

Multivariate statistische Verfahren für quantitative Variablen

Erwin Grüner*

Fachbereich Psychologie
Philipps-Universität Marburg

Version 1.0
(24.04.2007)

Inhaltsverzeichnis

1	Einführung	5
1.1	Wozu multivariate Verfahren?	5
1.2	Allgemeine Fragen	6
1.3	Übersicht über die multivariaten Verfahren	9
1.4	Mehrdimensionale Zufallsvariablen und deren Verteilung	9
1.5	Beschreibung mehrdimensionaler Datensätze	14
2	Regressionsanalyse	17
2.1	Einfache lineare Regressionsanalyse	17
2.2	Univariate multiple Regressionsanalyse (MR)	24
2.3	Multivariate multiple Regressionsanalyse (MMR)	33
3	Lineare Modelle	36
3.1	Das Allgemeine Lineare Modell	36
3.1.1	Einfache Varianzanalyse	36
3.1.2	Zweifaktorielle Varianzanalyse	40
3.1.3	Modell mit metrischen und Faktorvariablen	41
3.1.4	Kovarianzanalyse	41
3.2	Das Verallgemeinerte Lineare Modell	41
3.2.1	Klassische lineare Modelle	44
3.2.2	Logistische Regression	44
3.2.3	Poisson-Regression	44
3.3	Gemischte lineare Modelle	45
3.4	Generalized Estimation Equations (GEE-Modelle)	46
4	Multivariate Varianzanalyse (MANOVA)	48
4.1	Hotelling's T^2	48
4.2	MANOVA einfaktoriell	50
4.3	MANOVA mehrfaktoriell mit unabhängigen Gruppen	52
4.4	MANOVA mit Messwiederholungen	52
4.5	Überprüfung der Voraussetzungen für die MANOVA	52
4.6	MANOVA als ALM	53
5	Diskriminanzanalyse (DA)	56
6	Hauptkomponentenanalyse (PCA) und Faktorenanalyse (FA)	60
6.1	Hauptkomponentenanalyse	60
6.2	Faktorenanalyse	62
7	Kanonische Korrelationsanalyse (CA)	66
8	Vektor- und Matrizenrechnung	69
8.1	Skalare	69
8.2	Vektoren	69
8.2.1	Spalten- und Zeilenvektoren	69

8.2.2	Rechnen mit Vektoren	70
8.2.3	Vektoren geometrisch interpretiert	72
8.2.4	Länge, Winkel und Abstand von Vektoren	72
8.2.5	Orthogonale Projektion	74
8.2.6	Vektorräume	74
8.2.7	Gleichungen und deren geometrische Interpretation	76
8.3	Matrizen	77
8.3.1	Definitionen	77
8.3.2	Sonderformen von Matrizen	78
8.3.3	Spezielle Matrizen	78
8.3.4	Rechnen mit Matrizen	78
8.3.5	Eigenschaften und Sätze	79
8.3.6	Matrizeninversion	80
8.3.7	Rang einer Matrix	81
8.3.8	Spur und Norm einer Matrix	82
8.3.9	Determinante einer Matrix	82
8.3.10	Lineare Abbildungen	83
8.3.11	Das Eigenwertproblem	85
8.3.12	Quadratische Formen, Definitheit	88
8.3.13	Zerlegung einer symmetrischen Matrix	88
8.3.14	Eigenschaften symmetrischer Matrizen	89

1 Einführung

1.1 Wozu multivariate Verfahren?

- Viele Fragestellungen sind von Natur aus multivariat dadurch, dass mehrere Variablen gleichzeitig betrachtet werden.
- Solche Fragestellungen können univariat oft gar nicht oder nur in unzureichender Weise behandelt werden.
- Und wenn doch univariate Verfahren angewandt werden könnten, bringen multivariate Verfahren gegenüber univariaten Verfahren Vorteile:
 - Vermeidung der Alpha-Inflation
 - Größere Macht
 - Schwächere Voraussetzungen als univariate Verfahren

Vermeidung der Alpha-Inflation

- Problem bei multiplem Testen: Alpha-Inflation, d.h. das Risiko α für den Fehler erster Art wird größer als angenommen.
- Bei den multiplen Mittelwertsvergleichen (Tukey, Scheffé, ...) wird dieses Problem berücksichtigt.
- Darüber hinaus besteht die Möglichkeit, das Alpha zu korrigieren:
 - Bei n unabhängigen Tests: $\alpha' = 1 - (1 - \alpha)^{1/n}$
 - Bonferroni-Korrektur bei n Tests: $\alpha' = \alpha/n$

Größere Macht durch Einsatz multivariater Verfahren

- Multivariate Tests haben eine größere Macht, d.h. ein kleineres Risiko β für den Fehler zweiter Art.
- Beispielsweise 2 Variablen bei 2 Gruppen gemessen (siehe Abbildung 1.1): univariate Verteilungen überlappen sich mehr als die bivariaten; d.h. die univariaten t-Tests sind eventuell nicht signifikant, wohl aber die multivariate Version des t-Tests (Hotellings T^2):

Schwächere Voraussetzungen?

- Die ANOVA mit Messwiederholungen z.B. setzt die Sphärizität ("sphericity", "compound symmetry") voraus.

Bei einem r -stufigen Messwiederholungsfaktor muss für die $r \times r$ -Kovarianzmatrix Σ der einzelnen Messungen X_1, \dots, X_r in der Population folgendes gelten:

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

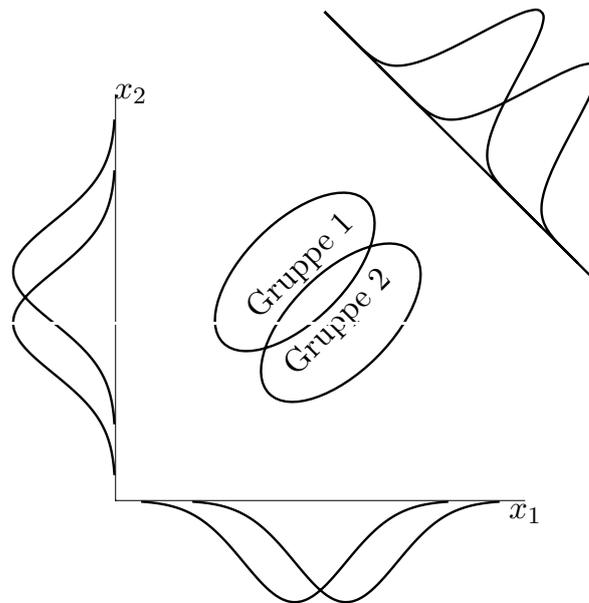


Abbildung 1.1: Macht von multivariaten Tests

- Enthält das Design auch unabhängige Gruppen, so muss dies für jede Gruppe gelten, und die einzelnen Kovarianzmatrizen müssen identisch sein.

Homogenitätsvoraussetzung bei der MANOVA

- Bei einer einfaktoriellen MANOVA mit k unabhängigen Gruppen und p Variablen müssen die Kovarianzmatrizen für die einzelnen Gruppen gleich sein, können aber beliebiges Aussehen haben:

$$\Sigma^{(1)} = \Sigma^{(2)} = \dots = \Sigma^{(k)} = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{bmatrix}$$

1.2 Allgemeine Fragen

- Einsatzzweck des Verfahrens: deskriptiv, exploratorisch oder konfirmatorisch?
- Ist das verwendete Modell kompensatorisch? Ist es linear?
- Enthält das Modell neben stochastischen auch deterministische Komponenten?
- Voraussetzungen bzgl. Skalenniveau, Verteilungsform, Parameter der Verteilung, Metrik?
- Stichprobenabhängigkeit?
- Ausreißer vorhanden?
- Kausalinterpretation von Ergebnissen möglich?

Verwendungszweck

1. Rein deskriptiv, d.h. man möchte nur die Daten mit Hilfe einfacher Statistiken beschreiben oder z.B. eine Kurve (eine Funktion) an die Daten anpassen.
2. Exploratorisch: die Analyse soll der Hypothesengenerierung dienen.

3. Konfirmatorisch, inferenzstatistisch: es soll ein mathematisch-statistisches Modell angepasst und getestet werden.

Linearität

- Viele der Modelle beinhalten eine lineare Funktion (eine Linearkombination) der Variablen.
- Damit wird auch eine lineare Beziehung zwischen Variablen angenommen.
- Ohnehin basieren viele der Verfahren auf Korrelationen.
- Da eine gewichtete Summe gebildet wird, kann die Minderleistung in einer Variable durch andere Variablen kompensiert werden.

Deterministische Komponenten

- Bei den unabhängigen Variablen unterscheidet man generell zwischen solchen mit festen und solchen mit zufälligen Effekten (fixed vs. random effects).
- Bei fixen Effekten interessiert man sich nur für die in den Daten realisierten Ausprägungen und man möchte nicht auf andere Ausprägungen generalisieren.
- Wird eine unabhängige Variable als solche mit zufälligen Effekten behandelt, so wirkt sich dies auf die Analyse aus.

Voraussetzungen

- Die multivariaten Verfahren machen Voraussetzungen hinsichtlich
 - Skalenniveau der Variablen
 - Verteilung (deren Form und Parameter) der Variablen
 - Unabhängigkeit der Beobachtungen
 - u.a.
- Manche statistischen Tests sind robust gegen die Verletzung von Voraussetzungen.
- In manchen Situationen bietet sich eine Transformation der Variablen an:
 - bei Verletzung von Verteilungsvoraussetzungen
 - bei Vorliegen einer nichtlinearen Beziehung

Das angepasste Modell gilt dann für die transformierte Variable.

- Die Ergebnisse der Verfahren sind abhängig von der gewählten Metrik. Es ist nicht nötig, dass die untersuchten Variablen in der Realität genau dieselben und genauso skaliert sind wie jene für die Analyse.

Beispielsweise postuliert Stevens(1950) in seinem Potenzgesetz den folgenden Zusammenhang zwischen der physikalischen Helligkeit X einer Lichtquelle und dem Helligkeitseindruck Y einer Versuchsperson:

$$Y = aX^b + c \quad \text{mit} \quad b = 0.35$$

Y ist also eine lineare Funktion von X^b und nicht von X . Man kann die Variable X durch die Variable $\tilde{X} = X^b$ substituieren und eine lineare Regression von Y auf \tilde{X} rechnen, um die Parameter a und c zu schätzen. Nach Rücktransformation von \tilde{X} nach X erhält man dann die Beziehung zwischen Y und X .

Versuchspersonenstichprobe

- Bei allen Verfahren wird eine sog. Zielfunktion maximiert bzw. minimiert.

1 Einführung

- Dies erfolgt für die aktuelle Stichprobe incl. aller Zufallskomponenten.
- Die gewonnenen Parameterschätzungen sind also maximiert bzw. minimiert lediglich für die aktuelle Stichprobe.
- Gegenmaßnahmen:
 - Biaskorrektur
 - Kreuzvalidierung
 - Verwendung von Moderatorvariablen
- Einfluss der Varianz: bei homogenen Stichproben erhält man niedrige Korrelationen (Stichwort: lokale stochastische Unabhängigkeit).

Variablenstichprobe

- Neben der Versuchspersonenstichprobe spielt auch die aktuelle Variablenauswahl eine Rolle.
- Bei den meisten Verfahren ändern sich die Ergebnisse, wenn Variablen weggelassen oder neue hinzugenommen werden.
- Probleme:
 - Lineare Abhängigkeit von Variablen
 - Multikollinearität

Ausreißer

- In der univariaten Statistik haben Ausreißer einen großen Einfluss auf Mittelwert und Varianz.
- Neben univariaten Ausreißern unterscheidet man auch multivariate Ausreißer; diese sind schwerer zu entdecken.
- In der multivariaten Statistik spielen Ausreißer eine noch größere Rolle, weil Korrelationen dafür noch sensitiver sind als Mittelwerte oder Varianzen.
- Der Einfluss von Ausreißern erkennt man an verschiedenen individuellen Einfluss-Statistiken.
- Es gibt robuste Schätzmethoden, die den Einfluss von Ausreißern ausschalten bzw. reduzieren.

Kausalschlüsse

- Korrelationen erlauben prinzipiell keine Kausalaussagen, und auf Korrelationen beruhen viele der multivariaten Verfahren.
- Auch wenn das verwendete Modell als Diagramm mit gerichteten Pfaden spezifiziert ist, ist eine Kausalinterpretation unzulässig. Dazu müssen externe Argumente herangezogen werden (z.B. die Tatsache, dass die Ursache zeitlich vor der Wirkung kommt).
- Bei Verwendung experimentell manipulierter Variablen sind jedoch Kausalschlüsse möglich.

1.3 Übersicht über die multivariaten Verfahren

Verfahren	Variablenset 1 (q Variablen)	Variablenset 2 (p Variablen)
Multiple Regression	metrisch, $q = 1$	beliebig, $p \geq 1$
Logistische Regression	kategorial, $q = 1$	beliebig, $p \geq 1$
Poisson-Regression	Zählvariable, $q = 1$	beliebig, $p \geq 1$
Multivariate Regression	metrisch, $q > 1$	beliebig, $p \geq 1$
Multivariate Varianzanalyse	kategorial, $q \geq 1$	metrisch, $p \geq 1$
Diskriminanzanalyse	kategorial, $q = 1$	metrisch, $p \geq 1$
Kanonische Korrelationsanalyse	metrisch, $q \geq 1$	metrisch, $p \geq 1$
Hauptkomponentenanalyse	metrisch, $q > 1$	-
Faktorenanalyse	metrisch, $q > 1$	-
Clusteranalyse	metrisch, $q > 1$	-

1.4 Mehrdimensionale Zufallsvariablen und deren Verteilung

Mehrdimensionale Zufallsvariablen

Wir betrachten gleichzeitig mehrere Zufallsvariablen X_1, \dots, X_p .

Der kürzeren Schreibweise wegen werden die Variablen meist zu einem p -dimensionalen Zufallsvektor \mathbf{X} zusammengefasst:

$$\mathbf{X} = (X_1, \dots, X_p).$$

Dabei sind die X_i eindimensionale Zufallsvariablen.

Uns interessiert deren gemeinsame Verteilung $f(X_1, \dots, X_p)$.

Ein mehrdimensionaler Zufallsvektor \mathbf{X} besitzt einen Erwartungswertvektor $\boldsymbol{\mu}$ und eine Kovarianzmatrix $\boldsymbol{\Sigma}$:

$$\begin{aligned} E(\mathbf{X}) &= \boldsymbol{\mu} = (\mu_1, \dots, \mu_p) = (E(X_1), \dots, E(X_p)) \\ \text{Cov}(\mathbf{X}) &= \boldsymbol{\Sigma} = E\{(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))'\} \end{aligned}$$

Neben der gemeinsamen Verteilung kann man jeweils auch Randverteilungen und bedingte Verteilungen betrachten.

Mehrdimensionale Verteilungen

Es gibt eine Reihe mehrdimensionaler Wahrscheinlichkeitsverteilungen, z.B.:

- Multinomialverteilung (multivariate Version der Binomialverteilung)
- Mehrdimensionale Normalverteilung
- Hotelling's T^2 -Verteilung (multivariate Version der t -Verteilung)
- Wishartverteilung (multivariate Version der χ^2 -Verteilung)
- Λ -Verteilung (multivariate Version der F-Verteilung)
- θ -Verteilung (Verteilung des größten Eigenwerts)

Mehrdimensionale Normalverteilung

Die Parameter der mehrdimensionalen Normalverteilung sind $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$. $\boldsymbol{\mu}$ ist der Erwartungswertvektor und $\boldsymbol{\Sigma}$ die Kovarianzmatrix des Zufallsvektors.

Die Dichte der p -dimensionalen Normalverteilung ist gegeben durch:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp \left[-0.5(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

In Verallgemeinerung zum eindimensionalen Fall taucht hier also anstelle der Varianz die Determinante und die Inverse der Kovarianzmatrix auf. D.h. für $p = 1$ ergibt sich:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]$$

Zweidimensionale Normalverteilung

Die zweidimensionale Normalverteilung ist definiert durch 5 Parameter: $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$.

Ihre Dichte ist gegeben durch:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

Die zweidimensionale Standardnormalverteilung hat nur einen Parameter, nämlich ρ , und die Dichte ist:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[x_1^2 - 2\rho x_1 x_2 + x_2^2 \right] \right\}$$

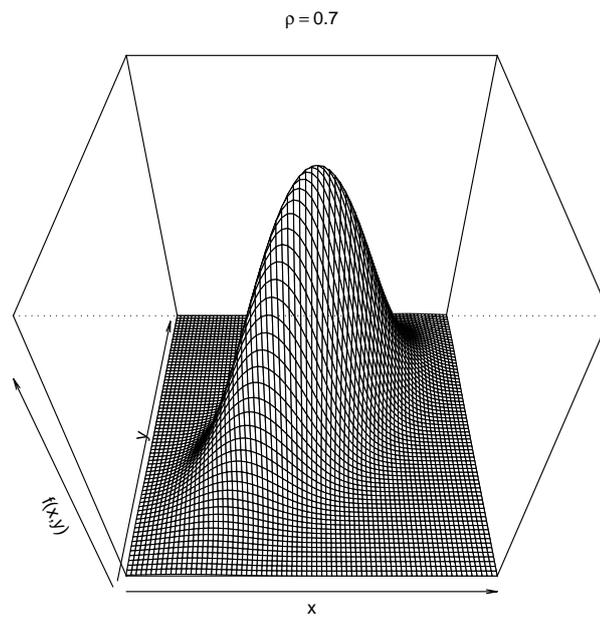


Abbildung 1.2: Zweidimensionale Normalverteilung

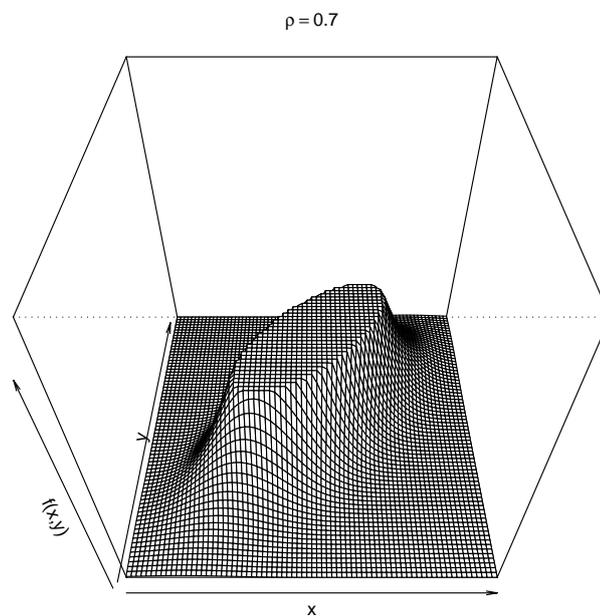


Abbildung 1.3: Zweidimensionale Normalverteilung (abgeschnitten)

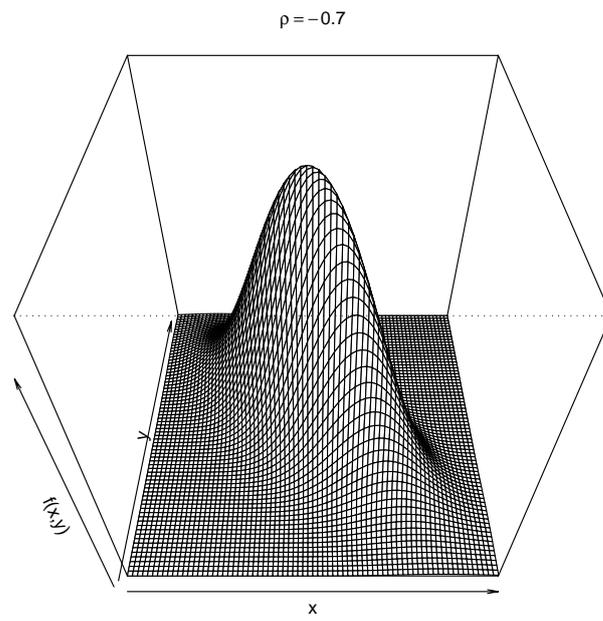


Abbildung 1.4: Zweidimensionale Normalverteilung

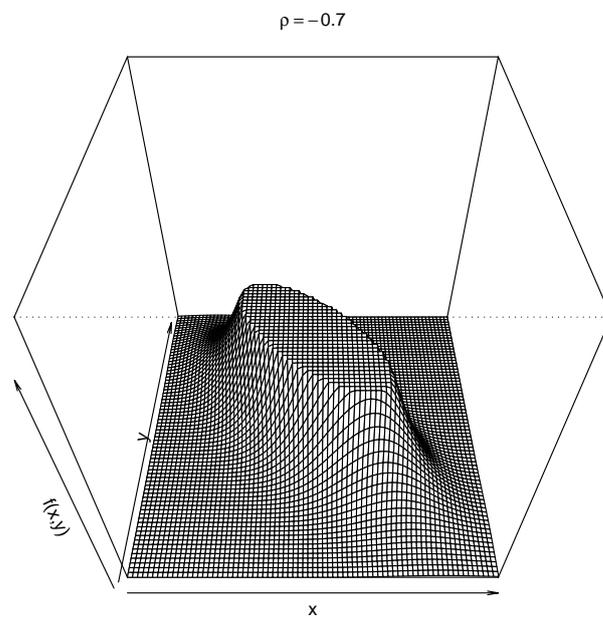


Abbildung 1.5: Zweidimensionale Normalverteilung (abgeschnitten)

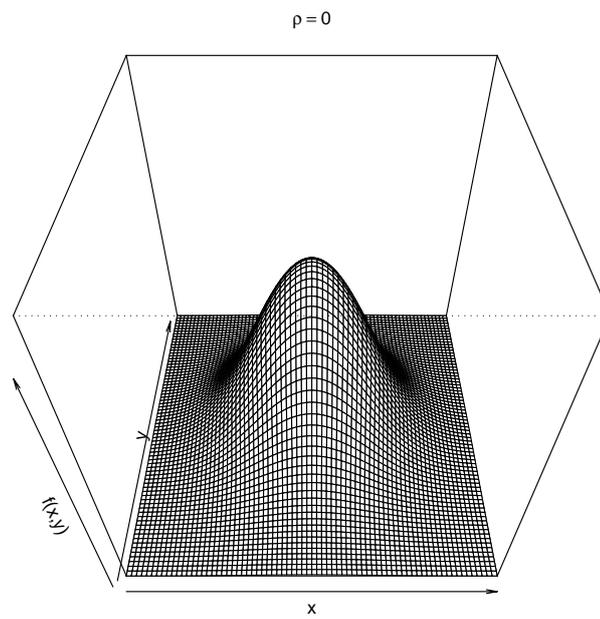


Abbildung 1.6: Zweidimensionale Normalverteilung

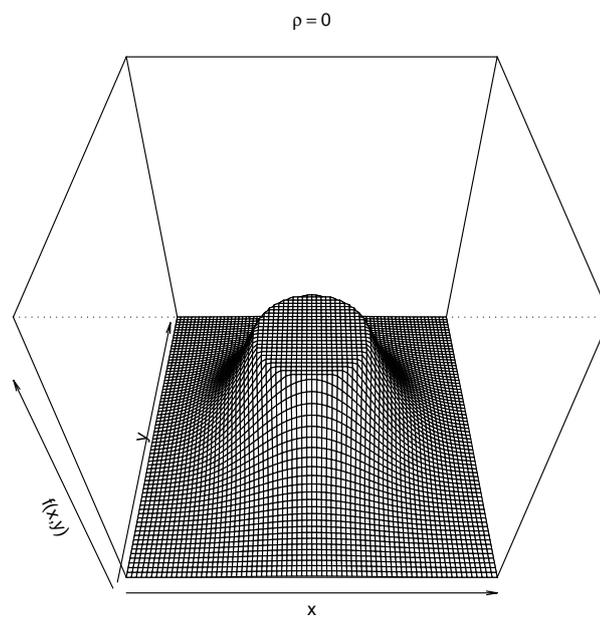


Abbildung 1.7: Zweidimensionale Normalverteilung (abgeschnitten)

1.5 Beschreibung mehrdimensionaler Datensätze

Kennwerte eines mehrdimensionalen Datensatzes

- Daten (n Zeilen/Vpn, p Spalten/Variablen) (x_{ij})
- Mittelwerte ($\bar{x}_{.j} = \frac{1}{n} \sum_{i=1}^n x_{ij}$)
- Abweichungen ($a_{ij} = x_{ij} - \bar{x}_{.j}$)
- Kreuzprodukte
 - der Rohwerte ($\sum_{i=1}^n x_{ij}x_{ik}$)
 - der Abweichungen ($\sum_{i=1}^n (x_{ij} - \bar{x}_{.j})(x_{ik} - \bar{x}_{.k})$)
- Kovarianzen ($s_{jk} = \frac{1}{n-1} \sum_{i=1}^n a_{ij}a_{ik} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})(x_{ik} - \bar{x}_{.k})$)
- Varianzen ($s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2$)
- Korrelationen ($r_{jk} = \frac{s_{jk}}{s_j s_k}$)

Die Kennwerte in Matrixschreibweise

- Datenmatrix: $\mathbf{X}_{n \times p}$
- Mittelwertsvektor: $\bar{\mathbf{x}}_{p \times 1} = \mathbf{X}'_{p \times n} \mathbf{1}_{n \times 1} / n$
- Abweichungsmatrix: $\mathbf{A}_{n \times p} = \mathbf{X}_{n \times p} - \mathbf{1}_{n \times 1} \bar{\mathbf{x}}'_{1 \times p}$
- Kreuzproduktmatrix ("SSCP"-Matrix)
 - der Rohwerte: $\mathbf{X}'_{p \times n} \mathbf{X}_{n \times p}$
 - der Abweichungen: $\mathbf{A}'_{p \times n} \mathbf{A}_{n \times p}$
- Kovarianzmatrix: $\mathbf{S}_{p \times p} = \frac{1}{n-1} \mathbf{A}'_{p \times n} \mathbf{A}_{n \times p}$
- Varianzvektor: $\mathbf{s}^2_{p \times 1} = \text{diag}(\mathbf{S}_{p \times p})$
- Vektor der Standardabweichungen: $\mathbf{s}_{p \times 1} = \sqrt{\text{diag}(\mathbf{S})}$
- Korrelationsmatrix: $\mathbf{R}_{p \times p} = \mathbf{S}_{p \times p} / \mathbf{s}_{p \times 1} \mathbf{s}'_{1 \times p}$

Graphische Darstellung von mehrdimensionalen Datensätzen

- Einfaches Streudiagramm (2 Variablen, 1 Gruppe)
- Überlagertes Streudiagramm (2 Variablen, mehrere Gruppen)
- 3D-Histogramm (2 Variablen)
- 3D-Streudiagramm (3 Variablen): statisch mit Perspektive oder rotierend ("Spinplot")
- Scatterplotmatrix (quadratische Anordnung von einfachen oder überlagerten Streudiagrammen)

Hinweis: In den folgenden Abbildungen wurden die Parameter der Ellipse jeweils aus den Daten geschätzt!

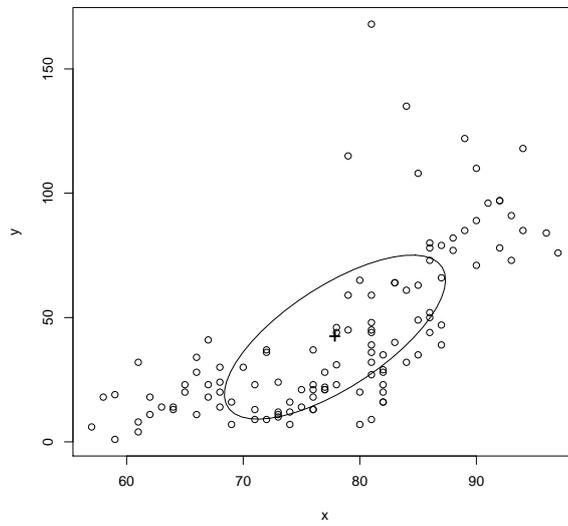


Abbildung 1.8: Einfaches Streudiagramm für Rohdaten mit NV-Ellipse

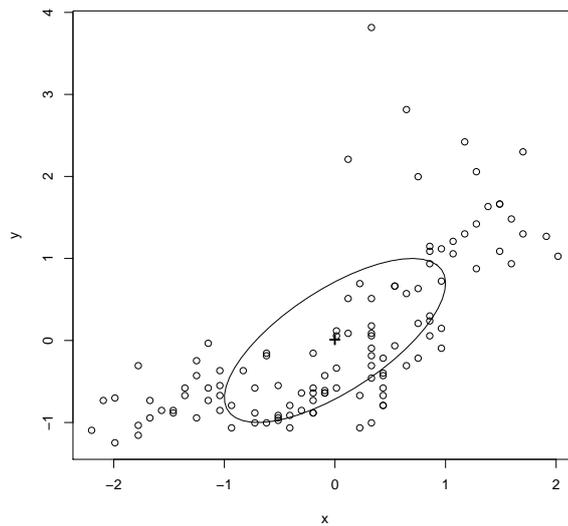


Abbildung 1.9: Einf. Streudiagramm für standardisierte Daten mit NV-Ellipse

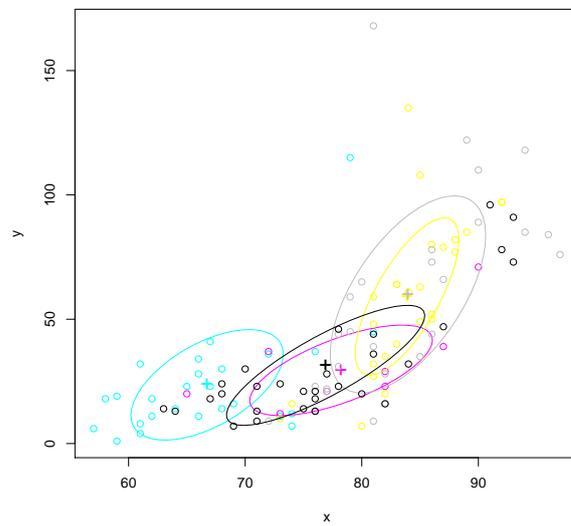


Abbildung 1.10: Überlagertes Streudiagramm für Rohdaten mit NV-Ellipsen

2 Regressionsanalyse

Zielsetzung

Es wird der Zusammenhang betrachtet zwischen

- einer *abhängigen*, zu erklärenden Variable (Kriteriumsvariable, Responsevariable, Zielvariable, Outcomevariable)
- und einer bzw. mehreren *unabhängigen*, erklärenden Variablen (Prädiktorvariablen, Regressoren, Designvariablen, Kovariablen).

Dabei unterscheidet man je nach Zielsetzung zwei verschiedene Ansätze:

- *deskriptiver Ansatz*: Beschreibung des Zusammenhangs bei beobachteten Variablen in einem Datensatz
- *inferenzstatistischer Ansatz*: Schätzen und Testen eines Modells.

2.1 Einfache lineare Regressionsanalyse

Regression als Kurvenanpassungsproblem

Gegeben sind n Datenpunkte (x_i, y_i) .

Ziel: eine Kurve konstruieren, die die Beziehung zwischen den x_i und y_i am besten repräsentiert.

- *Lineare Regression*: es wird ein Polynom ersten Grades, d.h. eine Gerade angepasst:

$$\hat{y}_i = a + bx_i$$

- *Nichtlineare Regression*: es wird irgendeine Funktion f angepasst, z.B. ein Polynom höheren Grades oder eine Exponentialfunktion:

$$\hat{y}_i = f(x_i)$$

Der Funktionstyp wird als bekannt vorausgesetzt und die Parameter werden geschätzt.

Die Anpassungsgüte wird mit Hilfe einer sog. *Zielfunktion* oder *Verlustfunktion* angegeben. Als häufigste Zielfunktion wird das Kriterium der kleinsten Quadrate verwendet: wir definieren den Vektor \mathbf{e} mit den Komponenten $e_i = y_i - \hat{y}_i$ und fordern, dass seine Länge minimal wird, bzw.

$$\mathbf{e}'\mathbf{e} = \sum (y_i - \hat{y}_i)^2 \rightarrow \text{Min}$$

Die Größe $\mathbf{e}'\mathbf{e}$ wird auch Residuenquadratsumme (RSS) genannt.

Als Lösung erhält man:

$$b = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}.$$

Einfache lineare Regression als mathematisches Modell

Zwischen einer beobachtbaren Variable Y und einer beobachtbaren Variable X besteht eine lineare Beziehung der Form

$$Y = \alpha + \beta X + \varepsilon$$

In Matrizenschreibweise:

$$Y = [1 \quad X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon$$

ANNAHMEN

- X ist ein feste, fehlerfrei gemessene Variable.
- α und β sind feste, unbekannte Parameter.
- ε ist die Fehlervariable oder Störgröße, eine nicht beobachtbare Zufallsvariable, für die gilt:

$$\varepsilon \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

d.h.

- ε ist normalverteilt mit Mittelwert 0 und konstanter Varianz σ_ε^2 („Homoskedastizität“) und
- verschiedene Realisierungen von ε sind unkorreliert:

$$\rho(\varepsilon_i, \varepsilon_j) = 0 \quad \text{für } i \neq j$$

LÖSUNG

Für die Modellparameter α und β ergeben sich nach dem OLS Kriterium¹ die folgenden erwartungstreuen Schätzer:

$$\hat{\beta} = b = \frac{s_{xy}}{s_x^2} = r_{xy} \frac{s_y}{s_x}$$

$$\hat{\alpha} = a = \bar{y} - b\bar{x}$$

s_x^2 , s_y^2 , s_{xy} und r_{xy} sind bekanntlich so definiert:

$$s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$s_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

¹OLS = Ordinary Least Squares = Kriterium der kleinsten Quadrate

Einzelne y -Werte können mit folgender Formel geschätzt werden:

$$\hat{Y} = a + bX$$

Man erhält somit als *geschätzte* („gefittete“) Werte:

$$\hat{y}_i = a + bx_i$$

und als *Residuen*:

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$$

DIE BEDEUTUNG VON α UND β

- Der Parameter α gibt den Wert der Responsevariable an, wenn die Prädiktorvariable den Wert 0 hat.
- Der Wert von β gibt an, um wieviel sich der bedingte Erwartungswert der Responsevariable ändert, wenn sich der Wert der Prädiktorvariable um 1 vergrößert.

STANDARDFEHLER DER SCHÄTZUNG

Als globales Abweichungsmaß wird die Residuenquadratsumme (RSS) verwendet:

$$RSS = \sum (y_i - \hat{y}_i)^2$$

Die Anzahl der Freiheitsgrade ist hier gleich $n - 2$ (nämlich Anzahl der Beobachtungen minus Anzahl der geschätzten Parameter). Also ist der mittlere quadratische Fehler (MSE):

$$MSE = s_{y.x}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2}$$

Dieser ist eine erwartungstreue Schätzung der Fehlervarianz σ_ε^2 .

Die Quadratwurzel aus MSE ergibt $s_{y.x}$ (kurz: s), den Standardfehler der Schätzung (Standardschätzfehler):

$$s = \hat{\sigma}_\varepsilon = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$$

Dieser kann auch nach folgender Formel berechnet werden:

$$s = s_y \sqrt{1 - r_{xy}^2}$$

STANDARDFEHLER UND KORRELATION DER SCHÄTZER

Die Standardfehler der geschätzten Parameter sind

$$\hat{SE}(\hat{\alpha}) = s \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \qquad \hat{SE}(\hat{\beta}) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Im allgemeinen sind die Schätzer $\hat{\alpha}$ und $\hat{\beta}$ nicht unabhängig. Ihre Kovarianz beträgt:

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{s^2 \bar{x}}{\sum (x_i - \bar{x})^2}$$

T-TESTS UND KONFIDENZINTERVALLE FÜR DIE PARAMETER

Die Hypothese $H_0 : \beta = \beta_0$ lässt sich mit Hilfe der folgenden t-Statistik testen:

$$t = \frac{\hat{\beta} - \beta_0}{\widehat{SE}(\hat{\beta})}$$

In analoger Weise lautet die t-Statistik für die Hypothese $H_0 : \alpha = \alpha_0$:

$$t = \frac{\hat{\alpha} - \alpha_0}{\widehat{SE}(\hat{\alpha})}$$

Unter H_0 haben beide Teststatistiken eine t-Verteilung mit $n - 2$ Freiheitsgraden.

Mit Hilfe des Standardfehlers lässt sich wie üblich für den Parameter auch jeweils ein Konfidenzintervall für die Parameter angeben.

Test des linearen Regressionsmodells: Quadratsummenzerlegung

Der Modelltest ist im allgemeinen der Test der Hypothese $H_0 : \beta = 0$, was gleichbedeutend ist mit $H_0 : \rho = 0$.

Dazu wird $\sum(y_i - \bar{y})^2$ (Quadratsumme von Y) zerlegt in:

- Quadratsumme der Regression: $\sum(\hat{y}_i - \bar{y})^2$ mit 1 Freiheitsgrad.
- Quadratsumme des Residuums: $\sum(y_i - \hat{y}_i)^2$ mit $n - 2$ Freiheitsgraden.

Die Varianzanalysetabelle sieht wie folgt aus:

Quelle	QS	df	MQ	F
Regression	$\sum(\hat{y}_i - \bar{y})^2$	1
Residuum	$\sum(y_i - \hat{y}_i)^2$	$n - 2$...	
Total	$\sum(y_i - \bar{y})^2$	$n - 1$		

Der Ausdruck $\sum(\hat{y}_i - \bar{y})^2$ ist die Abweichungs-QS der gefitteten Werte und damit gleich $(n - 1)s_{\hat{y}}^2$.

Weiterhin gelten folgende Beziehungen:

$$\sigma_\varepsilon^2 = \sigma_Y^2(1 - \rho^2) \quad \sigma_{\hat{y}}^2 = \sigma_Y^2\rho^2 \quad \rho^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_Y^2} = \frac{\sigma_{\hat{y}}^2}{\sigma_Y^2}$$

$$s^2 = s_Y^2(1 - r^2) \quad s_{\hat{y}}^2 = s_Y^2r^2 \quad r^2 = 1 - \frac{s^2}{s_Y^2} = \frac{s_{\hat{y}}^2}{s_Y^2}$$

In der F-Statistik wird letztlich der Anteil der aufgeklärten Varianz (r_{xy}^2) zum Anteil der Restvarianz ($1 - r_{xy}^2$) in Beziehung gesetzt (noch jeweils dividiert durch die Freiheitsgrade):

$$F = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2 / (n - 2)} = \frac{\frac{r^2}{1}}{\frac{1 - r^2}{n - 2}} = \frac{r^2}{1 - r^2} \cdot (n - 2)$$

Unter der Nullhypothese ist diese Statistik F-verteilt mit $df_1 = 1$ und $df_2 = n - 2$ Freiheitsgraden.

Der F-Wert hängt auch mit dem oben angegebenen t-Wert für die Hypothese $H_0 : \beta = 0$ zusammen:

$$F = \left(\frac{\hat{\beta}}{\widehat{SE}(\hat{\beta})} \right)^2$$

Konfidenz- und Vorhersageintervalle

Ein punktweises (isoliertes) *Konfidenzintervall* für die Regressionsgerade an der Stelle x ist wie folgt zu berechnen:

$$(a + b x) \pm t_{n-2; 1-\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Sollen simultan *mehrere* Konfidenzintervalle berechnet werden, so ist nach Bonferroni das α entsprechend zu korrigieren, d.h. durch die Anzahl der Konfidenzintervalle zu teilen.

Es lässt sich auch ein Konfidenzband für die gesamte Regressionsgerade angeben:

$$f(x) = (a + b x) \pm \sqrt{2F_{2, n-2; 1-\alpha}} s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Die Breite des Bandes nimmt zu, je weiter x von \bar{x} entfernt ist.

Will man eine Intervallschätzung nicht für den bedingten Mittelwert, sondern für individuelle Werte angeben (d.h. beispielsweise individuelle Werte vorhersagen), so berechnet man ein *Vorhersageintervall* (auch *Toleranzintervall* genannt):

$$(a + b x) \pm t_{n-2; 1-\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Dieses Intervall ist wesentlich breiter als das entsprechende Konfidenzintervall.

Auch hier gilt: sollen simultan mehrere Vorhersageintervalle angegeben werden, so muss die Bonferroni-Korrektur angewandt werden.

Regression mit standardisierten Variablen

Um den relativen Beitrag einer Prädiktorvariable für die Regression besser beurteilen zu können, werden oft standardisierte Variablen verwendet.

Bei standardisierten Variablen geht die Regressionsgerade durch den Ursprung (d.h. $a = 0$), und man erhält als Steigung

$$b^* = b \frac{s_x}{s_y} = r_{xy}$$

Diese Regressionsgewichte werden als *Standardregressionsgewichte* (oft auch als *Betagewichte*) bezeichnet.

Ausreißer

Datenpunkte, die weitab von der Verteilung der restlichen Datenpunkte liegen, werden als Ausreißer bezeichnet.

Bei der linearen Regression bezieht sich dies auf die relative Lage eines Datenpunktes bezüglich der entsprechenden bedingten Verteilung von Y .

Neben dem Umstand, dass Ausreißer die Gültigkeit des gesamten Modells in Frage stellen können, haben sie auch einen relativ starken Einfluss auf die Schätzung der Parameter.

Die spezielle Wirkung ist abhängig davon, in welchem Wertebereich von X der Datenpunkt liegt:

- Liegt ein Ausreißer mehr im mittleren Wertebereich von X , so wird er vor allem auf die Schätzung des Achsenabschnitts Einfluss nehmen.
- Ist er auch ein X -Ausreißer, so beeinflusst er in stärkerem Maße die Steigung.

IDENTIFIKATION VON AUSREISSERN

Zur Identifikation von Ausreißern werden im allgemeinen individuelle Einfluss-Statistiken verwendet.

Eine erste Möglichkeit ergibt sich durch Beurteilung der Residuen $e_i = y_i - \hat{y}_i$ bzw. der *standardisierten Residuen*:

$$\frac{y_i - \hat{y}_i}{s}$$

Eine bessere Möglichkeit bieten die *studentisierten Residuen*, da sie die sog. Hebelwirkung eines Datenpunktes mit berücksichtigen:

$$\frac{y_i - \hat{y}_i}{s\sqrt{1 - h_{ii}}}$$

Die Größe h_{ii} ist der *Hebelwert*, eine der Statistiken zur Schätzung des Einflusses individueller Datenpunkte:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_k - \bar{x})^2} \qquad h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum(x_k - \bar{x})^2}$$

Der Wert von h_{ii} ist beschränkt: $\frac{1}{n} \leq h_{ii} \leq 1$.

Übrigens gilt für die Varianzen und Kovarianzen der Residuen:

$$Var(\hat{\epsilon}_i) = \sigma^2(1 - h_{ii}) \qquad Cov(\hat{\epsilon}_i, \hat{\epsilon}_j) = -\sigma^2 h_{ij}$$

Dies besagt, dass die Residuen im Mittenbereich eine größere Variabilität aufweisen als am Rande.

Der Wert von h_{ii} gibt an, in welchem Maße der Wert von y_i für den Wert von \hat{y}_i beiträgt. Eine Änderung des y_i -Wertes um eine Größe Δy brächte eine entsprechende Änderung des \hat{y}_i -Wertes um $h_{ii}\Delta y$ mit sich.

Bei der einfachen linearen Regression wird im allgemeinen ein h -Wert von größer oder gleich $4/n$ als groß betrachtet.

Einflussreiche Datenpunkte

Eine Strategie, um einflussreiche Datenpunkte zu identifizieren, besteht darin, die Werte irgendwelcher Statistiken einmal unter Verwendung aller Datenpunkte und einmal unter Ausschluss des zu untersuchenden Datenpunktes zu berechnen.

Statistiken dieser Art sind z.B.²:

- Cook's Abstand

$$D_i = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}^{(-i)})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}^{(-i)})}{\hat{\sigma}^2}$$

In der Praxis gilt ein Datenpunkt als einflussreich, falls sein Cook-Abstand größer ist als die Summe der restlichen.

- DFFITS

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_i^{(-i)}}{(\hat{\sigma}^{(-i)})^2 h_{ii}}$$

Diese Größe ist die Differenz des gefitteten Wertes an der Stelle x_i mit und ohne den i -ten Datenpunkt. Für dieses Maß gelten ähnliche Kriterien wie für den Cook'schen Abstand, da beide miteinander verwandt sind.

²In den folgenden Formeln bedeutet der Ausdruck $(-i)$ die Berechnung der entsprechenden Größe ohne den i -ten Datenpunkt.

- 'Deleted Residuals'

$$e^{(-i)} = y_i - \hat{y}_i^{(-i)}$$

Damit lässt sich dann die PRESS-Statistik³ definieren:

$$PRESS = \sum_{i=1}^n [y_i - \hat{y}_i^{(-i)}]^2$$

- DFBETAS

$$DFBETAS_{ki} = \frac{\hat{\beta}_k - \hat{\beta}_k^{(-i)}}{\hat{\sigma}^{(-i)} \sqrt{c_{kk}}}$$

c_{kk} ist das k -te Diagonalelement von $(\mathbf{X}'\mathbf{X})^{-1}$.

Analyse der Residuen

Mit Hilfe der Residuen lassen sich verschiedene Modellvoraussetzungen der linearen Regression überprüfen.

Beispielsweise lassen sich Verletzungen der Linearität und Homoskedastizität in einem Streudiagramm der gefitteten Werte vs. Residuen erkennen.

Eine mögliche fehlende Unabhängigkeit der Residuen zeigt sich in einem (z.B. linearen) Trend, wenn man sie gegen ihren Index, d.h. die Zeit, aufträgt. Man könnte aber auch einen Test auf serielle Abhängigkeit der Residuen durchführen (Durbin-Watson Statistik).

Die Annahme der (globalen) Normalverteilung lässt sich mit einem Normalverteilungstest oder Normalverteilungplot ('QQ-Plot') überprüfen.

Lineare Regression und deren Voraussetzungen

Für die Residuen wurde Normalverteilung und Homoskedastizität vorausgesetzt.

Eine moderate Verletzung der Homogenitätsvoraussetzung wirkt sich nicht wesentlich auf die Schätzung der Parameter aus. Bei größeren Verletzungen könnte man *gewichtete* Kleinste-Quadrate verwenden (mit Gewichten, die invers zu den bedingten Varianzen sind).

Abweichungen von der Normalverteilung wirken sich nur auf die Tests und die Konfidenzintervalle aus. Geringe Abweichungen werden gemeinhin toleriert, wenn der Stichprobenumfang hinreichend groß ist.

Eine wichtige Voraussetzung besagt, dass die Daten eine Zufallsstichprobe aus der Population darstellen, auf die generalisiert werden soll. Ist dies nicht der Fall, so sind die Tests und Konfidenzintervalle ungültig (z.B. die Konfidenzintervalle zu eng).

Ist die Homogenitätsvoraussetzung verletzt, so kann man auch versuchen, durch eine geeignete Transformation der Responsevariable (z.B. Quadratwurzel- oder Logarithmustransformation) die Varianz zu stabilisieren.

Durch eine geeignete Transformation der Prädiktorvariable kann eventuell auch eine nichtlineare Beziehung linearisiert werden (z.B. Quadrat-, Logarithmus-, Quadratwurzeltransformation u.a.).

Weitere Formen der Regression

- *Regression durch den Ursprung*: das Modell enthält keinen Parameter α
- *Gewichtete Kleinste-Quadrate Regression*: die einzelnen Beobachtungen gehen mit unterschiedlichem Gewicht in die Analyse ein, was in folgenden Situationen sinnvoll sein könnte:
 - die X -Werte sind Mittelwerte von Teilstichproben,

³PRESS = „prediction sum of squares“

2 Regressionsanalyse

- die X -Werte sind unterschiedlich wichtig,
- die bedingten Verteilungen haben unterschiedliche Varianzen.
- *Loess-Kurven* (Lowess-Regression): lokale Regressionsfunktionen
- *Robuste Regressionsmethoden*: Ausreißer werden eliminiert oder mit niedrigem Gewicht versehen.

Residuen und partielle bzw. semipartielle Korrelation

Für die partielle Korrelation zwischen zwei Variablen X und Y unter Ausschaltung einer Kontrollvariable Z gilt:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Die Ausschaltung („Auspartialisierung“) einer Variable Z aus einer Variable X besteht darin, die Regressionsresiduen $e_{x.z} = X - (a_{x.z} + b_{x.z}Z)$ zu berechnen.

Die partielle Korrelation ist also eine Korrelation zwischen Residuen, d.h. im obigen Fall zwischen $e_{x.z}$ und $e_{y.z}$.

Bei der semipartiellen Korrelation wird die Kontrollvariable nur aus einer der beiden Variablen auspartialisiert, d.h. es wird $r(e_{x.z}, Y)$ bzw. $r(e_{y.z}, X)$ berechnet.

2.2 Univariate multiple Regressionsanalyse (MR)

Wir führen nun weitere unabhängige Variablen (Prädiktorvariablen, Designvariablen) ein und postulieren: Zwischen einer beobachtbaren Variable Y und mehreren beobachtbaren Variablen X_1, X_2, \dots, X_p besteht eine lineare Beziehung der Form⁴

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

In Matrizenschreibweise:

$$Y = [1 \quad X_1 \quad \dots \quad X_p] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon$$

Für die *Daten* von n Beobachtungen gilt also:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \dots & x_{ik} & \dots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Also kurzum:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

\mathbf{y} : Vektor der Responsewerte

\mathbf{X} : Matrix der Prädiktorvariablen mit einer zusätzlichen Einer-Spalte vorne

$\boldsymbol{\beta}$: der unbekannte Parametervektor

$\boldsymbol{\varepsilon}$: der unbekannte Fehlervektor

⁴Um die Notation zu vereinheitlichen, wird die additive Konstante jetzt mit β_0 und nicht wie bisher mit α bezeichnet.

ANNAHMEN

- X_1, \dots, X_p sind feste, fehlerfrei gemessene Variablen.
- β_0, \dots, β_p sind feste, unbekannte Parameter.
- $\boldsymbol{\varepsilon}$ ist ein unbekannter, n -dimensional multivariat normalverteilter Fehlervektor mit:

$$\boldsymbol{\varepsilon} = [\varepsilon_1 \dots \varepsilon_n]' \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

LÖSUNG

Die Zielfunktion ist wie bei der einfachen Regression das Kriterium der kleinsten Quadrate:

$$\mathbf{e}'\mathbf{e} = \sum_i (y_i - \hat{y}_i)^2 = \sum_i [y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})]^2 \rightarrow \text{Min}$$

Wenn man diese Größe partiell nach den einzelnen Parametern ableitet und die Ableitungen auf Null setzt, bekommt man ein Gleichungssystem mit $(p+1)$ Gleichungen und $(p+1)$ Unbekannten:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

$\mathbf{X}'\mathbf{X}$ ist die Matrix der (rohen) Kreuzprodukte.

Die letzte Gleichung lautet ausgeschrieben:

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ 1 & x_{31} & \dots & x_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Nach Ausführung der Matrizenmultiplikationen erhält man⁵:

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \dots & \sum x_{i1}x_{ip} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 & \dots & \sum x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip} & \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \dots & \sum x_{ip}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ip}y_i \end{bmatrix}$$

Ausgeschrieben ergibt dies das Gleichungssystem:

$$\begin{aligned} n \hat{\beta}_0 + \sum x_{i1} \hat{\beta}_1 + \dots + \sum x_{ip} \hat{\beta}_p &= \sum y_i \\ \sum x_{i1} \hat{\beta}_0 + \sum x_{i1}^2 \hat{\beta}_1 + \dots + \sum x_{i1}x_{ip} \hat{\beta}_p &= \sum x_{i1}y_i \\ \sum x_{i2} \hat{\beta}_0 + \sum x_{i2}x_{i1} \hat{\beta}_1 + \dots + \sum x_{i2}x_{ip} \hat{\beta}_p &= \sum x_{i2}y_i \\ \vdots & \\ \sum x_{ip} \hat{\beta}_0 + \sum x_{ip}x_{i1} \hat{\beta}_1 + \dots + \sum x_{ip}^2 \hat{\beta}_p &= \sum x_{ip}y_i \end{aligned}$$

⁵Summiert wird jeweils über den Index i .

2 Regressionsanalyse

Es geht also darum, dieses Gleichungssystem nach den Unbekannten $\hat{\beta}_0, \dots, \hat{\beta}_p$ aufzulösen.

Die Gleichung $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ lässt sich nach $\hat{\boldsymbol{\beta}}$ auflösen, indem man sie von links her mit der Inversen von $\mathbf{X}'\mathbf{X}$ multipliziert:

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})}_{=I} \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Als Lösung ergibt sich also:

$$\hat{\boldsymbol{\beta}} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Die Voraussetzung dafür ist, dass $(\mathbf{X}'\mathbf{X})^{-1}$ existiert, d.h. dass die Matrix $\mathbf{X}'\mathbf{X}$ nicht singulär ist.

Verwendet man anstelle der Matrix der Rohproduktschätzungen $\mathbf{X}'\mathbf{X}$ die Kovarianzmatrix der Prädiktorvariablen (\mathbf{S}_{XX} , kurz mit \mathbf{S} bezeichnet) sowie den Vektor der Kovarianzen der Prädiktorvariablen mit der Response (\mathbf{s}_{Xy} , kurz mit \mathbf{s}_y bezeichnet), so erhält man ebenfalls die Kleinsten Quadrate Schätzungen der Regressionsparameter β_1, \dots, β_p (auch hier gilt als Voraussetzung, dass \mathbf{S} nicht singulär ist):

$$[\hat{\beta}_1 \dots \hat{\beta}_p]' = [b_1 \dots b_p]' = \mathbf{S}^{-1} \mathbf{s}_y$$

Der Parameter β_0 muss extra geschätzt werden:

$$\hat{\beta}_0 = b_0 = \bar{y} - \sum b_k \bar{x}_k$$

Geht man hingegen von der Korrelationsmatrix der Prädiktorvariablen (\mathbf{R}_{XX} , kurz mit \mathbf{R} bezeichnet) sowie den Validitäten, d.h. dem Vektor der Korrelationen der Prädiktorvariablen mit der Response (\mathbf{r}_{Xy} , kurz mit \mathbf{r}_y bezeichnet) aus, so erhält man die standardisierten Regressionsgewichte („Betagewichte“):

$$\mathbf{b}^* = [b_1^* \dots b_p^*]' = \mathbf{R}^{-1} \mathbf{r}_y$$

Statistische Eigenschaften der Schätzer

Wir haben den Zusammenhang zwischen einer Response und mehreren Prädiktorvariablen wie folgt modelliert:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Unter Zugrundelegung der quadratischen Verlustfunktion ergab sich als Schätzer des Parametervektors $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Gelten die Voraussetzungen für das Modell, dann ist dieser Schätzer erwartungstreu⁶ und multivariat normalverteilt, d.h. dann gilt:

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$$

Damit lassen sich wieder Konfidenzintervalle und t-Tests für die einzelnen Regressionsparameter bilden.

Berechnung der gefitteten Werte

Die gefitteten Werte sind wie folgt zu berechnen:

$$\hat{y}_i = \sum_{k=0}^p b_k x_{ik}$$

In Matrixschreibweise:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{=: \mathbf{H}} \mathbf{y} = \mathbf{H}\mathbf{y}$$

Die Matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ wird auch „Hat-Matrix“ genannt.

⁶Man spricht hier auch von einem „BLUE“ (Best Linear Unbiased Estimator).

EIGENSCHAFTEN DER HAT-MATRIX

- Die Hat-Matrix ist quadratisch von der Ordnung n . Mit ihrer Hilfe wird der Vektor \mathbf{y} auf den Vektor $\hat{\mathbf{y}}$ abgebildet.
- Die Hat-Matrix ist symmetrisch:

$$\mathbf{H}' = (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' = (\mathbf{X}')'[(\mathbf{X}'\mathbf{X})^{-1}]'\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H}$$

(Hinweis: $\mathbf{X}'\mathbf{X}$ ist symmetrisch, also auch $(\mathbf{X}'\mathbf{X})^{-1}$).

- Sie ist idempotent:

$$\mathbf{H}^2 = \mathbf{H}\mathbf{H} = \mathbf{X} \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{=\mathbf{I}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H}$$

- Die Diagonalelemente der Matrix \mathbf{H} sind die Hebelwerte h_{ii} , die bekanntlich individuelle Einfluss-Statistiken darstellen.

Standardschätzfehler

Als globales Abweichungsmaß wird RSS, die Residuenquadratsumme, verwendet:

$$RSS = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \sum (y_i - \hat{y}_i)^2$$

Die Anzahl der Freiheitsgrade ist gleich $n - p - 1$. Also ist der mittlere quadratische Fehler (MSE):

$$MSE = \frac{RSS}{n - p - 1}$$

Dieser ist eine erwartungstreue Schätzung der Fehlervarianz σ^2 . Die Quadratwurzel aus MSE ergibt s , den Standardfehler der Schätzung:

$$s = \hat{\sigma} = \sqrt{\frac{RSS}{n - p - 1}}$$

Test des linearen Regressionsmodells: Varianzanalyse

Für den Modelltest wird wiederum eine Quadratsummenzerlegung durchgeführt.

Die totale Quadratsumme von Y wird zerlegt in die Quadratsumme der Regression („Quadratsumme auf der Regression“) und die Quadratsumme des Residuums („Quadratsumme um die Regression“):

$$\sum (y_i - \bar{y}.)^2 = \sum (\hat{y}_i - \bar{y}.)^2 + \sum (y_i - \hat{y}_i)^2$$

Die Freiheitsgrade werden in analoger Weise zerlegt:

$$\frac{df_{total}}{n - 1} = \frac{df_{regression}}{p} + \frac{df_{residuum}}{n - p - 1}$$

Wir erhalten also die folgende Varianzanalysetabelle:

Quelle	QS	df	MQ	F
Regression	$\sum (\hat{y}_i - \bar{y}.)^2$	p	$\frac{\sum (\hat{y}_i - \bar{y}.)^2}{p}$...
Residuum	$\sum (y_i - \hat{y}_i)^2$	$n - p - 1$	$\frac{\sum (y_i - \hat{y}_i)^2}{n - p - 1}$	
Total	$\sum (y_i - \bar{y}.)^2$	$n - 1$		

Unter der Nullhypothese ist die Statistik

$$F = \frac{\sum(\hat{y}_i - \bar{y})^2/p}{\sum(y_i - \hat{y}_i)^2/(n-p-1)}$$

F-verteilt mit $df_1 = p$ und $df_2 = n - p - 1$ Freiheitsgraden.

Multipler Korrelationskoeffizient

Die einfache Korrelation zwischen den beobachteten und den gefitteten Responsewerten, $\rho = \rho(Y, \hat{Y})$, wird *multipler Korrelationskoeffizient* genannt.

ρ^2 stellt den Anteil der durch die Regression erklärten Varianz dar und $1 - \rho^2$ den Anteil der Residualvarianz, und daraus folgt:

$$\sigma_\varepsilon^2 = \sigma_Y^2(1 - \rho^2)$$

In der Stichprobe wird der multiple Korrelationskoeffizient mit R bezeichnet. Um anzudeuten, dass es sich um die Regression von Y auf X_1, X_2, \dots, X_p handelt, kann man auch die Bezeichnung $R_{y.12\dots p}$ verwenden.

In der F-Statistik wird letztlich der Anteil der aufgeklärten Varianz ($R_{y.12\dots p}^2$) zum Anteil der Restvarianz ($1 - R_{y.12\dots p}^2$) in Beziehung gesetzt:

$$F = \frac{R_{y.12\dots p}^2/p}{(1 - R_{y.12\dots p}^2)/(n-p-1)}$$

Der multiple Korrelationskoeffizient R ist nicht erwartungstreu, denn ρ wird im Mittel überschätzt, gerade wenn viele Prädiktorvariablen verwendet werden. Daher sind verschiedene Biaskorrekturen vorgeschlagen worden, z.B.:

$$R_{korr}^2 = R^2 - \frac{p}{n-p-1}(1 - R^2)$$

Konfidenz- und Vorhersageintervalle

Ein punktweises *Konfidenzintervall* für die Regressionsfunktion an der Stelle \mathbf{x} ist wie folgt zu berechnen:

$$\mathbf{x}'\mathbf{b} \pm t_{n-p-1; 1-\alpha/2} s \sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

Hinweis: falls \mathbf{x}' eine Zeile aus der Datenmatrix \mathbf{X} ist, stellt der Ausdruck unter dem Wurzelzeichen das entsprechende Diagonalelement der Hat-Matrix dar.

Sollen simultan mehrere Konfidenzintervalle berechnet werden, so ist nach Bonferroni das α entsprechend zu korrigieren, d.h. durch die Anzahl der Intervalle zu teilen.

Es lässt sich auch ein *Konfidenzband* für die gesamte Regressionsgerade angeben:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{b} \pm \sqrt{pF_{p, n-p-1; 1-\alpha}} s \sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

Die Breite des Bandes nimmt zu, je weiter \mathbf{x} vom Mittelwertsvektor $\bar{\mathbf{x}} = [\bar{x}_1 \ \bar{x}_2 \ \dots \ \bar{x}_p]'$ entfernt ist.

Eine Intervallschätzung für einen *individuellen* Vektor \mathbf{x} wird durch ein *Vorhersage-* bzw. *Toleranzintervall* dargestellt:

$$\mathbf{x}'\mathbf{b} \pm t_{n-p-1; 1-\alpha/2} \hat{\sigma} \sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$$

Dieses Intervall ist wesentlich breiter als das entsprechende Konfidenzintervall.

Auch hier gilt: sollen simultan mehrere Vorhersageintervalle angegeben werden, so ist die Bonferroni-Korrektur anzuwenden.

Standardisierte Koeffizienten

Standardisierte Regressionskoeffizienten (sog. Betagewichte) b_i^* erhält man, wenn man durchweg standardisierte Variablen verwendet (siehe oben).

Sie hängen mit den unstandardisierten Regressionskoeffizienten b_i wie folgt zusammen:

$$b_i^* = b_i \frac{s_i}{s_y}$$

s_i ist die Standardabweichung der Variable X_i und s_y die Standardabweichung der Response.

Die standardisierten Regressionsgewichte erleichtern die Interpretation des relativen Beitrags der einzelnen Variablen insofern, als sie nicht von deren Varianz abhängen.

Mit Hilfe der standardisierten Regressionskoeffizienten lässt sich das Quadrat des multiplen Korrelationskoeffizienten auch als Linearkombination der Korrelationen der Prädiktorvariablen mit der Response⁷ ausdrücken:

$$R^2 = \mathbf{b}^* \mathbf{r}_y = \mathbf{r}_y' \mathbf{R}^{-1} \mathbf{r}_y$$

Interpretation der Regressionsgewichte

Da die Prädiktorvariablen normalerweise untereinander korrelieren, ist bei der Interpretation der Regressionsgewichte Vorsicht geboten.

Standardisierte Regressionsgewichte erleichtern zwar die Einschätzung des relativen Beitrags der einzelnen Prädiktorvariablen für die Regressionsgleichung, sie sind aber keine festen, isolierten Größen, sondern von der speziellen Menge der in der Regressionsgleichung verwendeten Variablen abhängig.

Wird also beispielsweise aus dieser Menge eine Prädiktorvariable entfernt oder eine hinzugenommen, so ändern sich normalerweise alle Regressionsgewichte.

Suppressoreffekte

Eine scheinbar paradoxe Situation liegt vor, wenn eine Prädiktorvariable ein hohes (meist negatives) Gewicht erhält, obwohl sie mit der Response nicht oder nur gering korreliert.

Eine solche Variable korreliert aber mit anderen Prädiktorvariablen hoch und unterdrückt bei diesen jenen Varianzanteil, der für die Regression irrelevant ist.

„Der Effekt einer Suppressorvariable besteht also darin, dass sie die Nützlichkeit anderer Prädiktorvariablen erhöht“ (Bortz).

Multiple Regression mit orthogonalen Prädiktorvariablen

Sind alle Prädiktorvariablen paarweise orthogonal, d.h. unkorreliert, so sind die standardisierten Regressionsgewichte gleich den Korrelationen zwischen den Prädiktorvariablen und der Response:

$$\mathbf{b}^* = \mathbf{r}$$

In diesem Falle gilt dann auch:

$$R^2 = \mathbf{r}' \mathbf{r} = \sum_{k=1}^p r_{ky}^2$$

⁷Diese Korrelationen könnte man auch als Validitäten der Prädiktorvariablen bezeichnen.

Multikollinearität und Toleranz

Multikollinearität entsteht, wenn die Prädiktorvariablen hoch miteinander korrelieren, so dass die Spalten der Matrix \mathbf{X} fast linear abhängig sind.

Die Parameterschätzungen sind dann instabil und haben einen großen Stichprobenfehler, denn es gilt:

$$[SE(b_k)]^2 = \frac{MSE}{(n-1)s_k^2} \cdot \frac{1}{1-R_k^2}$$

(MSE ist der mittlere quadratische Fehler, s_k^2 die Varianz von X_k und R_k der multiple Korrelationskoeffizient der Regression von X_k auf die restlichen X -Variablen.)

Der Ausdruck $1 - R_k^2$ stellt den durch die restlichen Prädiktorvariablen an X_k nicht erklärten Varianzanteil dar und wird als *Toleranz* bezeichnet. Deren Kehrwert ist der *Varianz-Inflations-Faktor* (VIF).

Variablenselektion und Test verschiedener Modelle

In der Praxis geht es oft darum, aus einer Menge von Prädiktorvariablen eine optimale Teilmenge auszuwählen, d.h. ein optimales Modell zu finden.

„Optimal“ ist ein Modell, wenn es einerseits möglichst viel Varianz aufklärt und andererseits möglichst sparsam ist.

Dies gelingt solange, wie redundante Prädiktoren (d.h. solche mit verschwindendem Gewicht) aus dem Modell entfernt werden können.

Normalerweise jedoch lassen sich die beiden Ziele nicht gleichzeitig erreichen, denn mit Hinzunahme einer weiteren Prädiktorvariable kann der multiple Korrelationskoeffizient nicht kleiner werden.

VERGLEICH HIERARCHISCHER MODELLE

Unter dem *Nullmodell* versteht man das Regressionsmodell, das lediglich die additive Konstante enthält.

Das *volle Modell* ist jenes, das sämtliche Prädiktorvariablen enthält.

Wenn wir ein Ausgangsmodell mit u Prädiktorvariablen um v zusätzliche Prädiktorvariablen erweitern, wird sich R^2 um einen bestimmten Betrag vergrößern. Dieser Zuwachs ($R_{change}^2 = R_{u+v}^2 - R_u^2$) lässt sich auf Signifikanz prüfen:

$$F = \frac{(R_{u+v}^2 - R_u^2)/v}{(1 - R_{u+v}^2)/(n - u - v - 1)}$$

Diese Größe ist unter $H_0: \rho_{change}^2 = 0$ F-verteilt mit $df_1 = v$ und $df_2 = n - u - v - 1$ Freiheitsgraden.

Man spricht hier auch von „hierarchischem Hypothesentesten“, da die zu vergleichenden Modelle in hierarchischer Beziehung zueinander stehen.

Es gibt verschiedene Strategien bei der hierarchischen Suche nach dem optimalen Modell:

- *Aufsteigendes Verfahren* („forward“): ausgehend vom Nullmodell wird Schritt für Schritt jeweils jene Prädiktorvariable hinzugenommen, für die R_{change}^2 am größten ist.
- *Absteigendes Verfahren* („backward“): ausgehend vom vollen Modell wird Schritt für Schritt jeweils jene Prädiktorvariable entfernt, für die R_{change}^2 am kleinsten ist.
- *Schrittweises Verfahren* („stepwise“): diese Strategie ist eine Mischung aus den beiden obigen Strategien, d.h. es wird zunächst das aufsteigende Verfahren angewandt und bei jedem Schritt geprüft, ob eine im augenblicklichen Modell enthaltene Prädiktorvariable wieder entfernt werden kann.

All diese Verfahren werden abgebrochen, sobald ein bestimmtes Kriterium nicht mehr erfüllt ist (z.B. wenn R_{change}^2 nicht mehr signifikant ist).

VERGLEICH NICHT-HIERARCHISCHER MODELLE

Der oben aufgeführte F-Test eignet sich nur für den Vergleich von hierarchischen Modellen, d.h. deren Prädiktorvariablen zueinander in Teilmengenrelation stehen.

Für den Vergleich von nicht-hierarchischen Modellen sind die folgenden Kriterien vorgeschlagen worden⁸:

- Mallow's C_p -Statistik:

$$C_p = (n-p-1) \left(\frac{MSE_p}{MSE_{full}} - 1 \right) + (p+1)$$

- Akaike's Informationskriterium:

$$AIC = n \ln \left(\frac{RSS_p}{n} \right) + 2(p+1)$$

Das Modell mit dem niedrigsten C_p bzw. AIC wird genommen.

Polynomiale Regression

Um eine kurvilineare Beziehung zwischen einer Prädiktorvariable und einer Response zu modellieren, besteht die Möglichkeit der Anwendung einer polynomialen Regression.

Dabei wird das folgende Modell angenommen:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_p X^p + \varepsilon$$

Wir definieren nun neue Variablen:

$$X_1 := X \quad X_2 := X^2 \quad \dots \quad X_p := X^p$$

und analysieren das folgende lineare Regressionsmodell:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Modelle mit Interaktionen

Auf ähnliche Weise wie bei der polynomialen Regression kann ein Modell mit Interaktionen analysiert werden:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p+1} X_1 X_2 + \dots + \varepsilon$$

Wir definieren neue Variablen:

$$X_1 := X_1 \quad X_2 := X_2 \quad \dots \quad X_{p+1} := X_1 X_2 \quad \dots$$

und analysieren ein lineares Regressionsmodell mit diesen Variablen.

Einflussreiche Datenpunkte

Eine Strategie, um einflussreiche Datenpunkte zu identifizieren, besteht auch bei der multiplen Regression darin, die Werte irgendwelcher Statistiken einmal unter Verwendung aller Datenpunkte und einmal unter Ausschluss des zu untersuchenden Datenpunktes zu berechnen.

Die Statistiken sind dieselben wie bei der einfachen linearen Regression⁹:

⁸Im folgenden bezieht sich der Index *full* auf das volle Modell und der Index *p* auf ein Modell mit *p* Prädiktorvariablen.

⁹In den folgenden Formeln bedeutet der Ausdruck $(^{-i})$ wiederum die Berechnung der entsprechenden Größe ohne den *i*-ten Datenpunkt.

2 Regressionsanalyse

- Cook's Abstand

$$D_i = \frac{(\hat{\mathbf{y}} - \hat{\mathbf{y}}^{(-i)})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}^{(-i)})}{p \hat{\sigma}^2}$$

p ist die Anzahl der Prädiktorvariablen. In der Praxis gilt ein Datenpunkt als einflussreich, falls sein Cook-Abstand größer ist als die Summe der restlichen.

- DFFITS

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_i^{(-i)}}{(\hat{\sigma}^{(-i)})^2 h_{ii}}$$

Diese Größe ist die Differenz des gefitteten Wertes an der Stelle x_i mit und ohne den i -ten Datenpunkt. Für dieses Maß gelten ähnliche Kriterien wie für den Cook'schen Abstand, da beide miteinander verwandt sind.

- 'Deleted Residuals'

$$e^{(-i)} = y_i - \hat{y}_i^{(-i)}$$

Damit lässt sich dann die PRESS-Statistik¹⁰ definieren:

$$PRESS = \sum_{i=1}^n [y_i - \hat{y}_i^{(-i)}]^2$$

- DFBETAS

$$DFBETAS_{ki} = \frac{\hat{\beta}_k - \hat{\beta}_k^{(-i)}}{\hat{\sigma}^{(-i)} \sqrt{c_{kk}}}$$

c_{kk} ist das k -te Diagonalelement von $(\mathbf{X}'\mathbf{X})^{-1}$.

Analyse der Residuen

In gleicher Weise wie der einfachen linearen Regression lassen sich mit Hilfe der Residuen verschiedene Modellvoraussetzungen überprüfen.

Beispielsweise zeigen sich Verletzungen der Linearität und Homoskedastizität in einem Streudiagramm der gefitteten Werte vs. Residuen.

Eine mögliche fehlende Unabhängigkeit der Residuen zeigt sich in einem (z.B. linearen) Trend, wenn man sie gegen ihren Index, d.h. die Zeit, aufträgt. Man könnte aber auch einen Test auf serielle Abhängigkeit der Residuen durchführen (Durbin-Watson Statistik).

Die Annahme der (globalen) Normalverteilung lässt sich mit einem Normalverteilungstest oder Normalverteilungplot ('QQ-Plot') überprüfen.

¹⁰PRESS = „prediction sum of squares“

2.3 Multivariate multiple Regressionsanalyse (MMR)

Zweck

Die multivariate multiple Regression (MMR) ist eine Verallgemeinerung der (univariaten) multiplen Regression.

Wir erweitern diese, indem wir nicht nur eine, sondern q Responsevariablen zulassen.

Für jede der Responsevariablen werden entweder (a) dieselben p oder (b) jeweils p_l verschiedene Prädiktorvariablen verwendet¹¹.

Das Modell enthält natürlich für jede Responsevariable einen eigenen Satz von Regressionsparametern.

Die Schätzungen dieser Gewichte sowie der Standardfehler sind zwar dieselben wie jene der separaten Regressionen, dennoch bietet die MMR den Vorteil, dass beim Modelltest auch die Kovarianzen zwischen den einzelnen Schätzern berücksichtigt werden können.

Modell

Das Modell der MMR sieht folgendermaßen aus¹²:

$$\mathbf{Y}_{n \times q} = \mathbf{X}_{n \times (p+1)} \mathbf{B}_{(p+1) \times q} + \boldsymbol{\varepsilon}_{n \times q}$$

\mathbf{Y} Matrix der Responsevariablen

\mathbf{X} Matrix der Prädiktorvariablen

\mathbf{B} Matrix der Regressionsparameter

$\boldsymbol{\varepsilon}$ Matrix mit normalverteilten, unabhängigen Zufallsvariablen:

$$\varepsilon_{il} \sim \mathcal{N}_n(\mathbf{0}, \sigma_l^2 \mathbf{I})$$

Parameterschätzung

Die Verlustfunktion bei der MMR entspricht dem univariaten Fall:

$$\text{sp}[(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})] = \|\mathbf{Y} - \mathbf{X}\mathbf{B}\|^2 \rightarrow \text{Min}$$

Als BLUE-Schätzer¹³ der Parametermatrix \mathbf{B} ergibt sich damit völlig analog zur multiplen Regression:

$$\hat{\mathbf{B}} = \mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Daraus sieht man, dass die l -te Spalte der Matrix \mathbf{B} wie folgt berechnet wird:

$$\hat{\boldsymbol{\beta}}_l = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}_l$$

Das bedeutet, dass hier q einzelne multiple Regressionen zusammengefasst werden.

Trotzdem bringt die MMR gegenüber einzelnen univariaten multiplen Regressionen einen Vorteil, denn dadurch, dass mehrere Responsevariablen gleichzeitig betrachtet werden, gehen auch deren Kovarianzen in die Analyse und in den Modelltest ein.

¹¹Im folgenden betrachten wir nur den Fall (a).

¹²Die Symbole \mathbf{B} und $\boldsymbol{\varepsilon}$ sollen die Großbuchstaben von $\boldsymbol{\beta}$ und $\boldsymbol{\varepsilon}$ darstellen.

¹³BLUE = Best Linear Unbiased Estimator

Modelltest

Die zu testende Hypothese betrifft normalerweise den linearen Zusammenhang zwischen den beiden Variablenmengen. Sie bezieht sich also nur auf die Matrix der Regressionsgewichte und nicht auf den Vektor der additiven Konstanten, d.h. auf die Matrix \mathcal{B} ohne die erste Zeile.

Dazu stellen wir die Parametermatrix \mathcal{B} auf folgende Weise dar:

$$\mathcal{B} = \begin{bmatrix} \boldsymbol{\beta}'_0 \\ \mathcal{B}_1 \end{bmatrix}$$

$\boldsymbol{\beta}'_0$ ist der Zeilenvektor der additiven Konstanten und \mathcal{B}_1 die Matrix der Regressionsgewichte.

Die Nullhypothese lautet dann:

$$H_0: \mathcal{B}_1 = \mathbf{0}$$

Für die Berechnung der Testgröße wird die gesamte Kovarianzmatrix (\mathbf{S}) partitioniert in die Kovarianzmatrizen der X - und Y -Variablen (\mathbf{S}_{XX} und \mathbf{S}_{YY}), sowie deren gemeinsame Kovarianzmatrix \mathbf{S}_{XY} bzw. \mathbf{S}_{YX} :

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix}$$

Die Testgröße ist *Wilks' Lambda*:

$$\Lambda = \frac{|\mathbf{S}|}{|\mathbf{S}_{XX}||\mathbf{S}_{YY}|}$$

Die Testgröße Λ lässt sich auch mit Hilfe der Matrizen \mathbf{X}_d und \mathbf{Y}_d (Abweichungsmatrizen der \mathbf{X} - bzw. \mathbf{Y} -Variablen) berechnen:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{T}|} = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|}$$

$$\begin{aligned} \mathbf{T} &= \mathbf{Y}'_d \mathbf{Y}_d \\ \mathbf{H} &= \mathbf{B}'_1 (\mathbf{X}'_d \mathbf{X}_d) \mathbf{B}_1 \\ \mathbf{E} &= \mathbf{T} - \mathbf{H} \end{aligned}$$

Die Matrizen \mathbf{T} , \mathbf{H} und \mathbf{E} sind SSCP-Matrizen und werden als „Totalmatrix“, „Hypothesenmatrix“ bzw. „Errormatrix“ bezeichnet.

Aus der Testgröße Λ lässt sich nach Bartlett eine Teststatistik ableiten, die asymptotisch χ^2 -verteilt ist:

$$V = -[(n-1) - (p+q)/2] \ln \Lambda \quad \text{mit} \quad df = q(p-1)$$

WILKS' LAMBDA

- Wilks' Lambda (auch U-Statistik genannt) ist eine Testgröße, die in der multivariaten Statistik eine wichtige Rolle spielt.
- Ihr Wertebereich liegt zwischen 0 und 1, wobei Werte nahe 0 für die Ablehnung von H_0 sprechen.
- Wilks' Lambda ist ein Quotient von Determinanten, kann aber auch als Funktion der Eigenwerte bestimmter SSCP-Matrizen ausgedrückt werden (s.u.).

- Die Verteilung der Testgröße unter H_0 ist die U-Verteilung, eine dreiparametrische Wahrscheinlichkeitsverteilung, die durch eine entsprechende F-Verteilung oder eine Chiquadratverteilung¹⁴ approximiert werden kann.

¹⁴siehe oben

3 Lineare Modelle

3.1 Das Allgemeine Lineare Modell

Im Allgemeinen Linearen Modell (ALM)¹ wird die Varianzanalyse als Spezialfall der multiplen linearen Regression dargestellt.

3.1.1 Einfache Varianzanalyse

Wir betrachten ein varianzanalytisches Design für m unabhängige Gruppen und einer abhängigen Variable Y .

Wir wollen testen, ob zwischen den m Populationen, aus denen die Gruppen stammen, ein Y -Mittelwertsunterschied besteht.

Für Y_{gi} , den Messwert der i -ten Vp in der g -ten Gruppe, nehmen wir das folgende Modell an:

$$Y_{gi} = \mu_g + \varepsilon_{gi}$$

μ_g : Erwartungswert von Y in der g -ten Population

ε_{gi} : normalverteilte, unabhängige Zufallsvariablen mit gleicher unbekannter Varianz σ^2 .

Varianzanalyse und Regression

Wir wollen nun das obige Modell als lineares Regressionsmodell schreiben:

$$\begin{aligned} Y_{gi} &= \mu_g + \varepsilon_{gi} \\ &= \sum \beta_k X_k + \varepsilon_{gi} \end{aligned}$$

$$\begin{aligned} \hat{Y}_{gi} &= \hat{\mu}_g \\ &= \sum \hat{\beta}_k X_k \end{aligned}$$

Die Prädiktorvariablen X_k werden durch die unabhängige Variable, nämlich zu welcher Gruppe eine bestimmte Vp gehört, definiert, sie sind also Codiervariablen.

Die einfachste Möglichkeit der Codierung besteht darin, für jede Gruppe eine eigene Variable zu verwenden, die den Wert 1 hat, wenn die Vp zu dieser Gruppe gehört, und sonst den Wert 0. Diese Form der Codierung heißt *Dummycodierung*.

Die Regressionsparameter β_g sind dann nichts anderes als die Populationsmittelwerte μ_g .

Wir schreiben nun diese Codiervariablen als Spalten einer Matrix, der sog. *Designmatrix*, hin.

Wenn wir die Vpn nach den Gruppen sortieren, können wir das varianzanalytische Modell in folgender Weise hinschreiben:

¹engl.: General Linear Model (GLM)

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ \vdots \\ y_{g1} \\ \vdots \\ y_{gn} \\ \vdots \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_g \\ \vdots \\ \mu_m \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n} \\ \vdots \\ \varepsilon_{g1} \\ \vdots \\ \varepsilon_{gn} \\ \vdots \\ \vdots \\ \varepsilon_{m1} \\ \vdots \\ \varepsilon_{mn} \end{bmatrix}$$

Also kurzum:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

\mathbf{y} : Wertevektor der abhängigen Variablen

\mathbf{X} : Designmatrix

$\boldsymbol{\mu}$: der unbekannte Parametervektor

$\boldsymbol{\varepsilon}$: der unbekannte Fehlervektor

Wir sehen also, dass wir eine lineare Regressionsgleichung vorliegen haben.

Das primäre Ziel der Varianzanalyse besteht allerdings nicht darin, die Werte Y_{gi} zu schätzen, sondern die Hypothese

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$$

zu testen.

Dafür müssen die Parameter μ_i geschätzt werden, wobei wieder das Kriterium der kleinsten Quadrate ins Spiel kommt:

$$\sum_g \sum_i (y_{gi} - \hat{\mu}_g)^2 \rightarrow \text{Min}$$

STRUKTURMODELL

Die Hypothese $H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$ wird üblicherweise mit Hilfe der sog. Effekte $\alpha_g = \mu_g - \mu$ formuliert:

$$H_0 : \alpha_g = 0 \quad \text{für alle } g$$

Dementsprechend soll nun eine Reparametrisierung des Modells vorgenommen werden: im sog. Strukturmodell der einfachen Varianzanalyse werden die Parameter μ (Mittelwert der Gesamtpopulation) und die Effekte α_g verwendet um zu beschreiben, wie der Wert der abhängigen Variable bei der Vp i in der Gruppe g zustande kommt:

$$Y_{gi} = \mu + \alpha_g + \varepsilon_{gi}$$

Für die Gesamtgruppe sieht dann diese Gleichung wie folgt aus²:

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ \vdots \\ y_{g1} \\ \vdots \\ y_{gn} \\ \vdots \\ \vdots \\ y_{m1} \\ \vdots \\ y_{mn} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_g \\ \vdots \\ \alpha_m \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n} \\ \vdots \\ \epsilon_{g1} \\ \vdots \\ \epsilon_{gn} \\ \vdots \\ \vdots \\ \epsilon_{m1} \\ \vdots \\ \epsilon_{mn} \end{bmatrix}$$

Also kurzum:

$$y = X\beta + \epsilon$$

- y: Wertevektor der abhängigen Variablen
- X: Designmatrix
- β: der unbekannte Parametervektor mit $\beta = [\mu \ \alpha_1 \ \alpha_2 \ \dots \ \alpha_m]'$
- ε: der unbekannte Fehlervektor

Diese Gleichung entspricht einer linearen Regressionsgleichung. Wir können also die Varianzanalyse wie eine multiple lineare Regressionsanalyse behandeln und die Schätzung der Parameter, die Hypothesentests usw. wie bei dieser durchführen.

Leider zeigt sich aber, dass die Spalten der Matrix X linear abhängig sind³ und daher $(X'X)^{-1}$ nicht existiert. Dies liegt daran, dass im varianzanalytischen Modell die Parameter einer Restriktion unterliegen:

$$\sum \alpha_g = 0$$

Man löst das Problem durch eine weitere Reparametrisierung. Dafür gibt es viele Möglichkeiten. Jede neue Parametrisierung bedeutet natürlich auch eine Änderung der Designmatrix.

- Vorschlag 1:

Wir streichen aus der Designmatrix X die letzte Spalte. Diese Maßnahme bedeutet folgende Parametrisierung:

$$\begin{aligned} \beta &= [\ \mu_m \quad \mu_1 - \mu_m \quad \mu_2 - \mu_m \quad \dots \quad \mu_{m-1} - \mu_m]' \\ &= [\mu + \alpha_m \quad \alpha_1 - \alpha_m \quad \alpha_2 - \alpha_m \quad \dots \quad \alpha_{m-1} - \alpha_m]' \end{aligned}$$

- Vorschlag 2 (Effektcodierung):

²Zur Vereinfachung der Notation wählen wir gleiches n in den Gruppen.
³Die erste Spalte von X ist die Summe der restlichen Spalten.

Wir verwenden die folgende Parametrisierung:

$$\boldsymbol{\beta} = [\mu \quad \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_{m-1}]'$$

Die Designmatrix sieht dann wie folgt aus:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & -1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & -1 & \dots & -1 \end{bmatrix}$$

CODIERUNGSSTANDARDS BEIM ALM

- *Abweichungskontraste*: Abweichungen jeweils einer Stufe vom Mittel der anderen werden betrachtet; die letzte Stufe entfällt, sie stellt sozusagen die Vergleichsstufe dar.
- *Differenzkontraste*: eine Stufe wird jeweils mit dem Mittel aller vorhergehenden Stufen verglichen.
- *Helmert-Kontraste*: eine Stufe wird jeweils mit dem Mittel aller nachfolgenden Stufen verglichen.
- *Einfache Kontraste*: jede Stufe wird mit einer Referenzstufe verglichen.
- *Wiederholungskontraste*: jede Stufe außer der ersten wird mit der vorhergehenden verglichen.
- *Polynomiale Kontraste*: es werden Koeffizienten orthogonaler Polynome verwendet.

Parameterschätzungen, gefittete Werte und Modelltest

Im allgemeinen interessieren die Werte der geschätzten Parameter nicht, sondern es geht nur um die Frage nach der Modellgeltung.

Will man aber spezielle Hypothesen testen (z.B. ob die Kontrollgruppe sich von den restlichen Gruppen unterscheidet), so kann dies durch geeignete Wahl der Designmatrix und der entsprechenden Parameter geschehen.

Auf jeden Fall erhält man als gefittete Werte unabhängig von der gewählten Parametrisierung:

$$\hat{y}_{gi} = \bar{y}_g.$$

Für den Test auf Modellgeltung wird wieder eine Quadratsummenzerlegung durchgeführt, d.h. die totale Quadratsumme in die Quadratsumme der Regression und die Quadratsumme der Residuen zerlegt:

$$\sum_g \sum_i (y_{gi} - \bar{y}_{..})^2 = \sum_g \sum_i (\hat{y}_{gi} - \bar{y}_{..})^2 + \sum_g \sum_i (y_{gi} - \hat{y}_{gi})^2$$

Die Freiheitsgrade werden in analoger Weise zerlegt:

$$\begin{matrix} df_{Total} & = & df_{Regression} & + & df_{Residuum} \\ n - 1 & & m - 1 & & n - m - 2 \end{matrix}$$

3.1.2 Zweifaktorielle Varianzanalyse

Bei einem mehrfaktoriellen Versuchsplan mit unabhängigen Gruppen codiert man die Haupteffekte jedes Faktors mit einer Codiervariablen weniger als der Faktor Stufen hat.

Interaktionen zwischen den Faktoren werden durch das elementweise Produkt der Codiervariablen der beteiligten Faktoren codiert.

Als erste Spalte der Designmatrix wird schließlich noch eine Einerspalte verwendet.

Das weitere Vorgehen ist das gleiche wie bei der multiplen Regression, mit dem Unterschied, dass neben dem globalen Modelltest vor allem auch Tests für die einzelnen Haupteffekte und Wechselwirkungen durchgeführt werden.

Ist das Design orthogonal (dies ist z.B. bei gleichem n in den Gruppen der Fall), so sind die einzelnen Tests unabhängig.

Ist es nicht orthogonal, so hängen die Ergebnisse der Tests von der Reihenfolge ab, in der die einzelnen Effekte berücksichtigt werden.

BEISPIEL

Wir betrachten ein dreifaktorielles Design mit 12 unabhängigen Gruppen und 6 Vpn pro Zelle.

Die Faktoren seien

- Faktor A mit 2 Stufen,
- Faktor B mit 3 Stufen,
- Faktor C mit 2 Stufen.

Wir verwenden einfache Kontraste. Die Designmatrix hat 72 Zeilen und 12 Spalten.

Effekt →	A	B ₁	B ₂	C	AB ₁	AB ₂	AC	B ₁ C	B ₂ C	AB ₁ C	AB ₂ C	
X ₀	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	
Gruppe:												
A ₁ B ₁ C ₁	1	1	1	0	1	1	0	1	1	0	1	0
A ₁ B ₁ C ₂	1	1	1	0	-1	1	0	-1	-1	0	-1	0
A ₁ B ₂ C ₁	1	1	0	1	1	0	1	1	0	1	0	1
A ₁ B ₂ C ₂	1	1	0	1	-1	0	1	-1	0	-1	0	-1
A ₁ B ₃ C ₁	1	1	-1	-1	1	-1	-1	1	-1	-1	-1	-1
A ₁ B ₃ C ₂	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1
A ₂ B ₁ C ₁	1	-1	1	0	1	-1	0	-1	1	0	-1	0
A ₂ B ₁ C ₂	1	-1	1	0	-1	-1	0	1	-1	0	1	0
A ₂ B ₂ C ₁	1	-1	0	1	1	0	-1	-1	0	1	0	-1
A ₂ B ₂ C ₂	1	-1	0	1	-1	0	-1	1	0	-1	0	1
A ₂ B ₃ C ₁	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1
A ₂ B ₃ C ₂	1	-1	-1	-1	-1	1	1	1	1	1	-1	-1

Jede Zeile wiederholt sich 6mal, nämlich für alle Vpn einer Gruppe.

3.1.3 Modell mit metrischen und Faktorvariablen

Mit Hilfe des ALM ist es auch kein Problem, ein lineares Modell anzupassen, das sowohl metrische Variablen als auch Faktorvariablen enthält.

Als Beispiel könnte man modellieren, wie Y von X abhängt und als weitere Prädiktorvariable das Geschlecht betrachten.

Die Matrix \mathbf{X} besteht dann aus 4 Spalten:

- Eine Spalte für μ .
- Eine Spalte für den Haupteffekt Geschlecht.
- Eine Spalte für den Haupteffekt der Variable X .
- Eine Spalte für die Interaktion X mal Geschlecht (für den Test auf Parallelität der Regressionsfunktionen).

Das weitere Vorgehen ist dann wie bisher.

3.1.4 Kovarianzanalyse

Im Unterschied zu dem zuletzt beschriebenen Modell verwendet man die Kovarianzanalyse oft nur dazu, die Fehlervarianz mit Hilfe einer metrischen Variable zu reduzieren.

Ähnlich wie bei einer Suppressorvariable wird damit der Varianzanteil, der für die Effekte der Faktorvariablen irrelevant ist, unterdrückt.

In einem ersten Schritt wird eine lineare Regression der Responsevariablen (Y) auf die Kovariablen (X) gerechnet und die Residuen ermittelt.

Anschließend wird eine Varianzanalyse mit den Residuen gerechnet.

EINFACHE KOVARIANZANALYSE

Vorgehen nach Bortz (1999, 5. Aufl.):

1. Rechne eine Regressionsanalyse von Y auf X ohne Berücksichtigung der Faktorstufen; die Quadratsumme und die Freiheitsgrade der Residuen dieser Regression bilden die Quadratsumme und die Freiheitsgrade Total.
2. Rechne eine Regressionsanalyse von Y auf X innerhalb der Zellen, d.h. mit den gepoolten Abweichungen der Einzelwerte von den Gruppenmittelwerten; die Quadratsumme der Residuen dieser Regression ergibt die Fehlerquadratsumme; deren Freiheitsgrade sind $n - m - p - 1$.
 - n Anzahl der Vpn insgesamt
 - m Anzahl der Gruppen
 - p Anzahl der Kovariablen
3. Die Quadratsumme und Freiheitsgrade des Faktors ergibt sich aus den bereits berechneten Größen.
4. Das weitere Vorgehen entspricht dem üblichen Schema.

3.2 Das Verallgemeinerte Lineare Modell

Wie wir gesehen haben, lassen sich verschiedene statistische Verfahren wie die lineare Regression, die Varianzanalyse und die Kovarianzanalyse als Spezialfälle des Allgemeinen Linearen Modells darstellen.

Dieses ist wiederum ein Spezialfall des Verallgemeinerten oder Generalisierten Linearen Modells (GLM)⁴.

⁴engl.: Generalized Linear Model (GLZ)

Bei diesem Modell werden einige der Voraussetzungen des Allgemeinen Linearen Modells fallen gelassen:

- Die Responsevariable braucht nicht normalverteilt und nicht stetig zu sein. Es wird lediglich angenommen, dass ihre Verteilung aus der Familie der Exponentialverteilungen stammt. Beispielsweise könnte sie eine dichotome Variable sein oder eine Zählvariable für das Auftreten eines seltenen Ereignisses.
- Es muss auch nicht Varianzhomogenität (Homoskedastizität der Residuen) vorausgesetzt werden.
- Die Beziehung zwischen der Responsevariablen und den Prädiktorvariablen darf auch nichtlinear sein.

DIE EXPONENTIALFAMILIE

Eine Verteilung gehört zur Familie der Exponentialverteilungen, wenn ihre Wahrscheinlichkeits- bzw. Dichtefunktion die folgende Struktur aufweist:

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

mit irgendwelchen speziellen Funktionen $a(\cdot)$, $b(\cdot)$ und $c(\cdot)$. Der Parameter θ , der sogenannte „kanonische Parameter“, ist ein Lageparameter und ϕ , der sog. „Skalenparameter“, ist ein Dispersionsparameter.

Einige wichtige Vertreter aus der Exponentialfamilie:

- Normalverteilung
- Poissonverteilung
- Binomialverteilung
- Gammaverteilung
- Inverse Normalverteilung

Die Wahrscheinlichkeits- bzw. Dichtefunktionen dieser Verteilungen lassen sich in der oben angegebenen allgemeinen Form ausdrücken.

KOMPONENTEN DES GLM

Das GLM besteht aus drei Komponenten:

1. Zufällige Komponente
2. Systematische Komponente mit dem linearen Prädiktor
3. Linkfunktion, die eine Verbindung zwischen der systematischen Komponente und dem Erwartungswert der zufälligen Komponente herstellt.

Im Detail handelt es sich beim GLM um ein Regressionsmodell der folgenden Art:

- Es enthält als *Zufallskomponente* die Responsevariable $\mathbf{Y} = [Y_1, \dots, Y_n]'$. Die Y_i sind unabhängig und besitzen eine Verteilung aus der Familie der Exponentialverteilungen. Der Erwartungswert von \mathbf{Y} ist $E(\mathbf{Y}) = \boldsymbol{\mu} = [\mu_1, \dots, \mu_n]'$ und für die Varianz wird angenommen:

$$Var(\mathbf{Y}) = \sigma^2 V(\boldsymbol{\mu})$$

- Es wird ein linearer Prädiktor definiert (*systematische Komponente*):

$$\boldsymbol{\eta} = \mathbf{X}'\boldsymbol{\beta}$$

- $\boldsymbol{\eta}$ linearer Prädiktor
- \mathbf{X} Vektor der Prädiktorvariablen
- $\boldsymbol{\beta}$ unbekannter Parametervektor

Alle GLM-Modelle sind lineare Modelle, d.h. die Parameter $\boldsymbol{\beta}$ gehen in die Gleichung linear ein.

- Der lineare Prädiktor $\boldsymbol{\eta}$ wird nun über eine *Responsefunktion* $h(\cdot)$ mit dem Erwartungswert $\boldsymbol{\mu}$ der Responsevariablen \mathbf{Y} verknüpft:

$$\boldsymbol{\mu} = h(\boldsymbol{\eta})$$

Die Umkehrfunktion von h ist die *Linkfunktion* $g(\cdot)$:

$$h^{-1}(\boldsymbol{\mu}) = g(\boldsymbol{\mu}) = \boldsymbol{\eta}$$

$$\boldsymbol{\mu} = g^{-1}(\boldsymbol{\eta})$$

Die Linkfunktion beschreibt die (normalerweise nichtlineare) Beziehung zwischen der Responsevariable und der linearen rechten Seite.

- Die Funktion $V(\cdot)$ ist die *Varianzfunktion*. Diese beschreibt den Zusammenhang zwischen dem Erwartungswert und der Varianz der Responsevariable.

LINKFUNKTIONEN

- Für jede Verteilung gibt es mehr als eine Linkfunktion.
- Verschiedene Linkfunktionen ergeben leicht unterschiedliche Parameterschätzungen.
- Unter den verschiedenen Linkfunktionen ist die sog. „kanonische“ oder natürliche Linkfunktion hervorzuheben. Bei dieser ist der lineare Prädiktor $\boldsymbol{\eta}$ gleich dem kanonischen Parameter $\boldsymbol{\theta}$.

SPEZIALFÄLLE DES GLM

Verteilung	Normalverteilung	Binomialverteilung	Poissonverteilung
Variablentyp	stetig	binär	Zählvariable
Notation	$N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$	$B(m, \pi)/m$	$P(\mu)$
Wertebereich der Response	$(-\infty, +\infty)$	$\frac{0(1)m}{m}$	$0(1)\infty$
Dispersionsparameter ϕ	$\boldsymbol{\sigma}^2$	$1/m$	1
$b(\boldsymbol{\theta})$	$\boldsymbol{\theta}^2/2$	$\log(1 + e^\theta)$	$\exp(\boldsymbol{\theta})$
$c(y, \phi)$	$-\frac{1}{2}(\frac{y^2}{\phi} + \log(2\pi\phi))$	$\log\binom{m}{my}$	$-\log(y!)$
$\boldsymbol{\mu}(\boldsymbol{\theta})$	$\boldsymbol{\theta}$	$e^\theta/(1 + e^\theta)$	$\exp(\boldsymbol{\theta})$
Kanonischer Link	Identität	logit	log
Varianzfunktion:	1	$\mu(1 - \mu)$	μ
$a(\phi)$:	$\boldsymbol{\sigma}^2$	1	1

Weitere Spezialfälle des GLM sind die *Gammaverteilung* (z.B. für die Analyse von Überlebenszeiten) und die *inverse Normalverteilung* (z.B. für die Analyse von stationären unabhängigen Zuwächsen).

GÜTE DER MODELLANPASSUNG

Als Maß für die Güte eines linearen Modells M sind verschiedene Statistiken vorgeschlagen worden. Am bekanntesten ist die sog. *Devianz* (engl.: deviance). Diese ist eine Verallgemeinerung der Fehlerquadratsumme (RSS) aus dem linearen Regressionsmodell:

$$dev(M) = 2[\log L(\hat{\boldsymbol{\beta}}_{max} | \mathbf{y}) - \log L(\hat{\boldsymbol{\beta}} | \mathbf{y})]$$

$\log L$ Log-Likelihood-Funktion

$\hat{\boldsymbol{\beta}}$ ML-Schätzung des Modellparameters $\boldsymbol{\beta}$ des Modells M

$\hat{\boldsymbol{\beta}}_{max}$ ML-Schätzung von $\boldsymbol{\beta}$ des saturierten Modells

Damit wird also der Wert der log-Likelihood-Funktion des zutestenden Modells mit dem Wert der log-Likelihood-Funktion des perfekten Modells verglichen.

3.2.1 Klassische lineare Modelle

Klassische lineare Modelle umfassen die lineare Regression, die Varianz- und Kovarianzanalyse. Modelle dieser Art werden auch mit „OLS-Regressionsmodelle“⁵ bezeichnet.

Ein OLS-Modell ist linear in den Parametern.

Die Residuen sind *normalverteilt* und deren Varianz⁶ ist unabhängig von den \hat{Y} -Werten und konstant (Homoskedastizität).

Der kanonische Link ist die Identität:

$$E(Y|\mathbf{x}) = \sum \beta_k X_k$$

D.h. Responsewerte und vorhergesagte Werte (Y und \hat{Y}) liegen auf derselben Skala.

3.2.2 Logistische Regression

Bei der logistischen Regression ist die Responsevariable dichotom, nicht aber der vorhergesagte Wert.

Die Wahrscheinlichkeitsfunktion für eine dichotome Variable ist die *Bernoulliverteilung*:

$$f(Y = y|\pi) = \pi^y(1 - \pi)^{1-y}$$

Deren Varianz ist $\pi(1 - \pi)$, d.h. sie hängt von π ab und ist nicht konstant.

Wir sagen nicht den Wert der Responsevariable (0 oder 1) vorher, sondern die Wahrscheinlichkeit π für den Wert 1.

Die Wahrscheinlichkeitsfunktion für Anteilswerte ist die *Binomialverteilung*:

$$f(Y = y|\pi, n) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

Für diese Wahrscheinlichkeit wird die logistische Funktion gewählt:

$$\hat{\pi} = \frac{e^{\sum \beta_k X_k}}{1 + e^{\sum \beta_k X_k}}$$

Der kanonische Link ist die Logit-Funktion:

$$\log\left(\frac{\hat{\pi}}{1 - \hat{\pi}}\right) = \sum \beta_k X_k$$

Eine Alternative zur Logitfunktion ist die Probitfunktion:

$$\Phi^{-1}(\hat{\pi}) = \sum \beta_k X_k$$

Sie ist dann zu verwenden, wenn die zugrunde liegende Variable eigentlich als normalverteilt angenommen werden soll.

3.2.3 Poisson-Regression

Bei der *Poissonregression* ist die Responsevariable eine Zählvariable.

Die *Poissonverteilung* ist eine Wahrscheinlichkeitsverteilung für eine diskrete Zufallsvariable Y (mit $Y \geq 0$).

⁵OLS=ordinary least squares

⁶D.h. auch die Varianz der bedingten Verteilung von Y gegeben \hat{Y}

Die Wahrscheinlichkeitsfunktion ist wie folgt definiert:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}$$

Erwartungswert und Varianz der Verteilung sind:

$$E(Y) = \lambda \quad \text{Var}(Y) = \lambda$$

d.h. Erwartungswert und Varianz sind gleich.

Bei der Poissonregression sagen wir nicht den Wert der Responsevariable selbst vorher, sondern deren Logarithmus, d.h. der kanonische Link ist die Logarithmus-Funktion:

$$\log(y_i) = \sum \beta_k X_k$$

Das bekannteste Beispiel einer Poissonregression ist das *loglineare Modell*. Ein solches Modell analysiert den Zusammenhang zwischen Variablen in einer mehrdimensionalen Kontingenztabelle. Die Logarithmen der erwarteten Zellenhäufigkeiten werden als Linearkombinationen von Effekten der unabhängigen Variablen modelliert. Neben Haupteffekten kann ein solches Modell auch Wechselwirkungseffekte enthalten.

Ein häufiges Problem bei der Poissonregression tritt dann auf, wenn die Varianzen größer oder kleiner sind als vorhergesagt (*overdispersion* bzw. *underdispersion*). Dann muss man in das Modell einen zusätzlichen Parameter aufnehmen.

3.3 Gemischte lineare Modelle

Eine weitere Verallgemeinerung der klassischen linearen Modelle stellen die gemischten linearen Modelle dar. Diese gehen von folgenden Annahmen aus:

- Die Daten sind normalverteilt.
- Die Mittelwerte bzw. Erwartungswerte der Daten sind lineare Funktionen einer gewissen Menge von festen Parametern.
- Den Varianzen und Kovarianzen der Daten liegen eine Menge von Parametern zugrunde (Kovarianzparameter).

Szenarien, in denen gemischte Modelle eingesetzt werden, sind:

- Messwiederholungen, Längsschnittstudien
- Geclusterte Daten (z.B. Schüler innerhalb Klassen und Klassen innerhalb Schulen)

Gemischte lineare Modelle enthalten verschiedene Arten von Parametern:

- Parameter für fixe Effekte
- Parameter für Random-Effekte (Varianzkomponenten, Kovarianzparameter)
- Modelle, die beide Arten von Parameters enthalten, heißen gemischte Modelle.
- Es gibt verschiedene Methoden, um die jeweiligen Parameter zu schätzen (z.B. REML=restricted maximum likelihood).

Ein gemischtes lineares Modell sieht wie folgt aus:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\zeta} + \boldsymbol{\varepsilon}$$

X bekannte Designmatrix für die fixen Effekte

$\boldsymbol{\beta}$ unbekannter Vektor der fixen Effekt-Parameter

Z bekannte Designmatrix für die Random-Effekte

$\boldsymbol{\zeta}$ unbekannter Vektor der Random Effekt-Parameter

$\boldsymbol{\varepsilon}$ unbekannter Zufallsvektor, dessen Elemente nicht unabhängig und homogen sein müssen

Beispiele für Fehler-Kovarianzmatrizen⁷:

- Diagonalmatrix

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

- Compound Symmetry

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

- AR(1)-Prozess

$$\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

- Unstrukturierte Kovarianzmatrix

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

3.4 Generalized Estimation Equations (GEE-Modelle)

Die GEE-Regressionsmodelle⁸ stellen eine weitere Verallgemeinerung der bisher betrachteten Modelle, nämlich des verallgemeinerten linearen Modells und des gemischten linearen Modells dar. Sie unterstützen die Analyse von Daten, die in

- Längsschnittstudien,
- bei Versuchsplänen mit genesteten Faktoren,
- bei Versuchsplänen mit geclusterten Gruppen,
- bei Versuchsplänen mit Messwiederholungen

anfallen. Bei Versuchsplänen dieser Art liegen Daten vor, die innerhalb einer Person bzw. innerhalb eines Clusters korreliert sind, also gegen die Voraussetzung der Unabhängigkeit der Fehlervariable verstoßen, eine Voraussetzung, die beim GLM gelten muss.

In diesem Punkt stellen die GEE-Regressionsmodelle eine Erweiterung des GLM in Richtung der gemischten linearen Modelle dar, ansonsten haben sie mit ihnen viele Gemeinsamkeiten:

- Die abhängige Variable muss nicht stetig sein, sondern kann z.B. eine dichotome oder eine Zählvariable sein.
- Als Verteilung der abhängigen Variable kommen außer der Normalverteilung die Binomialverteilung, die Poissonverteilung u.a. in Frage.
- Wie beim GLM wird ein linearer Prädiktor formuliert und die Beziehung zu der abhängigen Variable mit Hilfe einer Link-Transformationsfunktion hergestellt.

⁷Außer den aufgeführten Typen von Kovarianzmatrizen gibt es noch eine ganze Reihe weiterer Standardtypen.

⁸Lit.: Liang & Zeger (1986); Zeger & Liang (1986); Diggle, Liang & Zeger (1994)

Für die Korrelationen der Fehler kann der Benutzer wie bei den gemischten linearen Modellen verschiedene Strukturen annehmen wie z.B. Compound Symmetry oder AR(1) oder unstrukturiert.

4 Multivariate Varianzanalyse (MANOVA)

Zweck

Wie bei der univariaten Varianzanalyse (ANOVA) werden bei der multivariate Varianzanalyse (MANOVA) Mittelwertsunterschiede zwischen Populationen betrachtet.

Hierbei gehen aber gleichzeitig mehrere abhängige Variablen in die Analyse ein, d.h. es werden nicht einzelne Mittelwerte, sondern Mittelwertsvektoren verglichen.

Dadurch wird einerseits das Problem der Alpha-Inflation vermieden, andererseits normalerweise auch die Teststärke erhöht.

Modell

Wir betrachten ein varianzanalytisches Design mit m unabhängigen Stichproben und q abhängigen Variablen, d.h. wir betrachten q -dimensionale Vektoren \mathbf{Y}_{gi} .

Wir wollen testen, ob zwischen den m Populationen, aus denen die Stichproben stammen, ein \mathbf{Y} -Mittelwertsunterschied besteht.

Für \mathbf{Y}_{gi} , den Messwertvektor der i -ten Beobachtung in der g -ten Stichprobe, nehmen wir das folgende Modell an:

$$\mathbf{Y}_{gi} = \boldsymbol{\mu}_g + \boldsymbol{\varepsilon}_{gi}$$

$\boldsymbol{\mu}_g$ Erwartungswertvektor von \mathbf{Y}_{gi} in der g -ten Population

$\boldsymbol{\varepsilon}_{gi}$ multivariat normalverteilte Zufallsvariablen mit für alle Gruppen identischer Kovarianzmatrix $\boldsymbol{\Sigma}$

Anstelle der Populationsparameter $\boldsymbol{\mu}_g$ werden im *Strukturmodell* die sog. Effekte verwendet, z.B. im einfaktoriellen Fall heißt das Modell:

$$\mathbf{Y}_{gi} = \boldsymbol{\mu} + \boldsymbol{\alpha}_g + \boldsymbol{\varepsilon}_{gi}$$

$\boldsymbol{\mu}$ Erwartungswertvektor von \mathbf{Y} in der Gesamtpopulation

$\boldsymbol{\alpha}_g$ Vektor der Effekte der g -ten Population bzw. der g -ten Behandlung

4.1 Hotelling's T^2

Hotelling's T^2 für zwei unabhängige Stichproben

Hotelling's T^2 dient dem Mittelwertsvergleich von zwei Stichproben, d.h. testet die Hypothese:

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

Die Formel für die Testgröße entspricht jener des quadrierten univariaten t-Tests für unabhängige Stichproben:

$$t^2 = \frac{(\bar{y}_1 - \bar{y}_2)^2}{s^2(1/n_1 + 1/n_2)} = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' (\mathbf{s}^2)^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$

Also:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = \frac{n_1 n_2}{n_1 + n_2} D^2$$

- n_g Stichprobenumfang der Gruppe g
- $\bar{\mathbf{y}}_g$ Mittelwertsvektor der Gruppe g
- \mathbf{S} gepoolte Kovarianzmatrix innerhalb der Gruppen
- D^2 Mahalanobis-Distanz zwischen den Mittelwertsvektoren (s.u.)

Die Testgröße T^2 kann auch auf folgende Weise ausgedrückt werden:

$$T^2 = \frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{W}^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$

\mathbf{W} ist die gepoolte SSCP-Matrix innerhalb der Gruppen und wird wie folgt berechnet:

$$\mathbf{W} = \mathbf{W}^{(1)} + \mathbf{W}^{(2)}$$

Dabei ist $\mathbf{W}^{(1)}$ bzw. $\mathbf{W}^{(2)}$ die SSCP-Matrix innerhalb Gruppe 1 bzw. 2:

$$\mathbf{W}^{(g)} = (\mathbf{Y}_d^{(g)})' (\mathbf{Y}_d^{(g)}) \quad (g = 1, 2)$$

$\mathbf{Y}_d^{(g)}$ ist die Matrix der Abweichungen der Einzelwerte von den jeweiligen Gruppenmittelwerten:

$$\mathbf{Y}_d^{(g)} = \mathbf{Y}^{(g)} - \mathbf{1} \bar{\mathbf{y}}'^{(g)}$$

Für die Signifikanzprüfung wird T^2 in eine F-Statistik umgerechnet:

$$F = \frac{(n_1 + n_2 - 2) - (q - 1)}{q(n_1 + n_2 - 2)} T^2 = \frac{n_1 + n_2 - q - 1}{q} \frac{T^2}{n_1 + n_2 - 2}$$

Unter H_0 ist diese Testgröße F-verteilt mit $df_1 = q$ und $df_2 = n_1 + n_2 - q - 1$ Freiheitsgraden.

Mahalanobis Distanz

Die Mahalanobis-Distanz D^2 gibt den Abstand zwischen zwei Punkten \mathbf{x}_1 und \mathbf{x}_2 (Vektoren von Einzelwerten, Mittelwerten, ...) einer mehrdimensionalen Normalverteilung an. Sie stellt eine Verallgemeinerung des univariaten Falles dar, bei dem als Distanzmaß der z-Wert bzw. t-Wert verwendet wird, und ist folgendermaßen definiert:

$$D^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)$$

bzw. für die Stichprobe

$$D^2(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{x}_1 - \mathbf{x}_2)' \mathbf{S}^{-1} (\mathbf{x}_1 - \mathbf{x}_2)$$

- $\boldsymbol{\Sigma}$ Kovarianzmatrix der Normalverteilung
- \mathbf{S} empirische Kovarianzmatrix

Die Mahalanobis-Distanz spielt in der multivariaten Statistik eine wichtige Rolle.

Im Gegensatz zur Mahalanobis-Distanz geht in die Euklidische Distanz (die bekanntlich einen Spezialfall der Minkowski-Metrik darstellt) nicht die Kovarianzmatrix ein, d.h. sie ist unabhängig von der Verteilung der Datenpunkte:

$$d_2(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)' (\mathbf{x}_1 - \mathbf{x}_2)}$$

Geometrisch gesehen bilden die Punkte gleicher Mahalanobis-Distanz von einem Zentrum im 2-dimensionalen Raum eine Ellipse, die Punkte gleicher Euklidischer Distanz hingegen einen Kreis.

Hotelling's T^2 für eine Stichprobe

Bei Hotelling's T^2 für eine Stichprobe werden die Stichprobenmittelwerte gegen den Nullvektor bzw. gegen einen festen Vektor $\boldsymbol{\mu}_0$ getestet. Die Nullhypothese lautet:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

Als Testgröße verwendet man:

$$T^2 = n(n-1)(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}_0) = n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)' \mathbf{W}^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$$

Für die F-Statistik ergibt sich:

$$F = \frac{n-q+1}{(q-1)(n-1)} T^2$$

Unter H_0 ist diese Testgröße F-verteilt mit $df_1 = q-1$ und $df_2 = n-q+1$ Freiheitsgraden.

Hotelling's T^2 bei Messwiederholung

Hotelling's T^2 für eine Stichprobe ist ein Verfahren, das die sich auch bei einem Messwiederholungsdesign anwenden lässt.

Dazu bildet man für jede abhängige Variable die Differenzen von der ersten zur zweiten Messung und analysiert die Vektoren der Differenzen anstatt der Originalwertvektorenpaare.

4.2 MANOVA einfaktoriell

Wir betrachten nun ein einfaktorielles Design mit m unabhängigen Gruppen und q abhängigen Variablen und gehen von dem eingangs dargestellten Modell und den dort angenommenen Voraussetzungen aus.

Die Nullhypothese lautet:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_m = \boldsymbol{\mu}$$

Für die Testgrößen werden verschiedene SSCP-Matrizen benötigt¹:

- T** SSCP-Matrix **T**otal
- B** (bzw. **H**) SSCP-Matrix **B**etween (**H**ypothesis)
- W** (bzw. **E**) SSCP-Matrix **W**ithin (**E**rror)

$$t_{ij} = \sum_{k=1}^m \sum_{l=1}^{n_k} [y_{li}^{(k)} - \bar{y}_{.i}^{(\cdot)}] [y_{lj}^{(k)} - \bar{y}_{.j}^{(\cdot)}]$$

$$b_{ij} = h_{ij} = \sum_{k=1}^m n_k [\bar{y}_{.i}^{(k)} - \bar{y}_{.i}^{(\cdot)}] [\bar{y}_{.j}^{(k)} - \bar{y}_{.j}^{(\cdot)}]$$

$$w_{ij} = e_{ij} = \sum_{k=1}^m \sum_{l=1}^{n_k} [y_{li}^{(k)} - \bar{y}_{.i}^{(k)}] [y_{lj}^{(k)} - \bar{y}_{.j}^{(k)}]$$

Übrigens sei $n = \sum_g n_g$.

T, **B**, **W** in Matrixschreibweise:

$$\mathbf{T} = (\mathbf{Y}_d)' (\mathbf{Y}_d)$$

$$\mathbf{B} = (\bar{\mathbf{Y}}_d)' (\bar{\mathbf{Y}}_d)$$

$$\mathbf{W} = (\mathbf{Y}_d^{(k)})' (\mathbf{Y}_d^{(k)})$$

¹Übrigens wird die Matrix **B** in manchen Lehrbüchern auch mit **H** (für **H**ypothesis) bezeichnet und die Matrix **W** mit **E** (für **E**rror)

- \mathbf{Y}_d $N \times Q$ -Matrix der Abweichungen der Einzelwerte von den Gesamtmittelwerten
 $\bar{\mathbf{Y}}_d$ $N \times Q$ -Matrix der Abweichungen der Gruppenmittelwerte von den Gesamtmittelwerten
 (für alle Zeilen einer Gruppe identisch)
 $\mathbf{Y}_d^{(k)}$ $N \times Q$ -Matrix der Abweichungen der Einzelwerte von den Gruppenmittelwerten

Die am häufigsten verwendete Testgröße ist *Wilks' Lambda*:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{1}{|\mathbf{B}\mathbf{W}^{-1} + \mathbf{I}|} \quad 0 \leq \Lambda \leq 1$$

Diese Größe hängt auch mit den Eigenwerten λ_i von $\mathbf{B}\mathbf{W}^{-1}$ zusammen:

$$\Lambda = \prod \frac{1}{1 + \lambda_i}$$

Für die Bewertung der Testgröße gibt es zwei Approximationen:

- Bartlett's χ^2 (gut bei mittlerem bis großem Stichprobenumfang):

$$\chi^2 = -[(n-1) - (q+m)/2] \ln \Lambda \quad df = q(m-1)$$

- Rao's F (gut vor allem bei kleinem Stichprobenumfang).

Neben Wilks' Lambda gibt es noch weitere Alternativen:

- *Roy's Kriterium des größten Eigenwerts*: sei λ_{max} der größte Eigenwert von $\mathbf{B}\mathbf{W}^{-1}$, dann beträgt die Prüfgröße:

$$\frac{\lambda_{max}}{1 + \lambda_{max}}$$

- *Hotelling-Lawley Spur*: Summe der Eigenwerte von $\mathbf{B}\mathbf{W}^{-1}$.
- *Pillai-Bartlett Spur*: Summe der Eigenwerte von $\mathbf{B}\mathbf{T}^{-1}$.

Anschluss tests an die MANOVA

Nach Ablehnung der Nullhypothese beim globalen multivariaten Test (Λ) sind speziellere Anschluss tests durchführbar:

- T^2 für Gruppenpaare
- Nach signifikantem T^2 : t -Tests

Um die Alphainflation zu vermeiden, ist dabei jeweils die Bonferroni-Korrektur vorzunehmen.

Daneben können auch *Konfidenzintervalle* berechnet werden:

- Tukey's simultane Konfidenzintervalle
- Beliebig viele simultane Konfidenzintervalle nach Roy-Bose (Verallgemeinerung von Scheffé, sehr konservativ)

Geplante Vergleiche

Auch bei der MANOVA ist es möglich, mit Hilfe linearer Kontraste spezielle a priori Hypothesen zu prüfen.

Hier sind es allerdings nicht einzelne Mittelwerte, die in die Kontraste eingehen, sondern Mittelwertsvektoren.

Für die Kontrastkoeffizienten gelten dieselben Voraussetzungen und Eigenschaften wie im univariaten Fall ($\sum c_j = 0$; Kontraste orthogonal oder nicht?)

4 Multivariate Varianzanalyse (MANOVA)

Test eines einzelnen Kontrastes $\boldsymbol{\psi}$:

$$H_0: \boldsymbol{\psi} = \mathbf{0}$$

$$\hat{\boldsymbol{\psi}} = \sum c_j \bar{\mathbf{y}}_j$$

$$T^2 = \frac{1}{\sum c_j^2 / n_j} \hat{\boldsymbol{\psi}}' \mathbf{S}^{-1} \hat{\boldsymbol{\psi}}$$

$$F = \frac{n-m-q+1}{(n-m)q} T^2 \quad \text{mit} \quad df_1 = q \quad df_2 = n-m-q+1$$

4.3 MANOVA mehrfaktoriell mit unabhängigen Gruppen

Das Vorgehen beim Test der verschiedenen Effekte bei einer mehrfaktoriellen MANOVA (Haupteffekte, Wechselwirkungen, Test spezieller Hypothesen) ist völlig analog zum univariaten Fall, nur dass anstatt der Quadratsummen SSCP-Matrizen berechnet werden müssen.

Beispielsweise werden für den Test des Haupteffekts eines Faktors A die Differenzvektoren der Mittelwerte von A zu den Gesamtmittelwerten gebildet und daraus \mathbf{B}_A , die SSCP-Matrix Zwischen für den Faktor A , berechnet. Dabei ist zu berücksichtigen, dass alle Summanden innerhalb einer Stufe von A identisch sind. Also multipliziert man jeden Summanden mit der Anzahl der Beobachtungen, die in die Berechnung der jeweiligen Mittelwerte von A eingingen.

Die SSCP-Matrix für eine Wechselwirkung $A \times B$ ist analog zum univariaten Fall zu berechnen:

$$\mathbf{B}_{AB} = \mathbf{B}_{zwAB} - \mathbf{B}_A - \mathbf{B}_B$$

- \mathbf{B}_A SSCP-Matrix Between für den Faktor A
- \mathbf{B}_B SSCP-Matrix Between für den Faktor B
- \mathbf{B}_{AB} SSCP-Matrix für die Wechselwirkung AB
- \mathbf{B}_{zwAB} SSCP-Matrix Between für alle AB -Gruppen

Testgrößen

Jeder einzelne Effekt wird über ein entsprechendes Wilks' Lambda getestet.

Als Fehlerterm wird \mathbf{W} , die SSCP-Matrix innerhalb der Gruppen, verwendet.

4.4 MANOVA mit Messwiederholungen

Viele Computerprogramme zur ANOVA/MANOVA mit Messwiederholungen transformieren zunächst die abhängigen Variablen für die verschiedenen Messzeitpunkte mit Hilfe einer orthonormalen Transformationsmatrix.

Die erste dieser transformierten Variablen ist die normierte Summe der Originalvariable(n) zu allen Messzeitpunkten. Diese wird für das „Between-Design“ (Zwischen den Subjekten) verwendet.

Die weiteren transformierten Variablen bilden die Grundlage für die Analyse des „Within-Designs“ (Innerhalb der Subjekte, d.h. Zwischen den Messzeitpunkten).

4.5 Überprüfung der Voraussetzungen für die MANOVA

Box M

Die Gleichheit der Kovarianzmatrizen in den einzelnen Populationen ist eine der Voraussetzungen der MANOVA.

Diese Voraussetzung wird mit Box' M getestet²:

$$M = (n-q)\ln|\mathbf{S}| - \sum (n_j-1)\ln|\mathbf{S}_j|$$

Für diese Statistik gibt es Umrechnungsformeln in eine approximativ F-verteilte Testgröße.

Sphärizitäts Test

Eine Folgerung aus dem klassischen Modell einer ANOVA mit Messwiederholungen ist die Voraussetzung der sog. „compound symmetry“, d.h. für die Kovarianzmatrix der Messungen muss gelten:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

Diese Voraussetzung ist gleichbedeutend mit der Voraussetzung der „Sphärizität“, d.h. dass für die Kovarianzmatrix der *transformierten* Variablen gilt:

$$\mathbf{\Sigma} = \sigma^2\mathbf{I}$$

Ein Test, der diese Voraussetzung überprüft, ist Mauchley's Sphericity Test.

4.6 MANOVA als ALM

Das Modell der MANOVA kann auch mit Hilfe des Allgemeinen Linearen Modells in der mehrdimensionalen Version formuliert werden³:

$$\mathbf{Y}_{n \times q} = \mathbf{X}_{n \times p} \mathbf{\Psi}_{p \times q} + \mathbf{\mathcal{E}}_{n \times q}$$

- Y** bekannte Beobachtungsmatrix (abhängige Variablen)
- X** bekannte Designmatrix
- Ψ** unbekannte Parametermatrix
- ℰ** unbekannte Fehlermatrix

Man erkennt hier, dass es sich um ein mehrdimensionales multiples Regressionsmodell (MMR) handelt. Dabei entspricht die Parametermatrix **Ψ** der Matrix **B** bei der MMR.

Parametrisierung

Wie im univariaten Fall kann die Designmatrix **X** verschieden definiert sein. Dementsprechend fällt auch die Parametrisierung, d.h. die Parametermatrix **Ψ**, verschieden aus.

²**S** ist die gepoolte Kovarianzmatrix innerhalb der Gruppen und **S_j** die Kovarianzmatrix in der Gruppe *j*.

³Das Symbol **ℰ** soll den Großbuchstaben von **ε** darstellen.

Beispiele:

- Dummycodierung:

$$\mathbf{\Psi} = \begin{bmatrix} \boldsymbol{\mu}'_1 \\ \boldsymbol{\mu}'_2 \\ \vdots \\ \boldsymbol{\mu}'_m \end{bmatrix}$$

- Effektcodierung:

$$\mathbf{\Psi} = \begin{bmatrix} \boldsymbol{\mu}' \\ \boldsymbol{\alpha}'_1 \\ \vdots \\ \boldsymbol{\alpha}'_{m-1} \end{bmatrix}$$

Modelltest: Hypothesen

Die zu testenden Hypothesen beziehen sich auf Unterschiede zwischen den Mittelwertsvektoren.

Die Designmatrix \mathbf{X} enthält normalerweise als erste Spalte (Spalte 0) eine konstante Spalte mit Einsen:

$$\mathbf{x}_0 = \mathbf{1}'$$

Diese Spalte gilt dem Parameter $\boldsymbol{\mu}$. Dieser wird normalerweise nicht getestet, muss jedoch bei den Parameterschätzungen mit berücksichtigt werden.

Für die Berechnung der Testgrößen wird also die um die erste Spalte reduzierte Designmatrix $[\mathbf{x}_1 \dots \mathbf{x}_{m-1}]$ verwendet, die wir mit \mathbf{X}_1 bezeichnen wollen.

Entsprechend lassen wir auch bei der Matrix der Parameterschätzungen die erste Zeile weg und verwenden die reduzierte Matrix $\hat{\boldsymbol{\Psi}}_1 = [\hat{\boldsymbol{\psi}}'_1 \dots \hat{\boldsymbol{\psi}}'_{m-1}]'$.

Für die Testgröße werden dann folgende SSCP-Matrizen berechnet:

$$\begin{aligned} \mathbf{T} &= \mathbf{Y}'_d \mathbf{Y}_d \\ \mathbf{H} &= \hat{\boldsymbol{\Psi}}'_1 (\mathbf{X}'_1 \mathbf{X}_1) \hat{\boldsymbol{\Psi}}_1 \end{aligned}$$

\mathbf{T} „Totalmatrix“

\mathbf{H} „Hypothesenmatrix“ (entspricht der Matrix \mathbf{B}_Ψ)

\mathbf{E} „Errormatrix“ (entspricht der Matrix \mathbf{W})

Dann lautet die Testgröße wiederum:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|}$$

Wie im univariaten Fall lassen sich durch Verwendung einzelner X -Variablen (Spalten der Designmatrix) im Regressionsmodell differenzierte Hypothesen testen (z.B. geplante Vergleiche oder im mehrfaktoriellen Fall Haupteffekte und Wechselwirkungen).

Zweifaktorielle Varianzanalyse

Bei einem mehrfaktoriellen Versuchsplan mit unabhängigen Gruppen codiert man die Haupteffekte jedes Faktors mit einer Codiervariablen weniger als der Faktor Stufen hat.

Interaktionen zwischen den Faktoren werden durch das elementweise Produkt der Codiervariablen der beteiligten Faktoren codiert.

Als erste Spalte der Designmatrix wird schließlich noch eine Einerspalte verwendet.

Das weitere Vorgehen ist das gleiche wie bei der multiplen Regression, mit dem Unterschied, dass neben dem globalen Modelltest vor allem auch Tests für die einzelnen Haupteffekte und Wechselwirkungen durchgeführt werden.

Ist das Design orthogonal (dies ist z.B. bei gleichem n in den Gruppen der Fall), so sind die einzelnen Tests unabhängig.

Ist es nicht orthogonal, so hängen die Ergebnisse der Tests von der Reihenfolge ab, in der die einzelnen Effekte berücksichtigt werden (wie bei der multiplen Regression).

5 Diskriminanzanalyse (DA)

Einordnung

- Ähnlichkeit zur einfaktoriellen MANOVA: es liegt hier wie dort die gleiche Datenstruktur vor, nämlich unabhängige Stichproben aus m Populationen mit q mehrdimensional normalverteilten Variablen mit gleicher Kovarianzmatrix Σ , aber möglicherweise verschiedenen Erwartungswerten μ_g .
 - MANOVA: Test auf Mittelwertsunterschiede zwischen den Gruppen
 - DA: welches Gewicht haben die verwendeten abhängigen Variablen für die Unterscheidung der Gruppen?
- Ähnlichkeit zur multiplen Regression:
 - MR: Schätzen eines Kriteriums mit Hilfe von Prädiktorvariablen
 - DA: Zuordnung von Beobachtungen zu einer Population anhand verschiedener metrischer Merkmale (Klassifikationsproblem)

Zielsetzung

Bei der linearen Diskriminanzanalyse sollen vorgegebene Gruppen¹ anhand einer oder mehrerer Linearkombinationen der Variablen möglichst gut getrennt werden.

D.h. wir suchen Gewichtskoeffizienten, die eine maximale Trennung (Diskriminierung) der vorgegebenen Gruppen erlauben.

Dies bedeutet, dass wir im Variablenraum eine bzw. mehrere Richtungen (d.h. neue Variablen) \mathbf{v}_l suchen, auf denen die Projektionen der Personenvektoren für die einzelnen Gruppen sich möglichst wenig überlappen.

Diese Richtungen werden *Diskriminanzfunktionen* oder *Diskriminanzfaktoren* genannt.

Gibt es mehrere Diskriminanzfunktionen, so wird gefordert, dass sie paarweise orthogonal sind.

Die erste Diskriminanzfunktion wird so bestimmt, dass sie die bestmögliche Trennung erlaubt. Die weiteren Diskriminanzfunktionen sind so zu bilden, dass sie unter den verbleibenden möglichen Diskriminanzfunktionen jeweils die maximale Trennung bringen.

Zielfunktion und Lösung

Die Güte der Trennung hängt von zwei voneinander unabhängigen Bedingungen ab:

1. von den Abständen zwischen den Gruppenmittelwertend.h. von der Streuung zwischen den Gruppen
2. von der Streuung innerhalb der Gruppen

Das Zielkriterium ist ein Kompromiss aus beiden Bedingungen:

$$\lambda = \frac{\mathbf{v}'\mathbf{B}\mathbf{v}}{\mathbf{v}'\mathbf{W}\mathbf{v}} \rightarrow \text{Max}$$

unter der Restriktion $\mathbf{v}'\mathbf{v} = 1$.

¹Im Gegensatz zur Clusteranalyse ist die Zuordnung zu den Gruppe bekannt.

Die Lösung dieser Extremalaufgabe läuft auf das Eigenwertproblem hinaus, nämlich auf den Eigenvektor zum größten Eigenwert der Matrix $\mathbf{W}^{-1}\mathbf{B}$.

Bei mehr als zwei Gruppen gibt es in den meisten Fällen mehrere solcher Diskriminanzfunktionen, nämlich bei m Gruppen maximal $m-1$.

Die Diskriminanzfunktionen bilden zusammen den sog. „Diskriminanzraum“.

Hintergrund: die Ordnung der Matrix $\mathbf{W}^{-1}\mathbf{B}$ ist zwar q , für ihren Rang r gilt aber: $r \leq \min(q, m-1)$. Also hat sie auch nur r von 0 verschiedene Eigenwerte².

Für r verschiedene Diskriminanzfunktionen ergibt sich also die vollständige Lösung aus dem folgenden Ansatz:

$$(\mathbf{W}^{-1}\mathbf{B})\mathbf{V} = \mathbf{V} \text{diag}(\boldsymbol{\lambda})$$

$\text{diag}(\boldsymbol{\lambda})$ $r \times r$ -Diagonalmatrix mit den Eigenwerten
 \mathbf{V} $q \times r$ -Matrix mit den normierten Eigenvektoren

UNKORRELIERTHEIT DER DISKRIMINANZFUNKTIONEN

Gibt es mehr als eine Diskriminanzfunktion, so wird gefordert, dass die einzelnen Diskriminanzfunktionen unkorreliert sein sollen.

Die Eigenschaft der Orthogonalität der Eigenvektoren ist aber nur bei symmetrischen Matrizen erfüllt. Obwohl die Matrizen \mathbf{W}^{-1} und \mathbf{B} selbst symmetrisch sind, ist es ihr Produkt nicht.

Aus diesem Grunde verwendet man anstelle der ursprünglichen Gleichung

$$(\mathbf{W}^{-1}\mathbf{B})\mathbf{V} = \mathbf{V} \text{diag}(\boldsymbol{\lambda})$$

den folgenden Lösungsansatz:

$$(\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2})\mathbf{Q} = \mathbf{Q} \text{diag}(\boldsymbol{\lambda})$$

Die Matrix $\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$ ist symmetrisch und damit sind ihre Eigenvektoren \mathbf{q}_l orthogonal.

$\mathbf{W}^{-1}\mathbf{B}$ und $\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2}$ haben dieselben Eigenwerte λ_l , jedoch unterschiedliche Eigenvektoren.

\mathbf{V} hängt mit \mathbf{Q} in folgender Weise zusammen:

$$\mathbf{V} = \mathbf{W}^{-1/2}\mathbf{Q}$$

Die Spalten der Matrix \mathbf{V} sind die gesuchten Diskriminanzgewichte.

Test der Diskriminanzfunktionen

Die Eigenwerte λ_l der Matrix $\mathbf{W}^{-1}\mathbf{B}$ bilden die Grundlage mehrerer multivariater Tests auf Mittelwertsunterschiede.

Beispielsweise lässt sich Wilks' Lambda folgendermaßen ausdrücken:

$$\Lambda = \prod_{l=1}^r \frac{1}{1 + \lambda_l}$$

Nach Bartlett lässt sich diese Statistik umrechnen in

$$V = [n - 1 - (q + m)/2] \sum \ln(1 + \lambda_q)$$

V ist approximativ χ^2 -verteilt mit $q(m-1)$ Freiheitsgraden.

²Man beachte: die Matrix \mathbf{B} hat nicht vollen Rang.

5 Diskriminanzanalyse (DA)

Bartlett's Test wird als Residualtest schrittweise durchgeführt, d.h. zunächst gehen die Eigenwerte $\lambda_1, \dots, \lambda_r$ in die Testgröße ein, dann die Eigenwerte $\lambda_2, \dots, \lambda_r$ usw. Es werden also sukzessive immer nur die restlichen Eigenwerte getestet.

Wird der Test im Schritt l schließlich nicht mehr signifikant, so werden für die weitere Analyse nur die ersten $l-1$ Diskriminanzfunktionen verwendet.

Schrittweise Diskriminanzanalyse

Wie bei der multiplen Regressionsanalyse gibt es auch bei der Diskriminanzanalyse die Möglichkeit einer schrittweisen Selektion der Variablen.

In jedem Schritt wird jeweils jene Variable hinzugenommen, die die Zuordnung zu den Gruppen am meisten verbessert.

Diskriminanzscores

Die Werte der Datenpunkte auf den Diskriminanzfaktoren sind analog zur linearen Regression die gewichteten Summen der Variablen:

$$\mathbf{F} = \mathbf{XV}$$

Also ergibt sich für den Datenpunkt i der Wert im s -ten Diskriminanzfaktor durch:

$$f_{is} = \sum_{j=1}^q x_{ij}v_{js}$$

Ist $\bar{\mathbf{x}}^{(g)}$ der Mittelwertsvektor der g -ten Gruppe, so ist ihr Zentroid im Diskriminanzraum durch folgenden Vektor gegeben:

$$\bar{\mathbf{f}}^{(g)} = (\bar{\mathbf{x}}^{(g)'} \mathbf{V})' \quad \text{d.h.} \quad \bar{f}_s^{(g)} = (\bar{\mathbf{x}}^{(g)'} \mathbf{v}_s)'$$

Interpretation der Diskriminanzfaktoren

Für die Interpretation der Diskriminanzfunktionen können zunächst die Diskriminanzgewichte herangezogen werden. Diese stellen partielle Koeffizienten dar, bei denen der Einfluss der anderen Variablen ausgeschaltet ist.

Die Interpretation wird noch erleichtert, wenn man nicht die Rohgewichte, sondern die standardisierten Gewichte verwendet.

Für die *inhaltliche* Interpretation der Diskriminanzfaktoren eignen sich in erster Linie die Korrelationen zwischen den Variablen und den Diskriminanzfaktoren. Diese Korrelationen werden auch als „Strukturkoeffizienten“ oder „Faktorladungen“ bezeichnet. Diese Korrelationen haben eine höhere Stabilität als die Gewichte und zeigen die latenten Beziehungen zwischen den Originalvariablen und den Diskriminanzvariablen.

Rotation der Diskriminanzfaktoren

Die inhaltliche Interpretation eines Diskriminanzfaktors wird erleichtert, wenn einige Variablen ein hohes (auch negatives) Gewicht haben und alle anderen ein Nullgewicht.

Diese Situation kann man durch eine orthogonale Rotation herbeiführen. Dabei transformiert man die F -Matrix nach der Varimax-Methode mit einer orthogonalen Transformationsmatrix.

Man sollte bei der Rotation allerdings nicht alle, sondern nur die signifikanten Diskriminanzfaktoren einbeziehen.

Klassifikationsproblem

Man kann auf jeder einzelnen Diskriminanzfunktion einen Cutoff-Wert angeben, der die Zuordnung zu den einzelnen Populationen definiert.

Bei der Zuordnung von einzelnen Beobachtungen zu den Populationen geht man aber besser multivariat vor, d.h. man legt normalerweise die Mahalanobis-Distanzen der Beobachtungen zu den einzelnen Gruppenzentroiden zugrunde.

Bei der Klassifikation können auch die a priori Wahrscheinlichkeiten für die einzelnen Populationen ins Kalkül einbezogen werden.

Schließlich lässt sich auch noch wie bei jedem Entscheidungsproblem eine Auszahlungsmatrix in die Klassifikationsentscheidung einbeziehen.

Quadratische Diskriminanzanalyse

Eine Voraussetzung, die bisher gemacht wurde, ist die Homogenität der Kovarianzmatrizen.

Ist diese nicht erfüllt, so lässt sich keine lineare Diskriminanzfunktion, d.h. keine lineare Trennfunktion finden. Für diesen Fall steht aber als Methode die quadratische Diskriminanzanalyse zur Verfügung.

Gruppenzuordnung anhand der Mahalanobis-Distanz

Eine Alternative zu den bisher behandelten Ansätzen ergibt sich, wenn man für die Zuordnung zu den Gruppen die Mahalanobis-Distanz verwendet: man weist ein Element (Versuchsperson, Objekt, ...) jener Gruppe zu, zu der die Distanz am geringsten ist.

Hierbei können auch a-priori Wahrscheinlichkeiten berücksichtigt werden.

Häufig unterscheidet sich die auf diese Weise ermittelte Lösung von jener mit Hilfe der Diskriminanzanalyse gewonnenen.

6 Hauptkomponentenanalyse (PCA) und Faktorenanalyse (FA)

Hinsichtlich der Datenstruktur unterscheiden sich die Hauptkomponentenanalyse und die Faktorenanalyse von allen bisherigen Verfahren: es ist nur eine einzelne Menge von metrischen Variablen gegeben.

Diese Variablen sollen entweder auf eine geringere Anzahl von Komponenten reduziert werden (PCA) oder es soll das Zustandekommen der korrelativen Beziehungen zwischen den manifesten Variablen durch die Wirkung latenter Variablen erklärt werden (FA).

6.1 Hauptkomponentenanalyse

Ziel

Die Hauptkomponentenanalyse unterscheidet sich in ihrem Ziel und in ihren Ergebnissen von der Faktorenanalyse, ist jedoch hinsichtlich der mathematischen Behandlung mit ihr verwandt.

Das Hauptziel der Hauptkomponentenanalyse ist die Datenreduktion.

Aus den beobachteten ursprünglichen Variablen sollen durch geeignete Linearkombinationen einige wenige neue Variablen erzeugt werden, die die wesentliche Information des Datenbestandes beinhalten.

Dabei stellt man sich vor, dass der Großteil der Varianz des Datenbestandes durch einige wenige Komponenten aufgeklärt wird und der Rest Fehlervarianz ist.

Mathematisch geht es hierbei darum, das ursprüngliche Koordinatensystem im Personenraum schrittweise durch Rotation so zu positionieren dass auf den Koordinatenachsen sukzessiv jeweils maximale Varianz entsteht.

Spezielle Anwendungen der Hauptkomponentenanalyse

Häufig dient die Hauptkomponentenanalyse dazu, aus einer Vielzahl von Variablen, die eventuell auch noch hoch miteinander korreliert sind, in einem ersten Schritt einige wenige Variablen zu erzeugen, die dann im zweiten Schritt der eigentlichen Analyse zugeführt werden.

Beispiele hierfür sind:

- **Regression auf die Hauptkomponenten**

Hierbei werden für die Regression nicht die unabhängigen Variablen selbst, sondern deren Hauptkomponenten als Prädiktoren verwendet.

- **Partial Least Squares (PLS)**

PLS-Modelle sind eine Alternative für die multivariate multiple Regressionsanalyse oder die kanonische Korrelationsanalyse. Sie basieren auf den Hauptkomponenten sowohl der unabhängigen Variablen \mathbf{X} als auch der abhängigen Variablen \mathbf{Y} .

Für die Matrizen \mathbf{X} und \mathbf{Y} werden zunächst die Hauptkomponenten getrennt berechnet und anschließend wird ein Regressionsmodell zwischen den Scores der Hauptkomponenten (und nicht zwischen den Originaldaten) erstellt.

Dazu zerlegt man die Matrix \mathbf{X} in eine Scorematrix \mathbf{T} und eine Ladungsmatrix \mathbf{P} plus einer Fehlermatrix \mathbf{E} . Analog zerlegt man auch die Matrix \mathbf{Y} in die Matrizen \mathbf{U} und \mathbf{Q} und den Fehlerterm \mathbf{F} . Also:

$$\begin{aligned}\mathbf{X} &= \mathbf{TP}' + \mathbf{E} \\ \mathbf{Y} &= \mathbf{UQ}' + \mathbf{F}\end{aligned}$$

Diese zwei Gleichungen nennt man die „äußeren Beziehungen“.

Das Ziel von PLS ist es nun, diese beiden Datenräume miteinander zu verbinden, d.h. die Matrizen \mathbf{U} und \mathbf{T} zueinander in Beziehung zu setzen:

$$\mathbf{U} = \mathbf{BT}$$

Diese Gleichung nennt man die „innere Beziehung“.

Für die praktische Berechnung setzt man ein iteratives Verfahren ein: es wird jeweils die größte Hauptkomponente der einen Seite (X bzw. Y) als Scorevektor für die andere Seite verwendet, um deren Ladungsmatrix zu berechnen, und so geht es hin und her, bis das Verfahren schließlich konvergiert. Die Normen der Fehlermatrizen werden dabei minimiert.

Hauptachsentransformation

Die Grundlage für die Hauptkomponentenanalyse ist die Hauptachsentransformation

Diese beruht auf der Singulärwertzerlegung (bzw. Diagonalisierung) der Kovarianzmatrix:

$$\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$$

- \mathbf{S} Kovarianzmatrix
- $\mathbf{\Lambda}$ Diagonalmatrix der Eigenwerte der Kovarianzmatrix
- \mathbf{V} Eigenvektoren der Kovarianzmatrix

Der ursprüngliche Vektorraum mit der durch die Variablen gegebenen Basis wird in einen Vektorraum mit neuer Basis überführt:

$$\mathbf{x} \mapsto \mathbf{y} = \mathbf{V}'\mathbf{x}$$

Die dabei verwendete orthogonale Transformationsmatrix \mathbf{V} ist die Matrix der normierten Eigenvektoren der Kovarianzmatrix. Diese bilden die neue Basis, die sogenannten „Hauptkomponenten“.

Die Eigenwerte sind die Varianzen der Hauptkomponenten, d.h. die Kovarianzmatrix von \mathbf{y} wurde in die Diagonalmatrix

$$\mathbf{\Lambda} = \mathbf{V}'\mathbf{C}\mathbf{V}$$

überführt.

Hauptkomponenten

Die Summe der Eigenwerte der Kovarianzmatrix ergibt die Gesamtvarianz. Diese entspricht auch der Spur der Kovarianzmatrix.

Sortiert man die Eigenwerte nach ihrer Größe, dann erklärt die erste Hauptkomponente den größten Teil der Gesamtvarianz, die zweite Hauptkomponente den größten Teil der von der ersten Hauptkomponente nicht erklärten Gesamtvarianz usw.

Insgesamt ergeben sich so für einen p -dimensionalen Datensatz p Hauptkomponenten mit immer kleinerem Varianzanteil.

Bei der Hauptkomponentenanalyse werden nun nicht alle p , sondern nur die ersten r ($r \ll p$) Hauptkomponenten verwendet, die einen festgelegten Teil der Gesamtvarianz auf sich vereinigen.

6.2 Faktorenanalyse

Bei der Faktorenanalyse wird ein mathematisches Modell angepasst und getestet, in dem das Zustandekommen korrelativer Beziehungen innerhalb einer Menge von Zufallsvariablen durch die Wirkung von latenten Variablen, den sog. *Faktoren*, erklärt werden soll.

Als Nebeneffekt ergibt sich auch bei der Faktorenanalyse eine Datenreduktion, diese ist aber nicht ihr Hauptziel.

In Bezug auf die Zielsetzung kann man zwei verschiedene Arten von Faktorenanalysen unterscheiden:

- *Exploratorische Faktorenanalyse*:
Diese ist ein hypothesengenerierendes Verfahren, d.h. es sollen erst Faktoren gefunden werden. Das entsprechende mathematische Modell ist teilweise unbestimmt und enthält beispielsweise keine Angabe über die Anzahl der Faktoren.
- *Konfirmatorische Faktorenanalyse*:
Diese ist ein hypothesentestendes Verfahren, d.h. es soll ein konkret gegebenes faktorenanalytisches Modell angepasst und getestet werden. Konfirmatorische Faktorenanalysen sind Spezialfälle von linearen Strukturgleichungsmodellen (LISREL).

Grundgleichung der Faktorenanalyse

Im Populationsmodell der Faktorenanalyse wird der Wert der Zufallsvariable X_i als Funktion von latenten Zufallsvariablen beschrieben, nämlich als Linearkombination von sog. gemeinsamen Faktoren, sowie einem spezifischen Faktor:

$$X_l = \mu_l + \alpha_{l1}F_1 + \alpha_{l2}F_2 + \cdots + \alpha_{lr}F_r + \varepsilon_l$$

- μ_l Erwartungswert von Variable l
- F_s Faktorscore auf dem s -ten gemeinsamen Faktor
- α_{ls} Ladung der l -ten Variable auf dem s -ten Faktor
- ε_l spezifischer Faktor der l -ten Variable

In der Praxis interessiert man sich meist nicht für die Erwartungswerte μ_l , sondern geht von Abweichungswerten oder von standardisierten Variablen \mathbf{X}^* aus, so dass sich das obige Modell noch etwas vereinfacht¹.

In Matrixschreibweise in der Version für standardisierte Variablen lautet das mathematische Modell der orthogonalen Faktorenanalyse:

$$\mathbf{X}^* = \mathbf{F}^* \mathbf{A}^{*'} + \mathbf{E}^*$$

Annahmen des orthogonalen Faktormodells

- $E(\mathbf{X}) = \boldsymbol{\mu}$
- $\text{cov}(\mathbf{X}) = \boldsymbol{\Sigma}$
- $E(\mathbf{F}) = E(\boldsymbol{\varepsilon}) = \mathbf{0}$
- $\text{cov}(\mathbf{F}) = \mathbf{I}$
- $\text{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi} = \text{diag}(\boldsymbol{\psi})$ mit $\psi_l > 0$.
- $\text{cov}(\mathbf{F}, \boldsymbol{\varepsilon}) = \mathbf{0}$

Über die Verteilung von \mathbf{X} gibt es im Faktormodell normalerweise keine Annahmen. Unter der Zusatzannahme, dass die Faktorscores und die spezifischen Faktoren normalverteilte Zufallsvariablen sind, lassen sich ML-Schätzungen für die Parameter und Hypothesentests für die Modellgültigkeit ableiten.

¹Es interessieren meist nur die korrelativen Beziehungen zwischen den Variablen.

Fundamentaltheorem der Faktorenanalyse

Aus den Annahmen folgt für die Kovarianzmatrix (bzw. Korrelationsmatrix) von \mathbf{X} :

$$\Sigma = \mathcal{A}\mathcal{A}' + \Psi$$

bzw.

$$\mathcal{R} = \mathcal{A}^* \mathcal{A}^{*'} + \Psi^*$$

Bemerkung: da wir im Folgenden immer von standardisierten Variablen ausgehen wollen, sollen zur Vereinfachung der Schreibweise die Symbole \mathcal{A} , Ψ usw. (ohne „*“) verwendet werden. Also:

$$\mathcal{R} = \mathcal{A}\mathcal{A}' + \Psi$$

Varianzaufteilung

Die Gesamtvarianz einer Variable setzt sich zusammen aus der gemeinsamen Varianz mit den Faktoren („Kommunalität“) und der Einzelvarianz.

Bei der Einzelvarianz kann man unterscheiden zwischen der spezifischen Varianz („Spezifität“) und der Fehlervarianz².

Also:

$$\sigma_i^2 = h_i^2 + u_i^2 \quad \text{mit} \quad u_i^2 = b_i^2 + \varepsilon_i^2$$

Für die Extraktion der Faktoren wird nicht die Korrelationsmatrix \mathbf{R} genommen, sondern die sog. „reduzierte Korrelationsmatrix“ \mathbf{R}_h . Bei dieser sind die Diagonalelemente durch die Kommunalitäten ersetzt, d.h.

$$\mathbf{R}_h = \mathbf{R} - \hat{\Psi}$$

Schätzung der Parameter

Es gibt verschiedene iterative und nicht-iterative Verfahren zur Schätzung der Parameter des faktorenanalytischen Modells.

All diese Verfahren kommen aber nicht ohne Zusatzannahmen aus.

Meistens beziehen sich diese Zusatzannahmen auf die Anzahl der gemeinsamen Faktoren oder auf deren Anteil an der Gesamtvarianz der Variablen (Kommunalität).

Anzahl der zu extrahierenden Faktoren

Um zu entscheiden, wieviele gemeinsame Faktoren verwendet werden sollen, sind verschiedene Kriterien gebräuchlich:

- Eigenwertkriterium: verwende nur Faktoren mit einem Eigenwert ≥ 1 .
- Prozentsatzkriterium: verwende so viele Faktoren, dass sie einen bestimmten Prozentsatz der Gesamtvarianz aufklären (z.B. 75%).
- Screeplot-Kriterium: verwende alle Faktoren, deren Eigenwerte im Screeplot oberhalb des Knickes (Übergang zu einer Geraden) liegen.

²Die Fehlervarianz ist der unreliable Varianzanteil.

Faktorladungen

Die Faktorladung α_{ls} ist die Korrelation der Variable l mit dem Faktor s .

Ist das Kommunalitätenproblem gelöst, d.h. liegt eine Schätzung der reduzierten Korrelationsmatrix \mathcal{R}_h vor, so kann man die Faktorladungen mit Hilfe der Singulärwertzerlegung schätzen:

$$\mathbf{R}_h = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$$

Also:

$$\hat{\mathbf{A}} = \mathbf{A} = \mathbf{U}\mathbf{\Lambda}^{1/2}$$

mit $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$.

Nach dem Fundamentaltheorem der Faktorenanalyse gilt:

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

Orthogonale Rotation

Nach dem Fundamentaltheorem gilt

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

Es gilt aber auch

$$\mathbf{R}_h = \mathbf{A}\mathbf{I}\mathbf{A}' = \mathbf{A}\mathbf{T}\mathbf{T}'\mathbf{A}' = \mathbf{B}\mathbf{B}'$$

mit jeder orthonormalen Matrix \mathbf{T} . Die Ladungsmatrix der rotierten Faktorenlösung ist dann $\mathbf{B} = \mathbf{A}\mathbf{T}$.

Die verschiedenen Lösungen stellen unterschiedliche Rotationen und/oder Spiegelungen der ursprünglichen Lösung dar.

Die Transformationsmatrix \mathbf{T} für eine Rotation um den Winkel ϕ gegen den Uhrzeigersinn ist

$$\mathbf{T} = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}$$

Eine Rotation in einem mehrdimensionalen Koordinatensystem kann man sich als Hintereinanderschaltung von Rotationen von jeweils zwei Achsen vorstellen (die restlichen Achsen werden dann jeweils identisch abgebildet). Beispiel einer Rotation der ersten beiden Achsen im dreidimensionalen Raum:

$$\mathbf{T} = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So gibt es also unendlich viele gleichwertige Faktorenlösungen mit der Ausnahme, dass sich die aufgeklärten Varianzanteile unterschiedlich auf die Faktoren verteilen.

In der Praxis werden Lösungen bevorzugt, bei denen sich die Faktoren inhaltlich leicht interpretieren lassen.

Unter den verschiedenen Rotationskriterien ist Kaiser's *Varimax-Kriterium* das am häufigsten verwendete: hier wird die Varianz der Ladungsquadrate maximiert.

Ideale Bedingungen für die Interpretierbarkeit einer Faktorenlösung sind gegeben, wenn eine sog. *Einfachstruktur* vorliegt:

- Jeder Faktor lädt auf möglichst wenigen Variablen³.
- Jede Variable lädt möglichst nur auf einem Faktor.

Damit werden die Variablen in sich gegenseitig ausschließende Gruppen eingeteilt.

³D.h. deren Ladungen sind dem Absolutbetrag nach hoch und die Ladungen der restlichen Variablen nahe Null

Schiefwinklige Rotation

Zuweilen wird zum Zwecke einer besseren Interpretierbarkeit im Sinne einer Einfachstruktur anstelle einer orthogonalen eine schiefwinklige (oblique) Rotation der Faktoren durchgeführt.

Für eine solche Rotation sind verschiedene Kriterien vorgeschlagen worden (z.B. Oblimin oder Promax).

Bei einer schiefwinkligen Rotation geht allerdings die Eigenschaft der Unkorreliertheit der Faktoren verloren:

$$\mathbf{R}_h = \mathbf{A}\mathbf{T}\mathbf{T}'\mathbf{A}' = \mathbf{A}\mathbf{\Phi}\mathbf{A}'$$

T nicht-orthogonale Transformationsmatrix

Φ Korrelationsmatrix der Faktoren.

Außerdem sind die Ladungen nicht mehr als Korrelationen zwischen den Variablen und den Faktoren zu interpretieren.

7 Kanonische Korrelationsanalyse (CA)

Zweck

Mit Hilfe der kanonischen Korrelationsanalyse soll der Zusammenhang zwischen zwei Variablenmengen $\mathbf{X} = [X_1, X_2, \dots, X_p]'$ und $\mathbf{Y} = [Y_1, Y_2, \dots, Y_q]'$ untersucht werden.

Sie stellt damit eine Verallgemeinerung der multiplen Korrelation dar.

Von der Datenstruktur her ist die kanonische Korrelationsanalyse verwandt mit der multivariaten multiplen Regression.

Im Unterschied zu ihr ist die kanonische Korrelationsanalyse normalerweise ungerichtet, d.h. es gibt keine Prädiktorvariablen (unabhängige Variablen) und keine Responsevariablen (abhängige Variablen), sondern lediglich zwei unterschiedliche Variablenmengen.

Zielfunktion

Gesucht werden paarweise orthogonale Linearkombinationen der X -Variablen und solche der Y -Variablen:

$$\begin{array}{ll} U_1 = \mathbf{X}\mathbf{a}_1 & V_1 = \mathbf{Y}\mathbf{b}_1 \\ \dots & \dots \\ U_r = \mathbf{X}\mathbf{a}_r & V_r = \mathbf{Y}\mathbf{b}_r \end{array}$$

Dabei sind die Gewichte \mathbf{a}_1 und \mathbf{b}_1 so zu bestimmen, dass die Korrelation zwischen U_1 und V_1 maximal wird.

Weiterhin sind restlichen Linearkombinationen der Reihe nach so zu bilden, dass die Korrelation zwischen den entsprechenden Linearkombinationen jeweils maximal wird.

Es gilt übrigens

$$r \leq \min(p, q)$$

Die Variablen U_s und V_s werden als kanonische Funktionen oder auch als *kanonische Faktoren* bezeichnet.

Lösung

Für die Lösungsgleichung werden die gemeinsame Kovarianzmatrix \mathbf{S} der X - und Y -Variablen, sowie die partitionierten Kovarianzmatrizen \mathbf{S}_{XX} , \mathbf{S}_{YY} und \mathbf{S}_{XY} benötigt:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix}$$

Die Lösung der Extremalaufgabe läuft auf ein Eigenwertproblem hinaus, nämlich auf die Eigenwerte der Matrix $\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}$.

Die Matrix $\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}$ ist jedoch nicht symmetrisch und damit sind ihre Eigenvektoren nicht orthogonal.

Daher wird der folgende alternative Ansatz gewählt:

$$(\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1/2})\mathbf{Q} = \mathbf{Q}\text{diag}(\boldsymbol{\lambda})$$

¹ \mathbf{S}_{XX} und \mathbf{S}_{YY} dürfen nicht singulär sein.

λ ist der Vektor der Eigenwerte zur symmetrischen Matrix $\mathbf{S}_{XX}^{-1/2} \mathbf{S}_{XY} \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX} \mathbf{S}_{XX}^{-1/2}$ und \mathbf{Q} die Matrix der dazugehörigen Eigenvektoren.

Kanonischer Korrelationskoeffizient

Die Korrelationen r_c zwischen U_1 und V_1 , U_2 und V_2 usw. werden kanonische Korrelationskoeffizienten genannt.

Die kanonischen Korrelationskoeffizienten sind die positiven Quadratwurzeln der λ_s , der Eigenwerte der Matrix $\mathbf{S}_{XX}^{-1/2} \mathbf{S}_{XY} \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX} \mathbf{S}_{XX}^{-1/2}$.

Test der kanonischen Funktionen

Nach Bartlett lässt sich auch im Falle der kanonischen Korrelationsanalyse mit Hilfe der gefundenen Eigenwerte λ_s eine Teststatistik berechnen:

$$\chi_j^2 = -\left\{n - \frac{1}{2}(p + q + 1)\right\} \sum_{s=j+1}^r \ln(1 - \lambda_s) \quad \text{mit } df = (p - j)(q - j)$$

Also wird Bartlett's Test als Residualtest schrittweise durchgeführt, d.h. zunächst gehen die Eigenwerte $\lambda_1, \dots, \lambda_r$ in die Testgröße ein, dann die Eigenwerte $\lambda_2, \dots, \lambda_r$ usw. Es werden also sukzessive immer nur die restlichen Eigenwerte getestet.

Wird der Test im Schritt j schließlich nicht mehr signifikant, so werden für die weitere Analyse nur die ersten $j-1$ kanonischen Funktionen verwendet.

Kanonische Gewichte und Scores

Die kanonischen Gewichte der X -Variablen werden wie folgt berechnet:

$$\mathbf{A} = \mathbf{S}_{XX}^{-1/2} \mathbf{Q}$$

und jene der Y -Variablen wie folgt:

$$\mathbf{B} = \mathbf{S}_{YY}^{-1} \mathbf{S}_{YX} \mathbf{S}_{XX}^{-1/2} \mathbf{Q} \text{diag}(\mathbf{r}_c^{-1})$$

Mit Hilfe dieser Gewichte können die kanonischen Faktorscores berechnet werden:

$$\mathbf{V} = \mathbf{X}\mathbf{A} \quad \text{und} \quad \mathbf{W} = \mathbf{Y}\mathbf{B}$$

Interpretation der kanonischen Faktoren

Jedes Paar von kanonischen Variablen (U_1, V_1) , (U_2, V_2) , ... repräsentiert eine unabhängige Dimension in der Beziehung zwischen den beiden Variablenmengen.

(U_1, V_1) hat die höchstmögliche Korrelation und ist daher am wichtigsten, danach (U_2, V_2) usw.

Für die Interpretation der kanonischen Funktionen können zunächst die kanonischen Gewichte herangezogen werden.

Die Interpretation wird noch erleichtert, wenn man nicht die Rohgewichte, sondern die standardisierten Gewichte verwendet.

Für die *inhaltliche* Interpretation der Faktoren eignen sich in erster Linie die Korrelationen bzw. Kreuzkorrelationen zwischen den Variablen und den kanonischen Faktoren. Diese Korrelationen werden auch als „Strukturkoeffizienten“ oder „Faktorladungen“ bezeichnet.

Rotation der kanonischen Faktoren

Die inhaltliche Interpretation eines kanonischen Faktors wird erleichtert, wenn einige Variablen ein hohes (auch negatives) Gewicht haben und alle anderen ein Nullgewicht.

Diese Situation kann man durch eine orthogonale Rotation herbeiführen. Dabei transformiert man die **F**-Matrizen nach der Varimax-Methode mit einer orthogonalen Transformationsmatrix.

Man sollte bei der Rotation allerdings nicht alle, sondern nur die signifikanten Faktoren einbeziehen.

8 Vektor- und Matrizenrechnung

8.1 Skalare

Innerhalb der Vektor- und Matrizenrechnung bezeichnet man einzelne Zahlen als *Skalare*.

Als Symbole für Skalare werden meistens normalgedruckte Kleinbuchstaben verwendet.

8.2 Vektoren

In der multivariaten Analyse betrachtet man gleichzeitig mehrere Variablen.

Diese Variablen bzw. die Werte einer Person in diesen Variablen werden üblicherweise zu einem *Vektor* zusammengefasst.

Andersherum können auch die Werte verschiedener Personen in einer Variablen zu einem Vektor zusammengefasst werden.

Allgemein ist ein Vektor ein n -Tupel, d.h. eine geordnete Menge von Elementen.

Man ordnet die Elemente eines Vektors untereinander oder nebeneinander an und schließt sie in runden oder eckigen Klammern ein.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{bzw.} \quad [x_1 \ x_2 \ \cdots \ x_n]$$

Die einzelnen Elemente des Vektors nennt man seine *Komponenten*. Diese können beliebige Größen sein, im einfachsten Falle reelle Zahlen.

Die Anzahl der Komponenten nennt man die *Dimension* des Vektors.

Für die Benennung von Vektoren verwendet man üblicherweise fettgedruckte Kleinbuchstaben (z.B. \mathbf{x}).

8.2.1 Spalten- und Zeilenvektoren

Benutzt man Vektoren innerhalb der Matrizenrechnung, so muß man zwischen *Spaltenvektoren* und *Zeilenvektoren* unterscheiden.

Konventionsgemäß werden Vektoren in ihrer ursprünglichen Form als Spaltenvektoren aufgefaßt:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Der Vektor \mathbf{x} kann in folgender Weise als Zeilenvektor \mathbf{x}' (d.h. als die Transponierte von \mathbf{x} , s.u.) geschrieben werden:

$$\mathbf{x}' = [x_1 \ x_2 \ \cdots \ x_n]$$

8.2.2 Rechnen mit Vektoren

Gleichheit: zwei Vektoren sind gleich, wenn sie in ihrer Dimension und in allen Komponenten übereinstimmen.

Durch *Transposition* wird ein Spaltenvektor zu einem Zeilenvektor und umgekehrt. Der Operator ist das Zeichen $'$ oder \top (für die Transponierte von \mathbf{x} schreibt man also \mathbf{x}' bzw. \mathbf{x}^\top).

Multiplikation eines Vektors mit einem Skalar:

$$\mathbf{z} = c\mathbf{x} \quad \text{ist definiert durch} \quad z_i = cx_i$$

Die *Addition* bzw. *Subtraktion* zweier Vektoren wird komponentenweise durchgeführt:

$$\mathbf{z} = \mathbf{x} \pm \mathbf{y} \quad \text{ist definiert durch} \quad z_i = x_i \pm y_i$$

Dazu müssen die Vektoren die gleiche Dimension haben.

Manche Computerprogramme (z.B. SPSS, S-Plus) unterstützen noch eine in der Vektorrechnung üblicherweise nicht definierte Operation, nämlich die *Addition* bzw. *Subtraktion* eines Vektors und eines Skalars. Diese könnte auf folgende Weise definiert werden:

$$\mathbf{z} = \mathbf{x} \pm c \quad \text{definieren wir durch} \quad z_i = x_i \pm c$$

Inneres Produkt (auch *Skalarprodukt* genannt):

Es wird berechnet durch elementweise Multiplikation der beiden Vektoren und anschließendes Aufsummieren der Produkte. Das Ergebnis ist ein Skalar:

$$s = \mathbf{x}'\mathbf{y} \quad \text{ist definiert durch} \quad s = \sum_i x_i y_i$$

Voraussetzungen:

- Die Vektoren müssen die gleiche Dimension haben.
- Der erste Vektor muß ein Zeilenvektor und der zweite ein Spaltenvektor sein.

Beispiel:

$$\begin{bmatrix} 2 & 1 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} = 2 \times 3 + 1 \times (-2) + 6 \times 1 = 10$$

Umgekehrt kann man z.B. $2a + 3b$ als inneres Produkt schreiben:

$$\begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

Beispiele für das Rechnen mit Vektoren

MULTIPLIKATION EINES VEKTORS MIT EINEM SKALAR

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad 1.5\mathbf{x} = \begin{bmatrix} 3 \\ 1.5 \end{bmatrix} \qquad -\mathbf{x} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

ADDITION VON ZWEI VEKTOREN

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad \mathbf{x} + \mathbf{y} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

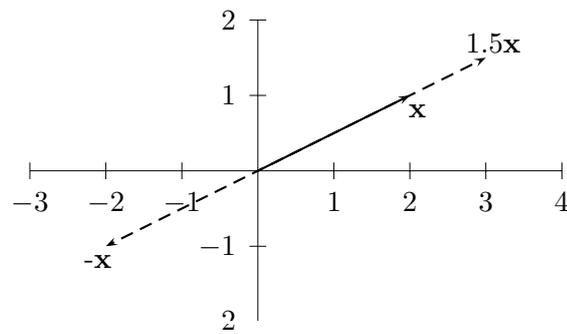


Abbildung 8.1: Multiplikation eines Vektors mit einem Skalar

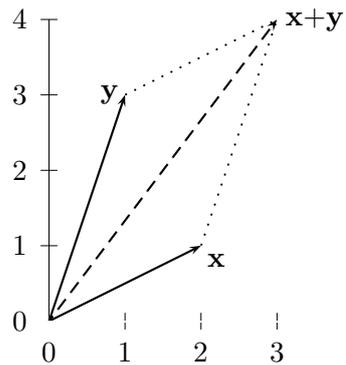


Abbildung 8.2: Addition zweier Vektoren

SUBTRAKTION VON ZWEI VEKTOREN

$$\mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

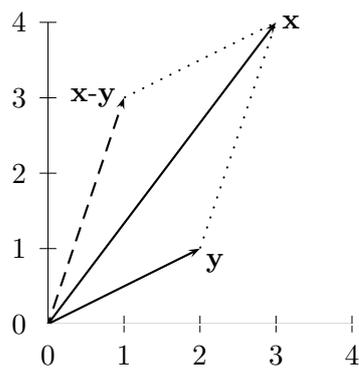


Abbildung 8.3: Vektordifferenz

INNERES PRODUKT EINES VEKTORES MIT SICH SELBST

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\mathbf{x}'\mathbf{x} = [2 \quad 1] \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2^2 + 1^2 = 5$$

INNERES PRODUKT ZWEIER VEKTOREN

$$\mathbf{x} = \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \mathbf{x}'\mathbf{y} = [2.5 \quad 0.5] \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 2.5 \times 1 + 0.5 \times 3 = 4$$

8.2.3 Vektoren geometrisch interpretiert

Häufig werden Vektoren durch gerichtete Pfeile von einem Startpunkt zu einem Zielpunkt dargestellt.

Ein solcher Pfeil hat eine *Länge* und eine *Richtung*.

Pfeile, die gleich lang sind und in die gleiche Richtung zeigen, werden als gleich betrachtet, d.h. durch Parallelverschiebung wird ein Vektor nicht verändert.

Entspringt ein Pfeil im Koordinatenursprung, so spricht man von einem *Ortsvektor*.

Stellt man einen Vektor mit n Komponenten als Ortsvektor dar, so kann man ihn geometrisch auch als einen Punkt im n -dimensionalen euklidischen Raum \mathbb{R}^n interpretieren, dessen Koordinaten den Vektorkomponenten entsprechen.

Der *Nullvektor*, d.h. der Vektor, dessen Elemente alle Null sind, entspricht dem Koordinatenursprung.

8.2.4 Länge, Winkel und Abstand von Vektoren

$\mathbf{x}'\mathbf{x}$, das innere Produkt eines Vektors \mathbf{x} mit sich selbst, ergibt das Quadrat seiner Länge. Für die *Länge* von \mathbf{x} verwenden wir die Bezeichnung $\|\mathbf{x}\|$. Es gilt also:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{\sum x_i^2}$$

Ein Vektor der Länge 1 heißt *normierter Vektor*.

Jeder Vektor (außer dem Nullvektor) lässt sich *normieren*, indem man ihn durch seine Länge dividiert:

$$\frac{\mathbf{x}}{\|\mathbf{x}\|}$$

Sind \mathbf{x} und \mathbf{y} zwei Vektoren gleicher Dimension, so ist das innere Produkt $\mathbf{x}'\mathbf{y}$ gleich dem Produkt der Längen der beiden Vektoren mal dem Cosinus des eingeschlossenen Winkels:

$$\mathbf{x}'\mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \angle(\mathbf{x}, \mathbf{y})$$

Der *Winkel* zwischen zwei Vektoren \mathbf{x} und \mathbf{y} lässt sich also über die folgende Beziehung berechnen:

$$\cos \angle(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

Ist $\cos \angle(\mathbf{x}, \mathbf{y}) = 1$ oder $\cos \angle(\mathbf{x}, \mathbf{y}) = -1$, so zeigen beide Vektoren in die gleiche bzw. entgegengesetzte Richtung, d.h. dann gilt $\mathbf{y} = \alpha \mathbf{x}$ mit einem Skalar α .

Ist $\cos \angle(\mathbf{x}, \mathbf{y}) = 0$, so sind die beiden Vektoren *orthogonal* zueinander. Dafür schreibt man auch $\mathbf{x} \perp \mathbf{y}$.

Der *Abstand* zwischen \mathbf{x} und \mathbf{y} entspricht der Länge ihrer Differenz:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

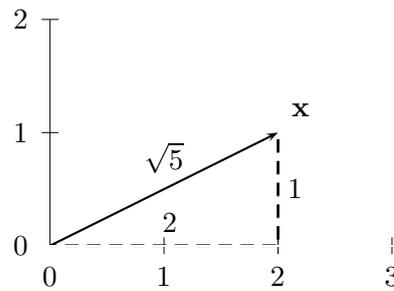


Abbildung 8.4: Länge eines Vektors

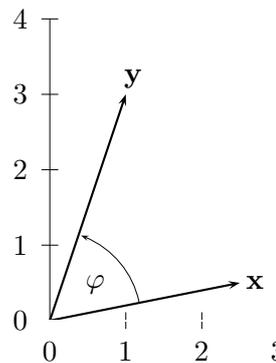


Abbildung 8.5: Winkel zwischen zwei Vektoren

HERLEITUNG DES WINKELS FÜR DEN ZWEIDIMENSIONALEN FALL

Wir betrachten zwei Vektoren $\mathbf{v}_1 = [x_1 \ y_1]'$ und $\mathbf{v}_2 = [x_2 \ y_2]'$. Dann ergibt sich für deren Abstand:

$$\begin{aligned} d^2 &= (x_1 - x_2)^2 + (y_1 - y_2)^2 \\ &= x_1^2 + x_2^2 - 2x_1x_2 + y_1^2 + y_2^2 - 2y_1y_2 \\ &= (x_1^2 + y_1^2) + (x_2^2 + y_2^2) - 2(x_1x_2 + y_1y_2) \\ &= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2(x_1x_2 + y_1y_2) \end{aligned}$$

Was ist nun der Ausdruck $x_1x_2 + y_1y_2$?

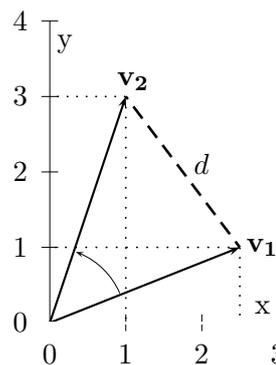


Abbildung 8.6: Winkel zwischen Vektoren

Nach dem Cosinussatz für beliebige Dreiecke $\triangle ABC$ gilt:

$$a^2 = b^2 + c^2 - 2bc \cos \alpha$$

Dies übertragen auf die obige Ableitung ergibt:

$$\begin{aligned} d^2 &= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2(x_1x_2 + y_1y_2) \\ &= \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 - 2\|\mathbf{v}_1\|\|\mathbf{v}_2\|\cos \angle(\mathbf{v}_1, \mathbf{v}_2) \end{aligned}$$

Also ist $\|\mathbf{v}_1\| \|\mathbf{v}_2\| \cos \angle(\mathbf{v}_1, \mathbf{v}_2) = (x_1 x_2 + y_1 y_2) = [x_1 \ y_1]' \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \mathbf{v}_1' \mathbf{v}_2$.

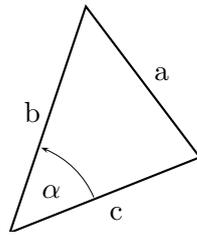


Abbildung 8.7: Cosinussatz

Richtungscosinus

Ein n -dimensionaler Vektor \mathbf{x} bildet zu den n Achsen n Winkel.

Die Cosinusse dieser Winkel werden *Richtungscosinusse* genannt. Sie geben die Richtung des Vektors \mathbf{x} an, unabhängig von seiner Länge.

Ein Vektor ist eindeutig definiert durch seine Länge und seine Richtungscosinusse.

Ist \mathbf{k} ein Vektor von Richtungscosinussen, so gilt

$$\mathbf{k}'\mathbf{k} = \sum k_i^2 = 1 \quad \text{d.h.} \quad \|\mathbf{k}\| = 1$$

Dies stellt auch eine Bedingung dar, die ein Vektor erfüllen muss, um in einem Raum bestimmter Dimension darstellbar zu sein. Beispielsweise kann man in der Ebene keinen Vektor zeichnen, der zu beiden Achsen jeweils einen Winkel von 50° bildet.

8.2.5 Orthogonale Projektion

Wir betrachten zwei n -dimensionale Vektoren \mathbf{x} und \mathbf{y} .

Für die *orthogonale Projektion* von \mathbf{y} auf \mathbf{x} schreiben wir $P_{\mathbf{x}}\mathbf{y}$. Es gilt:

$$P_{\mathbf{x}}\mathbf{y} = \alpha\mathbf{x} \quad \text{mit} \quad \alpha = \frac{\mathbf{y}'\mathbf{x}}{\|\mathbf{x}\|^2}$$

Die Länge der Projektion eines Vektors \mathbf{y} auf einen *Einheitsvektor* \mathbf{e}_i (Vektor mit einer Eins an i -ter Stelle und sonst alles Nullen) ist y_i .

Für die Projektion eines Vektors \mathbf{y} auf den *Einsvektor* (Vektor aus lauter Einsen) erhält man übrigens:

$$P_1\mathbf{y} = [\bar{y} \ \bar{y} \ \dots \ \bar{y}]'$$

8.2.6 Vektorräume

Ein n -dimensionaler *Vektorraum* V_n ist die Menge der n -dimensionalen Vektoren, die bezüglich Vektoraddition und Skalarmultiplikation abgeschlossen ist, d.h. für die gilt:

1. Ist $\mathbf{x} \in V_n$ und $\mathbf{y} \in V_n$, dann ist auch $\mathbf{z} = \mathbf{x} + \mathbf{y} \in V_n$.
2. Ist α ein beliebiger Skalar und $\mathbf{x} \in V_n$, dann ist auch $\mathbf{z} = \alpha\mathbf{x} \in V_n$.

Aus der Definition der Vektoraddition und Skalarmultiplikation folgen verschiedene Gesetze, die in einem Vektorraum gelten müssen, z.B. das Kommutativgesetz: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

In jedem Vektorraum gibt es einen Vektor \mathbf{o} , den *Nullvektor*, mit der Eigenschaft $\mathbf{x} + \mathbf{o} = \mathbf{x}$ für jedes \mathbf{x} .

Weiterhin existiert zu jedem Vektor \mathbf{x} genau ein Vektor \mathbf{v} mit der Eigenschaft $\mathbf{x} + \mathbf{v} = \mathbf{o}$. Dieser Vektor wird mit $-\mathbf{x}$ bezeichnet.

Variablenraum vs. Versuchspersonenraum

Eine $n \times m$ -Datenmatrix \mathbf{X} lässt sich auffassen als ein System von n m -dimensionalen Spaltenvektoren oder als ein System von m n -dimensionalen Zeilenvektoren. In beiden Fällen ist die gesamte in \mathbf{X} enthaltene Information repräsentiert.

Üblicherweise werden die Daten von n Versuchspersonen in m Variablen dargestellt als n Punkte in einem m -dimensionalen Koordinatensystem. Die m Dimensionen stellen die Variablen dar und die n Punkte die Versuchspersonen.

Umgekehrt kann man dieselben Daten auch darstellen als m Punkte in einem n -dimensionalen Koordinatensystem. Dann stellen die n Dimensionen die Versuchspersonen dar und die m Punkte die Variablen.

Beispiel

$$\mathbf{X} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 6 & 2 \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2] = \begin{bmatrix} \mathbf{p}'_1 \\ \mathbf{p}'_2 \\ \mathbf{p}'_3 \end{bmatrix}$$

mit

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$$

und

$$\mathbf{p}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \mathbf{p}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad \mathbf{p}_3 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

Vektorunterräume

Eine Teilmenge S eines Vektorraums V_n ist ein *Unterraum*, falls S selbst ein Vektorraum ist.

Ein n -dimensionaler Raum enthält neben $(n-1)$ -dimensionalen Unterräumen auch $(n-2)$ -, $(n-3)$ -, ..., 2 -, 1 -, 0 -dimensionale Unterräume.

Beispielsweise bilden die Vektoren der Form $\mathbf{z} = \alpha\mathbf{x} + \beta\mathbf{y}$ mit beliebigen Skalaren α und β und den beiden Vektoren $\mathbf{x} = [1 \ 0 \ 0]'$ und $\mathbf{y} = [0 \ 1 \ 0]'$ eine Ebene (d.h. einen zweidimensionalen Raum) im dreidimensionalen Raum V_3 .

$(n-1)$ -dimensionale Unterräume eines Vektorraums V_n werden oft auch als *Hyperebenen* bezeichnet.

Basis eines Vektorraums

Eine Menge von Vektoren $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ bilden eine *Basis* eines Vektorraums V_n , falls jeder beliebige Vektor $\mathbf{v} \in V_n$ sich als Linearkombination dieser Vektoren darstellen lässt:

$$\mathbf{v} = \sum \alpha_i \mathbf{y}_i$$

Man sagt auch, dass diese Vektoren „den Vektorraum V_n aufspannen“.

Sind die Basisvektoren paarweise orthogonal, so bilden sie eine *Orthogonalbasis* des Vektorraums.

Sind sie zusätzlich normiert, so bilden sie eine *Orthonormalbasis* des Vektorraums.

Gram-Schmidt'sches Orthogonalisierungsverfahren

Das Ziel dieses Verfahrens ist die Erzeugung von orthogonalen bzw. orthonormalen Vektoren $\mathbf{y}_1, \mathbf{y}_2, \dots$ aus k beliebigen Vektoren $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$.

Das Verfahren sieht wie folgt aus:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{x}_1 \\ \mathbf{y}_2 &= \mathbf{x}_2 - P_{\mathbf{y}_1} \mathbf{x}_2 && (\mathbf{y}_2 \perp \mathbf{y}_1) \\ \mathbf{y}_3 &= \mathbf{x}_3 - P_{\mathbf{y}_1} \mathbf{x}_3 - P_{\mathbf{y}_2} \mathbf{x}_3 && (\mathbf{y}_3 \perp \mathbf{y}_2 \perp \mathbf{y}_1) \\ \dots & \\ \mathbf{y}_i &= \mathbf{x}_i - \sum_{j=1}^{i-1} P_{\mathbf{y}_j} \mathbf{x}_i && (\mathbf{y}_i \perp \mathbf{y}_{i-1} \perp \dots \perp \mathbf{y}_1) \end{aligned}$$

Die \mathbf{y}_i müssen schließlich noch normiert werden.

Erhält man $\mathbf{y}_i = \mathbf{0}$, so ist \mathbf{y}_i zu eliminieren. Dann besteht eine lineare Abhängigkeit zwischen den Ausgangsvektoren.

Die Anzahl der von $\mathbf{0}$ verschiedenen Vektoren \mathbf{y}_i ist gleich der Dimension des von den \mathbf{x}_i aufgespannten Raumes.

Lineare Unabhängigkeit

Eine Menge von Vektoren $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$ heißen *linear abhängig*, falls es reelle Zahlen $\alpha_1, \alpha_2, \dots, \alpha_k$ gibt, die nicht alle gleich Null sind, so dass gilt:

$$\sum \alpha_i \mathbf{y}_i = \mathbf{0}$$

In diesem Falle lässt sich auch mindestens einer der Vektoren als Linearkombination der restlichen Vektoren darstellen.

Lässt sich die obige Gleichung nur erfüllen, wenn alle α_i gleich Null sind, so nennt man die Menge der Vektoren *linear unabhängig*.

Die lineare Unabhängigkeit bzw. lineare Abhängigkeit spielt in der multivariaten Statistik eine eminente Rolle.

8.2.7 Gleichungen und deren geometrische Interpretation

Eine Gleichung mit einer Unbekannten von der Form $ax = k$ repräsentiert einen Punkt auf einer durch x definierten Strahlengeraden.

Eine Gleichung mit 2 Unbekannten von der Form $a_1x_1 + a_2x_2 = k$ repräsentiert eine Gerade in der Ebene, die durch die Achsen x_1 und x_2 definiert ist (siehe Abbildung).

Eine Gleichung mit 3 Unbekannten repräsentiert eine Ebene in einem dreidimensionalen Raum.

Eine Gleichung mit n Unbekannten repräsentiert einen $(n-1)$ -dimensionalen Unterraum eines n -dimensionalen Raumes.

Schnittmengen von Unterräumen

In der Ebene schneiden sich 2 Geraden in einem Punkt, es sei denn, sie fallen zusammen oder sind parallel.

Im dreidimensionalen Raum schneiden sich zwei Ebenen in einer Geraden, es sei denn, sie fallen zusammen oder sind parallel.

Ebenso schneiden sich im dreidimensionalen Raum eine Ebene und eine Gerade in einem Punkt, es sei denn, die Gerade liegt in der Ebene oder ist parallel zu ihr.

Zwei Geraden schneiden sich im dreidimensionalen Raum normalerweise nicht.

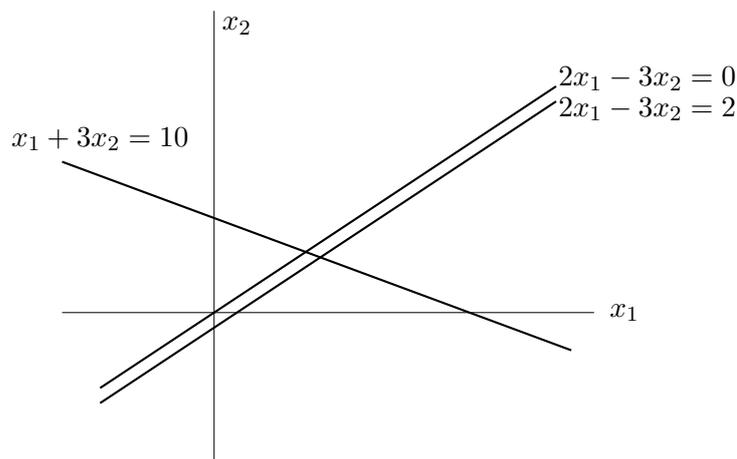


Abbildung 8.8: Gleichungen geometrisch betrachtet

Allgemein: abgesehen von Spezialfällen schneiden sich im n -dimensionalen Raum zwei Unterräume V_{n-p} und V_{n-q} in einem $V_{n-(p+q)}$.

Lösungsmenge von Gleichungen als Schnittmenge

Eine Gleichung mit n Unbekannten definiert einen $(n-1)$ -dimensionalen Unterraum in einem n -dimensionalen Raum.

Zwei Gleichungen mit n Unbekannten definieren einen $(n-2)$ -dimensionalen Unterraum in einem n -dimensionalen Raum (abgesehen von Sonderfällen).

Abgesehen von Sonderfällen definieren n Gleichungen mit n Unbekannten einen 0-dimensionalen Unterraum, d.h. einen Punkt in einem n -dimensionalen Raum. Dies ist gleichbedeutend mit einer eindeutigen Lösung des Gleichungssystems.

8.3 Matrizen

8.3.1 Definitionen

Eine *Matrix* ist ein System von $n \times m$ Elementen x_{ij} , die in einem rechteckigen Schema von n Zeilen und m Spalten angeordnet sind:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix}$$

Das Zahlenpaar (n, m) nennt man den *Typ* der Matrix. Ist $n = m$, so nennt man n auch die *Ordnung* der Matrix.

Die Elemente einer Matrix sind durch zwei Indizes identifiziert: den Zeilenindex und den Spaltenindex, wobei konventionsgemäß der erste Index der Zeilenindex ist.

Der Grenzfall einer Matrix mit nur einer Spalte stellt einen Spaltenvektor dar, eine solche mit nur einer Zeile einen Zeilenvektor.

Für die Benennung von Matrizen verwendet man üblicherweise fettgedruckte Großbuchstaben (z.B. \mathbf{X}). Um zusätzlich noch den Typ einer Matrix \mathbf{X} zu kennzeichnen, schreibt man auch $\mathbf{X}_{(n \times m)}$, $\mathbf{X}_{n \times m}$ oder $\mathbf{X}_{(n,m)}$.

Eine Matrix \mathbf{X} kann auch aufgefaßt werden als ein Zeilenvektor von m Spaltenvektoren oder als ein Spaltenvektor von n Zeilenvektoren.

8.3.2 Sonderformen von Matrizen

Quadratische Matrix: eine Matrix mit gleich vielen Zeilen wie Spalten.

Diagonalmatrix: eine Matrix, die außerhalb der Diagonale alles Nullen enthält.

Obere bzw. untere Dreiecksmatrix: eine Matrix, die unterhalb (bzw. oberhalb) der Diagonale lauter Nullen enthält.

Symmetrische Matrix: eine Matrix, die - an der Diagonale gespiegelt - gleich bleibt (d.h. es gilt: $x_{ij} = x_{ji}$ für alle i und j).

8.3.3 Spezielle Matrizen

Nullmatrix (Symbol: \mathbf{O}): alle Elemente sind gleich 0.

Einsmatrix: (Symbol: $\mathbf{1}$): alle Elemente sind gleich 1.

Skalarmatrix: alle Elemente sind identisch.

Einheitsmatrix: (Symbol: \mathbf{I}): eine Diagonalmatrix mit Einsen in der Diagonale.

8.3.4 Rechnen mit Matrizen

Gleichheit: zwei Matrizen \mathbf{A} und \mathbf{B} sind gleich, wenn sie vom gleichen Typ sind und in allen Elementen übereinstimmen:

$$a_{ij} = b_{ij} \quad \text{für alle } i \text{ und } j$$

Transposition: die Zeilen und Spalten werden vertauscht. Mit \mathbf{A}' bezeichnet man die Transponierte zu \mathbf{A} . Es gilt:

$$a'_{ij} = a_{ji} \quad \text{für alle } i \text{ und } j$$

Multiplikation einer Matrix und eines Skalars: die entsprechende Operation wird elementweise ausgeführt:

$$\mathbf{B} = c\mathbf{A} \quad \text{ist definiert durch} \quad b_{ij} = c a_{ij}$$

Addition und Subtraktion zweier Matrizen: die entsprechende Operation wird elementweise ausgeführt:

$$\mathbf{C} = \mathbf{A} \pm \mathbf{B} \quad \text{ist definiert durch} \quad c_{ij} = a_{ij} \pm b_{ij}$$

Die beiden Matrizen müssen vom gleichen Typ sein.

Manche Computerprogramme (z.B. SPSS, S-Plus) unterstützen noch eine in der Matrizenrechnung eigentlich nicht definierte Operation, nämlich die *Addition* bzw. *Subtraktion* einer *Matrix* und eines *Skalars*. Diese könnte auf folgende Weise definiert werden:

$$\mathbf{B} = \mathbf{A} \pm c \quad \text{definieren wir durch} \quad b_{ij} = a_{ij} \pm c$$

Rechnen mit Matrizen: Matrizenprodukt

Das Matrizenprodukt ist definiert durch die inneren Produkte der Zeilenvektoren der ersten Matrix mit den Spaltenvektoren der zweiten Matrix:

$$\mathbf{C} = \mathbf{AB} \quad \text{ist definiert durch} \quad c_{ij} = \sum_k a_{ik} b_{kj}$$

d.h. das Element c_{ij} der Produktmatrix \mathbf{C} ist gleich dem inneren Produkt der i -ten Zeile von \mathbf{A} mit der j -ten Spalte von \mathbf{B} .

Damit das Matrizenprodukt definiert ist, muss folgende Voraussetzung erfüllt sein: die Anzahl der Spalten von \mathbf{A} muß gleich der Anzahl der Zeilen von \mathbf{B} sein, d.h. für die Typen der beteiligten Matrizen gilt:

$$\mathbf{A}_{(n \times m)} \mathbf{B}_{(m \times p)} = \mathbf{C}_{(n \times p)}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2k} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ik} & \cdots & a_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kj} & \cdots & b_{kp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mj} & \cdots & b_{mp} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1j} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2j} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{ij} & \cdots & c_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nj} & \cdots & c_{np} \end{bmatrix}$$

8.3.5 Eigenschaften und Sätze

- Die Addition von Matrizen ist assoziativ und kommutativ:

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + \mathbf{B} + \mathbf{C} \quad \text{und} \quad \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

- Die skalare Multiplikation ist assoziativ und kommutativ:

$$\begin{aligned} c(d\mathbf{A}) &= (cd)\mathbf{A} = cd\mathbf{A} \\ c\mathbf{A} &= \mathbf{A}c \end{aligned}$$

- Das Matrizenprodukt ist assoziativ:

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{ABC}$$

- Das Matrizenprodukt ist *nicht kommutativ*, d.h. im allgemeinen gilt:

$$\mathbf{AB} \neq \mathbf{BA}$$

Es macht also normalerweise einen Unterschied, ob eine Matrix von links oder von rechts her mit einer anderen Matrix multipliziert wird (vorausgesetzt, die Matrizen sind überhaupt in beiden Reihenfolgen verknüpfbar).

- Für Matrizen gelten die Distributivgesetze:

$$\begin{aligned} c(\mathbf{A} + \mathbf{B}) &= c\mathbf{A} + c\mathbf{B} \\ (c + d)\mathbf{A} &= c\mathbf{A} + d\mathbf{A} \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{AC} + \mathbf{BC} \end{aligned}$$

- Die Nullmatrix ist das Nullelement der Matrizenmultiplikation:

$$\mathbf{AO} = \mathbf{OA} = \mathbf{O}$$

- Die Einheitsmatrix ist das Einselement der Matrizenmultiplikation:

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

- Die Transponierte der Transponierten ist die Originalmatrix:

$$(\mathbf{A}')' = \mathbf{A}$$

- Die Transponierte eines Matrizenprodukts ist das Produkt der Transponierten in umgekehrter Reihenfolge:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

Rechnen mit Matrizen: Kronecker-Produkt

Für eine beliebige $n \times m$ -Matrix \mathbf{A} und eine beliebige $r \times p$ -Matrix \mathbf{B} lassen sich weitere Produkte definieren, beispielsweise das Kronecker-Produkt $\mathbf{A} \otimes \mathbf{B}$. Dieses ergibt eine $nr \times mp$ -Matrix, bei der jedes Element von \mathbf{A} mit der Matrix \mathbf{B} multipliziert wird:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rp} \end{bmatrix} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix}$$

Das Kronecker-Produkt kommt unter anderem bei der Konstruktion von Designvariablen beim allgemeinen linearen Modell zum Einsatz. Das Kronecker-Produkt wird auch in der verallgemeinerten linearen Regressionsanalyse verwendet, um eine Kovarianzmatrix von korrelierten Störgrößen zu konstruieren, z.B. in Form einer blockdiagonalen Matrix.

8.3.6 Matrizeninversion

Eine quadratische Matrix \mathbf{A} heißt *invertierbar* oder *regulär*, falls eine Matrix \mathbf{A}^{-1} existiert, so daß gilt:

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

Die Matrix \mathbf{A}^{-1} heißt die *inverse Matrix* bzw. die *Inverse* zu \mathbf{A} .

Eine Matrix \mathbf{A} , die keine Inverse besitzt, heißt *nicht invertierbar* oder *singulär*.

Die Inverse einer Matrix spielt bei der Auflösung von linearen Gleichungssystemen eine wichtige Rolle:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

In Matrixschreibweise lautet dieses Gleichungssystem:

$$\mathbf{Ax} = \mathbf{b}$$

\mathbf{A} ist die quadratische Koeffizientenmatrix (deren Elemente sind Konstanten), \mathbf{x} der Vektor der Unbekannten und \mathbf{b} der Vektor der rechten Seiten (dessen Elemente sind ebenfalls Konstanten):

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Multipliziert man die Matrixgleichung $\mathbf{Ax} = \mathbf{b}$ von links her mit der Inversen von \mathbf{A} (falls sie existiert), so erhält man:

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b}$$

$\mathbf{A}^{-1}\mathbf{A}$ ist aber gleich \mathbf{I} , und $\mathbf{I}\mathbf{x}$ ist gleich \mathbf{x} , so daß sich schließlich als Lösung für das obige Gleichungssystem ergibt:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Die Berechnung der Inversen ist bei größeren Matrizen ziemlich rechenaufwendig. Zur Illustration soll die Inverse der folgenden 2×2 -Matrix berechnet werden:

$$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

Die Matrix \mathbf{A}^{-1} enthält 4 unbekannte Größen, die wir y_{11} , y_{12} , y_{21} und y_{22} nennen wollen:

Die Gleichung $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ ergibt in unserem Fall:

$$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Durch Ausmultiplizieren erhält man das folgende Gleichungssystem (4 Gleichungen mit 4 Unbekannten):

$$4y_{11} + 2y_{21} = 1; \quad 4y_{12} + 2y_{22} = 0; \quad 2y_{11} + 6y_{21} = 0; \quad 2y_{12} + 6y_{22} = 1$$

Dessen Lösung ist:

$$y_{11} = 6/20; \quad y_{12} = -2/20; \quad y_{21} = -2/20; \quad y_{22} = 4/20$$

d.h.

$$\mathbf{A}^{-1} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$$

Probe:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \frac{1}{20} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Allgemein gilt für die Inverse einer 2×2 -Matrix \mathbf{A} :

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Der Ausdruck $a_{11}a_{22} - a_{12}a_{21}$ ist die sog. *Determinante* von \mathbf{A} . Sie muss verschieden von Null sein, damit die Inverse von \mathbf{A} existiert.

8.3.7 Rang einer Matrix

Die maximale Anzahl linear unabhängiger Spaltenvektoren (Zeilenvektoren) einer Matrix \mathbf{A} heißt *Spaltenrang* (*Zeilenrang*) von \mathbf{A} .

Es gilt: Spaltenrang von $\mathbf{A} =$ Zeilenrang von \mathbf{A} .

Daraus folgt auch:

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}')$$

Daher spricht man einfach nur vom Rang der Matrix \mathbf{A} und bezeichnet diese Zahl mit $\text{rang}(\mathbf{A})$.

Eine quadratische Matrix der Ordnung n hat *vollen Rang*, wenn $\text{rang}(\mathbf{A}) = n$ ist. Nur dann ist sie auch invertierbar.

Übrigens gilt:

$$\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}'\mathbf{A}) = \text{rang}(\mathbf{A}\mathbf{A}')$$

8.3.8 Spur und Norm einer Matrix

Die Summe der Diagonalelemente einer $n \times n$ -Matrix \mathbf{A} nennt man ihre *Spur* (engl. *trace*):

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

Es gelten folgende Sätze:

$$\begin{aligned}\text{sp}(\mathbf{AB}) &= \text{sp}(\mathbf{BA}) \\ \text{sp}(\mathbf{A}') &= \text{sp}(\mathbf{A}) \\ \text{sp}(\mathbf{A}'\mathbf{A}) &= \text{sp}(\mathbf{AA}')$$

Die euklidische *Norm* einer $n \times m$ -Matrix \mathbf{A} ist die Quadratwurzel der Summe der Quadrate aller ihrer Elemente und wird mit $\|\mathbf{A}\|$ bezeichnet:

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}$$

Mit Hilfe der Spur kann man die Norm wie folgt definieren:

$$\|\mathbf{A}\| = \{\text{sp}(\mathbf{A}'\mathbf{A})\}^{1/2} = \{\text{sp}(\mathbf{AA}')\}^{1/2}$$

Unter anderem gilt:

$$\begin{aligned}\|\mathbf{A}\| &\geq 0 \\ \|\mathbf{A}\| &= 0 \quad \text{genau dann, wenn } \mathbf{A} = \mathbf{0}\end{aligned}$$

8.3.9 Determinante einer Matrix

Jeder $n \times n$ -Matrix \mathbf{A} ist ihre Determinante $|\mathbf{A}|$ (auch mit $\det(\mathbf{A})$ bezeichnet) zugeordnet. Dies ist eine reelle Zahl, die in folgender Weise definiert ist:

$$|\mathbf{A}| = \sum_{i_1, i_2, \dots, i_n} (-1)^k a_{1i_1} a_{2i_2} \dots a_{ni_n}$$

Dazu sind zunächst alle möglichen Produkte von jeweils n Elementen von \mathbf{A} zu bilden, bei denen jede Zeile und jede Spalte von \mathbf{A} durch genau ein Element vertreten ist (die Indizes i_1, i_2, \dots, i_n stellen jeweils eine Permutation der Zahlen $1, 2, \dots, n$ dar). Der Exponent k ist die Anzahl der Inversionen, die der jeweiligen Permutation entsprechen: das Vorzeichen eines Produkts hängt also davon ab, ob die jeweilige Permutation gerade oder ungerade ist. Schließlich werden alle auf diese Weise erhaltenen $n!$ Produkte aufsummiert.

Determinantensätze

- $|\mathbf{A}| = |\mathbf{A}'|$.
- $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$.
- Ist \mathbf{A} eine Dreiecksmatrix (oder Diagonalmatrix), so ist ihre Determinante das Produkt der Diagonalelemente.

- Sind zwei Zeilen oder Spalten von \mathbf{A} gleich, so ist $|\mathbf{A}| = 0$.
Allgemeiner: ist eine Zeile (Spalte) von \mathbf{A} gleich einer Linearkombination der anderen Zeilen (Spalten), so ist $|\mathbf{A}| = 0$.
- Die Vertauschung von zwei Zeilen (Spalten) von \mathbf{A} ändert $|\mathbf{A}|$ um den Faktor -1 .
- Die Multiplikation einer einzelnen Zeile (Spalte) von \mathbf{A} mit einem Faktor $c \neq 0$ ändert deren Determinante um den Faktor c .
- Die Multiplikation der Matrix \mathbf{A} mit einem Faktor $c \neq 0$ ändert deren Determinante um den Faktor c^n :
 $|c\mathbf{A}| = c^n |\mathbf{A}|$.

Wichtig ist vor allem der folgende Satz:

- Die Determinante einer Matrix \mathbf{A} ist genau dann von 0 verschieden, wenn \mathbf{A} invertierbar ist.
- D.h. umgekehrt:
Ist $|\mathbf{A}| = 0$, so ist \mathbf{A} nicht invertierbar, d.h. dann existiert \mathbf{A}^{-1} nicht.

Entwicklungssatz

Sei A_{ij} die zum Element a_{ij} gehörige Unterdeterminante (das ist jene $(n-1)$ -reihige Determinante, die man aus \mathbf{A} nach Streichen der i -ten Zeile und j -ten Spalte gewinnt), so gilt:

$$|\mathbf{A}| = \sum_{j=1}^n a_{ij} (-1)^{(i+j)} A_{ij}$$

für jedes i („Entwicklung nach der i -ten Zeile“) bzw.

$$|\mathbf{A}| = \sum_{i=1}^n a_{ij} (-1)^{(i+j)} A_{ij}$$

für jedes j („Entwicklung nach der j -ten Spalte“).

Der Entwicklungssatz ist für die praktische Berechnung von $|\mathbf{A}|$ sehr bedeutsam.

Die Ausdrücke $(-1)^{(i+j)} A_{ij}$ werden auch als *Cofaktoren* bezeichnet.

Determinanten von Matrizen bis zur Ordnung 3

$$|a| = a$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

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

8.3.10 Lineare Abbildungen

Durch die Vorschrift $y = f(x)$ wird einer reellen Zahl x eine reelle Zahl y zugeordnet. Man sagt, y sei eine Funktion von x . Die Vorschrift f gibt im speziellen Fall dann an, wie die Zahl y aus der Zahl x zu bilden ist. Allgemein versteht man unter einer Abbildung die eindeutige Zuordnung der Elemente einer Menge A zu Elementen einer Menge B . Wie für reelle Zahlen kann man auch für Vektoren den Funktions- bzw. Abbildungsbegriff einführen.

Durch die Vorschrift $\mathbf{y} = f(\mathbf{x})$ wird einem Vektor \mathbf{x} aus einem Vektorraum \mathbb{R}^n ein bestimmter Vektor \mathbf{y} aus einem Vektorraum \mathbb{R}^m zugeordnet bzw. er wird in ihn überführt (transformiert)¹. Man sagt, der Vektor \mathbf{x} wird in den Vektor \mathbf{y} abgebildet. Den Vektor \mathbf{y} nennt man den Bildvektor oder das Bild von \mathbf{x} , den Vektor \mathbf{x} das Urbild und f die Abbildung.

Wir wollen nun die Klasse der sog. *linearen* Abbildungen betrachten. Eine Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt linear, falls für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und alle $a \in \mathbb{R}$ die folgenden Bedingungen erfüllt sind:

1. $f(a\mathbf{x}) = af(\mathbf{x})$
2. $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$

d.h. die Proportionalität der Vektoren \mathbf{x} und $a\mathbf{x}$ soll auch in ihren Bildern erhalten bleiben und das Bild einer Summe von Vektoren soll stets gleich der Summe der Bilder sein.

Eine besonders einfache Abbildung ist die folgende:

$$f(\mathbf{x}) = c\mathbf{x}$$

mit beliebigem, aber festem reellen c . Diese Abbildung nennt man auch lineare Verstreckung. Insbesondere liefert $c = 1$ die *identische Abbildung* und $c = 0$ die *Nullabbildung*.

Lineare Abbildungen lassen sich nach bestimmten Gesetzen auch miteinander verknüpfen. Die wichtigste Verknüpfung besteht in der Hintereinanderschaltung, etwa einer ersten Abbildung f und einer zweiten Abbildung g . Das Ergebnis ist eine neue Abbildung h , die man auch das Produkt der beiden Einzelabbildungen nennt:

$$h(\mathbf{x}) = f(g(\mathbf{x}))$$

Das Produkt von Abbildungen ist abgesehen von Ausnahmefällen nicht kommutativ. $h = fg$ bedeutet: zuerst Anwenden der Abbildung g und dann Anwenden der Abbildung f auf das Ergebnis. Bei $h' = gf$ ist die Reihenfolge umgekehrt. Die beiden Abbildungen h und h' sind im allgemeinen verschieden. Beide sind aber linear.

Unter bestimmten Bedingungen ist eine lineare Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ umkehrbar, d.h. es gibt eine sog. *inverse* Abbildung g derart, dass die Produkte fg und gf die identische Abbildung liefern. Die Abbildung g macht die Abbildung f rückgängig und umgekehrt. Die zu f inverse Abbildung wird üblicherweise mit f^{-1} bezeichnet.

Zu einer linearen Abbildung f gibt es genau dann eine inverse Abbildung f^{-1} , wenn die Bildvektoren $f(\mathbf{b}_1), f(\mathbf{b}_2), \dots, f(\mathbf{b}_n)$ einer Basis $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ des Vektorraumes \mathbb{R}^n linear unabhängig sind, d.h. selbst wieder eine Basis von \mathbb{R}^n darstellen.

Schreiben wir nämlich einen beliebigen Vektor $\mathbf{x} \in \mathbb{R}^n$ in der Form:

$$\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_n\mathbf{b}_n$$

so geht er bei der Abbildung f in den Bildvektor \mathbf{y} über mit

$$\mathbf{y} = f(\mathbf{x}) = x_1f(\mathbf{b}_1) + x_2f(\mathbf{b}_2) + \dots + x_nf(\mathbf{b}_n)$$

Sind nun die Bilder der Basisvektoren $f(\mathbf{b}_k)$ linear abhängig, so gibt es Komponenten x_i derart, dass $f(\mathbf{x}) = \mathbf{0}$ bei $\mathbf{x} \neq \mathbf{0}$. Dann kann es aber keine Abbildung g geben mit $gf(\mathbf{x}) = \mathbf{x} \neq \mathbf{0}$. Sind aber die $f(\mathbf{b}_k)$ linear unabhängig, so spannt die Gesamtheit der Bilder $f(\mathbf{b}_k)$ auch wieder den ganzen \mathbb{R}^n auf, wenn \mathbf{x} den ganzen \mathbb{R}^n überstreicht. g ist dann jene eindeutige Abbildung, welche jedes der Bilder $f(\mathbf{b}_k)$ in das dazugehörige \mathbf{b}_k zurückverwandelt, d.h. gf ist die identische Abbildung (dafür schreibt man auch: $gf = 1$). Dann gilt aber auch $fg = 1$.

Eine Abbildung f , deren Bilder $f(\mathbf{b}_k)$ einer Basis linear abhängig sind, heißt ausgeartet oder *singulär*. Der lineare Raum der Bilder ist von geringerer Dimension als der lineare Raum der Urbilder, man sagt: der Raum

¹Wir wollen uns auf Vektorräume über dem Körper der reellen Zahlen beschränken.

\mathbb{R}^n wird auf einen in \mathbb{R}^n eingebetteten Unterraum geringerer Dimension abgebildet. Die Abbildung kann daher nicht mehr eindeutig und umkehrbar sein, da zu einem Bild mehrere Urbilder, d.h. „Originale“ gehören. Beispielsweise ist die Projektion des \mathbb{R}^3 auf eine Ebene ($=\mathbb{R}^2$) eine solche singuläre Abbildung.

Sind bei einer Abbildung die Bilder einer Basis hingegen linear unabhängig, so nennt man die Abbildung nicht ausgeartet, nicht singulär oder *regulär*. Nur eine solche Abbildung ist umkehrbar.

Eine lineare Abbildung $\mathbf{y} = f(\mathbf{x})$ lässt sich immer durch eine Matrixgleichung $\mathbf{y} = \mathbf{A}\mathbf{x}$ darstellen. Die Matrix \mathbf{A} ist die *Transformationsmatrix*, deren Elemente sich auf die fest gewählte Basis beziehen. Die Abbildung f liegt nämlich fest, wenn die Bilder $f(\mathbf{b}_k)$ der Basisvektoren \mathbf{b}_k gegeben sind. Die Spaltenvektoren \mathbf{a}_k der Transformationsmatrix \mathbf{A} sind nichts anderes als die Bilder der Basisvektoren.

Unter den linearen Abbildungen spielt die Klasse der *orthogonalen* Abbildungen eine besondere Rolle. Diese haben die Eigenschaft, dass sie die Längen und Winkel der Vektoren eines Vektorsystems nicht verändern. Dies lässt sich wie folgt ausdrücken: Eine Abbildung f heißt orthogonal, falls das Skalarprodukt zweier beliebiger Vektoren \mathbf{x} und \mathbf{y} aus dem Vektorraum gleich dem Skalarprodukt ihrer Bildvektoren $f(\mathbf{x})$ und $f(\mathbf{y})$ ist.

Die Spaltenvektoren einer orthogonalen Matrix \mathbf{B} bilden ein sog. *Orthonormalsystem* (eine sog. *Orthonormalbasis*). Es gilt:

$$\mathbf{B}'\mathbf{B} = \mathbf{I} \quad \text{bzw.} \quad \mathbf{B}^{-1} = \mathbf{B}'$$

Wegen $|\mathbf{B}'\mathbf{B}| = |\mathbf{B}'||\mathbf{B}| = |\mathbf{B}|^2 = |\mathbf{I}| = 1$ ist die Determinante einer orthogonalen Matrix \mathbf{B} entweder $+1$ oder -1 . Im ersten Fall spricht man von *eigentlichen* orthogonalen Abbildungen. Diese stellen reine *Drehungen* im \mathbb{R}^n dar. Im zweiten Fall spricht man von *uneigentlichen* orthogonalen Abbildungen. Diese repräsentieren *Drehungen mit Spiegelung*.

Das Produkt zweier Orthogonalmatrizen, d.h. die Hintereinanderausführung zweier orthogonaler Abbildungen, ist stets wieder eine Orthogonalmatrix, d.h. eine orthogonale Abbildung.

Durch eine orthogonale Abbildung wird ein System orthogonaler Einheitsvektoren, d.h. eine Orthonormalbasis, wieder in ein solches überführt.

Beispiele:

Drehung um den Winkel ϕ in der Ebene

$$\mathbf{A} = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}$$

$$|\mathbf{A}| = \cos^2 \phi + \sin^2 \phi = 1$$

Drehung um den Winkel ϕ mit Spiegelung in der Ebene

$$\mathbf{B} = \begin{bmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{bmatrix}$$

$$|\mathbf{B}| = -\cos^2 \phi - \sin^2 \phi = -1$$

Für die Matrix \mathbf{B} gilt zudem: $\mathbf{B}' = \mathbf{B}$, woraus folgt:

$$\mathbf{B}'\mathbf{B} = \mathbf{B}\mathbf{B} = \mathbf{B}^2 = \mathbf{I}$$

d.h. die zweimalige Ausführung der linearen Abbildung mit dieser Matrix führt zum Ausgangssystem zurück, kommt also der identischen Abbildung gleich. Derartige Matrizen nennt man *involutorisch*.

8.3.11 Das Eigenwertproblem

Gegeben sei eine quadratische Matrix \mathbf{A} der Ordnung n . Gesucht werden Zahlen λ und Vektoren \mathbf{v} , so daß gilt:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

bzw.

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

Lösung des Eigenwertproblems

Die geforderte Beziehung gilt für $\mathbf{v} \neq \mathbf{0}$ nur, wenn die „charakteristische Gleichung“ gilt:

$$|\mathbf{A} - \lambda \mathbf{I}| = 0$$

Die Entwicklung dieser Determinante führt auf ein Polynom n -ten Grades in λ . Dieses Polynom wird „charakteristisches Polynom“ genannt.

Die Aufgabe besteht zunächst darin, die Nullstellen $\lambda_1, \lambda_2, \dots, \lambda_n$ des charakteristischen Polynoms zu finden. Dies sind die *Eigenwerte* von \mathbf{A} .

Nun setzt man einen Eigenwert nach dem anderen in die Gleichung $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}$ ein und erhält jeweils ein homogenes Gleichungssystem von n Gleichungen mit n Unbekannten $v_{1i}, v_{2i}, \dots, v_{ni}$.

Die Lösungen dieses Gleichungssystems sind die *Eigenvektoren* von \mathbf{A} zum entsprechenden Eigenwert.

Da die Gleichungen aber linear abhängig sind, sind diese Lösungen nicht eindeutig, es gibt vielmehr unendlich viele Lösungen.

Um eine eindeutige Lösung zu erzielen, werden die Eigenvektoren normiert. Zuweilen werden sie auch auf eine Länge gebracht, die jeweils der Größe des zugehörigen Eigenwerts entspricht.

Übrigens kann man die obige Gleichung in folgender Weise schreiben, um alle Eigenwerte und Eigenvektoren einzubeziehen:

$$\mathbf{A}\mathbf{V} = \mathbf{A}\mathbf{V}$$

Dabei ist \mathbf{A} die Diagonalmatrix mit den Eigenwerten und \mathbf{V} eine Matrix, deren Spalten die dazugehörigen Eigenvektoren darstellen.

Beispiel

Es sollen die Eigenwerte und Eigenvektoren der folgenden Matrix berechnet werden:

$$\mathbf{A} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

Gesucht werden also Zahlen λ und Vektoren \mathbf{v} , so daß gilt:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

bzw.

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

Die charakteristische Gleichung ist:

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 4-\lambda & 2 \\ 2 & 6-\lambda \end{vmatrix} = 0$$

Die Entwicklung dieser Determinante führt auf eine quadratische Gleichung:

$$(4 - \lambda)(6 - \lambda) - 2 \times 2 = \lambda^2 - 10\lambda + 20 = 0$$

Die Wurzeln dieser Gleichung sind:

$$\lambda_1 = 5 + \sqrt{5} = 7.236068 \quad \lambda_2 = 5 - \sqrt{5} = 2.763932$$

Dies sind die beiden Eigenwerte von \mathbf{A} .

Um den Eigenvektor $\mathbf{v}_1 = [v_{11} \ v_{21}]'$ zum ersten Eigenwert zu erhalten, setze ich in die Gleichung

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

den Wert von λ_1 ein:

$$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = 7.236068 \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$$

und erhalte das folgende homogene Gleichungssystem:

$$\begin{aligned} 4v_{11} + 2v_{21} &= \lambda_1 v_{11} = 7.236068v_{11} \\ 2v_{11} + 6v_{21} &= \lambda_1 v_{21} = 7.236068v_{21} \end{aligned}$$

Also:

$$\begin{aligned} -3.236068v_{11} + 2v_{21} &= 0 \\ 2v_{11} - 1.236068v_{21} &= 0 \end{aligned}$$

Aus beiden Gleichungen ergibt sich als Lösung:

$$v_{11} = 0.618034v_{21}$$

Die Lösung ist nicht eindeutig. Ein Eigenvektor wäre z.B.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0.618034 \end{bmatrix}$$

Wir normieren nun \mathbf{v}_1 und bekommen eine eindeutige Lösung:

$$\mathbf{v}_1 = \begin{bmatrix} 0.8506508 \\ 0.5257311 \end{bmatrix}$$

Um den Eigenvektor zum zweiten Eigenwert zu erhalten, wiederholen wir das Verfahren, indem wir λ_2 in die Gleichung

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$

einsetzen.

Der normierte zweite Eigenvektor ist

$$\mathbf{v}_2 = \begin{bmatrix} -0.5257311 \\ 0.8506508 \end{bmatrix}$$

Übrigens sind die beiden Eigenvektoren sind zueinander orthogonal:

$$\mathbf{v}_1' \mathbf{v}_2 = 0$$

Einige Sätze

- Eine Matrix der Ordnung n hat genau n (nicht notwendigerweise verschiedene) Eigenwerte.
- Ist r der Rang der Matrix, so sind genau r Eigenwerte von Null verschieden.
- Bei einer Dreiecks- bzw. Diagonalmatrix sind die Diagonalelemente die Eigenwerte.
- Die Eigenwerte der Transponierten einer Matrix sind gleich den Eigenwerten der Originalmatrix.
- Die Eigenwerte der Inversen einer Matrix sind die Kehrwerte der Eigenwerte der Originalmatrix.
- Es gibt einen Zusammenhang zwischen den Eigenwerten, der Determinante und der Spur einer Matrix:
Sind λ_i die Eigenwerte der Matrix \mathbf{A} , so gilt

$$\prod \lambda_i = |\mathbf{A}|$$

sowie

$$\sum \lambda_i = \text{sp}(\mathbf{A})$$

8.3.12 Quadratische Formen, Definitheit

Wir definieren eine lineare Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}$:

Eine Funktion $q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$ mit einer symmetrischen Matrix \mathbf{A} und einem Vektor $\mathbf{x} \neq \mathbf{0}$ heißt *quadratische Form*.

Die Matrix \mathbf{A} heißt Matrix der Form.

Eine Funktion der Form $\mathbf{x}'\mathbf{A}\mathbf{y}$ mit einer symmetrischen Matrix \mathbf{A} und zwei Vektoren $\mathbf{x} \neq \mathbf{0}$ und $\mathbf{y} \neq \mathbf{0}$ heißt *bilineare Form*.

Eine Form bzw. Matrix heißt *positiv definit*, wenn gilt

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \quad \text{für alle } \mathbf{x} \neq \mathbf{0}$$

Sie heißt *positiv semidefinit*, wenn gilt

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \quad \text{für alle } \mathbf{x} \neq \mathbf{0}$$

Übrigens gilt:

Eine Matrix ist positiv definit bzw. semidefinit, wenn alle Eigenwerte positiv bzw. alle von Null verschiedenen Eigenwerte positiv sind.

8.3.13 Zerlegung einer symmetrischen Matrix

Quadratische Matrizen können auf verschiedene Weise zerlegt („faktoriert“), d.h. als Produkt bestimmter Matrizen dargestellt werden. Zerlegungen dieser Art sind z.B.:

- Singulärwertzerlegung
- Cholesky-Zerlegung
- LU-Zerlegung

Singulärwertzerlegung

Für jede reelle symmetrische $n \times n$ -Matrix \mathbf{A} existiert eine orthonormale Matrix \mathbf{U} und eine $n \times n$ -Diagonalmatrix $\mathbf{\Lambda}$, so dass gilt:

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

Umgekehrt kann man die Matrix \mathbf{A} folgendermaßen auf *Diagonalgestalt* bringen (*diagonalisieren*):

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

Die Spalten der Matrix \mathbf{U} enthalten die Eigenvektoren von \mathbf{A} und die λ_i deren Eigenwerte.

Die obige Gleichung kann man auch so schreiben:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{U}'$$

Damit haben wir die Möglichkeit, die Wurzel von \mathbf{A} zu definieren:

$$\mathbf{A}^{1/2} = \mathbf{U}\mathbf{\Lambda}^{-1/2}$$

Cholesky-Zerlegung

Die Cholesky-Zerlegung bestimmt zu einer symmetrischen positiv definiten Matrix \mathbf{A} eine untere Dreiecksmatrix \mathbf{L} , so dass

$$\mathbf{L}\mathbf{L}' = \mathbf{A}$$

Ausführlich geschrieben:

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{n1} \\ 0 & l_{22} & \cdots & l_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Durch sukzessives Ausmultiplizieren und Einsetzen² erhält man nacheinander die Werte der Komponenten der Matrix \mathbf{L} .

Mit dem Cholesky-Verfahren kann man auch auf recht einfachem Wege das lineare Gleichungssystem $\mathbf{A}\mathbf{x} = \mathbf{b}$ lösen.

8.3.14 Eigenschaften symmetrischer Matrizen

- Die Eigenwerte einer symmetrischen reellen Matrix sind reell.
- Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal.
- Eine symmetrische Matrix der Ordnung n besitzt stets ein kartesisches Koordinatensystem aus n Eigenvektoren.

Darüberhinaus haben *Kovarianzmatrizen* (Korrelationsmatrizen) noch einige spezielle Eigenschaften, die erwähnenswert sind:

- Eine Kovarianzmatrix ist positiv definit oder positiv semidefinit, d.h. besitzt keine negativen Eigenwerte.
- Die Eigenvektoren einer Kovarianzmatrix sind zueinander orthogonal.
- Diese bilden die Hauptachsen der Schnittellipsen (bzw. deren mehrdimensionalen Verallgemeinerung) der zwei- bzw. mehrdimensionalen Normalverteilung im Variablenraum.
- Die Eigenwerte sind proportional den Längen dieser Hauptachsen.
- Der Eigenvektor zum ersten, d.h. größten Eigenwert entspricht der Richtung der größten Varianz im Variablenraum. Die nächsten Eigenvektoren entsprechen der Richtung der größten Varianz im jeweils verbleibenden orthogonalen Unterraum.

²Skalarprodukt 1. Zeile mit 1. Spalte ergibt $l_{11}^2 = a_{11}$, also $l_{11} = \sqrt{a_{11}}$ usw.

Index

- Abbildung
 - identische, 84
 - lineare, 83
 - Null-, 84
 - orthogonale, 85
 - reguläre, 85
 - singuläre, 84
- Abhängigkeit
 - lineare, 76
- Akaike, 31
- ALM, 53
- AR(1)-Prozess, 47
- Auspartialisierung, 24
- Ausreißer, 21

- Bartlett, 34, 51, 57, 67
- Basis, 75
- Bernoulliverteilung, 44
- Betagewichte, 29
- Bilineare Form, 88
- Binomialverteilung, 42–44, 46
- Bonferroni-Korrektur, 5
- Box M, 52

- Cholesky-Zerlegung, 89
- Clusteranalyse, 56
- Codierung
 - Dummy-, 54
 - Effekt-, 54
- Compound Symmetry, 5, 47
- Cook's Abstand, 22, 32
- Cosinussatz, 73

- Deleted Residuals, 23, 32
- Designmatrix, 36
- Determinante, 81, 82
- Deviance, 43
- Devianz, 43
- DFBETAS, 23, 32
- DFFITS, 22, 32
- Diagonalgestalt, 88
- Diagonalisieren, 61, 88
- Diskriminanzanalyse, 56
 - quadratische, 59
- Diskriminanzfaktoren, 56
- Diskriminanzfunktionen, 56
- Diskriminanzraum, 57
- Diskriminanzscores, 58
- Distanz
 - euklidische, 49
 - Mahalanobis, 49
- Drehung, 85
- Dummycodierung, 36, 54
- Durbin-Watson Statistik, 23

- Effektcodierung, 38, 54
- Eigenvektor, 86
- Eigenwert, 86
- Eigenwertkriterium, 63
- Einfachstruktur, 64
- Einflussreiche Datenpunkte, 31
- Einheitsvektor, 74
- Einsvektor, 74
- Entwicklungssatz, 83
- Exponentialfamilie, 42

- Faktor
 - kanonischer, 66
- Faktorenanalyse, 62
 - exploratorische, 62
 - konfirmatorische, 62
- Faktorladungen, 58, 64, 67
- Faktorscores, 61, 62
- Form
 - bilineare, 88
 - quadratische, 88
- Fundamentaltheorem, 63
- Funktion
 - kanonische, 66

- Gammaverteilung, 42, 43
- GEE, 46
- Generalized Estimation Equations, 46
- Gleichung
 - charakteristische, 86
- GLM, 41
- Gram-Schmidt, 76

- Hat-Matrix, 26
- Hauptachsentransformation, 61

- Hauptkomponenten, 61
- Hauptkomponentenanalyse, 60
- Hebelwert, 22
- Homogenität, 59
- Homoskedastizität, 18, 23, 42, 44
- Hotelling's T^2 , 48, 50
- Hotelling-Lawley Spur, 51

- Informationskriterium, 31
- Inverse, 80

- Kanonische Korrelationsanalyse, 66
- Kommunalität, 63
- Konfidenzband, 21, 28
- Konfidenzintervall, 21, 28, 51
- Kontraste, 39
 - Abweichungs-, 39
 - Differenz-, 39
 - einfache, 39, 40
 - Helmert-, 39
 - polynomiale, 39
- Korrelation
 - partielle, 24
 - semipartielle, 24
- Korrelationskoeffizient
 - kanonischer, 67
 - multipler, 28
- Korrelationsmatrix, 89
 - reduzierte, 63
- Kovarianzanalyse, 41
- Kovarianzmatrix, 14, 89
 - AR(1)-Prozess, 46
 - Compound Symmetry, 46
 - Fehler-, 46
 - unstrukturierte, 46
- Kovarianzparameter, 45
- Kreuzproduktmatrix, 14
- Kronecker, 80
- Kurvenanpassung, 17

- Ladungsmatrix, 61, 64
- Lineare Kontraste, 51
- Linkfunktion, 43, 46
 - kanonische, 43–45
 - natürliche, 43
- LISREL, 62
- Loess-Kurven, 24
- Loglineares Modell, 45
- Lowess-Regression, 24

- Mahalanobis-Distanz, 49, 59
- Mallow's C_p -Statistik, 31
- MANOVA, 48

- Matrix, 77
 - Diagonal-, 78
 - Dreiecks-, 78
 - Einheits-, 78
 - Eins-, 78
 - inverse, 80
 - invertierbare, 80
 - Null-, 78
 - positiv definite, 88
 - positiv semidefinite, 88
 - quadratische, 78
 - reguläre, 80
 - singuläre, 80
 - Skalar-, 78
 - symmetrische, 78
- Matrixwurzel, 88
- Matrizenprodukt, 79
- Mauchley, 53
- Messwiederholungen, 52
- Minkowski-Metrik, 49
- Modelle
 - GEE-, 46
 - gemischte lineare, 45
- Multikollinearität, 30

- Norm, 82
- Normalverteilung, 10, 42, 43, 46
 - inverse, 42, 43
- Normalverteilungplot, 23, 32
- Normieren, 72
- Nullvektor, 72, 75

- Oblimin, 65
- OLS, 18
- OLS-Regression, 44
- Orthogonalbasis, 75
- Orthogonalisierungsverfahren, 76
- Orthonormalbasis, 75, 85
- Orthonormalsystem, 85
- Ortsvektor, 72
- Overdispersion, 45

- Parameter
 - Dispersions-, 42
 - kanonischer, 42, 43
- Partial Least Squares, 60
- Pillai-Bartlett Spur, 51
- PLS, 60
- Poissonregression, 44
- Poissonverteilung, 42–44, 46
- Polynom
 - charakteristisches, 86
 - positiv definit, 88

- positiv semidefinit, 88
- Potenzgesetz, 7
- PRESS, 23, 32
- Produkt
 - inneres, 70
 - Kronecker, 80
- Projektion
 - orthogonale, 74
- Promax, 65

- QQ-Plot, 23, 32
- Quadratische Form, 88

- Rang, 81
 - voller, 81
- Rao, 51
- Regression
 - gewichtete kleinste Quadrate, 23
 - nichtlineare, 17
 - polynomiale, 31
 - robuste, 24
- Regression auf die Hauptkomponenten, 60
- Regressionsanalyse
 - einfache lineare, 17
 - multivariate multiple, 33, 53
 - univariate multiple, 24
- Regressionskoeffizient
 - standardisierter, 29
 - unstandardisierter, 29
- REML, 45
- Residuen
 - standardisierte, 22
 - studentisierte, 22
- Responsefunktion, 43
- Richtungscosinus, 74
- Rotation, 58, 64, 68
- Roy, 51
- Roy-Bose, 51

- Scatterplotmatrix, 14
- Scheffé, 51
- Screetest, 63
- Singulärwertzerlegung, 61, 88
- Skalar, 69
- Skalarprodukt, 70
- Skalenparameter, 42
- Spaltenrang, 81
- Spezifität, 63
- Sphärizität, 5, 53
- Sphärizitätstest, 53
- Spiegelung, 85
- Spinplot, 14
- Spur, 82

- Hotelling-Lawley, 51
- Pillai-Bartlett, 51
- SSCP-Matrix, 14, 54
- Standardregressionsgewichte, 21
- Strukturgleichungsmodell, 62
- Strukturkoeffizienten, 58, 67
- Strukturmodell, 37, 48
- Suppressorvariable, 29

- Toleranz, 30
- Toleranzintervall, 28
- Transformation, 23
- Transformationsmatrix, 52, 61, 64, 68, 85
- Transposition, 78
- Tukey, 51

- U-Statistik, 34
- U-Verteilung, 35
- Unabhängigkeit
 - lineare, 76
- Underdispersion, 45
- Unterraum, 75

- Variablenraum, 75
- Varianz-Inflations-Faktor, 30
- Varianzanalyse, 36, 40
- Varianzfunktion, 43
- Varianzkomponenten, 45
- Varimax-Kriterium, 64, 68
- Vektor, 69
 - normierter, 72
 - orthogonaler, 72, 76
- Vektorabstand, 72
- Vektorlänge, 72
- Vektorraum, 74
- Vektorrichtung, 72
- Vektorunterraum, 75
- Vektorwinkel, 72
- Vergleich
 - geplanter, 51
- Versuchspersonenraum, 75
- Vorhersageintervall, 28

- Wilks' Lambda, 34, 51, 52

- Zeilenrang, 81