

Zusatzaufgaben zum Selbststudium

Aufgabe 1

Eine Familie mit n Schulkindern vergibt ein kleines Taschengeld an diese Schul Kinder entsprechend der Schulklasse, d.h. ein Kind bekommt einen Euro pro Woche, wenn es in der 1. Klasse ist, zwei Euro, wenn es in der 2. Klasse ist, usw. Betrachten Sie die folgenden Situationen:

- A** alle Kinder sind in der gleichen Klassenstufe s .
- B** Jede Klassenstufe von 1 bis n wird genau von einem Kind besucht, d.h. es kommen die Taschengeldhöhen $1, 2, 3, \dots$ bis n Euro jeweils genau einmal vor (man sagt, dass Taschengeld ist im statistischen Sinne gleichverteilt auf den Ausprägungen $\{1, 2, \dots, n\}$).
- a) Skizzieren Sie für beide Situationen die empirische Verteilungsfunktion des wöchentlichen Taschengeldes.
- b) In welcher Situation ist das Taschengeld gerechter verteilt? In welchem Fall ist die Konzentration geringer?
- c) Überlegen Sie sich, wie in den beiden Fällen die Ausdrücke $u_j = \sum_{l=1}^j f_l$ und
- $$v_j = \frac{\sum_{l=1}^j f_l \cdot a_l}{\sum_{l=1}^k f_l \cdot a_l}$$
- aussehen.
- d) Skizzieren Sie für beide Situationen die Lorenzkurven.

Aufgabe 2

Sind die folgenden Aussagen richtig?

- a) Die absolute kumulierte Häufigkeitsverteilung beschreibt die Verteilung der Daten vollständig.
- b) Der Mittelwert und die Varianz einer Verteilung beschreiben diese vollständig.
- c) Aus der empirischen Verteilungsfunktion ist der Mittelwert ableitbar.
- d) Aus der Lorenzkurve ist der Gini-Koeffizient ableitbar.
- e) Aus der Lorenzkurve ist die empirische Verteilungsfunktion ableitbar.
- f) Der Interquartilsabstand ist ein Lagemaß.
- g) Für die Varianzzerlegung gilt, dass die Gesamtvarianz immer größer oder gleich der Varianz innerhalb der Schichten ist.
- h) Für die Varianzzerlegung gilt, dass die Gesamtvarianz immer größer oder gleich der Varianz zwischen den Schichten ist.
- i) Aus dem Boxplot ist das 0%-Quantil ablesbar.
- j) Aus der empirischen Verteilungsfunktion ist die Varianz ableitbar.
- k) Aus der empirischen Verteilungsfunktion kann man die Stichprobengröße n ermitteln.
- l) Aus der empirischen Verteilungsfunktion kann man eine untere Schranke für die Stichprobengrößen ableiten.
- m) Der Median eines verhältnisskalierten Merkmals ist sinnvoll interpretierbar.
- n) Der Mittelwert eines nur ordinalskalierten Merkmals ist sinnvoll interpretierbar.
- o) Der Variationskoeffizient eines lediglich intervallskalierten Merkmals ist sinnvoll interpretierbar.
- p) $\left(\sum_{i=1}^n a_i\right) \cdot \left(\sum_{i=1}^n b_i\right) = \sum_{i=1}^n a_i \cdot b_i$.
- q) $\left(\sum_{i=1}^m a_i\right) \cdot \left(\sum_{j=1}^n b_j\right) = \sum_{i,j \in \{1, \dots, n\}} a_i \cdot b_j$.
- r) $\sum_{i=1}^n c \cdot a_i = c \cdot \sum_{i=1}^n a_i$.
- s) $a \leq b \implies a + c \leq b + c$.

- t) $a \leq b \implies a \cdot c \leq b \cdot c$.
- u) Der Median der aus der Größe X abgeleiteten Größe $a \cdot X + b$ hat (, falls er eindeutig bestimmt ist,) den Wert $a \cdot \text{Median}(X) + b$.
- v) Das 75% - Quantil der aus der Größe X abgeleiteten Größe $Y := a \cdot X + b$ hat (, falls es eindeutig bestimmt ist,) den Wert $a \cdot x_{0.75} + b$.
- w) Das 75% - Quantil der aus der Größe X abgeleiteten Größe $Y := a \cdot X + b$ hat (, falls es eindeutig bestimmt ist,) für negatives a den Wert $a \cdot x_{0.25} + b$.
- x) Das Histogramm ist längentreu.
- y) Die Summe aller kumulierten relativen Häufigkeiten ist 1.
- z) Zwei verschiedene Lorenzkurven führen zu verschiedenen Gini-Koeffizienten.

Aufgabe 3

Hier sehen Sie zehn zufällig ausgewählte Beobachtungen aus dem ALLBUS 2008, bereitgestellt von GESIS (Leibniz-Institut für Sozialwissenschaften). In diesem Datensatz ist das Geschlecht des Befragten (Geschlecht), der Wohnort in Ost- oder Westdeutschland (Ostwest), der Fernsehkonsum in Minuten (Fernsehen), das Alter in Jahren (Alter), das Einkommen in Euro (Eink.), das Gewicht in kg (Gewicht), die Größe des Befragten in cm (Grosesse) und der Body-Mass-Index (BMI) erfasst. Als Wiederholung können Sie die Tabelle auf der nächsten Seite ausfüllen. Verwenden Sie für die Berechnung von \tilde{s}^2 , \tilde{s}_{zw}^2 und \tilde{s}_{in}^2 Geschlecht als Schichtungsvariable. Berechnen Sie nur diejenigen Größen, die sinnvoll sind.

Geschlecht	Ostwest	Fernsehen	Alter	Eink.	Gewicht	Grosesse	BMI
Frau	West	60	43	860	67	164	24.9
Frau	West	180	67	1500	65	164	24.2
Frau	West	240	20	250	55	170	19.0
Frau	Ost	360	51	1200	85	151	37.3
Mann	Ost	240	56	1300	83	171	28.4
Mann	West	240	82	1000	92	168	32.6
Frau	West	150	27	345	60	164	22.3
Frau	West	120	26	700	55	175	18.0
Mann	Ost	270	52	308	100	176	32.3
Mann	West	240	70	1000	80	170	27.7

