Cook's D & Effect Size

what I've learned and how we've used them in our research / analyses

Estelle Higgins, Spring 2022

Cook's D

Cook's D (D*i*):

- Combines **leverage** + **residuals**

- Calculated by removing the i^{th} data point from the model and recalculating -- *i.e., how much do the values in the regression change without the i^{th} observation?*
  - Based on confidence ellipsoids

$$D_i = \frac{\sum_{j=1}^{n} (\widehat{Y}_j - \widehat{Y}_{j(i)})^2}{(p+1)\,\widehat{\sigma}^2}$$

- High Cook's D = strong influence on fitted data

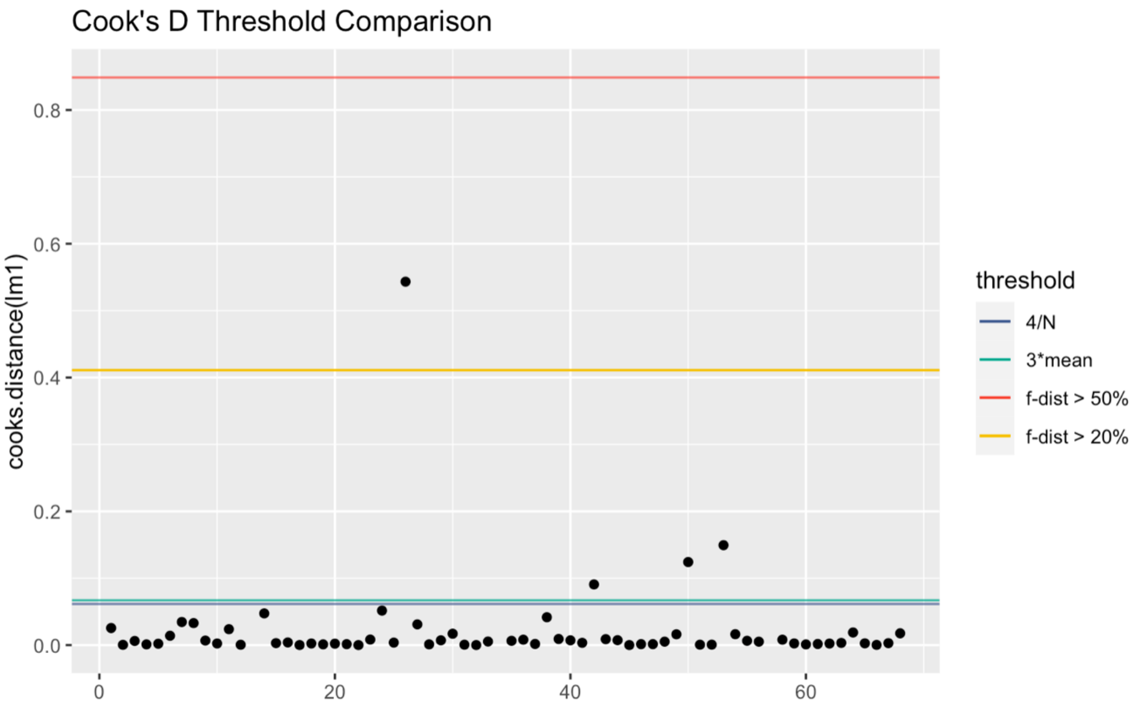
Cook's D

p = #parameters
n = #observations
k = #EVs

Cutoffs:

1. F-Distribution: D*i* > $F_{0.5}(p, n-p)$ *[Cook & Weisberg, 1982]* →
2. D*i* > 1 (large influence)
    a. some suggest D*i* > .5 (medium influence)
3. D*i* > 3 * mean(cook's d)
4. D*i* > 4/n or 4/n-k-1
5. Other F-Distribution Cutoffs?
    a. D*i* > $F_{0.1-0.2}(p, n-p)$

Cook’s D


Cook's D Threshold Comparison
cooks.distance(lm1)
0.0
0.2
0.4
0.6
0.8
0
20
40
60
threshold
4/N
3*mean
f-dist > 50%
f-dist > 20%

Cook's D

p = #parameters
n = #observations
α = percentile

Useful R functions:

- cooks.distance(*linear model*)
- plot(cooks.distance*(linear model))*
- F-Distribution quantile function: qf(α, p, n—p)
  - critical value from probability distribution, with α probability
- function I wrote / Ajay adapted: https://github.com/higgins5/lab-resources/blob/main/CooksD_funcs.Rmd

Cook's D

References / Resources:

- Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. New York: Chapman and Hall.
- Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics*, *19*(1), 15–18.
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). Applied linear statistical models (5th ed.). Boston: McGraw-Hill Irwin.
  - *See p. 403*

- UC Berkeley: https://www.stat.berkeley.edu/~spector/s133/Lr1.html
- SFU: https://www.sfu.ca/~lockhart/richard/350/97_1/examples/lecture21/lecture21.html
- Penn State: https://online.stat.psu.edu/stat462/node/173/

Effect Size

Effect Size: measure of the magnitude of experimental treatment – *importance* vs *statistical significance*

Families: correlation (*variance explained*); difference (*means*); categorical

- Many options within each family – *best practice?*

- Cohen's *d*:
    - .2 = small; .5 = medium; .8 = large
    - *Importance depends on what you're studying (e.g., 'small' effect size for mortality is invaluable)*

x = mean
s = standard deviation
n = #subj
t = treatment condition
c = comparison/control

$$d = \frac{\bar{x}_t - \bar{x}_c}{s_{pooled}} \qquad s_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Borenstein et al., 2009

Effect Size

What effect size metric should you use with repeated measures designs?

… debated. See this blog post.

- Cohen's *d* ignores repeated measures structure – limited alternatives
  - Selye et al. (2012) suggest correlation-based Cohen's f^2 – *requires R^2, complicated because of df in lmer*
  - lmer: can't convert t → d; *high df = uses measures instead of subjects*
  - Eta^2 (variance) often recommended, but not easily compared between studies

Effect Size

Hedge's g: adjusting for bias in small samples [Cohen's *d*]

- Cohen's *d* has slight bias, overestimating population standardized mean difference in small samples

- Multiply Cohen's *d* by correction factor "j" – *this is usually close to 1 unless df is very small*

$$J = 1 - \frac{3}{4df - 1}. \qquad g = J \times d,$$

df = degrees of freedom used to estimate pooled SD *(n1 + n2 – 2)*

Borenstein et al., 2009

Effect Size

Goldberg et al. (2016/ 2019):

- Compute effect size for each experimental condition (within-condition), using first minus last observation / pooled SD within full sample (*convert to hedge's g*)

$$d_{within} = \frac{M_{post} - M_{pre}}{SD_{pooled}}$$

- Then, quantify responsiveness to intervention: difference in pre-post effects *(Becker, 1988)*

$$\Delta = g_{within}^{M} - g_{within}^{c}$$

Effect Size

```{r}
CI_to_SD <- function(ci.upper,ci.lower,n){
  sd = sqrt(n) * ((ci.upper - ci.lower)/(3.92))
  print(sd)
}
# ci.upper = upper limit of confidence interval
# ci.lower = lower limit of confidence interval
# n = number of subj

calc.pooled.sd <- function(sdtreat,sdctrl,ntreat,nctrl){
  s = ((ntreat-1)*(sdtreat^2)) + ((nctrl-1)*(sdctrl)^2)
  sn = s / (ntreat + nctrl-2)
  s.pooled = sqrt(sn)
  print(s.pooled)
}
# sdtreat = standard deviation in treatment group (SD of the mean pre-post change)
# sdctrl = standard deviation in control group (SD of the mean pre-post change)
# ntreat = number in treatment group (final visit, because you use difference scores)
# nctrl = number in control group (final visit, because you use difference scores)

calc.cohens.d <- function(meantreat, meanctrl, sdpooled){
  cohensd = (meantreat - meanctrl)/sdpooled
  print(cohensd)
}
# meantreat = mean pre-post change in treatment group
# meanctrl = mean pre-post change in control group
# sdpooled = pooled standard deviation
```

Alternative?

$$d = \frac{\overline{Y}_{diff}}{S_{within}} = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{within}}. \quad (4.26)$$

This is the same formula as for independent groups (4.18). However, when we are working with independent groups the natural unit of deviation is the standard deviation within groups and so this value is typically reported (or easily imputed). By contrast, when we are working with matched groups, the natural unit of deviation is the standard deviation *of the difference scores*, and so *this* is the value that is likely to be reported. To compute d from the standard deviation of the differences we need to impute the standard deviation within groups, which would then serve as the denominator in (4.26).

Concretely, when working with a matched study, the standard deviation within groups can be imputed from the standard deviation of the difference, using

$$S_{within} = \frac{S_{diff}}{\sqrt{2(1-r)}}, \quad (4.27)$$

where r is the correlation between pairs of observations (e.g., the pretest-posttest

- Requires imputing/estimating pretest-posttest correlation "r"

Borenstein et al., 2009

Effect Size

References / Resources:

- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x
- *Introduction to Meta-Analysis* (1st ed.). (2009). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386
- Cochrane Handbook: https://training.cochrane.org/handbook/current
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.
- Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*(1), 1–11. https://doi.org/10.1037/cou0000131
- Goldberg, S. B., Tucker, R. P., Greene, P. A., Simpson, T. L., Hoyt, W. T., Kearney, D. J., & Davidson, R. J. (2019). What can we learn from randomized clinical trials about the construct validity of self-report measures of mindfulness? A meta-analysis. *Mindfulness*, *10*(5), 775–785. https://doi.org/10.1007/s12671-018-1032-y
- Selya, A., Rose, J., Dierker, L., Hedeker, D., & Mermelstein, R. (2012). A Practical Guide to Calculating Cohen's f2, a Measure of Local Effect Size, from PROC MIXED. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/article/10.3389/fpsyg.2012.00111
- Thalheimer, W., & Cook, S. (2002, August). How to calculate effect sizes from published research articles: A simplified methodology
- http://jakewestfall.org/blog/index.php/2016/03/25/five-different-cohens-d-statistics-for-within-subject-designs/