Hermann Singer

Multivariate Statistik

korrigierte 1. Auflage

6. August 2019

Der vorliegende Kurs wird am *Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung* angeboten.

Autor:
Univ.-Prof. Dr. Hermann Singer

Kursergänzende Lehrbücher und Multimedia-Übungen siehe Homepage des Lehrstuhls für angewandte Statistik und Methoden der empirischen Sozialforschung (`www.fernuni-hagen.de/ls_statistik/`)

Inhaltsverzeichnis

Vorwort

Der hier vorgelegte Kurs *Multivariate Statistik* ist als B-Modul im Bachelor-Studiengang Wirtschaftswissenschaft konzipiert.

Er setzt Kenntnisse der deskriptiven Statistik sowie der Wahrscheinlichkeitsrechnung und Inferenzstatistik voraus, wie sie in Bachelor-Studiengängen vermittelt werden (etwa im Modul *Grundlagen der Wirtschaftsmathematik und Statistik* des Bachelor-Studiengangs Wirtschaftwissenschaft der FernUniversität). Weiterhin sind Kenntnisse der Wirtschaftsmathematik erforderlich (Analysis und Matrix-Algebra). Methoden, die in diesen Kursen nicht enthalten sind, lassen sich anhand der Anhänge erarbeiten.

Eine Aufgabensammlung sowie Lösungen zu den Übungen sind in einem separaten Aufgabenheft enthalten.

Es ist nicht möglich, die Verfahren der multivariaten Analyse in ihrer Breite und statistischen Tiefe in einem Einführungskurs zu behandeln. Aus der Vielfalt der Literatur möchte ich die Bücher von Mardia et al. (1979); Nagl (1992); Fahrmeir et al. (1996, 2009) und Handl und Niermann (2010) zum vertieften Studium empfehlen.

Die Multimedia-Ausstattung des Kurses wird einerseits durch Standardsoftware wie SAS/JMP und SPSS, andererseits durch vom Lehrstuhl entwickelte interaktive Mathematica-Applets realisiert, die mit dem frei erhältlichen Mathematica-Player (bzw. Wolfram CDF-Player) in Realzeit abgespielt werden können (siehe http://www.fernuni-hagen.de/ls_statistik/lehre/).

Dies verbindet die Kombination von berufsbefähigenden Softwarekenntnissen (Marketing, Banken, etc.) mit speziellen Lern-Applets (etwa simultane Konfidenzintervalle, Simulation von normalverteilten Daten etc.).

Matrixberechnungen können auch mit der frei verfügbaren Statistik-Software R durchgeführt werden (http://www.r-project.org/).

Ich danke meinen Mitarbeitern vom Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung für vielfältige Unterstützung.

Mein besonderer Dank gilt Frau Dipl. Kffr. Marina Lorenz für die Erstellung der Übungsaufgaben und Lösungen, die sorgfältige Lektüre des Manuskripts und unzählige Verbesserungsvorschläge.

Herrn Thomas Feuerstack vom Zentrum für Medien und IT (ZMI) danke ich ganz herzlich für die Entwicklung einer LaTeX-Umgebung im FernUni-Design.

Hagen, im September 2012

Vorwort zur korrigierten 1. Auflage

Leider haben sich einige Fehler eingeschlichen, die nun in der korrigierten 1. Auflage beseitigt wurden. Eine aktuelle Fehlerliste ist auf der Seite des Lehrstuhls zu finden. Ich danke allen, die Ihre Korrekturen geschickt haben.

Bitte beachten Sie auch die Hinweise unter:

http://www.fernuni-hagen.de/ls_statistik/kurse/k00883.shtml.

Hagen, im Juni 2013

Kapitel 1

Fallstudien

1.1 Übungen mit SPSS und SAS/JMP

Damit die dargestellte Theorie für den Leser auch praktisch umgesetzt werden kann, wurde im Text eine Vielzahl von Beispielen integriert, die mit Hilfe der leicht bedienbaren Softwarepakete SPSS (Statistical Package for the Social Sciences) und SAS/JMP nachvollzogen werden können. Alle Verfahren sind mit Hilfe einer Menü-Steuerung zugänglich und können ohne textbezogene Kommandos durch Maus-Klicks bedient werden. JMP unterstützt auch eine interaktive Bearbeitung bereits bestehender Outputs (in diesen können zusätzliche Analysen durchgeführt und graphische Elemente, z.B. Konfidenzellipsen, hinzugefügt werden).

SPSS
SAS/JMP

Die selbsttätige Durchführung der Computerübungen ist für das Begreifen des Lehrstoffs unbedingt erforderlich. Es reicht nicht aus, den Text nur zu lesen. Sie sollten die in den Graphiken und Tabellen dargestellten Resultate auch selbst erzeugen. Die Datensätze und Programme sind in den Verzeichnissen `SPSS` und `JMP` der beigefügten Kurs-CD enthalten.

SPSS kann für die Studierenden der FernUniversität über ein Softwareportal des ZMI (Zentrum für Medien und IT) bezogen werden.
Bitte beachten Sie die Hinweise auf der Homepage des Lehrstuhls:

http://www.fernuni-hagen.de/ls_statistik/

Für die Nutzung von JMP verweise ich auf die Homepage von SAS/JMP:

http://www.jmp.com/

1.2 Einführung

Multivariate Statistik hat das Ziel, mehrere Variablen gleichzeitig zu betrachten. Im Gegensatz zur univariaten Analyse kann hier das gleichzeitige (simultane) Auftreten von Merkmalskombinationen untersucht werden. Zum Beispiel wird die Verteilung eines Merkmals getrennt nach Vorliegen eines anderen Kriteriums bestimmt und graphisch dargestellt.

Beispiel 1.1 (Credit-Scoring)

Credit-Scoring

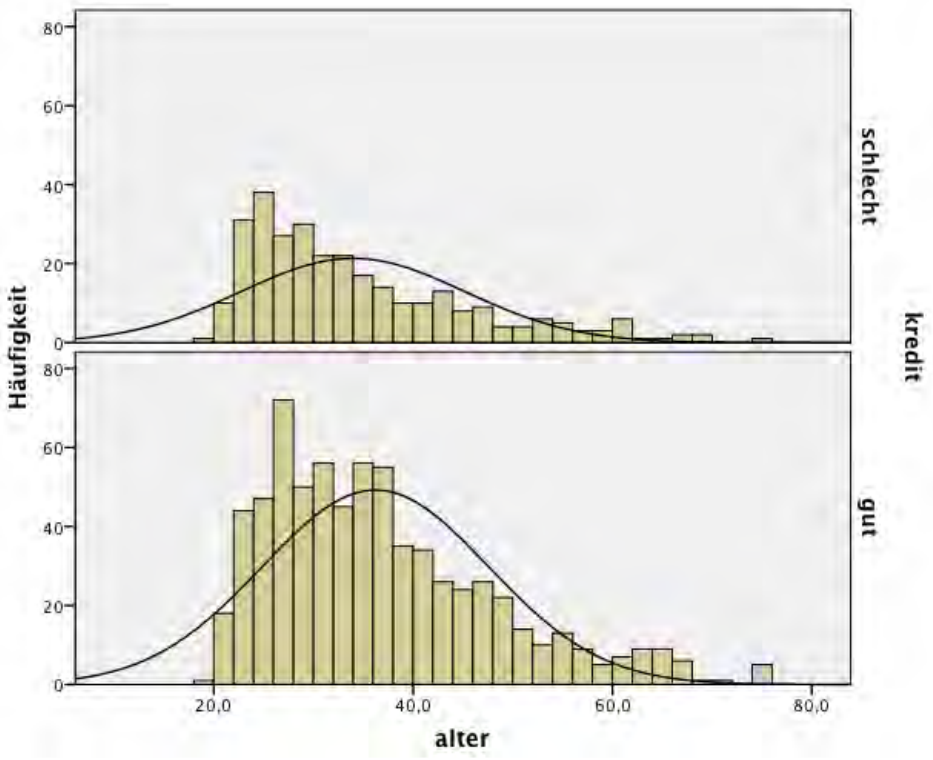
Im Rahmen des sog. Credit-Scorings wird nach Kunden-Merkmalen (etwa Alter, Bürge, Laufzeit etc.) gesucht, die sich in den Gruppen der guten bzw. schlechten Schuldner unterscheiden. Dann kann die Vergabe neuer Kredite an diesen Kriterien orientiert werden, um die Wahrscheinlichkeit zu minimieren, daß der Kredit nicht ordnungsgemäß zurückgezahlt wird (Abb. 1.1; Datensätze `Kreditrisiko.LMU.sav` bzw. `Kreditrisiko.LMU.jmp`). [1]

In den Graphiken wird also die Altersverteilung $f(\texttt{Alter}|\texttt{Kredit})$ in den Gruppen gute/schlechte Schuldner getrennt dargestellt. Wenn zwischen den Variablen `Alter` und `Kredit` ein Zusammenhang besteht, sollten sich die Alters-Verteilungen in den Gruppen `Kredit` = schlecht und `Kredit` = gut unterscheiden, etwa in der Verteilungsform, im Mittelwert oder in der Streuung. Ob ein statistisch bedeutsamer Unterschied vorliegt, muß getestet werden (etwa mit einem t-Test).

Umgekehrt kann die Wahrscheinlichkeit, daß eine Person ein guter oder schlechter Schuldner ist, als Funktion des Alters dargestellt werden. Dies führt auf eine Fragestellung, die im Rahmen der kategorialen Regression bzw. der Diskriminanzanalyse beantwortet werden kann. Meist werden weitere Variablen hinzugenommen, etwa Laufzeit des Kredits, Vorliegen eines Bürgen, Geschlecht etc.

■

[1] Vgl. auch
http://www.stat.uni-muenchen.de/service/datenarchiv/kredit/kredit.html



alter

kredit	N	Mittelwert	Standardabweichung	Schiefe	Kurtosis
schlecht	300	33,960	11,2252	1,155	,786
gut	700	36,220	11,3474	,988	,611
Insgesamt	1000	35,542	11,3527	1,025	,621

Abbildung 1.1: Altersverteilung in den Gruppen: schlechte und gute Kredite. Zum Vergleich: Normalverteilungen mit den gruppenspezifischen Mittelwerten und Standardabweichungen.

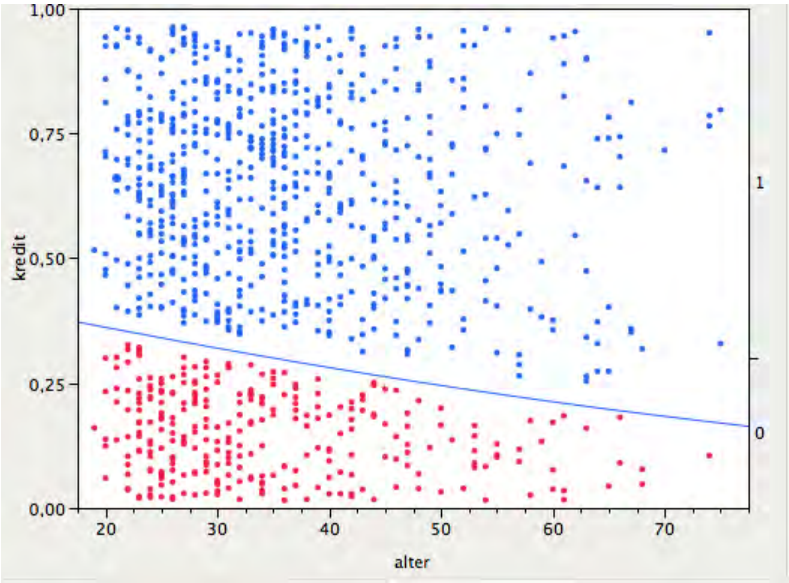


Abbildung 1.2: Logistische Regression: Geschätzte Wahrscheinlichkeit für schlechte und gute Kredite (rot/blau) als Funktion des Alters. Mit steigendem Alter sinkt die Wahrscheinlichkeit, dass der Kredit ausfällt, also $p(\texttt{Kredit} = \textit{schlecht}|\texttt{Alter})$.

Beispiel 1.2 (OECD-Daten)

Ein interessanter Datensatz, der auf OECD-Erhebungen beruht[2], soll im folgenden diskutiert werden. Er ist auf der Kurs-CD enthalten (als Excel-, SPSS- und JMP-Datei), kann jedoch auch im Internet als Excel-Tabelle gefunden werden (siehe http://oecdbetterlifeindex.org/). Die Excel-Tabelle kann in SPSS importiert und anschließend als SPSS-Datensatz (.sav) wieder gespeichert werden (siehe Abb. 1.3).

```
SPSS/Datei/Öffnen/Daten/Format Excel auswählen
```

Abb. 1.4 zeigt die sogenannte Daten- und Variablenansicht des Datensatzes `BetterLifeIndex.sav`. Der Datensatz wird mit Hilfe des Menüs

```
SPSS/Datei/Öffnen/Daten
```

[2]Organisation for Economic Cooperation and Development

Abbildung 1.3: Excel-Datensatz `BetterLifeIndex.xls` in SPSS importieren. Das Datenblatt `score1` enthält nur die Variablen-Namen und die Daten (vom Autor hinzugefügt).

in SPSS geöffnet.

Die ersten beiden Spalten (Variablen) sind nominal skaliert (Abkürzung und ausführliche Ländernamen), während die anderen Variablen quantitativer Natur sind. Es handelt sich um die Größen:

1. Abbreviation
2. Country
3. Rooms per person
4. Dwelling without basic facilities
5. Household disposable income
6. Household financial wealth
7. Employment rate
8. Long-term unemployment rate
9. Quality of support network
10. Educational attainment
11. Students reading skills
12. Air pollution
13. Consultation on rule-making

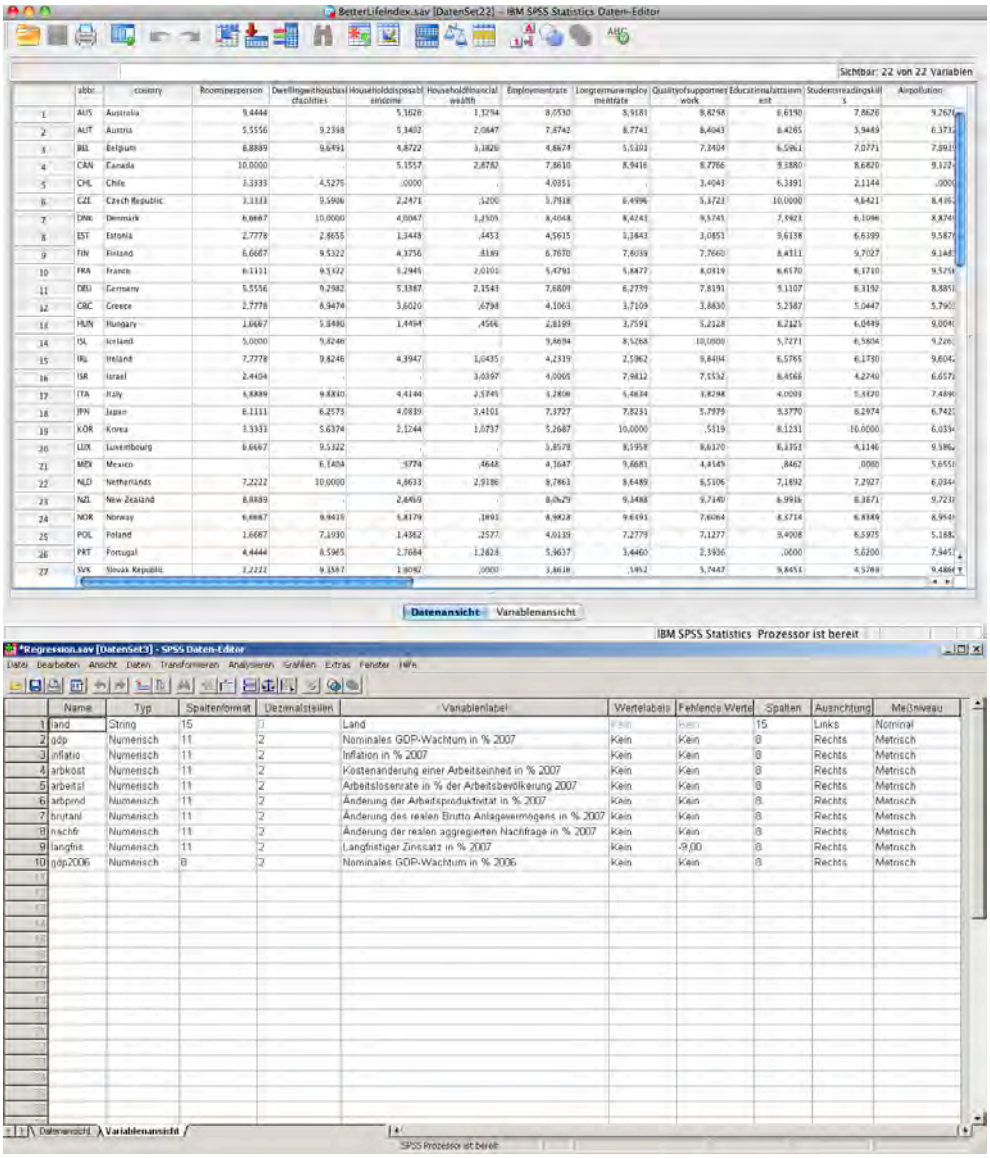


Abbildung 1.4: SPSS-Datensatz. Daten- und Variablenansicht. Datensatz `BetterLifeIndex.sav`

OECD Better Life Initiative

YOUR BETTER LIFE INDEX - List of indicators and definitions

TOPICS	INDICATORS	DEFINITIONS
Housing	Rooms per person	It signals whether the persons occupying a dwelling are living in crowded conditions. It is measured as the number of rooms in a dwelling divided by the number of persons living in the dwelling.
	Dwelling without basic facilities	It provides an assessment of the potential deficits and shortcomings of accommodation focusing on facilities for personal hygiene. One basic facility is considered here: a lack of indoor flushing toilet (measured as the percentage of dwellings not having indoor flushing toilet for the sole use of their household).
Income	Household disposable income	It includes income from work, property, imputed rents attributed to home owners and social benefits in cash, net of direct taxes and social security contributions paid by households; it also includes the social transfers in kind, such as education and health care, that households receive from governments. Income is measured net of the depreciation of capital goods that households use in production.
	Household financial wealth	It consists of various financial assets owned by households (e.g. cash, bonds and shares) net of all types of financial liabilities.
Jobs	Employment rate	It is the share of the working age population (people aged from 15 to 64 in most OECD countries) who are currently employed in a paid job. Employed persons are those aged 15 and over who declare having worked in gainful employment for at least one hour in the previous week, following the standard ILO definition.
	Long-term unemployment rate	It is the number of persons who have been unemployed for one year or more as a share of the labour force. Unemployed persons are those who are currently not working but are willing to do so and actively searching for jobs.
Community	Quality of support network	It shows the proportion of the population reporting that they have relatives, friends, or neighbours they can count on to help if they were in trouble.
Education	Educational attainment	It profiles the education of the adult population as captured through formal educational qualifications. Educational attainment is measured as the percentage of the adult population (15 to 64 years of age) holding at least an upper secondary degree, as defined by the OECD-ISCED classification.
	Students reading skills	It measures the capacity of students near the end of compulsory education to understand, use, reflect on and engage with written texts in order to achieve their own goals, to develop their knowledge and potential. This indicator comes from the 2009 edition of OECD's Programme for International Student Assessment (PISA), which focused on reading.
Environment	Air pollution	It refers to the population-weighted average concentrations of fine particles (PM10) in the air we breathe (measured in micro grams per cubic meter); data refer to residential areas of cities larger than 100,000 inhabitants. Particulate matters consist of small liquid and solid particles floating in the air, and include sulphate, nitrate, elemental carbon, organic carbon matter, sodium and ammonium ions in varying concentrations. Of greatest concern to public health are the particles small enough to be inhaled into the deepest parts of the lung.
Governance	Voter turnout	It measures the extent of electoral participation in major national elections. Only the number of votes casted over the population registered to vote are considered. The voting-age population is generally defined as the population aged 18 or more, while the registered population refers to the population listed on the voters' register. The number of votes casted are gathered from national statistics offices and national electoral management bodies.
	Consultation on rule-making	It describes the extent to which formal consultation processes are built-in at key stages of the design of regulatory proposals, and whether mechanisms exist for the outcome of that consultation to influence the preparation of draft primary laws and subordinate regulations. This indicator is a composite index aggregating various information on the openness and transparency of the consultation process used when designing regulations.
Health	Life expectancy	It is the standard measure of the length of people's life. Life-expectancy measures how long on average people could expect to live based on the age specific mortality rates currently prevailing. Life-expectancy can be [illegible]
	Self-reported health	It is based on questions of the type: "How is your health in general?". Data are based on general household surveys or on more detailed Health Interviews undertaken as part of the official surveys in various countries.
Life Satisfaction	Life Satisfaction	It measures overall life satisfaction as perceived by individuals. Life satisfaction measures how people evaluate their life as a whole rather than their current feelings. It is measured via the Cantril Ladder (also referred to as the Self-Anchoring Striving Scale), which asks people to rate how they value their life in terms of the best possible life (10) through to the worst possible life (0). The score for each country is calculated as the mean
Safety	Homicide rate	It measures the number of police-reported intentional homicides reported each year, per 100,000 people. The data come from the United Nations Office on Drugs and Crime (UNODC) and are based on national data collected from law enforcement, prosecutor offices, and ministries of interior and justice, as well as Interpol, Eurostat and regional crime prevention observatories.
	Assault rate	It is based on the percentage of people who declare that they have been victim of an assault crime in the last 12 months. The data presented here are drawn from the Gallup World Poll.
Work-life balance	Employees working very long hours	It shows the proportion of employees who usually work for pay for more than 50 hours per week. The data exclude self-employed workers who are likely to chose deliberately to work long hours.
	Employment rate of women with children	It shows the employment rate of mothers with children aged 6-14 years. The employment rate for all women of roughly the same age group is also shown to provide contextual information on labour market access for women overall (OECD, 2010).
	Time devoted to leisure and personal care	It presents data from national time use surveys on the hours devoted to leisure and personal care in a typical day.

Abbildung 1.5: OECD-Datensatz. Themen und ihre Indikatoren mit Definitionen.

14. Voter turnout
15. Life expectancy
16. Self-reported health
17. Life Satisfaction
18. Homicide rate
19. Assault rate
20. Employees working very long hours
21. Employment rate of women with children
22. Time devoted to leisure and personal care

Eine ausführliche Beschreibung der Variablen ist auf der oben genannten Internet-Seite und in der Excel-Tabelle enthalten (vgl. Abb. 1.5).

Um einen Überblick zu gewinnen, wird zuerst eine univariate Analyse durchgeführt, etwa durch eine graphische Darstellung der Häufigkeitsverteilung sowie durch die Berechnung von Statistiken wie Mittelwert, Median, Standardabweichung etc. Die Statistiken und Histogramme können durch die Menübefehle

```
SPSS/Analysieren/Deskriptive Statistiken/Häufigkeiten/...
```

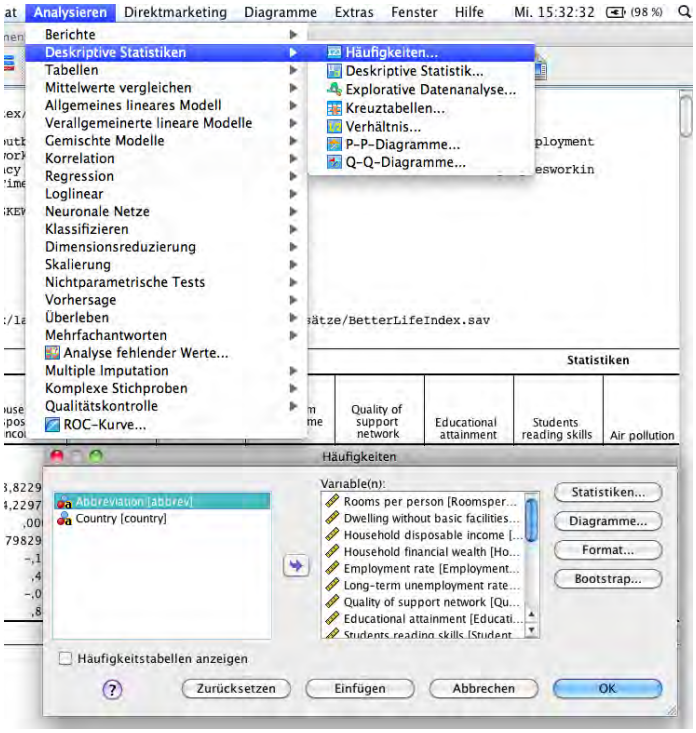
erzeugt werden (siehe Abb. 1.6, oben).

Im Rahmen einer multivariaten Analyse interessiert man sich eher für Zusammenhänge zwischen den Variablen. Grundlegend hierfür ist eine Tabelle von Korrelationskoeffizienten bzw. von Streudiagrammen.

Abb. 1.7 zeigt einen Ausschnitt der Korrelationstabelle aller Variablen. Signifikante Werte sind durch Sterne gekennzeichnet (* = signifikant auf dem $\alpha = 5\%$-Niveau). Informativ sind auch Streudiagramme, die in Matrixform ausgegeben werden (Abb. 1.8).

Eine übersichtliche Darstellung des Korrelationsmusters ist die Farbmatrix der Korrelationen und deren Clusterbildung durch Umordnen der Zeilen und Spalten nach Größe der Ähnlichkeit (Abb. 1.9). In SPSS kann eine Dendrogramm der Ähnlichkeit von Variablen erzeugt werden (siehe Abb. 1.23).

■



Statistiken

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth	Employment rate	Long-term unemployment rate	Quality of support network	Educational attainment	Students reading skills	Air pollution	Consultation on rule-making	Voter turnout	Life expectancy	Self-reported health	Life Satisfaction	Homicide rate	Assault rate	Employees working very long hours	Employment rate of women with children	Time devoted to leisure and personal care
N	Gültig	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
	Fehlend	2	4	4	5	0	1	0	0	0	0	0	0	0	0	0	0	0	3	5	9
Mittelwert		5,180	8,351	3,823	1,736	5,996	6,656	6,657	7,074	5,981	7,752	5,561	5,12	6,154	6,447	6,347	8,162	7,952	8,245	6,752	6,182
Median		5,556	9,532	4,230	1,283	5,911	7,278	7,580	7,758	6,172	8,348	5,461	5,51	6,978	6,826	6,935	8,793	8,284	8,837	6,930	6,007
Modus		6,667	10,00	,000[a]	,000[a]	,000[a]	,000[a]	,000[a]	,000[a]	,000[a]	,000[a]	7,368	0[a]	6,923	8,447	4,52[a]	8,966	6,42[a]	,000[a]	,000[a]	,000[a]
Standardabweichung		2,510	2,478	1,798	1,323	2,262	2,706	2,634	2,660	2,007	2,060	2,703	2,622	2,775	2,529	2,693	2,034	1,926	1,954	1,914	2,099
Schiefe		-,022	-1,95	-,154	,761	-,271	-1,05	-,959	-1,40	-,566	-1,73	-,234	-,027	-1,09	-,926	-,744	-2,74	-2,46	-2,91	-1,58	-,863
Standardfehler der Schiefe		,414	,427	,427	,434	,403	,409	,403	,403	,403	,403	,403	,403	,403	,403	,403	,403	,403	,421	,434	,464
Kurtosis		-,673	3,642	-,026	-,211	-,059	,302	,234	1,575	1,594	4,627	-,723	-1,07	,149	,569	-,234	8,152	8,281	10,35	4,422	2,155
Standardfehler der Kurtosis		,809	,833	,833	,845	,788	,798	,788	,788	,788	,788	,788	,788	,788	,788	,788	,788	,788	,821	,845	,902

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

Histogramm

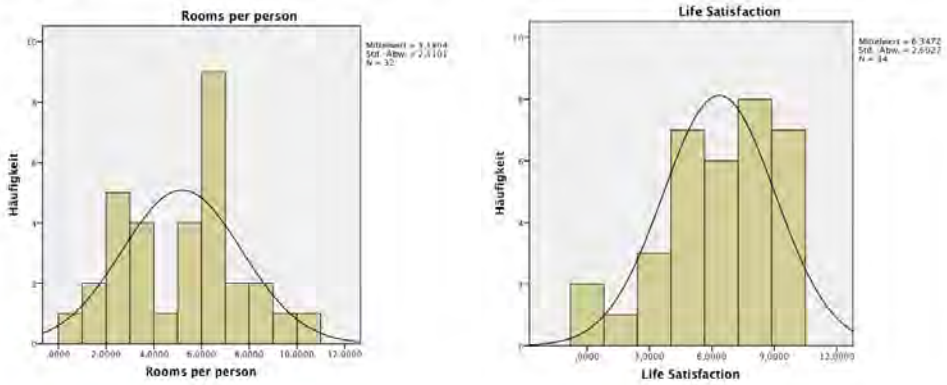
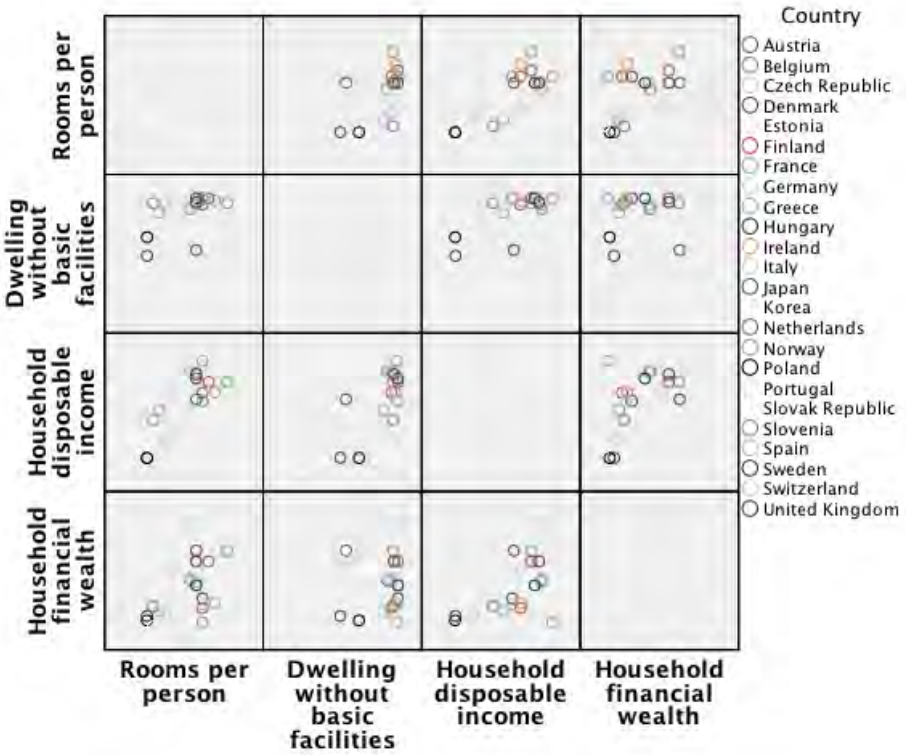


Abbildung 1.6: OECD-Datensatz. Auswahlmenü für Häufigkeiten (univariate Statistiken). In der Schaltfläche `Statistiken...` wurden verschiedene Lage- und Streuungsparameter sowie Schiefe und Kurtosis ausgewählt. Außerdem wurden Histogramme mit Normalverteilungskurve angefordert (Schaltfläche `Diagramme...`).

Korrelationen

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth	Employment rate	Long-term unemployment rate	Quality of support network	Educational attainment	Students reading skills	Air pollution	Consultation on rule-making	Voter turnout	Life expectancy	Self-reported health	Life Satisfaction	Homicide rate	Assault rate	Employees working very long hours	Employment rate of women with children	Time devoted to leisure and personal care
Rooms per person	Korrelation nach Pearson	1	,64**	,69**	,422*	,62**	,320	,69**	,092	,50**	,440*	,285	,317	,65**	,57**	,70**	,261	,242	,373*	,413*	,161
	Signifikanz (2-seitig)		,000	,000	,028	,000	,079	,000	,615	,004	,012	,114	,077	,000	,001	,000	,148	,182	,046	,029	,462
	N	32	28	28	27	32	31	32	32	32	32	32	32	32	32	32	32	32	29	28	23
Dwelling without basic facilities	Korrelation nach Pearson	,64**	1	,72**	,318	,59**	,075	,75**	,271	,221	,47**	,239	,074	,64**	,441*	,59**	,48**	,347	,65**	,53**	,459*
	Signifikanz (2-seitig)	,000		,000	,114	,001	,699	,000	,148	,240	,009	,204	,697	,000	,015	,001	,007	,060	,000	,005	,032
	N	28	30	27	26	30	29	30	30	30	30	30	30	30	30	30	30	30	27	26	22
Household disposable income	Korrelation nach Pearson	,69**	,72**	1	,71**	,58*	,267	,59**	,194	,374*	,440*	,162	,421*	,65**	,59**	,62**	,428*	,461*	,165	,201	,391
	Signifikanz (2-seitig)	,000	,000		,000	,001	,162	,001	,304	,042	,015	,392	,021	,000	,001	,000	,018	,010	,403	,326	,059
	N	28	27	30	28	30	29	30	30	30	30	30	30	30	30	30	30	30	28	26	24
Household financial wealth	Korrelation nach Pearson	,422*	,318	,71**	1	,395*	,304	,286	,154	,211	-,022	-,120	,214	,53**	,375*	,428*	,105	,191	-,093	,254	,090
	Signifikanz (2-seitig)	,028	,114	,000		,034	,108	,133	,424	,272	,910	,537	,264	,003	,045	,020	,587	,322	,646	,221	,682
	N	27	26	28	29	29	29	29	29	29	29	29	29	29	29	29	29	29	27	25	23
Employment rate	Korrelation nach Pearson	,62**	,59**	,58**	,395*	1	,55**	,64**	,396*	,45**	,386*	,300	,112	,61**	,387*	,63**	,351*	,326	,47**	,85**	,163
	Signifikanz (2-seitig)	,000	,001	,001	,034		,001	,000	,021	,007	,024	,085	,530	,000	,024	,000	,042	,060	,007	,000	,435
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
Long-term unemployment rate	Korrelation nach Pearson	,320	,075	,267	,304	,55**	1	,241	,125	,191	-,096	,337	,142	,347*	,356*	,59**	-,019	,059	-,184	,364	-,140
	Signifikanz (2-seitig)	,079	,699	,162	,108	,001		,177	,487	,287	,596	,055	,432	,048	,042	,000	,918	,745	,332	,052	,504
	N	31	29	29	29	33	33	33	33	33	33	33	33	33	33	33	33	33	30	29	25
Quality of support network	Korrelation nach Pearson	,69**	,75**	,59**	,286	,64**	,241	1	,382*	,253	,52**	,252	,163	,51**	,61**	,73**	,381*	,377*	,432*	,60**	,277
	Signifikanz (2-seitig)	,000	,000	,001	,133	,000	,177		,026	,149	,002	,151	,356	,002	,000	,000	,026	,028	,015	,001	,180
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
Educational attainment	Korrelation nach Pearson	,092	,271	,194	,154	,396*	,125	,382*	1	,46**	,280	,153	-,172	,109	,002	,271	,261	,56**	,49**	,56**	,185
	Signifikanz (2-seitig)	,615	,148	,304	,424	,021	,487	,026		,006	,108	,386	,331	,539	,991	,121	,136	,001	,005	,002	,375
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
Students reading skills	Korrelation nach Pearson	,50**	,221	,374*	,211	,45**	,191	,253	,46**	1	,45**	,373*	,076	,396*	,105	,181	,52**	,70**	,453*	,51**	,347
	Signifikanz (2-seitig)	,004	,240	,042	,272	,007	,287	,149	,006		,007	,030	,671	,020	,553	,307	,002	,000	,011	,005	,090
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
Air pollution	Korrelation nach Pearson	,440*	,47**	,440*	-,022	,386*	-,096	,52**	,280	,45**	1	,272	-,075	,185	,287	,227	,389*	,425*	,341	,458*	,251
	Signifikanz (2-seitig)	,012	,009	,015	,910	,024	,596	,002	,108	,007		,120	,671	,296	,100	,197	,023	,012	,060	,012	,226
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25
Consultation on rule-making	Korrelation nach Pearson	,285	,239	,162	-,120	,300	,337	,252	,153	,373*	,272	1	-,259	,082	,155	,199	,164	,345*	,084	,217	-,331
	Signifikanz (2-seitig)	,114	,204	,392	,537	,085	,055	,151	,386	,030	,120		,140	,645	,380	,260	,355	,046	,655	,258	,106
	N	32	30	30	29	34	33	34	34	34	34	34	34	34	34	34	34	34	31	29	25

Abbildung 1.7: Ausschnitt der Korrelationstabelle aller Variablen.



Korrelationen

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth
Rooms per person	Korrelation nach Pearson	1	,638**	,687**	,422*
	Signifikanz (2-seitig)		,000	,000	,028
	N	32	28	28	27
Dwelling without basic facilities	Korrelation nach Pearson	,638**	1	,725**	,318
	Signifikanz (2-seitig)	,000		,000	,114
	N	28	30	27	26
Household disposable income	Korrelation nach Pearson	,687**	,725**	1	,710**
	Signifikanz (2-seitig)	,000	,000		,000
	N	28	27	30	28
Household financial wealth	Korrelation nach Pearson	,422*	,318	,710**	1
	Signifikanz (2-seitig)	,028	,114	,000	
	N	27	26	28	29

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

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Abbildung 1.8: Streudiagramm-Matrix von 4 Variablen.

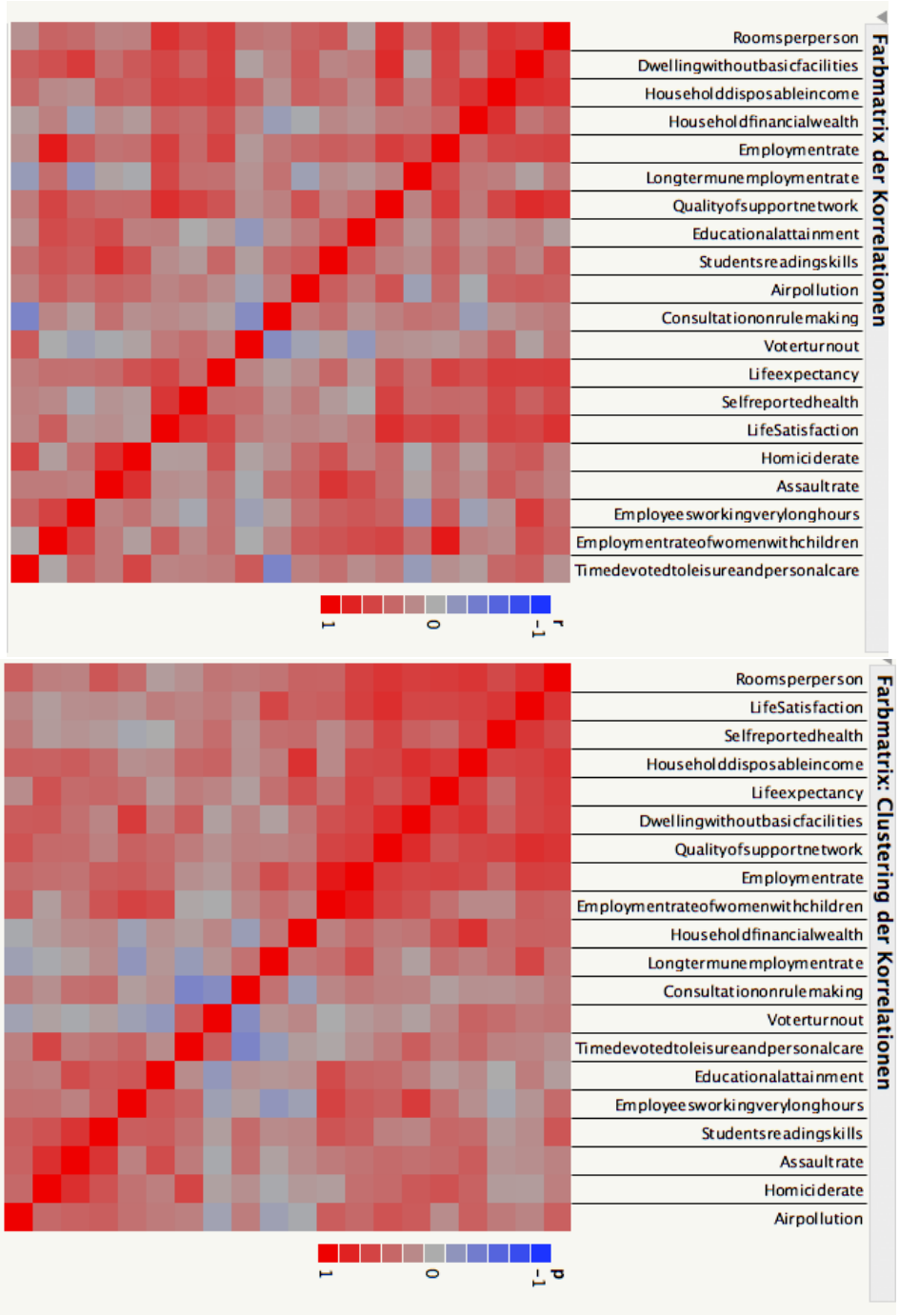


Abbildung 1.9: SAS/JMP: Farbmatrix der Korrelationen und Cluster (diagonales Ordnen).

Der Stoff wird in Aufgabe 1.1 vertieft (siehe separates Aufgabenheft).

1.3 Graphische Zusammenhänge zwischen Variablen

Im Gegensatz zur *univariaten* Statistik, bei der gleichzeitig nur *eine* Variable betrachtet wird (etwa ihre Verteilung, Mittelwert, Standardabweichung etc.), werden in der *multivariaten* Statistik mehrere Variablen *simultan* untersucht. Diese Betrachtungsweise führt über die bloße Wiederholung univariater Analysen hinaus, da auch die Zusammenhänge *zwischen* den Variablen analysiert werden. Im allgemeinen werden zuerst die Zusammenhänge graphisch dargestellt, wobei das sog. Streudiagramm (vgl. Abb. 1.8, oben) besonders einfach zu interpretieren ist. Man trägt die einzelnen Datenpunkte (N = Stichprobengröße)

Streudiagramm

$$(x_n, y_n), n = 1, ..., N \tag{1.1}$$

in ein $x - y$-Koordinatensystem ein, wobei dann, ähnlich wie auf einer Schießscheibe, ein Muster zu sehen ist. Wenn die Variablen einen deterministischen Zusammenhang aufweisen, etwa $y = a + bx$ oder $y = x^2$, würde man Muster sehen, die den üblichen Funktionsdarstellungen in der Mathematik ähneln. Bei empirischen Daten sind jedoch immer zufällige Einflußgrößen überlagert, die den Zusammenhang maskieren (etwa Rauschen bei Fernsehbildern, Meßfehler bei der Datenerhebung), oder es fehlt eine systematischer Zusammenhang.

Im allgemeinen hat man nicht nur 2, sondern p Variablen, sodaß die Zusammenhänge zwischen

$$(x_{n1}, x_{n2}, ..., x_{np}), n = 1, ..., N \tag{1.2}$$

dargestellt werden müssen, am besten in Form einer Tabelle von Streudiagrammen, wobei in der i-ten Zeile und der j-ten Spalte der Graphik das Streudiagramm zwischen den Zufalls-Variablen X_i und X_j mit den Meßwerten (Realisationen)

Matrix-Streudiagramm

$$(x_{ni}, x_{nj}), n = 1, ..., N \tag{1.3}$$

aufgetragen ist (Abb. 1.8).

Man kann auch versuchen, den 4-dimensionalen Datensatz anders zu visualisieren, etwa in Form einer 3-dimensionalen Punktewolke, die perspektivisch in 2 Dimensionen abgebildet ist (Abb. 1.10). Inwieweit dies zu einem verbesserten Verständnis für den Datensatz führt, mag dahingestellt sein. Programme wie SAS/JMP bieten die Möglichkeit, eine interaktive Rotation der Daten-Wolke durchzuführen (Abb. 1.10, unten) sowie Dichtekonturen (Ellipsoide bei Normalverteilung oder nichtparametrische Konturen) einzuzeichnen.

Es ist sinnvoll, die Daten in Form einer Tabelle aufzulisten, wobei üblicherweise jede Zeile einer Untersuchungseinheit (Person, Land, Firma, Zeitpunkt etc.) entspricht. Die Spalten enthalten die Variablen. Die Daten $x_{ni}, n = 1, ..., N, i = 1, ..., p$ werden dann als Daten-Matrix (Tabelle)

Daten-Matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... \\ x_{N1} & x_{N2} & ... & x_{Np} \end{bmatrix} : N \times p \tag{1.4}$$

dargestellt. Dies ist auch die Form der Speicherung in Dateien (englisch: files) für Statistikprogramme und Tabellenkalkulationen.

Häufig werden den Spalten noch Variablennamen (Land, Preis, Kosten für Werbung,Zahl der Vertreterbesuche etc.) und den Zeilen Namen für die statistischen Einheiten (Filiale 1,...,Filiale N) oder (Zeitpunkt 1,..., Zeitpunkt N) zugewiesen. Ein SPSS-Datensatz ist in Abb. 1.4 zu sehen (Daten- und Variablen-Ansicht).

Weiterhin können die Daten in Form von univariaten Histogrammen, Quantil-Graphiken etc. dargestellt werden (Abb. 1.6).

1.4 Deskriptive Statistik

Geht man über graphische Analysen hinaus, so berechnet man nicht nur die empirischen Stichprobenvarianzen $s_1^2, ..., s_p^2$ für die p Variablen (Realisationen) $x_1, ..., x_p$ einzeln, sondern auch die empirischen Kovarianzen

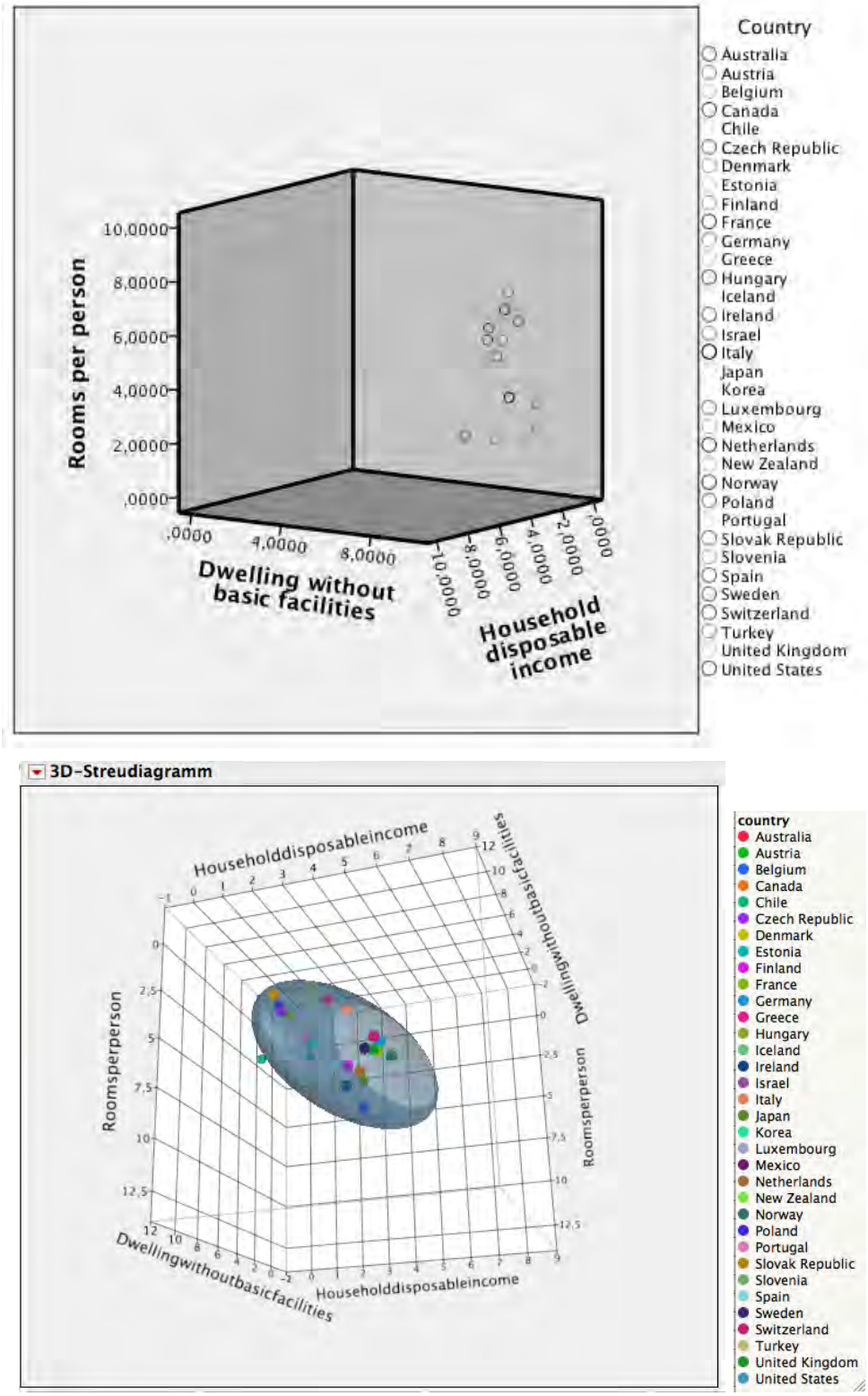


Abbildung 1.10: Oben: 3-D Streudiagramm der Variablen (SPSS).
Unten: Interaktives 3-D Streudiagramm der Variablen (SAS/JMP).

s_{ij} zwischen den Variablen x_i und x_j

$$s_{ij} = (N-1)^{-1} \sum_n (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j) \tag{1.5}$$

$$s_{ii} = s_i^2 = (N-1)^{-1} \sum_n (x_{ni} - \bar{x}_i)^2 \tag{1.6}$$

$$\bar{x}_i = N^{-1} \sum_n x_{ni}. \tag{1.7}$$

Zusammenhang

Korrelation

Diese geben den Zusammenhang (Kovarianz) s_{ij}, in standardisierter Form auch die sog. Korrelation

$$r_{ij} = \frac{s_{ij}}{s_i s_j} = r_{ji} \tag{1.8}$$

zwischen x_i und x_j an. In übersichtlicher Form läßt sich dieses Zusammenhangsmuster in Form einer Tabelle darstellen, die als Korrelations-Matrix $\mathbf{R}$ (für $p = 4$ Variablen)

Korrelationsmatrix

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{bmatrix} \tag{1.9}$$

bezeichnet wird. In der Diagonale stehen die Werte $r_{ii} = 1$, da $s_{ii} = s_i^2, i = 1, ..., p$ für die Einzelvarianzen gilt.

Beispiel 1.3 (Fortsetzung: OECD-Datensatz)

Die Zusammenhänge zwischen den Variablen lassen sich in Form einer Korrelationstabelle (Matrix) übersichtlich darstellen. In Abb. 1.8 sind die paarweisen Korrelationen

$$\mathbf{R} = \begin{bmatrix} 1.000 & .638 & .687 & .422 \\ .638 & 1.000 & .725 & .318 \\ .687 & .725 & 1.000 & .710 \\ .422 & .318 & .710 & 1.000 \end{bmatrix} \tag{1.10}$$

der Variablen als Tabelle dargestellt, zusammen mit univariaten Tests.

p-Wert

In der Tabelle stehen sog. p-Werte:[3] wenn p kleiner als ein vorgegebenes Signifikanzniveau α (z.B. 5%) ist, wird die Nullhypothese $\rho_{ij} = 0$ abgelehnt.

[3] die fälschlicherweise von SPSS als Signifikanz bezeichnet werden.

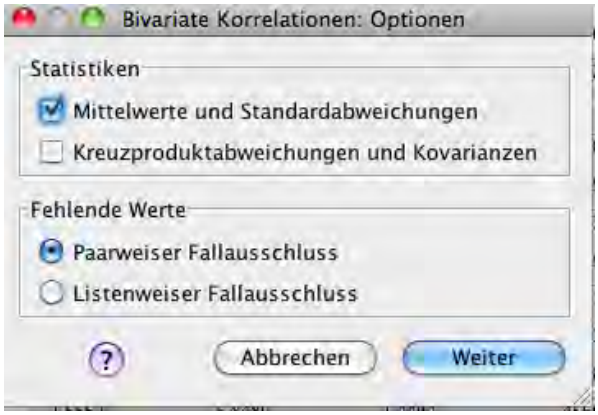


Abbildung 1.11: Paarweiser und listenweiser Fallausschluss

Die Stichprobengrößen der Spalten sind in der Praxis oft unterschiedlich, die nicht vorhandenen Werte werden als fehlende Daten bezeichnet. Sie sind in SPSS oder JMP standardmäßig als Punkt (.) codiert. Datenpaare (x_{in}, x_{jn}) mit fehlenden Werten auf einer Variable werden weggelassen (casewise deletion; paarweiser Fallausschluß).

fehlende Daten

casewise deletion

Man kann die Korrelation jedoch auch multivariat berechnen, wobei nur Beobachtungen mit vorhandenen Werten auf allen 4 Variablen benutzt werden (listwise deletion; listenweiser Fallausschluß). Diese Option kann auch in

listwise deletion

```
SPSS/Analysieren/Korrelation/Bivariate.../Optionen
```

angeklickt werden (Abb. 1.11).

■

Der Stoff wird in Aufgabe 1.2 vertieft.

1.5 Statistische Tests

Statistik

Bei der berechneten Korrelationsmatrix $\mathbf{R}$ handelt es sich um eine empirische Größe (Statistik), die statistischen Schwankungen unterworfen ist (etwa bei Berechnung in verschiedenen Zeiträumen). Will man über eine rein deskriptive Analyse hinaus gehen, können statistische Tests oder Vertrauensintervalle für die Korrelationen berechnet werden.

univariate Signifikanztests

Die einfachste Vorgehensweise besteht darin, die einzelnen Korrelationen r_{ij} zu prüfen. Man führt dann $p(p-1)/2$ univariate Signifikanztests aus, etwa für die Einzel-Hypothesen

$$H_{0,ij} : \rho_{ij} = 0 \tag{1.11}$$

$i = 1, ..., p; j = i + 1, ..., p$. Die Diagonale von $\mathbf{R}$ muß nicht getestet werden, da sie auf jeden Fall 1 ist. Außerdem ist die Korrelationsmatrix symmetrisch, sodaß man nur die obere (oder untere) Hälfte testen muß. Für $p = 4$ erhält man somit $4 * 3/2 = 6$ Tests.

Grundgesamtheit

Getestet wird (wie in jedem Test) nicht die berechnete Korrelation r_{ij}, sondern der hypothetische Korrelations-Parameter ρ_{ij} in der Grundgesamtheit.

Die empirische Korrelation r_{ij} ist hierfür eine Schätzung, die in großen Stichproben nur noch wenig vom (unbekannten) ρ_{ij} abweicht. Die Abweichung ist umgekehrt proportional zur Wurzel aus der Stichprobengröße N, also $N^{-1/2}$.

Man kann sich nun fragen, warum der multivariate Test nicht als Wiederholung univariater Vorgehensweisen aufgefaßt werden kann (etwa die $p(p-1)/2$-fache Wiederholung aller interessierenden Korrelationstests).

Signifikanzniveau

Ein Problem liegt darin, daß die einzelnen Tests mit einem Signifikanzniveau α, z.B. 5% ausgeführt werden. Dies bedeutet, daß die Nullhypothese fälschlicherweise verworfen wird, obwohl sie richtig ist. Dies geschieht mit einer Wahrscheinlichkeit α.

Wiederholt man mehrere Tests auf diesem Niveau, so ist das simultane Signifikanzniveau α^* für alle Tests wesentlich höher, etwa $\alpha^* \leq 6\alpha = 30\%$ für $k = 6$ Tests (siehe Übung 1.3). Dies bedeutet, daß die Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl H_0 richtig ist

$$\alpha^* = P(H_0 \text{ ablehnen} | H_0 \text{ richtig}) \tag{1.12}$$

(Fehler 1. Art), sehr groß werden kann.

Man kann sagen, daß bei der Ausführung vieler Tests signifikante Resultate zufällig entstehen, obwohl gar kein Effekt vorliegt (also Korrelation gleich Null).

Im Spezialfall voneinander unabhängiger Einzel-Tests kann man sogar zeigen, daß

$$\alpha^* = 1 - (1-\alpha)^k;\ k = \text{Anzahl der Tests} \tag{1.13}$$

gilt. Daher geht die simultane Irrtumswahrscheinlichkeit gegen 1, wenn man nur genügend viele Tests ausführt. In der Praxis sind $k = 100$ Tests keine Seltenheit. Dann findet man auf jeden Fall Effekte, obwohl keine vorliegen.

Übung 1.1 (Simultanes Signifikanzniveau)
Berechnen Sie die simultanen Irrtumswahrscheinlichkeiten für $\alpha = 5\%$ und $k = 2, ..., 10$ durchgeführte unabhängige Tests.
■

Übung 1.2 (Simultanes Signifikanzniveau: Herleitung)
Leiten Sie Gleichung 1.13 her.

Hinweis:
Für Ereignisse A gilt: $P(\bar{A}) = 1 - P(A)$ (Komplement) sowie $P(A \cap B) = P(A)P(B)$ bei Unabhängigkeit.
Definieren Sie die Ereignisse $E_i : H_{0i}$ beibehalten und $E = \bigcap_i E_i : H_0$ beibehalten und berechnen Sie $P(\bar{E})$.
Notation: $\bigcap_{i=1}^{I} E_i = E_1 \cap E_2 \cap ... \cap E_I$.
■

Da die Tests jedoch im allgemeinen nicht voneinander unabhängig sind, behilft man sich in der Praxis damit, die Signifikanzniveaus für die Einzelprüfungen abzusenken, etwa $\alpha \to \alpha/k$, sodaß $\alpha^* \leq \alpha = 5\%$ ist (Bonferroni-Methode).

Bonferroni-Methode

Allerdings werden dann die Einzeltests mit dem kleinen Wert α/k ausgeführt, was insgesamt zu einem **konservativen** Vorgehen führt. Dies bedeutet, daß das tatsächliche Signifikanzniveau α^* der Testprozedur deutlich kleiner als das vorgegebene $\alpha = 5\%$ sein kann. Man findet also

konservatives Testen

nichts, obwohl in den Daten Effekte vorliegen. Ein Tabelle mit adjustierten α-Niveaus für den t-Test ist in Kap. 12.4 zu finden. Die Quantile werden mit wachsendem k immer größer.

Übung 1.3 (Bonferroni-Ungleichung)
Zeigen Sie die Gültigkeit der Ungleichung

$$P(\bar{E}) = P(\bigcup_i \bar{E}_i) \leq \sum_i P(\bar{E}_i). \tag{1.14}$$

Dies bedeutet, daß die Wahrscheinlichkeit, die simultane Hypothese $H_0 = \bigcap_i H_{0i}$ abzulehnen, kleiner ist als die Summe der Wahrscheinlichkeiten, die Einzelhypothesen H_{0i} abzulehnen.

Hinweis:
Definieren Sie die Ereignisse $E_i : H_{0i}$ beibehalten und $E = \bigcap_i E_i : H_0$ beibehalten und berechnen Sie $P(\bar{E})$.
Für Ereignisse (Mengen) A, B gilt: $\overline{A \cap B} = \bar{A} \cup \bar{B}$ und $P(A \cup B) \leq P(A) + P(B)$.
■

Sphärizitäts-Test

Um ein konservatives Vorgehen zu vermeiden, sind exakte multivariate Tests notwendig, deren Signifikanzniveau genau gleich α ist, etwa der Korrelationstest (Sphärizitäts-Test)[4]

$$H_0 = \bigcap_{i<j} H_{0,ij} : \mathbf{P} \quad = \quad \mathbf{I}_p \tag{1.15}$$

wobei hierbei die gesamte Matrix

$$\mathbf{P} \quad = \quad \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{bmatrix} \tag{1.16}$$

[4]Der Name ergibt sich aus der Tatsache, daß die quadratische Form $\sum_{ij} x_i \rho_{ij} x_j = r^2$ eine p-dimensionale Kugel mit Radius r beschreibt, wenn ρ gleich der Einheitsmatrix ist. Für $p = 2$ ist $x_1^2 + x_2^2 = r^2$ ein Kreis mit Radius r.

Faktorenanalyse

[DatenSet1] /Users/hermannsinger/Library/texmf/tex/latex/Kurse/Multi_2012/Daten/SPSS-Datensätze/BetterLifeIndex.sav

Korrelationsmatrix

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth
Korrelation	Rooms per person	1,000	,638	,687	,422
	Dwelling without basic facilities	,638	1,000	,725	,318
	Household disposable income	,687	,725	1,000	,710
	Household financial wealth	,422	,318	,710	1,000
Signifikanz (1-seitig)	Rooms per person		,000	,000	,014
	Dwelling without basic facilities	,000		,000	,057
	Household disposable income	,000	,000		,000
	Household financial wealth	,014	,057	,000	

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,651
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	53,541
	df	6
	Signifikanz nach Bartlett	,000

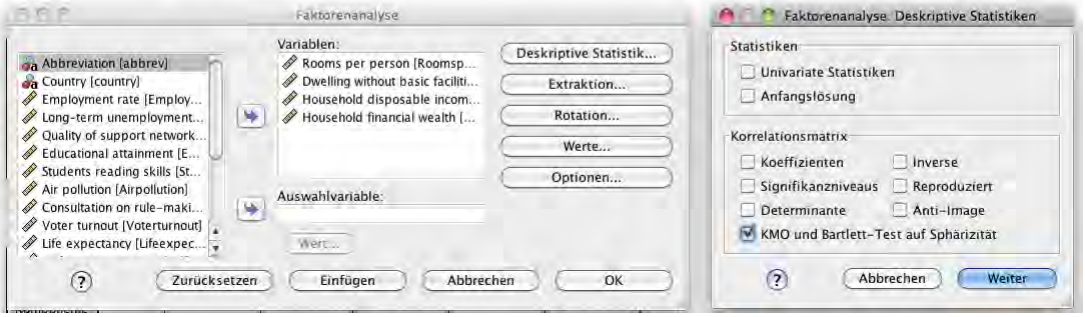


Abbildung 1.12: Bartlett-Test auf Sphärizität für die Korrelationsmatrix der OECD-Daten.

simultan auf Übereinstimmung mit der Einheitsmatrix

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.17}$$

Union-Intersection-Prinzip

getestet wird. Die Nullhypothese $H_0 = \bigcap_{i<j} H_{0,ij}$ läßt sich hierbei als Schnittmenge der univariaten Hypothesen auffassen (das sog. *Union-Intersection-Prinzip*).

In SPSS läßt sich dieser Test sehr einfach im Rahmen der Faktorenanalyse durchführen (Abb. 1.12). Da die meisten p-Werte der univariaten Tests in diesem Beispiel klein sind, ist es plausibel, daß auch der simultane Test hochsignifikant ausfällt.

mineralw	natrium	kalium	magnesiu	calcium	chlorid	sulfat	hydrogen	nitrat	nitrit	mangan	anionen	gesamt_m	minerali
Adelheidquelle	950,00	47,20	102,00	152,00	112,00	317,00	2937,0	,00	-9,00	-9,00	-9,00	4617,20	Extrem stark
Adelholzener	4,60	,50	31,00	94,00	3,60	10,00	430,0	4,00	-9,00	-9,00	-9,00	577,70	Gering
Bad Brückenauer	2,90	5,60	7,30	16,90	6,20	10,70	76,3	3,60	-9,00	-9,00	-9,00	129,50	Extrem gering
Bad Lauchstädt	58,20	22,30	44,20	112,00	37,00	287,00	325,0	,71	-9,00	-9,00	-9,00	886,41	Mittel
FB St. Anna	305,00	16,40	84,90	304,30	432,50	656,10	549,0	,50	-9,00	-9,00	-9,00	2348,70	Stark
Hirschquelle	261,00	11,60	29,20	216,00	41,50	85,00	1343,0	1,70	-9,00	-9,00	-9,00	1989,00	Mittel
Römerquell	12,40	2,00	55,90	401,00	19,70	495,00	878,0	1,00	-9,00	,02	,54	1865,56	Mittel
St. Gero	171,00	10,20	109,40	331,00	39,00	34,80	,1	8,28	-9,00	-9,00	-9,00	703,78	Gering
St. Christophorus	340,00	24,30	58,30	301,00	54,00	65,20	1990,0	2,60	-9,00	,05	,54	2835,99	Sehr stark
Staatl. Fachingen	574,00	14,50	60,50	99,00	129,00	37,00	1854,0	,00	-9,00	,32	,30	2768,62	Sehr stark

Abbildung 1.13: SPSS-Datensatz für den Mineraliengehalt von Heilwässern (`Heilwässer.sav`).

Übung 1.4 (Bonferroni-Adjustierung)
Das simultane Signifikanzniveau für den Korrelationstest wird auf $\alpha^* \leq \alpha = 5\%$ festgelegt.
Prüfen Sie die Einzelkorrelationen in Abb. 1.8 unter Einhaltung von α^*.
■

1.6 Mineral- und Heilwässer

Beispiel 1.4 (Gruppierung von Mineralwasser-Sorten)
Mineralwässer unterscheiden sich stark im Gehalt an Mineralien. Daher kann es von Interesse sein, aus der sehr großen Zahl von Sorten Gruppen zu bilden, um ein Angebots-Sortiment zu strukturieren. Abb. 1.13 zeigt eine Auswahl von Heilwässern (SPSS-Datensatz `Heilwässer.sav`). Eine wesentlich größere Zahl von Sorten ist im Datensatz `mineral.sav` enthalten (Quelle: Fristo-Getränkemarkt). Abb. 1.14 zeigt eine Scatterplot-Matrix für die Mineralien Natrium, Kalium, Magnesium, Kalzium sowie die Gesamt-Mineralien in mg/Liter bei Heilwässern. Der gesamte Datensatz ist in Abb. 1.15 gezeigt. Der Zusammenhang zwischen Natriumgehalt und Gesamt-Mineralisation ist in Abb. 1.16 für die Heilwässer (oben) und für den gesamten Datensatz dargestellt (unten). Die Korrelation aller Variablen (nur Heilwässer) zeigt Abb. 1.17. Das Muster ist recht unübersichtlich und reicht von stark positiven bis negativen Zusammenhängen. Viele Korrelationen sind aufgrund der kleinen Stichprobengrößen (durch fehlende Werte) nicht signifikant, auch bei Einzelbetrachtung ohne Bonferroni-Adjustierung. Die Korrelationsmatrix aller Sorten zeigt Abb. 1.18.

Dimensionsreduktion

Es ist evident, daß hier Methoden zur Reduktion der multivariaten Informationsfülle herangezogen werden müssen. Einerseits kann versucht

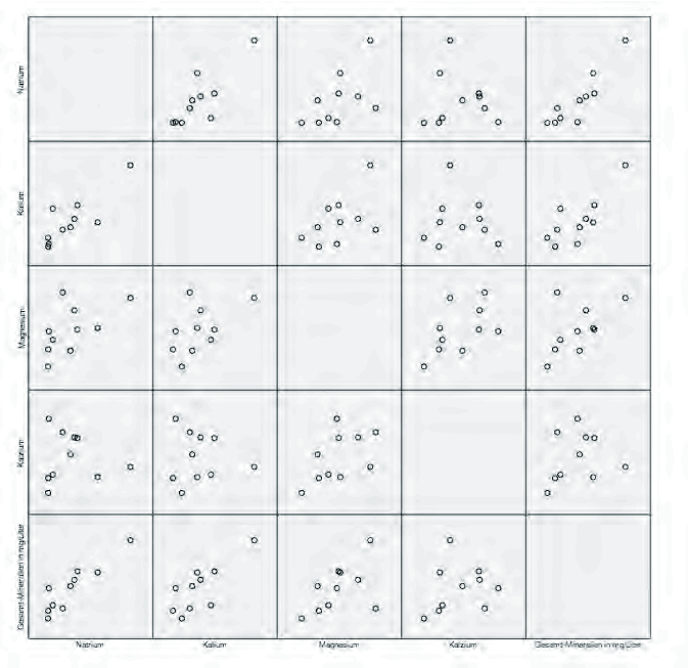


Abbildung 1.14: Scatter-Plots für den Mineraliengehalt von Heilwässern.

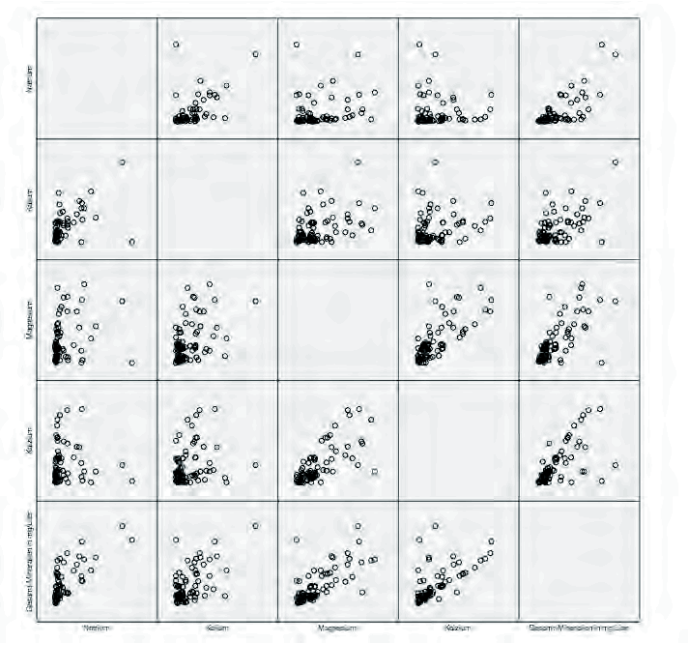
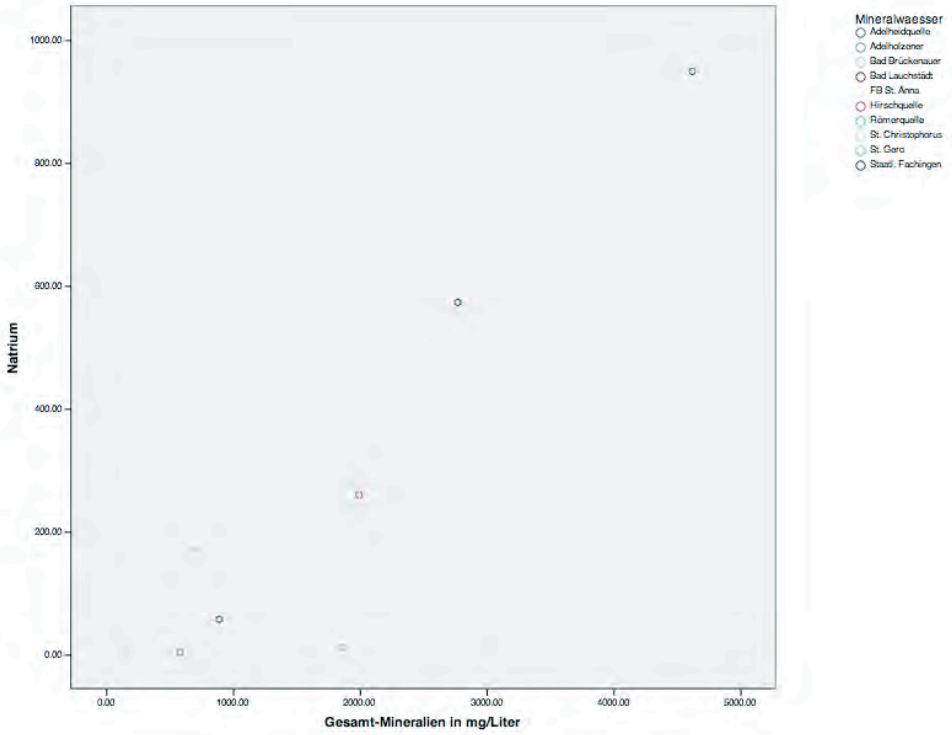


Abbildung 1.15: Scatter-Plots für den Mineraliengehalt von Mineralwässern.



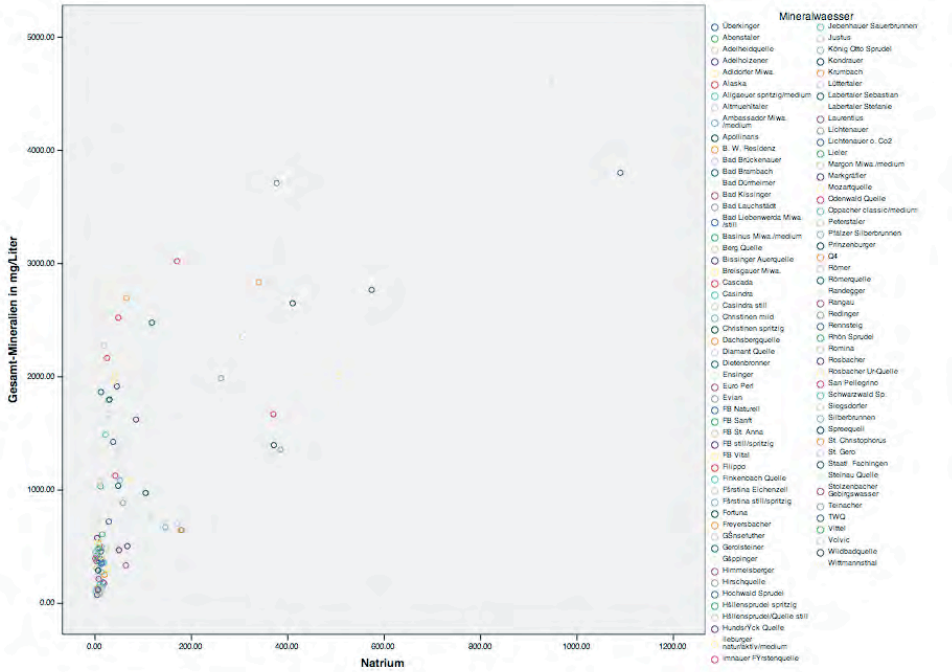


Abbildung 1.16: Zusammenhang zwischen Natriumgehalt und Gesamt-Mineralisation. Oben: Heilwässer, unten: Mineralwässer.

Correlations

		Natrium	Kalium	Magnesium	Kalzium	Chlorid	Sulfat	Hydrogen-Karbonat	Nitrat	Nitrit	Mangan	Anionen Fluorid	Gesamt-Mineralien in mg/Liter
Natrium	Pearson Correlation	1	,837**	,572	-,063	,334	,097	,873**	-,434	.[a]	,864	-,814	,909**
	Sig. (2-tailed)		,003	,084	,862	,346	,790	,001	,210	.	,337	,394	,000
	N	10	10	10	10	10	10	10	10	0	3	3	10
Kalium	Pearson Correlation	,837**	1	,522	-,071	,243	,204	,737*	-,416	.[a]	,160	-,070	,785**
	Sig. (2-tailed)	,003		,121	,846	,499	,572	,015	,232	.	,898	,956	,007
	N	10	10	10	10	10	10	10	10	0	3	3	10
Magnesium	Pearson Correlation	,572	,522	1	,529	,438	,366	,302	,110	.[a]	,897	-,853	,525
	Sig. (2-tailed)	,084	,121		,116	,206	,299	,397	,763	.	,291	,349	,119
	N	10	10	10	10	10	10	10	10	0	3	3	10
Kalzium	Pearson Correlation	-,063	-,071	,529	1	,238	,490	,011	,145	.[a]	-,971	,946	,204
	Sig. (2-tailed)	,862	,846	,116		,507	,151	,977	,690	.	,153	,211	,571
	N	10	10	10	10	10	10	10	10	0	3	3	10
Chlorid	Pearson Correlation	,334	,243	,438	,238	1	,674*	,083	-,382	.[a]	,976	-,952	,376
	Sig. (2-tailed)	,346	,499	,206	,507		,033	,820	,276	.	,141	,199	,284
	N	10	10	10	10	10	10	10	10	0	3	3	10
Sulfat	Pearson Correlation	,097	,204	,366	,490	,674*	1	,031	-,524	.[a]	-,621	,547	,332
	Sig. (2-tailed)	,790	,572	,299	,151	,033		,932	,120	.	,574	,632	,348
	N	10	10	10	10	10	10	10	10	0	3	3	10
Hydrogen-Karbonat	Pearson Correlation	,873**	,737*	,302	,011	,083	,031	1	-,558	.[a]	,481	-,400	,936**
	Sig. (2-tailed)	,001	,015	,397	,977	,820	,932		,094	.	,680	,738	,000
	N	10	10	10	10	10	10	10	10	0	3	3	10
Nitrat	Pearson Correlation	-,434	-,416	,110	,145	-,382	-,524	-,558	1	.[a]	-,734	,792	-,605
	Sig. (2-tailed)	,210	,232	,763	,690	,276	,120	,094		.	,476	,418	,064
	N	10	10	10	10	10	10	10	10	0	3	3	10
Nitrit	Pearson Correlation	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]
	Sig. (2-tailed)	.	.	.	.	.	.	.	.	.	.	.	.
	N	0	0	0	0	0	0	0	0	0	0	0	0
Mangan	Pearson Correlation	,864	,160	,897	-,971	,976	-,621	,481	-,734	.[a]	1	-,996	,525
	Sig. (2-tailed)	,337	,898	,291	,153	,141	,574	,680	,476	.		,058	,648
	N	3	3	3	3	3	3	3	3	0	3	3	3
Anionen Fluorid	Pearson Correlation	-,814	-,070	-,853	,946	-,952	,547	-,400	,792	.[a]	-,996	1	-,445
	Sig. (2-tailed)	,394	,956	,349	,211	,199	,632	,738	,418	.	,058		,706
	N	3	3	3	3	3	3	3	3	0	3	3	3
Gesamt-Mineralien in mg/Liter	Pearson Correlation	,909**	,785**	,525	,204	,376	,332	,936**	-,605	.[a]	,525	-,445	1
	Sig. (2-tailed)	,000	,007	,119	,571	,284	,348	,000	,064	.	,648	,706	
	N	10	10	10	10	10	10	10	10	0	3	3	10

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Abbildung 1.17: Korrelation zwischen allen Inhaltsstoffen (nur Heilwässer).

Correlations

		Natrium	Kalium	Magnesium	Kalzium	Chlorid	Sulfat	Hydrogen-Karbonat	Nitrat	Nitrit	Mangan	Anionen Fluorid	Gesamt-Mineralien in mg/Liter
Natrium	Pearson Correlation	1	,573**	,259**	,048	,513**	,216*	,669**	,017	,220	,594*	-,062	,702**
	Sig. (2-tailed)		,000	,008	,631	,000	,034	,000	,891	,601	,032	,707	,000
	N	103	87	103	103	95	97	103	64	8	13	39	103
Kalium	Pearson Correlation	,573**	1	,446**	,266*	,460**	,238*	,545**	,014	,281	,180	-,071	,606**
	Sig. (2-tailed)	,000		,000	,013	,000	,030	,000	,914	,500	,556	,673	,000
	N	87	87	87	87	84	84	87	61	8	13	38	87
Magnesium	Pearson Correlation	,259**	,446**	1	,668**	,278**	,524**	,543**	,100	,255	,265	-,151	,707**
	Sig. (2-tailed)	,008	,000		,000	,006	,000	,000	,433	,543	,381	,358	,000
	N	103	87	103	103	95	97	103	64	8	13	39	103
Kalzium	Pearson Correlation	,048	,266*	,668**	1	,243*	,838**	,263**	-,018	,163	-,091	-,126	,704**
	Sig. (2-tailed)	,631	,013	,000		,018	,000	,007	,886	,699	,767	,444	,000
	N	103	87	103	103	95	97	103	64	8	13	39	103
Chlorid	Pearson Correlation	,513**	,460**	,278**	,243*	1	,283**	,142	,067	,225	,010	-,069	,471**
	Sig. (2-tailed)	,000	,000	,006	,018		,007	,170	,600	,593	,974	,680	,000
	N	95	84	95	95	95	91	95	64	8	13	38	95
Sulfat	Pearson Correlation	,216*	,238*	,524**	,838**	,283**	1	,100	-,116	,137	-,257	-,091	,696**
	Sig. (2-tailed)	,034	,030	,000	,000	,007		,331	,367	,746	,397	,586	,000
	N	97	84	97	97	91	97	97	63	8	13	38	97
Hydrogen-Karbonat	Pearson Correlation	,669**	,545**	,543**	,263**	,142	,100	1	-,121	,604	,581*	-,136	,751**
	Sig. (2-tailed)	,000	,000	,000	,007	,170	,331		,343	,113	,037	,410	,000
	N	103	87	103	103	95	97	103	64	8	13	39	103
Nitrat	Pearson Correlation	,017	,014	,100	-,018	,067	-,116	-,121	1	,145	-,350	,015	-,093
	Sig. (2-tailed)	,891	,914	,433	,886	,600	,367	,343		,731	,241	,935	,464
	N	64	61	64	64	64	63	64	64	8	13	32	64
Nitrit	Pearson Correlation	,220	,281	,255	,163	,225	,137	,604	,145	1	,582	,628	,341
	Sig. (2-tailed)	,601	,500	,543	,699	,593	,746	,113	,731		,130	,095	,409
	N	8	8	8	8	8	8	8	8	8	8	8	8
Mangan	Pearson Correlation	,594*	,180	,265	-,091	,010	-,257	,581*	-,350	,582	1	,000	,444
	Sig. (2-tailed)	,032	,556	,381	,767	,974	,397	,037	,241	,130		1,000	,129
	N	13	13	13	13	13	13	13	13	8	13	13	13
Anionen Fluorid	Pearson Correlation	-,062	-,071	-,151	-,126	-,069	-,091	-,136	,015	,628	,000	1	-,078
	Sig. (2-tailed)	,707	,673	,358	,444	,680	,586	,410	,935	,095	1,000		,636
	N	39	38	39	39	38	38	39	32	8	13	39	39
Gesamt-Mineralien in mg/Liter	Pearson Correlation	,702**	,606**	,707**	,704**	,471**	,696**	,751**	-,093	,341	,444	-,078	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,464	,409	,129	,636	
	N	103	87	103	103	95	97	103	64	8	13	39	103

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Abbildung 1.18: Korrelation zwischen allen Inhaltsstoffen (alle Mineralwässer).

werden, die *Sorten* zu gruppieren, andererseits können ähnliche *Variablen* (Mineraliengehalt) gesucht werden. Abb. 1.19–1.20 zeigen eine Clusteranalyse mit den Sorten als Objekten. Als Variable wurde der Gesamtgehalt an Mineralien genommen. Abstände zwischen den Sorten sind somit die Differenzen des Mineraliengehalts. [5] Die durchgeführte Analyse berechnet die Abstände zwischen den Clustern als mittlere Differenz der Einzelabstände der Clusterelemente (sog. *Average Linkage-Methode*, von SPSS auch als *Linkage zwischen den Gruppen* bezeichnet). Das Verfahren ist hierarchisch und agglomerativ. Dies bedeutet, daß eine Abfolge von Partitionen (Aufteilungen) der Objektmenge konstruiert wird, wobei die Homogenität der Klassen schrittweise verringert wird (erhöht: divisiv). Das sog. Dendrogramm (Baumdiagramm) zeigt die einzelnen Cluster als Funktion der Homogenität der Klassen. Links ist die Zahl der Klassen maximal. Sehr ähnliche Objekte werden fusioniert, wodurch die Zahl der Klassen (Cluster) immer geringer wird. Die Abbildung zeigt eingefärbt eine 4-Cluster-Lösung. Man erkennt leicht die Wässer mit extremer Mineralisation (unten).

Clusteranalyse

Average Linkage

Dendrogramm

Nimmt man andere Distanzmaße, etwa *nächster Nachbar (Single Linkage)*, bei dem der Clusterabstand durch den geringsten Abstand der einzelnen Objekte definiert ist, so erhält man andere Dendrogramme. Insofern sind die Resultate immer etwas willkürlich und durch unterschiedliche Klassifizierungsmethoden und Abstandsmaße bedingt. Die Cluster lassen sich auch anschaulich im Raum der Variablen darstellen (hier Natrium vs. Gesamtmineralisation mit 90%-Konfidenzellipsen; Abb. 1.21). Demnach lassen sich die Wässer nach der Stärke ihrer Mineraliengehalte gruppieren, was die Strukturierung eines Getränke-Sortiments erleichtert.

Single Linkage

Bisher wurden die Sorten als p-dimensionale Objekte (Punkte) im Variablenraum (Mineralienkonzentration) betrachtet. Alternativ kann man die Variablen (Mineralien) als N-dimensionale Objekte betrachten, die nach Ähnlichkeit über die verschiedenen Sorten gruppiert werden. Dies bedeutet, daß in der Daten-Tabelle Zeilen und Spalten vertauscht werden (transponierte Datenmatrix), sodaß in den Zeilen die Mineralien (Natrium, Kalium, etc.) und in den Spalten die Sorten (Abentaler,.. etc.) auftreten. In SPSS ist die Option: `Cluster Cases` oder `Variables` möglich,

Variablen als Objekte

[5] Benutzt man mehrere Variablen $\mathbf{x}_n = [x_{n1}, ..., x_{np}]'$, so ist ein Abstandsmaß

$$d(n, m) = ||\mathbf{x}_n - \mathbf{x}_m|| \tag{1.18}$$

zugrundezulegen, etwa die euklidische Distanz, die der üblichen Entfernungs-Vorstellung entspricht. Vgl. Kap. 7.3.3.

Abbildung 1.19: Average Linkage-Clusteranalyse (JMP). Dendrogramm der Objekte (alle Mineralwässer). 4 Cluster sind eingefärbt.

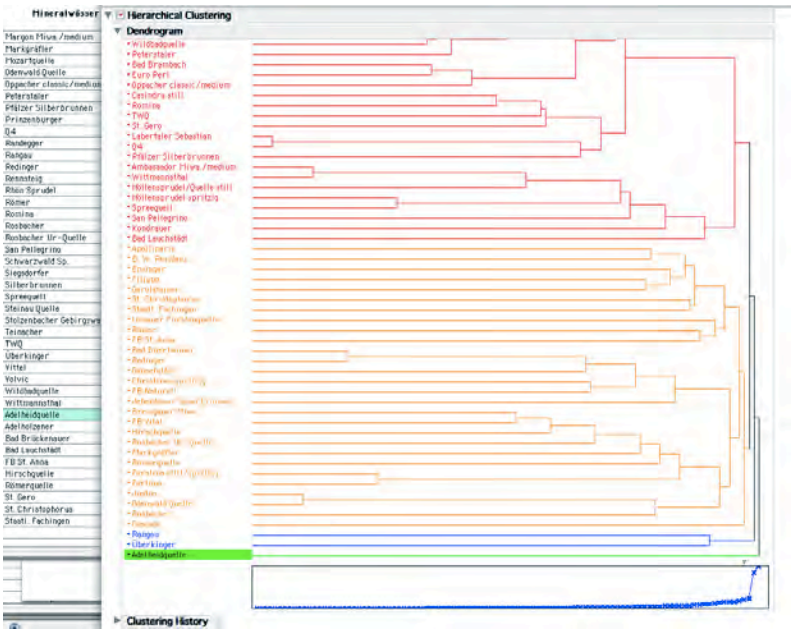


Abbildung 1.20: Single Linkage-Clusteranalyse (JMP). Dendrogramm der Objekte (alle Mineralwässer). 4 Cluster sind eingefärbt.

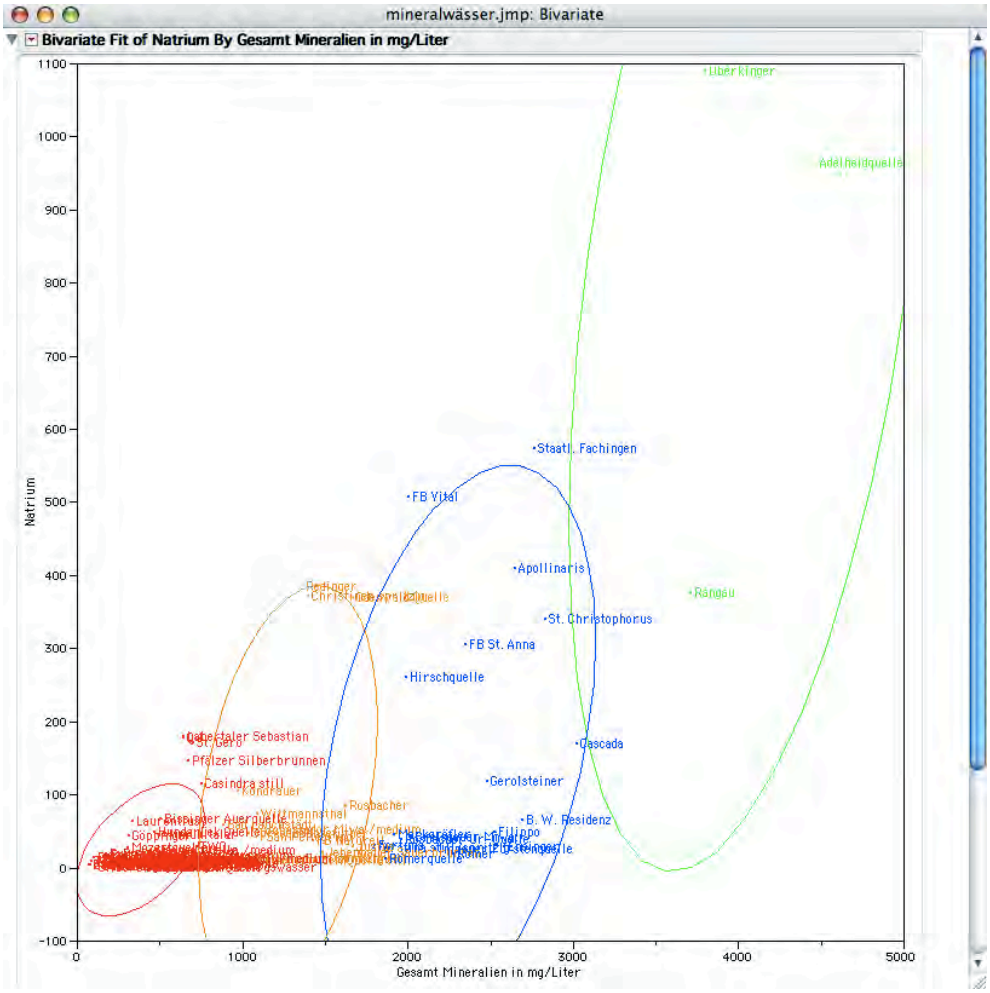


Abbildung 1.21: Average Linkage-Clusteranalyse (JMP). Streudiagramm der Objekte und 90%-Konfidenzellipsen (einige Mineralwässer sind bezeichnet).

ohne die Daten-Tabelle zu transponieren (Abb. 1.22).

Das Dendrogramm der Inhaltsstoffe ist in Abb. 1.23 gezeigt. Danach sind Natrium, Kalium und Chlorid in sehr ähnlicher Menge in den Sorten enthalten. Ein Blick auf die Korrelationstabelle (Abb. 1.18) zeigt, daß diese auch hoch korrelieren, also ähnlich sind und deshalb einen geringen quadrierten Abstand

Abstand

$$d(i,j)^2 = ||\mathbf{x}_{(i)} - \mathbf{x}_{(j)}||^2 := \sum_n (x_{ni} - x_{nj})^2 \tag{1.19}$$

$$= \sum_n (x_{ni}^2 + x_{nj}^2 - 2x_{ni}x_{nj}) \tag{1.20}$$

aufweisen ($\mathbf{x}_{(i)}$ Spalte i der Datenmatrix, N-dimensionaler Variablen-Raum).

Unter der Annahme zentrierter Variablen mit Mittelwert $\bar{x}_i = (1/N)\sum_n x_{ni} = 0$ gilt

$$d(i,j)^2 = (N-1)(s_i^2 + s_j^2 - 2r_{ij}s_is_j) \tag{1.21}$$

wobei die Stichprobenvarianzen $s_i^2 = (N-1)^{-1}\sum_n x_{ni}^2$ und Stichprobenkovarianzen $s_{ij} = (N-1)^{-1}\sum_n x_{ni}x_{nj}$ sowie die Korrelation $r_{ij} = s_{ij}/(s_is_j)$ eingesetzt wurden. Eine Korrelation von $r_{ij} = 1$ ergibt also $s_{ij} = s_is_j$ und somit

$$d(i,j)^2 = (N-1)(s_i - s_j)^2. \tag{1.22}$$

Bei gleichen Streuungen $s_i^2 = s_j^2 = s^2$ ist also der Abstand

$$d(i,j)^2 = 2(N-1)s^2(1 - r_{ij}) = 0, \tag{1.23}$$

wenn die Korrelation $r_{ij} = 1$ ist.

Eine *hohe Korrelation* bedeutet also einen *kleinen Abstand (hohe Ähnlichkeit)* der Objekte (Variablen).

Etwas überraschend ist, daß auch Nitrat mit den vorgenannten Stoffen clustert, obwohl eine sehr geringe Korrelation in Abb. 1.18 vorliegt. Das Rätsel löst sich, wenn man bedenkt, daß die Variablen Nitrit und Mangan sehr viele fehlende Daten aufweisen und sich daher die Clusteranalyse nur auf $N = 8$ Sorten stützt. Dies ist im Cluster-Output aufgelistet. Schließt man die Variablen Nitrit und Mangan aus der Analyse aus, so ergibt

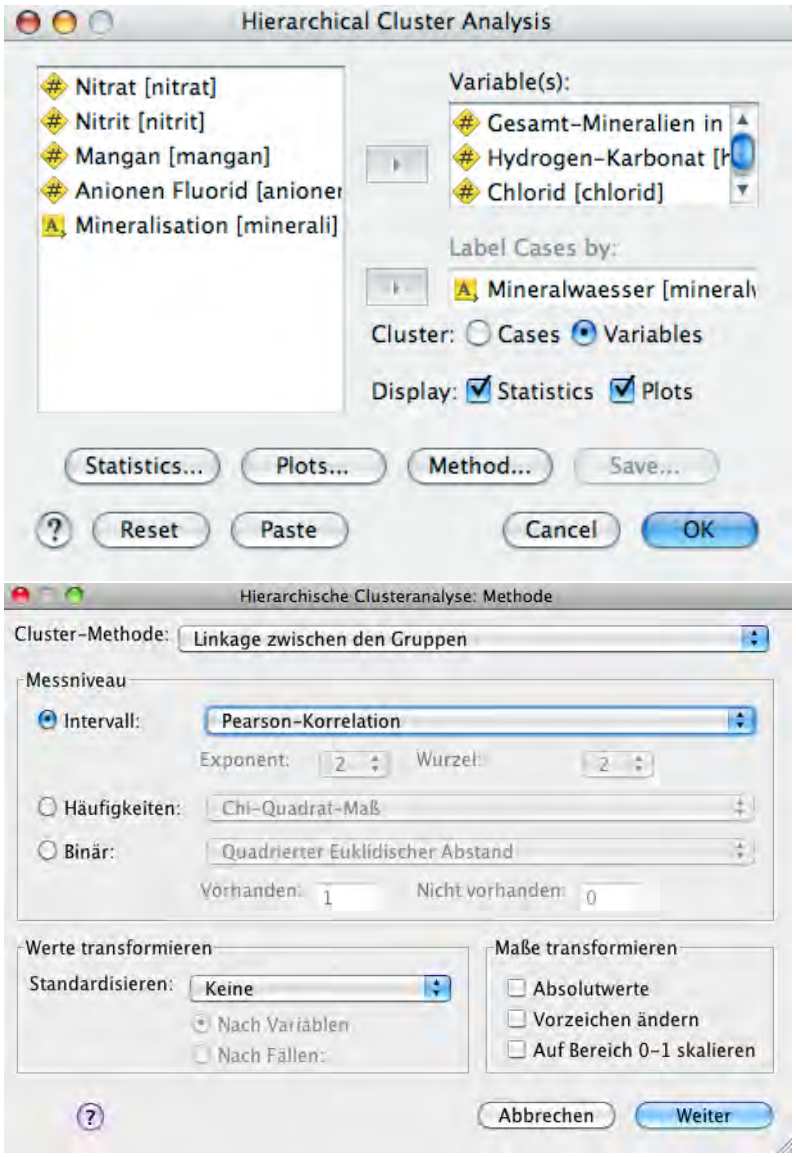
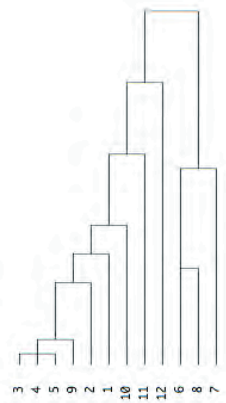


Abbildung 1.22: SPSS-Option zur Clusterung von Fällen (cases) oder Variablen (oben). Als Methode wurde Average Linkage und als Distanzmaß die Pearson-Korrelation benutzt (unten).

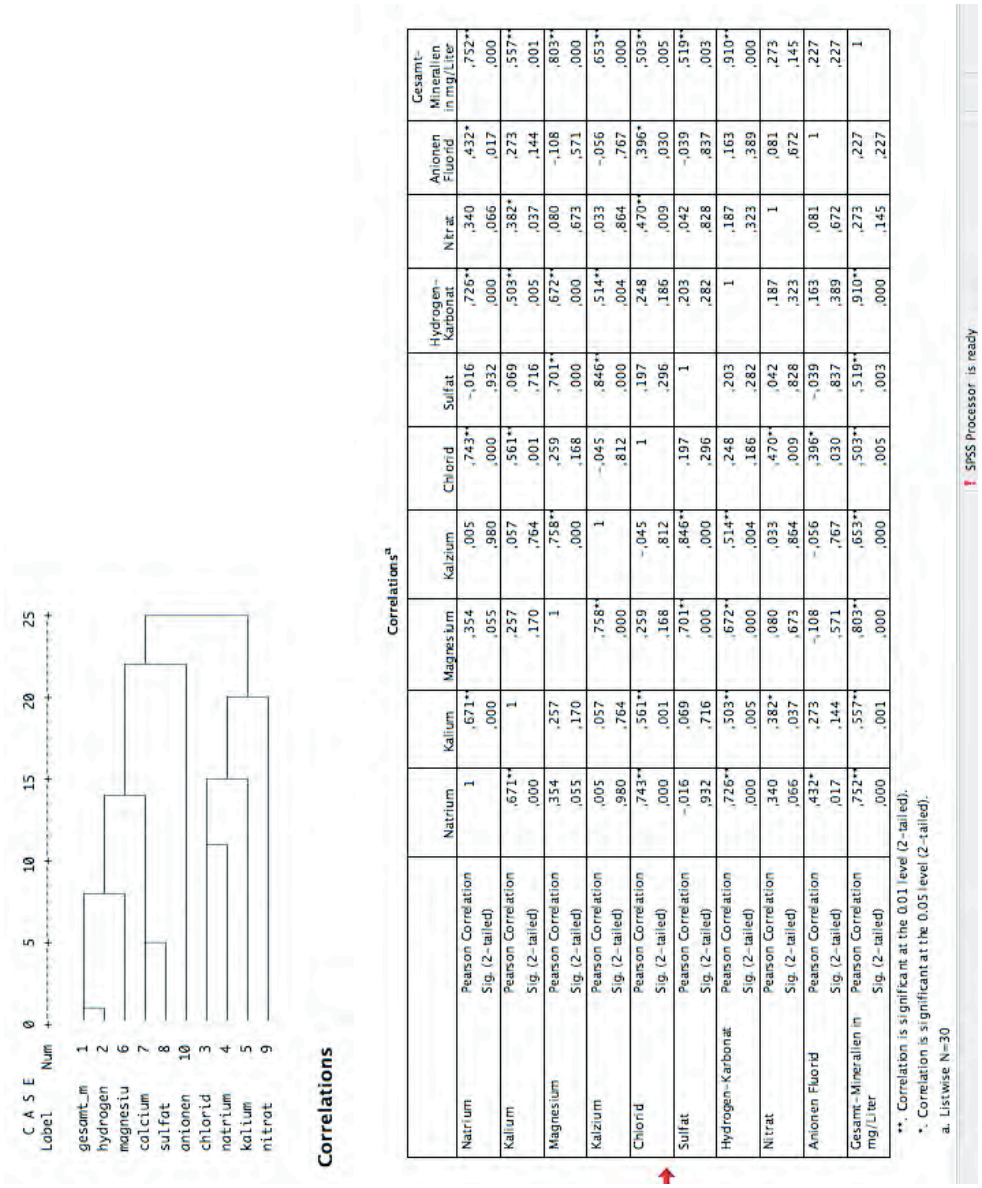


Correlations[a]

		Natrium	Kalium	Magnesium	Kalzium	Chlorid	Sulfat	Hydrogen-Karbonat	Nitrat	Anionen Fluorid	Gesamt-Mineralien in mg/Liter	Nitrit	Mangan
Natrium	Pearson Correlation	1	,995**	,425	-,370	,997**	,205	,878**	,991**	,823*	,847**	,220	-,286
	Sig. (2-tailed)		,000	,293	,367	,000	,627	,004	,000	,012	,008	,601	,492
Kalium	Pearson Correlation	,995**	1	,490	-,296	,997**	,271	,915**	,974**	,829*	,885**	,281	-,249
	Sig. (2-tailed)	,000		,218	,477	,000	,517	,001	,000	,011	,004	,500	,552
Magnesium	Pearson Correlation	,425	,490	1	,669	,491	,956**	,628	,326	,170	,833*	,255	-,434
	Sig. (2-tailed)	,293	,218		,070	,216	,000	,095	,431	,687	,010	,543	,282
Kalzium	Pearson Correlation	-,370	-,296	,669	1	-,301	,818*	-,032	-,472	-,454	,180	,163	-,150
	Sig. (2-tailed)	,367	,477	,070		,469	,013	,941	,237	,258	,669	,699	,723
Chlorid	Pearson Correlation	,997**	,997**	,491	-,301	1	,278	,893**	,980**	,801*	,883**	,225	-,318
	Sig. (2-tailed)	,000	,000	,216	,469		,505	,003	,000	,017	,004	,593	,443
Sulfat	Pearson Correlation	,205	,271	,956**	,818*	,278	1	,420	,101	-,083	,682	,137	-,458
	Sig. (2-tailed)	,627	,517	,000	,013	,505		,301	,812	,845	,062	,746	,253
Hydrogen-Karbonat	Pearson Correlation	,878**	,915**	,628	-,032	,893**	,420	1	,818*	,866**	,917**	,604	-,057
	Sig. (2-tailed)	,004	,001	,095	,941	,003	,301		,013	,005	,001	,113	,893
Nitrat	Pearson Correlation	,991**	,974**	,326	-,472	,980**	,101	,818*	1	,813*	,778*	,145	-,302
	Sig. (2-tailed)	,000	,000	,431	,237	,000	,812	,013		,014	,023	,731	,468
Anionen Fluorid	Pearson Correlation	,823*	,829*	,170	-,454	,801*	-,083	,866**	,813*	1	,620	,628	,198
	Sig. (2-tailed)	,012	,011	,687	,258	,017	,845	,005	,014		,101	,095	,638
Gesamt-Mineralien in mg/Liter	Pearson Correlation	,847**	,885**	,833*	,180	,883**	,682	,917**	,778*	,620	1	,341	-,375
	Sig. (2-tailed)	,008	,004	,010	,669	,004	,062	,001	,023	,101		,409	,360
Nitrit	Pearson Correlation	,220	,281	,255	,163	,225	,137	,604	,145	,628	,341	1	,582
	Sig. (2-tailed)	,601	,500	,543	,699	,593	,746	,113	,731	,095	,409		,130
Mangan	Pearson Correlation	-,286	-,249	-,434	-,150	-,318	-,458	-,057	-,302	,198	-,375	,582	1
	Sig. (2-tailed)	,492	,552	,282	,723	,443	,253	,893	,468	,638	,360	,130	

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. ListWise N=8

Abbildung 1.23: Average Linkage-Clusteranalyse der Inhaltsstoffe (Ähnlichkeit aller Variablen; $N = 8$ gültige Fälle). Der Terminus 'proximity matrix' wurde mit Näherungsmatrix falsch übersetzt, korrekt wäre *Ähnlichkeitsmatrix*.



		Natrium	Kalium	Magnesium	Kalzium	Chlorid	Sulfat	Hydrogen-Karbonat	Nitrat	Anionen Fluorid	Gesamt-Mineralien in mg/Liter
Natrium	Pearson Correlation	1	,671**	,354	,005	,743**	-,016	,726**	,340	,432*	,752**
	Sig. (2-tailed)		,000	,055	,980	,000	,932	,000	,066	,017	,000
Kalium	Pearson Correlation	,671**	1	,257	,057	,561**	,069	,503**	,382*	,273	,557**
	Sig. (2-tailed)	,000		,170	,764	,001	,716	,005	,037	,144	,001
Magnesium	Pearson Correlation	,354	,257	1	,758**	,259	,701**	,672**	,080	-,108	,803**
	Sig. (2-tailed)	,055	,170		,000	,168	,000	,000	,673	,571	,000
Kalzium	Pearson Correlation	,005	,057	,758**	1	-,045	,846**	,514**	,033	-,056	,653**
	Sig. (2-tailed)	,980	,764	,000		,812	,000	,004	,864	,767	,000
Chlorid	Pearson Correlation	,743**	,561**	,259	-,045	1	,197	,248	,470**	,396*	,503**
	Sig. (2-tailed)	,000	,001	,168	,812		,296	,186	,009	,030	,005
Sulfat	Pearson Correlation	-,016	,069	,701**	,846**	,197	1	,203	,042	-,039	,519**
	Sig. (2-tailed)	,932	,716	,000	,000	,296		,282	,828	,837	,003
Hydrogen-Karbonat	Pearson Correlation	,726**	,503**	,672**	,514**	,248	,203	1	,187	,163	,910**
	Sig. (2-tailed)	,000	,005	,000	,004	,186	,282		,323	,389	,000
Nitrat	Pearson Correlation	,340	,382*	,080	,033	,470**	,042	,187	1	,081	,273
	Sig. (2-tailed)	,066	,037	,673	,864	,009	,828	,323		,672	,145
Anionen Fluorid	Pearson Correlation	,432*	,273	-,108	-,056	,396*	-,039	,163	,081	1	,227
	Sig. (2-tailed)	,017	,144	,571	,767	,030	,837	,389	,672		,227
Gesamt-Mineralien in mg/Liter	Pearson Correlation	,752**	,557**	,803**	,653**	,503**	,519**	,910**	,273	,227	1
	Sig. (2-tailed)	,000	,001	,000	,000	,005	,003	,000	,145	,227	

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. Listwise N=30

Abbildung 1.24: Average Linkage-Clusteranalyse der Inhaltsstoffe (Ähnlichkeit der Variablen ohne Nitrit und Mangan; $N = 30$ gültige Fälle).

sich folgendes Bild mit $N = 30$ gültigen Fällen (Abb. 1.24). Dies zeigt, daß die Ergebnisse von Analysen sehr stark von der Zahl der gültigen Fälle (statistische Einheiten) abhängen und sich je nach Vorgehensweise unterscheiden. ■

Der Stoff wird in Aufgabe 1.3 vertieft.

1.7 Gesichtspunkte bei multivariaten Analysen

Um sich im Dickicht der verschiedenen Verfahren zurechtzufinden, gibt es einige Kriterien, die eine systematische Einordnung der Verfahren erlauben.

Wichtige Gesichtspunkte sind:

- Skalen-Niveau der Variablen: diskret, stetig oder gemischt.
- Symmetrische/Asymmetrische Zusammenhänge:

  gibt es abhängige (AV) und unabhängige Variablen (UV) oder sind alle Variablen gleichberechtigt?
- Interessiert man sich für die Zusammenhänge der Variablen oder der Objekte (statistische Einheiten)?
- Manifeste/Latente Variablen: sind die Variablen direkt beobachtbar oder nur hypothetische Konstrukte?

Nimmt man nur asymmetrische Zusammenhänge und unterscheidet das Skalenniveau, so ergibt sich die Tabelle

Asymmetrische Verfahren $Y = f(X)$		
	UV $= X$	
AV $= Y$	diskret	stetig
diskret	Kreuztabellen, log-lineare Modelle kategoriale Regression	kategoriale Regression Diskriminanz-Analyse
stetig	Varianz-Analyse	Regressions-Analyse

allgemeines lineares Modell

Kovarianzanalyse

Sind die abhängigen Variablen stetig und hat man gemischte stetige und diskrete unbhängige Variablen, so spricht man auch vom allgemeinen linearen Modell. Nimmt man bei der Varianz-Analyse (diskrete UV) noch stetige Kovariablen (d.h. weitere UV) hinzu, so ergibt sich das Modell der Kovarianzanalyse.

Clusteranalyse

Verfahren, bei denen Objekte (Zeilen der Datenmatrix) anhand der Spalten (Variablen) gruppiert werden, entstammen dem Bereich der Clusteranalyse.

Faktorenanalyse

Hat man eine große Zahl korrelierter Variablen, so kann eine Dimensionsreduktion auf wenige latente Faktoren angestrebt werden (Faktorenanalyse).

Strukturgleichungs-Modellierung

Auch sind Kombinationen von Regressions- und Faktorenanalyse möglich. Dies wird als Strukturgleichungs-Modellierung bezeichnet.

Kapitel 2

Multivariate Verteilungen und Zufallsvariablen

2.1 Die bivariate Normalverteilung

2.1.1 Gemeinsame Dichte

Ähnlich wie im univariaten Fall spielt die Normalverteilung bei mehrdimensionalen (multivariaten) Problemen eine wichtige Rolle. Sie kann durch den Erwartungswert $\boldsymbol{\mu}$ und die Varianz $\boldsymbol{\Sigma}$ in ihrer Lage und Streuung charakterisiert werden. Abb. 2.1 zeigt den bivariaten Fall ($p = 2$ Variablen). Es handelt sich um ein interaktives Applet, das auf der Homepage des Lehrstuhls zu finden ist. [1] Die theoretische Verteilung der Zufallsvariablen X, Y ist durch eine glockenförmige Dichtefunktion $f(x, y)$ gegeben, deren Parameter μ_x, μ_y die Lokalisation (Lage) und deren Varianzen $\sigma_{xx} = \sigma_x^2, \sigma_{yy} = \sigma_y^2$ und Kovarianzen σ_{xy} die Streuung auf der x- und y-Achse sowie den Zusammenhang der Variablen bestimmen.

Man schreibt kurz[2]

$$\mathbf{x} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{2.1}$$

[1] http://www.fernuni-hagen.de/ls_statistik/lehre/

[2] Normalerweise werden Zufallsvariablen $X = X(\omega)$ durch Großbuchstaben gekennzeichnet. Andererseits ist es üblich, Vektoren klein und Matrizen groß zu schreiben. Im Rahmen der multivariaten Analyse erscheint es daher sinnvoll, auch Zufallsvektoren $\mathbf{x} = \mathbf{x}(\omega)$ klein zuschreiben. In Zweifelsfällen kann der Zufall ω explizit als Argument angegeben werden, etwa $E[\mathbf{A}(\omega)]$ (Erwartungswert einer zufälligen Matrix). Weiterhin ist es in der Literatur üblich, Vektoren und Matrizen **fett** zu schreiben.

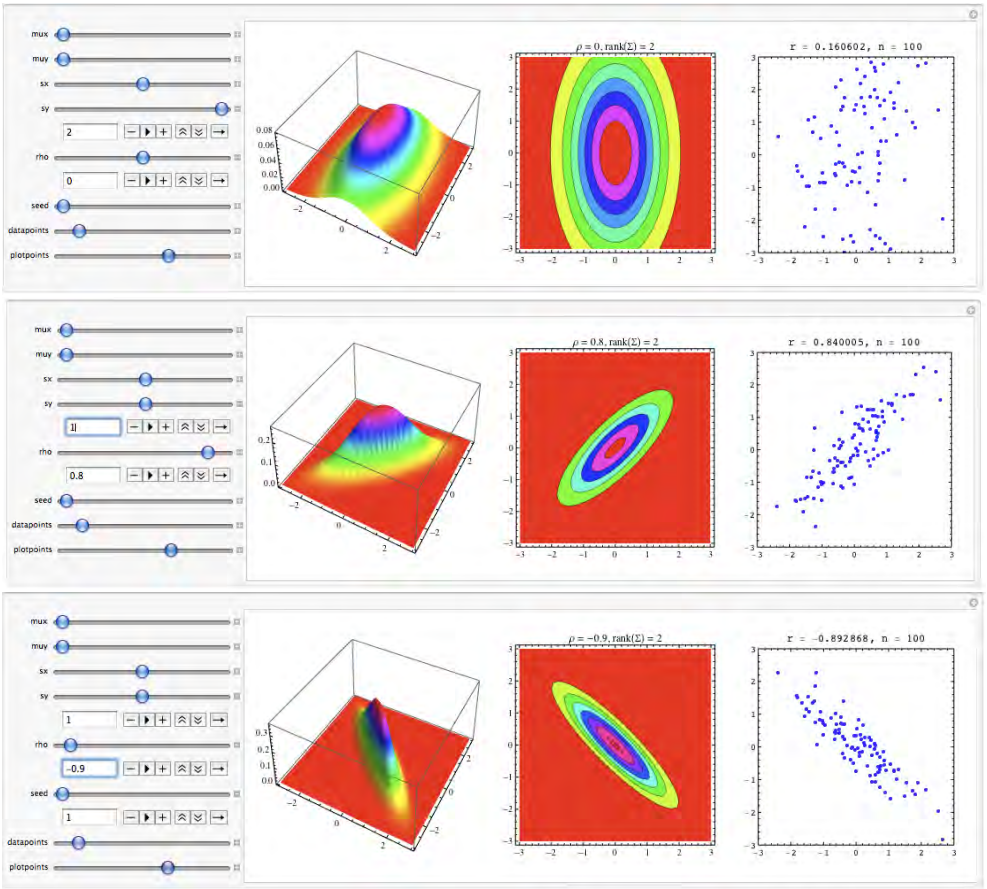


Abbildung 2.1: Bivariate Normalverteilungsdichte.
Obere Zeile: $\rho_{xy} = 0$, $\sigma_x = 1$, $\sigma_y = 2$. Mittlere Zeile: $\rho_{xy} = 0.8$ $\sigma_x = 1$, $\sigma_y = 1$. Untere Zeile: $\rho_{xy} = -0.9$ $\sigma_x = 1$, $\sigma_y = 1$.
Von Links: Regler, 3D-Graphik, Höhenlinien und simulierte Daten ($N = 100$).
http://www.fernuni-hagen.de/ls_statistik/lehre/

für die Zufallsvariablen bzw. die Verteilung und

$$\boldsymbol{\mu} = E[\mathbf{x}] = \begin{bmatrix} E(X) \\ E(Y) \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \tag{2.2}$$

sowie

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix} \tag{2.3}$$

$$= \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \tag{2.4}$$

$$= \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho_{xy} \\ \sigma_y\sigma_x\rho_{yx} & \sigma_y^2 \end{bmatrix} \tag{2.5}$$

für die Parameter der bivariaten Normalverteilung. Es gilt $\mathrm{Cov}(Y,X) = \mathrm{Cov}(X,Y)$.

Dabei wurden die Variablen und Erwartungswerte zu (2 mal 1) Spaltenvektoren $\mathbf{x} : 2 \times 1$ und $\boldsymbol{\mu} : 2 \times 1$ und die Varianzen und Kovarianzen zu einer (2 mal 2) Matrix $\boldsymbol{\Sigma} : 2 \times 2$ zusammengefasst.

Vektor
Matrix

Der Korrelationskoeffizient

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \rho_{yx} \tag{2.6}$$

Korrelationskoeffizient

bestimmt den (linearen) Zusammenhang der Variablen.

Die *Dichtefunktion* $f(x,y) = \phi(x,y)$ der bivariaten Normalverteilung ist durch folgenden Ausdruck gegeben:

$$\phi(x,y) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{2.7}$$

Normalverteilung (bivariate Dichte)

Hierbei ist $\det(\mathbf{A}) := |\mathbf{A}|$ die Determinante der Matrix $\mathbf{A}$ und $\mathbf{A}^{-1}$ die Inverse der Matrix. Weiterhin bedeutet $\mathbf{x}' = [x,y]$ den transponierten Spaltenvektor, also eine Zeile. Daher ist der Term im Exponent von der Form $Zeilenvektor \cdot Matrix \cdot Spaltenvektor = \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x}$. Der Punkt (wird meistens weggelassen) ist als Matrix-Produkt zu verstehen, d.h.

$$\begin{aligned} \mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} = \sum_{i=1}^{2}\sum_{j=1}^{2} x_i a_{ij} x_j &= x_1 a_{11} x_1 + x_1 a_{12} x_2 \\ &+ x_2 a_{21} x_1 + x_2 a_{22} x_2. \end{aligned} \tag{2.8}$$

Man erhält also als Ergebnis eine Zahl (quadratische Form).

Es ist unvermeidlich, einige Kenntnisse der Matrix-Rechnung zu erwerben, da die meisten Formeln so übersichtlicher geschrieben werden können. Weiterhin lassen sich viele Umformungen nur mit Hilfe von Matrix-Operationen bewältigen.
Eine Zusammenstellung von Methoden, die für diesen Kurs relevant sind, finden Sie im Anhang Matrix-Algebra (Kap. 9).
Weiterhin sind auf der Homepage des Lehrstuhls interaktive Applets verfügbar, mit denen Matrixoperationen ausgeführt werden können (vgl. Abb. 8.1).

Übung 2.1 (Quadratische Form)
i) Berechnen Sie die quadratische Form $\mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x}$ mit
$\mathbf{A} = \begin{bmatrix} 1 & .8 \\ .8 & 1 \end{bmatrix}$ und $\mathbf{x}' = [x, y]$.
ii) Welche Kurve definiert die Gleichung $\mathbf{x}' \cdot \mathbf{A} \cdot \mathbf{x} = 1$?
iii) Welche Kurve erhält man, wenn der Wert 0.8 durch 0 ersetzt wird ?
Hinweis: Betrachten Sie Abb. 2.1.

■

Setzt man die Parameter (2.2–2.3) der bivariaten Normalverteilung in (2.7) ein, so gilt explizit

$$\begin{aligned} \phi(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \\ &\times \exp\left\{ -\frac{1}{2(1-\rho_{xy}^2)} \right. \\ &\times \left. \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\}. \end{aligned} \tag{2.9}$$

Übung 2.2 (Berechnung der Determinante)
Zeigen Sie, daß $\det(\mathbf{\Sigma}) = \sigma_x^2\sigma_y^2(1-\rho_{xy}^2)$ gilt.

Hinweis:
Die Determinante von $\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho_{xy} \\ \sigma_x\sigma_y\rho_{xy} & \sigma_y^2 \end{bmatrix}$
ergibt sich aus der Formel $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - cb$.

■

Übung 2.3 (Berechnung der inversen Matrix)
i) Zeigen Sie, daß

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = (ad-cb)^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{2.10}$$

gilt.

Hinweis:
Multiplizieren Sie die rechte Seite von (2.10) mit $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Das Ergebnis ist $\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (Einheitsmatrix).
ii) Bestimmen Sie die Inverse der Kovarianzmatrix (2.5).

■

Der komplizierte Ausdruck (2.9) vereinfacht sich bei unkorrelierten Variablen. Setzt man den Korrelationskoeffizienten $\rho_{xy} = 0$, so ergibt sich der Ausdruck

$$\phi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \tag{2.11}$$

$$= \phi(x;\mu_x,\sigma_x^2)\cdot\phi(y;\mu_y,\sigma_y^2) \tag{2.12}$$

wobei

$$\phi(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{2.13}$$

die univariate Normalverteilungsdichte ist (siehe Abb. 2.1, oben).

Im unkorrelierten Fall ist also die bivariate Dichte das Produkt der univariaten Dichten, daher sind die Variablen X, Y auch voneinander unabhängig. Bei normalverteilten Variablen ist also Unabhängigkeit gleichbedeutend mit Unkorreliertheit.

Die Faktorisierung der Dichte ist analog zu $P(A \cap B) = P(A)P(B)$ bei unabhängigen Ereignissen.

In der Tat ist die Dichtefunktion durch die Ereignisse $A = \{x \leq X \leq x + dx\}$, $B = \{y \leq Y \leq y + dy\}$ und $P(A \cap B) \approx f(x, y)dxdy$ gegeben. Nimmt man also ein (kleines) Rechteck der Fläche $dxdy$ und eine Säule mit Höhe $f(x, y)$, so ist das *Volumen die Wahrscheinlichkeit*, daß man Werte im Rechteck $[x, x + dx] \times [y, y + dy]$ findet (Abb. 2.2).

2.1.2 Randverteilungen und bedingte Normalverteilung

Im letzten Abschnitt wurde der normalverteilte Vektor

$$\mathbf{x} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\Big(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \Big) \tag{2.14}$$

durch die gemeinsame Dichte $\phi(x, y) = \phi(x, y; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ charakterisiert. Man kann sich nun auch fragen, wie die einzelnen Variablen X bzw. Y verteilt sind.

2.1.2.1 Randverteilungen

Randverteilungen

Es ergibt sich das keinesfalls triviale Resultat, daß auch die sog. Randverteilungen $\phi(x; \mu_x, \sigma_x^2)$, $\phi(y; \mu_y, \sigma_y^2)$ Normalverteilungsdichten sind; auch dann, wenn die Variablen X, Y korreliert sind.

2.1.2.2 Bedingte Verteilung

bedingte Verteilung

Von großer praktischer Relevanz ist auch die Frage nach der bedingten Verteilung von Y, gegeben X. Hat man etwa den Wert der Körpergröße, so ist die bedingte Verteilung (Dichte) $f(\texttt{Gewicht}|\texttt{Größe})$ ein gutes Prognoseinstrument für das unbekannte Gewicht einer Person.

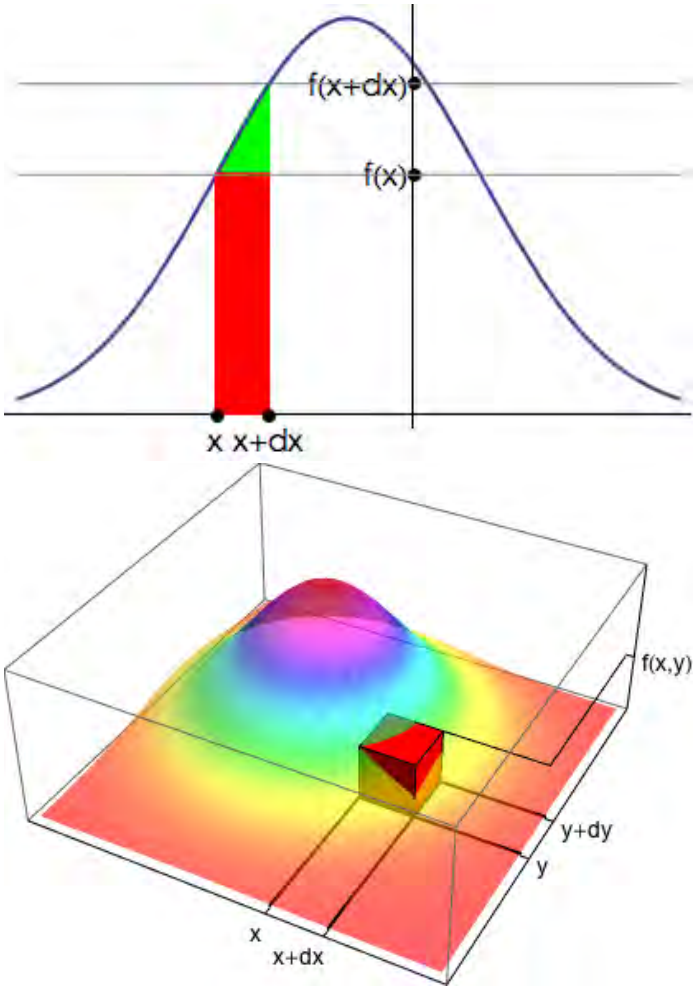


Abbildung 2.2: Dichtefunktionen und Wahrscheinlichkeiten.
http://www.fernuni-hagen.de/ls_statistik/lehre/

Mit Hilfe der Bayes-Formel[3]

Bayes-Formel

$$f(y|x) = \frac{f(x,y)}{f(x)} \tag{2.15}$$

läßt sich diese bedingte Dichte aus der gemeinsamen Dichte und der Randdichte berechnen (dies gilt ganz allgemein). Bei Normalverteilungen gilt das Resultat:

$$Y|X \sim N(\mu_{y|x}, \sigma_{y|x}) \tag{2.16}$$

wobei die bedingten Erwartungswerte und Varianzen durch

Satz von der Normalkorrelation

$$\mu_{y|x} = E[Y|X] = \mu_y + \sigma_{yx}\sigma_{xx}^{-1}(X - \mu_x) \tag{2.17}$$
$$\sigma_{y|x} = \mathrm{Var}[Y|X] = \sigma_{yy} - \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy} \tag{2.18}$$

gegeben sind. Dies ist der sog. *Satz von der Normalkorrelation*. Er besagt, daß sich der bedingte Erwartungswert

optimale Prognose

$$\hat{Y} = E[Y|X] \tag{2.19}$$

(optimale Prognose von Y durch X) bei Normalverteilungen als *lineare Funktion* von X schreiben läßt. Dies ist der tiefere Grund für die Benutzung der linearen Regressionsanalyse in vielen Fragestellungen. Die Streuung der Variablen Y wird durch die Kenntnis von X verringert.

Man kann auch in der üblichen Notation schreiben

$$E[Y|X] = \alpha + \beta X \tag{2.20}$$
$$\alpha = \mu_y - \beta\mu_x \tag{2.21}$$
$$\beta = \sigma_{yx}\sigma_{xx}^{-1}. \tag{2.22}$$

Beispiel 2.1 (Lineare Regression)

Setzt man $\boldsymbol{\mu} = [0,0]'$ und $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ (standardisierte Variable), so ergibt sich

$$\alpha = 0 \tag{2.23}$$
$$\beta = \rho \tag{2.24}$$
$$E[Y|X] = \rho X \tag{2.25}$$
$$\mathrm{Var}(Y|X) = 1 - \rho^2. \tag{2.26}$$

[3] $P(B|A) = P(A \cap B)/P(A)$.

Die Information, die in X über die Variable Y vorhanden ist, verringert deren Streuung. Ist die Korrelation $\rho = 1$, so gilt $\mathrm{Var}(Y|X) = 1 - \rho^2 = 0$. In diesem Fall verschwindet die bedingte Varianz.

■

Im folgenden werden folgende Rechenregeln oft benutzt:

Sind X und Y irgendwelche Zufallsvariablen (nicht unbedingt normalverteilt) und a, b Konstanten. Dann gilt (:= Definition)

Rechenregeln für Zufallsvariablen

$$
\begin{aligned}
E[aX+b] &= aE[X]+b && (2.27)\\
E[X+Y] &= E[X]+E[Y] && (2.28)\\
\mathrm{Var}(X) &:= E[X^2]-E[X]^2 && (2.29)\\
\mathrm{Cov}(X,Y) &:= E[XY]-E[X]E[Y] && (2.30)\\
\mathrm{Cov}(Y,X) &= \mathrm{Cov}(X,Y) && (2.31)\\
\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+2\,\mathrm{Cov}(X,Y)+\mathrm{Var}(Y) && (2.32)\\
\mathrm{Var}(X) &= \mathrm{Cov}(X,X) && (2.33)\\
\mathrm{Var}(aX+b) &= a^2\mathrm{Var}(X) && (2.34)\\
\mathrm{Cov}(aX,bY) &= ab\,\mathrm{Cov}(X,Y) && (2.35)\\
\mathrm{Cov}(a,X) &= 0 && (2.36)\\
\mathrm{Cov}(X+Y,Z) &= \mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z) && (2.37)
\end{aligned}
$$

Übung 2.4 (Bitte alles nachrechnen.)

i) Zeigen Sie, daß (2.27) und (2.28) gilt.

Hinweis:

Verwenden Sie die Linearität des Erwartungswerts $E[X] = \sum_l x_l p_l$ = Summe der Ausprägungen mal Wahrscheinlichkeiten $p_l = P(X = x_l)$.

Für stetige Variablen gilt $E[X] = \int x f(x) dx$.

ii) Zeigen Sie (2.31-2.37).

■

Der Stoff wird in Aufgabe 2.1 vertieft.

Mit obigen Rechenregeln ausgerüstet kann man folgendes berechnen: Die Prognose $\hat{Y}$ ist erwartungstreu und es gilt

$$E[\hat{Y}] = \mu_y \tag{2.38}$$
$$\mathrm{Var}(\hat{Y}) = \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy}. \tag{2.39}$$

Der Prognosefehler

Prognosefehler

$$\tilde{Y} = Y - \hat{Y} \tag{2.40}$$

hat den Erwartungswert $E[Y - \hat{Y}] = E[Y] - E[\hat{Y}] = 0$ und die Varianz $\mathrm{Var}(Y - \hat{Y}) = \mathrm{Var}(Y) - 2\,\mathrm{Cov}(Y, \hat{Y}) + \mathrm{Var}(\hat{Y})$. Setzt man die Varianz der Prognose ein und verwendet $\mathrm{Cov}(Y, \hat{Y}) = \mathrm{Cov}(Y, \mu_y + \sigma_{yx}\sigma_{xx}^{-1}(X - \mu_x)) = \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy} = \mathrm{Var}(\hat{Y})$, so findet man

$$\mathrm{Var}(\tilde{Y}) = \sigma_{yy} - \sigma_{yx}\sigma_{xx}^{-1}\sigma_{xy} \tag{2.41}$$
$$= \mathrm{Var}(Y) - \mathrm{Var}(\hat{Y}). \tag{2.42}$$

Außerdem ist die Prognose und der Prognosefehler unkorreliert (bei Normalverteilung sogar unabhängig), da $\mathrm{Cov}(\hat{Y}, \tilde{Y}) = \mathrm{Cov}(\hat{Y}, Y - \hat{Y}) = \mathrm{Cov}(\hat{Y}, Y) - \mathrm{Var}(\hat{Y}) = 0$.

geometrische Interpretation

Eine geometrische Interpretation besagt, daß Prognose und Prognosefehler orthogonal sind (Abb. 2.3).

Dies zeigt, daß man die Varianz der Variable Y in einen erklärten Teil (MQE) und einen Residualteil (MQR) zerlegen kann, die sog. Streuungszerlegung

Streuungszerlegung

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y - \hat{Y}) \tag{2.43}$$
$$MQT = MQE + MQR \tag{2.44}$$

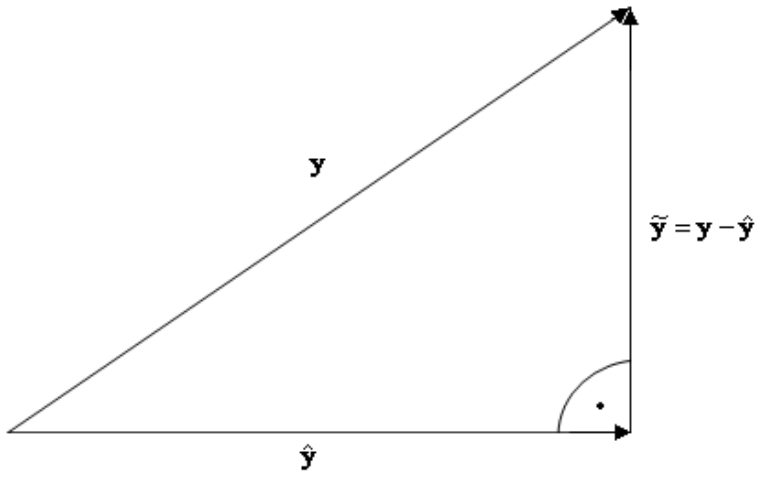


Abbildung 2.3: Geometrische Interpretation von Prognose und Prognosefehler (orthogonale Zufallsvariablen). Daher enthält die Streuungszerlegung von $\mathbf{y}$ keine Korrelationsterme $\mathrm{Cov}(\hat{\mathbf{y}}, \tilde{\mathbf{y}})$.

Beispiel 2.2 (Lineare Regression, Fortsetzung)
In diesem einfachen Fall ergibt sich (bitte nachrechnen)

$$\begin{aligned} E[\hat{Y}] &= 0 && (2.45)\\ \mathrm{Var}(\hat{Y}) &= \rho^2 && (2.46)\\ \mathrm{Var}(\tilde{Y}) &= 1-\rho^2 && (2.47)\\ \mathrm{Var}(Y) &= \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y-\hat{Y}) \\ 1 &= \rho^2 + (1-\rho^2) && (2.48) \end{aligned}$$

$\mathrm{Var}(\hat{Y})/\mathrm{Var}(Y)$ ist der Prozentanteil der Streuung von Y, der sich durch X erklären läßt. Bei $\rho = 1$ kann man also $1 = 100\%$ der Streuung erklären und die Varianz des Prognosefehlers verschwindet.

■

2.2 Die multivariate Normalverteilung

2.2.1 Gemeinsame Dichte

Der bisher diskutierte 2-dimensionale Fall kann unmittelbar auf p Variablen $X, Y, Z, W, ...$ verallgemeinert werden. In Anwendungen sind meistens $p > 2$ Variablen simultan zu betrachten. Schreibt man $\mathbf{x} =$

$[X_1, ..., X_p]' : p \times 1$ für den Zufallsvektor $\mathbf{x}$ (wird klein geschrieben, um eine Verwechslung mit der Matrix $\mathbf{X}$ zu vermeiden), so ist die p-variate Normalverteilungsdichte für $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ durch folgenden Ausdruck gegeben:

$$\phi(\mathbf{x}) = \det(2\pi\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}. \tag{2.49}$$

Hierbei ist $\mathbf{x} = [x_1, ..., x_p]'$ ein p-Vektor und

$$\boldsymbol{\mu} = E[\mathbf{x}] = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \tag{2.50}$$

sowie

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \dots & \mathrm{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \dots & \mathrm{Cov}(X_p, X_p) \end{bmatrix} \tag{2.51}$$

sind die Parameter (Vektoren und Matrizen) der p-variaten Normalverteilung. Als Abkürzung kann man auch $\sigma_{ij} = \mathrm{Cov}(X_i, X_j), i, j = 1, ..., p$ schreiben. Hierbei ist $\sigma_{ii} = \sigma_i^2 = \mathrm{Var}(X_i)$ die Varianz und $\sigma_i = \sqrt{\sigma_{ii}}$ die Standardabweichung.

Der Korrelationskoeffizient zwischen den Variablen X_i und X_j, $i, j = 1, ..., p$,

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \tag{2.52}$$

kann als Matrix $\mathbf{P}$ zusammengefasst werden. Schreibt man alle Standardabweichungen in eine Diagonalmatrix

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_p \end{bmatrix} = \mathrm{diag}(\sigma_1, ..., \sigma_p) \tag{2.53}$$

so ergibt sich die Faktorisierung

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{DPD} && (2.54) \\ \mathbf{P} &= \mathbf{D}^{-1}\mathbf{\Sigma}\mathbf{D}^{-1} && (2.55) \end{aligned}$$

mit $\mathbf{D}^{-1} = \text{diag}(\sigma_1^{-1}, ..., \sigma_p^{-1})$.

Übung 2.5 (Korrelationsmatrix)
Zeigen Sie durch Ausmultiplizieren, daß

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{\Sigma}\mathbf{D}^{-1}$$

gilt.
■

Mit dieser Notation kann man die quadratische Form im Exponent vereinfacht als $(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{z}'\mathbf{P}^{-1}\mathbf{z}$ schreiben. Hierbei sind

$$\begin{aligned} \mathbf{z} &= \mathbf{D}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ z_i &= (x_i - \mu_i)/\sigma_i, i = 1, ..., p \end{aligned}$$

(univariat) standardisierte Variablen.
Entsprechend gilt für die Determinante

$$\det(\mathbf{\Sigma}) = \det(\mathbf{D})\det(\mathbf{P})\det(\mathbf{D}) = \sigma_1^2 \dots \sigma_p^2 \det(\mathbf{P}) \qquad (2.56)$$

Hier wurde die Rechenregel

$$\begin{aligned} \det(\mathbf{AB}) &= \det(\mathbf{A})\det(\mathbf{B}) \\ \det(\text{diag}(\sigma_1, ..., \sigma_p)) &= \sigma_1\sigma_2 \dots \sigma_p = \prod_{i=1}^{p} \sigma_i \end{aligned}$$

verwendet.

Beispiel 2.3 (Bivariate Normalverteilung)
Für $p = 2$ ergibt sich ($\rho_{12} = \rho$):

$$\begin{aligned}
\mathbf{P} &= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} && (2.57)\\
\det(\mathbf{P}) &= 1 - \rho^2 && (2.58)\\
\mathbf{\Sigma} &= \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} && (2.59)\\
\det(\mathbf{\Sigma}) &= \det\left(\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right) \det\left(\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \det\left(\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right) && (2.60)\\
&= \sigma_1^2\sigma_2^2(1-\rho^2) && (2.61)
\end{aligned}$$

Die quadratische Form im Exponent ist

$$\begin{aligned}
\mathbf{z}'\mathbf{P}^{-1}\mathbf{z} &= [z_1, z_2](1-\rho^2)^{-1} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} [z_1, z_2]' && (2.62)\\
&= (1-\rho^2)^{-1}(z_1^2 - 2\rho z_1 z_2 + z_2^2) && (2.63)
\end{aligned}$$

Dies ist aber genau die explizite Form (2.9) der 2-dimensionalen Normalverteilung.
Bitte alles nachrechnen!

■

Eine multivariate Standardisierung ist durch die Transformation

multivariate Standardisierung

$$\begin{aligned}
\mathbf{x} &= \boldsymbol{\mu} + \mathbf{\Gamma}\mathbf{z} && (2.64)\\
\mathbf{z} &= \mathbf{\Gamma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) && (2.65)\\
\mathbf{z} &\sim N(\mathbf{0}, \mathbf{I}) && (2.66)
\end{aligned}$$

gegeben. Hierbei sind $\mathbf{z} = [Z_1, ..., Z_p]'$ unabhängige normalverteilte Variablen und $\mathbf{\Gamma}$, definiert durch

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}', \qquad (2.67)$$

Matrix-Wurzel

ist eine sog. Matrix-Wurzel. In der Tat gilt

$$E[\mathbf{x}] = \boldsymbol{\mu} + \boldsymbol{\Gamma} E[\mathbf{z}] = \boldsymbol{\mu} \tag{2.68}$$
$$\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Gamma}\mathrm{Var}(\mathbf{z})\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}\mathbf{I}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \boldsymbol{\Sigma} \tag{2.69}$$

Oben wurden folgende Rechenregeln benutzt:

Sind $\mathbf{x}$ und $\mathbf{y}$ irgendwelche vektoriellen Zufallsvariablen (nicht unbedingt normalverteilt) und $\mathbf{A}, \mathbf{B}, \mathbf{c}$ Konstanten (Matrizen mit passenden Dimensionen). Dann gilt (:= Definition)

Rechenregeln für Zufallsvektoren

$$E[\mathbf{A}\mathbf{x} + \mathbf{c}] = \mathbf{A}E[\mathbf{x}] + \mathbf{c} \tag{2.70}$$
$$E[\mathbf{x} + \mathbf{y}] = E[\mathbf{x}] + E[\mathbf{y}] \tag{2.71}$$
$$E[\mathbf{x}'] = (E[\mathbf{x}])' \tag{2.72}$$
$$\mathrm{Var}(\mathbf{x}) := E[\mathbf{x}\mathbf{x}'] - E[\mathbf{x}]E[\mathbf{x}'] \tag{2.73}$$
$$\mathrm{Cov}(\mathbf{x}, \mathbf{y}) := E[\mathbf{x}\mathbf{y}'] - E[\mathbf{x}]E[\mathbf{y}'] \tag{2.74}$$
$$\mathrm{Cov}(\mathbf{y}, \mathbf{x}) = \mathrm{Cov}(\mathbf{x}, \mathbf{y})' \tag{2.75}$$
$$\mathrm{Var}(\mathbf{x} + \mathbf{y}) = \mathrm{Var}(\mathbf{x}) + \mathrm{Cov}(\mathbf{x}, \mathbf{y}) + \mathrm{Cov}(\mathbf{y}, \mathbf{x}) + \mathrm{Var}(\mathbf{y}) \tag{2.76}$$
$$\mathrm{Var}(\mathbf{x}) = \mathrm{Cov}(\mathbf{x}, \mathbf{x}) \tag{2.77}$$
$$\mathrm{Var}(\mathbf{A}\mathbf{x} + \mathbf{c}) = \mathbf{A}\mathrm{Var}(\mathbf{x})\mathbf{A}' \tag{2.78}$$
$$\mathrm{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) = \mathbf{A}\ \mathrm{Cov}(\mathbf{x}, \mathbf{y})\ \mathbf{B}' \tag{2.79}$$
$$\mathrm{Cov}(\mathbf{c}, \mathbf{x}) = \mathbf{O} \tag{2.80}$$
$$\mathrm{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \mathrm{Cov}(\mathbf{x}, \mathbf{z}) + \mathrm{Cov}(\mathbf{y}, \mathbf{z}) \tag{2.81}$$

Übung 2.6 (Bitte alles nachrechnen.)

Hinweise:

i) Die Definitionen (mit :=) sind vorgegeben.

ii) Der Erwartungswert ist linear, da $E[\mathbf{x}] = \sum_l \mathbf{x}_l P(\mathbf{x} = \mathbf{x}_l)$ (diskrete Zufallsvariablen) bzw. $E[\mathbf{x}] = \int \mathbf{x} f(\mathbf{x}) d\mathbf{x}$ (stetige Zufallsvariablen).

In Komponenten kann man auch $E[X_i] = \int x_i f(x_1, ..., x_p) dx_1 ... dx_p, i = 1, ..., p$ schreiben. Der Matrix-Ausdruck $E[\mathbf{x}\mathbf{x}'] : p \times p$ ist in Komponenten $E[X_i X_j] = \int x_i x_j f(x_1, ..., x_p) dx_1 ... dx_p;\ i, j = 1, ..., p$.

■

Mit Hilfe der multivariaten Standardisierung $\mathbf{z} = \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ kann man die Normalverteilungsdichte (2.49) in der einfachen Form

$$\begin{aligned}\phi(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \det(2\pi\mathbf{I})^{-1/2}\exp\left\{-\tfrac{1}{2}\mathbf{z}'\mathbf{z}\right\}|\det(\boldsymbol{\Gamma})^{-1}| && (2.82)\\ &= \phi(\mathbf{z};\mathbf{0},\mathbf{I})|\det(\boldsymbol{\Gamma})^{-1}| && (2.83)\end{aligned}$$

schreiben ($|\cdot|$ = ist der Absolutbetrag). Hierbei wurden die Rechenregeln $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Gamma}\boldsymbol{\Gamma}')^{-1} = (\boldsymbol{\Gamma}')^{-1}\boldsymbol{\Gamma}^{-1}$, $\det(\boldsymbol{\Gamma}\boldsymbol{\Gamma}') = \det(\boldsymbol{\Gamma})\det(\boldsymbol{\Gamma}') = \det(\boldsymbol{\Gamma})^2$ und $\det(\boldsymbol{\Gamma}) = \det(\boldsymbol{\Gamma}')$ benutzt. Faßt man die Multiplikation von $(\mathbf{x}-\boldsymbol{\mu})$ mit $\boldsymbol{\Gamma}^{-1}$ als Drehung und Stauchung auf, so läßt sich die Normalverteilungsdichte $\phi(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ als Produkt von unabhängigen Normalverteilungsdichten in einem gedrehten und gestauchten Koordinatensystem auffassen, d.h. $\phi(\mathbf{z};\mathbf{0},\mathbf{I}) = \prod_{i=1}^{p}\phi(z_i;0,1)$ (vgl. Abb. 2.1).
Im univariaten Fall ergibt sich aus (2.83) die bekannte Formel $\phi(x;\mu,\sigma^2) = \phi(z;0,1)\sigma^{-1}$, $z = (x-\mu)/\sigma$.

2.2.2 Multivariate Randverteilungen und bedingte Normalverteilung

Die obigen Überlegungen zur Randverteilung und zur bedingten Verteilung (Abs. 2.1.2) übertragen sich analog auf den p-variaten Fall.
Teilt man den Zufallsvektor $\mathbf{z}' = [\mathbf{x}',\mathbf{y}']$ in Teilvektoren der Dimension p und q auf, so ist

$$\mathbf{z} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy}\\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy}\end{bmatrix}\right) \tag{2.84}$$

durch die gemeinsame Dichte $\phi(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x},\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})$ charakterisiert ($\mathbf{x} = [x_1,...,x_p]'$, $\mathbf{y} = [y_1,...,y_q]'$). Hierbei ist $\boldsymbol{\Sigma}$ eine Blockmatrix der Dimension $(p+q)\times(p+q)$.

2.2.2.1 Randverteilungen

Randverteilungen

Auch im allgemeinen Fall ergibt sich das Resultat, daß die Randverteilungen $\phi(\mathbf{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x)$, $\phi(\mathbf{y};\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_y)$ multivariate Normalverteilungsdichten sind. Man muß also aus der vollen Kovarianzmatrix nur die entsprechenden Zeilen und Spalten herausschneiden, um die Randdichte zu erhalten. Dies ist eine der vielen überraschenden und (für die Benutzung) einfachen Eigenschaften der Normalverteilung.

2.2.2.2 Bedingte Verteilungen

Mit Hilfe der Bayes-Formel

Bayes-Formel

$$f(\mathbf{y}|\mathbf{x}) = \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{x})} \tag{2.85}$$

läßt sich die bedingte Dichte wieder aus der gemeinsamen Dichte und der Randdichte berechnen (dies gilt ganz allgemein). Bei Normalverteilungen gilt das Resultat:

$$\mathbf{y}|\mathbf{x} \sim N(\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x}), \tag{2.86}$$

wobei die bedingten Erwartungswerte und Kovarianzen durch

Satz von der Normalkorrelation

$$\boldsymbol{\mu}_{y|x} = E[\mathbf{y}|\mathbf{x}] = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \tag{2.87}$$

$$\boldsymbol{\Sigma}_{y|x} = \mathrm{Var}[\mathbf{y}|\mathbf{x}] = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \tag{2.88}$$

gegeben sind. Der bedingte Erwartungswert

optimale Prognose

$$\hat{\mathbf{y}} = E[\mathbf{y}|\mathbf{x}] \tag{2.89}$$

(optimale Prognose von $\mathbf{y}$ durch $\mathbf{x}$) ist auch bei multivariaten Normalverteilungen eine *lineare Funktion* des Vektors $\mathbf{x}$. Dies motiviert auch in multivariaten Fragestellungen die Benutzung der linearen Regressionsanalyse. Die bedingte Kovarianzmatrix (2.88) hängt wie im bivariaten Fall nicht von den Daten ab.

Man kann auch in der üblichen Notation schreiben

$$E[\mathbf{y}|\mathbf{x}] = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{x} \tag{2.90}$$

$$\boldsymbol{\alpha} = \boldsymbol{\mu}_y - \boldsymbol{\beta}\boldsymbol{\mu}_x \tag{2.91}$$

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}. \tag{2.92}$$

Beispiel 2.4 (Lineare multiple Regression)

Setzt man
$\mathbf{z} = [X_1, X_2, Y]'$,
$\boldsymbol{\mu} = [0, 0, 0]'$ und $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & \rho_1 \\ 0 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{bmatrix}$
(standardisierte Variable, unkorrelierte Regressoren X_1, X_2), so ergibt sich

$$\begin{aligned}
\boldsymbol{\alpha} &= [0, 0, 0]' && (2.93)\\
\boldsymbol{\beta} &= \begin{bmatrix} \rho_1 & \rho_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \rho_1 & \rho_2 \end{bmatrix} && (2.94)\\
E[Y|X_1, X_2] &= \rho_1 X_1 + \rho_2 X_2 && (2.95)\\
\mathrm{Var}[Y|X_1, X_2] &= 1 - \begin{bmatrix} \rho_1 & \rho_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} && (2.96)\\
&= 1 - (\rho_1^2 + \rho_2^2) && (2.97)
\end{aligned}$$

Da eine Varianz ≥ 0 sein muß, gilt $\rho_1^2 + \rho_2^2 \leq 1$.
■

Die Prognose $\hat{\mathbf{y}}$ ist erwartungstreu und es gilt

$$\begin{aligned}
E[\hat{\mathbf{y}}] &= \boldsymbol{\mu}_y && (2.98)\\
\mathrm{Var}(\hat{\mathbf{y}}) &= \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}. && (2.99)
\end{aligned}$$

Prognosefehler

Der Prognosefehler

$$\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{y}} \qquad (2.100)$$

hat den Erwartungswert $E[\mathbf{y} - \hat{\mathbf{y}}] = E[\mathbf{y}] - E[\hat{\mathbf{y}}] = 0$ und die Varianz $\mathrm{Var}(\mathbf{y} - \hat{\mathbf{y}}) = \mathrm{Var}(\mathbf{y}) - \mathrm{Cov}(\mathbf{y}, \hat{\mathbf{y}}) - \mathrm{Cov}(\hat{\mathbf{y}}, \mathbf{y}) + \mathrm{Var}(\hat{\mathbf{y}})$.
Setzt man die Varianz der Prognose ein und verwendet

$$\begin{aligned}
\mathrm{Cov}(\mathbf{y}, \hat{\mathbf{y}}) &= \mathrm{Cov}(\mathbf{y}, \boldsymbol{\beta}\mathbf{x}) = \mathrm{Cov}(\mathbf{y}, \mathbf{x})\boldsymbol{\beta}' && (2.101)\\
&= \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy} = \mathrm{Cov}(\mathbf{y}, \hat{\mathbf{y}})' && (2.102)\\
&= \mathrm{Var}(\hat{\mathbf{y}}), && (2.103)
\end{aligned}$$

so findet sich

$$\mathrm{Var}(\tilde{\mathbf{y}}) = \mathbf{\Sigma}_{yy} - \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xy}. \tag{2.104}$$

Außerdem ist die *Prognose und der Prognosefehler unkorreliert* (bei Normalverteilung sogar unabhängig), da $\mathrm{Cov}(\hat{\mathbf{y}}, \tilde{\mathbf{y}}) = \mathrm{Cov}(\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}) = \mathrm{Cov}(\hat{\mathbf{y}}, \mathbf{y}) - \mathrm{Var}(\hat{\mathbf{y}}) = 0$.

geometrische Interpretation

In einer geometrischen Interpretation kann man analog zum bivariaten Fall sagen, daß Prognose und Prognosefehler *orthogonal* sind (Abb. 2.3). Dies zeigt, daß man die Varianz der Variable $\mathbf{y}$ in einen erklärten Teil (MQE) und einen Residualteil (MQR) zerlegen kann, die sog. Streuungszerlegung

Streuungszerlegung

$$\mathrm{Var}(\mathbf{y}) = \mathrm{Var}(\hat{\mathbf{y}}) + \mathrm{Var}(\mathbf{y} - \hat{\mathbf{y}}) \tag{2.105}$$

$$MQT = MQE + MQR \tag{2.106}$$

Beispiel 2.5 (Lineare multiple Regression, Fortsetzung)
In diesem einfachen Fall ergibt sich (bitte nachrechnen)

$$E[\hat{Y}] = 0 \tag{2.107}$$

$$\mathrm{Var}(\hat{Y}) = \rho_1^2 + \rho_2^2 \tag{2.108}$$

$$\mathrm{Var}(\tilde{Y}) = 1 - \rho_1^2 - \rho_2^2 \tag{2.109}$$

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y - \hat{Y})$$

$$1 = \rho_1^2 + \rho_2^2 + (1 - \rho_1^2 - \rho_2^2) \tag{2.110}$$

Bei $\rho_1^2 + \rho_2^2 = 1$ kann man also $1 = 100\%$ der Streuung von Y durch X_1 und X_2 erklären.

■

2.2.3 Simulation von multivariat normalverteilten Zufalls-Vektoren

Die Standardisierungs-Transformation

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{\Gamma}\mathbf{z} \tag{2.111}$$

$$\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}) \tag{2.112}$$

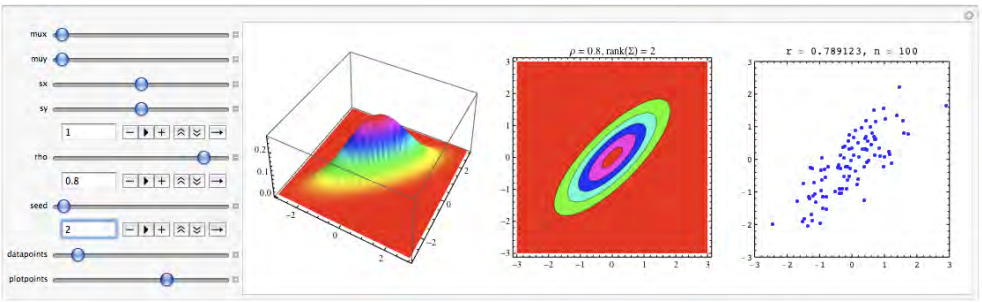


Abbildung 2.4: Simulation. Bivariate Normalverteilung. $\rho_{xy} = 0.8$ $\sigma_x = 1$, $\sigma_y = 1$. Random seed = 2.

ermöglicht die einfache Simulation von multivariat normalverteilten Zahlen $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Man benötigt nur einen Zufallsgenerator für univariate unabhängige Zahlen $z_i \sim N(0,1)$ sowie eine Faktorisierung der Kovarianz-Matrix $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$. Die in Abb. 2.1 simulierten Daten wurden so erzeugt (gleiche Zufallszahlen z_i, aber unterschiedliche Erwartungswerte und Kovarianzen bei $\mathbf{x}$). Nimmt man eine andere *random seed* (Initialisierung des Zufallsgenerators, vgl. Abb. 2.4), so ergeben sich unterschiedliche Streudiagramme, wobei die Daten aber aus der selben Verteilung gezogen wurden.

Beispiel 2.6 (Cholesky-Zerlegung)

Die Korrelations-Matrix $\mathbf{P} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ läßt sich als Produkt $\boldsymbol{\Pi} \cdot \boldsymbol{\Pi}'$ mit $\boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$ (untere Dreiecksmatrix) schreiben.

Für die Kovarianzmatrix gilt deshalb $\boldsymbol{\Sigma} = \mathbf{DPD} = \mathbf{D}\boldsymbol{\Pi} \cdot \boldsymbol{\Pi}'\mathbf{D} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$.

Im Beispiel ergibt sich

$$\boldsymbol{\Gamma} = \mathbf{D}\boldsymbol{\Pi} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} = \begin{bmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{bmatrix}.$$

Übung 2.7

Zeigen Sie explizit, daß man daraus wieder $\boldsymbol{\Sigma}$ erhält.

■

Für $\rho = 0.8, \sigma_1 = \sigma_2 = 1$ ergibt sich explizit die Cholesky-Wurzel

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 \\ 0.8 & 0.6 \end{bmatrix}.$$

■

Beispiel 2.7 (Simulation mit SPSS)
In Komponenten lautet die Standardisierungstransformation

$$x_1 = \mu_1 + \Gamma_{11} z_1 + \Gamma_{12} z_2 \tag{2.113}$$
$$x_2 = \mu_2 + \Gamma_{21} z_1 + \Gamma_{22} z_2 \tag{2.114}$$

Eine $N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}\right)$-Verteilung kann daher durch

$$x_1 = 1 + 1.0 z_1 + 0.0 z_2 \tag{2.115}$$
$$x_2 = 1 + 0.8 z_1 + 0.6 z_2 \tag{2.116}$$
$$\tag{2.117}$$

erzeugt werden.
Im Rahmen von SPSS lassen sich normalverteilte Zufallsvariablen mit Hilfe des Menüs

```
Transformieren/Variable berechnen
```

erzeugen (siehe Abb. 2.5-2.6). Die simulierten Variablen sind im Datensatz `Zufall.sav` enthalten.
■

Übung 2.8 (Simulation mit SPSS)
Erzeugen Sie bitte weitere Daten:

Starten Sie dazu SPSS zunächst mit einem leeren Datenblatt. Tragen Sie dann in der Variablenansicht zwei numerische Variablen mit dem Messniveau „Skala“ ein (z_1 und z_2). In der Datenansicht muss für beide Variablen im N-ten Datensatz (Zeile) ein beliebiger Wert eingetragen werden, um die Anzahl der Zufallsvariablen zu bestimmen.

i) Generieren Sie jetzt $N = 5$ normalverteilte Zufallsvariablen mit den im Beispiel angegebenen Parametern. Zeichnen Sie ein Streudiagramm (Menüleiste: Diagramme) sowohl von den standard-normalverteilten als auch von den transformierten Zufallszahlen. Wiederholen Sie mit $N = 50$.

ii) Variieren Sie jetzt die einzelnen Parameter. Verwenden Sie dabei weiterhin $N = 50$. Starten Sie mit $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 10$ und

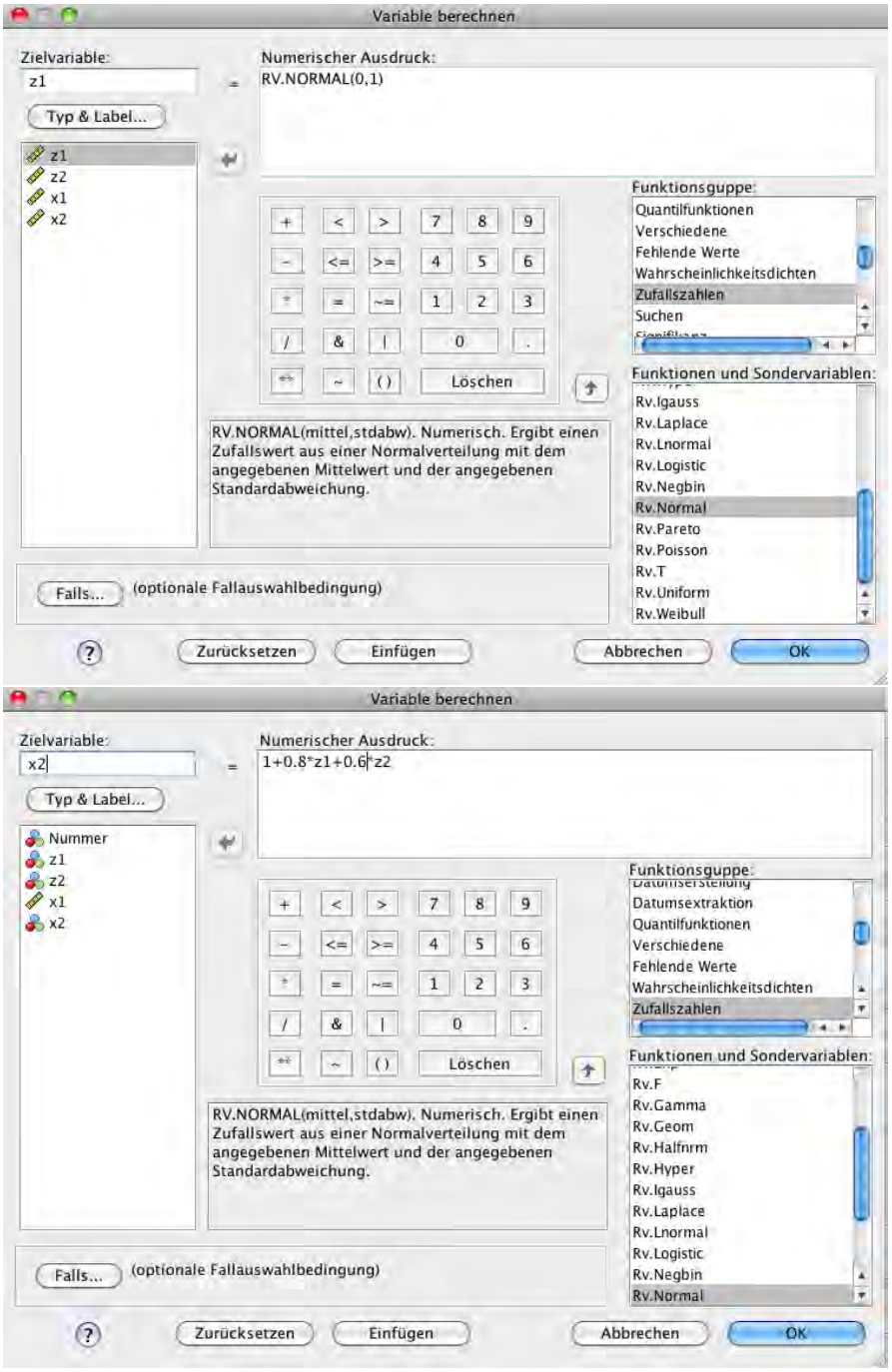
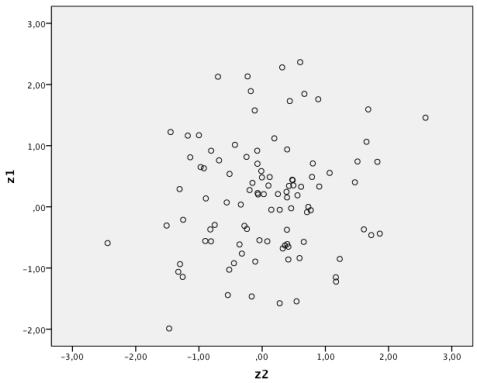
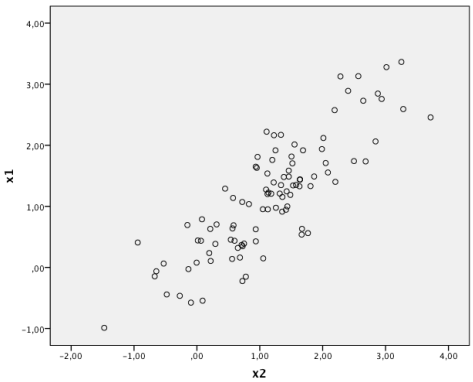


Abbildung 2.5: Simulation mit SPSS. Bivariate Normalverteilung. $\rho = 0.8$ $\sigma_1 = 1$, $\sigma_2 = 1$, $\mu_1 = 1, \mu_2 = 1$. Die Koeffizienten der Cholesky-Wurzel müssen von Hand eingegeben werden.





Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
z1	,1553	,93182	100
z2	,0720	,88577	100
x1	1,1553	,93182	100
x2	1,1674	,97306	100

Korrelationen

		z1	z2	x1	x2
z1	Korrelation nach Pearson	1	,137	1,000**	,841**
	Signifikanz (2-seitig)		,173	,000	,000
	N	100	100	100	100
z2	Korrelation nach Pearson	,137	1	,137	,651**
	Signifikanz (2-seitig)	,173		,173	,000
	N	100	100	100	100
x1	Korrelation nach Pearson	1,000**	,137	1	,841**
	Signifikanz (2-seitig)	,000	,173		,000
	N	100	100	100	100
x2	Korrelation nach Pearson	,841**	,651**	,841**	1
	Signifikanz (2-seitig)	,000	,000	,000	
	N	100	100	100	100

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

Abbildung 2.6: Oben: Streudiagramme. Links unkorrelierte Variaben $z1, z2$, rechts mit Korrelation $\rho = 0.8$. Unten: Deskriptive Statistiken.

$\rho = 0.8$.
Verändern Sie als erstes den Mittelwert zu $\mu_1 = 5$, $\mu_2 = 0$
Variieren Sie danach die Korrelation: $\rho = 0.6$, $\rho = 0$ und $\rho = -0.8$.
Ändern Sie die Standardabweichungen zu $\sigma_1 = 10$ und $\sigma_2 = 10$.

■

Der Stoff wird in Aufgabe 2.2 vertieft.

2.3 Schätzung der Parameter aus Daten

Analog zur univariaten Statistik ist es plausibel, die Mittelwerte

$$\bar{x}_i = N^{-1} \sum_{n=1}^{N} x_{ni} \tag{2.118}$$

und die empirischen Kovarianzen

$$s_{ij} = (N-1)^{-1} \sum_{n=1}^{N} (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j) \tag{2.119}$$

$$s_{ii} = s_i^2 = (N-1)^{-1} \sum_{n=1}^{N} (x_{ni} - \bar{x}_i)^2 \tag{2.120}$$

als Schätzungen für die unbekannten Parameter $\mu_i, \sigma_{ij}; i, j = 1, ..., p$ zu verwenden. Abkürzend kann man $\bar{\mathbf{x}}$ und $\mathbf{S}$ schreiben. Explizit gilt

Mittelwert

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \tag{2.121}$$

Stichprobenvarianz

$$\mathbf{S} = \frac{1}{N-1} \sum_{n=1}^{N} (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})'. \tag{2.122}$$

Mittelwert und Stichprobenvarianz kann mit Hilfe der Datenmatrix $\mathbf{X}$: $N \times p$ auch in der Form

$$\bar{\mathbf{x}} = \frac{1}{N}\mathbf{X}'\mathbf{1} \tag{2.123}$$

$$\mathbf{S} = \frac{1}{N-1}\left(\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n' - N\bar{\mathbf{x}}\bar{\mathbf{x}}'\right) \tag{2.124}$$

$$= \frac{1}{N-1}\left(\mathbf{X}'\mathbf{X} - \frac{1}{N}\mathbf{X}'\mathbf{1}\mathbf{1}'\mathbf{X}\right) \tag{2.125}$$

$$= \frac{1}{N-1}\mathbf{X}'\left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}'\right)\mathbf{X} \tag{2.126}$$

$$\mathbf{S} = \frac{1}{N-1}\mathbf{X}'\mathbf{H}\mathbf{X} \tag{2.127}$$

geschrieben werden. Die Matrix

$$\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}' : N \times N \tag{2.128}$$

Zentrierungsmatrix

wird als Zentrierungsmatrix bezeichnet. Sie bewirkt, daß bei einer Datenmatrix der Mittelwert $\mathbf{MX} = \frac{1}{N}\mathbf{1}\mathbf{1}'\mathbf{X}$ abgezogen wird.
Die Stichproben-Kovarianzmatrix enthält $p(p+1)/2$ Varianzen und Kovarianzen. Eine einzelne Größe für die Streuung kann man durch

1. $\det(\mathbf{S})$ (verallgemeinerte Varianz) und
2. $\mathrm{tr}(\mathbf{S})$ (totale Varianz)

verallgemeinerte Varianz

totale Varianz

gewinnen.
Die Stichproben-Korrelation $r_{ij} = \frac{s_{ij}}{s_i s_j}$ kann zu einer Korrelations-Matrix

Korrelationsmatrix

$$\mathbf{R} = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1}, \tag{2.129}$$

$\mathbf{D} = \text{diag}(s_1, ..., s_p)$ zusammengefaßt werden.

2.4 Maximum-Likelihood-Schätzung

Es ist sinnvoll, Schätzer mit Hilfe von Schätzprinzipien zu konstruieren. Das Prinzip der maximalen Likelihood (ML) kann auch auf den multivariaten Fall übertragen werden (Mardia et al., 1979, Kap. 4). Allgemein ist die Likelihood-Funktion[4] durch

Likelihood-Funktion

$$L(\boldsymbol{\theta}) := f(\mathbf{x}_1, \ldots, \mathbf{x}_N | \boldsymbol{\theta}) \tag{2.130}$$

gegeben, wobei die gemeinsame Dichtefunktion aller Daten $\mathbf{x}_1, \ldots, \mathbf{x}_N$ als Funktion eines Parameter-Vektors $\boldsymbol{\theta} = [\theta_1, ..., \theta_u]'$ aufgefaßt wird.

Maximum-Likelihood(ML)-Schätzer

Man bestimmt nun das Maximum der Likelihood-Funktion und nennt den maximierenden zufälligen $\boldsymbol{\theta}$-Wert Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\theta}}_{ML}$.
Der Begriff *Likelihood* wurde von Fisher eingeführt, da f als Funktion von $\boldsymbol{\theta}$ keine Wahrscheinlichkeit ist.

[4] Der senkrechte Strich in der Dichte f kann nur dann als Bedingung gelesen werden, wenn man, wie in der Bayes-Statistik, den Parameter als Größe mit Verteilung auffaßt. Hier ist die Notation als Dichte, berechnet mit dem Wert $\boldsymbol{\theta}$, zu verstehen (vgl. z.B. Fahrmeir et al., 2007, Kap. 9.3).

Beispiel 2.8 (ML-Schätzer bei univariater Normalverteilung)
Annahme:
Die univariaten Daten $X_1, \ldots, X_N$ sind unabhängig und identisch normalverteilt

$$X_n \sim N(\mu, \sigma^2). \tag{2.131}$$

Die gemeinsame Dichtefunktion aller Daten ist somit:

$$f(x_1, \ldots, x_N|\mu, \sigma^2) = \phi(x_1; \mu, \sigma^2) \ldots \phi(x_N; \mu, \sigma^2) \tag{2.132}$$

$$\phi(x; \mu, \sigma^2) := (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \tag{2.133}$$

Die Likelihood-Funktion

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2}\sum_n \frac{(X_n-\mu)^2}{\sigma^2}\right\} \tag{2.134}$$

$$= (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{N}{2\sigma^2}\left(S_*^2 + (\bar{X}-\mu)^2\right)\right\} \tag{2.135}$$

ist die Dichtefunktion $f(X_1, \ldots, X_N|\mu, \sigma^2)$, aufgefaßt als Funktion der Parameter und mit eingesetzten Daten.
Hier wurde der Mittelwert $\bar{X}$ und die Stichprobenvarianz (mittlere quadratische Abweichung)

$$S_*^2 = \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2 = \frac{N-1}{N}S^2 \tag{2.136}$$

substituiert

Übung 2.9
Führen Sie diese Umrechnung explizit durch.
Hinweis:
$(X_n - \mu) = (X_n - \bar{X} + \bar{X} - \mu)$. ■

Die Likelihood hängt also von den Daten nur über die (suffizienten) Statistiken $\bar{X}, S_*^2$ ab.[5]

[5] Die Statistik $\mathbf{t} = \mathbf{t}(\mathbf{X}); \mathbf{X} = [X_1, ..., X_N]$ wird als *suffizient* bezeichnet, wenn sich die Likelihood als $L(\boldsymbol{\theta}, \mathbf{X}) = g(\boldsymbol{\theta}, \mathbf{t})h(\mathbf{X})$ faktorisieren läßt. Dann hängt der ML-Schätzer von den Daten nur durch $\mathbf{t}(\mathbf{X})$ ab, also hier von $\bar{X}$ und S_*^2.

Das Maximum von L ist durch die Parameter-Schätzwerte

$$\hat{\mu} = \bar{X} \tag{2.137}$$
$$\widehat{\sigma^2} = S_*^2 \tag{2.138}$$

gegeben.

Übung 2.10 (ML-Schätzwerte)
Zeigen Sie dies.
Hinweis: Leiten Sie die Likelihood nach μ und σ^2 ab und setzen Sie die Ableitungen 0. Einfacher ist es, den Logarithmus $l = \log L$ (die sog. Log-Likelihood) abzuleiten. ■

Man erhält also bei der Maximum-Likelihood-Schätzung i.A. verzerrte Schätzer (hier $E[S_*^2] = \frac{N-1}{N} E[S^2] = \frac{N-1}{N}\sigma^2 \neq \sigma^2$).
■

Der Stoff wird in Aufgabe 2.3 vertieft.

Bei multivariater Normalverteilung der unabhängigen Daten $\mathbf{x}_1, ..., \mathbf{x}_N$ ergibt sich der Ausdruck

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= [\det(2\pi\boldsymbol{\Sigma})]^{-N/2} \exp\left\{-\frac{1}{2}\sum_n (\mathbf{x}_n - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})\right\} \\ &= [\det(2\pi\boldsymbol{\Sigma})]^{-N/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_n (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})'\right]\right\} \\ &= [\det(2\pi\boldsymbol{\Sigma})]^{-N/2} \exp\left\{-\frac{N}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{S}_* + (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})')\right]\right\} \end{aligned} \tag{2.139}$$

der analog zu (2.134) ist. Durch Maximierung nach $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ ergeben sich die Maximum-Likelihood-Schätzer

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \tag{2.140}$$

Mittelwert

$$\widehat{\boldsymbol{\Sigma}} = \mathbf{S}_* = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})' = \frac{N-1}{N}\mathbf{S} \tag{2.141}$$

Stichprobenvarianz (ML)

Man erhält also wie im univariaten Fall einen Varianzschätzer, der den Vorfaktor $1/N$ hat und verzerrt ist, d.h. $E[\mathbf{S}_*] = \frac{N-1}{N}\boldsymbol{\Sigma}$.

Übung 2.11 (Umformung der Likelihood-Funktion)
Die Ausdrücke in (2.139) ergeben sich mit Hilfe folgender Rechenregeln:

i) In der ersten Zeile wurde das Produkt der einzelnen Exponential-Terme zu einer Summe im Exponent zusammengefasst, d.h. $\exp(a)\exp(b) = \exp(a+b)$ etc.

ii) Die quadratische Form $(\mathbf{x}_n - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$ wurde als Spur (= engl. trace) $\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})']$ geschrieben. Dies ermöglicht, die Summe über n auszuführen und $\boldsymbol{\Sigma}^{-1}$ auszuklammern.

iii) In $\mathbf{x}_n - \boldsymbol{\mu} = \mathbf{x}_n - \bar{\mathbf{x}} + \bar{\mathbf{x}} - \boldsymbol{\mu}$ wurde der Mittelwert eingeschoben. Der Produkt-Term $(\mathbf{x}_n - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \boldsymbol{\mu})'$ verschwindet bei Summation über n.

Bitte alles explizit nachrechnen!

■

Aufgrund der Wichtigkeit der Spur hier noch die Definition und einige Rechenregeln:

Spur

$\mathbf{A} = (A_{ij})$ ist eine quadratische Matrix.

Die Spur (trace) ist die Summe der Diagonalelemente

$$\mathrm{tr}(\mathbf{A}) \quad := \quad \sum_i A_{ii} \tag{2.142}$$

Es gilt

$$\mathrm{tr}(\mathbf{AB}) \quad = \quad \mathrm{tr}(\mathbf{BA}) \tag{2.143}$$

und somit für die quadratische Form $\mathbf{x}'\mathbf{Ax}$

$$\mathbf{x}'\mathbf{Ax} \quad = \quad \mathrm{tr}(\mathbf{x}'\mathbf{Ax}) = \mathrm{tr}(\mathbf{Axx}') \tag{2.144}$$

Übung 2.12 (Spur)
Bitte (2.144) explizit nachrechnen!

Hinweis:
Schreiben Sie $\mathbf{C} = \mathbf{AB}$ in Komponenten: $C_{ik} = \sum_j A_{ij}B_{jk}$. Daher ist $\mathrm{tr}(\mathbf{C}) = \sum_i C_{ii} = \sum_i \sum_j A_{ij}B_{ji} = \sum_j \sum_i B_{ji}A_{ij} = \mathrm{tr}(\mathbf{BA})$.
■

Kapitel 3

Tests und Konfidenzintervalle

3.1 Allgemeine Bemerkungen

Im univariaten Fall konnten Hypothesen über skalare Parameter, etwa $H_0 : \mu = \mu_0$, mit Hilfe von Gauß- oder t-Tests überprüft werden. Im multivariaten Fall hat man das Problem, daß mehrere Mittelwerte $\mu_1, ..., \mu_p$ gleichzeitig (simultan) getestet werden müssen. In Kap. 1.5 wurde erwähnt, daß bei der Untersuchung auf korrelative Zusammenhänge mehrere Korrelationen gleichzeitig auf Signifikanz (d.h. $H_0 : \rho_{ij} = 0$ ablehnen) untersucht werden müssen. Setzt man die Nullhypothese aus dem Schnitt mehrerer univariater Hypothesen zusammen, so stellt sich bei einer Abfolge von univariaten Tests heraus, daß das simultane Signifikanzniveau (Fehler 1. Art) $\alpha^* = P(H_0 \text{ ablehnen}|H_0 \text{ richtig})$ größer als das α der Einzeltests werden kann. Eine Adjustierung der Einzeltests führte zu einer recht einfachen Lösung (Bonferroni-Ungleichung), jedoch kann der Test konservativ sein (bestehende Unterschiede werden durch die Testprozedur nicht entdeckt, da das gesamte Signifikanzniveau α^* zu klein ist.)

Daher ist es sinnvoll, multivariate Tests durchzuführen, die das geforderte simultane Signifikanzniveau (Fehler 1. Art) $\alpha^* = P(H_0 \text{ ablehnen}|H_0 \text{ richtig}) = \alpha$ exakt einhalten.

Hypothesen

Als *Hypothesen* werden im folgenden Teilräume des Parameter-Raums $\boldsymbol{\Theta}$ bezeichnet.

Etwa ist $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ ein einzelner Punkt im u-dimensionalen Raum $\boldsymbol{\Theta} = \mathbb{R}^u$.

Man kann auch allgemeiner $H_0 = \{\boldsymbol{\theta} | \boldsymbol{\theta} \in \boldsymbol{\Theta}_0\} \subset \boldsymbol{\Theta}$ schreiben, also die Menge der Parameterwerte, die eine bestimmte Bedingung erfüllen, z.B. ein einzelner Punkt $\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta}_0\} = \{[\theta_{01}, ..., \theta_{0u}]'\}$.

Systematische Prinzipien zur Konstruktion von Tests sind das Likelihood-Quotienten- sowie das Union-Intersection-Prinzip. Im ersten Fall wird der Likelihood-Quotient (LQ)

Likelihood-Quotient

$$\lambda = \frac{L(\hat{\boldsymbol{\theta}}_0; \mathbf{X})}{L(\hat{\boldsymbol{\theta}}_1; \mathbf{X})} \tag{3.1}$$

unter der H_0 sowie der H_1 berechnet. Hierbei sind $\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_1$ die Parameter-Schätzwerte, bei denen die Likelihood unter den Hypothesen maximal wird (vgl. Abs. 2.4). Daten, die eher für die H_1 sprechen, führen also zu kleinen Werten der LQ-Statistik.

Union-Intersection-Prinzip

Beim Union-Intersection-Prinzip wird die multivariate Nullhypothese als Schnittmenge

$$H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}} \tag{3.2}$$

univariater Hypothesen (Komponenten) geschrieben. Etwa ist H_{0a} : $\mathbf{a}'\boldsymbol{\mu} = \mathbf{a}'\boldsymbol{\mu}_0$ eine solche Hypothese.

Wählt man als $\mathbf{a} = \mathbf{e}_1 = [1, 0]'$ (Einheitsvektor in x-Richtung), so ist $H_{0\mathbf{e}_1}$: $\mathbf{e}_1'\boldsymbol{\mu} = \mathbf{e}_1'\boldsymbol{\mu}_0$ bzw. $\mu_1 = \mu_{10}$ eine univariate Hypothese für die 1. Komponente. Die Wahl $\mathbf{a} = [1, 1]'$ führt zu einer Linearkombination $\mu_1 + \mu_2 = \mu_{10} + \mu_{20}$. Derartige Linearkombinationen sind oft den Daten besser angepaßt, wenn die Variablen korreliert sind.

Die Nullhypothese wird beibehalten, wenn alle Komponenten $H_{0\mathbf{a}}$ beibehalten werden. Dagegen führt die Ablehnung schon einer Komponenten-Hypothese $H_{0\mathbf{a}}$ zur Ablehnung der H_0. Der Test von $H_{0\mathbf{a}}$ wird mit einer geeigneten univariaten Teststatistik durchgeführt.

Beispiel 3.1 (Mittelwerts-Test, $\boldsymbol{\Sigma}$ bekannt)

Will man die Hypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = [6, 6]'$ testen, so kann man die Komponenten $H_{0\mathbf{a}} : \mathbf{a}'\boldsymbol{\mu} = \mathbf{a}'\boldsymbol{\mu}_0 = 6a_1 + 6a_2$ einzeln abprüfen. Der Mittelwert $\bar{\mathbf{x}}$ ist normalverteilt $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}/N)$. Daher gilt für die Projektionen auf den Vektor $\mathbf{a}$: $\mathbf{a}'\bar{\mathbf{x}} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}/N)$.

Somit ist $Z_{\mathbf{a}} = (\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)/\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}/N}$ standardnormalverteilt. H_0 wird nur beibehalten, wenn alle Einzeltests nicht signifikant sind, also $|Z_{\mathbf{a}}| \leq z(1 - \alpha/2)$. Maximiert man über alle denkbaren $\mathbf{a}$, so muß auch

$\max_{\mathbf{a}} |Z_{\mathbf{a}}| \leq z(1-\alpha/2)$ gelten. Die Maximierung führt direkt zur Teststatistik nach dem Union-Intersection-Prinzip.

■

3.2 Ein-Stichproben-Fall

3.2.1 Test für den Erwartungswert $\boldsymbol{\mu}$ ($\boldsymbol{\Sigma}$ bekannt)

Dies ist die direkte Verallgemeinerung des Gauß-Tests bei univariaten normalverteilten Stichproben.

Bekanntlich ist der Mittelwert $\bar{X}$ der unabhängigen Daten $X_n \sim N(\mu, \sigma^2), n = 1, ..., N$ unter der Nullhypothese $\mu = \mu_0$ auch normalverteilt $N(\mu_0, \sigma^2/N)$. Quadrate von normalverteilten Größen sind χ^2-verteilt. Schreibt man $t^2 = (\bar{X} - \mu_0)^2/(\sigma^2/N) = Z^2$ mit der (unter H_0) standardisierten Variable $Z = (\bar{X} - \mu_0)/\sqrt{\sigma^2/N}$, so kann man die $\chi^2(1)$-Verteilung zum Hypothesentest verwenden.

Analog schreibt man im multivariaten Fall (p-dimensionale Daten):

1. Daten: $\mathbf{x}_n \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), n = 1, ..., N$
   (unabhängig und identisch verteilt).
2. Hypothesen: $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ gegen $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.
3. Teststatistik: $T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi^2(p)$ unter H_0.
4. Kritischer Wert: $\chi^2(1-\alpha, p)$.
5. Testentscheidung: Falls $T^2 > \chi^2(1-\alpha, p)$, H_0 ablehnen.

In der Tat ist $\mathbf{z} = \sqrt{N}\boldsymbol{\Gamma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$ (vgl. (2.65)) ein normalverteilter Zufallsvektor mit $\mathrm{Var}(\mathbf{z}) = N\boldsymbol{\Gamma}^{-1}(\boldsymbol{\Sigma}/N)(\boldsymbol{\Gamma}^{-1})' = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'(\boldsymbol{\Gamma}^{-1})' = \mathbf{I}$. Hierbei wurde $\mathbf{I} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}'(\boldsymbol{\Gamma}^{-1})'$ und $\mathrm{Var}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = \boldsymbol{\Sigma}/N$ eingesetzt.

Daher sind die Komponenten von $\mathbf{z}$ standardnormalverteilt $N(0,1)$ und es gilt $T^2 = \mathbf{z}'\mathbf{z} = \sum z_i^2 \sim \chi^2(p)$.

Beispiel 3.2 (Mittelwerts-Test)

Die Variablen `Lifeexpectancy, Selfreportedhealth` sollen bivariat auf den Erwartungswert $\boldsymbol{\mu}_0 = [6, 6]'$ getestet werden. Abb. 3.1 kann man die Mittelwerte sowie die empirischen Kovarianzen entnehmen. Die wahre Kovarianzmatrix $\boldsymbol{\Sigma}$ ist nicht bekannt. Wir nehmen daher zunächst an, daß $\boldsymbol{\Sigma}$ konstant und numerisch gleich der empirischen Kovarianzmatrix $\mathbf{S}$ ist (vgl. aber Abs. 3.2.3).

Somit hat man $N = 34$, $\bar{\mathbf{x}} = [6.154, 6.447]'$, $\boldsymbol{\Sigma} = \begin{bmatrix} 7.701 & 2.647 \\ 2.647 & 6.397 \end{bmatrix}$.

Die Teststatistik ist

$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi^2(2)$.

Man benötigt noch die Inverse der Kovarianz, d.h.

$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.151 & -0.063 \\ -0.063 & 0.182 \end{bmatrix}$.

Dies ergibt sich aus der Formel (2.10). Die Determinante ist $\det(\boldsymbol{\Sigma}) = 42.257$.

Insgesamt hat man also (die Realisation von T^2 wird klein geschrieben)

$t^2 = 34\,[0.154, 0.447] \begin{bmatrix} 0.151 & -0.063 \\ -0.063 & 0.182 \end{bmatrix} \begin{bmatrix} 0.154 \\ 0.447 \end{bmatrix} = 1.067.$

Der kritische Wert ist aber $\chi^2(0.95, 2) = 5.991$.

Damit muß $H_0 : \boldsymbol{\mu} = [6, 6]'$ auf dem 5%-Niveau beibehalten werden.

■

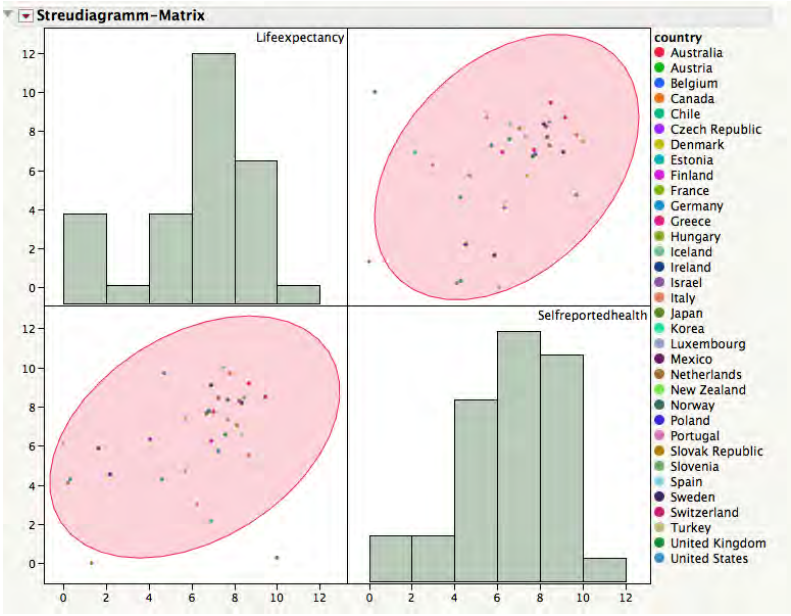
Der Stoff wird in Aufgabe 3.1 vertieft.

3.2.2 Konfidenzintervall für den Erwartungswert $\boldsymbol{\mu}$ ($\boldsymbol{\Sigma}$ bekannt)

Aus der Teststatistik T^2 läßt sich die Wahrscheinlichkeitsaussage

$$P\{N(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \chi^2(1 - \alpha, p)\} = 1 - \alpha \tag{3.3}$$

herleiten. Die quadratische Form definiert ein Konfidenz-Ellipsoid (Ellipse für $p = 2$) im Parameterraum mit Zentrum $\bar{\mathbf{x}}$ und Konfidenz-Niveau



Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
Life expectancy	6,153846	2,7750179	34
Self-reported health	6,446748	2,5291737	34

Korrelationen

		Life expectancy	Self-reported health
Life expectancy	Korrelation nach Pearson	1	,377*
	Signifikanz (2-seitig)		,028
	Quadratsummen und Kreuzprodukte	254,124	87,363
	Kovarianz	7,701	2,647
	N	34	34
Self-reported health	Korrelation nach Pearson	,377*	1
	Signifikanz (2-seitig)	,028	
	Quadratsummen und Kreuzprodukte	87,363	211,092
	Kovarianz	2,647	6,397
	N	34	34

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Abbildung 3.1: OECD-Daten. Oben (JMP): Streudiagramm und Histogramm der Variablen `Lifeexpectancy`, `Selfreportedhealth`, sowie 95%-Ellipsen der empirischen Verteilung $N(\bar{\mathbf{x}}, \mathbf{S})$. Unten (SPSS): Mittelwerte und Kovarianzen.

$1 - \alpha$.

Liegt der Punkt $\boldsymbol{\mu}_0$ innerhalb der Ellipse, so wird H_0 beibehalten, ansonsten abgelehnt.

Beispiel 3.3 (Mittelwerts-Test, Fortsetzung)

Betrachtet man Abb. 3.2, so würde H_0 bei einem Signifikanzniveau von $\alpha = 60\%$ (Konfidenzniveau von $1 - \alpha = 40\%$) abgelehnt (rote Ellipse), jedoch bei kleineren Niveaus beibehalten.

Die Überschreitungswahrscheinlichkeit (p-Wert) des Tests ist

p-Wert

$$p = P(T^2 > t^2 = 1.067) = 0.587. \tag{3.4}$$

Daher müßte $\alpha > p$ gewählt werden (etwa $\alpha = 60\%$), um ein signifikantes Ergebnis zu erhalten. Entsprechend gilt für das Konfidenzniveau $1 - \alpha < 1 - p = 0.413$.

Das Signifikanzniveau α muß jedoch *vor* Ausführung des Tests fixiert werden.
Übliche Praxis ist jedoch, den Test auszuführen und *nach* Betrachtung des p-Werts das Signifikanzniveau so zu wählen, daß man das gewünschte Ergebnis erhält.
Leider wird der p-Wert von SPSS als *Signifikanz* bezeichnet. Daher erscheinen die obigen Bemerkungen als wirkungslos.

Will man im Nachhinein die H_0 verwerfen, so wäre das zu wählende Signifikanzniveau ($\alpha > p = 0.587$, z.B. $\alpha = 0.6$) außerhalb der üblichen Werte (0.1, 0.05, 0.01). Mehr Spielraum zum „Erreichen“ des gewünschten Testergebnisses bleibt bei p-Werten, die kleiner als 0.1 sind. Es handelt sich jedoch bei diesem Vorgehen um eine Verfälschung der Testprozedur.

■

3.2.3 Test für den Erwartungswert $\boldsymbol{\mu}$ ($\boldsymbol{\Sigma}$ unbekannt)

In der Praxis ist $\boldsymbol{\Sigma}$ meistens unbekannt. Wird es durch $\mathbf{S}$ geschätzt, so ist die Teststatistik $T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$ unter H_0 nicht mehr $\chi^2(p)$-verteilt.

Im skalaren Fall $p = 1$ hat man den Quotient aus dem Quadrat einer normalverteilten Größe (d.h. $\chi^2(1)$-verteilt) und einer $\chi^2(N-1)$-verteilten

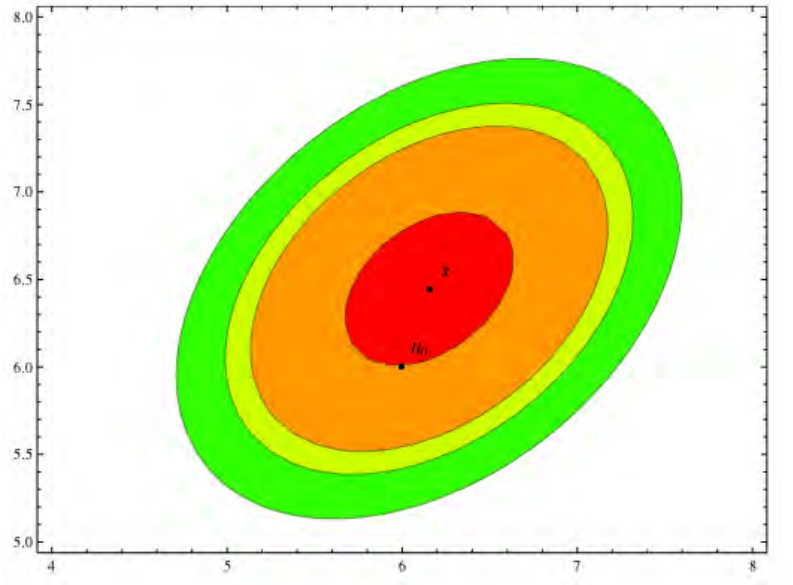


Abbildung 3.2: OECD-Daten. Bekanntes $\mathbf{\Sigma}$. Konfidenz-Ellipsen zu den Niveaus $1-\alpha = 0.4, 0.9, 0.95, 0.99$. Außerdem ist die Nullhypothese $H_0 : \boldsymbol{\mu}_0 = [6, 6]'$ eingezeichnet.

Stichprobenvarianz im Nenner. Dies führt also auf eine $F(1, N-1)$-Verteilung. Im skalaren Fall wird jedoch meistens der t-Test benutzt. In der Tat ist das Quadrat einer $t(N-1)$-verteilten Variable $F(1, N-1)$-verteilt.

Die Zusammenhänge zwischen den Testverteilungen N, χ^2, t und F sollten Ihnen bekannt sein. Siehe Kap. 10.1

Die Verteilung der Statistik

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim T^2(p, N-1) \tag{3.5}$$

Hotelling-T^2-Verteilung

wird als Hotelling-T^2-Verteilung bezeichnet. Sie hat, wie oben motiviert, einen engen Zusammenhang zur F-Verteilung. Es gilt der Zusammenhang

$$T^2(p,m) = \frac{mp}{m-p+1} F(p, m-p+1) \tag{3.6}$$

Setzt man $m = N - 1$, so läßt sich der Test mit Hilfe der $F(p, N - p)$-Verteilung durchführen. Der Testwert T^2 muß nur mit einem Faktor multipliziert werden:

$$\frac{N-p}{(N-1)p} \cdot T^2 := \tilde{T}^2 \sim F(p, N-p). \tag{3.7}$$

Beispiel 3.4 (Mittelwerts-Test, Fortsetzung)

In diesem Fall ist $N = 34$, $\bar{\mathbf{x}} = [6.154, 6.447]'$, $\mathbf{S} = \begin{bmatrix} 7.701 & 2.647 \\ 2.647 & 6.397 \end{bmatrix}$.

Die Teststatistik hat die Form

$$\tilde{T}^2 = \frac{N-p}{(N-1)p} \cdot N \cdot (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim F(p, N-p).$$

Man benötigt die Inverse der Stichproben-Kovarianz, d.h.

$\mathbf{S}^{-1} = \begin{bmatrix} 0.151 & -0.063 \\ -0.063 & 0.182 \end{bmatrix}$.

Dies ergibt sich aus der Formel (2.10). Die Determinante ist $\det(\mathbf{S}) = 42.257$.

Insgesamt hat man also (die Realisation von T^2 wird klein geschrieben)

$$\begin{aligned} \tilde{t}^2 &= \frac{34-2}{(34-1)2} \cdot t^2 \\ &= \frac{32}{66} \cdot 34 [0.154, 0.447] \begin{bmatrix} 0.151 & -0.063 \\ -0.063 & 0.182 \end{bmatrix} \begin{bmatrix} 0.154 \\ 0.447 \end{bmatrix} = 0.517. \end{aligned}$$

Der kritische Wert (95%-Quantil) ist aber $F(0.95, 2, 32) = 3.295$.

Damit muß $H_0 : \boldsymbol{\mu}_0 = [6, 6]'$ auf dem 5%-Niveau beibehalten werden.

Der p-Wert $p = 0.601$ ist nun etwas größer und $1 - p = 0.399$.

Ein Signifikanzniveau von $\alpha = 60\%$ reicht nun nicht mehr aus, um H_0 abzulehnen (vgl. Abb. 3.3). Der Unterschied zwischen den Quantilen der

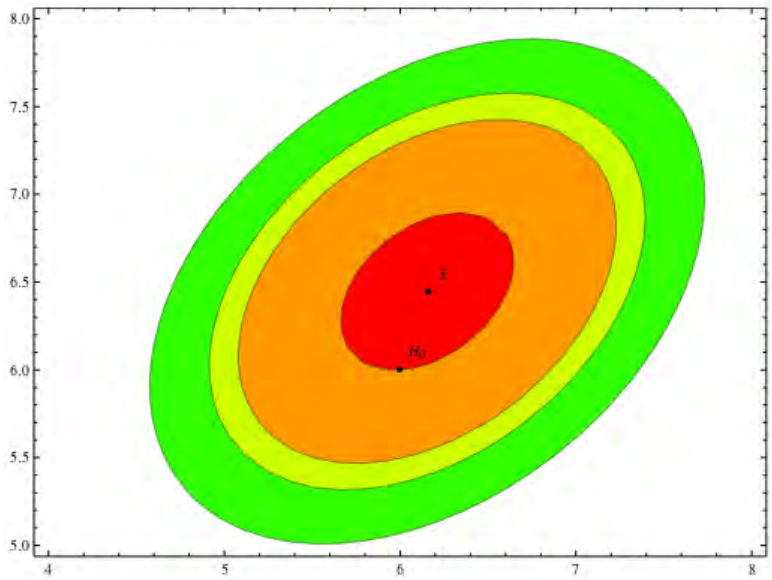


Abbildung 3.3: OECD-Daten. Unbekanntes $\mathbf{\Sigma}$. Konfidenz-Ellipsen zu den Niveaus $1-\alpha = 0.4, 0.9, 0.95, 0.99$. Außerdem ist die Nullhypothese H_0 : $\boldsymbol{\mu}_0 = [6,6]'$ eingezeichnet.

χ^2-Verteilung und der Hotelling-T^2-Verteilung ist in Abb. 3.4, unten) zu sehen. Die Quantile der Hotelling-T^2-Verteilung sind immer größer, da ja $\mathbf{\Sigma}$ nur geschätzt wurde (analog zur Normal- und t-Verteilung).

Wählt man als Nullhypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = [7, 5.5]'$, so ergibt sich

$$\tilde{t}^2 = \tfrac{32}{66} \cdot 34\,[-0.846, 0.947] \begin{bmatrix} 0.151 & -0.063 \\ -0.063 & 0.182 \end{bmatrix} \begin{bmatrix} -0.846 \\ 0.947 \end{bmatrix} = 6.135.$$

Damit muß H_0 auf dem 5%-Niveau abgelehnt werden (vgl. Abb. 3.6).

Der Stoff wird in Aufgabe 3.2 vertieft.

■

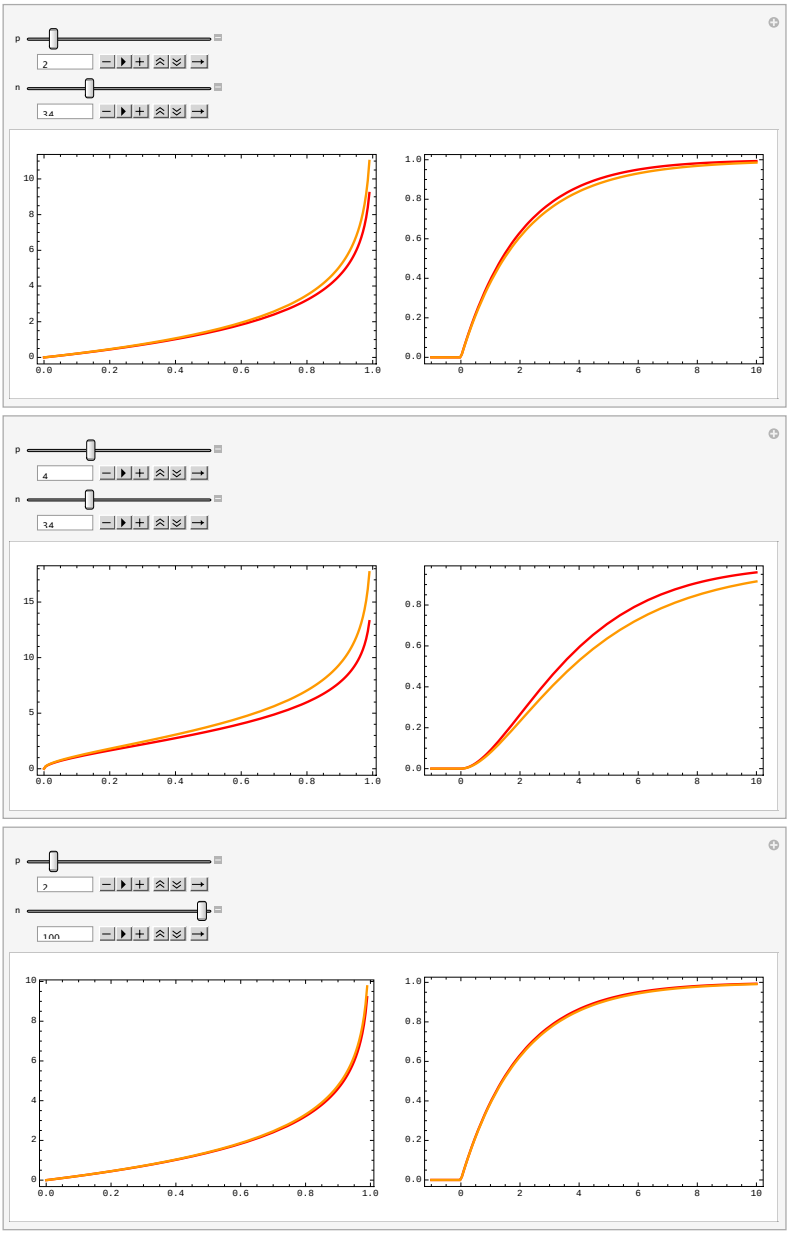


Abbildung 3.4: Quantile der χ^2-Verteilung (rot) und der Hotelling-T^2-Verteilung (orange) als Funktion von α sowie kumulierte Verteilungsfunktionen (rechts).

3.2.4 Simultane Tests und Konfidenzintervalle nach Bonferroni

Univariate Tests und Konfidenzintervalle können leicht auf multivariate Probleme angewandt werden, wenn die Einzeltests auf einem adaptierten Signifikanzniveau durchgeführt werden. Dazu wird die p-dimensionale Nullhypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ in der Form (Schnitt, Intersektion)

$$H_0 = \bigcap_i H_{0i} = \bigcap_i \{\boldsymbol{\mu} | \mu_i = \mu_{0i}\} \tag{3.8}$$

geschrieben. Die Alternative ist dann

$$H_1 = \bar{H}_0 = \bigcup_i \bar{H}_{0i} = \bigcup_i \{\boldsymbol{\mu} | \mu_i \neq \mu_{0i}\}. \tag{3.9}$$

Dies bedeutet, daß mindestens eine Komponente des Erwartungswerts vom hypothetischen Wert μ_{0i} abweicht.

Definiert man die Ereignisse (die i-te Komponente von $\boldsymbol{\mu}_0$ liegt im Konfidenzintervall)

$$E_i = \{\mu_{0i} \in [\bar{X}_i - z_i\sigma_i/\sqrt{N}, \bar{X}_i + z_i\sigma_i/\sqrt{N}]\}, \tag{3.10}$$

mit den Quantilen $z_i = z(1 - \alpha_i/2), i = 1, ..., p$, so ist das Eintreten von E_i gleichbedeutend mit dem Beibehalten der H_{0i}. Es gilt also

$$P(\bar{E}_i | H_0 \text{ richtig}) = \alpha_i \tag{3.11}$$

(Signifikanzniveau des i-ten Tests). Die Schnittmenge $E = \bigcap_i E_i$ ist mit dem Beibehalten der gesamten H_0 gleichbedeutend.

Der gesamte Test hat dann das Signifikanzniveau

$$P(\bar{E}|H_0) = P\left(\bigcup_i \bar{E}_i \middle| H_0\right) \leq \sum_i P(\bar{E}_i)|H_0) = \sum_i \alpha_i, \tag{3.12}$$

d.h. mindestens eine univariate Hypothese wird abgelehnt. Die Notation $P(\bar{E}|H_0)$ bedeutet, daß die Wahrscheinlichkeit unter Gültigkeit der H_0 berechnet wird (keine bedingte Wahrscheinlichkeit).

Will man ein bestimmtes Niveau α einhalten, so kann man $\sum_i \alpha_i = \alpha$ oder der Einfachheit halber $\alpha_i = \alpha/p$ wählen (Adjustierung des Fehlers 1. Art). Somit gilt

$$\alpha^* = P(H_0 \text{ ablehnen} | H_0 \text{ richtig}) \leq \alpha. \tag{3.13}$$

Der Test ist jedoch konservativ, da das tatsächliche α^* deutlich kleiner als α sein kann. Dieser Nachteil entfällt bei der Testprozedur mit Hilfe der T^2-Statistik.

In analoger Form erhält man simultane Konfidenzintervalle nach Bonferroni, die als Schnittmenge univariater Konfidenzintervalle entstehen. Schreibt man

simultane Konfidenzintervalle nach Bonferroni

$$P\left(\bigcap_i \{\mu_i \in [\bar{X}_i - z_i\sigma_i/\sqrt{N}, \bar{X}_i + z_i\sigma_i/\sqrt{N}]\}\right) \geq 1-\alpha \tag{3.14}$$

wobei die Umrechnung $P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$ benutzt wurde, so ist dies ein rechteckiger Bereich aus univariaten Vertrauensintervallen. Auch hier ist das tatsächliche Konfidenzniveau $1-\alpha^* \geq 1-\alpha$, d.h. der Bereich ist größer, als er sein müßte.

Alles Gesagte gilt in analoger Form mit der Ersetzung

$$z_i = z(1-\alpha_i/2) \quad \rightarrow \quad t_i = t(1-\alpha_i/2, N-1) \tag{3.15}$$

$$\sigma_i \quad \rightarrow \quad s_i \tag{3.16}$$

für den Fall mit unbekannter Kovarianzmatrix.

Man betrachtet also adjustierte Quantile $z(1-\alpha/(2p))$ oder $t(1-\alpha/(2p), N-1)$ bei $i = 1, ..., p$ Einzelvergleichen. Die entsprechenden Quantile der t-Verteilung sind in Kap. 12.4 tabelliert (Bonferroni-t-Statistik). Normalverteilungsquantile stehen in der letzten Zeile (Freiheitsgrade $\nu = \infty$ in der Tabelle).

Beispiel 3.5 (Mittelwerts-Test, Fortsetzung)

Abb. 3.5 zeigt Konfidenzellipsen und rechteckige Bonferroni z- und t-Intervalle, je nachdem, ob $\mathbf{\Sigma}$ als bekannt angenommen wird.

Die univariaten Intervalle sind ($\mathbf{\Sigma}$ bekannt, $\alpha = 0.05$):

$[\bar{x}_i - z_i\sigma_i/\sqrt{N}, \bar{x}_i + z_i\sigma_i/\sqrt{N}]$

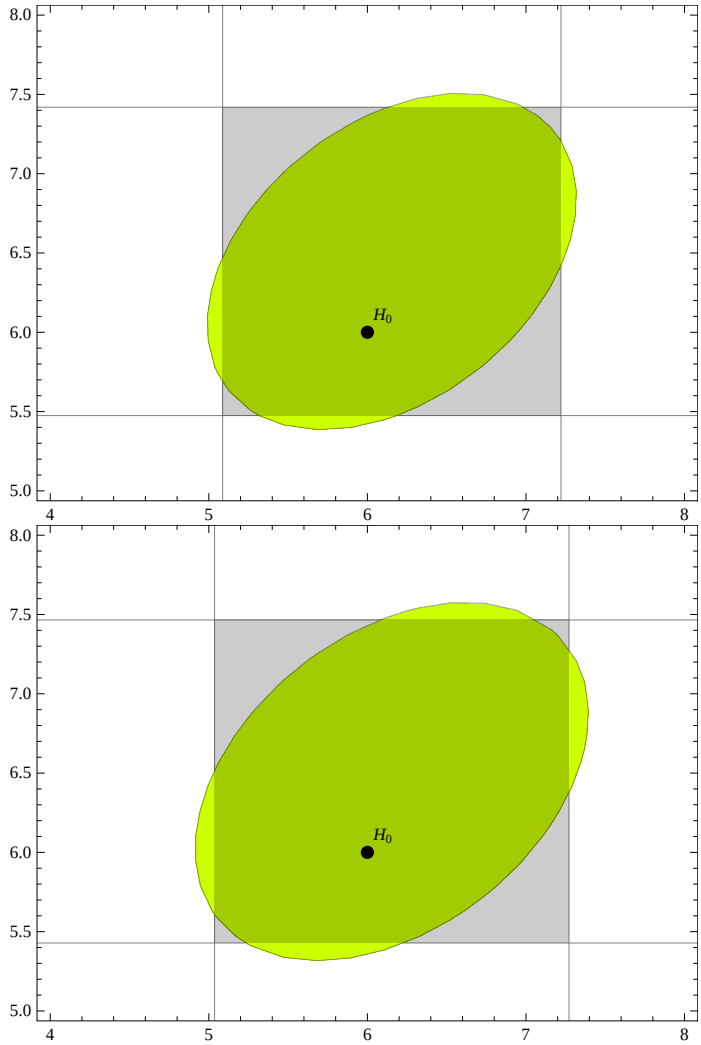


Abbildung 3.5: OECD-Daten. Konfidenz-Ellipsen zum Niveau $1 - \alpha = 0.95$ und rechteckige Bonferroni-Konfidenzintervalle. Außerdem ist die Nullhypothese $H_0 : \boldsymbol{\mu}_0 = [6, 6]'$ eingezeichnet. Oberes Bild: Bekanntes $\boldsymbol{\Sigma}$ (z-Intervalle), unten: unbekanntes $\boldsymbol{\Sigma}$ (t-Intervalle).

mit $\bar{x} = [6.154, 6.447]'$, $\sigma_1 = 2.776$, $\sigma_2 = 2.529$, $z(1 - \alpha/(2p)) = 2.24$. Adjustierte Quantile $t(1 - \alpha/(2k), \nu)$ sind in Kap. 12.4 zu finden. Für $\nu = \infty$ erhält man Normalverteilungsquantile (letzte Zeile).

Daraus findet man die Intervalle für μ_i:

[5.087, 7.221] und [5.475, 7.419].

■

Übung 3.1

Berechnen Sie die Intervalle für den Fall mit unbekannter Kovarianzmatrix.

■

Betrachtet man andere Hypothesen als bisher, so ergibt sich das etwas seltsame Ergebnis, daß

a) Hypothesen, die im exakten Test abgelehnt werden, bei der Bonferroni-Prozedur beibehalten werden (graue Flächen).

b) Hypothesen, die im exakten Test beibehalten werden, bei der Bonferroni-Prozedur abgelehnt werden (gelbe Flächen).

Während (a) plausibel ist (es handelt sich um einen konservativen Test), kann Fall b) dadurch motiviert werden, daß es sich um unterschiedliche Konstruktionen der Konfidenzintervalle handelt, die zu widersprüchlichen Resultaten führen können.

Eine ausführliche Diskussion simultaner Tests und Konfidenzintervalle ist in Miller (1981) und Mardia et al. (1979) enthalten.

3.2.5 Simultane Tests und Konfidenzintervalle nach dem Union-Intersection-Prinzip

Schreibt man die Nullhypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ in der Form

$$H_0 = \bigcap_{\mathbf{a}} H_{0\mathbf{a}} \tag{3.17}$$

mit den Komponenten $H_{0\mathbf{a}} : \mathbf{a}'\boldsymbol{\mu} = \mathbf{a}'\boldsymbol{\mu}_0$, so wird H_0 nur dann beibehalten, wenn dies auch für alle $H_{0\mathbf{a}}$ gilt. Diese Tests können mit Hilfe der z-Statistik

$$Z_{\mathbf{a}} = (\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu})/\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}/N} \sim N(0,1) \tag{3.18}$$

ausgeführt werden, wenn $\mathbf{\Sigma}$ bekannt ist. Ansonsten nimmt man

$$t_{\mathbf{a}} = (\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)/\sqrt{\mathbf{a}'\mathbf{S}\mathbf{a}/N} \sim t(N-1). \tag{3.19}$$

Maximiert man über alle denkbaren $\mathbf{a}$, so muß auch $\max_{\mathbf{a}} |Z_{\mathbf{a}}| \leq z(1-\alpha/2)$ gelten. Die Maximierung führt direkt zur Teststatistik nach dem Union-Intersection-Prinzip.

In der Tat gilt[1]

$$\begin{aligned} \max_{\mathbf{a}} Z_{\mathbf{a}}^2 &= \max_{\mathbf{a}} N(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)^2/\mathbf{a}'\mathbf{\Sigma}\mathbf{a} && (3.20)\\ &= N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = T^2 \sim \chi^2(p). && (3.21) \end{aligned}$$

Man erhält also die in Abs. 3.2.1 benutzte Teststatistik aus dem Union-Intersection-Prinzip. Der Likelihood-Quotienten-Test führt in diesem Fall auf das gleiche Ergebnis.

Der Vorteil des komponentenweisen Testens liegt in der Konstruktion von simultanen Konfidenzintervallen. Schreibt man

$$P\{Z_{\mathbf{a}}^2 \leq \chi^2(1-\alpha, p) \text{ für alle } \mathbf{a}\} = 1-\alpha \tag{3.22}$$

und löst dies nach $\boldsymbol{\mu}$ auf, so gilt simultan für alle $\mathbf{a}$

simultane Konfidenz-intervalle

$$P\{\bigcap_{\mathbf{a}} \mathbf{a}'\boldsymbol{\mu} \in [\mathbf{a}'\bar{\mathbf{x}} - b, \mathbf{a}'\bar{\mathbf{x}} + b]\} = 1-\alpha \tag{3.23}$$

wobei

$$b = b(\mathbf{a}) = \sqrt{N^{-1}\,\mathbf{a}'\mathbf{\Sigma}\mathbf{a}\;\chi^2(1-\alpha, p)} \tag{3.24}$$

die Breite des bei $\mathbf{a}'\bar{\mathbf{x}}$ zentrierten Konfidenzintervalls ist.

Bei unbekanntem $\mathbf{\Sigma}$ ergibt sich analog

$$\begin{aligned} \max_{\mathbf{a}} t_{\mathbf{a}}^2 &= \max_{\mathbf{a}} N(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)^2/\mathbf{a}'\mathbf{S}\mathbf{a} && (3.25)\\ &= N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) = T^2 \sim T^2(p, N-1). && (3.26) \end{aligned}$$

und somit (siehe Glg. 3.7)

$$b = b(\mathbf{a}) = \sqrt{\frac{(N-1)p}{(N-p)} N^{-1}\,\mathbf{a}'\mathbf{S}\mathbf{a}\;F(1-\alpha, p, N-p)} \tag{3.27}$$

[1] statt des Betrags nimmt man das Quadrat. Siehe auch Bsp. 3.7

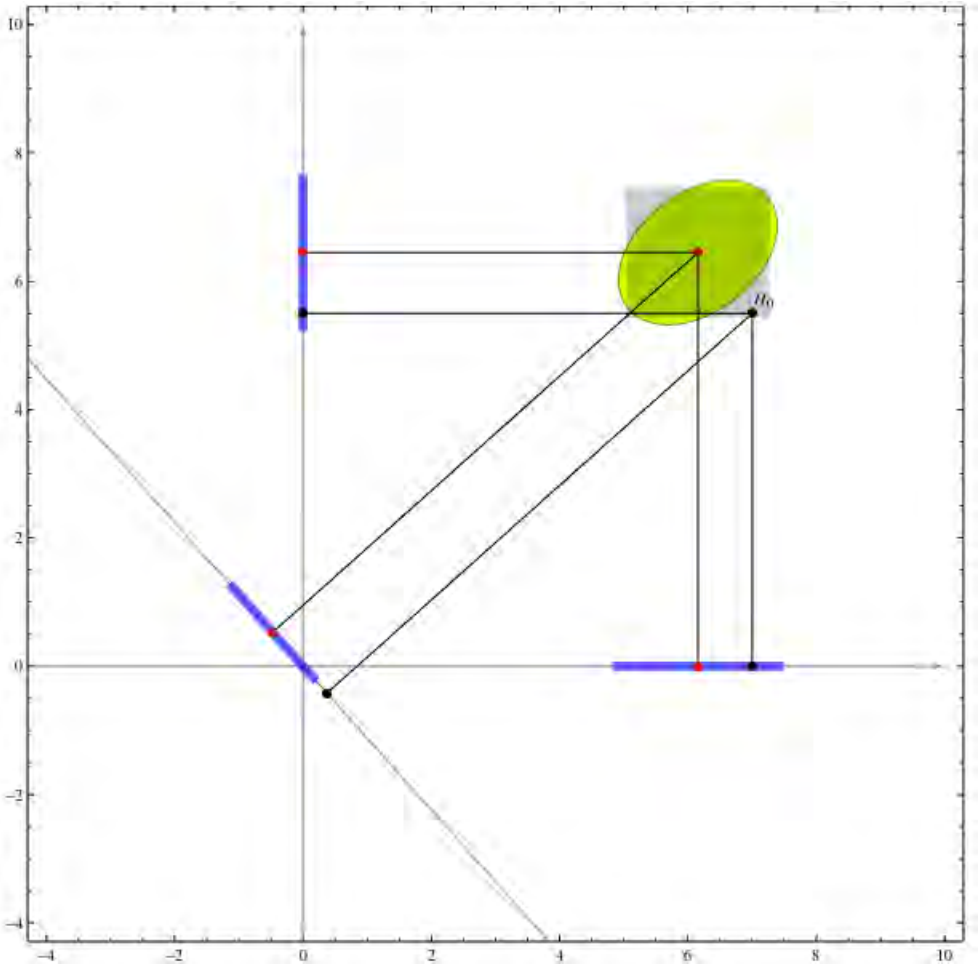


Abbildung 3.6: Simultane Konfidenz-Intervalle ($\mathbf{\Sigma}$ unbekannt). Projektionen auf die x- und y-Achse sowie auf die Verbindungslinie zwischen $\bar{\mathbf{x}}$ und $\boldsymbol{\mu}_0$. Außerdem ist die Konfidenz-Ellipse zum Niveau $1-\alpha = 0.95$, ein rechteckiges Bonferroni-Konfidenzintervall und die Nullhypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = [7, 5.5]'$ eingezeichnet. Da mindestens eine Projektion $\mathbf{a}'\boldsymbol{\mu}_0$ außerhalb des Konfidenz-Intervalls $\mathbf{a}'\bar{\mathbf{x}} \pm b(\mathbf{a})$ liegt, wird H_0 abgelehnt.

Beispiel 3.6 (Mittelwerts-Test, Fortsetzung)
Im bereits ausgiebig bekannten Beispiel war $N = 34$, $\bar{\mathbf{x}} = [6.154, 6.447]'$ und

$$\mathbf{S} = \begin{bmatrix} 7.701 & 2.647 \\ 2.647 & 6.397 \end{bmatrix}. \tag{3.28}$$

Daraus ergibt sich $\mathbf{a}'\bar{\mathbf{x}} = 6.154a_1 + 6.447a_2$ und $\mathbf{a}'\mathbf{S}\mathbf{a} = 7.701a_1^2 + 2.647a_1a_2 + 2.647a_2a_1 + 6.397a_2^2$. Weiterhin war $F(0.95, 2, 32) = 3.295$. Die Länge des 95%-Konfidenzintervalls ist somit

$$b = b(\mathbf{a}) = \sqrt{\frac{33 \cdot 2}{32 \cdot 34} \cdot \mathbf{a}'\mathbf{S}\mathbf{a} \cdot 3.295}. \tag{3.29}$$

In Abb. 3.6 wurde einige der simultanen Konfidenzintervalle ausgewählt, hier mit der Nullhypothese $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = [7, 5.5]'$. Es handelt sich um die Projektionen auf die x- und y-Achse ($\mathbf{a} = [1, 0]', [0, 1]'$) sowie auf die Verbindungslinie zwischen $\bar{\mathbf{x}}$ und $\boldsymbol{\mu}_0$. Dieser Vektor kann als Differenz von Mittelwert und Hypothese berechnet werden, d.h. $\mathbf{a} = \bar{\mathbf{x}} - \boldsymbol{\mu}_0 = [-0.846, 0.947]'$.

Da diese Projektion $\mathbf{a}'\boldsymbol{\mu}_0$ außerhalb des Konfidenz-Intervalls $\mathbf{a}'\bar{\mathbf{x}} \pm b(\mathbf{a})$ liegt, wird H_0 abgelehnt. Dies stimmt mit dem Testergebnis in Bsp. 3.4 überein.

Beliebige Projektionsrichtungen können mit Hilfe eines Applets eingestellt werden (Abb. 3.7). Der Vektor $\mathbf{a} = [\cos\phi, \sin\phi]'$ kann mit Hilfe des Winkels $0 \leq \phi \leq 2\pi$ verändert werden. ■

Beispiel 3.7 (Maximum der univariaten Statistik)
Die univariate Statistik $Z_{\mathbf{a}}^2$ ergab das Maximum

$$\max_{\mathbf{a}} Z_{\mathbf{a}}^2 = \max_{\mathbf{a}} N(\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0)^2/\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \tag{3.30}$$

$$= N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0). \tag{3.31}$$

Dies erhält man aus dem allgemeinen Resultat

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\mathbf{y})^2}{\mathbf{a}'\mathbf{B}\mathbf{a}} = \mathbf{y}'\mathbf{B}^{-1}\mathbf{y} \tag{3.32}$$

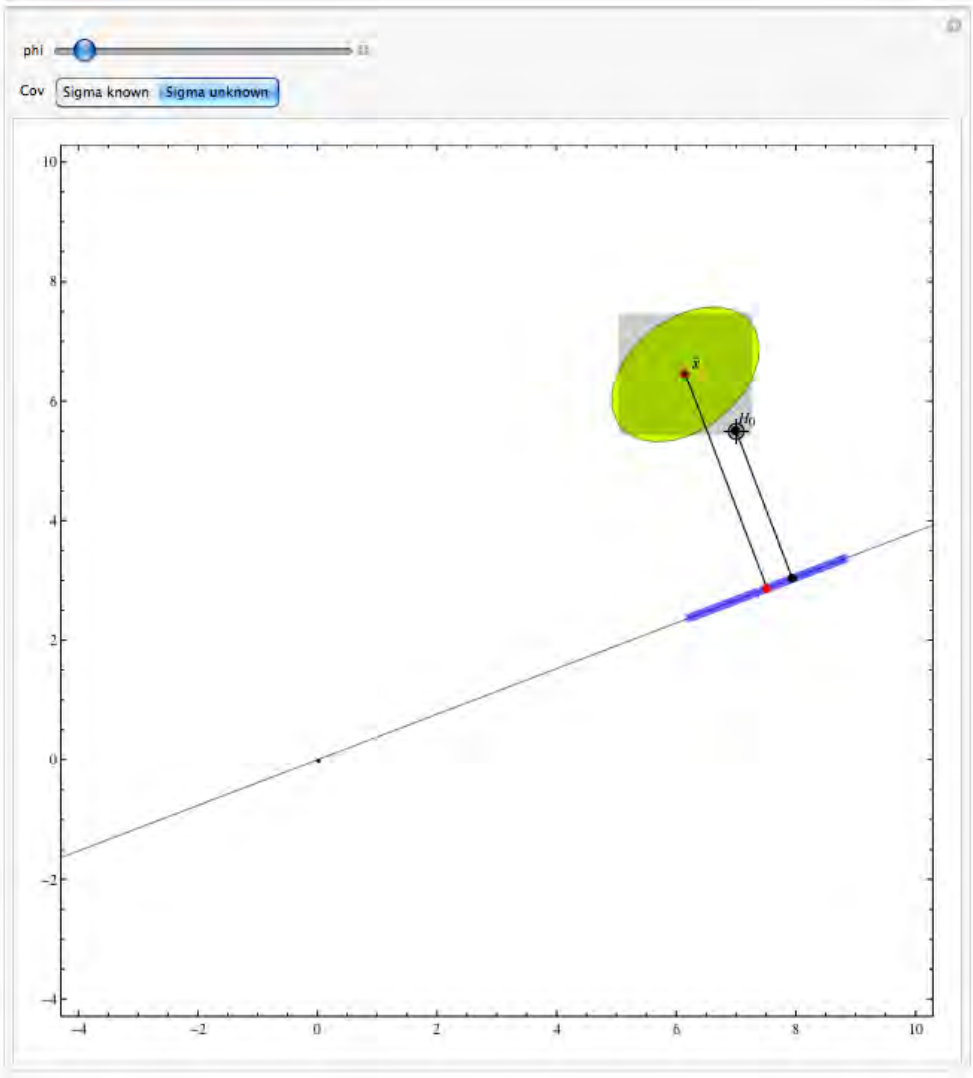


Abbildung 3.7: Applet für simultane Konfidenz-Intervalle.
http://www.fernuni-hagen.de/ls_statistik/lehre/

für beliebige Vektoren $\mathbf{y}$ und symmetrische positiv definite Matrizen $\mathbf{B}$[2].

Herleitung:
Ein ausführlicher Beweis findet sich in (Mardia et al., 1979, Kap. A.9).
Schreibt man $\mathbf{a}'\mathbf{y} = \sum_j a_j y_j$ und $\mathbf{a}'\mathbf{B}\mathbf{a} = \sum_{jk} a_j B_{jk} a_k$, so ergibt das Nullsetzen der Ableitung nach a_i (Quotientenregel)

$$\frac{\partial}{\partial a_i} \frac{(\mathbf{a}'\mathbf{y})^2}{\mathbf{a}'\mathbf{B}\mathbf{a}} = \frac{2y_i\mathbf{a}'\mathbf{y}}{\mathbf{a}'\mathbf{B}\mathbf{a}} - \frac{(\mathbf{a}'\mathbf{y})^2}{(\mathbf{a}'\mathbf{B}\mathbf{a})^2} 2 \sum_j B_{ij} a_j = 0. \tag{3.33}$$

In Vektor-Form gilt also ($\mathbf{a}'\mathbf{y}/\mathbf{a}'\mathbf{B}\mathbf{a}$ kürzen)

$$\mathbf{y} = \frac{\mathbf{a}'\mathbf{y}}{\mathbf{a}'\mathbf{B}\mathbf{a}} \mathbf{B}\mathbf{a}. \tag{3.34}$$

Diese Gleichung wird durch $\mathbf{a} = \mathbf{B}^{-1}\mathbf{y}$ gelöst. Einsetzen in $(\mathbf{a}'\mathbf{y})^2/(\mathbf{a}'\mathbf{B}\mathbf{a})$ ergibt $\mathbf{y}'\mathbf{B}^{-1}\mathbf{y}$.

■

3.2.6 Test für die Korrelationsmatrix P

Korrelationsmatrizen sind ein wichtiges Instrument, um einen Überblick der Zusammenhänge zwischen Variablen zu gewinnen (vgl. Abs. 1.5). Meistens werden die Einzelkorrelationen r_{ij} betrachtet und deren Signifikanz mit dem bekannten Korrelationstest beurteilt:
Bei bivariat normalverteilten Merkmalen

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho \\ \sigma_x\sigma_y\rho & \sigma_y^2 \end{bmatrix} \right). \tag{3.35}$$

ist unter

$$H_0 \;:\; \rho = 0 \tag{3.36}$$
$$H_1 \;:\; \rho \neq 0 \tag{3.37}$$

die Teststatistik

$$T = \sqrt{N-2}\frac{R}{\sqrt{1-R^2}} \sim t(N-2). \tag{3.38}$$

[2]d.h. alle Eigenwerte von $\mathbf{B}$ sind positiv.

p-Wert

t-verteilt mit $N-2$ Freiheitsgraden. Die von Programmpaketen berichteten Korrelationsmatrizen (vgl. Abb. 1.8) sind mit univariaten p-Werten und Signifikanz-Sternchen ausgestattet. Allerdings muss bei einem simultanen Test auf $\mathbf{P} = \mathbf{I}$, d.h. alle $\rho_{ij} = 0, i \neq j$, das Signifikanzniveau der Einzeltests adjustiert werden, um eine Inflation des Signifikanzniveaus zu vermeiden.

Alternativ kann ein Likelihood-Quotienten-Test für die Hypothese H_0 : $\mathbf{P} = \mathbf{I}$, $\boldsymbol{\mu}$ unbekannt, ausgeführt werden. Man kann zeigen (Mardia et al., 1979, S. 137), daß die Statistik

Test der Korrelations-Matrix

$$T = -N \log \det \mathbf{R} \tag{3.39}$$

asymptotisch (d.h. für große N) χ^2-verteilt ist mit $\mathrm{df} = p(p+1)/2 - p = p(p-1)/2$ Freiheitsgraden.

Sphärizitäts-Test

Da die quadratische Form $\mathbf{x}'\mathbf{I}\mathbf{x} = r^2$ eine (p-dimensionale) Kugel beschreibt, wird der Test als Sphärizitäts-Test bezeichnet. Man kann anstatt N den korrigierten Wert $N' = N - (2p+11)/6$ einsetzen, um eine verbesserte χ^2-Approximation zu erhalten.

Beispiel 3.8 (OECD-Daten)

In Abb. 1.8 wurde eine Korrelationsmatrix bei paarweisem Fallausschluß gezeigt. Dies führt auf die Schwierigkeit, daß unterschiedliche Fallzahlen für die paarweisen Korrelation auftreten. Im Rahmen einer multivariaten Vorgehensweise ist es sinnvoll, einen listenweisen Fallausschluß vorzunehmen (Abb. 3.8, $N = 24$).

Man erhält so $\det(\mathbf{R}) = 0.117$. Die Determinante wird von SPSS auch angegeben (Faktorenanalyse: Option Determinante, vgl. Abb. 1.12).

Die Freiheitsgrade ergeben sich zu $\mathrm{df} = p(p-1)/2 = 6$, die Fallzahl ist $N = 24$ und der korrigierte Wert lautet $N' = N - (2p+11)/6 = 20.833$.

Daraus findet man die Teststatistiken $t = 51.511$ und $t' = 44.714$. Die Diskrepanz zum Wert 44.719 ergibt sich aus Rundungsfehlern (Korrelationsmatrix mit 3 Kommastellen in Abb. 3.8).

Das Quantil $\chi^2(0.95, 6) = 12.592$ führt zur Ablehnung der Nullhypothese. Da es sich um einen Likelihood-Quotienten-Test handelt, können keine simultanen Konfidenzintervalle berechnet werden. Es erscheint jedoch sinnvoll, die p-Werte mit dem adjustierten Niveau $\alpha' = \alpha/6 = 0.05/6 = 0.0083$ zu vergleichen. ■

Korrelationen[c]

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth
Rooms per person	Korrelation nach Pearson	1	,504*	,793**	,526**
	Signifikanz (2-seitig)		,012	,000	,008
Dwelling without basic facilities	Korrelation nach Pearson	,504*	1	,670**	,251
	Signifikanz (2-seitig)	,012		,000	,237
Household disposable income	Korrelation nach Pearson	,793**	,670**	1	,614**
	Signifikanz (2-seitig)	,000	,000		,001
Household financial wealth	Korrelation nach Pearson	,526**	,251	,614**	1
	Signifikanz (2-seitig)	,008	,237	,001	

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

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

c. Listenweise N=24

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,680
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	44,719
	df	6
	Signifikanz nach Bartlett	,000

Abbildung 3.8: OECD-Daten. Oben: Korrelationsmatrix bei listenweisem Fallausschluß. Unten: Sphärizitäts-Test. Der Test wird von SPSS im Rahmen der Faktorenanalyse ausgeführt.

Übung 3.2

Finden Sie Korrelationen, die auch auf dem Niveau α' signifikant sind.

■

3.3 Zwei-Stichproben-Fall

3.3.1 Test für die Erwartungswerte μ_1, μ_2 $(\Sigma_1 = \Sigma_2)$

Wir betrachten 2 Gruppen, etwa gute und schlechte Schuldner (Datensatz `Kreditrisiko.LMU.sav`) und fragen uns, ob Sie sich auf den Variablen `Alter` und `Hoehe` (Höhe des Kredits) simultan unterscheiden. In Abb. 1.1 wurde die Variable `Alter` dargestellt. Ein gruppiertes bivariates Streudiagramm ist in Abb. 3.9 zu sehen. Gruppierte Mittelwerte und Korrelationen sind Abb. 3.10 zu entnehmen.

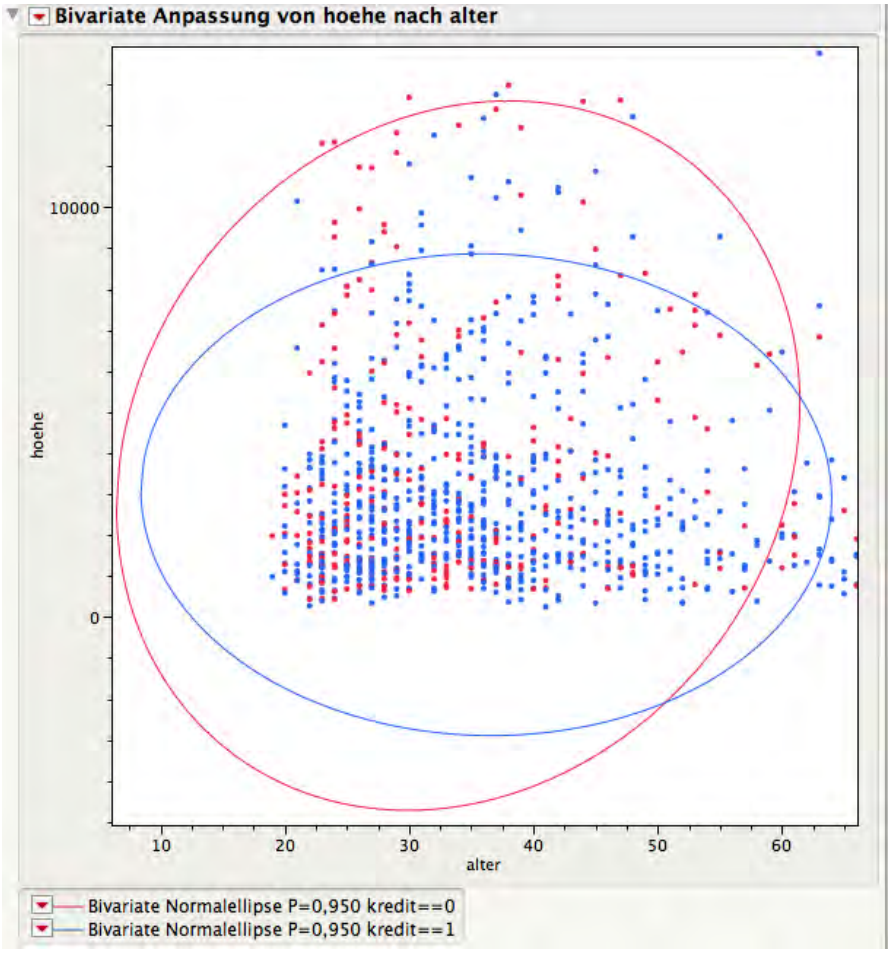


Abbildung 3.9: Kreditrisiko-Daten. Gruppiertes bivariates Streudiagramm der Variablen `Alter` und `Hoehe`

Dazu muß der Datensatz vorher gruppiert werden (Abb. 3.11).

Die Frage ist nun, ob sich die Gruppen auf den ausgewählten Variablen signifikant unterscheiden. Dies wäre ein Hinweis, daß im Umkehrschluß aus Kenntnis der Variablen `Alter` und `Hoehe` auf die Zugehörigkeit zur Gruppe `Kredit`=schlecht/gut geschlossen werden kann.

Im folgenden wird ein Test zur Überprüfung der Nullhypothese gleicher Gruppenmittelwerte bei unbekannten Kovarianzmatrizen diskutiert.

Es wird angenommen, daß die Kovarianz-Matrizen in beiden Gruppen gleich, aber unbekannt sind.

Deskriptive Statistiken

kredit		Mittelwert	Standardabweichung	N
schlecht	alter	33,960	11,2252	300
	hoehe	3938,1267	3535,81896	300
gut	alter	36,220	11,3474	700
	hoehe	2985,4429	2401,49558	700

Korrelationen

kredit			alter	hoehe
schlecht	alter	Korrelation nach Pearson	1	,148*
		Signifikanz (2-seitig)		,010
		Quadratsummen und Kreuzprodukte	37675,520	1760760,52
		Kovarianz	126,005	5888,831
		N	300	300
	hoehe	Korrelation nach Pearson	,148*	1
		Signifikanz (2-seitig)	,010	
		Quadratsummen und Kreuzprodukte	1760760,52	3,738E+9
		Kovarianz	5888,831	12502015,7
		N	300	300
gut	alter	Korrelation nach Pearson	1	-,014
		Signifikanz (2-seitig)		,703
		Quadratsummen und Kreuzprodukte	90006,120	-275448,20
		Kovarianz	128,764	-394,060
		N	700	700
	hoehe	Korrelation nach Pearson	-,014	1
		Signifikanz (2-seitig)	,703	
		Quadratsummen und Kreuzprodukte	-275448,20	4,031E+9
		Kovarianz	-394,060	5767181,01
		N	700	700

*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Abbildung 3.10: Kreditrisiko-Daten. Gruppierte Statistiken der Variablen `Alter` und `Hoehe` nach `Kredit`

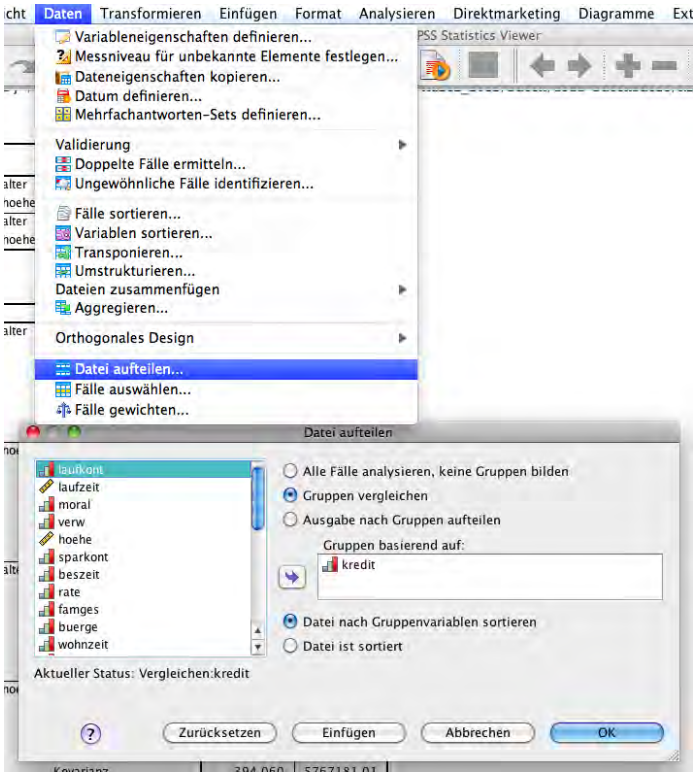


Abbildung 3.11: SPSS-Menü zur Gruppierung.

1. Daten:

   $\mathbf{x}_{1n} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), n = 1, ..., N_1$ (Gruppe 1)

   $\mathbf{x}_{2n} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), n = 1, ..., N_2$ (Gruppe 2)

   (unabhängig und identisch verteilt in den Gruppen, unabhängige Gruppen, $N = N_1 + N_2$).

2. Hypothesen: $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ gegen $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$; $\quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$.

3. Teststatistik:

$$T^2 = \frac{N_1 N_2}{N}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \sim T^2(p, N-2) \tag{3.40}$$

4. gepoolte Kovarianzmatrix $\mathbf{S} = \frac{1}{N-2}[(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2]$

5. Kovarianzmatrizen $\mathbf{S}_i = \frac{1}{N_i-1}\sum_{n=1}^{N_i}(\mathbf{x}_{in} - \bar{\mathbf{x}}_i)(\mathbf{x}_{in} - \bar{\mathbf{x}}_i)', i = 1, 2$
6. Kritischer Wert: $c = F(1 - \alpha, p, N - p - 1)$.
7. Testentscheidung: Falls $F = \frac{N-p-1}{(N-2)p}T^2 > c$, H_0 ablehnen.

Beweis-Skizze:

1. Die gepoolte Kovarianzmatrix $\mathbf{S}$ wird aus den gruppenspezifischen Kovarianzen durch Mittelung (proportional zur Gruppengröße) gewonnen. Da die beiden Mittelwerte geschätzt sind, verringert sich der Freiheitsgrad der Quadratsummen auf $N - 2$.
2. Die Differenz der Mittelwerte $\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$ ist unter H_0 normalverteilt $N(\mathbf{0}, \frac{N}{N_1N_2}\mathbf{\Sigma})$, da für die Mittelwerte $\bar{\mathbf{x}}_i \sim N(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i/N_i); i = 1, 2$ gilt. Dies erklärt den Faktor $1/N_1 + 1/N_2 = \frac{N}{N_1N_2}$.
3. Bei bekanntem $\mathbf{\Sigma}$ wäre $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'[\frac{N}{N_1N_2}\mathbf{\Sigma}]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ χ^2-verteilt mit p Freiheitsgraden. Wie im Einstichprobenfall ergibt die Ersetzung von $\mathbf{\Sigma}$ durch die Schätzung $\mathbf{S}$ eine Hotelling-$T^2(p, N - 2)$-Verteilung. Diese wird dann auf die F-Verteilung umgerechnet (Glg. 3.6 mit der Wahl $m = N - 2$). Dies ergibt den Faktor $\frac{N-2-p+1}{(N-2)p} = \frac{N-p-1}{(N-2)p}$.

Beispiel 3.9 (Kreditrisiko-Daten)
Am Anfang des Abschnitts wurde bereits der Unterschied in den Gruppen schlechte/gute Schuldner qualitativ diskutiert und mit entsprechenden Statistiken belegt. Aus Abb. 3.10 entnimmt man die Werte

$N_1 = 300, N_2 = 700, N = 1000$

$\bar{\mathbf{x}}_1 = [33.960, 3938.1267]'$

$\bar{\mathbf{x}}_2 = [36.220, 2985.4429]'$

$\mathbf{d} = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 = [-2.26, 952.684]'$

$$\mathbf{S}_1 = \begin{bmatrix} 126.005 & 5888.831 \\ 5888.831 & 12502015.7 \end{bmatrix}$$

$$\mathbf{S}_2 = \begin{bmatrix} 128.764 & -394.060 \\ -394.060 & 5767181.01 \end{bmatrix}$$

$$\mathbf{S} = ((N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2)/(N - 2) = \begin{bmatrix} 127.937 & 1488.29 \\ 1488.29 & 7784932.084 \end{bmatrix}.$$

Diese Matrix kann auch aus Abb. 3.12 abgelesen werden.

$T^2 = \frac{N_1N_2}{N}\mathbf{d}'\mathbf{S}^{-1}\mathbf{d} = 34.294$

Matrix der Quadratsummen und Kreuzprodukte für Residuen

		alter	hoehe
Quadratsumme und Kreuzprodukte	alter	127681,640	1485312,32
	hoehe	1485312,32	7,769E+9
Kovarianz	alter	127,938	1488,289
	hoehe	1488,289	7784932,08
Korrelation	alter	1,000	,047
	hoehe	,047	1,000

Basiert auf Typ III Quadratsumme

Multivariate Tests[a]

Effekt		Wert	F	Hypothese df	Fehler df	Sig.
Konstanter Term	Pillai-Spur	,901	4541,823[b]	2,000	997,000	,000
	Wilks-Lambda	,099	4541,823[b]	2,000	997,000	,000
	Hotelling-Spur	9,111	4541,823[b]	2,000	997,000	,000
	Größte charakteristische Wurzel nach Roy	9,111	4541,823[b]	2,000	997,000	,000
kredit	Pillai-Spur	,033	17,130[b]	2,000	997,000	,000
	Wilks-Lambda	,967	17,130[b]	2,000	997,000	,000
	Hotelling-Spur	,034	17,130[b]	2,000	997,000	,000
	Größte charakteristische Wurzel nach Roy	,034	17,130[b]	2,000	997,000	,000

a. Design: Konstanter Term + kredit

b. Exakte Statistik

Abbildung 3.12: Kreditrisiko-Daten. MANOVA-Output.

$F = \frac{N-p-1}{(N-2)p} T^2 = 17.13$
mit dem Quantil $F(0.95; 2, 997) = 3.005$. Damit muß H_0 abgelehnt werden.

Die Kovarianz-Matrizen können auch aus den Korrelationen und den Standardabweichungen berechnet werden (Übung!).

Betrachtet man den SPSS-Output in Abb. 3.12, so sieht man, daß sich die gepoolte Kovarianz $\mathbf{S}$ als residuale Quadratsumme SQR geteilt durch die Freiheitsgrade $N-2$ auffassen läßt ($= MQR$). Der F-Test ist in Spalte 3 der unteren Abbildung abgedruckt. Die Umrechnung der Wilks'-Λ-Statistik findet sich etwa in Mardia et al. (1979, S. 138 ff).

In der Tat kann die MANOVA-Spezifikation (Abb. 3.13)

MANOVA

$$\mathbf{x}_{in} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{in} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{in},\ n = 1, ..., N_i, i = 1, 2 \tag{3.41}$$

mit $\mathrm{Var}(\boldsymbol{\epsilon}_{in}) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ zum Test der Hypothese $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ benutzt werden. Sie testet im allgemeinen die Gleichheit in $I > 2$ Gruppen (vgl. Kap. 5). Die Angabe 'konstanter Term' im MANOVA-Output ist ein Test für die Konstante $\boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$ im Modell mit Effekt-Kodierung, d.h. $\boldsymbol{\alpha}_i = \boldsymbol{\mu}_i - \boldsymbol{\mu}$, $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 = \mathbf{0}$.

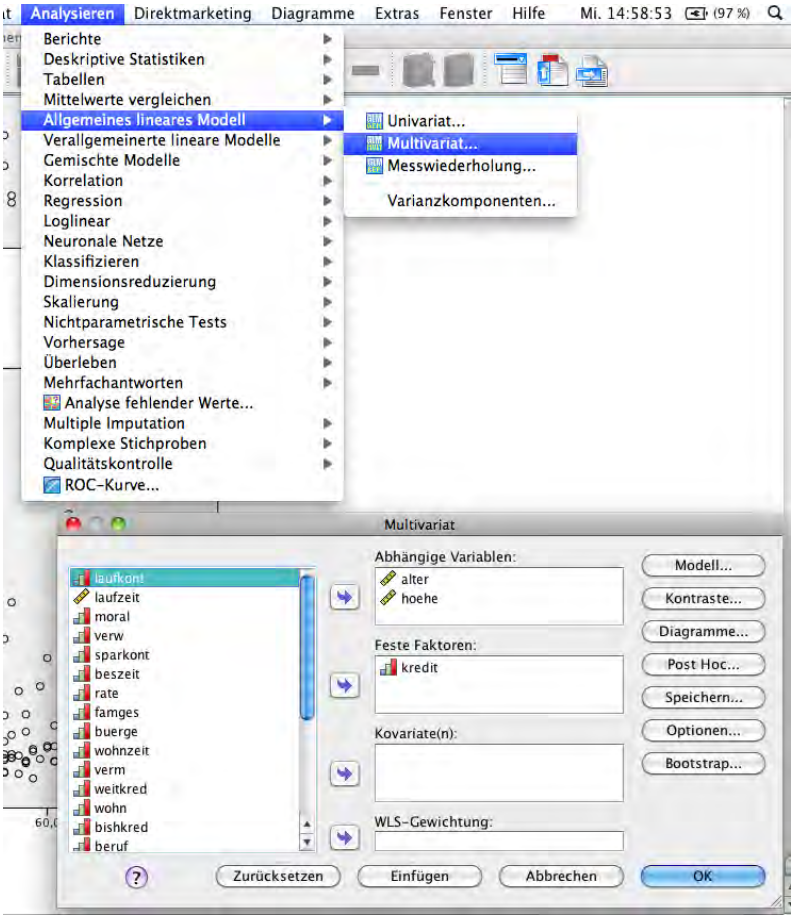


Abbildung 3.13: Kreditrisiko-Daten. MANOVA-Menü.

■

Übung 3.3 (Bonferroni-t-Test)

Prüfen Sie bitte den Mittelwertsunterschied univariat mit entsprechender Bonferroni-Adjustierung und vergleichen Sie.

Hinweis:

Starten Sie mit dem ungeteilten Datensatz. Führen Sie einen t-Test für unabhängige Stichproben durch:

SPSS/Analysieren/Mittelwerte vergleichen

■

Kapitel 4

Regressionsanalyse

Beispiel 4.1 (OECD-Daten)
Der Datensatz `BetterLifeIndex.sav` enthält eine Vielzahl von Variablen, deren Zusammenhänge bisher mit Korrelationen untersucht wurde (vgl. Abb. 1.9). Man kann auch versuchen, eine Zielvariable (abhängige Variable) aus anderen, sogenannten unabhängigen Variablen vorherzusagen. Im einfachsten Fall kommt hierfür ein lineares Modell in Frage. Als Beispiel wird im folgenden die Variable `Life Satisfaction` in Abhängigkeit der 3 Variablen `Employment rate`, `Employment rate of women with children` und `Air pollution` untersucht. Die 4-dimensionalen Daten sind in Abb. 4.1 als 2-dimensionale Projektionen dargestellt. Aufgrund des Korrelationsmusters würde man eine generell positive Wirkung der Variablen auf `Life Satisfaction` erwarten.
■

4.1 Modell und Parameterschätzung

In Erweiterung des einfachen Regressionsmodells

$$Y_n = \alpha + \beta X_n + \epsilon_n, n = 1, \ldots, N, \tag{4.1}$$

wird ein lineares Modell der Form

$$Y_n = \beta_0 + \beta_1 X_{n1} + ... + \beta_q X_{nq} + \epsilon_n, n = 1, \ldots, N, \tag{4.2}$$

multiples lineares Regressionsmodell

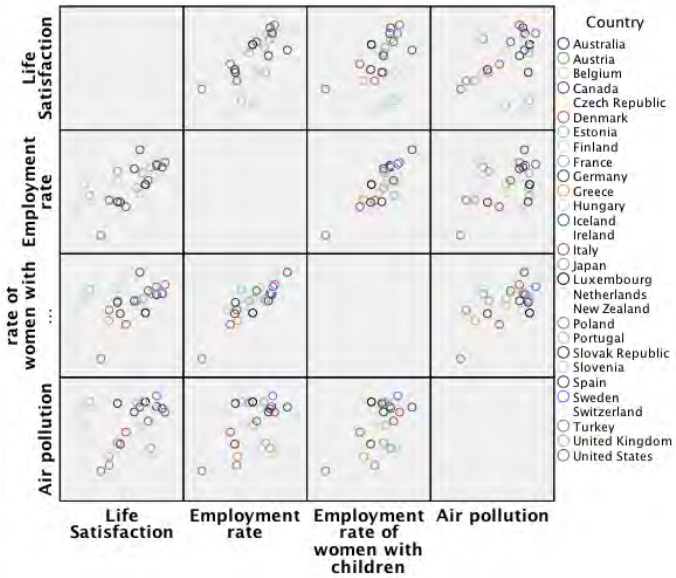


Abbildung 4.1: Streudiagramm der Variablen `Life Satisfaction`, `Employment rate`, `Employment rate of women with children` und `Air pollution`

.

Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
Life Satisfaction	6,218020	2,8067648	29
Employment rate	6,111071	2,3029350	29
Employment rate of women with children	6,752221	1,9139754	29
Air pollution	8,147611	1,5286449	29

Korrelationen

		Life Satisfaction	Employment rate	Employment rate of women with children	Air pollution
Korrelation nach Pearson	Life Satisfaction	1,000	,676	,449	,317
	Employment rate	,676	1,000	,855	,328
	Employment rate of women with children	,449	,855	1,000	,458
	Air pollution	,317	,328	,458	1,000
Sig. (Einseitig)	Life Satisfaction	.	,000	,007	,047
	Employment rate	,000	.	,000	,041
	Employment rate of women with children	,007	,000	.	,006
	Air pollution	,047	,041	,006	.
N	Life Satisfaction	29	29	29	29
	Employment rate	29	29	29	29
	Employment rate of women with children	29	29	29	29
	Air pollution	29	29	29	29

Abbildung 4.2: Statistiken der 4 Variablen.

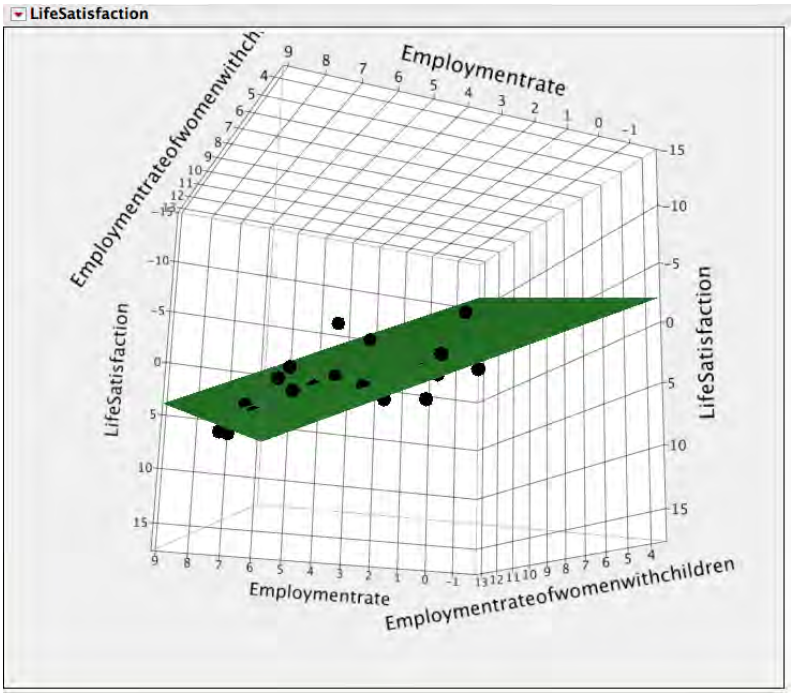


Abbildung 4.3: Daten und lineares Modell (Fläche). Die Graphik kann interaktiv gedreht werden (SAS/JMP).

angenommen. Die abhängige Variable Y_n wird also durch eine Linearkombination von p Variablen $X_{ni}, i = 1, ..., q$ erklärt. Die dabei auftretenden Fehler werden durch einen stochastischen Gleichungsfehler ϵ_n kompensiert, der weggelassene Regressoren und Spezifikationsfehler (z.B. Nichtlinearitäten) modelliert. Jeder statistischen Einheit n (Person, Firma, Land etc.) werden eigene Variable Y_n, X_{nq}, ϵ_n zugeordnet, die im einfachsten Fall für $n \neq n'$ als voneinander unabhängig angenommen werden.

Man kann sich das lineare Modell als Gerade, Fläche bzw. Hyperfläche vorstellen (Abb. 4.3).

Die Wahl eines linearen Modells mag angesichts der komplexen Zusammenhänge der Welt als trivial erscheinen, kann jedoch folgendermaßen motiviert werden:

1. Das lineare Modell ist die einfachste Näherung eines nichtlinearen Zusammenhangs $y = f(x)$.

2. Die optimale Prognose einer Variable Y aus anderen Variablen $X_1, ..., X_q$ ist linear, wenn alle Größen multivariat normalverteilt sind (siehe Abs. 2.2.2).

klassisches lineares Modell

Im klassischen linearen Modell werden folgende Annahmen getroffen:

abhängige Variablen

1. Die Regressanden Y_n (abhängige Variablen, AV) sind metrisch. Dies bedeutet, daß die Ausprägungen quantitativ interpretiert werden können. Es handelt sich also um Intervall- oder Verhältnisskalen. Qualitative Variablen führen auf sog. kategoriale Regressionsmodelle.

unabhängige Variablen

2. Die Regressoren X_{nj} (unabhängige Variablen, UV) sind metrisch oder Indikator-Variablen ($X_{nj} = 0$ oder 1). Sie werden als deterministisch angenommen (und im folgenden klein geschrieben). Dies bedeutet, daß die Werte unabhängig im Rahmem eines experimentellen Designs variiert werden können. Sind die Werte stochastisch, wie es bei Beobachtungsstudien meistens der Fall ist, sind die folgenden Resultate nur *bedingt* auf die beobachteten $X = x$-Werte zu interpretieren.

3. Die Fehler verschwinden im Mittel, d.h. $E[\epsilon_n] = 0$. Dies bedeutet, daß systematische Abweichungen von 0 schon im Modell, d.h. in β_0 berücksichtigt sind.

homoskedastische Fehler

4. Die Fehler sind homoskedastisch, d.h. $\mathrm{Var}[\epsilon_n] = \sigma^2$ (gleich für alle n).

unkorrelierte Fehler

5. $\mathrm{Cov}(\epsilon_n, \epsilon_m) = 0$ für $n \neq m$ (unkorrelierte Fehler). Die Fehler schwanken also unsystematisch für die einzelnen statistischen Einheiten.

6. Für Tests und Konfidenzintervalle wird angenommen, daß $\epsilon_n \sim N(0, \sigma^2)$ ist.

Fragestellungen

Folgende Fragestellungen sind von Interesse:

- Wie groß ist der Einfluß der erklärenden Variablen?
- Ist dieser Einfluß statistisch bedeutsam (signifikant)?
- Welche Variablen sollen aus einer großen Menge von erklärenden Größen ausgewählt werden?
- Wie kann y für neue Werte $\mathbf{x} = [x_1, ..., x_q]'$ prognostiziert werden?

Dazu muß der Gewichtsvektor $\boldsymbol{\beta}$ und die Fehlervarianz σ^2 geschätzt werden.
Zur Berechnung der Schätzer für die unbekannten Parameter $\beta_j, j = 0, ..., q$ und $\mathrm{Var}[\epsilon_n] = \sigma^2$ ist es nützlich, das multiple lineare Regressionsmodell (4.2) in Matrix-Form zu schreiben.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1q} \\ 1 & x_{21} & \dots & x_{2q} \\ \vdots & & & \vdots \\ 1 & x_{N1} & \dots & x_{Nq} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix} \tag{4.3}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.4}$$

Hierbei sind die Dimensionen der Matrizen:
$\mathbf{y} : N \times 1, \mathbf{X} = [\mathbf{1}, \mathbf{x}_{(1)}, ..., \mathbf{x}_{(q)}] : N \times (q+1), \boldsymbol{\beta} : (q+1) \times 1$ und $\boldsymbol{\epsilon} : N \times 1$.
Aufgrund der Unkorreliertheit der Fehler gilt $E[\boldsymbol{\epsilon}] = \mathbf{0}$ und $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_N$.
Bei der kleinste-Quadrate-Schätzung (KQ) wird $\boldsymbol{\beta}$ so bestimmt, daß die Abweichung der Daten vom linearen Prädiktor $\hat{\mathbf{y}} := \mathbf{X}\boldsymbol{\beta}$ minimal wird, d.h.

$$S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \sum_n \epsilon_n^2 \tag{4.5}$$

$$= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \sum_n (\mathbf{y}_n - \hat{\mathbf{y}}_n)^2 \tag{4.6}$$

$$\stackrel{!}{=} \text{Minimum} \tag{4.7}$$

Hierbei ist $\hat{\mathbf{y}}_n = \sum_j \mathbf{X}_{nj}\beta_j = \mathbf{X}_{n\cdot}\boldsymbol{\beta}$ der lineare Prädiktor des n-ten Datenpunkts ($\mathbf{X}_{n\cdot} = n$-te Zeile von $\mathbf{X}$).
Leitet man nach $\beta_j, j = 0, ..., q$ ab, so ergeben sich die sog. Normalgleichungen

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_j} = \sum_n 2(\mathbf{y}_n - \hat{\mathbf{y}}_n)\mathbf{X}_{nj} = 2(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{X}_{\cdot j} = 0 \tag{4.8}$$

Normalgleichungen

($\mathbf{X}_{\cdot j} = j$-te Spalte von $\mathbf{X}$) bzw.

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0. \tag{4.9}$$

Dies ergibt den KQ-Schätzer

KQ-Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{4.10}$$

Dies setzt voraus, daß man $\mathbf{X}'\mathbf{X} : (q+1) \times (q+1)$ invertieren kann. Dafür ist notwendig, daß man $N \geq q+1$ Datenpunkte hat. Wenn Spalten von $\mathbf{X}$ sehr ähnlich sind (kollineare Variablen), kann es numerische Probleme geben.

Mit Hilfe des KQ-Schätzers kann man prognostizierte Werte (y-Dach)

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} := \mathbf{P}\mathbf{y} \tag{4.11}$$

und Residuen (geschätzte Gleichungsfehler, Prognosefehler)

$$\begin{aligned} \hat{\boldsymbol{\epsilon}} &= \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} && (4.12) \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y} := \mathbf{Q}\mathbf{y} && (4.13) \end{aligned}$$

berechnen.

Die Matrizen $\mathbf{P} : N \times N$ und $\mathbf{Q} : N \times N$ erfüllen die Gleichungen

Projektionsmatrizen

$$\mathbf{PP} = \mathbf{P}; \quad \mathbf{QQ} = \mathbf{Q}; \quad \mathbf{PQ} = \mathbf{O}; \quad \mathbf{P} + \mathbf{Q} = \mathbf{I} \tag{4.14}$$

Dies bedeutet, daß sich ein Vektor $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}$ nicht mehr verändert, d.h. $\mathbf{P}\hat{\mathbf{y}} = \mathbf{PP}\mathbf{y} = \hat{\mathbf{y}}$. Man spricht auch von Projektionsmatrizen oder Projektoren.

hat-Matrix

$\mathbf{P}$ wird ebenfalls als hat-Matrix bezeichnet.[1]

[1] englisch y-hat für y-Dach

Aus $\mathbf{PQ} = \mathbf{O}$ läßt sich leicht ersehen, daß Prognose und Residuen senkrecht aufeinander stehen (vgl. Abb. 2.3). Es gilt nämlich

$$\hat{\mathbf{y}}'\hat{\boldsymbol{\epsilon}} = (\mathbf{Py})'\mathbf{Qy} = \mathbf{y}'\mathbf{PQy} = 0. \tag{4.15}$$

Zerlegt man den Beobachtungsvektor in

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\boldsymbol{\epsilon}} = \mathbf{Py} + \mathbf{Qy} \tag{4.16}$$

so findet man die Streuungszerlegung

$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \tag{4.17}$$

Der Schätzer für die Fehlervarianz ergibt sich aus den Residuen als gemittelte Quadratsumme

$$\widehat{\sigma^2} = \frac{1}{N-q-1}\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = \frac{1}{N-q-1}SQR \tag{4.18}$$

Die Eigenschaften der KQ-Schätzer lassen sich im sog. Gauß-Markoff-Theorem zusammenfassen:

Gauß-Markoff-Theorem

1. $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$
2. $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
3. $\hat{\boldsymbol{\beta}}$ ist BLUE (best linear unbiased estimator).

   D.h. der KQ-Schätzer hat die kleinste Varianz unter allen linearen, erwartungstreuen Schätzern.
4. $E[\widehat{\sigma^2}] = \sigma^2$.

1. Aus dem Modell $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ ergibt sich $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ und somit $E[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{y}] = \boldsymbol{\beta}$.
2. Weiterhin gilt $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Var}[\mathbf{y}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Setzt man $\mathrm{Var}[\mathbf{y}] = \sigma^2\mathbf{I}$ ein, so ergibt sich die Behauptung.

3. Siehe Fahrmeir et al. (1996, S. 99).

4. Die residuale Quadratsumme $SQR = \mathbf{y}'\mathbf{Q}\mathbf{y} = \mathrm{tr}[\mathbf{Q}\mathbf{y}\mathbf{y}']$ hat den Erwartungswert $E[SQR] = \mathrm{tr}[\mathbf{Q}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}')]$, wobei $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{y}) = \sigma^2\mathbf{I}_N$ und $\boldsymbol{\mu} = E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$. Nun gilt $\mathbf{Q}\mathbf{X} = (\mathbf{I} - \mathbf{P})\mathbf{X} = \mathbf{O}$, da $\mathbf{P}\mathbf{X} = \mathbf{X}$. Somit bleibt $E[SQR] = \mathrm{tr}[\mathbf{Q}\sigma^2\mathbf{I}_N] = \sigma^2\mathrm{tr}[\mathbf{Q}]$. Weiterhin ist $\mathrm{tr}[\mathbf{Q}] = \mathrm{tr}[\mathbf{I}_N - \mathbf{P}] = N - (q+1)$, $\mathrm{tr}[\mathbf{P}] = \mathrm{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \mathrm{tr}[\mathbf{I}_{q+1}] = q + 1$. Daher ist $SQR/(N - (q+1))$ erwartungstreu.

Der Stoff wird in Aufgabe 4.1 vertieft.

4.2 Zentriertes Modell und Varianz-Zerlegung

Schreibt man das Modell in der zur Einfachregression (4.1) analogen Form[2]

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} && (4.19) \\ &= [\mathbf{1}, \mathbf{X}] \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix} + \boldsymbol{\epsilon} && (4.20) \end{aligned}$$

$\boldsymbol{\beta} := [\beta_1, ..., \beta_q]' : q \times 1, \mathbf{X} = [\mathbf{x}_{(1)}, ..., \mathbf{x}_{(q)}] : N \times q$, so kann die Regression auch in der zentrierten (mittelwertskorrigierten) Form

$$\begin{aligned} \mathbf{y} &= \mathbf{1}(\alpha + \bar{\mathbf{x}}'\boldsymbol{\beta}) + \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\epsilon} && (4.21) \\ &= [\mathbf{1}, \mathbf{X}_*] \begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} + \boldsymbol{\epsilon}, && (4.22) \end{aligned}$$

$\mu := \alpha + \bar{\mathbf{x}}'\boldsymbol{\beta}$, spezifiziert werden. Hierbei sind die auf die Mittelwerte zentrierten Größen durch

$$\begin{aligned} \mathbf{X}_* &= \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}' && (4.23) \\ \mathbf{y}_* &= \mathbf{y} - \mathbf{1}\bar{y} = \mathbf{y} - \bar{\mathbf{y}} && (4.24) \end{aligned}$$

[2]Die Matrix $\mathbf{X}$ enthält im Gegensatz zum vorigen Abschnitt keine $\mathbf{1}$ als erste Spalte, wird aber der Einfachheit halber gleich notiert.

mit den Mittelwerten

$$\bar{\mathbf{x}}' = N^{-1}\mathbf{1}'\mathbf{X} : 1 \times q \tag{4.25}$$
$$\bar{y} = N^{-1}\mathbf{1}'\mathbf{y} : 1 \times 1 \tag{4.26}$$

definiert.[3] Da

$$[\mathbf{1}, \mathbf{X}_*]'[\mathbf{1}, \mathbf{X}_*] = \begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{X}_* \\ \mathbf{X}_*'\mathbf{1} & \mathbf{X}_*'\mathbf{X}_* \end{bmatrix} \tag{4.27}$$
$$= \begin{bmatrix} N & \mathbf{0}' \\ \mathbf{0} & \mathbf{X}_*'\mathbf{X}_* \end{bmatrix}, \tag{4.28}$$

mit dem Nullvektor $\mathbf{0} : q \times 1$, ergeben sich die KQ-Schätzer in der aus der Einfachregression bekannten Form [4]

$$\hat{\mu} = \bar{y} \tag{4.29}$$
$$\hat{\alpha} = \bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}} \tag{4.30}$$
$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_*. \tag{4.31}$$

Hierbei wurde $\mathbf{X}_*'\mathbf{y} = \mathbf{X}_*'\mathbf{y}_* + \mathbf{X}_*'\mathbf{1}\bar{y}$ und $\mathbf{X}_*'\mathbf{1} = \mathbf{0}$ (zentrierte Matrix) eingesetzt.
Mit Hilfe der Stichproben-Kovarianzmatrizen $\mathbf{S}_{xx} = (N-1)^{-1}\mathbf{X}_*'\mathbf{X}_*$ und $\mathbf{S}_{xy} = (N-1)^{-1}\mathbf{X}_*'\mathbf{y}_*$ gilt

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}. \tag{4.32}$$

Daher ist die Prognose $\hat{\mathbf{y}}$ durch

$$\hat{\mathbf{y}} = \mathbf{1}\hat{\alpha} + \mathbf{X}\hat{\boldsymbol{\beta}} \tag{4.33}$$
$$= \mathbf{1}(\bar{y} - \bar{\mathbf{x}}'\hat{\boldsymbol{\beta}}) + \mathbf{X}\hat{\boldsymbol{\beta}} \tag{4.34}$$
$$= \mathbf{1}\bar{y} + \mathbf{X}_*\hat{\boldsymbol{\beta}} \tag{4.35}$$
$$= \mathbf{1}\bar{y} + \mathbf{P}_*\mathbf{y}_* \tag{4.36}$$

[3] $\bar{y}$ ist im multiplen Regressionsmodell eine skalare Größe, jedoch im Fall von $p > 1$ abhängigen Variablen (multivariate Regression) ein $1 \times p$-Zeilenvektor $\bar{\mathbf{y}}' = N^{-1}\mathbf{1}'\mathbf{Y}$.

[4] $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, $\hat{\beta} = [\sum(x_n - \bar{x})^2]^{-1}\sum(x_n - \bar{x})(y_n - \bar{y})$

gegeben. Dabei wurde $\mathbf{X}_*\hat{\boldsymbol{\beta}} = \mathbf{X}_*(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*'\mathbf{y}_* = \mathbf{P}_*\mathbf{y}_*$ abgekürzt. Die (symmetrische) Projektionsmatrix

$$\mathbf{P}_* = \mathbf{X}_*(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_*' \tag{4.37}$$

enthält im Gegensatz zu $\mathbf{P}$ (siehe Glg. 4.11) nur die zentrierten Regressoren. Damit läßt sich der Prognosefehler als

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{1}\bar{y} - \mathbf{P}_*\mathbf{y}_* = (\mathbf{I} - \mathbf{P}_*)\mathbf{y}_* := \mathbf{Q}_*\mathbf{y}_* \tag{4.38}$$

schreiben. Es gilt wieder (vgl. 4.14)

$$\mathbf{P}_*\mathbf{P}_* = \mathbf{P}_* \quad \mathbf{Q}_*\mathbf{Q}_* = \mathbf{Q}_* \quad \mathbf{P}_*\mathbf{Q}_* = \mathbf{O} \quad \mathbf{P}_* + \mathbf{Q}_* = \mathbf{I} \tag{4.39}$$

Damit ergibt sich die

Varianzzerlegung

$$(\mathbf{y} - \mathbf{1}\bar{y})'(\mathbf{y} - \mathbf{1}\bar{y}) = (\hat{\mathbf{y}} - \mathbf{1}\bar{y})'(\hat{\mathbf{y}} - \mathbf{1}\bar{y}) + \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \tag{4.40}$$

$$SQT = SQE + SQR \tag{4.41}$$

$$\text{Totale Streuung} = \text{Erklärte Streuung} + \text{Residual-Streuung}$$

Herleitung 1:
Man schreibt $(\mathbf{y} - \mathbf{1}\bar{y})'(\mathbf{y} - \mathbf{1}\bar{y}) = \mathbf{y}_*'\mathbf{y}_* = \mathbf{y}_*'(\mathbf{P}_* + \mathbf{Q}_*)\mathbf{y}_*$ und setzt $\hat{\mathbf{y}} = \mathbf{1}\bar{y} + \mathbf{P}_*\mathbf{y}_*$ sowie $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{Q}_*\mathbf{y}_*$ ein (siehe 4.38-4.39).

Herleitung 2:
Die Streuungszerlegung ergab sich bereits in unzentrierter Form als $\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = \mathbf{y}'(\mathbf{P} + \mathbf{Q})\mathbf{y}$. Die zentrierten Größen $\mathbf{y}_* = \mathbf{y} - \mathbf{1}\bar{y}$ lassen sich mit Hilfe der Zentrierungsmatrix (siehe Glg. 2.128 und Kap. 9)

$$\mathbf{H} = \mathbf{I} - \mathbf{M} \tag{4.42}$$

und der Mittelwertsmatrix

$$\mathbf{M} = (1/N)\mathbf{1}\mathbf{1}', \tag{4.43}$$

kurz als

$$\mathbf{y}_* = \mathbf{H}\mathbf{y} = \mathbf{y} - \mathbf{M}\mathbf{y} = \mathbf{y} - \mathbf{1}\bar{y} \tag{4.44}$$

schreiben. Es gilt $\mathbf{HH} = \mathbf{H}, \mathbf{MM} = \mathbf{M}, \mathbf{HM} = \mathbf{O}$. Aus der Prognoseformel (4.36) $\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} = \mathbf{M}\mathbf{y} + \mathbf{P}_*\mathbf{y}_* = (\mathbf{M} + \mathbf{P}_*)\mathbf{y}$ ergibt sich die Beziehung zwischen den Projektoren

$$\mathbf{P} = \mathbf{M} + \mathbf{P}_* \tag{4.45}$$

da $\mathbf{P}_*\mathbf{y}_* = \mathbf{P}_*\mathbf{y}$ und $\mathbf{y}$ beliebig ist (Abb. 4.4). Durch Multiplikation mit $\mathbf{M}$ ergibt sich die Beziehung $\mathbf{MP} = \mathbf{MM} + \mathbf{MP}_* = \mathbf{M}$, da $\mathbf{MP}_* = \mathbf{O}$ und somit

$$\mathbf{MP} = \mathbf{M} = \mathbf{PM}. \tag{4.46}$$

Dies bedeutet, daß die Prognose des Mittelwerts gleich dem Mittelwert ist. Für die rechte Seite von (4.46) wurde die Symmetrie von $\mathbf{M}$ bzw. $\mathbf{P}$ benutzt. Außerdem ergibt sich $\mathbf{M} = (\mathbf{P} + \mathbf{Q})\mathbf{M} = \mathbf{M} + \mathbf{QM}$ und somit $\mathbf{QM} = \mathbf{O}$.
Damit gilt für die Streuungszerlegung $SQT = \mathbf{y}_*'\mathbf{y}_* = \mathbf{y}'\mathbf{H}\mathbf{y} = \mathbf{y}'\mathbf{H}(\mathbf{P} + \mathbf{Q})\mathbf{H}\mathbf{y}$. Setzt man noch $\mathbf{PH} = \mathbf{P} - \mathbf{PM} = \mathbf{P} - \mathbf{M}$ und $\mathbf{QH} = \mathbf{Q} - \mathbf{QM} = \mathbf{Q}$ ein, so findet man wieder $SQT = SQE + SQR$.
■
Im Fall der Einfachregression galt $\hat{y}(x) = \hat{\alpha} + \hat{\beta}x = \bar{y} + \hat{\beta}(x - \bar{x})$. Setzt man $x = \bar{x}$ ein, so ist $\hat{y}(\bar{x}) = \bar{y}$. Analog gilt hier

$$\hat{\mathbf{y}}(\mathbf{X}) = \mathbf{1}\bar{y} + \mathbf{X}_*\hat{\boldsymbol{\beta}} = \mathbf{1}\bar{y} + (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')\hat{\boldsymbol{\beta}} \tag{4.47}$$

$$\hat{\mathbf{y}}(\mathbf{1}\bar{\mathbf{x}}') = \mathbf{1}\bar{y}. \tag{4.48}$$

Man kann auch die Formeln $\mathbf{H1}\bar{\mathbf{x}}'$ und $\mathbf{H1} = \mathbf{0}$ verwenden.

Übung 4.1 (Zentrierung von 1)
Zeigen Sie, daß $\mathbf{H1} = \mathbf{0}$ und $\mathbf{M1} = \mathbf{1}$ gilt.
■

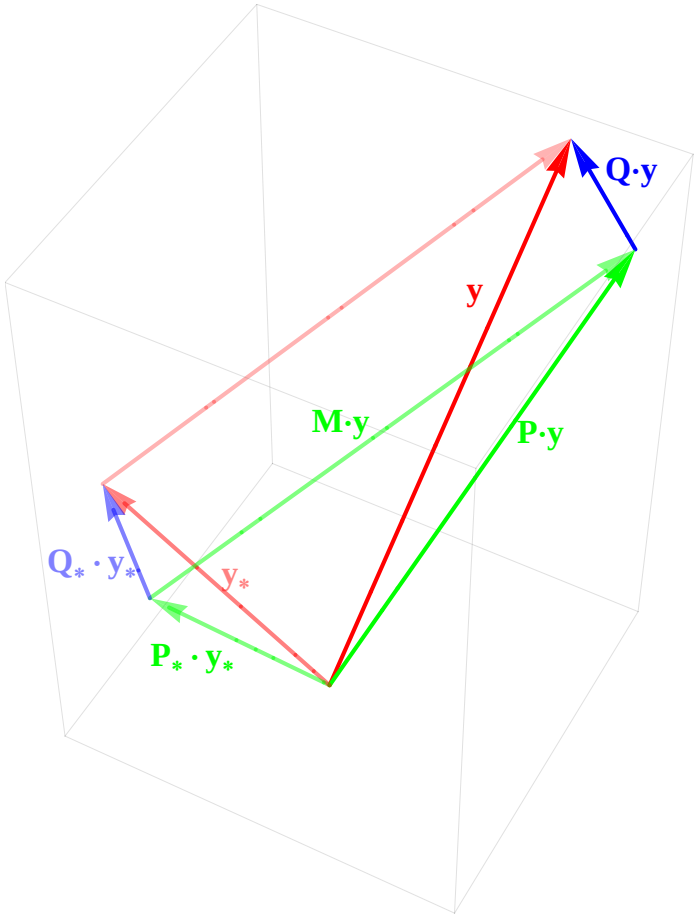


Abbildung 4.4: 3-dimensionale Darstellung der Varianzzerlegung.
http://www.fernuni-hagen.de/ls_statistik/lehre/

Beispiel 4.2 (Einfach-Regression)
Die obigen Formeln lassen sich anhand der Einfachregression

$$Y_n = \alpha + \beta x_n + \epsilon_n, n = 1, \ldots, N, \tag{4.49}$$

gut illustrieren. In Vektorform gilt

$$\mathbf{y} = \mathbf{1}\alpha + \mathbf{x}\beta + \boldsymbol{\epsilon}, \tag{4.50}$$

(d.h. $\mathbf{x} \equiv \mathbf{X}$).
Man findet daher die KQ-Schätzer (4.30-4.31)

$$\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta} \tag{4.51}$$
$$\hat{\beta} = (\mathbf{x}_*'\mathbf{x}_*)^{-1}\mathbf{x}_*'\mathbf{y}_* \tag{4.52}$$

mit $\mathbf{x}_* = \mathbf{x} - \mathbf{1}\bar{x}$. Daher ist $\mathbf{x}_*'\mathbf{x}_* = \sum_n (x_n - \bar{x})^2$ und $\mathbf{x}_*'\mathbf{y}_* = \sum_n (x_n - \bar{x})(y_n - \bar{y})$.
Die Projektionsmatrix $\mathbf{P}_*$ ist hier einfach

$$\mathbf{P}_* = \frac{\mathbf{x}_*\mathbf{x}_*'}{\sum_n (x_n - \bar{x})^2}. \tag{4.53}$$

■

4.3 Multiple Korrelation

multiple Korrelation

Der multiple Korrelationskoeffizient ist als Korrelation zwischen der Prognose $\hat{\mathbf{y}}$ und den beobachteten Daten $\mathbf{y}$ definiert.
Die mit $N-1$ multiplizierte Stichproben-Kovarianz zwischen zentrierten Daten und Prognose ist

$$(\mathbf{y} - \mathbf{1}\bar{y})'(\hat{\mathbf{y}} - \mathbf{1}\bar{y}) = \mathbf{y}_*'(\mathbf{P}_*\mathbf{y}_*) = SQE \tag{4.54}$$

und somit gleich der erklärten Streuung (siehe 4.36 und 4.40). Teilt man durch die Standardabweichungen $SQT^{1/2}$ und $SQE^{1/2}$, so ergibt sich

$$R(\mathbf{y}, \hat{\mathbf{y}}) = (SQE/SQT)^{1/2} \tag{4.55}$$

oder

$$R^2(\mathbf{y}, \hat{\mathbf{y}}) = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}. \tag{4.56}$$

Der Determinationskoeffizient (erklärte Streuung durch totale Streuung) ist also durch die quadrierte multiple Korrelation gegeben.

Übung 4.2 (Wertebereich des Determinationskoeffizienten)
Zeigen Sie,
i) daß $0 \leq R^2 \leq 1$ gilt.
ii) falls $R^2 = 1$, ist $\mathbf{y} = \hat{\mathbf{y}}$.

■

Übung 4.3 (Mittelwert der Prognose)
In (4.54) wurde benutzt, daß $\bar{\hat{\mathbf{y}}} = \mathbf{1}\bar{y}$ gilt.

Hinweis: Multiplizieren Sie $\hat{\mathbf{y}} = \mathbf{1}\bar{y} + \mathbf{P}_*\mathbf{y}_*$ von links mit $\mathbf{M}$ und verwenden Sie $\mathbf{M}\mathbf{P}_* = \mathbf{0}$ (da die Projektionsmatrix zentrierte Daten enthält).

■

Der Stoff wird in Aufgabe 4.2 vertieft.

4.4 Globaler F-Test und ANOVA-Tafel

Bei einer Regressionsanalyse ist von zentraler Bedeutung, ob die Regressoren zur Erklärung der abhängigen Variable überhaupt beitragen. Dies kann durch die Prüfung der Hypothese

$$H_0 : \beta_1 = \beta_2 = ... = \beta_q = 0 \tag{4.57}$$

entschieden werden. Unter Gültigkeit der H_0 sind alle $y_n = \beta_0 + \epsilon_n$ gleich, bis auf zufällige Schwankungen. Teilt man die Quadratsummen SQE und SQR durch die entsprechenden Freiheitsgrade, so ergibt sich ein F-Test für die Nullhypothese

$$
\begin{aligned}
F &= \frac{SQE/q}{SQR/(N-q-1)} = \frac{N-q-1}{q}\frac{SQE}{SQR} && (4.58)\\
&= \frac{N-q-1}{q}\frac{R^2}{1-R^2} && (4.59)\\
&\sim F(q, N-q-1). && (4.60)
\end{aligned}
$$

Eine große multiple Korrelation korrespondiert also zu einem großen F-Wert (Ablehnung der H_0).

Meistens wird die Testprozedur mit den 'Varianzquellen' in Form einer Varianzanalyse-Tafel (**an**alysis **of** **va**riance, ANOVA) zusammengestellt. Die gesamte Varianz wird in die Quellen

- Hypothese (lineare Prognoseregel)
- und Fehler (Residuen)

aufgeteilt ($\bar{\mathbf{y}} = \mathbf{1}\bar{y}$):

ANOVA-Tafel

Varianz-Quelle	SQ		df	F-Statistik
Hypothese	SQE	$(\hat{\mathbf{y}}-\bar{\mathbf{y}})'(\hat{\mathbf{y}}-\bar{\mathbf{y}})$	q	$F = \frac{SQE/q}{SQR/(N-q-1)}$ $\sim F(q, N-q-1)$
Residuen	SQR	$(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})$	$N-q-1$	
Total	SQT	$(\mathbf{y}-\bar{\mathbf{y}})'(\mathbf{y}-\bar{\mathbf{y}})$	$N-1$	

Der Stoff wird in Aufgabe 4.3 vertieft.

4.5 Tests und Konfidenzintervalle für einzelne Parameter

Nachdem die $H_0 : \beta_1 = \beta_2 = ... = \beta_q = 0$ abgelehnt wurde, ist von großem Interesse, welche Regressoren zur Ablehnung geführt haben.

Das Hypothesenpaar $H_0 : \beta_j = 0$, $H_1 : \beta_j \neq 0$ kann mit einem t-Test überprüft werden:

$$\frac{\hat{\beta}_j}{s_j} \sim t(N-q-1). \tag{4.61}$$

Hierbei ist $s_j := s_{\hat{\beta}_j}$ die geschätzte Standard-Abweichung von $\hat{\beta}_j$. Sie ergibt sich als Wurzel des j-ten Diagonalelements von $\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ (siehe Abs. 4.1), wenn man anstatt σ^2 den Schätzer (4.18) einsetzt.
Die t-Verteilung ergibt sich aus der Normalverteilung von $\hat{\beta}_j$ und der $\chi^2(N-q-1)$-Verteilung von $\widehat{\sigma^2}$.
Aus obigem Test kann man leicht die Gültigkeit der t-Intervalle

$$P\left\{\beta_j \in [\hat{\beta}_j - ts_j, \hat{\beta}_j + ts_j]\right\} = 1-\alpha;\ j = 0, ..., q \tag{4.62}$$

$t = t(1-\alpha/2, N-q-1); s_j = s_{\hat{\beta}_j}$ ableiten.
Wie schon in Kap. (3.2.4) ausführlich diskutiert, können simultane Konfidenzbereiche auf einfache Art als Schnittmenge

$$P\left\{\bigcap_{j\in J} \beta_j \in [\hat{\beta}_j - t_j s_j, \hat{\beta}_j + t_j s_j]\right\} \geq 1-\alpha \tag{4.63}$$

erhalten werden. Hierbei wurde $t_j = t(1-\alpha_j/2, N-q-1)$, $\alpha = \sum_{j\in J} \alpha_j$ und $J \subset \{0, ..., q\}$ gesetzt. Es ist also nicht nötig, alle Komponenten simultan zu betrachten.

Übung 4.4 (Größe der Konfidenzintervalle)
Welchen Vorteil hat es, wenn man nicht alle Parameter $\beta_j, j = 0, ..., q$ gleichzeitig in ein Konfidenzintervall einschließt?

■

4.6 Prognose von neuen Werten

Liegt ein neuer Beobachtungsvektor $\mathbf{x}_0 = [1, x_{01}, ..., x_{0q}]'$ vor, so ist eine wichtige Aufgabe, den unbekannten Wert Y_0 vorherzusagen. Legt man das bereits geschätzte lineare Modell zugrunde, so gilt für die Prognose und den Prognose-Fehler:

$$\hat{Y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \sum_{j=0}^{q} x_{0j}\hat{\beta}_j \tag{4.64}$$

$$Y_0 - \hat{Y}_0 = \hat{\epsilon}_0 = \mathbf{x}_0'\boldsymbol{\beta} + \epsilon_0 - \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \mathbf{x}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \epsilon_0 \tag{4.65}$$

Daher ist die Varianz des Prognosefehlers

$$\mathrm{Var}(\hat{\epsilon}_0) = \mathbf{x}_0'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 + \sigma^2 \tag{4.66}$$

da ϵ_0 und $\hat{\boldsymbol{\beta}}$ unkorreliert sind (warum?). Setzt man noch $\widehat{\sigma^2}$ ein, so ergibt sich das Prognose-Intervall für Y_0

Prognose-Intervall für individuelles Y_0

$$P\left\{Y_0 \in \hat{Y}_0 \pm t\,\hat{\sigma}\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}\right\} = 1 - \alpha. \tag{4.67}$$

$t = t(1 - \alpha/2, N - q - 1)$.
Anstatt einer Prognose kann man auch ein Konfidenzintervall für die Regressionsfunktion $E[Y|\mathbf{x}_0] = \mathbf{x}_0'\boldsymbol{\beta}$ ableiten. Ein Punktschätzer hiervon ist $\hat{E}[Y|\mathbf{x}_0] = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ mit Varianz $\mathrm{Var}(\hat{E}) = \mathbf{x}_0'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$. Dies ergibt das Konfidenzintervall an der Stelle $\mathbf{x}_0$

Konfidenzintervall für die Gerade

$$P\left\{\mathbf{x}_0'\boldsymbol{\beta} \in \mathbf{x}_0'\hat{\boldsymbol{\beta}} \pm t\,\hat{\sigma}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}\right\} = 1 - \alpha \tag{4.68}$$

$t = t(1 - \alpha/2, N - q - 1)$. Dieses zufällige Intervall für den deterministischen Wert $E[Y|\mathbf{x}_0] = \mathbf{x}_0'\boldsymbol{\beta}$ ist enger, da die Variabilität in ϵ_0 nicht berücksichtigt werden muß.

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	95,0% Konfidenzintervalle für B		Korrelationen		
		Regressionskoeffizient B	Standardfehler	Beta			Untergrenze	Obergrenze	Nullter Ordnung	Partiell	Teil
1	(Konstante)	,512	2,065		,248	,806	-3,742	4,765			
	Employment rate	1,393	,314	1,143	4,434	,000	,746	2,039	,676	,663	,587
	Employment rate of women with children	-,931	,402	-,635	-2,317	,029	-1,758	-,103	,449	-,420	-,307
	Air pollution	,427	,276	,233	1,547	,134	-,142	,996	,317	,296	,205

a. Abhängige Variable: Life Satisfaction

Abbildung 4.5: Geschätzte Regressionskoeffizienten.

Der Stoff wird in Aufgabe 4.4 vertieft.

Beispiel 4.3 (OECD-Daten, Fortsetzung)

Die Schätzungen für $\boldsymbol{\beta}$ lassen sich einer SPSS-Tabelle entnehmen (Abb. 4.5) Sie ergeben sich explizit aus den Daten (Tabelle 4.1)

und der Berechnung von

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 29. & 177.221 & 195.814 & 236.281 \\ 177.221 & 1231.51 & 1302.15 & 1476.28 \\ 195.814 & 1302.15 & 1424.75 & 1632.96 \\ 236.281 & 1476.28 & 1632.96 & 1990.55 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1.1029 & 0.0037 & -0.0296 & -0.1094 \\ 0.0037 & 0.0255 & -0.0274 & 0.0031 \\ -0.0296 & -0.0274 & 0.0417 & -0.0104 \\ -0.1094 & 0.0031 & -0.0104 & 0.0197 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 180.323 \\ 1224.38 \\ 1285.09 \\ 1507.25 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 0.512 \\ 1.393 \\ -0.931 \\ 0.427 \end{bmatrix}.$$

In den Datenmatrizen $\mathbf{X}, \mathbf{y}$ wurden alle Zeilen weggelassen, welche fehlende Werte enthalten (listenweiser Fallausschluß). Dadurch ergeben sich nur $N = 29$ Zeilen (der Datensatz hat insgesamt $N = 34$ Zeilen).

Die Schätzung für σ^2 ergibt sich aus der Berechnung der geschätzten Werte und der Residualwerte durch Berechnung der Quadratsumme SQR.

	$\mathbf{y}$	$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{y}_*$	$\hat{\mathbf{y}}$	$\hat{\epsilon}$
1	9.032	1.	8.053	7.433	9.263	2.814	8.765	0.267
2	8.387	1.	7.874	7.521	6.373	2.169	7.2	1.188
3	7.097	1.	4.867	6.189	7.894	0.879	4.902	2.194
4	9.677	1.	7.861	7.529	9.122	3.459	8.348	1.329
5	4.839	1.	5.792	7.435	8.437	−1.379	5.262	−0.423
6	10.	1.	8.405	8.556	8.875	3.782	8.044	1.956
7	1.29	1.	4.562	7.977	9.588	−4.928	3.535	−2.245
8	8.71	1.	6.767	8.322	9.148	2.492	6.098	2.612
9	6.774	1.	5.479	6.707	9.526	0.556	5.969	0.806
10	6.452	1.	7.681	6.7	8.886	0.234	8.768	−2.317
11	3.548	1.	4.106	4.42	5.79	−2.67	4.59	−1.041
12	0.	1.	2.82	5.571	9.004	−6.218	3.1	−3.1
13	7.097	1.	9.869	10.	9.226	0.879	8.89	−1.793
14	8.387	1.	4.232	4.976	9.604	2.169	5.876	2.511
15	5.484	1.	3.281	3.969	7.489	−0.734	4.586	0.898
16	4.516	1.	7.373	6.695	6.743	−1.702	7.428	−2.912
17	7.742	1.	5.858	5.305	9.586	1.524	7.827	−0.085
18	9.032	1.	8.786	8.089	6.034	2.814	7.797	1.236
19	8.065	1.	8.063	8.203	9.724	1.846	8.259	−0.195
20	3.548	1.	4.014	5.662	5.188	−2.67	3.048	0.5
21	0.645	1.	5.964	6.93	7.945	−5.573	5.761	−5.115
22	4.516	1.	3.862	6.509	9.487	−1.702	3.884	0.632
23	4.516	1.	6.163	8.058	6.372	−1.702	4.317	0.199
24	4.839	1.	3.796	5.213	6.661	−1.379	3.792	1.047
25	9.032	1.	8.184	8.332	10.	2.814	8.427	0.606
26	9.032	1.	10.	8.733	7.679	2.814	9.59	−0.558
27	2.581	1.	0.	0.	4.798	−3.637	2.561	0.019
28	7.419	1.	7.188	6.915	9.58	1.201	8.177	−0.758
29	8.065	1.	6.322	7.866	8.259	1.846	5.523	2.542
Mittelwerte	6.218	1.	6.111	6.752	8.148	0.	6.218	0.

Tabelle 4.1: Daten und Rechengrößen zur multiplen Korrelation (Datei `Bsp.4.3.xls`).

ANOVA[a]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Sig.
1	Regression	123,899	3	41,300	10,679	,000[b]
	Nicht standardisierte Residuen	96,683	25	3,867		
	Gesamt	220,582	28			

a. Abhängige Variable: Life Satisfaction

b. Einflußvariablen : (Konstante), Air pollution, Employment rate, Employment rate of women with children

Modellzusammenfassung[b]

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,749[a]	,562	,509	1,9665506

a. Einflußvariablen : (Konstante), Air pollution, Employment rate, Employment rate of women with children

b. Abhängige Variable: Life Satisfaction

Abbildung 4.6: ANOVA-Tafel und multiple Korrelation.

Dies ergibt $SQR = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = 96.683$ und $MQR = SQR/(N-q-1) = 96.683/(29-3-1) = 3.86732 = \widehat{\sigma^2}$. Diesen Wert findet man auch in der ANOVA-Tafel (Abb. 4.6).

Die totale und erklärte Quadratsumme ergibt sich aus $\mathbf{y}_* = \mathbf{y} - \mathbf{1}\bar{y}$ und $\hat{\mathbf{y}} - \mathbf{1}\bar{y}$ mit $\bar{y} = 6.21802$. Man findet $SQT = 220.582$, $SQE = 123.899$. Man kann natürlich auch die Formel $SQT = SQE + SQR$ ausnützen.

Daraus folgt die multiple Korrelation $R^2 = SQE/SQT = 0.562$.

Der globale F-Wert ist $F = (SQE/q)/(SQR/(n-q-1)) = 10.679$. Aufgrund des Quantils $F(1-0.05, 3, 25) = 2.99$ muß die $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ abgelehnt werden. Damit korrespondiert ein sehr kleiner p-Wert $P(F > 10.6791) = 0.000105359$.

Mit Hilfe des residualen Mittelwerts $MQR = \widehat{\sigma^2}$ kann man die Standardfehler von $\hat{\boldsymbol{\beta}}$ als Wurzel der Diagonalelemente $s_j = [\widehat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1}]_{jj}^{1/2}$ gewinnen. Es ergibt sich $\mathbf{s} = [2.06521, 0.314097, 0.401661, 0.276178]'$ und die t-Werte $\hat{\boldsymbol{\beta}}/\mathbf{s} = [0.247695, 4.43367, -2.3172, 1.54684]'$ (Abb. 4.5, Spalte 5).

Ohne Bonferroni-Adjustierung würde man t-Werte größer als 2 als Indiz für einen sigifikanten Einfluß werten ($t(0.975, 25) = 2.06$). *Mit* Adjustierung hat man $t(1-\alpha/(2(q+1)), 25) = 2.69$.

Weiterhin kann man aus $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \widehat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1} := s_{jk}$ die geschätzte Kor-

Korrelation der Schätzer

Korrelation				
Achsenabschnitt	1,0000	0,0222	-0,1381	-0,7415
Employmentrate	0,0222	1,0000	-0,8391	0,1380
Employmentrateofwomenwithchildren	-0,1381	-0,8391	1,0000	-0,3627
Airpollution	-0,7415	0,1380	-0,3627	1,0000

Abbildung 4.7: Geschätzte Korrelation $\widehat{\text{Corr}}(\hat{\boldsymbol{\beta}})$ der Schätzer.

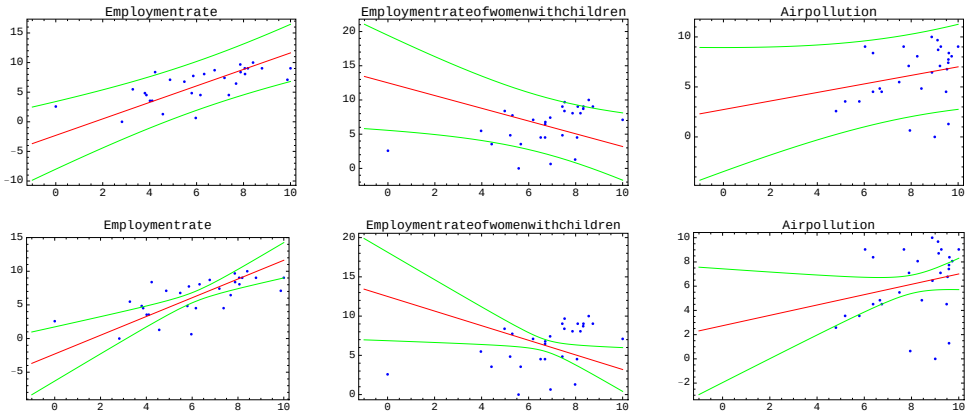


Abbildung 4.8: Oben: Individuelle Prognose-Intervalle $(1-\alpha = 95\%)$. Die Datenpunkte sind die Projektionen $[x_{nj}, y_n], j = 1, .., 3, n = 1, ..., N$. Unten: Konfidenzintervalle.

relationsmatrix der Schätzer bestimmen. Man berechnet einfach $r_{jk} = s_{jk}/(s_j s_k)$ und findet

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 4.265 & 0.014 & -0.115 & -0.423 \\ 0.014 & 0.099 & -0.106 & 0.012 \\ -0.115 & -0.106 & 0.161 & -0.04 \\ -0.423 & 0.012 & -0.04 & 0.076 \end{bmatrix}$$

und

$$\widehat{\text{Corr}}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 1. & 0.022 & -0.138 & -0.742 \\ 0.022 & 1. & -0.839 & 0.138 \\ -0.138 & -0.839 & 1. & -0.363 \\ -0.742 & 0.138 & -0.363 & 1. \end{bmatrix}.$$

Schließlich werden noch die individuellen Prognoseintervalle und Konfidenzintervalle für die Regressionsfläche (vgl. Glg. 4.11 und Abb. 4.3) berechnet. Da $\mathbf{x}_0 = [1, x_{01}, x_{02}, x_{03}]'$ ein 4-dimensionaler Vektor ist, wird zur graphischen Darstellung nur eine x-Koordinate benutzt und die anderen durch die Mittelwerte $[1, 6.11107, 6.75222, 8.14761] = N^{-1}\mathbf{1}'\mathbf{X}$ der $\mathbf{X}$-Matrix ersetzt.

Korrelationen

	LifeSatisfaction	Employmentrate	Employmentrateofwomenwithchildren	Airpollution
LifeSatisfaction	1,0000	0,6764	0,4488	0,3168
Employmentrate	0,6764	1,0000	0,8549	0,3282
Employmentrateofwomenwithchildren	0,4488	0,8549	1,0000	0,4583
Airpollution	0,3168	0,3282	0,4583	1,0000

Die Korrelationen werden anhand der Methode „Zeilenweise" geschätzt.

Partielle Korrelationen

	LifeSatisfaction	Employmentrate	Employmentrateofwomenwithchildren	Airpollution
LifeSatisfaction	.	0,6635	-0,4205	0,2955
Employmentrate	0,6635	.	0,8486	-0,2947
Employmentrateofwomenwithchildren	-0,4205	0,8486	.	0,4386
Airpollution	0,2955	-0,2947	0,4386	.

Geteilt in Bezug auf alle anderen Variablen

Abbildung 4.9: Korrelationen und partielle Korrelationen (SAS/JMP).

partielle Korrelationen

Vergleicht man die Korrelationstabelle 4.2 mit den Regressionskoeffizienten, zeigt sich ein negativer Einfluß der Variable `Employment rate of women with children` in Kontext der anderen Variablen. In der Korrelationstabelle wird der Einfluß isoliert betrachtet. Es ist daher interessant, auch die partiellen Korrelationen zu betrachten. Dies ist die Korrelation zwischen zwei Größen, wenn der Einfluß der anderen Variablen schon berücksichtigt wurde. In der Koeffiziententabelle (Abb. 4.5) sind in der vorletzten Spalte die partiellen Korrelationen tabelliert. Berücksichtigt man also die Wirkung von `Employment rate` (und `Air pollution`), so ist der Einfluß der Berufstätigkeit von Frauen mit Kindern negativ auf die Lebenszufriedenheit. Dieser Effekt zeigt sich auch ohne die Variable `Air pollution`. Die partiellen Korrelationen sind in Abb. 4.9 nochmals übersichtlich dargestellt.

Diagnose

Eine wichtige Thematik ist auch die Modellüberprüfung (Diagnose). Zwar war der F-Test signifikant, jedoch wurden bei der Modellspezifikation einige Annahmen gemacht, die verletzt sein könnten. Dies ist vor allem die Linearität des Modells sowie die Homoskedastizität bzw. Normalverteilungsanahme der Fehlerterme, die bei den Teststatistiken benutzt wurde.

Unter den getroffenen Annahmen muß gelten ($\hat{\boldsymbol{\epsilon}} = \mathbf{Qy}, \mathbf{QX} = \mathbf{O}$)

$$E[\hat{\boldsymbol{\epsilon}}] = \mathbf{Q}E[\mathbf{y}] = \mathbf{QX}\boldsymbol{\beta} = \mathbf{0} \tag{4.69}$$

$$\mathrm{Var}(\hat{\boldsymbol{\epsilon}}) = \mathbf{Q}\mathrm{Var}(\mathbf{y})\mathbf{Q} = \mathbf{Q}\sigma^2\mathbf{IQ} = \sigma^2\mathbf{Q} \tag{4.70}$$

$$\mathrm{Cov}(\hat{\mathbf{y}}, \hat{\boldsymbol{\epsilon}}) = \mathbf{P}\mathrm{Var}(\mathbf{y})\mathbf{Q} = \mathbf{P}\sigma^2\mathbf{IQ} = \sigma^2\mathbf{PQ} = \mathbf{O} \tag{4.71}$$

Daher sollten die Residuen um Null schwanken und nicht mit den prognostizierten Werten korrelieren. Das n-te Residuum weist die Varianz

$\text{Var}(\hat{\epsilon}_n) = \sigma^2 Q_{nn} = \sigma^2(1 - P_{nn})$ auf (siehe unten). Es ist außerdem normalverteilt $N(0, \sigma^2 Q_{nn})$.

Eine Inspektion von Abb. 4.10 zeigt, daß Residuen im mittleren Bereich der geschätzten y-Werte (also bei 0 in standardisierten Variablen) eher im positiven Bereich liegen. Die Beobachtung `country` = Portugal liegt weit im negativen Bereich (Ausreißer?). Eine Korrelation der Residuen mit den prognostizierten Werten ist jedoch nicht erkennbar.

Die Elemente P_{nm} der Hat-Matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ bestimmen den Einfluß (Hebelwirkung, leverage) des Datenpunkts y_m auf die Prognose $\hat{y}_n$. Es gilt $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}$ (siehe Glg. 4.11). In Komponenten hat man $\hat{y}_n = P_{nn}y_n + \sum_{m \neq n} P_{nm}y_m$. Insbesondere drückt das Diagnalelement P_{nn} den Einfluß den n-ten Datenpunkts auf seine eigene Prognose aus.

high leverage points

Beobachtungen mit hohen Werten P_{nn} werden als high leverage points bezeichnet. Welche Werte von P_{nn} sind aber hoch?

Da es sich bei $\mathbf{P}$ um eine Projektionsmatrix handelt, d.h. $\mathbf{P}^2 = \mathbf{P}$, gilt in Komponenten $P_{nn} = \sum_m P_{nm}^2 \geq P_{nn}^2$. Daher gilt $0 \leq P_{nn} \leq 1$. Außerdem ist die Summe der Diagonalelemente $\sum_n P_{nn} = \text{tr}(\mathbf{P}) = q + 1$. Die durchschnittliche Größe eines Diagonalelements ist also $(q + 1)/N$. Als Faustregel gilt $P_{nn} > 2(q + 1)/N = 0.276$. Die Türkei überschreitet diesen Wert (Tab. 4.2) und übt somit einen großen Einfluß auf die Modellanpassung aus. Eine ausführliche Diskussion dieser Thematik ist in Hoaglin und Welsch (1978), Fahrmeir et al. (1996, S. 117) und Fahrmeir et al. (2009, insb. Abs. 3.6) zu finden.

■

4.7 Regression mit quantitativen und qualitativen Regressoren: Heterogene Populationen

Das Einkommen Y von Personen hängt sicherlich vom Bildungsgrad x_1 ab (etwa gemessen in Jahren oder als höchster Abschluß), kann aber noch zusätzlich durch die Zugehörigkeit zu bestimmten Gruppen (etwa Geschlecht, $x_2 = 0$ (männlich), $x_2 = 1$ (weiblich)) verändert werden. Hat man diese moderierenden Variablen erhoben, so ist es sinnvoll, entweder die Analyse getrennt für die beiden Gruppen oder simultan durch Einbeziehung der moderierenden Variablen durchzuführen. Schreibt man

moderierende Variable

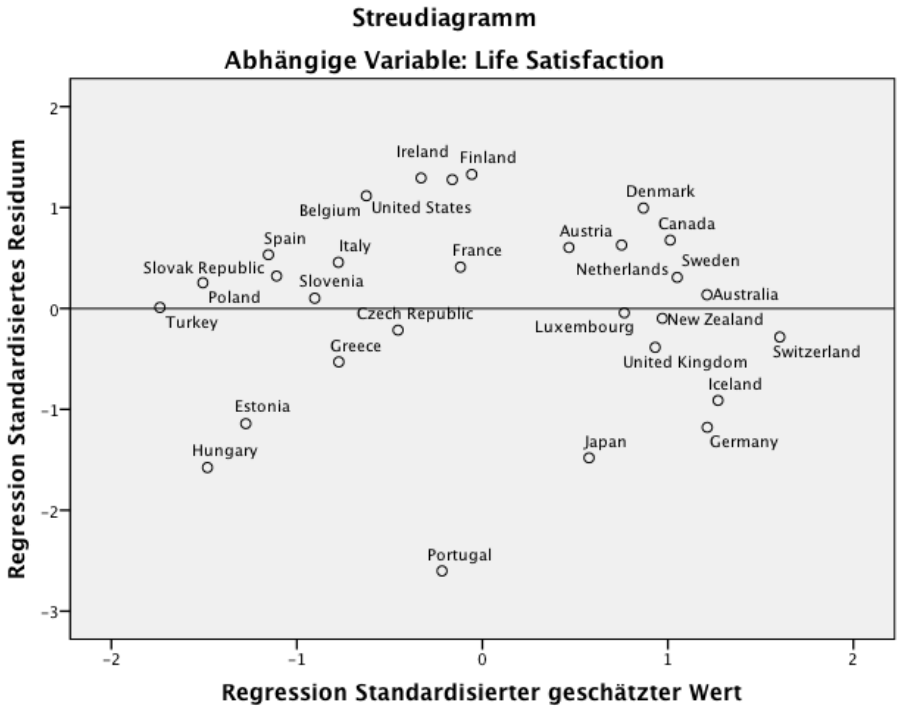


Abbildung 4.10: Residuen als Funktion des geschätzten Werts (oben). Unten: Normalverteilungs-Diagramm (P-P-Plot).

	P_{nn}	
21	0.04	Portugal
3	0.049	Belgium
5	0.065	Czech_Republic
6	0.071	Denmark
29	0.072	United_States
9	0.077	France
4	0.077	Canada
8	0.083	Finland
19	0.085	New_Zealand
24	0.092	Spain
25	0.1	Sweden
1	0.1	Australia
28	0.101	United_Kingdom
16	0.105	Japan
15	0.113	Italy
10	0.121	Germany
11	0.132	Greece
2	0.135	Austria
13	0.142	Iceland
14	0.152	Ireland
22	0.16	Slovak_Republic
12	0.174	Hungary
26	0.175	Switzerland
17	0.185	Luxembourg
18	0.208	Netherlands
23	0.212	Slovenia
20	0.215	Poland
7	0.253	Estonia
27	0.507	Turkey

Tabelle 4.2: Diagonalelemente P_{nn} der Hat-Matrix **P** nach Größe sortiert.

(Personen-Index n weggelassen)

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \tag{4.72}$$

so ergeben sich die gruppenspezifischen Regressionen

$$Y = \beta_0 + \beta_1 x_1 \qquad + \epsilon \qquad \text{(männlich)} \tag{4.73}$$
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 + \epsilon \qquad \text{(weiblich)} \tag{4.74}$$

Die gruppenspezifischen Achsenabschnitte sind also β_0 (männlich) und $\beta_0 + \beta_2$ (weiblich). Man kann auch schreiben $\beta_0(x_2) = \beta_0 + \beta_2 x_2$.
Auch die Steigung β_1 der Gerade könnte von x_2 abhängen. Analog kann man setzen: $\beta_1(x_2) = \beta_1 + \beta_3 x_2$. Die Steigung ist also β_1 oder $\beta_1 + \beta_3$.
Dies führt auf sog. Interaktionsterme der Form

Interaktionsterme

$$\beta_1(x_2)x_1 = \beta_1 x_1 + \beta_3\ x_2 \cdot x_1. \tag{4.75}$$

Insgesamt ergibt sich das Modell mit variablem Achsenabschnitt und Steigung

$$Y = (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3 x_2)x_1 + \epsilon \tag{4.76}$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2 + \epsilon \tag{4.77}$$

β_2 ist also die Änderung im Achsenabschnitt, während β_3 die Änderung der Steigung angibt. Diese Interpretation wird durch die 0-1-Dummy-Codierung ermöglicht.

Übung 4.5 (Codierung mit 1, 2)
Was passiert, wenn Sie die Variable `Geschlecht` mit den Werten 1, 2 codieren?
■

Ganz allgemein kann man die Parameter $\beta_i, i = 0, ..., q$ von Moderatorvariablen $z = 0, 1$ abhängen lassen. Folgende Notation erscheint als nützlich:

$$\beta_i(z) = \beta_i + \beta_{i1} z. \tag{4.78}$$

So entstehen Interaktions-Terme $\beta_i(z)x_i = \beta_i x_i + \beta_{i1}\ z \cdot x_i$. Auch die Parameter β_{i1} können wiederum von weiteren Variablen $z_2, ...$ abhängen, d.h. $\beta_{i1}(z_{2)} = \beta_{i1} + \beta_{i11} z_2$. Dies führt auf 3-fach-Interaktionen der Gestalt $z_1 \cdot z_2 \cdot x_i$. Hierbei wurde $z = z_1$ gesetzt.

Da die Analyse bedingt auf die Regressoren durchgeführt wird, kann man beliebige Interaktionsterme der Form $\beta_{ijk..}x_ix_jx_k...$ bilden, aber auch nichtlineare Terme $\beta_{ijk..}f(x_i, x_j, x_k...)$ sind möglich. Dabei muß die Funktion f vorgegeben werden. Solange der Parameter $\beta_{ijk..}$ nicht Teil der Funktion ist, kann man weiterhin mit dem linearen Regressionsmodell arbeiten. Hat man aber die Form $f(\boldsymbol{\beta}, x)$ so muß die Quadratsumme $S(\boldsymbol{\beta}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon}$ nichtlinear nach $\boldsymbol{\beta}$ optimiert werden (Minimum). Dann erhält man ein nichtlineares (in $\boldsymbol{\beta}$) Regressions-Modell.

Beispiel 4.4 (Moderator-Effekte)

Im folgenden werden mit SPSS simulierte Daten analysiert, die aus einer Regressionsgleichung mit Moderatoreffekten resultieren (Datensatz `Moderator.sav`). Die unabhängigen Variablen heißen x und $z = 0, 1$. Eine Betrachtung der Streudiagramme in Abb. 4.11 läßt einen Effekt der Variable z auf die Steigung vermuten. Ohne Berücksichtigung von z erhält man Schätzungen, die nahe bei 0 liegen und nicht signifikant sind (wenn man einen t-Wert von 2 zugrundelegt, also $\alpha = 0.05$). Bei getrennter Analyse ergeben sich signifikante Werte (außer für β_0 bei $z = 1$). Eine simultane Analyse mit dem Modell

$$Y = (\beta_0 + \beta_2 z) + (\beta_1 + \beta_3 z)x + \epsilon \tag{4.79}$$

$$= \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + \epsilon \tag{4.80}$$

ergibt die Schätzung

$$\hat{Y} = (2.283 - 0.251z) + (5.57 - 10.730z)x \tag{4.81}$$

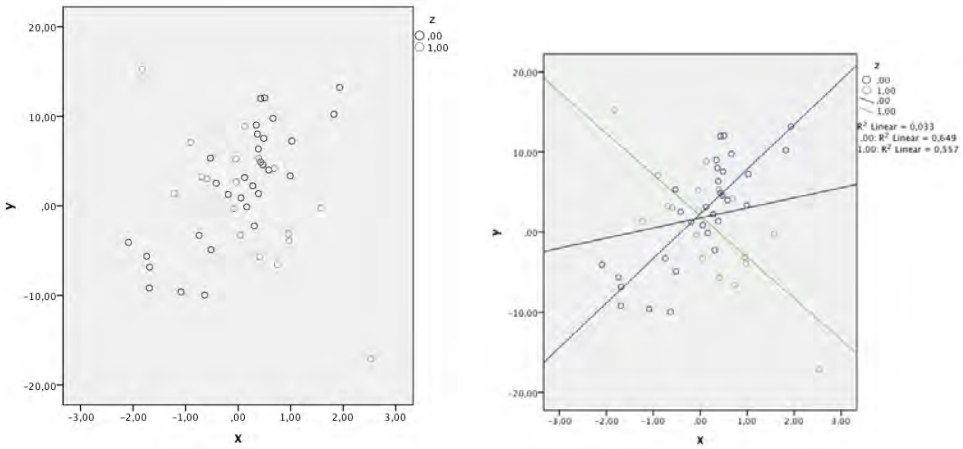
Die y-Werte wurden mit dem Kommando

```
COMPUTE y=1+2*z + (4-10*z)*x + 5*RV.NORMAL(0,1).
EXECUTE.
```

berechnet. Zuvor muß ein Syntax-Fenster mit

```
Datei/Öffnen/Syntax...
```

geöffnet werden. Man kann auch interaktiv arbeiten (Abb. 4.12). Die wahren Parameterwerte sind also $\beta_0 = 1, \beta_2 = 2, \beta_1 = 4, \beta_3 = -10, \sigma = 5$. Zuvor wurden die Variablen z und x erzeugt, und zwar mit den Zufallsgeneratoren `RV.Bernoulli(0.5)` und `RV.NORMAL(0,1)`. Anschließend erfolgte die Berechnung der Produktvariable xz mit Hilfe von



Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	Korrelationen		
		Regressionskoeffizient B	Standardfehler	Beta			Nullter Ordnung	Partiell	Teil
1	(Konstante)	1,774	,946		1,875	,067			
	x	1,258	,982	,182	1,281	,206	,182	,182	,182

a. Abhängige Variable: y

Koeffizienten[a]

z	Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	Korrelationen		
			Regressionskoeffizient B	Standardfehler	Beta			Nullter Ordnung	Partiell	Teil
,00	1	(Konstante)	2,283	,719		3,177	,004			
		x	5,570	,760	,806	7,326	,000	,806	,806	,806
1,00	1	(Konstante)	2,032	1,109		1,832	,085			
		x	-5,160	1,116	-,746	-4,625	,000	-,746	-,746	-,746

a. Abhängige Variable: y

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	Korrelationen		
		Regressionskoeffizient B	Standardfehler	Beta			Nullter Ordnung	Partiell	Teil
1	(Konstante)	2,283	,771		2,960	,005			
	x	5,570	,816	,805	6,827	,000	,182	,709	,624
	z	-,251	1,265	-,018	-,199	,843	-,093	-,029	-,018
	xz	-10,730	1,297	-,983	-8,274	,000	-,476	-,773	-,756

a. Abhängige Variable: y

Abbildung 4.11: Oben: Streudiagramm der Variablen y, x gruppiert nach $z = 0, 1$ (links) und mit linearer Anpassung (rechts). 2. Zeile: Regressionsanalyse ohne Gruppierung. 3. Zeile: Regressionsanalyse gruppiert nach z. 4. Zeile: Simultanes Modell $Y = (\beta_0 + \beta_2 z) + (\beta_1 + \beta_3 z)x + \epsilon$.

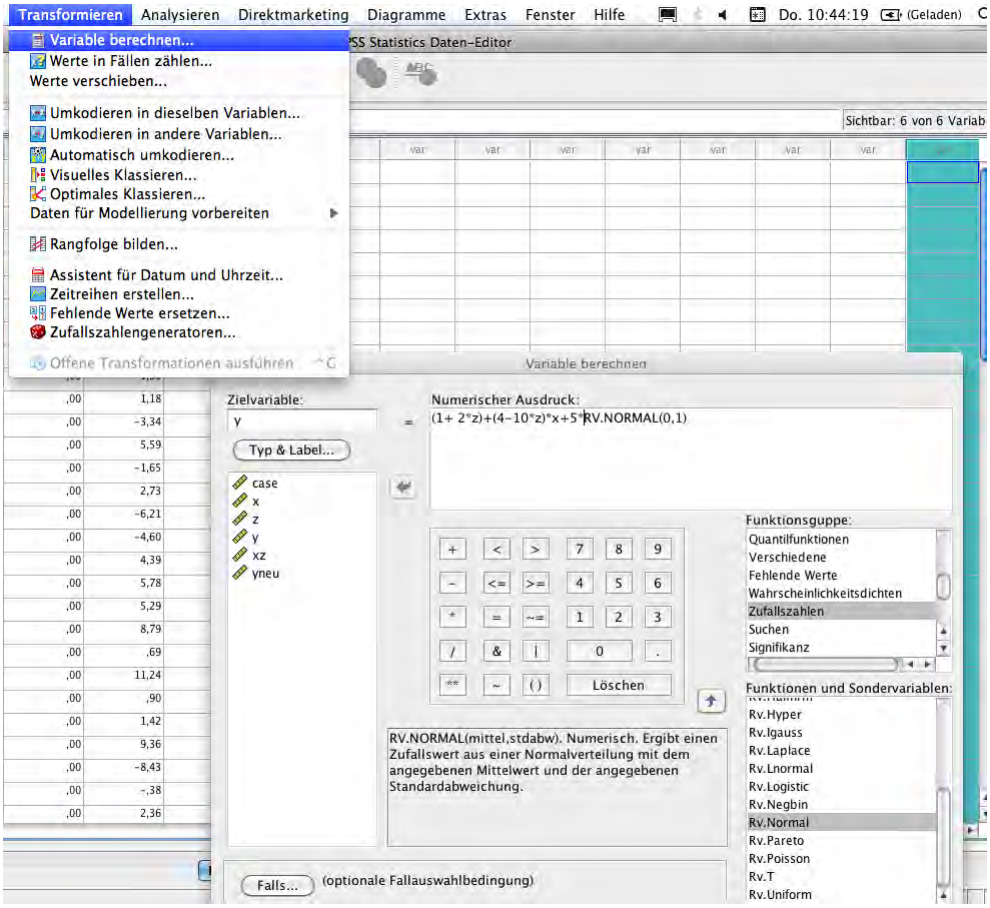


Abbildung 4.12: Berechnung der abhängigen Variablen y. Die wahren Parameterwerte sind $\beta_0 = 1, \beta_2 = 2, \beta_1 = 4, \beta_3 = -10, \sigma = 5$.

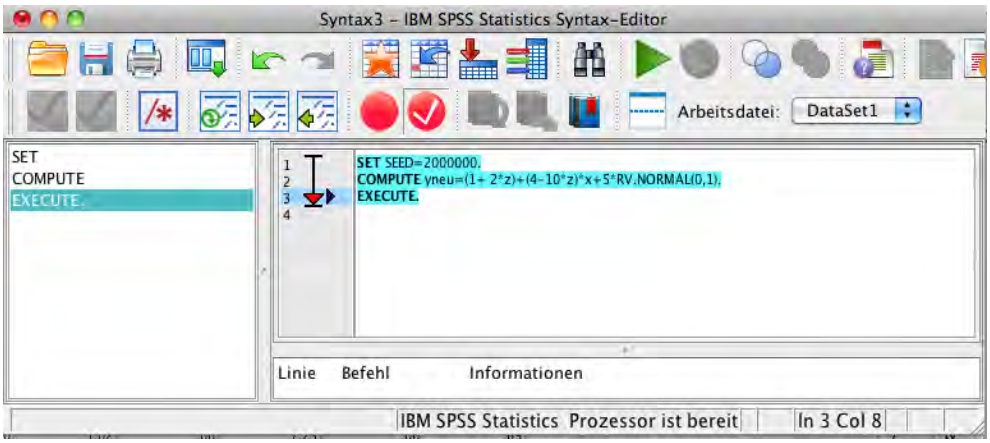


Abbildung 4.13: Berechnung der neuen abhängigen Variablen y_{neu} mit Syntax-Fenster. Der Zufallsgenerator wird mit `SET SEED=2000000.` initialisiert.

```
COMPUTE xz=x*z.
EXECUTE.
```

bzw. interaktiv (wie in Abb.4.12).

■

Übung 4.6 (Neuer Datenvektor y_{neu})

Berechnen Sie bitte einen neuen Datenvektor y_{neu} (Abb. 4.13) und führen Sie die obige Analyse mit diesen Daten durch. Warum ergeben sich andere Schätzwerte für $\boldsymbol{\beta}$ und σ?

■

Die bisherigen Analysen wurden mit dem Regressionsmodell `Analysieren/Regression/Linear/...` (Abb. 4.14) durchgeführt. Die Interaktionsvariable xz wurde dabei aus x und z als neue Variable erzeugt. Man kann auch das allgemeine lineare Modell benutzen, bei dem die Modellterme im Dialog eingegeben werden können (Abb. 4.15).

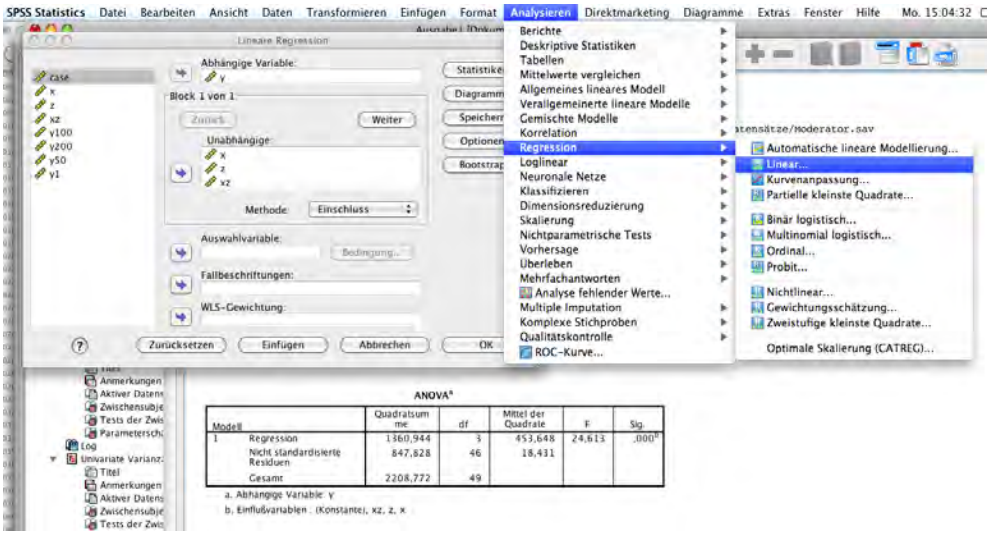
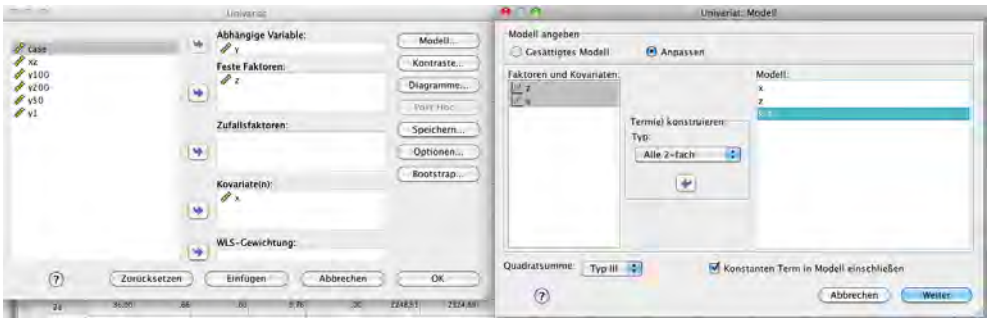


Abbildung 4.14: Regression mit Interaktionseffekten.



Parameterschätzer

Abhängige Variable: y

Parameter	Regressionskoeffizient B	Standardfehler	T	Sig.	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
Konstanter Term	2,032	1,002	2,027	,048	,014	4,049
x	-5,160	1,008	-5,118	,000	-7,189	-3,131
[z=,00]	,251	1,265	,199	,843	-2,294	2,797
[z=1,00]	0[a]	.	.	.	.	.
[z=,00] * x	10,730	1,297	8,274	,000	8,119	13,340
[z=1,00] * x	0[a]	.	.	.	.	.

a. Dieser Parameter wird auf Null gesetzt, weil er redundant ist.

Abbildung 4.15: Allgemeines lineares Modell mit Interaktionseffekten. Der Term -5.160 entspricht $\hat{\beta}_1+\hat{\beta}_3$, der konstante Term 2.032 entspricht $\hat{\beta}_0+\hat{\beta}_1$.

Kapitel 5

Varianzanalyse

5.1 Einleitung

Im Rahmen der Regressionsanalyse wurde angenommen, daß die abhängigen Variablen metrisch, die unabhängigen Variablen metrisch oder Indikator-Variablen (0-1) sind. In vielen Datensätzen sind jedoch alle x-Variablen kategorial und haben auch mehr als 2 Ausprägungen. In diesem Fall spricht man von Varianzanalyse. Bei q unabhängigen Variablen führt dies auf einen q-faktoriellen Versuchsplan. Diese Terminologie deutet darauf hin, daß man die Varianzanalyse bei geplanten Experimenten einsetzt, wo die Kombinationswirkung von Abstufungen der abhängigen Variablen untersucht wird, etwa in der Psychologie oder der Pharmakologie.

Auch in der Wirtschaftswissenschaft gibt es Anwendungen dieser Verfahren, z.B. in der Marktforschung. Etwa findet man unter `http://marktforschung.wikia.com/wiki/Varianzanalyse` die Ausführungen:

> So könnte man beispielsweise untersuchen, welche Wirkung verschiedene Formen der Werbung (z.B. Anzeigen in Zeitschriften, Plakate, Radiowerbespots etc.) auf die Verkaufszahlen eines bestimmten Produktes haben. Die abhängige Variable ist in diesem Fall die Anzahl der verkauften Einheiten oder der Gesamtumsatz, die unabhängige Variable - der Faktor - die durchgeführten Werbemaßnahmen. Es liegt eine einfaktorielle Varianzanalyse vor.
>
> Von Interesse könnte auch die Untersuchung der Wirkung zweier Faktoren auf den Verkauf sein, nämlich der Verpackung einer Ware und der Plazierung im Supermarktregal und zwar so-

wohl isoliert als auch gemeinsam. Da hier zwei Faktoren in die Varianzanalyse eingehen, handelt es sich um eine zweifaktorielle Varianzanalyse.

Im folgenden wird die Methodik in elementarer Form sowie in Form der Dummy- und Effekt-Kodierung erläutert.

5.2 Einfaktorielle Varianzanalyse mit fixen Effekten

5.2.1 Grundmodell

Im Fall *eines* Faktors (unabhängige Variable) handelt es sich nur um die Verallgemeinerung des Tests auf Gleichheit zweier Mittelwerte, d.h. man prüft $H_0 : \mu_1 = \mu_2 = ... = \mu_I$ in $i = 1, ..., I$ Populationen. Die unabhängige Variable $x = 1, ..., I$ erzeugt also eine Aufteilung der Stichprobe in I Subpopulationen.

Die Nullhypothese kann mit einer Reihe von t-Tests überprüft werden, wobei ein Signifikanzniveau von α durch Adjustierung der Einzeltests eingehalten werden kann. Die gesamte Testprozedur ist aber konservativ (vgl. Kap. 1.5). Hier wird ein simultaner Test hergeleitet.

Das univariate Modell hat die Form

$$Y_{ij} \sim N(\mu_i, \sigma^2), \quad i = 1, ..., I, j = 1, ..., J, \tag{5.1}$$

d.h. man hat gleich viele Beobachtungen J pro Subpopulation i (Faktorstufe i). Die statistischen Einheiten innerhalb Gruppe i werden mit dem Index j unterschieden.

In einer Notation, die eher der Regressionsanalyse ähnelt, kann man schreiben

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, ..., I, j = 1, ..., J. \tag{5.2}$$

Hierbei sind die Gleichungsfehler $\epsilon_{ij} \sim N(0, \sigma^2)$ voneinander unabhängig und identisch verteilt. Es gilt somit

$$E[Y_{ij}] = \mu_i \tag{5.3}$$

$$\mathrm{Cov}(Y_{ij}) = \sigma^2. \tag{5.4}$$

Die Parameter μ_i sind fix, aber unbekannt. Man spricht auch von fixen Effekten.

fixe Effekte

Die im letzten Kapitel eingeführte Matrix-Notation für das multiple Regressionsmodell ermöglicht die Darstellung

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \\ \vdots \\ Y_{I1} \\ \vdots \\ Y_{IJ} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_I \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1J} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2J} \\ \vdots \\ \epsilon_{I1} \\ \vdots \\ \epsilon_{IJ} \end{bmatrix} \tag{5.5}$$

In kompakter Form kann man schreiben[1]

$$\mathbf{y} = (\mathbf{I}_I \otimes \mathbf{1}_J)\boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{5.6}$$

$$= \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{5.7}$$

Die Notation $\mathbf{I}_I \otimes \mathbf{1}_J$ bedeutet, daß in der Einheitsmatrix $\mathbf{I}_I : I \times I$ jede Eins durch einen Vektor $\mathbf{1}_J$ und jede Null durch $\mathbf{0}_J$ ersetzt wird – es resultiert der obige explizite Ausdruck. Die Matrizen haben die Dimensionen $\mathbf{y} : IJ \times 1, \mathbf{I}_I \otimes \mathbf{1}_J : IJ \times I, \boldsymbol{\mu} : I \times 1, \boldsymbol{\epsilon} : IJ \times 1$. Die Matrix $\mathbf{X}$, welche die Struktur des Versuchsplans enthält, wird auch als Design-Matrix bezeichnet.

Design-Matrix

Mit Hilfe der Darstellung als multiples Regressionsmodell kann man die Parameter schätzen und Tests durchführen, etwa $H_0 : \mu_1 = \mu_2 = ... = \mu_I$.

5.2.2 Elementare Berechnung

Zunächst jedoch ein Zugang ohne Matrix-Operationen. Man kann durch einfaches Nachrechnen zeigen, daß eine Streuungszerlegung

[1] $\otimes$ ist das Kronecker-Produkt mit der Eigenschaft $\mathbf{A} \otimes \mathbf{B} = a_{ij}\mathbf{B}$, d.h. $(\mathbf{A} \otimes \mathbf{B})_{ik,jl} = a_{ij}b_{kl}$. Siehe Kap. 9.10

$$SQT = SQE + SQR \tag{5.8}$$

gilt, wobei

$$SQT = \sum_{ij}(Y_{ij} - \bar{Y}_{++})^2 \qquad \text{Totale Streuung} \tag{5.9}$$

$$SQE = \sum_{ij}(\bar{Y}_{i+} - \bar{Y}_{++})^2 \qquad \text{Erklärte Streuung} \tag{5.10}$$

$$SQR = \sum_{ij}(Y_{ij} - \bar{Y}_{i+})^2 \qquad \text{Residuale Streuung.} \tag{5.11}$$

Diese Zerlegung kann als Prognose-Regel für Y interpretiert werden:

Prognose-Regel

- **Falls Beobachtung zur Gruppe $x = i$ gehört:**
  Prognostiziere $\hat{Y} = \hat{\mu}_i = \bar{Y}_{i+}$ (Gruppenmittel).
  Prognose-Fehler: $F(+) = \sum_{ij}(Y_{ij} - \bar{Y}_{i+})^2 = SQR$.
  Dies ist die Streuung *innerhalb* der Gruppen.
- **ohne Berücksichtigung von x:**
  Prognostiziere $\hat{Y} = \hat{\mu} = \bar{Y}_{++}$ (totales Mittel).
  Prognose-Fehler: $F(-) = \sum_{ij}(Y_{ij} - \bar{Y}_{++})^2 = SQT$.
  Dies ist die Streuung ohne Berücksichtigung der Gruppen.

Die proportionale Fehler-Reduktion (PRE) bei Benutzung der Prognoseregel ist also

$$PRE = \frac{F(-) - F(+)}{F(-)} \tag{5.12}$$

$$= \frac{SQT - SQR}{SQT} = \frac{SQE}{SQT} \tag{5.13}$$

$$= R^2. \tag{5.14}$$

Determinationskoeffizient

Dies wird auch als Determinationskoeffizient bezeichnet. R kann eben-

falls als multiple Korrelation aufgefaßt werden (Korrelation zwischen den Daten $\mathbf{y}$ und der Prognose $\hat{\mathbf{y}}$).

Die Quadratsummen sind χ^2-verteilt und man kann zur Prüfung der H_0 einen F-Test durchführen.

Unter Gültigkeit der H_0 sind alle $\bar{Y}_{i+}$ ungefähr gleich dem Gesamtmittelwert $\bar{Y}_{++}$, sodaß SQE klein ist im Verhältnis zur Residualstreuung.

Unter H_0 ist somit $(SQE/(I-1))/(SQR/(IJ-I)) = MQE/MQR$ F-verteilt mit Freiheitsgraden $I-1, IJ-I$.

Übersichtlich schreibt man in Form einer ANOVA-Tabelle (Analysis of Variance)

ANOVA-Tafel

SQ	Formel	df	F-Statistik
SQE (zwischen)	$\sum_{ij}(\bar{Y}_{i+}-\bar{Y}_{++})^2$	$I-1$	$\frac{SQE/(I-1)}{SQR/(I(J-1))} \sim F(I-1, I(J-1))$
SQR (innerhalb)	$\sum_{ij}(Y_{ij}-\bar{Y}_{i+})^2$	$I(J-1)$	
SQT (total)	$\sum_{ij}(Y_{ij}-\bar{Y}_{++})^2$	$IJ-1$	

Für

$$F(I-1, I(J-1)) > f(1-\alpha; I-1, I(J-1)) \tag{5.15}$$

wird die H_0 auf dem Signifikanzniveau α verworfen.

Welche Mittelwertsunterschiede zu der Ablehnung der H_0 geführt haben, ist somit noch nicht geklärt.

Man führt daher im Anschluß an die signifikante Varianzanalyse t-Vergleiche (post hoc) durch, die aber mit adjustierten α-Fehlern gerechnet werden müssen, um das Signifikanzniveau des globalen Tests der H_0 einzuhalten.

Als Teststatistiken verwendet man die t-Brüche

$$T_{ii'} = \frac{\bar{Y}_{i+}-\bar{Y}_{i'+}}{S\sqrt{2/J}} \sim t(IJ-I). \tag{5.16}$$

Der Nenner ergibt sich aus der Überlegung

$$\mathrm{Var}(\bar{Y}_{i+}-\bar{Y}_{i'+}) = \mathrm{Var}(\bar{Y}_{i+}) + \mathrm{Var}(\bar{Y}_{i'+}) = 2\frac{\sigma^2}{J}, \tag{5.17}$$

Methode	1	2	3	4	5	6	7	8	$\bar{y}_i$	s_i
1	16	18	20	15	20	15	23	19	18.25	2.82
2	16	12	10	14	18	15	12	13	13.75	2.55
3	2	10	9	10	11	9	10	9	8.75	2.82
4	5	8	8	11	1	9	5	9	7.00	3.16

Tabelle 5.1: Abschlußtest von 4 Lehrmethoden.

da die Gruppen voneinander unabhängig sind und gleiche Varianzen aufweisen (Annahmen). Ersetzt man die Varianz durch die Schätzung

$$S^2 = \widehat{\sigma^2} = SQR/(I(J-1)) = MQR, \tag{5.18}$$

so ergibt sich obige t-Statistik.

Der Stoff wird in Aufgabe 5.1 vertieft.

Beispiel 5.1 (Vergleich von 4 Lehrmethoden)
Datensätze `Lehrmethoden.sav, Lehrmethoden.jmp`.

In einem Lehrgang wuden 4 verschiedene Unterrichtsmethoden benutzt. $N = 32$ Teilnehmer wurden auf die 4 Gruppen (Methoden) zufällig verteilt. Am Ende des Lehrgangs wurde ein Abschlußtest durchgeführt. Die entscheidende Frage lautet: Hat die Lehrmethode einen Einfluß auf den mittleren Punktwert?
Getestet wird also:

$$H_0: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ gegen} \tag{5.19}$$
$$H_1: \quad \mu_i \neq \mu_j \text{ für mindestens ein Paar } i, j. \tag{5.20}$$

Die Daten $y_{ij}, i = 1, \ldots, I = 4; j = 1, \ldots, J = 8$ sind in Tabelle 5.1 angegeben und in Abb. 5.1 graphisch dargestellt. Die Daten y_{ij} sind als Funktion der Lehrmethode i (x-Achse) als Punkte, Mittelwerte und Standardfehler der Mittelwerte $(s_i/\sqrt{J})$ aufgetragen. Um die Mittelwertsunterschiede zu betonen, wurden die Mittelwerte durch Linien verbunden, obwohl nur diskrete (nominale) Ausprägungen auf der x-Achse vorliegen. Aufgrund der Standardfehler unterscheiden sich vermutlich die Methoden 3 und 4 nicht (signifikant) voneinander.

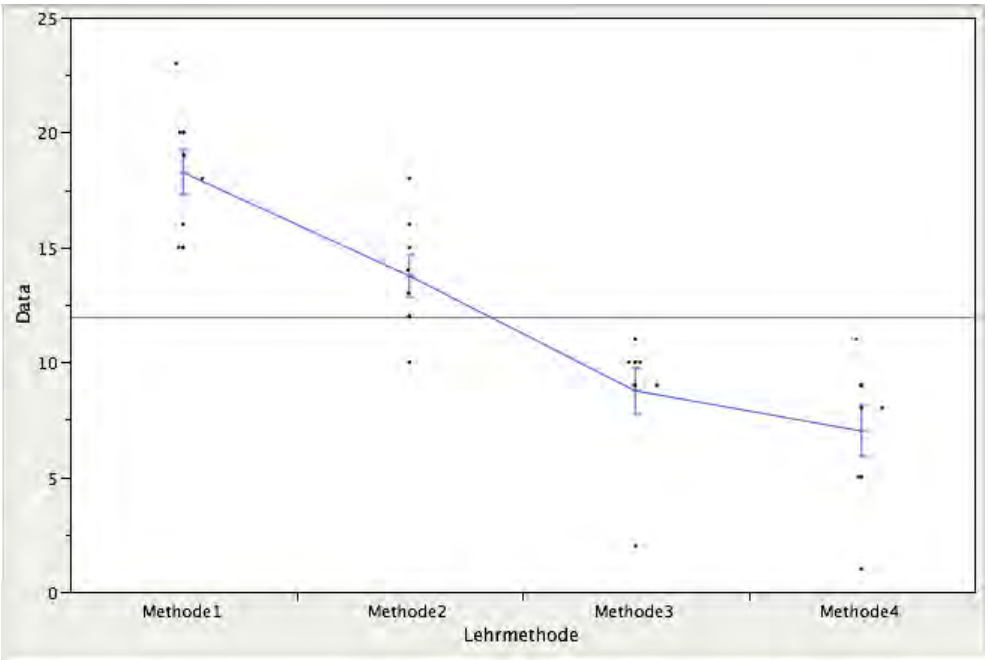


Abbildung 5.1: Abschlußtest von 4 Lehrmethoden als Funktion der Methode.

Durchführung des F-Tests:
Es gilt die vorteilhafte Umrechnung

$$SQE = \sum_{ij}(\bar{Y}_{i+} - \bar{Y}_{++})^2 = J\sum_i \bar{Y}_{i+}^2 - IJ\bar{Y}_{++}^2 \tag{5.21}$$

$$SQT = \sum_{ij}(Y_{ij} - \bar{Y}_{++})^2 = \sum_{ij} Y_{ij}^2 - IJ\bar{Y}_{++}^2. \tag{5.22}$$

Aus Tabelle 5.1 entnimmt man die Gruppenmittelwerte
$\bar{y}_{i+} = \{18.25, 13.75, 8.75, 7.00\}$ und erhält für den Gesamtmittelwert
$\bar{y}_{++} = \frac{1}{4}(18.25 + 13.75 + 8.75 + 7.00) = 11.94$.
Daraus findet man

$$\begin{aligned} SQE &= 8(18.25^2 + 13.75^2 + 8.75^2 + 7.00^2) - 32 \times 11.94^2 = 621.3750 \\ SQT &= 5408 - 32 \times 11.94^2 = 847.8750. \end{aligned}$$

Die residuale Quadratsumme ist somit
$SQR = SQT - SQE = 847.8750 - 621.3750 = 226.50$.
Für den F-Test ergibt sich ($IJ - I = 32 - 4 = 28$)

$$f(3, 28) = (621.3750/3)/(226.050/28) = 25.6049$$

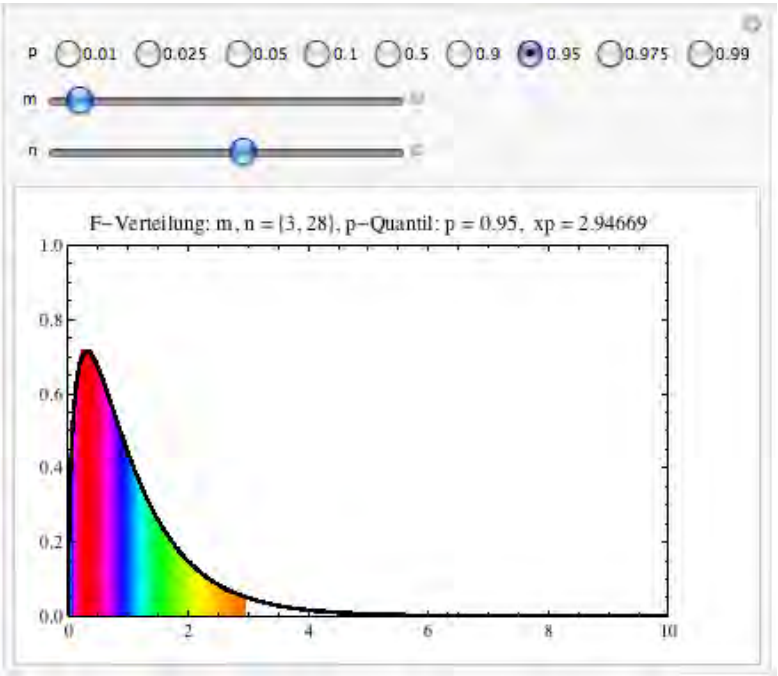


Abbildung 5.2: $f(0.95, 3, 28)$-Quantil.
http://www.fernuni-hagen.de/ls_statistik/lehre/

mit dem Quantil $f(.95, 3, 28) = 2.95$ (Abb. 5.2).

Damit muß H_0 abgelehnt werden: die Lehrmethoden sind also nicht gleich effektiv.

Der Determinationskoeffizient ist $R^2 = SQE/SQT = 621.357/847.875 = 0.733$, also wird 73% der Varianz im Testergebnis durch die Lehrmethode $X = i$ (unterschiedliche Gruppenmittelwerte) erklärt.

Alle wichtigen Resultate der 1-faktoriellen ANOVA sind nochmals im JMP-Output Abb. 5.3 zusammengefaßt. Um herauszufinden, welche der Gruppenunterschiede zu einem signifikanten Ergebnis geführt haben, können paarweise t-Tests gerechnet werden.

Im Beispiel ist $s^2 = SQR/(I(J-1)) = 226.50/28 = 8.0839; s = 2.8442$. Der Nenner der t-Statistik ist also $s\sqrt{2/J} = 2.8442\sqrt{2/8} = 1.4221$.

Die 6 t-Statistiken lauten (die signifikanten Vergleiche sind durch * gekennzeichnet)

Einfaktorielle ANOVA

Übersicht der Anpassung

r2	0,732862
r2 korrigiert	0,70424
Wurzel der mittleren quadratischen Abweichung	2,844167
Mittelwert der Zielgröße	11,9375
Beobachtungen (oder Summe Gewichte)	32

Varianzanalyse

Quelle	Freiheitsgrade	Summe Quadrate	Mittlere Quadrate	F-Wert	Wahrsch. > F
Lehrmethode	3	621,37500	207,125	25,6049	<,0001*
Fehler	28	226,50000	8,089		
K. Summe	31	847,87500			

Mittelwerte der einfaktoriellen ANOVA

Stufe	Anzahl	Mittelwert	Std.-Fehler	95% KI unten	95% KI oben
Methode1	8	18,2500	1,0056	16,190	20,310
Methode2	8	13,7500	1,0056	11,690	15,810
Methode3	8	8,7500	1,0056	6,690	10,810
Methode4	8	7,0000	1,0056	4,940	9,060

Std.-Fehler verwendet gepoolten Schätzer der Fehlervarianz

Abbildung 5.3: ANOVA (analysis of variance) für unterschiedliche Lehrmethoden (SAS/JMP).

$t_{ii'}$	18.25	13.75	8.75	7.00
18.25		3.1643*	6.6803*	7.91*
13.75			3.5159*	4.7465*
8.75				1.2306
7.00				

mit dem kritischen t-Wert $t(1-\alpha/(2k), df) = 2.83; k = 6, df = 28, \alpha = 0.05$ (Bonferroni-Adjustierung). Das nicht adjustierte Quantil ist dagegen $t(1-\alpha/2, df) = 2.048$, also wesentlich kleiner. Allerdings würde auch hier der Vergleich 3–4 nicht signifikant.

■

Der Stoff wird in Aufgabe 5.2 vertieft.

5.2.3 Berechnung mit dem linearen Modell

Der Test auf Gleichheit der Mittelwerte kann auch mit Hilfe der allgemeinen linearen Hypothese

$$H_0 : \mathbf{C}\boldsymbol{\mu} = \boldsymbol{\xi} \qquad H_1 : \mathbf{C}\boldsymbol{\mu} \neq \boldsymbol{\xi} \tag{5.23}$$

durchgeführt werden. Hierbei ist $\mathbf{C}$ eine Hypothesenmatrix und $\boldsymbol{\xi}$ ein konstanter Vektor. In diesem Fall setzt man

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} : (I-1) \times I && (5.24) \\ &= [\mathbf{I}_{I-1}, -\mathbf{1}_{I-1}] && (5.25) \end{aligned}$$

$$\boldsymbol{\xi} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} : (I-1) \times 1 \tag{5.26}$$

Ausmultiplizieren ergibt

$$\begin{aligned} \mu_1 - \mu_I &= 0 \\ \mu_2 - \mu_I &= 0 \\ &\vdots \\ \mu_{I-1} - \mu_I &= 0 \end{aligned}$$

Daher müssen alle Erwartungswerte μ_i gleich sein.
Setzt man als Regressormatrix $\mathbf{X} = \mathbf{I}_I \otimes \mathbf{1}_J$, so gilt

$$\mathbf{X}'\mathbf{X} = (\mathbf{I}_I \otimes \mathbf{1}'_J)(\mathbf{I}_I \otimes \mathbf{1}_J) = \mathbf{I}_I \otimes J = J\mathbf{I}_I \tag{5.27}$$

und

$$(\mathbf{X}'\mathbf{X})^{-1} = J^{-1}\mathbf{I}_I. \tag{5.28}$$

Daher ist der KQ-Schätzer für die Erwartungswerte in den Gruppen

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} && (5.29) \\ &= J^{-1}\mathbf{I}_I(\mathbf{I}_I \otimes \mathbf{1}'_J)\mathbf{y} && (5.30) \end{aligned}$$

oder explizit

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \bar{y}_{1+} \\ \bar{y}_{2+} \\ \vdots \\ \bar{y}_{I+} \end{bmatrix}. \tag{5.31}$$

Dies sind einfach die Gruppen-Mittelwerte, was völlig plausibel ist. Die formale Rechnung ist mit der Blockstruktur

$$(\mathbf{I}_I \otimes \mathbf{1}'_J)\mathbf{y} = \begin{bmatrix} \mathbf{1}'_J & \mathbf{0}' & \cdots & & & \mathbf{0}' \\ \mathbf{0}' & \mathbf{1}'_J & \mathbf{0}' & \cdots & & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{1}'_J & \mathbf{0}' & \cdots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \ddots & \mathbf{0}' & \mathbf{0}' \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0}' & \mathbf{0}' & \cdots & & & \mathbf{1}'_J \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \\ \mathbf{y}_I \end{bmatrix} \tag{5.32}$$

direkt nachvollziehbar.
Wie im Regressions-Kapitel gezeigt, ist die Kovarianz des Schätzers durch $\mathrm{Var}(\hat{\boldsymbol{\mu}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ gegeben. Hier gilt also explizit

$$\mathrm{Var}(\hat{\boldsymbol{\mu}}) = \frac{\sigma^2}{J}\mathbf{I}_I. \tag{5.33}$$

Dies stimmt mit der einfachen Berechnung $\mathrm{Var}(\bar{Y}_{i+}) = \mathrm{Var}(J^{-1}\sum_j Y_{ij}) = J^{-1}\sigma^2$ und $\mathrm{Cov}(\bar{Y}_{i+}, \bar{Y}_{j+}) = 0, i \neq j$ überein.
Ganz allgemein hat der KQ-Schätzer die Verteilung

$$\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{5.34}$$

Entsprechend ist unter $H_0 : \mathbf{C}\boldsymbol{\mu} = \boldsymbol{\xi}$

$$\mathbf{C}\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\xi}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'). \tag{5.35}$$

Man betrachtet daher die quadratische Form

$$Q = (\mathbf{C}\hat{\boldsymbol{\mu}} - \boldsymbol{\xi})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\mu}} - \boldsymbol{\xi}), \tag{5.36}$$

die $\sigma^2\chi^2$-verteilt ist mit $I-1$ Freiheitsgraden.
Die residuale Streuung

$$SQR = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'\mathbf{Q}\mathbf{y} \tag{5.37}$$

mit $\mathbf{Q} = \mathbf{I}_{IJ} - \mathbf{P}$, $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ hat $IJ - I$ Freiheitsgrade, da $\mathrm{tr}[\mathbf{I}_{IJ} - \mathbf{P}] = IJ - \mathrm{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = IJ - \mathrm{tr}[\mathbf{I}_I] = IJ - I$. Explizit gilt

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\mu}} = (\mathbf{I}_I \otimes \mathbf{1}_J)\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} \otimes \mathbf{1}_J \tag{5.38}$$

$$= \begin{bmatrix} \bar{y}_{1+}\mathbf{1}_J \\ \bar{y}_{2+}\mathbf{1}_J \\ \vdots \\ \bar{y}_{I+}\mathbf{1}_J \end{bmatrix} \tag{5.39}$$

und somit

$$SQR = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \sum_{ij}(y_{ij} - \hat{y}_{i+})^2. \tag{5.40}$$

Hierbei wurde die Rechenregel

$$(\mathbf{A} \otimes \mathbf{d})\mathbf{B} = (\mathbf{A} \otimes \mathbf{d})(\mathbf{B} \otimes 1) = \mathbf{AB} \otimes \mathbf{d} \tag{5.41}$$

benutzt (mit der Ersetzung $\mathbf{A} = \mathbf{I}_I, \mathbf{d} = \mathbf{1}_J, \mathbf{B} = \hat{\boldsymbol{\mu}}$).
Zur Ausführung des F-Tests

$$F = \frac{Q/(I-1)}{SQR/(IJ-I)} \sim F(I-1, IJ-I) \tag{5.42}$$

muß Q noch weiter ausgerechnet werden. Hier gilt speziell ($\boldsymbol{\xi} = \mathbf{0}$)

$$Q = (\mathbf{C}\hat{\boldsymbol{\mu}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\mu}}), \tag{5.43}$$

$$\mathbf{C}\hat{\boldsymbol{\mu}} = \begin{bmatrix} \bar{y}_{1+} - \bar{y}_{I+} \\ \bar{y}_{2+} - \bar{y}_{I+} \\ \vdots \\ \bar{y}_{I-1,+} - \bar{y}_{I+} \end{bmatrix}. \tag{5.44}$$

und

$$\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}' = J^{-1}[\mathbf{I}_{I-1}, -\mathbf{1}_{I-1}] \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix} \tag{5.45}$$

$$= J^{-1}(\mathbf{I}_{I-1} + \mathbf{1}_{I-1}\mathbf{1}'_{I-1}). \tag{5.46}$$

Dies ist proportional zu einer sogenannten Equikorrelations-Matrix $\mathbf{E}(\rho) = (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}'$ mit der Wahl $2\mathbf{E}(\frac{1}{2})$. Mit Hilfe der Formel

Equikorrelations-Matrix

$$(\mathbf{I}_{I-1} + \mathbf{1}_{I-1}\mathbf{1}'_{I-1})^{-1} = \mathbf{I}_{I-1} - I^{-1}\mathbf{1}_{I-1}\mathbf{1}'_{I-1} : (I-1) \times (I-1) \tag{5.47}$$

kann die Inverse berechnet werden, die aber wegen dem Faktor I^{-1} keine Zentrierungsmatrix $\mathbf{H}_{I-1}$ ist. Man findet jedoch

$$\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C} = J \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix} (\mathbf{I}_{I-1} - I^{-1}\mathbf{1}_{I-1}\mathbf{1}'_{I-1})[\mathbf{I}_{I-1}, -\mathbf{1}_{I-1}]$$

$$= J\,\mathbf{H}_I : I \times I \tag{5.48}$$

Dabei wurde $(\mathbf{I}_{I-1} - I^{-1}\mathbf{1}_{I-1}\mathbf{1}'_{I-1})\mathbf{1}_{I-1} = I^{-1}\mathbf{1}_{I-1}$ und $\mathbf{1}'_{I-1}I^{-1}\mathbf{1}_{I-1} = I^{-1}(I-1) = 1 - I^{-1}$ eingesetzt.
Damit ist die quadratische Form

$$Q = J\,\hat{\boldsymbol{\mu}}'\mathbf{H}_I\hat{\boldsymbol{\mu}} = \sum_{ij}(\bar{Y}_{i+} - \bar{Y}_{++})^2 = SQE \tag{5.49}$$

und man erhält den bekannten F-Test.

5.2.4 Effekt-Darstellung

Das Grundmodell (5.2) läßt sich umformulieren, wenn man die Effekte

Effekte

$$\alpha_i = \mu_i - \mu \tag{5.50}$$

$$\mu = I^{-1}\sum_i \mu_i \tag{5.51}$$

als Differenz vom globalen (theoretischen) Mittel einführt. Es gilt die Restriktion

$$\sum_i \alpha_i = \sum_i \mu_i - I\mu = 0. \tag{5.52}$$

Somit kann man in Effektdarstellung schreiben

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, ..., I, j = 1, ..., J. \tag{5.53}$$

Die H_0 lautet nun $\alpha_1 = \alpha_2 = ... = \alpha_I = 0$, bzw. $\mu_i = \mu$.

Reparametrisierung

Statt des Parametervektors $\boldsymbol{\mu} = [\mu_1, ..., \mu_I]' : I \times 1$ kann man $\tilde{\boldsymbol{\mu}} = [\mu, \alpha_1, ..., \alpha_{I-1}]' : I \times 1$ verwenden. Es handelt sich um eine sogenannte Reparametrisierung und es gilt folgende Umrechnung:

$$\tilde{\boldsymbol{\mu}} = \begin{bmatrix} I^{-1}\mathbf{1}' \\ \mathbf{H}_{-1} \end{bmatrix} \boldsymbol{\mu} := \mathbf{A}\boldsymbol{\mu}, \tag{5.54}$$

wobei $\mathbf{H} = \mathbf{I} - I^{-1}\mathbf{1}\mathbf{1}'$ und $\mathbf{H}_{-1} = \mathbf{H}_{1:I-1,1:I}$ eine Zentrierungsmatrix ohne letzte Zeile ist. Umgekehrt kann man schreiben

$$\boldsymbol{\mu} = [\mathbf{1}_I, \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix}]\tilde{\boldsymbol{\mu}} := \mathbf{A}^{-1}\tilde{\boldsymbol{\mu}} \tag{5.55}$$

Das lineare Modell in Effekt-Kodierung lautet explizit:

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1J} \\ Y_{21} \\ \vdots \\ Y_{2J} \\ \vdots \\ Y_{I-1,1} \\ \vdots \\ Y_{I-1,J} \\ Y_{I1} \\ \vdots \\ Y_{IJ} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ 1 & -1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & -1 & -1 & -1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_{I-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1J} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2J} \\ \vdots \\ \epsilon_{I-1,1} \\ \vdots \\ \epsilon_{I-1,J} \\ \epsilon_{I1} \\ \vdots \\ \epsilon_{IJ} \end{bmatrix}$$

Der Parameter α_I, der in $\tilde{\boldsymbol{\mu}}$ nicht vorkommt, ergibt sich als $\alpha_I = -\sum_{i=1}^{I-1} \alpha_i$. Dies wird durch die negativen Einsen der letzten J Zeilen bewirkt.

Etwas kompakter kann man schreiben

$$\mathbf{y} = \left[\mathbf{1}_I \otimes \mathbf{1}_J, \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix} \otimes \mathbf{1}_J \right] \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{\epsilon} \tag{5.56}$$

$$:= [\mathbf{X}_0, \mathbf{X}_\alpha] \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{\epsilon}. \tag{5.57}$$

Die Abkürzung

$$x^\alpha_{ii'} = \begin{cases} 1, & i = i' < I \\ -1, & i = I \\ 0, & sonst \end{cases} \tag{5.58}$$

Effektkodierung

$i = 1, ..., I$, $i' = 1, ..., I-1$ bzw. als Matrix

$$\mathbf{x}^\alpha = \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & \dots & -1 \end{bmatrix} : I \times (I-1) \tag{5.59}$$

ergibt die Darstellung

$$Y_{ij} = \mu + \sum_{i'=1}^{I-1} x_{ii'}^{\alpha} \alpha_{i'} + \epsilon_{ij}, \quad i = 1, ..., I, j = 1, ..., J. \tag{5.60}$$

oder

$$\mathbf{y} = [\mathbf{1}_I \otimes \mathbf{1}_J, \mathbf{x}^{\alpha} \otimes \mathbf{1}_J] \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{\epsilon}. \tag{5.61}$$

Das Modell in Effektdarstellung ergibt sich aus dem Grundmodell

$$\mathbf{y} = (\mathbf{I}_I \otimes \mathbf{1}_J)\boldsymbol{\mu} + \boldsymbol{\epsilon} \tag{5.62}$$

durch die Ersetzung (Reparametrisierung)

$$(\mathbf{I}_I \otimes \mathbf{1}_J)\mathbf{A}^{-1}\tilde{\boldsymbol{\mu}} = (\mathbf{I}_I \otimes \mathbf{1}_J)[\mathbf{1}_I, \mathbf{x}^{\alpha}]\tilde{\boldsymbol{\mu}} \tag{5.63}$$

$$= [\mathbf{1}_I \otimes \mathbf{1}_J, \mathbf{x}^{\alpha} \otimes \mathbf{1}_J]\tilde{\boldsymbol{\mu}} \tag{5.64}$$

$$= [\mathbf{X}_0, \mathbf{X}_{\alpha}]\tilde{\boldsymbol{\mu}}. \tag{5.65}$$

Hierbei wurde die Rechenregel

$$(\mathbf{A} \otimes \mathbf{d})[\mathbf{B}, \mathbf{C}] = [\mathbf{AB}, \mathbf{AC}] \otimes \mathbf{d} = [\mathbf{AB} \otimes \mathbf{d}, \mathbf{AC} \otimes \mathbf{d}] \tag{5.66}$$

benutzt.

Die KQ-Schätzungen der Effekte ergeben sich aus den Schätzungen für $\boldsymbol{\mu}$ durch die Umrechnung $\tilde{\boldsymbol{\mu}} = \mathbf{A}\boldsymbol{\mu}$. Explizit gilt

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} = \bar{Y}_{i+} - \bar{Y}_{++} \tag{5.67}$$

$$\hat{\mu} = I^{-1} \sum_i \hat{\mu}_i = I^{-1} \sum_i \bar{Y}_{i+} = \bar{Y}_{++} \tag{5.68}$$

Mit Hilfe der geschätzten Effekte gilt

$$SQE = \sum_{ij} (\bar{Y}_{i+} - \bar{Y}_{++})^2 = \sum_{ij} \hat{\alpha}_i^2 \tag{5.69}$$

(von den Effekten erklärte Quadratsumme).

Beispiel 5.2 (Vergleich von 4 Lehrmethoden)
Die Gruppenmittelwerte waren
$\bar{y}_{i+} = \{18.25, 13.75, 8.75, 7.00\}$
und der Gesamtmittelwert lautet
$\bar{y}_{++} = \frac{1}{4}(18.25 + 13.75 + 8.75 + 7.00) = 11.94.$
Daraus findet man die geschätzten Effekte
$\hat{\alpha}_i = \{6.31, 1.81, -3.19, -4.94\}$
und die erklärte Quadratsumme
$SQE = \sum_{ij} \hat{\alpha}_i^2 = 621.375$
Die Darstellung der Varianzanalyse mit Hilfe der Dummy- und Effekt-Kodierung soll im folgenden nochmals verdeutlicht werden. Abb. 5.4 zeigt den Datensatz mit Dummykodierung (Spalten mu1–mu4) und Effektkodierung (Spalten mu, alpha1–alpha3). Dies entspricht den Matrizen $\mathbf{I}_I \otimes \mathbf{1}_J$ und $[\mathbf{1}_I \otimes \mathbf{1}_J, \mathbf{x}^\alpha \otimes \mathbf{1}_J]$. Die Spalte x hat die nominale Kodierung 1–4.
Die Ergebnisse der KQ-Schätzungen sind in Abb. 5.5 dargestellt.

Wichtig ist, daß im Regressions-Programm der konstante Wert (Achsenabschnitt, intercept) gleich Null gesetzt wird, da schon 4 Parameter $(\mu_1 ... \mu_4)$ oder $(\mu, \alpha_1, ..., \alpha_3)$ geschätzt werden. Ansonsten wird vom Programm ein redundanter Parameter entfernt.
Alternativ kann man nur die Parameter $\mu_1 ... \mu_3$ oder $\alpha_1, ..., \alpha_3$ auswählen und das Regressionsmodell mit Konstante schätzen (Abb. 5.6). Die Konstante (Achsenabschnitt) korrespondiert in diesem Fall zu einer Referenzkategorie.

Man erkennt in Abb. 5.5 die schon vertrauten Werte für die Quadratsummen und die Parameterschätzungen. Die Schätzungen für α_i sind korreliert, da $\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu}$ gemeinsame Anteile hat.
Die geschätzte Kovarianzmatrix für $\hat{\boldsymbol{\mu}}$ ist

$$\begin{aligned}
\widehat{\mathrm{Var}}(\hat{\boldsymbol{\mu}}) &= \frac{MQR}{J}\mathbf{I}_I = \frac{SQR}{(IJ-I)J}\mathbf{I}_I && (5.70)\\
&= 226.5/(28 \cdot 8)\ \mathbf{I}_4 && (5.71)\\
&= \begin{bmatrix} 1.01116 & 0. & 0. & 0. \\ 0. & 1.01116 & 0. & 0. \\ 0. & 0. & 1.01116 & 0. \\ 0. & 0. & 0. & 1.01116 \end{bmatrix}. && (5.72)
\end{aligned}$$

	Lehrmethode	Data	x	mu1	mu2	mu3	mu4	mu	alpha1	alpha2	alpha3
1	Methode1	16	1	1	0	0	0	1	1	0	0
2	Methode1	18	1	1	0	0	0	1	1	0	0
3	Methode1	20	1	1	0	0	0	1	1	0	0
4	Methode1	15	1	1	0	0	0	1	1	0	0
5	Methode1	20	1	1	0	0	0	1	1	0	0
6	Methode1	15	1	1	0	0	0	1	1	0	0
7	Methode1	23	1	1	0	0	0	1	1	0	0
8	Methode1	19	1	1	0	0	0	1	1	0	0
9	Methode2	16	2	0	1	0	0	1	0	1	0
10	Methode2	12	2	0	1	0	0	1	0	1	0
11	Methode2	10	2	0	1	0	0	1	0	1	0
12	Methode2	14	2	0	1	0	0	1	0	1	0
13	Methode2	18	2	0	1	0	0	1	0	1	0
14	Methode2	15	2	0	1	0	0	1	0	1	0
15	Methode2	12	2	0	1	0	0	1	0	1	0
16	Methode2	13	2	0	1	0	0	1	0	1	0
17	Methode3	2	3	0	0	1	0	1	0	0	1
18	Methode3	10	3	0	0	1	0	1	0	0	1
19	Methode3	9	3	0	0	1	0	1	0	0	1
20	Methode3	10	3	0	0	1	0	1	0	0	1
21	Methode3	11	3	0	0	1	0	1	0	0	1
22	Methode3	9	3	0	0	1	0	1	0	0	1
23	Methode3	10	3	0	0	1	0	1	0	0	1
24	Methode3	9	3	0	0	1	0	1	0	0	1
25	Methode4	5	4	0	0	0	1	1	-1	-1	-1
26	Methode4	8	4	0	0	0	1	1	-1	-1	-1
27	Methode4	8	4	0	0	0	1	1	-1	-1	-1
28	Methode4	11	4	0	0	0	1	1	-1	-1	-1
29	Methode4	1	4	0	0	0	1	1	-1	-1	-1
30	Methode4	9	4	0	0	0	1	1	-1	-1	-1
31	Methode4	5	4	0	0	0	1	1	-1	-1	-1
32	Methode4	9	4	0	0	0	1	1	-1	-1	-1

Abbildung 5.4: Datensatz mit expliziter Dummy- und Effektkodierung (Datensatz `Lehrmethoden.jmp`).

Übersicht der Anpassung

r^2	0,732862
r^2 korrigiert	0,70424
Wurzel der mittleren quadratischen Abweichung	2,844167
Mittelwert der Zielgröße	11,9375
Beobachtungen (oder Summe Gewichte)	32

Varianzanalyse

Quelle	Freiheitsgrade	Summe Quadrate	Mittlere Quadrate	F-Wert
Modell	3	621,37500	207,125	25,6049
Fehler	28	226,50000	8,089	Wahrsch. > F
K. Summe	31	847,87500		<,0001*

Getestet gegen reduziertes Modell: Y = Mittelwert

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch.>|t|
mu1	18,25	1,005565	18,15	<,0001*
mu2	13,75	1,005565	13,67	<,0001*
mu3	8,75	1,005565	8,70	<,0001*
mu4	7	1,005565	6,96	<,0001*

Korrelation der Schätzer

Korr.

	mu1	mu2	mu3	mu4
mu1	1,0000	0,0000	0,0000	0,0000
mu2	0,0000	1,0000	0,0000	0,0000
mu3	0,0000	0,0000	1,0000	0,0000
mu4	0,0000	0,0000	0,0000	1,0000

Übersicht der Anpassung

r^2	0,732862
r^2 korrigiert	0,70424
Wurzel der mittleren quadratischen Abweichung	2,844167
Mittelwert der Zielgröße	11,9375
Beobachtungen (oder Summe Gewichte)	32

Varianzanalyse

Quelle	Freiheitsgrade	Summe Quadrate	Mittlere Quadrate	F-Wert
Modell	3	621,37500	207,125	25,6049
Fehler	28	226,50000	8,089	Wahrsch. > F
K. Summe	31	847,87500		<,0001*

Getestet gegen reduziertes Modell: Y = Mittelwert

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch.>|t|
mu	11,9375	0,502782	23,74	<,0001*
alpha1	6,3125	0,870845	7,25	<,0001*
alpha2	1,8125	0,870845	2,08	0,0467*
alpha3	-3,1875	0,870845	-3,66	0,0010*

Korrelation der Schätzer

Korr.

	mu	alpha1	alpha2	alpha3
mu	1,0000	0,0000	0,0000	0,0000
alpha1	0,0000	1,0000	-0,3333	-0,3333
alpha2	0,0000	-0,3333	1,0000	-0,3333
alpha3	0,0000	-0,3333	-0,3333	1,0000

Abbildung 5.5: KQ-Schätzungen mit Dummy- und Effektkodierung (ohne intercept).

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch.>|t|
Achsenabschnitt	7	1,005565	6,96	<,0001*
mu1	11,25	1,422083	7,91	<,0001*
mu2	6,75	1,422083	4,75	<,0001*
mu3	1,75	1,422083	1,23	0,2287

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch.>|t|
Achsenabschnitt	11,9375	0,502782	23,74	<,0001*
alpha1	6,3125	0,870845	7,25	<,0001*
alpha2	1,8125	0,870845	2,08	0,0467*
alpha3	-3,1875	0,870845	-3,66	0,0010*

Abbildung 5.6: KQ-Schätzungen mit Dummy- und Effektkodierung (mit intercept). Die Konstante korrespondiert zu einer Referenzkategorie ($i = 4$).

Dies ergibt die Standardfehler Std $= [1.00556, 1.00556, 1.00556, 1.00556]'$.
Daraus findet man unter Benutzung der Transformationsmatrix

$$\mathbf{A} = \begin{bmatrix} I^{-1}\mathbf{1}' \\ \mathbf{H}_{-1} \end{bmatrix} \tag{5.73}$$

$$= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \end{bmatrix} \tag{5.74}$$

die Effekte

$$\hat{\hat{\boldsymbol{\mu}}} = \mathbf{A}\hat{\boldsymbol{\mu}} = [11.9375, 6.3125, 1.8125, -3.1875]'. \tag{5.75}$$

Deren geschätzte Kovarianzmatrix ist

$$\widehat{\mathrm{Var}}(\hat{\hat{\boldsymbol{\mu}}}) = \mathbf{A}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\mu}})\mathbf{A}' \tag{5.76}$$

$$= \begin{bmatrix} 0.25279 & 0 & 0 & 0 \\ 0 & 0.758371 & -0.25279 & -0.25279 \\ 0 & -0.25279 & 0.758371 & -0.25279 \\ 0 & -0.25279 & -0.25279 & 0.758371 \end{bmatrix}, \tag{5.77}$$

die Standardfehler sind $[0.502782, 0.870845, 0.870845, 0.870845]$ und die Korrelations-Matrix ist

$$\widehat{\mathrm{Corr}}(\hat{\hat{\boldsymbol{\mu}}}) = \begin{bmatrix} 1. & 0. & 0. & 0. \\ 0. & 1. & -0.333333 & -0.333333 \\ 0. & -0.333333 & 1. & -0.333333 \\ 0. & -0.333333 & -0.333333 & 1. \end{bmatrix} \tag{5.78}$$

Schätzt man das Modell in Dummy-Codierung mit Achsenabschnitt (intercept), so ist die Designmatrix durch $\mathbf{X} = [\mathbf{1}_I, \mathbf{I}_{1:I,1:I-1}] \otimes \mathbf{1}_J$ gegeben, wobei die erste Spalte zum Achsenabschnitt korrespondiert. Die Parameter (bzw. Schätzungen) in Abb. 5.6 sind nicht die ursprünglichen Werte μ_i, sondern durch die Reparametrisierung $\mu_i^* = \mu_i - \mu_I$ als Abweichung von μ_4 gegeben. Daher ergeben sich auch andere Standardfehler, da die Schätzungen $\hat{\mu}_i^*$ korreliert sind.

■

Der Stoff wird in Aufgabe 5.3 vertieft.

5.2.5 Multiple Mittelwertsvergleiche

Wenn die Nullhypothese $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_I = 0$, bzw. $\mu_1 = \mu_2 = ... = \mu_I = \mu$ abgelehnt wird, weiß man nicht, welche Effekte die Ablehnung verursacht haben. Man weiß nur, daß mindestens *ein* Gleichheitszeichen nicht gilt.

Man betrachtet daher Hypothesen der Form

$$H_0 : \sum_i c_i \mu_i \quad = \quad \psi_0 \tag{5.79}$$

mit der Nebenbedingung $\sum_i c_i = 0$.

lineare Kontraste

Die Linearkombinationen

$$\psi = \mathbf{c}'\boldsymbol{\mu} \tag{5.80}$$

werden als lineare Kontraste bezeichnet.

Paarvergleiche

Paarvergleiche $\mu_k - \mu_l$ haben die spezielle Form

$$\mathbf{c} = [0, ..., 1, 0, ..., -1, 0, ..., 0]'. \tag{5.81}$$

Als Schätzfunktion für den Kontrast ergibt sich

$$\hat{\psi} = \mathbf{c}'\hat{\boldsymbol{\mu}} = \sum_i c_i \bar{Y}_{i+} \tag{5.82}$$

mit der Varianz $\mathrm{Var}(\hat{\psi}) = \sum_i c_i^2 \mathrm{Var}(\bar{Y}_{i+}) = \frac{\sigma^2}{J} \sum_i c_i^2$. Ersetzt man den Varianzparameter durch die Schätzung $MQR = SQR/(IJ - I)$, so ergibt sich die Teststatistik

$$\frac{\hat{\psi} - \psi_0}{\sqrt{\frac{MQR}{J} \sum_i c_i^2}} = \frac{\sum_i c_i \bar{Y}_{i+} - \psi_0}{\sqrt{\frac{SQR}{(IJ-I)J} \sum_i c_i^2}} \sim t(IJ - I) \tag{5.83}$$

und nach Umformung die Konfidenzintervalle ($\psi_0 \to \psi$)

$$P\{\hat{\psi} - t\sqrt{...} \leq \psi \leq \hat{\psi} + t\sqrt{...}\} = 1 - \alpha \tag{5.84}$$

mit dem t-Quantil $t = t(1-\alpha/2, IJ-I)$. Explizit kann man schreiben

$$\left(\sum_i c_i \bar{Y}_{i+}\right) \pm t(1-\alpha/2, IJ-I)\left[\frac{SQR}{(IJ-I)J}\sum_i c_i^2\right]^{1/2}. \qquad (5.85)$$

Im allgemeinen ist man jedoch an mehreren Kontrasten, z.B. allen Paarvergleichen $\mu_k - \mu_l, k < l$ interessiert. Damit ist jedoch, wie schon bekannt, eine Erhöhung des simultanen Signifikanzniveaus verbunden (Fehlerrate des gesamten Experiments).

Die einfachste Möglichkeit ist wieder, die R einzelnen t-Tests nach Bonferroni mit adjustierten Signifikanz-Niveaus $\alpha_r, \sum_{r=1}^{R} \alpha_r = \alpha$, durchzuführen.

Entsprechendes gilt für die t-Konfidenzintervalle:

$$\left(\sum_i c_i \bar{Y}_{i+}\right) \pm t(1-\alpha_r/2, IJ-I)\left[\frac{SQR}{(IJ-I)J}\sum_i c_i^2\right]^{1/2} \qquad (5.86)$$

Bonferroni-Konfidenz-Intervall

Eine exakte Alternative besteht darin, simultane Konfidenzintervalle nach Scheffé zu untersuchen.

Analog zur Konstruktion der simultanen Konfidenzintervalle in Kap. 3.2.5 betrachtet man alle möglichen Projektionen $\mathbf{c}'\boldsymbol{\mu}$ (Kontraste). Der F-Test für H_0 wird signifikant, wenn irgendeines der Intervalle

$$\left(\sum_i c_i \bar{Y}_{i+}\right) \pm [(I-1)F]^{1/2}\left[\frac{SQR}{(IJ-I)J}\sum_i c_i^2\right]^{1/2} \qquad (5.87)$$

Scheffé-Konfidenz-Intervall

$F = F(1-\alpha, I-1, IJ-I)$ den Wert 0 nicht überdeckt (vgl. Abb. 3.6).

Beispiel 5.3 (Vergleich von 4 Lehrmethoden)

Die Paarvergleiche nach Bonferroni wurden bereits in Bsp. 5.1 betrachtet.

Im folgenden werden Paarvergleiche nach der Scheffé-Methode berechnet.

Die Projektions-Vektoren haben die Form $\mathbf{c} = [0, ...1, 0, ..., -1, 0, ...0]$ mit $\sum_i c_i^2 = \mathbf{c}'\mathbf{c} = 2$.
Aus Tabelle 5.1 entnimmt man die Gruppenmittelwerte
$\bar{y}_{i+} = \{18.25, 13.75, 8.75, 7.00\}$ und den Gesamtmittelwert
$\bar{y}_{++} = \frac{1}{4}(18.25 + 13.75 + 8.75 + 7.00) = 11.94$.
Die residuale Quadratsumme war $SQR = SQT - SQE = 847.8750 - 621.3750 = 226.50$ und $I = 4$, $J = 8$, $IJ - I = 32 - 4 = 28$.
Wählt man $\alpha = 0.1$, ergibt sich für das F-Quantil $F(0.9, 3, 28) = 2.29$ (Abb. 5.7) und für $\alpha = 0.05$ der Wert $F(0.95, 3, 28) = 2.95$. Daraus findet man die Grenzen der Konfidenzintervalle
$\sqrt{3 \cdot 2.29}\sqrt{\frac{226.50 \cdot 2}{28 \cdot 8}} = 2.62 \cdot 1.42 = 3.73$ und
$\sqrt{3 \cdot 2.95}\sqrt{\frac{226.50 \cdot 2}{28 \cdot 8}} = 2.97 \cdot 1.42 = 4.23$.
Die Kontraste $\bar{y}_{i+} - \bar{y}_{i'+}$ lauten

$\bar{y}_{i+} - \bar{y}_{i'+}$	18.25	13.75	8.75	7.00
18.25		4.50	9.50	11.25
13.75			5.00	6.75
8.75				1.75
7.00				

Daher überdeckt das Intervall $1.75 \pm 3.73(4.23)$ den Wert 0. Deshalb würde man die Hypothese $\mu_3 - \mu_4 = 0$ beibehalten ($\alpha = 0.1$ und 0.05).
■

5.2.5.1 Herleitung der Scheffé-Intervalle

In Kap. 5.2.3 wurde die lineare Hypothese $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ mit Hilfe des linearen Modells geprüft. Da der KQ-Schätzer die Verteilung $\hat{\boldsymbol{\mu}} \sim N(\boldsymbol{\mu}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$ aufwies, ist

$$\mathbf{C}\hat{\boldsymbol{\mu}} \sim N(\mathbf{C}\boldsymbol{\mu}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'). \tag{5.88}$$

Man betrachtet daher die quadratische Form

$$Q = (\mathbf{C}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}))'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}), \tag{5.89}$$

die $\sigma^2\chi^2$-verteilt ist mit $I - 1$ Freiheitsgraden. Der Quotient

$$\frac{Q/(I-1)}{SQR/(IJ-I)} \sim F(I-1, IJ-I) \tag{5.90}$$

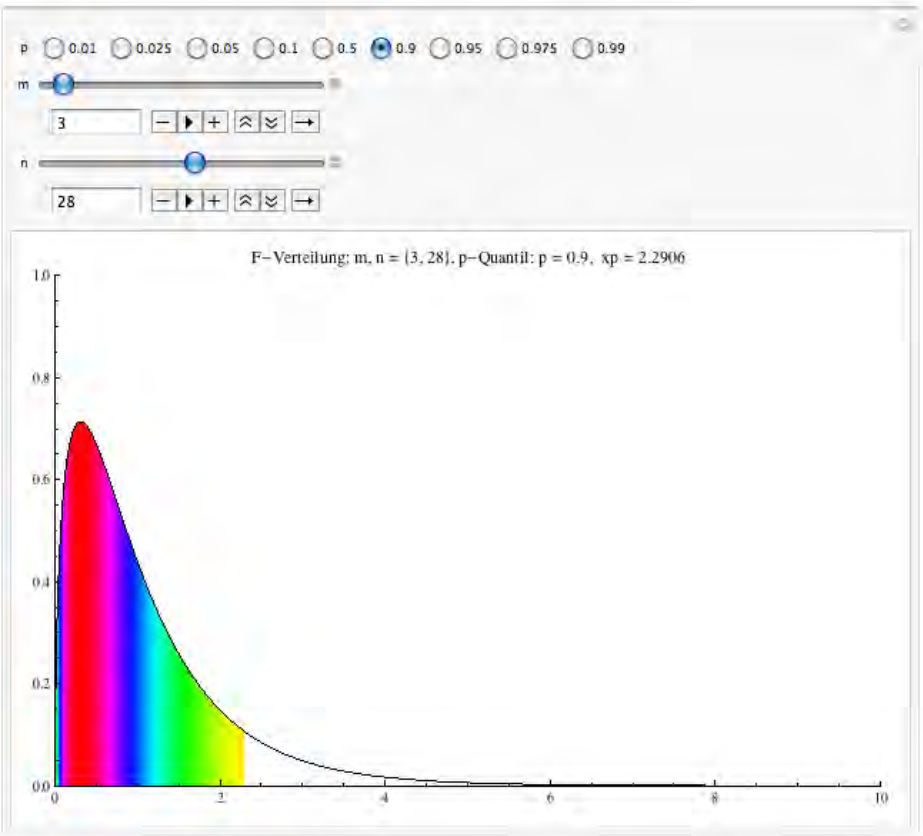


Abbildung 5.7: $f(0.9, 3, 28)$-Quantil.
(siehe http://www.fernuni-hagen.de/ls_statistik/lehre/).

ist also F-verteilt. Die Prüfung der Nullhypothese ist äquivalent zur Frage, ob das Ellipsoid

$$\frac{Q/(I-1)}{SQR/(IJ-I)} \leq F(1-\alpha, I-1, IJ-I) \tag{5.91}$$

den Wert $\boldsymbol{\mu} = \mathbf{0}$ überdeckt.
Die quadratische Form Q kann als Maximum der Projektionen

$$Z_{\mathbf{a}}^2 = \frac{|\mathbf{a}'\mathbf{C}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})|^2}{\mathbf{a}'\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\mathbf{a}} \tag{5.92}$$

im Sinne des Union-Intersection-Prinzips aufgefaßt werden (vgl. Kap. 3.2.5, Glg. 3.32). Setzt man $\mathbf{c}' = \mathbf{a}'\mathbf{C}$, so ergeben sich die simultanen Konfidenzintervalle

$$|\mathbf{c}'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})| \leq \sqrt{(I-1)F(1-\alpha, I-1, IJ-I)}\sqrt{\frac{SQR}{IJ-I}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}} \tag{5.93}$$

für alle $\mathbf{c}$. Es gilt $\sum_i c_i = 0$, wie es für Kontraste gelten muß. Dies folgt aus der Form (5.24) für $\mathbf{C}$.

Übung 5.1
Zeigen Sie, daß $\sum_i c_i = 0$ gilt.

Hinweis: Verwenden Sie $\mathbf{c}'\mathbf{1} = \mathbf{a}'\mathbf{C}\mathbf{1}$ und (5.24).
■

Im Fall der Varianzanalyse hat man

$$(\mathbf{X}'\mathbf{X})^{-1} = J^{-1}\mathbf{I}_I \tag{5.94}$$
$$\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} = J^{-1}\mathbf{c}'\mathbf{c} \tag{5.95}$$

und $\hat{\boldsymbol{\mu}} = [\bar{y}_{1+}, \bar{y}_{2+}, \ldots, \bar{y}_{I+}]'$.
Einsetzen ergibt die gesuchte explizite Form

$$|\mathbf{c}'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})| \leq \sqrt{(I-1)F(1-\alpha, I-1, IJ-I)}\sqrt{\frac{SQR}{J(IJ-I)}\mathbf{c}'\mathbf{c}}.$$

Wenn die Matrix $\mathbf{C}$ nur $d \leq I-1$ Zeilen hat, ergibt sich

$$|\mathbf{c}'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})| \leq \sqrt{d \cdot F(1-\alpha, d, IJ-I)}\sqrt{\frac{SQR}{J(IJ-I)}\mathbf{c}'\mathbf{c}}. \tag{5.96}$$

Bei nur *einem* Vergleich $d = 1$ erhält man $F(1-\alpha, 1, IJ - I) = t(1 - \alpha/2, IJ - I)^2$, also die Form der t-Intervalle.

Ein detaillierte Herleitung ist in Miller (1981, Kap. 2.2) enthalten.

Kapitel 6

Kategoriale Regression

6.1 Dichotome abhängige Variablen

6.1.1 Wahl der Response-Funktion

In den Kapiteln Regressionsanalyse und Varianzanalyse waren die abhängigen Variablen quantitativ (stetig, metrisch), während die unabhängigen Variablen sowohl quantitativ als auch kategorial sein konnten. Dies wurde durch Dummy- oder Effektkodierung umgesetzt.

Hier wird der Fall betrachtet, daß die abhängige Variable nur zwei (nominale) Ausprägungen hat, etwa die Präferenz für ein Produkt (ja/nein) oder die Variable `Kredit` = schlecht/gut (Bsp. 1.1). Man interessiert sich hier dafür, ob die die Rückzahlung eines Kredits von Merkmalen des Schuldners (Alter, Höhe des Kredits, Zahlungsmoral, Vorliegen eines Bürgen etc.) abhängt. Da die Ausprägung $y = 0, 1$ der abhängigen Variable nur nominal interpretiert werden kann, modelliert man die Wahrscheinlichkeit des Ereignisses $Y = y$ in Abhängigkeit von Regressorvariablen x, d.h. die bedingte Wahrscheinlichkeit $p(y|x) := P(Y = y|X = x)$. Aufgrund der Bayes-Formel gilt [1]

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}. \tag{6.1}$$

Daher ist die bedingte Wahrscheinlichkeit durch die Verteilung der Regressoren in den durch y gegebenen Gruppen, d.h. $p(x|y = 0), p(x|y = 1)$, die a priori-Wahrscheinlichkeiten der y-Kategorien $p(y = 0)$, $p(y = 1)$ und die Randverteilung der Regressoren $p(x)$ gegeben.

[1] $p(y|x) = P(Y = y|X = x), p(y) = P(Y = y)$. Für stetiges X gilt $p(x|y) = P(x \leq X \leq x + dx|Y = y)/dx$. p sind also Dichtefunktionen.

Man benötigt also im Beispiel die Altersverteilung in der Gruppen der guten/schlechten Schuldner (Abb. 1.1), die Wahrscheinlichkeiten, daß ein guter/schlechter Schuldner vorliegt, d.h. $p(\texttt{Kredit} = \textit{schlecht})$, $p(\texttt{Kredit} = \textit{gut})$ sowie die Altersverteilung

$$\begin{aligned} p(\texttt{Alter}) &= p(\texttt{Alter}|\textit{schlecht}) * p(\textit{schlecht}) \\ &+ p(\texttt{Alter}|\textit{gut}) * p(\textit{gut}). \end{aligned} \tag{6.2}$$

oder ganz allgemein

$$p(x) = \sum_y p(x|y)p(y) \tag{6.3}$$

(Satz von der totalen Wahrscheinlichkeit).

Nimmt man an, daß in den durch $y = 0, 1$ definierten Gruppen bedingte Normalverteilungen vorliegen (vgl. Abb. 1.1), d.h. $p(x|y = 0) = \phi(x; \mu_0, \sigma^2) = \phi_0$, $p(x|y = 1) = \phi(x; \mu_1, \sigma^2) = \phi_1$, und mit der Notation $p(y = 0) = p_0$, $p(y = 1) = p_1$, so ergibt sich

$$p(y = 1|x) = \frac{\phi_1 p_1}{\phi_1 p_1 + \phi_0 p_0} \tag{6.4}$$

$$= \frac{1}{1 + \frac{\phi_0}{\phi_1}\frac{p_0}{p_1}}. \tag{6.5}$$

Mit der expliziten Form $\phi(x; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2}(x - \mu)^2/\sigma^2)$ findet man $\frac{\phi_0}{\phi_1} = \exp\{[-(\mu_1 - \mu_0)x + \frac{1}{2}(\mu_1^2 - \mu_0^2)]/\sigma^2\}$, also eine Funktion der Form

$$p(y = 1|x) = \frac{1}{1 + c\exp(-\beta x)}. \tag{6.6}$$

Dies wird (für $c = 1$ und $\beta = 1$) als logistische Funktion bezeichnet. Bei ungleichen Varianzen $\sigma_0^2 \neq \sigma_1^2$ ergeben sich auch quadratische Terme in x (Singer, 2010).

Die eben durchgeführte einfache Ableitung motiviert die Benutzung der logistischen Funktion als Wahrscheinlichkeitsmodell für die dichotome Variable y. Definiert man die logistische Funktion

$$L(x) = \frac{1}{1+\exp(-x)} = \frac{\exp(x)}{1+\exp(x)}, \tag{6.7}$$

logistische Funktion

so kann man für die kategoriale Regression den Ansatz

$$p(y=1|\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})} = L(\mathbf{x}'\boldsymbol{\beta}) \tag{6.8}$$

Logit-Modell

machen. Hierbei ist

$$\eta := \mathbf{x}'\boldsymbol{\beta} = \beta_0 + x_1\beta_1 + ... + x_q\beta_q \tag{6.9}$$

ein linearer Prädiktor (analog zur multiplen Regression). Der konstante Term $\exp(-\beta_0) = c$ kann mit der Konstante in Glg. (6.6) identifiziert werden. Löst man nach $\mathbf{x}'\boldsymbol{\beta}$ auf, so findet man mit der Abkürzung $\pi := p(y=1|\mathbf{x})$

$$\mathbf{x}'\boldsymbol{\beta} = \log\left(\frac{\pi}{1-\pi}\right) = \text{logit}(\pi). \tag{6.10}$$

Logit-Funktion

Dies ist das logarithmierte Verhältnis der Wahrscheinlichkeiten (log odds ratio), $y = 1$ und $y = 0$ zu beobachten.[2] Es wird auch logarithmiertes relatives Risiko genannt. Das Verhältnis

$$\frac{\pi}{1-\pi} \tag{6.11}$$

odds ratio

wird als relatives Risiko (odds ratio) bezeichnet.

[2] $\log \equiv \ln$ bezeichnet hier den sog. *natürlichen Logarithmus*, d.h. wenn $y = \mathbf{e}^x$ mit der Eulerschen Zahl e=2.71828..., so gilt $x = \log y$ und umgekehrt $y = e^{\log y}$. Somit sucht man den Exponenten $x = \log y$, der wieder y ergibt.

Ganz allgemein kann man schreiben

Response-Funktion

$$p(y = 1|\mathbf{x}) \quad = \quad h(\mathbf{x}'\boldsymbol{\beta}) \tag{6.12}$$

Die Funktion $0 \leq h(\eta) \leq 1$ wird als Response (Antwort)-Funktion bezeichnet. Löst man nach $\eta = \mathbf{x}'\boldsymbol{\beta}$ auf, so gilt $\mathbf{x}'\boldsymbol{\beta} = g(\pi)$. Dies ist die sogenannte Link-Funktion.

Die Wahrscheinlichkeit $p(y = 1|\mathbf{x})$ kann als bedingter Erwartungswert der dichotomen Variable Y aufgefaßt werden, d.h.

$$E[Y|\mathbf{x}] = 1 \cdot p(y = 1|\mathbf{x}) + 0 \cdot p(y = 1|\mathbf{x}) = p(y = 1|\mathbf{x}) \tag{6.13}$$

Daher wird analog zum klassischen Regressionsmodell $Y = E[Y|\mathbf{x}] + \epsilon$ der bedingte Erwartungswert der Response-Variable modelliert.

Im einfachsten Fall kann man für h eine lineare Funktion wählen, d.h.

lineares Wahrscheinlichkeitsmodell

$$p(y = 1|\mathbf{x}) \quad = \quad \mathbf{x}'\boldsymbol{\beta} \tag{6.14}$$

Dies wird als lineares Wahrscheinlichkeitsmodell bezeichnet. Es hat den Nachteil, daß die Restriktionen $0 \leq p(y = 1|\mathbf{x}) \leq 1$ nicht unbedingt eingehalten werden. Ein stückweise lineares Modell ist durch die Gleichverteilung $U(x; [0, 1])$ gegeben. Dieses ist durch

$$U(x) = \begin{cases} 0, & x < 0 \\ x, & 0 < x < 1 \\ 1, & x > 1 \end{cases} \tag{6.15}$$

definiert.

Die Wahl $h = L$ führt auf das sog. Logit-Modell. Ein weitere gängige Möglichkeit ist die Wahl als kumulierte Verteilungsfunktion $F(x)$, etwa $\Phi(x)$ (Normalverteilung). Dies ist das sogenannte

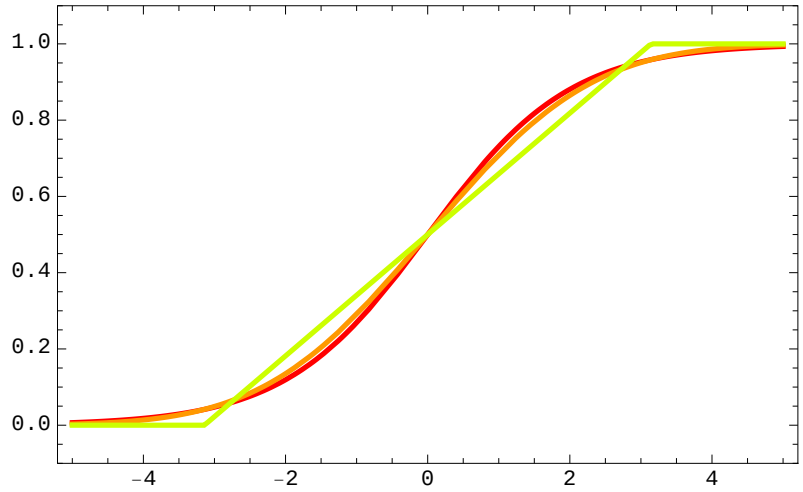


Abbildung 6.1: Responsefunktionen: Logistische (rot), Normalverteilung (orange), Gleichverteilung (grün). Die Varianzen wurden auf den Wert $\pi^2/3$ der logistischen Funktion adjustiert.

Probit-Modell

Probit-Modell

$$p(y = 1|\mathbf{x}) \quad = \quad \Phi(\mathbf{x}'\boldsymbol{\beta}). \tag{6.16}$$

Die unterschiedlichen Modell sind in Abb. 6.1 dargestellt. Zum besseren Vergleich wurden die Varianzen auf den Wert $\pi^2/3$ der logistischen Funktion adjustiert. Dies ist sinnvoll, da die Funktionen $h(\beta_0 + \beta_1 x) = \tilde{h}(\tilde{\beta}_0 + \tilde{\beta}_1 x)$ auf eine äquivalente Modellierung führen. Daher kann die Funktion verschoben und das Argument mit einem Faktor skaliert werden (vgl. Fahrmeir et al., 1996, S. 249). Die Unterschiede in den Funktionen sind recht gering, wobei die logistische Funktion im Gegensatz zur Normalverteilung leichter zu berechnen ist.

Generell muß die Response-Funktion zwischen 0 und 1 liegen, es ist nicht notwendig, daß es sich um eine kumulative Verteilungsfunktion handelt.

Man kann jedoch das binäre Regressions-Modell durch eine latente Variable $Y^* = \mathbf{x}'\boldsymbol{\beta}^* + \epsilon$ motivieren, die nicht direkt beobachtet werden

Variable	Beschreibung	Ausprägung	Punkte-bewertung	rel. Häufigk. in % bei	
				schlechten Krediten	guten Krediten
kredit	Dummy-Variable: 1 : Kredit wurde zurückgezahlt 0 : Kredit wurde nicht ordnungsgemäß zurückgezahlt				
laufkont	bestehendes lfd. Konto bei der Bank	kein Kontostand bzw. Debetsaldo	2	35.00	23.43
		0 <= ... < 200 DM	3	4.67	7.00
		... >= 200 DM oder Gehaltskonto seit mind. 1 Jahr	4	15.33	49.71
		kein lfd. Konto	1	45.00	19.86
moral	bisherige Zahlungsmoral	keine Kredite bisher / alle bisherigen Kredite zurückgezahlt	2	56.33	51.57
		frühere Kredite bei der Bank einwandfrei abgewickelt	4	16.67	34.71
		noch bestehende Kredite bei der Bank bisher einwandfrei	3	9.33	8.57
		früher zögernde Kreditführung	0	8.33	2.14
		kritisches Konto / es bestehen weitere Kredite nicht bei der Bank	1	9.33	3.00

Abbildung 6.2: Variablen im Datensatz `Kreditrisiko.LMU`.

Schwellwert-Modell

kann. Wenn sie einen Schwellwert τ übersteigt, ist $Y = 0$, ansonsten $Y = 1$. Da der latente Gleichungsfehler ϵ die Verteilung F aufweist, gilt $P(Y = 1) = P(Y^* < \tau) = P(\epsilon < -\mathbf{x}'\boldsymbol{\beta}^* + \tau) = F(-\mathbf{x}'\boldsymbol{\beta}^* + \tau) = h(\mathbf{x}'\boldsymbol{\beta})$. In diesem Fall ist also die Response-Funktion durch eine kumulative Verteilungsfunktion gegeben.

Beispiel 6.1 (Kreditrisiko)

In Bsp. 1.1 wurde bereits ein Datensatz diskutiert, bei dem die Rückzahlung eines Kredits in Abhängigkeit von Merkmalen des Kunden tabelliert ist (vgl. Abb. 6.2). [3]

Im einfachsten Fall kann man die relative Häufigkeit einer Rückzahlung in Abhängigkeit von nominalen Regressoren tabellieren, etwa der Variablen `Zahlungsmoral`. Dann erhält man eine Kreuztabelle bzw. ein Mosaik-Diagramm (Abb. 6.3). Faßt man die Variable `Zahlungsmoral` als Ordinalskala auf, so sinkt das Risiko eines Zahlungsausfalls (`Kredit = 0`, rot) mit höheren Werten.

Im Rahmen der kategorialen Regression kann man die Parameter eines logistischen Regressionsmodells schätzen. Im folgenden wird `Zahlungsmoral` als stetige Variable aufgefaßt. Die geschätzte Wahrscheinlichkeit

$$1 - \hat{\pi} = \hat{p}(y = 0|\mathbf{x}) = L(\mathbf{x}'\hat{\boldsymbol{\gamma}}) \tag{6.17}$$

[3] Datensätze `Kreditrisiko.LMU.sav` bzw. `Kreditrisiko.LMU.jmp`.
`http://www.stat.uni-muenchen.de/service/datenarchiv/kredit/kredit.html`

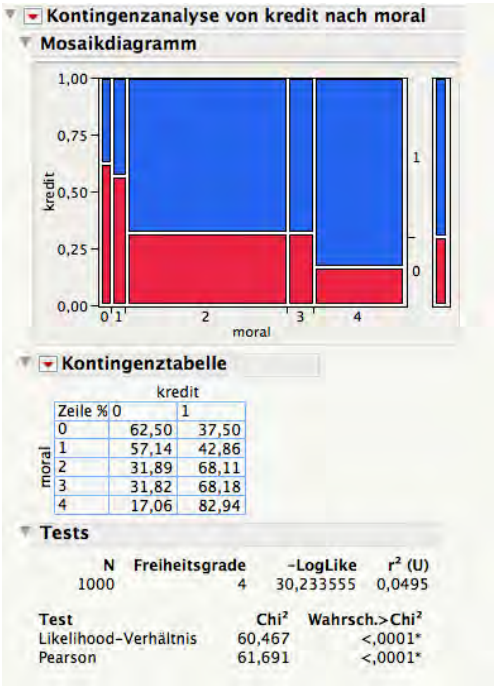


Abbildung 6.3: JMP: Mosaik-Plot der Variablen `Kredit` und `Zahlungsmoral`.

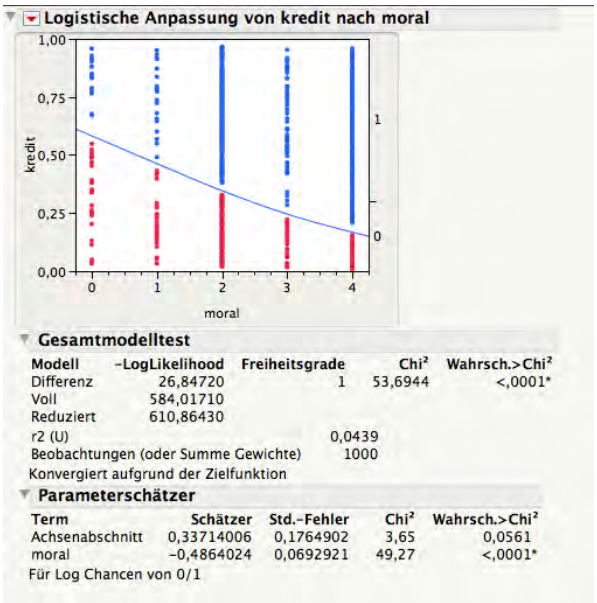


Abbildung 6.4: JMP: Logistische Regression der Variablen `Kredit` und `Zahlungsmoral`.

Variablen in der Gleichung

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1[a]	moral	,486	,069	49,279	1	,000	1,626
	Konstante	-,337	,176	3,650	1	,056	,714

a. In Schritt 1 eingegebene Variablen: moral.

Abbildung 6.5: SPSS: Logistische Regression der Variablen `Kredit` und `Zahlungsmoral`.

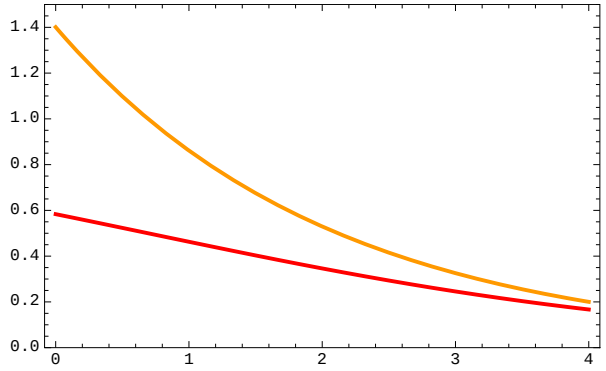


Abbildung 6.6: Response-Funktion $p(y = 0|\mathbf{x})$ (rot) und odds-ratio $p(y = 0|\mathbf{x})/p(y = 1|\mathbf{x})$ (orange) der Variablen `Kredit=0` und `Zahlungsmoral`.

ist als fallende Linie im Diagramm eingetragen (Abb. 6.4). Die Parameterschätzungen lauten $\hat{\boldsymbol{\gamma}} = [0.3371, -0.4864]'$.
Schätzt man das gleiche mit SPSS, so ergibt sich Abb. 6.5. Hier wird also

$$\hat{\pi} = \hat{p}(y = 1|\mathbf{x}) = L(\mathbf{x}'\hat{\boldsymbol{\beta}}) \tag{6.18}$$

berechnet ($\hat{\boldsymbol{\beta}} = -\hat{\boldsymbol{\gamma}}$). In den Log-odds (logarithmiertes relatives Risiko) ergibt sich der lineare Prädiktor $\mathbf{x}'\hat{\boldsymbol{\beta}}$, d.h.

$$\log \frac{\hat{\pi}}{1 - \hat{\pi}} = \text{logit}(\hat{\pi}) = \mathbf{x}'\hat{\boldsymbol{\beta}}. \tag{6.19}$$

Der Kehrwert ergibt $\log \frac{1-\hat{\pi}}{\hat{\pi}} = -\mathbf{x}'\hat{\boldsymbol{\beta}} = \mathbf{x}'\hat{\boldsymbol{\gamma}}$.
Die Parameter lassen sich also als Einflußgewichte der Variablen $\mathbf{x}$ auf die Log-odds interpretieren. Schreibt man

$$\frac{\pi}{1 - \pi} = e^{\mathbf{x}'\boldsymbol{\beta}} = e^{\beta_0} e^{\beta_1 x_1} ... e^{\beta_q x_q}, \tag{6.20}$$

so sieht man, daß die Regressoren (Kovariablen) in exponentiell-multiplikativer Form auf das Verhältnis der Wahrscheinlichkeiten wirken.

Im vorliegenden Fall gilt

$$\frac{\pi}{1-\pi} = e^{\beta_0} e^{\beta_1 x_1} \tag{6.21}$$

bzw. wenn man den Kreditausfall $p(y=0) = 1-\pi$ betrachtet

$$\frac{1-\pi}{\pi} = e^{-\beta_0} e^{-\beta_1 x_1} = e^{\gamma_0} e^{\gamma_1 x_1}. \tag{6.22}$$

Mit den Schätzern $\hat{\boldsymbol{\gamma}} = [0.3371, -0.4864]'$ sieht man, daß die Zahlungsmoral x_1 bei einer Änderung um 1 die odds-ratio um einen Faktor $\exp(-0.4864) = 0.614836$ verringert. In Abb. 6.6 ist die Wahrscheinlichkeit $1-\pi = p(y=0|\mathbf{x})$ und die odds-ratio $\frac{1-\pi}{\pi}$ über der Zahlungsmoral aufgetragen.
Die Modelle können in SPSS mit den Menü-Befehlen

```
Analysieren/Regression/Binär logistisch...
Analysieren/Regression/Multinomial logistisch...
Analysieren/Regression/Ordinal...
```

aufgerufen werden.

- Das binär logistische Modell läßt nur 2 Kategorien (0,1) für die abhängige Variable zu. Es wird $P(Y=1|\mathbf{x})$ modelliert.
- Beim multinomial logistischen Modell kann man auch abhängige Variablen mit mehr als 2 Kategorien analysieren und die Referenzkategorie explizit ausgewählen (siehe Abs. 6.2.1).
- Das ordinal logistische Modell nimmt an, dass die abhängige Variable ordinalskaliert ist (siehe Abs. 6.2.2).

■

Der Stoff wird in Aufgabe 6.1 vertieft.

6.1.2 Aufbau des Daten-Vektors

Der Vektor der unabhängigen Variablen $\mathbf{x}'_n = [x_{n1}, ..., x_{nq}], n = 1, ..., N$ kann, wie gesagt, metrische Variablen und kategoriale Größen enthalten. Folgende Fälle lassen sich unterscheiden ($j = 1, ..., q$, $n = 1, ..., N$ weglassen, $x = x_{nj}$):

- *Metrische* Variable: x ist der numerische Wert der Variablen.
- *Kategoriale* Variable: x hat $i = 1, .., I$ nominale Ausprägungen:

  1. Man benötigt $I - 1$ *Dummy-Variablen* x_i, um die I Ausprägungen zu kodieren.
     Die Ausprägung $x = i$ wird durch den 0-1-Vektor $\mathbf{x}' = [x_1, ..., x_{I-1}]$ mit

Dummy-Kodierung

$$x_i = \begin{cases} 1, & x = i \\ 0, & sonst \end{cases} \tag{6.23}$$

     kodiert, also $x_i(x) = \delta_{ix}$.[4]
     Beispielsweise hat man bei $I = 3$ Ausprägungen für x die Vektoren $\mathbf{x}'$

$$\begin{aligned} [1,0] &: \text{Kategorie 1 } (x = 1) \\ [0,1] &: \text{Kategorie 2 } (x = 2) \\ [0,0] &: \text{Kategorie 3 } (x = 3, \text{ Referenz-Kategorie}). \end{aligned}$$

  2. Man kann auch eine *Effektkodierung* $\mathbf{x}' = [x_1, ..., x_{I-1}]$

Effektkodierung

$$x_i = \begin{cases} 1, & x = i < I \\ -1, & x = I \\ 0, & sonst, \end{cases} \tag{6.24}$$

     vornehmen, im Beispiel

$$\begin{aligned} [1,0] &: \text{Kategorie 1 } (x = 1) \\ [0,1] &: \text{Kategorie 2 } (x = 2) \\ [-1,-1] &: \text{Kategorie 3 } (x = 3) \end{aligned}$$

[4] $\delta_{ix} = 1$ für $i = x$, 0 sonst, ist das Kronecker-Delta-Symbol. Siehe Glg. (11.3).

Auch *Interaktionen* zwischen Variablen lassen sich kodieren, etwa

$$x_3 = x_1 \cdot x_2 \tag{6.25}$$

$$= \begin{cases} x_2, & x_1 = 1 \\ 0, & x_1 = 0 \end{cases} \tag{6.26}$$

In Effektkodierung hat man für kategoriale Variablen $x_A = 1, ..., I$, $x_B = 1, ..., J$ den Interaktionsvektor $\mathbf{x}' = [x_{11}^{AB}, ..., x_{I-1,J-1}^{AB}]$ mit

Interaktionen

$$x_{ij}^{AB} = x_i^A x_j^B = \begin{cases} 1, & x_A = i, x_B = j \text{ oder } x_A = I, x_B = J \\ -1, & x_A = I, x_B = j \text{ oder } x_A = i, x_B = J \\ 0, & \textit{sonst.} \end{cases}$$

Die Effekt-Kodierung für die Ausprägungen $x_A = 1, ..., I$ kann man auch in Form einer Matrix (vgl. Glg. 6.24)

$$\mathbf{x}^\alpha = \begin{bmatrix} \mathbf{I}_{I-1} \\ -\mathbf{1}'_{I-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & \dots & -1 \end{bmatrix} : I \times (I-1) \tag{6.27}$$

schreiben. Die i-te Zeile der Matrix ergibt die Kodierung für die Ausprägung $x_A = i$.
Kombiniert man 2 Variablen $x_A = 1, ..., I$, $x_B = 1, ..., J$ additiv, so ergeben sich die Matrizen

$$\mathbf{x}^\alpha \otimes \mathbf{1}_J \;:\; IJ \times (I-1) \tag{6.28}$$

$$\mathbf{1}_I \otimes \mathbf{x}^\beta \;:\; IJ \times (J-1). \tag{6.29}$$

Die Interaktion zwischen den Variablen $x_A = 1, ..., I$, $x_B = 1, ..., J$ wird durch das Kronecker-Produkt

$$\mathbf{x}^\alpha \otimes \mathbf{x}^\beta : IJ \times (I-1)(J-1) \tag{6.30}$$

beschrieben. Diese Matrix ergibt aus dem Produkt aller $I-1$ Spalten von $\mathbf{x}^\alpha$ mit allen $J-1$ Spalten von $\mathbf{x}^\beta$. [5]

Analog ist die Dummy-Kodierung für die Ausprägungen $x_A = 1, ..., I$ in Form einer Matrix gegeben (vgl. Glg. 6.23)

$$\mathbf{x}^\alpha = \begin{bmatrix} \mathbf{I}_{I-1} \\ \mathbf{0}'_{I-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix} : I \times (I-1). \tag{6.31}$$

Gruppierte Daten

gruppierte Daten

Wenn ausschließlich qualitative Regressoren vorliegen, gibt es nur eine bestimmte Anzahl G von verschiedenen Ausprägungen der unabhängigen Variablen $\mathbf{x}_n$. Man spricht auch von gruppierten Daten. Dann kann anstatt der einzelnen Beobachtungen y_n für jede Gruppe $g = 1, ..., G$ ein Mittelwert $\bar{y}_g = p_g$ (relative Häufigkeit), eine Gruppengröße n_g sowie die unabhängige Variable $\mathbf{x}_g$ tabelliert werden, was die gleiche Information enthält. $p_g = f(Y = 1|\mathbf{x}_g)$ ist dann die relative Häufigkeit für die Ausprägung $Y = 1$ in Gruppe g. Siehe Bsp. 6.6.

Beispiel 6.2 (Datensatz für Effekt- und Dummy-Kodierung)

Nimmt man an, daß die kategorialen Variablen $x_A = 1, 2$ (z.B. Geschlecht) und $x_B = 1, 2, 3$ (z.B. Schulbildung) $I = 2$ und $J = 3$ Ausprägungen aufweisen, so hat ein Datensatz mit allen 6 Kombinationen in Effektkodierung folgende Form:

n	x_A	x_B	x_1^A	x_1^B	x_2^B	x_{11}^{AB}	x_{12}^{AB}
1	1	1	1	1	0	1	0
2	1	2	1	0	1	0	1
3	1	3	1	−1	−1	−1	−1
4	2	1	−1	1	0	−1	0
5	2	2	−1	0	1	0	−1
6	2	3	−1	−1	−1	1	1

(eine Beobachtung pro Kombination oder gruppierte Daten mit $G = 6$ Gruppen). In der ersten Spalte sind die 6 Beobachtungen fortlaufend

[5] Das Produkt der Spalten k, k' ist:
$(\mathbf{x}^\alpha \otimes \mathbf{1}_J)_{ij,k}(\mathbf{1}_I \otimes \mathbf{x}^\beta)_{ij,k'} = \mathbf{x}^\alpha_{ik} 1_j 1_i \mathbf{x}^\beta_{jk'} = \mathbf{x}^\alpha_{ik}\mathbf{x}^\beta_{jk'} = (\mathbf{x}^\alpha \otimes \mathbf{x}^\beta)_{ij,kk'}$.

numeriert. Spalte 2 und 3 enthält die Variablen x_A, x_B in der ursprünglichen nominalen Kodierung (1,2; 1,2,3). Spalte 4 zeigt die Effekt-Variable x_1^A, die bei Ausprägung $x_A = 1$ den Wert 1 und bei $x_A = 2$ den Wert –1 aufweist. Spalten 5 und 6 sind die Effekt-Variablen x_1^B und x_2^B mit $x_1^B(1) = 1$, $x_1^B(2) = 0$ und $x_1^B(3) = -1$ etc. Spalten 7 und 8 sind die Produkte der Spalten für x_1^A und x_1^B, x_2^B. Man hat nur $(I-1)(J-1) = 2$ Interaktionsvariablen. Die Effektparameter für die Spalten 4–8 sind $[\alpha_1, \beta_1, \beta_2, (\alpha\beta)_{11}, (\alpha\beta)_{12}]$. Die anderen Effekte ergeben sich aus den Restriktionen

$$\sum_i \alpha_i = 0 \tag{6.32}$$

$$\sum_j \beta_j = 0 \tag{6.33}$$

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0. \tag{6.34}$$

Der Wert des linearen Prädiktors $\mathbf{x}'\boldsymbol{\beta}$ ergibt sich beispielsweise in der Form ($x_A = 1, x_B = 2$, 2. Zeile)

$$\alpha_1 + \beta_2 + (\alpha\beta)_{12} \tag{6.35}$$

oder ($x_A = 1, x_B = 3$, 3. Zeile)

$$\alpha_1 - \beta_1 - \beta_2 - (\alpha\beta)_{11} - (\alpha\beta)_{12} = \alpha_1 + \beta_3 + (\alpha\beta)_{13}. \tag{6.36}$$

unter Benutzung der Restriktionen. Da die Effekte als Abstände von einem Mittelwert berechnet werden, wird normalerweise noch eine Spalte mit Einsen und ein Mittelwertsparameter μ hinzugefügt.

Die Matrixdarstellung der Effekte gibt

$$\mathbf{x}^\alpha = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \mathbf{x}^\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix},$$

$$\mathbf{x}^\alpha \otimes \mathbf{1}_3 \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \mathbf{1}_2 \otimes \mathbf{x}^\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \mathbf{x}^\alpha \otimes \mathbf{x}^\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}.$$

Zusammengefügt ist dies die oben beschriebene Form (Spalten 4–8).
Analog ergibt sich für die Dummy-Kodierung die Designmatrix

$$\left[\begin{array}{c|cc|c|cc|cc} n & x_A & x_B & x_1^A & x_1^B & x_2^B & x_{11}^{AB} & x_{12}^{AB} \\ \hline 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 & 0 & 1 & 0 & 1 \\ 3 & 1 & 3 & 1 & 0 & 0 & 0 & 0 \\ 4 & 2 & 1 & 0 & 1 & 0 & 0 & 0 \\ 5 & 2 & 2 & 0 & 0 & 1 & 0 & 0 \\ 6 & 2 & 3 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Dies läßt sich mit den Matrizen (vgl. 6.31)

$$\mathbf{x}^\alpha = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{x}^\beta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

leicht nachrechnen.

■

Übung 6.1
Berechnen Sie bitte die Designmatrix in Dummykodierung explizit.

■

Der Stoff wird in Aufgabe 6.2 vertieft.

6.1.3 Interpretation der Parameter

Die Interpretation der Parameter ist also schwieriger als im linearen Wahrscheinlichkeitsmodell. Hier wäre $\boldsymbol{\beta}$ direkt als Gewicht der Veränderung der Response-Wahrscheinlichkeit $p(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$ interpretierbar.
Man kann auch eine 2-stufige Interpretation vornehmen:

1. $\eta = \mathbf{x}'\boldsymbol{\beta}$:

   linearer Prädiktor: Effekte wie im linearen Modell.

2. $\pi = h(\eta)$:

   Transformation durch die Response-Funktion:

   nichtlineare Effekte für π.

   Wenn h eine Verteilungsfunktion ist, kann η als π-Quantil aufgefaßt werden, also $\eta = h^{-1}(\pi)$.

2-stufige Interpretation

Im Fall der logistischen Funktion ergibt sich für die odds ratio

$$\frac{\pi}{1-\pi} = e^{\mathbf{x}'\boldsymbol{\beta}} = e^{\beta_0} e^{\beta_1 x_1} \ldots e^{\beta_q x_q} \tag{6.37}$$

exponentiell-multiplikative Form

eine relativ einfache Interpretation (exponentiell-multiplikative Form).

Beispiel 6.3 (Kreditrisiko, Zahlungsmoral als nominale Variable)

In Beispiel 6.1 wurde die Variable `moral` als stetige Variable aufgefaßt und als Prädiktor $\beta_0 + \beta_1 * \texttt{moral}$ in die logistische Funktion eingesetzt. Man kann sich auch auf den Standpunkt stellen, daß es sich um eine nominale Größe mit den Ausprägungen 0,...,4 handelt (d.h. $I = 5$). Man benötigt also $I - 1 = 4$ Dummy- oder Effektvariablen `moral[0]`,...,`moral[3]`, die von den Programmen generiert werden. Man muß also die Designmatrix nicht selbst erzeugen. Der Output für JMP und SPSS ist in Abb. 6.7 gegeben. Hierbei ist zu beachten, daß die Programme unterschiedliche Kodierungen benutzen, nämlich die Effekt-Kodierung (JMP) und die Dummy-Kodierung (SPSS). Außerdem wird $p(y = 0|\mathbf{x})$ in JMP und $p(y = 1|\mathbf{x})$ in SPSS geschätzt. Mit Hilfe der Design-Matrizen (erste Spalte = Konstante)

$$\mathbf{X}_{\text{effekt}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{bmatrix} \tag{6.38}$$

Parameterschätzer

Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	-0,460704	0,1049155	19,28	<,0001*
moral[0]	0,97152965	0,2738745	12,58	0,0004*
moral[1]	0,7483861	0,2469965	9,18	0,0024*
moral[2]	-0,2982752	0,1273562	5,49	0,0192*
moral[3]	-0,301436	0,2059996	2,14	0,1434

Für Log Chancen von 0/1

Variablen in der Gleichung

		Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)
Schritt 1[a]	moral			55,809	4	,000	
	moral(1)	-2,092	,362	33,459	1	,000	,123
	moral(2)	-1,869	,328	32,500	1	,000	,154
	moral(3)	-,822	,181	20,602	1	,000	,440
	moral(4)	-,819	,277	8,766	1	,003	,441
	Konstante	1,581	,155	103,656	1	,000	4,860

a. In Schritt 1 eingegebene Variablen: moral.

Abbildung 6.7: Nominale Modellierung von `moral`. Parameterschätzungen mit JMP (oben) und SPSS (unten).

(Effekt-Kodierung) und

$$\mathbf{X}_{\text{dummy}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6.39}$$

(Dummy-Kodierung) können die Prädiktoren $\boldsymbol{\eta}_{\text{JMP}} = \mathbf{X}_{\text{effekt}}\boldsymbol{\gamma}$ und $\boldsymbol{\eta}_{\text{SPSS}} = \mathbf{X}_{\text{dummy}}\boldsymbol{\beta}$ berechnet werden. Die 5 Zeilen der Design-Matrizen entsprechen den Ausprägungen 0–4 der nominalen Variable `moral`.

Aus den Tabellen entnimmt man die Parameterschätzungen. Multiplikation mit der Designmatrix ergibt

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= [-0.460704, 0.97153, 0.748386, -0.298275, -0.301436]' \\ \hat{\boldsymbol{\eta}}_{\text{JMP}} &= [0.510826, 0.287682, -0.758979, -0.76214, -1.58091]' \\ \hat{\boldsymbol{\beta}} &= [1.581, -2.092, -1.869, -0.822, -0.819]' \\ \hat{\boldsymbol{\eta}}_{\text{SPSS}} &= [-0.511, -0.288, 0.759, 0.762, 1.581]' \end{aligned}$$

Man erhält also tatsächlich (bis auf das Vorzeichen und Rundungsfehler) die gleichen Werte.

Die geschätzten Responsefunktionen $L(\hat{\boldsymbol{\eta}})$ sind (JMP):

$$[0.625, 0.571429, 0.318868, 0.318182, 0.170667]'$$

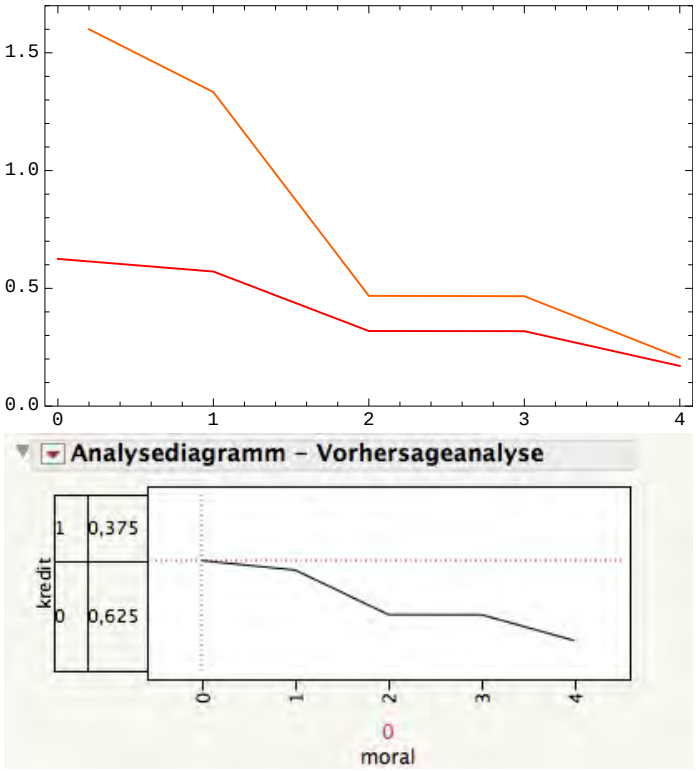


Abbildung 6.8: Nominale Modellierung von `moral`. Oben: Response-Funktion $p(y = 0|\mathbf{x})$ (rot) und odds ratio $p(y = 0|\mathbf{x})/p(y = 1|\mathbf{x})$ (orange) der Variablen `Kredit=0` und `Zahlungsmoral` = 0,...,4. Unten: von JMP berechnete Response-Funktion.

und in SPSS

$$[0.374959, 0.428494, 0.681137, 0.681788, 0.829346]'.$$

Dies entspricht aber gerade den Werten aus der Kontingenztabelle (Kreuztabelle) 6.3.
Die odds $\exp(\hat{\boldsymbol{\eta}})$ lauten:

$$[1.66667, 1.33333, 0.468144, 0.466667, 0.205788]'$$

und in SPSS

$$[0.599895, 0.749762, 2.13614, 2.14256, 4.85981]'.$$

Ein graphische Darstellung ist in Abb. 6.8 zu sehen.

Chancenverhältnisse

Für Chancen der kredit von 0 im Vergleich zu 1.

Chancen Verhältnisse für moral

Stufe1	/Stufe2	Verhältnis Chancen	Reziprok
1	0	0,8	1,25
2	0	0,2808864	3,5601578
2	1	0,351108	2,8481262
3	0	0,28	3,5714285
3	1	0,35	2,8571428
3	2	0,9968442	1,0031658
4	0	0,1234728	8,098948
4	1	0,154341	6,4791584
4	2	0,4395827	2,2748846
4	3	0,4409744	2,2677055

Abbildung 6.9: Quotienten der odds ratios.

Parameterschätzer

Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	0,51082562	0,3265986	2,45	0,1178
moral[1-0]	-0,2231436	0,4358899	0,26	0,6087
moral[2-1]	-1,0466613	0,3033489	11,90	0,0006*
moral[3-2]	-0,0031608	0,2471198	0,00	0,9898
moral[4-3]	-0,8188984	0,2765794	8,77	0,0031*

Für Log Chancen von 0/1

Abbildung 6.10: Schätzung mit ordinaler Kodierung von Zahlungsmoral.

Die Verhältnisse der odds ratios für die einzelnen Abstufungen von `Zahlungsmoral` sind in Abb. 6.9 gezeigt. Man kann diese Werte aus den odds [1.66667, 1.33333, 0.468144, 0.466667, 0.205788] leicht nachrechnen, etwa odds(1)/odds(0) = 1.33333/1.66667 = 0.8.

Auch andere Kodierungen sind möglich, etwa für ordinale Variablen. Hier lautet die Design-Matrix (1. Spalte = Konstante)

$$\mathbf{X}_{\text{ord}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{6.40}$$

Die Schätzungen des Parametervektors $\boldsymbol{\gamma}_{\text{ord}}$ sind in Abb. 6.10 zu sehen. Man erhält wieder die gleichen Prädiktoren $\hat{\boldsymbol{\eta}} = \mathbf{X}_{\text{ord}}\hat{\boldsymbol{\gamma}}_{\text{ord}}$. Bitte nachrechnen!

■

Beispiel 6.4 (Interaktion `Zahlungsmoral * laufendes Konto`)
Wird zur Variable 'Zahlungsmoral' = `moral` noch die Größe 'laufendes Konto' = `laufkont` hinzugefügt (vgl. Abb. 6.2), so ergeben sich ohne Interaktionseffekte die Resultate in Abb. 6.11. Obwohl keine Interaktionseffekte modelliert wurden, sind die Linien nicht parallel. Dies ist eine Folge der nichtlinearen Transformation $L(\mathbf{x}'\boldsymbol{\beta})$. Eine graphische Darstellung des linearen Prädiktors $\hat{\eta} = \mathbf{x}'\hat{\boldsymbol{\beta}}$ ist in Abb. 6.12 gezeigt.

Nimmt man Interaktionseffekte hinzu, so ergibt sich Abb. 6.13. Man erhält das Resultat, daß das Vorliegen eines Gehaltskontos die Wahrscheinlichkeit eines Zahlungsausfalls verringert. Es ergibt sich auch ein signifikanter Wechselwirkungseffekt zwischen Zahlungsmoral und Kontovariable, nämlich der Term `moral[3]*laufkont[2]`. Allerdings ist der gesamte Interaktionseffekt `moral*laufkont` nicht signifikant (Abb. 6.14–6.15).

Die Diagramme in den Abbildungen lassen sich aus den linearen Prädiktoren $\mathbf{X}\hat{\boldsymbol{\beta}}$ konstruieren, wobei die Designmatrizen in Effektkodierung die Form $\mathbf{X} = [\mathbf{1}_I \otimes \mathbf{1}_J, \mathbf{x}^\alpha \otimes \mathbf{1}_J, \mathbf{1}_I \otimes \mathbf{x}^\beta]$ und $\mathbf{X} = [\mathbf{1}_I \otimes \mathbf{1}_J, \mathbf{x}^\alpha \otimes \mathbf{1}_J, \mathbf{1}_I \otimes \mathbf{x}^\beta, \mathbf{x}^\alpha \otimes \mathbf{x}^\beta]$ (mit Interaktion) aufweisen. Hierbei sind

$$\mathbf{x}^\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 \end{bmatrix}, \tag{6.41}$$

$$\mathbf{x}^\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \tag{6.42}$$

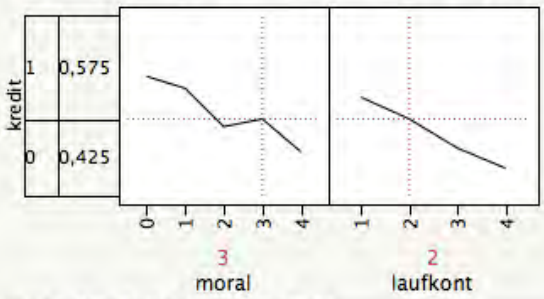
die Designmatrizen für die Effektkodierung der Variablen mit $I = 5, J =$

Parameterschätzer

Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	-0,5861399	0,1276783	21,08	<,0001*
moral[0]	0,84053661	0,2892649	8,44	0,0037*
moral[1]	0,55399236	0,2622362	4,46	0,0346*
moral[2]	-0,3066242	0,1351095	5,15	0,0232*
moral[3]	-0,1297255	0,2198736	0,35	0,5552
laufkont[1]	0,89561749	0,1317872	46,18	<,0001*
laufkont[2]	0,41447682	0,1341294	9,55	0,0020*
laufkont[3]	-0,309307	0,2393093	1,67	0,1962

Für Log Chancen von 0/1

Analysediagramm – Vorhersageanalyse



Wechselwirkungsanalyse

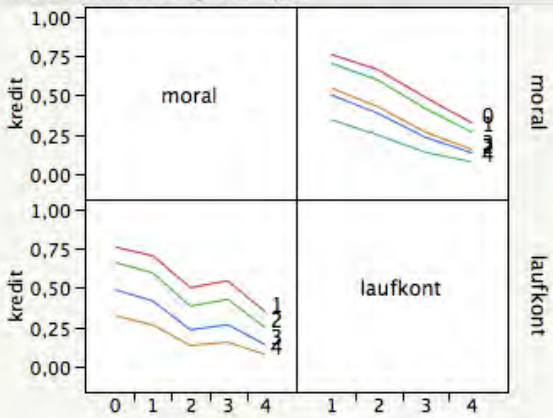


Abbildung 6.11: Zahlungsmoral und laufendes Konto ohne Interaktion.

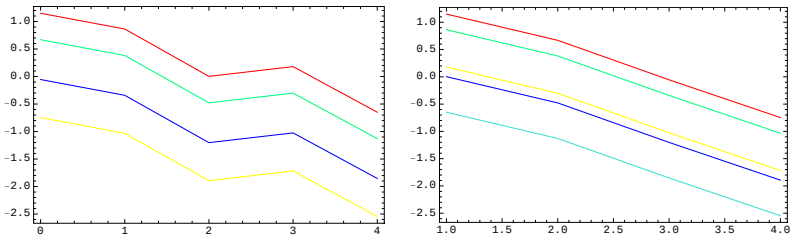


Abbildung 6.12: Zahlungsmoral und laufendes Konto ohne Interaktion: linearer Prädiktor η.

Parameterschätzer

Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	-0,5427105	0,1511559	12,89	0,0003*
moral[0]	0,78791785	0,3869407	4,15	0,0417*
moral[1]	0,40280659	0,3495936	1,33	0,2492
moral[2]	-0,4110978	0,1825697	5,07	0,0243*
moral[3]	0,08634101	0,326854	0,07	0,7917
laufkont[1]	1,00910578	0,21681	21,66	<,0001*
laufkont[2]	0,31703235	0,1965258	2,60	0,1067
laufkont[3]	-0,2761584	0,364418	0,57	0,4486
moral[0]*laufkont[1]	-0,0503403	0,5512546	0,01	0,9272
moral[0]*laufkont[2]	0,13090752	0,4904053	0,07	0,7895
moral[0]*laufkont[3]	0,03095107	0,9271869	0,00	0,9734
moral[1]*laufkont[1]	0,11162742	0,4638148	0,06	0,8098
moral[1]*laufkont[2]	0,22833671	0,4704501	0,24	0,6274
moral[1]*laufkont[3]	-0,2770849	0,8259347	0,11	0,7373
moral[2]*laufkont[1]	-0,005287	0,2549409	0,00	0,9835
moral[2]*laufkont[2]	0,19119087	0,2402912	0,63	0,4262
moral[2]*laufkont[3]	-0,2253205	0,4428783	0,26	0,6109
moral[3]*laufkont[1]	0,54587604	0,5141258	1,13	0,2883
moral[3]*laufkont[2]	-0,9215348	0,4092427	5,07	0,0243*
moral[3]*laufkont[3]	0,03938073	0,8165696	0,00	0,9615

Für Log Chancen von 0/1

Analysediagramm – Vorhersageanalyse

Wechselwirkungsanalyse

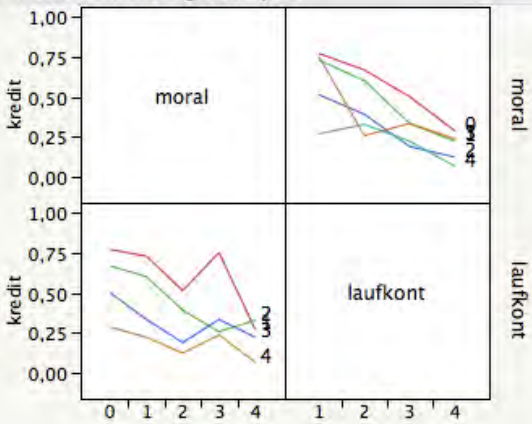


Abbildung 6.13: Zahlungsmoral und laufendes Konto mit Interaktion.

Effekt-Wald-Tests

Quelle	Anzahl Parameter	Freiheitsgrade	Wald-Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	35,1442042	<,0001*
laufkont	3	3	93,3801282	<,0001*

Effekt-Likelihood-Verhältnistests

Quelle	Anzahl Parameter	Freiheitsgrade	L-R Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	37,2469344	<,0001*
laufkont	3	3	108,115748	<,0001*

Abbildung 6.14: Zahlungsmoral und laufendes Konto ohne Interaktion.

Effekt-Wald-Tests

Quelle	Anzahl Parameter	Freiheitsgrade	Wald-Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	18,1771594	0,0011*
laufkont	3	3	38,752917	<,0001*
moral*laufkont	12	12	13,5497991	0,3304

Effekt-Likelihood-Verhältnistests

Quelle	Anzahl Parameter	Freiheitsgrade	L-R Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	16,3163905	0,0026*
laufkont	3	3	44,7861983	<,0001*
moral*laufkont	12	12	14,0350419	0,2985

Abbildung 6.15: Zahlungsmoral und laufendes Konto mit Interaktion.

4. Explizit ergibt sich

$$\mathbf{X} = \left[\begin{array}{c|cccc|ccc}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & -1 & -1 & -1 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 \\
1 & -1 & -1 & -1 & -1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 & -1 & 0 & 0 & 1 \\
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1
\end{array}\right] \tag{6.43}$$

für das Modell ohne Interaktion. Nimmt man eine Interaktion hinzu, ergeben sich $(I-1)(J-1) = 12$ weitere Spalten, explizit

$$\left[\begin{array}{c|cccc|ccc|cccccccccccc}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\
1 & -1 & -1 & -1 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\
1 & -1 & -1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 \\
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}\right]$$

Multipliziert man die Designmatrizen mit den Schätzungen, so ergeben sich die Prädiktoren η und daraus die Wahrscheinlichkeiten $p(y = 0|\mathbf{x}) = L(\eta)$. Diese sind in den Abbildungen 6.16 und 6.17 dargestellt. Eine grafische Darstellung ist den Abb. 6.11– 6.13 zu entnehmen.

■

	moral	laufkont	Wahrscheinlichkeit (kredit=0)	Wahrscheinlichkeit (kredit=1)
1	0	1	0,75951352	0,24048648
2	0	2	0,66125089	0,33874911
3	0	3	0,48627588	0,51372412
4	0	4	0,3216083	0,6783917
5	1	1	0,70338512	0,29661488
6	1	2	0,59443478	0,40556522
7	1	3	0,4154562	0,5845438
8	1	4	0,26251553	0,73748447
9	2	1	0,50071337	0,49928663
10	2	2	0,38265666	0,61734334
11	2	3	0,23110699	0,76889301
12	2	4	0,13084008	0,86915992
13	3	1	0,54481741	0,45518259
14	3	2	0,42521807	0,57478193
15	3	3	0,2640211	0,7359789
16	3	4	0,15230282	0,84769718
17	4	1	0,34328218	0,65671782
18	4	2	0,24419021	0,75580979
19	4	3	0,13544771	0,86455229
20	4	4	0,07275593	0,92724407

Abbildung 6.16: Prädiktion: Zahlungsmoral und laufendes Konto ohne Interaktion.

	moral	laufkont	Wahrscheinlichkeit (kredit=0)	Wahrscheinlichkeit (kredit=1)
1	0	1	0,76923077	0,23076923
2	0	2	0,66666667	0,33333333
3	0	3	0,5	0,5
4	0	4	0,28571429	0,71428571
5	1	1	0,72727273	0,27272727
6	1	2	0,6	0,4
7	1	3	0,33333333	0,66666667
8	1	4	0,22222222	0,77777778
9	2	1	0,5125	0,4875
10	2	2	0,39041096	0,60958904
11	2	3	0,18918919	0,81081081
12	2	4	0,12299474	0,87700526
13	3	1	0,75	0,25
14	3	2	0,25714286	0,74285714
15	3	3	0,33333333	0,66666667
16	3	4	0,23684211	0,76315789
17	4	1	0,26865672	0,73134328
18	4	2	0,32727273	0,67272727
19	4	3	0,22222222	0,77777778
20	4	4	0,06537808	0,93462192

Abbildung 6.17: Prädiktion: Zahlungsmoral und laufendes Konto mit Interaktion.

Beispiel 6.5 (Zahlungsmoral, Alter, Kredithöhe)
Hier noch ein gemischtes Modell, bei dem die stetigen Variablen Alter und Kredithöhe hinzugenommen wurden. Abb. 6.18 zeigt die Schätzungen der Parameter und Vorhersagediagramme der geschätzten Wahrscheinlichkeit $p(y = 0|\mathbf{x})$. Mit steigendem Alter sinkt die Wahrscheinlichkeit eines Zahlungsausfalls, während sie mit der Kredithöhe zunimmt. Die Schätzungen für die Effekte `moral[0]`,...,`moral[3]` weichen etwas von der vorigen Analyse ab, da ja weitere Variablen im Modell enthalten sind. Die Graphiken im Fenster Wechselwirkungsanalyse zeigen die Responsefunktion bei Konstanthaltung der anderen Variablen. Bei den stetigen Größen werden nur die Minimal- und Maximalwerte benutzt (etwa 19 Jahre und 75 Jahre). Alle Variablen sind signifikant. Nimmt man auch die Interaktionseffekte hinzu, ergeben sich komplexere Resultate. Zunächst ist nur noch der Haupteffekt `moral` signifikant ($p < 0.05$), aber auch die Interaktionen `moral*hoehe` und `moral*alter`. Die Diagramme zeigen deutliche Schereneffekte. Etwa ist bei den Ausprägungen `moral` $= 0, 1, 3$ eine Zunahme der Ausfallwahrscheinlichkeit mit dem Alter, bei `moral` $= 2, 4$ eine Abnahme mit dem Alter zu beobachten (Diagramm rechts oben). Betrachtet man die Abhängigkeit des Kreditausfallrisikos von der Kredithöhe, so ist bei 75-jährigen ein Abfall, bei 19-jährigen eine Zunahme zu beobachten. Diese Zusammenhänge sind jedoch im Diagramm nur für die Abstufung `moral` $= 0$ gezeigt. Eine andere Situation ergibt sich bei `moral` $= 4$ (Abb. 6.20). Hier bleibt die Wahrscheinlichkeit bei 75-jährigen fast konstant, während sie bei 19-jährigen stark ansteigt. Aufgrund der 2- und 3-fach-Interaktionen sind die Verhältnisse sehr verwickelt.

■

6.1.4 Schätzung der Parameter

Bisher wurde nur die Responsefunktion $h(\eta)$ und der lineare Prädiktor $\eta = \mathbf{x}'\boldsymbol{\beta}$ spezifiziert. Wie die unbekannten Parameter zu schätzen sind, blieb offen. Beim klassischen Regressionsmodell wird meistens die Methode der kleinsten Quadrate (KQ) betrachtet, also die Minimierung der quadrierten Differenz zwischen Daten und Prädiktor, d.h. $S(\boldsymbol{\beta}) = \sum_n (y_n - E[Y_n|\mathbf{x}_n, \boldsymbol{\beta}])^2$. Der Vektor mit dem minimalen Wert $\min_{\boldsymbol{\beta}} S(\boldsymbol{\beta})$ ist der KQ-Schätzer $\hat{\boldsymbol{\beta}}$. Im Fall der kategorialen Regression gilt

$$E[Y_n|\mathbf{x}_n, \boldsymbol{\beta}] = p(y_n = 1|\mathbf{x}_n, \boldsymbol{\beta}) := \pi_n(\boldsymbol{\beta}). \tag{6.44}$$

Man könnte daher das Minimum von $S(\boldsymbol{\beta}) = \sum_n (y_n - \pi_n(\boldsymbol{\beta}))^2$ suchen.

Parameterschätzer

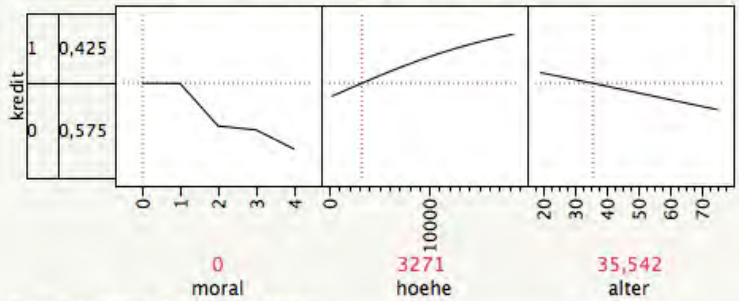
Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	-0,288875	0,2727496	1,12	0,2895
moral[0]	0,82700153	0,2800412	8,72	0,0031*
moral[1]	0,82008147	0,2503053	10,73	0,0011*
moral[2]	-0,2535063	0,1307309	3,76	0,0525
moral[3]	-0,3647237	0,2104846	3,00	0,0831
hoehe	0,00010684	0,000025	18,27	<,0001*
alter	-0,0164522	0,0068109	5,84	0,0157*

Für Log Chancen von 0/1

Effekt-Likelihood-Verhältnistests

Quelle	Anzahl Parameter	Freiheitsgrade	L-R Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	51,1357617	<,0001*
hoehe	1	1	18,4013443	<,0001*
alter	1	1	6,045636	0,0139*

Analysediagramm - Vorhersageanalyse



Wechselwirkungsanalyse

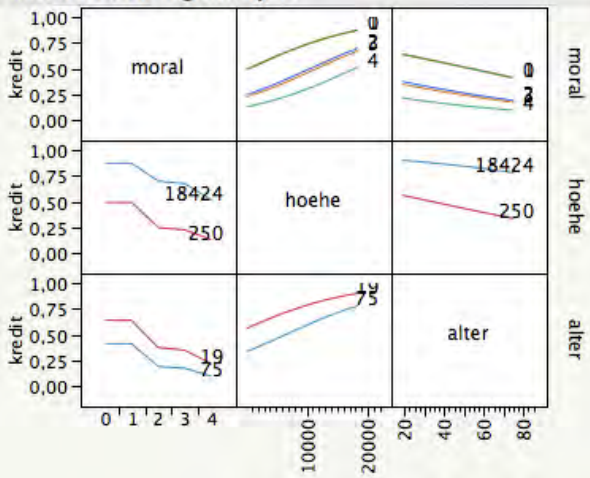


Abbildung 6.18: Schätzung mit 3 unabhängigen Variablen (Zahlungsmoral, Alter und Kredithöhe).

Parameterschätzer

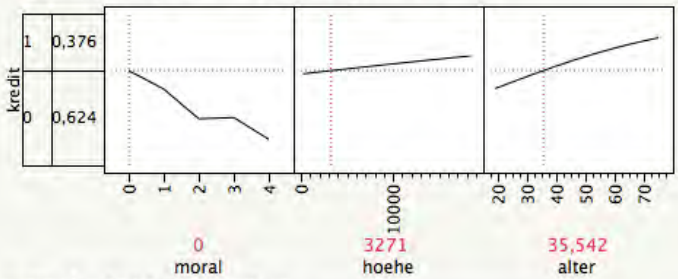
Term	Schätzer	Std.-Fehler	Chi²	Wahrsch.>Chi²
Achsenabschnitt	-0,7198606	0,4589439	2,46	0,1168
moral[0]	1,03569394	0,3162138	10,73	0,0011*
moral[1]	0,54121329	0,2827724	3,66	0,0556
moral[2]	-0,2851129	0,1420317	4,03	0,0447*
moral[3]	-0,2475158	0,227029	1,19	0,2756
hoehe	2,08785e-5	3,9118e-5	0,28	0,5935
moral[0]*(hoehe-3271,25)	8,27429e-6	7,4787e-5	0,01	0,9119
moral[1]*(hoehe-3271,25)	-0,000197	0,0001174	2,81	0,0935
moral[2]*(hoehe-3271,25)	0,00013359	4,7921e-5	7,77	0,0053*
moral[3]*(hoehe-3271,25)	-7,6727e-5	7,301e-5	1,10	0,2933
alter	0,00340882	0,0126927	0,07	0,7883
moral[0]*(alter-35,542)	0,02581243	0,0372166	0,48	0,4879
moral[1]*(alter-35,542)	0,01223758	0,028612	0,18	0,6689
moral[2]*(alter-35,542)	-0,0279132	0,014661	3,62	0,0569
moral[3]*(alter-35,542)	0,0377132	0,0217193	3,02	0,0825
(hoehe-3271,25)*(alter-35,542)	2,99766e-6	3,6567e-6	0,67	0,4123
moral[0]*(hoehe-3271,25)*(alter-35,542)	-5,1842e-6	6,8281e-6	0,58	0,4477
moral[1]*(hoehe-3271,25)*(alter-35,542)	1,83246e-5	0,0000108	2,88	0,0895
moral[2]*(hoehe-3271,25)*(alter-35,542)	1,85853e-6	4,4101e-6	0,18	0,6734
moral[3]*(hoehe-3271,25)*(alter-35,542)	-6,8932e-6	6,9757e-6	0,98	0,3231

Für Log Chancen von 0/1

Effekt-Likelihood-Verhältnistests

Quelle	Anzahl Parameter	Freiheitsgrade	L-R Chi-Quadrat	Wahrsch.>Chi²
moral	4	4	42,072309	<,0001*
hoehe	1	1	0,27876863	0,5975
moral*hoehe	4	4	12,1522099	0,0163*
alter	1	1	0,07302429	0,7870
moral*alter	4	4	12,41961	0,0145*
hoehe*alter	1	1	0,70461662	0,4012
moral*hoehe*alter	4	4	6,2016217	0,1846

Analysediagramm – Vorhersageanalyse



Wechselwirkungsanalyse

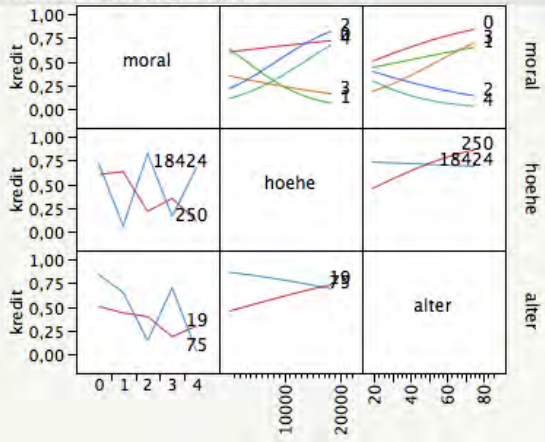


Abbildung 6.19: Analyse mit Interaktionseffekten (Zahlungsmoral = 0, Kredithöhe und Alter Durchschnittswerte).

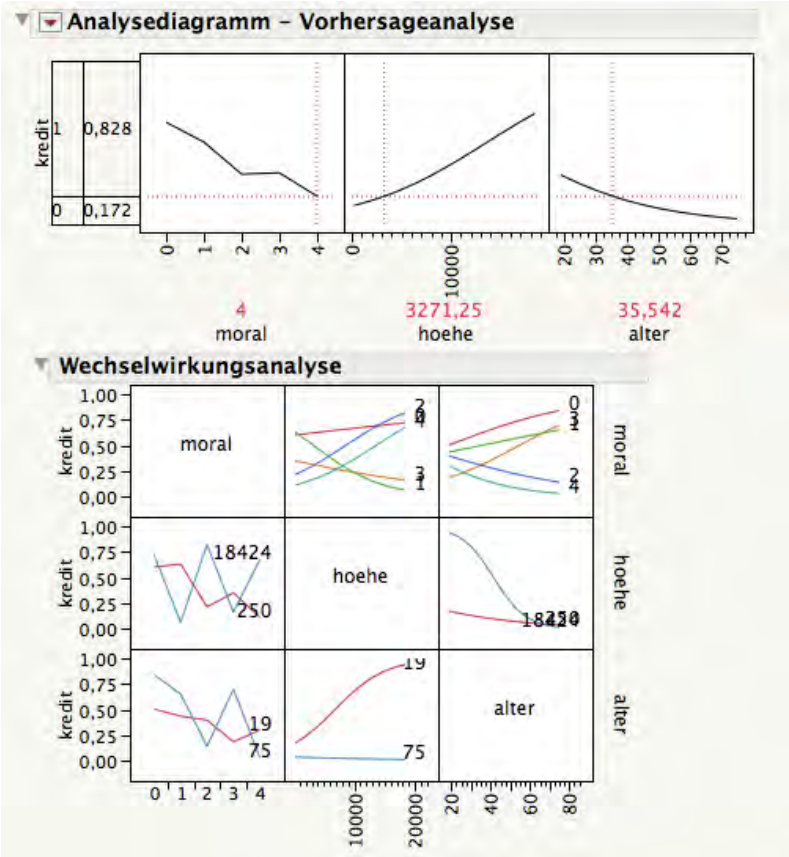


Abbildung 6.20: Analyse mit Interaktionseffekten (Zahlungsmoral = 4, Kredithöhe und Alter Durchschnittswerte).

Im Fall der kategorialen Regression wird jedoch meistens die Maximum-Likelihood-Methode angewandt. Hierbei wird die Wahrscheinlichkeit der Daten $p(y_1, ..., y_N|\boldsymbol{\beta}) := L(\boldsymbol{\beta})$ als Funktion des Parametervektors $\boldsymbol{\beta}$ betrachtet und maximiert. Der Wert, der die maximale Likelihood ergibt, wird als Maximum-Likelihood-Schätzer (ML) bezeichnet, also

Maximum-Likelihood(ML)-Schätzer

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}), \tag{6.45}$$

vgl. Mardia et al. (1979, Kap. 4), Fahrmeir et al. (1996, Kap. 3.2.3, 6.2.1) oder Fahrmeir et al. (2007, Kap. 9). Die Likelihood lautet hier konkret

$$L(\boldsymbol{\beta}) = \prod_n \pi_n^{y_n}(1-\pi_n)^{1-y_n}, \tag{6.46}$$

wobei $y_n = 0, 1$ und $\pi_n = h(\mathbf{x}_n'\boldsymbol{\beta})$ die Responsefunktion ist. Es handelt sich also um das Produkt der Wahrscheinlichkeiten der unabhängigen Messungen $y_1, ..., y_N$, da $\pi^0 = 1, \pi^1 = \pi$ gilt.

Meistens wird die Log-Likelihood $l = \log L$ betrachtet, die eine einfachere Struktur hat. Da die Logarithmus-Funktion monoton ist, erhält man die gleichen Schätzer. Es gilt

$$l(\boldsymbol{\beta}) = \sum_n y_n \log(\pi_n) + (1-y_n)\log(1-\pi_n). \tag{6.47}$$

Der Stoff wird in Aufgabe 6.3 vertieft.

Bei gruppierten Daten (nur G verschiedene Werte von $\mathbf{x}_n$) ist $\pi_n = h(\mathbf{x}_n'\boldsymbol{\beta})$ gleich für alle Beobachtungen aus Gruppe g. Schreibt man für diesen Wert der Einfachheit halber π_g, so läßt sich mit der Notation $\bar{y}_g = p_g = n_g^{-1}\sum_{n\in N_g} y_n$, n_g = Größe der Gruppe g, N_g = Indexmenge der Beobachtungen in Gruppe g, die Likelihood in gruppierter Form

$$l(\boldsymbol{\beta}) = \sum_g n_g p_g \log(\pi_g) + n_g(1-p_g)\log(1-\pi_g). \tag{6.48}$$

schreiben. Die Speicherung ist hier wesentlich effizienter.

Am Maximum ist die Steigung (Gradient) der Log-Likelihood gleich Null. Man kann also anstatt des Maximums die Nullstelle der Vektor-Funktion

Score-Funktion

$$\mathbf{s}(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \tag{6.49}$$

suchen. Sie wird auch als Score-Funktion bezeichnet. Da $\boldsymbol{\beta} = [\beta_0, ..., \beta_q]'$, erhält man $q+1$ nichtlineare Gleichungen. Diese werden in der Praxis numerisch gelöst, indem man iterativ die Nullstelle von $\mathbf{s}$ oder das Maximum von l sucht. Es handelt sich um Verfahren, die schon von Newton benutzt wurden.

Differenziert man nach $\beta_k, k = 0, ..., q$, so ergibt sich

$$s_k = \partial l(\boldsymbol{\beta})/\partial \beta_k = \sum_n \frac{y_n}{\pi_n}\frac{\partial \pi_n}{\partial \beta_k} - \frac{(1-y_n)}{(1-\pi_n)}\frac{\partial \pi_n}{\partial \beta_k}. \tag{6.50}$$

Die Ableitung des Prädiktors $\pi_n = h(\mathbf{x}_n'\boldsymbol{\beta})$ lautet

$$\frac{\partial \pi_n}{\partial \beta_k} = \frac{\partial h(\mathbf{x}_n'\boldsymbol{\beta})}{\partial \beta_k} \tag{6.51}$$

$$= h'(\mathbf{x}_n'\boldsymbol{\beta})x_{nk} := \pi_n' x_{nk}. \tag{6.52}$$

da $\partial(\mathbf{x}_n'\boldsymbol{\beta})/\partial\beta_k = x_{nk}$. Hierbei ist h' die Ableitung der Response-Funktion. Zusammengefaßt gilt also

$$s_k = \sum_n \frac{y_n - \pi_n}{\pi_n(1-\pi_n)}\frac{\partial \pi_n}{\partial \beta_k} \tag{6.53}$$

$$= \sum_n \frac{\pi_n' x_{nk}}{\pi_n(1-\pi_n)}(y_n - \pi_n) \tag{6.54}$$

$$= \sum_n \frac{h'(\mathbf{x}_n'\boldsymbol{\beta})x_{nk}}{h(\mathbf{x}_n'\boldsymbol{\beta})(1-h(\mathbf{x}_n'\boldsymbol{\beta}))}\left(y_n - h(\mathbf{x}_n'\boldsymbol{\beta})\right). \tag{6.55}$$

Speziell gilt für die logistische Funktion ($h = L$) die Formel

$$L(x) = \frac{1}{1+e^{-x}} \tag{6.56}$$

$$L'(x) = \frac{e^{-x}}{(1+e^{-x})^2} \tag{6.57}$$

$$L(1-L) = \frac{e^{-x}}{(1+e^{-x})^2} \tag{6.58}$$

und somit $L' = L(1-L)$. Man erhält also einfach

$$\partial l(\boldsymbol{\beta})/\partial \beta_k = \sum_n x_{nk}(y_n - L(\mathbf{x}_n'\boldsymbol{\beta})). \tag{6.59}$$

Beispiel 6.6 (Nominale Regressoren)
Nimmt man an, daß nur eine nominale Variable $x_n = 1, ..., I$ als unabhängige Variable vorliegt, so kann man eine Dummy-Codierung vornehmen. Die Design-Matrix hat die gruppierte Form (I Blöcke)

$$\mathbf{X} = \left[\begin{array}{c|ccccc}
1 & 1 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & & & \vdots \\
1 & 1 & 0 & 0 & \cdots & 0 \\
\hline
1 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
1 & 0 & 1 & 0 & \cdots & 0 \\
\hline
\vdots & & & & & \vdots \\
\hline
1 & 0 & 0 & 0 & \cdots & 1 \\
\vdots & \vdots & \vdots & & & \vdots \\
1 & 0 & 0 & 0 & \cdots & 1 \\
\hline
1 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
1 & 0 & 0 & 0 & \cdots & 0
\end{array}\right] : N \times I \tag{6.60}$$

Es gilt also $x_{n0} = 1$ (erste Spalte), $x_{nk} = 1$, wenn $x_n = k$, 0 sonst, $k = 1, ...I-1$. Die Zahl der Beobachtungen mit $x_n = k$ wird mit $n_k = \sum x_{nk}$ und der entsprechende Mittelwert der Response-Variable

mit $\bar{y}_k = p_k$ (relative Häufigkeit) bezeichnet. Für die Ausprägung $k = I$ setzt man $n_I = N - \sum_{k=1}^{I-1} n_k, \bar{y}_I = p_I$. Dann gilt für die Nullstelle der Score-Funktion ($k = 1, ..., I-1$), d.h. am Maximum der Likelihood

$$s_k = \sum_n x_{nk}(y_n - L(\mathbf{x}_n'\boldsymbol{\beta})) = \sum_{n, x_{nk}=1} (y_n - L(\beta_0 + \beta_k)) \tag{6.61}$$

$$= n_k\bar{y}_k - n_k L(\beta_0 + \beta_k) \stackrel{!}{=} 0. \tag{6.62}$$

Daraus ergibt sich der ML-Schätzer für $\boldsymbol{\beta}$

$$\bar{y}_k = p_k = L(\hat{\beta}_0 + \hat{\beta}_k) \tag{6.63}$$

$$\hat{\beta}_0 + \hat{\beta}_k = L^{-1}(p_k) = \log(p_k/(1-p_k)) = \text{logit}(p_k). \tag{6.64}$$

Für $k = 0$ findet man ($x_{n0} = 1$, erste Spalte von $\mathbf{X}$)

$$s_0 = \sum_n y_n - L(\mathbf{x}_n'\boldsymbol{\beta}) \tag{6.65}$$

$$= \sum_{k=1}^{I-1} [n_k p_k - n_k L(\beta_0 + \beta_k)] + [n_I p_I - n_I L(\beta_0)] \tag{6.66}$$

$$= n_I p_I - n_I L(\beta_0) = 0 \tag{6.67}$$

Die Terme in der Summe verschwinden wegen (6.63). Daher gilt

$$p_I = L(\hat{\beta}_0) \tag{6.68}$$

$$\hat{\beta}_0 = L^{-1}(p_I) = \log(p_I/(1-p_I)) = \text{logit}(p_I). \tag{6.69}$$

Dies wurde bereits in Bsp. 6.3 angemerkt. Die Response-Funktion stimmt also mit den relativen Häufigkeiten $p_k, k = 1, ..., I$ überein.

Man kann die Analyse natürlich auch gleich mit gruppierten Daten

durchführen. Man tabelliert hier

n	**p**	**X**					
n_1	p_1	1	1	0	0	$\cdots$	0
n_2	p_2	1	0	1	0	$\cdots$	0
$\vdots$							$\vdots$
n_{I-1}	p_{I-1}	1	0	0	0	$\cdots$	1
n_I	p_I	1	0	0	0	$\cdots$	0

(6.70)

■

Im allgemeinen muß der Maximum Likelihood-Schätzer iterativ durch Maximierung der Likelihood-Funktion berechnet werden. In den Software-Paketen wird hierfür ein Newton-Raphson-Algorithmus verwendet. Die 2. Ableitung der Likelihood-Funktion

Hesse-Matrix

$$\mathbf{H}_{kl}(\boldsymbol{\beta}) = \partial^2 l(\boldsymbol{\beta})/\partial\beta_k\partial\beta_l \tag{6.71}$$

wird hierbei als Gewichtsmatrix verwendet. Sie wird auch als Hesse-Matrix bezeichnet. Die Newton-Raphson-Iteration hat die Form

Newton-Raphson-Algorithmus

$$\boldsymbol{\beta}_{i+1} = \boldsymbol{\beta}_i - \mathbf{H}^{-1}(\boldsymbol{\beta}_i)\mathbf{s}(\boldsymbol{\beta}_i) \tag{6.72}$$

Man startet von einem Anfangswert $\boldsymbol{\beta}_0$ und iteriert so lange, bis sich $\boldsymbol{\beta}_i$ nur noch wenig ändert. Dann gilt $\mathbf{s}(\boldsymbol{\beta}_i) \approx \mathbf{0}$, sodaß man eine Nullstelle der Score-Funktion gefunden hat. Der Wert $\boldsymbol{\beta}_i \approx \hat{\boldsymbol{\beta}}$ wird als Maximum-Likelihood-Schätzwert genommen.

Man kann zeigen, daß die Kovarianzmatrix $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$ des ML-Schätzers asymptotisch (d.h. für $N \to \infty$) durch die *Inverse* der sog. Fisher-Informationsmatrix

Fisher-Informationsmatrix

$$\mathbf{F} := E[\mathbf{s}\mathbf{s}'] = -E[\mathbf{H}] \approx -\mathbf{H} \tag{6.73}$$

gegeben ist. Siehe Mardia et al. (1979, Kap. 4), Fahrmeir et al. (1996, Kap. 3.2.3).

Setzt man anstelle des unbekannten $\boldsymbol{\beta}$ die ML-Schätzung $\hat{\boldsymbol{\beta}}$ ein, so ist die Matrix $-\mathbf{H}$ bereits im letzten Schritt der Newton-Raphson-Iteration berechnet worden. $\mathbf{J} = -\mathbf{H}$ wird auch als *beobachtete* Fisher-Informationsmatrix bezeichnet. Man kann also die Standardfehler numerisch durch die Diagonale von $\mathbf{J}^{-1}$ gewinnen. Alternativ kann man die geschätzte Fisher-Informationsmatrix $\mathbf{F}(\hat{\boldsymbol{\beta}})$ benutzen (vgl. Bsp. 6.8).

Die ML-Schätzer sind asymptotisch normalverteilt, d.h.

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, \mathbf{F}^{-1}). \tag{6.74}$$

Man kann daher leicht Tests und Konfidenzintervalle konstruieren.

Mit Hilfe der Standardfehler $s_k = \sqrt{(\mathbf{J}^{-1})_{kk}}$ lassen sich etwa t-Werte $\hat{\beta}_k / s_k \overset{a}{\sim} N(0,1)$ berechnen.

Im Fall des Logit-Modells ergeben sich wieder besonders einfache Formeln. Aus der Score-Funktion (6.59) findet man durch Ableiten die Hesse-Matrix

$$H_{kl} = -\sum_n x_{nk} x_{nl} L'(\mathbf{x}_n' \boldsymbol{\beta}) \tag{6.75}$$

$$= -\sum_n x_{nk} x_{nl} L(\mathbf{x}_n' \boldsymbol{\beta})(1 - L(\mathbf{x}_n' \boldsymbol{\beta})). \tag{6.76}$$

In diesem Fall gilt $\mathbf{F} = E[-\mathbf{H}] = -\mathbf{H}$, da die Hesse-Matrix keine zufälligen Werte y_n enthält.

6.1.5 Asymptotische Tests für Parameter

Die allgemeine lineare Hypothese

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \boldsymbol{\xi}, \qquad H_1 : \mathbf{C}\boldsymbol{\beta} \neq \boldsymbol{\xi} \tag{6.77}$$

kann mit Hilfe einer Hypothesenmatrix $\mathbf{C} : r \times (q+1)$ formuliert werden. Es wird angenommen, daß die Matrix vollen Zeilenrang besitzt, d.h. $\text{rg}(\mathbf{C}) = r \leq q+1$. [6]

Speziell ($\mathbf{C} = \mathbf{I}_{q+1}$) erhält man die Hypothese $H_0 : \boldsymbol{\beta} = \boldsymbol{\xi}$. [7]

Die Hypothese kann mit Hilfe der Likelihood-Quotienten-Statistik

$$LQ = -2[l(\hat{\boldsymbol{\beta}}_0) - l(\hat{\boldsymbol{\beta}})] = -2 \log \frac{L(\hat{\boldsymbol{\beta}}_0)}{L(\hat{\boldsymbol{\beta}})} \tag{6.78}$$

Likelihood-Quotienten-Statistik

geprüft werden. Hierbei ist $\hat{\boldsymbol{\beta}}_0$ der ML-Schätzer von $\boldsymbol{\beta}$ unter H_0. Große Werte von LQ sprechen also gegen die Nullhypothese.

Alternativ kann die Wald-Statistik

$$W = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi})'(\mathbf{C}\mathbf{F}^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}) \tag{6.79}$$

Wald-Statistik

oder die Rao-Score-Statistik

$$S = \mathbf{s}'(\hat{\boldsymbol{\beta}}_0)\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}}_0)\mathbf{s}(\hat{\boldsymbol{\beta}}_0) \tag{6.80}$$

Rao-Score-Statistik

benutzt werden. Da am Maximum $\mathbf{s}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ gilt, kann die Score-Statistik als Abstand zwischen $\mathbf{s}(\hat{\boldsymbol{\beta}}_0)$ und $\mathbf{s}(\hat{\boldsymbol{\beta}})$ mit Gewichtsmatrix $\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}}_0)$ interpretiert werden.

Es gilt das Resultat: Die Teststatistiken sind asymptotisch äquivalent und Chi-Quadrat-verteilt, d.h.

$$LQ, W, S \overset{a}{\sim} \chi^2_r \tag{6.81}$$

[6] D.h. alle Zeilen von $\mathbf{C}$ sind voneinander linear unabhängig.

[7] Die Bezeichnung $\boldsymbol{\beta}_0$ wird für die ML-Schätzung unter H_0 reserviert.

Beispiel 6.7 (Test für *einen* Koeffizienten)
Soll nur der Parameter β_j getestet werden, so wählt man die Matrix $\mathbf{C} = [0, ..., 1, 0, ..., 0]$ und $\boldsymbol{\xi} = [\xi]$. Dies ergibt das Hypothesenpaar

$$H_0 : \beta_j = \xi, \qquad H_1 : \beta_j \neq \xi \tag{6.82}$$

Die Wald-Statistik ist dann

$$\begin{aligned} W &= (\hat{\beta}_j - \xi)(\mathbf{F}^{-1})_{jj}^{-1}(\hat{\beta}_j - \xi) && (6.83)\\ &= (\hat{\beta}_j - \xi)^2/s_j^2 && (6.84)\\ &\overset{a}{\sim} \chi^2(1). && (6.85) \end{aligned}$$

Hierbei ist $s_j^2 = [F^{-1}(\hat{\boldsymbol{\beta}})]_{jj}$ das j-te Diagonalelement der inversen geschätzten Fisher-Informationsmatrix (vgl. 6.74), also die asymptotische Kovarianzmatrix der Schätzer. Diese Chi-Quadrat-Werte werden von den Programmen ausgedruckt (vgl. Abb. 6.4, 6.5; $(.337/.176)^2 = 3.66635$).
Manchmal wird auch die t-Statistik angegeben, d.h.

$$t = (\hat{\beta}_j - \xi)/s_j \overset{a}{\sim} N(0, 1). \tag{6.86}$$

■

Ein simultaner Test für mehrere Parameter, etwa für die $I-1$ Dummyvariablen einer nominalskalierten Variable mit I Abstufungen führt auf den Test

$$H_0 : \beta_j = 0, j = 1, ..., I-1 \tag{6.87}$$

mit der Hypothesenmatrix

$$\mathbf{C} = [\mathbf{I}_{I-1}, \mathbf{O}] : (I-1) \times (q+1), \ \boldsymbol{\xi} = \mathbf{0}_{I-1}. \tag{6.88}$$

Die Wald-Statistik ist dann

$$W = [\hat{\beta}_1, ..., \hat{\beta}_{I-1}](\mathbf{C}\mathbf{F}^{-1}\mathbf{C}')^{-1}[\hat{\beta}_1, ..., \hat{\beta}_{I-1}]', \tag{6.89}$$

wobei $\mathbf{C}\mathbf{F}^{-1}\mathbf{C}'$ eine Teilmatrix der asymptotischen Kovarianzmatrix $\mathbf{F}^{-1}$ ist.

Der Stoff wird in Aufgabe 6.4 vertieft.

6.2 Modelle mit $R > 2$ Kategorien

Bisher wurden nur dichotome abhängige Variablen diskutiert. Eine Erweiterung auf R Antwortkategorien ist ohne weiteres möglich. Hierbei ist zu unterscheiden, ob man

- nominale Zielvariablen (multinomiale Modelle) oder
- ordinale Zielvariablen (kumulative oder Schwellwert-Modelle)

benutzt.

6.2.1 Multinomiales Modell

Hier wird angenommen, daß die Response-Variable $Y = 1, ..., R$ Kategorien (Ausprägungen) aufweist. Es wird ein Modell für alle $r = 1, ..., R$ Kategorien aufgestellt, d.h.

Response-Funktion

$$\pi_r = P(Y = r|\mathbf{x}) \quad = \quad h(\mathbf{x}'\boldsymbol{\beta}_r). \tag{6.90}$$

Für jede Kategorie r hat man also einen Parametervektor $\boldsymbol{\beta}_r$ und einen eigenen Regressionsansatz $\eta_r = \mathbf{x}'\boldsymbol{\beta}_r$.

Statt der nominalen Variable Y kann man einen Vektor von Dummy-Variablen verwenden, d.h. $\mathbf{y}' = [y_1, ..., y_{R-1}]$, wobei

$$y_r = \begin{cases} 1, & Y = r, r = 1, ..., R-1 \\ 0, & sonst \end{cases} \tag{6.91}$$

Die 'Referenzkategorie' $y_R = 1 - y_1 - ... - y_{R-1}$ ist gleich 1, wenn alle $y_r = 0$ sind, also für $Y = R$.

Man kann also sagen, daß $Y = r$ durch $\mathbf{y}' = [0, .., 1, 0, ..., 0]$ (1 in Spalte $r < R$, 0 sonst) und $Y = R$ durch $\mathbf{y}' = [0, .., 0]$ kodiert wird. Der Vektor

$$\mathbf{y} \sim M(1, \boldsymbol{\pi} = [\pi_1, ..., \pi_{R-1}]) \tag{6.92}$$

ist also multinomialverteilt mit Parametervektor $\boldsymbol{\pi}$. Dieser enthält nur $R - 1$ Komponenten, da $\pi_R = 1 - \pi_1 - ... - \pi_{R-1}$ durch die Normierung der Wahrscheinlichkeit redundant (überflüssig) ist.

Analog zum Fall $R = 2$ können verschiedene Modelle aufgestellt werden, etwa das multinomiale Logit-Modell mit der Response-Funktion

multinomiales Logit-Modell

$$\pi_r = p(y = r|\mathbf{x}) = \frac{\exp(\eta_r)}{1 + \sum_{r=1}^{R-1} \exp(\eta_r)} \tag{6.93}$$

$$= L_r(\eta_1, ..., \eta_{R-1}) \tag{6.94}$$

Löst man nach $\eta_r = \mathbf{x}'\boldsymbol{\beta}_r$ auf, so ergibt sich die Link-Funktion (Logit)

Link-Funktion (Logit)

$$\log \frac{\pi_r}{1 - \sum_r \pi_r} = \log(\frac{\pi_r}{\pi_R}) \tag{6.95}$$

$$= \eta_r = \mathbf{x}'\boldsymbol{\beta}_r \tag{6.96}$$

Man erhält also Log-Chancen (log-odds) bzgl. der Referenzkategorie R. Im Spezialfall $R = 2$ ergeben sich die Formeln des binären Logitmodells, d.h.

$$\pi_1 = p(y = 1|\mathbf{x}) = \frac{\exp(\eta_1)}{1 + \exp(\eta_1)} \tag{6.97}$$

$$\log(\frac{\pi_1}{\pi_2}) = \eta_1 = \mathbf{x}'\boldsymbol{\beta}_1 \tag{6.98}$$

Zu beachten ist, daß hier die Kodierung $y = 1, 2$ (bzw. im vorigen Abschnitt $y = 0, 1$) benutzt wurde. Man erhält also die Log-Chancen 1 gegen 2 (bzw. 0 gegen 1). Dies erklärt, warum die Parameter in den Abb. 6.4, 6.5 das umgekehrte Vorzeichen aufweisen.

In JMP wird immer R als Referenzkategorie benutzt, während man in SPSS die Referenzkategorie auswählen kann:

```
Analysieren/Regression/Multinomial logistisch...
```

6.2.2 Kumulative oder Schwellwert-Modelle

Bei nominalen Variablen ist für jede Ausprägung r ein Parametervektor $\boldsymbol{\beta}_r$ notwendig, was zu einer großen Anzahl von zu schätzenden Größen führt. Wenn die Zielvariable Y ordinal ist, kann die Ausprägung r so interpretiert werden, als ob eine zugrundeliegende latente Variable Y^* zwischen den zwei Schwellwerten $\theta_{r-1} < Y^* \leq \theta_r$ liegt. Man postuliert also Schwellwerte $\theta_r, r = 1, ..., R-1$, setzt formal $\theta_0 = -\infty, \theta_R = +\infty$ und definiert die Zuordnung

$$Y = r \quad \Leftrightarrow \quad \theta_{r-1} < Y^* \leq \theta_r \tag{6.99}$$

$$Y \leq r \quad \Leftrightarrow \quad Y^* \leq \theta_r \tag{6.100}$$

Das latente Modell wird als Regression

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \tag{6.101}$$

mit der Verteilungsfunktion $F_\epsilon(x) = P(\epsilon \leq x)$ für den Fehlerterm definiert. Damit ergeben sich die Response-Wahrscheinlichkeiten

$$\pi_r = P(Y = r) \tag{6.102}$$

$$= P(\theta_{r-1} < Y^* \leq \theta_r) \tag{6.103}$$

$$= P(\theta_{r-1} < \mathbf{x}'\boldsymbol{\beta} + \epsilon \leq \theta_r) \tag{6.104}$$

$$= P(\theta_{r-1} - \mathbf{x}'\boldsymbol{\beta} < \epsilon \leq \theta_r - \mathbf{x}'\boldsymbol{\beta}) \tag{6.105}$$

$$= F_\epsilon(\theta_r - \mathbf{x}'\boldsymbol{\beta}) - F_\epsilon(\theta_{r-1} - \mathbf{x}'\boldsymbol{\beta}) \tag{6.106}$$

Man benötigt also nur *einen* Parametervektor $\boldsymbol{\beta}$ sowie die Schwellwerte. Die Kurven in Abhängigkeit von $\mathbf{x}$ für unterschiedliches r unterscheiden sich nur durch die Schwellwerte, sind also gegeneinander verschoben.

Im Spezialfall einer logistischen Verteilung für ϵ erhält man das kumulative Logit-Modell, d.h.

$$\pi_r = L(\theta_r - \mathbf{x}'\boldsymbol{\beta}) - L(\theta_{r-1} - \mathbf{x}'\boldsymbol{\beta}) \tag{6.107}$$

Man kann auch schreiben

$$P(Y \leq r) = L(\theta_r - \mathbf{x}'\boldsymbol{\beta}) = \frac{\exp(\theta_r - \mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\theta_r - \mathbf{x}'\boldsymbol{\beta})} \tag{6.108}$$

und für die log-odds

$$\log \frac{P(Y \leq r)}{P(Y > r)} = \theta_r - \mathbf{x}'\boldsymbol{\beta}. \tag{6.109}$$

Beispiel 6.8 (Mineralwasser)

Im folgenden wird die Modellierung der ordinalen Variable 'Mineralisation' (mit den Abstufungen *Extrem gering, Gering, Mittel, Stark, Sehr stark, Extrem stark*) als Funktion der quantitativen Variable 'Natrium' (in mg Ionen) durchgeführt. Abb. 6.21 zeigt ein multinominales Logit-Modell, während in Abb. 6.22 ein kumulatives Logit-Modell geschätzt wird. Das letztere Modell benötigt nur die Schwellwerte (als Achsenabschnitt[y_r] bezeichnet) sowie einen Regressionskoeffizient β_1 für die Variable 'Natrium'. Die Konstante β_0 ist in den Schwellwerten enthalten, da die Responsefunktion von den Argumenten $\theta_r - \mathbf{x}'\boldsymbol{\beta} = \theta_r - \beta_0 - \beta_1 x_1$ abhängt und somit nur $\theta_r - \beta_0$ geschätzt werden kann. Der Regressionsparameter β_1 wird als $\hat{\beta}_1 = -.0096$ geschätzt. Man muß aber beachten, daß in JMP die Response-Wahrscheinlichkeit als $P(Y \leq r) = L(\theta_r + \mathbf{x}'\boldsymbol{\beta})$ definiert wird. Dagegen verwendet SPSS die oben benutzte Definition, sodaß man die Schätzung $\hat{\beta}_1 = 0.10$ erhält (Abb. 6.23). Daher sinkt (in beiden Fällen) mit höherem Natriumgehalt die relative Wahrscheinlichkeit $P(Y \leq r)/P(Y > r)$ (odds), daß höchstens Kategorie (Mineralisation) r vorliegt.

Dagegen wird bei der nominalen Modellierung für jede Abstufung r von 'Mineralisation' eine Konstante β_{r0} sowie ein Regressionskoeffizient β_{r1} benötigt (z.B. Achsenabschnitt[Sehr stark], Natrium[Sehr stark]). Die Wahrscheinlichkeit für extrem starke Mineralisierung nimmt mit steigendem Natriumgehalt stark zu (rechts oben in den Graphiken), während die anderen Wahrscheinlichkeiten gegen Null gehen. Die multinomiale Parametrisierung führt zu unterschiedlichen Kurven, während im kumulativen Modell nur Verschiebungen der logistischen Kurve durch unterschiedliche Schwellwerte zu sehen sind (vgl. Glg. 6.107).

Die Outputs enthalten auch Tests für das Gesamtmodell (Omnibustest), bei dem das Modell *mit* x-Variablen und das reduzierte Modell (nur mit Schwellwerten/intercepts) mit Hilfe eines Likelihood-Quotienten-Tests verglichen wird. Im ordinalen Fall ergibt sich die Differenz 47.922 mit einem Freiheitsgrad (da nur eine x-Variable vorliegt). Weiterhin werden die ML-Schätzungen mit den asymptotischen Standardfehlern und den χ^2-Werten angegeben. Vergleicht man die Standardfehler und χ^2-

Werte in Abb. 6.22 mit Abb. 6.23 (Spalte: Wald-Chi-Quadrat), so ergeben sich etwas unterschiedliche Werte. Dies liegt an der Art der Berechnung der asymptotischen Kovarianzmatrix des ML-Schätzers. Im Fall von SPSS (`Analysieren/Verallgemeinerte Lineare Modelle/...`) kann man als Option die Berechnung der Standardfehler mit Hilfe der Hesse-Matrix (Option Newton-Raphson) oder der geschätzten Fisher-Informationsmatrix auswählen (Abb. 6.24).

Die Modelle können in SPSS mit den Menü-Befehlen

```
Analysieren/Regression/Multinomial logistisch...
Analysieren/Regression/Ordinal...
Analysieren/Verallgemeinerte Lineare Modelle/...
```

aufgerufen werden (siehe Abb. 6.25).

■

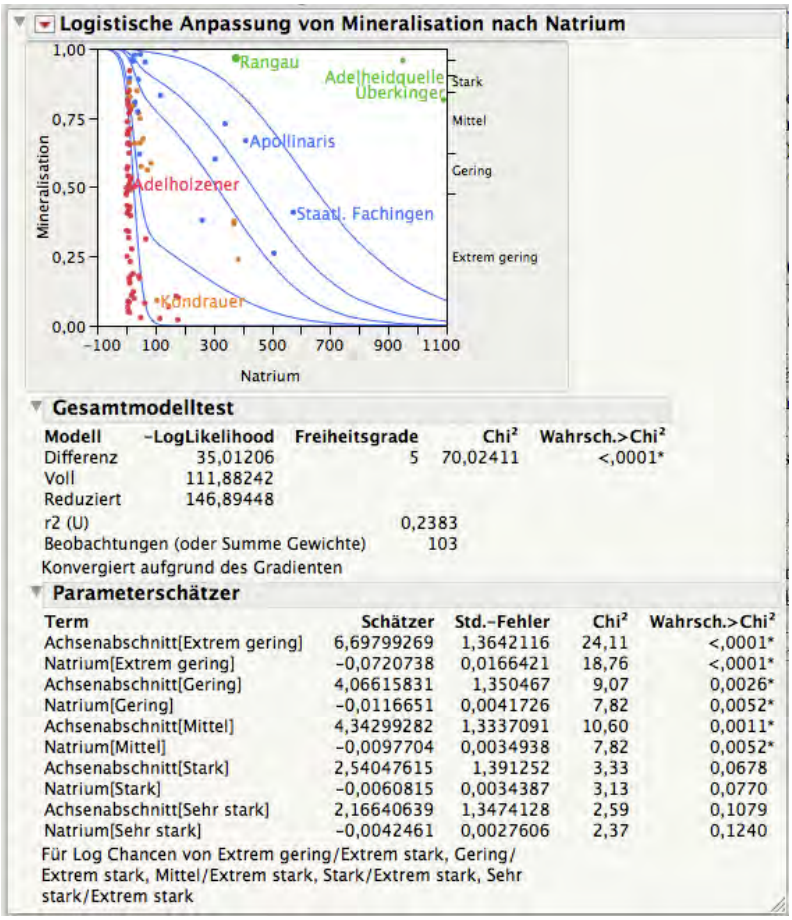


Abbildung 6.21: JMP: 'Mineralisation' (mit den Abstufungen *Extrem gering*, *Gering*, *Mittel*, *Stark*, *Sehr stark*, *Extrem stark*) als Funktion der quantitativen Variable 'Natrium'. Nominales Logit-Modell.

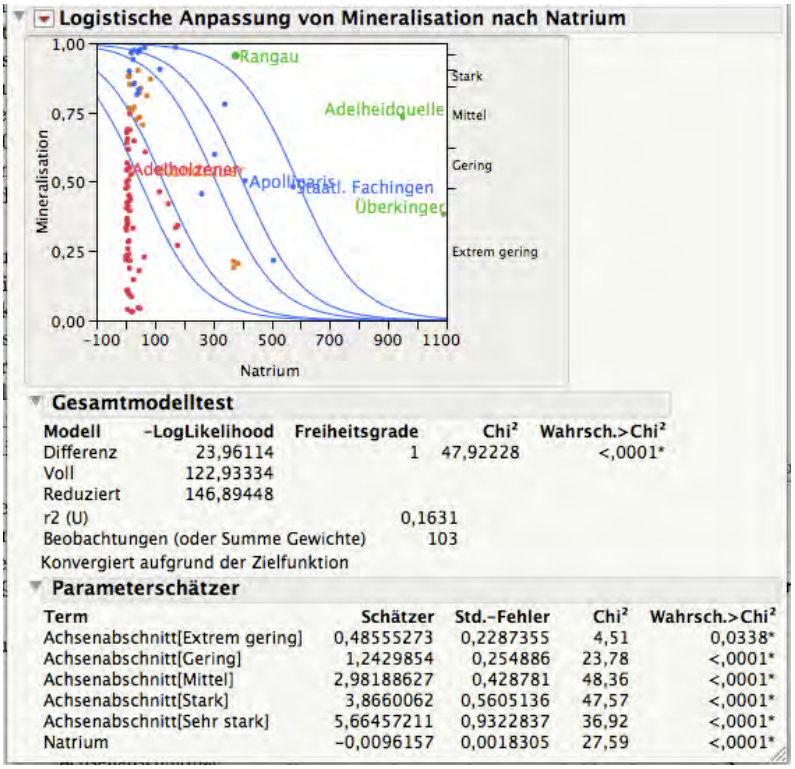


Abbildung 6.22: JMP: 'Mineralisation' (mit den Abstufungen *Extrem gering*, *Gering*, *Mittel*, *Stark*, *Sehr stark*, *Extrem stark*) als Funktion der quantitativen Variable 'Natrium'. Ordinales Logit-Modell.

Omnibus-Test[a]

Likelihood-Quotienten-Chi-Quadrat	df	Sig.
47,922	1	,000

Abhängige Variable: mineralinum
Modell: (Schwellenwert), natrium

a. Vergleicht das angepasste Modell mit dem Modell, bei dem ausschließlich Schwellenwerte verwendet werden.

Anpassungsgüte[a]

	Wert	df	Wert/df
Abweichung	234,776	474	,495
Skalierte Abweichung	234,776	474	
Pearson-Chi-Quadrat	456,645	474	,963
Skaliertes Pearson-Chi-Quadrat	456,645	474	
Log-Likelihood[b]	-122,933		
Akaike-Informations-Kriterium (AIC)	257,867		
AIC mit Korrektur für endliche Stichproben (AICC)	258,742		
Bayes-Informationskriterium (BIC)	273,675		
Konsistentes AIC (CAIC)	279,675		

Abhängige Variable: mineralinum
Modell: (Schwellenwert), natrium

a. Die Informationskriterien liegen in einem möglichst kleinem Format vor.

b. Der Kernel der Log-Likelihood-Funktion wird angezeigt und bei der Berechung der Informationskriterien verwendet.

Parameterschätzer

Parameter		Regressionskoeffizient B	Standardfehler	95% Wald-Konfidenzintervall		Hypothesentest		
				Unterer Wert	Oberer Wert	Wald-Chi-Quadrat	df	Sig.
Schwellenwert	[mineralinum=1,00]	,486	,2309	,033	,938	4,422	1	,035
	[mineralinum=2,00]	1,243	,2582	,737	1,749	23,182	1	,000
	[mineralinum=3,00]	2,982	,4235	2,152	3,812	49,584	1	,000
	[mineralinum=4,00]	3,866	,5446	2,799	4,933	50,397	1	,000
	[mineralinum=5,00]	5,665	,8991	3,903	7,427	39,698	1	,000
natrium		,010	,0017	,006	,013	31,116	1	,000
(Skala)		1[a]						

Abhängige Variable: mineralinum
Modell: (Schwellenwert), natrium

a. Auf den angezeigten Wert festgesetzt.

Parameterschätzer

Parameter		Regressionskoeffizient B	Standardfehler	95% Wald-Konfidenzintervall		Hypothesentest		
				Unterer Wert	Oberer Wert	Wald-Chi-Quadrat	df	Sig.
Schwellenwert	[mineralinum=1,00]	,486	,2287	,037	,934	4,506	1	,034
	[mineralinum=2,00]	1,243	,2549	,743	1,743	23,782	1	,000
	[mineralinum=3,00]	2,982	,4288	2,141	3,822	48,363	1	,000
	[mineralinum=4,00]	3,866	,5605	2,767	4,965	47,572	1	,000
	[mineralinum=5,00]	5,665	,9323	3,837	7,492	36,918	1	,000
natrium		,010	,0018	,006	,013	27,594	1	,000
(Skala)		1[a]						

Abhängige Variable: mineralinum
Modell: (Schwellenwert), natrium

a. Auf den angezeigten Wert festgesetzt.

Abbildung 6.23: SPSS: ordinales Logit-Modell. Oben links: Test für das Gesamtmodell, oben rechts: Likelihood und diverse Informationskriterien, Mitte: Standardfehler mit Hesse-Matrix, unten: Standardfehler mit Fisher-Informationsmatrix.
Menü: `Analysieren/Verallgemeinerte Lineare Modelle/....`

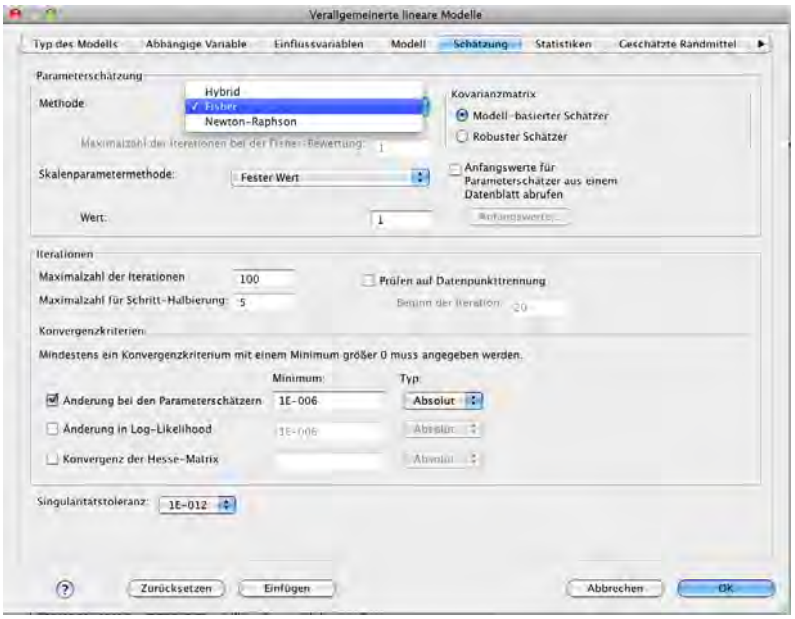


Abbildung 6.24: SPSS: Verallgemeinertes lineares Modell: Option für die Standardfehler.

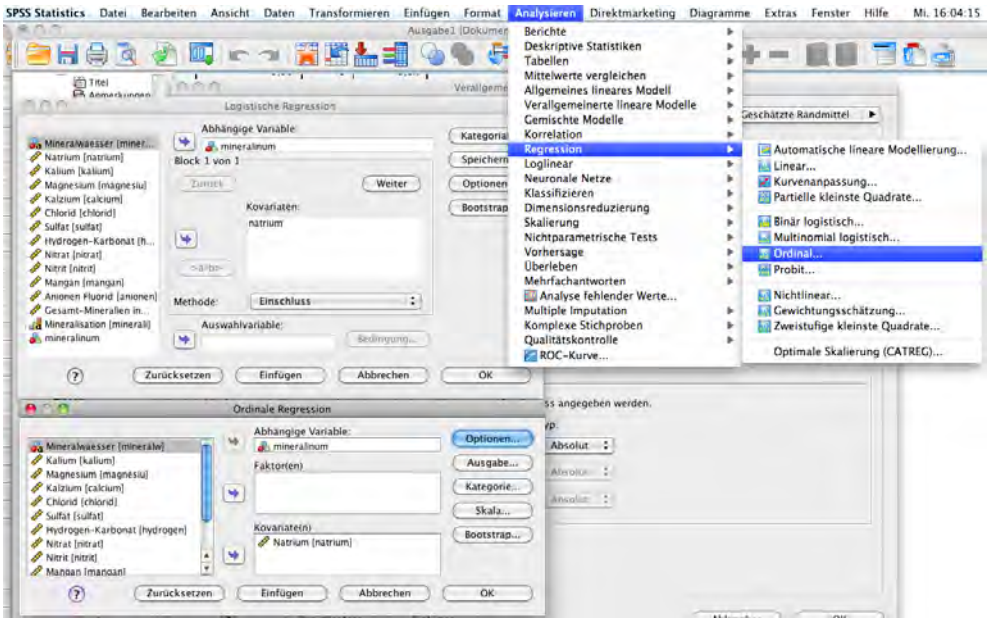


Abbildung 6.25: SPSS: Auswahlmenü für logistische Regression.

Kapitel 7

Cluster-Analyse

7.1 Übersicht

In den bisherigen Kapiteln wurde davon ausgegangen, daß die statistischen Einheiten bereits bestimmten Gruppen, etwa gute/schlechte Schuldner, zugeordnet werden konnten. Hier wird die umgekehrte Fragestellung behandelt, wie man die Objekte (Personen, Staaten, Firmen, Mineralwässer etc.) anhand ihrer Kennzahlen (Variablen) so gruppieren kann, daß die Objekte innerhalb einer Gruppe (Klasse) möglichst ähnlich sind, jedoch zwischen den Klassen maximale Unterschiede bestehen.

Meistens werden die Objekte als Zeilen der Datenmatrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]'$ aufgefaßt und Ähnlichkeiten zwischen den Zeilen $\mathbf{x}'_n$ und $\mathbf{x}'_m$ untersucht. Dazu wird ein Abstand $d(n, m) = ||\mathbf{x}_n - \mathbf{x}_m||$ benötigt. Die Merkmalsvektoren $\mathbf{x}_n$ werden dazu als Elemente eines p-dimensionalen Raums aufgefaßt.

Man kann jedoch auch Ähnlichkeiten zwischen den Spalten $\mathbf{x}_{(i)}$ und $\mathbf{x}_{(j)}$ der Datenmatrix untersuchen. Diese entsprechen den Variablen X_i und X_j. Man erhält so Cluster von ähnlichen Markmalen.

Cluster

Der englische Begriff Cluster kann mit *Traube, Büschel, Haufen, Schwarm* oder *Gruppe* übersetzt werden. Bei der Bildung von Klassen gibt es einen großen Spielraum, etwa bei der Auswahl von

- Variablen
- Abstandsmaßen
- Klassendistanzen
- Anzahl der Cluster etc.

Daher ist es nicht sinnvoll, von einer richtigen oder falschen, sondern von einer zweckmäßigen Gruppierung zu sprechen. Es sind also vor allem inhaltliche Kriterien von Bedeutung.

Zunächst soll die Terminologie eingeführt werden, die im folgenden benutzt wird.

Objektmenge

Die Menge der zu klassifizierenden Objekte (Objektmenge) wird als

$$I = \{I_1, ..., I_N\} \tag{7.1}$$

oder kurz

$$I = \{1, ..., N\} \tag{7.2}$$

bezeichnet. Die Merkmale $\mathbf{x}_n = [x_{n1}, ..., x_{np}]'$ der Objekte $n = 1, ..., N$ werden in einer Datenmatrix

$$\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]' = [\mathbf{x}_{(1)}, ..., \mathbf{x}_{(p)}] : N \times p \tag{7.3}$$

zusammengefaßt. Weiterhin können folgende Fälle unterschieden werden:

- Die Merkmale gehen direkt in die Klassifizierung ein.
- Es werden zuerst Abstände $d(n, m) = d_{nm} = \text{Abstand}(I_n, I_m)$ aus den Merkmalen berechnet.
- Die Abstände werden direkt erhoben.

Ziel der Clusteranalyse ist eine Klasseneinteilung (Partitionierung, Zerlegung)

Klasseneinteilung

$$C = \{C_1, ..., C_g\} \tag{7.4}$$

$$I = \bigcup_{k=1}^{g} C_k \tag{7.5}$$

$$C_k \cap C_j = \emptyset, k \neq j \tag{7.6}$$

der Menge der zu klassifizierenden Objekte in g disjunkte Klassen. Die verschiedenen Verfahrensweisen können folgendermaßen klassifiziert werden (Abb. 7.1). Hierbei werden bei den stochastischen Verfahren die Be-

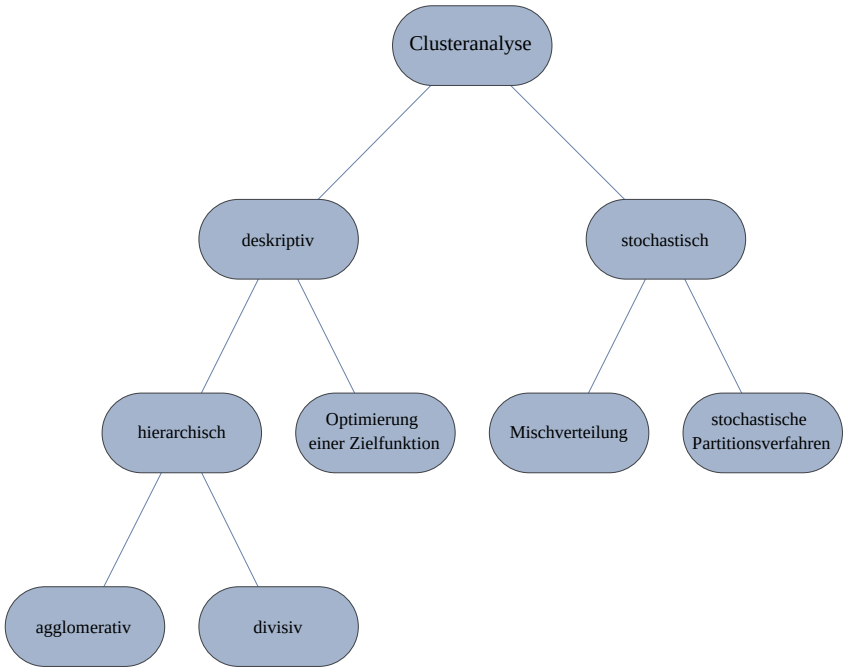


Abbildung 7.1: Clustermethoden im Überblick (vgl. Fahrmeir et al., 1996, Kap. 9.1).

obachtungen als Zufallvariable angesehen, deren Verteilung $p(\mathbf{x}|k)$ durch die unbekannte, nichtbeobachtbare Klassenzugehörigkeit k gegeben ist. Hierarchische Klassifikationsverfahren konstruieren eine Folge von Partitionen. Bei den agglomerativen Verfahren werden Klassen sukzessive zusammengelegt, während sie bei den divisiven Verfahren aufgeteilt werden. Auch ist die Klasseneinteilung durch Optimierung einer Zielfunktion möglich.

7.2 Ähnlichkeits- und Distanzmaße

Um die Ähnlichkeit von Objekten bzw. Clustern beurteilen zu können, muß ein Ähnlichkeits- oder Distanzmaß eingeführt werden. Der Begriff der Distanz wird hierbei analog zum räumlichen Abstand in metaphorischer Form benutzt.

Ein Ähnlichkeitsmaß (s = similarity) $s_{nm} = s(I_n, I_m) \geq 0$ ist eine Funktion der Objekte $I_n \in I$ mit den Eigenschaften

Ähnlichkeitsmaß

$$s_{nm} = s_{mn} \text{ (Symmetrie)} \tag{7.7}$$
$$s_{nm} \leq s_{nn} \text{ (n ist ähnlicher zu sich selbst)} \tag{7.8}$$

Man kann auch zusätzlich $s_{nm} \geq 0$ und $s_{nn} = 1$ fordern.

Ähnlichkeitsmatrix

$\mathbf{S} = s_{nm}$ heißt Ähnlichkeitsmatrix.

Analog definiert man das Distanzmaß $d_{nm} = d(I_n, I_m) \geq 0$ mit den Eigenschaften

Distanzmaß

Abstandsmaß

$$d_{nm} = d_{mn} \text{ (Symmetrie)} \tag{7.9}$$
$$d_{nm} \geq 0 \quad \text{(Positivität)} \tag{7.10}$$
$$d_{nn} = 0 \quad \text{(kein Abstand zu sich selbst)} \tag{7.11}$$

Es gilt also $\mathbf{D} = \mathbf{D}', \text{diag}(\mathbf{D}) = \mathbf{0}$ und $d_{nm} \geq d_{nn} = 0$.

Ein metrisches Distanzmaß ist durch die sog. Dreiecks-Ungleichung

Dreiecks-Ungleichung

$$d_{nm} \leq d_{nl} + d_{lm} \tag{7.12}$$

definiert. Dies entspricht der räumlichen Vorstellung (Abb. 7.2). Wie schon erwähnt, können Distanzen direkt über Urteile von Probanden erhoben oder indirekt aus den Merkmalen in der Datenmatrix berechnet werden.

Abstände zwischen den Clustern

Bisher wurden nur Distanzen zwischen Objekten betrachtet. Wenn die Objekte zu Klassen C_k gruppiert werden, ist es auch notwendig, Abstände zwischen den Clustern zu definieren.

Die entsprechenen Distanz- und Ähnlichkeitsmaße werden analog definiert, etwa $S_{kl} = S(C_k, C_l)$ zwischen den Clustern C_k, C_l. Zur Unterscheidung werden die Klassenabstände groß, die Objektabstände klein

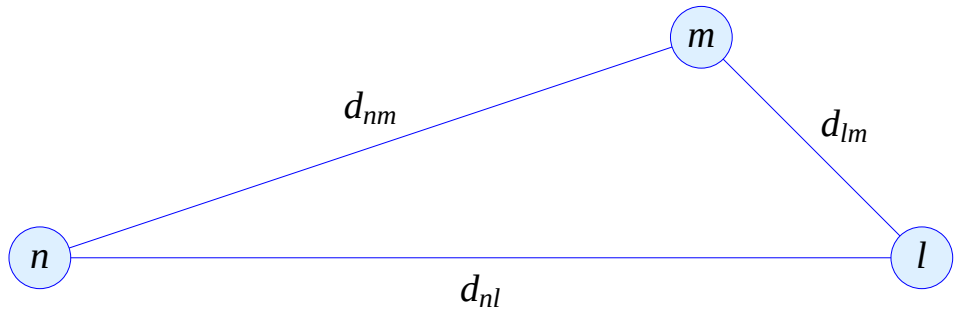


Abbildung 7.2: Dreiecks-Ungleichung.

geschrieben. Es handelt sich um eine Mengenfunktion, da die Argumente C_k aus den Objekten zusammengesetzt sind, etwa $C_k = \{I_1, I_4, I_5\}$.

Beispiel 7.1 (Potenzmenge)
Betrachtet man eine Menge von Objekten $I = \{1, 2, 3\}$, so sind alle möglichen Cluster durch die sog. Potenzmenge $P(I)$ (Menge aller Untermengen) gegeben, d.h.

$$P(I) = \Big\{\{\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\Big\}. \quad (7.13)$$

Der Abstand zwischen 2 Einzel-Objekten $\{n\}, \{m\} \in P(I)$ ist also $D(\{n\}, \{m\})$. Die Abstandsfunktion zwischen den Objekten $n, m \in I$ ist dagegen $d(n, m)$. Es gilt $D(\{n\}, \{m\}) = d(n, m)$.

■

7.3 Spezielle Distanzmaße

7.3.1 Nominalskalierte Merkmale

Bei binären Merkmalen kann man eine Kontingenztabelle

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	p

(7.14)

der Übereinstimmungen des p-dimensionalen Merkmalsvektors $\mathbf{x}_n = [0, 0, 1, ..., 0, 1, ...]'$ mit $\mathbf{x}_m = [1, 0, 1, ..., 0, 1, 1...]'$ betrachten. Hierbei sind a bzw. d die Übereinstimmungen der Einsen (Nullen), während b, c die

Nichtübereinstimmungen abzählen. Insgesamt hat man $p = a + b + c + d$ Einträge.

Daraus läßt sich ein normierter M-Koeffizient (Matching oder Übereinstimmung) berechnen, d.h.

Matching-Koeffizient

$$s_{nm} = \frac{a_{nm} + d_{nm}}{p}. \tag{7.15}$$

Beispielsweise ergibt $[\mathbf{x}_n, \mathbf{x}_m]' = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ den Wert $s_{nm} = 3/4$.

Alternativ kann man einen S-Koeffizienten (Similarity)

Similarity-Koeffizient

$$s^*_{nm} = \frac{a_{nm}}{a_{nm} + b_{nm} + c_{nm}}. \tag{7.16}$$

definieren, bei dem das gemeinsame Fehlen von Eigenschaften (also d) nicht gezählt wird.

Im obigen Beispiel hat man $s^*_{nm} = 2/4 = 1/2$.

Analog zum ϕ-Koeffizienten (Korrelation für binäre Merkmale) kann man einen Ähnlichkeitskoeffizienten der Gestalt

$$s^\circ_{nm} = \frac{a_{nm}d_{nm} - b_{nm}c_{nm}}{(a_{nm} + b_{nm})(c_{nm} + d_{nm})(a_{nm} + c_{nm})(b_{nm} + d_{nm})} \tag{7.17}$$

definieren. Hierbei wurde die Summation über die p Merkmale ausgeführt, d.h. $\widehat{\mathrm{Cov}}(\mathbf{x}_n, \mathbf{x}_m) = p^{-1} \sum_{i=1}^{p} (x_{ni} - \bar{x}_n)(x_{mi} - \bar{x}_m)$. Durch Ausnützen der 0-1-Codierung (d.h. $x_{ni}^2 = x_{ni}$) ergibt sich obiger Korrelationskoeffizient.

Übung 7.1

Bitte rechnen Sie die Werte für s_{nm} und s^*_{nm} nach und bestimmen Sie s°_{nm}.

■

7.3.2 Ordinalskalierte Merkmale

Bei geordneten Merkmalen sind die Abstände zwischen x_n und x_m Abstände in der Rangordnung. Codiert man die ordinale Variable x mit den Zahlen $1, ..., I$, so ist $x_n - x_m$ der ganzzahlige Abstand der Rangplätze der Variablen x_n, x_m. Die Rangplatz-Differenzen sind invariant gegenüber einer monotonen Transformation, die für ordinale Variablen zulässig ist.

Alternativ kann man der Ausprägung $x = i$ die binäre Hilfsvariable $\tilde{\mathbf{x}} = [1, 1, ..., 1, 0, ...]$ (1 bis Stelle i) zuordnen und dann mit den Methoden für nominale Merkmale arbeiten.

7.3.3 Quantitative Merkmale

Bei Intervall- und Verhältnisskalen sind lineare Transformationen $\tilde{x} = a + bx$ zulässig ($a = 0$ bei Verhältnisskalen).

Fordert man, daß die Maßeinheit keinen Einfluß auf die Distanz hat, so muß $d(\mathbf{x}_n, \mathbf{x}_m) = d(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_m)$ gelten mit $\tilde{x}_{ni} = bx_{ni}, i = 1, ..., p$. Ein solches Distanzmaß heißt skaleninvariant.

Skaleninvarianz

Eine mildere Forderung ist, daß sich die Maßeinheit der Distanz wie die Skala ändert, d.h.

$$d(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_m) = b \cdot d(\mathbf{x}_n, \mathbf{x}_m). \tag{7.18}$$

Da sich der Nullpunkt bei Intervallskalen ändern kann (da $a \neq 0$), sollte der Abstand nicht von der Wahl des Ursprungs abhängen. Man fordert daher Translationsinvarianz, d.h.

$$d(a + \mathbf{x}_n, a + \mathbf{x}_m) = d(\mathbf{x}_n, \mathbf{x}_m). \tag{7.19}$$

Translations-Invarianz

Die sog. L_q-Distanzen sind metrische Distanzen (erfüllen also die Dreiecksungleichung), sind translationsinvariant, jedoch nicht skaleninvariant. Sie sind als

$$d_q(\mathbf{x}_n, \mathbf{x}_m) = \left(\sum_{i=1}^{p} |x_{ni} - x_{mi}|^q \right)^{1/q} = ||\mathbf{x}_n - \mathbf{x}_m||_q \tag{7.20}$$

L_q-Distanz

definiert. Die Wahl $q = 2$ (2-Norm) entspricht der üblichen euklidischen Metrik (Distanz im Ortsraum)

$$d_2(\mathbf{x}_n, \mathbf{x}_m) = \left(\sum_{i=1}^{p} |x_{ni} - x_{mi}|^2 \right)^{1/2} = ||\mathbf{x}_n - \mathbf{x}_m||_2. \tag{7.21}$$

euklidische Distanz

In Matrixnotation kann man schreiben $d_2(\mathbf{x}_n, \mathbf{x}_m) = ||\mathbf{x}_n - \mathbf{x}_m||_2$ mit der 2-Norm $||\mathbf{x}||_2 = \sqrt{\mathbf{x}'\mathbf{x}}$.

Die Wahl $q = 1$ wird auch als City-Block-Metrik bezeichnet, da man sich

City-Block-Metrik

$$d_1(\mathbf{x}_n, \mathbf{x}_m) = \sum_{i=1}^{p} |x_{ni} - x_{mi}| = ||\mathbf{x}_n - \mathbf{x}_m||_1. \tag{7.22}$$

als Summe der Abstände in den Koordinaten-Richtungen $\mathbf{e}_i, i = 1, ..., p$ vorstellen kann (Abb. 7.3).

Um eine Skaleninvarianz zu erreichen, können die Daten standardisiert werden, d.h. $z_i = (x_i - \bar{x}_i)/s_i$.

Beispiel 7.2 (Abstände)

Die Datenmatrix der 5 Objekte $\mathbf{X} = \begin{bmatrix} 1 & 2 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 & 4 \end{bmatrix}'$

ergibt die euklidische Distanzmatrix

$$D_2 = \begin{bmatrix} 0 & \sqrt{5} & \sqrt{2} & \sqrt{13} & 3\sqrt{2} \\ \sqrt{5} & 0 & 1 & \sqrt{2} & \sqrt{5} \\ \sqrt{2} & 1 & 0 & \sqrt{5} & 2\sqrt{2} \\ \sqrt{13} & \sqrt{2} & \sqrt{5} & 0 & 1 \\ 3\sqrt{2} & \sqrt{5} & 2\sqrt{2} & 1 & 0 \end{bmatrix}.$$

(siehe Abb. 7.4). Da sich $\mathbf{x}_3 = \mathbf{x}_1 + \mathbf{a}$, $\mathbf{x}_4 = \mathbf{x}_2 + \mathbf{a}$ nur durch eine Translation $\mathbf{a} = [1, 1]'$ unterscheiden, gilt $\mathbf{x}_4 - \mathbf{x}_3 = \mathbf{x}_2 - \mathbf{x}_1$.

■

Übung 7.2 (City-Block-Metrik)

Berechnen Sie bitte die Abstände der Objekte mit der City-Block-Metrik.

■

Eigenschaften von L_q-Distanzen

Translations-Invarianz

1. *Translations-Invarianz*

   Transformiert man die Koordinaten der Objekte $\tilde{\mathbf{x}} = \mathbf{a} + \mathbf{x}$, $\tilde{\mathbf{y}} = \mathbf{a} + \mathbf{y}$, so gilt

   $$d_q(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = ||\mathbf{a} + \mathbf{x} - \mathbf{a} - \mathbf{y}||_q = ||\mathbf{x} - \mathbf{y}||_q = d_q(\mathbf{x}, \mathbf{y}). \tag{7.23}$$

   (siehe Abb. 7.4).

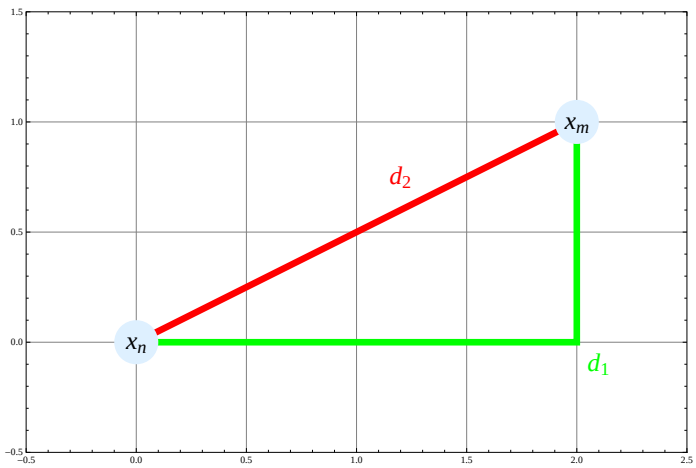


Abbildung 7.3: Vergleich von euklidischer Distanz d_2 und City-Block-Metrik d_1. Diese bleibt invariant, wenn andere kürzeste Wege entlang des Rasters genommen werden.

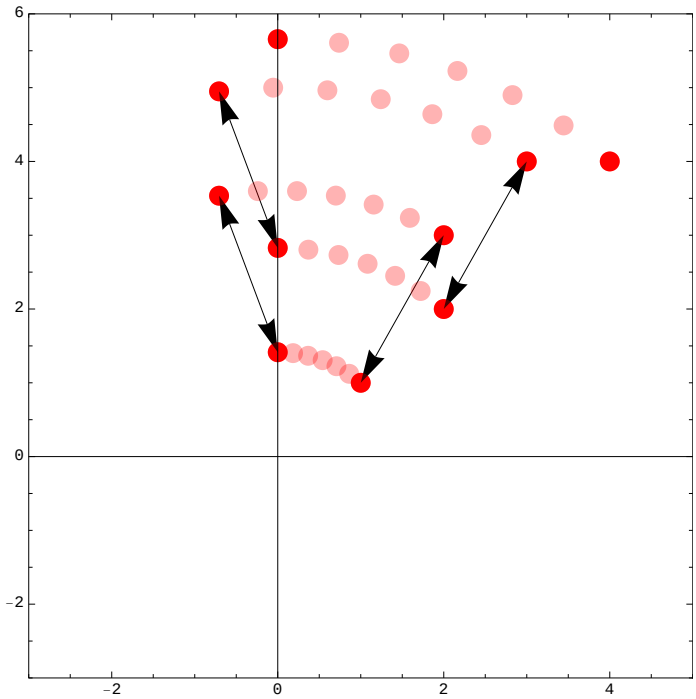


Abbildung 7.4: Daten und Abstände. Translationsinvarianz der Distanzen. Bei um ϕ rotierten Daten bleiben die Abstände invariant.

orthogonale Transformation

2. *Orthogonale Transformationen*

   (Drehung, Spiegelung) lassen die euklidische Distanz invariant.

   Die orthogonale Matrix $\mathbf{B}$, $\mathbf{B}' = \mathbf{B}^{-1}$ ergibt $\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x}$ und

$$||\tilde{\mathbf{x}}||_2 = \sqrt{\tilde{\mathbf{x}}'\tilde{\mathbf{x}}} = \sqrt{\mathbf{x}'\mathbf{B}'\mathbf{B}\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{x}} = ||\mathbf{x}||_2 \tag{7.24}$$

   (siehe Abb. 7.4).

Dreiecks-Ungleichung

3. *Dreiecks(Minkowski)-Ungleichung*

$$\begin{aligned} d_q(\mathbf{x},\mathbf{y}) &= ||\mathbf{x}-\mathbf{y}||_q = ||\mathbf{x}-\mathbf{z}+\mathbf{z}-\mathbf{y}||_q && (7.25)\\ &\leq ||\mathbf{x}-\mathbf{z}||_q + ||\mathbf{z}-\mathbf{y}||_q && (7.26)\\ &= d_q(\mathbf{x},\mathbf{z}) + d_q(\mathbf{z},\mathbf{y}). && (7.27) \end{aligned}$$

Satz von Pythagoras

4. *Satz von Pythagoras (Cosinus-Satz)*

   Für $\mathbf{c} = \mathbf{a} + \mathbf{b}$ gilt die Formel (euklidische Metrik, $q = 2$)

$$||\mathbf{c}||^2 = \mathbf{c}'\mathbf{c} = (\mathbf{a}+\mathbf{b})'(\mathbf{a}+\mathbf{b}) = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 + 2\mathbf{a}'\mathbf{b} \tag{7.28}$$

   Hierbei ist $\mathbf{a}'\mathbf{b} = ||\mathbf{a}|| \cdot ||\mathbf{b}|| \cdot \cos\phi$ das Skalarprodukt von $\mathbf{a}$ und $\mathbf{b}$, das bei orthogonalen Vektoren verschwindet.

Beispiel 7.3 (Abstände)

Die orthogonale Matrix $\mathbf{B} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$ mit $\phi = \pi/4$ ergibt die rotierte Datenmatrix

$$\tilde{\mathbf{X}}' = \mathbf{B}\mathbf{X}' = \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} & 0 \\ \sqrt{2} & \frac{5}{\sqrt{2}} & 2\sqrt{2} & \frac{7}{\sqrt{2}} & 4\sqrt{2} \end{bmatrix}$$

und die euklidische Distanzmatrix

$$\tilde{D}_2 = \begin{bmatrix} 0 & \sqrt{5} & \sqrt{2} & \sqrt{13} & 3\sqrt{2} \\ \sqrt{5} & 0 & 1 & \sqrt{2} & \sqrt{5} \\ \sqrt{2} & 1 & 0 & \sqrt{5} & 2\sqrt{2} \\ \sqrt{13} & \sqrt{2} & \sqrt{5} & 0 & 1 \\ 3\sqrt{2} & \sqrt{5} & 2\sqrt{2} & 1 & 0 \end{bmatrix}.$$

■

Übung 7.3 (Orthogonale Transformation)

Zeigen Sie, daß $\mathbf{B}\mathbf{B}' = \mathbf{I}$ und $|\mathbf{B}| = 1$ gilt und damit $\mathbf{B}$ eine orthogonale Matrix ist.

Hinweis: Verwenden Sie $\sin^2(\phi) + \cos^2(\phi) = 1$.

■

Die Mahalanobis-Distanz

Mahalanobis-Distanz

$$d_M(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}||_M \tag{7.29}$$

$$= \sqrt{(\mathbf{x}-\mathbf{y})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{y})} \tag{7.30}$$

läßt sich als euklidische Distanz mit Gewichtsmatrix $\mathbf{S}$ (empirische Kovarianzmatrix) auffassen. Sie ist skaleninvariant, da die transformierten Daten $\tilde{\mathbf{X}}' = \mathbf{C}\mathbf{X}'$ auf die Kovarianz-Matrix $\tilde{\mathbf{S}} = \mathbf{C}\mathbf{S}\mathbf{C}'$ führen. Daher gilt für $\tilde{\mathbf{x}} = \mathbf{C}\mathbf{x}$: $||\tilde{\mathbf{x}}||_M = \sqrt{\mathbf{x}'\mathbf{C}'\mathbf{C}'^{-1}\mathbf{S}^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{S}^{-1}\mathbf{x}} = ||\mathbf{x}||_M$. Dies hat auch $||\tilde{\mathbf{x}}-\tilde{\mathbf{y}}||_M = ||\mathbf{x}-\mathbf{y}||_M$ zur Folge.

Unter der Wurzel wurde die Rechenregel $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ benutzt.

Die Mahalanobis-Distanz berücksichtigt die Korrelationen der Merkmale, da man sie als euklidische Norm der standardisierten Variablen $\mathbf{z} = \mathbf{S}^{-1/2}\mathbf{x}$, $\mathbf{S}_z = \mathbf{S}^{-1/2}\mathbf{S}\mathbf{S}^{-1/2} = \mathbf{I}_p$ auffassen kann, d.h.

$$||\mathbf{z}|| = \sqrt{\mathbf{z}'\mathbf{z}} = \sqrt{\mathbf{x}'\mathbf{S}^{-1/2}\mathbf{S}^{-1/2}\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{S}^{-1}\mathbf{x}} = ||\mathbf{x}||_M. \tag{7.31}$$

7.4 Hierarchische Klassifikationsverfahren

Hier wird eine Folge von Partitionen (Zerlegungen) der Objektmenge $I = \{1, ..., N\}$ konstruiert, und zwar entweder agglomerativ oder divisiv. Bei den agglomerativen Verfahren werden die Klassen sukzessive zusammengelegt, während sie bei den divisiven Verfahren aufgeteilt werden. Die Klassifikationsergebnisse lassen sich anschaulich mit Hilfe eines Dendrogramms (Gabel-Diagrammms) darstellen (Abb. 7.5). Ausgehend von den einzelnen Ländern werden die ähnlichsten Objekte fusioniert (zuerst Australien und Kanada, dann Neuseeland und Schweden), die anschließend zusammengefaßt werden, usw. Der Index (Abstand, Unähnlichkeit, Inhomogenität) h der Cluster ist dabei nach rechts aufgetragen. In der Graphik wurden 4 Cluster eingefärbt.

Die Klassifikation hat folgende Eigenschaften:

- Jedem Wert der Indexfunktion h ist genau eine Partition zugeordnet, bei $h = 0$ hat man $C = \{\{1\}, ..., \{N\}\}$.
- Die Klassenzahl muß nicht vorgegeben werden.

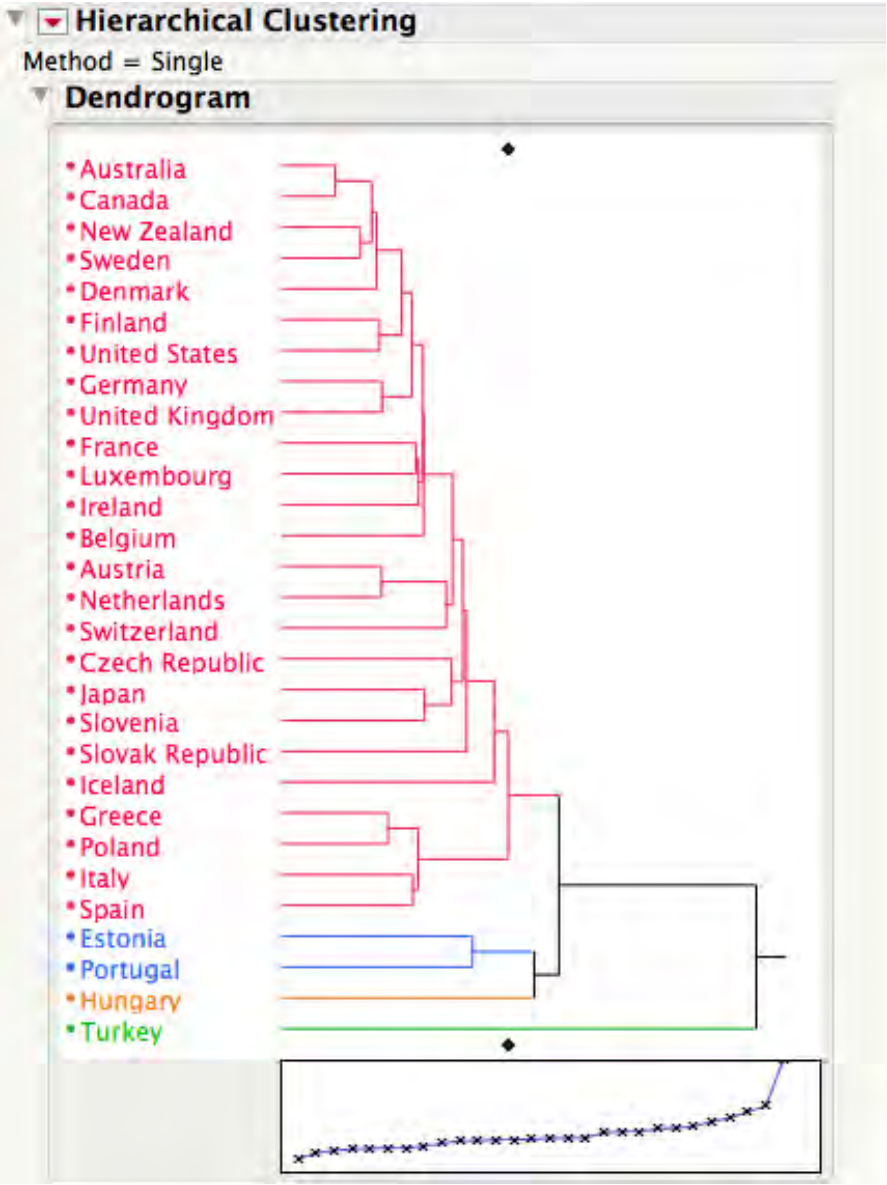


Abbildung 7.5: JMP: Dendrogramm der OECD-Länder. Zur Klassifikation wurden die Variablen `LifeSatisfaction`, `Employmentrate`, `Employmentrateofwomenwithchildren` und `Airpollution` verwendet.

- Im Dendrogramm mißt der Index h die Inhomogenität der Klassen. Aus dem Dendrogramm kann jeder Anwender ablesen, welche Klassenbildung seiner Homogenitätsvorstellung entspricht (Fahrmeir et al., 1996, Kap. 9, S. 454)
- Das Resultat hängt vom Distanzmaß ab (wird noch definiert). Hier wurde der Minimalabstand der Objekte zwischen den Klassen C_i, C_j als Abstand $D(C_i, C_j)$ genommen (single linkage).

7.4.1 Formale Beschreibung

Die Objektmenge wird wie oben als $I = \{I_1, ..., I_N\}$ oder kurz $I = \{1, ..., N\}$ bezeichnet. Die Menge aller Untermengen $P(I)$ ohne die leere Menge ist $P_*(I)$. Vgl. Bsp. 7.1.

Hierarchie

Eine Untermenge $H \subset P_*(I)$ heißt Hierarchie von I, wenn entweder

- $B \cap C = \emptyset$ oder
- $B \subset C$ oder
- $C \subset B$

für $B, C \in H$. Dies bedeutet, daß die Klassen entweder disjunkt oder vollständig ineinander enthalten sind.

Klassen

Die Elemente $C \in H$ heißen Klassen.

Wenn nun $I \in H$ und alle $\{I_j\} \in H, j = 1, ..., N$, so wird die Hierarchie als total bezeichnet. Insbesondere ist $H = P_*$ total.

Index

Eine Mengenfunktion $h \geq 0$ heißt Index zur Hierarchie H, wenn

$$h(C) \leq h(B) \quad \Leftrightarrow \quad C \subseteq B. \tag{7.32}$$

Ausserdem wird gefordert: $h(C) = 0 \Leftrightarrow$ alle Objekte in C sind gleich.

Der Index h läßt sich als Inhomogenität interpretieren, d.h. C mit $h(C) \leq h(B)$ ist homogener als B (Abb. 7.6).

7.4.2 Agglomerative Verfahren

Hier wird ausgehend von der feinsten Partition $C^0 = \{\{I_1\}, ..., \{I_N\}\}$ durch Zusammenfassen eine Hierarchie erzeugt. Die Vorgehensweise ist wie folgt:

- Man gibt sich ein Distanzmaß D oder Ähnlichkeitsmaß S vor.

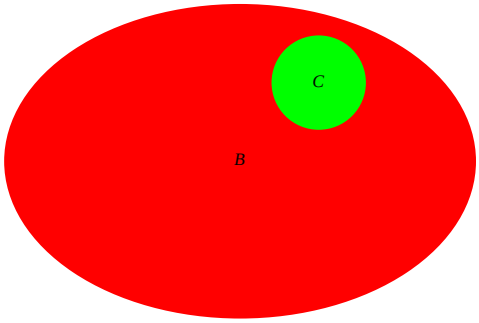


Abbildung 7.6: C ist homogener als B.

- Die Startpartition ist $C^0 = \{\{I_1\}, ..., \{I_N\}\}$.
- Die Klassen $C_i, C_j \in C^{\nu-1}$ mit dem kleinsten Abstand werden fusioniert. Man erhält eine neue Hierarchie C^ν.
- Wenn $C^\nu = \{I\}$ (nur noch eine Klasse): Ende der Klassifikation.

Der Indexwert der Fusionierung wird durch

Indexwert der Fusionierung

$$h_\nu = D_\nu = \min_{k \neq j, C_k, C_j \in C^{\nu-1}} D(C_k, C_j) \tag{7.33}$$

oder $\max S(C_k, C_j)$ angegeben. Es handelt sich also um den minimalen Klassenabstand aller Klassen in der Partition $C^{\nu-1}$.

Um den Algorithmus praktisch durchführen zu können, müssen Abstände zwischen den Klassen definiert werden. Da die Klassen Objekte $I_n \in I$ enthalten, können die Klassenabstände auf Objektabstände zurückgeführt werden. In folgenden werden die Methoden

- Single-Linkage (minimale Distanz)
- Complete-Linkage (maximale Distanz)
- Average-Linkage (durchschnittliche Distanz)
- Zentroid-Verfahren

behandelt.

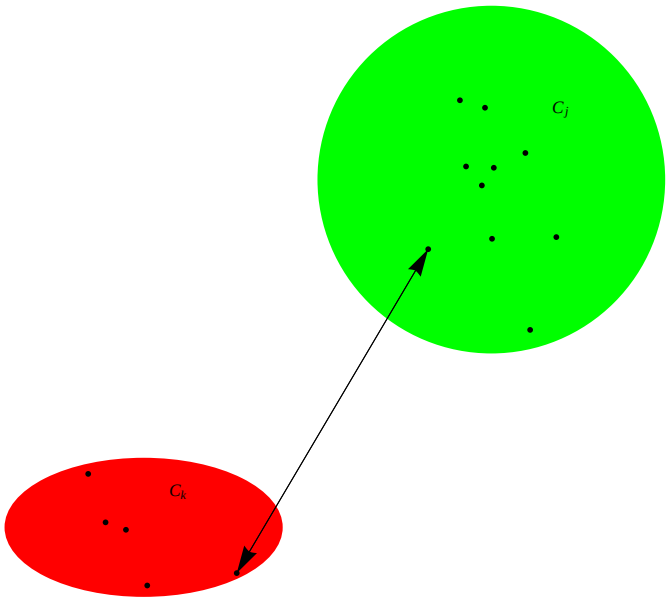


Abbildung 7.7: Abstand von 2 Klassen beim single-linkage-Verfahren.

7.4.2.1 Single-Linkage-Methode

Hierbei wird der Abstand zwischen 2 Klassen als der minimale Abstand der darin enthaltenen Objekte definiert, also

$$D(C_k, C_j) = \min_{n \in C_k, m \in C_j} d_{nm} \tag{7.34}$$

(vgl. Abb. 7.7). Die Fusion findet zwischen den Klassen mit minimalem Abstand und Indexwert

$$h_\nu = D_\nu = \min_{k \neq j, C_k, C_j \in C^{\nu-1}} D(C_k, C_j) = \min_{k \neq j} \min_{n,m} d_{nm} \tag{7.35}$$

statt.

Beispiel 7.4 (Single-Linkage-Methode)

Startpartition:
$C^0 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$

Distanzmatrix:

$$\left[\begin{array}{c|cccccc} & \{1\} & \{2\} & \{3\} & \{4\} & \{5\} & \{6\} \\ \hline \{1\} & 0 & \mathbf{2} & 6.25 & 13 & 9 & 29 \\ \{2\} & 2 & 0 & 13.25 & 13 & 5 & 17 \\ \{3\} & 6.25 & 13.25 & 0 & 9.25 & 15.25 & 45.25 \\ \{4\} & 13 & 13 & 9.25 & 0 & 4 & 20 \\ \{5\} & 9 & 5 & 15.25 & 4 & 0 & 8 \\ \{6\} & 29 & 17 & 45.25 & 20 & 8 & 0 \end{array}\right]$$

1. Fusion: Klassen $\{1\}, \{2\}$, Indexwert $h_1 = h(\{1\}, \{2\}) = 2$

$C^1 = \{\{1, 2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$

Folgetableau:

$$\left[\begin{array}{c|ccccc} & \{1,2\} & \{3\} & \{4\} & \{5\} & \{6\} \\ \hline \{1,2\} & 0 & 6.25 & 13 & 5 & 17 \\ \{3\} & 6.25 & 0 & 9.25 & 15.25 & 45.25 \\ \{4\} & 13 & 9.25 & 0 & \mathbf{4} & 20 \\ \{5\} & 5 & 15.25 & 4 & 0 & 8 \\ \{6\} & 17 & 45.25 & 20 & 8 & 0 \end{array}\right]$$

2. Fusion: Klassen $\{4\}, \{5\}$, $h_2 = h(\{4\}, \{5\}) = 4$

$C^2 = \{\{1, 2\}, \{3\}, \{4, 5\}, \{6\}\}$

$$\left[\begin{array}{c|cccc} & \{1,2\} & \{3\} & \{4,5\} & \{6\} \\ \hline \{1,2\} & 0 & 6.25 & \mathbf{5} & 17 \\ \{3\} & 6.25 & 0 & 9.25 & 45.25 \\ \{4,5\} & 5 & 9.25 & 0 & 8 \\ \{6\} & 17 & 45.25 & 8 & 0 \end{array}\right]$$

3. Fusion:
$C^3 = \{\{1, 2, 4, 5\}, \{3\}, \{6\}\}$, $h_3 = h(\{1, 2\}, \{4, 5\}) = 5$

$$\left[\begin{array}{c|ccc} & \{1,2,4,5\} & \{3\} & \{6\} \\ \hline \{1,2,4,5\} & 0 & \mathbf{6.25} & 8 \\ \{3\} & 6.25 & 0 & 45.25 \\ \{6\} & 8 & 45.25 & 0 \end{array}\right]$$

4. Fusion:
$C^4 = \{\{1, 2, 4, 5, 3\}, \{6\}\}$, $h_4 = h(\{1, 2, 4, 5\}, \{3\}) = 6.25$

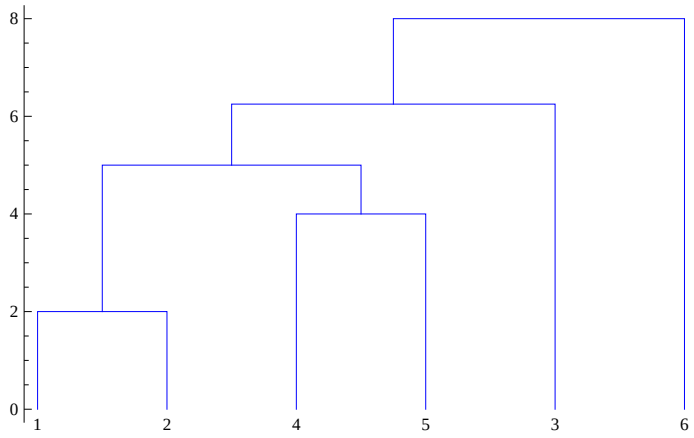


Abbildung 7.8: Dendrogramm beim single linkage-Verfahren.

$$\left[\begin{array}{c|cc} & \{1,2,4,5,3\} & \{6\} \\ \hline \{1,2,4,5,3\} & 0 & \mathbf{8} \\ \{6\} & 8 & 0 \end{array}\right]$$

5. Fusion (Ende):
$C^5 = \{\{1,2,4,5,3,6\}\}$, $h_4 = h(\{1,2,4,5,3\},\{6\}) = 8$

$$\left[\begin{array}{c|c} & \{1,2,4,5,3,6\} \\ \hline \{1,2,4,5,3,6\} & 0 \end{array}\right]$$

Das Dendrogramm der Indexwerte h_ν ist in Abb. 7.8 dargestellt.
■

Bemerkungen:

- Wenn h nicht berechnet wird, reicht die Matrix der Distanz-Rangordnungen zur Berechnung des Dendrogramms aus.
- Da nur der Minimalabstand zwischen den Klassen eine Rolle spielt, werden Klassen auch dann verbunden, wenn die anderen Objekte weit voneinander entfernt liegen (Verkettungseigenschaft).

Dies führt auf die Idee, den maximalen Abstand oder den Mittelwert aller Entfernungen zwischen den Objekten in beiden Klassen zur Grundlage der Klassendistanz zu machen.

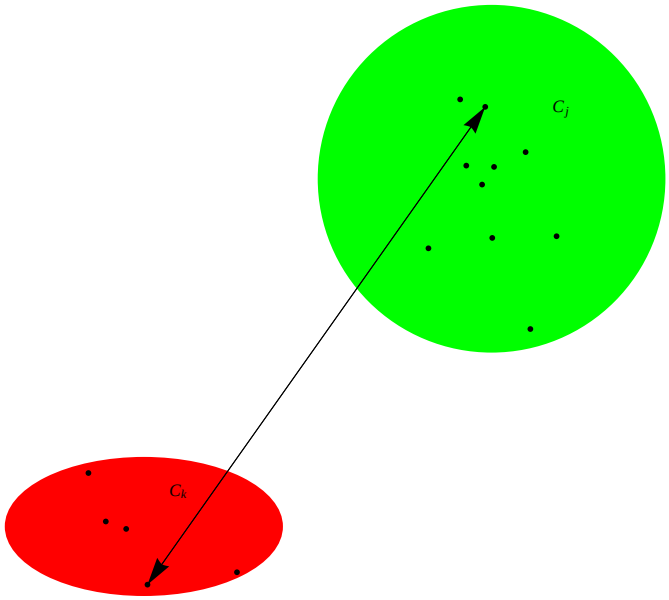


Abbildung 7.9: Abstand von 2 Klassen beim complete-linkage-Verfahren.

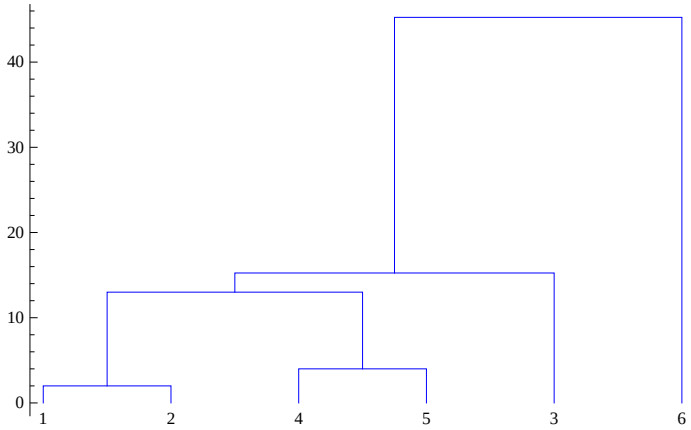


Abbildung 7.10: Complete-linkage-Verfahren. Dendrogramm aus Übung 7.4.

7.4.2.2 Complete-Linkage-Methode

Hierbei wird der Abstand zwischen 2 Klassen als der maximale Abstand der darin enthaltenen Objekte definiert, also

$$D(C_k, C_j) = \max_{n \in C_k, m \in C_j} d_{nm} \tag{7.36}$$

(vgl. Abb. 7.9). Die Fusion findet wieder zwischen den Klassen mit minimalem Abstand und zugehörigem Indexwert

$$h_\nu = D_\nu = \min_{k \neq j, C_k, C_j \in C^{\nu-1}} D(C_k, C_j) = \min_{k \neq j} \max_{n,m} d_{nm} \tag{7.37}$$

statt.

Übung 7.4 (Complete-Linkage-Methode)
Berechnen Sie die Partitionen, Indexwerte und Folgetableaus wie im vorigen Beispiel und zeichnen Sie das Dendrogramm.
■

7.4.2.3 Average-Linkage-Methode

Hierbei wird der Abstand zwischen 2 Klassen als der durchschnittliche Abstand der darin enthaltenen Objekte definiert, also

$$D(C_k, C_j) = \frac{1}{n_k n_j} \sum_{n \in C_k, m \in C_j} d_{nm} \tag{7.38}$$

(vgl. Abb. 7.11). Die Fusion findet wieder zwischen den Klassen mit minimalem Abstand und zugehörigem Indexwert

$$h_\nu = D_\nu = \min_{k \neq j, C_k, C_j \in C^{\nu-1}} D(C_k, C_j) = \min_{k \neq j} \frac{1}{n_k n_j} \sum_{n,m} d_{nm} \tag{7.39}$$

statt. Wie die Abbildung zeigt, ist die Berechnung (von Hand) relativ aufwendig.
Da die meisten Programme die Originaldaten benötigen, ist auf der CD ein Datensatz enthalten, der die Distanzmatrix ergibt.[1]. Der von SPSS erzeugte output ist in Abb. 7.13 zu sehen. Man sieht, daß die Distanzmatrix aus den quadrierten euklidischen Distanzen besteht. Außerdem wird die Fusionierungsgeschichte sowie die Indexwerte tabelliert.

[1] `Cluster-Bsp(Fahrmeir, S. 460).jmp` oder `Cluster-Bsp(Fahrmeir, S. 460).sav`

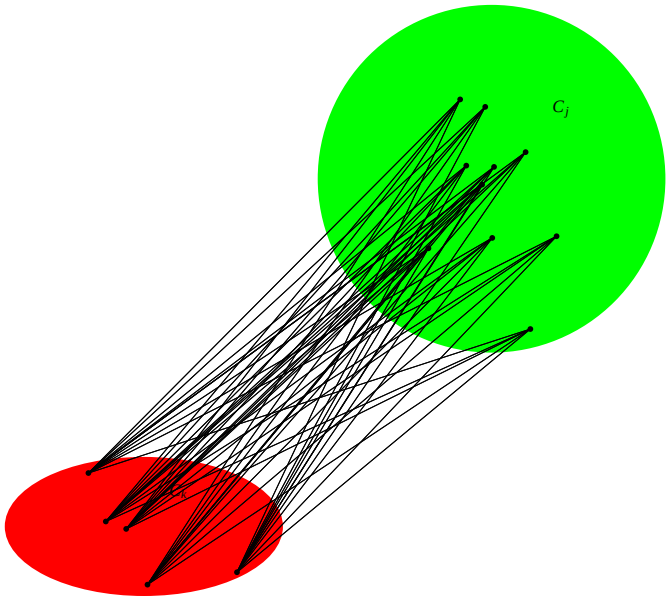


Abbildung 7.11: Abstand von 2 Klassen beim average-linkage-Verfahren.

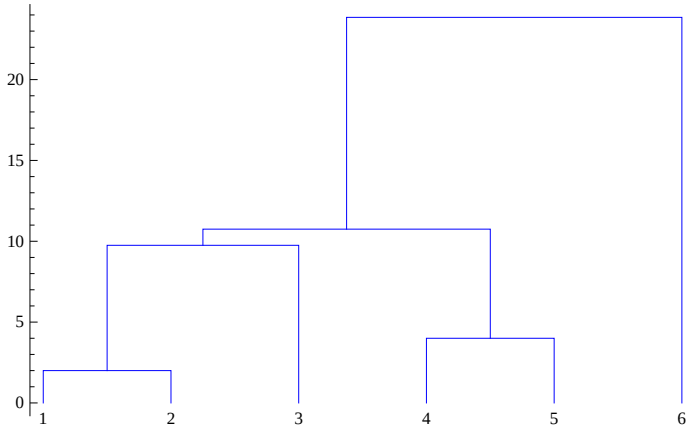


Abbildung 7.12: Dendrogramm beim average-linkage-Verfahren.

Proximity Matrix

Case	Squared Euclidean Distance					
	1	2	3	4	5	6
1	,000	2,000	6,250	13,000	9,000	29,000
2	2,000	,000	13,250	13,000	5,000	17,000
3	6,250	13,250	,000	9,250	15,250	45,250
4	13,000	13,000	9,250	,000	4,000	20,000
5	9,000	5,000	15,250	4,000	,000	8,000
6	29,000	17,000	45,250	20,000	8,000	,000

This is a dissimilarity matrix

Average Linkage (Between Groups)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	2,000	0	0	3
2	4	5	4,000	0	0	4
3	1	3	9,750	1	0	4
4	1	4	10,750	3	2	5
5	1	6	23,850	4	0	0

Dendrogram using Average Linkage (Between Groups)

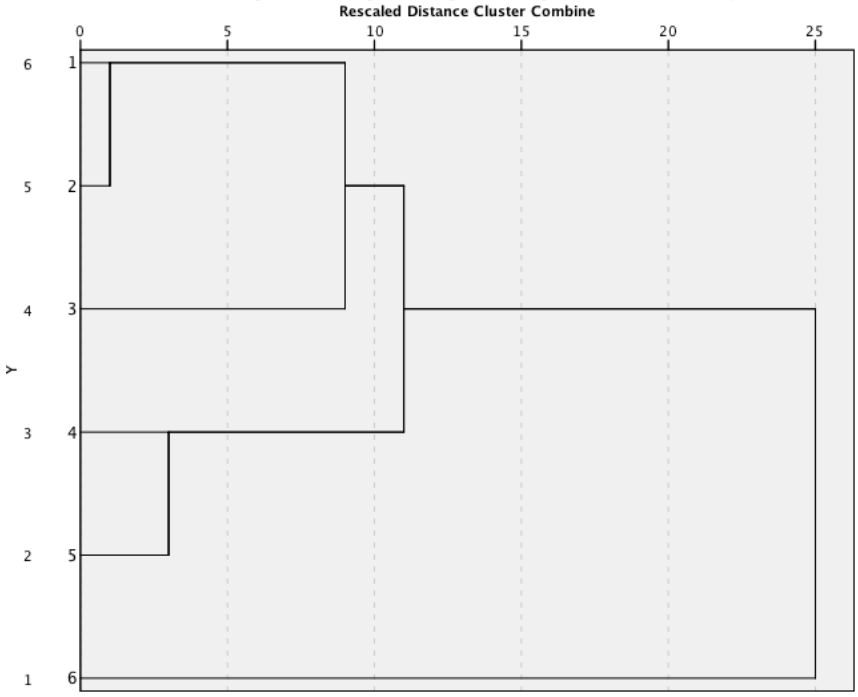


Abbildung 7.13: SPSS: Output und Dendrogramm beim average-linkage-Verfahren.

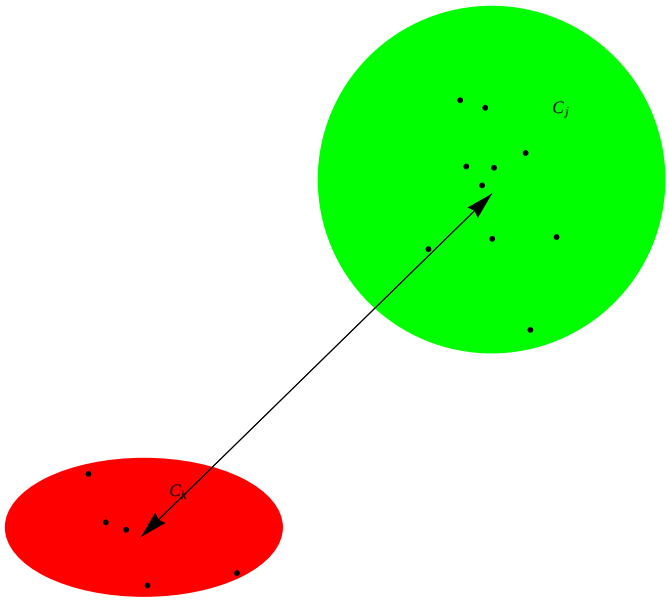


Abbildung 7.14: Abstand von 2 Klassen beim Zentroid-Verfahren.

7.4.2.4 Zentroid-Methode

Mit der Average-Linkage-Methode ist die sog. Zentroid-Methode verwandt, bei der die Klassenabstände als Differenzen der Klassenmittelwerte (Klassenschwerpunkte) berechnet werden. Man definiert den Abstand der Klassen C_k, C_j als

$$D(C_k, C_j) = ||\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_j||^2 \tag{7.40}$$

(quadrierte euklidische Distanz) wobei $\bar{\mathbf{x}}_k = \frac{1}{n_k} \sum_{n \in C_k} \mathbf{x}_n$ der Klassenmittelwert ist (Abb. 7.14).
Aufgrund der Definition gilt das Zentroid-Verfahren

- nur für metrische Variablen
- ist die Indexbedingung $h(C) \leq h(B) \Leftrightarrow C \subseteq B$ nicht unbedingt erfüllt, d.h. es kann vorkommen, daß die fusionierte Klasse $C_v \cup C_w$ homogener ist als C_v oder C_w.

  Es kann also gelten $C_v \subseteq (C_v \cup C_w) \Rightarrow h(C_v) \geq h(C_v \cup C_w)$ (entsprechend für C_w). Dies wird als Inversion bezeichnet (vgl. Abb. 7.15).

Inversion

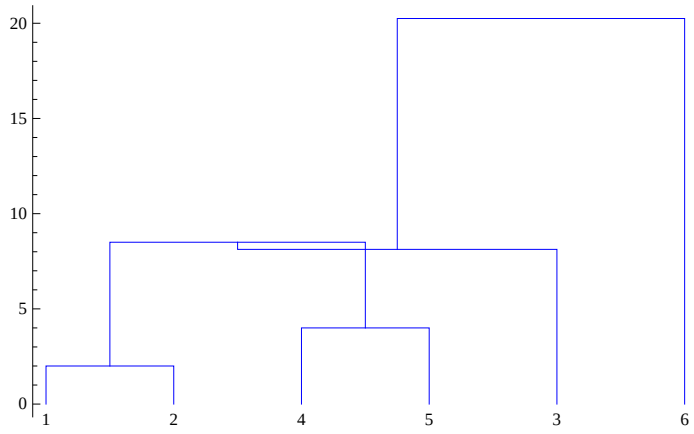


Abbildung 7.15: Dendrogramm beim Zentroid-Verfahren. Inversion: nach der Fusion mit 3 sinkt der Indexwert von 8.5 auf 8.125.

Der Zusammenhang mit der average-linkage-Methode (avl) wird durch folgende Umrechnung transparent (Distanzmaß = quadrierte euklidische Distanz):

$$
\begin{aligned}
D_{avl}(C_k, C_j) &= \frac{1}{n_k n_j} \sum_{n \in C_k, m \in C_j} ||\mathbf{x}_n - \mathbf{x}_m||^2 && (7.41)\\
&= \frac{1}{n_k n_j} \sum_{n,m} ||(\mathbf{x}_n - \bar{\mathbf{x}}_k) + (\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_j) - (\mathbf{x}_m - \bar{\mathbf{x}}_j)||^2 \\
&= ||\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_j||^2 + \frac{1}{n_j} \sum_m ||\mathbf{x}_m - \bar{\mathbf{x}}_j||^2 + \frac{1}{n_k} \sum_n ||\mathbf{x}_n - \bar{\mathbf{x}}_k||^2 \\
&= D_{Zentr}(C_k, C_j) + s_j^2 + s_k^2 && (7.42)
\end{aligned}
$$

Dies bedeutet, daß bei der average-linkage-Methode zusätzlich zum minimalen Abstand der Klassenmittelwerte (Zentroid) auch noch minimale Varianzen s_j^2, s_k^2 in den Klassen C_j, C_k verlangt werden. Die Klassen, die fusioniert werden, müssen also auch noch *kompakt* sein.

In obiger Umrechnung wurde benutzt, daß die gemischten Terme in $\sum_{n,m} ||\mathbf{a}_{nk} + \mathbf{b}_{kj} - \mathbf{c}_{mj}||^2 = \sum_{n,m} ||\mathbf{a}_{nk}||^2 + ||\mathbf{b}_{kj}||^2 + ||\mathbf{c}_{mj}||^2 + 2\mathbf{a}'_{nk}\mathbf{b}_{kj} - 2\mathbf{a}'_{nk}\mathbf{c}_{mj} - 2\mathbf{b}'_{kj}\mathbf{c}_{mj}$ verschwinden (vgl. Glg. 7.28). Z.B. gilt

$$
\begin{aligned}
\sum_{n,m} \mathbf{a}'_{nk}\mathbf{b}_{kj} &= \sum_{n,m} (\mathbf{x}_n - \bar{\mathbf{x}}_k)'(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_j) \\
&= \sum_m (n_k\bar{\mathbf{x}}_k - n_k\bar{\mathbf{x}}_k)'(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_j) = 0.
\end{aligned}
$$

Außerdem ist die quadrierte euklidische Norm ein Skalarprodukt $||\mathbf{x}||^2 = \mathbf{x}'\mathbf{x}$. Man hat z.B. $||\mathbf{x} - \mathbf{y}||^2 = \mathbf{x}'\mathbf{x} + \mathbf{y}'\mathbf{y} - 2\mathbf{x}'\mathbf{y}$ etc.

Der Stoff wird in den Aufgaben 7.1 und 7.2 vertieft.

Beispiel 7.5 (OECD-Daten)

In Abb. 7.5 wurde das Dendrogramm der OECD-Länder bzgl. der 4 Variablen `LifeSatisfaction`, `Employmentrate`, `Employmentrateofwomenwithchildren` und `Airpollution` gezeigt. Als Abstandsmaß wurde die single-linkage-Methode verwendet. Es ist auch instruktiv, die Objekte zusätzlich im Variablenraum darzustellen. In den Abb. 7.16–7.18 wird die single-linkage, average-linkage und die Zentroid-Methode verglichen. Letztere Verfahren ergeben sehr ähnliche Resultate. Im Dendrogramm der Zentroid-Methode ist auch eine Inversion zu erkennen. Die ersten beiden Cluster der average-linkage und Zentroid-Methode sind im single linkage-Verfahren zu einem großen Cluster (rot) zusammengefaßt. Dies ist wieder ein Ausdruck der Verkettungseigenschaft dieses Verfahrens.

Das doppelte Dendogramm in den Abbildungen 7.16 – 7.18 (oben) zeigt zusätzlich die Ähnlichkeit der 4 Variablen. Die beiden Arbeitslosigkeits-Variablen werden zuerst fusioniert, dann LifeSatisfaction und zuletzt Airpollution. Die Farbcodierung der Variablen zeigt einen Bereich von rot (hohe Werte) über grau zu niedrigen Ausprägungen (blau).

■

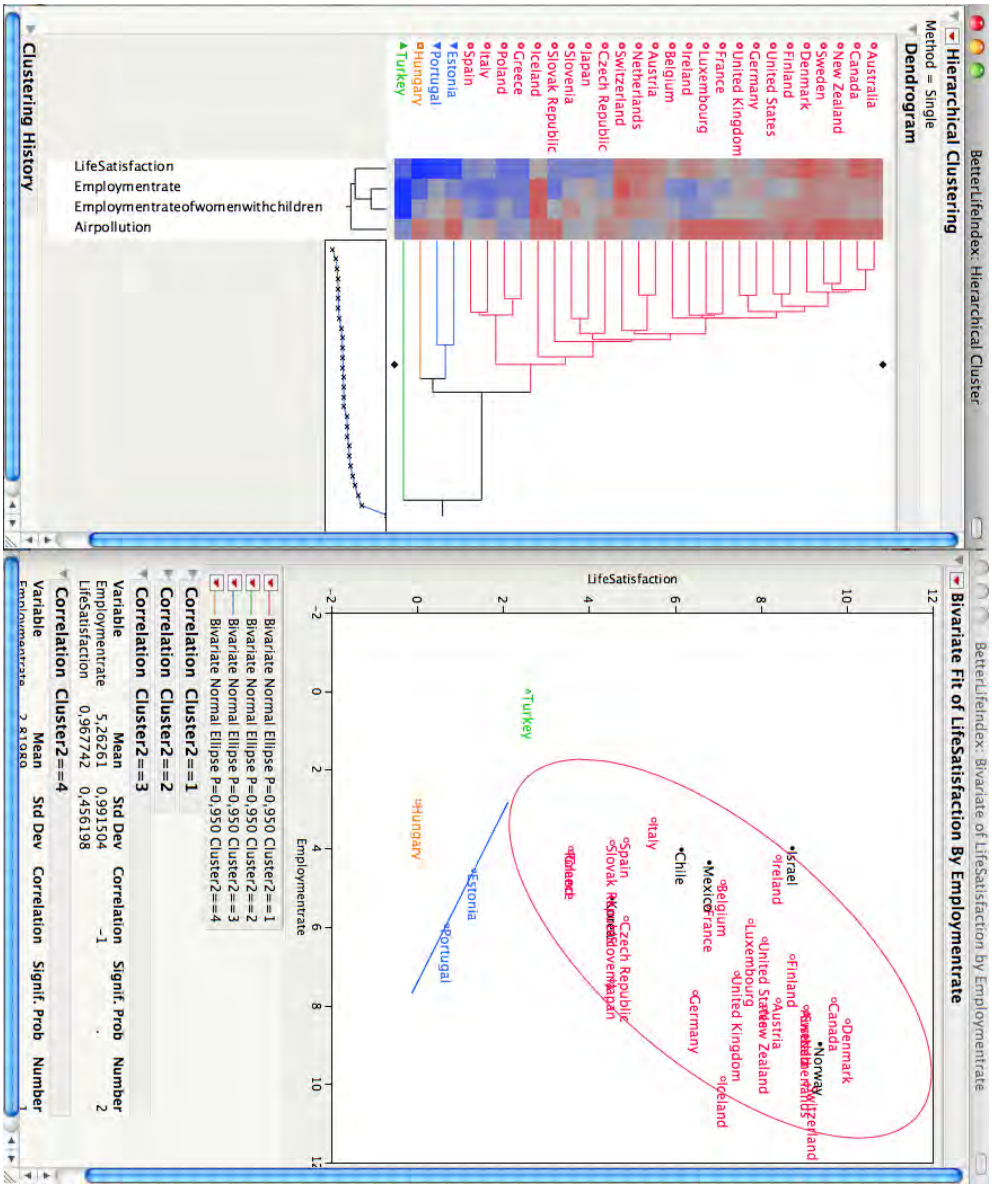


Abbildung 7.16: Single-linkage: 2-faches Dendrogramm der OECD-Länder und von 4 Variablen (links). 95%-Ellipsen der Cluster im Variablenraum (rechts).

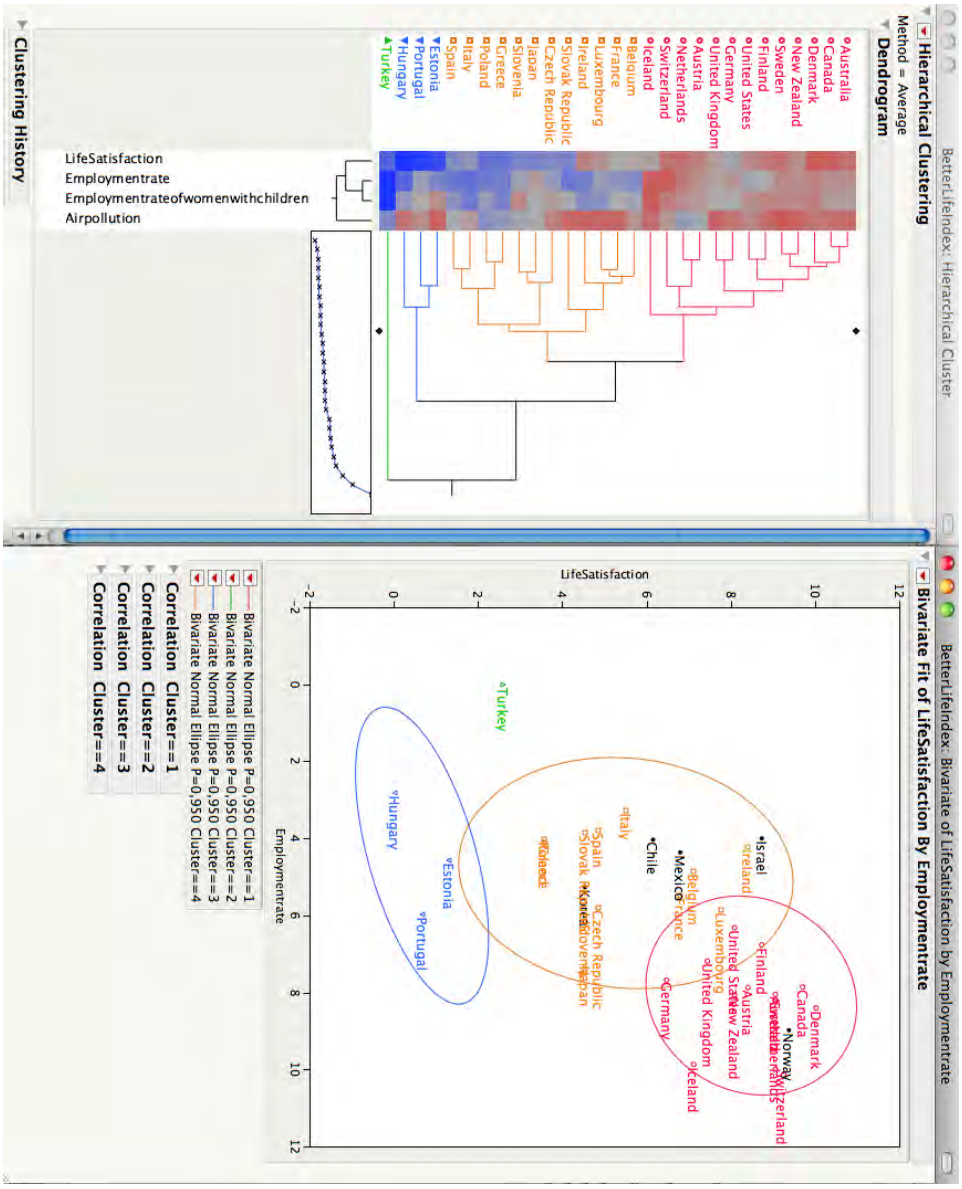


Abbildung 7.17: Average-linkage: 2-faches Dendrogramm der OECD-Länder und von 4 Variablen (links). 95%-Ellipsen der Cluster im Variablenraum (rechts).

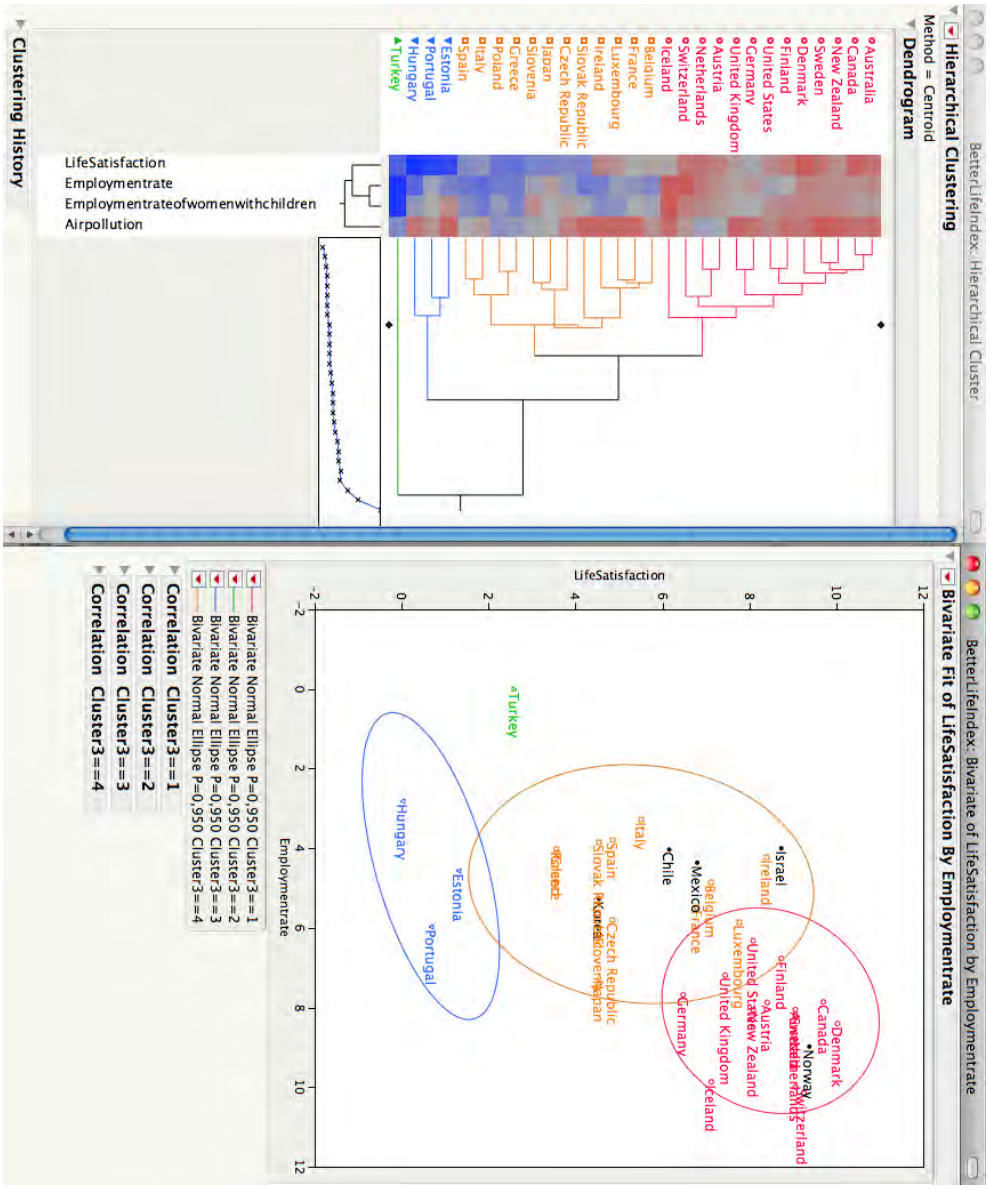


Abbildung 7.18: Zentroid: 2-faches Dendrogramm der OECD-Länder und von 4 Variablen (links). 95%-Ellipsen der Cluster im Variablenraum (rechts).

7.4.3 Cluster der Variablen

Variablen als Objekte

Bisher wurden die Objekte (Zeilen der Datenmatrix) gruppiert. Man kann aber auch, wie schon im letzten Beispiel sowie der Einleitung (Mineralwässer) erwähnt, die Spalten der Datenmatrix (Variablen) gruppieren. Man erhält so eine Übersicht, welche der Variablen in einem Datensatz ähnlich oder unähnlich sind. Natürlich läßt sich dies auch aus einer Korrelationsmatrix ablesen, wenn man die Korrelation als Ähnlichkeit interpretiert. Die Zeilen und Spalten können dann so umgeordnet werden, daß hochkorrelierte Variablen benachbart sind (vgl. Abb. 1.9). Andererseits kann die Korrelationsmatrix als Ähnlichkeitsmatrix benutzt werden und als Grundlage eines Dendrogramms dienen. In Abb. (7.19, oben) ist dies gezeigt (average linkage).

Alternativ kann auch die quadrierte euklidische Distanz der Spalten $\mathbf{x}_{(i)}, \mathbf{x}_{(j)}$ als Distanzmaß dienen. Man erhält

$$\begin{aligned} ||\mathbf{x}_{(i)} - \mathbf{x}_{(j)}||^2 &= (\mathbf{x}_{(i)} - \mathbf{x}_{(j)})'(\mathbf{x}_{(i)} - \mathbf{x}_{(j)}) && (7.43)\\ &= \mathbf{x}'_{(i)}\mathbf{x}_{(i)} + \mathbf{x}'_{(j)}\mathbf{x}_{(j)} - 2\mathbf{x}'_{(i)}\mathbf{x}_{(j)} && (7.44)\\ &= \sum_n (x_{ni}^2 + x_{nj}^2 - 2x_{ni}x_{nj}). && (7.45) \end{aligned}$$

Nimmt man an, daß die Variablen standardisiert sind (d.h. $\bar{x}_i = N^{-1}\sum_n x_{ni} = 0$ und $s_i^2 = (N-1)^{-1}\sum_n x_{ni}^2 = 1$, so ist die quadrierte euklidische Distanz

$$||\mathbf{x}_{(i)} - \mathbf{x}_{(j)}||^2 = d_{ij}^2 = 2(1 - r_{ij}) \qquad (7.46)$$

durch den (negativen) Korrelationskoeffizienten gegeben. Es handelt sich um eine monoton fallende Transformation des Ähnlichkeitsmaßes r_{ij} in ein Distanzmaß $d_{ij} = \sqrt{2(1 - r_{ij})}$. Eine perfekte Korrelation ergibt also eine Distanz von 0. Abb. (7.19, unten) zeigt das Dendrogramm der quadrierten euklidischen Distanz der standardisierten Variablen. Wenn die Variablen unstandardisiert verwendet werden, erhält man ein anderes Resultat (Abb. 7.20). In diesem Fall ist die euklidische Distanz auch durch die Mittelwerte und Standardabweichungen $\bar{x}_i, s_i$ sowie r_{ij} bestimmt.

Es ist auch instruktiv, das Dendrogramm mit der Farbmatrix (diagonales Ordnen; Abb. 1.9) zuvergleichen.

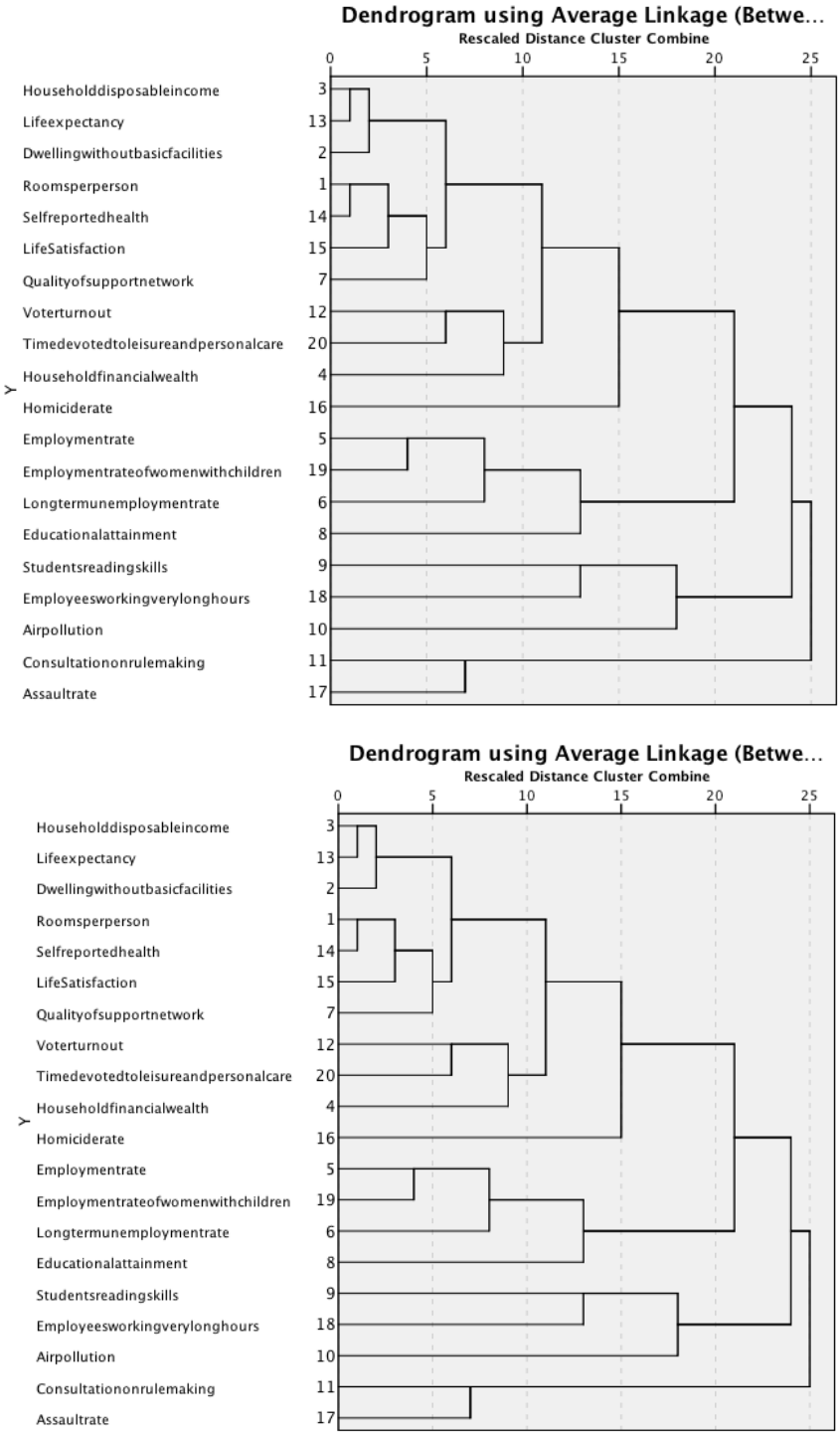


Abbildung 7.19: Dendrogramm der OECD-Variablen. Korrelationen als Ähnlichkeitsmaß (oben), quadrierte euklidische Distanz (standardisierte Variablen, unten).

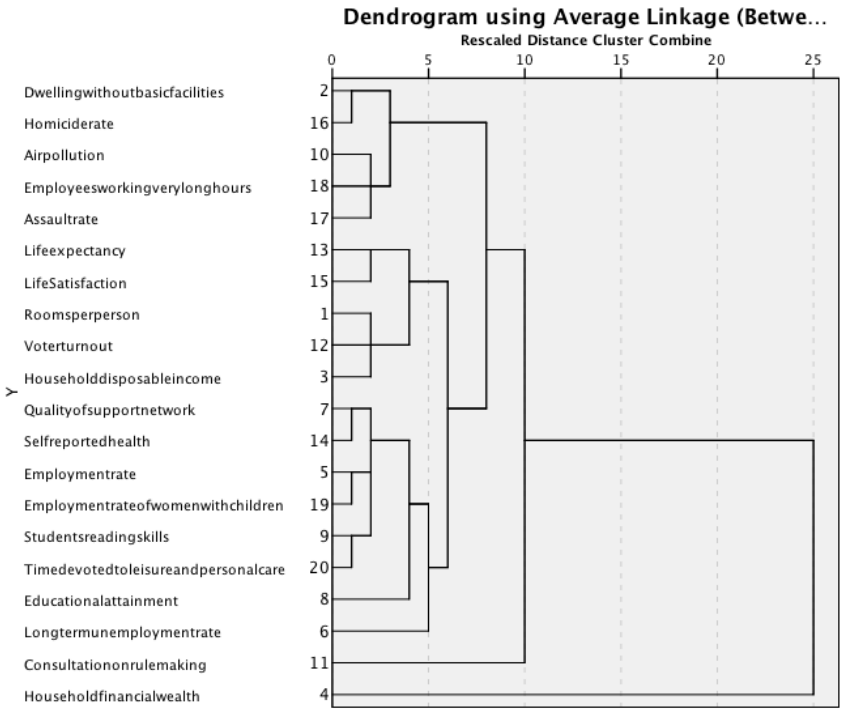


Abbildung 7.20: Dendrogramm der OECD-Variablen. Quadrierte euklidische Distanz (unstandardisiert).

Kapitel 8

Faktoren-Analyse

8.1 Modell

Im letzten Abschnitt wurde die Ähnlichkeit der Variablen im OECD-Datensatz mit Hilfe der Clusteranalyse untersucht. Hierbei diente die Korrelationsmatrix als Ähnlichkeitsmatrix. Hoch korrelierende Variablen wurden zu Clustern zusammengefaßt.

Im Modell der Faktorenanalyse wird versucht, diese Gruppen von ähnlichen Variablen durch latente, nicht beobachtbare Variablen, sog. Faktoren, zu erklären. Das Ziel ist hierbei, die Vielzahl von Variablen (etwa in einem Fragebogen oder Test) durch wenige, zugrundeliegende Variablen zu erklären.

Das Modell der Faktorenanalyse ist durch

Faktorenanalyse-Modell

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon}. \tag{8.1}$$

Faktor-Ladungs-Matrix
Faktoren

gegeben. Die Gewichts-Matrix $\mathbf{\Lambda} : p \times q$ wird als Faktor-Ladungs-Matrix bezeichnet und $\boldsymbol{\xi} : q \times 1$ ist ein Vektor von Faktoren (latenten Variablen, true scores).

Diese können entweder orthogonal ($\mathrm{Var}(\boldsymbol{\xi}) = \mathbf{I}$) oder schiefwinklig sein ($\mathrm{Var}(\boldsymbol{\xi}) = \mathbf{\Phi} \neq \mathbf{I}$).

Die üblichen Annahmen und Bezeichnungen sind:

- $\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\epsilon}) = \mathbf{O} : q \times p$ (Faktoren und Fehler sind orthogonal)

- $E[\boldsymbol{\xi}] = \mathbf{0}; E[\boldsymbol{\epsilon}] = E[\mathbf{x}] = \mathbf{0}$ (zentrierte Variablen)
- $\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Sigma} : p \times p$
- $\mathrm{Var}(\boldsymbol{\xi}) = \boldsymbol{\Phi} : q \times q$

  Falls $\boldsymbol{\Phi} = \mathbf{I}$: Faktoren sind rechtwinklig und standardisiert
- $\mathrm{Var}(\boldsymbol{\epsilon}) = \mathbf{V} : p \times p$ (Fehler-Kovarianz-Matrix).

Es handelt sich beim Faktorenmodell (8.1) trotz der identischen Form **nicht** um ein Regressionsmodell, da $\boldsymbol{\xi}$ nicht beobachtbar ist. Die beobachteten Variablen werden durch drei nichtbeobachtete Größen $\boldsymbol{\Lambda}, \boldsymbol{\xi}, \boldsymbol{\epsilon}$ erklärt, was durchaus ambitioniert ist und auch diverse Mehrdeutigkeiten zur Folge hat.

Die Faktoren (Zufallsvariablen) lassen sich schätzen, wenn man die optimale Prognose (Glg. 2.87)

Faktoren-Schätzung (Thompson)

$$\hat{\boldsymbol{\xi}} = E[\boldsymbol{\xi}|\mathbf{x}] = E[\boldsymbol{\xi}] + \mathrm{Cov}(\boldsymbol{\xi}, \mathbf{x})\mathrm{Cov}(\mathbf{x}, \mathbf{x})^{-1}(\mathbf{x} - E[\mathbf{x}]) \tag{8.2}$$

$$= \mathrm{Cov}(\boldsymbol{\xi}, \mathbf{x})\mathrm{Cov}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{x} \tag{8.3}$$

zugrundelegt (Regressions-Schätzer, Thompson 1951).

Alternativ kann man den GLS (Bartlett)-Schätzer[1]

Faktoren-Schätzung (Bartlett)

$$\hat{\boldsymbol{\xi}} = (\boldsymbol{\Lambda}'\mathbf{V}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\mathbf{V}^{-1}\mathbf{x}. \tag{8.4}$$

verwenden. Dieser ergibt sich aus dem Grundmodell

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon}. \tag{8.5}$$

durch Multiplizieren mit $\mathbf{V}^{-1/2}$ und Berechnen des KQ-Schätzers (vgl. Mardia et al., Kap. 9.7).

Setzt man noch

[1] GLS = generalized least squares

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\xi}, \mathbf{x}) &= \mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\Lambda}\boldsymbol{\xi}) = \mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi})\boldsymbol{\Lambda}' = \boldsymbol{\Phi}\boldsymbol{\Lambda}' && (8.6)\\
\mathrm{Cov}(\mathbf{x}, \mathbf{x}) &= \boldsymbol{\Lambda}\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi})\boldsymbol{\Lambda}' + \mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}) && (8.7)\\
&= \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{V} = \boldsymbol{\Sigma} && (8.8)
\end{aligned}$$

in (8.3) ein, so können die Faktorwerte als lineare Regression der beobachtbaren Werte $\mathbf{x}$ ausgedrückt werden.[2] Der geschätzte Faktor-Score ist also ein gewichtetes Mittel

$$\hat{\boldsymbol{\xi}} = E[\boldsymbol{\xi}|\mathbf{x}] = \boldsymbol{\Phi}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{V})^{-1}\mathbf{x} \tag{8.10}$$

der items.

Fundamentaltheorem

Die Varianzzerlegung (8.8) wird etwas hochtrabend als Fundamentaltheorem der Faktorenanalyse bezeichnet.

Die Modellparameter $(\boldsymbol{\Phi}, \boldsymbol{\Lambda}, \mathbf{V})$ sind unbekannt und müssen geschätzt werden.

8.2 Hauptkomponentenanalyse I

Die Hauptkomponentenanalyse ist eine deskriptive Methode, bei der die beobachtete Kovarianz- oder Korrelationsmatrix so zerlegt wird, daß eine geringere Zahl $q < p$ von orthogonalen Faktoren ausreicht, um die Beobachtungen zu erklären. Sind etwa 2 Variablen stark korreliert, so genügt im Grunde *eine* Dimension, um deren Variabilität zu beschreiben (Abb. 8.2). Aus dem Faktorenmodell

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{8.11}$$

ergibt sich die Kovarianzmatrix (Streuungszerlegung)

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \mathbf{V} \tag{8.12}$$

[2] Mit Hilfe der Formel

$$\boldsymbol{\Phi}\boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{V})^{-1} = (\boldsymbol{\Lambda}'\mathbf{V}^{-1}\boldsymbol{\Lambda} + \boldsymbol{\Phi}^{-1})^{-1}\boldsymbol{\Lambda}'\mathbf{V}^{-1} \tag{8.9}$$

ergibt sich eine analoge Form zum Bartlett-Schätzer. Für eine nichtinformative Prior-Verteilung der Faktoren, d.h. $\boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Phi} \to \infty)$ geht der Thompson-Schätzer in den Bartlett-Schätzer über.

der beobachtbaren Variablen. Hier wurde die Orthogonalität von $\boldsymbol{\xi}$ und $\boldsymbol{\epsilon}$ sowie die Orthonormalität von $\boldsymbol{\xi}$, d.h. $\mathrm{Var}(\boldsymbol{\xi}) = \boldsymbol{\Phi} = \mathbf{I}$, ausgenutzt.

Daher setzt sich $\boldsymbol{\Sigma}$ aus dem Quadrat der sogenannten Ladungsmatrix $\boldsymbol{\Lambda} : p \times q$ (Gewichtsmatrix der Faktoren) und einer Fehlermatrix $\mathbf{V}$ zusammen.

Die Diagonale der von den Faktoren erklärten Varianz wird als Kommunalität

Kommunalität

$$h_i^2 = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}')_{ii} = \sum_{j=1}^{q} \lambda_{ij}^2 \tag{8.13}$$

Einzelvarianz

bezeichnet, während die Diagonale $[v_{11}, \ldots, v_{pp}]$ von $\mathbf{V}$ spezifische oder Einzelvarianz heißt. Zusammenfassend gilt also die Zerlegung der Diagonale

$$\begin{aligned} \boldsymbol{\Sigma}_{ii} &= (\boldsymbol{\Lambda}\boldsymbol{\Lambda}')_{ii} + \mathbf{V}_{ii} && (8.14) \\ &= \sum_{j=1}^{q} \lambda_{ij}^2 + v_{ii}. && (8.15) \end{aligned}$$

Der erste Term enthält gemeinsame Ladungen, während der zweite Term nur für Variable i spezifische Anteile enthält.

Hauptachsen-Transformation

Man kennt zunächst nicht die Anzahl q der Faktoren $\boldsymbol{\xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q]'$.

Hier kann man sich die sogenannte Hauptachsen-Transformation zunutze machen. Es gilt für jede symmetrische Matrix die Eigenwert- oder Spektralzerlegung

Eigenwertzerlegung

$$\begin{aligned} \boldsymbol{\Sigma} &= \sum_{i=1}^{p} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' = \mathbf{PMP}' && (8.16) \\ \mathbf{P} &= [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_p] : p \times p && (8.17) \\ \mathbf{P}' &= \mathbf{P}^{-1} \text{ (orthogonale Matrix)} && (8.18) \\ \mathbf{M} &= \mathrm{Diag}(\mu_1, \ldots, \mu_p) : p \times p \text{ (Diagonalmatrix)} && (8.19) \end{aligned}$$

wobei die reellen Zahlen $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_p$ als Eigenwerte und die Vektoren $\boldsymbol{\psi}_i$ als Eigenvektoren bezeichnet werden. Es handelt sich dabei um Vektoren

Eigenwerte
Eigenvektoren

$$\boldsymbol{\Sigma}\boldsymbol{\psi}_i = \mu_i \boldsymbol{\psi}_i, \tag{8.20}$$

deren Richtung durch $\boldsymbol{\Sigma}$ nicht verändert wird (Matrizen drehen Vektoren).

Die Eigenwertzerlegung hat *zunächst keine statistische Interpretation*, sondern gilt ganz allgemein für quadratische Matrizen (siehe Abs. 8.3). Allerdings sind für *positiv definite* Matrizen die *Eigenwerte positiv*, was z.B. bei Kovarianz-Matrizen der Fall ist.

8.3 Mathematischer Einschub: Hauptachsentransformation

Im vorigen Abschnitt wurde die Eigenwertgleichung

Eigenwertgleichung

$$\boldsymbol{\Sigma}\boldsymbol{\psi}_i = \mu_i \boldsymbol{\psi}_i \tag{8.21}$$
$$\boldsymbol{\Sigma}\mathbf{P} = \mathbf{P}\mathbf{M} \tag{8.22}$$

zur Zerlegung der symmetrischen Matrix $\boldsymbol{\Sigma}$ benutzt. Daher muß die Gleichung

$$(\boldsymbol{\Sigma} - \mu_i \mathbf{I})\boldsymbol{\psi}_i = \mathbf{0} \tag{8.23}$$

gelöst werden, was nur für

$$\det(\boldsymbol{\Sigma} - \mu_i \mathbf{I}) = 0 \tag{8.24}$$

zu nichttrivialen Lösungen $\boldsymbol{\psi}_i \neq \mathbf{0}$ führt. Aus der Determinante ergibt sich eine Gleichung p-ten Grades für die Eigenwerte (Säkulargleichung). Da $\mathbf{P}$ die orthonormalen Eigenvektoren $\boldsymbol{\psi}_i$ enthält, gilt

Säkulargleichung

Eigenwertzerlegung

$$\begin{aligned} \mathbf{P}'\mathbf{P} &= \mathbf{P}\mathbf{P}' = \mathbf{I} && (8.25)\\ \mathbf{P}' &= \mathbf{P}^{-1} && (8.26)\\ \boldsymbol{\Sigma} &= \mathbf{P}\mathbf{M}\mathbf{P}' = \sum_{i=1}^{p} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i', && (8.27) \end{aligned}$$

Diagonalisierung

$\mathbf{P} = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_p] : p \times p, \mathbf{M} = \mathrm{Diag}(\mu_1, \ldots, \mu_p) : p \times p$ (Diagonalmatrix). Die Summendarstellung von $\boldsymbol{\Sigma}$ wird als *Eigenwertzerlegung* oder *Spektral-Darstellung* bezeichnet. Man spricht auch von *Diagonalisierung*, da $\mathbf{P}'\boldsymbol{\Sigma}\mathbf{P} = \mathbf{M}$ eine Diagonalmatrix ist.

Die Wichtigkeit dieser Formeln kann gar nicht überschätzt werden. Sie erlauben, eine Matrix als Überlagerung von Projektionen $\boldsymbol{\psi}_i\boldsymbol{\psi}_i'$ auf eindimensionale Unterräume darzustellen, mit den Eigenwerten (Spektrum) als Gewicht.

Ganz allgemein gilt für die Spur (= trace) der Matrix

Spur

$$\sum_i \sigma_{ii} := \mathrm{tr}(\boldsymbol{\Sigma}) = \sum_i \mu_i = \mathrm{tr}(\mathbf{M}), \qquad (8.28)$$

da $\mathrm{tr}(\boldsymbol{\Sigma}) = \mathrm{tr}(\mathbf{P}\mathbf{M}\mathbf{P}') = \mathrm{tr}(\mathbf{M}\mathbf{P}'\mathbf{P}) = \mathrm{tr}(\mathbf{M})$.

Beispiel 8.1 (Eigenwerte einer Korrelationsmatrix)
Für die (theoretische) Korrelationsmatrix[3]

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad (8.29)$$

[3] $\mathbf{P}$ wurde bereits für die Matrix der Eigenvektoren reserviert.

ergeben sich die Eigenwerte aus

$$\det(\begin{bmatrix} 1-\mu & \rho \\ \rho & 1-\mu \end{bmatrix}) = (1-\mu)^2 - \rho^2 \stackrel{!}{=} 0 \tag{8.30}$$

$$\mu_{1,2} = 1 \pm \rho \tag{8.31}$$

Die Summe der Eigenwerte ist also $2 = \mathrm{tr}(\mathbf{R})$ = Summe der Diagonale := Spur = trace. Ganz allgemein gilt

$$\sum_i R_{ii} := \mathrm{tr}(\mathbf{R}) = \sum_i \mu_i = p. \tag{8.32}$$

Die Eigenvektoren ergeben sich aus den Bedingungen

$$(\mathbf{R} - \mu_1 \mathbf{I}_2)\boldsymbol{\psi}_1 = \begin{bmatrix} -\rho & \rho \\ \rho & -\rho \end{bmatrix} \begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8.33}$$

$$(\mathbf{R} - \mu_2 \mathbf{I}_2)\boldsymbol{\psi}_2 = \begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix} \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{8.34}$$

Etwa löst

$$\begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{8.35}$$

$$\begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{8.36}$$

$$\tag{8.37}$$

obige Gleichungen. Das Betrags-Quadrat der Vektoren ist $[1,1][1,1]' = 2, [1,-1][1,-1]' = 2$, sodaß man

$$\boldsymbol{\psi}_1 = \begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = 1/\sqrt{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{8.38}$$

$$\boldsymbol{\psi}_2 = \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = 1/\sqrt{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{8.39}$$

als orthonormierte Eigenvektoren findet.

Übung 8.1

Zeigen Sie, daß $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2$ orthonormiert sind.

■

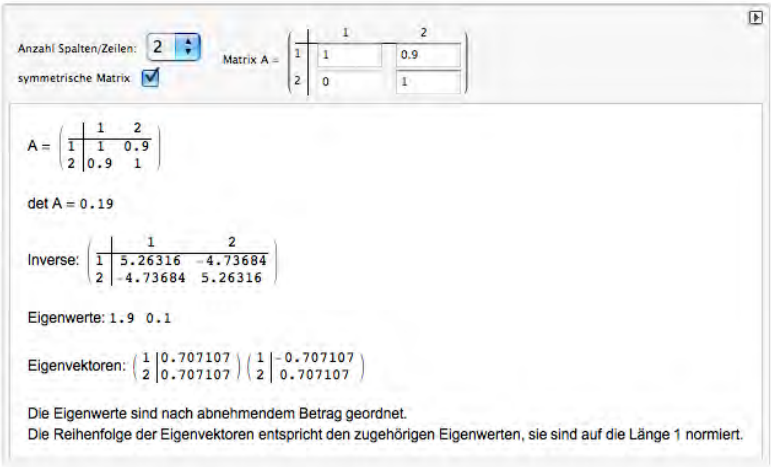


Abbildung 8.1: Applet zur Berechnung von Matrizen.
http://www.fernuni-hagen.de/ls_statistik/lehre/).

Es ist wichtig, daß die Eigenvektoren gar nicht von der Korrelation ρ abhängen. Sie zeigen immer in Richtung der Winkelhalbierenden der Quadranten.

Numerische Berechnungen können mit einem Mathematica-Applet erfolgen (Abb. 8.1)

Abb. 8.2 zeigt simulierte Daten aus einer bivariaten Normalverteilung

$$N\left(\mathbf{0}, \mathbf{R} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right). \tag{8.40}$$

Die Eigenwerte von $\mathbf{R}$ sind $1 \pm 0.9 = 1.9, 0.1$ und die orthogonale Matrix der Eigenvektoren lautet

$$\mathbf{P} = 2^{-1/2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{8.41}$$

$$\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2. \tag{8.42}$$

Im gedrehten Koordinatensystem gilt daher

$$\mathbf{y} = \mathbf{P}'\mathbf{x} = \begin{bmatrix} \boldsymbol{\psi}_1'\mathbf{x} \\ \boldsymbol{\psi}_2'\mathbf{x} \end{bmatrix} = 2^{-1/2} \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix} \tag{8.43}$$

und $\mathrm{Cov}(\mathbf{y}) = \mathbf{P}'\mathbf{R}\mathbf{P} = \mathbf{M} = \mathrm{diag}(1.9, 0.1)$.

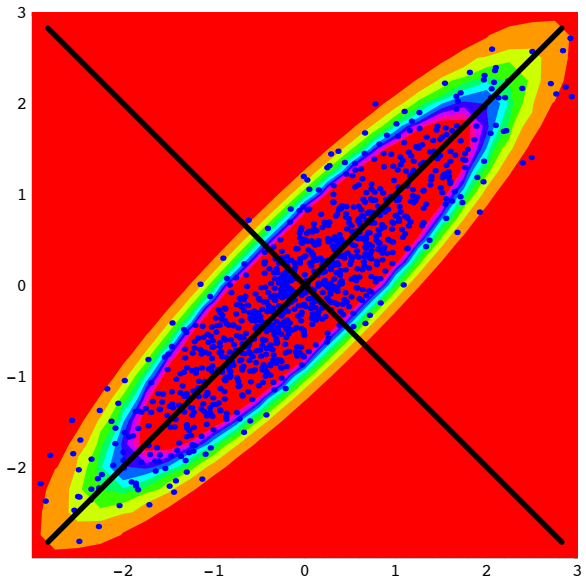


Abbildung 8.2: Simulierte normalverteilte Daten $\mathbf{x}_n, n = 1, ..., N = 1000$ mit Kovarianz-Matrix $\mathbf{R} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$. Die Hauptachsen zeigen in Richtung der Winkelhalbierenden.

Daher sind die gedrehten Koordinaten (Hauptkomponenten) y_1, y_2 unkorreliert.
Die quadratische Form (Ellipse) der Matrix $\mathbf{R}$

$$\mathbf{x}'\mathbf{R}\mathbf{x} = \sum_{ij} x_i \rho_{ij} x_j = x_1^2 + 2\rho x_1 x_2 + x_2^2 \tag{8.44}$$

ist diagonal im gedrehten System:

$$\begin{aligned} \mathbf{x}'\mathbf{R}\mathbf{x} &= \mathbf{x}'\mathbf{P}\mathbf{P}'\mathbf{R}\mathbf{P}\mathbf{P}'\mathbf{x} && (8.45) \\ &= \mathbf{y}'\mathbf{M}\mathbf{y} = \mu_1 y_1^2 + \mu_2 y_2^2 = (1+\rho)y_1^2 + (1-\rho)y_2^2. && (8.46) \end{aligned}$$

Die im Bild gezeigte Ellipse ist allerdings

$$\begin{aligned} \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} &= \mathbf{y}'\mathbf{M}^{-1}\mathbf{y} && (8.47) \\ &= \frac{y_1^2}{\mu_1} + \frac{y_2^2}{\mu_2} && (8.48) \\ &= \frac{y_1^2}{1+\rho} + \frac{y_2^2}{1-\rho} && (8.49) \\ &= \frac{y_1^2}{1.9} + \frac{y_2^2}{0.1}, && (8.50) \end{aligned}$$

da die (multivariat) standardisierte Variable

$$\mathbf{z} = \mathbf{R}^{-1/2}\mathbf{x} = \mathbf{M}^{-1/2}\mathbf{P}'\mathbf{x} \tag{8.51}$$

auf die quadratische Form

$$\mathbf{z}'\mathbf{z} = (\mathbf{M}^{-1/2}\mathbf{P}'\mathbf{x})'\mathbf{M}^{-1/2}\mathbf{P}'\mathbf{x} = \mathbf{x}'\mathbf{P}\mathbf{M}^{-1}\mathbf{P}'\mathbf{x} = \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} \tag{8.52}$$

führt.
Im Exponent der bivariaten Normalverteilung ($\mu_i = 0, \sigma_i^2 = 1$)

bivariate Normalverteilung

$$\begin{aligned} f(x_1, x_2) &= |2\pi\mathbf{R}|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}'\mathbf{R}^{-1}\mathbf{x}\right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right) \end{aligned} \tag{8.53}$$

steht aber genau der Ausdruck $\mathbf{x}'\mathbf{R}^{-1}\mathbf{x} = \mathbf{z}'\mathbf{z}$.
■

Übung 8.2
Zeigen Sie, daß $\det(\mathbf{R}) = |\mathbf{R}| = 1 - \rho^2$ und $\mathbf{R}^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$ gilt.
Berechnen Sie die quadratische Form $\mathbf{x}'\mathbf{R}^{-1}\mathbf{x}$ explizit.
■

8.4 Hauptkomponentenanalyse II

Vergleicht man nun (8.12) mit (8.16), d.h.

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{V} = \sum_{i=1}^{p} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \tag{8.54}$$

so kann man die Teilsumme

$$\mathbf{\Sigma}_1 = \sum_{i=1}^{q} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \tag{8.55}$$

mit $\mathbf{\Lambda\Lambda}'$ identifizieren und den Rest

$$\mathbf{\Sigma}_2 = \sum_{i=q+1}^{p} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \tag{8.56}$$

mit $\mathbf{V}$, d.h. $\mathbf{\Sigma} = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2$.[4] Die Idee ist, daß sich die Matrix $\mathbf{\Sigma}$ gut durch $\mathbf{\Sigma}_1$ approximieren läßt.
In Matrixform ergibt sich

$$\mathbf{\Sigma}_1 = \sum_{i=1}^{q} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \tag{8.57}$$

$$= \mathbf{P}_1 \mathbf{M}_1 \mathbf{P}_1' = \mathbf{\Lambda\Lambda}' \tag{8.58}$$

mit den Teilmatrizen

$$\mathbf{P}_1 = [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_q] = \mathbf{P}[\mathbf{I}_q, \mathbf{O}]' : p \times q \tag{8.59}$$

$$\mathbf{M}_1 = \mathrm{Diag}(\mu_1, \ldots, \mu_q) = \mathbf{M}[\mathbf{I}_q, \mathbf{O}]' : q \times q. \tag{8.60}$$

Setzt man also

$$\mathbf{\Lambda} = \mathbf{P}_1 \mathbf{M}_1^{1/2} \tag{8.61}$$

Faktorladungen (Hauptkomponenten-Analyse)

so ist das Produkt

$$\mathbf{\Lambda\Lambda}' = \mathbf{P}_1 \mathbf{M}_1^{1/2} (\mathbf{P}_1 \mathbf{M}_1^{1/2})' = \mathbf{P}_1 \mathbf{M}_1 \mathbf{P}_1' = \mathbf{\Sigma}_1. \tag{8.62}$$

Somit findet man eine Darstellung der Faktorladungen $\mathbf{\Lambda}$ durch die Eigenwerte und Eigenvektoren der Kovarianzmatrix.
Die optimale Zahl q kann durch Auftragen der Eigenwerte graphisch bestimmt werden (Abb. 8.3). Diese Graphik wird als scree plot bezeichnet. Meistens findet man einige große und viele kleine Eigenwerte, sodaß man am Knick des scree plot die optimale Anzahl von Faktoren (hier $q = 1$ oder $q = 5$?) ablesen kann.

scree plot

[4] Allerdings ist die Residualmatrix $\mathbf{\Sigma}_2$ i.a. nicht diagonal und auch singulär, da sie nur aus $p - q$ Dyaden $\boldsymbol{\psi}_i \boldsymbol{\psi}_i'$ besteht, d.h. $\mathrm{rg}(\mathbf{\Sigma}_2) = p - q$. Aufgrund der Nichtdiagonalität lassen sich keine Einzelvarianzen modellieren, die nur die Variable i beeinflussen.

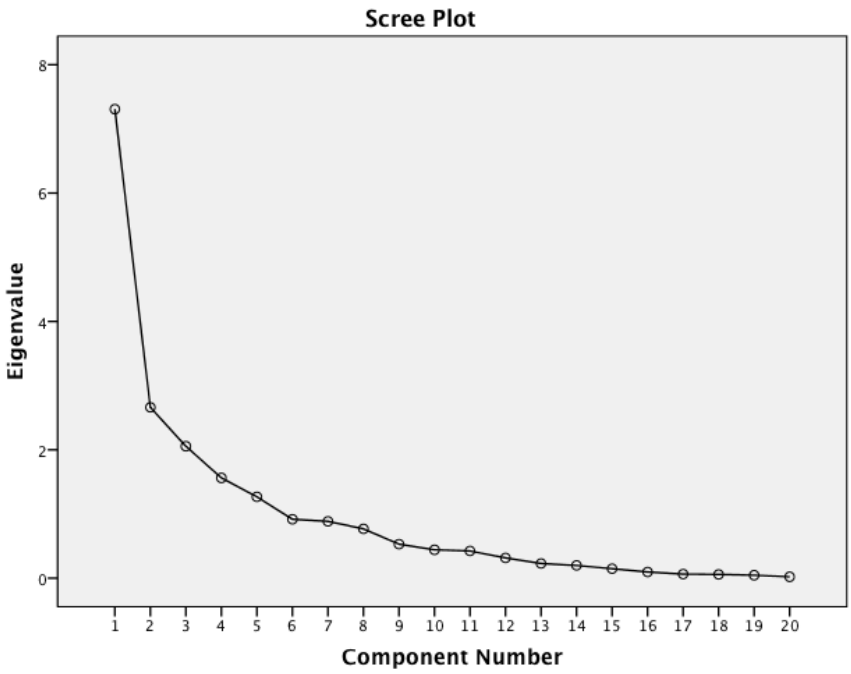


Abbildung 8.3: Scree plot der Eigenwerte (OECD-Daten).

Die geschätzten Faktorwerte ergeben sich durch die optimale lineare Prognose (8.3) nach Thompson. Durch eine einfache Umrechnung ergibt sich für die optimale lineare Prognose die explizite Form

$$
\begin{aligned}
E[\boldsymbol{\xi}|\mathbf{x}] &= \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1}\mathbf{x} && (8.63)\\
&= (\mathbf{M}_1^{1/2}\mathbf{P}_1')(\mathbf{P}\mathbf{M}^{-1}\mathbf{P}')\mathbf{x} && (8.64)\\
&= \mathbf{M}_1^{-1/2}(\mathbf{P}_1'\mathbf{x}) && (8.65)\\
&= \mathbf{M}_1^{-1/2}\mathbf{y}_1 && (8.66)
\end{aligned}
$$

oder

geschätzte Faktorwerte

$$E[\boldsymbol{\xi}|\mathbf{x}] = \mathbf{M}_1^{-1/2}\mathbf{y}_1 = \mathbf{M}_1^{-1/2}(\mathbf{P}_1'\mathbf{x}), \qquad (8.67)$$

da $\mathbf{P}_1'\mathbf{P} = [\mathbf{I}_q, \mathbf{O}] : q \times p$ und $[\mathbf{I}_q, \mathbf{O}]\mathbf{M}^{-1}\mathbf{P}' = [\mathbf{M}_1^{-1}, \mathbf{O}]\mathbf{P}' = \mathbf{M}_1^{-1}\mathbf{P}_1'$. Der Term $\mathbf{y}_1 = \mathbf{P}_1'\mathbf{x}$ ist aber die *Projektion* der Beobachtungen $\mathbf{x}$ auf die

ersten q Eigenvektoren $\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_q$, die sogenannten Hauptkomponenten

Hauptkomponenten

$$\mathbf{y}_1 = \mathbf{P}_1'\mathbf{x} = \begin{bmatrix} \boldsymbol{\psi}_1'\mathbf{x} \\ \vdots \\ \boldsymbol{\psi}_q'\mathbf{x} \end{bmatrix}. \tag{8.68}$$

Die geschätzten Faktoren sind also die mit $1/\sqrt{\mu_i}$ skalierten Hauptkomponenten, d.h. $\hat{\xi}_i = \boldsymbol{\psi}_i'\mathbf{x}/\sqrt{\mu_i}$

Die Skalierung ist nötig, da $\mathbf{y}_1$ nur orthogonal ist, d.h. $\mathrm{Cov}(\mathbf{y}_1) = \mathrm{Cov}(\mathbf{P}_1'\mathbf{x}) = \mathbf{P}_1'\boldsymbol{\Sigma}\mathbf{P}_1 = \mathbf{M}_1$, während die Faktoren sogar orthonormiert sind, d.h.

$$\begin{aligned} \mathrm{Cov}(\mathbf{M}_1^{-1/2}\mathbf{y}_1) &= \mathbf{M}_1^{-1/2}\mathrm{Cov}(\mathbf{y}_1)\mathbf{M}_1^{-1/2} \\ &= \mathbf{M}_1^{-1/2}\mathbf{M}_1\mathbf{M}_1^{-1/2} = \mathbf{I}_q. \end{aligned} \tag{8.69}$$

Die bisherige Argumentation erfolgte auf der Ebene der Kovarianz-Matrizen. Man kann aber auch direkt den Beobachtungsvektor

$$\mathbf{x} = \mathbf{P}\mathbf{y} = [\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_p] \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} \tag{8.70}$$

als Funktion der orthogonalen Hauptkomponenten $\mathbf{y} = \mathbf{P}'\mathbf{x}$ ausdrücken ($\mathrm{Cov}(\mathbf{x}) = \mathbf{P}\mathrm{Cov}(\mathbf{y})\mathbf{P}' = \mathbf{P}\mathbf{M}\mathbf{P}' = \boldsymbol{\Sigma}$) und somit eine Zerlegung

$$\mathbf{x} = \mathbf{P}_1\mathbf{y}_1 + \mathbf{P}_2\mathbf{y}_2 \tag{8.71}$$

$$= \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{8.72}$$

induzieren ($\mathbf{y}_1 = [y_1, ..., y_q]', \mathbf{y}_2 = [y_{q+1}, ..., y_p]'$; entsprechend für $\mathbf{P}$). Dann kann man durch Multiplizieren mit $\mathbf{P}_1'$ und Einsetzen von $\boldsymbol{\Lambda}$ direkt nach der orthonormalen Komponente ($\mathrm{Var}(\boldsymbol{\xi}) = \mathbf{I}_q$)

$$\boldsymbol{\xi} = \mathbf{M}_1^{-1/2}\mathbf{P}_1'\mathbf{x} = \mathbf{M}_1^{-1/2}\mathbf{y}_1 \tag{8.73}$$

auflösen, was mit (8.67) übereinstimmt. Somit sind bei der Hauptkomponenten-Analyse die Faktoren und deren Prognose identisch. In obigen Formeln wurden die Identitäten

$$\mathbf{P}_1'\mathbf{P}_1 = \begin{bmatrix} \boldsymbol{\psi}_1' \\ \vdots \\ \boldsymbol{\psi}_q' \end{bmatrix} [\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_q] = \mathbf{I}_q \tag{8.74}$$

$$\mathbf{P}_1'\mathbf{P}_2 = \begin{bmatrix} \boldsymbol{\psi}_1' \\ \vdots \\ \boldsymbol{\psi}_q' \end{bmatrix} [\boldsymbol{\psi}_{q+1}, \ldots, \boldsymbol{\psi}_p] = \mathbf{O}_{q\times p} \tag{8.75}$$

benutzt.

Bisher wurden die Eigenvektoren und Eigenwerte aus der theoretischen Kovarianzmatrix $\boldsymbol{\Sigma}$ berechnet. In der Praxis muß diese aus Daten $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]' : N \times p$ geschätzt werden (vgl. Abs. 2.3), d.h.

$$\hat{\boldsymbol{\Sigma}} = \mathbf{S} = \frac{1}{N-1}\mathbf{X}'\mathbf{H}\mathbf{X}. \tag{8.76}$$

Die Berechnungen erfolgen dann analog mit den Stichprobengrößen $\mathbf{S} = \hat{\mathbf{P}}\hat{\mathbf{M}}\hat{\mathbf{P}}'$.
Die Eigenvektoren und Eigenwerte sind somit Zufallsgrößen (vgl. Mardia et al., Kap. 8).

Beispiel 8.2 (Hauptkomponentenanalyse für 2 Variablen)
Die Formeln der Hauptkomponentenanalyse werden nun auf die oben diskutierte Korrelationsmatrix angewandt.
Nimmt man $q = 1$ Faktoren, so gilt

$$\boldsymbol{\Lambda} = \boldsymbol{\psi}_1\mu_1^{1/2} = \sqrt{\frac{1+\rho}{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{8.77}$$

Das Produkt

$$\begin{aligned}
\mathbf{\Lambda\Lambda}' &= \frac{1+\rho}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} && (8.78)\\
&= \frac{1+\rho}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} && (8.79)
\end{aligned}$$

ist der erste Teil $\mathbf{\Sigma}_1$ der Varianz $\mathbf{\Sigma}(= \mathbf{R})$.
Die Kommunalitäten sind somit

$$\begin{aligned}
h_i^2 = (\mathbf{\Lambda\Lambda}')_{ii} &= \frac{1+\rho}{2}(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix})_{ii};\ i = 1,2 && (8.80)\\
&= \frac{1+\rho}{2}[1,1] && (8.81)\\
&= [0.95, 0.95] && (8.82)
\end{aligned}$$

im Zahlenbeispiel. Die Gewichte zur Berechnung der Faktoren lauten

$$\begin{aligned}
\boldsymbol{\xi} &= \mathbf{M}_1^{-1/2}\mathbf{P}_1'\mathbf{x} && (8.83)\\
&= (1+\rho)^{-1/2}\boldsymbol{\psi}_1'\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} && (8.84)\\
&= (1+\rho)^{-1/2}\frac{1}{\sqrt{2}}[1,1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} && (8.85)\\
&= \frac{1}{\sqrt{2(1+\rho)}}(x_1 + x_2) && (8.86)\\
&= 0.512989(x_1 + x_2) && (8.87)
\end{aligned}$$

(standardisierte 1. Hauptkomponente).

■

Übung 8.3

Berechnen Sie die Restvarianz $\mathbf{\Sigma}_2 = \mathbf{V} = \mu_2\boldsymbol{\psi}_2\boldsymbol{\psi}_2'$.

■

Der Stoff wird in Aufgabe 8.1 vertieft.

Beispiel 8.3 (OECD-Daten)

In Abb. 8.3 wurden die Eigenwerte der Korrelationsmatrix von 20 Variablen der Größe nach dargestellt. Es zeigt sich, daß einige der Eigenwerte fast 0 sind. Dies hängt damit zusammen, daß der Datensatz nur $N = 32$ Länder enthält. Allerdings sind in vielen Variablen fehlende Werte enthalten. Bei listenweisem Fallausschluß (also nur Länder mit Werten auf allen 20 Variablen) bleibt nur eine Stichprobe der Größe $N' = 16$ übrig. Daher ist die empirische Korrelationsmatrix singulär (sie hat 4 Eigenwerte, die gleich 0 sind) und die Inverse existiert nicht. Man könnte sich auch durch die Analyse einer kleineren Zahl von Variablen behelfen.

Nimmt man paarweisen Fall-Ausschluß, so kommt man auf Fallzahlen zwischen $N' = 25$ und 32, jedoch ist die Korrelationsmatrix nicht positiv semidefinit (es gibt auch negative Eigenwerte). Dann verweigert SPSS die Berechnung der Hauptkomponenten-Analyse (Abb. 8.6, links).

Imputation von fehlenden Werten

Alternativ kann man die fehlenden Werte durch Mittelwerte ersetzen (Abb. 8.6, rechts). Dies ist eine sogenannte Imputationsmethode, die in der Praxis häufig verwendet wird. Diese Option wird im folgenden benutzt. Die entsprechende Korrelationsmatrix ist in den Abb. 8.4-8.5 tabelliert.

In Abb. 8.7 sind die Eigenwerte der Korrelationsmatrix aufgelistet. Nimmt man nur die Eigenwerte $\mu_i > 1$, so erklären diese ca. 74% der Gesamtstreuung ($\mathrm{tr}(\mathbf{\Sigma}) = \sum_i \mu_i, \mathrm{tr}(\mathbf{R}) = p = 20$). Die geschätzte Faktorladungsmatrix $\mathbf{\Lambda} = \mathbf{P}_1\mathbf{M}_1^{1/2} : 20 \times 5$ ist in Abb. 8.8 angegeben. Die Zeilen der Ladungsmatrix enthalten die Gewichte der einzelnen Faktoren zur Berechnung der Variablen. Eine 2-dimensionale Projektion ist in Abb. 8.9 dargestellt. Die Gewichte sind als Punkte mit dem Variablennamen gezeichnet.

Die Summe der quadrierten Zeilen von $\mathbf{\Lambda}$ ergibt die Kommunalitäten (Abb. 8.10). Bitte für einige Zeilen nachrechnen!

Weiterhin kann die Gewichtsmatrix zur Berechnung der Faktor-Scores, d.h. $\mathbf{M}_1^{-1/2}\mathbf{P}_1' : q \times p$ verwendet werden. In Abb. 8.11 ist die transponierte Matrix (Dimensionen $p \times q = 20 \times 5$) ausgedruckt.

Anhand der Faktorladungen oder der Gewichtsmatrix können die Faktoren $\boldsymbol{\xi}$ inhaltlich interpretiert werden. Zunächst handelt es sich nur um latente Variablen, die die beobachtbaren Werte $\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon}$ erklären. Der erste Faktor ξ_1 hat einen positiven Einfluß auf alle Variablen (erste Spalte von $\mathbf{\Lambda}$). Man könnte ihn als g (generellen)-Faktor interpretieren. Diese Terminologie ist der Intelligenzforschung entlehnt. Der zweite Faktor hat positive Auswirkungen auf die Variablen 'Assault rate' und 'Employees

Correlation Matrix[a]

		Rooms per person	Dwelling without basic facilities	Household disposable income	Household financial wealth	Employment rate	Long-term unemployment rate	Quality of support network	Educational attainment	Students reading skills	Air pollution	Consultation on rule-making
Correlation	Rooms per person	1,000	,506	,518	,322	,614	,309	,683	,083	,425	,432	,282
	Dwelling without basic facilities	,506	1,000	,536	,206	,564	,063	,711	,266	,204	,454	,208
	Household disposable income	,518	,536	1,000	,630	,471	,234	,505	,171	,354	,415	,148
	Household financial wealth	,322	,206	,630	1,000	,319	,294	,231	,136	,184	-,014	-,105
	Employment rate	,614	,564	,471	,319	1,000	,545	,640	,396	,451	,386	,300
	Long-term unemployment rate	,309	,063	,234	,294	,545	1,000	,235	,125	,180	-,072	,314
	Quality of support network	,683	,711	,505	,231	,640	,235	1,000	,382	,253	,517	,252
	Educational attainment	,083	,266	,171	,136	,396	,125	,382	1,000	,462	,280	,153
	Students reading skills	,425	,204	,354	,184	,451	,180	,253	,462	1,000	,451	,373
	Air pollution	,432	,454	,415	-,014	,386	-,072	,517	,280	,451	1,000	,272
	Consultation on rule-making	,282	,208	,148	-,105	,300	,314	,252	,153	,373	,272	1,000
	Voter turnout	,301	,068	,396	,196	,112	,138	,163	-,172	,076	-,075	-,259
	Life expectancy	,623	,621	,581	,478	,612	,346	,512	,109	,396	,185	,082
	Self-reported health	,559	,399	,582	,354	,387	,352	,615	,002	,105	,287	,155
	Life Satisfaction	,700	,553	,591	,410	,633	,592	,729	,271	,181	,227	,199
	Homicide rate	,172	,477	,421	,092	,351	-,017	,381	,261	,521	,389	,164
	Assault rate	,161	,332	,454	,173	,326	,055	,377	,555	,696	,425	,345
	Employees working very long hours	,350	,590	,090	-,058	,448	-,176	,382	,480	,411	,334	,081
	Employment rate of women with children	,390	,468	,119	,155	,802	,335	,517	,505	,345	,313	,182
	Time devoted to leisure and personal care	,137	,412	,333	,074	,139	-,118	,240	,150	,303	,171	-,273

a. Determinant = 3,703E-009

Abbildung 8.4: Korrelationsmatrix mit Imputation von fehlenden Werten (Teil 1).

		Consultation on rule-making	Voter turnout	Life expectancy	Self-reported health	Life Satisfaction	Homicide rate	Assault rate	Employees working very long hours	Employment rate of women with children	Time devoted to leisure and personal care
Correlation	Rooms per person	,282	,301	,623	,559	,700	,172	,161	,350	,390	,137
	Dwelling without basic facilities	,208	,068	,621	,399	,553	,477	,332	,590	,468	,412
	Household disposable income	,148	,396	,581	,582	,591	,421	,454	,090	,119	,333
	Household financial wealth	-,105	,196	,478	,354	,410	,092	,173	-,058	,155	,074
	Employment rate	,300	,112	,612	,387	,633	,351	,326	,448	,802	,139
	Long-term unemployment rate	,314	,138	,346	,352	,592	-,017	,055	-,176	,335	-,118
	Quality of support network	,252	,163	,512	,615	,729	,381	,377	,382	,517	,240
	Educational attainment	,153	-,172	,109	,002	,271	,261	,555	,480	,505	,150
	Students reading skills	,373	,076	,396	,105	,181	,521	,696	,411	,345	,303
	Air pollution	,272	-,075	,185	,287	,227	,389	,425	,334	,313	,171
	Consultation on rule-making	1,000	-,259	,082	,155	,199	,164	,345	,081	,182	-,273
	Voter turnout	-,259	1,000	,231	,366	,290	,059	-,023	-,091	-,006	,379
	Life expectancy	,082	,231	1,000	,377	,590	,519	,377	,326	,322	,246
	Self-reported health	,155	,366	,377	1,000	,681	,096	,143	-,030	,184	,199
	Life Satisfaction	,199	,290	,590	,681	1,000	,093	,172	,133	,431	,196
	Homicide rate	,164	,059	,519	,096	,093	1,000	,724	,325	,044	,517
	Assault rate	,345	-,023	,377	,143	,172	,724	1,000	,242	,154	,257
	Employees working very long hours	,081	-,091	,326	-,030	,133	,325	,242	1,000	,511	,369
	Employment rate of women with children	,182	-,006	,322	,184	,431	,044	,154	,511	1,000	,019
	Time devoted to leisure and personal care	-,273	,379	,246	,199	,196	,517	,257	,369	,019	1,000

Abbildung 8.5: Korrelationsmatrix mit Imputation von fehlenden Werten (Teil 2).

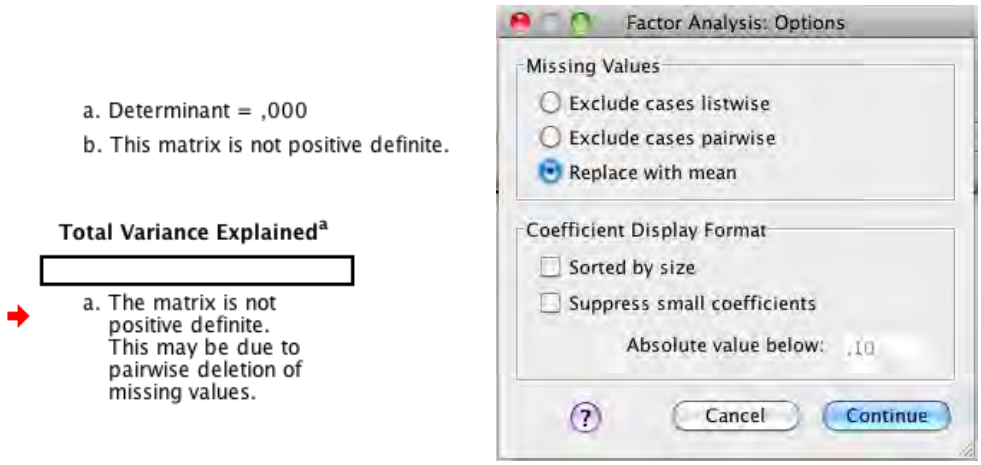


Abbildung 8.6: Links: Korrelationsmatrix mit paarweisem Fall-Ausschluß (OECD-Daten). Rechts: Option zur Imputation von fehlenden Werten.

working very long hours', während 'Self-reported health' negativ eingeht (nur Ladungen $|\lambda_{ij}| > 0.5$). Man könnte dies evtl. einen *Stress-Faktor* nennen. Dafür spricht auch und der negative Einfluß auf 'Life Satisfaction'.

Die Variablen 'Educational attainment', 'Students reading skills', 'Homicide rate' weisen positive Gewichte auf. Diese Variablen korrelieren hoch untereinander und mit 'Assault rate' (Abb. 8.5).

Etwas aus dem Rahmen fällt zunächst die Variable 'Long-term unemployment rate', die ein negatives Gewicht aufweist. Allerdings sind die Variablen 'Long-term unemployment rate' und 'Life Satisfaction' positiv korreliert ($r = .59$; vgl. auch Abb. 1.9, wo ein Korrelationscluster aus den obigen positiv gepolten Variablen zu sehen ist).

Insgesamt ist die Interpretation jedoch schwierig.

Der Stoff wird in Aufgabe 8.2 vertieft.

■

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7,308	36,538	36,538	7,308	36,538	36,538
2	2,661	13,307	49,845	2,661	13,307	49,845
3	2,056	10,279	60,123	2,056	10,279	60,123
4	1,562	7,810	67,933	1,562	7,810	67,933
5	1,268	6,338	74,271	1,268	6,338	74,271
6	,918	4,592	78,862			
7	,886	4,428	83,290			
8	,768	3,842	87,133			
9	,530	2,651	89,783			
10	,442	2,212	91,996			
11	,424	2,119	94,114			
12	,315	1,577	95,691			
13	,229	1,145	96,836			
14	,198	,992	97,828			
15	,147	,737	98,565			
16	,097	,484	99,049			
17	,064	,321	99,370			
18	,059	,295	99,665			
19	,046	,229	99,894			
20	,021	,106	100,000			

Extraction Method: Principal Component Analysis.

Abbildung 8.7: Tabelle der Eigenwerte (OECD-Daten).

Component Matrix[a]

	Component				
	1	2	3	4	5
Rooms per person	,756	-,274	-,093	-,107	-,230
Dwelling without basic facilities	,765	,102	,095	-,349	-,198
Household disposable income	,729	-,258	,352	,255	-,060
Household financial wealth	,433	-,417	,160	,237	,405
Employment rate	,827	-,037	-,331	-,098	,194
Long-term unemployment rate	,395	-,462	-,435	,318	,288
Quality of support network	,825	-,067	-,086	-,204	-,257
Educational attainment	,464	,484	-,239	,003	,389
Students reading skills	,599	,436	,033	,364	,178
Air pollution	,557	,348	-,010	,002	-,482
Consultation on rule-making	,323	,187	-,507	,462	-,399
Voter turnout	,238	-,480	,494	-,089	,121
Life expectancy	,750	-,165	,162	,033	,156
Self-reported health	,593	-,517	,078	,031	-,313
Life Satisfaction	,762	-,474	-,153	-,069	-,010
Homicide rate	,550	,459	,449	,247	-,019
Assault rate	,572	,503	,185	,500	,058
Employees working very long hours	,495	,530	-,041	-,529	,099
Employment rate of women with children	,595	,112	-,519	-,363	,283
Time devoted to leisure and personal care	,378	,169	,680	-,266	,129

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Abbildung 8.8: Faktorladungen (OECD-Daten). Die Summe der quadrierten Zeilen ergibt die Kommunalitäten.

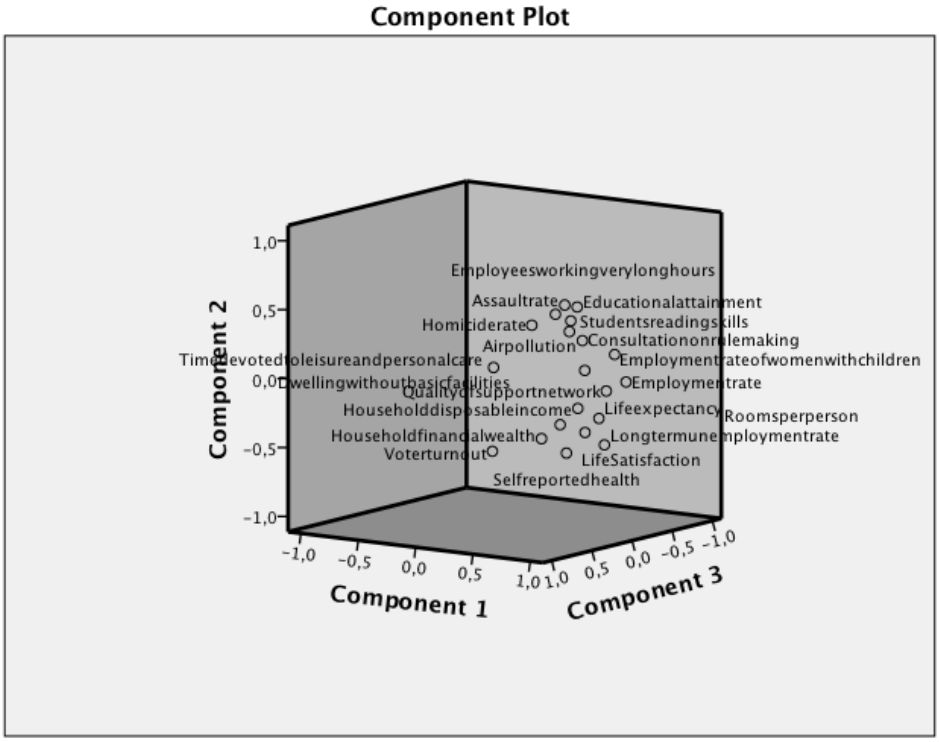


Abbildung 8.9: Faktorladungen (OECD-Daten). Projektion auf die Fläche.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,637
Bartlett's Test of Sphericity	Approx. Chi-Square	262,091
	df	190
	Sig.	,000

Communalities

	Initial	Extraction
Rooms per person	1,000	,720
Dwelling without basic facilities	1,000	,766
Household disposable income	1,000	,791
Household financial wealth	1,000	,607
Employment rate	1,000	,841
Long-term unemployment rate	1,000	,743
Quality of support network	1,000	,800
Educational attainment	1,000	,658
Students reading skills	1,000	,714
Air pollution	1,000	,663
Consultation on rule-making	1,000	,768
Voter turnout	1,000	,553
Life expectancy	1,000	,642
Self-reported health	1,000	,724
Life Satisfaction	1,000	,833
Homicide rate	1,000	,777
Assault rate	1,000	,868
Employees working very long hours	1,000	,816
Employment rate of women with children	1,000	,848
Time devoted to leisure and personal care	1,000	,721

Extraction Method: Principal Component Analysis.

Abbildung 8.10: Kommunalitätenschätzung $h_i^2 = \sum_j \lambda_{ij}^2$.

Component Score Coefficient Matrix

	Component				
	1	2	3	4	5
Rooms per person	,103	-,103	-,045	-,069	-,181
Dwelling without basic facilities	,105	,038	,046	-,223	-,156
Household disposable income	,100	-,097	,171	,164	-,048
Household financial wealth	,059	-,157	,078	,151	,320
Employment rate	,113	-,014	-,161	-,063	,153
Long-term unemployment rate	,054	-,173	-,212	,204	,227
Quality of support network	,113	-,025	-,042	-,131	-,203
Educational attainment	,064	,182	-,116	,002	,307
Students reading skills	,082	,164	,016	,233	,140
Air pollution	,076	,131	-,005	,002	-,380
Consultation on rule-making	,044	,070	-,246	,296	-,314
Voter turnout	,033	-,180	,240	-,057	,095
Life expectancy	,103	-,062	,079	,021	,123
Self-reported health	,081	-,194	,038	,020	-,247
Life Satisfaction	,104	-,178	-,074	-,044	-,008
Homicide rate	,075	,173	,218	,158	-,015
Assault rate	,078	,189	,090	,320	,046
Employees working very long hours	,068	,199	-,020	-,339	,078
Employment rate of women with children	,081	,042	-,252	-,233	,223
Time devoted to leisure and personal care	,052	,063	,331	-,170	,102

Extraction Method: Principal Component Analysis.

Abbildung 8.11: Gewichtsmatrix für die Faktorwerte (OECD-Daten).

8.5 Rotation der Faktoren

Aufgrund der Darstellung

$$\begin{aligned} \mathbf{x} &= \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon} && (8.88)\\ &= \mathbf{\Lambda}\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\xi} + \boldsymbol{\epsilon} && (8.89)\\ &= \tilde{\mathbf{\Lambda}}\tilde{\boldsymbol{\xi}} + \boldsymbol{\epsilon}. && (8.90) \end{aligned}$$

sind die Faktoren und Ladungen nicht eindeutig. Man kann das Produkt $\tilde{\boldsymbol{\xi}} = \mathbf{A}^{-1}\boldsymbol{\xi}$ als Rotation im Raum der Faktoren auffassen. Die neuen Faktoren haben die Kovarianzmatrix

$$\mathrm{Var}(\tilde{\boldsymbol{\xi}}) = \tilde{\mathbf{\Phi}} = \mathrm{Var}(\mathbf{A}^{-1}\boldsymbol{\xi}) = \mathbf{A}^{-1}\mathbf{\Phi}\mathbf{A}'^{-1} \tag{8.91}$$

und die transformierten Ladungen sind

$$\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}\mathbf{A}. \tag{8.92}$$

Die Kommunalitäten $h_i^2 = \sum_j \lambda_{ij}^2 = [\mathrm{diag}(\mathbf{\Lambda}\mathbf{\Lambda}')]_i$ transformieren sich zu

$$\mathrm{diag}(\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Lambda}}') = \mathrm{diag}(\mathbf{\Lambda}\mathbf{A}\mathbf{A}'\mathbf{\Lambda}'). \tag{8.93}$$

Die Rotationsmatrix $\mathbf{A}$ kann auf verschiedene Weise bestimmt werden. Fordert man, daß

orthogonale Matrix

$$\mathbf{A}^{-1} = \mathbf{A}' \tag{8.94}$$

(orthogonale Matrix), so bleiben die Winkel zwischen Objekten unverändert, d.h.

$$(\mathbf{A}^{-1}\boldsymbol{\xi}_n)'(\mathbf{A}^{-1}\boldsymbol{\xi}_m) = \boldsymbol{\xi}_n'\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\xi}_m = \boldsymbol{\xi}_n'\boldsymbol{\xi}_m, \tag{8.95}$$

da die Skalarprodukte $\mathbf{y}'\mathbf{x} = ||\mathbf{x}|| \cdot ||\mathbf{y}|| \cdot \cos\phi$ proportional zum Winkel sind.

Die Kommunalitäten bleiben unverändert, da $\mathbf{A}\mathbf{A}' = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

Für die Kovarianzmatrix der rotierten Faktoren gilt

$$\tilde{\mathbf{\Phi}} = \mathbf{A}^{-1}\mathbf{\Phi}\mathbf{A}. \tag{8.96}$$

Ähnlichkeitstransformation

Man spricht auch von einer Ähnlichkeitstransformation. Insbesondere bleiben die Eigenwerte gleich, während die Eigenvektoren rotiert sind.

Sind die ursprünglichen Faktoren orthogonal, d.h. $\mathbf{\Phi} = \mathbf{I}$, so gilt ebenfalls $\tilde{\mathbf{\Phi}} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

Varimax-Methode

Einfachstruktur

Es gibt eine Vielzahl von Methoden, um eine Rotationsmatrix zu finden. Bei der sog. Varimax-Methode (Kaiser) wird versucht, die Ladungsmatrix so zu bestimmen, daß die Ladungen auf einige Variablen sehr hoch, auf andere dagegen sehr niedrig sind (eine sog. Einfachstruktur). Hierbei wird das Kriterium (Varianz der quadrierten Ladungen)

$$\sum_{i=1}^{p}\sum_{j=1}^{q}(d_{ij}^2 - \overline{d_j^2})^2 \tag{8.97}$$

maximiert, wobei $d_{ij} = \tilde{\lambda}_{ij}/h_i$ die durch deren Kommunalität h_i normierte rotierte Faktorladung und $\overline{d_j^2} = p^{-1}\sum_i d_{ij}^2$ der Mittelwert der normierten rotierten Ladungsquadrate ist (vgl. Mardia et al., 1979, Kap. 9.6).

Beispiel 8.4 (OECD-Daten, rotierte Faktoren)

Rotiert man die Faktoren mit dem Varimax-Kriterium, so ergibt sich folgendes Ergebnis (Abb. 8.12). Eine 2-dimensionale Projektion der rotierten Ladungen ist in Abb. 8.13 zu sehen (vgl. Abb. 8.9). Die Drehmatrix ist in Abb. 8.14 abgedruckt.

Übung: Multiplizieren Sie die erste Zeile der Ladungsmatrix Abb. 8.8 von rechts mit $\mathbf{A}$.

Faktoren mit hohen Ladungen ($|\lambda_{ij}| > 0.5$) sind in Abb. 8.15 abgedruckt (nach Größe sortiert).

Man könnte von den 5 latenten Dimensionen 1. Lebensqualität, 2. Streß und Unsicherheit, 3. Arbeitsbelastung, 4. Wohlstand und 5. politische Partizipation sprechen. Hierbei scheint hoher Wohlstand mit hoher Langzeitarbeitslosigkeit einherzugehen.[5] Außerdem sind Wahlbeteiligung (Voter turnout) und 'Consultation on rule-making' negativ korreliert. Hohe Transparenz bei Gesetzesvorhaben geht also einher mit eher niedriger Wahlbeteiligung.

Die Gruppierung der Variablen in Abb. 1.5 läßt sich daher faktorenanalytisch nur teilweise bestätigen.

■

[5] Vermutlich nicht auf individueller Ebene.

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Rooms per person	,798	,096	,205	,177	-,027
Dwelling without basic facilities	,687	,255	,411	-,155	,190
Household disposable income	,625	,462	-,115	,353	,224
Household financial wealth	,224	,171	,000	,680	,256
Employment rate	,525	,206	,618	,363	-,095
Long-term unemployment rate	,251	-,048	,142	,755	-,295
Quality of support network	,799	,197	,350	,015	-,014
Educational attainment	-,064	,434	,665	,122	-,092
Students reading skills	,105	,758	,299	,166	-,100
Air pollution	,524	,442	,162	-,354	-,204
Consultation on rule-making	,281	,316	,024	,012	-,767
Voter turnout	,323	-,031	-,215	,293	,562
Life expectancy	,525	,344	,228	,372	,239
Self-reported health	,797	,012	-,146	,256	,043
Life Satisfaction	,755	-,007	,233	,456	,018
Homicide rate	,194	,826	,073	-,099	,206
Assault rate	,101	,909	,122	,082	-,095
Employees working very long hours	,194	,240	,753	-,342	,194
Employment rate of women with children	,295	-,027	,841	,188	-,128
Time devoted to leisure and personal care	,201	,381	,112	-,147	,708

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Abbildung 8.12: Rotierte Faktorladungen (OECD-Daten). Werte im Betrag > 0.5 sind rot markiert.

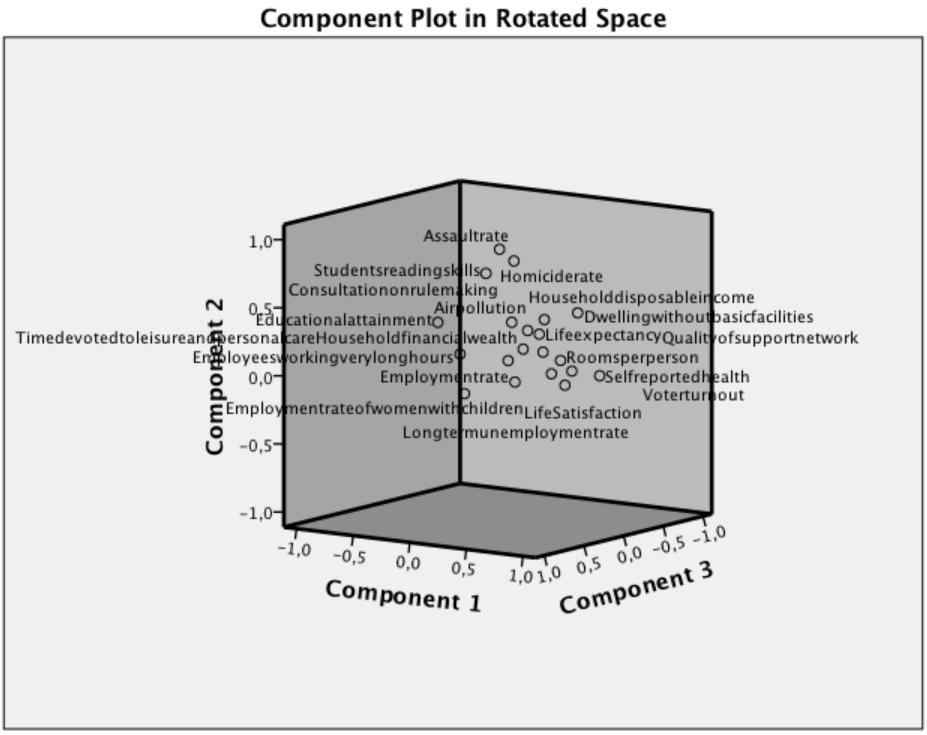


Abbildung 8.13: Rotierte Faktorladungen (OECD-Daten). Projektion auf die Fläche. Vgl. Abb. 8.9.

Component Transformation Matrix

Component	1	2	3	4	5
1	,728	,475	,419	,256	,059
2	-,393	,580	,396	-,575	-,144
3	,019	,360	-,453	-,162	,799
4	-,181	,553	-,517	,438	-,450
5	-,532	,046	,441	,621	,368

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Abbildung 8.14: Rotationsmatrix **A**.

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Quality of support network	,799				
Rooms per person	,798				
Self-reported health	,797				
Life Satisfaction	,755				
Dwelling without basic facilities	,687				
Household disposable income	,625				
Life expectancy	,525				
Air pollution	,524				
Assault rate		,909			
Homicide rate		,826			
Students reading skills		,758			
Employment rate of women with children			,841		
Employees working very long hours			,753		
Educational attainment			,665		
Employment rate	,525		,618		
Long-term unemployment rate				,755	
Household financial wealth				,680	
Consultation on rule-making					-,767
Time devoted to leisure and personal care					,708
Voter turnout					,562

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Abbildung 8.15: Rotierte Faktorladungen (OECD-Daten) nach Größe sortiert. Nur Werte im Betrag > 0.5.

Component Score Coefficient Matrix

	Component				
	1	2	3	4	5
Rooms per person	,224	-,073	-,021	-,050	-,051
Dwelling without basic facilities	,185	-,042	,085	-,198	,081
Household disposable income	,110	,141	-,180	,096	,065
Household financial wealth	-,091	,064	-,010	,358	,138
Employment rate	,015	-,040	,215	,130	-,035
Long-term unemployment rate	-,054	-,028	,045	,378	-,149
Quality of support network	,223	-,058	,034	-,133	-,039
Educational attainment	-,191	,109	,286	,122	-,003
Students reading skills	-,121	,275	,034	,113	-,059
Air pollution	,206	,094	-,083	-,290	-,159
Consultation on rule-making	,114	,122	-,134	-,055	-,453
Voter turnout	,059	-,030	-,095	,107	,280
Life expectancy	,031	,058	,026	,135	,114
Self-reported health	,264	-,061	-,179	-,018	-,037
Life Satisfaction	,156	-,105	,026	,117	-,011
Homicide rate	-,030	,301	-,087	-,055	,077
Assault rate	-,098	,359	-,078	,066	-,078
Employees working very long hours	-,010	-,043	,326	-,194	,141
Employment rate of women with children	-,039	-,146	,384	,074	-,016
Time devoted to leisure and personal care	-,004	,091	,030	-,088	,372

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Abbildung 8.16: Rotierte Gewichtsmatrix für die Faktorwerte (OECD-Daten).

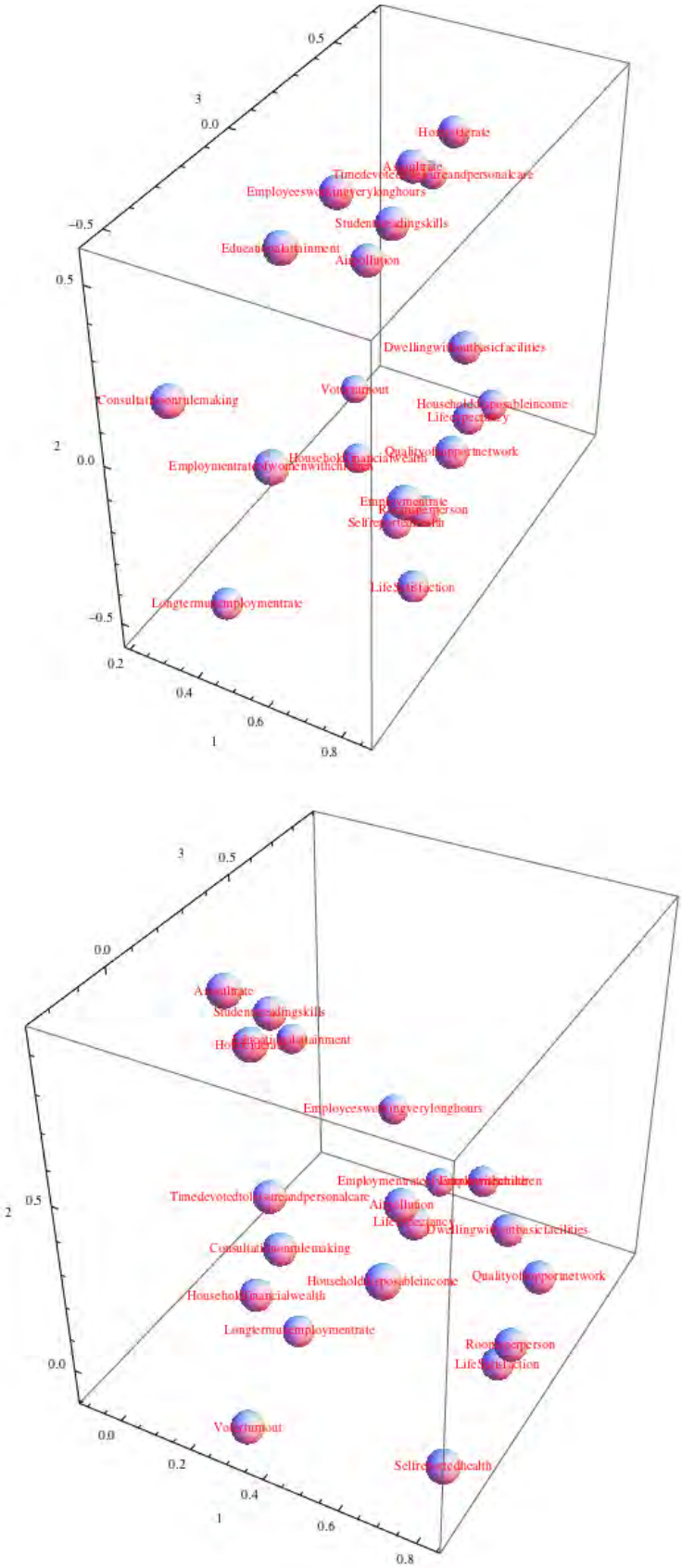


Abbildung 8.17: 3D-Darstellung der Faktorladungen (oben: ursprüngliche, unten: rotierte Ladungen).
http://www.fernuni-hagen.de/ls_statistik/lehre/

8.6 Maximum-Likelihood-Methode

8.6.1 Modell

Faktorenanalyse (Maximum-Likelihood)

Die Maximum-Likelihood(ML)-Methode geht von dem Grundmodell

$$\mathbf{x} = \mathbf{\Lambda}\boldsymbol{\xi} + \boldsymbol{\epsilon}. \tag{8.98}$$

aus, wobei nun probabilistische Anahmen gemacht werden. Unter Normalverteilungsannahme für $\boldsymbol{\xi}$ und $\boldsymbol{\epsilon}$ ist auch $\mathbf{x}$ normalverteilt mit Erwartungswert und Kovarianzmatrix

$$E[\mathbf{x}] = \boldsymbol{\mu} = \mathbf{0} \tag{8.99}$$

$$\mathrm{Var}(\mathbf{x}) = \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{V}. \tag{8.100}$$

Hierbei wurde wieder die Zentrierung sowie die Unkorreliertheit der Faktoren und Fehler angenommen ($\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\epsilon}) = \mathbf{O}$). Daher ist die gemeinsame Verteilung der Messungen $\mathbf{x}_n, n = 1, ..., N$ durch

$$f(\mathbf{x}_1, ..., \mathbf{x}_N) = \prod_{n=1}^{N} f(\mathbf{x}_n) \tag{8.101}$$

$$= \prod_{n=1}^{N} |2\pi\mathbf{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}_n'\mathbf{\Sigma}^{-1}\mathbf{x}_n\right) \tag{8.102}$$

gegeben. Interpretiert man die Wahrscheinlichkeits-Dichte der Daten als Funktion der freien Parameter $\boldsymbol{\theta}$ in $\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Sigma}(\mathbf{\Lambda}(\boldsymbol{\theta}), \mathbf{\Phi}(\boldsymbol{\theta}), \mathbf{V}(\boldsymbol{\theta}))$, so ergibt sich die Likelihood-Funktion[6]

Likelihood-Funktion

$$L(\boldsymbol{\theta}) = f(\mathbf{x}_1, ..., \mathbf{x}_N; \boldsymbol{\theta}). \tag{8.103}$$

Maximiert man diese bezüglich $\boldsymbol{\theta}$, so ergibt sich der Maximum-Likelihood(ML)-Schätzer

Maximum-Likelihood(ML)-Schätzer

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} f(\mathbf{x}_1, ..., \mathbf{x}_N; \boldsymbol{\theta}). \tag{8.104}$$

[6] Die immer wiederkehrende Behauptung, daß die ML-Methode nur bei Normalverteilung anwendbar sei, ist falsch. Es ergibt sich nur eine andere Form als (8.102). Vgl. etwa Abs. 6.1.4.

Meistens wird anstelle von L die Log-Likelihood $l = \log(L)$

Log-Likelihood

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= \sum_{n=1}^{N} \log(f(\mathbf{x}_n)) && (8.105)\\
&= -\frac{N}{2}\log|2\pi\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}\mathrm{tr}[\boldsymbol{\Sigma}^{-1}\mathbf{x}_n\mathbf{x}_n'] && (8.106)\\
&= -\frac{N}{2}\log|2\pi\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{N}{2}\mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}_*] && (8.107)
\end{aligned}
$$

benutzt $[\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}); \mathbf{x}'\mathbf{A}\mathbf{x} = \mathrm{tr}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}')]$.
Die Daten wurden hierbei in der Statistik

$$\mathbf{S}_* = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\mathbf{x}_n' \qquad (8.108)$$

komprimiert (ML-Stichproben-Kovarianz bei zentrierten Variablen). Bei nichtzentrierten Größen mit $E[\mathbf{x}] = \boldsymbol{\mu}$ nimmt man

$$\mathbf{S}_* = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})' = \frac{1}{N}\mathbf{X}'\mathbf{H}\mathbf{X}. \qquad (8.109)$$

da $E(\mathbf{x}_n - \bar{\mathbf{x}}) = \mathbf{0}$, wie gefordert. [7]
Im klassischen Modell der orthonormierten Faktoren kann zudem $\boldsymbol{\Phi} = \mathrm{Var}(\boldsymbol{\xi}) = \mathbf{I}_q$ gesetzt werden. Dann enthält das Modell nur die freien Parameter in $\boldsymbol{\Lambda}$ und $\mathbf{V}$. Die Berechnung der ML-Schätzer muß numerisch erfolgen, da es sich um ein nichtlineares Schätzproblem handelt.

8.6.2 Identifikation

Identifikations-problem

Zuvor muß aber geklärt werden, ob die Parameter identifizierbar sind (Identifikationsproblem). Im Fall von $p = 2$ manifesten Variablen und einem latenten Faktor ($q = 1$) gilt bei diagonaler Fehler-Varianz $\mathbf{V}$

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \mathbf{V} && (8.110)\\
\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} &= \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \end{bmatrix}\begin{bmatrix} \lambda_{11} & \lambda_{21} \end{bmatrix} + \begin{bmatrix} v_{11} & 0 \\ 0 & v_{22} \end{bmatrix} && (8.111)\\
&= \begin{bmatrix} \lambda_{11}^2 & \lambda_{11}\lambda_{21} \\ \lambda_{21}\lambda_{11} & \lambda_{21}^2 \end{bmatrix} + \begin{bmatrix} v_{11} & 0 \\ 0 & v_{22} \end{bmatrix}. && (8.112)
\end{aligned}
$$

[7] Glg. (8.107) ist dann die sog. konzentrierte Likelihood nach Einsetzen des ML-Schätzers $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$.

Aus den Parametern in $\mathbf{\Sigma}$ (Parameter der beobachtbaren Größen) müssen sich die freien Parameter der latenten Variablen *eindeutig* berechnen lassen, damit das Modell identifiziert ist. Aufgrund der Symmetrie hat man 3 Parameter in $\mathbf{\Sigma}$, jedoch 4 Parameter auf der rechten Seite.

Identifikation

Daher kann keine eindeutige Lösung gefunden werden – das Modell ist *nicht identifizierbar* und kann daher auch nicht geschätzt werden. Man muß Restriktionen setzen (etwa $\lambda_{11} = 1$) oder mehr Indikatoren messen. Dann ergibt sich

Restriktionen

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{21} \end{bmatrix} \begin{bmatrix} 1 & \lambda_{21} \end{bmatrix} + \begin{bmatrix} v_{11} & 0 \\ 0 & v_{22} \end{bmatrix} \tag{8.113}$$

$$= \begin{bmatrix} 1 & \lambda_{21} \\ \lambda_{21} & \lambda_{21}^2 \end{bmatrix} + \begin{bmatrix} v_{11} & 0 \\ 0 & v_{22} \end{bmatrix}. \tag{8.114}$$

woraus alle Parameter identifiziert sind.

Übung 8.4 (Identifikation)

Lösen Sie die Gleichungen nach λ und v auf.

■

Im allgemeinen hat man $p(p+1)/2$ freie Parameter in $\mathbf{\Sigma}$ sowie $pq + p$ freie Parameter in $\mathbf{\Lambda}$ und $\mathbf{V}$ (falls diagonal). Die Zahl der Gleichungen $p(p+1)/2$ muß also auf jeden Fall größer oder gleich $pq+p$ sein. Die Differenz $s_1(p, q)$ beider Größen ist als Tabelle aufgetragen ($p = 1, ..., 20; q = 1, ..., 5$), Tab. 8.1, oben. Führt man $q(q-1)$ Restriktionen durch

$$\mathbf{\Lambda}\ \mathrm{diag}(\sigma_{11}, ..., \sigma_{pp})^{-1}\ \mathbf{\Lambda}' = \text{diagonal} \tag{8.115}$$

ein (vgl. Mardia et al., Kap. 9.2), so ergibt sich die untere Tabelle für $s_2(p, q) = p(p+1)/2 - (pq + p) + q(q-1)/2$. Die Werte s_1, s_2 dürfen auf jeden Fall nicht negativ sein.

Hat man $s_1, s_2 > 0$, so ergibt das Faktor-Modell eine sparsamere Erklärung der Daten, als dies $\mathrm{Var}(\mathbf{x}) = \mathbf{\Sigma}$ leisten würde.

$s_1(p,q)$	1	2	3	4	5
1	-1	-2	-3	-4	-5
2	-1	-3	-5	-7	-9
3	0	-3	-6	-9	-12
4	2	-2	-6	-10	-14
5	5	0	-5	-10	-15
6	9	3	-3	-9	-15
7	14	7	0	-7	-14
8	20	12	4	-4	-12
9	27	18	9	0	-9
10	35	25	15	5	-5
11	44	33	22	11	0
12	54	42	30	18	6
13	65	52	39	26	13
14	77	63	49	35	21
15	90	75	60	45	30
16	104	88	72	56	40
17	119	102	85	68	51
18	135	117	99	81	63
19	152	133	114	95	76
20	170	150	130	110	90

$s_2(p,q)$	1	2	3	4	5
1	-1	-1	0	2	5
2	-1	-2	-2	-1	1
3	0	-2	-3	-3	-2
4	2	-1	-3	-4	-4
5	5	1	-2	-4	-5
6	9	4	0	-3	-5
7	14	8	3	-1	-4
8	20	13	7	2	-2
9	27	19	12	6	1
10	35	26	18	11	5
11	44	34	25	17	10
12	54	43	33	24	16
13	65	53	42	32	23
14	77	64	52	41	31
15	90	76	63	51	40
16	104	89	75	62	50
17	119	103	88	74	61
18	135	118	102	87	73
19	152	134	117	101	86
20	170	151	133	116	100

Tabelle 8.1: Zahl der Gleichungen minus Zahl der freien Parameter. Oben: ohne Restriktionen. Unten: mit Restriktionen (vgl. Haupttext)

Beispiel 8.5 (OECD-Daten (ML))

Zum Vergleich mit der Hauptkomponentenanalyse hier die Resultate der Maximum-Likelihood-Methode. Da man 20 Indikatoren hat, aber nur 5 Faktoren extrahiert wurden, ist das Modell identifiziert (bitte anhand von Tabelle 8.1 überprüfen).

Abb. 8.18 zeigt die nach dem Varimax-Kriterium rotierte Faktorladungsmatrix. Die Werte unterscheiden sich stark von den Resultaten der Hauptkomponentenanalyse. Entsprechendes gilt für die Schätzung der Kommunalitäten $h_i^2 = \sum_j \lambda_{ij}^2$ (Zeilensummen der quadrierten Ladungsmatrix), Abb. 8.19.

Man könnte von den 5 latenten Dimensionen 1. Lebensqualität, 2. Arbeitsbelastung, 3. Streß und Unsicherheit und 4. primitive Lebensverhältnisse sprechen. Der 5. Faktor ist schwer interpretierbar.

■

8.7 Methode der kleinsten Quadrate

Die Maximum-Likelihood-Methode setzt die Kenntnis der Verteilung von $\mathbf{x}$ voraus. Wenn diese nicht bekannt ist, kann man anstatt der möglicherweise inkorrekten gaußschen Likelihood-Funktion

$$l(\boldsymbol{\theta}) = -\frac{N}{2}\log|2\pi\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{N}{2}\mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}] \tag{8.116}$$

eine Diskrepanzfunktion

Diskrepanzfunktion (ungewichtete kleinste Quadrate)

$$F(\boldsymbol{\theta}) = \mathrm{tr}[(\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbf{S})^2] \tag{8.117}$$

$$= ||\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbf{S}||^2 = \sum_{ij}(\Sigma_{ij} - S_{ij})^2 \geq 0 \tag{8.118}$$

betrachten (quadrierter euklidischer Abstand von $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ und der Stichprobenkovarianzmatrix $\mathbf{S}$) und diese minimieren.[8]

Man spricht von einer ULS-Diskrepanzfunktion (ULS = unweighted least squares = ungewichtete kleinste Quadrate).

[8] Der Ausdruck $||\mathbf{A}||^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}') = \sum_{ij} A_{ij}^2$ ist die quadrierte euklidische Matrixnorm von $\mathbf{A}$. Hier gilt $\boldsymbol{\Sigma} - \mathbf{S} := \mathbf{A} = \mathbf{A}'$.

Rotated Factor Matrix[a]

	Factor				
	1	2	3	4	5
Rooms per person	,657	,360	,048	,228	-,025
Dwelling without basic facilities	,418	,408	,128	,634	-,128
Household disposable income	,683	-,011	,315	,306	,086
Household financial wealth	,499	,022	,140	-,022	,105
Employment rate	,473	,758	,305	,090	,278
Long-term unemployment rate	,575	,264	,153	-,451	,246
Quality of support network	,592	,423	,239	,362	-,251
Educational attainment	-,035	,423	,583	,056	-,316
Students reading skills	,080	,215	,677	,191	,127
Air pollution	,151	,268	,327	,377	-,108
Consultation on rule-making	,147	,159	,383	-,093	-,008
Voter turnout	,417	-,113	-,099	,094	,094
Life expectancy	,586	,219	,236	,376	,251
Self-reported health	,752	,063	,047	,085	-,153
Life Satisfaction	,867	,324	,097	,017	-,187
Homicide rate	,116	-,035	,562	,629	,329
Assault rate	,098	-,041	,950	,290	-,018
Employees working very long hours	-,122	,590	,120	,559	-,102
Employment rate of women with children	,139	,907	,192	-,020	-,038
Time devoted to leisure and personal care	,165	-,026	,073	,589	,039

Extraction Method: Maximum Likelihood.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 13 iterations.

Abbildung 8.18: Maximum-Likelihood-Methode. Rotierte Faktorladungen (OECD-Daten). Werte im Betrag > 0.5 sind rot markiert.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,637
Bartlett's Test of Sphericity	Approx. Chi-Square	262,091
	df	190
	Sig.	,000

Communalities[a]

	Initial	Extraction
Rooms per person	,883	,617
Dwelling without basic facilities	,899	,777
Household disposable income	,891	,666
Household financial wealth	,704	,280
Employment rate	,939	,977
Long-term unemployment rate	,766	,687
Quality of support network	,873	,780
Educational attainment	,789	,623
Students reading skills	,893	,564
Air pollution	,704	,355
Consultation on rule-making	,704	,203
Voter turnout	,627	,214
Life expectancy	,853	,652
Self-reported health	,650	,603
Life Satisfaction	,905	,901
Homicide rate	,892	,835
Assault rate	,850	,999
Employees working very long hours	,819	,701
Employment rate of women with children	,930	,881
Time devoted to leisure and personal care	,777	,382

Extraction Method: Maximum Likelihood.

a. One or more communalitiy estimates greater than 1 were encountered during iterations. The resulting solution should be interpreted with caution.

Abbildung 8.19: Maximum-Likelihood-Methode. Kommunalitätenschätzung $\sum_j \lambda_{ij}^2$.

Abbildung 8.20: GLS-Faktorenanalyse (Auswahldialog).

Allgemeiner definiert man eine Diskrepanzfunktion mit gewichteten kleinsten Quadraten (Fahrmeir et al., 1996, Kap. 11, S. 746–748 f.)

Diskrepanzfunktion (gewichtete kleinste Quadrate)

$$F(\boldsymbol{\theta}) \quad = \quad \mathrm{tr}\{[(\boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbf{S})\mathbf{G}]^2\} \tag{8.119}$$

und Gewichtungsmatrix $\mathbf{G} > 0$ (positiv definit und symmetrisch). F ist eine positiv definite quadratische Form.[9] Die Wahl $\mathbf{G} = \mathbf{S}^{-1}$ wird als GLS-Diskrepanzfunktion bezeichnet, vgl. Abb. 8.20 (GLS = generalized least squares = verallgemeinerte KQ-Methode).

Auch in diesem Fall stellt sich das Identifikationsproblem, nämlich ob sich die Parameter in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ eindeutig aus $\boldsymbol{\Sigma}$ berechnen lassen. Die Schätzung mit SPSS führt auf die analogen Probleme wie bei ML.

Der Stoff wird in Aufgabe 8.3 vertieft.

[9] Schreibt man für die symmetrischen Matrizen $a_{ij} := \sigma_{ij} - s_{ij}$, $a_{ij} = a_{ji}, g_{ij} = g_{ji}$ die Spur als $\mathrm{tr}[\mathbf{AGAG}] = \sum_{ijk} a_{ij} g_{jk} a_{kl} g_{li} = \sum_{ijk} a_{ij} g_{il} g_{jk} a_{lk} = \mathbf{a}'(\mathbf{G} \otimes \mathbf{G})\mathbf{a}$, $\mathbf{a} = [a_{11}, a_{12}, ..., a_{pp}]' := \mathrm{row}\mathbf{A}$ (zeilenweiser Vektor aus $\mathbf{A}$), so ist dies eine quadratische Form der positiv definiten Matrix $\boldsymbol{\Omega} = \mathbf{G} \otimes \mathbf{G}$ (vgl. Kap. 9.10, Glg. 9.188). Man kann auch eine allgemeine positiv definite Matrix $\boldsymbol{\Omega}$ als Gewichtsmatrix verwenden (Diskrepanzfunktionen von quadratischer Form).

Anhänge

Kapitel 9

Matrix-Algebra

In diesem Kapitel sollen einige Matrix-Methoden zusammengestellt werden, die für einen Kurs in multivariater Statistik unverzichtbar sind. Entsprechende Kapitel enthalten (wahrscheinlich) alle Lehrbücher zum Thema (Mardia et al., 1979; Fahrmeir et al., 1996). Bei speziellerem Interesse verweise ich auf Golub und Van Loan (1996); Magnus und Neudecker (1999).

9.1 Vektoren und Matrizen

9.1.1 Datenmatrix

Daten werden am übersichtlichsten in Form einer Tabelle zusammengestellt. Die Daten $x_{ni}, n = 1, ..., N, i = 1, ..., p$ werden dann als Daten-Matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... \\ x_{N1} & x_{N2} & ... & x_{Np} \end{bmatrix} : N \times p \tag{9.1}$$

Daten-Matrix

bezeichnet.

Die Matrix-Elemente (n-te Zeile, i-te Spalte) können als $(\mathbf{X})_{ni} = \mathbf{X}_{ni} = x_{ni}$ geschrieben werden.

Spalten der Matrix $\mathbf{X}$ (Variablen) lassen sich als Spalten-Vektoren $\mathbf{x}_{(i)}$:

$N \times 1, i = 1, ..., p$ auffassen. [1] Damit kann $\mathbf{X}$ in der Form

$$\mathbf{X} = [\mathbf{x}_{(1)}, ..., \mathbf{x}_{(p)}] \tag{9.2}$$

geschrieben werden. Auch Zeilen-Vektoren $\mathbf{x}'_n : 1 \times p, n = 1, ..., N$ kann man definieren (Beobachtungen der statistischen Einheit n). [2] Damit gilt

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_N \end{bmatrix}. \tag{9.3}$$

transponierte Matrix

Matrizen lassen sich transponieren. Man vertauscht einfach Zeilen und Spalten. Es gilt

$$(\mathbf{X}')_{in} = (\mathbf{X})_{ni}; i = 1, ..., p, n = 1, ..., N. \tag{9.4}$$

Die transponierte Matrix hat also p Zeilen und N Spalten. Man kann schreiben

$$\mathbf{X}' = [\mathbf{x}_1, ..., \mathbf{x}_N] : p \times N. \tag{9.5}$$

Beispiel 9.1 (Spalten, Zeilen und transponierte Daten-Matrix)

Die Daten-Matrix hat die Form $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$.

Der 2. Spaltenvektor ist somit $\mathbf{x}_{(2)} = \begin{bmatrix} 2 \\ 5 \\ 8 \\ 11 \end{bmatrix}$

und die 4. Zeile $\mathbf{x}'_4 = \begin{bmatrix} 10 & 11 & 12 \end{bmatrix}$.

Die transponierte Matrix ist durch $\mathbf{X}' = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix} = [\mathbf{x}_1, ..., \mathbf{x}_4]$

gegeben.

■

[1] Der Spalten-Index wird in Klammern geschrieben, um die Spalten von den Zeilen $\mathbf{x}'_n, n = 1, ..., N$ zu unterscheiden.

[2] Hier wird die Konvention benutzt, daß Vektor-Symbole wie $\mathbf{x}, \mathbf{y}, ...$ immer Spalten-Vektoren sind.

9.1.2 Allgemeine Definition von Vektoren und Matrizen

Ganz unabhängig von der Datenmatrix und ihren Zeilen und Spalten kann man Vektoren und Matrizen wie folgt definieren:
Eine $I \times J$-Matrix $\mathbf{A}$ ist durch das rechteckige Schema

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1J} \\ a_{21} & a_{22} & \dots & a_{2J} \\ \vdots & \vdots & & \vdots \\ a_{I1} & a_{I2} & \dots & a_{IJ} \end{bmatrix} : I \times J \tag{9.6}$$

Matrix

definiert.
Ein Spaltenvektor ist die $I \times 1$-Matrix [3]

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_I \end{bmatrix} = [x_1, ..., x_I]' \tag{9.8}$$

Spaltenvektor

Ein Zeilenvektor ist die $1 \times I$-Matrix

$$\mathbf{y}' = [y_1, ..., y_I] \tag{9.9}$$

Zeilenvektor

[3] Formal korrekt müßte man

$$\mathbf{x} = \begin{bmatrix} x_{11} \\ \vdots \\ x_{I1} \end{bmatrix} = [x_{11}, ..., x_{I1}]' \tag{9.7}$$

schreiben, da es sich um eine Matrix handelt. Meistens wird der Index 1 weggelassen.

Spalten und Zeilen von Matrizen lassen sich auch durch die Notation

$$\mathbf{s}_j = \mathbf{A}_{.j} \tag{9.10}$$
$$\mathbf{z}_i' = \mathbf{A}_{i.} \tag{9.11}$$

angeben. Der Punkt bedeutet, daß alle Werte des nicht notierten Index (also i bzw. j) durchlaufen werden.
Zeilen und Spalten einer Matrix $\mathbf{A} : I \times J$ können vertauscht werden. Die transponierte Matrix $\mathbf{A}' : J \times I$ ist das Schema

transponierte Matrix

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{I1} \\ a_{12} & a_{22} & \dots & a_{I2} \\ \vdots & \vdots & & \vdots \\ a_{1J} & a_{2J} & \dots & a_{IJ} \end{bmatrix} : J \times I \tag{9.12}$$
$$(\mathbf{A}')_{ij} = (\mathbf{A})_{ji} \tag{9.13}$$

Entprechend wird aus einem Spaltenvektor ein Zeilenvektor, d.h.

$$\mathbf{x}' = [x_1, ..., x_I]. \tag{9.14}$$

Eine (quadratische) Matrix, für die gilt

symmetrische Matrix

$$\mathbf{A}' = \mathbf{A} \tag{9.15}$$
$$(\mathbf{A}')_{ij} = (\mathbf{A})_{ij} \tag{9.16}$$

heißt symmetrisch.

Beispiel 9.2 (Kovarianzmatrix)
Die Kovarianzmatrix $\mathbf{\Sigma} = \mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$ ist symmetrisch.
■

9.1.3 Spezielle Matrizen

Matrizen und Vektoren mit einer speziellen Struktur werden häufig benutzt. Die Diagonalmatrix $\mathbf{D}$ enthält nur Werte auf der Diagonalen, d.h.

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & d_J \end{bmatrix} := \text{diag}(\mathbf{d}) : J \times J \tag{9.17}$$

Diagonalmatrix

Die Matrix mit Einsen auf der Hauptdiagonalen heißt Einheitsmatrix $\mathbf{I} = \mathbf{I}_J$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & 1 \end{bmatrix} = \text{diag}(\mathbf{1}) : J \times J \tag{9.18}$$

Einheitsmatrix

Wenn es keine Verwechslungen gibt, kann man den Index J weglassen.
Die Spalten von $\mathbf{I}$ sind die Einheitsvektoren, d.h.

$$\mathbf{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{I}_{.j} \tag{9.19}$$

(1 in Zeile j). Man kann auch $\mathbf{I} = [\mathbf{e}_1, ..., \mathbf{e}_J]$ schreiben.
Oft benutzt wird auch der Null-Vektor und der Eins-Vektor

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}. \tag{9.20}$$

Null-Vektor

Eins-Vektor

Null-Matrix

Eine Nullmatrix wird als $\mathbf{O} : I \times J$ notiert.

Im statistischen Kontext ist die sog. Equikorrelations-Matrix von Interesse, bei der alle Korrelationen gleich sind. Man kann schreiben

Equikorrelations-Matrix

$$\mathbf{E} = (1-\rho)\mathbf{I} + \rho\mathbf{J} : p \times p, \tag{9.21}$$

wobei $\mathbf{J} = \mathbf{1}\mathbf{1}'$ eine quadratische Matrix aus Einsen ist. Es gilt also $e_{ii} = 1, e_{ij} = \rho, i \neq j$. Für die inverse Matrix (siehe Abs. 9.7) findet man

$$\mathbf{E}^{-1} = (1-\rho)^{-1}\left(\mathbf{I} - \frac{\rho}{1+(p-1)\rho}\mathbf{J}\right). \tag{9.22}$$

Daher gilt für die Wahl $\rho = 1/2$:

$$\mathbf{I} + \mathbf{J} = 2\,\mathbf{E}(1/2) \tag{9.23}$$

$$(\mathbf{I} + \mathbf{J})^{-1} = [\mathbf{I} - \frac{1}{p+1}\mathbf{J}]. \tag{9.24}$$

9.2 Summen von Vektoren und Matrizen

Vektoren und Matrizen können komponentenweise addiert werden. Es gilt

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \tag{9.25}$$

$$c_{ij} = a_{ij} + b_{ij},\ i = 1, ..., I, j = 1, ...J. \tag{9.26}$$

Da Vektoren spezielle Matrizen sind, gilt auch

$$\mathbf{z} = \mathbf{x} + \mathbf{y} \tag{9.27}$$
$$z_i = x_i + y_i,\ i = 1, ..., I. \tag{9.28}$$

Der Stoff wird in Aufgabe 9.1 vertieft.

9.3 Produkte mit Skalaren

Vektoren und Matrizen können mit Skalaren (Zahlen) komponentenweise multipliziert werden.

$$\mathbf{B} = c\mathbf{A} \tag{9.29}$$
$$b_{ij} = c\, a_{ij},\ i = 1, ..., I, j = 1, ...J. \tag{9.30}$$

Der Vektor $c\mathbf{x} = cx_i$ hat die gleiche Richtung wie $\mathbf{x}$, nur eine c-fache Länge.

Der Stoff wird in Aufgabe 9.2 vertieft.

9.4 Produkte von Vektoren und Matrizen

Will man den Mittelwert einer Spalte (Variable i) der Datenmatrix berechnen, so muß man in dieser Spalte über alle Zeilen summieren (Spaltensumme). Etwa gilt $\bar{x}_i = N^{-1} \sum_{n=1}^{N} x_{ni}$. Die Summe kann man in Vektor-Notation als Produkt $\mathbf{1}'\mathbf{x}_{(i)}$ schreiben, wobei das Produkt explizit lautet

$$\mathbf{1}'\mathbf{x}_{(i)} = [1, ..., 1] \begin{bmatrix} x_{1i} \\ \vdots \\ x_{Ni} \end{bmatrix} = \sum_{n=1}^{N} x_{ni}. \tag{9.31}$$

Entsprechend kann durch

$$\mathbf{1}'\mathbf{X} = [1, ..., 1] \begin{bmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... \\ x_{N1} & x_{N2} & ... & x_{Np} \end{bmatrix} = [\mathbf{1}'\mathbf{x}_{(1)}, ..., \mathbf{1}'\mathbf{x}_{(p)}] \tag{9.32}$$

der Zeilen-Vektor aller Spaltensummen kompakt notiert werden.

Ganz allgemein definiert man das Produkt zweier Vektoren und/oder Matrizen als die Summe über die inneren Indizes.

Für 2 Vektoren gilt[4]

Skalarprodukt

$$\mathbf{x}'\mathbf{y} = \sum_{j=1}^{J} x_j y_j = [x_1, ..., x_J] \begin{bmatrix} y_1 \\ \vdots \\ y_J \end{bmatrix} \tag{9.34}$$

inneres Produkt

Das Ergebnis ist eine 1×1-Matrix (= Skalar = Zahl). Man spricht daher auch von einem Skalarprodukt (inneres Produkt).

Das Produkt aus Matrix $\mathbf{A}$ und dem Spalten-Vektor $\mathbf{x}$ hat die Form

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{9.35}$$

$$y_i = \sum_{j=1}^{J} a_{ij} x_j = (\mathbf{A}\mathbf{x})_i \tag{9.36}$$

Der Spalten-Index j von $\mathbf{A}$ wird 'wegsummiert', der Zeilenindex bleibt. Man erhält als Ergebnis einen Spaltenvektor.

[4] Formal korrekt müßte man

$$\mathbf{x}'\mathbf{y} = \sum_{j=1}^{J} x_{1j} y_{j1} = [x_{11}, ..., x_{1J}] \begin{bmatrix} y_{11} \\ \vdots \\ y_{J1} \end{bmatrix} : 1 \times 1 \tag{9.33}$$

schreiben.

Dieser hat i.A. eine andere Richtung und eine andere Länge als $\mathbf{x}$ (siehe Abs. 9.9).
Entsprechend gilt für die Linksmultiplikation

$$\mathbf{y}' = \mathbf{x}'\mathbf{A} \tag{9.37}$$

$$y_j = \sum_{i=1}^{J} x_i a_{ij} = (\mathbf{x}'\mathbf{A})_j \tag{9.38}$$

Der Zeilen-Index i von $\mathbf{A}$ wird 'wegsummiert', der Spaltenindex bleibt. Man erhält als Ergebnis einen Zeilenvektor.
Multipliziert man von links und rechts einen Vektor $\mathbf{x}$, so erhält man eine quadratische Form Q ($\mathbf{A}$ muß quadratisch und symmetrisch sein)

quadratische Form

$$Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} \tag{9.39}$$

$$= \sum_{i,j=1}^{I} x_i a_{ij} x_j. \tag{9.40}$$

Diese skalare Funktion kennen Sie bereits von der Dichtefunktion der Normalverteilung. $Q(\mathbf{x}) = r^2$ definiert eine Ellipsengleichung.
Setzt man $\mathbf{A} = \mathbf{I}$, so ist $Q = \mathbf{x}'\mathbf{x}$ das Quadrat der Länge (Norm) des Vektors. Man schreibt

Norm

$$||\mathbf{x}|| = \sqrt{\mathbf{x}'\mathbf{x}} \tag{9.41}$$

$$= \sqrt{\sum_{i=1}^{I} x_i^2}. \tag{9.42}$$

euklidischer Abstand

Der euklidische Abstand zweier Vektoren ist somit $||\mathbf{x} - \mathbf{y}||$.
Oft werden standardisierte Daten benutzt. Nimmt man als Gewichtsma-

trix $\mathbf{A} = \mathbf{\Sigma}^{-1}$ (inverse Kovarianz-Matrix), so ergibt sich die Norm

$$||\mathbf{x}||_{\mathbf{\Sigma}} = \sqrt{\mathbf{x}'\mathbf{\Sigma}^{-1}\mathbf{x}} \tag{9.43}$$

$$= \sqrt{\sum_{i,j}^{I} x_i a_{ij} x_j}. \tag{9.44}$$

Mahalanobis-Distanz

Der Abstand $||\mathbf{x} - \mathbf{y}||_{\mathbf{\Sigma}}$ wird auch als Mahalanobis-Distanz bezeichnet. Sie berücksichtigt die Korrelation in den Komponenten der Variablen $\mathbf{x}$.

positiv definit

Eine quadratische Form ist positiv definit, wenn $Q(\mathbf{x}) > 0$ für alle $\mathbf{x} \neq \mathbf{0}$. Entsprechend wird die symmetrische Matrix $\mathbf{A}$ als positiv definit bezeichnet, wenn $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$ positiv definit ist.

positiv semidefinit

Bemerkung: Gilt oben $\geq$, so nennt man die Matrix positiv semidefinit.

Mit Hilfe der Eigenwertzerlegung (Abs. 9.9) $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ kann man $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{P}\mathbf{\Lambda}\mathbf{P}'\mathbf{x} = \mathbf{y}'\mathbf{\Lambda}\mathbf{y} = \sum_i \lambda_i y_i^2$ schreiben. Also müssen für eine positiv definite Matrix alle Eigenwerte $\lambda_i > 0$ sein. Daher ist auch die Determinante $|\mathbf{A}| = \det(\mathbf{A}) = \prod_i \lambda_i > 0$ (Abs 9.6). Eine positiv definite Matrix $\mathbf{A}$ also auch nichtsingulär.

Für die Varianz der Zufallsvariable $\mathbf{a}'\mathbf{x}$ gilt ($\mathbf{a}$ beliebig):

$$\mathrm{Var}(\mathbf{a}\mathbf{x}) = \mathbf{a}'\mathrm{Var}(\mathbf{x})\mathbf{a} \geq 0 \tag{9.45}$$

Daher ist die Kovarianzmatrix $\mathbf{\Sigma} = \mathrm{Var}(\mathbf{x})$ positiv semidefinit.

Dividiert man einen Vektor durch seine Länge, so hat er die Länge 1. Gilt für das Skalarprodukt

orthogonale Vektoren

$$\mathbf{x}'\mathbf{y} = \sum_{j=1}^{J} x_j y_j = 0 \tag{9.46}$$

so sind die Vektoren orthogonal. Man kann dies aus der Darstellung $\mathbf{x}'\mathbf{y} = ||\mathbf{x}|| \cdot ||\mathbf{y}|| \cdot \cos\phi$ ablesen, da der Winkel $\pi/2 = 90°$ sein muß.
Gilt außerdem $||\mathbf{x}|| = ||\mathbf{y}|| = 1$, so sind die Vektoren orthonormal.

orthonormale Vektoren

Das Produkt der Matrizen $\mathbf{A}$ und $\mathbf{B}$ lautet

Matrix-Produkt

$$\mathbf{C} = \mathbf{AB} \tag{9.47}$$

$$c_{ik} = \sum_{j=1}^{J} a_{ij} b_{jk} = (\mathbf{AB})_{ik},\ i = 1, ..., I, k = 1, ...K. \tag{9.48}$$

Der innere Index j wird 'wegsummiert', die äußeren Indizes i, k bleiben. Die obigen Produkte kann man als Spezialfälle auffassen. Die Dimensionen der Matrizen sind

$$I \times K = (I \times J) \cdot (J \times K) \tag{9.49}$$

Der Stoff wird in den Aufgaben 9.3 und 9.4 vertieft.

Übung 9.1 (Produkt von Matrizen)

1. Wählen Sie $\mathbf{B} = \mathbf{x}$ als Vektor und schreiben Sie $\mathbf{Ax}$ in Komponenten. Vergleichen Sie Glg. 9.33.
2. Schreiben Sie das Produkt $\mathbf{D} = \mathbf{ABC}$ in Komponenten an.

■

Schreibt man die Matrizen mit Hilfe von Zeilen- und Spaltenvektoren, d.h.

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_I' \end{bmatrix}, \ \mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_K] \tag{9.50}$$

so gilt die Darstellung mit Skalarprodukten

$$c_{ik} = \mathbf{a}_i'\mathbf{b}_k, \ i = 1, ..., I, k = 1, ...K. \tag{9.51}$$

Man kann sich das Matrixprodukt also als Skalarprodukt der i-ten Zeile von $\mathbf{A}$ mit der k-ten Spalte von $\mathbf{B}$ vorstellen.

Übung 9.2 (Produkt von Matrizen)

Multiplizieren Sie $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ mit $\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

■

Die Multiplikation des Spalten-Vektors $\mathbf{x} : I \times 1$ mit dem Zeilenvektor $\mathbf{y}' : 1 \times J$ (eine sog. Dyade) wird als äußeres Produkt bezeichnet. Man erhält als Ergebnis eine $I \times J$-Matrix

äußeres Produkt

$$\mathbf{x}\mathbf{y}' : I \times J \tag{9.52}$$

$$(\mathbf{x}\mathbf{y}')_{ij} = x_i y_j, \ i = 1, ..., I, j = 1, ...J. \tag{9.53}$$

Beispiel 9.3 (Mittelwerte und Kovarianz-Matrizen)

Der Eins-Vektor kann zur Berechnung von Mittelwerten benutzt werden. Es gilt

$$\bar{\mathbf{x}} = N^{-1}\mathbf{X}'\mathbf{1} = N^{-1}\sum_n \mathbf{x}_n : p \times 1 \tag{9.54}$$

$$\bar{x}_i = N^{-1}\sum_n x_{ni} \tag{9.55}$$

da $\mathbf{X}' = [\mathbf{x}_1, ..., \mathbf{x}_N] : p \times N$.
Schreibt man die Matrix $\mathbf{M}$ als äußeres Produkt

$$\mathbf{M} := N^{-1}\mathbf{1}\mathbf{1}' := N^{-1}\mathbf{J} : N \times N \tag{9.56}$$

Mittelwerts-Matrix

so gibt die Multiplikation mit einem Vektor $\mathbf{x} : N \times 1$ einen Vektor von Mittelwerten, d.h.

$$\mathbf{M}\mathbf{x} = N^{-1}\mathbf{1}\mathbf{1}'\mathbf{x} = \mathbf{1}(N^{-1}\mathbf{1}'\mathbf{x}) = \mathbf{1}\bar{x} : N \times 1. \tag{9.57}$$

zentrierter Vektor

$$\mathbf{x}_* := \mathbf{x} - \mathbf{M}\mathbf{x} = (\mathbf{I} - \mathbf{M})\mathbf{x} \tag{9.58}$$

ist also ein zentrierter Vektor. Man definiert entsprechend die Zentrierungsmatrix $\mathbf{H}$ als

$$\mathbf{H} = \mathbf{I} - \mathbf{M} \tag{9.59}$$

$$\mathbf{H}\mathbf{x} = (\mathbf{I} - \mathbf{M})\mathbf{x} = \mathbf{x} - \mathbf{1}\bar{x} := \mathbf{x}_* \tag{9.60}$$

Zentrierungsmatrix

$$\mathbf{x}'\mathbf{H}\mathbf{x} = \mathbf{x}'\mathbf{x} - N\bar{x}^2 = \sum_n (x_n - \bar{x})^2. \tag{9.61}$$

Es gilt $\mathbf{M}\mathbf{M} = (N^{-1}\mathbf{1}\mathbf{1}')(N^{-1}\mathbf{1}\mathbf{1}') = N^{-2}\mathbf{1}(\mathbf{1}'\mathbf{1})\mathbf{1}' = N^{-1}\mathbf{1}\mathbf{1}' = \mathbf{M}$, da $\mathbf{1}'\mathbf{1} = N$. Man nennt eine Matrix mit der Eigenschaft $\mathbf{M}^2 = \mathbf{M}$ idempotent. Daher gilt $\mathbf{M}(\mathbf{M}\mathbf{x}) = \mathbf{M}\mathbf{x}$, d.h. ein gemittelter Vektor verändert sich nicht mehr. Man spricht auch von einer Projektion (Abb. 9.1) und nennt $\mathbf{M}$ und $\mathbf{H}$ Projektionsmatrizen .

idempotente Matrix

Zusammenfassend gilt

Projektionsmatrizen

$$\mathbf{M} = N^{-1}\mathbf{1}\mathbf{1}' : N \times N \tag{9.62}$$

$$\mathbf{H} = \mathbf{I} - \mathbf{M} \tag{9.63}$$

$$\mathbf{M}^2 = \mathbf{M} \tag{9.64}$$

$$\mathbf{H}^2 = \mathbf{H} \tag{9.65}$$

$$\mathbf{M}\mathbf{H} = \mathbf{O} \tag{9.66}$$

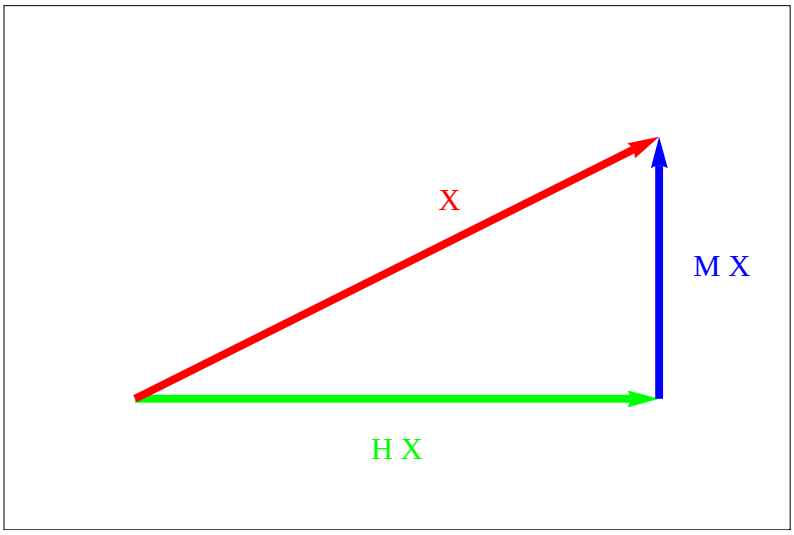


Abbildung 9.1: Orthogonal-Projektion von **X** auf **HX** und **MX**.

Ein zentrierter Vektor $\mathbf{Hx}$ bleibt also ein solcher. Wird er auch noch gemittelt, ergibt sich $\mathbf{0}$, d.h. $\mathbf{MHx} = \mathbf{Ox} = \mathbf{0}$.

Übung 9.3 (Eigenschaften von Projektoren)
Zeigen Sie, daß $\mathbf{H}^2 = \mathbf{H}$ und $\mathbf{MH} = \mathbf{O}$ gilt.
Hinweis: Multiplizieren Sie $(\mathbf{I} - \mathbf{M})(\mathbf{I} - \mathbf{M})$ und nutzen Sie $\mathbf{M}^2 = \mathbf{M}$.
■

Das Produkt von $\mathbf{M}$ mit der gesamten Datenmatrix $\mathbf{X} = [\mathbf{x}_{(1)}, ..., \mathbf{x}_{(p)}] : N \times p$ führt zu einer Matrix von Mittelwerten, d.h.

$$\mathbf{MX} = [\mathbf{1}\bar{x}_1, ..., \mathbf{1}\bar{x}_p] = \mathbf{1}[\bar{x}_1, ..., \bar{x}_p] = \mathbf{1}\bar{\mathbf{x}}'. \tag{9.67}$$

Entsprechend ist $\mathbf{HX} = [\mathbf{Hx}_{(1)}, ..., \mathbf{Hx}_{(p)}] = \mathbf{X}_*$ eine mittelwertskorrigierte Datenmatrix. Explizit gilt $\mathbf{HX} = [\mathbf{x}_{(1)} - \mathbf{1}\bar{x}_1, ..., \mathbf{x}_{(p)} - \mathbf{1}\bar{x}_p]$.
Das Produkt $\mathbf{X}_*'\mathbf{X}_* = \mathbf{X}'\mathbf{HX} : p \times p$ ist somit

$$\mathbf{X}_*'\mathbf{X}_* \;=\; \mathbf{X}'(\mathbf{I} - \mathbf{M})\mathbf{X} = \mathbf{X}'\mathbf{X} - \bar{\mathbf{x}}\mathbf{1}'\mathbf{1}\bar{\mathbf{x}}' = \mathbf{X}'\mathbf{X} - N\bar{\mathbf{x}}\bar{\mathbf{x}}'. \tag{9.68}$$

Dies ist aber $(N - 1)$ mal die Stichprobenkovarianzmatrix. Also gilt die einfache Formel

$$\mathbf{S} = \frac{1}{N-1}\mathbf{X}'\mathbf{H}\mathbf{X}. \tag{9.69}$$

Die übliche Form der Kovarianzmatrix benutzt die Zeilenvektoren $\mathbf{x}_n'$ der Datenmatrix. Schreibt man $\mathbf{X}' = [\mathbf{x}_1, ..., \mathbf{x}_N]$, $\mathbf{X}'\mathbf{H} = [\mathbf{x}_1 - \bar{\mathbf{x}}, ..., \mathbf{x}_N - \bar{\mathbf{x}}]$, wobei $\mathbf{X}'\mathbf{M} = \bar{\mathbf{x}}\mathbf{1}'$ benutzt wurde, so ergibt sich

$$\mathbf{S} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})' \tag{9.70}$$

mit $\bar{\mathbf{x}}' = N^{-1}\mathbf{1}'\mathbf{X} : 1 \times p$.

■

9.5 Rechenregeln für Matrizen

Es wird angenommen, daß alle Matrizen zueinander passen. Dann gelten folgende Rechenregeln

$$\begin{aligned}
\mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} && (9.71)\\
\mathbf{AB} &\neq \mathbf{BA} && (9.72)\\
\mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} && (9.73)\\
\mathbf{IA} &= \mathbf{AI} = \mathbf{A} && (9.74)\\
(\mathbf{A}')' &= \mathbf{A} && (9.75)\\
(\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' && (9.76)\\
(\mathbf{A} + \mathbf{B})' &= \mathbf{A}' + \mathbf{B}' && (9.77)\\
\mathbf{x}'\mathbf{y} &= \mathbf{y}'\mathbf{x} = (\mathbf{x}'\mathbf{y})' && (9.78)
\end{aligned}$$

9.6 Determinanten

Die Determinante einer quadratischen Matrix $\mathbf{A} : I \times I$ wird mit $\det(\mathbf{A})$ oder $|\mathbf{A}|$ bezeichnet und berechnet sich rekursiv aus

Determinante

$$I = 1 \quad : \quad |\mathbf{A}| = a_{11} \tag{9.79}$$

$$I > 1 \quad : \quad |\mathbf{A}| = \sum_{j=1}^{I} a_{ij}\,(-1)^{i+j}|\mathbf{A}_{ij}| \text{ für beliebiges } i. \tag{9.80}$$

(Entwicklung nach der i-ten Zeile). Hierbei ist $|\mathbf{A}_{ij}|$ die Determinante der Matrix, die man durch Streichen der i-ten Zeile und j-ten Spalte erhält (Minor von a_{ij}). $(-1)^{i+j}|\mathbf{A}_{ij}|$ heißt Kofaktor von a_{ij}.

Beispiel 9.4 (Determinante einer 2×2-Matrix)

Setzt man $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, so ist die Entwicklung nach der 1. Zeile:

$$|\mathbf{A}| = a_{11}(-1)^{1+1}|\mathbf{A}_{11}| + a_{12}(-1)^{1+2}|\mathbf{A}_{12}| \tag{9.81}$$

$$= a_{11}a_{22} - a_{12}a_{21} \tag{9.82}$$

■

Der Stoff wird in Aufgabe 9.5 vertieft.

Übung 9.4 (Determinante einer 3×3-Matrix)

Entwickeln Sie die Matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

nach der 1. Zeile.

Hinweis: Verwenden Sie das Resultat des vorigen Beispiels.

Allgemein gilt:

$$\begin{aligned}
\mathbf{A} &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\
|\mathbf{A}| &= a_{13}\left(a_{21}a_{32} - a_{22}a_{31}\right) \\
&+ a_{12}\left(a_{23}a_{31} - a_{21}a_{33}\right) \\
&+ a_{11}\left(a_{22}a_{33} - a_{23}a_{32}\right)
\end{aligned}$$

■

Rechenregeln:

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}| = |\mathbf{BA}| \tag{9.83}$$

$$|c\mathbf{A}| = c^I |\mathbf{A}| \tag{9.84}$$

$$|\mathbf{D}| = d_1 d_2 ... d_I \tag{9.85}$$

$$|\mathbf{A}| = \lambda_1 \lambda_2 ... \lambda_I \tag{9.86}$$

Hierbei ist $\mathbf{B} : I \times I$ und $\mathbf{D}$ ist eine Diagonalmatrix (nur Werte auf der Hauptdiagonalen) oder eine Dreiecksmatrix (alle Werte oberhalb oder unterhalb der Hauptdiagonalen sind Null). In der letzten Formel sind λ_i die Eigenwerte von $\mathbf{A}$ (siehe unten).

Der Stoff wird in Aufgabe 9.6 vertieft.

Übung 9.5 (Jacobi-Term der Normalverteilung)
Zeigen Sie, daß $|2\pi\mathbf{\Sigma}|^{-1/2} = (2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2}$ gilt ($\mathbf{\Sigma} : p \times p$).
■

9.7 Inverse Matrix

Betrachtet man das lineare Gleichungssystem $\mathbf{b} = \mathbf{Ax}$, so kann man nach $\mathbf{x}$ auflösen, wenn es eine inverse Matrix gibt, d.h. $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Man darf jedoch nicht einfach die Elemente von $\mathbf{A}$ invertieren.
Eine Matrix $\mathbf{A}^{-1}$ mit der Eigenschaft

inverse Matrix

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{9.87}$$

heißt inverse Matrix (Inverse) von $\mathbf{A}$.

Übung 9.6 (Elementweiser Kehrwert)

1. Multiplizieren Sie $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ mit $\begin{bmatrix} 1 & 1/2 \\ 1/3 & 1/4 \end{bmatrix}$.
2. Multiplizieren Sie $\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ mit $\begin{bmatrix} 1 & 0 \\ 0 & 1/4 \end{bmatrix}$.

■

Die inverse Matrix läßt sich wie folgt berechnen:

$$(\mathbf{A}^{-1})_{ij} = \frac{(-1)^{i+j}|\mathbf{A}_{ji}|}{|\mathbf{A}|}. \tag{9.88}$$

nichtsinguläre Matrix

Aus obiger Formel sieht man, daß $|\mathbf{A}| \neq 0$ sein muß (nichtsinguläre Matrix).

Beispiel 9.5 (Inverse einer 2×2-Matrix)

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{9.89}$$

■

Übung 9.7

Multiplizieren Sie mit $\mathbf{A}^{-1}$ mit $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

■

Der Stoff wird in Aufgabe 9.7 vertieft.

Rechenregeln:

$$
\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} && (9.90)\\
(c\mathbf{A})^{-1} &= c^{-1}\mathbf{A}^{-1} && (9.91)\\
\mathbf{D}^{-1} &= \mathrm{diag}(d_1^{-1}, d_2^{-1}, ...d_I^{-1}) && (9.92)\\
|\mathbf{A}^{-1}| &= |\mathbf{A}|^{-1} && (9.93)\\
(\mathbf{A}^{-1})' &= (\mathbf{A}')^{-1} && (9.94)
\end{aligned}
$$

Übung 9.8 (Beweis)
Beweisen Sie obige Aussagen durch Ausmultiplizieren sowie durch die Rechenregel (9.83).
■

Der Stoff wird in Aufgabe 9.8 vertieft.

Übung 9.9 (Inverse der Kovarianzmatrix)
Die Kovarianzmatrix kann als Produkt der Korrelationsmatrix mit den Standardabweichungen geschrieben werden.
Zeigen Sie, daß für $\mathbf{\Sigma} = \mathbf{DPD}$ gilt: $\mathbf{\Sigma}^{-1} = \mathbf{D}^{-1}\mathbf{P}^{-1}\mathbf{D}^{-1}$.
■

Übung 9.10
Zeigen Sie (9.22) durch Ausmultiplizieren von $\mathbf{E}$ und $\mathbf{E}^{-1}$.
■

9.8 Spur

Die Spur (trace) einer quadratischen Matrix ist die Summe der Diagonalelemente, d.h.

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{I} a_{ii}. \tag{9.95}$$

Der Stoff wird in Aufgabe 9.9 vertieft.

Rechenregeln:

$$\begin{aligned}
\mathrm{tr}(\mathbf{CD}) &= \mathrm{tr}(\mathbf{DC}) && (9.96)\\
\mathrm{tr}(\mathbf{CC}') &= \mathrm{tr}(\mathbf{C}'\mathbf{C}) = \sum_{ij} c_{ij}^2 := ||\mathbf{C}||^2 && (9.97)\\
\mathrm{tr}(\mathbf{A}+\mathbf{B}) &= \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}) && (9.98)\\
\mathrm{tr}(c\mathbf{A}) &= c\,\mathrm{tr}(\mathbf{A}) && (9.99)\\
\mathrm{tr}(\mathbf{x}'\mathbf{y}) &= \mathrm{tr}(\mathbf{y}\mathbf{x}') = \mathbf{x}'\mathbf{y} && (9.100)\\
\mathrm{tr}(\mathbf{x}'\mathbf{C}\mathbf{y}) &= \mathrm{tr}(\mathbf{C}\mathbf{y}\mathbf{x}') = \mathbf{x}'\mathbf{C}\mathbf{y} && (9.101)\\
\sum_n \mathrm{tr}(\mathbf{x}_n'\mathbf{C}\mathbf{y}_n) &= \mathrm{tr}(\mathbf{C}\sum_n \mathbf{y}_n\mathbf{x}_n') = \sum_n \mathbf{x}_n'\mathbf{C}\mathbf{y}_n && (9.102)\\
\mathrm{tr}(\mathbf{A}) &= \sum_i \lambda_i && (9.103)\\
\mathrm{tr}(\mathbf{P}) &= \mathrm{rg}(\mathbf{P}) && (9.104)
\end{aligned}$$

Bemerkungen:

1. Die Dimensionen der Matrizen sind: $\mathbf{A}, \mathbf{B} : I \times I$, $\mathbf{C} : I \times J$, $\mathbf{D} : J \times I$
2. $||\mathbf{C}||$ ist eine Matrix-Norm (vgl. $\mathbf{x}'\mathbf{x} = ||\mathbf{x}||^2 = \mathrm{tr}(\mathbf{x}\mathbf{x}')$).
3. Die Werte λ_i sind die Eigenwerte der Matrix $\mathbf{A}$.
4. $\mathbf{P}^2 = \mathbf{P}$ ist eine Projektionsmatrix und $\mathrm{rg}(\mathbf{P})$ ist der Rang der Matrix (Zahl der linear unabhängigen Zeilen- oder Spaltenvektoren, die die Matrix bilden; Abs. 9.11).

Übung 9.11 (Spur der Zentrierungsmatrix H)
Zeigen Sie, daß $\mathrm{tr}[\mathbf{M}] = 1$ und $\mathrm{tr}[\mathbf{H}] = N - 1$ gilt.
Hinweis:
Setzen Sie $\mathbf{H} = \mathbf{I} - \mathbf{M}$ ein und benutzen Sie die Rechenregeln.
■

9.9 Eigenwerte und Eigenvektoren

Die Eigenwertgleichung

$$\mathbf{A}\boldsymbol{\psi} = \lambda\boldsymbol{\psi}, \tag{9.105}$$

Eigenwertgleichung

$\lambda =$ Zahl, läßt sich so interpretieren, daß ein Vektor $\boldsymbol{\psi}$ gesucht wird, dessen Richtung sich nicht verändert, wenn er mit der Matrix $\mathbf{A}$ multipliziert wird. Schreibt man

$$(\mathbf{A} - \lambda\mathbf{I})\boldsymbol{\psi} = \mathbf{0}, \tag{9.106}$$

so muß

$$|\mathbf{A} - \lambda\mathbf{I}| := q(\lambda) = 0 \tag{9.107}$$

gelten, damit eine Lösung $\boldsymbol{\psi} \neq \mathbf{0}$ gefunden werden kann. Man erhält also eine Gleichung I-ter Ordnung in λ. Das Polynom $q(\lambda)$ heißt charakteristisches Polynom.

charakteristisches Polynom

Die Lösungen $\lambda_i, i = 1, ..., I$ sind die Eigenwerte von $\mathbf{A}$. Die zugehörigen Vektoren $\boldsymbol{\psi}_i$ mit

Eigenwerte

$$\mathbf{A}\boldsymbol{\psi}_i = \lambda_i\boldsymbol{\psi}_i \tag{9.108}$$

Eigenvektoren

heißen Eigenvektoren.
Schreibt man $q(\lambda) = \prod_i(\lambda_i - \lambda)$, so ergibt sich

$$q(0) = |\mathbf{A}| = \prod_i \lambda_i. \tag{9.109}$$

Beispiel 9.6 (Eigenwerte einer Korrelationsmatrix)
Für die (theoretische) Korrelationsmatrix[5]

$$\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{9.110}$$

ergeben sich die Eigenwerte aus der quadratischen Gleichung ($I = 2$ Lösungen)

$$\det(\begin{bmatrix} 1-\lambda & \rho \\ \rho & 1-\lambda \end{bmatrix}) = 0 = (1-\lambda)^2 - \rho^2 \tag{9.111}$$

$$\lambda_{1,2} = 1 \pm \rho \tag{9.112}$$

Die Summe der Eigenwerte ist also $2 = \mathrm{tr}(\mathbf{R})$ = Summe der Diagonale := Spur = trace.
Ganz allgemein gilt für Korrelationsmatrizen

$$\sum_i \mathbf{R}_{ii} := \mathrm{tr}(\mathbf{R}) = \sum_i \lambda_i = I. \tag{9.113}$$

Die Eigenvektoren ergeben sich aus den Bedingungen

$$(\mathbf{R} - \lambda_1 \mathbf{I})\boldsymbol{\psi}_1 = \begin{bmatrix} -\rho & \rho \\ \rho & -\rho \end{bmatrix} \begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{9.114}$$

$$(\mathbf{R} - \lambda_2 \mathbf{I})\boldsymbol{\psi}_2 = \begin{bmatrix} \rho & \rho \\ \rho & \rho \end{bmatrix} \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{9.115}$$

Etwa löst

$$\begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{9.116}$$

$$\begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{9.117}$$

$$\tag{9.118}$$

obige Gleichungen. Das Betrags-Quadrat der Vektoren ist $[1,1][1,1]' = 2, [1,-1][1,-1]' = 2$, sodaß man

$$\boldsymbol{\psi}_1 = \begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = 1/\sqrt{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{9.119}$$

$$\boldsymbol{\psi}_2 = \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = 1/\sqrt{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{9.120}$$

als orthonormierte Eigenvektoren findet, d.h. sie sind orthogonal und auf die Länge 1 normiert.

orthonormierte Eigenvektoren

Es ist wichtig, daß die Eigenvektoren gar nicht von der Korrelation ρ abhängen. Sie zeigen in Richtung der Winkelhalbierenden der Quadranten (Abb. 9.2).

Der Fall $\rho = 0$ erfordert eine spezielle Behandlung. In diesem Fall ist $\mathbf{R} = \mathbf{I}_2$ und man hat den doppelten Eigenwert $\lambda_{1,2} = 1$. Die Eigenwertgleichung $\mathbf{R}\boldsymbol{\psi} = \boldsymbol{\psi} = \lambda\boldsymbol{\psi}$ wird von beliebigen Vektoren erfüllt. Man kann also bei obiger Wahl bleiben.

■

Übung 9.12

Zeigen Sie, daß $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2$ orthonormiert sind.

■

Schreibt man alle Eigenvektoren in eine Matrix $\mathbf{P} = [\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_I]$, so kann die Eigenwertgleichung (9.108) in der Form

$$\mathbf{AP} = [\lambda_1\boldsymbol{\psi}_1, ..., \lambda_I\boldsymbol{\psi}_I] = \mathbf{P}\mathrm{diag}(\lambda_1, ..., \lambda_I) = \mathbf{P}\boldsymbol{\Lambda} \tag{9.121}$$

geschrieben werden. Alternativ ergibt sich

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} \tag{9.122}$$

$$\boldsymbol{\Lambda} = \mathbf{P}^{-1}\mathbf{AP}. \tag{9.123}$$

Bei symmetrischen Matrizen $\mathbf{A} = \mathbf{A}'$ muß sogar $\mathbf{P}^{-1} = \mathbf{P}'$ (orthogonale Matrix) gelten mit

symmetrische Matrix

$$\mathbf{PP}' = \mathbf{P}'\mathbf{P} = \mathbf{I}. \tag{9.124}$$

Dies bedeutet, daß die Vektoren $\boldsymbol{\psi}_i$ orthogonal sind. Das Matrix-Produkt $\mathbf{P}'\mathbf{P}$ setzt sich ja aus den Skalarprodukten $\boldsymbol{\psi}_i'\boldsymbol{\psi}_j = \delta_{ij}$ zusammen. Hier wurde die Notation

$$\delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{sonst} \end{cases}, \tag{9.125}$$

Kronecker-Delta-Symbol

benutzt.

[5] Die Notation $\mathbf{R}$ dient zur Unterscheidung von der Matrix der Eigenvektoren $\mathbf{P}$.

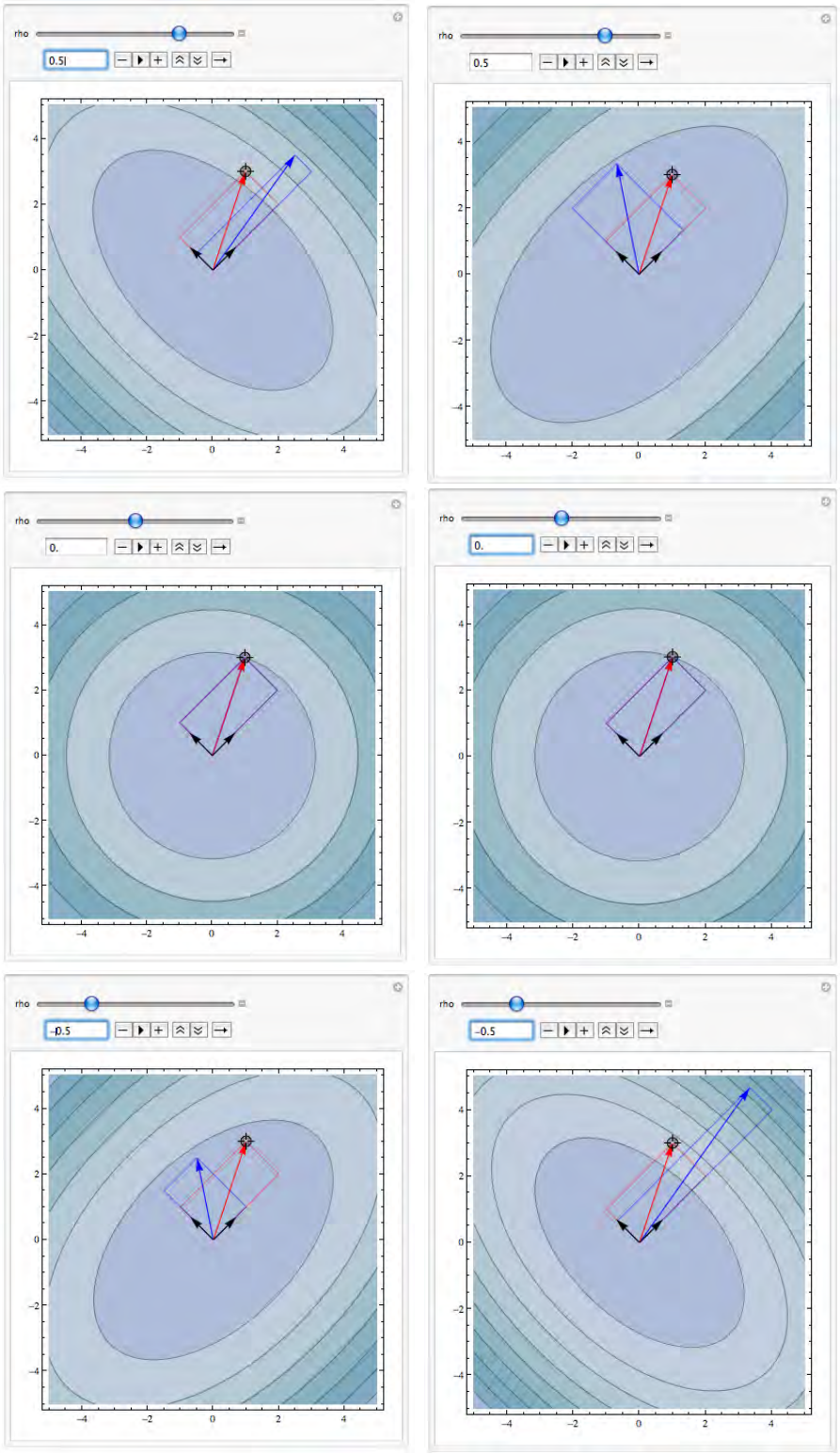


Abbildung 9.2: Eigenvektoren zur Korrelations-Matrix **R**. Die Hauptachsen zeigen in Richtung der Winkelhalbierenden (schwarze Pfeile). Ein Vektor $\mathbf{x} = [1, 3]'$ (rot) und seine Projektionen $y_i = \boldsymbol{\psi}_i'\mathbf{x}$ sind ebenfalls dargestellt. Die blauen Pfeile repräsentieren $\mathbf{Rx}$ (links) und $\mathbf{R}^{-1}\mathbf{x}$ (rechts). Die Ellipsen sind durch die quadratische Form $\mathbf{x}'\mathbf{Rx} = r^2$ (links) und $\mathbf{x}'\mathbf{R}^{-1}\mathbf{x} = r^2$ (rechts) gegeben. Von oben: $\rho = 0.5, 0, -0.5$.

Für *symmetrische Matrizen* gilt zusammenfassend:

1. Die Eigenwerte λ_i sind reell.
2. Die Eigenvektoren $\boldsymbol{\psi}_i$ sind paarweise orthogonal.
3. Es gilt

$$\begin{aligned} \mathbf{A} &= \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' && \text{(Spektralzerlegung)} && (9.126)\\ \boldsymbol{\Lambda} &= \mathbf{P}'\mathbf{A}\mathbf{P} && \text{(Diagonalisierung)} && (9.127)\\ \mathbf{P}\mathbf{P}' &= \mathbf{P}'\mathbf{P} = \mathbf{I} && \text{(orthogonale Matrix)} && (9.128) \end{aligned}$$

Hauptachsen-Transformation

Die Spektraldarstellung

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \sum_i \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \tag{9.129}$$

hat ihren Namen von den Eigenwerten (Spektrum) der Matrix. Die äußeren Produkte $\mathbf{P}_i = \boldsymbol{\psi}_i\boldsymbol{\psi}_i'$ erfüllen $\mathbf{P}_i^2 = \mathbf{P}_i$, $\mathbf{P}_i\mathbf{P}_j = \mathbf{O}, i \neq j$, sind also Projektor-Matrizen auf paarweise orthogonale Unterräume des I-dimensionalen Raums. Ein Vektor $\mathbf{x}$ wird auf $\mathbf{P}_i\mathbf{x} = \boldsymbol{\psi}_i(\boldsymbol{\psi}_i'\mathbf{x})$ abgebildet, zeigt also in Richtung des i-ten Eigenvektors.

Betrachtet man die quadratische Form (für einen beliebigen Vektor $\mathbf{x}$)

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \mathbf{x}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{x} = \mathbf{y}'\boldsymbol{\Lambda}\mathbf{y} && (9.130)\\ \mathbf{y} &= \mathbf{P}'\mathbf{x}, && (9.131) \end{aligned}$$

so sind die Komponenten von $\mathbf{y}$ die Projektionen auf die Eigenvektoren $\boldsymbol{\psi}_i$, d.h. $y_i = \boldsymbol{\psi}_i'\mathbf{x}$. Im Koordinatensystem der Eigenvektoren ist also die quadratische Form diagonal, d.h. es gilt

$$\sum_{ij} x_i a_{ij} x_j = \sum_i y_i \lambda_i y_i = \sum_i \lambda_i y_i^2. \tag{9.132}$$

Die Wirkung einer Matrix auf einen Vektor kann man sich mit Hilfe der Spektraldarstellung so vorstellen:

$$\begin{aligned}
\mathbf{A}\mathbf{x} &= \sum_i \lambda_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \mathbf{x} && (9.133)\\
&= \sum_i \boldsymbol{\psi}_i \lambda_i (\boldsymbol{\psi}_i' \mathbf{x}) = \sum_i \boldsymbol{\psi}_i \lambda_i y_i && (9.134)\\
\mathbf{x} &= \sum_i \boldsymbol{\psi}_i (\boldsymbol{\psi}_i' \mathbf{x}) = \sum_i \boldsymbol{\psi}_i y_i && (9.135)
\end{aligned}$$

Die Komponenten des Vektors im System der Eigenvektoren, d.h. $y_i = \boldsymbol{\psi}_i' \mathbf{x}$ (Projektionen) werden also durch die Eigenwerte λ_i skaliert. Dies ist in Abb. 9.2 durch die roten ($\mathbf{x}$) und blauen Pfeile ($\mathbf{A}\mathbf{x}$) und deren Projektionen auf die Eigenvektoren $\boldsymbol{\psi}_i$ dargestellt.

Die Darstellung des Vektors im System der Eigenvektoren (Glg. 9.135) ergibt sich aus der Identität $\mathbf{x} = \mathbf{P}\mathbf{P}'\mathbf{x} = \sum_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' \mathbf{x}$. Man kann auch schreiben $\sum_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i' = \mathbf{I} = \mathbf{P}\mathbf{P}'$. Daraus findet man $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{P}\mathbf{P}'\mathbf{x} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{x} = \sum_i \boldsymbol{\psi}_i \lambda_i (\boldsymbol{\psi}_i' \mathbf{x})$.

Wenn es sich bei $\mathbf{A}$ um eine Kovarianzmatrix $\boldsymbol{\Sigma}$ handelt, sind die Eigenwerte λ sogar positiv oder 0 (vgl. Glg. 9.45).

Die Komponenten des Vektors

Hauptkomponenten

$$\mathbf{y} = \mathbf{P}'\mathbf{x} \tag{9.136}$$

werden auch als Hauptkomponenten bezeichnet. Es handelt sich um die Projektion des Vektors $\mathbf{x}$ auf die Eigenvektoren $\boldsymbol{\psi}_i$, also $y_i = \boldsymbol{\psi}_i' \mathbf{x}$. Ist $\mathbf{x}$ zufällig mit $\mathrm{Var}(\mathbf{x}) = \boldsymbol{\Sigma}$, so gilt

$$\mathrm{Var}(\mathbf{y}) = \mathrm{Var}(\mathbf{P}'\mathbf{x}) = \mathbf{P}'\boldsymbol{\Sigma}\mathbf{P} = \boldsymbol{\Lambda}. \tag{9.137}$$

Die Hauptkomponenten sind also unkorreliert und haben die Varianz $\mathrm{Var}(y_i) = \lambda_i$.

Beispiel 9.7 (Eigenwerte einer Korrelationsmatrix, Fortsetzung)

Die orthogonale Matrix der Eigenvektoren lautet explizit

$$\mathbf{P} = 2^{-1/2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{9.138}$$

$$\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2. \tag{9.139}$$

Im gedrehten Koordinatensystem gilt daher

$$\mathbf{y} = \mathbf{P}'\mathbf{x} = \begin{bmatrix} \boldsymbol{\psi}_1'\mathbf{x} \\ \boldsymbol{\psi}_2'\mathbf{x} \end{bmatrix} = 2^{-1/2} \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}. \tag{9.140}$$

Nimmt man an, daß $\mathbf{x}$ ein zufälliger Vektor ist mit Kovarianzmatrix $\mathrm{Var}(\mathbf{x}) = \mathbf{R}$, so gilt

$$\mathrm{Var}(\mathbf{y}) = \mathbf{P}'\mathbf{R}\mathbf{P} = \boldsymbol{\Lambda} = \mathrm{diag}(1+\rho, 1-\rho). \tag{9.141}$$

Daher sind die Koordinaten (Hauptkomponenten) y_1, y_2 unkorreliert.
Die quadratische Form (Ellipse) der Matrix $\mathbf{R}$

$$\mathbf{x}'\mathbf{R}\mathbf{x} = \sum_{ij} x_i \rho_{ij} x_j = x_1^2 + 2\rho x_1 x_2 + x_2^2 \tag{9.142}$$

ist diagonal im gedrehten System:

$$\begin{aligned} \mathbf{x}'\mathbf{R}\mathbf{x} &= \mathbf{x}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{x} && (9.143) \\ &= \mathbf{y}'\boldsymbol{\Lambda}\mathbf{y} && (9.144) \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 && (9.145) \\ &= (1+\rho)y_1^2 + (1-\rho)y_2^2. && (9.146) \end{aligned}$$

Abb. 9.2 zeigt die quadratische Form der Matrix $\mathbf{R}$ für die Werte $\rho = -0.5, 0, 0.5$. Die Ellipse hat für positives ρ in Richtung $\boldsymbol{\psi}_1$ eine kürzere Hauptachse als in Richtung $\boldsymbol{\psi}_2$, da $1+\rho > 1-\rho$. Dies erscheint etwas überraschend, da man bei positiver Korrelation eine Ellipse in Richtung der 45°-Achse erwarten würde. Man muß sich aber klarmachen, daß im Fall der Normalverteilung die quadratische Form

$$\begin{aligned} \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} &= \mathbf{x}'\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'\mathbf{x} && (9.147) \\ &= \mathbf{y}'\boldsymbol{\Lambda}^{-1}\mathbf{y} && (9.148) \\ &= y_1^2/\lambda_1 + y_2^2/\lambda_2 && (9.149) \\ &= y_1^2/(1+\rho) + y_2^2/(1-\rho) && (9.150) \end{aligned}$$

betrachtet wird. Diese hat die gewohnte Form und Richtung (Abb. 9.2, rechte Spalte).

■

Im vorigen Beispiel wurde anstatt von $\mathbf{R}$ auch die inverse Matrix $\mathbf{R}^{-1}$ betrachtet. In der Tat hat $\mathbf{R}^{-1}$ die gleichen Eigenvektoren, jedoch die inversen Eigenwerte von $\mathbf{R}$.

Man kann ganz allgemein schreiben

$$\mathbf{A}^{-1}\boldsymbol{\psi}_i = \lambda_i^{-1}\boldsymbol{\psi}_i \tag{9.151}$$

wenn man (9.108) mit $\mathbf{A}^{-1}$ multipliziert. Daher gilt (bei symmetrischen Matrizen)

$$\mathbf{A}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}' \tag{9.152}$$

$$\boldsymbol{\Lambda}^{-1} = \mathbf{P}'\mathbf{A}^{-1}\mathbf{P}. \tag{9.153}$$

Die entsprechende quadratische Form ist

$$\mathbf{x}'\mathbf{A}^{-1}\mathbf{x} = \mathbf{x}'\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'\mathbf{x} = \mathbf{y}\boldsymbol{\Lambda}^{-1}\mathbf{y} = \sum_i y_i^2/\lambda_i. \tag{9.154}$$

Beispiel 9.8 (Normalverteilung)

Die Normalverteilung $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ hat die Dichte

$$\phi(\mathbf{x}) = |2\pi\boldsymbol{\Lambda}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{y}'\boldsymbol{\Lambda}^{-1}\mathbf{y}\right\} \tag{9.155}$$

$$= (2\pi^p\lambda_1...\lambda_p)^{-1/2}\exp\left\{-\frac{1}{2}\sum_i y_i^2/\lambda_i\right\} \tag{9.156}$$

$$\mathbf{y} = \mathbf{P}'(\mathbf{x}-\boldsymbol{\mu}) \tag{9.157}$$

Hierbei wurde die Kovarianzmatrix als $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ bzw. $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'$ dargestellt. Daher sind die Komponenten des Zufallsvektors $\mathbf{y}$ unkorreliert und es gilt $\mathrm{Var}(\mathbf{y}) = \mathbf{P}'\boldsymbol{\Sigma}\mathbf{P} = \boldsymbol{\Lambda}$.

Die Determinante lautet $|\boldsymbol{\Sigma}| = |\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'| = |\boldsymbol{\Lambda}\mathbf{P}'\mathbf{P}| = |\boldsymbol{\Lambda}|$.

Setzt man $\mathbf{z} = \boldsymbol{\Lambda}^{-1/2}\mathbf{y} = \boldsymbol{\Lambda}^{-1/2}\mathbf{P}'(\mathbf{x}-\boldsymbol{\mu})$, so gilt $\mathrm{Var}(\mathbf{z}) = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}$. $\mathbf{z}$ ist also eine standardisierte Variable. Auflösen nach $\mathbf{x}$ ergibt $\mathbf{x} = \boldsymbol{\mu} + (\boldsymbol{\Lambda}^{-1/2}\mathbf{P}')^{-1}\mathbf{z} = \boldsymbol{\mu} + \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$.

Daher kann $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}$ als Matrixwurzel gesetzt werden. Es gilt, wie behauptet, $\boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})' = \mathbf{P}\boldsymbol{\Lambda}^{1/2}(\mathbf{P}\boldsymbol{\Lambda}^{1/2})' = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \boldsymbol{\Sigma}$.

■

Für Potenzen von $\mathbf{A}$ gilt

$$\mathbf{A}^n \boldsymbol{\psi}_i = \lambda_i^n \boldsymbol{\psi}_i \tag{9.158}$$

wenn man (9.108) mit $\mathbf{A}^{n-1}$ multipliziert. Daher gilt (bei symmetrischen Matrizen)

$$\mathbf{A}^n \mathbf{P} = \mathbf{P} \boldsymbol{\Lambda}^n \tag{9.159}$$

$$\mathbf{A}^n = \mathbf{P} \boldsymbol{\Lambda}^n \mathbf{P}'. \tag{9.160}$$

Für eine Funktion $f(\mathbf{A}) = \sum_j c_j \mathbf{A}^j$ kann man schreiben

$$f(\mathbf{A}) = \sum_j c_j \mathbf{A}^j \tag{9.161}$$

$$= \sum_j c_j \mathbf{P} \boldsymbol{\Lambda}^j \mathbf{P}' \tag{9.162}$$

$$= \mathbf{P}(\sum_j c_j \boldsymbol{\Lambda}^j)\mathbf{P}' \tag{9.163}$$

$$= \mathbf{P} f(\boldsymbol{\Lambda}) \mathbf{P}' \tag{9.164}$$

Eine Matrixfunktion, etwa $\sqrt{\mathbf{A}}$, wird entsprechend durch

$$f(\mathbf{A}) = \mathbf{P} f(\boldsymbol{\Lambda}) \mathbf{P}' \tag{9.165}$$

Matrix-Funktion

definiert. Die Determinante erfüllt

$$|f(\mathbf{A})| = |\mathbf{P} f(\boldsymbol{\Lambda}) \mathbf{P}'| = |f(\boldsymbol{\Lambda})| = \prod_i f(\lambda_i). \tag{9.166}$$

Beispiel 9.9 (Matrix-Wurzel)
Die Wurzel aus der Kovarianz-Matrix $\boldsymbol{\Sigma}$ kann durch

$$\boldsymbol{\Sigma}^{1/2} = \mathbf{P} \boldsymbol{\Lambda}^{1/2} \mathbf{P}' \tag{9.167}$$

definiert werden. Sie ist symmetrisch und es gilt $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{\Sigma}$.
Man kann auch die asymmetrische Form

$$\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2} \tag{9.168}$$

verwenden. Es gilt $\mathbf{\Sigma}^{1/2}(\mathbf{\Sigma}^{1/2})' = \mathbf{P}\mathbf{\Lambda}^{1/2}(\mathbf{P}\mathbf{\Lambda}^{1/2})' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}' = \mathbf{\Sigma}$.

■

9.10 Kronecker-Produkte

Design-Matrizen im linearen Modell lassen sich mit Hilfe von Kronecker-Produkten aufbauen. Dabei wird jedes Matrix-Element von $\mathbf{A} : I \times J$ durch die Matrix $a_{ij}\mathbf{B} : K \times L$ ersetzt, also

Kronecker-Produkt

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & ... & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & ... & a_{2J}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & ... & a_{IJ}\mathbf{B} \end{bmatrix} : IK \times JL \tag{9.169}$$

In Komponenten gilt

$$(\mathbf{A} \otimes \mathbf{B})_{ik,jl} = a_{ij}b_{kl}. \tag{9.170}$$

Die Zeilen-Indizes i, k ergeben den neuen Zeilenindex ik, die Spalten-Indizes j, l den neuen Spaltenindex jl. Dabei wird der Multi-Index $ik = (i, k)$ von $1, ..., I \cdot K$ linear durchgezählt (analog für jl).

Beispiel 9.10 (Design-Matrix für Dummy-Codierung)
Die Design-Matrix in Dummy-Codierung lautet $\mathbf{I}_I \otimes \mathbf{1}_J$. Konkret ergibt sich für $I = 3, J = 2$ die Matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Der Multi-Index (i, k), $i = 1, ..., 3, k = 1, 2$, d.h. $(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)$ wird von 1,...,6 linear durchgezählt.

■

Folgende Rechenregeln werden im Text benutzt:

$$\begin{aligned}
\mathbf{A} &= \mathbf{A} \otimes 1 = 1 \otimes \mathbf{A} && (9.171)\\
[\mathbf{A}, \mathbf{B}] \otimes \mathbf{C} &= [\mathbf{A} \otimes \mathbf{C}, \mathbf{B} \otimes \mathbf{C}] && (9.172)\\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{AC}) \otimes (\mathbf{BD}) && (9.173)\\
(\mathbf{A} \otimes \mathbf{B})' &= \mathbf{A}' \otimes \mathbf{B}' && (9.174)
\end{aligned}$$

Daraus folgt ($\mathbf{d}$ = Vektor)

$$(\mathbf{A} \otimes \mathbf{d})\mathbf{B} = \mathbf{AB} \otimes \mathbf{d} \qquad (9.175)$$

$$\begin{aligned}
(\mathbf{A} \otimes \mathbf{d})[\mathbf{B}, \mathbf{C}] &= [\mathbf{AB} \otimes \mathbf{d}, \mathbf{AC} \otimes \mathbf{d}] && (9.176)\\
&= [\mathbf{AB}, \mathbf{AC}] \otimes \mathbf{d}. && (9.177)
\end{aligned}$$

Dies ergibt sich sofort aus

$$\begin{aligned}
(\mathbf{A} \otimes \mathbf{d})\mathbf{B} &= (\mathbf{A} \otimes \mathbf{d})(\mathbf{B} \otimes 1) && (9.178)\\
&= \mathbf{AB} \otimes \mathbf{d}1 && (9.179)\\
&= \mathbf{AB} \otimes \mathbf{d} && (9.180)
\end{aligned}$$

und

$$\begin{aligned}
(\mathbf{A} \otimes \mathbf{d})[\mathbf{B}, \mathbf{C}] &= (\mathbf{A} \otimes \mathbf{d})([\mathbf{B}, \mathbf{C}] \otimes 1) && (9.181)\\
&= \mathbf{A}[\mathbf{B}, \mathbf{C}] \otimes \mathbf{d}1 && (9.182)\\
&= [\mathbf{AB}, \mathbf{AC}] \otimes \mathbf{d} && (9.183)\\
&= [\mathbf{AB} \otimes \mathbf{d}, \mathbf{AC} \otimes \mathbf{d}] && (9.184)
\end{aligned}$$

Übung 9.13 (Bitte Glg. 9.171–9.174 nachrechnen)
Hinweise:

i) Schreiben Sie $\sum_{jl}(\mathbf{A} \otimes \mathbf{B})_{ik,jl}(\mathbf{C} \otimes \mathbf{D})_{jl,mn}$
in Komponenten als $\sum_{jl} a_{ij} b_{kl} c_{jm} d_{ln} = \sum_j a_{ij} c_{jm} \sum_l b_{kl} d_{ln}$.

ii) $a_{ij} = a'_{ji}$.

■

Der Stoff wird in Aufgabe 9.10 vertieft.

Lassen sich die quadratischen Matrizen $\mathbf{A}$ und $\mathbf{B}$ mit Hilfe der Spektralzerlegung als $\mathbf{A} = \mathbf{P\Lambda P}^{-1}$ und $\mathbf{B} = \mathbf{PMP}^{-1}$ schreiben, so gilt

$$\mathbf{A} \otimes \mathbf{B} = (\mathbf{P} \otimes \mathbf{P})(\mathbf{\Lambda} \otimes \mathbf{M})(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}) \tag{9.185}$$

und

$$\det(\mathbf{A} \otimes \mathbf{B}) = \det(\mathbf{P} \otimes \mathbf{P}) \det(\mathbf{\Lambda} \otimes \mathbf{M}) \det(\mathbf{P}^{-1} \otimes \mathbf{P}^{-1}) \tag{9.186}$$

$$= \det(\mathbf{\Lambda} \otimes \mathbf{M}) = \prod_{ij} \lambda_i \mu_j \tag{9.187}$$

Für symmetrische und positiv definite Matrizen $\mathbf{A} = \mathbf{P\Lambda P}'$ und $\mathbf{B} = \mathbf{PMP}'$ ist die Matrix $\mathbf{A} \otimes \mathbf{B}$ folglich positiv definit, da

$$Q = \mathbf{x}'(\mathbf{A} \otimes \mathbf{B})\mathbf{x} = \mathbf{x}'(\mathbf{P} \otimes \mathbf{P})(\mathbf{\Lambda} \otimes \mathbf{M})(\mathbf{P}' \otimes \mathbf{P}')\mathbf{x} \tag{9.188}$$

$$= \mathbf{y}'(\mathbf{\Lambda} \otimes \mathbf{M})\mathbf{y} = \sum_{ij} y_{ij}^2 \lambda_i \mu_j \tag{9.189}$$

und $\lambda_i \mu_j > 0$ gilt.

9.11 Rang einer Matrix

Die Zeilen oder Spalten einer Matrix $\mathbf{A} : I \times J$ können als Vektoren aufgefaßt werden. Ist z.B. $I = 2$ und $J = 3$, so kann die dritte Spalte als Linearkombination der ersten beiden Spalten geschrieben werden. Etwa gilt für die dritte Spalte von $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ die Darstellung

$$\begin{bmatrix} 3 \\ 6 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 5 \end{bmatrix}. \tag{9.190}$$

Die Spaltenvektoren sind also linear abhängig.

Man definiert: Die Vektoren $\mathbf{x}_1, ..., \mathbf{x}_J$ sind linear abhängig, wenn die Gleichung

$$\lambda_1 \mathbf{x}_1 + ... + \lambda_J \mathbf{x}_J \quad = \quad \mathbf{0} \tag{9.191}$$

lineare Unabhängigkeit

von Zahlen λ_j erfüllt wird, die nicht alle 0 sind. Ansonsten sind die $\mathbf{x}_j$ linear unabhängig. Schreibt man $\mathbf{A} = [\mathbf{x}_1, ..., \mathbf{x}_J]$ und $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_J]'$ so gilt $\mathbf{A}\boldsymbol{\lambda} = \mathbf{0}$, $\boldsymbol{\lambda} \neq \mathbf{0}$.

Rang

Als Rang $\mathrm{rg}(\mathbf{A})$ einer Matrix wird die maximale Zahl linear unabhängiger Spalten (Zeilen) bezeichnet.

Im obigen Beispiel ist offenbar $\mathrm{rg}(\mathbf{A}) = 2$. Die Vektoren $\mathbf{x}_1$ und $\mathbf{x}_2$ sind linear unabhängig.

Eigenschaften:

1. $\mathrm{rg}(\mathbf{A}) = \mathrm{rg}(\mathbf{A}\mathbf{A}') = \mathrm{rg}(\mathbf{A}'\mathbf{A})$
2. $\mathrm{rg}(\mathbf{A}) = \mathrm{rg}(\mathbf{B}\mathbf{A}\mathbf{C})$ für nichtsinguläre Matrizen $\mathbf{B}$, $\mathbf{C}$.
3. Der Rang einer symmetrischen Matrix ist gleich der Zahl der Eigenwerte ungleich 0.
4. Daher ist der Rang $\mathrm{rg}(\mathbf{P})$ einer Projektionsmatrix ($\mathbf{P}^2 = \mathbf{P}$) gleich ihrer Spur $\mathrm{tr}(\mathbf{P})$, da die Eigenwerte von $\mathbf{P}$ entweder 0 oder 1 sind.

Beweis: Mardia et al. (1979, Anhang A, S. 464)

Kapitel 10

Multivariate Verteilungen

10.1 Univariate Testverteilungen

Bei univariaten Tests werden häufig aus der Normalverteilung $N(\mu, \sigma^2)$ abgeleitete Teststatistiken benötigt, etwa die χ^2-, F- oder t-Verteilung. Auch im multivariaten Fall werden Teststatistiken benutzt, die aus der Normalverteilung $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ abgeleitet sind. Allgemein gilt:[1]

- Ist $X_i \sim N(0,1), i = 1, ..., m$ und paarweise unabhängig.
  Dann ist $Y = \sum_{i=1}^{m} X_i^2 \sim \chi^2(m)$
  chi-quadrat-verteilt mit m Freiheitsgraden.
- Ist $X \sim N(0,1)$ und unabhängig von $Y \sim \chi^2(m)$.
  Dann ist $T = \frac{X}{\sqrt{Y/m}} \sim t(m)$
  t-verteilt mit m Freiheitsgraden
  und $T^2 = \frac{X^2}{(Y/m)} \sim F(1, m)$.
- Ist $X \sim \chi^2(m)$ und unabhängig von $Y \sim \chi^2(n)$.
  Dann ist $F = \frac{X/m}{Y/n} \sim F(m, n)$
  F-verteilt mit m Zähler- und n Nenner-Freiheitsgraden.
- Ist $X \sim \chi^2(m)$ und unabhängig von $Y \sim \chi^2(n)$.
  Dann ist $B = \frac{X}{X+Y} \sim \beta(m/2, n/2)$
  β-verteilt mit Parametern $m/2$ und $n/2$.
  Es gilt $B = mF/(n + mF)$ oder $F = nB/(m(1 - B))$.

[1] vgl. http://www.fernuni-hagen.de/ls_statistik/lehre/.

10.2 Normalverteilung

Die Dichtefunktion einer normalverteilten vektoriellen Zufallsvariable $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ist

$$\phi(\mathbf{x}) = |2\pi\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}. \tag{10.1}$$

Hierbei ist $\mathbf{x} = [x_1, ..., x_p]'$ ein p-Vektor und

$$\boldsymbol{\mu} = E[\mathbf{x}] = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \tag{10.2}$$

$$\tag{10.3}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \dots & \mathrm{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \dots & \mathrm{Cov}(X_p, X_p) \end{bmatrix} \tag{10.4}$$

sind die Parameter (Vektoren und Matrizen) der p-variaten Normalverteilung. Es gilt

$$\boldsymbol{\mu} = E[\mathbf{x}] = \int \mathbf{x}\, \phi(\mathbf{x}) d^p\mathbf{x} \tag{10.5}$$

$$\boldsymbol{\Sigma} = \mathrm{Var}[\mathbf{x}] = \int (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})'\phi(\mathbf{x}) d^p\mathbf{x}, \tag{10.6}$$

$d^p\mathbf{x} = dx_1 ... dx_p$. Es handelt sich um p-dimensionale Integrale, d.h. $\int ... \int$. Einige *Eigenschaften:*

1. Lineare Transformationen $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ sind ebenfalls normalverteilt $N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

   Jeder Vektor aus Komponenten von $\mathbf{x}$ ist multivariat normalverteilt.

2. Die Projektion $\mathbf{a}'\mathbf{x}$ ist univariat normalverteilt $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$.

3. $\mathbf{x}$ ist p-variat normalverteilt, wenn $\mathbf{a}'\mathbf{x}$ univariat normalverteilt ist für alle Vektoren $\mathbf{a}$.

4. $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})'\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'(\mathbf{x}-\boldsymbol{\mu})$, wenn man die Spektralzerlegung $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'$ einsetzt.

   $\mathbf{y} = \mathbf{P}'(\mathbf{x}-\boldsymbol{\mu})$ sind die sog. Hauptkomponenten.

   Es gilt $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \mathbf{y}'\boldsymbol{\Lambda}^{-1}\mathbf{y} = \sum_i y_i^2/\lambda_i^2$.

5. Die Hauptkomponenten sind unkorreliert und es gilt

   $\mathrm{Var}(\mathbf{y}) = \mathbf{P}'\mathrm{Var}(\mathbf{x})\mathbf{P} = \boldsymbol{\Lambda}$.

   Daher ist $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_i y_i^2/\lambda_i^2 \sim \chi^2(p)$.

Betrachtet man einen partitionierten normalverteilten Vektor $\mathbf{z}' = [\mathbf{x}', \mathbf{y}']$ mit Teilvektoren der Dimension p und q, so ist

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \right) \tag{10.7}$$

durch die gemeinsame Dichte $\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ charakterisiert ($\mathbf{x} = [x_1, ..., x_p]'$, $\mathbf{y} = [y_1, ..., y_q]'$). Hierbei ist $\boldsymbol{\Sigma}$ eine Blockmatrix der Dimension $(p+q) \times (p+q)$. Es ergibt sich das Resultat, daß auch die Randverteilungen $\phi(\mathbf{x}; \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $\phi(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$ multivariate Normalverteilungsdichten sind. Man muß also aus der vollen Kovarianzmatrix nur die entsprechenden Zeilen und Spalten herausschneiden, um die Randdichte zu erhalten.

Randverteilungen

Für die bedingte Verteilung $f(\mathbf{y}|\mathbf{x})$ gilt

$$\mathbf{y}|\mathbf{x} \sim N(\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x}), \tag{10.8}$$

wobei die bedingten Erwartungswerte und Kovarianzen durch

$$\boldsymbol{\mu}_{y|x} = E[\mathbf{y}|\mathbf{x}] = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x) \tag{10.9}$$

$$\boldsymbol{\Sigma}_{y|x} = \mathrm{Var}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \tag{10.10}$$

Satz von der Normalkorrelation

gegeben sind.

Man erhält dieses Resultat aus der Zerlegung $\mathbf{y} = \mathbf{a}+\mathbf{B}\mathbf{x}+\boldsymbol{\epsilon}$, wobei $\mathbf{x}$ und $\boldsymbol{\epsilon}$ orthogonal sind, d.h. $\mathrm{Cov}(\mathbf{x}, \boldsymbol{\epsilon}) = \mathbf{0}$. Hierbei ist auch $\boldsymbol{\epsilon}$ normalverteilt

mit $E[\boldsymbol{\epsilon}] = \mathbf{0}$. Aus der Orthogonalitätsbedingung findet man $\mathrm{Cov}(\mathbf{y}, \mathbf{x}) = \mathbf{B}\mathrm{Cov}(\mathbf{x}, \mathbf{x})$, d.h. $\mathbf{B} = \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$. Außerdem muß $E[\mathbf{y}] = \mathbf{a} + \mathbf{B}E[\mathbf{x}]$ oder $\mathbf{a} = \boldsymbol{\mu}_y - \mathbf{B}\boldsymbol{\mu}_x$ gelten.

Die Varianzzerlegung $\mathrm{Var}(\mathbf{y}) = \mathbf{B}\mathrm{Var}(\mathbf{x})\mathbf{B}' + \mathrm{Var}(\boldsymbol{\epsilon})$ folgt ebenfalls aus der Orthogonalität.

Dies ergibt die bedingten Erwartungswerte

$$
\begin{aligned}
E[\mathbf{y}|\mathbf{x}] &= \mathbf{a} + \mathbf{B}\mathbf{x} + E[\boldsymbol{\epsilon}|\mathbf{x}] && (10.11)\\
&= \mathbf{a} + \mathbf{B}\mathbf{x} + E[\boldsymbol{\epsilon}] && (10.12)\\
&= \mathbf{a} + \mathbf{B}\mathbf{x} && (10.13)\\
&= \boldsymbol{\mu}_y + \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}_x) && (10.14)\\
&= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) && (10.15)
\end{aligned}
$$

da $\mathbf{x}$ und $\boldsymbol{\epsilon}$ orthogonal (und wegen der Normalverteilung sogar unabhängig) sind. Für die bedingte Varianz gilt

$$
\begin{aligned}
\mathrm{Var}(\mathbf{y}|\mathbf{x}) &= \mathrm{Var}(\mathbf{a} + \mathbf{B}\mathbf{x} + \boldsymbol{\epsilon}|\mathbf{x}) && (10.16)\\
&= \mathrm{Var}(\boldsymbol{\epsilon}|\mathbf{x}) && (10.17)\\
&= \mathrm{Var}(\boldsymbol{\epsilon}) && (10.18)\\
&= \boldsymbol{\Sigma}_{yy} - \mathbf{B}\boldsymbol{\Sigma}_{xx}\mathbf{B}' && (10.19)\\
&= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}. && (10.20)
\end{aligned}
$$

Betrachtet man eine Zufallsstichprobe $\mathbf{x}_1, ..., \mathbf{x}_N$, $\mathbf{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so wird $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]'$ als Daten-Matrix aus $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ bezeichnet. Es gilt für den Mittelwert $\bar{\mathbf{x}} = N^{-1}\mathbf{X}'\mathbf{1}$

$$
\bar{\mathbf{x}} \sim N(\boldsymbol{\mu}, N^{-1}\boldsymbol{\Sigma}) \qquad (10.21)
$$

10.3 Die Wishart-Verteilung

Die χ^2-Verteilung entsteht aus quadrierten normalverteilten Zufallsvariablen. Im multivariaten Fall ergeben sich häufig quadratische Formen der Gestalt $\mathbf{X}'\mathbf{A}\mathbf{X}$, etwa $\mathbf{X}'\mathbf{H}\mathbf{X} = (N-1)\mathbf{S}$ (Stichprobenkovarianz-Matrix).

Wishart-Verteilung

Setzt man $\mathbf{M} = \mathbf{X}'\mathbf{X}$, $\mathbf{X} : N \times p$ Datenmatrix aus $N(\mathbf{0}, \mathbf{\Sigma})$, so hat $\mathbf{M}$ eine Wishart-Verteilung $\mathbf{M} \sim W_p(\mathbf{\Sigma}, N)$ mit Skalen-Matrix $\mathbf{\Sigma}$ und N Freiheitsgraden.

Eigenschaften:

1. $p = 1$: $W_1(\sigma^2, N)$ ist die Verteilung von $\mathbf{x}'\mathbf{x} = \sum_n x_i^2, x_i \sim N(0, \sigma^2)$. Daher gilt $W_1(\sigma^2 = 1, N) = \chi^2(N)$ oder $W_1(\sigma^2, N) = \sigma^2\chi^2(N)$.
2. $E[\mathbf{M}] = \sum_n E[\mathbf{x}_n\mathbf{x}_n'] = N\mathbf{\Sigma}$
3. Wenn $\mathbf{M} \sim W_p(\mathbf{\Sigma}, N)$, dann $\mathbf{B}'\mathbf{M}\mathbf{B} \sim W_q(\mathbf{B}'\mathbf{\Sigma}\mathbf{B}, N)$ wobei $\mathbf{B} : p \times q$
4. Wenn $\mathbf{M} \sim W_p(\mathbf{\Sigma}, N)$, dann
$$\mathbf{a}'\mathbf{M}\mathbf{a}/\mathbf{a}'\mathbf{\Sigma}\mathbf{a} \quad \sim \quad \chi^2(N) \tag{10.22}$$
5. Wenn $\mathbf{M}_1 \sim W_p(\mathbf{\Sigma}, N_1)$, $\mathbf{M}_2 \sim W_p(\mathbf{\Sigma}, N_2)$, dann
$$\mathbf{M}_1 + \mathbf{M}_2 \sim W_p(\mathbf{\Sigma}, N_1 + N_2) \tag{10.23}$$
6. $\mathbf{X}'\mathbf{A}\mathbf{X} \sim W_p(\mathbf{\Sigma}, r)$, $r = \mathrm{tr}[\mathbf{A}]$, genau dann, wenn $\mathbf{A}^2 = \mathbf{A}$ und symmetrisch.
7. $(N-1)\mathbf{S} = \mathbf{X}'\mathbf{H}\mathbf{X} \sim W_p(\mathbf{\Sigma}, N-1)$, $N - 1 = \mathrm{tr}[\mathbf{H}]$, da $\mathbf{H}^2 = \mathbf{H}$.

Herleitung:

3. $\mathbf{B}'\mathbf{M}\mathbf{B} = \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} := \mathbf{Y}'\mathbf{Y}$. Daher ist $\mathbf{Y}' = [\mathbf{B}'\mathbf{x}_1, ..., \mathbf{B}'\mathbf{x}_N]$ und $\mathrm{Var}(\mathbf{B}'\mathbf{x}_n) = \mathbf{B}'\mathbf{\Sigma}\mathbf{B}$.

4. Setze $\mathbf{B} = \mathbf{a}$. Daher ist $\mathbf{a}'\mathbf{M}\mathbf{a} \sim W_1(\mathbf{a}'\mathbf{\Sigma}\mathbf{a}, N) = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}\ \chi^2(N)$.

6. $\mathbf{X}'\mathbf{A}\mathbf{X} = \sum_{\lambda_i=1} \mathbf{X}'\boldsymbol{\psi}_i\boldsymbol{\psi}_i'\mathbf{X}$. Wegen $\mathbf{A}^2 = \mathbf{A}$ sind die Eigenwerte 0 oder 1. Setzt man $\mathbf{y}_i = \mathbf{X}'\boldsymbol{\psi}_i = \sum_n \mathbf{x}_n\psi_{in} : p \times 1$, so gilt $\mathrm{Cov}(\mathbf{y}_i) = \sum_n \mathbf{\Sigma}\psi_{in}^2 = \mathbf{\Sigma}$. Daher ist $\sum_{\lambda_i=1} \mathbf{y}_i\mathbf{y}_i' \sim W(\mathbf{\Sigma}, r = \sum \lambda_i)$.

Der Mittelwert $\bar{\mathbf{x}}$ und die Stichproben-Kovarianzmatrix $\mathbf{S}$ erfüllen:

$$\bar{\mathbf{x}} \sim N(\boldsymbol{\mu}, N^{-1}\boldsymbol{\Sigma}) \tag{10.24}$$
$$(N-1)\mathbf{S} \sim W_p(\boldsymbol{\Sigma}, N-1) \tag{10.25}$$
$$\bar{\mathbf{x}} \text{ und } \mathbf{S} \text{ sind unabhängig.} \tag{10.26}$$

10.4 Die Hotelling-T^2-Verteilung

Die univariate t-Verteilung ergibt sich, wenn man in der Statistik $Z = (\bar{X} - \mu)/(\sigma/\sqrt{N})$ die unbekannte Standardabweichung durch S ersetzt. Dann ist $T = (\bar{X} - \mu)/(S/\sqrt{N}) \sim t(N-1)$ und $T^2 = (\bar{X} - \mu)/(S/N) \sim F(1, N-1)$ verteilt.

Im multivariaten Fall betrachtet man die quadratische Form $Q = m\,\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}$ mit $\mathbf{d} \sim N(\mathbf{0}, \mathbf{I})$ und $\mathbf{M} \sim W_p(\mathbf{I}, m)$.

Hotelling-T^2-Verteilung

Dann ist

$$Q = m\,\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} \sim T^2(p, m) \tag{10.27}$$

Hotelling-T^2-verteilt mit Parametern p und m.

Daraus folgt für $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ und $\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$:

$$m(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim T^2(p, m) \tag{10.28}$$

und

$$N(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim T^2(p, N-1). \tag{10.29}$$

Für $p = 1$ ergibt sich $d \sim N(0, 1)$ und $M \sim \chi^2(m)$, d.h. $Q = d^2/M \sim F(1, m)$. Die allgemeine Umrechnung lautet

$$T^2(p,m) = \frac{mp}{m-p+1} F(p, m-p+1). \tag{10.30}$$

Wählt man $m = N - 1$, so ergibt sich

$$\frac{(N-p)N}{p(N-1)} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim F(p, N-p). \tag{10.31}$$

10.5 Die Wilks'-Λ- und θ-Verteilung

Stichprobenvarianzen enthalten quadrierte normalverteilte Variablen und sind daher χ^2-verteilt. Bei Tests auf Gleichheit von Mittelwerten ergeben sich Quotienten von solchen Quadratsummen, die auf die F-Verteilung führen.

Im multivariaten Fall ergeben sich Wishart-Verteilungen für die Quadratsummen. Nimmt man an, daß $\mathbf{A} \sim W_p(\boldsymbol{\Sigma}, m)$ und $\mathbf{B} \sim W_p(\boldsymbol{\Sigma}, n)$ (voneinander unabhängig, $m \geq p$), so interessiert man sich für die Matrix $\mathbf{A}^{-1}\mathbf{B}$.

Man kann $\mathbf{A}$ als Residualstreuung und $\mathbf{B}$ als erklärte Streuung interpretieren.

Die Eigenwerte λ_i von $\mathbf{A}^{-1}\mathbf{B}$ sind größer oder gleich 0. Die gemeinsame Verteilung der Eigenwerte λ_i wird als $\Psi(p, m, n)$-Verteilung bezeichnet.

Die Statistik $\phi = |\mathbf{A}^{-1}\mathbf{B}| = |\mathbf{B}|/|\mathbf{A}|$ ist das Produkt von $F_i \sim F(n-i+1, m-i+1)$-verteilten Variablen.

Bei Likelihood-Quotienten-Tests ergeben sich Determinanten der Form $\Lambda = |\mathbf{A}|/|\mathbf{A}+\mathbf{B}| = |\mathbf{I}+\mathbf{A}^{-1}\mathbf{B}|^{-1}$. Dies kann mit Hilfe der Eigenwerte als $\prod_i (1+\lambda_i)^{-1}$ geschrieben werden. Der Quotient Λ ist analog zur Beta-Verteilung.

Man erhält so die Definition:

Wilks'-Lambda

Wenn $\mathbf{A} \sim W_p(\mathbf{I}, m)$ und $\mathbf{B} \sim W_p(\mathbf{I}, n)$, unabhängig und $m \geq p$, so heißt

$$\Lambda = |\mathbf{A}|/|\mathbf{A}+\mathbf{B}| = |\mathbf{I}+\mathbf{A}^{-1}\mathbf{B}|^{-1} \sim \Lambda(p, m, n) \tag{10.32}$$

Wilks'-Lambda-verteilt mit Parametern p, m, n.

Hierbei korrespondiert m zu den Freiheitsgraden der Residualstreuung und n zur erklärten Streuung. $n+m$ sind die Freiheitsgrade der totalen Streuung.

Eigenschaften:

1. Man kann auch $\mathbf{A} \sim W_p(\mathbf{\Sigma}, m)$ und $\mathbf{B} \sim W_p(\mathbf{\Sigma}, n)$ setzen, da sich die Varianz kürzt.
2. Λ ist gleich dem Produkt von unabhängigen Beta-verteilten Variablen $u_i \sim \beta((m+i-p)/2, p/2), i = 1, ..., n$.
3. $\Lambda = \prod_{i=1}^{p}(1+\lambda_i)^{-1}$. Vgl. (9.166).
4. $\Lambda(p, m, n) = \Lambda(n, m+n-p, p)$.
5. Für $m \to \infty$ gilt: $-[m - \frac{1}{2}(p-n+1)] \log \Lambda(p, m, n) \overset{a}{\sim} \chi^2(np)$ (Approximation von Bartlett)

Aufgrund der Darstellung $\Lambda = \prod_{i=1}^{p}(1+\lambda_i)^{-1}$ kann Wilks-Lambda als Funktion der Eigenwerte von $\mathbf{A}^{-1}\mathbf{B}$ geschrieben werden. Auch andere Statistiken können so dargestellt werden.

Beim Union-Intersection-Prinzip erhält man oft das Kriterium des größten Eigenwerts.

größter Eigenwert

Wenn $\mathbf{A} \sim W_p(\mathbf{I}, m)$ und $\mathbf{B} \sim W_p(\mathbf{I}, n)$, unabhängig und $m \geq p$, so heißt der größte Eigenwert θ von $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$ θ-verteilt, d.h. $\theta \sim \theta(p, m, n)$.

Eigenschaften:

1. Man kann auch $\mathbf{A} \sim W_p(\mathbf{\Sigma}, m)$ und $\mathbf{B} \sim W_p(\mathbf{\Sigma}, n)$ setzen, da sich die Varianz kürzt.

2. $\theta = \lambda_1/(1+\lambda_1)$. Hierbei ist $\lambda_1 > \lambda_2 > ... > \lambda_p$. Vgl. (9.165).

3. $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = (\mathbf{I}+\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}\mathbf{B}$. Daher hat die Matrix die gleichen Eigenvektoren wie $\mathbf{A}^{-1}\mathbf{B}$ und es gilt $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}\boldsymbol{\psi} = \lambda/(1+\lambda)\boldsymbol{\psi}$.

4. $|(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}| = |\mathbf{B}|/|\mathbf{A}+\mathbf{B}| = \phi\Lambda$ kann als erklärte Streuung/totale Streuung interpretiert werden.

Kapitel 11

Notation und Rechenregeln

In diesem Kapitel werden die im Kurs benutzte Notation sowie wichtige Rechenregeln für Ereignisse und Wahrscheinlichkeiten alphabetisch zusammengestellt. Wenn Ihnen diese nicht bekannt sind, konsultieren Sie bitte Ihren Bachelor-Statistikkurs bzw. entsprechende Lehrbücher (z.B. Fahrmeir et al., 2007).

11.1 Formeln (alphabetisch)

Bayes-Formel

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \text{ (bedingte Wahrscheinlichkeit)} \tag{11.1}$$

bedingte Wahrscheinlichkeit

$$= P(A) \text{ bei Unabhängigkeit} \tag{11.2}$$

Kronecker-Delta-Symbol

$$\delta_{jj'} = \begin{cases} 1 & j = j' \\ 0 & \text{sonst} \end{cases} \tag{11.3}$$

Dichtefunktion

$$f(x) = P(x \leq X \leq x + dx)/dx \quad \text{(stetige Zufallsvariablen)} \tag{11.4}$$

$$= P(X = x) \quad \text{(diskrete Zufallsvariablen)} \tag{11.5}$$

Dichtefunktion (bivariat)

$$f(x,y) = P(x \leq X \leq x+dx, y \leq Y \leq y+dy)/dxdy \quad \text{(stetige Zufallsvariablen)} \tag{11.6}$$

$$= P(X=x, Y=y) \quad \text{(diskrete Zufallsvariablen)} \tag{11.7}$$

$$f(x,y) = f_x(x) \cdot f_y(y) \text{ bei Unabhängigkeit} \tag{11.8}$$

$$f_x(x), f_y(y) : \text{Randverteilungen} \tag{11.9}$$

$$f_x(x) = \int f(x,y)dy \quad \text{(stetige Zufallsvariablen)} \tag{11.10}$$

$$= \sum_y f(x,y) \quad \text{(diskrete Zufallsvariablen)} \tag{11.11}$$

Ereignisse

$$A = \{\omega | \omega \in \Omega\} \tag{11.12}$$

$$\Omega = \text{Ergebnismenge} \tag{11.13}$$

$$P(\Omega) = 1 \text{ (sicheres Ereignis)} \tag{11.14}$$

$$\bar{A} : A \text{ tritt nicht ein} \tag{11.15}$$

$$P(\bar{A}) = 1 - P(A) \tag{11.16}$$

$$\emptyset = \bar{\Omega} \text{ (unmögliches Ereignis)} \tag{11.17}$$

$$P(\emptyset) = 1 - P(\Omega) = 0 \tag{11.18}$$

$$A \cap B : A \text{ und } B \text{ tritt ein} \tag{11.19}$$

$$P(A \cap B) = P(A) \cdot P(B) \text{ bei Unabhängigkeit} \tag{11.20}$$

$$A \cup B \quad A \text{ oder } B \text{ tritt ein} \tag{11.21}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{11.22}$$

$$= P(A) + P(B), \text{ falls } A \cap B = \emptyset \tag{11.23}$$

$$A \subset B : \text{Wenn } A \text{, dann auch } B \tag{11.24}$$

$$A \setminus B = A \cap \bar{B} \text{ (} A \text{ ohne } B\text{)} \tag{11.25}$$

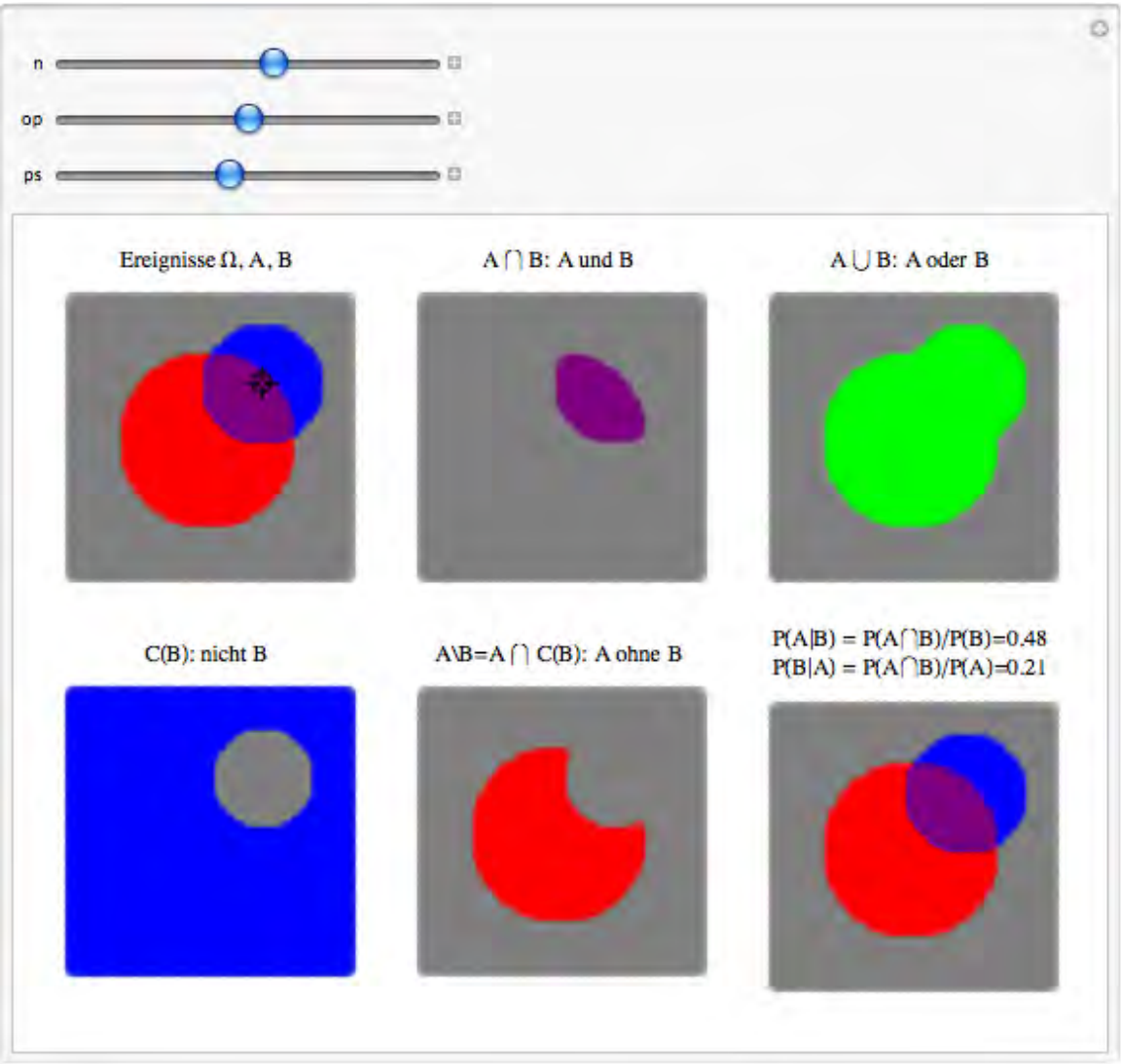


Abbildung 11.1: Ereignisse.
http://www.fernuni-hagen.de/ls_statistik/lehre/

Erwartungswert

$$E[X] = \mu \tag{11.26}$$

$$= \int x f(x) dx$$

(stetige Zufallsvariablen) (11.27)

$$= \sum_i x_i f(x_i)$$

(diskrete Zufallsvariablen) (11.28)

$$E[h(X)] = \int h(x) f(x) dx$$

(stetige Zufallsvariablen) (11.29)

$$= \sum_i h(x_i) f(x_i)$$

(diskrete Zufallsvariablen) (11.30)

Kovarianz

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{11.31}$$

$$E(X) = \mu_X \tag{11.32}$$

$$E(Y) = \mu_Y \tag{11.33}$$

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \tag{11.34}$$

$$= \int (x-\mu_X)(y-\mu_Y) f(x,y) dx dy$$

(stetige Zufallsvariablen) (11.35)

$$= \sum (x_i-\mu_X)(y_j-\mu_Y) f(x_i,y_j)$$

(diskrete Zufallsvariablen) (11.36)

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X) = E[X^2] - \mu^2 \tag{11.37}$$

$$\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y) \tag{11.38}$$

$$= \mathrm{Cov}(X,X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,Y) \tag{11.39}$$

$$= \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y) \tag{11.40}$$

$$\begin{aligned}
\mathrm{Cov}(aX, bY) &= a \cdot b\,\mathrm{Cov}(X, Y) && (11.41)\\
\mathrm{Cov}(X+Y, Z+W) &= \mathrm{Cov}(X, Z+W) + \mathrm{Cov}(Y, Z+W) \\
&= \mathrm{Cov}(X, Z) + \mathrm{Cov}(X, W) \\
&+ \mathrm{Cov}(Y, Z) + \mathrm{Cov}(Y, W) && (11.42)
\end{aligned}$$

Kovarianzmatrix

$$\begin{aligned}
\sigma_{ij} &= \mathrm{Cov}(X_i, X_j);\ i, j = 1, \ldots, p && (11.43)\\
\mathbf{\Sigma} &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} (p = 3) && (11.44)
\end{aligned}$$

Mittelwert

$$\begin{aligned}
\bar{X} &= \frac{1}{N} \sum_{n=1}^{N} X_n && (11.45)\\
\bar{\mathbf{x}} &= \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n && (11.46)
\end{aligned}$$

Normalverteilung (univariat)

$$\begin{aligned}
X &\sim N(\mu, \sigma^2) && (11.47)\\
\mu &= E(X) && (11.48)\\
\sigma^2 &= \mathrm{Var}(X) && (11.49)
\end{aligned}$$

Dichtefunktion:

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (11.50)$$

Normalverteilung (bivariat)

$$\mathbf{x} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{11.51}$$

$$\boldsymbol{\mu} = E[\mathbf{x}] = \begin{bmatrix} E(X) \\ E(Y) \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \tag{11.52}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix} \tag{11.53}$$

$$= \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \tag{11.54}$$

$$= \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho_{xy} \\ \sigma_x\sigma_y\rho_{xy} & \sigma_y^2 \end{bmatrix} \tag{11.55}$$

$$\tag{11.56}$$

Dichtefunktion:

$$\begin{aligned} f(x,y) &= \det(2\pi\boldsymbol{\Sigma})^{-1/2} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \\ &\times \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy}\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \end{aligned}$$

Spur

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{p} A_{ii} \text{ (Spur = engl. trace)} \tag{11.57}$$

Stichprobenvarianz

$$s^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^2 \tag{11.58}$$

$$\mathbf{S} = \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})' \tag{11.59}$$

Varianz

$$\begin{aligned}
\mathrm{Var}(X) &= \sigma^2 && (11.60)\\
&= \int (x-\mu)^2 f(x)dx \\
&\quad \text{(stetige Zufallsvariablen)} && (11.61)\\
&= \sum_i (x_i-\mu) f(x_i) \\
&\quad \text{(diskrete Zufallsvariablen)} && (11.62)
\end{aligned}$$

Unabhängigkeit

$$\begin{aligned}
P(A\cap B) &= P(A)\cdot P(B) && (11.63)\\
P(A|B) &:= \frac{P(A\cap B)}{P(B)} = P(A) && (11.64)
\end{aligned}$$

Verteilungsfunktion

$$\begin{aligned}
F(x) &= P(X\le x) && (11.65)\\
&= \int_{-\infty}^{x} f(y)dy \\
&\quad \text{(stetige Zufallsvariablen)} && (11.66)\\
&= \sum_{y_i\le x} f(y_i) \\
&\quad \text{(diskrete Zufallsvariablen)} && (11.67)
\end{aligned}$$

Wahrscheinlichkeit

$$\begin{aligned}
P(A) &= \text{Wahrscheinlichkeit des Ereignisses } A && (11.68)\\
&\quad [0\le P(A)\le 1] && (11.69)\\
P(A\cup B) &= P(A)+P(B)-P(A\cap B) && (11.70)\\
P(A\cap B) &= P(A)\cdot P(B) \text{ bei Unabhängigkeit} && (11.71)\\
P(A|B) &:= \frac{P(A\cap B)}{P(B)} \text{ (bedingte Wahrscheinlichkeit)} && (11.72)\\
&= P(A) \text{ bei Unabhängigkeit} && (11.73)
\end{aligned}$$

Zufallsvariable

$$X = X(\omega); \omega \in \Omega \tag{11.74}$$

$$X : \Omega \to \mathbb{R} \tag{11.75}$$

Zufallsvektor

$$\mathbf{x} = \mathbf{x}(\omega) = [X_1(\omega), \ldots, X_p(\omega)]'; \omega \in \Omega \tag{11.76}$$

$$\mathbf{x} : \Omega \to \mathbb{R}^p \tag{11.77}$$

11.2 Beispiel: symmetrischer Würfel

11.2.1 Einmaliges Würfeln

$$
\begin{aligned}
\Omega &= \{1,2,3,4,5,6\} = \{\omega_1,\ldots,\omega_6\} \\
&\quad \text{(Ergebnismenge)} \\
P(\Omega) &= 1 \text{ (sicheres Ereignis)} \\
P(\omega_i) &= 1/6 \\
\omega_i &= i \text{ (Elementarereignisse)} \\
X(\omega_i) &= i \text{ (Zufallsvariable)} \\
f(x_i) &= P(X = x_i) = 1/6; x_i = 1,2,3,4,5,6 \\
&\quad \text{(Dichtefunktion)} \\
F(x_i) &= \sum_{x_j \leq x_i} f(x_j) \\
&\quad \text{(Verteilungsfunktion)}
\end{aligned}
$$

$$
\begin{aligned}
E[X] &= \sum x_i f(x_i) = (1+2+3+4+5+6)/6 = 3.5 \\
A &= \{1,3,5\} \text{ (ungerade Zahlen)} \\
P(A) &= 1/2 \\
\bar{A} &= \{2,4,6\} \text{ (A tritt nicht ein, gerade Zahlen)} \\
P(\bar{A}) &= 1 - P(A) = 1 - 1/2 = 1/2 \\
\emptyset &= \bar{\Omega} \text{ (unmögliches Ereignis)} \\
P(\emptyset) &= 1 - P(\Omega) = 0 \\
B &= \{2,4,6\} \text{ (gerade Zahlen)} \\
A \cap B &= \emptyset \text{ (gerade und ungerade Zahl unmöglich)} \\
P(A \cap B) &= P(\emptyset) = 0 \\
A \cup B &= A \text{ oder } B \text{ tritt ein} \\
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= 1/2 + 1/2 - 0 = 1
\end{aligned}
$$

11.2.2 Zweimaliges Würfeln

$$
\begin{aligned}
\Omega &= \{(1,1),(1,2),(1,3),\ldots,(6,6)\} = \{\omega_{11},\ldots,\omega_{66}\} \\
&\quad \text{(Ergebnismenge)} \\
P(\Omega) &= 1 \text{ (sicheres Ereignis)} \\
P(\omega_{ij}) &= 1/36 = 1/6 \cdot 1/6 \text{ (Würfe unabhängig)} \\
\omega_{ij} &= (i,j) \text{ (Elementarereignisse, Tupel)} \\
X(\omega_{ij}) &= i+j = X_1(\omega_{ij}) + X_2(\omega_{ij}) \\
&= \text{ (Zufallsvariable: Würfelsumme)} \\
f(x) &= P(X=x), x = 2,\ldots,12 \text{ (Dichtefunktion)} \\
f(2) &= P(X=2) = 1/36 \\
f(3) &= P(X=3) = 2/36 \\
f(4) &= P(X=4) = 3/36 \\
f(5) &= P(X=5) = 4/36 \\
f(6) &= P(X=6) = 5/36 \\
f(7) &= P(X=7) = 6/36 \\
f(8) &= P(X=8) = 5/36 \\
f(9) &= P(X=9) = 4/36 \\
f(10) &= P(X=10) = 3/36 \\
f(11) &= P(X=11) = 2/36 \\
f(12) &= P(X=12) = 1/36 \\
F(x_i) &= \sum_{x_j \le x_i} f(x_j) \\
&\quad \text{(Verteilungsfunktion)}
\end{aligned}
$$

$$
\begin{aligned}
A &= \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \\
&\quad \text{(6 im ersten Wurf, 2. Wurf beliebig = 1. Zeile in Omega)} \\
P(A) &= 6/36 = 1/6 \\
B &= \text{ (6 im zweiten Wurf, 1. Wurf beliebig = 1. Spalte in Omega)} \\
A \cap B &= \{(6,6)\} \\
P(A \cap B) &= P(\{(6,6)\}) = 1/36 = P(A) \cdot P(B) \\
&\quad \text{(Ereignisse sind unabhängig)}
\end{aligned}
$$

In übersichtlicher Tabellenform kann man die Elementarereignisse als Tupel $\omega_{ij} = (i, j)$

ω_{ij}	1	2	3	4	5	6
1	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
2	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
3	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
4	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
6	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6\}$

und die Summenwerte als

$X(\omega_{ij})$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

anschreiben. Daraus ergeben sich unmittelbar die Wahrscheinlichkeiten $f(x) = P(X = x)$, z.B. $P(X = 7) = 6/36$.

In Abb. 11.2 sind die Ergebnismenge Ω sowie die Ereignisse $A = \{(3,2), \ldots, (5,4)\}$ (rot), $B = \{(3,4), \ldots, (4,5)\}$ (blau) sowie die Mengen $A \cap B = \{(3,4), (4,4)\}$ (lila), $A \cup B$ (grün), $\bar{B}$ (blau), $A \setminus B$ (rot) sowie die bedingten Wahrscheinlichkeiten $P(A|B) = P(A \cap B)/P(B) = (2/36)/(4/36) = 2/4 = .5$ und $P(B|A) = P(A \cap B)/P(A) = (2/36)/(9/36) = 2/9 = .22$ dargestellt.

Die Ereignisse A und B sind voneinander abhängig, da $P(A \cap B) = 2/36 \neq P(A) \cdot P(B) = 9/36 \cdot 4/36 = 1/36$.

Entsprechend gilt für die bedingten Wahrscheinlichkeiten: $P(A|B) = 2/4 \neq P(A) = 1/4$ und $P(B|A) = 2/9 \neq P(B) = 4/36$.

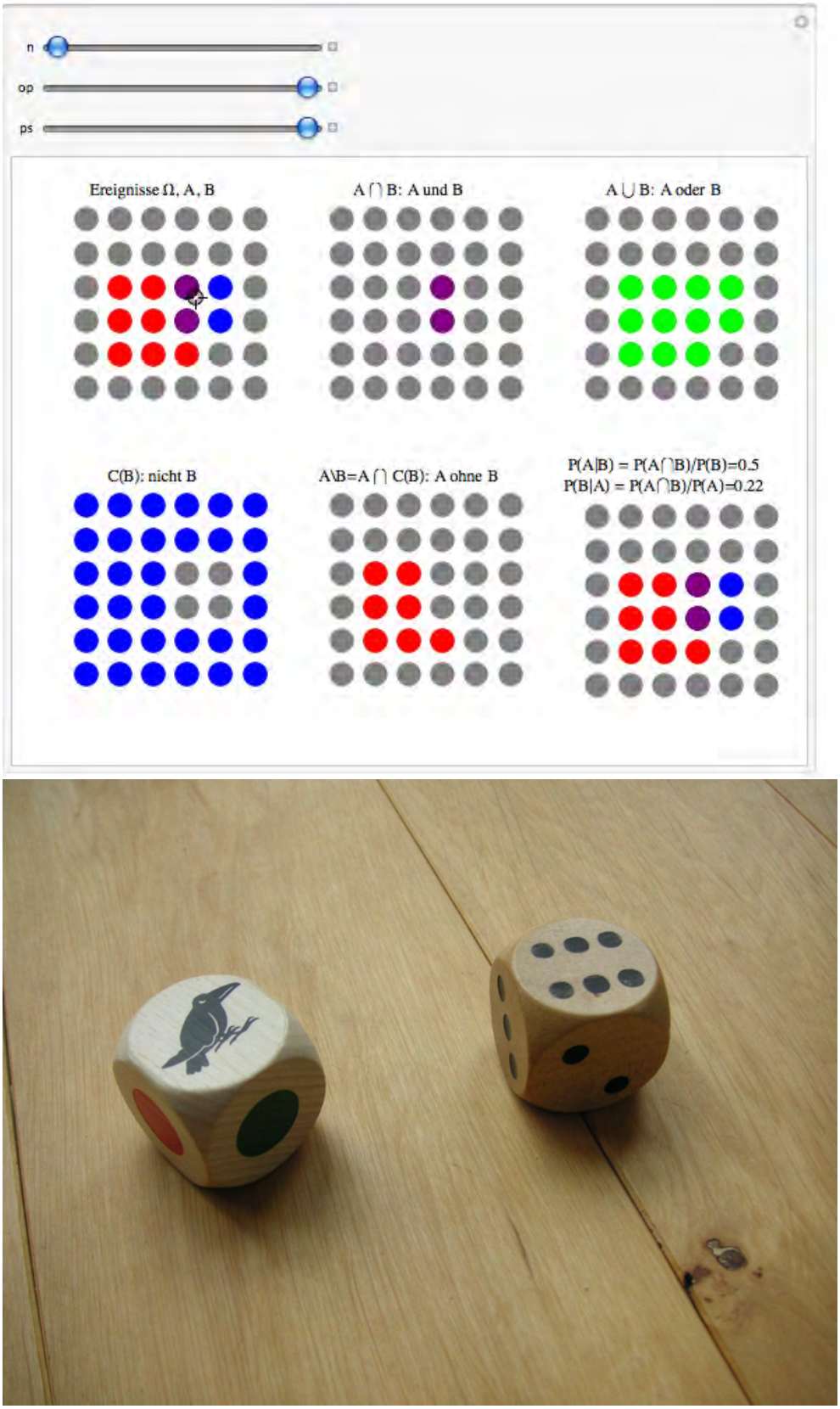


Abbildung 11.2: Ereignisse beim zweimaligen Würfeln.

Kapitel 12

Tabellen

12.1 Normalverteilung

Dichtefunktion der Normalverteilung

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

Verteilungsfunktion

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{\xi-\mu^2}{2\sigma^2}\right) d\xi$$

Erwartungswert

$$E(X) = \mu$$

Varianz

$$\mathrm{Var}(X) = \sigma^2$$

Parameter der Normalverteilung

$$\mu, \sigma^2$$

Man schreibt häufig $X \sim N(\mu, \sigma^2)$ und spricht von einer $N(\mu, \sigma^2)$-verteilten Zufallsvariablen.

(1) Ist die Zufallsvariable $X \sim N(\mu, \sigma^2)$, so ist die Zufallsvariable $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$. Es gilt also: Eine lineare Funktion einer normalverteilten Zufallsvariablen ist wieder eine normalverteilte Zufallsvariable. Aus dieser Eigenschaft folgt für $a = \frac{1}{\sigma}$ und $b = -\frac{\mu}{\sigma}$ speziell:

Ist die Zufallsvariable $X \sim N(\mu, \sigma^2)$, so ist

$$Z = \frac{X}{\sigma} - \frac{\mu}{\sigma} = \frac{X-\mu}{\sigma}$$

$N(0,1)$-verteilt.

Standard-normalverteilung

Eine Normalverteilung mit den Parametern $\mu = 0$ und $\sigma^2 = 1$ heißt Standardnormalverteilung.

Ist die Zufallsvariable $X \sim N(\mu, \sigma^2)$ und ist $Z \sim N(0,1)$, so folgt

$$P(x_1 \leq X \leq x_2) = P\left(\frac{x_1-\mu}{\sigma} \leq Z \leq \frac{x_2-\mu}{\sigma}\right).$$

(2) Sind X_1 und X_2 unabhängige Zufallsvariablen und $N(\mu_1, \sigma_1^2)$- bzw. $N(\mu_2, \sigma_2^2)$-verteilt, so ist die Summe der Zufallsvariablen $X = a_1X_1 + a_2X_2 + b \sim N(a_1\mu_1 + a_2\mu_2 + b, a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$ wieder normalverteilt.

Um die Handhabung und Anwendung von Tabellen der Normalverteilung zu erleichtern, gibt es verschiedene Spalten, in denen zu gegebenen z-Werten ($z = \frac{x-\mu}{\sigma}$) folgende Wahrscheinlichkeiten angegeben sind:

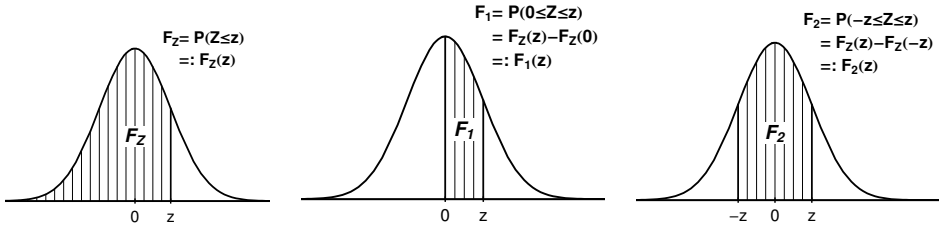
(1) $F_Z: \; P(Z \leq z)$ bzw. $P(X \leq \mu + z\sigma)$

(2) $F_1: \; P(0 \leq Z \leq z)$ bzw. $P(\mu \leq X \leq \mu + z\sigma)$

(3) $F_2: \; P(-z \leq Z \leq z)$ bzw. $P(\mu - z\sigma \leq X \leq \mu + z\sigma)$

Tabelle der Normalverteilung

Die folgende Abbildung gibt eine Veranschaulichung der Tabellenwerte:



In der angeführten Tabelle sind F_Z-, F_1- und F_2-Werte dargestellt.

Wegen der Symmetrie der Normalverteilung reicht es aus, die F_1-, F_2- und F_Z-Werte nur für positive Z-Werte zu tabellieren.

http://www.fernuni-hagen.de/ls_statistik/lehre/

z	F_Z	F_1	F_2	z	F_Z	F_1	F_2	z	F_Z	F_1	F_2
0,01	0,5040	0,0040	0,0080	0,51	0,6950	0,1950	0,3899	1,01	0,8438	0,3438	0,6875
0,02	0,5080	0,0080	0,0160	0,52	0,6985	0,1985	0,3969	1,02	0,8461	0,3461	0,6923
0,03	0,5120	0,0120	0,0239	0,53	0,7019	0,2019	0,4039	1,03	0,8485	0,3485	0,6970
0,04	0,5160	0,0160	0,0319	0,54	0,7054	0,2054	0,4108	1,04	0,8508	0,3508	0,7017
0,05	0,5199	0,0199	0,0399	0,55	0,7088	0,2088	0,4177	1,05	0,8531	0,3531	0,7063
0,06	0,5239	0,0239	0,0478	0,56	0,7123	0,2123	0,4245	1,06	0,8554	0,3554	0,7109
0,07	0,5279	0,0279	0,0558	0,57	0,7157	0,2157	0,4313	1,07	0,8577	0,3577	0,7154
0,08	0,5319	0,0319	0,0638	0,58	0,7190	0,2190	0,4381	1,08	0,8599	0,3599	0,7199
0,09	0,5359	0,0359	0,0717	0,59	0,7224	0,2224	0,4448	1,09	0,8621	0,3621	0,7243
0,10	0,5398	0,0398	0,0797	0,60	0,7257	0,2257	0,4515	1,10	0,8643	0,3643	0,7287
0,11	0,5438	0,0438	0,0876	0,61	0,7291	0,2291	0,4581	1,11	0,8665	0,3665	0,7330
0,12	0,5478	0,0478	0,0955	0,62	0,7324	0,2324	0,4647	1,12	0,8686	0,3686	0,7373
0,13	0,5517	0,0517	0,1034	0,63	0,7357	0,2357	0,4713	1,13	0,8708	0,3708	0,7415
0,14	0,5557	0,0557	0,1113	0,64	0,7389	0,2389	0,4778	1,14	0,8729	0,3729	0,7457
0,15	0,5596	0,0596	0,1192	0,65	0,7422	0,2422	0,4843	1,15	0,8749	0,3749	0,7499
0,16	0,5636	0,0636	0,1271	0,66	0,7454	0,2454	0,4907	1,16	0,8770	0,3770	0,7540
0,17	0,5675	0,0675	0,1350	0,67	0,7486	0,2486	0,4971	1,17	0,8790	0,3790	0,7580
0,18	0,5714	0,0714	0,1428	0,68	0,7517	0,2517	0,5035	1,18	0,8810	0,3810	0,7620
0,19	0,5753	0,0753	0,1507	0,69	0,7549	0,2549	0,5098	1,19	0,8830	0,3830	0,7660
0,20	0,5793	0,0793	0,1585	0,70	0,7580	0,2580	0,5161	1,20	0,8849	0,3849	0,7699
0,21	0,5832	0,0832	0,1663	0,71	0,7611	0,2611	0,5223	1,21	0,8869	0,3869	0,7737
0,22	0,5871	0,0871	0,1741	0,72	0,7642	0,2642	0,5285	1,22	0,8888	0,3888	0,7775
0,23	0,5910	0,0910	0,1819	0,73	0,7673	0,2673	0,5346	1,23	0,8907	0,3907	0,7813
0,24	0,5948	0,0948	0,1897	0,74	0,7704	0,2704	0,5407	1,24	0,8925	0,3925	0,7850
0,25	0,5987	0,0987	0,1974	0,75	0,7734	0,2734	0,5467	1,25	0,8944	0,3944	0,7887
0,26	0,6026	0,1026	0,2051	0,76	0,7764	0,2764	0,5527	1,26	0,8962	0,3962	0,7923
0,27	0,6064	0,1064	0,2128	0,77	0,7794	0,2794	0,5587	1,27	0,8980	0,3980	0,7959
0,28	0,6103	0,1103	0,2205	0,78	0,7823	0,2823	0,5646	1,28	0,8997	0,3997	0,7995
0,29	0,6141	0,1141	0,2282	0,79	0,7852	0,2852	0,5705	1,29	0,9015	0,4015	0,8029
0,30	0,6179	0,1179	0,2358	0,80	0,7881	0,2881	0,5763	1,30	0,9032	0,4032	0,8064
0,31	0,6217	0,1217	0,2434	0,81	0,7910	0,2910	0,5821	1,31	0,9049	0,4049	0,8098
0,32	0,6255	0,1255	0,2510	0,82	0,7939	0,2939	0,5878	1,32	0,9066	0,4066	0,8132
0,33	0,6293	0,1293	0,2586	0,83	0,7967	0,2967	0,5935	1,33	0,9082	0,4082	0,8165
0,34	0,6331	0,1331	0,2661	0,84	0,7995	0,2995	0,5991	1,34	0,9099	0,4099	0,8198
0,35	0,6368	0,1368	0,2737	0,85	0,8023	0,3023	0,6047	1,35	0,9115	0,4115	0,8230
0,36	0,6406	0,1406	0,2812	0,86	0,8051	0,3051	0,6102	1,36	0,9131	0,4131	0,8262
0,37	0,6443	0,1443	0,2886	0,87	0,8078	0,3078	0,6157	1,37	0,9147	0,4147	0,8293
0,38	0,6480	0,1480	0,2961	0,88	0,8106	0,3106	0,6211	1,38	0,9162	0,4162	0,8324
0,39	0,6517	0,1517	0,3035	0,89	0,8133	0,3133	0,6265	1,39	0,9177	0,4177	0,8355
0,40	0,6554	0,1554	0,3108	0,90	0,8159	0,3159	0,6319	1,40	0,9192	0,4192	0,8385
0,41	0,6591	0,1591	0,3182	0,91	0,8186	0,3186	0,6372	1,41	0,9207	0,4207	0,8415
0,42	0,6628	0,1628	0,3255	0,92	0,8212	0,3212	0,6424	1,42	0,9222	0,4222	0,8444
0,43	0,6664	0,1664	0,3328	0,93	0,8238	0,3238	0,6476	1,43	0,9236	0,4236	0,8473
0,44	0,6700	0,1700	0,3401	0,94	0,8264	0,3264	0,6528	1,44	0,9251	0,4251	0,8501
0,45	0,6736	0,1736	0,3473	0,95	0,8289	0,3289	0,6579	1,45	0,9265	0,4265	0,8529
0,46	0,6772	0,1772	0,3545	0,96	0,8315	0,3315	0,6629	1,46	0,9279	0,4279	0,8557
0,47	0,6808	0,1808	0,3616	0,97	0,8340	0,3340	0,6680	1,47	0,9292	0,4292	0,8584
0,48	0,6844	0,1844	0,3688	0,98	0,8365	0,3365	0,6729	1,48	0,9306	0,4306	0,8611
0,49	0,6879	0,1879	0,3759	0,99	0,8389	0,3389	0,6778	1,49	0,9319	0,4319	0,8638
0,50	0,6915	0,1915	0,3829	1,00	0,8413	0,3413	0,6827	1,50	0,9332	0,4332	0,8664

z	F_Z	F_1	F_2	z	F_Z	F_1	F_2	z	F_Z	F_1	F_2
1,51	0,9345	0,4345	0,8690	2,01	0,9778	0,4778	0,9556	2,51	0,9940	0,4940	0,9879
1,52	0,9357	0,4357	0,8715	2,02	0,9783	0,4783	0,9566	2,52	0,9941	0,4941	0,9883
1,53	0,9370	0,4370	0,8740	2,03	0,9788	0,4788	0,9576	2,53	0,9943	0,4943	0,9886
1,54	0,9382	0,4382	0,8764	2,04	0,9793	0,4793	0,9586	2,54	0,9945	0,4945	0,9889
1,55	0,9394	0,4394	0,8789	2,05	0,9798	0,4798	0,9596	2,55	0,9946	0,4946	0,9892
1,56	0,9406	0,4406	0,8812	2,06	0,9803	0,4803	0,9606	2,56	0,9948	0,4948	0,9895
1,57	0,9418	0,4418	0,8836	2,07	0,9808	0,4808	0,9615	2,57	0,9949	0,4949	0,9898
1,58	0,9429	0,4429	0,8859	2,08	0,9812	0,4812	0,9625	2,58	0,9951	0,4951	0,9901
1,59	0,9441	0,4441	0,8882	2,09	0,9817	0,4817	0,9634	2,59	0,9952	0,4952	0,9904
1,60	0,9452	0,4452	0,8904	2,10	0,9821	0,4821	0,9643	2,60	0,9953	0,4953	0,9907
1,61	0,9463	0,4463	0,8926	2,11	0,9826	0,4826	0,9651	2,61	0,9955	0,4955	0,9909
1,62	0,9474	0,4474	0,8948	2,12	0,9830	0,4830	0,9660	2,62	0,9956	0,4956	0,9912
1,63	0,9484	0,4484	0,8969	2,13	0,9834	0,4834	0,9668	2,63	0,9957	0,4957	0,9915
1,64	0,9495	0,4495	0,8990	2,14	0,9838	0,4838	0,9676	2,64	0,9959	0,4959	0,9917
1,65	0,9505	0,4505	0,9011	2,15	0,9842	0,4842	0,9684	2,65	0,9960	0,4960	0,9920
1,66	0,9515	0,4515	0,9031	2,16	0,9846	0,4846	0,9692	2,66	0,9961	0,4961	0,9922
1,67	0,9525	0,4525	0,9051	2,17	0,9850	0,4850	0,9700	2,67	0,9962	0,4962	0,9924
1,68	0,9535	0,4535	0,9070	2,18	0,9854	0,4854	0,9707	2,68	0,9963	0,4963	0,9926
1,69	0,9545	0,4545	0,9090	2,19	0,9857	0,4857	0,9715	2,69	0,9964	0,4964	0,9929
1,70	0,9554	0,4554	0,9109	2,20	0,9861	0,4861	0,9722	2,70	0,9965	0,4965	0,9931
1,71	0,9564	0,4564	0,9127	2,21	0,9864	0,4864	0,9729	2,71	0,9966	0,4966	0,9933
1,72	0,9573	0,4573	0,9146	2,22	0,9868	0,4868	0,9736	2,72	0,9967	0,4967	0,9935
1,73	0,9582	0,4582	0,9164	2,23	0,9871	0,4871	0,9743	2,73	0,9968	0,4968	0,9937
1,74	0,9591	0,4591	0,9181	2,24	0,9875	0,4875	0,9749	2,74	0,9969	0,4969	0,9939
1,75	0,9599	0,4599	0,9199	2,25	0,9878	0,4878	0,9756	2,75	0,9970	0,4970	0,9940
1,76	0,9608	0,4608	0,9216	2,26	0,9881	0,4881	0,9762	2,76	0,9971	0,4971	0,9942
1,77	0,9616	0,4616	0,9233	2,27	0,9884	0,4884	0,9768	2,77	0,9972	0,4972	0,9944
1,78	0,9625	0,4625	0,9249	2,28	0,9887	0,4887	0,9774	2,78	0,9973	0,4973	0,9946
1,79	0,9633	0,4633	0,9265	2,29	0,9890	0,4890	0,9780	2,79	0,9974	0,4974	0,9947
1,80	0,9641	0,4641	0,9281	2,30	0,9893	0,4893	0,9786	2,80	0,9974	0,4974	0,9949
1,81	0,9649	0,4649	0,9297	2,31	0,9896	0,4896	0,9791	2,81	0,9975	0,4975	0,9950
1,82	0,9656	0,4656	0,9312	2,32	0,9898	0,4898	0,9797	2,82	0,9976	0,4976	0,9952
1,83	0,9664	0,4664	0,9328	2,33	0,9901	0,4901	0,9802	2,83	0,9977	0,4977	0,9953
1,84	0,9671	0,4671	0,9342	2,34	0,9904	0,4904	0,9807	2,84	0,9977	0,4977	0,9955
1,85	0,9678	0,4678	0,9357	2,35	0,9906	0,4906	0,9812	2,85	0,9978	0,4978	0,9956
1,86	0,9686	0,4686	0,9371	2,36	0,9909	0,4909	0,9817	2,86	0,9979	0,4979	0,9958
1,87	0,9693	0,4693	0,9385	2,37	0,9911	0,4911	0,9822	2,87	0,9979	0,4979	0,9959
1,88	0,9699	0,4699	0,9399	2,38	0,9913	0,4913	0,9827	2,88	0,9980	0,4980	0,9960
1,89	0,9706	0,4706	0,9412	2,39	0,9916	0,4916	0,9832	2,89	0,9981	0,4981	0,9961
1,90	0,9713	0,4713	0,9426	2,40	0,9918	0,4918	0,9836	2,90	0,9981	0,4981	0,9963
1,91	0,9719	0,4719	0,9439	2,41	0,9920	0,4920	0,9840	2,91	0,9982	0,4982	0,9964
1,92	0,9726	0,4726	0,9451	2,42	0,9922	0,4922	0,9845	2,92	0,9982	0,4982	0,9965
1,93	0,9732	0,4732	0,9464	2,43	0,9925	0,4925	0,9849	2,93	0,9983	0,4983	0,9966
1,94	0,9738	0,4738	0,9476	2,44	0,9927	0,4927	0,9853	2,94	0,9984	0,4984	0,9967
1,95	0,9744	0,4744	0,9488	2,45	0,9929	0,4929	0,9857	2,95	0,9984	0,4984	0,9968
1,96	0,9750	0,4750	0,9500	2,46	0,9931	0,4931	0,9861	2,96	0,9985	0,4985	0,9969
1,97	0,9756	0,4756	0,9512	2,47	0,9932	0,4932	0,9865	2,97	0,9985	0,4985	0,9970
1,98	0,9761	0,4761	0,9523	2,48	0,9934	0,4934	0,9869	2,98	0,9986	0,4986	0,9971
1,99	0,9767	0,4767	0,9534	2,49	0,9936	0,4936	0,9872	2,99	0,9986	0,4986	0,9972
2,00	0,9772	0,4772	0,9545	2,50	0,9938	0,4938	0,9876	3,00	0,9987	0,4987	0,9973

12.2 χ^2-Verteilung

Sind $X_1, X_2, ..., X_\nu$ standardnormalverteilte, stochastisch unabhängige Zufallsvariablen, so ist die Zufallsvariable

χ^2-Verteilung

$$Y = \sum_{i=1}^{\nu} X_i^2$$

χ^2-verteilt mit ν Freiheitsgraden. Die Dichtefunktion lautet

Dichtefunktion

$$f_Y(y) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{\nu}{2})} y^{(\frac{\nu-2}{2})} e^{(-\frac{y}{2})} \qquad \text{für } y > 0.$$

Dabei ist $\Gamma(\frac{\nu}{2})$ die Eulersche **Gammafunktion** mit folgenden Eigenschaften:

$$\begin{aligned}
(1) \quad \Gamma\left(\frac{\nu}{2}\right) &= \left(\frac{\nu}{2} - 1\right)\Gamma\left(\frac{\nu}{2} - 1\right), \quad \nu > 2 \\
(2) \quad \Gamma\left(\frac{\nu}{2}\right) &= \left(\frac{\nu}{2} - 1\right)! \quad \nu = 2n, \quad n \in \mathbb{N} \\
(3) \quad \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi}
\end{aligned}$$

Erwartungswert

$$E(Y) = \nu$$

Varianz

$$\mathrm{Var}(Y) = 2\nu$$

Parameter der χ^2-Verteilung

ν (Anzahl der Freiheitsgrade)
Y heißt auch $\chi^2(\nu) -$ verteilt

Approximation

Für $\nu > 100$ ist eine $\chi^2(\nu)$-verteilte Zufallsvariable näherungsweise normalverteilt mit Erwartungswert ν und Varianz 2ν. Die Zufallsvariable $Z = \sqrt{2Y} - \sqrt{2\nu - 1}$ ist für $\nu > 30$ näherungsweise standardnormalverteilt.

Tabelle der χ^2-Verteilung

Für verschiedene Freiheitsgrade und für verschiedene Werte von $F_Y(y)$ (kumulierte Verteilung) sind die zugehörigen y-Werte (Quantile) tabelliert.

http://www.fernuni-hagen.de/ls_statistik/lehre/

	$F_Y(y)$																
ν	0,001	0,005	0,01	0,02	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,98	0,99	0,995	0,999
1	-	-	-	-	-	0,004	0,016	0,102	0,455	1,323	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	0,002	0,010	0,020	0,040	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	0,024	0,072	0,115	0,185	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	9,837	11,345	12,838	16,266
4	0,091	0,207	0,297	0,429	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	11,668	13,277	14,860	18,466
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,832	13,388	15,086	16,750	20,515
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	0,599	0,989	1,239	1,564	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	16,622	18,475	20,278	24,321
8	0,857	1,344	1,647	2,032	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	18,168	20,090	21,955	26,124
9	1,152	1,735	2,088	2,532	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	1,479	2,156	2,558	3,059	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	2,214	3,074	3,571	4,178	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	2,617	3,565	4,107	4,765	5,009	5,892	7,041	9,299	12,340	15,984	19,812	22,362	24,736	25,471	27,688	29,819	34,527
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	26,873	29,141	31,319	36,124
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547	11,037	14,339	18,245	22,307	24,996	27,488	28,259	30,578	32,801	37,698
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	4,416	5,697	6,408	7,255	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	30,995	33,409	35,718	40,791
18	4,905	6,265	7,015	7,906	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	33,687	36,191	38,582	43,819
20	5,921	7,434	8,260	9,237	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	35,020	37,566	39,997	45,314
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	36,343	38,932	41,401	46,796
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	7,529	9,260	10,196	11,293	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	8,085	9,886	10,856	11,992	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	40,270	42,980	45,558	51,179
25	8,649	10,520	11,524	12,697	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	41,566	44,314	46,928	52,619
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	20,843	25,336	30,435	35,563	38,885	41,923	42,856	45,642	48,290	54,051
27	9,803	11,808	12,878	14,125	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,195	44,140	46,963	49,645	55,475
28	10,391	12,461	13,565	14,847	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	45,419	48,278	50,994	56,892
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	46,693	49,588	52,335	58,301
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	47,962	50,892	53,672	59,702
31	12,196	14,458	15,655	17,042	17,539	19,281	21,434	25,390	30,336	35,887	41,422	44,985	48,232	49,226	52,191	55,002	61,098
32	12,810	15,134	16,362	17,783	18,291	20,072	22,271	26,304	31,336	36,973	42,585	46,194	49,480	50,487	53,486	56,328	62,487
33	13,431	15,815	17,073	18,527	19,047	20,867	23,110	27,219	32,336	38,058	43,745	47,400	50,725	51,743	54,775	57,648	63,869

	$F_Y(y)$																
ν	0,001	0,005	0,01	0,02	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,98	0,99	0,995	0,999
34	14,057	16,501	17,789	19,275	19,806	21,664	23,952	28,136	33,336	39,141	44,903	48,602	51,966	52,995	56,061	58,964	65,247
35	14,688	17,192	18,509	20,027	20,569	22,465	24,797	29,054	34,336	40,223	46,059	49,802	53,203	54,244	57,342	60,275	66,619
36	15,324	17,887	19,233	20,783	21,336	23,269	25,643	29,973	35,336	41,304	47,212	50,998	54,437	55,489	58,619	61,581	67,985
37	15,965	18,586	19,960	21,542	22,106	24,075	26,492	30,893	36,336	42,383	48,363	52,192	55,668	56,730	59,893	62,883	69,348
38	16,611	19,289	20,691	22,304	22,878	24,884	27,343	31,815	37,335	43,462	49,513	53,384	56,895	57,969	61,162	64,181	70,704
39	17,261	19,996	21,426	23,069	23,654	25,695	28,196	32,737	38,335	44,539	50,660	54,572	58,120	59,204	62,428	65,475	72,055
40	17,917	20,707	22,164	23,838	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	60,436	63,691	66,766	73,403
41	18,576	21,421	22,906	24,609	25,215	27,326	29,907	34,585	40,335	46,692	52,949	56,942	60,561	61,665	64,950	68,053	74,744
42	19,238	22,138	23,650	25,383	25,999	28,144	30,765	35,510	41,335	47,766	54,090	58,124	61,777	62,892	66,206	69,336	76,084
43	19,905	22,860	24,398	26,159	26,785	28,965	31,625	36,436	42,335	48,840	55,230	59,304	62,990	64,116	67,459	70,616	77,418
44	20,576	23,584	25,148	26,939	27,575	29,787	32,487	37,363	43,335	49,913	56,369	60,481	64,201	65,337	68,710	71,892	78,749
45	21,251	24,311	25,901	27,720	28,366	30,612	33,350	38,291	44,335	50,985	57,505	61,656	65,410	66,555	69,957	73,166	80,078
46	21,929	25,041	26,657	28,504	29,160	31,439	34,215	39,220	45,335	52,056	58,641	62,830	66,616	67,771	71,201	74,437	81,400
47	22,610	25,775	27,416	29,291	29,956	32,268	35,081	40,149	46,335	53,127	59,774	64,001	67,821	68,985	72,443	75,704	82,720
48	23,294	26,511	28,177	30,080	30,754	33,098	35,949	41,079	47,335	54,196	60,907	65,171	69,023	70,197	73,683	76,969	84,037
49	23,983	27,249	28,941	30,871	31,555	33,930	36,818	42,010	48,335	55,265	62,038	66,339	70,222	71,406	74,919	78,231	85,350
50	24,674	27,991	29,707	31,664	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	72,613	76,154	79,490	86,660
51	25,368	28,735	30,475	32,459	33,162	35,600	38,560	43,874	50,335	57,401	64,295	68,669	72,616	73,818	77,386	80,746	87,967
52	26,065	29,481	31,246	33,256	33,968	36,437	39,433	44,807	51,335	58,468	65,422	69,832	73,810	75,021	78,616	82,001	89,272
53	26,765	30,230	32,019	34,055	34,776	37,276	40,308	45,741	52,335	59,534	66,548	70,993	75,002	76,223	79,843	83,253	90,573
54	27,467	30,981	32,793	34,856	35,586	38,116	41,183	46,676	53,335	60,600	67,673	72,153	76,192	77,422	81,069	84,502	91,871
55	28,173	31,735	33,571	35,659	36,398	38,958	42,060	47,610	54,335	61,665	68,796	73,311	77,380	78,619	82,292	85,749	93,167
56	28,881	32,491	34,350	36,464	37,212	39,801	42,937	48,546	55,335	62,729	69,919	74,468	78,567	79,815	83,514	86,994	94,462
57	29,592	33,248	35,131	37,270	38,027	40,646	43,816	49,482	56,335	63,793	71,040	75,624	79,752	81,009	84,733	88,237	95,750
58	30,305	34,008	35,914	38,078	38,844	41,492	44,696	50,419	57,335	64,857	72,160	76,778	80,936	82,201	85,950	89,477	97,038
59	31,021	34,770	36,698	38,888	39,662	42,339	45,577	51,356	58,335	65,919	73,279	77,930	82,117	83,391	87,166	90,715	98,324
60	31,738	35,534	37,485	39,699	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	84,580	88,379	91,952	99,608
61	32,458	36,300	38,273	40,512	41,303	44,038	47,342	53,232	60,335	68,043	75,514	80,232	84,476	85,767	89,591	93,186	100,887
62	33,181	37,068	39,063	41,327	42,126	44,889	48,226	54,171	61,335	69,104	76,630	81,381	85,654	86,953	90,802	94,419	102,165
63	33,905	37,838	39,855	42,143	42,950	45,741	49,111	55,110	62,335	70,165	77,745	82,529	86,830	88,137	92,010	95,649	103,442
64	34,632	38,610	40,649	42,960	43,776	46,595	49,996	56,050	63,335	71,225	78,860	83,675	88,004	89,320	93,217	96,878	104,717
65	35,362	39,383	41,444	43,779	44,603	47,450	50,883	56,990	64,335	72,285	79,973	84,821	89,177	90,501	94,422	98,105	105,988
66	36,092	40,158	42,240	44,599	45,431	48,305	51,770	57,931	65,335	73,344	81,085	85,965	90,349	91,681	95,626	99,330	107,257

	$F_Y(y)$																
ν	0,001	0,005	0,01	0,02	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,98	0,99	0,995	0,999
67	36,826	40,935	43,038	45,421	46,261	49,162	52,659	58,872	66,335	74,403	82,197	87,108	91,519	92,860	96,828	100,554	108,525
68	37,561	41,714	43,838	46,244	47,092	50,020	53,548	59,814	67,335	75,461	83,308	88,250	92,688	94,037	98,028	101,776	109,793
69	38,298	42,493	44,639	47,068	47,924	50,879	54,438	60,756	68,334	76,519	84,418	89,391	93,856	95,213	99,227	102,996	111,055
70	39,036	43,275	45,442	47,893	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	96,387	100,425	104,215	112,317
71	39,776	44,058	46,246	48,720	49,592	52,600	56,221	62,641	70,334	78,634	86,635	91,670	96,189	97,561	101,621	105,432	113,577
72	40,520	44,843	47,051	49,548	50,428	53,462	57,113	63,585	71,334	79,690	87,743	92,808	97,353	98,733	102,816	106,647	114,834
73	41,263	45,629	47,858	50,377	51,265	54,325	58,006	64,528	72,334	80,747	88,850	93,945	98,516	99,904	104,010	107,862	116,092
74	42,009	46,417	48,666	51,208	52,103	55,189	58,900	65,472	73,334	81,803	89,956	95,081	99,678	101,074	105,202	109,074	117,347
75	42,757	47,206	49,475	52,039	52,942	56,054	59,795	66,417	74,334	82,858	91,061	96,217	100,839	102,243	106,393	110,285	118,599
76	43,507	47,996	50,286	52,872	53,782	56,920	60,690	67,362	75,334	83,913	92,166	97,351	101,999	103,410	107,582	111,495	119,850
77	44,257	48,788	51,097	53,705	54,623	57,786	61,586	68,307	76,334	84,968	93,270	98,484	103,158	104,576	108,771	112,704	121,101
78	45,011	49,581	51,910	54,540	55,466	58,654	62,483	69,252	77,334	86,022	94,374	99,617	104,316	105,742	109,958	113,911	122,347
79	45,764	50,376	52,725	55,376	56,309	59,522	63,380	70,198	78,334	87,077	95,476	100,749	105,473	106,906	111,144	115,116	123,595
80	46,520	51,172	53,540	56,213	57,153	60,391	64,278	71,145	79,334	88,130	96,578	101,879	106,629	108,069	112,329	116,321	124,839
81	47,276	51,969	54,357	57,051	57,998	61,262	65,176	72,091	80,334	89,184	97,680	103,010	107,783	109,231	113,512	117,524	126,084
82	48,036	52,767	55,174	57,890	58,845	62,132	66,076	73,038	81,334	90,237	98,780	104,139	108,937	110,393	114,695	118,726	127,324
83	48,795	53,567	55,993	58,730	59,692	63,004	66,976	73,985	82,334	91,289	99,880	105,267	110,090	111,553	115,876	119,927	128,565
84	49,558	54,368	56,813	59,570	60,540	63,876	67,876	74,933	83,334	92,342	100,980	106,395	111,242	112,712	117,057	121,126	129,802
85	50,320	55,170	57,634	60,412	61,389	64,749	68,777	75,881	84,334	93,394	102,079	107,522	112,393	113,871	118,236	122,324	131,043
86	51,084	55,973	58,456	61,255	62,239	65,623	69,679	76,829	85,334	94,446	103,177	108,648	113,544	115,028	119,414	123,522	132,276
87	51,850	56,777	59,279	62,098	63,089	66,498	70,581	77,777	86,334	95,497	104,275	109,773	114,693	116,184	120,591	124,718	133,511
88	52,617	57,582	60,103	62,943	63,941	67,373	71,484	78,726	87,334	96,548	105,372	110,898	115,841	117,340	121,767	125,912	134,746
89	53,385	58,389	60,928	63,788	64,793	68,249	72,387	79,675	88,334	97,599	106,469	112,022	116,989	118,495	122,942	127,106	135,977
90	54,156	59,196	61,754	64,635	65,647	69,126	73,291	80,625	89,334	98,650	107,565	113,145	118,136	119,648	124,116	128,299	137,208
91	54,925	60,005	62,581	65,482	66,501	70,003	74,196	81,574	90,334	99,700	108,661	114,268	119,282	120,801	125,289	129,490	138,437
92	55,698	60,815	63,409	66,330	67,356	70,882	75,100	82,524	91,334	100,750	109,756	115,390	120,427	121,953	126,462	130,681	139,667
93	56,471	61,625	64,238	67,179	68,211	71,760	76,006	83,474	92,334	101,800	110,850	116,511	121,571	123,105	127,633	131,871	140,894
94	57,246	62,437	65,068	68,028	69,068	72,640	76,912	84,425	93,334	102,850	111,944	117,632	122,715	124,255	128,803	133,059	142,118
95	58,022	63,250	65,898	68,879	69,925	73,520	77,818	85,376	94,334	103,899	113,038	118,752	123,858	125,405	129,973	134,247	143,343
96	58,800	64,063	66,730	69,730	70,783	74,401	78,725	86,327	95,334	104,948	114,131	119,871	125,000	126,554	131,141	135,433	144,566
97	59,577	64,878	67,562	70,582	71,642	75,282	79,633	87,278	96,334	105,997	115,223	120,990	126,141	127,702	132,309	136,619	145,789
98	60,356	65,693	68,396	71,434	72,501	76,164	80,541	88,229	97,334	107,045	116,315	122,108	127,282	128,849	133,476	137,803	147,009
99	61,136	66,510	69,230	72,288	73,361	77,046	81,449	89,181	98,334	108,093	117,407	123,225	128,422	129,996	134,641	138,987	148,230
100	61,918	67,328	70,065	73,142	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	131,142	135,807	140,170	149,449

12.3 F-Verteilung

Y_1 und Y_2 seien χ^2-verteilte, voneinander unabhängige Zufallsvariablen mit m bzw. n Freiheitsgraden. Die Zufallsvariable

F-Verteilung

$$X = \frac{Y_1/m}{Y_2/n}$$

ist F-verteilt.

Dichtefunktion

$$f_X(x; m, n) = \left(\frac{m}{2}\right)^{\frac{m}{2}} \cdot \left(\frac{n}{2}\right)^{\frac{n}{2}} \cdot \frac{\Gamma\left(\frac{m}{2} + \frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \cdot \frac{x^{(\frac{m}{2}-1)}}{\left(\frac{m}{2}x + \frac{n}{2}\right)^{\frac{m+n}{2}}}$$

Erwartungswert

$$E(X) = \frac{n}{n-2} \qquad n > 2$$

Varianz

$$\mathrm{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \qquad n > 4$$

Parameter der F-Verteilung

m, n (Anzahl der Freiheitsgrade)

Tabelle der F-Verteilung

In Abhängigkeit von m und n sind die Werte x der F-verteilten Zufallsvariablen tabelliert, bei denen die Verteilungsfunktion den Wert 0,95 bzw. 0,99 annimmt.

http://www.fernuni-hagen.de/ls_statistik/lehre/

$F(x) = 0,95$

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246	246	247	247	248	248	250	251	252	253	254	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	10,14	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66	8,62	8,59	8,58	8,55	8,54	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80	5,75	5,72	5,70	5,66	5,65	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,50	4,46	4,44	4,41	4,39	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,81	3,77	3,75	3,71	3,69	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,38	3,34	3,32	3,27	3,25	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,08	3,04	3,02	2,97	2,95	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,86	2,83	2,80	2,76	2,73	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,70	2,66	2,64	2,59	2,56	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65	2,57	2,53	2,51	2,46	2,43	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,47	2,43	2,40	2,35	2,32	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,38	2,34	2,31	2,26	2,23	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,31	2,27	2,24	2,19	2,16	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,25	2,20	2,18	2,12	2,10	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,19	2,15	2,12	2,07	2,04	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,15	2,10	2,08	2,02	1,99	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,11	2,06	2,04	1,98	1,95	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16	2,07	2,03	2,00	1,94	1,91	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,04	1,99	1,97	1,91	1,88	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,01	1,96	1,94	1,88	1,84	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15	2,13	2,11	2,10	2,08	2,07	1,98	1,94	1,91	1,85	1,82	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13	2,11	2,09	2,08	2,06	2,05	1,96	1,91	1,88	1,82	1,79	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11	2,09	2,07	2,05	2,04	2,03	1,94	1,89	1,86	1,80	1,77	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09	2,07	2,05	2,04	2,02	2,01	1,92	1,87	1,84	1,78	1,75	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,90	1,85	1,82	1,76	1,73	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,88	1,84	1,81	1,74	1,71	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,87	1,82	1,79	1,73	1,69	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03	2,01	1,99	1,97	1,96	1,94	1,85	1,81	1,77	1,71	1,67	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01	1,99	1,98	1,96	1,95	1,93	1,84	1,79	1,76	1,70	1,66	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92	1,90	1,89	1,87	1,85	1,84	1,74	1,69	1,66	1,59	1,55	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87	1,85	1,83	1,81	1,80	1,78	1,69	1,63	1,60	1,52	1,48	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84	1,82	1,80	1,78	1,76	1,75	1,65	1,59	1,56	1,48	1,44	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81	1,79	1,77	1,75	1,74	1,72	1,62	1,57	1,53	1,45	1,40	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79	1,77	1,75	1,73	1,72	1,70	1,60	1,54	1,51	1,43	1,38	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,59	1,53	1,49	1,41	1,36	1,30
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77	1,75	1,73	1,71	1,69	1,68	1,57	1,52	1,48	1,39	1,34	1,28
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87	1,83	1,80	1,78	1,75	1,73	1,71	1,69	1,67	1,66	1,55	1,50	1,46	1,37	1,32	1,25
140	3,91	3,06	2,67	2,44	2,28	2,16	2,08	2,01	1,95	1,90	1,86	1,82	1,79	1,76	1,74	1,72	1,70	1,68	1,66	1,65	1,54	1,48	1,44	1,35	1,30	1,23
160	3,90	3,05	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,81	1,78	1,75	1,73	1,71	1,69	1,67	1,65	1,64	1,53	1,47	1,43	1,34	1,28	1,21
180	3,89	3,05	2,65	2,42	2,26	2,15	2,06	1,99	1,93	1,88	1,84	1,81	1,77	1,75	1,72	1,70	1,68	1,66	1,64	1,63	1,52	1,46	1,42	1,33	1,27	1,20
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72	1,69	1,67	1,66	1,64	1,62	1,52	1,46	1,41	1,32	1,26	1,19
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72	1,69	1,67	1,64	1,62	1,60	1,59	1,57	1,46	1,39	1,35	1,24	1,17	1,00

$F(x) = 0,99$

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107	6126	6143	6157	6170	6181	6191	6201	6209	6260	6286	6302	6334	6350	6366
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0	26,9	26,9	26,8	26,8	26,8	26,7	26,7	26,5	26,4	26,4	26,2	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,5	14,4	14,3	14,2	14,2	14,2	14,1	14,1	14,0	14,0	13,8	13,7	13,7	13,6	13,5	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55	9,38	9,29	9,24	9,13	9,08	9,02
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40	7,23	7,14	7,09	6,99	6,93	6,88
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16	5,99	5,91	5,86	5,75	5,70	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36	5,20	5,12	5,07	4,96	4,91	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81	4,65	4,57	4,52	4,41	4,36	4,31
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41	4,25	4,17	4,12	4,01	3,96	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10	3,94	3,86	3,81	3,71	3,66	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86	3,70	3,62	3,57	3,47	3,41	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66	3,51	3,43	3,38	3,27	3,22	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51	3,35	3,27	3,22	3,11	3,06	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37	3,21	3,13	3,08	2,98	2,92	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26	3,10	3,02	2,97	2,86	2,81	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16	3,00	2,92	2,87	2,76	2,71	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08	2,92	2,84	2,78	2,68	2,62	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00	2,84	2,76	2,71	2,60	2,55	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94	2,78	2,69	2,64	2,54	2,48	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03	2,99	2,96	2,93	2,90	2,88	2,72	2,64	2,58	2,48	2,42	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98	2,94	2,91	2,88	2,85	2,83	2,67	2,58	2,53	2,42	2,36	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93	2,89	2,86	2,83	2,80	2,78	2,62	2,54	2,48	2,37	2,32	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,58	2,49	2,44	2,33	2,27	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85	2,81	2,78	2,75	2,72	2,70	2,54	2,45	2,40	2,29	2,23	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90	2,86	2,81	2,78	2,75	2,72	2,69	2,66	2,50	2,42	2,36	2,25	2,19	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87	2,82	2,78	2,75	2,71	2,68	2,66	2,63	2,47	2,38	2,33	2,22	2,16	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84	2,79	2,75	2,72	2,68	2,65	2,63	2,60	2,44	2,35	2,30	2,19	2,13	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81	2,77	2,73	2,69	2,66	2,63	2,60	2,57	2,41	2,33	2,27	2,16	2,10	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70	2,66	2,63	2,60	2,57	2,55	2,39	2,30	2,25	2,13	2,07	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52	2,48	2,45	2,42	2,39	2,37	2,20	2,11	2,06	1,94	1,87	1,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56	2,51	2,46	2,42	2,38	2,35	2,32	2,29	2,27	2,10	2,01	1,95	1,82	1,76	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,03	1,94	1,88	1,75	1,68	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31	2,27	2,23	2,20	2,18	2,15	1,98	1,89	1,83	1,70	1,62	1,54
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27	2,23	2,20	2,17	2,14	2,12	1,94	1,85	1,79	1,65	1,58	1,49
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33	2,29	2,24	2,21	2,17	2,14	2,11	2,09	1,92	1,82	1,76	1,62	1,55	1,46
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,27	2,22	2,19	2,15	2,12	2,09	2,07	1,89	1,80	1,74	1,60	1,52	1,43
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	2,28	2,23	2,19	2,15	2,12	2,09	2,06	2,03	1,86	1,76	1,70	1,56	1,48	1,38
140	6,82	4,76	3,92	3,46	3,15	2,93	2,77	2,64	2,54	2,45	2,38	2,31	2,26	2,21	2,17	2,13	2,10	2,07	2,04	2,01	1,84	1,74	1,67	1,53	1,45	1,35
160	6,80	4,74	3,91	3,44	3,13	2,92	2,75	2,62	2,52	2,43	2,36	2,30	2,24	2,20	2,15	2,11	2,08	2,05	2,02	1,99	1,82	1,72	1,66	1,51	1,42	1,32
180	6,78	4,73	3,89	3,43	3,12	2,90	2,74	2,61	2,51	2,42	2,35	2,28	2,23	2,18	2,14	2,10	2,07	2,04	2,01	1,98	1,81	1,71	1,64	1,49	1,41	1,30
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22	2,17	2,13	2,09	2,06	2,03	2,00	1,97	1,79	1,69	1,63	1,48	1,39	1,28
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18	2,13	2,08	2,04	2,00	1,97	1,93	1,90	1,88	1,70	1,59	1,52	1,36	1,25	1,00

12.4 Student-t-Verteilung

Es seien X und Y unabhängige Zufallsvariablen und X sei $N(0,1)$- und Y sei $\chi^2(\nu)$-verteilt. Dann ist

Student-t-Verteilung

$$T = \frac{X}{\sqrt{Y/\nu}}$$

Student-verteilt mit ν Freiheitsgraden.
Die Dichtefunktion lautet

Dichtefunktion der Student-t-Verteilung

$$f_T(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}(1+t^2/\nu)^{-(\nu+1)/2}$$

Erwartungswert

$$E(T) = 0 \quad \text{für} \quad \nu \geq 2$$

Varianz

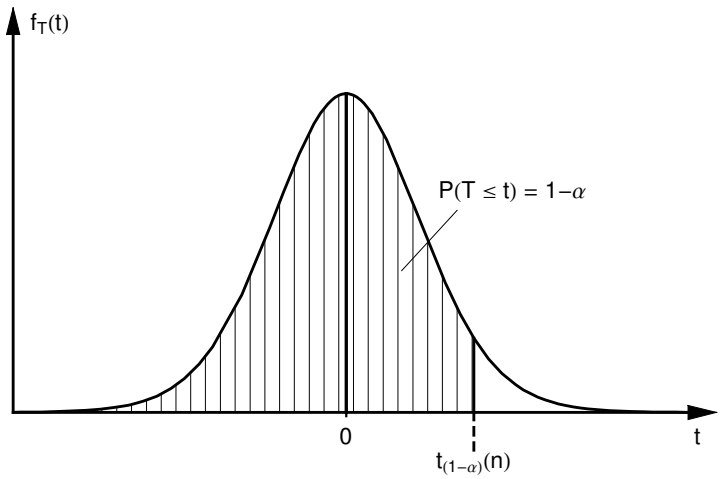
$$\mathrm{Var}(T) = \frac{\nu}{\nu-2} \qquad \text{für} \quad \nu \geq 3$$

Parameter der Student-Verteilung

ν (Anzahl der Freiheitsgrade)

Für verschiedene Freiheitsgrade ν und für verschiedene Wahrscheinlichkeiten $p = P(T \leq t)$ sind die entsprechenden t-Werte $t(1-\alpha, \nu)$ tabelliert.

Tabelle der Student-Verteilung



http://www.fernuni-hagen.de/ls_statistik/lehre/

Quantile der t-Verteilung $t(1-\alpha,\nu)$

	$P(T \leq t) =$							
ν	0,75	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	318,300	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	22,330	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,210	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,679	1,299	1,676	2,009	2,403	2,678	3,261	3,496
100	0,677	1,290	1,660	1,984	2,364	2,626	3,174	3,390
200	0,676	1,286	1,652	1,972	2,345	2,601	3,131	3,340
∞	0,675	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Bonferroni-t-Statistik: $t(1-\alpha/(2k),\nu), \alpha = 0.05$

$\alpha = 0.05$																	
	k																
ν	2	3	4	5	6	7	8	9	10	15	20	25	30	35	40	45	50
2	6.21	7.65	8.86	9.92	10.89	11.77	12.59	13.36	14.09	17.28	19.96	22.33	24.46	26.43	28.26	29.97	31.60
3	4.18	4.86	5.39	5.84	6.23	6.58	6.90	7.18	7.45	8.58	9.46	10.21	10.87	11.45	11.98	12.47	12.92
4	3.50	3.96	4.31	4.60	4.85	5.07	5.26	5.44	5.60	6.25	6.76	7.17	7.53	7.84	8.12	8.38	8.61
5	3.16	3.53	3.81	4.03	4.22	4.38	4.53	4.66	4.77	5.25	5.60	5.89	6.14	6.35	6.54	6.71	6.87
6	2.97	3.29	3.52	3.71	3.86	4.00	4.12	4.22	4.32	4.70	4.98	5.21	5.40	5.56	5.71	5.84	5.96
7	2.84	3.13	3.34	3.50	3.64	3.75	3.86	3.95	4.03	4.36	4.59	4.79	4.94	5.08	5.20	5.31	5.41
8	2.75	3.02	3.21	3.36	3.48	3.58	3.68	3.76	3.83	4.12	4.33	4.50	4.64	4.76	4.86	4.96	5.04
9	2.69	2.93	3.11	3.25	3.36	3.46	3.55	3.62	3.69	3.95	4.15	4.30	4.42	4.53	4.62	4.71	4.78
10	2.63	2.87	3.04	3.17	3.28	3.37	3.45	3.52	3.58	3.83	4.00	4.14	4.26	4.36	4.44	4.52	4.59
15	2.49	2.69	2.84	2.95	3.04	3.11	3.18	3.23	3.29	3.48	3.62	3.73	3.82	3.90	3.96	4.02	4.07
20	2.42	2.61	2.74	2.85	2.93	3.00	3.06	3.11	3.15	3.33	3.46	3.55	3.63	3.70	3.75	3.80	3.85
25	2.38	2.57	2.69	2.79	2.86	2.93	2.99	3.03	3.08	3.24	3.36	3.45	3.52	3.58	3.64	3.68	3.73
30	2.36	2.54	2.66	2.75	2.82	2.89	2.94	2.99	3.03	3.19	3.30	3.39	3.45	3.51	3.56	3.61	3.65
40	2.33	2.50	2.62	2.70	2.78	2.84	2.89	2.93	2.97	3.12	3.23	3.31	3.37	3.43	3.47	3.51	3.55
50	2.31	2.48	2.59	2.68	2.75	2.81	2.85	2.90	2.94	3.08	3.18	3.26	3.32	3.38	3.42	3.46	3.50
60	2.30	2.46	2.58	2.66	2.73	2.79	2.83	2.88	2.91	3.06	3.16	3.23	3.29	3.34	3.39	3.43	3.46
100	2.28	2.43	2.54	2.63	2.69	2.75	2.79	2.83	2.87	3.01	3.10	3.17	3.23	3.28	3.32	3.36	3.39
∞	2.24	2.39	2.50	2.58	2.64	2.69	2.73	2.77	2.81	2.94	3.02	3.09	3.14	3.19	3.23	3.26	3.29

Bonferroni-t-Statistik: $t(1-\alpha/(2k),\nu), \alpha = 0.01$

$\alpha = 0.01$																	
	k																
ν	2	3	4	5	6	7	8	9	10	15	20	25	30	35	40	45	50
2	14.09	17.28	19.96	22.33	24.46	26.43	28.26	29.97	31.6	38.71	44.70	49.98	54.76	59.15	63.23	67.07	70.70
3	7.45	8.58	9.46	10.21	10.87	11.45	11.98	12.47	12.92	14.82	16.33	17.6	18.71	19.7	20.60	21.43	22.20
4	5.60	6.25	6.76	7.17	7.53	7.84	8.12	8.38	8.61	9.57	10.31	10.92	11.44	11.9	12.31	12.69	13.03
5	4.77	5.25	5.60	5.89	6.14	6.35	6.54	6.71	6.87	7.50	7.98	8.36	8.69	8.98	9.24	9.47	9.68
6	4.32	4.70	4.98	5.21	5.40	5.56	5.71	5.84	5.96	6.43	6.79	7.07	7.31	7.52	7.71	7.87	8.02
7	4.03	4.36	4.59	4.79	4.94	5.08	5.20	5.31	5.41	5.80	6.08	6.31	6.50	6.67	6.81	6.94	7.06
8	3.83	4.12	4.33	4.50	4.64	4.76	4.86	4.96	5.04	5.37	5.62	5.81	5.97	6.11	6.23	6.34	6.44
9	3.69	3.95	4.15	4.30	4.42	4.53	4.62	4.71	4.78	5.08	5.29	5.46	5.60	5.72	5.83	5.92	6.01
10	3.58	3.83	4.00	4.14	4.26	4.36	4.44	4.52	4.59	4.85	5.05	5.20	5.33	5.44	5.53	5.62	5.69
15	3.29	3.48	3.62	3.73	3.82	3.90	3.96	4.02	4.07	4.27	4.42	4.53	4.62	4.70	4.77	4.83	4.88
20	3.15	3.33	3.46	3.55	3.63	3.70	3.75	3.80	3.85	4.02	4.15	4.24	4.32	4.39	4.44	4.49	4.54
25	3.08	3.24	3.36	3.45	3.52	3.58	3.64	3.68	3.73	3.88	4.00	4.08	4.15	4.21	4.27	4.31	4.35
30	3.03	3.19	3.30	3.39	3.45	3.51	3.56	3.61	3.65	3.80	3.90	3.98	4.05	4.11	4.15	4.20	4.23
40	2.97	3.12	3.23	3.31	3.37	3.43	3.47	3.51	3.55	3.69	3.79	3.86	3.92	3.98	4.02	4.06	4.09
50	2.94	3.08	3.18	3.26	3.32	3.38	3.42	3.46	3.50	3.63	3.72	3.79	3.85	3.90	3.94	3.98	4.01
60	2.91	3.06	3.16	3.23	3.29	3.34	3.39	3.43	3.46	3.59	3.68	3.75	3.81	3.85	3.89	3.93	3.96
100	2.87	3.01	3.10	3.17	3.23	3.28	3.32	3.36	3.39	3.51	3.60	3.66	3.72	3.76	3.80	3.83	3.86
∞	2.81	2.94	3.02	3.09	3.14	3.19	3.23	3.26	3.29	3.40	3.48	3.54	3.59	3.63	3.66	3.69	3.72

Literaturverzeichnis

Fahrmeir, L., Hamerle, A. und Tutz, G. (Hrsg.) (1996). *Multivariate Statistische Methoden*, 2. Aufl., de Gruyter, Berlin, New York.

Fahrmeir, L., Kneib, T. und Lang, S. (2009). *Regression*, 2. Aufl., Springer, Berlin, Heidelberg, New York.

Fahrmeir, L., Künstler, R., Pigeot, I. und Tutz, G. (2007). *Statistik*, 6. Aufl., Springer, Berlin, Heidelberg, New York.

Golub, G. und Van Loan, C. (1996). *Matrix Computations*, 3. Aufl., The Johns Hopkins University Press, Baltimore.

Handl, A. und Niermann, S. (2010). *Multivariate Analysemethoden: Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS*, Springer.

Hoaglin, D. und Welsch, R. (1978). The Hat Matrix in Regression and ANOVA, *The American Statistician* **32**(1): 17–22.

Magnus, J. R. und Neudecker, H. (1999). *Matrix Differential Calculus*, 2. Aufl., Wiley, New York.

Mardia, K., Kent, J. und Bibby, J. (1979). *Multivariate Analysis*, Academic Press, London.

Miller, R. (1981). *Simultaneous Statistical Inference*, Springer, New York.

Nagl, W. (1992). *Statistische Datenanalyse mit SAS*, Campus, Frankfurt/New York.

Singer, H. (2010). Maximum Entropy Inference for Mixed Continuous-Discrete Variables, *International Journal of Intelligent Systems* **25, 4**: 287 – 385.

Index