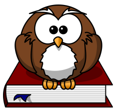
KOMPETENZHEFT – STATISTIK

Inhaltsverzeichnis



Kompetenzmaterialien – Statistik

In diesem Kompetenzheft wird ein möglicher Einstieg ins Thema „Statistik“ vorgestellt.

Die mit ☆ markierten Inhalte sind für besonders interessierte Personen gedacht.

Die folgenden Materialien sind für den Einsatz im Unterricht konzipiert:

✓ Arbeitsblatt – Relative Häufigkeiten und Baumdiagramme (Ausarbeitung)
✓ Arbeitsblatt – Statistische Kenngrößen und Boxplot (Ausarbeitung)
✓ Arbeitsblatt – Interpolation und Regression (Ausarbeitung)
✓ Technologieblatt – Regression (Ausarbeitung)

In der Aufgabensammlung – Statistik befinden sich passende Übungsaufgaben.

Wir freuen uns über Feedback an mmf@univie.ac.at.

Datum: 10. Dezember 2020.

1. Absolute/Relative Häufigkeiten & Diagramme

Arbeitsblatt – Relative Häufigkeiten und Baumdiagramme

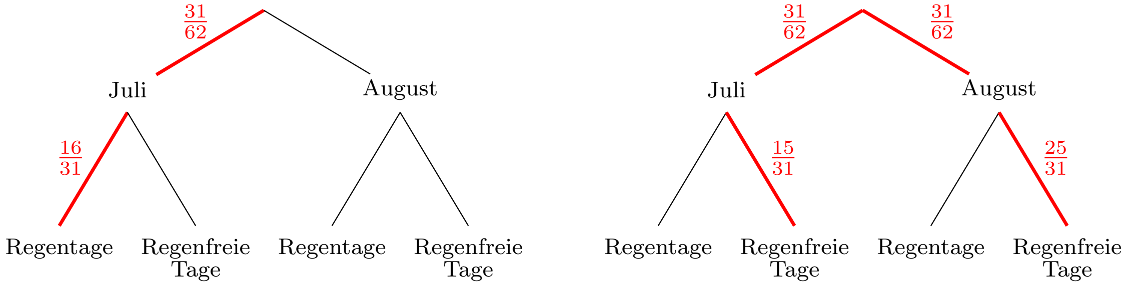
Auf dem AB – Relative Häufigkeiten und Baumdiagramme behandeln wir die folgenden Fragen:

Was sind **absolute Häufigkeiten**?

Was sind **relative Häufigkeiten**? Was sind **prozentuelle Häufigkeiten**?

Was sind **Baumdiagramme** für relative Häufigkeiten?

Was sind die **Pfadregeln** in Baumdiagrammen?



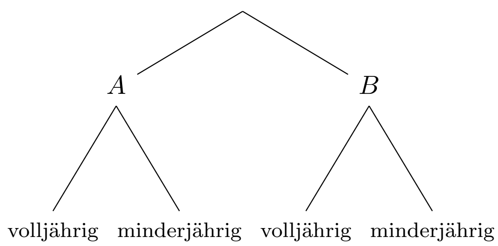
Beispiel 1.1. In einem Stadion sehen sich 21 600 Personen ein Match zwischen Team A und Team B an.
$\frac{11}{18}$ der Personen sind Fans von Team A, die anderen Personen sind Fans von Team B.
$\frac{7}{40}$ der Fans von Team A sind minderjährig.
Unter den Fans von Team B befinden sich 7140 volljährige Personen.

Im Folgenden sind mit „Personen im Stadion“ immer ausschließlich die Fans gemeint.

a) Vervollständige die Tabelle mit den absoluten Häufigkeiten.

	volljährig	minderjährig	Summe
A			
B			
Summe			

b) Beschrifte das dargestellte Baumdiagramm mit den relativen Häufigkeiten.



c) Welcher relative Anteil der Personen im Stadion sind Fans von Team A und minderjährig?

d) Welcher relative Anteil der Personen im Stadion ist volljährig?

e) Welcher relative Anteil der volljährigen Personen im Stadion sind Fans von Team B?

Lösung.

a) $21\,600 \cdot \frac{11}{18} = 13\,200$ der Personen im Stadion sind Fans von Team A. $13\,200 \cdot \frac{7}{40} = 2310$ dieser Fans sind minderjährig. Die anderen $13\,200 - 2310 = 10\,890$ Fans sind volljährig.

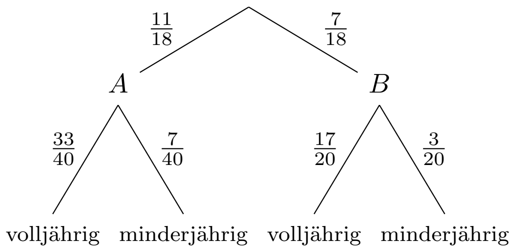
$21\,600 - 13\,200 = 8400$ der Personen im Stadion sind Fans von Team B. Da 7140 von ihnen volljährig sind, bleiben $8400 - 7140 = 1260$ dieser Fans, die minderjährig sind.

Die absoluten Häufigkeiten sind in der folgenden Tabelle zusammengefasst:

	volljährig	minderjährig	Summe
A	10 890	2310	13 200
B	7140	1260	8400
Summe	18 030	3570	21 600

b) Die *zwei* Entscheidungen Team A/Team B und volljährig/nicht volljährig stellen wir in einem *zweistufigen* Baumdiagramm dar.

Von der ersten Stufe kennen wir die relative Häufigkeit $\frac{11}{18}$ von Team A. Die relative Häufigkeit von Team B ist also $1 - \frac{11}{18} = \frac{7}{18}$.



Genauso können wir die relativen Häufigkeiten für die zweite Stufe von Team A berechnen.

Für die zweite Stufe von Team B verwenden wir, dass 7140 der 8400 Fans von Team B volljährig sind. Die relative Häufigkeit ist also $\frac{7140}{8400} = \frac{17}{20}$.

c) Wir haben zwei verschiedene Lösungswege:

1) Lösung mit absoluten Häufigkeiten: Insgesamt sind 21 600 Personen im Stadion. Davon sind 2310 Personen minderjährig und Fans von Team A. Der relative Anteil ist also $\frac{2310}{21600} = \frac{77}{720}$.

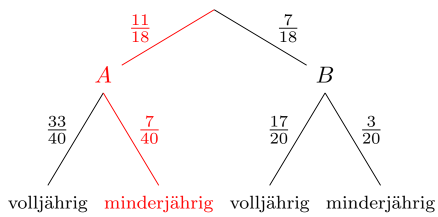
2) Lösung mit relativen Häufigkeiten:

Multiplikationssatz

$\frac{11}{18}$ *von* allen Personen sind Fans von Team A. *Davon* sind $\frac{7}{40}$ minderjährig.

Der relative Anteil aller Personen, die Fans von Team A und nicht volljährig sind, beträgt also

$$\frac{11}{18} \cdot \frac{7}{40} = \frac{77}{720}.$$



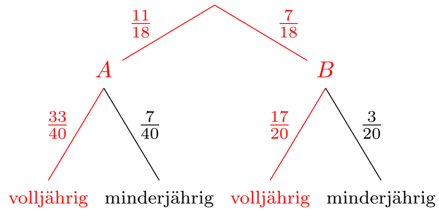
d) 1) Lösung mit absoluten Häufigkeiten: Von den 21 600 Personen im Stadion sind 18 030 Personen volljährig. Der gesuchte relative Anteil ist also $\frac{18030}{21600} = \frac{601}{720}$.

2) Lösung mit relativen Häufigkeiten:

Additionssatz

$\frac{11}{18} \cdot \frac{33}{40}$ aller Personen im Stadion sind Fans von Team A und volljährig. $\frac{7}{18} \cdot \frac{17}{20}$ aller Personen im Stadion sind Fans von Team B und volljährig. Der relative Anteil der volljährigen Personen im Stadion ist

$$\frac{11}{18} \cdot \frac{33}{40} + \frac{7}{18} \cdot \frac{17}{20} = \frac{601}{720}.$$



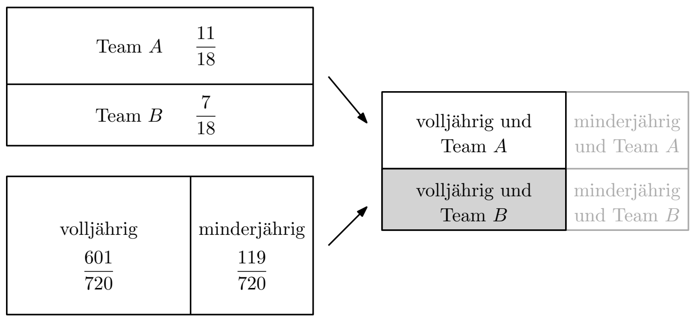
e) 1) Lösung mit absoluten Häufigkeiten: Diesmal ist nicht ein relativer Anteil an *allen* Personen im Stadion gesucht, sondern nur ein relativer Anteil an den volljährigen Personen. Unter den 18030 volljährigen Personen befinden sich 7140 Fans von Team B.
Der gesuchte relative Anteil ist also $\frac{7140}{18030} = \frac{238}{601}$.

2) Lösung mit relativen Häufigkeiten:

Einschränkung der Grundmenge

Von **d)** wissen wir, dass $\frac{601}{720}$ der Personen im Stadion volljährig sind.
Volljährig *und* Fan von Team B sind $\frac{7}{18} \cdot \frac{17}{20}$ der Personen im Stadion.
Der gesuchte relative Anteil ist also

$$\frac{\frac{7}{18} \cdot \frac{17}{20}}{\frac{601}{720}} = \frac{238}{601}.$$



□

Wenn wir die absoluten Häufigkeiten kennen, sind wir mit Lösungsmethode **1)** schneller. Kennen wir nur die relativen Häufigkeiten, sind diese Aufgaben mit Methode **2)** aber immer noch lösbar.

Entweder – Oder

Jede Verzweigung in einem Baumdiagramm zeigt eine *Aufteilung* in *verschiedene* Gruppen an. Weshalb können wir die Gruppen

A = „Jugendliche, die gerne Musik hören“ und B = „Jugendliche, die gerne Sport machen“

wohl *nicht* als Verzweigung in einem Baumdiagramm darstellen?

Beispiel 1.2. In einer Klasse wird eine Umfrage durchgeführt. Jede Schülerin und jeder Schüler markiert mit einem Strich auf der Tafel die Anzahl der eigenen Geschwister:

Anzahl Geschwister	Strichliste	Absolute Häufigkeit	Relative Häufigkeit	Prozentuelle Häufigkeit
0	~~IIII~~ III			
1	~~IIII~~ ~~IIII~~ II			
2	III			
3	II			
Summe				

1) Trage die absoluten, relativen und prozentuellen Häufigkeiten in der Tabelle ein.
2) Erstelle ein Säulen- oder Balkendiagramm mit den absoluten Häufigkeiten.
3) Erstelle ein Stabdiagramm mit den relativen Häufigkeiten.
4) Erstelle ein Kreisdiagramm mit den prozentuellen Häufigkeiten.

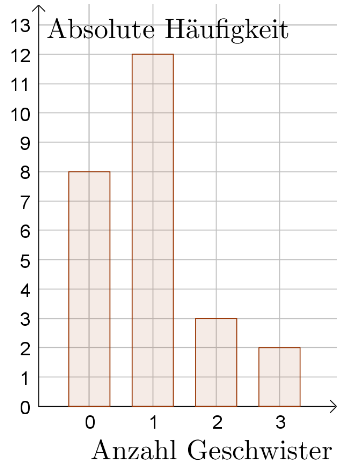
Lösung.

1)

Anzahl Geschwister	Strichliste	Absolute Häufigkeit	Relative Häufigkeit	Prozentuelle Häufigkeit
0	~~IIII~~ III	8	$\frac{8}{25} = 0{,}32$	32 %
1	~~IIII~~ ~~IIII~~ II	12	$\frac{12}{25} = 0{,}48$	48 %
2	III	3	$\frac{3}{25} = 0{,}12$	12 %
3	II	2	$\frac{2}{25} = 0{,}08$	8 %
Summe		25	$\frac{25}{25} = 1$	100 %

2) Für jede Ausprägung des Merkmals (hier: 0, 1, 2, 3) zeichnen wir ein Rechteck.

Bei einem **Säulendiagramm** tragen wir die Ausprägungen auf der waagrechten Achse und die Häufigkeiten auf der senkrechten Achse auf. Vergiss nicht auf die Beschriftung der Achsen.
Die Rechtecke wachsen also wie *Säulen* nach oben. Da die Anzahl der Geschwister nur ganzzahlig sein kann, lassen wir zwischen benachbarten Säulen ein wenig Platz.

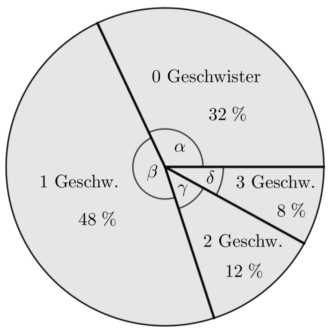


Bei einem **Balkendiagramm** vertauschen wir die beiden Achsen. Die Rechtecke liegen dann also wie *Balken* waagrecht am Boden.

3) Ein **Stabdiagramm** ist ein Säulendiagramm mit schmalen Säulen.

4) Um die prozentuellen Häufigkeiten in einem **Kreisdiagramm** darzustellen, berechnen wir jeweils den Zentriwinkel der Sektoren:

$$\alpha = 360° \cdot 32\,\% = 115{,}2°$$
$$\beta = 360° \cdot 48\,\% = 172{,}8°$$
$$\gamma = 360° \cdot 12\,\% = 43{,}2°$$
$$\delta = 360° \cdot 8\,\% = 28{,}8°$$



□

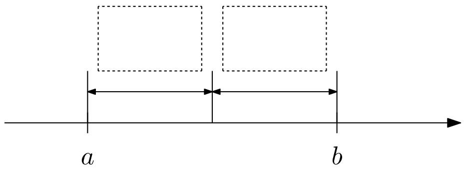
2. Statistische Kenngrössen & Boxplot

Arbeitsblatt – Statistische Kenngrößen und Boxplot

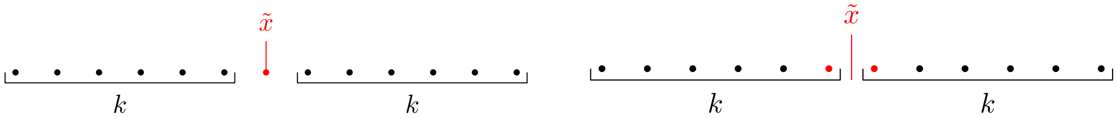
Auf dem AB – Statistische Kenngrößen und Boxplot behandeln wir die folgenden Fragen:

Was sind **arithmetisches Mittel**, **Varianz** und **Standardabweichung** einer Zahlenliste?

Warum liegt das arithmetische Mittel von zwei Zahlen auf der Zahlengerade genau in der Mitte zwischen den Zahlen?



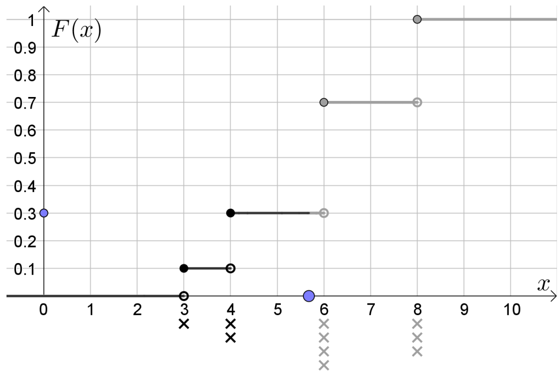
Was ist der **Median** einer Zahlenliste? Wie kann man ihn ermitteln?



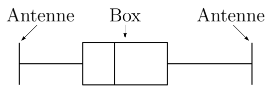
Was ist die **empirische Verteilungsfunktion** einer Zahlenliste?

Was ist ein ***p*-Quantil** einer Zahlenliste?

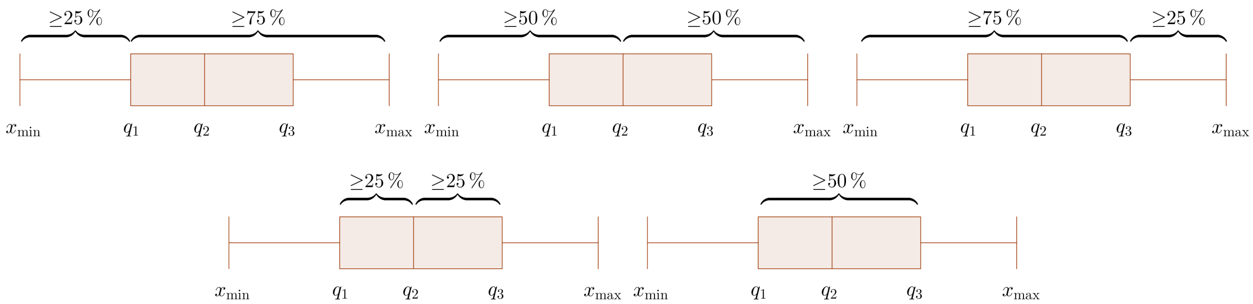
Was sind **Quartile** einer Zahlenliste? Welche Eigenschaften haben sie?



Wie kann man aus einer Zahlenliste einen **Boxplot** erstellen?



Welche Aussagen kann man mit einem Boxplot über die zugrunde liegende Zahlenliste treffen?



Was sind **Spannweite** und **Interquartilsabstand** einer Zahlenliste?

Beispiel 2.1. An sechs aufeinander folgenden Tagen wurde um 12 Uhr die Temperatur gemessen:

$$(21\,°\mathrm{C}, 24\,°\mathrm{C}, 25\,°\mathrm{C}, 24\,°\mathrm{C}, 17\,°\mathrm{C}, 19\,°\mathrm{C})$$

a) Berechne das arithmetische Mittel $\bar{x}$ und den Median $\tilde{x}$ der Messwerte.

Beim Abtippen der Messwerte ist ein Fehler passiert: $(212\,°\mathrm{C}, 24\,°\mathrm{C}, 25\,°\mathrm{C}, 24\,°\mathrm{C}, 17\,°\mathrm{C}, 19\,°\mathrm{C})$.

b) Berechne das neue arithmetische Mittel $\bar{x}_F$. Um wie viel Prozent ist $\bar{x}_F$ größer als $\bar{x}$?

c) Berechne den neuen Median $\tilde{x}_F$. Um wie viel Prozent ist $\tilde{x}_F$ größer als $\tilde{x}$?

Lösung.

a) Das arithmetische Mittel der Messdaten ist $\bar{x} = \dfrac{21 + 24 + 25 + 24 + 17 + 19}{6} = 21{,}66\ldots\,°\mathrm{C}$.

Der Median von $(17, 19, 21, 24, 24, 25)$ ist $\tilde{x} = \dfrac{21 + 24}{2} = 22{,}5\,°\mathrm{C}$.

b) Das neue arithmetische Mittel ist $\bar{x}_F = \dfrac{212 + 24 + 25 + 24 + 17 + 19}{6} = 53{,}5\ldots\,°\mathrm{C}$.

Es ist $\dfrac{\bar{x}_F}{\bar{x}} = 2{,}469\ldots = 246{,}9\ldots\,\%$, also ist $\bar{x_F}$ um $146{,}9\ldots\,\%$ größer als $\bar{x}$.

c) Der Median von $(17, 19, 24, 24, 25, 212)$ ist $\tilde{x}_F = 24\,°\mathrm{C}$.

Es ist $\dfrac{\tilde{x}_F}{\tilde{x}} = 1{,}066\ldots = 106{,}6\ldots\,\%$, also ist $\tilde{x}_F$ um $6{,}6\ldots\,\%$ größer als $\tilde{x}$. □

Ausreißer

Wenn eine Zahl in einer Zahlenliste stark vom Großteil der anderen der Zahlen abweicht, wird von einem **Ausreißer** gesprochen.

Eine mögliche Definition ist, dass ein Ausreißer mehr als das 1,5-fache vom Quartilsabstand außerhalb von $[q_1; q_3]$ liegt.

Einfluss von Ausreißern

Was ist der Median der Liste $(0, 0, 0, 100, 100, 100)$?
Die Zahl 420 wird zur Liste hinzugefügt. Was ist der Median der neuen Liste?

Vor dir liegt eine Liste mit mindestens 4 Zahlen.
Du berechnest den Median q_2, ein unteres Quartil q_1 und ein oberes Quartil q_3.
Eine Zahl wird zur Liste hinzugefügt.
Erkläre, warum der Median der neuen Liste in $[q_1; q_3]$ liegt.
Ein Ausreißer kann den Median also *nicht beliebig* stark verändern.

Erkläre, warum ein Ausreißer das arithmetische Mittel *beliebig* stark verändern kann.

Odd

„Mindestens 25 % der Werte sind $\leq q_1$, und mindestens 75 % der Werte sind $\geq q_1$.“
Wir haben gesehen, dass diese Forderung die Zahl q_1 nicht immer eindeutig festlegt.
Tatsächlich kannst du je nach Lehrbuch, Taschenrechner und Mathematik-Software unterschiedliche Berechnungsmethoden für q_1 und q_3 finden.

Vorsicht! Seit Jahrzehnten sind auch Methoden im Umlauf, die *nicht* zuverlässig 25 %-Quantile bzw. 75 %-Quantile liefern:

- Methode 1: Ist n ungerade, dann zähle den mittleren Wert zu keiner Hälfte dazu.
  Für die Liste $(1, 2, 3, 4, 5)$ wird mit dieser Methode also $q_1 = 1{,}5$ berechnet.
  Warum ist $q_1 = 1{,}5$ kein 25 %-Quantil? Allgemein kann diese Methode bei $n = 4 \cdot \ell + 1$ scheitern.
- Methode 2: Ist n ungerade, dann zähle den mittleren Wert zu beiden Hälfte dazu.
  Für die Liste $(1, 2, 3, 4, 5, 6, 7)$ wird mit dieser Methode also $q_1 = 2{,}5$ berechnet.
  Warum ist $q_1 = 2{,}5$ kein 25 %-Quantil? Allgemein kann diese Methode bei $n = 4 \cdot \ell + 3$ scheitern.
- Methode 3: Ist $n = 4 \cdot \ell + 1$, dann ist q_1 eine gewichtete Summe:
  q_1 ist 25 % vom ℓ-ten Wert plus 75 % vom $(\ell + 1)$-ten Wert der sortierten Liste.
  Für die Liste $(1, 2, 3, 4, 5)$ wird mit dieser Methode also $q_1 = 1{,}75$ berechnet.
  Warum ist $q_1 = 1{,}75$ kein 25 %-Quantil?

Berechnung der Quartile

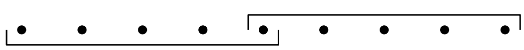
Wir können Quartile q_1, q_2 und q_3 einer aufsteigend sortierten Listen wie folgt berechnen:

1) Für q_2 verwenden wir den Median der Werte.

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

- Ist ***n* gerade**, dann teilen wir die Liste in der Mitte auf:

- Ist ***n* ungerade**, dann teilen wir die Liste so auf, dass auch beide Teillisten eine *ungerade* Anzahl an Werten enthalten. Dafür musst du den mittleren Wert entweder zu beiden Teillisten dazunehmen oder jeweils *nicht* dazunehmen:

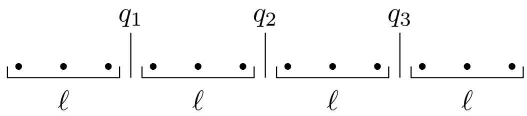

oder


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

Warum liefert diese Methode wirklich Quartile?

Die vorgestellte Berechnungsmethode liefert immer 25 %-, 50 %- bzw. 75 %-Quantile:

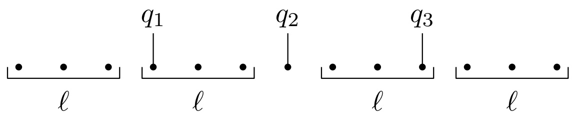
1) Wenn $n = 4 \cdot \ell$ ist, dann liefert unsere Methode die folgenden Werte:



Es ist $\lceil 25\,\% \cdot n \rceil = \ell$, $\lceil 50\,\% \cdot n \rceil = 2 \cdot \ell$ und $\lceil 75\,\% \cdot n \rceil = 3 \cdot \ell$.

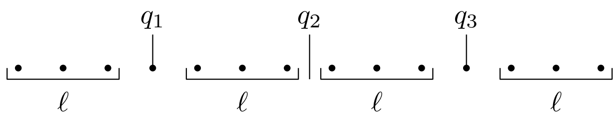
$\lceil x \rceil$ ist die kleinste ganze Zahl $\geq x$.

2) Wenn $n = 4 \cdot \ell + 1$ ist, dann liefert unsere Methode die folgenden Werte:



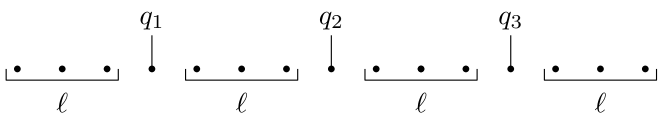
Es ist $\lceil 25\,\% \cdot n \rceil = \ell + 1$, $\lceil 50\,\% \cdot n \rceil = 2 \cdot \ell + 1$ und $\lceil 75\,\% \cdot n \rceil = 3 \cdot \ell + 1$.

3) Wenn $n = 4 \cdot \ell + 2$ ist, dann liefert unsere Methode die folgenden Werte:



Es ist $\lceil 25\,\% \cdot n \rceil = \ell + 1$, $\lceil 50\,\% \cdot n \rceil = 2 \cdot \ell + 1$ und $\lceil 75\,\% \cdot n \rceil = 3 \cdot \ell + 2$.

4) Wenn $n = 4 \cdot \ell + 3$ ist, dann liefert unsere Methode die folgenden Werte:

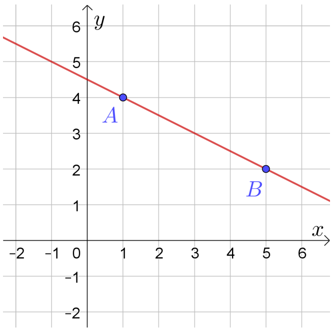


Es ist $\lceil 25\,\% \cdot n \rceil = \ell + 1$, $\lceil 50\,\% \cdot n \rceil = 2 \cdot \ell + 2$ und $\lceil 75\,\% \cdot n \rceil = 3 \cdot \ell + 3$.

3. Interpolation, Regression & Ausgleichsfunktionen

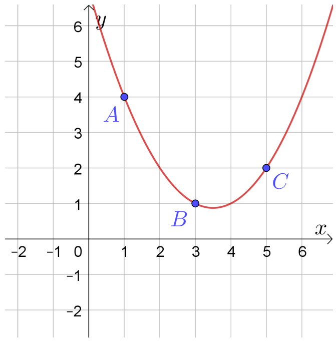
Arbeitsblatt – Interpolation und Regression

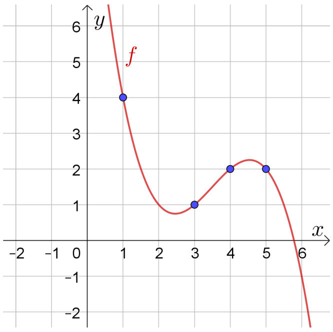
Auf dem Arbeitsblatt – Interpolation und Regression behandeln wir die folgenden Fragen:



Was ist **lineare Interpolation**?

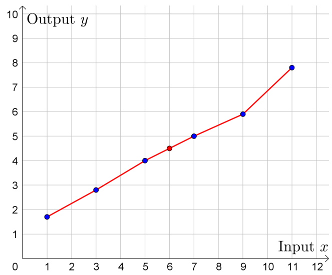
Was ist **quadratische Interpolation**?





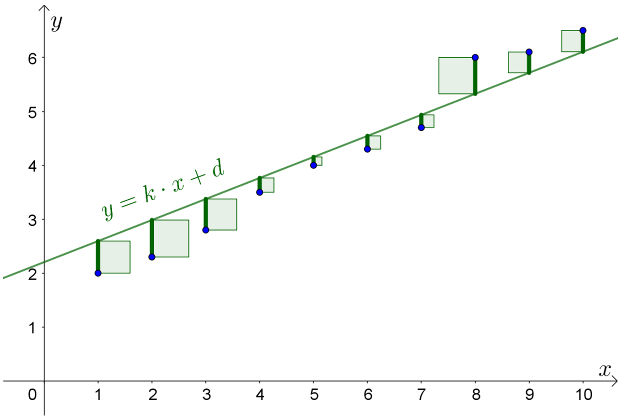
Was ist **Polynominterpolation**?

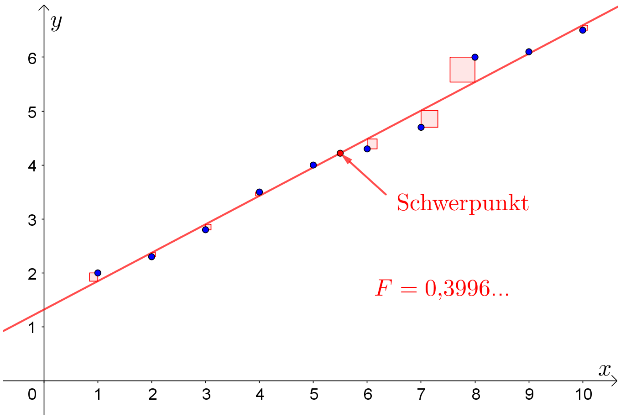
Was ist **stückweise lineare Interpolation**?



Wodurch unterscheiden sich **Interpolation** und **Extrapolation**?

Was ist **lineare Regression**? Was ist eine **Fehlerquadratsumme**?





Technologieblatt – Regression

Auf dem Technologieblatt – Regression behandeln wir die folgenden Fragen:

Wie kann man **Ausgleichsfunktionen** mit Technologieeinsatz ermitteln?

Was ist der **Korrelationskoeffizient** nach Pearson? Wie kann man ihn berechnen? Welche Eigenschaften hat er?