# Homework 3

*Emorie Beck*

*February 19, 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/homework3/popularity_2(1).csv
dat      <- read.csv(url(data_url)) %>% tbl_df %>%
  mutate(sex_c = as.numeric(scale(sex, scale = F)),
         sex = factor(sex, levels = 0:1, labels = c("Male", "Female")))
```

The popularity data include the teachers' ratings of the popularity of the students in their classrooms. For this assignment, you will test several hypotheses about the relationship between peer and teacher popularity ratings, using the file: popularity_2.csv. This file contains the following variables:

sex: Male=0, Female=1 texp: Teacher's experience in years popular: Average of other students' popularity ratings for the student, ranging from 0 to 10 popteach: Teacher's popularity rating for the student, ranging from 0 to 10

Conduct the following analyses:

# Question 1

1. First, modify the data in the following ways:

## Part A

Create the class average for teacher popularity ratings (call it mean_pop_t).

```
dat <- dat %>% group_by(class) %>% mutate(mean_pop_t = mean(popteach, na.rm = T)) %>% ungroup()
```

## Part B

Create a class-centered teacher popularity rating (call it popteach_c).

```
dat <- dat %>% mutate(popteach_c = popteach - mean_pop_t)
```

## Part C

Grand mean center the results of the calculation in (a). Call it gmc_mean_pop_t.

```
dat <- dat %>% mutate(gmc_mean_pop_t = mean_pop_t - mean(popteach, na.rm = T)) %>% ungroup()
```

## Part D

Create a grand-mean-centered version of texp called gmc_texp. Include the head( ) function to list the first 30 lines of the data frame to show the results of the required calculations.

```
dat <- dat %>% mutate(gmc_texp = as.numeric(scale(texp, scale = F)))

head(dat, n = 30)

## # A tibble: 30 x 11
##    pupil class sex      texp popular popteach  sex_c mean_pop_t popteach_c
##    <int> <int> <fct>   <int>   <dbl>   <int>  <dbl>      <dbl>      <dbl>
```

```
## 1     1     1 Female    24    6.30       6  0.495      4.90     1.10
## 2     2     1 Male      24    4.90       5 -0.505      4.90     0.1000
## 3     3     1 Female    24    5.30       6  0.495      4.90     1.10
## 4     4     1 Female    24    4.70       5  0.495      4.90     0.1000
## 5     5     1 Female    24    6.00       6  0.495      4.90     1.10
## 6     6     1 Male      24    4.70       5 -0.505      4.90     0.1000
## 7     7     1 Male      24    5.90       5 -0.505      4.90     0.1000
## 8     8     1 Male      24    4.20       5 -0.505      4.90     0.1000
## 9     9     1 Male      24    5.20       5 -0.505      4.90     0.1000
## 10    10    1 Male      24    3.90       3 -0.505      4.90    -1.90
## # ... with 20 more rows, and 2 more variables: gmc_mean_pop_t <dbl>,
## #   gmc_texp <dbl>
```

## Question 2

Then begin with the fully unconditional model for peer-rated popularity:

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

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

Compare it to the following model that includes popteach_c as a Level 1 predictor:

$$popular_{ij} = \beta_0 + \beta_1 * popteach\_c_{ij} + r_{ij}$$

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

$$\beta_1 = \gamma_{10} + u_{1j}$$

```r
mod0 <- lmer(popular ~ 1            + (1|class), data = dat)
mod1 <- lmer(popular ~ popteach_c + (popteach_c|class), data = dat)

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

tab0 <- table_fun(mod0) %>% mutate(model = "unconditional")
tab1 <- table_fun(mod1) %>% mutate(model = "conditional")

options(knitr.kable.NA = '')
tab0 %>% full_join(tab1) %>%
  mutate(term = str_replace_all(term, "\\_", "\\\\_")) %>%
  gather(key = est, value = value, b, CI) %>%
  unite(tmp, model, est, sep = ".") %>%
  mutate(type = factor(type, levels = c("Fixed Parts", "Random Parts", "Model Terms"))) %>%
  spread(key = tmp, value = value) %>%
  select(term, contains("uncond"), contains("cond")) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("Term", "b", "CI", "b", "CI")) %>%
  kable_styling(full_width = F) %>%
  column_spec(2:5, width = "2cm") %>%
  group_rows("Fixed Parts",1,2) %>%
  group_rows("Random Parts",3,4) %>%
  group_rows("Model Terms",5,6) %>%
  add_header_above(c(" " = 1, "Unconditional" = 2, "Conditional" = 2))
```

3

|       | Unconditional | | Conditional | |
|-------|---------------|---|-------------|---|
| Term  | b | CI | b | CI |
| **Fixed Parts** | | | | |
|    (Intercept) | 5.08 | [5.00, 5.20] | 5.08 | [4.96, 5.18] |
|    popteach_c | | | 0.73 | [0.71, 0.75] |
| **Random Parts** | | | | |
|    $\tau\_00$ | 0.70 | [0.59, 0.93] | 0.74 | [0.60, 0.92] |
|    $\tau\_11$ | | | 0.00 | [0.00, 0.01] |
| $R^2\_c$ | 0.36 | | 0.73 | |
| $R^2\_m$ | 0.00 | | 0.34 | |

## Part A

Are peer-reports and teacher-reports significantly related? Yes, a one unit increase in teacher ratings of popularity is associated with a $0.73$, $95\%$ CI $[0.71, 0.75]$.

## Part B

How much variance in the Level 1 model is accounted for by teacher popularity ratings?

```
vexpL1 <- (sigma(mod0)^2 - sigma(mod1)^2)/sigma(mod0)^2
```

$58\%$ of the variance in the Level 1 model is accounted for by teacher popularity ratings.

# Question 3

Now test two hypotheses about the ability of teachers as judges of their students' popularity.

To test these hypotheses, analyze the following model:

$$popular_{ij} = \beta_0 + \beta_1 * popteach\_c_{ij} + r_{ij}$$
$$\beta_0 = \gamma_{00} + \gamma_{01} * gmc\_texp_j + \gamma_{02} * gmc\_mean\_pop\_t_j + u_{0j}$$

$$\beta_1 = \gamma_{10} + \gamma_{11} * gmc\_texp_j + \gamma_{12} * gmc\_mean\_pop\_t_j + u_{1j}$$

```
mod2 <- lmer(popular ~ popteach_c + gmc_texp * popteach_c + gmc_mean_pop_t * popteach_c + (popteach_c |
tab2 <- table_fun(mod2)

tab2 %>% select(-type) %>%
  mutate(term = str_replace_all(term, "\\_", "\\\\_")) %>%
  kable(., "latex", booktabs = T, escape = F,
        col.names = c("Term", "b", "CI")) %>%
  kable_styling(full_width = F) %>%
  column_spec(2:3, width = "2cm") %>%
  group_rows("Fixed Parts",1,6) %>%
  group_rows("Random Parts",7,8) %>%
  group_rows("Model Terms",9,10) %>%
  add_header_above(c(" " = 1, "Popularity" = 2))
```

|  | Popularity | |
| Term | b | CI |
| --- | --- | --- |
| **Fixed Parts** | | |
| (Intercept) | 5.08 | [5.05, 5.10] |
| popteach_c | 0.72 | [0.70, 0.75] |
| gmc_texp | 0.00 | [-0.00, 0.01] |
| gmc_mean_pop_t | 0.98 | [0.94, 1.03] |
| popteach_c:gmc_texp | -0.01 | [-0.01, -0.00] |
| popteach_c:gmc_mean_pop_t | -0.01 | [-0.04, 0.02] |
| **Random Parts** | | |
| $\tau\_00$ | 0.00 | [0.00, 0.00] |
| $\tau\_11$ | 0.00 | [0.00, 0.01] |
| $R^2\_m$ | 0.73 | |
| $R^2\_c$ | 0.73 | |

## Part A

Teachers with more experience are better judges of student popularity. Teachers with more experience are worse judges of popularity, $\gamma_{11} = -0.01$, 95% CI $[-0.01, -0.00]$.

## Part B

It is more difficult to judge student popularity when the mean popularity of the class (as rated by teachers) is high than when it is low.
It is not more difficult to judge student popularity when the mean popularity is high v. low, $\gamma_{12} = -0.01$, 95% CI $[-0.04, 0.02]$.

## Part C

Illustrate the cross-level interaction, $\gamma_{11}$.

```
# get SD from unconditional model
mod.pt0 <- lmer(popteach_c ~ 1 + (1 | class), data = dat)
sd.pt0  <- sigma(mod.pt0)
sd.te   <- sd(dat$texp)


mod.pt1 <- lmer(popteach_c ~ 1 + gmc_texp + gmc_mean_pop_t + (1 | class), data = dat)

te.inc.1 <- sd.te * fixef(mod.pt1)["gmc_texp"]

tibble(
  class=1,
  popteach_c = seq(-1*te.inc.1 - sd.pt0,-te.inc.1 + sd.pt0, by = sd.pt0),
  gmc_texp = -sd.te,
  gmc_mean_pop_t = mean(dat$gmc_mean_pop_t),
  level = "-1 SD"
  ) %>% full_join(
tibble(
  class=1,
  popteach_c = seq(-sd.pt0,sd.pt0, by = sd.pt0),
```
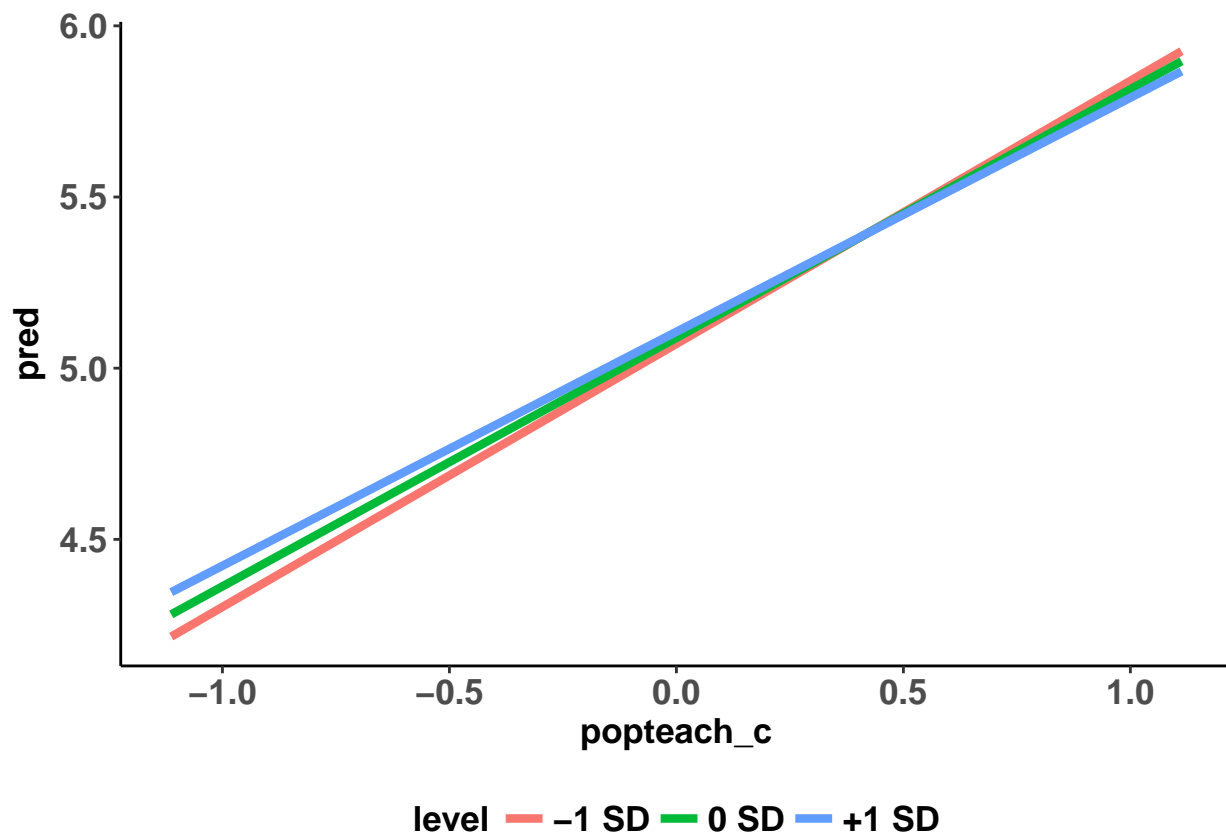
```
  gmc_texp = 0,
  gmc_mean_pop_t = mean(dat$gmc_mean_pop_t),
  level = "0 SD"
)) %>% full_join(
tibble(
  class=1,
  popteach_c = seq(te.inc.1 - sd.pt0,te.inc.1 + sd.pt0, by = sd.pt0),
  gmc_texp = sd.te,
  gmc_mean_pop_t = mean(dat$gmc_mean_pop_t),
  level = "+1 SD"
)) %>%
  mutate(pred = predict(mod2, .),
         level = factor(level, c("-1 SD", "0 SD", "+1 SD"))) %>%
  ggplot(aes(x = popteach_c, y = pred, color = level)) +
  geom_line(size = 1.5) +
  theme_classic() +
  theme(legend.position = "bottom",
        axis.text = element_text(face = "bold", size = rel(1.2)),
        axis.title = element_text(face = "bold", size = rel(1.2)),
        legend.text = element_text(face = "bold", size = rel(1.2)),
        legend.title = element_text(face = "bold", size = rel(1.2)))
```



## Part D

Illustrate the cross-level interaction, $\gamma_{12}$.

```r
# get SD from unconditional model
mod.pt0 <- lmer(popteach_c ~ 1 + (1 | class), data = dat)
sd.pt0  <- sigma(mod.pt0)
sd.mtp  <- sd(dat$gmc_mean_pop_t)

mod.pt1 <- lmer(popteach_c ~ 1 + gmc_texp + gmc_mean_pop_t + (1 | class), data = dat)

te.inc.1 <- sd.mtp * fixef(mod.pt1)["gmc_mean_pop_t"]

tibble(
  class=1,
  popteach_c = seq(te.inc.1 - sd.pt0,te.inc.1 + sd.pt0, by = sd.pt0),
  gmc_texp = mean(dat$gmc_texp),
  gmc_mean_pop_t = -sd.mtp,
  level = "-1 SD"
  ) %>% full_join(
tibble(
  class=1,
  popteach_c = seq(-sd.pt0, sd.pt0, by = sd.pt0),
  gmc_texp = mean(dat$gmc_texp),
  gmc_mean_pop_t = 0,
  level = "0 SD"
)) %>% full_join(
tibble(
  class=1,
  popteach_c = seq(-te.inc.1 - sd.pt0,-te.inc.1 + sd.pt0, by = sd.pt0),
  gmc_texp = mean(dat$gmc_texp),
  gmc_mean_pop_t = sd.mtp,
  level = "+1 SD"
)) %>%
  mutate(pred = predict(mod2, .),
         level = factor(level, c("-1 SD", "0 SD", "+1 SD"))) %>%
  ggplot(aes(x = popteach_c, y = pred, color = level)) +
  geom_line(size = 1.5) +
  theme_classic() +
  theme(legend.position = "bottom",
        axis.text = element_text(face = "bold", size = rel(1.2)),
        axis.title = element_text(face = "bold", size = rel(1.2)),
        legend.text = element_text(face = "bold", size = rel(1.2)),
        legend.title = element_text(face = "bold", size = rel(1.2)))
```
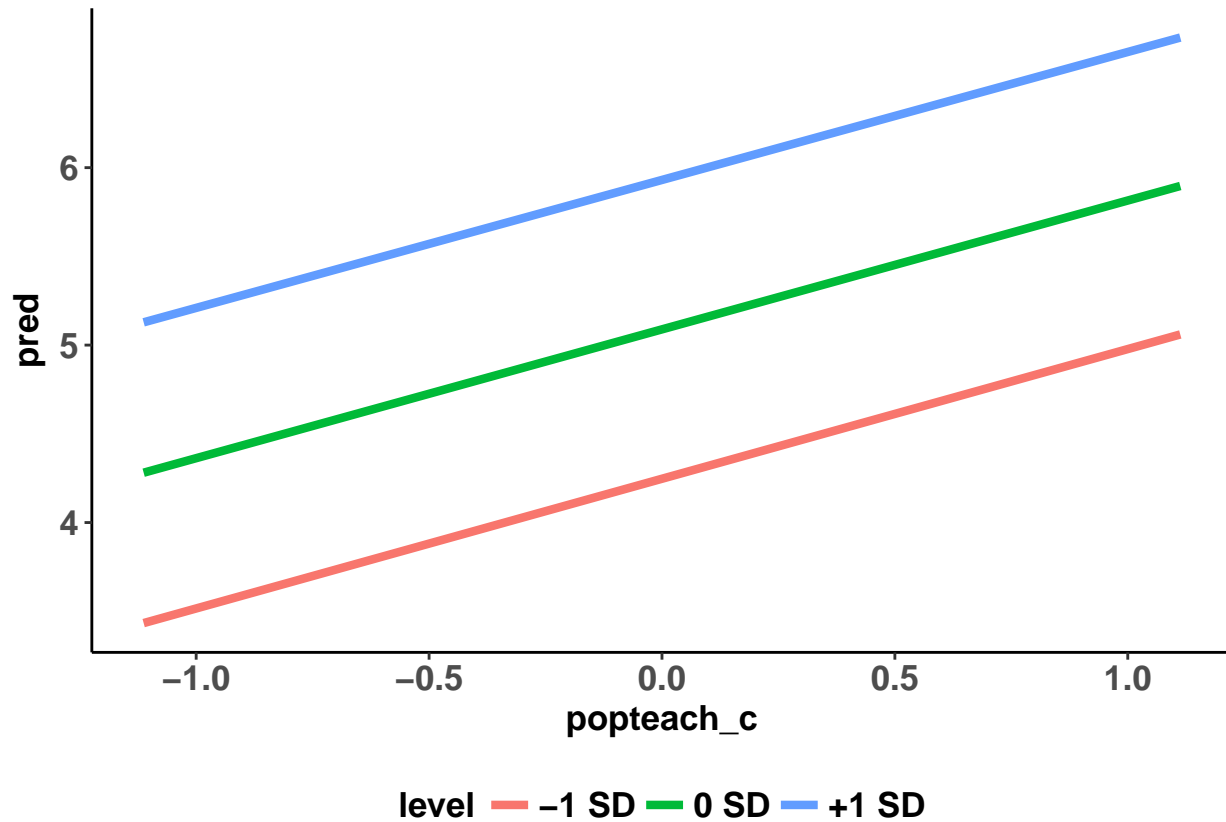
For (c) and (d) illustrate the interactions using a method of your choosing. If you treat the continuous predictors as discrete for the purpose of graphing, define "low" and "high" as + 1 standard deviation from the grand mean. Remember that the simple standard deviation taken across the entire data set may not be the best indicator of variability for Level 1 predictors. Note also that displaying the interactions most accurately involves accounting for the possible relationship between teacher popularity ratings and teacher experience and the possible relationship between teacher popularity ratings and classroom mean teacher-rated popularity. A separate document (Additional advice on figures for Homework 3) describes how to approach this if you want to give it a go. Simpler figures (without any such adjustments) are acceptable for this homework assignment.

## Question 4

Now test the hypothesis that it is more difficult for teachers to judge the popularity of boys than to judge the popularity of girls. If true, this would manifest itself in the form of a Teacher Rating x Sex of Student interaction—the relationship of teacher ratings to peer ratings would be different for male and female students. Test the following model:

$$popular_{ij} = \beta_0 + \beta_1 * sex\_c_{ij} + \beta_2 * popteach\_c_{ij} + \beta_3 * (sex\_c_{ij} : popteach\_c_{ij}) + r_{ij}$$

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

$$\beta_1 = \gamma_{10} + u_{1j}$$

$$\beta_2 = \gamma_{20} + u_{2j}$$

$$\beta_3 = \gamma_{30} + u_{3j}$$

```
mod3 <- lmer(popular ~ sex_c * popteach_c + (sex_c*popteach_c|class), data = dat)
tab3 <- table_fun(mod3)
```

## Part A

Compared to the fully unconditional model, how much Level 1 variance is accounted for by all Level 1 predictors?

```
vexpL1.4a <- (sigma(mod0)^2 - sigma(mod3)^2)/sigma(mod0)^2
```

62% of the variance in the Level 1 model is accounted for by all predictors.

## Part B

How much is attributable to the addition of the sex main effect and the interaction?

```
vexpL1.4b <- (sigma(mod1)^2 - sigma(mod3)^2)/sigma(mod1)^2
```

11% of the variance in the Level 1 model is accounted for by all predictors.

## Part C

Is there support for the hypothesis? We cannot conclude that it is more difficult for teachers to judge the popularity of boys than to judge the popularity of girls, $\gamma_{30} = 0.03$, 95% CI $[0.01, 0.12]$.

## Part D

Illustrate the interaction in this model.

```
crossing(sex_c = unique(dat$sex_c),
         popteach_c = seq(min(dat$popteach_c,na.rm =T), max(dat$popteach_c, na.rm = T), .1)) %>%
  mutate(pred = predict(mod3, newdata = ., re.form = NA)) %>%
  mutate(sex = mapvalues(sex_c, unique(sex_c), c("Male", "Female"))) %>%
  ggplot(aes(x = popteach_c, y = pred, color = sex)) +
  geom_line(size = 1.5) +
  theme_classic() +
  theme(legend.position = "bottom",
        axis.text = element_text(face = "bold", size = rel(1.2)),
        axis.title = element_text(face = "bold", size = rel(1.2)),
        legend.text = element_text(face = "bold", size = rel(1.2)),
        legend.title = element_text(face = "bold", size = rel(1.2)))
```

**sex** — Female — Male
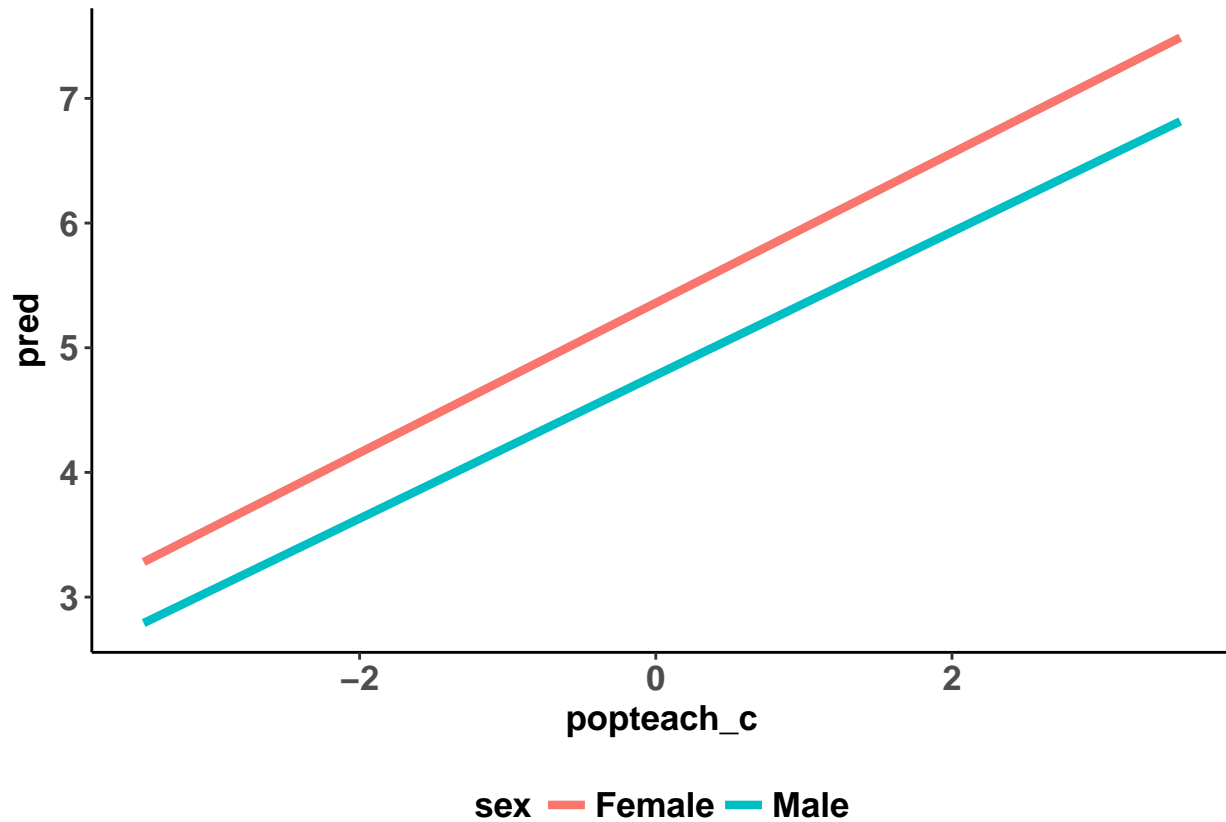
## Question 5

The parameter estimates and residuals can provide some very interesting information for these data. See if you can figure out how to determine answers to the following questions.

In general, once you have fit a model with lmer( ), the following functions can be used to get various bits of information:

`resid(fit_object)`: the Level 1 residuals $(r_{ij})$ `fitted(fit_object)`: the predicted outcomes at Level 1 $(Y_{ij})$ `ranef(fit_object)`: the Level 2 residuals $(u_j)$ `fixef(fit_object)`: the Level 2 fixed effects $(\gamma)$ `coef(fit_object)`: the Level 1 coefficients $(\beta_j)$

For the following, identify the requested individual by class number (and if relevant, by pupil number).

### Part A

Using the second model in Question 2, identify the teacher whose popularity ratings were most highly related to peer-rated popularity ratings. How high was that relationship?

$$popular_{ij} = \beta_0 + \beta_1 * popteach\_c_{ij} + r_{ij}$$
$$\beta_0 = \gamma_{00} + u_{0j}$$
$$\beta_1 = \gamma_{10} + u_{1j}$$

```r
(raneffs <- coef(mod1)[[1]] %>%
  data.frame() %>%
  mutate(class = rownames(.)) %>%
  tbl_df %>%
  arrange(desc(abs(popteach_c))))
```

```
## # A tibble: 100 x 3
##    X.Intercept. popteach_c class
##           <dbl>      <dbl> <chr>
## 1          2.68      0.792 82
## 2          4.13      0.782 100
## 3          4.18      0.776 48
## 4          3.53      0.775 97
## 5          4.00      0.772 8
## 6          3.64      0.772 55
## 7          4.30      0.771 31
## 8          3.69      0.769 14
## 9          4.00      0.766 86
## 10         4.45      0.765 71
## # ... with 90 more rows
```

The strongest relationship between between teacher and peer popularity ratings occurred in classroom 82, $\gamma_{10} + u_{1\ 82} = 0.7921922$.

## Part B

Using the model in Question 3, identify the teacher whose popularity ratings were the least related to peer-rated popularity when compared to what was expected for that teacher. How much poorer than expectation did this teacher perform?

$$popular_{ij} = \beta_0 + \beta_1 * popteach\_c_{ij} + r_{ij}$$

$$\beta_0 = \gamma_{00} + \gamma_{01} * gmc\_texp_j + \gamma_{02} * gmc\_mean\_pop\_t_j + u_{0j}$$

$$\beta_1 = \gamma_{10} + \gamma_{11} * gmc\_texp_j + \gamma_{12} * gmc\_mean\_pop\_t_j + u_{1j}$$

```r
(s.teach <- broom::augment(mod2) %>%
  mutate(pupil = 1:nrow(.)) %>%
  arrange(desc(abs(.resid))) %>%
  tbl_df %>%
  select(popular:.fixed, pupil))
```

```
## # A tibble: 2,000 x 9
##    popular popteach_c gmc_texp gmc_mean_pop_t class .fitted .resid .fixed
##      <dbl>      <dbl>    <dbl>          <dbl> <int>   <dbl>  <dbl>  <dbl>
## 1     2.40     0.0526  - 2.26         -0.113    56    4.98  -2.58   5.00
## 2     8.40     0.550    10.7          0.389     62    5.87   2.53   5.85
## 3     3.30     1.48    -11.3         -0.537     60    5.69  -2.39   5.70
## 4     2.00    -1.06    - 1.26        -0.00494   89    4.28  -2.28   4.29
## 5     6.40    -0.391   - 1.26        -0.669     88    4.14   2.26   4.13
## 6     5.50     0.800    10.7          2.14      35    7.69  -2.19   7.72
## 7     3.80     0.294     7.74         0.645     29    5.94  -2.14   5.93
## 8     3.80     0.600   - 0.263        0.340     73    5.82  -2.02   5.84
```

11

```
## 9     7.70     1.17     1.74     -0.227      91    5.71   1.99   5.69
## 10    6.50    -1.47     5.74       0.410      68    4.51   1.99   4.49
## # ... with 1,990 more rows, and 1 more variable: pupil <int>
```
```
# %>% group_by(class) %>% summarize(mean = mean(.resid)) %>% arrange(mean)
```

The weakest relationship between between teacher and peer popularity ratings compared to expectation occurred in classroom 56, $Y_{ij} - \hat{Y}_{ij} = -2.58$.

## Part C

Using the Level 2 residuals from the model in Question 3, which class exceeded their expected mean popularity the most? What was the simple (OLS) mean popularity of this class? What was the empirical Bayes estimate for the mean popularity of this class? Why are these two estimates (OLS and EB) for this class's mean popularity different?

$$\beta_1 = \gamma_{10} + \gamma_{11} * gmc\_texp_j + \gamma_{12} * gmc\_mean\_pop\_t_j + u_{1j}$$

```r
lm_fit <- function(data) {
  fit <- lm(popular ~ popteach_c, data)
  t(coef(fit)) %>% tbl_df %>% setNames(c("Intercept", "Slope"))
}

ols_mods <- broom::augment(mod2) %>% tbl_df %>%
  arrange(desc(.resid)) %>%
  group_by(class) %>%
  nest() %>%
  mutate(ols = map(data, lm_fit))

coef(mod2)[[1]] %>% data.frame() %>%
  mutate(class = rownames(.)) %>% tbl_df %>%
  arrange(desc(popteach_c))
```

```
## # A tibble: 100 x 7
##    X.Intercept. popteach_c gmc_texp gmc_mean_pop_t popteach_c.gmc_texp
##           <dbl>      <dbl>    <dbl>          <dbl>               <dbl>
## 1          5.07      0.737  0.00271          0.983            -0.00648
## 2          5.07      0.737  0.00271          0.983            -0.00648
## 3          5.06      0.735  0.00271          0.983            -0.00648
## 4          5.04      0.734  0.00271          0.983            -0.00648
## 5          5.08      0.733  0.00271          0.983            -0.00648
## 6          5.07      0.733  0.00271          0.983            -0.00648
## 7          5.06      0.733  0.00271          0.983            -0.00648
## 8          5.07      0.732  0.00271          0.983            -0.00648
## 9          5.09      0.732  0.00271          0.983            -0.00648
## 10         5.08      0.732  0.00271          0.983            -0.00648
## # ... with 90 more rows, and 2 more variables:
## #   popteach_c.gmc_mean_pop_t <dbl>, class <chr>
```

```r
raneffs <- ranef(mod2, condVar=TRUE)
Coef <- as.matrix(coef(mod2)$class[1:2])
# Then we select the posterior variances and covariances
# as an attribute and save it.
PVar <- attr(raneffs[[1]], which = "postVar")
```

```
dim(PVar) <- c(2*2 , 100)
Coef_Var_Cov <- t(PVar)
# And take it apart so each of the matrix elements for
# each school is a separate column.

L2_Resid <- ranef(mod2)[[1]]

Coef <- cbind(Coef,Coef_Var_Cov,L2_Resid)


# By_School <- ddply(HSB_Data, ~School, lm_fit)
# add to the Coef matrix and define as a dataframe.
Coef <- bind_cols(Coef, ols_mods %>% unnest(ols, .drop = T)) %>% tbl_df %>%
  setNames(c("EB_Intercept","EB_Slope","Intercept_Var","Covariance","Covariance_Too",
             "Slope_Var","Intercept_Res","Slope_Res","School","OLS_Intercept", "OLS_Slope")) %>%
  mutate(I_lwr = EB_Intercept-sqrt(Intercept_Var)*1.96,
         I_upr = EB_Intercept+sqrt(Intercept_Var)*1.96,
         S_lwr = EB_Slope-sqrt(Slope_Var)*1.96,
         S_upr = EB_Slope+sqrt(Slope_Var)*1.96)

Coef %>% select(EB_Intercept, EB_Slope)
```

```
## # A tibble: 100 x 2
##    EB_Intercept EB_Slope
##           <dbl>    <dbl>
## 1          5.09    0.726
## 2          5.06    0.725
## 3          5.10    0.728
## 4          5.05    0.720
## 5          5.08    0.723
## 6          5.07    0.724
## 7          5.07    0.726
## 8          5.09    0.730
## 9          5.06    0.733
## 10         5.09    0.712
## # ... with 90 more rows
```

Class 62 most exceeded their expected popularity by 0.01. The simple (OLS) mean popularity of this class
was 5.71 and the empirical Bayes estimate was 5.09. These two estimate are different because the MLM
expected values leverage information about the relationship between mean classroom popularlity and teaching
experience.


## Part D

Using the model in Question 4, which kid (identify the class number and kid number) exceeded expected
popularity the most? What was that child's actual peer-rated popularity?

```
(q5d <- broom::augment(mod3) %>%
  mutate(pupil = as.numeric(rownames(.))) %>%
  tbl_df %>%
  arrange(desc(.resid)))
```

```
## # A tibble: 2,000 x 14
##    popular  sex_c popteach_c class .fitted .resid .fixed   .mu .offset
```

```
##        <dbl>   <dbl>      <dbl> <int>   <dbl>   <dbl>   <dbl> <dbl>   <dbl>
##  1     8.40   0.495      0.550    62    6.30    2.10    5.69  6.30       0
##  2     6.00  -0.505     -0.550    27    3.95    2.05    4.46  3.95       0
##  3     8.40   0.495      1.91     46    6.43    1.97    6.51  6.43       0
##  4     6.60  -0.505     -1.71     29    4.74    1.86    3.80  4.74       0
##  5     7.40   0.495      1.79     48    5.63    1.77    6.44  5.63       0
##  6     6.50  -0.505     -1.47     68    4.73    1.77    3.93  4.73       0
##  7     4.90  -0.505     -1.20      2    3.13    1.77    4.09  3.13       0
##  8     5.30  -0.505     -1.87     99    3.58    1.72    3.71  3.58       0
##  9     7.40   0.495      0.857    70    5.72    1.68    5.87  5.72       0
## 10     7.70   0.495      1.17     91    6.04    1.66    6.06  6.04       0
## # ... with 1,990 more rows, and 5 more variables: .sqrtXwt <dbl>,
## #   .sqrtrwt <dbl>, .weights <dbl>, .wtres <dbl>, pupil <dbl>
dat <- dat %>% mutate(student = 1:dim(dat)[1])
```

Student 7 in class 62 exceeded their expected popularity by 2.1. Their rated popularity was 8.4.