

# Lagemaße



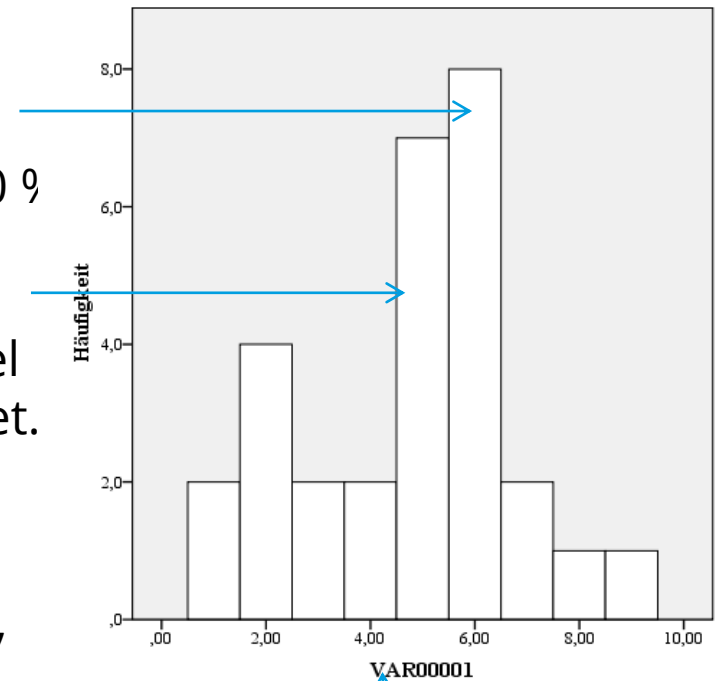
**Lagemaße:** beschreiben die Lage der Verteilung der Daten.

**Modalwert** (Modus): Der häufigste gemessene Wert in einer Stichprobe. Bei mehreren Werten mit gleicher maximaler Häufigkeit = kleinste Wert.

**Median:** Gib den Wert an, unterhalb dessen 50 % aller gemessenen Werte liegen (zweite Quartil der Stichprobe). Bei gerader Anzahl von Fällen wird der Median aus dem arithmetischen Mittel der benachbarten mittleren Messwerte gebildet.

**Mittelwert** (arithmetisches Mittel): Gibt das arithmetische Mittel der Messwerte an. Dieses berechnet sich aus der Summe der Messwerte, geteilt durch ihre Anzahl.  $\bar{x} = \frac{1}{n} \sum x_i$  Ausreißer „verfälschen“ den Mittelwert.

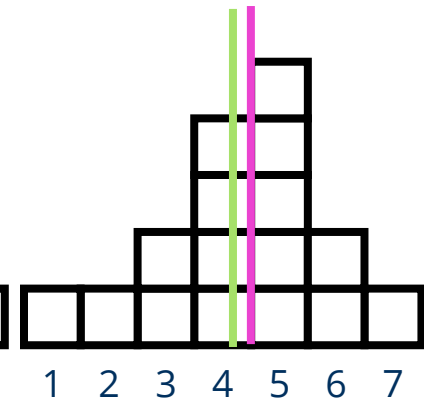
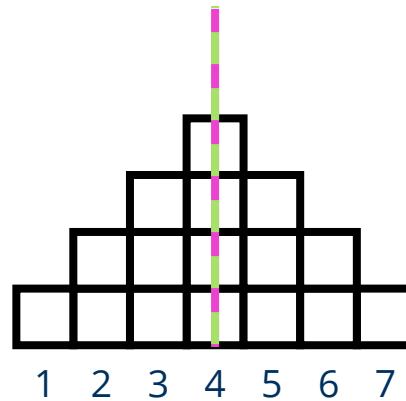
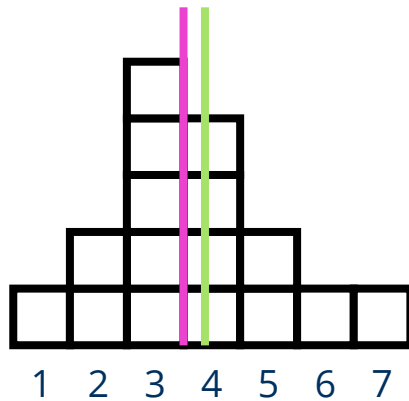
**Summe** aller Werte (selten nützlich)



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



# Verhältnis Mittelwert und Median



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



## **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:** 15 Minuten **Abgabeformat:** Exceldatei



# Quiz Lagemaße

Dauer: 15 Minuten Abgabeformat: Textdatei

	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. Der Mittelwert kann niemals gleich dem kleinsten oder größten Wert entsprechen.		
6. Median und Mittelwert können den gleichen Wert annehmen.		



# Warum: Lagemaße und Co?

Wir gewinnen einen ersten Einblick in die Daten.

Vergleich von Gruppen und Variablen.

Deskriptive Statistiken können auch verwendet werden um Fehler zu suchen

Häufigkeitstabellen

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

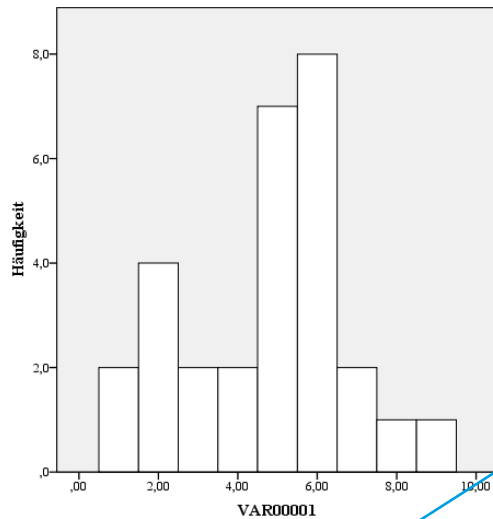
Angabe von Minimum, Maximum ...

Sortieren der Variable

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



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



Wichtigsten Perzentile:

25-Prozent-Perzentil (untere Quartil/1. Quartil): 25 % aller Fälle fallen sind haben Werte bis „3“.

50-Prozent-Perzentil (2. Quartil): 50 % aller Fälle nehmen Werte bis „5“ an.

75-Prozent-Perzentil (obere Quartil/3. Quartil): 75 % aller Fälle nehmen Werte bis „6“ an.

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



Perzentile sind eigentlich so was wie Streuungsmaße, wir brauchen Sie aber schon für die Diagramme, weshalb sie hier dabei sind.



# Übung Quantil:

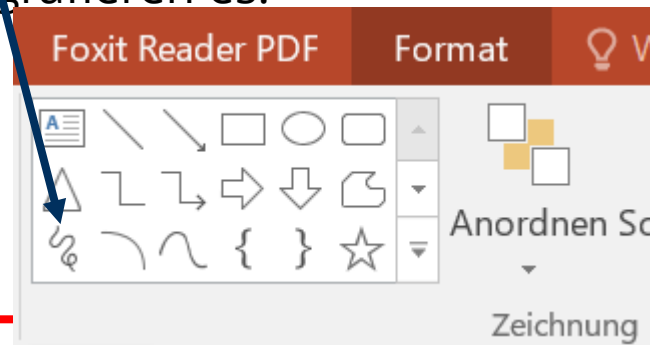
## Aufgabe Quantile

Berechnen Sie für diese Werte das 25, 50 und 75 Prozent Quantil.

Skizzieren Sie anschließend ein Boxplot.

Erfinden Sie einen Ausreiser und wiederholen Sie die beiden vorangegangenen Schritte.

**Dauer:** 15 Minuten **Abgabe:** Textdatei. Skizze entweder in Powerpoint mit der Zeichnung „Freihand“ oder anderen Formen, oder sie nehmen ein Blatt und fotografieren es.



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

# 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

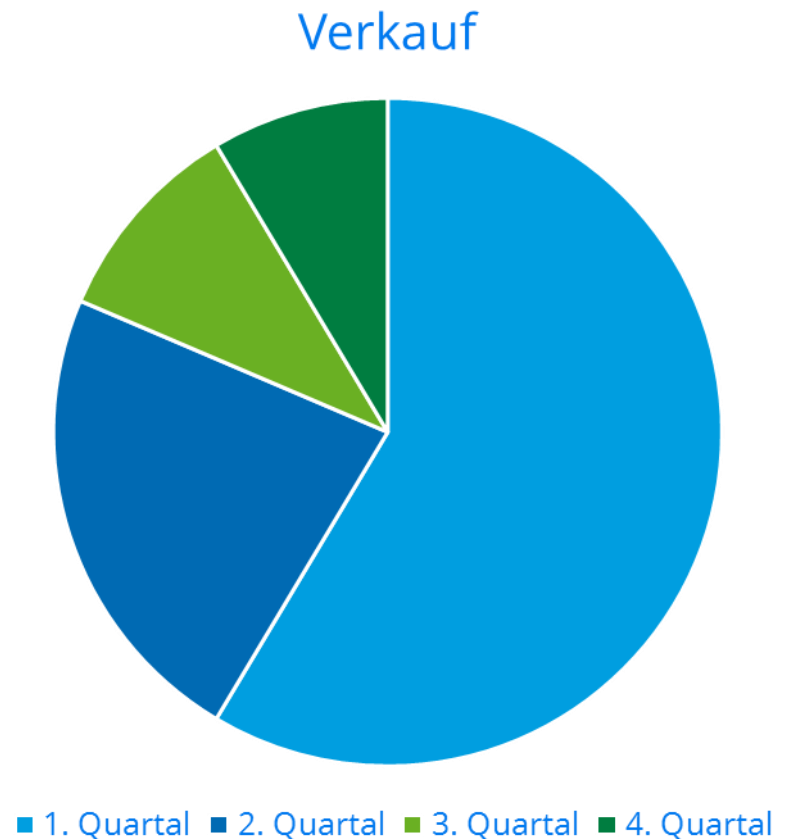
Hier habe ich die Tabelle erweitert.  
Nicht jede Grafik eignet sich für jedes Skalenniveau.

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



# Kreisdiagramm

Für Darstellung von Anteilen  
Für nominale, kategoriale und  
metrische Daten



# Häufigkeiten bei Mehrfachantworten

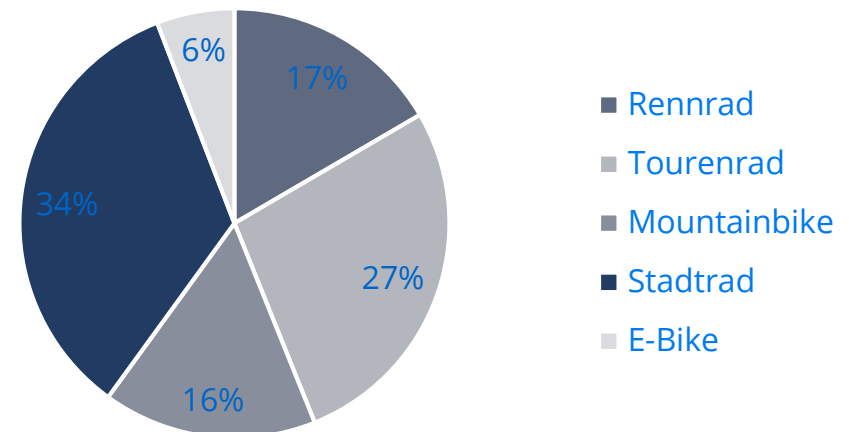
## Aufgabe Häufigkeiten Fahrräder:

Sie haben 100 Kunden eines Fahrradgeschäfts befragt, welche der Produkte sie in den letzten 5 Jahren genutzt haben. Dabei erhalten Sie folgende Nennungen: Rennrad 34, Tourenrad 56, Mountainbike 33, „Stadtrad“ 70 und E-Bike 12. Warum ist die klassische Häufigkeitstabelle für diese Daten nicht sinnvoll? Und warum ist die abgebildete Grafik falsch? Wählen Sie eine geeignete Darstellungsart.

**Dauer:** 10 Minuten

**Abgabe:** Text oder Exceldatei

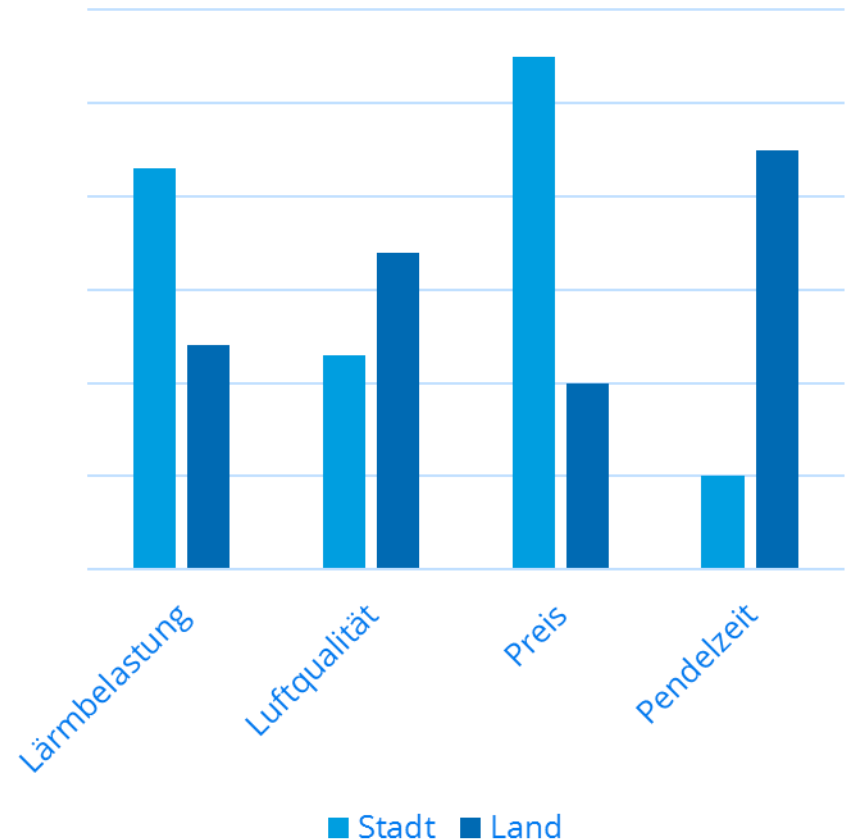
Verkauf



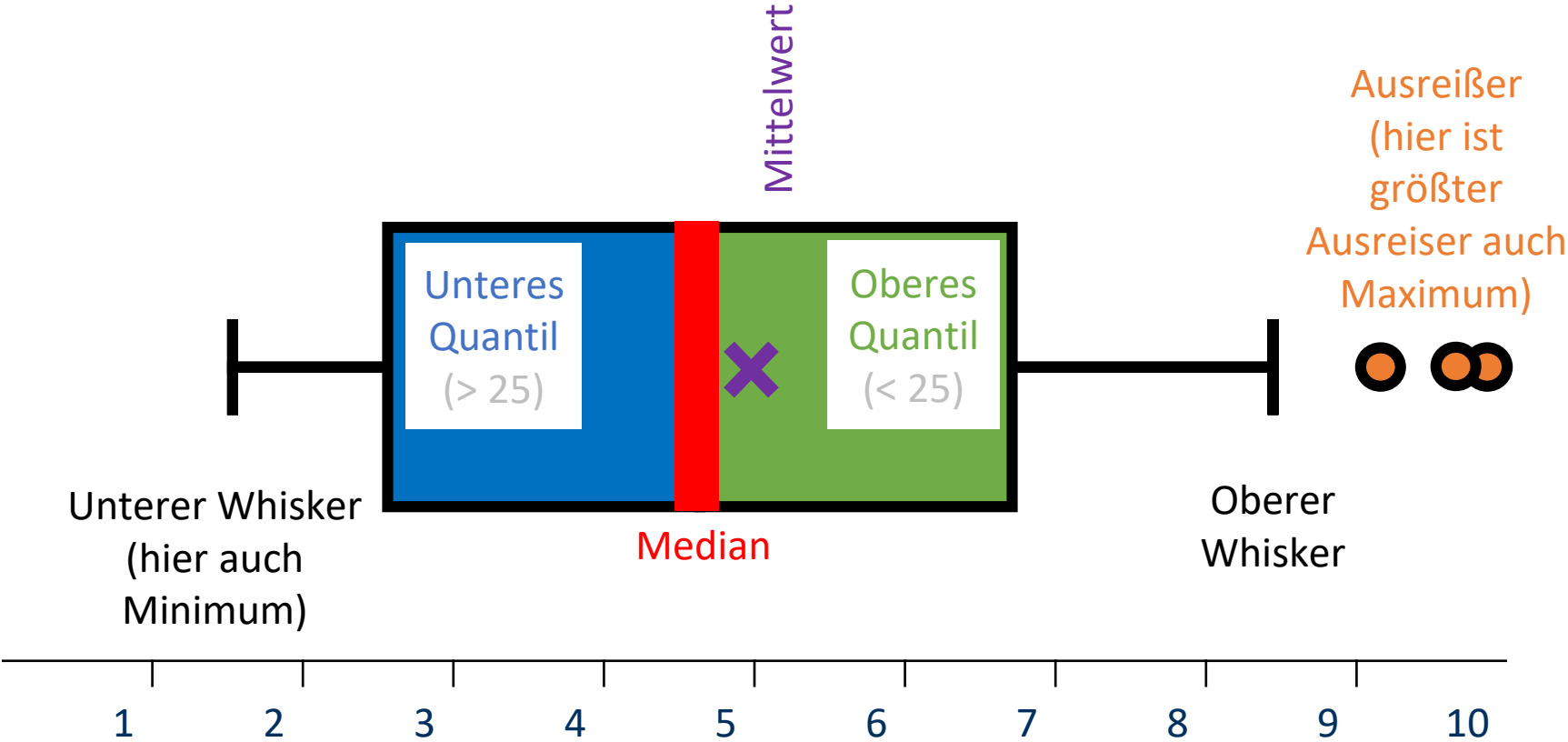
# Balkendiagramm (auch Säulendiagramm)

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.

Zufriedenheit mit Lebenssituation nach Wohnort



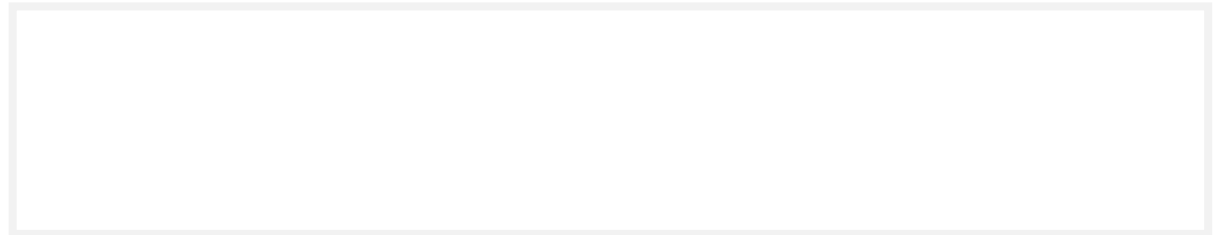
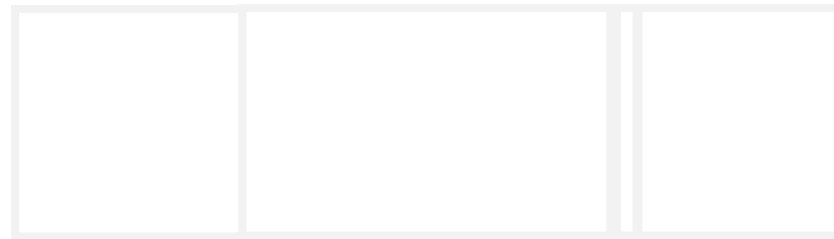
# Boxplots (Kastengrafik)



# Boxplot zeichnen

Unsere Datenreihe: 5 1 4 4 4 4 5 2 2 3 2 5

Median =



# Anwendungsbeispiel: Boxplots

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

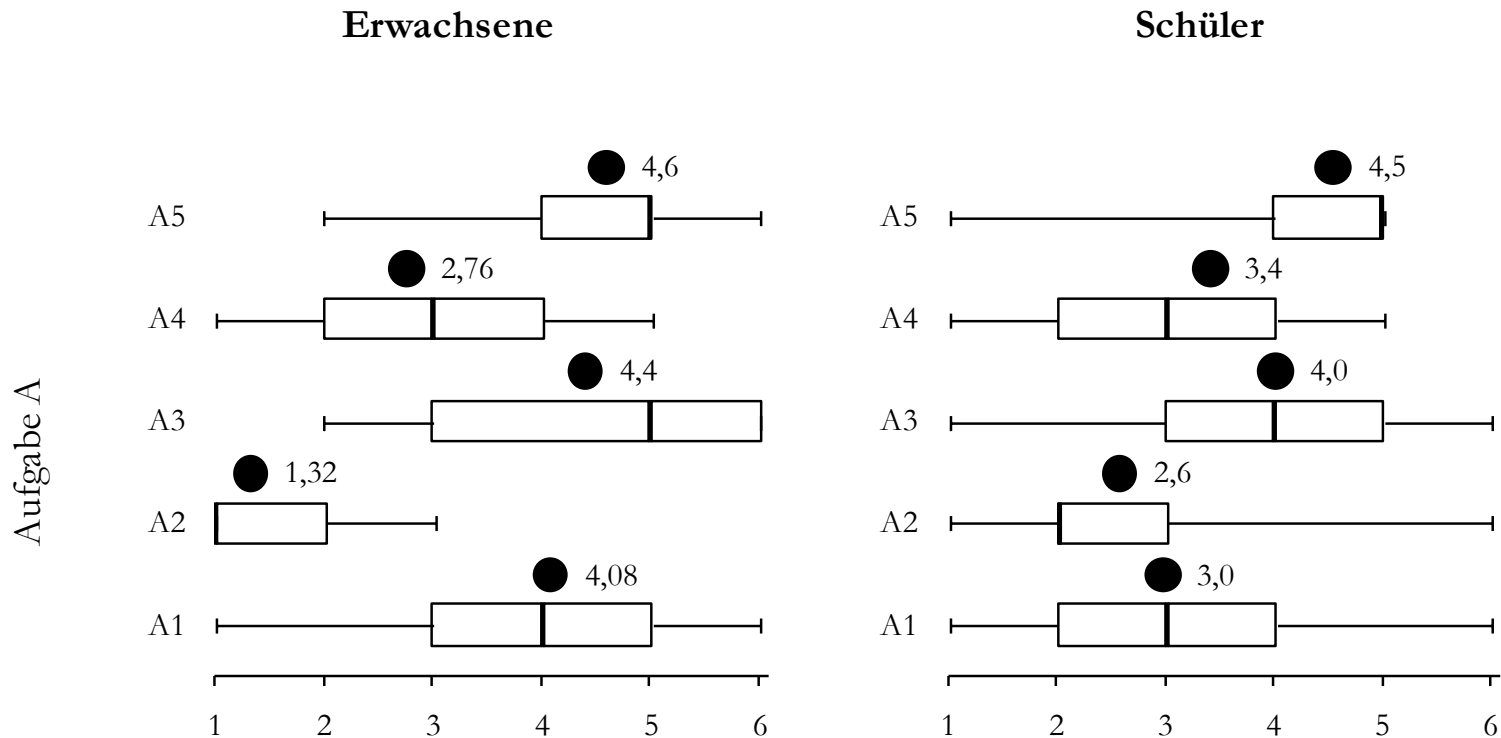


Abbildung: Boxplots für die einzelnen Verhaltensstrategien nach Erwachsenen und Schüler für die Aufgaben A-C (inklusive des Mittelwertes: der schwarzer Punkt oberhalb der jeweiligen Box).



## **Aufgabe Diagramm:**

Ihre Aufgabe ist es, ein Kreisdiagramm, ein Säulendiagramm und ein Boxplot in Excel zu erstellen. Die Daten müssen Sie sich dazu selbst ausdenken. Nehmen Sie pro Grafik mindestens 10 Fälle.

### **Hinweis:**

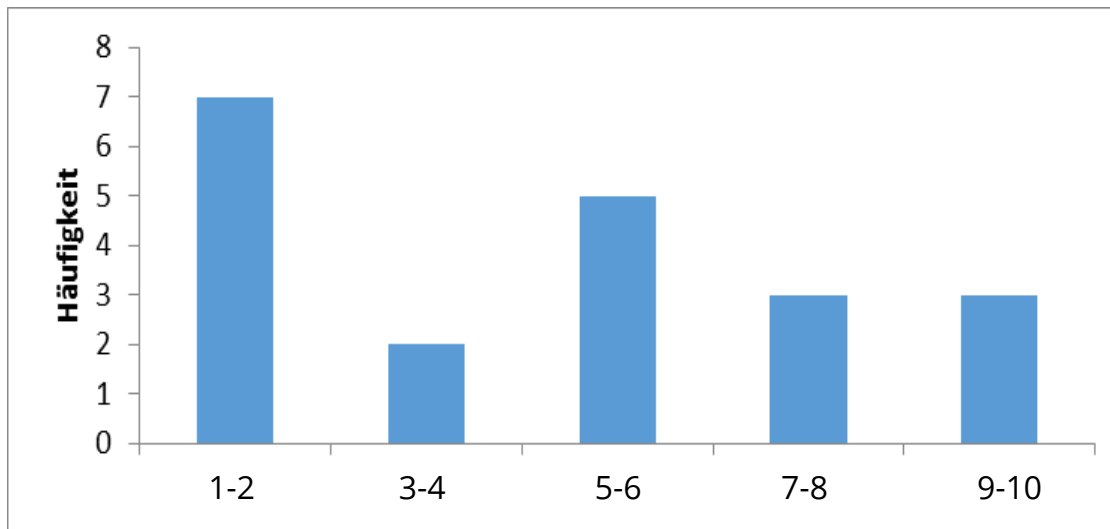
Im Internet finden Sie gute Anleitungen wie man in Excel Grafiken erstellt. Achten Sie darauf, dass das Skalenniveau für die jeweilige Grafik passend ist. Beachten Sie dazu die Folie zu „Skalenniveaus und Grafiken“.

**Dauer:** 30 Minuten **Abgabeformat:** Exceldatei

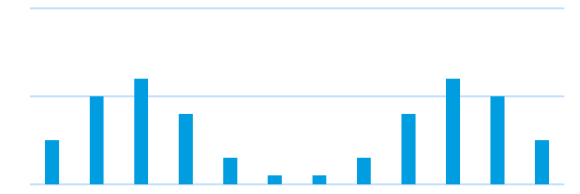
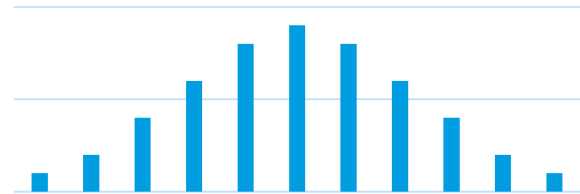
# Histogramm

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

Auf der Y-Achse werden nur die Häufigkeiten abgetragen.



# 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:

Zweigipflig

Glockenkurve

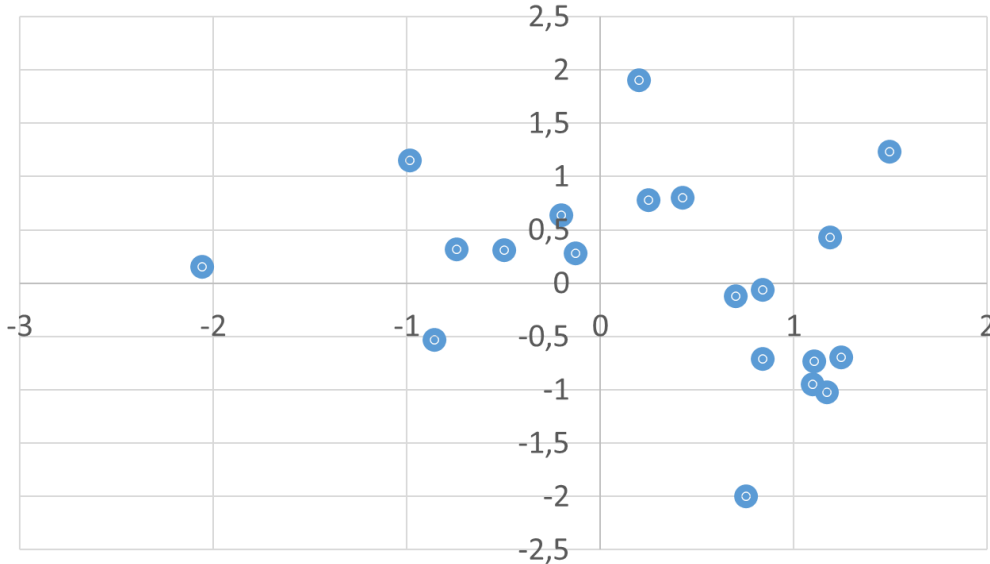
U-Förmige Verteilung

Gleichverteilung

**Dauer:** 1 Minute **Abgabeformat:** keins



# Streudiagramm (Punktdiagramm)



Die Werte zweier Variablen werden miteinander verglichen

Beide Variablen sind ordinal/metrisch

In Excel gibt es die Möglichkeit, eine Trendlinie einzufügen (→Diagramm-element hinzufügen)



# Liniendiagramme

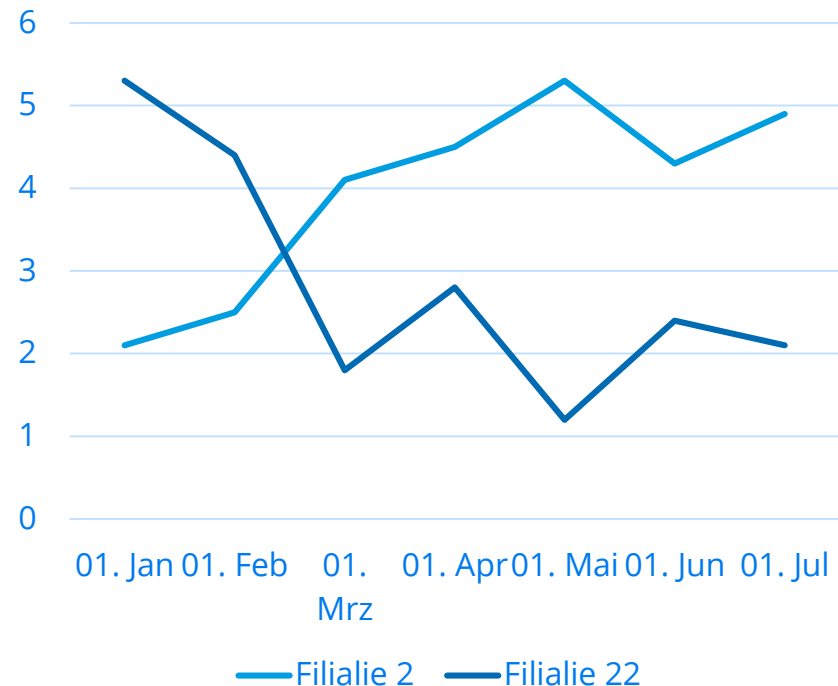
Liniendiagramme sind gedacht für zeitliche Entwicklungen

## Aufgabe Linien:

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

**Dauer:** 2 Minuten **Abgabeformat:** Textdatei/Exceldatei

### Umsatzentwicklung nach Filialien



# Übung: Skizzieren von Histogrammen

a)

$n =$	2
	1
	3
	2
	3
	2
	3
	4
	5
	6
	3
	4
	3

b)

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

c) Die Befragten sollten auf einer Skala von 1-5 antworten. Dabei wurde die Eins 5 mal die Zwei 4 mal und die Drei nur einmal angekreuzt. Die Optionen 4 und 5 wurden von niemanden gewählt.

- Aufgabe Histogramme:**
- Erstellen Sie für alle drei Daten ein Histogramm.
  - Interpretieren Sie die Verteilung. von c auch inhaltlich. Entwickeln Sie für eine von Ihnen ausgewählte Datenreihe auch ein Boxplot.
  - Hinweis: Beachten Sie, dass sie die notwendigen Kategorien für das Histogramm selbst definieren müssen.
  - a) b) und c) sind nur unterschiedliche Darstellungsformen
- Dauer:** 45 Minuten **Abgabeformat:** Exceldatei

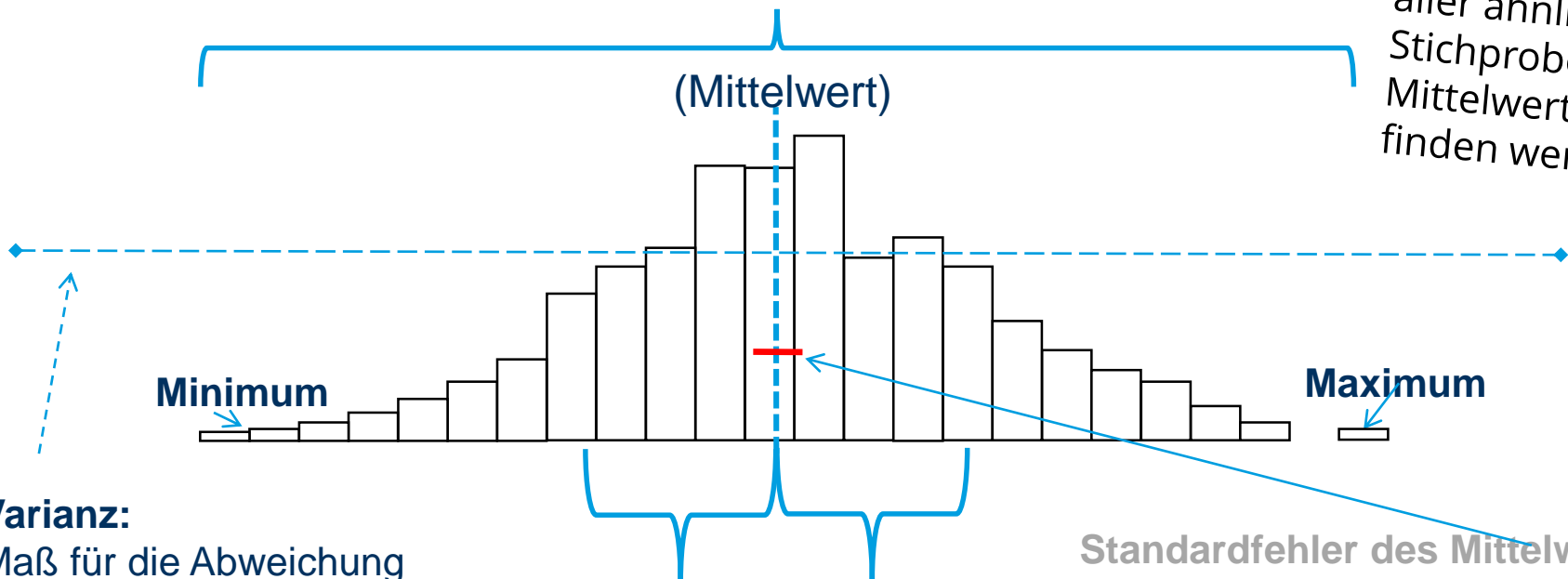
# Streuungsmaße



# Streuungsmaße: messen die Variationen und die Breite der Daten.

**Spannweite:** Differenz zwischen Minimum und Maximum

Falls wir, in 68 % aller ähnlicher Stichproben, den Mittelwert hier finden werden.



**Varianz:**  
Maß für die Abweichung aller Werte vom Mittelwert

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

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



# Formeln Varianz und Standardabweichung

## 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 Standardabweichung einnehmen?

Wie sieht die Kurve bei diesen Extremwerten aus?

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

**Dauer:** 3 Minuten **Abgabeformat:** Textdatei



**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:** 10 Minuten **Abgabeformat:** Textdatei

# Skalenniveaus

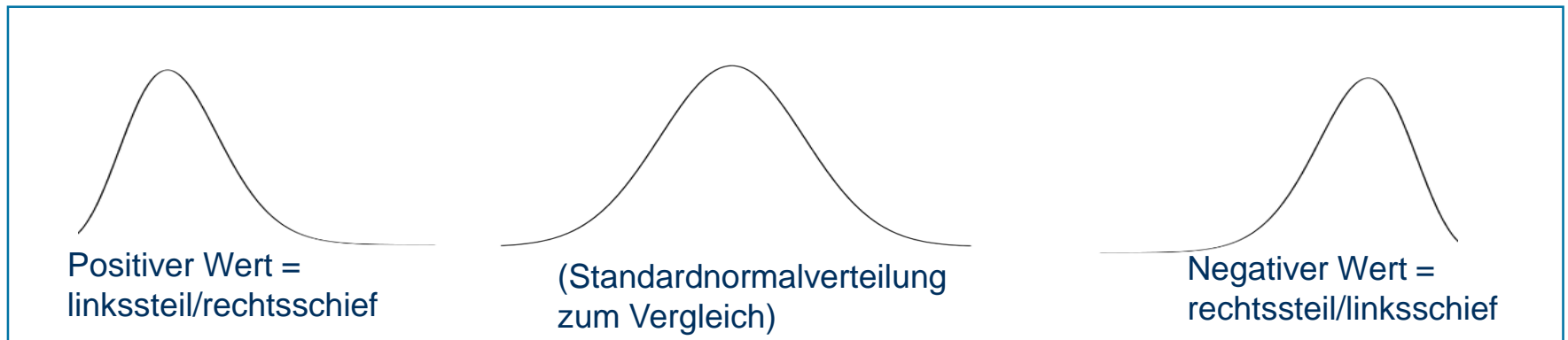
**Aufgabe**  
Beispiele Skala:  
Finden Sie jeweils  
2 Beispiele.

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

**Dauer: 3**  
**Minuten.**  
**Abgabe:**  
**Textdatei**

# Verteilungsmaße: Schiefe

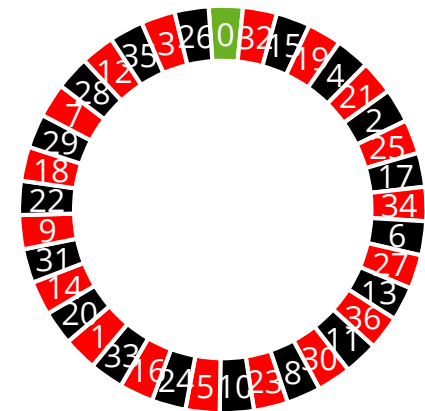
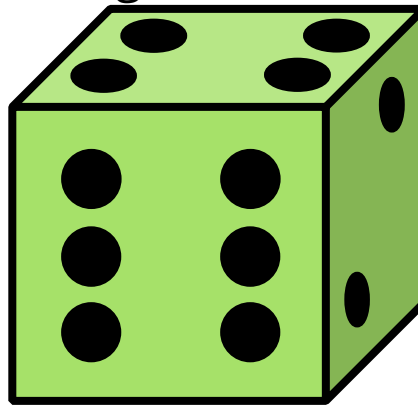
Schiefe = Maß für die Abweichung der Häufigkeitsverteilung von der Normalverteilung.



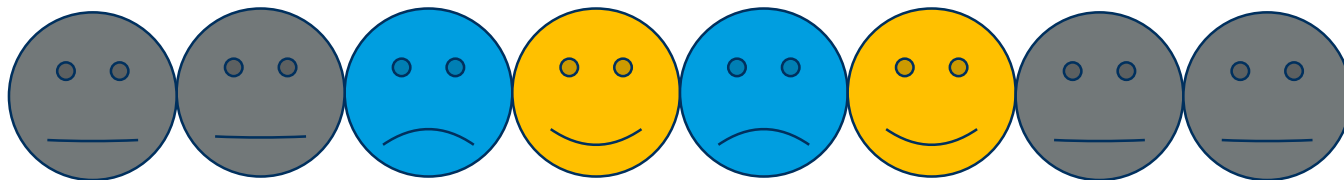
# Stichproben

# Von Wahrscheinlichkeit zur Statistik

Bei Münzen, Würfeln und Roulettespielen sind die Wahrscheinlichkeiten für das Auftreten eines konkreten Ereignisses **bekannt**.



Bei anderen Beispielen (aus Sozialwissenschaft) sind die Wahrscheinlichkeiten **nicht eindeutig bekannt**, dann reden wir von Statistik mit Stichproben.



# Stichproben, Grundgesamtheit

Das ist unsere Grundgesamtheit (Population), alle Personen eine Gruppe



Wir ziehen zufällig 10 Personen, das ist unsere Stichprobe:



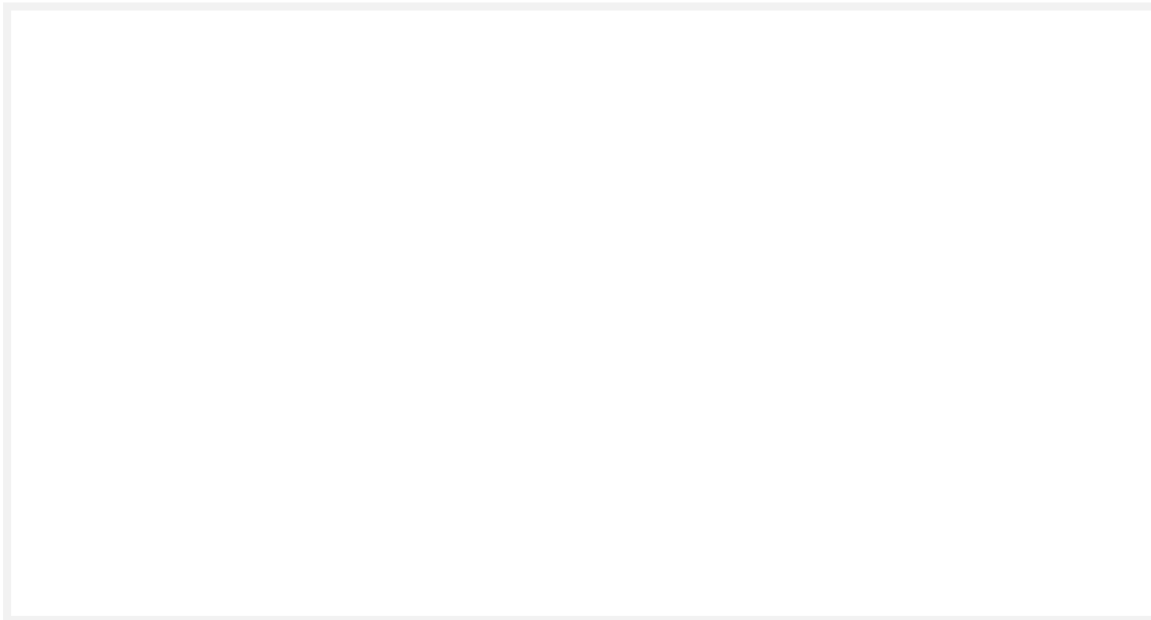
Wir analysieren diese Stichprobe und machen Rückschlüsse auf die Grundgesamtheit.

Wahrscheinlichkeit für gelb ist  $\frac{\text{Anzahl Gelber}}{\text{Anzahl der Elemente in der Stichprobe}} = \frac{8}{10} = 80 \%$

Warum stimmen die 80 % aus der Stichprobe mit der Grundgesamtheit nicht über ein? Wie sicher können wir die Rückschlüsse und Interpretationen aus der Stichprobe auf die Grundgesamtheit ziehen?

# Einführung in die Stichprobentheorie

## Warum werden Stichproben gezogen?



# Übung zu Grundgesamtheit und Stichprobe

Eine Politikerin möchte wissen, wie viel ältere Menschen wöchentlich Sport machen. Was ist ihre Grundgesamtheit?

Eine Wissenschaftlerin möchte den Zusammenhang untersuchen zwischen der Motivation und Arbeitsleistung. Sie macht Experimente mit Psychologiestudenten. Was ist ihre Grundgesamtheit und ist die Stichprobe geeignet? Welchen Einfluss kann die Stichprobe auf das Ergebnis haben?

Auf einer Internetseite werden Sie gefragt, wenn Sie wählen würden, wenn nächsten Sonntag Wahl wäre? Wer soll die Grundgesamtheit sein? Aus welcher Grundgesamtheit wird die Stichprobe tatsächlich gewählt?

Eine Initiative möchte das Gesundheitsverhalten von sehr armen Menschen analysieren. Sie ziehen dazu eine Stichprobe aus dem Melderegister. Ist es eine Zufallsauswahl? Wer ist die Grundgesamtheit, wenn erreichen Sie.

# Arten von Stichproben

Einfache Zufallsstichprobe:


Geschichtete Auswahl:

Klumpenstichprobe:

Mehrstufig Auswahl:

Bewusste Auswahl:

Schneeballsystem:

- 
- All diesen Verfahren haben eine Regel
  - Sie sind nicht willkürlich
  - Die Grundgesamtheit ist immer angebar

# Übung Stichproben

Sie messen den Benzinpreis in Dresden an 25 Tankstellen und kommen auf einen Mittelwert von 1,20 Euro. Sie schließen nun, dass in Sachsen der Mittelwert bei 1,20 Euro liegt. Was stimmt hierbei nicht?

Der Verein für eine „sportliche Stadt“ hat unter seinen Mitgliedern eine Umfrage zum Thema Ausbau der Sportstätten gemacht. Das Ergebnis: 90 % der Befragten sprechen sich für mehr Investition in die Sportstätten aus. Was stimmt hierbei nicht?

Eine Kosmetikartikelhersteller hat anhand einer Stichprobe ( $n = 7$ ) herausgefunden, dass sein Produkt zu einem jüngeren Äußeren führt und das dies auch zwei Stunden nach auftragen des Produkts noch so ist. Was stimmt hier nicht?

Ein Marktforschungsinstitut hat auf einer Einkaufsstraße 1000 Leute nach Ostern befragt und herausgefunden, dass kaum einer der Befragten in den nächsten 5 Tagen Schokolade essen möchte. Was stimmt hier nicht?

Ein Trainerin testet, ob Hallentraining oder Outdoortraining besser ist. Dazu lässt sie die Teilnehmenden freiwillig entscheiden, in welche Gruppe sie möchten. Was stimmt hier nicht?

Eine Firma wirbt mit einer Lieferzeit von 3 Tagen. Der Verbraucherschutz macht nach Beschwerden eine Stichprobe und kommt auf eine durchschnittliche Lieferzeit von 5 Tagen. Hat die Firma gelogen?

# Übung: Zufallsstichprobe ziehen

Unsere **Grundgesamtheit** sind die Zahlen 1-100. Wir ziehen daraus nun 30 Mal eine Stichprobe von 30 Zahlen und bilden daraus jeweils den Mittelwert. Beispielsweise können Sie diese Seite auf A4 drucken und die Zahlen einzeln ausschneiden. Anschließend alle Zahlen mischen und 30 Mal 30 Zahlen ziehen.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

# Wieso entstehen zufällige Ergebnisse?

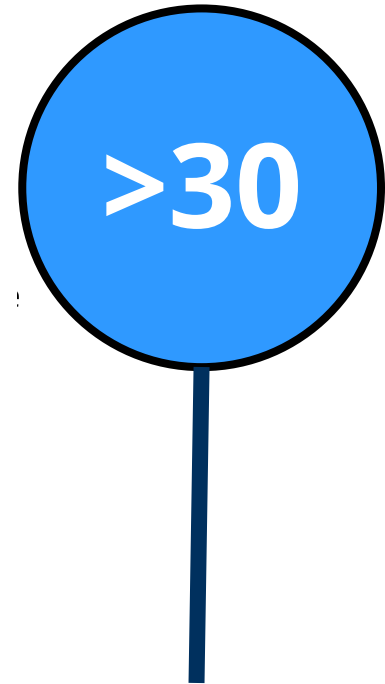
... weil keine Totalerhebung der Grundgesamtheit gemacht wird, sondern Stichproben gezogen werden.

Beispiel: In unserer Stichprobe erscheinen nur die roten Zahlen

Grundgesamtheit FH	Grundgesamtheit BA	Stichprobe FH	Stichprobe BA
500 €	500 €	500 €	500 €
500 €	500 €	500 €	500 €
500 €	<del>500 €</del>	500 €	
500 €	<del>500 €</del>	500 €	
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	1.000 €	1.000 €	1.000 €
1.000 €	<del>1.000 €</del>	1.000 €	
1.500 €	1.500 €	1.500 €	1.500 €
1.500 €	1.500 €	1.500 €	1.500 €
<del>1.500 €</del>	1.500 €		1.500 €
<del>1.500 €</del>	1.500 €		
2.000 €	2.000 €	2.000 €	2.000 €
<del>2.000 €</del>	2.000 €		2.000 €
<del>2.000 €</del>	2.000 €		2.000 €

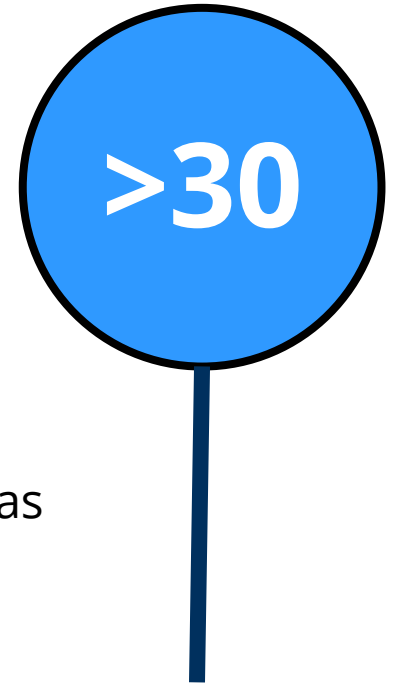
	FH	BA
Unsere Stichprobe		
In Wirklichkeit		

# Aus der Praxis: Stichprobengröße berechnen in der Praxis und warum wir das nicht lernen



## 2. Grund warum eine Stichprobe von $n > 30$ ?

In der vorhergehenden Übung haben wir 30 Stichproben gezogen mit einer jeweiligen Größe von 30? Warum?



Wir werden es (so gut wie) nie mit Variablen in den Sozialwissenschaften haben, die wirklich normalverteilt sind. Das heißt wir könnten viele Berechnungen (die wir später in der Veranstaltung machen wollen) nicht rechnen, weil dafür die Voraussetzung ist, dass die Daten normalverteilt sind. Das wär ungünstig. Daher gibt es eine Ausnahme (die empirische mathematisch auch begründet ist): Wenn Ihre Stichprobe mindestens 30 Fälle hat, dann kann man mindestens annehmen das ein aus so einer Stichprobe gewonnener Mittelwert normalverteilt ist und man schön weiter rechnen kann.

(Wenn es um die Berechnung von Anteilswerten ( $p$ ) geht gibt es eine andere Formel:  $p \cdot n > 5$ .)