Homework 4

Emorie Beck

February 26, 2018

Contents

Workspace

Packages

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

For this assignment, you will use the High School and Beyond data (HSB.csv) that were used for Homework 1.

Data

```
data_url <- "https://raw.githubusercontent.com/emoriebeck/homeworks/master/homework4/HSB(18).csv"
dat      <- read.csv(url(data_url)) %>% tbl_df
```

Question 1

Create four dummy codes to represent the four possible combinations of the variables, minority and female:

- minority girls (MG): minority = 1, female = 1
- minority boys (MB) minority = 1 and female = 0
- nonminority girls (NMG): minority = 0 and female = 1
- nonminority boys (NMB): minority = 0 and female = 0

```
dat <- dat %>%
  mutate(
    MG  = ifelse(female == 1 & minority == 1, 1, 0),
    MB  = ifelse(female == 0 & minority == 1, 1, 0),
    NMG = ifelse(female == 1 & minority == 0, 1, 0),
    NMB = ifelse(female == 0 & minority == 0, 1, 0)
  )
```

Question 2

Fit the following no-intercept model:

Part A

mathach ~-1 + MG + MB + NMG + NMB + (-1 + MG + MB + NMG + NMB|School).

```
source("https://raw.githubusercontent.com/emoriebeck/homeworks/master/table_fun.R")

mod2 <- lmer(mathach ~ -1 + MG + MB + NMG + NMB + (-1 + MG + MB + NMG + NMB|School), data = dat)
tab2 <- table_fun(mod2)
```

Part B

Construct weights for combining the fixed effect parameters from (a) that will test the main effect of Student Sex.

```
contra2b            <- matrix(c(-1,1,-1,1),nrow=1,ncol=4,byrow=TRUE)
row.names(contra2b) <- "female"

glht_mod2_b   <- glht(mod2, linfct = contra2b, alternative = "two.sided", rhs=0)
```

```
res_mod2_b    <- confint(glht_mod2_b, calpha = univariate_calpha())
res_mod2_b_df <- res_mod2_b$confint %>% data.frame()
```

There was a main effect of sex, $b = 2.61$, 95% CI $[1.8, 3.43]$.

Part C

Construct weights for combining the fixed effect parameters from (a) that will test the main effect of Minority Status.

```
contra2c            <- matrix(c(1,1,-1,-1),nrow=1,ncol=4,byrow=TRUE)
row.names(contra2c) <- "minority"

glht_mod2_c   <- glht(mod2, linfct = contra2c, alternative = "two.sided", rhs=0)
res_mod2_c    <- confint(glht_mod2_c, calpha = univariate_calpha())
res_mod2_c_df <- res_mod2_c$confint %>% data.frame()
```

There was a main effect of sex, $b = -7.51$, 95% CI $[-8.59, -6.43]$.

Part D

Construct weights for combining the fixed effect parameters from (a) that will test the Student Sex x Minority Status interaction.

```
contra2d            <- matrix(c(1,-1,-1,1),nrow=1,ncol=4,byrow=TRUE)
row.names(contra2d) <- "minority:female"

glht_mod2_d   <- glht(mod2, linfct = contra2d, alternative = "two.sided", rhs=0)
res_mod2_d    <- confint(glht_mod2_c, calpha = univariate_calpha())
res_mod2_d_df <- res_mod2_d$confint %>% data.frame()
```

There was no interaction between gender and minority status, $b = -7.51$, 95% CI $[-8.59, -6.43]$.

Part E

Construct weights for combining the fixed effect parameters from (a) that will test the Minority Status effect, but just for males.

```
contra2e            <- matrix(c(0,1,0,-1),nrow=1,ncol=4,byrow=TRUE)
row.names(contra2e) <- "Males: Minority v. Non-Minority"

glht_mod2_e   <- glht(mod2, linfct = contra2e, alternative = "two.sided", rhs=0)
res_mod2_e    <- confint(glht_mod2_e, calpha = univariate_calpha())
res_mod2_e_df <- res_mod2_e$confint %>% data.frame()
```

Minority and nonminority males differed, $b = -3.96$, 95% CI $[-4.62, -3.3]$.

Part F

Construct weights for combining the fixed effect parameters from (a) that will test the Student Sex effect, but just for minority students.

Table 1: Question 2G glht Results

Hypothesis	b	CI
female	2.61	[1.80, 3.43]
minority	-7.51	[-8.59, -6.43]
minority:female	0.40	[-0.35, 1.16]
Males: Minority v. Non-Minority	-3.96	[-4.62, -3.30]
Minorities: Females v. Males	-1.10	[-1.78, -0.43]

```
contra2f            <- matrix(c(1,-1,0,0),nrow=1,ncol=4,byrow=TRUE)
row.names(contra2f) <- "Minorities: Females v. Males"

glht_mod2_f   <- glht(mod2, linfct = contra2f, alternative = "two.sided", rhs=0)
res_mod2_f    <- confint(glht_mod2_f, calpha = univariate_calpha())
res_mod2_f_df <- res_mod2_f$confint %>% data.frame()
```

Male and female minorities differed in math achievement, $b = -1.1$, 95% CI $[-1.78, -0.43]$.

Part G

Use the resulting weight matrix in `glht()` to obtain the significance tests. Which effects are significant?

```
contra2 <- rbind(contra2b, contra2c, contra2d, contra2e, contra2f)

glht_mod2   <- glht(mod2, linfct = contra2, alternative = "two.sided", rhs=0)
res_mod2    <- confint(glht_mod2, calpha = univariate_calpha())
res_mod2_df <- res_mod2$confint %>% data.frame() %>% mutate(term = rownames(.))

options(knitr.kable.NA = '')
res_mod2_df %>%
  mutate(CI = sprintf("[%.2f, %.2f]", lwr, upr)) %>%
  select(term, Estimate, CI) %>%
  kable(., "latex", escape = F, booktabs = T, digits = 2,
        col.names = c( "Hypothesis", "b", "CI"),
        caption = "Question 2G glht Results") %>%
  kable_styling(full_width = F)
```

All the effects but the interaction are significant.

Question 3

Now fit this model:

Part A

mathach ~ 1 + minority + female + minority:female + (1 + minority + female + minority:female|School)

```
mod3 <- lmer(mathach ~ 1 + minority*female + (1 + minority * female | School), data = dat)
tab3 <- table_fun(mod3)
```

Table 2: Question 3B Results

	No Intercept Model		Traditional Model	
Term	b	CI	b	CI
(Intercept)			14.45	[14.16, 14.87]
female	2.61	[1.80, 3.43]	-1.51	[-1.76, -1.22]
minority	-7.51	[-8.59, -6.43]	-3.96	[-4.57, -3.24]
minority:female	0.40	[-0.35, 1.16]	0.40	[0.07, 0.92]

Part B

Do the main effects and interaction tests resemble the results from the previous analysis?

```
res_mod2_df %>%
  mutate(Estimate = sprintf("%.2f", Estimate),
         CI = sprintf("[%.2f, %.2f]", lwr, upr)) %>%
  select(term, Estimate, CI) %>%
  setNames(c("term", "b", "CI")) %>%
  filter(term %in% c("female", "minority", "minority:female")) %>%
  mutate(model = "No Intercept Model") %>%
  full_join(tab3 %>% filter(type == "Fixed Parts") %>%
              mutate(model = "Traditional Model") %>%
              select(-type)) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, model, est, sep = ".") %>%
  spread(key = tmp, value = value) %>%
  kable(., "latex", escape = F, booktabs = T, digits = 2,
        col.names = c( "Term", rep(c("b", "CI"), times = 2)),
        caption = "Question 3B Results") %>%
  kable_styling(full_width = F) %>%
  add_header_above(c(" " = 1, "No Intercept Model" = 2, "Traditional Model" = 2))
```

Part C

Construct weights for combining the fixed effect parameters from (a) that will reproduce the means for the four groups (MB, MG, NMB, and NMG).

```
contra3 <- matrix(c(
  1, 1, 0, 0,
  1, 1, 1, 1,
  1, 0, 0, 0,
  1, 0, 1, 0
), nrow=4,ncol=4,byrow=TRUE)
rownames(contra3) <- c("MB", "MG", "NMB", "NMG")
```

Part D

Use the weight matrix with `glht()`. How close are the means to those that were present in the fixed effect parameters for the model in Question 2?

Table 3: Question 3D Results

	glht Estimates		No Intercept Model	
Term	b	CI	b	CI
MB	10.49	[9.78, 11.20]	10.49	[9.78, 10.85]
MG	9.39	[8.66, 10.12]	9.39	[8.73, 10.32]
NMB	14.45	[13.98, 14.92]	14.45	[13.88, 14.63]
NMG	12.94	[12.52, 13.36]	12.94	[12.56, 13.27]

```
glht_mod3   <- glht(mod3, linfct = contra3, alternative = "two.sided", rhs=0)
res_mod3    <- confint(glht_mod3, calpha = univariate_calpha())
res_mod3_df <- res_mod3$confint %>% data.frame()

res_mod3_df %>%
  mutate(Estimate = sprintf("%.2f", Estimate),
         term = rownames(.),
         CI = sprintf("[%.2f, %.2f]", lwr, upr)) %>%
  select(term, Estimate, CI) %>%
  setNames(c("term", "b", "CI")) %>%
  mutate(model = "ghlt Estimates") %>%
  full_join(
    tab2 %>% filter(type == "Fixed Parts") %>%
             mutate(model = "No Intercept Model") %>%
             select(-type)
  ) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, model, est, sep = ".") %>%
  spread(key = tmp, value = value) %>%
  kable(., "latex", escape = F, booktabs = T, digits = 2,
        col.names = c( "Term", rep(c("b", "CI"), times = 2)),
        caption = "Question 3D Results") %>%
  kable_styling(full_width = F) %>%
  add_header_above(c(" " = 1, "glht Estimates" = 2, "No Intercept Model" = 2))
```

The means are nearly identical but have different confidence intervals.

Question 4

Fit the following model:

Part A

mathach ~ 1 + MB + NMG + NMB + (1 + MB + NMG + NMB|School).

```
mod4 <- lmer(mathach ~ 1 + MB + NMG + NMB + (1 + MB + NMG + NMB|School), data = dat)
tab4 <- table_fun(mod4)
```

Part B

What do the fixed effect parameters mean in this analysis?

The intercept (γ_{00}) represents the mean of the minority girls. The MB (γ_{10}) represenets the difference between minority girls and boys – minority boys have higher math achivement than minority girls on average. The NMG term (γ_{20}) represents the difference between minority and non-minority girls – non-minority girls have higher math achivement than minority girls on average. The NMB term (γ_{30}) represents the difference between minority girls and boys – minority boys have higher math achivement than minority girls on average.

Part C

Construct the weight matrix that will reproduce the means for the four groups (MB, MG, NMB, and NMG).

```
contra4c <- matrix(c(
   1, 1, 0, 0,
   1, 0, 0, 0,
   1, 0, 1, 0,
   1, 0, 0, 1
), nrow=4,ncol=4,byrow=TRUE)
rownames(contra4c) <- c("MB", "MG", "NMB", "NMG")
```

Part D

Construct the weight matrix necessary to reproduce the tests of the two main effects and the interaction.

```
contra4d <- matrix(c(
    0,  1, -1, 1, # b0 + b0 + b2 - (b0 + b1 + b0 + b3) --> b2 = b1 + b3
    0,  1, -1,-1, # b0 + b0 + b1 - (b0 + b2 + b0 + b3) --> b1 = b2 + b3
    0, -1, -1, 1  # b0 + b0 + b3 - (b0 + b1 + b0 + b2) --> b2 = b1 + b2
), nrow=3,ncol=4,byrow=TRUE)
rownames(contra4d) <- c("female", "minority", "minority:female")
```

Part E

Use the matrix in glht( ) and compare the results to those obtained in Question 2.

```
contra4 <- rbind(contra4c, contra4d)

glht_mod4   <- glht(mod4, linfct = contra4, alternative = "two.sided", rhs=0)
res_mod4    <- confint(glht_mod4, calpha = univariate_calpha())
res_mod4_df <- res_mod4$confint %>% data.frame() %>% mutate(term = row.names(.))

res_mod2_df %>%
  mutate(Estimate = sprintf("%.2f", Estimate),
         CI = sprintf("[%.2f, %.2f]", lwr, upr)) %>%
  select(term, Estimate, CI) %>%
  setNames(c("term", "b", "CI")) %>%
  filter(term %in% c("female", "minority", "minority:female")) %>%
  mutate(model = "Q2 Model") %>%
  full_join(
    tab2 %>% filter(type == "Fixed Parts") %>%
               mutate(model = "Q2 Model") %>%
```

Table 4: Question 4D Results

	Q2 Model		Q4 Model	
Term	b	CI	b	CI
female	2.61	[1.80, 3.43]	2.61	[1.80, 3.43]
MB	10.49	[9.78, 10.85]	10.49	[9.78, 11.20]
MG	9.39	[8.73, 10.32]	9.39	[8.66, 10.12]
minority	-7.51	[-8.59, -6.43]	-7.51	[-8.59, -6.43]
minority:female	0.40	[-0.35, 1.16]	0.40	[-0.35, 1.16]
NMB	14.45	[13.88, 14.63]	12.94	[12.52, 13.36]
NMG	12.94	[12.56, 13.27]	14.45	[13.98, 14.92]

```
            select(-type)
) %>%
full_join(
  res_mod4_df %>%
    mutate(model = "Q4 Model",
           Estimate = sprintf("%.2f", Estimate),
           CI = sprintf("[%.2f, %.2f]", lwr, upr)) %>%
    select(term, Estimate, CI, model) %>%
    setNames(c("term", "b", "CI", "model"))
) %>%
gather(key = est, value = value, b, CI) %>%
unite(tmp, model, est, sep = ".") %>%
spread(key = tmp, value = value) %>%
kable(., "latex", escape = F, booktabs = T, digits = 2,
      col.names = c( "Term", rep(c("b", "CI"), times = 2)),
      caption = "Question 4D Results") %>%
kable_styling(full_width = F) %>%
add_header_above(c(" " = 1, "Q2 Model" = 2, "Q4 Model" = 2))
```

Question 5

Compare the fit for the three models using the `anova( )` function. If you named the fit objects: Fit_1, Fit_2, and Fit_3, then use `anova(Fit_1, Fit_2, Fit_3)`. The result should not surprise you; why not?

```
anova(mod2, mod3, mod4)
```

```
## Data: dat
## Models:
## mod2: mathach ~ -1 + MG + MB + NMG + NMB + (-1 + MG + MB + NMG + NMB |
## mod2:     School)
## mod3: mathach ~ 1 + minority * female + (1 + minority * female | School)
## mod4: mathach ~ 1 + MB + NMG + NMB + (1 + MB + NMG + NMB | School)
##      Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mod2 15 46754 46857 -23362    46724
## mod3 15 46754 46857 -23362    46724     0      0     <2e-16 ***
## mod4 15 46754 46857 -23362    46724     0      0          1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In terms of fit statistics, these three models are identical because they are testing different linear combinations of the same data and hypotheses.