Übungsaufgaben Kapitel 3
Lösungen

Prof. Dr. Torben Kuhlenkasper

Aufgabe 1

Ein Statistiker sieht sich neun aufeinanderfolgende Sendungen der Show WER WIRD MILLIONÄR an und notiert sich am Ende der Sendung den realisierten Gesamtgewinn des Tages. Es ergaben sich folgende Werte (in Tausend EUR):

```
34   17   96   33   189   282   33   66   64
```

1. Bestimmen Sie den Mittelwert und den Median.

   Mittelwert:

$$\begin{aligned} \bar{x} &= \frac{1}{9} \cdot (34 + 17 + 96 + 33 + 189 + 282 \\ &+ 33 + 66 + 64) \\ &= \frac{814}{9} = 90.44 \end{aligned}$$

   Median:
   sortierter Datensatz:

```
17   33   33   34   64   66   96   189   282
```

$$x_{0.5} = x_{(k)}$$

   mit $k = (9+1)/2 = 5$

$$x_{0.5} = x_{(5)} = 64$$

2. Bestimmen Sie die Stichprobenvarianz und die Standardabweichung.

   Stichprobenvarianz:

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ &= \frac{1}{8}[(34-90.44)^2+(17-90.44)^2 \\ &+ \ ...+(64-90.44)^2] \\ &= \frac{1}{8}\cdot 62914.22 = 7864.28 \end{aligned}$$

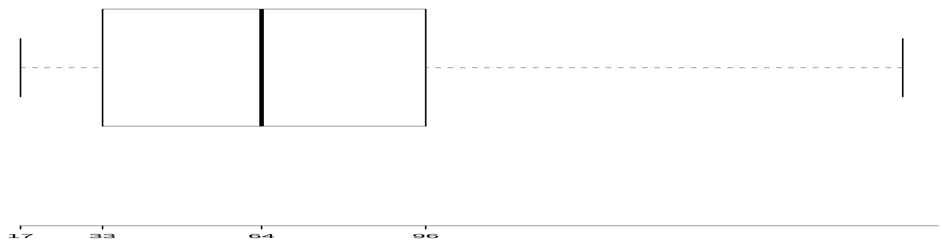
   Stichprobenstandardabweichung:

$$s=\sqrt{s^2}=\sqrt{7864.28}=88.68$$

3. Erstellen und interpretieren Sie den Boxplot.

   Für den Boxplot wird benötigt: Minimum, erstes Quartil, Median (=2. Quartil), drittes Quartil und Maximum

$$\begin{aligned} x_{min} &= 17 \\ x_{0.25} &: k = n\cdot p = 9\cdot 0.25 = 2.25 \\ x_{0.25} &= x_{(\lfloor 2.25\rfloor+1)} = x_{(3)} = 33 \\ x_{0.5} &= 64 \\ x_{0.75} &: k = n\cdot p = 9\cdot 0.75 = 6.75 \\ x_{0.75} &= x_{(\lfloor 6.75\rfloor+1)} = x_{(7)} = 96 \\ x_{max} &= 282 \end{aligned}$$



mit einer Ausnahme (282) deutet der Boxplot auf Symmetrie hin.

Aufgabe 2

Es gab vor Joachim Löw insgesamt 8 Trainer bzw. Teamchefs der deutschen Fußballnationalmannschaft. In der folgenden Tabelle ist die Anzahl der Länderspiele (x) zu finden, bei denen die jeweiligen Personen Bundestrainer waren.

Name	Anzahl
Sepp Herberger	162
Helmut Schön	139
Jupp Derwall	67
Franz Beckenbauer	66
Berti Vogts	102
Erich Ribbeck	24
Rudi Völler	53
Jürgen Klinsmann	34

1. Erstellen Sie den Boxplot der Daten, ohne Ausreißer zu berücksichtigen.

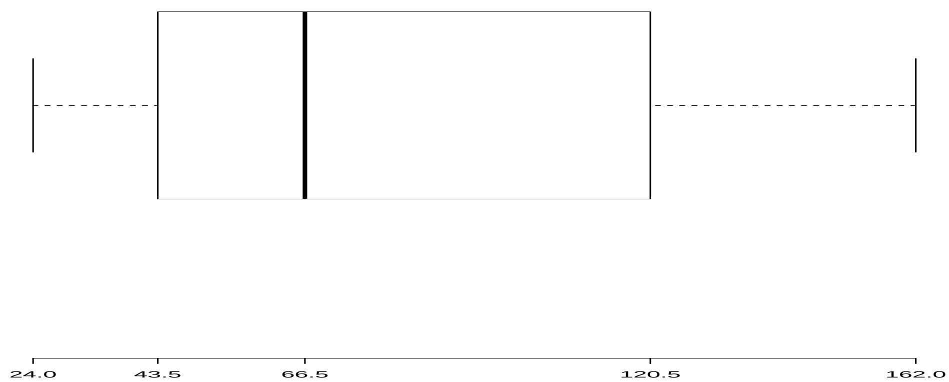
   sortierter Datensatz:

   ```
   24  34  53  66  67 102 139 162
   ```

$$(k = 8 \cdot 0.25) = 2 \rightarrow x_{0.25} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{34 + 53}{2} = 43.5$$

$$(k = 8 \cdot 0.5) = 4 \rightarrow x_{0.5} = \frac{x_{(4)} + x_{(5)}}{2} = \frac{66 + 67}{2} = 66.5$$

$$(k = 8 \cdot 0.75) = 6 \rightarrow x_{0.75} = \frac{x_{(6)} + x_{(7)}}{2} = \frac{102 + 139}{2} = 120.5$$



2. Berechnen Sie den Mittelwert und die Varianz.

   Rechenhilfe: $\overline{x^2} = 8669.375$

$$\bar{x} = \frac{1}{8} \cdot 647 = 80.88$$

$$\begin{aligned} d^2 &= \overline{x^2} - \overline{x}^2 \\ &= 8669.375 - 80.88^2 \\ &= 2127.801 \\ s^2 &= \frac{n}{n-1} \cdot d^2 \\ &= \frac{8}{7} \cdot 2127.801 \\ &= 2431.773 \end{aligned}$$

Aufgabe 3

Wir betrachten noch einmal den Datensatz und die Klasseneinteilung aus Aufgabe 5 auf dem 1. Aufgabenblatt.

1. Welche Wartezeit bis zum ersten Tor wurde in 25 Prozent der Spiele nicht überschritten? Bestimmen Sie diesen Wert

   - mit Hilfe der empirischen Verteilungsfunktion,

     es gilt

     $$x_p = x_{i-1}^* + \frac{p - F_n^*(x_{i-1}^*)}{f_i} \cdot \Delta_i$$

     hier ist $x_{0.25}$ gesucht (Datensatz sortieren)

     $$x_{0.25} = 15 + \frac{0.25 - 0.24}{0.32} \cdot 15 = 15.47$$

   - aus den Rohdaten.

     es gilt n=25 und p=0.25. Also ergibt sich $k = 25 \cdot 0.25 = 6.25$

     $$x_{0.25} = x_{(\lfloor 6.25 \rfloor + 1)} = x_{(7)} = 15$$

2. Bestimmen Sie aus den Rohdaten

   - das untere Quartil,

     das unterere Quartil oder erstes Quartil ist $x_{0.25} = 15$

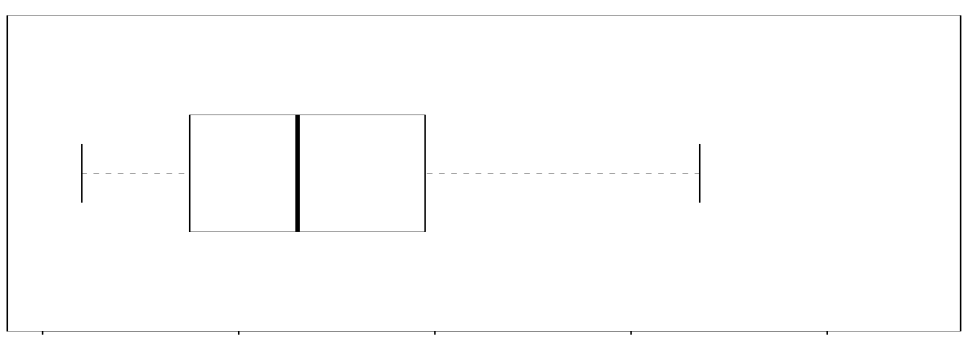
   - das obere Quartil,

     gesucht: $x_{0.75}$ mit $k = 25 \cdot 0.75 = 18.75$. Also ergibt sich hier $x_{(19)} = 39$

   - den Median.

     gesucht: $x_{0.5}$, n ist ungerade. Also: $k = \frac{25+1}{2} = 13$. Also ergibt sich hier $x_{(13)} = 26$

3. Zeichnen und interpretieren Sie den Boxplot.



Deutet auf eine symmetrische Verteilung hin.

4. Berechnen Sie den Mittelwert mit den klassierten Daten.

   es gilt für den Mittelwert bei klassierten Daten:

$$\bar{x} = \sum_{i=1}^{k} m_i \cdot f_i$$

   zunächst die Klassenmitten m_i bestimmen:

$$m_1 = \frac{0+15}{2} = 7.5 \quad m_2 = \frac{15+30}{2} = 22.5, \ldots, m_6 = \frac{75+90}{2} = 82.5$$

   es ergibt sich also:

$$\bar{x} = 7.5 \cdot 0.24 + 22.5 \cdot 0.32 + \ldots + 82.5 \cdot 0 = 29.1$$

Aufgabe 4

Gegeben sei folgender Datensatz:

```
2   1   18   3   6
```

1. Berechnen Sie für diesen Datensatz den Mittelwert und den Median.

   $\bar{x} = 6$

   $x_{0.5} = 3$ (Datensatz sortieren, $x_{(3)}$ auswählen)

2. Es stellt sich heraus, dass in obigem Datensatz keine 18, sondern 8 stehen sollte. Korrigieren Sie den Fehler und bestimmen Sie jetzt den Mittelwert und den Median.

   $\bar{x} = 4$

   $x_{0.5} = 3$ (Datensatz sortieren, wieder $x_{(3)}$ auswählen

3. Wie lauten der Mittelwert und Median, wenn Sie die 18 aus dem Datensatz herauslassen?

   $\bar{x} = 3$

   $x_{0.5} = 2.5$ (Datensatz sortieren, Mitte der zweiten und dritten Beobachtung nehmen)

4. Man sagt, dass der Mittelwert ausreißerempfindlich ist. Das heißt, dass eine extreme Beobachtung einen großen Einfluss auf den Wert des Mittelwerts hat. Gilt dies auch für den Median?

   Nein, der Median ist ein robustes und somit ausreißerunempfindliches Lagemaß. Der Median wird über die Position im geordneten Datensatz bestimmt. Der Mittelwert verteilt hingegen alle beobachteten Werte gleichmäßig auf den Stichprobenumfang n. Extreme Werte gehen mit ihrem vollen Wert in den Mittelwert ein.

Aufgabe 5

Sind die folgenden Aussagen wahr oder falsch?

- Ein Boxplot ist eine geeignete grafische Darstellung um die Verteilung eines Merkmals in verschiedenen Teilpopulationen zu vergleichen. (WAHR)
- Die Mitte der Box beim Boxplot entspricht dem Modus (Modalwert). (FALSCH)
- Die Länge der Box beim Boxplot ist gleich dem Quartilsabstand, d.h. der Differenz aus dem 75%- und dem 25%-Quantil. (WAHR)
- Beim Boxplot liegen etwa die Hälfte der Beobachtungen außerhalb der Box. Von diesen wiederum liegt etwa die älfte oberhalb, die andere Hälfte unterhalb der Box. (WAHR)
- Beim Boxplot liegen ungefähr 2/3 der Beobachtungen in der Box. (FALSCH)
- Ein Boxplot stellt ausschließlich die Daten zwischen dem ersten und dritten Quartil dar. (FALSCH)

Aufgabe 6

In ihrem ersten Job bekommen Sie gleich im ersten Jahr eine Gehaltssteigerung von 10%. Im darauffolgenden Jahr können Sie Ihr Gehalt um 6% steigern. Im dritten Jahr geht es der Firma schlecht und Sie müssen auf 4% ihres Gehalts verzichten.

Wie hat sich Ihr Gehalt in diesen drei Jahren durchschnittlich entwickelt?

Hier sind Veränderungsraten angegeben. Das arithmetische Mittel ist somit ungeeignet. Hier muss das geometrische Mittel verwendet werden.

es gilt:

$$\bar{v}_{geom} = \sqrt[n-1]{\prod_{i=1}^{n-1}(1+v_i)} - 1$$

Hier: n-1 = 3 (es bezieht sich auf $n = 4$ Jahre, somit drei Veränderungsraten)
$v_1 = 0.1$, $v_2 = 0.06$, $v_3 = -0.04$
Also:

$$\bar{v}_{geom} = \sqrt[3]{(1+0.1)\cdot(1+0.06)\cdot(1-0.04)} - 1 = 0.038$$

Im Durschnitt konnte er sein Gehalt um ca. 3.8% steigern.