CARL VON OSSIETZKY universität OLDENBURG

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 1 -15

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

Stand 15.04. 2005

Aufgabe 1)

a) $$\sum_{i=1}^{5} x_i = 5+3+2+1+6 = 17$$

b) $$\sum_{i=1}^{5}\left(x_i + y_i\right) = 5+2+3+3+2+4+1+1+6+0 = 27$$

c) $$\sum_{i=1}^{5}\left(x_i \cdot y_i\right) = 5\cdot 2+3\cdot 3+2\cdot 4+1\cdot 1+6\cdot 0 = 28$$

d) $$\sum_{i=1}^{5} x_i \cdot \sum_{i=1}^{5} y_i = 17\cdot 10 = 170$$

e) $$\sum_{i=1}^{5}\left(x_i + 2\right) = \sum_{i=1}^{5}\left(x_i\right) + \sum_{i=1}^{5}\left(2\right) = 17 + 5\ \ 2 = 27$$

f) $$\sum_{i=1}^{3} y_i + \sum_{i=3}^{5} y_i = 9+5 = 14$$

g) $$\sum_{i=1}^{10}\left(6a-2i+3\right) + \sum_{i=1}^{8} i - \sum_{i=2}^{5}\left(2a-2i\right) = 60a-110+30+36-8a+28 = 52a-16$$

h) $$\sum_{i=8}^{10}\left(2a+2b-1\right) + \sum_{i=1}^{14} b + \sum_{i=5}^{7}\left(5-i\right) = 6a+6b-3+14b+\left(15-18\right) = 6a+20b-6$$

i) $$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{5}(5+3+2+1+6) = \frac{17}{5} = 3{,}4$$

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{5}(2+3+4+1+0) = \frac{10}{5} = 2$$

j) $$VAR(X) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{X}^2$$

$$\begin{aligned} VAR(X) &= \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(x_i^{\,2} - 2x_i\overline{X} + \overline{X}^2) \\ &= \frac{1}{N}\sum_{i=1}^{N} x_i^{\,2} - 2\overline{X}\,\frac{1}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}\sum_{i=1}^{N}\overline{X}^2 \end{aligned}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^{\,2} - 2\overline{X}\cdot\overline{X} + \frac{1}{N}\cdot N\overline{X}^{\,2}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^{\,2} - 2\overline{X}\cdot\overline{X} + \overline{X}^{\,2}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^2 \;\; - \overline{X}^{\,2}$$

$$COV = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1}{N}\sum_{i=1}^{N} x_i\cdot y_i - \overline{X}\cdot\overline{Y}$$

$$COV = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i y_i - x_i\overline{Y} - y_i\overline{X} + \overline{X}\cdot\overline{Y})$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{Y}\,\frac{1}{N}\sum_{i=1}^{N} x_i - \overline{X}\,\frac{1}{N}\sum_{i=1}^{N} y_i + \frac{1}{N}\sum_{i=1}^{N}\overline{X\cdot Y}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{Y}\cdot\overline{X} - \overline{X}\cdot\overline{Y} + \frac{1}{N}\cdot N\overline{X}\cdot\overline{Y}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{Y}\cdot\overline{X} - \overline{X}\cdot\overline{Y} + \overline{X}\cdot\overline{Y}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{X}\cdot\overline{Y}$$

k) $$VAR(X) = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \;\; - \overline{X}^{\,2} = \frac{1}{5}(5^2 + 3^2 + 2^2 + 1^2 + 6^2) - (3{,}4)^2 = \frac{75}{5} - 11{,}56 = 3{,}44$$

$$COV = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{X}\cdot\overline{Y} = \frac{1}{5}(5\cdot 2 + 3\cdot 3 + 2\cdot 4 + 1\cdot 1 + 6\cdot 0) - (3{,}4\cdot 2) = \frac{28}{5} - 6{,}88 = -1{,}28$$

Aufgabe 2)

a)

(I)	Steuerklasse:	nominal	diskret
(II)	Geschlecht:	nominal	diskret
(III)	soziale Schicht:	ordinal	diskret
(IV)	Einkommenssteuer (-betrag):	Verhältnisskala	quasi-stetig
(V)	Temperatur		
	- in Celsius :	Intervallskala	stetig
	- in Fahrenheit:	Intervallskala	stetig
	- in Kelvin:	Verhältnisskala	stetig

(VI)	Windstärke (Bofors):	ordinal	diskret
(VII)	Körpergewicht:	Verhältnisskala	stetig
(VIII)	Schulnote:	ordinal	diskret
(IX)	Klausurpunkte:	ordinal/metrisch	diskret
(X)	Einwohnerzahl:	Absolutskala	quasi-stetig
(XI)	Semesterzahl:	Verhältnisskala	diskret
(XII)	Handelsklasse:	ordinal	diskret

Anmerkung: Als quasi-stetig kann man in der Statistik die, eigentlich diskreten, Merkmale behandeln, deren Einheit (etwa Cent oder EUR) im Vergleich zum Merkmalsbetrag sehr klein ist.

b) Monatsumsatz: Verhältnisskala

Durch die Aufstellung einer Rangordnung der Unternehmen einer Branche nach Jahresumsätzen gelangt man zu einem ordinal skalierten Merkmal. Eine solche Rangordnung sagt jedoch nichts darüber aus,

- wie sich der jeweilige Monatsumsatz zusammensetzt
- wie stabil der erzielte Umsatz ist
- wie hoch der tatsächliche Gewinn ausgefallen ist
- welche spezifischen Besonderheiten (z.B. Region) den einzelnen Unternehmen eigen sind.

c) Das Merkmal Geburtsjahr besitzt das Skalenniveau einer Intervallskala. Fasst man Abfolgen von Geburtsjahren zu Generationen zusammen, erhält man eine Ordinalskala. Geht es dagegen um die Zugehörigkeit zu Generation (68er, Nachkriegsgeneration) liegt es nahe, lediglich von einem nominalskalierten Merkmal zu sprechen, da diese Begriffe zumeist politisch-soziale Kontexte implizieren die sich nicht in eine Rangfolge bringen lassen. Derartige Zusammenfassungen erlauben zwar keine Rückschlüsse mehr über die Häufigkeit des Auftretens einzelner Geburtsjahre (Informationsverlust), zeichnen sich aber durch einen Gewinn an Übersichtlichkeit aus.

d) siehe Aufgabe a)

e) Die Frage nach den Urlaubsplänen ist zu allgemein gehalten und umfasst (ganz offensichtlich) mehrere Dimensionen. Die Erstellung einer sinnvollen Nomenklatur ist hier nahezu unmöglich, da sich die gegebenen Antworten nicht klar voneinander abgrenzen lassen. Um nicht im nachhinein Gefahr zu laufen, Äpfel mit Birnen zu vergleichen, hätte man hier (beispielsweise) direkt nach einem konkreten Urlaubsland fragen können.

Aufgabe 3)

a) + b)

Einkommen von...bis unter...EUR	absolute Häufigkeit	relative Häufigkeit	absolute Häufigkeit aufkumuliert	absolute Häufigkeit abkumuliert	relative Häufigkeit aufkumuliert	relative Häufigkeit abkumuliert
-	f_i	f_i'	$f_i^{\uparrow}$	$f_i^{\downarrow}$	$f_i'^{\uparrow}$	$f_i'^{\downarrow}$
0 - 250	23	0,14375	23	137	0,14375	0,85625
250 - 450	10	0,06250	33	127	0,20625	0,79375
450 - 550	7	0,04375	40	120	0,25000	0,75000
550 - 650	14	0,08750	54	106	0,33750	0,66250
650 - 750	29	0,18125	83	77	0,51875	0,48125
750 - 850	25	0,15625	108	52	0,67000	0,32500
850 - 950	14	0,08750	122	38	0,76250	0,23750
950 - 1050	10	0,06250	132	28	0,82500	0,17500
1050 - 1250	9	0,05625	141	19	0,88125	0,11875
1250 - 2000	15	0,09375	156	4	0,97500	0,02500
2000 und mehr	4	0,02500	160	0	1	0
Σ	160	1	-	-	-	-

Diskutiert werden sollten folgende Punkte:
- Problem der Klassifizierung bzw. Anzahl der Klassen
- Klassenbreite: gleich/ungleich
- offene Klassen: Vorteile/Nachteile
- Eindeutigkeit der Klassengrenzen
  von ... bis unter ... bzw. mehr als ... bis einschließlich ...
- Wahl der Klassenmittelpunkte: bei den gegebenen Häufungspunkten 500 EUR, 600 EUR usw.
- inhaltliche Genauigkeit der Angaben (wegen der Häufungspunkte sind es vermutlich nur ca.-Angaben)
- rechnerische Genauigkeit der rel. Häufigkeit (Stellen hinter dem Komma)
- Tabellenform: Überschrift, Kopfzeile, Vorspalte, Anmerkung, Quelle

Aufgabe 4)

a) Anzahl der Haushalte in Millionen nach Haushaltsgröße in Personen

Haushaltsgröße	Haushalte in Millionen		
in Personen	1900	1950	2000
1	0,9	3,2	13,8
2	1,8	4,2	12,7
3	2,1	3,8	5,6
4	2,1	2,7	4,4
5 und mehr	5,4	2,7	1,7
insgesamt	12,3	16,6	38,2
Anzahl der Personen	56,4	49,8	82,5

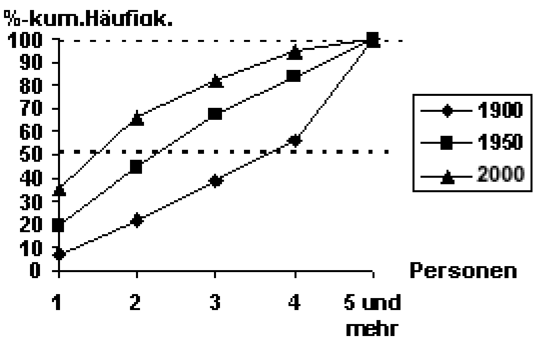
b) Tabellarische Darstellung der absoluten, relativen, prozentualen und aufkum. Häufigkeiten

Haushaltsgröße in Personen	Haushalte in Millionen					
	1900					
bis zu ... Personen	f_i	$f_i^{\uparrow}$	f_i'	$f_i'^{\uparrow}$	$f_i^{\%}$	$f_i^{\%\uparrow}$
1	0,9	0,9	0,0732	0,0732	7,32	7,32
2	1,8	2,7	0,1463	0,2195	14,63	21,95
3	2,1	4,8	0,1707	0,3902	17,07	39,02
4	2,1	3,9	0,1707	0,5609	17,07	56,09
5 und mehr	5,4	12,3	0,4390	˜ 1	43,90	˜ 100
insgesamt	12,3	-	˜ 1	-	˜ 100	-

Haushaltsgröße in Personen	Haushalte in Millionen					
	1950					
bis zu ... Personen	f_i	$f_i^{\uparrow}$	f_i'	$f_i'^{\uparrow}$	$f_i^{\%}$	$f_i^{\%\uparrow}$
1	3,2	3,2	0,1927	0,1927	19,27	19,27
2	4,2	7,4	0,2530	0,4457	25,30	44,57
3	3,8	11,2	0,2289	0,6746	22,89	67,46
4	2,7	13,9	0,1627	0,8373	16,27	83,73
5 und mehr	2,7	16,6	0,1627	1	16,27	100
insgesamt	16,6	-	1	-	100	-

Haushaltsgröße in Personen	Haushalte in Millionen					
	2000					
bis zu ... Personen	f_i	$f_i^{\uparrow}$	f_i'	$f_i'^{\uparrow}$	$f_i^{\%}$	$f_i^{\%\uparrow}$
1	13,8	13,8	0,3613	0,3613	36,13	36,13
2	12,7	26,5	0,3325	0,6938	33,25	69,38
3	5,6	32,1	0,1466	0,8404	14,66	84,04
4	4,4	36,5	0,1152	0,9556	11,52	95,56
5 und mehr	1,7	38,2	0,0445	˜ 1	4,45	˜ 100
insgesamt	38,2	-	˜ 1	-	˜ 100	-

Vergleich und Interpretation der Verteilungen



Die Graphik zeigt die aufkumulierten prozentualen Häufigkeit der obigen Tabellen. Bei 50% der kumulierten prozentualen Häufigkeit lässt sich beobachten, dass der Anteil der Single-Haushalte beständig zugenommen hat. So waren 1900 durchschnittlich 50% der Haushalte mit 3 bis 4 Personen besetzt. 1950 lag diese Zahl ungefähr bei 2 Personen und 1992 bei 1 bis 2 Personen.

c) (I) Entwicklung der durchschnittlichen Haushaltsgröße

Die Entwicklung der durchschnittlichen Haushaltsgröße ergibt sich aus dem Quotienten zwischen der Anzahl der Personen (56,6) und der Anzahl der Haushalte insgesamt (12,3) - Zahlenwerte siehe Tabelle Aufgabenteil a) - :

$$\frac{56,4}{12,3} = 4,6$$

Daraus ergibt sich folgende Tabelle

Entwicklung der durchschnittlichen Haushaltsgröße insgesamt			
	1900	1950	2000
durchschnittliche Haushaltsgröße in Personen	4,6	3,0	2,2

(II) Entwicklung der durchschnittlichen Größe der Haushalte mit fünf oder mehr Personen

Die durchschnittliche Größe der Haushalte mit fünf oder mehr Personen ergibt sich wie folgt...

Anzahl der Personen je Haushaltstyp und durchschnittlichen Haushaltsgröße der Haushalte mit 5 und mehr Personen

	1900	1950	2000
Personen insgesamt	56,4	49,8	82,5
- Personen in 1-Personen-Haushalten	0,9	3,2	13,8
- Personen in 2-Personen-Haushalten	3,6	8,4	25,4
- Personen in 3-Personen-Haushalten	6,3	11,4	16,8
- Personen in 4-Personen-Haushalten	8,4	10,8	17,6
Personen in 5 und mehr Personen-Haushalten	37,2	16	8,9
÷Haushalte mit 5 und mehr Personen (absolut)	5,4	2,7	1,7
durchschnittliche Größe der Haushalte mit 5 und mehr Personen	6,9	5,9	5,2

Aufgabe 5)

Arbeitstabelle für die prozentualen, auf- und abwärtskumulierten Häufigkeiten:

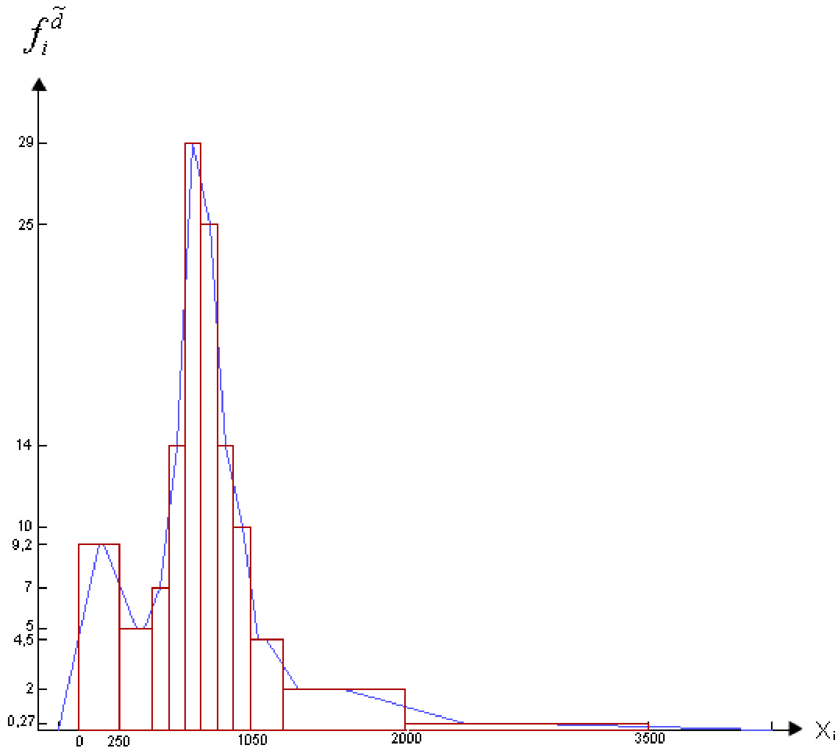
Einkommen von...bis unter...EUR	absolute Häufigkeit	relative Häufigkeit	prozentuale Häufigkeit	prozentuale Häufigkeit aufkumuliert	prozentuale Häufigkeit abkumuliert
-	f_i	f_i'	$f_i^{\%}$	$f_i^{\%\uparrow}$	$f_i^{\%\downarrow}$
0 - 250	23	0,14375	14,375	14,375	85,625
250 - 450	10	0,06250	6,25	20,625	79,375
450 - 550	7	0,04375	4,375	25	75
550 - 650	14	0,08750	8,75	33,75	66,25
650 - 750	29	0,18125	18,125	51,875	48,125
750 - 850	25	0,15625	15,625	67	32,5
850 - 950	14	0,08750	8,75	76,25	23,75
950 - 1050	10	0,06250	6,25	82,5	17,5
1050 - 1250	9	0,05625	5,625	88,125	11,875
1250 - 2000	15	0,09375	9,375	97,5	2,5
2000 - 3500	4	0,02500	2,5	100	0
Σ	160	1	100	-	-

Arbeitstabelle für das Histogramm:

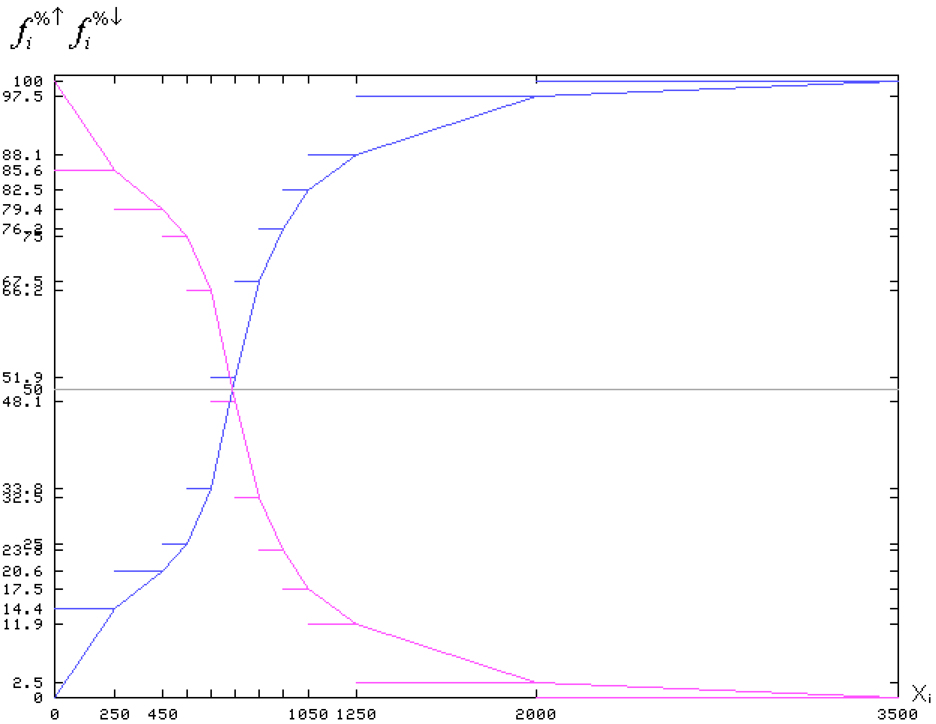
Einkommen von...bis unter...EUR	absolute Häufigkeit	Klassen-breite	Klassen-mitte	Häufigkeits-dichte	mod. Häufigkeitsdichte[1]
-	f_i	c_i	m_i	f_i^d	$f_i^{\tilde{d}}$
0 - 250	23	250	125	0,0920	9,2
250 - 450	10	200	350	0,0500	5
450 - 550	7	100	500	0,0700	7
550 - 650	14	100	600	0,1400	14
650 - 750	29	100	700	0,2900	29
750 - 850	25	100	800	0,2500	25
850 - 950	14	100	900	0,1400	14
950 - 1050	10	100	1000	0,1000	10
1050 - 1250	9	200	1150	0,0450	4,5
1250 - 2000	15	750	1625	0,0200	2
2000 - 3500	4	1500	2750	0,0027	0,27
Σ	160	-	-		-

[1] $f_i^{\tilde{d}} = \frac{f_i}{c_i} \cdot \tilde{c}$ (hier mit $\tilde{c} = 100$)

Histogramm und Polygonzug:



Treppenfunktion und Summenpolygone:



Aufgabe 6)

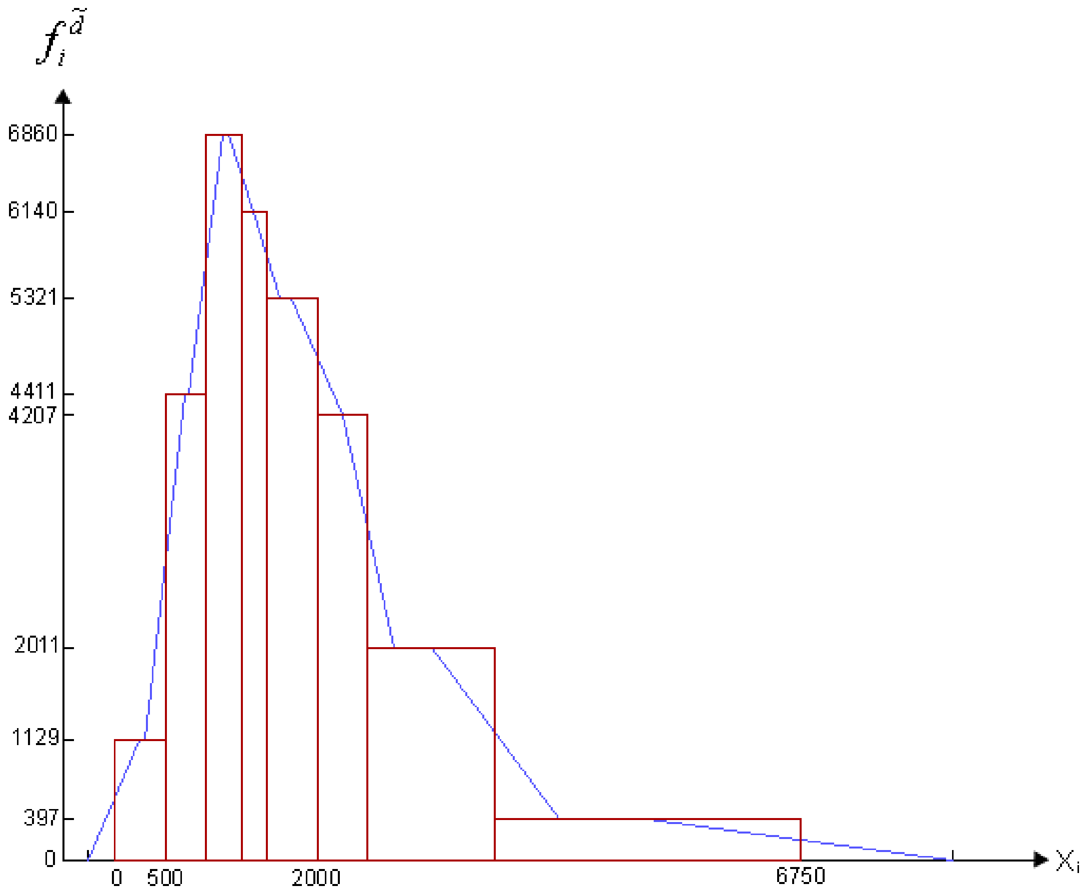
(I) Häufigkeitsverteilung des früheren Bundesgebiets

Arbeitstabelle

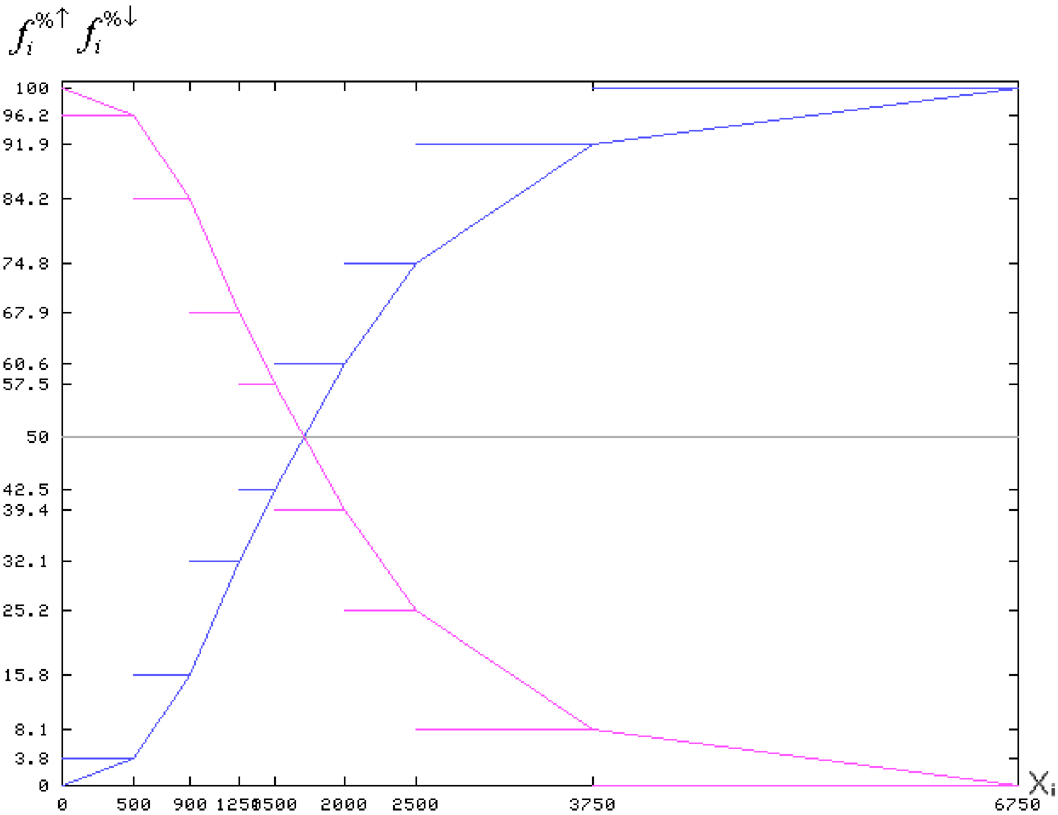
Einkommen in EUR	absolute Häufigkeit	relative Häufigkeit	aufkum. relative Häufigkeit	aufkum. Prozentuale Häufigkeit	Klassen-breite	Klassen-mitte	Häufigkeits dichte	modifizierte Häufigkeits dichte[1]
von...bis unter...	f_i	f_i'	$f_i'^{\uparrow}$	$f_i^{\%\uparrow}$	c_i	m_i	f_i^{d}	$f_i^{\tilde{d}}$
0-500	1129	0,0383	0,0383	3,8	500	250	2,2580	1129
500-900	3529	0,1197	0,1580	15,8	400	700	8,8225	4411,25
900-1250	4802	0,1629	0,3209	32,09	350	1075	13,7200	6860
1250-1500	3070	0,1042	0,4251	42,51	250	1375	12,2800	6140
1500-2000	5321	0,1805	0,6056	60,56	500	1750	10,6420	5321
2000-2500	4207	0,1427	0,7483	74,83	500	2250	8,4140	4207
2500-3750	5029	0,1706	0,9189	91,89	1250	3125	4,0232	2011,6
3750-6750	2385	0,0809	≈1	≈100	3000	5250	0,7950	397,5
Σ	29472	≈1	-	-	-	-	-	-

[1] Der modifizierten Häufigkeitsdichte liegt ein $\tilde{c} = 500$ zugrunde

Histogramm und Polygonzug:



Treppenfunktion und Summenpolygon:



Anmerkung: Statt der relativen, aufkumulierten Häufigkeiten sind in der Graphik die prozentualen, aufkumulierten Häufigkeiten eingezeichnet (die natürlich mit den relativen identisch sind, es ändert sich nur die Skalierung der Y-Achse). Der Vollständigkeit halber sind auch die abwärtskumulierten, prozentualen Häufigkeiten (pink) sowie die jeweiligen Summenpolygone mit eingezeichnet.

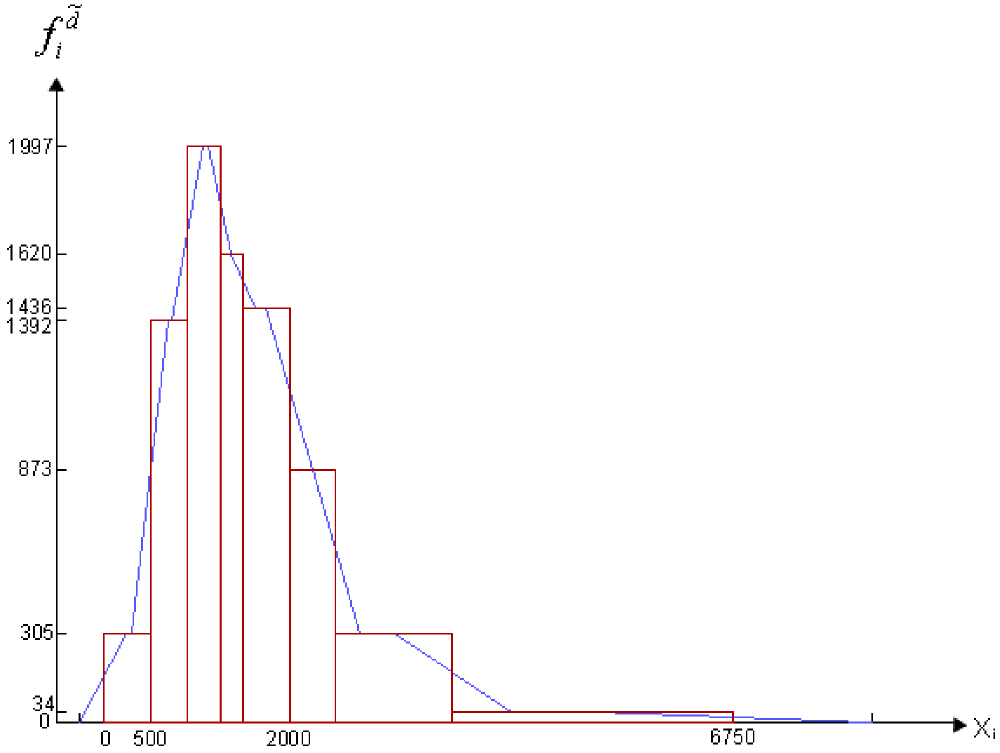
(II) Häufigkeitsverteilung der neuen Bundesländer

Arbeitstabelle

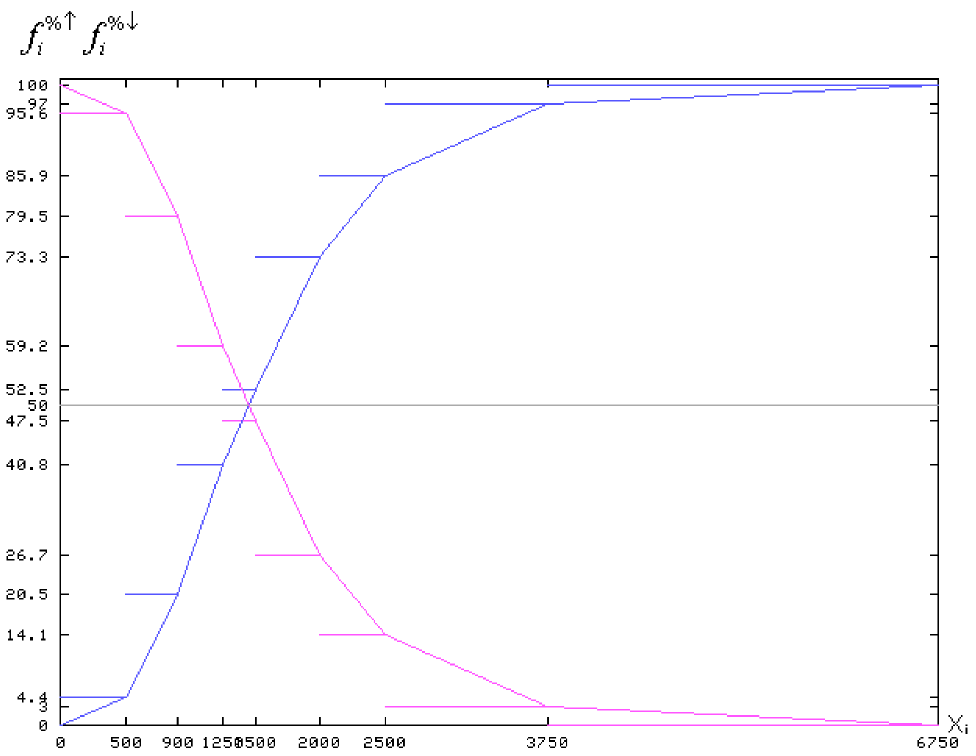
Einkommen in EUR	absolute Häufigkeit	relative Häufigkeit	aufkum. relative Häufigkeit	aufkum. prozentuale Häufigkeit	Klassen-breite	Klassen-mitte	Häufigkeits dichte	modifizierte Häufigkeits dichte[1]
von...bis unter...	f_i	f_i'	$f_i'^{\uparrow}$	$f_i^{\%\uparrow}$	c_i	m_i	f_i^{d}	$f_i^{\tilde{d}}$
0-500	305	0,0441	0,0441	4,41	500	250	0,6100	305
500-900	1114	0,1612	0,2053	20,53	400	700	2,7850	1392,5
900-1250	1398	0,2023	0,4076	40,76	350	1075	3,9943	1997,15
1250-1500	810	0,1172	0,5248	52,48	250	1375	3,2400	1620
1500-2000	1436	0,2078	0,7326	73,26	500	1750	2,8720	1436
2000-2500	873	0,1263	0,8589	85,89	500	2250	1,7460	873
2500-3750	770	0,1114	0,9703	97,03	1250	3125	0,6160	308
3750-6750	205	0,0297	≈1	≈100	3000	5250	0,0683	34,15
Σ	6911	≈1	-	-	-	-	-	-

[1] Der modifizierten Häufigkeitsdichte liegt ein $\tilde{c} = 500$ zugrunde

Histogramm und Polygonzug:



Treppenfunktion und Summenpolygon:



Anmerkung: Statt der relativen, aufkumulierten Häufigkeiten sind in der Graphik die prozentualen, aufkumulierten Häufigkeiten eingezeichnet (die natürlich mit den relativen identisch sind, es ändert sich nur die Skalierung der Y-Achse).

Der Vollständigkeit halber sind auch die abwärtskumulierten, prozentualen Häufigkeiten (pink) sowie die jeweiligen Summenpolygone mit eingezeichnet.

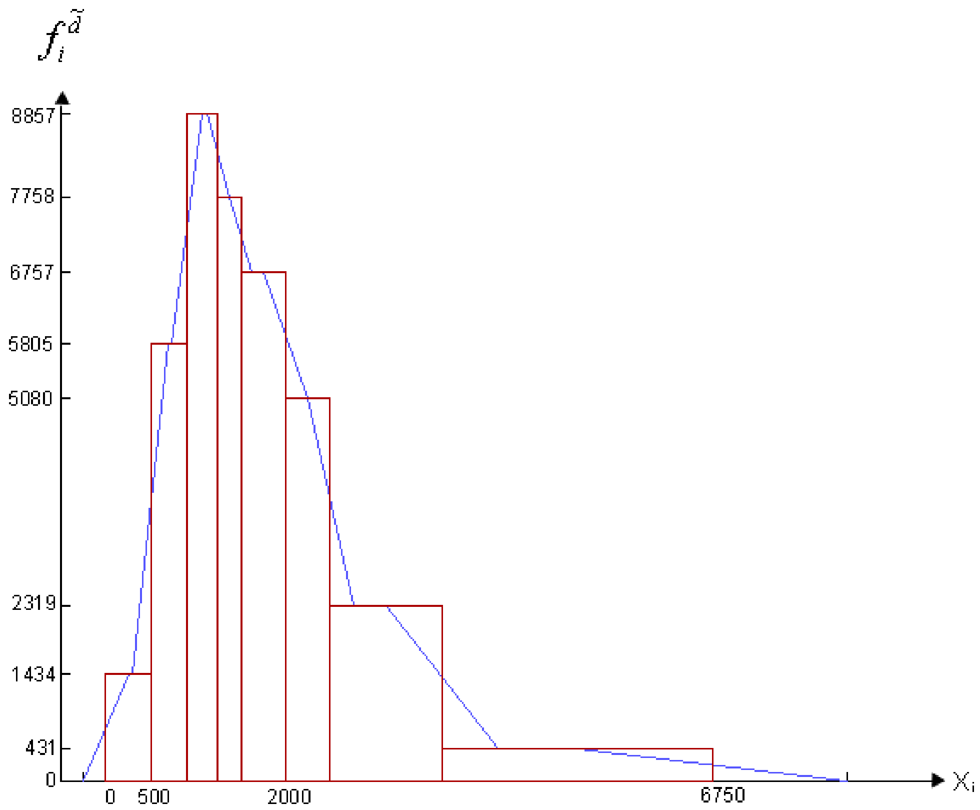
(III) Häufigkeitsverteilung des gesamten Bundesgebiets

Arbeitstabelle:

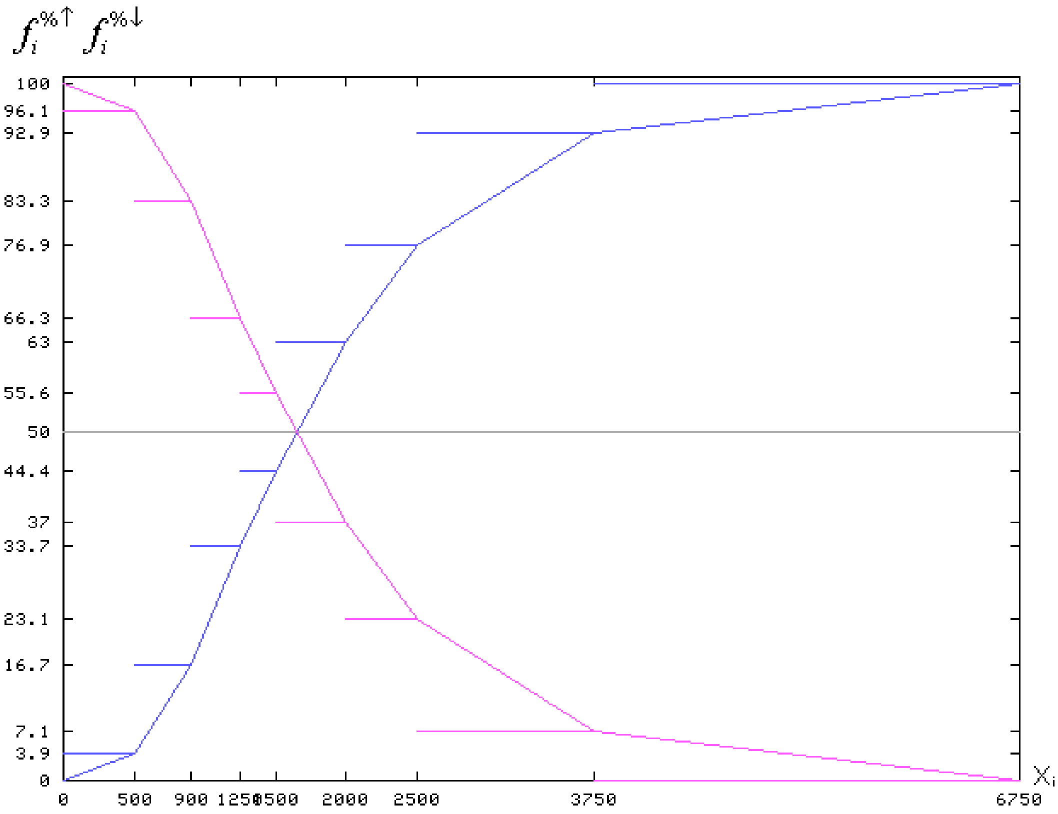
Einkommen in EUR	absolute Häufigkeit	relative Häufigkeit	aufkum. relative Häufigkeit	aufkum. Prozentuale Häufigkeit	Klassen-breite	Klassen-mitte	Häufigkeits dichte	modifizierte Häufigkeits dichte[1]
von...bis unter...	f_i	f_i'	$f_i'^{\uparrow}$	$f_i^{\%\uparrow}$	c_i	m_i	f_i^{d}	$f_i^{\tilde{d}}$
0-500	1434	0,0394	0,0394	3,94	500	250	2,8680	1434
500-900	4644	0,1276	0,1670	16,7	400	700	11,6100	5805
900-1250	6200	0,1704	0,3374	33,74	350	1075	17,7143	8857,15
1250-1500	3879	0,1066	0,4440	44,40	250	1375	15,5160	7758
1500-2000	6757	0,1857	0,6297	62,97	500	1750	13,5140	6757
2000-2500	5080	0,1396	0,7693	76,93	500	2250	10,1600	5080
2500-3750	5799	0,1594	0,9287	92,87	1250	3125	4,6392	2319,6
3750-6750	2590	0,0712	≈1	≈100	3000	5250	0,8633	431,65
Σ	36383	≈1	-	-	-	-	-	-

[1] Der modifizierten Häufigkeitsdichte liegt ein $\tilde{c} = 500$ zugrunde

Histogramm und Polygonzug:



Treppenfunktion und Summenpolygon:



Anmerkung: Statt der relativen, aufkumulierten Häufigkeiten sind in der Graphik die prozentualen, aufkumulierten Häufigkeiten eingezeichnet (die natürlich mit den relativen identisch sind, es ändert sich nur die Skalierung der Y-Achse). Der Vollständigkeit halber sind auch die abwärtskumulierten, prozentualen Häufigkeiten (pink) sowie die jeweiligen Summenpolygone mit eingezeichnet.

Aufgabe 7)

a) Bestimmung der Werte für ungruppierte Daten

(I) Bestimmung des Modus

Der Modus ist aus der Tabelle der Aufgabe 3 zu entnehmen. Er ist definiert als die Merkmalsausprägung, die am häufigsten vorkommt. Damit ergibt sich Mod = 0.

(II) Bestimmung des Median

Der Median ist definiert als die Merkmalsausprägung desjenigen Elementes, das in der Mitte der (der Größe nach geordneten) Beobachtungsreihe in der Mitte steht. Da N hier gerade ist, gilt:

$$Med = \frac{1}{2}\left(X_{\frac{N}{2}} + X_{\frac{N}{2}+1}\right) = \frac{1}{2}(731 + 734) = 732{,}5$$

(III) Bestimmung des arithmetischen Mittels

Die Berechnung des arithmetischen Mittels ist nur sinnvoll für metrisch skalierte Daten. Diese Eigenschaft liegt bei dem Merkmal „Höhe des Einkommens“ vor. Rechnerisch ergibt sich das arithmetische Mittels als

$$\bar{x} = \frac{1}{N} \cdot \sum x_i \cdot f_i = \frac{1}{160} \cdot 129755 = 810{,}969$$

b) Arbeitstabelle

Einkommen von...bis unter...EUR	absolute Häufigkeit	absolute Häufigkeit aufkum.	Klassen-breite	Klassen-mitte	Häufigkeits-dichte	mod. Häufigkeitsdichte mit $\tilde{c} = 100$	
-	f_i	$f_i^{\uparrow}$	c_i	m_i	f_i^d	$f_i^{\tilde{d}}$	$m_i \cdot f_i$
0 - 250	23	23	250	125	0,0920	9,2	2875
250 - 450	10	33	200	350	0,0500	5	3500
450 - 550	7	40	100	500	0,0700	7	3500
550 - 650	14	54	100	600	0,1400	14	8400
650 - 750	29	83	100	700	0,2900	29	20300
750 - 850	25	108	100	800	0,2500	25	20000
850 - 950	14	122	100	900	0,1400	14	12600
950 - 1050	10	132	100	1000	0,1000	10	10000
1050 - 1250	9	141	200	1150	0,0450	4,5	10350
1250 - 2000	15	156	750	1625	0,0200	2	24375
2000 - 3500	4	160	1500	2750	0,0027	0,27	11000
Σ	160	-	-	-	-	-	126900

b1)

Berechnung des arithmetischen Mittels:

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{k} m_i \cdot f_i = \frac{1}{160} \cdot 126900 = 793{,}125$$

feinberechneter Modus:

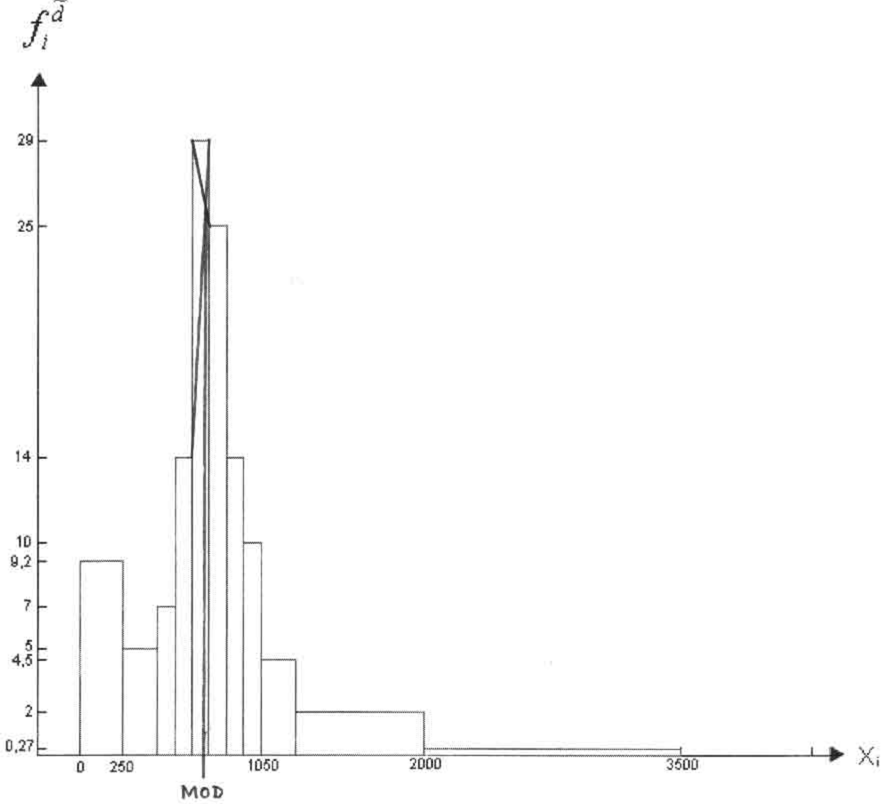
$$Mod = L_{Mod} + c_{Mod} \cdot \left[\frac{\tilde{f}^{d}_{Mod} - \tilde{f}^{d}_{Mod-1}}{2\tilde{f}^{d}_{Mod} - (\tilde{f}^{d}_{Mod-1} + \tilde{f}^{d}_{Mod+1})} \right] = 650 + 100 \cdot \left[\frac{29 - 14}{2 \cdot 29 - (14 + 25)} \right] = 728{,}9474$$

feinberechneter Median:

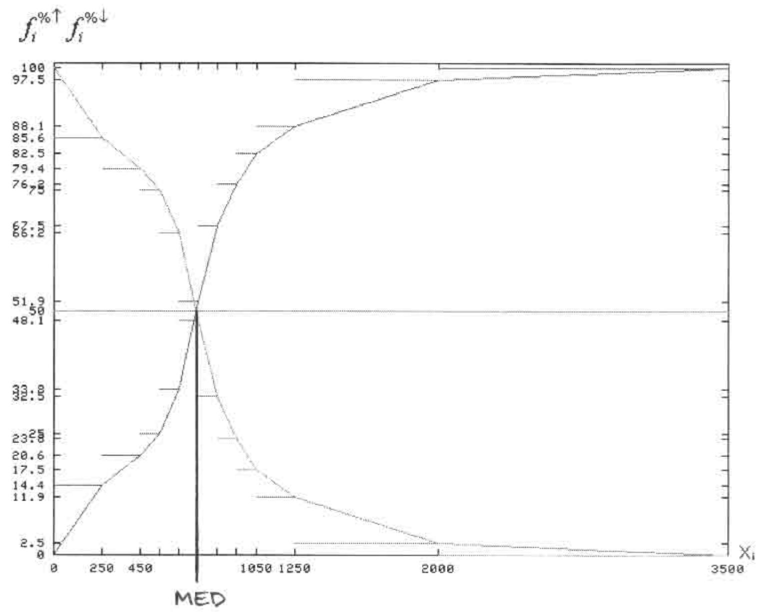
$$Med = L_{Med} + c_{Med} \cdot \left(\frac{\frac{N}{2} - (\Sigma f)_{L_{Med}}}{f_{Med}} \right) = 650 + 100 \cdot \left(\frac{80 - 54}{29} \right) = 739{,}6552$$

b2) Graphische Bestimmung des Modus und des Medians

(I) Bestimmung des Modus



(II) Bestimmung des Medians



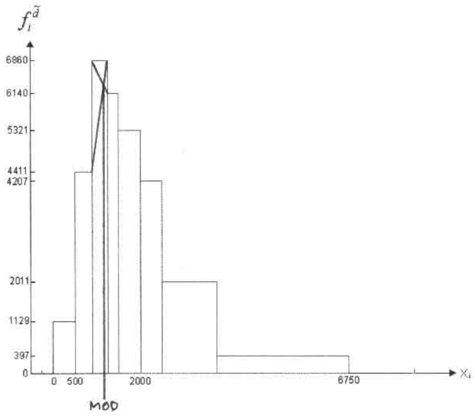
Vergleich der errechneten und grafisch ermittelten Mittelwerte			
Werte auf Basis der...	Modus	Median	$\overline{X}$
...gruppierten Tabelle	0	732,5	810,969
...klassierten Tabelle	728,9474	739,6552	793,125
...grafischen Bestimmung	ca. 730	ca. 750	-

Die Berechnung auf Basis der gruppierten Daten liefert exakte Werte. Durch den Informationsverlust der Klassierung weichen die feinberechneten Mittelwerte notwendigerweise davon ab, die Abweichung ist aber so gering, dass sie ihre Aussagekraft nicht verlieren. Der feinberechnete Modus besitzt zudem sicherlich mehr empirische Relevanz (im Sinne einer ‚Maßzahl der zentralen Tendenz') als der direkt aus der Tabelle abgelesene Wert von 0. Die grafisch ermittelten Mittelwerte liefern aufgrund von Zeichnungs- und Ableseungenauigkeiten per se nur Näherungswerte.

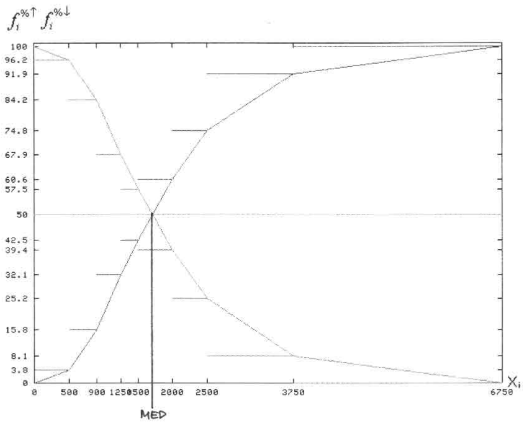
Aufgabe 8)

(I) Mittelwerte des früheren Bundesgebiets

Modus:



Median:



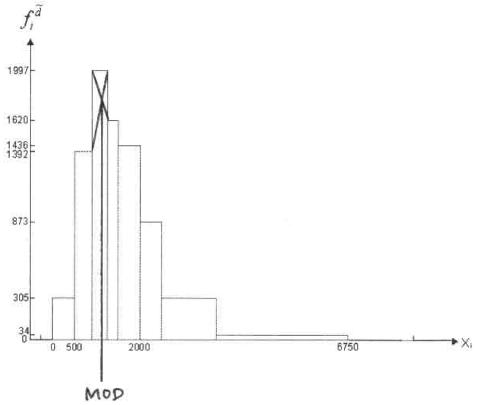
arithmetisches Mittel:

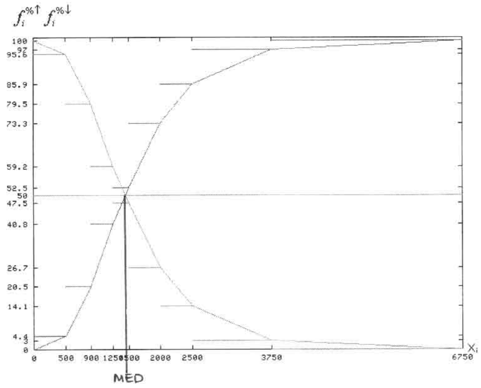
$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} m_i \cdot f_i = \frac{1}{29480} \cdot 59141325 = 2006{,}15$$

(II) Mittelwerte der neuen Bundesländer

Modus:

Median:





arithmetisches Mittel:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} m_i \cdot f_i = \frac{1}{6911} \cdot 11432400 = 1654{,}23$$

Vergleich der errechneten Werte			
	Modus	Median	$\overline{X}$
früheres Bundesgebiet	ca. 1200	ca. 1700	2006,15
neue Bundesländer	ca. 1150	ca. 1450	1654,23

Während das am häufigsten genannte Nettoeinkommen der privaten Haushalte in Ost und West noch annähernd gleich hoch ausfällt, zeigen sich bei den anderen Mittelwerten deutliche Unterschiede: In den alten Bundesländern verfügen rund 50 % der Haushalte über ein Nettoeinkommen von mehr (oder weniger) als 1700 € im Monat, in den neuen Bundesländern dagegen liegt der Median deutlich darunter. Hier verfügen 50 % der Haushalte über ein Nettoeinkommen von mehr als (weniger als) 1450 €. Daneben fällt das durchschnittliche Haushaltseinkommen in den alten Bundesländern um rund 350 € höher aus als im Osten.

Aufgabe 9)

$n = 3$ Jahre
$y_3 = 4000000$
$y_0 = 1000000$

$$\Rightarrow x_{GM} = \sqrt[n]{\frac{y_n}{y_0}} - 1 = \sqrt[3]{\frac{4000000}{1000000}} - 1 = \sqrt[3]{4} - 1 = 0{,}587 = 58{,}7\%$$

Aufgabe 10)

a) Durchschnittliche jährliche Zuwachsraten der 5-Jahresabschnitte

x_{GM}: jährliche Zuwachsrate
y_i: Bevölkerungszahl zum Zeitpunkt i

t: betrachteter Zeitraum

$$x_{GM} = \left(\sqrt[t]{\frac{y_n}{y_0}} - 1\right) \cdot 100$$

Arbeitstabelle:

Zeitraum \ Region	Bevölkerung der Welt	Bevölkerung der entwickelten Regionen	Bevölkerung der unterentwickelten Regionen
1975 - 1980	1,79%	0,70%	2,20%
1980-1985	1,77%	0,67%	2,15%
1985-1990	1,82%	0,65%	2,19%
1990-1995	1,77%	0,63%	2,11%
1995-2000	1,72%	0,46%	2,07%
1975 - 2000	1,78%	0,62%	2,14%

b) Identität von Zinseszinsformel und geometrischem Mittel

$$x_{GM} = \left(\sqrt[t]{\frac{y_n}{y_0}} - 1\right) \cdot 100 \Leftrightarrow y_n = y_0 \cdot \left(1 + \frac{x_{GM}}{100}\right)^t = y_0 \cdot (1+i)^t$$

mit i: Zinssatz

Aufgabe 11)

a) Ausgegangen wird von der Definition der Durchschnittskosten

$$Durchschnittskosten = \frac{Gesamtkosten}{Gesamtmenge} = \frac{160+180+210+250}{4000} \cdot \left[\frac{EUR}{l}\right] = 0{,}20 \cdot \left[\frac{EUR}{l}\right]$$

Dieser Wert entspricht dem arithmetischen Mittel des Literpreises. Der durchschnittliche Preis ist unabhängig von der gekauften Menge, wenn diese in jeder Periode gleich ist.

b) Auch hier wird von der Definition der Durchschnittskosten ausgegangen:

$$DK = \frac{GK}{GM} = \frac{800}{1250 + 1111{,}1 + 952{,}4 + 800} \cdot \left[\frac{EUR}{l}\right] = 0{,}195 \cdot \left[\frac{EUR}{l}\right]$$

Dieser Wert entspricht dem harmonischen Mittel der Literpreise:

$$\overline{x_{HM}} = \frac{4 \cdot 200}{\frac{200}{0{,}16} + \frac{200}{0{,}18} + \frac{200}{0{,}21} + \frac{200}{0{,}25}} \cdot \left[\frac{EUR}{l}\right] = \frac{4}{\frac{1}{0{,}16} + \frac{1}{0{,}18} + \frac{1}{0{,}21} + \frac{1}{0{,}25}} \cdot \left[\frac{EUR}{l}\right] = 0{,}195 \cdot \left[\frac{EUR}{l}\right]$$

Das Ergebnis ist unabhängig von dem Betrag, der in jeder Periode ausgegeben wird, wenn dieser in jeder Periode gleich hoch ist.

Aufgabe 12)

(a) Bestimmung der Streuungsparameter für die unklassierten Daten der Aufgabe 3):

● Quartile:

$$X_{\frac{N}{4}} \leq Q_1 \leq X_{\frac{N}{4}+1},$$

$$X_{\frac{N}{4}} = X_{40} = 500; X_{\frac{N}{4}+1} = X_{41} = 550$$

$$Q_1 = 525$$

$$Q_2 = Med = \frac{1}{2} \quad X_{\frac{N}{2}} + x_{\frac{N}{2}+1} \quad = \frac{1}{2}\left(731 + 734\right) = 732{,}5$$

$$X_{\frac{3N}{4}} \leq Q_3 \leq X_{\frac{3N}{4}+1},$$

$$X_{\frac{3N}{4}} = X_{120} = 915; X_{\frac{3N}{4}+1} = X_{121} = 920$$

$$Q_3 = 917{,}5$$

● Quartilsabstand:

$$\text{QA} = \text{Q}_3 - \text{Q}_1 = 917{,}5 - 525 = 392{,}5$$

● Semiquartilsabstand (ergänzend):

$$\text{SQA} = \frac{1}{2}\left(\text{Q}_3 - \text{Q}_1\right) = \frac{1}{2} \; 392{,}5 = 196{,}25$$

● Spannweite (ergänzend)

$$R = 3300 - 0 = 3300$$

Die Berechnung der Varianz und der Standardabweichung gelingt am einfachsten und übersichtlichsten durch die Aufstellung einer geeigneten **Arbeitstabelle**:

Einkommen	absolute Häufigkeit	-	-	-
x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
[...]	[...]	[...]	[...]	[...]
Summe	$\sum f_i$	$\sum x_i \cdot f_i$	-	$\sum x_i^2 \cdot f_i$

Für das arithmetische Mittel gilt: $\bar{x} = \frac{1}{N} \cdot \sum x_i \cdot f_i$

Für die Varianz gilt: $VAR(x) = \frac{1}{N} \cdot \sum_{j=1}^{k} \left(x_i - \bar{x}\right)^2 \cdot f_i = \frac{1}{N} \cdot \sum_{i=1}^{k} x_i^2 \cdot f_i - \bar{x}^2$

In der vorliegenden Aufgabe sind die Summenterme jedoch schon errechnet (Zwischenergebnis vom Aufgabenblatt). Das arithmetische Mittel wurde bereits in Aufgabe 7) mit $\bar{x} = 810{,}969$ für unklassierte Daten der Aufgabe 3) bestimmt.
Unter Berücksichtigung dieser Zwischenergebnisse ist die Berechnung der Varianz ohne explizite Aufstellung der Arbeitstabelle möglich:

$$VAR(x) = \frac{1}{160} \cdot 132789809 - 810{,}969^2 = 172265{,}5873$$

Damit ergibt sich für die Standardabweichung:

$$s = \sqrt{VAR(x)} = \sqrt{172265{,}5873} = 415{,}05$$

(b) Bestimmung der Streuungsparameter für die klassierten Daten aus Aufgabe 3) zum Vergleich:

- Quartile:

$$Q_1 = L + c \cdot \left(\frac{\frac{1}{4} \cdot n - \left(\sum f\right)_L}{f_{Q_1}} \right) = 450 + 100 \cdot \left(\frac{40 - 33}{7} \right) = 550$$

$$Q_2 = Med = 739{,}655$$

$$Q_3 = L + c \cdot \left(\frac{\frac{3}{4} \cdot n - \left(\sum f\right)_L}{f_{Q_3}} \right) = 850 + 100 \cdot \left(\frac{120 - 108}{14} \right) = 935{,}71$$

- Quartilsabstand:

$$QA = Q_3 - Q_1 = 935{,}71 - 550 = 385{,}71$$

- Semiquartilsabstand:

$$SQA = \frac{1}{2} \cdot \left(Q_3 - Q_1\right) = 192{,}86$$

Zur Berechnung der mittleren absoluten Abweichung MA, der Standardabweichung s sowie der Varianz VAR generieren wir der Übersicht halber die folgende Arbeitstabelle:

Einkommen von...bis unter	absolute Häufigkeit	Klassen-mitte	absolute Häufigkeit aufkum.	-	-	-	-
-	f_i	m_i	$f_i^{\uparrow}$	$\lvert m_i - \bar{x} \rvert$	$f_i \cdot \lvert m_i - \bar{x} \rvert$	m_i^2	$f_i \cdot m_i^2$
0 - 250	23	125	23	668,125	15366,875	15625	359375
250 - 450	10	350	33	443,125	4431,25	122500	1225000
450 - 550	7	500	40	293,125	2051,875	250000	1750000
550 - 650	14	600	54	193,125	2703,75	360000	5040000
650 - 750	29	700	83	93,125	2700,625	490000	14210000
750 - 850	25	800	108	6,875	171,875	640000	16000000
850 - 950	14	900	122	106,875	1496,25	810000	11340000
950 - 1050	10	1000	132	206,875	2068,75	1000000	10000000
1050 - 1250	9	1150	141	356,875	3211,875	1322500	11902500
1250 - 2000	15	1625	156	831,875	12478,125	2640625	39609375
2000 - 3500	4	2750	160	1956,875	7827,5	7562500	30250000
Summe	160	-	-	5156,875	54508,75	15213750	141686250

- Mittlere absolute Abweichung:

$$MA = \frac{1}{n} \cdot \sum_{i=1}^{k} f_i \cdot \left| m_i - \bar{x} \right| = \frac{1}{160} \cdot 54508{,}75 = 340{,}6797$$

- Bestimmung der Varianz

$$VAR(x) = s^2 = \frac{1}{n} \cdot \sum_{i=1}^{k} f_i \cdot m_i^2 - \bar{x}^2$$

$$VAR(x) = s^2 = \frac{1}{160} \cdot 141686250 - 793{,}125^2 = 885539{,}0625 - 629047{,}2656 = 256491{,}7969$$

- Bestimmung der Standardabweichung:

$$s = \sqrt{VAR(x)} = \sqrt{s^2} = 506{,}4502$$

- Bestimmung des Variationskoeffizienten:

$$V = \frac{s}{\bar{x}} = \frac{506{,}4502}{793{,}125} = 0{,}6386$$

c) Welche Informationen lassen sich aus den Maßzahlen gewinnen?

Lage- und Streuungsparameter dienen dazu, von der beobachteten Verteilung eines Merkmals zu abstrahieren bzw. die Gesamtheit der einzelnen Merkmalsausprägungen zu bestimmten Maßzahlen zu verdichten. Dabei haben die Streuungsparameter an sich noch keine Aussagekraft. Erst die Betrachtung mehrerer vergleichbarer Verteilungen liefert Erkenntnisse darüber, welche Besonderheiten eine bestimmte Merkmalsverteilung aufweist.

d) Vergleich der errechneten Maßzahlen

Maßzahl	unklassierte Daten	klassierte Daten
Q_1	525	550
Q_2	732,5	739,655
Q_3	917,5	935,71
QA	392,5	385,71
SQA	196,25	192,86
S	3300	-
$\bar{x}$	810,969	793,125
VAR	172265,6	256491,7969
s	415,05	506,4502
MA	-	347,6797
V	0,512	0,6386

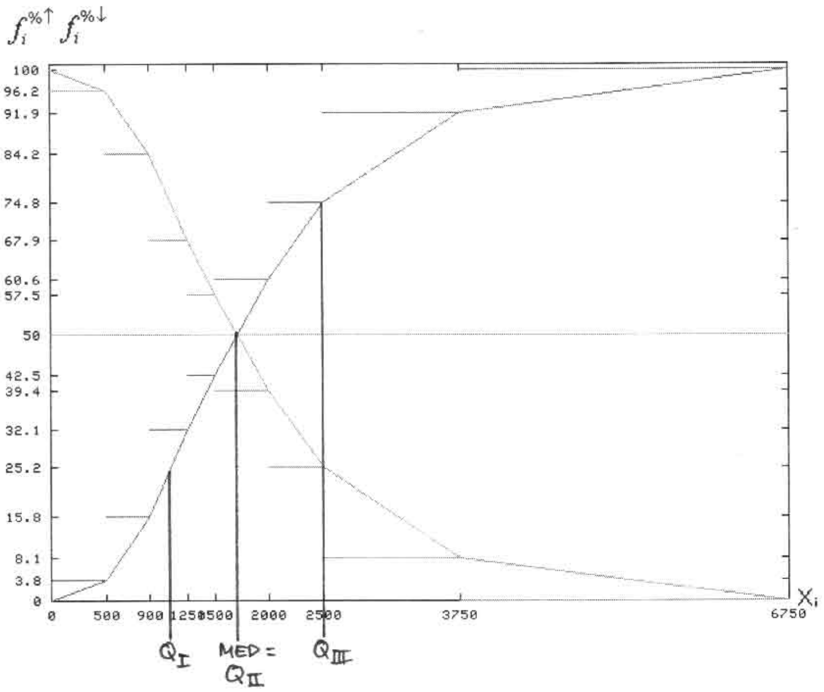
Aus der obigen Gegenüberstellung der unklassierten und klassierten Daten der Aufgabe 3 sollte folgendes deutlich werden:

- die für die klassierten Daten gezogenen Klassengrenzen sollten so bestimmt werden, dass der Informationsverlust möglichst gering gehalten wird. Man stelle sich dazu die Arbeitstabelle mit folgende Klassengrenzen vor [(0...1500),(1500...3000)]. Alle wichtigen Informationen, die sich im Bereich [400...1250] finden lassen, gehen damit verloren
- die für die klassierten Daten gezogenen Klassengrenzen sollten so bestimmt werde, dass die Verteilung der einzelnen Merkmalsausprägungen innerhalb der Klassen möglichst homogen ist.

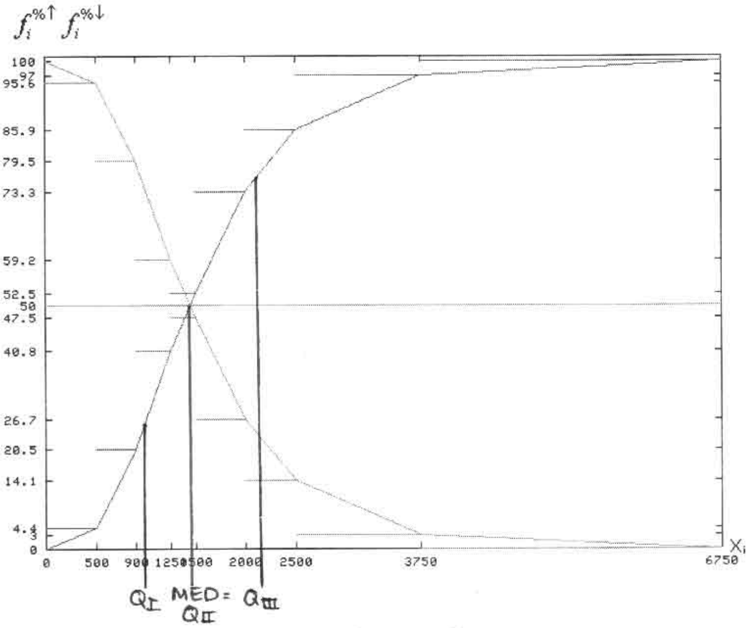
Aufgabe 13)

a) Grafische Ermittlung der Quartile:

alte Bundesländer:



neue Bundesländer:



Aus den Graphiken lassen sich folgende Werte ungefähr bestimmen:

Quartile			
	Q_I	Q_{II} = Med	Q_{III}
alte Bundesländer	1100	1700	2500
neue Bundesländer	1000	1450	2100

Daraus lassen sich die Semiquartilsabstände SAQ für die beiden Verteilungen berechnen:

SAQ (alte Bundesländer) $= \frac{(Q_{III} - Q_I)}{2} = \frac{2500 - 1100}{2} = 700$

SAQ (neue Bundesländer) $= \frac{(Q_{III} - Q_I)}{2} = \frac{2100 - 1000}{2} = 550$

Streuungsmaße			
	VAR	s	SQA
alte Bundesländer	1590868	1261	700
neue Bundesländer	982097	991	550

Die Varianz als Summe der quadrierten Abweichungen vom arithmetischen Mittel erlaubt keine direkte empirisch sinnvolle Interpretation. Mit Hilfe der Standardabweichung lässt sich die (durch die Varianz quantifizierte) Streuung aber immerhin in der Maßeinheit der Ausgangsdaten darstellen, auch wenn durch das „Wurzel ziehen“ die Quadrierung der Summanden nicht (korrekt) aufgehoben wird. Unter diesem Vorbehalt lässt sich beispielsweise über die Einkommensverteilung der alten Bundesländer sagen, dass die Einkommen um durchschnittlich 1261 € vom arithmetischen Mittel abweichen; der durchschnittliche Abstand zwischen dem Median und den Quartilen wiederum beläuft sich auf gut 700 €. Der direkte Vergleich beider Verteilungen wird durch den Variationskoeffizient ermöglicht (vgl. c))

c) relative Streuung und Schiefe der Verteilungen

Daten aus Aufgabe 8:

Modus, Median und arithmetisches Mittel der Verteilungen			
	Modus	Median	$\overline{X}$
früheres Bundesgebiet	ca. 1200	ca. 1700	2006,15
neue Bundesländer	ca. 1150	ca. 1450	1654,23

Der Variationskoeffizient gibt die relative Streuung an und erlaubt den direkten Vergleich der beiden Verteilungen.

alte Bundesländer $V = \frac{s}{\overline{X}} = \frac{1261}{2006{,}15} = 0{,}6103$

neue Bundesländer $V = \frac{s}{\overline{X}} = \frac{991}{1654{,}23} = 0{,}5991$

Die relative Streuung beträgt bei beiden Einkommensverteilungen rund 60%.

Die Schiefe der Verteilungen lässt sich über die ‚Fechnersche Lageregel' bestimmen, demnach gilt hier für beide Verteilungen:

$$\text{Mod} < \text{Med} < \overline{X}$$

Beide Verteilungen sind demnach rechtsschief (linkssteil).

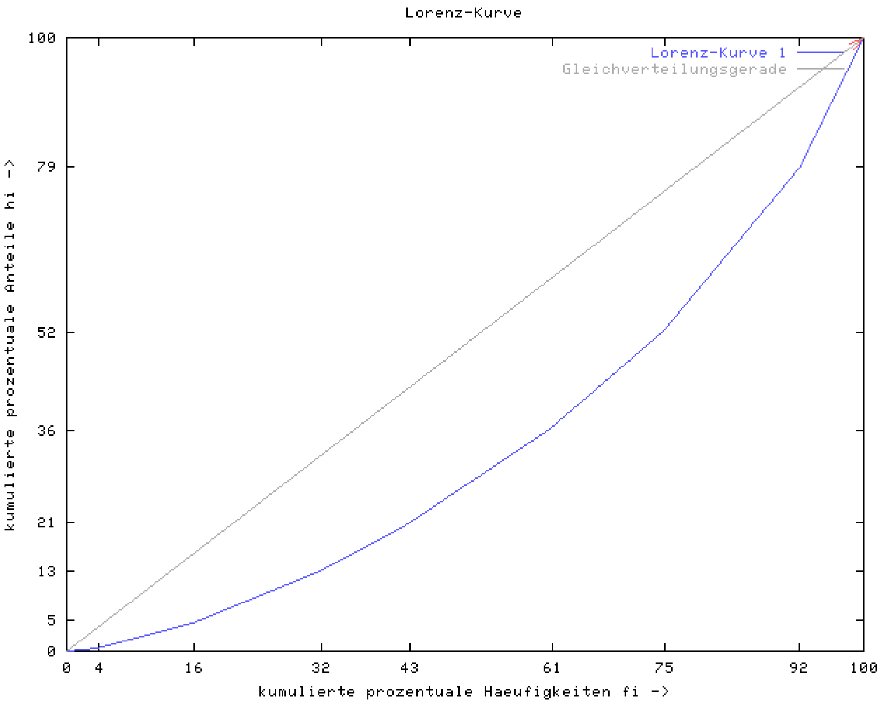
Aufgabe 14)

a) Entwicklung der Lorenzkurven und der Gini-Koeffizienten der relativen Konzentration auf Basis der Tabelle von Aufgabe 6.

alte Bundesländer:

Einkommen von... bis unter...EUR	m_i	f_i	$m_i \cdot f_i$	$h_i^{\%}$	$h_i^{\%\uparrow}$	$f_i^{\%}$	$f_i^{\%\uparrow}$	$\left(h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow}\right) \cdot f_i^{\%}$
0-500	250	1129	282250	0,4772	0,4772	3,8308	3,8308	1,8279
500-900	700	3529	2470300	4,1763	4,6535	11,9741	15,8048	61,4349
900-1250	1075	4802	5162150	8,7272	13,3807	16,2934	32,0983	293,8380
1250-1500	1375	3070	4221250	7,1365	20,5171	10,4167	42,5149	353,1019
1500-2000	1750	5321	9311750	15,7425	36,2596	18,0544	60,5694	1025,0721
2000-2500	2250	4207	9465750	16,0029	52,2625	14,2746	74,8439	1263,6155
2500-3750	3125	5029	15715625	26,5690	78,8315	17,0637	91,9076	2236,9426
3750-6750	5250	2385	12521250	21,1685	100	8,0924	100	1447,1806
$\sum$	-	29472	59150325	100	-	100	-	6683,0135

Graphische Darstellung der Lorenz-Kurve (alte Bundesländer):



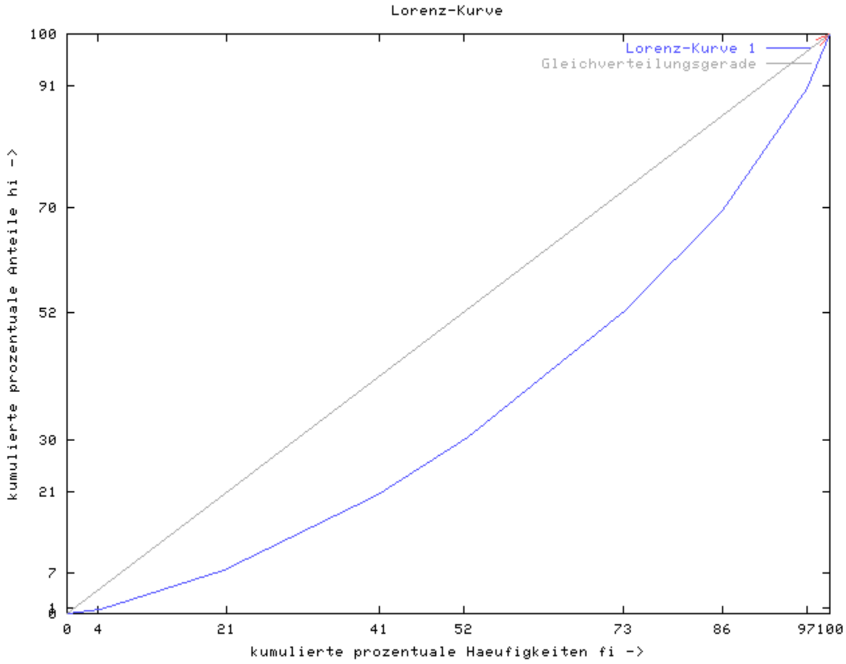
Als rechnerischen Wert für die Stärke der Konzentration erhalten wir:

$$G = 1 - \sum_{i=1}^{k} f_i^{'} \left(h_i^{'\uparrow} + h_{i-1}^{'\uparrow}\right) \;=\; 1 - \frac{\sum f_i^{\%} (h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow})}{10\,000} = 1 - \frac{6683{,}0135}{10\,000} = 1 - 0{,}6683 = 0{,}3317$$

neue Bundesländer:

Einkommen von... bis unter...EUR	m_i	f_i	$m_i \cdot f_i$	$h_i^{\%}$	$h_i^{\%\uparrow}$	$f_i^{\%}$	$f_i^{\%\uparrow}$	$\left(h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow}\right) \cdot f_i^{\%}$
0-500	250	305	76250	0,6670	0,6670	4,4133	4,4133	2,9435
500-900	700	1114	779800	6,8210	7,4879	16,1192	20,5325	131,4506
900-1250	1075	1398	1502850	13,1455	20,6335	20,2286	40,7611	568,8570
1250-1500	1375	810	1113750	9,7420	30,3755	11,7204	52,4816	597,8479
1500-2000	1750	1436	2513000	21,9814	52,3569	20,7785	73,2600	1719,0528
2000-2500	2250	873	1964250	17,1814	69,5383	12,6320	85,8921	1539,7849
2500-3750	3125	770	2406250	21,0476	90,5860	11,1417	97,0337	1784,0502
3750-6750	5250	205	1076250	9,4140	100	2,9663	100	565,3324
$\sum$	-	6911	11432400	100	-	100	-	6909,3193

Graphische Darstellung der Lorenz-Kurve (neue Bundesländer):



Als rechnerischen Wert für die Stärke der Konzentration erhalten wir:

$$G = 1 - \sum_{i=1}^{k} f_i^{'} \left(h_i^{'\uparrow} + h_{i-1}^{'\uparrow} \right) \; = \; 1 - \frac{\sum f_i^{\%} (h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow})}{10\,000} = 1 - \frac{6909{,}3193}{10\,000} = 1 - 0{,}6909 = 0{,}3091$$

Interpretation:
Beide Gini-Koeffizienten weisen eine eher schwache bis mittlere Konzentration der Merkmalssumme (Summe der monatlichen Einkommen) auf einzelne Merkmalsträger (Haushalte) aus. Im direkten Vergleich der Regionen finden sich kaum Unterschiede in der Konzentration. Von Gleichverteilung der Einkommen kann natürlich dennoch keine Rede sein. So verfügen beispielsweise in den alten Bundesländern rund 16 % der Haushalte mit dem geringstem Einkommen nur über einen Anteil von rund 5 % am Gesamteinkommen. Der Anteil (am Gesamteinkommen) der einkommensstärksten 8 % der Haushalte (in den alten Bundesländern) dagegen beläuft sich auf rund 21 %.

b) Entwicklung der Haushaltstruktur zwischen 1900 und 2000

Die Analyse der Entwicklung der Haushaltsstruktur (vgl. Aufgabe 4) mit den Methoden der absoluten Konzentration basiert

- auf der Anzahl (f_i) der Haushalte in den unterschiedlichen Haushaltsgrößen und
- der Anzahl (h_i) der Personen in den Haushalten unterschiedlicher Größe.

Die jeweiligen prozentualen Häufigkeiten für die Jahre 1900 und 2000 wurden aus den Tabellen der Musterlösungen zu den Aufgaben 4b und 4c (II) ermittelt und in die nachstehenden Arbeitstabellen eingesetzt

Arbeitstabelle 1900

Klasse	$f_i^{\%}$	$f_i^{\%\uparrow}$	$h_i^{\%}$	$h_i^{\%\uparrow}$	$\left(h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow}\right) \times f_i^{\%}$
1.	7.32	7.32	1.60	1.60	11.71
2.	14.63	21.95	6.38	7.98	140.16
3.	17.07	39.02	11.17	19.15	463.11
4.	17.07	56.09	14.89	34.04	907.95
5.	43.90	99.99	65.96	100.00	5884.36
Summe:	100	-	100	-	7407.29

Gini = 1 – 0.74 = 0.26

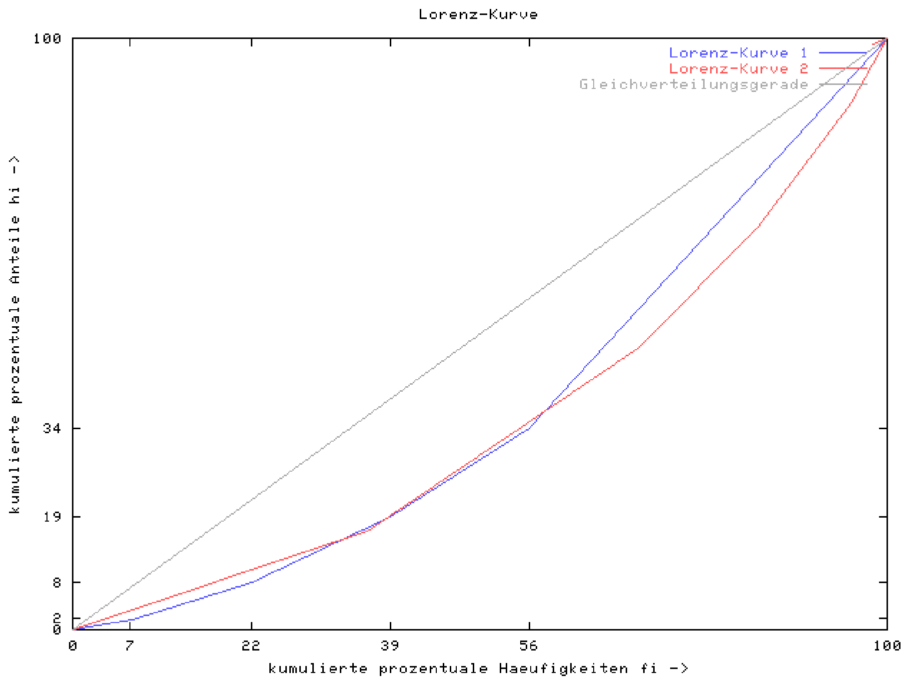
Der klassifikationsbedingte Sachverhalt, dass im Jahre 1900 66 % der Bevölkerung in einer Größenklasse leben, die 44 % der Haushalte umfasst (für 66% der Bevölkerung also eine Gleichverteilung gegeben ist), führt dazu, dass der Gini-Koeffizient für 1900 relativ klein ist und sogar unter dem des Jahres 2000 liegt (vgl. nächste Arbeitstabelle).
Aus den Spalten 3 und 5 der Arbeitstabellen übertragen wir die kumulierten Häufigkeiten in ein Koordinatensystem zur Erstellung der Lorenzkurve (vgl. nachstehende Abbildung). Die blaue Kurve gibt die Werte für 1900 wieder, die rote Lorenzkurve die des Jahres 2000. Dabei liegt die Kurve für 1900 zuerst unter dann über derjenigen des Jahres 2000.

Arbeitstabelle 2000

Klasse	$f_i^{\%}$	$f_i^{\%\uparrow}$	$h_i^{\%}$	$h_i^{\%\uparrow}$	$\left(h_i^{\%\uparrow} + h_{i-1}^{\%\uparrow}\right) \times f_i^{\%}$
1.	36.13	36.13	16.72	16.72	604.09
2.	33.25	69.38	30.79	47.51	2135.65
3.	14.66	84.04	20.37	67.88	1691.62
4.	11.52	95.56	21.33	89.21	1809.68
5.	4.45	100.01	10.78	99.99	841.94
Summe:	100	-	100	-	7082.98

Gini = 1 - 0.71 = 0.29

Abbildung: Konzentration der Personen auf die Haushalte (1900 und 2000)



Aufgabe 15: Absolute Konzentration

Zur Konzentrationsanalyse der Daten der nachstehenden Tabelle zu Aufgabe 15 werden die Ansätze der absolute Konzentration angewandt.

Unternehmen	$h_i^{\%}$	Rang
RWE Energie AG	26,1	1
Preussen Elektra AG	17,3	2
VEAG	10,3	3
Bayernwerk AG	7,7	4
VEW Energie AG	4,7	5
EnBW AG	4,1	6
HEW AG	3,2	7
BEWAG	2,4	9
NEW Neckerwerke Stuttgart	3,1	8
Isar-Amperwerke AG München	1,7	10
	80,5	

Als erstes, einfach aus der Originaltabelle zu bestimmendes Maß kann die **Konzentrationsrate** C_r für r = 10 bestimmt werden:

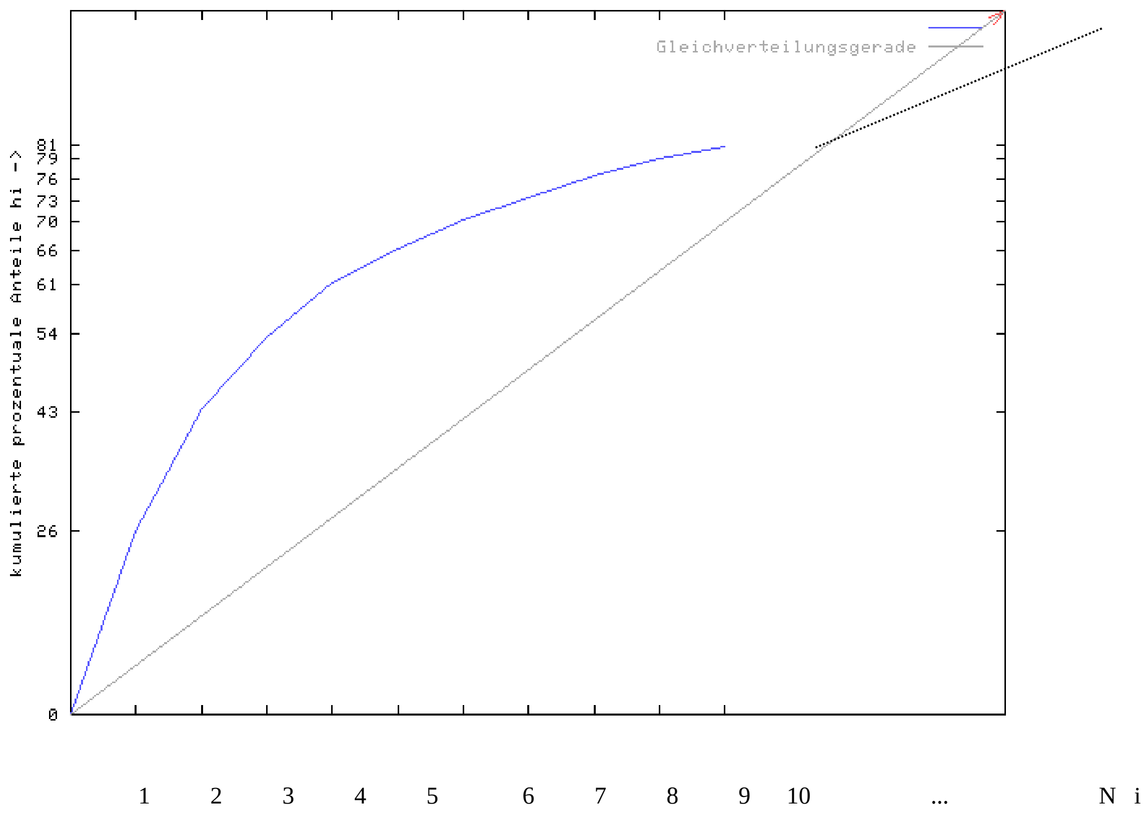
$$C_r = 0{,}805$$

Bei der Ermittlung der **Konzentrationskurve** und der **Indizes von Rosenbluth und Herfindahl** auf der Basis der gegebenen Informationen ist zu bedenken, dass keine Informationen über die Anzahl und die Anteile der Anbieter der restlichen 19,5 % des Marktaufkommens gegeben sind. Wenn das 10. Unternehmen einen Anteil von 1,7% hat müssen noch mindestens 12 weitere Anbieter mit jeweils 1,62% Anteil auf dem Markt vertreten sein. Auf die jeweiligen Beeinträchtigungen aufgrund dieser fehlenden Information wird im Analyseprozess konkret eingegangen.

Zwecks weiterer graphischer Darstellungen und Berechnungen müssen die Unternehmen gemäß ihrer Marktanteile sortiert und die relativen Anteile verwendet werden. Dabei wird die rechnerische Spaltensumme von 80,6 zugrundegelegt und die Tabelle um Arbeitspalten erweitert.

	$h_i^{\%}$	$h_i^{\%\vert}$	Rang = i	i * h_i	h_i^2
RWE Energie AG	26,1	26,1	1	0,26	0,068
Preussen Elektra AG	17,3	43,4	2	0,35	0,030
VEAG	10,3	53,7	3	0,31	0,011
Bayernwerk AG	7,7	61,4	4	0,31	0,006
VEW Energie AG	4,7	66,1	5	0,24	0,002
EnBW AG	4,1	70,2	6	0,25	0,002
HEW AG	3,2	73,4	7	0,22	0,001
NEW Neckerwerke Stuttgart	3,1	76,5	8	0,25	0,001
BEWAG	2,4	78,9	9	0,22	0,001
Isar-Amperwerke AG München	1,7	80,6	10	0,17	0,000
	80,6			2,56	0,121

Konzentrationskurve



Anmerkung: Die automatisch erstellte Graphik weist im Bereich $10 \leq i \geq N$ eine zu starke Steigung auf. N muss aus den erwähnten Gründen mindestens so groß sein, dass die Steigung gleich oder kleiner als im letzten Kurvenabschnitt ist.

Berechnung der Konzentrationsmaße

Rosenbluth - Index

$$C_R = \frac{1}{2 \cdot \sum i \cdot h_i - 1} = \frac{1}{2 * 2{,}56 - 1} = 0{,}24$$

wobei gilt: $\frac{1}{N} \leq C_R \leq 1$

Herfindahl - Index

$$C_H = \sum h_i^2 = 0{,}121$$

wobei gilt: $\frac{1}{N} \leq C_H \leq 1$

Interpretation:
Beide Konzentrationsmaße, der Rosenbluth-Index und der Herfindahl-Index, weisen auf eine schwache bis sehr schwache Konzentration hin. Es kann gesagt werden, dass der Markt unter den 10 bekannten Unternehmen relativ gleichmäßig aufgeteilt ist.
Anmerkung: Würden die Indizes auf der Basis aller Anbieter errechnet, würde sich der Rosenbluth-Index deutlich verringern $(\sum i\, h_i \uparrow)$ und der Herfindahl-Index würde steigen $(\sum h_i^2 \uparrow)$. Nimmt man an, dass die restlichen 19,4% des Angebots von 12 Unternehmen mit je 1,62% Anteil erbracht werden, sinkt der Rosenbluth-Index auf 0,094; der Herfindahl-Index hingegen steigt auf 0,127. Das zeigt, dass der Herfindahl-Index wesentlich geringer davon tangiert wird, dass nicht alle kleinen Anbieter in die Rechnung mit einbezogen wurden. Bei fehlender Markttransparenz im Bereich der kleinen Anbieter ist er also vorzuziehen.