



Psychotherapieforschung

MSc Klinische Psychologie und Psychotherapie

SoSe 2025

Prof. Dr. Dirk Ostwald

(6) Metaanalyse

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selbstkontrollfragen

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-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 potentiellen Mehrwert von Großforschung.
- Es ist intuitiv sinnvoll und plausibel nach der gesammelten Evidenz für etwas zu fragen.
- Metaanalysen wird das Potential 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.

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

Narrative Reviews

- Standard der Akkumulation wissenschaftlicher Evidenz bis etwas 1990
- Subjektive und intransparente Auswahl von Studien durch Expert:innen
- Nicht-explicite 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)

Metaanalysen in der Medizin

- Inhaltliche Beiträge, z.B. Peto and Parish (1980)
- Methodische Beiträge, z.B. DerSimonian and Laird (1986)
- 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 Metaanalysen-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.	

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

Versuche der Qualifikation von Biases

Einschätzung des Verzerrungsrisikos anhand des Cochrane Risk of Bias Tools

		Risk of bias domains					
		D1	D2	D3	D4	D5	Overall
Ref.	¹⁰						
Ref.	¹¹						
Ref.	¹²						
Ref.	¹³						
Ref.	¹⁴						
Ref.	¹⁵						
Ref.	¹⁶						
Ref.	¹⁷						
Ref.	¹⁸						
Ref.	¹⁹						
Ref.	²⁰						
Ref.	²¹						

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

High

Some concerns

Low

Schaeffele et al. (2024)

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



Trusted evidence.
Informed decisions.
Better health.

[Home](#)[Features](#)[Support?](#)[Pricing](#)[Trial](#)[My Reviews](#)

RevMan: Systematic review and meta-analysis tool for researchers worldwide

Streamline your systematic reviews and meta-analyses with Cochrane's Review Manager (RevMan). Access powerful tools for efficient data management, analysis, and reporting. Promoting high quality research worldwide.

[Pricing & subscription](#)

Intuitive and easy to use

Meta-analysis has never been so easy.

High quality support and guidance

Up-to date how-to articles and screencasts guides you to get started.

Promotes high quality research

Guides you to pre-define study PICO's and analysis criteria to make decisions before extracting data.

Collaborate on the same project

Access and work together on your reviews anywhere.

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 Filedrawer Problem

- Nicht alle 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)

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selbstkontrollfragen

Anwendungsbeispiel

Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression

“We used a database of randomised trials of psychotherapies for adult depression that has been described in a methods paper (Cuijpers et al. 2008a). The general methods we have used in this paper have been described in a manual that is freely available (Cuijpers, 2016b). In brief, the database was developed through a comprehensive literature search (from 1966 until 1 January 2017), and is updated every year. We searched major bibliographical databases (PsycINFO, PubMed, Embase, Cochrane Central Register of Controlled Trials). The full search string for PubMed is given in Appendix A.

In this database, we included all randomised trials in which at least one arm was a psychological treatment for adults (>18 years) with a depressive disorder according to a diagnostic interview or an elevated level of depressive symptomatology (as indicated by a score above a cut-off score on a validated self-report depression scale). In the current study, we only used trials that compared a psychotherapy for adult depression with a control group (waiting list, care-as-usual, placebo or other; this last category included control conditions that could not be categorised into one of the three categories, such as participation in online discussion forums, in workshops on other subjects, routine care in general medical care or an information booklet).

We calculated effect sizes (Hedges' g) for each comparison between a psychotherapy and a comparison group. We only included outcome measures that assess depressive symptoms. If there is more than one measure of depression, these are pooled within the study, before the overall effects are pooled across studies.”

Cuijpers et al. (2019)

Anwendungsbeispiel

Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression

Table 1. Effects of psychotherapies compared with control groups ($k=369$): Hedges' g^a

	k	g	95% CI
All studies	369	0.70	0.64–0.75
Extreme outliers excluded ($g > 2.0$)	352	0.61	0.57–0.66
One effect size per study (only highest)	289	0.70	0.65–0.76
One effect size per study (only lowest)	289	0.64	0.58–0.69
Only HAM-D	103	0.86	0.75–0.97
Only BDI	128	0.87	0.77–0.98
Only BDI-II	80	0.68	0.57–0.80
Subgroup analyses			
Type of therapy			
CBT	192	0.70	0.63–0.77
Behavioural activation	20	0.94	0.66–1.22
Interpersonal psychotherapy	25	0.60	0.40–0.80
Problem-solving therapy	27	0.77	0.54–1.01
Third wave therapies	18	0.77	0.58–0.97
Supportive counselling	19	0.58	0.42–0.75
Psychodynamic therapy	11	0.40	0.17–0.63
Other	57	0.68	0.54–0.82
Control group			
Waiting list	159	0.89	0.80–0.98
Care-as-usual	144	0.61	0.53–0.68
Other	66	0.51	0.40–0.62
Country			
Western	325	0.63	0.58–0.68
Non-Western	44	1.13	0.94–1.33
Risk of bias score			
0 (high risk)	14	1.11	0.87–1.36
1	122	0.92	0.79–1.05
2	63	0.73	0.60–0.86
3	62	0.71	0.58–0.85
4 (low risk)	108	0.46	0.41–0.52

BDI, Beck Depression Inventory; CBT, cognitive behavioural therapy; CI, confidence interval; N_{comp} , number of comparisons; NNT, numbers-needed-to-treat.

^aAccording to the random-effects model.

Cuijpers et al. (2019)

Anwendungsbeispiel

Is psychotherapy effective? A re-analysis of treatments for depression

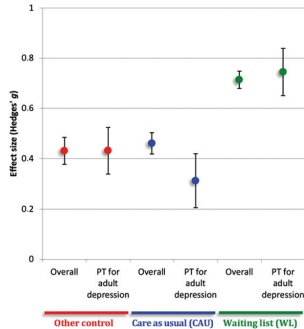


Fig. 1. Effect sizes for psychological interventions for depression. Error bars represent standard errors. PT = psychotherapy. Overall is based on all effect sizes without outliers and corrected for publication bias ($k = 146$ contrasts with WL, $k = 142$ contrasts with CAU, $k = 65$ contrasts with 'other' controls). PT for adult depression only includes (individual or group) psychotherapy for adults with a diagnosis of depression ($k = 30$ contrasts with WL, $k = 29$ contrasts with CAU, $k = 12$ contrasts with 'other control').

Munder et al. (2019)

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. If the population product moment r between neurophysiological and questionnaire measures of introversion-extroversion is .30, the ES is .30; if r is .60, so is the ES, a larger value and a larger departure from the null hypothesis, which here is $r = 0$.
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.

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."

Cohen (1988) Statistical power analysis for the behavioral sciences

Cohen's d in den Worten von Jacob Cohen

2.2 THE EFFECT SIZE INDEX: d

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 of the alternate 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, the latter also in the original measurement unit. For the two independent samples case, this is simply

$$(2.2.1) \quad d = \frac{m_A - m_B}{\sigma}$$

for the directional (one-tailed) case, and

$$(2.2.2) \quad d = \frac{|m_A - m_B|}{\sigma}$$

for the nondirectional (two-tailed) case,

where d = ES index for t tests of means in standard unit,

m_A, m_B = population means expressed in raw (original measurement) unit, and

σ = the standard deviation of either population (since they are assumed equal).

Cohen (1988) Statistical power analysis for the behavioral sciences

Cohen's d in den Worten der metaanalytischen Community

Computing d and g from studies that use independent groups

We can estimate the standardized mean difference (δ) from studies that used two independent groups as

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}. \quad (4.18)$$

In the numerator, $\bar{X}_1$ and $\bar{X}_2$ are the sample means in the two groups. In the denominator S_{within} is the within-groups standard deviation, pooled across groups,

$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \quad (4.19)$$

where n_1 and n_2 are the sample sizes in the two groups, and S_1 and S_2 are the standard deviations in the two groups. The reason that we pool the two sample

estimates of the standard deviation is that even if we assume that the underlying population standard deviations are the same (that is $\sigma_1 = \sigma_2 = \sigma$), it is unlikely that the sample estimates S_1 and S_2 will be identical. By pooling the two estimates of the standard deviation, we obtain a more accurate estimate of their common value.

The sample estimate of the standardized mean difference is often called Cohen's d in research synthesis. Some confusion about the terminology has resulted from the fact that the index δ , originally proposed by Cohen as a *population parameter* for describing the size of effects for statistical power analysis is also sometimes called d . In this volume we use the symbol δ to denote the effect size parameter and d for the sample estimate of that parameter.

Definition (Cohen's d im Parallelgruppen-Design)

Gegeben seien ein Parallelgruppen-Design mit einer Treatmentgruppe t und einer Kontrollgruppe c und es seien

$$y_{t1}, \dots, y_{tn_t} \text{ und } y_{c1}, \dots, y_{cn_c} \quad (1)$$

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

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

die Mittelwerte und die empirischen Varianzen beider Gruppen. Schließlich sei

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

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

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

Bemerkungen

- $\bar{y}_t - \bar{y}_c$ ist die *empirische Effektstärke*.
- d ist die mithilfe der gepoolten empirischen Standardabweichung *standardisierte empirische Effektstärke*.
- d misst den Unterschied der Mittelwerte in Einheiten der gepoolten Standardabweichung, z.B.

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

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

$$T = \sqrt{\frac{n_t n_c}{n_t + n_c}} d \Leftrightarrow d = \sqrt{\frac{n_t + n_c}{n_t n_c}} T. \quad (6)$$

- 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

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

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

Bestimmung von Cohen's d für den Beispieldatensatz

```
D      = read.csv("./6_Daten/parallelgruppendedesign.csv")      # Einlesen des Datensatzes
y_t    = D$dBDI[D$Group == "T"]                                # Daten der Treatmentgruppe
y_c    = D$dBDI[D$Group == "C"]                                # 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
s6_t   = var(y_t)                                               # empirische Varianz der Treatmentgruppe
s6_c   = var(y_c)                                               # empirische Varianz der Kontrollgruppe
s      = sqrt(((n_t-1)*s6_t + (n_c-1)*s6_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      : 3.07
Mittelwert Kontrollgruppe       : -0.4
Gepoolte empirische Standardabweichung : 3.82
Empirische Effektstärke         : 3.47
Cohen's d                       : 0.91
```

Hedge's g

Hedges' g ist ein durch einen Multiplikationsfaktor angepasstes Cohen's d , das von Larry V. Hedges (1981) und Larry V. 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 Metanalyse werden Hedges' g und sein Varianzschätzer meist zunächst für jede Einzelstudie bestimmt und dann über Studien, zum Beispiel mit den Fixed-Effects- und Random-Effects-Modellen aus Einheit (4) integriert.

Definition (Einzelstudienmodell)

Es seien y_{tj} für $j = 1, \dots, n_t$ und y_{cj} für $j = 1, \dots, n_c$ 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

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

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

$$\delta := \frac{\mu_t - \mu_c}{\sigma}. \quad (9)$$

Bemerkungen

- Das Einzelstudienmodell ist offenbar äquivalent zum Zweistichproben-T-Test-Modell.
- Mit der Theorem zur linear-affinen Transformation von normalverteilten Zufallsvariablen gilt äquivalent

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

- Die hier gewählte Formulierung entspricht Larry V. Hedges and Olkin (1985) p. 76 (vgl. Lin and Aloe (2021))
- Larry V. 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

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

wobei

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

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

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

der Treatment- und Kontrollgruppe definierte *gepoolte Stichprobenstandardabweichung*

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

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-Schätzer im ESM)

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

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

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

Remarks

- $\sqrt{n}d$ 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 Appendix.
- 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 Larry V. Hedges (1981) auf Johnson and Welch (1940).

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

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

Es ergibt sich also

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

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

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

Es ergibt sich also

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

Effektstärkeschätzung

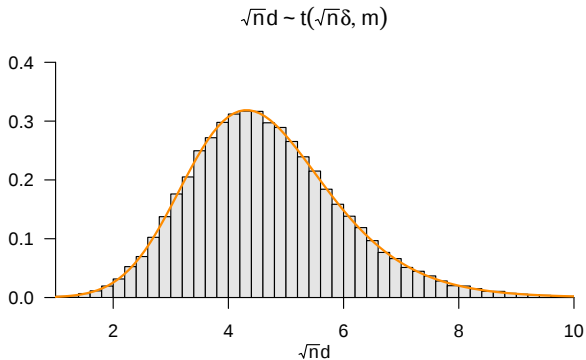
Verteilung von $\sqrt{nd}$

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                                  # Varianzparameter
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(NA, sim)                      # SES-Schätzerarrayinitialisierung
for(i in 1:sim){                             # Simulationsiterationen
  y       = matrix(rep(NA, n_t+n_c), ncol = 2) # Datenanarrayinitialisierung
  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 $\sqrt{nd}$

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



Theorem (Hegdes' g im ESM)

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

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

wie oben definiert. Dann ist

$$g := c_m d, \quad (21)$$

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

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

Beweis

Die Unverzerrtheit von d_u ergibt sich aus

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

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

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

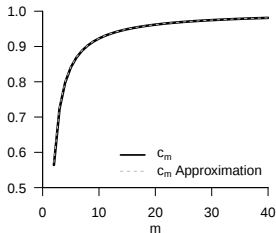
□

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

$$c_m \approx 1 - \frac{3}{4m - 1} \quad (25)$$

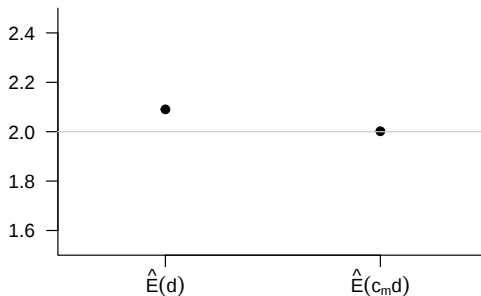
angenähert werden, vgl. Larry V. Hedges (1981), Borenstein (2009) und untenstehende Abbildung.



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

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 $n_t \rightarrow \infty$, $n_c \rightarrow \infty$ asymptotisch normalverteilt und es gilt speziell

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

Bemerkungen

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

Effektstärkeschätzung

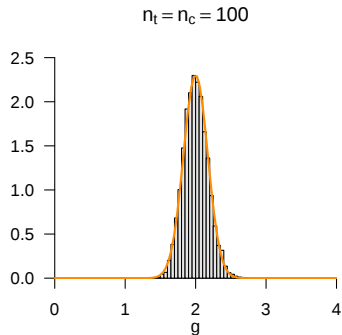
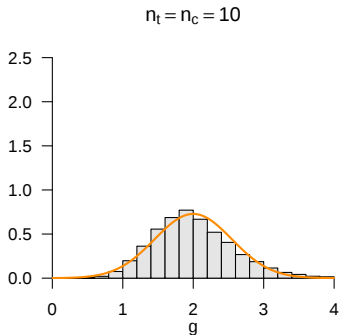
Asymptotische Normalverteilung von Hedges' g

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            # Varianzparameter
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) # Datenanarrayinitialisierung
    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

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

Dann wird der Varianzparameter der asymptotischen Verteilung von g die *asymptotische Varianz* von g genannt und mit

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

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

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

- $\hat{\mathbb{V}}_a(g)$ wird häufig für die metanalytische Modellbildung genutzt, vgl. Einheit (4) und 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 unverzerzte Schätzer der standardisierten Effektstärke δ und für ein Konfidenzlevel $\gamma \in]0, 1[$ sei

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

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

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

ein asymptotisches γ -Konfidenzintervall für die standardisierte Effektstärke δ .

Bemerkungen

- Das betrachtete Konfidenzintervall wurde von Larry V. 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

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

Dazu halten wir zunächst fest, dass

$$g \stackrel{a}{\sim} N(\delta, \mathbb{V}_a(g)). \quad (33)$$

Mit dem Theorem zur Z -Transformation gilt dann

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

Weiterhin gilt per Definition von z_γ , dass

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

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.

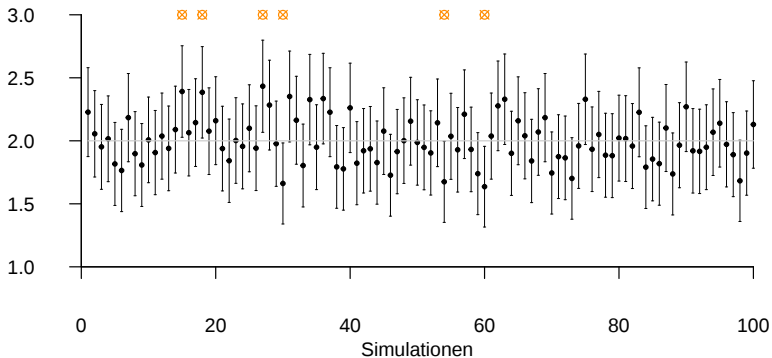
Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES

```
set.seed(0)
sim      = 1e2
mu_t     = 2
mu_c     = 0
sigma    = 1
delta    = (mu_t - mu_c)/sigma
gamma    = 0.95
z_gamma  = qnorm((1+gamma)/2)
n_t      = 100
n_c      = 100
n        = (n_t*n_c)/(n_t+n_c)
m        = n_t + n_c - 2
c_m      = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))
gs       = rep(NA, sim)
vs       = rep(NA, sim)
kappa    = array(rep(NA, sim*2), dim = c(2,sim))
for(i in 1:sim){
  y       = matrix(rep(NA,n_t+n_c), ncol = 2)
  y[,1]   = rnorm(n_t, mu_t, sigma)
  y[,2]   = rnorm(n_c, mu_c, sigma)
  ybar    = apply(y,2,mean)
  s2      = apply(y,2,var)
  s       = sqrt(((n_t-1)*s2[1] + (n_c-1)*s2[2])/m)
  gs[i]   = c_m*(ybar[1]-ybar[2])/s
  vs[i]   = ((n_t+n_c)/(n_t*n_c))*(gs[i]^2)/(2*(n_t+n_c))
  kappa[1,i] = gs[i] - sqrt(vs[i])*z_gamma
  kappa[2,i] = gs[i] + sqrt(vs[i])*z_gamma
}
```

Zufallsgeneratorstate
Anzahl Realisierungen
Erwartungswertparameter Treatment
Erwartungswertparameter Control
Varianzparameter
wahre, aber unbekannte, SES
Konfidenzlevel
Konfidenzlevelkonstante
Treatmentgruppenumfang
Kontrollgruppenumfang
Stichprobenumfangparameter
Freiheitsgradparameter
Biaskorrekturfaktor
Hedges' g Arrayinitialisierung
asymptotische Varianzarray
Konfidenzintervallarray
Simulationsiterationen
Datenanarrayinitialisierung
Datengeneration Treatmentgruppe
Datengeneration Kontrollgruppe
Gruppenmittelwerte
Gruppenvarianzen
gepoolte Standardabweichung
Hedges' g
Varianzschätzer
untere Konfidenzintervallgrenze
obere Konfidenzintervallgrenze

Effektstärkeschätzung

Simulation eines asymptotischen 95%-Konfidenzintervalls für die SES



Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selbstkontrollfragen

Definition (Fixed-Effects Modell der Metaanalyse)

Für Studien $i = 1, \dots, k$ 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

$$y_i \sim N(\delta, \sigma_i^2), \quad (37)$$

wobei δ die wahre, aber unbekannte, standardisierte Effektstärke (SES) bezeichnet und die $y_1, \dots, y_k$ 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 $\sigma_i^2, i = 1, \dots, k$ 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 Larry V. Hedges and Olkin (1985) und in Viechtbauer (2010) implementierten Schätzer

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

- 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-Effect-Modell der Metaanalyse mit wahrer, aber unbekannter, standardisierter Effektstärke δ . Dann ergibt sich der Maximum-Likelihood-Schätzer der standardisierten Effektstärke δ zu

$$\hat{\delta} := \frac{1}{\sum_{i=1}^k w_i} \sum_{i=1}^k w_i y_i \text{ mit den Wichtungparametern } w_i := \frac{1}{\sigma_i^2}, i = 1, \dots, k. \quad (39)$$

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 Wichtungparameter werden als reziproke Werte der studienspezifischen Varianz des standardisierten Effektschärkeschätzers auch *studienspezifische Präzisionen* des standardisierten Effektstärkeschätzers genannt.
- Sind die studienspezifischen Varianzen σ_i^2 alle identisch zu σ^2 , so ergibt sich mit $w := 1/\sigma^2$

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

Fixed-Effects-Modelle

Beweis

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

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

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

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

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

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

Es ergibt sich also die Maximum-Likelihood Bedingung

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

Auflösen ergibt dann

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

□

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

Gegeben sei das Fixed-Effect-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.
- Aussage (2) wird z.B. in der Herleitung des DerSimonian-Laird Schätzers benötigt.

Beweis

(1) Mit den Eigenschaften des Erwartungswerts gilt

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

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

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

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(NA, k)                    # wahre, bekannte, studienspezifische Varianzenarray
y_i        = rep(NA, 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                    # Wichtungparameter
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 Metanalyse und es sei

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

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

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

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 Effektstärkerschätzer y_i im Fixed-Effects-Modells tatsächlich auf ein und diesselbe Effektstärke δ zurückgehen, ist Q asymptotisch χ^2 -verteilt. Ansätze für Beweise dieser Tatsache finden sich in Cochran (1954) und Larry V. 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, Dollinger, and Bjørkestøl (2011b) und Kulinskaya, Dollinger, and Bjørkestøl (2011a).

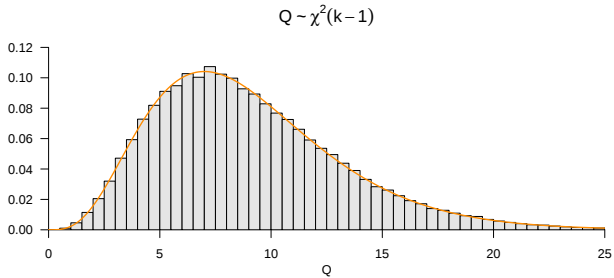
Verteilung der Q-Statistik

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

```
set.seed(1)                                # reproduzierbare Ergebnisse
sim      = 1e5                             # Anzahl Realisierungen
k        = 10                             # Anzahl an Studien
delta    = 1                              # wahre, aber unbekannte, Effektstärke
Q        = rep(NA, sim)                   # Q-Statistik Array
for(s in 1:sim){                          # Simulationsiterationen
  sigsq_r_i = rep(NA, k)                  # wahre, bekannte, studienspezifische Varianz
  y_i      = rep(NA, k)                  # Empirische Effektstärkenarray
  for(i in 1:k){                          # Studieniterationen
    sigsq_r_i[i] = runif(1, 1, 2)         # wahre, bekannte, studienspezifische Varianz
    y_i[i]      = rnorm(1, delta, sqrt(sigsq_r_i[i])) # empirische Effektstärke
    w_i        = 1/sigsq_r_i              # Wichtungparameter
    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 $\sigma_i \sim U(1, 2)$



Definition (I^2 -Statistik)

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

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

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 Schaeffele 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.

Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-Modelle

Selbstkontrollfragen

Definition (Random-Effects-Modell der Metaanalyse)

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

$$y_i \sim N(\delta_i, \sigma_i^2) \text{ und } \delta_i \sim N(\delta, \tau^2), \quad (51)$$

wobei

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

bezeichnen und die $\delta_1, \dots, \delta_k, y_1, \dots, y_k$ als unabhängige Zufallsvariablen 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 standardisierte Effektstärke wird auch als *Heterogenität* bezeichnet.

Bemerkungen (fortgesetzt)

- Für $i = 1, \dots, k$ impliziert die Definition die marginalen und die bedingten Wahrscheinlichkeitsdichtefunktionen

$$p(y_i|\delta_i) = N(y_i; \delta_i, \sigma_i^2) \text{ und } p(\delta_i) = N(\delta_i; \delta, \tau^2) \quad (52)$$

und damit die gemeinsame WDF

$$p(y_i, \delta_i) = p(y_i|\delta_i)p(\delta_i) = N(y_i; \delta_i, \sigma_i^2)N(\delta_i; \delta, \tau^2). \quad (53)$$

- Mit den Eigenschaften der multivariaten Normalverteilung impliziert dies die Marginalverteilung

$$p(y_i) = N(y_i; \delta, \sigma_i^2 + \tau^2) \quad (54)$$

- In struktureller Form impliziert die Definition, dass

$$y_i = \delta + \xi_i + \varepsilon_i \text{ mit } \xi_i \sim N(0, \tau^2) \text{ und } \varepsilon_i \sim N(0, \sigma_i^2) \quad (55)$$

- 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 additivem normalverteilten Fehlers mit unbekannter Varianz τ^2 als δ_i indirekt beobachtet. Direkt beobachtet wird dann mit y_i eben dieses δ_i unter dem Einfluss additivem normalverteilten Fehlers mit bekannter Varianz σ_i^2 .

Bemerkungen (fortgesetzt)

- Bei dem Random-Effects-Modell der Metaanalyse handelt es sich um das LMM in Kompaktdarstellung

$$y = X\beta + Zb + \varepsilon \quad (56)$$

mit

$$y := (y_1, \dots, y_k)^T, X := 1_k, \beta := \delta, Z := I_k \quad (57)$$

sowie

$$b \sim N(0_k, \tau^2 I_k) \text{ und } \varepsilon \sim N(0_k, \text{diag}(\sigma_1^2, \dots, \sigma_k^2)) \quad (58)$$

- Mit $A := 1_k$ und damit $X = ZA$ ergibt dabei eine Populationsparameterinterpretation für die wahre, aber unbekannte, standardisierte Effektstärke δ (vgl. Theorem zur Populationsparameterdarstellung eines Linear Mixed Models).

Definition (REM-DerSimonian-Laird-Schätzer)

Gegeben sei das Random-Effects-Modell der Metanalyse mit wahrer, aber unbekannter, standardisierter Effektstärke δ und wahrer, aber unbekannter, studienübergreifender Effektstärkevarianz (Heterogenität) τ^2 .

- Der *DerSimonian-Laird Schätzer* von τ^2 ist definiert als

$$\hat{\tau}_{DL}^2 = \frac{Q - (k - 1)}{c} \text{ mit } c := \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i}, \quad (59)$$

wobei

$$Q := \sum_{i=1}^k w_i (y_i - \hat{\delta})^2 \text{ und } \hat{\delta} := \frac{1}{\sum_{i=1}^k w_i} \sum_{i=1}^k w_i y_i \text{ mit } w_i := \frac{1}{\sigma_i^2} \text{ für } i = 1, \dots, k \quad (60)$$

die Q -Statistik bzw. den FEM-ML-SES-Schätzer bezeichnen.

- Der *DerSimonian-Laird Schätzer* von δ ist definiert als

$$\hat{\delta}_{DL} := \frac{1}{\sum_{i=1}^k \tilde{w}_i} \sum_{i=1}^k \tilde{w}_i y_i \text{ mit } \tilde{w}_i := \frac{1}{\hat{\tau}_{DL}^2 + \sigma_i^2} \text{ für } i = 1, \dots, k. \quad (61)$$

Bemerkungen

- Die REM-DerSimonian-Laird Schätzer gehen auf DerSimonian and Laird (1986) zurück.
- Es gibt viele verschiedene Schätzer für τ^2 und δ im REM, vgl. Viechtbauer (2005) und Veroniki et al. (2016).
- Heutzutage wird meist eine Linear-Mixed-Modell-ReML-basierte Schätzung bevorzugt.

Theorem (Motivation der DerSimonian-Laird Schätzung)

Gegeben sei das Random-Effects-Modell der Metanalyse mit wahrer, aber unbekannter, standardisierter Effektstärke δ und wahrer, aber unbekannter, studienübergreifender Effektstärkevarianz (Heterogenität) τ^2 . Weiterhin seien $\hat{\tau}_{DL}^2$ und $\hat{\delta}_{DL}$ die DerSimonian-Laird Schätzer von τ^2 und δ , respektive.

Dann gelten

- (1) $\hat{\tau}_{DL}^2$ ist ein Momentenmethodenschätzer von τ^2 basierend auf dem Erwartungswert der Q -Statistik.
- (2) $\hat{\delta}_{DL}$ ist ein Maximum-Likelihood Schätzer hinsichtlich der marginalen Datenverteilung mit $\tau^2 := \hat{\tau}_{DL}^2$.

Bemerkungen

Die Momentenmethode zur Schätzergewinnung besteht aus drei Schritten

- (1) Schreiben der Momente als Funktionen des zu schätzenden Parameters.
- (2) Ersetzen der Momente durch ihre Stichprobenäquivalente.
- (3) Auflösen der Gleichung hinsichtlich des zu schätzenden Parameters.

Nach DerSimonian and Laird (1986): "(...) Q ist used to derive a noniterative estimate of τ^2 by equating the sample statistic with the corresponding expected value. This yields a weighted estimator (...)."

Beweis

(1) Wir bestimmen zunächst den Erwartungswert der Q -Statistik als Funktion der nach Voraussetzung bekannten Werte von $\sigma_1^2, \dots, \sigma_k^2$ und des unbekannten Parameters τ^2 (vgl. Kacker (2004), DerSimonian and Kacker (2007))

$$\begin{aligned}\mathbb{E}(Q) &= \mathbb{E} \left(\sum_{i=1}^k w_i (y_i - \hat{\delta})^2 \right) \\ &= \sum_{i=1}^k w_i \mathbb{E} ((y_i - \hat{\delta})^2) \\ &= \sum_{i=1}^k w_i \mathbb{E} (((y_i - \hat{\delta}) - 0)^2) \\ &= \sum_{i=1}^k w_i \mathbb{E} (((y_i - \hat{\delta}) - (\mathbb{E}(y_i) - \mathbb{E}(\hat{\delta})))^2) \\ &= \sum_{i=1}^k w_i \mathbb{E} (((y_i - \hat{\delta}) - \mathbb{E}(y_i - \hat{\delta}))^2) \\ &= \sum_{i=1}^k w_i \mathbb{V}(y_i - \hat{\delta}) \\ &= \sum_{i=1}^k w_i (\mathbb{V}(y_i) + \mathbb{V}(\hat{\delta}) - 2\mathbb{C}(y_i, \hat{\delta})).\end{aligned}\tag{62}$$

Beweis (fortgeführt)

Dabei gilt nachdem Theorem zu den Eigenschaften des FEM-ML-SES-Schätzers, dass

$$\mathbb{V}(\hat{\delta}) = \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{(\sum_{i=1}^k w_i)^2}. \quad (63)$$

Mit der Eigenschaft der Kovarianz bei linear-affinen Transformationen sowie der Unabhängigkeit der y_i gilt ferner

$$\begin{aligned} \mathbb{C}(y_i, \hat{\delta}) &= \mathbb{C}\left(y_i, \frac{1}{\sum_{j=1}^k w_j} \sum_{j=1}^k w_j y_j\right) \\ &= \frac{1}{\sum_{j=1}^k w_j} \mathbb{C}\left(\sum_{j=1}^k w_j y_j, y_i\right) \\ &= \frac{\sum_{j=1}^k w_j \mathbb{C}(y_j, y_i)}{\sum_{j=1}^k w_j} \\ &= \frac{w_i \mathbb{C}(y_i, y_i)}{\sum_{j=1}^k w_j} \\ &= \frac{w_i \mathbb{V}(y_i)}{\sum_{j=1}^k w_j}. \end{aligned} \quad (64)$$

Beweis (fortgeführt)

Es ergibt sich also

$$\begin{aligned}\mathbb{E}(Q) &= \sum_{i=1}^k w_i \left(\mathbb{V}(y_i) + \mathbb{V}(\hat{\delta}) - 2\mathbb{C}(y_i, \hat{\delta}) \right) \\&= \sum_{i=1}^k w_i \left(\mathbb{V}(y_i) + \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{(\sum_{i=1}^k w_i)^2} - 2 \frac{w_i \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \right) \\&= \sum_{i=1}^k w_i \mathbb{V}(y_i) + \sum_{i=1}^k w_i \left(\frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{(\sum_{i=1}^k w_i)^2} \right) - 2 \sum_{i=1}^k w_i \frac{w_i \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \\&= \sum_{i=1}^k w_i \mathbb{V}(y_i) + \frac{\sum_{i=1}^k w_i \sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{(\sum_{i=1}^k w_i)^2} - 2 \sum_{i=1}^k w_i \frac{w_i \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \quad (65) \\&= \sum_{i=1}^k w_i \mathbb{V}(y_i) + \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} - 2 \sum_{i=1}^k w_i \frac{w_i \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \\&= \sum_{i=1}^k w_i \mathbb{V}(y_i) + \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} - 2 \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \\&= \sum_{i=1}^k w_i \mathbb{V}(y_i) - \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{\sum_{i=1}^k w_i}.\end{aligned}$$

Beweis (fortgeführt)

Mit $\mathbb{V}(y_i) = \tau^2 + \sigma_i^2$ und $w_i = 1/\sigma_i^2$ ergibt sich schließlich

$$\begin{aligned}
 \mathbb{E}(Q) &= \sum_{i=1}^k w_i \mathbb{V}(y_i) - \frac{\sum_{i=1}^k w_i^2 \mathbb{V}(y_i)}{\sum_{i=1}^k w_i} \\
 &= \sum_{i=1}^k w_i (\tau^2 + \sigma_i^2) - \frac{\sum_{i=1}^k w_i^2 (\tau^2 + \sigma_i^2)}{\sum_{i=1}^k w_i} \\
 &= \tau^2 \sum_{i=1}^k w_i + \sum_{i=1}^k w_i \sigma_i^2 - \tau^2 \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i} - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \\
 &= \tau^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + \left(\sum_{i=1}^k w_i \sigma_i^2 - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) \\
 &= \tau^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + \left(\sum_{i=1}^k \frac{\sigma_i^2}{\sigma_i^2} - \frac{\sum_{i=1}^k \frac{\sigma_i^2}{(\sigma_i^2)^2}}{\sum_{i=1}^k \frac{1}{\sigma_i^2}} \right) \\
 &= \tau^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + \left(\sum_{i=1}^k 1 - 1 \right) \\
 &= \tau^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + (k - 1).
 \end{aligned} \tag{66}$$

Beweis (fortgeführt)

Das Stichprobenäquivalent des Erwartungswerts der Q -Statistik ist der Mittelwert der Q -Statistik, allerdings bei nur einer Beobachtung. Gleichsetzen des Erwartungswerts von Q mit seinem Stichprobenäquivalent ergibt also folgende Bedingung für den Momentenmethodenschätzer von τ^2

$$Q = \mathbb{E}(Q) \Leftrightarrow Q = \hat{\tau}_{\text{DL}}^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + (k-1). \quad (67)$$

Auflösen nach $\hat{\tau}_{\text{DL}}^2$ ergibt dann den Momentenmethodenschätzer von τ^2

$$\hat{\tau}_{\text{DL}}^2 \left(\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i} \right) + (k-1) = Q \Leftrightarrow \hat{\tau}_{\text{DL}}^2 = \frac{Q - (k-1)}{c} \text{ mit } c := \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2 \sigma_i^2}{\sum_{i=1}^k w_i}. \quad (68)$$

(2) In Analogie zum Beweis des FEM-ML-SES-Schätzer führt die marginale Datenverteilung

$$y_i \sim N(\delta, \sigma_i^2 + \hat{\tau}_{\text{DL}}^2) \text{ unabhängig für } i = 1, \dots, n \quad (69)$$

auf die Maximum-Likelihood Bedingung

$$\sum_{i=1}^k \frac{1}{\sigma_i^2 + \hat{\tau}_{\text{DL}}^2} (y_i - \hat{\delta}_{\text{DL}}) = 0. \quad (70)$$

und Auflösen nach $\hat{\delta}_{\text{DL}}$ dann zur Aussage des Theorems. $\square$

Qualität der DerSimonian-Laird Schätzung

- Die Frequentistischen Eigenschaften der DerSimonian-Laird Schätzung sind Gegenstand aktueller Forschung.
- Vgl. Viechtbauer (2005), Jackson, Bowden, and Baker (2010), Veroniki et al. (2016), Langan et al. (2019).
- Generell lässt sich wohl festhalten, dass der DerSimonian-Laird Heterogenitätsschätzer verzerrt sein kann.
- Es wird eher die Restricted Maximum-Likelihood Schätzung von τ^2 empfohlen (vgl. Langan et al. (2019)).
- Die Restricted Maximum-Likelihood Schätzung von τ^2 ist die Default-Einstellung in *metafor*.
- Der DerSimonian-Laird Schätzer $\hat{\delta}_{DL}$ ist wohl unverzerrt (vgl. Jackson, Bowden, and Baker (2010)).
- Die Betrachtung von Metaanalysen als Linear Mixed Models wird Metaanalyse-Community-Mainstream.

Random-Effects-Modelle

Simulation mit $k := 10$, $\delta := 1$, $\tau^2 := 0.1$, $\sigma_i^2 \sim U(0, 0.2)$

```
set.seed(2)
k          = 10
delta      = 1
tausqr     = .1
sigsqr_i   = rep(NA, k)
delta_i    = rep(NA, k)
y_i        = rep(NA, k)
for(i in 1:k){
  delta_i[i] = rnorm(1, delta, sqrt(tausqr))
  sigsqr_i[i] = runif(1, .0, .2)
  y_i[i]     = rnorm(1, delta_i[i], sqrt(sigsqr_i[i]))}
w_i         = 1/sigsqr_i
delta_hat   = sum(w_i*y_i)/sum(w_i)
Q           = sum((w_i*(y_i-delta_hat)^2))
I2          = (Q - (k-1))/Q
c           = sum(w_i) - (sum(w_i**2)/sum(w_i))
tau_hat_dl  = (Q - (k-1))/c
w_i_dl      = 1/(sigsqr_i + tau_hat_dl)
delta_hat_dl = sum(w_i_dl*y_i)/sum(w_i_dl)
```

reproduzierbare Ergebnisse
Anzahl an Studien
wahre, aber unbekannte, SES
wahre, aber unbekannte, Heterogenität
Studienspezifische Varianzenarray
Latente Effektstärken Array
Empirische Effektstärken Array
Studieniterationen
latente SES
bekannte studienspezifische Varianz
empirische SES
Wichtungsparameter des FEM-ML-SES-Schätzers
FEM-ML-SES-Schätzer
Q-Statistik
I²-Statistik
c für DerSimonian-Laird
DerSimonian-Laird tau² Schätzer
Wichtungsparameter des REM-SES-Schätzers
REM-DL-SES-Schätzer Schätzer

```
Q-Statistik           : 16.0326
I2-Statistik          : 43.86
DerSimonian-Laird-Schätzer von tau2 : 0.0576
Maximum-Likelihood-SES des FEM       : 1.2404
DerSimonian-Laird-SES des REM         : 1.1598
```

Random-Effects-Modelle

Simulation mit $k := 10$, $\delta := 1$, $\tau^2 := 0.1$, $\sigma_i^2 \sim U(0, 0.2)$

Anwendung von metafor

```
library(metafor)                                # R Paket
res = rma(yi = y_i, vi = sigsq_i, method = "DL") # Daten und DerSimonian-Laird ("DL") Methode
print(res)                                       # Ergebnisausgabe
```

Random-Effects Model (k = 10; tau^2 estimator: DL)

```
tau^2 (estimated amount of total heterogeneity): 0.0576 (SE = 0.0645)
tau (square root of estimated tau^2 value):      0.2401
I^2 (total heterogeneity / total variability):    43.86%
H^2 (total variability / sampling variability):   1.78
```

Test for Heterogeneity:

Q(df = 9) = 16.0326, p-val = 0.0662

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub	
1.1598	0.1191	9.7360	<.0001	0.9263	1.3932	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theorem (Asymptotische Verteilung eines REM-SES-Schätzers)

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

$$\tilde{w}_i := \frac{1}{\tau^2 + \sigma_i^2} \text{ für } i = 1, \dots, k. \quad (71)$$

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

$$\hat{\delta} \overset{a}{\sim} N \left(\delta, \left(\sum_{i=1}^k \tilde{w}_i \right)^{-1} \right). \quad (72)$$

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.
- $\hat{\delta}$ muss nicht notwendigerweise der DerSimonian-Laird-Schätzer von δ sein.

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

Gegeben sei das Random-Effects-Modell der Metaanalyse mit bekannten studienspezifischen Varianzen σ_i^2 für $i = 1, \dots, k$, studienübergreifenden standardisierter Effektstärkevarianz (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 $\hat{\delta}$ die *asymptotische Varianz* von $\hat{\delta}$ genannt und mit

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

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

$$\hat{\mathbb{V}}_a(\hat{\delta}) = \left(\sum_{i=1}^k \hat{w}_i \right)^{-1} \quad \text{mit } \hat{w}_i := \frac{1}{\hat{\tau}^2 + \sigma_i^2} \text{ für } i = 1, \dots, k. \quad (74)$$

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

$$z_\gamma := \Phi^{-1} \left(\frac{1+\gamma}{2} \right) \quad (75)$$

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

$$\kappa := \left[\hat{\delta} - \sqrt{\mathbb{V}_a(\hat{\delta})} z_\gamma, \hat{\delta} + \sqrt{\mathbb{V}_a(\hat{\delta})} z_\gamma \right] \quad (76)$$

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

$$\mathbb{P}(\kappa \ni \delta) = \gamma. \quad (77)$$

Dazu halten wir zunächst fest, dass

$$\hat{\delta} \stackrel{a}{\sim} N\left(\delta, \mathbb{V}_a(\hat{\delta})\right). \quad (78)$$

Mit dem Theorem zur Z -Transformation gilt dann

$$Z_{\hat{\delta}} := \frac{\hat{\delta} - \delta}{\sqrt{\mathbb{V}_a(\hat{\delta})}} \sim N(0, 1). \quad (79)$$

Weiterhin gilt per Definition von z_γ , dass

$$\mathbb{P}\left(-z_\gamma \leq Z_{\hat{\delta}} \leq z_\gamma\right) = \gamma. \quad (80)$$

Beweis (fortgeführt)

Aus der Definition eines γ -Konfidenzintervalls folgt dann

$$\begin{aligned}\gamma &= \mathbb{P} \left(-z_\gamma \leq Z_{\hat{\delta}} \leq z_\gamma \right) \\&= \mathbb{P} \left(-z_\gamma \leq \frac{\hat{\delta} - \delta}{\sqrt{\mathbb{V}_a(\hat{\delta})}} \leq z_\gamma \right) \\&= \mathbb{P} \left(-z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq \hat{\delta} - \delta \leq z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \right) \\&= \mathbb{P} \left(-\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq -\delta \leq -\hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \right) \\&= \mathbb{P} \left(\hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \geq \delta \geq \hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \right) \\&= \mathbb{P} \left(\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \leq \delta \leq \hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \right) \\&= \mathbb{P} \left(\left[\hat{\delta} - z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})}, \hat{\delta} + z_\gamma \sqrt{\mathbb{V}_a(\hat{\delta})} \right] \ni \delta \right) \\&= \mathbb{P}(\kappa \ni \delta)\end{aligned} \tag{81}$$

und damit ist alles gezeigt.

Random-Effects-Modelle

Simulation mit $k := 10$, $\delta := 1$, $\tau^2 := 0.1$, $\sigma_i^2 \sim U(0, 0.2)$

```
set.seed(2)

k          = 10          # reproduzierbare Ergebnisse
delta      = 1           # Anzahl an Studien
tausqr     = .1          # wahre, aber unbekannte, SES
sigsqr_i   = rep(NA,n,k) # w.a.u. Heterogenität
delta_i     = rep(NA,n,k) # Studienspezifische Varianzenarray
y_i        = rep(NA,n,k) # Latente Effektstärken Array
for(i in 1:k){          # Empirische Effektstärken Array
  delta_i[i] = rnorm(1,delta, sqrt(tausqr)) # Studieniterationen
  sigsq_i[i] = runif(1,.0,.2)              # latente Studieneffektstärke
  y_i[i]     = rnorm(1,delta_i[i],sqrt(sigsqr_i[i])) # bekannte studienspezifische Varianz
w_i          = 1/sigsqr_i                 # empirische Effektstärke
delta_hat    = sum(w_i*y_i)/sum(w_i)      # Wichtungparameter des FEM-ML-SES-Schätzers
Q            = sum((w_i*(y_i-delta_hat)^2)) # FEM-ML-SES-Schätzer
I2           = (Q - (k-1))/Q              # Q-Statistik
c            = sum(w_i) - (sum(w_i**2)/sum(w_i)) # I^2-Statistik
tau_hat_dl   = (Q - (k-1))/c              # c für DerSimonian-Laird
w_i_dl       = 1/(sigsqr_i + tau_hat_dl)  # DerSimonian-Laird tau^2 Schätzer
delta_hat_dl = sum(w_i_dl*y_i)/sum(w_i_dl) # Wichtungparameter des REM-SES-Schätzers
delta_hat_dl_var = 1/sum(w_i_dl)          # REM-DL-SES-Schätzer
gamma        = 0.95                     # asymptotische Varianz des DL-SES-Schätzers
z_gamma      = qnorm((1 + gamma)/2)      # Konfidenzlevel
ci.lb        = delta_hat_dl-sqrt(delta_hat_dl_var)*z_gamma # z-Konfidenzkonstante
ci.ub        = delta_hat_dl+sqrt(delta_hat_dl_var)*z_gamma # untere KI Grenze
              # obere KI Grenze

estimate ci.lb ci.up
1.1598 0.9263 1.3932
```

Random-Effects-Modelle

Simulation mit $k := 10$, $\delta := 1$, $\tau^2 := 0.1$, $\sigma_i^2 \sim U(0, 0.2)$

Anwendung von metafor

```
library(metafor) # R Paket
res = rma(yi = y_i, vi = sigsqrr_i, method = "DL") # Daten und DerSimonian-Laird ("DL") Methode
print(res) # Ergebnisausgabe
```

Random-Effects Model (k = 10; tau^2 estimator: DL)

```
tau^2 (estimated amount of total heterogeneity): 0.0576 (SE = 0.0645)
tau (square root of estimated tau^2 value):      0.2401
I^2 (total heterogeneity / total variability):    43.86%
H^2 (total variability / sampling variability):    1.78
```

Test for Heterogeneity:

```
Q(df = 9) = 16.0326, p-val = 0.0662
```

Model Results:

estimate	se	zval	pval	ci.lb	ci.ub
1.1598	0.1191	9.7360	<.0001	0.9263	1.3932 ***

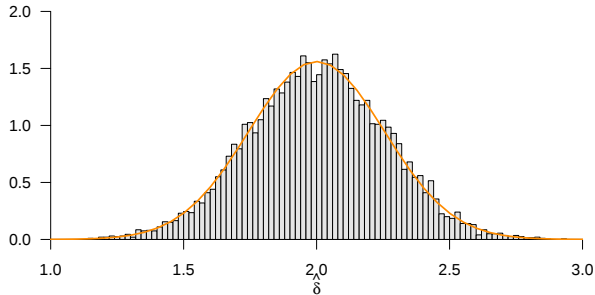
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Random-Effects-Modelle

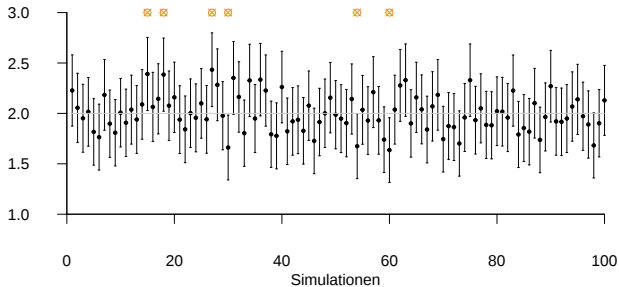
Simulation der asymptotischen Verteilung und des 95%-Konfidenzintervalls mit metafor

```
set.seed(2)                                # reproduzierbare Ergebnisse
k      = 10                                # Anzahl an Studien
delta  = 2                                  # wahre, aber unbekannte, SES
tausqr = .5                                # w.a.u. Heterogenität
sigsqr_i = runif(k,.1,.2)                  # bekannte studienspezifische Varianzen
sim     = 1e1                               # Anzahl Simulationen
delta_hat_dl = rep(NA, sim)                 # SES Array
kappa    = array(rep(NA, 2*sim), dim = c(2, sim)) # Konfidenzintervallarray
for(s in 1:sim){                           # Simulationsiterationen
  delta_i    = rep(NA, k)                  # Latente standardisierte Effektstärkenarray
  y_i        = rep(NA, k)                  # Empirische standardisierte Effektstärkenarray
  for(i in 1:k){                           # Studieniterationen
    delta_i[i] = rnorm(1, delta, sqrt(tausqr)) # latente Studieneffektstärke
    y_i[i]     = rnorm(1, delta_i[i], sqrt(sigsqr_i[i])) # empirische Effektstärke
  }
  res = rma(yi = y_i, vi = sigsqr_i, method = "DL") # metafor basierte Konfidenzintervallbestimmung
  delta_hat_dl[s] = res$b                  # Standardisierte Effektstärkeschätzer
  kappa[1, s] = res$ci.lb                  # untere Konfidenzintervallgrenze
  kappa[2, s] = res$ci.ub                  # obere Konfidenzintervallgrenze
}
```

Simulation der asymptotischen Verteilung und des 95%-Konfidenzintervalls mit `metafor`



Simulation der asymptotischen Verteilung und des 95%-Konfidenzintervalls mit `metafor`



Einführung

Effektstärkeschätzung

Fixed-Effects-Modelle

Random-Effects-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 Treatment- und Kontrollgruppen-Design 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.

Definition (Nichtzentrale t -Zufallsvariable)

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

$$p : \mathbb{R} \rightarrow \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{\nu-1}{2}} \Gamma(\frac{\nu}{2}) (\nu\pi)^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{\nu-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2} \left(t \left(\frac{\tau}{\nu}\right)^{\frac{1}{2}} - \mu\right)^2\right) d\tau. \quad (82)$$

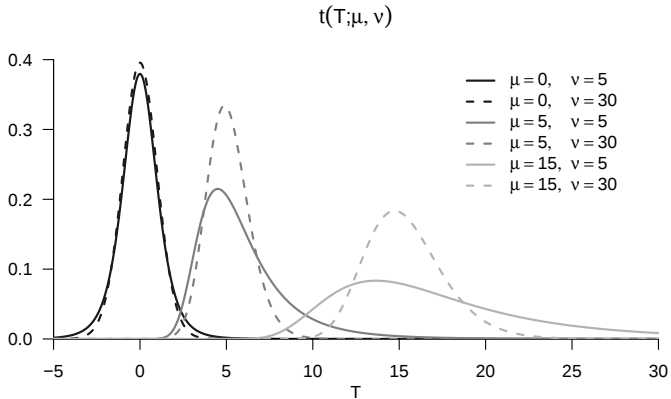
Dann sagen wir, dass T einer nichtzentralen t -Verteilung mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν unterliegt und nennen T eine *nichtzentrale t -Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν* . Wir kürzen dies mit $T \sim t(\mu, \nu)$ ab. Erwartungswert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

$$\mathbb{E}(T) = \mu \sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \quad \text{und} \quad \mathbb{V}(T) = \frac{\nu(1+\mu^2)}{\nu-2} - \frac{\mu^2\nu}{2} \left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \right)^2. \quad (83)$$

Bemerkung

- Nichtzentrale t -Zufallsvariablen sind für die Testgütefunktion des T-Tests essentiell.
- Für die Herleitung von Erwartungswert und Varianz verweisen wir auf Hogben, Pinkham, and Wilk (1961).

Wahrscheinlichkeitsdichtefunktionen spezieller nichtzentralen t -Zufallsvariablen.



- Borenstein, Michael, ed. 2009. *Introduction to Meta-Analysis*. Chichester, U.K: John Wiley & Sons.
- Cochran, William G. 1954. "Some Methods for Strengthening the Common χ^2 Tests." *Biometrics* 10 (4): 417. <https://doi.org/10.2307/3001616>.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates.
- 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>.
- DerSimonian, Rebecca, and Raghu Kacker. 2007. "Random-Effects Model for Meta-Analysis of Clinical Trials: An Update." *Contemporary Clinical Trials* 28 (2): 105–14. <https://doi.org/10.1016/j.cct.2006.04.004>.
- DerSimonian, Rebecca, and Nan Laird. 1986. "Meta-Analysis in Clinical Trials." *Controlled Clinical Trials* 7 (3): 177–88. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
- 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](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. Boca Raton: Chapman and Hall/CRC. <https://doi.org/10.1201/9781003107347>.
- Hedges, Larry V. 1981. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators," 23.

Referenzen II

- 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., 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.
- Jackson, Dan, Jack Bowden, and Rose Baker. 2010. "How Does the DerSimonian and Laird Procedure for Random Effects Meta-Analysis Compare with Its More Efficient but Harder to Compute Counterparts?" *Journal of Statistical Planning and Inference* 140 (4): 961–70. <https://doi.org/10.1016/j.jspi.2009.09.017>.
- 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>.
- Kacker, Raghu N. 2004. "Combining Information from Interlaboratory Evaluations Using a Random Effects Model." *Metrologia* 41 (3): 132–36. <https://doi.org/10.1088/0026-1394/41/3/004>.
- Kulinskaya, Elena, Michael B. Dollinger, and Kirsten Bjørkestøl. 2011a. "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>.
- . 2011b. "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>.

- Langan, Dean, Julian P. T. Higgins, Dan Jackson, Jack Bowden, Areti Angeliki Veroniki, Evangelos Kontopantelis, Wolfgang Viechtbauer, and Mark Simmonds. 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>.
- 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>.
- 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>.
- Munder, T., C. Flückiger, F. Leichenring, A. A. Abbass, M. J. Hilsenroth, P. Luyten, S. Rabung, C. Steinert, and B. E. Wampold. 2019. "Is Psychotherapy Effective? A Re-Analysis of Treatments for Depression." *Epidemiology and Psychiatric Sciences* 28 (03): 268–74. <https://doi.org/10.1017/S2045796018000355>.
- Page, Matthew J, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, 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](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. Hoboken, New Jersey: Wiley.

- Schaeuffele, Carmen, Laura E. Meine, Ava Schulz, Maxi C. Weber, Angela Moser, Christina Paersch, Dominique Recher, 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.
- Veroniki, Areti Angeliki, Dan Jackson, Wolfgang Viechtbauer, Ralf Bender, Jack Bowden, Guido Knapp, Oliver Kuss, Julian PT Higgins, Dean Langan, and Georgia Salanti. 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>.
- . 2010. "Conducting Meta-Analyses in R with the **Metafor** Package." *Journal of Statistical Software* 36 (3). <https://doi.org/10.18637/jss.v036.i03>.