

BSc PHY – Angewandte Statistik

LE1 Deskriptive Statistik

Patric Eichelberger & Aglaja Busch
aF&E Physiotherapie



patric.eichelberger@bfh.ch | aglaja.busch@bfh.ch



Moodlekurs

9. Mai 2025

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Intro

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Lernziele Deskriptive Statistik

- ▶ Die Studierenden können die Begriffe Population und Stichprobe abgrenzen.
- ▶ Die Studierenden können aus einer Population eine zufällige Stichprobe ziehen.
- ▶ Die Studierenden können den Datentyp eines Merkmals bestimmen.
- ▶ Die Studierenden können Daten mit Hilfe von Kennzahlen und Diagrammen beschreiben.
- ▶ Die Studierenden können die Symmetrie von Daten beurteilen.

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Population und Stichprobe

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

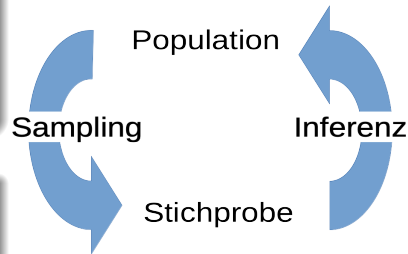
Statistik

Deskriptive Statistik

Mit Hilfe der Methoden der deskriptiven (**beschreibenden**) Statistik werden die Informationen aus Daten mit grafischen Methoden und Kennzahlen möglichst übersichtlich zusammengefasst.

Inferenzielle Statistik

Aufgrund von Modellvorstellungen wird in der inferenziellen (**schliessenden**) Statistik anhand von Stichproben auf Unterschiede und Zusammenhänge in einer Population geschlossen.



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen
Streuungskennzahlen
Grafische Darstellung

Visualisierung

Zusammenfassung

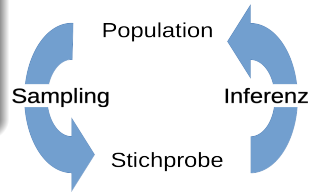
Population und Stichprobe

Sämtliche Mitglieder einer genau definierten Gruppe werden als **Population** (oder Grundgesamtheit) im statistischen Sinn bezeichnet.

Beispiele

- ▶ Alle Studierenden an der BFH Gesundheit
- ▶ Alle in der Stadt Bern wohnhaften Personen
- ▶ Alle Personen mit einer bestimmten Krankheit

Meistens wird bei einem Experiment nur ein Teil der Population, eine sogenannte **Stichprobe** untersucht. Diese Mitglieder der Population müssen zufällig aus der Population ausgewählt werden, damit man später auf die Population schliessen kann.



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

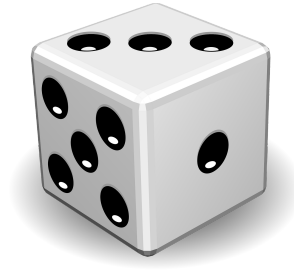
Grafische Darstellung

Visualisierung

Zusammenfassung

Einfache Stichprobe

- ▶ Eine einfache Stichprobe erhält man, indem man zufällig einige Nummern bzw. einige Personen aus der Population auswählt.
- ▶ Man verwendet Zufallszahlen aus Tabellen oder erzeugt sie mit einem Zufallsgenerator eines Computers.
- ▶ Im Web z.B. <https://www.randomizer.org/>



Quelle: OpenClipart-Vectors auf pixabay.com

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Abhängig/Unabhängig

Abhängige (verbundene/gepaarte) Stichprobe

Erhält man wenn man zum Beispiel wiederholte Messungen zu unterschiedlichen Zeitpunkten an den selben Personen durchführt.

Unabhängige (unverbundene, ungepaarte) Stichprobe

Wenn man zum Beispiel ein Merkmal zwischen Frauen und Männern vergleichen will sind die Gruppen unabhängig.

Ob die Gruppen abhängig oder unabhängig sind muss bei der statistischen Analyse berücksichtigt werden.

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung



Datentypen

- ▶ Generell unterscheidet man **qualitative** (nicht messbare) und **quantitative** Merkmale, wobei die Einteilung nicht immer eindeutig ist.
- ▶ Bei der Beschreibung von Daten ist auf ihre Art Rücksicht zu nehmen.

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Qualitative Merkmale

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Qualitative Merkmale

Qualitative Merkmale werden oft auch als kategorielle Merkmale bezeichnet:

- ▶ **Nominale** Merkmale: Merkmalswerte können nicht (natürlich) geordnet werden.

Beispiel

Geschlecht, Herkunft (Kanton, Land), Studienfach

- ▶ **Ordinale** Merkmale: Merkmalswerte können (natürlich) geordnet werden. Abstände zwischen Merkmalswerten sind irrelevant.

Beispiel

Schmerzen (schwach, mittel, stark), Kleidergrößen (S, M, L)

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen
Streuungskennzahlen
Grafische Darstellung

Visualisierung

Zusammenfassung

Häufigkeitstabellen

Eine gute Darstellungsmöglichkeit für nominale und ordinale Daten sind Häufigkeitstabellen.

Beispiel

$n = 88$ Personen wurden gefragt, wieviele Kaffees sie pro Tag trinken. In einer Häufigkeitstabelle wird zu jeder Kategorie (hier Anzahl Kaffees pro Tag) j die absolute Häufigkeit h_j (Anzahl Personen, die j Kaffees pro Tag trinken) und/oder die relative Häufigkeit $p_j = h_j/n$ eingetragen.

j	h_j	p_j
0	1	0.01
1	6	0.07
2	23	0.26
3	27	0.31
4	22	0.25
5	8	0.09
6	1	0.01
88		1

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

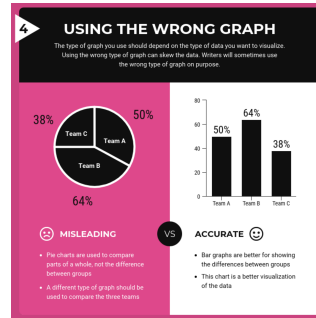
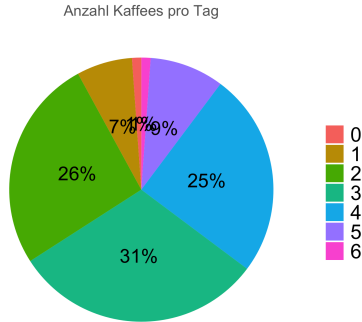
Visualisierung

Zusammenfassung

Kuchendiagramm

Eine weitere Darstellungsmöglichkeit für nominale oder ordinale Daten ist ein Kuchendiagramm.

- ▶ Unübersichtlich bei vielen Kategorien
- ▶ Für menschliches Auge schwierig, die verschiedenen Winkel zu vergleichen



Quelle: <https://venngage.com/blog/misleading-graphs/>

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

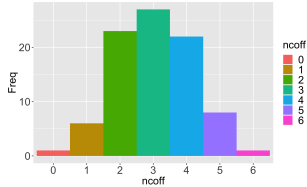
Visualisierung

Zusammenfassung

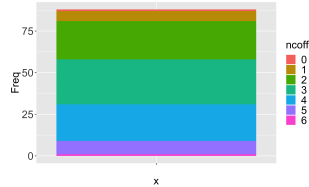
Balkendiagramm

Eine bessere Darstellungsmöglichkeit als das Kuchendiagramm für nominale oder ordinale Daten ist ein Balkendiagramm.

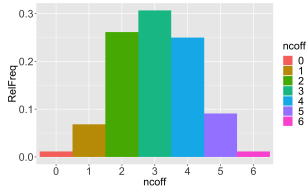
Balken absolute Häufigkeit



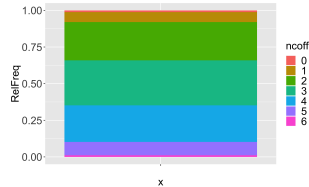
Gestapelt absolute Häufigkeit



Balken relative Häufigkeit



Gestapelt relative Häufigkeit



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Quantitative Merkmale

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Quantitative Merkmale

- ▶ **Ordinale** Merkmale (mit quantitativer Bedeutung): Merkmalswerte können (natürlich) geordnet werden.

Beispiel

Visuelle Analogskala (VAS), Intelligenzquotient (IQ), Schulnoten, Windstärken
Beaufortskala, Ränge im Sport

- ▶ **Rationalskalierte Merkmale**: 0 ist ein absoluter Wert

Beispiel

Alter, Körpergrösse, Einkommen

- ▶ **Intervallskalierte** Merkmale: 0 ist ein willkürlicher Wert

Beispiel

Temperatur in Grad Celsius oder Fahrenheit

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung



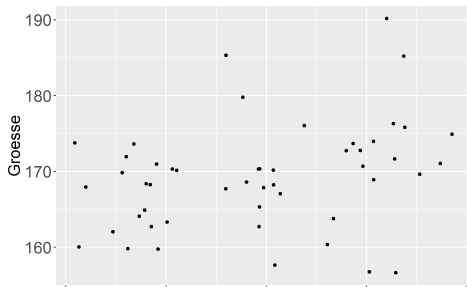
Quantitative Merkmale (cont.)

Vorliegend seien die Körpergrössen in cm von $n = 50$ Studierenden.

Fragen

- ▶ Wo ist die Mitte?
- ▶ Wie gross ist die Streuung?
- ▶ Welche Form hat die Verteilung der Daten?

- ▶ Lagekennzahlen
- ▶ Streuungskennzahlen
- ▶ Diagramme



Sample	Körpergrösse (cm)
x_1	174
x_2	168
x_3	185
x_4	161
x_5	173
...	...
x_n	...

Quantitative Merkmale

Lagekennzahlen

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Mittelwert

Stichprobe von n Beobachtungen x_1, x_2, \dots, x_n

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Sample	Körpergrösse (cm)
x_1	174
x_2	168
x_3	185
x_4	161
x_5	173
...	...
x_n	...

Beispiel

$x_1 = 174, x_2 = 168, x_3 = 185, x_4 = 161, x_5 = 173$

$$\bar{x} = \frac{174 + 168 + 185 + 161 + 173}{5} = 172.2$$

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

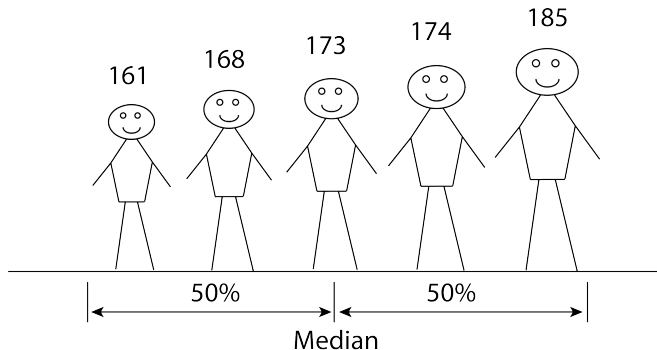
Grafische Darstellung

Visualisierung

Zusammenfassung

Median

Der Median teilt die Stichprobe in zwei Hälften. 50% der Daten sind kleiner als der Median, 50% der Daten sind grösser als der Median. Der Median ist sozusagen der mittlere Wert.



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Median: Berechnung ungerade Anzahl Datenpunkte

Stichprobe

Sample	Körpergrösse (cm)
x_1	174
x_2	168
x_3	185
x_4	161
x_5	173

Geordnete Stichprobe

Rang	Körpergrösse (cm)
$x_{(1)}$	161
$x_{(2)}$	168
$x_{(3)}$	173
$x_{(4)}$	174
$x_{(5)}$	185

Median n ungerade

$$\begin{aligned}\tilde{x} &= x_{(\frac{n+1}{2})} \\ &= x_{(\frac{5+1}{2})} = x_{(3)} = 173\end{aligned}$$

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

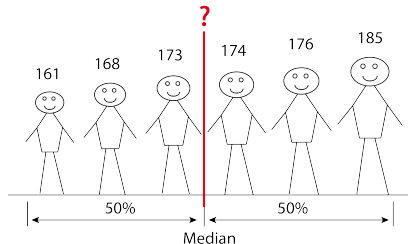
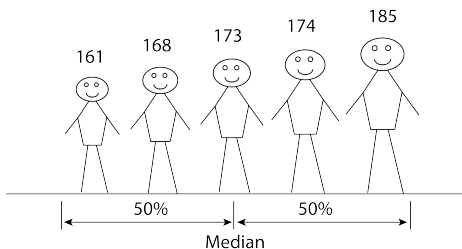
Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Median: Berechnung gerade Anzahl Datenpunkte



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Median: Berechnung gerade Anzahl Datenpunkte (cont.)

Stichprobe

Sample	Körpergrösse (cm)
x_1	174
x_2	168
x_3	185
x_4	161
x_5	173
x_6	176

Geordnete Stichprobe

Rang	Körpergrösse (cm)
$x_{(1)}$	161
$x_{(2)}$	168
$x_{(3)}$	173
$x_{(4)}$	174
$x_{(5)}$	176
$x_{(6)}$	185

Median n gerade

$$\begin{aligned}\tilde{x} &= \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) \\ &= \frac{1}{2}(x_{(\frac{6}{2})} + x_{(\frac{6}{2}+1)}) = \frac{1}{2}(x_{(3)} + x_{(4)}) = \frac{1}{2}(173 + 174) = 173.5\end{aligned}$$

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

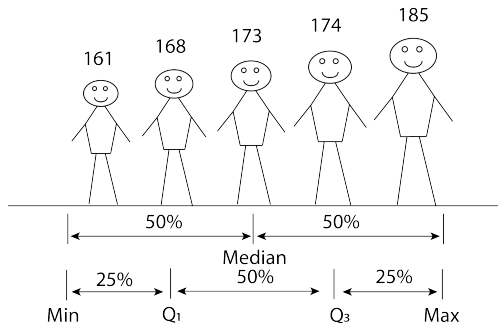
Grafische Darstellung

Visualisierung

Zusammenfassung

Minimum, Maximum und Quantile

- ▶ Das Minimum $x_{(1)}$ und das Maximum $x_{(n)}$ beschreiben die extreme Lage.
- ▶ Das untere/erste Quartil Q_1 teilt die Stichprobe im Verhältnis 1:3 (d.h. 25% aller Werte liegen darunter, 75% darüber), das obere/dritte Quartil Q_3 teilt die Stichprobe im Verhältnis 3:1 (d.h. 75% aller Werte liegen darunter, 25% darüber).



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

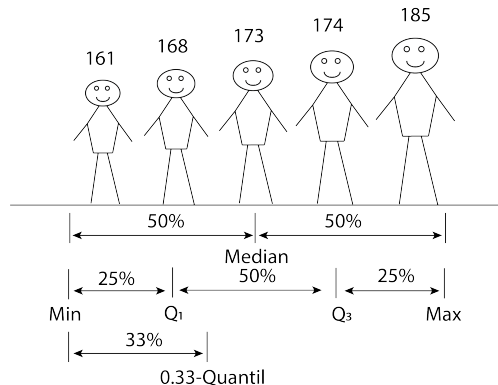
Grafische Darstellung

Visualisierung

Zusammenfassung

Minimum, Maximum und Quantile (cont.)

- ▶ Das p -Quantil x_p oder $100 \cdot p\%$ -Quantil (auch Perzentil genannt) teilt die Stichprobe im Verhältnis $p : (1 - p)$
- ▶ $100 \cdot p\%$ aller Werte liegen darunter, $100(1 - p)\%$ darüber)
- ▶ Median: $x_{0.5}$
- ▶ 1. Quartil: $x_{0.25}$
- ▶ 3. Quartil: $x_{0.75}$



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Berechnung Quantile

- ▶ $\lfloor x \rfloor$ bezeichnet die Abrundungsfunktion. Bsp.: $\lfloor 1, 2 \rfloor = 1$
- ▶ Gegeben sei eine geordnete Stichprobe (x_1, x_2, \dots, x_n) wo gilt $x_1 \leq x_2 \leq \dots \leq x_n$

Das empirische p -Quantil von (x_1, x_2, \dots, x_n) mit $p \in (0, 1)$ ist

$$x_p = \begin{cases} \frac{1}{2} (x_{n \cdot p} + x_{(n \cdot p) + 1}), & \text{wenn } n \cdot p \text{ ganzzahlig,} \\ x_{\lfloor n \cdot p + 1 \rfloor}, & \text{wenn } n \cdot p \text{ nicht ganzzahlig.} \end{cases}$$

- ▶ Dies ist **nur eine mögliche Berechnungsart!**
- ▶ R kennt in der Funktion `quantile(x,p,type)` neun verschiedene Typen ¹
- ▶ Die Typen 2 und 5 arbeiten nach dieser Formel

¹Hyndman (1996) Am Stat 50, 361-365

Berechnung Quantile: Beispiel

Geordnete Stichprobe: $X = \{161, 168, 173, 174, 176, 185\}$

▶ $x_{0.25}$ bei $n = 6$

▶ $n \cdot p = 6 \cdot 0.25 = 1.5$

▶ $n \cdot p$ ist nicht ganzzahlig

$$\begin{aligned}x_{0.25} &= x_{(\lfloor 1.5+1 \rfloor)} \\&= x_{(2)} \\&= 168\end{aligned}$$

▶ $x_{1/3}$ bei $n = 6$

▶ $n \cdot p = 6 \cdot 1/3 = 2$

▶ $n \cdot p$ ist ganzzahlig

$$\begin{aligned}x_{1/3} &= \frac{1}{2} (x_{(2)} + x_{(3)}) \\&= \frac{1}{2} (168 + 173) \\&= 170.5\end{aligned}$$

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Berechnung Quantile jamovi / R

Descriptives

Descriptives

	KG
Median	173.500
Minimum	161
Maximum	185
25th percentile	169.250
50th percentile	173.500
75th percentile	175.500

```
for (i in seq(1:9)) {  
  print(paste("##### Quantile type", i, " #####"))  
  xp = quantile(greven,seq(0, 1, 0.25),type=i)  
  print(xp)  
}
```

[1] "##### Quantile type 1 #####"
0% 25% 50% 75% 100%
161 168 173 176 185

[1] "##### Quantile type 2 #####"
0% 25% 50% 75% 100%
161.0 168.0 173.5 176.0 185.0

[1] "##### Quantile type 3 #####"
0% 25% 50% 75% 100%
161 168 173 174 185

[1] "##### Quantile type 4 #####"
0% 25% 50% 75% 100%
161.0 164.5 173.0 175.0 185.0

[1] "##### Quantile type 5 #####"
0% 25% 50% 75% 100%
161.0 168.0 173.5 176.0 185.0

[1] "##### Quantile type 6 #####"
0% 25% 50% 75% 100%
161.00 166.25 173.50 178.25 185.00

[1] "##### Quantile type 7 #####"
0% 25% 50% 75% 100%
161.00 169.25 173.50 175.50 185.00

[1] "##### Quantile type 8 #####"
0% 25% 50% 75% 100%
161.0000 167.4167 173.5000 176.7500 185.0000

[1] "##### Quantile type 9 #####"
0% 25% 50% 75% 100%
161.0000 167.5625 173.5000 176.5625 185.0000

Default

- ▶ R: Typen 2 und 5 entsprechen den manuellen Berechnungen
- ▶ jamovi rechnet nach Typ 7 (default)

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

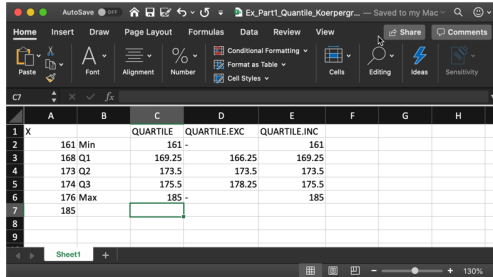
Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

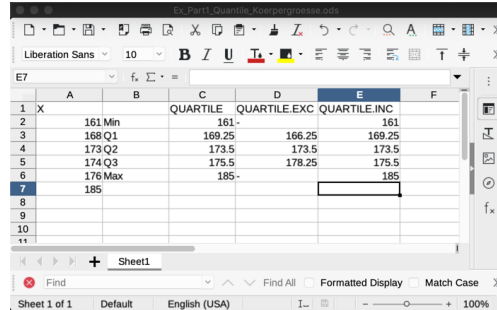
Berechnung Quantile MS Excel / LibreOffice Calc



The screenshot shows the MS Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F	G	H
1	X		QUARTILE	QUARTILE.EXC	QUARTILE.INC			
2		161 Min	161 -		161			
3		168 Q1	169.25	166.25	169.25			
4		173 Q2	173.5	173.5	173.5			
5		174 Q3	175.5	178.25	175.5			
6		176 Max	185 -		185			
7		185						
8								
9								

MS Excel



The screenshot shows the LibreOffice Calc interface with the following data in the spreadsheet:

	A	B	C	D	E	F
1	X		QUARTILE	QUARTILE.EXC	QUARTILE.INC	
2		161 Min	161 -		161	
3		168 Q1	169.25	166.25	169.25	
4		173 Q2	173.5	173.5	173.5	
5		174 Q3	175.5	178.25	175.5	
6		176 Max	185 -		185	
7		185				
8						
9						
10						
11						

LibreOffice Calc

- ▶ MS Excel und LibreOffice rechnen gleich
- ▶ QUARTILE() oder QUARTILE.INC() verwenden!

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Quantitative Merkmale

Streuungskennzahlen

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Spannweite und IQR

Spannweite

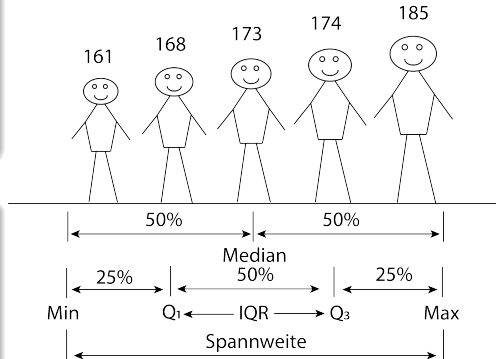
Die Spannweite (auch Variationsbreite genannt) misst den Wertebereich der Stichprobe:

$$\text{Spannweite} = \text{Maximum} - \text{Minimum}$$

IQR

Der Interquartilsabstand (auch Quartilsabstand genannt) ist durch die Differenz des oberen und des unteren Quartils definiert:

$$\text{IQR} = Q_3 - Q_1$$



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Varianz und Standardabweichung

Die Varianz ist ein Mass für die mittlere quadratische Abweichung der Daten zu ihrem Mittelwert.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}$$

Die Standardabweichung ist definiert als die Quadratwurzel der Varianz.

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

Beispiel Körpergrösse

$$s^2 = (174 - 172.2)^2 + (168 - 172.2)^2 + (185 - 172.2)^2 + (161 - 172.2)^2 + (173 - 172.2)^2 / 4 = 77.7$$

$$s = \sqrt{77.7} = 8.8$$

Sample	Körpergrösse (cm)
x_1	174
x_2	168
x_3	185
x_4	161
x_5	173
...	...
x_n	...

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Standardfehler

- ▶ Der Standardfehler ist in der Literatur auch hie und da zu finden.
- ▶ Alternative Bezeichnung: Standardfehler des Mittelwerts.
- ▶ Englische Bezeichnung ist standard error (SE) oder standard error of the mean (SEM).
- ▶ Wichtig im Zusammenhang mit der Berechnung von Vertrauensintervallen (folgt in Teil III Schätzung).

Definition

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

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Quantitative Merkmale

Grafische Darstellung

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

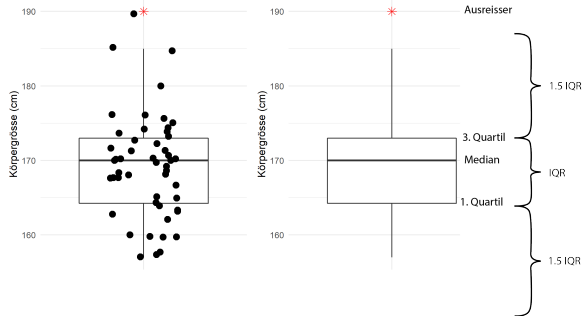
Grafische Darstellung

Visualisierung

Zusammenfassung

Boxplot

- ▶ Wichtige Darstellungsmöglichkeit für quantitative Daten.
- ▶ Auch Box-Whisker-Plot genannt.
- ▶ Als Richtwert ab einer Stichprobengröße von $n = 15$ einsetzen. Ansonsten Streudiagramm verwenden um einzelne Datenpunkte darzustellen.
- ▶ Überlagerung mit einzelnen Datenpunkten bei kleinen Stichproben sinnvoll.
- ▶ Symmetrie der Daten: Whiskers etwa gleich lang



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

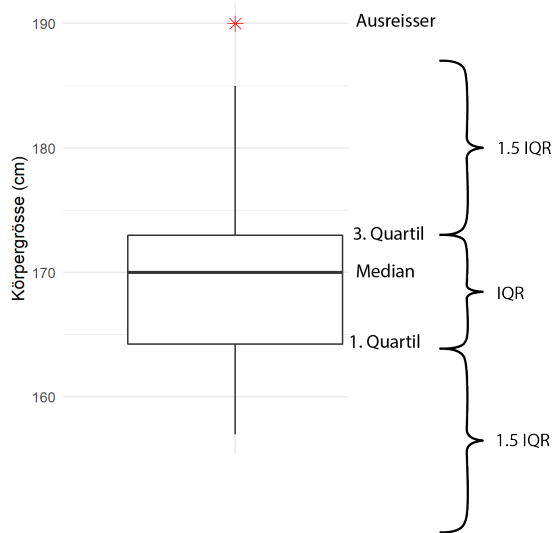
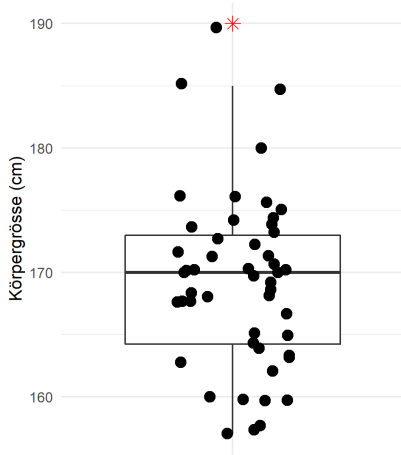
Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Boxplot (cont.)



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

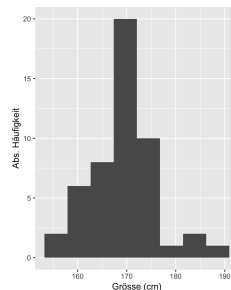
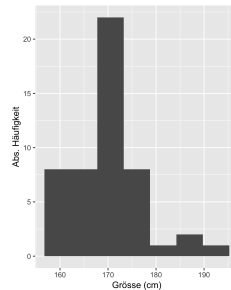
Grafische Darstellung

Visualisierung

Zusammenfassung

Histogramm

- ▶ Beurteilung von
 - ▶ Symmetrie/Asymmetrie
 - ▶ Spitzen/Täler
 - ▶ Ausreisser
- ▶ Sehr wichtig
 - ▶ Gute Klasseneinteilung und Anzahl Klassen (5 bis 20)
 - ▶ Schlechte Klassenwahl kann zu Verfälschungen führen
 - ▶ Bsp. rechts: oben 7 unten 8 Klassen
- ▶ Sinnvoll ab einer Stichprobengrösse von $n = 50$
- ▶ Zu jeder Klasse ein Balken mit absoluter oder relativer Häufigkeit



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

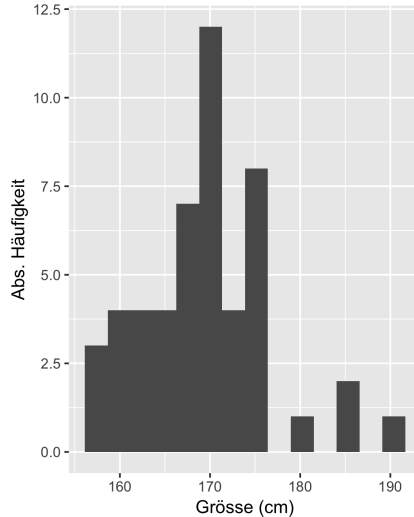
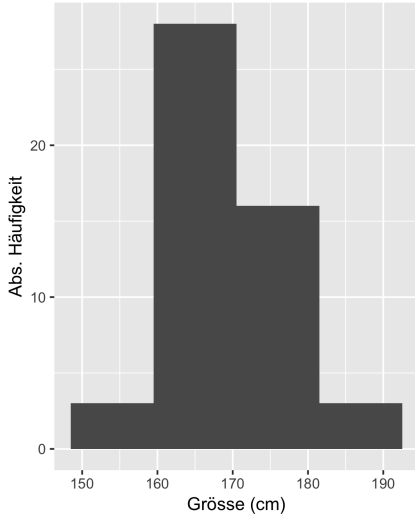
Grafische Darstellung

Visualisierung

Zusammenfassung

Histogramm (cont.)

Die gleichen Daten mit zu wenigen und zu vielen Klassen:



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

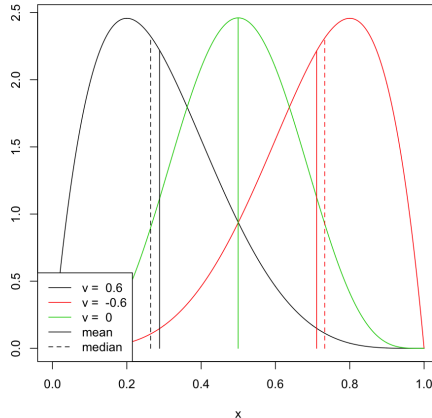
Grafische Darstellung

Visualisierung

Zusammenfassung

Schiefe (Skewness)

- ▶ Kennzahl welche die Symmetrie/Asymmetrie eines Datensatzes misst
- ▶ $v > 0$
 - ▶ Verteilung heisst rechtsschief oder linkssteil
 - ▶ "Gipfel" befindet sich links vom Mittelwert
- ▶ $v < 0$
 - ▶ Verteilung heisst linksschief oder rechtssteil
 - ▶ "Gipfel" befindet sich rechts vom Mittelwert



Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Visualisierung

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

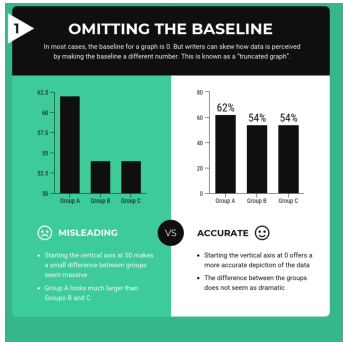
Streuungskennzahlen

Grafische Darstellung

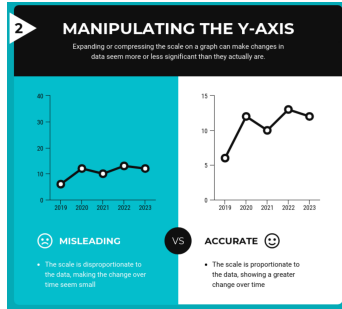
Visualisierung

Zusammenfassung

Stolperfalle y-Achse



- ▶ Start der y-Achse bei 50% stellt die Differenzen zwischen den Gruppen übertrieben dar.



- ▶ Dem Wertebereich unangepasste Skalierung verfälscht den Eindruck!

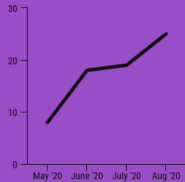
Bilder: venngage.com

Stolperfalle Cherry Picking

3

CHERRY PICKING DATA

Writers may only include certain data points on their graphs to reinforce their narratives. This can create a false impression of the data.



MISLEADING

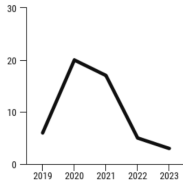
- Only a few months out of the year are graphed, depicting an upward trends

VS

ACCURATE



- A much wider date range is graphed, revealing an overall downward trend
- This graphs shows the bigger picture



- ▶ Reduktion der Darstellung auf einen ausgewählten Bereich der Daten kann zu grosser Verzerrung führen.

Quelle: venngage.com

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

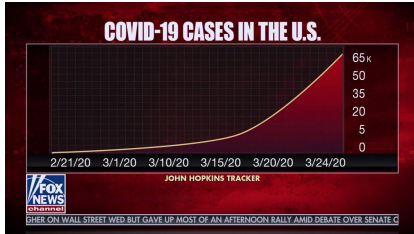
Visualisierung

Zusammenfassung

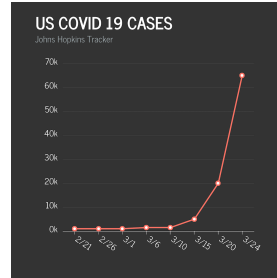


Bern University of Applied Sciences

Stolperfalle Manipulation x- und y-Achse



- ▶ Irreführung durch falsche Skalierung. Die Abstände der Datenpunkte sind jeweils unterschiedlich.
- ▶ y: Sprung von 0 auf 5, dann von 5 auf 20.
- ▶ x: Mal liegen mehr mal weniger Tage zwischen den Datenpunkten.



- ▶ Covid-Fälle vs. Zeit bei richtiger Achsskalierung.

Bilder: [venngage.com](https://www.venngage.com)

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

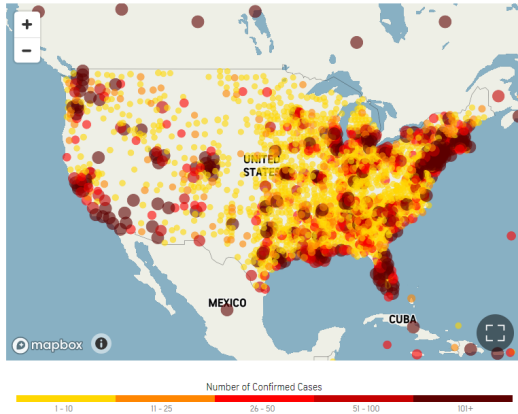
Grafische Darstellung

Visualisierung

Zusammenfassung

Die falschen Daten

CONFIRMED CASES BY REGION



Quelle: [venngage.com](https://www.venngage.com)

- ▶ Dargestellt sind absolute Häufigkeiten von Fällen für verschiedene Regionen.
- ▶ Es ist klar, dass es in grösseren Städten mehr Fälle gibt, weil dort mehr Personen leben. Die Grafik zeigt also eigentlich nichts anderes, als die Besiedlungsdichte.
- ▶ Richtig wäre die Darstellung der Anzahl Fälle bezogen auf die Anzahl Einwohner.
- ▶ Weiter ist die oberste Skalastufe mit 101k+ nicht sinnvoll. Es ist relevant ob es 1'000 oder 10'000 Fälle sind!

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

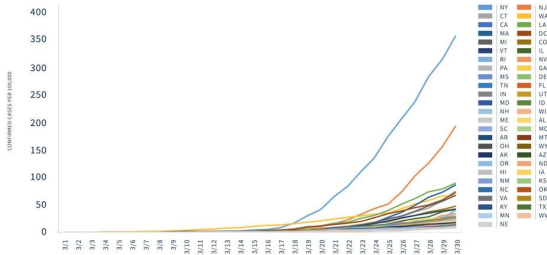
Grafische Darstellung

Visualisierung

Zusammenfassung

Zu viele Daten

CUMULATIVE CASES PER 100,000: ALL STATES



Quelle: venngage.com

- ▶ Du viele Datenreihen machen eine Grafik schwer zu lesen.
- ▶ Im Beispiel sehr deutlich, weil die y-Spannweite gleichzeitig sehr gross ist.

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

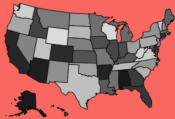
Visualisierung

Zusammenfassung

Entgegen der Intuition

5 GOING AGAINST CONVENTIONS

Over time, we have developed standards for how data is visualized. Flipping those conventions can make a graph confusing or misleading to readers.



Individuals per km
0 20 40 60

☹ MISLEADING

- Normally, darker shades are associated with density on a map but here, dark has been used to depict lower population density
- This graph can confuse and mislead readers, who expect dark to represent a higher population density

VS



Individuals per km
0 20 40 60

😊 ACCURATE

- This map follows the convention of using lighter shades for lighter density and darker shades for higher density
- Readers will intuitively know how to interpret the data

- ▶ Dunklere Farbtöne assoziieren wir mit höheren Werten.
- ▶ Die linke Abbildung ist irreführend, weil die Graustufen entgegen der etablierten Konvention sind.
- ▶ Konventionen sind beim Visualisieren von Daten einzuhalten.

Quelle: venngage.com

Intro

Population und Stichprobe

Qualitative Merkmale

Quantitative Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Zusammenfassung

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Grafische Darstellung

Visualisierung

Zusammenfassung

Zusammenfassung

- ▶ Es muss zwischen der Grundgesamtheit und einer betrachteten Stichprobe unterschieden werden.
- ▶ Eine Stichprobe ist nicht zwingend repräsentativ was die Gültigkeit einer Statistik stark beeinflussen kann.
- ▶ Wir unterscheiden zwischen beschreibender und schliessender Statistik sowie quantitativen (rational- und intervallskaliert) und qualitativen (nominal, ordinal) Merkmalen.
- ▶ Wir benötigen graphische Werkzeuge wie Balkendiagramm, Histogramm und Boxplot um Daten visuell zu beschreiben.
- ▶ Wir haben Lage- (Mittelwert, Median, Perzentile) und Streumasse (Varianz, Spannweite, IQR) gesehen um Daten zusammenfassend mit Zahlen zu beschreiben.

Intro

Population und
Stichprobe

Qualitative
Merkmale

Quantitative
Merkmale

Lagekennzahlen

Streuungskennzahlen

Graphische Darstellung

Visualisierung

Zusammenfassung