Skriptum für die Vorlesungen
Statistik I und II im Studienjahr 1996/97

Gerhard Arminger
und Mitarbeiter

Mit einem Beitrag von Prof. Dr. Ronge:
Die deutsche Universität — für Kulturfremde skizziert

© 1996 bei den Verfassern, Überarbeitete Fassung

Vorwort

Studiert man Sozial- oder Wirtschaftswissenschaften, stellt man sehr bald mit Erschrecken fest, daß das Fach Mathematik, dem man sich endgültig nach der Schule entronnen glaubte, in Mathematik- und Statistik-Lehrveranstaltungen der Hochschulen wieder auftaucht. Wenn man auch zunächst geneigt ist, dies der Böswilligkeit von Studienplanern und Hochschullehrern anzulasten, so überzeugt man sich durch einen Blick auf andere Fakultäten, daß neben den klassischen Naturwissenschaften auch andere Fächer, die von der Biologie und Medizin bis zu Geographie und Geschichtsforschung reichen, zunehmend von dieser Mathematisierung betroffen sind. Unter allen mathematischen Disziplinen, die in Substanzwissenschaften praktisch angewendet werden, zeichnen sich die Wahrscheinlichkeitstheorie und Statistik wohl durch die größte Verbreitung und Anwendungshäufigkeit aus.

Was hat das für einen Grund? Man erkannte, daß die einfachen Wenn-Dann-Beziehungen, aus denen wissenschaftliche Erklärungen bestehen, meistens unzulässige Vereinfachungen darstellen: Beim Beobachten und Experimentieren erleben wir immer wieder, daß Vorgänge, die unter scheinbar gleichen Bedingungen ablaufen, zu verschiedenen, aber ähnlichen Resultaten führen. Der Schritt vom Wenn zum Dann ist mit Ungewißheit belastet; der Zufall verdeckt die Struktur von Wirkungszusammenhängen. An die Stelle von Wenn-Dann-Aussagen treten Aussagen über Wahrscheinlichkeiten von Ereignissen.

Untersuchungsgegenstand der Statistik sind Vorgänge, deren Resultate nicht mit Sicherheit vorhersehbar sind und die man daher als *Zufallsexperimente* bezeichnet. In diesem Sinne ist jede Messung, deren Resultate streuen, z.B. die Ausbildung der individuellen Körpergröße oder das Steueraufkommen einer Region ein Zufallsexperiment. Bemerkenswert ist nun aber, daß die Ergebnisse solcher Zufallsexperimente nicht regellos (chaotisch) anfallen. Sie lassen vielmehr Gesetzmäßigkeiten erkennen, die freilich nicht als einfache Wenn-Dann-Aussagen darstellbar sind: Niemand weiß beispielsweise das Datum seines Todes. Eine Generation stirbt aber im Verlauf eines Jahrhunderts in ganz gesetzmäßiger Weise ab. Die Menschen sind verschieden groß, ihre Körpergrößen sind aber nicht regellos verteilt. Wir wissen, daß Zwerge und Riesen nicht häufiger sind als Mittelwüchsige. Extreme Resultate des Wachstumsvorganges sind seltener als Durchschnittsresultate. Die Gesetzmäßigkeiten *zufälliger Ereignisse* geben dem Unvorhersehbaren einen Rahmen, machen Unsicherheit kalkulierbar. Durch geeignete Maßnahmen kann man Unsicherheit verringern. Das Fachgebiet der Statistik umfaßt einen Großteil der dazu verwendeten Methoden.

Dieses Skriptum ist als Hilfsmittel zum leichteren Studium gedacht. Es ersetzt *nicht* den Besuch der Vorlesung und die regelmäßige Vorbereitung auf die Übungen, indem man *selbst* die gestellten Übungsaufgaben durchrechnet. Schriftliches Üben ist die wichtigste Voraussetzung für das Erlernen statistischer Methoden wie auch anderer Wissenschaften. Dies wurde bereits von Christian Fürchtegott Gellert (1715 - 1769) erkannt. Er schreibt in seiner Vorlesung: *Von den Fehlern der Studierenden bei der Erlernung der Wissenschaften, insbesonderheit der Akademien:* "Ja, meine Herren, daß wir unsere Kraft zu denken und unsere Gedanken ausdrücken, so wenig durch schriftliche Versuche stärken, dieses ist der letzte Fehler, den ich noch berühren will; ein unvergeblicher Fehler!". Die jetzige Studentengeneration ist nicht die erste, die mit Statistik zu kämpfen hat. Im Lehrplan des Vereinigten Friedrichswerdener und Friedrichstädter Gymnasiums Berlin, für die Prima im Jahre 1795/1796 findet man: Mittwoch: 10 - 11 Uhr: Geographie und Statistik. Zum Schluß sei Georg Christoph Lichtenberg mit der Zeitlosen Klage eines Mathematikprofesors zitiert: "Es ist unglaublich, wie unwissend die studierende Jugend auf Universitäten kommt. Wenn ich nur zehn Minuten rechne oder geometrisiere, so schläft ein viertel derselben sanft ein".

Inhaltsverzeichnis

1 Deskriptive Statistik **1**

1.1 Grundbegriffe 1

1.2 Empirische Verteilungsfunktion 6

1.3 Deskriptive Lagemaße 7

1.4 Streuungsmaße 9

2 Grundbegriffe der Wahrscheinlichkeitsrechnung **12**

2.1 Zufällige Ereignisse 12

2.2 Wahrscheinlichkeit 14

2.3 Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit 16

2.4 Zufallsvariable 17

3 Diskrete Verteilungen **19**

3.1 Grundlagen 19

3.2 Spezialfälle 23

4 Stetige Verteilungen **27**

4.1 Grundlagen 27

4.2 Spezialfälle 29

5 Mehrdimensionale Verteilungen **33**

5.1 Diskrete Verteilungen 33

5.2 Stetige Verteilungen 36

6 Grenzwertsätze **39**

6.1 Linearkombination von Zufallsvariablen 39

6.2 Stochastische Ungleichungen 40

6.3 Schwaches Gesetz der großen Zahlen 41

6.4 Zentraler Grenzwertsatz 42

7 Grundbegriffe der mathematischen Statistik **43**

7.1 Grundgesamtheit und Stichprobe 43

7.2 Stichprobenfunktionen 43

8 Punkt und Intervallschätzung **46**

8.1 Punktschätzung 46

8.2 Schätzverfahren . . . 47
8.3 Intervallschätzung . . . 50

9 Signifikanztests **53**
9.1 Aufbau von Signifikanztests . . . 53
9.2 Fehler erster Art und zweiter Art . . . 54
9.3 Signifikanztests für spezielle Fragestellungen . . . 57

10 Korrelation und Regression **62**
10.1 Einfache Korrelation . . . 62
10.2 Klassisches Regressionsmodell . . . 65
10.3 Einfache Zeitreihenanalyse . . . 73
10.4 Verallgemeinerungen des klassischen Regressionsmodells . . . 75
10.5 Varianz- und Kovarianzanalyse . . . 76
10.6 Multiple und partielle Korrelation . . . 79

11 Abhängigkeit zwischen qualitativen und ordinalen Merkmalen **80**
11.1 Assoziationsmaße für qualitative Merkmale . . . 80
11.2 Der χ^2 Test auf statistische Unabhängigkeit . . . 84
11.3 Assoziationsmaße und Tests für ordinale Merkmale . . . 85

12 Wirtschafts- und Sozialstatistik **87**
12.1 Datenbasis . . . 87
12.2 Bevölkerungsstatistik . . . 87
12.3 Erwerbsstatistik . . . 93
12.4 Indexrechnung . . . 95

Anhang **101**
Mengenlehre . . . 101
Das Summenzeichen . . . 102
Exponentialfunktion und Logarithmus . . . 103
Differential- und Integralrechnung . . . 104
Matrizenrechnung . . . 107

Literatur **115**

1 Deskriptive Statistik

1.1 Grundbegriffe

Bei statistischen Erhebungen (z.B. Volkszählung, Mikrozensus, Arbeitsstättenzählung, Einkommens- und Verbrauchsstichprobe) fallen in der Regel Tausende von Einzeldaten an. Diese im einzelnen unüberschaubare Datenmenge wird durch die Methoden der deskriptiven Statistik auf möglichst wenige, aber aussagefähige Zahlen reduziert. Wichtige Beispiele sind absolute und relative Häufigkeiten, empirische Verteilungsfunktionen, Mittelwerte und Indexzahlen. Die Darstellung der Daten durch Zahlen wird durch graphische Darstellungen und Tabellen unterstützt.

Grundlage aller statistischen Überlegungen sind die *statistischen Einheiten*, die als Träger *statistischer Merkmale* fungieren. Die für eine Untersuchung relevanten Einheiten faßt man zu einer *Grundgesamtheit* zusammen. Eine ausgewählte Teilmenge der Grundgesamtheit bezeichnet man als *Stichprobe*. Eine Grundgesamtheit ist nach sachlichen, zeitlichen, räumlichen und inhaltlichen Kriterien abzugrenzen. Von jedem beliebigen Objekt muß entschieden werden können, ob es zur Grundgesamtheit gehört oder nicht.

Beispiel: Die 'deutsche Bevölkerung' ist keine wohldefinierte Grundgesamtheit. Hingegen ist die Menge 'Einwohner der Bundesrepublik Deutschland am 1.1.1990 um 12 Uhr MEZ eine sachlich, räumlich und zeitlich genau abgegrenzte Menge von Individuen und kann daher als Grundgesamtheit dienen.

Statistische Mengen (Grundgesamtheiten oder Stichproben), die auf einen Zeitpunkt (z.B. auf einen Stichtag) bezogen sind, heißen *Bestandsmassen* (z.B. Einwohner der DDR am 1. Okt. 1990); Massen, die auf einen Zeitraum bezogen sind, heißen *Bewegungsmassen* (z.B. Geburten in der Bundesrepublik Deutschland vom 1.1.1989 bis zum 31.12.1989).

An jeder statistischen Einheit werden Merkmale oder Variable festgestellt, z.B. bei Personen das Alter, das Einkommen, der Beruf, das Geschlecht etc. Merkmale sind nur dann für statistische Zwecke brauchbar, wenn die Ausprägungen eines Merkmals zwei Eigenschaften aufweisen. Erstens, die Merkmalsausprägungen schließen einander aus. Zweitens, jeder statistischen Einheit kann eine Merkmalsausprägung zugeordnet werden.

Beispiel: Das Merkmal A Religionsbekenntnis mit den Ausprägungen $\{A_1$ = katholisch, A_2 = protestantisch$\}$ ist kein statistisches Merkmal, da es die zweite Bedingung nicht erfüllt. Hingegen ist das Merkmal B Religionsbekenntnis mit $\{B_1$ = katholisch, B_2 = protestantisch, B_3 = sonstige$\}$ als statistisches Merkmal zu verwenden.

Die Zuordnung von Merkmalsausprägungen zu statistischen Einheiten bezeichnet man als *Messung*. Die sorgfältige Durchführung einer Messung ist ein zentrales Problem jeder Wissenschaft. Die Messungen werden für jede statistische Einheit durchgeführt und auf einer *Urliste* notiert. Selbstverständlich werden derartige Urlisten heute als *Verbunde* oder *Sätze* auf einem elektronischen Datenträger abgelegt und verwahrt. Wir geben ein Beispiel für eine Urliste an. Es bezieht sich auf eine Stichprobe der Wohnbevölkerung in der Bundesrepublik Deutschland zum 1.1.1990.

Nummer	Geschlecht	Beruf	Alter	Kinderzahl
1	weiblich	Kauffrau	42	2
2	weiblich	Studentin	23	0
3	männlich	Schlosser	33	unbekannt
4	männlich	Beamter	59	4
⋮	⋮	⋮	⋮	⋮
n	weiblich	unbekannt	29	0

An diesem Beispiel ist zu erkennen, daß der Begriff der Messung in der Statistik allgemeiner ist als in der Umgangssprache, in der als Messung die Zuordnung einer Zahl zu einer Untersuchungseinheit verstanden wird. In der Statistik wird ausgehend von der Urliste ebenfalls jeder Person für jede Variable eine Zahl zugeordnet. Diese Zahlen werden

jedoch zunächst nur als Kodierungen verwendet und haben nur für bestimmte Variable eine numerische Bedeutung. Die Kodierung von Merkmalsausprägungen muß in einem Kodierungsschlüssel festgelegt werden. Bei dieser Kodierung können auch Zusammenfassungen der Merkmalsausprägungen erfolgen. Für das Beispiel verwenden wir folgende Kodierung:

Variable	Merkmalsausprägung	Kodierung
Geschlecht	männlich weiblich unbekannt	1 2 -9999
Beruf	Arbeiter Angestellter Sonstige unbekannt	1 2 3 -9999
Alter	x = Altersangabe in Jahren unbekannt	x -9999
Kinderzahl	x = Kinderzahl unbekannt	x -9999

Führt man diese Kodierung durch, erhält man die sogenannte Datenmatrix, bei der alle Messungen mit Zahlen angegeben sind und die fehlenden Werte mit -9999 gekennzeichnet sind.

Nummer	Geschlecht	Beruf	Alter	Kinderzahl
1	2	2	42	2
2	2	3	23	0
3	1	1	33	-9999
4	1	3	59	4
⋮	⋮	⋮	⋮	⋮
n	2	-9999	29	0

Skalenniveau

Für die Verwendung statistischer Maßzahlen, die im nächsten Abschnitt diskutiert wird, ist die Unterscheidung von Merkmalen nach ihrem Meß– oder Skalenniveau von großer Bedeutung.

- *Nominales Meßniveau:* Wenn sich die Merkmalsausprägungen eines statistischen Merkmals für eine statistische Untersuchung beliebig umordnen lassen, liegt ein nominal oder qualitativ skaliertes Merkmal vor. Typische Beispiele sind Geschlecht oder Beruf. Die Zahlen, die Merkmalsausprägungen dieser Variablen zugeordnet werden, haben keine inhaltliche Bedeutung, Rechnungen wie Addition und Subtraktion oder Vergleiche durch Ordnungsrelationen sind inhaltlich bedeutungslos.

- *Ordinales Meßniveau:* Wenn sich die Merkmalsausprägungen eines statistischen Merkmals nach einem Kriterium ordnen lassen, die Abstände zwischen den Merkmalsausprägungen aber nicht bekannt sind, so liegt ein ordinales Merkmal vor. Die Zahlen, die als Kodierungen diesen Merkmalsausprägungen zugeordnet sind, müssen zwar die Rangfolge der Merkmalsausprägungen wiedergeben, die Abstände zwischen den Kodierungen können aber beliebig gewählt werden. Typische Beispiele sind Schulnoten mit den Merkmalsausprägungen 'sehr gut', 'gut', 'befriedigend', 'ausreichend' und 'mangelhaft' oder Befragungen in den Sozial- und Wirtschaftswissenschaften, in denen Skalen mit Ausprägungen der Form 'stimme zu', 'teils teils', 'lehne ab', den befragten Personen vorgelegt werden. Wiederum sind Rechnungen wie Addition und Subtraktion bedeutungslos. Am besten macht man sich dieses Faktum bei den Schulnoten klar, für deren Merkmalsausprägungen die Kodierung $\{1, 2, 3, 4, 5\}$ genauso zulässig ist wie die Kodierung $\{0, \pi, 5, 99.9, 1000\}$. Die Berechnung eines Mittelwertes führt jedoch zu völlig unterschiedlichen Ergebnissen. Aus dieser Überlegung folgt, daß die Durchschnittsnoten, die für die Zuweisung von Studienplätzen berechnet werden, aus der Sicht des Statistikers unsinnig sind.

- *Rangskala:* Eng verwandt mit dem ordinalen Meßniveau sind die Rangskalen. Lassen sich die Elemente $1, 2, \ldots, n$ einer Grundgesamtheit oder Stichprobe nach einem Kriterium ordnen, erhält man die Ränge $r(1) \leq r(2) \leq \ldots \leq r(n)$. Eine Ordinalskala läßt sich daher als Rangskala mit Bindungen auffassen, in der einige Elemente die gleichen Ränge aufweisen können. Rangskalen mit keinen oder nur wenigen Bindungen treten in den Sozial- und Wirtschaftswissenschaften kaum auf.

- *Quantitatives Meßniveau:* Wenn sich die Merkmalsausprägungen eines statistischen Merkmals sowohl ordnen lassen als auch die Abstände zwischen den Merkmalsausprägungen angeben lassen, spricht man von quantitativen oder metrischen Merkmalen. Weisen sie darüber hinaus einen natürlichen Nullpunkt auf, liegt eine Ratio– oder Verhältnisskala vor. Die Kodierung dieser Merkmale muß sowohl die Ordnung als auch die Abstände zwischen den Merkmalen wiedergeben. Die Bildung von Summen und Differenzen ist inhaltlich bedeutungsvoll. Typische Beispiele metrisch skalierter Merkmale sind Häufigkeiten (z.B. Kinderzahl in einer Familie, Zahl der Autounfälle an einer Kreuzung) oder Variable mit beliebig feiner Einteilung der Merkmalsausprägungen wie Alter, Größe und Gewicht. Eine wichtige Unterscheidung ist die Unterteilung der metrischen Merkmale in *stetige Merkmale* (z.B. Alter, Größe, Gewicht), in denen die Merkmalsausprägungen beliebige Werte der reellen Zahlenachse annehmen können, und *diskrete Merkmale*, in denen nur bestimmte Merkmalsausprägungen in $\mathcal{R}$ (reelle Zahlen) angenommen werden können. Beispiele sind die oben genannten Häufigkeiten. Ist ein diskretes Merkmal sehr fein unterteilt (z. B. Geldbeträge in Pfennigen), wird das diskrete Merkmal wie ein stetiges Merkmal behandelt und daher als quasistetig bezeichnet.

Absolute und relative Häufigkeiten

Zur Analyse der Daten einer statistischen Erhebung faßt man zunächst für jedes einzelne Merkmal die Daten zusammen, indem aus jeweils einer Spalte der Datenmatrix die *absoluten Häufigkeiten* jeder Merkmalsausprägung berechnet werden. Früher erfolgte diese Berechnung durch Strichlisten, heute wird die elektronische Datenverarbeitung eingesetzt.

Beispiel: Für die qualitative Variable A Religionsbekenntnis mit den Ausprägungen und Kodierungen $\{A_1$ = römisch katholisch = $1, A_2$ = protestantisch = $2, A_3$ = konfessionslos = $3, A_4$ = sonstiges Bekenntnis = $4, A_5$ = unbekannt = $-9999\}$ liege folgende Kodierung der Urliste vor:

2 3 1 1 3 2 1 4 −9999 −9999 3 1 1 1 4 3 2 2 1 1 −9999 1 2 4 3 2

Die Berechnung der absoluten Häufigkeiten h_m für jede Merkmalsausprägung A_m ergibt die Tabelle:

Häufigkeitstabelle zum Religionsbekenntnis				
Merkmalsausprägung	Symbol	Kodierung	absolute Häufigkeit (h_m)	relative Häufigkeit ($\tilde{p}_m$)
römisch katholisch	A_1	1	9	0.346
protestantisch	A_2	2	6	0.231
konfessionslos	A_3	3	5	0.192
sonstiges Bekenntnis	A_4	4	3	0.115
unbekannt	A_5	-9999	3	0.115
Summe	$\sum$		26	1.000

Neben den absoluten Häufigkeiten $h_m, m = 1, \ldots, 5$ stehen die relativen Häufigkeiten $\tilde{p}_m$, die aus den h_m für alle K Merkmalsausprägungen berechnet werden:

$$\tilde{p}_m = \frac{h_m}{\sum_{m=1}^{M} h_m}$$

Die Gesamtzahl der Elemente wird mit N in der Grundgesamtheit und mit n in der Stichprobe bezeichnet, so daß gilt:

$$h_\cdot = \sum_{m=1}^{M} h_m = N \quad \text{in einer Grundgesamtheit}$$

$$h_\cdot = \sum_{m=1}^{M} h_m = n \quad \text{in einer Stichprobe}$$

Die Summe der relativen Häufigkeiten muß 1 ergeben. Die Bedeutung der relativen Häufigkeiten liegt in der Tatsache, daß mit ihnen Grundgesamtheiten oder Stichproben unterschiedlicher Größe verglichen werden können.

Die graphische Darstellung eines qualitativen oder ordinalen Merkmals oder eines metrischen Merkmals mit wenigen Ausprägungen erfolgt durch ein *Stabdiagramm* oder ein *Kreisdiagramm*. Im Stabdiagramm werden auf der Abszisse die Merkmalsausprägungen A_m und auf der Ordinate die relativen Häufigkeiten $\tilde{p}_m$ aufgetragen. Im Kreisdiagramm werden die Winkel α_m (in Grad) der Kreissektoren proportional zu den relativen Häufigkeiten $\tilde{p}_m$ gewählt:

$$\alpha_m = \tilde{p}_m \cdot 360$$

Beispiel: Auf die Frage nach ihrer Parteipräferenz gaben 50 Studenten Antworten, die sich in folgender Häufigkeitstabelle zusammenfassen lassen:

Tabelle zur Parteipräferenz				
Merkmalsausprägung	Symbol	h_i	$\tilde{p}_i$	α_i
CDU/CSU	A_1	21	0.42	151.5
SPD	A_2	19	0.38	136.8
FDP	A_3	4	0.08	28.8
GRÜNE	A_4	6	0.12	43.2

Im Unterschied zu qualitativen, ordinalen oder diskreten Merkmalen liegen bei stetigen oder quasistetigen metrischen Merkmalen häufig so viele verschiedene Merkmalsausprägungen vor, daß bei einer einfachen Häufigkeitsauszählung keine Reduktion der Datenfülle erreicht wird. In diesem Fall ist es erforderlich, die Merkmalsausprägungen zu Klassen zusammenzufassen, die ein neues metrisches Merkmal mit weniger Merkmalsausprägungen ergeben. Als Beispiel betrachten wir das Merkmal Brenndauer (in Stunden) an 200 Leuchtstoffröhren, für die z.B. folgende Werte auftreten:

127.53 144.27 443.17 99.40 207.89 ...

Zur Berechnung der absoluten und relativen Häufigkeiten werden für jede Klasse $k = 1, \ldots, K$ eine untere Klassengrenze a_k und eine obere Klassengrenze a_{k+1} bestimmt. Der realisierte Wert x_i der Stichprobe oder der Grundgesamtheit wird der Klasse k zugewiesen, wenn x_i ein Element des halboffenen Intervalls $(a_k, a_{k+1}]$ ist.

Tabelle: Brenndauer von Leuchtstoffröhren (in Stunden)					
untere Klassengrenze	obere Klassengrenze	Klassenmitte	h_k	$\tilde{p}_k$	$\tilde{f}_k$
0	100	50	27	0.135	0.135
100	200	150	49	0.245	0.245
200	300	250	37	0.185	0.185
300	400	350	28	0.140	0.140
400	1000	700	59	0.295	0.049
$\sum$			200	1.000	

Man beachte, daß die Abstände zwischen den Klassengrenzen nicht gleich sind. Sie betragen jeweils 100 in den ersten 4 Klassen und 600 in der fünften Klasse. Durch diese Klassenbildung läßt sich die Brenndauer als neues statistisches Merkmal mit 5 Ausprägungen auffassen, dessen Merkmalsausprägungen die Klassenmitten c_k sind, durch die das metrische Skalenniveau der Variablen Brenndauer beibehalten wird. (Die Klassenmitte wird als repräsentativer Wert einer Klasse gewählt.) Man beachte, daß das metrische Skalenniveau nur dann erhalten bleibt, wenn keine offenen Klassen (d.h. $a_1 = -\infty$ oder $a_K = +\infty$) verwendet werden. Offene Klassen sollten daher vermieden werden. Die Anzahl der Klassen sollte so gewählt werden, daß die Darstellung sowohl in Tabellen– als auch in graphischer Form übersichtlich bleibt. In Abhängigkeit von der Stichprobengröße sollen nach DIN 55302, Blatt 1 etwa 10 Klassen für $h_{\cdot} \sim 100$, 13 Klassen für $h_{\cdot} \sim 1000$ und 16 Klassen für $h_{\cdot} \sim 10.000$ gebildet werden.

Die graphische Darstellung erfolgt wiederum durch Stabdiagramme, bei denen die Klassenmitten c_k auf der Abszisse und die relativen Häufigkeiten $\tilde{p}_k$ auf der Ordinate für $k = 1, \ldots, K$ aufgetragen werden. Eine zweite Möglichkeit ist die Darstellung durch Histogramme. Das Histogramm besteht aus Rechtecken, die über den Intervallen $(a_k, a_{k+1}]$ errichtet werden. Die Fläche des Rechtecks entspricht der relativen Häufigkeit $\tilde{p}_k$. Da die Intervalle $(a_k, a_{k+1}]$ der Länge nach variieren können, müssen die Höhen $\tilde{f}_k$ der Rechtecke wie folgt berechnet werden:

$$\tilde{f}_k = \frac{\tilde{p}_k}{a_{k+1} - a_k} \qquad k = 1, \ldots, K$$

Die Rechteckshöhe ist nicht identisch mit der relativen Häufigkeit, da die Breite der Klasse berücksichtigt wird. In der letzten Tabelle ist die Dichte der einzelnen Klassen angegeben, so daß unmittelbar das Histogramm erstellt werden kann. Liegt eine offene Klasse vor, kann kein Histogramm gezeichnet werden.

1.2 Empirische Verteilungsfunktion

Ist ein Merkmal metrisch, so läßt sich aus den relativen Häufigkeiten die empirische Verteilungsfunktion berechnen. Das metrische Merkmal werde mit X bezeichnet, der Wert x ist ein beliebiger Wert aus $\mathcal{R}$. Die empirische Verteilungsfunktion $\tilde{F}_X(x)$ des Merkmals X an der Stelle x gibt an, wie groß die relative Häufigkeit ist, daß die Variable X einen Wert $\leq x$ annimmt. Die Verteilungsfunktion ist definiert durch

$$\tilde{F}_X(x) = \frac{1}{h} \sum_{m=1}^{M} h_m \cdot I_m(x)$$

Die Funktion $I_m(x)$ ist die Indikatorfunktion. Sie nimmt den Wert 1 an, wenn der zu h_m zugehörige Wert x_i (Merkmalsausprägung) der Variablen $X \leq x$ ist und nimmt den Wert 0 an, wenn $x_i > x$ ist.

Wenn alle Meßwerte x_i unterschiedlich sind, erhält man $h_m = 1$ und $M = h = n$. Daraus folgt:

$$\tilde{F}_X(x) = \frac{1}{M} \sum_{m=1}^{M} I_m(x)$$

Für die empirische Verteilungsfunktion der Brenndauer von Leuchtstoffröhren erhält man:

Tabelle: empirische Verteilungsfunktion der Brenndauer

untere Klassengrenze	obere Klassengrenze	Klassenmitte	$\tilde{p}_k$	$\tilde{F}_k(a_{k+1})$
0	100	50	0.135	0.135
100	200	150	0.245	0.380
200	300	250	0.185	0.565
300	400	350	0.140	0.705
400	1000	700	0.295	1.000

Man beachte, daß die empirische Verteilungsfunktion einer Klasse k immer an der oberen Klassengrenze a_{k+1} durch Summierung der relativen Häufigkeiten berechnet wird. Für die Werte der Klasse k, die unterhalb der oberen Klassengrenze a_{k+1} liegen, ist daher die Verteilungsfunktion zu groß. Für beliebige Werte x wird daher zwischen der

unteren und der oberen Klassengrenze (unter der Annahme, daß die Werte innerhalb einer Klasse gleich verteilt sind) linear interpoliert. Wenn x in der Klasse k liegt, gilt:

$$\tilde{F}_X(x) = \tilde{F}_X(a_k) + \frac{(\tilde{F}_X(a_{k+1}) - \tilde{F}_X(a_k))}{(a_{k+1} - a_k)}(x - a_k)$$

Man beachte, daß die empirische Verteilungsfunktion nur für metrische Variable definiert ist. Liegt eine ordinale Variable vor, so lassen sich zwar die K Merkmalsausprägungen $A_1 < A_2 < \ldots < A_K$ ordnen, die Lage dieser Merkmalsausprägungen auf $\mathcal{R}$ ist aber nicht bekannt. Trotzdem werden in manchen Anwendungen die relativen Häufigkeiten auch für ordinale Häufigkeiten wie im letzten Beispiel kumuliert. Diese kumulierte Funktion $H : \{A_1, \ldots, A_M\} \longrightarrow [0, 1]$ bezeichnet man als kumulierte relative Häufigkeit:

$$H(A_m) = \sum_{j=1}^{m} \tilde{p}_j \quad \text{für} \quad m = 1, \ldots, M$$

1.3 Deskriptive Lagemaße

Die gesamte statistische Information über ein Merkmal ist in den relativen Häufigkeiten enthalten. Dieser Informationsgehalt läßt sich jedoch häufig – ohne Informationsverlust – durch wenige Kennzahlen darstellen. Die wichtigsten dieser Kennzahlen sind Lage- und Streuungsmaße. Wir gehen zunächst auf die Lagemaße ein.

Der Modus

Der *Modus* oder *Modalwert* ist die häufigste Ausprägung einer Verteilung. Er wird mit M bezeichnet. Liegt eine metrische Variable in gruppierter Form vor, ist die häufigste Ausprägung die *Modalklasse*. Im Beispiel zur Parteipräferenz ist der Modus die Ausprägung CDU/CSU. Der Modus kann sowohl für qualitative als auch für ordinale als auch für rangskalierte als auch für metrische Variable verwendet werden.

Der Median

Als *Median* oder *Zentralwert* bezeichnet man den Wert $x_{0.5}$, für den gilt:

$$\tilde{F}_X(x_{0.5}) = 0.5$$

Der Median teilt die Grundgesamtheit oder Stichprobe in zwei gleiche Hälften. Die erste Hälfte besitzt Merkmalsausprägungen $\leq x_{0.5}$, die zweite Hälfte besitzt Merkmalsausprägungen $\geq x_{0.5}$. Zur Berechnung von $x_{0.5}$ werden die Werte $x_1, \ldots, x_{h.}$ einer metrischen Variablen zunächst geordnet, so daß gilt:

$$x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[i]} \leq \ldots \leq x_{[h.]}$$

Ist die Zahl $h.$ ungerade, so ist

$$x_{0.5} = x_{[(h.+1)/2]}$$

Ist die Zahl $h.$ gerade, so ist:

$$x_{0.5} = \frac{1}{2}\left(x_{[h./2]} + x_{[h./2+1]}\right)$$

Beispiel: Gegeben sei eine Stichprobe von monatlichen Einkommen (in DM) von Studenten:

898, 912, 719, 1032, 1716, 697, 981, 1413, 750, 637

Die geordnete Liste der Einkommen ist:

637, 697, 719, 750, 898, 912, 981, 1032, 1413, 1716

Die Stichprobengröße $h_{\cdot}$ ist 10. Der Median ist daher:

$$x_{0.5} = (x_{[5]} + x_{[6]})/2 = (898 + 912)/2 = 905$$

Liegen die Daten nur in klassifizierter Form wie im Beispiel über die Brenndauer von Leuchtstoffröhren vor, so muß zunächst die Klasse k bestimmt werden, in der der Median liegt. Diese Klasse heißt Medianklasse. Für sie gilt:

$$k \text{ ist Medianklasse} \Longleftrightarrow \tilde{F}_X(a_k) < 0.5 \leq \tilde{F}_X(a_{k+1})$$

Hat man die Medianklasse k ermittelt, kann unter der Annahme der Gleichverteilung der Werte innerhalb dieser Klasse der Median linear interpoliert werden:

$$x_{0.5} = a_k + \frac{(a_{k+1} - a_k)}{(\tilde{F}_X(a_{k+1}) - \tilde{F}_X(a_k))}(0.5 - \tilde{F}_X(a_k))$$

Beispiel: Aus der Häufigkeitstabelle der Brenndauer von Leuchtstoffröhren erhält man als Medianklasse $k = 3$, da gilt

$$F(a_3 = 200) = 0.380 < 0.5 \leq 0.565 = F(a_4 = 300)$$

Daraus läßt sich $x_{0.5}$ berechnen:

$$x_{0.5} = 200 + \frac{300 - 200}{0.565 - 0.380}(0.500 - 0.380) = 264.864$$

Liegt eine ordinale Skala vor, so läßt sich nur eine Ausprägung bestimmen, für die gilt:

$$H(A_{k-1}) < 0.5 \leq H(A_k)$$

Diese Ausprägung A_k kann als Medianausprägung oder kurz als Median des ordinalen Merkmals A bezeichnet werden.

Quantile

Der Begriff des Medians läßt sich auf den Begriff des α *Quantils* verallgemeinern. Gibt man einen Wert $\alpha \in [0, 1]$ vor, so läßt sich für ein metrisches Merkmal X der Wert x_α bestimmen, für den gilt:

$$\tilde{F}_X(x_\alpha) = \alpha$$

Wichtige Spezialfälle sind die Quartile $\{x_{0.25}, x_{0.50}, x_{0.75}\}$ und die Dezile $\{x_{0.1}, x_{0.2}, \ldots, x_{0.9}\}$. Die Quantile werden wie der Median durch Auszählen bei Vorliegen einer geordneten Liste $\{x_{[1]}, \ldots, x_{[h_\cdot]}\}$ oder durch lineare Interpolation bestimmt.

Beispiel: Das Unternehmen, das die im letzten Beispiel untersuchten Leuchtstoffröhren herstellt, möchte die Garantiezeit für die Brenndauer der Leuchtstoffröhren so festsetzen, daß maximal 15% der Röhren ersetzt werden müssen. Dieser Wert ist das 0.15 Quantil der Verteilung. Die Klasse, in der dieses Quantil liegt, ist $k = 2$, da gilt:

$$F(100) = 0.135 < 0.150 \leq 0.380 = F(200)$$

Das 0.15 Quantil wird durch lineare Interpolation ermittelt:

$$x_{0.15} = 100 + \frac{200 - 100}{0.380 - 0.135}(0.15 - 0.135) = 106.123$$

Das Unternehmen kann daher als Garantiedauer einen Wert von 106.1 Stunden festsetzen.

Das arithmetische Mittel

Das bekannteste Lagemaß für eine metrische Variable X ist das *arithmetische Mittel* $\bar{x}$. Die Beobachtungen $\{x_1, \ldots, x_n\}$ werden gemittelt, d.h.:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{(ungewichtetes Mittel)}$$

Treten bestimmte Merkmalsausprägungen häufiger als einmal auf, so läßt sich die Berechnung vereinfachen, indem die Merkmalsausprägungen x_m mit h_m multipliziert werden:

$$\bar{x} = \frac{1}{h} \sum_{m=1}^{M} x_m \cdot h_m = \sum_{m=1}^{M} x_m \cdot \tilde{p}_m \quad \text{(gewichtetes Mittel)}$$

Auf die letzte Gleichung muß immer dann zurückgegriffen werden, wenn die Daten nur in klassifizierter Form vorliegen. Dann sind die Werte x_n die Klassenmitten.

Beispiel: Die mittlere Brenndauer der Leuchtstoffröhren läßt sich als gewichtetes arithmetisches Mittel berechnen:

$$\bar{x} = 50 \cdot 0.135 + 150 \cdot 0.245 + 250 \cdot 0.185 + 350 \cdot 0.140 + 700 \cdot 0.295 = 345.25$$

Das arithmetische Mittel kann nicht für ordinale und qualitative Merkmale berechnet werden, da für diese die Addition nicht definiert ist.

Sowohl $x_{0.5}$ als auch $\bar{x}$ charakterisieren die Lage der Verteilung von X. Häufig sind $x_{0.5}$ und $\bar{x}$ die Werte, um die sich die meisten Werte der Verteilung anordnen. Das arithmetische Mittel ist zwar das gebräuchlichste Lagemaß; es empfiehlt sich aber, immer auch den Median zu berechnen, da dieser unempfindlicher gegenüber Ausreißern als das arithmetische Mittel ist.

1.4 Streuungsmaße

Zusätzlich zur Lage der Verteilung ist man an der Streuung der Verteilung interessiert. Die Streuung besagt, ob sich die Werte $x_i, i = 1, \ldots, n$ einer metrischen Variablen X eng um einen Wert gruppieren, oder ob sie weit von diesem Wert entfernt liegen.

Die Spannweite

Das einfachste Streuungsmaß ist die Spannweite R. Liegen die Daten als geordnete Liste $\{x_{[1]}, \ldots, x_{[n]}\}$ vor, so ist:

$$R = x_{[n]} - x_{[1]}$$

Die Spannweite hat den Nachteil, daß nur zwei extreme Werte zur Berechnung der Streuung verwendet werden, so daß nur ein kleiner Teil der Information der Daten ausgenützt wird.

Varianz– und Standardabweichung

Um alle Werte in die Berechnung der Streuung einzubeziehen, liegt es nahe, die Summe aller Differenzen zwischen $x_i, i = 1, \ldots, n$ und $\bar{x}$ zu bilden.

Es gilt jedoch:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \bar{x} \cdot n = 0$$

Dieser Nachteil läßt durch Verwendung von Absolutbeträgen oder Quadraten vermeiden. Als empirische *Varianz* s^2 wird die durchschnittliche Abweichung vom Mittelwert verwendet, bei der große Abweichungen überproportional gewichtet werden. Liegt eine Grundgesamtheit vor, so gilt:

$$s^2 = \frac{1}{N}\sum_{j=1}^{N}(x_j - \bar{x})^2$$

Für eine Stichprobe gilt:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Die unterschiedlichen Definitionen werden in der statistischen Methodenlehre begründet.

Sind die Daten gruppiert oder klassifiziert, müssen die obigen Formeln durch Gewichtung modifiziert werden:

$$s^2 = \sum_{k=1}^{K}(x_k - \bar{x})^2 \tilde{p}_k \quad \text{(für eine Grundgesamtheit)}$$

$$s^2 = \frac{n}{n-1}\sum_{k=1}^{K}(x_k - \bar{x})^2 \tilde{p}_k \quad \text{(für eine Stichprobe)}$$

Die Varianz ist immer positiv. Um auf die ursprüngliche Maßeinheit zu kommen, verwendet man die Standardabweichung:

$$s = \sqrt{s^2}$$

Beispiel: Die Varianz und die Standardabweichung der Brenndauer von Leuchtstoffröhren sind aus einer Stichprobe von $n = 200$ aus der angegebenen Häufigkeitstabelle zu berechnen.

Tabelle: Berechnung der Varianz				
x_k	$\tilde{p}_k$	$x_k\tilde{p}_k$	$(x_k-\bar{x})^2$	$(x_k-\bar{x})^2\tilde{p}_k$
50	0.135	6.75	87172.563	11768.296
150	0.245	36.75	38122.563	9340.028
250	0.185	46.25	9072.563	1678.424
350	0.140	49.00	22.563	3.159
700	0.295	206.50	125847.563	37125.031
$\sum$	1.000	$\bar{x}=345.25$		59914.937

Die empirische Varianz der Stichprobe beträgt:

$$s^2 = \frac{200}{199} \cdot 59914.937 = 60216.018$$

Die Standardabweichung ist dann

$$s = 245.389$$

In diesem Beispiel findet man daher eine starke Streuung der Brenndauern um den Mittelwert.

Man beachte, daß s^2 und s genauso wie $\bar{x}$ nur für metrische Variablen definiert sind.

Der Variationskoeffizient

Zum Zweck des Vergleichs von Streuungen aus verschiedenen Grundgesamtheiten oder Stichproben benötigt man ein dimensionsloses Maß der Streuung. Ein solches Maß ist der Variationskoeffizient für positive metrische Variable X.

$$v = \frac{S}{\bar{x}} \quad \text{(für } \bar{x} \text{ und } s \text{ aus der Grundgesamtheit oder aus der Stichprobe)}$$

Beispiel: Mittelwert und Standardabweichung des Brotpreises für ein kg Brot in der Bundesrepublik Deutschland betragen $\bar{x} = 3.25$ DM, $s^2 = 1.96\text{DM}^2$, in Österreich $\bar{x} = 19$ öS, $s^2 = 66.98\text{öS}^2$. In welchem Land streut der Brotpreis stärker? In beiden Ländern streut, gemessen am Durchschnitt, der Brotpreis ungefähr gleich, da gilt:

$$v_D = \frac{1.4}{3.25} \sim \frac{8.18}{19} = v_A$$

Der Quartilsabstand

Als Alternative zur Standardabweichung läßt sich auch der *Quartilsabstand* q verwenden:

$$q = (x_{0.75} - x_{0.25})/2$$

Das Maß q gibt die halbe Länge eines Intervalls an, auf dem die mittleren 50% der Verteilung liegen. Als Übung berechne man q für das Beispiel mit der Brenndauer der Leuchtstoffröhren.

2 Grundbegriffe der Wahrscheinlichkeitsrechnung

Die Wahrscheinlichkeitsrechnung ist ein Teilgebiet der Mathematik, das sich mit der Untersuchung der Gesetzmäßigkeiten von Ereignissen befaßt, deren Eintreffen vom Zufall abhängt. Zu den Grundbegriffen der Wahrscheinlichkeitsrechnung zählen Zufallsexperimente, Ereignisse und Wahrscheinlichkeit.

2.1 Zufällige Ereignisse

Untersuchungsgegenstand der Wahrscheinlichkeitsrechnung sind *Zufallsexperimente*. Unter einem Zufallsexperiment versteht man ein Experiment, das beliebig oft unter identischen Bedingungen wiederholt werden kann. Dieses Experiment hat eine bestimmte Anzahl unterschiedlicher Ergebnisse, die zufallsbedingt sind, d.h. im voraus nicht eindeutig bestimmt werden können.

Beispiele:

- Bei einem Würfelwurf ist nicht vorhersehbar, wieviele Augen $\{1, 2, \ldots, 6\}$ die Kopfseite des Würfels zeigen wird. Das Ergebnis eines Würfelwurfes – die oben liegende Augenzahl – bezeichnet man als zufälliges Ereignis.
- Besteht das Zufallsexperiment aus der Ziehung einer Karte aus einem Skat-Kartenspiel (32 Karten), so kann das zufällige Ereignis durch ein zweidimensionales Merkmal beschrieben werden, nämlich durch die Farbe $\{$Kreuz, Pik, Herz, Karo$\}$ und durch das Bild $\{7, 8, \ldots,$ König, As$\}$.
- In einem Materiallager bilden die Abgänge der verschiedenen Produkte pro Monat ein Zufallsexperiment. Dieses Experiment wiederholt sich monatlich. Das zufällige Ereignis (Versuchsausgang) wird mehrdimensional durch die Höhe der Abgänge der einzelnen Produkte beschrieben.

Allerdings ist anzumerken, daß gerade in den Sozial- und Wirtschaftswissenschaften auch nicht reproduzierbare Phänomene als Zufallsexperimente modelliert werden. Über die inhaltliche Bedeutung des zufälligen Ereignisses kommen wir jetzt zur mathematischen Begriffsbestimmung. Gegeben sei ein Zufallsexperiment, dessen Ausgang durch ein ein- bzw. mehrdimensionales Merkmal beschrieben wird.

Definition:

1. Jeder Wert, den ein Merkmal annehmen kann, (jeder mögliche Ausgang eines Zufallsexperiments) heißt *Elementarereignis* (Symbol ω).
2. Die Menge aller Elementarereignisse heißt *Ereignisraum* (Merkmalsraum, Stichprobenraum; Symbol Ω).
3. Jede Teilmenge A des Ereignisraums Ω nennt man ein *Ereignis*. $A \subset \Omega$.
4. Man spricht vom *Eintreffen des Ereignisses* A, wenn das bei einem Versuch realisierte Elementarereignis ω ein Element aus der Menge A ist ($\omega \in A$).

Ereignisse werden gewöhnlich mit großen lateinischen Buchstaben gekennzeichnet, die häufig noch mit einem Index versehen sind ($A_1, A_2 \ldots$), um die Ereignisse durchzunumerieren. Elementarereignisse werden dagegen mit kleinen griechischen Buchstaben gekennzeichnet, die auch indiziert sein können. $\omega_i \in A$ heißt: das Elementarereignis ω_i ist Element des Ereignisses A.

Beispiel: Wir betrachten zunächst ein Experiment mit einem Würfel. Die 6 Elementarereignisse sind die Augenzahlen $i = 1, \ldots, 6$. Der Ereignisraum Ω ist $\{1, 2, \ldots, 6\}$. Wir definieren die Ereignisse:

A_i : Es werden i Augen gewürfelt. $A_i = \{i\}$

A : Es wird eine gerade Augenzahl gewürfelt. $A = \{2, 4, 6\}$

B : Es wird eine ungerade Augenzahl gewürfelt. $B = \{1, 3, 5\}$

A_3 tritt also dann ein, wenn eine 3 gewürfelt wird, und A tritt ein, wenn eine 2, 4 oder 6 gewürfelt wird.

Beispiel: Wird ein Experiment mit zwei Würfeln durchgeführt, so sind die 36 Elementarereignisse die Kombinationen der Augenzahlen der beiden Würfel. Zum Beispiel bedeutet $\omega = (2, 4)$, daß der erste Würfel 2 und der zweite Würfel 4 Augen zeigt. Weitere Beispiele für Ereignisse sind:

A : Mit dem ersten Würfel wurden 3 Augen gewürfelt, $A = \{(3, j) | j = 1, \ldots, 6\}$.

B : Mit beiden Würfeln werden gerade Augenzahlen geworfen,
$B = \{(i, j) | i = 2, 4 \text{ oder } 6, j = 2, 4 \text{ oder } 6\}$.

Beispiel: Bei der Untersuchung über die Lebensdauer einer Glühbirne besteht die Menge der möglichen Ausgänge aus allen nichtnegativen reellen Zahlen: $\Omega = \{x | x \geq 0\}$. A sei das Ereignis, daß eine Glühbirne mindestens 100 Stunden brennt: $A = \{x | x \geq 100\}$. Wenn die Birne nach 90 Stunden ausfällt, so ist A nicht eingetroffen.

Verknüpfungsoperationen zwischen Ereignissen

Im folgenden seien $A, B, A_i \subset \Omega, i \in \mathcal{N}$, ohne daß dies extra erwähnt wird.

Definition:

1. Das Ereignis $A \cup B$ tritt genau dann ein, wenn entweder A oder B eintritt oder beide Ereignisse A und B gleichzeitig eintreten. $A \cup B$ heißt Vereinigungsereignis von A und B.
2. Das Ereignis $A \cap B$ tritt genau dann ein, wenn sowohl A als auch B eintritt. $A \cap B$ heißt Durchschnittsereignis von A und B.
3. Das Ereignis $\cap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \ldots$ tritt ein, wenn alle A_i eintreten (wenn das realisierte Elementarereignis Element aller A_i ist, für alle $i \in \mathcal{N}$ gilt: $\omega \in A_i$).

Beispiel: In einem Würfelexperiment mit einem Würfel seien die Ereignisse A und C wie folgt definiert: A : Es wird eine gerade Augenzahl gewürfelt, C : Es wird eine 2 oder 3 gewürfelt. Dann ist $A \cap C = \{2\}$ und $A \cup C = \{2, 3, 4, 6\}$.

Definition:

1. Zwei Ereignisse A und B heißen *gleich* (in Zeichen $A = B$), wenn A genau dann realisiert wird, wenn B realisiert wird. Dies impliziert, daß A und B dieselben Elementarereignisse enthalten.
2. Tritt mit dem Ereignis A auch das Ereignis B ein, so zieht das Ereignis A das Ereignis B nach sich. In Zeichen $A \subset B$.

Das Ereignis U_1, daß mit einem Würfel eine Augenzahl größer als 6 geworfen wird, kann ebenso unmögliche eintreten wie das Ereignis U_2, daß eine Zahl zwischen 1 und 2 gewürfelt wird. Nach Definition sind diese Ereignisse gleich. Es gibt nur ein unmögliches Ereignis.

Definition:

1. Das Ereignis $\{\ \}$ (oder auch $\emptyset$) heißt *unmögliches Ereignis* (es tritt nie ein).
2. Ω heißt das *sichere Ereignis* (es tritt immer ein).
3. Das Ereignis, das genau dann eintritt, wenn A *nicht* eintritt, heißt das zu A *komplementäre Ereignis* A^c.
4. Zwei Ereignisse A und B *schließen einander aus* (sind *unverträglich* oder *disjunkt*), wenn ihr gemeinsames Auftreten unmöglich ist, also wenn $A \cap B = \emptyset$.

5. Die Ereignisse $A_1, \ldots, A_n$ heißen *Zerlegung* von Ω, wenn in einem Versuch genau eines dieser Ereignisse realisiert werden muß. $A_1, \ldots, A_n$ bilden eine Zerlegung, wenn gilt:
   a) $A_1 \cup \ldots \cup A_n = \Omega$
   b) Ungleiche Ereignisse schließen einander paarweise aus, d.h. $A_i \cap A_j = \emptyset$ für alle Paare $i \neq j$.

Beispiel: Wir betrachten wieder ein Würfelexperiment mit $A = \{1, 3, 5\}$ und $B = \{2, 4, 6\}$. Die Ereignisse A und B schließen einander aus, da es keine Augenzahl gibt, die sowohl gerade als auch ungerade ist. Da aber entweder eine gerade oder ungerade Augenzahl auftreten muß, gilt $A^c = B$ und $B^c = A$. Da zueinander komplementäre Ereignisse wegen $A \cup A^c = \Omega$ immer eine Zerlegung bilden, sind A und B eine Zerlegung von Ω. $A_1, \ldots, A_6$ bilden eine weitere Zerlegung.

Potenzmenge

Sind in einem Zufallsexperiment nur endlich viele Versuchsausgänge möglich, so besteht die Menge aller Ereignisse aus der *Potenzmenge* von Ω, d.h. aus der Menge aller möglichen Teilmengen von Ω (die leere Menge $\emptyset$ und die Menge Ω selbst sind auch Teilmengen von Ω). Besteht Ω aus n Elementarereignissen, so gibt es 2^n verschiedene Ereignisse (Teilmengen).

Beispiel: $\Omega = \{1, 2, 3\}$. Daher ist die Potenzmenge
$\mathcal{P}(\Omega) = \{\{\emptyset\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.

2.2 Wahrscheinlichkeit

Mit der Ausnahme des unmöglichen Ereignisses besteht für alle Ereignisse die Möglichkeit, in einem Zufallsexperiment aufzutreten. Die Ereignisse besitzen nun einen bestimmten Wahrscheinlichkeitsgrad der Realisierung, der durch eine Zahl zwischen 0 und 1 repräsentiert wird. So weist im fairen Würfelexperiment z.B. das Ereignis A (gerade Augenzahl) eine größere Wahrscheinlichkeit auf als das Ereignis C (Augenzahl 2 oder 3). Wenn man ein Zufallsexperiment mit einem Zehn-Pfennig-Stück durchführt, so unterstellt man für 'Zahl' und 'Wappen' dieselbe Wahrscheinlichkeit. Üblicherweise wird ein Wahrscheinlichkeitsmaß P so normiert, daß die Wahrscheinlichkeit P des sicheren Ereignisses gleich 1 ist. Daher wird in einem Zufallsexperiment mit einer Münze den Ausprägungen Zahl und Wappen die Wahrscheinlichkeit 0.5 zugewiesen. In der Umgangssprache ist der Wahrscheinlichkeitsbegriff subjektiv. Wenn man Student S zwei Wochen vor der Statistik-Klausur sagt, er werde wahrscheinlich die Klausur bestehen, so ist dies nur eine qualitative Aussage, da eine quantitative Aussage über die Wahrscheinlichkeit des Ereignisses {S besteht Statistik-Klausur} fehlt. Bevor wir zur mathematischen Definition der Wahrscheinlichkeit kommen, werden zwei Interpretationsmodelle der Wahrscheinlichkeit vorgestellt.

A-priori-Modelle

A-priori-Modelle beruhen auf dem Prinzip vom unzureichenden Grund (auch Indifferenzprinzip): Hat man keine Veranlassung, einen bestimmten Ausgang eines Zufallsexperiments für wahrscheinlicher als einen anderen zu halten, so wird man alle Ausgänge für gleichmöglich ansehen. Besteht der Ereignisraum Ω aus N, $N < \infty$, Elementarereignissen, so ist die Wahrscheinlichkeit, daß ein bestimmtes Elementarereignis realisiert wird, gleich $1/N$. Somit berechnet man die Wahrscheinlichkeit $P(A)$ eines Ereignisses A durch $P(A) = K/N$, wobei K die Anzahl der in A enthaltenen Elementarereignisse ist. Der Vorteil der A-priori-Modelle besteht darin, daß man die Wahrscheinlichkeit durch Abzählen der Elementarereignisse ausrechnen kann. Der Nachteil liegt darin, daß es in den Wirtschafts- und Sozialwissenschaften selten Situationen gibt, auf die man das Indifferenzprinzip anwenden kann.

Die Häufigkeitsinterpretation

Unter der (empirischen) Wahrscheinlichkeit eines Ereignisses versteht man die relative Häufigkeit dieses Ereignisses in einer (theoretisch unendlich) langen Versuchsreihe. Die relative Häufigkeit $p_n = \frac{m}{n}$ ist der Quotient aus der

Anzahl m des Eintreffens des Ereignisses in n Versuchen. Nach einer weiteren Durchführung des Experiments erhält man:

$$p_{n+1} = \begin{cases} \frac{m+1}{n+1} = \frac{n}{n+1}\, p_n + \frac{1}{n+1} & \text{wenn das Ereignis eingetreten ist,} \\ \frac{m}{n+1} = \frac{n}{n+1}\, p_n & \text{wenn es nicht eingetreten ist.} \end{cases}$$

Man erkennt, daß die Änderung der relativen Häufigkeit durch den Ausgang eines weiteren Experiments umso geringer ist, je größer n ist. Die Fluktuation der Folge $p_1, p_2, \ldots$ nimmt ab. Strebt die relative Häufigkeit eines Ereignisses mit wachsendem n zu einem Grenzwert, so bezeichnet man diesen Grenzwert als Wahrscheinlichkeit dieses Ereignisses.

Münzversuch: p_n('Zahl')$\rightarrow P$('Zahl'), wenn $n \rightarrow \infty$.

Diese Interpretation bietet den Vorteil, daß man die Wahrscheinlichkeit durch eine endliche Messung approximieren kann. Es wird kein A-priori-Modell benötigt. Die Genauigkeit der Messung läßt sich durch Versuchswiederholungen beliebig heraufsetzen. Die Wahrscheinlichkeit kann auf diese Weise allerdings nur bestimmt werden, wenn das Zufallsexperiment beliebig oft wiederholbar ist. Darin besteht der Nachteil dieses Ansatzes.

Zur formalen Behandlung wird der Wahrscheinlichkeitsbegriff axiomatisch (d.h. durch Festlegung) eingeführt. Die Axiomatisierung der Wahrscheinlichkeit wurde durch den russischen Mathematiker Kolmogoroff (1933) durchgeführt, der die Wahrscheinlichkeit indirekt durch Angabe gewünschter Eigenschaften und Relationen definiert hat.

Definition: (Axiome von Kolmogoroff) Eine Funktion $P(A)$, die jedem Ereignis $A \subset \Omega$ einen Wert $P(A)$ so zuordnet, daß die folgenden Bedingungen gelten, heißt ein *Wahrscheinlichkeitsmaß* auf der Gesamtheit der zu einem Zufallsexperiment gehörenden Ereignisse.

1. $0 \leq P(A) \leq 1$ für alle Ereignisse A
2. $P(\emptyset) = 0,\ P(\Omega) = 1$
3. Schließen A und B einander aus, so gilt:

   $$P(A \cup B) = P(A) + P(B)$$

4. Sind die abzählbar unendlich vielen Ereignisse $A_1, A_2, \ldots$ paarweise disjunkt, so gilt:

   $$P\left(\cup_{i=1}^{\infty} A_i\right) = P\left(A_1 \cup A_2 \cup \cdots\right) = \sum_{i=1}^{\infty} P(A_i)$$

Bemerkung: Gilt für das Ereignis $P(A) = 0$, so folgt daraus *nicht* , daß A das unmögliche Ereignis ist. Aus den Axiomen von Kolmogoroff lassen sich folgende Eigenschaften der Wahrscheinlichkeit herleiten:

Satz:

1. Für das komplementäre Ereignis von A gilt: $P(A^c) = 1 - P(A)$
2. Gilt für zwei Ereignisse A und B $A \subset B$, so ist $P(A) \leq P(B)$
3. Für zwei beliebige Ereignisse A und B gilt: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Additionssatz)
4. Bilden die Ereignisse $A_1, \ldots, A_n$ eine Zerlegung von Ω, so gilt:

   $$\sum_{i=1}^{n} P(A_i) = 1$$

Beweis: Übung.

Beispiel: Besitzt jede Augenzahl in einem Würfelexperiment die Wahrscheinlichkeit 1/6, und ist E das Ereignis, daß keine 6 gewürfelt wird, so ist $P(E) = 1 - P(E^c) = 1 - 1/6$; denn E^c tritt ein, wenn eine 6 gewürfelt wird.

2.3 Bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

Bisher sind nur Wahrscheinlichkeiten von Ereignissen ohne Berücksichtigung weiterer Bedingungen behandelt worden. Oft interessiert man sich aber für die Wahrscheinlichkeit eines Ereignisses A unter der zusätzlichen Voraussetzung, daß ein bestimmtes Ereignis B eintritt oder schon eingetreten ist. Man möchte z.B. wissen, mit welcher Wahrscheinlichkeit ein Fernseher noch 2 Jahre funktioniert, wenn man weiß, daß er bereits 5 Jahre störungsfrei gelaufen ist. Ein zweites Beispiel ist die Berechnung der Wahrscheinlichkeit, daß *von den* Frauen der Belegschaft eines Unternehmens mindestens eine befördert wird. Das Ereignis A ist dann das Ereignis mindestens eine Person wird befördert" und B ist das Ereignis "weiblich". Muß man also bei der Berechnung von $P(A)$ eine Bedingung berücksichtigen, die einen Einfluß auf die Wahrscheinlichkeit von A ausübt, so spricht man von einer bedingten Wahrscheinlichkeit.

Definition:

$$P(A|B) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{falls } P(B) > 0 \\ 0 & \text{falls } P(B) = 0 \end{cases}$$

heißt *bedingte Wahrscheinlichkeit von A, gegeben, daß B eintrifft oder eingetroffen ist.*

Beispiel: Wie groß ist die Wahrscheinlichkeit in einem Würfelexperiment, mit einem Würfel eine Augenzahl kleiner als 3 zu werfen, wenn bekannt ist, daß eine gerade Augenzahl gewürfelt worden ist? Es ist $P(A|B)$ zu berechnen, wobei $A = \{1, 2\}$ und $B = \{2, 4, 6\}$ ist. Wegen $P(A \cap B) = P(\{2\}) = 1/6$ und $P(B) = 1/2$ folgt $P(A|B) = 1/3$.

Beispiel: Im Würfelexperiment mit zwei Würfeln besitzt jede Kombination von Augenzahlen die Wahrscheinlichkeit $1/36$. Es sei A_1 das Ereignis, daß mit dem ersten Würfel eine 1, und B, daß mit dem zweiten Würfel eine ungerade Augenzahl gewürfelt wird, also $A_1 = \{(1, j)|j = 1, \ldots, 6\}$, $B = \{(i, j)|i = 1, \ldots, 6, j = 1, 3, 5\}$. Es ist $P(A_1) = 1/6$ und $P(B) = 1/2$. Wegen $A_1 \cap B = \{(1, 1), (1, 3), (1, 5)\}$ folgt:

$$P(A_1|B) = \frac{1/12}{1/2} = \frac{1}{6}.$$

Das Ereignis B hat also keinen Einfluß auf die Wahrscheinlichkeit von A_1, was auch erwartet wird, da die Ereignisse A_1 und B zwei verschiedene Würfel betreffen.

Definition: Man bezeichnet zwei Ereignisse A, B eines Zufallsexperiments als *stochastisch unabhängig*, wenn das Eintreten des einen die Eintrittswahrscheinlichkeit des anderen nicht beeinflußt:

$$P(A|B) = P(A), \text{falls } P(B) > 0.$$

Aus dieser Definition und der Definition der bedingten Wahrscheinlichkeit folgt der nächste Satz.

Satz: A und B sind genau dann stochastisch unabhängig, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

Beweis:

$$P(A|B) = P(A) \iff \frac{P(A \cap B)}{P(B)} = P(A) \iff P(A \cap B) = P(A) \cdot P(B)$$

Dieser Satz zeigt, daß bei stochastisch unabhängigen Ereignissen die Wahrscheinlichkeit des gemeinsamen Eintretens beider Ereignisse gleich dem Produkt der Einzelwahrscheinlichkeiten ist.

Aus der allgeminen Definition der bedingten Wahrscheinlichkeit kann die folgende *Multiplikationsregel* hergeleitet werden:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Aus der Multiplikationsregel lassen sich die Formel für die vollständige Wahrscheinlichkeit und die Formel von Bayes herleiten.

Satz von der vollständigen Wahrscheinlichkeit: Bilden die Ereignisse $E_1, \ldots E_n$ eine Zerlegung von Ω, so gilt für ein beliebiges Ereignis A:

$$P(A) = \sum_{i=1}^{n} P(A|E_i) \cdot P(E_i)$$

Beweis: Übungsaufgabe

Satz von Bayes: Bilden die Ereignisse $E_1, \ldots, E_n$ eine Zerlegung von Ω, so gilt für ein beliebiges Ereignis A mit $P(A) > 0$:

$$P(E_i|A) = \frac{P(A|E_i) \cdot P(E_i)}{\sum_{j=1}^{n} P(A|E_j) \cdot P(E_j)} \text{ für } i = 1, \ldots, n$$

Beweis: Nach der Multiplikationsregel ist $P(A|E_i) \cdot P(E_i) = P(A \cap E_i)$ und nach dem Satz der vollständigen Wahrscheinlichkeit ist $\sum_{j=1}^{n} P(A|E_j) \cdot P(E_j) = P(A)$.

$P(E_i)$ wird als a priori Wahrscheinlichkeit des Ereignisses E_i und $P(E_i|A)$ wird als a posteriori Wahrscheinlichkeit von E_i bezeichnet. Dieser Satz kann also dazu verwendet werden, ein unbekannte a posteriori Wahrscheinlichkeit mit Hilfe von a priori Wahrscheinlichkeiten und bedingten Wahrscheinlichkeiten zu berechnen.

Beispiel: Eine Firma baut 3 verschiedene elektronische Bauteile. Durchschnittlich sind 2% der Bauteile des ersten Typs, 5% des zweiten und 3% des dritten Ausschuß. Wie groß ist die Wahrscheinlichkeit, daß ein zufällig aus der Produktion ausgewähltes Bauteil Ausschuß ist, wenn der Anteil des ersten Typs an dem Produktionsausstoß 20%, der des zweiten 30% und der des dritten 50% beträgt? Steht A für das Ereignis, daß ein Bauteil defekt ist, und E_i dafür, daß das Teil vom Typ i ist, so ergibt der Satz für die totale Wahrscheinlichkeit:

$$P(A) = \sum_{i=1}^{3} P(E_i) \cdot P(A|E_i) = 0.2 \cdot 0.02 + 0.3 \cdot 0.05 + 0.5 \cdot 0.03 = 0.034$$

Ein Kunde beschwert sich, daß das ihm gelieferte Bauteil defekt ist, ohne den Typ des Bauteils anzugeben. Wie groß ist die Wahrscheinlichkeit, daß das Bauteil eines des ersten Typs ist? Nach dem Satz von Bayes erhält man

$$P(E_1|A) = \frac{P(A|E_1) \cdot P(E_1)}{\sum_{i=1}^{3} P(A|E_i) \cdot P(E_i)} = \frac{0.02 \cdot 0.2}{0.034} = 0.118$$

2.4 Zufallsvariable

Werden allen möglichen Ausgängen eines Zufallsexperiments (allen Elementarereignissen) durch eine Funktion Zahlen zugeordnet, spricht man von einer eindimensionalen Zufallsvariablen, die wir mit X, Y oder Z bezeichnen. Wird ein Zahlentupel $(X_1, \ldots, X_k)$ zugeordnet, so sprechen wir von einer mehrdimensionalen Zufallsvariablen oder einem Zufallsvektor.

Beispiel: Eine Münze wird dreimal geworfen. Die Menge der Elementarereignisse ist:

$$\Omega = \{WWW,\ ZWW,\ WZW,\ WWZ,\ ZZW,\ ZWZ,\ WZZ,\ ZZZ\}$$

wobei W für Wappen und Z für Zahl stehen. Nach dem Indifferenzprinzip hat jedes Elementarereignis die Wahrscheinlichkeit $1/8$. Die Zufallsvariable X sei nun als die Häufigkeit von Wappen definiert. Der Wertebereich von X ist dann $\{0, 1, 2, 3\}$. Die Wahrscheinlichkeiten für die einzelnen Werte werden durch die Wahrscheinlichkeitsverteilung auf Ω *induziert*. Daher gilt: $P(X = x) = P(\{\omega \in \Omega | X(\omega) = x\})$.

x	0	1	2	3	Σ
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Weitere Beispiele für Zufallsvariable sind:

- die Brenndauer einer Glühbirne (stetig).
- die Anzahl der Auftragseingänge eines Betriebs während eines Monats (diskret).
- die Dauer einer Reparatur in einer Werkstatt (stetig).
- die Anzahl der abgeschlossenen Versicherungsverträge einer Agentur (diskret).

Definition:

1. Eine Zufallsvariable heißt diskret, wenn ihr Wertebereich endlich oder abzählbar unendlich ist.
2. Eine Zufallsvariable heißt stetig, wenn ihre möglichen Werte wenigstens ein Intervall der reellen Zahlen $\mathcal{R}$ ausfüllen und kein Elementarereignis positive Wahrscheinlichkeit besitzt.

Durch Zufallsvariable wird eine Wahrscheinlichkeit auf dem Wertebereich (üblicherweise Intervalle) induziert. Wir schreiben

$$P_X(I) = P(X \in I) = P(\{\omega | X(\omega) \in I\}).$$

für Teilmengen I von $\mathcal{R}$. Falls keine Mißverständnisse auftreten können, schreibt man auch $P(I)$ statt $P_X(I)$.

Beispiel: Beschreibt die Zufallsvariable X die Brenndauer einer Glühbirne in Stunden, so ist $P_X((100, \infty))$ die Wahrscheinlichkeit, daß die Glühbirne länger als 100 Stunden brennt. Man beachte, daß die Wahrscheinlichkeit auf dem beiderseitig offenen Intervall $(100, \infty)$ berechnet wird.

Beispiel: Eine Münze wird dreimal geworfen. X sei die Häufigkeit von Wappen und Y sei die Anzahl der Versuche, bevor das erste Wappen erscheint. Falls bei keinem Versuch Wappen geworfen wird, so soll Y gleich 3 gesetzt werden. Der Wertebereich des Zufallsvektors (X, Y) ist: {(0,3), (1,0), (1,1), (1,2), (2,0), (2,1), (3,0)}. Das Paar (2,0) tritt bei WZW und WWZ ein. (X, Y) besitzt folgende Wahrscheinlichkeitsverteilung:

	x			
y	0	1	2	3
0	-	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
1	-	$\frac{1}{8}$	$\frac{1}{8}$	-
2	-	$\frac{1}{8}$	-	-
3	$\frac{1}{8}$	-	-	-

3 Diskrete Verteilungen

3.1 Grundlagen

Der Wertebereich M einer diskreten Zufallsvariablen X ist abzählbar. Besitzt ein Zufallsexperiment als Menge der Ausgänge die Menge der ganzen Zahlen, so ist $\Omega = \mathcal{Z}$. Die Wahrscheinlichkeitsverteilung $P_X(A)$ auf M wird durch ihre *diskrete Dichte* (Wahrscheinlichkeitsfunktion) beschrieben:

$$p_j = P_X(\{j\}) = P(X = j) = P(\{\omega \in \Omega | X(\omega) = j\}) \quad \text{für alle } j \in M$$

p_j ist die Wahrscheinlichkeit, daß die Zufallsvariable X die Ausprägung j annimmt. Ist $A \subset M$, dann gilt:

$$P_X(A) = \sum_{j \in A} p_j$$

Für diskrete Dichten gilt (falls $\Omega = \mathcal{Z}$):

$$p_j \geq 0 \text{ für alle } j \in \mathcal{Z}$$

$$\sum_{j=-\infty}^{\infty} p_j = 1$$

Eine Verteilung wird durch die Verteilungsfunktion vollständig repräsentiert:

Definition: Ist $P_X(A)$ die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X, so heißt

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x), \; x \in \mathcal{R}$$

die *Verteilungsfunktion* von $P_X(A)$. Man beachte, daß $F_X(x)$ eine Stufenfunktion darstellt und auf $\mathcal{R}$ definiert ist. Im folgenden schreiben wir nur $P(A)$ und $F(x)$ statt $P_X(A)$ bzw. $F_X(x)$, da wir nur eine Zufallsvariable behandeln und daher Mißverständnisse ausgeschlossen sind.

Satz: Die Verteilungsfunktion einer diskreten Zufallsvariablen wird durch folgende Eigenschaften charakterisiert:

1. $F(x)$ steigt monoton
2. $F(x) = \sum_{j \leq x} p_j$
3. $\lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$
4. $F(j) - F(j-1) = p_j$ für $j \in \mathcal{Z}$

Beweis: Übung.

Beispiel: Die Verteilungsfunktion der Zufallsvariablen X, die die Anzahl der im dreifachen Münzwurf geworfenen Wappen beschreibt, ist:

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{8} & \text{für } 0 \leq x < 1 \\ \frac{4}{8} & \text{für } 1 \leq x < 2 \\ \frac{7}{8} & \text{für } 2 \leq x < 3 \\ 1 & \text{für } 3 \leq x \end{cases}$$

Zur Charakterisierung einer Zufallsvariablen genügen häufig einzelne Kennzahlen, sogenannte Verteilungsparameter.

Definition: Es sei $g(X) : \mathcal{R} \to \mathcal{R}$ eine reellwertige Funktion. Dann ist der Erwartungswert von $g(X)$ durch folgenden Ausdruck gegeben:

$$E(g(X)) = \sum_{j \in M} g(j) \cdot p_j$$

Beispiel: Eine Telefonvermittlung kann maximal 10 Gespräche pro Minute vermitteln. Wird die Anzahl der Anrufe durch die Zufallsvariable X beschrieben und ist $g(x) = 0$ falls $x \leq 10$ und $g(x) = 1$ falls $x > 10$, so beschreibt $g(X)$, ob die Vermittlung überlastet ist oder nicht. $E(g(X))$ ist in diesem Fall die Wahrscheinlichkeit, daß die Vermittlung mehr als 10 Anrufe erhält.

Definition:

1. $E(X)$ heißt Erwartungswert und wird mit dem Symbol μ bezeichnet.

   $$E(X) = \sum_{j \in M} x_j \cdot p_j$$

2. $E((X-\mu)^2)$ heißt Varianz (Streuung) von X und wird mit $V(X), \sigma^2(X)$ oder σ^2 bezeichnet.

   $$E((X-\mu)^2) = \sum_{j \in M} (x_j - \mu)^2 \cdot p_j$$

3. $\sigma(X) = \sqrt{\sigma^2(X)}$ heißt Standardabweichung von X.

Eine zusätzliche Charakterisierung der Verteilung einer Zufallsvariablen X läßt sich durch die sogenannten Momente vornehmen.

Definition:

$$E((X-a)^k) = \sum_{j \in M} (x_j - a)^k p_j \qquad k = 1, 2, \ldots \text{heißt } k\text{tes Moment um } a .$$

$E(X^k)$ heißt $k$tes gewöhnliches Moment $(a = 0)$.

$E((X-\mu)^k)$ heißt $k$tes zentrales Moment $(a = \mu)$.

Das erste gewöhnliche Moment $E(X)$ ist der Erwartungswert. Das zweite zentrale Moment $E((X-\mu)^2)$ ist die Varianz von X.

Definition: Der *Momentkoeffizient der Schiefe* einer Zufallsvariablen X wird durch $S(X)$ definiert:

$$S(X) = \frac{E((X-\mu)^3)}{\sigma^3(X)} = \frac{E(X^3) - 3E(X^2)\mu + 2\mu^3}{\sigma^3(X)}$$

Ist $S(X)$ negativ (positiv), so ist die Verteilung der Zufallsvariablen linksschief (rechtsschief). Ist ihr Wert Null, so liegt eine symmetrische Verteilung vor.

Der *Momentkoeffizient der Wölbung* wird durch $W(X)$ definiert:

$$W(X) = \frac{E((X-\mu)^4)}{\sigma^4(X)} - 3 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - \mu^4}{\sigma^4(X)} - 3$$

Ist $W(X) > 0$, heißt die Verteilung leptokurtisch. Ist $W(X) = 0$, heißt sie mesokurtisch und ist $W(X) < 0$, heißt sie platykurtisch.

Für Erwartungswerte und Varianzen gelten folgende Rechenregeln:

Satz: Für reellwertige Zufallsfunktionen $g_1(X)$ und $g_2(X)$ gilt, sofern die Erwartungswerte existieren:

1. $E(g_1(X) + g_2(X)) = E(g_1(X)) + E(g_2(X))$
2. $E(c \cdot g_1(X)) = c \cdot E(g_1(X))$ für jede Konstante $c \in \mathcal{R}$

Mit Hilfe des letzten Satzes lassen sich folgende Regeln für den Erwartungswert und die Varianz herleiten.

Satz: Es seien $a, b \in \mathcal{R}$. Dann gilt:

1. $E(a + bX) = a + bE(X)$ (Linearitätsregel).
2. $\sigma^2(a + bX) = b^2 \cdot \sigma^2(X)$.
3. $\sigma^2(X) = E(X^2) - \mu^2$ Verschiebungssatz.

Beweis: Der Beweis des ersten Teils des Satzes sei dem Leser überlassen. Wir beweisen den Verschiebungssatz:

$$\begin{aligned} \sigma^2(X) = E((X-\mu)^2) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Beispiel: Im A-priori Modell, das auch Laplace Modell genannt wird, gehen wir davon aus, daß der Ereignisraum endlich viele Elementarereignisse besitzt und jedes Elementarereignis dieselbe Chance des Eintretens besitzt. Ist die

Anzahl der Elementarereignisse gleich N, so besitzt jedes Elementarereignis die Wahrscheinlichkeit $\frac{1}{N}$. Beispiele für Laplace Modelle sind: Roulette mit $N = 37$, Würfelexperiment mit einem Würfel ($N = 6$), einfacher Münzwurf ($N = 2$) oder das zufällige Ziehen aus einer Menge mit N Elementen. Gehören die Elementarereignisse $x_j, j = 1, \ldots, N$ zur Menge der ganzen Zahlen, so wird eine diskrete Zufallsvariable definiert, deren Verteilungsfunktion und Parameter folgendermaßen berechnet werden:

$$F(x) = \frac{1}{N} \cdot (\text{Anzahl der } x_j \leq x)$$

$$E(X) = \frac{1}{N} \sum_{j=1}^{N} x_j$$

$$\sigma^2(X) = \frac{1}{N} \sum_{j=1}^{N} (x_j - \mu)^2 = \left[\frac{1}{N} \sum_{j=1}^{N} x_j^2 \right] - \mu^2$$

Im Würfelexperiment mit einem Würfel ist $\mu = 3.5$ und $\sigma^2 = 2.91\overline{6}$. Zusätzlich zum Erwartungswert ist noch der Median als Lagemaß gebräuchlich, der mit $\tilde{x}$ oder $x_{0.5}$ bezeichnet wird.

Definition: Der Median (bezeichnet mit $\tilde{x}$ oder $x_{0.5}$) teilt den Wertebereich von X in zwei Bereiche, die gleich wahrscheinlich sind, auf. Formal wird der Median $\tilde{x}$ dadurch definiert, daß folgende Gleichungen gleichzeitig erfüllt sind:

$$P(X \leq \tilde{x}) \geq 0.5 \text{ und } P(X \geq \tilde{x}) \geq 0.5$$

Beispiel: Gegeben sei X mit Verteilungsfunktion $F(x)$

$X = x$	1	2	3	4	5
$F(x) = P(X \leq x)$	0.15	0.25	0.45	**0.85**	*1.0*
$P(X \geq x)$	*1.0*	*0.85*	*0.75*	**0.55**	0.15

Der einzige Wert, der beide Gleichungen erfüllt, ist 4. Daher gilt: $\tilde{x} = 4$.

Beispiel: Gegeben sei X mit Verteilungsfunktion $F(x)$ (Würfelwurf).

X	1	2	3	4	5	6
$F(x) = P(X \leq x)$	1/6	2/6	**3/6**	**4/6**	*5/6*	*1*
$P(X \geq x)$	*1*	*5/6*	**4/6**	**3/6**	2/6	1/6

Die obigen Gleichungen werden in diesem Beispiel durch alle Werte $\tilde{x} \in [3, 4]$ erfüllt. Man spricht daher von einer Medianklasse. Als charakteristischer Wert der Medianklasse wird das arithmetische Mittel aus Unter- und

Obergrenze der Medianklasse ausgewählt und wiederum als Median bezeichnet. In diesem Beispiel gilt daher:

$$\tilde{x} = \frac{1}{2}(3+4) = 3.5$$

Definition: Sei $\alpha \in (0,1)$. Das α Quantil x_α der Verteilung von X wird durch die folgenden Gleichungen definiert:

$$P(X \leq x_\alpha) \geq \alpha \text{ und } P(X \geq x_\alpha) \geq 1-\alpha$$

Spezialfälle: Der Median ($\alpha = 0.5$), das untere Quartil ($\alpha = 0.25$), das obere Quartil ($\alpha = 0.75$) sowie die Dezile ($\alpha = 0.1, \alpha = 0.2 \cdots, \alpha = 0.9$).

3.2 Spezialfälle

Zur Darstellung diskreter Verteilungen muß der Binomial-Koeffizient eingeführt werden.

Kombinatorik

n Objekte lassen sich auf $1 \cdot 2 \cdot 3 \cdot 4 \ldots \cdot (n-1) \cdot n = n!$ (n-Fakultät) Arten anordnen. Jede Anordnung der n Objekte oder von n Zahlen wird als Permutation bezeichnet. $0!$ wird durch 1 festgelegt.

$$\tau = \begin{pmatrix} 1 & 2 & 3 & \ldots & j & \ldots & n \\ \tau(1) & \tau(2) & \tau(3) & \ldots & \tau(j) & \ldots & \tau(n) \end{pmatrix}$$

Beispiel: Die Permutationen von $\{1,2,3\}$ sind:

$$\tau_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$

$$\tau_4 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \tau_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \tau_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

Es gibt $3! = 1 \cdot 2 \cdot 3 = 6$ Permutationen. Einen Spezialfall erhält man, wenn die Plätze ringförmig verteilt sind. Beispielsweise, wenn man alle Möglichkeiten sucht n Personen an einen runden Tisch zu setzen. Diesen Spezialfall nennt man *Ringpermutation*. Er wird durch $(n-1)!$ berechnet.

Stehen weniger als n Plätze zur Verfügung, um die n Objekte anzuordnen, dann ergeben sich für $k < n$ Plätze

$$\frac{n!}{(n-k)!}$$

Permutationen. Man beachte, daß die *Reihenfolge der Plätze dabei beachtet wird*. Wie man erkennen kann, ist die gewöhnliche Permutation ($n!$) lediglich ein Spezialfall mit $n = k$.

Soll zusätzlich die *Reihenfolge der Plätze nicht beachtet werden*, dann spricht man von einer *Kombination*. Sie wird berechnet als:

$$\binom{n}{k} = \frac{n \cdot (n-1) \ldots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \quad n,k \geq 0, \binom{n}{k} = 0 \text{ für } k > n$$

$\binom{n}{k}$ wird als Binomialkoeffizient bezeichnet.

Durch Einsetzen erhält man die Regeln:

1. $\binom{n}{k} = \binom{n}{n-k}$

2. $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ (Pascal'sches Dreieck)

Beispiel: Wieviele Möglichkeiten gibt es, aus einem Verein mit 25 Mitgliedern einen Vorstand, der aus 3 Personen besteht, zu wählen (Ämterhäufung ausgeschlossen)?

$$n = 25,\ k = 3,\ \binom{n}{k} = \binom{25}{3} = \frac{25 \cdot 24 \cdot 23}{1 \cdot 2 \cdot 3} = 25 \cdot 4 \cdot 23 = 2300.$$

Sollen n Objekte auf n Plätzen verteilt werden, wobei k_1 Objekte des Typs 1, k_2 des Typs 2, ..., k_p Objekte des Typs p, mit $n = \sum_{i=1}^{p} k_i$, dann existieren

$$\frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_p!}$$

Permutationen. Man beachte, daß sich die Kombination als Spezialfall mit $p = 2$ darstellen läßt.

Von einer *Variation* spricht man, wenn n Objekte auf k Plätzen verteilt werden sollen, wobei jedes Objekt mehrere Plätze einnehmen darf. Die Anzahl der Variationsmöglichkeiten sind n^k. Beispielsweise beträgt die Anzahl der möglichen Ausgänge eines Wurfes mit zwei Würfeln $n^k = 6^2 = 36$.

Die Bernoulli-Verteilung

Modellexperiment: Ein Zufallsexperiment besteht aus einem einzigen Versuch, in dem ein bestimmtes Ereignis A eintritt oder nicht. Wir definieren die folgende Zufallsvariable:

$$X(w) = \begin{cases} 0 & \text{falls} \quad w \notin A \\ 1 & \text{falls} \quad w \in A \end{cases}$$

Hat A die Wahrscheinlichkeit π, so besitzt X die Dichte:

$$p_0 = 1 - \pi,\ p_1 = \pi$$

Die wichtigsten Parameter sind:

$$\mu = \pi, \quad \sigma^2 = \pi(1-\pi), \quad S(X) = (1-2\pi)/\sqrt{\pi(1-\pi)}$$

Die Binomialverteilung

Ein Bernoulli-Experiment wird n mal *unabhängig und unter gleichen Bedingungen* durchgeführt. Beispiel: Eine Urne enthält schwarze und weiße Kugeln im Verhältnis $\pi : (1-\pi)$. Der Urne werden n Kugeln *mit Zurücklegen* entnommen. Die Anzahl X der dabei gezogenen schwarzen Kugeln ist $B_{n,\pi}$ verteilt. Das Ziehen mit Zurücklegen sichert, daß jeder Versuch unter gleichen Bedingungen durchgeführt wird, d.h. bei jedem Versuch ist der Anteil an schwarzen Kugeln in der Urne konstant.

Eine diskrete Zufallsvariable X ist binomialverteilt $B_{n,\pi}$ mit den Parametern n und π, wenn die Wahrscheinlichkeitsfunktion definiert ist durch:

$$p_k = P(X = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}, \quad k = 0, \ldots, n, \quad n > 0, \quad 0 \le \pi \le 1$$

Die wichtigsten Parameter sind:

$$\mu = n\pi, \quad \sigma^2 = n\pi(1-\pi), \quad S(x) = (1-2\pi)/\sqrt{n\pi(1-\pi)}$$

Wird ein Versuch n mal unabhängig unter gleichen Bedingungen durchgeführt und kann in jedem Versuch das Ereignis A mit der Wahrscheinlichkeit π eintreten, so ist die Anzahl X der Versuche mit dem Eintreten von A $B_{n,\pi}$ verteilt.

Die Hypergeometrische Verteilung

Ein Bernoulli-Experiment wird n mal hintereinander durchgeführt wobei die Wahrscheinlichkeit π für das Eintreten eines Elementarereignisses A sich nach dem Schema des folgenden Modelexperimentes verändern kann: Eine Urne enthält N Kugeln, von denen A schwarz und $N-A$ weiß sind. Der Urne werden *ohne Zurücklegen* n Kugeln entnommen. Die Anzahl X der dabei gezogenen schwarzen Kugeln ist $H_{N,A,n}$ verteilt. Das Ziehen ohne Zurücklegen bewirkt, daß jede Ziehung unter verschiedenen Bedingungen erfolgt.

Die hypergeometrische Verteilung $H_{N,A,n}$ besitzt die Wahrscheinlichkeitsfunktion

$$p_k = P(X=k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}, k = 0, \ldots, n, \text{ mit } n \leq A \text{ und } n \leq N-A$$

Ihre wichtigsten Parameter sind:

$$\pi = \frac{A}{N}, \; E(X) = \mu = n\frac{A}{N} = n\pi, \; V(X) = \sigma^2 = n\frac{A}{N}(1-\frac{A}{N})(\frac{N-n}{N-1}) = n\pi(1-\pi)(\frac{N-n}{N-1})$$

Man beachte, daß der Erwartungswert zu dem der Binomialverteilung identisch ist, sich die Varianz jedoch um die sog. *Endlichkeitskorrektur* $(\frac{N-n}{N-1})$ unterscheidet.

Beispiel: In der Schule beträgt die Anzahl der Schüler in der ersten Klasse 120, davon 70 Knaben und 50 Mädchen. Für einen Schulversuch werden 12 Kinder ausgewählt. Wie wahrscheinlich ist es, daß exakt das gleiche Verhältnis Jungen zu Mädchen wie in der ersten Klasse auftritt? Dieses Modell entspricht dem Ziehen ohne Zurücklegen. X sei Anzahl der Knaben im Schulversuch. X ist $H_{120,70,12}$ verteilt.

$$P(x=7) = \frac{\binom{70}{7}\binom{50}{5}}{\binom{120}{12}} = \frac{1 \cdot 1987745 \cdot 10^9 \cdot 2.118760 \cdot 10^6}{1.0542857 \cdot 10^{16}} = 0.2409$$

Die Poisson-Verteilung

Eine Zufallsvariable X besitzt eine *Poissonverteilung* P_λ, wenn sie die Wahrscheinlichkeitsfunktion

$$p_k = P(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \; k = 0, 1, 2, \ldots,$$

mit einem Erwartungswert von $\lambda > 0$ besitzt. (λ wird hier *Intensitätsparameter* genannt). Die Varianz ist λ und der Momentkoeffizient der Schiefe ist $1/\sqrt{\lambda}$.

Beispiel: Eine Brandschutzversicherung hat ermittelt, daß in einem bestimmten Gebiet im langjährigen Durchschnitt $\lambda = 1.5$ Schadensfälle über 100.000 DM auftreten. Um die notwendigen Reserven zu kalkulieren, möchte sie die Anzahl c der Schadensfälle berechnen, so daß $P(X > c) \leq 0.05$ ist.

Es gilt: $P(X > c) = 1 - P(X \leq c) = 1 - F(c)$.

Daher muß c so bestimmt werden, daß $F(c) \geq 0.95$ ist. Zu diesem Zweck bildet man die Verteilungsfunktion.

$$F(c) = \sum_{k=0}^{c} P(X = k) = \sum_{k=0}^{c} e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=0}^{c} e^{-1.5} \cdot \frac{1.5^k}{k!}$$

Die Werte der Verteilungsfunktion der Poissonverteilung mit $\lambda = 1.5$ sind:

k	0	1	2	3	4
p_k	0.223	0.334	0.251	0.125	0.047
$F(k)$	0.223	0.557	0.808	0.933	0.980

$F(4) = 0.98 \geq 0.95$. Die Versicherung muß daher Reserven für 4 Schadensfälle aufbringen, um ihr Risiko unter 5% zu halten.

Im folgenden wollen wir die ersten zwei gewöhnlichen Momente der Poissonverteilung herleiten, um daraus μ und σ^2 zu berechnen. Für diese Rechnung wird die Reihenentwicklung von e^x verwendet:

$$e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!}$$

$$\begin{aligned} E(X) &= \sum_{j=-\infty}^{\infty} j \cdot p_j = \sum_{j=0}^{\infty} j \cdot e^{-\lambda} \frac{\lambda^j}{j!} = e^{-\lambda} \sum_{j=1}^{\infty} j \cdot \frac{\lambda^{j-1}}{j(j-1)!} \lambda \\ &= e^{-\lambda} \cdot \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} \\ &= e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{-\lambda} \lambda e^{\lambda} = \lambda \, . \end{aligned}$$

$$\begin{aligned} E(X^2) &= \sum_{j=0}^{\infty} j^2 e^{-\lambda} \frac{\lambda^j}{j!} = e^{-\lambda} \sum_{j=1}^{\infty} j \frac{\lambda^j}{(j-1)!} \\ &= e^{-\lambda} \sum_{j=1}^{\infty} ((j-1)+1)) \frac{\lambda^j}{(j-1)!} \\ &= e^{-\lambda} \left[\sum_{j=1}^{\infty} (j-1) \frac{\lambda^j}{(j-1)!} + \sum_{j=1}^{\infty} \frac{\lambda^j}{(j-1)!} \right] \\ &= e^{-\lambda} \left[\sum_{j=2}^{\infty} \frac{\lambda^j}{(j-2)!} + \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} \right] \\ &= e^{-\lambda} \left[\lambda^2 \sum_{j=2}^{\infty} \frac{\lambda^{j-2}}{(j-2)!} + \lambda e^{\lambda} \right] = e^{-\lambda} \left[\lambda^2 e^{\lambda} + \lambda e^{\lambda} \right] \\ &= \lambda^2 + \lambda \, . \end{aligned}$$

Daraus folgt: $\sigma^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Praktisch können die folgenden Näherungen verwendet werden:

Verteilung	Näherung	Voraussetzung
$H_{N,A,n}$	$B_{n,\frac{A}{N}}$	$n \leq \frac{N}{10}$
$H_{N,A,n}$	$P_{n\cdot\frac{A}{N}}$	$n \leq \frac{N}{10}$ und $\frac{A}{N} \leq \frac{1}{10}$
$B_{n,\pi}$	$P_{n\pi}$	$\pi \leq \frac{1}{10}$

Beispiel: Da im Beispiel für die hypergeometrische Verteilung die Parameter $n = 12$, $N = 120$ der $H_{N,A,n}$ -Verteilung die Bedingung $n \leq N/10$ erfüllen, können wir mit der Binomialverteilung die Wahrscheinlichkeit annähernd bestimmen. Mit $n = 12$ und $\pi = A/N = 7/12$ erhalten wir

$$P(X = 7) = \binom{12}{7} \cdot \left(\frac{7}{12}\right)^7 \cdot \left(\frac{5}{12}\right)^5 = 792 \cdot 0.023 \cdot 0.0126 = 0.2295 .$$

4 Stetige Verteilungen

4.1 Grundlagen

Der Wertebereich M einer stetigen Zufallsvariablen X ist gleich $\mathcal{R}$ (Menge der reellen Zahlen) oder ein Intervall von $\mathcal{R}$. Die Wahrscheinlichkeitsverteilung $P_X(A)$ wird für das Ereignis $A = (-\infty, x]$ durch eine *stetig differenzierbare Verteilungsfunktion* beschrieben:

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = P_X(A)$$

Satz: Für die Verteilungsfunktion $F_X(x)$ einer stetigen Zufallsvariablen X und die dazu korrespondierende Dichtefunktion $f_X(x)$ gilt:

1. $F(x)$ steigt monoton
2. $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1$
3. $F'(x) = \frac{dF(x)}{dx} = f(x)$
4. $f(x) \geq 0$
5. $\int_{-\infty}^{\infty} f(x)dx = 1$

Man beachte, daß die Dichtefunktion $f(x)$ *keine* Wahrscheinlichkeitsfunktion ist, wie bei diskreten Verteilungen, und daß $f(x)$ durchaus größer als 1 sein kann (z.B. bei der Dreiecksverteilung). Die Wahrescheinlichkeit $P(X = x)$ ist außerdem bei stetigen Verteilungen immer 0.

Ist ein Ereignis A ein Intervall $A = (a, b]$, so ist

$$P_X(A) = P(a < X \leq b) = \int_a^b f(x)dx.$$

Dies entspricht dem Flächeninhalt unter der Funktion $f(x)$ im Intervall $(a, b]$. Wir schreiben im folgenden nur $P(A)$ und $F(x)$ statt $P_X(A)$ und $F_X(x)$.

Beispiel: Eine Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} 1 & \text{für } x \in [0,1] \\ 0 & \text{sonst} \end{cases}$$

heißt über dem Intervall [0,1] gleichverteilt.

Definition: Es sei $g(x) : \mathcal{R} \longrightarrow \mathcal{R}$ eine reellwertige Funktion. Dann ist der Erwartungswert von $g(X)$ definiert durch:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Wichtige Spezialfälle:

1. $E(X)$ heißt Erwartungswert von X (Symbol μ).

   $$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

2. $E((X-\mu)^2)$ heißt Varianz von X (Symbole: $V(X)$, $\sigma^2(x)$ und σ^2).

   $$E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

   $\sigma(X) = \sqrt{\sigma^2(X)}$ heißt Standardabweichung von X

3. $E((X-a)^k)$ heißt $k$tes Moment um a.

   $$E((X-a)^k) = \int_{-\infty}^{\infty} (x-a)^k f(x)dx$$

   $E(X^k)$ heißt $k$tes gewöhnliches Moment ($a = 0$).
   $E((X-\mu)^k)$ heißt $k$tes zentrales Moment ($a = \mu$).

Die Sätze über die Rechenregeln für Erwartungswerte von diskreten Zufallsvariablen gelten auch für stetige Zufallsvariablen.

Beispiel: Für die Gleichverteilung über [0,1] wollen wir die ersten 4 Momente und die Momentenkoeffizienten der Schiefe und der Wölbung berechnen.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 xdx = 0.5\,.$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 x^2 dx = \frac{x^3}{3}\Big]_0^1 = \frac{1}{3}\,.$$

$$E(X^3) = \int_{-\infty}^{\infty} x^3 f(x)dx = \int_0^1 x^3 dx = \frac{x^4}{4}\Big]_0^1 = \frac{1}{4}\,.$$

$$E(X^4) = \int_{-\infty}^{\infty} x^4 f(x)dx = \int_0^1 x^4 dx = \frac{x^5}{5}\Big]_0^1 = \frac{1}{5}\,.$$

$$E((X-\mu)^2) = E(X^2) - \mu^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\,.$$

$$E((X-\mu)^3) = E(X^3) - 3E(X^2)\mu + 2\mu^3 = \frac{1}{4} - 3\frac{1}{3}\cdot\frac{1}{2} + 2\left[\frac{1}{2}\right]^3 = 0\,.$$

$$E((X-\mu)^4) = \int_0^1 \left[x - \frac{1}{2}\right]^4 dx = \frac{1}{5}\left[x - \frac{1}{2}\right]^5\Bigg]_0^1 = \frac{1}{5}\left[\frac{1}{2}\right]^5 - \frac{1}{5}\left[-\frac{1}{2}\right]^5 = 0.0125\,.$$

$$S(X) = 0,\ W(X) = \frac{0.0125}{0.0833^2} - 3 = -1.2\,.$$

Die Verteilung ist symmetrisch und platykurtisch.

Bemerkung: Quantile für stetige Variable werden analog zu Quantilen von diskreten Zufallsvariablen definiert.

Lineare Transformation stetiger Zufallsvariablen

Satz: Die Zufallsvariable $Y = aX + b$ mit $a \neq 0$ und b als Konstante besitzt die Verteilungsfunktion

$$F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right), & \text{falls } a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right), & \text{falls } a < 0 \end{cases}$$

Beweis:

1. $a > 0$:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)\,.$$

2. $a < 0$:

$$F_Y(y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right) = P\left(X > \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)\,.$$

4.2 Spezialfälle

Die Gleichverteilung

Wir haben bereits die Gleichverteilung über dem Intervall $[0, 1]$ kennengelernt. Diese Definition kann auf beliebige endliche Intervalle übertragen werden. Eine Zufallsvariable X heißt gleichverteilt auf dem Intervall $[a, b]$, wenn sie die Dichte besitzt:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Die Dichte ist daher konstant. Die wichtigsten Momente sind:

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}, \quad S(X) = 0$$

Um den Erwartungswert und die Varianz herzuleiten, transformieren wir X in $Y = (X-a)/(b-a)$. Y ist dann über $[0,1]$ gleichverteilt. Da $X = (b-a)Y+a$ ist, erhält man nach dem Satz über Dichten transformierter Zufallsvariablen.

$$E(X) = (b-a)E(Y) + a = \frac{(b-a)}{2} + a = \frac{(a+b)}{2}$$

$$\sigma^2(X) = (b-a)^2\sigma^2(Y) = \frac{1}{12}(b-a)^2$$

Die Verteilungsfunktion ist somit:

$$F(x) = \begin{cases} 0 & \text{für } x < a \\ (x-a)/(b-a) & \text{für } a \leq x \leq b \\ 1 & \text{für } x > b \end{cases}$$

Daher folgt für $a \leq a_1 \leq b_1 \leq b$:

$$P(a_1 \leq X \leq b_1) = (b_1 - a_1)/(b-a)$$

Die Wahrscheinlichkeit in diesem Intervall wird somit in vollem Umfang durch die Intervallänge $b_1 - a_1$ determiniert.

Die Exponentialverteilung

Eine Zufallsvariable besitzt eine Exponentialverteilung, wenn sie die folgende Dichte hat:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Der Parameter $\lambda > 0$ beschreibt die 'Sterbe-' oder allgemeiner die 'Übergangsrate'. Die Verteilungsfunktion ist gegeben durch:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Die zentralen Parameter sind:

$$\mu = \frac{1}{\lambda},\ \sigma^2 = \frac{1}{\lambda^2},\ S(X) = \frac{2}{\lambda^3}$$

Die Exponentialverteilung wird auch 'Verteilung ohne Gedächtnis' genannt. Es gilt nämlich für $s \geq 0$ und $t \geq 0$:

$$P(X \leq s+t | X \geq t) = P(X \leq s)$$

Beweis: Übung.

Die Normal- oder Gaußverteilung

Eine Zufallsvariable X genügt einer Normalverteilung $N(\mu, \sigma^2)$, wenn sie die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \text{ für } -\infty < x < \infty$$

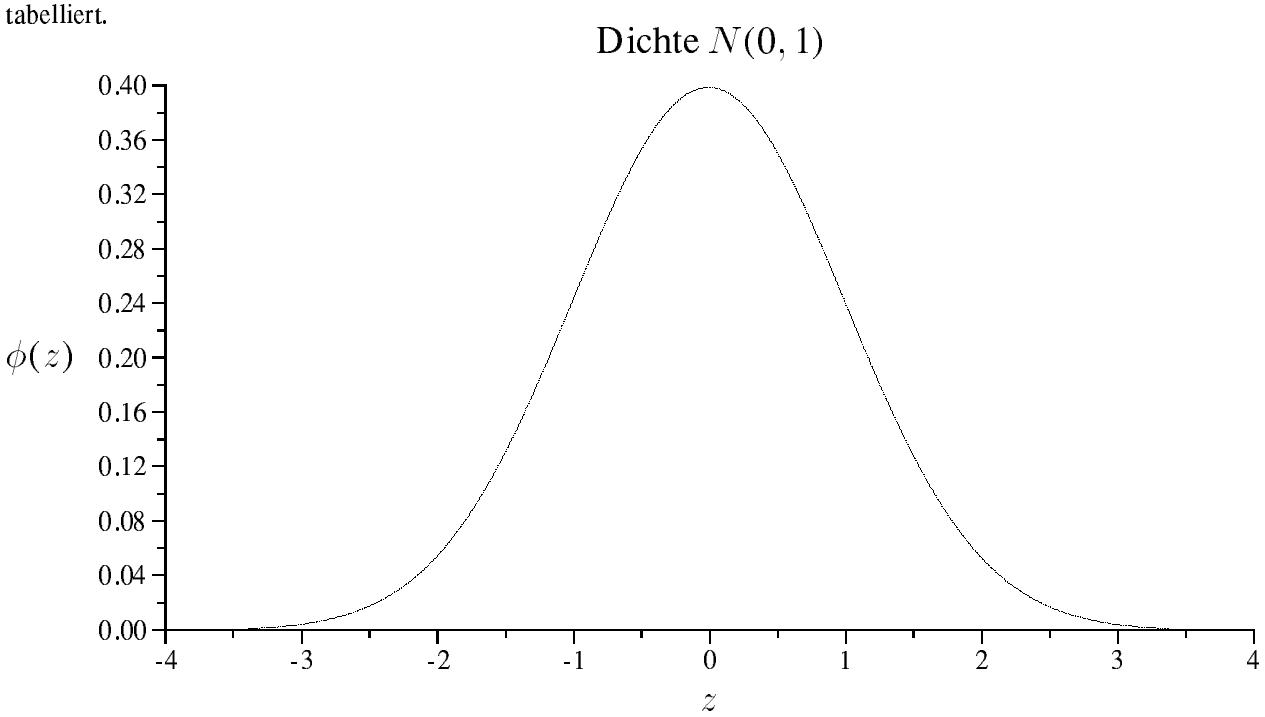
mit $\mu \in \mathcal{R}$ und $\sigma > 0$ besitzt.

Die Verteilung von X wird also durch 2 Parameter gekennzeichnet und zwar durch den Erwartungswert μ und die Varianz σ^2. Die Momentkoeffizienten der Schiefe und Wölbung sind Null.

Ist der Erwartungswert $\mu = 0$ und die Varianz $\sigma^2 = 1$, so nennt man X standardnormalverteilt ($N(0,1)$). Die Verteilungsfunktion der Standardnormalverteilung wird mit $\Phi(z)$ bezeichnet, die Dichte mit $\phi(z)$.

$$\Phi(z) = \int_{-\infty}^{z} \phi(z)dz = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx, \; -\infty < z < \infty$$

Da $\Phi(z)$ analytisch nicht exakt bestimmt werden kann, sind die Funktionswerte in Tabelle C2 im Anhang für $z \geq 0$ tabelliert.



Für negative z können die Symmetrieeigenschaften

1. $f(z) = f(-z)$
2. $P(Z \leq z) = P(Z \geq -z)$
3. $\Phi(-z) = 1 - \Phi(z)$

zur Berechnung verwendet werden.

Satz: Für die Verteilungsfunktion einer $N(\mu, \sigma^2)$ verteilten Zufallsvariablen gilt:

1. $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(z)$
2. $P(|X - \mu| > c) = 2(1 - \Phi\left(\frac{c}{\sigma}\right))$, für $c > 0$
3. $P(a \leq x \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$, für $a < b$

Beweis:

1. Wir wenden den Satz über die Dichte einer transformierten Zufallsvariable an, um zu zeigen, daß $\mu + \sigma \cdot Z$ und X dieselbe Verteilung besitzen, wobei $Z \sim N(0,1)$ und $X \sim N(\mu, \sigma^2)$ verteilt sind. Mit

   $x = g(z) = \mu + \sigma z$ und $g^{-1}(x) = \frac{x-\mu}{\sigma}$ folgt:

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

Daraus folgt: $X \sim N(\mu, \sigma^2)$.

2. $P(|X-\mu| > c) = P(-c > X-\mu) + P(X-\mu > c)$

$$= P\left(\frac{X-\mu}{\sigma} \leq \frac{-c}{\sigma}\right) + 1 - P\left(\frac{x-\mu}{\sigma} \leq \frac{c}{\sigma}\right) = \Phi\left(\frac{-c}{\sigma}\right) + 1 - \Phi\left(\frac{c}{\sigma}\right) = 2(1 - \Phi\left(\frac{c}{\sigma}\right)) .$$

3. $P(a \leq x \leq b) = P(x \leq b) - P(x < a)$

$$= P\left(\frac{x-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) - P\left(\frac{x-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) .$$

Für die α -Quantile gilt die folgende Regel: Ist z_α das α -Quantil der $N(0,1)$ Verteilung, so ist

$$x_\alpha = \mu + z_\alpha \cdot \sigma$$

das α -Quantil der $N(\mu, \sigma^2)$ -Verteilung. Man beachte, daß die Berechnung des α-Quantils einer stetigen Verteilung der Berechnung der Umkehrfunktion der entsprechenden Verteilungsfunktion entpricht ($\alpha = F^{-1}(x_\alpha)$). Die Multiplikation mit σ und die Addition von μ entspricht in diesem Fall der Umkehrung der Transformation ($\frac{x-\mu}{\sigma}$).

Beispiel: X sei $N(75, 36)$ verteilt. Gesucht ist $P(X \leq 87)$. Nach dem letzten Satz ist:

$$P(X \leq 87) = F(87) = \Phi\left(\frac{87-75}{6}\right) = \Phi(2) = 0.97725$$

Ferner ist das $\alpha = 0.99865$ Quantil zu bestimmen. Anwendung der Rechenregel für Quantile ergibt: $x_{0.99865} = \mu + \sigma \cdot z_{0.99865} = 75 + 6 \cdot 3 = 93$

Beispiel: Der Durchmesser von bestimmten Drehteilen aus einer automatischen Fertigung muß zwischen 9.5 und 12 cm liegen. Andernfalls gehört das Drehteil zum Ausschuß. Wie groß ist die Wahrscheinlichkeit, daß ein Drehteil den gestellten Anforderungen genügt, wenn der Durchmesser mit $\mu = 10.27$ und $\sigma^2 = 1.44$ normalverteilt ist.

$$\begin{aligned} P(9.5 \leq X \leq 12) &= \Phi\left(\frac{12-10.27}{1.2}\right) - \Phi\left(\frac{9.5-10.27}{1.2}\right) \\ &= \Phi(1.44) - \Phi(-0.63) = 0.925 - (1 - 0.736) = 0.661 \end{aligned}$$

Näherung der Poisson- und Binomialverteilung durch die Normalverteilung

Poisson: $P_\mu \sim N(\mu, \mu)$, für $\mu \geq 10$.

Binomial: $B_{n,\pi} \sim N(n\pi, n\pi(1-\pi))$ für alle $n\pi(1-\pi) \geq 10$.

Beispiel: In einer Telefonzentrale eines Konzerns werden durchschnittlich 25 Anrufe pro Minute gezählt. Wie groß ist die Wahrscheinlichkeit, daß mehr als 30 Anrufe in einer Minute gezählt werden, wenn die Anzahl der Anrufe poissonverteilt ist? Da $\mu = 25$ ist, ist die Anzahl X der Anrufe in der beobachteten Minute annähernd $N(25, 25)$ verteilt. Man erhält:

$$P(X > 30) = P\left(\frac{X-25}{5} > \frac{30-25}{5}\right) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

5 Mehrdimensionale Verteilungen

Bei vielen Fragestellungen der Statistik werden *mehrere* Zufallsvariablen als Ergebnis *eines* Zufallsexperiments betrachtet. So kann man etwa bei einer Untersuchung von Haushalten die Variablen Haushaltsgröße (X), Haushaltseinkommen (Y), Konsumausgaben (Z) usw. untersuchen.

5.1 Diskrete Verteilungen

Betrachtet man den Ausgang eines Zufallsexperiments, der durch das k dimensionale Merkmal $(X_1, \ldots, X_k)$ beschrieben wird, dann heißt die k dimensionale Zufallsvariable $(X_1, \ldots, X_k)$ diskret, falls ihr Wertebereich endlich oder abzählbar unendlich, z. B. gleich $\mathcal{Z}^k$ oder einer Teilmenge von $\mathcal{Z}^k$ ist. Die Zufallsvariable heißt stetig, falls ihr Wertebereich überabzählbar unendlich ist und kein Punkt aus dem $\mathcal{R}^k$ eine positive Wahrscheinlichkeit besitzt.

Beispiel: Ein Verlag publiziert 6 Wochenzeitschriften. $X_1, \ldots, X_6$ seien die Anzahlen der verkauften Zeitschriften pro Woche. $(X_1, \ldots, X_6)$ ist eine 6-dimensionale Zufallsvariable.

Beispiel: Beschreibt X_1 die Liegezeit und X_2 die zu löschende Ladung eines Schiffes, so ist (X_1, X_2) ein 2-dimensionaler stetiger Zufallsvektor. Der Vektor $X = (X_1, \ldots, X_k)$ heißt k dimensionale Zufallsvariable oder Zufallsvektor mit k Komponenten.

Die Wahrscheinlichkeitsfunktion eines diskreten k dimensionalen Zufallsvektors wird durch eine Dichte mit k Argumenten beschrieben.

$$p(x_1, x_2, \ldots, x_k) = P(X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_k = x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$$

Für eine diskrete Dichte gelten folgende Eigenschaften:

1. $0 \leq p(x_1, \ldots, x_k) \leq 1$
2. $\sum_{x_k \in \mathcal{Z}} \cdots \sum_{x_1 \in \mathcal{Z}} p(x_1, \ldots, x_k) = 1$
3. $P(A) = P((X_1, \ldots, X_k) \in A) = \sum_{(x_1 \ldots x_k) \in A} p(x_1, \ldots, x_k)$

Die *Verteilungsfunktion* ist gegeben durch:

$$F(x_1, \ldots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k) = \sum_{y_k \leq x_k} \cdots \sum_{y_1 \leq x_1} p(y_1, \ldots, y_k)$$

Randverteilungen

Wir beschränken uns auf den Fall $k = 2$; Randverteilungen höherer Ordnung werden analog gebildet. (X_1, X_2) sei eine 2-dimensionale diskrete Zufallsvariable mit der Dichte $p(x_1, x_2)$ und der Verteilungsfunktion $F(x_1, x_2)$. Die eindimensionalen Randdichten und Randverteilungen sind definiert durch:

$$\begin{aligned}
p_1(x_1) &= P(X_1 = x_1) = \sum_{x_2 = -\infty}^{\infty} p(x_1, x_2) \\
F_1(x_1) &= P(X_1 \leq x_1) = F(x_1, \infty) \\
p_2(x_2) &= P(X_2 = x_2) = \sum_{x_1 = -\infty}^{\infty} p(x_1, x_2) \\
F_2(x_2) &= P(X_2 \leq x_2) = F(\infty, x_2)
\end{aligned}$$

Sie entsprechen also den eindimensionalen Dichten und Verteilungen, wenn man die jeweils anderen Variablen unberücksichtigt läßt.

Bedingte Verteilungen und bedingte Dichten

Als Folge der Definition der bedingten Wahrscheinlichkeit

$$P(A|B) = P(A \cap B)/P(B)$$

erhält man die Definition einer bedingten Verteilung (zweidimensionaler Fall):

$$F_{X_1|X_2}(x_1|x_2) = \frac{F(x_1, x_2)}{F_2(x_2)}$$

Die bedingte Dichte ist gegeben durch:

$$p_{X_1|X_2}(x_1|x_2) = \frac{p(x_1, x_2)}{p_2(x_2)}$$

Beispiel: Für einen psychologischen Test, der aus zwei Teilen mit 2 bzw. 3 Aufgaben besteht, werden die Wahrscheinlichkeiten, daß jeweils 0 bis 2 bzw. 0 bis 3 Aufgaben gemeinsam gelöst werden, wie folgt angegeben:

	x_2				
x_1	0	1	2	3	$p_1(x_1)$
0	0.1	0.05	0.05	0	0.2
1	0.1	0.1	0.15	0.05	0.4
2	0.05	0.05	0.1	0.2	0.4
$p_2(x_2)$	0.25	0.2	0.3	0.25	1

Die Verteilungsfunktion von (X_1, X_2) an der Stelle $x_1 = 0, x_2 = 2$ ist $F(0, 2) = 0.1 + 0.05 + 0.05 = 0.2$.

Unabhängigkeit

Zwischen den Komponenten eines k dimensionalen Zufallsvektors können Zusammenhänge bestehen. Ein wichtiger Spezialfall ist die *stochastische Unabhängigkeit*. Gilt für einen diskreten Zufallsvektor $(X_1, \ldots, X_k)$ die Aussage

$$p(x_1, \ldots, x_k) = p_1(x_1) \cdots p_k(x_k) \quad ,$$

so heißen $X_1, \ldots, X_k$ stochastisch unabhängig.

Satz: $X_1, \ldots, X_k$ sind genau dann stochastisch unabhängig, wenn gilt:

$$F(x_1, \ldots, x_k) = F_1(x_1) \cdots F_k(x_k)$$

Im letzten Beispiel sind X_1 und X_2 nicht stochastisch unabhängig, denn es gilt:

$$F(0,2) = 0.2 \neq F_1(0) \cdot F_2(2) = 0.2 \cdot 0.75 = 0.15$$

Kovarianz und Korrelation

$(X_1, \ldots, X_k)$ sei ein k dimensionaler diskreter Zufallsvektor. Die Funktion

$$\mu_i = E(X_i) = \sum_{x_k \in \mathcal{Z}} \cdots \sum_{x_1 \in \mathcal{Z}} x_i \cdot p(x_1, \ldots, x_k) = \sum_{x_i \in \mathcal{Z}} x_i \cdot p_i(x_i)$$

heißt Erwartungswert der Zufallsvariablen X_i, $(i = 1, \ldots, k)$, wobei $p_i(x_i)$ die marginale Wahrscheinlichkeitsfunktion ist. Die Funktion

$$\sigma_{ii} = \sigma^2(X_i) = E((X_i - \mu_i)^2) = \sum_{x_i \in Z} (x_i - \mu_i)^2 p_i(x_i)$$

heißt Varianz von X_i . Die Funktion

$$\begin{aligned} \sigma_{ij} &= E((X_i - \mu_i)(X_j - \mu_j)) \\ &= \sum_{x_k \in \mathcal{Z}} \cdots \sum_{x_1 \in \mathcal{Z}} (x_i - \mu_i)(x_j - \mu_j) \cdot p(x_1, \ldots, x_k) \end{aligned}$$

heißt *Kovarianz* der Zufallsvariablen X_i und X_j $(i \neq j)$. Die Matrix $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,k}$ heißt *Kovarianzmatrix* der Zufallsvariablen $(X_1, \ldots, X_k)$. Für die Kovarianz gilt:

$$\begin{aligned} \sigma_{ij} &= E[(X_i - E(X_i)) \cdot (X_j - E(X_j))] \\ &= E[(X_i X_j - X_i \cdot E(X_j) - E(X_i) \cdot X_j + E(X_i) \cdot E(X_j)] \\ &= E(X_i X_j) - E(X_i) \cdot E(X_j) - E(X_i) \cdot E(X_j) + E(X_i) \cdot E(X_j) \\ &= E(X_i X_j) - E(X_i) \cdot E(X_j) \end{aligned}$$

Damit ergibt sich die zur praktischen Berechnung einfachere Formel

$$\sigma_{ij} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} x_i x_j p_{ij}(x_i, x_j) \right) - \mu_i \cdot \mu_j \quad .$$

Der Wert der Kovarianz hängt von den Einheiten ab, in denen x_i und x_j gemessen werden. Zur Normierung verwendet man den Korrelationskoeffizienten

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\, \sigma_{jj}}} \quad .$$

Damit gilt: $\rho_{ij} \in [-1, 1]$

Satz: Sind X_i und X_j stochastisch unabhängig (und damit auch *linear unabhängig*), so gilt:

$$\sigma_{ij} = 0 \quad \Longrightarrow \quad \rho_{ij} = 0$$

Besteht eine exakte lineare Beziehung zwischen X_i und X_j, so gilt:

$$|\rho_{ij}| = 1$$

Ein positiver Korrelationskoeffizient nahe bei 1 weist auf einen starken positiven linearen Zusammenhang hin, während ein negativer Korrelationskoeffizient auf einen negativen linearen Zusammenhang deutet. Zur inhaltlichen Interpretation von Korrelationskoeffizienten beachte man die Ausführungen auf Seite 63.

Für die beiden psychologischen Tests gilt:

$$\mu_1 = 1.2, \quad \mu_2 = 1.55, \quad \sigma_{11} = 0.56, \quad \sigma_{22} = 1.2475 \quad \text{und}$$

$$\sigma_{12} = E(X_1X_2) - \mu_1\mu_2 = 2.25 - 1.2 \cdot 1.55 = 0.39$$

Kovarianzmatrix und Korrelationskoeffizient:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.56 & 0.39 \\ 0.39 & 1.2475 \end{pmatrix}$$

$$\rho_{12} = \frac{0.39}{\sqrt{0.56 \cdot 1.2475}} = 0.4666$$

Die Kovarianzmatrix Σ enthält die Kovarianzen $\sigma_{ij} = COV(X_i, X_j), i = 1, \ldots, k, j = 1, \ldots, k$. Die Hauptdiagonalelemente entsprechen dabei den Varianzen $\sigma_{ii} = \sigma_i^2$. Daher ist die Kovarianzmatrix immer symmetrisch.

Die Multinomialverteilung

Beispiel: Ein wichtiges Beispiel für eine k dimensionale diskrete Verteilung ist die Multinomialverteilung mit der Dichte:

$$p(x_1, \ldots, x_k) = \frac{n!}{x_1!x_2! \ldots x_k!} \pi_1^{x_1} \pi_2^{x_2} \ldots \pi_k^{x_k}$$

mit $0 \leq x_i \leq n, \; x_1 + \ldots + x_k = n, \; 0 < \pi_i < 1$ und $\pi_1 + \ldots + \pi_k = 1$

Erwartungswerte: $\mu_i = n\pi_i$

Varianzen: $\sigma_{ii} = n\pi_i(1 - \pi_i) \;\; 1 \leq i \leq k$

Kovarianzen: $\sigma_{ij} = -n\pi_i\pi_j \;\; i, j = 1, \ldots, k$ und $i \neq j$

Als Spezialfall der Multinomialverteilung ergibt sich für $k = 2$ die Binomialverteilung.

5.2 Stetige Verteilungen

Die Wahrscheinlichkeitsverteilung einer stetigen k dimensionalen Zufallsvariablen $(X_1, \ldots, X_k)$ wird durch die *Dichte* $f(x_1, \ldots, x_k)$ beschrieben, für die gilt:

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_k \leq X_k \leq b_k) = \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} f(x_1, \ldots, x_k) dx_1 \ldots dx_k$$

Dabei seien $a_i \leq b_i$, für alle $a_i, b_i \in \mathcal{R}, i = 1, \ldots, n$.

Die Dichte $f(x_1, \ldots, x_k)$ erfüllt folgende Bedingungen:

1. $f(x_1, \ldots, x_k) \geq 0$
2. $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k) dx_1 \ldots dx_k = 1$

Die *Verteilungsfunktion* ist gegeben durch:

$$F(x_1, \ldots, x_k) = P(X_1 \leq x_1, \ldots, X_k \leq x_k) = \int_{-\infty}^{x_k} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_k) dt_1 \ldots dt_k$$

Randverteilungen

Wir beschränken uns wieder auf den Fall $k = 2$. Die Dichten und Verteilungsfunktionen der beiden Randverteilungen des stetigen Zufallsvektors (X_1, X_2) sind gegeben durch:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2; \quad F_1(x_1) = P(X_1 \leq x_1) = F(x_1, \infty)$$

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1; \quad F_2(x_2) = P(X_2 \leq x_2) = F(\infty, x_2)$$

Die zweidimensionale Gleichverteilung

Die Dichte der 2-dimensionalen gleichverteilten Zufallsvariablen (X_1, X_2) über dem Rechteck $[a_1, b_1] \times [a_2, b_2]$ ist wie folgt definiert:

$$f(x_1, x_2) = \begin{cases} \frac{1}{(b_1 - a_1)(b_2 - a_2)} & \text{für } a_1 \leq x_1 \leq b_1 \text{ und } a_2 \leq x_2 \leq b_2 \\ 0 & \text{sonst} \end{cases}$$

Die Randdichten sind gegeben durch:

$$f_1(x_1) = \int_{a_2}^{b_2} \frac{1}{(b_1 - a_1)(b_2 - a_2)} dx_2 = \frac{1}{(b_1 - a_1)(b_2 - a_2)} \int_{a_2}^{b_2} 1 \, dx_2 = \frac{1}{b_1 - a_1}$$

Analog erhält man $f_2(x_2) = \frac{1}{b_2 - a_2}$. Die Randverteilungen sind also eindimensionale Gleichverteilungen.

Unabhängigkeit

Die stetigen Zufallsvariablen $X_1, \ldots, X_k$ heißen *stochastisch unabhängig*, wenn gilt:

$$f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k)$$

Im letzten Beispiel sind X_1 und X_2 stochastisch unabhängig, da $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$.

Kovarianz und Korrelation

$(X_1, \ldots, X_k)$ sei ein stetiger Zufallsvektor. Erwartungswert und Varianz der Komponente X_i sind gegeben durch:

$$\mu_i = E(X_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} x_i f(x_1, \ldots, x_k) dx_k \ldots dx_1 = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i$$

$$\sigma_{ii} = E(X_i - \mu_i)^2 = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f(x_1, \ldots, x_k) dx_k \ldots dx_1 = \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i) dx_i$$

Die Kovarianz der beiden Variablen X_i, X_j ist gegeben durch:

$$\begin{aligned} \sigma_{ij} &= E((X_i - \mu_i)(X_j - \mu_j)) \\ &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \ldots, x_k) dx_k \ldots dx_1 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f_{ij}(x_i, x_j) dx_i dx_j \end{aligned}$$

Die Kovarianzmatrix von $(X_1, \ldots, X_k)$ ist gegeben durch:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}$$

Der Korrelationskoeffizient zwischen X_i und X_j ist definiert durch:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\, \sigma_{jj}}}$$

Die k dimensionale Normalverteilung

Als Beispiel für eine k dimensionale stetige Verteilung geben wir die Dichte der k dimensionalen Normalverteilung an. Sei X ein k dimensionaler Spaltenvektor mit Erwartungswert μ und Kovarianzmatrix Σ. Die Determinante wird mit $|\Sigma|$ und die Inverse der Kovarianzmatrix mit Σ^{-1} bezeichnet. Daher sei $|\Sigma| > 0$.

$$\mu = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}$$

Die Dichte der k variaten Normalverteilung im Punkt x ist dann gegeben durch:

$$f(x_1, \ldots, x_k) = (2\pi)^{\frac{-k}{2}} \cdot |\Sigma|^{\frac{-1}{2}} \cdot \exp\left\{ -\frac{1}{2}(x - \mu)^T \cdot \Sigma^{-1} \cdot (x - \mu) \right\} .$$

Bemerkung: Wie aus der Dichte ersichtlich ist, gilt im Fall der Normalverteilung:

$\sigma_{ij} = 0$ für $i \neq j \iff X$ stochastisch unabhängig.

Σ wird in diesem Fall eine Diagonalmatrix, d.h. die Dichte kann als Produkt der marginalen Dichten geschrieben werden.

$$f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k).$$

Im Falle einer bivariaten Normalverteilung, also $k = 2$, erhält man mit $\rho = \rho_{12}$ als Kovarianzmatrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Determinante:

$$|\Sigma| = \sigma_1^2\sigma_2^2 \cdot (1 - \rho^2)$$

Inverse:

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2 \cdot (1 - \rho^2)} \cdot \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

σ_i^2 bezeichnet die Varianzen von X_i, $i = 1, 2$. ρ ist der Korrelationskoeffizient von X_1 und X_2. Als Dichte einer bivariaten Normalverteilung erhält man:

$$f(x_1, x_2) = (2\pi)^{-1} \cdot |\Sigma|^{\frac{-1}{2}} \cdot \exp\left\{ -\frac{1}{2}(x_1 - \mu_1, x_2 - \mu_2) \cdot \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right\} =$$

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

6 Grenzwertsätze

6.1 Linearkombination von Zufallsvariablen

Ein Zufallsexperiment wird n-mal unabhängig wiederholt. Diese Standardformulierung bedeutet, daß entweder ein Experiment n-mal durchgeführt wird oder n Experimente (ohne gegenseitige Beeinflussung) ein einziges Mal durchgeführt werden. Man kann zeigen, daß diese unterschiedlichen Auffassungen mathematisch identisch sind. Formal wird dieses Experiment durch eine Folge von n stochastisch unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ beschrieben mit existierenden Erwartungswerten μ_i und Varianzen σ_i^2. Für die Linearkombination

$$Z = a_1X_1 + \ldots + a_nX_n$$

gilt der folgende Satz:

1. $E(Z) = \sum_{i=1}^{n} a_i\mu_i$

2. $V(Z) = \sum_{i=1}^{n} a_i^2\sigma_i^2$

Beweis: Übung.

Beispiel: Seien $X_1, \ldots, X_n$ stochastisch unabhängig identisch verteilt. Sei $\bar{x}$ das arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Setzen wir $a_1 = \ldots = a_n = \frac{1}{n}$, so ist

$$\bar{x} = \frac{1}{n} x_1 + \ldots + \frac{1}{n} x_n = a_1 x_1 + \ldots + a_n x_n$$

Also gelten die Aussagen:

$$E(\bar{x}) = \mu \cdot \sum_{i=1}^{n} a_i = \mu \cdot \sum_{i=1}^{n} \frac{1}{n} = \mu, \quad \sigma^2(\bar{x}) = \sigma^2 \cdot \sum_{i=1}^{n} a_i^2 = \sigma^2 \cdot \sum_{i=1}^{n} \frac{1}{n^2} = \frac{1}{n} \sigma^2$$

6.2 Stochastische Ungleichungen

Die Ungleichung von Markov

Satz: Es sei X eine nicht-negative Zufallsvariable mit Erwartungswert $E(X)$. Dann ist für jede positive Zahl ϵ die folgende Ungleichung erfüllt:

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}$$

Beweis:

1. X diskret

$$E(X) = \sum_{j=0}^{\infty} j \cdot p_j \geq \sum_{j \geq \epsilon} j \cdot p_j \geq \sum_{j \geq \epsilon} \epsilon \cdot p_j = \epsilon \sum_{j \geq \epsilon} p_j = \epsilon \cdot P(X \geq \epsilon)$$

2. X stetig

$$E(X) = \int_0^{\infty} x \cdot f(x) dx \geq \int_{\epsilon}^{\infty} x \cdot f(x) dx \geq \int_{\epsilon}^{\infty} \epsilon \cdot f(x) dx = \epsilon \cdot \int_{\epsilon}^{\infty} f(x) dx = \epsilon \cdot P(X \geq \epsilon)$$

Beispiel: Die durchschnittliche Anzahl von Anträgen bei einer Behörde ist 5. Wie groß ist die obere Schranke der Wahrscheinlichkeit, daß 10 oder mehr Anträge eintreffen?

$$P(X \geq 10) \leq \frac{E(X)}{10} = \frac{5}{10} = \frac{1}{2}$$

Die Ungleichung von Tschebyscheff

Satz: Existiert für eine Zufallsvariable X mit Erwartungswert $E(X)$ zusätzlich die Varianz $\sigma^2(X)$, so folgt für alle $\epsilon > 0$:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2(X)}{\epsilon^2}$$

Da die Ungleichung von Tschebyscheff eine Aussage über die Wahrscheinlichkeit einer Abweichung vom Mittelwert macht, muß für eine Abschätzung von $P(X \geq k)$ zuerst eine entsprechende Transformation durchgeführt werden. Da hier eine Maximalwahrscheinlichkeit berechnet wird, kann die Ungleichung von Tschebyscheff auch für einseitige Abweichungen vom Mittelwert verwendet werden.

Beweis: Es gilt $P(|X-\mu| \geq \epsilon) = P((X-\mu)^2 \geq \epsilon^2)$. Die Anwendung des Satzes von Markov auf die Zufallsvariable $(X-\mu)^2$ ergibt:

$$P((X-\mu)^2 \geq \epsilon^2) \leq \frac{E((X-\mu)^2)}{\epsilon^2} = \frac{\sigma^2(X)}{\epsilon^2}$$

Beispiel: Der Bedarf an Teilen eines bestimmten Typs in einem Produktionsbetrieb wird durch die Zufallsvariable X beschrieben. Aus Erfahrung ist bekannt, daß $E(X)$ gleich 45 und $\sigma^2(X) = 5$ ist. Mit welcher Wahrscheinlichkeit liegt der Bedarf zwischen 40 und 50 Teilen?

$$P(40 < X < 50) = P(|X-45| < 5) = 1 - P(|X-45| \geq 5)$$

$$P(|X-45| \geq 5) \leq \frac{5}{25} = 0.2$$

Die Wahrscheinlichkeit, daß der Bedarf zwischen 40 und 50 Teilen liegt, ist mindestens 0.8.

6.3 Schwaches Gesetz der großen Zahlen

Satz (Schwaches Gesetz der großen Zahlen): Existiert für stochastisch unabhängige und identisch verteilte Zufallsvariablen $X_i, i = 1, 2, \ldots$ der Erwartungswert μ und die Varianz σ^2, dann gilt für das arithmetische Mittel $\overline{x}$ und für beliebig kleine $\epsilon > 0$ die Aussage:

$$\lim_{n \to \infty} P(|\overline{x} - \mu| \geq \epsilon) = 0$$

Für $n \to \infty$ ist also die Wahrscheinlichkeit einer Abweichung des Wertes $\bar{x}$ vom Mittelwert μ um mehr als eine Konstante ϵ gleich Null. Man spricht von einer *Konvergenz nach Wahrscheinlichkeit*.

Dieser Satz ist ein Spezialfall des folgenden allgemeinen Satzes: $X_i, i \in \mathcal{N}$ ($\mathcal{N} = \{1, 2, 3, \ldots\}$), seien stochastisch unabhängige und identisch verteilte Zufallsvariablen. $g(x)$ sei eine reellwertige Funktion. $\mu_g = E(g(X_i))$ und $\sigma_g^2 = \sigma^2(g(X_i))$ existieren. Es sei $\bar{g}_x$ der Mittelwert: $\bar{g}_x = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$. Dann gilt die folgende Aussage für beliebig kleine $\epsilon > 0$.

$$\lim_{n \to \infty} P(|\bar{g}_x - \mu_g| \geq \epsilon) = 0$$

Beweis:

$$\sigma^2(\bar{g}_x) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2(g(X_i)) = \frac{1}{n^2} n \cdot \sigma_g^2 = \frac{1}{n}\sigma_g^2$$

Aus der Tschebyscheff-Ungleichung folgt:

$$P(|\bar{g}_x - \mu_g)| \geq \epsilon) \leq \frac{\sigma^2(\bar{g}_x)}{\epsilon^2} = \frac{1}{n \cdot \epsilon^2} \cdot \sigma_g^2$$

Daraus folgt:

$$\lim_{n \to \infty} P(|\bar{g}_x - \mu_g| \geq \epsilon) = 0$$

Ein Zufallsexperiment wird n mal unabhängig wiederholt. Tritt im i ten Versuch das Ereignis A auf, nimmt die Zufallsvariable X_i den Wert 1 an, sonst den Wert 0. Sei X_i Bernoulli verteilt mit $\mu = P(A) = \pi$ und $\sigma^2(X_i) = \pi(1-\pi)$. Die relative Häufigkeit von A in n Versuchen ist gegeben durch:

$$\tilde{p}_n = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Man beachte, daß $\tilde{p}_n$ die relative Häufigkeit ist, die wir bereits in der deskriptiven Statistik kennengelernt haben.

Satz (Theorem von Bernoulli): Es sei $\tilde{p}_n$ die relative Häufigkeit eines Ereignisses in n unabhängigen Wiederholungen eines Zufallsexperiments und π die Wahrscheinlichkeit des Ereignisses. Dann gilt:

$$\lim_{n\to\infty} P(|\tilde{p}_n - \pi| \geq \epsilon) = 0$$

Dies bedeutet, daß die Wahrscheinlichkeit einer beliebig kleinen Abweichung der relativen Häufigkeit von der Wahrscheinlichkeit eines Ereignisses bei wachsendem n immer kleiner wird. Daher läßt sich im Fall unabhängiger Zufallsexperimente der formal eingeführte Wahrscheinlichkeitsbegriff inhaltlich erklären. Auch in der Praxis kann man die Ergebnisse sinnvoll verwenden, denn aufgrund der Gesetze kann man Parameter einer Verteilung (konsistent – siehe später) schätzen.

6.4 Zentraler Grenzwertsatz

Das Gesetz der großen Zahl besagt, **daß** das arithmetische Mittel $\bar{x}$ gegen den Mittelwert μ konvergiert. Der Zentrale Grenzwertsatz gibt nun Auskunft darüber, **wie** $\bar{x}$ gegen μ konvergiert. Unter den gleichen Voraussetzungen wie im vorigen Abschnitt gilt nun:

Zentraler Grenzwertsatz von Lindeberg - Levy: $X_i, i \in \mathcal{N}$, sei eine Folge von stochastisch unabhängigen und identisch verteilten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2. Dann gilt für die Folge S_n der standardisierten Partialsummen der Zufallsvariablen X_i :

$$S_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}$$

$$\lim_{n\to\infty} F_{S_n}(x) = \Phi(x) \text{ für alle } x \in \mathcal{R}$$

Hier bezeichnet F_{S_n} die Verteilungsfunktion von S_n und Φ die Standardnormalverteilung. Es folgt, daß $\bar{x} \sim N(\mu, \sigma^2/n)$ für $n \to \infty$. Man spricht in solchen Fällen von einer *Konvergenz nach Verteilung*. Ab $n \geq 30$ ist S_n in guter Näherung normalverteilt. Für die Anwendbarkeit des zentralen Grenzwertsatzes ist nur die Existenz von Erwartungswert und Varianz wichtig, die Gestalt der Verteilung spielt keine Rolle.

Beispiel: Wir untersuchen die Wahrscheinlichkeitsverteilung der Anzahl X von Wappen in 10 Münzwürfen (Werfen mit einer fairen Münze). Da P ('Wappen')= 1/2 ist, ist $X \sim B_{10;0.5}$ verteilt. Wir erhalten die folgende Verteilungsfunktion und vergleichen sie mit der $N(5, 2.5)$ Verteilungsfunktion.

j	0	1	2	3	4	5	6	7	8	9
$F(j)$	0	0.01	0.05	0.17	0.38	0.62	0.83	0.95	0.99	1.00
$\Phi(\frac{(j-5)}{\sqrt{2.5}})$	0	0.01	0.03	0.10	0.26	0.50	0.74	0.90	0.97	0.99

7 Grundbegriffe der mathematischen Statistik

7.1 Grundgesamtheit und Stichprobe

Als Grundgesamtheit bezeichnen wir die Gesamtheit aller Merkmalsträger, die in einer statistischen Untersuchung auftreten können. Beispiele für Grundgesamtheiten sind

- die Zahl der Einwohner der Bundesrepublik Deutschland am 1.1.1990 (endliche Grundgesamtheit).
- Ein Würfelexperiment mit einem Würfel läßt sich beliebig oft unter den gleichen Bedingungen wiederholen (unendliche Grundgesamtheit).

Wird aus der Grundgesamtheit ein Element zufällig ausgewählt (d.h. jedes Element besitzt dieselbe Chance, ausgewählt zu werden) und der Wert des zu untersuchenden Merkmals gemessen, so kann der Wert x als Realisation einer Zufallsvariablen X aufgefaßt werden. Für ein Intervall I ist also $P(X \in I)$ die Wahrscheinlichkeit dafür, daß ein Element aus der Grundgesamtheit ausgewählt wird, dessen Merkmalswert in I liegt. Wir bezeichnen daher die Verteilung von X auch als Verteilung der Grundgesamtheit auf dem Merkmal X. Die Verteilungsparameter (Erwartungswert, Median, ...) und die Verteilungsfunktion der Grundgesamtheit charakterisieren die Zufallsvariable X. In den meisten Fällen sind diese Parameter jedoch unbekannt. Unsere Aufgabe ist es, Aussagen über diese Parameter zu machen.

Eine Möglichkeit zur Lösung dieser Aufgabe ist die Durchführung einer Totalerhebung, bei der bei jedem Objekt die Merkmalsausprägung gemessen wird. Allein aus finanziellen Gründen ist dies meistens nicht durchführbar. Eine Erhebung der Grundgesamtheit wird unsinnig, wenn mit der Untersuchung die Zerstörung des Objekts verbunden ist (Lebensdauer von Glühbirnen). Daher muß man sich oft damit begnügen, einige Objekte aus der Grundgesamtheit auszuwählen. Werden n Objekte herausgegriffen, so sprechen wir von einer *Stichprobe vom Umfang* n. Werden die n Objekte unabhängig voneinander gezogen, so daß jedes Element die gleiche Chance besitzt ausgewählt zu werden, sprechen wir von einer *Zufallsstichprobe*, die bei endlicher Grundgesamtheit einem Ziehen mit oder ohne Zurücklegen entspricht. In einer endlichen Grundgesamtheit ist bei einmaliger Ziehung die Wahrscheinlichkeit, daß das zufällig ausgewählte Element den Merkmalswert x besitzt, gleich der relativen Häufigkeit von x in der Grundgesamtheit. Weiter stimmt die aus allen x Werten der endlichen Grundgesamtheit gebildete empirische Verteilungsfunktion mit der Verteilungsfunktion der Grundgesamtheit überein.

7.2 Stichprobenfunktionen

Will man die durchschnittliche Kinderzahl der deutschen Familie feststellen, so kann man entweder eine Totalerhebung durchführen oder sich mit einer Stichprobe begnügen. In beiden Fällen wird das arithmetische Mittel berechnet. Dieses repräsentiert bei der Totalerhebung die mittlere Kinderzahl der Grundgesamtheit, bei der Stichprobe hingegen die mittlere Kinderzahl einer speziellen, aber zufällig ausgewählten Teilmenge der Grundgesamtheit. Der Stichprobenmittelwert kann sich deshalb vom exakten Mittelwert unterscheiden.

Bei wiederholter Stichprobenentnahme erhalten wir eine *Verteilung von Stichprobenmittelwerten*, die von der Verteilung der Grundgesamtheit abhängt. Die Analyse dieser Verteilung hilft uns, die Genauigkeit des Stichprobenverfahrens zu beurteilen bzw. Maßnahmen zur Verbesserung der Genauigkeit zu entwickeln.

Maßzahlen wie Erwartungswert und Varianz, die die Grundgesamtheit charakterisieren, werden als *Parameter* bezeichnet. Eine Zufallsvariable $Z = g(X_1, \ldots, X_n)$ (eine Funktion der Zufallsvariablen $X_1, \ldots, X_n$ der Stichprobe), heißt eine *Stichprobenfunktion* oder auch *Statistik*. Mit Hilfe von Statistiken kann auf die Parameter der Grundgesamtheit geschlossen werden. Dies ist Gegenstand der *Inferenzstatistik*.

Man kann Statistiken zu folgenden Zwecken benutzen:

1. Zur Schätzung von Parametern der Grundgesamtheit.
2. Zur Schätzung eines Intervalls, das mit einer vorgegebenen Sicherheitswahrscheinlichkeit den wahren Parameter überdeckt.
3. Zur Überprüfung einer Hypothese, ob ein Parameter in einer bestimmten Region liegt.
4. Zur Prognose.

Arithmetisches Mittel

Der Mittelwert einer Zufallsstichprobe (arithmetisches Mittel) wird mit $\bar{x}$ bezeichnet. Für $\bar{x}$ gilt:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad E(\bar{x}) = \mu, \quad \sigma^2(\bar{x}) = \frac{1}{n}\sigma^2(X).$$

Beweis: Die Zufallsvariablen X_i einer Zufallsstichprobe $(X_1 \ldots X_n)$ sind unabhängig voneinander und wie das Merkmal X verteilt. Daher gilt die Behauptung.

Ist die Grundgesamtheit $N \sim (\mu, \sigma^2)$ verteilt, so ist $\bar{x} \sim N(\mu, \sigma^2/n)$ verteilt.

Ist der Umfang n der Stichprobe hinreichend groß, so ist $\bar{x}$ annähernd normalverteilt. Dies gilt auch dann, wenn die Zufallsvariable X_i nicht normalverteilt ist (Zentraler Grenzwertsatz).

Ist eine Grundgesamtheit endlich, wird häufig eine Stichprobe *ohne* Zurücklegen erhoben. In diesem Fall sind die Zufallsvariablen X_i, die das Ergebnis der $i$ten Ziehung repräsentieren, nicht voneinander unabhängig. Besitzt die Grundgesamtheit N Elemente, so ist beim Ziehen ohne Zurücklegen

$$E(\bar{x}) = \mu, \ \sigma^2(\bar{x}) = \frac{\sigma^2(X)}{n} \cdot \frac{N-n}{N-1}.$$

Wegen $(N-n)/(N-1) < 1$ streut die Statistik $\bar{x}$ beim Ziehen ohne Zurücklegen weniger stark als beim Ziehen mit Zurücklegen, jedoch geht dieser Vorteil für große N wegen $\lim_{N\to\infty}(N-n)/(N-1) = 1$ verloren.

Beispiel: Das Durchschnittseinkommen einer Gruppe von unselbständig Erwerbstätigen beträgt 1900 DM im Monat mit Standardabweichung $\sigma = 200$DM. Man bestimme ein approximatives Intervall um μ, in dem mit Wahrscheinlichkeit 0.96 der Stichprobenmittelwert $\bar{x}$ einer Zufallsstichprobe vom Umfang $n = 400$ liegt: Wegen $n = 400 \geq 30$ ist $\bar{x}$ annähernd $N(1900, 100)$ verteilt. Gesucht ist c mit $P(1900 - c \leq \bar{x} \leq 1900 + c) = 0.96$.

$$\Phi\left(\frac{1900+c-\mu}{\sigma(\bar{x})}\right) - \Phi\left(\frac{1900-c-\mu}{\sigma(\bar{x})}\right) = 0.96 \Longleftrightarrow \Phi\left(\frac{c}{10}\right) - \Phi\left(\frac{-c}{10}\right) = 0.96$$

Aus der Normalverteilungstabelle erhält man $\Phi(2.06) = 0.98$. Daraus folgt: $\Phi(2.06) - \Phi(-2.06) = 0.96$ und somit $c = 10 \cdot 2.06 = 20.6$. Das gesuchte approximative Intervall ist daher $[1879.4, 1920.6]$.

Die relative Häufigkeit

Interessiert uns nur die Wahrscheinlichkeit π eines Ereignisses A und setzen wir $x_i = 1$, falls A im $i$ten Versuch realisiert wird, und $x_i = 0$ sonst, so erhält man als Spezialfall des arithmetischen Mittels die relative Häufigkeit eines Ereignisses A in einer Stichprobe vom Umfang n.

$$\tilde{p}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Der Erwartungswert von $\tilde{p}_n$ ist gleich der Wahrscheinlichkeit π von A:

$$E(\tilde{p}_n) = \pi$$

Da in einer Zufallsstichprobe jedes X_i die Varianz $\pi(1-\pi)$ besitzt, gilt:

$$\sigma^2(\tilde{p}_n) = \frac{\pi(1-\pi)}{n}.$$

Für $n\pi(1-\pi) \geq 10$ ist $\tilde{p}_n$ annähernd normalverteilt:

$$\tilde{p}_n \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

In einer endlichen Grundgesamtheit besitzt $\tilde{p}_n$ beim *Ziehen ohne Zurücklegen* den Erwartungswert π und die Varianz

$$\sigma^2(\tilde{p}_n) = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)$$

Stichprobenvarianz

Die Varianz einer Zufallsvariablen X wird aus einer Zufallsstichprobe durch die Stichprobenvarianz geschätzt:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Die Stichprobenvarianz s^2 besitzt in einer Zufallsstichprobe den Erwartungswert σ^2 (Beweis als Übung):

$$E(s^2) = \sigma^2$$

Ist die Grundgesamtheit normalverteilt, so ist der Ausdruck

$$(n-1)\cdot\frac{s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

χ^2 verteilt mit $m = n-1$ Freiheitsgraden (kurz χ^2_{n-1} verteilt).

Die χ^2 Verteilung besitzt eine positive Dichte $f(x)$ über $0 \leq x < \infty$ und hängt von einem Parameter m, $m = 1, 2, \ldots$ (Freiheitsgrade) ab. Sie besitzt den Erwartungswert $\mu = m$ und die Varianz $\sigma^2 = 2m$. Für eine χ^2_m verteilte Zufallsvariable X ist ab $m \geq 30$ der Ausdruck:

$\sqrt{2X} - \sqrt{2m-1}$ annähernd $N(0,1)$ verteilt.

Und für das α–Fraktil x_α der Verteilung von X gilt in diesem Fall:

$$x_\alpha \approx \frac{1}{2}(z_\alpha + \sqrt{2m-1})^2$$

Dabei ist z_α das α–Fraktil der $N(0,1)$ Verteilung.

Gewöhnliche Stichprobenmomente

Die gewöhnlichen Stichprobenmomente um 0

$$M_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

besitzen als Erwartungswert die gewöhnlichen Momente der Grundgesamtheit $E(X^k)$.

8 Punkt und Intervallschätzung

8.1 Punktschätzung

Eine Punktschätzung liegt vor, wenn aufgrund einer Stichprobe durch eine Statistik $g(X_1, \ldots, X_n)$ ein Schätzer $\hat{\vartheta}$ für einen unbekannten Parameter ϑ der Verteilung der Grundgesamtheit festgelegt wird. $\bar{x}$ und s^2 sind Punktschätzer für μ bzw. σ^2.

Beispiel: Der Median $x_{0.5}$ der Stichprobenwerte $x_1, \ldots, x_n$ ist ein Punktschätzer für den Median der Grundgesamtheit $x_{0.5}$. Er wird folgendermaßen berechnet: Zunächst werden die n Stichprobenwerte der Größe nach geordnet.

$$x_{[1]} \leq x_{[2]} \leq \ldots \leq x_{[n]}.$$

Dann ist $x_{0.5}$ definiert durch:

$$x_{0.5} = \begin{cases} x_{[\frac{n+1}{2}]} & \text{falls } n \text{ ungerade ist (der Wert in der Mitte)} \\ \frac{1}{2}(x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}) & \text{falls } n \text{ gerade ist (das artithmetische Mittel der beiden Werte in der Mitte)} \end{cases}$$

Ist die Verteilung einer Grundgesamtheit symmetrisch um μ, so ist $x_{0.5}$ auch ein inhaltlich sinnvoller Punktschätzer für den Erwartungswert μ.

Schätzer werden nach *Gütekriterien*, z. B. Erwartungstreue, Effizienz und Konsistenz, beurteilt.

Erwartungstreue (Unverzerrtheit)

Ein Punktschätzer $\hat{\vartheta}$ für einen Parameter ϑ heißt erwartungstreu, wenn gilt:

$$E(\hat{\vartheta}) = \vartheta$$

Beispiel:

$\hat{\vartheta} = \bar{x}$ ist erwartungstreu für μ .

$\hat{\vartheta} = s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ ist erwartungstreu für σ^2 .

$\hat{\vartheta} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ ist nicht erwartungstreu für σ^2 .

Effizienz

Es seien $\hat{\vartheta}_1$ und $\hat{\vartheta}_2$ erwartungstreu für ϑ. $\hat{\vartheta}_1$ ist wirksamer (effizienter) als $\hat{\vartheta}_2$, wenn er eine kleinere Varianz besitzt:

$$V(\hat{\vartheta}_1) < V(\hat{\vartheta}_2)$$

d.h. ein Schätzer ist umso wirksamer, je geringer seine Streuung ist.

Beispiel: In einer $N(\mu, \sigma^2)$ verteilten Grundgesamtheit sind sowohl $\bar{x}$ als auch $x_{0.5}$ erwartungstreue Schätzer für μ. Es gilt:

$$V(\bar{x}) = \frac{1}{n}\sigma^2 < V(x_{0.5}) = \frac{\pi}{2} \cdot \frac{1}{n}\sigma^2$$

Daher gilt: $\bar{x}$ ist wirksamer als $x_{0.5}$.

Mit Hilfe der sogenannten Informationsungleichung von Rao - Cramer kann man berechnen, wie groß die Varianz des wirksamsten aller unverzerrten Schätzer ist.

Konsistenz

$\hat{\vartheta}$ ist konsistent für ϑ, wenn gilt:

$$\lim_{n \to \infty} P(|\hat{\vartheta}_n - \vartheta| > \epsilon) = 0 \quad \text{für alle } \epsilon > 0$$

d.h. der Schätzer $\hat{\vartheta}$ konvergiert für $n \to \infty$ nach Wahrscheinlichkeit gegen den Wert ϑ.

8.2 Schätzverfahren

Momentenmethode

Lassen sich die unbekannten Parameter ϑ_i als Funktionen $g_i(m_1, \ldots, m_r)$ der (auch unbekannten) gewöhnlichen Momente $M_k = E(X^k)$ darstellen, so heißt die Stichprobenfunktion

$$\hat{\vartheta}_i = g_i(M_1, \ldots, M_r)$$

Momentenschätzer für ϑ_i, wobei

$$M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

die Stichprobenmomente sind. Die Momentenschätzer sind im allgemeinen nicht effizient.

Maximum-Likelihood-Methode

Wird eine Grundgesamtheit durch einen unbekannten Parameter ϑ charakterisiert, so hängt die Dichte vom unbekannten Parameter ϑ ab:

$f(x|\vartheta)$ im stetigen Fall

$p(j|\vartheta) = P(X = j|\vartheta)$ im diskreten Fall

Beispiel: Ist die Grundgesamtheit exponentialverteilt, so gilt:

$$f(x|\vartheta) = f(x|\lambda) = f(x) = \lambda e^{-\lambda x} \quad \text{für} \quad x \geq 0.$$

Ist die Grundgesamtheit poissonverteilt, so gilt:

$$p(x|\vartheta) = p(x|\lambda) = p_x = e^{-\lambda}\frac{\lambda^x}{x!}, \; x = 0, 1 \ldots.$$

Liegt eine unabhängige Zufallsstichprobe vom Umfang n vor, so besitzt $(X_1, \ldots, X_n)$ die Dichte

$f(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n)$ im stetigen Fall bzw.

$p(x_1, \ldots, x_n) = p_{x_1} \cdots p_{x_n}$ im diskreten Fall.

Bei der Likelihoodfunktion rückt der Parameter in den Vordergrund, die Werte der Stichprobe werden als gegeben aufgefaßt. Die Likelihoodfunktion stimmt mit der Dichte der Stichprobe überein. Im Gegensatz zur Dichte werden die Parameter als variabel und die Realisierungen x_i als fest aufgefaßt.

$$L(\vartheta) = L(\vartheta|x_1, \ldots, x_n) = \begin{cases} f(x_1|\vartheta) \cdots f(x_n|\vartheta) & \text{im stetigen Fall} \\ p(x_1|\vartheta) \cdots p(x_n|\vartheta) & \text{im diskreten Fall} \end{cases}$$

Likelihoodprinzip nach Birnbaum: Für das Stichprobenergebnis $(x_1, \ldots, x_n)$ wählt man denjenigen Wert $\hat{\vartheta}$ als Schätzwert für ϑ aus, für den die Likelihoodfunktion am größten ist. Der so konstruierte Schätzer heißt *Maximum-Likelihood-Schätzer.*

Im diskreten Fall ist für den so berechneten Wert $\hat{\vartheta}$ die Wahrscheinlichkeit für die Beobachtung $x_1, \ldots, x_n$ am größten, im stetigen Fall ist für $\hat{\vartheta}$ die Dichte $f(x_1, \ldots, x_n)$ an den beobachteten Werten am größten.

Beispiel: Eine Urne enthalte schwarze und weiße Kugeln mit Anteilen π und $1 - \pi$. Bei einer Zufallsstichprobe von n Kugeln aus der Urne werden k schwarze Kugeln gezogen. Die Anzahl X der schwarzen Kugeln ist $B_{n,\pi}$ verteilt. Daher ist die Likelihoodfunktion gegeben durch:

$$L(\pi) = \binom{n}{k} \cdot \pi^k (1 - \pi)^{n-k}$$

Der Wert π ist nun so zu bestimmen, daß $L(\pi)$ ein Maximum annimmt. Man beachte, daß bei der Bestimmung der Likelihood die Produktbildung der einzelnen Dichten der Zufallsvariablen berücksichtigt wird.

Bei der Bestimmung des Maximums geht man zweckmäßigerweise von der zur Basis e logarithmierten Likelihoodfunktion aus, um Produkte zu Summen zu transformieren, die einfacher differenziert werden können. Da der Logarithmus eine streng monotone Funktion ist, nimmt die logarithmierte Likelihoodfunktion an der gleichen Stelle wie die ursprüngliche Funktion das Maximum an. Man erhält:

$$\begin{aligned} \log L(\pi) &= \log \binom{n}{k} + \log \pi^k + \log (1 - \pi)^{n-k} \\ &= \log \binom{n}{k} + k \cdot \log \pi + (n - k) \log (1 - \pi) \end{aligned}$$

Ableiten nach π und Nullsetzen der Ableitung ergibt:

$$\frac{d \log L(\pi)}{d\pi} = 0 + k \cdot \frac{1}{\pi} + (n - k) \cdot \frac{1}{1 - \pi}(-1) = 0$$

Daraus folgt: $\hat{\pi} = \frac{k}{n}$. Die relative Häufigkeit $\tilde{p}_n = \frac{k}{n}$ ist daher der Maximum-Likelihood-Schätzer für π (es handelt sich tatsächlich um ein Maximum). $L(\pi)$ nimmt für $n = 10$ und $k = 3$ folgende Werte an:

π	0.1	0.2	0.3	0.4	0.5	0.7
$L(\pi)$	0.06	0.20	0.27	0.21	0.12	0.01

Beispiel: Der Maximum-Likelihood-Schätzer für den Parameter μ einer $\mathcal{N}(\mu, \sigma^2)$-Verteilung soll bestimmt werden, wobei σ^2 als bekannt vorausgesetzt wird. Die Likelihood-Funktion ist aufgrund der Unabhängigkeit der Stichprobe:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

Die Log-Likelihood-Funktion ist:

$$\log L(\mu) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x_i - \mu)^2}{2\sigma^2}$$

Die erste Ableitung nach μ ist:

$$\frac{\partial \log L(\mu)}{\partial \mu} = \sum_{i=1}^{n} -\frac{1}{2\sigma^2} \cdot 2 \cdot (x_i - \mu) \cdot (-1) = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2}$$

Nullsetzen und Auflösen nach μ ergibt:

$$\begin{aligned}
\frac{1}{\sigma^2} \sum_{i=1}^{n} x_i - \hat{\mu} &= 0 \\
\left(\sum_{i=1}^{n} x_i\right) - n\hat{\mu} &= 0 \\
n\hat{\mu} &= \sum_{i=1}^{n} x_i \\
\hat{\mu} &= \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

Methode der kleinsten Quadrate

Beispiel: Aus den Beobachtungswerten $x_1, \ldots, x_n$ soll der Erwartungswert μ der Grundgesamtheit so geschätzt werden, daß die Summe der Quadrate der Abstände der Beobachtungen x_i von μ durch

$$\sum_{i=1}^{n} (x_i - \mu)^2 \rightarrow \text{Min}_{\mu \in \mathcal{R}}$$

minimiert wird. Ableiten der Summe nach μ und Nullsetzen der Ableitung ergibt

$$\frac{d}{d\mu} \sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} 2(x_i - \mu) \cdot (-1) = 0.$$

Daraus folgt:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\bar{x}$ ist also der Kleinste-Quadrate-Schätzer für μ.

8.3 Intervallschätzung

Bei der Punktschätzung ist im allgemeinen die Wahrscheinlichkeit, daß der Schätzer $\hat{\vartheta}$ mit dem wahren Parameter ϑ übereinstimmt, gleich Null. Daher konstruieren wir ein Intervall, das zu vorgegebener Wahrscheinlichkeit $1-\alpha$ den wahren, aber unbekannten Parameter ϑ einschließt.

Definition: Unter einem *Konfidenzintervall* für ϑ zur Sicherheit $S = 1-\alpha,\ 0 < \alpha < 1$, verstehen wir ein Intervall $[\underline{\vartheta},\ \overline{\vartheta}]$, dessen Grenzen $\underline{\vartheta}$ und $\overline{\vartheta}$ Zufallsvariablen sind, für die gilt:

$$P(\underline{\vartheta} \leq \vartheta \leq \overline{\vartheta}) = 1-\alpha$$

Die *Sicherheit* $S = 1-\alpha$ heißt auch *Konfidenzniveau* oder *Vertrauenswahrscheinlichkeit*. Es gibt *ein-* und *zweiseitige* Konfidenzintervalle. Die Länge $2d = \overline{\vartheta} - \underline{\vartheta}$ heißt *Genauigkeit* der Konfidenzschätzung in einem zweiseitigen Konfidenzintervall.

Unter einem *einseitigen Konfidenzintervall* zur Sicherheit $S = 1-\alpha$ verstehen wir ein Zufallsintervall $(-\infty,\ \overline{\vartheta}]$ bzw. $[\underline{\vartheta}, \infty)$ mit

$$P(\underline{\vartheta} \leq \vartheta) = 1-\alpha \ \text{ bzw. } \ P(\vartheta \leq \overline{\vartheta}) = 1-\alpha\,.$$

Die Intervallgrenzen $\underline{\vartheta}$ und $\overline{\vartheta}$ sind also identisch mit den α- und $1-\alpha$-Quantilen der zugrundeliegenden Verteilung. Bei zweiseitigen Konfidenzintervallen, entsprechen hingegen sie den $\frac{\alpha}{2}$- und $1-\frac{\alpha}{2}$-Quantilen.

Ein zweiseitiges Konfidenzintervall $[\underline{\vartheta},\ \overline{\vartheta}]$ heißt *symmetrisch* um ϑ, falls gilt:

$$P(\underline{\vartheta} \leq \vartheta) = P(\overline{\vartheta} \geq \vartheta)$$

Konfidenzintervall für μ bei normalverteilter Grundgesamtheit und bekannter Varianz σ^2

Ein symmetrisches Konfidenzintervall für μ zu $S = 1-\alpha$ wird folgendermaßen konstruiert. Da jedes Element der Grundgesamtheit $\mathcal{N}(\mu, \sigma^2)$ verteilt ist, ist $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ verteilt. Daher ist die *Gaußstatistik*

$$\frac{\bar{x}-\mu}{\sigma} \cdot \sqrt{n} \sim N(0,1)$$

normalverteilt. Somit gilt:

$$P\left[z_{\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \leq z_{1-\frac{\alpha}{2}}\right] = 1-\alpha,$$

wobei $z_{\alpha/2}$ bzw. $z_{1-\alpha/2}$ die $\alpha/2$ bzw. $1-\alpha/2$ Quantile der $N(0,1)$ Verteilung sind. Daraus folgt:

$$z_{\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \leq z_{1-\alpha/2}$$

$$\iff z_{\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \text{ und } \frac{\bar{x}-\mu}{\sigma}\sqrt{n} \leq z_{1-\frac{\alpha}{2}}$$

$$\iff \mu \leq \bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \text{ und } \mu \geq \bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Wegen $z_{\alpha/2} = -z_{1-\alpha/2}$ erhält man das zweiseitige Konfidenzintervall zur Sicherheit $S = 1-\alpha$

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}})$$

Die *einseitigen Konfidenzintervalle* zur $S = 1-\alpha$ sind

$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha},\ \infty\right) \text{ und } \left(-\infty,\ \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\right]$$

Beispiel: Um den durchschnittlichen Benzinverbrauch pro 100 km eines neuen Modells zu ermitteln, läßt eine Automobilfirma mit 25 Versuchswagen Testfahrten durchführen. Die Firma interessiert sich für das Konfidenzintervall zu $S = 0.95$ für den durchschnittlichen Benzinverbrauch μ pro 100 km. Es wird angenommen, daß der Verbrauch normalverteilt ist mit $\sigma = 0.9\ell/100\text{km}$. Der Durchschnittsverbrauch aller 25 Testwagen war $9.1\ell/100\text{km}$. Daraus folgt:

$$\underline{\mu} = \bar{x} - z_{0.975}\frac{\sigma}{\sqrt{n}} = 9.1 - 1.96 \cdot \frac{0.9}{5} = 8.75$$

$$\overline{\mu} = \bar{x} + z_{0.975}\frac{\sigma}{\sqrt{n}} = 9.1 + 1.96 \cdot \frac{0.9}{5} = 9.45$$

Konfidenzintervall für μ bei normalverteilter Grundgesamtheit und unbekannter Varianz

Die unbekannte Varianz σ^2 wird durch die Stichprobenvarianz

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

geschätzt. Bei normalverteilter Grundgesamtheit ist die *t-Statistik*

$$t^\star = \frac{\bar{x}-\mu}{s}\sqrt{n}$$

t verteilt mit $n-1$ Freiheitsgraden.

Die *t Verteilung* (auch SStudent-Verteilung") besitzt eine Dichte $f(x)$ über $-\infty < x < \infty$ und einem Parameter $m = 1, 2, \ldots$, der als Freiheitsgrad bezeichnet und durch df (degrees of freedom) abgekürzt wird. Die t Verteilung ist symmetrisch um den Erwartungswert $\mu = 0$ und besitzt die Varianz $\sigma^2 = m/(m-2)$. Ist $m \geq 30$, so kann die t Verteilung durch die Normalverteilung angenähert werden. *Das Konfidenzintervall für* μ zur Sicherheit $S = 1-\alpha$ ist

$$\left(\bar{x} - t_{1-\frac{\alpha}{2};n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\frac{\alpha}{2};n-1} \cdot \frac{s}{\sqrt{n}}\right).$$

$t_{1-\frac{\alpha}{2};n-1}$ ist das $1-\alpha/2$ Quantil der t Verteilung mit $n-1$ Freiheitsgraden.

Die einseitigen Konfidenzintervalle sind

$$\left(\bar{x} - t_{1-\alpha;n-1} \cdot \frac{s}{\sqrt{n}}, \infty\right) \text{ bzw. } \left(-\infty, \bar{x} + t_{1-\alpha;n-1} \cdot \frac{s}{\sqrt{n}}\right)$$

Konfidenzintervall für μ bei großen Stichproben

Ist der Umfang n der Stichprobe ≥ 30, so ist $\bar{x}$ annähernd normalverteilt. Bei bekannter Varianz σ^2 lautet das Konfidenzintervall für μ zu $S = 1 - \alpha$

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

wobei $z_{1-\alpha/2}$ das $1 - \alpha/2$ Quantil der $N(0, 1)$ Verteilung ist.

Ist die Varianz σ^2 unbekannt, so kann σ^2 durch s^2 geschätzt werden, da s^2 für σ^2 konsistent ist. In diesem Fall ist das Konfidenzintervall für μ zu $S = 1 - \alpha$:

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right)$$

Konfidenzintervall für π eines Ereignisses A mit $P(A) = \pi$

Ist π die Wahrscheinlichkeit eines Ereignisses A, so ist $\tilde{p}_n$ (relative Häufigkeit von A in einer Stichprobe vom Umfang n) eine Statistik für π.

Für $n\pi(1 - \pi) \geq 10$ ist $\tilde{p}$ annähernd $N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$ verteilt.

Das Konfidenzintervall für π zur Sicherheit $S = 1 - \alpha$ ist

$$\left(\tilde{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} \leq \pi \leq \tilde{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right).$$

Bemerkung: Wie an den vorangegangenen Beispielen zu sehen ist, steigt bei gleichbleibendem Konfidenzniveau die Präzision der Schätzung mit wachsendem Stichprobenumfang. Wenn der Stichprobenumfang vervierfacht wird, verdoppelt sich die Präzision.

Konfidenzintervall für σ^2 bei normalverteilter Grundgesamtheit

Die Statistik

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

ist ein erwartungstreuer Schätzer für σ^2. Der Ausdruck

$$\frac{(n-1) \cdot s^2}{\sigma^2}$$

ist χ^2 verteilt mit $n-1$ Freiheitsgraden. Das zweiseitige Konfidenzintervall für σ^2 zu $S = 1-\alpha$ ist

$$\left(\frac{(n-1)\cdot s^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \leq \sigma^2 \leq \frac{(n-1)\cdot s^2}{\chi^2_{\frac{\alpha}{2};n-1}} \right)$$

Beispiel: Ein Papierband wird von einer Maschine in ca. 24 cm lange Stücke geschnitten. Um die Streuung der Länge zu ermitteln, wurde an 24 zufällig ausgewählten Stücken die Länge nachgemessen. Man erhielt $s^2 = 0.88$ cm^2. Es soll das Konfidenzintervall für σ^2 zu $S = 0.99$ berechnet werden:

$$\underline{\sigma}^2 = \frac{(n-1)s^2}{\chi^2_{0.995;23}} = \frac{23\cdot 0.88}{44.2} = 0.46$$

$$\overline{\sigma}^2 = \frac{(n-1)s^2}{\chi^2_{0.005;23}} = \frac{23\cdot 0.88}{9.26} = 2.19$$

9 Signifikanztests

Bisher haben wir Schätzverfahren (Punkt- und Intervallschätzungen) für unbekannte Parameter einer Grundgesamtheit behandelt. In vielen konkreten Problemstellungen muß jedoch zwischen zwei Hypothesen entschieden werden.

Beispiele:

- Hypothese: Ein pharmazeutisches Präperat hat die beabsichtigte Wirkung. Gegenhypothese: Es hat nicht die gewünschte Wirkung.
- Hypothese: Ein neues Verfahren zur Herstellung von Glühbirnen bewirkt eine Verlängerung der Lebensdauer gegenüber einem alten Verfahren. Gegenhypothese: Die nach dem neuen Verfahren hergestellten Birnen besitzen keine längere Lebensdauer.

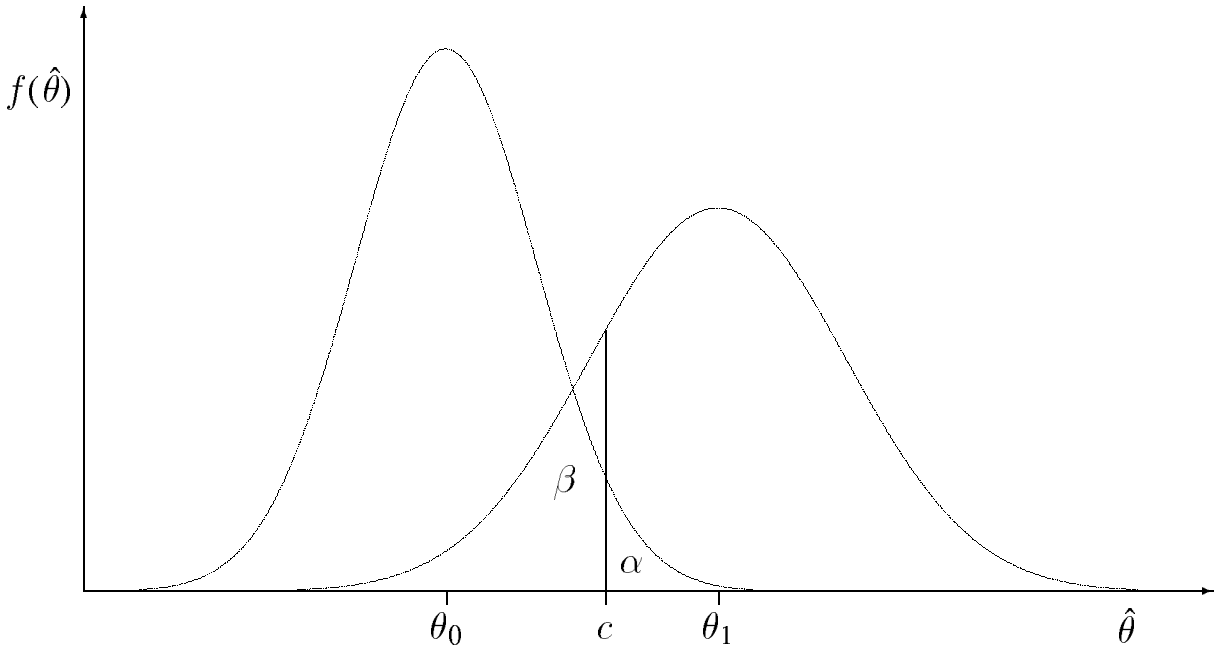
9.1 Aufbau von Signifikanztests

Eine *statistische Hypothese* ist eine Vermutung über die Wahrscheinlichkeitsverteilung der Grundgesamtheit, die in den meisten Anwendungen als Vermutung über einen unbekannten Parameter ϑ formuliert wird. Die zu untersuchende Hypothese wird als *Nullhypothese* H_0 bezeichnet, während die relevante Alternative als *Alternativhypothese* H_1 bezeichnet wird. Eine Entscheidungsvorschrift, die aufgrund der Werte einer Stichprobe angibt, wann für H_0 bzw. H_1 zu entscheiden ist, heißt *Signifikanztest*. Der *Annahmebereich* ist die Menge aller möglichen Datenkonstellationen, die zur Entscheidung für die Nullhypothese H_0 führen. Der *kritische Bereich* oder *Ablehnungsbereich* ist die Menge der Datenkonstellationen, die zur Annahme von H_1 führen.

Beispiel: Ist der Parameter ϑ einer Grundgesamtheit unbekannt und sind die möglichen Werte von ϑ nur ϑ_0 und ϑ_1, so lauten die Hypothesen:

$$H_0 : \vartheta = \vartheta_0\,,\quad H_1 : \vartheta = \vartheta_1$$

In der folgenden Abbildung sind die Dichten einer Statistik $\hat{\vartheta}$ für ϑ dargestellt (links die Dichte, wenn H_0 wahr ist).



Um eine Entscheidungsvorschrift festzulegen, wird die Menge aller Datenkonstellationen in E_0 und E_1 zerlegt:

$$E_0 = \left\{(x_1, \ldots, x_n) | \hat{\vartheta}(x_1, \ldots, x_n) < c\right\}$$

$$E_1 = \left\{(x_1, \ldots, x_n) | \hat{\vartheta}(x_1, \ldots, x_n) \geq c\right\}$$

Die einfachste Entscheidungsregel lautet wie folgt:

Wenn das Ereignis E_0 eintritt, entscheide für die Nullhypothese H_0.

Wenn E_1 eintritt, verwerfe H_0.

9.2 Fehler erster Art und zweiter Art

Die bei einem Test möglichen Fehler sind:

- Fehler erster Art (α Fehler): Man trifft eine Entscheidung für H_1, obwohl H_0 richtig ist.
- Fehler zweiter Art (β Fehler): Man trifft eine Entscheidung für H_0, obwohl H_1 richtig ist.

Im obigen Beispiel ist die Wahrscheinlichkeit für einen Fehler erster Art gleich der Wahrscheinlichkeit des Eintreffens von E_1, wenn H_0 wahr ist.

$$\alpha = P(E_1|H_0) = P(\hat{\vartheta} \geq c|H_0)$$

Die Wahrscheinlichkeit für einen Fehler zweiter Art ist gleich der Wahrscheinlichkeit des Eintreffens von E_0, wenn H_1 wahr ist.

$$\beta = P(E_0|H_1) = P(\hat{\vartheta} < c|H_1)$$

Die Wahrscheinlichkeiten für richtige Entscheidungen sind

$$1-\alpha = P(E_0|H_0) = P(\hat{\vartheta} < c|H_0), \text{ falls } H_0 \text{ wahr ist}$$

$$1-\beta = P(E_1|H_1) = P(\hat{\vartheta} \geq c|H_1), \text{ falls } H_1 \text{ wahr ist}$$

Die Größe c teilt die Verteilung von $\hat{\vartheta}$ im Verhältnis $(1-\alpha)$ zu α, wenn ϑ_0 der wahre Parameter ist. Dies impliziert die Verwendung der Null-Hypothese zur Konstruktion der kritischen Schwelle c.

Beispiel: Ein Unternehmen erhält eine große Warensendung vorgefertigter Teile, die entweder 5 % oder 10 % Ausschuß enthalten kann, je nachdem auf welcher Anlage des Zulieferers diese Teile hergestellt wurden. Die Frage ist, auf welcher Anlage die Warensendung produziert wurde. Daher muß zwischen $H_0 : \pi = 0.05$ und $H_1 : \pi = 0.1$ entschieden werden. Die Entscheidungsregel lautet: Ist der relative Anteil $\tilde{p}$ des Ausschusses in einer Stichprobe vom Umfang n größer/gleich c, so wird H_0 abgelehnt; ist p kleiner als c, so wird H_0 beibehalten.

Die Wahrscheinlichkeit für einen Fehler erster Art ist:

$$\alpha = P(\tilde{p} \geq c|\pi = 0.05) = \sum_{k \geq n \cdot c}^{n} \binom{n}{k} 0.05^k \cdot 0.95^{n-k}$$

Die Wahrscheinlichkeit für einen Fehler zweiter Art ist:

$$\beta = P(\tilde{p} < c|\pi = 0.1) = \sum_{k < n \cdot c} \binom{n}{k} 0.1^k \cdot 0.9^{n-k}$$

Entschließt sich das Unternehmen, die Lieferung zurückzuweisen, wenn in einer Stichprobe vom Umfang 20 der Ausschußanteil größer ist als $1/9$, so erhält man:

$$\alpha = P(\tilde{p} > \frac{1}{9}|\pi = 0.05) = \sum_{k=3}^{20} \binom{20}{k} \cdot 0.05^k \cdot 0.95^{20-k} = 1 - \sum_{k=0}^{2} \binom{20}{k} \cdot 0.05^k \cdot 0.95^{20-k}$$

$$= 1 - 0.3585 - 20 \cdot 0.05 \cdot 0.3774 - 190 \cdot 0.0025 \cdot 0.3972 = 1 - 0.9245 = 0.0755$$

$$\beta = P(\tilde{p} \leq \frac{1}{9}|\pi = 0.1) = \sum_{k=0}^{2} \binom{20}{k} 0.1^k \cdot 0.9^{20-k} = 0.1216 + 20 \cdot 0.1 \cdot 0.1351 + 190 \cdot 0.01 \cdot 0.1501 = 0.6769$$

Die Wahrscheinlichkeit, die Lieferung zurückzuweisen, obwohl $\pi = 0.05$ ist, ist gering, nämlich 0.0755. Hingegen ist die Wahrscheinlichkeit, die Lieferung anzunehmen, obwohl der relative Ausschußanteil 10 % beträgt, relativ groß (0.6769).

Bisher lag nur der Fall vor, daß sowohl H_0 als auch H_1 nur einen möglichen Wert des unbekannten Parameters umfassen. Gehören nun zu H_0 und H_1 mehrere mögliche Parameterwerte, so lassen sich die Wahrscheinlichkeiten für den Fehler erster und zweiter Art nicht mehr eindeutig bestimmen.

Durchläuft a den Bereich der möglichen Parameterwerte ϑ, so heißt die Funktion

$$G(a) = P(E_1|\vartheta = a) = P(\text{Ablehnung von } H_0 | \text{ der wahre Wert von } \vartheta \text{ ist } a)$$

Gütefunktion eines Tests. Die Gütefunktion gibt in Abhängigkeit von a die Wahrscheinlichkeit für die Annahme von H_1 an.

$$\alpha = \sup_{a \in H_0} G(a \leq \alpha)$$

ist die maximale Wahrscheinlichkeit, daß für H_1 entschieden wird, obwohl H_0 richtig ist. α heißt *Signifikanzniveau* oder *Irrtumswahrscheinlichkeit* des Tests.

$$\beta = \sup_{a \in H_1} (1 - G(a))$$

ist die maximale Wahrscheinlichkeit, daß für H_0 entschieden wird, obwohl H_1 richtig ist.

Wird eine Statistik $\hat{\vartheta}$ zur Festlegung des Annahme- und des Ablehnungsbereichs benutzt, so heißt $\hat{\vartheta}$ eine *Teststatistik* oder auch *Testfunktion*. Eine Teststatistik $\hat{\vartheta}(x_1, \ldots, x_n)$ heißt signifikant auf dem Niveau α, wenn das Signifikanzniveau α ist und die beobachteten Werte aus der Stichprobe zur Ablehnung von H_0 führen.

Man beachte: Die Beibehaltung der Nullhypothese bedeutet, daß die Testgröße $\hat{\vartheta}(x_1, \ldots, x_n)$ keinen Hinweis auf die Gültigkeit der Alternativhypothese H_1 liefert. Die Beibehaltung der Hypothese H_0 bedeutet weder, daß H_1 sicher falsch ist, noch daß H_1 mit Wahrscheinlichkeit $1 - \alpha$ falsch ist.

Beispiel: Einer normalverteilten Grundgesamtheit mit bekannter Varianz σ^2 wird eine Stichprobe vom Umfang n entnommen. Es soll geprüft werden, ob der Mittelwert μ aus der Grundgesamtheit größer ist als μ_0 .

$$H_0 : \mu \geq \mu_0 \text{ gegen } H_1 : \mu < \mu_0$$

Als Statistik für den Test wird der Stichprobenmittelwert $\bar{x}$ verwendet. Die Entscheidungsregel lautet: Ist $\bar{x} < c$, so wird H_1 angenommen, andernfalls wird H_0 beibehalten. Das Signifikanzniveau des Tests wird über die Gütefunktion bestimmt. Ist die Grundgesamtheit $N(\hat{\mu}, \sigma^2)$ verteilt, so ist $\bar{x} \sim N(\hat{\mu}, \sigma^2/n)$ verteilt. Daher ist

$$G(\hat{\mu}) \quad = \quad P(\bar{x} \leq c | \mu = \hat{\mu}) = P\left(\frac{\bar{x} - \hat{\mu}}{\sigma/\sqrt{n}} \leq \frac{c - \hat{\mu}}{\sigma/\sqrt{n}} \middle| \mu = \hat{\mu} \right) = \Phi\left(\frac{c - \hat{\mu}}{\sigma/\sqrt{n}} \right)$$

Für alle $\hat{\mu} > \mu_0$ gilt:

$$\frac{c - \hat{\mu}}{\sigma/\sqrt{n}} \leq \frac{c - \mu_0}{\sigma/\sqrt{n}} \text{ und somit } \Phi\left(\frac{c - \mu_0}{\sigma/\sqrt{n}} \right) \geq \Phi\left(\frac{c - \hat{\mu}}{\sigma/\sqrt{n}} \right),$$

Daraus folgt:

$$\alpha = \max_{\hat{\mu} \geq \mu_0} G(\hat{\mu}) = \Phi\left(\frac{c - \mu_0}{\sigma/\sqrt{n}} \right)$$

Der Schwellenwert c wird zu einem *vorgegebenem* Signifikanzniveau ermittelt: Wegen $\alpha = \Phi\left(\frac{c - \mu_0}{\sigma} \sqrt{n} \right)$ ist $\frac{c - \mu_0}{\sigma} \sqrt{n}$ gleich dem α Quantil z_α der $N(0, 1)$ Verteilung.

$$z_\alpha = \frac{c - \mu_0}{\sigma} \sqrt{n} \Longrightarrow c = \mu_0 + \frac{\sigma}{\sqrt{n}} \cdot z_\alpha = \mu - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$$

Die Entscheidungsregel im Test für $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ zum Signifikanzniveau α lautet bei normalverteilter Grundgesamtheit: Gilt

$$\bar{x} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha},$$

so wird H_0 abgelehnt, andernfalls wird H_0 beibehalten.

Bemerkung:

1. Sinnvoll wird ein statistischer Test nur dann verwendet, wenn das Ergebnis der vorliegenden Stichprobe der Null-Hypothese widerspricht. In diesem Fall ist zu prüfen, ob der Widerspruch nur zufällig ist oder ob er hinreicht, um die Null-Hypothese bei vorgegebener Irrtumswahrscheinlichkeit zu verwerfen.
2. Mit wachsendem Stichprobenumfang sinkt der Fehler zweiter Art, d.h. die Schärfe des Tests nimmt zu.

Auswahl der Hypothesen: Aufgrund von Bemerkung 1 sollte die für den Forscher zu überprüfende Hypothese die H_1-Hypothese sein. Im Zweifelsfalle, sind beide möglichen Hypothesen gegeneinander abzuwägen und diejenige als H_0-Hypothese zu wählen, die im Falle einer falschen Entscheidung den größeren Schaden anrichten würde (*worst-case-Prinzip*).

Beipiel: Es ist bekannt, daß eine bewährtes Medikament eine Krankheit in 90% aller Fälle zuverlässig bekämpft. Ein neues Medikament wird an 1 000 Probanden getestet und schneidet mit einer Zuverlässigkeit von $\tilde{p} = 0.99$ ab. Welcher Test sollte zu einem Signifikanzniveau von $\alpha = 0.05$ gewählt werden? Um eine solche Aufgabenstellung adäquat beurteilen zu können, sollte die H_0-Hypothese $H_0 : \pi \leq 0.9$ lauten, da in diesem Falle das neue Medikament schlechtere Resultate hervorbrächte als das alte. Der α Fehler (das neue Medikament wird als besser getestet, obwohl es in Wirklichkeit schlechter ist) kann so durch den Forscher kontrolliert werden, während der β Fehler (das neue Medikament wird als schlechter getestet, obwohl es in Wirklichkeit besser ist) nicht kontrolliert werden kann.

Definition: Eine statistische Fragestellung heißt *einseitig*, wenn sie die Form

1. $H_0 : \vartheta \leq \vartheta_0 \, , \; H_1 : \vartheta > \vartheta_0$
2. $H_0 : \vartheta \geq \vartheta_0 \, , \; H_1 : \vartheta < \vartheta_0$

besitzt. Eine statistische Fragestellung heißt *zweiseitig*, wenn sie folgende Form besitzt:

$H_0 : \vartheta = \vartheta_0 \, , \; H_1 : \vartheta \neq \vartheta_0$.

9.3 Signifikanztests für spezielle Fragestellungen

Test für μ bei normalverteilter Grundgesamtheit und bekannter Varianz (Gaußtest)

$H_0 : \mu \leq \mu_0 \, , \; H_1 : \mu > \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma} \geq z_{1-\alpha}$$

Gütefunktion: $G(a) = \Phi\left((a - \mu_0) \cdot \frac{\sqrt{n}}{\sigma} - z_{1-\alpha}\right)$.

$H_0 : \mu \geq \mu_0 \, , \; H_1 : \mu < \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n} \cdot \frac{\bar{x} - \mu_0}{\sigma} \leq -z_{1-\alpha}$$

Gütefunktion: $G(a) = \Phi\left((\mu_0 - a) \cdot \frac{\sqrt{n}}{\sigma} - z_{1-\alpha}\right)$

$H_0 : \mu = \mu_0 \, , \; H_1 : \mu \neq \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n} \cdot \frac{|\bar{x} - \mu_0|}{\sigma} \geq z_{1-\frac{\alpha}{2}}$$

Gütefunktion: $G(a) = \Phi\left((\mu_0 - a)\frac{\sqrt{n}}{\sigma} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left((a-\mu_0)\frac{\sqrt{n}}{\sigma} - z_{1-\frac{\alpha}{2}}\right)$

Test für μ bei normalverteilter Grundgesamtheit und unbekannter Varianz

$H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{\bar{x} - \mu_0}{s} \geq t_{1-\alpha;n-1}.$$

Gütefunktion: $G(a) = F\left((a-\mu_0)\cdot\frac{\sqrt{n}}{s} - t_{1-\alpha;n-1}\right)$

Hier ist $F(x)$ die Verteilungsfunktion der t Verteilung.

$H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{\bar{x} - \mu_0}{s} \leq -t_{1-\alpha;n-1}$$

Gütefunktion: $G(a) = F\left((\mu_0-a)\cdot\frac{\sqrt{n}}{s} - t_{1-\alpha;n-1}\right)$

$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{|\bar{x} - \mu_0|}{s} \geq t_{1-\frac{\alpha}{2};n-1}.$$

Gütefunktion: $G(a) = F\left((\mu_0 - a)\frac{\sqrt{n}}{s} - t_{1-\frac{\alpha}{2};n-1}\right) + F\left((a-\mu_0)\frac{\sqrt{n}}{s} - t_{1-\frac{\alpha}{2};n-1}\right)$

Test für μ bei großen Stichproben

Gegeben sei eine Zufallsstichprobe $x_i, i = 1, \ldots, n$ einer Zufallsvariablen mit einer beliebigen Verteilung und unbekannter Varianz σ^2, so daß gilt $X \sim (\mu, \sigma^2)$. Ist die Stichprobe $n \geq 30$, lassen sich approximative Signifikanztests wegen des zentralen Grenzwertsatzes analog zum Gaußtest durchführen:

$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{\bar{x} - \mu_0}{s} \geq z_{1-\alpha}$$

$H_0 : \mu \geq \mu_0, \quad H_1 : \mu < \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{\bar{x} - \mu_0}{s} \leq -z_{1-\alpha}$$

$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\sqrt{n}\frac{|\bar{x} - \mu_0|}{s} \geq z_{1-\alpha/2}$$

Tests für Mittelwertunterschiede bei unabhängigen Stichproben

Gegeben seien zwei voneinander unabhängig gezogene Stichproben $x_i, i = 1, \ldots, n_1$ der Zufallsvariablen $X \sim (\mu_1, \sigma_1^2)$ und $y_j, j = 1, \ldots, n_2$ der Zufallsvariablen $Y \sim (\mu_2, \sigma_2^2)$. Von Interesse ist die Hypothese $H_0 : \mu_1 = \mu_2$, d.h. die beiden Erwartungswerte sind gleich, gegen $H_1 : \mu_1 \neq \mu_2$. Je nach Annahmen über σ_1^2, σ_2^2 und die Verteilungsfunktion von X und Y erhält man unterschiedliche Teststatistiken, die hier mit D bezeichnet werden.

1. $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2), \sigma_1^2$ und σ_2^2 bekannt.

$$D = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2}}, \quad s_p^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

   Ist $|D| \geq z_{1-\alpha/2}$, wird H_0 mit Irrtumswahrscheinlichkeit α verworfen.

2. $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2), \sigma_1^2 = \sigma_2^2$ aber unbekannt.
   Wir definieren zunächst die Hilfsgröße s_p^2 (p für pooled), die ein Schätzer für die Varianz von $\bar{x} - \bar{y}$ ist.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}$$

$$D = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2}}$$

   Ist $|D| \geq t_{1-\alpha/2;n_1+n_2-2}$, wird H_0 mit Irrtumswahrscheinlichkeit α verworfen.

3. $X \sim (\mu_1, \sigma_1^2), Y \sim (\mu_2, \sigma_2^2), \sigma_1^2$ und σ_2^2 unbekannt, $n_1 \geq 30, n_2 \geq 30$.

$$D = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2}}, \quad s_p^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

   Ist $|D| \geq z_{1-\alpha/2}$, wird H_0 mit Irrtumswahrscheinlichkeit α verworfen.

Beispiel: Gegeben seien zwei Medikamente, die beide den Blutdruck senken. Zu prüfen ist, ob die beiden Medikamente gleichwertig sind, d.h. ob im Mittel der Unterschied 0 ist. Aus zwei unabhängigen klinischen Untersuchungen erhält man die Daten:

- Erste Klinik: $n_1 = 35, \quad \bar{x} = 147, \quad s_1^2 = 225$
- Zweite Klinik: $n_2 = 42, \quad \bar{y} = 137, \quad s_2^2 = 256$

Die Verteilung der Grundgesamtheit ist unbekannt. Die $H_0 : \mu_1 = \mu_2$ wird mit der Teststatistik

$$D = \frac{147 - 137}{\sqrt{225/35 + 256/42}} = 2.826 \geq z_{1-0.025} = 1.96$$

überprüft und mit Irrtumswahrscheinlichkeit $\alpha = 0.05$ zugunsten von $H_1 : \mu_1 \neq \mu_2$ verworfen.

Test für σ^2 bei normalverteilter Grundgesamtheit

Bei normalverteilter Grundgesamtheit ist

$$\frac{(n-1)s^2}{\sigma^2}$$

χ^2 verteilt mit $n-1$ Freiheitsgraden.

$H_0 : \sigma^2 \leq \sigma_0^2$, $H_1 : \sigma^2 > \sigma_0^2$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\alpha;n-1}.$$

$H_0 : \sigma^2 \geq \sigma_0^2$, $H_1 : \sigma^2 < \sigma_0^2$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\alpha;n-1}.$$

$H_0 : \sigma^2 = \sigma_0^2$, $H_1 : \sigma^2 \neq \sigma_0^2$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\frac{\alpha}{2};n-1} \text{ oder } \frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\frac{\alpha}{2};n-1}.$$

Beispiel: Um die Streuung der Abfüllmengen der auf einer neuen Anlage gefüllten Flaschen zu überprüfen, soll eine Stichprobe vom Umfang $n = 24$ gezogen werden. Zum Signifikanzniveau $\alpha = 0.01$ soll getestet werden, ob $\sigma^2 = 3$ cm^6 ist. $H_0 : \sigma^2 \geq 3$, $H_1 : \sigma^2 < 3$

Die Entscheidungsregel lautet: Die H_0 wird abgelehnt, wenn gilt:

$$\frac{23 \cdot s^2}{3} \leq \chi^2_{0.01;23}$$

In einer Stichprobe vom Umfang 24 erhält man für die Abfüllmenge die Statistiken:

$$\bar{x} = 501\text{cm}^3 \ , \ s^2 = 1.7\text{cm}^6$$

$$\chi^2_{0.01;23} = 10.2 \ , \ \frac{23 \cdot s^2}{3} = 13.03$$

Wegen $13.03 > \chi^2_{0.01;23} = 10.2$ kann H_0: $\sigma^2 \geq 3$ mit der Irrtumswahrscheinlichkeit $\alpha = 0.01$ nicht verworfen werden.

Test für eine unbekannte Wahrscheinlichkeit

Die relative Häufigkeit $\tilde{p}$ eines Ereignisses mit Wahrscheinlichkeit π ist eine Statistik für π. Für $n\pi(1-\pi) \geq 10$ gilt:

$$\tilde{p}_n \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

$H_0 : \pi \leq \pi_0$, $H_1 : \pi > \pi_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\tilde{p_n} \geq \pi_0 + z_{1-\alpha}\sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$H_0 : \pi \geq \pi_0$, $H_1 : \pi < \pi_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$\tilde{p}_n \leq \pi_0 - z_{1-\alpha}\sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$H_0 : \pi = \pi_0$, $H_1 : \pi \neq \pi_0$. H_0 wird zum Testniveau α verworfen, wenn gilt:

$$|\tilde{p}_n - \pi_0| \geq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Beispiel: Um zu erfahren, ob die Partei A bei einer Wahl die 5% Hürde überspringt, werden 1490 zufällig ausgewählte Personen nach ihrer Parteipräferenz gefragt. Es soll die Entscheidungsregel für $H_0 : \pi_0 \leq 0.05$ gegen $H_1 : \pi_0 > 0.05$ zum Niveau $\alpha \geq 0.05$ angegeben werden. π ist der Stimmanteil der Partei A bei der Wahl. Wegen $n\pi_0(1-\pi_0) = 1490 \cdot 0.05 \cdot 0.95 = 70.8 \geq 10$ kann der oben angeführte Test angewandt werden. H_0 ist zu verwerfen, wenn

$$\tilde{p}_n \geq 0.05 + z_{0.95} \cdot \sqrt{\frac{0.05 \cdot 0.95}{1490}} = 0.05 + 0.0093 = 0.0593$$

Die Stichprobe ergab einen Prozentanteil für A von 5.47. Die Hypothese, daß A unter der 5% Hürde bleibt, kann daher zum Signifikanzniveau $\alpha = 0.05$ nicht verworfen werden.

χ^2 Anpassungstest

Zu überprüfen ist, ob die diskrete unbekannte Verteilung einer Grundgesamtheit mit einer gegebenen hypothetischen Verteilung übereinstimmt. $\pi_i = P_0(E_i)$ seien die Wahrscheinlichkeiten für die durch Zerlegung gebildeten Ereignisse E_i, $i = 1, \ldots, r$ unter der Nullhypothese. Sind h_i die absoluten Häufigkeiten der Ereignisse E_i in einer Stichprobe vom Umfang n, so wird die Hypothese H_0, daß die Grundgesamtheit die Verteilung P_0 besitzt, mit der Irrtumswahrscheinlichkeit α abgelehnt, falls gilt:

$$\chi^{2*} = \sum_{i=1}^{r} \frac{(h_i - \hat{h}_i)^2}{\hat{h}_i} \geq \chi^2_{1-\alpha;r-1}$$

mit

$$\hat{h}_i = n\pi_i, \quad i = 1, \ldots, r$$

Voraussetzung: $\hat{h}_i \geq 5$ für alle $i = 1, \ldots, r$. χ^{2*} ist ein Maß für die Abweichung der beobachteten Verteilung von der hypothetischen Verteilung. Ein hoher χ^{2*}-Wert weist daher darauf hin, daß die Verteilungen nicht übereinstimmen.

Beispiel: Eine Lebensversicherung hat folgende Häufigkeit von Versicherungsfällen im Monat festgestellt:

i	0	1	2	3	≥ 4	Σ
h_i	45	70	48	21	16	200

Wir untersuchen zum Signifikanzniveau $\alpha = 0.05$, ob die Anzahl X der Versicherungsfälle pro Monat eine Poissonverteilung mit $\lambda = 1.5$ besitzt.

i	0	1	2	3	≥ 4	Σ
π_i	0.223	0.335	0.251	0.126	0.065	1
$\hat{h}_i = n\pi_i$	44.6	67	50.2	25.2	13	200

Die Werte für π_i stammen aus der Tabelle der Poissonverteilung mit $\lambda = 1.5$. Die Testfunktion ist

$$\chi^{2*} = \sum_i \frac{(h_i - \hat{h}_i)^2}{\hat{h}_i} = 0.0036 + 0.1343 + 0.0964 + 0.7 + 0.6923 = 1.626$$

Für die Teststatistik gilt: $\chi^{2*} < \chi^2_{0.95;4}$. Daher wird H_0: $\lambda = 1.5$ beibehalten.

10 Korrelation und Regression

10.1 Einfache Korrelation

Eine Hauptaufgabe jeder wissenschaftlichen Arbeit ist die Ermittlung von Zusammenhängen zwischen Variablen. Betrachtet man zunächst nur zwei metrische Merkmale, ist man an Richtung und Stärke des Zusammenhangs interessiert. Zu einer derartigen Charakterisierung läßt sich - unter später behandelten Beschränkungen - der Produktmomentkorrelationskoeffizient von Bravais–Pearson heranziehen. Der Korrelationskoeffizient in einer Grundgesamtheit wurde bereits behandelt. Im folgenden wird auf die Schätzung dieses Koeffizienten aus einer Zufallsstichprobe eingegangen.

Beispiel zur Korrelation: Ein Verband von Handelsfirmen ermittelt von 10 der ihm angeschlossenen Firmen den jährlichen Lagerumschlag X und den durchschnittlichen Kalkulationsaufschlag Y (in % des Einkaufspreises).

Firmen-Nr.	1	2	3	4	5	6	7	8	9	10	n = 10
X	8.5	7.8	7.5	6.2	6.5	6.0	5.6	4.6	4.0	3.3	$\bar{x} = 6.0$
Y	18.0	20.0	20.0	25.0	29.0	31.0	33.0	37.0	43.0	44.0	$\bar{y} = 30.0$

An den Firmen (den Elementen der Stichprobe) werden zwei Messungen vorgenommen. Der Betrachtung liegen also 10 Zahlenpaare $\{(x_i, y_i) : i = 1, \ldots, 10\}$ zugrunde. Diese Zahlenpaare lassen sich als 10 stochastisch unabhängige Realisierungen der Zufallsvariablen (X, Y) auffassen, für die eine gemeinsame Verteilung $F(x, y)$ mit Erwartungswerten μ_x und μ_y, Varianzen σ_x^2 und σ_y^2, Kovarianz σ_{xy} und Korrelation ρ_{xy} angenommen wird. Zur Erinnerung werden noch einmal die Definitionen von Kovarianz und Korrelationskoeffizient angegeben:

$$\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

Erwartungstreue und konsistente Schätzer für σ_{xy} und ρ_{xy} sind:

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

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

Der Korrelationskoeffizient der Stichprobe wird anhand des folgenden Schemas berechnet:

Nr.	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$	$(y - \bar{y})^2$
1	2.5	-12	6.25	-30.0	144
2	1.8	-10	3.24	-18.0	100
3	1.5	-10	2.25	-15.0	100
4	0.2	- 5	0.04	- 1.0	25
5	0.5	1	0.25	- 0.5	1
6	0.0	1	0.00	0.0	1
7	-0.4	3	0.16	-1.2	9
8	-1.4	7	1.96	-9.8	49
9	-2.0	13	4.00	-26.0	169
10	-2.7	14	7.29	-37.8	196
$\sum$	0	0	25.44	-139.3	794

$$s_{xy} = -15.444, \quad s_x^2 = 2.827, \quad s_y^2 = 88.222 \quad .$$

$$r_{xy} = \frac{-139.3}{\sqrt{25.44 \cdot 794}} = -0.980$$

Die Variablen Lagerumschlag und Kalkulationsaufschlag sind hoch negativ korreliert. Je öfter ein Lagerumschlag stattfindet, desto geringer ist der Kalkulationsaufschlag und umgekehrt. Man beachte, daß sich aus diesem Ergebnis keine Richtung eines kausalen Zusammenhanges ableiten läßt.

Wichtig sind die folgenden Eigenschaften des Korrelationskoeffizienten:

- $-1 \leq r \leq +1$, d.h. r ist normiert.
- Besteht zwischen y und x ein exakter linearer Zusammenhang ($y_i = a + bx_i$), so gilt:

$$r = \begin{cases} 1 & \text{wenn} \quad b > 0 \quad \text{(gleichsinnig)} \\ -1 & \text{wenn} \quad b < 0 \quad \text{(gegensinnig)} \end{cases}$$

- Sind x und y voneinander statistisch unabhängig, so ist ihre Kovarianz null und damit auch der Korrelationskoeffizient. Da der Korrelationskoeffizient eine Maßzahl des *linearen* Zusammenhangs ist, kann umgekehrt aus $r = 0$ nicht geschlossen werden, daß x und y statistisch unabhängig sind.
- Ein durch Korrelation nachgewiesener statistischer Zusammenhang zwischen Merkmalen wird vielfach als ursächlicher Zusammenhang verstanden. Einfache Beispiele zeigen, daß dies nicht ohne weiteres sinnvoll ist:

- zwischen der Anzahl besetzter Storchennester und der jährlichen Geburtenziffer hat man eine positive Korrelation festgestellt,
- zwischen der Anzahl der in England verkauften Radiogeräte und der Länge des Vornamens des Präsidenten der USA wurde eine starke Korrelation festgestellt (Zeitraum: 1923 – 1942),
- zwischen der Anzahl der Fernsehgenehmigungen und der Anzahl von hospitalisierten Geisteskranken hat man eine positive Korrelation festgestellt.

Bei der Deutung eines Korrelationskoeffizienten als Hinweis auf kausale Zusammenhänge muß man mindestens folgende *Interpretationsmodelle* in Betracht ziehen:

- Einseitige Steuerung: x wirkt auf y : $x \longrightarrow y$. Beispiel: Vererbung.
- Wechselseitige Steuerung: $x \longleftrightarrow y$. Beispiele: Kontakt fördert die Sympathie, Sympathie fördert den Kontakt; Löhne beeinflussen die Preise, Preise die Löhne.
- Drittseitige Steuerung: $z \longrightarrow y$ und $z \longrightarrow x$. Beispiel: Bis zum 20. Lebensjahr nimmt sowohl das Körpergewicht x als auch die Intelligenz y zu. In einer hinsichtlich des Lebensalters heterogenen Stichprobe erscheinen die schwereren Individuen als die intelligenteren (Scheinkorrelation; mittels des partiellen Korrelationskoeffizienten versucht man, solche drittseitige Steuerungen aufzuklären). Die oben angeführten Beispiele sind typische Fälle von Scheinkorrelation.
- Komplexe Steuerung: y hängt von p Variablen $(x_1, \ldots, x_p)$ ab. Abhängigkeiten dieser Art versucht man durch die multiple Korrelation zu erfassen.

Unter der Annahme einer bivariaten Normalverteilung für (X, Y) lassen sich folgende Tests zum Niveau α für den Korrelationskoeffizienten der Grundgesamtheit konstruieren:

Test der speziellen Hypothese H_0 : $\rho = 0$ (Unabhängigkeit) gegen

$$\left.\begin{array}{lll} H_1: \quad \rho > 0\ , & \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} & \geq \quad t_{n-2;1-\alpha} \\ H_1: \quad \rho < 0\ , & \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} & \leq \quad -t_{n-2;1-\alpha} \\ H_1: \quad \rho \neq 0\ , & \dfrac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} & \geq \quad t_{n-2;1-\alpha/2} \end{array}\right\} \Longrightarrow H_0 \text{ wird zum Testniveau } \alpha \text{ abgelehnt.}$$

Test der allgemeinen Hypothese $\rho = \rho_0$ gegen die folgenden Gegenhypothesen. Entwickle die Hilfsgrößen

$$z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right), \quad z_0 = \left[\frac{1}{2}\log\left(\frac{1+\rho_0}{1-\rho_0}\right)\right] + \frac{\rho_0}{2(n-1)}$$

Treffen folgende Aussagen zu,

$$\left.\begin{array}{llll} H_1: & \rho > \rho_0, & (z-z_0)\sqrt{n-3} & \geq \quad z_{1-\alpha} \\ H_1: & \rho < \rho_0, & (z-z_0)\sqrt{n-3} & \leq \quad z_{\alpha} \\ H_1: & \rho \neq \rho_0, & |z-z_0|\sqrt{n-3} & \geq \quad z_{1-\alpha/2} \end{array}\right\} \Longrightarrow H_0 \text{ wird zum Testniveau } \alpha \text{ abgelehnt.}$$

Beispiel: In einer Untersuchung an 67 Schülern wurde zwischen Intelligenzquotient und Mathematiknoten eine Korrelation von 0.38 festgestellt. Aus einer früheren Totalerhebung ist bekannt, daß der Korrelationskoeffizient ρ_0 der Grundgesamtheit den Wert 0.45 besitzt. Es ist zu überprüfen, ob der Korrelationskoeffizient der vorliegenden

Stichprobe mit der Hypothese $H_0 : \rho = 0.45$ in Einklang steht ($\alpha = 0.05$). Die Gegenhypothese ist $H_1 : \rho \neq 0.45$. Die Hilfsgrößen sind:

$$z_0 = \frac{1}{2} \log \frac{1.45}{0.55} + \frac{0.45}{132} = 0.4881 \quad ,$$

$$z = \frac{1}{2} \log \frac{1.38}{0.62} = 0.4 \quad ,$$

$$z^\star = |z - z_0| \sqrt{n-3} = 0.0881 \cdot 8 = 0.7048 \quad .$$

Da $z_{1-\alpha/2} = 1.96$, wird H_0 beibehalten.

10.2 Klassisches Regressionsmodell

Bei zahlreichen ökonomischen Fragestellungen untersuchen wir die Abhängigkeit einer Variablen Y von Kontroll- oder Einflußvariablen $X_1, X_2 \ldots X_p$. Ein klassisches, aber irreführendes Beispiel aus der Makroökonomie ist der vermutete Zusammenhang zwischen Konsum C_i und Volkseinkommen Y_i, der in einer vereinfachten Form durch die lineare Beziehung $C_i = a + bY_i + e_i, i = 1, \ldots n$ mit i als Index der Beobachtungsperiode dargestellt wird. Allgemein beschreiben wir einen derartigen Zusammenhang durch

$$y_i = f(x_{i1}, x_{i2}, \ldots x_{ip}) + e_i, i = 1, \ldots, n \quad .$$

e_i bezeichnet die Abweichung oder Fehler zwischem dem Funktionswert $f(x_{i1}, \ldots, x_{ip})$ und der Realisation y_i der abhängigen Variablen Y, die als Zufallsvariable interpretiert wird. Die parametrische Form der Regressionsfunktion wird durch ökonomische Überlegungen a priori festgelegt. Die Parameter sind dann direkt mit Hilfe der ökonomischen Theorie interpretierbar. Die Regressionsfunktion entspricht einer ex ante Gleichung der volkswirtschaftlichen Theorie, während die Koeffizienten (Parameter) der Regressionsfunktion ex post aus Daten geschätzt werden. Beispiele sind:

- Einfache lineare Funktion (eindimensionale Regression):

  $$f(x_i) = b_0 + b_1 x_i$$

- Allgemeine lineare Funktion (mehrdimensionale Regression):

  $$f(x_{i1} \ldots x_{ip}) = b_0 + b_1 x_{i1} + \ldots b_p x_{ip}$$

- Einfache Cobb–Douglas Produktionsfunktion, wobei A_i das Arbeitsvolumen, K_i den Kapitaleinsatz und α die Substitutionselastizität bedeutet. Die Cobb–Douglas Funktion ist eine homogene Funktion erster Ordnung mit konstanten Substitutionselastizitäten.

  $$f(A_i, K_i) = A_i^\alpha K_i^{1-\alpha}, \quad 0 \leq \alpha \leq 1.$$

- Verallgemeinerte Cobb–Douglas Produktionsfunktion. Die einfache Cobb–Douglas Produktionsfunktion wird einerseits durch die Einführung eines Koeffizienten γ für nicht durch Arbeit und Kapital erfaßte Faktoren (z. B. Rationalisierungsgrad der Volkswirtschaft), andererseits durch die Aufgabe der Annahme der Homogenität ersten Grades erweitert ($\alpha + \beta \neq 1$ ist zulässig).

  $$f(A_i, K_i) = \gamma \cdot A_i^\alpha K_i^\beta$$

- Rückführung der verallgemeinerten Cobb–Douglas Funktion auf ein lineares Modell:

  $$\log f(A_i, K_i) = \log \gamma + \alpha \log A_i + \beta \log K_i$$

Am letzten Beispiel erkennt man, daß sich nichtlineare Funktionen zum Teil durch Transformationen in lineare Funktionen überführen lassen. Ist dies nicht möglich, lassen sich die Parameterschätzer von nichtlinearen Modellen häufig durch wiederholte Anwendung transformierter linearer Regressionen berechnen. Dies erfordert Existenz und Stetigkeit der beiden ersten Ableitungen der nichtlinearen Regressionsfunktion bezüglich aller Parameter. Daher konzentrieren wir uns in diesem Skript auf die Schätzung der Parameter $b_0, \ldots b_p$ in linearen Funktionen:

$$y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + e_i$$

Statistische Aufgabe ist die Schätzung von $b_0, b_1, \ldots b_p$ und der Varianz σ^2 des Fehlers e_i, die Überprüfung auf signifikante Abweichungen von vorgegebenen Hypothesen über diese Parameter sowie die Beurteilung der Güte der Schätzung. Als Schätzverfahren verwenden wir den Kleinsten Quadrate-Schätzer und das Maximum Likelihood Schätzprinzip.

Zur Vereinfachung der Notation stellen wir das Modell für $i = 1, \ldots, n$ in Matrixform dar:

$$y = Xb + e$$

wobei y und e $n \times 1$ Vektoren, b ein $(p+1) \times 1$ Vektor und X eine $n \times (p+1)$ Matrix ist. Im einzelnen gilt:

$$y^T = (y_1, \ldots, y_n)$$

$$e^T = (e_1, \ldots, e_n)$$

$$b^T = (b_0, b_1, \ldots, b_p)$$

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix}$$

Die Matrix X wird als Datenmatrix der unabhängigen oder exogenen Variablen bezeichnet.

Zum Lösen der statistischen Aufgaben treffen wir zunächst folgende Annahmen (Klassisches Regressionsmodell):

1. X ist exogen und X ist nicht stochastisch.
2. $Rang(X) = p + 1$ für alle $n \in \mathcal{N}$ mit $n \geq p + 1$, d.h. X hat vollen Spaltenrang.
3. $\lim_{n \to \infty} \frac{1}{n} X^T X =: Q$ existiert und ist regulär.
4. $E(e) = 0$.
5. $V(e) = E(ee^T) = \sigma^2 I_n$ mit $\sigma^2 > 0$, wobei I_n die $n \times n$ Einheitsmatrix ist.
6. $e \sim \mathcal{N}(0, \sigma^2 I_n)$

Die explizite Verteilungsannahme 6 ist nur in Verbindung mit dem ML-Prinzip bzw. mit der Konstruktion exakter Tests und Konfidenzintervalle notwendig, während Annahme 3 die Grundlage für die Konsistenz der Schätzverfahren liefert. Zur Ableitung des Kleinsten Quadrate Schätzers genügt die Annahme 2. Der Kleinste Quadrate Schätzer der Parameter b ist erwartungstreu, wenn Annahme 4 zusätzlich erfüllt ist. Wir leiten zunächst den KQ Schätzer ab. Um alle vorhin angeführten Aufgaben lösen zu können, werden alle Annahmen 1) bis 5) angenommen.

Grundlage der KQ-Schätzung ist die Minimierung der Funktion:

$$M(b) = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i1} + \ldots b_p x_{ip}))^2 \longrightarrow \min_{b \in \mathcal{R}^{p+1}}$$

In Matrixschreibweise gilt:

$$M(b) = (y - Xb)^T(y - Xb) = e^T e \longrightarrow \min_{b \in \mathcal{R}^{p+1}}$$

Notwendige Bedingung für ein Minimum von $M(b)$ an der Stelle $\hat{b}$ ist, daß der Vektor der ersten partiellen Ableitungen nach b gleich 0 ist:

$$\left.\frac{\partial M(\hat{b})}{\partial b}\right|_{b=\hat{b}} = 0 \quad .$$

Die einzelnen ersten Ableitungen von $M(b)$ nach b_j werden 0 gesetzt:

$$\left.\frac{\partial M(b)}{\partial b_j}\right|_{b=\hat{b}} = 2\sum_{i=1}^{n}(y_i - (\hat{b}_0 + \hat{b}_1 x_{i1} \ldots + \hat{b}_p x_{ip})(-x_{ij})) = 0$$

Umformungen liefern die $p + 1$ Normalgleichungen NG_j, $j = 0, \ldots, p$:

$$\sum_{i=1}^{n} y_i x_{ij} = \hat{b}_0 \sum_{i=1}^{n} x_{ij} + \hat{b}_1 \sum_{i=1}^{n} x_{ij} x_{i1} \ldots + \hat{b}_p \sum_{i=1}^{n} x_{ij} x_{ip}$$

In kompakter Matrixschreibweise lautet diese Gleichung:

$$X^T y = (X^T X)\hat{b}$$

Aus Annahme 2 folgt die Invertierbarkeit von $X^T X$ und somit:

$$\hat{b} = (X^T X)^{-1} X^T y$$

$\hat{b}$ heißt Kleinster Quadrate Schätzer für b. Aus $\hat{b}$ lassen sich folgende Größen berechnen:

$\hat{y} = X\hat{b}$ lineare Prädiktoren (geschätzte Werte)

$\hat{e} = y - \hat{y}$ Residuen (geschätzte Fehler)

Als erwartungstreuer Varianzschätzer s^2_{KQ} für σ^2 wird in Verbindung mit dem KQ-Schätzer der Ausdruck

$$s^2_{KQ} = \frac{1}{n - (p+1)} \cdot \sum_{i=1}^{n} \hat{e}_i^2 == \frac{1}{n - (p+1)} \hat{e}^T \hat{e}$$

berechnet.

Beispiel zur Regression: Abhängigkeit der Importmengen von Bruttonationalprodukt und Preisentwicklung in England.

y_i = Mengenindex der Importe nach England zu konstanten Preisen des Jahres 1948.
x_{i1} = Bruttonationalprodukt zu konstanten Preisen von 1948.
x_{i2} = Quotient des Preisindexes für Importe und des allgemeinen Preisindexes.

Jahr	y_i	x_{i1}	x_{i2}	$\hat{y}_i$	$\hat{e}_i$
1948	100	100	100	98.49	1.51
1949	106	104	99	103.83	2.17
1950	107	106	110	107.80	-0.80
1951	120	111	126	116.44	3.56
1952	110	111	113	114.97	-4.97
1953	116	115	103	119.28	-3.28
1954	123	120	102	125.99	-2.99
1955	133	124	103	131.56	1.44
1956	137	126	98	133.72	3.28

Ein lineares Modell für diese ökonomische Beziehung sieht nun so aus:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i \quad .$$

Gesucht sind b_0, b_1 und b_2 sowie die Varianz des Fehlers. Die Matrix der exogenen Variablen ist nun gegeben durch:

$$X = \begin{pmatrix} 1 & 100 & 100 \\ 1 & 104 & 99 \\ 1 & 106 & 110 \\ 1 & 111 & 126 \\ 1 & 111 & 113 \\ 1 & 115 & 103 \\ 1 & 120 & 102 \\ 1 & 124 & 103 \\ 1 & 126 & 98 \end{pmatrix}$$

Die inverse Matrix, die zur Auflösung der Normalgleichungen benötigt wird, ist dann:

$$(X^T X)^{-1} = \begin{pmatrix} 44.796100 & -0.208234 & -0.19957 \\ -0.208234 & 0.001586 & 0.000274 \\ -0.19957 & 0.000274 & 0.001591 \end{pmatrix}$$

Der Vektor der Schätzer $\hat{b} = (X^T X)^{-1} X^T y$ ist:

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} -49.329 \\ 1.364 \\ 0.114 \end{pmatrix}$$

Die prognostizierten Werte für y_i sind nun:

$$\hat{y}_i = -49.329 + 1.364 x_{i1} + 0.114 x_{i2}$$

Die Eigenschaften des KQ Schätzers lassen sich bei Gültigkeit der obigen Annahmen durch folgenden *Satz* zusammenfassen:

1. Erwartungstreue von $\hat{b}$:

   $$E(\hat{b}) = b$$

2. Erwartungstreue von s^2_{KQ}:

   $$E(s^2_{KQ}) = \sigma^2$$

3. Varianz-Kovarianzmatrix von $\hat{b}$:

   $$V(\hat{b}) = E(\hat{b} - b)(\hat{b} - b)^T = \sigma^2(X^TX)^{-1} = \sigma^2(c_{jk})_{j,k=0,\ldots,p}$$

   Dabei ist c_{jk} das j,kte Element von $C = (X^TX)^{-1}$.

4. Konsistenz von $\hat{b}$:

   Für alle $\epsilon > 0$ gilt: $\lim_{n\to\infty} P(|\hat{b}_j - b_j| \leq \epsilon) = 1$

5. Konsistenz von s^2_{KQ}:

   Für alle $\epsilon > 0$ gilt: $\lim_{n\to\infty} P(|s^2_{KQ} - \sigma^2| \leq \epsilon) = 1$

6. Asymptotische Normalverteilung:

   $\dfrac{(\hat{b}_j - b_j)}{\sqrt{s^2_{KQ}c_{jj}}}$ ist asymptotisch N(0,1) verteilt.

7. Theorem von Gauß-Markov: Innerhalb der Klasse der erwartungstreuen und in y linearen Schätzer ist der KQS effizient, d.h. er besitzt unter den Schätzern, die sich in der Form $\tilde{b} = Ly + d$ darstellen lassen, die kleinstmögliche Varianz:

   $$var(\hat{b}_j) \leq var(\tilde{b}_j) \quad \text{für} \quad E(\tilde{b}_j) = b_j \quad .$$

Exemplarisch beweisen wir die Punkte 1, 3 und 4.

Beweis zu 1: Da X nicht stochastisch ist, gilt:

$$\begin{aligned} E(\hat{b}) &= E\left[(X^TX)^{-1}X^Ty\right] \\ &= E\left[(X^TX)^{-1}X^T(Xb + e)\right] \\ &= E\left[(X^TX)^{-1}(X^TX)b + (X^TX)^{-1}X^Te\right] \\ &= b + (X^TX)^{-1}X^TEe \\ &= b \end{aligned}$$

Beweis zu 3:

$$\begin{aligned} E(\hat{b} - b)(\hat{b} - b)^T &= E(X^TX)^{-1}X^Tee^TX(X^TX)^{-1} \\ &= (X^TX)^{-1}X^TE(ee^T)X(X^TX)^{-1} \\ &= (X^TX)^{-1}X^T\sigma^2I_nX(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1} \end{aligned}$$

Beweis zu 4:

$$E(\hat{b}_j^{(n)} - b_j)^2 = \sigma^2 c_{jj}^{(n)}$$

Die Ungleichung von Tschebysscheff liefert:

$$P(|\hat{b}_j^{(n)} - b_j| > \epsilon) \leq \frac{\sigma^2 c_{jj}^{(n)}}{\epsilon^2}$$

Wegen Annahme 3 ist $\lim_{n\to\infty}(\frac{1}{n}X^TX)^{-1}$ endlich. Daraus folgt:

$$\lim_{n\to\infty} s^2 c_{jj}^{(n)} = \lim_{n\to\infty} \underbrace{\frac{\sigma^2}{n}}_{\to 0} \cdot \underbrace{n \cdot (X^TX)^{-1}}_{<\infty} = 0$$

Dies impliziert die obige Behauptung.

Zur Bestimmung der Güte der Regression, d.h. wie gut die beobachteten Werte von y durch das Modell angepaßt werden, benötigt man ein auf $[0, 1]$ normiertes Maß. Geeignet hierfür ist der quadrierte Korrelationskoeffizient $R^2_{y\hat{y}}$ zwischen den beobachteten Werten y_i und den aus der Regression geschätzten Werten $\hat{y}_i = x_i^T\hat{b}$. Es kann gezeigt werden, daß $R^2_{y\hat{y}} = R^2_{y\cdot x_1 \ldots x_p}$ (wenn $x_{i0} = 1$) identisch ist mit der normierten Quadratsumme SSR/SST, die wir aus der folgenden Streuungszerlegung mit der Notation

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{Gesamtquadratsumme, sum of squares total,}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{Fehlerquadratsumme, sum of squares of error,}$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \quad \text{erklärte Quadratsumme, sum of squares residual,}$$

erhalten:

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

Die Berechnung von R^2 kann daher alternativ durch

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\hat{e}^T\hat{e}}{SST}$$

erfolgen. Der Beweis der Streuungszerlegung erfolgt in zwei Schritten. Zunächst zeigt man, daß $\hat{e}^TX = 0$ ist, d.h. Residuen und Regressoren sind orthogonal.

$$e^TX = (y - X\hat{b})^TX = y^TX - y^TX(X^TX)^{-1}X^TX = 0$$

Aus $e^TX = 0$ und $x_{i0} = 1$ folgt die Behauptung

$$\sum_{i=1}^{n} e_i = 0$$

Im zweiten Schritt wird die Gesamtquadratsumme erweitert und das Ergebnis des ersten Schritts verwendet:

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

Zu zeigen ist, daß der erste Summand gleich 0 ist.

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

Wiederum ist zu zeigen, daß der letzte Summand gleich 0 ist.

$$\sum_{i=1}^{n}e_i\hat{y}_i=\sum_{i=1}^{n}e_i\sum_{j=0}^{p}x_{ij}b_j=\sum_{j=0}^{p}b_j(\sum_{i=1}^{n}e_ix_{ij})\sum_{j=0}^{p}b_j(e^TX_{.j})=0$$

Zur Konstruktion von Konfidenzintervallen und Tests benötigen wir (bei kleinen Stichproben) die Annahme der Normalverteilung (Annahme 6). Aus dieser Annahme folgt:

- $\hat{b}_j \sim \mathcal{N}(b_j, \sigma^2 c_{jj}), \quad j=0,\ldots p$.
- Wird σ^2 durch s^2_{KQ} ersetzt, gilt $(\hat{b}_j-b_j)/\sqrt{s^2_{KQ}c_{jj}} \sim t_{n-p-1}$.

Daher erhalten wir für den ersten Fall (σ^2 bekannt) das zweiseitige Konfidenzintervall:

$$P(\hat{b}_j - z_{1-\alpha/2}\sqrt{\sigma^2 c_{jj}} \le b_j \le \hat{b}_j + z_{1-\alpha/2}\sqrt{\sigma^2 c_{jj}}) = 1-\alpha$$

Für den zweiten Fall (σ^2 durch s^2_{KQ} geschätzt) lautet das Konfidenzintervall:

$$P(\hat{b}_j - t_{n-(p+1);1-\alpha/2}\sqrt{s^2_{KQ}c_{jj}} \le b_j \le \hat{b}_j + t_{n-(p+1);1-\alpha/2}\sqrt{s^2_{KQ}c_{jj}}) = 1-\alpha$$

Die Konstruktion von einseitigen Konfidenzintervallen und allgemeinen Teststatistiken erfolgt in gleicher Weise wie beim Mittelwert $\bar{x}$. Exemplarisch konstruieren wir einen einseitigen Signifikanztest für einen einzelnen Koeffizienten b_j:

$H_0: b_j \le b_{j0}$ gegen $H_1: b_j > b_{j0}$. H_0 wird zum Testniveau α abgelehnt, wenn für die Teststatistik $t^\star$ gilt:

$$t^\star = \frac{\hat{b}_j - b_{j0}}{\sqrt{s^2_{KQ}c_{jj}}} > t_{n-(p+1);1-\alpha}$$

$H_0: b_j \ge b_{j0}$ gegen $H_1: b_j < b_{j0}$. H_0 wird zum Testniveau α abgelehnt, wenn für die Teststatistik $t^\star$ gilt:

$$t^\star = \frac{\hat{b}_j - b_{j0}}{\sqrt{s^2_{KQ}c_{jj}}} < -t_{n-(p+1);1-\alpha}$$

$H_0: b_j = b_{j0}$ gegen $H_1: b_j \ne b_{j0}$. H_0 wird zum Testniveau α abgelehnt, wenn für die Teststatistik $t^\star$ gilt:

$$t^\star = \frac{|\hat{b}_j - b_{j0}|}{\sqrt{s^2_{KQ}c_{jj}}} > t_{n-(p+1);1-\frac{\alpha}{2}}$$

Da ein mehrdimensionales Modell vorliegt, ist es von Interesse, Hypothesen über Submodelle, definiert durch Teilmengen von $\{b_0 \ldots b_p\}$ zu testen. Ohne Beschränkung der Allgemeinheit seien $b_{k+1}, \ldots b_p$ die zu testenden Parameter mit $H_0 : b_{k+1} = b_{k+2} = \ldots = b_p = 0$ gegen $H_1 : b_i \neq 0$ mindestens für ein $i \in \{k+1, \ldots p\}$.

Zur Berechnung der Teststatistik benutzen wir als Hilfsgrößen die Größen b, x, e, SST, SSE des ursprünglichen Modells, das $b_0, \ldots b_p$ enthält. Die Größen $b^{(1)}$, $x^{(1)}$, $e^{(1)}$, $SSR^{(1)}$, $SSE^{(1)}$ sind analog den obigen Bezeichnungen definiert, aber sie enthalten nur die unabhängigen Variablen $x_0, x_1, \ldots x_k$. Der Superindex (1) bezeichnet das entsprechende Submodell mit Parametern $b_0, \ldots b_k$. Unter Gültigkeit von H_0 sind folgende Größen χ^2 verteilt (Beweis in der mathematischen Statistik):

$$\begin{aligned} \frac{SSE}{\sigma^2} &\sim \chi^2_{n-(p+1)} \\ \frac{SSE^{(1)}}{\sigma^2} &\sim \chi^2_{n-(k+1)} \\ \frac{(SSE^{(1)} - SSE)}{\sigma^2} &\sim \chi^2_{p-k} \end{aligned}$$

Es läßt sich zeigen, daß der erste und der dritte Term stochastisch unabhängig sind, so daß aus der Standardtheorie der mathematischen Statistik folgt, daß die Teststatistik

$$F^\star = \frac{(SSE^{(1)} - SSE)/(p-k)}{SSE/(n-(p+1))} \sim F(p-k, n-(p+1))$$

unter H_0 einer F Verteilung mit $(p-k, n-(p+1))$ Freiheitsgraden folgt. Daraus läßt sich folgender zweiseitiger Test für $H_0 : b_{k+1} = \ldots = b_p = 0$ konstruieren. H_0 wird abgelehnt zum Niveau α, wenn $F^\star > F(p-k, n-(p+1); 1-\alpha)$. Die F-Werte sind den Tafelwerten der F-Verteilung zu entnehmen.

Ein wichtiger Spezialfall ist der Test auf Signifikanz aller Regressoren $x_1, \ldots, x_p$. Einsetzen in die obige Formel mit $R^2_{y\hat{y}}$ als multiplem Bestimmtheitsmaß liefert die Teststatistik

$$F^\star = \frac{R^2(n-(p+1))}{(1-R^2)p} \sim F(p, n-(p+1)) \quad ,$$

die unter der $H_0 : b_1 = b_2 = \ldots b_p = 0$ Hypothese F verteilt ist.

Beispiel: Für die Daten des letzten Beispiels lassen sich jetzt sofort Konfidenzintervalle für die Regressionskoeffizienten b_j berechnen. Als Sicherheitswahrscheinlichkeit wird $1 - \alpha = 0.95$ angenommen.

$$\begin{aligned} \bar{b}_1 &= \hat{b}_1 + t_{1-\alpha/2;n-(p+1)} s_{KQ} \sqrt{c_{11}} \\ &= 1.364 + 2.447 \cdot \sqrt{12.93} \cdot \sqrt{0.001586} \\ &= 1.7146 \end{aligned}$$

$$\begin{aligned} \underline{b}_1 &= \hat{b}_1 - t_{1-\alpha/2;n-(p+1)} s_{KQ} \sqrt{c_{11}} \\ &= 1.364 - 2.447 \cdot \sqrt{12.93} \cdot \sqrt{0.001586} \\ &= 1.0138 \end{aligned}$$

$$\begin{aligned} \bar{b}_2 &= \hat{b}_2 + t_{1-\alpha/2;n-(p+1)} s_{KQ} \sqrt{c_{22}} \\ &= 0.114 + 2.447 \cdot \sqrt{12.93} \cdot \sqrt{0.001591} \\ &= 0.4648 \end{aligned}$$

$$\begin{aligned} \underline{b}_2 &= \hat{b}_2 - t_{1-\alpha/2;n-(p+1)} s_{KQ} \sqrt{c_{22}} \\ &= 0.114 - 2.447 \cdot \sqrt{12.93} \cdot \sqrt{0.001591} \\ &= -0.2370 \end{aligned}$$

Die Annahme 6 (Normalverteilung) ist für die Konstruktion von Konfidenzintervallen und statistischen Tests nicht erforderlich, wenn eine große Stichprobe vorliegt. Der t-Test wird dann durch den z-Test ersetzt.

Beispiel: Test auf Gleichheit von Regressionskoeffizienten. Die Nullhypothese lautet $H_0 : b_j = b_k$ gegen $H_1 : b_j \neq b_k$. Als Teststatistik verwendet man die standardisierte Differenz:

$$D^\star = \frac{\hat{b}_j - \hat{b}_k}{\sqrt{s_D^2}}$$

Die Statistik s_D^2 ist ein konsistenter Schätzer der Varianz σ_D^2 der Differenz $\hat{b}_j - \hat{b}_k$:

$$s_D^2 = V(\hat{b}_j) + V(\hat{b}_k) - 2Kov(\hat{b}_j, \hat{b}_k) = s_{KQ}^2(c_{jj} + c_{kk} - 2c_{jk})$$

$Kov(\hat{b}_j, \hat{b}_k)$ ist die Kovarianz der Schätzer $\hat{b}_j$ und $\hat{b}_k$. Die H_0 wird auf dem Testniveau α verworfen, wenn die Teststatistik $|D^\star| \geq z_{1-\alpha/2}$ ist.

10.3 Einfache Zeitreihenanalyse

In einigen Untersuchungen sind nicht nur die Daten, sondern auch deren zeitliche Abfolge wichtig. Dies ist zum Beispiel notwendig, um Prozesse zu analysieren, und/oder um zukünftige Entwicklungen zu prognostizieren. Beispielsweise sind Regierungen daran interessiert, die Bevölkerungsentwicklung abzuschätzen, oder Banken sind daran interessiert, Aktienkurse zu prognostizieren. Ein weiteres Beispiel ist die Absatzprognose eines Händlers, um den Lagerbestand rechtzeitig aufstocken zu können.

Ein Datensatz heißt *Zeitreihe*, wenn er Informationen über die Zeit, in der die Daten angefallen sind enthält. Es kann sich dabei sowohl um Zeitpunkte, als auch um Zeitperioden (z.B. Monat, Jahr) handeln. Diese Information kann auf unterschiedliche Art kodiert sein:

- *Explizite Zeitinformation:* Die Zeitpunkte, bzw. Zeitperioden sind im Datensatz enthalten.

  Beispiel: Abverkaufszahlen eines Unternehmens für ein bestimmtes Produkt.

i	1	2	3	$\cdots$
Monat	10/1996	09/1996	11/1996	$\cdots$
Abverkauf (in Stück)	17	25	20	$\cdots$

- *Implizite Zeitinformation:* Die Daten liegen in chronologischer Reihenfolge mit gleichen zeitlichen Abständen vor. Die Position i innerhalb des Datensatzes reicht dann als Zeitinformation aus.

  Beispiel: Die Abverkaufszahlen aus dem ersten Beispiel lassen sich auch in dieser Form darstellen, da alle zeitlichen Abstände gleich groß sind und keine Lücken enthalten:

i	1	2	3	$\cdots$
Abverkauf (in Stück)	25	17	20	$\cdots$

Zur Zeitreihenanalyse verwendet man oft Regressionsmodelle der Form

$$y_t = g(t, y_{t-1}, y_{t-2}, \ldots, \vartheta) + \varepsilon_t$$

wobei t die Zeit, y_{t-1} den um eine Zeiteinheit verzögerten Wert und ϑ den zu schätzenden Parametervektor bezeichnet. Sehr oft kommen auch andere Verfahren zum Einsatz (z.B.: Exponentielle Glättung, Verfahren der bayesianischen Statistik, etc.), deren Darstellung jedoch den Rahmen dieses Skriptums sprengen würden. Hier wird die Zeitreihenanalyse exemplarisch mit Hilfe der *Trendanalyse* eingeführt. Andere Verfahren wie gleitende Mittelwerte oder autoregressive Prozesse werden hier nicht behandelt.

Linearer Trend

Man betrachte eine Zeitreihe $y_1, y_2, \ldots, y_T$. Das einfachste trendanalytische Modell ist:

$$y_t = b_0 + b_1 \cdot t + \varepsilon_t, \quad t = 1, \ldots, T$$

Wie unschwer zu erkennen ist, ist dieses Model äquivalent zu einem klassischen Regressionsmodell, bei dem die erklärende Variable die Zeit t ist, daher erfolgt die Berechnung analog. Dieses Modell ist interessant, weil es die Möglichkeit bietet zusätzliche Regressoren einzuführen. Weiß man zum Beispiel, daß durch die Urlaubszeit im August die Abverkaufszahlen deutlich niedriger sind, als in den übrigen Monaten, könnte man eine Dummyvariable AUG einführen, die dann den Wert 1 annimmt, wenn der aktuelle Zeitpunkt der August ist, und ansonsten 0 ist. Damit kann die Anpassungsgüte des Modells verbessert werden. Solche Effekte, die wie in diesem Beispiel periodisch wiederkehren, nennt man *saisonale Effekte*.

Die Regressionsfunktion wird *Trendfunktion* genannt und ist für das einfache Modell gegeben durch:

$$\hat{y}_t = \hat{b}_0 + \hat{b}_1 \cdot t$$

Die Prognose für den Zeitpunkt $T + k$ erhält man durch Einsetzen der Zeit:

$$\hat{y}_{T+k} = \hat{b}_0 + \hat{b}_1 \cdot (T + k)$$

Nichtlinearer Trend

In den meisten Fällen reicht die Verwendung eines linearen Modells nicht aus, da die meisten dynamischen Prozesse nichtlinear sind. Beispielsweise werden Analysen von ungebremsten Wachstumsprozessen in biologischen Populationen meist mit Hilfe exponentieller Modelle durchgeführt. Im Folgenden werden exemplarisch einige nichlineare Trendmodelle vorgestellt.

- *Exponentialfunktion:*

  $$y_t = \exp(b_0 + b_1 \cdot t) \cdot \varepsilon_t, \quad t = 1, \ldots, T, \quad t > 0$$

  Diese Gleichung kann durch logarithmieren in die lineare Form $\ln y = b_0 + b_1 \cdot t$ transformiert werden, wodurch die Berechnung der Parameter vereinfacht wird.

- *Zeitinverse Exponentialfunktion:*

  $$y_t = \exp(b_0 - b_1/t) \cdot \varepsilon_t, \quad t = 1, \ldots, T, \quad b > 0, \quad t > 0$$

  Die Trendfunktion hat einen S-förmigen Verlauf, und kann ebenfalls durch eine Logarithmierung in eine lineare Funktion überführt werden.

- *Logistische Funktion:*

$$y_t = \frac{s}{1 + \exp(b_0 - b_1 \cdot t)} + \varepsilon_t, \quad t = 1, \ldots, T, \quad b > 0, \quad s > 0,$$

wobei s als *Sättigungsniveau* bezeichnet wird und ebenfalls geschätzt werden muß. Die Trendfunktion ist ebenfalls S-förmig, läßt sich jedoch nicht in eine lineare Form transformieren. Daher erfolgt die Schätzung mit Hilfe numerischer Verfahren der nichtlinearen Regression.

10.4 Verallgemeinerungen des klassischen Regressionsmodells

Eine zentrale Annahme im klassischen Regressionsmodell ist die Annahme 5: $E(ee^T) = \sigma^2 I$. Diese Annahme impliziert, daß die Varianzen aller Fehlerkomponenten identisch (Homoskedastizität) und die Fehler verschiedener Stichprobenelemente unkorreliert sind. Diese Annahme ist in vielen Fällen nicht haltbar, wie folgende Beispiele zeigen:

Beispiel zur Heteroskedastizität: Wir untersuchen den Zusammenhang zwischen Einkommen Y_t und Konsum C_t durch Befragung von Personen $t = 1, \ldots, T$ in einer Querschnittsuntersuchung. Der vermutete Zusammenhang ist gegeben durch die Regressionsgleichung:

$$C_t = b_0 + b_1 Y_t + e_t$$

Bei dieser Gleichung ist zu beachten, daß aus Gründen der ökonomischen Tradition die abhängige Variable mit C_t und die unabhängige Variable mit Y_t bezeichnet wird. Werden b_0 und b_1 durch die Methode der kleinsten Quadrate geschätzt, unterstellt man gleiche Varianzen $E(e_t^2) = \sigma^2$ für $t = 1, \ldots T$. Die ökonomische Theorie vermutet jedoch, daß bei größeren Einkommen nicht nur der Konsum, sondern auch die Fehlervarianzen stärker ausgeprägt sind. Ein mögliches Modell ist gegeben durch $E(e_t^2) = \sigma^2 Y_t^2$ und $E(e_t e_t') = 0$ für $t \neq t'$.

Trotz dieser neuen Spezifikation der Varianz des Fehlers ist der KQ-Schätzer für b zwar noch erwartungstreu und konsistent, die Aussagen des Satzes in Abschnitt 11.2 bezüglich 2, 3, 5, 6 und 7 gelten jedoch nicht mehr, wie durch Einsetzen von $E(e_t^2) = \sigma^2 Y_t^2$ in den Beweis des Satzes gezeigt werden kann. Einen Schätzer mit den gleichen Eigenschaften wie denen des KQ-Schätzers im homoskedastischen Fall erhält man durch folgende Transformationen:

$$\frac{C_t}{Y_t} = b_1 + b_0 \frac{1}{Y_t} + e_t^\star \quad \text{mit} \quad e_t^\star = e_t \frac{1}{Y_t}$$

Dieser Schätzer heißt *Aitken-Schätzer*. Dies entspricht einem Modell, in dem alle Variablen mit dem Faktor $1/Y_t$ gewichtet werden. Nunmehr gilt: $E(e_t^{\star 2}) = \sigma^2$ für alle t. Daher erfüllt das transformierte Modell die Annahmen 1) bis 6) der klassischen Regression.

Beispiel zur Autokorrelation: Wir untersuchen den gleichen Zusammenhang wie oben, aber an die Stelle von individuellen Querschnittsdaten treten aggregierte Jahresdaten der volkswirtschaftlichen Gesamtrechnung, so daß eine Längsschnittuntersuchung vorliegt.

$$C_t = a_0 + a_1 Y_t + u_t \qquad t = 1, \ldots, T$$

Schätzt man mit KQS a_0 und a_1 sowie das Residuum $\hat{u}_t = C_t - (\hat{a}_0 + \hat{a}_1 Y_t)$ und trägt das Residuum gegen die Zeitachse auf, erhält man häufig ein zyklisches Muster. Gründe für dieses Muster sind häufig:

- Wichtige Variablen fehlen in der Regressionsgleichung (Fehlspezifikation).
- Die Fehler sind über die Zeit korreliert, d.h. $E(e_t e_{t+1}) \neq 0$. Im vorliegenden Beispiel spricht man von positiver Autokorrelation. Wir treffen folgende Annahme:

  $u_t = \rho u_{t-1} + v_t \quad (v_t \text{ sind iid})$

  In diesem Fall folgt u_t einem autoregressiven Prozeß erster Ordnung (AR(1)). Der Fehler v_t erfüllt die Annahmen des klassischen Regressionsmodells.

Im Fall der Autokorrelation wird angenommen, daß ρ entweder bereits bekannt ist oder durch $\hat{\rho}$ geschätzt wird. Der Schätzer $\hat{\rho}$ ist definiert durch:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2} \quad ,$$

Ist ρ bekannt oder durch $\hat{\rho}$ geschätzt, führt die nachfolgende Transformation der Bildung der ersten Differenzen wiederum zu Homoskedastizität und Unkorreliertheit des neuen Fehlers v_t.

$$\begin{aligned} C_t - C_{t-1} &= a_0 + a_1 Y_t + u_t - \rho a_0 - \rho a_1 Y_{t-1} - \rho u_t \\ &= (1-\rho) a_0 + a_1 (Y_t - \rho Y_{t-1}) + u_t - \rho u_{t-1} \\ &= (1-\rho) a_0 + a_1 (Y_t - \rho Y_{t-1}) + v_t \end{aligned}$$

Die letzte Gleichung ist äquivalent einer Regressionsgleichung in den ersten Differenzen mit den Annahmen des klassischen Regressionsmodells.

Das Vorliegen von Autokorrelation wird mit Hilfe der Durbin-Watson-Statistik getestet. Sowohl für Heteroskedastizität als auch für korrelierte Fehler werden in der Ökonometrie erheblich komplexere Modelle behandelt.

10.5 Varianz- und Kovarianzanalyse

Eine spezielle Form der Regressionsanalyse tritt auf, wenn die Regressoren nur die Werte 0 oder 1 annehmen. Dies ist insbesondere dann der Fall, wenn die unabhängige Variable nominal skaliert ist. Die Umsetzung einer nominal skalierten Variablen in Dummy-Variable (d.h. 0, 1 Variable) wird an folgenden Beispielen deutlich. Sei A eine qualitative Variable mit Ausprägungen $\{A_1, A_2, \ldots, A_L\}$. Die Ausprägung der qualitativen Variablen wird durch L Dummy-Variable repräsentiert. Die $l$te Dummy-Variable nimmt genau dann den Wert 1 an, wenn die $l$te Ausprägung A_l eintritt. Alle anderen Dummy-Variablen werden auf 0 gesetzt. Formal lautet das Regressionsmodell mit einer qualitativen unabhängigen Variablen dann:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + , \ldots, + b_l x_{il} + \ldots, + b_L x_{iL} + e_i$$

mit den Regressoren $x_{il} = 1$, wenn i die Ausprägung A_l hat, und $x_{il} = 0$ sonst.

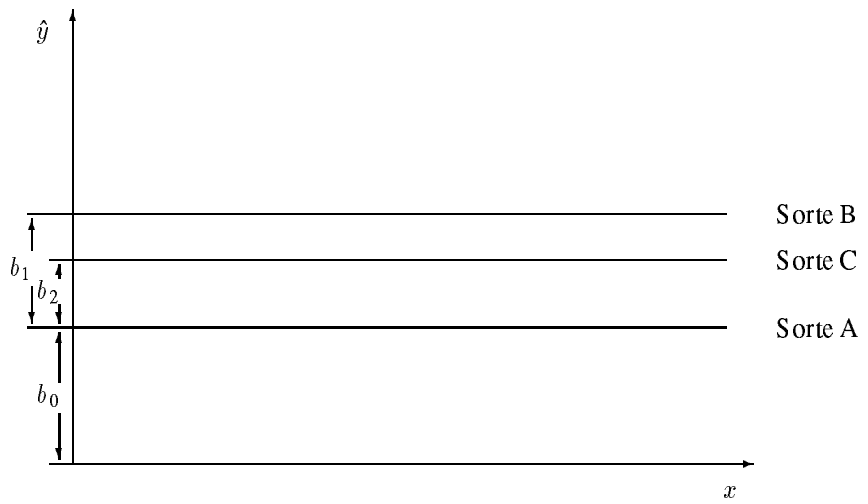
Beispiel zur einfachen Varianzanalyse: Zur Verbesserung der Versorgung in einem Entwicklungsland wurden drei verschiedene Sorten Mais gezüchtet. Diese Sorten A, B, C wurden auf 5 bzw. 7 bzw. 4 gleich großen Feldern angebaut, wobei sich folgende Erträge ergaben (in 100 kp):

Sorte A	Sorte B	Sorte C
13.3	8.3	11.2
11.8	10.3	10.9
10.7	9.1	13.4
9.1	10.2	12.7
12.0	8.8	
	12.5	
	11.9	

Der Ertrag ist abhängig von der Maissorte. Dann nimmt x_{il} jeweils den Wert 1 an, wenn der Ertrag y_i von der Sorte l stammt, sonst ist $x_{il} = 0$. Damit ist die Matrix der Regressoren (Designmatrix) X mit $n = 16$ und $L = p = 3$ gegeben durch:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \qquad y = \begin{pmatrix} 13.3 \\ 11.8 \\ 10.7 \\ 9.1 \\ 12.0 \\ 8.3 \\ 10.3 \\ 9.1 \\ 10.2 \\ 8.8 \\ 12.5 \\ 11.9 \\ 11.2 \\ 10.9 \\ 13.4 \\ 12.7 \end{pmatrix}$$

Unmittelbar einsichtig ist die Tatsache, daß die Matrix keinen vollen Spaltenrang besitzt, da $x_{.1} = x_{.2} + x_{.3} + x_{.4}$. Daher ist die Matrix $(X^T X)$ nicht invertierbar. Um in X vollen Spaltenrang zu erreichen, führen wir lineare Restriktionen ein, die auch als Reparametrisierungsbedingungen bezeichnet werden. Lineare Restriktionen können auf vielfältige Weise eingeführt werden. Wir beschränken uns hier auf die einfachste Restriktion, indem wir eine linear abhängige Spalte streichen und damit den dieser Spalte entsprechenden Parameter implizit auf 0 setzen. In unserem Beispiel führen wir dies für die erste Ausprägung, nämlich Sorte A, ein und streichen die zweite Spalte von X. Die neue Matrix der Regression X hat damit die Ordnung $n \times 3$ und ist von vollem Spaltenrang. Der Parameter b_1 wird implizit auf 0 gesetzt, die Parameter b_2 und b_3 sind als Mittelwertsdifferenzen zur ersten Gruppe interpretierbar. Der Mittelwert der ersten Gruppe wird durch b_0 geschätzt.



In den Anwendungen treten jedoch in der Regel sowohl nominala als auch metrisch skalierte Variable auf. Berücksichtigt man die oben angeführten Restriktion für nominale Variable, so läßt sich die gesamte Regressormatrix einfach durch Anfügen der Datenvektoren für die metrischen Variablen erzeugen. Im gemischten Fall spricht man von Kovarianzanalyse. Die Berechnungen erfolgen in genau gleicher Weise wie im klassischen Regressionsmodell.

Beispiel zur Kovarianzanalyse: Wir untersuchen die Abhängigkeit der Größe von Schiffsbesatzungen Y der englischen Handelsmarine um 1870 von der Antriebsart mit den Kategorien P_1 = unbekannt, P_2 = Segel, P_3 = Dampf

und der Tonnage T eines Schiffes. Die Daten sind in Rohform:

P	T	Y	P	T	Y
1	44	3	3	357	10
1	144	6	3	1080	16
1	150	5	3	1027	22
2	236	8	1	45	2
3	739	16	1	62	3
3	970	15	2	68	2
3	2371	23	3	2507	22
3	309	5	2	138	2
3	679	13	3	502	18
2	26	4	3	1501	21
3	1272	19	3	2750	24
3	3246	33	3	192	9
3	1904	19			

Entsprechend diesen Variablen wählen wir als Regressionsmodell

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$$

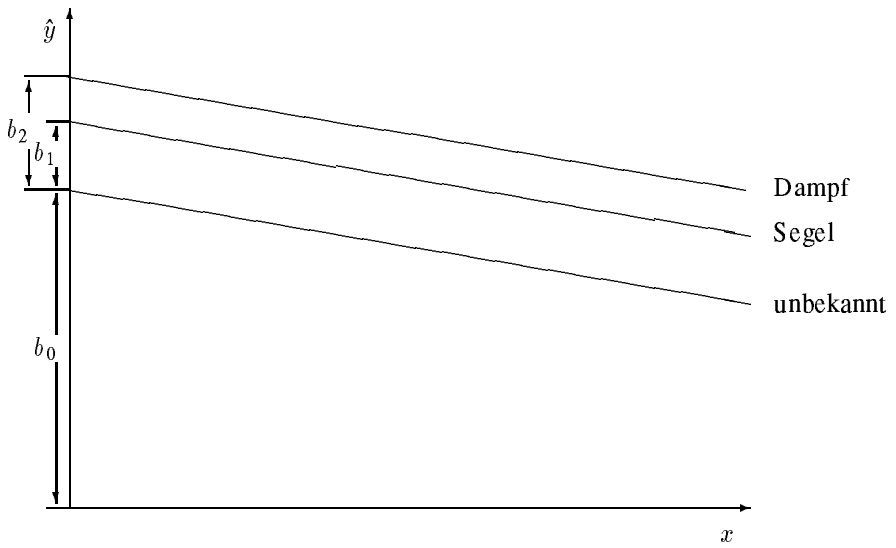
mit den Dummyvariablen $x_{i1} = 1$ für $P = 1$, $x_{i2} = 1$ für $P = 2$ und $x_{i3} = 1$ für $P = 3$ sowie der metrischen Variablen x_{i4} für Tonnage. Damit ist die Datenmatrix X gegeben durch:

b_0	b_1	b_2	b_3	b_4
1	1	0	0	44
1	1	0	0	144
1	1	0	0	150
1	0	1	0	236
1	0	0	1	739
1	0	0	1	970
1	0	0	1	2371
1	0	0	1	309
1	0	0	1	679
1	0	1	0	26
1	0	0	1	1272
1	0	0	1	3246
1	0	0	1	1904
1	0	0	1	357
1	0	0	1	1080
1	0	0	1	1027
1	1	0	0	45
1	1	0	0	62
1	0	1	0	68
1	0	0	1	2507
1	0	1	0	138
1	0	0	1	502
1	0	0	1	1501
1	0	0	1	2750
1	0	0	1	192

Um vollen Spaltenrang zu erreichen, führen wir als lineare Restriktion $b_1 = 0$ ein und streichen somit die zweite Spalte der Datenmatrix und berechnen aus der reduzierten Matrix den KQ-Schätzer:

$$\hat{b}^T = (0.959,\ 0.253,\ 3.617,\ -0.0283)$$

Die Berechnung von s^2_{KQ}, R^2, Konfidenzintervallen und Tests bleibt als Übungsaufgabe dem Leser überlassen.



10.6 Multiple und partielle Korrelation

Unter multipler Korrelation versteht man den Zusammenhang zwischen einer metrischen Variablen y und einer Linearkombination $\hat{y} = b_0 + b_1 X_1 + \ldots + b_p X_p$ von anderen Variablen. Die Gewichte $b_1, \ldots, b_p$ sind so gewählt, daß $r_{y\hat{y}}$ maximal wird. Die Gewichte stimmen mit den KQ-Schätzern $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_p$ der linearen Regression überein. $r_{y\hat{y}}$ stimmt mit $R_{y \cdot x_1 \ldots x_p}$ überein. Die Berechnung erfolgt mit der im vorhergehenden Abschnitt angegebenen Streuungszerlegung basierend auf den Residuen $\hat{e}_1, \ldots, \hat{e}_n$.

Wie bereits bei der Interpretation der einfachen Korrelation bemerkt wurde, ist es oft ratsam, Scheinzusammenhänge zu untersuchen. Ein geeignetes Instrument ist der partielle Korrelationskoeffizient, der mit folgendem Beispiel eingeführt wird.

Beispiel: Bei einer Stichprobe von Schulkindern im Alter von 6 - 10 Jahren findet man zwischen manueller Geschicklichkeit und Körpergewicht eine Korrelation von $r = 0.45$. Dies widerspricht der Erfahrung, die einen negativen Zusammenhang vermuten läßt. Die Erklärung dieses Resultates besteht darin, daß sowohl manuelle Geschicklichkeit als auch Körpergewicht positiv mit dem Lebensalter korreliert. Um die Einwirkung des Alters auszuschalten, teilt man die Kinder in Gruppen gleichen Alters und untersucht die Korrelation zwischen manueller Geschicklichkeit und Körpergewicht in jeder dieser hinsichtlich des Alters homogenen Gruppen.

Allgemein läßt sich der partielle Korrelationskoeffizient über die Residuen des Regressionsmodells definieren. Ohne Beschränkung der Allgemeinheit betrachten wir die ersten beiden Variablen X_1, X_2 aus einer Menge von Zufallsvariablen $\{X_1, X_2, \ldots, X_p\}$. Unterstellt man, daß die Variablen $X_i, i = 1, 2$ eine stochastische Linearkombination über die Regressionsgleichung

$$X_i = b_0^{(i)} + \sum_{j=3}^{p} b_j^{(i)} X_j + e^{(i)}$$

erfüllen, so ist die partielle Korrelation zwischen X_1 und X_2 unter Elimination von $X_3, \ldots, X_p$ durch den Korrelationskoeffizienten $\rho_{e_1 e_2}$ definiert. Er wird mit $\rho_{1,2\cdot 3,\ldots,p}$ bezeichnet. Die Schätzung erfolgt durch Berechnung der geschätzten Residuen

$$\hat{e}_i = X_i - (b_0^{(i)} + \sum_{j=3}^{p} b_j^{(i)} X_j) \quad \text{mit} \quad i = 1, 2$$

Die praktische Berechnung des geschätzten partiellen Korrelationskoeffizienten für den Fall $p = 3$ erfolgt durch die Formel

$$r_{1,2\cdot 3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

Im obigen Beispiel sind für die Variablen X_1 (manuelle Geschicklichkeit), X_2 (Körpergewicht) und X_3 (Alter) folgende Korrelationen bekannt:

$$r_{12} = 0.45, \quad r_{13} = 0.60, \quad r_{23} = 0.85$$

Einsetzen in die letzte Formel liefert:

$$r_{12\cdot 3} = \frac{0.45 - 0.60 \cdot 0.85}{\sqrt{1 - 0.60^2} \cdot \sqrt{1 - 0.85^2}} = -0.142$$

Mit Hilfe der Definition über die Residuen lassen sich Rekursionsformeln für den allgemeinen Fall $p \geq 3$ entwickeln.

Ausgehend von obigem Schätzer für die partielle Korrelation kann auch ein Signifikanztest zum Niveau α auf partielle Unkorreliertheit bzw. Unabhängigkeit von X und Y gegeben U durchgeführt werden. Die Hypothese $H_0 : \rho_{(12\cdot 3)} = 0$ wird zugunsten der Alternative $H_1 : \rho_{(12\cdot 3)} \neq 0$ zum Testniveau α verworfen, falls gilt:

$$\frac{|r_{12\cdot 3}| \cdot \sqrt{n-3}}{\sqrt{1 - r_{12\cdot 3}^2}} > t_{n-3;1-\alpha/2}$$

11 Abhängigkeit zwischen qualitativen und ordinalen Merkmalen

11.1 Assoziationsmaße für qualitative Merkmale

Liegen zwei oder mehr nominal skalierte Merkmale vor, wird zunächst - ähnlich der Korrelation - die Stärke des Zusammenhangs zwischen den Merkmalen analysiert. Ausgangspunkt der Analyse ist in der Regel eine zwei- oder höherdimensionale Kontingenztabelle. Wir beschränken uns auf zweidimensionale Kontingenztabellen mit empirischen Häufigkeiten $h_{ij}, i = 1, \ldots r, j = 1, \ldots s$. Um analoge Begriffsbildungen zur Korrelation bilden zu können, führen wir zunächst den Begriff vollständiger Abhängigkeit ein, der bei Kontingenztabellen jedoch nur für quadratische Tabellen definiert werden kann.

Definition: Zwei Merkmale A, B mit jeweils r Merkmalsausprägungen heißen vollständig abhängig, wenn bei geeigneter Anordnung der Merkmalsausprägungen nur die Diagonalzellen der Verteilungstabelle besetzt sind, d.h.

$$h_{ij} = 0 \quad \text{für} \quad i \neq j, \quad i, j = 1, 2, \ldots, r$$

Im folgenden werden die Bezeichnungen

$$h_{i\cdot} = \sum_j h_{ij}; \quad h_{\cdot j} = \sum_i h_{ij}; \quad n = \sum_i \sum_j h_{ij}$$

verwendet.

Beispiel: Wir betrachten zwei Merkmale A und B mit jeweils $r = 2$ Ausprägungen. Die linke Tabelle zeigt vollständige Abhängigkeit, die rechte zeigt Unabhängigkeit.

		B	$\sum$
A	.70	.00	.70
	.00	.30	.30
$\sum$	.70	.30	1.00

		B	$\sum$
A	.49	.21	.70
	.21	.09	.30
$\sum$	.70	.30	1.00

Zur Analyse der allgemeinen Kontingenztabelle benützen wir folgende Bezeichnungen:

h_{ij} beobachtete Häufigkeit

$\frac{h_{i\cdot} \cdot h_{\cdot j}}{n} = \hat{h}_{ij}$ bei Unabhängigkeit erwartete Häufigkeit

$\Delta_{ij} = h_{ij} - \hat{h}_{ij}$ Abweichungen

Die Abweichungen Δ_{ij} sind die Grundlage der Messung von Abhängigkeit. Die Summe der Differenzen ergeben zeilen- und spaltenweise die Summe 0:

$$\sum_j \Delta_{ij} = \sum_j h_{ij} - \frac{h_{i\cdot}}{n} \sum_j h_{\cdot j} = h_{i\cdot} - h_{i\cdot} \frac{n}{n} = 0.$$

$$\sum_i \Delta_{ij} = \sum_i h_{ij} - \frac{h_{\cdot j}}{n} \sum_i h_{i\cdot} = h_{\cdot j} - h_{\cdot j} \frac{n}{n} = 0.$$

Die Abweichungen geben an, wie sehr man die Elemente der Grundgesamtheit (oder Stichprobe) umverteilen muß, damit aus der bei Unabhängigkeit erwarteten Verteilung die beobachtete Verteilung entsteht.

Definition: Das Assoziationsmaß 'Chi-Quadrat' (auch 'quadratische Kontingenz') ist gegeben durch

$$\chi^2 = \sum_i \sum_j \frac{\Delta_{ij}^2}{\hat{h}_{ij}} = \sum_i \sum_j \frac{(h_{ij} - \hat{h}_{ij})^2}{\hat{h}_{ij}}$$

Folgende Umformungen sind für die praktische Rechnung bequemer:

$$\chi^2 = \sum_i \sum_j \frac{h_{ij}^2}{\hat{h}_{ij}} - n = n \left(\sum_i \sum_j \frac{h_{ij}^2}{h_{i\cdot} h_{\cdot j}} - 1 \right)$$

Beweis der Umformung:

$$\sum_i \sum_j \frac{1}{h_{ij}^e} (h_{ij} - \hat{h}_{ij})^2 = \sum_i \sum_j \frac{1}{\hat{h}_{ij}} \left[h_{ij}^2 - 2 h_{ij} \hat{h}_{ij} + \hat{h}_{ij}^2 \right]$$

$$= \sum_i \sum_j \frac{h_{ij}^2}{\hat{h}_{ij}} - 2 \sum_i \sum_j h_{ij} + \sum_i \sum_j \hat{h}_{ij} = \sum_i \sum_j \frac{h_{ij}^2}{\hat{h}_{ij}} - n \quad .$$

Das χ^2-Maß hängt von n, also dem Umfang der Grundgesamtheit oder Stichprobe ab. Aus diesem Grunde eignet es sich nicht zum Vergleich des Abhängigkeitsgrades bei verschieden großen Grundgesamtheiten oder Stichproben. Man hat daher Koeffizienten vorgeschlagen, die (1) von n nicht abhängen, (2) bei statistischer Unabhängigkeit null werden und (3) bei vollständiger Abhängigkeit den Wert 1 annehmen.

Wir betrachten zunächst *Vierfelderkoeffizienten* für Merkmale mit jeweils $r = 2$ Merkmalsausprägungen.

	B_1	B_2	$\sum$
A_1	a	b	a + b
A_2	c	d	c + d
$\sum$	a + c	b + d	h..

Für diese Tabelle gelten somit die Bezeichnungen:

$$h_{11} = a, \quad h_{12} = b, \quad h_{21} = c, \quad h_{22} = d$$

$$n = a + b + c + d, \quad \hat{h}_{11} = \frac{(a+c)(a+b)}{n}, \quad \hat{h}_{12} = \frac{(a+b)(b+d)}{n},$$

$$\hat{h}_{21} = \frac{(a+c)(c+d)}{n}, \quad \hat{h}_{22} = \frac{(c+d)(b+d)}{n}$$

Einsetzen ergibt nach einiger Rechnung:

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

Damit erhält man den Kontingenz- oder Phi-Koeffizienten:

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

wobei $0 \leq |\phi| \leq 1$ gilt, d.h. dieser Koeffizient ist normiert.

Zusätzlich definieren wir den Assoziationskoeffizienten κ, der auch *Yule'sches Assoziationsmaß* heißt:

$$\kappa = \frac{ad-bc}{ad+bc} \quad .$$

In unserem Beispiel erhält man für die linke Vierfeldertafel:

$$\phi = \frac{(.70)(.30)}{\sqrt{(.70)(.70)(.30)(.30)}} = 1$$

$$\kappa = \frac{(.70)(.30)-0}{(.70)(.30)+0} = 1 \quad ,$$

also beide Male vollständige Abhängigkeit.

Nun behandeln wir Maße für den Zusammenhang in einer *allgemeinen* Kontingenztabelle. Zu diesem Zweck geben wir vier Maße an:

- Der allgemeine Kontingenzkoeffizient C:

$$C = \sqrt{\frac{\chi^2}{n}}$$

- Das Maß C hat den Nachteil, daß es größer als 1 werden kann. Der korrigierte Kontingenzkoeffizient C_{cor} ist hingegen immer kleiner als 1:

$$C_{cor} = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

- Zwei andere Wege der Normierung werden in den nächsten beiden Koeffizienten eingeschlagen. Der erste Koeffizient ist das Kontingenzmaß von Tschuprow:

$$T = \sqrt{\frac{\chi^2}{n\sqrt{(s-1)(r-1)}}}$$

- Der zweite Koeffizient ist das Kontingenzmaß von Cramér.

$$V = \sqrt{\frac{\chi^2}{n \cdot min\{(s-1),(r-1)\}}}$$

Der Größe nach sind diese Koeffizienten in der Regel nach der Reihenfolge $C \geq C_{cor} \geq T \geq V$ geordnet. Diese Maße sind 0, wenn die beiden Merkmale unabhängig sind. V ist T überlegen, weil der Koeffizient auch dann den Wert 1 annehmen kann, wenn s und r ungleich sind.

Faßt man in einer Kontingenztafel Spalten und/oder Zeilen zusammen, vergröbert man also die Gruppen, so kann der χ^2-Wert nicht größer werden.

Bei der Interpretation weisen Werte unter 0.2 auf schwache, Werte zwischen 0.2 und 0.5 auf mittelstarke und Werte, die größer als 0.5 sind, auf starke Zusammenhänge hin. Die Art des Zusammenhangs ist jedoch nur durch Betrachtung der bedingten Verteilungen zu erfassen.

Beispiel: Zusammenhang zwischen Religionsbekenntnis von Bräutigam und Braut bei allen Ehen, die 1957 vor Wiener Standesämtern geschlossen wurden.

	E	K	S	O	$\sum$
Evangelisch	344	728	22	44	1138
Katholisch	693	9916	97	293	10999
Sonstige	27	248	134	22	431
ohne Bekenntnis	108	812	31	197	1148
$\sum$	1172	11704	284	556	13716

Zunächst wird χ^2 und dann T berechnet:

$$\chi^2 = 13716 \cdot (\frac{344^2}{1172 \cdot 1138} + , \ldots , + \frac{197^2}{556 \cdot 1148} - 1) = 3223.7$$

$$T = \sqrt{\frac{3223.7}{13716 \cdot \sqrt{(4-1)(4-1)}}} = 0.279$$

Das Resultat deutet auf einen mittelstarken Zusammenhang zwischen den Religionsbekenntnissen der Brautleute hin.

11.2 Der χ^2 Test auf statistische Unabhängigkeit

Wir haben bereits den χ^2 Test auf Übereinstimmung einer empirischen Verteilung mit einer theoretischen Verteilung kennengelernt. Bei der Analyse von Kontingenztabellen zweier Merkmale A und B richtet sich das Interesse auf die Überprüfung der H_0 : *A und B sind voneinander statistisch unabhängig* gegen H_1 : *A und B sind statistisch abhängig*. Als Teststatistik verwendet man das Kontingenzmaß χ^2, das hier mit $\chi^{2\star}$ bezeichnet wird. Unter der Nullhypothese folgt die Teststatistik $\chi^{2\star}$ einer χ^2 Verteilung mit $(r-1)(s-1)$ Freiheitsgraden. Als Beispiel überprüfen wir die Nullhypothese: *Das Religionsbekenntnis der Brautleute ist statistisch unabhängig* an den Daten des letzten Beispiels mit einem Irrtumsniveau von $\alpha = 0.01$. Der kritische Wert ist $\chi^2_{9,0.99} = 21.7$, die Teststatistik beträgt $\chi^{2\star} = 3223.7$. Die Nullhypothese wird daher zum vorgegebenen Testniveau verworfen.

Genauso wie im Fall von metrischen Variablen können bei qualitativen Merkmalen Scheinzusammenhänge auftreten, die durch dritte Variable verursacht werden. Man untersucht daher anstelle von zweidimensionalen häufig drei- und höherdimensionale Kontingenztabellen.

Als *Beispiel* für eine dreidimensionale Kontingenztabelle betrachten wir den Zusammenhang zwischen Rauchen, Lungenkrebs und Vogelhaltung. Die beiden folgenden Tabellen zeigen die Häufigkeitsverteilung von Lungenkrebs und Vogelhaltung nach Rauchern und Nichtrauchern getrennt.

Nichtraucher			
Vogel-	Lungenkrebs		
haltung	ja	nein	$\sum$
ja	6	36	42
nein	5	162	167
$\sum$	11	198	209

Raucher			
Vogel-	Lungenkrebs		
haltung	ja	nein	$\sum$
ja	93	67	160
nein	119	179	298
$\sum$	212	246	458

Für die Nichtraucher erhält man $\chi^{2\star} = 8.58$ und den korrigierten Kontingenzkoeffizient $C_{cor} = 0.2$. Für die Raucher erhält man $\chi^{2\star} = 13.86$ und den korrigierten Kontingenzkoeffizient $C_{cor} = 0.17$. Die beiden korrigierten Kontingenzkoeffizienten zeigen einen schwachen bis mittelstarken Zusammenhang zwischen Lungenkrebs und Vogelhaltung, wobei der Zusammenhang bei den Nichtrauchern etwas stärker ausgeprägt ist. Die Hypothese, daß kein Zusammenhang zwischen Lungenkrebs und Vogelhaltung besteht, muß bei Rauchern wie bei Nichtrauchern sogar bei der sehr kleinen Irrtumswahrscheinlichkeit $\alpha = 0.005$ verworfen werden ($\chi^{2\star}_{1,0.995} = 7,88$). Weiter läßt sich aus den Tabellen ablesen, daß nur ca. 5% der untersuchten Nichtraucher, aber ca. 46% der Raucher an Lungenkrebs erkrankt sind. Bezieht man jedoch zusätzlich die Variable Vogelhaltung in die Analyse mit ein, so stellt man fest, daß ca. 14% der Nichtraucher und ca. 59% der Raucher, die einen Vogel halten, unter Lungenkrebs leiden. Betrachtet man ausschließlich die an Krebs erkrankten Personen, stellt man fest, daß sogar ca. 55% der an Krebs erkrankten Nichtraucher und ca. 44% der Raucher einen Vogel besitzen. Man muß somit von der Möglichkeit ausgehen, daß neben Rauchen auch Vogelhaltung das Krebsrisiko erhöht.

11.3 Assoziationsmaße und Tests für ordinale Merkmale

Als Maß der Stärke des Zusammenhangs zwischen ordinal skalierten Variablen X und Y werden folgende Koeffizienten verwendet:

Der Rangkorrelationskoeffizient von Kendall

Die Beobachtungen $(x_i, y_i), i = 1, \ldots, n$ werden so angeordnet, daß die x Werte nach wachsender Größe geordnet sind. Danach stellt man fest, wieviele Paare rechts von i eine y Komponente aufweisen, die größer ist als y_i; die Anzahl solcher Paare sei mit S_i bezeichnet. Dann heißt der Koeffizient

$$\hat{\tau} = \frac{4\sum_{i=1}^{n} S_i}{n(n-1)} - 1, \quad -1 \leq \hat{\tau} \leq 1$$

Schätzer für den Rangkorrelationskoeffizienten τ einer Grundgesamtheit, der nach dem Statistiker Kendall als Kendall's τ bezeichnet wird.

Beispiel: Wir betrachten $n = 5$ Paare von Rangdaten, die nach x geordnet sind.

i	1	2	3	4	5
x	1	3	4	8	9
y	2	1	5	3	4
S_i	3	3	0	1	0

Wegen $n = 5$ und $\sum_i^n S_i = 7$ erhalten wir

$$\tau = \frac{4 \cdot 7}{5 \cdot 4} - 1 = 0.4$$

Der Kendall'sche Rangkorrelationskoeffizient verteilt sich bei Unabhängigkeit der beiden Merkmale und Vorliegen von mindestens 10 Beobachtungspaaren annähernd normal mit Mittelwert 0 und Varianz

$$\sigma^2 = \frac{2(2n+5)}{9n(n-1)}$$

Damit lassen sich Teststatistiken zur Überprüfung der $H_0 : \tau = 0$ gegen $H_1 : \tau \neq 0$ formulieren.

Der Rangkorrelationskoeffizient von Spearman

Ein weiterer Rangkorrelationskoeffizient wurde von Spearman vorgeschlagen und wird mit ρ_s in der Grundpopulation bezeichnet. Zur Schätzung aus einer Stichprobe schreibt man die Meßdaten ihrer Größe nach auf und ordnet ihnen die natürlichen Zahlen als Rangnummern zu.

$$x : x_1 < x_2 <, \ldots, < x_n \quad r(x) = 1, 2, \ldots n, \quad y : y_1 < y_2 <, \ldots, < y_n \quad r(y) = 1, 2, \ldots, n$$

Dem i-ten Paar mit den Komponenten (x_i, y_i) ist nun das Rangnummernpaar $(r(x_i), r(y_i))$ zugeordnet. Wir bilden die Rangnummerndifferenz $d_i = r(x_i) - r(y_i)$.

Die Größe

$$r_s = 1 - \frac{1}{n(n^2-1)} 6 \sum_{i=1}^{n} d_i^2 \quad , -1 \leq r_s \leq 1$$

heißt Spearman'scher Rangkorrelationskoeffizient der Stichprobe.

Beispiel: Bestimmung von r_s aus den Daten des letzten Beispiels.

i	1	2	3	4	5
d_i	-1	1	-2	1	1
d_i^2	1	1	4	1	1

Aus $n = 5$, $\sum_i^n d_i^2 = 8$ berechnet man

$$r_s = 1 - \frac{6 \cdot 8}{5 \cdot 24} = 0.6 \quad .$$

Der Spearman'sche Rangkorrelationskoeffizient r_s ist, falls $\rho_s = 0$ gilt, in Stichproben von mindestens 20 Beobachtungspaaren so verteilt, daß die Transformation

$$u = r_s \sqrt{n-2} \sim \mathcal{N}(0,1)$$

asymptotisch standard normal verteilt ist. Daraus lassen sich wieder Teststatistiken zur Überprüfung der $H_0 : \rho_s = 0$ gegen die $H_1 : \rho_s \neq 0$ berechnen.

Kendalls τ wird vor allem dann verwendet, wenn Bindungen auftreten, wenn also eine Merkmalsausprägung in x oder y nicht nur einmal vorkommt. Da in diesem Fall keine natürliche Ordnung hergestellt werden kann, ist die Verwendung von r_s problematisch. Um in solchen Fällen dennoch r_s berechnen zu können, werden bei gleichen Werten *mittlere Ränge* vergeben. Die Vorgehensweise ist dem folgenden Beispiel zu entnehmen.

Beispiel: Wir betrachten folgende Rangdaten.

i	1	2	3	4	5	6	7	8
x	1	3	4	8	9	3	7	6
y	2	1	5	3	4	4	4	6

x nimmt zweimal den Wert 3 an und bei y kommt der Wert 4 sogar dreimal vor. Mit Hilfe von mittleren Rängen, kann eine modifizierte Rangtabelle erzeugt werden, wobei die mittleren Ränge hier fett dargestellt werden:

i	1	2	3	4	5	6	7	8
$r(x)$	1	**2.5**	4	7	8	**2.5**	6	5
$r(y)$	2	1	7	3	**5**	**5**	**5**	8

12 Wirtschafts- und Sozialstatistik

12.1 Datenbasis

Wirtschaftswissenschaft als Erfahrungswissenschaft ist auf umfangreiche Informationen angewiesen, die sowohl von der amtlichen Statistik als auch von öffentlichen und privaten Instituten gesammelt und zur Verfügung gestellt werden (verfügbare Daten). Da verfügbare Daten zur Beantwortung spezifischer Fragen oft nicht ausreichen, müssen eigene Daten im Forschungsprozeß erhoben werden (nicht verfügbare Daten). Einen Überblick über die verfügbaren Daten und deren Produzenten geben Hujer und Cremer (1978), die auch die internationale Statistik sowie Sozialindikatoren behandeln.

Die genauesten Daten der allgemeinen Statistik liefern die in ca. 10-Jahres-Abständen durchgeführten Volkszählungen, in denen an einem Stichtag alle Haushalte der Bundesrepublik Deutschland erhoben und befragt werden (Totalerhebung). Die wesentlichen Merkmale, die dabei erhoben werden, sind Geschlecht, Alter, Familienstand, Staatsangehörigkeit, Beteiligung am Erwerbsleben, ausgeübter Beruf, Stellung im Beruf und Wirtschaftszweig. Volkszählungen werden nur alle 10 Jahre duchgeführt, da sie hohe Kosten verursachen, die Datenaufbereitung sehr lange dauert und die Belastung für die gesamte Bevölkerung sehr groß ist. Da jedoch aktuelle Daten über die Entwicklung der Wirtschaft, der Verkehrssituation usw. benötigt werden, wird jährlich eine Stichprobe, der Mikrozensus, erhoben, bei der ein kleiner Teil der Bevölkerung (1%, 0.25% bzw. 0.1%) befragt wird. Die Ergebnisse werden nach den Regeln der schließenden Statistik auf die Bevölkerung hochgerechnet. Beim Mikrozensus werden neben dem Grundprogramm, also den Fragen aus der Volkszählung, zusätzliche Fragen über aktuelle Probleme (Zusatzprogramm) gestellt (z. B. zum Pendlerproblem).

Die Statistik der Bevölkerungsbewegung und Familienstandsänderung wird von den Standesämtern (Geburt, Tod, Eheschließung) sowie den Gerichten (Scheidung) erstellt, während die Wanderung der Bevölkerung durch die Einwohnerämter (Meldescheine) erfaßt wird. Die Arbeitsmarktstatistik obliegt Arbeitsämtern und der Bundesanstalt für Arbeit.

Gleichzeitig mit den Volkszählungen werden Arbeitsstätten-, Gebäude- und Wohnungszählungen durchgeführt. Daher müßte man exakt von einer Volks-, Berufs-, Einkommens-, Wohnungs-, Arbeitsstätten- und Gebäudezählung sprechen.

12.2 Bevölkerungsstatistik

Grundbegriffe

In der Bevölkerungsstatistik werden drei Konzepte des Bevölkerungsbegriffs unterschieden:

- Konzept der Staatsangehörigkeit.
- Konzept der Nation. Hier handelt es sich um historische und ideologische Konzepte.
- Konzept der Wohnbevölkerung (Inlandsbevölkerung).

In der amtlichen Statistik wird nur von der Wohnbevölkerung ausgegangen, um für infrastrukturelle Maßnahmen wie Wohnungsbau, Straßenbau, Krankenhäuser, Schulen usw. Plandaten zur Verfügung zu stellen. Die (Wohn-) Bevölkerung umfaßt alle natürliche Personen, die sich überwiegend in der Bundesrepublik Deutschland aufhalten und damit durch die Meldebehörden erfaßt sind. Zur Bevölkerung gehören daher im Bundesgebiet gemeldete Gastarbeiter (einschließlich der Staatenlosen), während permanent im Ausland lebende deutsche Staatsbürger ihr nicht angehören. Nicht zur Bevölkerung gehören die Angehörigen der ausländischen Stationierungskräfte sowie der ausländischen diplomatischen und konsularischen Vertretungen mit ihren Familienangehörigen.

Die wichtigste ökonomische Einheit ist der Haushalt. Als Haushalt gilt jede Personengemeinschaft, die zusammen wohnt und eine gemeinsame Hauswirtschaft führt (natürlich auch einzelne Personen, die alleine wohnen und wirtschaften). Dabei wird zwischen Privathaushalten und Anstaltshaushalten (z.B. Heime, Gemeinschaftsunterkünfte, Gefängnisse etc.) unterschieden.

In der Bevölkerungsstatistik verwendet man Kennzahlen zur Charakterisierung der Bevölkerungsentwicklung, die als Raten und Ziffern (= Rate $\cdot$ 1000) bezeichnet werden. Alle Kennzahlen beziehen sich auf die Wohnbevölkerung. Wichtige Raten sind:

1. *Rohe Geburtenrate:*

$$r_j = \frac{b_j}{L_j}$$

b_j ist die Anzahl der Geburten im Jahr j, d. h. zwischen den Stichtagen des Jahres j und des Jahres $j+1$.
L_j ist der Durchschnittsbestand der Wohnbevölkerung im Jahr j.

2. *Altersspezifische Sterberate*

$$m_{aj} = \frac{d_{aj}}{L_{aj}}$$

d_{aj} ist die Anzahl der Sterbefälle von Personen im Jahre j, die das Alter von a Jahren erreichten.
L_{aj} ist der durchschnittliche Bestand von a-jährigen im Jahr j.

3. *Rohe Fruchtbarkeitsrate (Rohe Fertilitätsrate)*

$$f_j = \frac{b_j}{F_j}$$

F_j ist die durchschnittliche Anzahl der Frauen im gebärfähigen Alter im Jahre j. Das gebärfähige Alter ist statistisch durch die Untergrenze 15 und durch die Obergrenze 45 Jahre festgelegt.

4. *Altersspezifische Fruchtbarkeitsrate*

$$f_{aj} = \frac{b_{aj}}{F_{aj}}$$

b_{aj} ist die Anzahl der Geburten im Erhebungszeitraum $(j, j+1)$, deren Mütter bei der Geburt a Jahre alt sind.
F_{aj} ist die durchschnittliche Anzahl der Frauen, die im Erhebungszeitraum das Alter von a Jahren erreicht haben.

5. *Altersspezifische Rate von Mädchengeburten*

$$f^w_{aj} = \frac{b^w_{aj}}{F_{aj}}$$

Mit diesen Raten läßt sich allerdings noch nicht feststellen, ob eine Bevölkerung langfristig wächst, stabil bleibt oder schrumpft. Zur Analyse der langfristigen Bevölkerungsentwicklung werden folgende Kennzahlen abgeleitet. Die rohe Fruchtbarkeitsrate f_j läßt sich zerlegen in

$$f_j = \frac{b_j}{F_j} = \frac{\sum_a b_{aj}}{F_j} = \sum_a \frac{b_{aj}}{F_{aj}} \cdot \frac{F_{aj}}{F_j} = \sum_a f_{aj}\,\pi_{aj}$$

f_{aj} ist die altersspezifische Fruchtbarkeitsrate.
π_{aj} ist der Anteil von Frauen im Alter a.

Diese Zerlegung bildet die Grundlage der Definition der totalen Fruchtbarkeitsrate, die auch als Fertilitätsrate (TFR) bezeichnet wird:

$$TFR_j = \sum_a f_{aj}$$

Bei der Berechnung der totalen Fertilitätsrate wird angenommen, daß sich die altersspezifischen Fruchtbarkeitsraten im Laufe der Zeit nicht ändern (unechte Längsschnittanalyse). Berücksichtigt man nur die Mädchengeburten, erhält man den Bruttoreproduktionsindex (BRI):

$$BRI_j = \sum_a f_{aj}^w \sim \frac{1}{2} TFR_j$$

Bei der Berechnung von TFR und BRI wird angenommen, daß eine Generation von Frauen in den 31 Jahren der Gebärfähigkeit zur Gänze erhalten bleibt. Diese Annahme ist unrealistisch. Daher wird die Überlebensrate von gebärfähigen Frauen jeder Altersstufe ($p_{aj}^w, a = 15, 16, \ldots, 45$), die aus den Sterbetafeln der amtlichen Statistik entnommen werden kann, berücksichtigt. Analog zu den konstanten altersspezifischen Fruchtbarkeitsraten werden konstante altersspezifische Mortalitätsraten angenommen. Die Korrektur des BRI durch Überlebensraten führt zur Definition des Nettoreproduktionsindex:

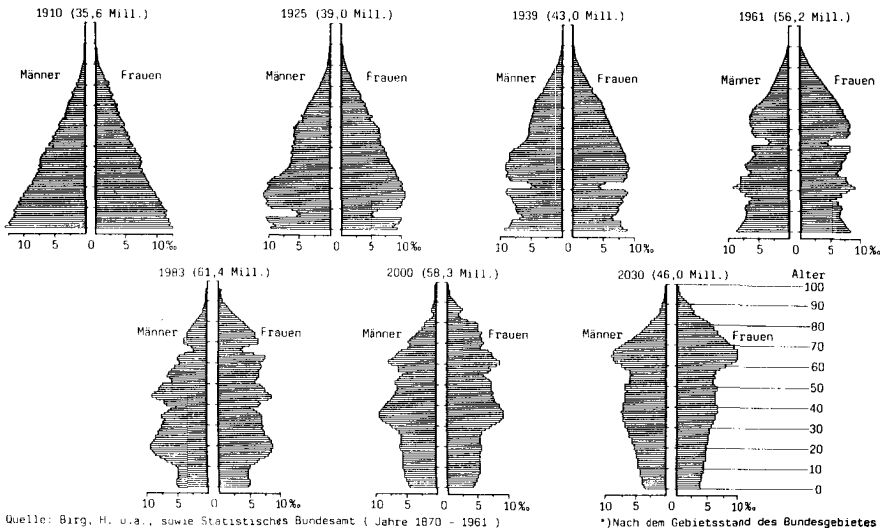
$$NRI_j = \sum_a f_{aj}^w p_{aj}^w$$

Der NRI wird üblicherweise wie folgt interpretiert: Ist der NRI = 1, bleibt die Bevölkerung stabil. (In Wirklichkeit sinkt sie langsam, da der NRI die Sterberate der neugeborenen Mädchen nicht berücksichtigt.) Ist der NRI > 1, so wächst die Bevölkerung, andernfalls schrumpft sie. Die Richtigkeit dieser Interpretation hängt von dem Ausmaß ab, in dem die getroffenen Annahmen approximativ erfüllt sind.

Die Bevölkerungspyramide

Die Geburtenzahl hängt neben dem Zeugungsverhalten auch von der Altersstruktur einer Bevölkerung ab. Dieser Effekt wird anhand einiger Bevölkerungspyramiden für Deutschland illustriert. Bei einer Bevölkerungspyramide ist auf der senkrechten Achse das Alter abgetragen: Unten stehen die Personen im ersten Lebensjahr; nach oben geht es bis zum Alter 100. Die Länge eines Balkens nach rechts entspricht der relativen Stärke der entsprechenden Frauenaltersgruppe in Promille, die Länge eines Balkens nach links der relativen Stärke der entsprechenden Männeraltersgruppe. Insgesamt kumuliert sich die Fläche zu 1000 Promille auf. Nach oben hin wird die Bevölkerungspyramide durch die Sterblichkeit allmählich dezimiert. Bei wachsenden Bevölkerungen ist die "Pyramide" pyramidenartig oder pfeilförmig. Bei Bevölkerungen mit konstant bleibendem Umfang ist die "Pyramide" glockenförmig, bei schrumpfenden urnenförmig.

Die folgende Abbildung repräsentiert die Altersstruktur der Bevölkerung im Deutschen Reich und in der Bundesrepublik Deutschland (ohne die neuen Länder) von 1910 bis 2030 (bis 1983 real, anschließend geschätzt).



Quelle: Birg, H. u.a., sowie Statistisches Bundesamt (Jahre 1870 - 1961)

*)Nach dem Gebietsstand des Bundesgebietes

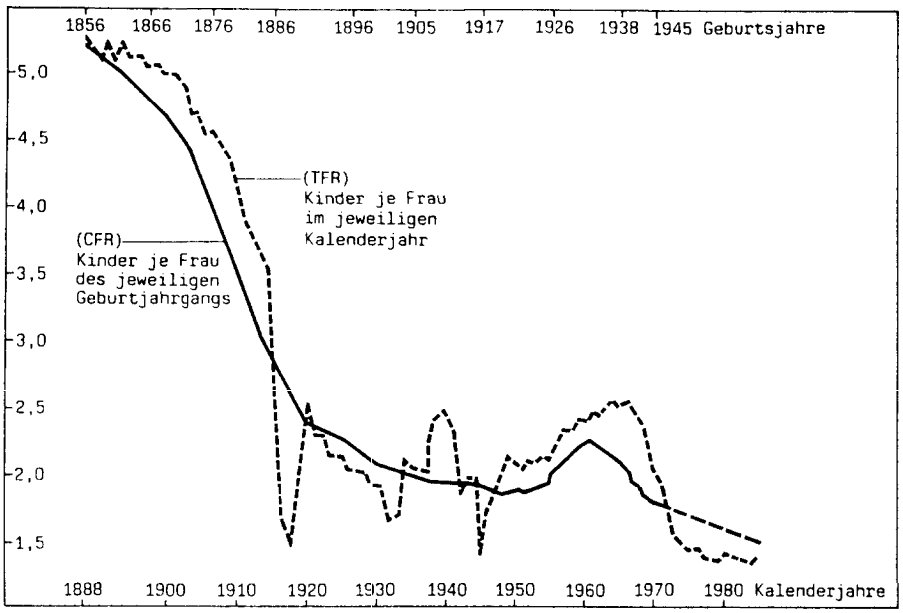
Diese Abbildung enthält links den Altersaufbau der deutschen Bevölkerung von 1910, der durch ein starkes Wachstum und keine wesentlichen Störungen gekennzeichnet ist. Bei der danebenstehenden Bevölkerungspyramide von 1925 sind deutlich drei Effekte erkennbar:

- Die gering besetzte Altersgruppe der 7 bis 10jährigen Jungen und Mädchen. Dieser Effekt geht auf die Geburtenausfälle während des ersten Weltkriegs zurück.
- In der Altersgruppe der 25 bis 50jährigen gibt es deutlich weniger Männer als Frauen. Dieser Effekt ist auf die Gefallenen des ersten Weltkriegs zurückzuführen.
- In der Altersgruppe der über 70jährigen gibt es weitaus mehr Frauen als Männer. Dieser Effekt geht auf die in etwa um 7 Jahre höhere Lebenserwartung der Frauen zurück.

Im Altersaufbau von 1939 sind diese Effekte um 14 Jahre nach oben gewandert. Am Sockel der Pyramide sind zusätzlich die nach dem ersten Weltkrieg einsetzenden starken Geburtenrückgänge zu erkennen, die ihren Tiefpunkt während der Weltwirtschaftskrise erreichten. Die beiden folgenden Alterspyramiden zeigen den Aufbau der bundesdeutschen Bevölkerung 1961 mit einer starken Verbreiterung der Basis sowie den Geburtenrückgang ab 1968, der sich in der schrumpfenden Basis des Jahres 1983 ausdrückt. Die letzten beiden Pyramiden sind (realistische) Prognosen für die bundesdeutsche Bevölkerung der Jahre 2000 und 2030 (ohne die Bevölkerung der Länder Mecklenburg, Brandenburg, Sachsen–Anhalt, Sachsen, Thüringen und Berlin). Hier wird sowohl das Schrumpfen der Bevölkerung als auch die zunehmende Überalterung deutlich. Die fruchtbarste Phase der Frauen liegt zwischen dem zwanzigsten und dem dreißigsten Lebensjahr. Offensichtlich geht auch bei konstantem generativen Verhalten die Geburtenzahl zurück, wenn eine 'Beule' in der Bevölkerungspyramide auftritt, d.h. wenn eine Altersgruppe von Frauen schwächer besetzt ist. Ein Teil der hohen Geburtenzahlen zu Beginn der sechziger Jahre ist durch die starke Besetzung der Frauenjahrgänge zu erklären, die am 31.12.1988 zwischen 45 und 55 Jahre alt waren. Andererseits ist ein Teil des Geburtenrückganges ab Mitte der sechziger Jahre darauf zurückzuführen, daß zu dieser Zeit die durch den zweiten Weltkrieg dezimierte Altersgruppe die fruchtbarste Phase durchwanderte.

Der Geburtenrückgang in der Bundesrepublik Deutschland

Die Wohnbevölkerung in der BRD ist seit 1950 von 50,3 Mio. bis 1973 auf 62,1 Mio. gewachsen. Seitdem fällt sie, wenn auch sehr langsam. Bei der deutschen Bevölkerung setzt der Rückgang bereits 1971 ein. Er wird bis 1973 nur durch einen größeren Zuzug von Ausländern und deren höhere Geburtenzahl überdeckt. Der Geburtenrückgang setzt bereits viel früher ein, wie aus folgender Graphik ersichtlich ist.



Quelle: P. Marschalck, Bevölkerungsgeschichte Deutschlands, Frankfurt 1984, ergänzt durch Daten aus H. Birg, u.a., Kohortenanalyse der Geburtenentwicklung in der Bundesrepublik Deutschland, IBS-Materialien Nr.10, Universität Bielefeld 1984.

Die außerordentlich niedrige Fertilitätsrate TFR von 1916 bis 1919 ist auf den starken Geburtenausfall während des ersten Weltkriegs zurückzuführen. Die relativ hohe Fertilitätsrate TFR von 1955 bis 1968 ist sowohl auf die hohe Geburtenzahl zwischen 1934 und 1942 (Echoeffekt) als auch auf die relativ hohe altersspezifische Fruchtbarkeitsrate dieser Frauengeneration zurückzuführen. Die rohe Geburtenziffer beträgt 18.3 Lebendgeborene pro 1000 Einwohner im Jahr 1963 und 9.6 im Jahr 1977. Von besonderem Interesse ist die Entwicklung der altersspezifischen Geburtenraten, die seit 1964 in allen Altersgruppen eine stark fallende Tendenz aufweisen, wobei in den letzten Jahren auch eine Verschiebung zu höherem Alter der Mütter festzustellen ist. Die totale Fertilitätsrate TFR ist von 2.54 im Jahr 1964 auf 1.45 im Jahr 1975 abgesunken. Zur langfristigen Bestandserhaltung der Bevölkerung ist unter Berücksichtigung der Sterblichkeit ein Wert der TFR von 2.2 erforderlich. Einen Vergleich mit anderen Ländern zeigt folgende Aufstellung:

Totale Fertilitätsrate einiger ausgewählter Industrieländer						
Jahr	BRD	DDR	Öster–reich	Frank–reich	USA	UdSSR
1950	2.10	2.35	-	2.93	3.09	-
1955	2.14	2.38	2.22	2.68	3.58	-
1960	2.37	2.35	2.59	2.73	3.65	2.82
1965	2.50	2.48	2.68	2.84	2.91	2.46
1970	2.01	2.19	2.32	2.47	2.48	2.39
1975	1.45	1.54	1.84	1.93	1.77	2.41
1980	1.45	1.94	1.68	1.96	1.82	2.28
1985	1.31	1.80	1.51	1.80	1.80	

Als Indikatoren und Faktoren für den Geburtenrückgang lassen sich folgende Fakten angeben:

- *Reduktion der Familiengröße:*

Anzahl der Kinder pro 100 Ehen			
Jahr	1966	1972	1975
ohne Kinder	15	22	24
mit 1 Kind	18	28	31
mit 2 Kindern	31	35	33
mit 3 Kindern	20	11	10
mit 4 und mehr Kindern	16	4	2

- *Siedlungsweise:* In den Großstädten liegt die Geburtenziffer traditionell niedriger als in kleinen Gemeinden. Hier findet allerdings zur Zeit eine Angleichung statt.
- *Allgemeine Säkularisierung:* Schon 1970 wiesen diejenigen Ehen die höchste Kinderzahl auf, in denen beide Partner katholisch waren (2.132 Kinder), während für Ehen, in denen beide Partner keiner Konfession angehörten, ein Durchschnittswert von 1.593 Kindern festgestellt wurde.
- *Wandel der Berufsstruktur:* Selbständige Landwirte wiesen 1970 im Durchschnitt 2.671 Kinder aus erster Ehe auf, während die Angestellten im Durchschnitt 1.622 Kinder hatten.

Auswirkungen des Geburtenrückgangs

Unter der Annahme konstanter Fruchtbarkeitsraten läßt sich die Entwicklung der Bevölkerung aufgrund der bekannten Überlebenswahrscheinlichkeiten schätzen. Ergebnisse dieser Schätzungen wurden in einer der vorhergehenden Abbildungen gezeigt. Die voraussichtliche Zusammensetzung der Bevölkerung läßt sich auf spezielle Gruppen aufgliedern, die von besonderem gesellschaftspolitischen Interesse sind. Die Nachfrage nach den Dienstleistungen des Schul- und Ausbildungssystems wird wesentlich von der Anzahl der Jugendlichen in bestimmten Altersgruppen bestimmt. Insgesamt ist langfristig mit einer Abnahme des Bildungsbedarfs zu rechnen. Trotzdem kann es in den nächsten Jahren noch zu schweren Belastungen des Bildungssystems durch den Zustrom von Aus- und Übersiedlern mit einer hohen Anzahl schulpflichtiger und bildungswilliger Kinder kommen, wenn nicht die Ausbildungskapazitäten verstärkt werden. Insbesondere im universitären Bereich zeichnet sich aufgrund der steigenden Anzahl von Abiturienten sowie des steigenden Anteils von Studierwilligen unter den Abiturienten keine Entlastung innerhalb des nächsten Jahrzehnts ab.

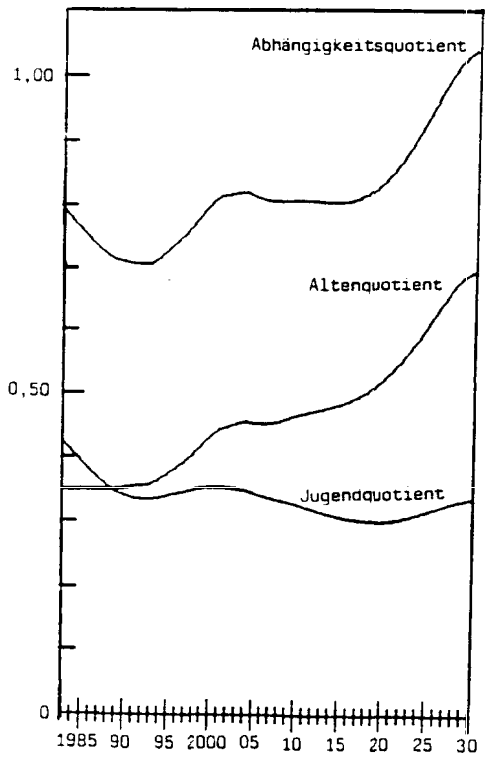
Von zentraler Bedeutung für das Angebot an Arbeitskräften ist das Erwerbspotential der Bevölkerung, das parallel zum Sinken der Gesamtbevölkerung im Prognosezeitraum bis 2030 von ca. 30 Millionen im Jahr 1990 auf 20 Millionen im Jahr 2030 fällt. Insgesamt wird dem Zusammenhang zwischen demographischer Struktur und Arbeitsangebot in der gesamten Diskussion um Arbeitslosigkeit und Beschäftigung viel zu wenig Beachtung geschenkt. Die derzeitige seit Jahren anhaltende hohe Arbeitslosigkeit von ca. 2 Millionen Personen ist zu einem hohen Ausmaß auf das Anwachsen des Bevölkerungsanteils der 20 bis 60jährigen von 30.9 Millionen im Jahre 1970 auf 34.5 Millionen im Jahre 1985 zurückzuführen. Diese Erhöhung des Erwerbspotentials konnte auch durch die zusätzliche Bereitstellung von Millionen von Arbeitsplätzen im letzten Jahrzehnt nicht vollständig aufgefangen werden.

Zum Abschluß gehen wir auf einen wichtigen Faktor der ökonomischen Stabilität ein, nämlich auf das Verhältnis von ökonomisch abhängigen Personen (Jugendliche unter 20 und Personen über 60 Jahre) zu den Personen im erwerbsfähigen Alter (Erwerbsfähige). Die für diesen Vergleich wichtigen Kennzahlen sind:

1. Der Jugendquotient wird durch das Verhältnis zwischen der Zahl der Jugendlichen und der Zahl der Erwerbsfähigen definiert.
2. Der Altenquotient wird durch das Verhältnis zwischen der Zahl der Senioren und der Zahl der Erwerbsfähigen definiert.
3. Der Abhängigkeitsquotient wird durch das Verhältnis zwischen der Zahl der Jugendlichen und Senioren und der Zahl der Erwerbsfähigen definiert.

Die nächste Abbildung zeigt die Prognose der Entwicklung dieser drei Kennzahlen im Zeitraum 1983 bis 2030. Dem stagnierenden Jugendquotienten steht ein dramatisch anwachsender Altenquotient und damit ein steil ansteigender Abhängigkeitsquotient gegenüber. Die Konsequenzen für das System der sozialen Sicherung sind offensichtlich. Die Belastung der 20 bis 60jährigen durch die Erfüllung des Generationenvertrages, auf dem das derzeitige System der sozialen Sicherung beruht, steigt erheblich oder die Zuwendungen an die über 60jährigen sinken erheblich. Natürlich sind alle Formen des Kompromisses zwischen diesen beiden Extremen denkbar. Es ist jedoch klar, daß die 1989 im Bundestag verabschiedete Rentenreform nur ein bescheidener Vorgeschmack der Kürzungen und Belastungen sind, die ab dem Jahr 2000 auf die Bundesrepublik Deutschland und andere Industrieländer zukommen.

Abbildung: Entwicklung des Jugendquotienten, des Altenquotienten und des Abhängigkeitsquotienten 1983 bis 2030.



12.3 Erwerbsstatistik

In der Erwerbsstatistik unterscheidet man zwischen der Beteiligung am Erwerbsleben (Erwerbskonzept) und der Hauptquelle des Lebensunterhalts (Unterhaltskonzept). *Erwerbspersonen* sind alle Personen mit Wohnsitz im Bundesgebiet (Inländerkonzept), die eine unmittelbar oder mittelbar auf Erwerb gerichtete Tätigkeit ausüben oder suchen, unabhängig von der Bedeutung des Ertrags für ihren Lebensunterhalt und ohne Rücksicht auf die von ihnen tatsächlich geleistete oder vertragsmäßig zu leistende Arbeitszeit. *Erwerbstätige* sind Personen, die in einem Arbeitsverhältnis stehen (einschließlich Soldaten und mithelfender Familienangehöriger) oder selbständig ein Gewerbe oder eine Landwirtschaft betreiben oder einen freien Beruf ausüben. *Erwerbslose* sind Personen ohne Arbeitsverhältnis, die sich um eine Arbeitsstelle bemühen, unabhängig davon, ob sie beim Arbeitsamt gemeldet sind. *Nichterwerbspersonen* sind alle Personen, die keinerlei auf Erwerb gerichtete Tätigkeit ausüben oder suchen. (Dazu gehören auch Studenten und Rentner.) Das Unterhaltskonzept gliedert nach Unterhalt vor allem aus Erwerbstätigkeit, Arbeitslosengeld, Rente u. dgl. (Vermögenserträge) sowie durch Angehörige. *Arbeitslose* sind Personen, die sich als Arbeitssuchende beim Arbeitsamt gemeldet haben. Offene Stellen sind zu besetzende Arbeitsplätze, die durch Arbeitgeber beim Arbeitsamt gemeldet sind. Die Arbeitslosenquote ist die Anzahl der Arbeitslosen dividiert durch die Anzahl der abhängigen Erwerbstätigen (ohne Soldaten). Es gibt mehrere Definitionen für die Arbeitslosenquote. Die Zahl, die variiert, ist die, durch die dividiert wird. Man kann die Arbeitslosenquote z. B. verringern, indem man nicht durch die Anzahl der abhängigen Erwerbstätigen sondern aller Erwerbstätigen dividiert. Da diese Quote ein starkes politisches Mittel ist, sollte man immer beachten, wie sie definiert ist. Dieses muß man auch beim Vergleich zwischen verschiedenen Ländern tun.

Die folgenden Tabellen fassen einige Ergebnisse der Erwerbs– und Beschäftigtenstatistik zusammen.

Tabelle: Wohnbevölkerung nach dem Erwerbskonzept in 1000 *(Alte Bundesländer)*					
	1989	1990	1991	1992	1993
Wohnbevölkerung	62 063	63 253	64 074	64 865	65 532
Erwerbspersonen, davon	29 799	30 369	30 663	30 938	30 935
Selbständige und mithelfende Familienangehörige	3 011	3 026	3 054	3 067	3 064
Beschäftigte Arbeitnehmer	24 750	25 460	25 920	26 063	25 692
Arbeitslose (als Ersatz für Erwerbslose)	2 038	1 883	1 689	1 808	2 270

Die Verlagerung der Erwerbstätigkeit vom primären und sekundären Sektor in den tertiären Sektor läßt sich an folgender Tabelle ablesen.

Tabelle: Erwerbstätige nach Wirtschaftsbereichen in 1000 *(Alte Bundesländer)*					
	1989	1990	1991	1992	1993
Land- und Forstwirtschaft	1 028	995	970	930	879
Warenproduzierendes Gewerbe	10 997	11 309	11 457	11 303	10 777
Handel und Verkehr	5 158	5 314	5 547	5 655	5 605
Dienstleistungsunternehmen	4 992	5 294	5 592	5 841	5 986
Unternehmen gesamt	22 175	22 912	23 560	23 729	23 247

Von besonderem Interesse sind die Statistiken der Arbeitsämter. Die Entwicklung der Arbeitslosenzahlen in den letzten Jahren läßt sich aus der folgenden Tabelle ablesen.

Tabelle: Arbeitslose *(Alte Bundesländer)*				
Jahresdurchschnitt	Bundesgebiet insgesamt	Arbeitslosenquote	Frauen	Ausländer
1989	2 037 781	7.9	967 978	232 512
1990	1 883 147	7.2	915 404	202 975
1991	1 689 365	6.3	791 688	208 094
1992	1 808 310	6.6	825 531	254 201
1993	2 270 349	8.2	993 261	344 840

Am Ende dieses Abschnitts geben wir einen Überblick über die Bruttoverdienste von Arbeitern und Angestellten an (alte Bundesländer).

Tabelle: Durchschnittliche Bruttojahresverdienste in DM für 1990

	Arbeiter Männer	Arbeiter Frauen	Angestellte Männer	Angestellte Frauen
Bergbau	48 722		76 452	52 034
Grundstoff- u. Produktionsgüterindustrie	50 542	37 081	77 061	53 589
Investitionsgüterindustrie	50 522	35 636	76 936	49 205
Verbrauchsgüterindustrie	45 726	30 895	66 371	43 450
Nahrungs- u. Genußmittelindustrie	47 711	30 884	66 567	45 780

12.4 Indexrechnung

Indizes sind Kennzahlen zur Charakterisierung der zeitlichen Entwicklung quantitativer Größen. Typische Beispiele sind Preis- und Umsatzindizes.

Einfache Indizes

Gegeben ist eine Zeitreihe von Größen. Die Größe wird mit G bezeichnet, während die Zeitreihe durch die Folge $G_0, G_1, \ldots G_t$ repräsentiert wird. *Definition*: Das Verhältnis

$$I_{0-t}(G) = \frac{G_t}{G_0}$$

heißt Meßzahl oder einfacher Index von G auf Basis 0. Dabei werden die Bezeichnungen G_t für absolute Werte, 0 für den Basiszeitpunkt (Zeitraum) und t für den Berichtszeitpunkt (Zeitraum) verwendet. Meßzahlen werden häufig in Prozenten angegeben:

$$I_{0-t}(G) = \frac{G_t}{G_0} \cdot 100$$

Ein zentrales Problem ist die Umstellung von Indizes auf einen neuen Basiszeitraum. Dieser Vorgang wird als Umbasierung von Meßzahlreihen bezeichnet. Gegeben seien die absoluten Werte $G_0, G_1, \ldots G'_t, \ldots G_t$ und die Indizes $1, I_{0-1}(G), \ldots, I_{0-t'}(G), \ldots, I_{0-t}(G)$ auf Basis 0. Ein Index kann ohne Kenntnis der absoluten Werte auf die neue Basis t' umgestellt werden:

$$I_{t'-t}(G) = \frac{I_{0-t}(G)}{I_{0-t'}(G)}$$

Die letzte Formel folgt aus der Identität:

$$I_{t'-t}(G) = \frac{G_t}{G_{t'}} = \frac{G_t/G_0}{G_{t'}/G_0} = \frac{I_{0-t}(G)}{I_{0-t'}(G)}$$

Daraus ergibt sich die Kettenformel:

$$I_{0-t}(G) = I_{0-t'}(G) I_{t'-t}(G)$$

Preis- und Mengenindexzahlen

Zur Zusammenfassung der Entwicklung mehrerer Größen $G^{(i)}, i = 1, \ldots m$ werden zusammengesetzte (gewichtete) Indizes benutzt. Folgende Bezeichnungen werden verwendet:

Symbol	Bezeichnung
p	Preis
q	Menge
$u = p \cdot q$	Wert (Umsatz, Ausgaben)
0	Basiszeitpunkt
1	Berichtszeitpunkt
$p_t^{(i)}$	Preis pro Einheit der $i$ten Ware zum Zeitpunkt t
$q_t^{(i)}$	Menge der $i$ten Ware zum Zeitpunkt t

Ein Preisindex aller betrachteten Waren in einem Warenkorb $\{q_t^{(1)}, q_t^{(2)}, \ldots q_t^{(m)}\}$ zum Zeitpunkt t läßt sich auf folgende Arten konstruieren.

1. Arithmetisches Mittel der Preismeßzahlen:

$$I_{0-1}(p) = \frac{1}{m} \sum_{i=1}^{m} \frac{p_1^{(i)}}{p_0^{(i)}}$$

   Problem: Keine Berücksichtigung der Mengen.

2. Index des mit den Mengen gewichteten arithmetischen Mittels der Preise:

$$I_{0-1}(p) = \frac{\sum_i^m p_1^{(i)} q_1^{(i)} / \sum_i^m q_1^{(i)}}{\sum_i^m p_0^{(i)} q_0^{(i)} / \sum_i^m q_0^{(i)}} = \frac{\sum_i^m p_1^{(i)} q_1^{(i)}}{\sum_i^m p_0^{(i)} q_0^{(i)}} \cdot \frac{\sum_i^m q_0^{(i)}}{\sum_i^m q_1^{(i)}} \quad .$$

   Problem: unterschiedliche Mengen werden berücksichtigt (Änderung des Konsumverhaltens).

3. Mengen werden konstant gehalten: Werden die Mengen der Basisperiode konstant gehalten, erhält man den Preisindex von *Laspeyres*:

$$I_{0-1}^{L}(p) = \frac{\sum_i^m p_1^{(i)} q_0^{(i)}}{\sum_i^m p_0^{(i)} q_0^{(i)}} \quad .$$

   Werden die Mengen der Berichtsperiode konstant gehalten, erhält man den Preisindex von *Paasche*:

$$I_{0-1}^{P}(p) = \frac{\sum_i^m p_1^{(i)} q_1^{(i)}}{\sum_i^m p_0^{(i)} q_1^{(i)}} \quad .$$

Ähnlich wie bei der Konstruktion eines Preisindex kann bei der Konstruktion eines Mengenindex verfahren werden.

1. Outputmeßzahl

$$I_{0-1}(q) = \frac{\sum_i^m q_1^{(i)}}{\sum_i^m q_0^{(i)}}$$

2. Umsatzmeßzahl

$$I_{0-1}(p \cdot q) = \frac{\sum_i^m q_1^{(i)} p_1^{(i)}}{\sum_i^m q_0^{(i)} p_0^{(i)}}$$

3. Mengenindex nach *Laspeyres*

$$I_{0-1}^L(q) = \frac{\sum_i^m p_0^{(i)} q_1^{(i)}}{\sum_i^m p_0^{(i)} q_0^{(i)}}$$

4. Mengenindex nach *Paasche*

$$I_{0-1}^P(q) = \frac{\sum_i^m p_1^{(i)} q_1^{(i)}}{\sum_i^m p_1^{(i)} q_0^{(i)}}$$

Durch die beiden letzten Indizes wird die Änderung von Warenkörben zu konstanten Preisen gemessen.

Beispiel zur Indexrechnung: Im Zeitraum von 4 Jahren hat man folgende Preis- und Mengenentwicklung beim durchschnittlichen Verbrauch von 3 Gütern gefunden:

Zeitpunkt	$t = 0$		$t = 1$		$t = 2$		$t = 3$	
	$q_0^{(i)}$	$p_0^{(i)}$	$q_1^{(i)}$	$p_1^{(i)}$	$q_2^{(i)}$	$p_2^{(i)}$	$q_3^{(i)}$	$p_3^{(i)}$
Zigaretten (Stück)	476	0.12	553	0.11	598	0.13	709	0.16
Limonade (Liter)	21	1.1	25	1.25	30	1.2	29	1.2
Kaffee (kg)	0.6	12	0.8	13	1.2	14	1.3	15

Preisindizes von Laspeyres für Basiszeitpunkt 0 und Berichtszeiten 2 und 3:

$$I_{0-2}^L(p) = \frac{476 \cdot 0.13 + 21 \cdot 1.2 + 0.6 \cdot 14}{476 \cdot 0.12 + 21 \cdot 1.1 + 0.6 \cdot 12} = 1.092$$

$$I_{0-3}^L(p) = \frac{476 \cdot 0.16 + 21 \cdot 1.2 + 0.6 \cdot 15}{476 \cdot 0.12 + 21 \cdot 1.1 + 0.6 \cdot 12} = 1.262$$

Preisindex und Mengenindex von Paasche für Basiszeit 0 und Berichtszeit 1:

$$I_{0-1}^{P}(p) = \frac{553 \cdot 0.11 + 25 \cdot 1.25 + 0.8 \cdot 13}{553 \cdot 0.12 + 25 \cdot 1.1 + 0.8 \cdot 12} = 0.9905$$

$$I_{0-1}^{P}(q) = \frac{553 \cdot 0.11 + 25 \cdot 1.25 + 0.8 \cdot 13}{476 \cdot 0.11 + 21 \cdot 1.25 + 0.6 \cdot 13} = 1.1859$$

Umsatzmeßzahl für Basiszeitpunkt 1 und Berichtszeit 2:

$$I_{1-2}(p \cdot q) = \frac{598 \cdot 0.13 + 30 \cdot 1.2 + 1.2 \cdot 14}{553 \cdot 0.11 + 25 \cdot 1.25 + 0.8 \cdot 13} = 1.2738$$

Erweiterung des Indexschemas

Bei der Berechnung des Preisindex nach Laspeyres wird von der Annahme ausgegangen, daß die Warenkörbe zum Basiszeitpunkt und zum Berichtszeitpunkt qualitativ und mengenmäßig gleich sind. Problematisch ist daher die Berücksichtigung von Waren, die erst nach dem Basiszeitpunkt auf dem Markt eingeführt werden (z. B. CD–Player). Sei 0 der Basiszeitpunkt mit Warenkorb $\{q_0^{(1)}, q_0^{(2)}, \ldots, q_0^{(m)}\}$. Sei 1 der Zeitpunkt der Einführung einer neuen Ware und 2 sei der Berichtszeitpunkt. Da die Ware $m + 1$ zu 0 noch nicht existiert hat, ist sie (aber nur sie) im Warenkorb mit der zum Zeitpunkt 1 gültigen Menge repräsentiert. Somit gilt zum Zeitpunkt 1 der Warenkorb $\{q_0^{(1)}, q_0^{(2)}, \ldots, q_0^{(m)}, q_1^{(m+1)}\}$.

Man berechnet einen Index I_{0-1} mit dem alten Warenkorb

$$I_{0-1} = \frac{\sum_{i}^{m} p_1^{(i)} q_0^{(i)}}{\sum_{i}^{m} p_0^{(i)} q_0^{(i)}}$$

sowie einen Index I_{1-2} mit dem um die neue Ware erweiterten Korb:

$$I_{1-2} = \frac{\sum_{i}^{m} p_2^{(i)} q_0^{(i)} + p_2^{(m+1)} q_1^{(m+1)}}{\sum_{i}^{m} p_1^{(i)} q_0^{(i)} + p_1^{(m+1)} q_1^{(m+1)}}$$

Der gesuchte Index wird durch Verkettung ermittelt:

$$I_{0-2}^{v} = I_{0-1} \cdot I_{1-2} \quad .$$

Durch dieses Verfahren wird erstens gewährleistet, daß die Preisentwicklung der ursprünglichen m Waren normal nach Laspeyres berechnet wird. Zweitens wird für die zum Zeitpunkt 1 eingeführte Ware ein fiktiver Preis $p_0^{(m+1)}$ ermittelt, für den gilt:

$$\frac{p_1^{(m+1)}}{p_0^{(m+1)}} = I_{0-1}$$

Damit wird angenommen, daß die (hypothetische) Preisentwicklung des Gutes $m + 1$ mit der durchschnittlichen Preisentwicklung der Waren 1 bis m übereinstimmt.

Beispiel zur Erweiterung: Im vorigen Beispiel wird zum Zeitpunkt $t = 2$ ein neues Gut mit $q_2^{(4)} = 2.2$ und $p_2^{(4)} = 2.13$ bzw. $p_3^{(4)} = 2.29$ eingeführt. Man berechne $I_{0-3}^L(p)$ unter Berücksichtigung des neuen Gutes.

$$I_{0-3}^L(p) = I_{0-2}^L(p) \cdot I_{2-3}^L(p) \quad \text{mit}$$

$$I_{0-2}^L(p) = 1.092$$

$$I_{2-3}^L(p) = \frac{476 \cdot 0.16 + 21 \cdot 1.2 + 0.6 \cdot 15 + 2.2 \cdot 2.29}{476 \cdot 0.13 + 21 \cdot 1.2 + 0.6 \cdot 14 + 2.2 \cdot 2.13} = 1.152$$

$$I_{0-3}^L(p) = 1.092 \cdot 1.152 = 1.258$$

Ausgewählte Indizes der wirtschaftlichen Entwicklung

Die Indexrechnung dient vor allem der übersichtlichen Darstellung der Entwicklung der gesamten Volkswirtschaft sowie einzelner Sektoren bzw. Branchen im Hinblick auf eine Reihe von Variablen, deren wichtigste in folgenden Teilbereichen zusammengefaßt werden (vgl. Abels (1976)):

1. *Preisentwicklung:* Preisindizes für die Lebenshaltung - Indizes der Erzeugnisse industrieller Produkte.
2. *Nachfrageentwicklung:* Umsatzindizes des Auftragseinganges - Indizes des Auftragsbestandes.
3. *Produktionsentwicklung:* Industrielle Produktion und Produktionswerte - Indizes der industriellen Nettoproduktion - Indizes der industriellen Bruttoproduktion.
4. *Einkommensentwicklung*: Indizes der Effektivverdienste - Indizes der Tarifverdienste.
5. *Arbeitsproduktivität:* Produktivitätsindizes.
6. *Außenhandel:* Außenhandelswerte - Außenhandelsvolumen - Außenhandelsindizes.

Als Indikator der (Verbraucher)-Preisentwicklung werden Preisindizes sowohl für den durchschnittlichen Haushalt (2.7 Personen, 0.7 Kinder unter 18 Jahren) als auch typische Haushalte wie Haushalte von Einzelpersonen, Vierpersonenhaushalte (Eltern, zwei Kinder), etc. berechnet. Grundlage dieser Indizes sind einerseits die Ergebnisse der im Abstand von 3 bis 5 Jahren durchgeführten Einkommens- und Verbrauchsstichprobe, andererseits aber auch laufende Wirtschaftsberechnungen ausgewählter privater Haushalte.

Preisindex (Laspeyres) für die Lebenshaltung aller privaten Hauptgruppen (Basis 1985)						
	Gewichtung	1989	1990	1991	1992	1993
Lebenshaltung insgesamt	1 000	104.2	107.0	110.7	115.1	119.9
insgesamt ohne saison- und witterungsabhängige Nahrungsmittel	984.05	104.2	106.9	110.5	115.1	120.0
Nahrungsmittel und Genuß	229.89	102.6	105.6	108.6	112.1	114.1
Kleidung, Schuhe	69.47	106.0	107.5	110.1	113.3	116.4
Wohnungsmieten	117.77	108.8	112.5	117.3	123.8	131.1
Energie ohne Kraftstoffe	72.52	82.0	85.5	89.4	89.8	91.0
Waren u. Dienstleistungen für die Haushalte	72.21	104.9	107.3	110.5	114.4	118.0
Waren u. Dienstleistungen für Gesundheits- u. Körperpflege	40.99	108.6	110.3	113.8	117.9	122.4
Waren u. Dienstleistungen für Verkehrszwecke	144.03	103.3	106.1	112.1	117.1	121.7
Waren u. Dienstleistungen für Bildungs- u. Unterhaltungszwecke	83.71	103.8	106.7	108.1	112.7	115.9
Persönliche Ausstattung	109.41	113.1	115.3	118.0	124.3	135.2

Anhang

Mengenlehre

Eine Menge ist eine Zusammenfassung von Objekten. Eine Menge ist definiert, wenn von jedem beliebigen Objekt feststeht, ob es zur Menge gehört oder nicht. Die zur Menge gehörenden Objekte heißen *Elemente* dieser Menge. $a \in A$ bedeutet, daß a ein Element der Menge A ist. $a \notin A$ heißt, daß a kein Element der Menge A ist. Die Menge, die kein Element enthält, heißt *leere Menge* und wird mit $\{\ \}$ oder $\emptyset$ bezeichnet. $A = \{a_1, \ldots, a_n\}$ bedeutet, daß A aus den Elementen $a_1, \ldots, a_n$ besteht. Ist eine Menge A dadurch bestimmt, daß ihre Elemente die Eigenschaft E besitzen, so schreibt man $A = \{a|E(a)\}$.

Zwei Mengen A und B heißen gleich ($A = B$), wenn sie die selben Elemente enthalten. A heißt *Teilmenge* von B, $A \subset B$, wenn jedes Element von A auch zu B gehört. A ist genau dann gleich B, wenn gilt: $A \subset B$ und $B \subset A$.

Die *Vereinigung* von A und B ist die Menge der Elemente, die zu A oder B gehören:

$$A \cup B = \{a|a \in A \text{ oder } a \in B\}$$

Der *Durchschnitt* von A und B ist die Menge der Elemente, die sowohl zu A als auch zu B gehören:

$$A \cap B = \{a|a \in A \text{ und } a \in B\}$$

Die Vereinigung der Mengen $A_n, n \in \mathcal{N}$, ist die Menge der Elemente, die mindestens zu einer A_n gehören.

$$\cup_{n \in \mathcal{N}} A_n = \{a|a \in A_n \text{ für mindestens ein } n \in \mathcal{N}\}$$

Der Durchschnitt der Mengen $A_n, n \in \mathcal{N}$, ist die Menge der Elemente, die zu allen A_n gehören.

$$\cap_{n \in \mathcal{N}} A_n = \{a|a \in A_n \text{ für alle } n \in \mathcal{N}\}$$

A und B heißen *disjunkt*, wenn gilt: $A \cap B = \emptyset$

Für beliebige Teilmengen $A \subset \Omega$ ist $\bar{A} = \{a|a \notin A \text{ und } a \in \Omega\}$ das *Komplement* von A.

$A_1, \ldots, A_n$ bilden eine *Zerlegung* der Menge Ω, wenn gilt:

1. $A_i \cap A_j = \emptyset$ für $i \neq j$ $1 \leq i, j \leq n$
2. $A_1 \cup \ldots \cup A_n = \Omega$.

Einige Rechenregeln für Mengen:

$$\begin{aligned}
A \cup B &= B \cup A \\
(A \cup B) \cup C &= A \cup (B \cup C) \\
A \cap B &= B \cap A \\
(A \cap B) \cap C &= A \cap (B \cap C) \\
A \cup \emptyset &= A \\
A \cap \emptyset &= \emptyset \\
\overline{A \cup B} &= \overline{A} \cap \overline{B} \\
\overline{A \cap B} &= \overline{A} \cup \overline{B}
\end{aligned}$$

Produkte von Mengen

n nicht notwendig verschiedene Elemente $a_1, \ldots, a_n$ in einer bestimmten Reihenfolge bilden ein *n-Tupel* $(a_1, \ldots, a_n)$. (a_1, a_2) heißt ein *Paar*, (a_1, a_2, a_3) ein *Tripel*.

Ist $a_1 \neq a_2$, so gilt $(a_1, a_2) \neq (a_2, a_1)$, da das Tupel die Reihenfolge von a_1 und a_2 eindeutig festlegt.

$$A \times B = \{(a, b) | a \in A \text{ und } b \in B\}$$

heißt *kartesisches Produkt* der Mengen A und B.

Für $A \neq B$ ist $A \times B \neq B \times A$. Ist $A_1 \subset A_2$ und $B_1 \subset B_2$, so ist $A_1 \times B_1 \subset A_2 \times B_2$.

Das Summenzeichen

Ein Hilfsmittel zur Vereinfachung umfangreicher Formeln ist das *Summenzeichen* Σ. Man setzt:

$$a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i$$

i heißt Summationsindex; die Menge $(1, 2, \ldots, n,)$, über die der Summationsindex läuft, heißt *Summationsbereich*.

Offenbar gilt dann: Die Summe ist unabhängig von der Wahl des Summationsindex.

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j$$

Ein allen Summanden gemeinsamer Faktor c kann vor die Summe gezogen werden (Distributivgesetz).

$$\sum_{i=1}^{n} c a_i = c \sum_{i=1}^{n} a_i$$

Summen mit gemeinsamen Summationsbereich können zusammengezogen werden.

$$\sum_{i=1}^{n} a_i + \sum_{j=1}^{n} b_j + \sum_{k=1}^{n} c_k = \sum_{i=1}^{n} (a_i + b_i + c_i)$$

Hat man alle Elemente des zweifachen indizierten Zahlenschemas $(a_{ij} : i = 1 \ldots m; j = 1 \ldots n)$:

$$\begin{array}{llll} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{array}$$

zu summieren, dann läßt sich das mit Hilfe einer Doppelsumme leicht formulieren. Die Summe S ist gegeben durch:

$$S = \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij}) = \sum_{j=1}^{n} (\sum_{i=1}^{m} a_{ij})$$

Im ersten Fall wird zuerst, bei festem Zeilenindex i, über den Spaltenindex j summiert, und dann die so gewonnenen Zeilensummen von $i = 1$ bis $i = n$ aufsummiert. Im zweiten Fall geht man umgekehrt vor. Beide Formeln liefern aber offenbar das gleiche Resultat, nämlich die gewünschte Summe S aller a_{ij}. Es gilt die Regel:

$$\sum_{i=1}^{m}(\sum_{j=1}^{n} a_{ij}) = \sum_{j=1}^{n}(\sum_{i=1}^{m} a_{ij})$$

Da es auf die Summationsreihenfolge nicht ankommt, läßt man die (überflüssigen) Klammern fort und schreibt:

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

Selbstverständlich können diese Betrachtungen auf $k$fach indizierte Summanden ausgedehnt werden. Man hat dann Summanden der Form:

$$\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}\cdots\sum_{j_k=1}^{m_k} a_{j_1,j_2,\ldots j_k}$$

Hat man das Produkt der beiden Summen $\sum_{i=1}^{m} a_i$ und $\sum_{j=1}^{n} b_j$ zu bilden, so werden folgende Regeln (Produktregel) verwendet:

$$(\sum_{i=1}^{m} a_i)(\sum_{j=1}^{n} b_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j$$

Hingegen gilt im allgemeinen (mit Ausnahme von Spezialfällen):

$$(\sum_{i=1}^{m} a_i)(\sum_{j=1}^{m} b_j) \neq \sum_{i=1}^{m} a_i b_i$$

Für Dreifachprodukte gilt:

$$(\sum_{i=1}^{m} a_i)(\sum_{j=1}^{n} b_j)(\sum_{k=1}^{p} c_k) = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p} a_i b_j c_k$$

Für Potenzen gilt:

$$(\sum_{i=1}^{m} a_i)^2 = (\sum_{i=1}^{m} a_i)(\sum_{j=1}^{m} a_j) = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j$$

Allgemein:

$$(\sum_{i=1}^{m} a_i)^k = \sum_{i_1=1}^{m}\cdots\sum_{i_k=1}^{m} a_{i_1}\cdots a_{i_k}$$

Exponentialfunktion und Logarithmus

$\exp(x) = e^x$ heißt Exponentialfunktion und ist definiert auf R. Es gilt: $\exp(1) = e = 2.718281\ldots$

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{2\cdot 3} + \frac{x^4}{2\cdot 3\cdot 4} + \ldots \text{ für } -\infty < x < \infty$$

Ableitung: $(e^x)' = e^x$
Dabei bezeichnet $f(x)'$ die erste Ableitung von f nach x.

Additionstheorem: $e^{x+y} = e^x e^y$

Der natürliche Logarithmus $\ln x$ ist definiert als die Umkehrfunktion der Exponentialfunktion, also durch

$$\exp(\ln x) = \ln e^x = x$$

Da e^x nur positive Werte annehmen kann, ist der Definitionsbereich von $\ln x$ die positive reelle Zahlengerade $(0, \infty)$.

ln besitzt die folgende Reihenentwicklung:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots \text{ für } -1 < x \leq 1$$

Ableitung: $(\ln x)' = \frac{1}{x}$

Rechenregel: $\ln(x \cdot y) = \ln x + \ln y$

Mit $\log x$ wird meistens die Umkehrfunktion von 10^x bezeichnet.

Differential- und Integralrechnung

Differentialrechnung

$f(x)$ sei eine stetige Funktion mit dem Definitionsbereich

$I = (a, b), (a, \infty), (-\infty, b)$ oder $\mathcal{R}$. Existiert

$$\lim_{x \to x_0;\, x \neq x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

so heißt $f(x)$ an der Stelle x_0 differenzierbar.

$$f'(x_0) = \lim_{x \to x_0;\, x \neq x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

heißt Ableitung (oder Differentialquotient) von $f(x)$ an der Stelle x_0.

Ist $f(x)$ in jedem Punkt $x_0 \in I$ differenzierbar, so heißt $f(x)$ differenzierbar in I und $f'(x)$ oder $\frac{d}{dx} f(x)$ Ableitung von $f(x)$ in I. Anschaulich ist $f'(x_0)$ die Steigung $\frac{\Delta y_0}{\Delta x_0}$ der Tangente an der durch die Gleichung $y = f(x)$ bestimmten Kurve im Punkt x_0. Einige wichtige Regeln sind:

$$(c \cdot f(x))' = c \cdot f'(x), c \in \mathcal{R}$$

$$(f_1(x) + f_2(x))' = f_1'(x) + f_2'(x),$$

$$(f_1(x) \cdot f_2(x))' = f_1'(x) f_2(x) + f_1(x) f_2'(x)$$

$$\left(\frac{f_1(x)}{f_2(x)}\right)' = \frac{f_1'(x) f_2(x) - f_2'(x) \cdot f_1(x)}{f_2^2(x)} \quad \text{für} \quad f_2(x) \neq 0$$

Sind f_1 und f_2 zwei differenzierbare Funktionen, für die der Definitionsbereich von f_2 den Wertebereich von f_1 enthält, so gilt die Kettenregel:

$$(f_2(f_1(x)))' = f_2'(f_1(x)) \cdot f_1'(x)$$

Beispiele:

$f(x)$	$f'(x)$	Definitionsbereich
c	0	$\mathcal{R}$
x	1	$\mathcal{R}$
x^2	$2x$	$\mathcal{R}$
x^n	$n\,x^{n-1}$	$\mathcal{R}$ für $n \neq 0$
e^x	e^x	$\mathcal{R}$
ln x	$\frac{1}{x}$	$(0, \infty)$
a^x	$a^x \ln a$	$\mathcal{R}$ für $a > 0$
e^{-x^2}	$e^{-x^2} \cdot (-2x)$	$\mathcal{R}$

Integralrechnung

Zunächst sei $f(x)$ eine positive stetige Funktion über dem Intervall $[a, b]$. $\int_a^b f(x)dx$ entspricht der Fläche unter der Kurve $(x, f(x))$ über dem Intervall $[a, b]$.

$f(x)$ sei nun eine beliebige Funktion. $F(x)$ heißt Stammfunktion von $f(x)$, falls in dem Definitionsbereich von $f(x)$ gilt:

$$F'(x) = f(x)$$

Jede stetige Funktion $f(x)$ besitzt eine Stammfunktion $F(x)$. Zwei Stammfunktionen einer Funktion $f(x)$ unterscheiden sich nur um eine additive Konstante.

Beispielsweise sind $x^3 + \frac{x^2}{2} + 4$ und $x^3 + \frac{x^2}{2}$ Stammfunktionen von $3x^2 + x$.

Ist $F(x)$ eine Stammfunktion von $f(x)$ und liegt $[a, b]$ im Definitionsbereich von $F(x)$, so ist

$$\int_a^b f(x)dx = F(b) - F(a).$$

Existieren $\lim_{b \to \infty} F(b)$ oder $\lim_{a \to -\infty} F(a)$ oder beide Grenzwerte so gilt

$$\int_a^\infty f(x)dx = \lim_{b \to \infty} F(b) - F(a) \text{ bzw.}$$

$$\int_{-\infty}^b f(x)dx = F(b) - \lim_{a \to -\infty} F(a) \text{ bzw.}$$

$$\int_{-\infty}^\infty f(x)dx = \lim_{b \to \infty} F(b) - \lim_{a \to -\infty} F(a)$$

Beispiel: Das bestimmte Integral $\int_2^4 x\,dx$ ist zu berechnen. $F(x) = x^2/2$ ist eine Stammfunktion von $f(x) = x$. Daher gilt

$$\int_2^4 x\,dx = F(4) - F(2) = \frac{4^2}{2} - \frac{2^2}{2} = 6$$

Einige wichtige Regeln: Es seien $F(x)$ und $G(x)$ die Stammfunktionen von $f(x)$ bzw. $g(x)$ und $a, b, c, d \in \mathcal{R}$

$$\int_a^b c \cdot f(x) + d \cdot g(x)dx = c\int_a^b f(x)dx + d \cdot \int_a^b g(x)dx$$

$$\int_a^z f(x)dx + \int_z^b f(x)dx = \int_a^b f(x)dx \text{ für } a \le z \le b$$

$$\int_a^b f(x)g(x)dx = (F(b) \cdot g(b) - F(a) \cdot g(a)) - \int_a^b F(x)g'(x)dx$$

$f(x)$	$F(x)$	Beispiel
x^n	$\frac{x^{n+1}}{n+1}$	$\int_{-1}^4 x^3 dx = \frac{4^4}{4} - \frac{(-1)^4}{4} = 63\frac{3}{4}$
e^{cx}	$\frac{e^{cx}}{c}$ für $c \neq 0$	$\int_0^{10} e^{-x}dx = -e^{-10} - (-e^0) = 1 - e^{10}$

Beispiele für partielle Integration: Zu berechnen ist:

$$\int_0^3 x^2 e^{2x} dx$$

Hier setzt man $f(x) = e^{2x}$ und $g(x) = x^2$ und erhält:

$$\begin{aligned}\int_0^3 x^2 e^{2x} dx &= \left(3^2 \cdot \frac{e^{2\cdot 3}}{2} - 0^2 \cdot \frac{e^{2\cdot 0}}{2}\right) - \int_0^3 2x \cdot \frac{e^{2x}}{2} dx \\ &= 4.5 \cdot e^6 - \int_0^3 x e^{2x} dx\end{aligned}$$

Neuerliche Anwendung der partiellen Integration mit $f(x) = e^{2x}$ und $g(x) = x$ ergibt:

$$\begin{aligned}\int_0^3 x e^{2x} dx &= \left(3 \cdot \frac{e^{2\cdot 3}}{2} - 0 \cdot \frac{e^{2\cdot 0}}{2}\right) - \int_0^3 1 \cdot \frac{e^{2x}}{2} dx \\ &= 1.5 \cdot e^6 - \left(\frac{e^{2\cdot 3}}{4} - \frac{e^{2\cdot 0}}{4}\right) \\ &= 1.5e^6 - 0.25e^6 + 0.25\end{aligned}$$

Somit ist

$$\int_0^3 x^2 e^{2x} dx = 3.25e^6 - 0.25$$

Matrizenrechnung

Begriff der Matrix

Eine Matrix ist ein Rechteckschema von Zahlen; z. B.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} -1 & 3 & 5 \\ 8 & 10 & 76 \end{pmatrix} \quad .$$

Matrizen werden mit großen Buchstaben bezeichnet A, B, Σ usw.. Eine Matrix besitzt m Zeilen und n Spalten. Man schreibt auch, die Matrix hat die Ordnung $m \times n$. Allgemein wird die $m \times n$ Matrix A so dargestellt:

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

Ein einzelnes Element bezeichnet man mit a_{ij}, i bezeichnet die Zeile und j die Spalte, in der sich das Element befindet.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad \Longrightarrow \quad a_{11} = 1, \quad a_{22} = 5, \quad a_{23} = 6, \quad a_{31} = 7 \quad .$$

Wir nennen Matrizen, die nur aus einer Zeile oder einer Spalte bestehen, Zeilenvektoren bzw. Spaltenvektoren. Die in der $i$ten Zeile von A stehenden Elemente $a_{i1}, \dots, a_{in}$ bilden somit den $i$ten Zeilenvektor.

$$a_{i.} = (a_{i1}, \dots, a_{in})$$

von A. Die in der $j$ten Spalte von A stehenden Elemente $A_{1j}, \dots, a_{mj}$ bilden den j-ten Spaltenvektor

$$a_{.j} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Definition: Zwei $m \times n$ Matrizen A und B heißen gleich, wenn die einander entsprechenden Elemente gleich sind, d.h. $A = B \Longleftrightarrow a_{ij} = b_{ij} \;\; (i = 1, \dots, m; \; j = 1, \dots, n)$.

Definition: Eine Matrix, deren Elemente sämtlich gleich Null sind, heißt Nullmatrix 0.

Definition: Eine quadratische Matrix, deren Elemente außerhalb der Hauptdiagonalen Null sind ($a_{ij} = 0$ für $i \neq j$), wird Diagonalmatrix genannt:

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

Die Elemente der Hauptdiagonale können ebenfalls gleich Null sein.

Definition: Eine quadratische Matrix, deren Elemente außerhalb der Hauptdiagonalen Null und deren Diagonalelemente alle gleich 1 sind, heißt Einheitsmatrix I.

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Ist $m = n = 1$, d.h. A hat nur ein einziges Element, so ist a_{11} ein Skalar (eine reelle Zahl). Schreibweise: $A_{11} = (a_{11}) = (a) = a$ (Skalare werden mit kleinen Buchstaben bezeichnet.)

Einfache Rechenregeln

Für das allgemeine Rechnen mit Matrizen werden die folgenden Regeln gesetzt:

Definition: Sind $A = (a_{ij})$ und $B = (b_{ij})$ zwei Matrizen von je m-Zeilen und n-Spalten, so wird als Summe (Differenz) von A, B die $m \times n$-Matrix $C = A \pm B = (c_{ij})$ mit $c_{ij} = a_{ij} \pm b_{ij}$ erklärt. Bemerkung: Die Summe (Differenz) zweier Matrizen mit ungleicher Ordnung ist nicht definiert!

$$\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \pm \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} \pm b_{11} & \cdots & a_{1n} \pm b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & \cdots & a_{mn} \pm b_{mn} \end{pmatrix} \quad .$$

Die Addition ist kommutativ:

$$A + B = B + A \quad .$$

Die Addition ist assoziativ:

$$A + (B + C) = (A + B) + C = A + B + C \quad .$$

Setzt man in der Summendefinition $B = A$ und schreibt, wie naheliegend, $A + A = 2A$, so kommt man verallgemeinernd zur nächsten Regel.

Definition: Das Produkt kA oder Ak einer $m \times n$-Matrix A mit einer Zahl k (einem Skalar) ist die $m \times n$-Matrix, bei der jedes Element das k-fache des entsprechenden von A ist

$$kA = Ak = \begin{pmatrix} ka_{11} & \cdots & ka_{1n} \\ \vdots & \ddots & \vdots \\ ka_{m1} & \cdots & ka_{mn} \end{pmatrix} \quad .$$

Für das Zahlenprodukt einer Matrix gilt:

- $kA + kB = k(A + B)$
- $kA + lA = (k + l)A$
- $k(lA) = (klA) = (lk)A = l(kA)$

Transponierte Matrix, symmetrische Matrix

Für Operationen mit Matrizen ist es erforderlich, Zeilen und Spalten der Matrix zu vertauschen. Dazu führen wir die Transposition von Matrizen ein.

Definition: Die Transponierte A^T einer $(m \times n)$ Matrix A ist diejenige $(n \times m)$ Matrix, die aus A durch Vertauschen der Zeilen und Spalten hervorgeht. Bezeichnen wir die Elemente von A^T mit a_{ij}^T und die von A wie üblich mit a_{ij} so gilt:

$$a_{ij} = (a^T)_{ji} \quad (i = 1, \ldots, n, k = 1, \ldots, m) \quad .$$

Offenbar besteht die Beziehung $(A^T)^T = A$

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}, \quad A^T = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}$$

Definition: Eine quadratische Matrix A heißt symmetrisch, wenn gilt $A^T = A$, d.h. $a_{ij} = a_{ji}$ $(i, j = 1, \ldots, n)$. Die nächste Matrix ist ein Beispiel.

$$A = \begin{pmatrix} -1 & 3 & -1 \\ 3 & 4 & 5 \\ -1 & 5 & 0 \end{pmatrix} = A^T$$

Matrizenmultiplikation

Definition: Das Produkt AB der $(m \times n)$-Matrix A mit der $n \times p$-Matrix B ist diejenige $(m \times p)$-Matrix $C = (c_{il})$, für die gilt:

$$c_{il} = \sum_{j=1}^{n} a_{ij} b_{jl} \quad (i = 1, \ldots m; l = 1, \ldots, p).$$

Damit sehen wir, daß eine Multiplikation nicht zwischen beliebigen Matrizen möglich ist, sondern daß die Spaltenzahl von A mit der Zeilenzahl von B übereinstimmen muß.

$$A = \begin{pmatrix} 2 & 4 & 3 \\ -1 & 0 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 3 \\ 2 & -4 \\ 3 & -2 \end{pmatrix} \quad .$$

A ist ein (2×3)-Matrix und B eine (3×2)-Matrix. Daher existiert AB, und zwar gilt:

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

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

$$\begin{pmatrix} 15 & -16 \\ 16 & -13 \end{pmatrix}$$

Unter Verwendung der Zeilenvektoren von $a_{i\cdot}$ von A und der Spaltenvektoren $b_{\cdot j}$ von B kann man die Multiplikation auch in der Form

$$c_{il} = a_{i\cdot} \cdot b_{\cdot l} = \sum_{j=1}^{n} a_{ij} \cdot b_{jl}$$

schreiben.

Für die Matrizenmultiplikation gelten die folgenden Regeln:

1. $A(BC) = (AB)C$ (Assoziativgesetz)
2. $A(B + C) = AB + AC$ (Distributivgesetz)
3. $(A + B)C = AC + BC$ (Distributivgesetz)
4. $k(AB) = (kA) \cdot B = A(kB)$ (Assoziativgesetz für die Skalarmultiplikation)

Eine Matrix heißt *idempotent*, wenn gilt:

$$A^2 = A \cdot A = A$$

Für das Rechnen mit transponierten Matrizen gelten die folgenden Gesetze:

1. $A^{T^T} = A$ zweimalige Transposition hebt sich auf.
2. $(A + B)^T = A^T + B^T$.
3. $(kA)^T = kA^T$ k ein Skalar.
4. $(AB)^T = B^T A^T$.

Für jede beliebige Matrix A sind die Matrizen AA^T und $A^T A$ stets definiert, quadratisch und symmetrisch. Der Beweis ergibt sich aus (1) und (4).

Spur einer Matrix

Die Summe der Diagonalelemente einer quadratischen Matrix wird Spur (englisch *trace*) der Matrix genannt.

$$Spur(A) = tr(A) = \sum_{i=1}^{n} a_{ii} \quad .$$

Die Spur eines Skalars ist der Skalar selbst. Für die Spur eines Produktes gilt:

$$SpurAB = SpurBA$$

Determinante einer Matrix

Definition: Sei $A = (a_{ij})$ eine quadratische Matrix der Ordnung n. Als Determinante von A bezeichnet man

1. $\det(A) = a_{11}$, falls $n = 1$

2. $\det(A) = a_{11} \cdot a_{22} - a_{12}a_{21}$, falls $n = 2$

3. $\det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij} \det(A_{ij})$ für beliebiges $i = 1, 2 \ldots, n$, wobei die $(n-1 \times n-1)$–Matrix A_{ij} aus A durch Streichung der i-ten Zeile $A_{i.}$ und der j-ten Spalte A_{j} hervorgeht. Durch wiederholte Anwendung von (3) kann man rekursiv alle rechts stehenden Determinanten auf den Fall (2) zurückführen.

Beispiel: Fall einer (3x3)-Matrix A; Entwicklung nach (3) für $i = 1$.

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

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

$$a_{11} \cdot a_{22} \cdot a_{33} - a_{11} \cdot a_{23} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} - a_{13} \cdot a_{22} \cdot a_{31} \quad .$$

Eigenschaften von Determinanten: Sei A eine quadratische Matrix der Ordnung n und $\det(A)$ die Determinante von A. Dann gilt:

1. $\det(A) = \det(A^T)$.
2. Vertauscht man in A zwei Zeilen (bzw. Spalten), so ändert $\det(A)$ das Vorzeichen.
3. Addiert man zu einer Zeile (bzw. Spalte) von A eine beliebige Linearkombination der anderen Zeilen (bzw. Spalten), so ändert sich $\det(A)$ nicht.
4. Multipliziert man die Elemente einer Zeile (bzw. Spalte) von A mit einem Skalar k, so wird $\det(A)$ mit k multipliziert.
5. Sind in A zwei Zeilen (bzw. Spalten) gleich, so gilt $\det(A) = 0$.
6. $\det(I) = 1$.

Lineare Unabhängigkeit von Vektoren und Rang einer Matrix

Ein Vektor b heißt Linearkombination der Vektoren $a_1, a_2, \ldots, a_n$, wenn es (reelle) Zahlen $k_1, k_2, \ldots, k_n$ gibt, so daß gilt:

$$b = k_1a_1 + k_2a_2 + \ldots + k_na_n = \sum_{i=1}^{n} k_ia_i$$

1. Fall: Die Vektoren $a_1, a_2, \ldots, a_n$ heißen linear unabhängig, wenn

$$\sum_{i=1}^{n} k_i \cdot a_i = 0$$

nur für $k_i = 0$ für alle $i = 1, 2, \ldots, n$ gilt, d.h. kein a_i läßt sich dann als Linearkombination der übrigen a_i darstellen.

2. Fall: Ist dagegen mindestens ein $k_i \neq 0$, so läßt sich schreiben:

$$a_i = -\sum_{j=1, j \neq i}^{n} \frac{k_j}{k_i} a_j$$

und a_i ist als Linearkombination der übrigen a_j von diesen linear abhängig. Es besteht dann also mindestens eine lineare Beziehung oder lineare Abhängigkeit zwischen $a_1, a_2, \ldots, a_n$. Entsprechendes gilt für Zeilenvektoren.

Die Maximalzahl der linear unabhängigen Spaltenvektoren heißt Spaltenrang von A und die Maximalzahl der linear unabhängigen Zeilenvektoren heißt Zeilenrang von A. Der Spaltenrang von A ist immer gleich dem Zeilenrang von A. Diese eindeutig bestimmte Zahl heißt Rang von A und wird mit $rg(A)$ bezeichnet. Für eine $n \times m$-Matrix gilt: $rg(A) \leq min\{n, m\}$. Ist $rg(A) = min\{n, m\}$, so besitzt A vollen Rang. Eine quadratische Matrix mit vollem Rang heißt regulär ($rg(A) = n$), anderenfalls singulär ($rg(A) < n$).

Wichtige Rechenregeln für den Rang von Matrizen sind:

$$rg(A) = rg(A^T)$$

$$rg(AB) \leq min\{rg(A), rg(B)\}$$

$$rg(A^T A) = rg(A) = rg(AA^T)$$

$rg(BA) = rg(A) = rg(AC)$ für reguläre Matrizen B und C

Lösen von linearen Gleichungssystemen und inverse Matrix

Definition: Unter einem linearen Gleichungssystem verstehen wir ein System von n Gleichungen mit m Unbekannten:

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1m}x_m & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2m}x_m & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \ldots & + & a_{nm}x_n & = & b_n \end{array}$$

In Matrixnotation:

$$\begin{array}{ccc} A & x & = & b \\ (n \times m) & (m \times 1) & & (n \times 1) \end{array}$$

wobei gilt:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{lm} \\ a_{21} & a_{22} & \cdots & \\ \vdots & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad .$$

Der Vektor x heißt Lösung des linearen Gleichungssystems. Ist $b = 0$, so spricht man von einem homogenen linearen Gleichungssystem. Im Fall $b \neq 0$ spricht man von einem inhomogenen linearen Gleichungssystem.

Das Lösen von linearen Gleichungssystemen erfolgt nach dem Gauß'schen Eliminationsverfahren.

1. Vertausche die Gleichungen (Zeilen) so, daß die erste Unbekannte x_1 einen von Null verschiedenen Koeffizienten erhält. Damit gilt: $a_{11} \neq 0$ nach Vertauschung.
2. Für jedes $i > 1$ wird die i-te Gleichung L_i durch $-a_{11}L_1 + a_{11}L_i$ ersetzt. Symbolisch: $L_i \leftarrow (-a_{11}L_1 + a_{11}L_i)$

Ergebnis: Die erste Gleichung bleibt erhalten, alle anderen Gleichungen enthalten die Variable x_1 nicht mehr. Dieser Prozeß wird wiederholt. Dabei werden sukzessiv die Unbekannten eliminiert.

Beispiel: Wir reduzieren das folgende System

$$\begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ x & + & 3y & + & z & = & 11 \\ 2x & + & 5y & - & 4z & = & 13 \\ 2x & + & 6y & + & 2z & = & 22 \end{array}$$

durch die Operationen $L_2 \leftarrow (-L_1 + L_2)$, $L_3 \leftarrow (-2L_1 + L_3)$ und $L_4 \leftarrow (-2L_1 + L4)$, und anschließend durch die Operationen $L_3 \leftarrow (L_2 - L_3)$ und $L_4 \leftarrow (-2L_2 + L_4)$.

$$\begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ x & + & 3y & + & z & = & 11 \\ 2x & + & 5y & - & 4z & = & 13 \\ 2x & + & 6y & + & 2z & = & 22 \end{array} \quad \Rightarrow \quad \begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ & & y & + & 4z & = & 7 \\ & & y & + & 2z & = & 5 \\ & & 2y & + & 8z & = & 14 \end{array} \quad \Rightarrow$$

$$\begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ & & y & + & 4z & = & 7 \\ & & & & 2z & = & 2 \\ & & & & 0 & = & 0 \end{array} \quad \Rightarrow \quad \begin{array}{rcrcrcr} x & + & 2y & - & 3z & = & 4 \\ & & y & + & 4z & = & 7 \end{array}$$

Existiert eine Lösung, so heißt das lineare Gleichungssystem konsistent, anderenfalls inkonsistent. Das System $Ax = b$ ist genau dann konsistent, wenn $rg(A, b) = rg(A)$.

1. Ist die Koeffizientenmatrix eines konsistenten Systems $Ax = b$ quadratisch und besitzt sie vollen Rang, d.h. $rg(A) = n$, dann gilt:

   $$x = A^{-1}b$$

   ist die eindeutig bestimmte Lösung des Gleichungssystems ist $Ax = b$. Die Matrix A^{-1} heißt die inverse Matrix von A. Sie ist eindeutig bestimmt.

2. Gegeben sei eine $n \times m$ Koeffizientenmatrix A des konsistenten Systems $Ax = b$ mit $rg(A) = m$ und $m \leq n$, d.h. die Anzahl der Unbekannten ist kleiner gleich der Anzahl der Gleichungen. Das System ist eindeutig lösbar, denn wegen $(rg(A) = m)$ besitzt A vollen Spaltenrang. Da $A^T A$ regulär ist, existiert $(A^T A)^{-1}$. Durch Multiplikation der Gleichung $Ax = b$ von links mit A^T erhält man:

   $$A^T Ax = A^T b \quad .$$

   Daraus folgt die eindeutige Lösung:

   $$x = (A^T A)^{-1} A^T b \quad .$$

3. Ist der $rg(A) = r < m$, dann besitzt das System unendlich viele Lösungen. Es können dann $(m - r)$ Komponenten von x willkürlich bestimmt werden und die r verbleibenden Komponenten von x sind eindeutig festgelegt.

4. Ein homogenes lineares Gleichungssystem $Ax = 0$ ist konsistent, da es stets die triviale Lösung $x = 0$ besitzt. Für den Fall $rg(A) = m$ hat $Ax = 0$ wegen der ersten beiden Punkte nur die Lösung $x = 0$. Damit $Ax = 0$ auch nichttrivial lösbar ist, muß wegen des dritten Punktes $rg(A) < m$ erfüllt sein. Falls A quadratisch ist, bedeutet dies: $\det(A) = 0$

Die Berechnung der inversen Matrix

Definition: A sei eine quadratische Matrix von vollem Rang; dann existiert eine Matrix A^{-1} der gleichen Ordnung mit $A^{-1}A = AA^{-1} = I$. A^{-1} heißt die zu A inverse Matrix oder Kehrmatrix.

Satz: A und B seien quadratische Matrizen der gleichen Ordnung mit Inversen A^{-1} und B^{-1}. Dann gilt:

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(k \cdot A)^{-1} = k^{-1}A^{-1}$, k ist ein Skalar
4. $(A^T)^{-1} = (A^{-1})^T$

Satz: Für quadratische Matrizen gelten unter den angegebenen Bedingungen die folgenden Äquivalenzen:

1. Die Matrix A ist regulär.
2. Die Matrix A besitzt eine Kehrmatrix A^{-1}
3. $\det A \neq 0$.

Die Matrix A wird wie folgt invertiert: Der Gauß'sche Algorithmus wird auf das Gleichungssystem $AX = I$ angewandt. An die Stelle von x im üblichen Gleichungssystem $Ax = b$ tritt die $j$te Spalte von A^{-1}, an die Stelle von b die $j$te Spalte der Einheitsmatrix. Die notwendigen Zeilentransformationen werden simultan durchgeführt.

Beispiel:

$$A = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{pmatrix} \quad ,$$

gesucht ist A^{-1}.

$$AI = \left(\begin{array}{ccc:ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 3 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc:ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc:ccc} 1 & 0 & 3 & 4 & 0 & 3 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & -1 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc:ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right) \quad .$$

Daher gilt für die inverse Matrix:

$$A^{-1} = \begin{pmatrix} 7 & -3 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Literatur

Wirtschafts– und Sozialstatistik

Abels, H. (1985): *Wirtschafts- und Bevölkerungsstatistik*, 2. Auflage, Wiesbaden.

Hujer, R. und Cremer, R. (1978): *Methoden der empirischen Wirtschaftsforschung*, München.

Krug, W und Nourney, M. (1987): *Wirtschafts- und Sozialstatistik*, 2. Auflage, München, Wien.

Lippe, Von der, P. (1990): *Wirtschaftsstatistik*, 4. Auflage, München.

Statistisches Bundesamt (1983): *Fachserie 1, Reihe 1, Gebiet und Bevölkerung*, Wiesbaden.

Statistisches Bundesamt (1989): *Statistisches Jahrbuch für die Bundesrepublik Deutschland*, Wiesbaden.

Statistisches Bundesamt (1989): *Wirtschaft und Statistik 12/1989*, Wiesbaden.

Demographie

Birg, H. und Koch, H. (1987): *Der Bevölkerungsrückgang in der Bundesrepublik Deutschland.* Frankfurt/Main.

Birg, H. (1989): Die demographische Zeitwende. *Spektrum der Wissenschaft* 1/89, 40-49.

Findl, P., Holzmann, R. und Münz, R. (1987): *Bevölkerung und Sozialstaat. Szenarien bis 2050. Bd. 2*, Wien.

Teitelbaum, S. und Winter, M. (1985): *The Fear of Population Decline*. San Diego: Academic Press.

Wattenberg, J. (1987): *The Birth Dearth*. New York.

Statistische Methodenlehre

Bamberg, B. und Baur, F. (1989): *Statistik*, 6. Auflage, München.

Fahrmeir, L. und Hamerle, A. (1984): *Multivariate statistische Verfahren*, Berlin.

Floud, R. (1979): *An Introduction to Quantitative Methods for Historians.* London.

Hübler, O. (1989): *Ökonometrie*, Stuttgart.

Lippe, Von der, P. (1979): *Klausurtraining in Statistik*, München.

Schlittgen, R. (1990): *Einführung in die Statistik*, 2. Auflage, München.

Mathematik für Wirtschaftswissenschaftler

Hamerle, A. und Kemény, P. (1981): *Einführung in die Mathematik für Sozialwissenschaftler*, München.