Saarland University, Medical Faculty
PhD-Program

Medical statistics

PD Dr. Stefan Gräber
Institut für Medizinische Biometrie,
Epidemiologie und Medizinische Informatik
Universität des Saarlandes, Homburg

Content overview (1)

- Study question and hypothesis
- Study design
- Basic terms
- Descriptive statistical analysis (frequency tables, summarizing measures, graphical presentation)

Content overview (2)

- Confirmatory statistical analysis (statistical test, confidence interval)
- Special approaches (regression analysis, survival analysis)

Study question and hypotheses

- Study question is the clear definition of issue which shall be answered with the study.
- Hypotheses will be derived from the question. They are more specific and may be confirmed or rejected by a statistical test.
- Collecting data without study question is unscientific!

Examples (1)

- Which treatment of varicosis* (stripping or ELT) is better?
- Does drug A reduce the systolic blood pressure?

- The infection rate of ELT is 3% and of stripping 15%.
- The average reduction of blood pressure by drug A is 10 mmHg.

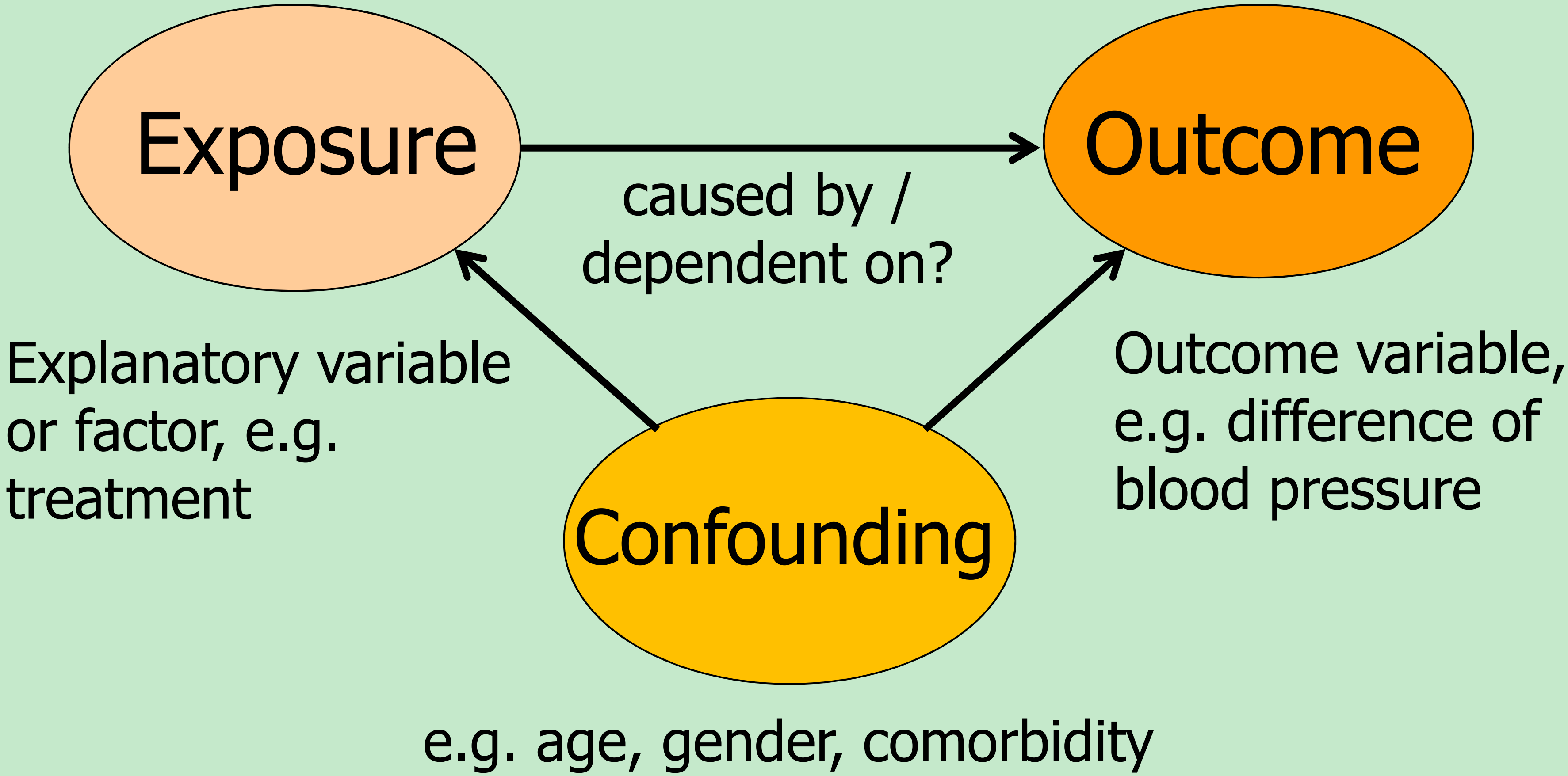
* Varicose veins (commonly on the leg) are veins that have become enlarged and tortuous, because the leaflet valves to prevent blood from flowing backwards are insufficient.

Examples (2)

- Is overweight a risk factor for stroke?
- How is the prognosis after the complete ectomy of a colon tumor?

- When BMI > 30 the risk is twice as much when BMI $\leq$ 30.
- After 5 years 80% of patients are still alive.

Statistical model



How to get data?

- Experiment
- Survey
- Trial
- Study

Study type

- If exposure is manipulated, you perform an experimental or interventional study (trial).
- If exposure is not manipulated, you perform a non-experimental or observational study.

Study type: Interventional

- Controlled
  - Randomized: Randomized controlled trial (RCT)
  - Non-randomized (quasi-experimental)
- Non-controlled

Example: Hypericum study (1)

- Objective: To investigate the efficacy of hypericum extract LI160 (St John's wort) compared with placebo in patients with mild or moderate major depression.
- Design: Randomized controlled double blind multicenter trial
- Setting: 3 psychiatric primary care units

Example: Hypericum study (2)

- Participants: 89 adult outpatients with mild or moderate depression (Hamilton score < 17)
- Interventions: LI160 or placebo three times a day for four weeks
- Main outcome variable: Change in Hamilton score from baseline to day 28

Hamilton score

Hamilton Rating Scale for Depression

Anwendung: Depressives Syndrom, für psychiatrische Patienten entwickelt, aber auch für Patienten mit anderen Diagnosen verwendbar
Bereich: Überwiegend wissenschaftliche Untersuchungen
Dauer: 5–15 Min.
Ergebnisbereich: 0–55 Punkte; eine hohe Punktzahl charakterisiert einen hohen Schwegrad der Depression
Literatur: Hamilton M. Development of a rating scale for primary depressive illness. Br J Clin Psychol 1967; 6: 278–296.

1. Depressed mood

Sad, hopeless, helpless, worthless

0 Absent
1 Gloomy attitude, pessimism, hopelessness
2 Occasional weeping
3 Frequent weeping
4 Patient reports highlight these feeling states in his/her spontaneous verbal and non-verbal communication.

2. Feelings of guilt

0 Absent
1 Self-reproach, feels he/she has let people down
2 Ideas of guilt or rumination over past errors or sinful deeds
3 Present illness is punishment
4 Hears accusatory or denunciatory voices and/or experiences threatening visual hallucinations. Delusions of guilt

3. Suicide

0 Absent
1 Feels life is not worth living
2 Wishes he/she were dead, or any thoughts of possible death to self
3 Suicide, ideas or half-hearted attempt
4 Attempts at suicide (any serious attempt rates 4)

4. Insomnia, early

0 No difficulty falling asleep
1 Complaints of occasional difficulty in falling asleep i. e. more than half-hour
2 Complaints of nightly difficulty in falling asleep

5. Insomnia, middle

0 No difficulty
1 Patient complains of being restless and disturbed during the night
2 Waking during the night – any getting out of bed rates 2 (except voiding bladder)

6. Insomnia, late

0 No difficulty
1 Waking in the early hours of the morning but goes back to sleep
2 Unable to fall asleep again if he/she gets out of bed

7. Work and activities

0 No difficulty
1 Thoughts and feelings of incapacity related to activities: work of hobbies
2 Loss of interest in activity – hobbies or work – either directly reported by patient or indirectly seen in listlessness, in decisions and vacillation (feels he/she has to push self to work or activities)
3 Decrease in actual time spent in activities or decrease in productivity. In hospital, rate 3 if patient does not spend at least three hours a day in activities
4 Stopped working because of present illness. In hospital rate 4 if patient engages in no activities except supervised ward chores

8. Retardation

Slowness of thought and speech; impaired ability to concentrate; decreased motor activity.

0 Normal speech and thought
1 Slight retardation at interview
2 Obvious retardation at interview
3 Interview difficult
4 Interview impossible

9. Agitation

0 None
1 Fidgetness
2 Playing with hands, hair, obvious restlessness
3 Moving about; can't sit still
4 Hand wringing, nail biting, hair pulling, biting of lips, patient is on the run

10. Anxiety, psychic

Demonstrated by:

- *subjective tension and irritability, loss of concentration*
- *worrying about minor matters*
- *apprehension*
- *fears expressed without questioning*
- *feelings of panic*
- *feeling jumpy*

0 Absent
1 Mild
2 Moderate
3 Severe
4 Incapacitating

11. Anxiety, somatic

Physiological concomitants of anxiety such as:

- *gastrointestinal: dry mouth, wind, indigestion, diarrhoea, cramps, belching*
- *cardiovascular: palpitations, headaches*
- *respiratory: hyperventilation, sighing*
- *urinary frequency*
- *sweating*
- *giddiness, blurred vision*
- *tinnitus*

0 Absent
1 Mild
2 Moderate
3 Severe
4 Incapacitating

12. Somatic symptoms; gastrointestinal

0 None
1 Loss of appetite but eating without encouragement
2 Difficulty eating without urging. Requests or requires laxation or medication for GI symptoms

13. Somatic symptoms; general

0 None
1 Heaviness in limbs, back or head; backaches, headaches, muscle aches, loss of energy, fatiguability
2 Any clear-cut symptom rates 2

14. Genital symptoms

Symptoms such as: loss of libido, menstrual disturbances:

0 Absent
1 Mild
2 Severe

15. Hypochondriasis

0 Not present
1 Self-absorption (bodily)
2 Preoccupation with health
3 Strong conviction of some bodily illness
4 Hypochondriacal delusions

16. Loss of weight

Rate either „A" or „B":

A. When rating by history:

0 No weight loss
1 Probable weight loss associated with present illness
2 Definite (according to patient) weight loss

B. Actual weight changes (weekly):

0 Less than 1 lb (0.5 kg) weight loss in one week
1 1–2 lb (0.5–1.0 kg) weight loss in week
2 Greater than 2 lb (1 kg) weight loss in week
3 Not assessed

17. Insight

0 Acknowledges being depressed and ill
1 Acknowledges illness but attributes cause to bad food, overwork, virus, need for rest, etc.
2 Denies being ill at all

Kommentar: Die Skala versucht, den Schwegrad einer Depression zu erfassen. Dafür werden 17 unterschiedliche Items bewertet. Es handelt sich um eine Fremdbewertungsskala. Sie erfordert psychiatrische Vorkenntnisse bei dem Benutzer. Sie ist weitverbreitet und auch deshalb von Bedeutung.

Change in Hamilton score

Treatment arm	Baseline Mean ± SD	After 4 weeks Mean ± SD
Hypericum (N=42)	15,57 ± 4,10	7,10 ± 3,11
Placebo (N=47)	14,96 ± 4,82	10,45 ± 3,60
p-value Mann-Whitney U-test	0,531	0,000

Responder = Patient whose Hamilton score after 4 weeks was ≤ 8 or decreased at least 50%

Responder rate

Arm * Responder Kreuztabelle

			Responder no	Responder yes	Gesamt
Arm	Hypericum	Anzahl	3	39	42
		% innerhalb von Arm	7,1%	92,9%	100,0%
	Placebo	Anzahl	17	30	47
		% innerhalb von Arm	36,2%	63,8%	100,0%
Gesamt		Anzahl	20	69	89
		% innerhalb von Arm	22,5%	77,5%	100,0%

OR = 0,136
[0,036; 0,506]

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)	Punkt-Wahrscheinlichkeit
Chi-Quadrat nach Pearson	10,727[a]	1	,001	,002	,001	
Kontinuitätskorrektur[b]	9,125	1	,003			
Likelihood-Quotient	11,714	1	,001	,001	,001	
Exakter Test nach Fisher				,002	,001	
Zusammenhang linear-mit-linear	10,606[c]	1	,001	,002	,001	,001
Anzahl der gültigen Fälle	89					

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 9,44.
b. Wird nur für eine 2x2-Tabelle berechnet
c. Die standardisierte Statistik ist -3,257.

Study type: Observational

- Controlled
  - Cohort-study
  - Case-control-study
- Non-controlled
  - Cohort-study
  - Cross-sectional study (survey)

Example cohort study (1)

- Study question: What are the causes and risks for cardiovascular disease in USA?
- Participants: Start of the study 1948 with 5209 men and women aged 30 - 62 years from Framingham (Massachusetts)

Link: http://www.framinghamheartstudy.org/

Example cohort study (2)

- Procedure: Every two years comprehensive medical check and interview about life style
- Results: Identification of the most important risk factors, like hypertension, hypercholesterolemia, smoking, overweight, diabetes

Example case-control study (1)

- Objective: To investigate the association between migraine and stroke in young women
- Participants: 291 women aged 20 - 44 years with stroke compared with 736 age and hospital matched controls
- Main outcome variable: self reported history of headaches

Example case-control study (2)

Table 2 Adjusted odds ratios* (95% confidence intervals) for types of stroke associated with personal or family history of migraine

	Ischaemic stroke†		Haemorrhagic stroke‡		All stroke§	
Variable	Odds ratios (95% CI)	No of cases/ controls	Odds ratios (95% CI)	No of cases/ controls	Odds ratios (95% CI)	No of cases/ controls
Simple	2.97 (0.66 to 13.5)	7/9	1.84 (0.77 to 4.39)	14/15	2.25 (1.10 to 4.63)	21/23
Classical	3.81 (1.26 to 11.5)	19/17	0.86 (0.44 to 1.67)	24/46	1.62 (0.98 to 2.67)	50/65
Migraine (total)	3.54 (1.30 to 9.61)	26/26	1.10 (0.63 to 1.94)	38/61	1.78 (1.14 to 2.77)	71/88
Family history of migraine§	4.99 (2.03 to 12.3)	23/26	2.30 (1.35 to 3.90)	41/50	2.55 (1.67 to 3.90)	65/76

*Reference group: women with no personal history of migraine.
†Adjusted for high blood pressure, education, smoking categories, family history of migraine (not in §), alcohol consumption, and social class.
‡Adjusted for high blood pressure, body mass index, smoking categories, and family history of migraine (not in §).
§Ischaemic, haemorrhagic, and unclassified stroke.

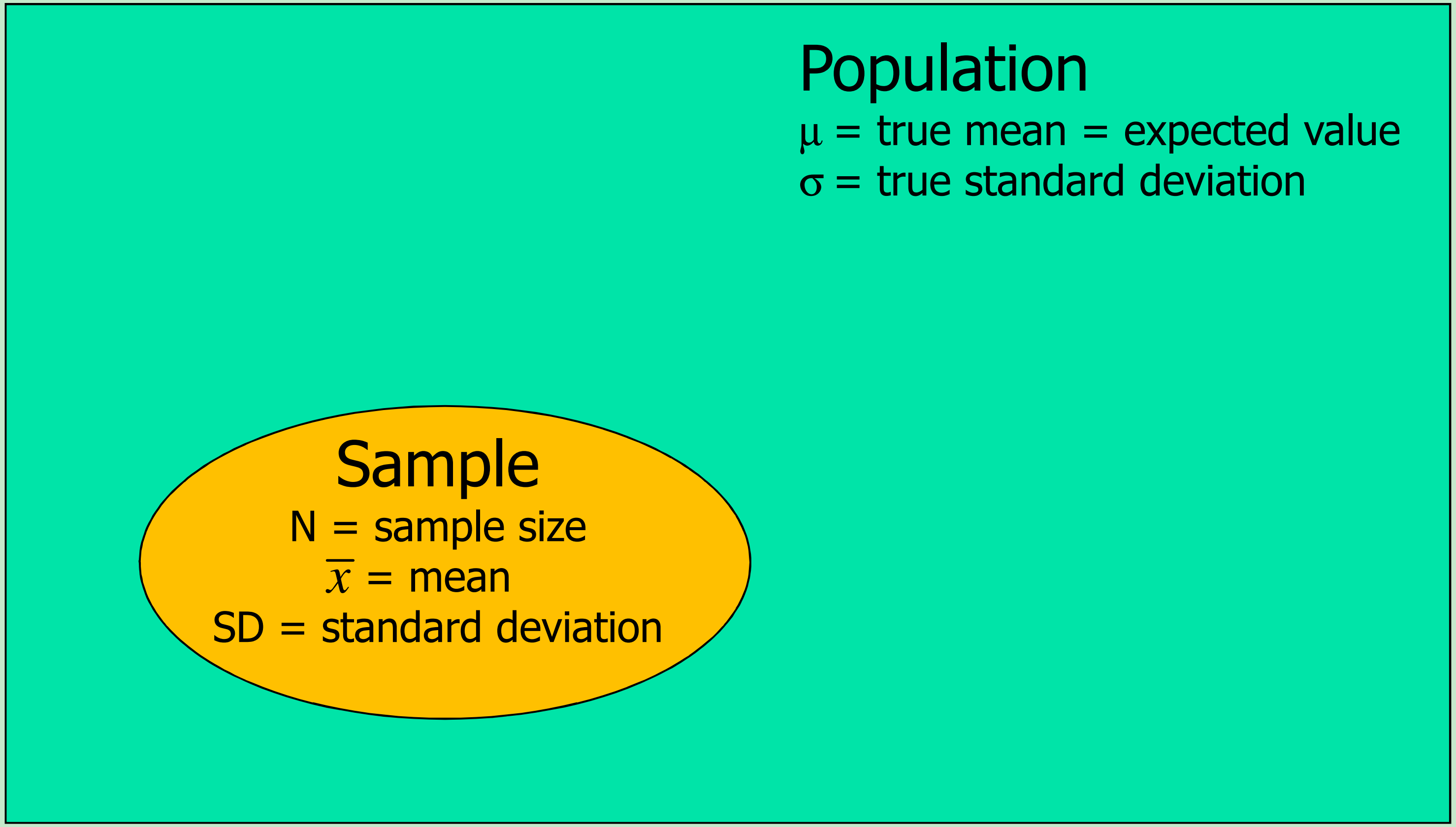
Study protocol

- Main study question, hypotheses
- Including and excluding criteria
- Outcome and explanatory variables
- Study type
- Follow-up, duration
- Statistical analysis methods
- …

Population and sample

- Population covers the entire group of individuals in whom you are interested.
- Due to size or inaccessibility of population almost always a subset can be investigated: The sample is the subset of individuals that are included in the study.
- Census: the sample consists of all members of the population.

Population and sample



Example

- Sample: 200 patients with hypertension, i.e. sample size N = 200.
- When the superiority of drug A is proved (based on sample data), potentially all patients with hypertension (=population) could be prescribed the drug, i.e. the result of sample is generalized to the population (statistical inference).

Statistical inference

- Inference from the (special) sample to the (general) population.
- Prerequisite: Random sample (also called representative sample), i.e. each object has the same chance to be selected for the sample.

Principles of statistical inference

- You want to prove a hypothesis: statistical test (hypotheses refer ever to population!)
- You want to estimate the true value of a parameter: estimation
- The certainty of a statistical result is ever lower than 100% (except for census)!

Oberservational unit / variable

- Observational unit is the object of a study, e.g. patient, animal, blood sample, …
- For each object the (for answering the study question relevant) properties have to be defined and measured as variables.

Values of variables

For each object the variable has a characteristic value, e.g.

Variable	Value
Gender	Female
Size	1,72 m
Weight	69 kg
Number of pregnancies	2
Blood pressure	120/70 mmHg
…	…

Types of variable

- Categorical (qualitative)
  - Nominal: categories are mutually exclusive and unordered, e.g. gender, eye colour Dichotomous or binary: two categories only, e.g. dead or alive, relapse y/n
  - Ordinal: categories are mutually exclusive and ordered, e.g. disease stage, education level, quality of life

Coding

- = (arbitrary) assignment of natural numbers to the categories
- Examples:
  - Variable gender: male = 1, female = 2
  - Variable histological type: epithelial = 1, intermediate = 2, anaplastic = 3, other = 4

Definition of categories Example: Lung cancer

- Yes / no
- Epithelial / mesothelial / other / no
- Using the WHO-classification

Table 1.–The 1999 World Health Organization/International Association for the Study of Lung Cancer Histological Classification of Lung and Pleural Tumours

1 Epithelial Tumours
- 1.1. Benign
  - *1.1.1. Papillomas*
    - 1.1.1.1. Squamous cell papilloma
      - Exophytic
      - Inverted
    - 1.1.1.2. Glandular papilloma
    - 1.1.1.3. Mixed squamous cell and glandular apilloma
  - *1.1.2. Adénomas*
    - 1.1.2.1. Alveolar adenoma
    - 1.1.2.2. Papillary adenoma
    - 1.1.2.3. Adenomas of salivary-gland type
      - Mucous gland adenoma
      - Pleomorphic adenoma
      - Others
    - 1.1.2.4. Mucinous cystadenoma
    - 1.1.2.5. Others
- 1.2. Preinvasive lesions
  - *1.2.1. Squamous dysplasia/Carcinoma in situ*
  - *1.2.2. Atypical adenomatous hyperplasia*
  - *1.2.3. Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia*
- 1.3. Malignant
  - *1.3.1. Squamous cell carcinoma*
    - Variants
    - 1.3.1.1. Papillary
    - 1.3.1.2. Clear cell
    - 1.3.1.3. Small cell
    - 1.3.1.4. Basaloid
  - *1.3.2. Small cell carcinoma*
    - Variant
    - 1.3.2.1. Combined small cell carcinoma
  - *1.3.3. Adénocarcinoma*
    - 1.3.3.1. Acinar
    - 1.3.3.2. Papillary
    - 1.3.3.3. Bronchioloalveolar carcinoma
      - 1.3.3.3.1. Non-mucinous (Clara/pneumocyte type II)
      - 1.3.3.3.2. Mucinous
      - 1.3.3.3.3. Mixed mucinous and non-mucinous or intermediate cell type
    - 1.3.3.4. Solid adenocarcinoma with mucin
    - 1.3.3.5. Adenocarcinoma with mixed subtypes
    - 1.3.3.6. Variants
      - 1.3.3.6.1. Well-differentiated fetal adenocarcinoma
      - 1.3.3.6.2. Mucinous ("colloid") adenocarcinoma
      - 1.3.3.6.3. Mucinous cystadenocarcinoma
      - 1.3.3.6.4. Signet-ring adenocarcinoma
      - 1.3.3.6.5. Clear cell adernocarcinoma
  - *1.3.4. Large cell carcinoma*
    - Variants
    - 1.3.4.1. Large cell neuroendocrine carcinoma
      - 1.3.4.1.1. Combined large cell neuroendocrine carcinoma
    - 1.3.4.2. Basaloid carcinoma
    - 1.3.4.3. Lymphoepithelioma-like carcinoma
    - 1.3.4.4. Clear cell carcinoma
    - 1.3.4.5. Large cell carcinoma with rhabdoid phenotype
  - *1.3.5. Adenosquamous carcinoma*
  - *1.3.6. Carcinomas with pleomorphic, sarcomatoid or sarcomatous elements*
    - 1.3.6.1. Carcinomas with spindle and/or giant cells
      - 1.3.6.1.1. Pleomorphic carcinoma
      - 1.3.6.1.2. Spindle cell carcinoma
      - 1.3.6.1.3. Giant cell carcinoma
    - 1.3.6.2. Carcinosarcoma
    - 1.3.6.3. Pulmonary blastoma
    - 1.3.6.4. Others

Table 1. Continued

  - *1.3.7. Carcinoid tumour*
    - 1.3.7.1. Typical carcinoid
    - 1.3.7.2. Atypical carcinoid
  - *1.3.8. Carcinomas of salivary-gland type*
    - 1.3.8.1. Mucoepidermoid carcinoma
    - 1.3.8.2. Adenoid cystic carcinoma
    - 1.3.8.3. Others
  - *1.3.9. Unclassified carcinoma*

2 Soft Tissue Tumours
- 2.1 Localized fibrous tumour
- 2.2 Epithelioid hemangioendothelioma
- 2.3 Pleuropulmonary blastoma
- 2.4 Chondroma
- 2.5 Calcifying fibrous pseudotumour of the pleura
- 2.6 Congenital peribronchial myofibroblastic tumour
- 2.7 Diffuse pulmonary lymphangiomatosis
- 2.8 Desmoplastic small round cell tumour
- 2.9 Other

3 Mesothelial Tumours
- 3.1 Benign
  - *3.1.1 Adenomatoid tumour*
- 3.2 Malignant
  - *3.2.1 Epithelioid mesothelioma*
  - *3.2.2 Sarcomatoid mesothelioma*
    - 3.2.2.1 Desmoplastic mesothelioma
  - *3.2.3 Biphasic mesothelioma*
  - *3.2.4 Other*

4 Miscellaneous Tumours
- 4.1 Hamartoma
- 4.2 Sclerosing hemangioma
- 4.3 Clear cell tumour
- 4.4 Germ cell neoplasms
  - *4.4.1 Teratoma, mature or immature*
  - *4.4.2 Malignant germ cell tumour*
- 4.5 Thymona
- 4.6 Melanoma
- 4.7 Others

5 Lymphoproliferative Disease
- 5.1 Lymphoid interstitial pneumonia
- 5.2 Nodular lymphoid hyperplasia
- 5.3 Low-grade marginal zone B-cell lymphoma of the mucosa-associated lymphoid tissue
- 5.4 Lymphomatoid granulomatosis

6 Secondary Tumours

7 Unclassified Tumours

8 Tumour-like Lesions
- 8.1 Tumourlet
- 8.2 Multiple meningothelioid nodules
- 8.3 Langerhans cell histiocytosis
- 8.4 Inflammatory pseudotumour (Inflammatory myofibroblastic tumour)
- 8.5 Organizing pneumonia
- 8.6 Amyloid tumour
- 8.7 Hyalinizing granuloma
- 8.8 Lymphangioleiomyomatosis
- 8.9 Multifocal micronodular pneumocyte hyperplasia
- 8.10 Endometriosis
- 8.11 Bronchial inflammatory polyp
- 8.12 Others

emerged that would necessitate a change. An example of this is "small-cell lung carcinoma". As compared to the previous edition, changes include a better definition of pre-invasive lesion, a reclassification of adenocarcinoma, the description of two new tumour types as variants of large cell carcinoma, large cell neuroendocrine carcinoma (LCNEC) and basaloid

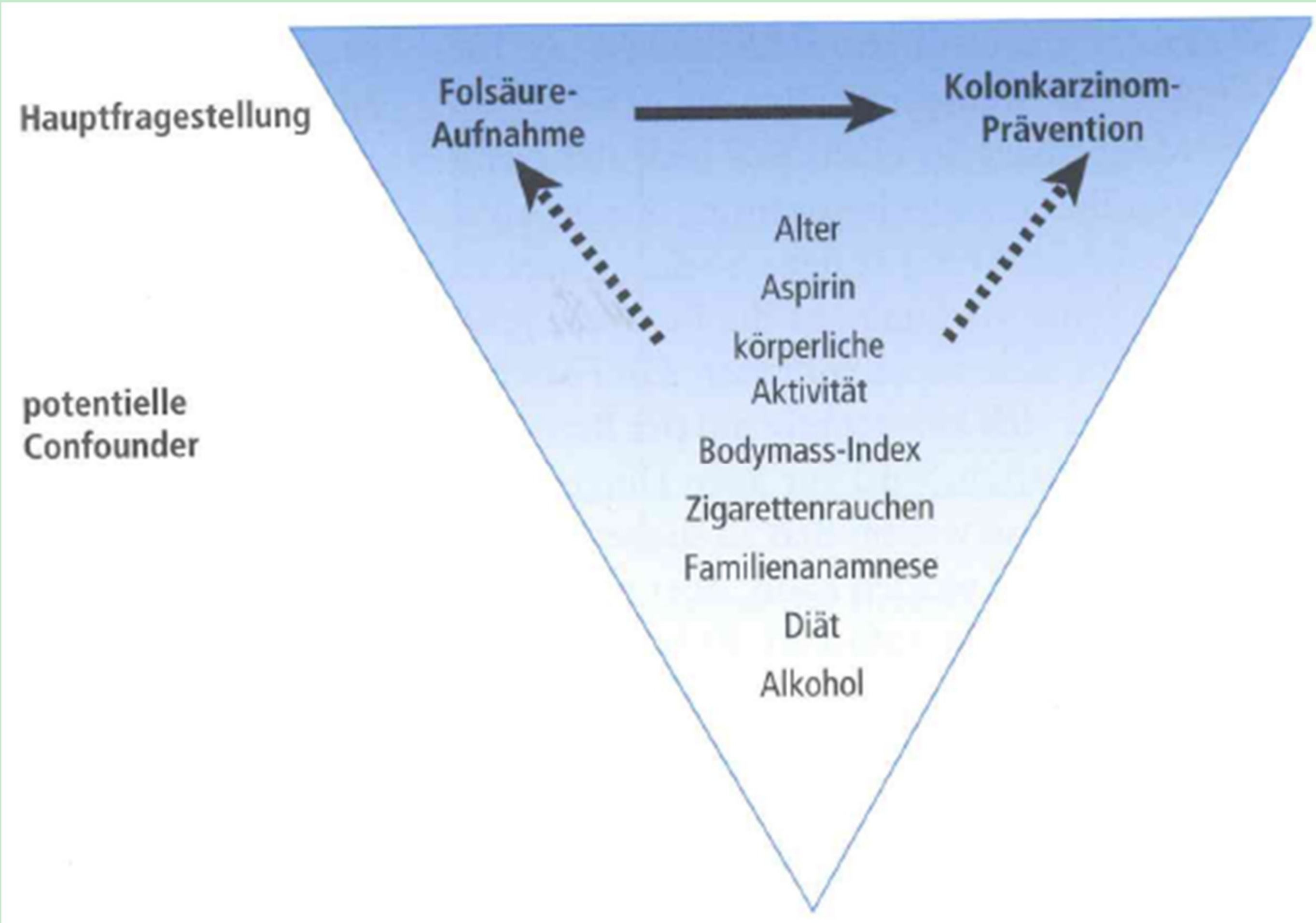
Types of variable

- Numerical (quantitative)
  - Counts (discrete): integer values, e.g. number of pregnancies, number of siblings
  - Continuous (measured): takes any value in a range of values (interval), e.g. blood pressure in mmHg, weight in kg, thickness in mm, age in years

Purpose of variables

- Identification
- Outcome variable
- Explanatory variable
  - Factor (qualitative)
  - Covariable (quantitative)
  - Confounder

Example confounding



Data recording

- = measuring and documentation of the values of all variables for each object
- Generating a rectangular structure (spreadsheed)
- Important: The measurements at different objects have to be indepen-dent (statistical independence)!

Checking data quality

Haarfarbe

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	hellblond	9	8,4	8,4	8,4
	dunkelblond	38	35,5	35,5	43,9
	rot / rotblond	4	3,7	3,7	47,7
	braun	45	42,1	42,1	89,7
	schwarz	9	8,4	8,4	98,1
	8	1	,9	,9	99,1
	keine Haare	1	,9	,9	100,0
	Gesamt	107	100,0	100,0	

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Größe	107	155	1175	180,78	97,376
Gültige Werte (Listenweise)	107				

Frequency

- Absolute frequency = number of occurrence of a value in a sample, e.g. 14 persons have blue eyes.
- Relative frequency = number of occurrence of a value in a sample / sample size, e.g. 14 persons of 57 have blue eyes, i.e. 24,6%.

A frequency distribution

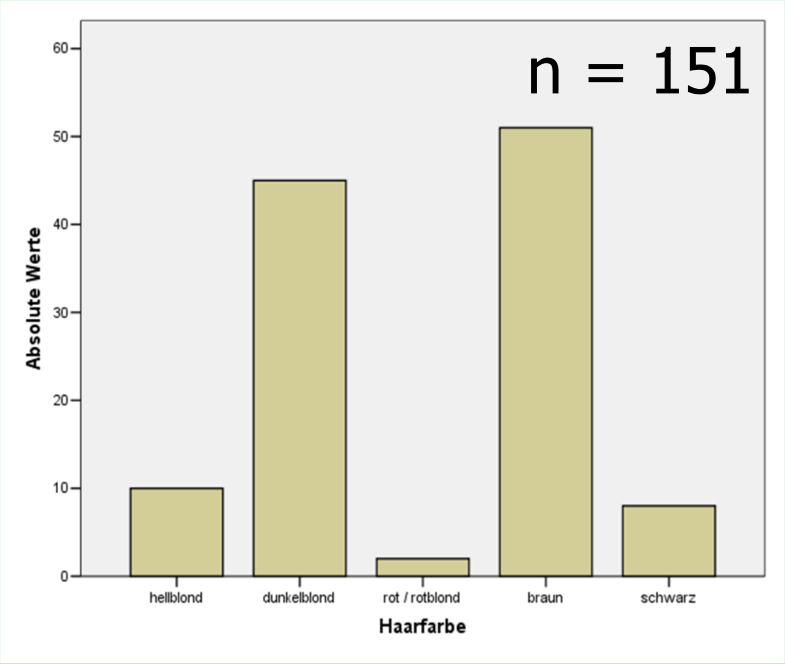
- describes how the frequencies are distributed on all (in the sample) occurring values.
- Presentation as frequency table or diagram

Frequency table

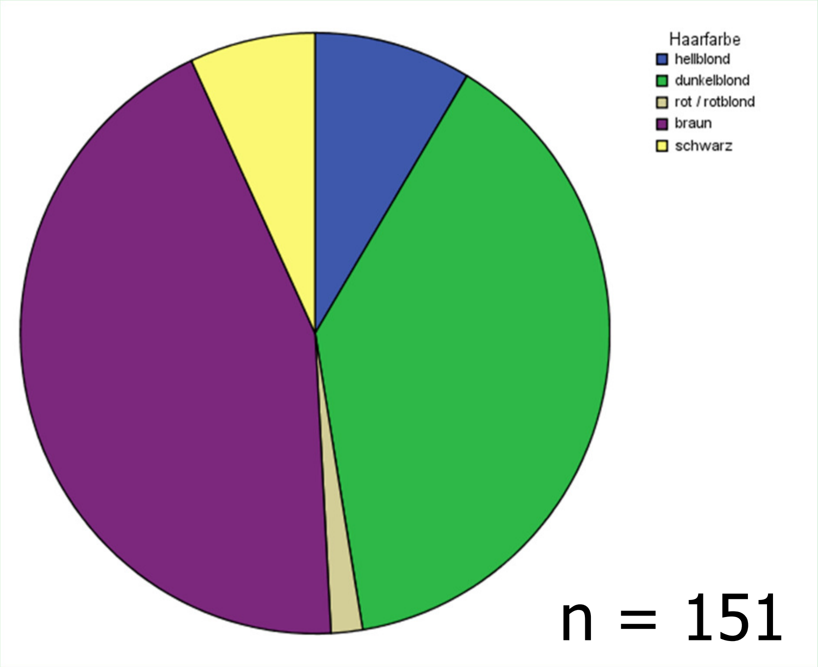
Haarfarbe

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	hellblond	10	8,5	8,6	8,6
	dunkelblond	45	38,5	38,8	47,4
	rot / rotblond	2	1,7	1,7	49,1
	braun	51	43,6	44,0	93,1
	schwarz	8	6,8	6,9	100,0
	Gesamt	116	99,1	100,0	
Fehlend	System	1	,9		
Gesamt		117	100,0		

Graphical data presentation



Bar chart



Pie chart

Frequency table

Alter

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	20,00	4	3,4	3,4	3,4
	21,00	28	23,9	24,1	27,6
	22,00	29	24,8	25,0	52,6
	23,00	20	17,1	17,2	69,8
	24,00	7	6,0	6,0	75,9
	25,00	7	6,0	6,0	81,9
	26,00	9	7,7	7,8	89,7
	27,00	4	3,4	3,4	93,1
	28,00	2	1,7	1,7	94,8
	30,00	2	1,7	1,7	96,6
	32,00	2	1,7	1,7	98,3
	38,00	1	,9	,9	99,1
	41,00	1	,9	,9	100,0
	Gesamt	116	99,1	100,0	
Fehlend	System	1	,9		
Gesamt		117	100,0		

Summarizing measures (parameters)

Central tendency / location

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

Median, quantiles, mode

Variation

Range = maximum - minimum

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

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

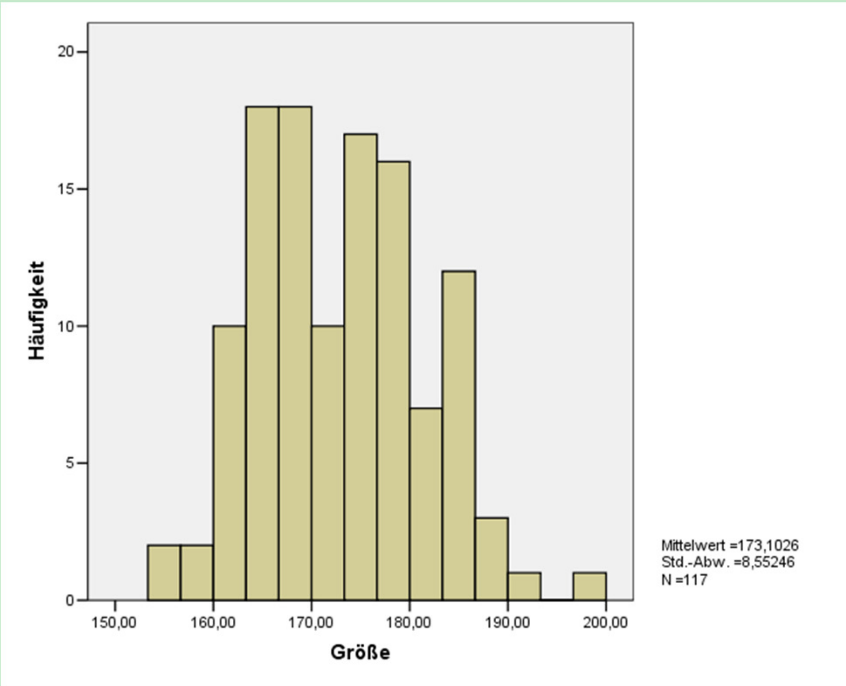
Interquartile distance = 0,75-Q – 0,25-Q

Summarizing measures

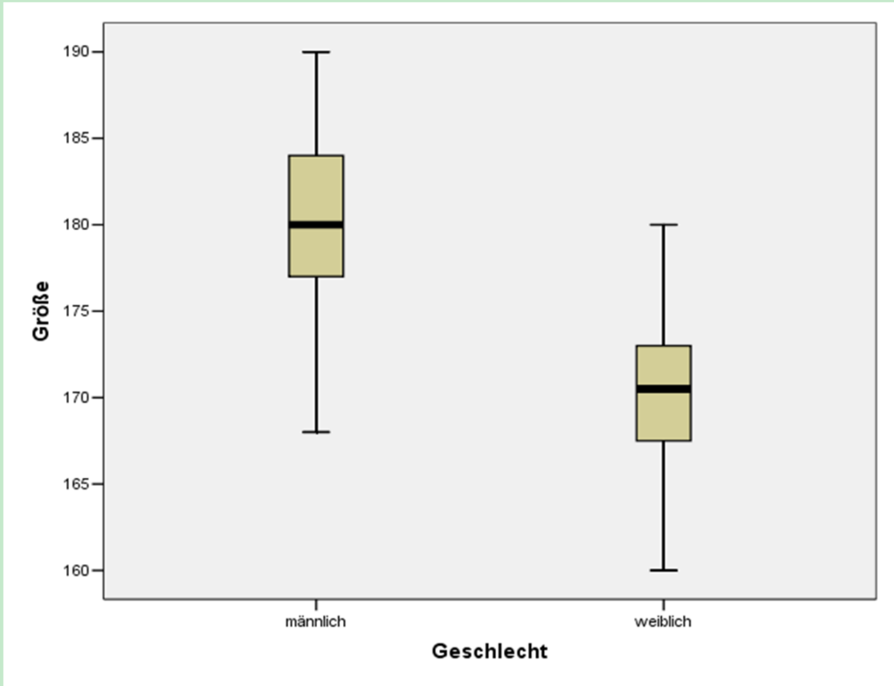
Statistiken

		Alter	Größe	Gewicht
N	Gültig	116	117	116
	Fehlend	1	0	1
Mittelwert		23,3621	173,1026	66,6888
Standardabweichung		3,23136	8,55246	12,41771
Varianz		10,442	73,145	154,200
Minimum		20,00	155,00	45,00
Maximum		41,00	200,00	98,70
Perzentile	25	21,0000	165,0000	58,0000
Median	50	22,0000	173,0000	64,0000
	75	24,0000	179,0000	75,7500

Graphical data presentation

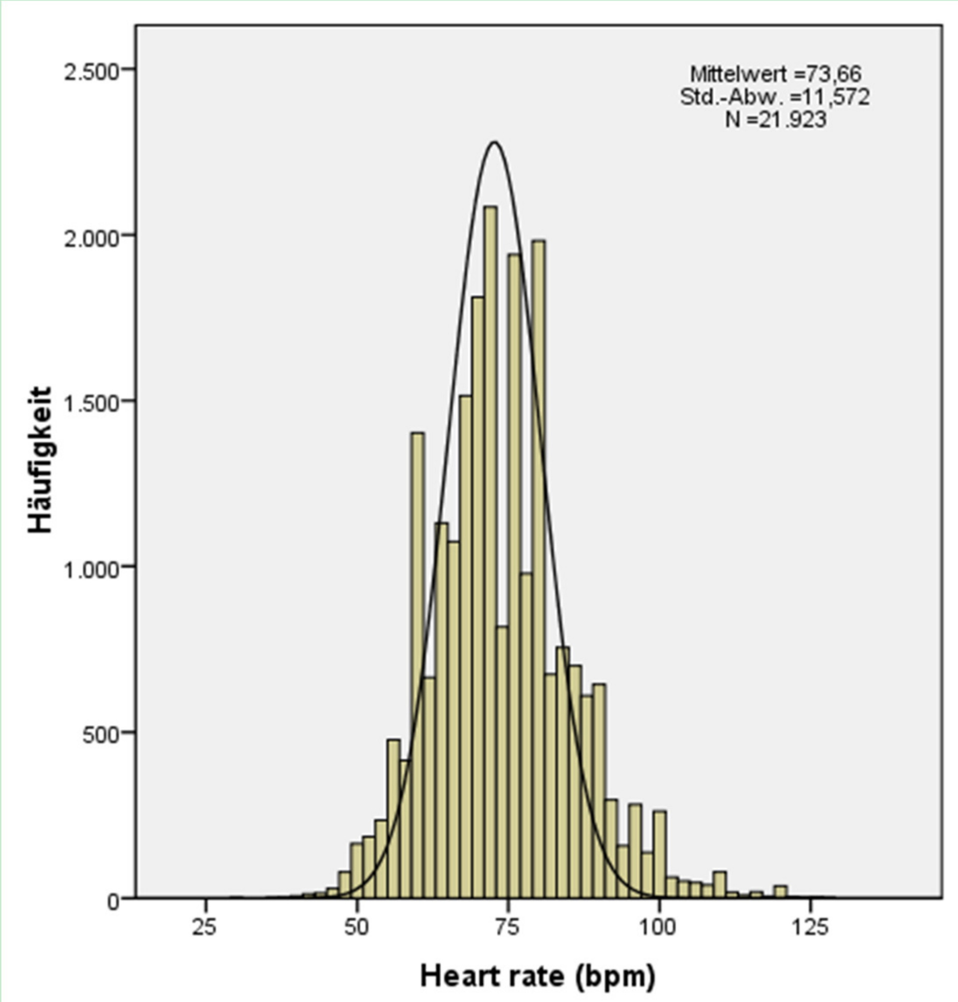


Histogram



Boxplot

Histogram with density function of normal distribution



Normal distribution (1)

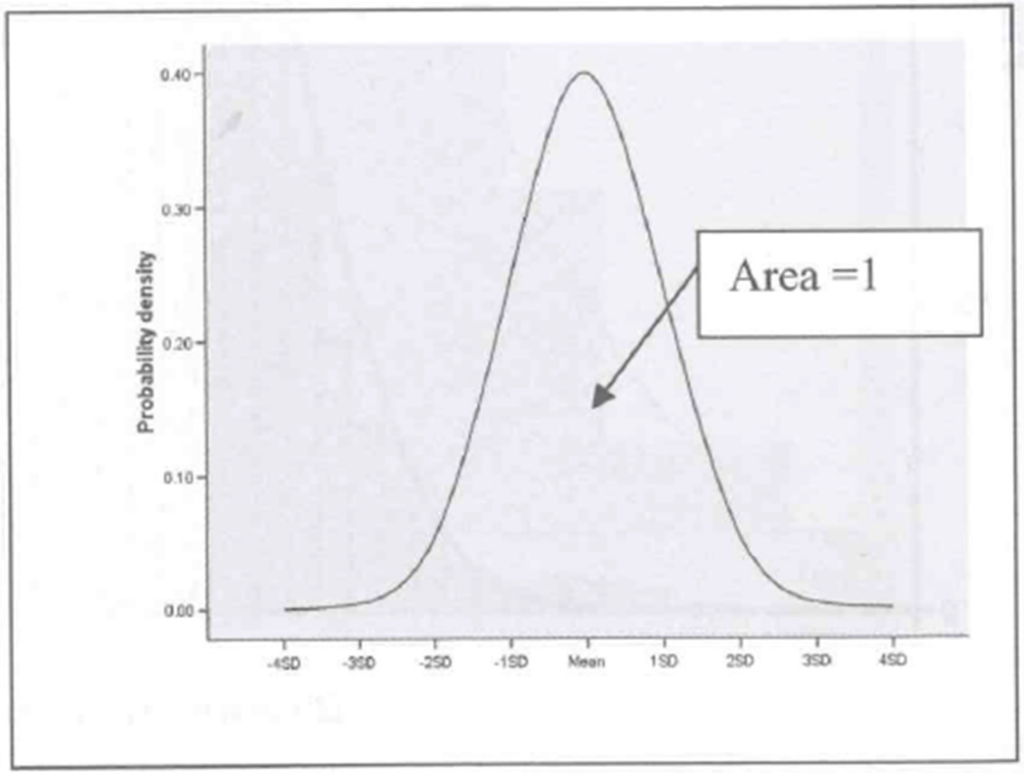
Total area under the curve = 1 (or 100%).

Bell shaped and symmetrical about its mean.

The peak of the curve lies above the mean.

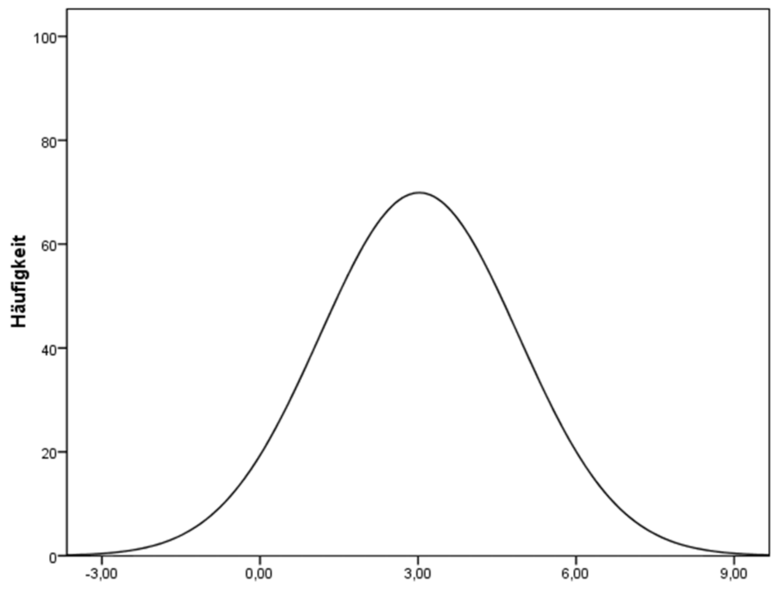
Any position along the horizontal axis can be expressed as a number of SDs away from the mean.

The mean and median coincide.



Normal distribution (2)

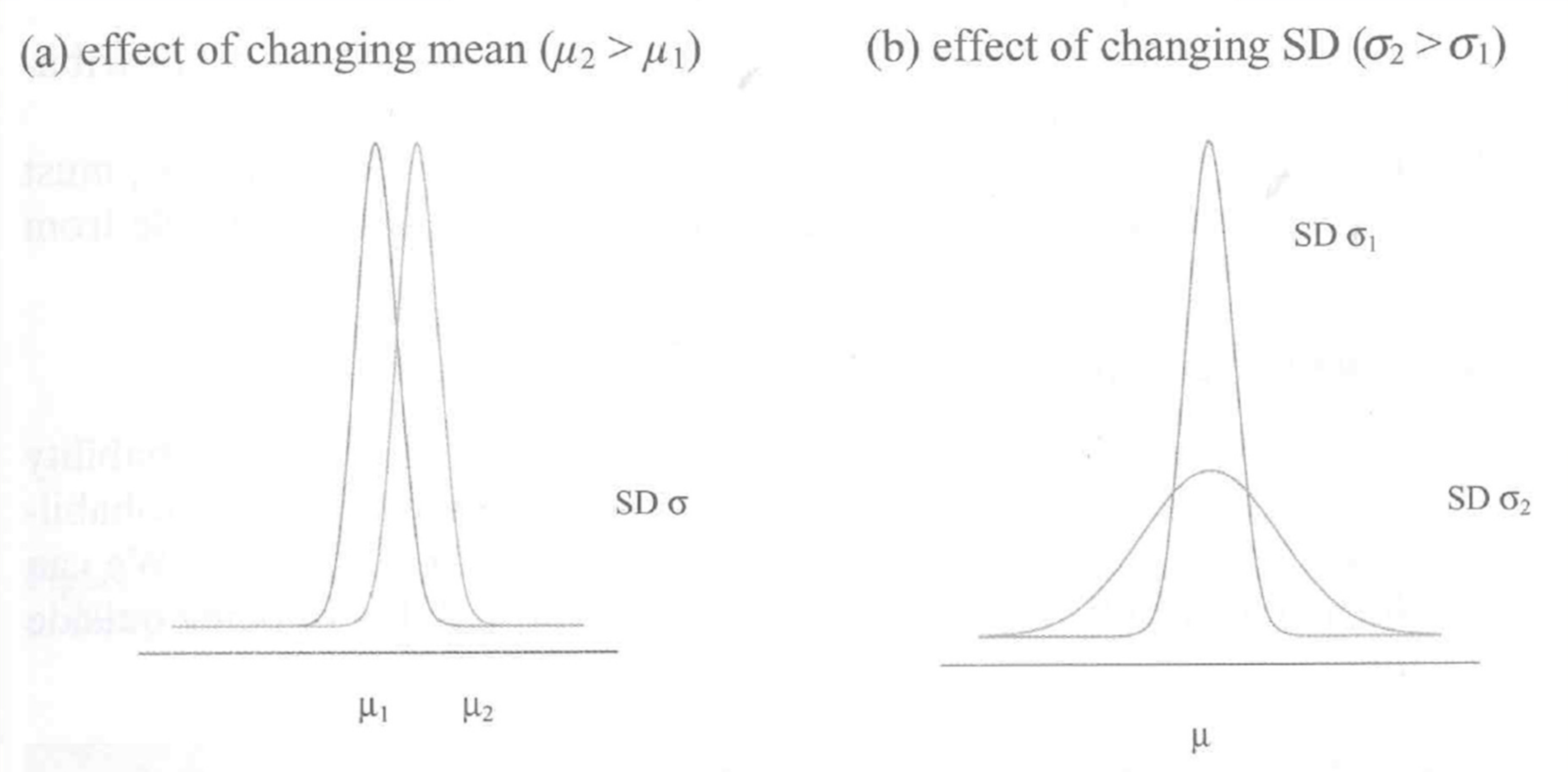
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$



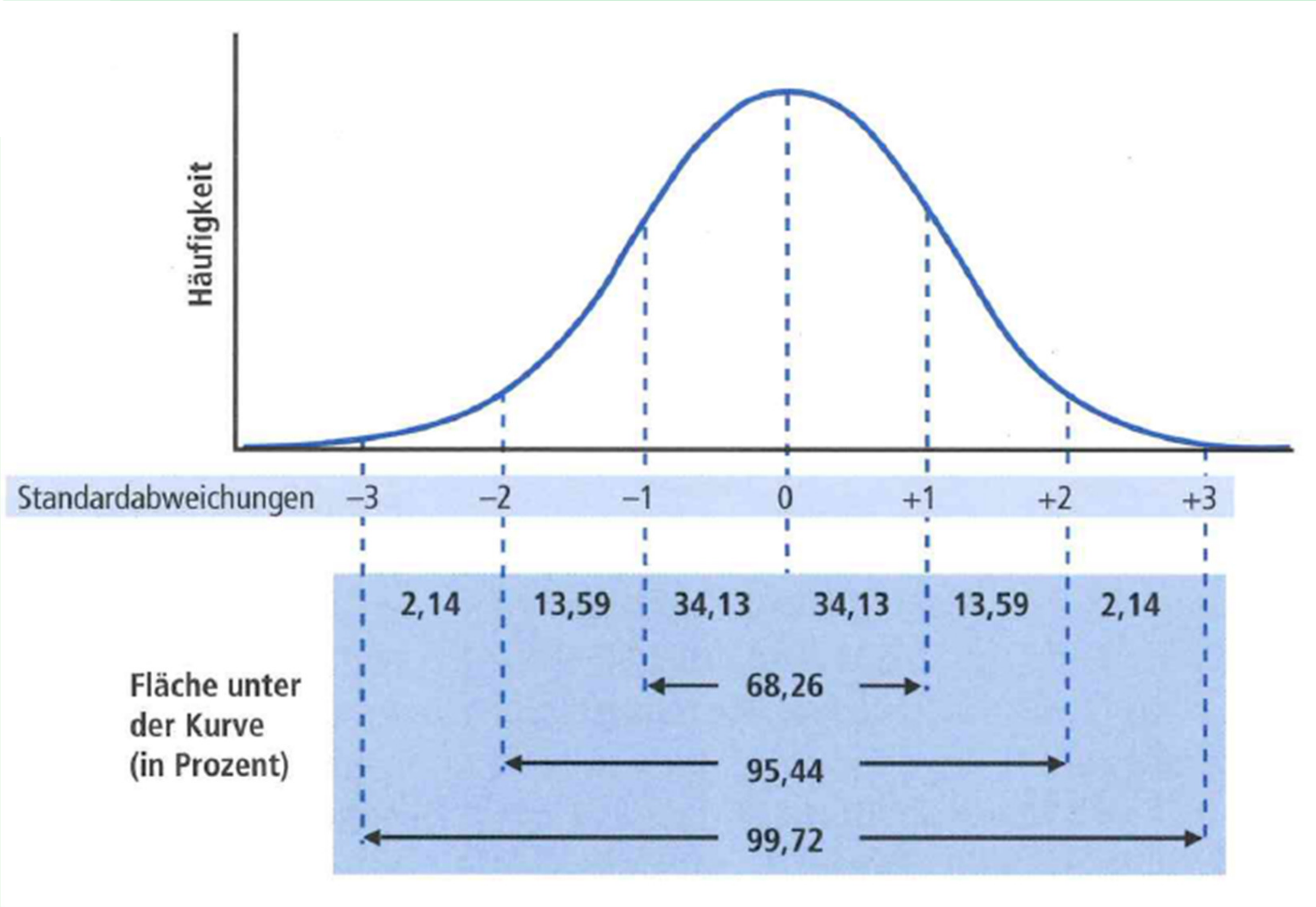
Density function of normal distribution with $\mu = 3$ and $\sigma = 4$

Normal distribution (3)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$



Normal distribution (4)



Normal distribution (5)

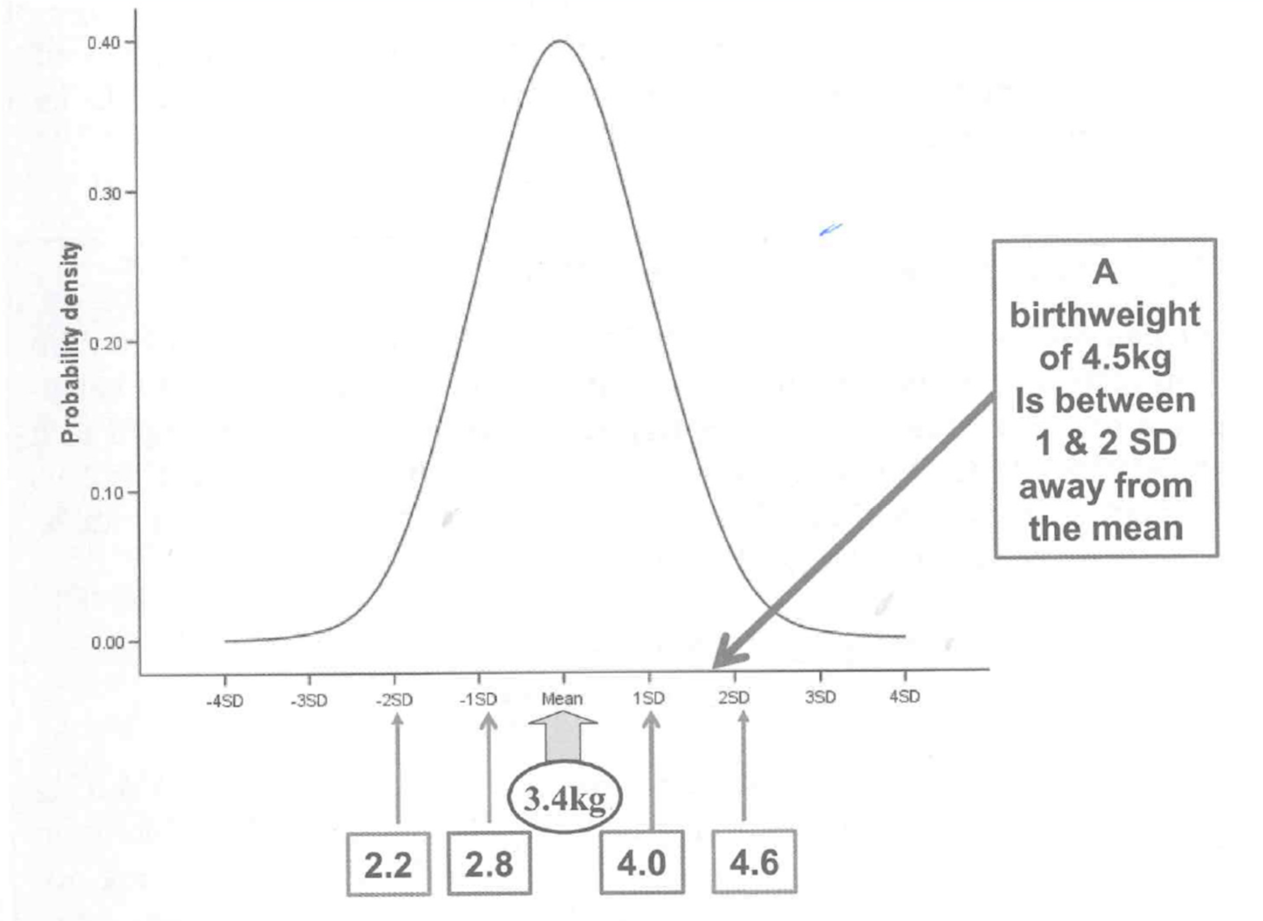
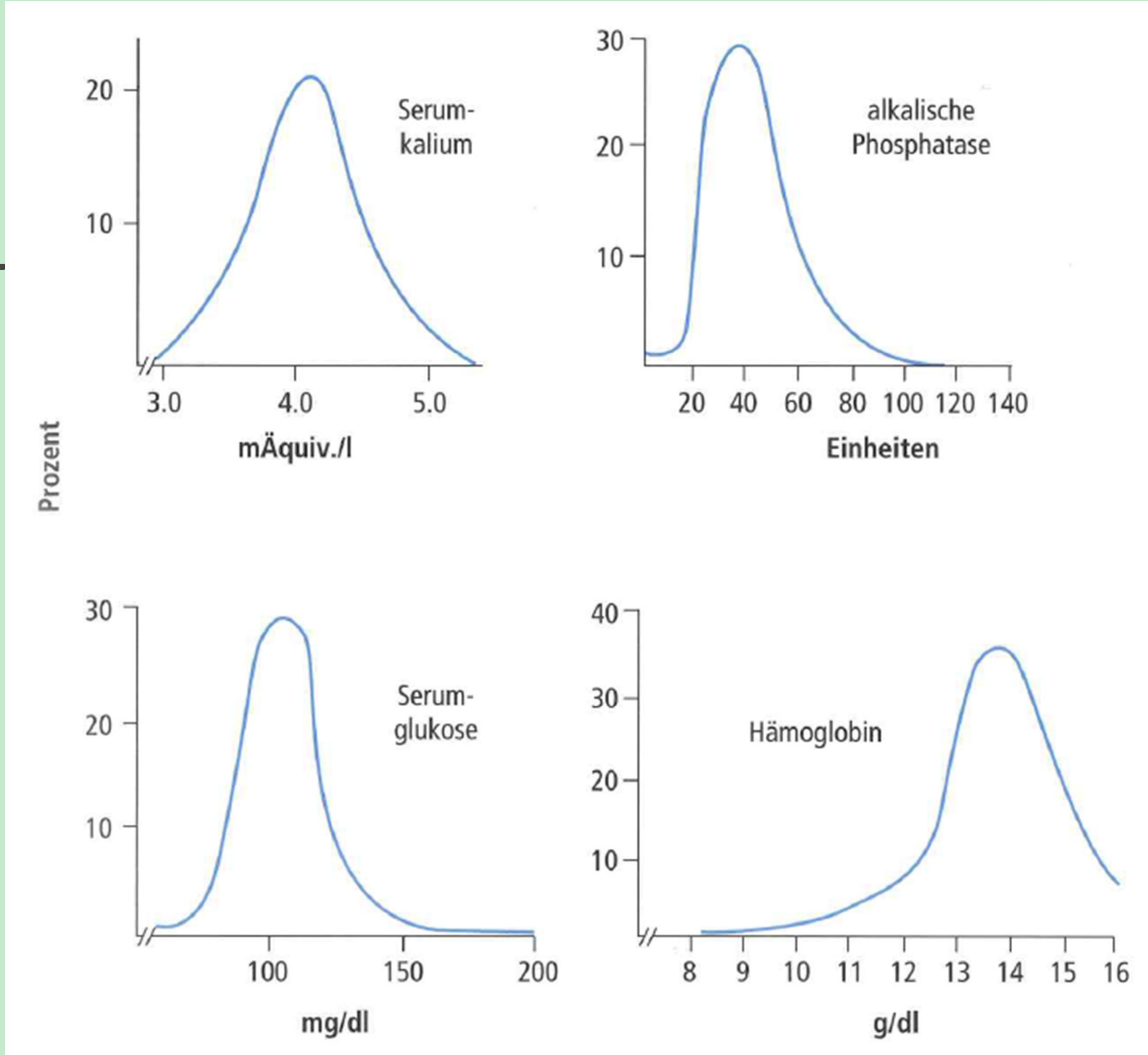


Figure 5.9 Normal distribution curve for birthweight with a mean of 3.4 kg and SD of 0.6 kg

Types of distribution



Analysis of two variables

- Both variables are qualitative: contingency table
- One variable is qualitative, one variable is quantitative: break-down table
- Both variables are quantitative: scattergram, correlation coefficient

Contingency table

Geschlecht * Haarfarbe Kreuztabelle

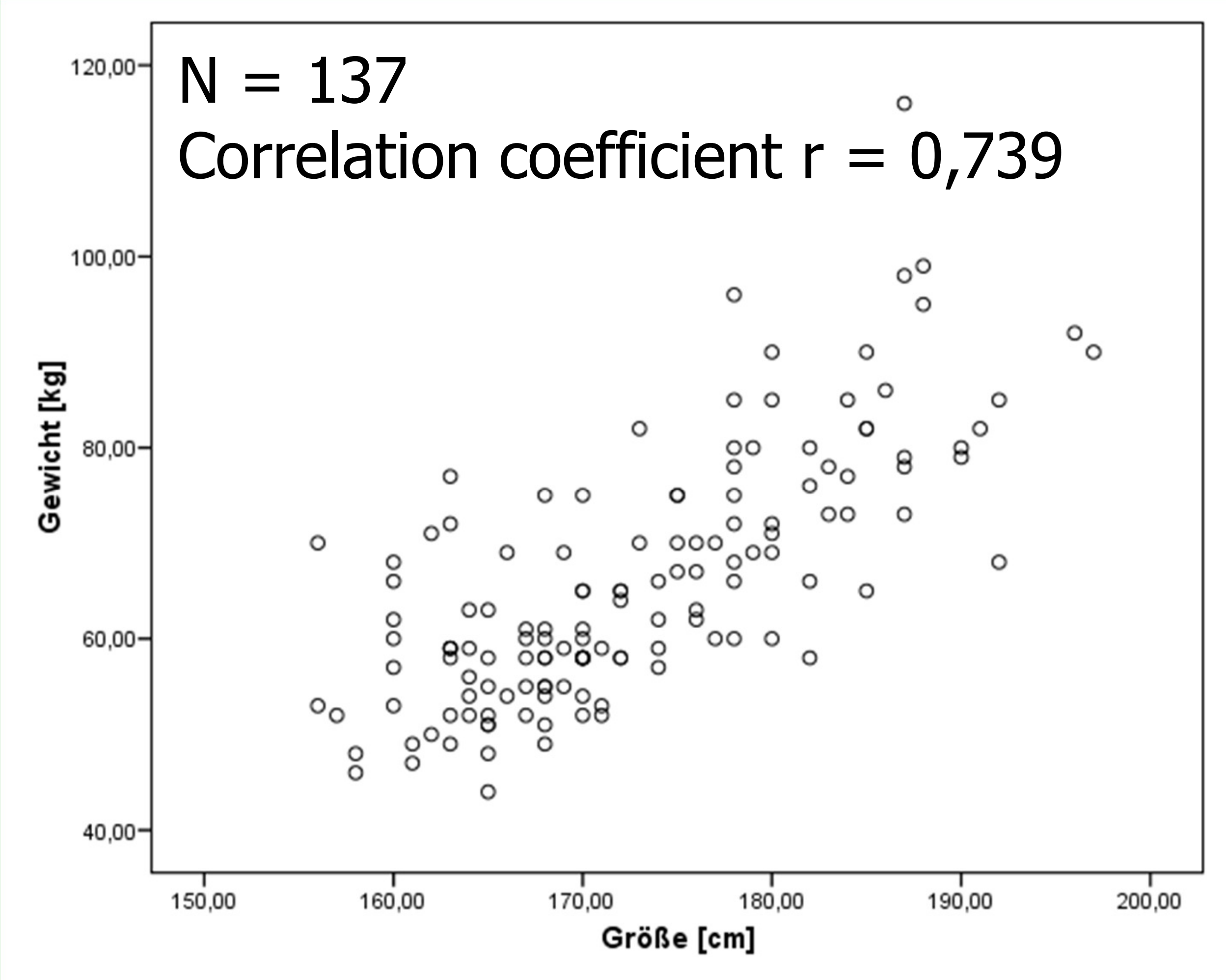
			Haarfarbe						Gesamt
			hellblond	dunkelblond	rot / rotblond	braun	schwarz	keine Haare	
Geschlecht	männlich	Anzahl	3	17	3	14	6	1	44
		% von Geschlecht	6,8%	38,6%	6,8%	31,8%	13,6%	2,3%	100,0%
		% von Haarfarbe	20,0%	30,4%	75,0%	25,5%	75,0%	100,0%	31,7%
		% der Gesamtzahl	2,2%	12,2%	2,2%	10,1%	4,3%	,7%	31,7%
	weiblich	Anzahl	12	39	1	41	2	0	95
		% von Geschlecht	12,6%	41,1%	1,1%	43,2%	2,1%	,0%	100,0%
		% von Haarfarbe	80,0%	69,6%	25,0%	74,5%	25,0%	,0%	68,3%
		% der Gesamtzahl	8,6%	28,1%	,7%	29,5%	1,4%	,0%	68,3%
Gesamt		Anzahl	15	56	4	55	8	1	139
		% von Geschlecht	10,8%	40,3%	2,9%	39,6%	5,8%	,7%	100,0%
		% von Haarfarbe	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
		% der Gesamtzahl	10,8%	40,3%	2,9%	39,6%	5,8%	,7%	100,0%

Break-down table

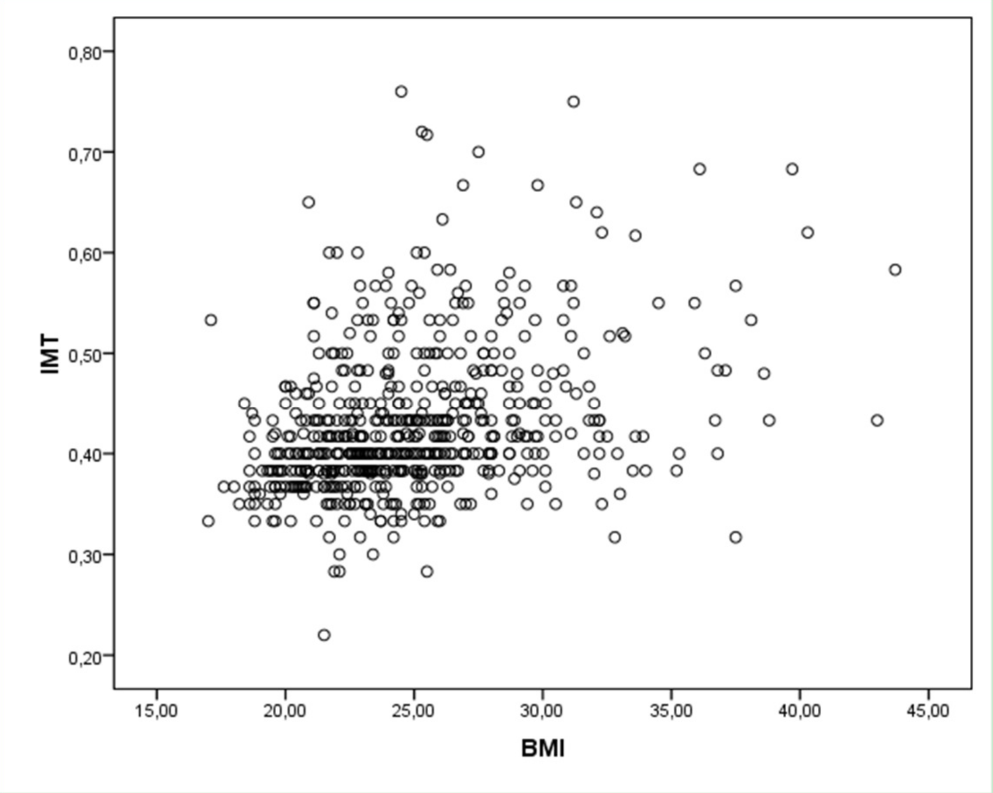
Deskriptive Statistik

Geschlecht		N	Minimum	Maximum	Mittelwert	Standardabweichung
männlich	Alter	44	20,00	30,00	23,0682	2,07306
	Größe	44	170,00	197,00	182,7727	6,30933
	Gewicht	44	58,00	116,00	79,8636	10,91317
	Gültige Werte (Listenweise)	44				
weiblich	Alter	95	19,00	32,00	22,7579	2,77801
	Größe	95	156,00	185,00	168,1474	6,18331
	Gewicht	93	44,00	80,00	59,4516	7,65778
	Gültige Werte (Listenweise)	93				

Scattergram (1)



Scattergram (2)



Korrelationen

		BMI	IMT
BMI	Korrelation nach Pearson	1	,339**
	Signifikanz (2-seitig)		,000
	N	581	580
IMT	Korrelation nach Pearson	,339**	1
	Signifikanz (2-seitig)	,000	
	N	580	580

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Correlation coefficient (CC)

- If both variables are (nearly) normally distributed you calculate the Pearson CC, else the Spearman CC.

Interpretation of CC

Sign positive: trend is positive
Sign negative: trend is negative

Absolute value	Interpretation (rule of thumb)
0 – 0,3	No correlation
0,3 – 0,6	Weak correlation
0,6 – 0,8	Moderate correlation
> 0,8	Strong correlation

Hypotheses

- Statistical hypothesis = assumption about a circumstance in the population
- Hypotheses are defined using the outcome variable and the clinical meaningfully (relevant) difference.
- Example: The mean reduction of drug A is 20 mmHg and of drug B 10 mmHg, i.e. the clinical meaningfully difference is 20 - 10 = 10 mmHg

Kind of hypotheses

- Null hypothesis H_0 = status quo / no difference / no change / no dependency (converse of the alternative hypothesis)
- Alternative hypothesis H_1 = possible innovation / issue to be proved (the study hypothesis) / difference / change / dependency (converse of null hypothesis)

Example H_0

- Null hypothesis: Drug A and drug B have the same effect, i.e.
  the mean reduction of blood pressure in the two groups is equal, i.e.
  $\mu_A = \mu_B$, i.e.
  $\delta = \mu_A - \mu_B = 0$

μ = true mean of blood pressure difference

Example H_1

- Alternative hypothesis: Drug A and drug B have different effects, i.e.
  the mean reduction of blood pressure in the two groups is not equal, i.e.
  $\mu_A \neq \mu_B$, i.e.
  $\delta = \mu_A - \mu_B \neq 0$

μ = true mean of blood pressure difference

One-sided hypotheses

- H_0: The mean reduction of blood pressure in group A is lower or equal as in group B, i.e.
  $\mu_A \leq \mu_B$, i.e. $\delta = \mu_A - \mu_B \leq 0$
- H_1: The mean reduction of blood pressure in group A is greater as in group B, i.e.
  $\mu_A > \mu_B$, i.e. $\delta = \mu_A - \mu_B > 0$

μ = true mean of blood pressure difference

Statistical test

- = statistical procedure to confirm or reject the null hypothesis
- The result is called statistically significant, if the null hypothesis is rejected.

Errors with statistical test

Result of test (based on sample data)	Population*	
	H_0 is true	H_1 is true
Test confirms H_0	✓	Type II error
Test rejects H_0	Type I error	✓

* We don‵t really know whether H_0 is true or false!

Type I error

- = probability of rejecting H_0 although H_0 is true.
- The type I error is controlled by the significance level α, i.e. α is the probability of making type I error.
- Usual values for α are 1% (0,01), 5% (0,05) or 10% (0,1).

Type II error

- = β = probability of confirming H_0 although H_0 is false.
- Power = 1 - β = probability of rejecting H_0 (= obtaining a „statistically significant“ result) when H_0 is truly false.

Type II error

- The type II error cannot be controlled because H_1 cannot be specified.
- Example H_1: Drug A and drug B have different effects, i.e. the mean reduction of blood pressure in the two groups is not equal, i.e. $\mu_A \neq \mu_B$, i.e. $\delta = \mu_A - \mu_B \neq 0$, but the true value of δ is unknown.

μ = true mean of blood pressure difference

Decision

- Comparison of p-value with the significance level α:
  - If $p > \alpha$: confirmation of H_0
  - If $p \leq \alpha$: rejection of H_0
- When you perform a statistical test with statistical software, the p-value will be calculated and printed.

General procedure

- Definition of outcome variable and hypotheses
- Choice of significance level
- Choice of appropriate test
- Performing the test with data
- Reading off the p-value and decision
- Interpretation of result

Example 1

- Does drug A reduce the systolic blood pressure for patients with hypertension?
- The average reduction of blood pressure by drug A is 10 mmHg.

Example 1: data structure

Variable →

Patient ↓

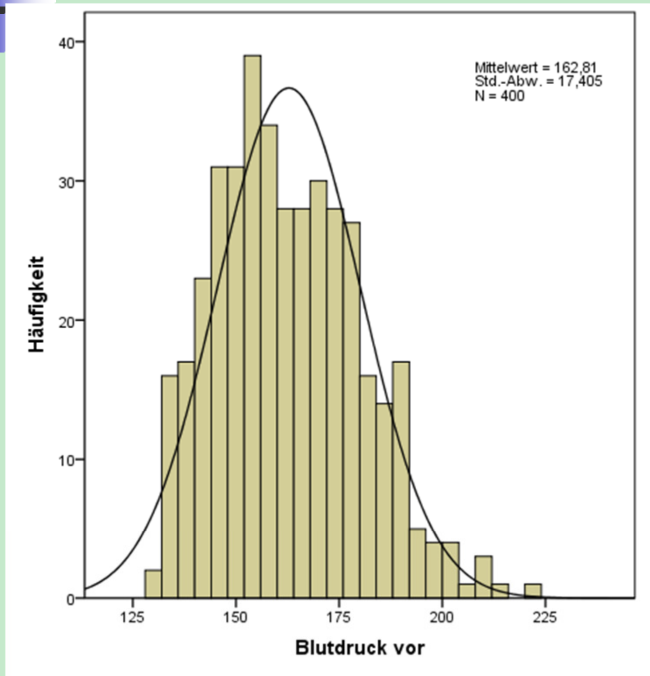
ID	Gender	Age	Size	Weight	SBP before	SBP after
1	m	63	180	93,0	160	140
2	m	72	183	79,7	150	145
3	f	83	165	78,0	170	172
4	m	74	175	90,5	160	130
5	m	52	176	72,4	190	180
6	f	61	165	64,0	150	155
7	f	71	173	83,0	165	145
8	m	79	180	92,3	185	175
9	m	65	177	66,5	170	175
10	...	...	...	...	...	...

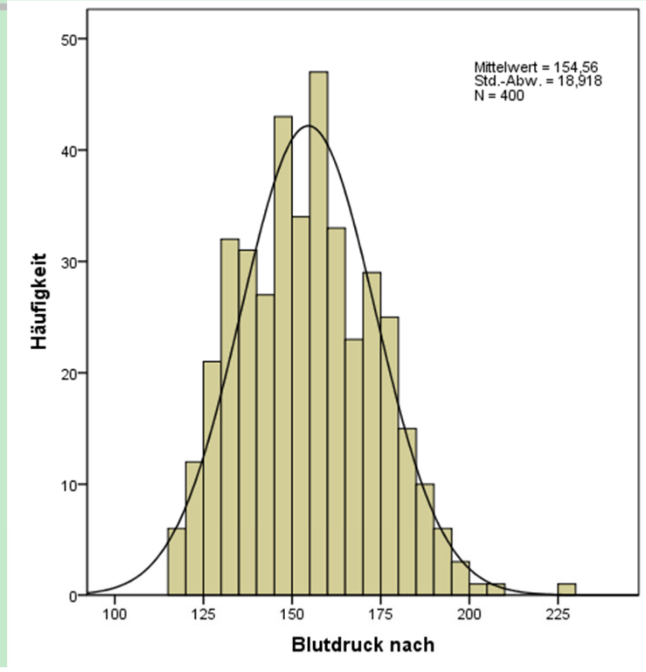
SBP = systolic blood pressure

Example 1: hypotheses

- H_0: „The mean of systolic blood pressure before and after treatment is equal." H_1: „The mean of systolic blood pressure before and after treatment is not equal."
- μ = true mean of blood pressure
- H_0: $\mu_{before} = \mu_{after}$, H_1: $\mu_{before} \neq \mu_{after}$
  oder
  H_0: $\mu_{before} - \mu_{after} = 0$,
  H_1: $\mu_{before} - \mu_{after} \neq 0$

Example 1: normal distribution check





Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von Blutdruck vor ist normal mit Mittelwert 162.81 und Standardabweichung 17.41.	Kolmogorov-Smirnov-Test einer Stichprobe	,126	Nullhypothese behalten.
2	Die Verteilung von Blutdruck nach ist normal mit Mittelwert 154.56 und Standardabweichung 18.92.	Kolmogorov-Smirnov-Test einer Stichprobe	,474	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

Example 1: t-test for paired samples

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Blutdruck vor	162,81	400	17,405	,870
	Blutdruck nach	154,56	400	18,918	,946

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
					95% Konfidenzintervall der Differenz				
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Untere	Obere			
Paaren 1	Blutdruck vor - Blutdruck nach	8,252	6,958	,348	7,568	8,936	23,718	399	,000

Mean of blood pressure difference p-value

Decision: $p < 0,05 \Rightarrow H_0$ is rejected

Example 1: Wilcoxon signed rank test

p-value

Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Der Medianwert der Unterschiede zwischen Blutdruck vor und Blutdruck nach ist gleich 0.	Wilcoxon-Vorzeichen-Rang-Test verbundener Stichproben	,000	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

Example 2

- There are differences between drug A and B in reducing the systolic blood pressure for patients with hypertension?
- The average reduction of blood pressure by drug A is 10 mmHg, by drug B 15 mmHg.

Example 2: data structure

Variable →

Patient ↓

ID	Treatment group	Gender	Age	Size	Weight	Systolic blood pressure before	Systolic blood pressure after	Difference
1	A	f	63	180	93,0	160	140	20
2	A	m	72	183	79,7	150	145	5
3	A	f	83	165	78,0	170	172	-2
…	…	…	…	…	…	…	…	…
61	B	f	61	165	64,0	150	155	-5
62	B	f	71	173	83,0	165	145	20
63	B	m	79	180	92,3	185	175	10
…	…	…	…	…	…	…	…	…

Example 2: hypotheses

- H_0: „The mean of blood pressure difference in both groups is equal."
  H_1: „The mean of blood pressure difference in both groups is not equal. "
- μ = true mean of blood pressure difference
- H_0: $\mu_{group\ A} = \mu_{group\ B}$
  H_1: $\mu_{group\ A} \neq \mu_{group\ B}$

Example 2: t-test for independent samples

Gruppenstatistiken

	Behandlungsgruppe	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Blutdruckdifferenz vor-nach	Medikament A	200	8,3890	6,93480	,49036
	Medikament B	200	8,1142	6,99578	,49468

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit					95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Blutdruckdifferenz vor-nach	Varianzen sind gleich	,001	,980	,395	398	,693	,27480	,69654	-1,09455	1,64415
	Varianzen sind nicht gleich			,395	397,969	,693	,27480	,69654	-1,09455	1,64415

p-value

Decision: $p > 0{,}05 \Rightarrow H_0$ is confirmed

Example 3: hypotheses

- Comparison of bone density in mice administrated with three different vitamin D concentrations
- μ = true mean of bone density
- H_0: $\mu_{dietgroup1} = \mu_{dietgroup2} = \mu_{dietgroup3}$
  H_1: $\mu_{dietgroup1} \neq \mu_{dietgroup2}$ or
  $\mu_{dietgroup1} \neq \mu_{dietgroup3}$ or
  $\mu_{dietgroup2} \neq \mu_{dietgroup3}$

Example 3: oneway analysis of variance

ONEWAY deskriptive Statistiken

Knochendichte [mg/mm^3]

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
100 IE Vit D/kg food	34	,6392	,04677	,00802	,6229	,6555	,56	,73
600 IE Vit D/kg food	36	,6572	,06181	,01030	,6363	,6781	,57	,76
2400IE Vit D/kg food	42	,7577	,04928	,00760	,7423	,7731	,64	,85
Gesamt	112	,6894	,07500	,00709	,6754	,7035	,56	,85

ONEWAY ANOVA

Knochendichte [mg/mm^3]

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	,319	2	,159	56,884	,000
Innerhalb der Gruppen	,306	109	,003		
Gesamt	,624	111			

p-value global test

Post-Hoc-Tests

Mehrfachvergleiche

p-values single tests

Bonferroni correction

Knochendichte [mg/mm^3]
Bonferroni

(I) Diätgruppe	(J) Diätgruppe	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
100 IE Vit D/kg food	600 IE Vit D/kg food	-,01796	,01266	,477	-,0487	,0128
	2400IE Vit D/kg food	-,11847*	,01221	,000	-,1482	-,0888
600 IE Vit D/kg food	100 IE Vit D/kg food	,01796	,01266	,477	-,0128	,0487
	2400IE Vit D/kg food	-,10051*	,01202	,000	-,1297	-,0713
2400IE Vit D/kg food	100 IE Vit D/kg food	,11847*	,01221	,000	,0888	,1482
	600 IE Vit D/kg food	,10051*	,01202	,000	,0713	,1297

*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Example 4

- Which treatment of varicosis* (stripping or ELT) is better?
- The infection rate of ELT is 3% and of stripping 15%.

* Varicose veins (commonly on the leg) are veins that have become enlarged and tortuous, because the leaflet valves to prevent blood from flowing backwards are insufficient.

Example 4: hypotheses

- H_0: „Infection rate and OP-method are independent."
  H_1: „Infection rate and OP-method are dependent."
- H_0: $P(\text{Infection}|\text{OP1})=P(\text{Infection}|\text{OP2})$
  H_1: $P(\text{Infection}|\text{OP1})\neq P(\text{Infection}|\text{OP2})$

OP1=ELT, OP2=Stripping
P(Infection|OP1) = (conditional) probability, that an infection occurs, if OP-method 1 was applied

Chi-squared test (χ^2 test)

- is a (non-parametric) test for categorical outcome variables resp. for dependencies in contingency tables (comparison of proportions).
- Requirement: the cell counts may not be to small, otherwise the Fisher exact test may be used.

Example 4: chi-squared test (1)

OP * Infektion Kreuztabelle

			Infektion		
			ja	nein	Gesamt
OP	ELT	Anzahl	51	267	318
		% innerhalb von OP	16,0%	84,0%	100,0%
	Stripping	Anzahl	36	204	240
		% innerhalb von OP	15,0%	85,0%	100,0%
Gesamt		Anzahl	87	471	558
		% innerhalb von OP	15,6%	84,4%	100,0%

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	,112[a]	1	,738	,814	,416
Kontinuitätskorrektur[b]	,047	1	,828		
Likelihood-Quotient	,112	1	,738	,814	,416
Exakter Test nach Fisher				,814	,416
Zusammenhang linear-mit-linear	,112[c]	1	,738	,814	,416
Anzahl der gültigen Fälle	558				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 37,42.

b. Wird nur für eine 2x2-Tabelle berechnet

c. Die standardisierte Statistik ist ,334.

Example 4: chi-squared test (2)

OP * Infektion Kreuztabelle

			Infektion		
			ja	nein	Gesamt
OP	ELT	Anzahl	51	267	318
		% innerhalb von OP	16,0%	84,0%	100,0%
	Stripping	Anzahl	17	204	221
		% innerhalb von OP	7,7%	92,3%	100,0%
Gesamt		Anzahl	68	471	539
		% innerhalb von OP	12,6%	87,4%	100,0%

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	8,237[a]	1	,004	,005	,003
Kontinuitätskorrektur[b]	7,498	1	,006		
Likelihood-Quotient	8,691	1	,003	,004	,003
Exakter Test nach Fisher				,004	,003
Zusammenhang linear-mit-linear	8,222[c]	1	,004	,005	,003
Anzahl der gültigen Fälle	539				

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 27,88.

b. Wird nur für eine 2x2-Tabelle berechnet

c. Die standardisierte Statistik ist 2,867.

General procedure

- Definition of outcome variable and hypotheses
- Choice of significance level
- Choice of appropriate test
- Performing the test with data
- Reading off the p-value and decision
- Interpretation of result

The choice of statistical test depends on

- Type and distribution of outcome variable
- Kind of hypothesis
- Number of groups
- Paired (related) or independent samples

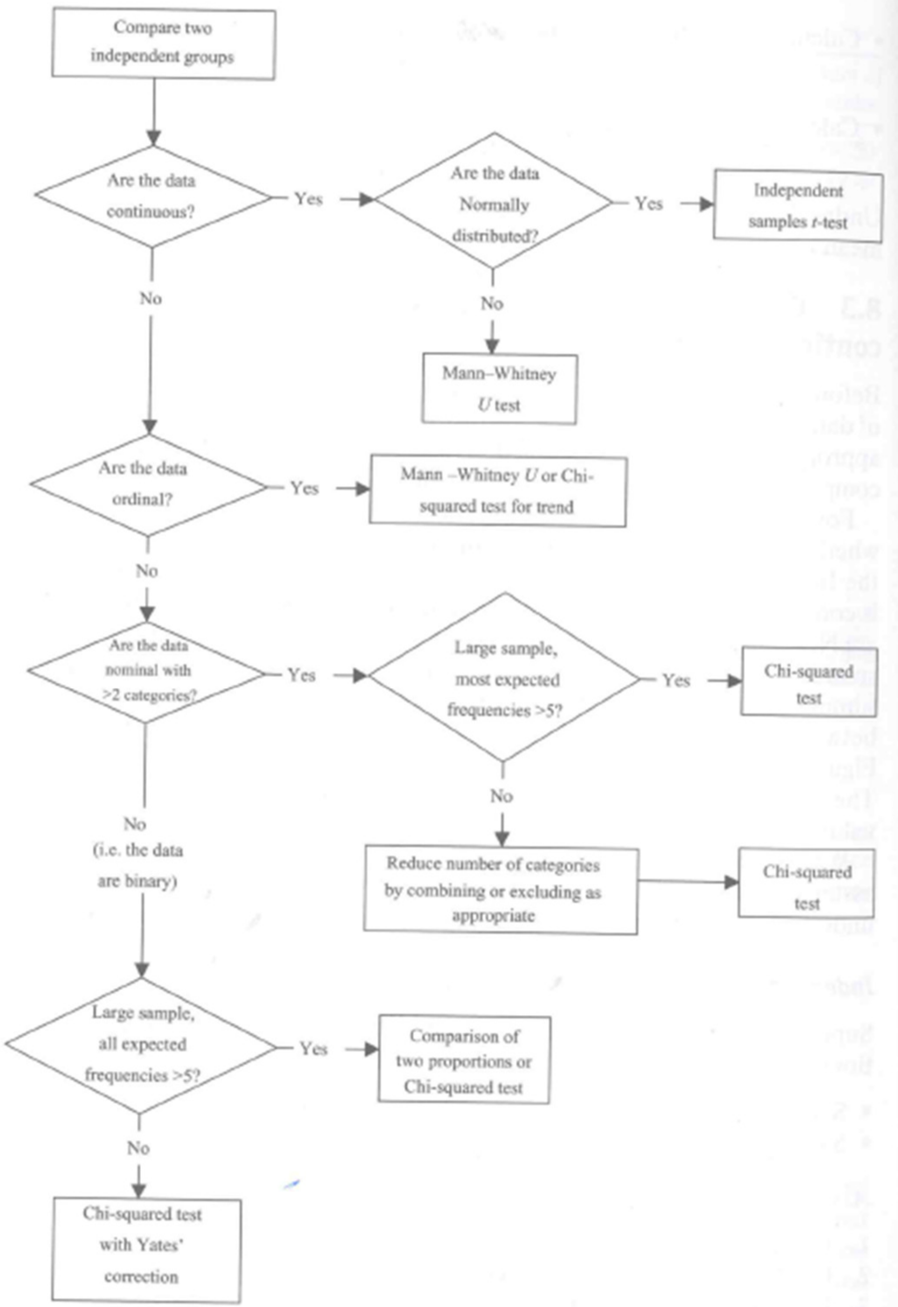


Figure 8.3 Statistical methods for comparing two independent groups or samples

Tests for comparing two or more groups of continuous data

	Outcome variable is normal distributed (parametric tests)	Outcome variable is not normal distributed (non-parametric tests)
Two independent groups	t-test for independent samples	Mann-Whitney U test
Two dependent groups	t-test for dependent (paired) samples	Wilcoxon signed rank test, sign test
Three and more independent groups	Analysis of variance for independent samples	Kruskal-Wallis test
Three and more dependent groups	Analysis of variance for dependent samples (repeated measurements)	Friedman test

Normal distribution check

- Comparing mean and median
- Interpretation of skewness
- Diagram (histogram, boxplot)
- Kolmogorov-Smirnov test

Example (1)

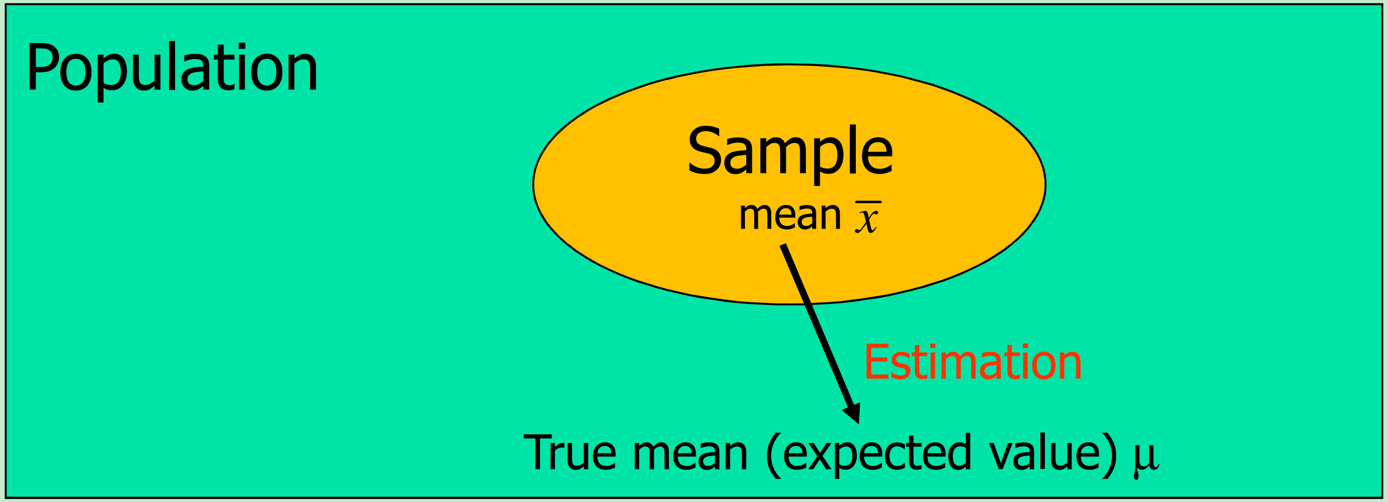
- Outcome variable is continuous and (approximately) normal distributed
- H_0: $\mu_A = \mu_B$, i.e. comparing means
- Two independent groups
- Appropriate statistical test: t-test for independent samples

Example (2)

- Explanatory and outcome variable are dichotomous
- H_0: Therapy and outcome variable are independent, i.e. comparing rates / proportions
- Appropriate statistical test: Chi-squared test

Parameter estimation

- Inference from the value of a parameter (summarizing measure based on data) to the (unknown) value in the population



A confidence interval (CI)

- consists of a lower and upper limit and describes the precision of estimation.
- The limits of CI include the true (but unknown) parameter value with a fixed confidence probability, e.g. 95%.
- The limits are calculated from data.
- Parameters are e.g. mean, odds ratio, correlation coefficient

Example CI

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
					95% Konfidenzintervall der Differenz				
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Untere	Obere			
Paaren 1	Blutdruck vor - Blutdruck nach	8,252	6,958	,348	7,568	8,936	23,718	399	,000

Mean of blood pressure difference

$$UG = \overline{x} - \frac{t_{n-1;0,975} \cdot s}{\sqrt{n}} \qquad OG = \overline{x} + \frac{t_{n-1;0,975} \cdot s}{\sqrt{n}}$$

Using CI for statistical test

- If H_0 has the form "Parameter has a defined value (e.g. $\mu = 0$)", one may perform a test by checking whether the CI for the parameter contains the defined value:
  - If the CI contains the defined value, H_0 will be confirmed.
  - If the CI does not contain the defined value, H_0 will be rejected.

Test using CI: example 1

H_0: true mean of blood pressure difference = 0

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
					95% Konfidenzintervall der Differenz				
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	Untere	Obere			
Paaren 1	Blutdruck vor - Blutdruck nach	8,252	6,958	,348	7,568	8,936	23,718	399	,000

The 95%-KI for the true mean of blood pressure difference does not contain the value 0, i.e. the null-hypothesis $\mu = 0$ is rejected at a significance level of 5%.

Test using CI: example 2

H_0: true odds ratio = 1

Täglicher Alkoholkonsum	Ösophagus-Karzinom ja	Ösophagus-Karzinom nein
≥ 80 g	96	109
< 80 g	104	666
Gesamt	200	775

Odds ratio: $OR = \frac{96 \cdot 666}{104 \cdot 109} = 5,64$ 95%-KI: [4,0;7,95]

The CI does not contain the value 1, i.e. a daily intake of more than 80 g alcohol is a statistically significant risk for esophageal carcinoma!

CI versus test

- A statistical test provides a decision pro or contra H_0, i.e. "Is there a statistically significant difference or not?"
- A CI provides a test decision and additionally information about the size of the difference!

Interpretation of statistical significance

- The number of tests affects the level of significance.
- If the sample size is very large, small differences may become significant.
- Significant results are not obvious clinically relevant.
- Significant results do not prove necessarily a causal correlation.