Arithmetisches Mittel

Für das **arithmetische Mittel** $\bar{x}$ von n Zahlen $(x_1, x_2, \ldots, x_n)$ gilt:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Berechne die Summe aller Zahlen und dividiere durch die Anzahl.

Arithmetisches Mittel

a) Berechne das arithmetische Mittel von $(7, -2, 0, 3, 1, 3, 9)$.

$$\bar{x} = \frac{7 + (-2) + 0 + 3 + 1 + 3 + 9}{7} = \frac{21}{7} = 3$$

b) Das arithmetische Mittel von 30 Zahlen ist $\bar{x} = 14{,}7$. Wir fügen die Zahl 5 als 31. Zahl hinzu. Berechne das neue arithmetische Mittel $\bar{x}_{\text{neu}}$. Um wie viel Prozent ist $\bar{x}_{\text{neu}}$ kleiner als $\bar{x}$?

$$x_1 + x_2 + \cdots + x_{30} = 30 \cdot \bar{x}$$

$$\bar{x}_{\text{neu}} = \frac{30 \cdot \bar{x} + x_{31}}{31} = 14{,}38... \quad \Longrightarrow \quad \frac{\bar{x}_{\text{neu}}}{\bar{x}} = 0{,}9787... = 97{,}87...\%.$$

Das neue arithmetische Mittel ist also um $100\,\% - 97{,}87...\% = 2{,}12...\%$ kleiner als $\bar{x}$.

An den Schrauben drehen

Wie verändert sich das arithmetische Mittel von 20 Zahlen, wenn …

… jede der Zahlen um 42 vergrößert wird? Es wird um 42 größer.

… jede der Zahlen mit 3 multipliziert wird? Es wird mit 3 multipliziert.

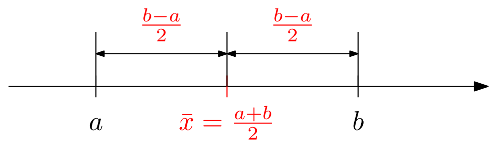
… das Vorzeichen aller Zahlen umgedreht wird? Das Vorzeichen des a. M. dreht sich um.

… das arithmetische Mittel von jeder Zahl abgezogen wird? Das neue a. M. ist 0.

… eine der Zahlen um 10 000 vergrößert wird? Es wird um 500 größer.

Arithmetisches *Mittel*

Rechts sind zwei Zahlen a und b mit $a < b$ auf der Zahlengerade dargestellt.
Erkläre, warum das arithmetische Mittel $\bar{x}$ von a und b genau in der Mitte von a und b auf der Zahlengerade liegt.



$$a + \frac{b-a}{2} = \frac{2 \cdot a}{2} + \frac{b-a}{2} = \frac{a+b}{2} = \bar{x}$$

Varianz & Standardabweichung

Für die **Varianz** s^2 von n Zahlen $(x_1, x_2, \ldots, x_n)$ mit arithmetischem Mittel $\bar{x}$ gilt:

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

Die Varianz s^2 ist also das arithmetische Mittel der quadratischen Abweichungen von $\bar{x}$.
Warum gilt $s^2 \geq 0$?
Unter welcher Bedingung gilt $s^2 = 0$?

Für die **Standardabweichung** s dieser n Zahlen gilt:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

s hat die gleiche Einheit wie die Zahlen x_i.
Das arithmetische Mittel von $(0, 0)$ ist 0.
Das arithmetische Mittel von $(-42, 42)$ ist 0.
Die Standardabweichung von $(0, 0)$ ist 0.
Die Standardabweichung von $(-42, 42)$ ist 42.

Varianz & Standardabweichung

Berechne das arithmetische Mittel der 5 Zahlen $(5, 7, 1, 3, 4)$: $\bar{x} = \frac{5+7+1+3+4}{5} = 4$

Berechne die Varianz und Standardabweichung dieser 5 Zahlen:

$$s^2 = \frac{1+9+9+1+0}{5} = 4$$

$$s = \sqrt{4} = 2$$

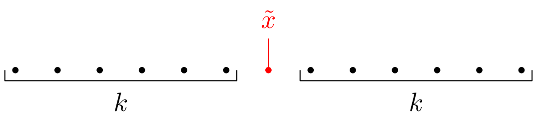
x_i	5	7	1	3	4
$x_i - \bar{x}$	1	3	-3	-1	0
$(x_i - \bar{x})^2$	1	9	9	1	0

Median

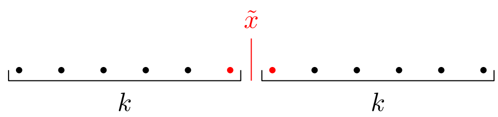
Den **Median $\tilde{x}$** von n reellen Zahlen ermitteln wir folgendermaßen:

Zuerst sortieren wir die Zahlen in aufsteigender Reihenfolge.

Ist n **ungerade**, dann ist $\tilde{x}$ der **mittlere Wert**:



Ist n **gerade**, dann ist $\tilde{x}$ das **arithmetische Mittel der beiden mittleren Werte**:



Median

a) Ermittle jeweils den Median der aufsteigend sortierten Zahlenliste.

i) $(-4, 1, 2, \mathbf{3}, 7, 7, 8) \implies \tilde{x} = 3$

ii) $(-4, 1, \mathbf{2}, \mathbf{7}, 7, 8) \implies \tilde{x} = \frac{2+7}{2} = 4{,}5$

b) In der Tabelle sind die absoluten Häufigkeiten der erreichten Noten bei einer Prüfung aufgelistet. Ermittle den Median der erreichten Noten bei dieser Prüfung.

Note	1	2	3	4	5
Häufigkeit	4	8	5	4	2

$< \underbrace{1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2}_{11 \text{ Zahlen}}, \mathbf{2}, \underbrace{3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5}_{11 \text{ Zahlen}} >$

Oder schneller: Die aufsteigend sortierte Notenliste besteht aus $4+8+5+4+2 = 23$ Zahlen. Die 12. Note ist genau in der Mitte.

Der Median ist also $\tilde{x} = 2$.

50 %-Eigenschaft

Zur Berechnung des Medians sortieren wir die Zahlenliste. Deshalb hat er die folgende Eigenschaft:
Mindestens 50 % der Zahlen in der Liste sind kleiner oder gleich dem Median, und
mindestens 50 % der Zahlen in der Liste sind größer oder gleich dem Median.
Können es *mehr* als 50 % der Zahlen sein? Begründe deine Antwort.

Ja, z.B. in der Liste $(42, 42, 42)$ sind *alle* Werte $\leq$ dem Median 42 und *alle* Werte $\geq$ dem Median.

An den Schrauben drehen

Wie verändert sich der Median von 20 verschiedenen Zahlen, wenn …

… jede der Zahlen um 42 vergrößert wird? Er wird um 42 größer.

… jede der Zahlen mit 3 multipliziert wird? Er wird mit 3 multipliziert.

… die fünf kleinsten Zahlen jeweils um 60 verkleinert werden? Er bleibt gleich.

… die drittgrößte Zahl um 10 000 vergrößert wird? Er bleibt gleich.

Empirische Verteilungsfunktion

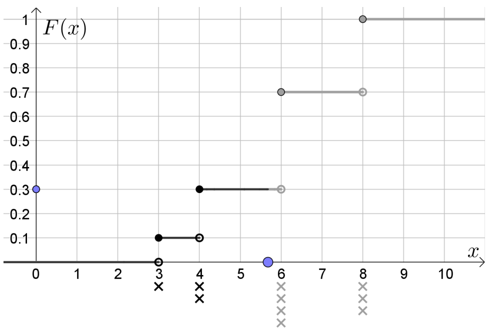
Die Liste $(3,4,4,6,6,6,6,8,8,8)$ hat $n = 10$ Einträge.

Die Funktion F mit

$$F(x) = \frac{\text{Anzahl Einträge} \leq x}{n}$$

heißt **empirische Verteilungsfunktion**.
Vervollständige die Wertetabelle von F.

x	1	2	3	4	5	6	7	8
$F(x)$	0	0	0,1	0,3	0,3	0,7	0,7	1



p-Quantil

Gegeben ist eine Liste mit n Zahlen und eine Zahl p in $[0;1]$.
Eine Zahl x_p heißt ***p*-Quantil**, wenn sie die folgenden beiden Eigenschaften hat:

$$\frac{\text{Anzahl Einträge} \leq x_p}{n} \geq p \quad \text{und} \quad \frac{\text{Anzahl Einträge} \geq x_p}{n} \geq 1-p$$

Zum Beispiel ist der Median $\tilde{x}$ jeder Zahlenliste ein 50 %-Quantil:

$$\frac{\text{Anzahl Einträge} \leq \tilde{x}}{n} \geq 0{,}5 \quad \text{und} \quad \frac{\text{Anzahl Einträge} \geq \tilde{x}}{n} \geq 0{,}5$$

p-Quantil

Die Liste $(2,4,5,5,7,7,7,8,9,9)$ besteht aus $n = 10$ Zahlen.

a) Erkläre, warum die Zahl 4 ein 20 %-Quantil der Liste ist.
Es sind $\frac{2}{10} = 20\,\%$ der Werte ≤ 4 und $\frac{9}{10} = 90\,\% \geq 80\,\%$ der Werte ≥ 4.

b) Erkläre, warum die Zahl 4,2 ein 20 %-Quantil der Liste ist. *p*-Quantile müssen also nicht eindeutig sein.
Es sind $\frac{2}{10} = 20\,\%$ der Werte $\leq 4{,}2$ und $\frac{8}{10} = 80\,\%$ der Werte $\geq 4{,}2$.

c) Erkläre, warum die Zahl 4 auch ein 17 %-Quantil der Liste ist.
Es sind $\frac{2}{10} \geq 17\,\%$ der Werte ≤ 4 und $\frac{9}{10} \geq 83\,\%$ der Werte ≥ 4.

d) Die Zahl 7,3 ist ein 70 %-Quantil dieser Liste.

e) Für welche Zahlen p ist die Zahl 8 ein p-Quantil dieser Liste?
Es sind $\frac{8}{10} = 80\,\%$ der Werte ≤ 8 und $\frac{3}{10} = 30\,\%$ der Werte ≥ 8.
Also ist für alle Zahlen p in $[70\,\%; 80\,\%]$ die Zahl 8 ein p-Quantil dieser Liste.

Quartile

Die 25 %-Quantile, 50 %-Quantile und 75 %-Quantile heißen auch **Quartile**. Viertel ↔ quarter

Jedes **25 %-Quantil** heißt auch **erstes/unteres Quartil** und wird mit q_1 abgekürzt.

Jedes **50 %-Quantil** heißt auch **zweites/mittleres Quartil** und wird mit q_2 abgekürzt.

Jedes **75 %-Quantil** heißt auch **drittes/oberes Quartil** und wird mit q_3 abgekürzt.

Boxplot

Die Körpergröße von $n = 80$ Personen wurde gemessen.
Die Messwerte (in cm) sind in der folgenden Liste aufsteigend sortiert:

155	156	156	158	158	158	159	161	162	162	162	163	163	163	164	164	165	165	165	165
166	166	166	166	166	166	166	167	167	167	167	168	168	168	169	169	169	169	169	169
170	170	170	170	170	171	171	171	171	172	172	172	172	172	173	173	174	174	174	174
174	174	175	176	176	177	178	178	178	179	179	180	180	180	180	182	184	188	188	190

Um diese Daten in einem **Boxplot** zu veranschaulichen, ermitteln wir fünf Kenngrößen:

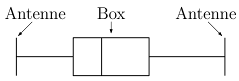
Kleinster Wert:	Unteres Quartil:	Mittleres Quartil:	Oberes Quartil:	Größter Wert:
$x_{\min} = 155\,\text{cm}$	$q_1 = 165{,}5\,\text{cm}$	$q_2 = 169{,}5\,\text{cm}$	$q_3 = 174\,\text{cm}$	$x_{\max} = 190\,\text{cm}$

In diesem Beispiel ist für q_1 *jede* Zahl im Intervall $[165\,\text{cm}; 166\,\text{cm}]$ richtig. Wir haben den Median der ersten 40 Messwerte gewählt.
In diesem Beispiel ist für q_2 *jede* Zahl im Intervall $[169\,\text{cm}; 170\,\text{cm}]$ richtig. Wir haben den Median aller Messwerte gewählt.
In diesem Beispiel ist für q_3 *nur* die Zahl $174\,\text{cm}$ richtig.

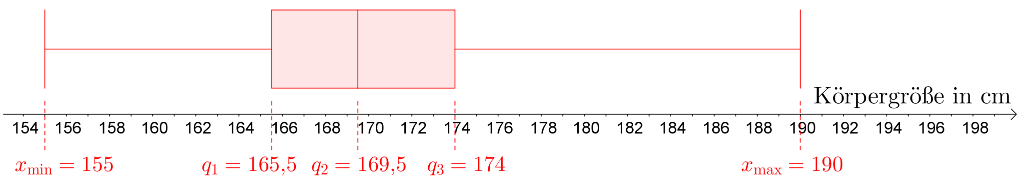
1) An den Stellen $x_{\min}$ und $x_{\max}$ zeichnen wir jeweils eine **Antenne** ein.

2) Die **Box** zeichnen wir als Rechteck über dem Intervall $[q_1; q_3]$ ein.

3) An der Stelle q_2 wird die Box durch einen senkrechten Strich in 2 Teile geteilt.



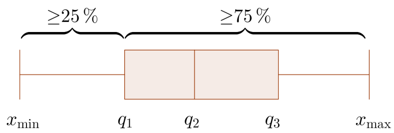
Zeichne den Boxplot ein:



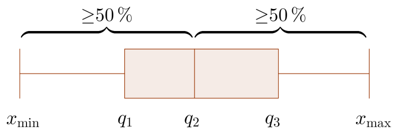
Eigenschaften der Quartile

Die Quartile q_1, q_2, q_3 teilen das Intervall $[x_{\min}; x_{\max}]$ in Intervalle mit den folgenden Eigenschaften auf:

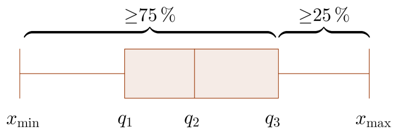
1) Das Intervall $[x_{\min}; q_1]$ enthält mindestens 25 % der Werte.
Das Intervall $[q_1; x_{\max}]$ enthält mindestens 75 % der Werte.
So sind 25 %-Quantile definiert.



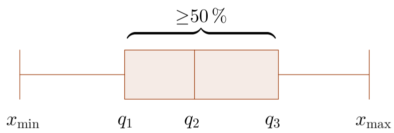
2) Das Intervall $[x_{\min}; q_2]$ enthält mindestens 50 % der Werte.
Das Intervall $[q_2; x_{\max}]$ enthält mindestens 50 % der Werte.
So sind 50 %-Quantile definiert.



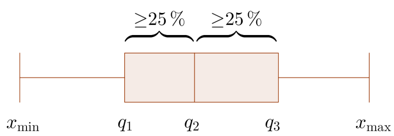
3) Das Intervall $[x_{\min}; q_3]$ enthält mindestens 75 % der Werte.
Das Intervall $[q_3; x_{\max}]$ enthält mindestens 25 % der Werte.
So sind 75 %-Quantile definiert.



4) Das Intervall $[q_1; q_3]$ enthält mindestens 50 % der Werte.
Es sind höchstens 25 % der Werte größer als q_3, weil mindestens 75 % der Werte $\leq q_3$ sind. Genauso sind höchstens 25 % der Werte kleiner als q_1. Also sind mindestens 50 % der Werte im Intervall $[q_1; q_3]$.



5) Das Intervall $[q_1; q_2]$ enthält mindestens 25 % der Werte.
Das Intervall $[q_2; q_3]$ enthält mindestens 25 % der Werte.
Es sind höchstens 25 % der Werte kleiner als q_1, weil mindestens 75 % der Werte $\geq q_1$ sind. Es sind höchstens 50 % der Werte größer als q_2, weil mindestens 50 % der Werte $\leq q_2$ sind. Also sind mindestens 25 % der Werte in $[q_1; q_2]$.



Berechnung von Quartilen MmF

Wir können Quartile q_1, q_2 und q_3 einer aufsteigend sortierten Zahlenliste wie folgt berechnen:

i) Für q_2 verwenden wir den Median der Werte.

ii) Dann teilen wir die Liste folgendermaßen in zwei gleich große Teillisten:

- Ist $\boldsymbol{n}$ **gerade**, dann teilen wir die Liste in der Mitte auf:
- Ist $\boldsymbol{n}$ **ungerade**, dann teilen wir die Liste so auf, dass auch beide Teillisten eine *ungerade* Anzahl an Werten enthalten: oder

  Dafür musst du den mittleren Wert entweder zu beiden Teillisten dazunehmen oder jeweils *nicht* dazunehmen.

iii) Für q_1 verwenden wir den Median der linken Teilliste.
Für q_3 verwenden wir den Median der rechten Teilliste.

Mit dieser Methode erhalten wir zuverlässig richtige Werte für q_1 und q_3. Eine Erklärung dafür findest du im KH – Statistik.

Odd MmF

Seit Jahrzehnten sind auch Berechnungsmethoden für q_1 und q_3 im Umlauf, die *nicht* zuverlässig 25 %-Quantile bzw. 75 %-Quantile liefern.

Auch GeoGebra berechnet zum Beispiel für $(1, 2, 3, 4, 5)$ den Wert $q_1 = 1{,}5$. Warum ist 1,5 *kein* 25 %-Quantil von $(1, 2, 3, 4, 5)$?

Spannweite & Interquartilsabstand MmF

Die **Spannweite** $\boldsymbol{R}$ (engl. range) einer Zahlenliste ist der Abstand zwischen dem kleinsten und dem größten Wert der Liste:

$$\boldsymbol{R = x_{\max} - x_{\min}}$$

Die Spannweite ist im Boxplot also der Abstand zwischen den beiden Antennen.

Der **Interquartilsabstand** $\boldsymbol{I}$ (kurz: Quartilsabstand) ist der Abstand zwischen dem ersten Quartil und dem dritten Quartil:

$$\boldsymbol{I = q_3 - q_1}$$

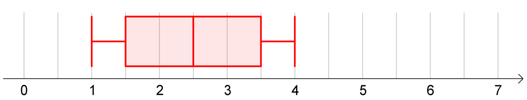
Der Quartilsabstand ist im Boxplot also die Breite der Box.

Boxplot MmF

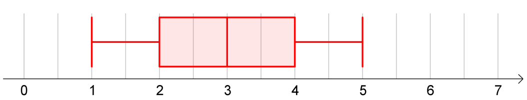
Berechne jeweils die Kenngrößen, und zeichne den zugehörigen Boxplot unten ein.

	$x_{\min}$	q_1	q_2	q_3	$x_{\max}$	R	I
$A = (1, 2, 3, 4)$	1	1,5	2,5	3,5	4	3	2
$B = (1, 2, 3, 4, 5)$	1	2	3	4	5	4	2
$C = (1, 2, 3, 4, 5, 6)$	1	2	3,5	5	6	5	3
$D = (1, 2, 3, 4, 5, 6, 7)$	1	2	4	6	7	6	4

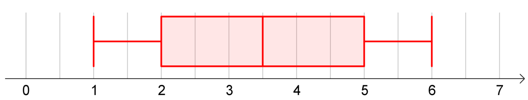
Boxplot von A:



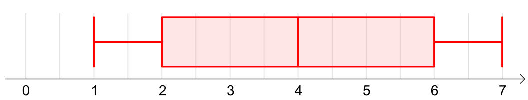
Boxplot von B:



Boxplot von C:



Boxplot von D:

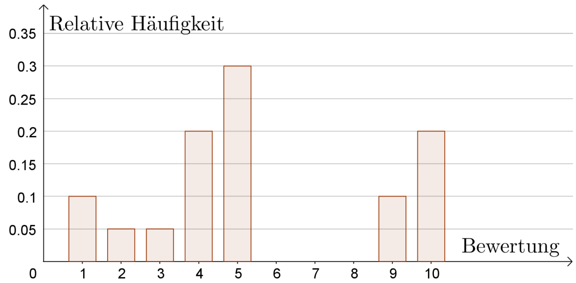


Säulendiagramm → Boxplot

Auf einer Website werden Filme bewertet.

Jede Bewertung ist eine natürliche Zahl von 1 bis 10.

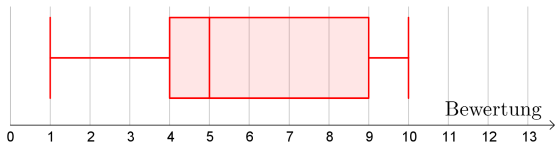
Rechts ist das Säulendiagramm mit den relativen Häufigkeiten der Bewertungen für einen bestimmten Film dargestellt.



1) Ermittle richtige Werte für die zugehörigen statistischen Kenngrößen x_{min}, q_1, q_2, q_3 und x_{max}.

$x_{\text{min}} = 1 \qquad q_1 = 4 \qquad q_2 = 5 \qquad q_3 = 9 \qquad x_{\text{max}} = 10$

2) Zeichne unten einen Boxplot für die Bewertungen dieses Films ein.



Welche Auswirkung haben die hohen relativen Häufigkeiten von 4 und 5 auf diesen Boxplot?

Welche Auswirkung haben die relativen Häufigkeiten von 6, 7 und 8 (0 %) auf diesen Boxplot?

Geometrisches Mittel

Für das **geometrische Mittel $\bar{x}_{\text{geo}}$** von n positiven Zahlen $(x_1, x_2, \dots, x_n)$ gilt:

$$\bar{x}_{\text{geo}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Geometrisches Mittel

Der Wert einer Immobilie ist vom Jahr 2019 auf das Jahr 2020 um 7 % gestiegen.
Der Wert dieser Immobilie ist vom Jahr 2020 auf das Jahr 2021 um 11 % gestiegen.

1) Um wie viel % ist der Wert dieser Immobilie vom Jahr 2019 auf das Jahr 2021 gestiegen?

2) Um wie viel % ist der Wert dieser Immobilie vom Jahr 2019 auf das Jahr 2021 pro Jahr durchschnittlich gestiegen?

1) $1{,}07 \cdot 1{,}11 = 1{,}1877$

Der Wert ist vom Jahr 2019 auf das Jahr 2021 um 18,77 % gestiegen.

2) $W \cdot q^2 = W \cdot 1{,}1877 \implies q = \sqrt{1{,}1877} = 1{,}0898\ldots$

Der Wert ist pro Jahr durchschnittlich um 8,98... % gestiegen.

Geometrisches Mittel

Für n beliebige positive Zahlen x_i gilt: $\dfrac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$

Das arithmetische Mittel von n positiven Zahlen ist stets größer oder gleich dem geometrischen Mittel dieser Zahlen.

Beweise den Fall $n = 2$, also: $\dfrac{x_1 + x_2}{2} \geq \sqrt{x_1 \cdot x_2}$

$$\frac{x_1 + x_2}{2} \geq \sqrt{x_1 \cdot x_2} \iff \frac{(x_1 + x_2)^2}{4} \geq x_1 \cdot x_2 \iff x_1^2 + 2 \cdot x_1 \cdot x_2 + x_2^2 \geq 4 \cdot x_1 \cdot x_2$$
$$\iff x_1^2 - 2 \cdot x_1 \cdot x_2 + x_2^2 \geq 0 \iff (x_1 - x_2)^2 \geq 0 \checkmark$$