Technische Universität München

PROTECT Behavioral Health Promotion & Technology Lab

Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

Effektstärken

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

Effektstärken: Hintergrund

Hintergrund (I)

- Der Begriff **"Effektstärke"** (*effect size*) ist **nicht klar definiert**; manche verstehen darunter nur **standardisierte Mittelwertsunterschiede** (Cohen's d); andere bevorzugen **breitere Definitionen**.
- Die **"engere" Definition** ist nur **schwer haltbar**, da auch Korrelationen, Odds Ratios, z-Werte etc. Richtung und Stärke eines Effekt ausdrücken können (und teils auch ineinander transformierbar sind).
- Im Kontext von RCTs werden Effektstärken genutzt, um die **Größe des Interventionseffekts** zu quantifizieren und **vergleichbar** zu machen.
- Eine praktische Schwierigkeit stellt dabei die **korrekte Berechnung von Konfidenzintervallen** dar. Für viele Effekstärken existieren **geschlossene Formeln zur Berechnung der sampling-Varianz**; diese beziehen aber nicht die **Imputationssicherheit** mit ein!

Hintergrund (II)

Eine elegante Form der Berechnung von Effekstärken und deren 95%-Konfidenzintervalle bei MI stellt die Nutzung der **"natürlichen" Interpretation des β-Gewichts der Treatmentvariable** in (G)LM dar.

Es sei $f(\beta)$ eine je nach Linkfunktion variierende Transformationsfunktion (häufig die Exponentialfunktion), und $\hat{\theta}$ die zu berechnende Effektstärke:

$$
\begin{aligned}
\hat{\theta} &= f(\hat{\beta}_{\text{treat}}) \\
\hat{\theta}_{\text{lower}} &= f\left(\hat{\beta}_{\text{treat}} - t_{\nu_{(\text{MI})},0.975} \times \text{S.E.}_{\hat{\beta}_{\text{treat}}}\right) \\
\hat{\theta}_{\text{upper}} &= f\left(\hat{\beta}_{\text{treat}} + t_{\nu_{(\text{MI})},0.975} \times \text{S.E.}_{\hat{\beta}_{\text{treat}}}\right)
\end{aligned}
$$

Viele Effektstärken lassen sich direkt aus GLM ableiten!

Verteilung	**Link-Typ**	**Linkfunktion**	**Support (y)**	**exp(β_{treat})**	`family`
Normal	Identity	μ	$\mathbb{R}$	$\log_e$(MD)	`gaussian(„identity“)`
Binomial	Logit*	$\log_e\left(\frac{\mu}{1-\mu}\right)$	$\frac{0,1,\ldots,m}{m}$	OR	`binomial(„logit“)`
	Log	$\log_e(\mu)$	$\frac{0,1,\ldots,m}{m}$	RR	`binomial(„log“)`
	Complem. Log-Log	$\log_e[-\log_e(1-\mu)]$	$\frac{0,1,\ldots,m}{m}$	HR	`binomial(„cloglog“)`
Negativ Binomial	Log	$\log_e\left(\frac{\mu}{\mu+k}\right)$	$\mathbb{N}_0$	(I)RR	log (in `MASS::glm.nb`)
Poisson	Log	$\log_e(\mu)$	$\mathbb{N}_0$	(I)RR	`poisson(„log“)`
Gamma	Inverse	$-\mu^{-1}$	$\mathbb{R}_+$	-	`gamma(„inverse“)`

Beispiel: Standardisierte Mittelwertsunterschiede (Cohen's d)

$$d = \frac{\mu_1 - \mu_1}{s_{\text{pooled}}} = \frac{\mu_1 - \mu_1}{\sqrt{\frac{(n_1-1)s_1^2+(n_0-1)s_0^2}{(n_1-1)+(n_0-1)}}}$$

$\rightarrow$ Hier kann die gepoolte Standardabweichung als "Offset" genutzt werden, um d direkt aus einem linearen Model zu berechnen.

```
# Der Wert von s_pooled wurde mit 6.32 berechnet
with(implist, lm(I(pss.1/6.32) ~ 1 + group + pss.0)) %>%
  testEstimates() -> m

m; confint(m)
```

Ergebnis: d=-0.92, 95%CI: -1.16—-0.68.

```
#>           Estimate Std.Error  t.value       df  P(>|t|)    RIV    FMI
#> (Intrcpt)  1.486     0.400    3.719  4878.106    0.000  0.075  0.071
#> group     -0.918     0.122   -7.548  4141.218    0.000  0.082  0.077
#> pss.0      0.085     0.015    5.517  6989.062    0.000  0.062  0.059
#>
#> Unadjusted hypothesis test as appropriate in larger samples.
#>
#>                   2.5 %     97.5 %
#> (Intercept)  0.70280613  2.2697069
#> group       -1.15677072 -0.6797633
#> pss.0        0.05486783  0.1153496
```

Beispiel: Odds Ratio (OR)

Für eine logistische Regression mit dummy-kodierter Treatmentvariable gilt:

$$\exp(\hat{\beta}_{\text{treat}}) = \widehat{\text{OR}}$$

Wir können also in R die **Exponentialfunktion** exp nutzen, um direkt die **OR und das Konfidenzintervall** zu berechnen:

```
with(implist, glm(ri ~ 1 + group + pss.0, binomial("logit"))) %>%
  testEstimates() -> mi.logreg

c(mi.logreg$estimates[2,1], confint(mi.logreg)[2,]) %>% exp()
```

```
#>        2.5 % 97.5 %
#> 5.958 3.366 10.546
```

Cohen's d von Within-Group-Vergleichen

Cohen's d von Within-Group-Vergleichen (I)

- Cohen's d ist bei Zwischengruppenvergleichen (z.B $\hat{\mu}_{\text{treat}} - \hat{\mu}_{\text{control}}$) eindeutig definiert.
- Strittiger ist, wie Cohen's d bei **Within-Group-Vergleichen** (z.B $\hat{\mu}_{\text{pre}} - \hat{\mu}_{\text{post}}$) zu berechnen ist, und ob dann noch von "dem" Cohen's d die Rede ist.
- Das Kernproblem liegt dabei in der **passenden Berechnung von s_{pooled}** zur Standardisierung des Mittelwertsunterschieds.

Five different "Cohen's d" statistics for within-subject designs

March 25, 2016 effect size

siehe Westfall (2016).

Die "agnostische" Methode

Eine Möglichkeit zur Berechnung von d_{within} ist:

$$d_{\text{within}} = \frac{\hat{\mu}_{\text{pre}} - \hat{\mu}_{\text{post}}}{s_{\text{pooled}}} = \frac{\hat{\mu}_{\text{pre}} - \hat{\mu}_{\text{pre}}}{\sqrt{(s^2_{\text{pre}} + s^2_{\text{post}})/2}}$$

Unter Annahme dass $n_{\text{pre}} = n_{\text{post}}$.

Hier wird schlicht "ignoriert," dass zu beiden Messzeitpunkten die gleichen Personen gemessen wurden (abhängige Stichprobe). Die Standardisierung erfolgt durch die **mittlere Standardabweichung der beiden Messzeitpunkte**.

Die "modellbasierte" Methode

"Effect size is how large the true condition effect is relative to the true amount of variability in this effect across the population. Measures of true effect and true amount of variability are ***only defined in statistical models****. They don't really exist accept within the context of a model."* (Rouder, 2016)

Nach dieser Logik ist s_{pooled} die **Standardabweichung der Residuale** des dem Design **zugrundeliegenden Modells**! Bei within-group-Vergleichen ist dies typischerweise eine *repeated-measures ANOVA* oder ein gemischtes Modell (da mehrere Messzeitpunkte pro Person).

Cave

Bei reinen prä-post-Vergleichen führt dies zu **nahezu identischen Ergebnissen** wie das "agnostische" Vorgehen. Der "modellbasierte" Ansatz ist aber **breiter einsetzbar** (z.B. repeated-measures ANOVA mit zwei Gruppen, AB-Designs, ...)

Beispiel: Berechnung von d_{within} aus rmANOVA

```
(m <- aov(pss ~ time + Error(id), data.ig)$Within)
cbind(coef(m), confint(m))
```

```
#>                     time Residuals
#> Sum of Squares  4811.095  7394.886
#> Deg. of Freedom        1       261
#>
#> Residual standard error: 5.322866
#> Estimated effects are balanced
#>                     2.5 %    97.5 %
#> time1 -8.537879 -9.828029 -7.247729
```

$$d_{\text{within}} = \frac{\texttt{time1}}{\texttt{Residual standard error}} = \frac{-8.538}{5.323} \approx -1.6$$

Referenzen

Rouder, J. (2016). *The effect-size puzzler, the answer.* https://web.archive.org/web/20200212203059/http://jeffrouder.blogspot.com/2016/03/the-effect-size-puzzler-answer.html.

Westfall, J. (2016). *Five different "Cohen's d" statistics for within-subject designs.* http://jakewestfall.org/blog/index.php/2016/03/25/five-different-cohens-d-statistics-for-within-subject-designs/.