

Vorsicht, Boxplots!

Wie aussagekräftig sind Kastenschaubilder?

Dr. Bernhard Salzger

St. Pölten, 17. Feber 2016

Grundlagen

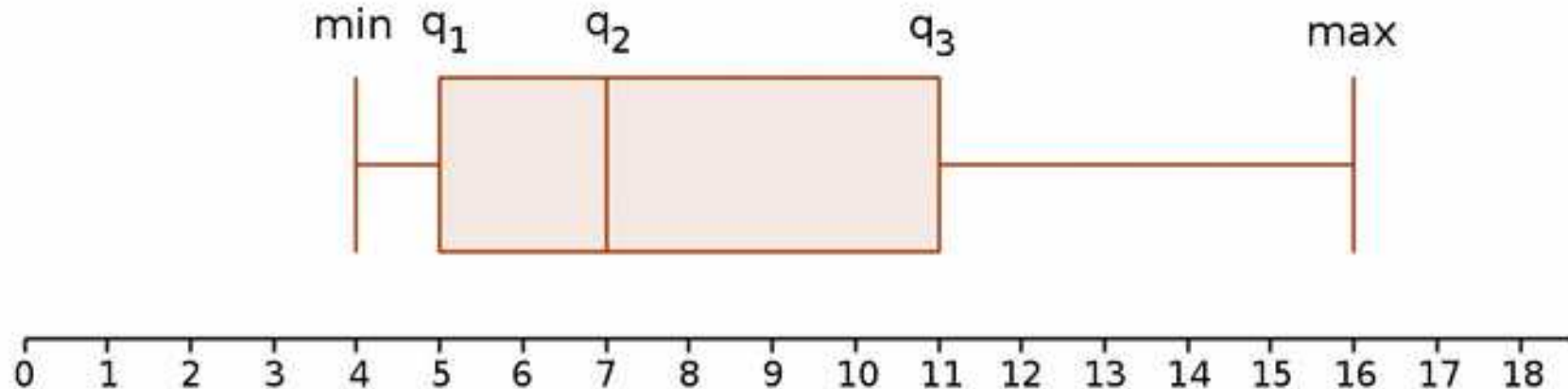
Ein **Kastenschaubild (Boxplot)** ist eine grafische Darstellung einer Zusammenfassung fünf relevanter Kennzahlen (**Fünfhahenzusammenfassung**):

- Minimum (min)
- 1. Quartil (q_1)
- 2. Quartil bzw. Median (q_2)
- 3. Quartil (q_3)
- Maximum (max)

Grafisch erkennbar sind damit zwei weitere Kennzahlen:

- die Spannweite (Differenz: $\max - \min$)
- der Quartilabstand (Differenz: $q_3 - q_1$)

Grundlagen



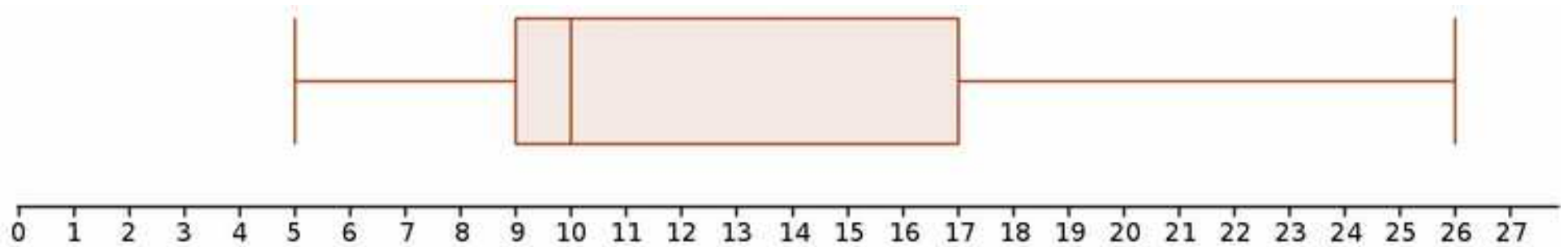
Der Bereich von q_1 bis q_3 , in dem sich etwa 50 % der Daten der geordneten Liste befinden, wird als Rechteck dargestellt.

Darin befindet sich der Median q_2 .

Links von q_1 und rechts von q_3 führen Strecken zum Minimum (min) bzw. zum Maximum (max) der Daten.

Eine Aufgabe am Anfang

Gegeben ist das folgende Kastenschaubild.



Kreuzen Sie die zutreffende(n) Aussage(n) an!

Die Spannweite beträgt 21.

Etwa 50 % der Daten liegen im Intervall $[9; 17]$.

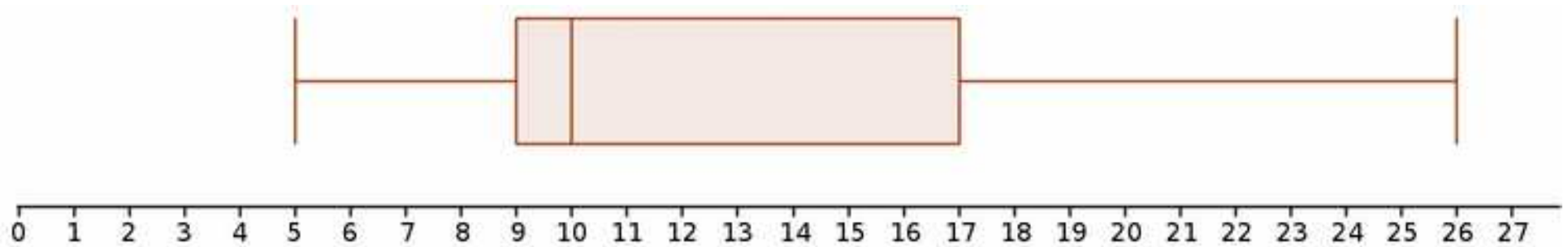
Etwa 75 % der Daten liegen im Intervall $[5; 17]$.

Fast 100 % der Daten liegen im Intervall $[9; 17]$.

Etwa 25 % der Daten sind größer als 17.

Eine Aufgabe am Anfang

Gegeben ist das folgende Kastenschaubild.

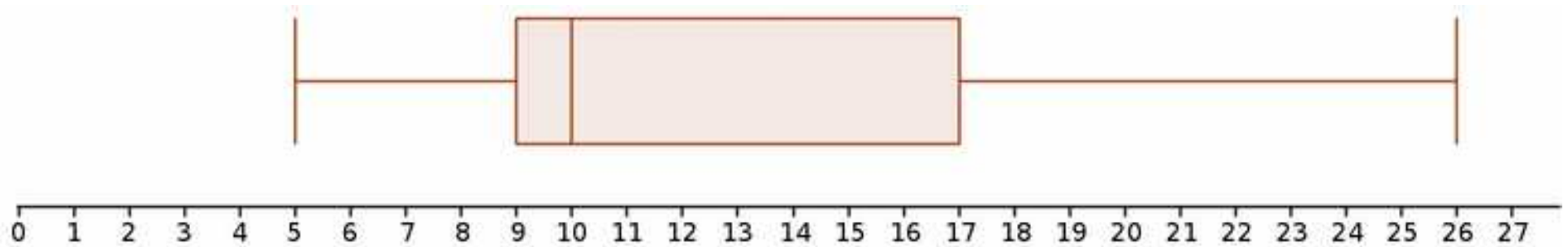


Kreuzen Sie die zutreffende(n) Aussage(n) an!

- Die Spannweite beträgt 21.
- Etwa 50 % der Daten liegen im Intervall $[9; 17]$.
- Etwa 75 % der Daten liegen im Intervall $[5; 17]$.
- Fast 100 % der Daten liegen im Intervall $[9; 17]$.
- Etwa 25 % der Daten sind größer als 17.

Eine Aufgabe am Anfang

Gegeben ist das folgende Kastenschaubild.



Kreuzen Sie die zutreffende(n) Aussage(n) an!

Die Spannweite beträgt 21.



Etwa 50 % der Daten liegen im Intervall $[9; 17]$.



Etwa 75 % der Daten liegen im Intervall $[5; 17]$.



Fast 100 % der Daten liegen im Intervall $[9; 17]$.

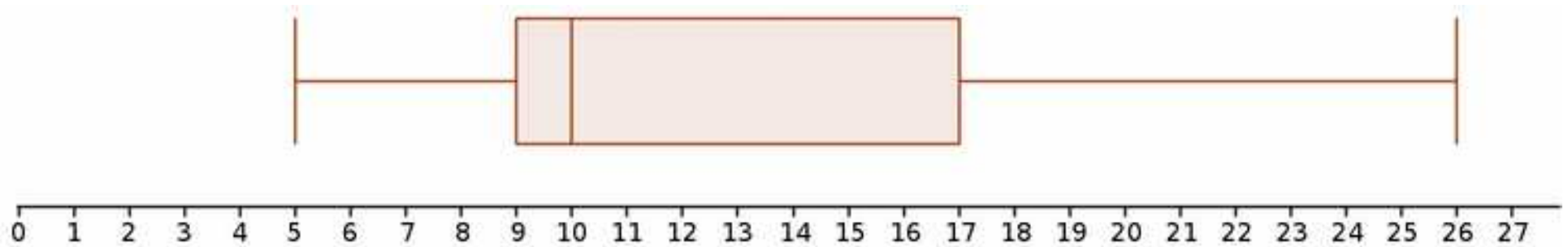


Etwa 25 % der Daten sind größer als 17.



Eine Aufgabe am Anfang

Gegeben ist das folgende Kastenschaubild.



Kreuzen Sie die zutreffende(n) Aussage(n) an!

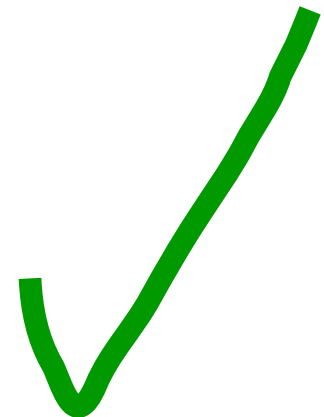
Die Spannweite beträgt 21.

Etwa 50 % der Daten liegen im Intervall $[9; 17]$.

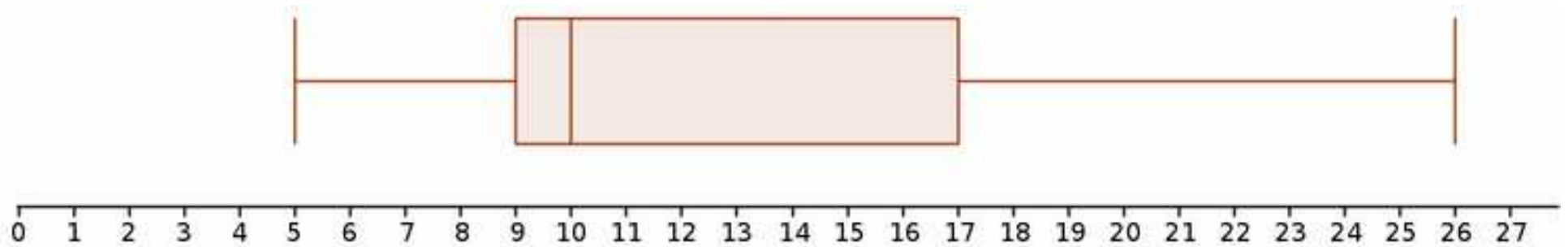
Etwa 75 % der Daten liegen im Intervall $[5; 17]$.

Fast 100 % der Daten liegen im Intervall $[9; 17]$.

Etwa 25 % der Daten sind größer als 17.



Eine Aufgabe am Anfang



Die geordnete Liste für diese Darstellung sieht nämlich so aus:

5, 9, 9, 9, 9, 9, 9, 9, 9, 9, 10, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 26

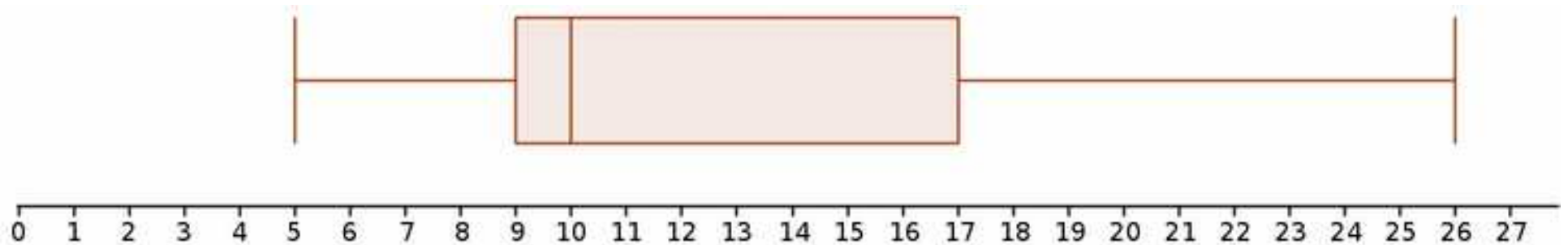
Minimum 1. Quartil Median 3. Quartil Maximum

... zugegeben, die Datenliste

5, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 10, 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 26

wirkt ein bisschen konstruiert ...

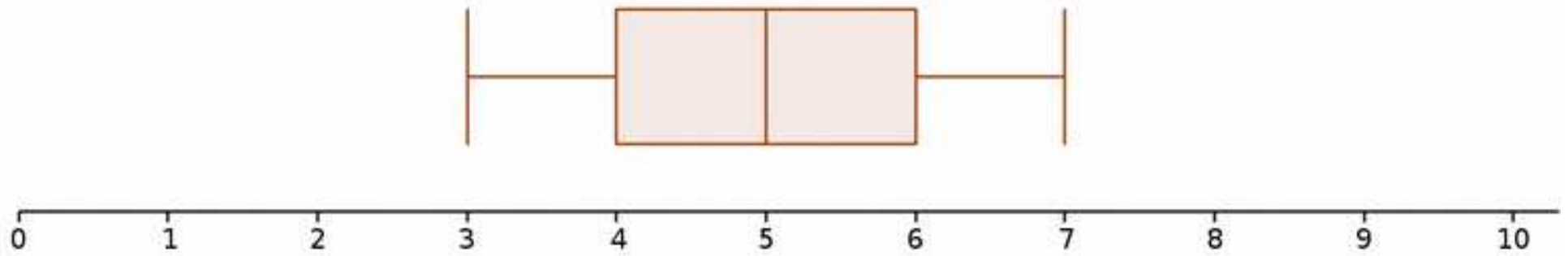
... aber das Kastenschaubild dazu



ist völlig korrekt.

Interpretationsversuche (Fall 1)

Gegeben ist das folgende Kastenschaubild:



Was ist diesem zu entnehmen?

Die geordnete Liste dazu sieht so aus:

3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 7

3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7

3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7

oder so:

oder so:

...

Fazit:

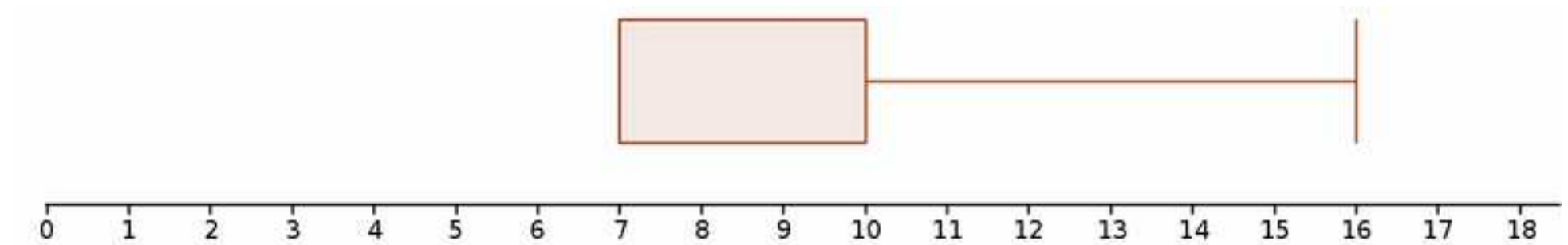
Ein symmetrisch anmutendes Kastenschaubild muss nicht zwangsläufig auf eine symmetrisch aufgebaute geordnete Datenliste verweisen.

Folgerung:

Identischen Kastenschaubildern können völlig unterschiedliche Datenlisten zugrunde liegen, die zufälligerweise dieselben fünf Kennzahlen aufweisen.

Interpretationsversuche (Fall 2)

Gegeben ist das folgende Kastenschaubild:



Was ist diesem zu entnehmen?

Die geordnete Liste dazu sieht so aus:

7, 7, 7, 7, 7, 7, 8, 12, 16

Es fallen also \min , q_1 und q_2 zusammen.

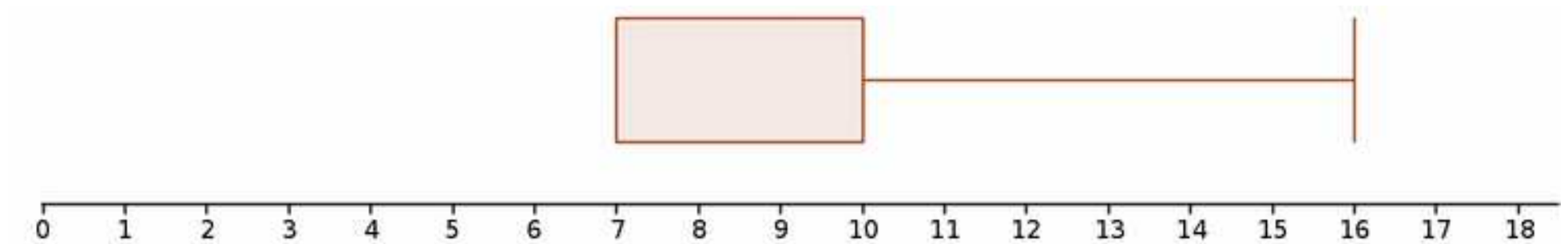
Es ist q_3 das arithmetische Mittel aus 8 und 12.

... zugegeben, die Datenliste

7, 7, 7, 7, 7, 7, 8, 12, 16

ist überschaubar ...

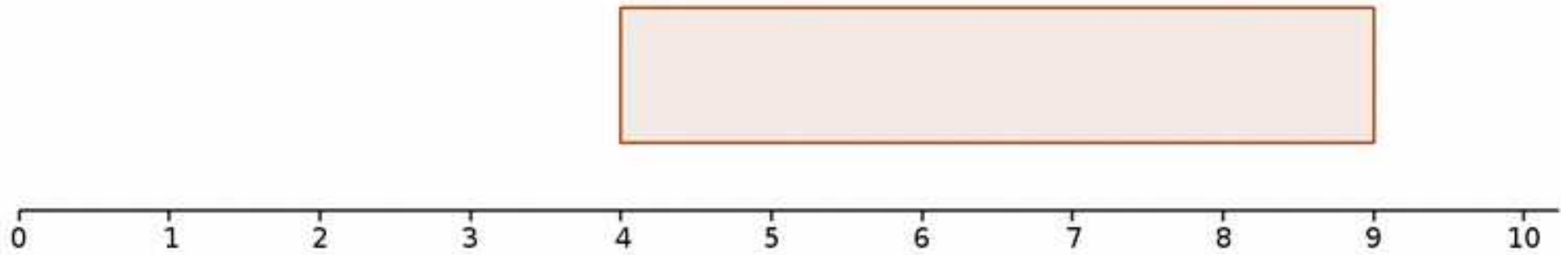
... aber das Kastenschaubild dazu



ist völlig korrekt.

Interpretationsversuche (Fall 3)

Gegeben ist das folgende Kastenschaubild:



Was ist diesem zu entnehmen?

Sicher ist:

100 % der Daten liegen im Intervall $[4; 9]$, dh. $\min = 4$, $\max = 9$.

Alles andere ...

Die geordnete Datenliste sieht so aus:

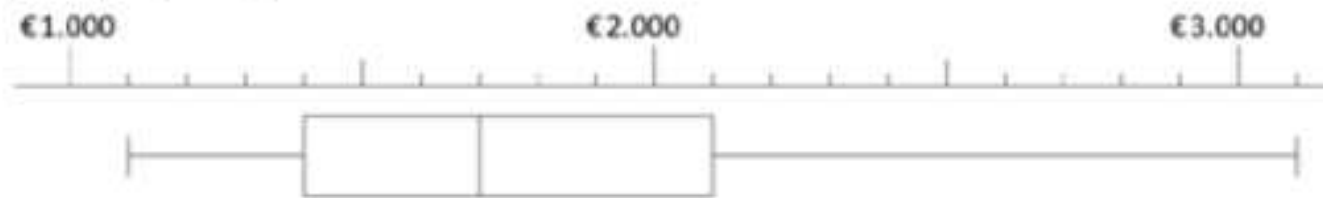
4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 7, 8, 9, 9, 9, 9, 9, 9, 9

Erkenntnis:

Nicht alle Datenlisten eignen sich dazu, als Kastenschaubild veranschaulicht zu werden.

Aufgabe zur Maturavorbereitung (BIFIE)

Die Nettogehälter von 44 Angestellten einer Firmenabteilung werden durch folgendes Kastenschaubild (Boxplot) dargestellt:



Aufgabenstellung:

Kreuzen Sie die beiden zutreffenden Antworten an!

22 Angestellte verdienen mehr als € 2.400.	<input type="checkbox"/>
Drei Viertel der Angestellten verdienen € 2.100 oder mehr.	<input type="checkbox"/>
Ein Viertel aller Angestellten verdient € 1.400 oder weniger.	<input type="checkbox"/>
Es gibt Angestellte, die mehr als € 3.300 verdienen.	<input type="checkbox"/>
Das Nettogehalt der Hälfte aller Angestellten liegt im Bereich [€ 1.400; € 2.100].	<input type="checkbox"/>

Aufgabe zur Maturavorbereitung (BIFIE)

Es ergibt sich:

Keine der angegebenen Antworten ist richtig.

Aber ...

Aufgabe zur Maturavorbereitung (BIFIE)

Lösungsweg

Ein Viertel aller Angestellten verdient € 1.400 oder weniger.	<input checked="" type="checkbox"/>
Das Nettogehalt der Hälfte aller Angestellten liegt im Bereich [€ 1.400; € 2.100].	<input checked="" type="checkbox"/>

Lösungsschlüssel

Ein Punkt ist nur dann zu geben, wenn genau zwei Antworten angekreuzt sind und beide Kreuze richtig gesetzt sind.

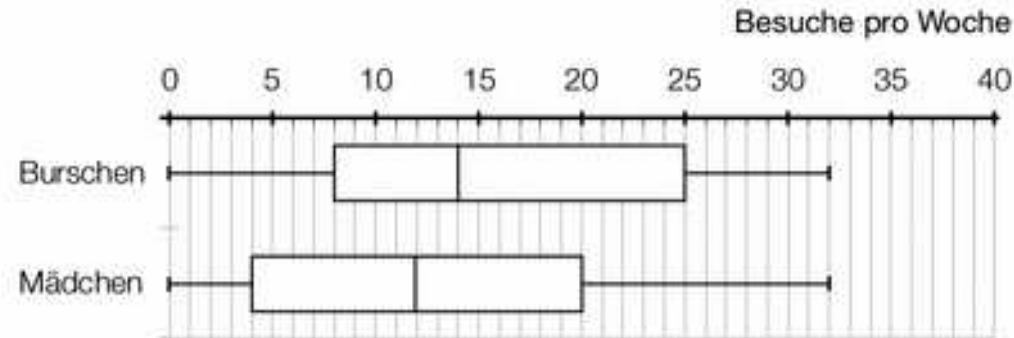
Erkenntnis:

Größte Vorsicht bei der Formulierung der Antwortmöglichkeiten ist geboten.

Maturaaufgabe 2015

Internetplattform

Die Nutzung einer bestimmten Internetplattform durch Jugendliche wird für Mädchen und Burschen getrennt untersucht. Dabei wird erfasst, wie oft die befragten Jugendlichen diese Plattform pro Woche besuchen. Die nachstehenden Kastenschaubilder (Boxplots) zeigen das Ergebnis der Untersuchung.



Aufgabenstellung:

Kreuzen Sie die beiden zutreffenden Aussagen an!

Der Median der Anzahl von Besuchen pro Woche ist bei den Burschen etwas höher als bei den Mädchen.	<input type="checkbox"/>
Die Spannweite der wöchentlichen Nutzung der Plattform ist bei den Burschen größer als bei den Mädchen.	<input type="checkbox"/>
Aus der Grafik kann man ablesen, dass genauso viele Mädchen wie Burschen die Plattform wöchentlich besuchen.	<input type="checkbox"/>
Der Anteil der Burschen, die mehr als 20-mal pro Woche die Plattform nützen, ist zumindest gleich groß oder größer als jener der Mädchen.	<input type="checkbox"/>
Ca. 80% der Mädchen und ca. 75% der Burschen nützen die Plattform genau 25-mal pro Woche.	<input type="checkbox"/>

Maturaaufgabe 2015

Lösungserwartung:

Der Median der Anzahl von Besuchen pro Woche ist bei den Burschen etwas höher als bei den Mädchen.	<input checked="" type="checkbox"/>
Der Anteil der Burschen, die mehr als 20-mal pro Woche die Plattform nützen, ist zumindest gleich groß oder größer als jener der Mädchen.	<input checked="" type="checkbox"/>

Lösungsschlüssel:

Ein Punkt ist genau dann zu geben, wenn ausschließlich die beiden laut Lösungserwartung richtigen Aussagen angekreuzt sind.

Wie können sinnvolle Aufgaben aussehen?

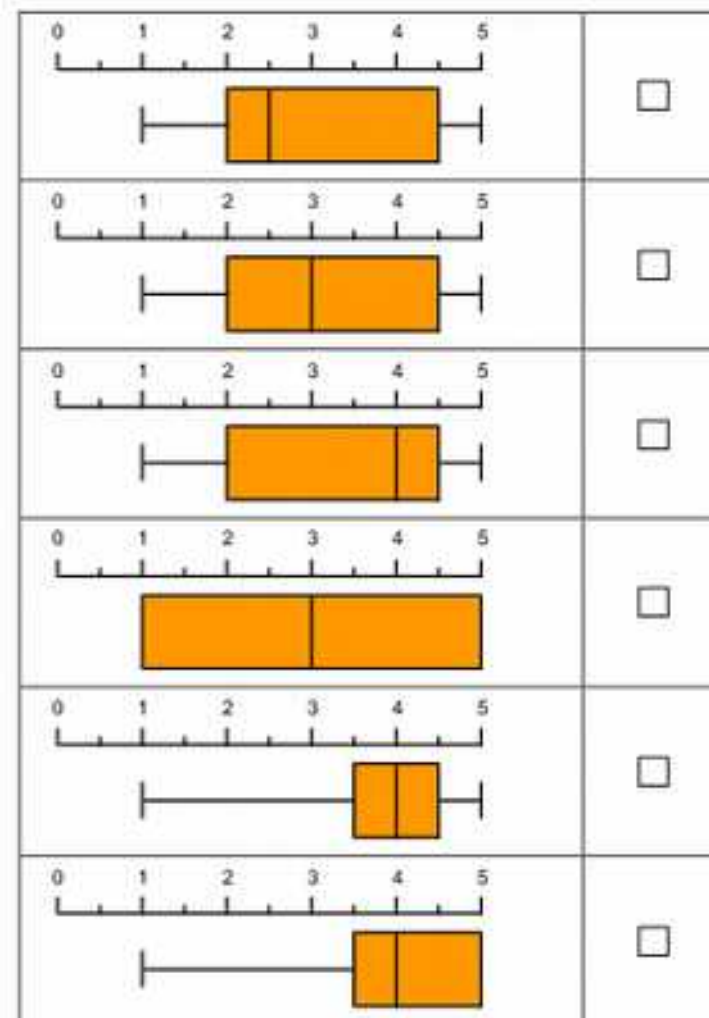
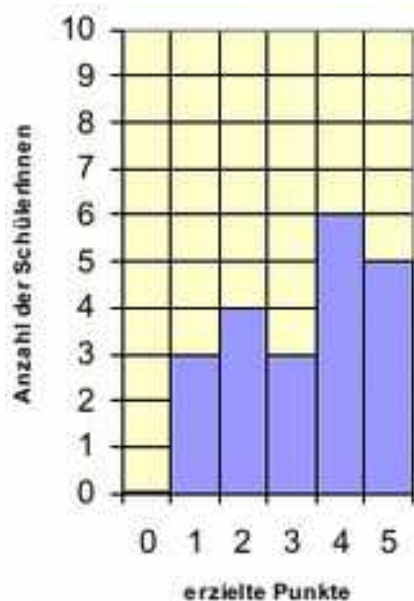
BIFIE Übungsaufgaben:

Ein Test enthält fünf Aufgaben, die jeweils nur mit einem Punkt (alles richtig) oder keinem Punkt (nicht alles richtig) bewertet werden. Die nebenstehende Grafik zeigt das Ergebnis dieses Tests für eine bestimmte Klasse.

Aufgabenstellung:

Welches der folgenden Kastenschaubilder (Boxplots) stellt die Ergebnisse des Tests richtig dar?

Kreuzen Sie das zutreffende Kastenschaubild an!

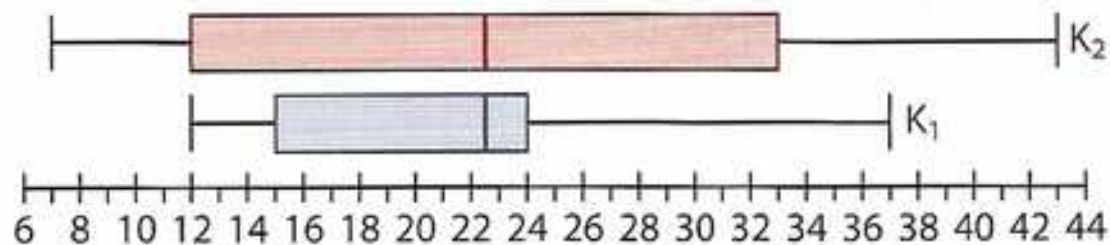


Wie können sinnvolle Aufgaben aussehen?

Bleier et al., Dimensionen 6:

Die folgenden beiden Boxplots veranschaulichen die Ausgaben der Schüler/innen der Klasse K_1 und Klasse K_2 bei der letzten Sportwoche.

Kreuze die zutreffende(n) Aussage(n) an.

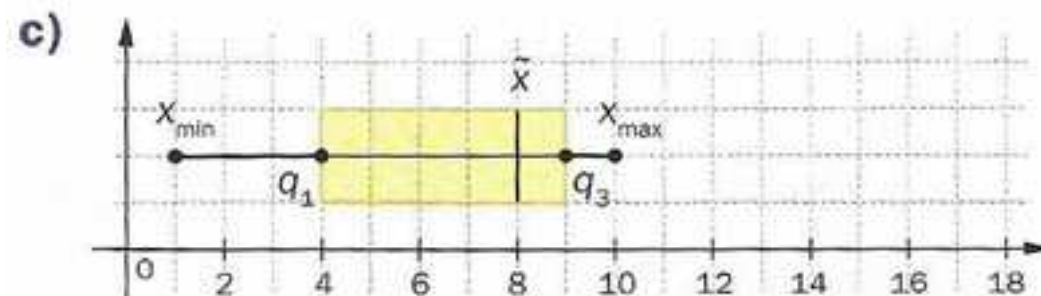
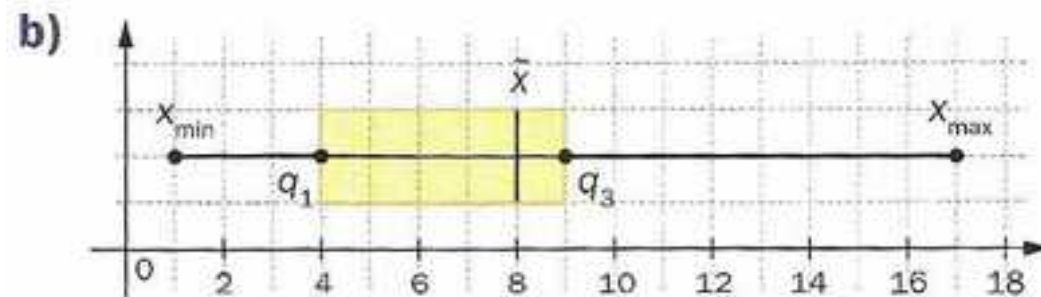
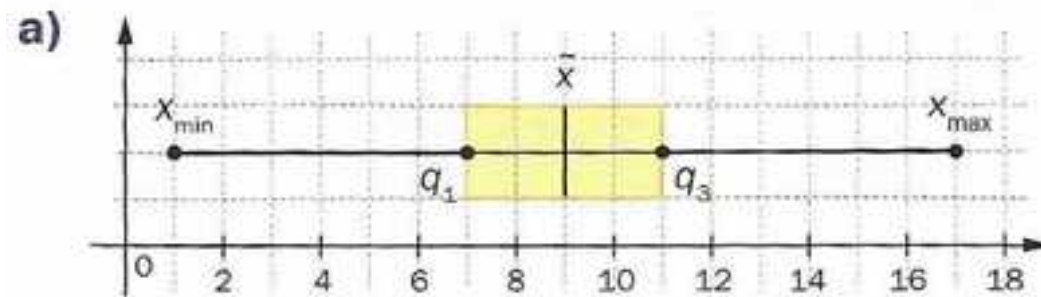


<input type="checkbox"/>	Der maximal ausgegebene Betrag ist in Klasse K_2 um 6 € größer als in Klasse K_1 .
<input type="checkbox"/>	In beiden Klassen haben 50% der Schüler/innen bis zu 22,50 € ausgegeben.
<input type="checkbox"/>	Der Quartilsabstand bei Klasse K_2 ist kleiner als bei Klasse K_1 .
<input type="checkbox"/>	Der kleinste Betrag, den ein Schüler bzw. eine Schülerin der Klasse K_1 ausgegeben hat, entspricht dem 1. Quartil von Klasse K_2 .
<input type="checkbox"/>	Die Spannweite der Ausgaben ist in beiden Klassen gleich groß.

Wie können sinnvolle Aufgaben aussehen?

Brand et al., Thema Mathematik 6:

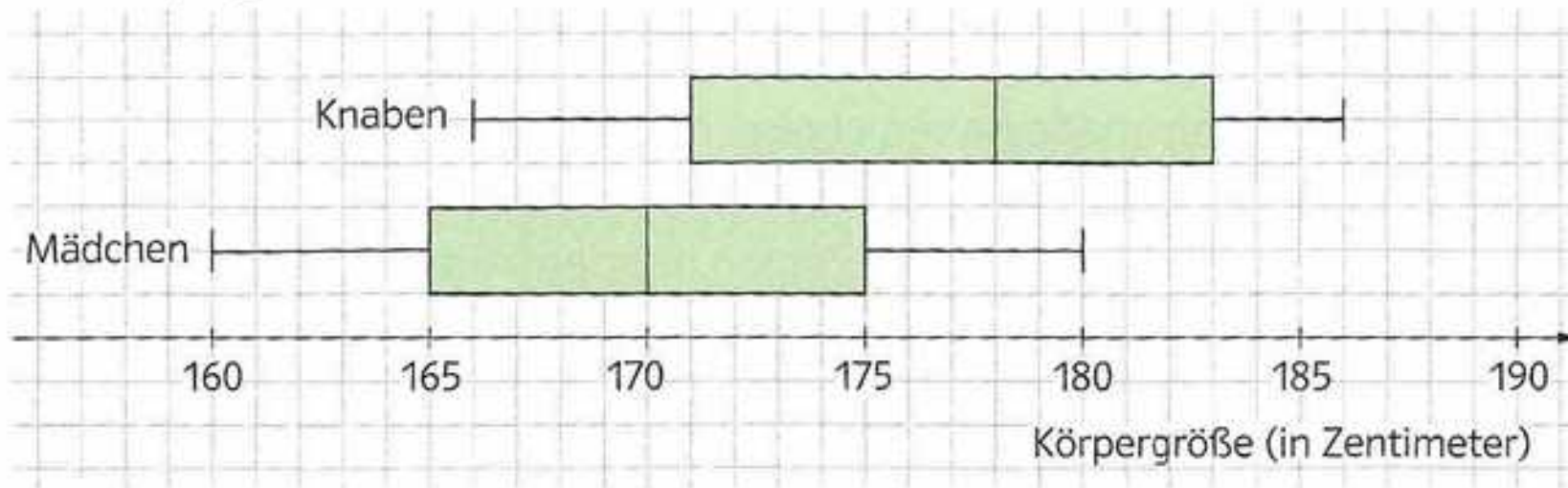
Was sagen die folgenden Boxplots aus? Lies so viele statistische Kennzahlen wie möglich ab.



Wie können sinnvolle Aufgaben aussehen?

Malle et al., Mathematik verstehen 6:

Die Körpergrößen von Schülern und Schülerinnen wurden ermittelt:

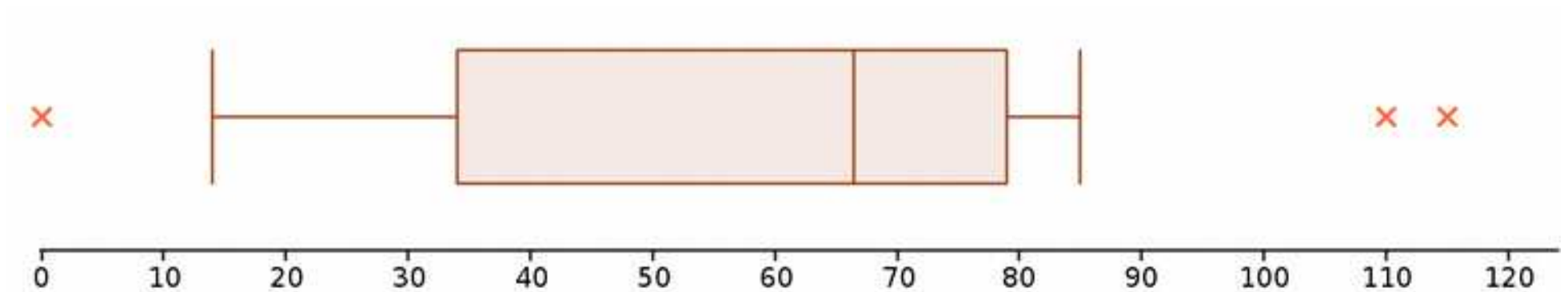


Kreuze die beiden Aussagen an, die mit Sicherheit der Grafik entnommen werden können:

- Genau die Hälfte der Knaben ist 171 cm bis 183 cm groß.
- Genau ein Mädchen hat eine Körpergröße von 160 cm.
- Mindestens 50% der Knaben sind höchstens 178 cm groß.
- Mindestens 75% der Mädchen sind mindestens 165 cm groß.
- Mindestens drei Viertel der Knaben sind kleiner als das größte Mädchen.

Problem: Ausreißer

Die Darstellung von Ausreißern ist in den Maturaformaten nicht vorgesehen.



Sie böten aber ein zweckmäßigeres und teilweise wirklichkeitsgetreueres Bild zum Interpretieren.

Worauf sollte geachtet werden?

Der Einsatz von Kastenschaubildern ist dann sinnvoll, wenn bei metrischen Merkmalen **viele Messwerte** vorliegen, zB: Körpergrößen, Massen etc.

Bei Testergebnissen hat ein Boxplot nur dann Sinn, wenn eine **große Spannweite** der zu erreichenden Punkte vorliegen.

Aussagen **müssen** Begriffe wie „ca.“, „etwa“, „ungefähr“ etc. beinhalten, damit sie als richtig eingestuft werden kann, zB: „Ca. 50 % der Daten reichen vom 1. bis zum 3. Quartil.“

Zwei oder drei Boxplots sollten zum **Vergleich** stehen.

Zu vermeiden sind:

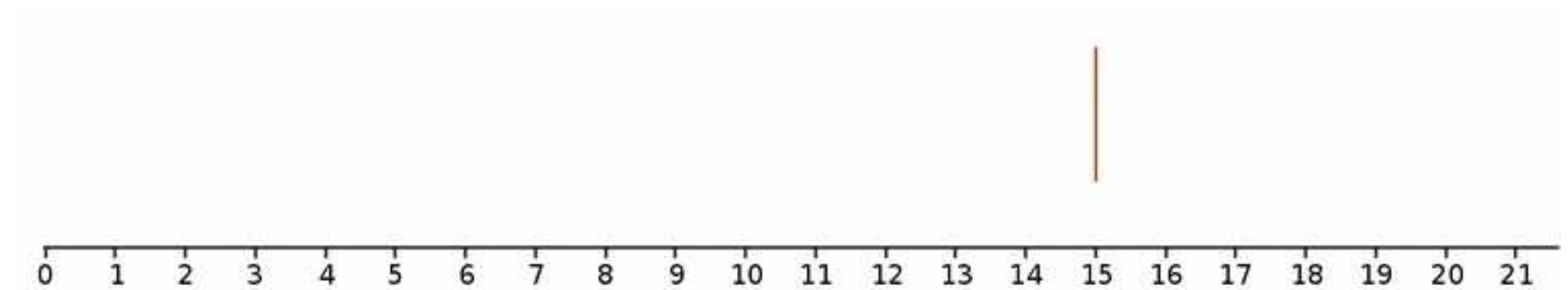
- wenige Daten in einer Liste
- Listen mit „künstlich entstandenen“ Daten, zB Gehälter in einer Firma
(1400, 1400, 2300, 2300, 2300, 2300, 2300, 2300, 2300, 4000)
- damit verbunden: viele gleiche Werte in einer Liste
- unklare Aussagekraft

Bonmot zum Schluss

Die folgende Datenliste liegt vor:

15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15

Fertigen Sie dazu ein Kastenschaubild an!



Gratulation!

Vielen Dank für die Aufmerksamkeit!

Quellen:

Bleier G., Lindenberg J., Lindner A., Süß-Stepancik E.: Dimensionen Mathematik 6.
Wien: Dorner 2015.

Brand C., Dorfmayr A., Lechner J., Mistlbacher A., Nussbaumer A.: Thema Mathematik
für die 6. Klasse AHS. Linz: Veritas 2010.

Malle G., Koth M., Woschitz H., Malle S., Salzger B., Ulovec A.: Mathematik verstehen 6.
Wien: öbv 2015.

BIFIE: Übungsaufgaben zur Vorbereitung auf die standardisierte kompetenzorientierte
schriftliche Reifeprüfung in Mathematik (AHS).
(www.bifie.at/system/files/dl/srdp_ma_uebungsaufgaben_gesamt_2014-11-21.pdf)

BIFIE: Standardisierte kompetenzorientierte schriftliche Reifeprüfung AHS 11. Mai 2015
Mathematik Teil-1-Aufgaben
(www.bifie.at/system/files/dl/KL15_PT1_AHS_MAT_T1_CC_AU_0.pdf)

BIFIE: Standardisierte kompetenzorientierte schriftliche Reifeprüfung AHS 11. Mai 2015
Mathematik Teil-1-Aufgaben Korrekturheft
(www.bifie.at/system/files/dl/KL15_PT1_AHS_MAT_T1_CC_LO_0.pdf)