

von der Lippe, Peter

Book — Accepted Manuscript (Postprint)

Deskriptive Statistik: Formeln, Aufgaben, Klausurtraining

Suggested Citation: von der Lippe, Peter (2006) : Deskriptive Statistik: Formeln, Aufgaben, Klausurtraining, ISBN 978-3-486-57863-8, Oldenbourg, München

This Version is available at:

<https://hdl.handle.net/10419/41601>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

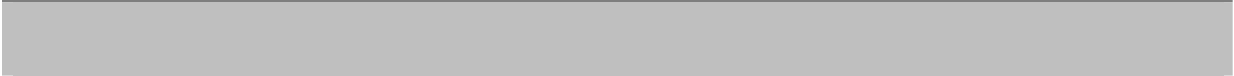


Peter von der Lippe

Deskriptive Statistik

Formeln, Aufgaben, Klausurtraining

Ursprünglich verlegt bei Oldenbourg, hier
in überarbeiteter Form als download zur
Verfügung gestellt



Oldenbourg

von der Lippe: Deskriptive Statistik Inhalt von Teil I (Formelteil)

Kap. 1 Gegenstand und Grundbegriffe der Statistik	4
Kap. 2 Daten, Maßzahlen und Axiomatik	5
Kap. 3 Eindimensionale Häufigkeitsverteilungen	7
Kap. 4 Mittelwerte und andere Lagemaße	10
Kap. 5 Streuung, Schiefe, Wölbung	14
Kap. 6 Konzentrations- und Disparitätsmessung	22
Kap. 7 Zweidimensionale Häufigkeitsverteilungen	26
Kap. 8 Regressionsanalyse	30
Kap. 9 Verhältniszahlen, Wachstumsraten und Aggregation	34
Kap. 10 Indextheorie	40
Kap. 11 Einführung in die Zeitreihenanalyse	47
Kap. 12 Bestandsanalyse und Tafelrechnung	50

Vorwort (zugleich eine Art Einführung)

Die Entstehungsgeschichte dieses Buches ist ähnlich der des Buches “Induktive Statistik” in der gleichen Reihe. Der Formel- und Aufgabenteil ist auch hier aus einer früheren Veröffentlichung hervorgegangen, wurde aber noch einmal überarbeitet. Ferner ist ein Teil “Klausuraufgaben” dem bisher im Oldenbourg Verlag erschienenen Buch “Klausurtraining in Statistik”, 1.-4. Auflage entnommen worden, einem Buch, das somit in dem vorliegenden Buch sowie in dem Buch “Induktive Statistik” aufgegangen ist. Schließlich enthält dieses Buch als vierten Teil auch einige effektiv in letzter Zeit von uns an der Universität-Gesamthochschule Essen gestellte Klausuraufgaben.

Die Erfahrung hat gezeigt, dass es für das Erlernen der Statistik von großer Wichtigkeit ist, sich selbstständig mit Kenntnis des Vorlesungsstoffs an das Lösen von Aufgaben zu machen. Dabei besteht auch ein Unterschied zwischen Übungsaufgaben, die sich jeweils auf einen Ausschnitt des (gerade gelernten) Stoffes beziehen und auch oft den Charakter von (in der Vorlesung benutzten) Demonstrationsbeispielen haben einerseits und Klausuraufgaben andererseits.

Das Buch ist gedacht als Begleitlektüre zu Vorlesungen und Übungen, wie sie üblicherweise unter dem Titel “Deskriptive Statistik” oder “Statistik I” an den meisten Hochschulen für Wirtschaftswissenschaftler angeboten werden. **Wenn** entsprechende Veranstaltungen besucht werden, sollte das Buch ausreichend sein zur Klausurvorbereitung. Dazu sind jedoch noch einige (etwas persönliche) Anmerkungen zum Was und Wie des Statistikstudiums angebracht.

Es wird nicht selten versucht, die Statistik als bloße Anwendung der Wahrscheinlichkeitsrechnung darzustellen oder die Unterscheidung zwischen Deskription und Induktion aufzulösen. Von dem, was man unter “Deskriptive Statistik” verstehen kann, bleiben dann allenfalls Gegenstände, wie sie hier in Kap. 3-5 sowie 7 und 8 (oder in Teilen dieser Kapitel) behandelt werden, übrig und sie werden quasi als Einführungen in bzw. Vorbemerkungen zu Darstellungen der entsprechenden Konzepte der Wahrscheinlichkeitsrechnung betrachtet. Ein solches Verständnis von Statistik wird m. E. weder der Leistungsfähigkeit der Statistik noch den Bedürfnissen der Nutzer von Statistik(en) in der Praxis (insbesondere auch der Wirtschaftspraxis) gerecht. Es mag auch mitverantwortlich sein für den Eindruck mancher Studenten, aber auch mancher Professoren der Wirtschaftswissenschaft, die Statistik sei eine mathematische Hexerei, die sich immer mehr in den Elfenbeinturm zurückzieht, und sie sei deswegen

eigentlich entbehrlich bzw. man könne sie sich ohne Mitwirkung von Statistikern von Fall zu Fall selbst aneignen. Nach unserem Verständnis ist aber Statistik nicht nur ein Teil der Mathematik und sie bietet viele Methoden zur Erkenntnisgewinnung aufgrund zahlenmäßiger Informationen, die nicht notwendig immer auf Wahrscheinlichkeitsüberlegungen beruhen. Gerade für Anwender aus der Wirtschaft sind “rein” beschreibende Methoden mindestens genau so wichtig wie stochastisch fundierte Methoden, und man kann sie nicht richtig verstehen und interpretieren, wenn man sie nur als Rechenaufgaben auffaßt. Man sollte also “Deskriptive Statistik” (und auch die hieran – was v.a. Kap. 10 und 12 zeigt – angrenzende “Wirtschaftsstatistik”) als selbständige Gegenstände betrachten, die es wert sind, sich mit ihnen zu beschäftigen.

Mehr Daten, mehr Rechenfähigkeiten und auch mehr Zwang, etwas empirisch “belegen” zu wollen, führt nicht nur zu mehr Anwendung der Statistik, sondern auch zu mehr Fehlanwendung. Dabei kann mit der Art, wie man Statistik lernt, schon der Grundstein für Fehlanwendung gelegt werden. Statistik kann man weder durch bloßes Hören von Vorlesungen (oder gar Auswendiglernen von Begriffen) lernen, noch durch (Nach-) Rechnen von Aufgaben, die einem vorgerechnet werden. Man kann nicht mit ihr umgehen, wenn man nur im Abstrakten bleibt oder nur lernt, Zahlen in Formeln einzusetzen. Sowohl Vorlesungen (wofür der Formelteil quasi ein Notizgerüst liefert) als auch Übungen (also Aufgaben) sind notwendig und der Reiz (aber leider auch die Schwierigkeit für viele) besteht darin, beides zu verbinden, Methoden und ihre (Rechen-) Ergebnisse. Die Fähigkeit, Methoden und Anwendungen zu verbinden, eine Anwendbarkeit zu erkennen und ein Ergebnis zu interpretieren, verlangt Kenntnisse **und** Übung, Verstehen und auch Phantasie. So etwas zu erlernen kann einem niemand abnehmen; man kann nur versuchen, es zu erleichtern.

Welche Art von Übungsaufgabe man als hilfreich empfindet, um für Statistik motiviert zu werden oder vielleicht auch die angesprochenen Fähigkeiten zu erwerben, ist sicher zum großen Teil Geschmacksache. Für viele sind dafür tatsächliche Anwendungen mit großen und evtl. auch unhandlichen Datensätzen aus Betrieben besonders motivierend. Wir glauben jedoch, dass es ein Schritt weiter ist, angeregt zu werden, sich selbst “Aufgaben” auszudenken. Wer Anwendungen anderer studiert, wird daraus viel lernen, wer aber Spaß daran findet, auch eigene Anwendungen zu konstruieren, könnte einen Schritt mehr Souveränität (und damit auch Motivation) gewinnen. Auf längere Sicht wird man nur das wirklich können, was einem auch Freude macht. Solche Überlegungen stecken auch hinter der Art der Aufgaben, die hier zusammengestellt sind.

Im Unterschied zum Buch “Induktive Statistik” kann hier auch auf einen Begleittext verwiesen werden, der die mit der Formelsammlung präsentierten Stichworte durch Erläuterungen verbindet:

P. v.d. Lippe: Deskriptive Statistik, Reihe UTB
(Uni-Taschenbücher) Bd. 1632, Stuttgart, Jena, 1993.

Die Nummerierung von Formeln und Definitionen im vorliegenden Buch nimmt darauf Bezug.

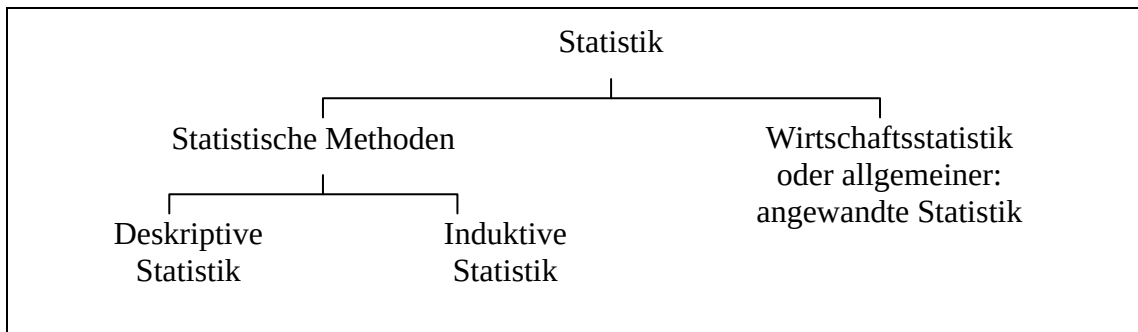
Bei der Vorbereitung des vorliegenden Buches wurde ich erneut von Herrn Dipl. Volkswirt Andreas Kladroba sehr tatkräftig unterstützt. Nach Einstellung von Herrn Dipl. Volkswirt Michael Westermann konnte die nicht unerhebliche Arbeit an der Überarbeitung und Neugestaltung der Texte auch etwas geteilt werden. Beide Mitarbeiter, Herr Kladroba und Herr Westermann, haben nicht nur mit viel Engagement die Veröffentlichung vorbereitet, sondern auch inhaltlich viel beigetragen aufgrund ihrer großen Erfahrungen mit Übungen und Tutorien sowie Klausuren. Ich danke ihnen sehr herzlich dafür. Ich danke auch Frau stud. rer. pol. Karla Behal und Frau stud. rer. pol. Alexandra Werner für die Arbeiten am PC, die sich wieder, wie beim Buch “Induktive Statistik” als aufwendiger und schwieriger herausstellten als wir zunächst dachten.

Essen, den 11.02.99

Kapitel 1: Gegenstand und Grundbegriffe der Statistik

Statistik ist die Lehre von Methoden zur Gewinnung, Charakterisierung und Beurteilung von zahlenmäßigen Informationen über die Wirklichkeit (Empirie).

Übersicht 1.1 Aufbau des Faches Statistik



Def. 1.1: Einheit, Masse

- a) Statistische Einheiten (Elemente, Merkmalsträger) sind Träger von Informationen, bzw. Eigenschaften, die im Rahmen einer empirischen Untersuchung von Interesse sind.
- b) Eine statistische Masse (Kollektiv, Population) ist eine hinsichtlich sachlicher, räumlicher und zeitlicher Kriterien sinnvoll gebildete Gesamtheit von statistischen Einheiten.
- c) Unter dem Umfang einer Masse versteht man die Anzahl ihrer Einheiten (Elemente).

Def. 1.2: Merkmal

Ein Merkmal ist eine Eigenschaft einer statistischen Einheit, die bei einer statistischen Untersuchung interessiert. Es hat endlich und unendlich viele Merkmalsausprägungen (mögliche Realisationen, Modalitäten). Ein Merkmal ist somit eine Menge von Merkmalsausprägungen. Ein Merkmalswert ist eine an einer statistischen Einheit ermittelte Merkmalsausprägung.

Def. 1.3: diskret und stetig

Eine metrisch skalierte Variable X mit den Ausprägungen $x_1, x_2, \dots, x_m$ heißt diskret, wenn X nur endlich viele oder abzählbar unendlich viele reelle Werte x_j annehmen kann, und in jedem endlichen Intervall $a < x < b$ der reellen Zahlengeraden nur endlich viele Werte liegen können. Gilt entsprechend "überabzählbar unendlich viele Werte", so liegt eine stetige (kontinuierliche) Variable vor.

Def. 1.4: Messung

Unter einer Messung versteht man die Abbildung eines empirischen Relativs in ein numerisches Relativ, d.h. die Zuordnung von Zahlen zu Merkmalsausprägungen, so dass die für die Merkmalsausprägungen der empirischen Objekte geltenden Relationen auch für die hierfür verwendeten Zahlen gelten.

Skala (Name, Typ)	definiert ist zusätzlich	zulässige Transformationen	anschauliches Beispiel	Mittelwert
Nominalskala	Äquivalenzrelation ($=, \neq$)	ein-eindeutige Transformation	Postleitzahlen Steuerklasse	Modus
Ordinalskala	Ordnungsrelation ($>, <$)	streng monoton steigend	Windstärke (Beaufort)	Median
Intervallskala	Maßeinheit und Nullpunkt*	linear $y_v = a + bx_v$	Temperatur in Grad Celsius	$\bar{x}$
Ratio- bzw. Verhältnisskala	natürl. Nullpunkt (Maßeinheit noch willkürlich)	proportional $y_v = bx_v$ ($a = 0$)**	Temperatur in Kelvin, Körpergröße	$\bar{x}_G$ $\bar{x}_H$
Absolutskala	auch natürliche Maßeinheit	identisch $y_v = x_v$ ($b = 1$)	Häufigkeit	

* beides (Nullpunkt und Maßeinheit) noch willkürlich.

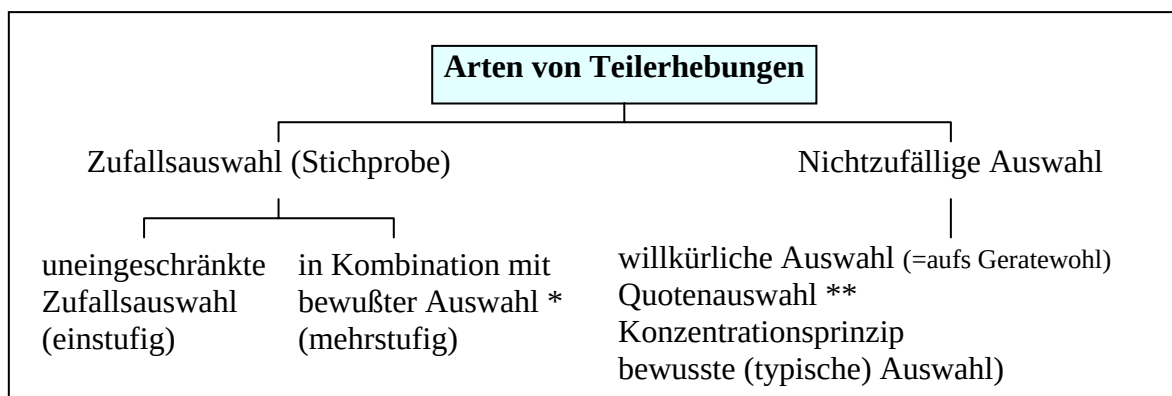
** d.h. der Nullpunkt ist nicht mehr willkürlich (er kann nicht durch $a \neq 0$ verschoben werden), wohl aber die Maßeinheit (weshalb $b \neq 1$ sein kann). Man kann sinnvoll Verhältnisse x_1/x_2 (Proportionen, engl. "ratios") bilden (denn $y_1/y_2 = x_1/x_2$).

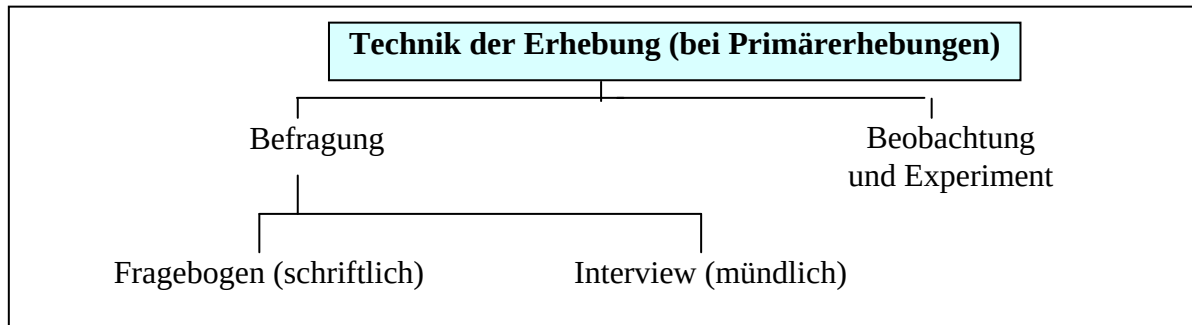
Kapitel 2: Daten, Maßzahlen und Axiomatik

Def. 2.1: Daten, Datensatz

Statistische Daten sind der Ausgangspunkt weitergehender statistischer Auswertungen. Es sind Zahlenangaben über Merkmalsausprägungen, die an Einheiten beobachtet bzw. "gemessen" worden sind. Alle sachlich zusammengehörigen und einer statistischen Auswertung zugrunde zu legenden Daten bilden einen Datensatz.

Übersicht 2.2: Methoden der Datengewinnung





* Geschichtete Stichprobe, Klumpenauswahl (z.B. area sample) usw.

** "Repräsentativer Bevölkerungsquerschnitt" (übliches Verfahren der Markt-, Meinungs- und Umfrageforschung)

Def. 2.2: Maßzahl

- a) Eine Funktion f , die den reellen Beobachtungswerten $x_1, x_2, \dots, x_n$ des Merkmals (der Variablen) X eine reelle Zahl M zuordnet,

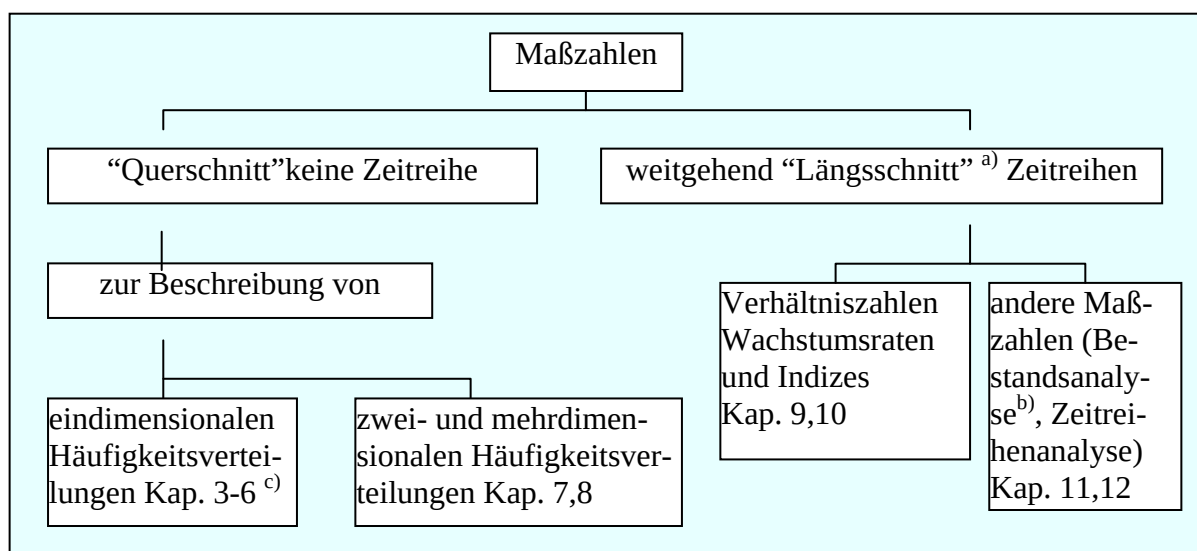
$$(2.1) \quad f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad M = f(x_1, x_2, \dots, x_n)$$

heißt (ungewogene) Maßzahl (Kennzahl), sofern sie bestimmten Axiomen genügt.

- b) Entsprechend ist eine gewogene Maßzahl eine Funktion g , die den reellen Beobachtungswerten $x_1, x_2, \dots, x_m$ des Merkmals X und den dazu korrespondierenden Gewichten $g_1, g_2, \dots, g_m$ eine reelle Zahl G zuordnet,

$$(2.2) \quad g: \mathbb{R}^{2m} \rightarrow \mathbb{R}, \quad G = g[(x_1, g_1), (x_2, g_2), \dots, (x_m, g_m)]$$

Übersicht 2.3: Arten von Maßzahlen



a) Viele, aber nicht alle Methoden sind auf Zeitreihen (nicht verwechseln mit "Längsschnittsdaten") bezogen. Bestimmte Verhältniszahlen, wie Gliederungs- und Beziehungszahlen beziehen sich auf Querschnittsdaten.

b) Kennzahlen der Bestandsanalyse wie z.B. Durchschnittsbestand, Umschlagshäufigkeit, mittlere Verweildauer dienen der Beschreibung von Abläufen, die zu Bestandsänderungen führen.

c) Die Berechnung vieler der in den Kap. 3 bis 6 dargestellten Maßzahlen ist nicht auf eindimensionale Häufigkeitsverteilungen beschränkt. Sie werden auch auf andere Arten von Daten angewandt, z.B. zeitliche Mittelwerte.

Axiome

Axiome sind formale Kriterien, die eine Klasse von Maßzahlen insgesamt erfüllt, wodurch sich diese Klasse auch von einer anderen Klasse von Maßzahlen unterscheidet.

Normierung von Maßzahlen

Wenn eine Maßzahl M den minimalen Wert M_u und den maximalen Wert M_o annimmt, so kann man leicht aus M durch eine Lineartransformation eine auf einen bestimmten Wertebereich normierte Maßzahl M^* erhalten. So erhält man z.B. - wie leicht zu beweisen ist - eine Maßzahl M^* , die zwischen M_u^* als kleinstem und M_o^* als größtem Wert schwankt, mit der folgenden Lineartransformation:

$$(2.3) \quad M^* = M_u^* + (M - M_u) \frac{M_o^* - M_u^*}{M_o - M_u}$$

$$(2.3a) \quad M^* = \frac{M - M_u}{M_o - M_u} \quad (2.3b) \quad M^* = \frac{2(M - M_u)}{M_o - M_u} - 1$$

(2.3a) Normierung von M^* auf den Wertebereich $0 \leq M^* \leq 1$

(2.3b) Normierung von M^* auf den Wertebereich $-1 \leq M^* \leq +1$

Kapitel 3: Eindimensionale Häufigkeitsverteilungen

Def. 3.1: Häufigkeiten

Seien $x_1, x_2, \dots, x_m$ (gruppierte Daten) die m realisierbaren Ausprägungen eines diskreten Merkmals X , dann heißt die Anzahl der Beobachtungseinheiten mit der i -ten Ausprägung,

$$(3.1) \quad n_i = n(x_i) \quad \text{absolute Häufigkeit} \quad (i = 1, 2, \dots, m)$$

und mit $n = \sum n_i$ (Gesamthäufigkeit, Umfang der Beobachtungsgesamtheit) der Quotient

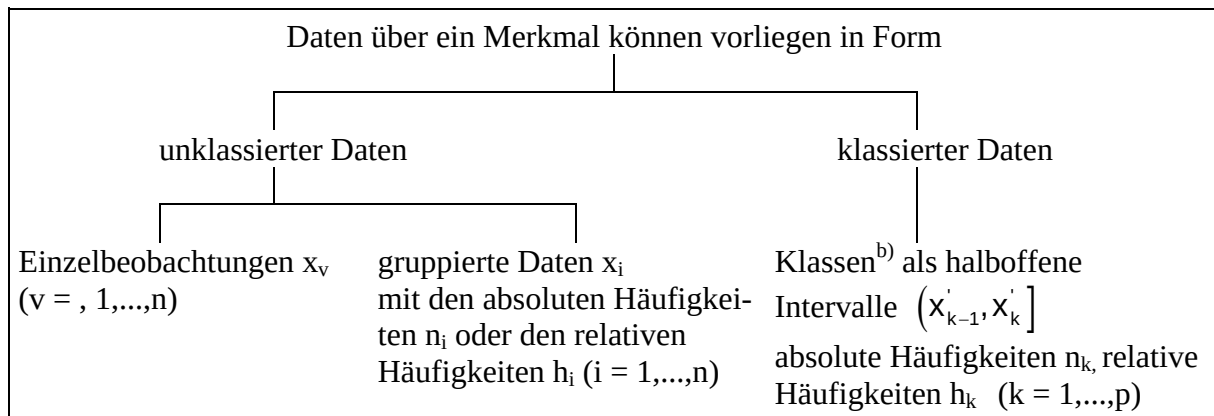
$$(3.2) \quad h_i = h(x_i) = n_i/n \quad \text{relative Häufigkeit}$$

der i -ten Ausprägung des Merkmals X . Es gilt $0 \leq h_i \leq 1$ und (wegen $n = \sum n_i$) $\sum h_i = 1$.

Def. 3.2: Häufigkeitsverteilung

Das m -Tupel $[(x_1, n_1), (x_2, n_2), \dots, (x_m, n_m)]$ heißt absolute Häufigkeitsverteilung und entsprechend ist $[(x_1, h_1), (x_2, h_2), \dots, (x_m, h_m)]$ die (relative) Häufigkeitsverteilung eines Merkmals X . Eine Häufigkeitsverteilung ist also eine Zuordnung von Häufigkeiten (h_i oder n_i) zu Merkmalsausprägungen x_i . Graphische Darstellung durch ein Histogramm (Balken-, Block-, Stabdiagramm).

Übersicht 3.1:



a) In späteren Abschnitten (insbes. im Kap. 8 und 10) wird gelegentlich auch x_i anstelle von x_v verwendet.

b) Es sei verabredet dass x'_k die Obergrenze der k-ten Klasse (d.h. der k-ten der p aneinander grenzenden Größenklassen) ist, so dass x'_{k-1} die Obergrenze der (k-1)-ten Klasse und damit die Untergrenze der k-ten Klasse ist.

Def. 3.3: Summenhäufigkeit, Verteilungsfunktion

Die Summe N_i der absoluten Häufigkeiten n_j ($j = 1, 2, \dots, i$) aller Merkmalsausprägung x_j eines mindestens ordinalskalierten Merkmals, die kleiner oder gleich x_i sind

$$(3.3) \quad N_i = N(x_i) = n(X \leq x_i) = \sum_{j=1}^i n_j$$

heißt absolute kumulierte Häufigkeit (absolute Summenhäufigkeit). Entsprechend heißt

$$(3.4) \quad H_i = H(x_i) = h(X \leq x_i) = \sum_{j=1}^i h_j = N_i/n$$

relative kumulierte Häufigkeit (relative Summenhäufigkeit).

Die Funktion

$$(3.5) \quad H(x) = \begin{cases} 0 & \text{für } x < x_1 \\ H_j & \text{für } x_j \leq x < x_{j+1} \\ 1 & \text{für } x \geq x_m \end{cases}$$

der reellen Variable X heißt (empirische) Verteilungsfunktion oder (relative) Summenhäufigkeitskurve des diskreten Merkmals X .

Def. 3.4: Resthäufigkeit

Die Summe N_i^- der absoluten Häufigkeiten n_j ($j = i+1, i+2, \dots, m$) aller Merkmalsaus-

prägungen, die größer als x_i sind, $N_i^- = N^-(x_i) = n(x > x_i) = \sum_{j=i+1}^m n_j = n - N_i$

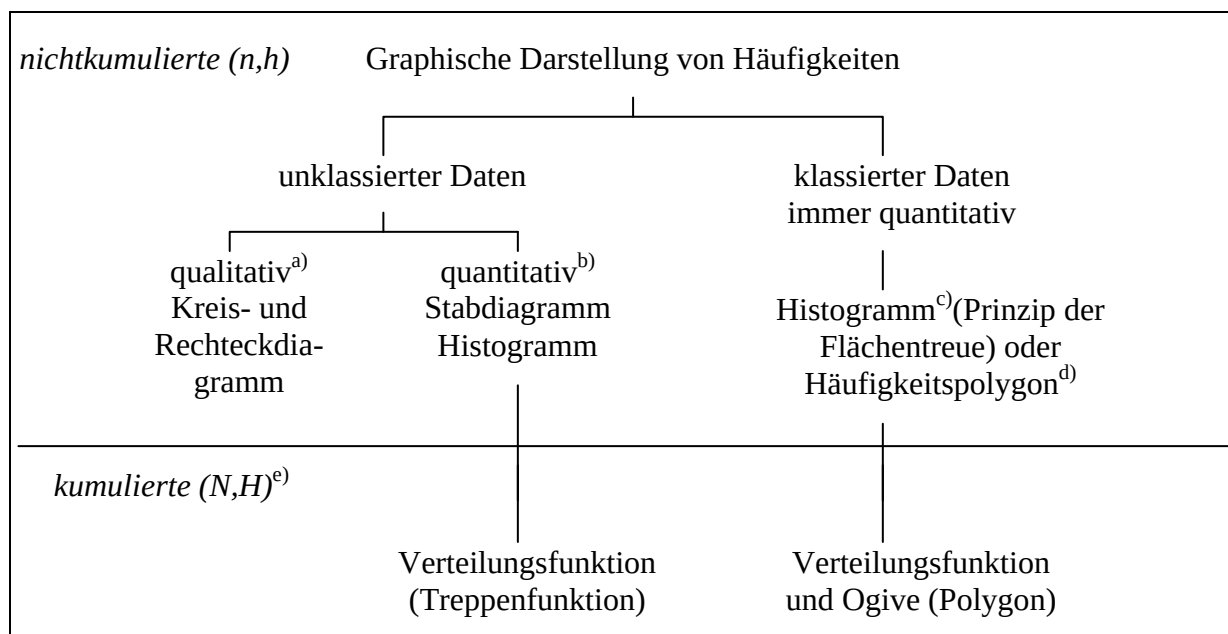
heißt absolute Resthäufigkeit. Entsprechend: $H_i^- = 1 - H_i$ (relative) Resthäufigkeit und $H^-(x) = 1 - H(x)$ relative Resthäufigkeitsfunktion.

Def. 3.5: Klassierung

- In einer klassierten Verteilung wird die Variable X in p Intervalle (Klassen) $(x'_{k-1}, x'_k]$ eingeteilt (linksseitig offene Intervalle) mit $k = 1, 2, \dots, p$ wobei x'_k die Obergrenze der k -ten Größenklasse ist.
- Die Differenz $b_k = x'_k - x'_{k-1}$ heißt Klassenbreite und die Größe $m_k = \frac{1}{2}(x'_{k-1} + x'_k)$ heißt Klassenmitte der k -ten Klasse.
- Die Anzahl n_k der Einheiten, die in die k -te Klasse "fallen" $n_k = n(x'_{k-1} < x \leq x'_k)$ ist die absolute Klassenhäufigkeit und der Anteil $h_k = n_k/n$ heißt relative Klassenhäufigkeit.
- Der Quotient $h_k^* = h_k/b_k$ (Häufigkeit je Klassenbreite) ist die Häufigkeitsdichte oder einfach die Dichte.

Graphische Darstellung von Häufigkeitsverteilungen und Summenhäufigkeiten

Übersicht 3.2: Graphiken



a) kategorial, nominalskaliert;

b) in diesem Fall Stäbe, Säulen oder (nicht notwendig aneinander angrenzende) Blöcke gleicher Breite;

c) bei gleichen Breiten (äquidistante Klassen) ist die Höhe und bei ungleichen Breiten die Fläche der aneinander angrenzenden Blöcke proportional zur absoluten oder relativen Häufigkeit;

d) lineare Verbindung der Blockmitten (auch Kurvendigramm genannt);

e) kumulierte Häufigkeiten (Summenhäufigkeiten) gem. Def. 3.3 (bei Resthäufigkeiten [Def. 3.4] erhält man jeweils fallende Treppenkurven).

Kapitel 4: Mittelwerte und andere Lagemaße

Def. 4.1: Mittelwertaxiome

Mittelwerte M sind Verteilungsmaßzahlen, die unter Berücksichtigung des Skalenniveaus die folgenden Axiome M1 bis M5 erfüllen:

M1	Einschränkung: Es gilt bei der Größe nach geordneten Einzelwerten $x_{(1)} \leq M \leq x_{(n)}$ bzw. bei Merkmalsausprägungen $x_1 \leq M \leq x_m$.
M2	Ergänzung: Tritt zu den n Beobachtungswerten $x_1, x_2, \dots, x_n$ mit dem Mittelwert $M(x_1, \dots, x_n) = M_n$ ein weiterer Wert x_{n+1} hinzu, so soll für den "neuen" Mittelwert $M(x_1, \dots, x_{n+1}) = M_{n+1}$ gelten: wenn $x_{n+1} \leq M_n$ dann $M_{n+1} \leq M_n$ wenn $x_{n+1} \geq M_n$ dann $M_{n+1} \geq M_n$
M3	Transformation: Für den Mittelwert M^* der transformierten Beobachtungswerte $x_v^* = f(x_v)$ soll gelten: $M^* = f(M)$. Dabei ist f eine auf dem Skalenniveau des Merkmals X zulässige Transformation.
M4	Monotonie: Bei den Merkmalen X und Y mit den Beobachtungsvektoren (Vektoren der Beobachtungswerte) $\mathbf{x}$ und $\mathbf{y}$ soll die Mittelwertfunktion monoton zunehmen in Bezug auf die Beobachtungswerte bzw. Merkmalsausprägungen. Für $\mathbf{x} \geq \mathbf{y}$ gilt $M(\mathbf{x}) \geq M(\mathbf{y})$.
M5	Unabhängigkeit von den absoluten Häufigkeiten: Für ein reelles k und mit den Vektoren $\mathbf{x}$ der Merkmalsausprägungen und $\mathbf{n}$ der absoluten Häufigkeiten gilt $M(\mathbf{x}, \mathbf{n}) = M(\mathbf{x}, k \cdot \mathbf{n})$ (d.h. eine Ver- k -fachung der absoluten Häufigkeiten verändert den Mittelwert nicht).

Def. 4.2: Arithmetisches Mittel

$$(4.4) \quad \bar{x} = \frac{1}{n} \sum_{v=1}^n x_v \quad \text{Berechnung aus Einzelbeobachtungen,}$$

ungewogenes arithmetisches Mittel

oder

$$(4.5) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^m x_i n_i = \sum_{i=1}^m x_i h_i \quad \text{Berechnung aus Merkmalsausprägungen}$$

gewogenes arithmetisches Mittel

Satz 4.1: Schwerpunkteigenschaft des arithmetischen Mittels

$$\sum_{v=1}^n (x_v - \bar{x}) = 0 \quad \text{bzw.} \quad \sum_{i=1}^m (x_i - \bar{x}) h_i = 0.$$

Satz 4.2: Minimumeigenschaft

Die Funktion $Q(M) = \sum_v (x_v - M)^2$ besitzt ein Minimum an der Stelle $M = \bar{x}$, d.h. für alle $M \neq \bar{x}$ ist $\sum_v (x_v - M)^2 > \sum_v (x_v - \bar{x})^2$.

Satz 4.3: Lineartransformation des arithmetischen Mittels

Das arithmetische Mittel erfüllt das Mittelwertaxiom M3 für lineare Transformationen.

Aus $x_v^* = a + b \cdot x_v$ folgt

$$(4.6) \quad \bar{x}^* = a + b \cdot \bar{x} \quad (a, b \text{ reelle Zahlen}).$$

Arithmetisches Mittel bei klassierten Daten

Sofern die Klassenmittelpunkte $\bar{x}_k$ ($k = 1, 2, \dots, p$) bekannt sind, berechnet man den Gesamtmittelwert $\bar{x}$ gem. Gl. 4.9:

$$(4.9) \quad \bar{x} = \sum_{k=1}^p \bar{x}_k h_k$$

Andernfalls verwendet man die Klassenmitten m_k und erhält den geschätzten Gesamtmittelwert m (als Schätzung von $\bar{x}$) mit:

$$(4.10) \quad m = \sum_{k=1}^p m_k h_k$$

Im Allgemeinen wird m von $\bar{x}$ verschieden sein. Die Näherung wird umso besser sein, je mehr sich die Beobachtungswerte (symmetrisch) um die Klassenmitten m_k verteilen.

Def. 4.3: Geometrisches Mittel

Die Maßzahl

$$(4.11) \quad \bar{x}_G = \left(\prod_{v=1}^n x_v \right)^{1/n} \quad (\text{bei Einzelbeobachtungen, "ungewogen"}),$$

(das Produktzeichen $\prod$ bedeutet $\prod x_v = x_1 x_2 \dots x_n$), bzw.

$$(4.12) \quad \bar{x}_G = \left(\prod_{i=1}^m x_i \right)^{h_i} \quad (\text{gruppierte Daten, "gewogen"})$$

heißt geometrisches Mittel (der positiven Merkmalswerte $x > 0$). Hieraus folgt unmittelbar

$$(4.13) \quad \log \bar{x}_G = \frac{1}{n} \sum_{v=1}^n \log x_v$$

und entsprechend bei gruppierten Daten, so dass der Logarithmus des geometrischen Mittels gleich ist dem arithmetischen Mittel der logarithmierten Merkmalswerte. Das geometrische Mittel wird deshalb auch logarithmisches Mittel genannt.

Def. 4.4: harmonisches Mittel

Die Maßzahl

$$(4.14) \quad \bar{x}_H = \frac{n}{\sum_v \frac{1}{x_v}} \quad (\text{bei der Berechnung aus Einzelbeobachtungen})$$

$$(4.15) \quad \bar{x}_H = \frac{n}{\sum_i n_i/x_i} = \frac{n}{\sum_i h_i/x_i} \quad (\text{bei gruppierten Daten, [Häufigkeitsverteilung]})$$

heißt harmonisches Mittel ($x \neq 0$).

Es gilt: Der reziproke Wert von $\bar{x}_H$ ist das arithmetische Mittel der reziproken Werte (also der Werte $1/x_v$).

Def. 4.5: quadratisches- und antiharmonisches Mittel

- a) Das quadratische Mittel wird aus Einzelwerten ("ungewogen") mit

$$(4.18) \quad \bar{x}_Q = +\sqrt{\frac{1}{n} \sum x_v^2}$$

bzw. bei gruppierten Daten (Merkmalsausprägungen, "gewogen") mit

$$(4.19) \quad \bar{x}_Q = \sqrt{\sum x_i^2 h_i} \quad \text{berechnet.}$$

- b) Die Maßzahl

$$(4.20) \quad \bar{x}_A = \bar{x}_Q^2 / \bar{x}$$

heißt antiharmonisches Mittel.

Def. 4.6: Potenzmittel

$$(4.21) \quad \bar{x}_{p,r} = \left[\frac{1}{n} (x_1^r + x_2^r + \dots + x_n^r) \right]^{\frac{1}{r}} = \left(\frac{1}{n} \sum_{v=1}^n x_v^r \right)^{\frac{1}{r}} \quad (\text{ungewogene Berechnung}), \text{ bzw.}$$

$$(4.22) \quad \bar{x}_{p,r} = (x_1^r h_1 + x_2^r h_2 + \dots + x_m^r h_m)^{\frac{1}{r}} = \left(\sum_{v=1}^m x_i^r h_i \right)^{\frac{1}{r}} \quad (\text{gewogene Berechnung})$$

Spezialfälle:	$r = -1$	harmonisches Mittel
	$r \rightarrow 0$	geometrisches Mittel
	$r = 1$	arithmetisches Mittel
	$r = 2$	quadratisches Mittel

Ungleichung von Cauchy

$$(4.23) \quad \bar{x}_H \leq \bar{x}_G \leq \bar{x} \leq \bar{x}_Q$$

Def. 4.7: Median

Das Merkmal X sei mindestens ordinalskaliert. Dann ist der Zentralwert (Median) $Z = \tilde{x}_{0,5}$

- a) bei Einzelbeobachtungen die Maßzahl

$$(4.24) \quad Z = \tilde{x}_{0,5} = \begin{cases} x_{((n+1)/2)} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2} [x_{(n/2)} + x_{(n/2+1)}] & , \text{ falls } n \text{ gerade.} \end{cases}$$

Der Median ist der Wert, der in einer der Größe nach geordneten Reihe $x_{(1)} \leq x_{(2)} \leq \dots x_{(n)}$ in der Mitte, d.h. an der $\frac{1}{2}(n+1)$ -ten Stelle steht (bzw. die Interpolation zwischen dem $n/2$ -ten Wert und dem darauf folgenden Wert an der Stelle $n/2 + 1$).

- b) bei gruppierten Werten (Häufigkeitsverteilung, Merkmalsausprägungen) gilt entsprechend für den Median

$$(4.25) \quad Z = \tilde{x}_{0,5} = \begin{cases} x_i & , \text{ falls } H_{i-1} < 0,5 \text{ und } H_i > 0,5 \\ \frac{1}{2}[x_i + x_{i+1}] & , \text{ falls } H_i = 0,5. \end{cases}$$

- c) bei klassierten Daten wird der Median aus der Summenhäufigkeitskurve bestimmt (zur Interpolation vgl. Gl. 4.26).

Interpolation des Medians

$$(4.26) \quad \tilde{x}_{0,5} = x'_{k-1} + b_k (0,5 - H_{k-1}) / h_k$$

Dabei gilt: k = Medianklasse

b_k = Breite der Medianklasse

x'_{k-1} = Obergrenze der k-1-ten Klasse (= Untergrenze der k-ten Klasse)

Def. 4.8: Quantil

Das Merkmal X sei mindestens ordinalskaliert. $[c]$ bedeutet "ganze Zahl, die kleiner oder gleich c ist" (Gaußklammer). Dann heißt die Maßzahl

$$(4.27) \quad \tilde{x}_p = \begin{cases} x_{([np+1])} & , \text{ wenn } np \text{ nicht ganzzahlig ist} \\ \frac{1}{2}(x_{[np]} + x_{[np+1]}) & , \text{ wenn } np \text{ ganzzahlig ist} \end{cases}$$

p -Quantil ($0 < p < 1$).

Quantile bei klassierten Daten

$$(4.26a) \quad \tilde{x}_p = x'_{k-1} + b_k (p - H_{k-1}) / h_k \quad \text{für das interpolierte } p\text{-Quantil.}$$

"Mittelpunkt" des Streubereichs (midrange)

$$(4.28) \quad \tilde{x}_M = \frac{1}{2}(x_{\min} + x_{\max})$$

Def. 4.9: Modus

Existiert bei einer diskreten Variable (einem diskreten Merkmal) X mit den Merkmalsausprägungen x_i genau ein Merkmalswert x_{i^*} dergestalt, dass

$$(4.29) \quad h(x = x_{i^*}) = \max_i h(x_i),$$

so ist dieser Wert der Modus $D = \bar{x}_{\text{mod}}$ (oder der Modalwert, der dichteste oder häufigste Wert), also $D = x_{i^*}$.

Der Modus ist derjenige Merkmalswert, der in einer Häufigkeitsverteilung am häufigsten (absolute oder relative Häufigkeit) vorkommt.

Kapitel 5: Streuung, Schiefe, Wölbung

Konstruktionsprinzipien für Streuungsmaße

1. Mittelwert aus Abständen (Abweichungen) der einzelnen Beobachtungen von einem Lageparameter (vgl. Übersicht 5.1)
2. Abstand zweier Ordnungsstatistiken untereinander (z.B. Spannweite)
3. Mittlerer Abstand der Merkmalswerte untereinander (z.B. Ginis Maß)

Übersicht 5.1: Einige Streuungsmaße nach dem Konstruktionsprinzip Nr. 1

Abweichung vom	Mittel der Abweichung	(absolutes) Streuungsmaß
arithmet. Mittel arithmet. Mittel*) Median**) Median**)	Quadratisches Mittel Arithmetisches Mittel Arithmetisches Mittel Median	Standardabweichung Varianz durchschn. Abweichung Medianabweichung

*) quadrierte Abweichungen vom arithmetischen Mittel

**) absolute Abweichungen vom Median (Zentralwert)

Axiomatik absoluter Streuungsmaße

Absolute Streuungsmaße (S) sind Verteilungsmaßzahlen, die unter Berücksichtigung des Skalenniveaus die Axiome S1 bis S4 erfüllen.

S1	Ein absolutes Streuungsmaß S soll den Wert Null annehmen, falls $x_1 = x_2 = \dots = x_n = \bar{x}$ gilt, d.h. wenn alle Merkmalswerte identisch sind.
S2	Sofern mindestens zwei Merkmalswerte x_i und x_j voneinander verschieden sind, ist $S > 0$ ($i, j = 1, 2, \dots, n$).
S3	Ersetzt man den Beobachtungswert x_k aus der Folge der Beobachtungen x_v ($v = 1, 2, \dots, n$) durch den neuen Wert x_p , so dass die Summe der absoluten Abweichungen von x_p von allen übrigen Werten größer ist als die Summe der absoluten Abweichungen von x_k von allen übrigen Werten, so soll das Streuungsmaß S nicht abnehmen.
S4	Invarianz gegenüber Verschiebungen des Nullpunkts (Translationen) aber nicht gegenüber Maßstabsänderungen: Falls S die Maßeinheit der Merkmalswerte $x_1, x_2, \dots, x_n$ hat, dann soll für die Streuung S_y der mit $y_v = a + bx_v$ transformierten Variablen X gelten: $S_y = b S_x$, wobei $ b > 0$. Für ein absolutes Streuungsmaß mit der quadrierten Maßeinheit der Merkmalswerte soll dann gelten $S_y = b^2S_x$.

Def. 5.1: Relative Streuung

Die Maße der relativen Streuung (S_r) sind definiert als Quotienten eines absoluten Streuungsmaßes S und eines Mittelwertes M (wenn $M \neq 0$), $S_r = S/M$ sofern S die Maßeinheit der Merkmalswerte hat.

Def. 5.2: Varianz und Standardabweichung

a) Die Varianz s^2 eines mindestens intervallskalierten Merkmals X ist, wenn sie aus den einzelnen Merkmalswerten $x_1, x_2, \dots, x_n$ berechnet wird (ungewogener Ansatz), gegeben durch

$$(5.2) \quad s^2 = \frac{1}{n} \sum_{v=1}^n (x_v - \bar{x})^2$$

und wenn sie aus einer Häufigkeitsverteilung (nicht aber bei klassierter Verteilung), d.h. aus den Merkmalsausprägungen $x_1, x_2, \dots, x_m$ berechnet wird (gewogener Ansatz), gilt

$$(5.3) \quad s^2 = \frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^2 n_i = \sum_{v=1}^n (x_v - \bar{x})^2 h_i.$$

b) Die positive Quadratwurzel aus der Varianz heißt Standardabweichung s

$$(5.4) \quad s = +\sqrt{s^2}.$$

Varianz des lineartransformierten Merkmals X

Mit $y_v = a + bx_v$ für alle v und $b \neq 0$ ist die Varianz s_y^2 des zum Merkmal (zur Variablen) Y transformierten Merkmals X durch

$$s_y^2 = \frac{1}{n} \sum_{v=1}^n (y_v - \bar{y})^2 = \frac{1}{n} \sum_{v=1}^n [a + bx_v - (a + b\bar{x})]^2 = b^2 s_x^2$$

und die Standardabweichung s_y durch $s_y = |b| s_x$ gegeben. Mithin ist das Axiom S4 erfüllt.

Verschiebungssatz

$$(5.5) \quad s^2 = \frac{1}{n} \sum_{v=1}^n x_v^2 - \bar{x}^2 \quad (\text{bei Einzelbeobachtungen}) \text{ bzw.}$$

$$(5.6) \quad s^2 = \frac{1}{n} \sum_{i=1}^m x_i^2 n_i - \bar{x}^2 = \sum_{i=1}^m x_i^2 h_i - \bar{x}^2 \quad (\text{bei einer Häufigkeitsverteilung})$$

Steinerscher Verschiebungssatz

$$(5.7) \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 - (\bar{x} - c)^2$$

Hierbei ist c eine beliebige reelle Zahl. Der erste Summand auf der rechten Seite von Gl. 5.7 ist die um c berechnete Varianz, die man mit s_c^2 bezeichnen kann. Zwischen s^2 (oder s_x^2) und s_c^2 besteht nach Gl. 5.7 die folgende Beziehung:

$$(5.7a) \quad s_x^2 = s_c^2 - (\bar{x} - c)^2$$

Mit $c = 0$ erhält man Gl. 5.5 und 5.6 als Spezialfall.

Streuungszerlegung

$$(5.8) \quad s^2 = s_{\text{ext}}^2 + s_{\text{int}}^2.$$

Die externe und die interne Varianz sind jeweils gewogene Mittelwerte. Und zwar ist die externe Varianz,

$$(5.9) \quad s_{\text{ext}}^2 = \sum_{k=1}^r h_k (\bar{x}_k - \bar{x})^2 \quad \text{mit } h_k = n_k/n$$

ein gewogenes Mittel der quadrierten Abstände zwischen den r Mittelwerten der Teilgesamtheiten ($\bar{x}_k$; das müssen nicht Mittelwerte von Klassen, also von Teilgesamtheiten im Sinne einer klassierten Verteilung sein) und dem Gesamtmittelwert $\bar{x}$. Die interne Varianz ist demgegenüber das gewogene Mittel der Varianz s_k^2 innerhalb der Teilgesamtheiten

$$(5.10) \quad s_{\text{int}}^2 = \sum_{k=1}^r h_k s_k^2$$

mit den relativen Häufigkeiten h_k als Gewichte.

Varianz bei klassierten Daten

$$(5.11) \quad s^2 = \sum_{k=1}^r h_k (\bar{x}_k - \bar{x})^2 + \sum_{k=1}^r h_k s_k^2 = s_{\text{ext}}^2 + s_{\text{int}}^2$$

wobei s_k^2 die Varianz innerhalb der k -ten Klasse ist. Gl. 5.11 ist also ein Spezialfall von Gl. 5.8-5.10 (Teilgesamtheiten als aufeinanderfolgende Größenklassen). Bei unbekannten Klassenmittelwerten gilt als Näherung für s_{ext}^2 :

$$(5.11a) \quad s_m^2 = \sum (m_k - m)^2 h_k$$

(mit m [m_k] als wahren oder geschätzten Gesamtmittelwert [Klassenmittelwert])

Sheppard-Korrektur

Sie geht davon aus, dass die Varianz durch s_m^2 häufig systematisch überschätzt wird. Deshalb ist eine bessere Approximation für s^2 gegeben, wenn man $SK = b^2/12$ (SK = Sheppard-Korrektur) von s_m^2 subtrahiert.

Def. 5.3: Durchschnittliche und Medianabweichung

a) Mit $a_1, a_2, \dots, a_n$ seien die absoluten Abweichungen der Merkmalswerte $x_1, x_2, \dots, x_n$ eines mindestens intervallskalierten Merkmals X vom Median $\tilde{x}_{0,5}$ bezeichnet

$$(5.23) \quad a_v = |x_v - \tilde{x}_{0,5}| \quad v=1,2,\dots,n$$

und $a_1, a_2, \dots, a_m$ seien die entsprechenden absoluten Abweichungen der Merkmalsausprägungen $x_1, x_2, \dots, x_m$.

$$(5.24) \quad a_i = |x_i - \tilde{x}_{0,5}| \quad i=1,2,\dots,m.$$

Dann ist das arithmetische Mittel der absoluten Abweichungen vom Median

$$(5.25) \quad d_x = \frac{1}{n} \sum_{v=1}^n a_v \quad \text{bei Einzelwerten}$$

$$(5.26) \quad d_x = \sum_{i=1}^m a_i h_i \quad \text{bei Häufigkeitsverteilungen}$$

die durchschnittliche Abweichung (vom Median). üblich ist auch die Bezeichnung mittlere oder mittlere absolute Abweichung (mean absolute deviation) .

- b) Der Median (Zentralwert) der n absoluten Abweichungen a_v heißt Medianabweichung m_x . Bei Einzelwerten ist m_x der $(n+1)/2$ - te Wert, bzw. der Mittelwert aus dem $n/2$ - ten und dem folgenden Wert in einer der Größe nach geordneten Folge der absoluten Abweichungen a_v :

$$(5.27) \quad m_x = \begin{cases} a_{(n+1)/2} & , \text{ falls } n \text{ ungerade ist} \\ \frac{1}{2} [a_{(n/2)} + a_{(n/2+1)}] & , \text{ falls } n \text{ gerade ist.} \end{cases}$$

- c) Ein selteneres, in erster Linie in der Technik angewandtes Streuungsmaß ist $a_{\max}$, die maximale absolute Abweichung a_v . Da das Maximum ein Grenzfall des Potenzmittels ist, kann man auch die maximale Abweichung als Streuungsmaß nach dem Konstruktionsprinzip Nr. 1 auffassen.

Verschiedentlich wird auch anstelle von d_x die weniger übliche mittlere absolute Abweichung um $\bar{x}$ verwendet, die wir d_x^* nennen wollen:

$$(5.28) \quad d_x^* = \begin{cases} \frac{1}{n} \sum_{v=1}^n |x_v - \bar{x}| & \text{bei Einzelwerten} \\ \sum_{i=1}^n |x_i - \bar{x}| h_i & \text{bei Häufigkeitsverteilungen} \end{cases}$$

Aus der Minimumeigenschaft von $\tilde{x}_{0,5}$ folgt $d_x \leq d_x^*$.

Def. 5.4: Spannweite, Quartilsabstand, Quantilsabstände

- a) Die Differenz zwischen dem (der) größten und kleinsten Beobachtungswert (Merkmalsausprägung) heißt Spannweite R (range, Wertebereich, Variationsbreite). Sie ist bei Einzelwerten durch

$$(5.31) \quad R = x_{(n)} - x_{(1)}$$

und bei Häufigkeitsverteilungen durch die Differenz zwischen kleinster und größter Merkmalsausprägung gegeben (die Berechnung von R ist jedoch vorwiegend bei Einzelwerten üblich).

- b) Der Quartilsabstand $Q_{0,25}$ (Interquartilsabstand IQR) ist die Differenz zwischen dem dritten und ersten Quartil (Gl. 5.32) und der mittlere Quartilsabstand $\bar{Q}_{0,25}$ (Semiquartilsabstand) ist durch Gl. 5.33 gegeben:

$$(5.32) \quad Q_{0,25} = Q_3 - Q_1 \quad \text{und} \quad (5.33) \quad \bar{Q}_{0,25} = \frac{1}{2} Q_{0,25} = \frac{1}{2} [(Q_3 - Q_2) + (Q_2 - Q_1)]$$

- c) Der Quantilsabstand (Interquantilsabstand) Q_p ist die Differenz zwischen dem $(1-p)$ -Quantil $\tilde{x}_{1-p}$ und dem p -Quantil $\tilde{x}_p$,

$$(5.34) \quad Q_p = \tilde{x}_{1-p} - \tilde{x}_p \quad \text{mit } 0 < p < 0,5.$$

Analog zu Gl. 5.33 heißt dann die Maßzahl $\bar{Q} = \frac{1}{2}Q_p$ mittlerer Quantilsabstand (Semi-quantilsabstand).

Größenbeziehung zwischen d_x , s und R : $d_x \leq s \leq R$.

Def. 5.5: Ginis Streuungsmaß

Für die Merkmalswerte $x_1, x_2, \dots, x_n$ eines metrisch skalierten Merkmals X ist Ginis Dispersionsmaß (auch mittlere Differenz genannt) gegeben durch

$$(5.39) \quad S_G = \frac{2}{n(n-1)} \sum_{v < w} |x_v - x_w|$$

(bei Einzelwerten $v, w = 1, 2, \dots, n$) und bei einer Häufigkeitsverteilung durch

$$(5.40) \quad S_G = \frac{2}{n(n-1)} \sum_{i < j} |x_i - x_j| n_{ij} \quad R |x_i - x_j| n_{ij}.$$

Seltener ist das Maß

$$(5.40a) \quad S_G^* = \frac{1}{n^2} \sum \sum |x_v - x_w| \quad v, w = 1, \dots, n$$

Variationskoeffizient

$$(5.51) \quad V = \frac{s}{\bar{x}} \quad (\text{Standardabweichung/arithmetisches Mittel})$$

Def. 5.8: Quartilsdispersionskoeffizient

Setzt man den mittleren Quartilsabstand $\bar{Q}_{0,25} = \frac{1}{2}(Q_3 - Q_1)$ als Maß der absoluten Streuung ins Verhältnis zum Wert $\frac{1}{2}(Q_1 + Q_3)$, den man als eine Art Mittelwert interpretieren kann (analog zu Gl. 4.28), so erhält man QD, den Quartilsdispersionskoeffizient

$$(5.52) \quad QD = (Q_3 - Q_1) / (Q_3 + Q_1).$$

Der Quartilsdispersionskoeffizient kann auch mit dem Median ($\tilde{x}_{0,5} = Q_2$) berechnet werden, man erhält dann

$$(5.52a) \quad QD^* = (Q_3 - Q_1) / Q_2.$$

Auf der Basis des Medians lassen sich auch andere Maße der relativen Streuung konstruieren, etwa

$$(5.52b) \quad RD = d_x / Q_2 = d_x / \tilde{x}_{0,5},$$

eine relativierte durchschnittliche Abweichung.

Def. 5.9: Momente

a) Mit der beliebigen reellen Konstanten a ist der folgende Ausdruck definiert als das k -te Moment um a :

- bei Einzelwerten (ungewogene Berechnung)

$$(5.53) \quad m_{k(a)} = \frac{1}{n} \sum_{v=1}^n (x_v - a)^k \quad [\text{k-tes Moment um } a]$$

- bei Häufigkeitsverteilungen (gewogene Berechnung)

$$(5.54) \quad m_{k(a)} = \frac{1}{n} \sum_{j=1}^m (x_j - a)^k n_j = \sum (x_j - a)^k h_j$$

b) Spezialfälle: Anfangsmomente (oder Momente um Null) und zentrale Momente sind Spezialfälle des Moments um a (Übers. 5.2).

c) Von geringerer Bedeutung sind absolute Momente: analog Gl. 5.53 ist das k -te absolute Moment um a definiert als

$$(5.53a) \quad m_{k(a)}^* = n^{-1} \sum |x_v - a|^k \quad [\text{k-tes absolutes Moment um } a].$$

Bei einer geraden Zahl sind die absoluten Momente gleich den "gewöhnlichen Momenten" [= Momente im Sinne von a) bzw. b)]

d) Für mehrdimensionale Verteilungen sind Produktmomente definiert (vgl. Kap. 7).

Bei proportionaler Transformation $y_v = bx_v$ gilt für zentrale Momente und Anfangsmomente: $z_{k(y)} = b^k z_{k(x)}$.

Def. 5.10: Achsensymmetrie

Die Häufigkeitsverteilung des metrisch skalierten Merkmals X heißt symmetrisch bezüglich des Medians $\tilde{x}_{0,5}$, falls für alle Werte einer reellen Konstante c gilt

$$(5.57) \quad h(\tilde{x}_{0,5} - c) = h(\tilde{x}_{0,5} + c) \quad c > 0.$$

Dabei ist $h(\tilde{x}_{0,5} - c)$ die relative Häufigkeit der Merkmalsausprägung $\tilde{x}_{0,5} - c$ und $h(\tilde{x}_{0,5} + c)$ ist entsprechend definiert. Eine Verteilung ist schief oder asymmetrisch, wenn Gl. 5.57 nicht gilt. Diese Definition ist jedoch nicht generell brauchbar für die Konstruktion von Schiefemaßen.

Fechnersche Lageregel

$$(5.59) \quad \text{linkssteil: } \bar{x}_{\text{mod}} < \tilde{x}_{0,5} < \bar{x} \quad \text{rechtssteil: } \bar{x} < \tilde{x}_{0,5} < \bar{x}_{\text{mod}}$$

($\bar{x}_{\text{mod}}$ = Modus, $\tilde{x}_{0,5}$ = Median)

Def. 5.12: Schiefemaße

a) Die von Bowley und Fisher eingeführte Momentschiefe (Momentkoeffizient der Schiefe) lautet:

$$(5.60) \quad SK_M = \frac{z_3}{s^3} \quad (z_3 \text{ ist das dritte zentrale Moment})$$

linkssteil: $SK_M > 0$

rechtssteil: $SK_M < 0$

symmetrisch: $SK_M = 0$, da $z_3 = 0$.

b) Als Quantilskoeffizient der Schiefe wird bezeichnet:

$$(5.61) \quad SK_{Q,p} = \frac{(\tilde{x}_{1-p} - Q_2) - (Q_2 - \tilde{x}_p)}{\tilde{x}_{1-p} - \tilde{x}_p} \quad p < \frac{1}{2}$$

wobei $Q_2 = \tilde{x}_{0,5} = \text{Median}$; der bekannteste spezielle Koeffizient ($p = \frac{1}{4}$) ist der Quartilskoeffizient der Schiefe (nach Yule und Bowley):

$$(5.62) \quad SK_Q = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}, \text{ mit } -1 \leq SK_Q \leq +1.$$

c) Auf der Fechnerschen Lageregel beruhen die folgenden, von (Yule und Pearson) vorgeschlagenen Schiefemaße (Pearsonsche Schiefemaße):

$$(5.63) \quad SK_{p1} = \frac{\bar{x} - \bar{x}_{\text{mod}}}{s} \quad (\bar{x}_{\text{mod}} = \text{Modus})$$

$$(5.64) \quad SK_{p2} = \frac{3(\bar{x} - \tilde{x}_{0,5})}{s} \quad (\tilde{x}_{0,5} = \text{Median})$$

Der zweite Koeffizient hat Vorteile, weil der Modus evtl. schwer zu bestimmen ist.

Def. 5.13: Symmetrisierende Potenztransformation

Die Variable X wird in die Variable Y nach Maßgabe einer Potenztransformation transformiert, wenn gilt

$$(5.66) \quad y_v = (x_v + c)^p \quad \text{für } p \neq 0 \quad \text{und} \quad y_v = \ln(x_v + c)^p \quad \text{für } p = 0.$$

Def. 5.14: Wölbungsmaße

a) Beim Wölbungsmaß W_M wird das vierte zentrale Moment durch die quadrierte Varianz (denn $(s^2)^2 = s^4$) geteilt:

$$(5.67) \quad W_M = \frac{Z_4}{s^4} - 3.$$

b) Weniger bekannt sind Wölbungsmaße auf der Basis von Quantilen, etwa ein Quantilskoeffizient W_Q der Wölbung:

$$(5.68) \quad W_Q = 1 - \frac{\tilde{x}_{1-p} - \tilde{x}_p}{\tilde{x}_{1-q} - \tilde{x}_q}, \quad \text{mit } 0 < q < p < \frac{1}{2}.$$

- $W_M = 0$ bei der Normalverteilung, bzw. einer Häufigkeitsverteilung die genauso gewölbt ist wie die Normalverteilung (man sagt dann, sie sei mesokurtisch)
- $W_M > 0$ bei Häufigkeitsverteilungen, die vergleichsweise steiler als die Normalverteilung gewölbt sind (leptokurtisch = hochgewölbt, spitz)
- $W_M < 0$ bei Häufigkeitsverteilungen, die vergleichsweise flacher als die Normalverteilung gewölbt sind (platykurtisch = flachgewölbt).

Übersicht 5.2: Momente

Moment um a
a ist eine beliebige reelle Konstante

$$(5.53) \quad m_{k(a)} = \frac{1}{n} \sum_{v=1}^n (x_v - a)^k \quad [\text{k-tes Moment um a}]$$

(ungewogene Berechnung, die gewogene Berechnung erfolgt analog, vgl. Def. 5.9)

Spezialfälle

Anfangsmoment
a = 0

zentrales Moment
a = $\bar{x}$

k-tes Anfangsmoment

$$(5.54) \quad m_k = \frac{1}{n} \sum x_v^k$$

$$(5.54a) \quad m_k = \sum x_i^k h_i$$

k-tes zentrales Moment

$$(5.55) \quad z_k = \frac{1}{n} \sum (x_v - \bar{x})^k$$

$$(5.55a) \quad z_k = \sum (x_i - \bar{x})^k h_i$$

Spezialfälle:

$$m_0 = 1$$

$$m_1 = \bar{x}$$

$$z_1 = 0 \quad (\text{Schwerpunkteigenschaft!})$$

$$z_2 = s^2 \quad (\text{Varianz})$$

Zusammenhänge zwischen Anfangs- und zentralen Momenten:

$z_2 = m_2 - m_1^2$ (Verschiebungssatz für die Varianz) Analog folgt:

$$z_3 = m_3 - 3m_1m_2 + 2(m_1)^3$$

$$z_4 = m_4 - 4m_1m_3 + 6(m_1)^2m_2 - 3(m_1)^4$$

Kapitel 6: Konzentrations- und Disparitätsmessung

Def. 6.1: Anteile, Merkmalsanteile

1. die Anzahl n (absolute Konzentration) der Merkmalsträger, bzw. die Anteile h_i an der Gesamtheit der Merkmalsträger (relative Konzentration, Disparität)
2. die Merkmalsanteile q_i , d.h. die Anteile an dem Merkmalsbetrag (Summe der Merkmalsbeträge der zu verteilenden Größe).

Übersicht 6.1

Darstellung	(absolute) Konzentration	(relative) Konzentration = Disparität
a) graphisch	Konzentrationskurve	Lorenzkurve
b) Maße summarisch	Rosenbluth-Index Herfindahl-Index	Gini-Koeffizient Variationskoeffizient
diskret	concentration ratios	Maximaler Nivellierungssatz

Def. 6.2: Disparitäts- und Gleichheitsmaß

Ist D ein Disparitätsmaß, so ist $G=1-D$ ein Gleichheitsmaß

Lorenzkurve und Gini-Koeffizient bei Einzelbeobachtungen

Def. 6.7: Lorenzkurve, Gini-Koeffizient bei Einzelbeobachtungen

a) Lorenzkurve

Wird der Merkmalsanteil des i -ten Merkmalsträgers bei einer Ordnung nach zunehmender Größe

$$(6.18) \quad q_i = x_{(i)} / \sum x_{(j)} = x_{(i)} / \sum x \quad i, j=1, 2, \dots, n$$

genannt, dann ist

$$(6.19) \quad Q_i = \sum_{j=1}^i q_j$$

der kumulierte Anteil der i kleinsten Merkmalsträger am Merkmalsbetrag.

Die lineare Verbindung der Punkte $P_i(H_i, Q_i)$ mit den kumulierten relativen Häufigkeiten H_i im H-Q-Kordinatensystem heißt Lorenzkurve.

Für die H_i gilt im Fall von Einzelbeobachtungen: $H_i = i/n$.

b) Gini-Koeffizient

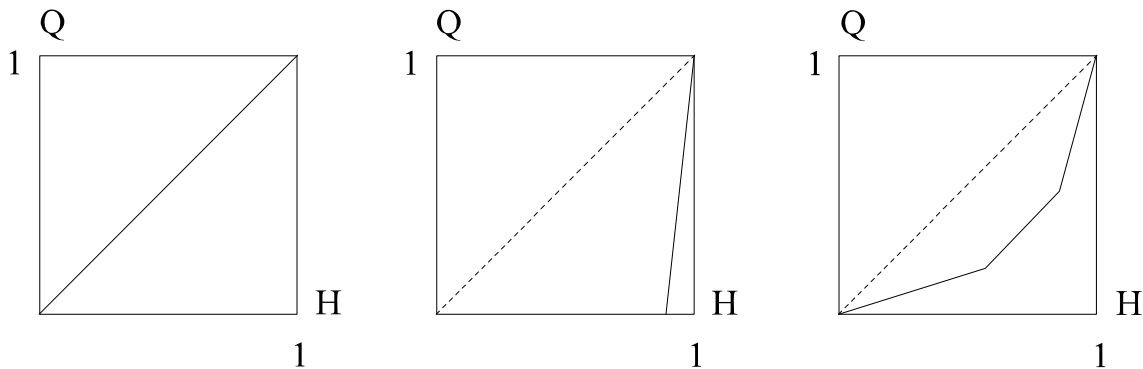
Die Größe

$$(6.20) \quad D_G = \sum_{i=1}^n \frac{2i-n-1}{n} q_i \quad 0 \leq D_G \leq 1 - \frac{1}{n}$$

heißt Disparitätskoeffizient von Gini (oder einfach Gini-Koeffizient). Zur entsprechenden Formel bei gruppierten und klassierten Daten vgl. Gl. 6.28 und 6.29.

Zusammenhang von q_j und h_j : (6.21) $\frac{q_j}{h_j} = \frac{x_j}{\bar{x}}.$

Abb. 6.6: Lorenzkurve und extreme Fälle von Disparität



Fläche zwischen der Lorenzkurve und der Gleichverteilungsgeraden, F

$$(6.22) \quad F = \sum_{i=1}^n \frac{2i-n-1}{2n} q_i.$$

Daraus folgt, dass der Gini-Koeffizient das Verhältnis zwischen F und der Dreiecksfläche unterhalb der Gleichverteilungsgeraden ist (diese Dreiecksfläche beträgt $\frac{1}{2}$), so dass $D_G = 2F$.

Steigung der Lorenzkurve

$$(6.23) \quad q_i/h_i = \frac{x_{(i)}/n\bar{x}}{1/n} = \frac{x_i}{\bar{x}} \quad (\text{Steigung der Lorenzkurve, vgl. Gl. 6.21. Danach ist nicht nur die Lorenzkurve sondern auch die Steigung der Lorenzkurve monoton steigend})$$

Beziehung zwischen H und Q

$H_i \geq Q_i$ ($H_i = Q_i$ gilt außer bei egalitärer Verteilung nur für $i=0$ und $i=n$), denn

$$H_i = \frac{i}{n} \geq \frac{i\bar{x}_i}{n\bar{x}} = Q_i \quad (\text{Lorenzkurve ist monoton steigend})$$

wobei $\bar{x}_i$ der mittlere Merkmalsbetrag der ersten i Merkmalsträger ist, der wegen der Reihenfolge der Merkmalsträger notwendig stets kleiner ist als der Mittelwert $\bar{x}$, der sich auf alle n Merkmalsträger bezieht.

Gini-Koeffizient bei einer linearen Transformation

Eine Lineartransformation der Merkmalswerte mit $y_i = a + bx_i$ wirkt auf die Merkmalsanteile wie folgt

$$q_i^* = a/n\bar{y} + b(\bar{x}/\bar{y})q_i,$$

so dass man für das Disparitätsmaß von Gini (vgl. Gl. 6.20) erhält

$$(6.26) \quad D_G^* = b(\bar{x}/\bar{y})D_G, \quad \text{da } \sum (2i - n - 1) = 2 \sum i - n^2 - n, \quad \text{für } i = 1, 2, \dots, n.$$

Bei proportionaler Transformation (Axiom K1) gilt $a=0$ und wegen $y=bx$ auch $D_G^* = D_G$. Bei einer Niveauänderung im Sinne des Axioms K3 gilt $b=1$ und folglich

$$(6.26a) \quad D_G^* = \frac{\bar{x}}{a + \bar{x}} D_G, \quad \text{so dass bei } a > 0 \text{ gilt } D_G^* < D_G.$$

Ginis "mittlere Differenz" und Ginis Disparitätsmaß

$$(6.27) \quad S_G^* = \sum_i \sum_k |x_i - x_k| / n^2 \quad i, k = 1, 2, \dots, n \quad (\text{siehe Gl. 5.40a})$$

Zusammenhang zwischen D_G und S_G^* : $D_G = S_G^* / 2\bar{x}$.

Daraus folgt übrigens auch, dass man D_G darstellen kann als

$$(6.27a) \quad D_G = \left(\sum_i \sum_k |q_i - q_k| \right) / 2n \quad i, k = 1, 2, \dots, n.$$

Lorenzkurve und Gini-Koeffizient bei gruppierten und klassierten Daten

Def. 6.8: Lorenzkurve bei gruppierten und klassierten Daten

Die lineare Verbindung der Punkte $P_i(H_i, Q_i)$ ($i=0, 1, \dots, m$) mit $P_0(0,0)$ und $P_m(1,1)$ heißt Lorenzkurve.

Gini-Koeffizient bei gruppierten Daten*)

$$(6.28) \quad D_G = 1 - \sum h_i(Q_i + Q_{i-1}) \quad \text{oder} \quad (6.29) \quad D_G = \sum q_i(H_i + H_{i-1}) - 1$$

$i = 1, 2, \dots, m$ ($Q_0 = H_0 = 0$)

*) Wenn bei klassierten Daten die Disparität innerhalb der Klasse berücksichtigt werden soll, vgl. P. von der Lippe, Deskriptive Statistik, UTB Nr. 1632, Gl. 6.29.

Gini-Koeffizient bei zwei Klassen (Spezialfall)

Die Lorenzkurve hat dann nur drei Punkte: $P_0(0,0)$, $P_1(h,q)$ und $P_2(1,1)$. Ginis Dispersions-

maß ist dann der senkrechte Abstand zwischen dem Punkt $P_1(h, q)$ und der Gleichverteilungsgeraden, also die Strecke $D_G = h - q$.

Normiertes Quadrat des Variationskoeffizienten als Disparitätsmaß

$$(6.30) \quad NV = V^2/n \quad \text{normiertes Quadrat des Variationskoeffizienten } V$$

Axiome für Disparitäts- und Konzentrationsmaße

Ist speziell ein Konzentrationsmaß gemeint, so wird es hier K genannt, ein Disparitätsmaß im allgemeinen Sinne heißt entsprechend D und bei einer Aussage, die sich sowohl auf Konzentrations- als auch auf Disparitätsmaße bezieht soll das entsprechende Maß C genannt werden. Die ersten drei Axiome (K1 bis K3) gelten für Konzentrations- und Disparitätsmaße in der gleichen Weise, bei den zweiten drei Axiomen (K4 bis K6) unterscheidet sich das Verhalten von Konzentrationsmaßen einerseits und Disparitätsmaßen andererseits.

K1	Unabhängigkeit von der Maßeinheit: Ein Konzentrations- oder Disparitätsmaß C soll invariant sein bei proportionaler Transformation: Ist $y_i = bx_i$ ($b > 0$), so ist $C(y) = C(x)$.
K2	Verschiebungsprobe (Transfer): Wird ein Betrag d mit $0 < d < h/2$ transferiert von einem Merkmalsträger i (mit dem Merkmalsbetrag $x_{(i)}$) zum Merkmalsträger j mit $x_{(j)} = x_{(i)} - h$, also $x_{(j)} < x_{(i)}$, so soll C abnehmen (regressiver [egalasierender, negativer, d.h. die Konzentration verringernder] Transfer). Die Umkehrung sollte entsprechend bei einem progressiven [positiven] also die Konzentration (und damit auch das Konzentrationsmaß) erhöhenden Transfer ("von arm zu reich") gelten.
K3	Verschiebung, Niveauänderung: Sei $y_i = a + x_i$, dann ist bei egalitärer Verteilung des Merkmals X die Konzentration des Merkmals Y gleich, also $C(y) = C(x)$ und in den sonstigen Fällen soll gelten $C(y) = \begin{cases} < C(x) & , \text{ wenn } a > 0 \text{ (abnehmende Konzentration)} \\ > C(x) & , \text{ wenn } a < 0 \text{ (zunehmende Konzentration)} \end{cases}$
K4	Proportionalitätsprobe, Anzahleffekt: Ersetzt man jeden einzelnen Merkmalsträger i mit dem Anteil q_i am Merkmalsbetrag durch $k > 1$ gleich große Merkmalsträger mit den Anteilen q_i/k , so soll für das neue Disparitätsmaß D^* gelten: $D^* = D$ (Disparität bleibt unverändert) und für das "neue" Konzentrationsmaß K^* im Vergleich zum "alten" $K^* = K/k$ (Fall der Dekonzentration). Entsprechend soll im "umgekehrten" Fall einer Fusion von k gleich großen Einheiten zu einer Einheit gelten $D^* = D$ und $K^* = kK$.
K5	Ergänzungsprobe, Nullergänzung, Disparitätseffekt: Fügt man einer Verteilung m Einheiten, deren Merkmalsbeträge jeweils Null sind ("Nullträger") hinzu, so soll gelten $K^* = K$ und $D^* > D$.
K6	Wertebereiche: Als Wertebereiche sollen $1/n \leq K \leq 1$ (für Konzentrationsmaße) und $0 \leq D \leq 1 - 1/n$ (für Disparitätsmaße) gelten.

Kapitel 7: Zweidimensionale Häufigkeitsverteilungen

Def. 7.1: Verbundene Beobachtungen

a) im Falle von Einzelbeobachtungen:

Wird jede Einheit $v = 1, 2, \dots, n$ mit zwei Merkmalen, d.h. einem Tupel (x_v, y_v) , mit drei Merkmalen [einem Tripel (x_v, y_v, z_v)] oder mit p Merkmalen (p -Tupel) beschrieben, so spricht man von verbundenen Beobachtungen (im Rahmen einer zwei-, drei-, ..., p -dimensionalen Messung) [im folgenden Beschränkung auf $p = 2$ Dimensionen].

b) bei gruppierten Daten:

Das Merkmal X habe die Ausprägungen $x_1, x_2, \dots, x_m$ oder allgemein x_i ($i=1, 2, \dots, m$) und das Merkmal Y habe die Ausprägungen y_j ($j=1, 2, \dots, k$). Dann ist n_{ij} die Anzahl der Einheiten mit den Ausprägungen $X = x_i$ und $Y = y_j$ (also die Anzahl gleicher Wertetupel). Wie im Falle der eindimensionalen Häufigkeitsverteilung $n(\dots)$ eine Funktion ist, die Merkmalsausprägungen eine absolute Häufigkeit zuordnet, so soll $n(\dots)$ hier einer Kombination von Merkmalsausprägungen eine absolute Häufigkeit zuordnen:

$$(7.1) \quad n_{ij} = n(X=x_i \text{ und } Y=y_j) \quad (i = 1, \dots, m \text{ und } j=1, \dots, k).$$

Für die relativen Häufigkeiten gilt analog zur eindimensionalen Häufigkeitsverteilung

$$(7.2) \quad h_{ij} = n_{ij}/n \quad \text{mit} \quad n = \sum_i \sum_j n_{ij} = \sum_{i,j} n_{ij}.$$

c) bei klassierten Daten gilt b) analog.

Def. 7.2: Zweidimensionale Häufigkeitsverteilung (joint distribution)

Eine zweidimensionale Häufigkeitsverteilung ist eine Zuordnung der gemeinsamen absoluten (n_{ij}) oder relativen (h_{ij}) Häufigkeiten zu den Ausprägungen x_i des Merkmals (der Variablen) X und y_j des Merkmals (der Variablen) Y nach Art nachfolgender Tabelle (Matrix). Bei kategorialen (nominalskalierten) Merkmalen spricht man auch von einer Kontingenztafel.

Zweidimensionale Häufigkeitsverteilung
(relative Häufigkeiten)

Merkmal X	Merkmal Y						Randverteilung von X
	y ₁	y ₂	...	y _j	...	y _k	
x ₁	h ₁₁	h ₁₂	...	h _{1j}	...	h _{1k}	
x ₂	h ₂₁	h ₂₂	...	h _{2j}	...	h _{2k}	
x _i	h _{i1}	h _{i2}	...	h _{ij}	...	h _{ik}	
x _m	h _{m1}	h _{m2}	...	h _{mj}	...	h _{mk}	
	Randverteilung von y						

Der Begriff Kontingenztafel wird von vielen Autoren auch bei metrisch skalierten Variablen benutzt. Die absoluten oder relativen Häufigkeiten heißen auch gemeinsame Häufigkeiten und die gesamte Häufigkeitsverteilung auch gemeinsame Häufigkeitsverteilung. Die Größen x_i ($i=1, 2, \dots, m$), bzw. y_j ($j=1, 2, \dots, k$) können Merkmalsausprägungen (gruppierte Daten) oder Größenklassen der Merkmale X und Y (klassierte Daten) bezeichnen.

Verteilungen		
eine zweidimensionale gemeinsame Verteilung h_{ij} (auch kumulierte Verteilung H_{ij}) von x_i, y_j	eindimensionale Verteilungen	
	zwei Randverteilungen Def. (7.3)	$m+k$ bedingte Verteilungen Def. (7.4)
Beschreibende Kennzahlen		
Kovarianz Def.(7.7) Korrelationskoeffizient Def.(7.8)	Mittelwerte $\bar{x}, \bar{y}$ und Varianzen der Randverteilungen	Bedingte Mittelwerte, Regressionslinie Def.(7.6)

Def. 7.3: Randverteilungen (marginal distribution)

Da die Ausprägung x_i bei den Kombinationen $(x_i, y_1), (x_i, y_2), \dots, (x_i, y_k)$ also allen Merkmalskombinationen der i -ten Zeile der zweidimensionalen Häufigkeitsverteilung (Kontingenztafel) vorliegt, ist die Randhäufigkeit $h_{i.}$ definiert als Zeilensumme

$$(7.4) \quad h_{i.} = \sum_{j=1}^k h_{ij} = h(X = x_i).$$

Die als Summen von Zeilen gebildeten Randhäufigkeiten $h_{1.}, h_{2.}, \dots, h_{m.}$ stellen die Randverteilung $h_x(x)$ der Variablen X dar.

Entsprechend bilden die als Summen von Spalten definierten Randhäufigkeiten $h_{.1}, h_{.2}, \dots, h_{.k}$ die Randverteilung $h_y(y)$ des Merkmals (der Variablen) Y , wobei gilt:

$$(7.5) \quad h_{.j} = \sum_{i=1}^m h_{ij} = h(Y = y_j).$$

Die Randverteilungen ausgedrückt in absoluten Häufigkeiten $n_x(x)$ mit den über k Spalten summierten absoluten Häufigkeiten einer Zeile

$$(7.4a) n_{i.} = n_{i1} + n_{i2} + \dots + n_{ik}$$

und die Randverteilung $n_y(y)$ mit den k absoluten Häufigkeiten n_j sind entsprechend definiert. Die beiden Randverteilungen (in relativen Häufigkeiten) sind in der folgenden Tabelle besonders durch Einrahmung markiert:

Merkmal X	Merkmal Y						Summe $h_x(x)$
	y_1	y_2	...	y_j	...	y_k	
x_1	h_{11}	h_{12}	...	h_{1j}	...	h_{1k}	$h_{1.}$
x_2	h_{21}	h_{22}	...	h_{2j}	...	h_{2k}	$h_{2.}$
x_i	h_{i1}	h_{i2}	...	h_{ij}	...	h_{ik}	$h_{i.}$
x_m	h_{m1}	h_{m2}	...	h_{mj}	...	h_{mk}	$h_{m.}$
$h_y(y)$	$h_{.1}$	$h_{.2}$	...	$h_{.j}$	...	$h_{.k}$	1

Die Spaltensumme $h_x(x)$ ist die Randverteilung von X und die Zeilensumme $h_y(y)$ ist die Randverteilung von Y .

Def. 7.4: bedingte Verteilung (conditional distribution)

Die durch Gl. 7.6 definierten bedingten relativen Häufigkeiten $h_{ij|j}$ stellen die bedingte Häufigkeitsfunktion (-verteilung) von X , gegeben $Y = y_j$ dar

$$(7.6) \quad h_{ij} = \frac{h_{ij}}{h_{.j}} = \frac{n_{ij}}{n_{.j}} = h(x | Y = y_j).$$

Analog ist die bedingte Häufigkeitsfunktion (-verteilung) von Y definiert durch die relativen Häufigkeiten der Ausprägung $y_1, y_2, \dots, y_k$ (allgemein: y_j) "gegeben $X = x_i$ " (oder: bedingt durch x_i , oder: wenn $X = x_i$)

$$(7.7) \quad h_{ji} = \frac{h_{ij}}{h_{i.}} = \frac{n_{ij}}{n_{i.}} = h(y | X = x_i).$$

Def. 7.5: Unabhängigkeit

Unabhängigkeit lässt sich auf zwei Arten definieren:

1. Sind die k bedingten Verteilungen h_{ij} des Merkmals X bei allen Ausprägungen y_j ($j = 1, 2, \dots, k$) des Merkmals Y identisch, so sind X und Y unabhängig (gleichzeitig gilt: Gleichheit der m bedingten Verteilungen h_{ji} des Merkmals Y also Unabhängigkeit von X und Y, [Unabhängigkeit ist eine symmetrische Relation]).
2. Im Falle der Unabhängigkeit ergeben sich die absoluten, bzw. relativen gemeinsamen Häufigkeiten aus den entsprechenden Häufigkeiten der Randverteilungen gem.

$$(7.8) \quad n_{ij} = \frac{n_{i.} \cdot n_{.j}}{n} \quad \text{bzw. (7.8a)} \quad h_{ij} = h_{i.} \cdot h_{.j}.$$

Unabhängigkeit impliziert Unkorreliertheit aber nicht umgekehrt, d.h. Unkorreliertheit kann bestehen, obgleich die Variablen X und Y nicht unabhängig sind.

Mittelwert und Varianz der Randverteilungen

Mittelwert $\bar{x}$ der Randverteilung $h_x(x)$

$$(7.9) \quad \bar{x} = \sum_i x_i h_{i.} = \sum_i \sum_j x_i h_{ij}$$

und die Varianz

$$(7.10) \quad s_x^2 = \sum_i x_i^2 h_{i.} - \bar{x}^2.$$

Die entsprechenden Parameter der Randverteilung $h_y(y)$ sind analog definiert.

Parameter der bedingten Verteilungen

- a) Die wichtigsten Parameter der bedingten (Häufigkeits-) Verteilungen sind die bedingten Mittelwerte

$$(7.11) \quad \bar{x} | y = \bar{x}(y_j) = \sum_{i=1}^m x_i h_{ij}$$

$$(7.12) \quad \bar{y} | x = \bar{y}(x_i) = \sum_{j=1}^k y_j h_{ji}$$

- b) Seltener ist die Berechnung der bedingten Varianzen (notwendig zur Berechnung des Korrelationsverhältnisses)

Def. 7.6: empirische Regressionslinie

Die lineare Verbindung der bedingten Mittelwerte $\bar{x}|y$ ist die Regressionslinie (empirische Regressionslinie) der Variablen X. Entsprechend ist die lineare Verbindung der Punkte $P(x, \bar{y}|x)$ die Regressionslinie der Variablen Y.

Der Begriff Regressions"linie" soll deutlich machen, dass die Punkte nicht notwendig auf einer Geraden liegen müssen. Es sind also Regressionslinie und Regressionsgerade (Kap. 8) zu unterscheiden.

Def. 7.7: Kovarianz

Die Kovarianz ist als beschreibende Kennzahl einer zweidimensionalen Verteilung definiert als

$$(7.13) \quad s_{xy} = \frac{1}{n} \sum_{v=1}^n (x_v - \bar{x})(y_v - \bar{y}) \quad \text{bei } n \text{ Einzelbeobachtungen}$$

bzw. bei gruppierten Daten

$$(7.14) \quad s_{xy} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^k (x_i - \bar{x})(y_j - \bar{y})n_{ij} \quad (7.14a) \quad s_{xy} = \sum_{i=1}^m \sum_{j=1}^k (x_i - \bar{x})(y_j - \bar{y})h_{ij}$$

mit absoluten Häufigkeiten

mit relativen Häufigkeiten

Kovarianz bei Lineartransformation

$$(7.15) \quad s_{x^*y^*} = bds_{xy} \quad , \text{ wenn } x^* = a + bx \text{ und } y^* = c + dy$$

Verschiebungssatz für die Kovarianz

Auch für die Kovarianz gilt der Verschiebungssatz:

$$(7.13a) \quad s_{xy} = \frac{1}{n} \sum_v x_v y_v - \bar{x} \bar{y}$$

bzw. bei gruppierten Daten

$$(7.14a) \quad s_{xy} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^k x_i y_j n_{ij} - \bar{x} \bar{y} \quad (7.14b) \quad s_{xy} = \sum_{i=1}^m \sum_{j=1}^k x_i y_j h_{ij} - \bar{x} \bar{y}$$

mit absoluten Häufigkeiten

mit relativen Häufigkeiten

$$\text{oder: } s_{xy} = \overline{xy} - \bar{x} \bar{y}$$

Hierin ist $\overline{xy}$ der Mittelwert des Produkts der x und y Werte und $\bar{x} \bar{y}$ ist das Produkt der Mittelwerte.

Die damit gegebene Beziehung zwischen dem Anfangsproduktmoment $\overline{xy}$ und dem zentralen Produktmoment s_{xy} führt auch wegen der Schwerpunkteigenschaft des arithmetischen Mittels zu folgenden Darstellungen der Kovarianz:

$$(7.17) \quad s_{xy} = \frac{1}{n} \sum_v (x_v - \bar{x})y_v = \frac{1}{n} \sum_v (y_v - \bar{y})x_v .$$

Satz 7.2:

Verschwindet eine der Varianzen (etwa $s_x^2 = 0$), so ist auch die Kovarianz null. Die Umkehrung des Satzes gilt nicht, d.h. $s_{xy} = 0$ ist verträglich mit $s_x^2 > 0$ und $s_y^2 > 0$.

Äquivalent ist die folgende Formulierung: Die Kovarianz einer Variablen mit einer Konstanten k ist stets Null, also $s_{xk} = 0$ oder $s_{yk} = 0$

Satz 7.3: Schwarz'sche Ungleichung

$$(7.18) \quad 0 \leq (s_{xy})^2 \leq s_x^2 s_y^2$$

Def. 7.8: Korrelationskoeffizient

Der Korrelationskoeffizient nach Bravais-Pearson (auch Produkt-Moment-Korrelationskoeffizient oder im Folgenden einfach Korrelationskoeffizient genannt) ist das Verhältnis aus Kovarianz (vgl. Def. 7.7) und dem Produkt der Standardabweichungen.

$$(7.20) \quad r_{xy} = s_{xy}/s_x s_y$$

$$(7.20a) \quad -1 \leq r_{xy} \leq +1 \quad (\text{wegen 7.18}).$$

Somit ist r_{xy} die auf den Wertebereich von -1 bis +1 normierte Kovarianz s_{xy} (während s_{xy} nicht beschränkt ist).

Def. 7.9: Scheinkorrelation, spurious correlation

Sind zwei Variablen X und Y hoch miteinander korreliert, weil sie gemeinsam abhängig sind von einer dritten Variablen Z , so spricht man von Scheinkorrelation.

Kapitel 8: Regressionsanalyse

Def. 8.1: Zusammenhang, Arten von Regressionsfunktionen

a) Ist Y funktional (deterministisch) abhängig von X , d.h. $y = f(x)$ [Y ist eine Funktion von X] so ist jedem Wert von X ein und nur ein Wert von Y zugeordnet. Bei einer stochastischen Beziehung ist diese Funktion, die Regressionsfunktion, von einer Störgröße (Restgröße, Residuum) U überlagert (i.d.R. additiv), so dass für eine einzelne Beobachtung gilt $y_v = f(x_v) + u_v$. Nach der Art der Regressionsfunktion (d.h. des funktionalen Teils der stochastischen Beziehung) unterscheidet man:

b) einfache und multiple Regression:

Bei der einfachen Regression werden nur zwei Variablen X und Y betrachtet. Von multipler Regression spricht man, wenn es eine abhängige Variable Y und mehrere unabhängige Variablen $X_1, X_2, X_3, \dots, X_p$ gibt.

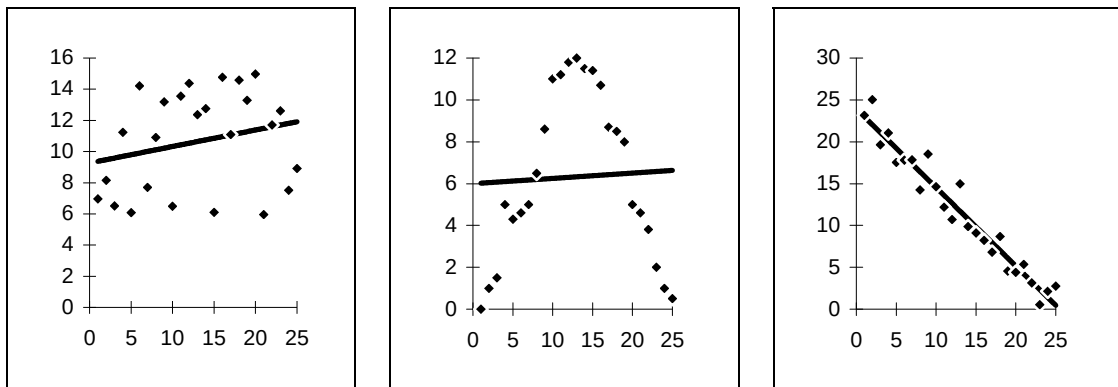
c) lineare und nichtlineare Regression:

Eine Regressionsfunktion ist linear (in den Variablen und in den Parametern), wenn gilt:

$\hat{y}_v = a + bx$ [a und b heißen Regressionskoeffizienten] (einfache lineare Regression) oder

$\hat{y}_v = b_0 + b_1x_{1v} + b_2x_{2v} + \dots + b_px_{pv}$ (multiple lineare Regression, p Regressoren), andernfalls ist sie nichtlinear.

Abb. 8.1: Verschiedene Streuungsdiagramme



In Abb. 8.1 sind beispielhaft drei Streuungsdiagramme (mit Regressionsgeraden $\hat{y}$) gegenübergestellt. Wie leicht zu sehen ist, kann man aus der ersten (linken) Punktwolke auf einen relativ geringen positiven ($r = +0,2408$) Zusammenhang, aus der zweiten Punktwolke auf einen parabolischen und aus der dritten Punktwolke auf einen beträchtlichen negativen ($r = -0,9727$) linearen Zusammenhang der Variablen X und Y schließen.

Def. 8.2: Regressionsgerade

- a) Die lineare Regressionsfunktion (Regressionsgerade) zur Bestimmung von Y (abhängige Variable) durch X (unabhängige Variable) lautet:

$$(8.1) \quad \hat{y}_v = a + bx$$

dabei ist $\hat{y}_v$ der Regresswert für die v-te Beobachtung (Einheit) mit $v = 1, 2, \dots, n$ und für die einzelne Beobachtung (x_v, y_v) gilt:

$$(8.1a) \quad y_v = \hat{y}_v + u_v = a + bx_v + u_v,$$

d.h. die geschätzte Störgröße u_v für die v-te Beobachtung ist der senkrechte Abstand zwischen y_v und $\hat{y}_v$ im x,y-Koordinatensystem.

- b) Die Größen a und b werden Regressionskoeffizienten genannt, wobei a den Ordinatenabstand und b die Steigung der Regressionsgeraden angibt. Es gilt, die Parameter a und b (mit der Methode der kleinsten Quadrate) sowie s_u^2 (Varianz der Störgröße) zu schätzen.

- c) Der Zusammenhang zwischen abhängiger und unabhängiger Variable ist rein rechnerisch vertauschbar, d.h. neben der Regressionsgeraden nach Gl. 8.1 ist auch

$$(8.2) \quad \hat{x}_v = c + dy_v \text{ zu berechnen, wobei für } x_v \text{ gilt:}$$

$$(8.2a) \quad x_v = c + dy_v + v_v.$$

Die Störgröße V ist jeweils der waagrechte Abstand zwischen einem Beobachtungspunkt

x_v und $\hat{x}_v$ im x,y-Koordinatensystem.

Schätzung der Koeffizienten bei der linearen, einfachen Regression

Nach der "Methode der kleinsten Quadrate" erhält man die Normalgleichungen:

(8.4a) $an + b \sum x_v = \sum y_v$	1. Normalgleichung
(8.4b) $a \sum x_v + b \sum x_v^2 = \sum x_v y_v$	2. Normalgleichung

Wird dieses Normalgleichungssystem nach a und b aufgelöst so erhält man als Schätzwerte zur Bestimmung der Regressionskoeffizienten a und b:

$$(8.5a) \quad a = \frac{\sum x_v^2 \sum y_v - \sum x_v \sum x_v y_v}{n \sum x_v^2 - (\sum x_v)^2}$$

$$(8.6a) \quad b = \frac{\sum (x_v - \bar{x})(y_v - \bar{y})}{\sum (x_v - \bar{x})^2} = \frac{s_{xy}}{s_x^2}.$$

Wie man leicht sieht, gilt aufgrund der ersten Normalgleichung:

$$(8.6b) \quad a = \bar{y} - b\bar{x}$$

Man erhält die entsprechenden Formeln zur Bestimmung von c und d indem man in den Normalgleichungen bzw. in den Formeln für a und b x und y vertauscht.

Korrelationskoeffizienten r_{xy}

$$(8.7) \quad r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} = \frac{s_{xy}}{s_x s_y} = \begin{cases} \sqrt{b \cdot d} & \text{wenn } b, d > 0 \\ -\sqrt{b \cdot d} & \text{wenn } b, d < 0 \end{cases}$$

Varianzzerlegung

(8.8) $\frac{1}{n} \sum (y_i - \bar{y})^2$	$=$	$\frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$	$+$	$\frac{1}{n} \sum (y_i - \hat{y}_i)^2$
totale Varianz	$=$	erklärte Varianz	$+$	Residualvarianz
s_y^2	$=$	$s_{\hat{y}}^2$	$+$	s_u^2

Bestimmtheitsmaß B_{yx} und Unbestimmtheitsmaß U_{yx}

$$B_{xy} = \frac{\text{erklärte Varianz}}{\text{totale Varianz}} = \frac{s_{\hat{y}}^2}{s_y^2} \quad 0 \leq B_{yx} \leq 1$$

$$U_{yx} = \frac{\text{Residualvarianz}}{\text{totale Varianz}} = \frac{s_u^2}{s_y^2} = 1 - B_{yx} \quad 0 \leq U_{yx} \leq 1$$

Speziell für die einfache lineare Regression gilt für das Bestimmtheits - und Unbestimmtheitsmaß:

1. Symmetrie: $B_{yx} = B_{xy}$ mit $B_{xy} = \frac{s_{\hat{x}}^2}{s_x^2}$

2. Das Bestimmtheitsmaß B_{yx} ist das Quadrat des Korrelationskoeffizienten ($B_{yx} = r_{xy}^2$).

$$(8.9) \quad B_{yx} = \frac{s_{xy}^2}{s_x^2 s_y^2} = b \cdot d = r_{xy}^2$$

Die mit x erklärte Varianz ist $s_{\hat{x}}^2 = d^2 s_y^2$ so dass $B_{xy} = d^2 s_y^2 / s_x^2 = s_{xy}^2 / s_x^2 s_y^2 = b \cdot d = r_{xy}^2 = B_{yx}$.

Man kann zeigen:

1. Für den Winkel α zwischen den Regressionsgeraden gilt:

$$(8.10) \quad \tan(\alpha) = \frac{s_{xy}(1 - r^2)}{r^2(s_x^2 + s_y^2)}$$

2. Die Steigung der Regressionsgeraden $\hat{x}$ im x,y-Koordinatensystem ist betragsmäßig stets größer als die Steigung b der

Eigenschaften der KQ-Schätzung

(8.11) $\sum u_v = \bar{u} = 0$, die geschätzte Regressionsgerade verläuft durch den Schwerpunkt

(8.12) $\sum y_v = \sum \hat{y}_v$ und somit $\bar{y} = \bar{\hat{y}}$.

(8.13) $\sum x_v u_v = 0$ und $s_{ux} = r_{ux} = 0$.

(8.14) $\sum y_v u_v = r_{\hat{y}u} = 0$.

(8.15) $\sum y_v u_v = \sum u_v^2$ und $s_u^2 = (\sum u_v^2)/n = s_{uy}$. Hieraus folgt $r_{yu} = s_u/s_y$ und damit auch

(8.16) $(r_{yu})^2 = s_u^2/s_y^2 = 1 - (r_{xy})^2 = U_{xy}$

(8.17) $r_{xy} = r_{\hat{y}y}$.

Kapitel 9: Verhältniszahlen, Wachstumsraten und Aggregation

Def. 9.1: (Verhältniszahlen)

- a) Kennzahlen, die als Quotient gebildet sind heißen Verhältniszahlen. Man unterscheidet zwischen Gliederungszahlen, Beziehungszahlen und Messzahlen, je nachdem, wie Zähler und Nenner des Quotienten definiert sind. Auch Wachstumsfaktoren und Wachstumsraten sind als Quotienten Verhältniszahlen im weiteren Sinne (vgl. Übers. 9.1).
- b) Bei Gliederungszahlen G_i ist der Zähler eine Teilmenge des Nenners. Die Gesamtheit (Nennermenge) wird nach einem i.d.R. kategorialen (nominalskalierten) Merkmal in m Teilmassen zerlegt. Mit dem Umfang n_i der i -ten Teilgesamtheit und n der Gesamtheit bzw. den Merkmalssummen S_i und S ist eine Gliederungszahl

$$(9.1) \quad G_i = \frac{n_i}{n} \quad \text{oder} \quad G_i = \frac{S_i}{S}$$

Eine Gliederungszahl (Quote, Anteilswert) G_i ist "dimensionslos" (genauer: G_i hat keine Maßeinheit). In der Praxis wird G_i mit 100 multipliziert und hat dann die Maßeinheit "Prozent".

- c) Bei Beziehungszahlen sind Zähler und Nenner Umfänge oder Merkmalssummen von selbstständigen Massen, die jedoch in sinnvoller Beziehung zueinander stehen sollten. Die Beziehungszahl ist deshalb auch i.d.R. nicht dimensionslos. Je nachdem, ob die Zählermasse als von der Nennermasse "verursacht" gelten kann oder nicht unterscheidet man zwischen Verursachungszahlen und Entsprechungszahlen.
- d) Eine Messzahl setzt einen (meist aktuellen) Wert y_t ins Verhältnis zum Basiswert y_0 , wobei t die "Berichtsperiode" und 0 die (meist zurückliegende) "Basisperiode" (Referenzperiode) ist. Eine dem räumlichen Vergleich dienende Messzahl ist analog definiert. Auch Messzahlen sind wie Gliederungszahlen dimensionslos, weil Kenngrößen (Umfänge, Merkmalsbeträge) gleichartiger Massen ins Verhältnis gesetzt werden. Indexzahlen (Kap. 10) sind zusammengefasste Messzahlen. Wachstumsraten und -faktoren werden in Def. 9.3 definiert.

Eigenschaften von Gliederungszahlen

Es ergibt sich als unmittelbare Folgerung aus Gl. 9.1:

$$(9.2) \quad 0 \leq G_i \leq 1 \quad \text{und} \quad (9.3) \quad \sum G_i = 1 \quad (i = 1, 2, \dots, m).$$

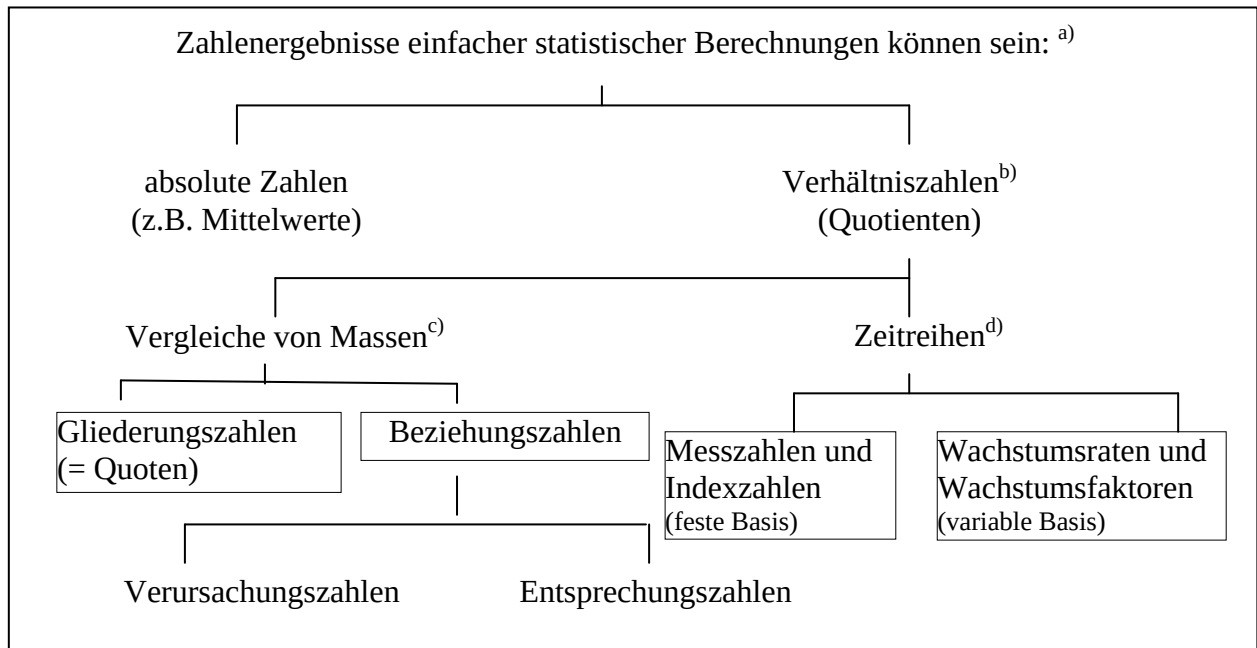
Eigenschaften von Beziehungszahlen

1. Dimension: Anders als Gliederungs- und Messzahlen haben Beziehungszahlen meist eine Maßeinheit.
2. Umkehrbarkeit: Beziehungszahlen sind grundsätzlich umkehrbar.
3. Zusammenhang mit Mittelwerten: Beziehungszahlen sind nichts anderes als Mittelwerte wenn eine Merkmalssumme (Zähler) zu einer entsprechend abgegrenzten Personengesamtheit (Nenner) ins Verhältnis gesetzt wird.

Alle Verhältniszahlen sind ferner Mittelwerte durch Aggregation, d.h. sie sind in dem Sinne Mittelwerte dass eine auf die Gesamtmasse bezogene Verhältniszahl ein Mittel der entsprechenden Verhältniszahlen der Teilmassen ist (so ist z.B. die rohe Todesrate ein Mittel der altersspezifischen Todesraten).

Def. 9.2: (Simpson Paradoxon)

Die Tatsache, dass ein Mittelwert oder eine Verhältniszahl (z.B. eine Quote, ein Anteilswert) für eine Gesamtheit A größer sein kann, als für eine andere Gesamtheit B, obgleich diese Größe (Mittelwert oder Verhältniszahl) in allen Teilgesamtheiten von A kleiner ist als in denen von B, ist bekannt als "Simpson-Paradoxon" (nach Th. Simpson 1710 - 1761).

Übersicht 9.1: Arten von Verhältniszahlen

- a) Zahlergebnisse statistischer Berechnungen können auch Schätzwerte für die Parameter eines Modells sein, z.B. Regressionskoeffizienten.
 b) Die englischen Begriffe sind ratios (Verhältniszahlen), rates (Beziehungszahlen), proportions (Gliederungszahlen) und relatives (Messzahlen).
 c) ohne Zeitbezug (Querschnittsdaten).
 d) Darstellung eines zeitlichen Ablaufs.

Messzahlen

Die Messzahl m_{0t} (d.h. zur Basis 0, Berichtszeit t) einer Variablen Y ist nach Def. 9.1 die Größe:

$$(9.6) \quad m_{0t} = y_t / y_0,$$

bei der diskreten Zeitvariable $t = 0, 1, 2, \dots, T$ bzw. die mit 100 multiplizierten Größen

$$(9.6a) \quad m_{0t}^* = 100m_{0t} = 100y_t / y_0.$$

Die Größe t kann, muss aber nicht "Zeit" bedeuten. Messzahlen können z.B. auch dem räumlichen Vergleich dienen, wenn 0 das Basisland und t das Vergleichsland ist.

Übersicht 9.3: Eigenschaften von Messzahlen

Eigenschaft	Inhalt der Forderung
Identität	$m_{00} = m_{tt} = 1$ ($m_{00}^* = m_{tt}^* = 100$) Identität von Basis- und Berichtsperiode
Dimensionalität	$m(\alpha)_{0t} = \alpha y_t / \alpha y_0 = m_{0t} = y_t / y_0$ Unabhängigkeit von der Maßeinheit der Messwerte
Zeitungkehrbarkeit (Reversibilität)	$m_{t0} = m_{0t}^{-1}$ Vertauschung von Basis- und Berichtsperiode ($m_{t0} m_{0t} = 1$)
Zirkularität (Transitivität, Verkettbarkeit)	für je drei Perioden 0, s und t gilt $m_{0t} = m_{0s} m_{st}$ (= Verkettung; Folgerung : $m_{st} = m_{0t} / m_{0s}$ [= Umbasierung])
Faktorumkehrprobe	ist für alle Perioden die Größe W das Produkt aus P und Q so gilt für die entsprechenden Messzahlen $m_{0t}^W = m_{0t}^P \cdot m_{0t}^Q$ *)

*) eine Wertmesszahl ist das Produkt aus Preis- und Mengenummesszahl.

Umbasierung und Verkettung:

Umbasierung (Basiswechsel) ist die Umkehrung der Verkettung. Mit den Perioden 0, s und t (etwa 1980, 1985 und 1990) bedeutet

Umbasierung: die bisherige Messzahl m_{0t} ist auf die neue Basis s umzustellen (um sie z.B. mit anderen Messzahlen der Basis s vergleichen zu können). Es ist also die Messzahl m_{st} zu bestimmen.

Verkettung: zwei Messzahlenreihen zur Basis 0 und s sind zu einer langen Reihe zusammenzufügen (die Reihe mit der Basis 0 ist mindestens bis s geführt worden).

Lösung:

a) Messzahlen m_{0t} , m_{st} :

Umbasierung: $m_{st} = m_{0t} / m_{0s}$

Verkettung: $m_{0t} = m_{0s} \cdot m_{st}$

b) Messzahlen m_{0t}^* , m_{st}^* (mit 100 multiplizierte Messzahlen):

Umbasierung: $m_{st}^* = (m_{0t}^* / m_{0s}^*) \cdot 100$

Verkettung: $m_{0t}^* = (m_{0s}^* \cdot m_{st}^*) / 100$

Def. 9.3: Wachstumsrate und Wachstumsfaktor bei diskreter Zeit t

- a) Mit der diskreten Zeitvariable $t = 0, 1, 2, \dots, T$ erhält man für die Wachstumsrate und den Wachstumsfaktor (auch Gliedziffer oder Kettenindex genannt) der Zeitreihe y_t (d.h. der Zahlenfolge $y_0, y_1, \dots, y_t, \dots, y_T$) die folgenden Ausdrücke:

$$(9.7) \quad r_t = (y_t - y_{t-1})/y_{t-1} = w_t - 1 \quad (r_t: \text{Wachstumsrate})$$

$$(9.8) \quad w_t = y_t/y_{t-1} = r_t + 1 \quad (w_t: \text{Wachstumsfaktor})$$

- b) Für ein Wachstum mit konstanter Wachstumsrate [z.B. Verzinsung mit Zinseszins] r ($r_t = r$ für alle t) gilt:

$$(9.9) \quad y_t = y_0 \cdot w^t = y_0 \cdot (1+r)^t \quad (\text{Wachstum mit konstanter Rate } r).$$

Bei variierenden Wachstumsraten r_t lautet die Wachstumsgleichung:

$$(9.10) \quad y_T = y_0(1+r_1)(1+r_2)\dots(1+r_T) = y_0 \prod_{t=1}^T (1+r_t) = y_0 \prod_{t=1}^T w_t.$$

- c) Als mittlere Wachstumsrate r soll diejenige konstante Wachstumsrate bezeichnet werden, die über den gleichen Zeitraum von 0 bis T zum gleichen Wachstum von y_0 zu y_T geführt hätte wie die tatsächlichen (unterschiedlichen) Wachstumsraten $r_1, r_2, \dots, r_T$. Daraus folgt, dass r aus dem geometrischen Mittel der Wachstumsfaktoren w_t zu berechnen ist

$$(9.11) \quad r = (w_1 w_2 \dots w_T)^{1/T} - 1 = \left[\prod_{t=1}^T w_t \right]^{1/T} - 1.$$

Mittlere Wachstumsrate

Die mittlere Wachstumsrate ist nach Def. 9.3 aus dem *geometrischen* Mittel der Wachstumsfaktoren zu bestimmen, nicht aber als *arithmetisches* Mittel der Wachstumsraten.

$$(9.11) \quad r = (y_t/y_0)^{1/t} - 1 \quad (\text{mittlere Wachstumsrate}),$$

bzw. in Prozent:

$$(9.11a) \quad r = [(y_t/y_0)^{1/t} - 1] \cdot 100.$$

Def. 9.4: Wachstumsrate bei stetiger Zeit

- a) Die Wachstumsrate $r(t)$ einer stetigen Funktion $y = y(t)$ ist

$$(9.14) \quad r(t) = \frac{y'(t)}{y(t)} = \frac{dy/dt}{y} = \frac{d \ln(y)}{dt}.$$

- b) Bei konstanter Wachstumsrate $r(t) = \alpha$ (für jeden Wert von t) ist die stetige Zeitreihe $y(t)$ gegeben mit

$$(9.15) \quad y(t) = y(0)e^{\alpha t} = y(0)\exp(\alpha t).$$

Beziehung zwischen den Wachstumsraten α (stetige Zeit) und r (diskrete Zeit)

$$(9.18) \quad e^\alpha = w = 1+r, \quad \text{so dass gilt} \quad (9.19) \quad \alpha = \ln(1+r).$$

Man erhält somit im Zusammenhang mit der Reihenentwicklung von e^α und $\ln(1+r)$ die folgenden Umrechnungen

$$(9.20) \quad r = e^\alpha - 1 = \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \frac{\alpha^4}{4!} + \dots$$

für die Umrechnung von α in r (so dass $\alpha < r$) und

$$(9.21) \quad \alpha = \ln(1+r) = r - \frac{r^2}{2!} + \frac{r^3}{3!} - \frac{r^4}{4!} + \dots$$

für die Umrechnung von r nach α .

Wie man sieht gilt nur bei kleinen Wachstumsraten $r \approx \alpha$.

Übersicht 9.4: Wachstumsraten von Produkten, Quotienten und Kehrwerten

	diskrete Zeit	stetige Zeit
Produkt $z = xy$	$w_z = w_x w_y$	$r_z(t) = r_x(t) + r_y(t)$
Quotient $z = x/y$	$w_z = w_x / w_y$	$r_z(t) = r_x(t) - r_y(t)$
Kehrwert $z = 1/y$	$w_z = 1/w_y$	$r_z(t) = -r_y(t)$

Def. 9.5: (Struktureffekt, Standardisierung)

Nach Gl. 9.22 ist eine aggregierte (für die Gesamtmasse errechnete) Beziehungszahl $Q = X/Y$ das gewogene arithmetische Mittel der Teil-Beziehungszahlen $Q_j = x_j/y_j$ ($j=1,2,\dots,J$)

$$(9.22) \quad Q = \sum Q_j g_{yj}.$$

Daraus folgt: Zwei Beziehungszahlen Q_A und Q_B für Gesamtheiten A und B, die sich jeweils in J Teilmassen gliedern lassen, können sich unterscheiden aufgrund unterschiedlicher

- a) Teil-Beziehungszahlen Q_{Aj} , Q_{Bj}
- b) Gewichte der Nennermasse g_{Ayj} , g_{Byj} .

Die Unterschiedlichkeit aufgrund von a) gilt als "echter" Unterschied, diejenige aufgrund von b) wird als Struktureffekt gedeutet. Um die echten Unterschiede herauszuarbeiten, vergleicht man nicht Q_A mit Q_B , sondern

$$(9.24) \quad Q_A^* = \sum Q_{Aj} g_j^* \quad \text{mit} \quad Q_B^* = \sum Q_{Bj} g_j^*,$$

d.h. man vergleicht Beziehungszahlen, die unter Zugrundelegung der gleichen Gewichte (Standardgewichte) g_j^* berechnet sind. Die Größen Q^* heißen dann standardisierte Beziehungszahlen.

Übersicht 9.5.: Wachstumsraten ausgewählter Funktionen

Man beachte:

1. $cy(t)$ und $y(t)$ haben die gleiche Wachstumsrate $r(t)$ [c : Konstante].
2. Hat $y(t)$ die Wachstumsrate $r(t)$, so hat $[y(t)]^{-1}$ die Wachstumsrate $-r(t)$.

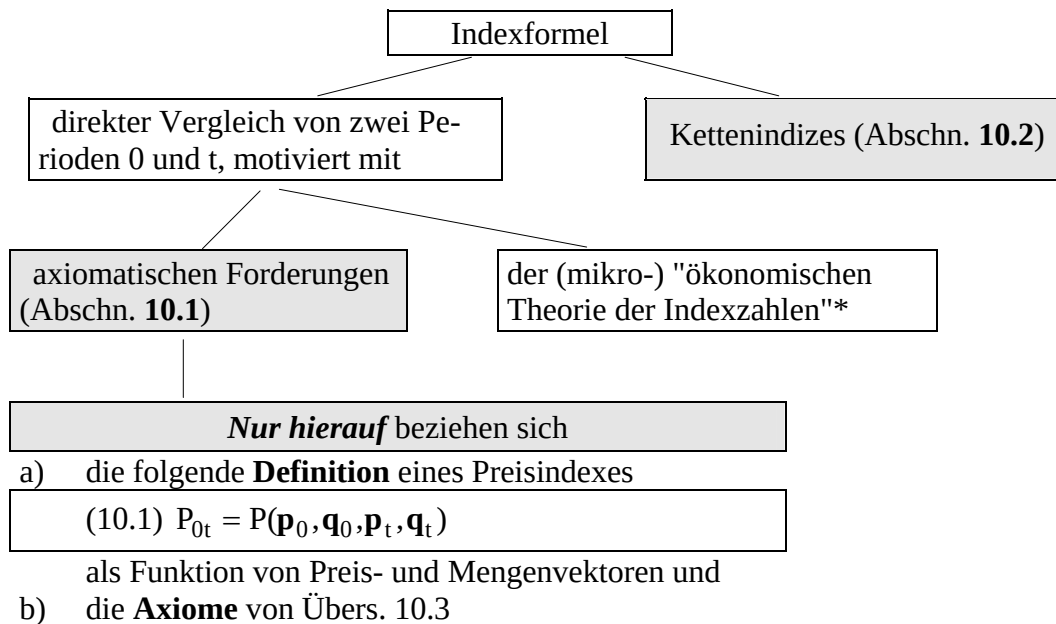
	Funktion $y(t)$	Ableitung	Wachstumsrate
1	$(a+bt)^\alpha$ ⁽²⁾	$\alpha b(a+bt)^{\alpha-1}$	$\alpha b(a+bt)^{-1}$
1a	$a=1$: Gerade $y = a+bt$	b	$b/(a+bt) = r_G$
1b	$a=-1$: $1/(a+bt)$	$-b(a+bt)^{-2}$	$-b/(a+bt) = -r_G$
1c	$a=1/2$: $\sqrt{a+bt}$	$1/2b(a+bt)^{-1/2}$	$b/2(a+bt) = 1/2r_G$
1d	Potenzfunktion bt^α	$b\alpha t^{\alpha-1}$	α/t (hyperbolisch)
2	Parabel ⁽³⁾ $a+bt+ct^2$	$b+2ct$	$(b+2ct)/(a+bt+ct^2)$
3	$a \cdot \exp(bt^\alpha)$	$y \cdot \alpha b t^{\alpha-1}$	$\alpha b t^{\alpha-1}$
3a	$\alpha=1$: ae^{bt}	$y b$	b
	oder: ar^t mit $r=e^b$	$y \cdot \ln(r)$	$b = \ln(r)$
3b	$a = -1$: $ae^{b/t}$	$y b/t^2$	$-b/t^2$
4	$k+be^{ct}$ oder $y=k+br^t$ mit $r=e^c$ (k : Sättigungsniveau)	cbe^{ct} strebt gegen 0 wenn $r < 1$, $c=\ln r < 0$	$(-c)[(k-y)/y] = -cR$ speziell: $c = -1$, dann $r(t) = R$;
5	$k + b/(c+t)$ (Hyperbel)	$-b/(c+t)^2$	$-b/[k(c+t)^2+b(c+t)]$
6	$k(t+a)/(t+b)$ ($b>a$)	$k(b-a)/(t+b)^2$	$(b-a)/(t+a)(t+b)$
7	$\exp(K+br^t)$ mit $r=e^c$ oder: $\ln(y) = K+be^{ct}$ $k=e^K$ Sättigungsniveau	$ya/(b+t)^2$	$a/(b+t)^2$
8	$k/(1+e^{-bt})$ $a,b,k>0$ k : Sättigungsniveau	$by(k-y)/k$	$b+\beta y$ ($\beta = -b/k$) ⁽⁴⁾
9	$\ln(y)_K = K - a/(b+t)$ $k=e^K$ Sättigungsniveau	$ya/(b+t)^2$	$a/(b+t)^2$

- (1) $r(t) = y'/y$
- (2) In dieser allgemeinen Form liegt eine Polynomfunktion vom Grade α vor, wobei $\alpha \in \mathbb{Z}^+$ ist. Bei $\alpha=1/2$ liegt eine Wurfelfunktion und bei ganzzahligem α und $a > 0$ eine Potenzfunktion vor (Fall 1d).
- (3) Kann entsprechend verallgemeinert werden wie Funktion Nr. 1.
- (4) Kennzeichnend für die logistische Funktion: $r(t) = f[y(t)]$ (f : linear).

Kapitel 10: Indexzahlen

Indexzahlen sind Maßzahlen (beschreibende Kennzahlen) für den Vergleich einer Gesamtheit von Erscheinungen. Indizes sind Maße der aggregierten Veränderung, z.B. ein Preisindex ist i.d.R. ein summarisches (zusammengefasstes) Maß von Preisveränderungen (im zeitlichen Vergleich)¹, etwa ein Mittelwert von Preis- Messzahlen für $i = 1, 2, \dots, n$ Waren.

10.1 Prinzipien der Konstruktion von Indexformeln



* wird hier nicht behandelt

10.1 Direkte Indexformeln

a) Vorläufer (historische ungewogene Indizes)

Mit den ungewogenen arithmetischen Mitteln der einzelnen Preise $\bar{p}_t = \sum p_{it}/n$ und $\bar{p}_0$ entsprechend erhält man die Preisindexformel

$$(10.2) \quad P_{0t}^D = \frac{\bar{p}_t}{\bar{p}_0} = \frac{\sum p_{it}}{\sum p_{i0}} \quad \text{von (D =) Dutot}$$

eine Verhältniszahl (Messzahl) von Mittelwerten: sie ist nicht sinnvoll, weil sie nicht die Kommensurabilität, Axiom P5 in Übers. 10.3 erfüllt. Es gilt

eine Messzahl von Mittelwerten erfüllt nicht Minimalforderungen an Indizes (z.B. Axiom P5), wohl aber ein Mittelwert von Messzahlen.

Ein ungewogenes arithmetisches Mittel von Preismesszahlen ist die Preisindexformel

$$(10.3) \quad P_{0t}^C = \frac{1}{n} \sum \frac{p_{it}}{p_{i0}} \quad \text{von (C =) Carli.}$$

¹ zumindest direkte Indizes (vgl. Übers. 10.1) werden auch für den interregionalen (z.B. internationalen) Vergleich benutzt.

Ein ungewogenes geometrisches Mittel von Preismesszahlen ist die Formel von **Jevons**

$$P_{0t}^J = \left[\frac{p_{1t}}{p_{10}} \cdot \frac{p_{2t}}{p_{20}} \cdot \dots \cdot \frac{p_{nt}}{p_{n0}} \right]^{\frac{1}{n}}.$$

b) Aktuelle Indexformeln (insbes. Laspeyres und Paasche)

Preisindizes nach Laspeyres und Paasche haben eine doppelte Interpretation, als

- gewogenes Mittel von Preismesszahlen (**Messzahlenmittelwertformel**) und als
- Verhältnis von Ausgaben- bzw. Einnahmenaggregate (**Aggregatformel**)².

Zur Vereinfachung ist im Folgenden das Subskript i (Warenart) weggelassen worden. Die Vektorschreibweise zeigt, dass die Indizes *lineare* Indizes sind (Preisindizes linear in den Preisen, Mengenindizes linear in den Mengen):

Formel von	Messzahlenmittelwertformel	Aggregatformel
Laspeyres (L)	(10.4) $P_{0t}^L = \frac{\sum p_t \cdot p_0 q_0}{p_0 \sum p_0 q_0}$ gewogenes arithmetisches Mittel Gewichte: Ausgabenanteile zur Basiszeit	(10.5) $P_{0t}^L = \frac{\sum p_t q_0}{\sum p_0 q_0} = \frac{\mathbf{p}_t \cdot \mathbf{q}_0}{\mathbf{p}_0 \cdot \mathbf{q}_0}$ Zähler: fiktives Ausgabenaggregat Nenner: tatsächliche Ausgaben
Paasche (P)	(10.6) $P_{0t}^P = \frac{\sum p_t \cdot p_0 q_t}{p_0 \sum p_0 q_t}$ oder: gewogenes harmonisches Mittel, Gewichte: Ausgabenanteile zur Berichtszeit	(10.7) $P_{0t}^P = \frac{\sum p_t q_t}{\sum p_0 q_t} = \frac{\mathbf{p}_t \cdot \mathbf{q}_t}{\mathbf{p}_0 \cdot \mathbf{q}_t}$ Zähler: tatsächliche Ausgaben, Nenner: fiktives Ausgabenaggregat

In manchen Lehrbüchern (nicht in der Praxis) spielt auch der Preis- (oder gar der Mengen-) index nach **Low** eine gewisse Rolle. Ein solcher Index erfüllt jedoch nicht die Axiome von Übers 10.3. Kein Index, der Durchschnittspreise verwendet, kann kommensurabel sein. Schon wegen der Unmöglichkeit, kg-, Liter-, Stück-Mengen usw. zu einer "Gesamtmenge" zu addieren, sind Durchschnittsmengen auch meist gar nicht definiert.

Wertindex (z.B. Lebenshaltungskostenindex im Unterschied zum Preisindex für die Lebenshaltung nach Laspeyres)

$$(10.8) \quad W_{0t} = \frac{\sum p_{it} q_{it}}{\sum p_{i0} q_{i0}} \quad \text{oder einfach} \quad W_{0t} = \frac{\sum p_t q_0}{\sum p_0 q_0}$$

Mengenindizes gewinnt man aus Preisindizes durch Vertauschen von Mengen und Preisen (vgl. Übers. 10.2):

$$(10.9) \quad Q_{0t}^L = \frac{\sum q_t p_0}{\sum q_0 p_0} \quad (\text{Laspeyres}) \quad \text{und} \quad (10.10) \quad Q_{0t}^P = \frac{\sum q_t p_t}{\sum q_0 p_t} \quad (\text{Paasche})$$

$$\text{Wertindex als Indexprodukt} \quad (10.11) \quad W_{0t} = P^L Q^P = P^P Q^L \quad (\text{Produkttest})^3$$

- _____

² Der berühmte "Idealindex" von I. Fisher oder auch Kettenindizes aller Art besitzen keine der beiden Interpretationen.

³ Das Indexpaar Laspeyres-Paasche erfüllt den Produkttest, nicht jedoch die anspruchsvollere (und in ihrer Bedeutung meist völlig überschätzte) Faktorumkehrbarkeit.

10.2 Übersicht über die Indexformeln

Preise p , Mengen q , Subskripte t = Berichtszeit, 0 = Basiszeit, Summierung über alle n Waren

$$\text{Wertindex } W_{0t} = \frac{\sum p_t q_0}{\sum p_0 q_0}$$

$$\text{Laspeyres Preisindex } P_{0t}^L = \frac{\sum p_t q_0}{\sum p_0 q_0}$$

Verwendung für: spezielle Preisniveaus
(z.B. Preisindizes für die Lebenshaltung)

$$\text{Paasche Preisindex } P_{0t}^P = \frac{\sum p_t q_t}{\sum p_0 q_t}$$

Verwendung: Preisbereinigung (Deflationierung, z.B. des Sozialprodukts)

Vertauschung von Preisen und Mengen in den Formeln führt zu den entsprechenden Mengenindizes Q^L und Q^P also:

$$\text{Laspeyres Mengenindex } Q_{0t}^L = \frac{\sum q_t p_0}{\sum q_0 p_0}$$

$$\text{Paasche Mengenindex } Q_{0t}^P = \frac{\sum q_t p_t}{\sum q_0 p_t}$$

Es gilt die grundlegende Formel als Basis für die Preisbereinigung:

$$(10.11) \quad W_{0t} = P_{0t}^L Q_{0t}^P = P_{0t}^P Q_{0t}^L$$

Preisbereinigung (Deflationierung; auch Realwert- oder Volumenrechnung genannt).

aus einem	ist zu errechnen ein	Vorgehensweise
Wert = $\sum p_t q_t$ (einer nominalen Größe, zu jeweiligen Preisen)	Volumen = $\sum p_0 q_t$ (eine reale Größe, zu konstanten Preisen des Basisjahres)	Division durch einen ¹⁾ Paasche Preisindex $V_t = \frac{W_t}{P_{0t}^P}$
Wertindex $W_{0t} = \frac{\sum p_t q_0}{\sum p_0 q_0}$	Laspeyres-Mengenindex* $Q_{0t}^L = \frac{\sum q_t p_0}{\sum q_0 p_0}$	Division durch einen ¹⁾ Paasche Preisindex $Q_{0t}^L = \frac{W_{0t}}{P_{0t}^P}$

1) sich auf das gleiche Aggregat beziehende

2) als Maß für die Veränderung von Volumen

Strukturelle Konsistenz (der Deflationierung)

Gilt für nominale Teilaggregate $W = W_1 + W_2 + \dots + W_m$ und soll dann für die realen Teilaggregate $V_j = W_j/P_j$ ($j = 1, 2, \dots, m$) gelten

$$\frac{W}{P} = V = \frac{W_1}{P_1} + \dots + \frac{W_m}{P_m} = V_1 + \dots + V_m, \text{ dann (10.12) } P^{-1} = \sum \frac{W_j}{W} P_j^{-1}$$

d.h. dann muss der Gesamt-Deflator P ein harmonisches Mittel der m Teil-Deflatoren sein (Gewichte $W_j/\Sigma W_j = W_j/W$), also ein direkter Paasche Preisindex. Deflationierung mit einem

anderen Index als P^P liefert strukturell inkonsistente Ergebnisse (Volumen addieren sich nicht in gleicher Weise wie Werte, das Ergebnis der Deflationierung ist abhängig vom Aggregationsgrad).

Additive Konsistenz (der Indexformel)

Wenn ein Gesamtindex zu Teilaggregaten $j = 1, \dots, m$ zerlegt werden kann, dann soll sich der Gesamtindex aus den Teilindizes in der gleichen Weise zusammensetzen, wie der Gesamtindex aus den Messzahlen. Im Falle von P^L gilt z.B. der folgende Zusammenhang

$$(10.13) \quad P_{0t}^L = \sum_j \frac{W_{j,0}}{\sum_j W_{j,0}} P_{j,0t}^L$$

d.h. der Gesamtindex P_{0t}^L ist das arithmetische Mittel der m Teil-Indizes $P_{j,0t}^L$ mit den Wertanteilen zur Basiszeit als Gewichte. Für den Paasche Preisindex P^P gilt Gl. 10.12. Lineare (=additive) Indizes (vgl. Übers. 10.3) sind additiv konsistent. Die Umkehrung gilt nicht.

Formel von Ladislaus v. Bortkiewicz

(Größenrelation zwischen Laspeyres- und Paasche-Preisindex)

Die Kovarianz von Preis- (b_i) und Mengenummesszahlen (c_i) mit den Gewichten g_i (Ausgabenanteile zur Basiszeit) lautet: $C = \sum (b_i - P^L)(c_i - Q^L)g_i = Q^L(P^P - P^L)$.

Daraus folgt $W = P^L Q^L + C$, denn mit $g_i = \frac{p_{i0} q_{i0}}{\sum p_{i0} q_{i0}}$, $b_i = \frac{p_{it}}{p_{i0}}$, und $c_i = \frac{q_{it}}{q_{i0}}$ gilt

$$P_{0t}^L = \sum g_i b_i \quad \text{und} \quad Q_{0t}^L = \sum g_i c_i \quad \text{sowie} \quad W_{0t} = \sum b_i c_i g_i = P^L Q^P = P^P Q^L$$

Dann gilt für die Kovarianz

$$(10.14) \quad C = Q^L(P^P - P^L) = P^L(Q^P - Q^L) \text{ also}$$

wenn negative Kovarianz $C < 0$ dann $P^L > P^P$ und $Q^L > Q^P$
--

wenn positive Kovarianz $C > 0$ dann $P^L < P^P$ und $Q^L < Q^P$
--

c) Einige Axiome und ein Axiomensystem (von Eichhorn/Voeller)

Zu einigen fundamentalen Forderungen an sinnvolle Indexformeln (Index-"axiome") vgl. Übers. 10.3. Wichtige Axiome, die erst in neuerer Zeit mehr beachtet werden sind ferner die Aggregationseigenschaften, wie z.B. strukturelle - und additive Konsistenz (s.o.).

Eine große Rolle spielen jedoch auch immer noch Axiome (oder "Proben", "Tests"), die aus der Indexphilosophie von Irving Fisher stammen wie:

Zeitumkehrbarkeit (Z)

Vertauschung von Basis- und Berichtsperiode führt zum reziproken Preisindex

(10.14) $P_{0t} P_{t0} = 1$ (Zeitumkehrprobe). Nicht erfüllt vom Paar Laspeyres/Paasche, denn

$$P_{0t}^L P_{t0}^P = P_{0t}^P P_{t0}^L = 1.$$

Faktorumkehrprobe (F)

Die Wertsteigerung kann in das Produkt einer *nach der gleichen Indexformel* berechneten Preis- und Mengenkompente zerlegt werden. Fisher's "Idealindex"

$$(10.15) \quad P_{0t}^F = \sqrt{P_{0t}^L P_{0t}^P}$$

das geometrische Mittel aus der Laspeyres- und Paasche Preisindexformel (bei Mengenindizes analog Q^F als geometr. Mittel aus Q^L und Q^P) erfüllt F (und Z, nicht aber T), denn

$$(10.16) \quad W_{0t} = P_{0t}^F Q_{0t}^F$$

Zirkularität (Verkettbarkeit, Transitivität, T)

Nach dieser Forderung (auch "Rundprobe" ["circular test"]) soll für *beliebige*, Einteilungen des Intervalls $[0,t]$ in $[0,s]$ und $[s,t]$ also für *jedes* s gelten:

$$(10.17) \quad P_{0t} = P_{0s} P_{st}.$$

Die Einteilung in zwei Teilintervalle mit $0 < s < t$ ist nicht zwingend (es könnten auch drei oder mehr Teilintervalle sein, etwa $0 < r < s < t$, so dass gilt $P_{0t} = P_{0r} P_{rs} P_{st}$, oder auch $0 > s > t$). T wird oft dahingehend missverstanden, dass ein als Produkt definierter Index, wie der Kettenindex "verkettbar" sei. Dabei wird auch vergessen, dass bei Gl. 10.17 betont werden muss "für jedes s ".

Wenn Identität gilt, dann folgt Z aus T (Umkehrung gilt nicht)..

Umbasierung und Verkettung (vgl. Kap. 9)

Von Zeitumkehrbarkeit und Verkettbarkeit als Axiome ist zu unterscheiden, dass entsprechende Berechnungen [als Hilfslösungen] vorgenommen werden:

	gegeben	gesucht	Lösung
Umbasierung (rescaling)	ein Index zur Basis 0 für die Perioden 0, ..., s, ..., t	ein Index zur Basis s (meist: zur aktuelleren Basis s)	(10.18) $P_{st} = \frac{P_{0t}}{P_{0s}}$
Verkettung (splicing)	ein Index zur Basis 0 (berechnet mindestens bis zur Periode s) und einer zur Basis s	Bildung einer langen Reihe zur Basis 0 aus zwei oder mehreren sich überlappenden Indexreihen	(10.19) $P_{0t} = P_{0s} P_{st}$ (für $0 < s < t$)

Man sieht, dass die Rechenoperationen äquivalent sind und beide (Gl. 10.18 und 19) auf dem simplen "Dreisatz": $P_{0t}/P_{0s} = P_{st}/P_{ss}$ (mit $P_{ss} = 1$) beruhen, der jedoch - genau genommen - nicht zutreffend ist, wenn Verkettbarkeit (wie z.B. bei P^L und P^P) nicht erfüllt ist

Additivität (= Linearität) der Indexfunktion

(als spezielle Form der Monotonie) bedeutet in der Notation der Übersicht 10.3:

Fall a) unterschiedliche Preise in der Berichtsperiode:

$$P(\mathbf{p}_0, \mathbf{p}_t^*) = P(\mathbf{p}_0, \mathbf{p}_t) + P(\mathbf{p}_0, \Delta \mathbf{p}_t^*) \quad \text{wenn für } \mathbf{p}_t^*, \mathbf{p}_t \text{ und } \Delta \mathbf{p}_t^* \text{ gilt: } \mathbf{p}_t^* = \mathbf{p}_t + \Delta \mathbf{p}_t^*$$

und entsprechend

Fall b) unterschiedliche Preise in der Basisperiode:

$$[P(\mathbf{p}_0^*, \mathbf{p}_t)]^{-1} = [P(\mathbf{p}_0, \mathbf{p}_t)]^{-1} + [P(\Delta \mathbf{p}_0^*, \mathbf{p}_t)]^{-1} \quad \text{wenn entsprechend gilt: } \mathbf{p}_0^* = \mathbf{p}_0 + \Delta \mathbf{p}_0^*.$$

Die Indizes von Laspeyres und Paasche sind additiv.

Übersicht 10.3: Axiomensystem von Eichhorn und Voeller

Notation:

Preis- und Mengenvektoren (jeweils n Komponenten [Waren]) $\mathbf{p}_0, \mathbf{q}_0, \mathbf{p}_t, \mathbf{q}_t$. Die Indexfunktion $P: \mathbb{R}^{4n} \Rightarrow \mathbb{R}$ sollte danach die folgenden Axiome erfüllen:

P1	Monotonie: a) in Berichtspreisen $P(\mathbf{p}_0, \mathbf{p}_t^*) > P(\mathbf{p}_0, \mathbf{p}_t)$, wenn $p_{it}^* \geq p_{it}$ und für mindestens eine Ware i gilt: $p_{it}^* > p_{it}$ b) in Basispreisen $P(\mathbf{p}_0, \mathbf{p}_t) > P(\mathbf{p}_0^*, \mathbf{p}_t)$ wenn analog gilt: $p_{i0}^* \geq p_{i0}$ und $p_{i0}^* > p_{i0}$ für mindestens ein i (eine Ware)
P2	Lineare Homogenität: ^{a)} $P(\mathbf{p}_0, \mu \mathbf{p}_t) = \mu P(\mathbf{p}_0, \mathbf{p}_t)$ mit $\mu \in \mathbb{R}_+$ (nicht zu verwechseln mit Proportionalität: $P(\mathbf{p}_0, \mu \mathbf{p}_0) = \mu$, wobei $p_{it} = \mu p_{i0}$ für alle i)
P3	Identität: ^{b)} $P(\mathbf{p}_0, \mathbf{p}_t) = 1$ wenn $p_{it} = p_{i0}$ für alle i also $\mathbf{p}_t = \mathbf{p}_0$
P4	Dimensionalität: $P(\mu \mathbf{p}_0, \mu \mathbf{p}_t) = P(\mathbf{p}_0, \mathbf{p}_t)$ mit $\mu \in \mathbb{R}_+$ (Unabhängigkeit von der Währungseinheit der Preise)
P5	Kommensurabilität: $P(\Lambda \mathbf{p}_0, \Lambda \mathbf{p}_t, \Lambda^{-1} \mathbf{q}_0, \Lambda^{-1} \mathbf{q}_t) = P(\mathbf{p}_0, \mathbf{p}_t, \mathbf{q}_0, \mathbf{q}_t)$ mit $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ und $ \lambda_i > 0$ (Unabhängigkeit von der Mengeneinheit, auf die sich die Preisnotierung bezieht).

a) Unter Homogenität vom Grade -1 versteht man die Forderung $P(\mu \mathbf{p}_0, \mathbf{p}_t) = \mu^{-1} P(\mathbf{p}_0, \mathbf{p}_t)$. Sie ist erfüllt, wenn P2 und P4 gelten.

b) Axiome P2 und P3 stellen zusammen sicher, dass die sog. Proportionalitätsprobe erfüllt ist.

10.2 Kettenindizes

Die Standardkritik am direkten Preisindex nach Laspeyres P_{0t}^L (als Maß der Inflation) bzw. an Volumen, die man als Ergebnis einer Deflationierung mit einem (direkten) Paasche Preisindex erhält ist, dass in P_{0t}^L die Mengen, bzw. in P_{0t}^P die Preise (allgemein die Gewichte) für eine gewisse Zeit (in Interesse des reinen Preisvergleichs) konstant gehalten werden und dass das Wägungsschema veraltet. Es müsse stattdessen jeweils mit möglichst aktuellen Gewichten gerechnet werden. Nicht viel mehr als dies steckt hinter der in neuerer Zeit vehement wiederbelebten Forderung nach Kettenindizes⁴.

Die Definition eines Kettenindexes umfasst stets zwei Elemente (Übers. 10.4),

- die **Kette** $\bar{P}_{0t}^C$ ($C = \text{chain}$), das konstantes Element: Zwei-Periodenvergleich (zwischen 0 und t) indirekt als Produkt $P_{0t} = P_{01} P_{12} \dots P_{t-1,t}$ (analog zur Verkettung)
- (das variable Element) das **Kettenglied** $P_t^C = P_{t-1,t}$ (link), das je nach verwendeter Indexformel unterschiedlich ist, z.B. nach Laspeyres, Paasche usw. (Übers. 10.4)

Befürworter von Kettenindizes vergleichen meist P_t^C (statt $\bar{P}_{0t}^C$) mit P_{0t} . Dem Vorteil, dass $\bar{P}_{0t}^{LC}$ (auch) von den aktuelleren Mengen q_{t-1} abhängt, nicht nur von den "veralteten" Mengen

• _____

⁴ Sie sind (leider) in internationalen Empfehlungen für die Verwendung in der amtlichen Statistik vorgeschrieben worden.

10.4 Definition von Kettenindizes

Kettenindex

Definition der Kettenglieder (links)	Verkettung zur Kette
$P_t^C = P_{t-1,t}$ ein Index (genügt Axiomen)	$\bar{P}_{0t}^C = P_1^C P_2^C \dots P_t^C$ kein Index*
Beispiele: Laspeyres $P_t^{LC} = \frac{\sum p_t q_{t-1}}{\sum p_{t-1} q_{t-1}}$, Paasche $P_t^{PC} = \frac{\sum p_t q_t}{\sum p_{t-1} q_t}$	Beispiele: $\bar{P}_{0t}^{LC} = P_1^{LC} P_2^{LC} \dots P_t^{LC}$ $\bar{P}_{0t}^{PC} = P_1^{PC} P_2^{PC} \dots P_t^{PC}$
ein Warenkorb q_{t-1} bzw. q_t **	viele Warenkörbe $q_0, q_1, \dots$

* muss nicht Axiome (z.B. im Sinne von Übers. 10.3) erfüllen, selbst wenn das einzelne Kettenglied dies tut.

** bei unterjährigem Vergleich zum Vorjahr (ein Monat verglichen mit dem gleichem Monat im Vorjahr) jedoch schon beim einzelnen Kettenglied **zwei** Warenkörbe.

q_0 wie in P_{0t} stehen folgende **Nachteile** gegenüber:

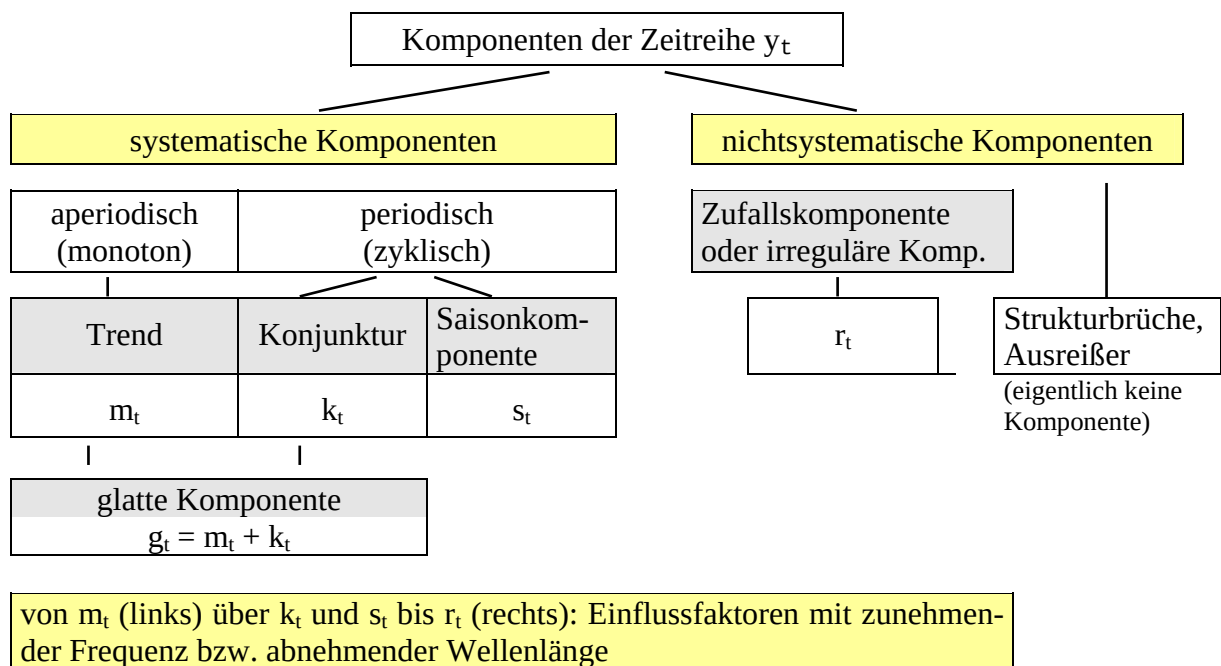
1. Kettenindizes erlauben keine **Interpretation** im Sinne des reinen Preisvergleichs, als Messzahlenmittelwert oder als Verhältnis von Aggregaten. Axiome sind auf sie nicht anwendbar: Trotz gleicher Preise in Periode 0 und 2 muss nicht gelten $P_{02} = 1$ (Identität verletzt, ebenso können Monotonie und andere Axiome von Übers. 10.3 verletzt sein).
2. Verkettung als Form der **zeitlichen Aggregation** ist **pfadabhängig**: ein Kettenindex ist kein Zwei-Perioden-**Vergleich**, sondern ein summarisches Maß für die Gestalt einer Zeitreihe (für einen **Verlauf**). Das Ergebnis für das Intervall von 0 bis t ist i.d.R. unterschiedlich, je nach dem, wie es in Teilintervalle zerlegt wird und wie sich Preise und Mengen in den Zwischenperioden 1,...,t-1 entwickeln. Bei zyklischer Bewegung der Preise (der Verlauf zwischen 0 und t wiederholt sich) kann die Kette für Periode 2t, 3t, ... im Wert ständig zunehmen (wenn der Index $P_{0t} > 1$ ist) oder abnehmen (wenn $P_{0t} < 1$), selbst dann wenn die Preise in 0, t, 2t,... alle gleich sind.
3. Ungünstige Aggregationseigenschaften: **additive** - und (bei Deflationierung) **strukturelle Konsistenz** nicht erfüllt. Volumen V_t nicht nur abhängig von q_t und p_0 , auch von Preisen $p_1, \dots, p_t$, so dass man kaum von "in konstanten" Preisen sprechen kann.
4. Erheblicher **Mehraufwand** für Datenbeschaffung (häufigere Feststellung der Warenkörbe).

Kapitel 11: Einführung in die Zeitreihenanalyse

1. Komponentenmodell

Def. 11.1: (Zeitreihe, Ursprungswerte)

Eine Folge von Beobachtungswerten y_t mit $t = 1, 2, \dots, T$ und einer natürlichen Ordnung dergestalt, dass die Werte in der Reihenfolge $y_1, y_2, \dots$ beobachtet wurden, heißt "Zeitreihe". Die (meist diskrete) Variable t ist i.d.R. die Zeit, wobei die Werte t Zeitpunkte oder Zeitintervalle darstellen. Die noch nicht durch eine Zeitreihenanalyse bearbeiteten (z.B. transformierten) Beobachtungswerte y_t heißen "Ursprungswerte".



Verknüpfung (Überlagerung) der Komponenten

Man unterscheidet zwei Grundmodelle:

- 1) additive Überlagerung: wenn bei steigendem bzw. fallendem Trend die zyklischen Einflüsse große Ausschläge besitzen, so dass die Schwankungen k_t , s_t und r_t vom Niveau der Zeitreihe y_t unabhängig sind. Es gilt dann:

$$(11.1) \quad y_t = m_t + k_t + s_t + r_t \quad (\text{additive Überlagerung})$$

- 2) multiplikative Überlagerung: wenn z.B. die Schwankungen der zyklischen Komponenten, insbesondere der Saisonkomponente mit steigendem Niveau der Zeitreihe zunehmen:

$$(11.2) \quad y_t = m_t k_t s_t r_t \quad (\text{multiplikative Überlagerung})$$

2. Methoden zur Berechnung des Trends (der glatten Komponente)

1. Trendberechnung mit der Methode der kleinsten Quadrate

Bei einem Trend mit einer Funktion bestimmten Typs (lineare -, Exponential-, Potenzfunktion usw.) können die Parameter nach der Methode der kleinsten Quadrate bestimmt werden. Der Regressand (die abhängige Variable) ist wie in Kap. 8 die Variable Y mit den Beobachtungen y_t und an die Stelle der unabhängigen Variable X tritt bei der Trendfunktion die Zeit t . Die Gesamtheit der übrigen Kom-

ponenten mit den Messwerten k_t , s_t und r_t ist das Residuum u_t .

Die Parameter a und b eines linearen Trends $y_t = m_t = a + bt$ ($t = 1, 2, \dots, T$) werden mit den Normalgleichungen wie folgt

(11.3a)	$aT + b \sum t = \sum y_t$	1. Normalgleichung
(11.3b)	$a \sum t + b \sum t^2 = \sum ty_t$	2. Normalgleichung

Zweckmäßig ist es, für t die Werte $\dots -2, -1, 0, +1, +2, \dots$ zu vergeben, so dass $\sum t = 0$ (statt $T(T+1)/2$ wie bei $t=1, 2, \dots, T$) ist und a und b direkt aus jeweils einer der beiden Normalgleichungen zu bestimmen ist: $a = (\sum y)/T$ und $b = (\sum ty)/T^2$.

2. Trendberechnung mit der Methode der gleitenden Durchschnitte

Gleitende Durchschnitte sind eine Folge von arithmetischen Mitteln, die aus jeweils p aufeinanderfolgenden Werten y_t der Zeitreihe gebildet werden.

Def. 11.2: (Gleitende Durchschnitte)

a) Der dem Ursprungswert y_t zugeordnete gleitende p -gliedrige Durchschnitt lautet bei ungeradzahligem $p = 2k+1$

$$(11.4) \quad \tilde{y}_t = \frac{1}{p} \sum_{h=-k}^k y_{t+h} \quad (p = 2k + 1, \text{ ungeradzahlig}) \quad \text{oder}$$

$$\tilde{y}_t = (y_{t-k} + y_{t-k+1} + \dots + y_t + \dots + y_{t+k-1} + y_{t+k})/p.$$

b) Bei geradzahligem $p = 2k$ wäre der Durchschnitt $\tilde{y}_t$ der Periode $t - \frac{1}{2}$ und $\tilde{y}_{t+1}$ der Periode $t + \frac{1}{2}$ zuzuordnen. Es liegt daher nahe einen ungewogenen Durchschnitt hieraus zu berechnen. Dieser der Periode t zugeordnete zentrierte gleitenden Durchschnitt lautet: $\frac{1}{2}$

$$(11.5) \quad \tilde{y}_t^z = \frac{1}{p} \left(\sum_{h=-(k-1)}^{k-1} y_{t+h} + \frac{y_{t-k} + y_{t+k}}{2} \right) \quad (p = 2k, \text{ geradzahlig}) \quad \text{etwa bei } p = 4$$

$$\tilde{y}_t^z = \frac{1}{4} \left(\frac{1}{2} y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2} y_{t+2} \right)$$

Am Anfang und Ende fallen beim gleitenden Durchschnitt jeweils k Glieder weg. Der erste gleitende Durchschnitt fällt auf den $k+1$ -ten Wert.

	$p = 2k + 1$ (ungerade)	$p = 2k$ (gerade)
es fallen weg	$k = (p-1)/2$	$k = p/2$
der erste Wert	$k + 1 = (p + 1)/2$	$k + 1 = p/2 + 1$

3. Berechnung der Saisonkomponente

Konstante (starre) Saisonfigur (Saisonnormale) bei **additiver** Überlagerung

Ursprungswerte y und trendbereinigte Werte y^* (bzw. bereinigt von glatter Komponente) für die Jahre $j = 1, 2, \dots, J$ und Unterzeitraum $z = 1, 2, \dots, Z$ (Z = Anzahl der Unterzeiträume, bei Quartalen $Z = 4$, bei Monatsdaten $Z = 12$)

$$(11.7) \quad y_{jz}^* = y_{jz} - g_{jz}, \quad \text{etwa mit } g_{jz} = \tilde{y}_{jz}$$

nicht-normierte Saisonnormale

$$(11.8) \quad S_z = \frac{1}{J} \sum_{j=1}^J y_{jz}^* \quad \text{mit dem Mittel} \quad \bar{S} = \frac{\sum_z S_z}{Z}$$

normierte (auf einen mittleren Wert 0) Saisonnormale (11.10) $S_z^* = S_z - \bar{S}$.

Bei multiplikativer Überlagerung Division statt Subtraktion

$$(11.7a) \quad y_{jz}^* = y_{jz} / g_{jz},$$

jedoch S_z und mittlere Saisonnormale $\bar{S}$ auch als arithmet. Mittel und

$$(11.10a) \quad S_z^* = S_z / \bar{S}$$

(Division durch mittlere nicht normierte Saison), normiert auf ein Mittel von 1.

4. Hinweise auf weiterführende Verfahren

4.1 Exponential Smoothing (exponentielles Glätten)

1. Prognose als gewogenes Mittel aus den letzten Werten

$$(11.11) \quad y_{t+1}^p = \alpha y_t (1 - \alpha) y_t^p \quad \text{mit: } 0 < \alpha < 1.$$

2. Prognose als Mittel aller vergangener Beobachtungen

$$(11.12) \quad \begin{aligned} y_{t+1}^p &= \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots + \alpha(1 - \alpha)^n y_{t-n} + (1 - \alpha)^{n+1} y_{t-n}^p \\ &= \sum_{i=0}^n \alpha(1 - \alpha)^i y_{t-i} + (1 - \alpha)^{n+1} y_{t-n}^p. \end{aligned}$$

3. Prognose als partielle (mit α gewogene) Korrektur einer Fehlschätzung F

$$(11.13) \quad y_{t+1}^p = y_t^p + \alpha(y_t - y_t^p) = y_t^p + \alpha F.$$

4.2 Filter, Operatoren, Polynome

1. Ein "Filter" verwandelt eine Zeitreihe y_t (input) in eine transformierte Zeitreihe (output) z_t . Einfache lineare Filter sind z.B. gleitende Mittelwerte oder Differenzenbildung (Output $z_t = y_t - y_{t-1}$). Ein nichtlinearer Filter ist z.B. die Bildung von Wachstumsraten $r_t = (y_t - y_{t-1})/y_{t-1}$.
2. Operatoren: Verschiebungen der Variable t bewirkt der Backshift- oder Lag-Operator: $Ly_t = y_{t-1}$, $L^2 y_t = y_{t-2}$ usw. Nicht auf t , sondern auf die Inputvariable wirken der Vorwärtsdifferenzenoperator (delta Δ) mit $\Delta y_t = y_{t+1} - y_t$, bzw. die Rückwärtsdifferenzen (nabla ∇) $\nabla y_t = y_t - y_{t-1}$. Hintereinanderausführen heißt Potenzieren des Operators $\nabla^2 y_t = \nabla y_{t+1} - \nabla y_t = y_{t+2} - 2y_{t+1} + y_t$. Man beachte, dass $\nabla^2 y_t$ nicht identisch ist mit $y_{t+2} - y_t$. Vor- und Rückwärtsdifferenzen für mehrere Perioden, etwa $\nabla_4 y_t = y_t - y_{t-4}$ oder $\nabla_{12} y_t = y_t - y_{t-12}$ beim Vorjahresvergleich mit Quartals- oder Monatsdaten sind "saisonale Differenzen".
3. Der Ausdruck $A_p(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_p t^p$ ist ein Polynom in t vom Grade p und $A_p(\cdot)$ heißt Polynomenoperator. Ein autoregressives Schema (eine linear-rekursive Funktion) ist $A_p(L)y_t = a_0 + a_1 L + a_2 L^2 + \dots + a_p L^p y_t$. Lineare Filter kann man als Lagpolynome darstellen und Polynome in t als linear-rekursive Funktionen. Einem Polynom $y = A_p(t)$ ist eine linear rekursive Funktion $B_{p+1}(L)y$ äquivalent:

p=1: das der Funktion $y_t = a_0 + a_1 t$ (Polynom vom Grade 1) äquivalente Lagpolynom ist $y_t = 2y_{t-1} - y_{t-2}$ mit den Anfangswerten $y_0 = a_0$ und $y_1 = a_0 + a_1$.

p=2: der Funktion $y_t = a_0 + a_1 t + a_2 t^2$ äquivalent ist $y_t = 3y_{t-1} - 3y_{t-2} + y_{t-3}$ mit den Anfangswerten $y_0 = a_0$, $y_1 = a_0 + a_1 + a_2$ und $y_2 = a_0 + 2a_1 + 4a_2$.

Kapitel 12: Bestandsanalyse und Tafelrechnung

12.1. Bestands- und Bewegungsmassen

Def. 12.1: (Bestandsmasse, Bewegungsmasse, Verweildauer)

- a) Eine statistische Masse, deren Einheiten ($i=1,2,\dots,n$) jeweils gemeinsam zu einem bestimmten Zeitpunkt t_j in einem Bestand (über eine nicht näher bestimmte Zeit) verweilen, heißt **Bestandsmasse** (engl. stock).
Der Umfang der Bestandsmasse zum Zeitpunkt t_j heißt Bestand $B(t_j) = B_j$. Er ist zu jedem Zeitpunkt $t = t_j$ durch die Bestandsfunktion $B(t)$ gegeben. Die Zeit kann als diskrete ($t = t_0, t_1, \dots, t_j, \dots, t_m$) oder stetige Variable betrachtet werden.
- b) Eine statistische Masse, deren Einheiten dadurch charakterisiert sind, dass sie zu einem bestimmten Zeitpunkt ihren Zustand ändern (was ein "Ereignis" darstellt), heißt **Bewegungsmasse** (Ereignismasse, Stromgröße, engl. flow).
Der Umfang einer Bewegungsmasse ist die Anzahl derartiger Ereignisse in einem gegebenen Zeitraum (Zeitintervall). Zustandsänderung kann insbesondere bedeuten: Zugang zu oder Abgang von einer Bestandsmasse.
- c) Jede Einheit einer Bewegungsmasse ($i=1,2,\dots,n$) ist durch Zugangszeit (t_{zi}) und Abgangszeit (t_{ai}) gekennzeichnet. Der Zeitraum zwischen Zu- und Abgangszeit $d_i = t_{ai} - t_{zi}$ heißt **Verweildauer**.

Methoden der Erhebung von Bestands- und Bewegungsmassen:

1. Feststellung der Bewegungen (Bewegungsmassen)

- a) durch individualisierte Erhebung aller Verläufe, d.h. für jede Einheit werden Zugangs- und Abgangszeit festgestellt (= Längsschnitts- oder Verlaufsanalyse);
- b) laufende Registrierung aller Bestandsveränderungen und Auswertung der über ein Beobachtungsintervall (von t_0 bis t_j) kumulierten Zugänge (Z_{oj}) und Abgänge (A_{oj}), d.h. der Bruttoströme.
- c) Feststellung der Bestandsveränderungen (d.h. der Salden- oder Nettoströme $Z_{oj}-A_{oj}$).

2. Feststellung der Bestände (Bestandsmassen)

- a) durch periodische Inventuren (Zählen oder Messen)
- b) durch **Fortschreibung** für das Intervall $[t_0, t_j]$: (12.1) $B_j = B_0 + Z_{oj} - A_{oj} \quad (j = 0, 1, \dots, m)$
In Gl. 12.1 ist B_0 der Anfangsbestand, B_j der Bestand zum Zeitpunkt t_j , Z_{oj} die Anzahl der Zugänge und A_{oj} die Anzahl der Abgänge im Beobachtungsintervall $[t_0, t_j]$.
- c) Bei Kenntnis sämtlicher individueller Verläufe (wie in 1a), also bei Längsschnittsdaten, ist der Bestand zu jedem beliebigen Zeitpunkt bekannt.

Querschnittsanalysen sind die Kombination 1b + 2a
Längsschnittsanalysen die Kombination 1a + 2c.

Beckersches Diagramm, Bestandsfunktion und Zeitmengenfläche

1. Beckersches Diagramm: Eine graphische Darstellung der individuellen Verläufe ist das Beckersche Diagramm (Abb. 12.1 für Aufgabe 12.1 [siehe unten Aufgabenteil]).

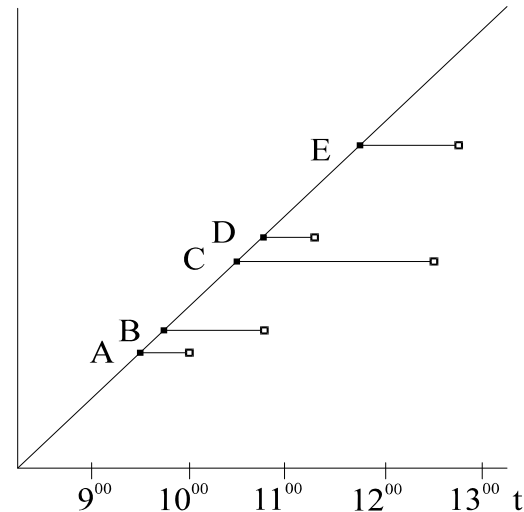
2. Bestandsfunktion: Es ist leicht zu sehen, wie aus dem Beckerschen Diagramm (oberer Teil von Abb. 12.1) die Bestandsfunktion $B(t)$ (t stetig), bzw. B_j (Bestände zu den diskreten Zeitpunkten t_j) herzuleiten ist. Mit jedem Zugang (Abgang) einer Einheit erhöht (verringert) sich

die Bestandsfunktion um 1.

3. Zeitmengenfläche: Die schraffierte Fläche unter der Bestandsfunktion heißt Zeitmengenfläche F , oder genauer F_{om} wenn die Fläche "über" dem Intervall $[t_o, t_m]$ betrachtet wird.

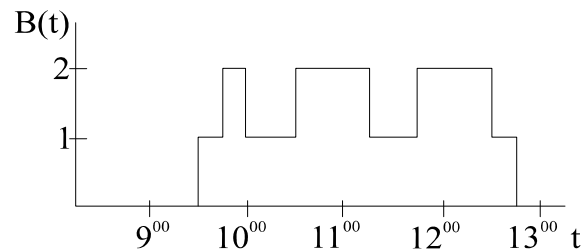
Abb. 2.1 Beckersches Diagramm und Bestandsfunktion (Bsp.)

	Zeitpunkt	
	Zugang	Abgang
A	09 ³⁰	10 ⁰⁰
B	09 ⁴⁵	10 ⁴⁵
C	10 ³⁰	12 ³⁰
D	10 ⁴⁵	11 ¹⁵
E	11 ⁴⁵	12 ⁴⁵



Def. 12.2: (offene-, geschlossene Masse)

Eine Bestandsmasse heißt geschlossen bezüglich des Zeitintervalls $[t_o, t_m]$, wenn keine ihrer Einheiten vor t_o zugegangen ist und nach t_m abgeht (endgültig aus dem Bestand ausscheidet). Eine Masse, die nicht beidseitig geschlossen ist, heißt offene Masse. Man kann auch halbseitig und beidseitig offene Massen unterscheiden.



12.2. Kennzahlen der Dynamik eines Bestands: Durchschnittsbestand, durchschnittliche Verweildauer, Umschlagshäufigkeit

Berechnung der Kennzahlen (vgl. Übers. 12.1)

a) bei Kenntnis der individuellen Verläufe (Längsschnittsdaten)

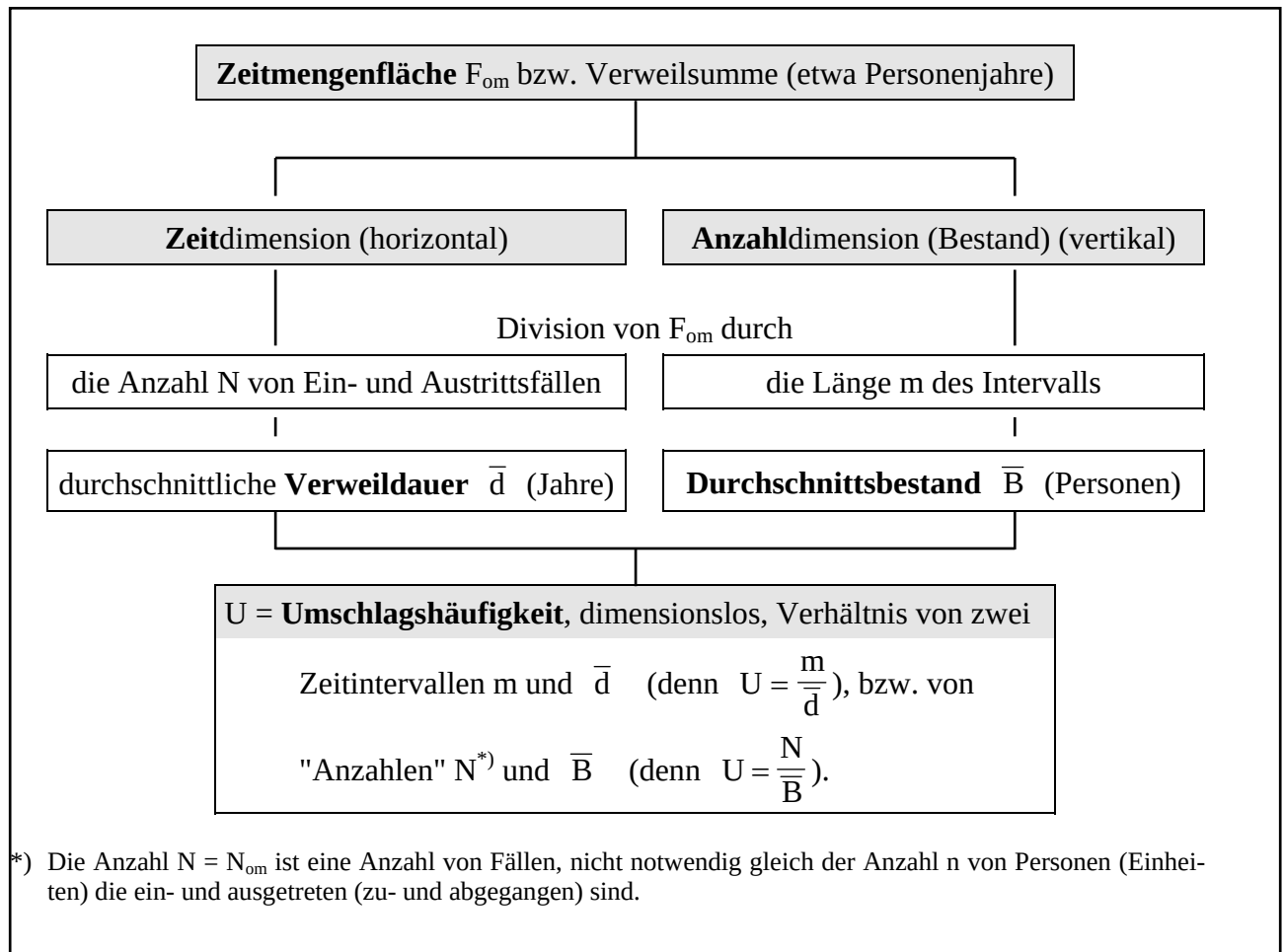
$$(12.3) \quad \bar{B} = \frac{F_{om}}{m} \quad (\text{Durchschnittsbestand})$$

$$(12.4) \quad \bar{d} = \frac{\sum d_i}{N_{om}} \quad (\text{durchschnittliche Verweildauer}).$$

Bei geschlossener Masse ist Zeitmengenfläche = Verweilsumme (und deshalb $\bar{d} = \frac{F_{om}}{N_{om}}$)

$$(12.5) \quad U = \frac{m}{\bar{d}} = \frac{N_{om}}{\bar{B}} \quad (\text{Umschlagshäufigkeit})$$

Übers. 12.1: Zusammenhänge zwischen Kennzahlen zur Beschreibung der Bestandsentwicklung



b) bei Querschnittsdaten

Zeitmengenfläche und Durchschnittsbestand

Finden die Bestandsänderungen ausschließlich genau zu den Beobachtungszeitpunkten t_j ($j = 1, \dots, m$) statt, dann ist die Zeitmengenfläche $F_{om} = \sum_j B_{j-1} (t_j - t_{j-1})$. Sind die Beobachtungszeitpunkte t_j (mit $j = 1, 2, \dots, m$) äquidistant, so dass $t_j - t_{j-1} = 1$ (für alle j) und $t_m - t_0 = m$, so gilt:

$$(12.10) \quad F_{om} = \frac{1}{2} B_0 + B_1 + \dots + B_{m-1} + \frac{1}{2} B_m \quad (\text{chronologisches Mittel})$$

woraus $\bar{B}$ wieder mit Gl. 12.3 zu errechnen ist.

Durchschnittliche Verweildauer und Umschlagshäufigkeit

Es ist nicht mehr von $\sum d_i = F_{om}$ auszugehen. Vielmehr ist F_{om} zu korrigieren um die Zeiten, welche

- die B_0 Einheiten des Anfangsbestands vor t_0 bereits dem Bestand angehört hatten (*Aufbauzeiten* d_0) und die Zeiten, welche die
- B_m Einheiten des Endbestands nach t_m dem Bestand noch angehören werden (*Abbauzeiten* d_m).

F_{om} ist also mit (geschätzten) durchschnittlichen Auf- und Abbaueiten zu G_{om} (geschätzte Verweilsomme Σd_i) zu korrigieren

$$(12.6) \quad G_{om} = B_o \bar{d}_o + F_{om} + B_m \bar{d}_m$$

$$(12.8) \quad \bar{d} = \frac{G_{om}}{N_{om}}.$$

Es sind jetzt Annahmen über die mittlere Aufbau- und Abbaueit nötig. üblich ist die Annahme $\bar{d}_o = \delta \bar{d}$ und $d_m = (1 - \delta) \bar{d}$ mit $0 < \delta < 1$ liefert das

$$(12.11) \quad \bar{d} = \frac{F_{om}}{\delta Z_{om} + (1 - \delta) A_{om}} \quad \text{und mit } \delta = 1/2 \text{ die bekannten Formel}$$

$$(12.12) \quad \bar{d} = \frac{2m\bar{B}}{Z_{om} + A_{om}}.$$

12.3. Stationäre Bevölkerung und Tafelrechnung

Def. 12.3: (Kohorte, Abgangsordnung, stationäre Bevölkerung)

- Eine Zugangskohorte oder einfach **Kohorte** ist die Gesamtheit der gleichzeitig (zum gleichen Zeitpunkt t_j , bzw. im gleichen Intervall geringer Länge $[t_{j-1}, t_j]$) zugehenden Einheiten. Der Umfang dieser Masse, d.h. die Anzahl der zugehenden Einheiten ist l_o .
- Die **Abgangsordnung** l_x (wobei $x = 0, 1, \dots, w$ das Alter, d.h. die Anzahl der vollendeten Jahre ist) ist die Anzahl der Überlebenden des Alters x . Es ist der Restbestand einer Geburtskohorte des Umfangs l_o nach Vollendung von x Jahren. l_x ist monoton fallend.
- Bei einer **stationären Bevölkerung** (Sterbetafelbevölkerung) wird jede Kohorte (jeder Geburtsjahrgang) in jedem aufeinanderfolgenden Intervall (in allen folgenden Jahren) durch eine gleich große Kohorte (so dass für die Zugänge Z gilt $Z_{j-1,j} = l_o$ für alle j) mit gleicher Abgangsordnung ersetzt (d.h. gleicher "Struktur"; l_x ist nicht von j sondern nur von x abhängig).

Def. 12.4 (Tafelfunktionen l , q , p , d , L):

- Die einjährige Sterbewahrscheinlichkeit q_x der x -jährigen ist die (bedingte) Wahrscheinlichkeit dafür, dass eine Person, die das Alter von x erreicht hat, das Alter von $x+1$ nicht mehr erreichen wird (mit $x = 0, 1, \dots, w$ für das Alter in vollendeten Jahren). Die einjährige Überlebenswahrscheinlichkeit p_x ist demzufolge $p_x = 1 - q_x$. Auch p_x ist eine bedingte Wahrscheinlichkeit.
- Sämtliche Sterbetafelfunktionen sind allein Funktionen des Alters x und sie sind mit der Folge der Sterbewahrscheinlichkeiten q_x und dem willkürlich gewählten Anfangsbestand (Geburten) l_o eindeutig gegeben:

die Absterbeordnung l_x ist ausgehend von einem fiktiven Anfangsbestand von $l_o = 100.000$ Personen rekursiv zu berechnen mit

$$(12.18) \quad l_{x+1} = l_x p_x = l_x (1 - q_x).$$

Entsprechend ist die Anzahl $d_x \geq 0$ der im Altersintervall $(x, x+1)$ gestorbenen Personen

$$(12.19) \quad d_x = l_x q_x = l_x - l_{x+1}.$$

Wie man leicht sieht, ist $\sum_x d_x = l_0$.

- d) Mit L_x wird die Anzahl der von allen Überlebenden x -jährigen Personen bis zum Alter $x+1$ durchlebten Jahre (die Anzahl der im Intervall $(x, x+1)$ verlebten Personenjahre [eine Zeitmengenfläche, bzw. lineare Interpolation der Abgangsordnung l_x]) bezeichnet.

$$(12.20) \quad L_x = \frac{1}{2}(l_x + l_{x+1}).$$

Def. 12.5: (Tafelfunktionen T , T^* , e , e^*)

- a) Die Tafelfunktion T_x , die Zahl der von den Überlebenden des Alters x noch zu durchlebenden Jahre ist die Summe der Größen $L_x, L_{x+1}, L_{x+2}, \dots, L_w$.

$$(12.24) \quad T_x = \sum_{y=x}^w L_y \quad x \leq y \leq w.$$

$$(12.25) \quad T_x^* = \sum_{y=x}^w l_y = T_x + \frac{1}{2}l_x$$

Die Größen T_x und T_x^* sind Verweilsommen; Maßeinheit: "Personenjahre".

- b) Dividiert man T_x bzw. T_x^* durch die Anzahl der Überlebenden des Alters x , also durch l_x , so erhält man mit

$$(12.26) \quad e_x = \frac{T_x}{l_x} = \frac{T_x^*}{l_x} - \frac{1}{2} = e_x^* - \frac{1}{2}$$

die (mittlere, durchschnittliche) weitere **Lebenserwartung** einer x -jährigen Person (spricht man von "der" Lebenserwartung, so ist e_0 gemeint). Die Größe e_0 ist zugleich das durchschnittliche Sterbealter der stationären Bevölkerung, eine Verweilsomme, und es gilt

$$\text{Bestand } (T_0) = \text{Zugang } (l_0) \cdot \text{durchschnittliche Verweildauer } (e_0)$$

bei einer **stationären** Bevölkerung.

Ende des Formelteils

Teil II: Übungsaufgaben Deskriptive Statistik

Aufgaben zum Kapitel 1

Aufgabe 1.1

Welche der nachstehenden Massen sind Bestandsmassen und welche sind Bewegungsmassen?

Auftragseingänge, Anlagevermögen, Sozialprodukt, Nettoinvestition, landwirtschaftliche Nutzfläche, Ehescheidungen, Steuereinnahmen, Sparvolumen, Schulden, Neuverschuldung, Baufertigstellungen, Gutschriften auf ein Bankkonto, Geburten.

Aufgabe 1.2

Das Merkmal "Art der Erkrankung" wird üblicherweise durch ärztliche Diagnose erhoben.
Ist es:

- a) quantitativ oder qualitativ?
- b) häufbar oder nicht häufbar?
- c) intensiv oder extensiv?
- d) manifest oder latent?



Aufgabe 1.3

Die Pizzeria P (des Eigentümers P) hat zwei Lokale (L1 und L2), bei denen man Mittags- und Abendessen (M, A) einnehmen kann, wobei es jedoch jeweils nur die folgenden Gerichte gibt: Pizza, Spaghetti, Ravioli und Canneloni. Es ergab sich, dass von den 4764 Gästen der Pizzeria insgesamt 5000 Gerichte im letzten Monat (April) wie folgt bestellt wurden:

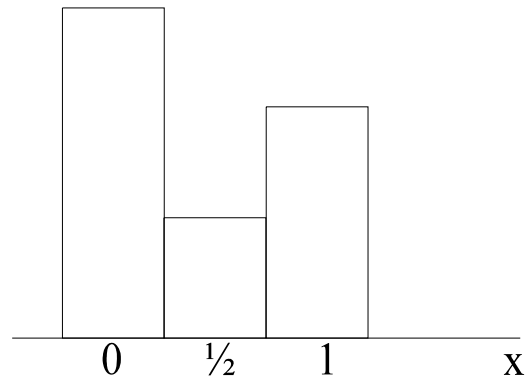
	L1		L2		insgesamt
	M	A	M	A	
Pizza	400	600	600	800	2400
Sonstige	700	1100	400	400	2600
Summe	1100	1700	1000	1200	5000

1. Wieviel Merkmale werden in dieser Statistik dargestellt, wie heißen sie und welche Merkmalsausprägungen werden in der Tabelle dargestellt?
2. Was (Masse, Einheit, Merkmal usw.) ist im Falle dieser Statistik
 - a) die in der Statistik mitgezählte Pizza, die Herr Schulze am 16. April zum Abendessen im Lokal L1 gegessen hat?
 - b) die Angabe "Pizza"?
 - c) die Angabe des Eigentümers P der Pizzeria?
 - d) die Zahl 5000?
 - e) das Lokal L2?

- f) die insgesamt $1700 + 1200 = 2900$ Gerichte, die abends ausgegeben wurden?
 g) die insgesamt 2800 Gerichte im Lokal L1?
 h) der Monat April?

Aufgabe 1.4

Stultitia (= Torheit) und ihre Ziehmutter und Nymphe Apaedia (= Beschränktheit), Tochter des Pan und Schutzgöttin der Statistiker, haben zusammen eine Volkszählung im Olymp veranstaltet. Sie zählten dabei 39 Götter, 13 Halbgötter (deshalb $x = \frac{1}{2}$) und 26 Nymphen (= niedrige Naturgötter) und ähnliche Dienstgrade. Gegen die graphische Darstellung ihrer Ergebnisse sind gewichtige Einwendungen zu erheben! Welche?



Aufgabe 1.5

Gegeben seien die folgenden Merkmalswerte

$$x_1 = 10 \quad ; \quad x_2 = 15 \quad ; \quad x_3 = 25 \quad ; \quad x_4 = 30$$

sowie die transformierten Werte

$$x_1^* = 15 \quad ; \quad x_2^* = 20 \quad ; \quad x_3^* = 30 \quad ; \quad x_4^* = 35$$

Welche Transformation $x^* = f(x)$ liegt hier vor und bei welcher Skalenart ist diese Art der Transformation zulässig:

Transformation (T)		Skalenniveau	
	monotone T	Nominalskala	
	lineare T	Ordinalskala	
	symmetrische T	Intervallskala	
	Intervalltransf.	Ratioskala	
	proportionale T	Absolutskala	

Aufgabe 1.6

Welches Skalenniveau wird mindestens vorausgesetzt bei der Berechnung der nachfolgend genannten Mittelwerte?

N = Nominalskala

O = Ordinalskala

I = Intervallskala

R = Ratioskala

A = Absolutskala

Bitte Buchstabe in die freie Fläche eintragen

Modus	
Median	
arithmetisches Mittel	
geometrisches Mittel	

Aufgabe 1.7

Schauspieler S (Rollenfach: jugendlicher Naturbursche) ist auf Tarzan-Filme spezialisiert. Gelegentlich spielt er auch in Krimis und Heimatfilmen mit. Sein Produzent führte die folgende Statistik:

Art des Films	Anzahl der Filme	darunter mit S
Tarzan	10	9
Krimis	12	2
Heimatfilme	8	1
Sex-Filme	82	0
Summe	112	12

- a) Kreuzen Sie jeweils an, ob eine der folgenden Methoden sinnvoll auf die Tabelle angewendet werden kann.



Berechnung bzw. Darstellung von	sinnvoll	sinnlos
1. arithm. Mittel	[]	[]
2. Modus (dichtester Wert)	[]	[]
3. Kreisdiagramm	[]	[]
4. Block-, Balkendiagramm	[]	[]

- b) Für die Merkmale "Art des Films" (A) und "Anzahl der Filme" (Z) gilt (Richtiges ankreuzen)

das Merkmal ist	gilt für Z	gilt für A
1. stetig	[]	[]
2. diskret	[]	[]
3. nominalskaliert	[]	[]
4. metrischskaliert	[]	[]

Aufgabe 1.8

Statistiker mögen festgestellt haben, dass zahlreiche Menschen Emotionen gegenüber Statistik und Mathematik haben. Schon der Gedanke an diese Dinge versetzt sie derartig in Angst (A) und Schrecken, dass sie keines klaren Gedankens mehr fähig sind. Nur wenige Menschen reagieren ohne Angst (N). Die Emotion sei zudem nicht mit Alter (in vollendeten Jahren) oder Bildungsstand korreliert. Als Beweis für diese Behauptung betrachte man die folgenden Daten über neun Personen:

Aufgaben zum Kapitel 2

Aufgabe 2.1

Für Umzugsunternehmen und Hausratsversicherungen etc. entwickelte Diplom Kaufmann K aus E die folgende Klassifikation von Einrichtungsgegenständen:

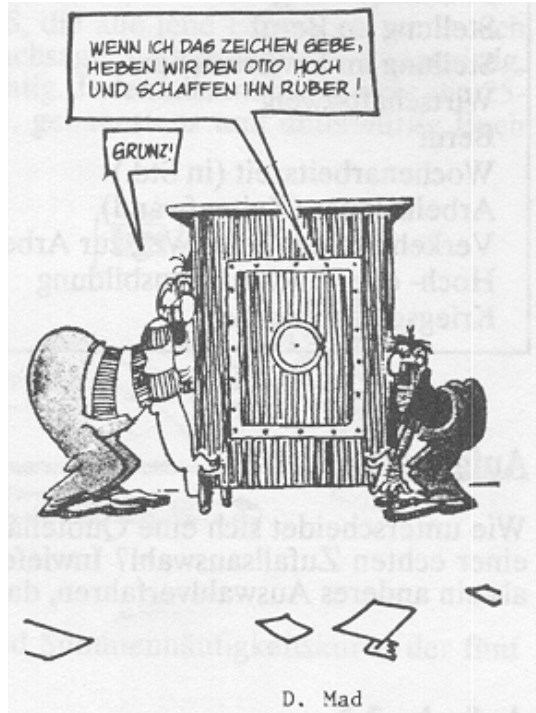
$M_x = \{\text{Schwermöbel, Leichtmöbel, langlebige Gebrauchsgegenstände, kurzlebige Verbrauchsgüter}\}$

Man ordne die folgenden Gegenstände aus dem Büro des Rechtsanwalts R diese Merkmalsausprägungen zu:

$X = \{\text{Gesetzesbücher, Tisch, Stuhl, Telefon, Geldschrank, Papierkorb, Teppich, Scheckbuch, Bleistifte, Whisky - Flasche}\}$

Erläutern Sie anhand dieses Beispiels die Begriffe Masse, Einheit, Merkmal und Merkmalsausprägung!

Welche Einwände lassen sich gegen die Klassifikation von K erheben? Ist sie vollständig und widerspruchsfrei?



Aufgabe 2.2

Nennen Sie beispielhaft einige Merkmalsausprägungen und den Skalentyp (Nominal-, Ordinal-, Intervall-, Ratioskala) bei der Messung der folgenden Merkmale (Merkmale der Volkszählungen):

Alter, Familienstand, Staatsangehörigkeit, Beruf, Stellung im Beruf, Schulbildung, Stellung im Erwerbsleben, Wochenarbeitszeit.

Aufgabe 2.3

In der Volks- und Berufszählung vom 27.05.1970 wurden folgende Merkmale erhoben, für die der Skalentyp durch ein Buchstabensymbol anzugeben ist:

N = Nominalskala

O = Ordinalskala

I = Intervallskala

R = Ratioskala

A = Absolutskala

Merkmal	Skalentyp
Alter	
Familienstand	
Eheschließungsjahr	
Religionszugehörigkeit	
Staatsangehörigkeit	
Stellung im Beruf	
Stellung im Erwerbsleben	
Wirtschaftszweig	
Beruf	
Wochenarbeitszeit (in Std.)	
Arbeitsdauer (Zeitaufwand)	
Verkehrsmittel beim Weg zur Arbeit	
Hoch- oder Fachschulausbildung	
Kriegsgefangenschaft	

Aufgabe 2.4

Wie unterscheidet sich eine Quotenauswahl ("Repräsentativer Bevölkerungsquerschnitt") von einer echten Zufallsauswahl? Inwiefern ist eine echte Zufallsauswahl (= Stichprobe) "besser" als ein anderes Auswahlverfahren, das nicht vom Zufall bestimmt ist?

Aufgabe 2.5

Ein Betrieb hat zwei Zweigniederlassungen (A und B). Für die Anzahl der Beschäftigten, das Durchschnittseinkommen (E), Gewinn (G) und Kapital (K) gilt:

Beschäftigte	E	G	K
A $n_A = 400$	1560	300	2000
B $n_B = 300$	1720	800	2000

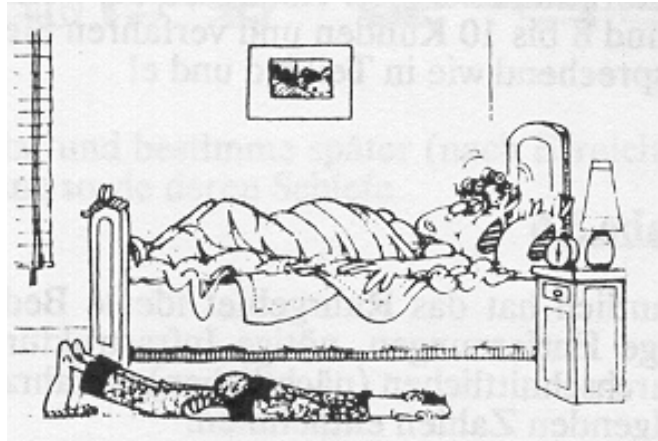
- Wie viele und welche Merkmale sind in dieser Tabelle dargestellt?
- Was sind die Merkmalsausprägungen?
- Welche Größen sind absolute Häufigkeiten und welche sind Merkmalssummen?
- Wie kann man welches Merkmal geeignet graphisch darstellen?
- Wie groß ist das Durchschnittseinkommen der Beschäftigten im Gesamtunternehmen?

Aufgaben zum Kapitel 3

Aufgabe 3.1

Diplom-Kaufmann K aus E heiratete eine Statistikerin S, die alle jene Eigenschaften in sich vereinigte, die man gemeinhin den bösen Statistikern nachsagt: S ist pedantisch, kleingeistig, humorlos und bei alle dem noch dominant und streitsüchtig. Jedenfalls musste K über die 25-tägige Hochzeitsreise, die die beiden in 6 Hotels führte, genauestens und unterwürfig Buch führen:

Gütekategorie d. Hotels	Aufenthalts- dauer (Tage)
4 (****)	2
3 (***)	5
2 (**)	8
1 (*)	7
0 (kein Stern)	3



a) Zeichnen Sie die (relative) Häufigkeitsverteilung und Summenhäufigkeitskurve der fünf Güteklassen!

b) Was ist im Falle von a)

- * die Zählereinheit (Erhebungseinheit)
- * die Häufigkeit
- * das Merkmal dieser Tabelle?

Zu bedenken sind Begriffe wie Güteklassen, Hotels, Sterne, Tage usw.

Aufgabe 3.2

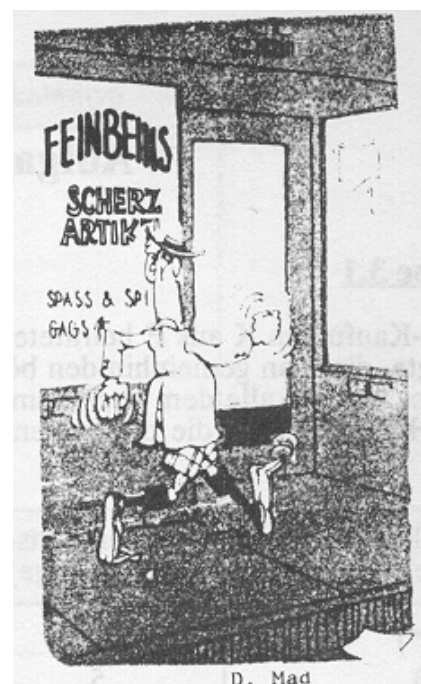
Das Bruttoinlandsprodukt des Euro-Währungsgebiets gliederte sich 1997 wie folgt (in Mrd. ECU, Angaben des EZB-Monatsberichts Jan. 99): Privater Verbrauch 2.854, Staatsverbrauch 722, Investitionen 930, Saldo Ausfuhr abzügl. Einfuhr 146. Zeichnen Sie ein Kreisdiagramm.

Aufgabe 3.3

In den letzten 30 Tagen betrat jeweils pro Tag nur die folgende leider zu geringe Anzahl von Kunden das bedeutende Fachgeschäft F:

2, 4, 6, 4, 7, 7, 5, 4, 3, 5,
5, 8, 6, 3, 5, 2, 9, 4, 5, 6,
8, 3, 10, 5, 4, 3, 7, 4, 6, 4

- a) Welches Skalenniveau hat die hier betrachtete Variable?
- b) Geben Sie die absoluten, relativen und prozentualen Häufigkeiten an!
- c) Erstellen Sie eine geordnete statistische Reihe!



- d) Zeichnen Sie das Stabdiagramm!
- e) Geben Sie die kumulierten absoluten, relativen und prozentualen Häufigkeiten an!
- f) Bestimmen Sie die klassierte Verteilung mit den Klassen 0 bis 4 Kunden, über 4 bis 7 Kunden und über 7 bis 10 Kunden und verfahren Sie entsprechend wie in Teil b, d und e!

Aufgabe 3.4

Bekanntlich hat das Ruhrgebiet ideale Bedingungen für die Ausbreitung des Vampirismus (geringe Entfernungen, nötige Infrastruktur, Ballung der Bevölkerung). Aus einer Statistik der durchschnittlichen (nächtlichen) Anfahrtzeit eines Vampirs zum nächsten Opfer kann man die folgenden Zahlen entnehmen:

Fahrzeit (Minuten)	Anzahl der Opfer
0 bis unter 5	100
5 bis unter 10	400
10 bis unter 20	300
20 bis unter 60	150
über eine Stunde	50

Stellen Sie diese Daten mit Angabe der relativen Häufigkeiten geeignet graphisch dar! Bestimmen Sie die Summenhäufigkeitskurve und zeichnen Sie diese!

Aufgabe 3.5

Die Zahl der in einem Einzelhandelsgeschäft zwischen 9 und 9.30 Uhr eintreffenden Kunden wurde an 20 Tagen ($t = 1, \dots, 20$) registriert. Man erhielt folgende Einzelwerte für die Anzahl x der Kunden (darunter y Frauen):

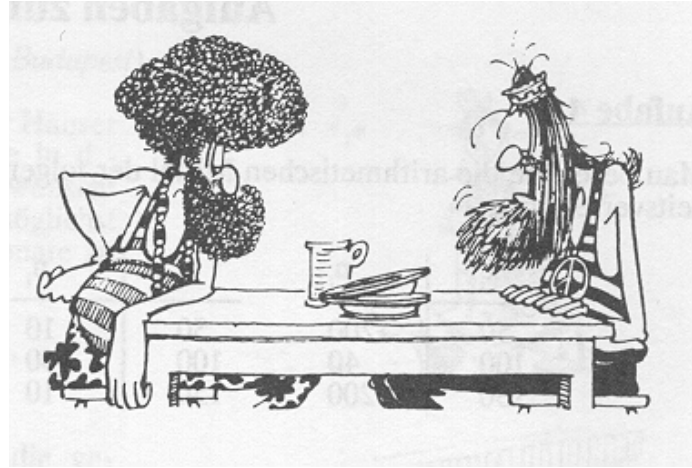
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x	0	1	1	2	4	1	0	1	3	3	1	0	2	3	2	1	1	2	0	2
y	0	0	0	1	1	1	0	0	1	2	0	0	1	1	0	0	0	1	0	1

- a) Stellen Sie die Daten als Zeitreihe graphisch dar.
- b) Gruppieren Sie die Daten und ermitteln Sie die absoluten, relativen Häufigkeiten, sowie die kumulierten absoluten und relativen Häufigkeiten von x .

Aufgabe 3.6

In einem Studentenheim gäbe es ausschließlich Wohngemeinschaften für $x_i = 2, 3, 4$ und 6 Studenten mit den Häufigkeiten n_i :

x_i	n_i
2	40
3	30
4	25
6	5



Man stelle die Häufigkeitsverteilung graphisch dar und bestimme später (nach Erreichen von Kap. 5) Mittelwert und Varianz dieser Verteilung sowie deren Schiefe

Aufgabe 3.7

Bei der Fußballweltmeisterschaft 1986 in Mexiko wurden von den 24 teilnehmenden Mannschaften in der Vorrunde insgesamt 83 Tore geschossen. Jede Mannschaft absolvierte drei Spiele, wofür sich folgende Verteilung der Tore erstellen lässt (von...bis unter...):

Anzahl der Tore	0 - 2	2 - 4	4 - 6	6 - 10
absolute Häufigkeit	5	8	8	3

a) Das betrachtete Merkmal ist

☐	diskret	☐	stetig	☐	extensiv	☐	häufbar
--------------------------	---------	--------------------------	--------	--------------------------	----------	--------------------------	---------

b) Stellen Sie die Häufigkeitsverteilung geeignet graphisch dar!

c) Wieviel Prozent der teilnehmenden Mannschaften haben sieben oder weniger, wie viel mehr als vier Tore geschossen?

d) Angenommen, Sie kennen die genaue Verteilung der Anzahl der Tore:

Anzahl der Tore	0	1	2	3	4	5	6	7	8	9
absolute Häufigkeit	1	4	5	3	4	4	1	0	0	2

Welche Veränderungen würden sich gegenüber der ursprünglichen Aufgabenstellung hinsichtlich

- der empirischen Verteilungsfunktion
- des arithmetischen Mittels
- des Medians ergeben?

Aufgaben zum Kapitel 4

Aufgabe 4.1

Man berechne die arithmetischen Mittel der folgenden Verteilungen und zeichne die Häufigkeitsverteilungen!

x_i	n_i
50	200
100	40
150	200

x_i	n_i
50	10
100	60
150	10

x_i	n_i
70	60
80	80
440	10

Aufgabe 4.2

Man berechne das arithmetische Mittel für

- a) die Zahlenfolge 1, 3, 5, 6, 9, 9, 12, 13, 15, 17 und für die klassierte Verteilung (von...bis unter)

x_j	n_j
1 - 5	3
6 - 12	4
13 - 17	3

- b) für die Zahlenfolge 1, 6, 6, 11, 1, 1, 6, 11, 6, 11 und die unklassierten Verteilung (gruppierte Daten)

x_i	n_i
1	3
6	4
11	3

Aufgabe 4.3

Sie kaufen an einem Stand zu 1 DM Apfelsinen (3 Stück zu 1 DM) und an einem anderen Stand ebenfalls für 1 DM, wobei Sie jedoch 5 Stück erhalten. Wie groß ist dann der Durchschnittspreis der Apfelsinen?

Aufgabe 4.4

Die 200 Beschäftigten einer Arbeitsstätte erhalten einen monatlichen Durchschnittslohn von 2.200 DM. Aufgrund einer Lohnverhandlung soll das Monatsgehalt jedes Beschäftigten um 10% angehoben werden, und es soll ein einmaliges Urlaubsgeld in Höhe von 120 DM gewährt werden. Wie groß ist das durchschnittliche Gehalt?

Aufgabe 4.5

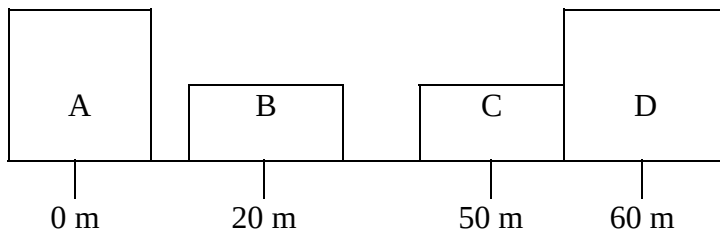
(Der optimale Standort des Stehgeigers von Budapest)

László Varga möchte den Bewohnern der Häuser A bis D der Bartók-Straße seine Sonate in d-moll op. 125 zu Gehör bringen. Dabei wünscht der Tonkünstler, dass alle 25 Familien möglichst gleich gut die Gelegenheit haben, die Sonate zu hören und zu würdigen.

An welcher Stelle sollte sich Varga stellen,

- um im "Schwerpunkt" zu stehen
- um von den potentiellen Hörern die geringst mögliche Entfernung zu haben

wenn die Straße wie folgt aussieht:



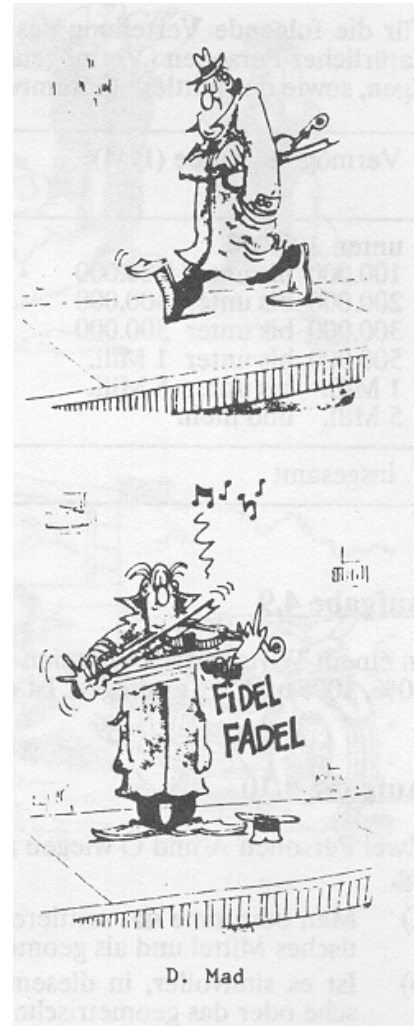
In den Häusern A und D wohnen jeweils 7 Familien, im Haus B 6 und im Haus C genau 5 Familien. Man erkläre auch anhand dieser Aufgabe, was mit der "Schwerpunkteigenschaft" des arithmetischen Mittels gemeint ist!

Aufgabe 4.6

Dem geisteskranken Diplom-Kaufmann K aus E sind von seinem früheren Studium nur noch Kenntnisse aus den oft als geisttötend empfundenen Fächern Buchhaltung und Statistik verblieben, wobei sich jedoch K häufig verrechnet. Er verbucht jeden Tag die Uhrzeitangaben im Radio in einem Staffellokonto, zählt die Zeitangaben zusammen und berechnet den Mittelwert. Danach zieht K von jeder Zeitangabe den Mittelwert ab und addiert die Ergebnisse. Wie kann man leicht feststellen, ob K sich verrechnet hat?

Aufgabe 4.7

Für die folgende Verteilung des Gesamtvermögens unbeschränkt vermögenssteuerpflichtiger natürlicher Personen (Vermögenssteuerstatistik 1983) ist eine graphische Darstellung anzufertigen, sowie das mittlere Gesamtvermögen zu berechnen.)



Vermögensgruppe (DM)	Steuer- pflichtige	Gesamtvermögen (Mill. DM)
unter 100.000	24.725	2.156
100.000 bis unter 200.000	136.557	20.919
200.000 bis unter 300.000	134.444	33.334
300.000 bis unter 500.000	163.020	63.092
500.000 bis unter 1 Mill.	126.625	87.010
1 Mill. bis unter 5 Mill.	72.576	134.846
5 Mill. und mehr	9.312	138.080
insgesamt	667.259	479.437

Aufgabe 4.8

In einem Wirtschaftszweig seien die Löhne und Gehälter in den letzten vier Jahren um 10%, 20%, 10% und 5% gestiegen. Ist es sinnvoll, hier das geometrische Mittel zu berechnen?

Aufgabe 4.9

Zwei Personen A und O wiegen $x_A = 35$ kg und $x_O = 140$ kg.

- Man bestimme das mittlere Gewicht als arithmetisches Mittel und als geometrisches Mittel.
- Ist es sinnvoller, in diesem Falle das arithmetische oder das geometrische Mittel zu berechnen? (Deuten Sie anschaulich die Rechenergebnisse!)



- Warum ist das geometrische Mittel kleiner als das arithmetische Mittel? Wann wären beide Mittelwerte gleich groß?

Aufgabe 4.10

Auf einer Rennstrecke von 12 km Länge erreicht der Rennfahrer R verschiedene Durchschnittsgeschwindigkeiten und zwar in den drei Runden

erste Runde	240 km/h
zweite Runde	200 km/h
dritte Runde	160 km/h

Die Gesamtfahrzeit für drei Runden (36 km) beträgt 11,1 Min (also 0,185 Std.). Wie groß ist die Durchschnittsgeschwindigkeit?

Aufgabe 4.11

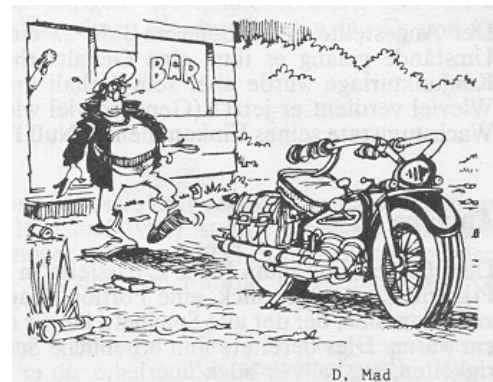
Der statistisch wenig vorgebildete Bauunternehmer B beschäftigt zwei Betonarbeiter, die jeweils bestimmte Flächen F auszubetonieren haben. Der eine braucht für eine Fläche 3 Stunden, der andere 5 Stunden. B glaubt nun, dass sie zusammen im Durchschnitt jeweils 4 Stunden an einer Fläche F betonieren werden. Er irrt sich! Warum?



Aufgabe 4.12

Der Rocker R kam leider nie in den Genuss, eine Statistikvorlesung zu hören. Es gelingt ihm deshalb nicht das folgende Problem zu lösen:

R möchte auf der Hin- und Rückfahrt zu seiner 4 km entfernten Stammkneipe eine Durchschnittsgeschwindigkeit von 60 km/h fahren. Dabei traut er sich auf dem Rückweg nur eine Geschwindigkeit von 30 km/h zu. Muß er deshalb auf dem Hinweg 90 km/h fahren?



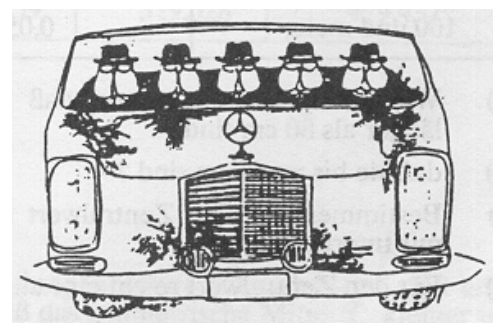
Aufgabe 4.13

Berechnen Sie den Zentralwert

- a) im Falle der Reihe ($n=7$) 5, 8, 10, 12, 13, 16, 20
- b) im Falle der Reihe ($n=8$) 5, 8, 10, 12, 13, 16, 20, 21

Aufgabe 4.14

Im fünfköpfigen Vorstand der X - AG sitzen verdiente Männer im Alter von 48, 53, 53, 55 und 62 Jahren. Man plant eine Geschäftsreise nach Bangkok. Das älteste Vorstandsmitglied kann jedoch leider nicht mitreisen, weil ihm sein Arzt wegen seines hohen Blutdrucks eindringlich von der weiten Reise abgeraten hat. Außerdem war seine Frau aufgrund ihrer in der Illustrierten gewonnenen Erkenntnisse dagegen. An seiner Stelle kann nun ein junger dynamischer Prokurist im Alter von 35 Jahren mitreisen.



Wie ändert sich der Zentralwert und das arithmetische Mittel der Altersverteilung der reisenden Geschäftsleute?

Aufgabe 4.15

Man berechne das arithmetische Mittel und den Zentralwert (mit Interpolation) der normalerweise geleisteten Wochenarbeitszeit in der BRD 1996 (Erwerbstätige in 1000)¹!

Stunden	Männer	Frauen	gesamt
unter 21	283	2051	2334
21 bis unter 36	1556	2644	4200
36 bis unter 41	14486	8010	22496
41 und mehr	4036	1171	5207

¹ Quelle: StJB 1997

Geben Sie auch das 1. und 3. Quartil an! Bei der Berechnung des arithmetischen Mittels sind die Klassengrenzen mit 15 und 50 Stunden zu schließen.

Aufgabe 4.16

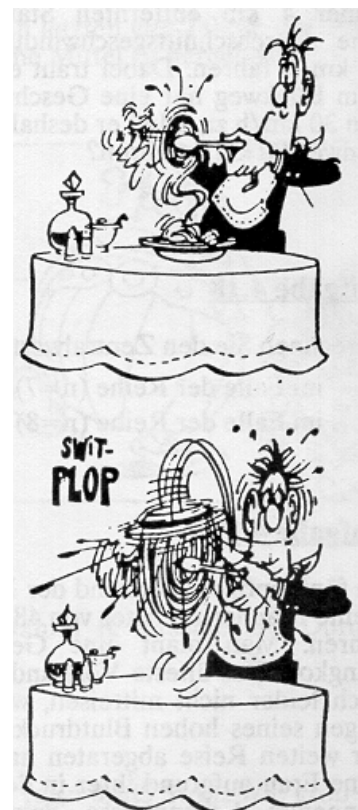
Der Angestellte A verdiene im Jahr ein Monatsgehalt von 1200 DM. Durch glückliche Umstände gelang es ihm, eine Gehaltserhöhung von 20% zu erhalten. Wegen schlechter Konjunkturlage wurde aber sein Gehalt im darauffolgenden Jahr wieder um 20% gekürzt. Wieviel verdient er jetzt? Wie groß ist die durchschnittliche Wachstumsrate seines Einkommens? (Null Prozent?)

Aufgabe 4.17

Dem Diplomkaufmann K aus E passierte in einer Pizzeria das Missgeschick, eine Portion Spaghetti zu bekommen, bei der alle Spaghetti länger als 80 cm waren. Dies bereitete ihm erhebliche Schwierigkeiten, weshalb er auch überlegte, ob er nicht doch besser Ravioli hätte essen sollen. Für die Häufigkeitsverteilung der Spaghettilänge in cm gilt:

Klasse x	h
unter 20	0,15
$20 < x \leq 40$	0,20
$40 < x \leq 60$	0,30
$60 < x \leq 80$	0,20
$80 < x \leq 100$	0,10
100 und mehr	0,05

- Wie häufig kommt es vor, dass Spaghetti länger als 80 cm sind;
- dass sie bis zu 60 cm sind?
- Bestimmen Sie den Zentralwert (Median) mit Interpolation.
- Für den Zentralwert reicht eine.....Skala der Daten aus.



Aufgabe 4.18

Gegeben sei folgende klassierte Verteilung (b_i = Klassenbreite)

x_i	n_i	h_i	H_i	b_i
0 - 20	10			
20 - 30	30			
30 - 60	20			
60 - 100	40			
Σ				

der stündlich vertelefontierten Einheiten eines eifrigen Wertpapierberaters, der seine "Telefonitis" 100 Stunden lang aufzeichnen ließ. Berechnen Sie die Quartile Q_1 , Q_2 und Q_3 anhand der Interpolations - Formel.

Aufgabe 4.19

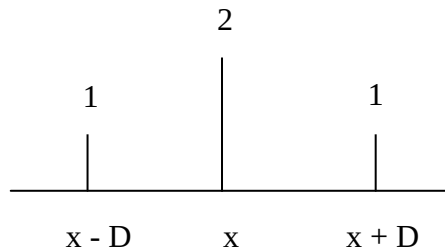
Für die folgende Verteilung des Gesamtvermögens unbeschränkt vermögensteuerpflichtiger natürlicher Personen (Vermögensteuerstatistik 1977 und 1983) ist eine graphische Darstellung anzufertigen, sowie das mittlere Gesamtvermögen 1977 und 1983 zu berechnen (mit den geschätzten Klassenmittelwerten [Klassenmitten m_k] und den wahren Klassenmittelwerten x_k).

Vermögensklasse von - bis unter -	Steuerpflichtige		Gesamtvermögen ¹⁾	
	1977	1983	1977	1983
< 100.000 DM	26.057	24.725	2.247	2.156
100 - 200.000	114.327	136.557	17.435	20.919
200 - 300.000	102.829	134.444	25.435	33.334
300 - 500.000	118.208	163.020	45.592	63.092
1/2 - 1 Mill.	87.145	126.625	59.841	87.010
1 - 5 Mill.	49.741	72.576	92.973	134.846
> 5 Mill. DM	5.753	9.312	76.105	138.080
insgesamt	504.060	667.259	319.628	479.437

¹⁾ in Mill. DM; Quelle: Stat. Jahrbuch 1982, S. 440 und 1989, S. 451

Aufgabe 4.20

- a) Gegeben sei die untenstehende Häufigkeitsverteilung (absolute Häufigkeiten 1, 2, 1 und die Merkmalswerte $x-D$, x , $x+D$). Zeigen Sie, daß das geometrische Mittel $\bar{x}_G$ kleiner ist als $\bar{x}$, also $\bar{x}_G < \bar{x}$!



- b) Berechnen Sie die Varianz für obige Häufigkeitsverteilung!
- c) Ein Flugzeug fliege von A nach B bei Rückenwind mit einer Geschwindigkeit von $90+D$ Knoten (wegen des Rückenwindes ist es schneller als 90 Knoten, $D>0$). Auf dem Rückflug von B nach A hat es gegen den Wind zu fliegen, so dass es nur eine Geschwindigkeit von $90-D$ Knoten erreichen konnte (langsamer als 90 Knoten). Auf der Gesamtstrecke (hin und zurück) erreichte es eine Durchschnittsgeschwindigkeit von 80 Knoten. Wie groß war die Geschwindigkeit des Gegenwinds (Windkomponente)?

Aufgabe 4.21

iplom-Kaufmann K aus E hat ein etwas ungewöhnliches Zahnleiden, weshalb er schon oft von den übrigen Ökonomen (die ja meist wenig feinführend sind) seines Betriebes geärgert wurde. Man übertrug ihm wegen der Arbeit mit dem Locher die Registratur. Aber auch bei dem ihm unterstellten Damen des Betriebs fiel es ihm schwer eine Autorität aufzubauen.



Er nahm sich deshalb vor, mit statistischen Mitteln hart durchzugreifen:

Seine Mitarbeiterinnen A bis D brauchten sehr unterschiedlich lange um ein und dieselbe Arbeit durchzuführen:

Zeitbedarf (in Minuten): 15, 10, 20, 15

- a) Welche Mittelwerte könnte man für die vier Zeitangaben berechnen, wenn es gilt, die in der Zeitangabe (Skala) enthaltene Information voll auszunutzen?
- b) Welche Größenbeziehungen gelten unter den von Ihnen vorgeschlagenen Mittelwerten?

c) Vorgriff auf Kap. 5:

Zeigen Sie, dass die durchschnittliche absolut genommene Abweichung der Zeiten der vier Damen untereinander 5 Minuten beträgt! Wie heißt die hiermit errechnete Maßzahl?

Aufgabe 4.22

Beim Bau eines Flugplatzes spielt die "Flughafenbezugstemperatur" (aerodrome reference temperature ART) eine gewisse Rolle. Sie ist im Gesetz definiert als:

"das monatliche Mittel der täglichen mittleren Temperatur - über 24 Stunden bei halbstündiger Messung - des heißesten Monat des Jahres plus ein Drittel des Unterschieds dieser Temperatur zu dem Mittel der täglichen Maximaltemperatur des gleichen Monats"

Interpretieren Sie diesen Gesetzestext und versuchen Sie den Sachverhalt in einer Formel auszudrücken!

Aufgabe 4.23

Das folgende Beispiel zeigt, dass man evtl. durch Unkenntnis einfachster statistischer Begriffe viel schwächer argumentiert, als es an sich möglich wäre:

"1835 Mark im Monat aber war das durchschnittliche Nettoeinkommen der 22,8 Millionen Arbeitnehmer im vergangenen Jahr; das heißt, daß allein elf Millionen Arbeitnehmer wesentlich weniger verdient haben als 1800 Mark"

Herbert Ehrenberg (früherer Arbeitsminister): Unaufhaltsamer Marsch in die Depression, in: DER SPIEGEL, 27.12.1983

a) Nehmen Sie Stellung zu dieser Schlussweise!

b) Angenommen (was der damaligen Realität etwa entsprach) $x = 1800$ und $s_x = 1600$ und die Schiefe betrage $SK_{p_2} = 15/16 = 0,9375$ (Schiefemaß Gl. 5.64), dann verdienen 50 % der Arbeitnehmer weniger als DM.

Aufgaben zum Kapitel 5

Aufgabe 5.1

Elf Moskauer Frauen seien Ende des vorigen Jahrhunderts nach der Zahl ihrer Kinder befragt worden. Dabei ergab sich die folgende Reihe: 0, 9, 3, 2, 0, 1, 1, 1, 4, 6, 6. Wie ändert sich die Streuung, gemessen anhand

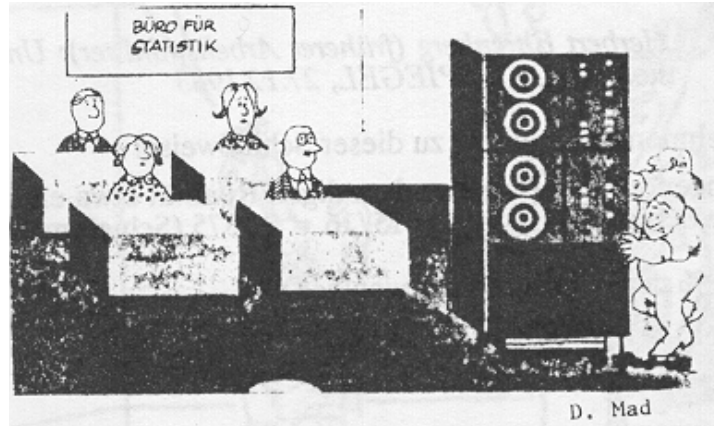
- a) des mittleren Quartilsabstands
- b) der Spannweite
- c) der Varianz und Standardabweichung,

wenn anstelle der oben als elfte Frau genannten Frau (mit 6 Kindern) Feodora Wassilet befragt worden wäre, die seinerzeit 69 Kinder zur Welt brachte und schon mit 56 Jahren starb?

Aufgabe 5.2

In einer statistischen Behörde gilt die Anzahl der pro Tag zu den Akten gegangenen veröffentlichungsreifen Tabellen als Produktivitätsmaßstab. Man will nun die ersten 100 Tage des Einsatzes eines Computers mit 440 früheren Tagen vergleichen. Es ergaben sich folgende Verteilungen:

x_i	früher	jetzt
1	60	5
2	160	10
3	110	25
4	0	20
5	60	0
6	50	0
8	0	40

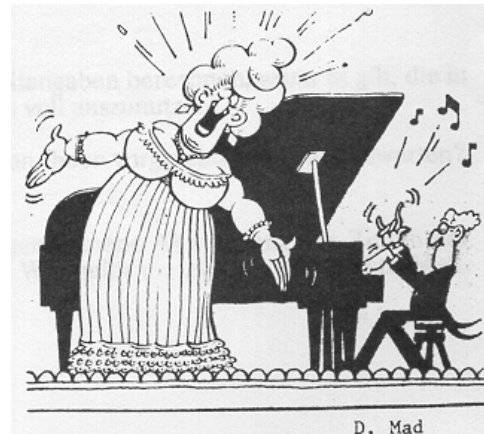


Es stellte sich heraus, dass sich die Produktivität durch den Computereinsatz etwas erhöhte. Wie steht es aber mit der Streuung? Man berechne zu diesem Zweck die durchschnittliche Abweichung um den Zentralwert (d_Z) und um das arithmetische Mittel (d_x^*)!

Aufgabe 5.3

Der Pianist P und die Sopranistin S besuchten auf ihrer Tournee die Zentren deutschen Kulturlebens. Ihre Konzerte waren jedoch sehr unterschiedlich besucht, nämlich in:

München	70 Besucher
Marburg	60 Besucher
Bottrop	20 Besucher
Essen	10 Besucher.



Man berechne Ginis Dispersionskoeffizient (Streuungsmaß)!

Aufgabe 5.4

- Ist es möglich, dass d_x^* und d_Z ungefähr gleich sind, obgleich $\bar{x}$ und $Z = \tilde{x}_{0,5}$ sehr unterschiedlich sind? Wenn ja, warum gilt dann stets $d_x^* \leq d_x$?
- Kann es sein, dass man mit d_x zwischen unterschiedlichen Graden der Streuung weniger differenzieren kann als mit d_x^* ?

Anmerkung zu b)

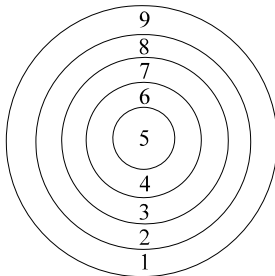
Man betrachte hierzu das folgende Beispiel von Notenverteilungen von vier Schülern (jeweils drei Klausuren):

Schüler	Noten
A	3, 4, 5
B	2, 4, 5
C	1, 4, 4
D	1, 4, 5

Anmerkung: Es soll hier, wie bei Aufg. 5.1 davon abgesehen werden, dass Schulnoten (Zensuren) eigentlich nicht intervall-, sondern nur ordinalskaliert sind.

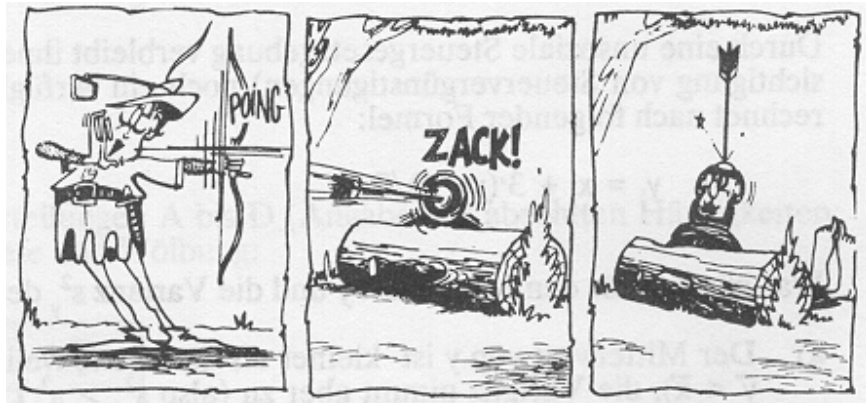
Aufgabe 5.5

Der Bogenschütze B schießt auf eine Zielscheibe, die wie folgt aufgebaut ist:



Er erzielte mit neun Schüssen die folgenden Ergebnisse:

2, 3, 3, 4, 5, 6, 7, 7, 8



Ein anderer Schütze C erzielte mit acht Schüssen folgende Ergebnisse:

3, 4, 4, 5, 5, 5, 6, 8

Wessen Treffsicherheit ist beständiger? Beantworten Sie diese Frage durch Berechnung der Standardabweichungen!

Aufgabe 5.6

a) Berechnen Sie die Varianz der Zahlenfolge 1, 1, 3, 5, 6, 7, 8, 10.5, 11, 11.5 !

b) Berechnen Sie die Varianz der (klassierten) Verteilung mit der Formel $s^2 = \sum h_k (\bar{x}_k - \bar{x})^2$

Größenklasse k	$\bar{x}_k$	n_i
0 bis unter 6	2,5	4
6 bis unter 10	7	3
10 und mehr	11	3

c) In den Teilen a) und b) handelt es sich um die gleichen Daten! Warum weichen die Er-

gebnisse voneinander ab? (Hinweis: Man beachte Formel 5.11!) Wie groß ist die interne Varianz in diesem Beispiel?

Aufgabe 5.7

"Denn wer da hat, dem wird gegeben, dass er die Fülle habe, wer aber nicht hat, von dem wird genommen, das er hat" (Matth. 13, Vers 12)

Fünf Personen haben die Einkommen x vor Steuerabzug

300 400 600 700 1000

Durch eine unsoziale Steuergesetzgebung verbleibt ihnen nach Steuerabzug (bzw. bei Berücksichtigung von Steuervergünstigungen) noch ein verfügbares Einkommen in Höhe von y , berechnet nach folgender Formel:

$$y_i = x_i + 3(x_i - \bar{x}) / 7$$

Was gilt nun für den Mittelwert $\bar{y}$ und die Varianz s_y^2 der Größe y ?

- Der Mittelwert von y ist kleiner als der von x , weil ja eine Steuer abgezogen wird (also $\bar{y} < \bar{x}$), die Varianz nimmt aber zu (also $s_y^2 > s_x^2$), weil die Steuergesetzgebung unsozial ist.
- Es gilt $\bar{y} = \bar{x} = 600$, und die Varianz wird aber doppelt so groß, genau $s_y^2 = \left(\frac{10}{7}\right)^2 \cdot s_x^2$.
- Die Varianz von x ist positiv und die von y ist negativ. Die Mittelwerte sind aber eigenartigerweise gleich und zwar $\bar{y} = \bar{x} = 600$.
- Alle Antworten a) bis c) sind falsch.

Aufgabe 5.8

Um seine Belastung mit Operationen statistisch zu durchleuchten, stellt der berühmte Gehirnochirurg G die Verteilung der Anzahl x der Operationen pro Woche fest:

Klasse	$\bar{x}_k$	n_k	s_k
unter 6	2,5	4	11/4
6 bis 10	7	3	2/3
über 10	11	3	2/3

Man berechne die Gesamtvarianz unter Berücksichtigung der internen Varianz!



Aufgabe 5.9

In einem bestimmten Wirtschaftszweig mit 500.000 Beschäftigten betrage der Durchschnittslohn 880,- DM bei einer Standardabweichung von 200,- DM. Zwischen den Tarifpartnern wird eine Lohnerhöhung um 20% sowie die Zahlung eines einmaligen (jährlichen) Urlaubs-

geldes in Höhe von 600,- DM vereinbart. Berechnen Sie

- den neuen durchschnittlichen Monatslohn ohne Urlaubsgeld sowie die entsprechende Standardabweichung!
- den neuen durchschnittlichen Monatslohn einschließlich des monatlichen Anteils am Urlaubsgeld sowie die entsprechende Standardabweichung!

Aufgabe 5.10

Man beurteile die folgenden vier Verteilungen A bis D (Angabe der absoluten Häufigkeiten in der Tabelle) hinsichtlich ihrer Schiefe und Wölbung:

x_i	A	B	C	D
60	0	0	4	4
65	4	12	8	4
70	40	24	20	20
75	24	48	24	44
80	20	24	40	20
85	8	12	4	4
90	4	0	0	4

Aufgabe 5.11

Gegeben sei folgende Verteilung:

x_i	1	2	3
h_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Man bestimme das erste, zweite und dritte

- Anfangsmoment
- zentrale Moment!

Aufgabe 5.12

Der folgende Ausdruck $S_H = \left(\sum (x_v - \bar{x}_H)^{-1} / n \right)^{-1} = n / \sum (x_v - \bar{x}_H)^{-1}$ soll als Streuungsmaß vorgeschlagen werden:

- Nach welchem Prinzip ist dieses Streuungsmaß konstruiert (Mittelwert von Abweichungen?)?
- Prüfen Sie ob S_H die Axiome für Streuungsmaße erfüllt und
- kreuzen Sie in der folgenden Tabelle "ja" (J) oder "nein" (N) an:

Das Streuungsmaß S_H ist	J	N
1. untauglich, weil S_H immer negativ ist und eine Streuung nicht negativ sein kann		
2. untauglich, weil S_H wegen der Schwerpunkteigenschaft immer Null ist		
3. untauglich, weil S_H nicht zu berechnen ist, wenn die Merkmalswerte x_i negativ sind		
4. tauglich, weil S_H alle die Axiome für Streuungsmaße erfüllt		
5. tauglich, weil S_H fast alle die Axiome für Streuungsmaße erfüllt		
6. tauglich, weil S_H das harmonische Mittel der Abweichungen vom harmonischen Mittel ist.		

Aufgabe 5.13:

- a) Warum taugt das Kriterium $\sum(x_i - \bar{x})$ nicht als Streuungsmaß?
- b) Bei welcher Datenkonstellation ergibt sich stets $s^2 = \sum(x_i - \bar{x})^2 = 0$?
- c) Welche der nachfolgenden Aussagen sind richtig? (Richtige Antwort ankreuzen:)

Die Varianz beschreibt

☐	die Summe der Abweichungen der Einzelbeobachtungswerte untereinander
☐	das arithmetische Mittel der Abweichungsquadrate der Einzelwerte vom Gesamtmittelwert
☐	die Streuung der Einzelwerte um das Zentrum der Verteilung
☐	die durchschnittliche Differenz zwischen der geringsten und der maximalen Abweichung der Einzelwerte vom Mittelwert

Aufgabe 5.14

In einem indianischen Dorf gibt es vier Arten von Zelten: 2-, 4-, 6- und 12 Personenzelte. In jedem Zeltyp leben insgesamt 120 Personen. Die Verteilung ist also wie folgt:

Zeltyp	Häufigkeit	Personen
2		120
4		120
6		120
10		120
Summe	120	480



Man bestimme das arithmetische Mittel, die Varianz und die Schiefe der Anzahl X der Personen je Zelt (der Häufigkeitsverteilung von X). Die Verteilung ist (Richtiges ankreuzen):

	symmetrisch
	zweigipflig
	zweidimensional
	asymmetrisch

	linkssteil
	stetig
	rechtsschief
	unkonzentriert

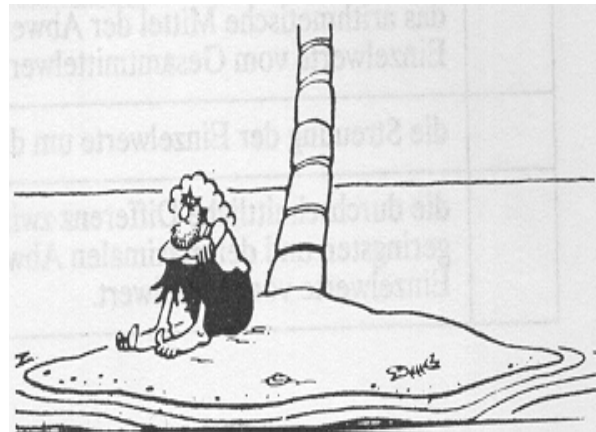
	rechtssteil
	diskret

Aufgabe 5.15

Der nordsibirische Großfürst Fjodor gebietet über ein Reich, welches aus 200 Inseln besteht, wovon jedoch nicht weniger als 120 unbewohnt sind. Außerdem liegt sein Reich an einem gefährlichen arktischen Seeweg.

Deshalb gehören auch 40 Schiffbrüchige zu seinen 10200 Untertanen. Die Verteilung der Bevölkerung auf die Inseln hat mithin folgende Gestalt:

Nr.	Größenklasse Einwohner	Anzahl der Inseln
1	0	120
2	1	40
3	2 bis 100	20
4	über 100	20
		200



- Man bestimme die mittlere Einwohnerzahl je Insel im Fürstentum (arithmetisches Mittel) und den Zentralwert sowie den mittleren Quartilsabstand.
- Wie groß ist die mittlere Einwohnerzahl der 20 größten Inseln mit jeweils über 100 Einwohnern (wenn in der Klasse 3 im Mittel 51 Einwohner auf einer Insel wohnen)?

Aufgabe 5.16

Gegeben sei die folgende klassierte Verteilung:

Größenklasse	Einzelwerte
1 bis unter 4	1, 2, 3
4 bis unter 10	6, 6, 7, 9
10 bis unter 14	13, 10, 13

Berechnen Sie

- Mittelwerte und Varianzen der einzelnen Größenklassen!
 - Gesamtmittelwert und Gesamtvarianz!
 - Die Varianz der Mittelwerte der drei Größenklassen!
- Welcher Zusammenhang besteht zwischen den Ergebnissen?

Aufgabe 5.17

Für die Lohnempfänger eines Unternehmens mit zwei Betrieben liegen folgende Angaben zur Einkommensverteilung vor:

Betrieb	Anzahl der Lohnempfänger	x_i	s_i^2
1	2500	1200	160000
2	7500	1600	200000

Berechnen Sie $\bar{x}$ und s^2 der Einkommensverteilung für das gesamte Unternehmen!

Aufgabe 5.18

In einem Unternehmen fallen beim Absatz von 5 ausgewählten Erzeugnissen unterschiedliche Verpackungskosten an.

Erzeugnis	Verpackungskosten pro Stück	relativer Anteil der abgesetzten Menge
A	4,15 DM/Stück	10%
B	3,25 DM/Stück	17%
C	2,95 DM/Stück	39%
D	2,85 DM/Stück	21%
E	3,50 DM/Stück	13%

- Berechnen Sie die durchschnittlich anfallenden Verpackungskosten pro Stück dieser 5 Erzeugnisse insgesamt.
- Geben Sie Median und Modus an!
- Berechnen Sie Standardabweichung und Variationskoeffizient!
- In welchem Bereich müsste aufgrund der Fechnerschen Lageregel der Momentenkoeffizient der Schiefe liegen?

Aufgabe 5.19

Eine Computerauswertung zur Dickenmessung (in mm) von Drähten ergibt für zwei zu vergleichende Prüfproben u.a. folgende statistische Maßzahlen:

Prüfprobe I		Prüfprobe II	
Mittelwert	11,6	Mittelwert	10,8
Varianz	23,1	Standardabweichung	0,9
Schiefe	0	Modus	9,5
Wölbung	-0,78	Schiefe	+0,95

- Interpretieren Sie die Maßzahlen sachbezogen (Angabe der Einheiten)!
- Vergleichen Sie beide Verteilungen und skizzieren Sie diese grob!
- Welchen Wert haben Median und Modus der Prüfprobe I?

Aufgabe 5.20

Diplom Kaufmann K aus E buchte eine besonders preisgünstige Mittelmeerkreuzfahrt. Widrige Umstände und erlittene Unbill ließen in ihm jedoch den Entschluss reifen, solche Reisen hinfert nicht mehr zu unternehmen: nach drei Tagen erreichte er als einer der wenigen Überlebenden eine Kaimauer des Hafens von Genua:



Im anschließenden Schadensersatzprozess legte das Reisebüro ein Umfrageergebnis vor, wonach die Befragten vor und nach der heutzutage so beliebten "Abenteuerreisen" eine sehr unterschiedliche Neigung zu solchen Veranstaltungen zu erkennen gaben, und zwar anhand einer 7-Punkte Skala mit 5-er Schritten von 0 (wenig beliebt) bis 30 (sehr beliebt):

vorher		nachher	
x_i	n_i	x_i	n_i
0	0	0	2
5	4	5	4
10	40	10	10
15	24	15	12
20	20	20	20
25	8	25	2
30	4	30	0

Kann man der Statistik entnehmen:

- ein (verglichen mit vorher) höheres Niveau der Beliebtheit nach einer solchen Reise (Mittelwert!)?
- eine geringe Aussagefähigkeit, etwa wegen einer nach Teilnahme an einer solchen Reise zunehmenden Streuung?
- eine Veränderung der Meinungen aufgrund unterschiedlicher Schiefe?

Aufgaben zum Kapitel 6

Aufgabe 6.1

150 Angehörige eines primitiven Volksstammes gehen auf die Jagd nach Federvieh. Ihre Beute beträgt 300 Wildgänse. Durch das an sich nur bei primitiven Völkern bekannte Gerangel um Geld, Gut und Prestige entstand eine etwas ungleiche Verteilung der Beute. Durch Eingreifen des Häuptlings konnte jedoch noch verhindert werden, dass einer leer ausging. Es bekamen jeweils n_i Personen x_i Gänse:

x_i	1	2	3	4
n_i	60	45	30	15



Man zeichne die Lorenzkurve. Wie sähe die Lorenzkurve aus, wenn jeder von der Beute gleichviel bekommen hätte?

Aufgabe 6.2

Drei Unternehmen A, B, C teilen sich den heiß umkämpften Markt für X zu gleichen Teilen. Durch die Fusion von A mit B steigt die relative Konzentration (Disparität) gemessen mit dem Gini-Koeffizienten D_G

von $D_G =$
auf $D_G = \dots\dots\dots!$

Vier Unternehmen A, B, C, D haben gleiche Marktanteile auf dem Y-Markt. Durch die Fusion von A mit B steigt D_G

von $D_G =$
auf $D_G = \dots\dots\dots!$

Aufgabe 6.3

Diplom-Kaufmann K aus E hat als Einzelhandelsunternehmer der Schreibwarenbranche zwar einen bescheidenen, aber doch von Null verschiedenen Anteil am Produktivvermögen. Als politisch engagierter Akademiker interessiert ihn jedoch die ungleiche Vermögensverteilung im ganzen, denn in ihr sieht er den Hauptgrund seines mangelnden unternehmerischen Erfolgs.

Aus einem Buch¹⁾ erfährt er folgende Daten: ($\bar{x}$ = Klassenmitte in DM, n = Anzahl der Personen in Mill.)

Vermögensgruppe	$\bar{x}$	n
nur Schulden	-7000	3
kein Vermögen	0	1
0 bis 16000	5000	26
16000 bis 100000	50000	28
über 100000 DM	435000	2

Zahlen vereinfacht nach Mierheim/Wicke, Die personelle Vermögensverteilung, Berlin 1979, S.60



- a) Hiernach haben 3 Mill. Personen kein positives Nettovermögen (= Forderungen - Schulden), sondern nur Schulden. Ihr Anteil am Vermögen ist folglich (richtige Antwort ankreuzen)

☐	negativ	☐	null	☐	positiv
--------------------------	---------	--------------------------	------	--------------------------	---------

und die Lorenzkurve, die sich aus obigen Daten berechnen lässt

☐	verläuft auch im negativen Bereich ($Q < 0$)
☐	verläuft nur im positiven Bereich ($Q \geq 0$)
☐	schneidet die Abszisse
☐	schneidet die Gleichverteilungsgerade.

- b) Man bestimme die Lorenzkurve und Ginis Disparitätsmaß für die obigen Daten!

Aufgabe 6.4

Im Lande A erhalten 70% der Bevölkerung 10% des Volksvermögens und im Lande B entfällt auf 80% der Bevölkerung genau 20% des Volksvermögens. In welchem Land ist die Vermögenskonzentration gemessen anhand des Gini-Koeffizienten D_G größer?

Aufgabe 6.5

Trotz neuer praxisrelevanter wissenschaftlicher Erkenntnisse konnte Prinzessin Rana von Esculenta (E) nicht umhin, an die Existenz des Froschkönigs zu glauben. Sie war jedoch gleichwohl rationalem Denken insofern aufgeschlossen, als sie bestrebt war, ihre Kuss-Aktivität zunächst auf solche Teiche zu konzentrieren, in denen eine größere Anzahl von Fröschen zu erwarten war. Sie beauftragte deshalb den Statistiker Pedro de las Tablas mit einer Schätzung des Froschbestands in Esculenta:

	Frösche im Teich	Anzahl der Teiche
1	0	120
2	1 bis 100	60
3	über 100	20

Pedro de las Tablas vermutet, dass sich in den Teichen der Größenklasse 2 und 3 jeweils insgesamt 3000 Frösche befinden, so dass das Fürstentum Esculenta insgesamt 6000 Frösche besitzt.

a) an bestimme (berechne und zeichne) die Lorenzkurve!

b) Berechnen Sie Ginis Disparitätsmaß D_G (Konzentrationsverhältnis von Gini)!



Aufgabe 6.6

In einem islamischen Dorf lebt eine Familie mit 8 Kindern, davon sind 4 männlich und 4 weiblich. Das an die Kinder zu vererbende Vermögen von 1200 Dinare soll getreu nach den Regeln des Korans vermacht werden:

"Und wenn die Geschwister Männer und Frauen sind, so soll ein Mann so viel erhalten wie zwei Frauen".

(Sure 4, Vers 175)

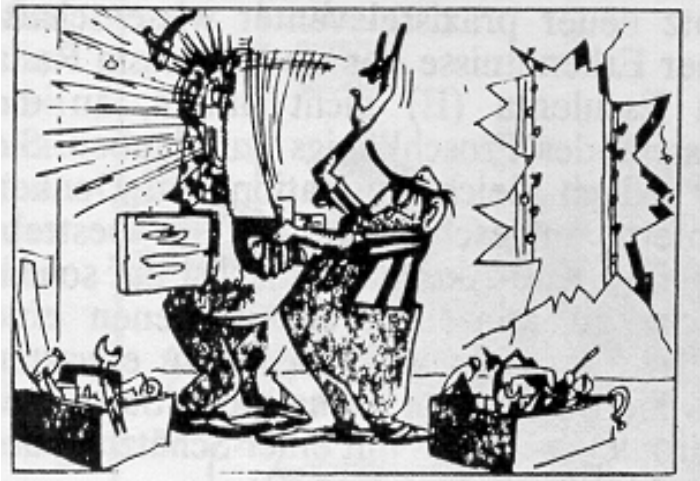
Man bestimme die sich bei Befolgung des islamischen Erbrechts ergebende Lorenzkurve!

Aufgabe 6.7

In einem Betrieb mit 20 Arbeitnehmern gibt es nur Zank und Krach, weil es drei Lohngruppen L_1 , L_2 und L_3 mit sehr unterschiedlichen Löhnen gibt: $L_1 < L_2 < L_3$. Der Durchschnittslohn beträgt 900. Von der Lorenzkurve seien zwei Punkte bekannt

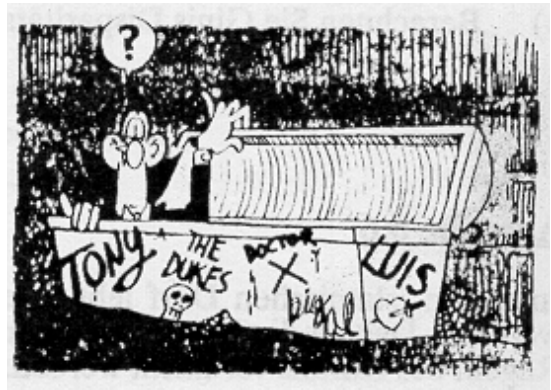
Punkt	H(x)	Q(x)
A	0,4	0,2
B	0,8	0,6

- a) Man bestimme die drei Löhne L_1 , L_2 , L_3 .
- b) Man bestimme den Variationskoeffizient V der Löhne.
- c) Man bestimme das Disparitätsmaß D_G von Gini für die obigen Daten (Punkte A und B).
- d) Angenommen vom Punkt B sei nur die Koordinate H bekannt ($H = 0,8$), aber nicht der Q-Wert. Bestimmen Sie den Punkt B und D_G so, dass die Voraussetzungen der Lorenzkurve nicht verletzt werden und die Disparität maximal wird.



Aufgabe 6.8

Zu den Vorteilen einiger neuer Bundesländer gehört die sich z.T. schon rasch entwickelnde Infrastruktur und die Ballung (Konzentration) der Wohnbevölkerung auf einige Großstädte. Angenommen, eines Tages sei völlig unerwartet Graf Dracula aus Polen kommend in Sachsen eingereist, vor allem wegen der genannten Vorteile unter dem Aspekt des Vampirismus. Aufgrund gewisser Eingewöhnungsschwierigkeiten und weil Sachsens Kultur dem Vampirismus schon immer fremd gegenüberstand, konzentrierte sich die Aktivität des Grafen und seiner Kollegen auf die Gebiete Görlitz und Bautzen. Einer Statistik des Sächsischen Statistischen Landesamtes konnte man entnehmen:



Gebietseinheit	Einwohner in 1000	Vampire
DD Dresden	520	80
L Leipzig	540	40
GR Görlitz	80	200
BZ Bautzen	50	120
übriges Sachsen	3810	60
insgesamt	5 Mill.	500

Geben Sie die Punkte der Lorenzkurve an, d.h. bestimmen Sie die Werte H_i und Q_i um die Lorenzkurve zu zeichnen (Skizze genügt) und berechnen Sie das Disparitätsmaß von Gini!

Aufgabe 6.9

Gleich nachdem die Regierung, unter dem Einfluss der entsprechenden Lobby, zur Überzeugung gelangte, die Professoren seien unterbezahlt, entbrannte ein dermaßen heftiger Streit unter den drei, etwa gleich stark vertretenen Professorentypen, dass die zusätzlichen Gelder nur noch nach dem Gießkannenprinzip verteilt werden konnten. Bisher mögen die Gesamteinkünfte X der Professoren im Verhältnis $1/4 : 1/3 : 5/12$ auf die drei Professorenarten aufgeteilt worden sein, und es werde nur ein Betrag in Höhe von $1/4 \cdot x$ zusätzlich nach dem Gießkannenprinzip verteilt. Wie ändern sich die Anteile der drei Gruppen, und wie verringert sich die Disparität (relative Konzentration) der Einkommen?

Aufgabe 6.10

Gegeben sei die folgende Einkommensverteilung:

Einkommensklasse	Anzahl d. Personen
0 - 1000	100
1000 - 2000	50
2000 - 3000	30
3000 und mehr	20

In jeder Einkommensklasse ist das Gesamteinkommen 80.000 DM.

- Berechnen (tabellarische Darstellung) und zeichnen Sie die Lorenzkurve. Berechnen Sie außerdem die Steigung der Lorenzkurve für jede der 4 Klassen.
- Berechnen Sie den Gini-Koeffizienten.
- Ermitteln Sie den Anteil der 25% reichsten Einkommensbezieher am Gesamteinkommen.
- Welche der nachfolgenden Aussagen trifft zu (mehrere richtige Antworten möglich)?
Die Lorenzkurve kann die Gleichverteilungsgerade (GVG)

☐	schneiden
☐	tangieren

☐	nicht schneiden
☐	nicht tangieren

Aufgabe 6.11

- Man bestimme anhand der Tabelle von Aufgabe 5.15 die Größen H und Q für die Lorenzkurve.
- Wenn man die erste Größenklasse (Nr. 1) der Inseln bei der Berechnung der Lorenzkurve wegließe, dann wird die Konzentration (Disparität)

☐	sich vergrößern	☐	sich verringern
☐	gleichbleiben	☐	kann man nicht sagen

- Kann man dies tun, d.h. einfach die erste Größenklasse weglassen, weil auf diesen Inseln

niemand wohnt?

- d) Wenn man die vierte Größenklasse weiter untergliedern würde, dann würde sich die Konzentration in der Regel

	sich vergrößern		sich verringern
--	-----------------	--	-----------------

Sie könnte allerdings auch gleichbleiben und zwar dann, wenn.....:

- e) Dank der weit über die Grenzen Russlands gepriesenen Großmut Fjodors wurden alle Schiffbrüchigen, die allein auf einer Insel auf Rettung warteten von der fürstlichen Marine gefunden und in die Hauptstadt (1000 Einwohner) gebracht. Wie ändert sich die Konzentration?

Aufgabe 6.12

Gegeben seien die folgenden offenbar unsinnigen Daten:

	Anteil der Betriebe (h_i)	Umsatzanteile (q_i)
Kleinbetriebe	0,5	0,2
Mittelbetriebe	0,2	0,6
Großbetriebe	0,3	0,2

Wie kann man leicht zeigen, dass die Einteilung in Klein-, Mittel- und Großbetriebe nicht stimmen kann?

Aufgabe 6.13

Es sei x das Durchschnittseinkommen ($\bar{x} = 3000,-$ DM) und es gelte $H(\bar{x}) = 0,75$. Ferner sei Z der Zentralwert (Median) und die Steigung der Lorenzkurve im Punkt $H(x=Z)$ ist 0,8.

Man bestimme den Median und stelle fest, ob (nach der Fechnerschen Lageregel) die Einkommensverteilung links- oder rechtssteil ist.

Aufgabe 6.14

Gegeben seien drei Einkommensklassen $E_1 < E_2 < E_3$ und ein Durchschnittseinkommen in Höhe von DM 2000,-. Es gelte $h_1 = 0,5$, $h_2 = 0,4$ und $h_3 = 0,1$ und für die Steigungen s der Lorenzkurve $s_1 = 0,4$, $s_2 = 0,75$ und $s_3 = 5$. Man bestimme die Einkommen E_1 , E_2 und E_3 !

Aufgabe 6.15

Es gibt gute Gründe dafür, anzunehmen, dass der Mann, der (die Frau, die) das erste Blasinstrument erfand in einer Zeit lebte, in welcher die Freizeit (F), die Musikalität (M), die Intelligenz (I) und der Wohlstand (W) recht ungleich verteilt waren.

- a) Für einige dieser Merkmale wäre es nicht sinnvoll, die Lorenzkurve zu bestimmen. Für welche Merkmale? Warum?
- b) Das damals übliche Höhlenleben bedingte es, dass sich z.T. größere Gemeinschaften bildeten. Eine steinzeitliche Volkszählung der Bewohner (= Bew.) der Höhlen in der Region R ergab:

Höhlen mit Bew.	Anzahl der Höhlen
1 - unter 6	50
6 - unter 10	20
10 - unter 30	10
30 - 70	20

Man bestimme die Lorenzkurve und mache eine Skizze ihres Verlaufs!

- c) Wie verändert sich die relative Konzentration durch folgende Vorgänge:
- Der Mann, der die Flöte erfand, möge ein Einzelgänger gewesen sein (eine Höhle für eine Person bewohnt haben). Es spricht für das kulturelle Niveau der Region R, dass man ihn ehrenvoll in eine Höhle zu 60 Personen aufnahm. In seine frühere Höhle zog der Urheber obiger Statistik, der aus seiner Höhle zu 5 Personen hinausgeworfen wurde.
 - Später gründet man eine eigene Höhle für die 100 Flötenspieler, die alle aus Höhlen mit 3, 4 oder 5 Personen kamen.
- d) Wieder einige Jahre später kam es zu einem Aufstand der 1400 Unmusikalischen gegen die 100 "privilegierten" Flötenspieler, der zu einer Revolution führte. Flöten wurden verboten, 500 Menschen starben, die verbliebenen 1000 Menschen bildeten neue Großhöhlen (H_1 bis H_5) wie folgt

Höhle	H_1	H_2	H_3	H_4	H_5
Bewohner	200	100	300	100	300

Man bestimme die Lorenzkurve!



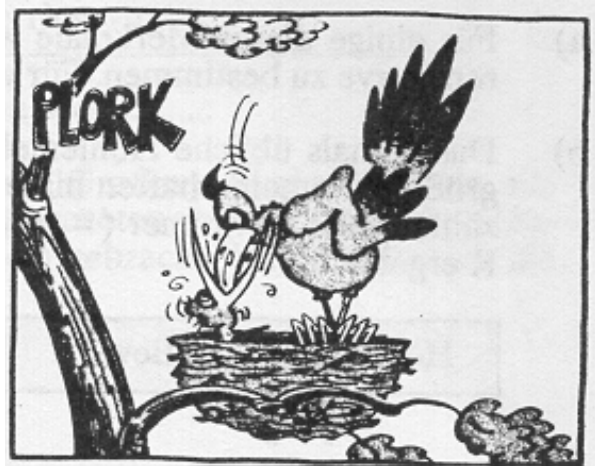
Aufgabe 6.16

"Je mehr man altert, desto mehr überzeugt man sich, dass Seine heilige Majestät der Zufall gut drei Viertel der Geschäfte dieses miserablen Universums besorgt."

(Friedrich der Große)

Bekanntlich hat die reaktionäre Natur den Kindersegen recht ungleich verteilt, so dass die Kinderzahl je Elternpaar auch bei den Vögeln erheblich schwankt. Hinzu kommt, dass gelegentlich auch Eier aus dem Nest fallen, so dass dieses ganz ohne Jungtiere sein mag. Für 10 der Nester im Garten des Diplom Kaufmanns K aus E möge jeweils die folgende Anzahl von Jungen gezählt worden sein:

0, 3, 2, 2, 3, 1, 0, 1, 1, 2



a) Man bestimme den Zentralwert und das arithmetische Mittel.

b) Die Verteilung ist (Richtiges ankreuzen):

☐	symmetrisch	☐	linkssteil
☐	rechtssteil	☐	linksschief

c) Bei Gleichverteilung der Jungtiere müssten in jedem Nest Junge sein.

d) Man zeichne die Lorenzkurve und berechne das Konzentrationsverhältnis von Gini!.

e) Der Zufall will es, dass die Kinderzahl in der Natur nicht bei allen Eltern gleich ist, sondern schwankt. Das hat zur Folge, dass die Varianz

☐	null	☐	negativ	☐	positiv
--------------------------	------	--------------------------	---------	--------------------------	---------

und deshalb notwendig auch die Konzentration (nach Gini)

☐	null	☐	negativ	☐	positiv
☐	beliebig (es gibt keinen Zusammenhang mit der Varianz)				

sein muß

Aufgabe 6.17

Gegeben sei die klassierte Verteilung

Größenklasse j	Klassenmittel x_i	relative Häufigkeit (h_j)
0 bis 8	6	0,5
9 bis 16	10	0,3
über 16	20	0,2

Man berechne den Gini-Koeffizient D_G für diese Verteilung!

Aufgabe 6.18

In einem Unternehmen mit 50 Beschäftigten ergab sich folgende Struktur der Verdienste: (x_j = Klassenmittelpunkt; n_j = Zahl der Lohn- und Gehaltsempfänger)

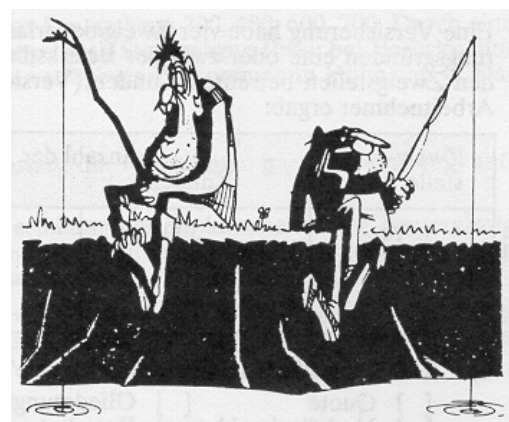
j	Verdienstverteilung		Lorenzkurve	
	x_j	n_j	H_j	Q_j
1	500	6	0,120	0,054
2	700	3	0,180	0,091
3	900	6	0,300	0,188
4	1100	15	0,600	0,484
5	1300	11	0,820	0,740
6	1500	4	0,900	0,848
7	1700	5	1	1

- Zeichnen Sie die Lorenzkurve!
- Fassen Sie jeweils mehrere Größenklassen zusammen und zeichnen Sie dann erneut die Lorenzkurve!

Aufgabe 6.19

Am Teich T_5 sitzen die beiden Angler A_1 und A_2 , die beide noch nie etwas von der Lorenzkurve gehört haben. Sie wissen deshalb nicht, dass sich der Fischbestand ihres Dorfes auf die Teiche T_2 und T_4 konzentriert. Es gibt 5 Teiche, die alle gleich groß sind und ein Statistiker hat den Fischbestand wie folgt geschätzt:

Teich	T_1	T_2	T_3	T_4	T_5
Fische	100	800	250	800	50



- Man bestimme den Zentralwert und das arithmetische Mittel

b) Das Merkmal “Anzahl der Fische” ist (Richtiges ankreuzen)

☐	intensiv	☐	diskret	☐	nominalskaliert
☐	extensiv	☐	stetig	☐	metrisch skaliert

Die Verteilung der Fische ist (Richtiges ankreuzen)

☐	rechtssteil	☐	symmetrisch	☐	linkssteil
☐	eindimensional	☐	zweidimensional		
☐	klassiert	☐	zweigipflig		

c) Man bestimme die Lorenzkurve und zeichne diese.

d) Zeigen Sie, dass Gini's Disparitätsmaß in diesem Fall den Wert 0,44 annimmt!

Aufgabe 6.20

Hinsichtlich der Gestalt einer Lorenzkurve (LK), welche die Preisgeldkonzentration der 20 bestplatzierten Tennisspieler der ATP-Weltrangliste darstelle, seien lediglich die Steigungen der Lorenzkurve in den vier Klassen bekannt, sowie der jeweilige Klassenumfang n_i :

Klasse	Steigung der LK	n_i	h_i
I	1/4	8 Spieler	
II	1	4 Spieler	
III	3/2	4 Spieler	
IV	2	4 Spieler	

a) Berechnen Sie die Lorenzkurve und den Gini-Koeffizienten.

b) Wie hoch ist die Preisgeldsumme der Klasse der bestverdienend Spieler, sofern insgesamt 10 Mio. DM an Preisgeld ausgeschüttet worden sind?

Aufgabe 6.21

Eine Versicherung habe vier Zweigniederlassungen und erwägt, ob sie nicht aus Rationalisierungsgründen eine oder zwei der Bezirksdirektionen schließen sollte. Eine Statistik der von den Zweigstellen betreuten Kunden (Versicherungsnehmer) und der in ihnen beschäftigten Arbeitnehmer ergab:

Zweigstelle	Anzahl der		Kunden je Beschäftigte
	Kunden	Beschäftigten	
A	300	15	20
B	900	60	15
C	100	10	10
D	700	35	20

a) Die Kennzahl "Kunden je Beschäftigte" ist eine

☐	Quote
☐	Verhältniszahl

☐	Gliederungszahl
☐	Rate

☐	Beziehungszahl
☐	Verursachungszahl

b) Um darzustellen, wie "ungleich" (im Sinne der relativen Konzentration = Disparität) die Zweigstellen in bezug auf die Anzahl der betreuten Kunden sind, geht man wie folgt vor (die Stellen [] ankreuzen):

1) man ordnet die Zweigstellen nach der Anzahl der

☐	Beschäftigten
--------------------------	---------------

☐	Kunden
--------------------------	--------

☐	Kunden je Beschäftigten
--------------------------	-------------------------

2) und setzt dann für die relativen Häufigkeiten h_i

☐	$h_1 = 0,05$, $h_2 = 0,15$, $h_3 = 0,35$, $h_4 = 0,45$
☐	$h_1 = h_2 = h_3 = h_4 = 0,25$
☐	$h_1 = 0,083$, $h_2 = 0,125$, $h_3 = 0,2917$, $h_4 = 0,5$

und für die Anteile q_i die Werte

☐	$q_1 = 0,05$, $q_2 = 0,15$, $q_3 = 0,35$, $q_4 = 0,45$
☐	$q_1 = 0,083$, $q_2 = 0,125$, $q_3 = 0,2917$, $q_4 = 0,5$
☐	$q_1 = 0,1538$, $q_2 = 0,2308$, $q_3 = q_4 = 0,3077$

c) Berechnen Sie nun die Lorenzkurve in der von Ihnen angegebenen Weise!

Aufgabe 6.22

Im Garten des Vogelfreundes V befinden sich 10 Nester. Da V sehr auf den Nachwuchs der Vogelfamilien achtet, zählte er eines Tages die Anzahl der Jungen in den Nestern (Variable X). Er erhielt die folgenden Werte 0,3,2,2,3,1,0,1,1,2.

Man zeichne die Lorenzkurve und berechne das Konzentrationsverhältnis von Gini!

Es gilt, mit obigen Zahlen $\bar{x} = 1,5$ und $s_x^2 = 1,05$.

Wenn davon auszugehen ist, dass zwei Drittel der ausgeschlüpften Jungen den diesjährigen unfreundlichen Sommer überleben, wie groß sind dann Mittelwert und Varianz der Anzahl x^* der Überlebenden Vogelkinder?



Aufgabe 6.23*Motto: Matth. 13, Vers 12*

- a) Vier Personen haben die Einkommen x_v vor Steuerabzug 300, 400, 600, 700. Durch eine unsoziale Steuergesetzgebung verbleibt ihnen nach Steuerabzug (bzw. bei Berücksichtigung von Steuervergünstigungen) noch das verfügbare Einkommen in Höhe von y_v . Dabei berechnet sich y_v wie folgt

$$y_v = x_v + \frac{x_v - \bar{x}}{2} \quad v = 1, \dots, 4$$

Man bestimme die Werte $\bar{y}$ und s_y^2 (= Varianz der Einkommen nach Steuerabzug) aufgrund von $\bar{x}$ und s_x^2 !

- b) Zeichnen Sie die Lorenzkurve für die Verteilung der Einkommen x vor Steuerabzug (Einkommensverteilung der vier Personen von Teil a) und der Einkommen y nach Steuerabzug! Wie groß ist Ginis Konzentrationsmaß?

c) Vorgriff auf Kap. 8: Der statistisch wenig vorgebildete Diplom-Kaufmann K aus E berechnete die Korrelation zwischen den Einkommen vor und nach Steuerabzug. Er erhielt $r_{xy} = +0,95$. Hat er richtig gerechnet? (Begründung)? Wie groß ist in diesem Fall die Kovarianz zwischen x und y ?

Aufgabe 6.24

Gegeben sei die folgende Verteilung der Einkommen (x):

von...bis zu...	Anzahl der Personen
0-1000	100
1000-2000	50
2000-3000	30
3000 und mehr	20

Jede Einkommensklasse verfügt über ein Gesamteinkommen von 80.000 Mark. Berechnen Sie die Lorenzkurve und die Steigung der Lorenzkurve für jede Einkommensgrößenklasse! In welcher Größenklasse befindet sich das Durchschnittseinkommen $\bar{x}$ und wie groß ist $\bar{x}$?

Aufgabe 6.25

Im Rahmen einer Untersuchung zur Situation in der Landwirtschaft einer Region wurde u.a. festgestellt, dass dort von 50 vorhandenen Betrieben die vier größten eine Fläche von insgesamt 520 ha, die 32 kleinsten Betriebe hingegen zusammen nur 164 ha bewirtschaften. Für die übrigen Betriebe liegt die landwirtschaftliche Nutzfläche bei insgesamt 316 ha.

- a) Zeichnen Sie die Lorenzkurve.
- b) Berechnen Sie den Gini-Koeffizient.

Aufgabe 6.26

Gegeben sei eine Einkommensverteilung mit nur zwei Klassen

Einkommensklasse Größe der Einkommen	Anteil in % an den Einkommensbeziehern	Anteil in vH am Ge- samteinkommen
untere Einkommen x_0	h	q
obere Einkommen x_1	$1 - h$	$1 - q$

Man zeige oder widerlege, dass unter diesen Voraussetzungen

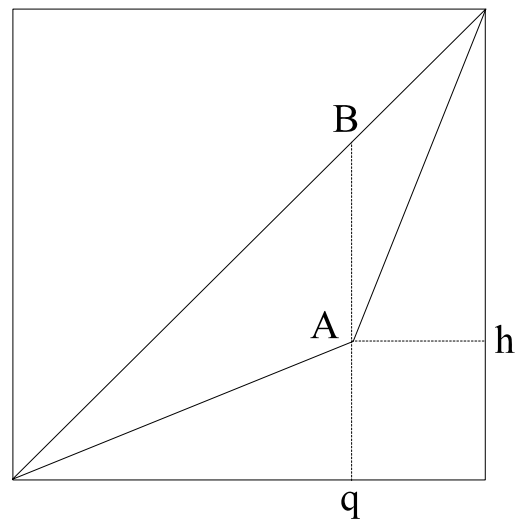
- a) das Disparitätsmaß D_G von Gini nichts anderes ist, als die Länge der Strecke AB (vgl. Skizze), dass also gilt:

$$D_G = h - q$$

- b) der Variationskoeffizient $V = s/\bar{x}$ unabhängig von der Größe der Einkommen x_0 bzw. x_1 beträgt:

$$V = D_G / \sqrt{h(1-h)}$$

so dass gilt: $V \leq 2D_G$

**Aufgabe 6.27**

Für eine Lorenzkurve gilt, dass die Steigung an der Stelle $H = 0,5$ kleiner ist als 1, nämlich 0,8. Kann man daraus folgern, ob die zugrundeliegende Verteilung des Merkmals x links- oder rechtssteil oder symmetrisch ist? Geben Sie eine Begründung an!

Aufgaben zum Kapitel 7**Aufgabe 7.1**

Bei drei Klausuren (A, B, C) mit jeweils 200 Klausurteilnehmern ergaben sich die folgenden Zusammenhänge zwischen Geschlecht und Klausurerfolg:

M = männlich W = weiblich B = Bestanden N = nicht bestanden

Klausur A		
	B	N
M	105	45
W	35	15

Klausur B		
	B	N
M	140	10
W	0	50

Klausur C		
	B	N
M	90	60
W	50	0

Man bestimme die Randverteilungen (was fällt dabei auf) und zeige, bei welcher Klausur ein

Zusammenhang zwischen Geschlecht und Klausurleistung besteht und welcher Art dieser ist! Was bedeuten "verbundene Beobachtungen"?

Aufgabe 7.2

Der Student S glaubt wieder einmal, eine Recht-Klausur ganz astrein gelöst zu haben. Mit seiner Selbsteinschätzung (Variable X), die mehr oder weniger gefühlsmäßig und zufällig, weniger aus tiefer juristischer Einsicht erfolgt, liegt er zwar oft in der Tendenz ganz richtig. Die genaue Klausurnote (Y) erscheint ihm aber fast immer rätselhaft und unerklärlich. So wie es ihm geht, ergeht es jedoch auch seinen 35 Mitstudenten. Dass die Noten bei den Rechtsklausuren irgendwie mysteriös sind glauben inzwischen alle. Das zeigt sich auch bei der Gegenüberstellung von X und Y für alle 36 Studenten:

Variable X	Variable Y				
	1	2	3	4	5
1	1	2	3	2	0
2	1	2	2	1	0
3	0	1	2	2	1
4	0	0	3	4	3
5	0	2	1	1	2

Bestimmen Sie die Randverteilungen und deren Mittelwerte sowie die empirischen Regressionslinie.

Aufgabe 7.3

Psychologen mögen festgestellt haben, dass ein reichhaltiges Warenangebot der Kauflust der Kundschaft förderliche sei, während gähnende Leere auf den Regalen eines Einzelhandelsgeschäfts im allgemeinen nicht besonders anregend wirke. Zwischen dem Wert (in 1000 DM) des Warensortiments (W) und den durchschnittlichen täglichen Verkäufen (V) bestehe der folgende Zusammenhang:

V	4	6	10	6	4
W	400	600	700	800	1000

Man zeichne das Streudiagramm und berechne die Kovarianz zwischen V und W sowie die Varianzen von V und W.



Aufgabe 7.4

Einer fehlgeschlagenen Intrige bei Hofe hat es Graf Giselher von Gelsenkirchen zu verdanken, dass er in einem Burgverlies schmachtet. Statt vor dem Verwaltungsgericht Gelsenkirchen zu klagen, (diese neuzeitliche Denkweise war Giselher noch vollkommen fremd) machte er sich daran, die meterdicke Wand zu durchbohren. Es gibt Tage, an denen er $y = 1$, $y = 2$ und $y = 3$ Zentimeter der Wand wegschaben kann. Zwischen dem Zeitaufwand X des Schabens (in Stunden) und der Zentimeterleistung Y des Verdünnens der Wand bestehe aufgrund von 10 Tagen Beobachtung, über die der Graf Aufzeichnungen machte - folgender Zusammenhang:

$x \backslash y$	1	2	3
6	1	1	0
10	1	2	1
15	0	2	2

- a) Man bestimme und zeichne (Skizze) die empirische Regressionslinie zur Schätzung von y ! Die Regressionslinie verläuft

	linear
--	--------

	nichtlinear
--	-------------

Die Korrelation r_{xy} ist folglich

	positiv
--	---------

	negativ
--	---------

- b) Man bestimme die Kovarianz zwischen X und Y !

- c) Wenn der Graf 1 cm Wand einmal in 6 Stunden und einmal in 10 Stunden abschabt, braucht er dann im Mittel 8 oder 7,5 Stunden? Zeigen Sie, dass das harmonische Mittel in der Tat 7,5 ist!



Aufgabe 7.5

Zwischen der Anzahl X der Feuerwehrlöschzüge und der Höhe des Brandschadens in 1.000.000 DM (Variable Y) bestand bei 9 Großbränden folgender Zusammenhang (beide Variablen von... bis unter...):



Y \ X	1-3	3-5	5-7	≥ 7
0 - 10	1	1	0	0
10 - 15	0	2	2	1
≥ 15	0	0	0	2

- a) Vergleichen Sie die Zahlenangaben mit den folgenden Einzelbeobachtungen (Sind die Angaben identisch?):
 (x_v, y_v) : (2,4), (3,4), (3,10), (4,10), (5,10), (6,10), (7,10), (7,16), (8,16)
- b) Bestimmen Sie die Randverteilungen und den Korrelationskoeffizient!
- c) Erklären Sie, warum Sie vermuten, dass zwischen der Anzahl der Feuerwehrlöschzüge und der Höhe des Brandschadens ein Zusammenhang bestehen dürfte!
- d) Bei wie viel Bränden waren bis zu 4 Löschzüge am Brandort und ein Schaden von bis zu 14 Millionen DM?

Aufgabe 7.6

Im Fach F prüfen seit Jahren die Professoren X und Y, wobei die Studenten "logischerweise" lieber zu X gehen, weil X milder sei. In einem Termin haben X und Y gemeinsam geprüft, so dass man die von ihnen vergebenen Noten im einzelnen miteinander vergleichen kann. Es ergab sich folgendes Bild:

Variable X	Variable Y				
	1	2	3	4	5
1	1	1	0	0	0
2	1	2	2	1	0
3	0	1	2	2	1
4	0	0	0	1	1

Man bestimme:

- Mittelwerte und Varianzen der Randverteilungen
- die Kovarianz und die Korrelation
- die bedingten relativen Häufigkeiten
- die bedingten Mittelwerte
- die Regressionsfunktionen x und y!
- Man zeichne die empirischen Regressionslinien (Verbindungen der bedingten Mittelwerte) und vergleiche sie mit den unter e) berechneten Regressionsgeraden!

Aufgabe 7.7

Gegeben sei die folgende zweidimensionale Verteilung:

Variable X	Variable Y			
	2	3	4	Σ
2	0,2	0,5	0,1	0,8
3	0,1	0,1	0	0,2
Σ	0,3	0,6	0,1	1

Man bestimme die Randverteilungen, die Regressionslinien und die Regressionsgeraden (vgl. Kap. 8) sowie Kovarianz und Korrelation. Sind die Regressionsgeraden und die Regressionslinien identisch?

Aufgabe 7.8

Es sei folgender Zusammenhang zwischen dem Alter X des Mannes und Y der Frau bei 25 Ehepaaren festgestellt worden:

X (Ehemann)	Y (Alter der Ehefrau)				
50	40	41	42	43	44
51	41	43	44	45	46
52	41	44	45	48	52
53	43	46	47	49	55
54	44	46	49	51	60

Zeichnen Sie das Streudiagramm und führen Sie Berechnungen in der Art von Aufgabe 7.6 durch!

Aufgabe 7.9

Gegeben sei die folgende zweidimensionale Verteilung mit $\bar{x} = 45$ und $\bar{y} = 30$:

x\y	24	28,2	35,7
28,5	5	15	10
37,5	7	21	14
42,0	12	36	24
54,5	18	54	36

Bestimmen Sie die Regressionslinie $x|y$ (Verbindung der bedingten Mittelwerte von x)
Hinweis: Schauen Sie sich die Zahlen genau an, Sie ersparen sich dadurch viel Rechenarbeit!

Was können Sie zum Verlauf der Regressionslinie $y|x$ (Verbindung der bedingten Mittelwerte von y) im Vergleich zu $x|y$ aussagen, ohne viel zu rechnen?

Aufgaben zum Kapitel 8

Aufgabe 8.1

(Pompadur und Pompamoll)

König Egon der XIII, auch der "Labile" genannt, hatte zwei Mätressen, die Pompadur (D) und die Pompamoll (M), die miteinander heftig um die Gunst des Königs konkurrierten. Dass sie jeweils verschiedene Seiten des empfindsamen Gemüts des Königs ansprachen und für ihn deshalb komplementär waren, steht seit der These des berühmten Historikers H in allen Lehrbüchern. H's jüngerer Kollege h glaubt dies jedoch aufgrund einer seinerzeit von der Hofschranze S verfassten Notiz empirisch widerlegen zu können. Aus dieser Notiz geht hervor, wie Egon seine Freizeit (gemessen in Stunden) in den letzten 10 Wochen des Jahres 17.. auf die Damen aufteilte:

Dur (D)	40	30	20	10	40	30	50	50	60	70
Moll (M)	30	10	30	40	20	30	50	30	40	20

Berechnen Sie die Korrelation und die Regressionsgeraden zwischen D und M, und folgern Sie daraus, ob H oder h recht hat!

Man bestimme die Korrelation und die Regressionsgeraden für die folgende Variante dieser Aufgabe (Pompadur und Pompamoll)!

Dur (D)	40	30	20	10	40
Moll (M)	30	10	30	40	20

Aufgabe 8.2

Antworten Sie auf die folgende Frage mit "richtig" oder "falsch": Die Korrelation (berechnet für eine Stichprobe von $n = 20$) zwischen den Punktzahlen X und Y von zwei Klausuraufgaben sei nur 0,224, so dass die Bestimmtheit r^2 nur 0,05 beträgt.

Bedeutet dies, dass:

- nur 5% der Punkte genau auf der Regressionsgerade liegen und 95% oberhalb oder unterhalb der Regressionsgeraden?
- nur bei einem von 20 Klausurteilnehmern ($1/20=0,05$) die Punktzahl der Aufgabe Y von der Aufgabe X abwich?
- nur bei einem von 20 Klausurteilnehmern ($1/20=0,05$) die Punktzahl in der Aufgabe Y von der abwich, die aufgrund der Punktzahl in Aufgabe X zu erwarten war?

Aufgabe 8.3

Es sei X der Intelligenzquotient (IQ) des Vaters und Y der des Sohnes. Psychologen fanden heraus, dass der IQ (praktisch in allen Generationen) mit Mittelwert 100 und Standardabweichung 16,4 symmetrisch verteilt ist, und dass die Korrelation zwischen dem IQ des Vaters und des Sohnes $r_{xy} = +0,5$ ist:

- a) Man bestimme die Kovarianz und
- b) die Regressionsgerade $y = a + bx$.
- c) Welcher IQ ist für den Sohn zu erwarten, wenn der Vater einen IQ von 75 (d.h. leichte Debität) und welche, wenn der Vater einen IQ von 130 (überragende Intelligenz) hat?
- d) Kann man aufgrund der Ergebnisse schließen, dass durch die Vererbung ein unaufhaltsamer Trend zum Mittelmaß besteht, so dass es nach einigen Generationen nur noch Personen mit einem IQ von 100 gibt (zu einem ähnlichen Schluss gelangten amerikanische Wirtschaftsforscher bei der langfristigen Analyse der Gewinne von 46 Unternehmen "The Triumph of Mediocrity in Business")?



Aufgabe 8.4

Zwischen dem Hopfenpreis in DM je Handelseinheit und dem Bierpreis (für Endverbraucher) in Pfennig je Liter bestehe folgender Zusammenhang:

Hopfenpreis	6	8	5	7	4
Bierpreis	120	150	125	145	110

- a) Bestimmen Sie eine lineare Schätzfunktion für den Bierpreis bei gegebenem Hopfenpreis!
- b) Welcher Bierpreis ist bei einem Hopfenpreis von 6,50 DM zu erwarten?
- c) Wie groß ist der Korrelationskoeffizient nach Bravais-Pearson (Produkt-Moment-Korrelation)?
- d) Wie würden sich die Werte der Regressions- und des Korrelationskoeffizienten verändern, wenn der Hopfenpreis in britischem Pfund, statt in DM, angegeben würde?

Aufgabe 8.5

Nachdem Andrea (A) zwei Jahre mit Charlie (C) ging, haben sie sich `ne echt besitzhafte Identität aufgebaut, aus der sich A nun emanzipieren will. Sie ist jetzt mehr so auf Bernd (B) drauf, kann aber noch nicht total auf B einflippen. Und weil ihr bisheriger Typ C die Trennungsverarbeitung erst einmal konkret abgecheckt haben will und das, was zwischen A und B emotional so läuft noch nicht so auffangen kann, haben sie jetzt alle drei beschlossen, das Problem bis spätestens zum nächsten Jahr zu dritt ganz konkret aufzuarbeiten.

Andrea teilte deshalb in den folgenden 12 Monaten ihre Zeit - sofern sie überhaupt Bock auf einen Typ hatte - auf B und C wie folgt auf:

Monat	1	2	3	4	5	6	7	8	9	10	11	12
Charlie(x_C)	3	3	6	9	3	3	3	6	9	6	12	12
Bernd (x_B)	18	18	18	12	18	9	12	12	18	12	9	3

- a) Man bestimme die zweidimensionale Häufigkeitsverteilung von B und C sowie die Randverteilungen. Man erhält für die Standardabweichungen $s_C = 3,34$ und $s_B = 4,66$ sowie für die Kovarianz $s_{BC} = -8,56$. Bestimmen Sie die Regressionsgeraden und den Korrelationskoeffizienten (r_{BC}).
- b) Bestimmen Sie die Regressionslinien.

Aufgabe 8.6

Diplom-Kaufmann K aus E hat zwar einen sehr kleinen Bekanntenkreis, der ihm aber doch sehr ans Herz gewachsen ist. Er bestehe aus $n=8$ Personen, 3 männlichen und einer weiblichen Arbeitskollegin, seiner Gattin, einem früheren Schulkameraden und zwei Freundinnen.

K ist jedoch ständig bemüht, die stark schwankende Zahl seiner Freundinnen y zu erhöhen, was erhebliche Telefonkosten x (in DM) verursacht. Es gelte folgender Zusammenhang:

$$\hat{y} = 0,5 + 0,01x \quad \text{und} \quad \hat{x} = 92,5 + 5y$$

- a) Man kann zeigen, dass K eine mittlere Anzahl $\bar{y} = 1,5$ von Freundinnen hat und im Mittel $\bar{x} = 100$ DM Telefonkosten aufwendet. Zeigen Sie, wie man aus den obigen Angaben zu diesem Ergebnis gelangen kann!
- b) Aus den obigen Angaben lässt sich auch folgern, dass x und y (Richtiges ankreuzen):

☐ positiv	☐ negativ
----------------------------------	----------------------------------

miteinander korrelieren.

Man kann ferner folgern, dass x und y stark miteinander korrelieren, also $r_{xy} \approx 1$

☐ ja	☐ nein
-----------------------------	-------------------------------

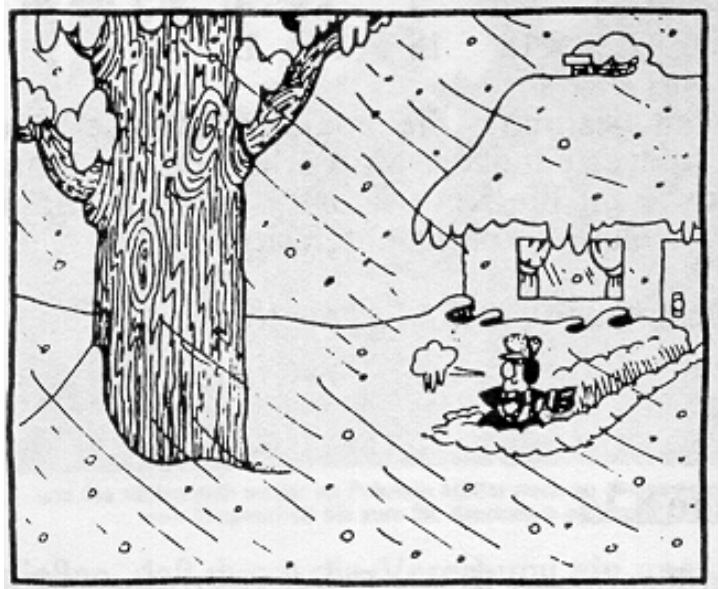


Aufgabe 8.7

Der statistisch geschulte Diplom Kaufmann K aus E stellte die folgenden Zusammenhänge zwischen der Außentemperatur X (in Grad Celsius) und der Dauer seines Weges zum Arbeitsplatz Y (in Minuten) fest:

X	Y
-20	60

-10	40
0	35
10	20
20	20



- a) Bestimmen Sie die Regressionsgerade $\hat{y} = a + bx$!
- b) K braucht bei 35° Hitze nur Minuten zum Arbeitsplatz und bei -30° Kälte nur Minuten.
- c) Führen Sie die Berechnungen unter a und b auch durch für die folgende Variante der Aufgabe:

X	-20	-10	0	10	20
Y	30	25	20	15	10

Aufgabe 8.8

Ein Drogist machte die Erfahrung, dass sich dann weniger Kunden nach dem Kauf eines Insektenbekämpfungsmittels über die Wirkungslosigkeit beschwerten, wenn er den Anteil x eines bestimmten Wirkstoffes erhöhte. Er machte bei 5 Präparaten (A bis E) folgende Erfahrungen:

x = Anteil des Wirkstoffes (in vH) y = Anteil der Beschwerden

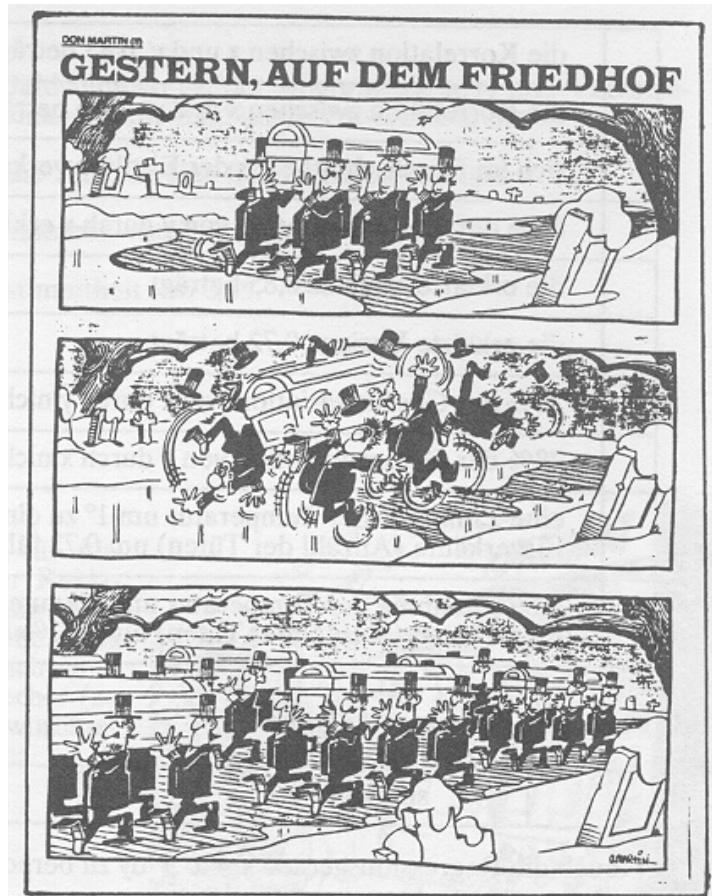
Präparat	x	y
A	2	10
B	3	7
C	4	6
D	5	5
E	6	2



- a) Berechnen Sie die Regressionsgerade $\hat{y} = a + bx$!
- b) Wie groß ist das Bestimmtheitsmaß? Interpretieren Sie die Aussage des Bestimmtheitsmaßes!

Aufgabe 8.9 (Hebelpunkt)

Normalerweise besteht kein ausgeprägter Zusammenhang zwischen der Beschaffenheit der Wege (gemessen am Rutschkoeffizient X) und der Sterblichkeit (Anzahl der Todesfälle Y) in der Gemeinde G. Die folgenden statistischen Aufzeichnungen umfassen 8 "normale" Tage sowie einen etwas ungewöhnlichen Tag (UT): Man bestimme die Regressionsfunktion $\hat{y} = a + bx$ und den Korrelationskoeffizienten mit und ohne UT und zeichne das Streuungsdiagramm.



X	3	4	5	6	4	5	6	7	8
Y	3	2	1	2	4	5	4	3	36

Aufgabe 8.10

Als K sein Diplomexamen an der Uni E gewonnen hatte, war ihm noch nicht klar, dass er längere Zeit als Diplom-Kaufmann K aus E keine angemessene Stellung finden würde. Er musste deshalb auf "Eisverkäufer" umsitzen. Als besonders wertvoll für seine Berufspraxis erwiesen sich seine mühsam erworbenen Statistikkenntnisse. So stellte er an 10 Frühlingstagen folgenden Zusammenhang zwischen der Temperatur (X) und den verkauften Eistüten pro Stunde (Y) fest:

X	12	16	17	18	20	20	21	23	24	29
Y	7	5	8	11	10	9	8	14	12	16

- Berechnen Sie die Regressionsgerade $\hat{y} = a + bx$ nach der Methode der kleinsten Quadrate sowie die Störgröße u an allen 10 Tagen!
- Die Varianzen der Größen sind bei den obigen Daten $s_y^2 = 10$ und $s_x^2 = 20$. Bestimmen Sie in Verbindung mit Teil a) dieser Aufgabe den Anteil an der Gesamtvarianz der Variable Y der durch die Regressionsgerade "erklärt" wird. Wie nennt man diesen Varianzanteil?

- c) Die Bestimmtheit in dieser Aufgabe beträgt 0,72. Das bedeutet, dass

	die Korrelation zwischen x und y 0,85 beträgt
	die Korrelation zwischen x und y -0,85 beträgt
	sich im Durchschnitt 85% der Eistüten verkaufen lassen
	85% der Gesamtvariation von y durch x erklärt wird
	die erklärte Varianz 0,85 beträgt
	die erklärte Varianz 0,72 beträgt
	28% der Gesamtvariation von x durch y nicht erklärt wird
	28% der Gesamtvariation von y durch x nicht erklärt wird
	eine Zunahme der Temperatur um 1° zu einer Zunahme der Eisverkäufe (Anzahl der Tüten) um 0,72 führt
	eine Zunahme der Temperatur um 1% zu einer Zunahme der Eisverkäufe (Anzahl der Tüten) um 0,72% führt



- d) Es wäre bei dieser Aufgabe

	sinnvoll
--	----------

	nicht sinnvoll
--	----------------

auch die Regressionsgerade $\hat{x} = c + dy$ zu berechnen.

Aufgabe 8.11

Diplom Kaufmann K aus E hatte jedes Spiel der 32 Mannschaften der Fußballweltmeisterschaft mehrmals im Fernsehen verfolgt. Sein Geisteszustand, gemessen anhand einer Punktzahl y in einem psychologischen Test hat sich dadurch ersichtlich verschlechtert. Psychologen haben festgestellt, dass sich der Wert y zu 81% durch die Anzahl x der betrachteten Fußballspiele erklären (bestimmen) läßt. Neben dieser Bestimmtheit von 0,81 ist noch bekannt, dass die Kovarianz zwischen x und y -5,4 und die Varianz von y genau 4 beträgt.

- a) Man bestimme die Residualvarianz, die erklärte Varianz und den Korrelationskoeffizienten!



- b) Wie groß ist die Varianz x der betrachteten Spiele?

- c) Die Regressionsfunktion lautet $\hat{y}_i = a + bx_i$. Berechnen Sie b!

Aufgabe 8.12

Die Deutsche Bundesbank korrelierte Geldvolumen (in der Abgrenzung M1) und Volkseinkommen für die Jahre 1960 bis 1970 miteinander. Sie fand unter Verwendung der

- vierteljährlichen Niveaudaten (absolute Zahlen) eine Bestimmtheit von 0,98
- halbjährlichen Zuwächsen eine Bestimmtheit von 0,37 und
- vierteljährlichen Zuwächsen eine Bestimmtheit von 0,25.

Erklären Sie das Ergebnis!

Aufgabe 8.13

Dem Hersteller eines Kaffeeautomaten für Autobahnraststätten ist es trotz zahlreicher Reklamationen bisher noch nicht gelungen, ein technisch ausgereifteres Modell auf den Markt zu bringen. Bei der bisherigen Ausführung schwankt die Menge X der abgegebenen Kaffeebecher (zu durchschnittlich 60 Pfennig) je eingeworfener Geldsumme Y (in DM) nicht unerheblich:

Berechnen Sie den Korrelationskoeffizient nach Bravais-Pearson und das Bestimmtheitsmaß!



X	22	18	24	17	25	14	25	28	26	21
Y	36	30	39	31	38	21	41	47	40	37

Aufgabe 8.14

Prüfen Sie jeweils ob die folgenden beiden Funktionen Regressionsgeraden sein können und bestimmen Sie den Korrelationskoeffizienten r:

a) $\hat{y} = 2 + 3x$ und $\hat{x} = -1 + y/3$

b) $\hat{y} = 2 + 3x$ und $\hat{x} = -(2/3) + y/3$

Aufgabe 8.15

Im Rahmen einer Anhörung zur Durchsetzung von Sparmaßnahmen im Polizeidienst behauptet der amerikanische Serienheld und Frauenliebling Don Johannes, dass die Anzahl der von ihm für Miami Nice gelösten Kriminalfälle in höchstem Maße vom jeweiligen Einsatz seines

Ferraris als Dienstfahrzeug abhängen. Tatsächlich konnte in Miami an fünf zufällig ausgewählten Tagen folgender Zusammenhang zwischen der Anzahl der gelösten Fälle (Y) und der dabei eingesetzten Ferraris (X) festgestellt werden:

i	x_i	y_i
1	1	2
2	2	5
3	2	4
4	3	6
5	2	3

- a) Ermitteln Sie die Koeffizienten a und b einer Kleinste-Quadrate-Regressionsfunktion. Wie viel Verbrechen würde Don J. aufklären, wenn er den Ferrari fünf mal benutzen würde?
- b) Angenommen, an einem Tag, der nicht in der obigen Regression berücksichtigt wurde, wären bei 4 Ferrari-Einsätzen tatsächlich 9 Verbrechen aufgeklärt worden. Welche Abweichungen (Gesamtabweichung, erklärte Abweichung und Zufallsabweichung) würden sich unter Zugrundelegung der unter a) ermittelten Regressionsfunktion für diese Beobachtung ergeben? Zeichnen Sie die Regressionsfunktion und die von Ihnen errechneten Abweichungen in das nachstehende Koordinatensystem ein (Beschriftung!).
- c) Ermitteln Sie zunächst s_y^2 und s_x^2 für die Beobachtungsdaten von Teil a). Wie groß ist die erklärte Varianz $s_y^2 \cdot r_{xy}^2$ bei einem Korrelationskoeffizient von $r_{xy} = 0,9$?
- d) Berechnen Sie unter Verwendung der soeben ermittelten Ergebnisse auf zweifache Weise s_{xy}^2 .

Aufgabe 8.16

Die zu bestimmende Regressionsgerade $\hat{x} = c + dy$ verläuft durch die Punkte $P_1(x_1 = 2 | y_1 = 6)$ und $P_2(x_2 = 4 | y_2 = 3)$. Könnte für die Steigung der anderen Regressionsgerade $\hat{y} = a + bx$ gelten (mehrere Antworten möglich; richtige Antworten ankreuzen):

☐	b = -1,5	☐	b = -0,3	☐	b = -2
☐	b = +1,5	☐	b = 0	☐	b = -0,1

Aufgabe 8.17 und 8.18 nicht aus der Buchveröffentlichung übernommen

Aufgabe 8.19

Der Zusammenhang zwischen dem Werbebudget X und dem Umsatz Y eines ausgewählten Produktes "P" wird von der Controlling-Abteilung eines Konsumgüterherstellers für die letzten 12 Monate untersucht.

Man erhielt für die 12 Monate folgende Größen:

- Werbeausgaben: insgesamt 120 000 DM im untersuchten Jahr.
 - Der durchschnittliche monatliche Umsatz des Produktes "P" in DM beträgt das zehnfache des durchschnittlichen Werbeetats pro Monat.
 - Die Standardabweichung der monatlichen Werbeausgaben im Untersuchungszeitraum beträgt 9 000 DM.
 - Standardabweichung der Umsatzdaten 50 000 DM.
 - Korrelationskoeffizient $r = 0,9$.
1. Berechnen Sie die Regressionsgerade für den Fall eines einfachen linearen Zusammenhanges der beiden Merkmale auf Monatsebene!
 2. Schätzen Sie den Umsatz bei einem Werbebudget von 11 000 DM in einem Monat!
 3. Angenommen, in einem bestimmten Monat, der bisher nicht in der Regressionsrechnung berücksichtigt wurde, wäre beim Einsatz eines Werbebudgets von 12 000 DM ein Umsatz von 120 000 DM erzielt worden. Welche Abweichungen (**Gesamtabweichung, erklärte Abweichung und Zufallsabweichung**) würden sich unter Zugrundelegung der ermittelten Regressionsfunktion für diese Beobachtung ergeben? Zeichnen Sie die Regressionsfunktion und die von Ihnen errechneten Abweichungen in ein Koordinatensystem ein.
 4. Berechnen Sie für die ursprünglichen Beobachtungsdaten die erklärte Varianz, das Bestimmtheitsmaß und das Unbestimmtheitsmaß!

Aufgabe 8.20

Die Mietervereinigung der Stadt DD hat am Jahresende für 600 Mietwohnungen nachfolgende Tabelle über den Mietpreis Y (in DM) in Abhängigkeit von der Wohnfläche X (in m²) veröffentlicht.

Wohnfläche	Mietpreis von... bis unter...			
	0 – 200	200 - 400	400 - 600	600 - 1000
0 - 40	100	50	0	0
40 - 80	80	220	10	0
80 - 120	0	20	40	80

- a) Bestimmen Sie die relativen Häufigkeiten und die absoluten Summenhäufigkeiten der zweidimensionalen Verteilung!
- b) Berechnen Sie die Randverteilungen und deren Maßzahlen!
- c) Geben Sie die bedingten Verteilungen an!
- d) Zeichnen Sie die empirische Regressionslinie!
- e) Berechnen Sie Kovarianz, Korrelationskoeffizient und Bestimmtheitsmaß!
- f) Wie lautet die Regressionsgerade $\hat{y} = a + bx$?

Aufgaben zum Kapitel 9

Aufgabe 9.1

In einer Arbeitsstätte seien 1998 im Jahresdurchschnitt $n=200$ Beschäftigte tätig gewesen, an die eine Lohn- und Gehaltssumme von $S=3$ Mill. DM gezahlt wurde. Dabei bestand folgende Aufgliederung nach dem Geschlecht:

Geschlecht	Beschäftigte		Lohnsumme	
	absolut	in vH	in 1000 DM	in vH
männlich	150		2400	
weiblich	50		600	
insgesamt	200		3000	

Man trage die fehlenden Werte für eine Verhältniszahl ein! Um welche Art von Verhältniszahl handelt es sich dabei?

Aufgabe 9.2

a) In der Bundesrepublik (alte Bundesländer) galten 1997 etwa die folgenden Zahlen

Ehescheidungen	163 Tausend	Eheschließungen	374 Tausend
Wohnbevölkerung	68 Millionen	Bestand an Ehen	15 Millionen

Die Scheidungsquote (Ehescheidungen auf 1000 Einwohner) betrug demnach 2,4. Bedeutet dies, dass nur ca. 0,2% der Ehen vor dem Scheidungsrichter enden? (Begründung!)

b) Welche Aussage könnte man mit einer Kennziffer "Bestand an Ehen/Ehescheidungen" erzielen? Kann man aus den mitgeteilten Zahlen etwas aussagen über die Verweildauer in der Ehe (= Ehedauer)?

Aufgabe 9.3

Die Umsatzentwicklung eines Einzelhandelsgeschäfts ergab bei zwei Warengruppen die folgenden Werte (in 1000 DM):

	1992	1993	1994	1995
Gruppe A	220	240	264	286
Gruppe B	440	484	528	594
insgesamt	660	724	792	880

a) Man berechne die Messziffernreihe für die Umsätze in den beiden Warengruppen und für den Gesamtumsatz (1992=100)!

b) Warum bewegt sich die Messziffer für die Gesamtumsätze stets in der Mitte der Messziffern für die beiden Warengruppen?

Aufgabe 9.4

Einer amerikanischen Statistik zufolge ergeben sich folgende Zahlen über die Unfallhäufigkeit von Männern und Frauen

Autounfall	Männer	Frauen	Summe
wenigstens einmal	3.122	2.255	5.377
nie	3.958	4.695	8.653
Summe	7.080	6.950	14.030

Kann man aufgrund dieser Zahlen schließen, dass Frauen bessere (sicherere) Autofahrer sind als Männer?

Es gibt allerdings auch Daten für zwei Teilgesamtheiten

1. häufiges Fahren (mehr als 10.000 Meilen)

Autounfall	Männer	Frauen	Summe
wenigstens einmal	2.605	996	3.601
nie	2.405	919	3.324
Summe	5.010	1.915	6.925

2. seltenes Fahren (< 10.000 Meilen)

Autounfall	Männer	Frauen	Summe
wen. einmal	517	1.259	1.776
nie	1.553	3.776	5.329
Summe	2.070	5.035	7.105



Aufgabe 9.5

Zwei Unternehmen (X und Y) mit jeweils den beiden Geschäftszweigen A und B erzielen in den beiden Geschäftszweigen jeweils die folgende gleiche Rendite (also sowohl beim Unternehmen X als auch beim Unternehmen Y):

$$R_A = 0,3 \quad R_B = 0,05$$

Das gesamte Unternehmen X erzielt aber eine Rendite von 10%, während das Unternehmen Y genau 25% erzielt, also

$$R_X = 0,1 \quad R_Y = 0,25,$$

obgleich $R_{XA} = R_{YA} = 0,3$ und $R_{XB} = R_{YB} = 0,05$.

Wie ist das möglich?



Aufgabe 9.6

Die Sterberate [früher "Sterbeziffer", d.h. die Zahl der Gestorbenen je 1.000 Lebende] anglikanischer Geistlicher ist viel höher (0,55%) als die der Bergarbeiter (0,15%):

L_x = Lebende des Alters x (Bestand in der Mitte des Jahres **)

D_x = Gestorbene im Alter von x im Jahre **

Altersklasse	Geistliche		Bergarbeiter	
	L_x	D_x	L_x	D_x
unter 50	10000	10	90000	90
über 50	90000	540	10000	60
insgesamt	100000	550	100000	150

Kann man aus den Angaben schließen, dass der Beruf des untertage arbeitenden Bergmanns "gesünder" ist als der des Geistlichen?

Bestimmen Sie das Durchschnittsalter der Lebenden und der Gestorbenen der beiden Berufsgruppen (Klassenobergrenze 100 Jahre).

Aufgabe 9.7

Einer monatlichen Wachstumsrate von 2% entspricht eine jährliche Wachstumsrate von $12 \cdot 2 = 24\%$. Ist das richtig? Wie verhält es sich bei einer monatlichen Wachstumsrate von 20%? (Jährlich 240%?)

Aufgabe 9.8

In der Bundesrepublik Deutschland stieg die Anzahl der Studenten von 1973 bis 1981 um 54,2%. Wie groß ist die durchschnittliche jährliche Wachstumsrate? Wann ist danach, ausgehend von 1973, mit einer Verdoppelung der Anzahl der Studenten zu rechnen?

Aufgabe 9.9

Der Verhaltensforscher V referiert im Fernsehen über die Lernfähigkeit von Papageien in Abhängigkeit von der Rasse. Er fand für den Schnöselspecht (S) folgenden Zusammenhang zwischen der Anzahl y_s der gelernten Worte und der Zeit t in Wochen (t : stetig)

$$y_s = e^{\sqrt{t}} \quad (t > 1)$$

während für den gemeinen Hauspapageien (P) gelten soll

$$y_p = (2t - 1)^2$$

- Für r_s gilt mit zunehmender Zeit t , dass diese Wachstumsrate (Richtiges ankreuzen):

☐	laufend abnimmt
☐	konstant bleibt
☐	laufend zunimmt

☐	negativ wird
☐	nur positiv ist
☐	gegen Null strebt

- Die Wachstumsrate r_p beim Hauspapagei

☐	kann auch über 100 % sein
☐	ist stets größer als r_s
☐	ist im allgemeinen kleiner als r_s

Aufgabe 9.10

Der häufig von skurrilen Vorstellungen geplagte Statistiker S wird nach dem Besuch einer einschlägigen Filmvorführung den Alptraum nicht los, dass Graf Dracula von den Toten auferstehen könnte. Bekanntlich wird ja auch ein friedlicher Bürger durch den Biss eines Vampirs selbst zum Vampir.

S geht davon aus, dass der "Durchschnittsvampir" pro Monat zwei Menschen das Blut aussaugt.

Wie lange wird es dauern, bis nach Draculas Wiedererscheinen eine Bevölkerung vom Umfang

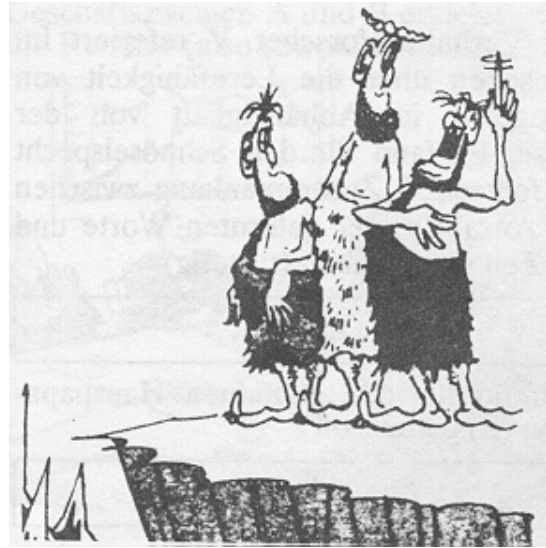
- einer Großstadt mit 700.000 Menschen
- der [alten] Bundesrepublik (60 Millionen Einwohner)

vollständig vom Vampirismus befallen sein wird?



Aufgabe 9.11 (Guru)

Trotz seines Studiums an der Uni Essen gelang es Diplom-Kaufmann K aus E nicht, eine Tätigkeit zu finden, die seinen gehobenen geistigen und materiellen Ansprüchen genügt. Er beschloss daher, eine Sekte zu gründen, um von den Zuwendungen seiner Jünger zu leben und ist jetzt der Guru G. Er fing zunächst bescheiden an mit nur zwei Jüngern (vgl. Bild). Er ging davon aus, dass seiner Lehre gemäß jeder Jünger in jedem Monat zwei neue Jünger hinzugewinnt (Austritte aus der Sekte waren nicht vorgesehen):



Die Realität entwickelte sich jedoch anders. Die Anzahl der Jünger stieg in zwei Quartalen nur von 2 auf 32 und zwar in den einzelnen Monaten 1 bis 6 wie folgt: 2, 5, 8, 20, 35, 32

Man beantworte die folgenden Fragen:

- Wie groß ist die tatsächliche mittlere monatliche Wachstumsrate der Anzahl J der Jünger?
- Wie groß ist die mittlere monatliche Wachstumsrate der Anzahl der Jünger, von der der Guru G ausging?
- Wie groß ist die aufs Quartal bezogene Wachstumsrate der Anzahl der Jünger, von der Guru G ausging?
- Wäre die Annahme des Gurus zutreffend gewesen: wie viel Jünger hätte es dann nach zwei Quartalen geben müssen?

Aufgabe 9.12 (fliegender Teppich)

Diplom Kaufmann K aus E gründete 1990 eine Spedition, deren Spezialität u.a. der Transport in den Nahen Osten war (Abtlg. 1 des Betriebs), was sich als eine beachtliche Marktlücke herausstellte. Durch navigatorische Fehlleistungen größeren Ausmaßes konnten jedoch einige Transporte nicht erfolgreich abgewickelt werden, so dass die Rentabilität zunächst (1991) noch zurückging, was auch für Abtlg. 2 des Betriebs (Polentransporte) galt. Gleichwohl konnte K die Rentabilität des Gesamtunternehmens steigern, wie die folgenden Zahlen (alle in 1000 DM) zeigen:

1990			1991		
Abt.	Gewinn	Kapital	Abt.	Gewinn	Kapital
1	32	400	1	126	1800
2	40	800	2	24	600

- a) Man berechne die Rentabilitäten der einzelnen Abteilungen und des Gesamtunternehmens 1990 und 1991. Wie ist die Zunahme der Gesamtrentabilität trotz abnehmender Rentabilitäten der beiden Abteilungen zu erklären?

- b) Wenn der Kapitaleinsatz um 10% (20%) steigt und der Gewinn um 25% (50%) dann nimmt die Rentabilität zu um $25 - 10 = 15\%$ (bzw. um $50 - 20 = 30\%$)?
- c) Die Rentabilität des Gesamtunternehmens ist ein gewogenes Mittel der Rentabilitäten der einzelnen Abteilungen, und zwar (richtige Art des Mittels und Art der Gewichtung ankreuzen)

Mittel	Gewinn- anteile	Kapital- anteile
arithmetisch		
harmonisch		

- d) Durch Kauf eines fliegenden Teppichs im Werte von 1,6 Mill. DM glaubt K die Rentabilität der Abteilung 1 verdoppeln zu können. Wie groß sind Gewinn und Rentabilität 1992 unter sonst gleichen Umständen?

- e) Die Rentabilität ist eine (Richtiges ankreuzen):

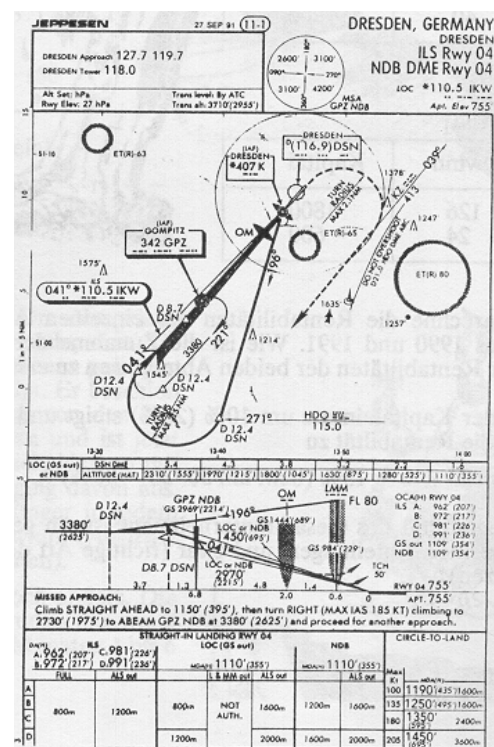
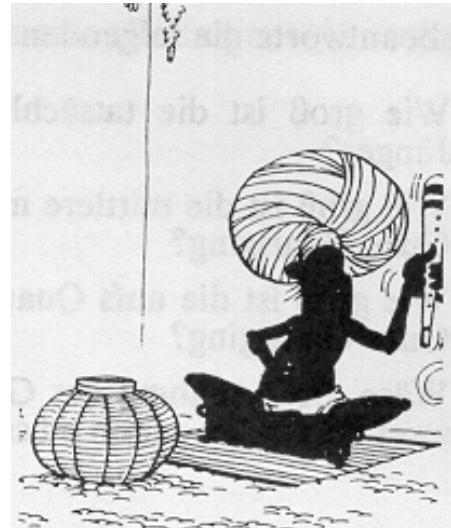
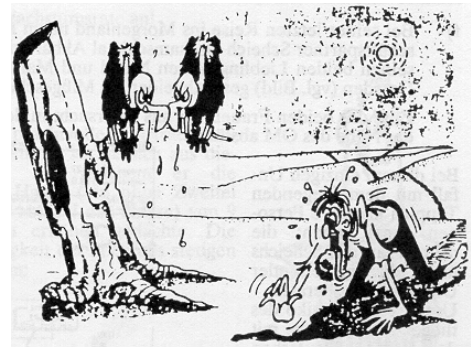
	Verhältniszahl		Gliederungszahl
	Rate		Beziehungszahl
	Quote		Messzahl
	Entsprechungszahl		Maßzahl

- f) Bei seiner letzten Reise ins Morgenland nahm K auf seinem Rückflug auch seinen Geschäftspartner Scheich Dschamschid al Akbar mit nach Dresden. Der Scheich wollte mit seinen beiden Lieblingsfrauen Nahid und Mahnoz Porzellan kaufen. Beim Anflug auf Dresden geschah leider ein Missgeschick:

Eine der beiden Frauen biss aus Eifersucht in den fliegenden Teppich, worauf dieser in der Nähe des OM abstürzte.

Bei diesem einzigen Unfall mit dem fliegenden Teppich kamen 2 Personen ums Leben; die dritte Frau des Scheichs und ein mitreisender Unternehmensberater. Um das Todesrisiko des fliegenden Teppichs mit dem der Deutschen Reichsbahn zu vergleichen ist die richtige Bezugsgröße:

- die Anzahl der transportierten Personen,
- die geflogenen (gefahrenen) Kilometer,



- die Personenkilometer,
- die Zeit,
- die Anzahl der Flüge (Bahnfahrten),
- die Anzahl der Fahrzeuge (Züge bzw. Flugzeuge)?

Aufgabe 9.13

Die Beat-Band B ist es gewohnt bei ihren Konzerten mit 8 Sattelschleppern voll Elektronik anzureisen. Dabei ist vermutlich der Zenit ihres künstlerischen Schaffens noch nicht erreicht. Die Anzahl F ihrer Fans und Konzertbesucher wächst von Jahr zu Jahr gemäß der Funktion (t = stetig)

$$F(t) = 4000 + 160t^3 - 30t^4$$

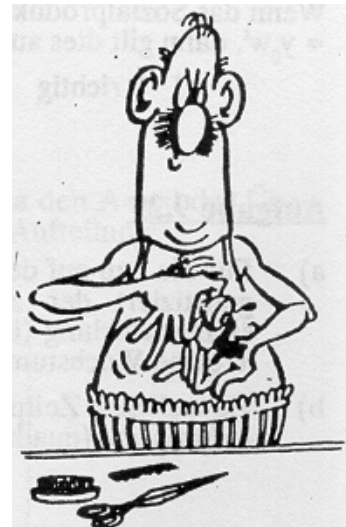
so dass die Band zur Zeit (im Jahre) $t = 4$ ihren Höhepunkt erreicht, nach dem es allerdings sehr dramatisch abwärts geht.

- Man zeige, dass die Band tatsächlich bei $t = 4$ den Gipfel ihrer Beliebtheit erreicht!
- Wie groß ist dann (zu diesem Zeitpunkt) die Wachstumsrate der Anzahl F der Fans und Konzertbesucher?
- Geben Sie die Funktion für die Wachstumsrate an!

Aufgabe 9.14

Der Rentner R glaubt sich durch eine prachtvolle Haartracht seinen Lebensabend verschönern zu sollen. Er kaufte sich aus diesem Grund ein Haarwuchsmittel, mit welchem er die ursprüngliche Gesamtlänge L seiner Haare (ein ohne Zweifel stetiges Merkmal, im Unterschied zur Anzahl der Haare) von 9 cm auf ein eindrucksvolles Maß zu erhöhen gedachte. Die Haarlänge entwickelt sich in Abhängigkeit der ebenfalls stetigen Zeit t (in Monaten) gemäß der Funktion:

$$L(t) = 9 + t^2/12 - t$$



- Man bestimme eine Funktion für die Messziffern zu Basis $t = 9$
- und eine Funktion für die Wachstumsrate $r_L(t)$ der Haarlänge
- Wie groß ist die Wachstumsrate $r_L(t)$ zur Zeit $t = 48$ (also nach genau 4 Jahren)?

Aufgabe 9.15

Gegeben ist die Funktion $x(t)$ mit (t : stetig)

$$x(t) = t^2 + 5 \cdot \sin(t) + 2 \cdot \cos(30t) + 3 \cdot \sin(2t) + 5 \cdot \sin(t)$$

Bestimmen Sie die Funktion $r_x(t)$ der Wachstumsrate von x !

Aufgabe 9.16

Der Bruch Schuldenstand/Sozialprodukt, heißt auch "Schuldenquote" und ist eine:

	Gliederungszahl		Beziehungszahl		Quote
	Verhältniszahl		Messzahl		Maßzahl

Angenommen, die Staatsverschuldung nimmt - was den tatsächlichen Verhältnissen des Jahrzehnts 1970-1980 in etwa entspricht - jahresdurchschnittlich um 15% zu. Wie lange dauert es dann jeweils, bis sich der Schuldenstand verdoppelt?

Ausgehend von $x_0 = 625$ erhält man, sofern gilt $y = 500 + 4x$ nach $t = 10$ Jahren einen Schuldenstand $x_t = x_{10}$ in Höhe von ... und ein Sozialprodukt in Höhe von ...

Die durchschnittliche jährliche Wachstumsrate des Sozialprodukts beträgt demnach, und sie ist somit

	kleiner		größer
--	---------	--	--------

als diejenige der Staatsverschuldung x .

Vorgriff auf Kap. 11

Wenn das Sozialprodukt y mit einer konstanten Wachstumsrate r zunimmt, also $y_t = y_0(1+r)^t = y_0w_t$, dann gilt dies auch für die gleitenden dreigliedrigen Durchschnitte des Sozialprodukts:

	richtig		falsch
--	---------	--	--------

Aufgabe 9.17

- a) Für die neu auf dem ostdeutschen Markt eingeführte Zigarettensmarke "Stasi-Light" prognostiziert der zuständige Marketing-Stratege *Egon Kranz* eine monatliche Absatzentwicklung (in Tsd. Stück), die der stetigen Funktion $y_t = 5 - 2t + t^2$ gehorcht. Welche Wachstumsrate ergibt sich daraufhin nach genau 2 Monaten ($t=2$)?
- b) Zu welchem Zeitpunkt t beträgt die Wachstumsrate genau 50%?

Aufgabe 9.18

Der Umsatz einer Unternehmung aus der chemischen Industrie weist folgende Wachstumsraten (gegenüber dem Vorjahr in Prozent) auf:

Jahr	1994	1995	1996	1997	1998
Wachstumsrate	+13,0%	+9,0%	-4,5%	+2,0%	+8,0%

- a) Bestimmen Sie die durchschnittliche jährliche Wachstumsrate der Unternehmung innerhalb der letzten fünf Jahre (Ansatz!).
- b) Das Chemieunternehmen legt zu Beginn des Jahres einen Teil seines Vermögens in Festgeld zum Zinssatz i_d an und vereinbart dabei eine jahresendliche Verzinsung. Als Alterna-

tive besteht die Möglichkeit einer jeweils halbjährigen Verzinsung des Kapitals zu i_s . Gilt bei identischem Guthaben am Ende der Laufzeit:

	$i_s = i_d$
--	-------------

	$i_s > i_d$
--	-------------

	$i_s < i_d$
--	-------------

Aufgabe 9.19

Angenommen, die Anzahl der Kunden nimmt jährlich um 2% zu und die Anzahl der Beschäftigten jährlich um 1,5% (jeweils konstante jährliche Wachstumsrate). Wie steigt oder sinkt dann eine mit der Kennzahl "Kunden je Beschäftigten" gemessene Art "Produktivität" eines Unternehmens, wenn man

- eine diskrete Variable Zeit annimmt (wie es hier bei größeren Intervallen von jeweils einem Jahr geboten wäre)?
- eine stetige Variable Zeit annimmt (also die Aussage über 2% bzw. 1,5% für kleine Intervalle gelten würde)?

Aufgaben zum Kapitel 10

Aufgabe 10.1

Fassen Sie die folgenden vier Preismesszahlen (1991=100) zu einem Index zusammen

Preise für	1997	Ausgabenanteil in %
Nahrungs- und Genussmittel	118,6	41,3
Kleidung, Schuhe	111,0	13,6
Wohnungsmiete	141,6	35,3
Energie	105,1	9,8

Quelle: http://www.statistik-bund.de/basis/bd_ueber.htm

- indem Sie einen ungewogenen Durchschnitt bilden!
- indem Sie mit den Ausgabenanteilen der Haushalte gewichten!
- Welcher der beiden Ansätze a) oder b) ist zu bevorzugen?

Die Zahlen beziehen sich auf den Preisindex für die Lebenshaltung aller Privaten Haushalte.

Aufgabe 10.2

Um seinen notleidenden staatlichen Dienstleistungsbetrieben finanziell auf die Sprünge zu helfen, plant ein Minister eine Gebührenerhöhung bei zwei von 200 Gebührenarten (A und B) und zwar um 50% (bei A) und um 100% (bei B). Die Ausgabenanteile für die Dienstleistungen A und B waren bei den Konsumenten bisher 10 bzw. 20%.



Wie groß ist der Preisindex nach Laspeyres?

Je nachdem, wie die Verbraucher reagieren, kann der Paasche Index zwischen und schwanken.

Aufgabe 10.3

Angenommen, das Sozialprodukt sei nominal (zu jeweiligen Preisen) um 10% gestiegen, real (zu konstanten Preisen eines Basisjahres) aber nur um 6%. Welchen Wert nimmt dann der Preisindex des Sozialprodukts (ein Preisindex nach Paasche) an?

Aufgabe 10.4

Gegeben seien zur Basiszeit (0) und zur Berichtszeit (t) die Preise zweier Waren, jeweils bezogen auf ein Kilogramm:

Ware	0	t
Tomaten (Salatware)	3	4
helles Mischbrot	2	4

Man berechne Durchschnittspreise p_0 und p_t sowie den von Dutot (1738) vorgeschlagenen Index $\bar{p}_t/\bar{p}_0$. Wie ändert sich der Preisindex nach Dutot, wenn man die Tomatenpreise nicht auf der Basis von kg sondern von Pfund notiert?

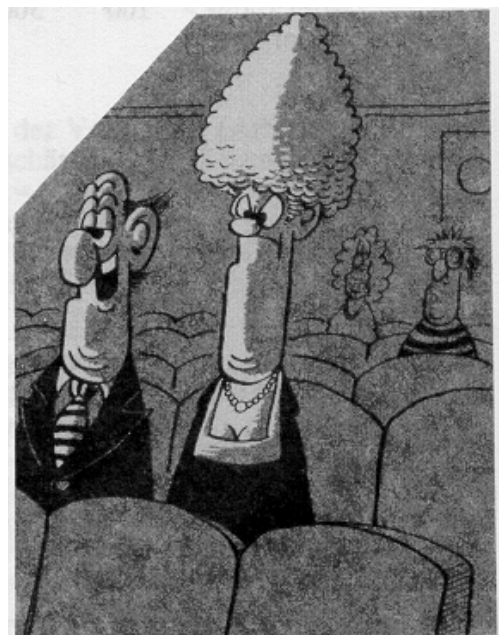
Aufgabe 10.5

Die Preise für die Lebenshaltung steigen um 20%. Gleichwohl bleiben die Lebenshaltungskosten gleich. Ist das möglich?

Aufgabe 10.6

Diplom-Kaufmann K aus E und Gattin gehen leidenschaftlich gern ins Kino. Von Zeit zu Zeit schätzen sie etwas Bildendes im "Filmkunst", und sie lassen sich auch schon mal politisieren im "Alternativkino". Die Ausgaben des Ehepaares für Kinobesuche sind von 1988 bis 1993 nominal um 40% und real um 20% gestiegen. Für die Eintrittspreise der Kinos gelte 1978 und 1983:

Nr.	Kino	88	93
1	Filmkunst FIKU	15	12
2	Alternativ AKI	9	12
3	Kolossal KOKI	12	24
4	Bahnhof BAKI	20	24



- a) Man berechne den Preisindex nach Laspeyres, wenn sich die Ausgabenanteile für Kinobesuche bei dem Ehepaar 1988 wie folgt verhalten: 1:3:2:1 (= Aufteilung der Ausgaben auf die vier Kinos).

b) Berechnen Sie den Preis- und Mengenindex nach Paasche!

c) Angenommen, es seien nur die oben mitgeteilten Preise bekannt und man kann nur davon ausgehen, dass K seine heimlichen Besuche im Bahnhofskino 1993 unterlässt und auch nicht mehr das Alternativkino aufsucht. Wie groß ist dann der Paasche Preisindex mindestens, und wie groß ist er höchstes ?

Aufgabe 10.7

Zwischen 1970 und 1980 hat sich das wertmäßige Bruttosozialprodukt verdoppelt; das volumenmäßige (in Preisen von 1970) Sozialprodukt ist dagegen von 1970 bis 1980 nur um 1/3 gestiegen. Der "Preisindex des Sozialprodukts" 1970 = 100 beträgt somit 1980 (Richtiges ankreuzen):

☐	150
--------------------------	-----

☐	166,67
--------------------------	--------

☐	133,33
--------------------------	--------

☐	66,67
--------------------------	-------

Aufgabe 10.8

Der Private Verbrauch (aus der Volkswirtschaftlichen Gesamtrechnung) habe sich nominal (zu jeweiligen Preisen) um 50% erhöht. Der Preisindex des Privaten Verbrauchs (ein Paasche Preisindex, 1980 = 100) "stehe" auf 125. Um wie viel hat sich der Private Verbrauch real (zu konstanten Preisen des Basisjahres 1980) erhöht?

Aufgabe 10.9

Bestimmen Sie die fehlenden Werte in der folgenden Tabelle, die sich durch Neuberechnung (neuer Warenkorb etc.) eines Index ergab:

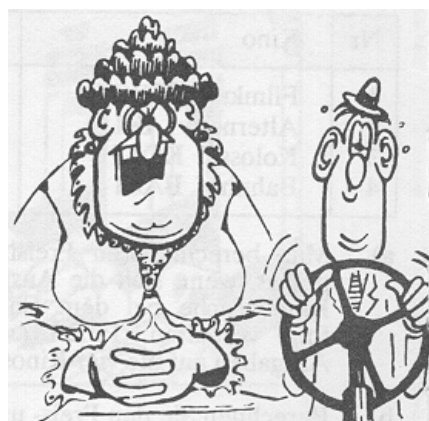
Jahr	1	2	3	4	5
alter Index	100	300	600		
neuer Index			900	1200	1800

Aufgabe 10.10

Diplom-Kaufmann K aus E leidet erheblich unter seiner dominanten Gattin, die ihm die Freude am Autofahren völlig verleidet, weil sie ständig sehr heftig Klage führt über

- die steigenden Benzinpreise und deshalb auch steigenden Lebenshaltungskosten
- die unkonzentrierte Fahrweise ihres Gatten.

Für die Benzinpreise (Ware A) und für die Preise (p) von zwei weiteren wichtigen Waren B und C sowie für die Mengen (q) mögen zur Basiszeit (0) und zur Berichtszeit t folgende Zahlen gelten:



Ware	p_0	p_t	q_0	q_t
A	0,9	1,5	10	12
B	2	2	4	6
C	5	3	3	8

- a) Man bestimme die Preisindizes nach Laspeyres und Paasche zur Basis $t = 0$!
- b) Angenommen, es gelten nicht die oben angegebenen Mengen zur Zeit $t=1$, sondern irgendwelche anderen Mengen. Man kann dann trotzdem sagen, dass der Preisindex nach Paasche mindestens den Wert..... haben müsste und andererseits könnte der Paasche-Preisindex maximal nur den Wert..... annehmen!
- c) Kann man generell sagen, dass der Laspeyres-Preisindex stets größer sei als der Paasche-Preisindex? Bedenken Sie bei Ihrer Antwort, dass K trotz seiner katastrophalen Fahrweise (und Gattin) auf das Autofahren (und damit den Benzinverbrauch) kaum verzichten kann!
- d) Vorgriff auf Kap. 11 (vgl. Aufg. 11.6)

Steigende Lebenshaltungskosten machten Diplom-Kaufmann K aus E allmählich den Gar-aus. Er verlor Auto und Gattin und fristet einsam sein Dasein zur Untermiete. In den Jahren 1994, 95 und 96 stellte er im März, Juli und November jeweils seine Kosten zusammen und stellte dabei fest, dass diese einen abnehmenden Trend haben, so dass er schon hofft, sich eines Tages wieder ein Auto und eine Frau leisten zu können. Er kam zu folgenden statistischen Daten:

	1994	1995	1996
März	69	63	57
Juli	71	65	59
November	64	58	52



Berechnen Sie gleitende Mittelwerte zu drei Perioden. Die Reihe der gleitenden Mittelwerte ist eine Gerade. Wie lautet die Funktion dieser Geraden?

Aufgabe 10.11

Gegeben seien die folgenden Daten über die Verteilung der Verdienste (Arbeitseinkommen) nach Gehaltsgruppen (GG) und über die Anteile der Beschäftigten an diesen Gehaltsgruppen eines Betriebes A im Jahre $t=0$ (Basisjahr) und im Jahr $t=1$ (Berichtsjahr) [vgl. auch Aufgabe 6.37]:

GG	Durchschnittsgehalt		Beschäftigtenanteil	
	$t=0$	$t=1$	$t=0$	$t=1$
1	600	720	0,5	0,1
2	1000	1100	0,3	0,4
3	2000	2000	0,2	0,5

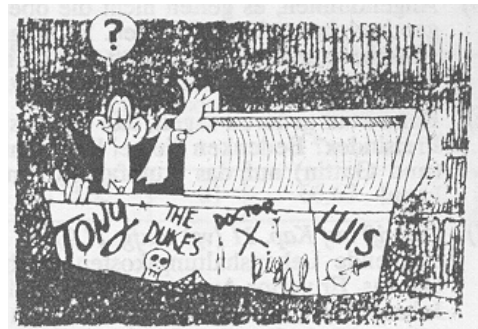
Berechnen Sie einen Laspeyres - Gehaltsindex mit der Beschäftigtenstruktur der Zeit $t=0$ als Basis (Gewichtung mit den Anteilen der Beschäftigten zur Basiszeit $t=0$)!

Erklären Sie folgenden Zusammenhang: Der oben berechnete Index zeigt einen Anstieg von nur 13%, das Durchschnittsgehalt des Betriebs ist dagegen, wie zu zeigen ist, um 51,12% gestiegen. Warum?

Aufgabe 10.12

Von vampirstatistisch großer Bedeutung ist die Entwicklung der Preise für die Lebenshaltung eines städtischen (Großstadt in NRW) Ein-Personen-Vampirhaushalts mit mittlerem Einkommen 1790 = 100. Dabei kommt es dem Statistiker sehr entgegen, dass der Warenkorb eines Vampirhaushalts leider nicht sehr viel Abwechslung bietet und folglich nur aus drei Waren besteht, deren Preise sich wie folgt entwickelt hatten:

Jahr	Nahrungsmittel		Wohnen (Tiefbau)
	Blut	andere	
1790	12	40	200
1890	26	52	850
1990	82	68	12800



Die Ausgaben teilten sich 1790, im Verhältnis 3:1:2 auf die Waren Blut, andere Nahrungsmittel und Wohnen auf.

Es kann gar keine Frage sein, dass dies kaum noch für die heutige Verbrauchsstruktur der Vampirhaushalte repräsentativ ist. Verstädterung, Bodenspekulation und ähnliche Erscheinungen haben nämlich den für Vampire geeigneten Wohnraum enorm verknüpft und verteuert. Man berechne gleichwohl den Preisindex für die Lebenshaltung zur Basis 1790 = 100 nach der Formel von Laspeyres mit den obigen Daten!

Aufgabe 10.13

Trotz verfeinerter Techniken der Beeinflussung gelingt es der Hausfrau H nicht, ihren Ehemann zu einer Anhebung des Haushaltsgeldes zu bewegen, weil dieser bei Problemen der Lebenshaltungskosten regelmäßig von großer Müdigkeit befallen wird. Es ist ihm insbesondere nicht klar zu machen, dass trotz Reduktion der Mengen die Lebenshaltungskosten durch steigende Preise zunehmen können.



Gegeben seien die folgenden Zahlen über Preise und Mengen dreier Waren zu den Zeiten 0 und t :

Ware	Preise		Mengen	
	0	t	0	t
A	10	15	60	50
B	25	20	40	70
C	30	40	80	60

Man berechne die Preisindizes nach Laspeyres und Paasche sowie die Zunahme der Lebenshaltungskosten!

Aufgabe 10.14

Der Haushalt des in erschütternder Armut lebenden arbeitslosen Diplom-Kaufmann K aus E (vgl. Bild) konsumiert nur zwei Waren A und B die zusammen den "Warenkorb" des Preisindex für die Lebenshaltung ausmachen, mit folgenden Mengen und Preisen:



	t = 0		t = 1	
	Preise	Mengen	Preise	Mengen
A	5	40	2	200
B	20	110	10	40

- a) Man bestimme den Wertindex W_{01} , den Laspeyres- (P_{01}^L) und den Paasche-Preisindex (P_{01}^P).
- b) Die Lebenshaltungskosten sind stärker zurückgegangen als die Preise. Daraus folgt, dass für die Mengenindizes nach Laspeyres (Q^L) und nach Paasche (Q^P) gelten muß (richtiges ankreuzen):

☐	Q^L und Q^P sind kleiner als 100 %
☐	$Q^L < P^P$, da $P^P < W$
☐	da $P^L > P^P$, muß $Q^L < Q^P$ sein
☐	da $P^L > W$, muß $Q^P < 1$ sein

Aufgabe 10.15

Eine Studie zur Wirtschaftsförderung empfahl zur Sanierung der Region "nördlicher Rio Grande do Sul" eine erhebliche Subventionierung der Schrumpfkopfproduktion oder eine gezielte Produktinnovation (Plastik- oder Holzköpfe) und neue Vertriebsformen (Versandhandel). Die Erzeugerpreisstatistik "Schrumpfköpfe von Europäern" gliederte nach

M = Missionare
 A = übrige Akademiker auf Forschungsreise
 U = Urlauber

und stellte folgende Zahlen fest:

	p_0	p_t	q_0	q_t
M	50	40	100	30
A	20	24	50	25
U	30	1	800	1200

p = Preis pro Stück

q = abgesetzte Mengen (Stückzahl)

a) Man berechne den Preisindex nach Laspeyres sowie den Mengenindex nach Paasche und den Wertindex zur Basis 0!

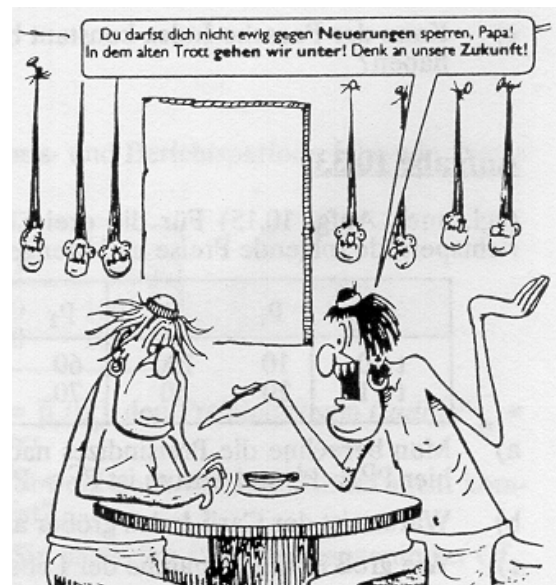
b) Geben Sie eine Empfehlung für eine mögliche Spezialisierung der Schrumpfkopfhersteller auf der Basis von Umsatzmesszahlen!

c) Für Schrumpfköpfe vom Typ A liegen folgende Angaben über die Angebotsfunktion $p_i = a + b x_i + u_i$ (p = Preis, x = Menge) vor:

$$a = 0, \sum x_i p_i = 3000, \sum x_i^2 = 5000 \quad \sum p_i^2 = 2000$$

Zeigen Sie, dass für $b = 0,6$ die Summe der Quadrate der Abweichungen $\sum u_i^2$ in der Tat ein Minimum darstellt! Wie groß ist dann $\sum u_i^2$?

Wie groß wäre $\sum u_i^2$, wenn man $b = 1$ statt $b = 0,6$ annehmen würde?



Aufgabe 10.16

Bekanntlich wird selbst der sog. "Schlangenfraß" (SF) immer teurer und auch die Anschaffungskosten (SA) für Schlangen leiden unter der allgemeinen Teuerung, so dass es nicht überrascht, wenn der Preisindex für die Schlangenbeschwörung von der Inflation in Mitleidenschaft gezogen wird. Es ist zudem auch wirtschaftspolitisch verständlich, dass der Lohn der Schlangenbeschwörer (SL) bei den Tarifrunden angemessen berücksichtigt werden sollte, so dass sich schließlich folgendes Bild ergibt:

t	Preise/Löhne			Mengen		
	SF	SA	SL	SF	SA	SL
0	4	100	55	10	8	3
1	4,8	120	66	5	8	6



- a) Man berechne einen Preisindex nach Laspeyres und nach Paasche!
- b) Es gibt Fälle, in denen es für das Ergebnis unwichtig ist, ob man nach der Laspeyres- oder nach der Paasche-Formel rechnet, man erhält stets das gleiche Zahlenergebnis. Wann ist das der Fall?
- c) Kann der Paasche-Index konstant bleiben, obgleich sich die Preise der Waren verändert haben?

Aufgabe 10.17

Für die drei Güter eines Warenkorbes galten für Basis- und Berichtsperiode folgende Preise und Mengen:

	p_1	q_1	p_2	q_2	p_3	q_3
$t=0$	10	100	60	100	20	100
$t=1$	20	50	70	80	30	60

- a) Man berechne die Preisindizes nach Carli P^C , Dutot P^D und Laspeyres P^L . Warum gilt hier $P^D = P^L$ und warum ist $P^C > P^L$?
- b) Warum ist der Carli-Index größer als der Laspeyres-Index also $P^C > P^L$?
- c) Wie groß ist die Zunahme der Lebenshaltungskosten?

Aufgabe 10.18

Ein Warenkorb bestehe nur aus zwei Waren (A und B). Es sei bekannt, dass der Paasche-Index und der Laspeyres-Index gleichermaßen den Wert 150 angenommen habe. Über die Mengen zur Zeit $t=0$ und $t=1$ sei jedoch nichts bekannt.

Die Preise lauteten

Gut	$t = 0$	$t = 1$
A	20	30
B	40	60

Welche Mengen könnten zur Zeit $t=0$ und $t=1$ nachgefragt worden sein?

Aufgabe 10.19

Gegeben sei ein Index, der aus drei Sektorenindizes besteht. Der Stand des Gesamtindex und der Sektorenindizes sowie die Gewichte der Sektoren sind der folgenden Tabelle zu entnehmen:

Sektor	Stand des Indexes zur Zeit t	Gewichte
1	140	(?)
2	150	0,2
3	(?)	0,3
Gesamtindex	148	

- a) Die fehlenden Angaben (?) sind zu ergänzen!
- b) Die Sektoren 1 und 3 werden zu einem Hauptsektor A und der Sektor 2 zu einem Hauptsektor B zusammengefasst; man berechne die Hauptsektorenindizes!
- c) Man berechne mit den Hauptsektorenindizes vom Teil b den Gesamtindex!

Aufgabe 10.20

Für die drei Güter eines Warenkorbes wurden in Basis- und Berichtsperiode folgende Preise und Mengen notiert:

	p_1	q_1	p_2	q_2	p_3	q_3
$t=0$	10	100	60	100	20	100
$t=1$	20	100	70	100	30	100

- a) Berechnen Sie den Preisindex von Dutot, den Preisindex von Carli sowie den Preisindex von Laspeyres.
- b) Wie erklären Sie die Identität der Ergebnisse, obwohl doch der Dutot-Index nicht kommensurabel ist und der Carli-Index ungewogen ist?
- c) Berechnen Sie den Wertindex W_{0t} und leiten Sie dann den Paasche-Mengenindex Q_{0t}^P aus Ihren bisherigen Berechnungen ab.

Aufgabe 10.21

Man zeige dass der Preisindex von Carli

- a) nicht verkettbar ist
- b) nicht der Zeitumkehrprobe genügt [d.h. dass nicht gilt: $P_{0t} = (P_{t0})^{-1}$]
- c) nicht der Additivität beim Basispreiswechsel genügt (Monotonieaxiom)!

Aufgabe 10.22

Gegeben seien Preise und Mengen von vier Waren zu zwei Zeitpunkten.

Gut	Preise		Mengen	
	$t = 0$	$t = 1$	$t = 0$	$t = 1$
A	10	14	40	48
B	20	28	30	24
C	30	48	20	16
D	40	64	10	12

Wie groß ist P_{01}^L bzw. P_{01}^P und warum gilt $P_{01}^L = P_{01}^P$?

Aufgabe 10.23

Für 4 Wirtschaftsgüter A, B, C und D sind die Umsätze für 1980 bekannt; ebenso das Verhältnis der 1992 und 1980 abgesetzten Gütermengen (Mengenmesszahlen $m_{80,92} = q_{92}/q_{80}$)

Gut	Umsatz*	Mengenmesszahl
A	10	1,5
B	8	1
C	12	2,0
D	20	1,4

*in Mill. DM

Man berechne aus diesen Angaben einen geeigneten Mengenindex!

Der Gesamtumsatz betrug 1992 genau 90 Mill. DM. Berechnen Sie den Paasche-Preisindex für 1992 zur Basis 1980!

Aufgaben zum Kapitel 11**Aufgabe 11.1**

Aus dem Statistischen Jahrbuch der Bundesrepublik Deutschland sind folgende Zahlen über die Anzahl der rechtskräftig Verurteilten (in Tausend) zu entnehmen:

Jahr	Anzahl	Jahr	Anzahl
1974	699	1981	747
1975	665	1982	772
1976	699	1983	785
1977	723	1984	753
1978	739	1985	720
1979	719	1986	705
1980	732	1987	691

Stellen Sie die Zeitreihe grafisch dar und beurteilen Sie, ob ein ansteigender oder absteigender Trend vorliegt und ob dieser Trend gegebenenfalls durch eine zyklische Bewegung überlagert ist!

**Aufgabe 11.2** über Jürgen v. d. Lippe

Bekanntlich gibt es Fernsehkünstler, die es verstehen, junge Menschen zu begeistern. Andererseits ist jedoch das jugendliche Gemüt empfindsam und ungeduldig und es neigt zum Perfekten und zu raschem Wandel. Es überrascht daher nicht, dass die Mitgliederzahl (in 1000)

des "von der Lippe Fan Clubs" (vdLFC) gewissen Schwankungen unterworfen ist, denn nicht alle vom Meister vorgetragenen Songs konnten Gefallen finden.

Für die Mitgliederzahl des vdLFC galten in den letzten 8 Monaten die folgenden Zahlen:

15, 20, 25, 30, 29, 31, 39 und 38.

Man berechne gleitende Mittelwerte zu je drei Monaten sowie einen Prognosewert für den neunten Monat ($a = 0,2$)!

Aufgabe 11.3

Der Kartenvorverkauf für das Konzert des berühmten Dirigenten D ergab in den ersten 15 Tagen die folgenden Werte ($t = 0, 1, \dots, 14$) für die Anzahl K_t der verkauften Karten:

40, 50, 60, 100, 110, 120, 160, 170, 180, 220, 230, 240, 280, 290, 300

a) Stellen sie die Zeitreihe K_t graphisch dar

b) Berechnen und interpretieren Sie

- gleitende 3-Tages-Durchschnitte der Zeitreihe K_t
- einen linearen Trend $K_t = a + bt$ nach der Methode der kleinsten Quadrate!



Aufgabe 11.4

Für vier Jahre liegen in einer Zeitreihe die vom Trend bereinigten Werte vor. Zeichnen Sie den Verlauf der trendbereinigten Werte sowie die Zeitreihe der Saisonnormalen.

Monat	Jahr 1	Jahr 2	Jahr 3	Jahr 4
Jan	-10,7	-11,6	-5,2	-15,0
Februar	-7,8	-6,4	-8,2	-2,4
März	7,4	0,7	10,2	11,0
April	-0,6	3,7	7,2	11,1
Mai	-0,4	9,6	-1,2	-9,8
Juni	6,6	-11,6	-5,3	4,2
Juli	-7,4	-1,2	3,3	0,1
August	-10,4	-11,3	-17,5	-20,1
September	2,8	2,0	8,8	8,5
Oktober	17,1	18,3	20,7	11,8
November	10,7	3,6	1,9	4,9
Dezember	-1,6	-8,7	-8,1	3,6

Ist die Annahme einer starren Saisonfigur in Form einer Saisonnormalen für diese Zeitreihe gerechtfertigt?

Aufgabe 11.5

Gegeben sei die folgende Zeitreihe:

t	-2	-1	0	+1	+2
y_t	10	12	15	17	16

Man berechne den Trend mit der Methode der kleinsten Quadrate und vergleiche die so erhaltenen fünf Trendwerte mit gewogenen arithmetischen Mitteln der fünf Ursprungswerte y_t der obigen Zeitreihe, wenn man die folgenden fünf Gewichtungsschemen benutzt:

-2	-1	0	+1	+2
0,6	0,4	0,2	0	-0,2
0,4	0,3	0,2	0,1	0
0,2	0,2	0,2	0,2	0,2
0,0	0,1	0,2	0,3	0,4
-0,2	0	0,2	0,4	0,6

Was fällt bei der Betrachtung der fünf Gewichtungsschemen auf?

Aufgabe 11.6

Boxprofi B aus St. Louis hat in drei Jahren neun Kämpfe absolviert und dabei jeweils nach Punkten gewonnen. Aus statistischen Gründen kämpfte er jeweils zu äquidistanten Zeitpunkten und zwar im April, August und Dezember in den Jahren 1989, 1990 und 1991. Die Punkte waren dabei:

	Jahr		
Monat	89	90	91
April	69	63	57
August	71	65	59
Dezember	64	58	52



- Berechnen Sie gleitende Mittelwerte zu drei Perioden! Hinweis: die Reihe der gleitenden Mittelwerte ist eine Gerade. Wie lautet die Funktion dieser Gerade?
- Berechnet man jetzt einen linearen Trend mit der Methode der kleinsten Quadrate, so stellt sich heraus, dass der so errechnete Trend eine betragsmäßig größere Steigung hat als der unter a) berechnete Trend.

Das ist nicht erstaunlich, weil die Verfahren von verschiedenen Modellannahmen (Voraussetzungen) ausgehen. Welche sind das?

- c) Mit der obigen Zeitreihe lassen sich zwei Komponenten errechnen, nämlich der Trend T (bzw. die glatte Komponente) und die Saisonkomponente S. Wodurch sind die Komponenten formal (mathematische Eigenschaften!) gekennzeichnet und wie sind sie aufgrund der Berechnungen unter a) rechnerisch bestimmt?

	formale Kennzeichen	Rechenergebnisse (aus a))
T		
S		

- d) Angenommen, die Punktzahl, die der Boxer erhält, schwankt ganz zufällig zwischen 59 und 61. Die ersten drei Werte seien

59, 61, 60,

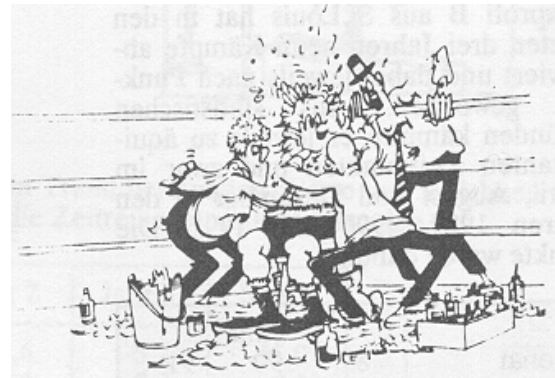
so dass der erste gleitende Durchschnitt von drei Perioden 60 ist. Der nächste gleitende Durchschnitt ist dann

mindestens:	höchstens:
-------------	------------

Der gleitende Durchschnitt wird also weniger schwanken können als die Ursprungswerte. Ferner gilt: Phasen des Anstiegs (bzw. Abstiegs) der Zahlenwerte des gleitenden Durchschnitts halten in der Regel (Richtiges ankreuzen)

☐	länger
☐	kürzer
☐	genauso lang

an, wie die der Ursprungswerte und es können in den gleitenden Mittelwerten Zyklen auftreten, die in den Ursprungswerten nicht auftreten. Das nennt man- Effekt.



Aufgabe 11.7

Angenommen, die Umsätze eines Unternehmens haben sich in den letzten sechs Quartalen wie folgt entwickelt:

Quartal	1	2	3	4	5	6
Umsatz	1800	1600	1900	2100	1400	2000

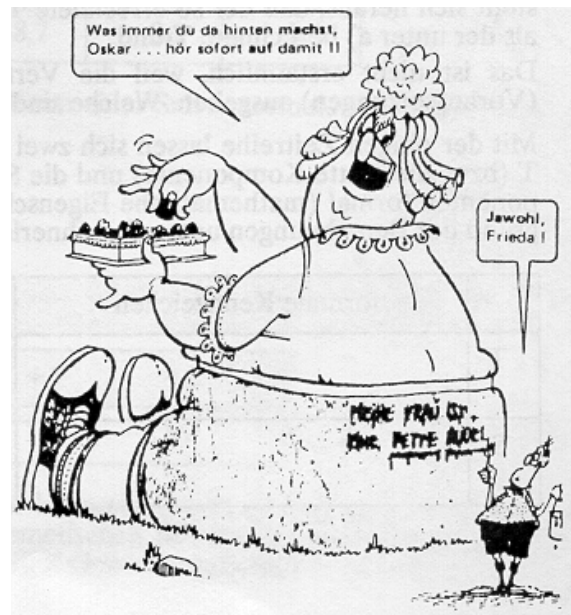
Berechnen Sie zentrierte gleitende Vier-Quartals-Durchschnitte!

Aufgabe 11.8

Die Hausfrau H wusste stets kulinarischen Genuss zu schätzen und entwickelte unterdessen ein Raumbedürfnis welches hienieden sonst nicht schicklich ist.

Sie trachtete deshalb hinfort danach, durch Schlankheitsmittel ihre Proportionen auf ein gefälligeres Maß zu reduzieren.

Dabei gebrach es ihr jedoch an der gebotenen Konsequenz, so dass ihr Gewicht y (in kg) stark schwankte und sich eine nachhaltige Reduktion nicht einstellen wollte, wie die folgenden Zahlen zeigen:



t	0	1	2	3	4	5	6
t^*	-3	-2	-1	0	+1	+2	+3
y_t	120	130	125	120	130	125	120

Man bestimme

- Gleitende Durchschnitte zu jeweils 3 Perioden,
- einen linearen Trend mit der Methode der kleinsten Quadrate und
- einen Prognosewert für die Periode 7 mit der Methode des exponentiellen Glättens ($\alpha = 0,2$)!

Aufgabe 11.9

Der Verkauf von Sonnenöl (y_t) der Marke "Roberto Blanco" hat sich in den letzten 13 Perioden wie folgt entwickelt (in Tsd. Flaschen):

t	1	2	3	4	5	6	7	8	9	10	11	12	13
y_t	22	25	28	25	28	31	28	31	34	31	34	37	34
$\tilde{y}_t$		25	26	27	28								
$y_t - \tilde{y}_t$													

- Berechnen Sie unter Bestimmung der Zykluslänge p die gleitenden Durchschnitte y_t sowie die trendbereinigten Werte ($y_t - \tilde{y}_t$).
- Ermitteln Sie außerdem die lineare Trendfunktion $y_t = a + bt$ mit der Methode der kleinsten Quadrate.

Aufgabe 11.10

Das Stahlwerk S ist in den letzten Jahren von einer wachsenden Zahl von Besuchern besichtigt worden, darunter auch ein zunehmender Anteil A von Frauen. In der gleichen Zeit haben sich auch Umsatz U und Kapitalstock K der Firma beträchtlich erhöht. Für die Anzahl B der Besucher gelte:

Jahr	Besucher
1990	150
1991	200
1992	250
1993	300
1994	290
1995	310
1996	390
1997	380



- a) Man bestimme gleitende Mittelwerte von jeweils drei Jahren!
- b) Die gleitenden Mittelwerte stellen dar

☐	die Saisonkomponente
☐	die Konjunkturkomponente
☐	den Trend, bzw. die glatte Komponente
☐	die vom Trend bereinigten Werte

- c) Welche Größe erhält man, wenn man die Abweichungen von den gleitenden Mittelwerten berechnet und wie groß sind diese Abweichungen?
- d) Geben Sie durch Ankreuzen an, um was es sich bei den genannten Größen handelt:

K	☐	Bestandsmasse	☐	Bewegungsmasse
U	☐	Bestandsmasse	☐	Bewegungsmasse
B	☐	Bestandsmasse	☐	Bewegungsmasse
A	☐	Gliederungszahl	☐	Beziehungszahl
U/K	☐	Gliederungszahl	☐	Beziehungszahl

- e) Um herauszubekommen, in welchen Jahren sich die Besucherzahl besonders häuft (konzentriert) berechne ich (Richtiges ankreuzen):

☐	die Lorenzkurve, weil sich die Besucherzahl auf bestimmte Jahre konzentriert
☐	die Abweichungen vom Trend, weil dies zeigt, in welchem Jahr besonders ungewöhnliche Verhältnisse vorlagen

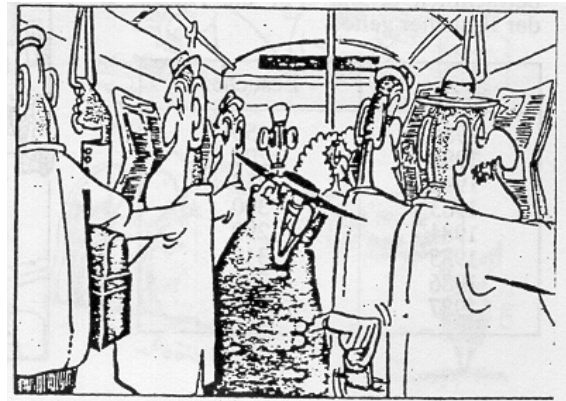
	den Modus
	die Korrelation mit dem Umsatz, weil die Besucherzahl hiervon abhängig ist

Aufgabe 11.11

Die Verkehrsbetriebe der Großstadt L sind besorgt über die häufige Meldung von Diebstählen in den Straßenbahnen. Es besteht die Vermutung, dass das Auftreten von Taschendieben saisonabhängig und außerdem abhängig von der Anzahl der Verkehrsteilnehmer (Fahrgäste) ist. Eine Statistik für zwei Jahre ergab die folgenden Zahlen über Diebstähle und Straßenbahnbenutzer (Fahrgäste):

Jahr /Quartal	Diebstähle	Fahrgäste ^{*)}
1/ 1	100	20
2	80	18
3	0	10
4	140	24
2/ 1	100	20
2	80	18
3	0	10
4	140	24

^{*)} in Tausend



a) Man bestimme für die Anzahl der Diebstähle zentrierte gleitende Mittelwerte zu je vier Werten.

b) Die Rechnung zeigt (Richtiges ankreuzen), dass

☐	es keine Saisonkomponente gibt
☐	es keinen Trend gibt
☐	der Trend ansteigt
☐	der Trend absinkt
☐	die Saisonkomponente konstant 80 ist

c) Die Anzahl der Diebstähle je Straßenbahnbenutzer ist

☐	konstant	☐	☐	eine Verhältniszahl
☐	eine Messzahl	☐	☐	ein Mittelwert
☐	eine Gliederungszahl	☐	☐	eine Beziehungszahl

d) Man zeichne das Streudiagramm der Anzahl der Straßenbahnbenutzer (X) und der Diebstähle (Y) [Achsen beschriften!] und beantworte die folgenden Fragen:

Die Regression ist

☐	linear
☐	positiv

☐	nicht linear
☐	negativ

Der Korrelationskoeffizient ist schätzungsweise:

e) Bestimmen Sie die klassierte Verteilung der Diebstähle und der Fahrgäste [Straßenbahnbenutzer] (Einheiten sind jetzt die Quartale):

Diebstähle	n_i	h_i
0 bis unter 90		
90 bis 150		

Fahrgäste	n_i	h_i
0 bis 19		
über 19		

Kann man aus der Ähnlichkeit der beiden Verteilungen bereits schließen ob die Anzahl der Diebstähle und der Straßenbahnbenutzer evtl. hoch miteinander korreliert ist? Begründung!

Aufgabe 11.12

Für einen in der Innenstadt angesiedelten Supermarkt wurden in den letzten drei Jahren folgende trendbereinigte Umsätze (in Mio. DM) ermittelt:

Quartal	Jahr		
	1996	1997	1998
I	-4	+1	-3
II	+8	+4	+6
III	-5	-3	-4
IV	+12	+8	+10

Ermitteln Sie die

- nicht - normierte
- normierte

Saisonfigur sowie die Restkomponente!

Aufgaben zum Kapitel 12

Aufgabe 12.1

Wegen des zur Nachsaison unsicheren Wetters ist das Badevergnügen oft von nur kurzer Dauer. Andererseits ergreifen jedoch die Urlauber angesichts der ihnen entstandenen Kosten und Mühen jede sich bietende Gelegenheit, den Strand aufzusuchen. Am Strand von Kata-pulco gab es mithin an einem Vormittag (von $t_0 = 9$ bis $t_m = 13$ Uhr) ein ständiges Kommen und Gehen von (zwecks Rechenvereinfachung) nur fünf Urlaubern A,...,E. Für die Zeiten galt:

	Zugang	Abgang
A	09 ³⁰	10 ⁰⁰
B	10 ³⁰	11 ⁰⁰
C	10 ⁴⁵	12 ³⁰
D	09 ⁴⁵	11 ¹⁵
E	11 ⁴⁵	12 ⁴⁵

- a) Man zeichne das Becker'sche Diagramm und berechne die folgenden Kennzahlen der Bestandsanalyse:

Zeitmengenfläche F_{om}	
Durchschnittsbestand	
mittlere Verweildauer	
Umschlagshäufigkeit	

- b) Erklären Sie in eigenen Worten, was mit der Kennziffer "Umschlagshäufigkeit" ausgesagt wird:

- c) Berechnen Sie Anfangs- (B_o) und Endbestand (B_m) sowie die mittlere Aufbau- (d_o) und Abbauzeit (d_m) für das Beispiel 12.1, wenn man als Beobachtungsintervall die Zeit von 10 bis 11 Uhr zugrunde legt. Berechnen Sie ferner die durchschnittlichen Verweildauern d_N und d_Z .



Aufgabe 12.2

Für die Bundesrepublik wurden die folgenden Zahlen zur Arbeitslosenstatistik ermittelt,⁵ für die folgendes zu berechnen ist:

- der Durchschnittsbestand (der Arbeitslosigkeit),
- die durchschnittliche Verweildauer (in der Arbeitslosigkeit) sowie
- die Umschlagshäufigkeit des Arbeitslosenbestandes!

•

⁵ Quelle: Kühl, J., 15 Jahre Massenarbeitslosigkeit, Aspekte einer Halbzeitbilanz, in: Aus Politik und Zeitgeschichte, Beilage zur Wochenzeitschrift "Das Parlament", 16.9.88

Jahr	Bestand	Zugang	darunter aus ET ^{*)}	Abgang
1982	1833244	3706655	2833726	3187165
1983	2258235	3704185	2698263	3578551
1984	2265559	3672791	2612203	3696594
1985	2304014	3750240	2658574	3728294
1986	2228004	3637266	2553496	3766214
1987	2228788	3726460	2606148	3636411

*) ET = Erwerbstätigkeit;

Aufgabe 12.3

Trotz intensivster ärztlicher Bemühungen haben sich die fünf ersten Exemplare einer neu gezüchteten Pferderasse als nicht sonderlich überlebensfähig herausgestellt. Das Pferd Egon konnte noch nicht einmal seinen ersten Geburtstag feiern. Das Alter (in vollendeten Jahren), das die fünf Pferde erreichten betrug leider nur:

Pferd	Alter
Egon	0
Doris	1
Boris	2
Clara	2
Augustus	3



Man bestimme aus diesen Daten alle Sterbetafel-funktionen l_x , d_x , T_x und e_x , sowie L_x , T_x^* und e_x^* ! Interpretieren Sie die Größe T_x als die Anzahl der von den l_x Tieren, die das Alter von x erreicht haben, insgesamt noch zu durchlebenden Jahre!

Aufgabe 12.4

Ein Lager werde zur Zeit t_0 mit vier Waren (A,...,D) gefüllt und der Lagerbestand soll während der ganzen Beobachtungszeit von $t_0 = 0$ bis $t_m = 8$ konstant 4 betragen:

- alle vier Waren haben die gleiche Verweildauer von 4 Perioden
- zwei der vier Waren (A, B) haben eine Verweildauer von 2 Perioden und zwei Waren (C, D) von 4 Perioden.

Man bestimme die den beiden Teilen zugrundeliegenden Verteilungen der Verweildauer sowie die Umschlagshäufigkeit des Lagers in beiden Fällen.

Aufgabe 12.5

Gegeben seien die folgenden vier Fälle von Bestandsveränderungen (jeweils eine geschlossene Masse):

	Fall 1		Fall 2		Fall 3		Fall 4	
Einheit	t_{Zi}	t_{Ai}	t_{Zi}	t_{Ai}	t_{Zi}	t_{Ai}	t_{Zi}	t_{Ai}
A	1	4	1	3	1	3	1	2
B	2	3	2	4	2	4	2	5
C	3	4	3	4	3	4	4	6
D	4	6	4	6	4	7	5	6
E	5	7	5	7	5	6	5	7

Wie unterscheiden sich diese Fälle hinsichtlich

- der zeitlichen Verteilung der Zu- und Abgänge,
- der Bestandsfunktion,
- der Verweildauerverteilung und
- des Becker'schen Diagramms?

Aufgabe 12.6

Gegeben seien die folgenden Absterbeordnungen l_x (fiktive Zahlen), die mit A, B und C bezeichnet werden sollen sowie eine weitgehend unbekannte Absterbeordnung D:

A		B		C		D	
x	l_x	x	l_x	x	l_x	x	l_x
0	50	0	50	0	50	0	50
1	40	1	45	1	32	1	?
2	30	2	38	2	24	2	?
3	20	3	27	3	12	3	?
4	10	4	19	4	3	4	?
5	0	5	0	5	0	5	0

Als lineare Absterbeordnung bezeichnet man den Fall, d.h. die Absterbeordnung (Richtiges ankreuzen):

☐ A

☐ B

☐ C

Die größte Lebenserwartung eines Nulljährigen e_0 erhält man bei:

☐ A

☐ B

☐ C

Die kleinste Lebenserwartung eines Nulljährigen e_0 erhält man bei:

☐ A

☐ B

☐ C

über die Absterbeordnung l_x im Fall D ist fast nichts bekannt (fast überall ein ? statt einer Zahl). Durch eine einfache Überlegung kann man aber feststellen, wie groß T_x für $x = 0$ also T_0 mindestens und höchstens sein kann (und damit auch e_0):

T_0 ist höchstens	T_0 ist mindestens
e_0 ist höchstens	e_0 ist mindestens

Von den Tafelfunktionen q_x , T_x und e_x lässt sich folgendes sagen [für beliebige Werte von x (Alter) bei einer Sterbetafel, wobei davon auszugehen ist, dass die Menschen zwischen 0 und ca. 100 Jahre alt werden und die Sterbetafel mit $l_0 = 100.000$ beginnt]:

- A die Funktion kann nicht zunehmen (sie ist monoton fallend)
- B sie kann nur Werte zwischen 0 und 1 annehmen
- C sie nimmt Werte zwischen 0 und ca. 100 an
- D sie nimmt Werte zwischen 0 und maximal 100.100000 an

Tragen Sie die Buchstaben (es kann auch mehr als ein Buchstabe richtig sein) in die freien Felder ein:

Beispiel: q_x : A, C

q_x	
T_x	
e_x	

Ende des Aufgabenteils

Lösungen der Übungsaufgaben Deskriptive Statistik

Kapitel 1

- 1.1 Bestandsmasse (stock S), Bewegungsmasse (flow F): Auftrag (F), Anlagev. (S), Sozialp. (F), Nettoinv. (F), landw. Nutzfl. (S), Ehescheid. (F), Steuerein. (F), Sparv.(S), Schulden (S), Neuversch. (F), Baufertigst. (F), Gutschriften (F), Geburten (F).
- 1.2 qualitativ, häufbar, intensiv, latent.
- 1.3 1. drei: Art des Gerichts (Pizza, Sonstige), Essenszeit (M,A), Nr. des Lokals (L1,L2)
2. a) Einheit
b) Merkmalsausprägung (auch e,h)
c) Identifikationsmerkmal bzw. -ausprägung (auch h)
d) Umfang der Masse (Teilmassen f,g).
- 1.4 Angemessen ist Kreisdiagramm; Winkel: G:180°, H:60°, N:120°.
- 1.5 lineare Transformation, Intervallskala
- 1.6 N,O,I,R
- 1.7 a) sinnvoll: Modus, Kreisdiagramm
b) Z = diskret, metrisch, A = nominalskaliert.
- 1.8 Alter: Ratioskala $\bar{x} = 20$
Bild.: Nominalskala, Modus (aber hier alle drei Ausprägungen gleich häufig)
Emotion: Nominalskala, Modus = A
Bildungsstand (2), mittl. Alter (5), Alter 25 (3), Person F(4)
Daten (1), Bildung gering (3) .

Kapitel 2

- 2.1 Klassifikation ist unsinnig, weil die systematischen Positionen nicht erschöpfend und nicht sich gegenseitig ausschließend sind.
- 2.2 Alter und Wochenarbeitszeit Ratioskala; Rest Nominalskalen.
- 2.3 R,N,I,N,N,N,N,N,R,R,N,N(O),N .
- 2.4 Welche Einheit im Rahmen der Quote ausgewählt wird, bleibt dem Interviewer überlassen; Auswahlfehler ist kein Zufallsfehler (keine Anwendung der Wahrscheinlichkeitsrechnung!).
- 2.5 Anzahl der Beschäftigten, E, G und K (4 Merkmale) gegliedert nach Zweigniederlassung. Ausprägungen sind Geldbeträge bzw. Anzahl der Personen, Summenvariablen G, K; Häufigkeiten sind n_A, n_B . Da Gliederung nach nominalskaliertem Merkmal (Ausprägung A,B), keine Häufigkeitsverteilung!); $\bar{x} = 1628,57$.

Kapitel 3

- 3.1 Rangskala der Güteklassen 0,1,...

x_i	0	1	2	3	4
h_i	0,12	0,28	0,32	0,2	0,08
H_i	0,12	0,4	0,72	0,92	1

Zähleinheit: Übernachtung; Merkmale: Dauer des Aufenthalts nach Art des Hotels.

3.2 Winkel für das Kreisdiagramm: Privater Verbrauch 220,86°, Staatsverbrauch 55,87°, Investitionen 71,97°, Außenbeitrag 11,3°

3.3

x_i	n_i	N_i	$h_i \cdot 100$	$H_i \cdot 100$
2	2	2	6,67%	6,67%
3	4	6	13,33%	20,00%
4	7	13	23,33%	43,33%
5	6	19	20,00%	63,33%
6	4	23	13,33%	76,67%
7	3	26	10,00%	86,67%
8	2	28	6,67%	93,33%
9	1	29	3,33%	96,67%
10	1	30	3,33%	100%

f) klassierte Verteilung

	n_i	N_i	$h_i \cdot 100$	$H_i \cdot 100$
$0 < x \leq 4$	13	13	43,33	43,33
$4 < x \leq 7$	13	26	43,33	86,67
$7 < x \leq 10$	4	30	13,33	100

3.4

Fahrzeit	n_i	b_i	n_i/b_i	H_i
$0 \leq x < 5$	100	5	20	0,1
$5 \leq x < 10$	400	5	80	0,5
$10 \leq x < 20$	300	10	30	0,8
$20 \leq x < 60$	150	40	3,75	0,95
$60 \leq x < 120$	50	60	0,83	1

Die Blöcke für das Histogramm müssen eine Höhe haben, die proportional zu den Werten n_i/b_i (Häufigkeit je Klassenbreite) sind. Da diese Werte sehr unterschiedlich sind, empfiehlt es sich, die letzte Klasse als beispielsweise $20 \leq x < 60$ zu definieren (obgleich x tatsächlich größer ist, sogar über 60), damit das Bild noch vernünftig darstellbar ist. Dann wäre $h_4/b_4 = 200/40 = 5$, was als Höhe ohnehin wenig ist gegenüber Werten wie z.B. 80. Man könnte außerdem die ersten beiden Klassen zusammenfassen. Man erhält dann $n/b = 50$. Die letzte Klasse muss willkürlich geschlossen werden (hier zunächst mit 2 Stunden!). Die Abgrenzung der Intervalle für die Summenhäufigkeitskurve H_i ist eigentlich $a < x \leq b$ und nicht wie hier $a \leq x < b$. Die Höhen H_i sind obiger Tab. zu entnehmen. Hier ist nicht durch die Klassenbreiten zu dividieren!

3.5

x_i	n_i	h_i	N_i	H_i
0	6	0,3	6	0,3
1	5	0,25	11	0,55
2	5	0,25	16	0,8
3	3	0,15	19	0,95
4	1	0,05	20	1,0

Frauen (h_i, N_i, H_i entsprechend)

y_i	n_i
0	11
1	8
2	20

- 3.6 Da $\sum n_i = n = 100$, sind die angegebenen absoluten Häufigkeiten zugleich die prozentualen Häufigkeiten. Ein Balkendiagramm ist leicht zu zeichnen.
 $\bar{x} = 3$, $s_x^2 = 1,1$, Verteilung ist linkssteil $SK_M = 1,0401$
- 3.7 a) nur "stetig" ist falsch (also R, F, R, R); Merkmal: Anzahl der Tore.
 c) Approximative Lösung:

$$h(x \leq 7) = h(x < 6) + \frac{1}{4} h(6 \leq x < 10) = 0,875 + \frac{1}{4} \cdot 0,125 = 0,90625$$

$$h(x > 4) = h(6 \leq x < 10) + \frac{1}{2} h(4 \leq x < 6) = 0,125 + \frac{1}{2} \cdot \frac{1}{3} = 0,291\bar{6}$$
- e) für Teil c erhält man jetzt $h(x \leq 7) = 22/24$, $h(x > 4) = 7/24$. Funktion $H(x)$ für jedes angegebene x zu bestimmen.
 $\bar{x} = 3,458$, $\tilde{x}_{0,5} = 3$ (keine Interpolation innerhalb einer Klasse nötig!).

Kapitel 4

- 4.1 $\bar{x}$ ist stets 100, obgleich die Gestalt der Verteilung sehr unterschiedlich ist.
- 4.2 a) $\bar{x} = 9$, $\hat{\bar{x}} = 9,5$
 b) ungewogene (Einzelwerte) und gewogene Berechnung führen zu $\bar{x} = 6$.
- 4.3 Man sollte erkennen und interpretieren: weil die Ausgabe (Menge·Preis) konstant (jeweils 1DM) ist, gilt: ungewogenes harmon. Mittel der Preise ($1/3$ und $1/5$ DM) = gewogenes (mit Mengen 3 bzw. 5) arithmet. Mittel der Preise = 0,25.
- 4.4 Linearkombination: DM 2430,-.
- 4.5 $\bar{x} = 31,6$; $\tilde{x}_{0,5} = 20$ Summe der negativen und der positiven Abweichungen von $\bar{x}$ ist gleich (nämlich 290,8).
- 4.6 Es sollte jeweils Null das Ergebnis sein (Schwerpunkteigenschaft von $\bar{x}$).
- 4.7 Wahres Mittel $\bar{x} = 718.517$ DM (aufgrund der wahren Klassenmittelwerte bzw. der Gesamtangaben 479.437 und 667.259). Geschätzter Mittelwert mit den geschätzten Klassenmitteln (50.000, 150.000 usw., letzte Klasse [willkürlich] 6 Mill.) ergibt $\hat{\bar{x}} = 706.034$ DM. Graphik optisch nicht schön, da sehr unterschiedlich breite Klassen!
- 4.8 Geometr. Mittel der Wachstumsfaktoren: 1,11119; also 11,12% ist die durchschnittl. Wachstumsrate, nicht $(10+20+10+5)/4 = 11,25\%$ (das wäre zu groß!).

- 4.9 $\bar{x} = 87,5$ (sinnvoll und üblich); $\bar{x}_G = 70$ (leicht zu interpretieren: Ist $x_0 = 4x_A$, so sollte das Gewicht einer mittleren Person $2 \cdot x_A$ sein). Allgemein gilt $\bar{x}_G \leq \bar{x}$ ($\bar{x}_G = \bar{x}$ wenn $x_A = x_0$).
- 4.10 $\bar{x} = 200$, bei 0,185 Std. bedeutet dies 37 statt 36 km; richtig ist $\bar{x}_H = 194,59$ km/h als Durchschnittsgeschwindigkeit (dann 36 km bei 0,185 Std.). Ungewogene Mittel weil die drei Strecken gleich lang sind.
- 4.11 Flächenleistungen sind $1/3$ und $1/5$ Fläche pro Stunde. Das ungewogene harmonische Mittel aus 3 und 5 ist 3,75 also $1/3,75 = 0,267$ Fläche pro Stunde. Die Durchschnittsleistung von $1/4$ Fläche ist zu gering ($0,25 < 0,267$).
- 4.12 90 km/h ist falsch, weil die Durchschnittsgeschwindigkeit nicht $\bar{x} = \frac{1}{2}(90+30) = 60$ ist, sondern das harmonische Mittel. Mit 30 km/h braucht R für 4 km genau 8 Minuten. Mit 60 km/h müsste er hin und zurück in 8 Minuten gelangt sein. Die mittlere Geschwindigkeit bei 30 und 90 km/h ist nicht 60 km/h, sondern 45 km/h.
- 4.13 a) 12, b) $(12+13)/2 = 12,5$.
- 4.14 Median: vorher und nachher 53; $\bar{x}$ vorher: 54,2 nachher: 48,8.
- 4.15

	1. Quartil	Median	3. Quartil	arith. Mittel
Männer	37,12	38,88	40,64	38,84
Frauen	29,05	37,45	39,56	34,16
gesamt	36,45	38,35	40,25	36,94

- 4.16 arithmet. Mittel 0% offensichtl. unsinnig. Gehalt ist geringer geworden. Mittlere Wachstumsrate -2,02%.
- 4.17 a) 15%, b) 65%, c) 50, d) Ordinalskala.
- 4.18 $Q_1 = 25, Q_2 = Z = \tilde{x}_{0,5} = 45, Q_3 = 75$.
- 4.19 Vgl. Aufg. 4.7. Bei gleicher Klasseneinteilung und gleichen Klassenmitten erhält man $\hat{\bar{x}}(1977) = 675.598$ und $\hat{\bar{x}}(1983) = 706.034$. Die wahren Klassenmitten sind durch Angabe der Gesamtvermögen zu bestimmen. Man erhält $\bar{x}(1977) = 634.107$ und $\bar{x}(1983) = 718.517$.
- 4.20 a) $\bar{x} = x$ so dass $\bar{x}^4 = x^4$. Dagegen ist $(\bar{x}_G)^4 = x^2(x^2 - D^2)$.
b) $s_x^2 = D^2/2$
c) die Durchschnittsgeschwindigkeit H (harmon. Mittel) kann nicht größer als 90 sein, weil man die folgende quadratische Gleichung erhält $D^2 = 90(90-H)$. So gilt etwa $H = 80 \rightarrow D = 30$ oder $H = 87,5 \rightarrow D = 15$.
- 4.21 a) alle Mittelwerte, die Spezialfall des Potenzmittels sind, etwa $\bar{x}_H = 14,1176, \bar{x}_G = 14,5648, \bar{x} = 15, x_Q = 15,41$.
b) vgl. Gl. 4.23
c) $n(n-1)/2 = 6$, so dass nach Gl. 5.39 gilt $S_G = 30/6 = 5$.
- 4.22 Begriffe "monatl. Mittel" und "heißester Monat" sind nicht eindeutig. Es gilt: tägl. Messungen $i=1, \dots, 48$; Tage des Monats $k=1, \dots, n_j$; Monate $j=1, \dots, 12$; heißester Monat habe Subskript m. Mit
- $$T_m = \max_j \frac{1}{48n_j} \sum_j \sum_k x_{jk} \quad (\text{monatl. Mittel...}) \quad \text{und} \quad M_m = \frac{1}{n_m} \sum_k \max_i x_{ik} \quad (\text{Mittel der}$$
- tägl. Maximaltemp.) gilt $ART = T_m + (T_m - M_m)/3$.
- 4.23 Dem Zitat liegt eine Verwechslung von arithmet. Mittel $\bar{x}(1835)$ und Median (elf Millionen, also 50%) zugrunde. Bei Linkssteilheit, die üblicherweise gegeben ist, ist $\tilde{x}_{0,5} < \bar{x}$, bei den Angaben gem. Gl. 5.64 1300. Das hieße, die Hälfte hat nicht weniger als 1800, sondern weniger als 1300 DM verdient. Ehrenberg hätte besser (in

seinem Sinne) argumentieren können, wenn er nicht $\bar{x}$ als $\tilde{x}_{0,5}$ interpretiert hätte.

Kapitel 5

- 5.1 $\bar{Q}_{0,25}$ ändert sich nicht (2,5); R (Spannweite) steigt von 9 auf 69; s^2 steigt von 7,82 auf 370,2.
- 5.2 d um Z (d_x) früher 2, jetzt 3, d um $\bar{x}$ (d_x^*) früher 1,25, jetzt 5.
- 5.3 $220/6 = 36,67$.
- 5.4 a) nein (bei Berechnung aus Einzelwerten) $d_z = d_x \leq d_{\bar{x}} = d_x^*$.
b) Bei Schüler B und C: $\tilde{x}_{0,5} = Z = 4$ gleich, d_x jeweils 1 aber d_x^* bei B: 1,111 bei C: 1,333 ($s_B^2 = 1,55$, $s_C^2 = 2$).
- 5.5 $s_B^2 = 4$; $s_B = 2$; $s_C^2 = 2$; $s_C = 1,4142$.
- 5.6 a) $s_x^2 = 54,85 - (6,4^2) = 13,89$
b) nur externe Varianz 12,54.
c) gleiche Daten bei b: dann innere Varianzen $s_1^2 = 11/4$, $s_2^2 = 2/3$, $s_3^2 = 1/6$; interne Varianz: 1,35, Gesamtvarianz $s^2 = 12,54 + 1,35 = 13,89$.
- 5.7 $\bar{x} = 600$, $\bar{y} = 600$; Lineartransformation, b ist richtig.
- 5.8 $s_{\text{int}}^2 = 1,5$, $s_{\text{ext}}^2 = 12,54$, $s_{\text{ges}}^2 = 14,04$.
- 5.9 $\bar{x} = 880$, $s_x = 200$, Lineartransformation $y = 50 + 1,2x$, also $\bar{y} = 1106$, $s_y = 240$.
- 5.10

	A	B	C	D
Schiefe	+0,694	0	-0,694	0
Wölbung	-0,222	-0,5	-0,222	+0,935

- 5.11 Anfangsmomente $m_1 = 2$, $m_2 = 4,5$, $m_3 = 11$; zentrale Momente $z_1 = z_3 = 0$, $z_2 = 1/2$.
- 5.12 Konstruktionsprinzip 1; erfüllt alle Axiome; 4J,6J, alle anderen Antworten N.
- 5.13 a) immer Null (Schwerpunkt); b) $n = 1$ oder alle $x_v = \bar{x}$; c) F,R,F (Zentrum: $\tilde{x}_{0,5}$), F.
- 5.14 $\bar{x} = 4$, $s^2 = 8$, $z_3 = 40$, Schiefe 1,7678; linkssteil, diskret, rechtsschief, asymmetrisch
- 5.15 $\bar{x} = 10200/200 = 51$, Median (Zentralwert) $Q_2 = Q_1 = 0$, $Q_3 = 1$, also $\bar{Q}_{0,25} = 1/2$, in Klasse 4 457 Einw. im Mittel.
- 5.16 a) Mittelwerte: 2,7,12; Varianzen 2/3, 1,5, 2
b) $\bar{x} = 7$ (Gl. 4.9), $s^2 = 1,4 + 15 = 16,4$ (Gl. 5.11)
c) $s_{\text{ext}}^2 = 15$.
- 5.17 $\bar{x} = 1.500$, $s^2 = 220.000$.
- 5.18 $\bar{x} = 3,1715$; Median (ohne Interpolation) und Modus: 2,95; $s = 0,3894$, $V = 0,1228$, linkssteil (positive Schiefe).
- 5.19 a) Maßeinheiten: mm bei Mittelwert, Modus, Standardabweichung, $(\text{mm})^2$ bei Varianz, Variationskoeffizienten: I: 0,414, II: 0,083.
b) Verteilung I ist symmetrisch und streut erheblich mehr als die leicht linkssteile Verteilung II bei etwa gleichem Mittelwert.
c) beide 11,6.
- 5.20 1. nein, $\bar{x}$ jeweils 15
2. nein, s^2 jeweils 36
3. Schiefe vorher +0,694 (A), nachher -0,694 (C).

Kapitel 6

6.1

H_i	0,4	0,7	0,9	1,0
Q_i	0,2	0,5	0,8	1,0

$$D_G = 0,27$$

6.2 von 0 auf 1/6 (in beiden Fällen).

6.3 die ersten beiden Klassen zusammenfassen und $\bar{x} = 0$ setzen.

a) null, 2

b)

H_i	0,0667	0,5	0,9667	1
Q_i	0	0,054	0,6375	1

6.4 Disparität in A und B gleich: $D_G = 0,6$.

6.5

H_i	0,6	0,9	1
Q_i	0	0,5	1

$$D_G = 0,7$$

6.6

H_i	0,5	1	
Q_i	1/3	1	

$$D_G = 1/6$$

6.7 a) $L_1 = 450$, $L_2 = 900 = \bar{L}$, $L_3 = 1800$ b) $s_L^2 = 243.000$; $V = \sqrt{0,3}$ c) $D_G = 0,28$

d) 0,4 .

6.8 Städte unabhängig von der Einwohnerzahl nach Anzahl der Vampire ordnen:

	Q_i
L	0,08
übr. S.	0,2
DD	0,36
BZ	0,6
GR	1

$$D_G = 0,32$$

6.9

H_i	1/3	2/3	1
vorher Q_i	3/12	7/12	1
nachher Q_i	4/15	9/15	1

$$D_G = 1/9 \text{ (vorher)}$$

$$D_G = 4/45 \text{ (nachher)}$$

6.10

H_i	0,5	0,75	0,9	1
Q_i	0,25	0,5	0,75	1
Steig.	0,5	1	1,667	2,5

$D_G = 0,325$; Anteil: 0,5; nicht schneiden, tangieren (sie kann mit der Gleichverteilungsgerade identisch sein)

6.11

Klasse	1	2	3	4
H_i	0,6	0,8	0,9	1
Q_i	0	0,004	0,11	1

$$D_G = 0,8768$$

b) Konzentration verringert sich.

c) darf man nicht .

d)

H_i	0,8	0,9	1
Q_i	0	0,106	1

$$D_G = 0,8788$$

- 6.12 Danach müsste die durchschnittl. Anzahl der Beschäftigten bei z.B. einer Gesamtzahl 10.000 Beschäftigten und 100 Betrieben sein:

Klein: 40, Mittel: 300, Groß: 66,7

Die Daten sind also nicht richtig geordnet!

- 6.13 Bei $\bar{x} = 3000$ ist die Steigung der Lorenzkurve 1 wegen $s_i = \frac{x_i}{\bar{x}}$; beim Median Z

ist Steigung kleiner (0,8), also $Z < \bar{x}$ (linkssteil!).

- 6.14 aus $q_i = s_i h_i$ folgt

	E_1	E_2	E_3
q_i	0,2	0,3	0,5

wegen $\bar{E} = 2000$ gilt z.B. bei 100 Beschäftigten für die Gesamteinkommen der drei Gruppen: 40.000, 60.000, 100.000 und damit für die Durchschnitte in den Gruppen: $E_1 = 40.000/50 = 800$, $E_2 = 1.500$, $E_3 = 10.000$.

- 6.15 a) Eigentlich nur für W, Merkmale I und M nicht metrisch skaliert und nicht extensiv

b)

H_i	0,5	0,7	0,8	1,0
Q_i	0,11	0,22	0,35	1,0

c) 1. Statistiker bleibt in der gleichen Klasse; Flötenerfinder verringert Anzahl der Bewohner in Höhlenklasse 1 und damit auch q_1 zu Gunsten von q_4 ; Disparität wird größer (Konzentration der Bewohner auf größere Höhlen).

2. In Höhlenklasse 1 vorher ca. 150 Personen, jetzt nur noch 50; neue Höhlenklasse mit 100 Personen (über 70!). Disparität wird größer!

d) Einzelangaben sortieren!

Höhle	H_2, H_4	H_1	H_3, H_5
n_i	2	1	2
h_i	0,4	0,2	0,4
q_i	0,2	0,2	0,6

also Lorenzkurve:

H_i	0,4	0,6	1
Q_i	0,2	0,4	1

- 6.16 a) $\tilde{x}_{0,5} = 1,5$ $\bar{x} = 1,5$ b) symmetrisch c) 1,5
d)

H_i	0,2	0,5	0,8	1,0
Q_i	0	0,2	0,6	1,0

$$D_G = 0,38$$

- e) Varianz und Disparität nie negativ! Richtige Antwort ist "positiv", $D_G > 0$.

6.17 $D_G = 0,26$.

- 6.18 Durch Zusammenfassung von Klassen rückt der Polygonzug der "Lorenzkurve" näher an die Gleichverteilungsgerade heran, D_G sinkt.

6.19 a) $\tilde{x}_{0,5} = 250$, $\bar{x} = 400$

- b) extensiv, diskret, metrisch; linkssteil ($\tilde{x}_{0,5} < \bar{x}$, aber Modus: 800!), eindimensional.

c)

H_i	0,2	0,4	0,6	1,0
Q_i	0,025	0,075	0,2	1,0

- 6.20 Steigung = q_i/h_i ; aus n_i erhält man h_i und daraus q_i und Q_i . Lorenzkurve

H_i	0,4	0,6	0,8	1,0
Q_i	0,1	0,3	0,6	1,0

$$D_G = 0,38$$

$q_4 = 0,4$, also erhält die Klasse der besten Spieler zusammen 4 Mill.DM (im Durschn. 1Mill.DM).

- 6.21 a) Beziehungs-, Verhältniszahl

- b) 1. Kundenzahl

2. $h_1 = h_2 = h_3 = h_4 = 0,25$; $q_1 = 0,05$, $q_2 = 0,15$, $q_3 = 0,35$, $q_4 = 0,45$

c)

	C	A	D	B
H_i	0,25	0,5	0,75	1,0
Q_i	0,05	0,2	0,55	1,0

$$D_G = 0,35$$

6.22

H_i	0,2	0,5	0,8	1,0
Q_i	0	0,2	0,6	1,0

$$D_G = 0,38; \text{ Lineartransformation! } \bar{x} = \frac{2}{3}1,5 = 1; \quad s^2 = \frac{4}{9}1,05 = 0,467.$$

6.23 $\bar{y} = 500 = \bar{x}$, $s_y^2 = 56.250 = (1,5)^2 s_x^2$

H_i	0,25	0,5	0,75	1	D_G
Q_i vor Steuer	0,15	0,35	0,65	1	0,175
Q_i nach Steuer	0,1	0,275	0,6	1	0,2625

- 6.24 Aus den Angaben folgt, dass alle $q_i = 1/4$ sind. Es ließen sich auch leicht die mittleren Einkommen je Klasse $\bar{x}_i$ ($i=1,\dots,4$) berechnen. Steigungen (s_i)

h_i	0,5	0,25	0,15	0,1
-------	-----	------	------	-----

$s_i=(1/4)/h_i$	0,5	1	1,67	2,5
-----------------	-----	---	------	-----

$\bar{x} = 1.600 = \bar{x}_2$ (Man verifiziert leicht, dass $\bar{x}_i/\bar{x} = s_i$).

6.25

H_i	0,64	0,92	1
Q_i	0,164	0,48	1

$D_G = 0,5963$

6.26 $D_G = 1 - [h_q - (1-h)(1+q)] = h-q$; $s_x^2 = h(1-h)(x_1-x_0)^2$ da gilt $q = h x_0/\bar{x}$ und $1-q = (1-h) x_1/\bar{x}$ erhält man $(x_1-x_0)^2 = \bar{x}^2[(h-q)/h(1-h)]^2$ und $V^2 = s_x^2/\bar{x}^2 = D_G^2/h(1-h)$. Die Ungleichung gilt, weil $h(1-h)$ maximal 1/4 sein kann.

6.27 Da die Steigung monoton steigend ist und bei $x = \bar{x}$ genau 1 ist, muss das Einkommen $x < \bar{x}$ sein, wenn die Steigung 0,8 ist. Der Punkt der Lorenzkurve, der dem Medianeinkommen zugeordnet ist, muss also links vom Punkt liegen, der $\bar{x}$ zugeordnet ist. Also ist $\tilde{x}_{0,5} < \bar{x}$ und die Verteilung von x linkssteil.

Kapitel 7

7.1 Zusammenhang erkennbar an geschlechtsspezifischen "Durchfallquoten" (bedingte Mittelwerte). Bestimmung der Randverteilungen elementar (wird auch aus Platzgründen bei den anderen Aufgaben nicht angegeben).

7.2 bed. Mittelwert von

y	2,75	2,5	3,5	4	3,5
x	1,5	2,714	2,727	3,1	4,167

7.3 $S_{vw} = 0$, $s_v^2 = 4,8$, $s_w^2 = 40.000$.

7.4 $\bar{y}|x: 1,5 \quad 2 \quad 2,5$; $\bar{x}|y: 8 \quad 56/5 \quad 40/3$; nichtlinear, positiv, $s_{xy} = 1,28$; $\bar{x}_H = 2/(1/6+1/10) = 7,5$.

7.5 a) identisch b) $r_{xy} = 5/6$ c) Scheinkorrelation d) 4

7.6 $\bar{y} = 3$, $s_y^2 = 1,5$, $\bar{x} = 2,5$, $s_x^2 = 0,75$, $\hat{x} = 1 + 0,5y$, $\hat{y} = 0,5 + x$ (identisch mit Regressionslinien), $s_{xy} = 3/4$, $r_{xy} = \sqrt{1/2}$.

7.7 $\hat{y} = 3,625 - 0,375x$, $\hat{x} = 2,667 - 0,167y$ (identisch mit Regressionslinien); $s_{xy} = -0,06$, $r_{xy} = -1/4$.

7.8 $\bar{x} = 52$, $\bar{y} = 46$, $s_x^2 = 2$, $s_y^2 = 20,96$, $S_{xy} = 4$, $r_{xy} = 0,6178$,
 $\hat{y} = 43,22 + 0,19x$, $\hat{x} = -58 + 2y$

Regressionslinie $\bar{y}|x$ ist eine Gerade (42, 44, 46,...) aber nicht identisch mit der Regressionsgeraden, zweite Regressionslinie keine Gerade.

7.9 Offenbar Unabhängigkeit (eine Spalte ist ein Vielfaches einer anderen!), $\bar{y}|x = \bar{y}$ für alle Werte von x und $\bar{x}|y$ entsprechend konstant $\bar{x}$. Also $r_{xy} = 0$.

Kapitel 8

- 8.1 $\Sigma D = 400$, $\bar{D} = 40$ Kovarianz $s_{DM} = 0$
 $\Sigma M = 300$, $\bar{M} = 30$ damit auch Korrelation $r_{DM} = 0$
 $\Sigma DM = 12000$
 Variante: $\bar{D} = 28$, $\bar{M} = 30$, $s_{DM} = -68$, $r_{DM} = -0,5718$
- 8.2 alle Antworten falsch.
- 8.3 a) $\frac{1}{2}(16,4)^2$
 b) $\hat{y} = 50 + \frac{1}{2}x$
 c) wenn $x = 75$ dann $\hat{y} = 87,5 > x$, wenn $x = 130$ dann $\hat{y} = 115 < x$.
 d) nein; deterministische Interpretation einer stochastischen Beziehung; r^2 ist auch nur 25%.
- 8.4 a) $\hat{y} = 70 + 10x$
 b) $\hat{y} = 135$
 c) $r = 0,933$
 d) r ändert sich nicht, wohl aber a und b.
- 8.5 a) Regressionsgeraden

$$\left. \begin{aligned} \hat{x}_B &= 18,046 - 0,7673 \cdot x_C \\ \hat{x}_C &= 11,473 - 0,3942 \cdot x_B \end{aligned} \right\} r_{BC} = -0,5497$$
- b)
- | $\bar{x}_B$ (Bernd) | |
|---------------------|-------------------|
| x_C | $\bar{x}_B x_C$ |
| 3 | 15 |
| 6 | 14 |
| 9 | 15 |
| 12 | 6 |
- | $\bar{x}_C$ (Charlie) | |
|-----------------------|-------------------|
| x_B | $\bar{x}_C x_B$ |
| 3 | 12 |
| 9 | 7,5 |
| 12 | 6 |
| 18 | 4,8 |
- 8.6 a) Schnittpunkt der beiden Regressionsgeraden
 b) geringe positive Korrelation $r = +0,2236$.
- 8.7 a) $\hat{y} = 35 - x$ b) bei 35° : $\hat{y} = 0$ Minuten; bei -30° : 65 Minuten(!),
 c) Variante: $\hat{y} = 20 - \frac{1}{2}x$; alle Punkte liegen genau auf der Regressionsgerade ($r_{xy} = -1$).
- 8.8 a) $\hat{y} = 13,2 - 1,8x$
 b) $r^2 = 0,95294$ ($s_y^2 = 6,48$ das ist 95,3% von $s_y^2 = 6,8$).
- 8.9 ohne UT: $\hat{y} = \bar{y} = 3$ und $\hat{x} = \bar{x} = 5$, $r_{xy} = 0$
 mit UT: $\hat{y} = -16,8 + 4,4x$, $r_{xy} = 0,6286$.
- 8.10 a) $\hat{y} = -2 + 0,6x$
- | Tag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----|-----|------|------|-----|---|----|------|-----|------|-----|
| u | 1,8 | -2,6 | -0,2 | 2,2 | 0 | -1 | -2,6 | 2,2 | -0,4 | 0,6 |
- b) $\Sigma u = 0$, $\Sigma u^2 = 28$, $s_u^2 = 28/10 = 2,8$; $s_y^2 = 10 = s_y^2 + s_u^2 = 7,2 + 2,8$, $r_{xy}^2 = 0,72$,
 $1 - r_{xy}^2 = 0,28$ erklärter Anteil ist 72%.
- c) R,F,F,F,F,F,R,R,F,F
 d) nicht sinnvoll (nicht kausal zu interpretieren!).

- 8.11 a) Residualvarianz $s_u^2 = 0,76$, erklärte Varianz $s_{\hat{y}}^2 = 3,24$, Korrelation $r_{xy} = -0,9$
 b) $s_x^2 = 9$
 c) $b = -0,6$.
- 8.12 Typisches Beispiel für Zeitreihenkorrelation: r^2 ist hoch weil Geldmenge und Sozialprodukt einen gemeinsamen [ansteigenden] Trend haben; wird dieser durch Bildung von Zuwächsen weitgehend eliminiert, dann verringert sich auch die Korrelation.
- 8.13 $r = 0,96056$, $r^2 = 0,9267$, ($\hat{y} = 2,144 + 1,539x$).
- 8.14 a) nein $\hat{x}$ ist eine Parallele von $\hat{y}$ (also nicht einfach $r = \sqrt{3 \cdot \frac{1}{3}} = 1$ rechnen!)
 b) ja, $r^2 = 1$ (die Geraden "fallen zusammen").
- 8.15 a) $\hat{y} = 2x$ wenn $x = 5$ dann $\hat{y} = 10$.
 b) wenn $x = 4$, dann $\hat{y} = 8$, $y = 9$ also $u = y - \hat{y} = 1$ (Zufallsabw., Residualabw.), Gesamtabweichung $y - \bar{y} = 9 - 5$ (erklärt davon 4, nicht erklärt 1).
 c) $s_y^2 = 2$; erklärt $s_{\hat{y}}^2 = 0,9^2 \cdot 2 = 1,62$; $s_x^2 = 0,4$; $s_{xy} = \frac{1}{n} \sum xy - \bar{x} \cdot \bar{y} = \frac{44}{5} - 2 \cdot 4 = 0,8$ demnach ist r^2 eigentl. 0,8 nicht 0,81, wie in der Aufgabe angegeben; aber mit 0,81 lässt sich leichter rechnen (Quadratzahl!). Mit den vorher ermittelten Ergebnissen $s_y^2 = 1,62$, $s_{xy}^2 = r^2 s_x^2 s_y^2 = 0,81 \cdot 0,4 \cdot 2 = 0,648 \approx 0,64 \rightarrow \sqrt{0,64} = 0,8$.
- 8.16 $\sqrt{bd} = 6 - \frac{2}{3}y$, also muss auch die Steigung b der anderen Regressionsgeraden negativ sein. Außerdem muss $r = \sqrt{bd}$ zwischen -1 und 0 liegen. Möglich ist dann
 $-1,5 \rightarrow r = -1$
 $-0,3 \rightarrow r = -\sqrt{0,2}$
 $-0,1 \rightarrow r = -\sqrt{1/15}$;
 die anderen Werte scheiden aus (etwa bei $b = -2$ erhielte man $\sqrt{bd} = \sqrt{1,33}$).
- 8.17 Man kann zeigen, dass die Steigung der Regressionsgeraden $\hat{x}$ betragsmäßig größer
 8.18 sein muss als die Steigung von $\hat{y}$. Also gilt L und bei 8.44 entsprechend auch L (links).
- 8.19 1. $\hat{y} = 50.000 + 5x$
 2. 105.000
 3. Gesamt: 20.000; erklärt und nicht erklärt jeweils 10.000
 4. erklärte Varianz 2.025 Mill.DM², $r^2 = 0,81$, $1-r^2=0,19$.
- 8.20 b) $x = \text{Fläche}$, $y = \text{Miete}$; $\bar{x} = 59,33$, $s_x^2 = 772,89$, $\bar{y} = 323,33$, $s_y^2 = 48122,22$
 d) Kausal interpretierbar nur $\bar{y}|x$: 166,67 254,84 642,86
 e) $s_{xy} = 4.548,89$, $r_{xy} = 0,74589$
 f) $\hat{y} = -25,877 + 5,8856x$.

Kapitel 9

- 9.1 Quoten: Beschäftigte 0,75 und 0,25, Lohnsumme 0,8; 0,2.
 9.2 a) Nein, es sind die Schwankungen eines Jahres; Schluss ist schon deshalb unzulässig, weil die Verweildauer in der Ehe nicht ein Jahr ist; auch Unterschied zwischen Quer- und Längsschnittanalyse.
 b) Analoge Kennzahl wie "Reichweite des Auftragsbestands" (Bestandsmasse / Abgangsmasse), daher auch ähnliche Interpretation denkbar.
 9.3 a) Messziffern

A	100	109,1	120	130
B	100	110	120	135
insges.	100	109,7	120	133,3

b) weil die Messziffer der Gesamtumsätze ein gewogenes Mittel der Messziffern für A und B ist (Gewichte: Umsatzanteil zur Basiszeit also 1/3 und 2/3).

9.4 Demonstrationsbeispiel für Scheinkorrelation. Anteil der "Verunfallten": Männer 44%, Frauen 32%. In den Teilgesamtheiten häufiges und seltenes Fahren sind die Quoten aber gleich.

9.5 Möglich, weil Kapitalstruktur von X und Y verschieden sind. Bei X ist 20% des Kapitals in A und 80% in B investiert und bei Y genau umgekehrt.

9.6 Strukturabhängigkeit (Altersstruktur) der rohen Todesrate, die ein gewogenes Mittel der altersspezifischen Todesraten (0,1 und 0,6) ist. $\bar{x}$ ist 70 und 30.

9.7 26,8% (statt 24%) und 791,6% (statt 240%).

9.8 5,56%, Verdoppelung nach 12,8 Jahren.

9.9 $r_s = \frac{1}{2\sqrt{t}}$ und $r_p = \frac{4}{2t-1}$

r_s : laufend abnehmen, nur positiv, gegen Null streben

r_p : R,F,R (wenn $t \geq 17$).

9.10 Wachstum mit konstanter Wachstumsrate (Dracula): $w_t = w = 3$ also $700.000 = 3^t \rightarrow t = 12,25$. Bei 60 Mill. ist $t = 16,3$ Monate .

9.11 1) 74,11%, 2) 200%, 3) 2600%, 4) 1458 (= $2 \cdot 3^6$).

9.12 a)

	1990	1991
1	0,08	0,07
2	0,05	0,04
	0,06	0,0625

Zunahme wegen veränderter Kapitalstruktur (mehr Kapital im rentableren Bereich).

b) 13,64 statt 15%, 25 statt 30%

c) arithmetisch, Kapitalanteile ; harmonisch, Gewinnanteile

d) Gewinn 500, Rent. 0,125

e) R F

R R

F F

F R

f) Für das Risiko evtl. zu sterben, ist wohl die Länge und Dauer des Transports weniger bedeutsam als die Häufigkeit, so dass für die meisten Menschen die erste und die letzte Beziehung relevant sein dürfte.

9.13 a) Maximum von $F(t)$ bei $t=4$, dann $r_F(4) = 0$ c) $r_F(t) = (480t^2 - 120t^3)/F(t)$.

9.14 a) $L(t)/6,75$

b) $(2t-12)/(108+t^2-12t)$

c) $84/1836 \approx 0,0458$.

9.15
$$r(t) = \frac{x'(t)}{x(t)} = \frac{2t + 5 \cdot \cos(t) - 60 \cdot \sin(30t) + 6 \cdot \cos(2t) + 5 \cdot \cos(t)}{t^2 + 5 \cdot \sin(t) + 2 \cdot \cos(30t) + 3 \cdot \sin(2t) + 5 \cdot \sin(t)}$$

9.16 Verhältnis-, Beziehungs-, Maßzahl; $t = \ln 2 / \ln 1,15 = 4,96$

- $x_{10} \approx 625 \cdot 4 = 2500$, dann $y_{10} = 500 + 4 \cdot 2500 = 10.500$; $\sqrt[10]{\frac{10500}{3000}} = \sqrt[10]{3,5} = 1,1335$, also 13,4% (!); sie ist geringer. Wenn $y_t = y_0 w^t$ ist die Folge der gleitenden Durchschnitte $\tilde{y}_t : 1/3 y_0 (1+w+w^2) = \tilde{y}_1, \tilde{y}_2 = w \tilde{y}_1, \tilde{y}_3 = w^2 \tilde{y}_1$ usw. Die Aussage ist also richtig.
- 9.17 a) 0,4
b) t ist zu finden, so dass gilt $(2t-2)/(5-2t+t^2) = 0,5$. Mit der Hilfsangabe $[-b \pm \sqrt{b^2 - 4ac}]/2a$ erhält man aus $t^2 - 6t + 9 = at^2 - bt + c = 0$ $t = 3$.
- 9.18 a) $(1,13 \cdot \dots \cdot 1,08)^{1/5} = 1,2958^{1/5} = 1,0532$ also 5,32%
b) $i_s < i_d$.
- 9.19 diskret: 0,4926%, stetig: 0,5%.

Kapitel 10

- 10.1 a) 119,075
b) 124,36
c) b) ist zu bevorzugen.
- 10.2 1,25; Paasche liegt zwischen 1 und 2 (100 und 200).
- 10.3 $1,1/1,06 = 1,03774$.
- 10.4 Kilopreis: $\bar{p}_t = 4, \bar{p}_0 = 2,5, P_{0t}^D = 4/2,5 = 1,6$
Pfundpreis für Tomaten: $\bar{p}_t = 3, \bar{p}_0 = 1,75, P_{0t}^D = 1,7143$.
- 10.5 Ja, Mengen reduziert $Q^P = 1/1,2 = 0,833$.
- 10.6 a) $P_{0t}^L = 10/7 = 1,4286$
b) $P_{0t}^P = W_{0t}/Q_{0t}^L = 1,4/1,2 = 1,167; Q^P = W_{0t}/P_{0t}^L = 0,98$
c) mind. 0,8, höchstens 2.
- 10.7 150 (=1,5).
- 10.8 $1,5/1,25 = 1,2$.
- 10.9 Verkettung des alten Indexes mit dem neuen; Alter Index fortgeführt: 800, 1200.
- 10.10 a) $P_{0t}^L = 1, P_{0t}^P = 0,8599$
b) $0,6 \leq P_{0t}^P \leq 1,67$
c) nein; gerade nicht bei preisunelastischer Nachfrage, wie z.B. Kfz-Nutzung
d) die gleitenden Mittelwerte liegen auf der Geraden $70-2t$ wenn März 1994 bedeutet $t = 1$.
- 10.11 $L_{0t}^L = 1,2 \cdot 0,5 + 1,1 \cdot 0,3 + 1 \cdot 0,2 = 1,13$ (L=Lohnindex); ihm liegt eine konstante Beschäftigungsstruktur (diejenige der Basisperiode) zugrunde. Die Durchschnittslöhne betrugen (bei der jeweiligen Beschäftigungsstruktur) zur Zeit $t=0$: 1000 und bei $t=1$: 1512, sie stiegen also um 51,2%, obgleich die Löhne sich maximal (bei GG) um 20% erhöhten. Grund: Veränderung der Beschäftigtenstruktur zugunsten der höheren GG.
- 10.12 $1790 = 0, 1890 = 1, 1990 = 2; P_{01}^L = 2,7, P_{02}^L = 25,02$.
- 10.13 $P^L = 1,225, P^P = 1,12346, W = 1,1375$ (Zunahme 13,75%).
- 10.14 a) $W = 1/3, P^L = 0,49167, P^P = 4/9 = 0,444$
b) da $P^L < W < 1$ muß auch $Q^L < 1$ sein, da $P^L/P^P = Q^L/Q^P > 1$ muss auch $Q^P < Q^L < 1$ sein, alle anderen Antworten sind falsch.
- 10.15 a) $P^L = 0,2, Q^P = 0,5, W = 0,2 \cdot 0,5 = 0,1$
b) Spezialisierung auf A
c) aus $\Sigma u^2 = \Sigma p^2 - 2b \Sigma xp + b^2 \Sigma x^2$ folgt, dass b minimal ist, wenn $b = \Sigma xp / \Sigma x^2$, dann ist $\Sigma u^2 = 200$. Bei $b = 1$ ist $\Sigma u^2 = 1000$.
- 10.16 a) $P^L = P^P = 1,2$, alle Preismesszahlen 1,2 (deshalb auch alle als Mittelwerte von

- Messzahlen darstellbare Indizes gleich!).
- b) das wäre so ein Fall (anderer Fall: gleiche Mengen bei 0 und t, anderer Fall → Aufg. 10.45).
- c) natürlich.
- 10.17 a) $P^C = 1,55$, $P^D = 4/3 = P^L$ da alle Mengen $q_{i0} = 100$ ($i = 1, 2, 3$); $P^C > P^L$ Ausgabenanteil niedrig (1/9) bei Gut 1 (hohe Preismesszahl 2) hoch (2/3) bei Gut 2 (niedrige Preismesszahl 7/6).
- c) $0,9\bar{3}$ also Zunahme um -6,6% (Abnahme).
- 10.18 Mengen beliebig. Es gilt bei allen Mengen $P^L = P^P = 1,5$ weil beide Preismesszahlen gleich sind (=1,5).
- 10.19 a) Gewicht: 0,5, Index: 160, b) $1000/7 = 142,86$, c) 148.
- 10.20 a) $P^D = 4/3$, $P^C = 14/9$, $P^L = P^P = 4/3$.
- b) weil alle Mengen zur Basis- und zur Berichtszeit gleich (100) sind; $P^C \neq P^L$ weil die Ausgabenanteile zur Basiszeit nicht jeweils 1/3 sondern 1/9, 6/9 und 2/9 sind.
- c) offensichtlich ist $Q^P = Q^L = 1$, so dass $W_{0t} = P_{0t}^L = 4/3$.
- 10.21 Bei Verkettung und Zeitumkehrprobe in die entsprechende Formel einsetzen. Dutots Index erfüllt beide Kriterien, Carlis Index nicht.
- Additivität, als Konkretisierung von Axiom P1b (Übers.10.4) bedeutet: ist $P(\mathbf{p}_0, \mathbf{p}_t)$ ein Preisindex mit den Preisvektoren $\mathbf{p}_0$ (Basispreise) und $\mathbf{p}_t$ (Berichtspreise) und ist $\mathbf{p}_0^* = \mathbf{p}_0 + \Delta$ so muss gelten $[P(\mathbf{p}_0^*, \mathbf{p}_t)]^{-1} = [P(\mathbf{p}_0, \mathbf{p}_t)]^{-1} + [P(\Delta, \mathbf{p}_t)]^{-1}$ (wird von P^C nicht erfüllt, wohl aber von P^D , P^L und P^P).
- 10.22 Gleichheit von Paasche- und Laspeyres-Index weil Preis- und Mengennmesszahlen nicht korreliert sind (Gewichtung der Messzahlen mit Ausgabenanteilen zur Basiszeit A: 0,2, B: 0,3, C: 0,3 und D: 0,2) Indexwerte: 1,5.
- 10.23 $Q_{0t}^L = 1,5$, $W_{0t} = 90/50 = 1,8$, $P^P = W_{0t}/Q_{0t}^L = 1,2$.

Kapitel 11

- 11.1 Die graphische Darstellung zeigt, dass z.B. ein parabolischer Trend (mit $r^2 = 0,6453$) den Daten besser angepasst ist als ein linearer Trend ($y_t = 523,41 + 2,5033t$ mit $t = 74,75, \dots$ und $r^2 = 0,1026$).
- 11.2 $\tilde{y} =$ gleitender 3-er Durchschnitt, $y^p =$ Prognose mit $\alpha = 0,2$
- | | | | | | | | | | |
|---------------|----|----|----|------|-------|--------|-------|--------|--------|
| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| $\tilde{y}_t$ | | 20 | 25 | 28 | 30 | 33 | 36 | | |
| y_t^p | 15 | 15 | 16 | 17,8 | 20,24 | 21,992 | 23,79 | 26,835 | 29,068 |
- (für $t=9$)
- 11.3 50, 70, ..., 290 also $\tilde{K}_t = 30 + 20t$; mit der Meth. der kl. Quadrate erhält man $\hat{K}_t = 32,5 + 19,643t$.
- 11.4 Nicht normierte Saisonkoeffizienten: Januar -10,6, Febr. -6,2, März 7,3, April 5,4, Mai -0,5, Juni -1,5, Juli -1,3, Aug. -14,8, Sept. 5,5, Okt. 17,0, Nov. 5,3, Dez. -3,7; Mittelwert 0,15833.
- 11.5 Ausführliche Lösung im UTB-Buch (Bsp. 11.4).
- 11.6 $\tilde{y}$ -Werte folgen der Funktion $72 - 2t$ (Apr. 89: $t=1$), lokales versus globales Trendmodell; Trend: monoton (hier: linear), Saison Zyklus mit Länge 1 Jahr; die Zykluswerte 3 (Aug.), -2 (Dez.), -1 (April) wiederholen sich. Mindestens 60, höchstens 60,33; länger, Slutsky-Yule-Effekt.
- 11.7 1800, 1800 ($t=3$ und $t=4$).

- 11.8 $\tilde{y}_1$ bis $\tilde{y}_5$ jeweils 125; $\hat{y} = 124,286 - 0,1786 t^*$; 123,145 .
- 11.9 $\tilde{y}_t = 23 + t$ ($p=3$); Zyklus: 0, 2, -2; mit $t^* = -6, -5, \dots, +5 + 6$: $\hat{y} = 29,846 + 1,022 \cdot t^*$ oder $22,69 + 1,022 \cdot t$.
- 11.10 a) 200, 250, 280, 300, 330, 360;
 b) Trend;
 c) trendbereinigte Werte: 0, 0, 20, -10, -20, 30;
 d) K: Bestandsmasse, U: Bewegungsmasse; B: Bewegungsmasse, A: Gliederungszahl, U/K: Beziehungszahl;
 e) Abw. v. Trend.
- 11.11 a) jeweils 80;
 b) kein Trend (konstant 80);
 c) Verhältnis-, Beziehungszahl;
 d) linear, $r \approx +1$;
 e) n_i jeweils 4, h_i jeweils $\frac{1}{2}$. Nein; verbundene Beobachtungen nötig.

Kapitel 12

- 12.1 a) $F_{om} = 5,25$; $\bar{B} = 5,25/4 = 1,3125$; $\bar{d} = 5,25/5 = 1,05$, $U = 3,8095$
 c) $B_0 = 1$ (Person D), $B_m = 2$ (Person C,D), $\bar{d}_0 = 0,25$, $\bar{d}_m = 0,875$, $\bar{d}_N = 3,75/3 = 1,25$; Zugänge im Intervall $[10,11]$: 2, deren durchschn.Verweildauer: 1,125 Std.
- 12.2 nach Gl.12.12: $\bar{B} = 2.217.365,6$; $Z_{om} = 22.197.597$, $A_{om} = 21.593.229$ also $\bar{d} = 0,50635$ Jahre und $U = m/\bar{d} = 5/0,506 = 9,8745$.

12.3

x	l_x	d_x	T_x^*	e_x^*
0	5	1	13	2,6
1	4	1	8	2
2	3	2	4	1,33
3	1	1	1	1

$e_x = e_x^* - \frac{1}{2}$; $T_2^* = 4$ heißt jeweils 1 Jahr für Boris und Clara, 2 für Augustus.

12.4 Fall a)

Fall a)	
d_i	4
n_i	8

Fall b)		
d_i	2	4
n_i	8	4

$U = 8/4 = 2$

$\bar{d} = 2,67$ (harmonisches Mittel von 2 und 4) $U = 8/2,67 = 3$.

12.5 Ausführliche Lösung im UTB-Buch (Beispiel 12.2).

12.6 A,B,C, $T_{o \max} 250$ ($e_0:5$), $T_{o \min} 50$ ($e_0:1$), $q_x:B$; $T_x:A,D$; $e_x:A,C$.

c) Bilden Sie Größenklassen für das Alter (von ... bis unter...) 0 - 20, 20 - 40, 40- 60, 60 - 80 und stellen Sie die klassierte Häufigkeitsverteilung sowie die Summenhäufigkeitskurve mit den absoluten Häufigkeiten graphisch dar!

d) Anstelle eines 73 jährigen Christen habe sich ein 38 jähriger Moslem in der Wüste verirrt. Nach einem allgemeinen Verständnis des Begriffs „Streuung“ müsste sich damit die Streuung

erhöht
verringert } haben .

Es gibt Streuungsmaße, bei denen sich dies darin ausdrückt, dass sich ihr Zahlenwert verändert und solche, die dies nicht zum Ausdruck bringen. Welche Streuungsmaße ändern sich und welche nicht?

Aufgabe 2

Der Geschäftsmann G hat für die 18 Löcher beim Golfspiel die folgende Anzahl x von Schlägen benötigt:

5, 3, 6, 2, 5, 4, 2, 3, 8, 1, 5, 5, 4, 4, 3, 3, 5, 4

a) Das Merkmal x ist (Richtiges ankreuzen) :

- | | |
|--|---|
| ☐ intensiv | ☐ extensiv |
| ☐ diskret | ☐ stetig |
| ☐ eine absolute Zahl | ☐ eine Verhältniszahl (relative Zahl) |

b) Die Folge der oben angegebenen 18 Zahlen bildet

- | | |
|---|---|
| ☐ eine Einheit | ☐ ein Merkmal |
| ☐ eine Bestandsmasse | ☐ eine Merkmalsausprägung |
| ☐ eine Bewegungsmasse | ☐ eine Verteilung |

Für das Merkmal x ist eineSkala definiert!



c) Bestimmen Sie den Zentralwert und das arithmetische Mittel sowie die Varianz!

d) Angenommen der Geschäftsmann G spiele zusammen mit seinem armen Bruder, Diplom-Kaufmann K aus E, der bei jedem Loch genau doppelt so viele Schläge benötigt (Merkmal y). Man gebe unter diesen Voraussetzungen die Regressionsgerade $\hat{y} = a + b \cdot x$ an!

e) Man zeichne die Regressionsgerade und gebe an, wie groß die Korrelation r_{xy} zwischen den beiden Variablen ist!

f) Angenommen der arme Bruder des Geschäftsmannes G benötigt nicht genau doppelt so viele Schläge wie G, sondern die folgende Anzahl der Schläge:

$y = 10, 5, 11, 4, 10, 7, 3, 14, 4, 5, 8, 6, 7, 6, 7, 9, 9, 8$

Man zeichne das Streudiagramm mit den Koordinaten x und y!

Aufgabe 3

Das Luxusrestaurant R hat durch einen neuen Geschäftsführer einen beträchtlichen Umsatzrückgang erlebt. Außerdem hat sich die soziale Struktur der Gäste stark verändert zugunsten von Gästen der „Unterschicht“, während bisher mehr Personen der „Oberschicht“ dort verkehrten. Der Anteil y (in Prozent) der Oberschichtgäste hat sich in den letzten 6 Monaten wie folgt verringert:

Monat	1	2	3	4	5	6
y	85	68	57	55	38	27



- Man berechne gleitende 3-Monats-Durchschnitte für die Größe y!
- Die Werte für die gleitenden Durchschnitte liegen offenbar auf einer Geraden. Wie lauten diese? Warum ist das so?
- Würde man zu der gleichen Geraden gelangen, wenn man einen linearen Trend mit der Methode der kleinsten Quadrate berechnen würde? Begründung!
- Während früher die meisten Gäste teure Menüs bestellten, begnügen sich jetzt die neuen Gäste vorwiegend mit billiger Suppe. Für Preise und Mengen liegen zu zwei Zeitpunkten folgende Angaben vor:

	t = 0		t = 1	
	Preise	Mengen	Preise	Mengen
Suppen	5	40	2	200
Menüs	20	110	10	40

Man bestimme eine Messzahl M für den Umsatz zur Basis t = 0 und berechne den Preisindex nach Laspeyres P_1^L zur Basis t = 0 und den Preisindex nach Paasche P_1^P zur Basis t = 0.

- Der Umsatz ist stärker zurückgegangen als die Preise. Daraus folgt, dass für die Mengenindizes nach Laspeyres (Q^L) und nach Paasche (Q^P) gelten muss (Richtiges ankreuzen)
 - ☐ Q^L und Q^P sind kleiner als 100 %
 - ☐ $Q^L < P^P$, da $P^P < M$
 - ☐ da $P^L > P^P$, muss $Q^L < Q^P$ sein
 - ☐ da $P^L > M$, muss $Q^P < P^L$ sein

Aufgabe 4

König Egon XIII „der Labile“ hatte zwei Maitressen, die Pompadur (D) und die Pompamoll (M), die miteinander heftig um die Gunst des Königs wetteiferten. Aus einer seinerzeit von der Hofschranze H verfassten Notiz geht hervor, dass Egon seine Freizeit in den letzten 40 Tagen des Jahres 1742 wie folgt verteilt hatte (Größenklassen z.B. von 2 bis unter 4 Stunden etc.)

		Stunden bei D (Variable X_D)		
		0 - 2	2 - 4	4 - 6
Stunden	0 - 2	2	4	3
bei M	2 - 4	6	12	9
(X_M)	4 - 6	2	4	3

- Man bestimme die bedingten Mittelwerte (empirische Regressionslinien)!
- Wie groß ist angesichts der Gestalt der Regressionslinien die Korrelation zwischen den Variablen X_D und X_M (Anzahl der Stunden bei D bzw. bei M)!
- Während der Zeit seiner Liaison mit D stiegen die Kosten y_t der Hofhaltung sprunghaft an. Für die sechs Jahre ab 1743 ergaben sich folgende Zahlen:

$$5^2 = 25, \quad 5^3 = 125, \quad 5^4 = 625, \dots, \quad 5^7 = 78125$$

Wie sieht diese Zeitreihe in halblogarithmischer Darstellung aus!

- Kann es sein, dass sich die Kosten der Hofhaltung laufend erhöhen, während ein entsprechender Preisindex nach Laspeyres konstant bleibt oder gar sinkt?
- Der König besucht mit D, um der eifersüchtigen M zu entgehen, sein 12 km weit entferntes Lustschloß „Egonsburg“. Auf dem Heimweg fährt seine Droschke eine Geschwindigkeit von 12 km/h, auf dem Rückweg dagegen, weil ein Empfang zur Eile mahnt, mit 30 km/h. Man berechne die Durchschnittsgeschwindigkeit

- als arithmetisches Mittel
- als harmonisches Mittel!

Welcher Mittelwert ist hier sinnvoller! (Begründung!)

h) die Anzahl „6“ bei den Einfällen

i) ein einzelner Einfall

j) 10 Bemühungen um Einfälle

e) Durch diese Zuordnung von Zahlen zu Begriffen sind die Zahlen im Sinne einer

- ☐ Intervallskala
- ☐ Nominalskala
- ☐ Ordinalskala
- ☐ überhaupt keiner Skala benutzt worden

Aufgabe 6

„Und wenn die Geschwister Männer und Frauen sind, so soll ein Mann so viel erhalten wie zwei Frauen“

Sure 4, Vers. 175

Ein Versuch eines islamischen und eines christlichen Dorfes ergab folgende Daten:

	christliches Dorf		islamisches Dorf	
Altersklasse der Frauen	Anzahl der		Anzahl der	
	Frauen	Geburten	Frauen	Geburten
15 - 30	400	44	720	96
30 - 45	600	48	480	48

a) Man berechne die Fruchtbarkeitsraten (Anzahl der Geburten auf 1000 Frauen im gebärfähigen Alter) für beide Dörfer. Hierbei handelt es sich um eine(n)

- ☐ Messzahl
- ☐ Maßzahl
- ☐ Gliederungszahl
- ☐ Beziehungszahl
- ☐ Verursachungszahl
- ☐ Quote
- ☐ Mittelwert
- ☐ Streuung
- ☐ Verhältniszahl

b) Man bestimme das Durchschnittsalter der gebärfähigen Frauen in beiden Dörfern als arithmetisches Mittel.

c) Worauf ist es zurückzuführen, dass die Fruchtbarkeit im islamischen Dorf offenbar größer ist als im christlichen Dorf? Wie kann man feststellen, ob der Unterschied tatsächlich ein echter Unterschied dergestalt ist, dass die islamischen Frauen fruchtbarer sind als die christlichen?

d) In den beiden Dörfern ist das Vermögen gleichmäßig verteilt. Nach christlichem Erbrecht erhalten Knaben und Mädchen ein gleich großes Erbe. Welche Vermögenskonzentration (Gemessen am Ginischen Konzentrationsverhältnis) entsteht jedoch, wenn das Vermögen aller Familien des islamischen Dorfes getreu nach den oben dargestellten Regeln des Korans auf 72 Knaben und 72 Mädchen verteilt wird?

e) ... auf 108 Knaben und 36 Mädchen verteilt wird?

Aufgabe 7

Eine Autovermietung habe die folgende Verteilung ihrer 900 Kunden bezüglich der gefahrenen Kilometer je Wagen festgestellt:

	gefahrte Kilometer je Wagen	Anzahl der Mieter
Größenklasse	x_j	n_j
1	0 bis unter 85	0
2	85 bis unter 115	200
3	115 bis unter 145	500
4	145 bis unter 175	200

- a) Berechnen Sie Mittelwert und Varianz der Größe x !
- b) Die festen Mieteinnahmen pro Tag und Wagen betragen 35,- DM und die Einnahmen pro gefahrenem Kilometer 0,50 DM. Wie groß sind dann die durchschnittlichen Mieteinnahmen $\bar{y}$ pro Tag und Wagen und die (externe) Standardabweichung s_y !
- c) Durch eine Preiserhöhung im Mietwagenwesen erhöhe sich der kilometerabhängige Betrag von 0,5 auf 0,8 DM. Wie ändert sich der Variationskoeffizient?
- d) Die gefahrenen Kilometer x und Mieteinnahmen y korrelieren mit
 $r_{xy} = \dots\dots\dots!$
- e) Ist die Verteilung der gefahrenen Kilometer
- ☐ linkssteil
 - ☐ rechtssteil
 - ☐ symmetrisch?
- f) Wenn man die Größenklassen 3 und 4 zusammenfasst, werden sich folgende Größen wie folgt verändern:

	bleibt gleich	wird größer	wird kleiner
arithm. Mittel			
Varianz			
Konzentration			
Mieteinnahmen			
Zentralwert			
Anzahl der Mieter			

Aufgabe 8

Der Student S glaubt wieder einmal, eine Klausur astrein gelöst zu haben. Mit seiner Selbsteinschätzung (Variable x), die mehr oder weniger gefühlsmäßig und zufällig (mangels tieferer

Einsicht) erfolgt, liegt er jedoch oft nicht richtig. Insbesondere bei den Recht-Klausuren erscheint ihm das tatsächliche Ergebnis (Variable y) meist überraschend und unerklärlich, nachdem er meinte, er habe den Fall Spitzenmäßig gepackt. Die letzten 8 Klausuren brachten folgende geschätzte (x) und tatsächliche (y) Noten:

x	1	3	4	5	4	2	3	2
y	2	4	3	5	5	3	5	5

- Berechnen Sie den Korrelationskoeffizienten r_{xy} und begründen Sie, warum diese Aufgabe die Fragestellung der Korrelationsanalyse beschreibt!
- Welche Werte $\hat{y}$ erhält man bei Kenntnis der Regressionsgeraden $\hat{y} = a + b \cdot x$ (also einer korrigierten Selbsteinschätzung x)?
- Zeigen Sie, dass $\hat{y}$ eine bessere Schätzung für y ist als die Selbsteinschätzung x des Studenten S!
- Kann man aufgrund der folgenden Angaben für 25 Klausuren des S darauf schließen, dass zwischen der Art des Faches und der Treffsicherheit der Selbsteinschätzung ein Zusammenhang besteht!

Klausur	S hat sich vorher eingeschätzt	
	besser	schlechter oder gerade richtig
Recht	4	2
andere Klausuren	10	9

Aufgabe 9

Ein Unternehmen betätigt sich in zwei Wirtschaftszweigen (Branche A und B). Umsätze (U) und Kapitaleinsatz (K) mögen sich wie folgt entwickelt haben

Jahr	Branche A		Branche B	
	U	K	U	K
1976	10	50	6	50
1977	14	70	4,2	35
1978	20	100	3	25

- Man berechne die Umsatzrentabilitäten $\frac{U}{K}$ für die beiden Branchen und für das gesamte Unternehmen (warum ist letztere gestiegen?)!
- Man berechne Messzahlen für die gesamte Umsatzentwicklung (Basis 1976 = 100)!
- Im Jahr 1979 habe sich in der Branche B gegenüber 1978 der Umsatz um 50 % und der

Kapitaleinsatz um 20 % erhöht, in der Branche A dagegen der Kapitaleinsatz um 30 %, der Umsatz aber nur um 55 %. In welcher Branche ist die Umsatzrentabilität stärker gestiegen?

d) Die Umsatzrentabilität ist eine (Zutreffendes ankreuzen)

- ☐ Verhältniszahl
- ☐ Gliederungszahl
- ☐ Quote
- ☐ Beziehungszahl
- ☐ Rate
- ☐ Wachstumsrate
- ☐ Messzahl
- ☐ Maßzahl

e) Man berechne einen Umsatzindex nach Laspeyres für die Jahre 1977 und 1978 aus den Umsatzmesszahlen der beiden Branchen durch Gewichtung mit den Umsätzen zur Basiszeit 1976. Wie unterscheidet sich ein so berechneter Umsatzindex von der Umsatzmesszahl des Teil b!

Aufgabe 10

Die Ehefrau des sanftmütigen Diplom-Kaufmann K aus E beklagt sich bei ihrem Mann vehement über die erheblich gestiegenen Lebenshaltungskosten. K verweist demgegenüber darauf, dass der Preisindex für die Lebenshaltung gesunken sei.

- a) Unter welchen Voraussetzungen ist es möglich, dass die Lebenshaltungskosten steigen, der Preisindex für die Lebenshaltung aber sinkt!
- b) Der Haushalt des K konsumierte von den vier Verbrauchsgruppen (Waren A, B, C, D) die folgenden Mengen



	A	B	C	D
zur Basiszeit	100	50	200	300
zur Berichtszeit	80	50	150	565

Die Preise waren:

	A	B	C	D
zur Basiszeit	10	8	15	2
zur Berichtszeit	9	7	12	2

Man berechne die Zunahme der Lebenshaltungskosten!

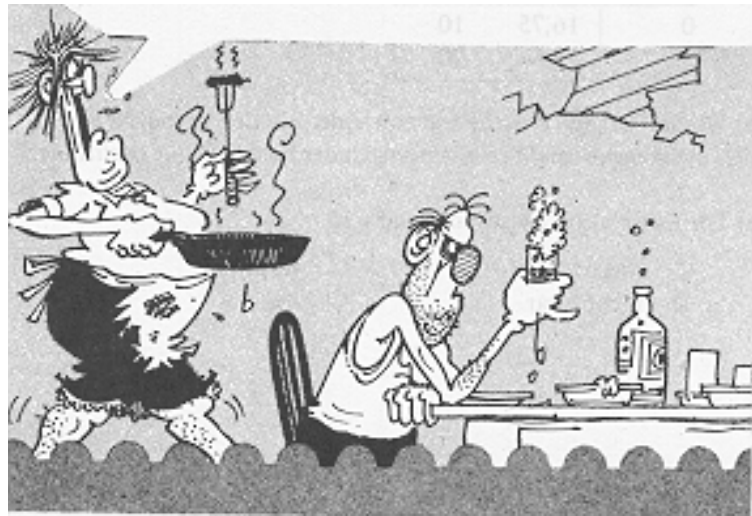
- c) Man berechne den Laspeyres - Preisindex (Zahlenangaben hier und im folgenden von Teil b)!
 d) Man berechne den Paasche-Mengenindex!
 e) Der Paasche - Preisindex könnte, wenn der Haushalt des K sein Verbraucherverhalten entsprechend einrichten würde, nur Werte annehmen zwischen ... und ...

Er wird in der Regel ☐ höher
☐ niedriger
☐ genauso hoch

sein als/wie der Laspeyres-Preisindex. Begründung?

Aufgabe 11

Der Haushalt des arbeitslosen Diplomkaufmanns K aus E habe sich in den letzten 4 Monaten jeweils auf dem gleichen niedrigen Nutzenniveau bewegt und dabei von zwei Gütern die folgenden Mengen x , y konsumiert (wegen der folgenden Berechnungen seien auch noch die natürlichen Logarithmen dieser Mengen mitgeteilt):



Monat	x	y	ln x	ln y
0	4	3,3	1,4	1,2
1	3,3	5	1,2	1,6
2	2	7,4	0,7	2
3	8	2	2,1	0,7

- a) Man zeichne ein Streudiagramm mit den Achsen x und y . Die zwischen diesen Punkten verlaufende Indifferenzkurve (Regressionsfunktion) wird vermutlich sein:

linear }
 konvex } vom Ursprung des Koordinatensystems gesehen.
 konkav }

b) Man schätze die Indifferenzkurve als Regressionsfunktion

$$\hat{y} = a \cdot x^b,$$

indem man diese Gleichung durch geeignete Transformation der Variablen linearisiert.
Hinweis: es genügt, die Normalgleichungen zu bestimmen!

c) Für die Periode 3 berechne man einen Preisindex zur Basis 0 nach Laspeyres und nach Paasche, wenn folgende Preise gegeben seien

Monat	Gut x	Gut y
0	16,75	10
3	13,5	20

d) Warum ist der Paasche-Index kleiner als der Laspeyres-Index? Wie verhalten sich die Mengen- und Preisänderungen der beiden Güter zueinander?

e) Die Korrelation zwischen x und y ist

- ☐ linear ☐ positiv
☐ nichtlinear ☐ negativ

Aufgabe 12

Die Sterbestatistik des Landes X erfasse die drei Merkmale: Alter, Todesursache und Geburtsort (A = Abendland, M = Morgenland).

Hinsichtlich der Todesursache wird unterschieden:

S = Altersschwäche

N = sonstige natürliche Todesursachen

U = unnatürlicher Tod

Man erhielt für das Jahr $t = 5$ folgende Daten für die Todesfälle in 1000:

Alter	Todesursache			Summe	darunter A
	S	N	U		
0 bis unter 20	0	5	10	15	10
20 bis unter 40	0	15	20	35	20
40 bis unter 60	0	20	30	50	40
60 bis unter 80	30	60	10	100	90
80 und mehr	70	30	0	100	100
Summe	100	130	70	300	260

- a) Diese Tabelle enthält Merkmale und mehrdimensionale Häufigkeitsverteilung(en) (gemeinsame Verteilungen von Merkmalen). Geben Sie die Skalen und die geeigneten graphischen Darstellungen für diese Merkmale an!
- b) Bilden Sie eine klassierte Verteilung mit relativen Häufigkeiten für das Merkmal Alter mit den Klassen „0 bis unter 40 Jahren“, „40 bis unter 60“ und „60 und mehr Jahre“ und bestimmen Sie den Zentralwert (mit Interpolation).
- c) Um herauszubekommen, auf welche Altersklassen sich die Todesfälle besonders konzentrieren und das Sterberisiko besonders groß ist, berechne ich (Richtiges ankreuzen)
- ☐ die Lorenzkurve, weil sich die Sterbefälle auf bestimmte Todesursachen konzentrieren
 - ☐ die Abweichung vom Trend, weil dies zeigt, in welchem Jahr besonders ungewöhnliche Verhältnisse vorlagen
 - ☐ den Modus der Altersverteilung
 - ☐ die Varianz, weil keine Konzentration vorliegt, wenn sie klein ist
 - ☐ die Korrelation mit der Todesursache, weil die Sterblichkeit hiervon abhängig ist.
- d) Bekanntlich sterben mehr Menschen im Bett als an irgendeinem anderen Ort. Kann man daraus schließen, dass das Bett der gefährlichste Aufenthaltsort ist? Wenn nein, warum nicht?
- e) Um das Jahr $t = 9$ erlebte das Land X eine Revolution, was dazu führte, dass die Anzahl der unnatürlich Gestorbenen sprunghaft anstieg. Diese Anzahl U_t folge ziemlich genau im Zeitablauf der folgenden Funktion:
- $$U_t = 86 - (9 - t)^2$$
- Man bestimme eine Funktion für
- die Wachstumsrate von U_t
 - Messzahlen von U_t zur Basis $t = 3$
- f) Die Wachstumsraten der Messzahlenreihen sind
- größer } als die der Reihe U_t
 - kleiner }
 - gleich denen der Ursprungsreihe U_t

Aufgabe 13

Im Entwicklungsland E gäbe es zwei soziale Klassen: Arme Schlucker (A) und dekadente Bourgeois (B), deren Warenkörbe sich erheblich unterscheiden. Die Ausgabenanteile für vier Warenarten und die Preise in E-Dollar waren zur Basiszeit ($t = 0$):

Warenart	Preis	Ausgabenanteile in vH	
		Klasse A	Klasse B

Miete	10	40	10
Gebrauchsgüter	2,5	10	40
Lebensmittel	5	40	20
Kleidung	10	10	30

- a) Die Inflation führte dazu, dass die Armen Schlucker zur Zeit $t = 1$ für ihren Warenkorb 95 und die Bourgeois sogar 504 E - Dollar mehr zahlen mussten als zur Zeit $t = 0$, denn die Preise waren wie folgt gestiegen:

Miete um 150 % Lebensmittel um 80 %
 Gebrauchsgüter um 20 % Kleidung um 10 %

Kann man daraus folgern, dass die Inflation die Reichen mehr schädigt als die Armen?

- b) Man berechne aus den soweit gemachten Angaben den Preisindex nach Laspeyres für die Klasse A und die Klasse B!

- c) Die Zunahme der „Lebenshaltungskosten“ ergibt sich

- ☐ aus der unter a) genannten Wertsteigerung der Warenkörbe um 95 bzw. 504 E - Dollar
- ☐ aus den unter b) berechneten Preisindizes
- ☐ weder aus Teil a) noch aus Teil b) dieser Aufgabe

- d) Die Ausgabenanteile des Haushalts B mögen sich verändert haben und zur Zeit $t = 1$ folgende sein:

Miete 15 %, Gebrauchsgüter 45 %, Lebensmittel 18 %, Kleidung 22 % und die Gesamtausgaben mögen 2000 E - Dollar betragen. Berechnen Sie den Paasche - Preisindex!

- e) Wie ist es zu erklären, dass der Preisindex nach Paasche für die Bourgeoisie noch niedriger ist als der Preisindex nach Laspeyres?

Aufgabe 14

Die britischen Flugzeughersteller machten in den Jahren 1960 bis 1970 die Profite (Gewinne) y und erhielten folgende Subventionen x (beide Variablen x und y in Millionen Pfund Sterling)

x	9,7	5,1	8,8	10,0	4,0	13,5	10,3	15,7	19,4	23,2	14,2
y	3	11	12	12	12	20	33	52	70	83	80

- a) Man berechne die Regressionsfunktion $\hat{y} = a + bx$!

- b) Berechnen Sie den Korrelationskoeffizienten r_{xy} !

- c) Berechnen Sie die Varianz der Störgröße u zur Beurteilung der Güte der Anpassung!

- d) Eine dritte Variable z seien die Regierungskäufe für zivile und militärische Zwecke. Es ließ sich errechnen:

$$r_{xz} = 0,3840$$

$$r_{yz} = 0,4425$$

Berechnen Sie die partiellen Korrelationen

$$r_{xy.z}, r_{yz.x} \text{ und } r_{xz.y}$$

und interpretieren Sie das Ergebnis!

e) Berechnen Sie die multiple Bestimmtheit $R^2_{y \cdot xz}$ und interpretieren Sie das Ergebnis!

Lösungen zum Klausurtraining

Lösung zu Aufgabe 1

a)

Merkmal	Skalentyp	Mittelwerte
G = Geschlecht R = Religion	Nominalskala	Modus (Dichtester Wert) bei G: männlich bei R: christlich
A = Alter	metrische Skala (Ratioskala)	alle Mittelwerte berechenbar (Modus aber bei den konkreten Zahlen nicht sinnvoll)

Bei A sind die Mittelwerte

Zentralwert: der $\frac{n+1}{2}$ te Wert (also bei $n = 7$) der vierte Wert in der Reihe

15, 17, 19, (21), 30, 35, 73

arithmet. Mittel: $\frac{210}{7} = 30$

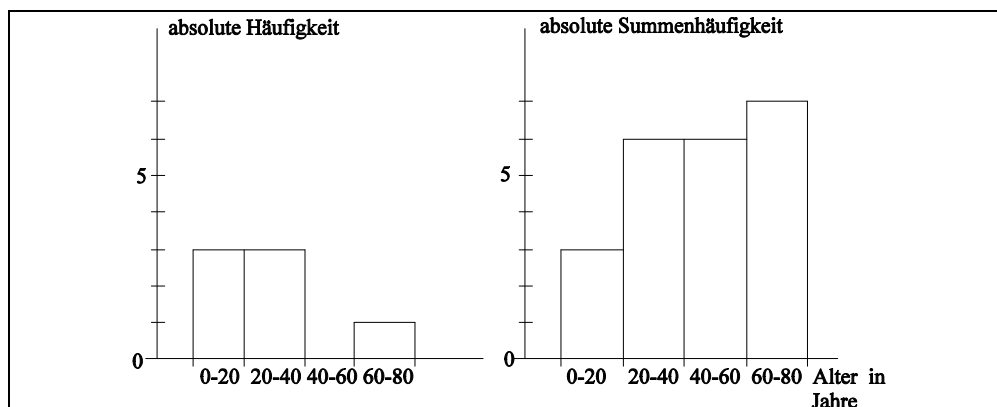
geometr. Mittel: $\sqrt[7]{15 \cdot 17 \cdot \dots \cdot 73} = 25,891$

harmonisches Mittel: $\frac{7}{\frac{1}{15} + \frac{1}{17} + \dots + \frac{1}{73}} = 23,229$

b) 1, 4, 3, 4, 5, 3, 2

c)

Klasse		absolute		relative	
Nr.	Alter	Häufigkeit	Summenhäufigkeit	Häufigkeit	Summenhäufigkeit
1	0 - 20	3	3	0,43	0,43
2	20 - 40	3	6	0,43	0,86
3	40 - 60	0	6	0	0,86
4	60 - 80	1	7	0,14	1,00



d) verringert haben

es verringert sich	es ändert sich nicht
Spannweite, mittl. Abweichung, Varianz, Standardabweichung und Variationskoeffizient	mittlerer Quartilsabstand

Der Variationskoeffizient sinkt von 2,96 auf 2,61, was aber nicht selbstverständlich ist, da sich sowohl die Standardabweichung als auch das arithmetische Mittel verringert haben.

Lösung zu Aufgabe 2

- a) Das Merkmal ist **extensiv**, weil beim Golf am Ende die **Gesamtzahl** der Schläge gewertet wird. X ist selbstverständlich **diskret** und stellt eine **absolute** Anzahl dar; auch wenn man von „Schlägen je Loch“ spricht, ist dies keine Beziehungszahl. X wird auch nicht durch Division, sondern direkt durch Zählen ermittelt.
- b) Bewegungsmasse, Verteilung, Ratioskala.
- c) $Z = 4$, $\bar{x} = 4$, $s^2 = \frac{46}{18} = 2,56$.
- d) Die Anzahl y ist stets genau doppelt so groß wie die Anzahl x. Es besteht also ein funktionaler Zusammenhang
 $\hat{y}_i = 2x_i \quad (i = 1, 2, \dots, 18) \quad \text{und} \quad \hat{y}_i = y_i$
- e) Die Regressionsgerade verläuft durch den Ursprung mit einer Steigung von 2. Da ein streng funktionaler (nicht durch eine Zufallsvariable gestörter) Zusammenhang besteht, ist r_{xy} notwendig genau + 1.

Lösung zu Aufgabe 3

- a) Die gleitenden Durchschnitte lauten

Monat (t)	2	3	4	5
y_t	70	60	50	40

- b) Die Gerade lautet $90 - 10t$. Um diese Gerade liegt ein regelmäßiger dreigliedriger Zyklus mit den Abweichungen +5, -2, -3

t	Gerade	Abweichung u_t	y_t	
1	80	+5	85	Dieser Zyklus wird mit dreigliedrigen gleitenden Durchschnitten vollständig eliminiert.
2	70	-2	68	
3	60	-3	57	
4	50	+5	55	
5	40	-2	38	
6	30	-3	27	

- c) Man gelangt nicht zur selben Geraden, weil der Ausgleich der Abweichungen ($\sum u_t = 0$) eine notwendige aber nicht hinreichende Bedingung für eine Regressionsgerade ist. Für den linearen Ansatz $y = a + bt$ erhält man folgende Normalgleichungen:

$$(1) 330 = 6a + 21b$$

$$(2) 964 = 21a + 91b$$

Die Gerade $\hat{y} = 90 - 10t$ erfüllt Gl. 1 (weil bei dieser Geraden wie in Teil b gezeigt $\sum u_t = 0$ ist), nicht aber Gl. 2. Die Normalgleichungen führen zu

$$\hat{y} = 93,2 - 10,914t$$

Die Summe der Quadrate der Abweichungen von dieser Geraden ist 61,37, während sie bei der Geraden $90 - 10t$ einen höheren Betrag, nämlich 76 annimmt.

d)

$$M = \frac{400 + 400}{200 + 2200} = \frac{1}{3}$$

$$P_1^L = \frac{2 \cdot 40 + 10 \cdot 110}{2400} = 0,4917 \text{ also } 49,2 \text{ vH}$$

$$P_1^P = \frac{800}{5 \cdot 200 + 20 \cdot 40} = \frac{4}{9} = 0,4444 \text{ also } 44,4 \text{ vH}$$

e) Es muss gelten $M = P^L Q^P = P^P Q^L$. Daraus folgt unmittelbar, dass die Antwortmöglichkeiten 1 und 3 richtig sind. Die zweite Antwort widerspricht offenbar den Fakten, da ja $P^P > M$. Aus obiger Gleichung errechnet sich sofort $Q^P = \frac{0,33}{0,49} \approx 0,67$, was der Behauptung der vierten Antwortmöglichkeit widerspricht.

Lösung zu Aufgabe 4

- a) Alle bedingten Mittelwerte der Variable X_M sind 3 Stunden und alle der Variable X_D sind 3,22 Stunden. Die Regressionslinien verlaufen also parallel zu den Achsen.

$$\hat{X}_D = 3,22 \text{ und } \hat{X}_M = 3,$$

d.h. dass die Steigung der Regressionsgeraden jeweils Null ist.

- b) Aus diesem Grunde sind die Variablen nicht miteinander korreliert. Die Kovarianz ist null.

- c) Die Zahlen stellen eine exponentielle Entwicklung $y_t = 5^t$ dar. In halblogarithmischer Dar-

stellung ist dies die Gerade $\ln y_t = 0,699t$. Die Wachstumsrate (bei stetiger Zeit t) ist dann konstant

$$\frac{y'}{y} = \frac{d \ln y}{dt} = \ln 5 = 1,61 \text{ also } 161\%.$$

- d) Beim Preisindex nach Laspeyres wird mit konstanten Mengen der Basiszeit gewichtet, so dass die gefragte Entwicklung eintreten kann, wenn gleichzeitig die Mengen erheblich zunehmen. Das folgt auch daraus, dass ein Wertindex (d.h. hier Messzahl der Kosten) das Produkt aus Laspeyres - Preisindex und Paasche - Mengenindex ist.
- e) Das arithmetische Mittel von 12 und 30 ist 21 km/h. Für die Gesamtstrecke von 2 mal 12 km (also 24 km) würde er danach 1,143 Stunden benötigt haben, während er wirklich 1,4 Stunden benötigte und somit eine Durchschnittsgeschwindigkeit von 17,14 km/h fuhr, was das harmonische Mittel ist $\frac{2}{\frac{1}{12} + \frac{1}{30}} = 17,14$, was zeigt, dass dies hier das sinnvoller anzuwendende Mittel ist als das arithmetische Mittel.

Lösung zu Aufgabe 5

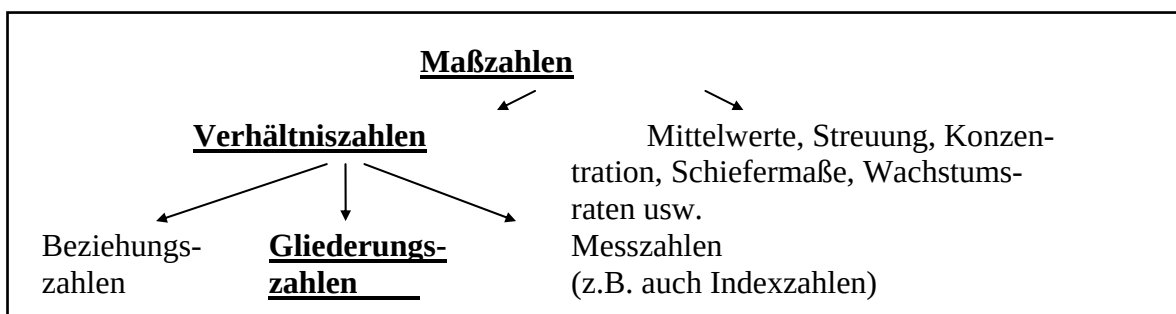
- a) Die Daten lassen sich als Vierfeldertafel wie folgt darstellen

	Einfälle	keine Einfälle	Σ
Spaziergang	a = 6	b = 5	11
Büro	c = 4	d = 5	9
Σ	10	10	20

Folglich sind die relativen Häufigkeiten für Einfälle

im Büro: 0,4
bei Spaziergängen: 0,6
insgesamt: 0,5

- b) Das Begriffschema ist



Maßzahlen sind alle kennzeichnenden Größen (Statistiken, Kennzahlen etc.), die in der beschreibenden Statistik verwendet werden. Die unterstrichenen Begriffe sind also richtig, alle anderen falsch.

- c) Vgl. Teil a) dieser Lösung. Die Häufigkeiten sind a, b, c, d. Da keine Rangordnung zwischen den Merkmalsausprägungen besteht, ist das Vorzeichen der Vierfelderkorrelation (-assoziati) nicht zu interpretieren, d.h. es ist hier irrelevant, ob Φ positiv oder negativ ist. Man erhält

$$\Phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} = \frac{10}{\sqrt{9900}} \approx 0,1$$

- d) a) $\rightarrow$ 1 b) $\rightarrow$ 2 c) $\rightarrow$ 4
 d) $\rightarrow$ 4 e) $\rightarrow$ 1 f) $\rightarrow$ 1
 g) $\rightarrow$ 3 h) $\rightarrow$ 4 i) $\rightarrow$ 2
 j) $\rightarrow$ 5

- e) Die Zahlen sind im Sinne einer **Nominalskala** benutzt worden.

Lösung zu Aufgabe 6

- a) Rohe (gesamte) Fruchtbarkeitsraten f:

$$\text{christliches Dorf} \quad f_c = \frac{92}{1000} \cdot 1000 = 92$$

$$\text{islamisches Dorf} \quad f_i = \frac{144}{1200} \cdot 1000 = 120$$

Maßzahl, Beziehungszahl (speziell hier auch Verursachungszahl genannt), Verhältniszahl (als Oberbegriff). Zum Begriffsschema vgl. auch Aufgabe 5!

- b) $\bar{x}_c = 22,5 \cdot 0,4 + 37,5 \cdot 0,6 = 31,5$
 $\bar{x}_i = 22,5 \cdot 0,6 + 37,5 \cdot 0,4 = 28,5$

- c) Sowohl auf eine günstigere Altersstruktur als auch auf einen echten Fruchtbarkeitsunterschied, der deutlich wird bei den altersspezifischen Fruchtbarkeitsraten (mit 1000 multipliziert)

	christlich	islamisch
15 - 30	11	13,3
30 - 45	8	10

Die rohen Fruchtbarkeitsraten kann man wegen der unterschiedlichen Altersstruktur nicht vergleichen. Gewichtet man beide Dörfer mit derselben Altersstruktur (z.B. mit der des islamischen Dorfes), so erhielte man folgende standardisierte Fruchtbarkeitsraten

$f_i^* = 120, f_c^* = 98$, die Ausdruck des echten Fruchtbarkeitsunterschieds sind.

- d) Anteile der Mädchen bzw. Knaben 0,5. Anteile am zu verteilenden Vermögen $\frac{1}{3}$ bzw. $\frac{2}{3}$.

Die Lorenzkurve besteht somit aus folgenden Punkten $P_i (H_i, Q_i)$:

$$P_0(0, 0), P_1\left(\frac{1}{2}, \frac{1}{3}\right), P_2(1, 1);$$

das Konzentrationsverhältnis ist $R = \frac{1}{6} = 0,16667$.

- e) Punkte der Lorenzkurve $P_0, P_1^*\left(\frac{1}{4}, \frac{1}{7}\right), P_2$, da die Mädchen vom Gesamtvermögen des Betrages $\frac{7}{4} \cdot V$ den Betrag $\frac{1}{4} \cdot V$ und die Jungen $\frac{3}{4} \cdot 2 \cdot V$ erhalten. Folglich haben sie die Anteile $\frac{1}{7}$ bzw. $\frac{6}{7}$.

Die Konzentration ist entsprechend kleiner geworden.

$$R = \frac{3}{28} = 0,10714$$

Die Ungleichverteilung zwischen Jungen und Mädchen führt dann zur größten Konzentration, wenn Knaben- und Mädchengeburten jeweils gleichwahrscheinlich sind. Gäbe es z. B. $\frac{3}{4}$ Mädchen und $\frac{1}{4}$ Jungen, so wäre die Konzentration $R = 0,15$.

Lösung zu Aufgabe 7

- a) $\bar{x} = 130, s_x^2 = 400, s_x = 20$. Dabei kann nur die externe Varianz (zwischen den Klassen) berücksichtigt werden. Die wahre Varianz s_x^2 ist um die (nicht mitgeteilte) interne Varianz größer.

- b) $y = 35 + 0,5x$ folglich $\bar{y} = 35 + 0,5 \cdot 130 = 100$ und $s_y^2 = 0,25 \cdot s_x^2 = 100, s_y = 10$

- c) Durch Änderung von b ändert sich der Variationskoeffizient nicht. Er bleibt 0,1, denn $\frac{0,8 \cdot 10}{0,8 \cdot 100} = 0,1$.

- d) $r_{xy} = +1$ weil der Zusammenhang streng funktional und linear (vgl. Teil b) ist.

- e) Die Verteilung ist symmetrisch. Es ist auch $Z = \bar{x} = 130$. Ferner ist das dritte zentrale Moment

$$\frac{1}{900} \cdot [(100 - 130)^3 \cdot 200 + (160 - 130)^3 \cdot 200] = 0$$

und folglich die Schiefe null.

- f) Der Zentralwert steigt von 13 auf 136,43 (die Verteilung ist auch nicht mehr symmetrisch). Die Varianz verringert sich, weil sich die externe Varianz wie folgt verringert

$$s_x^{*2} = \frac{1}{900} (900 \cdot 200 + 225 \cdot 700) = 375$$

Die Konzentration wird notwendig geringer (es entfällt einer der Punkte auf der Lorenzkurve), die übrigen verändern sich wie folgt:

	bleibt gleich	wird größer	wird kleiner
arithm. Mittel	$\bar{x}$		

Varianz			x
Konzentration			x
Mieteinnahmen	x		
Zentralwert		x	
Anzahl der Mieter	x		

Lösung zu Aufgabe 8

a) Arbeitstabelle:

$$\sum (x - \bar{x})^2 = 12$$

$$\sum (y - \bar{y})^2 = 10$$

$$\sum (x - \bar{x})(y - \bar{y}) = 6$$

$$\text{Folglich ist } r = \frac{6/8}{\sqrt{12/8 \cdot 10/8}} = 0,5477.$$

Die hier vorliegende Fragestellung ist ein Problem der Korrelationsanalyse, da beide zu betrachtenden Variablen X und Y als Zufallsvariablen aufgefaßt werden können⁶.

b) Berechnung der Koeffizienten der Regressionsgleichung

$$\left. \begin{array}{l} b = 0,5 \\ a = 2,5 \end{array} \right\} \hat{y} = 2,5 + 0,5x$$

Berechnung der $\hat{y}$ - Werte

x	1	3	4	5	4	2	3	2
$\hat{y}$	3	4	4,5	5	4,5	3,5	4	3,5

c) $\hat{y}$ sei eine bessere Schätzung für y als x:

Bestätigung der Aussage über die Berechnung der Summe der Quadrate der Abweichungen

	Quadrate der Abweichungen								Σ
$(x - y)^2$	1	1	1	0	1	1	4	9	18
$(\hat{y} - y)^2$	1	0	2,25	0	0,25	0,25	1	2,25	7

Dieses Ergebnis ist auch aus allgemeinen Erwägungen zu erwarten, da $\hat{y}$ ja der Funktionswert auf der Regressionsgeraden (Regresswert) ist und die Regressionsgerade so bestimmt wird, dass $\sum (\hat{y} - y)^2$ minimiert wird. Deshalb muss auch $\sum (\hat{y} - y)^2 < \sum (x - y)^2$ erfüllt sein.

⁶ Das heißt natürlich nicht, dass die Klausurbewertung nach dem Zufallsprinzip erfolgt, sondern nur, dass sie **aus der Sicht der Studenten** nicht vorherbestimmt und kontrolliert ist.

- d) Die bedingten relativen Häufigkeiten h dafür, sich vorher besser einzuschätzen als die Klausur dann tatsächlich war, sind

$$h_{\text{Recht}} = \frac{4}{6} = 0,667 \quad \text{und} \quad h_{\text{and. Klausuren}} = \frac{10}{19} = 0,526.$$

Sie sind also verschieden, so dass ein Zusammenhang (Assoziation) zwischen der Art des Faches und der Treffsicherheit der Selbsteinschätzung besteht. Die Vierfelderkorrelation ist hier

$$\phi = 0,12076 \neq 0, \text{ so dass keine Unabhängigkeit besteht.}$$

Lösung zu Aufgabe 9

a)

Jahr	Umsatzrentabilitäten			Kapitaleinsatz	Anteil der Branche A
	A	B	insgesamt		
76	0,2	0,12	0,160	100	1/2
77	0,2	0,12	0,173	105	2/3
78	0,2	0,12	0,184	125	4/5

Zunahme der Umsatzrentabilität im Gesamtunternehmen (ein gewogenes Mittel!) allein deshalb, weil der Kapitalanteil in der rentableren Branche gestiegen ist (Investitionen zugunsten der rentableren Branche).

b)

Jahr	Umsätze		Messzahlen des Kapitals vergl. Teil a)
	Absolut	1976 = 100	
76	16	100	
77	18,2	113,75	
78	23	143,75	

- c) In Branche B, weil für die Wachstumsrate der Umsatzrentabilität **näherungsweise** (nur bei kleinen, nicht wie im Beispiel großen Wachstumsraten) gilt:

Wachstumsrate des Umsatzes minus Wachstumsrate des Kapitaleinsatzes,

also Branche B: $0,5 - 0,2 = 0,3$

Branche A: $0,55 - 0,3 = 0,25$.

Die genauen Werte sind nicht 30 % und 25 % sondern 25 % und 19,2 %

	Branche A			Branche B		
	1978	1979	W-rate	1978	1979	W-rate
U	20	31	55%	3	4,5	50%
K	100	130	30%	25	30	20%

U/K	0,2	0,24	19,2%	0,12	0,15	25%
-----	-----	------	-------	------	------	-----

Die Umsatzrentabilität ist in Branche B stärker gestiegen.

d) Verhältniszahl, Beziehungszahl, Maßzahl, also: R, F, F, R, F, F, F, R.

e) Index für 1977

$$\frac{14 \cdot 10 + 4,2 \cdot 6}{10 \cdot 10 + 6 \cdot 6} = 1,215 \quad \text{also } 121,5 \text{ vH}$$

für das Jahr 1978

$$\frac{20 \cdot 10 + 3 \cdot 6}{136} = 1,603 \quad \text{also } 160,3 \text{ vH}$$

Ein solcher Umsatzindex (nach Art der amtlichen Statistik) steigt stärker an als die Messzahl der Umsätze (Teil b), weil er von Änderungen der Umsatzstruktur, gemessen an den Umsatzanteilen in vH

Jahr	A	B
1976	62,5	37,5
1977	76,9	23,1
1978	87,0	13,0

zugunsten der umsatzstärkeren Branche A nicht beeinflusst wird.

Lösung zu Aufgabe 10

a) Es sind sehr genau zwei Dinge auseinanderzuhalten (was häufig im ökonomischen „Alltagssprachgebrauch“ übersehen wird):

- Der Preisindex (nach Laspeyres) für die Lebenshaltung mißt die Zunahme der Preise bei konstanten Mengen.
- Die Lebenshaltungskosten sind Preise multipliziert mit den jeweiligen Mengen.

Wenn die Mengensteigerungen größer sind als die Preissteigerungen - oder die Preise sogar sinken - sind die Voraussetzungen für den beschriebenen Tatbestand gegeben.

b) Berechnung der Veränderung der Lebenshaltungskosten:

$$\frac{\sum p_t \cdot q_t}{\sum p_0 \cdot q_0} = \frac{80 \cdot 9 + 50 \cdot 7 + 150 \cdot 12 + 565 \cdot 2}{100 \cdot 10 + 50 \cdot 8 + 200 \cdot 15 + 300 \cdot 2} = \frac{4000}{5000} = \frac{4}{5} = 0,8$$

Es liegt eine Abnahme der Lebenshaltungskosten um 20 % vor. Den Wert 0,8 bezeichnet man als Wertmesszahl (W).

c) Berechnung des Laspeyres-Preisindex (n = 4 Warenarten):

$$P_{0t}^L = \frac{\sum_{j=1}^n p_{jt} \cdot q_{j0}}{\sum_{j=1}^n p_{j0} \cdot q_{j0}} = \frac{9 \cdot 100 + 7 \cdot 50 + 12 \cdot 200 + 2 \cdot 300}{10 \cdot 100 + 8 \cdot 50 + 15 \cdot 200 + 2 \cdot 300} = \frac{4250}{5000} = 0,85$$

d) Berechnung des Paasche-Mengenindex:

$$Q^P \cdot p^L = W = Q^P \cdot 0,85 = 0,8 \rightarrow Q^P = \frac{0,8}{0,85} = 0,9411$$

e) Die Preismesszahlen der vier Warenarten sind:

Ware	Preismesszahl
A	$9/10 = 0,9$
B	$7/8 = 0,875$
C	0,8
D	1

Wenn der Haushalt von K zur Berichtszeit ausschließlich das Gut C konsumiert, erhält man die Untergrenze des Paasche-Preisindex mit 0,8. Entsprechend ist die Obergrenze genau 1 unter der Voraussetzung, dass der Haushalt seine Konsumstruktur dergestalt ändert, dass er ausschließlich das Produkt D konsumiert. Man erhält also folgende Einschränkung $0,8 \leq P^P \leq 1,0$.

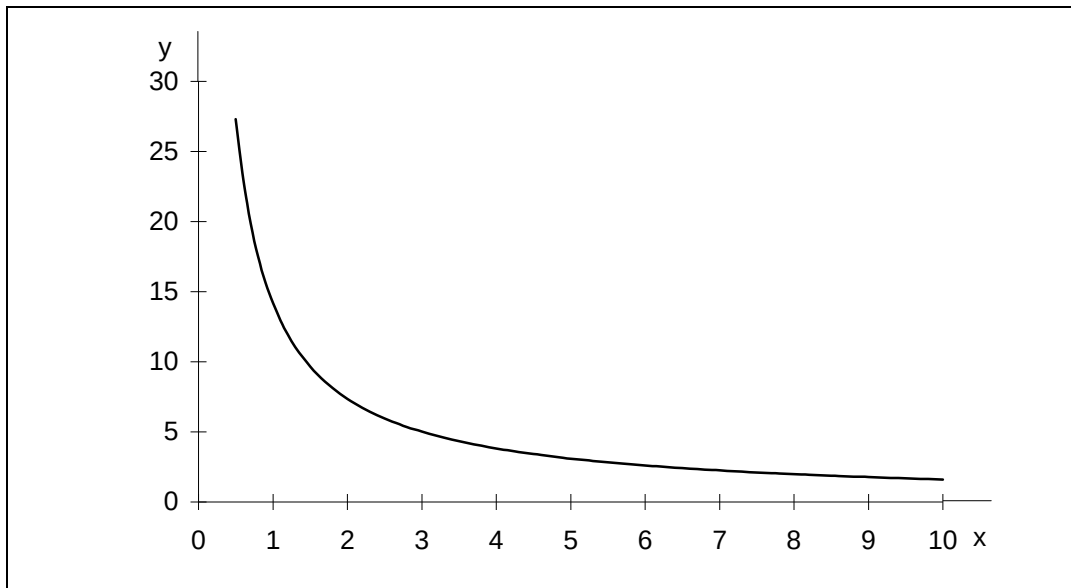
Wie immer die Konsumstruktur des Haushalts zur Berichtszeit sein mag, der Paasche-Preisindex kann nicht kleiner sein als 0,8 (also 80%) und nicht größer sein als 1. Rechnet man mit den Zahlen von Teil b, so erhält man 0,8734, einen Wert also, der innerhalb der oben angegebenen Schranken liegt.

Der Paasche-Preisindex wird in der Regel niedriger sein, als der Laspeyres-Preisindex.

Begründung: – unterschiedliche Gewichtung
– Substitutionen dergestalt, dass vom relativ teurer werdenden Produkt relativ weniger konsumiert wird.

Lösung zu Aufgabe 11

a) Vgl. Abb. auf der nächsten Seite für die unter b) zu schätzende Indifferenzkurve. Die Indifferenzkurve verläuft konvex vom Ursprung aus gesehen.



b) $\ln \hat{y} = \ln a + b \ln x$

Man gehe zweckmäßig wie folgt vor (Arbeitstabelle):

$\ln x$	$\ln y$	$(\ln x)^2$	$\ln x \cdot \ln y$
1,4	1,2	1,96	1,68
1,2	1,6	1,44	1,92
0,7	2	0,49	1,4
2,1	0,7	4,41	1,47
5,4	5,5	8,3	6,47

Die Normalgleichungen lauten dann

$$5,5 = 4 \ln a + b \cdot 5,4$$

$$6,47 = 5,4 \ln a + b \cdot 8,3$$

Folglich ist $\ln a = 2,65$ (also $a = 14,175$) und $b = -0,946$.

Die Indifferenzkurve $\hat{y} = 14,175x^{-0,946}$ verläuft also tatsächlich konvex.

c) Laspeyres Index $\frac{120}{100} = 1,2$, Paasche Index $\frac{148}{154} = 0,961$

d) Weil die Indifferenzkurve konvex gekrümmt ist. Wäre sie (was rationalem Verbraucherverhalten wenig entspräche) linear, so wären die Indizes gleich, wäre sie konkav, so müsste der Laspeyres-Index kleiner sein als der Paasche-Index. Bei rationaler Substitution ist der Laspeyres-Index jedoch stets größer als der Paasche Index. Rechnet man die Angaben für die Periode 3 in Prozent derjenigen für Periode 1 so ergeben sich folgende Wachstumsraten

	Gut x	Gut y
Preisänderung	- 19,4%	100%
Mengenänderung	100%	- 39,4%

so dass das erheblich stärker verteuerte Gut y auch weniger nachgefragt wird als das im

Preis gesunkene Gut x vermehrt nachgefragt wird (Substitution an einer konvexen Indifferenzkurve).

- e) Weil die unter b) bestimmte Indifferenzkurve nichtlinear ist, liegt ein Fall nichtlinearer Regression (und damit auch Korrelation) vor. Der Korrelationskoeffizient wird dann als positive Wurzel des Bestimmtheitsmaßes berechnet und man könnte (allgemein bei nichtlinearer Regression) deshalb auch nicht sinnvoll von negativer Korrelation sprechen. Es ist also stets $r > 0$ auch wenn eine Regressionsfunktion eine monoton fallende Kurve darstellt. Richtig ist also:

Die Korrelation zwischen x und y ist

- ☐ linear ☒ positiv
☒ nichtlinear ☐ negativ

Lösung zu Aufgabe 12

- a) 3 Merkmale und 2 zweidimensionale Häufigkeitsverteilungen

Merkmal	Skala	graph. Darstellung
Alter	Ratioskala	Histogramm bzw. Alterspyramide
Geburtsort, Todesursache	Nominalskala	Kreisdiagramm

- b)

Alter X	kumulierte relative Häufigkeit
0 - 40	1/6
40 - 60	1/3
über 60	1

← Medianklasse

bei einer Klassenobergrenze von $X_{j0} = 100$:

$$Z = X_{ju} + \frac{0,5 - H_{j-1}}{h_j} b_j$$

$$Z = 60 + \frac{0,5 - 0,33}{0,66} \cdot 40 = 70$$

bei $X_{j0} = 90$: $\rightarrow Z = 67,5$

Da die Medianklasse die letzte Größenklasse ist („über 60 Jahre“), hängt das Interpolationsergebnis in diesem Fall davon ab, bei welchem Wert (z. B. bei 100 oder bei 90 Jahren) diese Klasse geschlossen wird.

- c) Modus und evtl. altersspezifische Sterberaten (was jedoch neben den mitgeteilten Zahlen für die Gestorbenen auch die entsprechenden Zahlen für die Lebenden der betreffenden Altersklassen voraussetzt). Da keine Zeitreihe vorliegt und das Alter nicht sinnvoll als „verteilt“ bzw. „umverteilt“ gedacht werden kann, sind die Antwortmöglichkeiten 1 und 2 falsch. Die Antworten 4 und 5 entsprechen nicht der Fragestellung.

- d) Nein, dieser Schluss ist unzulässig. Es handelt sich hier um ein Problem der Scheinkorrelation. Sowohl das Sterben als auch der Aufenthalt im Bett werden von anderen Faktoren beeinflusst, wie etwa der Altersstruktur, der Gesundheit u. ä.
- e) $U_t = 5 + 18t - t^2$
 Folglich ist die Wachstumsrate von U_t zu berechnen als

$$\frac{U'_t}{U_t} = \frac{18 - 2t}{5 + 18t - t^2} \quad \text{oder} \quad \frac{2A}{86 - A^2}, \text{ wenn } A = 9 - t \text{ ist. Der Wert für } U_t \text{ zur Zeit } t = 3 \text{ ist}$$

$$U_3 = 50. \text{ Setzt man ihn 100, so ist die Funktion der Messzahlen einfach } 10 + 36t - 2t^2.$$
- f) gleich.

Lösung zu Aufgabe 13

- a) Der Schluss ist falsch, weil die Warenkörbe auch zur Basiszeit einen ganz unterschiedlichen Wert hatten. Deshalb kann man nicht absolute Ausgaben vergleichen, sondern stets nur relative (z. B. Indizes) Größen. Im Beispiel ist (wie noch unter Teil b) deutlich wird) mit folgenden Gesamtausgaben gerechnet worden (in E-Dollar)

Klasse A: 100
 Klasse B: 1200.

In Verbindung mit den Ausgabenanteilen und Preisen des Basisjahres ($t = 0$) lassen sich daraus leicht Ausgaben und Mengen für die einzelnen Warenarten errechnen.

- b) Die Preisindizes ergeben sich als arithmetische Mittel der Preismesszahlen (aufgrund der angegebenen Wachstumsraten) gewogen mit den Ausgabenanteilen der Basisperiode, also

für Klasse A: $2,5 \cdot 0,4 + 1,2 \cdot 0,1 + 1,8 \cdot 0,4 + 1,1 \cdot 0,1 = 1,95$
 für Klasse B entsprechend: $0,25 + 0,48 + 0,36 + 0,33 = 1,42$

Hieraus errechnet sich, dass die Gesamtausgaben zur Zeit $t = 0$ für Klasse A 100 und für Klasse B 1200 waren (denn 42% von 1200 sind 504). Man beachte, dass der Preisindex für A stärker gestiegen ist als für B. Es gilt also gerade nicht, was unter a) behauptet wurde.

- c) F, F, R, denn die Lebenshaltungskosten ergeben sich aufgrund der laufenden Preise (die bekannt sind) und der laufenden (zur Zeit $t = 1$ tatsächlich verbrauchten) Mengen. Man kennt aber aus den Teilen a und b nur die Verbrauchsstruktur der Basiszeit.
- d) Aus den Anteilen und dem Gesamtwert von 2000 sowie den aus Teil a und b bekannten Preisen lässt sich errechnen

	Ausgaben	Preise	Mengen
Miete	300	25	12
Gebrauchsg.	900	3	300
Lebensmittel	360	9	40
Kleidung	440	11	40

Daraus erhält man einen Paasche-Preisindex von

$$\frac{2000}{10 \cdot 12 + 2,5 \cdot 300 + 5 \cdot 40 + 10 \cdot 40} = 1,361$$

- e) Die Haushalte der Klasse B konnten hohen Preissteigerungen (Miete, Lebensmittel) durch Abnahme (Lebensmittel) oder Konstanz der Verbrauchsmengen (Miete) ausweichen und bei weniger stark im Preis gestiegenen Gütern (insbesondere Gebrauchsgütern) ihren Konsum erheblich ausweiten.

Lösung zu Aufgabe 14

- a) Normalgleichungen

$$\left. \begin{array}{l} 11a + 133,9b = 388 \\ 133,9a + 1964,61b = 6204,7 \end{array} \right\} \begin{array}{l} a = -18,6176 \\ b = 4,4271 \end{array}$$

somit $\hat{y} = -18,6176 + 4,4271x$.

$$b) \quad r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} = 0,8417$$

wobei $(\sum x)^2 = 17929,21$

$$\sum y^2 = 22944$$

$$(\sum y)^2 = 150544$$

- c) $\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-2}$ worin $\hat{u}$ die Störgröße in der Stichprobenregressionsfunktion ist.

Man erhält $\hat{\sigma}^2$ auch durch Berechnung von r^2 . Für $\hat{u}$ errechnet sich:

y	3	11	12	12	12	20	33	52	70	83	80
$\hat{y}$	24,3	3,9	20,3	25,6	-0,9	41,2	27	50,9	67,3	84,1	44,2
u	-21,3	7,1	-8,3	-13,6	12,9	-21,2	6	1,1	2,7	-1,1	35,8

Man erhält $\sum \hat{u}^2 = 2698,575$ und folglich $\hat{\sigma}^2 = 299,84$.

$$d) \quad r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} = 0,8114$$

$$r_{yz.x} = \frac{r_{yz} - r_{yx} \cdot r_{zx}}{\sqrt{(1 - r_{yx}^2)(1 - r_{zx}^2)}} = 0,2393$$

$$r_{xz.y} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{(1 - r_{xy}^2)(1 - r_{zy}^2)}} = 0,0238$$

Die Gegenüberstellung von $r_{xy} = 0,8417$ und $r_{xy} = 0,8114$ zeigt, dass die Beziehung zwischen den Variablen x und y auf eine echte Korrelation schließen lässt und nicht auf eine Scheinkorrelation vermittelt durch z.

$$\text{e) } R^2_{y,xz} = \frac{r^2_{xy} + r^2_{zy} - 2r_{xy} \cdot r_{zy} \cdot r_{xz}}{1 - r^2_{xz}} = 0,7253$$

Die multiple Bestimmtheit beträgt 72,53%. Offenbar trägt z wesentlich weniger zur Erklärung von y bei als x ($r_{yz} = 0,4425 < r_{yx} = 0,8417$). Wäre $r_{xz} = 0$ so wäre die multiple Bestimmtheit $R^2_{y,xz} = r^2_{xy} + r^2_{zy} = 0,9043$. Der Unterschied zu der geringeren tatsächlichen Bestimmtheit von 72,53% liegt also an der relativ hohen Korrelation $r_{xz} = 0,3840$, was auch sachlich zu begründen ist: wenn die Regierung Entwicklungsaufträge erteilt, wird sie in der Regel auch Zusagen machen über finanzielle Hilfen.

Teil IV: Drei Musterklausuren

Hauptklausur SS 97

Aufgabe 1:

15 Schwimmer verschiedenen Alters erbrachten in Wettkämpfen über 50 m Brustschwimmen die folgenden Zeiten:

Alter	Zeiten in sec.		
28	37.00	40.00	42.00
30	39.00	40.00	45.00
32	40.00	38.00	42.00
34	42.00	44.00	45.00
36	42.00	46.00	48.00

- a) Stellen Sie eine Häufigkeitstabelle auf, und bestimmen Sie die Randhäufigkeiten.
(3 Punkte)
- b) Bestimmen Sie die Kovarianz zwischen dem Alter der Schwimmer (X) und der Zeit in sec. (Y). Sind die Variablen unabhängig?
(6 Punkte)
- c) Der Schwimmlehrer A vermutet einen linearen Zusammenhang zwischen dem Alter und der Leistungsfähigkeit der Schwimmer:

$$y_i = a + bx_i$$
 Schätzen Sie die unbekannten Regressionsparameter a und b und bestimmen Sie die Regressionsgerade $\hat{y}$.
(4 Punkte)
- d) Bestimmen Sie die mittlere Schwimmzeit, unter der Bedingung, daß nur die 30jährigen Schwimmer betrachtet werden.
(4 Punkte)

Aufgabe 2:

1. Der Chef eines Unternehmens möchte sich über die täglich anfallenden Anfahrtswege seiner Mitarbeiter informieren. Vom Personalbüro wird ihm dazu folgende Tabelle übersandt:

km	Anzahl der Beschäftigten			
[0,1)	7			
[1,5)	24			
[5,15)	35			
[15,30)	18			
[30,50)	16			

- a) Wie groß sind das arithmetische Mittel (näherungsweise!), der interpolierte Median und der Modus der klassierten Verteilung?

(3+3+3=9 Punkte)

- b) Bestimmen Sie näherungsweise die Varianz der klassierten Verteilung. Warum kann die Varianz nur näherungsweise bestimmt werden?

(4 Punkte)

- c) Beurteilen Sie aufgrund der Lagemaße die Schiefe der Verteilung.

(2 Punkte)

2.

- a) Wie groß ist die folgende Kovarianz C zwischen Preismeßzahlen a_i und Mengenmeßzahlen b_i , gewogen mit den Ausgabenanteilen g_i zur Basiszeit?

$$C = \sum (a_i - P_{0t}^L)(b_i - Q_{0t}^L)g_i \quad \text{mit: } g_i = \frac{P_{i0}Q_{i0}}{\sum P_{i0}Q_{i0}}$$

(2 Punkte)

- b) Nach dem hiermit bestimmten Zusammenhang (nach L. v. Bortkiewicz) läßt sich aus den folgenden Angaben

- nominale Zunahme 50% (also $W_{0t} = 1,5$)
- reale (volumenmäßige) Zunahme 20% (also $Q_{0t}^L = 1,2$)
- Kovarianz $C = +0,12$

der Laspeyres-Preisindex wie folgt berechnen:

(2 Punkte)

Wie groß ist P_{0t}^L , wenn für die Kovarianz gilt $C = -0,24$ (die anderen Angaben bleiben unverändert)?

(1 Punkt)

- c) Nennen Sie drei Situationen, in denen ein Lasperes-Index den gleichen Zahlenwert an-

nimmt wie ein Paasche-Index, etwa $P_{0t}^L = P_{0t}^P = 1,4$?

(3 Punkte)

Aufgabe 3:

1. Fünf Personen A, B, C, D und E treffen sich zu einer Pokerpartie. Zu Beginn ($t = 1$) hat jeder 200 DM. Nach einer Stunde ($t = 2$) sieht die Verteilung des Geldes wie folgt aus:

Person	A	B	C	D	E
DM	50	100	150	200	500

Das Spiel endet nach zwei Stunden ($t = 3$) damit, daß E den gesamten Einsatz gewonnen hat und die anderen vier „pleite“ sind.

Beschreiben Sie die Disparität zu den drei Zeitpunkten sowohl graphisch als auch numerisch. (12 Punkte)

2. In den zwei Stunden werden insgesamt 10 Spiele gemacht. E hat nach den einzelnen Spielen jeweils folgende Geldbeträge vor sich liegen:

Spiel	1	2	3	4	5	6	7	8	9	10
DM	180	220	260	230	250	300	500	480	600	1000

- a) Geben Sie die durchschnittliche Wachstumsrate des Geldbetrags an.

(2 Punkte)

- b) Bestimmen Sie den Trend in dieser Zeitreihe mit Hilfe eines gleitenden Durchschnitts ($p = 3$) sowie die trendbereinigten Werte. (6 Punkte)

Hauptklausur SS 98

Aufgabe 1

- a) Gegeben sei die folgende Verteilung:

Klasse von... bis unter...	n_k	$\bar{x}_k$	s_k^2				
0 - 20	20	12	0,3				
20 - 35	12	25	1,4				
35 - 50	6	42	2,6				
50 - 90	10	70	3				
90 - 120	2	100	4,2				

- i) Stellen Sie die relativen Häufigkeiten graphisch dar.

(2 Punkte)

- ii) Berechnen Sie das arithmetische Mittel $\bar{x}$.

(2 Punkte)

- iii) Berechnen Sie die Standardabweichung. (4 Punkte)
- iv) Bestimmen Sie den Median und den Modus der Verteilung. (3 + 2 Punkte)
- b) Student S arbeitet ein halbes Jahr als studentische Hilfskraft. Dabei verdient er im Monat (in DM):

Januar	Februar	März	April	Mai	Juni
467	523	399	602	488	478

Berechnen Sie die Spannweite, den Quartilsabstand und das Gini-Dispersionsmaß.
(1 + 3 + 3 Punkte)

Aufgabe 2:

- a) Bei einer Lotterie mit 350 Teilnehmern gibt es drei verschiedene Gewinnbeträge $G_1 < G_2 < G_3$ (keine Nieten!). Dabei ist G_3 doppelt so hoch wie G_2 und dieser wiederum doppelt so hoch wie G_1 . Den Betrag G_1 erhalten 200 Teilnehmer, den Betrag G_2 100 Teilnehmer und den Betrag G_3 entsprechend 50 Teilnehmer. Insgesamt werden 600.000 DM ausgeschüttet.
- i) Bestimmen Sie die drei Gewinnbeträge. (3 Punkte)
- ii) Zeichnen Sie die Lorenzkurve. (3 Punkte)
- iii) Berechnen Sie den Gini-Koeffizienten. (3 Punkte)
- b) Bei einer Umfrage wurden 100 Personen gefragt, ob sie ein Auto (Variable X) und einen Videorekorder (Variable Y) besitzen. Dabei antworteten:

Auto	
ja	60
nein	40

Videorekorder	
ja	45
nein	55

- i) Bestimmen Sie für die Variable X ein sinnvolles Lagemaß. (2 Punkte)
- ii) Angenommen, es handele sich bei den beiden Verteilungen um die Randverteilungen einer zweidimensionalen Häufigkeitsverteilung. Wie müßte diese aussehen, wenn die beiden Variablen unabhängig wären? (2 Punkte)
- iii) Wieviel Prozent der Autobesitzer besitzen in diesem Fall (Unabhängigkeit) auch einen Videorekorder? (2 Punkte)
- iv) Handelt es sich bei der Variable Y um eine Bestands- oder Bewegungsmasse und um eine diskrete oder stetige Variable? Wie ist sie skaliert? (3 Punkte)

- v) Zur leichteren Auswertung ordnet der Befrager der Antwort „nein“ den Wert 0 und der Antwort „ja“ den Wert 1 zu. Welcher Transformation dürfte er die Variablen unterziehen? Nennen Sie ein Beispiel für diese Form der Transformation. (2 Punkte)

Aufgabe 3

- a) Zwei Variablen X und Y seien in der Form linear voneinander abhängig, daß folgende Regressionsgerade gilt:

$$\hat{y}_i = 100 + 5 \cdot x_i$$

Desweiteren ist $s_y^2 = 4000$ und $s_x^2 = 100$.

- i) Wie groß ist das Bestimmtheitsmaß? (4 Punkte)
- ii) Bestimmen Sie die Kovarianz. (3 Punkte)

- b) Für zwei Variablen X und Y gelten die Regressionsgeraden

$$\hat{y}_i = 3,5 + 0,5 \cdot x_i \quad \text{und}$$

$$\hat{x}_i = \frac{32}{22} + \frac{13}{22} y_i.$$

- i) Bestimmen Sie den Korrelationskoeffizienten. (3 Punkte)
- ii) Bestimmen Sie $\bar{x}$ und $\bar{y}$. (4 Punkte)
- c) Ein Warenkorb enthält fünf Produkte (A, B, C, D, E). Die folgende Tabelle gibt einen Überblick über die Preismeßzahlen (m_{ot}) der einzelnen Produkte, sowie über ihren Anteil an den Gesamtausgaben zur Basisperiode (g_i).

Produkt	m_{ot}	g_i
A	1,1	0,1
B	0,9	0,3
C	1,2	0,05
D	1,15	0,45
E	0,95	0,1

- i) Berechnen Sie einen Preisindex nach Laspeyres. (3 Punkte)
- ii) Der dazugehörige Mengenindex nach Laspeyres betrage $Q_{ot}^L = 1,1$, die Kovarianz zwischen Preis- und Mengenmeßzahl sei $C = -0,25$. Wie hoch ist der Wertindex? (3 Punkte)

Nachklausur SS 98

Aufgabe 1

- a) Die Werbeagentur des Netzanbieters „Schnurlos glücklich“ schätzt den diesjährigen Absatz von Handys anhand der folgenden stetigen Funktion:

$$y(t) = \frac{t^2}{4} - 2t + 10$$

wobei $y(t)$ die im Zeitpunkt t verkaufte Anzahl von Handys (in Tausend) darstellt und t die Anzahl der Monate.

- i) Welche Wachstumsrate ergibt sich nach genau einem Jahr? (3 Punkte)
- ii) Ein Mitarbeiter des Netzanbieters ist für die Überwachung der Vorhersagen der Werbeagentur zuständig. Ein Jahr nach der Prognose liegen ihm folgende Monatsverkaufszahlen vor.

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	10	8	7	6	4	5	8	10	12	14	16	20
r_t												
w_t												

Berechnen Sie die tatsächlichen Wachstumsraten und -faktoren. Wie hoch ist die mittlere Wachstumsrate? (2 + 2 Punkte)

- iii) Nach wieviel Jahren hat sich die gesamte Absatzmenge dieses Jahres bei einem konstanten Wachstum von 10% auf 300.000 erhöht? (4 Punkte)
- b) Das Personalbüro von „Schnurlos glücklich“ hat die Personalentwicklung der Firma getrennt nach Angestellten (x) und Arbeitern (y) ermittelt. Dummerweise sind die Daten nicht vollständig. Es ist jedoch bekannt, daß die mittlere Anzahl der Angestellten 577,50 beträgt.

t	1	2	3	4	5	6	7	8
Arbeiter (y_t)	68	75	84	93	86	72	66	54
Angestellte (x_t)	?	512	537	539	574	608	654	700

- i) Bestimmen Sie die Kovarianz zwischen der Anzahl der Arbeiter und der der Angestellten. (5 Punkte)
- ii) Wie groß ist die Korrelation zwischen den Beschäftigten dieser beiden Berufsgruppen? (4 Punkte)

Aufgabe 2:

- a) Der Chef eines Unternehmens möchte sich über die täglich anfallenden Anfahrtswege seiner Angestellten informieren. Vom Personalbüro erhält er die folgenden Werte:

km von ... bis unter ...	Anzahl der Angestellten
0 - 1	49

1 - 5	168
5 - 15	245
15 - 30	126
30 - 50	112

Wie groß sind das arithmetische Mittel, der Median und der Modalwert der Entfernung zum Arbeitsplatz. Entscheiden Sie mit Hilfe der Fechner'schen Lageregel, ob die Häufigkeitsverteilung symmetrisch, linkssteil oder rechtssteil ist.

(7 Punkte)

- b) Für die drei Güter eines Warenkorbes wurde in den Jahren 1989 bis 1993 folgende Preise und Mengen notiert:

	Gut 1		Gut 2		Gut 3	
Jahr	Preis	Menge	Preis	Menge	Preis	Menge
1989	10	100	60	100	30	100
1990	12	110	60	97	19	120
1991	12	100	50	100	18	100
1992	15	80	65	90	24	90
1993	20	50	70	80	30	60

- i) Berechnen Sie für 1993 zur Basis 1989 die Preisindizes von Laspeyres und Paasche.
(6 Punkte)
- ii) Berechnen Sie den Wertindex ($0 = 1989$, $t = 1993$) und leiten Sie daraus den Paasche-Mengenindex ab.
(5 Punkte)
- iii) Zeigen Sie mit Hilfe der Ergebnisse aus i) und ii), daß Indizes nach Paasche die Faktorumskehrprobe **nicht** erfüllen.
(3 Punkte)

Aufgabe 3:

- a) Bei der Fußballweltmeisterschaft in Frankreich gab es einen Spieler, der sechs Tore geschossen hat, zwei mit fünf Treffern, drei mit vier, sieben mit drei Toren, 23 Spieler trafen zweimal und 72 Aktive je einmal ins Schwarze. Es gab also insgesamt 108 Torschützen, die 167 Treffer erzielten.

Anmerkung: Spieler, die gar nicht getroffen haben, werden natürlich nicht berücksichtigt.

- i) Geben Sie in der untenstehenden Tabelle die Punkte der Lorenzkurve und die Steigungen der einzelnen Abschnitte an (nicht zeichnen!)
(5 Punkte)
- ii) Berechnen Sie den Gini-Koeffizienten.
(3 Punkte)

Hinweis: Rundung auf vier Nachkommastellen.

FRANCE 98 Fußball-Weltmeisterschaft		
Torschützen		
Suker	Kroatien	6
Batistuta	Argentinien	5
Vieri	Italien	5
Hernandez	Mexiko	4
Ronaldo	Brasilien	4
Salas	Chile	4
Bebeto	Brasilien	3
Bergkamp	Niederlande	3
Bierhoff	Deutschland	3
C. Sampaio	Brasilien	3
Henry	Frankreich	3
Klinsmann	Deutschland	3
Rivaldo	Brasilien	3
23	Spieler erzielten	2 Tore
72	Spieler erzielten	1 Tor

WAZ vom 14.07.98

b) Gegeben seien folgende Umsatzzahlen eines Unternehmens (in 1000 DM):

Jahr	1989	1990	1991	1992	1993	1994	1995	1996	1997
Umsatz	187	175	184	205	193	226	217	229	202

- i) Stellen Sie die Zeitreihe graphisch dar. (2 Punkte)
- ii) Ermitteln Sie die Trendgerade mittels der Methode der kleinsten Quadrate. (4 Punkte)
- iii) Geben Sie mit Hilfe der Trendgeraden aus Aufgabe ii) eine Schätzung für das Jahr 1997 ab. (2 Punkte)
- iv) Geben Sie mit Hilfe der Methode des exponentiellen Glättens eine Prognose für das Jahr 1998 ab. Verwenden Sie dabei den unter iii) errechneten Schätzwert und den tatsächlich beobachteten Wert für das Jahr 1997. Gewichten Sie beide Werte gleich. (3 Punkte)

Lösungen zu den Musterklausuren

Hauptklausur SS 97

Aufgabe 1

a)

		Alter					
		28	30	32	34	36	Σ
Zeit	37	1					1
	38			1			1
	39		1				1
	40	1	1	1			3
	42	1		1	1	1	4
	44				1		1
	45		1		1		2
	46					1	1
	48					1	1
	Σ	3	3	3	3	3	15

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

$$\bar{x} = \frac{1}{n} \sum x_i n_i = \frac{1}{15} \cdot 3 \cdot (28 + 30 + 32 + 34 + 36) = 32$$

$$\bar{y} = \frac{1}{n} \sum y_i n_i = \frac{1}{15} (37 + 38 + 39 + 3 \cdot 40 + 4 \cdot 42 + 44 + 2 \cdot 45 + 46 + 48) = 42$$

$$\begin{aligned} s_{xy} &= \frac{1}{15} (28 \cdot 37 + 28 \cdot 40 + 28 \cdot 42 + 30 \cdot 39 + 30 \cdot 40 + 30 \cdot 45 + 32 \cdot 40 + 32 \cdot 38 + 32 \cdot 42 \\ &\quad + 34 \cdot 42 + 34 \cdot 44 + 34 \cdot 45 + 36 \cdot 42 + 36 \cdot 46 + 36 \cdot 48) - 32 \cdot 42 \\ &= \frac{1}{15} 20242 - 1344 = 5,4\bar{6} \end{aligned}$$

Da $s_{xy} \neq 0$ ist, sind die Variablen X und Y abhängig.

$$c) \quad b = \frac{s_{xy}}{s_x^2}$$

$$s_x^2 = \frac{1}{n} \sum x_i^2 n_i - \bar{x}^2 = \frac{1}{15} \cdot 3 (28^2 + 30^2 + 32^2 + 34^2 + 36^2) - 32^2$$

$$= 1032 - 1024 = 8$$

$$b = \frac{5,4\bar{6}}{8} = 0,68\bar{3}$$

$$a = \bar{y} - b \cdot \bar{x} = 42 - 0,68\bar{3} \cdot 32 = 20,1\bar{3}$$

$$\hat{y}_i = 20,1\bar{3} - 0,68\bar{3} \cdot x_i$$

$$d) (\bar{x}|y=30) = \frac{1}{3}(39 + 40 + 45) = \frac{124}{3} = 41,3$$

Aufgabe 2

1.

	m_i	n_i	$m_i n_i$	H_i	n_i/b_i
[0,1)	0,5	7	3,5	0,07	7
[1,5)	3	24	72	0,31	6
[5,15)	10	35	350	0,66	3,5
[15,30)	22,5	18	405	0,84	1,2
[30,50)	40	16	640	1	0,8
Summe	-	100	1470,5	-	-

$$a) \bar{x} = \frac{1}{n} \sum m_i n_i = \frac{1}{100} 1470,5 = 14,705$$

$$\tilde{x}_{0,5} = x'_{k-1} + b_k \frac{0,5 - H_{k-1}}{h_k} = 5 + 10 \frac{0,5 - 0,31}{0,35} = 5 + 5,429 = 10,429$$

$$x_{\text{mod}} = 0,5$$

$$b) s^2 = \frac{1}{n} \sum m_i^2 n_i - \bar{m}^2 = \frac{1}{100} (0,5^2 \cdot 7 + 3^2 \cdot 24 + 10^2 \cdot 35 + 22,5^2 \cdot 18 + 40^2 \cdot 16) - 14,705^2$$

$$= 384,3025 - 14,705^2 = 168,065$$

Die Varianz kann nur näherungsweise berechnet werden, weil die interne Varianz unbekannt ist.

$$c) x_{\text{mod}} < \tilde{x}_{0,5} < \bar{x} \Rightarrow \text{Die Verteilung ist linkssteil.}$$

2.

$$a) C = Q_{0t}^L (P_{0t}^P - P_{0t}^L) \text{ oder } C = P_{0t}^L (Q_{0t}^P - Q_{0t}^L)$$

$$b) C = Q_{0t}^L (P_{0t}^P - P_{0t}^L)$$

$$\Rightarrow P_{0t}^L = -\frac{C}{Q_{0t}^L} + P_{0t}^P = \frac{W_{0t}}{Q_{0t}^L} - \frac{C}{Q_{0t}^L} = \frac{1,5}{1,2} - \frac{0,12}{1,2} = 1,25 - 0,1 = 1,15$$

2. Variante:

$$P_{0t}^L = \frac{1,5}{1,2} + \frac{0,24}{1,2} = 1,25 + 0,2 = 1,45$$

- c) 1. alle Preise steigen um 40%
2. alle Mengen bleiben gleich ($q_{it} = q_{i0} \forall i$)
3. Die Kovarianz zwischen den Preis- und Mengenmeßzahlen ist Null.

Aufgabe 3

1.

Für $t = 1$: $D_G = 0$

Für $t = 2$:

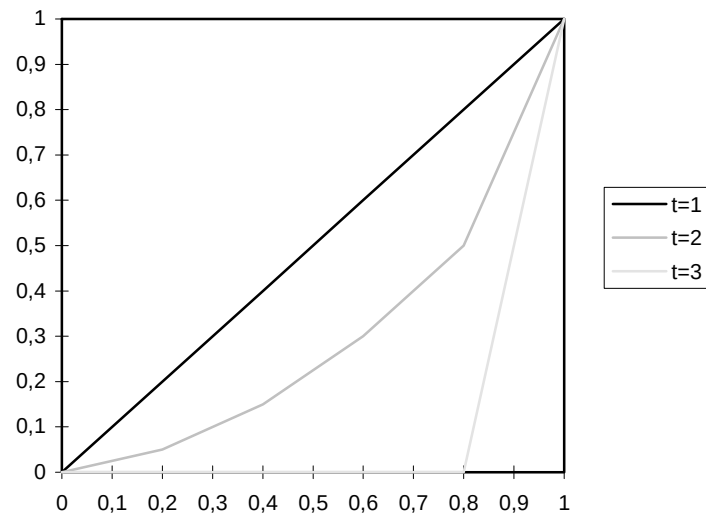
	x_i	h_i	H_i	q_i	Q_i	$(H_i + H_{i-1})q_i$
A	50	0,2	0,2	0,05	0,05	0,01
B	100	0,2	0,4	0,1	0,15	0,06
C	150	0,2	0,6	0,15	0,3	0,15
D	200	0,2	0,8	0,2	0,5	0,28
E	500	0,2	1	0,5	1	0,9
Summe	1000	1	-	1	-	1,4

$$D_G = \sum (H_i + H_{i-1})q_i - 1 = 1,4 - 1 = 0,4$$

Für $t = 3$

	x_i	h_i	H_i	q_i	Q_i	$(H_i + H_{i-1})q_i$
A	0	0,2	0,2	0	0	0
B	0	0,2	0,4	0	0	0
C	0	0,2	0,6	0	0	0
D	0	0,2	0,8	0	0	0
E	1000	0,2	1	1	1	1,8
Summe	1000	1	-	1	-	1,8

$$D_G = \sum (H_i + H_{i-1})q_i - 1 = 1,8 - 1 = 0,8$$



2.

a) $\bar{w} = \sqrt[10]{1000/200} = 1,1746 = 17,46\%$

b)

y	$\tilde{y}$	$y - \tilde{y}$
200	-	-
180	200	-20
220	220	0
260	236,67	23,33
230	246,67	-16,67
250	260	-10
300	350	-50
500	426,67	73,33
480	526,67	-46,67
600	693,33	-93,33
1000	-	-

Hauptklausur SS 98

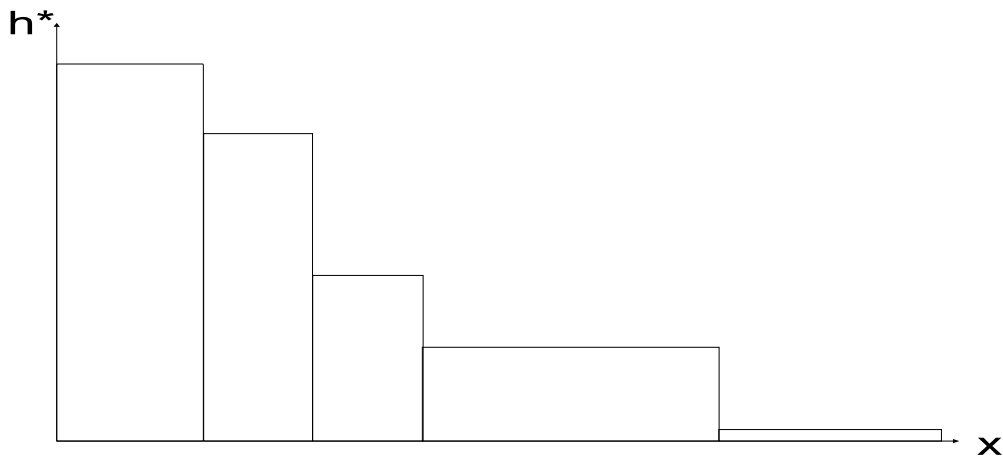
Aufgabe 1

a)

	n_k	h_k	h_k^*	$(\bar{x}_k - \bar{x})^2 h_k$	H_k
0 - 20	20	0,4	0,02	190,79	0,4
20 - 35	12	0,24	0,016	18,75	0,64
35 - 50	6	0,12	0,008	7,99	0,76
50 - 90	10	0,2	0,005	261,51	0,96

90 - 120	2	0,04	0,0013	175,09	1
Σ	50	1		654,13	

i)



$$\text{ii) } \bar{x} = \sum \bar{x}_k h_k = 12 \cdot 0,4 + 25 \cdot 0,24 + 42 \cdot 0,12 + 70 \cdot 0,2 + 100 \cdot 0,04 = 33,84$$

$$\text{iii) } s_{\text{int}}^2 = \sum s_k^2 h_k = 0,3 \cdot 0,4 + 1,4 \cdot 0,24 + 2,6 \cdot 0,12 + 3 \cdot 0,2 + 4,2 \cdot 0,04 = 1,536$$

$$s_{\text{ext}}^2 = 654,13 \quad \text{vgl. Tabelle}$$

$$s^2 = s_{\text{int}}^2 + s_{\text{ext}}^2 = 1,536 + 654,13 = 655,666$$

$$s = \sqrt{655,666} = 25,61$$

$$\text{iv) } x_{\text{mod}} = 10$$

$$\tilde{x}_{0,5} = 20 + 15 \frac{0,5 - 0,4}{0,24} = 26,25$$

$$\text{b) } R = 602 - 398 = 203$$

$$\tilde{x}_{0,25} = x_{([6 \cdot 0,25 + 1])} = x_{(2)} = 467$$

$$\tilde{x}_{0,75} = x_{([6 \cdot 0,75 + 1])} = x_{(5)} = 523$$

$$Q_{0,25} = 523 - 467 = 56$$

	399	467	478	488	523	602
399	-	68	79	89	124	203
467	-	-	11	21	56	135
478	-	-	-	10	45	124
488	-	-	-	-	35	114
523	-	-	-	-	-	79
602	-	-	-	-	-	-

$$S_G = \frac{2}{n(n-1)} \sum_{i < j} |x_i - x_j| n_{ij} = \frac{2}{6 \cdot 5} 1193 = 79,53$$

Aufgabe 2:

a)

i) $G_3 = 2G_2 = 4G_1$

$$600.000 = 200G_1 + 100G_2 + 50G_3$$

$$= 200G_1 + 100 \cdot 2G_1 + 50 \cdot 4G_1$$

$$= 600G_1$$

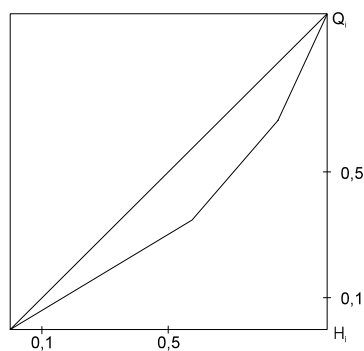
$$\Rightarrow G_1 = 1000$$

$$G_2 = 2000$$

$$G_3 = 4000$$

ii)

	x_i	n_i	h_i	$x_i n_i$	q_i	Q_i	H_i
G_1	1000	200	0,5714	200000	$0,\bar{3}$	$0,\bar{3}$	0,5714
G_2	2000	100	0,2857	200000	$0,\bar{3}$	$0,\bar{6}$	0,8571
G_3	4000	50	0,1429	200000	$0,\bar{3}$	1	1



iii) $D_G = \sum q_i (H_i + H_{i-1}) - 1$

	H_i	$H_i + H_{i-1}$	$(H_i + H_{i-1})q_i$
G_1	0,5714	0,5714	0,1905
G_2	0,8571	1,4285	0,4762
G_3	1	1,8571	0,619
			1,2857

$$D_G = 1,2857 - 1 = 0,2857$$

b)

i) $x_{\text{mod}} = \text{"ja"}$

ii)

x\y	ja	nein	Σ
ja	27	33	60
nein	18	22	40
Σ	45	55	100

iii) $h(Y = \text{ja} | X = \text{ja}) = h(Y = \text{ja}) = 0,45 = 45\%$

iv) Bestandsmasse, diskret, nominalskaliert

v) ein-eindeutige Transformation, denkbar wäre auch z.B. ja = 185 und nein = -32

Aufgabe 3:

a)

$$i) \quad B_{yx} = \frac{s_{\hat{y}}^2}{s_y^2}$$

$$\hat{y} = 100 + 5x$$

$$\Rightarrow s_{\hat{y}}^2 = 5^2 s_x^2 = 25 \cdot 100 = 2500$$

$$B_{xy} = \frac{2500}{4000} = 0,625$$

$$ii) \quad s_{xy} = r_{xy} s_x s_y = \sqrt{0,625} \sqrt{100} \sqrt{4000} = 500$$

b)

$$i) \quad r_{xy} = \sqrt{bd} = \sqrt{0,5 \cdot \frac{13}{22}} = 0,54$$

ii) Die Geraden schneiden sich im Schwerpunkt $(\bar{x}, \bar{y})$

$$I. \quad \bar{y} = 3,5 + 0,5\bar{x} \Rightarrow \bar{x} = 2\bar{y} - 7$$

$$II. \quad \bar{x} = \frac{32}{22} + \frac{13}{22}\bar{y}$$

$$2\bar{y} - 7 = \frac{32}{22} + \frac{13}{22}\bar{y}$$

$$2\bar{y} - \frac{13}{22}\bar{y} = \frac{32}{22} + 7$$

$$\bar{y} = 6 \quad \bar{x} = 2 \cdot 6 - 7 = 5$$

c)

$$\text{i) } P_{0t}^L = 1,1 \cdot 0,1 + 0,9 \cdot 0,3 + 1,2 \cdot 0,05 + 1,15 \cdot 0,45 + 0,95 \cdot 0,1 = 1,0525$$

$$\text{ii) } W = P_{0t}^L Q_{0t}^L + C = 1,0525 \cdot 1,1 - 0,25 = 0,90775$$

Nachklausur SS 98Aufgabe 1

a)

$$i) \quad y'(t) = \frac{1}{2}t - 2$$

$$r(t) = \frac{y'(t)}{y(t)} = \frac{\frac{1}{2}t - 2}{\frac{t^2}{4} - 2t + 10}$$

$$r(12) = \frac{\frac{12}{2} - 2}{\frac{12^2}{4} - 2 \cdot 12 + 10} = \frac{4}{22} = 0,1818$$

ii)

t	y _t	r _t	w _t
1	10	-	-
2	8	-0,2	0,8
3	7	-0,125	0,875
4	6	-0,143	0,857
5	4	-0,333	0,667
6	5	0,25	1,25
7	8	0,6	1,6
8	10	0,25	1,25
9	12	0,2	1,2
10	14	0,167	1,167
11	16	0,143	1,143
12	20	0,25	1,25
Summe	120		

$$\bar{w}_t = \sqrt[12]{20/10} = \sqrt[12]{2} = 1,065$$

$$\bar{r}_t = 1,065 - 1 = 0,065 \hat{=} 6,5\%$$

iii) aktueller Gesamtabsatz: 120.000 Stück

$$120.000 \cdot 1,065^t = 300.000$$

$$\Leftrightarrow 1,065^t = 2,5$$

$$\Leftrightarrow t \cdot \ln 1,065 = \ln 2,5$$

$$\Rightarrow t = \frac{\ln 2,5}{\ln 1,065} = 14,55$$

$$\text{b) i) } \bar{x} = \frac{1}{8}(x_1 + 4124)$$

$$\Rightarrow x_1 = 577,5 \cdot 8 - 4124 = 496$$

$$\bar{y} = \frac{1}{T} \sum y_t = \frac{1}{8}(68 + 75 + 84 + 93 + 86 + 72 + 66 + 54) = 74,75$$

$$s_{xy} = \frac{1}{T} \sum x_t y_t - \bar{x} \bar{y} = \frac{1}{8} 341667 - 577,5 \cdot 74,75 = -484,75$$

$$\text{ii) } r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$s_x^2 = \frac{1}{T} \sum x_t^2 - \bar{x}^2 = \frac{1}{8} 2703906 - 577,5^2 = 4482$$

$$s_y^2 = \frac{1}{T} \sum y_t^2 - \bar{y}^2 = \frac{1}{8} 45806 - 74,75^2 = 138,1875$$

$$r_{xy} = -\frac{484,75}{\sqrt{4482 \cdot 138,1875}} = -0,616$$

Aufgabe 2

a)

km	m_i	n_i	$m_i n_i$	h_i	H_i	h_i^*
0 - 1	0,5	49	24,5	0,07	0,07	0,07
1 - 5	3	168	504	0,24	0,31	0,06
5 - 15	10	245	2450	0,35	0,66	0,035
15 - 30	22,5	126	2835	0,18	0,84	0,012
30 - 50	40	112	4480	0,16	1	0,008
Summe		700	10293,5	1		

$$\bar{x} = \frac{1}{n} \sum m_i n_i = \frac{1}{700} 10293,5 = 14,705$$

$$\tilde{x}_{0,5} = x'_{k-1} + \frac{b_k(0,5 - H_{k-1})}{h_k} = 5 + \frac{10(0,5 - 0,31)}{0,35} = 10,43$$

$$x_{\text{mod}} = 0,5$$

Wegen $x_{\text{mod}} < \tilde{x}_{0,5} < \bar{x}$ ist die Verteilung linkssteil.

b)

$$\text{i) } P_{89,93}^L = \frac{\sum p_t q_0}{\sum p_0 q_0} = \frac{20 \cdot 100 + 70 \cdot 100 + 30 \cdot 100}{10 \cdot 100 + 60 \cdot 100 + 30 \cdot 100} = \frac{12000}{10000} = 1,2$$

$$P_{89,93}^P = \frac{\sum p_t q_t}{\sum p_0 q_t} = \frac{20 \cdot 50 + 70 \cdot 80 + 30 \cdot 60}{10 \cdot 50 + 60 \cdot 80 + 30 \cdot 60} = \frac{8400}{7100} = 1,183$$

$$\text{ii) } W_{89,93} = \frac{\sum p_t q_t}{\sum p_0 q_0} = \frac{8400}{10000} = 0,84 \quad Q_{89,93}^P = \frac{W_{89,93}}{P_{89,93}^L} = \frac{0,84}{1,2} = 0,7$$

$$\text{iii) } P_{89,93}^P Q_{89,93}^P = 1,183 \cdot 0,7 = 0,8281 \neq W_{89,93} = 0,84$$

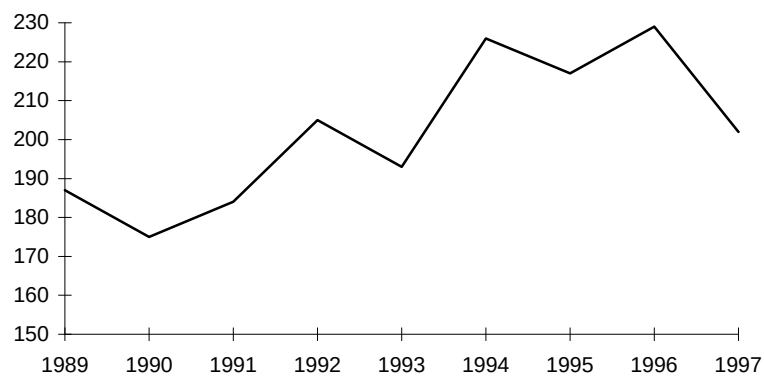
Aufgabe 3

a) i + ii)

x_i	n_i	h_i	H_i	q_i	Q_i	s_i	$H_i + H_{i-1}$	$(H_i + H_{i-1})q_i$
1	72	0,6667	0,6667	0,4311	0,4311	0,6466	0,6667	0,2874
2	23	0,2130	0,8797	0,2754	0,7065	1,2930	1,5464	0,4259
3	7	0,0648	0,9445	0,1257	0,8322	1,9398	1,8242	0,2293
4	3	0,0278	0,9723	0,0719	0,9041	2,5863	1,9168	0,1378
5	2	0,0185	0,9908	0,0599	0,9640	3,2378	1,9631	0,1176
6	1	0,0093	1	0,0359	1	3,8602	1,9908	0,0715
Summe	108	1		1				1,2695

$$D_G = 1,2695 - 1 = 0,2695$$

b) i)



ii) Annahme: $t = t^* = -4, -3, \dots, 3, 4$

$$\hat{y}_{t^*} = a + b \cdot t^*$$

$$a = \frac{\sum y_t}{T} = \frac{1818}{9} = 202$$

$$b = \frac{\sum ty}{\sum t^2} = \frac{309}{60} = 5,15$$

$$\hat{y}_{t^*} = 202 + 5,15 \cdot t^*$$

$$\text{iii) } \hat{y}_4 = 202 + 5,15 \cdot 4 = 222,6$$

$$\text{iv) } y_5^p = 0,5 \cdot y_4 + 0,5 \cdot y_4^p = 0,5 \cdot 202 + 0,5 \cdot 222,6 = 212,3$$