Angenommen bei der Aufgabe 3b) von Übungsblatt 3, es hätte einen Wert 570 statt 500 gegeben. Dann wäre der ja als Ausreißer einzustufen. Bis wohin hätte ich dann den oberen Zaun (Whisker) zeichnen müssen?

Antwort:
Den oberen Whisker hätte man bis zu 300 zeichnen müssen. Das ist der größte beobachtete Wert, der nicht als Ausreißer eingestuft wird (Siehe Folie 3.16).

Frage:

Übungsblatt 3: Wie beschriftet man beim Boxplot die Achsen? Mit x und y?

Antwort:

- Einfacher Boxplot: die x-Achse enthält keine Beschriftung, die y-Achse wird mit dem Namen des untersuchten Merkmals beschriftet.
- Gruppierter Boxplot: auf der x-Achse ist die Variable aufgetragen, welche die Gruppen definiert, die y-Achse wird mit dem Namen des Merkmals beschriftet (vgl. Angabe zur Aufgabe 4b).

Frage:

Übungsblatt 3: Was ist der Unterschied zwischen Quantil und Quartil?

Antwort:

Das α-Quantil einer Verteilung trennt die Daten in zwei Teilen so, dass etwa $\alpha \cdot 100\%$ der Daten darunter und $(1 - \alpha) \cdot 100\%$ darüber liegen. Damit ist der Median das 0,5-Quantil.

Wichtige Quantile sind etwa die *Dezile* mit $\alpha = 0{,}1, 0{,}2, \ldots, 0{,}9$.

Die Quantile werden oft auch in *Prozentangaben* umschrieben. Man sagt zum Beispiel 90%-Quantil für $\alpha = 0{,}9$ (siehe auch Folie 3.11).

Die *Quartile* sind Quantile, die für die Charakterisierung einer Verteilung und die Erstellung von Box-Plots eine besondere Bedeutung haben (vgl. Folie 3.13). Es gibt drei Quartile: das untere Quartil (25%-Quantil), den Median (50%-Quantil) und das obere Quartil (75%-Quantil).

Übungsblatt 5: Was ist der Unterschied zwischen den Notationen i und j, z.B. bei Aufgabe 2a oder 2b?

Antwort:

i und j sind Indizes.

Bei Aufgabe 2a wird nur ein Index verwendet, nämlich der Summationsindex i.

$$G = \frac{2\sum_{i=1}^{n} i x_{(i)} - (n+1)\sum_{i=1}^{n} x_{(i)}}{n \sum_{i=1}^{n} x_{(i)}}$$

Bei Aufgabe 2b hätte man bei der Berechnung des Gini-Koeffizienten auch den Summationsindex i verwenden können (siehe Formelsammlung S.3).

$$G = 1 - \frac{1}{n}\sum_{j=1}^{n}(v_{j-1} + v_j) = 1 - \frac{1}{n}\sum_{i=1}^{n}(v_{i-1} + v_i)$$

Bei der Ermittlung der Werte v_i für die Lorenzkurve in Aufgabe 2b werden zwei verschiedene Indizes benötigt.

$$v_i = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}}, \quad i = 1, \ldots, n$$

Beispiel:

$$v_3 = \frac{\sum_{j=1}^{3} x_{(j)}}{\sum_{j=1}^{5} x_{(j)}} = \frac{20 + 40 + 50}{255}$$

Übungsblatt 4, Aufgabe 2: Angenommen die absoluten Häufigkeiten für die Klasse > 4 wären ungleich Null, z.B. 2. Wie berechnet man das dann gegeben der Tatsache, dass die Klasse nun offen ist?

Antwort:

Bei einer Häufigkeit von 2 für die Ausprägung '>= 4' in der Häufigkeitstabelle ergibt sich für die Teilaufgaben:

a) Modus: $\bar{x}_M = 2$

   Median: Da $n = 22$ gerade:

$$\tilde{x}_{0.5} = \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}) = \frac{1}{2}(x_{(11)} + x_{(12)}) = \frac{1}{2}(2 + 2) = 2$$

b) Unteres Quartil: Da $n \cdot \alpha = 22 \cdot 0{,}25 = 5{,}5$ nicht ganzzahlig:

$$\tilde{x}_{0{,}25} = x_{(6)} = 0$$

   Oberes Quartil: Da $n \cdot \alpha = 22 \cdot 0{,}75 = 16{,}5$ nicht ganzzahlig:

$$\tilde{x}_{0{,}75} = x_{(17)} = 2$$

d)

$$F(2{,}5) = \sum_{a_j \leq 2{,}5} f(a_j) = f(0) + f(1) + f(2) = \frac{6}{22} + \frac{4}{22} + \frac{7}{22} = \frac{17}{22} = 0{,}77$$

Das arithmetische Mittel, sowie die empirische Varianz und die empirische Standardabweichung können nicht berechnet werden, da die Werte $x_{(21)}$ und $x_{(22)}$ unbekannt sind.