wildmeta: Cluster Wild Bootstrapping for Meta-Analysis

ESMAR 2023

Megha Joshi, James E. Pustejovsky & Pierce Cappelli

2023-02-15

Dependence

- Typical meta-analytic techniques (like meta-regression) involves the assumption that effect sizes are independent

- However, common for each primary study to yield more than one effect size or studies to be nested in some way creating dependence

- Example: Kalaian and Raudenbush (1996) data from `clubSandwich` with effects of SAT (US standardized test for college admissions) coaching

    - 47 studies, 67 effect sizes
    - Effects in same study from math and verbal sections of the test

Data SATcoaching

Show 10 entries Search:

	study	d	V	test	hrs
1	Burke (A)	0.5	0.0825	Verbal	50
2	Burke (B)	0.74	0.0855	Verbal	50
3	Coffin	0.33	0.2534	Math	18
4	Coffin	-0.23	0.2517	Verbal	18
5	Davis	0.13	0.0933	Math	15
6	Davis	0.13	0.0933	Verbal	15
7	Frankel	0.34	0.0451	Math	30
8	Frankel	0.13	0.0445	Verbal	30
9	Kintisch	0.06	0.0527	Verbal	20
10	Whitla	-0.11	0.0401	Math	10

Showing 1 to 10 of 67 entries

Handling Dependence

- Ignoring dependence leads to incorrect standard errors, incorrect inference from hypothesis tests
- Robust variance estimation (RVE) with small sample correction recommended (Tipton, 2015; Tipton and Pustejovsky, 2015)
- Such correction controls Type 1 error rates adequately but low power especially for multiple-contrast hypothesis tests (Joshi, Pustejovsky and Beretvas, 2022; Tipton and Pustejovsky, 2015)
    - Example of multiple contrast hypothesis test: To what extent do the effects of SAT coaching vary across math and verbal sections of the test?

Cluster Wild Bootstrapping

- We examined an alternative method cluster wild bootstrapping (CWB) (Joshi, Pustejovsky and Beretvas, 2022)
- Bootstrapping involves estimating unknown quantities by re-sampling from original data many times (Boos et al., 2013)
- CWB involves re-sampling residuals by multiplying them by random cluster-level weights (Cameron, Gelbach, and Miller 2008)
- CWB maintains adequate Type 1 error rates and has more power than RVE small-sample correction (Joshi, Pustejovsky and Beretvas, 2022)

CWB Algorithm

1. Fit a null model and a full model on the original data
2. Obtain residuals from the null model
3. Generate an auxiliary random variable that has mean of 0 and variance of 1 and multiply the residuals by the random variable (e.g., Rademacher weights) set to be constant within clusters (CWB)
4. Obtain new outcome scores by adding the transformed residuals to the predicted values from the null model fit on the original data
5. Re-estimate the full model with the new calculated outcome scores and obtain the test statistic
6. Repeat steps 3-5 R times. Calculate p-value:

$$p = \frac{1}{R}\sum_{r=1}^{R} I\left(F^{(r)} > F\right)$$

wildmeta

wildmeta

- The main function in the package is `Wald_test_cwb()`
- Works with meta-regressions models fit using `robumeta::robu()`, `metafor::rma.mv()` and `metafor::rma.uni()`

```
Wald_test_cwb(
  full_model,
  constraints,
  R,
  cluster = NULL,
  auxiliary_dist = "Rademacher",
  adjust = "CR0",
  type = "CR0",
  test = "Naive-F",
  seed = NULL
)
```

Data SATcoaching

Show 10 entries Search:

	study	d	V	test	hrs
1	Burke (A)	0.5	0.0825	Verbal	50
2	Burke (B)	0.74	0.0855	Verbal	50
3	Coffin	0.33	0.2534	Math	18
4	Coffin	-0.23	0.2517	Verbal	18
5	Davis	0.13	0.0933	Math	15
6	Davis	0.13	0.0933	Verbal	15
7	Frankel	0.34	0.0451	Math	30
8	Frankel	0.13	0.0445	Verbal	30
9	Kintisch	0.06	0.0527	Verbal	20
10	Whitla	-0.11	0.0401	Math	10

Showing 1 to 10 of 67 entries

robumeta Model

```r
library(wildmeta)
library(clubSandwich)
library(robumeta)

robu_model <- robu(d ~ 0 + test + hrs,
                   studynum = study,
                   var.eff.size = V,
                   small = FALSE,
                   data = SATcoaching)
```

robumeta Results

```
robu_model
```

```
## RVE: Correlated Effects Model
##
## Model: d ~ 0 + test + hrs
##
## Number of studies = 46
## Number of outcomes = 65 (min = 1 , mean = 1.41 , median = 1 , max = 2 )
## Rho = 0.8
## I.sq = 0
## Tau.sq = 0
##
##              Estimate  StdErr t-value dfs P(|t|>)  95% CI.L 95% CI.U Sig
## 1   testMath  0.05955 0.04570    1.30  43  0.1994 -0.032603  0.15171
## 2 testVerbal  0.05820 0.03769    1.54  43  0.1298 -0.017804  0.13421
## 3        hrs  0.00362 0.00206    1.75  43  0.0867 -0.000544  0.00778   *
## ---
## Signif. codes: < .01 *** < .05 ** < .10 *
## ---
```

robumeta

```r
robu_res <- Wald_test_cwb(full_model = robu_model,
                          constraints = constrain_equal(1:2),
                          R = 999,
                          seed = 20220209)

robu_res
```

```
##   Test Adjustment CR_type Statistic   R    p_val
## 1  CWB        CR0     CR0   Naive-F 999 0.975976
```

metafor

```r
library(metafor)

rma_model <- rma.mv(yi = d ~ 0 + test + hrs,
                    V = V,
                    random = ~ test | study,
                    data = SATcoaching)

rma_res <- Wald_test_cwb(full_model = rma_model,
                         constraints = constrain_equal(1:2),
                         R = 999,
                         seed = 20210314)

rma_res
```

```
##   Test Adjustment CR_type Statistic   R     p_val
## 1  CWB        CR0     CR0   Naive-F 999 0.9627767
```

Parallel Processing

```
library(future)

if (parallelly::supportsMulticore()) {
  plan(multicore)
} else {
  plan(multisession)
}

nbrOfWorkers()
```

```
## system
##      4
```

```
system.time(
  robu_res_para <- Wald_test_cwb(full_model = robu_model,
                                 constraints = constrain_equal(1:2),
                                 R = 1999,
                                 seed = 20230202)
)
```

```
##    user  system elapsed
##   2.843   0.204  40.956
```

`robumeta` Parallel

```
robu_res_para
```

```
##   Test Adjustment CR_type Statistic    R    p_val
## 1  CWB        CR0     CR0   Naive-F 1999 0.975988
```

How Many Bootstraps?

- Higher number of bootstraps better ~ precision, power
- Computationally intensive
- 1,999 or higher
- See Davidson & MacKinnon (2000) for guidance on number of bootstraps

Links

wildmeta: https://meghapsimatrix.github.io/wildmeta/index.html

My website: https://meghapsimatrix.com

THANK YOU!!

References

Boos, D. D., & others. (2003). Introduction to the bootstrap world. Statistical Science, 18(2), 168–174.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. The Review of Economics and Statistics, 47.

Davidson, R., & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps?. Econometric Reviews, 19(1), 55-68.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. Research Synthesis Methods, 1(1), 39–65.

Joshi, M., Pustejovsky, J. E., & Beretvas, S. N. (2022). Cluster wild bootstrapping to handle dependent effect sizes in meta-analysis with small number of studies. Research Synthesis Methods. Research Synthesis Methods.

References

Kalaian, H. A. & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. Psychological Methods, 1(3), 227-235. doi: 10.1037/1082-989X.1.3.227

Tipton, E., & Pustejovsky, J. E. (2015). Small-Sample Adjustments for Tests of Moderators and Model Fit Using Robust Variance Estimation in Meta-Regression. Journal of Educational and Behavioral Statistics, 40 (6), 604–634.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. Psychological Methods, 20(3), 375–393.