Musterlösungen

zu den Aufgaben aus

Statistische Methoden

in den Wirtschafts- und Sozialwissenschaften

von Prof. Dr. Hans Peter Litz

Oldenbourg-Verlag München, 3.Auflage 2003

Teil I

Konzeptionelle Grundlagen und statistische Deskription

Aufgaben 16 – 28

Stand 25.04. 2005

(überarbeitet und aktualisiert von Daniel Fokken, Frank Huisken und Stefan Mahlstedt)

Aufgabe 16: Eindimensionale Häufigkeiten, bedingte relative Häufigkeiten, Sätze der Unabhängigkeit

Arbeitstabelle

f_{ij} $f_{ij}^{'}$ $f^{'}(Y_i \mid X_j)$	Gruppe A	Gruppe B	$\sum$
Qualifikation erreicht	70 0,35 0,7	55 0,275 0,55	125 0,625
Qualifikation nicht erreicht	30 0,15 0,3	45 0,225 0,45	75 0,375
$\sum$	100 0,5	100 0,5	200 1,0

a) nach Satz 1:

$$f'(Y_1 \mid X_1) = \frac{f_{11}}{f_{.1}} = \frac{70}{100} = 0{,}7$$

$$f'(Y_1 \mid X_2) = \frac{f_{12}}{f_{.2}} = \frac{55}{100} = 0{,}55$$

$$f'(Y_2 \mid X_1) = \frac{f_{21}}{f_{.1}} = \frac{30}{100} = 0{,}3$$

$$f'(Y_2 \mid X_2) = \frac{f_{22}}{f_{.2}} = \frac{45}{100} = 0{,}45$$

Satz 1 ist demnach nicht erfüllt!,
d.h. die beiden Variablen sind voneinander unabhängig.

b) nach Satz 2:

$$\frac{f_{11}}{f_{.1}} = \frac{70}{100} = 0{,}7 \neq \frac{f_{1.}}{N} = \frac{125}{200} = 0{,}625$$

$$\frac{f_{12}}{f_{.2}} = \frac{55}{100} = 0{,}55 \neq \frac{f_{1.}}{N} = \frac{125}{200} = 0{,}625$$

$$\frac{f_{21}}{f_{.1}} = \frac{30}{100} = 0{,}3 \neq \frac{f_{2.}}{N} = \frac{75}{200} = 0{,}375$$

$$\frac{f_{22}}{f_{.2}} = \frac{45}{100} = 0{,}45 \neq \frac{f_{2.}}{N} = \frac{75}{200} = 0{,}375$$

Satz 2 wird nicht erfüllt!,
d.h. beide Variablen sind voneinander unabhängig

c) nach Satz 4: Bei Unabhängigkeit würde gelten

$$\frac{f_{11}}{N} = \frac{f_{.1}}{N} \cdot \frac{f_{1.}}{N}$$

tatsächlich gilt jedoch $\frac{70}{200} \neq \frac{100}{200} \cdot \frac{125}{200} \Leftrightarrow 0{,}35 \neq 0{,}3125$

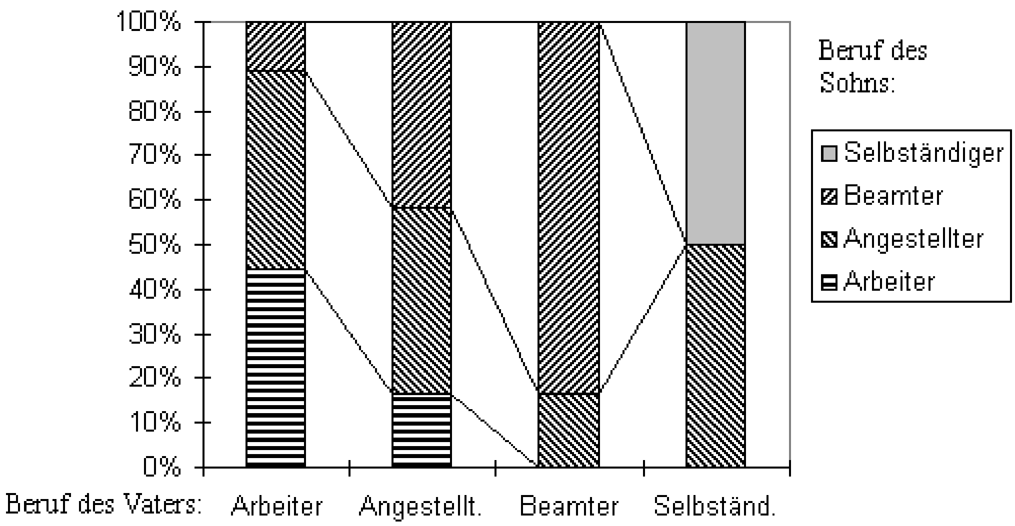
D.h. beide Variablen sind voneinander abhängig.

Aufgabe 17: Indifferenztabelle, Stabdiagramm

a) Graphische Darstellung der Kontingenztabelle:

V \ S	Arbeiter	Angestellter	Beamter	Selbstständige	$\Sigma f_{i.}=$
Arbeiter	**40**	10	0	0	50
Angestellter	40	**25**	5	10	80
Beamter	10	25	**25**	0	60
Selbstständige	0	0	0	**10**	10
$\Sigma f_{.j}=$	90	60	30	20	200

Graphische Darstellung der bedingten Häufigkeiten:



b) Ist der Berufsstand der Söhne unabhängig vom Berufsstand des Vaters, so gilt für die Hauptdiagonale nach Satz 4 der statistischen Unabhängigkeit:

$$\frac{f_{ij}}{N} = \frac{f_{i.}}{N} \cdot \frac{f_{.j}}{N} \Leftrightarrow f_{ij} = f_{.i.} \cdot \frac{f_{.j}}{N}$$

Bei Unabhängigkeit ergäbe sich: $f_{11} = 22{,}5 = 50 \cdot \frac{90}{200} = 22{,}5$

$$f_{22} = 24{,}0 = 80 \cdot \frac{60}{200} = 24{,}0$$

$$f_{33} = 9{,}0 = 60 \cdot \frac{30}{200} = 9{,}0$$

$$f_{44} = 1{,}0 = 10 \cdot \frac{20}{200} = 1{,}0$$

Aufgabe 18: Chi-Quadrat, Zusammenhangsmaße für nominal-skalierte Tabellen

Indifferenztabelle

	Gruppe A	Gruppe B	Summe
Qualifikation erreicht	$\frac{125 \cdot 100}{200} = 62{,}5$	$\frac{125 \cdot 100}{200} = 62{,}5$	125
Qualifikation nicht erreicht	$\frac{75 \cdot 100}{200} = 37{,}5$	$\frac{75 \cdot 100}{200} = 37{,}5$	75
Summe	100	100	200

Anmerkung: Festlegung der Freiheitsgrade: $FG = (z-1) \cdot (s-1) = 1$

Da die Tabelle einen Freiheitsgrad hat, genügt die Berechnung eines Feldes. Die anderen Häufigkeiten ergeben sich aus den Randhäufigkeiten.

$$\chi^2 = \sum_{b,e=1}^{z \cdot s} \frac{(f_b - f_e)^2}{f_e} = \frac{(70-62{,}5)^2}{62{,}5} + \frac{(55-62{,}5)^2}{62{,}5} + \frac{(30-37{,}5)^2}{37{,}5} + \frac{(45-37{,}5)^2}{37{,}5}$$

$$\chi^2 = 0{,}9 + 0{,}9 + 1{,}5 + 1{,}5 = 4{,}8$$

$$\varphi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{4{,}8}{200}} = 0{,}155$$

$$V = \sqrt{\frac{\chi^2}{N \cdot \min(z-1, s-1)}} = \sqrt{\frac{4{,}8}{200 \cdot 1}} = 0{,}155$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{4{,}8}{204{,}8}} = 0{,}153$$

$$C_{\max} = \sqrt{\frac{z-1}{z}} = \sqrt{\frac{1}{2}} = 0{,}7071$$

$$C_{korr} = \frac{C}{C_{\max}} = \frac{0{,}153}{0{,}7071} = 0{,}2164$$

Diskussion der Ergebnisse nach Interesse und Feedback im Tutorium.

Aufgabe 19: Zusammenhangsmaße für nominal-skalierte Daten

Hypothese: Es besteht ein Zusammenhang zwischen dem Beruf des Vaters und dem Beruf des Sohnes

a) Berechnung der Stärke des Zusammenhangs auf der Basis von χ^2 - Maßen

Vater / Sohn	Arbeiter	Angestellter	Beamter	Selbständiger	$\sum$
Arbeiter	22,5	15	7,5	5	50
Angestellter	36	24	12	8	80
Beamter	27	18	9	6	60
Selbständiger	4,5	3	1,5	1	10
$\sum$	90	60	30	20	200

$$\chi^2 = \frac{(40-22{,}5)^2}{22{,}5} + \frac{(10-15)^2}{15} + \frac{(0-7{,}5)^2}{7{,}5} + \frac{(0-5)^2}{5}$$

$$+\frac{(40-36)^2}{36} + \frac{(25-24)^2}{24} + \frac{(5-12)^2}{12} + \frac{(10-8)^2}{8}$$

$$+\frac{(10-27)^2}{27} + \frac{(25-18)^2}{18} + \frac{(25-9)^2}{9} + \frac{(0-6)^2}{6}$$

$$+\frac{(0-4{,}5)^2}{4{,}5} + \frac{(0{,}3)^2}{3} + \frac{(0-1{,}5)^2}{1{,}5} + \frac{(10-1)^2}{1}$$

$$\chi^2 = 13{,}61 + 1{,}67 + 7{,}5 + 5 + 0{,}44 + 0{,}04 + 4{,}08 + 0{,}5 + 10{,}7 + 2{,}72 + 28{,}44 + 6 + 4{,}5 + 3 + 1{,}5 + 81$$

$$\chi^2 = 170{,}71$$

$$\phi = \sqrt{\frac{\chi^2}{N}} = \sqrt{\frac{170{,}71}{200}} = 0{,}92$$

$$V = \sqrt{\frac{\chi^2}{N \quad min(z-1, s-1)}} = \sqrt{\frac{170{,}71}{200 \quad 3}} = 0{,}533$$

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{170{,}71}{170{,}71 + 200}} = 0{,}679$$

$$C_{max} = \sqrt{\frac{z-1}{z}} = \sqrt{\frac{2}{3}} = 0{,}816$$

$$C_{korr} = \frac{C}{C_{max}} = \frac{0{,}679}{0{,}816} = 0{,}83$$

Ergebnis: Es besteht ein hoher statistischer Zusammenhang zwischen dem Beruf des Sohnes und dem Beruf des Vaters.

b) Berechnung der Stärke des Zusammenhangs auf der Basis von PRE - Maßen λ :

Empirische Verteilung:

	Vater					
		Arbeiter	Angestellter	Beamter	Selbständiger	$\sum f_{i.}$
Sohn	Arbeiter	40	10	0	0	50
	Angestellter	40	25	5	10	80
	Beamter	10	25	25	0	60
	Selbständiger	0	0	0	10	10
	$\sum f_{.j}$	90	60	30	20	200

Erwartete Verteilung:

	Vater					
		Arbeiter	Angestellter	Beamter	Selbständiger	$\sum f_{i.}$
Sohn	Arbeiter	22,5	15	7,5	5	50
	Angestellter	36	24	12	8	80
	Beamter	27	18	9	6	60
	Selbständiger	4,5	3	1,5	1	10
	$\sum f_{.j}$	90	60	30	20	200

$F_1 = 120$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $\chi^2 = 170{,}71$

$F_2 = 100$ $\quad\quad (EV = 120)$ $\quad\quad \varphi = 0{,}92$

$\lambda_{y=f(x)} = 0{,}1667$ $\quad\quad (EV = 0)$

$$F_1 = N - \max(f_1) = 200 - 80 = 120$$

$$F_2 = \sum_{j=1}^{s}\left[f_{.j} - \max(f_{ij})\right] = (90-40) + (60-25) + (30-25) + (20-10) = 100$$

$$\lambda_{y=f(x)} = \frac{F_1 - F_2}{F_1} = \frac{120-100}{120} = 0{,}1\overline{6}$$

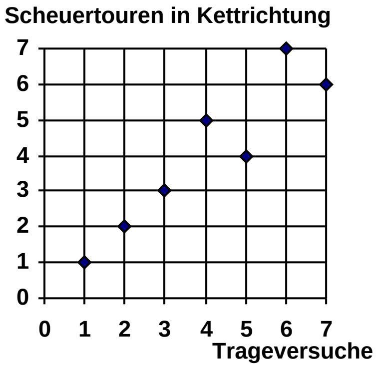
Aufgabe 20: Zusammenhangsmaße für ordinalskalierte Tabellen

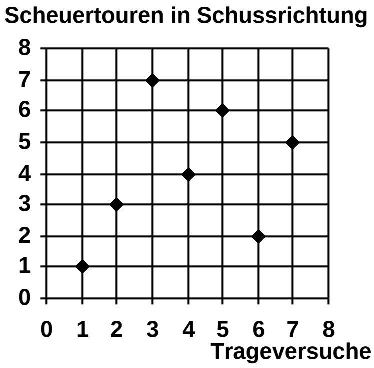
Um den geeigneten Kurzzeitversuch (Kettrichtung oder Schussrichtung) zu bestimmen, muss ein Rangvergleich angestellt werden. Das Skalenniveau ist ordinal, daher lässt sich ein Konkordanzmaß errechnen.

zunächst werden die Werte der Scheuertouren in eine Rangskala transformiert, damit sie mit den Rangzahlen im Trageversuch vergleichbar sind:

Gewebeart	Trageversuch	Rangzahl für Scheuertouren	
i		Kettrichtung	Schussrichtung
1	6	7	2
2	5	4	6
3	7	6	5
4	4	5	4
5	3	3	7
6	2	2	3
7	1	1	1

Graphische Darstellung:





Nun lässt sich die Anzahl der konkordanten und diskordanten Paare ermitteln:

$$N_p = \frac{N \cdot (N-1)}{2} = 21 \quad \text{(Anzahl der Paare insgesamt)}$$

Jetzt lassen sich die diskordanten und konkordanten Paare graphisch ermitteln:

Kettrichtung: $N_d = 2$ $N_k = 19$

Schussrichtung: $N_d = 8$ $N_k = 13$

Da keine Ties vorhanden sind, eignet sich der Wert τ_a.

$$\tau_a = \frac{N_k - N_d}{N_p} = \frac{19-2}{21} = 0{,}81 \quad \text{(für Kettrichtung)}$$

$$\tau_a = \frac{13-8}{21} = 0{,}24 \quad \text{(für Schussrichtung)}$$

Der Wert 1 würde bedeuten, dass ein Scheuertourenversuch die gleiche Rangfolge erbringt, wie ein Trageversuch. Da der Versuch in Kettrichtung mit 0,81 näher an 1 liegt als der Versuch in Schussrichtung, ist dieser der geeignetere Kurzzeitversuch.

Alternativ kann der Rankkorrelationskoeffizient nach Spearmann für die beiden Versuchsanordnungen der Scheuertouren in Kettrichtung und in Schussrichtung berechnet werden:

i) Für die Scheuertouren in Kettrichtung rechnen wir:

Gewebeart i	1	2	3	4	5	6	7
Langzeit - Tragezeitversuch	6	5	7	4	3	2	1
Scheuertouren in Kettrichtung	7	4	6	5	3	2	1
Rangdifferenz	-1	1	1	-1	0	0	0
Quadrierte Rangdifferenz $\left(d_i^2\right)$	1	1	1	1	0	0	0

Wir erhalten als Rangkorrelationskoeffizienten:

$$r_s = 1 - 2 \cdot \frac{\sum d_i^2}{\frac{1}{3} \cdot \left(N^3 - N\right)} = 1 - 2 \cdot \frac{4}{\frac{1}{3}\left(7^3 - 7\right)} = 0{,}9286$$

=> großer positiver Zusammenhang!

ii) Für die Scheuertouren in Schussrichtung rechnen wir:

Gewebeart i	1	2	3	4	5	6	7
Langzeit - Tragezeitversuch	6	5	7	4	3	2	1
Scheuertouren in Kettrichtung	2	6	5	4	7	3	1
Rangdifferenz	4	-1	2	0	-4	-1	0
Quadrierte Rangdifferenz $\left(d_i^2\right)$	16	1	4	0	16	1	0

Wir erhalten als Rangkorrelationskoeffizienten:

$$r_s = 1 - 2 \cdot \frac{\sum d_i^2}{\frac{1}{3} \cdot \left(N^3 - N\right)} = 1 - 2 \cdot \frac{38}{\frac{1}{3}\left(7^3 - 7\right)} = 0{,}3214$$

=> geringer positiver Zusammenhang!

Fazit: Die hohe Korrelation, die sich für die Versuchsanordnung ‘Scheuertouren in Kettrichtung’ ergab, lässt vermuten, dass sich die Langzeitversuche durch Kurzzeitversuche ersetzen lassen. Die Konkordanzmaße und die Rangkorrelationskoeffizienten fallen jeweils ähnlich aus.

Aufgabe 21: Zusammenhangsmaße für ordinal-skalierte Tabellen

a) Graphische Darstellung der relativen bedingten Häufigkeiten:

Arbeitstabelle:

			Bildung		
		untere	mittlere	obere	$\sum$
Schicht	untere	435	35	9	479
	mittlere	399	239	75	713
	obere	33	62	72	167
	$\sum$	867	336	156	1359

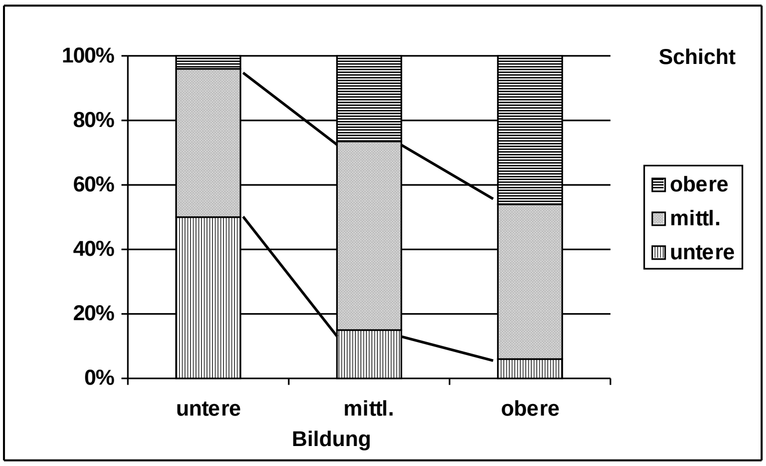
Interpretation:

Je höhere die Bildung, desto höher ist tendenziell die soziale Schicht, die eine der befragten Personen erreichen kann.

andere Interpretation möglich:

- je höher die Schicht, in der eine Person angesiedelt ist, desto höher ist tendenziell die Bildung der befragten Person
- Schicht und Bildung bedingen sich gegenseitig

Graphische Darstellung:



Die Segmentierung ändert sich mit dem Wechsel der Bedingungen relativ stark. Das deutet auf einen Zusammenhang zwischen Bildung und Schichtzugehörigkeit hin

(evtl. Diskussion über Schicht als Merkmal).

b) Zunächst muss die Anzahl aller möglichen Paare ermittelt werden (*Ordinalskala!*):

N_d= 9(399+33+239+62)+75(33+62)+35(399+33)+239(33)= 36.729

N_k=435(239+62+75+72)+35(75+72)+399(62+72)+239(72)= 270.699

T_y=435(35+9)+35(9)+399(239+75)+239(75)+33(62+72)+62(72)= 171.552

T_x=435(399+33)+399(33)+35(239+62)+239(62)+9(75+72)+75(72)= 233.163

T_{xy}=0,5(435*434+399*398+33*32+35*34+239*238+62*61+9*8
+75*74+72*71)= <u>210.618</u>

N_p= 922.761

Da hier von einem asymmetrischen Zusammenhang der Art Bildung = f(Schicht) ausgegangen werden, kann ist es am sinnvollsten, Somer`s $d_{y=f(x)}$ zu errechnen:

$$d_{y=f(x)} = \frac{N_k - N_d}{N_k + N_d + T_y} = \frac{270699 - 36729}{270699 + 36729 + 171552} = 0{,}49$$

Da Somer´s $d_{y=f(x)}$ zwischen -1 und +1 liegen muss bedeutet 0,49 einen mittelstarker positiven Zusammenhang (vgl. Graphik)

Alternativ kann nochmals mit χ^2 - Maßen gerechnet werden. Dazu ist im Tutorium zu diskutieren, warum diese Alternative auch zulässig ist.

=> χ^2 – *Zusammenhangsmaß*!

Indifferenztabelle:

	untere Bil.	mittlere Bil.	obere Bil.	Summe
untere Sch.	305.587196	118.428256	54.9845475	479
mittlere Sch.	454.871965	176.282561	81.8454746	713
obere Sch.	106.540839	41.2891832	19.1699779	167
Summe	867	336	156	1359

$$\chi^2 = \sum \frac{(f_b - f_e)^2}{f_e} = \frac{(435 - 305{,}59)^2}{305{,}59} + \ldots + \frac{(72 - 19{,}17)^2}{19{,}17} = 388{,}52$$

$$V = \sqrt{\frac{\chi^2}{N \cdot \min(z-1, s-1)}} = \sqrt{\frac{388{,}52}{1359 \cdot 2}} = 0{,}38$$

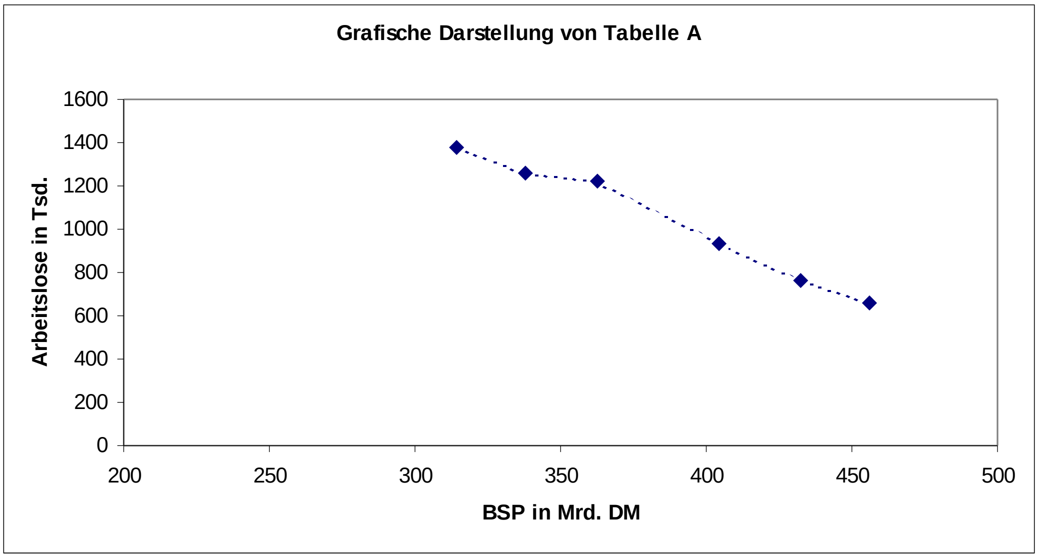
$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{388{,}52}{388{,}52 + 1359}} = 0{,}47$$

$$C_{\max} = \sqrt{\frac{z-1}{z}} = 0{,}82$$

$$C_{korr} = \sqrt{\frac{C}{C_{\max}}} = \sqrt{\frac{0{,}47}{0{,}82}} = 0{,}573$$

Aufgabe 22

zu a.)



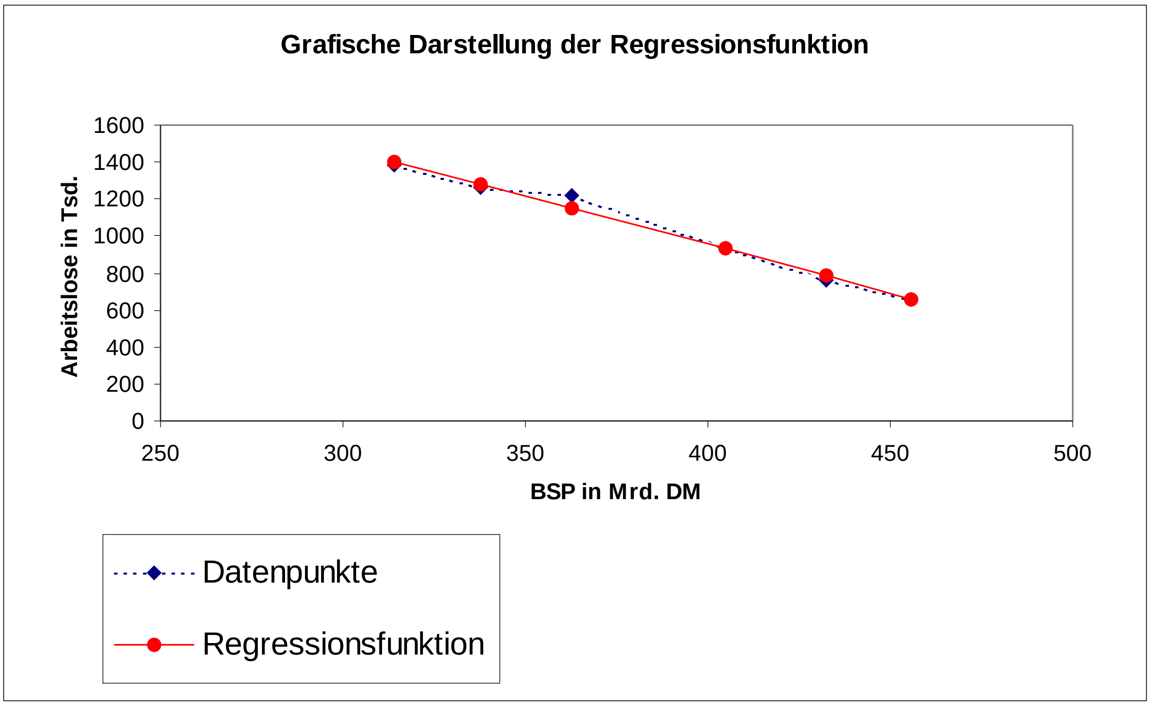
zu b.)

$y_i^c = a - b \cdot x_i$

Die Regressionsgleichung weist eine negative Steigung auf, d.h. zwischen den Variablen besteht folgender negativer, linearer Zusammenhang: Mit zunehmendem BSP sinkt die Zahl der Arbeitslosen.

zu c.)

i	Jahr	x	y	x^2	y^2	x*y	y_i^c
1	1952	314,15	1380	98690,2225	1904400	433527	1405,09
2	1953	337,82	1260	114122,352	1587600	425653,2	1280,831
3	1954	362,88	1220	131681,894	1488400	442713,6	1149,275
4	1955	404,64	930	163733,53	864900	376315,2	930,051
5	1956	432,48	760	187038,95	577600	328684,8	783,9011
6	1957	455,92	660	207863,046	435600	300907,2	660,8496
∑	-	2307,89	6210	903129,99	6858500	2307801	-



$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{6}\cdot 2307{,}89 = 384{,}65 \text{ BSP in Mrd. DM}$$

$$\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{6}\cdot 6210 = 1035 \text{ Tsd. Arbeitslose}$$

$$\bar{x}^2 = 147954{,}34$$

$$\bar{y}^2 = 1071225$$

Die Steigung b der Regressionsfunktion ergibt sich als:

$$b = \frac{\frac{1}{n}\sum x_i \cdot y_i - \bar{x}\cdot\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{\frac{1}{6}\cdot 2307801 - 384{,}65\cdot 1035}{\frac{1}{6}\cdot 903129{,}99 - 147954{,}34} = -5{,}25$$

Für den Achsenabschnitt erhält man :

$$a = \bar{y} - b\cdot\bar{x} = 1035 - (-5{,}25)\cdot 385{,}65 = 3054{,}26$$

Daraus ergibt sich die Regressionsfunktion:

$$y_i^c = 3054{,}26 - 5{,}25x$$

Die Hypothese wird durch die Grafik anscheinend bestätigt.

Bei der Interpretation von a ist Vorsicht angebracht. Formal besagt a, dass bei einem BSP von 0 Mrd. DM etwa 3 Mio. Arbeitslose zu erwarten wären. Da ein derartiges BSP fiktiv ist, macht auch die dabei zu erwartende Zahl von Arbeitslosen empirisch keinen Sinn.

Der relativ hohe, negative Steigungsfaktor b deutet darauf hin, dass ein starker negativer Zusammenhang zwischen BSP und Arbeitslosigkeit besteht. Der Koeffizient b besagt, dass im Beobachtungszeitraum ein Anstieg des BSP um 1 Mrd. DM zu einem Rückgang der Arbeitslosigkeit von jeweils etwa 5 Tsd. Personen geführt hat.

zu d.)

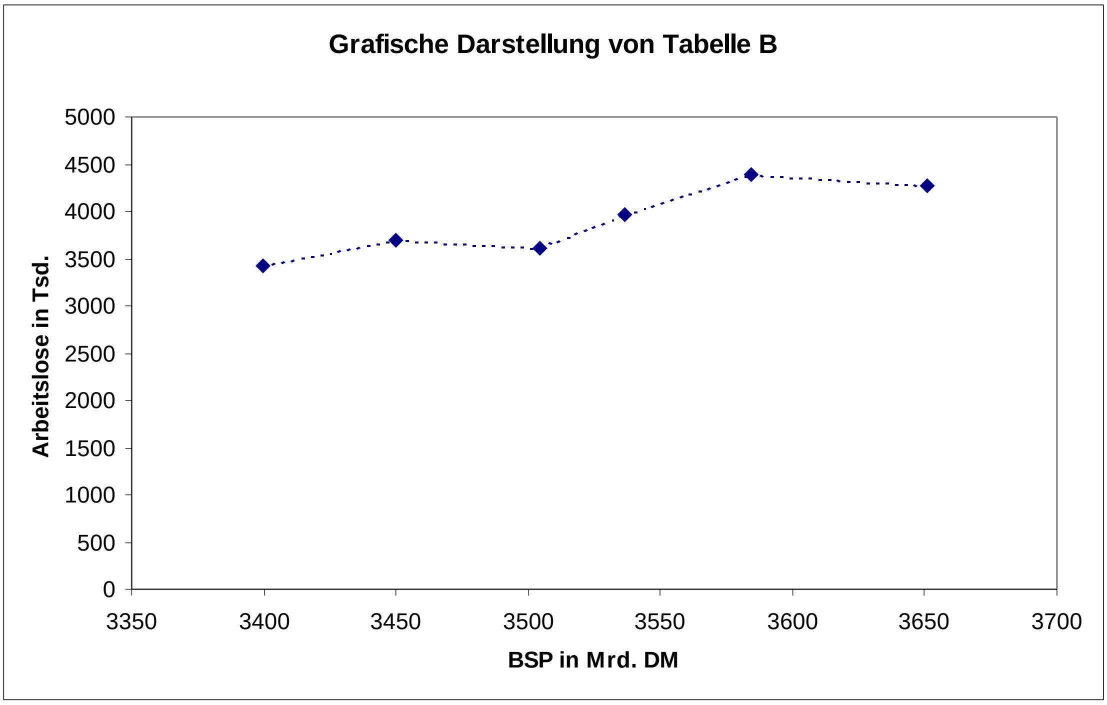
Bei einem BSP von 1000 Mrd. DM würde sich die Arbeitslosigkeit wie folgt ergeben:

$$y_i^c = 3054{,}26 - 5{,}25\ \ x = 3054{,}26 - 5{,}25\ \ 1000 = -2195{,}74$$

Bei einem BSP von 1000 Mrd. DM wäre nach der Regressionsfunktion mit einer negativen Zahl von Arbeitslosen zu rechnen. Auch hier ist anzumerken, dass eine Interpretation der Funktion und eine Prognose weit über den Beobachtungszeitraum hinaus zu gravierenden Fehlschlüssen führen kann.

zu e.)

i	Jahr	x	y	x^2	y^2	x*y	y_i^c
1	1993	3399,7	3419,1	11557960,09	11690244,81	11623914,27	3425,28
2	1994	3449,7	3698,1	11900430,09	13675943,61	12757335,57	3618,28
3	1995	3504,4	3611,9	12280819,36	13045821,61	12657542,36	3829,42
4	1996	3536,5	3965,1	12506832,25	15722018,01	14022576,15	3953,33
5	1997	3584,2	4384,5	12846489,64	19223840,25	15714924.9	4137,45
6	1998	3650,7	4279,3	13327610,49	18312408,49	15622440,51	4394,14
∑		21125,2	23358	74420141,92	91670276,78	82398733,76	
$\bar{x}$		3520,87	3893				



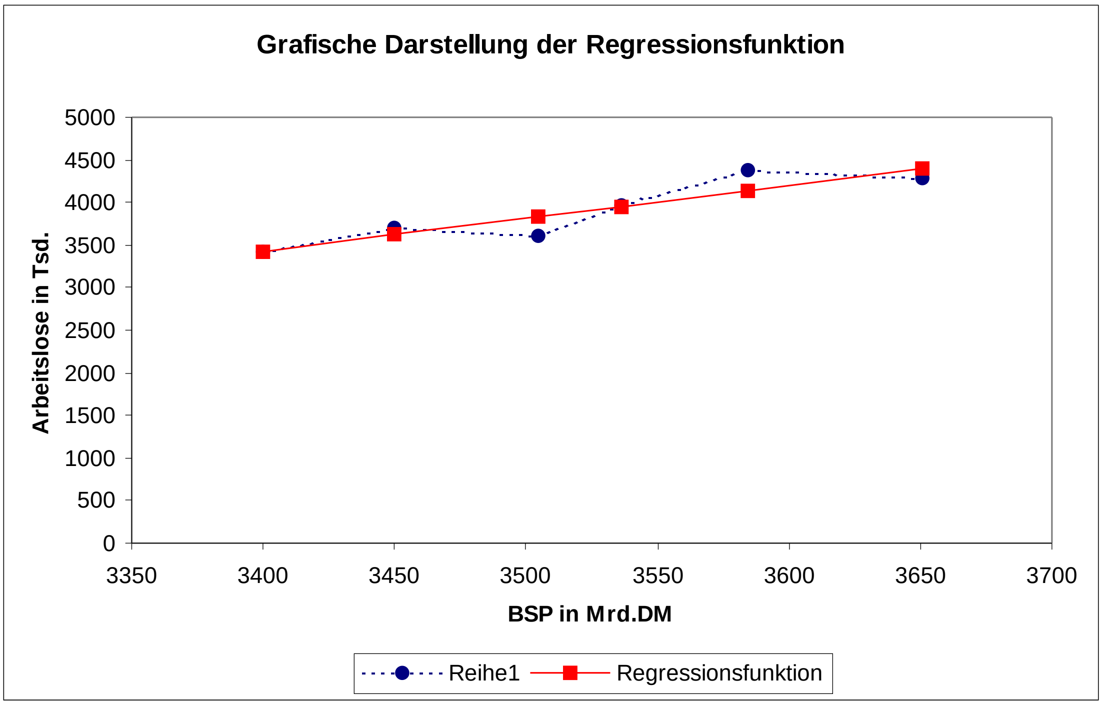
Die Steigung b der Regressionsfunktion ergibt sich als:

$$b = \frac{\frac{1}{n}\sum x_i \cdot y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{\frac{1}{6} \cdot 82398733{,}8 - 3520{,}87 \cdot 3893}{\frac{1}{6} \cdot 74420141{,}9 - 12396525{,}56} = 3{,}86$$

Für den Achsenabschnitt erhält man :

$$a = \bar{y} - b \cdot \bar{x} = 3893 - (3{,}86) \cdot 3520{,}87 = -9697{,}56$$

Die Regressionsfunktion ergibt sich somit als: $y_i^c = -9697{,}56 + 3{,}86 \cdot x$



zu f)

Um die beiden unterschiedlichen Ergebnisse (´50er Jahre = negativer Zshg., ´90er Jahre positiver Zshg.) besser interpretieren zu können, müssen weitere Faktoren bedacht werden, welche BSP und Arbeitslosigkeit bestimmen. So muss vor allem die völlig unterschiedliche Wirtschaftssituation Deutschlands in den verschiedenen Jahrzehnten berücksichtigt werden.

Außerdem wird hier Problematik der zugrunde liegenden Fragestellung (monokausale Erklärung der Arbeitslosenzahl) der Variablen deutlich. Es darf vermutet werden, dass die Zunahme des BSP in den 50er Jahren durch eine Extensivierung der Produktion (Wachstum bei in etwa konstanter Produktivität), in den 80er Jahren dagegen durch eine Intensivierung der Produktion (Wachstum bei hohen Produktivitätsfortschritten) zustande gekommen ist. Diese Frage ließe sich aber nur unter Hinzuziehung der Variablen *Beschäftigungszahl* untersuchen.

Die Berechnung einer Regressionsfunktion für den gesamten Zeitraum wäre nicht sinnvoll, da es, wie gesagt, im zeitlichen Verlauf zu Strukturbrüchen gekommen ist. Der Zusammenhang über den langen Zeitraum ist außerdem nicht linear und daher mit den hier zur Verfügung stehenden Mitteln nicht schätzbar.

zu g)

Hier wird die Regressionsfunktion mit Hilfe der Jahreszahlen (t = 0...6) und dem BSP errechnet. Es wird folgende Arbeitstabelle zugrunde gelegt:

i	Jahr	t	BSP (x_i) Mrd. DM	t^2	$t \cdot x_t$
1	1993	0	3399,7	0	0
2	1994	1	3449,7	1	3449,7
3	1995	2	3504,4	4	7008,8
4	1996	3	3536,5	9	10609,5
5	1997	4	3584,2	16	14336,8
6	1998	5	3650,7	25	18253,5
$\sum$	-	15	21125,2	55	24029,4

$\bar{t} = 2{,}5 \qquad \bar{t}^2 = 6{,}25 \qquad \bar{x} = 3520{,}87$

Es ist die Regressionsbeziehung zwischen der Zeit und dem BSP zu berechnen.

$$b = \frac{\frac{1}{n}\cdot\sum t\cdot x_t - \bar{t}\cdot\bar{x}}{\frac{1}{n}\cdot\sum t^2 - \bar{t}^2} = \frac{\frac{1}{6}\cdot 53658{,}3 - 2{,}5\cdot 3520{,}87}{\frac{1}{6}\cdot 55 - 6{,}25} = 48{,}25$$

$$a = \bar{x} - b\cdot\bar{t} = 3520{,}87 - 48{,}25\cdot 2{,}5 = 3400{,}24$$

Der Prognosewert ergibt sich über das Einsetzen in die Regressionsgleichung::

$$\hat{x} = a + b\cdot t$$

$$\hat{x}_{2000} = 3400{,}24 + 48{,}25\cdot 7 = 3737{,}99$$

Im Jahr 2000 ist mit einem Bruttosozialprodukt in Höhe von 3737,99 Mrd. DM zu rechnen

Das BSP ist damit im Vergleich zu 1998 um 87,29 Mrd. DM angestiegen.

Zu h) (Achtung Fehler im Buch: Bezugsjahr für die Prognose 2000)

$$\bar{x}_g = \sqrt[6]{\frac{3650{,}7}{3399{,}7}} - 1)\ 100 = 1{,}0119\ \_(= 1{,}19\%)$$

Bei einem Wachstum von 1,19% ergibt sich ein BSP von 3650,7*1,0119=3694,3 Mrd. DM.

Durch das geometrische Mittel ergibt sich für das Jahr 2000 ein Zuwachs von 3650,7 auf 3694,3 Mrd. DM BSP.

Zu i)

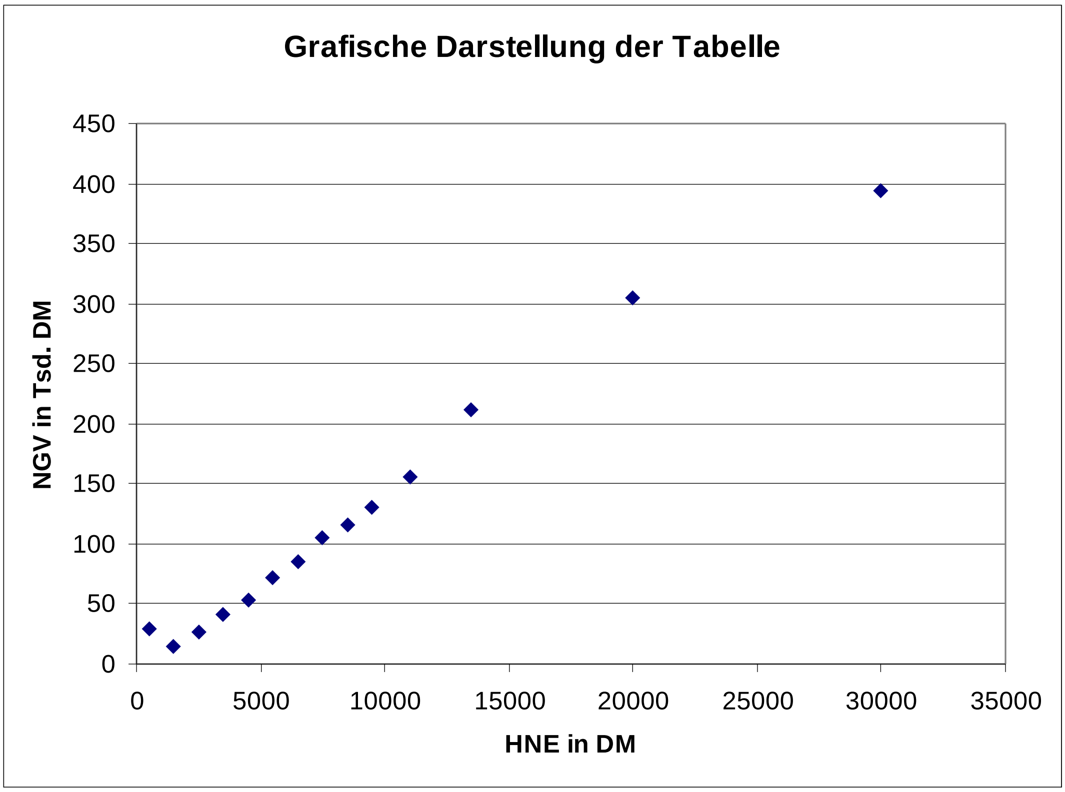
Eine Prognose auf der Basis des geometrischen Mittels orientiert sich an den durchschnittlichen prozentualen Zuwächsen pro Jahr und prognostiziert somit gleiche relative Entwicklungen. In der Regressionsanalyse wird ein linearer Trend ermittelt, der zugleichen absoluten jährliche Zuwächsen führt.

Aufgabe 23

zu a.)

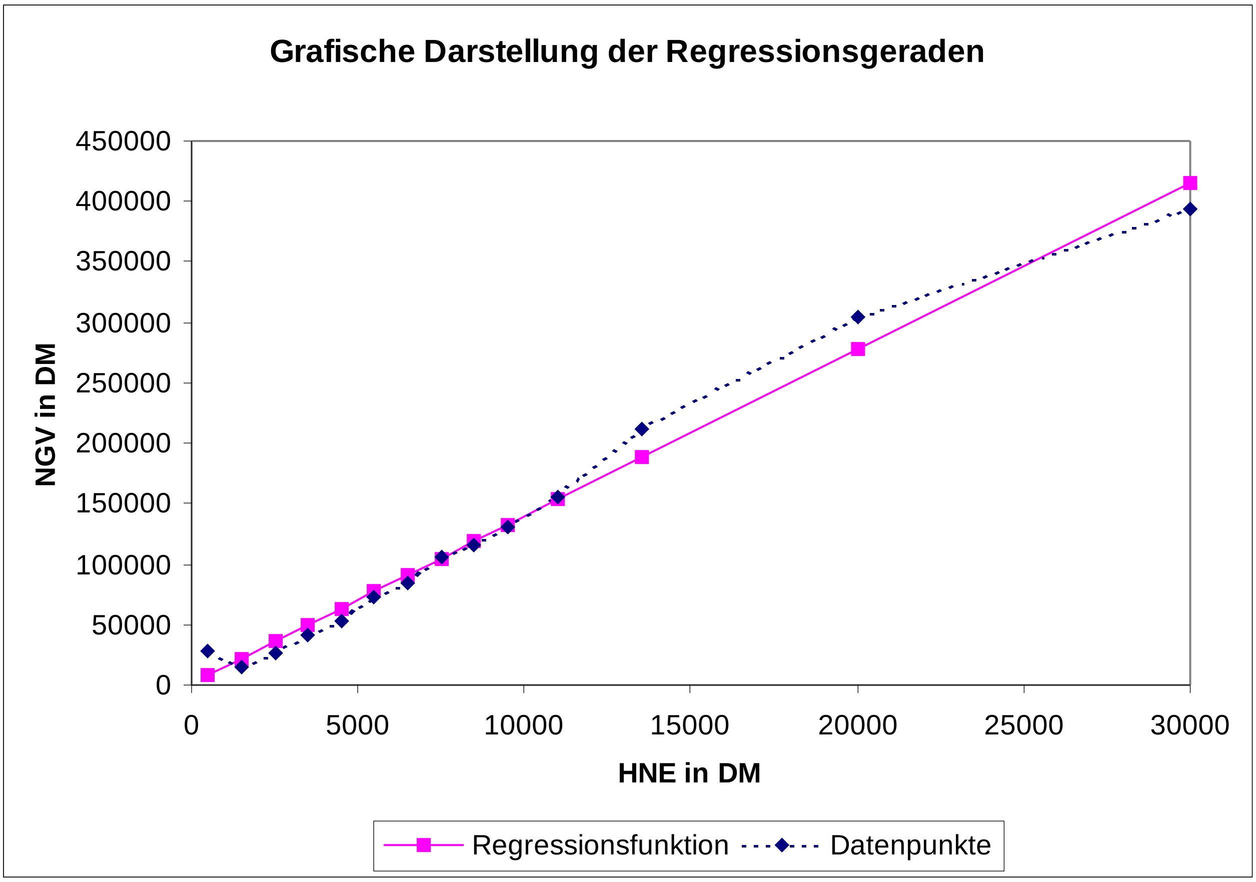
Je höher das Haushaltsnettoeinkommen (HNE), desto höher das durchschnittliche Nettogeldvermögen (NGV) der HH: NGV= f(HNE)

Zu b.) Auf der Basis der Klassenmitten der Haushaltseinkommen ergibt sich:



zu c.)

i	x_i	y_i	x^2	y^2	X*y	y_i^c
1	500	28940	250000	837523600	14470000	6914,2985
2	1500	14290	2250000	204204100	21435000	20742,8955
3	2500	26343	6250000	693953649	65857500	34571,4925
4	3500	41639	12250000	1733806321	145736500	48400,0895
5	4500	52626	20250000	2769495876	236817000	62228,6865
6	5500	72023	30250000	5187312529	396126500	76057,2835
7	6500	85080	42250000	7238606400	553020000	89885,8805
8	7500	105458	56250000	11121389764	790935000	103714,4775
9	8500	116074	72250000	13473173476	986629000	117543,0745
10	9500	130839	90250000	17118843921	1242970500	131371,6715
11	11000	155257	121000000	24104736049	1707827000	152114,567
12	13500	211244	182250000	44624027536	2851794000	186686,0595
13	20000	304589	400000000	92774458921	6091780000	276571,94
14	30000	393630	900000000	154944576900	11808900000	414857,91
∑	124500	1738032	1935750000	376826109042	26914298000	2311603,143



$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{14}\cdot 124500 = 8892{,}86$$

$$\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{14}\cdot 1738032 = 124145{,}14$$

$$\bar{x}^2 = 79082958{,}9796$$

Die Steigung b der Regressionsfunktion ergibt sich als:

$$b = \frac{\frac{1}{n}\sum x_i \cdot y_i - \bar{x}\cdot\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{\frac{1}{14}\cdot 26914298000 - (8892.86\cdot 124145.14)}{\frac{1}{14}\cdot 1935750000 - 79082958.9796} = 13{,}83$$

Für den Achsenabschnitt erhält man :

$$a = \bar{y} - b\cdot\bar{x} = 124145{,}14 - 13{,}83\cdot 8892{,}86 = 1156{,}89$$

Die Regressionsfunktion ergibt sich somit als:

y_i^c=1156,89 + 13,83 x_i

zu d.) Da zur Berechnung der Regressionsgerade klassierter Daten die Klassenmitte m_i benutzt wurde und je Einkommensklasse die durchschnittlichen Vermögensgegenstände zugrunde lagen, ist die geringe Abweichung der Daten von der Geraden verständlich. Die Grafik bestätigt die Hypothese relativ gut.

Aufgabe 24

`Der Korrelationskoeffizient Pearson´s *r* für Tabelle A:

$$r = \frac{COV(X,Y)}{s_x \cdot s_y} = \frac{\frac{1}{n} \cdot \sum x_i \cdot y_i - \bar{x} \cdot \bar{y}}{\sqrt{\frac{1}{n} \cdot \sum x_i^2 - \bar{x}^2} \cdot \sqrt{\frac{1}{n} \cdot \sum y_i^2 - \bar{y}^2}}$$

$$r = \frac{\frac{1}{6} \cdot 2307801 - 384{,}65 \cdot 1035}{\sqrt{\frac{1}{6} \cdot 903129{,}99 - 147954{,}34} \cdot \sqrt{\frac{1}{6} \cdot 6858500 - 1071225}} = -0{,}992$$

$\Rightarrow r^2 = 0{,}985$

Pearson´s r für Tabelle B:

$$r = \frac{\frac{1}{6} \cdot 82398733{,}76 - 3520{,}87 \cdot 3893}{\sqrt{\frac{1}{6} \cdot 74420141{,}92 - 12396525{,}56} \cdot \sqrt{\frac{1}{6} \cdot 91670276{,}78 - 15155449}} = 0{,}91$$

$\Rightarrow r^2 = 0{,}83$

Da Pearson´s r im Wertebereich $-1 \leq r \leq 1$ liegt, ist für Tabelle A von einem sehr hohen negativen Zusammenhang zu sprechen. Der Determinationskoeffizient r^2 beträgt 0,985. Rein theoretisch bedeutet dies, dass sich 98,46% der Arbeitslosigkeit aus dem BSP prognostizieren lassen (98,46% erklärte Varianz, Rest nicht erklärte Varianz).

Tabelle B weist mit einem Wert von r=0,91 ebenfalls einen starken, hier allerdings positiven Zusammenhang zwischen den Variablen BSP und Arbeitslosenzahlen auf.

Eine ausführliche Interpretation dieser auf den ersten Blick widersprüchlichen Ergebnisse findet sich unter Aufgabe 22 f).

Aufgabe 25

a)

Für die klassierten Daten ist statt x_i die Klassenmitte m_i zu nehmen. Dann ergeben sich folgende Werte (siehe auch: Arbeitstabelle zu Aufgabe 23 c):

Formeln: $VAR(X) = \frac{1}{N}\sum_{i=1}^{N} X_i^{\,2} - \bar{X}^2$ und $s = \sqrt{VAR(x)}$

$$VAR(X) = \frac{1}{N}\sum_{i=1}^{N} X_i^{\,2} - \bar{X}^2 = \frac{1}{14} \cdot 1935750000 - 79084737{,}56 = 59183119{,}58$$

$$s_x = \sqrt{VAR(X)} = \sqrt{59183119{,}58} = 7693{,}06$$

$$VAR(Y) = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2 = \frac{1}{14} \cdot 376826109042 - 15412015785{,}62 = 11504134860{,}24$$

$$s_y = \sqrt{VAR(Y)} = \sqrt{11504134860{,}24} = 107257{,}33$$

s_x= 7693,06; s_y= 107257,33 b= 13,83

Aus $b \cdot \frac{s_x}{s_y} = r$ ergibt sich $r = 13{,}83 \cdot \frac{7693{,}06}{107257{,}33} = 0{,}99$ und damit ein sehr starker positiver Zusammenhang.

b)

Die Varianz der abhängigen Variable y, die durch die unabhängige Variable x erklärt wird (= erklärte Varianz) lässt sich durch den Determinationskoeffizienten r^2 errechnen. Sie beträgt hier 0,98, also 98%.

c)

Da die Kovarianz sich verändert hätte, wäre r wahrscheinlich niedriger ausgefallen. Da Aggregatdaten bzw. Durchschnittswerte die individuellen Strömungen bereits ausgeschlossen haben, ist der hier erzielte hohe Determinationskoeffizient zum großen Teil dadurch bedingt. Individuelle Einkommens - und Vermögensdaten führen zwar nicht unbedingt zu einer anderen Regressionsfunktion, sicher aber zu einem wesentlich niedrigeren Korrelationskoeffizienten.

Aufgabe 26

a) Arbeitstabelle:

Produkt i	2000		2002	
	Preis (p_{0i})	Menge (m_{0i})	Preis (p_{2i})	Menge (m_{2i})
A	1,2	1000	2,0	1200
B	1,5	400	1,8	1000
C	0,5	1000	1,5	800

$$V = \frac{\sum_{i=1}^{N} p_{2i} \cdot m_{2i}}{\sum_{i=1}^{N} p_{0i} \cdot m_{0i}} \cdot 100 = \frac{2 \cdot 1200 + 1{,}8 \cdot 1000 + 1{,}5 \cdot 800}{1{,}2 \cdot 1000 + 1{,}5 \cdot 400 + 0{,}5 \cdot 1000} \cdot 100 = 234{,}783$$

Mengenindex nach Laspeyres:

$$Q_L = \frac{1200 \cdot 1{,}2 + 1000 \cdot 1{,}5 + 800 \cdot 0{,}5}{1000 \cdot 1{,}2 + 400 \cdot 1{,}5 + 1000 \cdot 0{,}5} \cdot 100 = 145{,}217$$

Mengenindex nach Paasche:

$$Q_P = \frac{1200 \cdot 2{,}0 + 1000 \cdot 1{,}8 + 800 \cdot 1{,}5}{1000 \cdot 2{,}0 + 400 \cdot 1{,}8 + 1000 \cdot 1{,}5} = 127{,}96$$

Preisindex nach Laspeyres:

$$P_L = \frac{2{,}0 \cdot 1000 + 1{,}8 \cdot 400 + 1{,}5 \cdot 1000}{1{,}2 \cdot 1000 + 1{,}5 \cdot 400 + 0{,}5 \cdot 1000} = 234{,}78$$

Preisindex nach Paasche:

$$P_P = \frac{2{,}0 \cdot 1200 + 1{,}8 \cdot 1000 + 1{,}5 \cdot 800}{1{,}2 \cdot 1200 + 1{,}5 \cdot 1000 + 0{,}5 \cdot 800} = 161{,}677$$

Aufgabe 27

Jahr	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Messzahl	100	104,5	106,4	102,7	92,7	111,8	89,1	97,3	72,7	79,1

a) Umbasierung der Messzahlen zum Basisjahr 2000

Bei der Umbasierung werden alle Indexwerte durch den Indexwert des neuen Basisjahres (hier: 97,3) geteilt und mit 100 multipliziert:

Beispiel für das Jahr 1993: $\frac{100}{97{,}3} \cdot 100 = 102{,}8$

Beispiel für das Jahr 1994: $\frac{104{,}5}{97{,}3} \cdot 100 = 107{,}4$

Umbasierter Index mit Basisjahr 2000:

Jahr	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Messzahl	102,8	107,4	109,4	105,5	95,3	114,9	91,6	100	74,7	81,3

b) Berechnung des Lieferumfangs der letzten zehn Jahre

1997 betrug der Lieferumfang 87 Tonnen.

Über die Formel der Mengenmessziffer lässt sich durch Umformung der Lieferumfang q_0 des Basisjahres 1993 bestimmen:

$$MZ_{q,t} = \frac{q_t}{q_0} \cdot 100$$

Werte des Jahres 1997 einsetzen:

$$92{,}7 = \frac{87}{q_0} \cdot 100 \quad \text{daraus folgt} \quad q_0 = \frac{87 \cdot 100}{92{,}7} = 93{,}9$$

Im Basisjahr 1993 wurden 93,9 Tonnen Kohle geliefert. Jetzt lassen sich die Lieferungen der einzelnen Jahre genau bestimmen:

Beispiel für das Jahr 1994:

$$MZ_{q,t} = \frac{q_t}{q_0} \cdot 100$$

Durch Umformung und Einsetzen der Werte ergibt sich: $q_t = \frac{104{,}5 \cdot 93{,}9}{100} = 98{,}1$

1994 wurden 98,1 Tonnen Kohle geliefert.

Jahr	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	Σ
Messzahl	100	104,5	106,4	102,7	92,7	111,8	89,1	97,3	72,7	79,1	-
Lieferung in Tonnen	93,9	98,1	99,9	96,4	87,0	105,0	83,7	91,4	68,3	74,3	898

Der Lieferumfang der letzten zehn Jahre beträgt 898 Tonnen.

Aufgabe 28

Jahr	BSP in jeweiligen Preisen	BSP in Preisen von 1995
1995	3.504,4	3.504,4
1996	3.570,9	3.536,5
1997	3.648,6	3.584,2
1998	3.758,6	3.650,7
1999	3.845,9	3.703,3
2000	3.946,9	3.815,5

a)

Indexreihe für das BSP in jeweiligen Preisen zum Basisjahr 1995

Die Indexwerte erhält man, indem man das BSP der jeweiligen Jahre durch das BSP des Jahres 1995 teilt und mit 100 multipliziert.

Beispiel für das Jahr 1996:

$$\frac{3570{,}9}{3504{,}4} \cdot 100 = 101{,}9$$

Jahr	BSP in jeweiligen Preisen	Indexreihe des BSP
1995	3.504,4	100,0
1996	3.570,9	101,9
1997	3.648,6	104,1
1998	3.758,6	107,3
1999	3.845,9	109,7
2000	3.946,9	112,6

b)

Das BSP beschreibt den Geldwert aller innerhalb eines Jahres im Inland erzeugten Güter und Dienstleistungen ($\sum_{i=1}^{N} q_{ti} \cdot p_{ti}$). Setzt man diesen Wert ins Verhältnis zum BSP eines Basisjahres ($\sum_{i=1}^{N} q_{0i} \cdot p_{0i}$) und multipliziert mit 100 (wie in Aufgabenteil a) geschehen), erhält man den jeweiligen Wert eines Volumenindexes: $V = \frac{\sum_{i=1}^{N} q_{ti} \cdot p_{ti}}{\sum_{i=1}^{N} q_{0i} \cdot p_{0i}} \cdot 100$

c)

Umbasierung der Indexreihe auf das Basisjahr 2000

Jahr	Indexreihe des BSP (Basisjahr 1995)	Umbasierter Index (Basisjahr 2000)
1995	100,0	88,8
1996	101,9	90,5
1997	104,1	92,5
1998	107,3	95,3
1999	109,7	97,4
2000	112,6	100,0

Bei der Umbasierung werden alle Indexwerte durch den Indexwert des neuen Basisjahres (hier: 112,6) geteilt und mit 100 multipliziert.

Beispiel für das Jahr 1995: $\frac{100}{112{,}6} \cdot 100 = 88{,}8$

Beispiel für das Jahr 1996: $\frac{101{,}9}{112{,}6} \cdot 100 = 90{,}5$

d)

Indexreihe der Preisentwicklung nach Paasche

Beim Preisindex nach Paasche wird das BSP in jeweiligen Preisen ins Verhältnis gesetzt zum BSP in den Preisen des Basisjahres 1995 und mit 100 multipliziert:

$$P_P = \frac{\sum_{i=1}^{N} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{N} p_{0i} \cdot q_{ti}} \cdot 100$$

Beispiel für das Jahr 1996:

$$P_P = \frac{\sum_{i=1}^{N} p_{ti} \cdot q_{ti}}{\sum_{i=1}^{N} p_{0i} \cdot q_{ti}} \cdot 100 = \frac{3570{,}9}{3536{,}5} \cdot 100 = 100{,}97$$

Entsprechend ergibt sich folgende Indexreihe (gerundet):

Jahr	BSP in jeweiligen Preisen	BSP in Preisen von 1995	Preisindex nach Paasche
1995	3.504,4	3.504,4	100,0
1996	3.570,9	3.536,5	100,1
1997	3.648,6	3.584,2	101,8
1998	3.758,6	3.650,7	103,0
1999	3.845,9	3.703,3	103,9
2000	3.946,9	3.815,5	103,4