

## Lösungsnotizen

## Übungsblatt Ferien

Statistik 1 für Studierende der Soziologie, des Nebenfachs Statistik, der Medieninformatik und der Cultural and Cognitive Linguistics  
(Augustin, Berger, Brandt, Endres, Plaß)

WiSe 2014/15

### Unverbindliche Lösungsnotizen!

Lösungsnotizen Aufgabe 1 ../../Aufgabensammlung/Loesungen/L\_Z\_3.tex

Variable	Skala	Modus	Median	arithmetisches Mittel	empirische Varianz ( $\hat{s}^2$ )	Varianz zwischen ( $\hat{s}_{zw}^2$ )	Varianz innerhalb ( $\hat{s}_{in}^2$ )	Anteil erklärter Varianz	Standardabweichung
Geschlecht	nominal	Frau	-	-	-	-	-	-	-
Ostwest	nominal	West	-	-	-	-	-	-	-
Fernsehen	Verhältnis	240	240	210	6480	937.5	5542.5	0.14	80.5
Alter	Verhältnis	alle	51.5	49.4	380.44	162.24	218.2	0.43	19.5
Einkommen	Verhältnis	1000 Euro	930	846.3	172375.21	2068.33	170306.88	0.01	415.18
Gewicht	Verhältnis	55 kg	73.5	74.2	228.56	141.14	87.42	0.62	15.12
Groesse	Verhältnis	164 cm	169	167.3	46.21	10.4	35.81	0.23	6.8
BMI	Verhältnis	alle	26.3	26.67	34.62	8.54	26.08	0.25	5.88

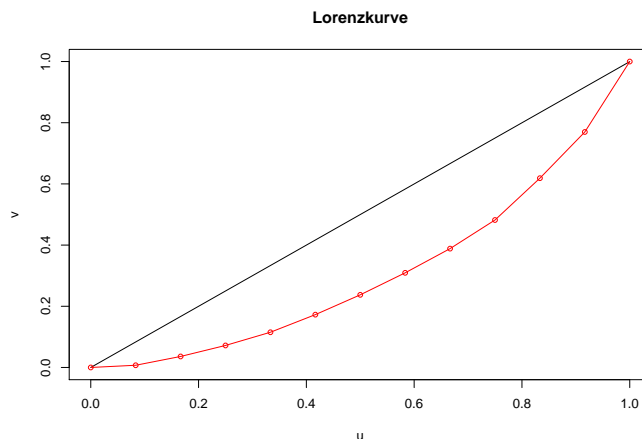
**Lösungsnotizen Aufgabe 2** ../../Aufgabensammlung/Loesungen/L\_Z\_2\_kurz.tex

- a) Ja.
- b) Nein.
- c) Ja.
- d) Ja.
- e) Nein.
- f) Nein.
- g) Ja.
- h) Ja.
- i) Ja
- j) Ja
- k) Nein.
- l) Ja.
- m) Ja.
- n) Nein.
- o) Nein.
- p) Nein.
- q) Ja.
- r) Ja.
- s) Ja.
- t) Nein
- u) Ja.
- v) Nein.
- w) Ja.
- x) Nein.
- y) Nein.
- z) Gilt im Allgemeinen nicht.

**Lösungsnotizen Aufgabe 3** ../../Aufgabensammlung/Loesungen/L\_Z\_4.tex

a) Urliste

(i) Lorenzkurve:



(ii) Gini-Koeffizient

$$G = \frac{2 \cdot (1 \cdot 1 + 2 \cdot 4 + \dots + 12 \cdot 32)}{12 \cdot 139} - \frac{13}{12} = 0.38$$

$$G^* = \frac{12}{11} \cdot 0.38 = 0.42$$

(iii) Interpretation:

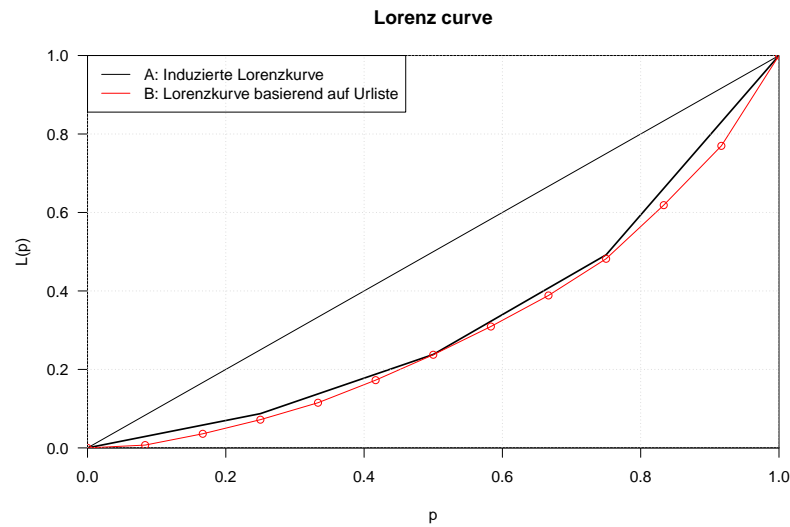
Die Lorenzkurve zeigt beispielsweise, dass die 50% ärmsten Marskönige 23% der grünen Taler besitzen. Mit Hilfe des Gini-Koeffizienten (kann Werte zwischen 0 und  $\frac{n-1}{n} = \frac{11}{12} \approx 0.92$ ) sowie des normierten Gini-Koeffizienten (kann Werte zwischen 0 und 1 annehmen) lässt sich auf eine mittlere Konzentration schließen.

b) Quantilsbasierte Daten

Quartil	1	2	3	4
Quantil	$\alpha_{0.25}$	$\alpha_{0.5}$	$\alpha_{0.75}$	$\alpha_1$
Wert	5.5	9.5	16.0	32.0

(i) Induzierte Lorenzkurve

$$\begin{aligned}
 u_l^* = \alpha_l &\Rightarrow u_1^* = 0.25, & u_2^* = 0.5, & u_3^* = 0.75, & u_4^* = 1 \\
 v_l^* = \sum_{r \leq l} z_r^* &\Rightarrow v_1^* = 0.087, & v_2^* = 0.238, & v_3^* = 0.492, & v_4^* = 1.000
 \end{aligned}$$



(ii) Induzierter Gini-Koeffizient:

$$G^* = 1 - (0.25 \cdot (0 + 0.087) + 0.25 \cdot (0.087 + 0.238) + 0.25 \cdot (0.238 + 0.492) + 0.25 \cdot (0.492 + 1)) = 0.34$$

(iii) Vergleich der Ergebnisse:

Die induzierte Lorenzkurve liegt immer oberhalb der Lorenzkurve basierend auf der Urliste.

c) Herfindahl-Index, Konzentrationsrate:

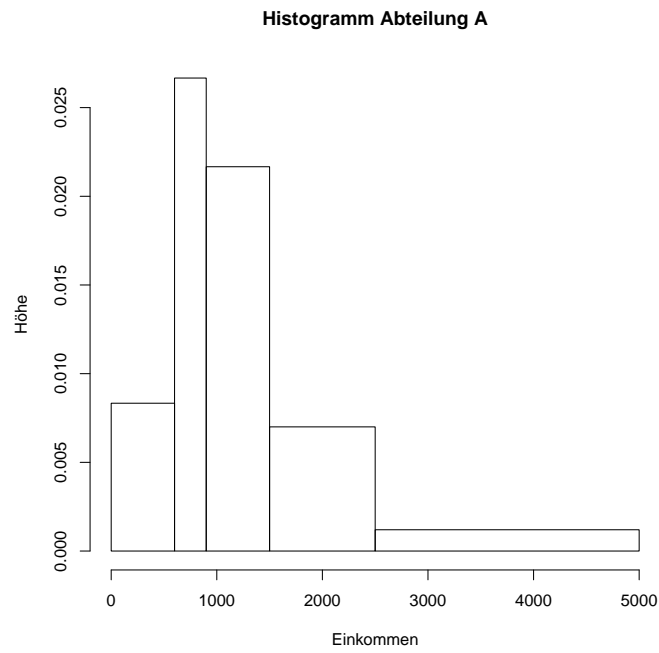
$$H = \left(\frac{1}{139}\right)^2 + \left(\frac{4}{139}\right)^2 + \dots + \left(\frac{32}{139}\right)^2 = 0.13$$

$$CR_3 = \frac{19}{139} + \frac{21}{139} + \frac{32}{139} = 0.52$$

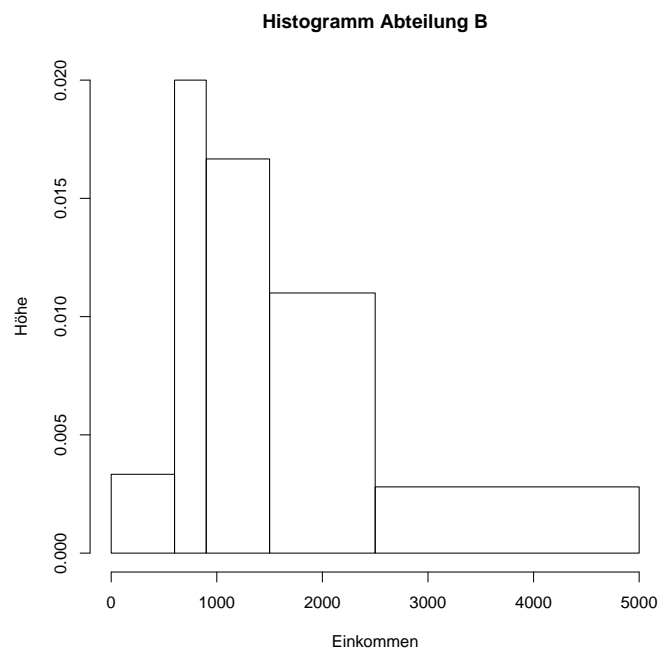
Der Herfindahl-Index sowie die Konzentrationsrate stellen absolute Konzentrationsmaße dar, wohingegen der Gini-Koeffizient ein relatives Konzentrationsmaß ist.

**Lösungsnotizen Aufgabe 4** ../../Aufgabensammlung/Loesungen/L\_Z\_6.tex

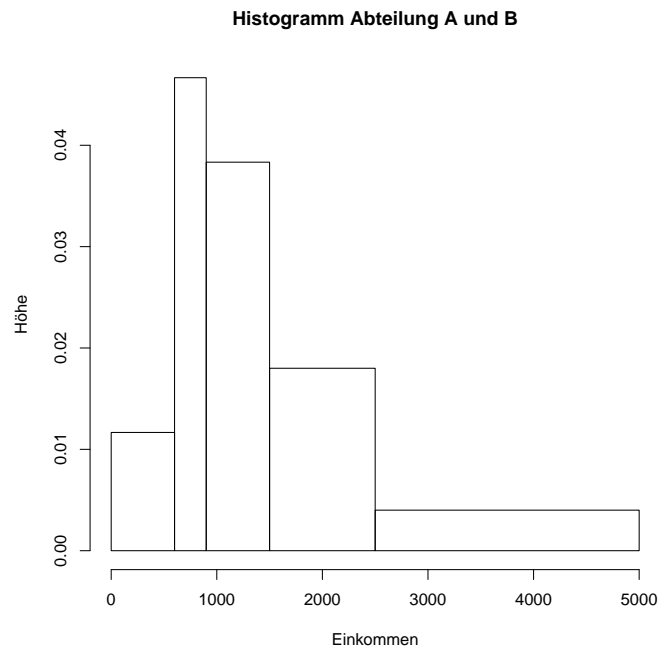
a) Histogramm Abteilung A:



Histogramm für Abteilung B:



Histogramm für Abteilung A und B:



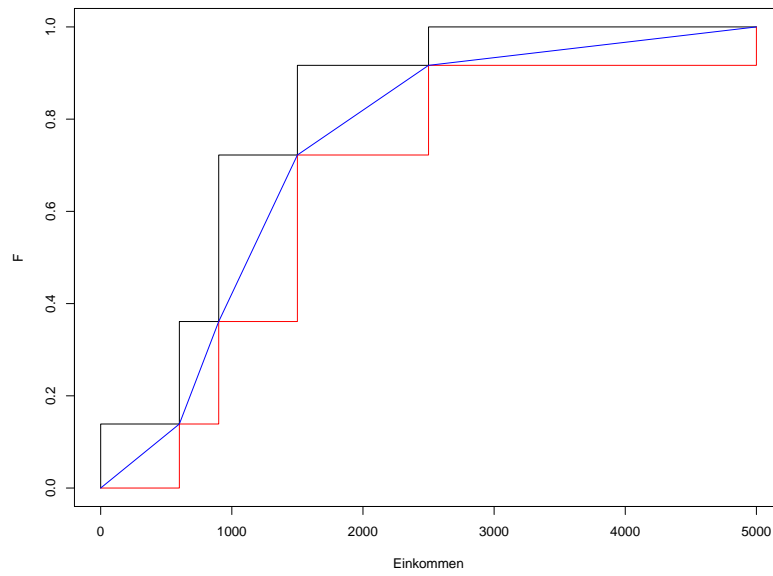
Interpretation:

In beiden Abteilungen typische Einkommensverteilung: hauptsächlich mittlere Einkommen, wenig hohe Einkommen, linkssteil  $\hat{=}$  rechtsschief. In Abteilung A niedrigere Einkommen im Vergleich zu Abteilung B.

b) Berechnung der durchschnittlichen Einkommen

- Abteilung A:  $\bar{x}_{grupp,A} = 1343.06$
- Abteilung B:  $\bar{x}_{grupp,B} = 1815.28$
- Beide Abteilungen zusammen:  $\bar{x}_{grupp,zs} = 1579.17$

c) Verteilungsfunktion

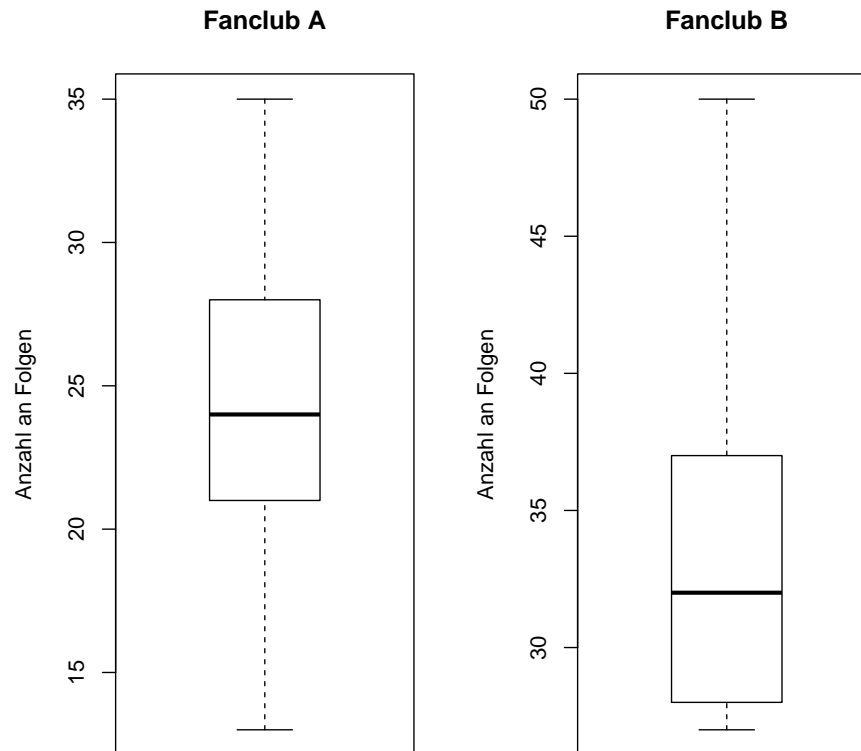


d) Bestimmung von Median und Interquartilsabstand anhand der Verteilungsfunktion

- Median: Wert der x-Achse mit  $F(x) = 0.5$
- Interquartilsabstand: Bestimmung des 25%-Quantils sowie des 75%-Quantils (x-Werte an der Stelle  $F(x) = 0.25$  und  $F(x) = 0.75$ ) und Berechnung der Differenz dieser Werte ( $x_{0.75} - x_{0.25}$ ).

**Lösungsnotizen Aufgabe 5** ../../Aufgabensammlung/Loesungen/L\_Z\_5.tex

- Zeichnen der Boxplots für Fanclub A und Fanclub B:



- Interpretation:

**Fanclub A:** Der Median beträgt 24. Die 50% mittleren Daten liegen zwischen 21 und 28. Es gibt keine Ausreißer.

**Fanclub B:** Der Median beträgt 32. Die 50% mittleren Daten liegen zwischen 28 und 37. Es gibt keine Ausreißer.

Insgesamt haben die Mitglieder aus Fanclub A weniger Folgen gesehen als die aus Fanclub B. Der Anteil der Gesamtvarianz der gesehenen Folgen, der durch die Schichtung in die beiden Fanclubs erklärt wird, ist vermutlich als einigermaßen groß einzustufen.