Homework 2

Emorie Beck

January 31, 2018

Contents

Workspace

Packages

```
library(psych)
library(lme4)
library(knitr)
library(kableExtra)
library(plyr)
library(tidyverse)
```

Data

```
data_url <- "https://raw.githubusercontent.com/emoriebeck/homeworks/master/homeowrk2/HSB.csv"
dat      <- read.csv(url(data_url)) %>% tbl_df
```

Local Functions

```
table_fun <- function(model){
  fixed <- broom::tidy(model) %>% filter(group == "fixed") %>%
    select(term, estimate)
  ## add random effects ##
  rand <- VarCorr(model)[[1]]
  rand <- rand[1:nrow(rand), 1:nrow(rand)]
  colnames(rand)[colnames(rand) == "(Intercept)"] <- "Intercept"
  rownames(rand)[rownames(rand) == "(Intercept)"] <- "Intercept"
  vars <- rownames(rand)
  rand[upper.tri(rand)] <- NA
  rand <- data.frame(rand) %>% mutate(var1 = rownames(.)) %>%
    gather(key = var2, value = estimate, -var1, na.rm = T) %>%
    mutate(var1 = mapvalues(var1, vars, 0:(length(vars)-1)),
           var2 = mapvalues(var2, vars, 0:(length(vars)-1))) %>%
    filter(var1 == var2) %>%
    unite(var, var1, var2, sep = "") %>%
    mutate(var = sprintf("$\\tau_{%s}$", var))
  ## get confidence intervals ##
  CI <- data.frame(confint.merMod(model, method = "boot", nsim = 10, oldNames = F)) %>%
    mutate(term = rownames(.)) %>% setNames(c("lower", "upper", "term"))

  CI %>% filter(term == "sigma") %>%
    mutate(estimate = sigma(model),
           term = "$\\sigma^2$",
           type = "Residuals")

  ## Get ICC & R2 values ##
  ICC <- reghelper::ICC(model)
  R2 <- MuMIn::r.squaredGLMM(model)

  ## format the fixed effects
  fixed <- fixed %>% left_join(CI %>% filter(!grepl(".sig", term))) %>%
    mutate(type = "Fixed Parts")

  rand <- rand %>%
    left_join(
      CI %>% filter(grepl("sd", term)) %>%
        mutate(lower = lower^2, upper = upper^2,
               var = mapvalues(term, unique(term), 0:(length(unique(term))-1)),
               var = sprintf("$\\tau_{%s%s}$", var, var)) %>% select(-term)) %>%
    mutate(type = "Random Parts") %>% rename(term = var)

  mod_terms <- tribble(
    ~term, ~estimate, ~type,
    # "ICC", ICC, "Model Terms",
    "$R^2_m$", R2[1], "Model Terms",
    "$R^2_c$", R2[2], "Model Terms"
  )

  tab <- fixed %>%
    full_join(rand) %>%
```

```
    mutate(CI = sprintf("[%.2f, %.2f]", lower, upper)) %>%
    select(-lower, -upper) %>%
    full_join(mod_terms) %>%
    mutate(estimate = sprintf("%.2f", estimate)) %>%
    dplyr::rename(b = estimate) %>%
    select(type, everything())
  return(tab)
}
```

Question 1

Begin by testing the fully unconditional model:

$$mathach_{ij} = \beta_0 + r_{ij}$$

$$\beta_0 = \gamma_{00} + u_{0j}$$

Calculate the intraclass correlation to determine how much of the variance in math achievement resides at Level 2 (the school level).

```
lmerICC <- function(obj) {
  v <- as.data.frame(VarCorr(obj))
  v$vcov[1]/sum(v$vcov)
}

mod0 <- lmer(mathach ~ 1 + (1 | School), data = dat)
lmerICC(mod0)
```

```
## [1] 0.180353
```

Question 2

Modify the model to include student minority status (minority: 1=minority, 0=other):

$$mathach_{ij} = \beta_{0j} + \beta_{1j} minority_{ij} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

```
mod1  <- lmer(mathach ~ minority + (minority | School), data = dat)
tidy1 <- broom::tidy(mod1)
tab   <- table_fun(mod1)

options(knitr.kable.NA = '')
tab %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c("type", "Term", "b", "CI")) %>%
  collapse_rows(1)
```

type	Term	b	CI
Fixed Parts	(Intercept)	13.68	[13.39, 13.95]
	minority	-3.75	[-3.99, -3.34]
Random Parts	τ_{00}	5.62	[4.97, 6.52]
	τ_{11}	3.24	[0.96, 4.20]
Model Terms	R^2_m	0.06	
Model Terms	R^2_c	0.21	

Part A

a) Is math achievement significantly related to minority status?

Part B

(b) What is the expected (mean) level of math achievement for non-minority students?

Part C

(c) What is the expected (mean) level of math achievement for minority students?

Part D

(d) How much Level 1 variance is accounted for by this model compared to the fully unconditional model?

Question 3

Now add student sex (female: male=0, female=1) and group-centered SES to the Level 1 model:

$$mathach_{ij} = \beta_{0j} + \beta_{1j} minority_{ij} + \beta_{2j} female_{ij} + \beta_{3j}(ses_{ij} - meanses_j) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

```r
dat <- dat %>%
  mutate(GC.ses = ses - meanses)

mod2  <- lmer(mathach ~ minority + female + GC.ses + (minority + female + GC.ses | School), data = dat)
tidy2 <- broom::tidy(mod2)
tab2  <- table_fun(mod2)

tab2 %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c("type", "Term", "b", "CI")) %>%
  collapse_rows(1)
```

type	Term	b	CI
Fixed Parts	(Intercept)	14.15	[13.67, 14.62]
	minority	-3.20	[-3.56, -2.84]
	female	-1.26	[-1.61, -1.00]
	GC.ses	1.87	[1.66, 1.99]
Random Parts	τ_{00}	6.23	[4.96, 7.96]
	τ_{11}	1.98	[1.37, 2.71]
	τ_{22}	0.88	[0.41, 1.42]
	τ_{33}	0.40	[0.18, 0.77]
Model Terms	R^2_m	0.10	
Model Terms	R^2_c	0.25	

Part a

(a) Is there a significant sex difference in math achievement, controlling for minority status and SES? Yes, females have lower math achievement than males, controlling for SES and minority status, $\gamma_{20} = -1.26$. ## Part b

(b) Is the effect of student-level SES significant? Explain how the coefficient for this effect (β_{3j}) should be interpreted.
Controlling for gender and minority status, a one unit increase in SES is associated with a 1.87 increase in math achievement.
Part c

(c) What is the expected (mean) level of math achievement for minority male students with SES equal to their school average?
$Y_{ij} = \gamma_{00} + \gamma_{10} * 1 + \gamma_{20} * 0 + \gamma_{30} * 0$
$Y_{ij} = \gamma_{00} + \gamma_{10}$ $Y_{ij} = 14.15 + -3.20$ $Y_{ij} = 10.95$

Part d

(d) How much Level 1 variance is accounted for by this model compared to the fully unconditional model? 594.2% of the variance is accounted for by the fully conditional model.

Part e

(e) Does this model provide a significantly better fit than the previous model?

```r
(c1 <- anova(mod1, mod2))
```

```
## Data: dat
## Models:
## mod1: mathach ~ minority + (minority | School)
## mod2: mathach ~ minority + female + GC.ses + (minority + female + GC.ses |
## mod2:     School)
##      Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod1  6 46811 46852 -23400    46799
## mod2 15 46460 46563 -23215    46430 369.33      9  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, the deviance of the model that includes gender and SES have smaller deviance than the model that does not, $\chi^2(9) = 369.33$, $p < .001$.

Part f

(f) Explain why there are 9 degrees of freedom for the χ^2 test in the previous question.
The smaller model (that did not include gender and SES) had 6 degrees of freedom, while the larger model had 15 degrees of freedom. The deviance test we conducted is χ^2 distributed with degrees of freedom equal to the degrees of freedom of the larger minus the smaller model.

Question 4

Now add sector (1=Catholic, 0=Public) to the model:

$$mathach_{ij} = \beta_{0j} + \beta_{1j} minority_{ij} + \beta_{2j} female_{ij} + \beta_{3j}(ses_{ij} - meanses_j) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma 01 sector_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma 11 sector_j + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma 21 sector_j + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + \gamma 31 sector_j + u_{3j}$$