$ = Multiplikation Spalte mal Reihe von \mathbf{A} und \mathbf{B}

Rechenregeln für Transponierte

$$\begin{aligned} (\mathbf{A}')' &= \mathbf{A} \\ (\mathbf{A} + \mathbf{B})' &= \mathbf{A}' + \mathbf{B}' \\ (\alpha \mathbf{A})' &= \alpha \mathbf{A}' \\ (\mathbf{AB})' &= \mathbf{B}' \mathbf{A}' \\ (\mathbf{ABC})' &= \mathbf{C}' \mathbf{B}' \mathbf{A}' \end{aligned}$$

6.2.5 Diagonalmatrizen

Bei einer *symmetrischen* Matrix handelt es sich um einen Sonderfall einer quadratischen Matrix ($n \times n$ -Matrix). Hierbei sind die Matrix und ihre Transponierte identisch, es gilt: $\mathbf{A} = \mathbf{A}'$. Eintrag a_{ij} und a_{ji} sind identisch, also $a_{ij} = a_{ji}$.

Bei einer *Diagonalmatrix* handelt es sich um eine Matrix, in der alle Wert, bis auf die der *Hauptdiagonalen* gleich Null sind. Es gilt also:

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

Bei einer Skalarmatrix handelt es sich um eine Diagonalmatrix, die durch den Skalar gebildet wird. Sie steht für den Skalar und unmegekehrt. Es gilt:

$$\boldsymbol{\omega} = \begin{pmatrix} \omega & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega \end{pmatrix} = \omega$$

$$\boldsymbol{\omega}\mathbf{A} = \mathbf{A}\boldsymbol{\omega} = \omega\mathbf{A} = \mathbf{A}\omega$$

Ein Spezialfall einer Diagonalmatrix ist die sogenannte Einheits- oder Identitätsmatrix, abgekürzt mit \mathbf{E} oder \mathbf{I} . In ihr stehen nur Einsen auf der Hauptdiagonalen.

$$\mathbf{E} \text{ bzw. } \mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

6.2.6 Die Spur einer Matrix

Als Spur einer Matrix $\text{sp}(\mathbf{A})$ oder $\text{tr}(\mathbf{A})$, Abkürzung des englischen “trace” für “Spur”) bezeichnet man die Summe der Elemente der Hauptdiagonalen.

Sie ist definiert als $\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$.

In der Matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

beträgt die Spur $1 + 1 + 1 = 3$

In der Matrix

$$\begin{pmatrix} 5 & -5 & 3 \\ 2 & 3 & 4 \\ 6 & -7 & 4 \end{pmatrix}$$

beträgt die Spur $5 + 3 + 4 = 12$

Rechenregeln für die Spur einer Matrix

$$\begin{array}{l|l}
 \begin{array}{l}
 \operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}) \\
 \operatorname{tr}(\alpha \mathbf{A}) = \alpha \operatorname{tr}(\mathbf{A}) \\
 \operatorname{tr}(\mathbf{BCA}) = \operatorname{tr}(\mathbf{CAB}) = \operatorname{tr}(\mathbf{ABC}) \\
 \text{(i. Allg.) } \operatorname{tr}(\mathbf{AB}) \neq \operatorname{tr}(\mathbf{A})\operatorname{tr}(\mathbf{B})
 \end{array}
 &
 \begin{array}{l}
 \operatorname{tr}(\mathbf{A}') = \operatorname{tr}(\mathbf{A}) \\
 \operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA}) \\
 \operatorname{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \operatorname{tr}(\mathbf{A})
 \end{array}
 \end{array}$$

Für partionierte Matrizen

$$\operatorname{tr} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \operatorname{tr}(\mathbf{A}_{11}) + \operatorname{tr}(\mathbf{A}_{22})$$

6.2.7 Determinante

Die Determinante einer 2×2 -Matrix errechnet sich wie folgt:
für die Matrix \mathbf{A} :

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \rightarrow \det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

Wie aber berechnet man die Determinante von höher dimensionierten Matrizen? Hier als Beispiel eine 3×3 -Matrix:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Es bestehen nun mehrere Möglichkeiten, die Determinante einer 3×3 -Matrix zu berechnen (Laplace'scher Entwicklungssatz). Wir können die Determinante über die gewichtete Summe der Elemente einer Reihe oder Spalte (egal welcher) bestimmen.

- Für jedes Element der gewählten Spalte oder Zeile, hier a_{11} , a_{21} und a_{31} , wird ein Gewicht berechnet.
- Dieses Gewicht ist die Determinante einer 2×2 -Matrix.
- Diese 2×2 -Matrix erhält man, wenn man alle Elemente streicht, die in der gleichen Spalte sowie Zeile stehen, wie das Element, für das man das Gewicht berechnen will.

Also berechnen wir wie folgt für die Matrix \mathbf{A} :

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} = a_{22}a_{33} - a_{23}a_{32} = \alpha$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} = a_{12}a_{33} - a_{13}a_{32} = \beta$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix} = a_{12}a_{23} - a_{13}a_{22} = \gamma$$

Die Determinanten der Restmatrizen werden Kofaktoren der Einzelemente genannt. Das Vorzeichen des Kofaktors erhält man, indem man den Spalten/Zeilen-Index des Einzelements addiert. Bei einer geraden Summe des Indexes ergibt sich ein positives Vorzeichen, bei einer ungeraden Summe ein negatives Vorzeichen. Also:

$$\begin{aligned} \alpha &\rightarrow a_{11} = a_{\text{gerade}} && \text{da } 1 + 1 = 2, \text{ also: } + \\ \beta &\rightarrow a_{12} = a_{\text{ungerade}} && \text{da } 1 + 2 = 3, \text{ also: } - \\ \gamma &\rightarrow a_{13} = a_{\text{gerade}} && \text{da } 1 + 3 = 4, \text{ also: } + \end{aligned}$$

Zusammengefasst ergibt sich die Determinante der 3×3 -Matrix aus:

$$\text{Det}(\mathbf{A}) = a_{11} \cdot \alpha - a_{12} \cdot \beta + a_{13} \cdot \gamma$$

Ebenso lässt sich die Determinante einer 4×4 -Matrix berechnen, allerdings ist dies verschachtelter, und somit aufwendiger. Hier ist es erforderlich, aus der 4×4 -Matrix auf analoge Weise erst vier 3×3 -Matrizen zu extrahieren, um danach mit der eben dargestellten Methode aus diesen 3×3 -Matrizen deren Determinanten zu errechnen. Also wird dieses Verfahren sehr schnell sehr aufwendig.

Eine alternative Berechnung der Determinante einer 3×3 -Matrix funktioniert folgendermaßen (Regel von Sarrus): Die Spalten der Matrix

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

stellt man wie folgt angeordnet dar:

$$\begin{array}{ccccc} a_1 & b_1 & c_1 & a_1 & b_1 \\ a_2 & b_2 & c_2 & a_2 & b_2 \\ a_3 & b_3 & c_3 & a_3 & b_3 \end{array}$$

Nun werden Elemente nach einem bestimmten Muster multipliziert und addiert bzw. subtrahiert.

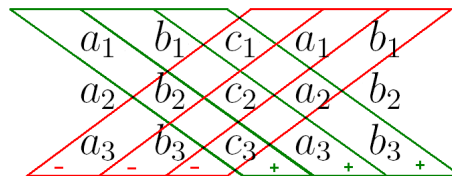


Abbildung 6.4: Rechenschema

Wir rechnen:

$$\det(\mathbf{A}) = a_1 b_2 c_3 + a_3 b_1 c_2 + a_2 b_3 c_1 - a_3 b_2 c_1 - a_1 b_3 c_2 - a_2 b_1 c_3$$

Eine weitere Möglichkeit besteht darin, die Matrix in Stufenform zu bringen. Hierbei ist das Vertauschen von zwei Zeilen oder das Multiplizieren einer Zeile mit einer Zahl (z.B. mit (-1)) nun aber nicht erlaubt bzw. verändert den Wert der Determinante.

Beispiel:

Die Matrix \mathbf{X}

$$\mathbf{X} = \begin{pmatrix} 4 & 7 & 6 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 & 4 & 5 \\ 4 & 7 & 7 & 5 & 6 & 6 \\ 0 & 2 & 1 & 8 & 6 & 6 \\ 0 & 4 & 2 & 16 & 16 & 14 \\ 8 & 14 & 13 & 7 & 9 & 12 \end{pmatrix}$$

wird Schrittweise in Stufenform gebracht. Diese Matrix hat eine Determinante mit dem Wert 192. Auf die ausführliche Darstellung der Berechnung der Determinante sowie der Herstellung der triagonalisierten Matrix (Stufenform) wird an dieser Stelle verzichtet. Wir erhalten folgende Matrix als Ergebnis:

$$\begin{pmatrix} 4 & 7 & 6 & 2 & 3 & 4 \\ 0 & 2 & 1 & 5 & 4 & 5 \\ 0 & 0 & 1 & 3 & 3 & 2 \\ 0 & 0 & 0 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Das Produkt der Hauptdiagonalen $4 \times 2 \times 1 \times 3 \times 4 \times 2 = 192$ ergibt wiederum $\det(\mathbf{A}) = 192$.

Regeln für Determinanten

Für eine $n \times n$ -Matrix \mathbf{A} gilt:

1. Wenn alle Elemente in einer Zeile oder Spalte von \mathbf{A} gleich 0, dann $\det(\mathbf{A}) = 0$
2. $\det(\mathbf{A}) = \det(\mathbf{A}')$
3. Wenn alle Elemente in einer Zeile oder Spalte von \mathbf{A} mit φ multipliziert werden gilt $\varphi \det(\mathbf{A}) = 0$
4. Wenn zwei Spalten oder Zeilen von \mathbf{A} vertauscht werden wechselt $\det(\mathbf{A})$ das Vorzeichen, der Absolutwert bleibt identisch.
5. Wenn zwei Zeilen oder Spalten von \mathbf{A} proportional sind, dann $\det(\mathbf{A}) = 0$
6. $\det(\mathbf{A})$ bleibt unverändert, wenn das Vielfache einer Zeile oder Spalte zu einer anderen Zeile oder Spalte von \mathbf{A} addiert wird.
7. Wenn \mathbf{B} ebenfalls eine $n \times n$ -Matrix ist, dann:
 $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$
8. Für $\varphi \in \mathbb{R}$ gilt $\det(\varphi \mathbf{A}) = \varphi^n \det(\mathbf{A})$

6.2.8 Adjunkte

Um die Adjunkte (abgekürzt mit: $\text{adj}(\mathbf{A})$) einer Matrix zu bestimmen müssen wir folgendes berechnen:

- Für jedes Matrixelement wird der Kofaktor bestimmt

- Jedes mit Element der Matrix wird durch seinen Kofaktor ersetzt
- Die Kofaktoren werden mit (+1) multipliziert, wenn die Indexsumme gerade ist, mit (-1), wenn die Indexsumme negativ ist.
- Danach wird die Matrix der Kofaktoren transponiert.

Beispiel:

$$\begin{pmatrix} 2 & 1 & 2 \\ 2 & 0 & 0 \\ 4 & 2 & 2 \end{pmatrix}$$

Wir ersetzen die ursprünglichen Elemente der Matrix durch die zugehörigen Kofaktoren. Dies geschieht durch die oben dargestellte Methode. Die Matrix der Kofaktoren sieht so aus:

$$\begin{pmatrix} 0 & 4 & 4 \\ -2 & -4 & 0 \\ 0 & -4 & -2 \end{pmatrix}$$

Nun wird für eine gerade Indexsumme mit (+1) multipliziert, mit (-1) bei ungerader Indexsumme. Also ergibt sich folgendes Schema:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightarrow \begin{pmatrix} \text{gerade} & \text{ungerade} & \text{gerade} \\ \text{ungerade} & \text{gerade} & \text{ungerade} \\ \text{gerade} & \text{ungerade} & \text{gerade} \end{pmatrix} \rightarrow \begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Also ergibt sich folgendes:

$$\begin{pmatrix} +(0) & -(4) & +(4) \\ -(-2) & +(-4) & -(0) \\ +(0) & -(-4) & +(-2) \end{pmatrix} \rightarrow \begin{pmatrix} 0 & -4 & 4 \\ 2 & -4 & 0 \\ 0 & 4 & -2 \end{pmatrix}$$

Schlussendlich transponieren wir diese Matrix und erhalten so die Adjunkte.

$$\begin{pmatrix} 0 & -4 & 4 \\ 2 & -4 & 0 \\ 0 & 4 & -2 \end{pmatrix} \xrightarrow{\text{transponieren}} \begin{pmatrix} 0 & 2 & 0 \\ -4 & -4 & 4 \\ 4 & 0 & -2 \end{pmatrix}$$

6.2.9 Inverse

Bleibt noch die *Inverse* einer Matrix, auch *Reziprokmatrix* genannt. Sie ist nur für quadratische Matrizen definiert. Die Inverse von \mathbf{A} wird mit \mathbf{A}^{-1}

abgekürzt. Es gilt: $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$. Wenn man sich die Zahlen, mit denen wir tagtäglich rechnen als eindimensionale, also 1×1 -Matrizen vorstellt, dann ist die Inverse zu $7 \frac{1}{7}$, oder anders geschrieben 7^{-1} , da $7 \cdot \frac{1}{7} = 1$ ist. \mathbf{I} ist die Identitätsmatrix im eindimensionalen Raum. Sie besteht nur aus einem Eintrag, nämlich 1, da dieser einzige Eintrag gleichzeitig die gesamte Hauptdiagonale ist. Für höherdimensionale Räume wird die Berechnung der Inversen aufwendiger, sofern sie überhaupt existiert.

- Die Inverse einer Matrix existiert nur für *quadratische* Matrizen, da nur quadratische Matrizen eine Determinante haben, die zur Berechnung der Inversen notwendig ist. Vorsicht: nicht jede quadratische Matrix besitzt eine Inverse!
- Die Inverse existiert nur, wenn die Determinante der Matrix von Null verschieden ist. Solche Matrizen heißen regulär oder nicht-singulär. Quadratische Matrizen besitzen also nicht immer eine Inverse sondern *können* Inversen besitzen. Müssen sie aber nicht.
- Eine Matrix mit $\det(\mathbf{A}) = 0$ heißt singulär.
- Eine $\det(\mathbf{A}) = 0$ resultiert dann, wenn man eine Zeile oder Spalte als Linearkombination einer oder mehrerer Spalten darstellen kann.

Berechnung der Inversen:

$$\mathbf{A}^{-1} = \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})}$$

Hier mag man sich noch einmal vor Augen führen, dass die Adjunkte einer Matrix wieder eine Matrix ist, die Determinante einer Matrix jedoch keine Matrix sondern ein Skalar. Ein Beispiel:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 2 \\ 2 & 0 & 0 \\ 4 & 2 & 2 \end{pmatrix}; \text{adj}(\mathbf{A}) = \begin{pmatrix} 0 & 2 & 0 \\ -4 & -4 & 4 \\ 4 & 0 & -2 \end{pmatrix}; \det(\mathbf{A}) = 4$$

Wir setzen diese Werte ein:

$$\mathbf{A}^{-1} = \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})} = \text{adj}(\mathbf{A}) \cdot \det(\mathbf{A})^{-1}$$

$$\mathbf{A}^{-1} = \begin{pmatrix} 0 & 2 & 0 \\ -4 & -4 & 4 \\ 4 & 0 & -2 \end{pmatrix} \cdot 4^{-1} = \begin{pmatrix} 0 & 0.5 & 0 \\ -1 & -1 & 1 \\ 1 & 0 & -0.5 \end{pmatrix}$$

wir erinnern uns, dass nun gilt $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$

$$\begin{pmatrix} 2 & 1 & 2 \\ 2 & 0 & 0 \\ 4 & 2 & 2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0.5 & 0 \\ -1 & -1 & 1 \\ 1 & 0 & -0.5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Wir sehen hier, warum nur Matrizen mit einer $\det(\mathbf{A}) \neq 0$ eine Inverse besitzen: ganz einfach deshalb, weil die Berechnung *möglich* ist. Matrizen mit einer $\det(\mathbf{A})=0$ stoßen bei der Berechnung der Inversen auf das altbekannte Problem einer Division durch Null, dem einen “großen Verbot” aus Schultagen neben dem Wurzelziehen aus negativen Zahlen. Es liegt also schlichtweg daran, das wir auf dem Rechenweg in einer Sackgasse enden.

Eine alternative Methode die Inverse einer Matrix auszurechnen funktioniert folgendermaßen: Wir schreiben links die Matrix, z.B. \mathbf{A} , die wir invertieren wollen und rechts die gleichdimensionierte Einheitsmatrix \mathbf{I}

$$(\mathbf{A}|\mathbf{I})$$

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 4 & 2 & 2 & 0 & 0 & 1 \end{array} \right) \begin{array}{l} \text{Zeile I} \\ \text{Zeile II} \\ \text{Zeile III} \end{array}$$

Nun formen wir die linke Seite schrittweise so um, dass sie zur Einheitsmatrix wird. Dadurch verändert sich die Einheitsmatrix auf der rechten Seite so, dass sie zur Inversen wird. Resultiert auf der linken Seite eine komplette Nullzeile oder Nullspalte, so hat die Matrix \mathbf{A} nicht vollen Rang, und sie ist nicht invertierbar.

Als 1. Schritt subtrahieren wir Zeile I von der Zeile II

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 4 & 2 & 2 & 0 & 0 & 1 \end{array} \right) -I$$

Als 2. Schritt subtrahieren wir 2·I von III

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 1 & 0 \\ 4 & 2 & 2 & 0 & 0 & 1 \end{array} \right) -2 \cdot I$$

Als 3. Schritt multiplizieren wir III mit (-1)

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 1 & 0 \\ 0 & 0 & -2 & -2 & 0 & 1 \end{array} \right) \quad (-1)$$

4. subtrahieren wir III von I

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 & 0 & -1 \end{array} \right) \quad -III$$

5. addieren wir III zu II

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & -2 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 & 0 & -1 \end{array} \right) \quad +III$$

6. addieren wir II zu I

$$\left(\begin{array}{ccc|ccc} 2 & 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 2 & 2 & 0 & -1 \end{array} \right) \quad +II$$

7. dividieren wir I und III durch 2 und multiplizieren II mit (-1)

$$\left(\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 2 & 2 & 0 & -1 \end{array} \right) \quad \begin{array}{l} \div 2 \\ (-1) \\ \div 2 \end{array}$$

Das Ergebnis sieht wie folgt aus:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & 1 & 0 & -0.5 \end{array} \right)$$

$(\mathbf{I}|\mathbf{A}^{-1})$

Wir erreichen hier wiederum, wie in der vorherigen Rechnung:

$$\text{adj}(\mathbf{A}) = \begin{pmatrix} 0 & 0.5 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & -0.5 \end{pmatrix}$$

Die *Division* zweier Matrizen kann als Multiplikation zweier Matrizen aufgefasst werden. Dazu benötigt man die Inverse.

$$\frac{\mathbf{A}}{\mathbf{B}} = \mathbf{A} \cdot \mathbf{B}^{-1}, \text{ deshalb auch } \frac{\mathbf{A}}{\mathbf{A}} = \mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$$

Rechenregeln für Inverse

Sofern die Matrizen \mathbf{A} , \mathbf{B} und \mathbf{C} invertierbar sind:

$$\begin{aligned} (\mathbf{A}^{-1})^{-1} &= \mathbf{A} \\ \mathbf{A}\mathbf{V} &= \mathbf{I}, \text{ wenn } \mathbf{V} = \mathbf{A}^{-1} \\ (\mathbf{A}\mathbf{B})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\ (\mathbf{A}\mathbf{B}\mathbf{C})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\ (\mathbf{A}')^{-1} &= (\mathbf{A}^{-1})' \\ (\varphi\mathbf{A})^{-1} &= \varphi^{-1}\mathbf{A}^{-1} \end{aligned}$$

Lösen von Gleichungen mittels Inversen:

$$\mathbf{A}\mathbf{X} = \mathbf{B} \Leftrightarrow \mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

$$\mathbf{X}\mathbf{A} = \mathbf{B} \Leftrightarrow \mathbf{X} = \mathbf{B}\mathbf{A}^{-1}$$

6.2.10 Der Rang einer Matrix

Der *Rang* ist innerhalb der Mathematik ein Begriff aus der linearen Algebra. Man ordnet ihn einer linearen Abbildung oder einer Matrix zu. Übliche Abkürzungen sind $\text{rang}(\mathbf{A})$ oder $\text{rg}(\mathbf{A})$. Bei einer linearen Abbildung ist der Rang als Dimension des Bildes dieser Abbildung definiert. Zu einer Matrix existiert ein *Zeilenrang* und ein *Spaltenrang*. Der Zeilenrang ist die Dimension des von den Zeilenvektoren aufgespannten Vektorraumes und entspricht der Anzahl der unabhängigen Zeilenvektoren. Entsprechendes gilt für den Spaltenrang. Man kann zeigen, dass der Zeilenrang und der Spaltenrang identisch sind. Man spricht deshalb vom Rang einer Matrix. Fasst man eine Matrix als Abbildungsmatrix einer linearen Abbildung auf, so besitzen beide -die Matrix und die lineare Abbildung- den gleichen Rang.

Um den Rang einer Matrix zu bestimmen, formt man sie mittels gauss'schem Eliminationsverfahren in eine äquivalente Matrix in Stufenform um. Die Anzahl der von Null verschiedenen Zeilen ergibt den Rang der Matrix. Eine $n \times n$ -Matrix heißt regulär, wenn sie vollen Rang hat, also wenn $\text{rg}(\mathbf{A}) = n$.

Beispiel :

$$\mathbf{A} = \begin{pmatrix} 3 & 4 & 6 \\ 0 & 3 & 2 \\ 0 & 6 & 5 \end{pmatrix} \sim \begin{pmatrix} 3 & 4 & 6 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \text{rang}(\mathbf{A}) = 3$$

$$\mathbf{B} = \begin{pmatrix} 2 & 1 & 6 \\ 0 & 6 & 4 \\ 0 & 3 & 2 \end{pmatrix} \sim \begin{pmatrix} 2 & 1 & 6 \\ 0 & 6 & 4 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \text{rang}(\mathbf{B}) = 2$$

Die einzige Matrix mit dem Rang 0 ist die Nullmatrix $\mathbf{0}$. Für eine $m \times n$ Matrix gilt: $\text{rang}(\mathbf{A}) \leq \min(m, n)$.

Alle reduzierten Korrelationsmatrizen, die von einem gemeinsamen Faktor gebildet werden haben eine Gemeinsamkeit: ihr Rang ist 1. Wenn die Matrix von zwei gemeinsamen Faktoren gebildet wird ist ihr Rang 2. Ein $\text{rg}(\mathbf{A})=1$ bedeutet, dass alle Spalten durch eine andere Spalte fehlerfrei reproduziert werden können. ein $\text{rg}(\mathbf{A})=2$ bedeutet, dass alle Spalten durch eine linearkombination von zwei anderen Spalten "vorhergesagt" werden können. Wenn wir wissen, dass k gemeinsame Faktoren gegeben sind, können wir daraus schleissen, dass der Rang der reduzierten korrelationsmatrix ebenfls k ist. Bei zwei oder mehr Faktoren sind jedoch zusätzliche Annahmen nötig, um das Modell zu formulieren. Sind die Faktoren korreliert, und mit welchen Variablen stehen die Faktoren in Beziehung? Verschmutzung der Daten durch Sampling- oder Messfehler können ebenfalls problematisch sein.

6.2.11 Idempotente Matrix

Eine quadratische Matrix heisst idempotent, wenn gilt: $\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}$.

Für idempotente Matrizen \mathbf{X} und \mathbf{Y} gilt:

$$\mathbf{X}\mathbf{Y} = \mathbf{A}\mathbf{X} \rightarrow \mathbf{X}\mathbf{Y} \text{ idempotent}$$

$$\mathbf{I} - \mathbf{X} \rightarrow \text{idempotent}$$

$$\mathbf{X}(\mathbf{I} - \mathbf{X}) = (\mathbf{I} - \mathbf{X})\mathbf{X} = \mathbf{0}$$

6.2.12 Diverses

Gramian Matrix

Als *Gramian Matrix* bezeichnet man eine quadratische Matrix, wenn sie symmetrisch ist, und alle Eigenwerte ≥ 0 sind. Korrelations- und Kovarianzmatrizen sind immer gramian.

Spektralzerlegung einer Matrix

Die Matrix \mathbf{R} sei Reell und Symmetrisch. Dann kann sie in die drei Matrizen \mathbf{A} , $\mathbf{\Lambda}$ und \mathbf{A}' zerlegt werden. Dabei handelt es sich bei \mathbf{A} um eine Matrix, deren Spalten aus den Eigenvektoren von \mathbf{R} bestehen. Die Matrix $\mathbf{\Lambda}$ ist eine Diagonalmatrix, deren Hauptdiagonale die Eigenwerte von \mathbf{R} enthält. Es gilt:

$$\mathbf{R} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}'$$

Die Matrix

$$\mathbf{R} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}$$

lässt sich zerlegen in $\mathbf{A}\mathbf{\Lambda}\mathbf{A}'$, also:

$$\mathbf{A}\mathbf{\Lambda}\mathbf{A}' = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

mit den entsprechenden Werten:

$$\mathbf{A}\mathbf{\Lambda}\mathbf{A}' = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}$$

6.3 Appendix C - Determinanten

Wieso können wir uns die Determinante als Flächeninhalt, bzw. Volumen vorstellen? Im \mathbb{R}^2 spannen zwei Vektoren \mathbf{a} und \mathbf{b} einen Raum auf.

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}; \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

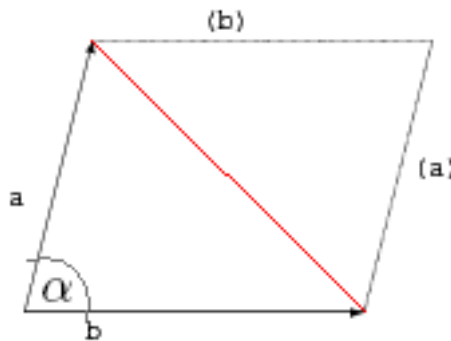


Abbildung 6.5: Flächeninhalt im \mathbb{R}^2

Für den Flächeninhalt A des von den Vektoren \mathbf{a} und \mathbf{b} aufgespannten Parallelogramms gilt $|\mathbf{a}| \cdot |\mathbf{b}| \cdot |\sin \alpha|$. Es gilt:

$$A = |\mathbf{a}| |\mathbf{b}| \sin \alpha$$

also

$$A^2 = \mathbf{a}^2 \mathbf{b}^2 \sin^2 \alpha$$

da $\sin^2 \alpha = 1 - \cos^2 \alpha$ folgt nach Ersetzen

$$A^2 = \mathbf{a}^2 \mathbf{b}^2 (1 - \cos^2 \alpha)$$

durch Ausmultiplizieren

$$A^2 = \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a}^2 \mathbf{b}^2 \cos^2 \alpha)$$

Wenn die Vektoren \mathbf{a} und \mathbf{b} den Winkel α ($0^\circ \leq \alpha \leq 180^\circ$) bilden, dann ist: $\mathbf{a} \bullet \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \alpha$, also folgt:

$$A^2 = \mathbf{a}^2 \mathbf{b}^2 - (\mathbf{a} \bullet \mathbf{b})^2$$

Ebenso gilt im \mathbb{R}^2 : $\mathbf{a} \bullet \mathbf{b} = a_1 b_1 + a_2 b_2$, also:

$$A^2 = (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2$$

nach Auflösen der binomischen Formel und Subtraktion folgt:

$$A^2 = a_1^2 b_2^2 + a_2^2 b_1^2 - 2a_1 b_2 a_2 b_1$$

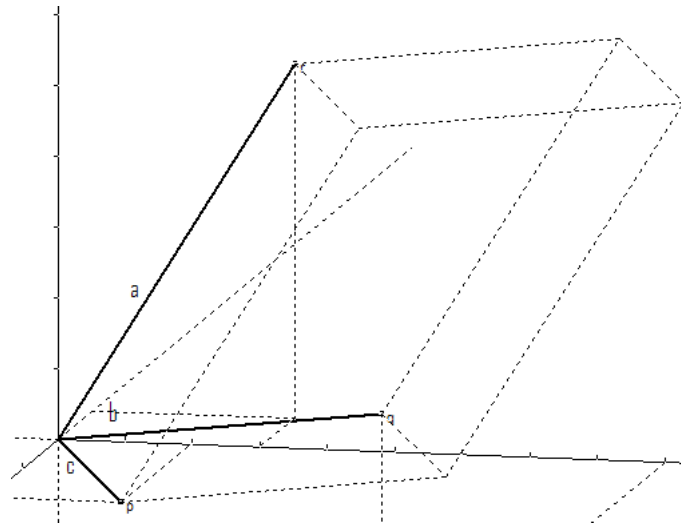
erneutes Zusammenfassen zu einer binomischen Formel:

$$A^2 = (a_1 b_2 - a_2 b_1)^2$$

führt zu:

$$A = |a_1 b_2 - a_2 b_1|$$

Diesen Ausdruck $a_1 b_2 - a_2 b_1$ nennt man Determinante einer 2×2 -Matrix. Im \mathbb{R}^3 errechnet sich das Volumen eines von drei Vektoren \mathbf{a} , \mathbf{b} und \mathbf{c} aufgespannten Spats, oder Raumes über das sogenannte "Spatprodukt".

Abbildung 6.6: Volumen im \mathbb{R}^3

Man kann zeigen, dass das Spatprodukt $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ identisch mit der Determinante einer 3×3 -Matrix ist, nämlich:

$$a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1 - a_1 b_3 c_2 - a_2 b_1 c_3$$

worauf an dieser Stelle jedoch verzichtet wird.

6.4 Appendix D - Eigenwerte

- Wir stellen uns eine 3x3-Matrix vor, benannt mit \mathbf{A} .

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 2 \\ 2 & 3 & 0 \\ 1 & 2 & 2 \end{pmatrix} \quad (6.1)$$

- Die Determinante $\det(\mathbf{A})$ dieser Matrix lässt sich geometrisch als Volumen interpretieren. Das Volumen ist maximal für unkorrelierte Variablen und nimmt bei steigender Korrelation ab bis zum Wert 0. Die Determinante beträgt in unserem Beispiel $\det(\mathbf{A}) = 2$
- Eine symmetrische Matrix hat so viele Eigenwerte λ_j , wie sie Variablen hat. Also hier drei: $\lambda_1, \lambda_2, \lambda_3$. Die Eigenwerte λ_j haben folgende Werte: +5,64575; +0,354249 und +1
- Wir wissen ebenfalls, dass die $\det(\mathbf{A})$ dem Produkt der Eigenwerte λ_j entspricht, also

$$\prod_{i=1}^n \lambda_j = \det(\mathbf{A})$$

In unserem Beispiel also : $5,64575 \cdot 0,354249 \cdot 1 = 2$

- Die Summe der λ_j ist gleich der Summe der Elemente der Hauptdiagonalen. Sie ist ebenfalls gleich der Spur der Matrix, also:

$$\sum_{i=1}^n \lambda_j = \sum_{i=1}^n a_{ii} = Sp(\mathbf{A})$$

in unserem Beispiel: $5,65 + 0,35 + 1 = 2 + 3 + 2 = 7$

Eine Überlegung ist folgende: wenn sich $\det(\mathbf{A})$ als Volumen interpretieren lässt, und $\det(\mathbf{A})$ mit dem Produkt der Eigenwerte zusammenhängt, nämlich folgendermaßen: $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_j$, dann liegt der Gedanke nahe, dass es sich bei den Eigenwerten um eine Art "Achsen" handelt, die das Volumen des Raumes, den die Determinante einnimmt, bestimmen. So wie ein Quader durch die Kanten a,b und c definiert ist, und sein Volumen dem Produkt Länge mal Höhe mal Breite, also $a \cdot b \cdot c$ entspricht, so entspricht die $\det(\mathbf{A})$ dem Produkt der Eigenwerte. Wenn wir nun nach dem Kaiserkriterium nur Faktoren mit $\lambda_j > 1$ extrahieren, so sehen wir folgendes: Es existieren so viele λ_j wie Variablen in der Korrelationsmatrix vorhanden sind. Da

$$\sum_{i=1}^n \lambda_j = \sum_{i=1}^n a_{ii}$$

und in einer Korrelationsmatrix alle $a_{ii} = 1$ (Korrelation einer Variablen mit sich selber: $r_{xx} = 1$), also für n Variablen

$$\sum_{i=1}^n a_{ii} = n \cdot 1 = n$$

ergibt sich: In einer 5×5 -Matrix ist die $\sum_{i=1}^n \lambda_j = 5$. Wenn die Variablen wechselseitig unkorreliert sind, also alle $r_{xy} = 0$, dann sind alle $\lambda_j = 1$. Sobald Korrelationen unter den Variablen in der Matrix vorhanden sind verändern sich die λ_j , und zwar der Gestalt, dass die Summe konstant bleibt, aber nicht mehr alle $\lambda_j=1$, sondern manche größer und manche kleinere Werte annehmen. So haben die 5 λ_j in einer unkorrelierten Matrix (Einheitsmatrix I) die Werte:

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1 ; \sum_{j=1}^5 \lambda_j = 5$$

In einer Korrelationsmatrix, die ihren Namen verdient, in der also nicht nur die Elemente der Hauptdiagonalen 1 sind und alle anderen gleich 0 sind, können die λ_j zum Beispiel folgendermaßen aussehen:

$$\lambda_1 = 2,4 \quad \lambda_2 = 1,5 \quad \lambda_3 = 0,6 \quad \lambda_4 = 0,3 \quad \lambda_5 = 0,2 ; \sum_{j=1}^5 \lambda_j = 5$$

Wir würden hier also den 5-dimensionalen Raum auf einen 2-dimensionalen "einschrumpfen", da von den 5 Achsen nur 2 nach unserem Kriterium als bedeutend eingestuft werden.

In Graphik 6.7 können wir uns anschauen, was dies in 3 Dimensionen bedeuten kann: Der Linke Würfel zeigt uns einen Raum, der von unkorrelierten Variablen aufgespannt wird. Der Quader rechts daneben zeigt, wie wir uns den Bedeutungsverlust einer Variablen, gemessen über die λ_j intuitiv vorstellen können. In unserem linken Beispiel können wir keine Achse weglassen, da jede bedeutend ist. Im rechten Beispiel jedoch sehen wir, dass wir die kürzeste Achse "weglassen" können, ohne allzuviel an Information preiszugeben, da die Werte generell in einem so schmalen Raum angeordnet sind, der eher an eine Ebene erinnert, als an einen "echten Raum" mit großer Ausdehnung in

Länge, Breite, und Höhe. Wir können dort also die dritte, “unbedeutenste” Achse eigentlich weglassen.

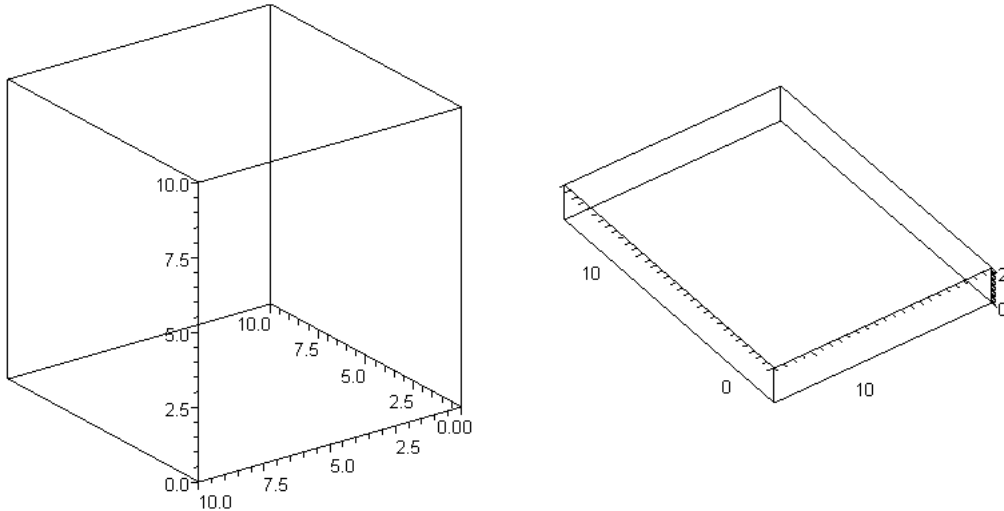


Abbildung 6.7: gleiche $\sum_{i=1}^n \lambda_j$ bei unterschiedlichen $\prod_{i=1}^n \lambda_j = \det(\mathbf{A})$

Unser Beispiel bleibt nur solange intuitiv eingängig, solange keiner unserer λ_j negativ ist. Glücklicherweise sind Kovarianz- und Korrelationsmatrizen (nicht reduzierte Korrelationsmatrizen) gramian, und besitzen wie im vorigen Abschnitt angegeben immer Eigenwerte ≥ 0 . Eine quadratische symmetrische Matrix, in der dies der Fall ist, heisst positiv definit, oder positiv semidefinit. Es gibt folgende Möglichkeiten: Eine quadratische symmetrische Matrix heisst

- positiv definit, falls alle λ_j größer als Null sind;
- positiv semidefinit, falls alle λ_j größer oder gleich Null sind;
- negativ definit, falls alle λ_j kleiner als Null sind;
- negativ semidefinit, falls alle λ_j kleiner oder gleich Null sind
- indefinit, falls positive und negative λ_j existieren.

Literaturverzeichnis

Arminger, Gerhard: Studienskripte zur Soziologie: Statistik für Soziologen 3: Faktorenanalyse. Stuttgart: Teubner, 1979

Basilevsky, Alexander: Statistical Factor Analysis and Related Methods - Theory and Applications, Wiley series in probability and mathematical statistics. New York usw.: John Wiley and Sons, 1994

Backhaus, Klaus / **Erichson**, Bernd / **Plinke**, Wulff / **Weiber**, Rolf: Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. Belin, Heidelberg: Springer Verlag, 2003

Bortz, Jürgen: Statistik - für Human- und Sozialwissenschaftler, 6. vollständig überarbeitete und aktualisierte Auflage. Heidelberg: Springer Medizin Verlag, 2004

Fahrmeir, Ludwig / **Hamerle**, Alfred / **Tutz**, Gerhard (Hrsg.): Multivariate statistische Verfahren, 2. über. Aufl. Berlin, New York: Walter de Gruyter, 1996

Harman, Harry H.: Modern Factor Analysis. Chicago, London: The University of Chicago Press, 1976

Kim, Jae-On / **Mueller**, Charles W.: Introduction to Factor Analysis - What is it and how to do it Sage University Paper series on Quantitative Applications in the Social Sciences, 07-013. Beverly Hills and London: Sage Publications, 1978

Kim, Jae-On / **Mueller**, Charles W.: Factor Analysis - Statistical Methods and Practical Issues Sage University Paper series on Quantitative Applications in the Social Sciences, 07-014. Beverly Hills and London: Sage Publications, 1978

Kohler, Ulrich / **Kreuter**, Frauke: Datenanalyse mit Stata - Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung. München, Wien: R. Oldenbourg Verlag, 2001

Litz, Hans Peter: Multivariate statistische Methoden - und ihre Anwendung in den Wirtschafts- und Sozialwissenschaften. München, Wien: R. Oldenbourg Verlag, 2000

Mulaik, S. A.: The Foundations of Factor Analysis. New York: McGraw-Hill, 1972

Überla, Karl: Faktorenanalyse - Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. Berlin, Heidelberg, New York: Springer Verlag, 1977

Internet

Lang, Stefan:

<http://www.stat.uni-muenchen.de/~lang/skript/matrixalgebra.pdf>

Böker, Fred:

<http://www.statoek.wiso.uni-goettingen.de/veranstaltungen/Multivariate/Daten/Matrix.pdf>

Abbildungsverzeichnis

1.2	Annahmen	5
1.1	2-Faktor 5-Variablen Modell	6
1.3	Ablauf der Faktorenanalyse	8
1.4	Flächeninhalt und lineare Abhängigkeit	9
1.5	$f(x) = \sin x$	10
1.6	Problem der Faktorkorrelation	11
1.7	Problem der Faktorladungen	12
1.8	Problem der Faktorzahl	12
1.9	Kombination der Probleme	13
2.1	Korrelation r_{12}	19
3.1	Regression	23
3.2	Principal Axis	24
3.3	Dreidimensionales kartesisches Koordinatensystem	25
3.4	Dyadisches Produkt: Reproduzierte Korrelationsmatrix	26
3.5	Skalarprodukt: Eigenwert	27
3.6	Eigenwert	28
3.7	Multivariate Normalverteilung	33
3.8	Titanic Baumdiagramm	35
3.9	Maximum Likelihood	37
3.10	Bivariate lineare Regression	38
3.11	Varianzzerlegung für einen Datenpunkt	39
3.12	Varianzzerlegung für alle Datenpunkte	39
3.13	verschiedene Regressionsgeraden	41
3.14	Bestimmung des kleinsten Fehlerquadrates	41
4.1	Screeplot-Beispiel	47
4.2	Geometrische Interpretation von Faktorladungen und Kom- munalitäten	50
4.3	Orthogonale Rotation	51

4.4	Varimax	52
4.5	Quartimax	53
4.6	Oblique Rotation	55
4.7	Beispiel: Strukturgleichungsmodell 1	56
4.8	Beispiel: Strukturgleichungsmodell 2	56
4.9	Idee Referenzachsen	57
5.1	Korrelation der Erziehungsziel-Variablen	65
5.2	Erklärte Gesamtvarianz	66
5.3	Screepplot der Erziehungsvariablen	66
5.4	Faktormatrix Erziehungsvariablen	67
5.5	Rotierte Faktoren in SPSS	68
5.6	Rotierte Lösung (nicht SPSS)	68
5.7	Rotierte Faktormatrix Erziehungsvariablen	69
5.8	Korrelationsmatrix der Vertrauensvariablen	70
5.9	Erklärte Gesamtvarianz der Vertrauensvariablen	70
5.10	Screepplot Vertrauensvariablen	71
5.11	Faktormatrix Vertrauensvariablen	71
5.12	Rotierte Faktormatrix Vertrauensvariablen	72
5.13	Dreidimensionaler unrotierter Faktorraum (nicht SPSS)	73
5.14	Dreidimensionaler rotierter Faktorraum (nicht SPSS)	73
5.15	Dreidimensionaler rotierter Faktorraum	74
5.16	Dreidimensionaler rotierter Faktorraum, positiver Bereich (nicht SPSS)	74
6.1	χ^2 in STATA	76
6.2	χ^2 in SPSS	77
6.3	Matrixmultiplikation	79
6.4	Rechenschema	86
6.5	Flächeninhalt im \mathbb{R}^2	95
6.6	Volumen im \mathbb{R}^3	97
6.7	gleiche $\sum_{i=1}^n \lambda_j$ bei unterschiedlichen $\prod_{i=1}^n \lambda_j = \det(\mathbf{A})$	100