

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB1

Deskriptive Statistik I

Dr. rer. med. habil. Marita Ziepert /
Dr. rer. nat. Meinhard Mende,
IMISE, Universität Leipzig

imise.

Lernziele Deskriptive Statistik I

Am Ende dieser Vorlesung werden Sie

- wissen, dass Wissenschaft mit Modellen arbeitet,
- die Begriffe Grundgesamtheit und Stichprobe kennen,
- das Skalenniveau von Merkmalen bestimmen
- sowie die Bedingungen für Codelisten anwenden können.

Ziel deskriptiver Statistik

- Merkmale einer Grundgesamtheit (GGS),
  z. B. der BMI (kg / m²)
- sollen aus den Daten einer Stichprobe
- anhand weniger Kenngrößen beschrieben
  z. B.: Mittelwert ± Standardabweichung
- oder durch eine Grafik charakterisiert werden,
- sodass man auf den ersten Blick typische Eigenschaften der GGS erkennen kann

Klassische Anwendung: Tabelle 1 medizinischer Publikationen
(bei randomisierten Studien: Baseline-Tabelle)

V 2

2.1 Wissenschaft arbeitet mit Modellen

Wir denken in Modellen

- Leben ist sehr komplex.
- Um es zu erforschen muss man abstrahieren: Unwichtiges weglassen, vereinfachen, sich auf das Wesentliche konzentrieren.
- Man bildet ein Modell mit vorgegebenen Bedingungen.

George E. P. Box:
Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.
Box GEP, Draper NR. Empirical model-building and response surfaces. John Wiley & Sons 1987



Verschiedene Zwecke – verschiedene Modelle



- Den Maler interessiert vor allem die Farbe:



- Der Imker interessiert sich vor allem für die Blüten



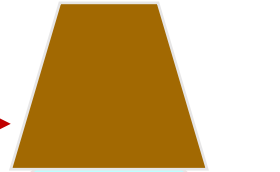
- und für den Landwirt sind die Erträge wesentlich

Wirklichkeit – und Modell

Aufgabe: Bestimme die Volumina von Bauch und Hut des Schneemannes!

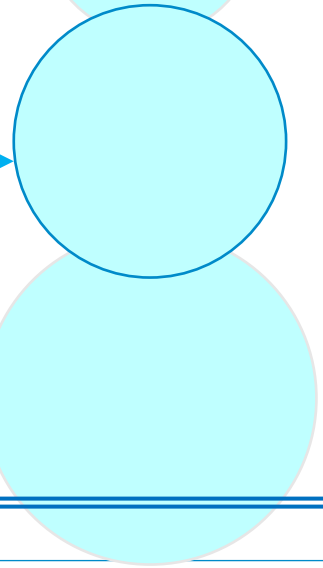


Modellierung



$$V = \frac{\pi \cdot h}{3}(R^2 + R \cdot r + r^2)$$

Modellierung



$$V = \frac{4\pi \cdot r^3}{3}$$

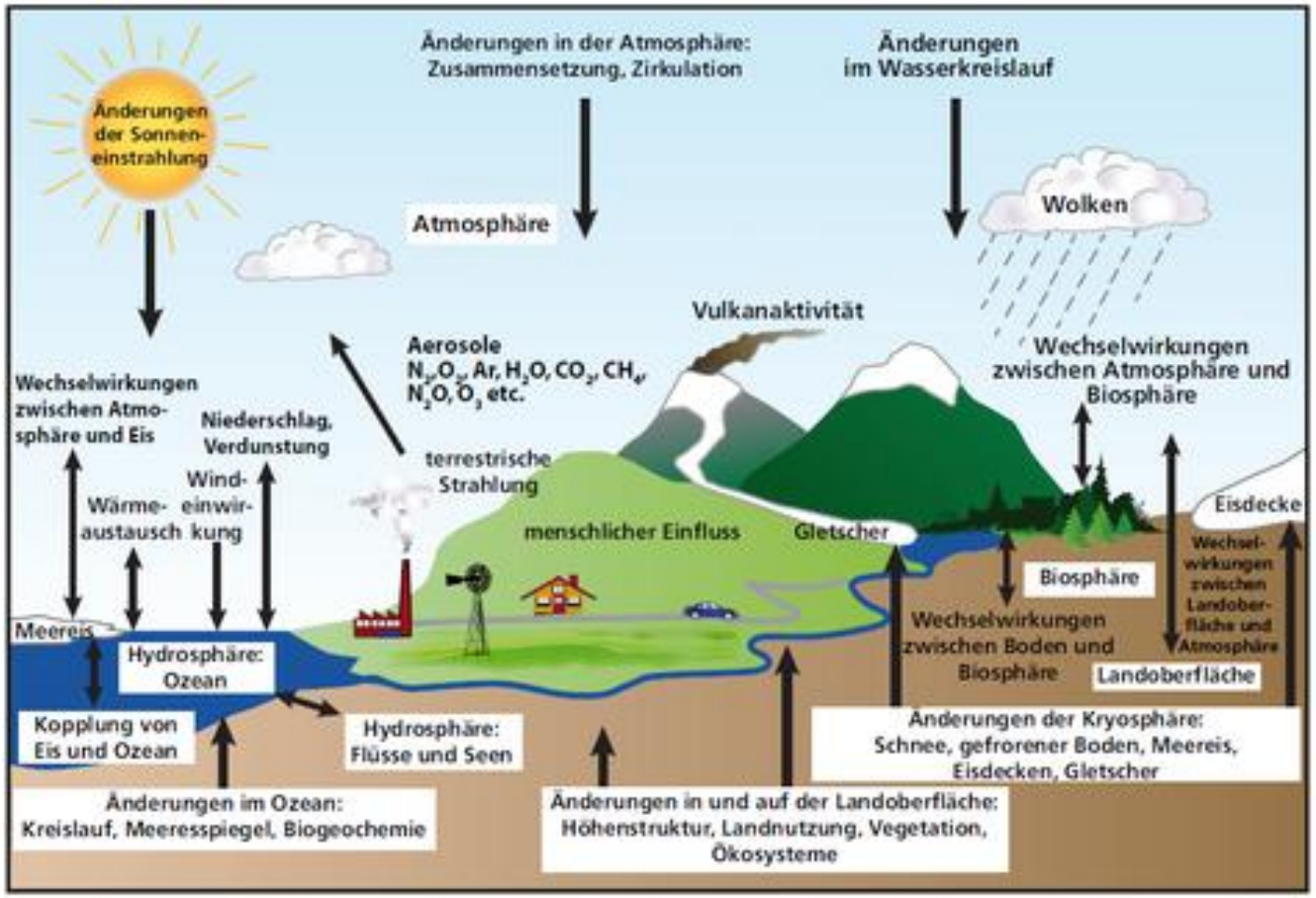
Interpretation

Ein extrem komplexes Modell (Geostatistik)

Wetter / Klima: chaotisches System

- Rückkopplungen
- nichtlineare Zusammenhänge

Zwischenstaatlicher Ausschuss für Klimaänderungen. Klimaänderung 2007: Wissenschaftliche Grundlagen, Beitrag der Arbeitsgruppe I zum Vierten Sachstandsbericht.



Since Newton, mankind has come to realize that the laws of physics are always expressed in the language of differential equations.

Steven Strogatz. Three blue one brown.
https://www.youtube.com/watch?v=p_di4Zn4wz4&list=PLhTS_25Wqvk2R0PMzRH3s4kWF08LD3z5h

V 2

2.2 Grundgesamtheit und Stichprobe

B 2.1 Die LIFE-Adult-Studie

Ziel: Erforschung der Ursachen von Zivilisationskrankheiten

Studientyp:

- Populations-basierte epidemiologische Kohortenstudie
- zunächst Querschnitt, dann Nachbeobachtung

Erfasste Merkmale:

- anthropometrische Größen, 3D-Body-Scan
- Labor, Herzecho, Hirn-MRT usw.

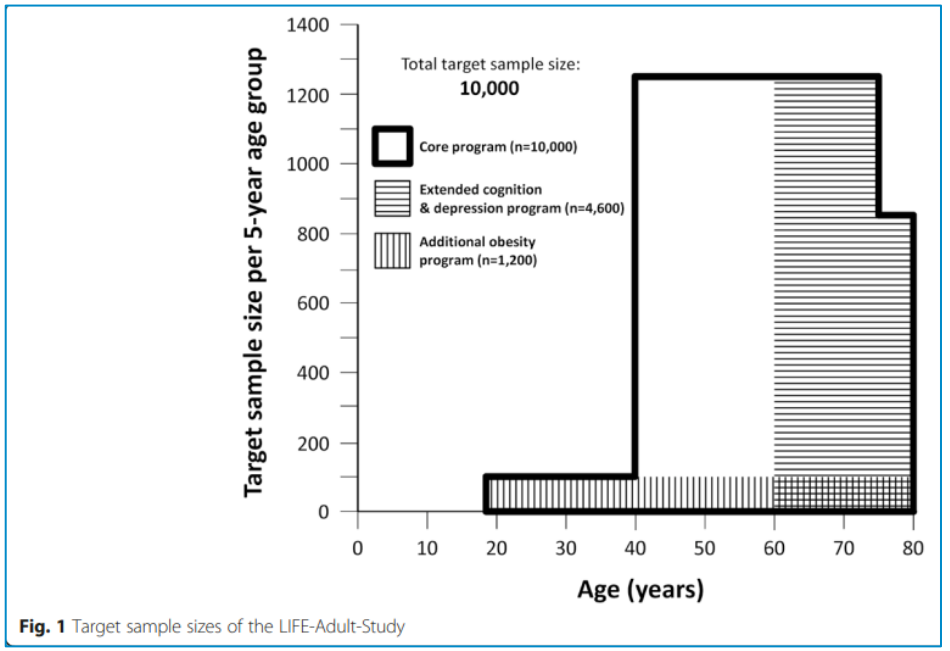


Fig. 1 Target sample sizes of the LIFE-Adult-Study

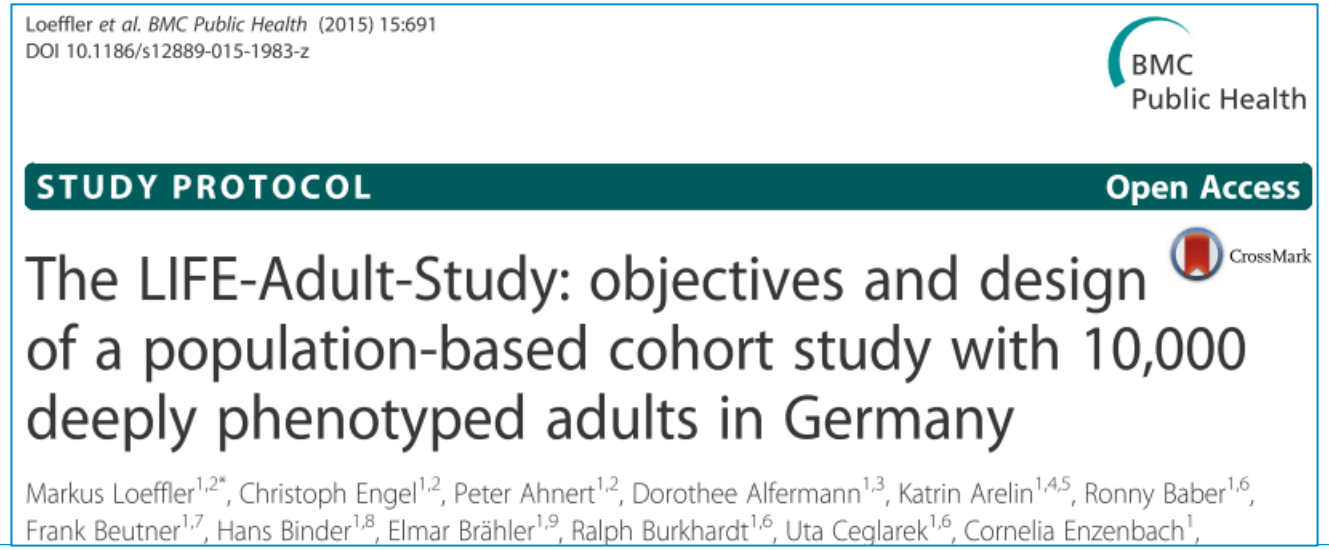
Löffler M et al. BMC Public Health 2015;15:691-704

LIFE-Adult-Studie: Grundgesamtheit

Eine Grundgesamtheit (GGS) ist die Anzahl Einheiten, zu denen eine Aussage getroffen werden soll (https://de.statista.com/statistik/lexikon/definition/60/grundgesamtheit/, 2026-09-14)

Beispiel: LIFE-Adult-Studie

- GGS: Alle Leipziger 2011 - 2014 zwischen 18 und 79 Jahren



Loeffler *et al. BMC Public Health* (2015) 15:691
DOI 10.1186/s12889-015-1983-z

BMC Public Health

STUDY PROTOCOL Open Access

CrossMark

The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany

Markus Loeffler[1,2*], Christoph Engel[1,2], Peter Ahnert[1,2], Dorothee Alfermann[1,3], Katrin Arelin[1,4,5], Ronny Baber[1,6], Frank Beutner[1,7], Hans Binder[1,8], Elmar Brähler[1,9], Ralph Burkhardt[1,6], Uta Ceglarek[1,6], Cornelia Enzenbach[1],

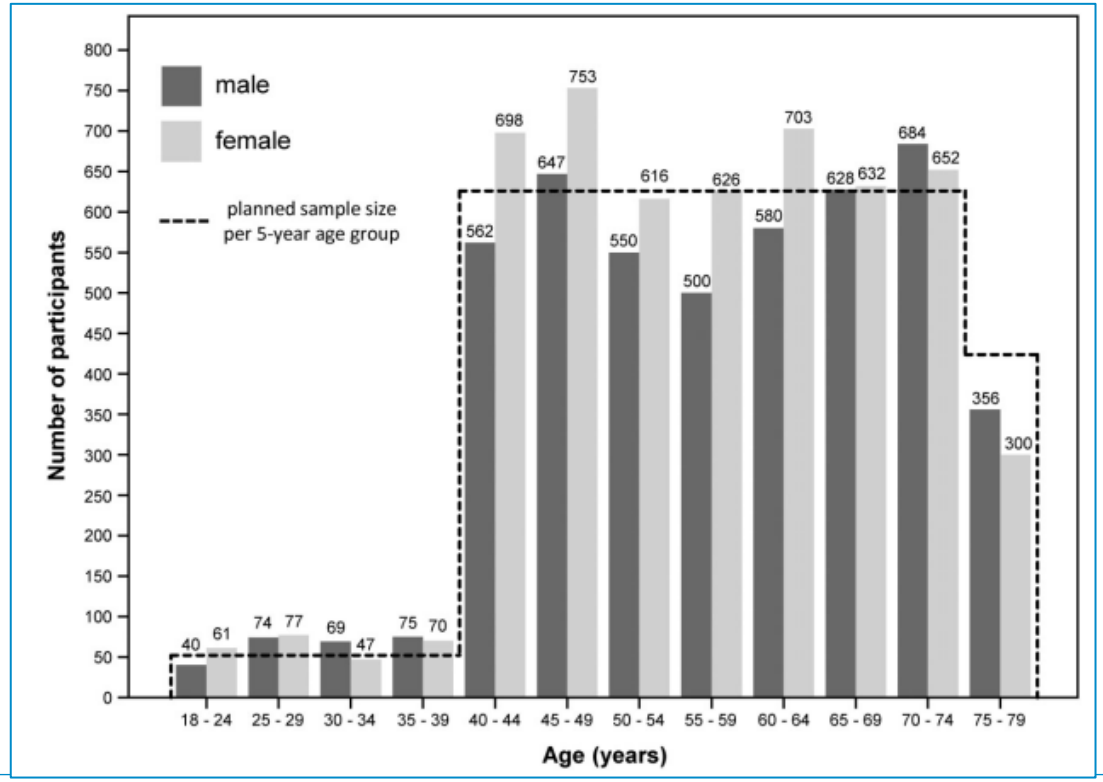
Stichprobe

Da es i. d. R. zu aufwändig ist, alle Probanden der GGS zu messen, erfasst man eine (zufällige) Auswahl, eine Stichprobe.

Beispiel LIFE-Adult:

– 10 000 Probanden

Engel C et al. International Journal of Epidemiology. 2023;52:e66–e79



Szenario: Bundestagswahl
Was sind Grundgesamtheit und evtl. Stichprobe?

Repräsentativität

- Eine Stichprobe muss repräsentativ sein, damit die aus ihren Daten gewonnenen Schlüsse auch auf die GGS übertragen werden können (externe Validität).
- Repräsentativität erzielt man durch möglichst zufällige Auswahl von Probanden aus der Kohorte (→ V 6).

Unstatistik des Monats März 2019:

- Nach einer online-Studie der Firma Duden Learnattack hätten gern 9 von 10 Lehrern mehr digitale Technologien zur Unterstützung des Unterrichts …
- Was sagen Sie dazu?

https://www.mpib-berlin.mpg.de/unstatistik-mehr-digital-gepraegter-unterricht (2026-09-14)



Zum Beispiel 1.1

Welches ist in der Ex-DHF-Studie die Grundgesamtheit?

– stabile Probanden mit Herzinsuffizienz mit erhaltener Pumpfunktion

Welches war die Stichprobe?

– N = 322 Probanden von Kliniken aus Göttingen, Berlin, Graz, München u. a.



V 2

2.3 Mess- / Skalenniveaus

Merkmale und Ausprägungen

- In wissenschaftlichen Untersuchungen werden Objekte oder Individuen, die sogenannten Beobachtungseinheiten,

im Hinblick auf vorgegebene Fragestellungen untersucht.

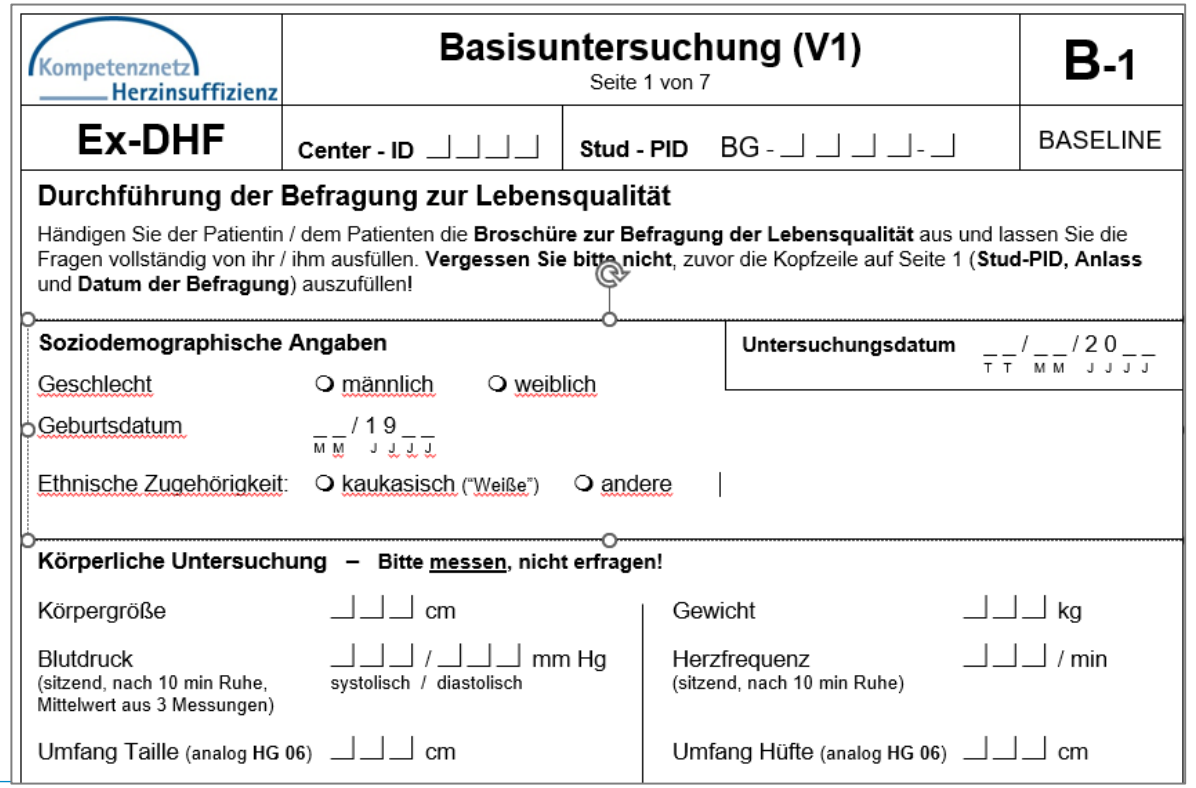
  - z.B.: Patienten, Hüftprothesen, Arztpraxen

- Diese Beobachtungseinheiten sind Träger von Merkmalen,
  - z.B.: Größe, Alter, NYHA-Klasse, Komorbiditäten
- die in unterschiedlichen Ausprägungen vorliegen
  - z.B.: Körpergröße, Alter, BMI, …
  - NYHA I – IV, CCS-Klasse …
  - Diabetes mellitus, Z. n. Myokardinfarkt …

Case Report Form (CRF)

Der Erfassungsbogen (CRF) ist ein wichtiges Werkzeug zur Datenerhebung auf Papier oder online am Bildschirm.

- Wir erkennen Merkmale auf ganz verschiedenem Skalenniveau:
  - Körpergröße
  - Geschlecht
  - ethnische Herkunft
  - Geburtsdatum



Kompetenznetz Herzinsuffizienz

Basisuntersuchung (V1) Seite 1 von 7 — **B-1**

Ex-DHF | **Center - ID** _ _ _ _ | **Stud - PID** BG - _ _ _ _ - _ | BASELINE

Durchführung der Befragung zur Lebensqualität

Händigen Sie der Patientin / dem Patienten die **Broschüre zur Befragung der Lebensqualität** aus und lassen Sie die Fragen vollständig von ihr / ihm ausfüllen. **Vergessen Sie bitte nicht**, zuvor die Kopfzeile auf Seite 1 (**Stud-PID, Anlass** und **Datum der Befragung**) auszufüllen!

Soziodemographische Angaben

Untersuchungsdatum _ _ / _ _ / 2 0 _ _ (T T / M M / J J J J)

Geschlecht ❍ männlich ❍ weiblich

Geburtsdatum _ _ / 1 9 _ _ (M M / J J J J)

Ethnische Zugehörigkeit: ❍ kaukasisch ("Weiße") ❍ andere

Körperliche Untersuchung – Bitte messen, nicht erfragen!

Körpergröße	_ _ _ cm	Gewicht	_ _ _ kg
Blutdruck (sitzend, nach 10 min Ruhe, Mittelwert aus 3 Messungen)	_ _ _ / _ _ _ mm Hg (systolisch / diastolisch)	Herzfrequenz (sitzend, nach 10 min Ruhe)	_ _ _ / min
Umfang Taille (analog HG 06)	_ _ _ cm	Umfang Hüfte (analog HG 06)	_ _ _ cm

Daten

Die bei der Untersuchung an den Beobachtungseinheiten festgestellten Ausprägungen eines Merkmals sind die Daten. Diese sind Gegenstand der statistischen Auswertung.

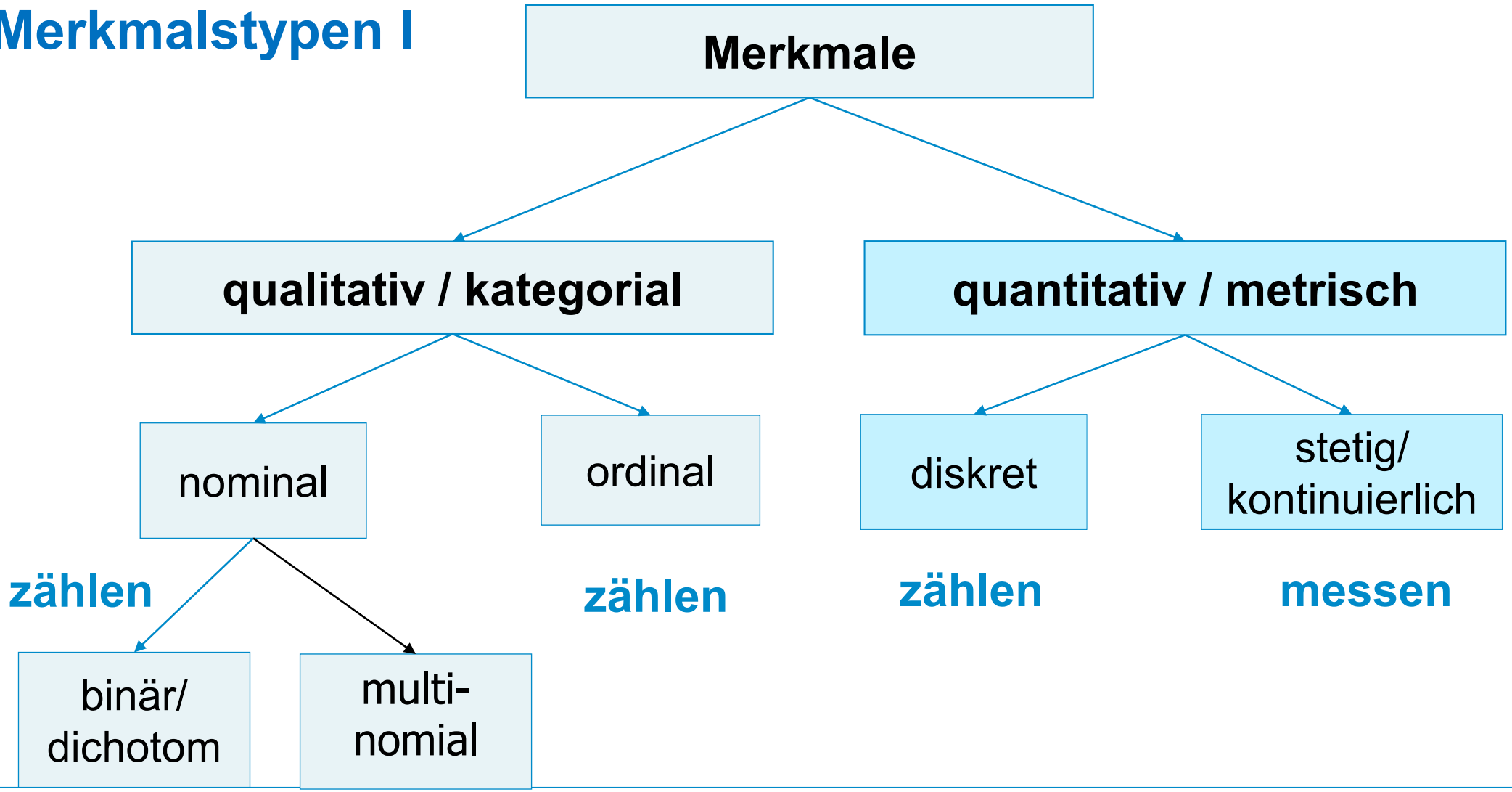
- Vor Studienbeginn wird für jedes betrachtete Merkmal die Liste möglicher Ausprägungen festgelegt (Codelisten):
- Sie muss jede denkbare Ausprägung enthalten und
- je zwei verschiedene Ausprägungen müssen sich gegenseitig ausschließen.

(Die Liste der Ausprägungen muss disjunkt und vollständig sein.)

z. B.: Erkrankungen des Herz-Kreislauf-Systems

1 – ja, 2 – nein, 3 – unbekannt

Merkmalstypen I

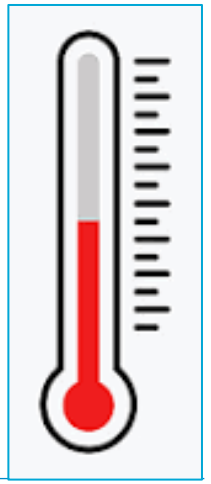


Datentyp – Mess- oder Skalenniveau

- Art und Anzahl der Ausprägungen (= der Wertebereich) bestimmen das Mess- oder Skalenniveau, Synonyme für den Datentyp des Merkmals.
- Davon wiederum hängen ganz wesentlich die statistischen Auswertungsmethoden ab, die man anwendet.

Eine entscheidende Frage vor Erhebung und Auswertung lautet also:

Auf welcher Skala (Zählen, Messen mit Maßeinheit usw.) hat das Merkmal seine Werte?



https://de.pngtree.com/free-png-vectors/thermometer (2026-09-14)

Merkmalstypen II

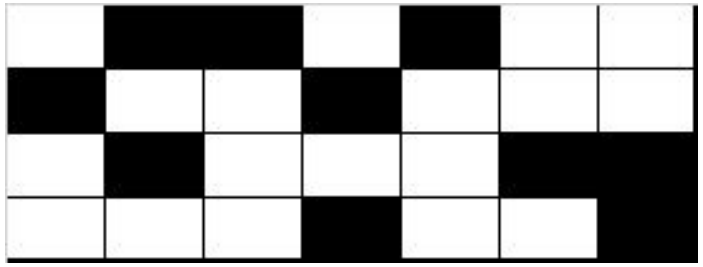
Ein Merkmal heißt qualitativ (bzw. kategorial),

- wenn seine verschiedenen Ausprägungen begrifflich voneinander unterschiedene Kategorien sind,
- die sich gegenseitig ausschließen, also disjunkt sind und
- alle denkbaren Fälle abdecken; z. B.
  - Diagnose Diabetes mellitus (j / n)
  - Genotyp (z. B. AA, AT, TT)

Zählen

- Qualitative Merkmale mit nur zwei Ausprägungen nennt man binär (bzw. dichotom).
  - durch die Ausprägung *unbekannt* wird das Merkmal trinär / multinomial

Binär – dichotom



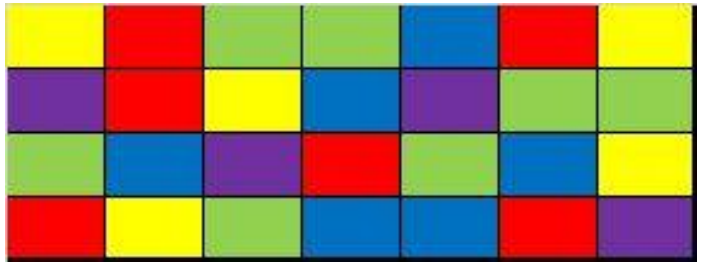
Binäre Merkmale haben nur zwei Ausprägungen

- Diabetes mellitus nein – ja
- Diagnose richtig – falsch
- Ruhedyspnoe vorhanden – nicht vorhanden
- Gesundheitsstatus gesund – krank
- Ladung positiv – negativ
- Wahrheit wahr – falsch

Der Informationsgehalt einer binären Variablen ist 1 Bit.

- Binäre Codierung ist Grundlage der digitalen Welt.

Qualitativ – nominal



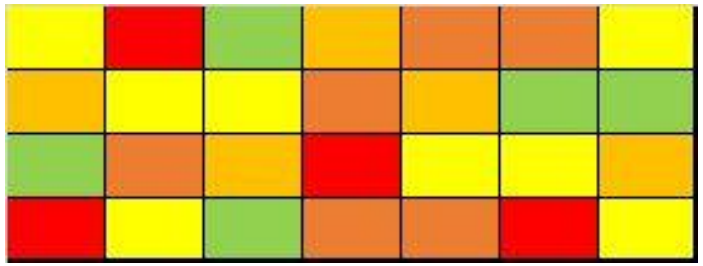
- syn. multinomial

m gleichwertige Ausprägungen (m > 2), z. B.

- Augenfarbe: blau – grau – braun – sonstige
- Familienstand: ledig – verheiratet – geschieden – verwitwet
- Hypertonie: nein – ja – unbekannt

Qualitativ – ordinal

m Ausprägungen (m > 2) mit einer natürlichen Rangfolge, z. B.





- NYHA-Klasse: I – II – III – IV
- Schmerzen: keine – leichte – mäßige – starke …

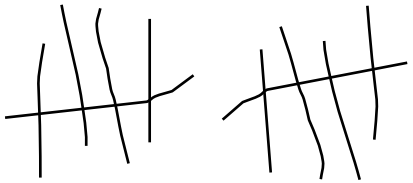
Die Abstände zwischen den Stufen ordinaler Merkmale sind undefiniert / ggf. ungleich.

Zählen

Quantitativ – diskret

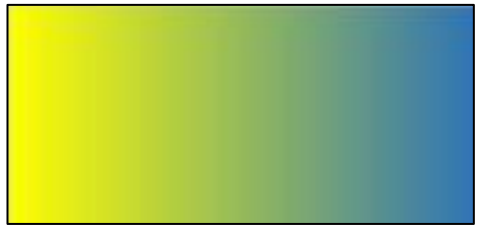
Ein Merkmal heißt quantitativ diskret,

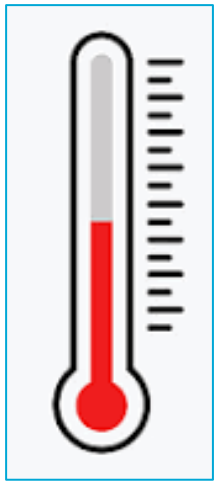
- wenn es bestimmte Eigenschaften zählt, z. B.
- Parität: 0 – 1 – 2 – 3 – …
- Anzahl Stents: 0 – 1 – 2 – 3 – …
- Tage im Krankenhaus: 0 – 1 – 2 – 3 – …



Zählen

(Quantitativ) Metrisch





- syn. kontinuierlich, stetig, engl. scale

Merkmale, die durch Messung einer Größe entstehen

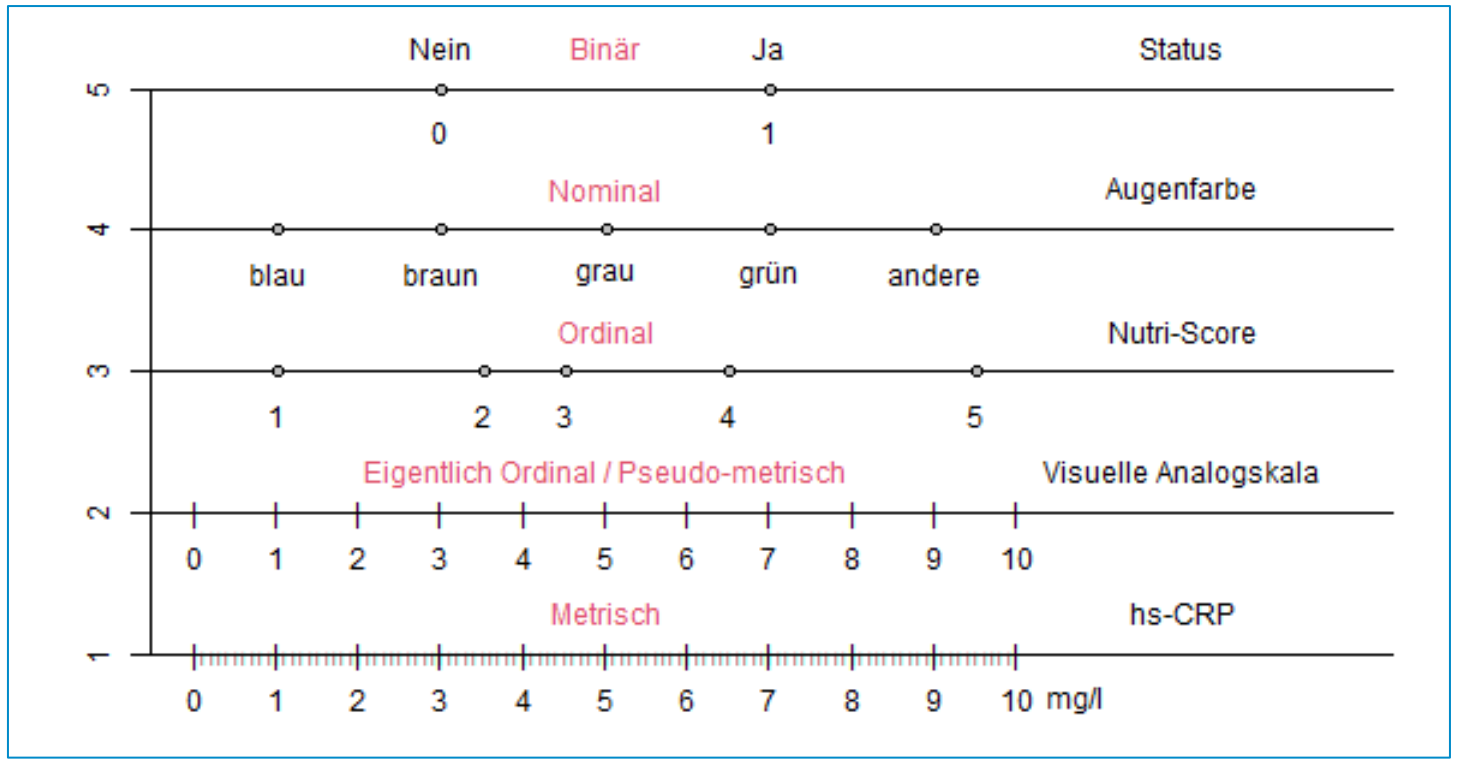
- alle Werte in einem Bereich sind möglich,
- es gibt (meist) eine Maßeinheit
- Messen nur in technisch vorgegebener Genauigkeit (Granularität)

Beispiele:

- Körpertemperatur: Maßeinheit °C
- hs-CRP im Serum (nmol / l)

Messen

Zusammenfassung: Mess- / Skalenniveaus



Diskret / Kategorial



Kontinuierlich / Metrisch

- Hinweis: Auf höchstmöglichem Skalenniveau arbeiten!

Übung

Welches Skalenniveau (welchen Merkmalstyp) haben die folgenden Variablen der Ex-DHF-Studie:

- Body-Maß-Index (BMI) _____
- Raucher (ja / nein) _____
- Wie viele Monate Diagnose Herzinsuffizienz? _____
- LV-EF _____
- Anzahl Geschwister _____
- NYHA-Klasse (I, II, III, IV) _____
- Endpunkt nach 1 Jahr _____
  (worsened – unchanged – improved)

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

Vielen Dank!

Dr. rer. nat. Meinhard Mende

Senior-Wissenschaftler, Institut für Medizinische Informatik, Statistik und Epidemiologie und Zentrum für Klinische Studien

Härtelstraße 16-18, 04107 Leipzig

Meinhard.mende@zks.uni-leipzig.de

imise.