Exercise 5

Giulia Bertoldo

4/4/2022

Dataset description

Geeraert, Van den Noortgate, Grietens, and Onghena (2004) performed a meta-analysis to study the effects of early prevention programs for families with young children at risk for physical child abuse and neglect. Data are given in the file 'ABUSE.csv', and include the study number, the number of the outcome (most of the studies reported on multiple outcomes), the inverse of the sampling variance (W) and a variable indicating whether the outcomes were related to child abuse/neglect directly (such as reports of child abuse; X=1) versus to risk factors (such as a measure of social support; X=2).

```r
# Load packages
library(readxl)
library(metafor)
```

```
## Loading required package: Matrix
```

```
##
## Loading the 'metafor' package (version 3.0-2). For an
## introduction to the package please type: help(metafor)
```

```r
# Import data
df1 <- read_excel('data/ABUSE.xlsx')
str(df1)
```

```
## tibble [587 x 5] (S3: tbl_df/tbl/data.frame)
##  $ study  : num [1:587] 1 1 1 1 1 1 1 1 2 2 ...
##  $ outcome: num [1:587] 1 2 3 4 5 6 7 8 9 10 ...
##  $ ES     : num [1:587] -0.5743 0 -0.2095 -0.0896 -0.3005 ...
##  $ W      : num [1:587] 5.51 10.55 4.26 10.1 5.87 ...
##  $ X      : num [1:587] 1 1 1 1 1 1 1 2 1 1 ...
```

Question a

Perform a three-level meta-analysis, taking into account that effects within studies are correlated.

```r
# Add column with sampling variance
df1$var <- 1/df1$W

# Three level meta-analysis
multilevel3 <- rma.mv(yi = ES,
                      V = var,
                      random = list(~1 | study/outcome),
                      data = df1)
summary(multilevel3)
```

```
##
## Multivariate Meta-Analysis Model (k = 587; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -220.7918   441.5836   447.5836   460.7035   447.6248
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed         factor
## sigma^2.1  0.0653  0.2556     40     no          study
## sigma^2.2  0.0476  0.2182    587     no  study/outcome
##
## Test for Heterogeneity:
## Q(df = 586) = 3332.0520, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.2864  0.0450  6.3688  <.0001  0.1982  0.3745  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We have 587 outcomes and 40 studies
- Overall effect size estimate is 0.2863551 ($p = 1.9046669 \times 10^{-10}$) indicating that the program has a positive effect, on average.
- Homogeneity test: We reject the null hypothesis of homogeneity. There is significant heterogeneity. We can decompose the variance in two parts:
  - variance between studies: 0.0653335
  - variance between outcomes within study: 0.0476137 Hence, there is slightly more variance between studies than between outcomes within study.
- Addition: Use I^2 to estimate how much heterogeneity there is at each level

Question b

Compare the results with the results of an ordinary (two-level) meta-analysis, considering all effect sizes as coming from independent studies. Are you surprised?

```r
# Two level meta-analysis
multilevel2 <- rma.mv(yi = ES, V = var, data = df1)
summary(multilevel2)
```

```
##
## Multivariate Meta-Analysis Model (k = 587; method: REML)
##
##     logLik    Deviance         AIC         BIC        AICc
## -1197.6579   2395.3158   2397.3158   2401.6891   2397.3226
##
## Variance Components: none
##
## Test for Heterogeneity:
## Q(df = 586) = 3332.0520, p-val < .0001
##
## Model Results:
##
## estimate      se     zval    pval   ci.lb   ci.ub
##   0.1568  0.0045  34.7451  <.0001  0.1480  0.1657  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The pooled effect size is slightly smaller in the two-level meta-analysis: 0.1568369, but still statistically significant
- The standard error of the pooled effect size is smaller in the two level meta-analysis: 0.0045139. This can be expected because the two-level meta-analysis overestimates the amount of available information by treating outcomes as coming from independent studies.

Question c (MODERATOR: NO INTERCEPT)

Estimate the mean effect for direct measures and the mean effect on risk factors.

```r
# Testing the moderator effect of X
moderator <- rma.mv(yi = ES, V = var, data = df1,
                    random = ~1 | study/outcome,
                    mods = ~ factor(X)-1)

summary(moderator)
```

```
##
## Multivariate Meta-Analysis Model (k = 587; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -220.4295   440.8591   448.8591   466.3455   448.9280
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed         factor
## sigma^2.1  0.0669  0.2587     40     no          study
## sigma^2.2  0.0473  0.2175    587     no  study/outcome
##
## Test for Residual Heterogeneity:
## QE(df = 585) = 3319.2195, p-val < .0001
##
## Test of Moderators (coefficients 1:2):
## QM(df = 2) = 40.9815, p-val < .0001
##
## Model Results:
##
##             estimate      se    zval    pval   ci.lb   ci.ub
## factor(X)1    0.2567  0.0550  4.6646  <.0001  0.1488  0.3645  ***
## factor(X)2    0.2949  0.0461  6.4008  <.0001  0.2046  0.3852  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- *mods* $= \sim$ *factor(X)-1* : We include the moderator X as a categorical variable and we do not include an intercept in the model (-1). In this way we obtain the mean effect for each category.
- The variance between studies is still larger than the variance between outcomes
- The test of residual heterogeneity shows that after accounting for the moderator, there is still significant heterogeneity
- The mean effect for the direct effect is 0.256688 and we reject the null hypothesis that the effect is equal to zero
- The mean effect for the risk factors is 0.2949158 and we reject the null hypothesis that the effect is equal to zero

Question d: MODERATOR (WITH INTERCEPT)

Also estimate and test the difference between these means.

```
# Testing the moderator effect of X
moderator2 <- rma.mv(yi = ES, V = var, data = df1,
                     random = ~1 | study/outcome,
                     mods = ~ factor(X))

summary(moderator2)
```

```
##
## Multivariate Meta-Analysis Model (k = 587; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -220.4295   440.8591   448.8591   466.3455   448.9280
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed         factor
## sigma^2.1  0.0669  0.2587     40     no          study
## sigma^2.2  0.0473  0.2175    587     no  study/outcome
##
## Test for Residual Heterogeneity:
## QE(df = 585) = 3319.2195, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 0.9659, p-val = 0.3257
##
## Model Results:
##
##             estimate      se    zval    pval    ci.lb   ci.ub
## intrcpt       0.2567  0.0550  4.6646  <.0001   0.1488  0.3645  ***
## factor(X)2    0.0382  0.0389  0.9828  0.3257  -0.0380  0.1145
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By including the intercept, one of the categories is considered as a reference category, and an estimate is given for the contrast between the reference category and the second category. Note that the model fit is exactly the same. Both models are indeed equivalent (the parameters of one model can be derived exactly from the parameters from the other model). We see that the average effect for the risk factors is 0.0382 units higher. The difference is statistically not significant (p = .33, as obtained both from the Q-test for the moderator and from the test of the contrast).