20.02.20 **Übung 1 (für Pharma/Geo/Bio/Stat)** Uni Basel

Besprechung der Lösungen: 25./26. Februar 2020 in den Übungsstunden

Aufgabe 1
Bestimmen Sie zu den folgenden Datensätzen je den Mittelwert $\overline{x}$, den Median $\tilde{x}$, sowie die Quartile $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ und zeichnen Sie den Boxplot.

(a) 5, 6, 1, 2, 4, 6, 6, 6, 7, 8, 5, 4, 9, 11, 4, 6, 7

(b) 5, 6, 1, 2, 4, 6, 6, 6, 7, 8, 5, 4, 9, 12, 4, 6, 7

Aufgabe 2
Gegeben ist der folgende Datensatz:

$$8,\ 9,\ 3,\ 2,\ 14,\ 4,\ 10,\ 6,\ 3$$

Verändern Sie genau eine Zahl des Datensatzes so, dass $\overline{x} = 7$, aber $\tilde{x}$, $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ unverändert bleiben.

Aufgabe 3
Bei 16 PKWs desselben Typs wurde der Benzinverbrauch pro 100 km gemessen. Dabei ergab sich die folgende Urliste (in Liter pro 100 km):

$10,8$	$9,9$	$10,2$	$10,4$	$11,3$	$9,3$	$9,1$	$10,4$	$10,1$	$11,7$
$10,9$	$10,4$	$10,8$	$11,3$	$11,5$	$9,8$				

(a) Berechnen Sie das arithmetische Mittel, den Median, die Quartile und die Standardabweichung.

(b) Welcher Benzinverbrauch wurde von genau 75 % der PKWs nicht überschritten?

(c) Bestimmen Sie das Quantil $\tilde{x}_{0,9}$. Was sagt dieses Quantil aus?

Aufgabe 4
An einem bestimmten Tag wird das Gewicht von 30 Versuchstieren gemessen. Man erhält die folgende Urliste (in kg):

$12,16$	$11,53$	$14,02$	$11,85$	$10,94$	$11,83$	$12,94$	$11,46$	$13,15$	$12,70$
$10,88$	$13,24$	$14,04$	$10,95$	$14,78$	$12,39$	$13,69$	$11,82$	$14,28$	$12,96$
$13,24$	$13,42$	$12,23$	$15,04$	$11,34$	$12,28$	$13,42$	$13,93$	$14,73$	$11,28$

(a) Erstellen Sie eine Tabelle, in welcher Sie die Messwerte in Klassen (10 kg bis unter 11 kg, 11 kg bis unter 12 kg, usw.) einteilen. Bestimmen Sie die absoluten und relativen Häufigkeiten der Messwerte in den Klassen.

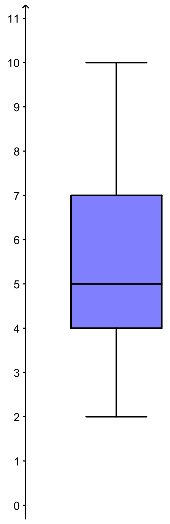
(b) Erstellen Sie (von Hand oder mit GeoGebra, Excel, usw.) ein Histogramm mit den absoluten Häufigkeiten.

Zusatzaufgaben

Die Zusatzaufgaben werden (in der Regel) nicht in den Übungsstunden besprochen. Bei den Ergebnissen finden Sie deshalb ausführliche Lösungen. Teilweise sind diese Aufgaben schwieriger als die Aufgaben für die Übungsstunden.

Aufgabe 5

Geben Sie (mindestens) zwei verschiedene Datensätze an, welche den abgebildeten Boxplot haben. Finden Sie insbesondere einen Datensatz mit 8 Daten.



Aufgabe 6

Fünfzehn befragte Personen gaben ihre monatlichen Ausgaben in CHF wie folgt an:

300	950	750	1200	250	4000	300	800
700	5000	900	1400	1450	300	1300	

(a) Berechnen Sie das arithmetische Mittel $\overline{x}$ und den Median $\tilde{x}$.

(b) Welche Masszahl repräsentiert die angegebenen Ausgaben besser, $\overline{x}$ oder $\tilde{x}$?

(c) Bestimmen Sie das Quantil $\tilde{x}_{0,4}$. Welche Bedeutung hat dieses Quantil?

Aufgabe 7

Seien $x_1, \ldots, x_n$ gegebene Zahlen und $\overline{x}$ ihr Mittelwert. Beweisen Sie, dass die Funktion

$$f(x) = \sum_{i=1}^{n} (x_i - x)^2$$

ein Minimum in $x = \overline{x}$ hat (vgl. Skript, Seite 6).

Lösungshinweise

Aufgabe 1

Mittelwert $\overline{x}$ gemäss Definition auf Seite 5 des Skripts; für den Median $\tilde{x}$, sowie die Quartile $\tilde{x}_{0,25}$ und $\tilde{x}_{0,75}$ müssen die Zahlen zuerst der Grösse nach geordnet werden (für die Berechnung der Quartile kann der Satz 1.1 (Seite 9) für $\alpha = 0,25$, bzw. $\alpha = 0,75$ benutzt werden). Beim Boxplot aufpassen auf Ausreisser.

Aufgabe 2

Wie gross muss die Summe der Zahlen sein, damit $\overline{x} = 7$?

Aufgabe 3

(a) Die Berechnung des arithmetischen Mittels, des Medians und der Quartile erfolgt wie in Aufgabe 1. Die Standardabweichung s gemäss Definition auf Seite 13 des Skripts.

(b) Dies entspricht dem dritten Quartil $\tilde{x}_{0,75}$.

(c) Satz 1.1 anwenden mit $n = 16$ und $\alpha = 0,9$. Die Bedeutung steht in der Definition, Seite 9.

Aufgabe 4

Wie im Beispiel auf Seite 5 des Skripts.

Aufgabe 5 (Zusatzaufgabe)

Suchen Sie zunächst nach einem Datensatz mit 5 Daten.

Aufgabe 6 (Zusatzaufgabe)

(b) Denken Sie an das Beispiel auf Seite 7 des Skripts.

(c) Zur Bestimmung von $\tilde{x}_{0,4}$ benutzt man am besten Satz 1.1 (Seite 9) und zur Interpretation die Definition (Seite 9).

Aufgabe 7 (Zusatzaufgabe)

Aus der Bedingung $f'(x) = 0$ an eine Extremalstelle folgt $x = \overline{x}$. Warum ist $\overline{x}$ ein Minimum und nicht ein Maximum?

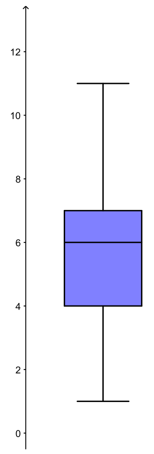
Ergebnisse

Aufgabe 1

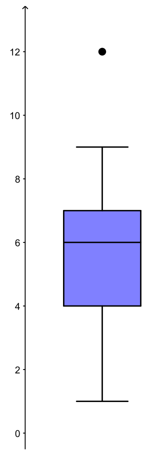
Mittelwerte: (a) $\overline{x} = 5,71$ (b) $\overline{x} = 5,77$

Der Median und die Quartile sind für (a) und (b) gleich: $\tilde{x} = 6$, $\tilde{x}_{0{,}25} = 4$, $\tilde{x}_{0{,}75} = 7$

Der Interquartilsabstand ist $\tilde{x}_{0{,}75} - \tilde{x}_{0{,}25} = 3$. Ausreisser und Extremwerte sind deshalb Zahlen $< 4 - 3 \cdot 1,5 = -0,5$ und $> 7 + 3 \cdot 1,5 = 11,5$. Die Zahl 12 im Datensatz (b) ist also ein Ausreisser, der Datensatz (a) hat keine Ausreisser.



(a)



(b)

Aufgabe 2

Eine der beiden grössten Zahlen muss um 4 erhöht werden, damit $\overline{x} = 7$. Also ersetzt man entweder die (grösste) Zahl 14 durch 18 oder die (zweitgrösste) Zahl 10 durch 14.

Aufgabe 3

(a) $\overline{x} = 10,49$; $\tilde{x} = 10,4$; $\tilde{x}_{0{,}25} = \frac{1}{2}(9,9 + 10,1) = 10,0$; $\tilde{x}_{0{,}75} = \frac{1}{2}(10,9 + 11,3) = 11,1$; $s = 0,76$

(b) Als Antwort kann jede Zahl im Intervall $[10{,}9\,;\,11{,}3)$ angegeben werden. Dies entspricht dem dritten Quartil $\tilde{x}_{0{,}75}$ (wobei es hier keinen Grund gibt, den Mittelwert der Intervallgrenzen zu wählen, die Angabe der Zahl 10,9 ist sinnvoller).

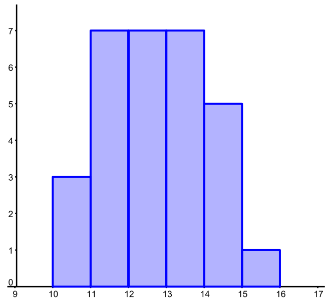
(c) $\tilde{x}_{0{,}9} = 11,5$ (die 15. geordnete Zahl)
Höchstens 90 % der PKWs haben einen Benzinverbrauch $< \tilde{x}_{0{,}9}$ und höchstens 10 % der PKWs haben einen Benzinverbrauch $> \tilde{x}_{0{,}9}$.

Aufgabe 4

(a)

<table>
<tr><th rowspan="2">j</th><th rowspan="2">Klasse j
in cm</th><th rowspan="2">Strichliste</th><th colspan="2">Häufigkeiten</th></tr>
<tr><th>absolut h_j</th><th>relativ f_j</th></tr>
<tr><td>1</td><td>$[10, 11)$</td><td>|||</td><td>3</td><td>$0,1$</td></tr>
<tr><td>2</td><td>$[11, 12)$</td><td>||||| ||</td><td>7</td><td>$0,2\overline{3}$</td></tr>
<tr><td>3</td><td>$[12, 13)$</td><td>||||| ||</td><td>7</td><td>$0,2\overline{3}$</td></tr>
<tr><td>4</td><td>$[13, 14)$</td><td>||||| ||</td><td>7</td><td>$0,2\overline{3}$</td></tr>
<tr><td>5</td><td>$[14, 15)$</td><td>|||||</td><td>5</td><td>$0,1\overline{6}$</td></tr>
<tr><td>6</td><td>$[15, 16)$</td><td>|</td><td>1</td><td>$0,0\overline{3}$</td></tr>
<tr><td>Summe</td><td></td><td></td><td>30</td><td>1</td></tr>
</table>

(b) Mit GeoGebra:



Aufgabe 5 (Zusatzaufgabe)
Dem Boxplot entnehmen wir: $\tilde{x} = 5$, $\tilde{x}_{0,25} = 4$, $\tilde{x}_{0,75} = 7$, $x_{\text{min}} = 2$, $x_{\text{max}} = 10$. Der Boxplot passt also zum Beispiel zum Datensatz

2, 4, 5, 7, 10

Diesen Datensatz kann man zum Beispiel um 4 weitere Daten ergänzen:

2, 3, 4, 4, 5, 6, 7, 8, 10

Bei einem Datensatz mit 8 Daten liegen der Median und die Quartile jeweils zwischen je zwei Daten. Beispiel:

2, 4, 4, 4, 6, 6, 8, 10

Aufgabe 6 (Zusatzaufgabe)

(a) $\overline{x} = 1307$, $\tilde{x} = 900$

(b) Der Median $\tilde{x}$ repräsentiert die Ausgaben besser, da der Ausreisser 4000 und der Extremwert 5000 das arithmetische Mittel $\overline{x}$ zu stark erhöhen.

(c) Mit $\alpha = 0,4$ und $n = 15$ ist $n\alpha = 15 \cdot 0,4 = 6$. Nach Satz 1.1 ist $\tilde{x}_{0,4} = \frac{1}{2}(750+800) = 775$. Dies bedeutet, dass höchstens 40 % (bzw. ein Anteil von 0, 4) der befragten Personen weniger als 775 CHF und höchstens 60 % mehr als 775 CHF monatlich ausgeben (tatsächlich sind es genau 40 %, bzw. 60 %).

Aufgabe 7 (Zusatzaufgabe)
Es gilt

$$0 = f'(x) = \sum_{i=1}^{n} 2(x_i - x) \cdot (-1) \,.$$

Division durch -2 ergibt

$$0 = \sum_{i=1}^{n} (x_i - x) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x = \sum_{i=1}^{n} x_i - nx$$

und wir erhalten

$$x = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x} \,.$$

Also ist $x = \overline{x}$ eine Extremalstelle von f.

Da f eine quadratische Funktion mit positivem Leitkoeffizienten n ist, ist der Graph von f eine nach oben geöffnete Parabel, so dass die Extremalstelle von f eine Minimalstelle ist.

Oder man berechnet die zweite Ableitung $f''(x) = 2\sum_{i=1}^{n} 1 = 2n$. Da sie positiv für $x = \overline{x}$ (bzw. für alle x) ist, ist $\overline{x}$ eine Minimalstelle.