

# Beschreibende Statistik (Analyse empirischer Daten)

Die empirisch gewonnenen Daten (Beobachtungswerte, Merkmalswerte)  $x_i$  werden in einer Liste (Urliste)  $[x_1, x_2, \dots, x_n]$  eingetragen.

## Grundbegriffe

<u>Arithmetisches Mittel</u>	ist die Summe der Werte $x_i$ dividiert durch deren Anzahl $n$ .
<u>Zentralwert (Median)</u>	ist derjenige Wert, der in der Mitte der aufsteigend sortierten Urliste steht. Bei einer geraden Anzahl $n$ der Werte wird als Median das arithmetische Mittel der beiden mittleren Werte (der sortierten Liste) berechnet.
<u>Modalwert (Modus)</u>	ist derjenige Wert, der die größte (relative oder absolute) Häufigkeit aufweist.
<u>Spannweite</u>	ist die Differenz aus dem größten und dem kleinsten Wert: $R = x_{max} - x_{min}$ ( $R$ engl. range)

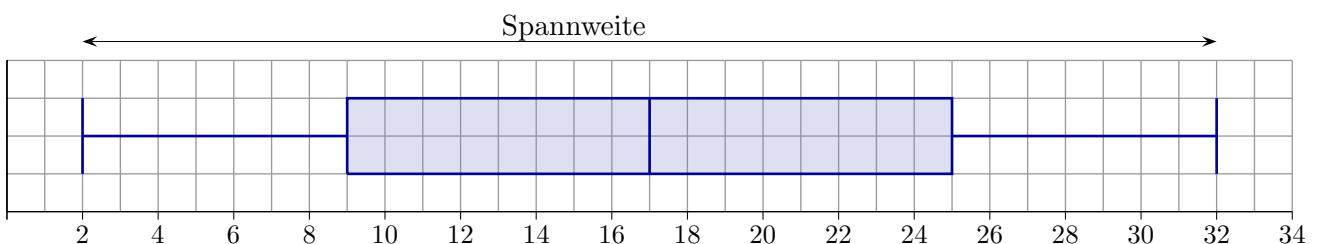
Die Spannweite ist zwar ein anschauliches Streuungsmaß, enthält allerdings keine näheren Angaben über die Verteilung der Werte und ist zudem anfällig gegenüber so genannten Ausreißern, also einzelnen ungewöhnlich niedrigen oder hohen Werten. Besser geeignet sind daher meist so genannte Quartils-Angaben (quarter engl. ein Viertel). Hierbei sortiert man zunächst alle Werte ihrer Größe nach und untergliedert diese dann in vier Bereiche.

$$L = [2, 4, 6, 8, \underbrace{10, 12, 14, 16}_{\text{unteres Quartil}}, \underbrace{17, 18, 19, 20}_{\text{Median}}, \underbrace{21, 22, 23, 24}_{\text{oberes Quartil}}, 26, 28, 30, 32]$$

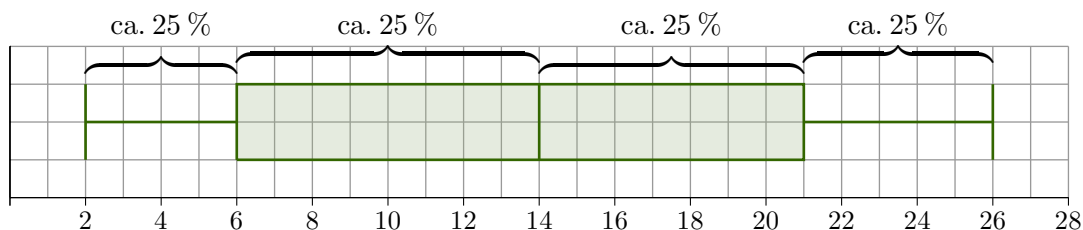
Beginne mit dem Median.

Das untere Quartil ist der Median der Werte, die links vom Median liegen.

Das obere Quartil ist der Median der Werte, die rechts vom Median liegen.



$L = [2, 4, 6, 6, 10, 12, 14, 16, 18, 20, 22, 24, 26]$ 
minimaler Wert 2, maximaler Wert 26  
unteres  
Quartil
Median
oberes  
Quartil



Der Anteil der Bereiche (hier  $[2; 6]$ ,  $[6; 14]$ ,  $[14; 21]$ ,  $[21; 26]$ ) beträgt jeweils ca. 25 %, häufig etwas mehr, da linke und rechte Grenze (und umgekehrt) in benachbarten Bereichen gleich sein können (Ausnahme hier  $[21; 26]$  mit 23 %). In Formulierungen (Interpretationen) wird vereinfachend (genau) 25 %, 50 % und 75 % verwendet.

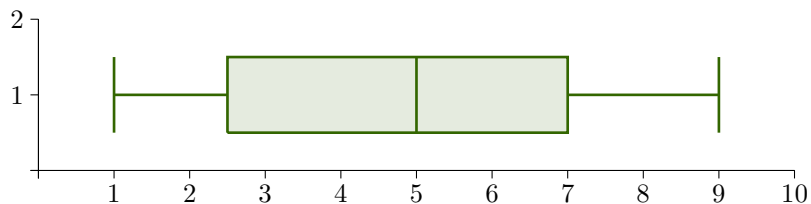
Der Quartilsabstand ist die Differenz aus dem oberen und dem unteren Quartil.

GeoGebra

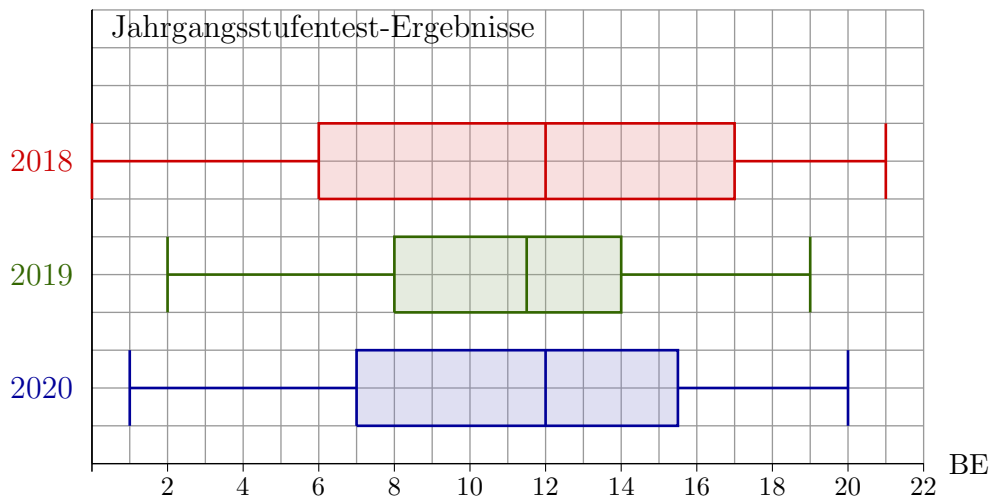
$L = \{4, 7, 1, 7, 3, 9, 2, 6, 5\}$

`Boxplot(1, 0.5, L)`

`Sortiere(L)`



Die Abbildung zeigt für drei Jahre die Mathematik-Test-Ergebnisse der jeweiligen achten Klassen eines Gymnasiums. Es konnten stets 21 Bewertungseinheiten (BE) erreicht werden.



Mögliche Interpretationen:

- a) 2019 wurden an dieser Schule Ergebnisse im Bereich von 2 bis 19 BE erzielt.
- b) 2020 erzielten ca. drei Viertel der Schülerinnen und Schüler 7 BE oder mehr.
- c) Die vollen 21 BE wurden an dieser Schule nur im Jahr 2018 erzielt.
- d) Der Median war 2018 und 2020 gleich.
- e) Die Ergebnisse umspannten 2019 einen kleineren Bereich als in den anderen beiden Jahren.
- f) Die Ergebnisse der jeweils mittleren Hälfte umspannten 2019 einen kleineren Bereich als in den anderen beiden Jahren.
- g) Mit 16 BE konnte man sich an dieser Schule in den Jahren 2019 und 2020 zu den besten 25% zählen, im Jahr 2018 dagegen nicht.

falsch wäre:

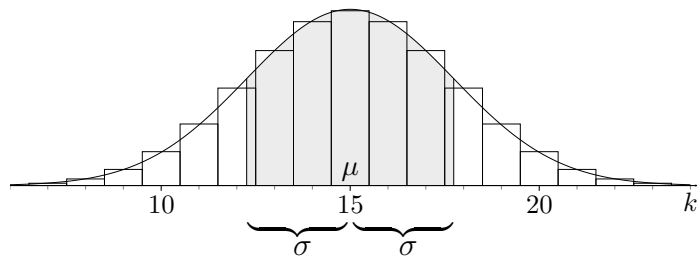
- a) 2020 erzielte genau eine Schülerin bzw. ein Schüler genau 20 BE.
- b) 2018 erzielte genau eine Schülerin bzw. ein Schüler genau 12 BE.

# Standardabweichung

Als Maß für die Streuung wird gewöhnlich die Wurzel aus der mittleren quadratischen Abweichung vom Mittelwert angegeben. Diese Größe wird Standardabweichung  $s$  ( $\sigma$ ) genannt:

$$\text{Standardabweichung } s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

siehe: [Standardabweichung](#)



Wird durch das Histogramm einer Binomialverteilung eine Ausgleichskurve gelegt (Gauß'sche Glockenkurve), so gibt  $\sigma$ , wie wir später sehen werden, die Entfernung der Wendepunkte der Kurve vom Erwartungswert  $\mu$  an.  $\sigma$  kann zudem leicht bestimmt werden:  $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$ .

Dann sind Aussagen wie „Im Abstand von maximal einer Standardabweichung vom Erwartungswert liegen etwa 2/3 aller Daten.“ möglich.

Aufgaben:

1. Seite 7 Aufg. 4a) L Seite 7
2. Seite 6 Aufg. 3a) L Seite 6
3. Seite 4 Aufg. 1a) L Seite 4
4. Seite 5 Aufg. 1 L Seite 4
5. Seite 10 Aufg. 6 L Seite 10