Evaluation und Metaanalyse

MSc Klinische Psychologie und Psychotherapie

SoSe 2024

Prof. Dr. Dirk Ostwald

(3) Effektstärkeschätzung

Motivation

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Mittelwertbildung für Hedges' g

Selbstkontrollfragen

Motivation

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Mittelwertbildung für Hedges' g

Selbstkontrollfragen

Motivation

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)

Motivation

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 HAMD		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
Depression; N_{comp}, number of comparisons; NNT, numbers-needed-to-treat.
[a]According to the random-effects model.

Cuijpers et al. (2019)

Motivation

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

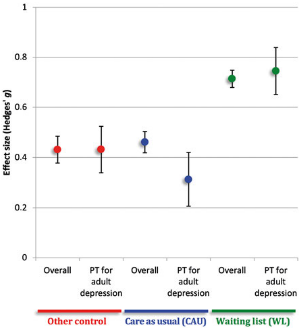
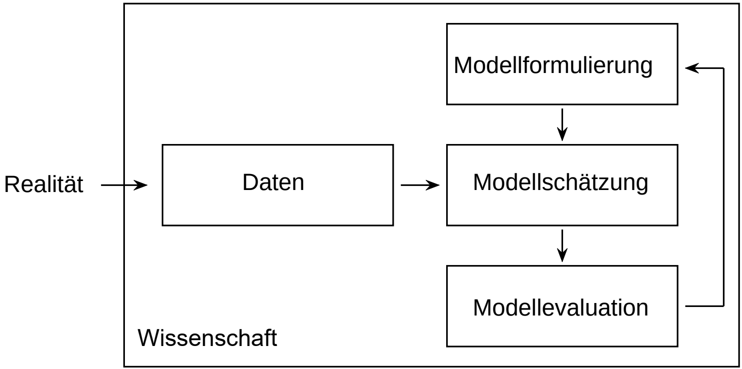
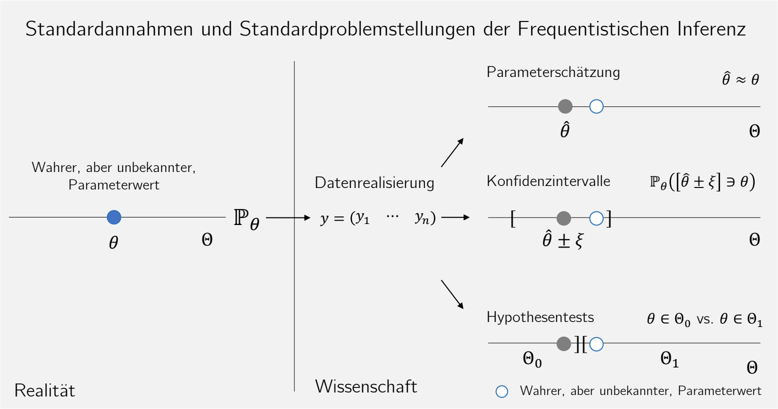


Fig. 1. Effect sizes for psychological interventions for depression. Error bars represent standard errors. PT = psychotherapy. Overall is based on all effect sizes without outliers and corrected for publication bias ($k = 146$ contrasts with WL, $k = 142$ contrasts with CAU, $k = 65$ contrasts with 'other' 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)


Modellformulierung
Realität
Daten
Modellschätzung
Modellevaluation
Wissenschaft


Standardannahmen und Standardproblemstellungen der Frequentistischen Inferenz
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Realität
Datenrealisierung
$y = (y_1 \quad \cdots \quad y_n)$
Wissenschaft
Parameterschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervalle
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentests
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Wahrer, aber unbekannter, Parameterwert

Motivation

Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein Frequentistisches Inferenzmodell mit Stichprobenvariablen $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Es wird angenommen, dass ein konkreter Datensatz $y \in \mathbb{R}^n$ eine der möglichen Realisierungen von $y = (\upsilon_1, ..., \upsilon_n)^T$ ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

Datensatz (1) : $y^{(1)} = \begin{pmatrix} y_1^{(1)} & ... & y_n^{(1)} \end{pmatrix}$ mit $\bar{y}^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

Datensatz (2) : $y^{(2)} = \begin{pmatrix} y_1^{(2)} & ... & y_n^{(2)} \end{pmatrix}$ mit $\bar{y}^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$

Datensatz (3) : $y^{(3)} = \begin{pmatrix} y_1^{(3)} & ... & y_n^{(3)} \end{pmatrix}$ mit $\bar{y}^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$

Datensatz (4) : $y^{(4)} = \begin{pmatrix} y_1^{(4)} & ... & y_n^{(4)} \end{pmatrix}$ mit $\bar{y}^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}$, $\bar{y}^{(2)}$, $\bar{y}^{(3)}$, $\bar{y}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}$?

Wenn eine statistische Methode im Sinne der Frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Standardproblemstellungen Frequentistischer Inferenz

(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für wahre, aber unbekannte, Parameterwerte oder Funktionen dieser abzugeben, typischerweise mithilfe der Daten.

(2) Konfidenzintervalle

Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der angenommenen Verteilung der Daten eine quantitative Aussage über die mit Schätzwerten assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Ziel des Hypothesentestens ist es, basierend auf der angenommenen Verteilung der Daten in einer möglichst zuverlässigen Form zu entscheiden, ob ein wahrer, aber unbekannter Parameterwert in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes liegt.

Motivation

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Mittelwertbildung für Hedges' g

Selbstkontrollfragen

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Hedges' g ist ein durch einen Multplikationsfaktor angepasstes Cohen's d, das von Hedges (1981) und Hedges & 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 Effekstärke, wird *Bias* genannt.

Im betrachtenten 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.

Effektstärkeschätzung in Einzelstudien mit Hedges' g

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

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

und die wahre, aber unbekannte, standardisierte Effekstärke ist definiert als

$$
\delta := \frac{\mu_t - \mu_c}{\sigma}. \tag{2}
$$

Bemerkungen

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

$$
y_{tj} \sim N(\mu_t, \sigma^2) \text{ für } j = 1, ..., n_t \text{ und } y_{cj} \sim N(\mu_c, \sigma^2) \text{ für } j = 1, ..., n_c \tag{3}
$$

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

Definition (Effektstärkeschätzer im Einzelstudienmodell)

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

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

wobei

$$\bar{y}_t := \frac{1}{n_t}\sum_{j=1}^{n_t} y_{tj} \text{ und } \bar{y}_c := \frac{1}{n_c}\sum_{j=1}^{n_c} y_{tc} \tag{5}$$

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

$$s_t^2 := \frac{1}{n_t - 1}\sum_{j=1}^{n_t}\left(y_{tj} - \bar{y}_t\right)^2 \text{ und } s_c^2 := \frac{1}{n_c - 1}\sum_{j=1}^{n_c}\left(y_{cj} - \bar{y}_c\right)^2 \tag{6}$$

der Treatment- und Kontrollgruppe definierte *gepoolte Stichprobenstandardabweichung*

$$s := \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}. \tag{7}$$

bezeichnet.

Bemerkungen

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

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Theorem (Eigenschaften des Effektstärkeschätzers im ESM)

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

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2}\Gamma((m-1)/2)}. \tag{8}$$

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 is also ein verzerrter Schätzer von von δ.

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Beweis

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

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

$$\mathbb{E}(\sqrt{n}d) = \sqrt{n}\delta\sqrt{\frac{m}{2}}\,\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}. \tag{9}$$

Es ergibt sich also

$$\mathbb{E}(d) = \frac{\sqrt{n}}{\sqrt{n}}\mathbb{E}(d) = \frac{1}{\sqrt{n}}\mathbb{E}(\sqrt{n}d) = \delta\sqrt{\frac{m}{2}}\,\frac{\Gamma((m-1)/2)}{\Gamma(m/2)} = \frac{1}{c_m}\,\delta. \tag{10}$$

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

$$\begin{aligned}
\mathbb{V}(\sqrt{n}d) &= \frac{m(1+(\sqrt{n}\delta)^2)}{m-2} - \frac{(\sqrt{n}\delta^2)m}{2}\left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}\right)^2 \\
&= \frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{m}{2}\left(\frac{\Gamma((m-1)/2)}{\Gamma(m/2)}\right)^2 \\
&= \frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{1}{c_m^2}.
\end{aligned} \tag{11}$$

Es ergibt sich also

$$\mathbb{V}(d) = \frac{n}{n}\mathbb{V}(d) = \frac{1}{n}\mathbb{V}(\sqrt{n}d) = \frac{1}{n}\left(\frac{m}{m-2}(1+n\delta^2) - n\delta^2\,\frac{1}{c_m^2}\right) = \frac{m}{(m-2)n}(1+n\delta^2) - \delta^2\,\frac{1}{c_m^2}. \tag{12}$$

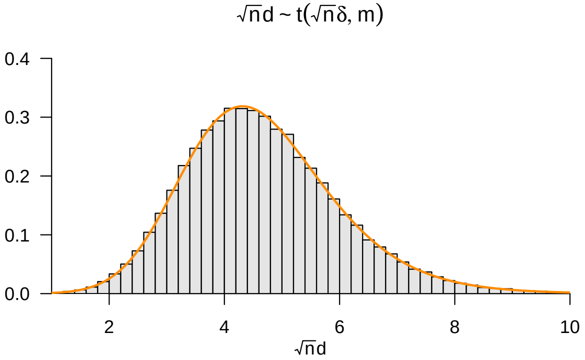
Verteilung von $\sqrt{n}d$

10^5 Effektstärkeschätzerrealisierungen 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, standardisierte Effektstärke
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)                               # Effektstärkenschätzerarrayinitialisierung
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
   ds[i] = (ybar[1]-ybar[2])/s}                        # Effektstärkenschätzerrealisierung
```

Verteilung von $\sqrt{n}d$

10^5 Effektstärkeschätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$



Theorem (Unverzerrter Effektstärkeschätzer im ESM - Hedges' g)

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

$$n := \frac{n_t n_c}{n_t + n_c}, m := n_t + n_c - 2 \text{ und } c_m = \frac{\Gamma(m/2)}{\sqrt{m/2}\,\Gamma((m-1)/2)}. \tag{13}$$

wie oben definiert. Dann ist

$$g := c_m d, \tag{14}$$

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

$$\mathbb{V}(g) < \mathbb{V}(d). \tag{15}$$

Beweis

Die Unverzerrtheit von d_u ergibt sich aus

$$\mathbb{E}(g) = \mathbb{E}(c_m d) = c_m \mathbb{E}(d) = c_m \frac{1}{c_m} \delta = \delta. \tag{16}$$

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

$$\mathbb{V}(g) = \mathbb{V}(c_m d) = c_m^2 \mathbb{V}(d) < \mathbb{V}(d). \tag{17}$$

□

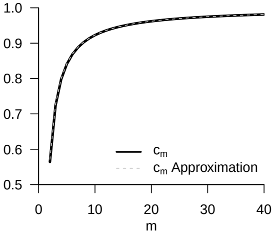
Effektstärkeschätzung in Einzelstudien mit Hedges' g

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

$$c_m \approx 1 - \frac{3}{4m-1} \tag{18}$$

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

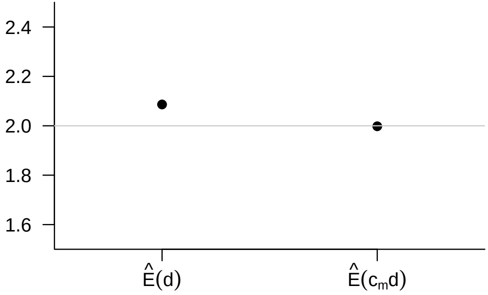


- Lin & Aloe (2021) zeigen in Appendix A1, dass $c_m \to 1$ für $m \to \infty$.
- Die Biaskorrektur ist am wichtigsten bei Gesamtgruppengrößen $n_t + n_c < 30$.
- Die Biaskorrektur ist am wichtigsten bei Treatment- und Kontrollgruppengrößen $n_t = n_c < 15$.

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Geschätzte Erwartungswerte der verzerrten und unverzerrten Effektstärkeschätzer im ESM

10^5 Effektstärkeschätzerrealisierungen mit $\delta = 2, n_t = 10, n_c = 10, m = 18, n = 5$



Theorem (Asymptotische Normalverteilung von Hedges' g)

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

$$g \overset{a}{\sim} N\left(\delta, \frac{n_t + n_c}{n_t n_c} + \frac{\delta^2}{2(n_t + n_c)}\right). \tag{19}$$

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 & Olkin (1985) auf Johnson & Welch (1940).
- Insbesondere ist die *Asymptotische Varianz* von g also gegeben durch

$$\mathbb{V}_a(g) = \frac{n_t + n_c}{n_t n_c} + \frac{\delta^2}{2(n_t + n_c)}. \tag{20}$$

- Ein häufig genutzter Schätzer für die studienspezifische Varianz von Hedges' g ist

$$\hat{\mathbb{V}}_a(g) = \frac{n_t + n_c}{n_t n_c} + \frac{g^2}{2(n_t + n_c)}. \tag{21}$$

- $\hat{\mathbb{V}}_a(g)$ ird häufig für die metananalytische Modellbildung genutzt, vgl. Einheit (4) und Lin & Aloe (2021).

Asymptotische Normalverteilung von Hedges' g

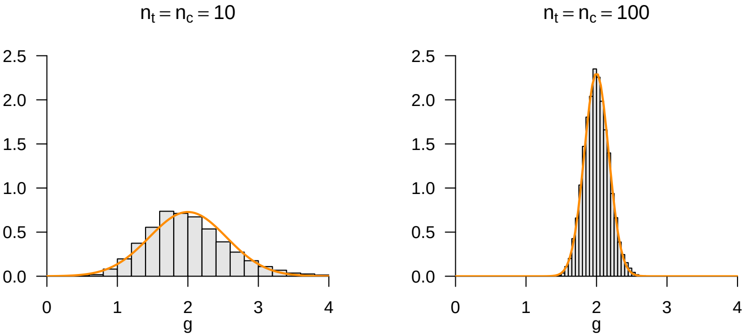
10^5 Effektstärkeschätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang

```
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, standardisierte Effektstärke
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]      = ((n_t+n_c)/(n_t*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
```

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Asymptotische Normalverteilung von Hedges' g

10^5 Effektstärkeschätzerrealisierungen mit $\delta = 2$ für kleinen und großen Stichprobenumfang



Effektstärkeschätzung in Einzelstudien mit Hedges' g

Hedges' g für den Beispieldatensatz

```
D        = read.csv("./3_Daten/Zwei_Gruppen_Design.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
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                                       # Effektstärke
d        = (bar_y_t - bar_y_c)/s                                   # Cohen's d/standardisierter Effekstärkenschätzer
m        = n_t + n_c - 2                                           # Freiheitsgradparameter
c_m      = (gamma(m/2))/(sqrt(m/2)*gamma((m-1)/2))                 # Biaskorrekturfaktor
g        = c_m*d                                                   # Hedges' g
v        = ((n_t+n_c)/(n_t*n_c))+(g^2)/(2*(n_t+n_c))               # Geschätzte Asymptotische Varianz von Hedges' g
```

```
Mittelwert Treatmentgruppe                    : 3.0667
Mittelwert Kontrollgruppe                      : -0.4
Gepoolte empirische Standardabweichung         : 3.8197
Effektstärke                                   : 3.4667
Cohen's d                                      : 0.9076
Biaskorrekturfaktor                            : 0.9729
Hedges' g                                      : 0.883
Geschätzte Asymptotische Varianz von Hedges' g  : 0.1463
```

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Hedges' g für den Beispieldatensatz mit `metafor` (Viechtbauer (2010))

```
library(metafor)
es  =  escalc("SMD", m1i = bar_y_t, m2i = bar_y_c, sd1i =  sqrt(s2_t), sd2i =  sqrt(s2_c), n1i = n_t, n2i = n_c)
print(es)
```

```
      yi     vi
1 0.8830 0.1463
```

```
Hedges' g                                     : 0.883
Geschätzte Asymptotische Varianz von Hedges' g  : 0.1463
```

Motivation

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Mittelwertbildung für Hedges' g

Selbstkontrollfragen

Mittelwertbildung für Hedges' g

Die im vorherigen Abschnitt betrachtete Entwicklung von Hedges' g bezieht sich auf eine einzelne Studie. Zu metaanalytischen Zwecken werden die dort eingeführten Werte von Hedges' g und seiner Varianz meist im Rahmen von Fixed-Effects- und insbesondere Random-Effects-Modellen, wie in Einheit (4) diskutiert, integriert.

Allerdings zeigen Hedges (1981) und Hedges & Olkin (1985) auch, dass, wiederum in einem Normalverteilungsmodell der Daten, die Mittelwertbildung von mehreren Hedges' gs über Studien hinweg zu einem unverzerrten Schätzer der wahren, aber unbekannten, Effektstärke führt. Diesem Ansatz wollen wir in diesem Abschnitt nachgehen, auch wenn in der Anwendung der integrierten Betrachtung von studienspezifischen Hedges' gs und ihrer Varianzen mithilfe der in Einheit (4) betrachteten metaanalytischen Fixed- und Random-Effects sicherlich vorherrscht.

Definition (Multiplestudienmodell)

Für Studien $i = 1, ..., k$ seien y_{itj} mit $j = 1, ..., n_{it}$ und y_{icj} mit $j = 1, ..., n_{ic}$ die Zufallsvariablen, die die primären Ergebnismaße der $i$ten Studie für die $j$te experimentelle Einheit in der Treatment und der Kontrollgruppe modellieren, respektive. Dann ist das *Multiplestudienmodell (MSM)* gegeben durch

$$
\begin{aligned}
y_{itj} &:= \mu_{it} + \varepsilon_{itj} \text{ mit } \varepsilon_{itj} \sim N(0, \sigma_i^2) \text{ für } j = 1, ..., n_{it} \\
y_{icj} &:= \mu_{ic} + \varepsilon_{icj} \text{ mit } \varepsilon_{icj} \sim N(0, \sigma_i^2) \text{ für } j = 1, ..., n_{ic}
\end{aligned}
\tag{22}
$$

wobei die wahre, aber unbekannte, standardisierte Effektstärke

$$
\delta := \frac{\mu_{it} - \mu_{ic}}{\sigma_i}
\tag{23}
$$

für alle $i = 1, ..., k$ konstant sei.

Bemerkung

- Das Multiplestudienmodell entspricht $i = 1, ..., k$ Einzelstudienmodellen mit (möglicherweise) studienspezifischen wahren, aber unbekannten, Treatment- und Kontrollgruppenerwartungswertparameter μ_{it} und μ_{ic} sowie (möglicherweise) studienspezifischen wahren, aber unbekannten, Varianzparametern σ_i^2, aber identischer wahrer, aber unbekannter, standardisierter Effektstärke δ.

Definition (Effekstärkeschätzer des Multiplestudienmodells)

Gegeben sei ein Multiplestudienmodell bestehend aus k Einzelstudienmodellen mit wahrer, aber unbekannter, Effektstärke $\delta \in \mathbb{R}$. Für $i = 1, ..., k$, seien $\bar{y}_{it}$, $\bar{y}_{ic}$ s_i die Stichprobenmittel der Treatment- und Kontrollgruppe und die gepoolte Stichprobenstandardabweichungen $i$ten Studie, respektive, und sei $m_i := n_{it} + n_{ic} - 2$. Es seien weiterhin

$$d_i := \frac{\bar{y}_{ie} - \bar{y}_{ic}}{s_i} \text{ und } g_i := c_{m_i} d_i \tag{24}$$

die verzerrten und unverzerrten Effektstärkeschätzer der $i$ten Studie, respektive. Dann heißt

$$\bar{g} := \frac{1}{k} \sum_{i=1}^{k} g_i \tag{25}$$

der *Effektstärkeschätzer des Multiplenstudienmodells* für die Effektstärke δ.

Bemerkungen

- $\bar{g}$ ist der Mittelwert der Hedges' g_i's der Einzelstudien.

Theorem (Eigenschaften des Effekstärkeschätzers im MSM)

Gegeben sei ein Multiplestudienmodell bestehen aus k Einzelstudien mit wahrer aber unbekannter, Effektstärke $\delta \in \mathbb{R}$ und es sei $\bar{g}$ der Effektstärkeschätzer für δ. Dann hat $\bar{g}$ folgende Eigenschaften:

(1) $\mathbb{E}(\bar{g}) = \delta$.

(2) $\mathbb{V}(\bar{g}) = \frac{1}{k^2} \sum_{i=1}^{k} \left(\frac{m_i c_{m_i}^2}{(m_i - 2) n_i} (1 + n_i \delta^2) - \delta^2 \right)$.

Bemerkungen

- Aussage (1) des Theorems besagt, dass $\bar{g}$ ein unverzerrter Schätzer von δ ist.

Mittelwertbildung für Hedges' g

Beweis

Für den Erwartungswert von $\bar{g}$ gilt

$$\mathbb{E}(\bar{g}) = \mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k} g_i\right) = \frac{1}{k}\left(\sum_{i=1}^{k}\mathbb{E}(g_i)\right) = \frac{1}{k}k\delta = \delta. \tag{26}$$

Für die Varianz von $\bar{g}$ gilt

$$\begin{aligned}
\mathbb{V}(\bar{g}) &= \frac{1}{k^2}\left(\sum_{i=1}^{k} c_{m_i}^2\,\mathbb{V}(d_i)\right) \\
&= \frac{1}{k^2}\left(\sum_{i=1}^{k} c_{m_i}^2\left(\frac{m_i}{(m_i-2)n_i}(1+n_i\delta^2) - \delta^2\frac{1}{c_{m_i}^2}\right)\right) \\
&= \frac{1}{k^2}\left(\sum_{i=1}^{k}\frac{m_i c_{m_i}^2}{(m_i-2)n_i}(1+n_i\delta^2) - \delta^2\right)
\end{aligned}$$

□

Motivation

Effektstärkeschätzung in Einzelstudien mit Hedges' g

Mittelwertbildung für Hedges' g

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition von Cohen's d im Treatment- und Kontrollgruppen-Design wieder.
2. Geben Sie die Definition des Einzelstudienmodells (ESMs) wieder.
3. Geben Sie die Definition des Effektstärkeschätzers im ESM wieder.
4. Geben Sie Aussage (1) und (2) des Theorems zu den Eigenschaften des Effektstärkeschätzers im ESM wieder.
5. Geben Sie das Theorem zum Unverzerrten Effektstärkeschätzer im ESM - Hedges' g wieder.
6. Geben Sie das Theorem zur Asymptotischen Normalverteilung von Hedges' g wieder.

Definition (Nichtzentrale t-Zufallsvariable)

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

$$p : \mathbb{R} \to \mathbb{R}_{>0}, t \mapsto p(t) := \frac{1}{2^{\frac{\nu-1}{2}} \Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{1}{2}}} \int_0^\infty \tau^{\frac{\nu-1}{2}} \exp\left(-\frac{\tau}{2}\right) \exp\left(-\frac{1}{2}\left(t\left(\frac{\tau}{\nu}\right)^{\frac{1}{2}} - \mu\right)^2\right) d\tau. \quad (27)$$

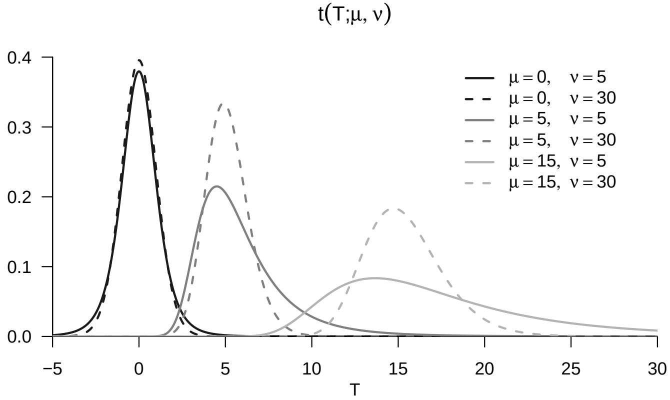
Dann sagen wir, dass T einer nichtzentralen t-Verteilung mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν unterliegt und nennen T eine *nichtzentrale t-Zufallsvariable mit Nichtzentralitätsparameter μ und Freiheitsgradparameter ν*. Wir kürzen dies mit $T \sim t(\mu, \nu)$ ab. Erwartungswert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

$$\mathbb{E}(T) = \mu\sqrt{\frac{\nu}{2}} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \text{ und } \mathbb{V}(T) = \frac{\nu(1+\mu^2)}{\nu-2} - \frac{\mu^2\nu}{2}\left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2. \quad (28)$$

Bemerkung

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

Wahrscheinlichkeitsdichtefunktionen spezieller nichtzentralen t-Zufallsvariablen.



Referenzen

Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (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

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

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.

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

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

Lin, L., & Aloe, A. M. (2021). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine*, *40*(2), 403–426. https://doi.org/10.1002/sim.8781

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P., Rabung, S., Steinert, C., & Wampold, B. E. (2019). Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiology and Psychiatric Sciences*, *28*(03), 268–274. https://doi.org/10.1017/S2045796018000355

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