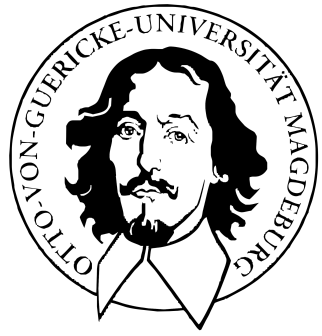


Interventionsforschung

MSc Klinische Psychologie und Psychotherapie

MSc Umweltpsychologie / Mensch-Technik-Interaktion

SoSe 2026

Prof. Dr. Dirk Ostwald

(7) Metaanalyse

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Definitionsversuch

Metaanalyse

... ist eine Gruppe quantitativer Methoden zur Kombination von Evidenz

... ist die Analyse der Resultate statistischer Analysen

... ist überwiegend frequentistisch geprägt

Glass (1976) „Primary, Secondary, and Meta-Analysis of Research“

„My major interest currently is in what we have come to call — not for want of a less pretentious name — the **meta-analysis of research**. The term is a bit grand, but it is precise, and apt, and in the spirit of ‚meta-mathematics,‘ ‚meta-psychology,‘ and ‚meta-evaluation.‘ Meta-analysis refers to the analysis of analyses. I use it to refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.“

Glass (1976)

Ethischer Imperativ

- Die Öffentlichkeit investiert substantielle Mittel in die (außer)universitäre Forschung.
- Neben wenigen Großforschungsprojekten gibt es viele kleine Forschungsprojekte.
- Die Kleinforschungsprojekte sind oft nicht zentralisiert und interessens- und zufallsgeleitet.
- Empirische Wissenschaft basiert auf der Replikation oder Nichtreplikation von Erkenntnissen.
- Die Analyse von Kleinforschungsergebnissen hat den potenziellen Mehrwert von Großforschung.
- Es ist intuitiv sinnvoll und plausibel nach der gesammelten Evidenz für etwas zu fragen.
- Metaanalysen wird das Potenzial zugeschrieben, aussagekräftiger als Einzelstudien zu sein.
- Entscheidungsrelevante Aussagen werden gerne aufgrund großer Datenmengen getroffen.
- Forschung hat die Pflicht, integrierte Evidenz dem Wohle der Gesellschaft bereitzustellen.
- Gegeben ein vernünftiges FDM sind spezielle Metaanalyse-Methoden eigentlich überflüssig.

Borenstein (2009), Harrer et al. (2021)

Einführung

Narrative Reviews

- Standard der Akkumulation wissenschaftlicher Evidenz bis etwa 1990
- Subjektive und intransparente Auswahl von Studien durch Expert:innen
- Nicht-explizite und qualitative Gewichtung von Studien
- Relativ gut möglich bei eher geringer Anzahl von zu inkludierenden Studien

Systematische Reviews und Metaanalysen

- Standard der Akkumulation wissenschaftlicher Evidenz seit etwa 2000
- Transparente Auswahl von Studien anhand festgelegter Kriterien
- Explizite, quantitative Wichtung von Studien mithilfe metaanalytischer Statistik
- Hilfreich bei einer großen Anzahl von zu inkludierenden Studien durch Automatisierung

Borenstein (2009), Harrer et al. (2021)

Metaanalysen in der Psychotherapieforschung

Eysenck (1952)

"A survey was made of reports on the improvement of neurotic patients after psychotherapy, and the results compared with the best available estimates of recovery without benefit of such therapy. **The figures fail to support the hypothesis that psychotherapy facilitates recovery from neurotic disorder**. In view of the many difficulties attending such actuarial comparisons, no further conclusions could be derived from the data whose shortcomings highlight the necessity of properly planned and executed experimental studies into this important field."

Smith (1977)

"Results of nearly 400 controlled evaluations of psychotherapy and counseling were coded and integrated statistically. **The findings provide convincing evidence of the efficacy of psychotherapy**. On the average, the typical therapy client is better off than 75% of untreated individuals. Few important differences in effectiveness could be established among many quite different types of psychotherapy. More generally, virtually no difference in effectiveness was observed between the class of all behavioral therapies (systematic desensitization, behavior modification) and the nonbehavioral therapies (Rogerian, psychodynamic, rational-emotive, transactional analysis, etc.)."

Cuijpers et al. (2019)

"In the 1950s, Eysenck suggested that psychotherapies may not be effective at all. Twenty-five years later, the first meta-analysis of randomised controlled trials showed that the effects of psychotherapies were considerable and that Eysenck was wrong. However, since that time methods have become available to assess biases in meta-analyses. We examined the influence of these biases on the effects of psychotherapies for adult depression, including risk of bias, publication bias and the exclusion of waiting list control groups. The unadjusted effect size of psychotherapies compared with control groups was g = 0.70 (limited to Western countries: g = 0.63) (…). Only 23% of the studies could be considered as a low risk of bias. **When adjusting for several sources of bias, the effect size across all types of therapies dropped to g = 0.31.**"

Borenstein (2009), Harrer et al. (2021). Kennedy et al. (2026)

Metaanalysen in der Medizin

- Inhaltliche Beiträge, z.B. Peto and Parish (1980)
- Richtlinien, z.B. Moher et al. (2009)

Cochrane Collaboration

- Stiftung zur Förderung der evidenzbasierten metaanalytischen Medizin seit 1993
- www.cochrane.org

Campbell Collaboration

- Stiftung zur Förderung der evidenzbasierten metaanalytischen Sozialwissenschaft seit 2000
- www.campbellcollaboration.org

Typische Charakteristika des Metaanalyse-Genres

- Orientierung am PRISMA Statement (Page et al. (2021))
- Spezifikation von Suchtermen für Studiendatenbanken wie Web of Science
- Spezifikation des Studienscreenings
- Extraktion von Effektstärken (Cohen's d)
- Korrektur von Effektstärken (Hedges' g)
- Angabe eines "mittleren" Hedges' g
- Mehrebenenanalyse mithilfe eines Linear Mixed Models
- Versuche der Qualifikation und evtl. Quantifizierung und Korrektur von Biases

Borenstein (2009), Harrer et al. (2021)

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Checklist

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Checklist

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	
	23b	Discuss any limitations of the evidence included in the review.	
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

Versuche der Qualifikation von Biases

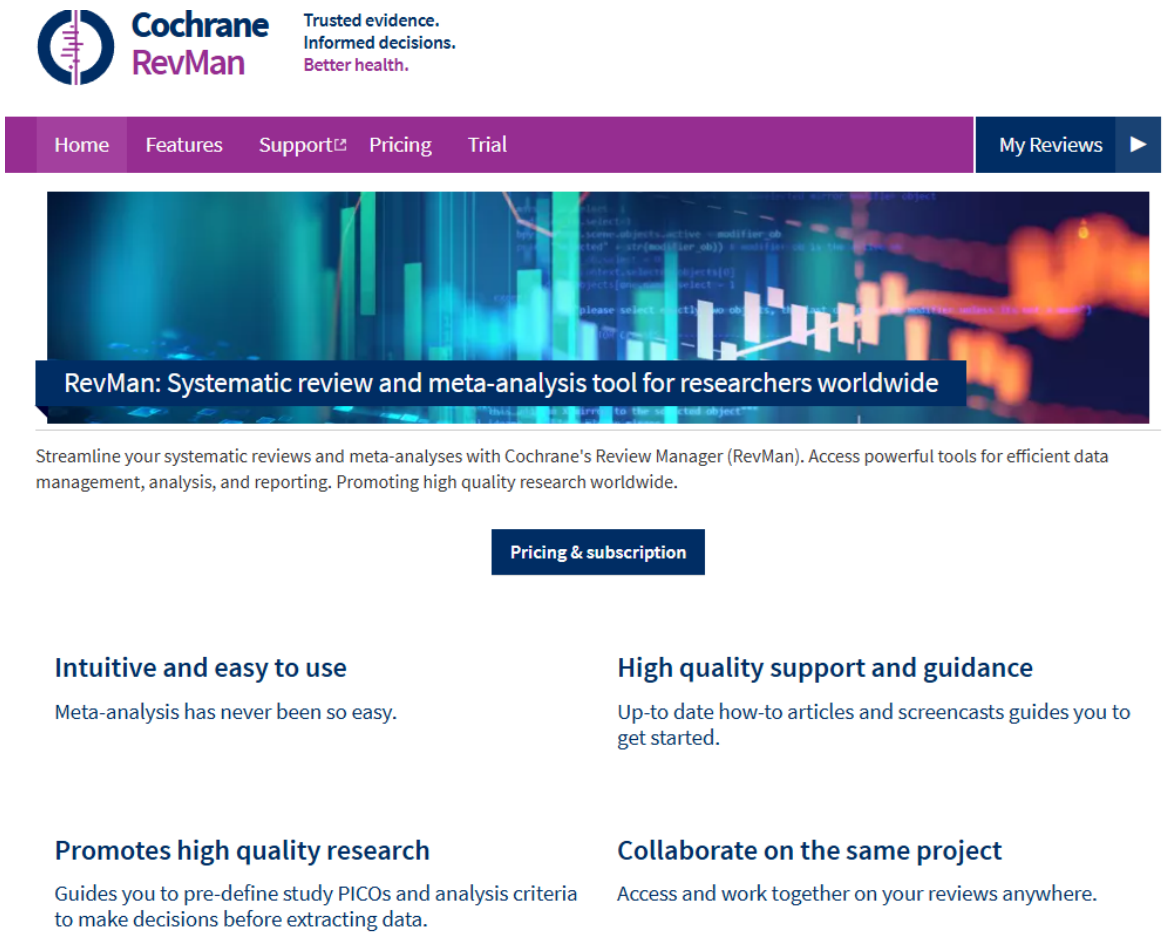
Einschätzung des Verzerrungsrisikos anhand des Cochrane Risk of Bias Tools

Domains:
D1: Bias arising from the randomization process.
D2: Bias due to deviations from intended intervention.
D3: Bias due to missing outcome data.
D4: Bias in measurement of the outcome.
D5: Bias in selection of the reported result.

Judgement
- X High
- − Some concerns
- \+ Low

Schaeuffele et al. (2024)

Cochrane Risk of Bias Tool als Teil der kommerziellen Cochrane RevMan Software



Typische Probleme des Metaanalyse-Genres

Apples and oranges

- Metaanalysen integrieren Studien, die sich in experimentellen Details unterscheiden
- Die zulässige experimentelle Varianz hängt von der metaanalytischen Fragestellung ab

Garbage In, Garbage Out

- Die Qualität eines metaanalytischen Ergebnisses hängt von der Qualität der Primärstudien ab
- Bei geringer Primärstudienqualität kann eine Metaanalyse zumindest diese Erkenntnis bringen

The File Drawer Problem

- Nicht jede themenrelevante Forschung wird tatsächlich auch publiziert
- Insbesondere sind meist hochwertig angefertigte Nullresultatstudien unterrepräsentiert

Researcher Agendas

- Jeder Studienreport, auch eine Metaanalyse, ist letztlich ein Kommunikationsakt
- Die Kommunikationssender sind bewusst oder unbewusst nicht frei von eigenen Motiven

Borenstein (2009), Harrer et al. (2021)

Anwendungsbeispiel

Zhang et al. (2026) Efficacy of Acceptance and Commitment Therapy for Self-Injurious Thoughts and Behaviors: A Systematic Review and Meta-Analysis

- Fragestellung: Wirksamkeit von Acceptance and Commitment Therapy (ACT) zur Reduktion selbstverletzender Gedanken und Verhaltensweisen (SITBs). ACT soll psychologische Flexibilität stärken, also Akzeptanz schwieriger Erfahrungen und werteorientiertes Handeln fördern.
- SITBs umfassen Suizidgedanken, Suizidpläne und -versuche, Selbstverletzung und nonsuizidale Selbstverletzung (NSSI). Der folgende Datensatz trennt diese Outcomes für die aus Figure 2 extrahierten Einzelstudieneffekte.
- Eingeschlossen wurden randomisierte kontrollierte ACT-Studien mit SITB-Outcome Messung nach Interventionsende; insgesamt 48 Studien ($N = 4719$), davon 35 in der Metaanalyse. Kontrollbedingungen waren u.a. Warteliste, Treatment-as-usual, aktive Kontrolle oder Placebo; kombinierte ACT-Interventionen wurden ausgeschlossen.
- Tabelle 2 berichtet Hedges' g aus Random-Effects-Modellen, Heterogenität (Q, I^2) und Egger-Tests; negative Effekte sprechen für ACT gegenüber Kontrollbedingungen. Die Effekte werden für Postintervention und Follow-up ausgewiesen, für Follow-up lagen nur für einige Outcomes ausreichend Studien vor.

Anwendungsbeispiel

Symptom	Study	g	SE	Vi	CIl	CIu
Suicide ideation	Abyar 2018	-0.95	0.44	0.19	-1.81	-0.09
Suicide ideation	Azar 2023	-1.26	0.52	0.27	-2.28	-0.23
Suicide ideation	Barnes 2021	0.07	0.30	0.09	-0.52	0.66
Suicide ideation	Demehri 2022	-0.24	0.36	0.13	-0.94	0.46
Suicide ideation	El-Sayed 2023	-1.58	0.29	0.08	-2.15	-1.01
Suicide ideation	Ghadam 2020	-0.31	0.43	0.19	-1.16	0.53
Suicide ideation	Gould 2024	-0.63	0.90	0.81	-2.39	1.14
Suicide ideation	Khorani 2020	-2.57	0.42	0.18	-3.40	-1.74
Suicide ideation	Kumpula 2019	0.22	0.06	0.00	0.10	0.35
Suicide ideation	Lang 2017	0.11	0.21	0.04	-0.30	0.51
Suicide ideation	Shareh 2022	-0.61	0.26	0.07	-1.12	-0.10
Suicide ideation	Tighe 2017	0.00	0.26	0.07	-0.50	0.50
Suicide ideation	WangB 2022	-1.29	0.24	0.06	-1.77	-0.81
Suicide ideation	WangR 2017	-0.44	0.26	0.07	-0.95	0.06
Suicide attempt	Geng 2021	-0.79	0.37	0.14	-1.52	-0.06
Suicide attempt	Gong 2020	-0.70	0.32	0.10	-1.34	-0.07
Suicide attempt	Gould 2024	-0.63	0.90	0.81	-2.39	1.14
Suicide attempt	Parsa 2023	-0.26	0.54	0.29	-1.31	0.79
Self-harm	Azar 2023	-0.98	0.50	0.25	-1.97	0.00
Self-harm	Derakhshan 2020	-2.30	0.46	0.21	-3.20	-1.39
Self-harm	Ghanbari 2020	-0.61	0.32	0.10	-1.23	0.01
Self-harm	Ghorbanzadeh 2021a	-2.39	0.47	0.22	-3.31	-1.47
Self-harm	Shirazi 2022	-1.54	0.30	0.09	-2.13	-0.94
NSSI	Chen 2024	-0.66	0.32	0.10	-1.29	-0.04
NSSI	Fang 2021	-0.93	0.45	0.20	-1.81	-0.04
NSSI	Hu 2022	0.27	0.40	0.16	-0.51	1.05
NSSI	WangJ 2023a	-1.08	0.49	0.24	-2.04	-0.11
NSSI	WangJ 2023b	-0.84	0.36	0.13	-1.54	-0.14
NSSI	Yuan 2023	-0.38	0.24	0.06	-0.85	0.08
NSSI	Yuan 2024	-0.58	0.24	0.06	-1.04	-0.11
NSSI	Zhang 2023	-1.05	0.38	0.14	-1.79	-0.30

Anwendungsbeispiel

Table 2. Effects, heterogeneity, and publication bias across outcomes

Outcome	Sample size		Effect size		Heterogeneity		Egger's test	
	k	*n*	Hedges' *g* [95% CI]	*p* value	*Q*	I^2, %	*z*	*p* value
Post-intervention								
Suicide ideation[a]	14	2,652	−0.64 [−1.06, −0.22]	0.003	126.37	89.75	−1.48	0.138
Suicide attempt	4	541	−0.66 [−1.08, −0.23]	0.002	0.71	0.00	0.40	0.686
Self-harm[b]	5	181	−1.53 [−2.22, −0.84]	<0.001	15.27	74.10	−0.90	0.367
NSSI[c]	8	374	−0.59 [−0.82, −0.36]	<0.001	8.98	3.68	−0.99	0.322
Overall SITBs	35	3,876	−0.99 [−1.30, −0.67]	<0.001	369.58	90.59	−2.26	0.024
Follow-up								
Suicide ideation	5	260	−2.15 [−3.58, −0.73]	0.003	50.64	95.03	−4.73	<0.001
NSSI[d]	3	174	−1.18 [−1.71, −0.64]	<0.001	5.47	61.53	−0.08	0.935
Overall SITBs	9	458	−1.52 [−2.45, −0.59]	0.001	76.85	94.56	−2.69	0.007

No studies reported on suicide attempt and self-harm at follow-up. *k*, number of studies; *n*, number of participants; 95% CI, 95% confidence interval; NSSI, nonsuicidal self-injury; SITB, self-injurious thoughts and behaviors. [a]Exclude outlier Ghorbanzade et al. [102]. Before excluding the outlier, the effect size was *g* = −0.82 (95% CI = −1.33 to −0.30; see online suppl. C1 for further details). [b]Exclude outlier Mollashahi et al. [118]. Before excluding the outlier, the effect size was *g* = −1.89 (95% CI = −2.78 to −1.00; see online suppl. C3 for further details). [c]Exclude outlier Xue et al. [112]. Before excluding the outlier, the effect size was *g* = −0.81 (95% CI = −1.23 to −0.39; see online suppl. C4 for further details). [d]Exclude outlier Hu [108]. Before excluding the outlier, the effect size was *g* = −0.79 (95% CI = −1.65 to 0.07; see online suppl. C4 for further details).

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Effektstärkeschätzung

Cohen's d

- Standardisierte Mittelwertsdifferenz
- Mit Stichprobenumfangsfaktor multiplizierter T-Wert des Zweistichproben-T-Tests

Hedges' g

- Schätzfehler-korrigiertes Cohen's d für Studien mit geringem Stichprobenumfang
- Ab einem Stichprobenumfang > 30 im Wesentlichen mit Cohen's d identisch

Cohen's d in den Worten von Jacob Cohen

Thus, we see that the absence of the phenomenon under study is expressed by a null hypothesis which specifies an exact value for a population parameter, one which is appropriate to the way the phenomenon under study is manifested. Without intending any necessary implication of causality, **it is convenient to use the phrase "effect size" to mean "the degree to which the phenomenon is present in the population," or "the degree to which the null hypothesis is false."**

Whatever the manner of representation of a phenomenon in a particular research in the present treatment, the null hypothesis always means that the effect size is zero. By the above route, it can now readily be made clear that when the null hypothesis is false, it is false to some specific degree, i.e., the effect size (ES) is some specific nonzero value in the population. The larger this value, the greater the degree to which the phenomenon under study is manifested. Thus, in terms of the previous illustrations:

1. If the percentage of males in the population of psychiatric patients bearing a diagnosis of paranoid schizophrenia is 52%, and the effect is measured as a departure from the hypothesized 50%, the ES is 2%; if it is 60%, the ES is 10%, a larger ES.
2. If children of multiple births have a population mean IQ of 96, the ES is 4 IQ units (or- 4, depending on directionality of significance criterion); if it is 92, the ES is 8 (or - 8) IQ units, i.e., a larger ES.
3. (...)
4. If the population of consumers preferring brand A has a median annual income $700 higher than that of brand B, the ES is $700. If the population median difference and hence the ES is $1000, the effect of income on brand preference would be larger.

Cohen (1988) Statistical power analysis for the behavioral sciences

Cohen's d in den Worten von Jacob Cohen

Thus, whether measured in one unit or another, whether expressed as a difference between two population parameters or the departure of a population parameter from a constant or in any other suitable way, the ES can itself be treated as a parameter which takes the value zero when the null hypothesis is true and some other specific nonzero value when the null hypothesis is false, and in this way the ES serves as an index of degree of departure from the null hypothesis.

(...)

As noted above (Section 1.4), we need a "pure" number, one free of our original measurement unit, with which to index what can be alternately called the degree of departure from the null hypothesis or the ES (effect size) we wish to detect. This is accomplished by standardizing the raw effect size as expressed in the measurement unit of the dependent variable by dividing it by the common standard deviation of the measures in their respective populations.

$$d = \frac{m_A - m_B}{\sigma} \qquad \text{directional (one-tailed)}$$

$$d = \frac{|m_A - m_B|}{\sigma} \qquad \text{nondirectional (two-tailed)}$$

$d :=$ ES index for t tests of means in standard units

$m_A, m_B :=$ population means in the original measurement unit

$\sigma :=$ standard deviation of either population

Cohen (1988) Statistical power analysis for the behavioral sciences

Cohen's d in den Worten der metaanalytischen Community

Consider a study with two independent groups. Let μ_1 and σ_1 be the true mean and standard deviation of the first group and let μ_2 and σ_2 be the true mean and standard deviation of the second group. If $\sigma_1 = \sigma_2 = \sigma$, then the population standardized mean difference is defined as

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}$$

and can be estimated by

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}, \quad S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

The sample estimate of the standardized mean difference is often called Cohen's d in research synthesis. The parameter is denoted by δ, and d is the sample estimate of that parameter.

Borenstein (2009) Introduction to Meta-Analysis

Definition (Cohen's d im Parallelgruppendesign)

Gegeben sei ein Parallelgruppendesign mit einer Treatmentgruppe t und einer Kontrollgruppe c und es seien

$$y_{t1}, ..., y_{tn_t} \text{ und } y_{c1}, ..., y_{cn_c} \tag{1}$$

die skalaren reellen Werte des primären Ergebnismaßes jeder Gruppe. Es seien weiterhin

$$\bar{y}_t := \frac{1}{n_t}\sum_{j=1}^{n_t} y_{tj}, \quad s_t^2 := \frac{1}{n_t - 1}\sum_{j=1}^{n_t}\left(y_{tj} - \bar{y}_t\right)^2 \text{ und } \bar{y}_c := \frac{1}{n_c}\sum_{j=1}^{n_c} y_{cj}, \quad s_c^2 := \frac{1}{n_c - 1}\sum_{j=1}^{n_c}\left(y_{cj} - \bar{y}_c\right)^2 \tag{2}$$

die Mittelwerte und die empirischen Varianzen beider Gruppen. Schließlich sei

$$s := \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}} \tag{3}$$

die *gepoolte empirische Standardabweichung* der beiden Gruppen. Dann ist *Cohen's d im Parallelgruppendesign* definiert als

$$d := \frac{\bar{y}_t - \bar{y}_c}{s}. \tag{4}$$

Effektstärkeschätzung

Bemerkungen

- $\bar{y}_t - \bar{y}_c$ ist die *empirische Effektstärke*.
- d ist die mithilfe der gepoolten empirischen Standardabweichung *standardisierte empirische Effektstärke*.
- Im hier gewählten Vorzeichen entspricht die empirische Effektstärke dem Unterschied Treatment - Kontrolle.
- d misst den Unterschied der Mittelwerte in Einheiten der gepoolten Standardabweichung, z.B.

$$\bar{y}_t - \bar{y}_c = 0 \Leftrightarrow d = 0 \quad \bar{y}_t - \bar{y}_c = s \Leftrightarrow d = 1 \quad \bar{y}_t - \bar{y}_c = 2s \Leftrightarrow d = 2. \tag{5}$$

- Für $\mu_0 = 0$ gilt im Sinne der Zweistichproben-T-Teststatistik, dass

$$T = \sqrt{\frac{n_t n_c}{n_t + n_c}}\, d \Leftrightarrow d = \sqrt{\frac{n_t + n_c}{n_t n_c}}\, T. \tag{6}$$

- Ersetzt man in der Definition von Cohen's d die gepoolte empirische Standardabweichung s durch die empirische Standardabweichung der Kontrollgruppe, so erhält man das nach Glass (1976) benannte

$$\Delta := \frac{\bar{y}_t - \bar{y}_c}{\sqrt{s_c^2}}. \tag{7}$$

- Glass Δ wird heutzutage allerdings eher selten als Effektstärkenmaß im metaanalytischen Kontext betrachtet.

Effektstärkeschätzung

Bestimmung von Cohen's d für Abyar et al. (2018) anhand simulierter Rohdaten

Group	BSSI
ACT	12.49
ACT	12.77
ACT	13.05
ACT	13.32
ACT	13.60
ACT	13.88
ACT	14.15
ACT	14.43
ACT	14.71
ACT	14.99
ACT	15.26
ACT	15.54
WLC	13.47
WLC	13.75
WLC	14.03
WLC	14.31
WLC	14.58
WLC	14.86
WLC	15.14
WLC	15.42
WLC	15.69
WLC	15.97
WLC	16.25
WLC	16.53

Effektstärkeschätzung

Bestimmung von Cohen's d für Abyar et al. (2018) anhand simulierter Rohdaten

```
D       = read.csv("7-Daten/abyar-data.csv")                    # Einlesen der simulierten Rohdaten
y_t     = D$BSSI[D$Group == "ACT"]                              # Daten der Treatmentgruppe
y_c     = D$BSSI[D$Group == "WLC"]                              # Daten der Kontrollgruppe
n_t     = length(y_t)                                           # Umfang Treatmentgruppe
n_c     = length(y_c)                                           # Umfang Kontrollgruppe
bar_y_t = mean(y_t)                                             # Mittelwert der Treatmentgruppe
bar_y_c = mean(y_c)                                             # Mittelwert der Kontrollgruppe
s2_t    = var(y_t)                                              # empirische Varianz der Treatmentgruppe
s2_c    = var(y_c)                                              # empirische Varianz der Kontrollgruppe
s       = sqrt(((n_t-1)*s2_t + (n_c-1)*s2_c)/(n_t + n_c -2))    # gepoolte Standardabweichung
es      = bar_y_t - bar_y_c                                     # empirische Effektstärke
d       = (bar_y_t - bar_y_c)/s                                 # Cohen's d
```

```
Mittelwert Treatmentgruppe             : 14.02
Mittelwert Kontrollgruppe              : 15
Gepoolte empirische Standardabweichung : 1
Empirische Effektstärke                : -0.98
Cohen's d                              : -0.98
Publiziertes Hedges' g                 : -0.95
```

Effektstärkeschätzung

Bestimmung von Cohen's d für Abyar et al. (2018) anhand simulierter Rohdaten mit `metafor`

```
library(metafor)                                          # R Paket
E       = escalc(                                         # Effektstärke
measure = "SMD",                                          # Standardized Mean Difference
m1i     = mean(y_t),                                      # Mittelwert Treatmentgruppe
sd1i    = sd(y_t),                                        # Standardabweichung Treatmentgruppe
n1i     = n_t,                                            # Umfang Treatmentgruppe
m2i     = mean(y_c),                                      # Mittelwert Kontrollgruppe
sd2i    = sd(y_c),                                        # Standardabweichung Kontrollgruppe
n2i     = n_c,                                            # Umfang Kontrollgruppe
correct = FALSE)                                          # keine Hedges-Korrektur
cat("Cohen's d (metafor):", round(E$yi, digits = 2))      # Ausgabe von Cohen's d
```

```
Cohen's d (metafor): -0.98
```

Hedges' g

Hedges' g ist ein durch einen Multiplikationsfaktor angepasstes Cohen's d, das von Hedges (1981) und Hedges and Olkin (1985) eingeführt wurde. Grundlage für Hedges' g ist die frequentistische Verteilung von Cohen's d in einem Normalverteilungsmodell der Daten einer einzelnen Studie. In diesem Modell ist Cohen's d ein verzerrter Schätzer der wahren, aber unbekannten, standardisierten Effektstärke, der Erwartungswert von Cohen's d entspricht also nicht der wahren, aber unbekannten, standardisierten Effektstärke. Die Differenz zwischen dem Erwartungswert und der wahren, aber unbekannten, standardisierten Effektstärke wird *Bias* genannt.

Im betrachteten Normalverteilungsmodell kann der Bias von Cohen's d analytisch bestimmt und numerisch approximiert werden. Die Bekanntheit des Bias erlaubt dann seine Korrektur durch Skalierung von Cohen's d und der zur Biaskorrektur skalierte Wert von Cohen's d wird *Hedges' g* genannt. Dabei ist der Bias von Cohen's d für kleine Stichprobenumfänge am größten.

Letztlich ist Hedges' g einfach der Tatsache geschuldet, dass Cohen's d eng mit der T-Statistik verwandt ist, die bekanntlich bei Nicht-Zutreffen der Nullhypothese nichtzentral-t-verteilt ist und der Erwartungswert einer nichtzentralen-t-verteilten Zufallsvariable nicht mit ihrem Nichtzentralitätsparameter identisch ist. Dabei ist die Nicht-Normalität der Verteilung von Cohen's d für größere Stichprobenumfänge vernachlässigbar, was wiederum asymptotische Schätzungen der Varianz von Hedges' g ermöglicht.

Im Kontext von Metaanalysen werden Hedges' g und sein Varianzschätzer meist zunächst für jede Einzelstudie bestimmt und dann über Studien, z.B. mit den Fixed-Effects- und Random-Effects-Modellen der Metaanalyse, integriert.

Definition (Einzelstudienmodell)

Es seien y_{tj} für $j = 1, ..., n_t$ und y_{cj} für $j = 1, ..., n_c$ die Zufallsvariablen, die die primären Ergebnismaße der $j$ten experimentellen Einheit in der Treatment- und der Kontrollgruppe einer Einzelstudie modellieren, respektive. Dann ist das Einzelstudienmodell (ESM) gegeben durch

$$\begin{aligned} y_{tj} &:= \mu_t + \varepsilon_{tj} \text{ mit } \varepsilon_{tj} \sim N(0, \sigma^2) \text{ für } j = 1, ..., n_t \\ y_{cj} &:= \mu_c + \varepsilon_{cj} \text{ mit } \varepsilon_{cj} \sim N(0, \sigma^2) \text{ für } j = 1, ..., n_c \end{aligned} \tag{8}$$

und die wahre, aber unbekannte, standardisierte Effektstärke (SES) ist definiert als

$$\delta := \frac{\mu_t - \mu_c}{\sigma}. \tag{9}$$

Bemerkungen

- Das Einzelstudienmodell ist äquivalent zum aus Einheit 2 bekannten Parallelgruppendesign-Modell.
- Unter der 0/1-Kodierung des Parallelgruppendesign-Modells gilt $\beta_0 = \mu_c$ und $\beta_1 = \mu_t - \mu_c$
- Cohen's d standardisiert den entsprechenden Mittelwertunterschied.
- Mit dem Theorem zur linear-affinen Transformation von normalverteilten Zufallsvariablen gilt äquivalent

$$y_{tj} \sim N(\mu_t, \sigma^2) \text{ für } j = 1, ..., n_t \text{ und } y_{cj} \sim N(\mu_c, \sigma^2) \text{ für } j = 1, ..., n_c \tag{10}$$

- Die hier gewählte Formulierung entspricht Hedges and Olkin (1985) p. 76 (vgl. Lin and Aloe (2021))
- Hedges (1981) formuliert ein äquivalentes Modell basierend auf einer nicht-standardisierten Effektstärke.

Definition (SES-Schätzer im Einzelstudienmodell)

Gegeben sei ein Einzelstudienmodell mit wahrer, aber unbekannter, standardisierter Effektstärke $\delta \in \mathbb{R}$. Dann ist der *Standardisierte Effektstärkeschätzer (SES-Schätzer)* d von δ definiert als

$$d := \frac{\bar{y}_t - \bar{y}_c}{s}, \tag{11}$$

wobei

$$\bar{y}_t := \frac{1}{n_t} \sum_{j=1}^{n_t} y_{tj} \text{ und } \bar{y}_c := \frac{1}{n_c} \sum_{j=1}^{n_c} y_{cj} \tag{12}$$

die Stichprobenmittel der Treatment- und der Kontrollgruppe bezeichnen und s die mithilfe der Stichprobenvarianzen

$$s_t^2 := \frac{1}{n_t - 1} \sum_{j=1}^{n_t} \left(y_{tj} - \bar{y}_t\right)^2 \text{ und } s_c^2 := \frac{1}{n_c - 1} \sum_{j=1}^{n_c} \left(y_{cj} - \bar{y}_c\right)^2 \tag{13}$$

der Treatment- und Kontrollgruppe definierte *gepoolte Stichprobenstandardabweichung*

$$s := \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}. \tag{14}$$

bezeichnet.

Bemerkungen

- Der SES-Schätzer ist offenbar mit Cohen's d identisch.
- Für den SES-Schätzer werden in δ lediglich μ_t, μ_c und σ durch ihre empirischen Äquivalente ersetzt.

Theorem (Eigenschaften des SES-Schätzers im ESM)

Gegeben sei ein Einzelstudienmodell, es sei d der Schätzer der wahren, aber unbekannten, standardisierten Effektstärke δ und es seien

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2}\Gamma((m-1)/2)}. \tag{15}$$

Dann hat d folgende Eigenschaften:

(1) $\sqrt{n}d \sim t(\sqrt{n}\delta, m)$,

(2) $\mathbb{E}(d) = \frac{1}{c_m}\delta$, und

(3) $\mathbb{V}(d) = \frac{m}{(m-2)n}(1 + n\delta^2) - \frac{1}{c_m^2}\delta^2$.

Bemerkungen

- $\sqrt{n}d$ ist eine nichtzentral-t-verteilte Zufallsvariable.
- Der Nichtzentralitätsparameter von $\sqrt{n}d$ ist $\sqrt{n}\delta$, der Freiheitsgradparameter ist m.
- Für die Definition nichtzentral-t-verteilter Zufallsvariablen und ihre Kennzahlen verweisen wir auf den Anhang.
- Für $\delta \neq 0$ und $c_m \neq 1$ impliziert (2), dass $\mathbb{E}(d) \neq \delta$.
- d ist also ein verzerrter Schätzer von δ.

Beweis

(1) Wir verzichten auf einen Beweis und verweisen wie Hedges (1981) auf Johnson and Welch (1940).

(2) Mit dem Erwartungswert einer nichtzentral-t-verteilten Zufallsvariablen ergibt sich

$$\mathbb{E}(\sqrt{n}d) = \sqrt{n}\delta\sqrt{\frac{m}{2}}\,\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}\,. \tag{16}$$

Es ergibt sich also

$$\mathbb{E}(d) = \frac{\sqrt{n}}{\sqrt{n}}\mathbb{E}(d) = \frac{1}{\sqrt{n}}\mathbb{E}(\sqrt{n}d) = \delta\sqrt{\frac{m}{2}}\,\frac{\Gamma((m-1)/2)}{\Gamma(m/2)} = \frac{1}{c_m}\,\delta. \tag{17}$$

(3) Mit der Varianz von nichtzentral-t-verteilten Zufallsvariablen ergibt sich

$$\begin{aligned}
\mathbb{V}(\sqrt{n}d) &= \frac{m(1+(\sqrt{n}\delta)^2)}{m-2} - \frac{(\sqrt{n}\delta)^2 m}{2}\left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}\right)^2 \\
&= \frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{m}{2}\left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}\right)^2 \\
&= \frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{1}{c_m^2}\,.
\end{aligned} \tag{18}$$

Es ergibt sich also

$$\mathbb{V}(d) = \frac{n}{n}\mathbb{V}(d) = \frac{1}{n}\mathbb{V}(\sqrt{n}d) = \frac{1}{n}\left(\frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{1}{c_m^2}\right) = \frac{m}{(m-2)n}(1+n\delta^2) - \delta^2\,\frac{1}{c_m^2}\,. \tag{19}$$

Effektstärkeschätzung

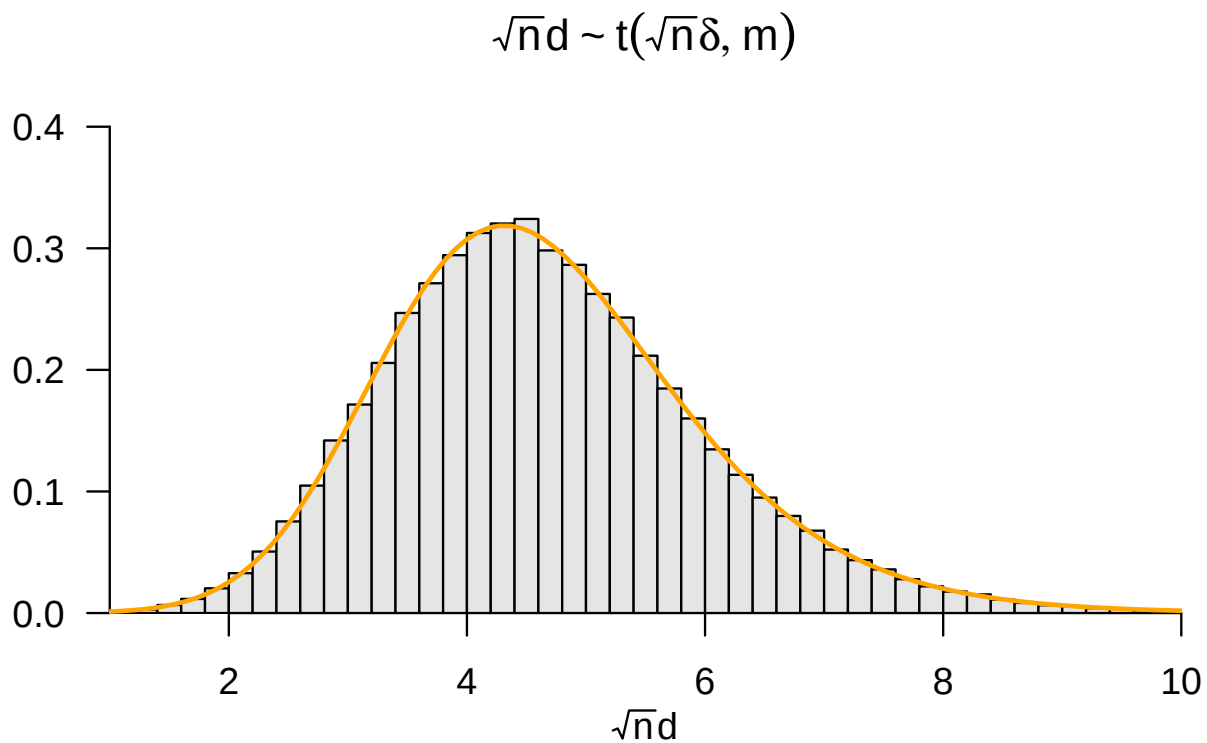
Verteilung von $\sqrt{n}d$

10^5 SES-Schätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$

```
sim      = 1e5                                  # Anzahl Realisierungen
n_t      = 10                                   # Treatmentgruppenumfang
n_c      = 10                                   # Kontrollgruppenumfang
mu_t     = 2                                    # Erwartungswertparameter Treatment
mu_c     = 0                                    # Erwartungswertparameter Control
sigma    = 1                                    # Standardabweichungsparameter
delta    = (mu_t - mu_c)/sigma                  # wahre, aber unbekannte, SES
n        = (n_t*n_c/(n_t+n_c))                  # Stichprobenumfangparameter
m        = n_t + n_c - 2                        # Freiheitsgradparameter
c_m      = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))  # Biaskorrekturfaktor
ds       = rep(NaN, sim)                        # SES-Schätzerarrayinitialisierung
for(i in 1:sim){                                # Simulationsiterationen
   y     = matrix(rep(NaN,n_t+n_c), ncol = 2)   # Datenarrayinitialisierung
   y[,1] = rnorm(n_t, mu_t, sigma)              # Datengeneration Treatmentgruppe
   y[,2] = rnorm(n_c, mu_c, sigma)              # Datengeneration Kontrollgruppe
   ybar  = apply(y,2,mean)                      # Gruppenmittelwerte
   s2    = apply(y,2,var)                       # Gruppenvarianzen
   s     = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m)  # gepoolte Standardabweichung
   ds[i] = (ybar[1]-ybar[2])/s}                 # SES-Schätzerrealisierung
```

Verteilung von $\sqrt{n}d$

10^5 SES-Schätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$



Theorem (Hedges' g im ESM)

Gegeben sei ein Einzelstudienmodell, d sei der verzerrte Schätzer von δ und es seien

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2}\Gamma((m-1)/2)}. \tag{20}$$

wie oben definiert. Dann ist

$$g := c_m d, \tag{21}$$

genannt *Hedges' g*, ein unverzerrter Schätzer von δ und es gilt

$$\mathbb{V}(g) < \mathbb{V}(d). \tag{22}$$

Beweis

Die Unverzerrtheit von g ergibt sich aus

$$\mathbb{E}(g) = \mathbb{E}(c_m d) = c_m \mathbb{E}(d) = c_m \frac{1}{c_m}\delta = \delta. \tag{23}$$

Die Varianzreduktion von g in Bezug auf d folgt aus der Tatsache, dass $c_m < 1$ für alle m und somit

$$\mathbb{V}(g) = \mathbb{V}(c_m d) = c_m^2 \mathbb{V}(d) < \mathbb{V}(d). \tag{24}$$

□

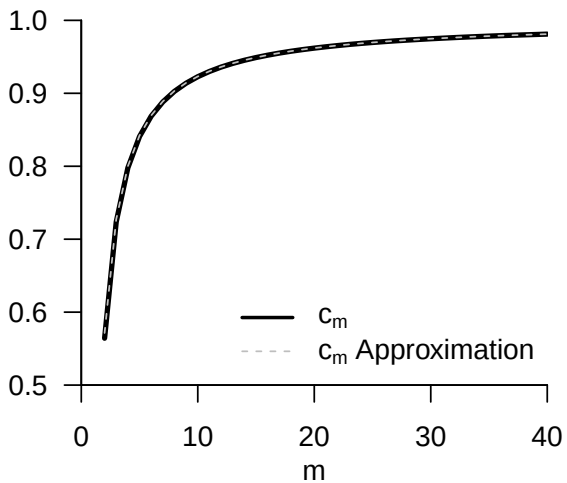
Effektstärkeschätzung

Bemerkungen

- Man beachte, dass der Korrekturfaktor c_m durch $n_t + n_c - 2$ vom Gesamtstichprobenumfang abhängig ist.
- Der Wert des Korrekturfaktors für einen gegebenen Wert von m kann sehr genau durch

$$c_m \approx 1 - \frac{3}{4m - 1} \tag{25}$$

angenähert werden, vgl. Hedges (1981), Borenstein (2009) und untenstehende Abbildung.

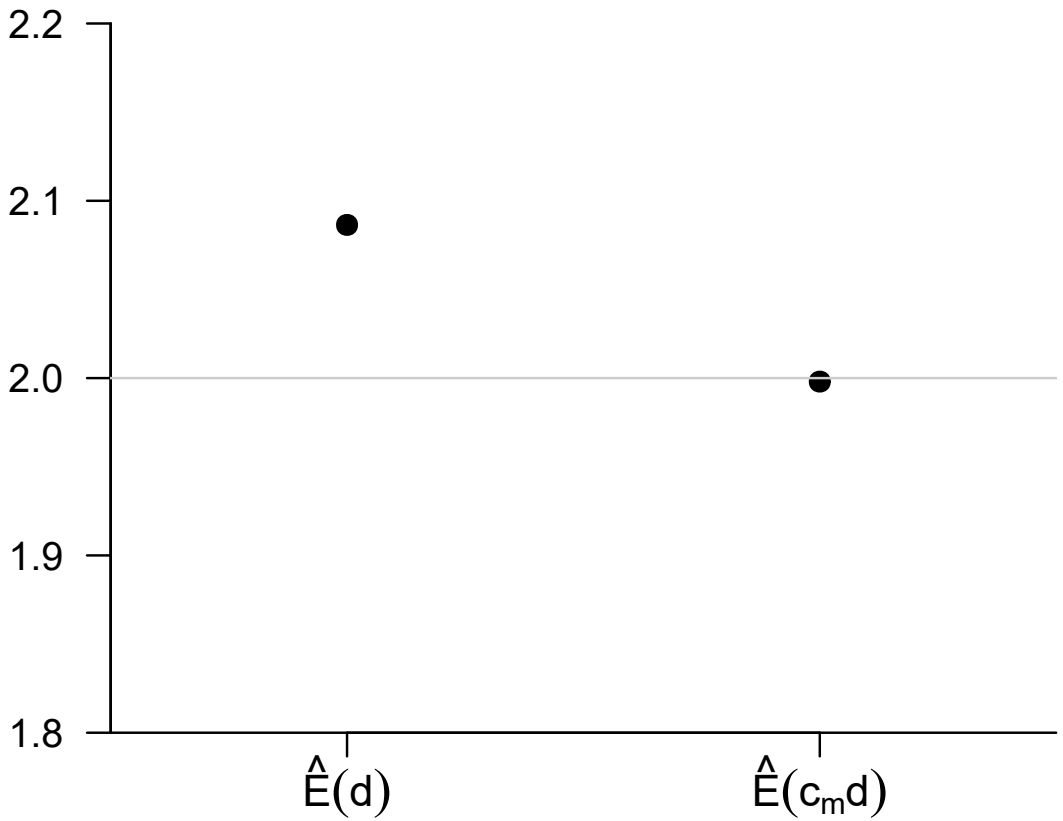


- Lin and Aloe (2021) zeigen in Appendix A1, dass $c_m \to 1$ für $m \to \infty$.
- Die Biaskorrektur ist am wichtigsten bei Gesamtgruppengrößen $n_t + n_c < 30$.
- Die Biaskorrektur ist am wichtigsten bei Treatment- und Kontrollgruppengrößen $n_t = n_c < 15$.

Effektstärkeschätzung

Geschätzte Erwartungswerte der verzerrten und unverzerrten SES-Schätzer im ESM

10^5 SES-Schätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$



Theorem (Asymptotische Normalverteilung von Hedges' g)

Gegeben sei ein Einzelstudienmodell und g sei der unverzerrte Schätzer von δ. Dann ist g bei einem konstanten Verhältnis von n_t und n_c für $n_t \to \infty$, $n_c \to \infty$ asymptotisch normalverteilt und es gilt speziell

$$g \overset{a}{\sim} N\left(\delta, \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)}\right). \tag{26}$$

Bemerkungen

- Wir erinnern daran, dass *Asymptotische Schätzereigenschaften* sich auf große Stichprobenumfänge beziehen.
- Das Theorem ist äquivalent dazu, dass die nichtzentrale-t-Verteilung für $\nu \to \infty$ die Normalverteilung annähert.
- Wir verzichten auf einen Beweis und verweisen wie Hedges and Olkin (1985) auf Johnson and Welch (1940).

Effektstärkeschätzung

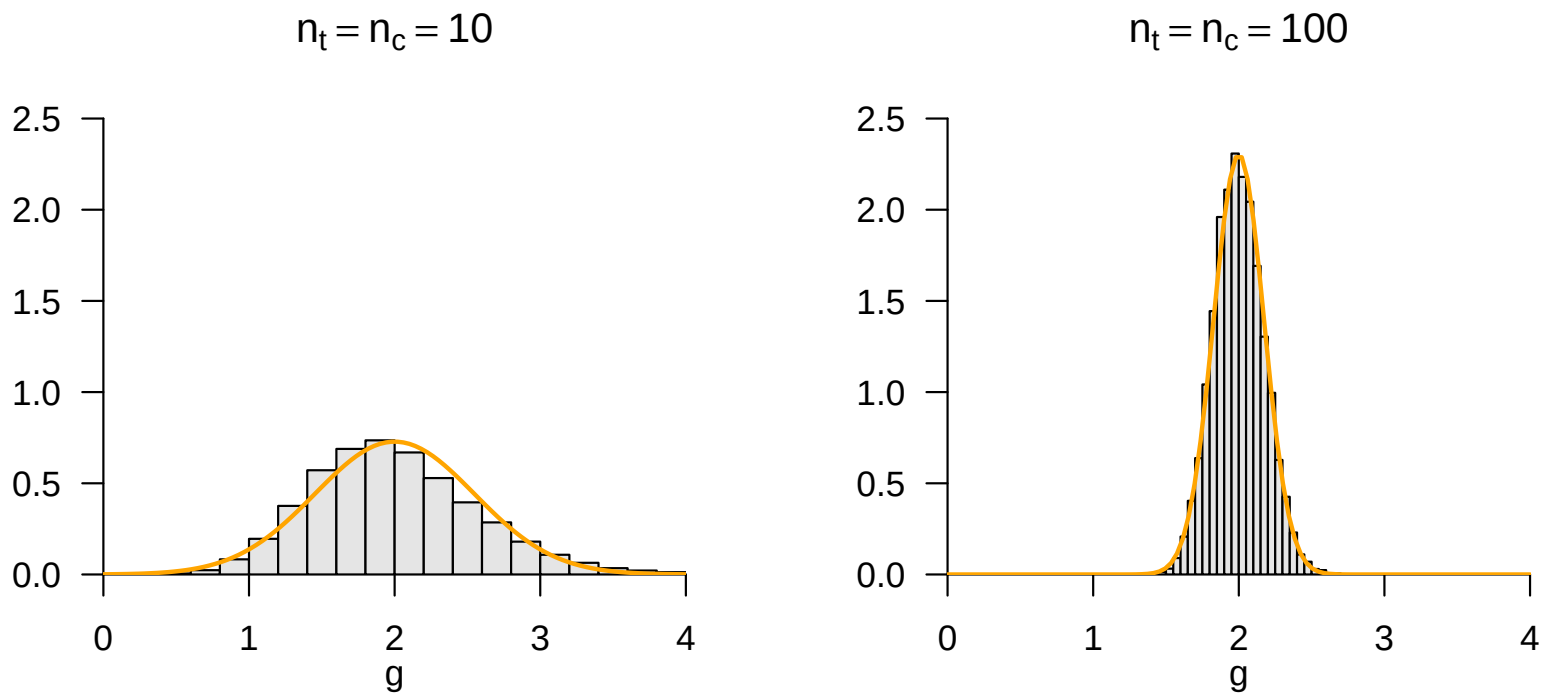
Asymptotische Normalverteilung von Hedges' g

10^4 SES-Schätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang

```
sim            = 1e4                                  # Anzahl Realisierungen
mu_t           = 2                                    # Erwartungswertparameter Treatment
mu_c           = 0                                    # Erwartungswertparameter Control
sigma          = 1                                    # Standardabweichungsparameter
delta          = (mu_t - mu_c)/sigma                  # wahre, aber unbekannte, SES
ng             = c(10,100)                            # Stichprobenumfänge
gs             = matrix(rep(NaN, 2*sim), ncol = 2)    # Hedges' g Arrayinitialisierung
vgs            = rep(NaN,2)                           # asymptotische Varianzarray
for (j in 1:2){                                       # Stichprobenumfangsiterationen
   n_t         = ng[j]                                # Treatmentgruppenumfang
   n_c         = ng[j]                                # Kontrollgruppenumfang
   n           = (n_t*n_c/(n_t+n_c))                  # Stichprobenumfangparameter
   m           = n_t + n_c - 2                        # Freiheitsgradparameter
   c_m         = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))  # Biaskorrekturfaktor
   vgs[j]      = (1/n_t)+(1/n_c)+(delta^2)/(2*(n_t+n_c))  # Asymptotische Varianz von Hedges' g
   for(i in 1:sim){                                   # Simulationsiterationen
      y        = matrix(rep(NaN,n_t+n_c), ncol = 2)   # Datenarrayinitialisierung
      y[,1]    = rnorm(n_t, mu_t, sigma)              # Datengeneration Treatmentgruppe
      y[,2]    = rnorm(n_c, mu_c, sigma)              # Datengeneration Kontrollgruppe
      ybar     = apply(y,2,mean)                      # Gruppenmittelwerte
      s2       = apply(y,2,var)                       # Gruppenvarianzen
      s        = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m)  # gepoolte Standardabweichung
      gs[i,j] = c_m*(ybar[1]-ybar[2])/s}}             # Hedges' g
```

Asymptotische Normalverteilung von Hedges' g

10^5 SES-Schätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang



Definition (Asymptotische Varianz und Varianzschätzer von Hedges' g)

Gegeben sei ein Einzelstudienmodell und für den

$$g \overset{a}{\sim} N\left(\delta, \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)}\right). \tag{27}$$

Dann wird der Varianzparameter der asymptotischen Verteilung von g als *asymptotische Varianz* von g bezeichnet und mit

$$\mathbb{V}_a(g) = \frac{1}{n_t} + \frac{1}{n_c} + \frac{\delta^2}{2(n_t + n_c)}. \tag{28}$$

bezeichnet. Ein häufig genutzter Schätzer für die asymptotische Varianz von Hedges' g ist

$$\hat{\mathbb{V}}_a(g) = \frac{1}{n_t} + \frac{1}{n_c} + \frac{g^2}{2(n_t + n_c)}. \tag{29}$$

- $\hat{\mathbb{V}}_a(g)$ wird häufig für die metaanalytische Modellbildung genutzt, vgl. Lin and Aloe (2021).
- Bei konstantem Wert von Hedges' g nimmt $\hat{\mathbb{V}}_a(g)$ offenbar bei steigenden Werten von n_t und n_c ab.

Theorem (Asymptotisches Konfidenzintervall für die SES)

Gegeben sei ein Einzelstudienmodell, g sei der unverzerrte Schätzer der standardisierten Effektstärke δ und für ein Konfidenzlevel $\gamma \in]0, 1[$ sei

$$z_\gamma := \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \tag{30}$$

wobei Φ^{-1} die inverse kumulative Verteilungsfunktion der Standardnormalverteilung bezeichne. Dann ist

$$\kappa := \left[g - \sqrt{\mathbb{V}_a(g)} z_\gamma, g + \sqrt{\mathbb{V}_a(g)} z_\gamma\right] \tag{31}$$

ein asymptotisches γ-Konfidenzintervall für die standardisierte Effektstärke δ.

Bemerkungen

- Das betrachtete Konfidenzintervall wurde von Hedges and Olkin (1985) (S. 86) vorgeschlagen.
- Wir nutzen hier γ für das Konfidenzlevel, da δ schon belegt ist.
- κ entspricht einem Konfidenzintervall für den Erwartungswert der Normalverteilung bei bekannter Varianz.
- In der Anwendung wird man $\mathbb{V}_a(g)$ durch seinen Schätzer $\hat{\mathbb{V}}_a(g)$ ersetzen.
- Asymptotische Konfidenzintervalle werden auch *Konfidenzintervalle von Wald-Typ* genannt.

Beweis

Wir müssen zeigen, dass

$$\mathbb{P}(\kappa \ni \delta) = \gamma. \tag{32}$$

Dazu halten wir zunächst fest, dass

$$g \overset{a}{\sim} N\left(\delta, \mathbb{V}_a(g)\right). \tag{33}$$

Mit dem Theorem zur Z-Transformation gilt dann

$$Z_g := \frac{g - \delta}{\sqrt{\mathbb{V}_a(g)}} \sim N(0, 1). \tag{34}$$

Weiterhin gilt per Definition von z_γ, dass

$$\mathbb{P}\left(-z_\gamma \leq Z_g \leq z_\gamma\right) = \gamma. \tag{35}$$

Beweis (fortgeführt)

Aus der Definition eines γ-Konfidenzintervalls folgt dann

$$\begin{aligned}
\gamma &= \mathbb{P}\left(-z_\gamma \leq Z_g \leq z_\gamma\right) \\
&= \mathbb{P}\left(-z_\gamma \leq \frac{g-\delta}{\sqrt{\mathbb{V}_a(g)}} \leq z_\gamma\right) \\
&= \mathbb{P}\left(-z_\gamma\sqrt{\mathbb{V}_a(g)} \leq g-\delta \leq z_\gamma\sqrt{\mathbb{V}_a(g)}\right) \\
&= \mathbb{P}\left(-g-z_\gamma\sqrt{\mathbb{V}_a(g)} \leq -\delta \leq -g+z_\gamma\sqrt{\mathbb{V}_a(g)}\right) \\
&= \mathbb{P}\left(g+z_\gamma\sqrt{\mathbb{V}_a(g)} \geq \delta \geq g-z_\gamma\sqrt{\mathbb{V}_a(g)}\right) \\
&= \mathbb{P}\left(g-z_\gamma\sqrt{\mathbb{V}_a(g)} \leq \delta \leq g+z_\gamma\sqrt{\mathbb{V}_a(g)}\right) \\
&= \mathbb{P}\left(\left[g-\sqrt{\mathbb{V}_a(g)}z_\gamma, g+\sqrt{\mathbb{V}_a(g)}z_\gamma\right] \ni \delta\right) \\
&= \mathbb{P}(\kappa \ni \delta)
\end{aligned} \tag{36}$$

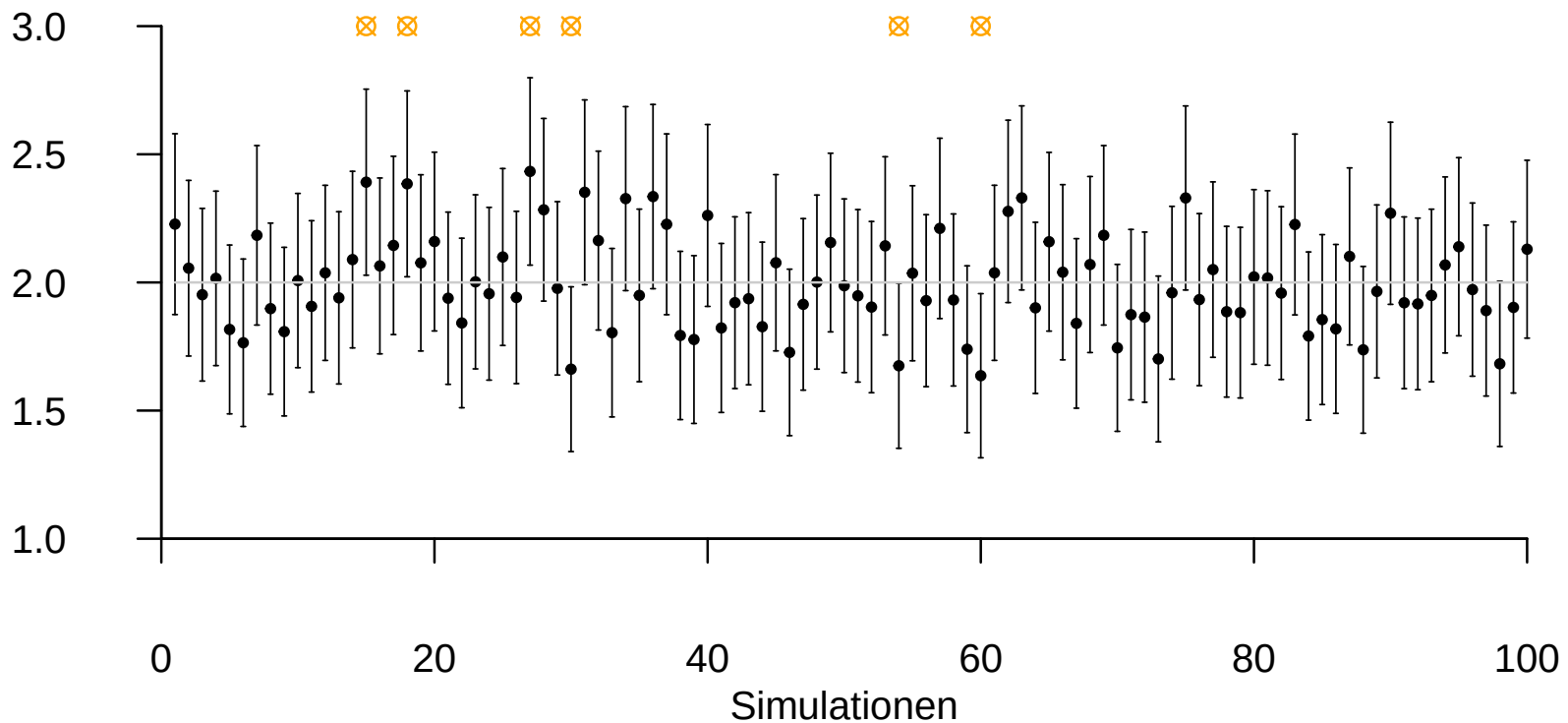
und damit ist alles gezeigt.

Effektstärkeschätzung

Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES

```
set.seed(0)                                                  # Zufallsgeneratorstate
sim             = 1e2                                        # Anzahl Realisierungen
mu_t            = 2                                          # Erwartungswertparameter Treatment
mu_c            = 0                                          # Erwartungswertparameter Control
sigma           = 1                                          # Standardabweichungsparameter
delta           = (mu_t - mu_c)/sigma                        # wahre, aber unbekannte, SES
gam             = 0.95                                       # Konfidenzlevel
zga             = qnorm((1+gam)/2)                           # Konfidenzlevelkonstante
n_t             = 100                                        # Treatmentgruppenumfang
n_c             = 100                                        # Kontrollgruppenumfang
n               = (n_t*n_c/(n_t+n_c))                        # Stichprobenumfangparameter
m               = n_t + n_c - 2                              # Freiheitsgradparameter
c_m             = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))    # Biaskorrekturfaktor
gs              = rep(NaN, sim)                              # Hedges' g Arrayinitialisierung
vs              = rep(NaN, sim)                              # asymptotische Varianzarray
kappa           = array(rep(NaN, sim*2), dim = c(2,sim))     # Konfidenzintervallarray
for(i in 1:sim){                                             # Simulationsiterationen
   y            = matrix(rep(NaN,n_t+n_c), ncol = 2)         # Datenarrayinitialisierung
   y[,1]        = rnorm(n_t, mu_t, sigma)                    # Datengeneration Treatmentgruppe
   y[,2]        = rnorm(n_c, mu_c, sigma)                    # Datengeneration Kontrollgruppe
   ybar         = apply(y,2,mean)                            # Gruppenmittelwerte
   s2           = apply(y,2,var)                             # Gruppenvarianzen
   s            = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m)    # gepoolte Standardabweichung
   gs[i]        = c_m*(ybar[1]-ybar[2])/s                    # Hedges' g
   vs[i]        = ((n_t+n_c)/(n_t*n_c))+(gs[i]^2)/(2*(n_t+n_c))  # Varianzschätzer
   kappa[1,i]   = gs[i] - sqrt(vs[i])*zga                    # untere Konfidenzintervallgrenze
   kappa[2,i]   = gs[i] + sqrt(vs[i])*zga                    # obere  Konfidenzintervallgrenze
}
```

Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES



Effektstärkeschätzung

Bestimmung von Hedges' g und 95%-KI für Abyar et al. (2018)

```
D           = read.csv("7-Daten/abyar-data.csv")                    # Rohdaten
y_t         = D$BSSI[D$Group == "ACT"]                              # Daten der ACT-Gruppe
y_c         = D$BSSI[D$Group == "WLC"]                              # Daten der Kontrollgruppe
n_t         = length(y_t)                                           # Umfang ACT-Gruppe
n_c         = length(y_c)                                           # Umfang Kontrollgruppe
m           = n_t + n_c - 2                                         # Freiheitsgradparameter
c_m         = gamma(m/2)/(sqrt(m/2)*gamma((m-1)/2))                 # Biaskorrekturfaktor
s           = sqrt(((n_t-1)*var(y_t) + (n_c-1)*var(y_c))/m)         # gepoolte SD
d           = (mean(y_t) - mean(y_c))/s                             # Cohen's d
g           = c_m*d                                                 # Hedges' g
v_g         = ((n_t+n_c)/(n_t*n_c))+(g^2)/(2*(n_t+n_c))             # Varianz aus Zhang-KI
gam         = 0.95                                                  # Konfidenzlevel
zga         = qnorm((1 + gam)/2)                                    # z-Konstante
ci.lb       = g - sqrt(v_g)*zga                                     # untere KI-Grenze
ci.ub       = g + sqrt(v_g)*zga                                     # obere  KI-Grenze
R           = data.frame(g = g, Vi = v_g, CIl = ci.lb, CIu = ci.ub) # Ergebnistabelle
```

```
     g   Vi  CIl   CIu
 -0.95 0.19 -1.8 -0.11
```

Effektstärkeschätzung

Bestimmung von Hedges' g und 95%-KI mit `metafor`

```
library(metafor)                                                        # R Paket
D           = read.csv("7-Daten/abyar-data.csv")                        # Rohdaten
E           = escalc(                                                   # Effektstärke
measure     = "SMD",                                                    # standardisierte Mittelwertsdifferenz
m1i         = mean(y_t),                                                # Mittelwert ACT-Gruppe
sd1i        = sd(y_t),                                                  # Standardabweichung ACT-Gruppe
n1i         = n_t,                                                      # Umfang ACT-Gruppe
m2i         = mean(y_c),                                                # Mittelwert Kontrollgruppe
sd2i        = sd(y_c),                                                  # Standardabweichung Kontrollgruppe
n2i         = n_c)                                                      # Umfang Kontrollgruppe
E$vi        = A$Vi                                                      # Zhang-Varianz
E$ci.lb     = E$yi - zga*sqrt(E$vi)                                     # untere KI-Grenze
E$ci.ub     = E$yi + zga*sqrt(E$vi)                                     # obere  KI-Grenze
R           = data.frame(g = E$yi, Vi = E$vi, CIl = E$ci.lb, CIu = E$ci.ub) # Ergebnistabelle
```

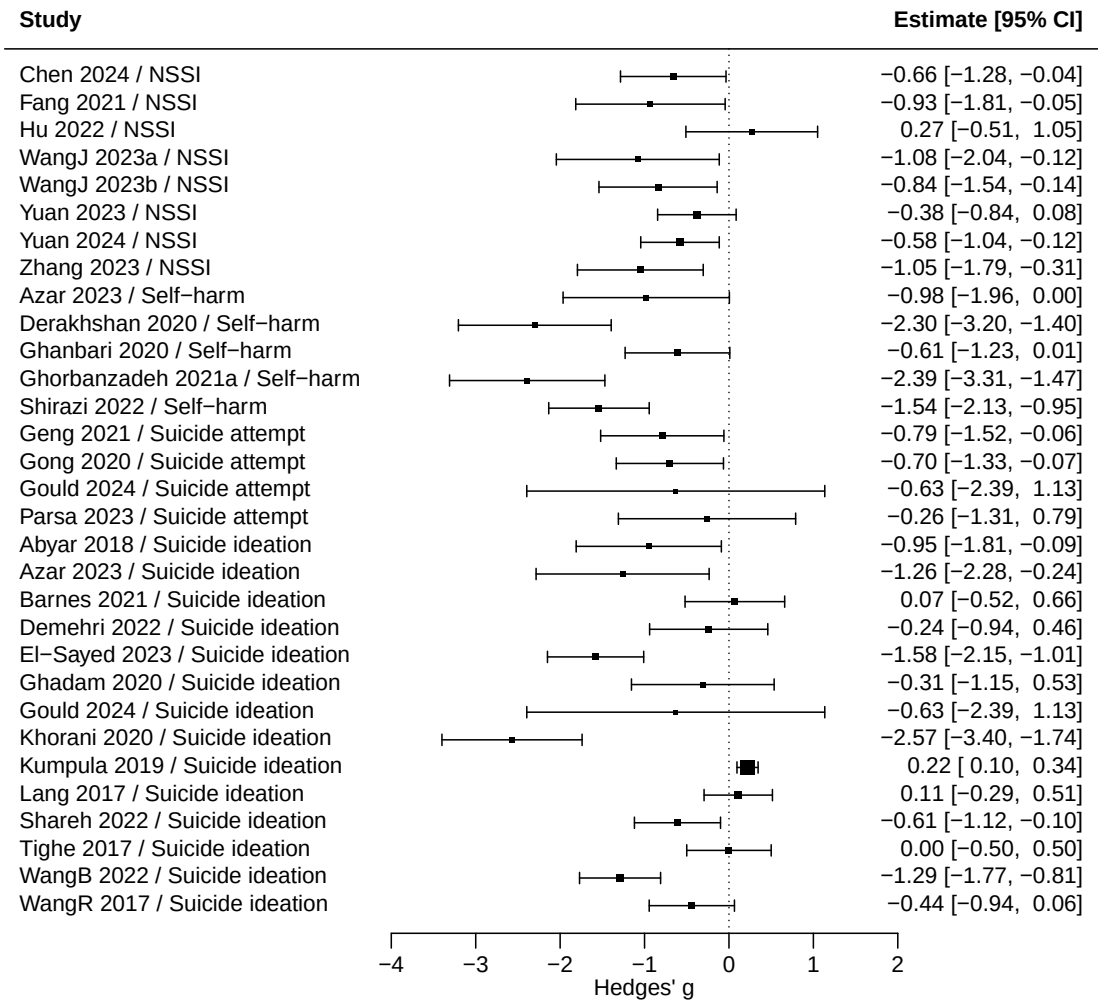
```
     g   Vi  CIl   CIu
 -0.95 0.19 -1.8 -0.09
```

Anwendungsbeispiel

Forest Plot der Einzelstudieneffekte

```
library(metafor)                                                  # R Paket
D               = read.csv(paste0("7-Daten/act-sitb-data.csv"))   # Datensatz
D               = D[D$Symptom != "STBI", ]                        # Figure 2-Datensatz
D               = D[order(D$Symptom, D$Study), ]                  # Sortierung
D$Label         = paste(D$Study, D$Symptom, sep = " / ")          # Studienlabel
pdf(                                                              # PDF-Device
file            = "7-Abbildungen/ivf-7-act-sitb-forest.pdf",      # Dateiname
width           = 8,                                              # Breite
height          = 7)                                              # Höhe
par(                                                              # Grafikparameter
family          = "sans",                                         # Schriftfamilie
mar             = c(3.2, 5, 1, 2),                                # Ränder
mgp             = c(1.8, .5, 0),                                  # Achsenbeschriftungsparameter
las             = 1,                                              # Achsenbeschriftungsstil
bty             = "l")                                            # Boxenstil
forest(                                                           # Forest Plot
x               = D$g,                                            # Hedges' g
vi              = D$Vi,                                           # Varianz
slab            = D$Label,                                        # Studienlabels
xlab            = "Hedges' g",                                    # x-Achsenbeschriftung
refline         = 0,                                              # Referenzlinie
alim            = c(-4, 2),                                       # x-Achsenlimits
at              = -4:2,                                           # x-Achsenbeschriftungspositionen
cex             = .9)                                             # Textgröße
invisible(dev.off())                                              # Device schließen
```

Anwendungsbeispiel



Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Definition (Fixed-Effects-Modell der Metaanalyse)

Für Studien $i = 1, \ldots, k$ seien y_i und σ_i^2 bekannte studienspezifische standardisierte Effektstärkeschätzer und ihre Varianzschätzer, respektive. Dann ist das *Fixed-Effects-Modell der Metaanalyse (FEM)* gegeben durch

$$y_i \sim N(\delta, \sigma_i^2), \tag{37}$$

wobei δ die wahre, aber unbekannte, standardisierte Effektstärke (SES) bezeichnet und die $y_1, \ldots, y_k$ als unabhängige Zufallsvariablen angenommen werden.

Bemerkungen

- y_i entspricht meist dem Wert von Hedges' g_i.
- Manchmal kann y_i auch dem Wert von Cohen's d_i entsprechen.
- $y_i \sim N(\delta, \sigma_i^2)$ approximiert die nichtzentrale-t-Verteilung von y_i bei großen Stichprobenumfängen.
- σ_i^2 entspricht meist dem Wert eines Schätzers der Varianz von Hedges' g_i.
- Manchmal kann σ_i^2 auch dem Wert eines Schätzers der Varianz von Cohen's d_i entsprechen.
- Das Fixed-Effects-Modell der Metaanalyse nimmt explizit an, dass die $\sigma_i^2, i = 1, \ldots, k$ bekannt sind.
- Es gibt verschiedenste Möglichkeiten, die Varianz von Hedges' g_i bzw. Cohen's d_i in Studie i zu schätzen.
- Wir betrachten nur den von Hedges and Olkin (1985) und in Viechtbauer (2010) implementierten Schätzer

$$\sigma_i^2 := \hat{\mathbb{V}}_a(g_i) = \frac{1}{n_t} + \frac{1}{n_c} + \frac{g_i^2}{2(n_t + n_c)}. \tag{38}$$

- Lin and Aloe (2021) diskutieren verschiedene Varianzschätzer für Hedges' g_i bzw. Cohen's d_i.
- Ziel der Fixed-Effects-Modellierung der Metaanalyse ist es, δ basierend auf den y_i und σ_i^2 zu schätzen.

Theorem (FEM-ML-SES-Schätzer)

Gegeben sei das Fixed-Effects-Modell der Metaanalyse mit wahrer, aber unbekannter, standardisierter Effektstärke δ. Dann ergibt sich der Maximum-Likelihood-Schätzer der standardisierten Effektstärke δ zu

$$\hat{\delta} := \frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i y_i \text{ mit den Wichtungsparametern } w_i := \frac{1}{\sigma_i^2}, i = 1, \ldots, k. \tag{39}$$

Bemerkungen

- $\hat{\delta}$ genügt der Intuition, dass Studien mit kleiner studienspezifischer Varianz des standardisierten Effektstärkeschätzers, z.B. hervorgerufen durch einen großen studienspezifischen Stichprobenumfang, mehr zu der übergeordneten Effektstärkeschätzung beitragen sollten als solche mit hoher studienspezifischer Varianz des standardisierten Effektstärkeschätzers.
- Die Wichtungsparameter werden als reziproke Werte der studienspezifischen Varianz des standardisierten Effektstärkeschätzers auch *studienspezifische Präzisionen* des standardisierten Effektstärkeschätzers genannt.
- Sind die studienspezifischen Varianzen σ_i^2 alle identisch zu σ^2, so ergibt sich mit $w := 1/\sigma^2$

$$\hat{\delta} = \frac{1}{\sum_{i=1}^{k} w} \sum_{i=1}^{k} w y_i = \frac{w}{w \sum_{i=1}^{k} 1} \sum_{i=1}^{k} y_i = \frac{1}{k} \sum_{i=1}^{k} y_i. \tag{40}$$

Beweis

Ohne zwischen y_i als Zufallsvariable oder Wert zu unterscheiden hat die Likelihood-Funktion des FEMs die Form

$$L : \mathbb{R} \to \mathbb{R}_{>0}, \delta \mapsto L(\delta) := \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \delta)^2\right). \tag{41}$$

Die Log-Likelihood-Funktion des FEMs hat damit die Form

$$\ell : \mathbb{R} \to \mathbb{R}, \delta \mapsto \ell(\delta) := -\frac{k}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{k} \ln \sigma_i^2 - \sum_{i=1}^{k} \frac{1}{2\sigma_i^2}(y_i - \delta)^2. \tag{42}$$

Die erste Ableitung der Log-Likelihood-Funktion des FEMs hat damit die Form

$$\frac{d}{d\delta}\ell(\delta) = -\sum_{i=1}^{k} \frac{1}{2\sigma_i^2}\frac{d}{d\delta}(y_i - \delta)^2 = \sum_{i=1}^{k} \frac{1}{\sigma_i^2}(y_i - \delta). \tag{43}$$

Es ergibt sich also die Maximum-Likelihood-Bedingung

$$\sum_{i=1}^{k} \frac{1}{\sigma_i^2}(y_i - \hat{\delta}) = 0. \tag{44}$$

Auflösen ergibt dann

$$\sum_{i=1}^{k} \frac{1}{\sigma_i^2} y_i - \sum_{i=1}^{k} \frac{1}{\sigma_i^2}\hat{\delta} = 0 \Leftrightarrow \hat{\delta} = \frac{\sum_{i=1}^{k} \frac{1}{\sigma_i^2} y_i}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}} \Leftrightarrow \hat{\delta} = \frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i y_i. \tag{45}$$

□

Theorem (Eigenschaften des FEM-ML-SES-Schätzers)

Gegeben sei das Fixed-Effects-Modell der Metaanalyse mit wahrer, aber unbekannter, standardisierter Effektstärke δ und es sei $\hat{\delta}$ der Maximum-Likelihood-Schätzer von δ. Dann gelten

(1) $\mathbb{E}(\hat{\delta}) = \delta$

(2) $\mathbb{V}(\hat{\delta}) = \frac{\sum_{i=1}^{k} w_i^2 \sigma_i^2}{(\sum_{i=1}^{k} w_i)^2}$

Bemerkungen

- Aussage (1) besagt insbesondere, dass $\hat{\delta}$ ein unverzerrter Schätzer von δ ist.

Beweis

(1) Mit den Eigenschaften des Erwartungswerts gilt

$$\mathbb{E}(\hat{\delta}) = \mathbb{E}\left(\frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i y_i\right) = \frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i \mathbb{E}(y_i) = \frac{\sum_{i=1}^{k} w_i \delta}{\sum_{i=1}^{k} w_i} = \frac{\delta \sum_{i=1}^{k} w_i}{\sum_{i=1}^{k} w_i} = \delta. \tag{46}$$

(2) Mit den Eigenschaften der Varianz bei unabhängigen Zufallsvariablen gilt

$$\mathbb{V}(\hat{\delta}) = \mathbb{V}\left(\frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i y_i\right) = \frac{1}{(\sum_{i=1}^{k} w_i)^2} \mathbb{V}\left(\sum_{i=1}^{k} w_i y_i\right) = \frac{\sum_{i=1}^{k} w_i^2 \mathbb{V}(y_i)}{(\sum_{i=1}^{k} w_i)^2} = \frac{\sum_{i=1}^{k} w_i^2 \sigma_i^2}{(\sum_{i=1}^{k} w_i)^2}. \tag{47}$$

Fixed-Effects-Modelle

Simulation mit $k := 10$, $\delta := 1$ und $\sigma_i^2 \sim U(1, 2)$

```
set.seed(1)                                          # reproduzierbare Ergebnisse
k               = 10                                 # Anzahl an Studien
delta           = 1                                  # wahre, aber unbekannte, Effektstärke
sigsqr_i        = rep(NaN,k)                         # wahre, bekannte, studienspezifische Varianzenarray
y_i             = rep(NaN,k)                         # Empirische Effektstärkenarray
for(i in 1:k){                                       # Studieniterationen
    sigsqr_i[i] = runif(1,1,2)                       # wahre, bekannte, studienspezifische Varianz
    y_i[i]      = rnorm(1, delta, sqrt(sigsqr_i[i]))}  # studienspezifischer Effektstärkeschätzer
w_i             = 1/sigsqr_i                         # Wichtungsparameter
delta_hat       = sum(w_i*y_i)/sum(w_i)              # Fixed-Effects-Modell-basierter Effektstärkeschätzer
```

Theorem (Q-Statistik)

Gegeben sei das Fixed-Effects-Modell der Metaanalyse und es sei

$$Q := \sum_{i=1}^{k} w_i (y_i - \hat{\delta})^2 \text{ mit } \hat{\delta} := \frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i y_i \text{ und } w_i := \frac{1}{\sigma_i^2}, i = 1, \ldots, k. \tag{48}$$

Dann ist Q asymptotisch χ^2-verteilt mit Freiheitsgradparameter $k - 1$

$$Q \overset{a}{\sim} \chi^2(k-1). \tag{49}$$

Bemerkungen

- Die Q-Statistik ist die anhand der studienspezifischen Präzisionen gewichtete Summe der quadrierten Abweichungen der studienspezifischen Effektstärkeschätzer vom studienübergreifenden Effektstärkeschätzer. Intuitiv sprechen hohe Werte von Q also gegen die Annahme einer festen Effektstärke über Studien.
- Unter der Annahme, dass die studienspezifischen Effektstärkeschätzer y_i im Fixed-Effects-Modell tatsächlich auf ein und dieselbe Effektstärke δ zurückgehen, ist Q asymptotisch χ^2-verteilt. Ansätze für Beweise dieser Tatsache finden sich in Cochran (1954) und Hedges (1982).
- Ist unter $Q \sim \chi^2(k-1)$ die Wahrscheinlichkeit für einen beobachteten Wert oder einen extremeren als den beobachteten Wert der Q-Statistik klein, so spricht dies intuitiv gegen die Annahme des Fixed-Effects-Modells (nach Schaeuffele et al. (2024): "A P value of the Q statistic below 0.05 indicates heterogeneity.") Für eine vertiefte Diskussion, siehe Kulinskaya et al. (2011b) und Kulinskaya et al. (2011a).

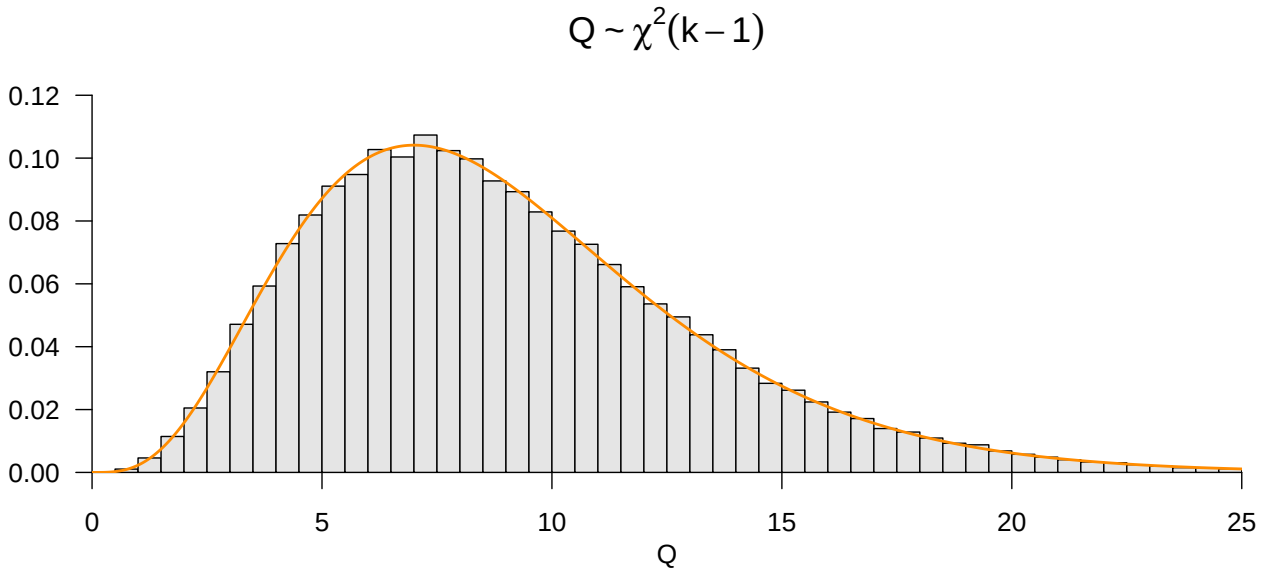
Verteilung der Q-Statistik

10^5 Q-Statistikrealisierungen für $k := 10$, $\delta := 1$ und $\sigma_i^2 \sim U(1, 2)$

```
set.seed(1)                                                 # reproduzierbare Ergebnisse
sim                 = 1e5                                   # Anzahl Realisierungen
k                   = 10                                    # Anzahl an Studien
delta               = 1                                     # wahre, aber unbekannte, Effektstärke
Q                   = rep(NaN,sim)                          # Q-Statistik Array
for(s in 1:sim){                                            # Simulationsiterationen
    sigsqr_i        = rep(NaN,k)                            # wahre, bekannte, studienspezifische Varianz
    y_i             = rep(NaN,k)                            # Empirische Effektstärkenarray
    for(i in 1:k){                                          # Studieniterationen
        sigsqr_i[i] = runif(1,1,2)                          # wahre, bekannte, studienspezifische Varianz
        y_i[i]      = rnorm(1, delta, sqrt(sigsqr_i[i]))}   # empirische Effektstärke
    w_i             = 1/sigsqr_i                            # Wichtungsparameter
    delta_hat       = sum(w_i*y_i)/sum(w_i )                # Fixed-Effects-Modell-basierter Effektstärkeschätzer
    Q[s]            = sum(w_i*(y_i - delta_hat)**2)         # Q-Statistik
}
```

Verteilung der Q-Statistik

10^5 Q-Statistikrealisierungen für $k := 10$, $\delta := 1$ und $\sigma_i^2 \sim U(1, 2)$



Definition (I^2-Statistik)

Gegeben sei das Fixed-Effects-Modell der Metaanalyse. Dann ist die I^2-Statistik definiert als

$$I^2 := \max\left\{0, \frac{Q-(k-1)}{Q}\right\}, \tag{50}$$

wobei k die Studienanzahl und Q die oben definierte Q-Statistik bezeichnen.

Bemerkungen

- Die I^2-Statistik wurde von Higgins and Thompson (2002) vorgeschlagen. Der Erwartungswert einer $\chi^2(k-1)$-verteilten Zufallsvariable ist $k-1$. I^2 misst also in standardisierter Form und unter Annahme des Fixed-Effects-Modells approximativ, inwieweit $Q > \mathbb{E}(Q)$ ist. (nach Schaeuffele et al. (2024): "I^2 ranges from 0 to 100%, with 25% representing low, 50% moderate and 75% high heterogeneity.") Generell, also für beliebige Effektstärkemaße außerhalb des Fixed-Effects-Modells, muss Q jedoch nicht $\chi^2(k-1)$ verteilt sein. Hoaglin (2016) gibt eine kritische Diskussion der Verwendung der Q- und I^2 Statistiken in der Metaanalyse.

Anwendungsbeispiel

Maximum-Likelihood-basierte standardisierte Effektstärkeschätzung im Fixed-Effects-Modell

```
D   = read.csv("7-Daten/act-sitb-data.csv")                                  # Einlesen des Datensatzes
S   = c("Suicide ideation","Suicide attempt","Self-harm","NSSI","STBI") # Symptomreihenfolge
gam = 0.95                                                                  # Konfidenzlevel
zga = qnorm((1 + gam)/2)                                                    # z-Konfidenzkonstante
R   = data.frame()                                                          # Ergebnistabelle
for(s in S){                                                                # Symptomiteration
    Ds            = D[D$Symptom == s, ]                                    # Symptomdatensatz
    k             = nrow(Ds)                                                # Anzahl an Studien
    y_i           = Ds$g                                                    # Empirische Effektstärken Array
    sigsqr_i      = Ds$Vi                                                   # Studienspezifische Varianzenarray
    w_i           = 1/sigsqr_i                                              # Wichtungsparameter
    delta_hat     = sum(w_i*y_i)/sum(w_i)                                   # FEM-ML-SES-Schätzer
    delta_hat_var = 1/sum(w_i)                                              # Asymptotische Varianz
    Q             = sum((w_i*(y_i-delta_hat)^2))                            # Q-Statistik
    I2            = max(0, (Q - (k-1))/Q)                                   # I^2-Statistik
    ci.lb         = delta_hat-sqrt(delta_hat_var)*zga                       # untere KI Grenze
    ci.ub         = delta_hat+sqrt(delta_hat_var)*zga                       # obere  KI Grenze
    R             = rbind(R,  data.frame(                                   # Ergebnisspeicherung
    Symptom       = s,                                                      # Symptom
    k             = k,                                                      # Anzahl an Studien
    estimate      = delta_hat,                                              # Effektstärkeschätzer
    ci.lb         = ci.lb,                                                  # untere KI-Grenze
    ci.ub         = ci.ub,                                                  # obere  KI-Grenze
    Q             = Q,                                                      # Q-Statistik
    I2            = 100*I2))}                                               # I^2-Statistik
```

```
          Symptom  k estimate ci.lb  ci.ub     Q I2
 Suicide ideation 14   -0.069 -0.17  0.031 124.9 90
  Suicide attempt  4   -0.655 -1.08 -0.232   0.7  0
       Self-harm  5   -1.420 -1.76 -1.084  15.4 74
             NSSI  8   -0.592 -0.82 -0.367   9.0 22
             STBI 35   -0.384 -0.47 -0.301 364.8 91
```

Anwendungsbeispiel

Maximum-Likelihood-basierte standardisierte Effektstärkeschätzung im Fixed-Effects-Modell mit `metafor`

```
library(metafor)                                                          # R Paket
D = read.csv("7-Daten/act-sitb-data.csv")                                 # Einlesen des Datensatzes
S = c("Suicide ideation","Suicide attempt","Self-harm","NSSI","STBI")     # Symptomreihenfolge
R = data.frame()                                                          # Ergebnistabelle
for(s in S){                                                              # Symptomiteration
    Ds       = D[D$Symptom == s, ]                                        # Symptomdatensatz
    res      = rma(yi = Ds$g, vi = Ds$Vi, method = "FE")                  # Fixed-Effects-Modell
    R        = rbind(R, data.frame(                                       # Ergebnisspeicherung
    Symptom  = s,                                                         # Symptom
    k        = res$k,                                                     # Anzahl an Studien
    estimate = as.numeric(res$b),                                         # Effektstärkeschätzer
    ci.lb    = res$ci.lb,                                                 # untere KI-Grenze
    ci.ub    = res$ci.ub,                                                 # obere  KI-Grenze
    Q        = res$QE,                                                    # Q-Statistik
    I2       = res$I2))}                                                  # I^2-Statistik
```

Symptom	k	estimate	ci.lb	ci.ub	Q	I2
Suicide ideation	14	-0.069	-0.17	0.031	124.9	90
Suicide attempt	4	-0.655	-1.08	-0.232	0.7	0
Self-harm	5	-1.420	-1.76	-1.084	15.4	74
NSSI	8	-0.592	-0.82	-0.367	9.0	22
STBI	35	-0.384	-0.47	-0.301	364.8	91

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Definition (Random-Effects-Modell der Metaanalyse)

Für Studien $i = 1, \ldots, k$ seien y_i und σ_i^2 bekannte studienspezifische standardisierte Effektstärkeschätzer bzw. ihre Varianzschätzer. Dann ist das *Random-Effects-Modell der Metaanalyse* gegeben durch

$$y_i|\delta_i \sim N\left(\delta_i, \sigma_i^2\right) \text{ und } \delta_i \sim N\left(\delta, \tau^2\right), \tag{51}$$

wobei

- δ die *wahre, aber unbekannte, standardisierte Effektstärke*,
- τ^2 die *wahre, aber unbekannte, studienübergreifende standardisierte Effektstärkenvarianz* und
- δ_i die Zufallsvariable der *latenten standardisierten Effektstärke der i-ten Studie*

bezeichnen, wobei die δ_i als unabhängig und die y_i bedingt auf die δ_i als unabhängig angenommen werden.

Bemerkungen

- Für die y_i und σ_i^2 ergeben sich die gleichen Bemerkungen wie für das Fixed-Effects-Modell.
- Die studienübergreifende Varianz der standardisierten Effektstärken wird auch als *Heterogenität* bezeichnet.

Bemerkungen (fortgesetzt)

- Für $i = 1, \ldots, k$ impliziert das Random-Effects-Modell der Metaanalyse die bedingten WDFen

$$p(y_i|\delta_i) = N(y_i; \delta_i, \sigma_i^2) \tag{52}$$

und die marginale WDF

$$p(\delta_i) = N(\delta_i; \delta, \tau^2) \tag{53}$$

- Das Random-Effects-Modell der Metaanalyse entspricht also der gemeinsamen WDF

$$p(y_i, \delta_i) = p(y_i|\delta_i)p(\delta_i) = N(y_i; \delta_i, \sigma_i^2)N(\delta_i; \delta, \tau^2). \tag{54}$$

- Mit den Eigenschaften der multivariaten Normalverteilung impliziert dies die marginalen WDFen

$$p(y_i) = N(y_i; \delta, \sigma_i^2 + \tau^2) \tag{55}$$

und damit die Marginalverteilungen

$$y_i \sim N(\delta, \sigma_i^2 + \tau^2) \tag{56}$$

der studienspezifischen standardisierten Effektstärkeschätzer.

Bemerkungen (fortgesetzt)

- Weiterhin gilt mit dem Theorem zur linear-affinen Transformation normalverteilter Zufallsvariablen

$$\delta_i \sim N(\delta, \tau^2) \Leftrightarrow \delta_i = \delta + b_i \text{ mit } b_i \sim N(0, \tau^2) \tag{57}$$

  und

$$y_i|\delta_i \sim N(\delta_i, \sigma_i^2) \Leftrightarrow y_i = \delta_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma_i^2) \tag{58}$$

- Zusammengefasst ergibt sich also die strukturelle Form

$$y_i = \delta + b_i + \varepsilon_i \text{ mit } b_i \sim N(0, \tau^2) \text{ und } \varepsilon_i \sim N(0, \sigma_i^2) \tag{59}$$

- Generativ betrachtet nimmt das Random-Effects-Modell also eine wahre, aber unbekannte, standardisierte Effektstärke δ an. Diese wahre, aber unbekannte, standardisierte Effektstärke wird in der i-ten Studie unter dem Einfluss eines additiven normalverteilten Fehlers b_i mit unbekannter Varianz τ^2 als δ_i indirekt beobachtet.
- Direkt beobachtet wird dann mit y_i eben dieses δ_i unter dem Einfluss eines additiven nullzentrierten normalverteilten Fehlers mit bekannter Varianz σ_i^2.

Bemerkungen (fortgesetzt)

Random-Effects-Modell der Metaanalyse in Kompaktdarstellung eines LMMs

Wie oben gesehen gilt für das REM der Metaanalyse in struktureller Form

$$y_i = \delta + b_i + \varepsilon_i \text{ mit } b_i \sim N(0, \tau^2) \text{ und } \varepsilon_i \sim N(0, \sigma_i^2) \text{ mit } i = 1, ..., k. \tag{60}$$

Mit der Definition von

$$X := 1_k, \quad \beta := \delta, \quad Z := I_k, \quad b := (b_1, ..., b_k)^T, \quad \varepsilon := (\varepsilon_1, ..., \varepsilon_k)^T, \tag{61}$$

$$\Sigma_b := \tau^2 I_k \text{ und } \Sigma_\varepsilon := \text{diag}\left(\sigma_1^2, ..., \sigma_k^2\right) \tag{62}$$

ergibt sich in Bezug auf die Definition des LMMs in Kompaktdarstellung aus Vorlesung 4

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_k, \Sigma_b) \text{ und } \varepsilon \sim N(0_k, \Sigma_\varepsilon) \tag{63}$$

$$\Leftrightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \delta + \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{pmatrix} \tag{64}$$

Die Random Effects b_i modellieren hier also studienspezifische Abweichungen vom Populationseffekt δ und die Fehlerterme ε_i studienspezifische Schätzfehler mit bekannter Varianz σ_i^2.

Weil mit $A := 1_k$ gilt, dass $X = ZA = I_k 1_k$, ist hier auch die Populationsparameterdarstellung anwendbar.

Bemerkungen (fortgesetzt)

Random-Effects-Modell der Metaanalyse in LMM Populationsparameterdarstellung

Mit der Definition von

$$u := (\delta_1, ..., \delta_k)^T, A := 1_k, \beta := \delta \text{ und } b \sim N(0, \tau^2 I_k), \tag{65}$$

also

$$u := A\beta + b \Leftrightarrow \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_k \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \delta + \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix} \tag{66}$$

ergibt sich unter Definition von

$$Z := I_k, \Sigma_b := \tau^2 I_k \text{ und } \Sigma_\varepsilon := \text{diag}\left(\sigma_1^2, ..., \sigma_k^2\right). \tag{67}$$

die Populationsparameterdarstellung des LMMs

$$y = Zu + \varepsilon \text{ mit } u \sim N\left(1_k\delta, \Sigma_b\right), \quad \varepsilon \sim N\left(0_k, \Sigma_\varepsilon\right). \tag{68}$$

Wie unter dem Random-Effects-Modell der Metaanalyse konzeptualisiert sind hier also die latenten Studieneffektstärken $u_i = \delta_i$ Realisierungen aus einer Population von Studieneffekten mit Erwartungswert δ und Varianz τ^2 und die beobachteten Studieneffektstärken sind eben diese unter der Addition von studienspezifischen Schätzfehlern ε_i mit Erwartungswert 0 und bekannter Varianz σ_i^2.

Parameterschätzung im Random-Effects-Modell

- Zur Schätzung der Heterogenität τ^2 existiert im Bereich der Metaanalyse eine Vielzahl von Verfahren. Beispiele sind Momentenmethoden wie DerSimonian-Laird, Hedges oder Hunter-Schmidt, Maximum-Likelihood-Schätzung, Restricted Maximum-Likelihood-Schätzung, Empirical-Bayes-Schätzung, Sidik-Jonkman-Schätzung und Paule-Mandel-Schätzung (vgl. Viechtbauer (2005), Veroniki et al. (2016), Langan et al. (2019)).
- Basierend auf einem Heterogenitätsschätzer $\hat{\tau}^2$ wird analog zum FEM-ML-SES-Schätzer ein inverse-Varianz-gewichteter REM-SES-Schätzer definiert als

$$\hat{\delta}_{\hat{\tau}^2} := \frac{1}{\sum_{i=1}^{k} \hat{\tilde{w}}_i} \sum_{i=1}^{k} \hat{\tilde{w}}_i y_i \text{ mit } \hat{\tilde{w}}_i := \frac{1}{\hat{\tau}^2 + \sigma_i^2} \text{ für } i = 1, ..., k. \tag{69}$$

- In der heutigen Anwendung wird häufig ReML bevorzugt.
- Das Random-Effects-Modell der Metaanalyse wird dabei als Linear Mixed Model betrachtet und mit den in Vorlesung 4 diskutierten Verfahren zur Parameterschätzung in LMMs geschätzt.

Theorem (Asymptotische Verteilung eines REM-SES-Schätzers)

Gegeben sei das Random-Effects-Modell der Metaanalyse mit bekannten studienspezifischen Varianzen σ_i^2 für $i = 1, \ldots, k$, studienübergreifender standardisierter Effektstärkenvarianz (Heterogenität) τ^2 und es sei $\hat{\delta}$ ein Schätzer der wahren, aber unbekannten, standardisierten Effektstärke δ. Schließlich seien

$$\tilde{w}_i := \frac{1}{\tau^2 + \sigma_i^2} \text{ für } i = 1, \ldots, k. \tag{70}$$

Dann ist $\hat{\delta}$ asymptotisch normalverteilt und es gilt speziell

$$\hat{\delta} \overset{a}{\sim} N\left(\delta, \left(\sum_{i=1}^{k} \tilde{w}_i\right)^{-1}\right). \tag{71}$$

Bemerkungen

- Das Theorem ist ein Spezialfall der asymptotischen Verteilung von Fixed-Effects-Parameterschätzern in LMMs.
- Rencher and Christensen (2012) 17.5.1 zitieren Fuller and Battese (1974) für einen Beweis.

Definition (Asymptotische Varianz für einen REM-SES-Schätzer)

Gegeben sei das Random-Effects-Modell der Metaanalyse mit bekannten studienspezifischen Varianzen σ_i^2 für $i = 1, ..., k$, studienübergreifender standardisierter Effektstärkenvarianz (Heterogenität) τ^2 und es sei $\hat{\delta}$ ein Schätzer der wahren, aber unbekannten, standardisierten Effektstärke δ. Dann wird der Varianzparameter der asymptotischen Verteilung von $\hat{\delta}$ als *asymptotische Varianz* von $\hat{\delta}$ bezeichnet und mit

$$\mathbb{V}_a(\hat{\delta}) = \left(\sum_{i=1}^{k} \tilde{w}_i \right)^{-1} \text{ mit } \tilde{w}_i := \frac{1}{\tau^2 + \sigma_i^2} \text{ für } i = 1, ..., k. \tag{72}$$

bezeichnet. Ein häufig genutzter Schätzer für die asymptotische Varianz von $\hat{\delta}$ ist

$$\hat{\mathbb{V}}_a(\hat{\delta}) = \left(\sum_{i=1}^{k} \hat{\tilde{w}}_i \right)^{-1} \text{ mit } \hat{\tilde{w}}_i := \frac{1}{\hat{\tau}^2 + \sigma_i^2} \text{ für } i = 1, ..., k. \tag{73}$$

für einen Schätzer $\hat{\tau}^2$ der studienübergreifenden standardisierten Effektstärkevarianz τ^2.

Bemerkungen

- Die Definitionen sind analog zu denen zur asymptotischen Varianz von Hedges' g im Einzelstudienmodell.

Theorem (Asymptotisches Konfidenzintervall für die SES)

Gegeben sei das Random-Effects-Modell der Metaanalyse, $\hat{\delta}$ sei ein REM-basierter Schätzer der standardisierten Effektstärke δ, $\mathbb{V}_a(\hat{\delta})$ sei seine Varianz und für ein Konfidenzlevel $\gamma \in]0, 1[$ sei

$$z_\gamma := \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \tag{74}$$

wobei Φ^{-1} die inverse kumulative Verteilungsfunktion der Standardnormalverteilung bezeichne. Dann ist

$$\kappa := \left[\hat{\delta} - \sqrt{\mathbb{V}_a(\hat{\delta})}z_\gamma, \hat{\delta} + \sqrt{\mathbb{V}_a(\hat{\delta})}z_\gamma\right] \tag{75}$$

ein asymptotisches γ-Konfidenzintervall für die standardisierte Effektstärke δ.

Bemerkungen

- Wir nutzen hier γ für das Konfidenzlevel, da δ schon belegt ist.
- κ entspricht einem Konfidenzintervall für den Erwartungswert der Normalverteilung bei bekannter Varianz.
- In der Anwendung wird man $\mathbb{V}_a(\hat{\delta})$ durch seinen Schätzer $\hat{\mathbb{V}}_a(\hat{\delta})$ ersetzen.

Beweis

Wir müssen zeigen, dass

$$\mathbb{P}(\kappa \ni \delta) = \gamma. \tag{76}$$

Dazu halten wir zunächst fest, dass

$$\hat{\delta} \overset{a}{\sim} N\left(\delta, \mathbb{V}_a(\hat{\delta})\right). \tag{77}$$

Mit dem Theorem zur Z-Transformation gilt dann

$$Z_{\hat{\delta}} := \frac{\hat{\delta} - \delta}{\sqrt{\mathbb{V}_a(\hat{\delta})}} \sim N(0, 1). \tag{78}$$

Weiterhin gilt per Definition von z_γ, dass

$$\mathbb{P}\left(-z_\gamma \leq Z_{\hat{\delta}} \leq z_\gamma\right) = \gamma. \tag{79}$$

Beweis (fortgeführt)

Aus der Definition eines γ-Konfidenzintervalls folgt dann

$$
\begin{aligned}
\gamma &= \mathbb{P}\left(-z_\gamma \leq Z_{\hat{\delta}} \leq z_\gamma\right) \\
&= \mathbb{P}\left(-z_\gamma \leq \frac{\hat{\delta} - \delta}{\sqrt{\mathbb{V}_a(\hat{\delta})}} \leq z_\gamma\right) \\
&= \mathbb{P}\left(-z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq \hat{\delta} - \delta \leq z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}\right) \\
&= \mathbb{P}\left(-\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq -\delta \leq -\hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}\right) \\
&= \mathbb{P}\left(\hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \geq \delta \geq \hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}\right) \\
&= \mathbb{P}\left(\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq \delta \leq \hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}\right) \\
&= \mathbb{P}\left(\left[\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}, \hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}\right] \ni \delta\right) \\
&= \mathbb{P}(\kappa \ni \delta)
\end{aligned}
\tag{80}
$$

und damit ist alles gezeigt.

Anwendungsbeispiel

ReML-basierte standardisierte Effektstärkeschätzung im Random-Effects-Modell

```
library(metafor)                                                    # R Paket
D = read.csv("7-Daten/act-sitb-data.csv")                           # Einlesen des Datensatzes
S = c("Suicide ideation","Suicide attempt","Self-harm","NSSI","STBI")  # Symptomreihenfolge
R = data.frame()                                                    # Ergebnistabelle
for(s in S){                                                        # Symptomiteration
    Ds       = D[D$Symptom == s, ]                                  # Symptomdatensatz
    res      = rma(yi = Ds$g, vi = Ds$Vi, method = "REML")          # ReML
    R        = rbind(R, data.frame(                                 # Ergebnistabelle
    Symptom  = s,                                                   # Symptom
    k        = res$k,                                               # Anzahl an Studien
    estimate = as.numeric(res$b),                                   # Standardisierte-Effektstärke-Schätzer
    ci.lb    = res$ci.lb,                                           # untere Konfidenzintervallgrenze
    ci.ub    = res$ci.ub,                                           # obere  Konfidenzintervallgrenze
    Q        = res$QE,                                              # Q-Statistik
    I2       = res$I2,                                              # I^2-Statistik
    tau2     = res$tau2))}                                          # ReML tau^2 Schätzer
```

Symptom	k	estimate	ci.lb	ci.ub	Q	I2	tau2
Suicide ideation	14	-0.64	-1.06	-0.22	124.9	89.7	0.5134
Suicide attempt	4	-0.65	-1.08	-0.23	0.7	0.0	0.0000
Self-harm	5	-1.53	-2.22	-0.84	15.4	74.3	0.4459
NSSI	8	-0.60	-0.83	-0.36	9.0	4.1	0.0047
STBI	35	-0.99	-1.30	-0.67	364.8	90.5	0.7341

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Selektionsbias-Modelle

Motivation

- In der wissenschaftlichen Literatur dominieren statistisch signifikante Ergebnisse ($p < 0.05$).
- Studien mit nicht-signifikanten Ergebnissen werden seltener veröffentlicht.
- Dadurch entsteht ein Publikationsbias in Effektstärkenschätzungen.

Fragestellungen

- Wie lässt sich dieser Publikationsbias quantitativ schätzen?
- Wie kann dieser Publikationsbias in der Metaanalyse berücksichtigt werden?

Graphische Verfahren (z.B. Light and Pillemer (1984), Egger et al. (1997), Sterne and Egger (2001))

- Visualisierung von Publikationsbias durch Funnel Plots.
- Quantifizierung von Publikationsbias durch den Funnel Plot-basierten Egger-Test.

Modellbasierte Verfahren (z.B. Hedges (1992), Copas (1999), Mavridis et al. (2014))

- Publikationsselektion wird im Metaanalysemodell explizit berücksichtigt.
- Der Selektionsmechanismus wird als Modellparameter geschätzt.
- Die Effektstärke wird unter Berücksichtigung der Selektion geschätzt.

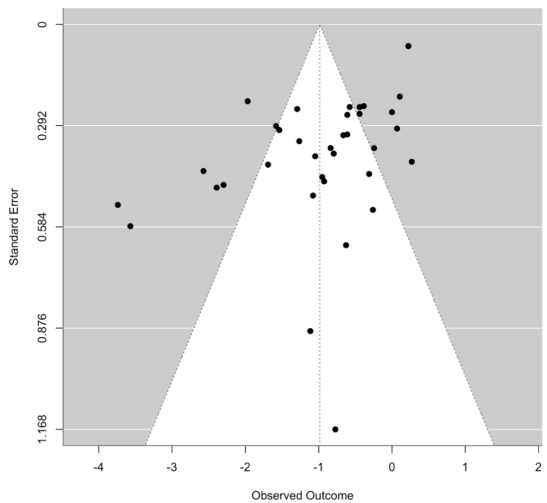
$\Rightarrow$ Integrierte Einführung in graphische und modellbasierte Verfahren am Beispiel von Hedges (1992)

Selektionsbias-Modelle

Anwendungsbeispiel nach Zhang et al. (2026)

Additionally, we assessed possible publication bias visually and quantitatively using funnel plots and Egger's test (...) For publication bias, Egger's tests indicated no publication bias for suicide ideation, suicide attempt, self-harm, NSSI (ps $>$ 0.05), supporting the robustness of these findings. However, evidence of publication bias was detected for overall SITBs (p $<$ 0.05), suggesting that this result should be interpreted with caution. The funnel plots of suicide ideation and overall SITBs are shown in online supplement E (online suppl. Fig. S14 and S15)

Figure S15. The funnel plot for overall SITBs at post-intervention



Selektionsbias-Modelle

Grundidee von Funnel Plots

- Darstellung studienspezifischer Effektstärkeschätzungen gegen ein Maß ihrer Unsicherheit/Studiengröße.
- Kleine Studien haben eine höhere Effektstärkerschätzervarianz (Standardfehler) als große Studien.
- Ohne systematische Selektion sollten daher kleine Studien stärker streuen als große Studien.
- Gleichzeitig sollten kleine und große Studien im Mittel ähnliche Effektstärkenschätzungen liefern.
- Insgesamt sollte sich eine trichterförmige Verteilung ergeben, die symmetrisch um die mittlere Effektstärke liegt.
- Asymmetrien im Funnel Plot können *small-study effects* anzeigen, Selektionsbias ist eine mögliche Erklärung.

Funnel Plot Varianten

- Es gibt nicht den Funnel Plot, sondern verschiedene Varianten mit unterschiedlichen Achsen.
- Die Wahl der Achsen hängt von genutztem Effektmaß und Fragestellung ab (vgl. Sterne and Egger (2001)).

Selektionsbias-Modelle

Achsen in Funnel Plots

Outcome Variable	Effektstärkemaß (x-Achse)	Unsicherheitsmaß (y-Achse)
Kontinuierlich	Mittelwertsdifferenz	Standardfehler, Varianz, Stichprobengröße
Kontinuierlich	Hedge's g	Standardfehler, Varianz, Stichprobengröße
Binär	Log Odds Ratio	Standardfehler, Varianz, Stichprobengröße
Binär	Log Risk Ratio	Standardfehler, Varianz, Stichprobengröße

- Die x-Achse zeigt das Effektstärkemaß auf der Skala, auf der das Metaanalysemodell spezifiziert ist.
- Die y-Achse zeigt typischerweise den Standardfehler, häufig wird sie invertiert dargestellt.
- Für Ratio-Maße wie Odds Ratios und Risk Ratios wird typischerweise die Log-Skala verwendet.
- Alternativ werden Präzision, Varianz oder Stichprobengröße auf der y-Achse verwendet.
- Wir betrachten im Folgenden konkret

x-Achse	$y_i := g_i$	als Effektstärkeschätzer von Studie i
y-Achse	$\hat{\sigma}_i := \sqrt{\hat{\mathbb{V}}_a(g_i)}$	als Effektstärkeschätzerunsicherheitsschätzer Studie i

Selektionsbias-Modelle

`metafor` Funnel Plots

Typische `metafor` Funnel Plots visualisieren die Effektstärkeschätzer und ihre Varianzschätzer in der (y_i, σ)-Ebene

Weiterhin zeigen die Funnel Plots die von σ-abhängigen Grenzen des Intervalls $\hat{\delta} \pm 1.96\sqrt{\hat{\tau}^2 + \hat{\sigma}^2}$

Oft sieht man auch Varianten mit $\hat{\tau}^2 := 0$, auch wenn dies nicht der tatsächliche Schätzwert war.

Die typischen `metafor` Funnel Plots appellieren also an die folgende visuelle Intuition:

Ohne P-wert-basierte Selektion

- sollten die Punkte (y_i, σ_i) symmetrisch um $\hat{\delta}$ verteilt sein
- sollten im Mittel 95/100 Punkte in $\hat{\delta} \pm 1.96\sqrt{\hat{\tau}^2 + \sigma^2}$ liegen.

Liegen also mehr Studiendatenpunkte auf einer Seite von $\hat{\delta}$ und/oder viele von ihnen außerhalb von $\hat{\delta} \pm 1.96\sqrt{\hat{\tau}^2 + \sigma^2}$, so spricht dies gegen die Abwesenheit von P-Wert-basierten Selektionsmechanismen oder anderen Prozessen, die die Verteilung der y_i verzerren.

Eleganter wäre insgesamt sicherlich eine Visualisierung der (y_i, σ_i^2) vor dem Hintergrund einer gemeinsamen WDF.

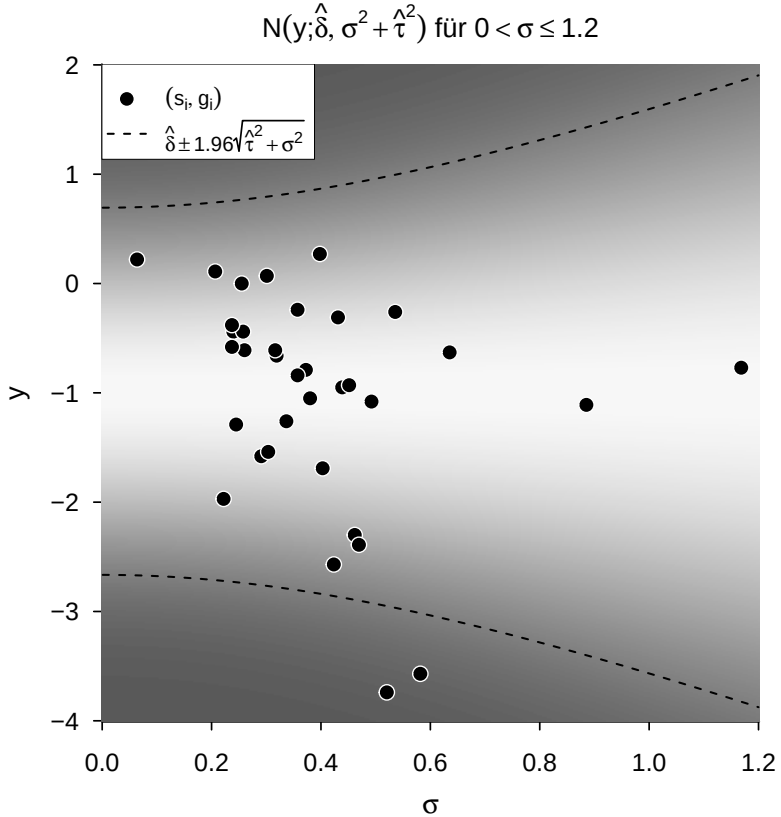
Selektionsbias-Modelle

Funnel Plot im `metafor` Stil der Daten aus Zhang et al. (2026) (“Random-Effects-Funnel-Plot”)

```r
library(metafor)                                                    # R Paket für Metaanalyse
library(latex2exp)                                                  # R Paket für LaTeX-Formeln in Plots
D           = read.csv("7-Daten/act-sitb-data.csv")                 # Einlesen des Datensatzes
D           = D[D$Symptom == "STBI", ]                              # Auswahl der Symptomreihe
R           = rma(yi = D$g, vi = D$Vi, method = "REML")             # ReML-basierte Metaanalyse
hat_delta   = as.numeric(R$b)                                       # \hat{\delta}
tau2        = R$tau2                                                # \hat{\tau}^2
res         = 401                                                   # Auflösung
y           = seq(-4.0, 2.0, length.out = res)                      # y-Achse Werte
sigma       = seq(1e-3, 1.2, length.out = res)                      # x-Achse Werte
z           = array(rep(NaN, res^2), dim = c(res,res))              # z-Werte Array
for(i in 1:length(sigma)){                                          # x-Achsen Iteration
    for(j in 1:length(y)){                                          # y-Achsen Iteration
        z[i,j] = dnorm((y[j]-hat_delta)/sqrt(tau2+sigma[i]^2))}}     # N(y; hat_delta, tau2 + sigma^2)
ub          = hat_delta + 1.96 * sqrt(tau2 + sigma^2)               # obere Intervallgrenze
lb          = hat_delta - 1.96 * sqrt(tau2 + sigma^2)               # untere Intervallgrenze
```

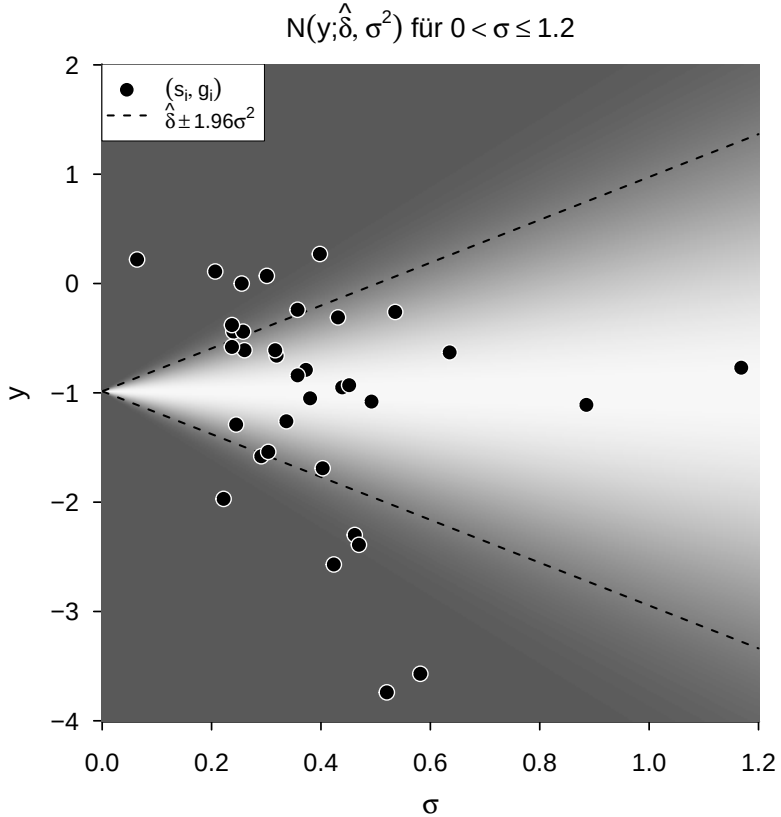
Selektionsbias-Modelle

Funnel Plot im `metafor` Stil der Daten aus Zhang et al. (2026) (“Random-Effects-Funnel-Plot”)



Selektionsbias-Modelle

Funnel Plot im `metafor` Stil der Daten aus Zhang et al. (2026) mit $\hat{\tau}^2 := 0$ ("Fixed-Effects-Funnel-Plot")



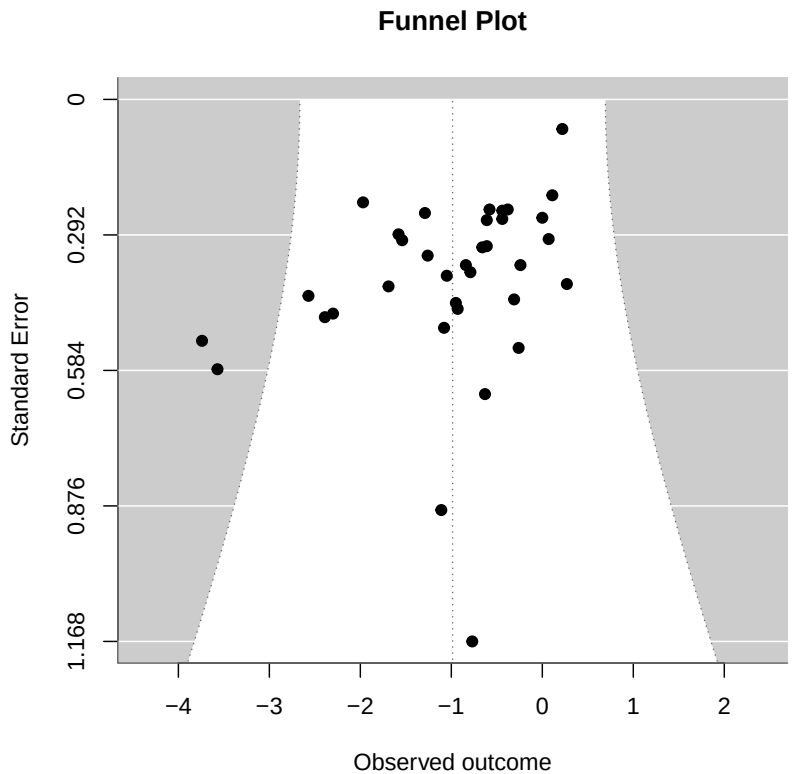
Selektionsbias-Modelle

Funnel Plot mit `metafor` für die Daten aus Zhang et al. (2026) (“Random-Effects-Funnel-Plot”)

```r
library(metafor)                                                                  # R Paket
pdf(                                                                              # PDF Ausgabe
file        = "./7-Abbildungen/ivf-7-zhang-funnel-plot-metafor-random-effects.pdf", # PDF Dateiname
width       = 6,                                                                  # PDF Breite
height      = 6)                                                                  # PDF Höhe
D           = read.csv("7-Daten/act-sitb-data.csv")                               # Einlesen des Datensatzes
D           = D[D$Symptom == "STBI", ]                                            # Auswahl der Symptomreihe
R           = rma(yi = D$g, vi = D$Vi, method = "REML")                           # ReML-basierte Metaanalyse
funnel(                                                                           # metafor Funktion
R,                                                                                # Analyseobjekt
xlab        = "Observed outcome",                                                 # x-Achse
ylab        = "Standard Error",                                                   # y-Achse
main        = "Funnel Plot",                                                      # Titel
refline     = as.numeric(R$b),                                                    # Intervallgrenzen
addtau2     = TRUE)                                                               # "Random Effects Funnel"
dev.off()                                                                         # PDF Finalisierung
```

Selektionsbias-Modelle

Funnel Plot mit `metafor` für die Daten aus Zhang et al. (2026) (“Random-Effects-Funnel-Plot”)



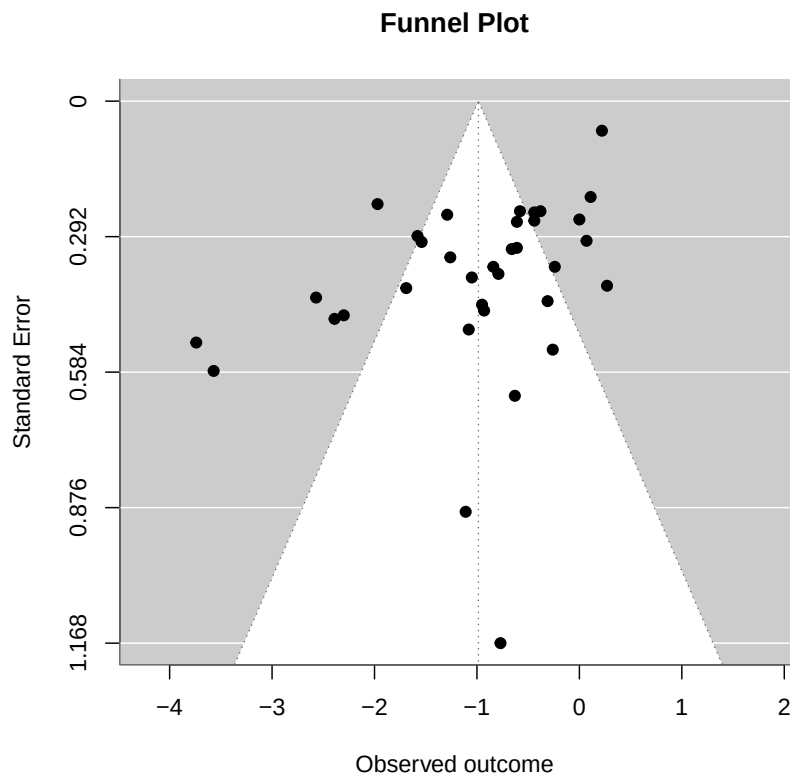
Selektionsbias-Modelle

Funnel Plot mit `metafor` für die Daten aus Zhang et al. (2026) für $\hat{\tau}^2 := 0$

```
library(metafor)                                                                  # R Paket
pdf(                                                                              # PDF Ausgabe
file        = "./7-Abbildungen/ivf-7-zhang-funnel-plot-metafor-fixed-effects.pdf", # PDF Dateiname
width       = 6,                                                                  # PDF Breite
height      = 6)                                                                  # PDF Höhe
D           = read.csv("7-Daten/act-sitb-data.csv")                               # Einlesen des Datensatzes
D           = D[D$Symptom == "STBI", ]                                            # Auswahl der Symptomreihe
R           = rma(yi = D$g, vi = D$Vi, method = "REML")                           # ReML-basierte Metaanalyse
funnel(                                                                           # metafor Funktion
R,                                                                                # Analyseobjekt
xlab        = "Observed outcome",                                                 # x-Achse
ylab        = "Standard Error",                                                   # y-Achse
main        = "Funnel Plot",                                                      # Titel
refline     = as.numeric(R$b))                                                    # Intervallgrenzen
dev.off()                                                                         # PDF Finalisierung
```

Selektionsbias-Modelle

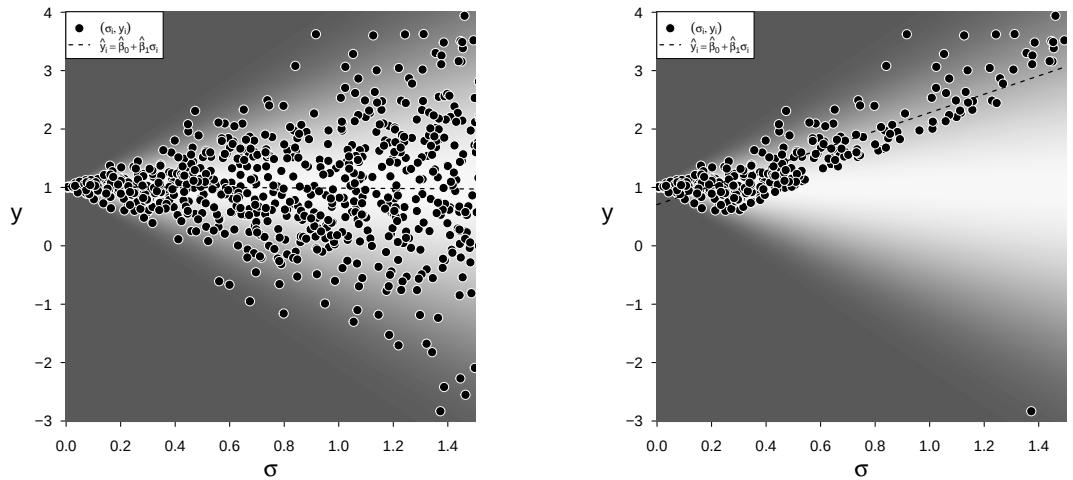
Funnel Plot mit `metafor` für die Daten aus Zhang et al. (2026) für $\hat{\tau}^2 := 0$ ("Fixed-Effects-Funnel-Plot")



Selektionsbias-Modelle

Visuelle Intuition zum Egger-Test

- Simulation von $\sigma_i \sim U(0, 2)$ und $y_i|\sigma_i \sim N(\delta, \sigma_i^2)$ mit $\delta = 1, i = 1, ..., k$
- Die linke Abbildung zeigt $k = 1.000$ simulierten Studiendatenpunkte (σ_i, y_i)
- Die rechte Abbildung zeigt die simulierten Studienpunkte mit P-Wert$_i = 2\Phi\left(-|\frac{y_i}{\sigma_i}|\right) \leq 0.05$



- Bei P-Wert-Selektion werden Studien mit großem σ_i nur veröffentlicht, wenn sie ein großes y_i haben.
- Dies führt zu einer positiven Kovariation zwischen Standardfehler und Effektstärkeschätzer.
- Der Egger-Test testet diese positive Kovariation formal im Sinne einer einfachen linearen Regression

Definition (Egger-Test-Modell in Effektstärken-Standardfehler-Form)

Vor dem Hintergrund der Definition des Fixed-Effects-Modells der Metaanalyse hat das Egger-Test-Modell in Effektstärken-Standardfehler-Form die strukturelle Form

$$y_i = \delta + \beta_1 \sigma_i + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma_i^2) \text{ unabhängig für } i = 1, ..., k, \tag{81}$$

die Datenverteilungsform

$$y_i \sim N(\delta + \beta_1 \sigma_i, \sigma_i^2) \text{ unabhängig für } i = 1, ..., k, \tag{82}$$

und die Designmatrixform

$$y = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} 1 & \sigma_1 \\ 1 & \sigma_2 \\ \vdots & \vdots \\ 1 & \sigma_k \end{pmatrix} \begin{pmatrix} \delta \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad \text{mit } \varepsilon \sim N\left(0_k, \text{diag}(\sigma_1^2, ..., \sigma_k^2)\right). \tag{83}$$

Bemerkung

- Das Modell ist als Erweiterung des Fixed-Effects-Modell um einen Zusammenhang von y_i und σ_i zu verstehen
- Im Fixed-Effects-Modell der Metaanalyse gilt $y_i \sim N(\delta, \sigma_i^2)$, im obigen Modell $y_i \sim N(\delta + \beta_1 \sigma_i, \sigma_i^2)$
- Ohne Selektion gilt $\beta_1 = 0$, d.h. die Effektstärkeschätzer y_i hängen nicht linear von den Standardfehlern σ_i ab.
- Für $\beta_1 = 0$ ist das obige Modell also genau das Fixed-Effects-Modell der Metaanalyse.
- Im Gegensatz zum ALM gilt bezüglich der Fehlerterme $\mathbb{C}(\varepsilon) = \text{diag}(\sigma_1^2, ..., \sigma_k^2) \neq \sigma^2 I_k$.
- Die analytische Verteilungstheorie des ALM ist also in obigem Modell nicht direkt anwendbar.

Definition (Egger-Test-Modell in Standard-Normal-Deviant-Form)

Vor dem Hintergrund der Definition des Fixed-Effects-Modells der Metaanalyse hat das Egger-Test-Modell in Standard-Normal-Deviant-Form die strukturelle Form

$$\frac{y_i}{\sigma_i} = \beta_0 + \frac{\delta}{\sigma_i} + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0,1) \text{ unabhängig für } i = 1, \ldots, k, \tag{84}$$

die Datenverteilungsform

$$\frac{y_i}{\sigma_i} \sim N\left(\beta_0 + \frac{\delta}{\sigma_i}, 1\right) \text{ unabhängig für } i = 1, \ldots, k, \tag{85}$$

und die Designmatrixform

$$y = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} \frac{y_1}{\sigma_1} \\ \frac{y_2}{\sigma_2} \\ \vdots \\ \frac{y_k}{\sigma_k} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{\sigma_1} \\ 1 & \frac{1}{\sigma_2} \\ \vdots & \vdots \\ 1 & \frac{1}{\sigma_k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \delta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad \text{mit } \varepsilon \sim N\left(0_k, I_k\right). \tag{86}$$

Bemerkung

- Offenbar gilt hier $\mathbb{C}(\varepsilon) = I_k$, d.h., die analytische Verteilungstheorie des ALM ist direkt anwendbar.

Theorem (Äquivalenz von Steigungs- und Interzeptparametern in den Egger-Test-Modellen)

Gegeben sei die Designmatrixform des Egger-Test-Modells in Effektstärken-Standardfehler-Form

$$y = X\beta + \varepsilon \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} 1 & \sigma_1 \\ 1 & \sigma_2 \\ \vdots & \vdots \\ 1 & \sigma_k \end{pmatrix} \begin{pmatrix} \delta \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad \text{mit } \varepsilon \sim N\left(0_k, \text{diag}(\sigma_1^2, ..., \sigma_k^2)\right). \tag{87}$$

Weiterhin seien

$$D := \text{diag}(\sigma_1^2, ..., \sigma_k^2) \text{ mit } \sigma_i > 0 \text{ für } i = 1, ..., k \tag{88}$$

und

$$W := \text{diag}(\sigma_1, ..., \sigma_k)^{-1} = D^{-1/2}. \tag{89}$$

Dann gilt, dass die Designmatrixform der Standard-Normal-Deviant-Form des Egger-Test-Modells gegeben ist durch

$$Wy = WX\tilde{\beta} + W\varepsilon \text{ mit } W\varepsilon \sim N(0_k, I_k) \text{ und } \tilde{\beta} = \begin{pmatrix} \beta_1 \\ \delta \end{pmatrix}. \tag{90}$$

Bemerkungen

- Die Standard-Normal-Deviant-Form ist also die durch W-transformierte Effektstärken-Standardfehler-Form.
- Der Interzept-Parameter der SND-Form ist identisch mit dem Steigungsparameter der ESF-Form.
- In der Standard-Normal-Deviant-Form ist die analytische Verteilungstheorie des ALMs direkt anwendbar.

Beweis

Es sei

$$y = X\beta + \varepsilon \quad \text{mit } \varepsilon \sim N(0_k, D) \tag{91}$$

die Designmatrixform des Egger-Test-Modells in Effektstärken-Standardfehler-Form und es sei

$$Wy = W(X\beta + \varepsilon) = WX\beta + W\varepsilon \tag{92}$$

seine durch W transformierte Form. Dann gilt

$$Wy = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_k} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{pmatrix} \frac{y_1}{\sigma_1} \\ \frac{y_2}{\sigma_2} \\ \vdots \\ \frac{y_k}{\sigma_k} \end{pmatrix}. \tag{93}$$

Mit der Definition von X und β gilt weiterhin

$$WX\beta = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_k} \end{pmatrix} \begin{pmatrix} 1 & \sigma_1 \\ 1 & \sigma_2 \\ \vdots & \vdots \\ 1 & \sigma_k \end{pmatrix} \begin{pmatrix} \delta \\ \beta_1 \end{pmatrix} \tag{94}$$

Beweis (fortgeführt)

und damit

$$
\begin{aligned}
W X \beta &= \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_k} \end{pmatrix} \begin{pmatrix} 1 & \sigma_1 \\ 1 & \sigma_2 \\ \vdots & \vdots \\ 1 & \sigma_k \end{pmatrix} \begin{pmatrix} \delta \\ \beta_1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{\delta}{\sigma_1} + \beta_1 \\ \frac{\delta}{\sigma_2} + \beta_1 \\ \vdots \\ \frac{\delta}{\sigma_k} + \beta_1 \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 + \frac{\delta}{\sigma_1} \\ \beta_1 + \frac{\delta}{\sigma_2} \\ \vdots \\ \beta_1 + \frac{\delta}{\sigma_k} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{1}{\sigma_1} \\ 1 & \frac{1}{\sigma_2} \\ \vdots & \vdots \\ 1 & \frac{1}{\sigma_k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \delta \end{pmatrix}.
\end{aligned}
\tag{95}
$$

Also gilt im Egger-Test-Modell in Standard-Normal-Deviant-Form $\beta_0 = \beta_1$ und der Parameter δ ist in beiden Modellen identisch.

Beweis (fortgeführt)

Nach dem Theorem zur linear-affinen Transformation normalverteilter Zufallsvariablen gilt schließlich

$$W\varepsilon \sim N\left(W0_k, WDW^T\right) \Leftrightarrow W\varepsilon \sim N\left(0_k, I_k\right) \tag{96}$$

denn

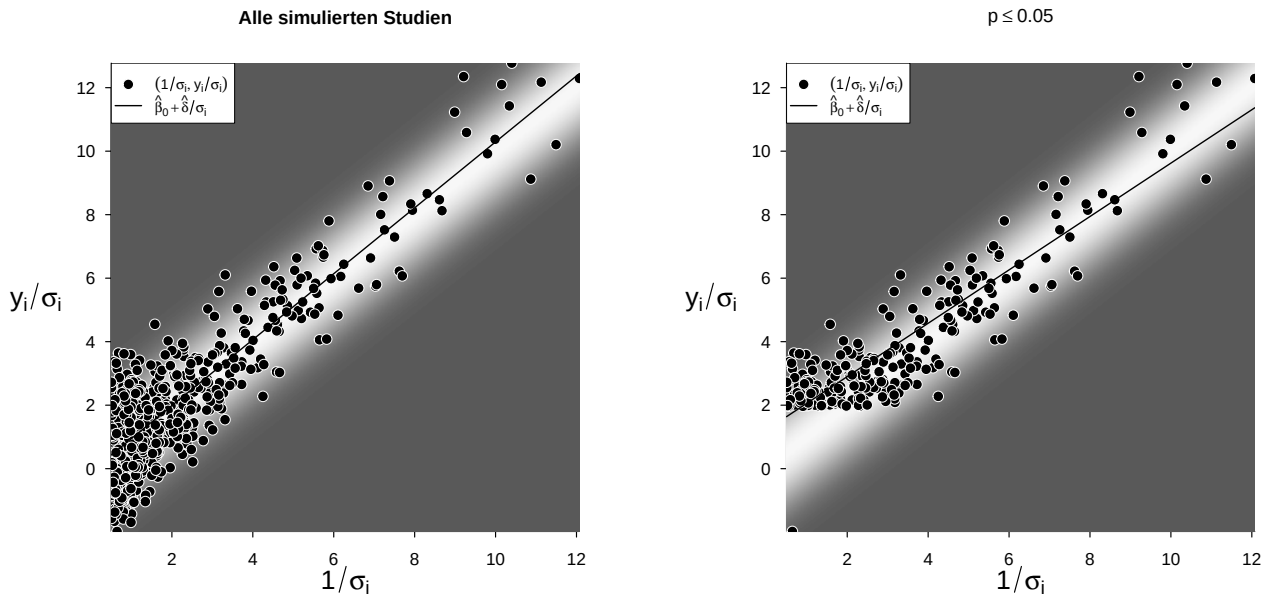
$$\begin{aligned} N(0_k, WDW^T) &= N\left(0_k, \text{diag}\left(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_k}\right) \text{diag}\left(\sigma_1^2, ..., \sigma_k^2\right) \text{diag}\left(\frac{1}{\sigma_1}, ..., \frac{1}{\sigma_k}\right)\right) \\ &= N(0_k, I_k). \end{aligned} \tag{97}$$

Die transformierten Fehler sind also sphärisch multivariat normalverteilt und die ALM-Verteilungstheorie ist auf die Standard-Normal-Deviant-Form direkt anwendbar.

□

Visuelle Intuition zum Egger-Test in Standard-Normal-Deviant-Form

- Simulation von $\sigma_i \sim U(0.08, 2)$ und $y_i|\sigma_i \sim N(\delta, \sigma_i^2)$ mit $\delta = 1, i = 1, ..., k$
- Transformation in Standard-Normal-Deviant-Form mit $x_i := 1/\sigma_i$ und $z_i := y_i/\sigma_i$
- Dann gilt $z_i|x_i \sim N(\delta x_i, 1)$ und der Egger-Test prüft den Interzeptparameter der Regression $z_i = \beta_0 + \delta x_i + \varepsilon_i$



Definition (Egger-Test)

Gegeben sei das Egger-Test-Modell in Standard-Normal-Deviant-Form

$$\frac{y_i}{\sigma_i} = \beta_0 + \frac{\delta}{\sigma_i} + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0,1) \text{ unabhängig für } i = 1, \ldots, k. \tag{98}$$

Dann ist der *Egger-Test* definiert als der Hypothesentest der einfachen Nullhypothese

$$H_0 : \beta_0 = 0. \tag{99}$$

Mit dem Kleinste-Quadrate-Schätzer $\hat{\beta}_0$ des Interzepts, der Designmatrix X des Egger-Test-Modells in Standard-Normal-Deviant-Form und der bekannten Fehlervarianz $\mathbb{V}(\varepsilon_i) = 1$ ist die Egger-Teststatistik gegeben durch

$$Z_{\mathsf{E}} := \frac{\hat{\beta}_0}{\sqrt{\mathbb{V}(\hat{\beta}_0)}} = \frac{\hat{\beta}_0}{\sqrt{\left((X^T X)^{-1}\right)_{11}}}. \tag{100}$$

Bemerkungen

- Der Test wurde von Egger et al. (1997) vorgeschlagen.
- Modifikationen für den Random-Effects-Kontext werden u.a. bei Lin and Chu (2018) diskutiert.
- Die Egger-Teststatistik entspricht der bekannten T-Statistik als Verhältnis von Schätzer und Standardfehler.
- Im Egger-Test-Modell ist allerdings $\sigma^2 = 1$ als bekannt vorausgesetzt.
- In klassischer Nomenklatur handelt es sich also um einen Z-Test.

Theorem (Egger-Test)

Gegeben sei das Egger-Test-Modell in Standard-Normal-Deviant-Form und die Definition des Egger-Tests und $\hat{\beta}_0, \hat{\delta}$ seien die Kleinste-Quadrate-Schätzer der Parameter β_0, δ. Dann gelten

(1) Interzeptparameterschätzer

$$\hat{\beta}_0 = \frac{1}{k} \sum_{i=1}^{k} \left(\frac{y_i - \hat{\delta}}{\sigma_i} \right) \tag{101}$$

(2) Verteilung der Egger-Teststatistik bei Zutreffen der Nullhypothese $H_0 : \beta_0 = 0$

$$Z_{\mathsf{E}} \sim N(0, 1). \tag{102}$$

Bemerkungen

- Der Interzeptparameter ergibt sich also als der Mittelwert der standardisierten Residuen der SND-Regression.
- Unter der Nullhypothese ist die Egger-Teststatistik standardnormalverteilt.

Beweis

Wir setzen für diesen Beweis

$$z_i := \frac{y_i}{\sigma_i} \quad \text{und} \quad x_i := \frac{1}{\sigma_i}. \tag{103}$$

Dann hat das Egger-Test-Modell in Standard-Normal-Deviant-Form die Form eines einfachen linearen Regressionsmodells

$$z_i = \beta_0 + \delta x_i + \varepsilon_i. \tag{104}$$

Nach dem Theorem zur Ausgleichsgerade gilt für den Kleinste-Quadrate-Schätzer

$$\hat{\beta}_0 = \bar{z} - \hat{\delta}\bar{x}, \tag{105}$$

wobei $\bar{z} := k^{-1} \sum_{i=1}^{k} z_i$ und $\bar{x} := k^{-1} \sum_{i=1}^{k} x_i$. Damit folgt

$$\hat{\beta}_0 = \bar{z} - \hat{\delta}\bar{x} = \frac{1}{k}\sum_{i=1}^{k} z_i - \hat{\delta}\frac{1}{k}\sum_{i=1}^{k} x_i = \frac{1}{k}\sum_{i=1}^{k}\left(z_i - \hat{\delta}x_i\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{y_i}{\sigma_i} - \hat{\delta}\frac{1}{\sigma_i}\right) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{y_i - \hat{\delta}}{\sigma_i}\right). \tag{106}$$

Beweis (fortgeführt)

Für die Verteilungsform schreiben wir das Egger-Test-Modell in Standard-Normal-Deviant-Form als

$$z = X \begin{pmatrix} \beta_0 \\ \delta \end{pmatrix} + \varepsilon \quad \text{mit } \varepsilon \sim N(0_k, I_k). \tag{107}$$

Nach dem Theorem zur frequentistischen Verteilung des Betaparameterschätzers gilt dann

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\delta} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \delta \end{pmatrix}, (X^T X)^{-1} \right). \tag{108}$$

Insbesondere gilt also bei Zutreffen der Nullhypothese $H_0 : \beta_0 = 0$ die Verteilung

$$\hat{\beta}_0 \sim N\left(0, \left((X^T X)^{-1}\right)_{11}\right). \tag{109}$$

Mit dem Theorem zur linearen-affinen Transformation eines normalverteilten Zufallsvektors folgt dann

$$Z_{\mathsf{E}} = \frac{1}{\sqrt{\left((X^T X)^{-1}\right)_{11}}} \hat{\beta}_0 \sim N\left(0, \sqrt{\left((X^T X)^{-1}\right)_{11}} \left((X^T X)^{-1}\right)_{11} \sqrt{\left((X^T X)^{-1}\right)_{11}}\right) = N(0,1). \tag{110}$$

Selektionsbias-Modelle

Anwendung des Egger-Tests auf die STBI-Daten aus Zhang et al. (2026) mit `lm()`

```
D           = read.csv("7-Daten/act-sitb-data.csv")         # Datensatz
D           = D[D$Symptom == "STBI", ]                      # STBI-Daten
sigma_i     = sqrt(D$Vi)                                    # Standardfehler
L           = lm(I(g/sigma_i) ~ I(1/sigma_i), data = D)     # SND-Regression
X           = model.matrix(L)                               # Designmatrix
V_beta      = solve(t(X) %*% X)                             # Varianz bei sigma^2 = 1
beta_0_hat  = coef(L)[1]                                    # Egger-Interzept
SE_beta_0   = sqrt(V_beta[1, 1])                            # bekannter SE
Z_E         = beta_0_hat/SE_beta_0                          # Z-Statistik
R           = data.frame(                                   # Ergebnistabelle
beta_0_hat  = beta_0_hat,                                   # Egger-Interzept
SE          = SE_beta_0,                                    # SE bei sigma^2 = 1
Z           = Z_E,                                          # Z-Statistik
p           = 2*pnorm(-abs(Z_E)))                           # Z-Test p-Wert
```

```
 beta_0_hat    SE     Z        p
      -3.83 0.288 -13.3 2.83e-40
```

Selektionsbias-Modelle

Anwendung des Egger-Tests auf die STBI-Daten aus Zhang et al. (2026) mit `metafor`

- `metafor` unterstellt einen unbekannten Varianzparameter im Egger-Test
- Entsprechend schätzt `metafor` diesen auch und nutzt eine T-Statistik

```
library(metafor)                                        # R Paket
D           = read.csv("7-Daten/act-sitb-data.csv")     # Datensatz
D           = D[D$Symptom == "STBI", ]                  # STBI-Daten
G           = regtest(                                  # klassischer Egger-Test
D$g,                                                    # Effektstärkeschätzer
vi          = D$Vi,                                     # Effektstärkenvarianzen
model       = "lm",                                     # klassische Option
predictor   = "sei",                                    # Standardfehlerprädiktor
ret.fit     = TRUE)                                     # Fit zurückgeben
C           = coef(summary(G$fit))                      # Koeffiziententabelle
R           = data.frame(                               # Ergebnistabelle
beta_0_hat  = C[2, 1],                                  # Egger-Interzept
SE          = C[2, 2],                                  # Standardfehler
Z_E         = G$zval,                                   # Z_E-Statistik
df          = G$dfs,                                    # Freiheitsgrade
p           = G$pval)                                   # p-Wert
```

```
 beta_0_hat    SE   Z_E df        p
      -3.83 0.688 -5.56 33 3.52e-06
```

Selektionsbias-Modelle

Abgleich mit Zhang et al. (2026) und `metafor` Default

```
library(metafor)                                                  # R Paket
D       = read.csv("7-Daten/act-sitb-data.csv")                   # Datensatz
D       = D[D$Symptom == "STBI", ]                                # STBI-Daten
G_lm    = regtest(D$g, vi = D$Vi, model = "lm",  predictor = "sei") # klassisch
G_rma   = regtest(D$g, vi = D$Vi, model = "rma", predictor = "sei") # metafor default
R       = data.frame(                                             # Abgleich
Variante = c("metafor lm", "metafor rma", "@zhang2026"),          # Variante
Z       = c(G_lm$zval, G_rma$zval, -2.26),                        # Teststatistik
p       = c(G_lm$pval, G_rma$pval, 0.024))                        # P-Wert
```

```
    Variante     Z        p
  metafor lm -5.56 3.52e-06
 metafor rma -2.25 2.46e-02
  @zhang2026 -2.26 2.40e-02
```

⇒ Der von Zhang et al. (2026) berichtete STBI-Egger-Test wird durch `metafor::regtest(..., model = "rma")` reproduziert.

Definition (Mixed-Effects-Egger-Test-Modell)

Das Mixed-Effects-Egger-Test-Modell hat die strukturelle Form

$$y_i = \delta + \beta_0 \sigma_i + b_i + \varepsilon_i \text{ mit } b_i \sim N(0, \theta) \text{ und } \varepsilon_i \sim N(0, \sigma_i^2) \text{ unabhängig für } i = 1, \ldots, k. \tag{111}$$

Mit $\beta := (\delta, \beta_0)^T$, $\sigma := (\sigma_1, \ldots, \sigma_k)^T$, $X := (1_k, \sigma)$, $Z := I_k$, $D := \text{diag}(\sigma_1^2, \ldots, \sigma_k^2)$ die Designmatrixform

$$y = X\beta + Zb + \varepsilon \text{ mit } b \sim N(0_k, \theta I_k) \text{ und } \varepsilon \sim N(0_k, D) \text{ unabhängig.} \tag{112}$$

Äquivalent gilt für die marginale Datenverteilung in Designmatrixform

$$y \sim N(X\beta, V_\theta) \text{ mit } V_\theta := Z\theta I_k Z^T + D = \theta I_k + D. \tag{113}$$

Bemerkungen

- Das Modell wurde von Sterne und Eggers in Rothstein et al. (2005) vorgeschlagen und ist in `metafor` implementiert.
- Für $\theta = 0$ reduziert sich die marginale Datenverteilung auf das Egger-Test-Modell in ESF-Form.

Definition (Mixed-Effects-Egger-Test)

Gegeben sei das Mixed-Effects-Egger-Test-Modell mit marginaler Datenverteilung

$$y \sim N(X\beta, V_\theta) \text{ mit } \beta := (\delta, \beta_0)^T \text{ und } V_\theta := \theta I_k + D. \tag{114}$$

Dann ist der *Mixed-Effects-Egger-Test* definiert als der Hypothesentest der einfachen Nullhypothese

$$H_0 : \beta_0 = 0. \tag{115}$$

Mit dem GLS-Schätzer

$$\hat{\beta} := \left(X^T V_\theta^{-1} X\right)^{-1} X^T V_\theta^{-1} y \tag{116}$$

und $c := (0, 1)^T$ ist die Mixed-Effects-Egger-Teststatistik gegeben durch

$$Z_{\text{ME}} := \frac{c^T \hat{\beta}}{\sqrt{c^T \left(X^T V_\theta^{-1} X\right)^{-1} c}} = \frac{\hat{\beta}_0}{\sqrt{\left(\left(X^T V_\theta^{-1} X\right)^{-1}\right)_{22}}}. \tag{117}$$

Theorem (Mixed-Effects-Egger-Test)

Gegeben sei das Mixed-Effects-Egger-Test-Modell mit bekannter marginaler Kovarianzmatrix V_θ und die Definition des Mixed-Effects-Egger-Tests. Dann gelten

(1) Verteilung des GLS-Schätzers

$$\hat{\beta} \sim N\left(\beta, \left(X^T V_\theta^{-1} X\right)^{-1}\right). \tag{118}$$

(2) Verteilung der Mixed-Effects-Egger-Teststatistik bei Zutreffen der Nullhypothese $H_0 : \beta_0 = 0$

$$Z_{\text{ME}} \sim N(0, 1). \tag{119}$$

Bemerkung

- Bei unbekanntem θ wird in der Anwendung θ durch $\hat{\theta}$ ersetzt.

Beweis

Aus der marginalen Datenverteilung des Mixed-Effects-Egger-Test-Modells folgt

$$y = X\beta + \eta \quad \text{mit} \quad \eta \sim N(0_k, V_\theta). \tag{120}$$

Nach dem GLS-Theorem gilt für

$$\hat{\beta} = \left(X^T V_\theta^{-1} X\right)^{-1} X^T V_\theta^{-1} y \tag{121}$$

die Verteilung

$$\hat{\beta} \sim N\left(\beta, \left(X^T V_\theta^{-1} X\right)^{-1}\right). \tag{122}$$

Mit $c := (0, 1)^T$ gilt also

$$c^T \hat{\beta} \sim N\left(c^T \beta, c^T \left(X^T V_\theta^{-1} X\right)^{-1} c\right). \tag{123}$$

Bei Zutreffen der Nullhypothese $H_0 : \beta_0 = 0$ gilt $c^T \beta = 0$, und damit folgt mit dem Theorem zur linearen-affinen Transformation eines normalverteilten Zufallsvektors

$$Z_{\text{ME}} = \frac{c^T \hat{\beta}}{\sqrt{c^T \left(X^T V_\theta^{-1} X\right)^{-1} c}} \sim N(0, 1). \tag{124}$$

□

Selektionsbias-Modelle

Anwendung des Mixed-Effects-Egger-Tests auf die STBI-Daten aus Zhang et al. (2026) mit `metafor`

```r
library(metafor)                                                  # R Paket
D           = read.csv("7-Daten/act-sitb-data.csv")               # Datensatz
D           = D[D$Symptom == "STBI", ]                            # STBI-Daten
D$sei       = sqrt(D$Vi)                                          # Standardfehler
M           = rma(                                                # metafor Mixed-Effects
yi          = g,                                                  # Effektschätzer
vi          = Vi,                                                 # Sampling Varianzen
mods        = ~ sei,                                              # Egger-Moderator
data        = D,                                                  # Datensatz
method      = "REML")                                             # ReML-Schätzung
G           = regtest(                                            # metafor regtest
D$g,                                                              # Effektstärkeschätzer
vi          = D$Vi,                                               # Effektstärkenvarianzen
model       = "rma",                                              # Mixed-Effects-Option
predictor   = "sei")                                              # Standardfehlerprädiktor
C_M         = coef(summary(M))                                    # metafor Koeffizienten
R           = data.frame(                                         # Ergebnistabelle
Variante    = c("metafor rma", "regtest rma", "@zhang2026"),      # Variante
tau2_ReML   = c(M$tau2, M$tau2, NA),                              # Heterogenität
beta_0_hat  = c(C_M["sei", 1], C_M["sei", 1], NA),                # Egger-Steigung
SE          = c(C_M["sei", 2], C_M["sei", 2], NA),                # Standardfehler
Z_ME        = c(C_M["sei", 3], G$zval, -2.26),                    # Wald-Z-Statistik
p           = c(C_M["sei", 4], G$pval, 0.024))                    # p-Wert
```

```
    Variante tau2_ReML beta_0_hat    SE  Z_ME      p
 metafor rma     0.597      -2.07 0.921 -2.25 0.0246
 regtest rma     0.597      -2.07 0.921 -2.25 0.0246
  @zhang2026        NA         NA    NA -2.26 0.0240
```

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selektionsbias-Modelle

Selbstkontrollfragen

Selbstkontrollfragen

1. Erläutern Sie, warum Metaanalysen aus ethischer Sicht sinnvoll sind.
2. Erläutern Sie den Unterschied zwischen Narrativen Reviews und Systematischen Reviews/Metaanalysen.
3. Nennen und erläutern Sie vier typische Probleme des Metaanalyse-Genres.
4. Geben Sie die Definition von Cohen's d im Parallelgruppendesign wieder.
5. Geben Sie die Definition des Einzelstudienmodells (ESMs) wieder.
6. Geben Sie die Definition des Effektstärkeschätzers im ESM wieder.
7. Geben Sie Aussage (1) und (2) des Theorems zu den Eigenschaften des Effektstärkeschätzers im ESM wieder.
8. Geben Sie das Theorem zum Unverzerrten Effektstärkeschätzer im ESM - Hedges' g wieder.
9. Geben Sie das Theorem zur Asymptotischen Normalverteilung von Hedges' g wieder.
10. Geben Sie die Definition des Fixed-Effects-Modells der Metaanalyse wieder und erläutern Sie es.
11. Geben Sie das Theorem zum FEM-ML-SES-Schätzer wieder.
12. Geben Sie das Theorem zur Q-Statistik wieder.
13. Geben Sie die Definition der I^2-Statistik wieder.
14. Geben Sie die Definition des Random-Effects-Modells der Metaanalyse wieder und erläutern Sie es.
15. Erläutern Sie das Vorgehen zur Schätzung der Parameter des Random-Effects-Modells der Metaanalyse.
16. Erläutern Sie die Grundidee eines Funnel-Plots im Kontext der Metaanalyse.

Definition (Nichtzentrale t-Zufallsvariable)

T sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{\nu-1}{2}} \Gamma\left(\frac{\nu}{2}\right) (\nu\pi)^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{\nu-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2}\left(t\left(\frac{\tau}{\nu}\right)^{\frac{1}{2}} - \mu\right)^2\right) d\tau. \quad (125)$$

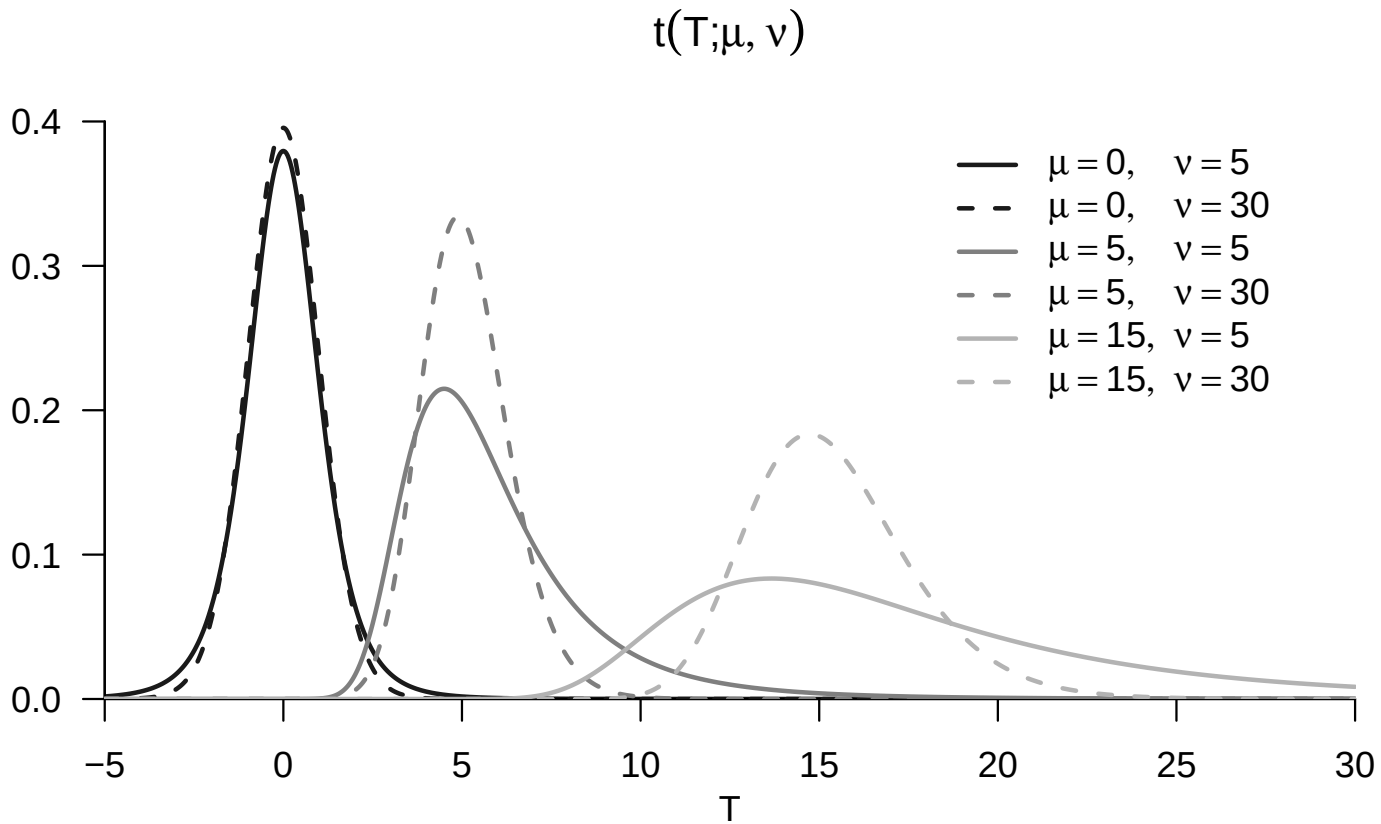
Dann sagen wir, dass T einer nichtzentralen t-Verteilung mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν unterliegt und nennen T eine *nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν*. Wir kürzen dies mit $T \sim t(\mu, \nu)$ ab. Erwartungswert und Varianz einer nichtzentral t-verteilten Zufallsvariablen T mit Nichtzentralitätsparameter μ und Freiheitsgradparameter $\nu > 1$ ergeben sich zu

$$\mathbb{E}(T) = \mu\sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \text{ und } \mathbb{V}(T) = \frac{\nu(1+\mu^2)}{\nu-2} - \frac{\mu^2\nu}{2}\left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2. \quad (126)$$

Bemerkung

- Nichtzentrale t-Zufallsvariablen sind für die Testgütefunktion des T-Tests essenziell.
- Für die Herleitung von Erwartungswert und Varianz verweisen wir auf Hogben et al. (1961).

Wahrscheinlichkeitsdichtefunktionen spezieller nichtzentraler t-Zufallsvariablen.



Referenzen I

Abyar, Z., B. Makvandi, S. Bakhtiarpoor, F. Naderi, and F. Hafezi. 2018. "Comparing the Effectiveness of Acceptance and Commitment Therapy and Mindfulness on the Suicidal Ideation Among Depressed Women." *Women Study* 9 (1): 1–23.

Borenstein, Michael, ed. 2009. *Introduction to Meta-Analysis*. John Wiley & Sons.

Cochran, William G. 1954. "The Combination of Estimates from Different Experiments." *Biometrics* 10 (1): 101. https://doi.org/10.2307/3001666.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. L. Erlbaum Associates.

Copas, John B. 1999. "What Works?: Selectivity Models and Meta-Analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 162 (1): 95–109. https://doi.org/10.1111/1467-985X.00123.

Cuijpers, P., E. Karyotaki, M. Reijnders, and D. D. Ebert. 2019. "Was Eysenck Right After All? A Reassessment of the Effects of Psychotherapy for Adult Depression." *Epidemiology and Psychiatric Sciences* 28 (1): 21–30. https://doi.org/10.1017/S2045796018000057.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *BMJ* 315 (7109): 629–34. https://doi.org/10.1136/bmj.315.7109.629.

Eysenck, H J. 1952. "The Effects of Psychotherapy: An Evaluation." *Journal of Consulting Psychology* 16 (5): 319–24.

Fuller, Wayne A., and George E. Battese. 1974. "Estimation of Linear Models with Crossed-Error Structure." *Journal of Econometrics* 2 (1): 67–78. https://doi.org/10.1016/0304-4076(74)90030-X.

Glass, Gene V. 1976. *Primary, Secondary, and Meta-Analysis of Research*. 7.

Harrer, Mathias, Pim Cuijpers, Toshi A. Furukawa, and David D. Ebert. 2021. *Doing Meta-Analysis with R: A Hands-On Guide*. 1st ed. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347.

Referenzen II

Hedges, Larry V. 1981. *Distribution Theory for Glass's Estimator of Effect Size and Related Estimators*. 23.

Hedges, Larry V. 1982. "Fitting Categorical Models to Effect Sizes from a Series of Experiments." *Journal of Educational Statistics* 7 (2): 119. https://doi.org/10.2307/1164961.

Hedges, Larry V. 1992. "Modeling Publication Selection Effects in Meta-Analysis." *Statistical Science* 7 (2): 246–55. https://doi.org/10.1214/ss/1177011364.

Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. Academic Press.

Higgins, Julian P. T., and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21 (11): 1539–58. https://doi.org/10.1002/sim.1186.

Hoaglin, David C. 2016. "Misunderstandings about *Q* and 'Cochran's *Q* Test' in Meta-Analysis." *Statistics in Medicine* 35 (4): 485–95. https://doi.org/10.1002/sim.6632.

Hogben, D., R. S Pinkham, and M. B. Wilk. 1961. *The Moments of the Non-Central t-Distribution*. 5.

Johnson, N. L., and B. L. Welch. 1940. "Applications of the Non-Central t-Distribution." *Biometrika* 31 (3/4): 362. https://doi.org/10.2307/2332616.

Kennedy, James E., Richard Wiseman, and Caroline Watt. 2026. "The Emerging Disenchantment with Retrospective Meta-Analysis." *Nature Human Behaviour* 10 (6): 1019–21. https://doi.org/10.1038/s41562-026-02463-y.

Kulinskaya, Elena, Michael B. Dollinger, and Kirsten Bjørkestøl. 2011a. "On the Moments of Cochran's *Q* Statistic Under the Null Hypothesis, with Application to the Meta-Analysis of Risk Difference." *Research Synthesis Methods* 2 (4): 254–70. https://doi.org/10.1002/jrsm.54.

Kulinskaya, Elena, Michael B. Dollinger, and Kirsten Bjørkestøl. 2011b. "Testing for Homogeneity in Meta-Analysis I. The One-Parameter Case: Standardized Mean Difference." *Biometrics* 67 (1): 203–12. https://doi.org/10.1111/j.1541-0420.2010.01442.x.

Referenzen III

Langan, Dean, Julian P. T. Higgins, Dan Jackson, et al. 2019. "A Comparison of Heterogeneity Variance Estimators in Simulated Random-Effects Meta-Analyses." *Research Synthesis Methods* 10 (1): 83–98. https://doi.org/10.1002/jrsm.1316.

Light, Richard J., and David B. Pillemer. 1984. *Summing Up: The Science of Reviewing Research*. Harvard University Press.

Lin, Lifeng, and Ariel M. Aloe. 2021. "Evaluation of Various Estimators for Standardized Mean Difference in Meta-Analysis." *Statistics in Medicine* 40 (2): 403–26. https://doi.org/10.1002/sim.8781.

Lin, Lifeng, and Haitao Chu. 2018. "Quantifying Publication Bias in Meta-Analysis." *Biometrics* 74 (3): 785–94. https://doi.org/10.1111/biom.12817.

Mavridis, Dimitris, Nicky J. Welton, Alex Sutton, and Georgia Salanti. 2014. "A Selection Model for Accounting for Publication Bias in a Full Network Meta-Analysis." *Statistics in Medicine* 33 (30): 5399–412. https://doi.org/10.1002/sim.6321.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and 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.

Page, Matthew J, Joanne E McKenzie, Patrick M Bossuyt, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *BMJ*, March, n71. https://doi.org/10.1136/bmj.n71.

Peto, R, and S Parish. 1980. "Aspirin After Myocardial Infarction." *The Lancet* 315 (8179): 1172–73. https://doi.org/10.1016/S0140-6736(80)91626-8.

Rencher, Alvin C., and William F. Christensen. 2012. *Methods of Multivariate Analysis*. Third Edition. Wiley Series in Probability and Statistics. Wiley.

Referenzen IV

Rothstein, Hannah, A. J. Sutton, and Michael Borenstein. 2005. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Wiley.

Schaeuffele, Carmen, Laura E. Meine, Ava Schulz, et al. 2024. "A Systematic Review and Meta-Analysis of Transdiagnostic Cognitive Behavioural Therapies for Emotional Disorders." *Nature Human Behaviour* 8 (3): 493–509. https://doi.org/10.1038/s41562-023-01787-3.

Smith, Mary Lee. 1977. "Meta-Analysis of Psychotherapy Outcome Studies." *American Psychologist* 32 (9): 752–60.

Sterne, Jonathan A C, and Matthias Egger. 2001. "Funnel Plots for Detecting Bias in Meta-Analysis: Guidelines on Choice of Axis." *Journal of Clinical Epidemiology* 54: 1046–55.

Veroniki, Areti Angeliki, Dan Jackson, Wolfgang Viechtbauer, et al. 2016. "Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis." *Research Synthesis Methods* 7 (1): 55–79. https://doi.org/10.1002/jrsm.1164.

Viechtbauer, Wolfgang. 2005. "Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model." *Journal of Educational and Behavioral Statistics* 30 (3): 261–93. https://doi.org/10.3102/10769986030003261.

Viechtbauer, Wolfgang. 2010. "Conducting Meta-Analyses in *R* with the **Metafor** Package." *Journal of Statistical Software* 36 (3). https://doi.org/10.18637/jss.v036.i03.

Zhang, Tao, Bryant Pui Hung Hui, Déborah Ducasse, et al. 2026. "Efficacy of Acceptance and Commitment Therapy for Suicide and Self-Harm: A Systematic Review and Meta-Analysis." *Psychotherapy and Psychosomatics* 95 (3): 193–213. https://doi.org/10.1159/000548398.