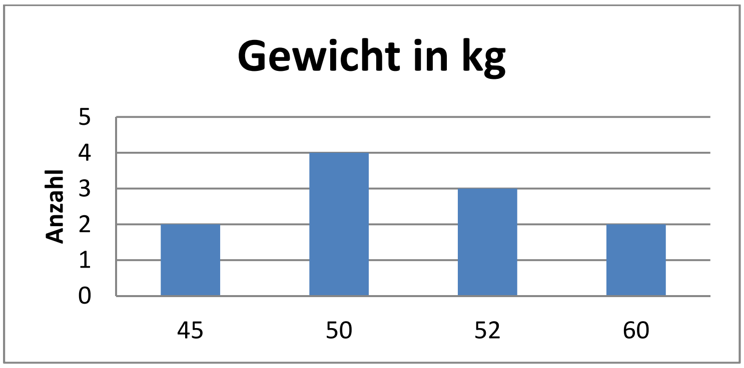
Aufgaben zur Statistik

1) Gegeben ist die Stichprobe: 5€, 15€, 8€, 10€, 12€.
Es soll das Minimum, das Maximum, der Median und der Mittelwert bestimmt werden, sowie die Varianz und die Standardabweichung.

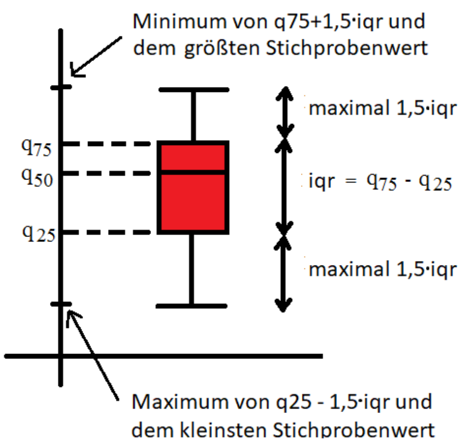
2) Gegeben ist die Stichprobe: 180, 160, 150, 160, 170, 180, 175, 150.
Gesucht werden das Minimum, das Maximum, der Median und der Mittelwert.
Ändert sich der Median, wenn die beiden 180 jeweils durch eine 200 ersetzt werden?

3) Gegeben ist das folgende Balkendiagramm, in dem das Ergebnis einer Umfrage nach dem Körpergewicht dargestellt wird:



a) Wie viele Personen wogen 50 kg und wie groß ist der relative bzw. prozentuale Anteil?
b) Wie viele Personen wurden befragt?
c) Wie groß ist der Median und wie ist dieser zu interpretieren?
d) Gesucht werden der Mittelwert, das Minimum und das Maximum.

4) Es wurden 10 Personen nach dem Monatsnettoeinkommen gefragt. Die Umfrage ergab folgende Werte: 1200€, 1500€, 1600 €, 1800 €, 2500 €, 2500 €, 3000 €, 3600 €, 4000 €, 6800 €.
a) Gesucht werden der Median, das untere Quartil (Q1 oder Q25 bzw. 25 % Quartil), das obere Quartil (Q3 oder Q25 bzw. 75 % Quartil), sowie das Minimum und das Maximum. Diese Kenngrößen sollen in einem Boxplot dargestellt werden (1 cm entsprechen 500 €).
b) Was fällt beim Vergleich des Medians mit dem Mittelwert (arithmetisches Mittel) auf?
Boxplot:

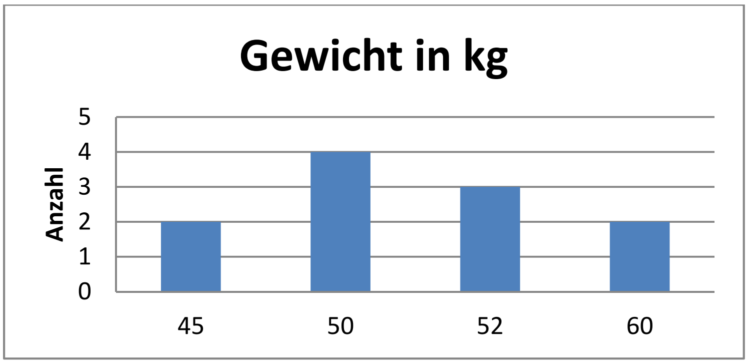


Lösungen

1) Die sortierte Stichprobe wird für den Median benötigt: 5€, 8€, 10€, 12€, 15€.
Das Minimum beträgt 5€ und das Maximum 15€.
Median: Anzahl n = 5. Die Anzahl ist ungerade, womit sich ein Wert in der Mitte befindet. Wir bestimmen die Position in der sortierten Stichprobe: 5 · 0,5 = 2,5. Ergibt sich keine ganze Zahl - wie hier -, müssen wir immer aufrunden und der Medien ist bei n = 5 der 3. Wert in der sortieren Stichprobe: Median = 10 €
Der Mittelwert bzw. das arithmetische Mittel ist die Summe aller Wert dividiert durch die Anzahl:
$\bar{x}$ = (5 € + 8€ + 10€ + 12 € + 15€) / 5 = 10€
Die Varianz (wir verwenden, wie in der Schulmathematik üblich, den Faktor 1/n):
$s^2 = ((5€ - 10€)^2 + (8€ - 10€)^2 + (10€ - 10€)^2 + (12€ - 10€)^2 + (15€ - 10€)^2) / 5 = 11{,}6€^2$
Zur Berechnung der Varianz wurde von jedem Stichprobenwert x_i der Mittelwert $\bar{x}$ subtrahiert und das Ergebnis quadriert: $(x_i - \bar{x})^2$. Die Summe dieser Quadratwerte durch n dividiert ergibt die Varianz. Siehe: http://mathe-total.de/Stochastik/Stichproben-und-deren-Kenngroessen.pdf
Die Standardabweichung: $s = \sqrt{s^2} = \sqrt{11{,}6€^2} \approx 3{,}41€$

2) Sortierte Stichprobe: 150, 150, 160, 160, 170, 175, 180, 180.
Das Minimum ist 150 und das Maximum 180.
Median: Anzahl n = 8. Die Anzahl ist gerade und 8 · 0,5 = 4. Wenn sich eine ganze Zahl ergibt, wie immer bei einer geraden Anzahl, muss der Mittelwert der beiden Werte in der Mitte gebildet werden, also hier aus dem 4. und dem 5. Wert (also aus dem (n/2)-ten und (n/2+1)-ten Wert).
Median = (160 + 170) / 2 = 165
Mittelwert: $\bar{x}$ = (150 + 150 + 160 + 160 + 170 + 175 + 180 + 180) / 8 = 165,625
Für die Berechnung des Mittelwerts muss natürlich nicht sortiert werden. Der Median ändert sich nicht, wenn wir die 180 jeweils durch eine 200 ersetzen, denn auch bei der Stichprobe 150, 150, 160, 160, 170, 175, 200, 200 ist der Median 165. Der Mittelwert würde sich aber ändern.

3) Wir können die absoluten Häufigkeiten im Diagramm ablesen:



Es ergibt sich die folgende Häufigkeitstabelle:

Gewicht in kg	absolute Häufigkeit
45	2
50	4
52	3
60	2

Die sortiere Stichprobe wäre 45, 45, 50, 50, 50, 50, 52, 52, 52, 60, 60 (alles in kg).

a) 4 Personen wogen 50 kg und die relative Häufigkeit dafür ist absolute Häufigkeit durch die Anzahl dividiert, also $4/11 \approx 0{,}364 = 36{,}4\ \%$.
b) Wie viele Personen wurden befragt? $n = 2 + 4 + 3 + 2 = 11$
c) Der Median ist der 6. Wert in der sortierten Stichprobe (da $11 \cdot 0{,}5 = 5{,}5$ und 5,5 aufgerundet 6 ergibt). Der 6. Wert ist 50 kg, was auch an der Tabelle gesehen werden kann. Der Median ist damit 50 kg und mindestens die Hälfte der Werte sind kleiner oder gleich 50 kg. Damit wiegen mindestens die Hälfte der Personen höchstens 50 kg.
d) Mittelwert: $\bar{x} = (2 \cdot 45\text{ kg} + 4 \cdot 50\text{ kg} + 3 \cdot 52\text{ kg} + 2 \cdot 60\text{ kg}) / 11 \approx 51{,}45\text{ kg}$
Der Mittelwert liegt „nahe“ beim Median, da das Balkendiagramm „fast“ symmetrisch ist.
Minimum: 45 kg
Maximum: 60 kg

4) Die Stichprobe ist bereits sortiert: 1200€, 1500€, 1600 €, 1800 €, 2500 €, 2500 €, 3000 €, 3600 €, 4000 €, 6800 €.

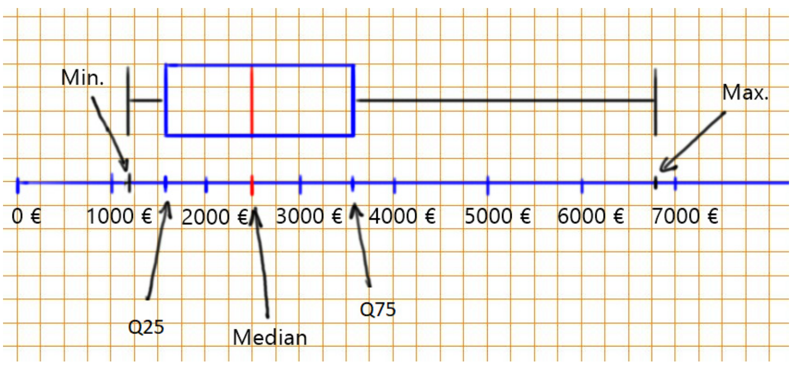
a) n = 10 und gerade, so dass wir für den Median wieder den Mittelwerte aus dem 5. Wert (da $10 \cdot 0{,}5 = 5$ ganzzahlig ist) und dem 6. Wert der sortierten Stichprobe berechnen müssen.
Median = (2500 € + 2500 €) / 2 = 2500 €. Der Median ist das mittlere Quartil (bzw. das 50% Quartil).

Wir bestimmen die Position des unteren Quartils: $10 \cdot 0{,}25 = 2{,}5$. Hätte sich eine ganze Zahl ergeben, z.B. 2, dann wäre das untere Quartil der Mittelwert des 2. und 3. Wertes der sortierten Stichprobe. Da sich keine ganze Zahl ergibt, runden wir auf und Q25 = 1600 € (der 3. Wert).

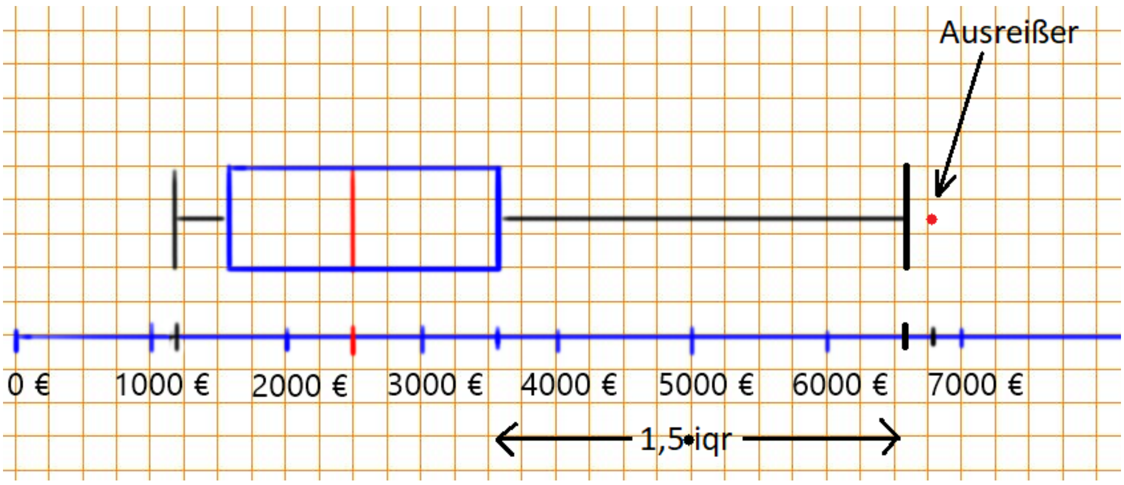
Zur Position des oberen Quartils: $10 \cdot 0{,}75 = 7{,}5$, also der 8. Wert (bzw. der 3. Wert von hinten). Damit gilt: Q75 = 3600 €

Das Minimum beträgt 1200 € und das Maximum beträgt 6800 €.

Zum Boxplot: 1 cm entspricht 500 €, 2 cm entsprechen damit 1000 € und 0,2 cm entsprechen 100 €. Wir wählen eine Boxhöhe von 1,5 cm (ist relativ egal).



Die Box zeigt den Bereich der „mittleren 50 %“, denn mindestens 25 % der Personen haben bis zu 1600 € verdient und mindestens 75 % bis zu 3600 €. Die schwarzen senkrechten Striche links und rechts kennzeichnen oben noch das Minimum und das Maximum und die rote Linie in der Box den Median (mindestens 50 % haben bis zu 2500 € verdient). Wir haben beim obigen einfacheren Boxplot die „Whiskers“ jeweils bis zum minimalen oder maximalen Wert gezeichnet. Der größte Wert ist hier ziemlich weit von den mittleren 50% entfernt. Hierzu sei folgendes bemerkt: Ist der kleinste bzw. größte Wert weiter als 1,5·(Q75 – Q25) von Q75 bzw. Q25 entfernt, dann gehen die Whiskers in der Regel nur bis zum Minimum bzw. Maximum. Details hierzu sind unter http://www.mathe-total.de/Buecher/Einstieg-in-die-angewandte-Statistik/Einstieg-in-die-Datenanalyse-mit-SPSS.pdf auf Seite 24 zu sehen. Wir sehen damit oben eine einfachere Variante eines Boxplot. Q75 – Q25 ist der Interquartilsabstand (iqr). Genau genommen sind die Whiskers maximal 1,5·(Q75 – Q25) = 1,5·(3600 € -1600 €) = 3000 € lang. Damit würde der obere Whisker nur bis zu Q75 + 1,5·(Q75 – Q25) = 3600 € + 3000 € = 6600 € gehen und nicht bis zum Maximum von 6800 €. Der untere Whisker kann bis zu Q25 - 1,5·(Q75 – Q25) = 1600 € - 3000 € = -1400 € gehen. Da aber das Minimum bei 1200 € liegt, geht dieser nur bis zu den Minimum. Analog würde der obere Whisker nur bis zum Maximum gehen, wenn dieses kleiner als 6600 € wäre. Hier mit angepasstem oberem Whisker:



b) Es soll der Median mit dem Mittelwert (arithmetisches Mittel) verglichen werden:
Der Median lag bei 2500 € und der Mittelwert beträgt:
(1200€ + 1500€ + 1600 € + 1800 € + 2500 € + 2500 € + 3000 € + 3600 € + 4000 € + 6800 €)/10 = 2850 €

Der Mittelwert liegt durch den großen Wert 6800 €, der relativ weit von den restlichen Werten entfernt liegt, 350 € über dem Median. Wäre der größte Wert statt 6800 € z.B. 10000 € gewesen, dann läge der Mittelwert mit 3170 € noch weiter über dem Median, der dann immer noch 2500 € betragen würde.

Theoretische Bemerkung:
Bei einer stetigen (theoretischen) Verteilung beträgt die Wahrscheinlichkeit genau 50%, dass ein Wert kleiner oder gleich dem Median ist. Mathematisch gesprochen beträgt die Wahrscheinlichkeit, dass die Zufallsvariable X einen Wert kleiner oder gleich dem Median annimmt, genau 50%:
$P(X \leq \text{Median}) = 50\%$. Bei einer Stichprobe und auch bei diskreten Verteilungen sind es nicht immer genau 50%.
Bei der Stichprobe 150, 160, 180, 190 wäre der Median (160 + 180) / 2 = 170 und es wären genau 50% der Wert kleiner oder gleich dem Median. Wenn aber die Stichprobe 150, 160, 160, 190 wäre, dann würden mindestens 50% kleiner oder gleich dem Median von 160 sein, genau genommen 3/4 bzw. 75% aller Werte.
Bei der Stichprobe 5€, 8€, 10€, 12€, 15€ aus Aufgabe 1 sind auch mindestens 50% der Werte kleiner oder gleich dem Median von 10 €, hier genau genommen 3/5 bzw. 60% aller Wert.