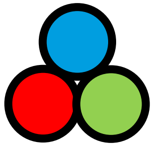
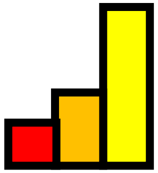
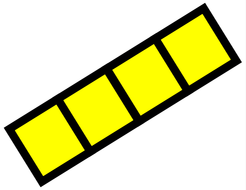
Diagramme

Wozu Diagramme und beschreibende Statistik?

Um einen erster Eindruck über Daten zu gelangen, Tendenzen erkennen. Fehler in Daten erkennen. Hintergrundinformationen zur eigenen Statistik geben

Skalenniveaus und Grafiken

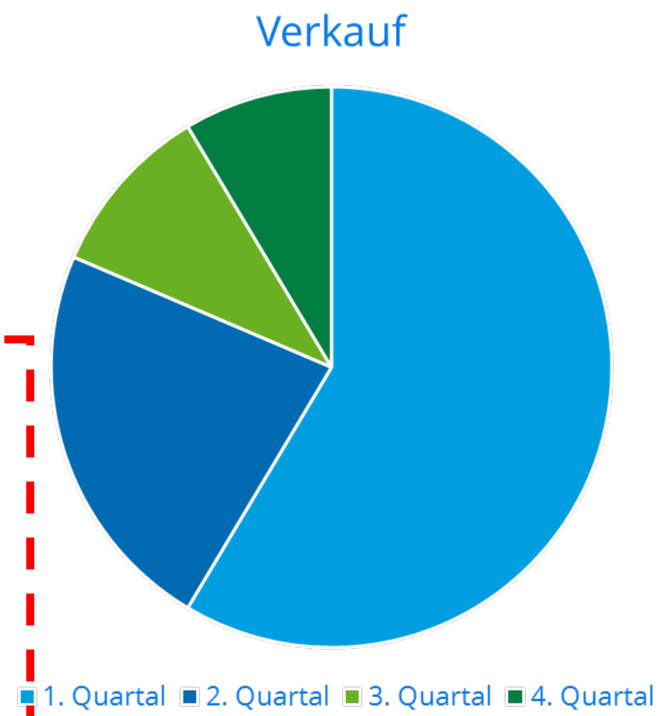
Hier habe ich die Tabelle erweitert. Nicht jede Grafik eignet sich für jedes Skalenniveau.

Grafik	Skala	Merkmale	Grafik
	Nominal	Klassifizierung qualitativer Merkmale	Kreisdiagramme, Balkendiagram, Häufigkeitstabelle
	Ordinal	Rangwerte mit Ordinalzahlen	Balkendiagram, Häufigkeitstabelle, Box Plot
	Metrisch	Gleiche Abstände	Histogramm, Box Plot

Kreisdiagramm

Für Darstellung von Anteilen

Für nominale, kategoriale und metrische Daten



Aufgabe: Kreisdiagramm

Skizzieren Sie ein Kreisdiagramm für eine beliebige nominelle, kategoriale und (gruppierte) metrische Variablen mit mit jeweils 10 Fällen.

- Nominal: Berufsgruppen von Befragten: Anzahl der Nennungen
- Kategorial: Anteil an Schulabschlüsse von Befragten
- Metrisch: Alter von Befragten

Zeit: 10 Minuten Abgabeformat: Zeichenprogramm

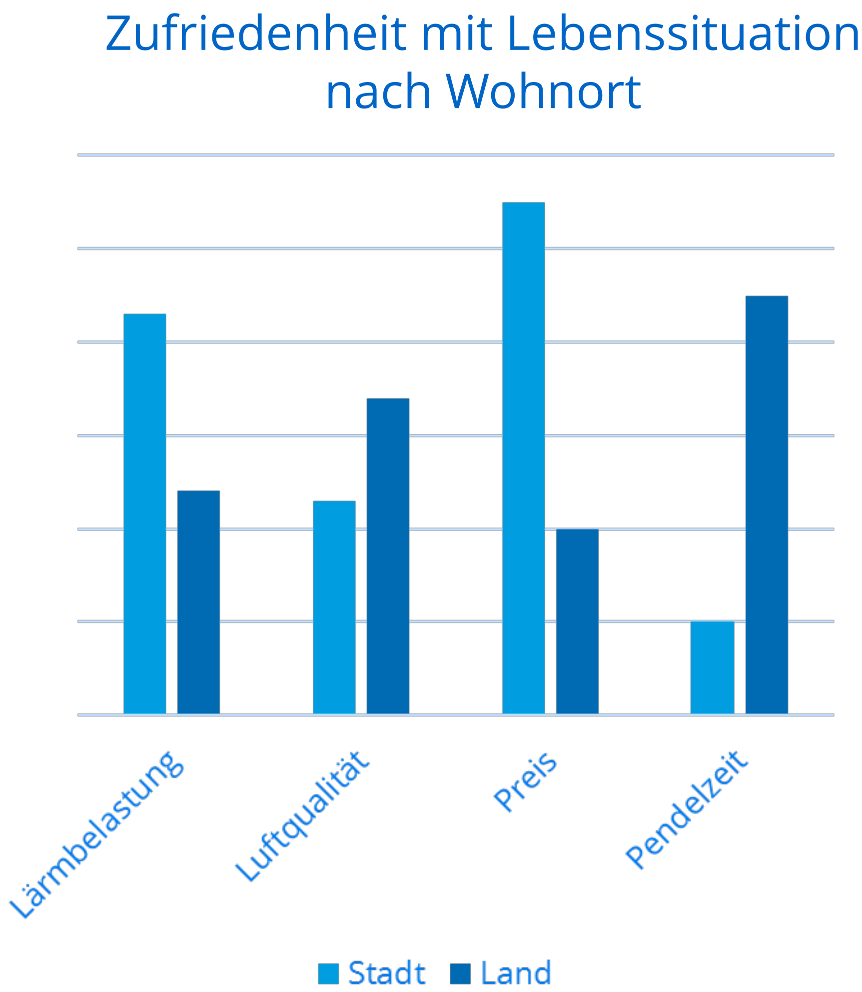
Balkendiagramm (auch Säulendiagramm)

Zeigt die Häufigkeit einer (oder mehrere) Variable in verschiedenen Kategorien an. Nominal und ordinale Variablen eignen sich sehr gut, es können aber auch metrische dargestellt werden.

Aufgabe: Balkendiagramm

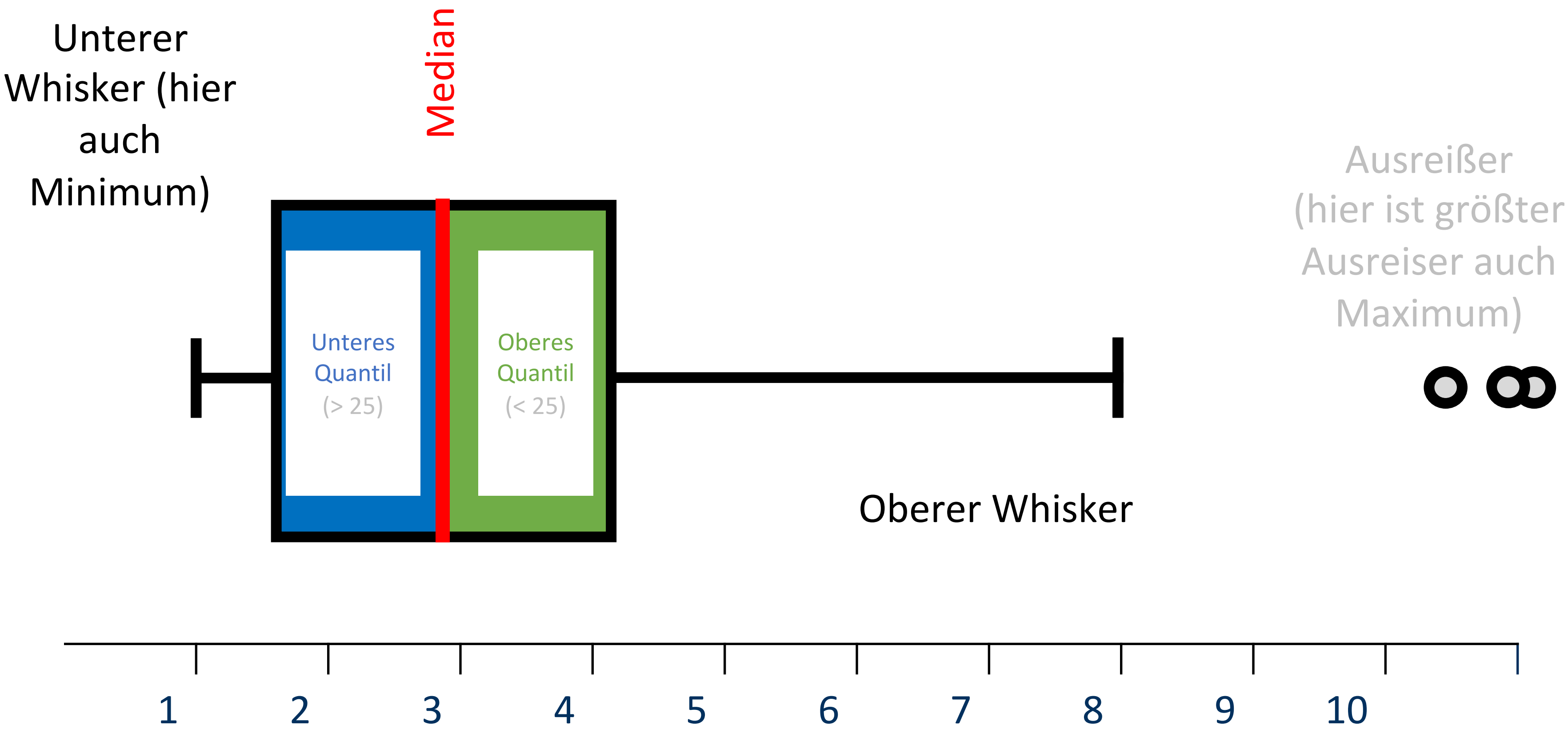
Skizieren Sie ein Balkendiagramm in den einmal die Anzahl und einmal der Mittelwert jeweils einer Variable abgebildet wird. Erstellen Sie 10 Fälle.

Zeit: 10 Minuten Abgabeformat: Zeichenprogramm

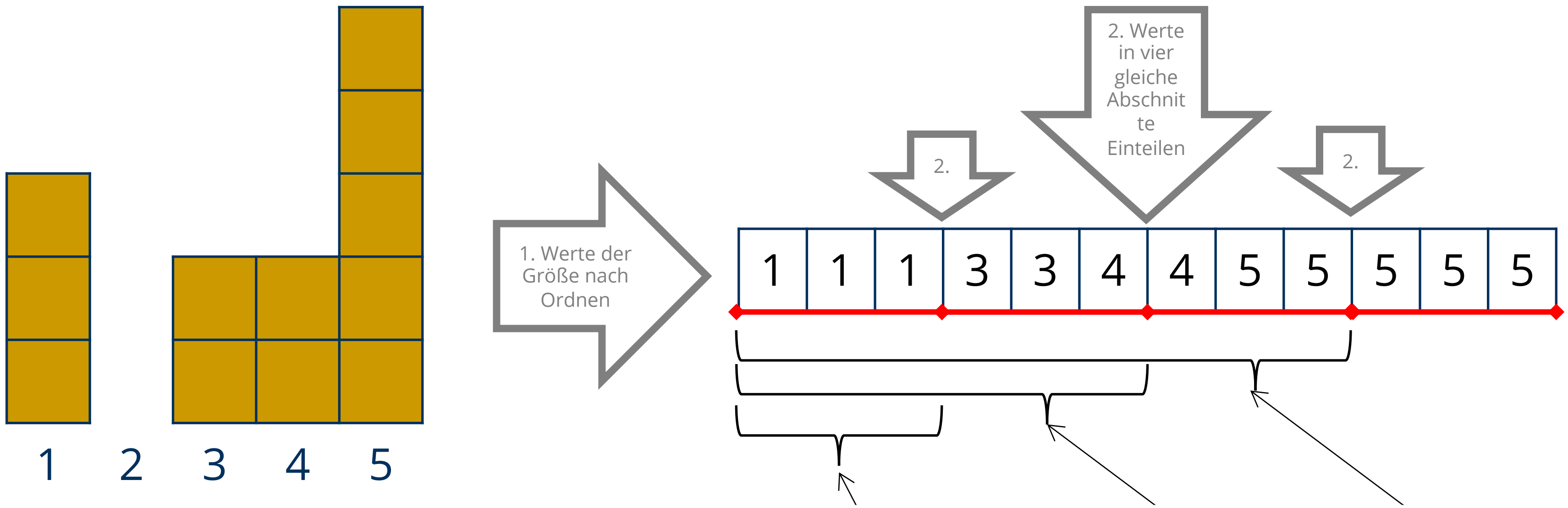


Boxplots (Kastengrafik)

Eine übliche Definition ist, dass die Antennen maximal bis zum 1,5-fachen der Boxgröße gehen (also 1,5-fache des Interquartilabstands
Hinweis für Fortgeschrittene: Was ein Ausreiße ist, wird unterschiedlich definiert.
https://de.wikipedia.org/wiki/Box-Plot
Für unsere Veranstaltung behandeln wir die Ausreißer nicht gesondert.



Quantile: unterteilen die *geordneten* Daten in Gruppen. In jeder Gruppe ist jeweils ein bestimmter Prozentsatz der Daten enthalten



Wichtige Quantile:

- 25-Prozent-Perzentil (untere Quartil/1. Quartil): 25 % aller Fälle liegen unter dem Werte „1“.
- 50-Prozent-Perzentil (2. Quartil): 50 % aller Fälle nehmen Werte bis „4“ an.
- 75-Prozent-Perzentil (obere Quartil/3. Quartil): 75 % aller Fälle nehmen Werte bis „5“ an.

Quantlie sind eigentlich Streuungsmaße, wir brauchen Sie aber schon für die Diagramme, weshalb sie hier dabei sind.

Übung Quantil:

Ermitteln Sie für folgende Werte das 25 %, 50 % und 75 % Prozent Quantil.

Reihe A:

1 1 1 1 1 1 2 2 2 2 3 3 3 3 3 4 4 4 4 5 5 5 5 5

Reihe B:

23 34 45 12 23 12 45 23 12 19 13 15

Reihe C:

Ihre eigene. Lösen von Kommilitonen.

Dauer: 4 Minuten **Abgabe:** Zeichenprogramm

Boxplot zeichnen mit 15 Daten (einfachste Version)

Rohdaten: 4 2 2 5 2 4 4 4 4 5 3 5 2 5 1

Geordnete Daten: 1 2 2 2 2 3 4 4 5 5 5 5

Quantile:

Beispiele für Reihen

Merkregel für die Farbreihen	Globaler Median <u>real</u> *Wird bei der Berechnung der Quartile <u>nicht</u> berücksichtigt.*	Globaler Median wird errechnen *Anzahl der Wert ist durch 2 teilbar.*
25% und 75 % Quartil existiert <u>real</u>		
25% und 75 % Quartil wird errechnet		

1

1 | 2

1 2 3

1 | 2 | 3 | 4

1 | 2 **3** 4 | 5

1 **2** 3 | 4 **5** 6

1 **2** 3 **4** 5 **6** 7

1 2 | 3 4 | 5 6 | 7 8

1 2 | 3 4 **5** 6 7 | 8 9

1 2 **3** 4 5 | 6 7 **8** 9 0

1 2 **3** 4 5 **6** 7 8 **9** 0 1

1 2 3 | 4 5 6 | 7 8 9 | 0 1 2

1 2 3 | 4 5 6 **7** 8 9 0 | 1 2 3

1 2 3 **4** 5 6 7 | 8 9 0 **1** 2 3 4

1 2 3 **4** 5 6 7 **8** 9 0 1 **2** 3 4 5

1 2 3 4 | 5 6 7 8 | 9 0 1 2 | 3 4 5 6

1 2 3 4 | 5 6 7 8 **9** 0 1 2 3 | 4 5 6 7

1 2 3 4 **5** 6 7 8 9 | 0 1 2 3 **4** 5 6 7 8

1 2 3 4 **5** 6 7 8 9 **0** 1 2 3 4 **5** 6 7 8 9

1 2 3 4 5 | 6 7 8 9 0 | 1 2 3 4 5 | 6 7 8 9 0

1 2 3 4 5 | 6 7 8 9 0 **1** 2 3 4 5 6 | 7 8 9 0 1

1 2 3 4 5 **6** 7 8 9 0 1 | 2 3 4 5 6 **7** 8 9 0 1 2

1 2 3 4 5 **6** 7 8 9 0 1 **2** 3 4 5 6 7 **8** 9 0 1 2 3

1 2 3 4 5 6 | 7 8 9 0 1 2 | 3 4 5 6 7 8 | 9 0 1 2 3 4

Legende
Der Median findet sich im Zentrum und ist fett oder (wenn die Zahl errechnet wird) als | markiert.
Das 25% und 72%-Quartil sind jeweils auch fett oder als | markiert.
Zahlen ab 10 sind in dieser schematischen Darstellung ohne ihre erste Ziffer dargestellt. Also 10 = 0; 11 = 1 usw. Ab 20 gilt das selbe.

Anwendungsbeispiel: Boxplots

Dies ist ein Beispiel aus meiner Arbeit. Die Punkte sind die Mittelwerte. Boxplots geben eine weiteren guten Blick auf Daten. Vor allem wenn man mehrere vergleichen will. Sie sind jedoch anspruchsvoller und werden seltener veröffentlicht.

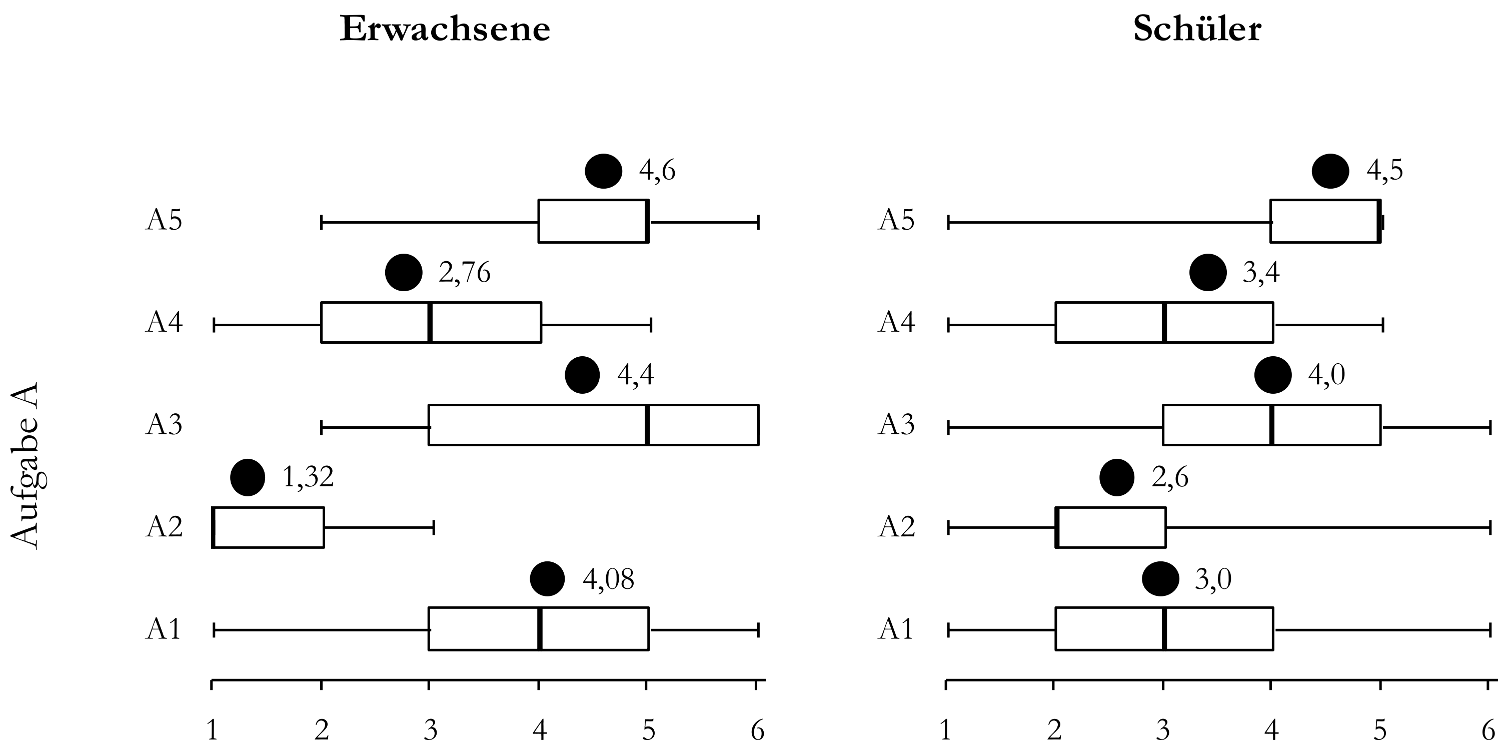


Abbildung: Boxplots für die einzelnen Verhaltensstrategien nach Erwachsenen und Schüler für die Aufgaben A–C (inklusive des Mittelwertes: der schwarzer Punkt oberhalb der jeweiligen Box).

Histogram

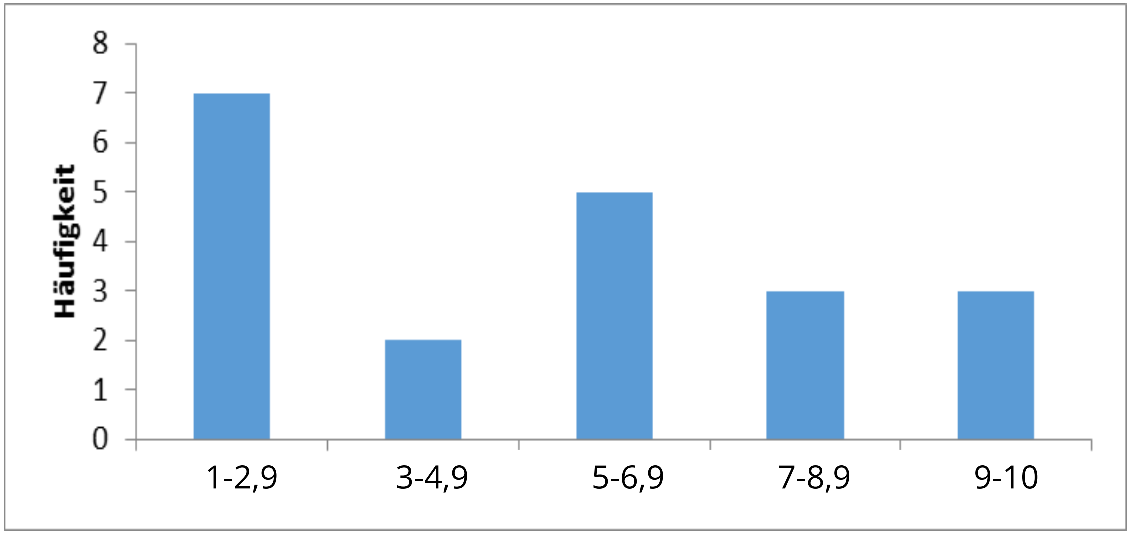
Ordnet die Werte einer Variablen der Größe nach (auf der X-Achse) und zeigt zu jedem Wert die Häufigkeit. Skalenniveau: ordinal/metrisch.

Y-Achse: Häufigkeiten

Da Histogramme für metrische stetige Daten gedacht sind, können auch die Zahlen 1,5; 1,51; 1,52 usw. auftreten. Da solche Werte meist nur einmal auftauchen, werden die Wert gruppiert. Im Folgenden z.B. werden alle Werte von 1-2,9 zusammengefasst

Geordnete Rohdaten

1,0	1,2	1,5	1,51	1,52
2,6	2,7	3,4	4,0	5,0
5,1	5,2	5,4	6,1	7,8
8,1	8,12	9,3	9,2	9,8



Übung: Skizzieren von Histogrammen

a)

1,3
2,1
1,2
2,4
1,6
1,0
3,4
4,8
1,1
2,1
1,7
1,8
1,1
2,5
3,7

b)

Werte	Häufigkeit
18	2
19	7
20	5
21	5
22	3
23	7
24	5
27	2
30	1
31	3
39	1

Aufgabe Histogramme:

1. Erstellen Sie anhand der Daten ein Histogramm.
2. Interpretieren Sie die Verteilung inhaltlich.

Hinweis: Beachten Sie, dass sie die notwendigen Kategorien für das Histogramm selbst definieren müssen.

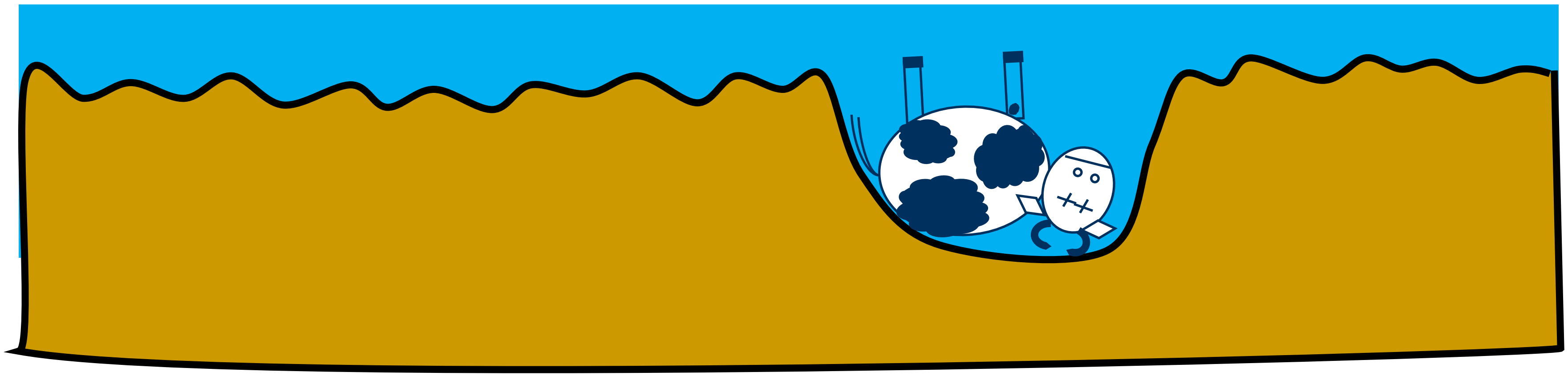
3. a) und b) sind nur unterschiedliche Darstellungsformen. Wie hießen die noch mal?

4. Was ist der Unterschied zwischen einem Histogramm und einem Balken/Säulendiagramm?

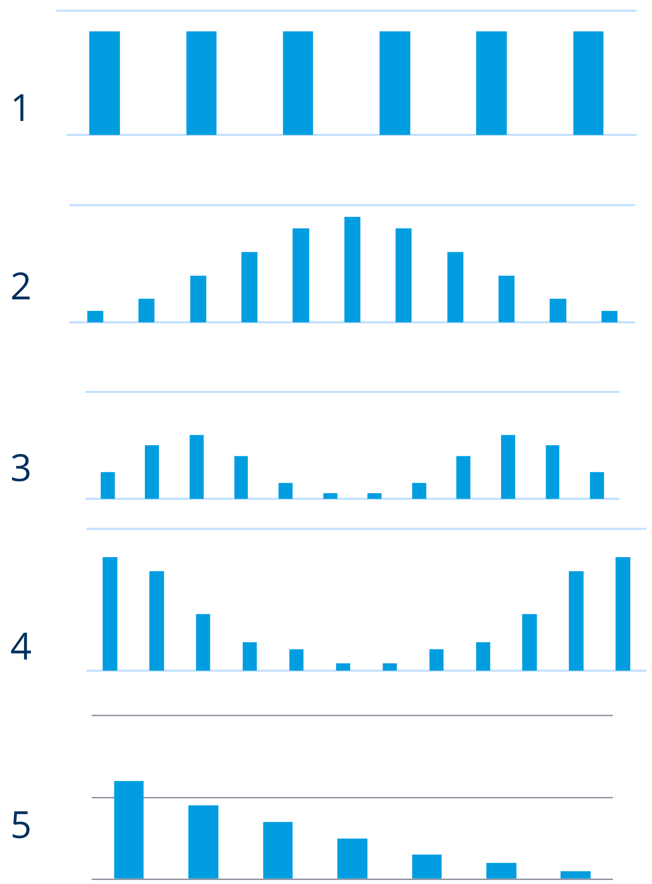
Dauer: 20 Minuten **Abgabeformat:** Zeichenprogramm

Verteilungen verstehen: Histogramme

Im Durchschnitt war der See einen Meter tief, und trotzdem ist die Kuh ertrunken.



Verschiedene Häufigkeitsverteilungen in Histogrammen



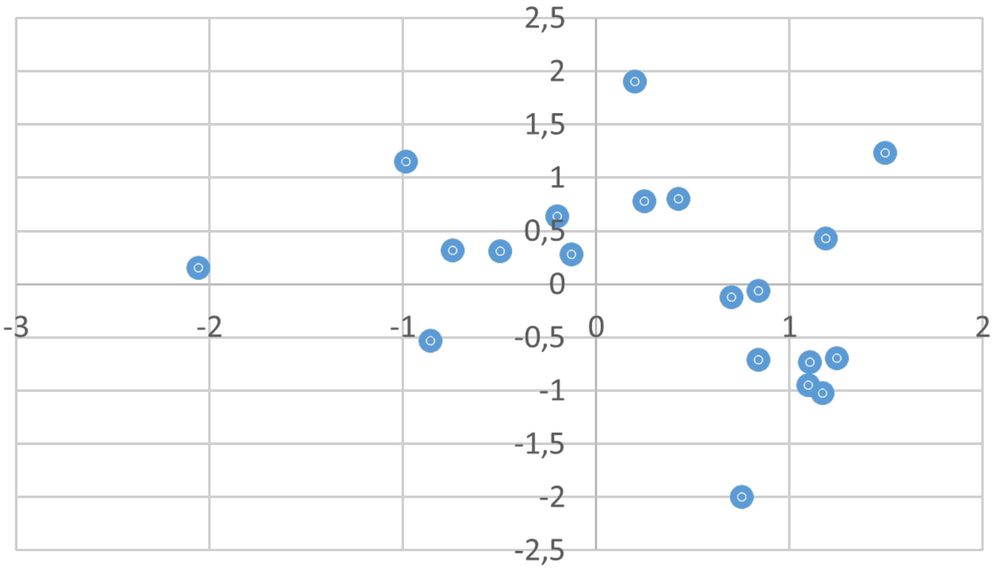
Hier sind einige typische Verteilungsformen: Es handelt sich dabei um idealförmige Verteilungen, die Sie fast nie in der Realität beobachten werden.

Aufgabe Histogramm und Verteilungen:
Ordnen Sie die die Namen richtig den Verteilungen zu:

A Zweigipflig
B Schiefe Verteilung
C Glockenkurve
D U-Förmige Verteilung
E Gleichverteilung

Dauer: 1 Minute **Abgabeformat:** mündlich

Streudiagram (Punktdiagram)



Die Werte **zweier** Variablen werden miteinander verglichen

Beide Variablen sind ordinal/metrisch

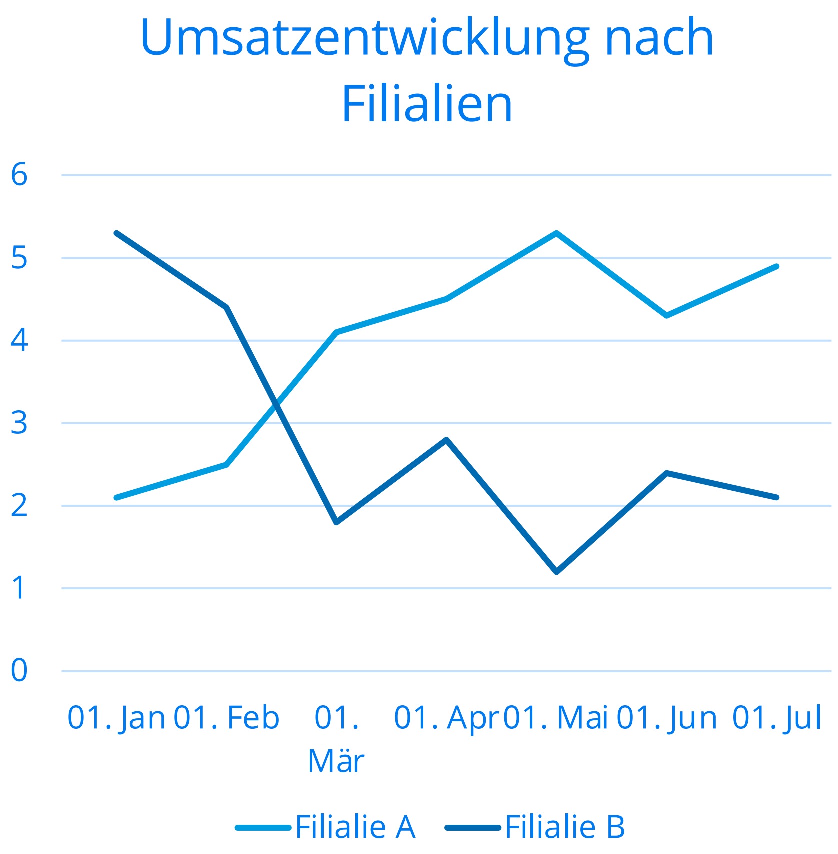
Liniendiagramme

Liniendiagramme sind gedacht für zeitliche Entwicklungen

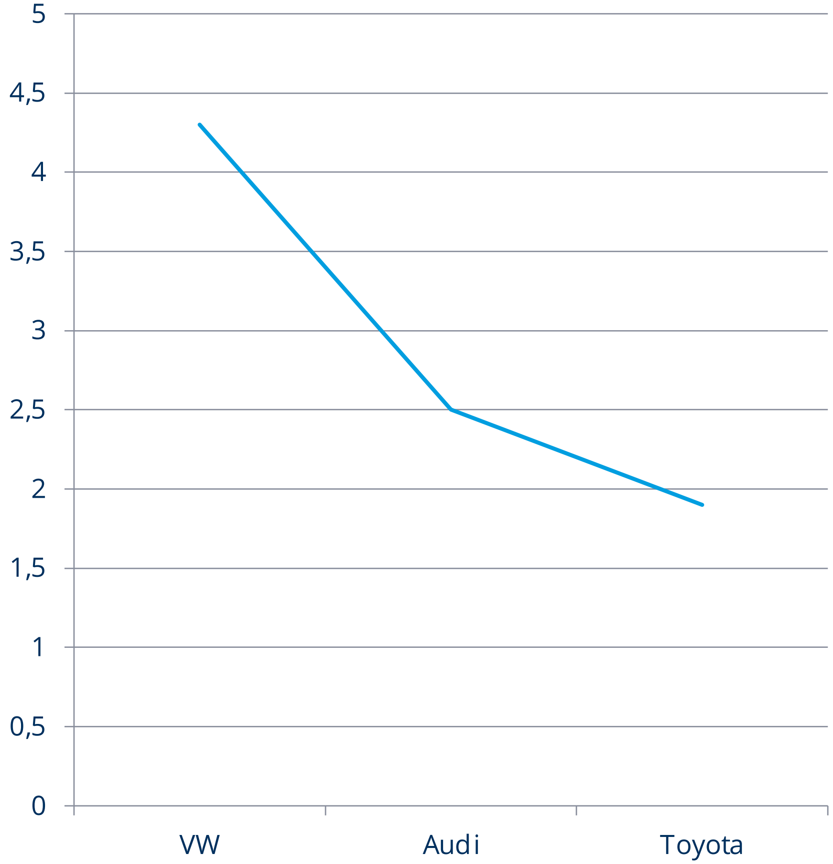
Aufgabe Linien:

Welche Schlussfolgerung lässt sich für die Praxis aus diesem Diagramm ziehen?

Dauer: 2 Minuten **Abgabeformat:** mündlich



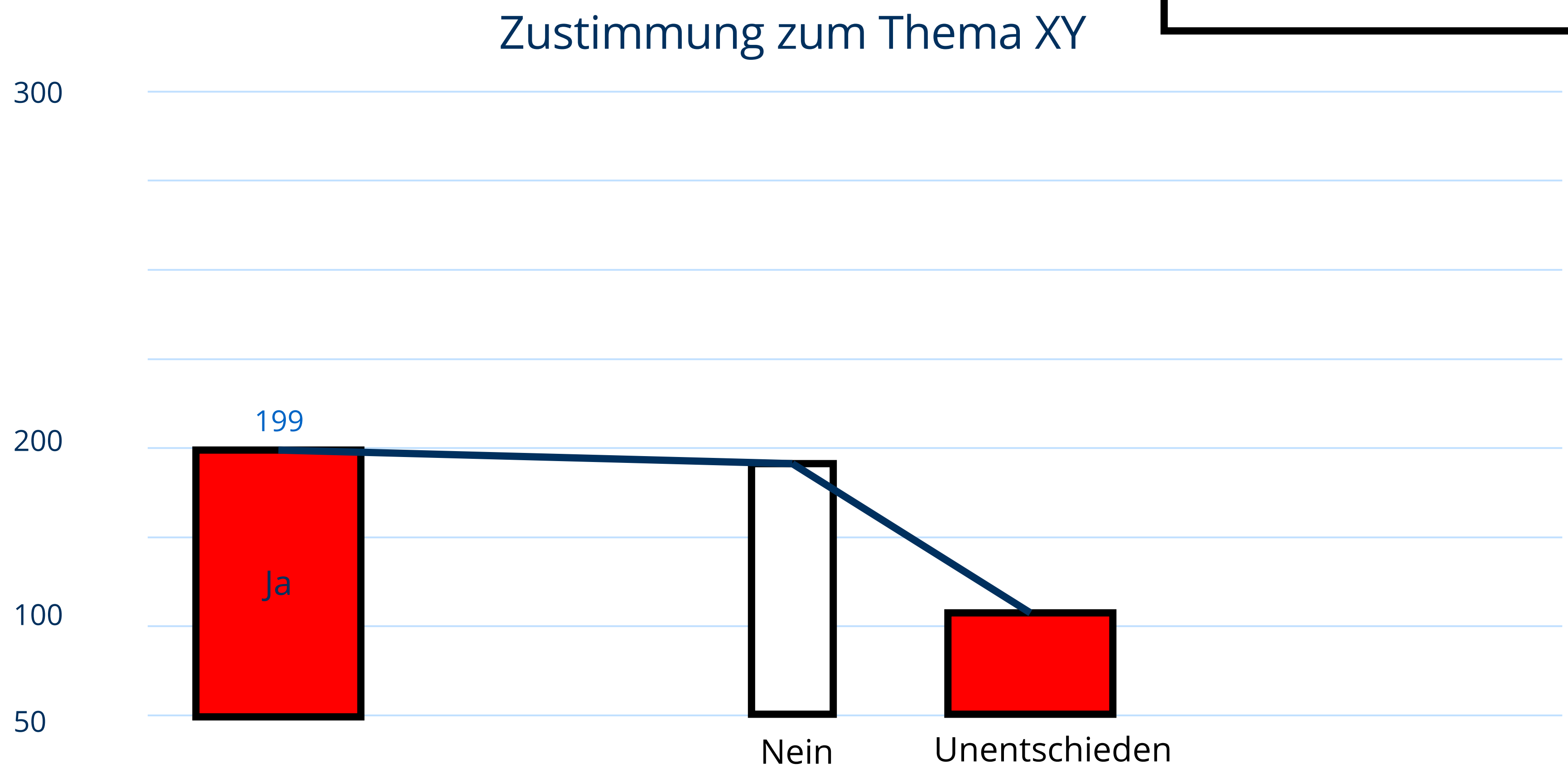
Schlechte Beispiele: Grafik



Was ist hier das Problem?

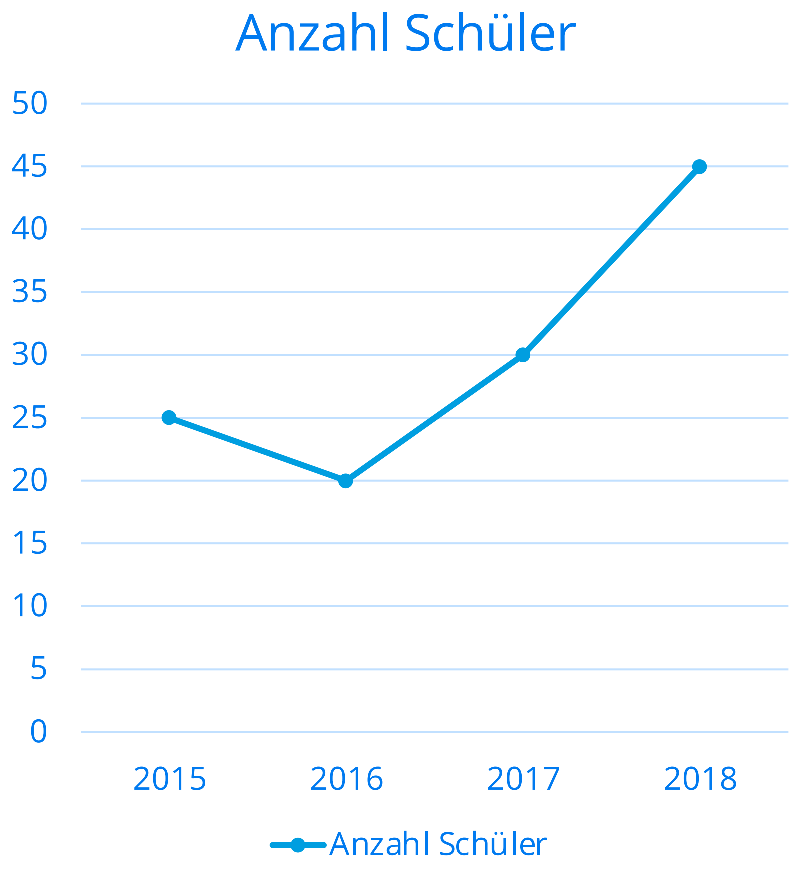
Schlechte Beispiele: Grafiken vermeiden

10 Fehler gibt es in dieser Grafik. Finden Sie diese.



Schlechte Beispiele: Trends zurechtbiegen

Was könnte die nebenstehende Grafik verheimlichen?



Aufgabe: Grafiken verwenden

Suchen Sie auf der Seite des Statistischen Bundesamtes eine Grafik aus. Sie müssen Sich dafür in die Themen durchklicken.

Erläutern Sie die Grafik Ihrer Gruppe. Teilen Sie dafür in Ihrem Gruppenchat den Link und geben Sie gegebenfalls weiter Anweisungen, um Ihre Grafik zu erreichen.

Bearbeiten Sie bei folgende Aufgaben

a) Um welche Art von Grafik handelt es sich (Säulen, Linien, Histogramm, Kreis)

b) Was ist auf den Achsen abgebildet (bei Kreisdiagramme, was bedeuten die „Kuchenteile“ usw.)

c) Ziehen sie eine (interessante) Schlussfolgerung (für Sie, Ihre Gruppe oder einen fiktiven Arbeitgeber/Kunden) aus der Grafik.

https://www.dashboard-deutschland.de/#/themen

Dauer: 20 Minuten **Abgabe:** Gruppendiskussion

Lagemaße bei Stichproben

Warum Lagemaße und Co?

Wir gewinnen einen ersten Einblick in die Daten.

Deskriptive Statistiken können einfache Fehler finden (z.B. unglaubliches minimales Alter von -10)

Suche nach bestimmten Werten (wie „k.A.“, „999“)

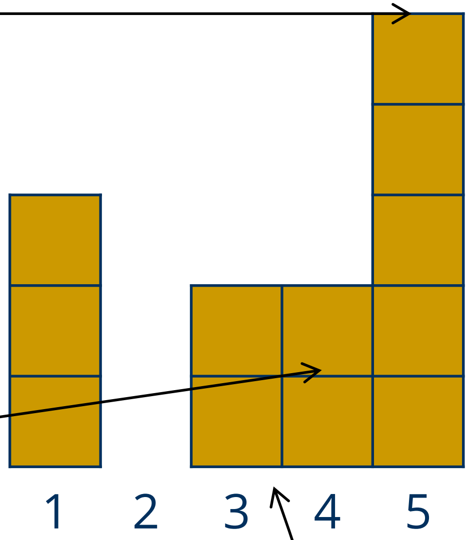
Angabe von Minimum, Maximum …

Kontrolle von Dezimalstellen („.“ statt „,“ bei Import fremder Daten) oder Datumsangaben empfiehlt sich immer in Excel

Lagemaße

Modalwert (Modus): Der häufigste gemessene Wert. (Gibt es mehr als einen Wert der maximal häufig gemessen wurde, wird der kleinste Wert verwendet.

Median: Gibt den Wert an, unterhalb dessen 50 % alle gemessenen Werte liegen. Bei gerader Anzahl von Fällen wird der Median aus dem arithmetischen Mittel der benachbarten Messwerte gebildet:

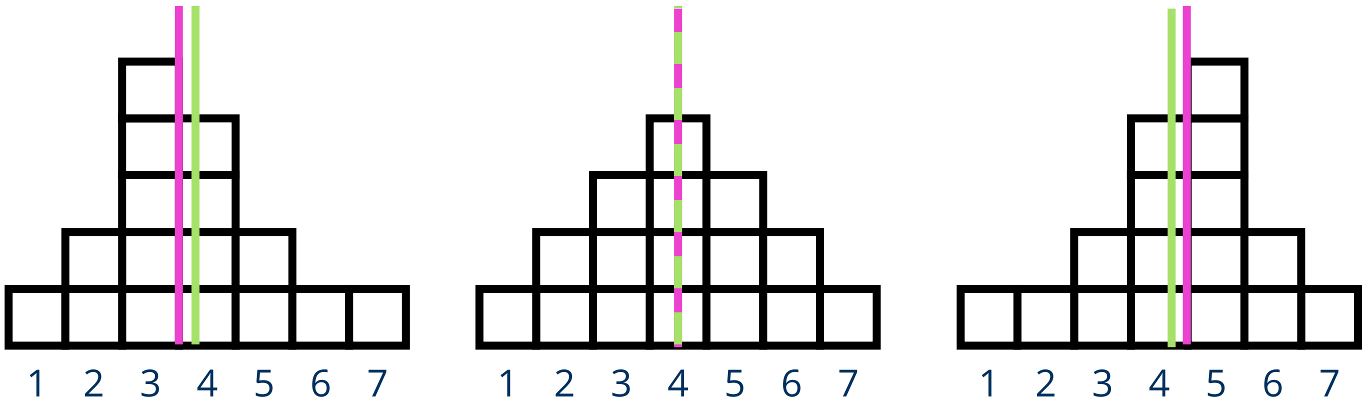


Beispiel:

1 1 1 3 3 4 4 5 5 5 5 5 (Median = 4)

Mittelwert (arithmetisches Mittel). Wird berechnet aus der Summe aller Messwerte geteilt durch die Anzahl der Fälle (hier 3,5).

Verhältnis Mittelwert und Median



Modalwert	3	4	5
Median	3,5	4	4,5
Mittelwert	3,7	4	4,3

Übung: Manipulation des Mittelwerts

Mittelwert Notendurchschnitt von zwei Klassen. Welche Schlussfolgerung könnten Sie ziehen? Ist diese sinnvoll?

Klasse 1	
1	
1	
1	
1	
1	
5	
Mittelwert:	
Median:	

Klasse 2	
1	
1	
1	
2	
2	
3	
Mittelwert:	
Median:	

Aufgabe Lagemaße Selbsttest

Gehen Sie auf folgende Seite und machen Sie einen kurzen interaktiven Test zu den Lagemaßen.
https://www.learningsnacks.de/#/welcome?q=statistik&channel=Learning%20Snacks
Dauer: 5 Minuten **Abgabe:** Keine

Aufgabe Lagemaße:
Erstellen sie bitte Ihre eigene metrische Variable (mit zwanzig Fällen) in Excel. Errechnen Sie für diese Variable jeweils, den Modus, Median, Mittelwert und die Summe.
Dauer: 10 Minuten **Abgabeformat:** Tabelle

Lagemaße: Quiz

Dauer: 15 Minuten **Abgabeformat:** mündlich

	Ja	Nein
1. Der Mittelwert ist ein Wert, der auch in den Daten vorkommen muss.		
2. Der Median ist ein Wert, der nicht in den Daten vorkommen muss.		
3. Um den Mittelwert zu ermitteln, müssen die Daten der Größe nach geordnet werden.		
4. Extremwerte / Ausreißer beeinflussen den Mittelwert stärker als den Median.		
5. Median und Mittelwert können den gleichen Wert annehmen.		
6. Der Mittelwert kann niemals gleich dem kleinsten oder größten Wert entsprechen.		

Mittelwert: Aufgabe Mittelwert interpretieren

Aussage aus dem Radio: „Die meisten Vereine in Sachsen haben weniger als 100 Vereinsmitglieder, damit sind sie kleiner als die durchschnittliche Vereinsgröße in Deutschland.“

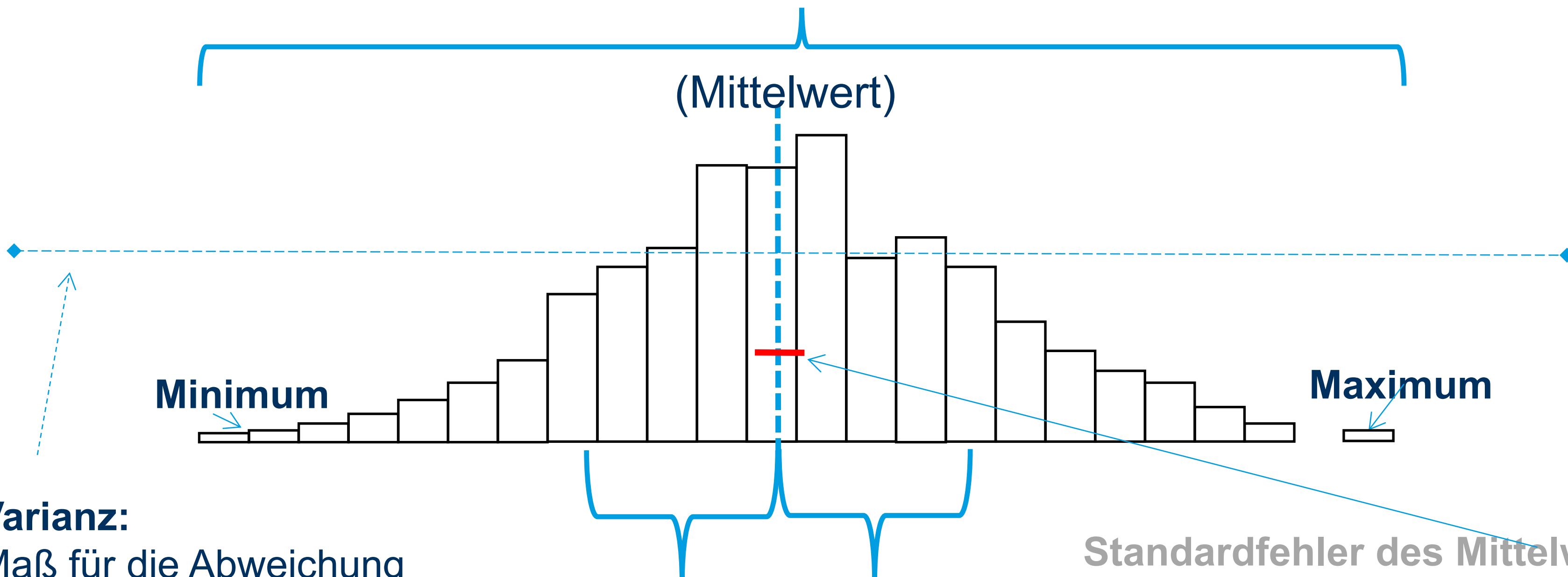
a) Welche zwei Werte werden hier miteinander verglichen?
b) Nennen Sie drei Beispiele für einen Vereine mit über 1 Millionen Mitgliedern? Und nennen Sie 3 Beispiele für Vereine mit unter 100 Mitgliedern.
c) Konstruieren Sie ein Histogramm in dem die Anzahl an Mitgliedern in den Abständen von 0 bis 5.000.000 abgebildet werden kann.
d) Zeichnen Sie Daten in Ihr Histogramm ein (für gesamt Deutschland) in dem die Mehrheit der Vereine unter 100 Mitglieder hat aber die durchschnittliche Anzahl der Vereinsmitglieder über 100 Mitglieder liegt.

Dauer: 10 Minuten **Abgabeformat:** Zeichenprogramm

Streuungsmaße bei Stichproben

Streuungsmaße: messen die Variationen und die Breite der Daten.

Spannweite: Differenz zwischen Minimum und Maximum



Varianz:
Maß für die Abweichung aller Werte vom Mittelwert

Standardfehler des Mittelwerts:
Maß für die Streuung des Mittelwerts (selten genutzt)

Standardabweichung:
Maß für die Streuung der Werte um den Mittelwert (68 % aller Fälle liegen hier im Bereich des Mittelwerts +/- Standardabweichung)

Varianz und Standardabweichung: Formeln

Varianz

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

Leider ist die Einheit der Varianz nicht die gleiche wie der ursprüngliche Wert, daher können wir sie nicht interpretieren.

Standardabweichung

Die Standardabweichung ist interpretierbar.

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

Standardfehler des Mittelwerts:

In welchem Bereich wird der Mittelwert sehr wahrscheinlich (68 %) wirklich liegen.

$$\frac{s}{\sqrt{n}}$$

Varianz und Standardabweichung berechnen mit Zwischenschritten

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

x_i
3
4
2
1
3

Fragen zum Mittelwert und Standardabweichung

Öffnen Sie die interaktive Normalverteilungsgrafik zum Mittelwert und Standardabweichung https://matheguru.com/stochastik/normalverteilung.html

Welche Extremwerte kann die Standartabweichung einnehmen?

Wie sieht die Kurve bei diesen Extremwerten aus?

Wie verändert sich die Normalverteilung, wenn der Mittelwert verändert wird?

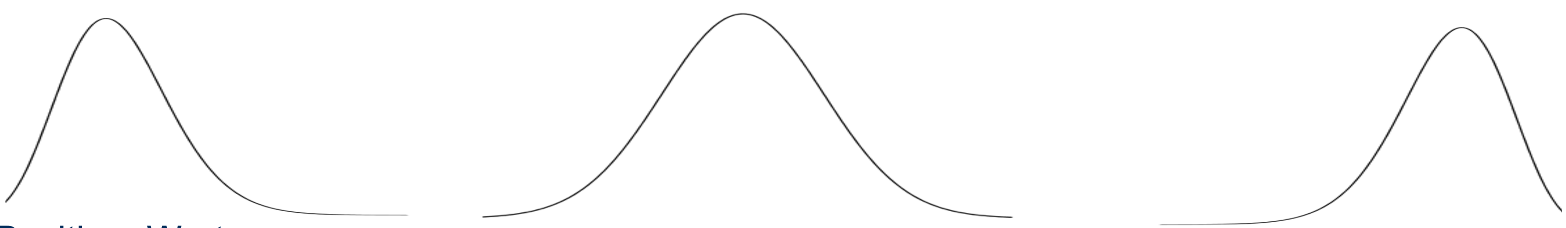
Dauer: 3 Minuten **Abgabeformat:** keine

Aufgabe: Varianz und Standardabweichung
Erfinden Sie eine metrische Variable mit nur 5 Fällen. Berechnen Sie für diese Fälle manuell die Varianz, die Standardabweichung und den Standardfehler des Mittelwerts. Geben Sie für die Varianzberechnung mindestens fünf Zwischenschritte an.
Dauer: 25 Minuten **Abgabeformat:** Tabelle

Verteilungsmaße: Schiefe

Die Standardnormalverteilung werden wir später noch näher kennen lernen.

Schiefe = Maß für die Abweichung der Häufigkeitsverteilung von der Normalverteilung.



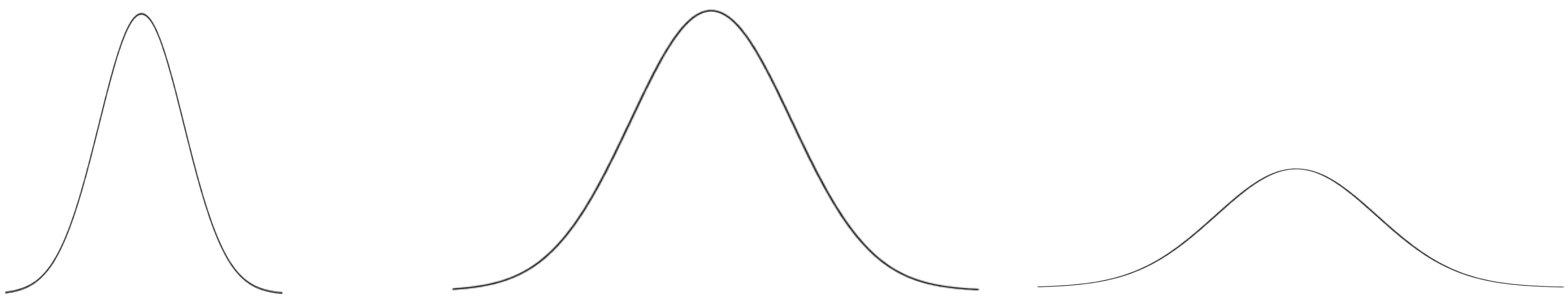
Positiver Wert = linkssteil/rechtsschief

(Standardnormalverteilung zum Vergleich)

Negativer Wert = rechtssteil/linksschief

Verteilungsmaße: Wölbung (Kurtoisis)

Die Standardnormalverteilung werden wir später noch näher kennen lernen.

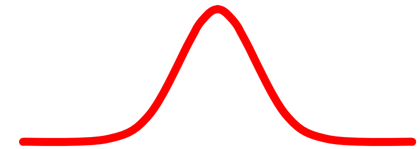


Kurtosis:
hoher Wert, > 3: Verteilung ist steiler als die Normalverteilung

(Standardnormalverteilung zum Vergleich)
Wert = 3

Kurtosis:
geringer Wert, < 3
Verteilung ist flacher als die Normalverteilung

Verteilungsmaße:
Schiefe und Wölbung, Formeln



Die Normalverteilung kann abweichen von der Idealform. Dafür gibt es zwei Messwerte: Schiefe und Wölbung.

Was heißt das, wie lautet die Formel zur Berechnung?

Erstellen Sie eine Beispielaufgabe.

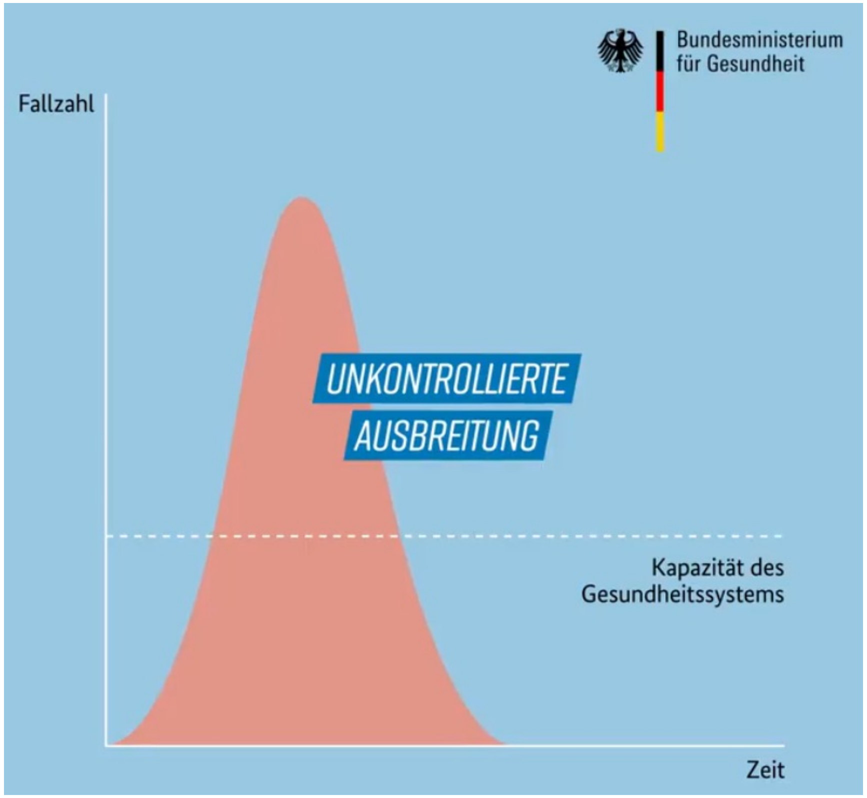
Schiefe

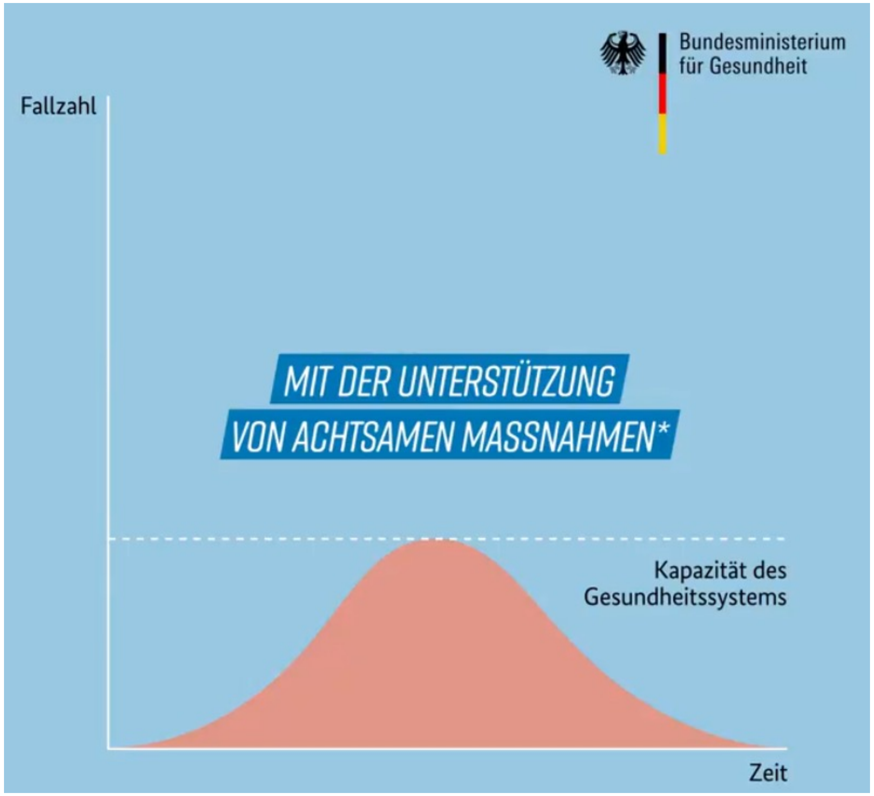
https://www.statistik-nachhilfe.de/ratgeber/statistik/deskriptive-statistik/masszahlen/parameter-der-form/schiefe

Wölbung

https://www.statistik-nachhilfe.de/ratgeber/statistik/deskriptive-statistik/masszahlen/parameter-der-form/woelbung-exzess-kurtosis

Anwendungsbeispiel: Flatten the curve





Verteilung Stauchen: Abrupte und verzögerte Ausbreitung des COVID-19 Erregers.

Quelle: Bundesministerium für Gesundheit
https://twitter.com/BMG_Bund/status/1237121721481781249

Konfidenzintervalle: Warum

Aussage: Im Durchschnitt machen die Deutschen 32,0 Überstunden im Jahr.

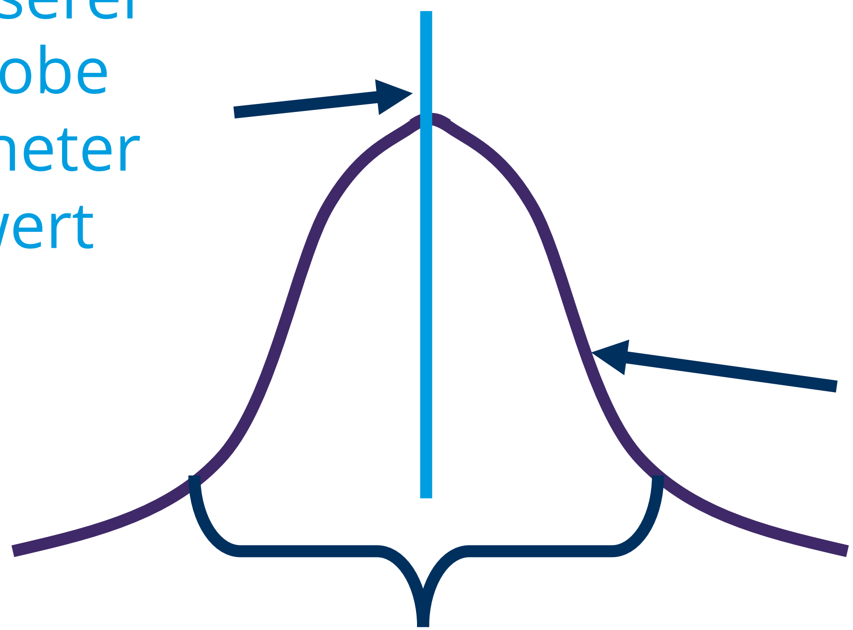
Problem: das trifft nicht für jeden zu, besser wäre die Angabe eines Intervalls. Es ist unwahrscheinlich, dass ein Mittelwert* aus einer Stichprobe mit dem „wahren“ Mittelwert in der Grundgesamtheit übereinstimmt.

Lösung: Deshalb gibt man einen Bereich an, in dem sich der „wahre“ Mittelwert“ befindet. In diesem Bereich ist der Mittelwert dann zu einem bestimmten Prozentsatz zu finden – meist 95 Prozent. Genau gesagt, er ist dort in 95 % alle so gemachten Stichproben zu finden).

Der Bereich heißt **Konfidenzintervall**

*anstatt des Mittelwertes können auch für andere Werte Konfidenzintervalle gebildet werden (Anteilswerte, Regressionskoeffizienten)

Konfidenzintervall: Darstellung



Im Konfidenzintervall befindet sich der „wahre“ Mittelwert mit einer Wahrscheinlichkeit von 95 %.

Konfidenzintervalle: Berechnung

$$Konfidenzintervall\ f\text{ür}\ den\ Stichprobenmittelwert = \pm Z * \frac{s}{\sqrt{n}}$$

Z ist wieder der Z-Wert aus der Normalverteilungstabelle für das gewünschte Wahrscheinlichkeitsintervall. Meist:

90 %: Z =1,64

95 %: Z = 1,96

99 %: Z = 2,58

Später lernen wie beim Rechnen mit der Normalverteilung, woher diese Werte von 1,96 usw. kommen

Konfidenzintervalle: Übungsaufgabe

Übungsaufgabe: Gesucht ist das 95 % Konfidenzintervall für eine Stichprobenmittelwert mit einer Standardabweichung von 30 und dem Mittelwert von 100, bei einem Stichprobenumfang von 900.

$$Fehlergrenze = \pm 1{,}96 * \underbrace{\frac{30}{\sqrt{900}}}_{=1} = 1{,}96$$

Nun ziehen Sie die Fehlergrenze einmal vom Mittelwert (100) ab und Sie erhalten

→ Die untere Grenze des Konfidenzintervalls = 98,04

Um die obere Grenze zu erhalten addieren Sie das Ergebnis einmal auf den Mittelwert hinzu:

→ Obere Grenze des Konfidenzintervalls = 101,96

Konfidenzintervalle Berechnung des 95 %

Daten: 1, 0, -1, 0

Stichprobengröße:

Mittelwert:

Varianz:

Standardabweichung:

Konfidenz:

Untere Grenze:

Obere Grenze:

Konfidenzintervalle: Interpretation

A. Achtung man interpretiert das Konfidenzintervall nicht so:

Mit 95 % Sicherheit liegt der wahre Mittelwert in dem Konfidenzintervall.

Sondern:

- In 95 % aller vergleichbar ermittelten Stichproben liegt der Mittelwert zu 95 % in diesem Intervall.
- Oder: *Wir wissen/glauben/vermuten* … mit 95 % Sicherheit, das der wahre Mittelwert in den Grenzen des Konfidenzintervalls liegt.

B. Beachten Sie, den Unterschied zwischen Fehlergrenze und obere/untere Schranke/Grenze des Konfidenzintervalls. (siehe vorherige Folie)

C. Die Einheiten des Konfidenzintervalls sind gleich die Einheiten des Stichprobenmittelwerts. Z.B. beim Mittelwert der Körpergröße z.B. in Zentimetern, kann das Konfidenzintervall z.B. 10 Zentimeter betragen.

Konfidenzintervall: Übung

Berechnen Sie das 95 % Konfidenzintervall für Rennstrecken von 100 Sportlern in einer Woche:

Mittelwert: 200 km

Standardabweichung: 40 km

Erläutern Sie, was das Konfidenzintervall bedeutet.

Konfidenzintervalle: Übung Abhängigkeit von Verteilung

Nutzen Sie interaktive Grafik: https://seeing-theory.brown.edu/frequentist-inference/index.html#section2

Scrollen Sie zu „Confidence Intervall“

Suchen Sie als Verteilungsformen die Normalverteilung aus.

Drücken Sie nun auf „Start Sampling“.

Beobachten Sie nun etwas eine halbe Minute wie das Programm zufällige Stichproben (n = 5) aus der Verteilung zieht und jeweils das Konfidenzintervall um den Mittelwert herum berechnet.

Wählen Sie im Anschluss eine neu Verteilung (z.B. uniform) und starten das Experiment erneut. Erhöhen Sie auch mal die Stichprobengröße.

Beschreiben Sie, wie sich die Konfidenzintervalle zwischen beiden Verteilungsformen unterscheiden.

Konfidenzintervalle: Eigenschaften der Fehlergrenze

$$Fehlergrenze\ f\text{ür}\ Mittelwert = \pm Z * \frac{s}{\sqrt{n}}$$

Wenn das n – also die Stichprobe – größer wird, wird der Fehler immer kleiner.

Daraus lässt sich ableiten, dass größere Stichproben zu verlässiger Aussagen über den gefundenen Mittelwert zulassen.

1. Was passiert mit der Fehlergrenze, wenn das Konfidenzniveau (Z) größer wird?
2. Was passiert mit der Fehlergrenze, wenn die Standardabweichung größer wird?
3. Kann es eine Fehlergrenze geben in der zu 100 % der Wert liegt?

Konfidenzintervall: Quiz

	a	b
Sie haben für eine Stichprobenmittelwert zwei Konfidenzintervalle errechnet ein a.) 95 % und b.) ein 99 %. Welches ist breiter?		
Sind breitere Konfidenzintervalle a.) eher nicht erstrebenswerter b.) erstrebenswerter als schmalere Konfidenzintervalle		
Mit dem Konfidenzintervall von 95 % wissen wir a.) zu 100 % das der Mittelwert der Stichprobe in diesem Intervall liegt b.) zu 95 % das der Mittelwert der Stichprobe in diesem Intervall liegt.		
Bei einer Stichprobe mit dem Mittelwert von 50 cm und einer Fehlergrenze von 2 cm hat das Konfidenzintervall a.) eine Breite von 2 cm oder b.) eine Breite von 4 cm.		
Um die Fehlergrenze bei einem 95 % Konfidenzintervall zu verringern kann man a.) die Stichprobe erhöhen oder b.) die Standardabweichung erhöhen.		

Konfidenzintervall: Konfidenz in Excel

=KONFIDENZ.NORM(Alpha;Standardabweichung;Umfang)

Der Standardwert für Alpha ist 0,05. Am Ende der Rechnung erhalten wir so das 95-Prozent-Konfidenzintervall.

Die Standardabweichung errechnet sich aus den Rohdaten mit der Funktion =STABW.S(Bereich)

Der Umfang ist einfach die Stichprobengröße.

	A	B	C	D
1	Rohdaten	Werte	Ergebnis	Formel
2	1	Mittelwert:	4,5	=MITTELWERT(A2:A9)
3	2	Standardabweichung:	2,449	=STABW.S(A2:A9)
4	3	Stichprobengröße:	8	=ANZAHL(A2:A9)
5	4	Konfidenz für 95 %:	1,697	=KONFIDENZ.NORM(0,05;C3;8)
6	5	Konfidenz für 99 %:	2,234	=2,58*C3/WURZEL(C4)
7	6			
8	7			
9	8			

Konfidenzintervall: Konfidenzintervall in Excel

Da das Konfidenzintervall ein Intervall ist muss die „Konfidenz“ noch einmal auf den Mittelwert addiert und einmal subtrahiert werden.

Beispiel:

Konfidenz = 1,687 (siehe vorherige Folie)

Mittelwert = 4,5

Obere Grenze: = Mittelwert + Konfidenz

= 4,5 + 1,697 = 6,197

Untere Grenze: = Mittelwert – Konfidenz

= 4,5 – 1,697 = 2,803

Konfidenzintervalle: Konfidenz mit Datenanalyse

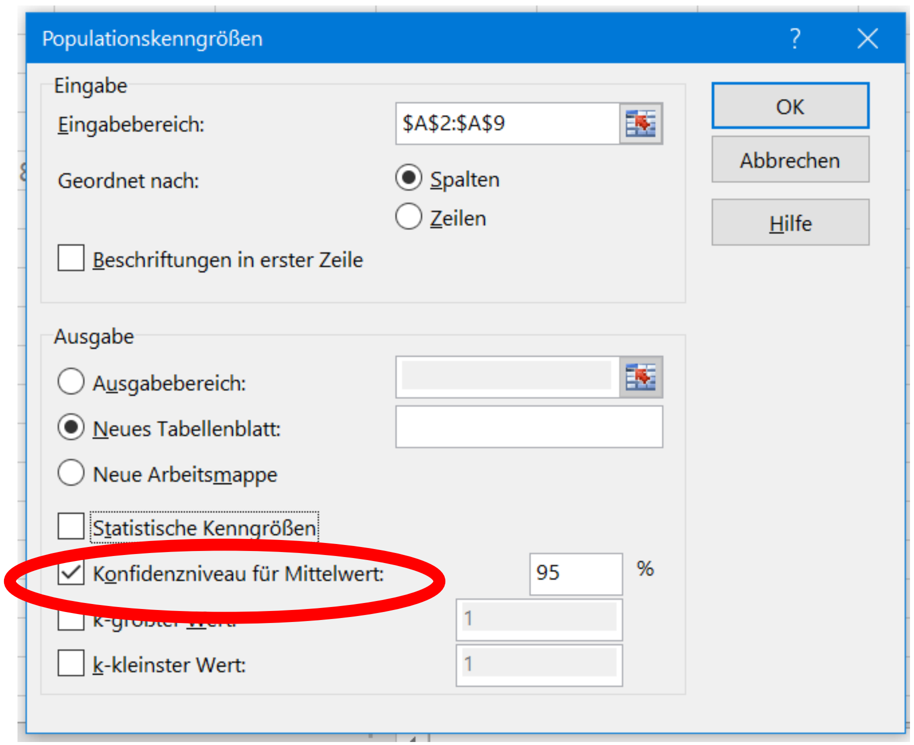
Daten → Datenanalyse → Populationskenngrößen

Den haken bei Konfidenzniveau für Mittelwert setzen und die 95 % lassen.

Diese Berechnete Konfidenz unterscheidet sich von der Funktion =KONFIDENZ.NORM()

Sie entspricht aber der Funktion

=KONFIDENZ.T()



Beachten Sie, dass Sie die Konfidenz wieder addieren und subtrahieren müssen um das Konfidenzintervall zu erhalten

Übung: Unterschied Konfidenzintervall zu Standardabweichung

Konfidenzintervall

Standardabweichung

Macht eine Aussage über den Mittelwert.

Macht eine Aussage über alle Daten.

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

$$= \pm Z * \frac{s}{\sqrt{n}}$$

Wird auf den Mittelwert auf addiert und ab subtrahiert.

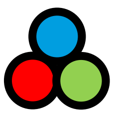
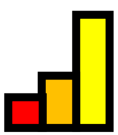
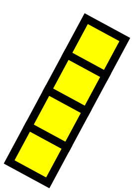
Muss geteilt werden, und dann um den Mittelwert auf addiert und ab subtrahiert.

Bezieht sich meist auf 68 % Prozent von etwas.

Bezieht sich meist auf 95 % von etwas.

Skalenniveaus

Aufgabe
Beispiele Skala:
Finden Sie jeweils 2 Beispiele.

Grafik	Skala	Merkmale	Lagemaße	Streuungsmaße, u.a.	Beispiel
	Nominal	Klassifizierung qualitativer Merkmale	**Modus**	*(kein Streuungsmaß)*	
	Ordinal	Rangwerte mit Ordinalzahlen	Modus, **Median**	**Minimum, Maximum,** Perzentile	
	Metrisch	Gleiche Abstände	Modus, Median, **Arithmetisches Mittel**	Min, Max; Perzentile, Spannweite, **Standard-abweichung**, Varianz, Schiefe	

Dauer: 3 Minuten.
Abgabe: Textdatei

Erwartungswert und Varianz bei bekannten Wahrscheinlichkeiten

Unendliches Zufallsexperiment

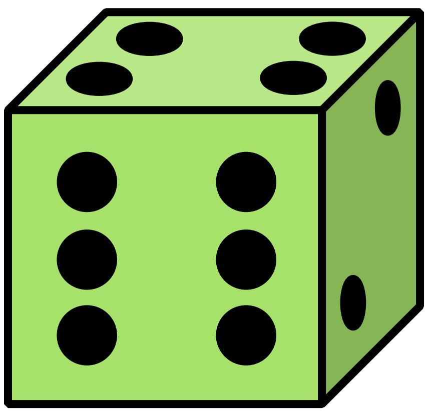
Wir wiederholen das Experiment unendlich oft (→ Gesetz der großen Zahlen).

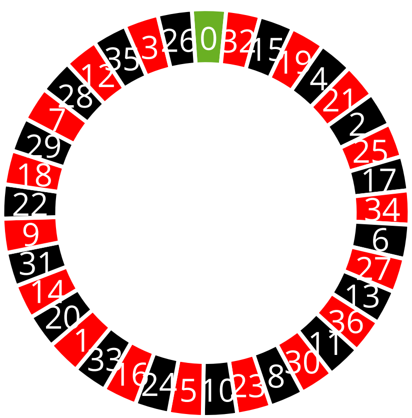
Welchen durchschnittlichen Wert erhalten wir, wenn wir unendlich oft, eine Münze werfen, würfeln oder Roulette spielen?

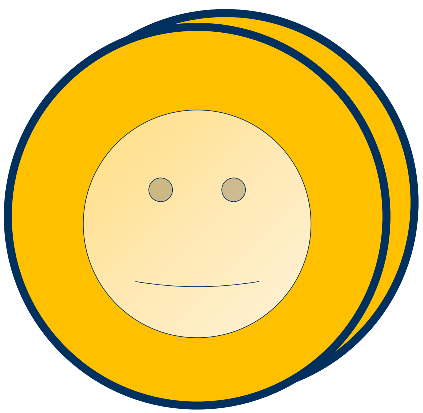
Würfel: Mittelwert =

Roulette: Mittelwert =

Definition für Münze notwendig: 0 = Kopf, 1 = Zahl. Mittelwert =







Erwartungswert

Allgemeine Formel für EX (Erwartungswert), X = Zahl

$\mathbf{EX} = \sum_{\mathbf{k}} \mathbf{x_k} * \mathbf{p_k}$ (p = Wahrscheinlichkeit für jedes einzelne X)

Für den Würfel sieht die Formel für den Erwartungswert so aus:

$$EX = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6}$$

$$EX = 3{,}5$$

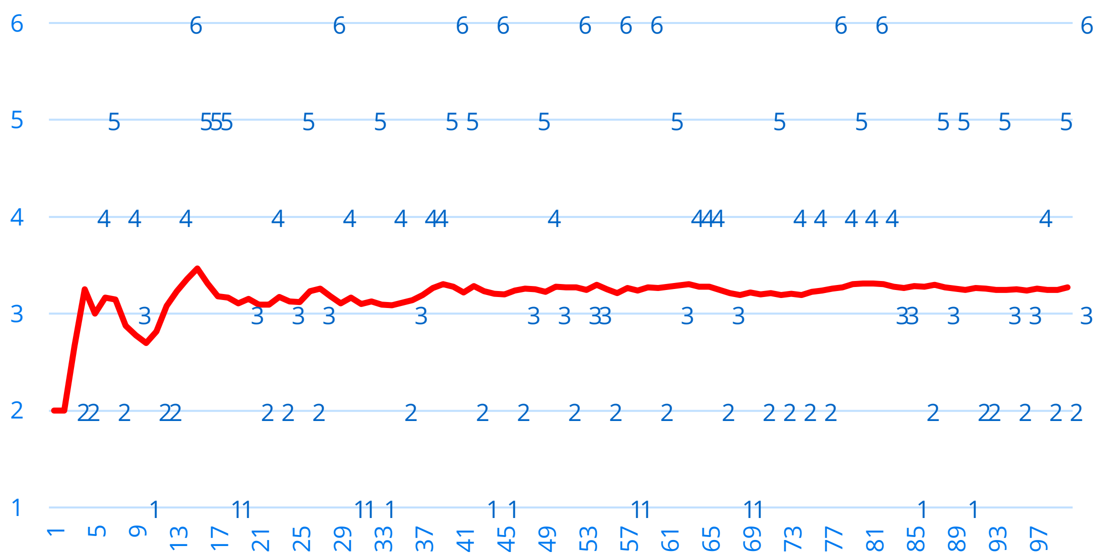
Welche Zahl kann man beim Roulette erwarten?

$$EX = 0 * \frac{1}{37} + 1 * \frac{1}{37} + 2 * \frac{1}{37} + 3 * \frac{1}{37} + 4 * \frac{1}{37} \; \dots \; 36 * \frac{1}{37}$$

Erwartungswert und Zentrale Grenzwertsatz

Der Erwartungswert gibt an, welcher Wert zu erwarten ist, wenn das „Experiment“ unendlich oft wiederholt wird.

Die Grafik zeigt die Ergebnisse eines Würfelexperiments. Auf der X-Achse sind die 100 Ergebnisse der Reihe nach dargestellt. Auf der Y-Achse ist die Höhe der Augenzahl dargestellt. Die rote Linie zeigt den Durchschnitt, jeweils vom ersten Wurf an. Am Ende nähert sich dieser Durchschnitt dem Erwartungswert.

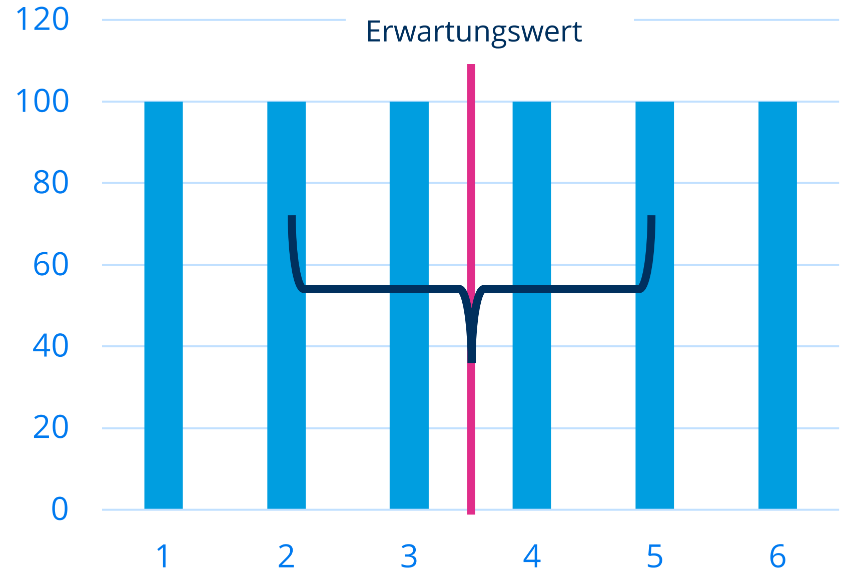


Führen Sie dieses Experiment selbst online durch: https://seeing-theory.brown.edu/basic-probability/index.html Scrollen Sie runter zu „**Expectation**“ und klicken Sie links „Roll 100 times. Wie viele Male müssen Sie auf den Button drücken, bis der Mittelwert sehr nah am Erwartungswert ist?

Varianz bei bekannten Wahrscheinlichkeiten

Erinnerung: Die Varianz zeigt an, wie sehr der Erwartungswert streut.

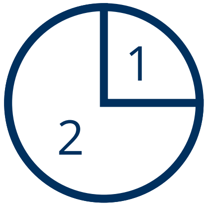
$$Var(X) = \sum_k (x_k - EX)^2 p_k$$



Errechnung der Varianz beim Würfel:

Aufgabe Erwartungswert

A. Wir haben ein Glücksrad. ¼ der Fläche entspricht dem Wert 1 und ¾ der Fäche dem Wert 2. Wie hoch ist der Erwartungswert für das Glücksrad?



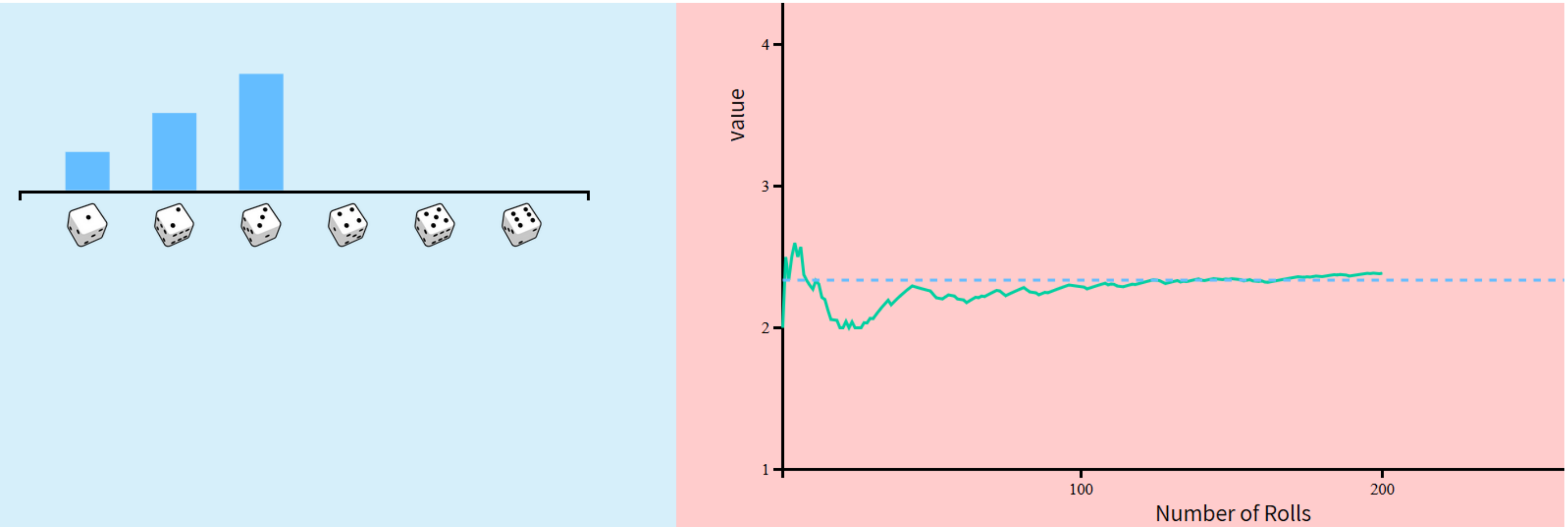
B. Ein Eisladen verkauft täglich
9 Eiswaffeln der Größe 1,
6 mal der Größe 2 und
3 Mal der Größe 3.
Wie lautet der Erwartungswert für die Größe vom Eis?
Wie lautet die Varianz?

Und ganz einfach gedacht:

C. Man würfelt 120 Mal. Wie groß ist der Erwartungswert für die 6?

D. 700 Mal wird eine Münze geworfen. Wie oft wird man Kopf werfen?

Veranschaulichung des Erwartungswert bei verschieden Wahrscheinlichkeiten



Gehen Sie auf folgende Seite: https://seeing-theory.brown.edu/basic-probability/index.html

Scrollen Sie bis hin zu „Expectations“

Dann verändern Sie bei den Würfeln die Wahrscheinlichkeiten. 4, 5 und 6 auf 0. Bei 1 auf 0,17, bei 2 auf 0,33 und bei 3 auf 0,5. (Wenn Sie ein Balken verändern, verändern sich die anderen automatisch, so dass Sie mehrfach die Veränderungen wiederholen müssen.)

Abgrenzung des Erwartungswert und Varianz bei bekannten Wahrscheinlichkeiten zu Mittelwert und Varianz bei Stichproben

Erwartungswert von einer Stichprobe entspricht dem Mittelwert.

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

Um Varianz zu schätzen verwendet man Streuung

$$s_{xx} \frac{1}{n.1}\sum_{k=1}^{n}(x_k - \bar{x})$$

$$s_x = \sqrt{s_{xx}}$$

Der Unterschied zwischen Erwartungswert und Varianz bei bekannten Wahrscheinlichkeiten: Bei bekannten wird der Wert multipliziert mit der Wahrscheinlichkeit für einen Wert. Bei Stichproben geht man davon aus, dass alle gefundene Werte die gleiche Wahrscheinlichkeit haben. Statt mit der Wahrscheinlichkeit zu multiplizieren wird durch n geteilt.

Erwartungswert und Varianz bei anderen Verteilungen

Wir haben hier den Erwartungswert und die Varianz bei bekannten Wahrscheinlichkeiten behandelt.

Für andere Verteilungsformen gibt es andere Formeln:

https://de.wikipedia.org/wiki/Liste_univariater_Wahrscheinlichkeitsverteilungen