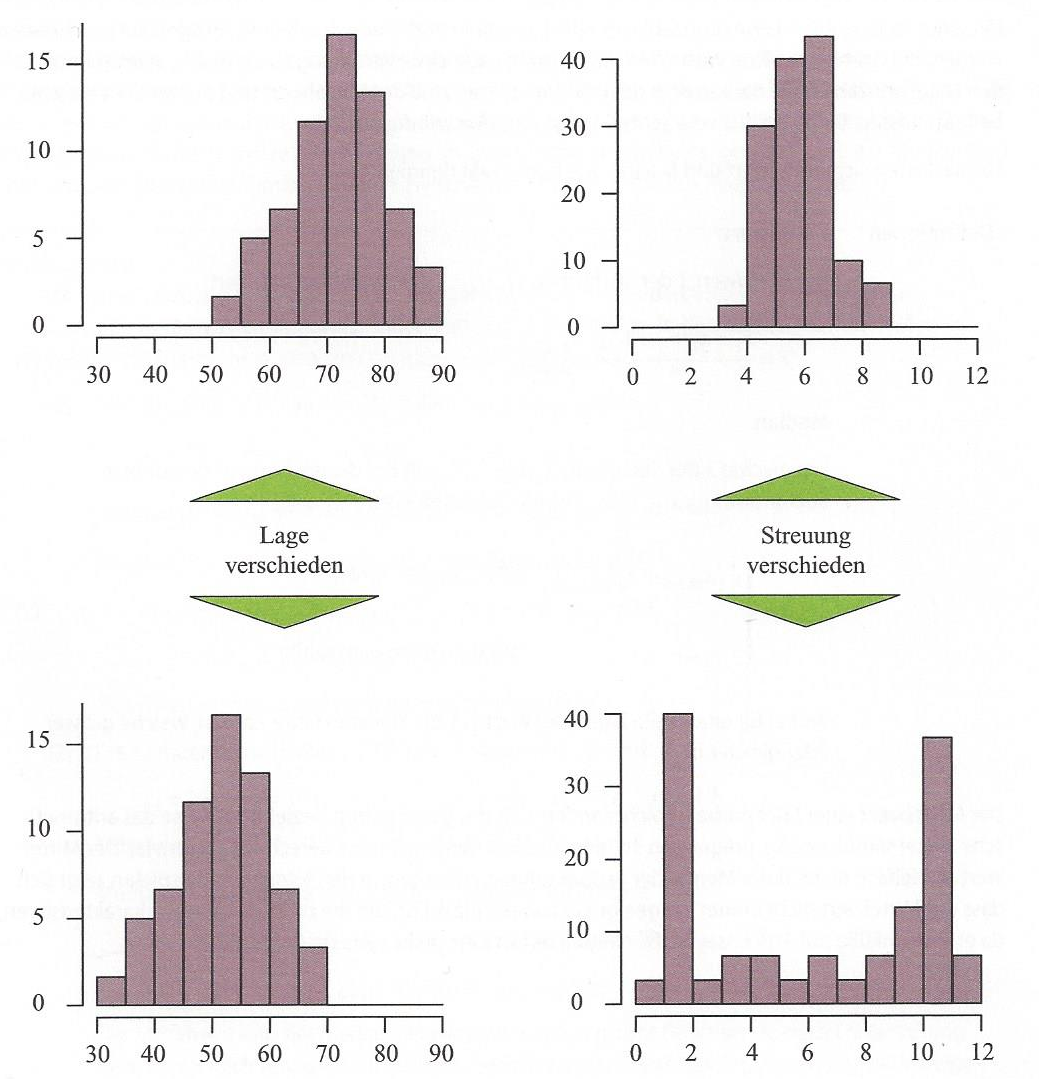
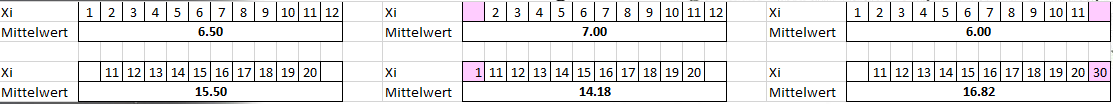
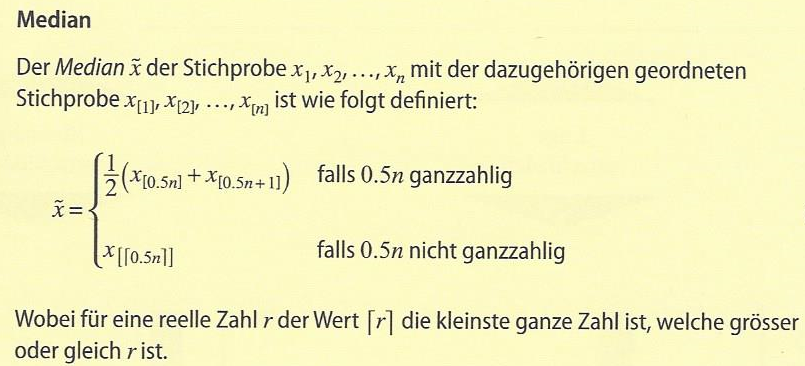
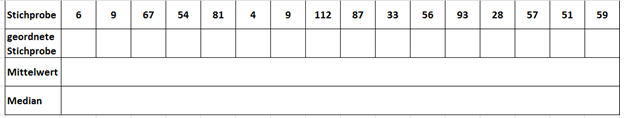
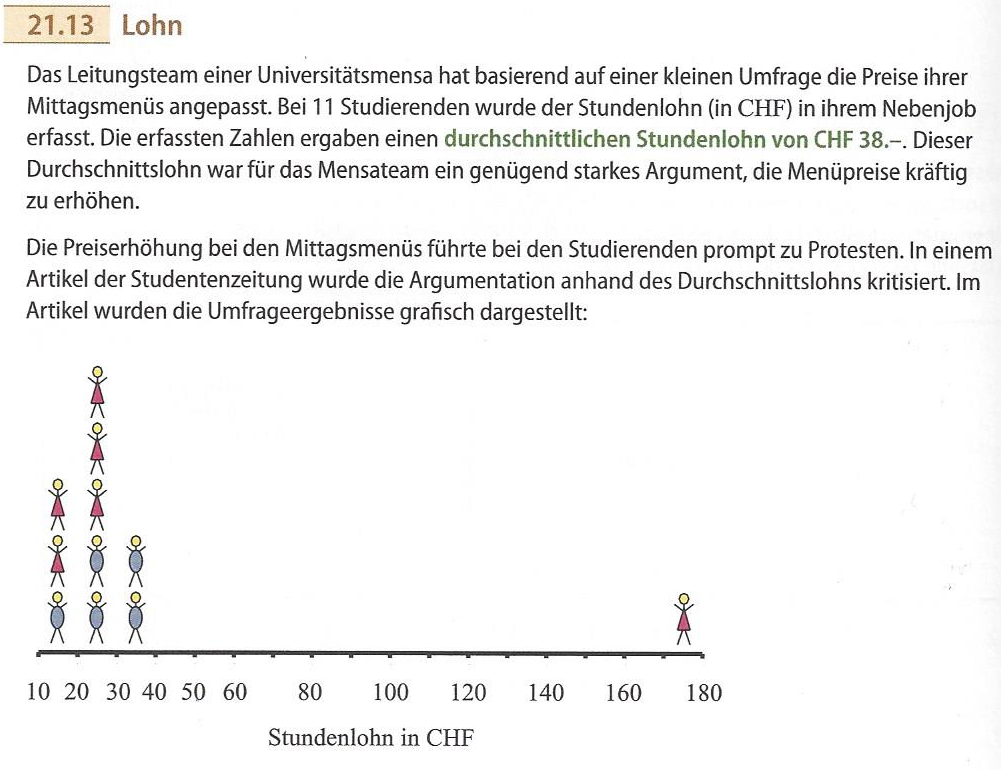
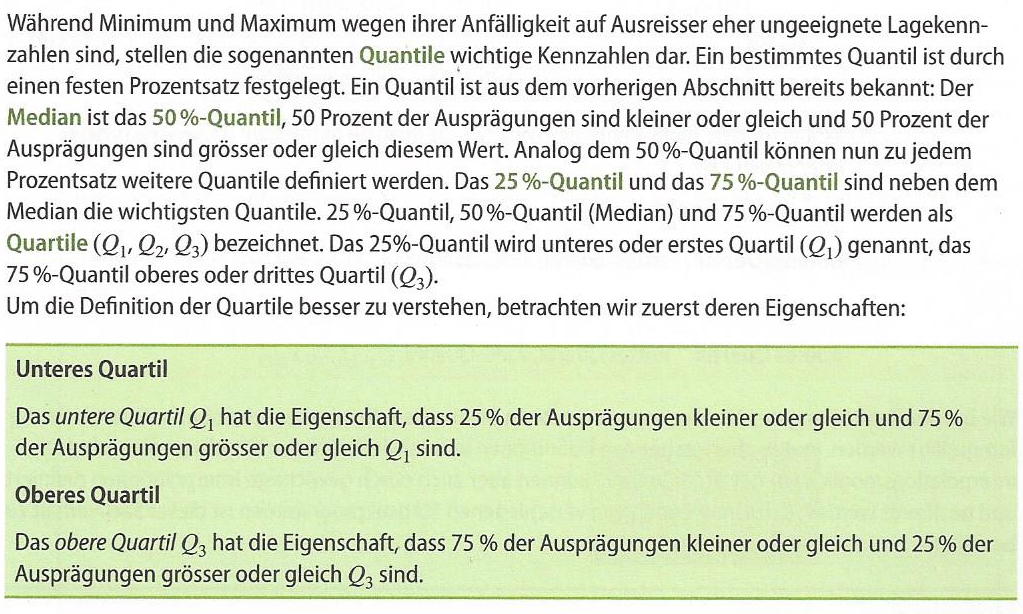
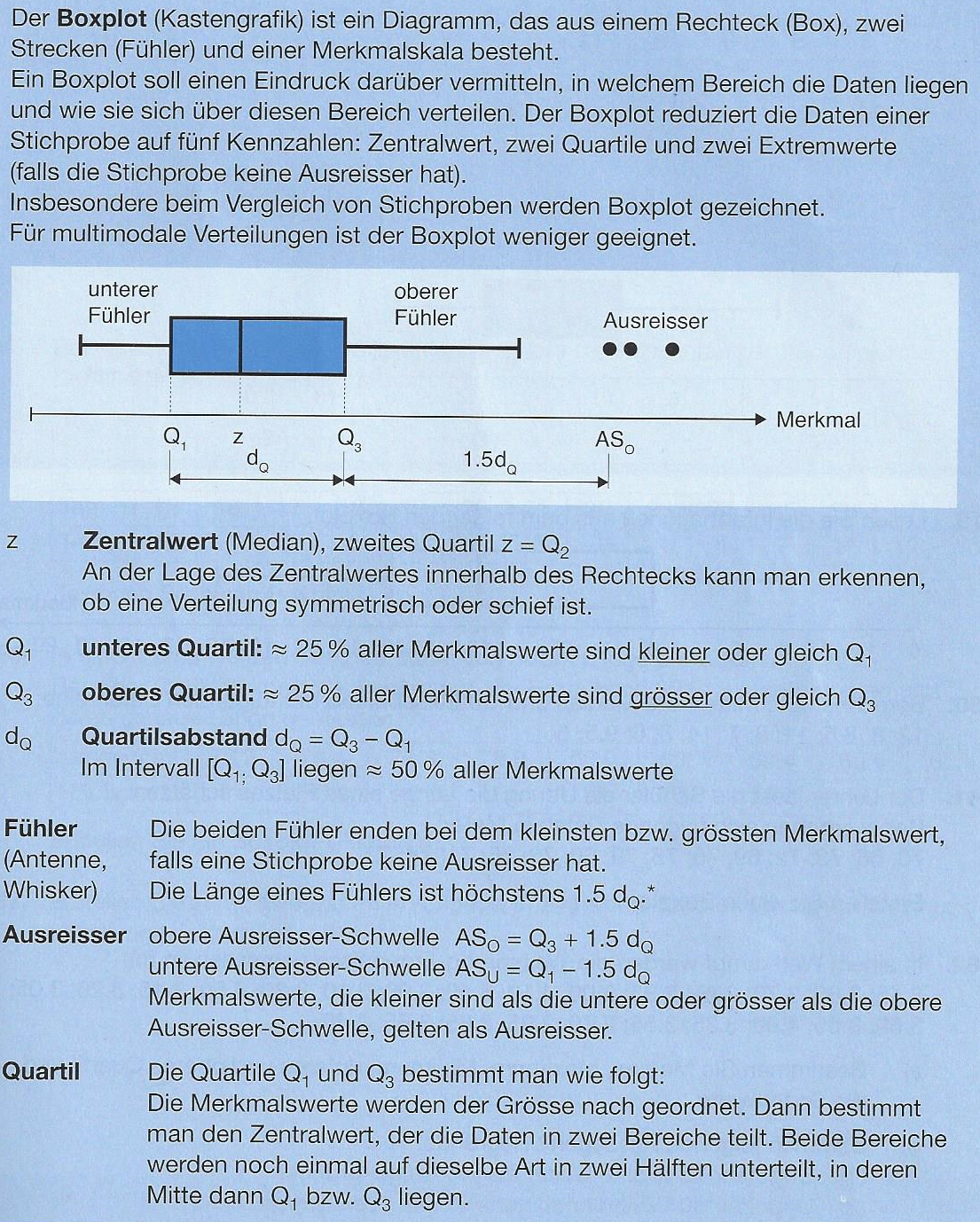
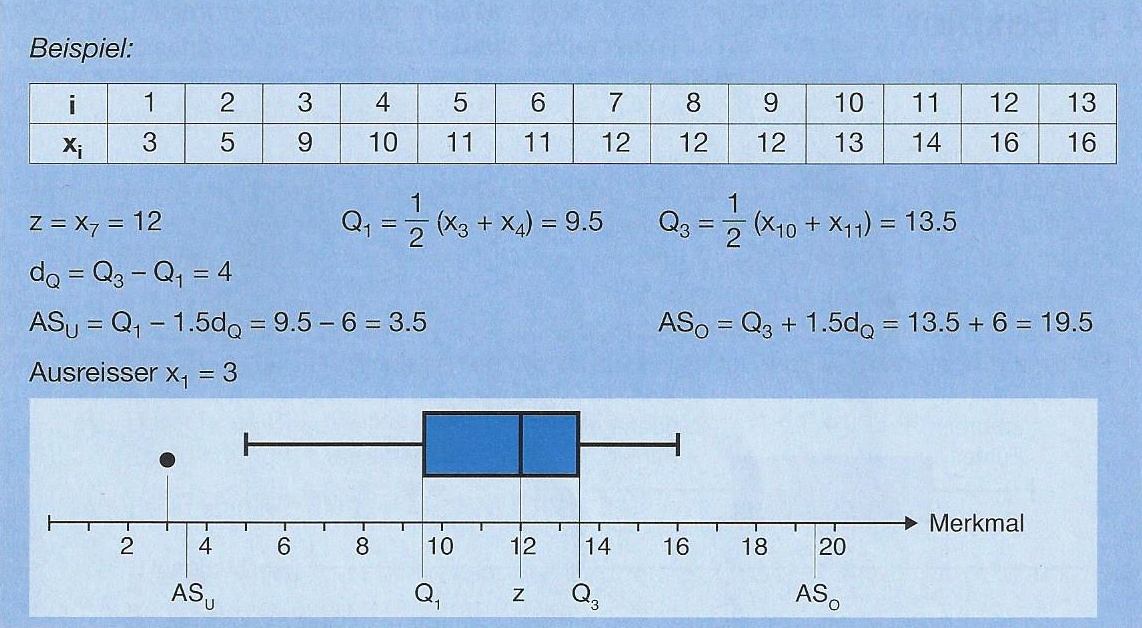
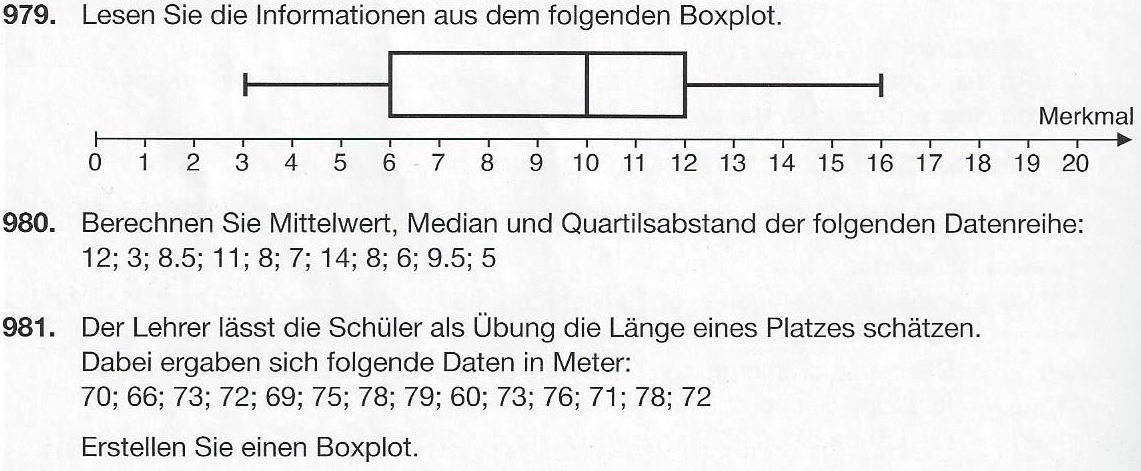
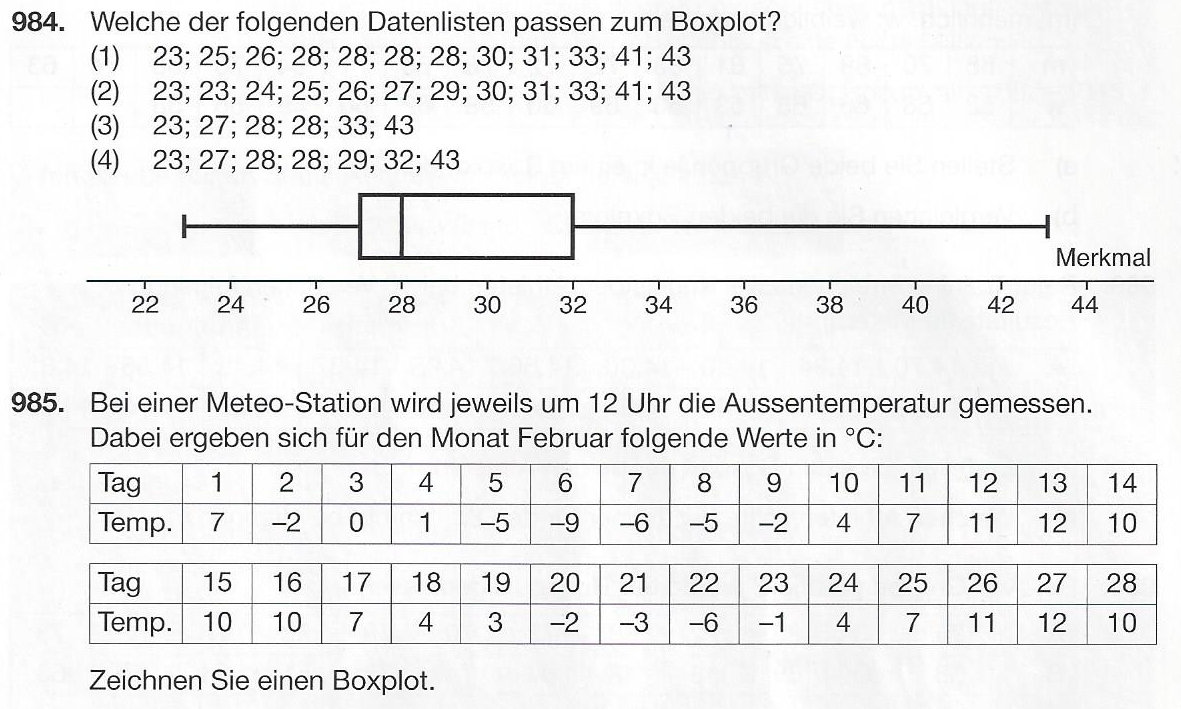
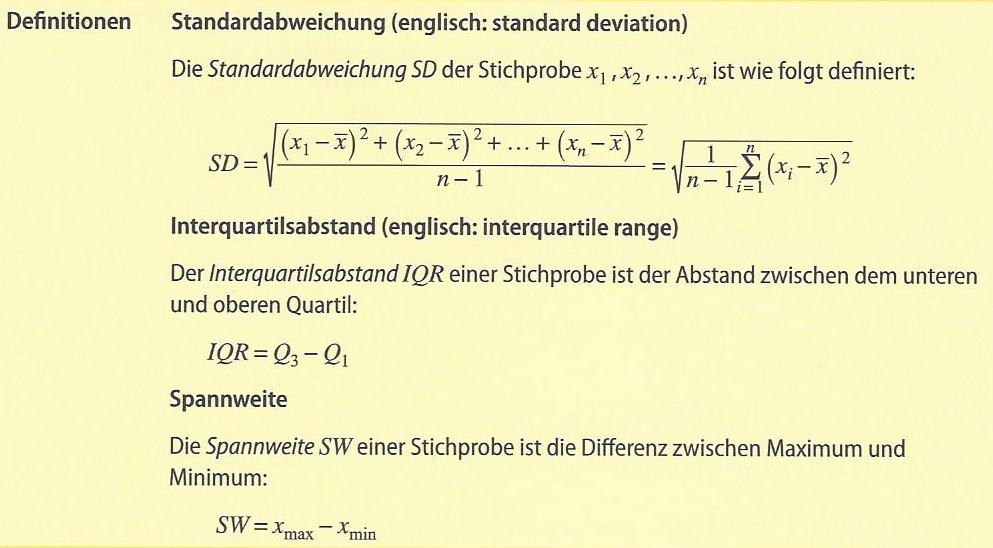
**DATENANALYSE I V**  BBB / BM15 /27. April 2017  
  
**Kennzahlen und Boxplot (mit Übungen)**Durch Kennzahlen können *Eigenschaften (Ausprägungen) von Verteilungen charakterisiert* werden. Mit Kennzahlen werden ausschliesslich *quantitative Merkmale* beschrieben. Wir werden Kennzahlen kennen lernen, welche die *Lage* und die *Streuung von Verteilungen* beschreiben:  
  
  
  
Robustheit: Eine Kennzahl wird als robust bezeichnet, wenn das Verändern, Hinzufügen oder Weglassen einer einzelnen Beobachtung ihren Wert nicht stark beeinflusst.  
 **1. Lagekennzahlen**  
Mit einer Lagekennzahl wird die Lage (Position) einer Verteilung durch eine einzige Zahl beschrieben. Die zentrale Lage einer Verteilung wird durch den *Mittelwert* oder den *Median* beschrieben:  
  
**1.1 Mittelwert**  
Der Mittelwert einer Stichprobe x1, x2, x3, ….. , xn ist wie folgt definiert: x1 + x2 + x3 + ….. + xn  
  
 (arithmetisches Mittel)  
  
Der Mittelwert einer Stichprobe ist nichts anderes als der Durchschnitt beziehungsweise das arithmetische Mittel sämtlicher Einzeldaten (Ausprägungen) des untersuchten Merkmals. Der Mittelwert ist vielfach nicht in der Menge der erfassten Daten enthalten. Er ist nicht immer die geeignete Lagekennzahl, um die zentrale Lage eines Merkmals zu charakterisieren, da er sehr anfällig auf Ausreisser ist. Der Mittelwert ist eine *nicht robuste* Kennzahl. Bp:  
  
  
  
 **1.2 Median**  
  
Bei schiefen Verteilungen ist der sogenannte Median besser geeignet, um die zentrale Lage der Verteilung zu beschreiben. Er ist eine *robuste* Lagekennzahl und ist wie folgt definiert:  
  
  
  
Der Median hat die Eigenschaft, dass 50% der Einzeldaten (Ausprägungen) kleiner oder gleich und 50% der Einzeldaten (Ausprägungen) grösser oder gleich sind.   
  
Um den Median zu bestimmen, muss die Stich probe der Grösse nach geordnet werden: Aus der Stichprobe wird also eine geordnete Stichprobe erstellt.  
  
Ist der Stichprobenumfang n eine ungerade Anzahl, ist der Median derjenige Wert, welcher genau in der mittleren Position der Stichprobe liegt, ist der Stichprobenumfang eine gerade Anzahl, so ist der Median der Durchschnitt der beiden in der Mitte liegenden Werte.

Bp:  
  
  
  
Median und Mittelwert können also bei bestimmten Stichproben zusammenfallen.  
  
  
  
**Übung 1:**  
Berechnen sie den Mittelwert und den Median (auf eine Stelle nach dem Komma):  
  
  
  
 **Übung 2: Beispiel 21.13 «Lohn» : Veranschaulichung Mittelwert - Median**  
  
Zur Analyse der graphischen Darstellung in der Studentenzeitung sollen nun Berechnungen gemacht werden. Zuerst sollen Mittelwert und Median in den beiden Merkmalen Geschlecht (männlich und weiblich) berechnet werden. (Hier mehr als Übung gedacht)  
  
Stundenlöhne der 5 Studenten: x1 = 27, x2 = 18, x3 = 23, x4 = 34, x5 = 32  
  
Mittelwert:   
  
Für den Median bilden wir zuerst die geordnete Stichprobe:  
 x[1] = 18, x[2] = 23, x[3] = 27, x[4] = 32, x[5] = 34  
  
Median:  ***= 27***  
  
Mittelwert und Median liegen hier sehr nahe zusammen.   
  
  
Stundenlöhne der 5 Studentinnen: x1 = 23, x2 = 171, x3 = 26, x4 = 17, x5 = 29, x6 =19  
  
Mittelwert:   
  
Für den Median bilden wir zuerst die geordnete Stichprobe:  
 x[1] = 17, x[2] = 19, x[3] = 23, x[4] = 26, x[5] = 29, x[6] = 171  
  
Median:  ***= 24,5***  
  
Mittelwert und Median liegen hier deutlich auseinander. 🡪 Je nach Fragestellung und Argumentation müsste man sich genau überlegen, welche Lagekennzahl als Argumentationshilfe beigezogen wird!  
  
  
  
  
  
Stundenlöhne aller Studenten: (Stundenlöhne siehe oben)

Mittelwert:   
  
Median:  ***= 26***

Die Mensaleitung hätte sich diesen Wiederstand ersparen können, wenn sie von Beginn weg mehr als nur eine Lagekennzahl betrachtet hätte: Der Ausreisser von 171 verzerrt das Bild des Mittelwertes. Der Median ist hier viel aussagekräftiger!

Bei quantitativen Merkmalen (Gewicht, Stundenlohn, Durchmesser, etc.) können Mittelwert und Median immer berechnet werden. Von einem nominalen Merkmal (männlich, weiblich) können weder Mittelwert noch Median berechnet werden.  
  
  
**1.3 Modus**  
Der Modus xmod ist derjenige Wert (Ausprägung), welcher in der Stichprobe am häufigsten vorkommt.  
  
  
**2. Weitere Kennzahlen**  
  
Das **Minimum xmin**eines Merkmals einer Stichprobe ist der kleinste auftretende Wert (Ausprägung).  
Das **Maximum xmax**eines Merkmals einer Stichprobe ist der grösste auftretende Wert (Ausprägung).

  
  
  
  
  
  
  
  
**Der Boxplot** (aus Frommenweiler S. 263)  
  
  
  
  
  
  
  
  
  
  
Übungsaufgaben:  
  
  
  
  
 **3. Streuungskennzahlen**Bis jetzt haben wir die *Lage* der Verteilung von Daten beschrieben (Lagekennzahlen). Neben der Lage ist auch die *Streuung* von Interesse. Wir definieren drei Streuungskennzahlen:  
  
**3.1 Standardabweichung**  
**3.2 Interquartilsabstand**  
**3.3 Spannweite**  
  
Der Interquartilsabstand *IQR* ist die Breite des Balkens vom Boxplot!  
  
**Weitere Boxplot-Aufgaben:**

**Aufgabe 1**

Eine Pharmafirma möchte wissen, ob ihr neu entwickeltes Medikament gegen Bluthochdruck (hypertension) besser wirkt als ein älteres Medikament. Dazu wurde eine Gruppe von 21 Leuten untersucht. Davon bekamen 10 Teilnehmer das ältere Medikament, die restlichen 11 das neue Medikament. Der Blutdruck wurde jeweils zu Beginn des Tests und nach 2 Wochen gemessen. Erfasst wurde dann jeweils die Höhe der Blutdrucksenkung:

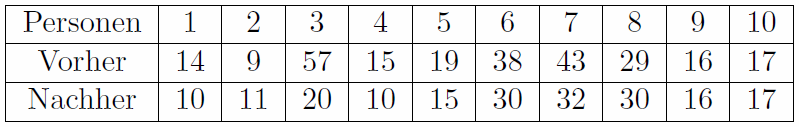
- mit dem alten Medikament: 7; 15; 11; 15; 25; 4; 19; 10; 8; 31

- mit dem neuen Medikament: 11; 16; 25; 9; 36; 40; 48; 4; 26; 36; 14

Prüfen Sie mit Hilfe von Boxplots, ob das neue Medikament eine bessere Wirkung hat als das alte Medikament. Begründen Sie Ihre Aussage.

**Aufgabe 2**

Bei Piloten kommt es darauf an, dass sie möglichst schnell auf optische Signale reagieren. Um die Reaktionszeit zu verbessern, wurde eine Trainingsmethode entwickelt. Eine Untersuchung an 10 Personen ergab folgende Reaktionszeiten vor Anwendung und nach Anwendung der Trainingsmethode (in Sekunden):



Untersuchen Sie mit einer geeigneten Statistik-Methode, ob das Training eine Verbesserung der Reaktionszeit (zumindest in der Stichprobe) bewirkt! Fertigen Sie gegebenenfalls Zeichnungen an und begründen Sie Ihre Aussagen!

**Aufgabe zur Erstellung eines Boxplots:**

1. **Aufgabe**

Die Körpergewichte der Lernenden einer Klasse sind nach Geschlechtern aufgeteilt (m, w).

Angaben in kg.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| m | 67 | 60 | 70 | 78 | 84 | 68 | 67 | 70 | 73 | 72 | 68 | 75 | 76 |  |
| w | 52 | 55 | 63 | 63 | 63 | 57 | 58 | 55 | 51 | 60 | 64 | 51 | 54 | 59 |

Berechnen Sie nach Geschlechter getrennt alle bekannten Kennzahlen und zeichnen Sie für jedes Geschlecht einen Boxplot.

1. **Lösung:** Daten sortieren!

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| m |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| w |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Kennzahl** | **m** | **w** |
| Median |  |  |

**Boxplot**

**Bestimmung der Antennen**

|  |  |  |
| --- | --- | --- |
|  | **m** | **w** |
| Untere Antenne |  |  |
| Obere Antenne |  |  |

Grafik

