Teil III Klausurtraining

Aufgabe 1

Aus der Statistik des Scheichtums S ergab sich, dass folgende 7 Personen sich im Jahre .... in der Wüste verirrt haben:

Person	Geschlecht	Alter (Jahre)	Religion
1	m	17	C
2	m	29	M
3	m	73	C
4	w	35	C
5	m	21	C
6	m	15	M
7	w	19	A

Zeichen:
m = männlich, w = weiblich
C = Christ, M = Moslem
A = Andere Religion

a) Geben Sie für jedes Merkmal den Skalentyp an und bestimmen Sie die geeignete Art von Mittelwert (Zentralwert, arithmet. Mittel usw.), der jeweils dem Skalentyp angemessen ist:

	Skalentyp	Mittelwert
Geschlecht		
Alter		
Religion		

b) Die folgenden Begriffe (Nr. 1 bis 5) möge man den folgenden auf die Aufgabe bezogenen Sachverhalten zuordnen, indem man die richtige Nummer in den dafür vorgesehenen freien Kasten einträgt:

Masse	1	Personen die sich verirrt haben	
Merkmal	2	einzelne Person, die sich verirrt hat	
Merkmalsausprägung	3	Moslem	
Einheit	4	Person 2	
Maßzahl	5	Mittleres Alter	
		weiblich	
		Geschlecht	

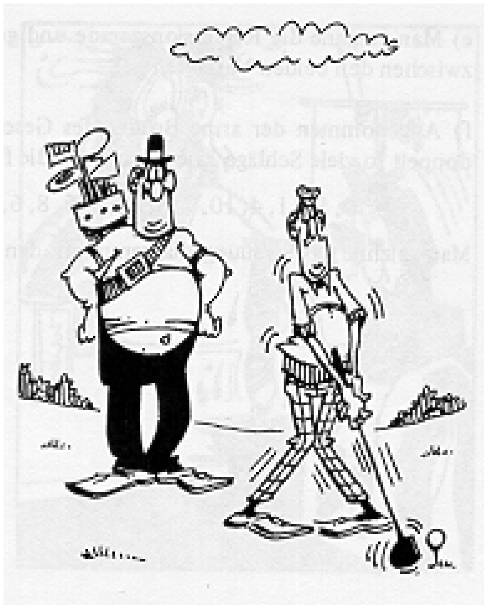
c) Bilden Sie Größenklassen für das Alter (von ... bis unter...) 0 - 20, 20 - 40, 40- 60, 60 - 80 und stellen Sie die klassierte Häufigkeitsverteilung sowie die Summenhäufigkeitskurve mit den absoluten Häufigkeiten graphisch dar!

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

$$\left.\begin{array}{l}\text{erhöht}\\ \text{verringert}\end{array}\right\}\ \text{haben}.$$

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) :

O intensiv	O extensiv
O diskret	O stetig
O eine absolute	O eine Verhältniszahl
O Zahl	(relative Zahl)

b) Die Folge der oben angegebenen 18 Zahlen bildet

O eine Einheit	O ein Merkmal
O eine Bestands-masse	O eine Merkmals-ausprägung
O eine Bewegungs-masse	O eine Verteilung

Für das Merkmal x ist eine . . . . .Skala 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!

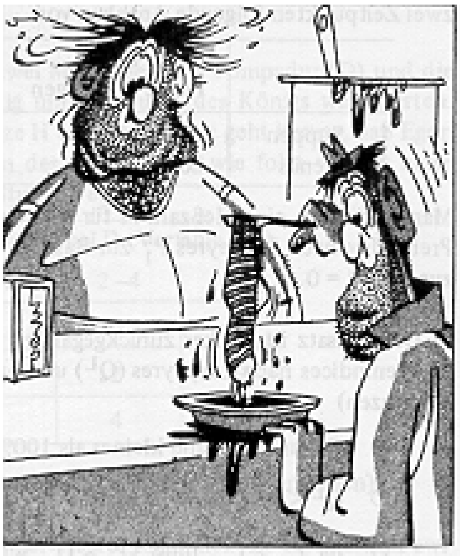
e) Man zeichne die Regressionsgerade und gebe an, wie groß die Korrelation r_{xy} zwischen den beiden Variablen ist!

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 Streuungsdiagramm 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

a) Man berechne gleitende 3-Monats-Durchschnitte für die Größe y!

b) Die Werte für die gleitenden Durchschnitte liegen offenbar auf einer Geraden. Wie lauten diese? Warum ist das so?

c) Würde man zu der gleichen Geraden gelangen, wenn man einen linearen Trend mit der Methode der kleinsten Quadrate berechnen würde? Begründung!

d) 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.

e) 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

a) Man bestimme die bedingten Mittelwerte (empirische Regressionslinien)!

b) 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)!

c) 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:

$5^2 = 25,\qquad 5^3 = 125,\qquad 5^4 = 625, \ldots\ldots\ldots, 5^7 = 78125$

Wie sieht diese Zeitreihe in halblogarithmischer Darstellung aus!

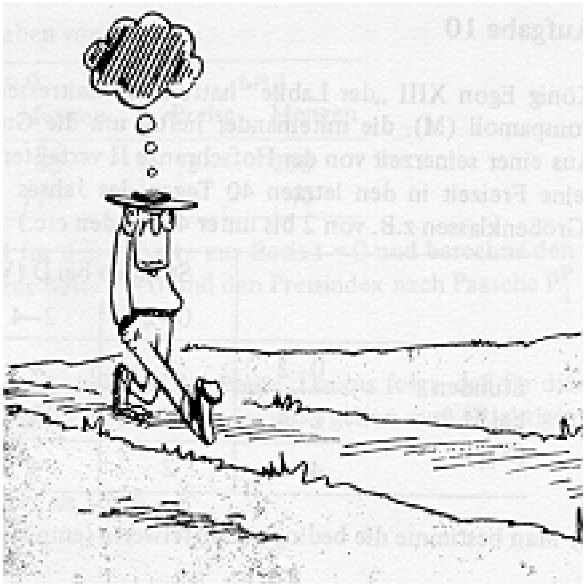
d) Kann es sein, dass sich die Kosten der Hofhaltung laufend erhöhen, während ein entsprechender Preisindex nach Laspeyres konstant bleibt oder gar sinkt?

e) 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!)

Aufgabe 5



Der Werbetexter W glaubt daran, auf längeren Fußmärschen über Felder relativ häufiger werbewirksame Einfälle zu haben als am Schreibtisch. Von 10 bedeutenden Einfällen kamen ihm 6 bei Spaziergängen und nur 4 im Büro. Andererseits gibt es auch 10 Fälle krampfhaften aber erfolglosen Bemühens um Einfälle, 5 davon im Büro und 5 bei Spaziergängen.

a) Man bestimme die relativen Häufigkeiten werbewirksamer Einfälle von W

im Büro:

bei Spaziergängen:

insgesamt:

b) Bei dieser Maßzahl handelt es sich um (Richtiges ankreuzen) :

○ eine Konzentrationszahl
○ eine Beziehungszahl
○ einen Index
○ eine Gliederungszahl
○ eine Verhältniszahl
○ eine Korrelation

c) Man bestimme eine zweidimensionale Verteilung zur Beschreibung des obigen Zusammenhangs zwischen dem Aufenthaltsort und der Fähigkeit Einfälle zu haben und berechne die Vierfelderkorrelation

d) Die folgenden Begriffe (Nr. 1 bis 5)

1 Merkmal
2 Einheit
3 Maßzahl
4 Merkmalsausprägung
5 Masse

möge man den folgenden 10 Worten aus dem Text der Aufgabe zuordnen indem man die richtige Nummer in den dafür vorgesehenen freien Kasten einträgt

a) Aufenthaltsort	
b) Werbetexter	
c) am Schreibtisch (bzw. im Büro)	
d) werbewirksam	
e) Art des Einfalls	
f) Häufigkeit des Einfalls	

g) Vierfelderkorrelation	
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

- O Intervallskala
- O Nominalskala
- O Ordinalskala
- O 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	Anzahl der		Anzahl der	
der Frauen	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)

O Messzahl	O Beziehungszahl	O Mittelwert
O Maßzahl	O Verursachungszahl	O Streuung
O Gliederungszahl	O Quote	O 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:

Größenklasse	gefahrene Kilometer je Wagen x_j	Anzahl der Mieter 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 $\overline{y}$ 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} = ...!

e) Ist die Verteilung der gefahrenen Kilometer
- O linkssteil
- O rechtssteil
- O 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

a) Berechnen Sie den Korrelationskoeffizienten r_{xy} und begründen Sie, warum diese Aufgabe die Fragestellung der Korrelationsanalyse beschreibt!

b) Welche Werte $\hat{y}$ erhält man bei Kenntnis der Regressionsgeraden $\hat{y} = a + b \cdot x$ (also einer korrigierten Selbsteinschätzung x)?

c) Zeigen Sie, dass $\hat{y}$ eine bessere Schätzung für y ist als die Selbsteinschätzung x des Studenten S!

d) 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!

	S hat sich vorher besser eingeschätzt	S hat sich vorher schlechter oder gerade richtig eingeschätzt
Klausur		
Recht	4	2
andere Klausuren	10	9

Aufgabe 9

Ein Unternehmen betätige sich in zwei Wirtschaftszweigen (Branche A und B). Umsätze (U) und Kapitaleinsatz (K) mögen sich wie folgt entwickelt haben

	Branche A		Branche B	
Jahr	U	K	U	K
1976	10	50	6	50
1977	14	70	4,2	35
1978	20	100	3	25

a) Man berechne die Umsatzrentabilitäten $\frac{U}{K}$ für die beiden Branchen und für das gesamte Unternehmen (warum ist letztere gestiegen?)!

b) Man berechne Messzahlen für die gesamte Umsatzentwicklung (Basis 1976 = 100)!

c) 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)

- O Verhältniszahl
- O Gliederungszahl
- O Quote
- O Beziehungszahl
- O Rate
- O Wachstumsrate
- O Messzahl
- O 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)!

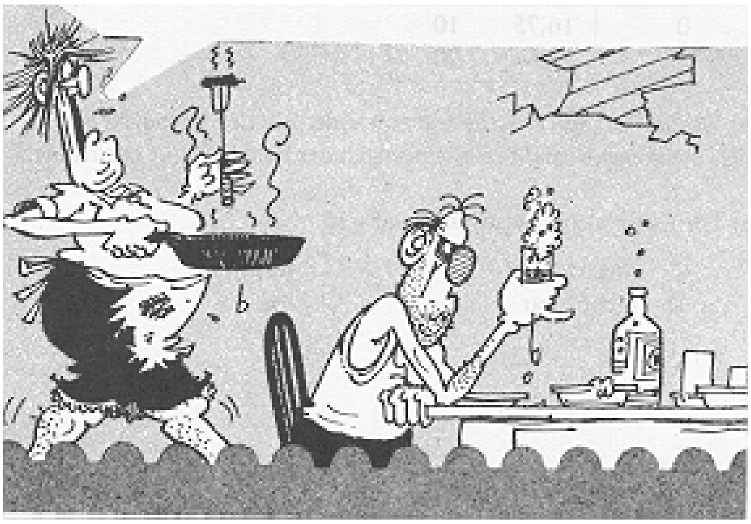
d) Man berechne den Paasche-Mengenindex!

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
- O höher
- O niedriger
- O 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 Streuungsdiagramm mit den Achsen x und y. Die zwischen diesen Punkten verlaufende Indifferenzkurve (Regressionsfunktion) wird vermutlich sein:

$$\left.\begin{array}{r}\text{linear}\\[1ex]\text{konvex}\\[1ex]\text{konkav}\end{array}\right\}\text{ vom Ursprung des Koordinatensystems gesehen.}$$

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

O	linear	O	positiv
O	nichtlinear	O	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:

	Todesursache				
Alter	S	N	U	Summe	darunter A
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)

- O die Lorenzkurve, weil sich die Sterbefälle auf bestimmte Todesursachen konzentrieren
- O die Abweichung vom Trend, weil dies zeigt, in welchem Jahr besonders ungewöhnliche Verhältnisse vorlagen
- O den Modus der Altersverteilung
- O die Varianz, weil keine Konzentration vorliegt, wenn sie klein ist
- O 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

$\left.\begin{matrix}\text{größer}\\ \text{kleiner}\end{matrix}\right\}$ als die der Reihe U_t

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):

		Ausgabenanteile in vH	
Warenart	Preis	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
- O aus der unter a) genannten Wertsteigerung der Warenkörbe um 95 bzw. 504 E - Dollar
- O aus den unter b) berechneten Preisindizes
- O 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

$r_{xy.z}$, $r_{yz.x}$ 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: $\frac{210}{7} = 30$

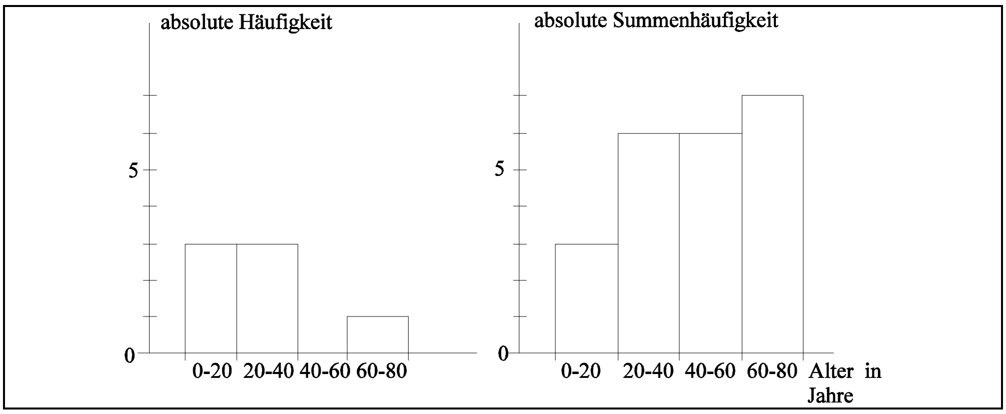
geometr. Mittel: $\sqrt[7]{15 \cdot 17 \cdot \ldots \cdot 73} = 25{,}891$

harmonisches Mittel: $\frac{7}{\frac{1}{15}+\frac{1}{17}+\ldots+\frac{1}{73}} = 23{,}229$

b) 1, 4, 3, 4, 5, 3, 2

c)

Klasse		absolute		relative	
Nr.	Alter	Häufigkeit	Summen-häufigkeit	Häufigkeit	Summen-häufigkeit
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

$$\hat{y}_i = 2x_i \quad (i = 1, 2, \ldots\ldots, 18) \quad \text{und} \quad \hat{y}_i = y_i$$

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 - 10 t. 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
2	70	-2	68
3	60	-3	57
4	50	+5	55
5	40	-2	38
6	30	-3	27

Dieser Zyklus wird mit dreigliedrigen gleitenden Durchschnitten vollständig eliminiert.

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 $\Sigma u_t = 0$ 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 Darstellung 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**
  - **Verhältniszahlen**
    - Beziehungszahlen
    - **Gliederungszahlen**
    - Messzahlen (z.B. auch Indexzahlen)
  - Mittelwerte, Streuung, Konzentration, Schiefermaße, Wachstumsraten usw.

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 (-assoziation) 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)	→	1	b)	→	2	c)	→	4
d)	→	4	e)	→	1	f)	→	1
g)	→	3	h)	→	4	i)	→	2
j)	→	5						

e) Die Zahlen sind im Sinne einer **Nominalskala** benutzt worden.

Lösung zu Aufgabe 6

a) Rohe (gesamte) Fruchtbarkeitsraten f:

christliches Dorf $\quad f_C = \frac{92}{1000} \cdot 1000 = 92$

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. ¾ Mädchen und ¼ 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

$$\frac{1}{900}\cdot\left[(100-130)^3\cdot 200 + (160-130)^3\cdot 200\right] = 0$$

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	x		
Varianz			x
Konzentration			x
Mieteinnahmen	x		
Zentralwert		x	
Anzahl der Mieter	x		

Lösung zu Aufgabe 8

a) Arbeitstabelle:

$$\Sigma(x-\bar{x})^2 = 12$$
$$\Sigma(y-\bar{y})^2 = 10$$
$$\Sigma(x-\bar{x})(y-\bar{y}) = 6$$

Folglich ist $r = \dfrac{6/8}{\sqrt{12/8 \cdot 10/8}} = 0{,}5477$.

Die hier vorliegende Fragestellung ist ein Problem der Korrelationsanalyse, da beide zu betrachtenden Variablen X und Y als Zufallsvariablen aufgefaßt werden können[1].

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

[1] 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.

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.

d) Die bedingten relativen Häufigkeiten h dafür, sich vorher besser einzuschätzen als die Klausur dann tatsächlich war, sind

$$h_{\text{Recht}} = \frac{4}{6} = 0{,}667 \text{ und } h_{\text{and. Klausuren}} = \frac{10}{19} = 0{,}526.$$

Sie sind also verschieden, so dass ein Zusammenhang (Assoziation) zwischen der Art des Faches und der Treffsicherheit der Selbsteinschätzung besteht. Die Vierfelderkorrelation ist hier

$\phi = 0{,}12076 \neq 0$, so dass keine Unabhängigkeit besteht.

Lösung zu Aufgabe 9

a)

	Umsatzrentabilitäten				Anteil der
Jahr	A	B	insgesamt	Kapitaleinsatz	Branche A
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)

	Umsätze	
Jahr	Absolut	1976 = 100
76	16	100
77	18,2	113,75
78	23	143,75

Messzahlen des Kapitals vergl. Teil a)

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 K	20 100	31 130	55% 30%	3 25	4,5 30	50% 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

$$\frac{14 \cdot 10 + 4{,}2 \cdot 6}{10 \cdot 10 + 6 \cdot 6} = 1{,}215 \text{ also } 121{,}5 \text{ vH}$$

für das Jahr 1978

$$\frac{20 \cdot 10 + 3 \cdot 6}{136} = 1{,}603 \text{ also } 160{,}3 \text{ vH}$$

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

$$0{,}8 \leq P^{P} \leq 1{,}0 .$$

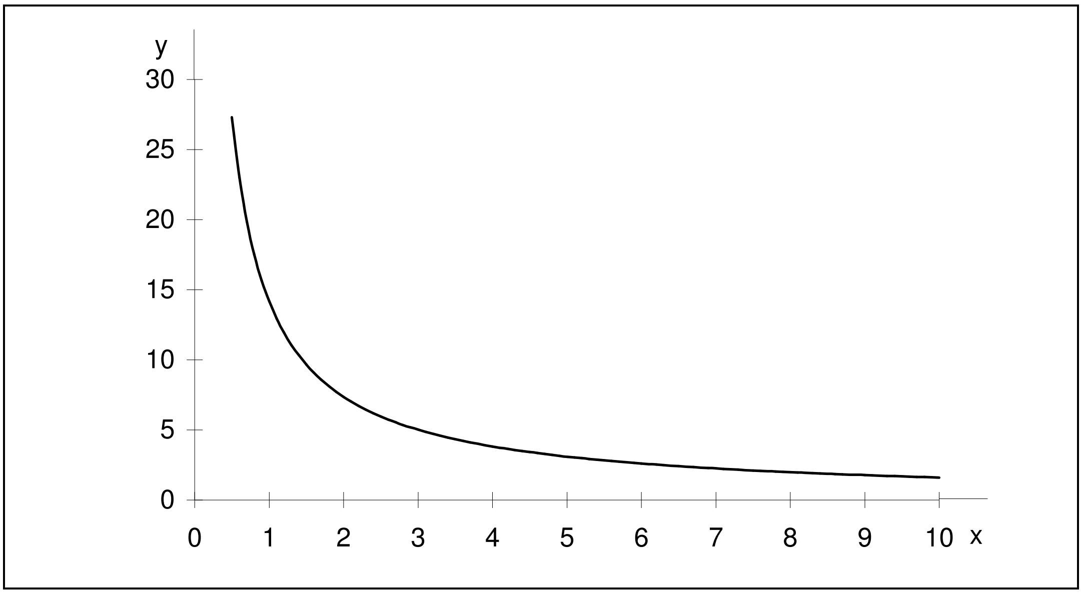
Wie immer die Konsumstruktur des Haushalts zur Berichtszeit sein mag, der Paasche-Preisindex kann nicht kleiner sein als 0,8 (also 80%) und nicht größer sein als 1. Rechnet man mit den Zahlen von Teil b, so erhält man 0,8734, einen Wert also, der innerhalb der oben angegebenen Schranken liegt.

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 $\hat{y} = 14{,}175 x^{-0{,}946}$ 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 $X_{j0} = 90$: → $Z = 67{,}5$

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
$\frac{U'_t}{U_t} = \frac{18-2t}{5+18t-t^2}$ oder $\frac{2A}{86-A^2}$, wenn $A = 9 - t$ ist. Der Wert für U_t zur Zeit t = 3 ist $U_3 = 50$. 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

$$\left.\begin{array}{l} 11a + 133{,}9b = 388 \\ 133{,}9b + 1964{,}61b = 6204{,}7 \end{array}\right\} \begin{array}{l} a = -18{,}6176 \\ b = 4{,}4271 \end{array}$$

somit $\hat{y} = -18{,}6176 + 4{,}4271x$.

b) $$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} = 0{,}8417$$

wobei $\left(\sum x\right)^2 = 17929{,}21$

$\sum y^2 = 22944$

$\left(\sum y\right)^2 = 150544$

c) $\hat{\sigma}^2 = \dfrac{\sum \hat{u}^2}{n-2}$ 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) $$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{\left(1 - r_{xz}^2\right)\left(1 - r_{yz}^2\right)}} = 0{,}8114$$

$$r_{yz.x} = \frac{r_{yz} - r_{yx} \cdot r_{zx}}{\sqrt{\left(1 - r_{yx}^2\right)\left(1 - r_{zx}^2\right)}} = 0{,}2393$$

$$r_{xz.y} = \frac{r_{xz} - r_{xy} \cdot r_{zy}}{\sqrt{\left(1 - r_{xy}^2\right)\left(1 - r_{zy}^2\right)}} = 0{,}0238$$

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.

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.