Statistik 1 für WInf, WI(MB), WI(ET), WI(BI)
Übung 1, Lösungsvorschlag

Gruppenübung

G 1 Beschreibende Statistik

Bei einer Klausur konnten maximal 50 Punkte erreicht werden. In der folgenden Tabelle ist zu jeder Note die Punktzahl p angegeben, die zum Erhalt dieser Note mindestens erreicht werden musste.

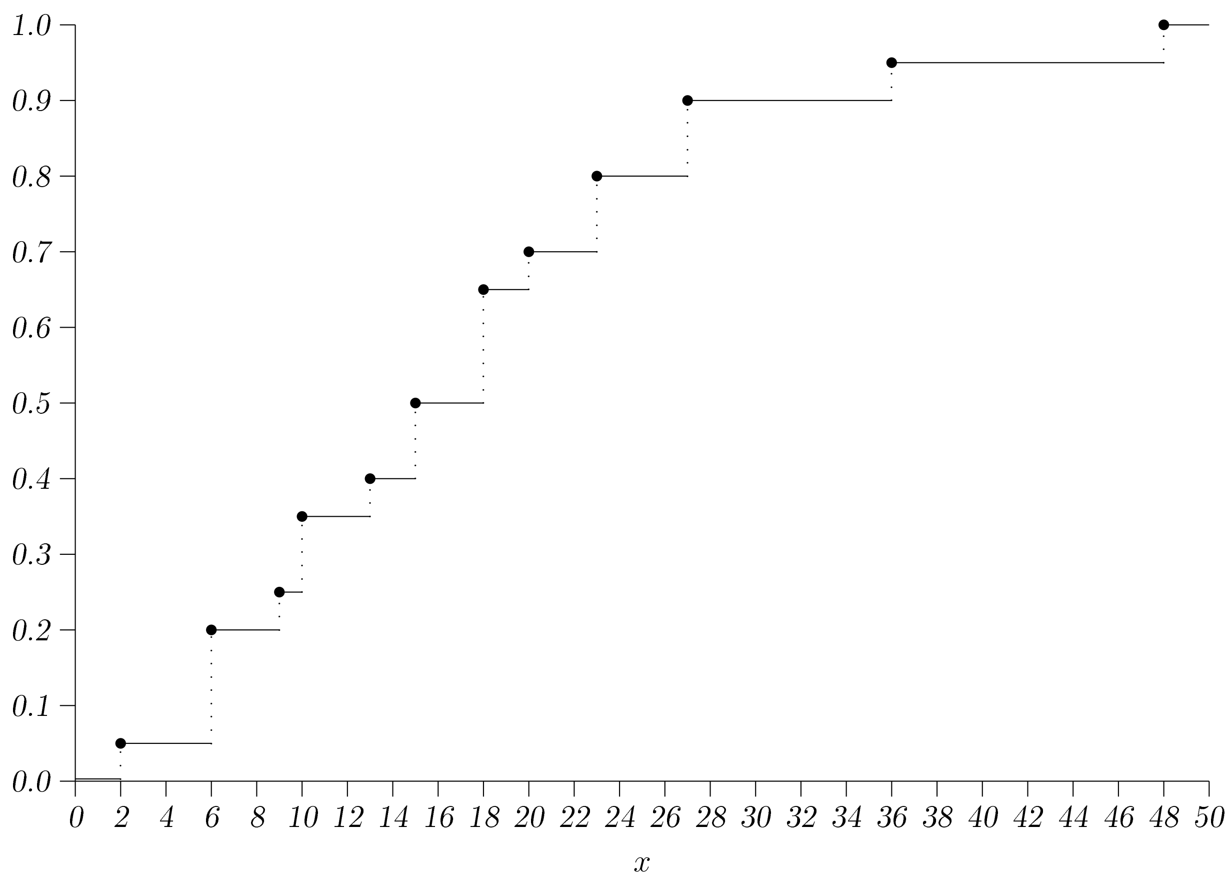
Note	1	2	3	4	5
p	35	20	15	10	0

Folgende Punktzahlen $p_1, p_2, \ldots, p_{20}$ wurden von den 20 Teilnehmern erzielt:

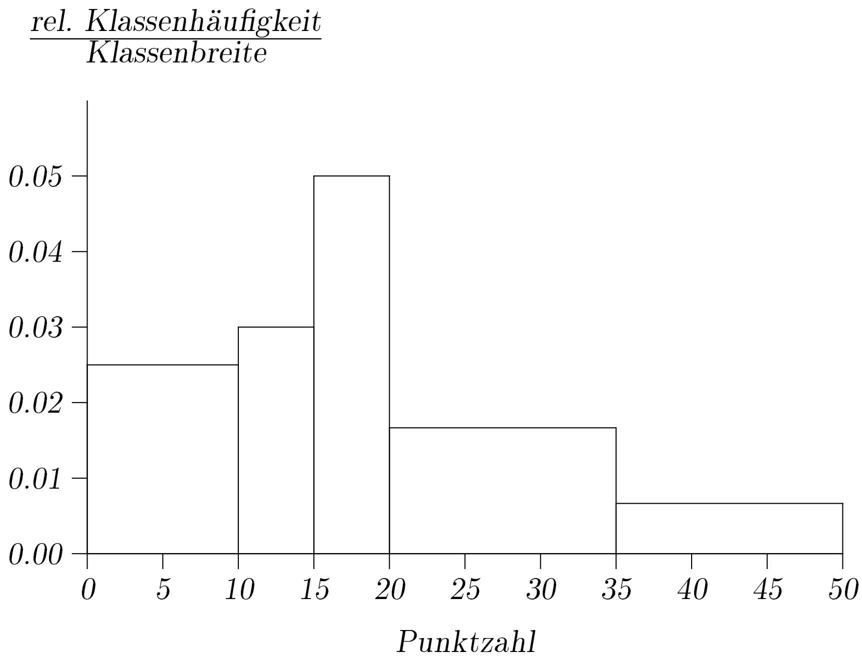
2	6	6	6	9	10	10	13	15	15
18	18	18	20	23	23	27	27	36	48

a) Skizzieren Sie die empirische Verteilungsfunktion der gegebenen Messreihe $p_1, p_2, \ldots, p_{20}$.

Empirische Verteilungsfunktion $F_{(20)}(x)$:



b) Erstellen Sie zu den Daten ein Stabdiagramm. Verwenden Sie dabei die Klasseneinteilung $[0, 10)$, $[10, 15)$, $[15, 20)$, $[20, 35)$, $[35, 50]$ (also entsprechend der Notengebung).



c) Bestimmen Sie die folgenden Charakteristika der Messreihe:

arithmetisches Mittel, Modalwert und Median,
empirische Varianz und empirische Standardabweichung,
0.33-Quantil und Quartilabstand.

Lösung:

$$\bar{p} = \frac{1}{20}\sum_{i=1}^{20} p_i = \frac{350}{20} = 17.5 \qquad \textit{(arithmetisches Mittel)}$$

$$\tilde{p} = p_{(10)} = \frac{15+18}{2} = 16.5 \qquad \textit{(Median)}$$

$$p_{0.33} = p_{(7)} = 10 \qquad \textit{(0.33-Quantil)}$$

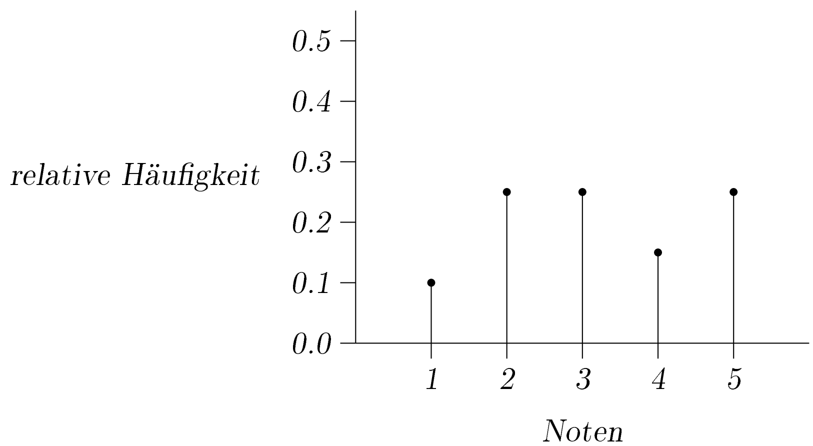
$$q = p_{(15)} - p_{(5)} = 23 - 9 = 14 \qquad \textit{(Quartilabstand)}$$

$$S^2 = \frac{1}{20}\left(\sum_{i=1}^{20} p_i^2 - 20\bar{p}^2\right) = 118,75 \qquad \textit{(empirische Varianz)}$$

$$S = \sqrt{S^2} = \sqrt{118,75} \approx 10.90 \qquad \textit{(empirische Standardabweichung)}$$

Es gibt zwei Modalwerte, nämlich 6 und 18. Der Quartilabstand ist nicht eindeutig, da alle Werte aus dem Intervall $[9, 10]$ 0.25-Quantile sind.

d) Stellen Sie die relative Häufigkeit der einzelnen Noten in einem Stabdiagramm dar.



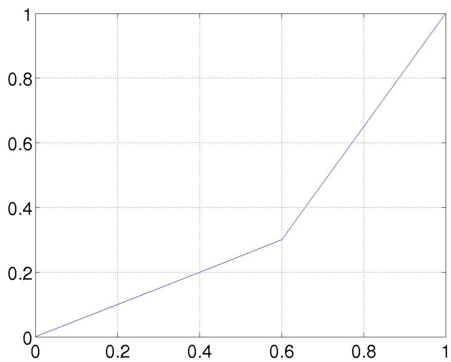
G 2 Konzentrationsmaße

Fünf Hersteller bestimmter Großgeräte lassen sich hinsichtlich ihrer Marktanteile in zwei Gruppen aufteilen: Drei Hersteller besitzen jeweils gleiche Marktanteile von 10 Prozent, der Rest des Marktes teilt sich unter den verbleibenden Herstellern gleichmäßig auf. Zeichnen Sie die zugehörige Lorenzkurve, und berechnen Sie den Gini-Koeffizienten.

Es gibt drei Hersteller mit 10 % Marktanteil, zwei weitere mit einem Marktanteil von 35 %. Für die Lorenzkurve gehen wir also von der geordneten Stichprobe 10,10,10,35,35 aus. Die Stichprobenlänge ist $n = 5$. Daraus erstellt man die folgende Tabelle:

k	$\frac{k}{n}$	$\frac{x_k}{\sum_{i=1}^n x_i}$	$v_k = \frac{\sum_{i=1}^k x_i}{\sum_{i=1}^n x_i}$
1	0.2	0.1	0.1
2	0.4	0.1	0.2
3	0.6	0.1	0.3
4	0.8	0.35	0.65
5	1.0	0.35	1.0

Die Punkte $\left(\frac{k}{n}, v_k\right)$ liefern die Lorenzkurve:



Als Gini-Koeffizient erhält man

$$\begin{aligned} G &= \frac{2\sum_{i=1}^n i x_i}{n\sum_{i=1}^n x_i} - \frac{n+1}{n} = \frac{2}{n}\sum_{i=1}^n i \frac{x_i}{\sum_{j=1}^n x_j} - \frac{n+1}{n} \\ &= \frac{2}{5}(1\cdot 0.1 + 2\cdot 0.1 + 3\cdot 0.1 + 4\cdot 0.35 + 5\cdot 0.35) - \frac{6}{5} \\ &= 0.3 \end{aligned}$$

Hausübung

H 1 In 40 Haushalten wurde die Anzahl der vorhandenen Elektrogeräte ermittelt. Es ergaben sich die folgenden Werte:

11	20	7	14	13	12	13	9	15	8
12	18	6	10	19	5	8	11	14	11
8	15	12	17	12	6	7	12	9	6
14	12	11	15	14	9	7	15	11	13

a) Bestimmen Sie den Median zu dieser Messreihe.

b) Veranschaulichen Sie die Messreihe mit einem Stabdiagramm.

c) Erstellen Sie die zugehörige empirische Verteilungsfunktion und lesen Sie daraus die empirischen Quantile $x_{0.1}$ und $x_{0.7}$ ab.

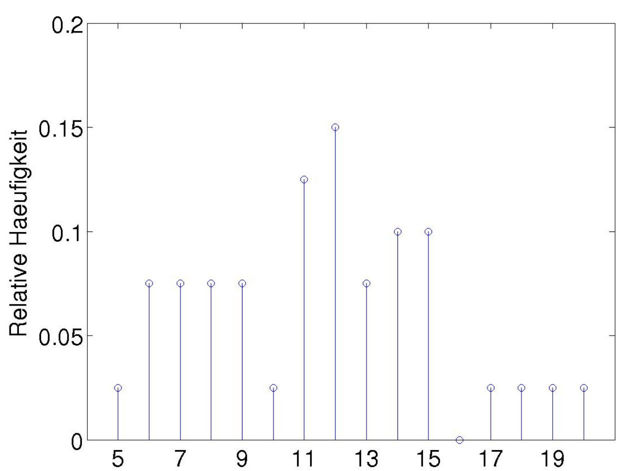
d) Erstellen Sie einen Boxplot.

e) Ermitten Sie die empirische Varianz, die Standardabweichung und die Stichprobenvarianz.

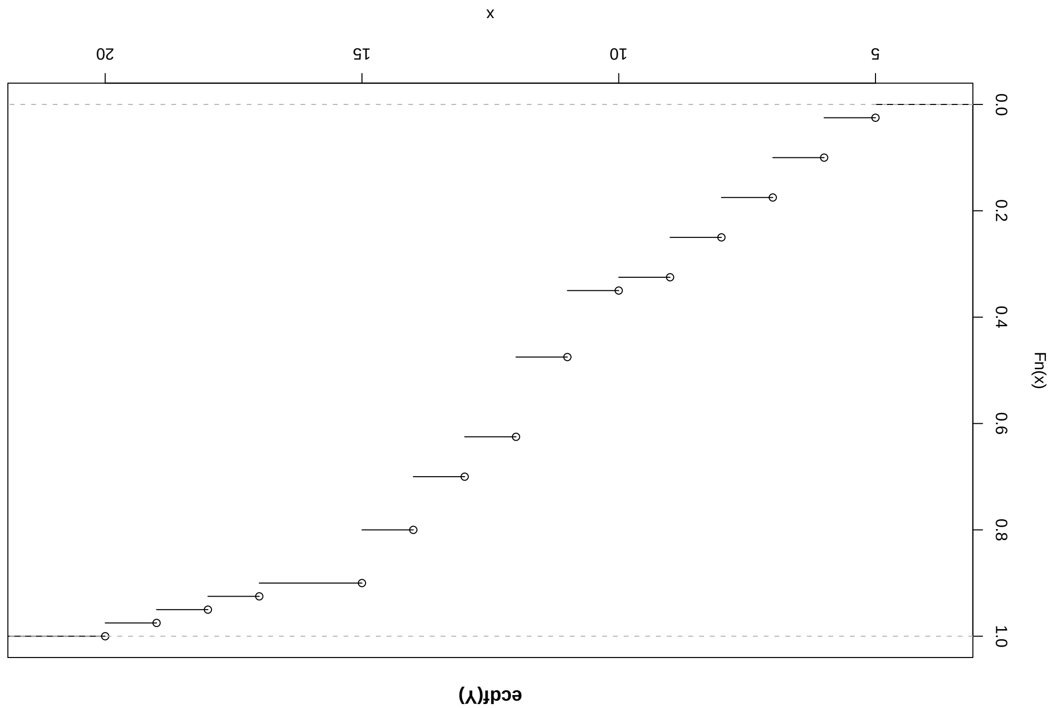
Die geordnete Messreihe ist:

5	*6*	*6*	*6*	*7*	*7*	*7*	*8*	*8*	*8*
9	*9*	*9*	*10*	*11*	*11*	*11*	*11*	*11*	*12*
12	*12*	*12*	*12*	*12*	*13*	*13*	*13*	*14*	*14*
14	*14*	*15*	*15*	*15*	*15*	*17*	*18*	*19*	*20*

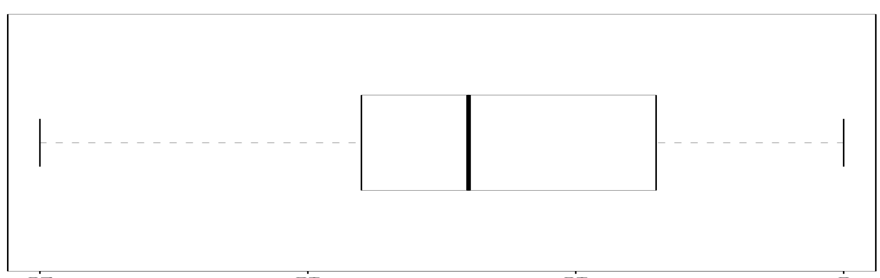
Stabdiagramm:



Empirische Verteilungsfunktion:



Boxplot:



Nach Ordnen der Stichprobe erhält man den Median $x_{Med} = \frac{x_{(20)}+x_{(21)}}{2} = 12$.
Jeder Wert zwischen zwischen $x_{(4)} = 6$ *und* $x_{(5)} = 7$ *ist ein 0.1-Quantil,* $x_{0.1} \in [6, 7]$, *und jeder Wert zwischen* $x_{(28)} = 13$ *und* $x_{(29)} = 14$ *ist ein 0.7-Quantil,* $x_{0.7} \in [13, 14]$.
Für den Boxplot benötigt man noch ein 0.25- und ein 0.75-Quantil, z.B. $x_{0.25} = 8$, $x_{0.75} = 14$, *und das arithmetische Mittel* $\bar{x} = \frac{461}{40} = 11.525$.

Weiterhin ist die empirische Varianz (=mittlere quadrat. Abweichung) gleich 13.399, *die Standardabweichung ist gleich* 3.661, *und als Stichprobenvarianz erhält man* 13.743.

H 2 Standardisierung einer Messreihe

Bestimmen Sie diejenige lineare Transformation

$$y_i = a x_i + b, \qquad a > 0,$$

durch die sich aus einer Messreihe $x_1, \ldots, x_n$ mit $s_x^2 > 0$ eine transformierte Messreihe $y_1, \ldots, y_n$ ergibt, für die

$$\overline{y} = 0 \qquad \text{und} \qquad s_y^2 = 1$$

gelten.

Für eine lineare Transformation $y_i = ax_i + b$, $i = 1, \ldots, n$, gelten $\bar{y} = a\bar{x} + b$ und $s_y^2 = a^2 s_x^2$.
Aus $\bar{y} = 0$ und $s_y^2 = 1$ folgen also

$$a^2 = \frac{1}{s_x^2}\,, \qquad b = -\frac{\bar{x}}{s_x}\,,$$

und somit

$$y_i \;=\; \frac{x_i - \bar{x}}{s_x}\,.$$

H 3 In einer Straße wohnen 20 Familien mit Kindern: 2 Familien mit 4 Kindern, 4 Familien mit 3 Kindern, 6 Familien mit 2 Kindern und 8 Familien mit einem Kind.

a) Berechnen Sie die Anteile x_1, x_2, x_3 bzw. x_4 der Kinder, die in Einkind-, Zweikind-, Dreikind- bzw. Vierkind-Familien wohnen.

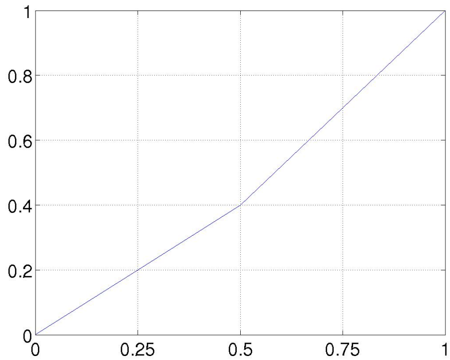
b) Stellen Sie die Konzentration der Messreihe x_i in einer Lorenzkurve (mit Achsenbeschriftung) dar. Berechnen Sie den Gini-Koeffizienten.

c) Ist es möglich, die Kinder so auf die 20 Familien zu verteilen, dass der Gini-Koeffizient gleich Null wird?

In der Straße wohnen insgesamt 40 Kinder. Daraus ermitteln wir $x_1 = \frac{8}{40} = 0.2$, $x_2 = \frac{6\cdot 2}{40} = 0.3$, $x_3 = \frac{4\cdot 3}{40} = 0.3$ und $x_4 = \frac{2\cdot 4}{40} = 0.2$. Daraus erstellt man die folgende Tabelle (für die geordnete Stichprobe $x_{(i)}$, $i = 1, \ldots, 4$):

k	$\frac{k}{n}$	$\frac{x_{(k)}}{\sum_{i=1}^n x_{(i)}}$	$v_k = \frac{\sum_{i=1}^k x_{(i)}}{\sum_{i=1}^n x_{(i)}}$
1	*0.25*	*0.2*	*0.2*
2	*0.5*	*0.2*	*0.4*
3	*0.75*	*0.3*	*0.7*
4	*1.0*	*0.3*	*1.0*

Die Punkte $\left(\frac{k}{n}, v_k\right)$ liefern die Lorenzkurve:



Als Gini-Koeffizient erhält man

$$\begin{aligned} G &= \frac{2\sum_{i=1}^n i x_i}{n\sum_{i=1}^n x_i} - \frac{n+1}{n} = \frac{2}{n}\sum_{i=1}^n i\frac{x_i}{\sum_{j=1}^n x_j} - \frac{n+1}{n} \\ &= \frac{2}{4}(1\cdot 0.2 + 2\cdot 0.2 + 3\cdot 0.3 + 4\cdot 0.3) - \frac{5}{4} \\ &= 0.1 \end{aligned}$$

Soll der Gini-Koeffizient den Wert 0 annehmen, so müssten die Anteile x_i*,* $i = 1,\dots,4$*, jeweils identisch gleich 0.25 sein. Das hieße, je 10 Kindern leben in Einkind-, Zweikind-, Dreikind- bzw. Vierkindfamilien. Da die Anzahl derjenigen Kindern, welche in Dreikindfamilien leben, durch 3 teilbar sein muss (analog für Vierkindfamilien etc.), ist ein Gini-Koeffizient von 0 in der gegebenen Situation unmöglich.*

H 4 Konzentrationsmaße

In einer Gemeinde gibt es 10 Landwirte. Ordnen Sie den unten abgebildeten Lorenzkurven die entsprechenden Situationen zu.

A: Fünf Landwirte halten sich keine Milchkühe, alle anderen haben jeweils denselben Milchkuhbestand

B: Ein Landwirt besitzt die Hälfte des gesamten Milchkuhbestandes der Gemeinde, die übrigen Kühe sind auf die restlichen Landwirte gleichmäßig aufgeteilt

C: Ein Landwirt ist im Besitz aller Milchkühe

D: Ein Landwirt war im Besitz aller Milchkühe. Inzwischen hat er die Hälfte der Kühe jedoch an einen anderen Landwirt der Gemeinde verkauft.

Bild (1) gehört zu Situation A, (3) zu B, (5) zu C und (9) zu D.

Beachte insbesondere, dass der Bereich auf der x-Achse bei einer Lorenzkurve von 0 bis 1 geht, nicht bis 10 wie in manchen der abgebildeten Kurven.