Multilevel Modeling

Contents

```
library(mosaic)
library(dplyr)
library(ggplot2)

#install.packages("lme4")
library(lme4)
```

Load the Hox Data

We will use the `hox` dataset that has the example from the slides. Each pupil has a popularity rating, `popular`, and extroversion rating, `extrav`, their gender, `sex`, the class room their in, `class`, and their teachers' years of experience, `texp`. We will be exploring popularity as a function of extroversion, sex, and teacher's experience.

```
hox <- read.csv("/Users/randigarcia/Desktop/Data/hox.csv", header=TRUE)

glimpse(hox)
```

```
## Observations: 2,000
## Variables: 7
## $ pupil    <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16...
## $ class    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
```

```
## $ extrav   <int> 5, 7, 4, 3, 5, 4, 5, 4, 5, 5, 5, 5, 5, 5, 5, 6, 4, 4,...
## $ sex      <int> 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0,...
## $ texp     <int> 24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 24, 2...
## $ popular  <dbl> 6.3, 4.9, 5.3, 4.7, 6.0, 4.7, 5.9, 4.2, 5.2, 3.9, 5.7...
## $ popteach <int> 6, 5, 6, 5, 6, 5, 5, 5, 5, 3, 5, 5, 5, 6, 5, 5, 2, 3,...
```

Exploratory Data Analysis

Do some preliminary data analysis to get familiar with the data set.

- What are the descriptives for `popular`?
- Make some visualizations. How about sex?
- How many classrooms do we have (hint: the `distinct()` function might help)?
- What are the gender percentages by classroom?
- Which classrooms have the top 5 most popular teachers (variable 'popteach')?

```
#Exploratory data analysis
```

Random Intercept, No Predictors

First let's fit the most basic multilevel model, the random intercept model with no predictors.

```
mlm <- lmer(popular ~ 1 + (1 | class), data = hox)

summary(mlm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ 1 + (1 | class)
##    Data: hox
##
## REML criterion at convergence: 6330.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5655 -0.6975  0.0020  0.6758  3.3175
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class    (Intercept) 0.7021   0.8379
##  Residual             1.2218   1.1053
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
```

```
##             Estimate Std. Error t value
## (Intercept)  5.07786    0.08739    58.1
```

Note that model objects have handy properties:

```
names(summary(mlm))
```

```
##  [1] "methTitle"    "objClass"     "devcomp"      "isLmer"
##  [5] "useScale"     "logLik"       "family"       "link"
##  [9] "ngrps"        "coefficients" "sigma"        "vcov"
## [13] "varcor"       "AICtab"       "call"         "residuals"
## [17] "fitMsgs"      "optinfo"
```

```
#which do you think will confirm your number of groups from above?
```

Intraclass Correlation (ICC)

The intraclass correlation (ICC) can be computed as:

```
ran_eff <- as.data.frame(summary(mlm)$varcor) %>%
  select(-var1, -var2)

icc <- ran_eff[1,2]/(ran_eff[1,2] + ran_eff[2,2])

icc
```

```
## [1] 0.3649386
```

Level 1 Predictor, Fixed

Does the pupil's extroversion predict their popularity? Now, let's add 1 level 1 predictor, the pupil's extroversion score. We'll ass the fixed component only, not the random piece.

```
mlm_pred1 <- lmer(popular ~ extrav + (1 | class), data = hox)

summary(mlm_pred1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ extrav + (1 | class)
##    Data: hox
##
## REML criterion at convergence: 5832.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.0644 -0.7267  0.0165  0.7088  3.3587
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class    (Intercept) 0.8406   0.9168
##  Residual             0.9304   0.9646
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.54214    0.14113   18.01
## extrav       0.48631    0.02015   24.13
##
## Correlation of Fixed Effects:
##        (Intr)
## extrav -0.745
```

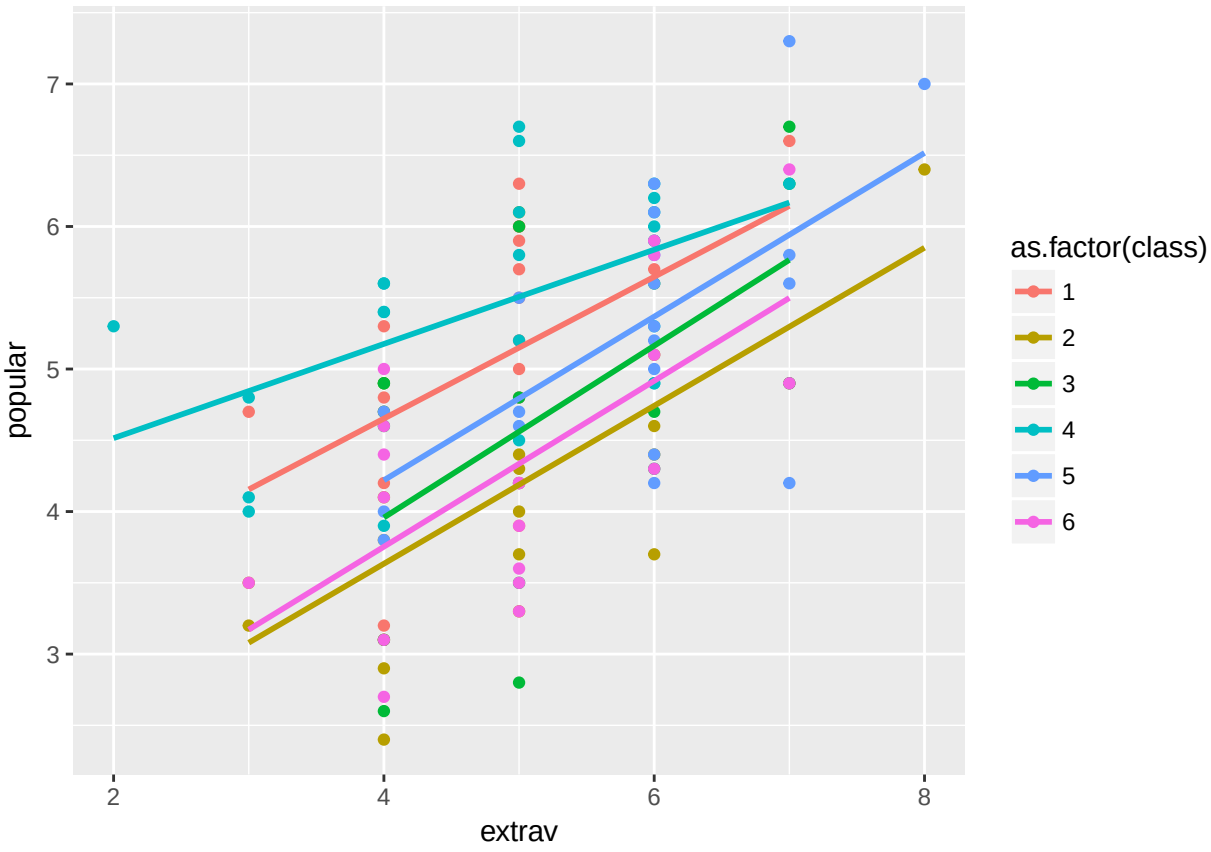
Level 1 predictor, Random

Spaghetti Plot

Spaghetti plot shows effects for different classrooms.

```
hox_small <- hox %>%
  filter(class >= 1 & class <= 6)

ggplot(hox_small, aes(extrav, popular,
                          group = as.factor(class),
                          color = as.factor(class))) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```



Try faceting by class instead.

```
#Alt spaghetti plot
```

For every 1 unit increase in extroversion we'd expect a 0.49 point increase in popularity. Does this size of this popularity effect differ across classrooms? Let's ass the random component of the extroversion effect.

```
mlm_pred1_ran <- lmer(popular ~ extrav + (extrav | class), data = hox)

summary(mlm_pred1_ran)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ extrav + (extrav | class)
##    Data: hox
##
## REML criterion at convergence: 5779.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1961 -0.7291  0.0146  0.6816  3.2217
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
```

```
##  class    (Intercept) 2.99952  1.7319
##           extrav      0.02599  0.1612   -0.97
##  Residual             0.89492  0.9460
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.46102    0.20315   12.11
## extrav       0.49286    0.02546   19.36
##
## Correlation of Fixed Effects:
##        (Intr)
## extrav -0.917
```

The standard deviation of the extroversion effect is 0.16. There is a negative correlation between the classroom's intercept and the effect of extroversion.

Level 2 predictor

Does the teacher's experience have an effect on the pupil's popularity? Let's ass this level 2 predictor, it can only be fixed.

```
mlm_pred2 <- lmer(popular ~ texp + (1 | class), data = hox)

summary(mlm_pred2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ texp + (1 | class)
##    Data: hox
##
## REML criterion at convergence: 6313.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5326 -0.6963 -0.0005  0.6896  3.3455
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class    (Intercept) 0.5427   0.7367
##  Residual             1.2218   1.1053
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
```

```
## (Intercept)  4.19668    0.18609  22.552
## texp         0.06164    0.01183   5.212
##
## Correlation of Fixed Effects:
##      (Intr)
## texp -0.909
```

Yes, more experienced teachers seem to have more popular pupils. Next we we can have level 1 and level 2 predictors, the level 1 predictor is random.

```
mlm_pred12_ran <- lmer(popular ~ extrav + texp + (extrav | class), data = hox)

summary(mlm_pred12_ran)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ extrav + texp + (extrav | class)
##    Data: hox
##
## REML criterion at convergence: 5749.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0472 -0.7305  0.0110  0.6777  3.2032
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  class    (Intercept) 1.55946  1.2488
##           extrav      0.03534  0.1880   -0.87
##  Residual             0.89059  0.9437
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.01683    0.23072   4.407
## extrav       0.50215    0.02763  18.173
## texp         0.09777    0.01026   9.530
##
## Correlation of Fixed Effects:
##        (Intr) extrav
## extrav -0.719
## texp   -0.702  0.103
```

Cross-Level Interaction

Is the effect of extroversion stronger in classrooms with more experienced teachers? Finally, we can estimate the cross-level interaction of extroversion and teacher's experience.

```
mlm_inter <- lmer(popular ~ extrav*texp + (extrav | class), data = hox)

summary(mlm_inter)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: popular ~ extrav * texp + (extrav | class)
##    Data: hox
##
## REML criterion at convergence: 5694.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2505 -0.7267  0.0162  0.6861  3.1895
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  class    (Intercept) 0.5289833 0.72731
##           extrav      0.0003337 0.01827  -1.00
##  Residual             0.8898028 0.94329
## Number of obs: 2000, groups:  class, 100
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -1.210965   0.321040  -3.772
## extrav       0.891087   0.044954  19.822
## texp         0.251926   0.019746  12.758
## extrav:texp -0.027462   0.002872  -9.563
##
## Correlation of Fixed Effects:
##             (Intr) extrav texp
## extrav      -0.874
## texp        -0.918  0.805
## extrav:texp  0.773 -0.898 -0.859
```

See if you can make a graph to get a sense of this interaction. Hint: first create a variable that is the median split of `texp`.

```
#graph the interaction
```

Check the residuals.

```
#model diagnostics here.
```