Homework 2

Emorie Beck

February 19

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 %>%
  mutate(female = factor(female, levels = 0:1, labels = c("Male", "Female")),
         minority = factor(minority, levels = 0:1, labels = c("Non-Minority","Minority")),
         sector = factor(sector, levels = 0:1, labels = c("Public", "Catholic")))
```

Local Functions

I don't like stargazer or the other packages that are supposed to create tables from merMod objects, so I wrote my own. The only feature I haven't implemented yet is including the ICC becuase I would need to refit the unconditional model.

```
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)
tidy0 <- broom::tidy(mod0)
lmerICC(mod0)
```

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

Question 2

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

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

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

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

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

options(knitr.kable.NA = '')
tab %>% select(-type) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c( "Term", "b", "CI")) %>%
  group_rows("Fixed", 1,2) %>%
  group_rows("Random", 3,4) %>%
  group_rows("Model Terms", 5,6) %>%
  kable_styling(latex_options = c("repeat_header"),full_width = F)
```

Term	b	CI
Fixed		
(Intercept)	13.68	[13.36, 14.10]
minorityMinority	-3.75	[-4.14, -3.64]
Random		
τ_{00}	5.62	[4.48, 6.83]
τ_{11}	3.24	[1.45, 6.11]
R_m^2	0.06	
R_c^2	0.21	

Part A

Is math achievement significantly related to minority status?
Yes, match minority status predicts math achievement, $\gamma_{10} = -3.75$, 95% CI $= [-4.14, -3.64]$.

Part B

What is the expected (mean) level of math achievement for non-minority students?
The expected mean level of math achievement for non-minority students is $\gamma_{00} = 13.68$, 95% CI $= [13.36, 14.10]$.

Part C

What is the expected (mean) level of math achievement for minority students?
The expected mean level of math achievement for non-minority students is 9.93

Part D

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

```r
d_var <- function(mod0, mod1){(sigma(mod0)^2 - sigma(mod1)^2)/sigma(mod0)^2}
sig.d <- d_var(mod0, mod1)
```

This model explains 4%‘ of more of the variance in math achievement than the fully unconditional model.

Question 3

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

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

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

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

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

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

```
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 %>% select(-type) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c( "Term", "b", "CI")) %>%
  group_rows("Fixed", 1,4) %>%
  group_rows("Random", 5,8) %>%
  group_rows("Model Terms", 9,10) %>%
  kable_styling(latex_options = c("repeat_header"),full_width = F)
```

Term	b	CI
Fixed		
(Intercept)	14.15	[13.56, 14.43]
minorityMinority	-3.20	[-3.60, -2.70]
femaleFemale	-1.26	[-1.53, -0.94]
GC.ses	1.87	[1.69, 2.12]
Random		
τ_{00}	6.23	[4.28, 6.89]
τ_{11}	1.98	[0.60, 2.96]
τ_{22}	0.88	[0.35, 1.06]
τ_{33}	0.40	[0.10, 0.62]
R^2_m	0.10	
R^2_c	0.25	

Part 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, $\gamma_{20} = -1.26, 95\% CI[-1.53, -0.94]$.

Part 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 (95% CI $[1.69, 2.12]$) increase in math achievement.

Part 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

```r
sig.d <- d_var(mod0, mod2)
```

How much Level 1 variance is accounted for by this model compared to the fully unconditional model?
9.81% 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

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:

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

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

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

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

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

```r
mod3  <- lmer(mathach ~ minority*sector + female*sector + GC.ses*sector +
                (minority + female + GC.ses | School), data = dat)
tidy3 <- broom::tidy(mod3)
tab3  <- table_fun(mod3)

tab3 %>% select(-type) %>%
  kable(., "latex", escape = F, booktabs = T,
        col.names = c( "Term", "b", "CI")) %>%
  group_rows("Fixed", 1,8) %>%
  group_rows("Random", 9,12) %>%
  group_rows("Model Terms", 13,14) %>%
  kable_styling(latex_options = c("repeat_header"),full_width = F)
```

Term	b	CI
Fixed		
(Intercept)	13.15	[12.30, 13.26]
minorityMinority	-4.14	[-4.36, -3.34]
sectorCatholic	2.31	[2.06, 3.31]
femaleFemale	-1.30	[-1.91, -0.66]
GC.ses	2.30	[2.05, 2.45]
minorityMinority:sectorCatholic	1.98	[1.42, 2.70]
sectorCatholic:femaleFemale	0.05	[-1.04, 0.52]
sectorCatholic:GC.ses	-0.98	[-1.33, -0.57]
Random		
τ_{00}	4.91	[3.28, 6.52]
τ_{11}	0.76	[0.16, 2.41]
τ_{22}	1.00	[0.15, 2.24]
τ_{33}	0.18	[0.09, 0.83]
R^2_m	0.14	
R^2_c	0.24	

Part a

Does sector significantly predict the Level 1 intercepts (β_{0j})? If so, provide an interpretation of the relationship.

Sector predicts the level 1 intercepts, $\gamma 01 = 2.31$, bootstrapped 95% CI $$. In other words, Catholic schools have higher average math achievement than public schools by 2.31.

Part b

Does sector significantly predict the Level 1 slope for minority (β_{1j})? If so, provide an interpretation of the relationship.

Sector is also a significant predictor of the level 1 slope for minority – that is, students in Catholic schools tend to have math avhievement scores that are higher on average than public schools, $\gamma 01 = 2.31$, bootstrapped 95% CI $[2.06, 3.31]$ on average.

Part c

Does sector significantly predict the Level 1 slope for ses (β_{3j})? If so, provide an interpretation of the relationship.

Sector significantly predicts the level 1 slope for ses, $\gamma 01 = -0.98$, bootstrapped 95% CI $[-1.33, -0.57]$. In other words sector moderates the relationship between SES and math. The relationship between math achievement and SES is different in Catholic and public schools, such that higher SES has a less positive effect on math achievement among students at Catholic schools than among students at public schools.

Part d

How much Level 2 variance is accounted for by this model compared to the model (Question 3) that does not contain sector? Note that this will require calculating four values, one for each of the four Level 2 equations. You will find something unusual when you do this; comment on why you think the odd result is occurring.

```
tidy2 %>% filter(grepl("sd_", term)) %>% select(term, estimate, group) %>% mutate(Model = "Q3 Model") %
full_join(tidy3 %>% filter(grepl("sd_", term)) %>% select(term, estimate, group) %>% mutate(Model = "Q4
  mutate(estimate = estimate^2) %>%
  group_by(Model) %>% mutate(estimate = estimate/sum(estimate)) %>%
  spread(key = Model, value = estimate) %>%
  mutate(diff = (`Q3 Model`-`Q4 Model`)/`Q3 Model`) %>%
  kable(., "latex", booktabs = T, digits = 2,
        col.names = c("Term", "Source", "Q3 Model", "Q4 Model", "Difference in Variance"))
```

Term	Source	Q3 Model	Q4 Model	Difference in Variance
sd_(Intercept).School	School	0.14	0.12	0.16
sd_femaleFemale.School	School	0.02	0.02	-0.20
sd_GC.ses.School	School	0.01	0.00	0.52
sd_minorityMinority.School	School	0.04	0.02	0.59
sd_Observation.Residual	Residual	0.79	0.84	-0.06

Across the models, we see that some sources of variance increase and some decrease when sector is included at Level 2. The smaller model accounted for more variance (variance decreased) in random intecepts as well as random slopes of group centered SES and minority status, while the variance accounted for by random slopes for gender increased. In other words, we are explaining *less* variance at Level 2 than we were previously. This makes sense when you interpret the τ matrix as conditional variances (conditional on the new Level 2 predictor sector). What this means is that when we include sector in the model, we reduce differences across schools, at least for mean math achievement, SES, and minority status.

Part e

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

```
(c2 <- anova(mod2, mod3))
```

```
## Data: dat
## Models:
## mod2: mathach ~ minority + female + GC.ses + (minority + female + GC.ses |
## mod2:     School)
## mod3: mathach ~ minority * sector + female * sector + GC.ses * sector +
## mod3:     (minority + female + GC.ses | School)
##      Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod2 15 46460 46563 -23215    46430
## mod3 19 46375 46506 -23168    46337 92.723      4  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes, the deviance of the model that includes sector in Level 2 is smaller deviance than the model that does not, $\chi^2(4) = 92.72$, $p < .001$.