



von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
senerscheinung  
Merkmale au

# **Statistik**

Terminologie

# Terminologie

- > **Merkmalsträger** (*unit*)
  - Untersuchte statistische Einheit  
z. B. Mensch, Baum, Computer, Software, ...
  
- > **Merkmal** (*attribute, data*)
  - Eigenschaft des Merkmalträgers  
z. B. Größe, Laufzeit, Fehleranzahl, ...
  
- > (Merkmals-) **Ausprägung**
  - Konkreter Wert eines Merkmals  
z. B. Messdaten, Beobachtungen
  
- > **Datensatz** (*data set*)
  - Zusammengehörige Merkmale eines/mehrerer Merkmalsträger(s)

# Terminologie

## Typen von Merkmalen

### > Qualitative Merkmale

- Eigenschaft  
z. B. Geschlecht, Farbe, ...

### > Quantitative Merkmale

- Messbar  
z. B. Schuhgröße, Gewicht, Laufzeit, ...
- Unterscheidung in
  - Diskret: Abzählbar (z. B. Zahl der Augen eines Würfels, natürliche Zahlen)
  - Stetig: Alle Zwischenwerte realisierbar (z. B. rationale Zahlen)

# Terminologie

Stanley Smith Stevens (1906 – 1973), Psychologe

## Skalenniveaus von Merkmalen



### Nominal

- Nominalskala
- Sammlung von Kategorien
- **Keine** intrinsische Ordnung
- Beispiele:
  - Geschlecht
  - Marke
  - Partei
  - Hochschule



### Ordinal

- Ordinalskala
- Sammlung von Kategorien
- Intrinsische Ordnung
- Sortierbar
- Abstände müssen **nicht** gleich sein!
- Beispiele:
  - Größe: S,M,L,XL
  - Noten: 1,2,3,4,5
  - Medaillen: Gold, Silber, Bronze



### Kardinal

- Kardinalskala
  - Intervallskaliert
  - Verhältnisskaliert
- Sammlung geordneter Kategorien
- Abstände (Intervalle) **immer** gleich!
- Beispiele:
  - Größe
  - Temperatur

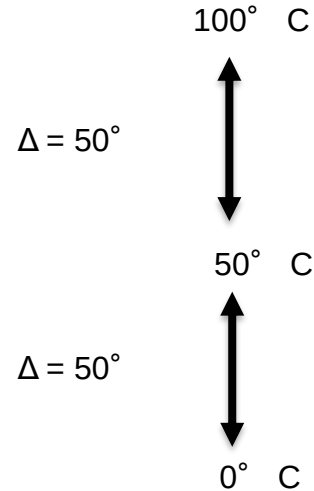
# Terminologie

## Skalenniveaus



### Kardinalskala

- Intervallskaliert
  - Gleiche Abstände
  - Beispiel
    - Celsius-Skala



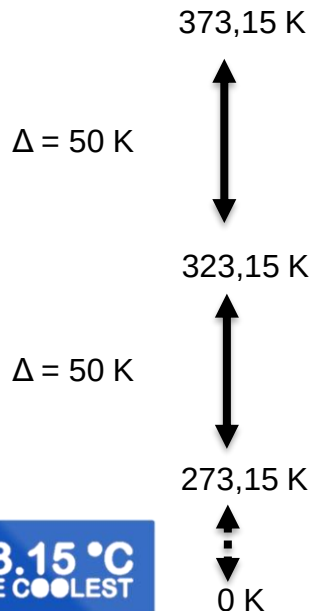
# Terminologie

## Skalenniveaus



### Kardinalskala

- Intervallskaliert
  - Gleiche Abstände
  - Beispiel
    - Celsius-Skala
- Verhältnisskaliert
  - Sinnvoller Nullpunkt
  - Beispiel
    - Kelvin Skala



$$\frac{373,15\text{K}}{323,15\text{K}} = 1,1547$$

$$\frac{100^{\circ}\text{C}}{50^{\circ}\text{C}} = 2$$

$$\frac{323,15\text{K}}{273,15\text{K}} = 1,1830$$

$$\frac{50^{\circ}\text{C}}{0^{\circ}\text{C}} = \text{undefiniert}$$

# Terminologie

## Skalenniveaus – Überblick

### > Nominalskala

- Qualitative Ausprägungen (z. B. männlich, weiblich)
- Falls Zahlen: nur Bezeichnungsfunktion (z. B. Artikelnummer, Postleitzahlen)

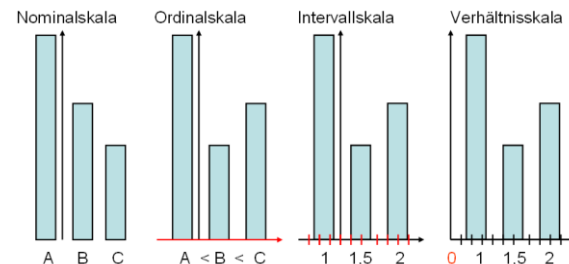
### > Ordinalskala

- Rangbildung möglich (Rangordnung, sortierbar) (z. B. Rangfolge bei Wettkampf, Schulnoten)
- Auch: sehr zufrieden > eher zufrieden > eher unzufrieden > sehr unzufrieden
- Mathematische Operationen i.d.R. nicht sinnvoll

### > Kardinalskala

- **Intervallskala**  
(z. B. Temperatur in ° C, Intelligenzquotient, Zeitpunkte)
- **Verhältnisskala**  
Absoluter (natürlicher) Nullpunkt vorhanden  
(z. B. Temperatur in K, Alter)

| Skala      | Kategorien                                                                          | Geordnet                            | Gleiche Intervalle                  | Sinnvoller Nullpunkt                |
|------------|-------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Nominal    |  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| Ordinal    |  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| Interval   |  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| Verhältnis |  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |



Wikipedia, Qwqchris~commonswiki, CC BY-SA 3.0

# Terminologie

## Skalendegression und Skalenprogression

- > Ziel der Skalierung
  - Angemessene Abbildung der Information (Daten)
  - Keine Über-/Unterschätzungen

### > Skalendegression

- Alle Merkmale können nominal skaliert werden
- Kardinale Merkmale können ordinal skaliert werden
- Aber: Informationsverlust!

### > Skalenprogression

- Nominale Merkmale dürfen **nicht** ordinal/kardinal skaliert werden!
- Ordinale Merkmale dürfen **nicht** kardinal skaliert werden!
- Gefahr der Fehlinterpretation, da Interpretation von Merkmalen ohne Information

| Skala                                                                                          | Kategorien                          | Geordnet                            | Gleiche Intervalle                  | Sinnvoller Nullpunkt                |
|------------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Nominal     | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | <input type="checkbox"/>            |
| Ordinal     | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |
| Intervall   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| Verhältnis  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |








von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
erscheinungen  
Merkmalen

# **Statistik**

Einführung in JMP (cont'd)

# Arbeiten mit Daten

## Merkmalseigenschaften – Typ, Skalenniveau, Einheit



The first screenshot shows the 'Spalten' (Columns) menu with 'Spalteninfo...' selected. The second screenshot shows the 'Spalteninfo' dialog box for 'Größe (in cm) 2' with 'Numerisch' selected for 'Datentyp' and 'Stetig' for 'Modellierungstyp'. The third screenshot shows the 'Spalteneigenschaften' (Column Properties) dialog box with 'Einheiten' (Units) set to 'cm'.

**Spalteninfo Dialog:**

- Spaltenname: Größe (in cm) 2
- Datentyp: Numerisch
- Modellierungstyp: Stetig
- Format: Bestes, Breite 12

**Spalteneigenschaften Dialog:**

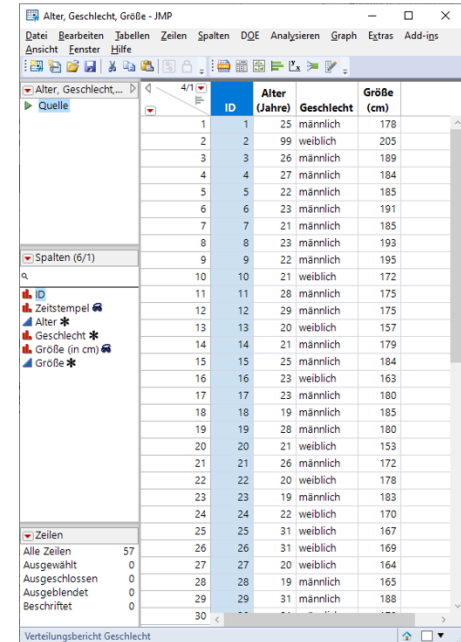
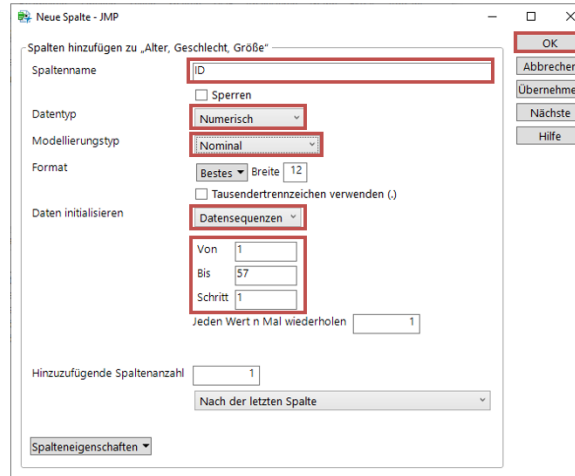
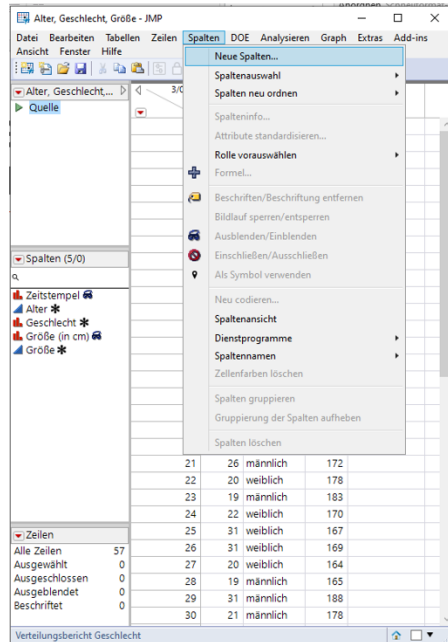
- Einheiten: cm

**Data Table:**

| Größe (in cm) | Größe (in cm) 2 |
|---------------|-----------------|
| 26 m          | 189             |
| 27 m          | 184             |
| 22 m          | 185             |
| 23 m          | 191             |
| 21 m          | 185             |
| 23 m          | 193             |
| 22 m          | 195             |
| 21 w          | 172             |
| 28 m          | 175             |
| 29 m          | 175             |
| 20 w          | 157             |
| 21 m          | 179             |
| 25 m          | 184             |
| 23 w          | 163             |
| 23 m          | 180             |
| 19 m          | 185             |
| 28 m          | 180             |
| 21 w          | 153             |
| 26 m          | 172             |
| 22            | 178             |
| 23            | 183             |
| 24            | 170             |
| 25            | 167             |
| 26            | 169             |
| 27            | 164             |
| 28            | 165             |
| 29            | 188             |
| 30            | 178             |
| 31            | 148             |

# Arbeiten mit Daten

## Spalten hinzufügen und initialisieren



# Arbeiten mit Daten

## Bereinigen / Fehleingaben ausschließen



**Fehleingabe selektieren**

Verteilung - JMP

Zeigt ein Histogramm und univariate Kenngrößen für jede Variable an.

Spalten auswählen: 6 Spalten

- ID
- Alter
- Geschlecht
- Größe (in cm)
- Größe

Ausgewählte Spalten den Rollen zuordnen

Y, Spalten: Alter, Geschlecht, Größe

Gewichtung: numerisch optional

Häufigkeit: numerisch optional

Nach: optional

☒ Nur Histogramme

Verteilungsbericht Geschlecht

Alter, Geschlecht, Größe - Verteilung - JMP

Verteilungen

Alter: 35, 30, 25, 20

Geschlecht: weiblich, männlich, divers

Größe: 210, 200, 190, 180, 170, 160

Alter, Geschlecht, Größe - JMP

Quelle

| ID | Alter (Jahre) | Geschlecht  | Größe (cm) |
|----|---------------|-------------|------------|
| 1  | 2             | 99 weiblich | 170        |
| 2  | 2             | 99 weiblich | 205        |
| 3  | 2             | 20 männlich | 169        |
| 4  | 4             | 27 männlich | 184        |

Alter, Geschlecht, Größe - JMP

Spalten

- Neue Spalten...
- Spaltenauswahl
- Spalten neu ordnen
- Spalteninfo...
- Attribut standardisieren...
- Rolle vorauswählen
- Formel...
- Beschriften/Beschriftung entfernen
- Bildlauf sperren/entsperren
- Ausblenden/Einblenden
- Einschließen/Ausschließen**
- Als Symbol verwenden
- Neu codieren...
- Spaltenansicht
- Dienstrechnungen

**Statistik**

Terminologie (cont'd)

# Terminologie

## > **Grundgesamtheit** (*population*)

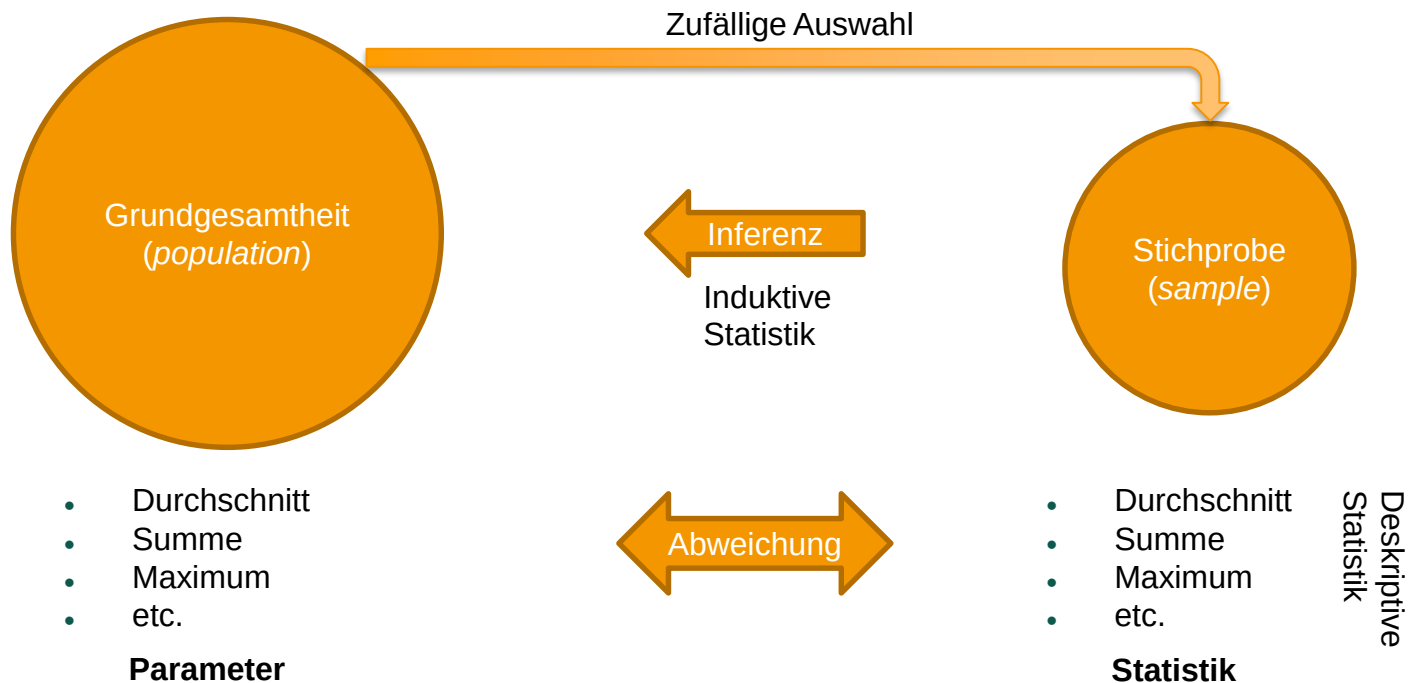
- Menge aller Individuen von Interesse (Merkmalsträger)  
z. B. Menschen in Deutschland, Studierende FKI, Bäume im Park, installierte Apps, ...
- Umgang abhängig von Anforderung
- In der Regel nicht komplett untersuchbar

## > **Stichprobe** (*sample*)

- Auswahl von Merkmalsträgern (*unit*) aus Grundgesamtheit (*population*)
- Ziel: Repräsentation der Grundgesamtheit
  - Aber: Keine exakte Repräsentation, d.h. immer Fehler/Abweichung
- i.d.R. zufällige Auswahl
- i.d.R. Umfang viel kleiner als Grundgesamtheit

# Terminologie

## Grundgesamtheit – Stichprobe



# Terminologie

## Messung

- > Zuweisung von Zahlen oder Bezeichnungen zu (i.d.R.) physikalischen Phänomenen nach bestimmten Regeln
- > Gütekriterien für Messungen
  - **Reliabilität**, Zuverlässigkeit (*reliability*)
    - Konsistenz der Messung bei gleichen Bedingungen
  - **Validität**, Aussagekraft (*validity*)
    - Richtigkeit gemessener Werte in Bezug auf Merkmal
- > Messfehler
  - $X_i = \tau_i + \varepsilon$  (gemessener Wert = echter Wert + zufälliger Fehler)



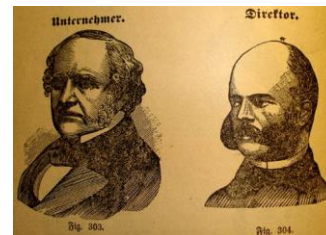
## Terminologie

### Messung – Pseudowissenschaften: fehlende Validität!

*„Als jüngsthin Dr. Gall das hiesige Narrenhaus, Bisetre besuchte, begleitete ihn ein Wahnsinniger, bey dem er keine Kennzeichen des Wahnsinns weder in seinen Reden, noch an seinem Schädel entdecken konnte. Er sagte es ihm. Der Wahnsinnige antwortete: Wundern Sie sich nicht, daß sie an dem Kopfe, der jetzt auf meinen Schultern sitzt, keine Kennzeichen des Wahnsinns antreffen. Es ist ein fremder Kopf, den man mir aufgesetzt hat, nachdem der meinige in der Revolutionszeit durch die Guillotine abgeschlagen worden war.“*

Augsburgische Ordinari Postzeitung, Nro. 175, Freytag, den 22. Jul., Anno 1808, S. 1

Phrenologie (F.J. Gall, 1785-1828),  
Kraniologie/Kraniometrie  
→ Pseudowissenschaft!

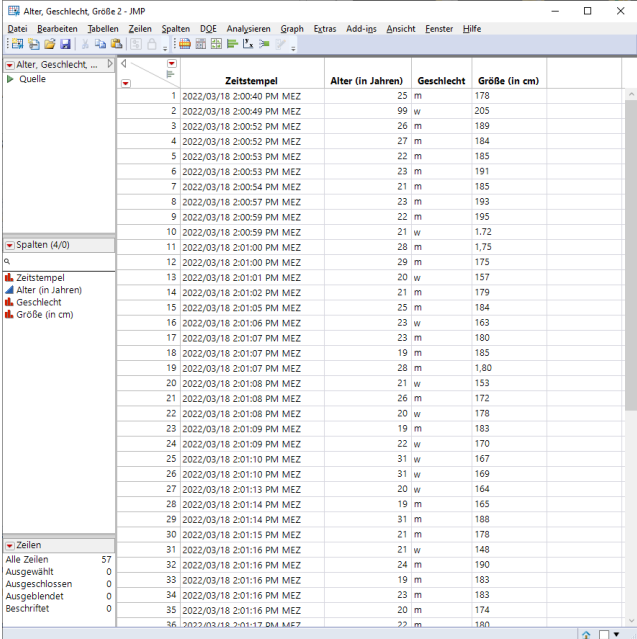


Summi: Phrenologie, gemeinfrei

# Terminologie

## Ergebnis von Messungen: Urliste, Urwerte

- > Beobachtungsreihe, Urdaten, Primärdaten, Rohdaten (raw data)
- > Ergebnis einer Datenerhebung (Messung)
- > Liste von Urwerten von Merkmalsträgern
  - Sortiert
  - Unsortiert
- > Unverarbeitet (inkl. Ausreiser, Fehleingaben)
- > I.d.R. unübersichtlich
  - I.d.R. keine Aussagen möglich



|    | Zeitstempel               | Alter (in Jahren) | Geschlecht | Größe (in cm) |
|----|---------------------------|-------------------|------------|---------------|
| 1  | 2022/03/18 2:00:40 PM MEZ | 25                | m          | 178           |
| 2  | 2022/03/18 2:00:49 PM MEZ | 99                | w          | 205           |
| 3  | 2022/03/18 2:00:52 PM MEZ | 26                | m          | 189           |
| 4  | 2022/03/18 2:00:52 PM MEZ | 27                | m          | 184           |
| 5  | 2022/03/18 2:00:53 PM MEZ | 22                | m          | 185           |
| 6  | 2022/03/18 2:00:53 PM MEZ | 23                | m          | 191           |
| 7  | 2022/03/18 2:00:54 PM MEZ | 21                | m          | 185           |
| 8  | 2022/03/18 2:00:57 PM MEZ | 23                | m          | 193           |
| 9  | 2022/03/18 2:00:59 PM MEZ | 22                | m          | 195           |
| 10 | 2022/03/18 2:00:59 PM MEZ | 21                | w          | 172           |
| 11 | 2022/03/18 2:01:00 PM MEZ | 28                | m          | 175           |
| 12 | 2022/03/18 2:01:00 PM MEZ | 29                | m          | 175           |
| 13 | 2022/03/18 2:01:01 PM MEZ | 20                | w          | 157           |
| 14 | 2022/03/18 2:01:02 PM MEZ | 21                | m          | 179           |
| 15 | 2022/03/18 2:01:05 PM MEZ | 25                | m          | 184           |
| 16 | 2022/03/18 2:01:06 PM MEZ | 23                | w          | 163           |
| 17 | 2022/03/18 2:01:07 PM MEZ | 23                | m          | 180           |
| 18 | 2022/03/18 2:01:07 PM MEZ | 19                | m          | 185           |
| 19 | 2022/03/18 2:01:07 PM MEZ | 28                | m          | 180           |
| 20 | 2022/03/18 2:01:08 PM MEZ | 21                | w          | 153           |
| 21 | 2022/03/18 2:01:08 PM MEZ | 26                | m          | 172           |
| 22 | 2022/03/18 2:01:08 PM MEZ | 20                | w          | 178           |
| 23 | 2022/03/18 2:01:09 PM MEZ | 19                | m          | 183           |
| 24 | 2022/03/18 2:01:09 PM MEZ | 22                | w          | 170           |
| 25 | 2022/03/18 2:01:10 PM MEZ | 31                | w          | 167           |
| 26 | 2022/03/18 2:01:10 PM MEZ | 31                | w          | 169           |
| 27 | 2022/03/18 2:01:13 PM MEZ | 20                | w          | 164           |
| 28 | 2022/03/18 2:01:14 PM MEZ | 19                | m          | 165           |
| 29 | 2022/03/18 2:01:14 PM MEZ | 31                | m          | 188           |
| 30 | 2022/03/18 2:01:15 PM MEZ | 21                | m          | 178           |
| 31 | 2022/03/18 2:01:16 PM MEZ | 21                | w          | 148           |
| 32 | 2022/03/18 2:01:16 PM MEZ | 24                | m          | 190           |
| 33 | 2022/03/18 2:01:16 PM MEZ | 19                | m          | 183           |
| 34 | 2022/03/18 2:01:16 PM MEZ | 23                | m          | 183           |
| 35 | 2022/03/18 2:01:16 PM MEZ | 20                | m          | 174           |
| 36 | 2022/03/18 2:01:17 PM MEZ | 22                | m          | 180           |



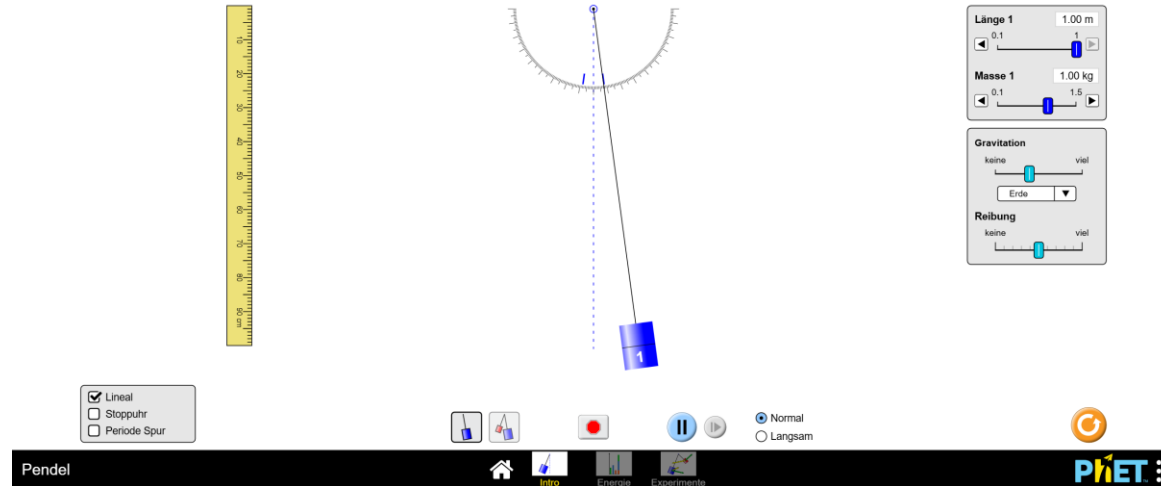
von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
enerscheinung  
Merkmalen

# **Statistik**

Live Experiment

# Experimentelle Bestimmung der Schwerebeschleunigung $g$

## Fadenpendel



<https://forms.gle/sfwpWdTMMejh4uRt6>

[https://phet.colorado.edu/sims/html/pendulum-lab/latest/pendulum-lab\\_all.html?locale=de](https://phet.colorado.edu/sims/html/pendulum-lab/latest/pendulum-lab_all.html?locale=de)

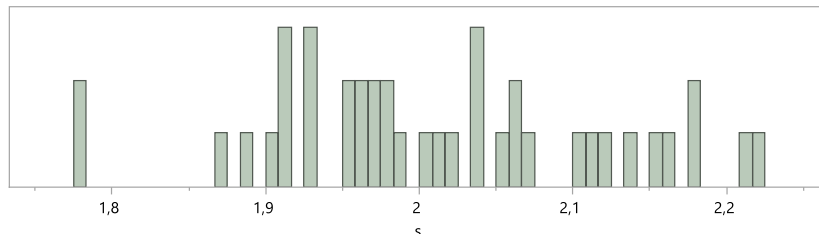
> Schwingungsdauer  $T$  Fadenpendel (Auslenkung  $\leq 8^\circ$ )

$$T = 2\pi \sqrt{\frac{l}{g}} \Rightarrow g = \frac{4\pi^2}{T^2}$$

# Experimentelle Bestimmung der Schwerebeschleunigung $g$

## Fadenpendel

- > Verteilung der experimentell bestimmten Pendelperiodendauer

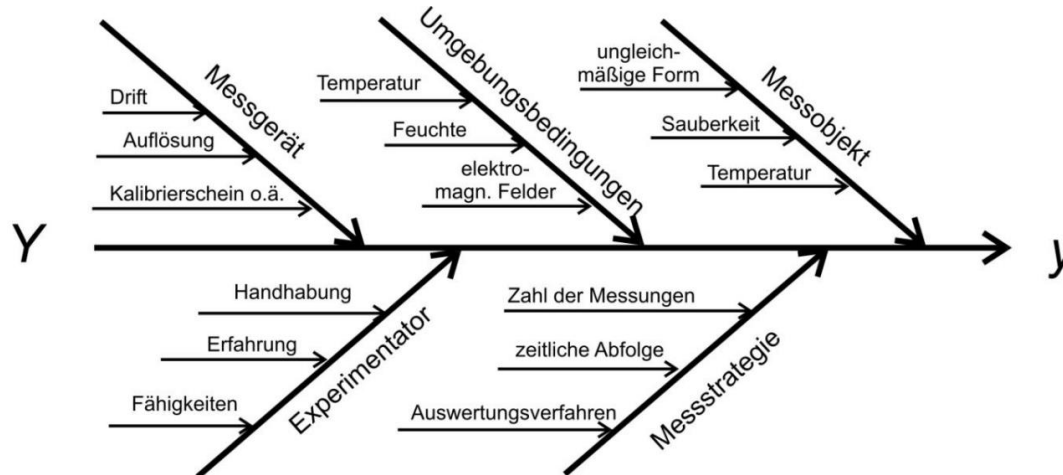


- > Jede Messung physikalischen Größen ist mit Unsicherheit behaftet (Messfehler, Messabweichung, Messunsicherheit)
  - Streuung
    - $X_i = \tau_i + \varepsilon$
- > Fehler  $\varepsilon$  besteht aus
  - Systematischen Fehler
  - Zufällige (statistische) Fehler
- > Messfehler bereits von Carl Friedrich Gauß beschrieben (Fehlerfortpflanzung)

# Experimentelle Bestimmung der Schwerebeschleunigung $g$

## Fadenpendel

> Ichikawa-Diagramm der Quellen für Messunsicherheit

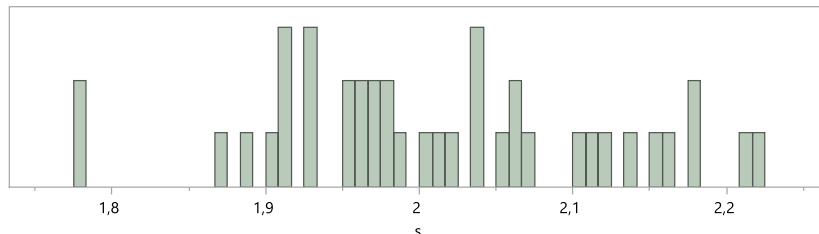


[https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung\\_8/8.4\\_mathematische\\_modellierung/277\\_PTB\\_SEMINAR/VORTRAEGE/11\\_Mieke\\_-\\_Berechnung\\_der\\_Messunsicherheit\\_nach\\_GUM\\_Kurzfassung\\_in\\_20.pdf](https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_8/8.4_mathematische_modellierung/277_PTB_SEMINAR/VORTRAEGE/11_Mieke_-_Berechnung_der_Messunsicherheit_nach_GUM_Kurzfassung_in_20.pdf)

# Experimentelle Bestimmung der Schwerebeschleunigung $g$

## Fadenpendel

- > Verteilung der experimentell bestimmten Pendelperiodendauer



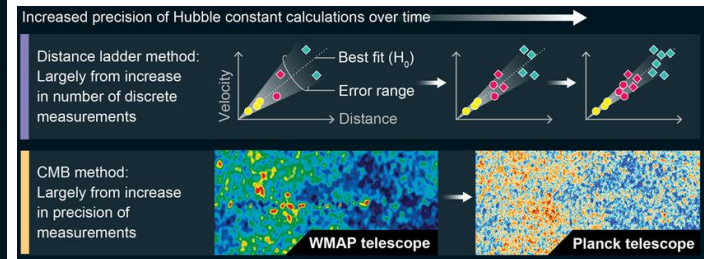
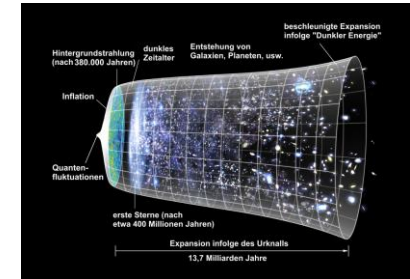
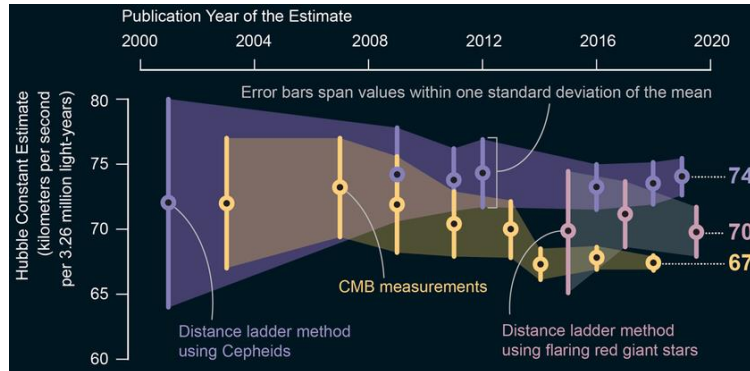
- > Kann der „wahre“ Wert gemessen werden?
  - Karl Pearson (\* 1857, † 1936, Statistiker): Nein
    - Angabe nur als Verteilung möglich
- > Arithmetischer Mittelwert bester Schätzer für „wahren“ Wert
  - Messwiederholungen
- > Zusätzlich: Angabe Messunsicherheit
  - z. B. mittlere Abweichung / Standardabweichung
- > Vgl. ISO/BIPM-Leitfaden „*Guide to the Expression of Uncertainty in Measurement*“ (GUM)



# Die Hubble-Konstante – Messunsicherheit oder neue Physik?

## Widersprüchliche Ergebnisse

- > Hubble-Konstante – Fundamentalste Größe der Kosmologie
  - Beschreibt Rate der Expansion des Universums
  - Messung über systematische Erfassung von Entfernung und scheinbarer Geschwindigkeit von astronomischen Objekten
- > Widersprüchliche Messergebnisse (und Fehlerbalken!)



<https://www.deutschlandfunk.de/hubble-konstante-was-stimmt-nicht-mit-der-expansion-des-100.html>

<https://www.scientificamerican.com/article/how-a-dispute-over-a-single-number-became-a-cosmological-crisis/>





von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
senerscheinung  
Merkmale au

# **Statistik**

Take Home

## Take Home

- Arbeiten mit JMP
  - Benutzeroberfläche
  - Importieren von Daten
  - Bereinigen / Neu kodieren
  - Merkmalseigenschaften
- Terminologie
  - Skalenniveaus
    - Nominal
    - Ordinal
    - Kardinal
  - Grundgesamtheit / Stichprobe
  - Messungen
- <https://community.jmp.com/t5/Getting-Started-with-JMP/Getting-Started-with-JMP-On-Demand-Course/ta-p/621172>



von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
nerscheinungen  
Merkmale

# **Statistik**

## Deskriptive Statistik

# Einführung

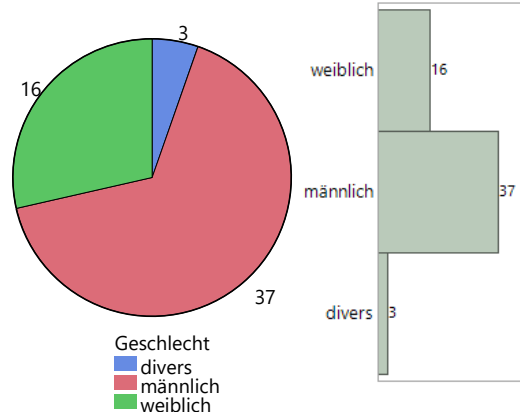
- **Beschreibende Statistik**
- Ziel: Übersichtliche Darstellung von empirischen Daten
  - Tabellen
  - Diagramme / Grafiken
  - Kennzahlen (Statistik)
    - Lageparameter
    - Streuungsparameter
    - Gestaltparameter
- Vergleich zu induktiver Statistik
  - Keine Aussagen über Grundgesamtheit
  - Keine Überprüfung von Hypothesen

|                  | Aggregation der Daten | Übersichtlichkeit | Informationsgehalt |
|------------------|-----------------------|-------------------|--------------------|
| <b>Tabelle</b>   | Niedrig               | Niedrig           | Hoch               |
| <b>Diagramm</b>  | Mittel                | Mittel            | Mittel             |
| <b>Parameter</b> | Hoch                  | Hoch              | Niedrig            |

# Häufigkeitsverteilung

- Funktion, die jedem vorkommenden (und möglichen) Wert dessen Häufigkeit zuordnet
  - Absolute Häufigkeit
  - Relative Häufigkeit
- Darstellung als
  - Tabelle
  - Diagramm
- Funktionsgleichung

| Klasse   | Absolute Häufigkeit | Relative Häufigkeit     |
|----------|---------------------|-------------------------|
| divers   | 3                   | $\frac{3}{56}$ 0,05357  |
| männlich | 37                  | $\frac{37}{56}$ 0,66071 |
| weiblich | 16                  | $\frac{16}{56}$ 0,28571 |
| Summe    | 56                  | 1                       |

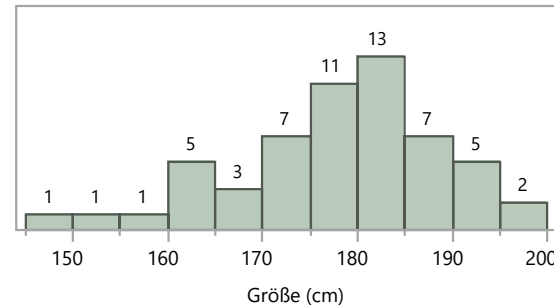


$$h(Klasse) = \begin{cases} Klasse = "divers" \Rightarrow 3 \\ Klasse = "männlich" \Rightarrow 37 \\ Klasse = "weiblich" \Rightarrow 16 \end{cases}$$

$$f(Klasse) = \begin{cases} Klasse = "divers" \Rightarrow \frac{3}{56} \\ Klasse = "männlich" \Rightarrow \frac{37}{56} \\ Klasse = "weiblich" \Rightarrow \frac{16}{56} \end{cases}$$

## Häufigkeitsverteilung kardinalskaliertter Daten

- Klassierung (Klassenbildung, *binning*)
  - Zusammenfassung „ähnlicher“ Merkmalswerte in disjunkte Klassen
- Hilfreich
  - für Übersichtlichkeit
  - bei große Anzahl verschiedener Werte
  - zur Untersuchung stetiger Variablen mit Methoden für diskrete Variablen
- Nachteil: Informationsverlust
- Tabelliert
- **Histogramm**

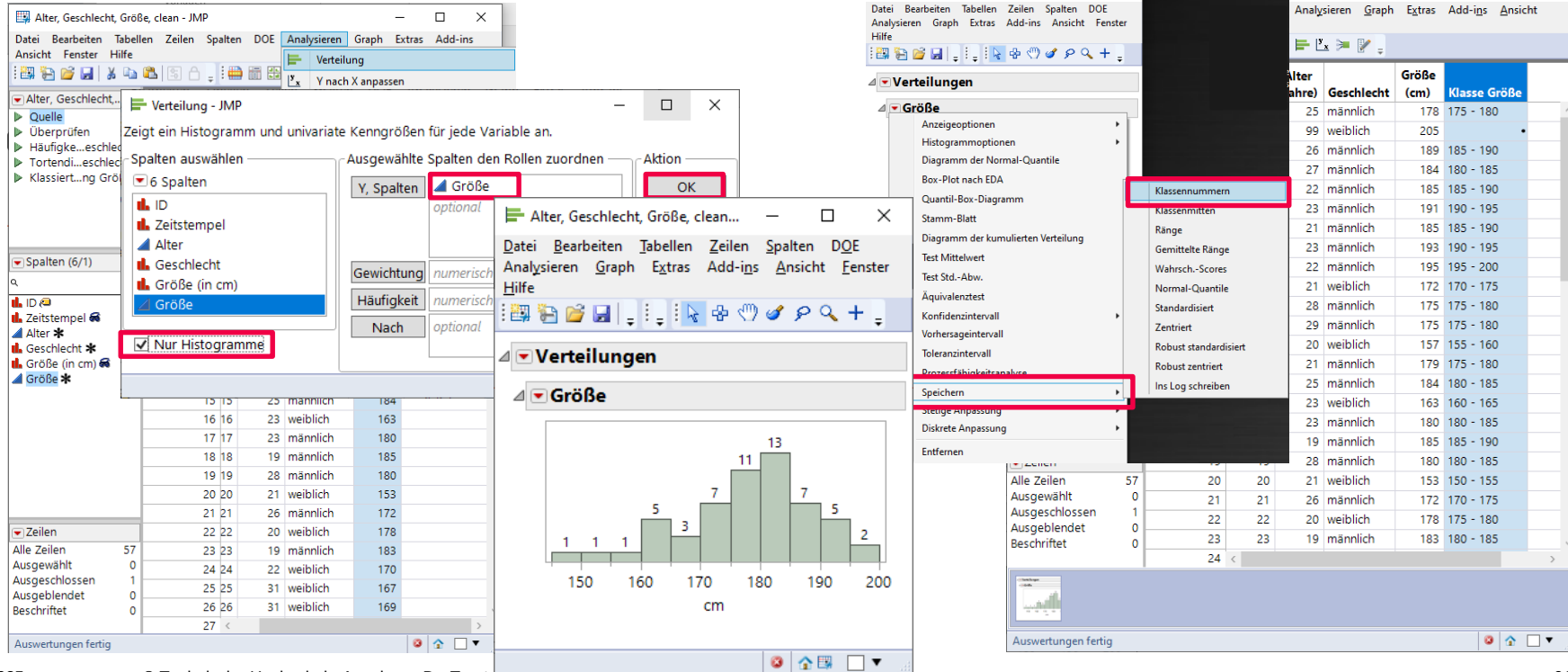


## Häufigkeitsverteilung kardinalskaliertter Daten

- Klassierung
  - Klassengrenzen (obere, untere):  $x_j^u \leq x < x_j^o$  bzw.  $x_j^u < x \leq x_j^o; j = 1, \dots, k$
  - Klassenbreite:  $\Delta x_j = x_j^o - x_j^u$ , i.d.R. konstant:  $\Delta x = x^o - x^u$
  - Anzahl Klassen
    - Sturges (i.d.R. zu klein bei  $n > 200$ ):  $k = 1 + \log_2(n) \approx 1 + 3,3 * \log_{10}(n)$
    - Jenks-Caspall
    - ...
  - Klassenmitte:  $x_j = \frac{x_j^o - x_j^u}{2}$
- Absolute Häufigkeit  $h(x_j)$ : Anzahl Merkmale je Klasse
- Absolute Häufigkeitsdichte:  $\hat{h}(x_j) = \frac{h(x_j)}{x_j^o - x_j^u}$

# Häufigkeitsverteilung kardinalskalierteter Daten

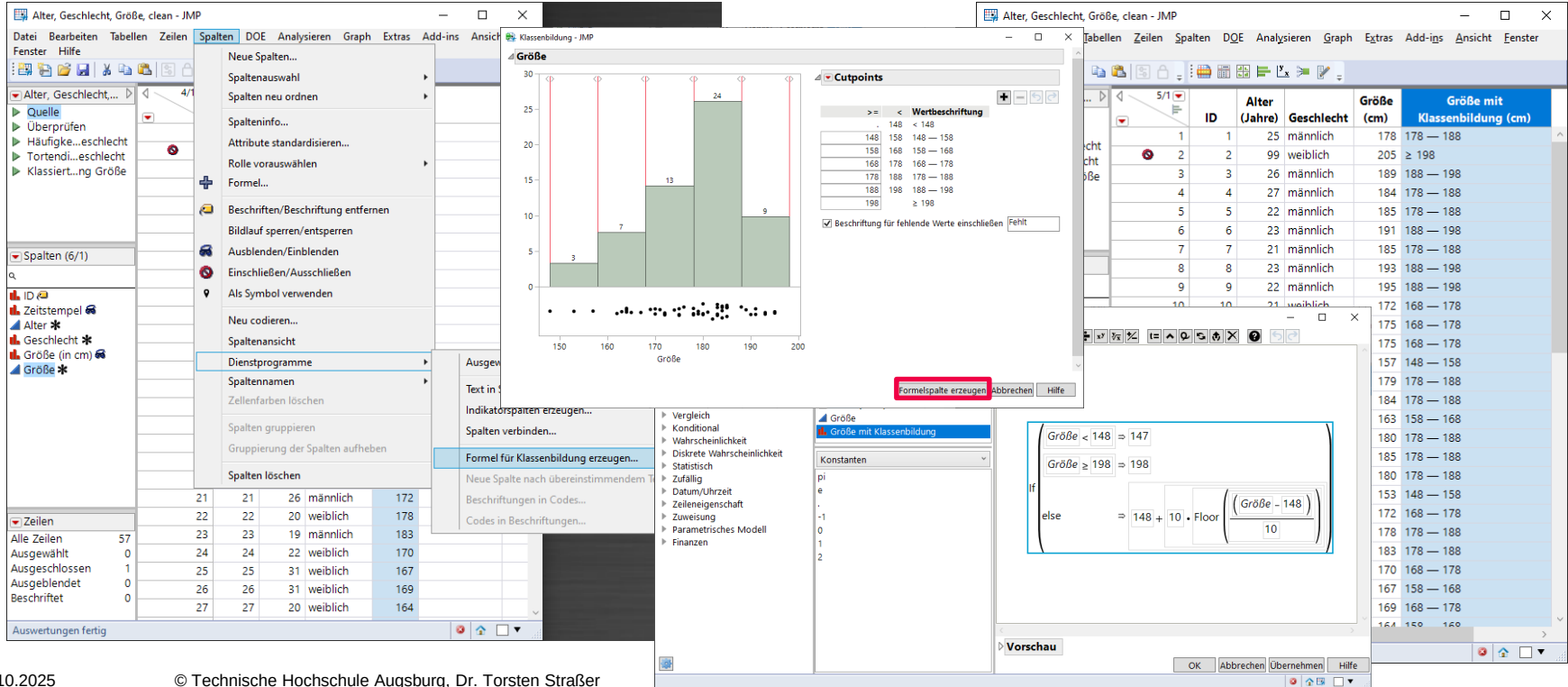
## Daten klassieren – Verteilung





# Häufigkeitsverteilung kardinalskaliertter Daten

## Daten klassieren – Dienstprogramme – Formel



# Häufigkeitsverteilung kardinalskalierteter Daten

## Darstellung Histogramm (GraphBuilder)

Alter, Geschlecht, Größe, clean - JMP

Datei Bearbeiten Tabellen Zeilen Spalten DOE Analysieren Graph Extras Add-ins Ansicht Fenster Hilfe

Quelle

- Überprüfen
- Häufigkeitsverteilung
- Tortendiagramm
- Klassierte Größe
- Histogramm

Spalten (7/1)

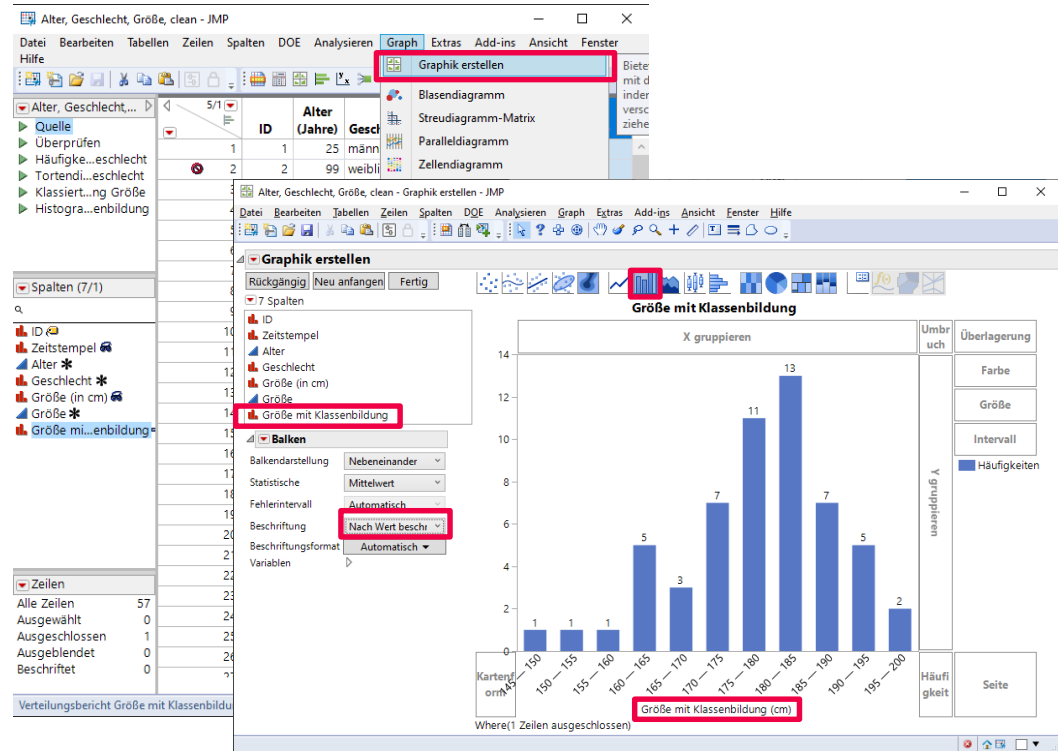
- ID
- Zeitstempel
- Alter \*
- Geschlecht \*
- Größe (in cm)
- Größe \*
- Größe mit Klassenbildung \*

Zeilen

- Alle Zeilen 57
- Ausgewählt 0
- Ausgeschlossen 1
- Ausgeblendet 0
- Beschriftet 0

| ID | Alter (Jahre) | Geschlecht | Größe (cm) | Größe mit Klassenbildung (cm) |
|----|---------------|------------|------------|-------------------------------|
| 1  | 1             | männlich   | 178        | 175 — 180                     |
| 2  | 2             | weiblich   | 205        | ≥ 200                         |
| 3  | 3             | männlich   | 189        | 185 — 190                     |
| 4  | 4             | männlich   | 184        | 180 — 185                     |
| 5  | 5             | männlich   | 185        | 185 — 190                     |
| 6  | 6             | männlich   | 191        | 190 — 195                     |
| 7  | 7             | männlich   | 185        | 185 — 190                     |
| 8  | 8             | männlich   | 193        | 190 — 195                     |
| 9  | 9             | männlich   | 195        | 195 — 200                     |
| 10 | 10            | weiblich   | 172        | 170 — 175                     |
| 11 | 11            | männlich   | 175        | 175 — 180                     |
| 12 | 12            | männlich   | 175        | 175 — 180                     |
| 13 | 13            | weiblich   | 157        | 155 — 160                     |
| 14 | 14            | männlich   | 179        | 175 — 180                     |
| 15 | 15            | männlich   | 184        | 180 — 185                     |
| 16 | 16            | weiblich   | 163        | 160 — 165                     |
| 17 | 17            | männlich   | 180        | 180 — 185                     |
| 18 | 18            | männlich   | 185        | 185 — 190                     |
| 19 | 19            | männlich   | 180        | 180 — 185                     |
| 20 | 20            | weiblich   | 153        | 150 — 155                     |
| 21 | 21            | männlich   | 172        | 170 — 175                     |
| 22 | 22            | weiblich   | 178        | 175 — 180                     |
| 23 | 23            | männlich   | 183        | 180 — 185                     |
| 24 | 24            | weiblich   | 170        | 170 — 175                     |
| 25 | 25            | weiblich   | 167        | 165 — 170                     |
| 26 | 26            | weiblich   | 169        | 165 — 170                     |
| 27 | 27            | weiblich   | 164        | 160 — 165                     |

Verteilungsbericht Größe mit Klassenbildung



# Häufigkeitsverteilung kardinalskaliertter Daten

## Darstellung Tabelle

**Alter, Geschlecht, Größe, clean - JMP**

| ID | Alter (Jahre) | Geschlecht | Größe (cm) | Größe mit Klassenbildung (cm) |
|----|---------------|------------|------------|-------------------------------|
| 1  | 1             | 25         | männlich   | 178 175 — 180                 |
| 2  | 2             | 99         | weiblich   | 205 ≥ 200                     |
| 3  | 3             | 26         | männlich   | 189 185 — 190                 |
| 4  | 4             | 27         | männlich   | 184 180 — 185                 |
| 5  | 5             | 22         | männlich   | 185 185 — 190                 |
| 6  | 6             | 23         | männlich   | 191 190 — 195                 |
| 7  | 7             | 21         | männlich   | 185 185 — 190                 |
| 8  | 8             | 23         | männlich   | 193 190 — 195                 |
| 9  | 9             | 22         | männlich   | 195 195 — 200                 |
| 10 | 10            | 21         | weiblich   | 172 170 — 175                 |
| 11 | 11            | 28         | männlich   | 175 175 — 180                 |
| 12 | 12            | 29         | männlich   | 175 175 — 180                 |
| 13 | 13            | 20         | weiblich   | 157 155 — 160                 |
| 14 | 14            | 21         | männlich   | 179 175 — 180                 |
| 15 | 15            | 25         | männlich   | 184 180 — 185                 |
| 16 | 16            | 23         | weiblich   | 163 160 — 165                 |
| 17 | 17            | 23         | männlich   | 180 180 — 185                 |
| 18 | 18            | 19         | männlich   | 185 185 — 190                 |
| 19 | 19            | 28         | männlich   | 180 180 — 185                 |
| 20 | 20            | 21         | weiblich   | 153 150 — 155                 |
| 21 | 21            | 26         | männlich   | 172 170 — 175                 |
| 22 | 22            | 20         | weiblich   | 178 175 — 180                 |
| 23 | 23            | 19         | männlich   | 183 180 — 185                 |
| 24 | 24            | 22         | weiblich   | 170 170 — 175                 |
| 25 | 25            | 31         | weiblich   | 167 165 — 170                 |
| 26 | 26            | 31         | weiblich   | 169 165 — 170                 |
| 27 | 27            | 20         | weiblich   | 164 160 — 165                 |

**Zusammenfassung - JMP**

Erstellt eine neue Datentabelle mit statistischen Kenngrößen.

Spalten auswählen  
☒ 7 Spalten  
☒ ID  
☒ Zeitstempel  
☒ Alter  
☒ Geschlecht  
☒ Größe (in cm)  
☒ Größe m...nbildung

☐ Kenngrößen für Randverteilungen berechnen  
 Geben Sie einen Wert in % für die Quantilberechnung ein  
 25  
 Benennungsformat für die Ergebnisspalte  
 Kenngröße(Spalte)

Name der Ausgabetable:

☒ Mit ursprünglicher Datentabelle verknüpfen  
☐ Aufforderung zum Speichern beim Schließen von Zusammenfassungstabellen

☐ Dialogfeld geöffnet lassen  
☐ Skript in Quelltable speichern

**Alter, Geschlecht, Größe, clean Nach (Größe mit Klassenbildung) - JMP**

| Größe mit Klassenbildung | Anzahl Zeilen | N(Größe mit Klassenbildung) |
|--------------------------|---------------|-----------------------------|
| 1 145 — 150              | 1             | 1                           |
| 2 150 — 155              | 1             | 1                           |
| 3 155 — 160              | 1             | 1                           |
| 4 160 — 165              | 5             | 5                           |
| 5 165 — 170              | 3             | 3                           |
| 6 170 — 175              | 7             | 7                           |
| 7 175 — 180              | 11            | 11                          |
| 8 180 — 185              | 13            | 13                          |
| 9 185 — 190              | 7             | 7                           |
| 0 190 — 195              | 5             | 5                           |
| 1 195 — 200              | 2             | 2                           |
| 2 ≥ 200                  | 0             | 0                           |

# Häufigkeitsverteilung kardinalskaliertter Daten

## Darstellung Tabelle



The screenshot displays the JMP software interface with two windows open, showing frequency distribution tables for cardinal data.

**Left Window: Alter, Geschlecht, Größe, clean Nach (Größe mit Klassenbildung) - JMP**

| Größe mit Klassenbildung | Anzahl Zeilen | N(Größe mit Klassenbildung) |
|--------------------------|---------------|-----------------------------|
| 1 145 — 150              | 1             | 1                           |
| 2 150 — 155              | 1             | 1                           |

**Right Window: Alter, Geschlecht, Größe, clean Nach (Größe mit Klassenbildung) - JMP**

| Größe mit Klassenbildung | Anzahl Zeilen | Absolute Häufigkeit | Relative Häufigkeit |
|--------------------------|---------------|---------------------|---------------------|
| 1 145 — 150              | 1             | 1                   | 0,0179              |
| 2 150 — 155              | 1             | 1                   | 0,0179              |
| 3 155 — 160              | 1             | 1                   | 0,0179              |
| 4 160 — 165              | 5             | 5                   | 0,0893              |
| 5 165 — 170              | 3             | 3                   | 0,0536              |
| 6 170 — 175              | 7             | 7                   | 0,1250              |
| 7 175 — 180              | 11            | 11                  | 0,1964              |
| 8 180 — 185              | 13            | 13                  | 0,2321              |

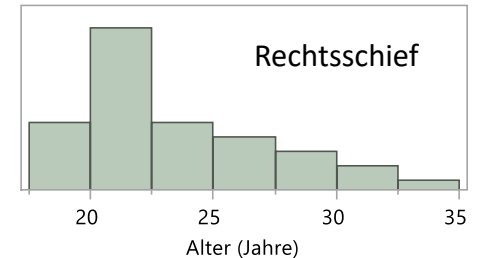
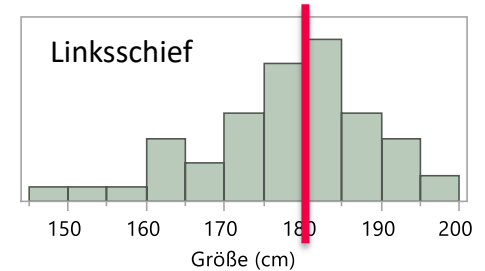
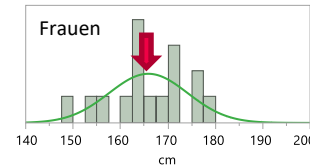
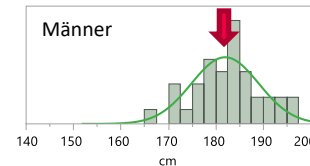
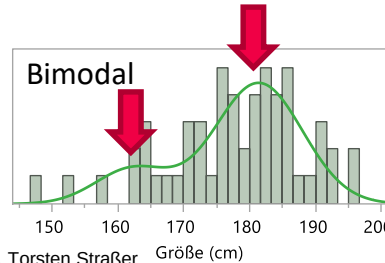
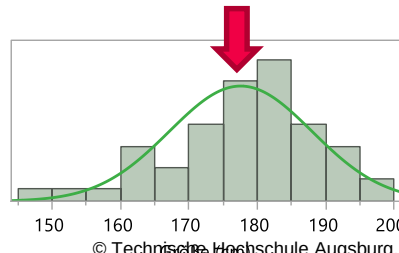
The 'Formel...' menu option is highlighted in the left window. The right window shows a formula editor with the following formula:

$$\text{Col Sum}(\text{Absolute Häufigkeit})$$

# Deskriptive Statistik

## Die Form der Häufigkeitsverteilung (*The Shape of Data*)

- Eigenschaften von Verteilungen
  - Symmetrie
    - Symmetrisch
    - Asymmetrisch
  - Schiefe (*skewness*)
    - Stärke der Asymmetrie
    - Linksschief, negative Schiefe
    - Rechtsschief, positive Schiefe
  - Unimodal / multimodal



# Beschreibung von Häufigkeitsverteilungen

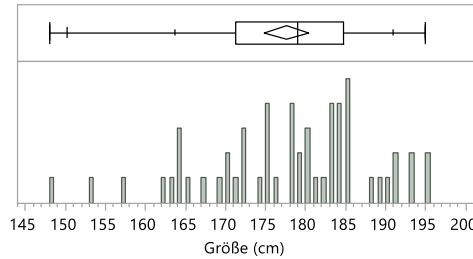
- Beschreibende Statistik
- Ziel: Übersichtliche Darstellung von empirischen Daten
  - Tabellen ✓
  - Diagramme / Grafiken ✓
  - Kennzahlen (Parameter) ⇐
    - Lageparameter
    - Streuungsparameter
    - Gestaltparameter

|                  | Aggregation der Daten | Übersichtlichkeit | Informationsgehalt |
|------------------|-----------------------|-------------------|--------------------|
| <b>Tabelle</b>   | Niedrig               | Niedrig           | Hoch               |
| <b>Diagramm</b>  | Mittel                | Mittel            | Mittel             |
| <b>Parameter</b> | Hoch                  | Hoch              | Niedrig            |

## Lageparameter – Modus, Median, Mittelwert

- Beschreibung zentrale Tendenz (Zentrum) einer (Häufigkeits-)Verteilung
  - Wo befindet sich ein Großteil der Merkmalswerte einer Verteilung?
- **Modus, Modalwert**  $x_{Mod}$ 
  - Häufigster Wert
  - Anwendung auch bei nominal skalierten Daten
- **Median**  $x_{Med}$ 
  - „Mittlerer“ Wert: Gleichviele Werte ober- und unterhalb
  - Median = 50%-Quantile
  - Urliste aufsteigend sortieren (nach Rang geordnet):  $x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}$
  - Mindestens ordinal skaliert
  - $$x_{Med} = \begin{cases} x_{[\frac{n+1}{2}]} & \text{falls } n \text{ ungerade} \\ \frac{x_{[\frac{n}{2}]} + x_{[\frac{n}{2}+1]}}{2} & \text{falls } n \text{ gerade} \end{cases}$$
- **Arithmetischer Mittelwert** (Durchschnittswert)  $\bar{x}$ 
  - $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
  - Sinnvoll nur bei kardinalskalierten Werten
  - Meist nur sinnvoll bei symmetrischer Verteilung (Schiefe  $\approx 0$ )
  - Nicht robust gegen Ausreißer

# Lageparameter – Modus, Median, Mittelwert

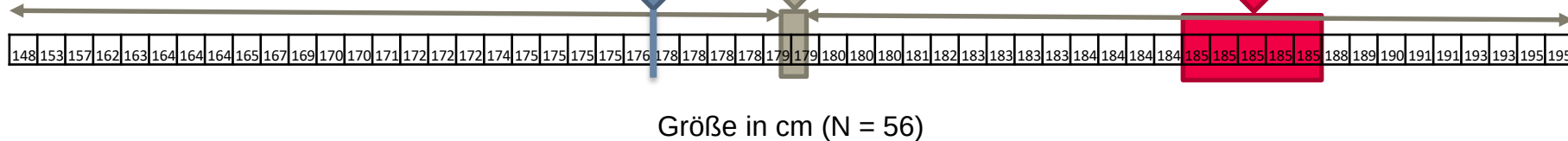


|            |        |
|------------|--------|
| Modus      | 185    |
| Median     | 179    |
| Mittelwert | 177,57 |
| N          | 56     |

$$\text{Median} = \frac{x_{[\frac{56}{2}]} + x_{[\frac{56}{2}+1]}}{2} = \frac{x_{[28]} + x_{[29]}}{2} = \frac{179 + 179}{2} = 179$$

$$\text{Mittelwert} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 177,57$$

$$\text{Modus} = \max h(x_j) = h(185) = 5$$

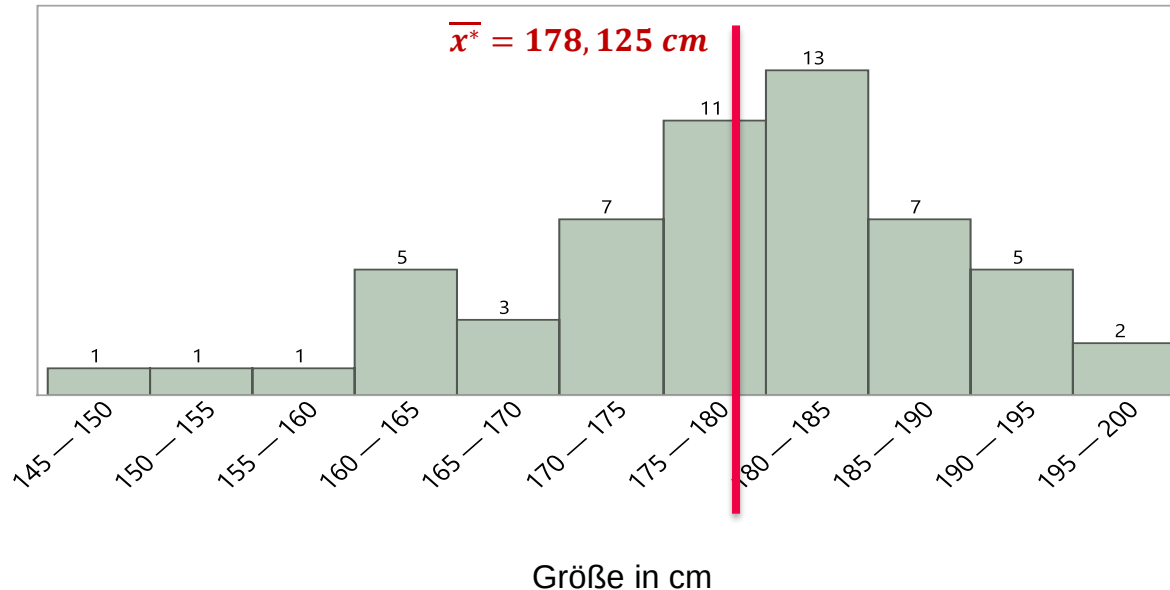




## Lageparameter – Geschätztes arithmetisches Mittel bei klassierten Daten

$$\bar{x}^* = \frac{1}{n} \sum (\text{Klassenmitte} * \text{Klassenhäufigkeit})$$

$$= \frac{1}{56} * (147,5 * 1 + 152,5 * 1 + 157,5 * 1 + 162,5 * 5 + 167,5 * 3 + 172,5 * 7 + 177,5 * 11 + 182,5 * 13 + 187,5 * 7 + 192,5 * 5 + 197,5 * 2)$$
$$= 178,125$$



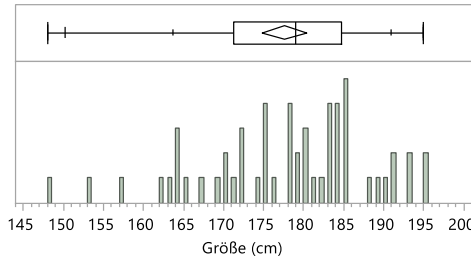
# Streuungsparameter

- Beschreiben Streubreite einer (Häufigkeits-)Verteilung um Lageparameter
- Voraussetzung: Kardinalskalierte Daten
- **Spannweite** (*range*)
  - $SP = \max(x_i) - \min(x_i)$
  - Nicht robust gegen Ausreißer

## Streuungsparameter um arithmetischen Mittelwert

- Mittlere absolute Abweichung (*mean absolute deviation*)
  - $MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$
- Summe der Abweichungsquadrate (*sum of squared differences*)
  - $SQ_x = \sum_{i=1}^n (x_i - \bar{x})^2$
- **Empirische Varianz (*variance*)**
  - $\tilde{s}_x^2 = \frac{1}{n} SQ_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
- **Empirische Standardabweichung (*standard deviation*)**
  - $\tilde{s} = \sqrt{\tilde{s}_x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
- Empirischer Variationskoeffizient (*coefficient of variation*)
  - $v = \frac{\tilde{s}}{\bar{x}}$

# Lageparameter – Modus, Median, Mittelwert

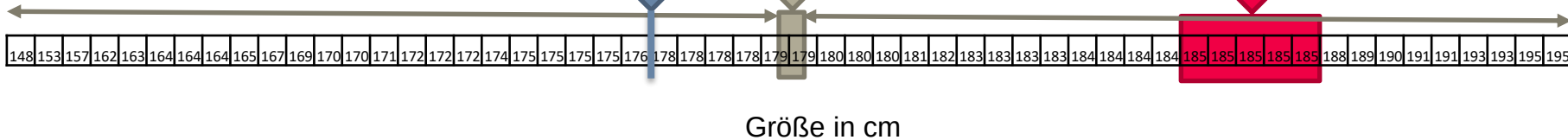


|            |        |
|------------|--------|
| Modus      | 185    |
| Median     | 179    |
| Mittelwert | 177,57 |
| N          | 56     |

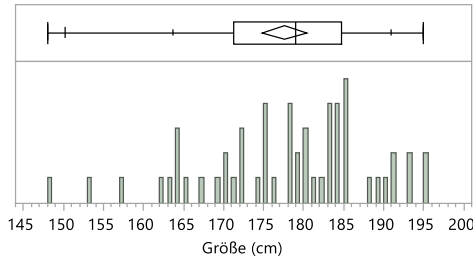
$$\text{Median} \quad \frac{x_{\lfloor \frac{56}{2} \rfloor} + x_{\lfloor \frac{56}{2} \rfloor + 1}}{2} = \frac{x_{[28]} + x_{[29]}}{2} = \frac{179 + 179}{2} = 179$$

$$\text{Mittelwert} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 177,57$$

$$\text{Modus} \quad \max h(x_j) = h(185) = 5$$



# Median und Quartile

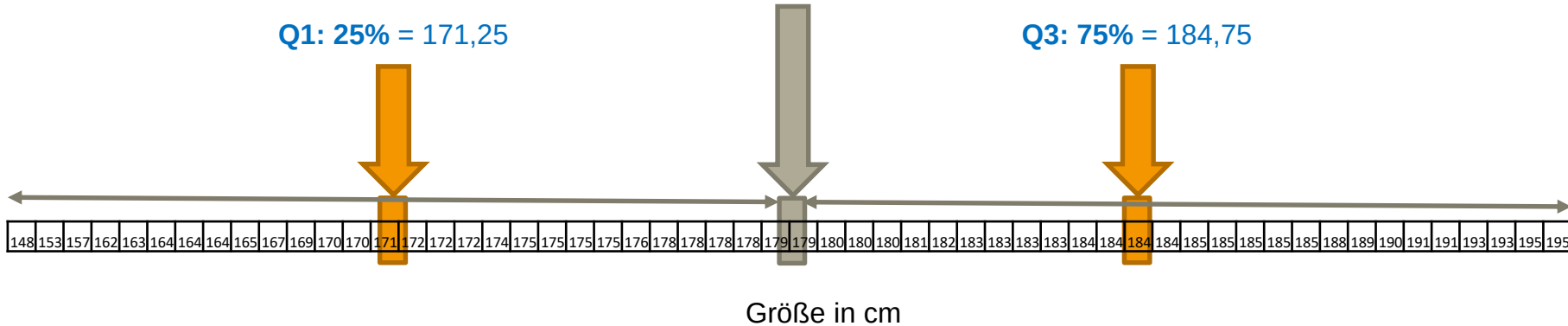


Median (Q2): 50%

$$\frac{x_{[\frac{56}{2}]} + x_{[\frac{56}{2}+1]}}{2} = \frac{x_{[28]} + x_{[29]}}{2} = \frac{179 + 179}{2} = 179$$

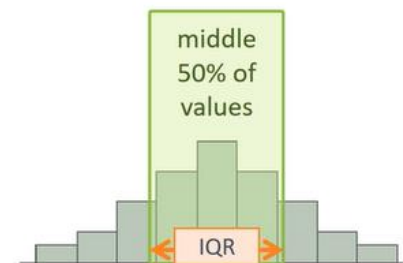
Q1: 25% = 171,25

Q3: 75% = 184,75

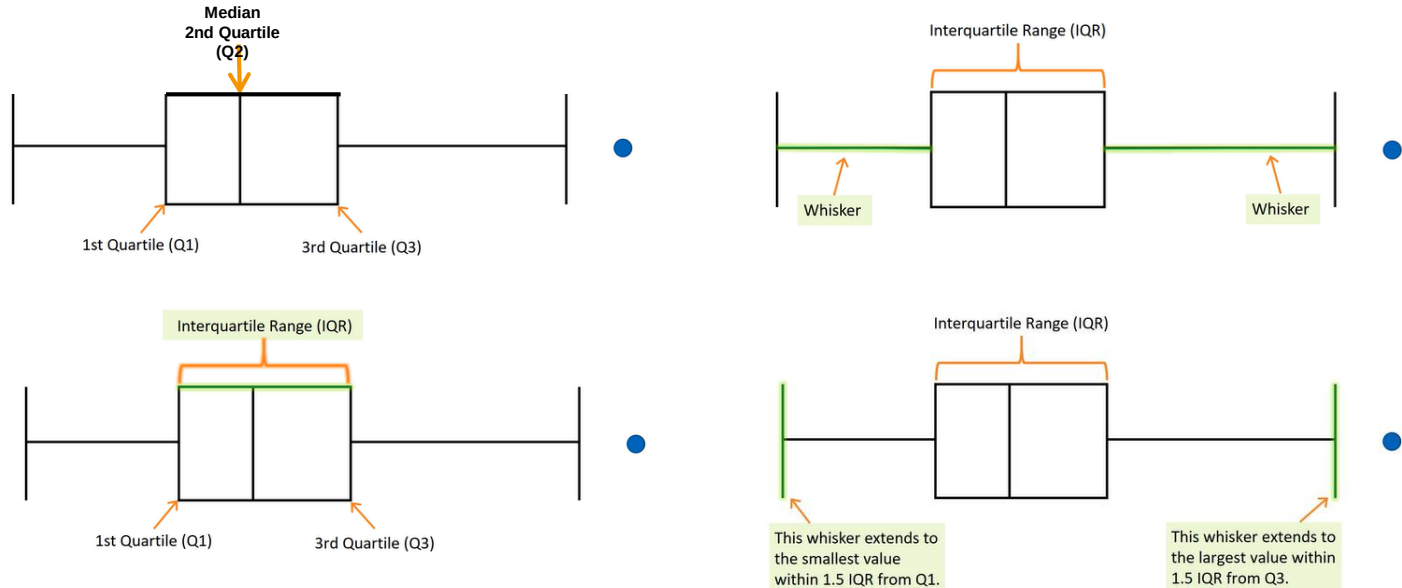


## Streuungsparameter um Median

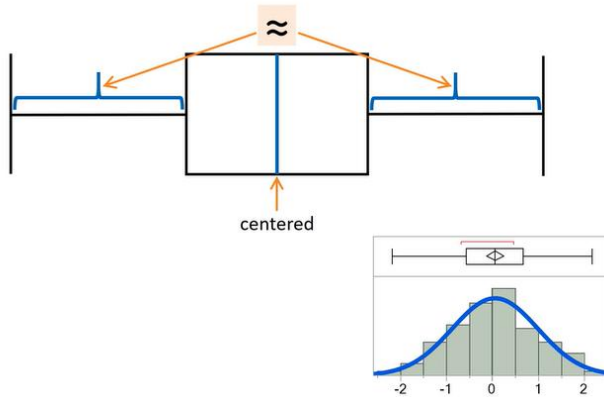
- **p-Quantil**
  - Teilt Werte in Anteil  $< p$  und Anteil  $> p$
  - **Median**  $x_{Med}, Q_{0,5}$ :  $p = 0,5$
  - Oberes **Quartil** ( $p = 0,75$ ), unteres **Quartil** ( $p = 0,25$ )
  - Terzile ( $1/3, 2/3$ ), Quintile ( $0,2; 0,4; 0,6; 0,8$ ), Dezile ( $0,1$  Schritte)
  - **Perzentile** ( $0,01$  Schritte, z. B. 5% oder 95% Perzentil)
- **Interquartilsabstand** (*interquartile range*)
  - enthält 50% aller Werte
  - $IQR = Q_{0,75} - Q_{0,25}$
- Mittlere absolute Abweichung vom Median (*mean absolute deviation from the median*)
  - $MD = \frac{1}{n} \sum_{i=1}^n |x_i - x_{Med}|$



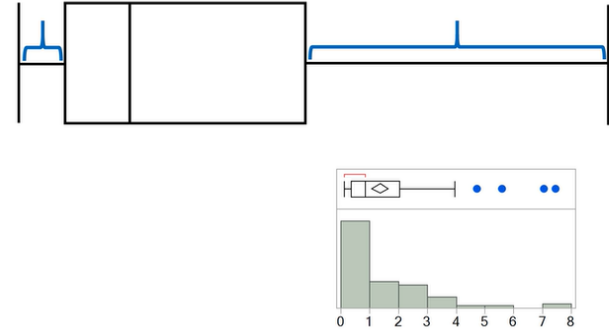
## Graphische Darstellung einer Verteilung: Box-Plots



## Graphische Darstellung einer Verteilung: Box-Plots



Verteilung ist symmetrisch



Verteilung ist asymmetrisch



# Streuungsparameter um Stichprobenmittelwert

## Stichprobenvarianz und Stichprobenstandardabweichung



- Stichprobenvarianz

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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Anzahl Freiheitsgrade  
 $n - 1$

- Stichprobenstandardabweichung

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

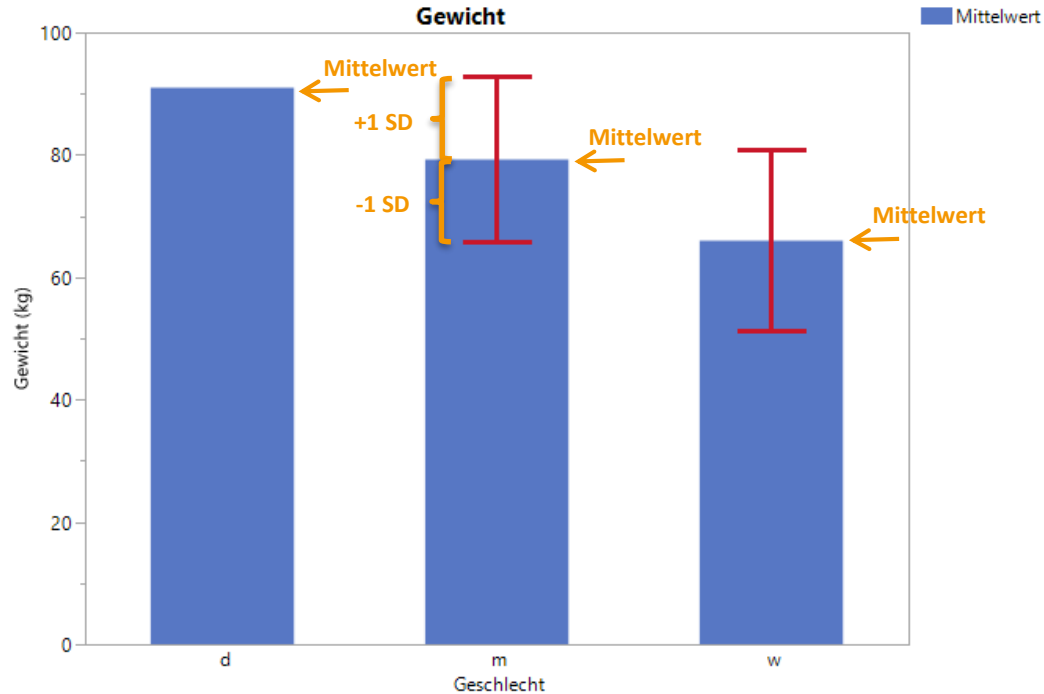
| Größe | Mittelwert | Größe-Mittelwert |
|-------|------------|------------------|
| 148   | 156,6      | -8,6             |
| 153   | 156,6      | -3,6             |
| 157   | 156,6      | 0,4              |
| 162   | 156,6      | 5,4              |
| 163   | 156,6      | 6,4              |
| Summe |            | 0                |

$$(-8,6) + (-3,6) + 0,4 + 5,4 + 6,4 = 0$$

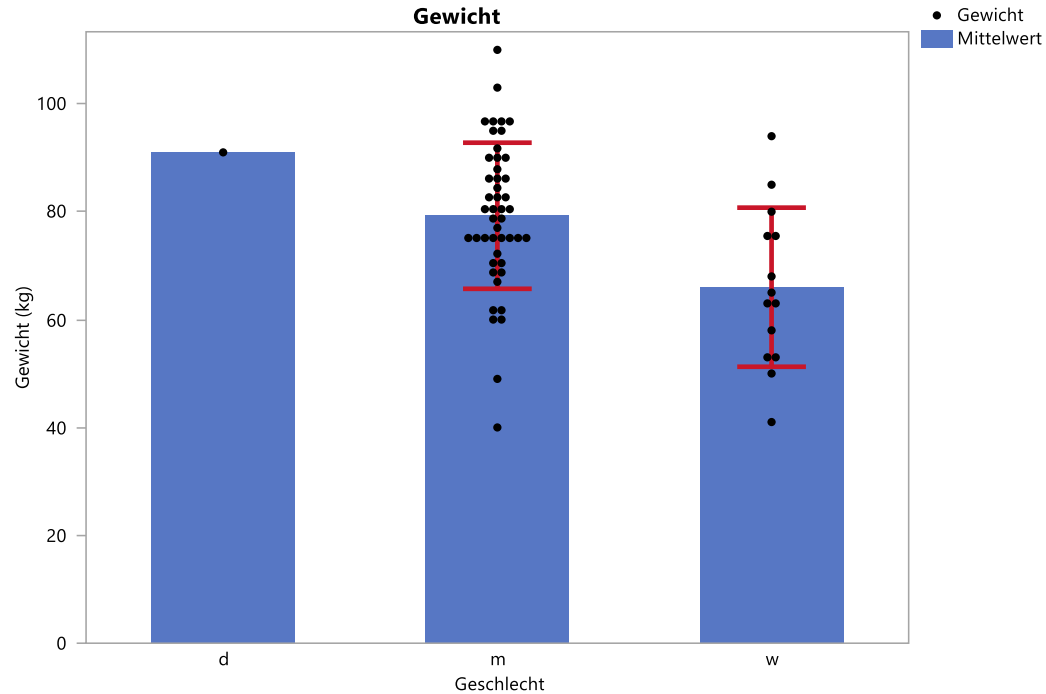
Frei!

Fest!

# Graphische Darstellung einer Verteilung: Balkendiagramm



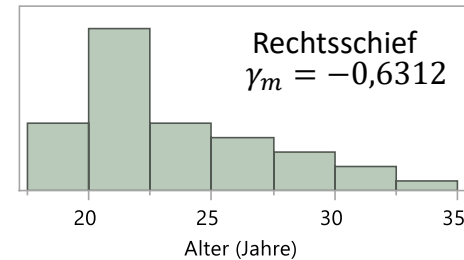
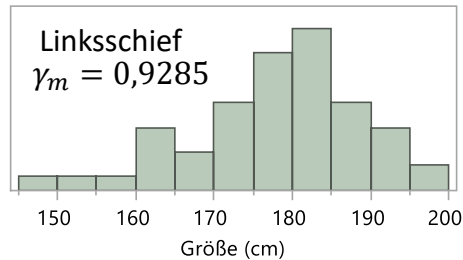
# Graphische Darstellung einer Verteilung: Balkendiagramm



# Gestaltparameter



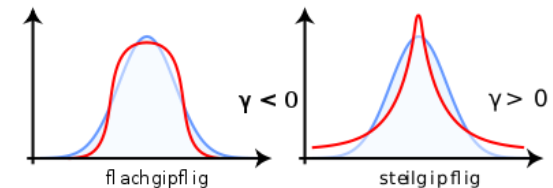
- Schiefe (*skewness*)
  - Maßzahl für Symmetrie einer (Häufigkeits-)Verteilung



- Kurtosis (Wölbung, *kurtosis*)
  - Maßzahl für Steilheit bzw. „Spitzigkeit“ einer (Häufigkeits-)Verteilung

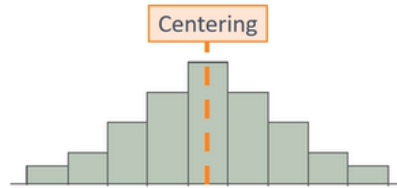
$$w = 0,2595$$

$$w = 0,1746$$



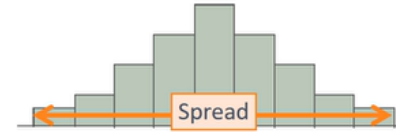
# Beschreibung von Häufigkeitsverteilungen

## Lageparameter



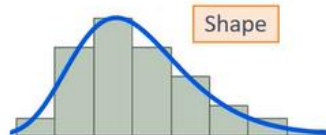
- Modus
- Median
- Mittelwert

## Streuungsparameter

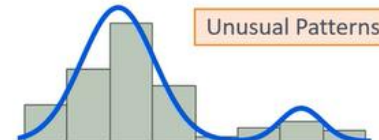


- Spannweite
- Quantilen (Quartile, Perzentile), IQR
- Varianz / Standardabweichung

## Gestaltparameter



- Schiefe
- Kurtosis (Wölbung)



- Unimodal / Multimodal



von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
erscheinungen  
Merkmalen

# **Statistik**

## Häufigkeitsverteilungen

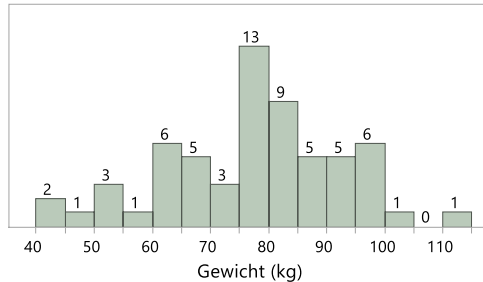
# Häufigkeitsverteilungen

## Absolute und relative Häufigkeit



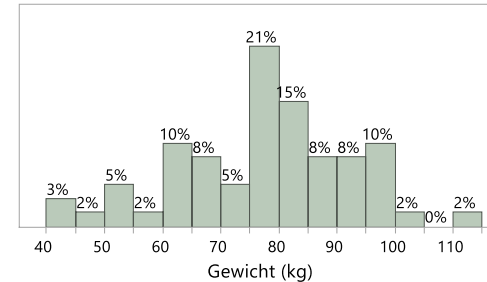
- Absolute Häufigkeit

$$h(x_i) = h_i \quad 1 < i \leq n$$



- Relative Häufigkeit

$$f(x_i) = \frac{h(x_i)}{n} = \frac{h_i}{n} \quad 1 < i \leq n$$



$n$ : Anzahl Klassen (Merkmalsausprägungen)  
 $x_i$ :  $i^{\text{te}}$ -Klasse  
 $h_i$ : Anzahl Merkmalsträger in Klasse  $i$

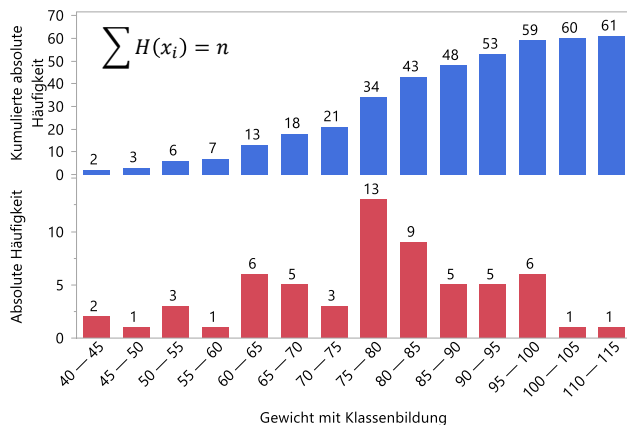
# Häufigkeitsverteilungen

## Kumulierte absolute und relative Häufigkeit



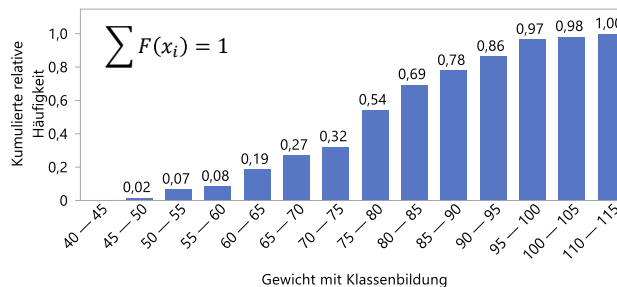
- Kumulierte absolute Häufigkeit

$$H(x_i) = \sum_{k=1}^i h(x_k) \quad 1 < i \leq n$$



- Kumulierte relative Häufigkeit

$$F(x_i) = \sum_{k=1}^i f(x_k) = \sum_{k=1}^i \frac{h(x_k)}{n} \quad 1 < i \leq n$$



$n$ :

Anzahl Klassen (Merkmalsausprägungen)

$x_i$ :

$i^{\text{te}}$ -Klasse

$h_i$ :

Anzahl Merkmalsträger in Klasse  $i$



# Häufigkeitsverteilungen

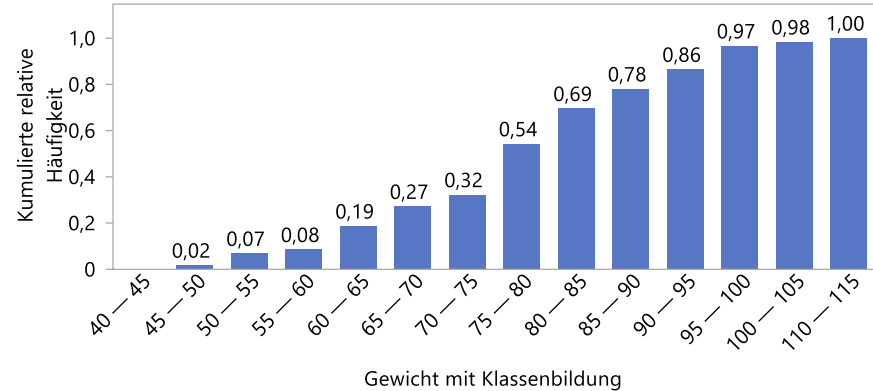
## Kumulierte absolute und relative Häufigkeit



- Anteil  $< 80$  kg?  
 $F(75 - 80) = 0,54 \triangleq 54\%$

- Anteil  $\geq 80$  kg?  
 $1 - F(75 - 80)$   
 $= 1 - 0,54 = 0,46 \triangleq 46\%$

- Anteil  $\geq 50$  und  $< 100$  kg?  
 $F(75 - 80) - F(50 - 55)$   
 $= 0,54 - 0,07 = 0,47 \triangleq 47\%$

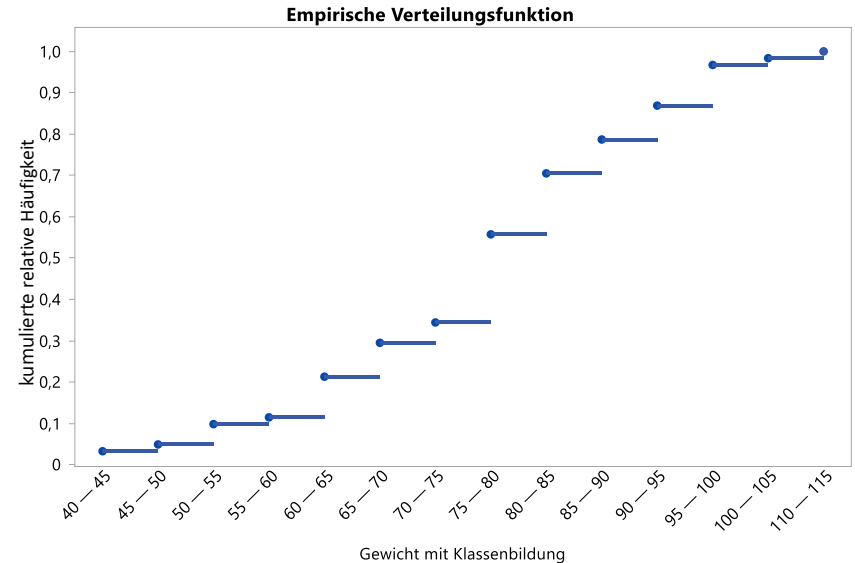


# Häufigkeitsverteilungen

## Empirische Verteilungsfunktion / Verteilungsfunktion der Stichprobe



- Summenhäufigkeitsfunktion
- $$F_n(x) = \frac{\text{Anzahl Werte} \leq x}{n} = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i \leq x\}}$$
- Ordnet jedem X Anteil der Stichprobenwerte  $\leq X$  zu
- Auch bei Ordinalskalierung





von  
nd  
al-  
tigter  
Rolle.  
Schaus  
die,  
senerscheinung  
Merkmale au

# **Statistik**

Take Home

## Take Home

- Arbeiten mit JMP
  - <https://community.jmp.com/t5/Getting-Started-with-JMP/Getting-Started-with-JMP-On-Demand-Course/ta-p/621172>
- Terminologie
  - Skalenniveaus
  - Grundgesamtheit / Stichprobe
  - Messungen
- Deskriptive Statistik
  - Häufigkeitsverteilungen
  - Beschreibung von Häufigkeitsverteilungen
    - Gestalt
    - Lageparameter
    - Streuungsparameter
  - Darstellung
    - Box-Plot
    - Balkendiagramm
  - Häufigkeiten
    - Absolute / relative Häufigkeit
    - Kumulierte Häufigkeiten