Arithmetisches Mittel

Das **arithmetische Mittel** $\bar{x}$ von n Zahlen $\langle x_1, x_2, \dots, x_n\rangle$ ist:

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

„Zähle alle Zahlen zusammen und dividiere durch die Anzahl.“

Arithmetisches Mittel

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

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

… jede der Zahlen mit 3 multipliziert wird?

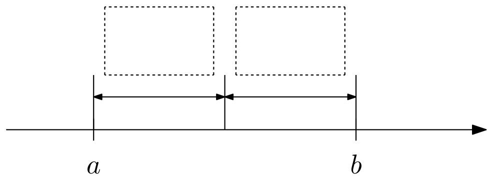
… das Vorzeichen aller Zahlen umgedreht wird?

… das arithmetische Mittel von jeder Zahl abgezogen wird?

… eine der Zahlen um 10 000 vergrößert wird?

Arithmetisches *Mittel*

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



Varianz & Standardabweichung

Die **Varianz** $\boldsymbol{s^2}$ von n Zahlen $\langle x_1, x_2, \dots, x_n\rangle$ mit arithmetischem Mittel $\bar{x}$ ist:

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

Die Varianz ist also das arithmetische Mittel der quadratischen Abweichungen von $\bar{x}$.

Warum ist die Varianz immer größer oder gleich Null? Kann die Varianz gleich Null sein?

Die **Standardabweichung** $\boldsymbol{s}$ dieser n Zahlen ist:

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

s hat die gleiche Einheit wie die Werte x_i.

Varianz & Standardabweichung

Berechne das arithmetische Mittel der 5 Zahlen $\langle 5, 7, 1, 3, 4 \rangle$: $\bar{x} =$ ____________________

Berechne die Varianz dieser 5 Zahlen:

$s^2 =$ ____________________

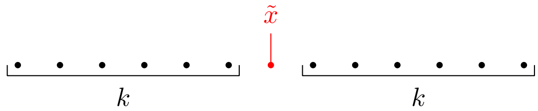
x_i	5	7	1	3	4
$x_i - \bar{x}$					
$(x_i - \bar{x})^2$					

Berechne die Standardabweichung dieser 5 Zahlen: $s =$ __________

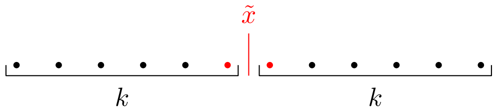
Median

Den **Median** $\tilde{x}$ von n reellen Zahlen berechnen 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 den Median der gegebenen Zahlen.

i) $\langle -4, 1, 2, 3, 7, 7, 8 \rangle$: $\tilde{x} =$ __________ **ii)** $\langle -4, 1, 2, 7, 7, 8 \rangle$: $\tilde{x} =$ __________

b) In der Tabelle siehst du die Häufigkeiten der Noten bei einer Prüfung.
Ermittle den Median der erhaltenen Noten bei dieser Prüfung.

Note	1	2	3	4	5
Häufigkeit	4	8	5	4	2

50 %-Eigenschaft

Erkläre, weshalb der Median einer Liste von Zahlen stets die folgende Eigenschaft hat:
Mindestens 50 % der Zahlen sind kleiner oder gleich dem Median, und
mindestens 50 % der Zahlen sind größer oder gleich dem Median.

Wie können es *mehr* als 50 % sein?

An den Schrauben drehen

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

… jede der Zahlen um 42 vergrößert wird?

… jede der Zahlen mit 3 multipliziert wird?

… die fünf kleinsten Zahlen jeweils um 60 verkleinert werden?

… die drittgrößte Zahl um 10 000 vergrößert wird?

Empirische Verteilungsfunktion

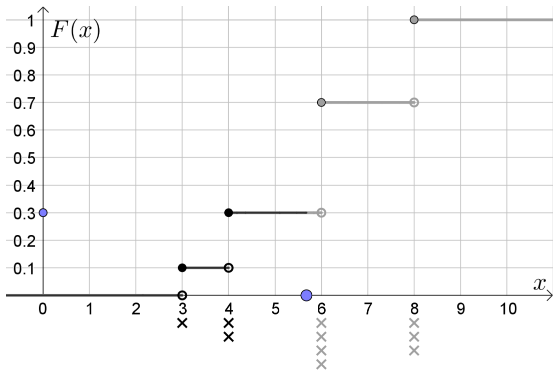
Die Liste $\langle 3, 4, 4, 6, 6, 6, 6, 8, 8, 8\rangle$ hat $n = 10$ Einträge.

Die Funktion F mit

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

heißt **empirische Verteilungsfunktion** der Liste.
Fülle die Wertetabelle von F aus:

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



***p*-Quantil**

Vor dir liegt eine Liste mit n Zahlen. p ist eine Zahl in $[0; 1]$.
Eine Zahl $\boldsymbol{x_p}$ heißt ***p*-Quantil**, wenn sie die folgenden beiden Eigenschaften hat:

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

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

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

50 %-Eigenschaft vom Median

***p*-Quantil**

Gegeben ist die Liste $\langle 2, 4, 5, 5, 7, 7, 7, 8, 9, 9\rangle$ mit 10 Werten.

1) Erkläre, warum die Zahl 4 ein 20 %-Quantil der Liste ist.

2) Erkläre, warum auch die Zahl 4,2 ein 20 %-Quantil der Liste ist.

3) Erkläre, warum die Zahl 4 auch ein 17 %-Quantil der Liste ist.

4) Die Zahl 7,3 ist ein ______ %-Quantil dieser Liste.

5) Für welche Zahlen p ist die Zahl 8 ein p-Quantil dieser Liste?

Quartile

Die 25 %-Quantile, 50 %-Quantile und 75 %-Quantile heißen auch kurz **Quartile**.

Merkhilfe: *quarter* ist englisch für *ein Viertel*.

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

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

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

Beispiel: Die Körpergröße von $n = 80$ Personen wurde gemessen.
Die Messwerte (in cm) haben wir für dich 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

Stelle die Messwerte in einem Boxplot dar.

Boxplot

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

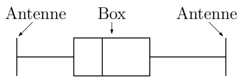
Kleinster Wert:	Unteres Quartil:	Mittleres Quartil:	Oberes Quartil:	Größter Wert:
$x_{\text{min}} = 155\,\text{cm}$	$q_1 = 165{,}5\,\text{cm}$	$q_2 = 169{,}5\,\text{cm}$	$q_3 = 174\,\text{cm}$	$x_{\text{max}} = 190\,\text{cm}$

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 $[165\,\text{cm}; 166\,\text{cm}]$ richtig. Wir haben den Median aller Messwerte gewählt.
In diesem Beispiel ist für q_3 *nur* die Zahl $174\,\text{cm}$ richtig.

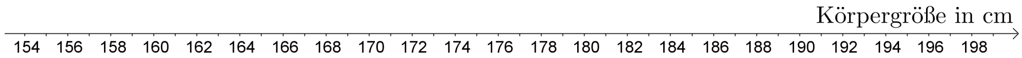
1) An den Stellen x_{min} und x_{max} zeichnen wir jeweils eine **Antenne** ein.

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

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



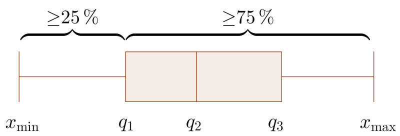
Zeichne den Boxplot ein:



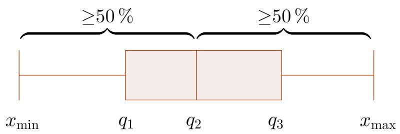
Eigenschaften der Quartile

q_1, q_2 und q_3 teilen das Intervall $[x_{\text{min}}; x_{\text{max}}]$ in Teilintervalle mit folgenden Eigenschaften auf:

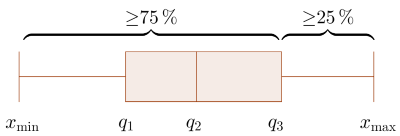
1) Das Intervall $[x_{\text{min}}; q_1]$ enthält mindestens 25 % der Werte.
Das Intervall $[q_1; x_{\text{max}}]$ enthält mindestens 75 % der Werte.
So sind 25 %-Quantile definiert.



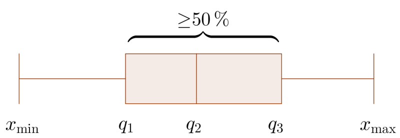
2) Das Intervall $[x_{\text{min}}; q_2]$ enthält mindestens 50 % der Werte.
Das Intervall $[q_2; x_{\text{max}}]$ enthält mindestens 50 % der Werte.
So sind 50 %-Quantile definiert.



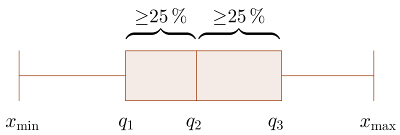
3) Das Intervall $[x_{\text{min}}; q_3]$ enthält mindestens 75 % der Werte.
Das Intervall $[q_3; x_{\text{max}}]$ enthält mindestens 25 % der Werte.
So sind 75 %-Quantile definiert.



4) Das Intervall $[q_1; q_3]$ enthält mindestens 50 % der Werte.
Erklärung: Es sind weniger als 25 % der Werte größer als q_3, weil mindestens 75 % der Werte kleiner oder gleich q_3 sind. Genauso sind weniger als 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.
Versuche diese Eigenschaft wie **4)** zu erklären.



Berechnung der Quartile

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

1) Für q_2 verwenden wir den Median der Werte.

2) Dann teilen wir die Liste in zwei gleich große Teillisten wie folgt:

- 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.

3) q_1 ist der Median der linken Teilliste. q_3 ist der Median der rechten Teilliste.

Odd

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

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

Die vorgestellte Methode liefert zuverlässig richtige Werte für q_1 und q_3.

Eine Erklärung dafür findest du im Kompetenzheft – Statistik I.

Spannweite & Interquartilsabstand

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:

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

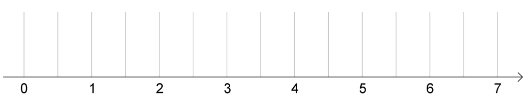
Der Quartilsabstand ist im Boxplot also die Breite der Box.

Boxplot

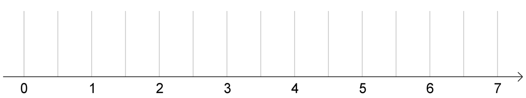
Berechne jeweils die Kenngrößen und zeichne den zugehörigen Boxplot.

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

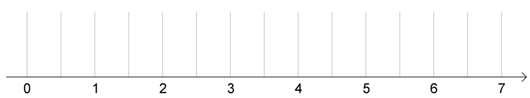
Boxplot von A:



Boxplot von B:



Boxplot von C:



Boxplot von D: