

Medizinische Statistik



ao. Univ.-Prof. Mag. Dr. Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
 Gesundheitsökonomie, Medizinische Universität Innsbruck

Statistik in medizinischen Top Journals



The American Statistician, February 2007, Vol. 61, No. 1

The Use of Statistics in Medical Research: A Comparison of *The New England Journal of Medicine* and *Nature Medicine*

Alexander M. STRASAK, Qunruz ZAMAN, Gerhard MARINELL, Karl P. PFEIFFER, and Hanno ULMER

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 4

Grundlagen



Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
 Gesundheitsökonomie, Medizinische Universität Innsbruck

Methodik der Fallstudie



- Alle Originalarbeiten publiziert im ersten Halbjahr 2004:
 - Vol. 350 No. 1–26 of NEJM
 - Vol. 10 No. 1–6 of NatMed
- wurde für die bibliometrische Analyse ausgewählt
- Editorials, Letters, Case Reports wurden nicht analysiert
- Zusätzlich wurden die Wiener klinische und die Wiener medizinische Wochenschrift untersucht:

Austrian Journal of Statistics
 Volume 36 (2007), Number 2, 141–152

**The Use of Statistics in Medical Research:
 A Comparison of Wiener Klinische Wochenschrift
 and Wiener Medizinische Wochenschrift**

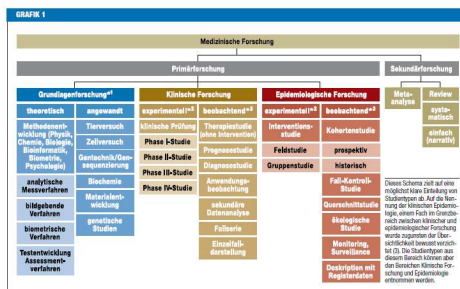
Alexander M. Strask¹, Qunruz Zaman¹, Gerhard Marinell¹,
 Karl P. Pfeiffer¹, and Hanno Ulmer²
¹Dept. of Medical Statistics, Informatics and Health Economics,
 Innsbruck Medical University, Austria
²Inst. of Statistics, University of Innsbruck, Austria

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 5

Medizinische Forschung



Unterschiedliche Versuchsdesigns
 * häufig synonym verwendet: Experimentelle Forschung, ** analoger Begriff: nicht interventionell, *** analoger Begriff: nicht interventionell/nicht experimentell

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 3

Statistische Verfahren



Kategorien nach Emerson/Colditz 1985	New England Journal of Medicine (n = 91)		Nature Medicine (n = 34)	
Types and Frequencies of Statistical Methods;	n	%	n	%
No statistical methods	2	2.2	1	2.9
Descriptive statistics only	3	3.3	5	14.7
Inferential methods	86	94.5	28	82.4
t-tests	32	35.2	14	41.2
Contingency table analysis				
Basic (χ^2 , Fisher's Exact test)	42	46.2	0	0.0
Advanced	6	6.6	0	0.0
Non-parametric tests	24	26.4	7	20.6
Analysis of Variance				
Basic (one-way ANOVA)	6	6.6	9	26.5
Advanced	6	6.6	1	2.9
Correlation coefficients	12	13.2	2	5.9
Regression				
Basic (single-linear regression)	4	4.4	1	2.9
Advanced	27	29.7	0	0.0
Epidemiologic methods	25	27.5	0	0.0
Survival Analysis	39	42.9	4	11.8
Other methods	15	16.5	5	14.7
Unidentified method/test	1	1.1	10	29.4
Confidence intervals	61	67.0	0	0.0

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 6

Nature Medicine versus New England Journal of Medicine



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- In über 95% der Originalarbeiten wurden statistische Methoden verwendet
- Statistische Methodik unterscheidet sich zwischen
 - Grundlagenwissenschaft (Nature Medicine)
 - Klinische Forschung (NEJM)
- Methodik komplexer in NEJM
- Keine Fallzahlschätzung bzw. Poweranalyse in NatMed
- Dokumentation der Methoden mangelhaft, fehlt fast völlig in Nature Medicine
- χ^2 Test, Überlebenszeitanalyse in NEJM, t-test in Nature Medicine

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 7

Quellen, Internet, Software



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- British Medical Journal Statistics at Square One
- Deutsches Ärzteblatt Bewertung wissenschaftlicher Publikationen
- Deutsche Medizinische Wochenschrift Statistik-Serie
- Jumbo Münster
- Graphpad Quickcalcs
- R, Stata, SPSS, SAS, Statistica, GraphPad, MedCalc
- nQuery Advisor, East, PASS, StudySize

09.01.2025

hanno.ulmer@i-med.ac.at

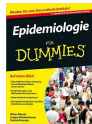
Seite 10

Literaturhinweise



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Grundlagenwissenschaft
- Klinische Forschung
- Epidemiologische Forschung



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 8

Häufig verwendete statistische Methoden



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Studienplanung
 - Fragestellung
 - Hypothesen
 - Studiendesign
 - Fallzahlschätzung
 - Datenerhebung
- Statistische Auswertung
 - Deskriptive Statistik
 - Inferenzstatistik I: Schätzen von Parametern mittels Konfidenzintervallen
 - Inferenzstatistik II: Unterschiede, Hypothesenprüfung mittels Signifikanztests
 - Inferenzstatistik III: Zusammenhänge, Korrelations- und Regressionsanalysen

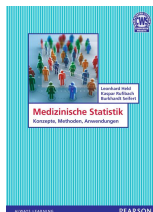
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 11

Prof. Dr. Leonhard Held / Prof. Dr. Burkhardt Seifert / Dr. Kaspar Rufibach Medizinische Statistik

Inhalte aus Kapitel 1



Prof. Dr. Leonhard Held / Prof. Dr. Burkhardt Seifert /
Dr. Kaspar Rufibach
Medizinische Statistik

ISBN 978-3-8689-4100-5
448 Seiten | 2-farbig
Juli 2013
€ 34,95 [D] | € 36,00 [A] | SFR 46,70

www.pearson-studium.de
www.pearson.ch

Beispielstudie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Didgeridoo playing as alternative treatment for obstructive sleep
apnoea syndrome: randomised controlled trial

W. A. D. J. van der Lely, M. A. J. van der Lely, M. A. J. van der Lely, M. A. J. van der Lely, M. A. J. van der Lely

Abstract Objective To assess the effect of didgeridoo playing on obstructive sleep apnoea syndrome (OSAS) in a randomised controlled trial.

Design Randomised controlled trial.

Setting General practice, North Devon, UK.

Participants 20 patients with OSAS, aged 45-65 years, who were referred to a general practitioner for treatment.

Intervention Didgeridoo playing for 15 minutes, 3 times a week, for 12 weeks.

Measurements and Main Results The primary outcome was the change in the apnoea-hypopnoea index (AHI) from baseline to 12 weeks.

Conclusion Didgeridoo playing for 15 minutes, 3 times a week, for 12 weeks, significantly reduced the AHI in patients with OSAS.

Keywords Didgeridoo, obstructive sleep apnoea, randomised controlled trial.

DOI: 10.1136/bmj-2013-076666

BMJ 2013;347:f1111

Copyright 2013 BMJ Group. All rights reserved. No reuse allowed without permission.

For further information on this trial, please visit the trial website: <http://www.bmj.com/lookup/doi/10.1136/bmj-2013-076666>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

For more information on this journal, please visit the journal website: <http://www.bmj.com/>

Beispielstudie



Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 13

Medizinproduktstudie

Onkoly & Neurology
00:00-00 © 2013, Onkoly & Neurology, Inc.

First European Multicenter Results With a New Transcutaneous Bone Conduction Hearing Implant System: Short-Term Safety and Efficacy

*Georg Sprinzl, †Thomas Lenarz, ‡Arneborg Ernst, §Rudolf Hagen,
*Astrid Wolf-Magele, †Hamidreza Mojjallal, ‡Ingo Todt, §Robert Mlynski,
and ||Mario D. Wolfmann

*Department of Otorhinolaryngology, Medical University Innsbruck, Austria; †Department of Otorhinolaryngology, Hannover Medical School, Germany; ‡Clinic for Ears, Nose, and Throat, Unfallkrankenhaus Berlin, Germany; §Department of Otorhinolaryngology, Plastic, Aesthetic and Reconstructive Head and Neck Surgery, University Clinic Würzburg, Germany; and ||Traum MED-EL, Innsbruck, Austria

hanno.ulmer@i-med.ac.at

16

„Wie lese ich eine Studie in 10 Minuten“

- Grundlegender Aufbau eines Studienberichts (papers):

IMRaD Schema

- Weitere Strukturierungen
- CONSORT Statement für Interventionsstudien (RCTs)
- STROBE Statement für Beobachtungsstudien (epidemiologische Studien: Kohortenstudie, Fall-Kontroll-Studie u. Querschnittstudie)
- STARD Statement für diagnostische Studien

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 14

Medizinproduktstudie

TABLE 1. Demographic data and medical parameter disease factors of the 12 study participants

Demographics				Disease factors and medical history				
Subject no.	Age at surgery	Sex	Study site	Implanted car	No. previous car surgeries	Duration of HL (yr)	Type of HL	Etiology
1	69	M	Berlin	R	2	60	CHL	Cholesteatoma
2	69	F	Berlin	R	4	60	CHL	Cholesteatoma
3	44	F	Berlin	R	2	9	Mixed	Otosclerosis
4	28	M	Hannover	R	2	15	CHL	COM
5	65	F	Hannover	R	1	2	CHL	Glioma tumor
6	65	F	Hannover	L	1	1	Mixed	Chronic mastoiditis
7	63	F	Hannover	L	3	22	Mixed	COM
8	35	M	Würzburg	R	5	35	CHL	Cholesteatoma
9	20	F	Innsbruck	L	2	20	CHL	Atresia auris
10	19	F	Innsbruck	R	2	19	Mixed	Cholesteatoma
11	28	F	Innsbruck	R	0	28	Mixed	Atresia auris
12	27	F	Innsbruck	R	1	27	CHL	Atresia auris

hanno.ulmer@i-med.ac.at

17

	Item number	Descriptor	Reported on page number
Title and abstract	1	How participants were allocated to interventions (eg, "random allocation", "randomized", or "randomly assigned").	
Introduction	2	Scientific background and explanation of rationale.	
Methods	3	Eligibility criteria for participants and the settings and locations where the data were collected.	
Participants	4	Baseline characteristics of the participants, including details of any restriction (eg, blocking, stratification).	
Interventions	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (eg, multiple observations, training of assessors, etc).	
Sample size	7	How sample size was determined (eg, when applicable, explanation of any interim analyses and stopping rules).	
Randomization	8	Method used to generate the random allocation sequence, including details of any restriction (eg, blocking, stratification).	
Sequence generation	9	Method used to implement the random allocation sequence (eg, numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Allocation concealment	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Implementation	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were aware of group assignment. If not, how the success of masking was assessed.	
Blinding (masking)	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.	
Statistical methods	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group, report the numbers of participants randomly assigned, reaching intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Results	14	Baseline demographic and clinical characteristics of each group.	
Participant flow	15	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat". State the results in absolute numbers when possible (eg, 10/20, not 50%).	
Recruitment	16	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (eg, 95% CI).	
Baseline data	17	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.	
Numbers analyzed	18	All important adverse events or side-effects in each intervention group.	
Outcomes and estimation	19		
Ancillary analyses	20		
Adverse events	21		
Discussion	22		
Interpretation	23		
Generalizability	24		
Overall evidence	25		

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 15

Medizinproduktstudie

Study Design

The study was a prospective, single-subject repeated-measures design, in which each subject served as his/her own control. Performance on audiometric tests preoperatively was compared with the aided 3 month postoperative condition using the Bonebridge. This type of design has been applied frequently to the evaluation of implantable hearing devices in multicenter clinical trials (35–37). It minimizes the effect of variability inherent to the population to the evaluation of treatment outcomes. Standardized evaluation methods were used to assure the reliability of the data across different investigational centers.

hanno.ulmer@i-med.ac.at

18



Statistical analyses were performed using IBM SPSS Statistics 19 (IBM, Armonk, NY, USA). One-way repeated-measure ANOVAs with time as factor were performed (significance was accepted at $p \leq 0.05$) and followed by post hoc pairwise comparisons to examine significant differences between the single test intervals. For each ANOVA, Mauchly's test of sphericity was applied. If sphericity could not be assumed, a Greenhouse-Geisser correction was used as part to the ANOVA. P-values of the pairwise comparisons were adjusted with the Holm-Sidak method. Box-Whisker Plots represent the whole data set. Whiskers extend to the maximum value within 1.5 times the interquartile range (IQR) above the third quartile or the minimum value within 1.5 times the IQR below the first quartile. Values outside this range are considered to be outliers, depicted as individual dots. Tukey box-whisker plots were generated using GraphPad Prism 5 (<http://www.graphpad.com>).

hanno.ulmer@i-med.ac.at

19

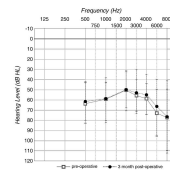


FIG. 5. Mean air conduction thresholds for the implanted ear: preoperative unaided testing compared with 3-month postoperative tests. Error bars represent ± 1 SD ($n = 12$).

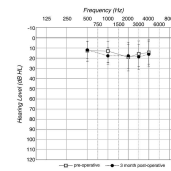


FIG. 6. Mean bone conduction thresholds for the implanted ear: preoperative unaided testing compared with 3-month postoperative tests. Error bars represent ± 1 SD ($n = 12$).

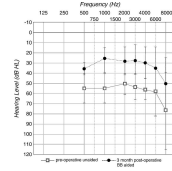


FIG. 7. Mean soundfield thresholds (binaural hearing) for the implanted ear: preoperative unaided testing compared with 3-month postoperative tests. Error bars represent ± 1 SD ($n = 12$).

TABLE 2. F statistics and p values from analysis of variance of audiometric tests (preoperative, 1-month postoperative and 3-month postoperative; $n = 12$)

	500 Hz	1 kHz	2 kHz	3 kHz	4 kHz	6 kHz	8 kHz
AC	$F_{(2,21)} = 0.284$	$F_{(2,21)} = 0.68$	$F_{(2,21)} = 0.55$	$F_{(2,21)} = 0.48$	$F_{(2,21)} = 0.52$	$F_{(2,21)} = 0.40$	$F_{(2,21)} = 0.38$
BC	$F_{(2,21)} = 0.41$	$F_{(2,21)} = 1.00$	$F_{(2,21)} = 0.38$	$F_{(2,21)} = 1.16$	$F_{(2,21)} = 0.35$	$F_{(2,21)} = 0.22$	$F_{(2,21)} = 0.25$
Soundfield	$F_{(2,21)} = 14.7$	$F_{(2,21)} = 40.0$	$F_{(2,21)} = 46.1$	$F_{(2,21)} = 40.0$	$F_{(2,21)} = 46.8$	$F_{(2,21)} = 40.0$	$F_{(2,21)} = 46.8$
	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$	$p = 0.000$

hanno.ulmer@i-med.ac.at

22

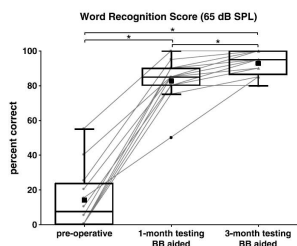


FIG. 3. Word recognition scores in quiet (Freiburger monosyllables) for the implanted ear: preoperative, 1-month postoperative and 3-month postoperative. Both postoperative scores are significantly improved from preoperative scores ($p < 0.0001$) and from each other ($p = 0.010$), $n = 12$, BB = Bonebridge.

hanno.ulmer@i-med.ac.at

20



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Placebo-Controlled Trial of Cytisine for Smoking Cessation

Robert West, Ph.D., Witold Zatonski, M.D., Magdalena Cedzynska, M.A.,
Dorota Lewandowska, Ph.D., M.D., Joanna Pazik, Ph.D., M.D.,
Paul Aveyard, Ph.D., M.D., and John Stapleton, M.Sc.

hanno.ulmer@i-med.ac.at

23

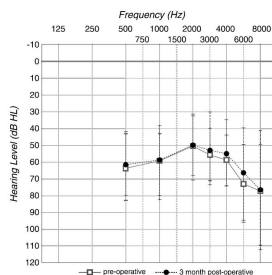


FIG. 5. Mean air conduction thresholds for the implanted ear: preoperative unaided testing compared with 3-month postoperative tests. Error bars represent ± 1 SD ($n = 12$).

hanno.ulmer@i-med.ac.at

21

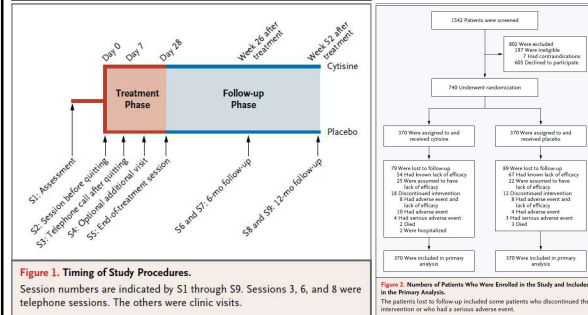


Figure 1. Timing of Study Procedures.

Session numbers are indicated by S1 through S9. Sessions 3, 6, and 8 were telephone sessions. The others were clinic visits.

hanno.ulmer@i-med.ac.at

24



Table 1. Characteristics of the Study Participants.*

Characteristic	Cytisine (N = 370)	Placebo (N = 370)
Male sex — no. (%)	183 (49.5)	161 (43.5)
Age — yr	47.8±12.6	48.5±12.6
Married — no. (%)†	190 (51.4)	207 (56.1)
Employment involving manual labor — no. (%)‡	196 (54.3)	178 (50.0)
Tried to stop smoking previously — no. (%)	307 (83.0)	301 (81.4)
No. of cigarettes smoked daily	23.0±8.7	22.5±9.6
Carbon monoxide in exhaled breath — ppm	19.2±8.7	18.2±9.0
Duration of smoking — yr	28.1±11.6	28.6±11.7
FTND score§	6.3±2.1	6.1±2.2
Beck Depression Inventory score¶	10.5±7.5	10.7±7.9

hanno.ulmer@i-med.ac.at

25



STATISTICAL ANALYSIS

With the use of previous trial data as a guide, we estimated that we would need to enroll 740 participants (370 in each group) to detect a between-group difference of 6 percentage points (6% vs. 12%) for the primary outcome, with 80% power and at an alpha level of 0.05.

The analyses of outcomes were based on the intention-to-treat principle, with treatment considered to have failed in participants who were lost to follow-up.²¹ The absolute percentage-point difference between participants who met the criteria for abstinence in the two groups was tested with the use of Fisher's exact test. The relative rate of abstinence (the percentage of patients in the cytisine group who met the abstinence criteria divided by the percentage in the placebo group) was also calculated. The 95% confidence interval was calculated for all measures. The relative rates and percentage-point differences were calculated for adverse events reported by 10 or more participants. Logistic regression was used to examine efficacy, with adjustment for baseline characteristics.

hanno.ulmer@i-med.ac.at

28



Table 2. Effect of Cytisine on Smoking Cessation.*

Outcome	Cytisine (N = 370)	Placebo (N = 370)	Percentage-Point Difference (95% CI)	Relative Rate (95% CI)‡
percent (number)				
Primary outcome: abstinence for 12 mo	8.4 (31)	2.4 (9)	6.0 (2.7–9.2)§	3.4 (1.7–7.1)
Abstinence for 6 mo	10.0 (37)	3.5 (13)	6.5 (2.9–10.1)‡	2.9 (1.5–5.3)
Point prevalence at 12 mo	13.2 (49)	7.3 (27)	5.9 (1.6–10.3)§	1.8 (1.2–2.8)

hanno.ulmer@i-med.ac.at

26

Deskriptive Statistik



Hanno Ulmer

hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck



Table 3. Adverse Events Reported by 10 or More Study Participants.*

Event	Cytisine (N = 370)	Placebo (N = 370)	Percentage-Point Difference (95% CI)†	Relative Rate (95% CI)‡
percent (number)				
Any gastrointestinal event	13.8 (51)	8.1 (30)	5.7 (1.2 to 10.2)§	1.7 (1.1 to 2.6)
Upper abdominal pain	3.8 (14)	3.0 (11)	0.8 (–1.8 to 3.4)	1.3 (0.6 to 2.8)
Nausea	3.8 (14)	2.7 (10)	1.1 (–1.5 to 3.6)	1.4 (0.6 to 3.1)
Dyspepsia	2.4 (9)	1.1 (4)	1.4 (–0.5 to 3.2)	2.2 (0.7 to 7.2)
Dry mouth	2.2 (8)	0.5 (2)	1.6 (0 to 3.3)	4.0 (0.9 to 18.7)
Any psychiatric event	4.6 (17)	3.2 (12)	1.4 (–1.4 to 4.2)	1.4 (0.7 to 2.9)
Dizziness	2.2 (8)	1.1 (4)	1.1 (–0.7 to 2.9)	2.0 (0.6 to 6.6)
Somnolence	1.6 (6)	1.1 (4)	0.5 (–1.1 to 2.2)	1.5 (0.4 to 5.3)
Any nervous system event	2.7 (10)	2.4 (9)	0.3 (–2.0 to 2.6)	1.1 (0.5 to 2.7)
Headache	1.9 (7)	2.2 (8)	–0.3 (–2.3 to 1.8)	0.9 (0.3 to 2.4)
Skin and subcutaneous tissue	1.6 (6)	1.4 (5)	0.3 (–1.5 to 2.0)	1.2 (0.4 to 3.9)

hanno.ulmer@i-med.ac.at

27

Deskriptive Statistik,
Merkmalstypen

DEFINITION 2.1 Qualitative Daten

Bei qualitativen Variablen werden die Individuen bestimmten Kategorien zugeordnet, z. B. (rot, grün, blau), (0, A, B, AB) oder (männlich, weiblich). Synonyme für „qualitativ“ sind *kategorisch*, *kategorial* oder *diskret*. Auch die Bezeichnung *Faktor* oder *Faktorvariable* ist gebräuchlich für eine qualitative Variable.

DEFINITION 2.2 Quantitative Daten

Quantitative Variablen können nur numerische Werte annehmen. Sie heißen deshalb auch *numerisch* oder *stetige* Daten.

DEFINITION 2.3 Zählraten

Zählraten sind spezielle diskrete Daten, die zählen, wie oft ein Ereignis aufgetreten ist. Zählraten können folglich die Werte 0, 1, 2, ... annehmen.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 30

Beispieldaten, BMJ 2005

Patient	Geschlecht (in Jahren)	Alter	Körpergröße (in m)	Behandlungsgruppe	Epworth-Index
1	Mann	45	1.74	Kontrollgruppe	11
2	Frau	41	1.70	Didgeridoo	11
3	Mann	41	1.80	Kontrollgruppe	5
4	Mann	39	1.79	Didgeridoo	8
5	Mann	33	1.80	Kontrollgruppe	1
6	Frau	55	1.67	Didgeridoo	9
7	Frau	50	1.65	Kontrollgruppe	10
8	Frau	65	1.64	Kontrollgruppe	11
9	Mann	43	1.76	Didgeridoo	7
10	Mann	59	1.85	Didgeridoo	9

Tabelle 2.1: Daten der ersten 10 Patienten der Didgeridoo-Studie.

Variable	Datentyp	Kategorien
Geschlecht	binär	Mann / Frau
Alter (in Jahren)	stetig	
Körpergröße (in m)	stetig	
Behandlungsgruppe	binär	Didgeridoo / Kontrolle
Epworth-Index	approximativ stetig	

Tabelle 2.2: Datentypen einiger Variablen der Didgeridoo-Studie.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 31

Didgeridoo playing as alternative treatment for obstructive sleep apnoea syndrome: randomised controlled trial
Milo A Puhon, Alex Suarez, Christian Lo Cascio, Alfred Zahn, Markus Heitz, Otto Braendli, BMJ 2005

Deskription quantitativer Daten

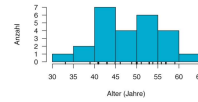


Abbildung 2.5: Histogramm des Alters der Patienten der Didgeridoo-Studie.

DEFINITION 2.5 Normalverteilung

Die drei Haupteigenschaften der Normalverteilung sind:

- Symmetrie, d.h., Abweichungen nach links und nach rechts vom Zentrum sind gleich häufig und gleich groß.
- „Große“ Abweichungen vom Zentrum sind selten.
- Die Normalverteilung lässt sich durch lediglich zwei Parameter eindeutig beschreiben: den Mittelwert und die Standardabweichung.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 34

Deskription qualitativer Daten

	abs. Häufigkeit	rel. Häufigkeit (%)
Didgeridoo	14	56
Kontrollgruppe	11	44
Total	25	100

Tabelle 2.4: Anzahl Patienten in den Behandlungsgruppen der Didgeridoo-Studie.

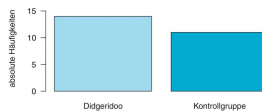


Abbildung 2.1: Balkendiagramm der Behandlungsfrequenzen.

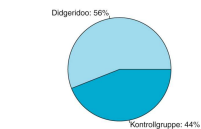


Abbildung 2.2: Kuchendiagramm der Behandlungsfrequenzen.

	Didgeridoo	Kontrollgruppe	Total
Frauen	12	9	21
Männer	2	2	4
Total	14	11	25

Tabelle 2.5: Anzahl Patienten in den Behandlungsgruppen der Didgeridoo-Studie nach Geschlecht.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 32

Deskription quantitativer Daten

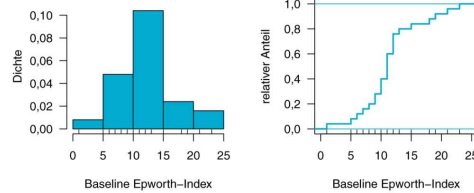


Abbildung 2.10: Histogramm und empirische Verteilungsfunktion der Baseline Epworth-Index Daten.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 35

Deskription quantitativer Daten

Intervall	(30, 35]	(35, 40]	(40, 45]	(45, 50]	(50, 55]	(55, 60]	(60, 65]
Anzahl Werte im Intervall	1	2	7	4	6	4	1
Relativer Anteil	4 %	8 %	28 %	16 %	24 %	16 %	4 %

Tabelle 2.3: Verteilung des Alters in der Didgeridoo-Studie.

DEFINITION 2.14 Histogramm

Mit einem Histogramm wird die Verteilung von stetigen Daten visualisiert. Dazu wird der Bereich der Daten in gleiche, anliegende aber sich nicht überlappende Intervalle (Zellen, Klassen) zerlegt. Dann zählt man die Anzahl der Beobachtungen in jedem Intervall und erstellt ein Balkendiagramm.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 33

Statistische Kennwerte

DEFINITION 2.6 Mittelwert

Der Mittelwert (engl. mean) von n Beobachtungen x_1, x_2, \dots, x_n ist

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

DEFINITION 2.7 Perzentil

Eine Zahl p_k heißt k -tes Perzentil (k bezeichnet hierbei eine ganze Zahl zwischen 1 und 99) einer Variablen, wenn mindestens k % der Beobachtungen der Variable kleiner oder gleich p_k und mindestens $(100 - k)$ % größer oder gleich p_k sind.

DEFINITION 2.12 Interquartilsabstand

Der Interquartilsabstand (engl. interquartile range, IQR) ist die Differenz zwischen dem 75. und dem 25. Perzentil.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 36

Statistische Kennwerte



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

DEFINITION 2.13 Spannweite

Die Spannweite (engl. range) ist die Differenz zwischen Maximum und Minimum und gibt den Bereich an, in dem die Daten liegen.

DEFINITION 2.8 Varianz

Die Varianz (engl. variance) ist die mittlere quadratische Abweichung der Beobachtungen vom Mittelwert:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{(n-1)}$$

DEFINITION 2.9 Standardabweichung

Die Standardabweichung (SD) ist die Wurzel aus der Varianz:

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

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 37

Vergleiche mit Boxplot



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

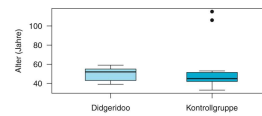


Abbildung 2.9: Boxplots des Alters aller Patienten nach Behandlungsgruppe mit den fehlerhaften Daten aus Abbildung 2.3 rechts.

DEFINITION 2.15 Boxplot

Die „Box“ im **Boxplot** (engl. box and whiskers plot) gibt den Bereich vom 25. zum 75. Perzentil an, der horizontale Strich in der Box den Median. Die Stäbe (engl. whiskers), die aus der Box herausragen, sind nicht einheitlich definiert. Bei einfachen Boxplots reichen sie zum Minimum und zum Maximum. Eine verbreitete Definition, die auf John W. Tukey zurückgeht, besteht darin, die Länge der Whiskers auf maximal das 1,5-fache der Boxlänge zu beschränken. Beobachtungen außerhalb dieses Bereichs werden als „Ausreißer“ gekennzeichnet.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 40

Streuung



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

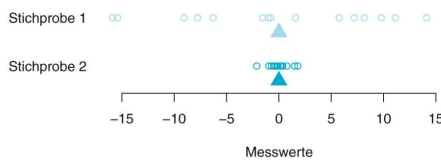


Abbildung 2.4: Zwei Stichproben mit demselben Mittelwert, aber verschiedener Streuung.

DEFINITION 2.11 Variationskoeffizient

Der **Variationskoeffizient** (CV) ist der Quotient von Standardabweichung und Mittelwert.

$$CV = s/\bar{x}$$

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 38

Deskriptive Vergleiche Didgeridoo Studie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

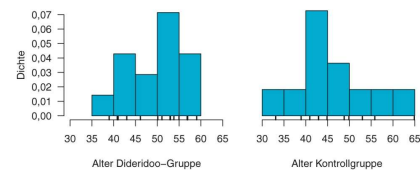


Abbildung 2.6: Histogramme des Alters aller Patienten in der Didgeridoo-Studie, nach Behandlungsgruppe.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 41

Boxplot



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

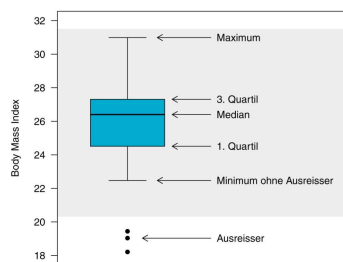


Abbildung 2.7: Boxplot des Body-Mass-Index in der Didgeridoo-Studie. Das graue Band beschreibt den Bereich 1. Quartil $-1,5 \cdot IQR$ bis zum 3. Quartil $+1,5 \cdot IQR$.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 39

Deskriptive Vergleiche Didgeridoo Studie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

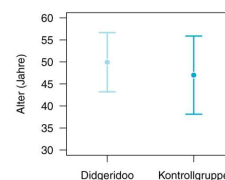


Abbildung 2.12: Fehlerbalken-Diagramm des Alters nach Behandlungsgruppe (Mittelwert \pm Standardabweichung).

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 42

Deskriptive Vergleiche Didgeridoo Studie

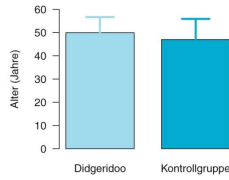


Abbildung 2.11: Balkendiagramme des Alters nach Behandlungsgruppe (Mittelwert \pm Standardabweichung).

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 43

Korrelationskoeffizient

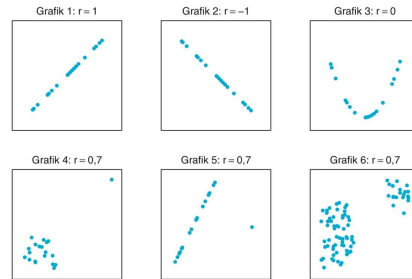


Abbildung 6.3: Illustration von speziellen Korrelationen.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 46

Streudiagramm und Korrelation

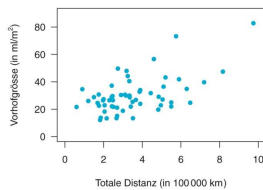


Abbildung 6.1: Streudiagramm der Größe des linken Vorhofs gegen die totale Distanz.

Variable	n	Min.	Median	\bar{x}	Max.	s
Totale Distanz (in 100 000 km)	61	0,6	3,1	3,4	9,8	1,8
Vorhofgröße (in ml/m²)	61	12,1	27,1	30,2	82,8	13,0

Tabelle 6.1: Deskriptive Statistiken in der Rennfahrerstudie.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 44

Regression

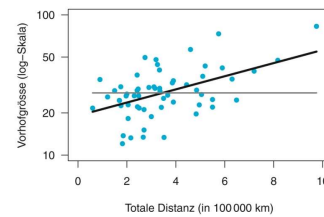


Abbildung 6.9: Streudiagramm der logarithmierten Größe des linken Vorhofs gegen die totale Distanz mit Regressionsgerade.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 47

Korrelationskoeffizient

DEFINITION 6.1 Pearson-Korrelation

Die Korrelation einer Stichprobe von n Paaren von Beobachtungen $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ist

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \cdot s_y} \quad (6.1)$$

wobei

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{und} \quad s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (6.2)$$

die geschätzten Standardabweichungen der Variablen x und y bezeichnen,

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

die geschätzte Kovarianz zwischen beiden Variablen.

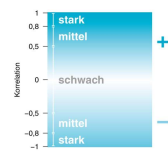


Abbildung 6.2: Die Bewertung von Korrelationskoeffizienten.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 45

Deskriptive Statistik

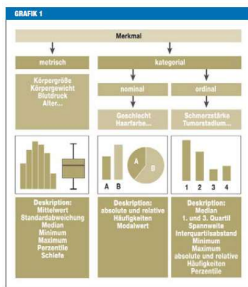
Qualitative Merkmale	Quantitative Merkmale normalverteilt	beliebig verteilt
n, %	Arithmetischer Mittelwert	Median
	Standardabweichung	Perzentile
Kreis-, Balkendiagramm	Histogramm	Boxplot
Beispiele:		
Geschlecht, Ansprechen auf Therapie, etc.	Blutdruck	CRP, GGT Schweregrade

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 48

Merkmalstypen und statistische Maßzahlen, Grafiken



09.01.2025

hanno.ulmer@i-med.ac.at

Table 1. Baseline Characteristics of Patients in the Intention-to-Treat Population.*

Characteristic	Cisplatin (N=204)	Fluorouracil plus Leucovorin (N=493)
Sex (%)		
Male	54	54
Female	46	46
Age (yr)		
Median	62	63
Range	25-80	22-82
Age group (%)		
<70 yr	81	79
≥70 yr	19	21
ECOG performance score (%)		
0	85	85
1	15	15
Neutrophils - (%)		
N1	69	71
N2	31	29
Tumor stage (%)		
T1 or 2	10	10
T3	76	76
T4	14	14
Carcinembryonic antigen level (%)		
<4 U/L	83	85
≥4 U/L	9	7
Missing data	8	8

Mittelwert vs. Median - Beispiel

Wann verwendet man den Mittelwert, wann den Median?

Daten: 12
14
16
18
20

Mittelwert: $\frac{12+14+16+18+20}{5} = 16$
Median: mittlerer Wert der Rangliste: 16

→ Bei Vorliegen einer Normalverteilung sind Mittelwert und Median gleich.

Daten mit einem Ausreißer: 12
14
16
18
20
40

Mittelwert: $\frac{12+14+16+18+20+40}{6} = 20$
Median: mittlerer Wert der Rangliste – bei einer geraden Anzahl an Werten wird der Mittelwert der beiden mittleren Werte berechnet → $\frac{16+18}{2} = 17$

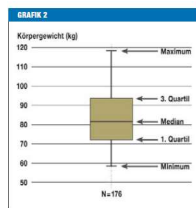
→ Bei Vorliegen von nicht normalverteilten Daten sind Mittelwert und Median nicht gleich.

→ Median ist robust gegen Ausreißer, Mittelwert nicht.

hanno.ulmer@i-med.ac.at

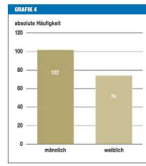
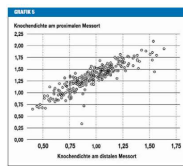
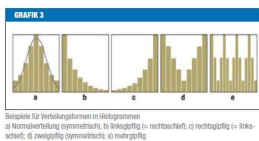
52

Statistische Grafiken



09.01.2025

hanno.ulmer@i-med.ac.at

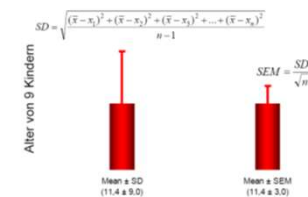


Seite 50

Standardabweichung oder Standardfehler?

- Standardfehler beschreibt nicht die Daten, sondern gibt die Genauigkeit des Mittelwerts als Schätzwerts an.

$SD > SEM$



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 53

Deskriptive Statistik

KASTEN

Mittelwert	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Varianz	$\text{Var} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Standardabweichung	$s = \sqrt{\text{Var}}$
Schiefe	$g = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$
Median	$\tilde{x} = x_{(n+1)/2}$ falls n ungerade $\tilde{x} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$ falls n gerade
Spannweite	$R = x_{\max} - x_{\min}$

mit n = Stichprobenumfang beziehungsweise Fallzahl
 x_i = Messwert für i-tes Stichprobenelement
beziehungsweise -ten Fall, wobei $i = 1, \dots, n$
 $x_{(i)}$ = bezeichnet den i-ten Wert in der aufsteigend geordneten Reihe der Messwerte, wobei $i = 1, \dots, n$

09.01.2025

hanno.ulmer@i-med.ac.at

Median oder Mittelwert?

St. Lange¹, R. Bender²
¹ Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr Universität Bochum
² Fakultät für Gesundheitswissenschaften, AC (epidemiologische und medizinische Statistik, Universität Bielefeld)

Tab.2 Übersetzungen (deutsch - englisch)

(arithmetischer) Mittelwert	(arithmetisch) mean
Median	Median
Ausreißer	Outlier
Stichprobenumfang	Sample size
schiefe Verteilung	Skewed distribution
zensierte Daten	Censored data

Variabilitätsmaße

St. Lange¹, R. Bender²
¹ Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr Universität Bochum
² Fakultät für Gesundheitswissenschaften

Tab.1 Übersetzungen (deutsch - englisch)

Spannweite	range
Standardabweichung	standard deviation
Varianz	variance
Standardfehler des Mittelwertes	standard error of the mean
Variabilitätsmaß	measure of variability
Spannweite	range
Interquartilsabstand	interquartile range
Summe der Abweichungsquadrate	sum of squares

Übung: Statistische Maßzahlen

- Berechnen Sie bitte den arithmetischen:
 - Mittelwert
 - Median
 - Varianz
 - Standardabweichung
 - Spannweite
 - Interquartilsabstand
 - und den Variationskoeffizient

Erstellen Sie Histogramme und Boxplots für:

17.11.2010

54

- Punktezahlen von 20 Studenten
6,3,7,5,6,4,4,6,7,3,5,9,6,4,2,7,5,5,8,6
- Anzahl der Angestellten in 20 Apotheken
2,3,3,3,4,4,4,4,5,5,5,5,5,6,6,6,8,10,15
- Krankheitstage von 20 Personen
0,0,0,0,0,0,0,1,1,1,1,1,2,2,2,3,3,15,76

17.11.2010

55

Links, survey tools

- MSIG services
<https://redcap.i-med.ac.at/>
<https://sosci.i-med.ac.at/admin/>
<https://grafstat.i-med.ac.at/grafstat20/index.htm>
- eCRF MUI
<https://www.askimed.com/>
- Google Forms
<https://www.google.com/forms/about/>
- LimeSurvey
<https://www.limesurvey.org/>
- MedUni Graz und Wien
<https://clincase.com/>

Questionnaires, Fragebogen

ao. Univ.-Prof. Dr. Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Interview

Basiert auf einem ausgearbeitetem Fragebogen. Fragen, Antworten und erläuternder Text sollten so vorgelesen werden, wie sie auf dem Papier stehen.

Vorteile:

- Hohe Antwortrate
- Hohe Komplexität möglich
- Selektive Nichtbeantwortung einzelner Fragen gering
- Lange Interviewdauer möglich

Nachteile:

- Teuer
- *Social Desirability Bias* groß
- Einfluss durch den Interviewer
- Mangel an kompetenten Interviewern
- Interviewerschulung ist erforderlich



Fragebogen

sind wesentlicher Bestandteil von Studien:

- in Meinungsumfragen, mit unterschiedlichen Zielgruppen (Kunden, PatientInnen, ExpertInnen etc.)
- in Surveys wie z.B. Gesundheitsstudien
- in epidemiologischen Studien wie Fall-Kontroll oder Kohortenstudien
- in klinischen Verlaufs- und Interventionsstudien (RCTs)
- in Register und Biobankstudien

Fragebogen werden heutzutage meistens über Internet appliziert, klassische Papierform und Telefoninterviews sind selten geworden

Telefon (chat) Interview

Mischform, angesiedelt zwischen Fragebogenerhebung und persönlichem Interview

Vorteile:

- Geringe Kosten
- *Social Desirability Bias* geringer
- Keine Beeinflussung durch Außenstehende
- Selektive Nichtbeantwortung einzelner Fragen gering

Nachteile:

- Komplexität geringer als bei persönlichem Interview
- „Offene“ Fragen können seltener gestellt werden
- Ablehnung einfacher wie bei persönlichen Interview

Fragebogen = Messgerät



Prinzipiell beliebig gestaltbar, allerdings sind gewisse formale und inhaltliche Regeln hilfreich

Fragebogen=Messgerät

- 1) Jeder sollte alle Fragen verstehen
- 2) Jeder sollte immer alle Fragen beantworten

Die Erstellung eines Fragebogens ist zeitintensiv und erfordert in der Regel die Zusammenarbeit von verschiedenen ExpertInnen

Formale Regeln:



- Zu Beginn des Textes soll eine kurze erklärende Einleitung stehen.
- Es soll ersichtlich sein, wer diese Studie durchführt.
- Es soll darauf hingewiesen werden, dass Daten vertraulich und anonym behandelt werden.
- Im Wesentlichen unterscheidet man zwischen drei Textformen: den Fragen, den Antworten und dem erklärenden bzw. überleitenden Text
- Die drei Textformen sollten sich im Schriftformat unterscheiden: zum Beispiel
 - alle Fragen Times New Roman (fett)
 - alle Antworten Arial
 - erklärender/überleitender Text Courier (kursiv).

Inhaltliche Regeln:



- Nicht zu viel vom Befragten annehmen/voraussetzen bzw. verlangen, das verunsichert den Befragten.
- Die Frage muss gut verständlich sein.
- Unklare Fragen- bzw. Antwortkonzepte („viel“, „wenig“) sollten vermieden werden.
- Wenn möglich, nur „geschlossene“ Fragen verwenden, es ist also nur eine limitierte Anzahl von Antworten möglich und vorgegeben.
- Offene Fragen nur für einfache Konzepte mit sehr vielen Antwortmöglichkeiten, wie z.B. die Frage nach dem Lebensalter.

Formale Regeln:



- Antwortmöglichkeiten sollten immer untereinander, also vertikal, aufgelistet werden. Nicht aus Platzgründen horizontale Auflistung verwenden – zu unübersichtlich.
- Die Fragen sollten fortlaufend durchnummeriert sein, ebenso die Seiten des Fragebogens.
- Die Antworten sollten auch nummeriert sein. Das erleichtert die Eingabe, wenn die Antworten nicht von einem Computer eingelesen werden.
- Die Fragen sollten, wenn möglich, eine logische Reihenfolge haben.
- Besser mit interessantem Thema starten, auch wenn man es gewohnt ist, mit „langweiligen“ demographischen Fragen (Alter, Geschlecht...) zu starten.
- Am Ende des Textes sollte der Befragte die Möglichkeit bekommen, „sonstige Anmerkungen“ als Freitext zu hinterlassen.
- Zuletzt sollte man sich für die Teilnahme bedanken und Kontaktdaten für etwaige Fragen angeben.

Inhaltliche Regeln:



- Antwortmöglichkeiten sollten allumfassend sein, das heißt, es sollte fast immer eine Kategorie „anderes“ oder „unbekannt“ geben.
- Jede Frage sollte jeweils nur ein Konzept beinhalten. Das erreicht man am besten, indem man Fragen, die „und“ bzw. „oder“ enthalten, vermeidet.
- „Bedrohliche“ Fragen erfordern ein spezielles Vorgehen. Als bedrohlich werden zum Beispiel Fragen nach dem Alkoholkonsum, dem Einkommen oder dem Sexualverhalten empfunden. Die Frage nach dem Einkommen sollte nicht als offene Frage gestellt werden, sondern in Form von Antwortmöglichkeiten in breiten Kategorien angeboten werden. Also nicht: „Wie hoch ist ihr monatliches Bruttogehalt?“, sondern „Ist Ihr monatliches Bruttogehalt (1) <750€, (2) 750-1500€, (3) 1500-2250€ ...“ Fragen nach Verhaltensformen, die von der Gesellschaft nicht akzeptiert sind, sollten mit erklärendem Begleittext „geladen“ und so abgeschwächt werden. Wenn bedrohliche Fragen verwendet werden müssen, empfiehlt es sich, Beispiele in der Literatur zu suchen.

Testung



Zuerst Fragebogen TESTEN:

- an Kollegen
- an nicht medizinisch ausgebildetem Personal
- an für die Zielgruppe repräsentativen Testpersonen

TIPP: Fragebögen verwenden, die bereits getestet und eventuell sogar vielfach angewandt wurden.
z.B. SF-36, WHOQOL, EORTC QLQ

Systematische Literatursuche, um zu erheben, ob es bereits ein geeignetes Instrument gibt.

Case Study 1

- Data example: BR23_Beispielsav
- Syntax: BR23_Beispielsps
- For scoring details and programming codes please see EORTC Scoring Manual
<https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf>
- QLQ-BR23 Scores

BRBI	Body image	BRCT	Systemic therapy
BRSEF	Sexual functioning	BRRT	Breast symptoms
BRSEE	Sexual enjoyment	BRSY	Arm symptoms
BRFU	Future perspective	BRHL	Hair loss
- A higher score represents a higher ("better") level of functioning, or a higher ("worse") level of symptoms.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 73

Case study 2: CINDI survey

C. RAUCHEN

1. Wie oft rauchen Sie Zigaretten?

1. nie
2. selten
3. oft

2. Wie viele Zigaretten rauchen Sie durchschnittlich pro Tag?

1. nie
2. weniger als 10
3. 10 bis 19
4. 20 bis 29
5. 30 bis 39
6. 40 bis 49
7. 50 oder mehr

D. ALKOHOLKONSUM

1. Wie oft trinken Sie Alkohol?

1. nie
2. selten
3. oft

2. Wie viel Alkohol trinken Sie durchschnittlich pro Woche?

1. nie
2. weniger als 100 ml
3. 100 bis 200 ml
4. 200 bis 300 ml
5. 300 bis 400 ml
6. 400 bis 500 ml
7. 500 ml oder mehr

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 76

Case Study 2: CINDI survey

WHO **Scoring Manual**

1. Wie oft rauchen Sie Zigaretten?

1. nie
2. selten
3. oft

2. Wie viele Zigaretten rauchen Sie durchschnittlich pro Tag?

1. nie
2. weniger als 10
3. 10 bis 19
4. 20 bis 29
5. 30 bis 39
6. 40 bis 49
7. 50 oder mehr

WHO **Scoring Manual**

1. Wie oft trinken Sie Alkohol?

1. nie
2. selten
3. oft

2. Wie viel Alkohol trinken Sie durchschnittlich pro Woche?

1. nie
2. weniger als 100 ml
3. 100 bis 200 ml
4. 200 bis 300 ml
5. 300 bis 400 ml
6. 400 bis 500 ml
7. 500 ml oder mehr

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 74

Case Study 2: CINDI survey

- Fragebogen in SPSS codieren (Variablenamen, Labels, etc.)
- Dateneingabe
- Plausibilitätskontrollen
- Deskriptive Auswertungen
- Mehrfachantworten
- Kreuztabellen
- Signifikanztests

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 77

Case study 2: CINDI survey

WHO **Scoring Manual**

1. Wie oft rauchen Sie Zigaretten?

1. nie
2. selten
3. oft

2. Wie viele Zigaretten rauchen Sie durchschnittlich pro Tag?

1. nie
2. weniger als 10
3. 10 bis 19
4. 20 bis 29
5. 30 bis 39
6. 40 bis 49
7. 50 oder mehr

WHO **Scoring Manual**

1. Wie oft trinken Sie Alkohol?

1. nie
2. selten
3. oft

2. Wie viel Alkohol trinken Sie durchschnittlich pro Woche?

1. nie
2. weniger als 100 ml
3. 100 bis 200 ml
4. 200 bis 300 ml
5. 300 bis 400 ml
6. 400 bis 500 ml
7. 500 ml oder mehr

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 75

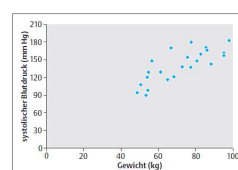
Inferenzstatistik III: Zusammenhänge, Korrelations- und Regressionsanalysen

(Lineare) Regression/Korrelation

St. Lange¹, R. Bender²

¹ Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr-Universität Bochum
² Fakultät für Gesundheitswissenschaften, AG Epidemiologie und medizinische Statistik, Universität Bielefeld

Neben der univariaten, das heißt auf ein einzelnes Merkmal bezogenen Analyse von Daten aus einer klinischen Studie, ist man häufig daran interessiert, den Zusammenhang zwischen zwei (bivariat) oder mehreren (multivariat) Variablen zu betrachten. Bei Betrachtung von zwei quantitativen Merkmalen bietet sich als anschauliche, graphische Darstellungsweise die Punktwolke an, bei der die Wertepaare durch einen Punkt in einem Koordinatensystem abgebildet werden (Abb. 1). Damit wird sofort visuell erfassbar, ob überhaupt ein Zusammenhang besteht, und wenn ja, wie stark er ist. Tab. 1 enthält die Werte für den systolischen Blutdruck und das Körpergewicht von 24 zufällig ausgewählten Patienten einer dermatologischen Ambulanz. Abb. 1 zeigt die dazugehörige Punktwolke, die einen recht deutlichen Zusammenhang zwischen den beiden Merkmalen erkennen lässt.



09.01.2025

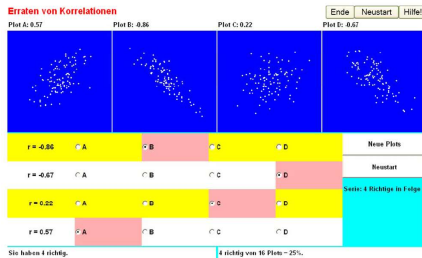
hanno.ulmer@i-med.ac.at

Seite 78

Inferenzstatistik III: Zusammenhänge, Korrelations- und Regressionsanalysen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

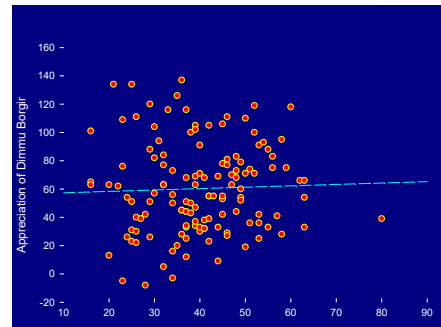
hanno.ulmer@i-med.ac.at

Seite 79

Very small relationship



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

Seite 82

Correlation



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Measuring Relationships
 - Scatterplots
 - Covariance
 - Pearson's Correlation Coefficient
- Nonparametric measures
 - Spearman's Rho
 - Kendall's Tau
- Interpreting Correlations
 - Causality

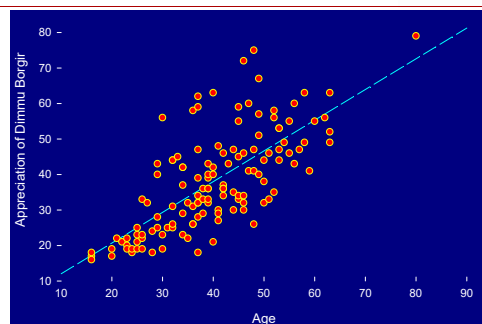
09.01.2025

Seite 80

Positive relationship



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

Seite 83

What is a Correlation?



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- In statistics, **dependence** or **association** is any statistical relationship, whether causal or not, between two random variables or bivariate data.
 - This means that the marginal distribution of a random variable A is different from the distribution of A knowing B
- Correlation** is a way of measuring the relationship between two metric or ordinal variables
- Correlation is a measure of association
- In **common usage**, it most often refers to the extent to which two variables have a **linear relationship** with each other
- (Linear) correlation => Dependence
- But: Dependence \neq (Linear) correlation

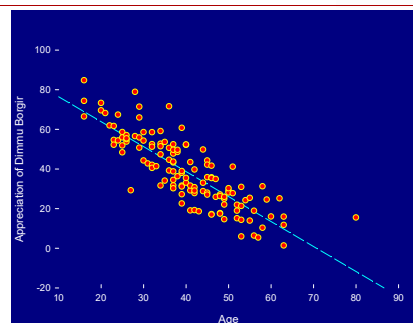
09.01.2025

Seite 81

Negative relationship



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

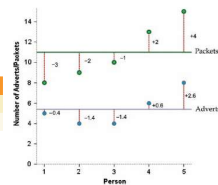
Seite 84

Measuring (linear) relationships

- We need to see whether as one variable increases, the other increases, decreases or stays the same.
- This can be done by calculating the **covariance**.
 - We look at how much each score deviates from the mean.
 - If both variables deviate from the mean by the same amount, they are likely to be related.

- Example:

Participant:	1	2	3	4	5	Mean	s
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92



09.01.2025

Seite 85

Pearson correlation coefficient (1)

- It depends upon the units of measurement.
 - E.g. The Covariance of two variables measured in Miles might be 4.25, but if the same scores are converted to Km, the Covariance is 11.
- One solution: standardise it!
 - Divide by the standard deviations of both variables.
- The standardised version of Covariance is known as the **Pearson correlation coefficient**.

$$r = \frac{Cov_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

In the example:

$$r = \frac{Cov_{xy}}{s_x s_y} = \frac{4.25}{1.67 \times 2.92} = .87$$

09.01.2025

Seite 88

Covariance

- The **variance** tells us by how much scores deviate from the mean for a single variable.
- It is closely linked to the sum of squares.

$$\text{Variance} = \frac{\sum (x_i - \bar{x})^2}{N-1} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N-1}$$

- Covariance** is similar – it tells us by how much scores on two variables differ from their respective means.

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

09.01.2025

Seite 86

Pearson Correlation Coefficient (2)

- Measure of linear dependence
- It varies between -1 (perfect negative) and +1 (perfect positive)
 - 0 = no relationship
- It is an effect size
 - ± 1 = small effect
 - ± 3 = medium effect
 - ± 5 = large effect
- Coefficient of determination, r^2**
 - By squaring the value of r you get the proportion of variance in one variable shared by the other.
- In our example: 0.87^2 , i.e. 75.7% of the variability in "Packets bought" can be explained by "Adverts watched"

09.01.2025

Seite 89

Example

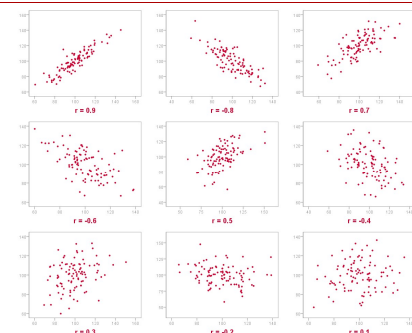
Participant:	1	2	3	4	5	Mean	s
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92

$$\begin{aligned} cov(x, y) &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1} \\ &= \frac{(-0.4)(-3) + (-1.4)(-2) + (-1.4)(-1) + (0.6)(2) + (2.6)(4)}{4} \\ &= \frac{1.2 + 2.8 + 1.4 + 1.2 + 10.4}{4} \\ &= \frac{17}{4} \\ &= 4.25 \end{aligned}$$

09.01.2025

Seite 87

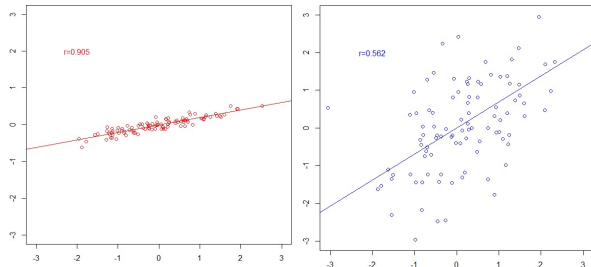
Examples



09.01.2025

Seite 90

Steepness of slope \neq Strength of correlation!



09.01.2025

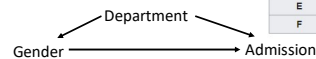
Seite 91

Correlation/Association and causality (2)

- Due to confounders, the estimated of the effect can be distorted so severely that even the estimate shows even in the opposite direction than the true effect
- Simpson's paradox**
- Appleton et al. (The American Statistician 1996;50(4))
- UC Berkeley gender bias**

	Applicants	Admitted
Men	8442	44%
Women	4321	35%

Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	625	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

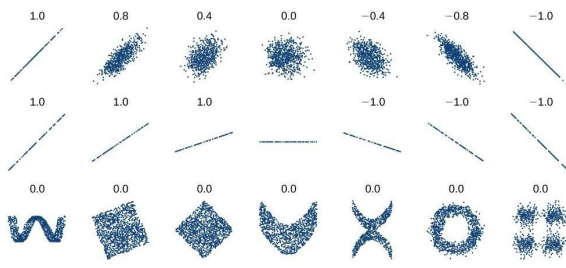


09.01.2025

Seite 94

Non-linear dependencies

- Pearson correlation coefficient not meaningful for non-linear dependencies!



09.01.2025

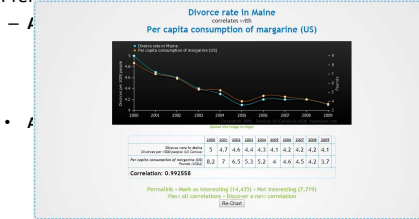
Seite 92

Correlation/Association and causality (3)

- In many experiments/studies, the main focus is on establishing **causal relationships** rather than mere associations

http://tylervigen.com/view_correlation?id=1703

- Prerequisite:

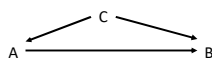


09.01.2025

Seite 95

Correlation/Association and causality (1)

- Direction of causality:**
 - Correlation coefficients say nothing about which variable causes the other to change
- The third-variable problem:**
 - in any correlation, causality between two variables cannot be automatically assumed because there may be other measured or unmeasured variables ("**confounders**") affecting the results.



09.01.2025

Seite 93

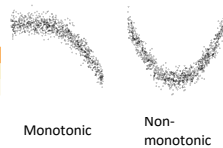
Nonparametric Correlation

- Spearman's Rho**
 - Pearson's correlation on the ranked data
- Kendall's Tau**
 - Better than Spearman's for small samples
- Rho and Tau also appropriate for ordinal data if the assumed dependence is monotonic

Participant	1	2	3	4	5	Mean	s
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92

$R = 0.871$

$Rho = 0.667$



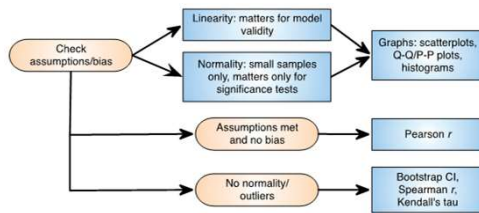
Monotonic

Non-monotonic

09.01.2025

Seite 96

Conducting Correlation Analysis



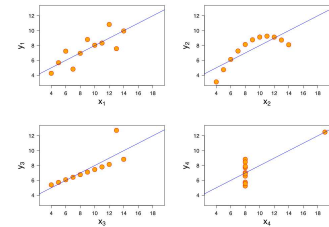
09.01.2025

Seite 97

Importance of looking at a set of data graphically

Anscombe's quartet

- All four sets (x_1 and y_1 , x_2 and y_2 , x_3 and y_3 , x_4 and y_4) are identical when examined using only summary statistics
- Mean of x : 9, Variance of x : 11
- Mean of y : 7.5, Variance of y : 4.125
- Correlation between x and y : 0.816
- Coefficient of determination: 0.67
- Linear regression line: $y=3+0.5x$
- However, they vary considerably when graphed



09.01.2025

Seite 100

Reporting the results of a correlation analysis

TABLE 7.2 An example of reporting a table of correlations

	Exam Performance	Exam Anxiety	Revision Time
Exam Performance	1	-.44*** [-.564, -.301]	.40*** [.245, .524]
Exam Anxiety	103	1	-.71*** [-.863, -.492]
Revision Time	103	103	1

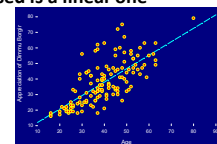
ns = not significant ($p > .05$), * $p < .05$, ** $p < .01$, *** $p < .001$. BCa bootstrap 95% CIs reported in brackets.

09.01.2025

Seite 98

(Simple) linear regression (1)

- A procedure to predict the value of one variable Y from another variable X
 - X – independent variable, predictor
 - Y – dependent variable, outcome
 - Because there is only one predictor -> simple linear regression
 - More than one predictor -> multiple linear regression
- It is a hypothetical model of the relationship between two variables.
- The model used is a linear one



09.01.2025

Seite 101

Exercise

Analyse the data in Data_correlation.csv!

09.01.2025

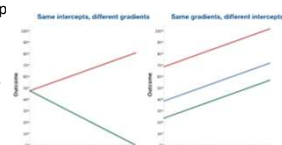
Seite 99

(Simple) linear regression (2)

- Therefore, we describe the relationship using the equation of a straight line

$$Y_i = b_0 + b_1X_i + \varepsilon_i$$

- b_1
 - Regression coefficient for the predictor
 - Gradient (slope) of the regression line
 - Direction/Strength of Relationship
- b_0
 - Intercept (value of Y when $X = 0$)
 - Point at which the regression line crosses the Y -axis (ordinate)



09.01.2025

Seite 102

The Method of Least Squares (1)

- From all possible lines choose the one which **minimizes the sum of the squared residuals** (differences between the prediction from the line and the true y-value)

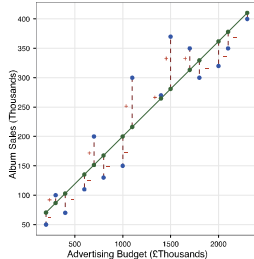


FIGURE 8.4
A scatterplot of some data with a line representing the general trend. The vertical lines (dotted) represent the differences (or residuals) between the line and the actual data.

09.01.2025

Seite 103

Example

$$\bar{x} = 162.85$$

$$\bar{y} = 49.03$$

PERSON i	1	2	3	4	5	6	7	8	9	10
$(x_i - \bar{x})$	-6.55	-3.95	-2.05	16.75	-6.25	2.25	3.05	-6.15	4.95	-2.05
$(y_i - \bar{y})$	-1.93	-2.23	0.27	4.17	-1.33	-0.03	1.57	-1.93	2.67	-1.23
$(x_i - \bar{x}) \cdot (y_i - \bar{y})$	12.64	8.81	-0.55	69.85	8.31	-0.07	4.79	11.87	13.22	2.52
$(x_i - \bar{x})^2$	42.90	15.60	4.20	280.56	39.06	5.06	9.30	37.82	24.50	4.20

$$b_1: \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{131.39}{463.2} = 0.2836$$

$$b_0: 49.03 - 0.2836 \cdot 162.85 = 2.84$$

09.01.2025

Seite 106

The Method of Least Squares (2)

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2 \rightarrow \text{Min!}$$

Coefficients b_0 and b_1 of this minimization task are given by:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{x,y}}{s_{x,x}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

b_1 can also be formulated with the Pearson correlation coefficient:

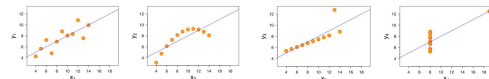
$$b_1 = r_{xy} \cdot \frac{s_y}{s_x}$$

09.01.2025

Seite 104

How Good is the model?

- The regression line is only a model based on the data.
- This model might not reflect reality.
- We need some way of testing how well the model fits the observed data.**
 - Graphically: Deviation from linearity?
 - Diagnostics of residuals
 - Coefficient of determination** via decomposition of variance



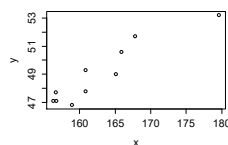
09.01.2025

Seite 107

Example

PERSON i	1	2	3	4	5	6	7	8	9	10
KÖRPERGRÖSSE x	156.3	158.9	160.8	179.6	156.6	165.1	165.9	156.7	167.8	160.8
RINGGRÖSSE y	47.1	46.8	49.3	53.2	47.7	49.0	50.6	47.1	51.7	47.8

- Calculate variances of x and y , covariance and the Pearson correlation coefficient!
- Calculate the regression line according to the least squares approach!
- Calculate the variance of the residuals $(\hat{y} - y)$!

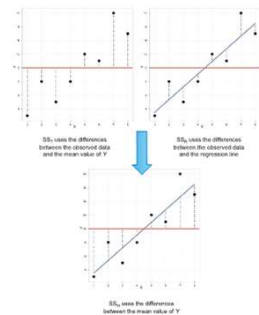


09.01.2025

Seite 105

Decomposition of variance (1)

- SS_{tot}
 - Total variability (variability between scores and the mean)
- SS_{res}
 - Residual/Error variability (variability between the regression model and the actual data)
- SS_{reg}
 - Model variability (difference in variability between the model and the mean)



09.01.2025

Seite 108

Decomposition of variance (2)

- $SS_{tot} = \sum_i (y_i - \bar{y})^2$
- $SS_{reg} = \sum_i (\hat{y}_i - \bar{y})^2$
- $SS_{res} = \sum_i (y_i - \hat{y}_i)^2$
- \hat{y}_i ... predicted values
- One has: $SS_{tot} = SS_{reg} + SS_{res}$
- If the model results in better prediction than using the mean, then we expect SS_{reg} to be much greater than SS_{res}
- The coefficient of determination, R^2 , is defined as $1 - \frac{SS_{res}}{SS_{tot}}$ (or $\frac{SS_{reg}}{SS_{tot}}$), i.e. the proportion of the variance of the outcome variable accounted for by the regression model
- R^2 equals the square of the Pearson correlation coefficient between y and \hat{y} (or – equivalently – between x and y)
- $0 \leq R^2 \leq 1$



$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

09.01.2025

Seite 109

Multiple regression: an example

- A record company boss was interested in predicting album sales from advertising.
- Data
 - 200 different album releases
- Outcome variable:
 - Sales (CDs and Downloads) in the week after release
- Predictor variables
 - The amount (in £s) spent promoting the album before release (see last lecture)
 - Number of plays on the radio (new variable)

09.01.2025

Seite 112

Exercise: Album_Sales.sav

- A record company boss was interested in predicting album sales from advertising.
- Data
 - 200 different album releases
- Outcome variable:
 - Sales (CDs and Downloads) in the week after release
- Predictor variable:
 - The amount (in £s) spent promoting the album before release.

Analyse the data and perform a regression analysis (in SPSS)!

09.01.2025

Seite 110

Multiple Regression as an Equation

- With multiple regression the relationship is described using a variation of the equation of a straight line

$$y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i$$

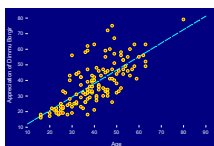
- b_0 is the intercept
 - the intercept is the value of the Y variable when all $X_s = 0$
- b_1 is the regression coefficient for variable 1
- b_2 is the regression coefficient for variable 2
- b_n is the regression coefficient for n^{th} variable.
- ε_i is the residual (error term)

09.01.2025

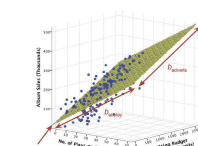
Seite 113

What is Multiple Regression?

- (Simple) Linear Regression is a model to predict the value of one variable from another
- **Multiple Regression** is a natural extension of this model:
 - We use it to predict values of an outcome from *several* predictors
 - It is a hypothetical model of the relationship between several variables



Simple regression - line



Multiple regression with 2 predictors - plane

FIGURE 6.3
Scatterplot of the relationship between album sales, advertising budget and radio plays

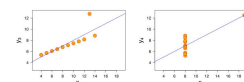
09.01.2025

Seite 111

Model fit (1)

- As in the simple linear regression model
- $SS_{tot} = SS_{reg} + SS_{res}$
- Coefficient of determination, R^2
 - The proportion of variance accounted for by the model.
- R^2 equals the square of the Pearson correlation coefficient between y and \hat{y}
- $0 \leq R^2 \leq 1$
- **Diagnostics of residuals:**
 - Standardized residuals
- **Influential cases:**
 - Cook's distance

$$R^2 = \frac{SS_{reg}}{SS_{tot}}$$

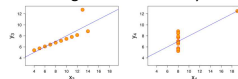


09.01.2025

Seite 114

Model fit (2)

- **Diagnostics of residuals:**
 - **Standardized residuals**
 - In an average sample, 95% of standardized residuals should lie between ± 2 .
 - 99% of standardized residuals should lie between ± 2.5 .
 - Outliers: any case for which the absolute value of the standardized residual is 3 or more, is likely to be an outlier.
- **Influential cases:**
 - **Cook's distance**
 - Measures the influence of a single case on the model as a whole.
 - Weisberg (1982): absolute values greater than 1 may be cause for concern



09.01.2025

Seite 115

Method 2: Stepwise entry

- Variables are entered into the model based on mathematical criteria
- Computer selects variables in steps.
- **Step 1:** Choose the predictor that can explain the most variance in the outcome variable.
- **Step 2:** Having selected the 1st predictor, a second one is chosen from the remaining predictors.
- The semi-partial correlation is used as a criterion for selection.
 - The semi-partial correlation measures the relationship between two variables controlling for the effect that a third variable has on only one of the others
 - It measures the *unique contribution* of a predictor to explaining the variance of the outcome
- **Drawbacks of stepwise entry:**
 - Rely on a mathematical criterion.
 - Variable selection may depend upon only slight differences in the semi-partial correlation
 - These slight numerical differences can lead to major theoretical differences
 - Should be used only for exploration

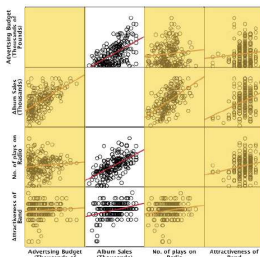
09.01.2025

Seite 118

Which variables to include?

- First step: scatterplot

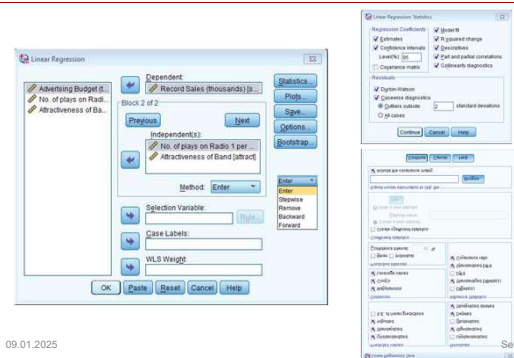
FIGURE 8.14
Matrix scatterplot of the relationships between advertising budget, airplay, and album sales



09.01.2025

Seite 116

Doing Multiple Regression in SPSS (1)



09.01.2025

Seite 119

Method 1: Forced entry

- All variables are entered into the model simultaneously
- The results obtained depend on the variables entered into the model
 - It is important, therefore, to have good theoretical reasons for including a particular variable

09.01.2025

Seite 117

Doing Multiple Regression in SPSS (2)

Model Summary ^a									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.719 ^a	.515	.495	65.592	.315	99.187	1	136	.000
2	.815 ^b	.665	.640	47.087	.330	96.447	2	136	.000

a. Predictors: (Constant), Advertising Budget (Thousands of Pounds)
b. Predictors: (Constant), Advertising Budget (Thousands of Pounds), Attractiveness of Band, No. of plays on Radio

Coefficients ^a									
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	15.0% Confidence Interval for B	Lower Bound	Upper Bound		
1	(Constant)		114.140	7.337	17.799	.000	119.278	149.002	
	Advertising Budget (Thousands of Pounds)	.696	.610	.578	9.979	.000	.877	1.113	
2	(Constant)		-26.613	27.350	-1.534	.127	-60.810	7.604	
	Advertising Budget (Thousands of Pounds)	.685	.607	.511	12.241	.000	.871	.899	
	No. of plays on Radio	1.187	.274	.512	12.121	.000	2.870	1.915	
	Attractiveness of Band	11.646	3.438	.192	4.548	.000	8.278	15.014	

a. Dependent Variable: Album Sales (Thousands)

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

$$\text{Sales} = -26.6 + 0.085\text{adverts} + 3367\text{plays} + 11.1\text{attractiveness}$$

09.01.2025

Seite 120

How to interpret beta values?

- **Beta values:**
 - the change in the outcome associated with a unit change in the predictor
 - $b_1 = 0.085$: as advertising increases by £1, album sales increase by 0.085 units.
 - $b_2 = 3367$: each time (per week) a song is played on the radio its sales increase by 3367 units.
- **Standardised beta values:**
 - tell us the same but expressed as standard deviations
 - $\beta_1 = 0.511$: As advertising increases by 1 standard deviation, album sales increase by 0.511 of a standard deviation.
 - $\beta_2 = 0.512$: When the number of plays on the radio increases by 1 SD its sales increase by 0.512 standard deviations.

09.01.2025

Seite 121

Generalization (2)

- **No Multicollinearity:**
 - Predictors must not be highly correlated.
 - Collinearity diagnostics
- **Homoscedasticity:**
 - For each value of the predictors the variance of the error term should be constant.
 - plot ZRESID against ZPRED
- **Independent Errors:**
 - For any pair of observations, the error terms should be uncorrelated
- **Normally-distributed Errors**
 - Normal probability plot

09.01.2025

Seite 124

Reporting the model

TABLE 8.2 Linear model of predictors of album sales, with 95% bias corrected and accelerated confidence intervals reported in parentheses. Confidence intervals and standard errors based on 1000 bootstrap samples

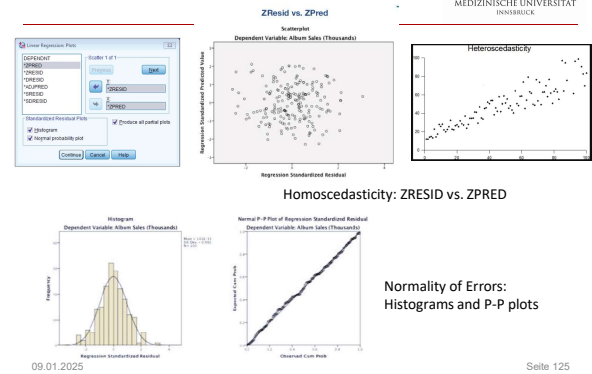
	<i>b</i>	<i>SE B</i>	β	<i>p</i>
Step 1				
Constant	134.14 (120.11, 148.79)	7.95		$p = .001$
Advertising Budget	0.10 (0.08, 0.11)	0.01	.58	$p = .001$
Step 2				
Constant	-26.61 (-55.40, 8.60)	16.30		$p = .097$
Advertising Budget	0.09 (0.07, 0.10)	0.01	.51	$p = .001$
Plays on BBC Radio 1	3.37 (2.82, 3.91)	0.32	.51	$p = .001$
Attractiveness	11.09 (6.28, 15.89)	2.22	.19	$p = .001$

Note. $R^2 = .34$ for Step 1; $\Delta R^2 = .33$ for Step 2 ($p < .001$).

09.01.2025

Seite 122

Regression plots



09.01.2025

Seite 125

Generalization (1)

- When we run regression, we hope to be able to generalize the sample model to the entire population
- To do this, several assumptions must be met
- Violating these assumptions stops us generalizing conclusions to our target population
- **Variable Type:**
 - Outcome must be continuous
 - Predictors can be continuous or dichotomous
- **Non-Zero Variance:**
 - Predictors must not have zero variance
- **Linearity:**
 - The relationship we model is, in reality, linear
- **Independence:**
 - All values of the outcome should come from a different person

09.01.2025

Seite 123

Collinearity diagnostics

		Coefficients ^a			Collinearity Statistics	
Model		Zero-order	Partial	Part	Tolerance	VIF
1	Advertising Budget (Thousands of Pounds)	.578	.578	.578	1.000	1.000
2	Advertising Budget (Thousands of Pounds)	.578	.659	.577	.986	1.013
	No. of plays on Radio	.599	.655	.598	.959	1.043
	Attractiveness of Band	.326	.309	.166	.963	1.038

a. Dependent Variable: Album Sales (Thousands)

- Tolerance should be more than 0.2 (Menard, 1995)
- VIF should be less than 10 (Myers, 1990)

09.01.2025

Seite 126

Signifikanztests und Konfidenzintervalle



Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Ziel der schließenden Statistik



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Aufgrund der Stichprobe Aussagen über die Grundgesamtheit machen
- Von der Stichprobe auf die Grundgesamtheit schließen
- **Grundgesamtheit**
 - ist eine entsprechend dem jeweiligen Untersuchungsziel abgegrenzte Menge von Personen oder Objekten, über die man eine Aussage machen will.
- **Stichprobe**
 - ist eine (kleinere) Menge von Einheiten, die aus der Grundgesamtheit ausgewählt werden und die man misst oder beobachtet, um aus ihnen Schlüsse mit Gültigkeit auch für den nicht ausgewählten Teil der Grundgesamtheit zu ziehen.
- **Repräsentativ**
 - hinsichtlich der für die Untersuchung relevanten Merkmale die Struktur der Grundgesamtheit widerspiegeln

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 130

Statistische Auswertung



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Deskriptive Statistik
- Inferenzstatistik I: Schätzen von Parametern mittels Konfidenzintervallen
- Inferenzstatistik II: Unterschiede, Hypothesenprüfung mittels Signifikanztests
- Inferenzstatistik III: Zusammenhänge, Korrelations- und Regressionsanalysen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 128

Grundlagen der schließenden Statistik: Verteilungsannahmen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Gleichverteilung
 - Bsp.: Würfel, Münze
- Binomialverteilung
 - Bsp.: Ziehen mit Zurücklegen
- Poissonverteilung
 - Bsp.: Seltene Ereignisse mit grossen Stichprobenumfängen und konstanter Wahrscheinlichkeit
- Exponentialverteilung
 - Bsp.: Überlebenszeiten
- Gauß- oder Normalverteilung
 - Wichtigste Verteilung siehe die folgenden Folien
- Testverteilungen
 - Chi-Quadrat-Verteilung
 - T-Verteilung
 - F-Verteilung
 - Anwendung bei Signifikanztests und Konfidenzintervallen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 131

Schließende Statistik



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

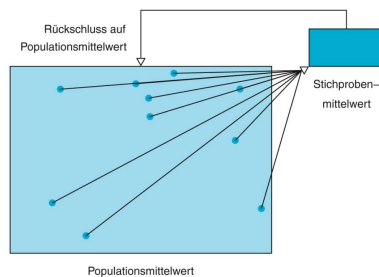


Abbildung 4.1: Das Prinzip des statistischen Schließens.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 129

Normalverteilung



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

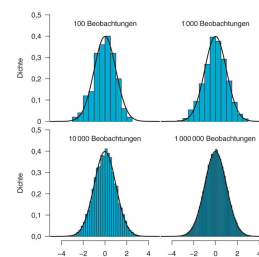


Abbildung 3.4: Histogramme für 100, 1000, 10000 und 1000000 Beobachtungen aus einer Normalverteilung.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 132

Dichtefunktion

DEFINITION 3.6 Dichtefunktion

Liegen von einer stetigen Variablen sehr viele, mit beliebiger Genauigkeit gemessene Beobachtungen vor und beschreiben wir diese Daten mit einem Histogramm, dann könnten wir eine extrem kleine Intervalllänge verwenden. Durch die sehr große Zahl der Beobachtungen ließe sich das Histogramm nicht mehr von einer glatten Kurve unterscheiden. Falls wir auf der y-Achse die Dichte-Skala abtragen, nennt man ein derartiges „Histogramm“ eine **Wahrscheinlichkeitsdichte**.

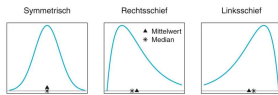


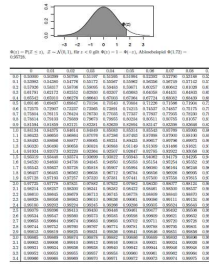
Abbildung 3.5: Dichtefunktion einer symmetrischen, einer rechts- und einer linkschiefen Verteilung.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 133

Quantile der Standardnormalverteilung



Wichtige Quantile der Standardnormalverteilung

Quantil	0.5	0.75	0.9	0.95	0.975	0.99	0.995
Quantil	0	0.67449	1.28156	1.64489	1.96000	2.57583	2.57583

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 136

Gauss- oder Normalverteilung

DEFINITION 3.7 Normalverteilung

Die Dichtefunktion der Normalverteilung lautet:

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

Dabei ist $\pi = 3,14159 \dots$ die Kreiszahl π und \exp die Exponentialfunktion. Der Erwartungswert μ und die Standardabweichung σ sind die Parameter der Normalverteilung.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 134

Die 68-95-99.7% Regel:

Für alle Normalverteilungen gilt

- $\approx 68\%$ aller Beobachtungen liegen innerhalb einer Standardabweichung um den Mittelwert (zwischen $\mu - \sigma$ und $\mu + \sigma$)
- $\approx 95\%$ aller Beobachtungen liegen innerhalb zweier Standardabweichung (zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$)
- 99.7% aller Beobachtungen liegen innerhalb dreier Standardabweichung (zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 137

Gauss- oder Normalverteilung

- $-\infty < X < +\infty$
- Parameter Erwartungswert μ und Varianz σ^2

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

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

$$N(\mu, \sigma^2)$$

- tabelliert ist die Standardnormalverteilung $N(0,1)$
 - $N(0,1)$...Mittelwert 0, Streuung 1
 - Da eine Transformation möglich ist

$$Z = \frac{X - \mu}{\sigma}$$

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 135

Rolle der Normalverteilung

- Viele Messwerte sind annähernd normalverteilt
 - Z.B.: Gewicht einer bestimmten Gruppe
- Manche Messwerte können durch Transformation an eine Normalverteilung angenähert werden
 - Auch diskrete Verteilungen, wie z.B. die Binomialverteilung können durch die Normalverteilung approximiert werden
- Wichtige (und günstige) statistisch-theoretische Eigenschaften
- Die Normalverteilung ist Grundlage wichtiger Testverteilungen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 138

Exercise 1



Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university?

$$Z = (585 - 500) / 100 = 0.85$$

$$P = [\text{area to the left of } z = 0.85] = 0.8023 = 80.23\%$$

Tom scored better than 80.23% of the students who took the test and he will be admitted to this University.

09.01.2025

Seite 139

Der Weg zum Konfidenzintervall



- Punktschätzer aus einer Stichprobe liefert nur einen einzigen Wert
 - Verschiedene Stichproben aus ein und derselben Grundgesamtheit liefern unterschiedliche numerische Werte für den zu schätzenden Parameter
- Berücksichtigung der Standardabweichung der Schätzstatistik (= Standardfehler)

DEFINITION 4.1 Standardfehler

Der Standardfehler eines Schätzers einer Kennzahl der Population ist dessen geschätzte Standardabweichung. Der Standardfehler des Mittelwertes ist s/\sqrt{n} , wobei s die Stichprobenstandardabweichung und n der Stichprobenumfang ist. Bei anderen Kennzahlen ergeben sich analoge Formeln für die zugehörigen Standardfehler.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 142

Exercise 2



The length of life of an instrument produced by a machine has a normal distribution with a mean of 12 months and standard deviation of 2 months. Find the probability that an instrument produced by this machine will last a) less than 7 months.) between 7 and 12 months.

$$P(x < 7) = P(z < -2.5) = 0.0062$$

$$P(7 < x < 12) = P(-2.5 < z < 0) = 0.4938$$

09.01.2025

Seite 140

Inferenzstatistik I: Schätzen von Parametern (Konfidenzintervalle)



■ Beispiel: Rule of three (adverse Events)

- Keine AE in $n=10$ Patienten
95% KI (0-30%)
- Keine AE in $n=100$ Patienten
95% KI (0-3%)
- Keine AE in $n=1000$ Patienten
95% KI (0-0,3%)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 143

Inferenzstatistik I: Schätzen von Parametern (Konfidenzintervalle)



- Ein Konfidenzintervall ist ein Vertrauensbereich oder eigentlich besser ein Unsicherheitsbereich für die Schätzung eines bestimmten, nicht bekannten Parameters der Grundgesamtheit.
- Konfidenzintervall – ein Bereich, in welchem der zu schätzende, unbekannte Parameter der Grundgesamtheit mit Wahrscheinlichkeit $(1 - \alpha)$ liegt
- Der interessierende Parameter kann ein Anteil, ein Mittelwert, ein relatives Risiko etc. sein.
- Ein 95%Konfidenzintervall beispielsweise enthält den gesuchten Parameter mit einer Wahrscheinlichkeit von 95%.

09.01.2025

hanno.ulmer@i-med.ac.at

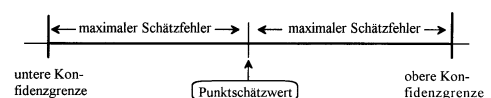
Seite 141

Konfidenzintervalle



“A confidence interval is used when estimating an unknown parameter from sample data. The interval gives a range for the parameter – and a confidence level that the range covers the true value.”

Freedman et al. (1991), p. 385.



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 144

(1-α)-Konfidenzintervall



- Punktschätzer aus einer Stichprobe liefert nur einen einzigen Wert
 - Verschiedene Stichproben aus ein und derselben Grundgesamtheit liefern unterschiedliche numerische Werte für den zu schätzenden Parameter
- Berücksichtigung der Standardabweichung der Schätzstatistik
- Konfidenzintervall – ein Bereich, in welchem der zu schätzende, unbekannte Parameter der Grundgesamtheit mit Wahrscheinlichkeit (1- α) liegt

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 145

Konfidenzintervall für den Mittelwert μ



- Bei **unbekannter** Varianz σ^2 , geschätzt durch s^2
- Gegeben: X_1, \dots, X_N unabhängige normalverteilte Zufallsvariable $N(\mu, \sigma^2)$
- Obere bzw. untere Grenze des Konfidenzintervalls

$$G_{\text{oben, unten}} = \bar{X} \pm t_{(N-1), 1-\alpha/2} s / \sqrt{N}$$

- $t_{(N-1), 1-\alpha/2} \dots$ t-Verteilung mit N-1 Freiheitsgraden
- Allgemein gilt: $t_{(N-1), 1-\alpha} > z_{1-\alpha}$

$$N = (2t_{(N-1), 1-\alpha/2} s / L)^2$$

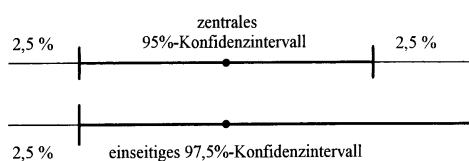
t...Quantil der t-Verteilung für N-1 Freiheitsgrade und 1-α

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 148

Zwei- und Einseitige Konfidenzintervalle



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 146

Konfidenzintervall für einen Anteil π



- Wahrscheinlichkeit π wird durch p aus einer Stichprobe vom Umfang N geschätzt
 - Das Konfidenzintervall ist asymmetrisch
- Approximation durch Normalverteilung
 - Wenn $N > 30$

$$\left[p - z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{N}}, p + z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{N}} \right]$$

- Approximation nach Wald
- Approximation nach Pearson-Clopper

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 149

(1-α)-Konfidenzintervall



- Einseitig
 - Bestimmung einer Ober- bzw. Untergrenze
- Zweiseitig
 - symmetrisch
- Üblich Werte für α
 - α=0.05, α=0.01, α=0.001
- Grundsätzlich kann für alle statistisch geschätzten Parameter ein Konfidenzintervall berechnet werden
 - Mittelwert, Standardabweichung, Häufigkeit, Korrelationskoeffizient, ...

09.01.2025

Statistik in der medizinischen
Forschung

147

95% Konfidenzintervalle

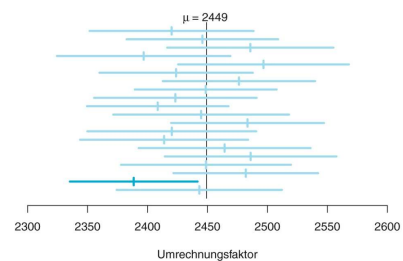


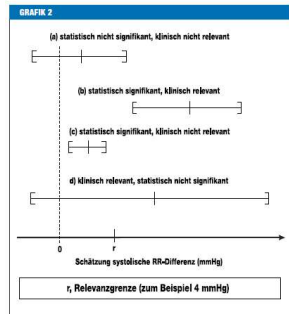
Abbildung 4.4: Illustration des Konfidenzniveaus. Von 20 95%-Konfidenzintervallen überdeckt eines (das dunkelblaue) den wahren Wert nicht.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 150

Klinische Relevanz versus statistische Signifikanz



09.01.2025

Statistische Signifikanz und klinische Relevanz

Seite 151

Inferenzstatistik II: Signifikanztests

- Ergebnis eines Signifikanztests ist die **Teststatistik**.
- Der Wertebereich der Teststatistik wird in zwei Teilmengen zerlegt, den **Verwerfungsbereich** und den **Annahmebereich**. Wenn die Prüfgröße in den Verwerfungsbereich fällt, wird die Nullhypothese verworfen, ansonsten wird sie behalten.
- Man wählt den Verwerfungsbereich so, dass unter H_0 seine Wahrscheinlichkeit unter einen vorgegebenen Wert α fällt. α , das sogenannte **Signifikanzniveau** des Tests, ist damit die **Obergrenze** für die Wahrscheinlichkeit, den **Fehler 1. Art** zu begehen. α wird vom Versuchsleiter vorgegeben. **Übliche Werte für α sind 0.05, 0.01 und 0.001.**
- Welches α man wählt, hängt von den Konsequenzen ab, die der Fehler 1. Art hat. Der naheliegende Wunsch, $\alpha = 0$ zu wählen, scheitert daran, dass dann β , die Wahrscheinlichkeit für den **Fehler 2. Art**, groß wird.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 154

Beispiel: Schätzen von Parametern (Konfidenzintervalle)

Table 2. Effect of Cytisine on Smoking Cessation.^a

Outcome	Cytisine (N=370)	Placebo (N=370)	Percentage-Point Difference (95% CI)	Relative Rate (95% CI) [†]
	percent (number)			
Primary outcome: abstinence for 12 mo	8.4 (31)	2.4 (9)	6.0 (2.7–9.2)§	3.4 (1.7–7.1)
Abstinence for 6 mo	10.0 (37)	3.5 (13)	6.5 (2.9–10.1)‡	2.9 (1.5–5.3)
Point prevalence at 12 mo	13.2 (49)	7.3 (27)	5.9 (1.6–10.3)§	1.8 (1.2–2.8)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 152

Definitionen

DEFINITION 5.1 Nullhypothese

In der **Nullhypothese** H_0 formulieren wir das Gegenteil der wissenschaftlichen Hypothese und nehmen an, dass Unterschiede, die wir in einer Stichprobe beobachten, lediglich durch Stichprobenvariation zustandekommen sind.

DEFINITION 5.2 p-Wert

Der **p-Wert** ist die Wahrscheinlichkeit, unter Annahme von H_0 ein Resultat so groß wie das beobachtete oder noch extremer zu erhalten. „Extremer“ bedeutet dabei „weiter weg vom Nullwert“.

DEFINITION 5.3 Statistische Power

Die **Power** eines Hypothesentests ist definiert als $1 - \beta$, also die Wahrscheinlichkeit, tatsächlich auf die Alternativhypothese zu entscheiden, wenn diese gilt. Das deutsche Wort für Power ist **Trennschärfe**, welches aber selten verwendet wird. Eine hohe Power eines Hypothesentests entspricht folglich einer hohen Wahrscheinlichkeit, bei Vorliegen der Alternative auch tatsächlich auf diese zu entscheiden, das heißt einen tatsächlich vorhandenen Effekt auch zu entdecken.

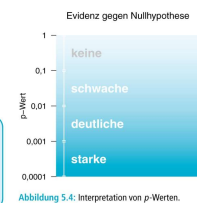


Abbildung 5.4: Interpretation von p-Werten.

Seite 155

Inferenzstatistik II: Signifikanztests

- Eine vorgegebene Annahme (**Nullhypothese H_0**) wird anhand von Daten überprüft. Wenn die Daten "stark" von dem abweichen, was man unter der Nullhypothese erwartet, lässt man die Nullhypothese fallen.
- Im statistischen Test wird dieses Vorgehen formalisiert.
- Nachdem die **Nullhypothese H_0** und die **Alternativhypothese H_1** so formuliert sind, dass sie sich gegenseitig ausschließen und keine dritte Möglichkeit zulassen, ergibt sich ein einfaches Entscheidungsschema mit 4 Möglichkeiten
- Der **Fehler 1. Art** ist der Fehler, die Nullhypothese zu **verwerfen**, obwohl sie **richtig** ist.
- Der **Fehler 2. Art** ist der Fehler, die Nullhypothese zu **behalten**, obwohl sie **falsch** ist.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 153

Mögliche Entscheidungen

In der Stichprobe erhalten wir ein...	In der Population ist...	...
... nicht signifikantes Testresultat ($p > \alpha$)	... H_0 wahr	korrekter Entscheid
... signifikantes Testresultat ($p \leq \alpha$)	... H_1 wahr	falscher Entscheid
		$1 - \alpha$
		β
		korrekter Entscheid
		α
		$1 - \beta$

Tabelle 5.2: Mögliche Entscheidungen in einem Hypothesentest. In der zweiten Zeile jedes Eintrags steht jeweils die auf die Tabellenspalten bedingte Wahrscheinlichkeit für die entsprechende Entscheidung.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 156

Type I versus Type II Error

		Decide for	
		H_0	H_1
Reality	H_0	Correct decision	Wrong decision: Type I error (α)
	H_1	Wrong decision: Type II error (β)	Correct decision: Power ($1-\beta$)

Vorgehen bei der Hypothesenüberprüfung

- Formulierung der Hypothese, Auswahl des Tests
- Datenerhebung
- Analyse der Daten mit Methoden der beschreibenden Statistik (Häufigkeiten, Mittelwert, ...)
- Überprüfung der Daten auf Ihre Verteilung
 - Grafische Darstellung mittels Boxplot und Histogramm
 - Statistische Überprüfung mit dem Kolmogorov-Smirnov Test
- Adaptierung der Testauswahl
- Durchführung des Tests
- Interpretation des Ergebnisses

Parallelgruppenstudie

- z.B. Untersuchung der Wirksamkeit eines neuen blutdrucksenkenden Medikaments im Vergleich zu einer Standardtherapie oder Placebo (=Kontrolle)
 - **Nullhypothese H_0 :** die beiden Therapien sind im Mittel gleich wirksam
 - z.B. die Änderung des systolischen Blutdrucks ist im Mittel in beiden Gruppen gleich
 - Die Ungültigkeit der Nullhypothese ist zu beweisen
 - **Alternativhypothese H_1 :** die beiden Therapien sind im Mittel unterschiedlich stark wirksam
- **Voraussetzung:** Die beiden Gruppen stimmen in den wesentlichen Merkmalen überein – siehe Randomisierung
 - Nur die Therapien sind unterschiedlich
 - Strukturgleichheit der Gruppen

Formulating hypothesis & statistical tests

Steps in conducting a statistical test:

- Quantify the scientific problem from a clinical / biological perspective
- Formulate the model assumptions (distribution of the variable of interest)
- Formulate the problem as a statistical testing problem: Nullhypothesis versus alternative hypothesis
- Define the „error“ you are willing to tolerate
- Calculate the appropriate test statistic
- Decide for the null hypothesis or against it

Formulating Hypothesis
&
Test statistics
&
p-values

Die statistischen Hypothesen

- **Alternativhypothese H_A**
Forschungshypothese

ungerichtet $H_1: \mu_A \neq \mu_B$
gerichtet $H_1: \mu_A > \mu_B$

- **Nullhypothese H_0**
geht davon aus, dass das, was mit der Alternativhypothese behauptet wird, nicht zutrifft.

$$H_0: \mu_A = \mu_B$$

Es gibt keinen Behandlungseffekt

Principles of statistical testing

Different approaches:

- Test for superiority, standard hypothesis testing
- Test for non-inferiority

the difference between the new treatment and the standard is less than the smallest clinically meaningful difference, define delta, use confidence intervals

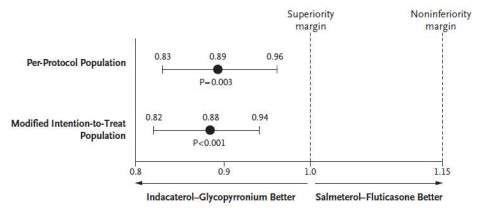
- Test for equivalence

To demonstrate the difference between the new treatment and standard treatment has no clinical importance, define delta

Three primary measures of interest:
a point estimate,
a confidence interval, and
a p-value

Example

A Rate Ratio for All Exacerbations



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 163

Student's t-Test für unabhängige Stichproben (1)

- Parametrischer Test zum Vergleich von Mittelwerten

$$H_0: \mu_A = \mu_B \quad H_1: \mu_A \neq \mu_B$$

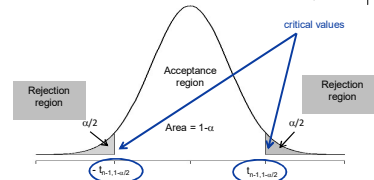
- Teststatistik:

$$t = \frac{\bar{X} - \bar{Y} - (\mu_A - \mu_B)}{\sqrt{\frac{1}{N_A} + \frac{1}{N_B} \frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2}}}$$

- Falls die Nullhypothese gilt, ist die Teststatistik $t(N_A + N_B - 2)$ -verteilt

- Entscheidung:

- H_0 wird mit Irrtumswahrscheinlichkeit α verworfen, wenn: $|t| > t_{(N_A + N_B - 2), 1 - \alpha/2}$



Rejection region: „ H_0 wird abgelehnt, da Daten zu stark vom dem abweichen, was man sich unter H_0 erwarten würde“

09.01.2025

Seite 166

2008: 100 Jahre Student's t-Test: William Sealy Gosset

Volume 11 MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

Abstract.
The question was first suggested by a student of a "population" of means, which might be considered as the mean of a series of observations, each of which is a mean of a series of observations. The question was first suggested by a student of a "population" of means, which might be considered as the mean of a series of observations, each of which is a mean of a series of observations. The question was first suggested by a student of a "population" of means, which might be considered as the mean of a series of observations, each of which is a mean of a series of observations.



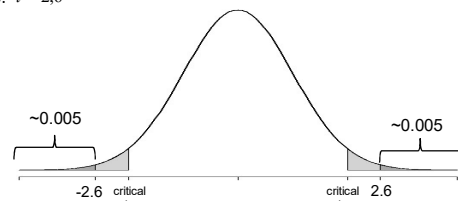
1939 The first GUINNESS® brewery overseas is built at Park Royal, London. William Sealy Gosset, one of the most famous statisticians, is appointed Head Brewer.

hanno.ulmer@i-med.ac.at

Seite 164

Student's t-Test für unabhängige Stichproben (2)

- Wie gelangt man von der Teststatistik \hat{t} zum genauen p-Wert?
- Z.B. $\hat{t} = 2,6$



- P-value (two-sided) = 0.005 + 0.005 = 0.01 (= Area under the curve)

09.01.2025

Seite 167

Therapie	Syst. Blutdruck
1 Therapie A	169
2 Therapie A	139
3 Therapie A	137
4 Therapie A	152
5 Therapie A	142
6 Therapie A	163
7 Therapie A	183
8 Therapie A	143
$\mu_1 = 8 \quad \bar{X}_1 = 153,5$	
1 Therapie B	155
2 Therapie B	154
3 Therapie B	176
4 Therapie B	158
5 Therapie B	156
6 Therapie B	170
7 Therapie B	168
8 Therapie B	179
9 Therapie B	167
10 Therapie B	142
11 Therapie B	163
12 Therapie B	144
$\mu_2 = 12 \quad \bar{X}_2 = 161$	

t-Test quantitatives Merkmal

H_0 : Es besteht kein Therapieunterschied bzgl. mittleren Blutdrucks ($\mu_A = \mu_B$)

H_1 : Es besteht ein Unterschied ($\mu_A \neq \mu_B$)

Teststatistik: $\hat{t} = 1,193$

Syst. Blutdruck bei Therapie A größer bzw. kleiner als bei Therapie B => 2-seitiger t-Test

Testentscheidung: $\hat{t} = 1,193 < 2,101 = t_{18, 0,975} \quad p = 0,248$

$t_{\alpha/2, n-1}$ mit n = Anzahl der Freiheitsgrade; α = Signifikanzniveau

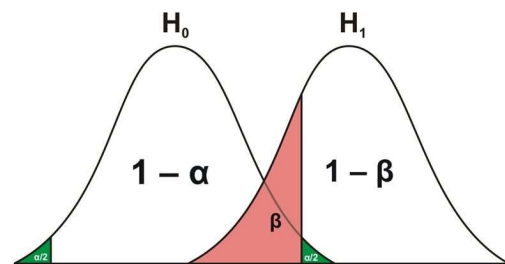
=> Nullhypothese kann nicht abgelehnt werden

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 165

Fehler 1. und 2. Art - Zusammenhang



09.01.2025

Seite 168

Was ist der p-Wert

- Wahrscheinlichkeit, daher $0 < p\text{-Wert} \leq 1$
- Er deutet an, wie wahrscheinlich es ist, ein Stichprobenergebnis wie beobachtet oder extremer zu erhalten, wenn die **Nullhypothese** wahr ist.
- Mit dem p-Wert wird also angedeutet, wie extrem das Ergebnis ist: je kleiner der p-Wert, desto mehr spricht das Ergebnis gegen die Nullhypothese
- $p\text{-Wert} = P(\text{Daten so extrem wie beobachtet} | H_0)$
- Häufige Fehlinterpretation: **p-Wert gibt an, wie wahrscheinlich die Nullhypothese bei Erhalt eines Stichprobenergebnisses wie beobachtet ist, d.h. $P(H_0 | \text{Daten so extrem wie beobachtet} | H_0)$**
- Die Nullhypothese wird verworfen, wenn der p-Wert kleiner als das vom Anwender festgelegte **Signifikanzniveau** α ist (oft 0.01, 0.001, oder 0.05)
- Wenn die Nullhypothese zugunsten der Alternativhypothese verworfen wird, wird das Resultat als statistisch signifikant bezeichnet.
- Die Größe des p-Werts gibt **keine** Aussage über die Größe des wahren Effekts

09.01.2025

Seite 169

Student's t-Test für unabhängige Stichproben (3)

- Ein zwei-Stichproben t-Test zum Vergleich der Mittelwerte in zwei Gruppen liefert einen Wert der Teststatistik von 2,6, was einen p-Wert von 0.01 liefert.
- Richtige Interpretation:
 - Falls das Experiment 100 Mal wiederholt wird, d.h. 100 Mal zufällig eine Stichprobe gezogen wird, dann Teststatistik berechnet wird etc., und falls es in Wirklichkeit keinen Unterschied zwischen den beiden Gruppen gibt (Mittelwerte sind gleich), dann kann man erwarten, dass nur eine dieser 100 Teststatistiken einen Wert $\geq |2.6|$ aufweist
- Falsche Interpretation:
 - **Die Nullhypothese ist zu 99% falsch**

<https://www.graphpad.com/quickcalcs/ttest1.cfm>

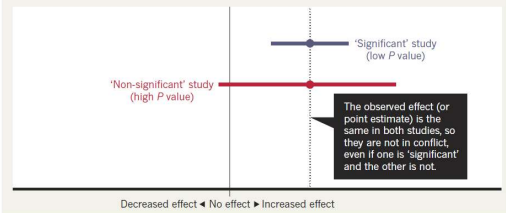
09.01.2025

Seite 172

Falsche Schlußfolgerungen

BEWARE FALSE CONCLUSIONS

Studies currently dubbed 'statistically significant' and 'statistically non-significant' need not be contradictory, and such designations might cause genuine effects to be dismissed.



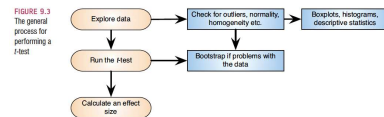
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 170

Student's t-Test für unabhängige Stichproben (4)

- Damit die Teststatistik unter H_0 auch tatsächlich t-verteilt ist und der Test valide ist, müssen folgende Voraussetzungen erfüllt sein:
 - Jede der beiden Grundgesamtheiten sollten normalverteilt sein, Abweichungen bei größeren Stichprobenumfängen ($N \geq 30$) zulässig (zentraler Grenzwertsatz)
 - \rightarrow Graphische Überprüfung mittels Histogramm
 - \rightarrow Kolmogoroff-Smirnov-Test
 - H_0 : Daten sind normalverteilt vs. H_1 : Daten sind nicht normalverteilt
 - Gleiche Varianzen in den beiden Gruppen
 - \rightarrow Levene-Test



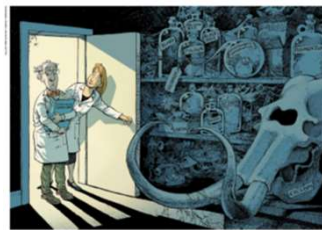
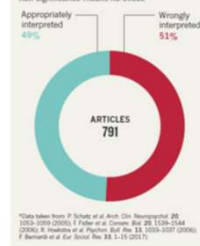
09.01.2025

Seite 173

Falsche Schlußfolgerungen

WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-significance means no effect.



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 171

Gepaarter Zwei-Stichproben t-Test

- Vorher-Nachher-Messungen
- Gruppe 1 (vorher) nicht unabhängig von Gruppe 2 (nachher)
- Modifikation der t-Tests notwendig
- \rightarrow Abhängiger/Gepaarter t-Test
- Berechnung der Differenz D für jedes Paar

$$H_0: \mu_D = 0 \quad H_1: \mu_D \neq 0$$

$$t = \frac{\bar{D} - \mu_D}{s_D / \sqrt{N}} \sim t_{n-1}$$

09.01.2025

Seite 174

Verbundene versus unverbundene Stichproben

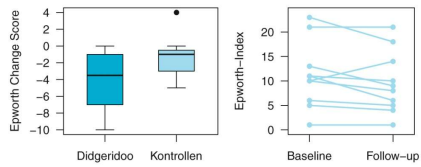


Abbildung 4.5: Illustration von unverbundenen (links) und verbundenen (rechts, nur Kontrollgruppe) Stichproben in der Didgeridoo-Studie.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 175

ANOVA

- Einfache ANOVA, ein Faktor mit ≥ 2 Stufen
- Mehrweg-ANOVA, mehrere Faktoren mit ≥ 2 Stufen
- Kovarianzanalyse, auch stetige Einflussfaktoren
- Multivariate ANOVA, mehrere abhängige Variablen
- ANOVA für Messwiederholungen, aufeinanderfolgende Beobachtungen sind abhängig

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 178

Beispiel

- Um eine neue Therapie zur Senkung des Cholesterinspiegels zu testen, werden bei zehn Probanden vor und nach der Behandlung die Cholesterinwerte bestimmt. Es ergeben sich die folgenden Messergebnisse:

Vor der Behandlung:	223	259	248	220	287	191	229	270	245	201
Nach der Behandlung:	220	244	243	211	299	170	210	276	252	189
Differenz:	3	15	5	9	-12	21	19	-6	-7	12

$$\bar{d} = 5,9 \quad s_d = 11,3866$$

<https://www.graphpad.com/quickcalcs/ttest1.cfm>

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{5,9}{11,3866 / \sqrt{10}} = 1,6385$$

Teststatistik

$$t(0,975; 9) = 2,2622$$

Kritischer Wert

$|t| \leq t(0,975; 9)$ Die Nullhypothese, dass die Erwartungswerte der Cholesterinwerte vor und nach der Behandlung gleich sind, kann nicht abgelehnt werden. Wenn die Behandlung überhaupt einen Effekt hat, so ist dieser nicht groß genug, um ihn mit einem so kleinen Stichprobenumfang zu entdecken.

09.01.2025

Seite 176

ANOVA, post hoc Vergleiche

- Bonferroni
- Bonferroni-Holm
- Holm-Sidak
- Tukey
- Scheffe
- LSD
- Dunnett

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 179

ANOVA

- Analysis of Variance - Streuungszerlegung
- Parametrischer Test zum Vergleich von Mittelwerten
- T-test = Zweigruppenvergleich, ANOVA mehrere Gruppen möglich
- Mittelwertsvergleich erfolgt durch die Zerlegung der Varianz in:
 - Streuung zwischen den Gruppen
 - Streuung innerhalb der Gruppen
- Voraussetzungen
 - Normalverteilung
 - Varianzhomogenität

09.01.2025

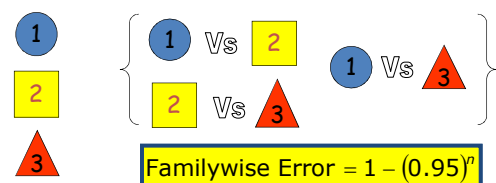
hanno.ulmer@i-med.ac.at

Seite 177

Why Not Use Lots of t-Tests?

If we want to compare several means why don't we compare pairs of means with t-tests?

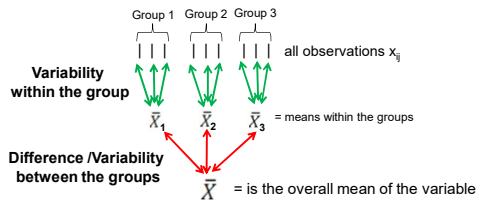
- Can't look at several independent variables.
- Inflates the Type I error rate.



Copyright
Andy Field

Analysis of Variance (ANOVA)

- Situation: Compare the means of k samples (k>2)
- Assumption: normal distribution of the population, $\sigma = \sigma_1 = \sigma_2 = \dots = \sigma_k$
- Hypothesis: $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ versus $H_1: \mu_i \neq \mu_j (i \neq j)$: At least two of the means differ
- Nowadays, linear mixed effects models are preferred instead of ANOVA



Beispiel

Smoking status	Current Smoker	Ex-Smoker	Never Smoker	Row Total
Gender				
Men	144	310	268	722
Women	117	143	475	735
Column Total	261	453	743	1457

Teststatistik: $121.9218 \sim \chi^2 ((2-1) \cdot (3-1)) = \chi^2 (2)$

Kritischer Wert von $\chi^2 (2) = 5.99$

→ Nullhypothese (Rauchverhalten unterscheidet sich nicht zwischen Männern und Frauen) kann abgelehnt werden ($p = 3.3e-27$)

09.01.2025

Seite 184

Pearson's Chi-Quadrat Test

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Korrelation und Regression



Karl Pearson (1857-1936)

Fisher's Exakter Test

Varianzanalyse



Sir Ronald Aylmer Fisher (1890-1962)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 182

Übung: Führen Sie einen Chi-Quadrat Test durch

Table 2. Effect of Cytisine on Smoking Cessation.*

Outcome	Cytisine (N=370)	Placebo (N=370)	Percentage-Point Difference (95% CI)	Relative Rate (95% CI)†
	percent (number)			
Primary outcome: abstinence for 12 mo	8.4 (31)	2.4 (9)	6.0 (2.7-9.2)‡	3.4 (1.7-7.1)
Abstinence for 6 mo	10.0 (37)	3.5 (13)	6.5 (2.9-10.1)‡	2.9 (1.5-5.3)
Point prevalence at 12 mo	13.2 (49)	7.3 (27)	5.9 (1.6-10.3)‡	1.8 (1.2-2.8)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 185

Chi-Quadrat-Test qualitatives Merkmal

Blutdruck	Therapie A	Therapie B	Gesamt
Hypertonie	$n_{1,1}=300$	$n_{1,2}=150$	$n_{1,}=450$
Normotonie	$n_{2,1}=300$	$n_{2,2}=250$	$n_{2,}=550$
Gesamt	$n_{,1}=600$	$n_{,2}=400$	$n=1000$

H_0 : Es besteht kein Zusammenhang zwischen Therapie und Blutdruck ($p_i = p_j$)

H_1 : Es besteht ein Zusammenhang ($p_i \neq p_j$)

Teststatistik: $\chi^2 = n \cdot \frac{(n_{2,1} \cdot n_{1,2} - n_{1,1} \cdot n_{2,2})^2}{(n_{2,} \cdot n_{1,} \cdot n_{,2} \cdot n_{,1})} = 1000 \cdot \frac{(300 \cdot 250 - 300 \cdot 150)^2}{450 \cdot 550 \cdot 600 \cdot 400} = 15,15$

Testentscheidung: $\chi^2 = 15,15 > 3,841 = \chi^2_{0,95}$ ⇒ Ablehnen der Nullhypothese

χ^2_{n-1} mit n = Anzahl der Freiheitsgrade; α = Signifikanzniveau

09.01.2025

hanno.ulmer@i-med.ac.at

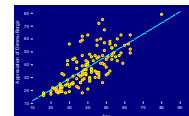
Seite 183

Regression – significance testing (1)

$$y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + \epsilon_i$$

$$SS_{tot} = SS_{reg} + SS_{res}$$

$$R^2 = \frac{SS_{reg}}{SS_{tot}} \dots \text{Coefficient of determination}$$



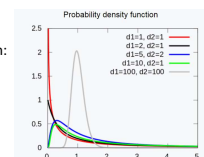
Using the regression model, can we significantly better predict values of the outcome than using the mean?

$H_0: R^2=0$ (alternativ: $\beta_1 = \beta_2 = \dots = \beta_p = 0$)

$H_1: R^2 \neq 0$

Test statistic: $F = \frac{MS_{reg}}{MS_{res}}$

F-distribution:



MS ... Mean Squares (averages of total values)

$F \sim F(n, N-(n+1))$ -distributed

ANOVA test – Analysis Of Variance

09.01.2025

Seite 186

Regression – significance testing (2)



OUTPUT 8.2

ANOVA ^a					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	433687.833	1	433687.833	99.587	.000 ^b
Residual	862264.167	198	4354.870		
Total	1295952.00	199			

a. Dependent Variable: Album Sales (thousands)
b. Predictors: (Constant), Advertising Budget (thousands of pounds)

Critical values for the F-distribution:

http://www.socr.ucla.edu/applets.dir/f_table.html

Critical value for $F(1,198)$ for $\alpha=0.05$: ~ 3.9

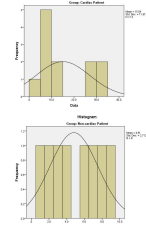
09.01.2025

Seite 187

Nicht-parametrische Tests Mann-Whitney U Test (1)



- 20 Patienten einer Klinik werden untersucht. 12 davon sind in kardiologischer Behandlung, während 8 dies nicht sind. Sie alle beantworten einen Fragebogen zum allgemeinen Wohlbefinden (Werte von 0 bis 35, 0 steht für ein sehr hohes, 35 für ein sehr geringes Wohlbefinden). Es soll geprüft werden, ob es Unterschiede hinsichtlich der zentralen Tendenz des Wohlbefindens zwischen den Herzpatienten und den übrigen Patienten gibt.
- Mann-Whitney.sav



- T-Test: Nur zulässig, falls Daten normalverteilt
- Falls Voraussetzungen für T-Test nicht erfüllt, dann:
- Mann-Whitney U Test:** zulässig für alle Verteilungen, solange $F_Y(x) = F_X(x - a)$ gilt (Verteilung in den beiden Gruppen bis auf Verschiebung gleich)

$$H_0 : a = 0 \text{ vs. } H_1 : a \neq 0$$

09.01.2025

190

Regression – significance testing (3)



- To test the significance of individual regression coefficients
 - t-test

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

- Test statistic: $T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t(N-2)\text{-distributed}$

Coefficients ^a							
Model		Unstandardized Coefficients	Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta		Lower Bound	Upper Bound
1	(Constant)	114.146	7.337		15.589	99.528	148.002
	Advertising Budget (Thousands of Pounds)	-.098	.010	-.17	1.639	-.215	-.081
2	(Constant)	-20.633	17.356		-.118	-54.930	13.664
	Advertising Budget (Thousands of Pounds)	.085	.007	.11	12.243	.000	.071
	No. of plays on Radio	1.187	.278	.43	4.271	.000	.630
	Advertisements of Band	13.068	2.638	.50	4.955	.000	7.775

a. Dependent Variable: Album Sales (Thousands)

09.01.2025

Seite 188

Nicht-parametrische Tests Mann-Whitney U Test (2)



- MWU-Test rechnet mit den Rängen, nicht mit den Messwerten selbst

ID	Gruppe	Wohlbefinden	Ränge Gruppe 1	Ränge Gruppe 2
5	1	0	1	
6	2	1		2
14	2	2		3
9	2	3		4
18	2	4		5
10	1	5	6	
19	1	5.5	7	
1	2	6		8
8	2	6.5		9
17	1	7	10	
15	2	7.5		11
11	1	8	12	
3	2	8.5		13
2	1	9	14	
20	1	11	15	
12	1	13	16	
16	1	28	17	
4	1	29	18	
7	1	32	19	
13	1	33	20	
Rangsumme			155	55

- Berechnung der Rangsummen in beiden Gruppen
- Teststatistik U: $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$
 n_1 =Stichprobengröße der Gruppe mit der größeren Rangsumme
 n_2 =Stichprobengröße der Gruppe mit der kleineren Rangsumme
 R_1 =größere der beiden Rangsummen
- Im Bsp.: $U = 12 \cdot 8 + \frac{12(12+1)}{2} - 155 = 19$
- Stichprobe hinreichend groß ($n_1+n_2>30$): U annähernd normalverteilt
- Im Bsp.: $z = \frac{U - \frac{n_1 n_2}{2}}{\sigma_U} = \frac{19 - 12 \cdot 8}{\sqrt{12 \cdot 8 \cdot (12+8+1)}} = -2.34$
- $p < 0.05$, da $-2.34 < -1.96$

09.01.2025

Seite 191

Regression – significance testing (4)

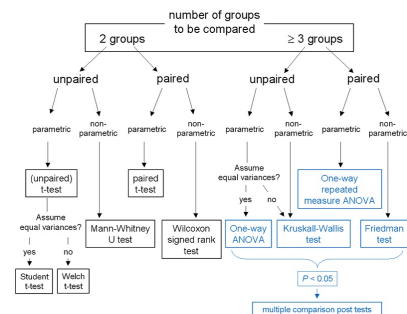


- Prerequisites that F-test (ANOVA) and t-test for regression are valid
 - Homoscedasticity:
 - For each value of the predictors the variance of the error term should be constant.
 - Independent Errors:
 - For any pair of observations, the error terms should be uncorrelated
 - Normally-distributed Errors
 - Normal probability plot

09.01.2025

Seite 189

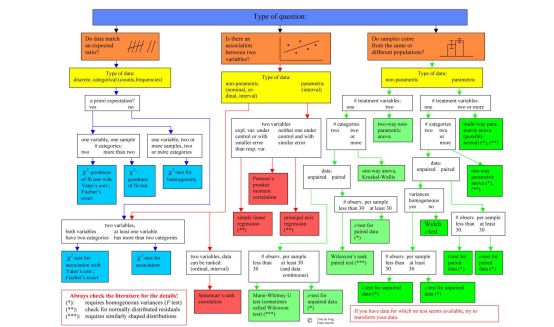
Overview of statistical tests (2)



09.01.2025

Seite 192

Overview of statistical tests (3)



09.01.2025

Seite 193

Multiples Testen

- „The multiple comparison problem involves the repeated testing of a series of hypotheses and the resultant increasing probability of a type I error.“ Van Belle (2002), p. 149.

- P (Fehler 1. Art)

$$\alpha = 1 - (1 - 0.05)^n$$

- z.B. Bonferroni-Korrektur

$$\alpha_i = \frac{0.05}{n}$$

09.01.2025

Statistik in der medizinischen
Forschung

196

Common statistical tests

	Quantitative outcome variable		Qualitative outcome variable	
	Normal distribution	Any other distribution	Expected frequency in each cell of the crosstable „high“	Expected frequency in each cell of the crosstable „low“
Compare 2 groups	t-test	Wilcoxon-test / Mann-Whitney U-Test	Chi-Square	Fishers exact test
Compare >2 groups	Analysis of Variance (ANOVA)	Kruskal-Wallis-Test	Chi-Square	

Testing measures of location:
Does the mean/median differ between groups

Testing frequencies in a crosstable:
Are the rows and columns independent from each other?

Sample size estimation

Question: How many individuals do you have to include in your study to get a reliable result ?

→ We want to **maximize the probability** for rejecting H_0 if H_1 is true

→ while keeping the **Type I error α** fixed

What do you have to know to calculate the sample size needed?

1. Power (typically set to 80% or 90%)
2. Type I error α (typically set to $\alpha = 0.05$)
3. The difference you want to find (for t-tests: the mean difference between groups)
4. standard deviation / measure of variance

Wichtige Signifikanztests

		Zielvariable(Outcome)		
		Qualitativ	Quantitativ	
			Normalverteilung	Beliebige Verteilung
Vergleich zweier Gruppen	Unverbunden	Chi-Quadrat Test, Fisher Test	t-Test für unverbundene Stichproben	Mann-Whitney U Test
	Verbunden	McNemar Test	t-Test für verbundene Stichproben	Wilcoxon Test
Vergleich von mehr als zwei Gruppen	Unverbunden	Chi-Quadrat Test	Einfache Varianzanalyse	Kruskal-Wallis Test
	Verbunden	Q-Test von Cochran	Varianzanalyse für Meßwiederholungen	Friedman Test

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 195



Sample size estimation

Example

- Hypothesis: $H_0: \mu_A = \mu_B$ versus $H_1: \mu_A \neq \mu_B \rightarrow$ two-sided t-test
- You consider a difference of 10 as relevant
- From former studies, you know, that the standard deviation is ~ 15 mmHG
- So far, you have recruited 20 patients (10 in each treatment arm)

→ What is your power?

<http://campus.uni-muenster.de/fileadmin/einrichtung/imib/lehre/skripte/biomatthe/biofallz.html>

Sample size estimation



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Fallzahlsschätzung für unverbundene Stichproben und stetige Zielgrößen

Ende Neustart Hilfe

- Fallzahlberechnung für vorgegebene Power
 - Powerberechnung für vorgegebene Fallzahl
 - Entdeckbare Differenz für vorgegebene Fallzahl und Power
- Eingabe von μ_1 : Eingabe von μ_2 :
- Eingabe von σ : Differenz Delta:
- Einseitiger Test
 - Zweiseitiger Test
- Eingabe von α (Standard ist 0.05):
- Eingabe der Power (Standard ist 0.80):
- Die Fallzahl für jede Gruppe ist:
- Berechne

The multiple testing problem



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

The Bonferroni correction method:

- Control the comparison-wise error rate: Reject H_0 if $p < \alpha$
- Control the family-wise error rate (including k tests): Reject H_0 if $p < \alpha/k$
→ **Advantage: simple**
- Problem: Bonferroni-correction increases the probability of a type II error
→ the power of detecting a true association is reduced → **Disadvantage: too conservative**

k	α/k ($\alpha=0.05$)
1	0.05
5	0.01
10	0.005
100	0.0005

→ $0.05/5=0.01$

202

The multiple testing problem

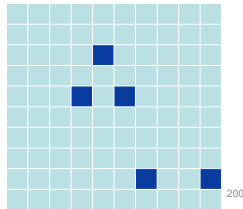


MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

The situation:

- Consider a dataset with 100 independent parameters, which do not play a role in the etiology of the disease of interest (what you don't know, of course)
- 100 statistical tests are performed with a significance level of $\alpha=0.05$
- The tests are constructed in that way, that maximum 5 of 100 tests reject the Nullhypothesis, although it is true (which is the case in this example)

→ You expect 5 tests to be significant just by chance



200

Problem of Multiplicity



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Number of tests	α level (in each test)	Global α level (risk of making at least one type 1 error)
1	0.05	0.05
2	0.05	0.10
3	0.05	0.14
5	0.05	0.23
14	0.05	0.51
100	0.05	0.994

Already by 14 tests we have over 50% chance of falsely rejecting H_0 in at least one of the tests.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 203

The multiple testing problem



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- The probability to get at least one Type I error increases with increasing number of tests.
- Family-wise error rate** (the error rate for the complete family of tests performed): $\alpha^* = 1 - (1 - \alpha)^k$, with α being the **comparison-wise error rate**

The probability to get one or more false discoveries (Type I error)

k	α^* ($\alpha=0.05$)
1	0.05
5	0.226
10	0.401
100	0.994

→ The significance level has to be modified for multiple testing situations

201

Das Problem der Multiplizität



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

ICH E9:

„Multiplicity may arise, for example, from multiple primary variables, ... multiple comparisons of treatments, repeated evaluations over time and/or interim analyses.“

- Mehrere Zielkriterien
- Mehr als 2 Behandlungsgruppen
- Zielkriterium an mehreren Zeitpunkten gemessen
- Zwischenauswertungen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 204

How to correct for multiplicity?



Bonferroni correction:

Create a **corrected significance level α/N** and test each of the analyses on this new level.

Example: We have 5 tests ($N=5$) and we wish to have an overall $\alpha=0.05$. Conduct each test on the corrected significance level $\alpha=0.05/5=0.01$.

Advantages:

- Easy to implement.
- No order of objectives

Disadvantages:

- Very conservative

How to correct for multiplicity?



Bonferroni-Holm

Conduct all tests (N tests). Order all the p-values from smallest to biggest: $P_{(1)}, P_{(2)}, \dots, P_{(N)}$

Test the smallest p-value ($P_{(1)}$) on $P_{(1)} < \alpha/(N+1)$ and the m^{th} p-value on $P_{(m)} < \alpha/(N+1-m)$. STOP at first non significant p-value (this p-value and all bigger p-values will be considered non-significant).

Advantages:

- No order of objectives.
- Not as conservative as Bonferroni.

Disadvantages:

- Hard to implement/explain.

How to correct for multiplicity?



Gate keeping procedures:

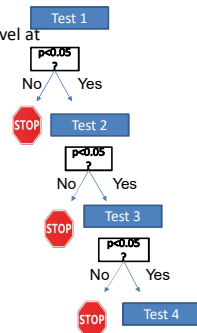
Order objectives/analyses. Test each level at
If significant => move to next level.
If non significant => STOP!

Advantages:

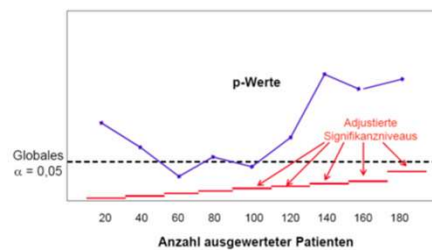
- Easy to implement.
- All tests on the same α level

Disadvantages:

- Must be pre-planned.
- Requires order of analyses.
- Some analyses may not be conducted.



Beispiel für Adjustierung des Fehlers 1. Art in Zwischenauswertungen

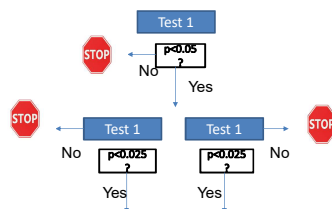


09.01.2025

hanno.ulmer@i-med.ac.at

Seite 209

Combined Bonferroni and Gate keeping



Subgruppenanalysen

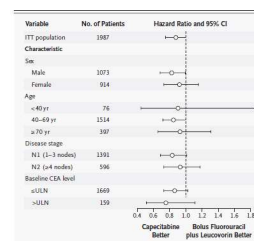


Figure 2. Subgroup Analysis of Disease-Free Survival in the Capecitabine Group as Compared with the Fluorouracil-plus-Leucovorin Group (Intention-to-Treat Population).
Data on carcinoembryonic antigen (CEA) levels were missing for 159 patients who were therefore not included in the analysis for this variable. ITT denotes intention to treat, N1 nodal status, and ≥ 5.0 ng/mL upper limit of normal.

09.01.2025

Seite 210

Lösungen für das Multiplizitätsproblem



- die Studie als explorativ zu definieren – data mining, fishing for significance erlaubt
- 1 primärer Zielparameter wird konfirmatorisch ausgewertet, alle übrigen explorativ
- Alfa-Adjustierung: Aufteilung des Signifikanzniveaus (erhöht Fallzahl)
Extrembeispiel: Biomarkersuche in Proteomics und Genomics
- A priori Ordnung der Hypothesen (bis zur ersten Nichtsignifikanz mit $\alpha=0,05$ testen)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 211

Theory of ANOVA (2)



$$F = \frac{MS_M}{MS_R} \text{ follows a } F(k-1, N-k) \text{ distribution}$$

Decision about $H_0 (\mu_1 = \mu_2 = \dots = \mu_k)$ by comparing F with the critical value of the $F(k-1, N-k)$ distribution

Assumptions which have to be fulfilled for the ANOVA (the same as for the t-test)

- data normally distributed
- homogeneity of variances

09.01.2025

Seite 214

ANOVA



- Null Hypothesis:
 - Like a t-test, ANOVA tests the null hypothesis that the means are the same.
- Experimental Hypothesis:
 - The means differ.
- ANOVA is an Omnibus test
 - It test for an overall difference between groups.
 - It tells us that the group means are different.
 - It doesn't tell us exactly which means differ.
- $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
- H_1 : es gibt i und j mit $\mu_i \neq \mu_j$

09.01.2025

Seite 212

ANOVA by hand (1)



- Testing the effects of viagra on libido using three groups:
 - Placebo (Sugar Pill)
 - Low Dose Viagra
 - High Dose Viagra
- The outcome/dependent variable (DV) was an objective measure of libido.

TABLE 11.1 Data in Viagra.sav

	Placebo	Low Dose	High Dose
3	3	5	7
2	2	2	4
1	4	5	
1	2	3	
4	3	6	
\bar{X}	2.20	3.20	5.00
s^2	1.30	1.30	1.58
s^2	1.70	1.70	2.50
Grand mean = 3.467 Grand SD = 1.767			
Grand variance = 3.124			

09.01.2025

Seite 215

Theory of ANOVA (1)



- Same concept as shown for the F-test for regression
- Decomposition of the total variability (SS_T) into (i) variability between groups (SS_M) and (ii) variability within groups (SS_R)

$$SS_T = \sum (x_i - \bar{x}_{grand})^2 \quad df_T = (N-1) \quad SS_T = SS_M + SS_R$$

$$SS_M = \sum n_i (\bar{x}_i - \bar{x}_{grand})^2 \quad df_M = (k-1) \quad MS_M = \frac{SS_M}{df_M}$$

$$SS_R = \sum (x_i - \bar{x}_i)^2 \quad df_R = (n_1-1) + \dots + (n_k-1) \quad MS_R = \frac{SS_R}{df_R}$$

- k ... number of groups, n_i ... size of group i, N ... $n_1 + \dots + n_k$
- SS ... Sum of squares
- MS ... Mean squares

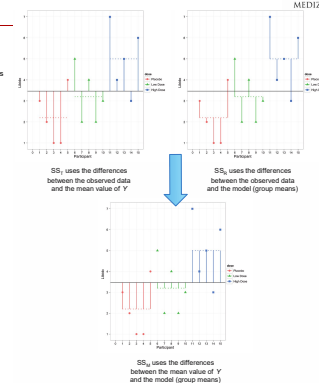
09.01.2025

Seite 213

ANOVA by hand (2)



FIGURE 11.3 Graphical representation of the different sums of squares in ANOVA designs



09.01.2025

Seite 216

ANOVA by hand (3)

Step 1:

TABLE 11.1 Data in Viagra.sav

	Placebo	Low Dose	High Dose
8	2	2	4
1	4	5	
1	2	3	
4	3	5	
\bar{x}	2.20	3.20	5.00
s	1.70	1.50	1.58
s^2	1.70	1.70	2.50
Grand mean	3.467		
Grand variance	3.124		

Step 3:

$$SS_R = s_{\text{group1}}^2(n_1 - 1) + s_{\text{group2}}^2(n_2 - 1) + s_{\text{group3}}^2(n_3 - 1)$$

$$= (1.70)(5 - 1) + (1.70)(5 - 1) + (2.50)(5 - 1)$$

$$= (1.70 \times 4) + (1.70 \times 4) + (2.50 \times 4)$$

$$= 6.8 + 6.8 + 10$$

$$= 23.60$$

Step 2:

$$SS_M = \sum n_i (\bar{x}_i - \bar{x}_{\text{grand}})^2$$

$$SS_M = 5(2.2 - 3.467)^2 + 5(3.2 - 3.467)^2 + 5(5.0 - 3.467)^2$$

$$= 5(-1.267)^2 + 5(-0.267)^2 + 5(1.533)^2$$

$$= 8.025 + 0.355 + 11.755$$

$$= 20.135$$

Step 4: Double check

$$SS_T = 3.124(15 - 1)$$

$$= 43.74$$

$$SS_T = SS_M + SS_R$$

$$43.74 = 20.14 + 23.60$$

$$43.74 = 43.74$$

09.01.2025

Seite 217

Post-hoc tests

- SPSS has 18 types of Post Hoc tests!

- Bonferroni (conservative option)
- Bonferroni-Holm
- Holm-Sidak
- Tukey HSD
- Scheffe
- LSD
- Dunnnett

Multiple Comparisons

Dependent Variable: Libido		Mean Difference (I-J)	Sig.	95% Confidence Interval	
(I) Dose of Viagra	(J) Dose of Viagra			Lower Bound	Upper Bound
Tukey HSD	Placebo - Low Dose	-1.000	.887	-3.37	1.37
	Placebo - High Dose	-2.800	.021	-5.17	-.43
	Low Dose - High Dose	-1.800	.887	-4.17	.57
Games-Howell	Placebo - Low Dose	-1.000	.825	-3.36	1.36
	Placebo - High Dose	-2.800	.017	-5.44	-.16
	Low Dose - High Dose	-1.800	.917	-4.44	.84
Dunnnett (vs control)	Placebo - Low Dose	-2.800	.017	-5.79	-.16
	Placebo - High Dose	-1.000	.887	-.87	4.44
	Low Dose - High Dose	2.800	.887	-.87	4.44

* The mean difference is significant at the 0.05 level.
a. Dunnnett's tests treat one group as a control, and compare all other groups against it.

09.01.2025

Seite 220

ANOVA by hand (4)

Step 5: Calculate mean squared errors

$$MS_M = \frac{SS_M}{df_M} = \frac{20.135}{2} = 10.0675$$

$$MS_R = \frac{SS_R}{df_R} = \frac{23.60}{12} = 1.9667$$

Step 6:

$$F = \frac{MS_M}{MS_R} = \frac{10.067}{1.967} = 5.12$$

Summary table:

Source	SS	df	MS	F
Model	20.14	2	10.067	5.12*
Residual	23.60	12	1.967	
Total	43.74	14		

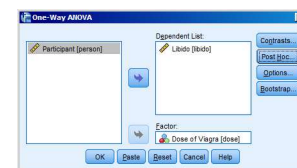
Step 7: Critical value of F(2,12): 3.89

5.12 > 3.89 – therefore H_0 can be rejected

09.01.2025

Seite 218

ANOVA in SPSS



ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.133	2	10.067	5.119	.025
Within Groups	23.600	12	1.967		
Total	43.733	14			

09.01.2025

Seite 221

Follow-up tests

- The F-ratio does not tell us specifically which group means differ from which
- We need additional tests to find out where the group differences lie
- > **Post Hoc tests**
- Compare each mean against all others.
- Multiple tests
- To control the family-wise error rate, stricter criteria to accept an effect as significant must be used
- Simplest example is the Bonferroni method

$$\text{Bonferroni } \alpha = \frac{\alpha}{\text{Number of Tests}}$$

09.01.2025

Seite 219

ANOVA - Variants

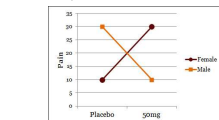
- Till now: only one factor (group variable)
 - Simple ANOVA / One-way ANOVA
- Extension: **several factors** each with >=2 levels
 - Two-way/Three-way etc. ANOVA
 - Several Independent Variables is known as a factorial design
- ANCOVA**: Analysis of Covariance, also continuous/metric independent variables are allowed
- Multivariate ANOVA**: several dependent variable
- ANOVA for repeated measurements**: subsequent observations are dependent

09.01.2025

Seite 222

Two-way (independent) ANOVA

- Two Independent Variables
- Several independent variables is known as a **factorial design**
- Benefit** of factorial designs
 - We can look at how variables **interact**
- Interactions**
 - Show how the effects of one IV might depend on the effects of another
 - Are often more interesting than main effects.
- Examples:



09.01.2025

Seite 223

Two-way ANOVA – example (2)

- Step 1: Calculate SS_T**

$$SS_T = s_{grand}^2(N - 1) = 190.78(48 - 1) = 8966.66$$

- Step 2: Calculate SS_M**

$$SS_M = \sum n_i (\bar{x}_i - \bar{x}_{grand})^2$$

$$SS_M = 40(56.625 - 58.33)^2 + 8(66.875 - 58.33)^2 + 8(57.5 - 58.33)^2 + 8(56.25 - 58.33)^2 + 8(2.2925)^2 + 8(8.5437)^2 + 8(4.177 - 58.33)^2 + 8(6.825)^2 + 8(-11.7675)^2 + 42(130 - 584.1302) + 120(1112 - 584.1302) + 5(3112 - 4124.1302) = 5479.167$$

TABLE 13.3 Data for the beer-goggles effect

Alcohol Gender	None	2 Pints	4 Pints
Female	65	70	55
Male	55	65	30
Female	60	80	70
Male	65	70	55
Female	60	70	55
Male	55	75	80
Female	60	75	80
Male	55	65	50
Total	405	535	500
Mean	60.625	66.875	57.50
Variance	24.55	196.70	59.00

Grand mean: 58.33

09.01.2025

Seite 226

Variance decomposition

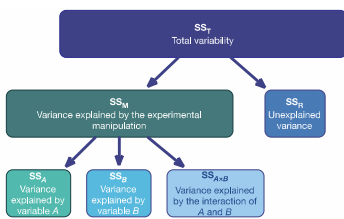


FIGURE 13.3 Breaking down the variance in two-way ANOVA

09.01.2025

Seite 224

Two-way ANOVA – example (3)

- Step 2a: Calculate SS_A**

$$SS_{gender} = 24(60.21 - 58.33)^2 + 24(56.46 - 58.33)^2 = 24(1.88)^2 + 24(-1.87)^2 = 84.8256 + 83.9256 = 168.75$$

A ₁ : Female		
65	70	55
70	65	65
60	60	70
60	70	55
60	65	55
55	60	60
60	60	50
55	50	50

Mean Female = 60.21

A ₂ : Male		
50	45	30
55	60	30
80	85	30
65	65	55
70	70	35
75	70	20
75	80	45
65	60	40

Mean Male = 56.46

09.01.2025

Seite 227

Two-way ANOVA – example (1)

- Field (2013): Testing the effects of Alcohol and Gender on 'the beer-goggles effect':
 - IV 1 (Alcohol): None, 2 pints, 4 pints
 - IV 2 (Gender): Male, Female
- Dependent Variable (DV) was an objective measure of the attractiveness of the partner selected at the end of the evening.

TABLE 13.4 Data for the beer-goggles effect

Alcohol Gender	None	2 Pints	4 Pints
Female	65	70	55
Male	55	65	30
Female	60	80	70
Male	65	70	55
Female	60	70	55
Male	55	75	80
Female	60	75	80
Male	55	65	50
Total	405	535	500
Mean	60.625	66.875	57.50
Variance	24.55	196.70	59.00

Grand mean: 58.33

09.01.2025

Seite 225

Two-way ANOVA – example (4)

- Step 2b: Calculate SS_B**

$$SS_{alcohol} = 16(63.75 - 58.33)^2 + 16(64.6875 - 58.33)^2 + 16(46.5625 - 58.33)^2 = 16(5.42)^2 + 16(6.3575)^2 + 16(-11.7675)^2 = 470.0224 + 646.6849 + 2215.5849 = 3332.292$$

- Step 2c: Calculate SS_{AxB}**

$$SS_{AxB} = SS_M - SS_A - SS_B$$

$$SS_{AxB} = SS_M - SS_A - SS_B = 5479.167 - 168.75 - 3332.292 = 1978.125$$

B ₁ : None	B ₂ : 2 Pints	B ₃ : 4 Pints
65	50	55
70	55	30
60	80	70
60	65	55
60	70	55
55	75	60
60	75	50
55	65	50
Mean None = 63.75	Mean 2 Pints = 64.6875	Mean 4 Pints = 56.5625

09.01.2025

Two-way ANOVA – example (5)

- Step 3: Calculate SS_R

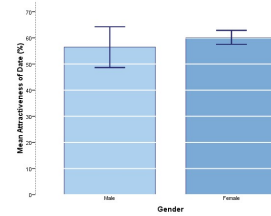
$$SS_R = s_{\text{group1}}^2(n_1 - 1) + s_{\text{group2}}^2(n_2 - 1) + s_{\text{group3}}^2(n_3 - 1) \dots s_{\text{group n}}^2(n_n - 1)$$

$$\begin{aligned} SS_R &= s_{\text{group1}}^2(n_1 - 1) + s_{\text{group2}}^2(n_2 - 1) + s_{\text{group3}}^2(n_3 - 1) \\ &\quad + s_{\text{group4}}^2(n_4 - 1) + s_{\text{group5}}^2(n_5 - 1) + s_{\text{group6}}^2(n_6 - 1) \\ &= (24.55 \times 7) + (106.7 \times 7) + (42.86 \times 7) \\ &\quad + (156.7 \times 7) + (50 \times 7) + (117.41 \times 7) \\ &= 171.85 + 746.9 + 300 + 1096.9 + 350 + 821.87 \\ &= 3487.52 \end{aligned}$$

09.01.2025

Seite 229

Interpretation: Main Effect Gender



There was a no significant main effect of gender on the attractiveness of selected mates, $F(1, 42) = 2.03, p = .161$.

09.01.2025

Seite 232

Two-way ANOVA – example (6)

Tests of Between-Subjects Effects

Dependent Variable: Attractiveness of Date					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Gender	168.750	1	168.750	2.032	.161
Alcohol	3332.292	2	1666.146	20.065	.000
Gender * Alcohol	1978.125	2	989.062	11.911	.000
Error	3487.500	42	83.036		

a. R Squared = .611 (Adjusted R Squared = .565)

F-statistic for each factor (gender, alcohol, gender*alcohol):

$$\frac{MS_{\text{factor}}}{MS_{\text{error}}} \quad \text{F-distributed}$$

09.01.2025

Seite 230

Interpretation: Interaction

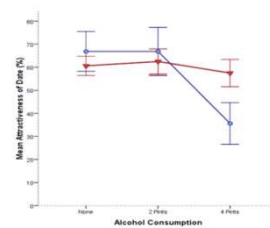


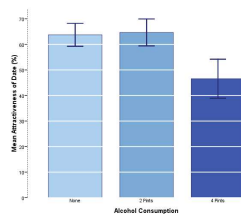
FIGURE 13.13
Graph of the interaction of gender and alcohol consumption in mate selection

There was a significant interaction between the amount of alcohol consumed and the gender of the person selecting a mate, on the attractiveness of the partner selected, $F(2, 42) = 11.91, p < .001$.

09.01.2025

Seite 233

Interpretation: Main Effect Alcohol

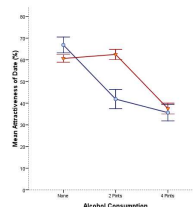


There was a significant main effect of the amount of alcohol consumed at the night-club, on the attractiveness of the mate that was selected, $F(2, 42) = 20.07, p < .001$.

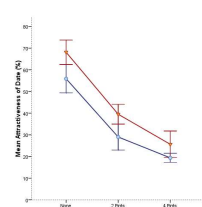
09.01.2025

Seite 231

Is there likely to be a significant interaction effect?



Yes



No

09.01.2025

Seite 234

Fallzahlsschätzung



- Je größer die Fallzahl desto geringer die Wahrscheinlichkeit für den Fehler 2. Art, desto höher die Präzision der Schätzung
- Fallzahlsschätzung bedingt bereits die Definition des Testproblems durch Operationalisierung der Fragestellung, durch Formulierung von Null- und Alternativhypothese
- Fallzahlsschätzung bedingt das Festlegen einer speziellen Alternative. „Es gibt keinen Unterschied zwischen den Gruppen“ genügt nicht. Die Größe des zu erwartenden Unterschieds muss festgelegt werden.
- Fallzahlsschätzung bedingt bereits die Auswahl des statistischen Tests, die Definition von Fehler 1. und 2. Art

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 235

Sample size estimation



Fallzahlsschätzung für unverbundene Stichproben und stetige Zielgrößen

- ☐ Fallzahlberechnung für vorgegebene Power
 - ☐ Powerberechnung für vorgegebene Fallzahl
 - ☐ Entdeckbare Differenz für vorgegebene Fallzahl und Power
- Eingabe von μ_1 : Eingabe von μ_2 :
- Eingabe von σ : Differenz Delta:
- ☐ Einseitiger Test
 - ☒ Zweiseitiger Test
- Eingabe von α (Standard ist 0.05):
- Eingabe der Power (Standard ist 0.80):
- Die Fallzahl für jede Gruppe ist:
-

Sample size estimation



Question: How many individuals do you have to include in your study to get a reliable result ?

→ We want to **maximize the probability** for rejecting H_0 , if H_1 is true

→ while keeping the **Type I error α** fixed

What do you have to know to calculate the sample size needed?

1. Power (typically set to 80% or 90%)
2. Type I error α (typically set to $\alpha = 0.05$)
3. The difference you want to find (for t-tests: the mean difference between groups)
4. standard deviation / measure of variance

Vergleich von Häufigkeiten zweier unverbundener Stichproben



In einer Karzinomstudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte Rezidivrate von 50% mit einer neuen Therapie (T) auf 40% zu senken.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Rezidivrate' sowie den weiteren Festlegungen (zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung: $p_1 = 0.5$, $p_2 = 0.4$, $\alpha = 0.05$, $1-\beta = 0.80$.

ergibt eine Fallzahl von 388 Patienten für jede Behandlungsgruppe.

Sample size estimation



Example

- Hypothesis: $H_0: \mu_A = \mu_B$ versus $H_1: \mu_A \neq \mu_B$ → two-sided t-test
 - You consider a difference of 10 as relevant
 - From former studies, you know, that the standard deviation is ~ 15 mmHG
 - So far, you have recruited 20 patients (10 in each treatment arm)
- What is your power?

<http://campus.uni-muenster.de/fileadmin/einrichtung/imib/lehre/skripte/biomathe/bio/fallz.html>

Vergleich von Mittelwerten zweier unverbundener Stichproben



In einer Hypertoniestudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte durchschnittliche Blutdrucksenkung von 15 mm mit einer neuen Therapie (T) auf 20 mm zu senken.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Senkung des Blutdrucks' sowie den weiteren Festlegungen (Standardabweichung = 15, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung: $\mu_1 = 15$, $\mu_2 = 20$, $\sigma = 15$, $\alpha = 0.05$, $1-\beta = 0.80$

ergibt eine Fallzahl von 142 Patienten für jede Behandlungsgruppe.

Vergleich von Mittelwerten zweier verbundener Stichproben



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

In einer Phase-II-Studie soll überprüft werden, ob sich ein neues Medikament zur Blutdrucksenkung eignet. Geeignet ist das Medikament dann, wenn bei Hypertonikern mit einem durchschnittlichen systolischen Blutdruck von 150 mm eine Senkung um mindestens 10 mm erreicht wird.

Für diese klinische Studie lautet das Zielkriterium 'Differenz des Blutdrucks vor und nach Behandlung'. Mit den weiteren Festlegungen (Standardabweichung = 15, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%) ergeben sich folgende Parameter für die Fallzahlberechnung:
 $\mu_1 = 150$, $\mu_2 = 140$, $\sigma = 15$, $\alpha = 0.05$, $1-\beta = 0.80$

ergibt eine Fallzahl von 20 Patienten.

Übung: Berechnen Sie die statistische Power



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Table 2. Effect of Cytisine on Smoking Cessation.*

Outcome	Cytisine (N=370)	Placebo (N=370)	Percentage-Point Difference (95% CI)	Relative Rate (95% CI)†
percent (number)				
Primary outcome: abstinence for 12 mo	8.4 (31)	2.4 (9)	6.0 (2.7–9.2)§	3.4 (1.7–7.1)
Abstinence for 6 mo	10.0 (37)	3.5 (13)	6.5 (2.9–10.1)§	2.9 (1.5–5.3)
Point prevalence at 12 mo	13.2 (49)	7.3 (27)	5.9 (1.6–10.3)§	1.8 (1.2–2.8)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 244

Vergleich von Überlebenszeiten zweier unverbundener Stichproben



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

In einer Karzinomstudie erhofft man sich, die bisher mit der Standardtherapie (S) erreichte mediane Überlebenszeit von 36 Monaten mit einer neuen Therapie (T) auf 48 Monate zu erhöhen.

Für eine kontrollierte klinische Studie mit den beiden Therapiegruppen S und T und dem Zielkriterium 'Überlebenszeit' sowie den weiteren Festlegungen (Rekrutierungszeit=24 Monate, Nachbeobachtungszeit = 36 Monate, zweiseitiger Test, Irrtumswahrscheinlichkeit 5%, Power 80%)

ergibt eine Fallzahl von 349 Patienten für jede Behandlungsgruppe.

Kategoriale Daten und 4-Felder Tafeln



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Dr. Hanno Ulmer
 hanno.ulmer@imed.ac.at
 Innsbruck, Oktober 2010

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Checklist für Fallzahl-schätzung (Testproblem)



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

- Studiendesign mit
- Auswahl des Hauptzielkriteriums
- Zu erwartender Unterschied und Angabe eines Variationsmaßes
- Begründung dafür – Literatur oder Vorstudie
- Fehler 1. Art (üblicherweise 0,05)
- Fehler 2. Art (0,1 oder 0,2)
- Auswahl des statistischen Tests
- Falls mehrere Hypothesen formuliert werden, Korrektur des Fehler 1. Art oder Hierarchisierung der Hypothesen
- Drop-Out Rate berücksichtigen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 243

Zusammenhänge qualitativ: Vierfelder Tafel



MEDIZINISCHE UNIVERSITÄT INNSBRUCK

Interventionsstudie

Tab. 1 Vierfeldertafel zur Untersuchung des Effekts einer intensivierten Insulintherapie auf die Entwicklung einer Neuropathie in 5 Jahren bei 622 Diabetikern (7)

		Neuropathie		Summe
		ja	nein	
Gruppe	Kontrolle	52	255	307
	Intervention	21	294	315
Summe		73	549	622

Diagnostische Studie

Tab. 2 Vierfeldertafel zur Untersuchung des Hämoccult-Test zum Screening auf ein kolorektales Karzinom bei 7493 Personen (1)

		kolorektales Karzinom		Summe
		ja	nein	
Hämoccult-Test	+	22	418	440
	-	10	7043	7053
Summe		32	7461	7493

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 246

Therapiestudie, epidemiologische Studie



	Outcome positiv	Outcome negativ	
Exposition positiv	a	b	Rel. Risiko = $\frac{a/b}{c/d}$
Exposition negativ	c	d	Odds Ratio = $\frac{a*d}{b*c}$
	Abs. Risiko Reduk. = $\frac{a}{a+b} - \frac{c}{c+d}$	Number needed to treat (NNT) = 1/ARR	

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 247

Logistische Regressionsanalyse



- Untersuchung / Modellierung von prognostischen Faktoren für binäre Outcomes bei gleichzeitiger Adjustierung für Störfaktoren („confounders“)
- Multivariate Erweiterung zur Kontingenztafelanalyse und Chi-Quadrat Test
- Berechnet wird das adjustierte Relative Risiko oder das adjustierte Odds Ratio

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 250

Tab. 1 Vierfeldertafel zur Untersuchung des Effekts einer intensiven Insulintherapie auf die Entwicklung einer Neuropathie in 5 Jahren bei 622 Diabetikern (7)

		Neuropathie ja	Neuropathie nein	Summe
Gruppe	Kontrolle	52	255	307
	Intervention	21	294	315
Summe		73	549	622



	Outcome positiv	Outcome negativ	
Exposition positiv	52 (16,9%)	255	RR = 2,5 $= \frac{a/b}{c/d}$
Exposition negativ	21 (6,7%)	294	OR = 2,9 $= \frac{a/c}{b/d} = \frac{a*d}{b*c}$
	ARR = 10,2% $= \frac{a}{a+b} - \frac{c}{c+d}$	NNT = 9,8 NNT = 1/ARR	

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 248

Kurzgefasst: Mit Hilfe der multiplen logistischen Regression lässt sich der Einfluss erklärender Variablen (Risikofaktoren) auf eine binäre Zielvariable (z.B. Krankheit ja/nein) untersuchen. Aus den Regressionskoeffizienten lassen sich adjustierte Odds Ratios als Maß für die Stärke des Zusammenhangs berechnen.



Tab. 2 Multiple logistische Regressionsanalyse für die Entwicklung einer diabetischen Nephropathie nach 6 Jahren bei 480 Typ 1 Diabetikern.

Risikofaktor	Regressionskoeffizient	Standardfehler	p-Wert	Differenz für Odds Ratio	Odds Ratio	95% Konfidenzintervall
Altersanstieg	-8,980	1,736	0,0001			
HbA _{1c}	+0,464	0,091	0,0001	1%	1,59	1,33 - 1,90
diast. Blutdruck	+0,048	0,019	0,0148	5 mm Hg	1,27	1,05 - 1,54
Diabetisdauer	+0,004	0,018	0,8220	5 Jahre	1,02	0,85 - 1,22
Geschlecht	-0,025	0,249	0,9212	männl. vs. weibl.	0,98	0,60 - 1,59

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 251

Relative Risiken



- **Relatives Risiko (RR):** Das Verhältnis der Risiken für ein bestimmtes Ereignis (z.B. Neuropathie) in zwei Vergleichsgruppen
- **Odds ratio (OR):** «Odds» entspricht der «Chance», dass ein bestimmtes Ereignis eintritt, bezeichnet also etwas Ähnliches wie ein Risiko, wird aber (wie beim Pferderennen) als Verhältnis (z.B. 1:9) angegeben.
- **Hazard Ratio (HR):** «Hazard» bezeichnet die Wahrscheinlichkeit, dass ein bestimmtes Ereignis eintritt. Dabei wird der Zeitpunkt berücksichtigt, wann das Ereignis eintritt. Im Gegensatz zum relativen Risiko wird also mit einer «hazard ratio» nicht nur ein Ausbleiben, sondern auch ein späteres Eintreffen eines Ereignisses als Effekt erfasst.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 249

Diagnostische Studie



	Goldstandard positiv	Goldstandard negativ	
Neues Verfahren positiv	a	b	Pos. Präd. W. $= \frac{a}{a+b} * 100$
Neues Verfahren negativ	c	d	Neg. Präd. W. $= \frac{c}{c+d} * 100$
	Sensitivität $= \frac{a}{a+c} * 100$	Spezifität $= \frac{d}{b+d} * 100$	

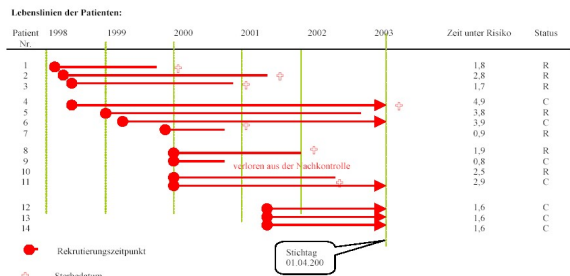
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 252

Zensierte Daten

Kaplan-Meier-Schätzung der Überlebensfunktion an einem Beispiel mit 14 Patienten



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 259

Log-Rank Test

- Vergleich von Überlebenszeiten
 - H_0 : Es besteht eine Gleichverteilung der Überlebenszeiten
 - H_1 : Die Überlebenszeiten sind unterschiedlich verteilt
- Einschränkungen
 - Probleme bei Überschneidungen der Kurven
 - Alternative Peto-Test

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 262

Überlebenszeitanalyse (Survival Analysis)

- Kaplan-Meier Methode

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53 457-481.

- Log-Rank Test
- Cox Proportional Hazards Modelle

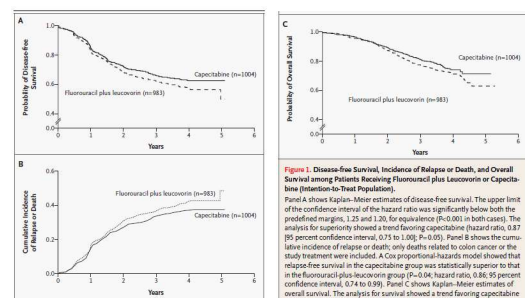
Cox, D. R. (1972). Regression Models and Life Tables (with discussion). J. R. Statistic. Soc. 34: 187-220.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 260

Disease-free/Overall Survival Incidence of Relapse/Death



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 263

Kaplan-Meier Methode

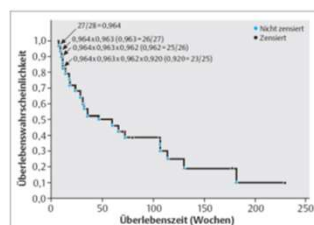


Abb.1 Kaplan-Meier Kurve für die Überlebenszeit der 28 Zungenkrebspatienten mit diploidem Tumor. Es wird die Wahrscheinlichkeit gezeigt, dass ein Patient eine Zeit (in Wochen) überlebt.

Tab. 1 Überlebenszeit (Wochen) von 28 Männern mit Zungenkrebs diploidem DNA-Tumortyp - Daten aus Table 1

Verstorben	Zensiert
1	18
2	23
3	26
4	27
5	30
6	42
7	56
8	62

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 261

Cox Proportional Hazards Regressionsanalyse

- Untersuchung / Modellierung von prognostischen Faktoren für Ereigniszeiten bei gleichzeitiger Adjustierung für Störfaktoren („Confounder“)
- Multivariate Erweiterung zu Kaplan-Meier und Log-Rank Test
- Berechnet wird die (adjustierte) Hazard Ratio als Maß für das relative Risiko

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 264

COX Modell für Versuche bis zur Lebendgeburt

Variablen	Signifikanz	Hazard Ratio	(95% CI)
Alter 30 – 34,9 Jahre	0,008	0,82	(0,70 - 0,95) vs. <30J
Alter 35 - < 39,9 Jahre	<,001	0,58	(0,46 - 0,65)
Alter 40 +	<,001	0,15	(0,11 - 0,22)
Blastozystentransfer	<,01	2,13	(1,90 - 2,40)
Endometriose	0,51	0,95	(0,80 - 1,11)
PCO	0,24	0,92	(0,80 - 1,06)

Mit Hilfe des Cox-Modells läßt sich der Einfluß von erklärenden Variablen auf eine Ereigniszeit untersuchen. Aus den Regressionskoeffizienten lassen sich adjustierte Hazard Ratios für die Stärke des Zusammenhangs berechnen.

09.01.2025

hanno.ulmer@i-med.ac.at

265

DOE Grundlagen

- Input: Beeinflussbare Faktoren x_1, x_2, x_3
- Output Y: Response, wird gemessen
- Noise: bekannte oder unbekannte Störungen
- Bestimmung der Koeffizienten c_1, c_2, c_3 mittels Varianzanalyse (ANOVA)

$$\text{Modell } y = f(x_1, \dots, x_n) = c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_{12} \cdot x_1 \cdot x_2 + \dots + c_{11} \cdot x_1^2$$



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 268

Design of Experiments

Dr. Hanno Ulmer

hanno.ulmer@imed.ac.at
Innsbruck, Oktober 2010

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Geschichte von DOE

- **The agricultural origins, 1918 – 1940s**
 - R. A. Fisher & his co-workers
 - Profound impact on agricultural science
 - Factorial designs, ANOVA
- **The first industrial era, 1951 – late 1970s**
 - Box & Wilson, response surfaces
 - Applications in the chemical & process industries
- **The second industrial era, late 1970s – 1990**
 - Quality improvement initiatives in many companies
 - Taguchi and robust parameter design, process robustness
- **The modern era, beginning circa 1990**
 - Wide use of computer technology in DOE
 - Expanded use of DOE in Six-Sigma and in business
 - Use of DOE in computer experiments

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 269

DOE Grundlagen

- DOE = Design of Experiments
- Idee:
 - Simultane Bewertung von mehreren Einflussfaktoren
 - Auswahl der Zielvariable
 - Planung des Experiments, Anzahl der Durchläufe
 - Verwendung von ANOVA/Regressionsanalyse
- Anwendungen: Naturwissenschaftliche Experimente, Marktforschung, Simulationsexperimente
- Software (MODDE, SAS JMP,...) erlaubt
 - Auswahl des experimentellen Designs
 - Analyse der Ergebnisse des Experiments

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 267

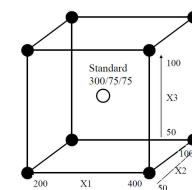
CakeMix Beispiel mit MODDE

Overview of DOE - CakeMix application

- Three factors varied: Flour (200-400g), Shortening (50-100g), and Eggpowder (50-100g)
- Response: Taste of resulting cake

Cake Mix Experimental Plan

Cake No	Flour	Shortening	Egg Powder	Taste
1	200	50	50	3.52
2	400	50	50	3.66
3	200	100	50	4.74
4	400	100	50	5.20
5	200	50	100	5.38
6	400	50	100	5.90
7	200	100	100	4.36
8	400	100	100	4.86
9	300	75	75	4.73
10	300	75	75	4.61
11	300	75	75	4.68



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 270

CakeMix Beispiel mit MODDE



Overview of steps in DOE - part I

1. Define Factors

Name	Alias	Units	Type	Units	Settings	Transform	Pre
1. Temperature	T	°C	Quantitative	Controlled: 20 to 400	None	Free	
2. Humidity	H	%	Quantitative	Controlled: 50 to 100	None	Free	
3. Humidity	H	%	Quantitative	Controlled: 50 to 100	None	Free	

2. Define Response(s)

Name	Alias	Units	Transform	MSD Scale	YLS Scale	Type
1. Error	T	None	None	Unit Variance	Regular	Y

3. Create Design (Make experiments)

Exp. No.	Exp. Name	Run Order	Temp	Humidity	Humidity	Temp
1	1	1	20	50	50	2.02
2	2	2	40	50	50	3.44
3	3	3	20	100	100	4.74
4	4	4	40	100	100	5.1
5	5	5	20	50	100	5.19
6	6	6	40	50	100	5.2
7	7	7	20	100	50	4.19
8	8	8	40	100	50	4.19
9	9	9	20	50	50	4.19
10	10	10	40	50	50	4.19
11	11	11	20	50	75	4.19
12	12	12	40	50	75	4.19

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 271

DOE Beispiel Crash Test



Beispiel Crash-Test



- Jeder Faktor hat mehrere Level
 - Stetige (z.B. Geschwindigkeit: 0 – 100 km/h)
 - Diskrete (z.B. Airbag: an/aus)
- Einteilung der stetigen Level in sinnvolle Stufen
 - Z.B. Geschwindigkeit niedrig, mittel, hoch
- Zielgröße: Kraftwirkung auf Dummy.
- Faktoren, Level, gewünschte Genauigkeit werden im Versuchsplan festgehalten

09.01.2025

hanno.ulmer@i-med.ac.at

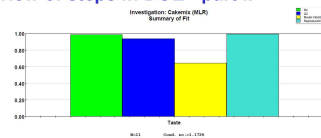
Seite 274

CakeMix Beispiel mit MODDE

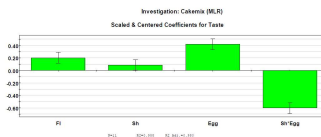


Overview of steps in DOE - part II

4. Make Model



5. Interpret Model



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 272

Statistik und Gender Medicine



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Dr. Hanno Ulmer
hanno.ulmer@imed.ac.at

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Designs



- Full Factorial Design:
Alle Level Kombinationen werden getestet
- Latin Square Design:
nur ein zufällige Teilmenge aller Level-Kombinationen werden getestet
- Plackett-Burman Designs:
sehr effizient, große Anzahl von Hauptfaktoren wird mit so wenig Durchgängen wie möglich untersucht
- Box-Behnken, Central Composite, Orthogonal Arrays, Center Point Designs, etc.

09.01.2025

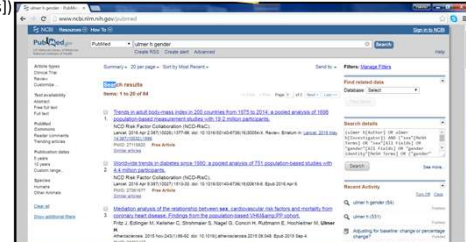
hanno.ulmer@i-med.ac.at

Seite 273

Pubmed



(ulmer h[Author] OR ulmer h[Investigator]) AND ("sex"[MeSH Terms] OR "sex"[All Fields] OR "gender"[All Fields] OR "gender identity"[MeSH Terms] OR "gender identity"[All Fields] AND "identity"[All Fields] OR "gender identity"[All Fields])



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 276

Verwendung

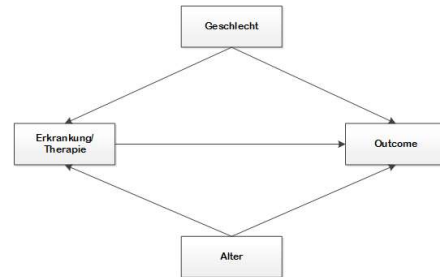
- sex-specific
- by sex
- age-sex-groups
- Independent of sex
- sex-matched
- after adjustment for sex, adjusting for sex
- stratified by sex
- age- and sex standardized

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 277

DAG: Geschlecht und Alter als Confounder



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 280

Verwendung

- Auswertung separat für Männer und Frauen, Unterschiede zwischen Männer und Frauen
sex-specific, by sex, sex-groups
- Geschlecht als Störvariable (Confounder) und Berücksichtigung dieses Einflusses
adjusting for sex, independent of sex, stratified by sex, sex standardized
- Im Studiendesign (z.B. Fall-Kontroll Studie) wird jedem männlichen Fall eine männliche Kontrolle und jedem weiblichen Fall eine weibliche Kontrolle zugewiesen
sex-matched
- Randomisierung führt zu gleicher Anzahl von Männern und Frauen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 278

Confounding/Modification/Mediation

- Epidemiologen sprechen von Verschleierung oder **Confounding**, wenn die Assoziation zwischen einer Exposition und einem Outcome durch eine Störgröße überlagert oder verzerrt wird. Diese Störgröße heißt Confounder (Zitat aus Razum O et al. Epidemiologie für Dummies).

Zu unterscheiden von:

- Wenn sich die Stärke einer Exposition und einem Outcome verändert wenn eine oder mehrere weitere Variablen hinzukommen, dann spricht man von **Effektmodifikation** (Cholesterin Beispiel).
- Intermediärvariablen (**Mediatoren**) sind Zwischenstufen in der Kausalkette Exposition und Outcome (KHK Beispiel).

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 281

Für den Statistiker

- Ist Geschlecht in der Regel oft ein Störfaktor (Confounder) oder Effektmodifikator (effect modifier)
- Alter und Geschlecht sind die wichtigsten Einflussfaktoren in der Medizin,
- deren Einfluss muss entweder im Studiendesign durch Randomisierung und Matching oder nachträglich in der Analyse durch separate Auswertung, Standardisierung, Adjustierung oder Stratifizierung (Gewichtung) berücksichtigt werden
- Voraussetzung: Variable Geschlecht muss vorhanden sein!

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 279

Geschlechtsunterschiede und Kausalität

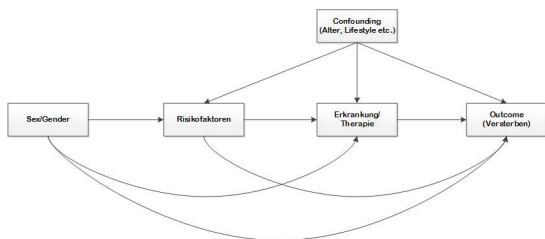
- Geschlechtsunterschiede können nur beobachtet werden
- Geschlechtsunterschiede können zwar in einem RCT beobachtet werden, aber nicht per se durch einen RCT untersucht werden
- Der Faktor Geschlecht kann nicht randomisiert werden
- Aber: Geschlecht ist ab Geburt defacto natürlich randomisiert
- Für die Gender Medizin Forschung gelten die Limitationen der Beobachtungsstudien,
Cave: Selection Bias und Informations Bias!
- Confounding ist per Definition nicht möglich
- Außer Frage stehende Fakten: Männer Prostatakrebs, etc.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 282

DAG: Geschlecht als Untersuchungsobjekt



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 283

Systematische Fehler, Confounding

- **Selection Bias**
 - Nichterreichen von Berufstätigen bei Telefonumfragen
- **Information Bias**
 - Fehlklassifikation / Fehldiagnosen,
 - Messfehler
- **Confounding**
 - Mangelnde Berücksichtigung von Störgrößen
 - Scheinassoziation von Alkohol und Lungenkarzinom über Rauchen erklärbar

09.01.2025

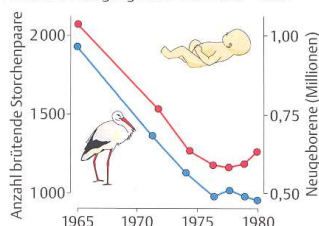
hanno.ulmer@i-med.ac.at

Seite 286

Glauben Sie den Ergebnissen einer Studie?

- **Konzepte zur Kausalität z.B. von** Henle&Koch (1880), Hill (1965) oder Rothman (1976)
- **Assoziation** (negativer oder positiver Zusammenhang)
- **Kausalität** (Risikofaktor als (Mit-)Ursache einer Krankheit)

Korrelation Abnahme brütender Storchpaare/ Geburtenrückgang in der BRD 1965–1980



hanno.ulmer@i-med.ac.at

Seite 284

Systematische versus zufällige Fehler

- Systematische Fehler führen zu einer Verzerrung (Bias) der Effektschätzer (RR, OR etc.)
- Zufällige Fehler (durch zu geringe Fallzahl, siehe Kapitel Fallzahlschätzung) erniedrigen die Präzision
- Systematische Fehler erniedrigen die Validität
- Die Validität hat Priorität vor der Präzision.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 287

Statistische Assoziation oder Kausalität ??

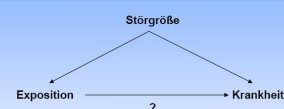
- Kausalitätskriterien nach
- Sir Austin Bradford-Hill:
 - Temporalität
 - Konsistenz (Meta-Analysen, Systematische Reviews)
 - Biologischer Gradient
 - Stärke des Effekts (z.B. doppeltes Risiko)
 - u.a.



hanno.ulmer@i-med.ac.at

Seite 285

Störgrößen (Confounder)



z.B. unterschiedliches Alter in den Therapiegruppen
Alter beeinflusst den Blutdruck und womöglich die Therapie.

Mögliche Lösung: Adjustierung für Alter
mittels multivariater Analyse

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 288

Simpson's Paradoxon Reserpin-Beispiel

		Brustkrebs		
		ja	nein	
Reserpin	ja	32	57	32/89 = 36 %
	nein	149	351	149 / 500 = 30 %

OR = (32/57) / (149/351)
= 0,56 / 0,42 = 1,3

28

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 289

Logistische Regressionsanalyse



Modell mit Reserpin:

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)
reserpin	,280	,242	1,339	1	,247	1,323	Lower .824 Upper 2,123

Modell mit Reserpin und Alter:

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)
reserpin	-,179	,255	,491	1	,483	,836	Lower ,507 Upper 1,379
alter50	1,474	,205	51,497	1	,000	4,367	2,920 6,532

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 292

Alter ≤ 50

		Brustkrebs		
		ja	nein	
Reserpin	ja	2	14	2/16 = 13 %
	nein	42	221	42 / 263 = 16 %

OR = 0,14 / 0,19 = 0,75

Alter > 50

		Brustkrebs		
		ja	nein	
Reserpin	ja	30	43	30/73 = 41 %
	nein	107	130	107 / 237 = 45 %

OR = 0,70 / 0,82 = 0,85

29

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 290

Logistische Regression



Untersuchungsziel: hat eine/mehrere (unabhängige) Variable/n X
einen Einfluss auf eine andere (abhängige) Variable Y?

- ▲ Abhängige Variable ist binär (Ausprägungen z. B. ja/nein)
- ▲ Unabhängige Variablen sind intervallskaliert oder als Dummy-Variablen codiert
- ▲ Unterschied zur linearen Regression: Y kann nur die Werte 0 oder 1 annehmen
- ▲ Hintergrund der logistischen Regression: Untersuchung des Zusammenhang $p = P(Y=1)$ und der unabhängigen Variable(n) X
- ▲ Es wird nicht der Wert der abhängigen Variablen vorhergesagt, sondern die Wahrscheinlichkeit, dass die abhängige Variable den Wert 1 annimmt

Cave: erwarteter kausaler Zusammenhang zwischen unabhängiger und abhängiger Variable muss theoretisch erklärbar sein

2

Logistische Regressionsanalyse



- Untersuchung / Modellierung von prognostischen Faktoren für binäre Outcomes bei gleichzeitiger Adjustierung für Störfaktoren („confounders“)
- Multivariate Erweiterung zur Kontingenztafelanalyse und Chi-Quadrat Test
- Berechnet wird das adjustierte Odds Ratio
- Cytisine for smoking cessation: Adjustment for all baseline characteristics shown in Table 1 had a negligible effect.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 291

Logistische Regression – statistischer Hintergrund



- ▲ Logistische Regressionsanalyse basiert auf der Maximum-Likelihood-Schätzung (MLE) und unterscheidet sich von der Methode der kleinsten Quadrate (lineare Regressionsanalyse)
- ▲ Ziel der Analyse: Identifikation einer Funktionskurve zu finden, die möglichst gut zu den Daten passt
- ▲ Funktion ist eine logistische Funktion (bei linearen Regressionsanalyse eine Gerade)
- ▲ Werte der logistischen Funktion werden als Wahrscheinlichkeit interpretiert (dass die abhängige Variable y den Wert 1 annimmt – gegeben die unabhängigen Variablen x_i)
- ▲ Wert nahe bei 0 bedeutet, dass das Eintreten von y (y = 1) sehr unwahrscheinlich ist; Wert nahe bei 1, dass das Eintreten von y sehr wahr

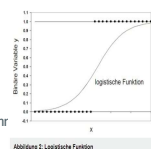


Abbildung 2: Logistische Funktion

2

Logistische Regression – statistischer Hintergrund



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

▲ Logistische Regressionsfunktion:

$$P(y=1) = \frac{1}{1+e^{-z}}$$

$P(y=1)$ = Wahrscheinlichkeit, dass $y = 1$
 e = Basis des natürlichen Logarithmus, Eulersche Zahl
 z = Logit (lineares Regressionsmodell der unabhängigen Variablen)

z , der sogenannte "Logit", stellt dabei ein lineares Regressionsmodell dar:

$$z = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \dots + \beta_k \cdot x_k + \epsilon$$

x_k = unabhängige Variablen

β_k = Regressionskoeffizienten

ϵ = Fehlerwert

$$P(y=1) = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + \epsilon)}}$$

2

Logistische Regression – statistische Signifikanz



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

▲ 1. Schritt: Überprüfen ob das Regressionsmodell insgesamt signifikant ist: Chi-Quadrat-Test

▲ Prüft ob das Modell insgesamt einen Erklärungsbeitrag leistet

▲ Modellgüte: Passung zwischen Modell und Daten ("Goodness of fit"): Analog zum R-Quadrat der linearen Regression – verschiedene Pseudo-R-Quadrate

▲ 2. Schritt: Überprüfen ob Regressionskoeffizienten (Betas) ebenfalls signifikant sind

Wald-Test für jeden der Regressionskoeffizienten

Variablen in der Gleichung									
	Regressionskoeffizient B	Standardfehler	Wald	df	Sig.	Exp(B)	95% Konfidenzintervall für Exp(B)		
Schritt 1 ^a									
Einkommen	.022	.008	14.651	1	.000	.979	.968	.990	
Risikobereitschaft	.340	.008	15.541	1	.000	1.416	1.181	1.683	
Interesse	.085	.018	23.036	1	.000	1.089	1.052	1.127	
Konstante	-1.668	.279	35.731	1	.000	.189			

a. In Schritt 1 eingetragene Variablen: Einkommen, Risikobereitschaft, Interesse.

▲ Exp(B) = entlogarithmierter logit-Koeffizienten – wenn Konfidenzintervall von Exp(B) den Wert 1 nicht einschließt: signifikanter Einfluss

▲ Risikobereitschaft: Exp(B) > 1 positiver Zusammenhang: steigt Risikobereitschaft um eine Einheit steigt die relative Wahrscheinlichkeit, dass eine Person bereits einmal Aktien gekauft hat, um 41.6% (1.416 – 1 = .416)

2
9
8

Logistische Regression - Interpretation



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

▲ Zusammenhang zwischen unabhängigen Variablen und abhängiger Variable wird mittels sogenannter "Odds" interpretiert

▲ Odds: Wahrscheinlichkeit, dass das Ereignis eintritt, in Relation zum Nichteintreffen des Ereignisses

$$Odds = \frac{P(y \text{ trifft ein})}{P(y \text{ trifft nicht ein})} = \frac{P(y \text{ trifft ein})}{1 - P(y \text{ trifft ein})}$$

Odds Ratio = $Exp(B) = e^B$ = Odds nach dem Anstieg von x um eine Einheit
 Odds vor dem Anstieg von x um eine Einheit

$$= \frac{Odds_{nach}}{Odds_{vor}}$$

2

Prädiktionsmodelle



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

"Das Entwickeln und Validieren von Prädiktionsmodellen hat sich in den letzten 15 Jahren methodisch weiterentwickelt.

Standardlehrbuch ist Ewout Steyerberg's Clinical Prediction Models

mit "checklist for developing valid prediction models".

Wichtige Themen sind "model performance", "discrimination", "calibration", und "clinical usefulness".

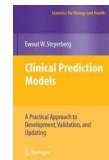
Reporting von Prädiktionsstudien:

TRIPOD statement: "transparent reporting of a multivariable prediction model for individual prognosis or diagnosis".

Br J Surg. 2015 Feb;102(3):148–58. doi: 10.1002/bjs.9736.

Empfehlenswert ist auch diese BMJ Serie:

<https://www.bmj.com/content/338/bmj.b375>
<https://www.bmj.com/content/338/bmj.b604.long>
<https://www.bmj.com/content/338/bmj.b605.long>
<https://www.bmj.com/content/338/bmj.b606>



2
9
9

Logistische Regression - Interpretation



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

▲ Odds Ratio einer unabhängigen Variablen = Veränderung der relativen Wahrscheinlichkeit von $y = 1$ an, wenn diese unabhängige Variable um eine Einheit steigt

▲ Odds Ratio einer unabhängigen Variablen ist der Faktor, um den sich die Odds verändern, wenn diese Variable um eine Einheit ansteigt

▲ Beträgt eine Odds Ratio (Exp(B)) = 1 ergibt sich keine Veränderung ($Odds_{nach} = Odds_{vor}$)

▲ Odds Ratio > 1 ergibt eine Zunahme der Odds ($Odds_{nach} > Odds_{vor}$)

▲ Odds Ratio < 1 ergibt eine Abnahme der Odds ($Odds_{nach} < Odds_{vor}$)

▲ Zusammenhang Odds Ratios und Regressionskoeffizienten: Odds Ratio = $Exp(B) = e^B$

▲ Odds Ratio = 1 wenn Regressionskoeffizient = 0, > 1 wenn Regressionskoeffizient positiv ist, < 1 wenn Regressionskoeffizient negativ ist

2

Vergleichbarkeit: Patientenflussdiagramm



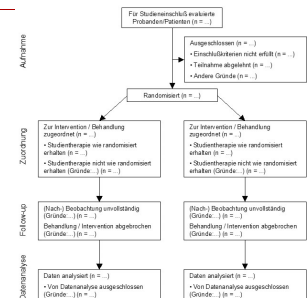
MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

• Intention-To-Treat

versus

• Per-Protocol

- Vergleichbarkeit der Gruppen:
- Strukturgleichheit
- Beobachtungsgleichheit
- Behandlungsgleichheit



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 300

Effektmodifikation (Interaktion)



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Existieren in verschiedenen Strata (Schichten) einer Variablen unterschiedliche Effektschätzer,
- so spricht man von **Effektmodifikation bzw. Interaktion**.
- Die **Schichtvariable** wird als Effektmodifikator (effect modifier) bezeichnet
- Reine Effektmodifikation führt nicht zu einer Verzerrung des Effektmaßes und gehört damit nicht zu den Fehlern in epidemiologischen Studien
- Modellierung durch Aufnahme von multiplikativen Termen in Regressionsmodellen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 301

Grundlagen der Epidemiologie

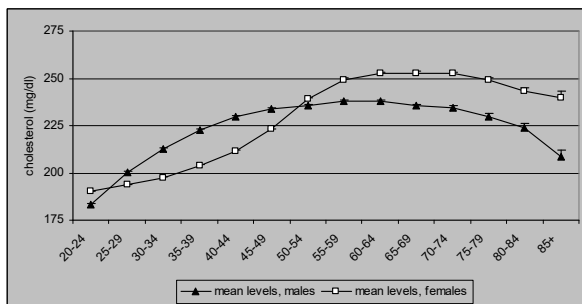


MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Hanno Ulmer

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Mittleres Gesamtcholesterin für Männer und Frauen nach Alter bei der Erstuntersuchung



302

Epidemiologie, Definition



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

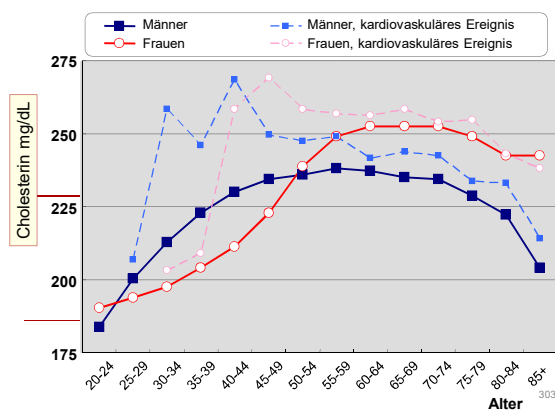
- Die **Epidemiologie** (von griech. *epi* „auf, über“, *demos* „Volk“, *logos* „Lehre“, ursprünglich: „Seuchenkunde“) ist jene wissenschaftliche Disziplin, die sich mit den Ursachen und Folgen sowie der Verbreitung von gesundheitsbezogenen Zuständen und Ereignissen in Populationen beschäftigt. Die Epidemiologie untersucht somit jene Faktoren, die zu Gesundheit und Krankheit von Individuen und Populationen beitragen und ist deshalb die Basis aller Maßnahmen, die im Interesse der Volksgesundheit unternommen werden.
- Im Gegensatz dazu kümmert sich die Medizin darum, dem einzelnen Menschen in einem konkreten Krankheitsfall zu helfen.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 305

Mittleres Gesamtcholesterin für Männer und Frauen mit kardiovaskulärer Mortalität



303

Bereiche/Teilgebiete der Epidemiologie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Infektionsepidemiologie, Epidemiologie allergischer und dermatologischer Erkrankungen, Epidemiologie der Arbeitswelt, Epidemiologische Methoden, Ernährungsepidemiologie, Genetische Epidemiologie, Herz-Kreislauf-Epidemiologie, Krebs Epidemiologie, etc.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 306

Schlüsselfragen der Epidemiologie



- Was?
 - Um welches Gesundheitsproblem geht es? Spezifizierung
- Wann?
 - Zu welchem Zeitpunkt oder in welchem Zeitraum?
- Wo?
 - An welchen Orten tritt das Problem auf?
- Wer?
 - Wer ist von dem Problem betroffen? Geschlecht, Alter, Sozialstatus...
- Warum?
 - Welche Ursachen gibt es für das Problem?

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 307

Definition des Schlaganfalls



Transitorische Ischämische Attacke

Die TIA ist ein reversibler ischämischer Schlaganfall, gekennzeichnet durch schlagartig auftretendes fokales neurologisches Defizit mit vollständiger Rückbildung in Minuten bis wenigen Stunden, spätestens innerhalb von 24 Stunden

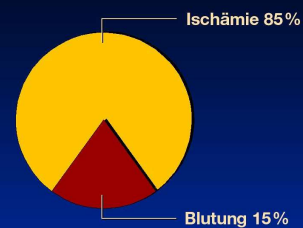
Herz-Kreislauf-Epidemiologie: Beispiel Schlaganfalls



Hanno Ulmer

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Schlaganfall



Definition des Schlaganfalls



Allgemein

Akutes fokales neurologisches Defizit auf Grund eines umschriebenen Durchblutungsmangels oder einer Blutung des Gehirns

Epidemiologie



Prävalenz, Inzidenz und Letalität des Schlaganfalls in Deutschland

Prävalenz	Inzidenz
400.000 - 600.000	120.000 - 200.000
(600-700/10.000 Einw.)*	(15-25/10.000/Jahr)

Letalität**
15.000 - 30.000/Jahr
(10 - 20%)

* > 65 Jahre

** innerh. v. 30 Tagen im Krankenhaus

Epidemiologie



Risikofaktor	Ischämie	Blutung	
		intrazerebral	subarachnoidal
Hypertonie	++	++	+
Zigarettenrauchen	++	+ -	++
Diabetes mellitus	++	0	0
Alkoholabusus	+ -	++	+
Fettstoffwechselstörung	+	0	0
Herzkrankheiten	++	0	0

++ enge Beziehung
+ mäßige Beziehung
+ - Beziehung nicht eindeutig
0 keine Beziehung

Epidemiologie

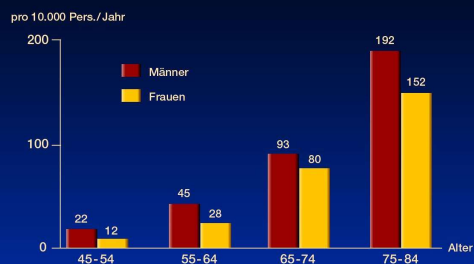


Risikofaktor	Relatives Risiko
Hypertonie	bis 5fach
Diabetes mellitus	bis 3fach
Zigarettenrauchen	bis 2fach
Alkoholabusus	bis 4fach *
Fettstoffwechselstörung	bis 3fach
Herzkrankheiten	bis 4fach
Vorhofflimmern	bis 17fach **

* unter Einschuß der Hirnblutung
** bei Kombination

Epidemiologie

Inzidenz des ersten Schlaganfalls*



Framingham-Studie

Grundbegriffe der Epidemiologie



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Hanno Ulmer

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Epidemiologie



Risikofaktor	Prävalenz
Hypertonie	25 - 40 %
Diabetes mellitus	4 - 8 %
Zigarettenrauchen	20 - 40 %
Alkoholabusus	6 - 30 %
Fettstoffwechselstörung	5 - 30 %
Herzkrankheiten	10 - 20 %
Vorhofflimmern	1 %

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Inzidenz: Maß für die Anzahl der Neuerkrankungen in
einem definierten Zeitraum

Prävalenz: Maß für die Anzahl von Erkrankten zu
einem definierten Zeitpunkt

Mortalität: Maß für die Anzahl der Todesfälle

Altersstandardisierung

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Inzidenz = Anzahl der Neuerkrankungen im Beobachtungszeitraum / Anzahl der Personen unter Risiko (zu Beginn des Zeitraums)

Beispiel: In einer Stadt leben 100.000 Frauen. Aktuell leiden 800 von ihnen an Brustkrebs. Von den anderen 99.200 erkranken im Laufe eines Jahres 110 an Brustkrebs.

Inzidenz = $110/99.200 = 0,001109$ oder 110,9 pro 100.000 Frauen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 319

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Gesamtmortalität = Anzahl der Todesfälle in einem Zeitraum / Bevölkerungsumfang

Beispiel: Im Laufe eines Jahres versterben in der Beispielstadt 100 Frauen.

Gesamtmortalität = $100/100.000 = 0,001$ oder 100 pro 100.000 Frauen

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 322

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Prävalenz = Anzahl der Erkrankungsfälle in der Bevölkerung / Bevölkerungsumfang

Beispiel: Aktuell sind 800 von 100.000 Frauen an Brustkrebs erkrankt.

Prävalenz = $800/100.000 = 0,008$ oder 800 pro 100.000 Frauen

Prävalenz = Inzidenz x Krankheitsdauer

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 320

Epidemiologische Maßzahlen

Ursachenspezifische Mortalität = Anzahl der Todesfälle nach Ursache in einem Zeitraum / Bevölkerungsumfang

Beispiel: Von den 100 Todesfällen sind 40 auf Brustkrebs zurückzuführen. Somit beträgt die

Brustkrebsmortalität = $40/100.000 = 0,0004$ oder 40 von 100.000 Frauen.

09.01.2025

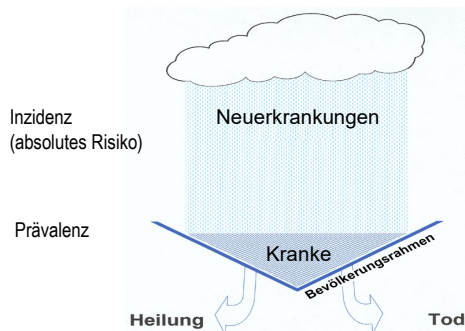
hanno.ulmer@i-med.ac.at

Seite 323

Prävalenz versus Inzidenz



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Altersspezifische Mortalität = Anzahl der Todesfälle in einer bestimmten Altersklasse / Bevölkerungsumfang in dieser Altersklasse

Beispiel: In der betrachteten Stadt sind 16.000 Frauen zwischen 55 und 60 Jahre alt. In dieser Altersklasse versterben im Laufe des Jahres 10 Frauen.

Altersspezifische Mortalität = $10/16.000 = 0,000625$ oder 6,25 von 100.000 Frauen.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 324

Epidemiologische Maßzahlen



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Letalität („Tödlichkeit einer Erkrankung“) = Anzahl der Todesfälle nach Ursache in einem Zeitraum / Anzahl der Neuerkrankungen an dieser Ursache im selben Zeitraum

Beispiel: Von den 110 neu an Brustkrebs erkrankten Frauen sterben im Laufe des Jahres 10 Frauen.

Letalität = $10/110 = 0,091$ oder 9,1%

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 325



Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle



Altersstandardisierung - Beispiel

Schritt 2: Berechnung der altersspezifischen Mortalitätsraten

Altersklasse	Kreuzberg			Zehlendorf		
	Population	Todesfälle	Mortalität*	Population	Todesfälle	Mortalität*
0-19	34.000	34	1,0	18.000	5	0,3
20-44	75.500	135	1,8	32.000	30	0,9
45-64	27.000	299	11,1	29.000	165	5,7
65 und älter	15.000	1.167	77,8	20.000	1.585	79,3
Insgesamt	151.500	1.635		99.000	1.758	

Rohe Mortalität:

Kreuzberg $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner

Zehlendorf $\frac{1.758}{99.000} = 18,0$ pro 1.000 Einwohner

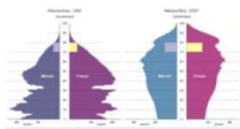
* alle Raten pro 1.000 Personen

Altersstandardisierung



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Ermöglicht den Vergleich von Bevölkerungen mit unterschiedlicher Altersstruktur, indem verzerrende Alterseinflüsse beseitigt werden.



Aufgrund des demographischen Wandels nimmt die Anzahl der über 70-jährigen stark zu. Dies muss bei temporalen Vergleichen berücksichtigt werden

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 326



Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle



Altersstandardisierung - Beispiel

Schritt 3: Wahl einer Standardpopulation

Altersklasse	Standard	Kreuzberg			Zehlendorf		
		Population	Todesfälle	Mortalität*	Population	Todesfälle	Mortalität*
0-19	385.000	34.000	34	1,0	18.000	5	0,3
20-44	850.500	75.500	135	1,8	32.000	30	0,9
45-64	540.000	27.000	299	11,1	29.000	165	5,7
65 und älter	356.000	15.000	1.167	77,8	20.000	1.585	79,3
Insgesamt	2.131.500	151.500	1.635		99.000	1.758	

Rohe Mortalität:

Kreuzberg $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner

Zehlendorf $\frac{1.758}{99.000} = 18,0$ pro 1.000 Einwohner

* alle Raten pro 1.000 Personen

Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle

Altersstandardisierung - Beispiel

Schritt 1: Berechnung der rohen Mortalitätsraten

Altersklasse	Kreuzberg			Zehlendorf		
	Population	Todesfälle		Population	Todesfälle	
0-19	34.000	34		18.000	5	
20-44	75.500	135		32.000	30	
45-64	27.000	299		29.000	165	
65 und älter	15.000	1.167		20.000	1.585	
Insgesamt	151.500	1.635		99.000	1.758	

Rohe Mortalität:

Kreuzberg $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner

Zehlendorf $\frac{1.758}{99.000} = 18,0$ pro 1.000 Einwohner

Beispiel zur Altersstandardisierung, entnommen einem Vortrag von A. Dauts, Tumorzentrum Erlangen-Nürnberg



Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle



Altersstandardisierung - Beispiel

Schritt 4: Anwendung der altersspezifischen Mortalitätsraten auf die fiktive Standardpopulation

Altersklasse	Standard	Kreuzberg			Zehlendorf		
		Population	Todesfälle	Mortalität*	Population	Todesfälle	Mortalität*
0-19	385.000	34.000	34	1,0	18.000	5	0,3
20-44	850.500	75.500	135	1,8	32.000	30	0,9
45-64	540.000	27.000	299	11,1	29.000	165	5,7
65 und älter	356.000	15.000	1.167	77,8	20.000	1.585	79,3
Insgesamt	2.131.500	151.500	1.635		99.000	1.758	

Rohe Mortalität:

Kreuzberg $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner

Zehlendorf $\frac{1.758}{99.000} = 18,0$ pro 1.000 Einwohner

* alle Raten pro 1.000 Personen

Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle

Altersstandardisierung - Beispiel

Schritt 5: Berechnung der altersstandardisierten Mortalitätsraten

Altersklasse	Standard	Kreuzberg				Zehlendorf			
		Population	Todesfälle	Mortalität	Erw. Fälle	Population	Todesfälle	Mortalität	Erw. Fälle
0-19	385.000	34.000	34	1,0	385	18.000	5	0,3	116
20-44	850.500	75.500	135	1,6	1.531	32.000	30	0,9	765
45-64	540.000	27.000	299	11,1	5.994	29.000	165	5,7	3.078
65 und älter	356.000	15.000	1.167	77,8	27.697	20.000	1.585	79,3	28.231
Insgesamt	2.131.500	151.500	1.635	16,7	35.607	99.000	1.758	16,1	32.190

Rohe Mortalität: $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner
 Altersstandardisierte Mortalität: $\frac{35.607}{2.131.500} = 16,7$ pro 1.000 Einwohner
 Kreuzberg

Rohe Mortalität: $\frac{1.785}{99.000} = 18,0$ pro 1.000 Einwohner
 Altersstandardisierte Mortalität: $\frac{32.190}{2.131.500} = 15,1$ pro 1.000 Einwohner
 Zehlendorf

* alle Raten pro 1.000 Personen

Die wichtigsten Studientypen

Dr. Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und Gesundheitsökonomie, Medizinische Universität Innsbruck

Bevölkerungsbezogenes Krebsregister Bayern - Registerstelle

Altersstandardisierung - Beispiel

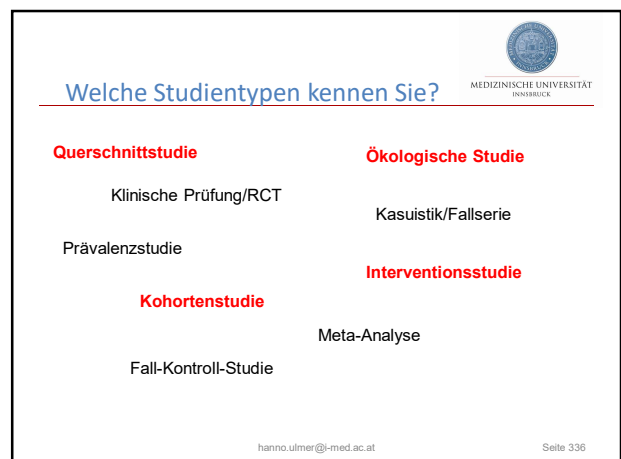
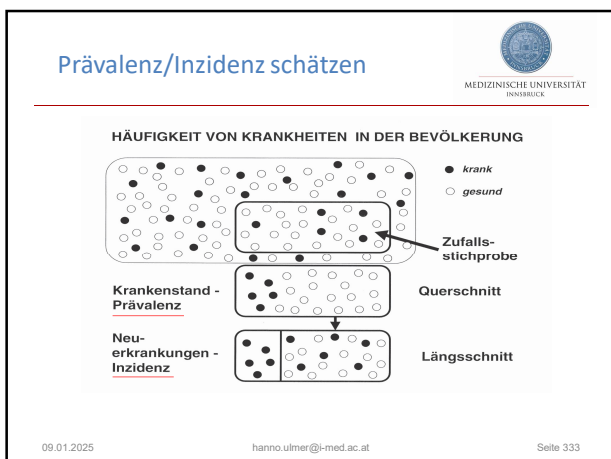
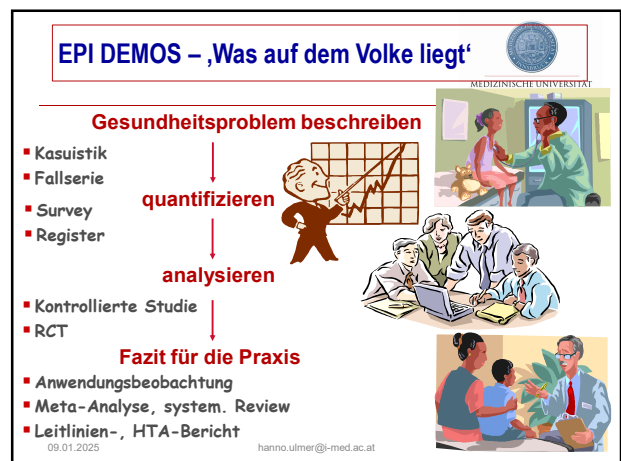
Der Vergleich der altersstandardisierten mit der rohen Mortalität zeigt die verzerrenden Alterseinflüsse.

Altersklasse	Standard	Kreuzberg				Zehlendorf			
		Population	Todesfälle	Mortalität	Erw. Fälle	Population	Todesfälle	Mortalität	Erw. Fälle
0-19	385.000	34.000	34	1,0	385	18.000	5	0,3	116
20-44	850.500	75.500	135	1,6	1.531	32.000	30	0,9	765
45-64	540.000	27.000	299	11,1	5.994	29.000	165	5,7	3.078
65 und älter	356.000	15.000	1.167	77,8	27.697	20.000	1.585	79,3	28.231
Insgesamt	2.131.500	151.500	1.635	16,7	35.607	99.000	1.758	15,1	32.190

Rohe Mortalität: $\frac{1.635}{151.500} = 10,8$ pro 1.000 Einwohner
 Altersstandardisierte Mortalität: $\frac{35.607}{2.131.500} = 16,7$ pro 1.000 Einwohner
 Kreuzberg

Rohe Mortalität: $\frac{1.785}{99.000} = 18,0$ pro 1.000 Einwohner
 Altersstandardisierte Mortalität: $\frac{32.190}{2.131.500} = 15,1$ pro 1.000 Einwohner
 Zehlendorf

* alle Raten pro 1.000 Personen



Hierarchie von Medizinischen Studien



09.01.2025

hanno.ulmer@i-med.ac.at

337

Kohortenstudie

Kohortenstudie 1

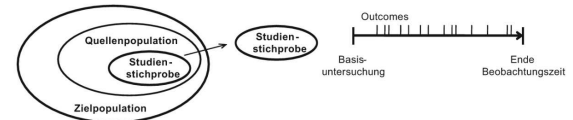


Abbildung 14.2: Grundstruktur einer Kohortenstudie.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 340

RCT: Randomisierte kontrollierte Studie

- **Klinische Studien sind ein Experiment**
 - Einflussfaktor wird gesteuert
 - Alle anderen Faktoren sollen möglichst konstant gehalten werden
- **Randomisierung**
 - Zufällige Behandlungszuteilung
 - Ausschluss von Verzerrungen (Bias) durch Selektion
- **Verblindung**
 - Ausschluss von Verzerrungen (Bias) durch Information

Modul 2.02

hanno.ulmer@i-med.ac.at

338

Kohortenstudie

Kohortenstudie 2

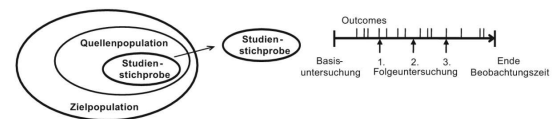


Abbildung 14.3: Eine Kohortenstudie mit wiederholten Untersuchungen.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 341

Experiment vs. Erhebung

- Zu untersuchender Faktor
(z.B. Wirksamkeit einer Therapie oder eines Medikaments, schädlicher Einfluss von Rauchen oder Übergewicht, Schützender Einfluss von Obst/Gemüse, Sport...)

Intervention

Beobachtung

Der Faktor wird gezielt eingesetzt und vorgegeben. Wirksamkeit (und Verträglichkeit) in Hinblick auf die Zielerkrankung werden geprüft.

Der Faktor wird beobachtet und Zusammenhänge mit bzw. sein Einfluss auf das Auftreten von Krankheit geprüft.

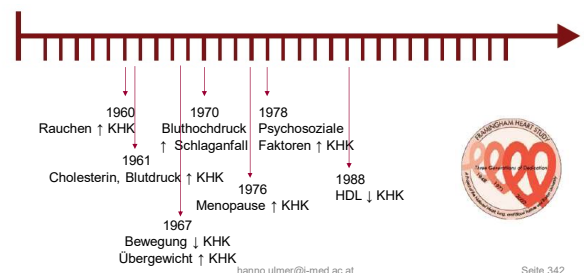
Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 339

Kohorten Studie: Framingham Heart Study

1948 Rekrutierung von 5209 gesunden Probanden (30-62J)



hanno.ulmer@i-med.ac.at

Seite 342

Fall-Kontroll Studie

Fall Kontroll Studie

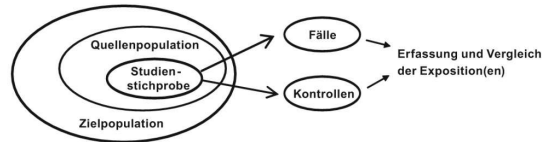


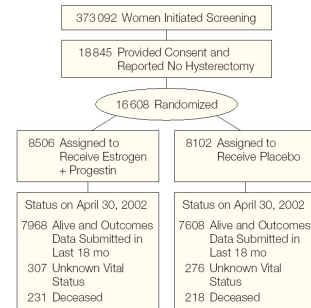
Abbildung 14.4: Grundstruktur einer Fall-Kontrollstudie.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 343

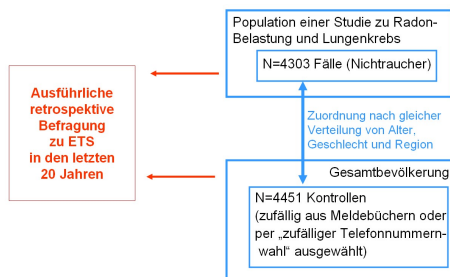
Interventionsstudie: Women's Health Initiative



hanno.ulmer@i-med.ac.at

Seite 346

Fall-Kontroll Studie: Passivrauchen (ETS) und Lungenkrebs



Kreuzer, Krauss, Kreienbrock, Jöckel, Wichmann in AJE, 1999

hanno.ulmer@i-med.ac.at

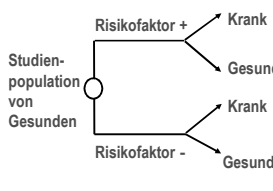
Seite 344

Meta-Analyse

Dr. Hanno Ulmer
hanno.ulmer@i-med.ac.at

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

Kohortenstudie prospektiv

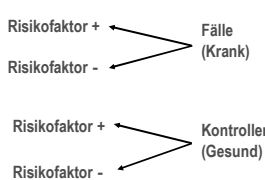


	Kranke	Gesunde	
Risikofaktor +	A	B	A + B
Risikofaktor -	C	D	C + D

$$\text{Relatives Risiko} = \frac{A / A+B}{C / C+D}$$

09.01.2025

Fall-Kontroll-Studie retrospektiv



	Kranke	Gesunde	
Risikofaktor +	A	B	A + B
Risikofaktor -	C	D	C + D

$$\text{Odds Ratio} = \frac{(A / A+B) / (B / A+B)}{(C / C+D) / (D / C+D)} = \frac{A \times D}{B \times C}$$

hanno.ulmer@i-med.ac.at

Gliederung

- Allgemeiner Teil
- Praktische Meta-Analyse mit MedCalc
- Übung Lionheart-Levorep

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 348

Literaturhinweise, Quellen

- Marcus Müllner - Erfolgreich wissenschaftlich arbeiten in der Klinik; Evidence Based Medicine-Springer Wien 2005, S.125 ff.
- George Davey Smith, Matthias Egger – Meta-Analysen, pharma-kritik Jahrgang 14 , Nummer 14, 1992
- Hansueli Stamm, Thomas M. Schwarb: Meta-analyse- Eine Einführung; ZfP 1995
- A. Ziegler, S. Lange, R. Bender; Systematische Übersichten und Meta-Analysen, Deutsche Medizinische Wochenschrift, 129. Ausgabe, 2004.
- Leo Held, Burkhard Seifert, Kaspar Rufibach Medizinische Statistik, Pearson 2013.
- <https://training.cochrane.org/interactivelearning>

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 349

Zweck

- Voraussetzung, es wurden zwei oder mehr Studien mit ähnlicher Fragestellung durchgeführt und/oder publiziert
- Stichprobenumfang der Einzelstudien zu klein
- Vorhandene Studien liefern inkonsistente Ergebnisse
- Untersuchung von Einflüssen und ihre Stärke auf das Ergebnis
- Grundlage für künftige Forschungstätigkeit
- Ermittlung des Publikationsbias

Definition und Idee

- Die Meta-Analyse ist die quantitative Kombination der Resultate mehrerer Einzelstudien
- Idee ca. 100 Jahre alt, erste Analyse 1955 (Becher et al JAMA 1955)
- Begriff erstmals 1979 durch Psychologen Gene V. Glass verwendet
- Mit der Entwicklung von geeigneteren statistischen Methoden haben Meta-Analysen seit dem Anfang der achtziger Jahre zunehmend an Bedeutung gewonnen. Der meta-analytische Ansatz blieb jedoch von heftiger Kritik nicht verschont. Während die einen die Meta-Analyse als «objektive, quantitative Methode» rühmen, wird das Verfahren von anderen als «statistischer Trick» bezeichnet, der «ungerechtfertigte Annahmen macht und zu unzulässigen Verallgemeinerungen führt».

Beispiel

Tab. 1 Untersuchung der Mortalität durch Prophylaxe mit Lidocain im akuten Myokardinfarkt (Quelle: Referenz [18], siehe auch [20]).

Quelle	Anzahl randomisierter Patienten		Anzahl verstorbener Patienten	
	Lidocain	Kontrolle	Lidocain	Kontrolle
1	39	43	2	1
2	44	44	4	4
3	107	110	6	4
4	103	100	7	5
5	110	106	7	3
6	154	146	11	4
Gesamt	557	549	37	21

09.01.2025

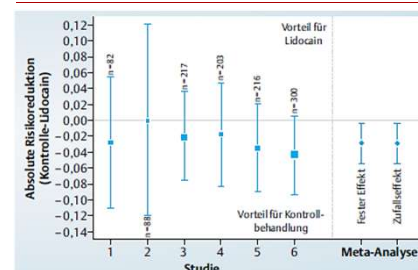
hanno.ulmer@i-med.ac.at

Seite 353

Ziel

- Einzelergebnisse inhaltlich homogener Primärstudien werden zusammengefasst und empirisch ausgewertet.
- Ziel ist eine Effektgrößenschätzung. Es soll untersucht werden, ob ein Effekt vorliegt und wie groß dieser ist.
- Effektgrößen können sein: Mittelwerte, Differenzen, relative oder absolute Risiken, Odds Ratio, etc.
- Mit Hilfe der Meta-Analyse lassen sich mehrere geeignete Einzelstudien statistisch zusammenfassen, diese können dabei verschieden gewichtet werden.
- Gewinn neuer Erkenntnisse aus alten Daten

Beispiel



Wirksamkeit von Lidocain zur Reduktion von Mortalität im akuten Myokardinfarkt [A. Ziegler, S. Lange, R. Bender; Systematische Übersichten und Meta-Analysen.]

Beispiel, Heterogenität

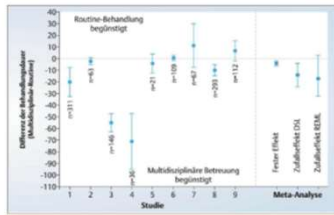


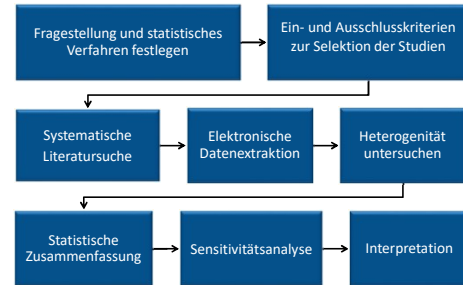
Abb. 2 Betreuung von Schlaganfallpatienten durch Team von Spezialisten mehrerer Disziplinen im Vergleich zum Routinemanagement: x-Achse stellt die einzelnen Studien sowie die Meta-Analysen dar; y-Achse Dauer des Krankenhausaufenthalts in Tagen (LOS). Für jede einzelne Studie und die Meta-Analysen sind die geschätzte LOS (Punkt) mit dem dazugehörigen 95% Konfidenzintervall (Schnurstriche) dargestellt.

09.01.2025

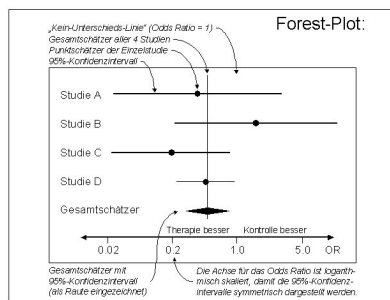
hanno.ulmer@i-med.ac.at

Seite 355

Ablauf der Meta-Analyse



Meta-Analyse



Modul 2.02

hanno.ulmer@i-med.ac.at

356

Meta-Analysis Protocol (FDA guidance)

- The planned purpose of the meta-analysis
- The background information available at the time of protocol development that motivated the meta-analysis
- The design features of the meta-analysis, including outcome definition and ascertainment, exposure periods and assessment, comparator drugs, and target subject population
- A description of the search strategy that will be used to identify candidate trials and the criteria that will be applied for trial selection
- The analysis strategy for conducting the meta-analysis, including planned subgroup analyses and sensitivity analyses

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 359

Stufen einer Meta-Analyse

- Definition der Hypothese
 - Definition der abhängigen und unabhängigen Variablen
 - Standardisierung der Terminologie
- Definition der Keywords für Literatursuche und Datensammlung
 - MeSH
- Definition von Ein- und Ausschlusskriterien für Studien
- Analyse und Validierung der Ergebnisse
 - Mit geeigneter Software

Modul 2.02

hanno.ulmer@i-med.ac.at

357

Methods: Search strategy

Search strategy

Key terms:

“physical activity”, “exercise”, “increase”, “brief intervention”, “counselling”, “systematic review”, “meta analysis”.

Period covered: 1854 - October 2011.

Databases

- CINAHL
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effects (DARE) on Cochrane Library and Centre for Reviews and Dissemination (CRD)
- Health Technology Assessment database on Cochrane Library and Centre for Reviews and Dissemination (CRD)
- Embase
- MEDLINE
- PsycINFO
- SSCI-Expanded
- SSCI SIGN
- Hand search of first authors' (LL) personal collections of articles

VBI VERY SMALL INTERVENTIONS

Hierarchie von Medizinischen Studien



Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 361

Fragen zur Meta-Analyse

- Ist die Qualität der berücksichtigten Studien zufriedenstellend?

Zwischen verschiedenen Studien bestehen zum Teil erhebliche Qualitätsunterschiede, die in einer Meta-Analyse zunächst nicht berücksichtigt werden. Als Mindestanforderung gilt, dass nur korrekt randomisierte Studien mit vollständigen Angaben über alle am Anfang in die Studie aufgenommenen Personen meta-analysiert werden. Der Randomisierungsvorgang, der leider oft ungenügend beschrieben ist, verdient deshalb besondere Aufmerksamkeit. Daneben ist es auch von Bedeutung, dass in den verschiedenen Studien tatsächlich gleiche oder vergleichbare Therapien und Beurteilungsverfahren eingesetzt wurden. Bei vielen Medikamentenstudien muss z.B. gefordert werden, dass die Bewertung der Therapieergebnisse blind erfolgt.

- Wurden alle relevanten Studien berücksichtigt?

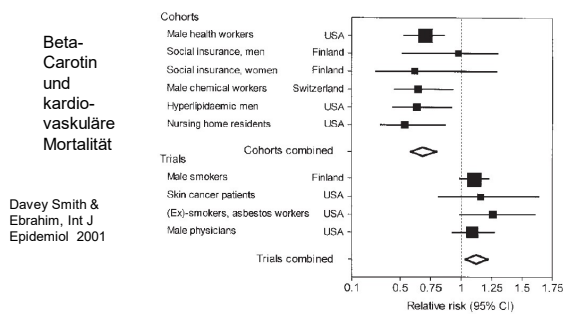
Da «positive» Resultate mit grösserer Wahrscheinlichkeit veröffentlicht werden als Studien, die keinen Effekt zeigen, sollten auch unpublizierte Studien berücksichtigt werden. Dies gilt natürlich nur, soweit die erwähnten Qualitätskriterien erfüllt sind. Wenn allenfalls in Frage kommende Studien nicht in eine Meta-Analyse aufgenommen werden, so sollten sie erwähnt und die entsprechende Entscheidung begründet werden.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 364

Konsistenz



hanno.ulmer@i-med.ac.at

Seite 362

Fragen zur Meta-Analyse

- Haben die berücksichtigten Studien teilweise entgegengesetzte Resultate erbracht?

Der Variabilität der Ergebnisse wird oft nicht genügend Beachtung geschenkt. Das Problem der Heterogenität von Studienresultaten kann nicht durch die Anwendung eines statistischen Tests gelöst werden. Wenn stark voneinander abweichende Resultate vorliegen, geben klinische und biologische Überlegungen den Ausschlag, ob überhaupt eine Meta-Analyse sinnvoll ist.

- Wie «robust» sind die Ergebnisse der Meta-Analyse?

Bei der Durchführung einer Meta-Analyse gibt es eine Reihe von teilweise arbiträren Entscheidungen (Ausschluss von Studien, Wahl der statistischen Methoden, Interpretation bestimmter Resultate). Das Resultat einer Meta-Analyse sollte von solchen Entscheidungen einigermaßen unabhängig sein. Wenn also die gleiche Meta-Analyse z.B. mit einer etwas anderen Studienauswahl oder mit einem anderen statistischen Verfahren durchgeführt wird, sollte sie ungefähr dasselbe Resultat ergeben.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 365

Evidenzgrade von Studien

Stufe	Für Therapie, Prävention, Ätiologie, Nebenwirkungen
1a	Systematisches Review mit homogenen RCTs
1b	Einzelne RCT (mit engen Konfidenzintervallen)
1c	Sonderfälle
2a	Systematisches Review mit homogenen Kohortenstudien
2b	Einzelne Kohortenstudie oder methodisch schwache RCT (z.B. <80% Nachbeobachtung)
2c	"Outcome"-Forschung, Ökologische Studien
3a	SR mit homogenen Fall-Kontroll-Studien
3b	Einzelne Fall-Kontroll-Studie
4	Fallserien (+ qualitativ schlechte Kohorten- u. Fall-Kontroll-Studien)
5	Expertenmeinung ohne explizite Bewertung der Evidenz oder basierend auf physiologischen Grundlagenmodellen

Modul 2.02

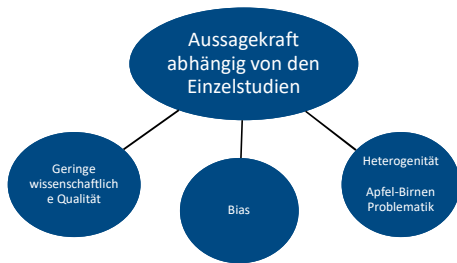
hanno.ulmer@i-med.ac.at

Seite 363

Stärken

- Mathematische-statistische Zusammenfassung von Einzelstudien
- Im Gegensatz dazu steht der narrative Review
- Meta-Analyse objektiver durch ihre Festlegung von Kriterien für die Auswahl von Primärstudien (gegebenenfalls sinkt aber die Studienanzahl)
- Relativ kostengünstig
- Erhöhung der Validität und Trennschärfe
- Ermittlung welche Eigenschaften zu welchen Effektstärken führen

Schwächen



Stratifying the analysis by trial

- An important principle involved in estimating risk from a meta-analysis is that the randomized comparisons of the individual trials should be maintained when analyzing the combined data.
- In other words, when comparing drug A to drug B, subjects randomly assigned to drug A in a single trial are compared to subjects assigned to drug B from the same trial and not to subjects from other trials. In the statistics literature, this is referred to as stratifying the analysis by trial.

Meta-Analyse

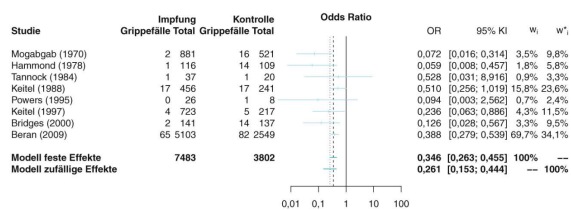


Abbildung 15.1: Forest-Plot der Meta-Analyse zur Grippeimpfung.

Modul 2.02

Tabelle 15.1: Daten aus Powers (1995), mit angepassten Werten zur Berechnung des Odds Ratios in Klammern.

Simpson's Paradoxon

Table 1. An Illustration of Simpson's Paradox from Incorrect Pooling of Data

Trial	Events	Drug A		Events	Drug B	
		Patients	Risk		Patients	Risk
1	1	100	1.0%	2	200	1.0%
2	1	100	1.0%	2	200	1.0%
3	200	1200	16.7%	50	300	16.7%
4	2	200	1.0%	2	200	1.0%
Total	204	1600	12.8%	56	900	6.2%

The hypothetical example in Table 1 illustrates an extreme example of Simpson's paradox in which, for each of four trials, the estimated risk of a safety event is identical for both Drug A and Drug B. With simple pooling, however, the risk for Drug A appears to be more than twice as high as that for Drug B (12.8 percent vs. 6.2 percent)

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 371

Publications Bias

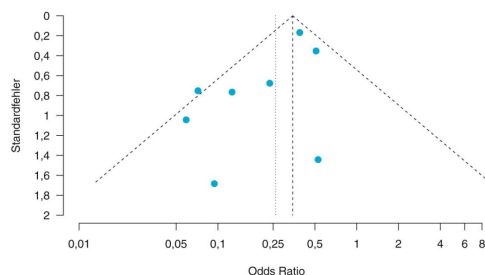


Abbildung 15.3: Funnel-Plot zur Beurteilung von Publikations-Bias.

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 369

Mantel-Haenszel Methode – stratifizierte Analyse

	Erfolgreich		Total
	Ja	Nein	
Therapie A	273 (78 %)	77	350
Therapie B	289 (83 %)	61	350

Tabelle 7.3: Analyse des Erfolgs zweier Therapien zur Nierensteinentfernung.

	Kleine Nierensteine			Große Nierensteine		
	Ja	Nein	Total	Ja	Nein	Total
Therapie A	81 (93 %)	6	87	102 (73 %)	71	263
Therapie B	234 (87 %)	36	270	55 (69 %)	25	80

Tabelle 7.4: Subgruppenanalyse des Erfolgs zweier Therapien zur Nierensteinentfernung in Abhängigkeit von der Nierensteingröße.

$$Odds\ Ratio_{Mantel-Haenszel} = \frac{81 \times 36 + 102 \times 25}{357 \times 343} = 1,45 (p=0,12)$$

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 372

Sensitivity Analysis



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- The goal of any sensitivity analysis should not be to search for additional findings, but to support and understand the primary findings of the meta-analysis.
- For example, a meta-analysis that included one very large study contributing a large proportion of subjects and events could raise a concern that it was overly influencing the meta-analytic results. A sensitivity analysis that excluded that study would have reduced numbers of subjects and events and lower power to yield a significant finding, but a risk estimate that was consistent with the original estimate would add to the weight of evidence of the finding

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 373

MedCalc statistical software: Meta-analysis Introduction



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- A meta-analysis integrates the quantitative findings from separate but similar studies and provides a numerical estimate of the overall effect of interest (Petrie et al., 2003).
- Different weights are assigned to the different studies for calculating the summary or pooled effect. The weighting is related with the inverse of the standard error (and therefore indirectly to the sample size) reported in the studies. Studies with smaller standard error and larger sample size are given more weight in the calculation of the pooled effect size.

09.01.2025

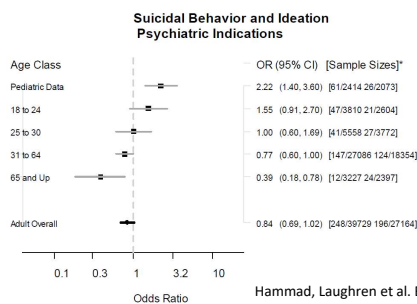
hanno.ulmer@i-med.ac.at

Seite 376

FDA Meta-Analysis of Antidepressants and Suicidal Behaviour



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 374

The effect of interest can be



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- The effect of interest can be:
- an average of a continuous variable
- a correlation between two variables
- an odds ratio, suitable for analyzing retrospective studies
- a relative risk (risk ratio) or risk difference, suitable for analyzing prospective studies
- a proportion
- the area under the ROC curve

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 377

Cochrane



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK



Das Cochrane-Logo spiegelt die Ergebnisse eines Systematischen Cochrane Reviews mit Kultcharakter wider. In dem Review von 1989 ging es um die Frage, ob die Reifung der Lungen bei Frühgeborenen durch die Gabe von Kortikosteroiden unterstützt werden kann. Der Cartoonist David Mostyn kreierte aus den Studienergebnissen ein prägnantes Logo.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 375

Fixed and random effects model



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Under the fixed effects model, it is assumed that the studies share a common true effect, and the summary effect is an estimate of the common effect size.
- Under the random effects model the true effects in the studies are assumed to vary between studies and the summary effect is the weighted average of the effects reported in the different studies (Borenstein et al., 2009).
- The random effects model will tend to give a more conservative estimate (i.e. with wider confidence interval), but the results from the two models usually agree when there is no heterogeneity.
- When heterogeneity is present (see below) the random effects model should be the preferred model.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 378

Heterogeneity

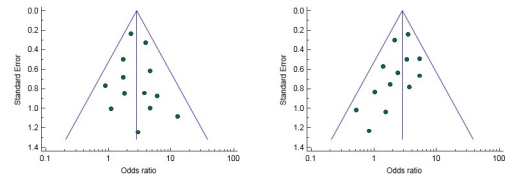
- Cochran's Q is the weighted sum of squares on a standardized scale. It is reported with a P value with low P-values indicating presence of heterogeneity. This test however is known to have low power to detect heterogeneity and it is suggested to use a value of 0.10 as a cut-off for significance (Higgins et al., 2003). Conversely, Q has too much power as a test of heterogeneity if the number of studies is large.
- I^2 statistics is the percentage of observed total variation across studies that is due to real heterogeneity rather than chance. It is calculated as $I^2 = 100\% \times (Q - df)/Q$, where Q is Cochran's heterogeneity statistic and df the degrees of freedom. Negative values of I^2 are put equal to zero so that I^2 lies between 0% and 100%. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity (Higgins et al., 2003).

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 379

Funnel plot



Publication bias results in asymmetry of the funnel plot. If publication bias is present, the smaller studies will show the larger effects. See Sterne et al. (2011) for interpreting funnel plot asymmetry. The funnel plot may not always be a reliable tool, in particular when the number of studies included in the analysis is small.

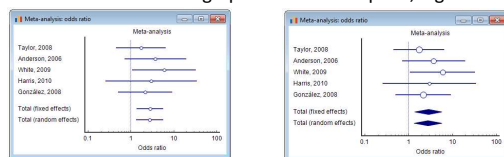
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 382

Forest plot

The results of the different studies, with 95% CI, and the overall effect (under the fixed and random effects model) with 95% CI are illustrated in a graph called "forest plot", e.g.:



In this example the markers representing the effect size all have the same size. Optionally, the marker size may vary in size according to the weights assigned to the different studies. In addition, the pooled effects can be represented using a diamond. The location of the diamond represents the estimated effect size and the width of the diamond reflects the precision of the estimate.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 380

References

- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Chichester, UK: Wiley.
- Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315: 629–634.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560.
- Petrie A, Bulman JS, Osborn JF (2003) Further statistics in dentistry. Part 8: systematic reviews and meta-analyses. *British Dental Journal* 194:73–78.
- Sterne JA, Egger E (2001) Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *Journal of Clinical Epidemiology* 54:1046–1055.
- Sterne JA, Sutton AJ, Ioannidis JP et al. (2011) Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* 7:177–188.
- Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from the retrospective analysis of disease. *Journal of the National Cancer Institute* 22: 719–748.
- Hedges LV, Olkin I (1985) Statistical methods for meta-analysis. London: Academic Press.
- Zhou XH, NA Obuchowski, DK McClish (2002) Statistical methods in diagnostic medicine. New York: Wiley.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 383

Funnel plot

- A funnel plot (Egger et al., 1997) is a graphical tool for detecting bias in meta-analysis.
- In a funnel plot treatment effect is plotted on the horizontal axis and MedCalc plots the standard error on the vertical axis (Sterne & Egger, 2001).
- The vertical line represents the summary estimated derived using fixed-effect meta-analysis.
- Two diagonal lines represent (pseudo) 95% confidence limits (effect ± 1.96 SE) around the summary effect for each standard error on the vertical axis. These show the expected distribution of studies in the absence of heterogeneity or of selection bias. In the absence of heterogeneity, 95% of the studies should lie within the funnel defined by these diagonal lines.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 381

Meta-analysis in MedCalc

- [Continuous measure](#)
- [Correlation](#)
- [Proportion](#)
- [Relative risk](#)
- [Risk difference](#)
- [Odds ratio](#)
- [Area under ROC curve](#)
- [Generic inverse variance method](#)

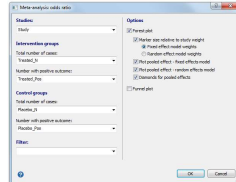
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 384

Meta-analysis: odds ratio

MedCalc uses the Mantel-Haenszel method (Mantel & Haenszel, 1959) for calculating the weighted pooled odds ratio under the fixed effects model. Next the heterogeneity statistic is incorporated to calculate the summary odds ratio under the random effects model (DerSimonian & Laird, 1986).



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 385

Meta-analysis: relative risk

Study	Treated_Positive	Treated_Total	Controls_Positive	Controls_Total
1 佐藤, 2012	95	103	47	104
2 渡辺, 2008	119	127	34	129
3 山本, 2013	51	223	12	76
4 長谷川, 2006	122	139	61	142
5 田村, 2006	47	53	10	51
6 小野, 2004	121	135	29	68
7 菊地, 2008	337	378	170	376

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 388

Meta-analysis: odds ratio

Study	Treated_N	Treated_Pos	Placebo_N	Placebo_Pos
1 Taylor, 2008	73	15	23	3
2 Anderson, 2006	35	7	32	2
3 White, 2009	20	8	20	2
4 Harris, 2010	12	3	10	1
5 González, 2008	42	6	42	3

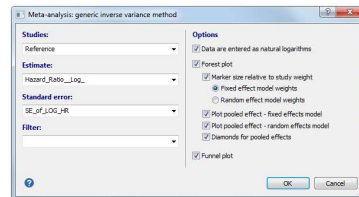
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 386

Meta-analysis: hazard ratio (generic inverse variance method)

Estimates and their standard errors are entered directly. For ratio measures of intervention effect, the data should be entered as natural logarithms (for example as a log Hazard ratio and the standard error of the log Hazard ratio). In the inverse variance method the weight given to each study is the inverse of the variance of the effect estimate (i.e. one over the square of its standard error). Thus larger studies are given more weight than smaller studies, which have larger standard errors.



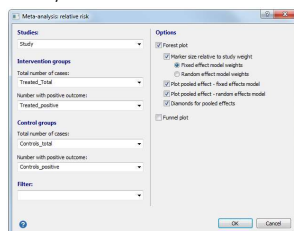
09.01.2025

hanno.ulmer@i-med.ac.at

Seite 389

Meta-analysis: relative risk and risk difference

MedCalc uses the Mantel-Haenszel method (based on Mantel & Haenszel, 1959) for calculating the weighted pooled relative risk and risk difference under the fixed effects model. Next the heterogeneity statistic is incorporated to calculate the summary relative risk under the random effects model (DerSimonian & Laird, 1986).



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 387

Meta-analysis: hazard ratio (generic inverse variance method)

Study	Hazard_Ratio_Log	SE_of_LOG_HR
1 Study 1	-0.077	0.212
2 Study 2	0.012	0.221
3 Study 3	0.323	0.426
4 Study 4	0.154	0.230
5 Study 5	0.051	0.348
6 Study 6	-0.661	0.232
7 Study 7	-0.199	0.337
8 Study 8	0.040	0.245
9 Study 9	0.305	0.432

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 390

Meta-analysis: continuous measure



For meta-analysis of studies with comparison of means between treated cases and controls, MedCalc uses the Hedges g statistic as a formulation for the standardized mean difference under the fixed effects model. Next, the heterogeneity statistic is incorporated to calculate the summary standardized mean difference under the random effects model (DerSimonian & Laird, 1986).

The standardized mean difference Hedges g is the difference between the two means divided by the pooled standard deviation, with a correction for small sample bias.

09.01.2025

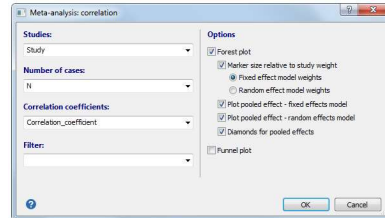
hanno.ulmer@i-med.ac.at

Seite 391

Meta-analysis: correlation



MedCalc uses the Hedges-Olkin (1985) method for calculating the weighted summary Correlation coefficient under the fixed effects model, using a Fisher Z transformation of the correlation coefficients. Next, the heterogeneity statistic is incorporated to calculate the summary Correlation coefficient under the random effects model (DerSimonian and Laird, 1986).

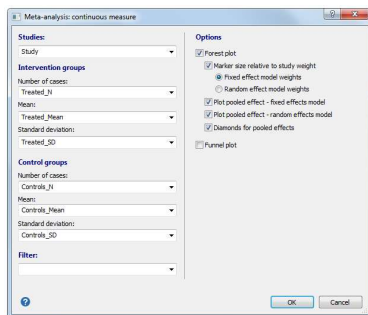


09.01.2025

hanno.ulmer@i-med.ac.at

Seite 394

Meta-analysis: continuous measure



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 392

Meta-analysis: correlation



Meta-analysis - Correlation				
Study				
A	B	C	D	
Study	Correlation_coefficient	N		
1 Moore, 2006	0.56	133		
2 Davis, 2008	0.43	149		
3 Thomas, 1999	0.53	131		
4 Miller, 2012	0.51	120		
5 Williams, 2012	0.66	111		
6 Young, 2013	0.46	152		
7 Baker, 2009	0.33	60		
8 Adams, 2006	0.38	122		

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 395

Meta-analysis: continuous measure



Meta-analysis - Continuous measure							
Study							
A	B	C	D	E	F	G	
Study	Treated_N	Treated_Mean	Treated_SD	Controls_N	Controls_Mean	Controls_SD	
1 Smith, 1998	40	23.52	1.38	40	20.12	3.36	
2 Brown, 2003	162	25.6	2.3	175	23.5	1.4	
3 Wilson, 1997	36	21.7	2.1	36	19.7	2.9	
4 Emerson, 2012	20	23.13	2.12	23	23.21	3	
5 Jones, 2006	25	25.3	2.2	25	24.9	2.6	
6							

09.01.2025

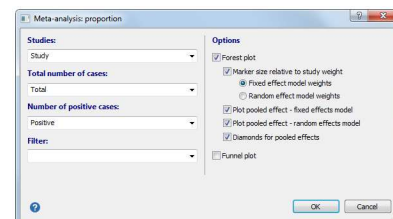
hanno.ulmer@i-med.ac.at

Seite 393

Meta-analysis: proportions



MedCalc uses a Freeman-Tukey transformation (arcsine square root transformation; Freeman and Tukey, 1950) to calculate the weighted summary proportion under the fixed and random effects model (DerSimonian & Laird, 1986).



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 396

Meta-analysis: proportions

Meta-analysis - Proportions				
A	Study			
	Study	Positive	Total	
1	Ward, 1999	226	324	
2	Li, 2012	48	61	
3	Peterson, 2009	43	59	
4	Kim, 2005	191	295	
5	Chang, 2008	2532	5033	
6	Reyes, 2013	64	110	
7	Gauthier, 2013	100	183	
8	Garcia, 1007	35	45	

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 397

Lionheart+Levorep

Meta-analysis: relative risk

Meta-analysis: relative risk									
Variable for studies									
1. Intervention groups									
Variable for total number of cases									
Variable for number of positive cases									
2. Control groups									
Variable for total number of cases									
Variable for number of positive cases									
Study	Intervention	Controls	Relative risk	95% CI	z	P	Weight (%)	Fixed	Random
Lionheart	15/48	8/21	0.820	0.412 to 1.632			90.80	80.81	
LevoRep	1/63	4/57	0.226	0.0260 to 1.965			9.20	19.39	
Total (fixed effects)	16/111	12/78	0.658	0.341 to 1.269	-1.250	0.211	100.00	100.00	
Total (random effects)	16/111	12/78	0.639	0.227 to 1.800	-0.848	0.397	100.00	100.00	

Test for heterogeneity

Q	1.3331
DF	1
Significance level	P = 0.2483
I ² (inconsistency)	24.98%
95% CI for I ²	0.00 to 0.00

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 400

Meta-analysis: AUC

MedCalc uses the methods described by Zhou et al. (2002) for calculating the weighted summary Area under the ROC curve under the fixed effects model and random effects model.

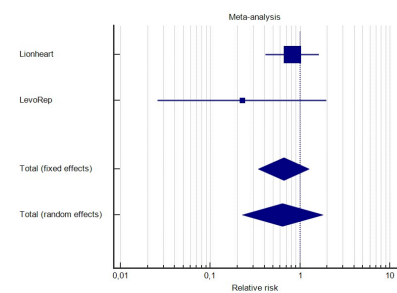
Meta-analysis: area under ROC curve	
Studies:	Options
Reference	<input checked="" type="checkbox"/> Forest plot
Area under ROC curve (AUC):	<input checked="" type="checkbox"/> Marker size relative to study weight
AUC	<input type="checkbox"/> Fixed effect model weights
Standard error of AUC:	<input type="checkbox"/> Random effect model weights
SE	<input checked="" type="checkbox"/> Plot pooled effect - fixed effects model
Filters:	<input checked="" type="checkbox"/> Plot pooled effect - random effects model
	<input checked="" type="checkbox"/> Diamonds for pooled effects
	<input type="checkbox"/> Funnel plot

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 398

Lionheart+Levorep



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 401

Meta-analysis: AUC

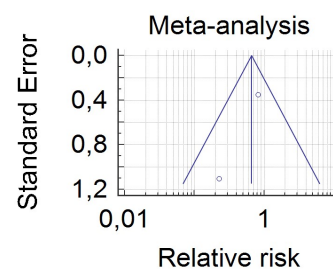
Meta-analysis - ROC area				
A	Reference			
	A	B	C	D
	Reference	AUC	SE	
1	Study 1	0.789	0.057	
2	Study 2	0.724	0.025	
3	Study 3	0.851	0.028	
4	Study 4	0.876	0.029	
5	Study 5	0.782	0.102	
6	Study 6	0.702	0.056	
7	Study 7	0.652	0.038	

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 399

Lionheart+Levorep



09.01.2025

hanno.ulmer@i-med.ac.at

Seite 402

Lionheart+Levorep



Study	Levo_N	Levo_Deaths	Placebo_N	Placebo_Deaths
Lionheart	48	15	21	8
LevoRep	63	1	57	4

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 403

Table 1. Checklist of items to include when reporting a systematic review or meta-analysis.

Item	Checklist item	Reported on page
1	Identify the report as a systematic review, meta-analysis, or both.	
2	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
3	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
4	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
5	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
6	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
7	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
8	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
9	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
10	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
11	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
12	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
13	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
14	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
15	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
16	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
17	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
18	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
19	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
20	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
21	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
22	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
23	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
24	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
25	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
26	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
27	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
28	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
29	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
30	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
31	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
32	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
33	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
34	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
35	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
36	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
37	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
38	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
39	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
40	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
41	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
42	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
43	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
44	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
45	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
46	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
47	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
48	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
49	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	
50	Provide a structured summary including: objectives, background, methods, results, conclusions, and a declaration of interests.	

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine 6(7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097>



Lionheart+Levorep



Meta-analysis: risk difference

Variable for studies		study						
1. Intervention groups								
Variable for total number of cases		Levosimendan_N						
Variable for number of positive cases		Levosimendan_Deaths						
2. Control groups								
Variable for total number of cases		Placebo_N						
Variable for number of positive cases		Placebo_Deaths						
Study	Intervention	Controls	Risk Difference	95% CI	Z	P	Weight (%)	
								<div><div></div><div>Facet</div><div>Random</div></div>
Lionheart	1546	821	-0.0985	-0.314 to 0.117			8.15	8.15
LevoRep	193	457	-0.0543	-0.127 to 0.0188			91.85	91.85
Total (fixed effects)	16111	12778	-0.0569	-0.153 to 0.0394	-1.224	0.221	100.00	100.00
Total (random effects)	16111	12778	-0.0566	-0.126 to 0.0146	-1.580	0.121	100.00	100.00

Test for heterogeneity

Q

0.02123

df

1

Significance level

P = 0.8842

df (uncertainty)

0.05%

95% CI for I²

0.00 to 0.00

Q	0.02123
DF	1
Significance level	P = 0.8842
I ² (inconsistency)	0.00%
95% CI for I ²	0.00 to 0.00

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 404

PRISMA for Individual Patient Data



PRISMA for Individual Patient Data systematic reviews (PRISMA-IPD)
 PRISMA-IPD was published in 2015 and provides guidelines for reporting systematic reviews and meta-analyses of IPD. Systematic reviews and meta-analyses of IPD aim to collect, check, and reanalyze individual-level data from all studies addressing a particular research question.

Statement paper:

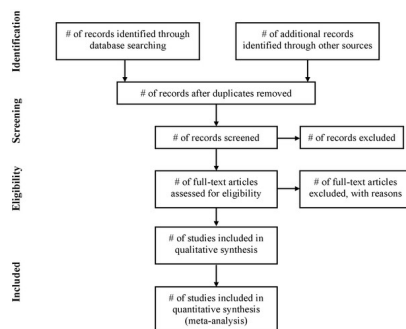
Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF; PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015;313(16):1657-1665.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 407

Figure 1. Flow of information through the different phases of a systematic review.



Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLOS Medicine 6(7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097>



Effect size calculator



Practical Meta-Analysis Effect Size Calculator David B. Wilson, Ph.D., George Mason University:

<http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 408

Deklaration von Helsinki Auszug

- Forscher muss Ethikkommission informieren über
 - Finanzierung
 - Sponsoren
 - institutionelle Verbindungen
 - potentielle Interessenskonflikte
 - Anreize für Versuchspersonen

09.01.2025

hanno.ulmer@i-med.ac.at

415

Weitere wichtige Richtlinien



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- ICH E6 Guideline for good clinical practice
 - Beschreibung der wichtigsten Elemente
 - Richtlinie muß für Zulassungsstudien befolgt werden
 - Bezieht sich auf Aspekte zur Durchführung klinischer Studien
 - Beschreibt die Rolle der Ethikkommission
 - Beschreibt Aufgaben des Prüfarztes, des Sponsors und des Monitors

09.01.2025

hanno.ulmer@i-med.ac.at

418

International Conference of Harmonisation (ICH)



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Mitglieder:

- Kommission der Europäischen Gemeinschaft
- European Federation of Pharmaceutical Industries and Associations (EFPIA)
- Ministry of Health, Labor and Welfare (MHLW)
- Japan Pharmaceutical Manufacturers Association (JPMA)
- US Food and Drug Administration (FDA)
- Pharmaceutical Research and Manufacturers of America (PhRMA)

09.01.2025

hanno.ulmer@i-med.ac.at

416

CONSORT-Richtlinien



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- **CONSORT (Consolidated Standards of Reporting Trials)**
- 22 internationale Standards und Richtlinien zur Darstellung von Ergebnissen in wissenschaftlichen Arbeiten
 - Studienplanung
 - Durchführung
 - Statistische Analyse
 - Interpretation
- Flussdiagramm über die Rekrutierung der Patienten
- Transparenz einer Studie gewährleisten
- www.consort-statement.org/
 - Revidierte Version 2001

Modul 2.02

hanno.ulmer@i-med.ac.at

419

Richtlinie ICH E9 (Statistical Principles for Clinical Trials, 1998)



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

- Harmonisierung statistischer Vorgehensweisen
 - Biostatistical Methodology in Clinical Trials in Application for Marketing Authorisations (Committee for Proprietary Medicinal Products, 1994)
 - Guidelines on Statistical Analysis of Clinical Studies (Japanese Ministry of Health and Welfare, 1992)
 - Guideline for the Format and Content of Clinical and Statistical Sections of a New Drug Application (U.S. Food and Drug Administration, 1988)

09.01.2025

hanno.ulmer@i-med.ac.at

417

Randomisierte Klinische Studien (Randomised Clinical Trials)



MEDIZINISCHE UNIVERSITÄT
INNSBRUCK

Dr. Hanno Ulmer
hanno.ulmer@i-med.ac.at
Innsbruck, Oktober 2010

Department für Medizinische Statistik, Informatik und
Gesundheitsökonomie, Medizinische Universität Innsbruck

RCTs

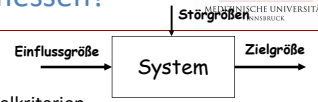
- **Randomisiert**
 - Zufällige Zuteilung zu einer Therapieform
 - Stratifizierung
 - Unterteilung in Subgruppen nach bestimmten Merkmalen, wie Alter, Ausgangswerte, ...
- **Kontrolliert**
 - Mindestens eine Kontrollgruppe
 - z.B.: Vergleich mit Standardtherapie, Placebo-Kontrolle
 - Statistische Vergleichbarkeit der wesentlichen Merkmale zwischen Gruppen
- **Verblindet – wenn möglich**
 - Offen
 - Einfach verblindet: Patient weiß nicht, welche Therapie er bekommt
 - Doppel-blind: Patient und Arzt wissen nicht, welche Therapie Patient bekommt

Modul 2.02

hanno.ulmer@i-med.ac.at

421

Was muss man messen?

- 
- ```

graph LR
 IG[Einflussgröße] --> S[System]
 S --> Z[Zielgröße]
 SG[Störgröße] --> S

```
- **Zielgröße(n)**
    - Haupt- und Nebenzielkriterien
      - In engem Zusammenhang mit dem Studienziel
      - Klinisch relevant
  - **Einflussgröße(n)**
    - Können Auswirkungen auf die Zielgröße(n) haben
  - **Störgröße(n)**
    - Nicht von primären Studieninteresse
    - Können jedoch das Studienergebnis wesentlich beeinflussen bzw. verzerren („Bias“!)

Modul 2.02

hanno.ulmer@i-med.ac.at

424

## Einige einfache Studienpläne

- **Parallelgruppen**
  - Zwei oder mehrere unabhängige, aber vergleichbare Gruppen
    - Werden zeitlich parallel behandelt
    - Mit unterschiedlichen Therapieformen
- **Cross-Over**
  - Eine Gruppe mit zwei oder mehreren aufeinander folgenden Therapien behandeln
  - Jeder Proband erhält jede Therapie aber in unterschiedlicher Reihenfolge
    - Randomisierung der Reihenfolge
      - z.B.: ABC, ACB, BAC, BCA, CAB, CBA
      - Carry-over-Effekte beachten
- **Faktorielle Pläne**
  - Kombination von zwei oder mehreren Einflussfaktoren
    - Wechselwirkungen überprüfbar

Modul 2.02

hanno.ulmer@i-med.ac.at

422

## Primäre und sekundäre Zielgrößen

- Wenn möglich nur ein einziges Hauptzielkriterium (=primäre Zielgröße)
  - Blutdrucksenkung, Überlebenszeit, ...
  - Kann multidimensional sein
  - Fragebogen, QALY, DALY, ...
- Im Studienprotokoll operationalisieren
  - Messbar machen
- Mehrere sekundäre Zielvariablen sind zulässig
- Surrogatkriterien

Modul 2.02

hanno.ulmer@i-med.ac.at

425

## Parallelgruppenstudie

- z.B. Untersuchung der Wirksamkeit eines neuen blutdrucksenkenden Medikaments im Vergleich zu einer Standardtherapie (=Kontrolle)
  - Manchmal Placebo
- **Nullhypothese  $H_0$** : die beiden Therapien sind im Mittel gleich wirksam
  - z.B. die Änderung des diastolischen Blutdrucks ist im Mittel in beiden Gruppen gleich hoch
  - Die Ungültigkeit der Nullhypothese ist zu beweisen
- **Alternativhypothese  $H_1$** : die beiden Therapien sind im Mittel unterschiedlich stark wirksam
- **Voraussetzung**: Die beiden Gruppen stimmen in den wesentlichen Merkmalen überein – siehe Randomisierung
  - Nur die Therapien sind unterschiedlich
  - Strukturgleichheit der Gruppen

Modul 2.02

hanno.ulmer@i-med.ac.at

423

## Vergleichbarkeit der Gruppen

- **Bezüglich der**
  - Gemessenen Merkmale
    - Dokumentieren
    - Bei der Auswertung berücksichtigen
  - Nicht-gemessenen Merkmale
  - Unbekannten Merkmale
    - Können möglicherweise zu Verzerrungen führen
- Randomisierung
- Stratifizierung
- Paarbildung
  - Zwillinge, Matching

Modul 2.02

hanno.ulmer@i-med.ac.at

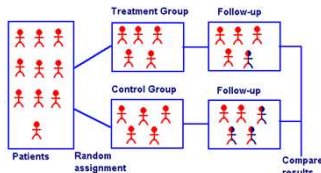
426



## Randomisierung



- **Zufällige Zuordnung eines Subjektes/Objektes einer Stichprobe zu einer der Gruppen des Einflussfaktors**
  - z.B. mit Hilfe eines Zufallsgenerators
- **ZIEL:**
  - Vermeidung eines Selektions-Bias
    - Jedes Objekt hat die gleiche Chance zufällig einer Gruppe zugeteilt zu werden
  - Strukturgleichheit (Zusätzliche Stratifizierung ist möglich)



Modul 2.02

hanno.ulmer@i-med.ac.at

427

## Fallzahl



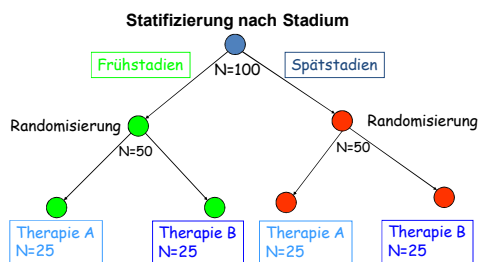
- **Fixe, berechnete Fallzahl**
  - Basierend auf Annahmen über Unterschiede/Behandlungseffekte
  - Fehler 1. und 2. Art (z.B. Signifikanzniveau=0.05, Power=80%)
- **Sequentielle Pläne**
  - 1 bis K geplante Zwischenauswertungen
    - **Interimsanalysen, Adaptive Studienpläne**
  - Möglicherweise frühere Entscheidung
- **Berücksichtigung der Ausfallrate bei statistischen Analyse**
  - **Intention-to-treat** Analyse: alle **eingeschlossenen** Fälle
  - **Per Protokoll** Analyse: alle **abgeschlossenen** Fälle

Modul 2.02

hanno.ulmer@i-med.ac.at

430

## Randomisierung / Stratifizierung



Modul 2.02

hanno.ulmer@i-med.ac.at

428

## Patientenflussdiagramm

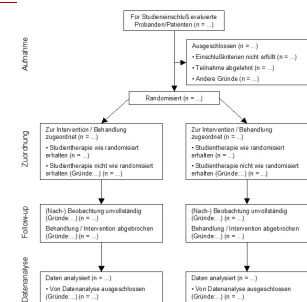


- **Intention-To-Treat**

versus

- **Per-Protocol**

- **Vergleichbarkeit der Gruppen:**
  - Strukturgleichheit
  - Beobachtungsgleichheit
  - Behandlungsgleichheit



Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 431

## Verblindung



- **Ausschaltung von Verzerrungen/systematischen Fehlern (Bias) aufgrund von Vorinformationen**
  - Selektions-Bias, Beobachter-Bias
- **Ausprägungen des Einflussfaktors sind für den Beobachter unbekannt**
  - **Doppel-Blindstudie**
    - Patient und Arzt wissen nicht, welche Therapieform angewendet wird
    - Nicht immer möglich!!!
- **Beschreibung der Art der Verblindung**
  - äußere Form, Geschmack, Farbe
  - sichtbare Unterschiede bei Behandlung (z.B. Farbe des Urins)
- **Beurteilung des Outcomes**
  - unabhängige Beurteilung durch Dritte
- **Was passiert, wenn Code gebrochen/entziffert wird?**

Modul 2.02

hanno.ulmer@i-med.ac.at

429

## Checklist für Fallzahl-schätzung (Testproblem)



- Fehler 1. Art (üblicherweise 0,05)
- Fehler 2. Art (0,1 oder 0,2)
- Auswahl des Hauptzielkriteriums
- Zu erwartender Unterschied und Angabe eines Variationsmaßes
- Begründung dafür – Literatur oder Vorstudie
- Auswahl des statistischen Tests
- Falls mehrere Hypothesen formuliert werden, Korrektur des Fehler 1. Art oder Hierarchisierung der Hypothesen
- Drop-Out Rate berücksichtigen

Modul 2.02

hanno.ulmer@i-med.ac.at

Seite 432



## Vorschlag „Studienprotokoll – Statistik“



### ad 9 Statistik

#### 9.1 Fallzahlplanung

Die Fallzahl wird in der Regel berechnet aus der primären Zielvariablen, dem klinisch relevanten Unterschied, der in der Studie nachgewiesen werden soll, dem statistischen Verfahren, das dazu verwendet wird, sowie aus dem Fehler 1. Art  $\alpha$  und der Power  $1-\beta$ , die erzielt werden soll. Die Fallzahl ist außerdem abhängig von der Anzahl der Gruppen, die verglichen werden sollen. Angaben zu diesen Parametern sowie zum Verfahren der Fallzahlplanung sind erforderlich.

#### 9.2 Randomisierung

Bei der Beschreibung der Randomisierung sind Angaben erforderlich zur Art der Randomisierung, zur Anzahl der Gruppen und der erforderlichen Strata und ggf. zum organisatorischen Ablauf der Randomisierung. Dabei sind besonders bei verblindeten Studien Maßnahmen aufzulisten, die die Verblindung gewährleisten.

#### 9.3 Statistische Methoden

Zielgrößen, Definition von Auswertungskollektiven, Datenanalyse, Zwischenauswertungen, Verweis auf ICH-GCP E9: Statistical Principles for Clinical Trials

09.01.2025

hanno.ulmer@i-med.ac.at

433

## Regressionsanalyse



Regressionsanalyse als statistische Methode um Zusammenhänge zu beschreiben :

Exposition (z.B. Risikofaktor, Therapie) --> Outcome (z.B. Erkrankung)

Multivariable Analyse:

k unabhängige Variable (Prädiktoren) --> 1 abhängige Variable (Outcome)

Verschiedene Arten und Berechnungsweisen von Regressionsanalysen:

Bereits 1805 wurde beispielsweise die Methode der kleinsten Quadrate von Legendre publiziert.

436

## Übung: EK1 Statistik ausfüllen



Anforderungen für die Statistik

Statistik

Angaben zur Studiendesign und Statistik

Design:

☐ randomisiert ☐ nicht randomisiert ☐ parallelisiert ☐ überkreuzt ☐ Interventionsstudien

☐ kontrolliert ☐ nicht kontrolliert ☐ ja ☐ nein

☐ Interventionsstudie ☐ ja ☐ nein

☐ Beobachtungsstudie ☐ ja ☐ nein

☐ Verblindung ☐ ja ☐ nein

☐ Haupt- und Nebeneffekte: ☐ ja ☐ nein

☐ Neben- und Alternativhypothese: ☐ ja ☐ nein

☐ Null- und Alternativhypothese: ☐ ja ☐ nein

☐ statistische Analyse: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

☐ Zweisamkeit: ☐ ja ☐ nein

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 434

## Die wichtigsten Regressionsanalysen



Multivariable Analyse: Regression von k unabhängigen auf 1 abhängige Variable

Abhängige Variable ist metrisch (stetig): Linear Regression

z.B. Geschlecht, Alter, BMI --> systolischer Blutdruck

Geschätzt (berechnet) wird das (standardisierte) Beta

Abhängige Variable ist kategorial: Logistische Regression

z.B. Geschlecht, Alter, BMI --> KHK in den nächsten 10 Jahren

Geschätzt wird das Odds Ratio

Abhängige Variable ist eine Ereigniszeit: Cox proportional hazards Regression

z.B. Geschlecht, Alter, BMI --> Zeit bis KHK (survival analysis)

Geschätzt wird das Hazard Ratio

437

## Confounding, Moderation, Mediation anhand einer Fallstudie erklärt



Epidemiology • Volume 26, Number 2, March 2015

Mediated Effect of Overweight and Obesity on CHD

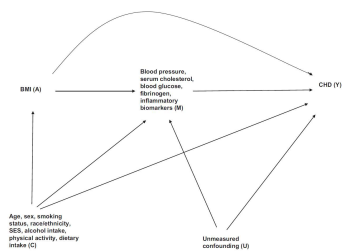


FIGURE 1. Causal diagram of the relation among BMI (A), metabolic risk factors, prothrombotic and inflammatory biomarkers (B), and CHD (Y) with measured confounders (C) and unmeasured confounding (U). Measured confounders were pre-baseline variables.

09.01.2025

hanno.ulmer@i-med.ac.at

Seite 435

## Confounding, Moderation, Mediation



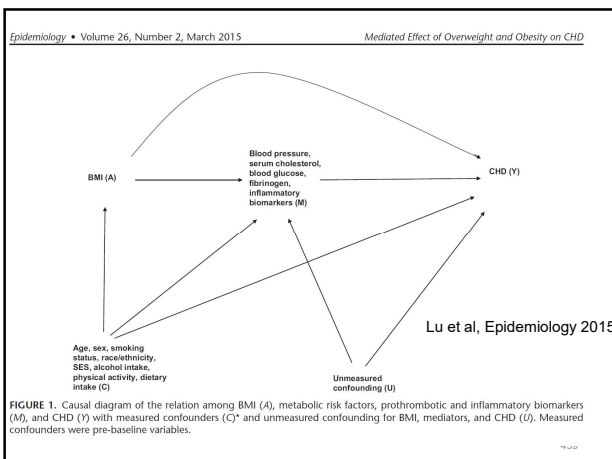
Die Regressionsanalyse ermöglicht es, den Effekt einer Prädiktorvariable (z.B. Adipositas) auf eine Zielvariable (z.B. KHK) unter Berücksichtigung von 'dritten Faktoren' (z.B. Geschlecht, Alter, Rauchen, Blutdruck, Cholesterin, Diabetes, etc.) abzuschätzen.

Diese 'dritten Faktoren' können als Confounder, Moderatoren, oder Mediatoren agieren, je nach angenommenen kausalen Wirkzusammenhang.

Die drei Konzepte werden nun am Beispiel (engl.) BMI ---> KHK illustriert.

438





## Example

Relationship between BMI and CHD incidence:

Crude Hazard Ratio

Obesity (30+ kg/m<sup>2</sup>) versus normal weight (20-25 kg/m<sup>2</sup>)

HR = 2.54 95%CI (2.32-2.78)

Sex, age and smoking adjusting hazard ratio:

HR = 1.60 95%CI (1.46-1.75)

Adjusted = controlled for confounding

Calculated with Cox proportional hazards regression analysis

442

## Example data

Vorarlberg Health Examinations (VHM&PP)

|                                                    |                                             |
|----------------------------------------------------|---------------------------------------------|
| Sex (male, female)                                 | categorical                                 |
| Age in years                                       | continuous                                  |
| Year of examination                                | continuous                                  |
| Body mass index in kg/m <sup>2</sup>               | continuous                                  |
| Systolic blood pressure in mmHG                    | continuous                                  |
| Total cholesterol in mg/dl                         | continuous                                  |
| Fasting glucose in mg/dl                           | continuous                                  |
| Smoking (current or past, never)                   | categorical                                 |
| Coronary heart disease mortality (ICD-10: I20-I25) | time to event<br>continuous and categorical |

440

## Confounding

Three essential characteristics:

The confounder is associated with the exposure of interest (BMI)

The confounder is associated with the disease (CHD)

The confounder is not in the causal pathway leading from the exposure of interest (BMI) to the disease of interest (CHD)

443

## Confounding

Confounding:

A "mixing of the effect" of the exposure-disease relationship with a third (or more) factors

BMI -----> CHD

< ----- sex, age, smoking ----- >

441

## Methods for Preventing Confounding in Study Designs

1. Stringent inclusion criteria to narrow the variability between study participants

2. Randomization (intervention/RCT only)  
In an optimal RCT, study groups only differ regarding the intervention

3. Matching (observational studies):

Simple Matching e.g. for age and sex in case-controls studies  
versus

Propensity Score Matching (involves logistic regression analysis)

Very popular in clinical research:  
Blackstone EH. Comparing apples and oranges. J Thoracic and Cardiovascular Surgery 2002; 1: 8-15.

An example:  
Ruttmann E et al. Second internal thoracic artery versus radial artery in coronary artery bypass grafting: a long-term, propensity score-matched follow-up study. Circulation. 2011 20;124(12):1321-9.

444



## Effect Modification/Moderation



Effect modification occurs when the association between the exposure (BMI) and the disease (CHD) varies by levels of a third factor.

How to assess: include interaction terms into the regression model

Interaction age\*obesity  $p < 0.001$

Young: BMI ----- > CHD  
Old: BMI ----- > CHD

445

## Example



Mediators in the relationship between BMI and CHD incidence:

Sex, age and smoking adjusting hazard ratio:

Total effect of BMI (obesity versus normal weight) on CHD:  
HR = 1.70 95%CI (1.57-1.85)

Direct effect of BMI on CHD  
HR = 1.30 95%CI (1.15-1.47)

Indirect effect mediated by blood pressure, cholesterol and glucose  
HR = 1.31 95%CI (1.16-1.48) (95%CI estimated by Bootstrap)

HRs ... multiplicative, do not add

448

## Example



Relationship between BMI and CHD incidence moderated by age:  
Interaction age\*obesity  $p < 0.001$

Obesity (30+ kg/m<sup>2</sup>) versus normal weight (20-25 kg/m<sup>2</sup>)  
Sex, age and smoking adjusting hazard ratio:

<50 years of age:

HR = 3.13 95%CI (2.27-4.31)

50+ years of age:

HR = 1.51 95%CI (1.37- 1.66)

446

## Example



Mediators in the relationship between BMI and CHD incidence:

Effect of BMI on CHD mediated by blood pressure, cholesterol and glucose

PERM (Percentage of excess risk mediated) =  
 $(1.70-1.30)/(1.70-1) \times 100$   
= 57% (approximate formula)

Global Burden of Metabolic Risk Factors for Chronic Diseases  
Collaboration. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lancet. 2014 Mar 15;383(9921):970-83

449

## Mediation



Mediation occurs if factors, like confounders, are associated with the exposure of interest (BMI) and the disease (CHD), but are **in the causal pathway** leading from the exposure to the disease.

These factors are called mediators:

BMI ---- > blood Pressure, cholesterol, diabetes ---- > CHD

447

## Mediation Techniques



Traditional approach:

Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986 Dec;51(6):1173-82

New approaches:

Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. Am J Epidemiol. 2014 Feb 15;179(4):513-8.

VanderWeele T. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press 2015.

New approaches applied on BMI ---- > CHD problem:

Lu Y, Hajifathalian K, Rimm EB, Ezzati M, Danaei G. Mediators of the effect of body mass index on coronary heart disease: decomposing direct and indirect effects. Epidemiology. 2015 Mar;26(2):153-62.

450



## Do risk factors explain the sex/gender gap in mortality from coronary heart disease?

Josef Fritz, Michael Edlinger, Cecily Kelleher, Susanne Strohmaier, Gabriele Nagel, Hans Concin, Elfriede Ruttman, Margarethe Hochleitner, Hanno Ulmer

Medical University of Innsbruck, Austria, University College Dublin, Ireland  
University of Oslo, Norway, University of Ulm, Germany  
Agency for Preventive and Social Medicine, Bregenz, Austria

451

## Material and methods

- We used prospective cohort data from the Vorarlberg Health Monitoring and Promotion Programme (VHM&PP), Austria
- A total of 172,262 individuals underwent baseline health examinations with fasting measurements of RFs
- There were 3,892 CHD deaths during a follow-up of 14.6 years

- For data analysis, we used a recently developed statistical mediation method (Lange et al, Am J Epidemiol 2014)
- Designed for survival data
- Allowing breakdown into single components of the indirect sex effect (that is explained by the RFs)

454

## Purpose

- Age and sex are the strongest predictors of coronary heart disease mortality
- Premature CHD (I20-I25) deaths – before age 65 - in Europe: 330,000 death cases, of which 77% in males, 23% in females (Nichols et al, Eur Heart J 2014)
- Aim of study: to estimate, how much of this large sex difference is explained by the major risk factors (RFs):
  - systolic blood pressure
  - total cholesterol
  - fasting glucose
  - smoking

452

## Results and conclusions

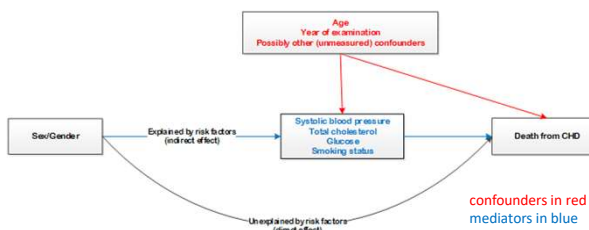
- The mortality difference between sexes decreased with age
- <50 years: HR=4.7 (95%CI 3.5-6.1)  
≥50 years: HR=1.9 (95%CI 1.7-2.1)

- The extent to which risk factors contributed varied with age
- <50 years: the 4 RF explained 41% (95%CI 27-54%) of sex effect  
≥50 years: the 4 RF explained 8% (95%CI 4-12%) of sex effect

- In younger individuals, the female survival advantage was explained to a substantial part through the pathways of the 4 major risk factors
- As blood pressure and cholesterol were the strongest factors, our results correspond to the oestrogen/testosterone thesis

455

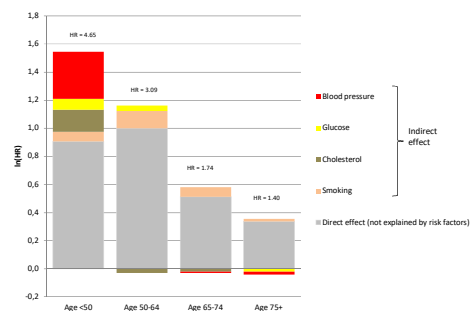
Figure 1. Underlying model



We assume that the 4 RFs are in the causal chain between sex and mortality, mediating the total sex effect, e.g. male sex causes hypertension, and hypertension causes CHD

453

Figure 2. What risk factors explain



456



Statistical errors in medical research –  
a review of common pitfalls

Alexander M. Strausk, Qamaruz Zaman, Karl P. Pfeiffer, Georg Gübel, Hanno Ulmer

**Table 3**Errors related to the  
documentation of  
statistical methods  
applied.

Failure to specify/define all tests used clear and correctly

Failure to state number of tails

Failure to state if test was paired or unpaired

Wrong names for statistical tests

Referring to unusual or obscure methods without explanation  
or referenceFailure to specify which test was applied on a given set of data  
if more than one test was done

"Where appropriate" statement

## Appendix (SMW-Artikel)

**Table 1**Statistical errors and  
deficiencies related  
to the design of a  
study.

Study aims and primary outcome measures not clearly stated or unclear

Failure to report number of participants or observations (sample size)

Failure to report withdrawals from the study

No a priori sample size calculation/effect-size estimation (power calculation)

No clear a priori statement or description of the Null-Hypothesis under investigation

Failure to use and report randomisation

Method of randomisation not clearly stated

Failure to use and report blinding if possible

Failure to report initial equality of baseline characteristics and comparability of study groups

Use of an inappropriate control group

Inappropriate testing for equality of baseline characteristics

**Table 4**Statistical errors and  
deficiencies related  
to the presentation  
of study data.

Inadequate graphical or numerical description of basic data

Mean but no indication of variability of the data

Giving SE instead of SD to describe data

Use of mean (SD) to describe non-normal data

Failure to define a notion for describing variability or use  
of unlabeled error bars

Inappropriate and poor reporting of results

Results given only as p-values, no confidence intervals given

Confidence intervals given for each group rather than  
for contrasts"p = NS", "p < 0.05" or other arbitrary thresholds instead  
of reporting exact p-valuesNumerical information given to an unrealistic level  
of precision**Table 2**Statistical errors and  
deficiencies related  
to data analysis.

Use of wrong statistical tests

Incompatibility of statistical test with type of data examined

Unpaired tests for paired data or vice versa

Inappropriate use of parametric methods

Use of an inappropriate test for the hypothesis under  
investigation

Inflation of Type I error

Failure to include a multiple-comparison correction

Inappropriate post-hoc Subgroup analysis

Typical errors with Student's t-test

Failure to prove test assumptions

Unequal sample sizes for paired t-test

Improper multiple pair-wise comparisons of more than  
two groups

Use of an unpaired t-test for paired data or vice versa

Typical errors with  $\chi^2$ -tests

No Yates-continuity correction reported if small numbers

Use of chi-square when expected numbers in a cell are &lt; 5

No explicit statement of the tested Null-Hypotheses

Failure to use multivariate techniques to adjust for confounding  
factors**Table 5**Statistical errors and  
deficiencies related  
to the interpretation  
of study findings.

Wrong interpretation of results

"non significant" interpreted as "no effect",  
or "no difference"

Drawing conclusions not supported by the study data

Significance claimed without data analysis or statistical test  
mentioned

Poor interpretation of results

Disregard for Type II error when reporting non-significant  
resultsMissing discussion of the problem of multiple significance  
testing if doneFailure to discuss sources of potential bias and confounding  
factors