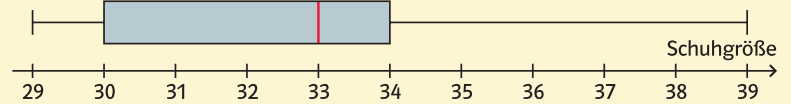


## 4 Boxplots



29 - 29 - 31 - 33 - 33 - 33 - 33 - 35 - 39

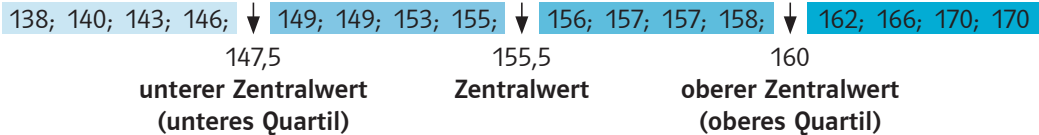
Marina und Peter haben die Schuhgrößen ihrer Freunde notiert. Die Angaben sind in der Liste sortiert. Im Diagramm ist eine andere Darstellung gewählt. Vergleiche und erkläre.



Mittelwerte reichen zur genauen Beschreibung von Zahlenlisten („Daten“) nicht immer aus. Oft muss man zusätzlich wissen, wie die Daten um den Mittelwert verstreut liegen und wo ein einzelner Wert im Vergleich zu den anderen liegt. Diese Information kann man **Boxplots** entnehmen. Wie man sie erstellt, wird an einem Beispiel erläutert.

Maika und ihre 15 Mitschülerinnen messen ihre Körpergrößen (in cm). Es ergibt sich die geordnete Zahlenliste 138, 140, 143, 146, 149, 149, 153, 155, 156, 157, 157, 158, 162, 166, 170, 170.

Maika ist 146 cm groß. Ihre Größe liegt damit unterhalb des Zentralwertes von 155,5 cm; sie gehört zu „den Kleineren“. Um beurteilen zu können, ob sie besonders klein ist, zerlegt man die Gesamtheit der Körpergrößen in vier Abschnitte. Dazu bestimmt man auch für die untere und die obere Datenhälfte die Zentralwerte. Sie liegen bei 147,5 cm und 160 cm. Wie die Darstellung zeigt, liegt Maikas Größe damit im unteren Viertel:



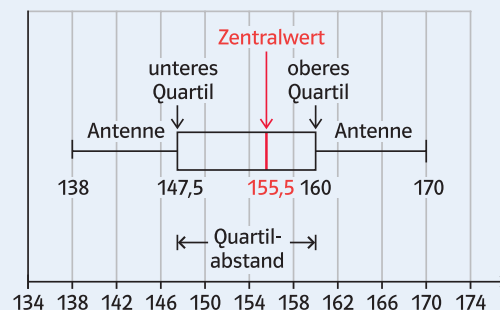
← unteres Viertel → ← Mitte → → oberes Viertel →

Mithilfe der Quartile kann man die Daten durch einen Boxplot übersichtlich darstellen:

- Markiere auf einer Skala die Quartile und zeichne dazwischen ein Rechteck (die „Box“).
- Markiere im Innern der Box den Zentralwert durch einen Strich.
- Verlängere die Box durch zwei „Antennen“ bis zum größten und kleinsten Datenwert.

### Boxplot

- In der Box liegt ungefähr die Hälfte aller Daten, im Bereich einer jeden Antenne ungefähr je ein Viertel.
- Die Länge der Box heißt Quartilabstand. Je kürzer die Box ist, desto weniger sind die Daten um den Zentralwert verstreut.



**Beispiel 1 Boxplots und Balkendiagramme**

Fig. 1 zeigt die Einnahmen der Klasse 5 c bei einem Sponsorenlauf als Balkendiagramm. Zeichne den zugehörigen Boxplot und vergleiche mit dem Balkendiagramm.

**Lösung**

Man entnimmt dem Balkendiagramm die Einnahmen der insgesamt 22 Schüler sortiert:

13 14 15 15 16 17 17 17 17 18 18 18 18 18 19 19 20 21 21 21 25

Der Zentralwert ist 18, die Quartile sind 17 und 19. Damit ergibt sich der Boxplot aus Fig. 2. Dem Säulendiagramm kann man alle Informationen zu den Einnahmen der Schüler entnehmen. So erkennt man, dass die Einnahmen 17€ oder 18€ am häufigsten auftreten. Bei beiden Darstellungen erkennt man, dass die Einnahmen insgesamt zwischen 13€ und 25€ liegen. Der Boxplot ist übersichtlicher und fasst die Daten zusammen. Man erkennt dort sofort, dass etwa die Hälfte der Schüler zwischen 17€ und 19€ bekommt.

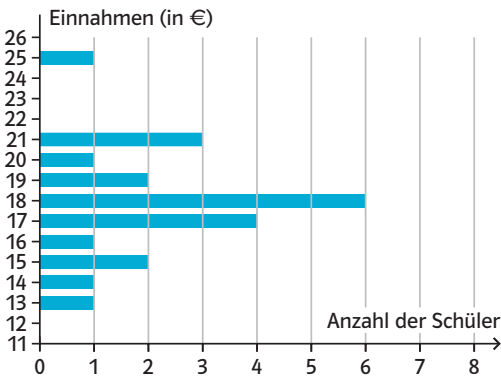


Fig. 1

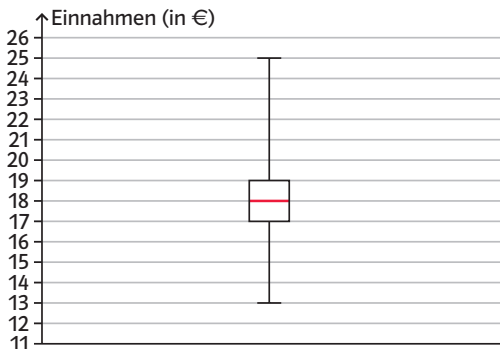


Fig. 2

**Beispiel 2 Reaktionszeiten**

Mark nimmt an einem Reaktionstest teil. Wenn aus seinem Computer ein Signal ertönt, muss er auf „Stopp“ klicken. Die benötigte Zeit (in ms) wird gestoppt. Dies wird 400-mal wiederholt. Fig. 4 zeigt die ersten und die letzten 20 Reaktionszeiten.

- a) Veranschauliche die ersten 20 Reaktionszeiten durch einen Boxplot.
- b) Vergleiche mit einem Boxplot zu den letzten 20 Zeiten.

**Lösung**

a) Man sortiert die 20 Versuchszeiten.

589,3; 590,0; 604,0; 617,1; 617,2;  
 618,8; 631,6; 632,2; 645,0; 646,5;  
 646,8; 661,5; 673,3; 675,3; 685,9;  
 686,5; 687,7; 688,4; 698,9; 700,8

Dann bestimmt man den Zentralwert zu 646,65, das untere Quartil zu 618 und das obere Quartil zu 686,2. Anschließend verlängert man die hierdurch festgelegte Box durch die Antennen bis zum kleinsten bzw. zum größten Wert (Fig. 3 links).

b) Der Zentralwert der Reaktionszeiten hat sich von 646,75 auf 590,73 verkleinert. Mark reagiert also deutlich schneller. Außerdem ist die Box kürzer geworden (ca. 30 statt 68); die Reaktionszeiten streuen deutlich weniger um den Zentralwert. Mark hat während der Testdurchführung seine Reaktionszeit deutlich verbessert.

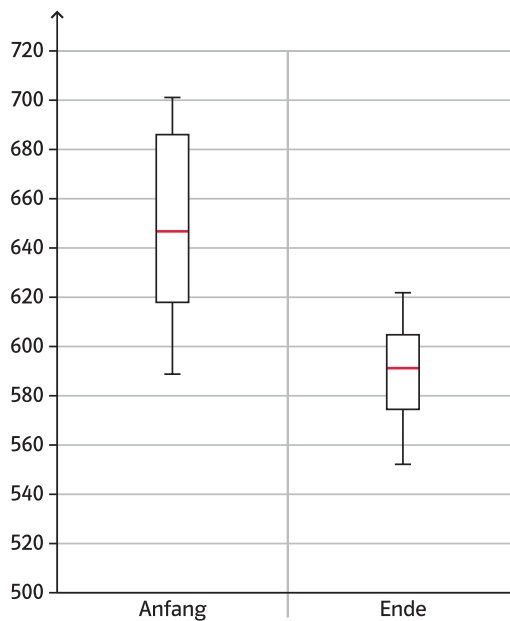


Fig. 3

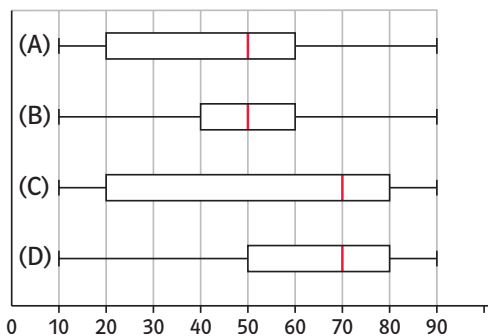
	erste 20 Zeiten	letzte 20 Zeiten
1	698,9	564,0
2	685,9	593,3
3	673,3	607,3
4	700,8	567,3
5	675,3	552,1
6	604,0	620,7
7	661,5	596,4
8	646,8	576,0
9	617,1	553,0
10	631,6	621,7
11	687,7	594,3
12	645,0	578,3
13	686,5	604,0
14	646,5	569,8
15	589,3	607,3
16	647,2	607,1
17	590,0	588,2
18	688,4	579,8
19	618,8	581,0
20	632,2	603,3

Fig. 4

## Aufgaben

1 Zu den vier Datenreihen sind die Boxplots gezeichnet worden. Sortiere, ordne zu und begründe deine Entscheidung.

- (1) 15, 40, 10, 60, 80, 75, 70, 70, 65, 80, 70, 90, 80  
 (2) 20, 30, 50, 50, 70, 10, 10, 20, 60, 60, 90, 20, 50  
 (3) 10, 40, 40, 50, 55, 65, 30, 80, 50, 90, 55, 50, 50  
 (4) 40, 10, 30, 85, 70, 85, 90, 15, 15, 25, 75, 70, 75



2 Zeichne Boxplots zu den Zahlenlisten.

a) 2, 17, 11, 34, 35, 21, 40, 5

b) 42, 56, 63, 34, 21, 65, 60, 99, 70, 66, 61

3 Nenne drei Zahlenreihen,

a) mit beliebiger Länge, die zum Boxplot in Fig. 1 passen.

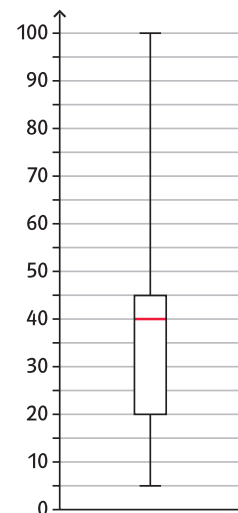


Fig. 1

4 Acht Schüler haben ihren Pulsschlag 15 s lang gezählt. Zuerst haben sie ihren Ruhepuls ermittelt, dann jeweils den Puls nach 10, 20 bzw. 30 Kniebeugen.

a) Veranschauliche die Daten der vier Spalten durch je einen Boxplot.

b) Untersuche, ob der Puls nach jeweils 10 Kniebeugen durchschnittlich um den gleichen Wert anwächst.

c) Führe dieses Experiment selbst durch und fasse die „am eigenen Leib gemachten“ Beobachtungen zusammen.

b) mit der Länge 15,

	Anzahl an Pulsschlägen in 15 s			
	in Ruhe	nach 10 Kniebeugen	nach 20 Kniebeugen	nach 30 Kniebeugen
Dennis	19	30	39	42
Oliver	22	35	39	41
Michael	21	34	38	40
Stefan	16	25	28	30
Tobi	18	32	33	35
Sven	19	28	32	35
Marc	23	35	40	40
Thorsten	29	35	36	40

5 Beim Sportfest der Geschwister-Scholl-Schule haben die Klassenstufen 5 und 6 einen Weitsprung-Wettbewerb veranstaltet. Der Boxplot in Fig. 2 veranschaulicht die Sprungweiten (in m), die erzielt wurden. Lies aus dem Boxplot möglichst viele Informationen ab.

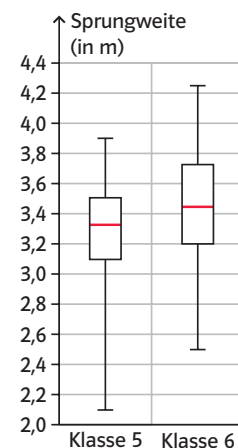


Fig. 2

6 Lance Armstrong hat 2005 die Tour de France gewonnen. Für 3391,1 km brauchte er 86 Stunden, 15 Minuten und 2 Sekunden. Fig. 3 zeigt, um wie viele Minuten die übrigen 144 Rennfahrer am Ende hinter Lance Armstrong lagen. Iker Flores wurde Letzter.

a) Wo findet man im Boxplot Lance Armstrong und Iker Flores wieder? Wo liegt Jan Ullrich, der 4 Minuten Rückstand auf Armstrong hatte?

b) Lies aus dem Boxplot möglichst viele weitere Informationen ab.

c) Wie würde der Boxplot aussehen, wenn außer Armstrong und Flores alle anderen etwa gleich lange gebraucht hätten?

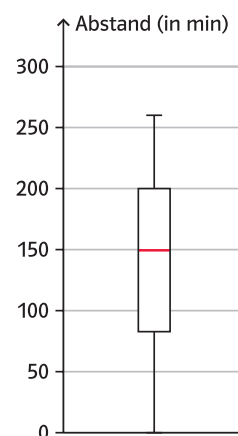


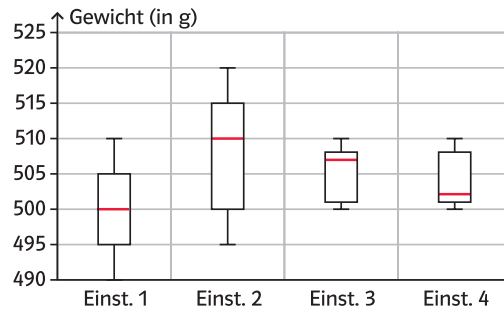
Fig. 3

**Bist du schon sicher?**

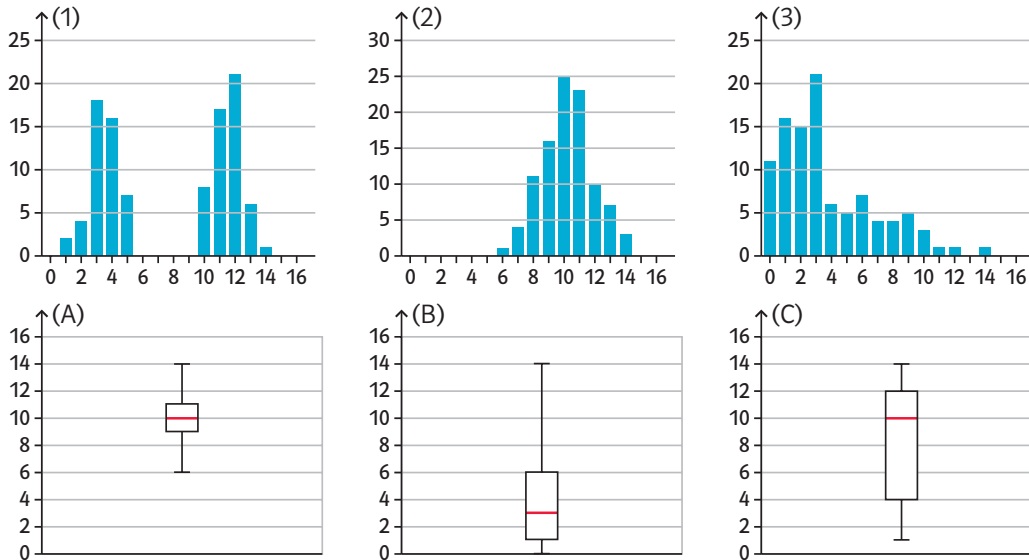
- 7 Zeichne einen Boxplot zu der Zahlenliste 1, 3, 4, 4, 4, 5, 6, 6, 6, 6, 7, 8, 16, 20. Muss der Zentralwert immer im Innern der Box liegen? Begründe. Kann es Boxplots „ohne Antennen“ geben? Begründe.

→ Lösung | Seite 208

- 8 Eine Maschine verpackt Pralinenpackungen zu 500 g. Vier verschiedene Einstellungen lieferten bei Kontrollen des Verpackungsgewichts die Boxplots.
- Welche Einstellung ist besonders günstig für die Pralinenfirma, welche für den Kunden?
  - Wie müsste der Boxplot einer für den Produzenten idealen Verpackungsmaschine aussehen?

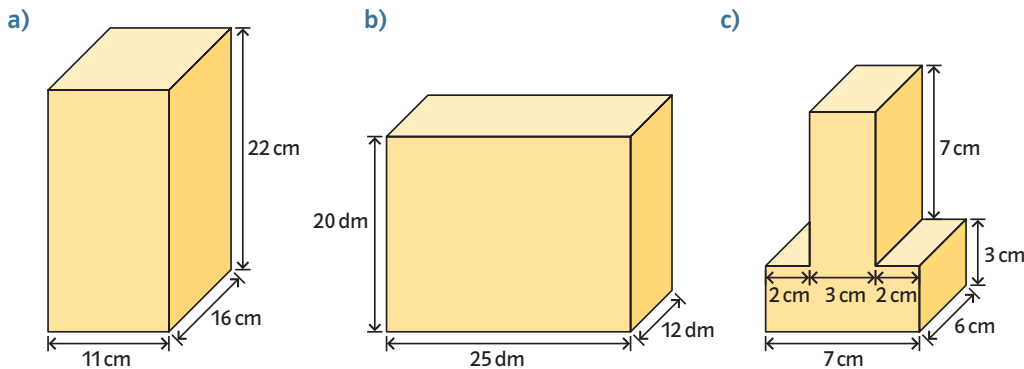


- 9 Bei einem Test gab es maximal 16 Punkte. Die Grafiken zeigen die Ergebnisse in drei Gruppen.
- Beschreibe in Worten, wie der Test ausgefallen ist.
  - Welcher Boxplot gehört zu welchem Säulendiagramm? Begründe deine Antwort.



**Kannst du das noch?**

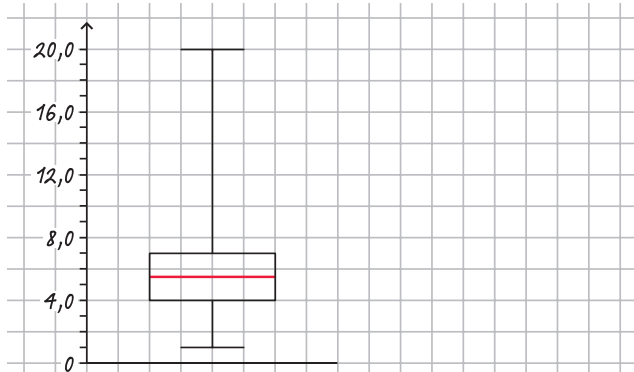
- 10 Berechne das Volumen und den Oberflächeninhalt.



→ Lösung | Seite 208

Kapitel VI, Bist du schon sicher?, Seite 173

7



Der Zentralwert liegt immer innerhalb der Box. Dies ergibt sich aus dem allgemeinen Aufbau des Boxplots. Es kann Boxplots ohne Antennen geben, wenn die Daten nur aus einem einzigen Kennwert bestehen.

Kapitel VI, Kannst du das noch?, Seite 173

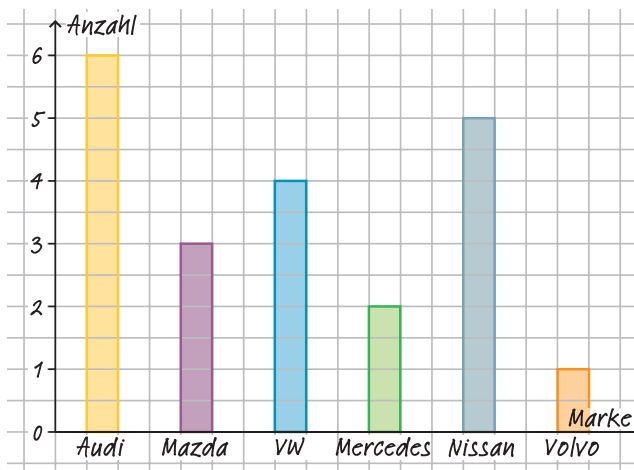
10

- a)  $V = 11\text{ cm} \cdot 16\text{ cm} \cdot 22\text{ cm} = 3872\text{ cm}^3$   
 $O = 2 \cdot (11\text{ cm} \cdot 16\text{ cm} + 16\text{ cm} \cdot 22\text{ cm} + 22\text{ cm} \cdot 11\text{ cm}) = 1540\text{ cm}^2$
- b)  $V = 25\text{ dm} \cdot 12\text{ dm} \cdot 20\text{ dm} = 6000\text{ dm}^3$   
 $O = 2 \cdot (25\text{ dm} \cdot 20\text{ dm} + 25\text{ dm} \cdot 12\text{ dm} + 12\text{ dm} \cdot 20\text{ dm}) = 2080\text{ dm}^2$
- c)  $V = 7\text{ cm} \cdot 6\text{ cm} \cdot 3\text{ cm} + 3\text{ cm} \cdot 6\text{ cm} \cdot 7\text{ cm} = 252\text{ cm}^3$   
 $O = 2 \cdot (7\text{ cm} \cdot 3\text{ cm} + 3\text{ cm} \cdot 7\text{ cm} + 6\text{ cm} \cdot 3\text{ cm} + 2\text{ cm} \cdot 6\text{ cm} + 6\text{ cm} \cdot 7\text{ cm}) + 3\text{ cm} \cdot 6\text{ cm} + 7\text{ cm} \cdot 6\text{ cm} = 288\text{ cm}^2$

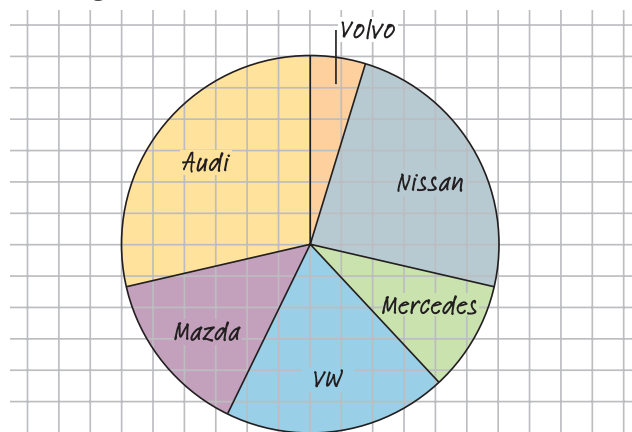
Kapitel VI, Training Runde 1, Seite 179

1

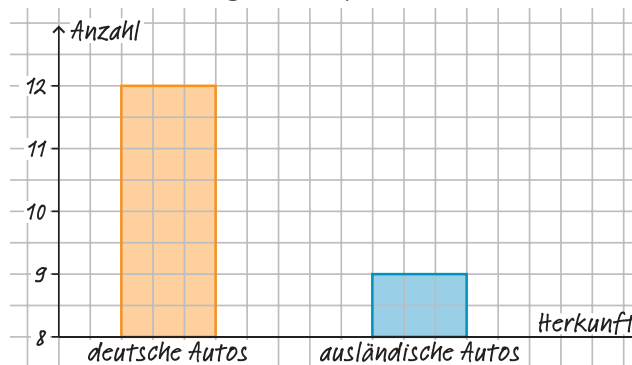
a) Säulendiagramm



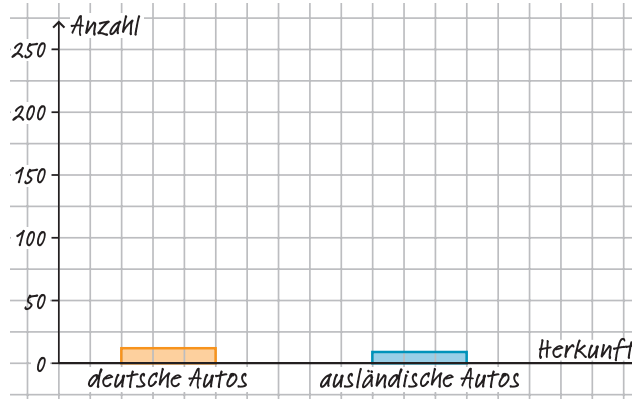
Kreisdiagramm



b) individuelle Lösung, zum Beispiel:



c) individuelle Lösung, zum Beispiel:



2

- a) Zentralwert für Deutsch: 2  
 Mittelwert für Deutsch: 2,6  
 Zentralwert für Englisch: 3  
 Mittelwert für Englisch: 2,6  
 Für Lea wäre in Deutsch der Zentralwert günstiger, in Englisch der Mittelwert.
- b) jeweils die Note 3

### 4 Boxplots

Seite 170

**Einstiegsaufgabe**

Die obere Grafik entspricht einer ungeordneten Liste; die mittlere Grafik zeigt die zugehörige geordnete Liste. Die ungeordnete Liste zeigt die gesammelten Daten. Die geordnete Liste zeigt den niedrigsten und höchsten Wert auf einen Blick; außerdem ermöglicht sie recht einfach die Berechnung des Zentralwertes.

Die untere Grafik veranschaulicht

- den kleinsten und den höchsten Wert und die dazwischenliegende Spannweite,
- den mittleren Bereich, in dem sich etwa 50% der Werte sammeln (hier sind 5 von den insgesamt 9 Werten enthalten),
- den Wert, der in der Mitte aller Messergebnisse liegt - den Zentralwert (hier 33),
- den oberen und unteren Bereich der Ergebnisse.

Die untere Grafik ermöglicht, Messergebnisse zu strukturieren und Aussagen über ihre Streuung bzw. die Häufigkeit ihres Auftretens zu treffen.

Seite 172

**1** Es ist hilfreich, als Erstes die Listen zu ordnen; dann ist es leicht, die Kennwerte zu erkennen.

- (1): 10, 15, 40, 60, 65, 70, 70, 75, 80, 80, 90
- (2): 10, 10, 20, 20, 20, 30, 50, 50, 50, 60, 60, 70, 90
- (3): 10, 30, 40, 40, 50, 50, 50, 50, 55, 55, 65, 80, 90
- (4): 10, 15, 15, 25, 30, 40, 70, 70, 75, 75, 85, 85, 90

Alle Listen haben 10 als kleinsten Wert und 90 als größten Wert.

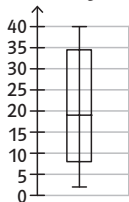
Boxplot (A) gehört zu (2) (Zentralwert: 50, unteres Quartil: 20, oberes Quartil: 60).

Boxplot (B) gehört zu (3) (Zentralwert: 50, unteres Quartil: 40, oberes Quartil: 60).

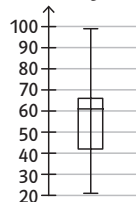
Boxplot (C) gehört zu (4) (Zentralwert: 70, unteres Quartil: 20, oberes Quartil: 80).

Boxplot (D) gehört zu (1) (Zentralwert: 70, unteres Quartil: 50, oberes Quartil: 80).

**2 a)** kleinster Wert: 2  
 größter Wert: 40  
 Zentralwert: 19  
 unteres Quartil: 8  
 oberes Quartil: 34,5



**b)** kleinster Wert: 21  
 größter Wert: 99  
 Zentralwert: 61  
 unteres Quartil: 42  
 oberes Quartil: 66



**3** kleinster Wert: 5, größter Wert: 100, Zentralwert: 40, unteres Quartil: 20, oberes Quartil: 45

a) individuelle Lösung, z. B.:

5, 10, 20, 20, 35, 45, 45, 45, 60, 100

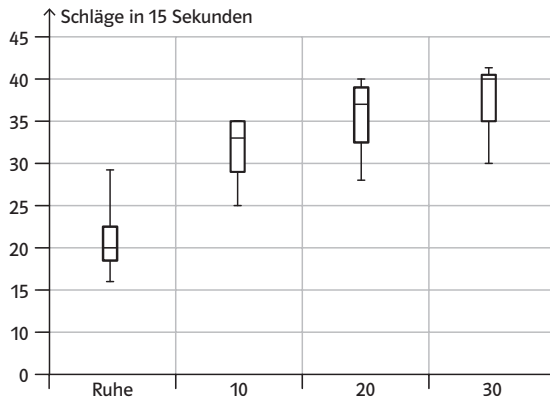
5, 15, 20, 25, 25, 40, 41, 44, 45, 70, 100

b) individuelle Lösung z. B.:

5, 5, 15, 20, 25, 30, 40, 40, 40, 40, 42, 45, 60, 65, 100

5, 10, 20, 20, 30, 30, 30, 40, 41, 42, 42, 45, 60, 70, 100

**4 a)**



b) Der Pulsschlag nimmt unterschiedlich zu. Nach 10 Kniebeugen steigt er durchschnittlich um 10,9 Schläge, nach 20 Kniebeugen um 3,9 Schläge und nach 30 Kniebeugen um 2,3 Schläge.

c) individuelle Lösung

**5**

- Die Sechstklässler springen im Mittel weiter als die Fünftklässler.
- Es gibt „viele“ Fünftklässler, die weiter springen als Sechstklässler.
- Der Hauptteil der Fünftklässler springt zwischen 3,10m und 3,50m.
- Bei den Sechstklässlern liegt der Hauptteil zwischen 3,20m und 3,75m.
- Die Sechstklässler springen zwischen 2,50m und 4,25m weit, die Fünftklässler zwischen 2,10m und 3,90m.
- Der Zentralwert beträgt bei den fünften Klassen ca. 3,3m, bei den sechsten Klassen ca. 3,4m.
- Über die Klassenstärke kann man keine Aussage machen.

**6 a)** Lance Armstrong findet man beim kleinsten Wert, denn er selbst kann auf sich keinen Rückstand haben. Iker Flores findet man beim größten Wert, da er Letzter wurde und sein Abstand zu Lance Armstrong daher am größten war.

Jan Ulrich würde sich mit vier Minuten Rückstand am unteren Ende der unteren Antenne nur kurz über Lance Armstrong befinden.

- b)
- Iker Flores benötigte ca. 260 Minuten mehr als Lance Armstrong.
  - Der Zentralwert liegt 150 Minuten über der Zeit von Lance Armstrong.
  - Die Hälfte der Fahrer benötigte zwischen 80 und 200 Minuten mehr als Lance Armstrong. (Das Mittelfeld erstreckte sich über 120 Minuten.)
- c) Die Box würde sehr klein werden, denn die Rückstände der meisten Fahrer lägen in einem sehr kleinen Zeitabschnitt. Der kleinste und der größte Wert (und somit auch die Antennen) würden sich jedoch nicht ändern.

Seite 173

- 8 a) Die Einstellungen 1 bzw. 4 sind für die Pralinenfirma am günstigsten.  
Bei Einstellung 1 sind die Packungen am leichtesten, was günstig für den Erlös der Firma ist. Es könnte aber Reklamationen wegen zu geringem Verpackungsgewicht geben. Möchte die Pralinenfirma dies ausschließen, so sollte sie Einstellung 4 wählen.  
Die Einstellungen 2 bzw. 3 sind günstiger für den Kunden. Bei Einstellung 2 können die Kunden am meisten Ware bekommen. Hier ist aber auch das Risiko relativ groß, zu wenig Ware zu erhalten. Soll dies ausgeschlossen werden, so bietet sich Einstellung 3 an.
- b) Bei einer idealen Maschine wären alle Packungen genau 500g schwer. Der Boxplot hätte dann keine Antennen und die Box hätte die Länge 0.

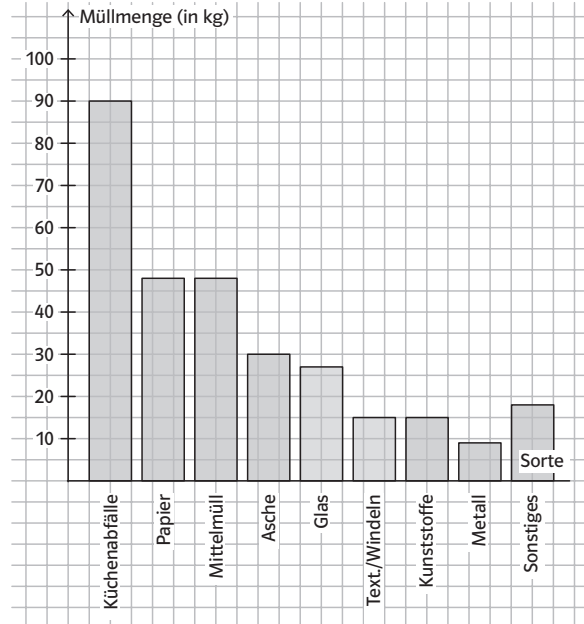
- 9 a) Säulendiagramm (1)  
In dieser Gruppe gab es keine Teilnehmer, die im Test zwischen 6 und 9 Punkte geschrieben haben. Die Ergebnisse teilen sich in fast gleich große Gruppen auf, von denen eine zwischen 1 und 5 Punkten (mit größter Häufigkeit von 3 Punkten) und die andere zwischen 10 und 14 Punkten (mit größter Häufigkeit von 12 Punkten) liegt. Das Maximum liegt bei 14 Punkten, das Minimum bei einem Punkt.  
Säulendiagramm (2)  
Der Test ist in dieser Gruppe am besten ausgefallen, da das arithmetische Mittel am höchsten ist. Keiner der Teilnehmer hat weniger als 6 Punkte in dem Test erhalten. Die höchste erreichte Punktzahl ist 14.  
Säulendiagramm (3)  
Die Punktzahlen, die erreicht wurden, haben eine Spanne von 0 bis 14 Punkte. Das arithmetische Mittel ist gering im Vergleich zu den beiden anderen Gruppen.
- b) Das Säulendiagramm (2) entspricht dem Boxplot (A), das Säulendiagramm (3) entspricht dem Boxplot (B), das Säulendiagramm (1) entspricht dem Boxplot (C) (vergleiche mit den jeweiligen Beschreibungen oben).

Vertiefen und Vernetzen

Seite 174

- 1 a) Mathematik:  
Mittelwert: 2,8, Zentralwert: 3  
Latein:  
Mittelwert: 2,8, Zentralwert: 2
- b) Die Differenzen der Einzelnoten vom Mittelwert sind im Fach Latein erheblich größer als im Fach Mathematik.

2 a)



b) Durch die Darstellung in der Zeitung wird der Eindruck erweckt, der Anteil der Küchenabfälle sei so groß wie alle anderen Anteile zusammen, obwohl er nur 30% beträgt. Grund dafür ist die Darstellung der Daten als Volumina, die dazu führt, dass große Zahlen vergrößert erscheinen, während kleine Zahlen verkleinert erscheinen. Das Säulendiagramm zeigt die tatsächlichen Unterschiede korrekt.

- 3 a) Vermutlich wird das Diagramm von Annika eine geringere Punktzahl erhalten. Das Diagramm gibt an, wie viel die Schultaschen der einzelnen Schüler zum Gesamtgewicht beitragen. Es gibt aber keine Auskunft darüber, wie schwer die Schultaschen im Durchschnitt sind. Deshalb erscheint es zunächst sinnlos.
- b) Das Diagramm von Annika ist sinnvoller, wenn z. B. ein bestimmtes Gesamtgewicht nicht überschritten werden darf oder ein Teil des Gewichts abgewogen werden soll. Dann könnte man z. B. das Gewicht von Uschis Schultasche mit dem Gewicht von Annikas und Jessis Schultasche vergleichen: Das Gewicht von Uschis Schultasche ist ungefähr (da auf ganze Prozent gerundet wurde) genauso schwer wie das Gewicht von Annikas und Jessis Schultaschen zusammen.