Behavioral Data Sciece Week 6 - Mixed Models

Jonathan Gragg

10/3/2021

Summary

In my findings the use of the first 12 questions isn't extremely effect. Adding in a mixed model to randomize the intercept based on LeaderID it gets slightly more effective. The most effective model is a random slope model using the average score and leader.

Analysis

Loading in packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.0.3      v dplyr   1.0.2
## v tidyr   1.1.1      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(haven)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
library(lmerTest)
```

```
##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##     lmer

## The following object is masked from 'package:stats':
##
##     step
```

loading in the data

```
data <- read_dta('BDS-Leadership-DataSet.dta')
#selecting relevant columns
data <- data %>% select(leaderID:lvi12,effect)
#creating score based on responses to 1-12
data$score <- rowMeans(data[,3:14],na.rm = T)
glimpse(data)
```

```
## Rows: 26,327
## Columns: 16
## $ leaderID       <dbl> 2, 2, 2, 2, 2, 2, 3, 3, 3, 6, 6, 6, 6, 7, 7, 7, 7, 7...
## $ OrganizationID <dbl> 41, 41, 41, 41, 41, 41, 41, 41, 41, 82, 82, 82, 82, ...
## $ lvi01          <dbl> 1, 0, 0, -2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, ...
## $ lvi02          <dbl> 0, 0, 0, -2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, ...
## $ lvi03          <dbl> 0, 1, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, ...
## $ lvi04          <dbl> 2, 1, 0, 0, -1, -1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,...
## $ lvi05          <dbl> 0, 0, 0, -3, 0, 0, 1, 0, 0, -1, 0, 0, 0, 0, 0, 0, 0,...
## $ lvi06          <dbl> 0, 0, 0, -1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, ...
## $ lvi07          <dbl> 0, 1, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, ...
## $ lvi08          <dbl> 1, 1, 0, -2, 0, 0, 0, -1, 0, -1, 0, 0, -1, 3, 0, 0, ...
## $ lvi09          <dbl> 0, 0, 0, -3, 1, 0, 1, 0, 0, 0, -1, -1, -1, 3, 0, 0, ...
## $ lvi10          <dbl> 0, 0, 0, -3, 0, 0, 0, 1, 0, -1, -1, 0, 0, 2, 0, 0, 1...
## $ lvi11          <dbl> 0, 0, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 0, 0, 0, ...
## $ lvi12          <dbl> 0, -1, 0, -2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2,...
## $ effect         <dbl> 8.50, 8.50, 9.00, 8.00, 8.50, 7.50, 8.00, 9.00, 9.10...
## $ score          <dbl> 0.33333333, 0.25000000, 0.00000000, -1.75000000, 0.0...
```
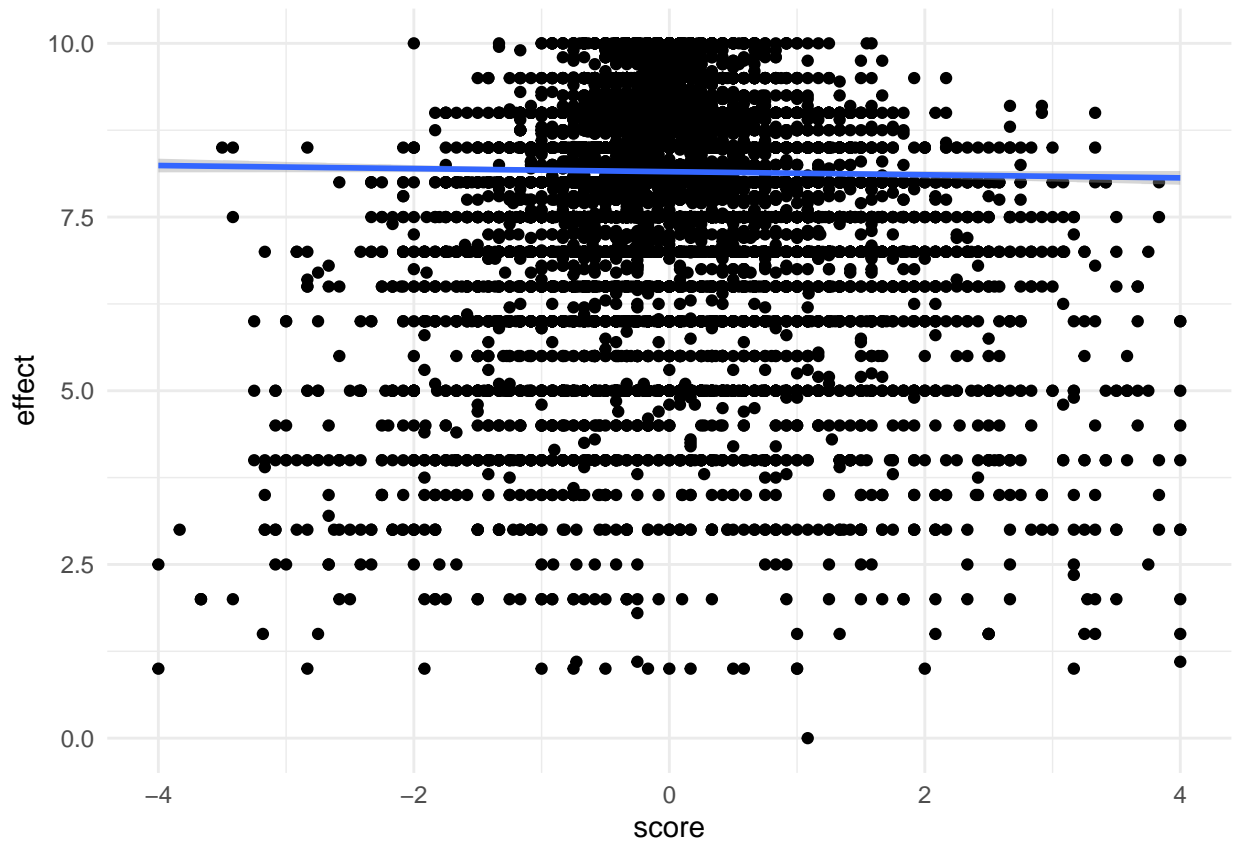
```
data %>%
  select(effect, score, leaderID, OrganizationID) %>%
  cor() %>%
  ggcorrplot::ggcorrplot()
```
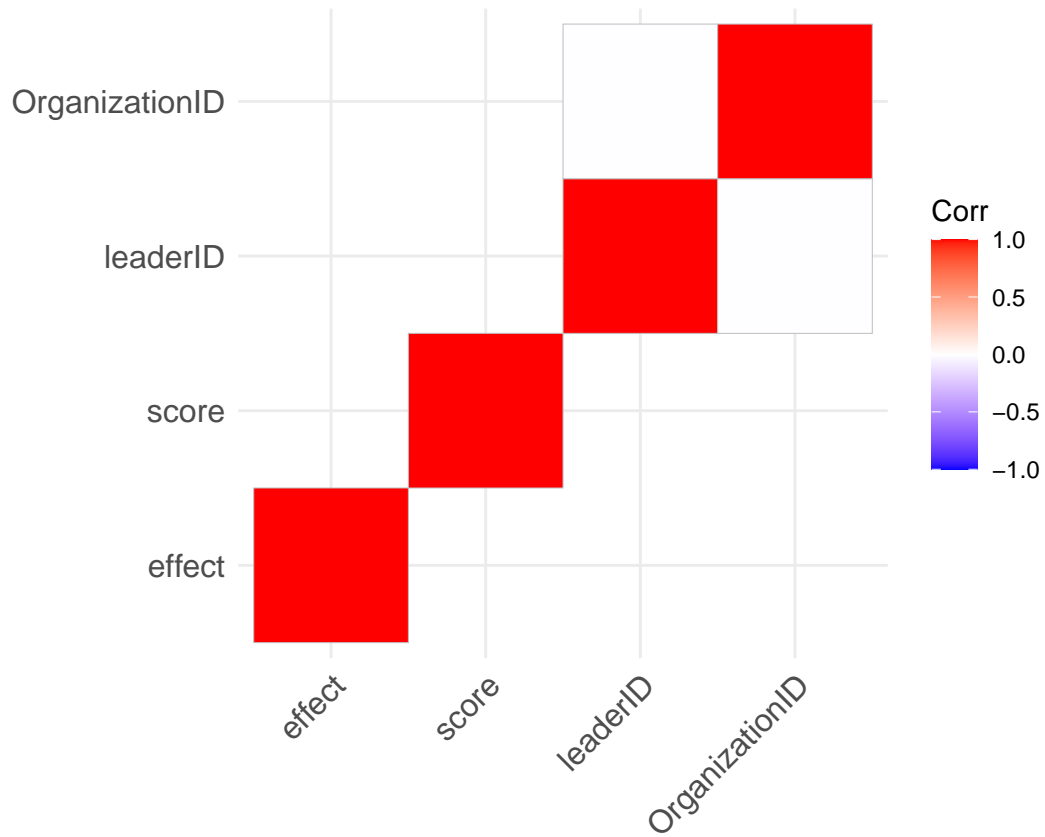
```
slim = lm(effect ~ score, data = data)

ggplot(data, aes(score, effect)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_minimal()
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

```
#Ordinary least squares model
mod_ols <- lm(effect~score, data = data)
summary(mod_ols)
```

```
##
## Call:
## lm(formula = effect ~ score, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1288 -0.6436  0.3398  0.8472  1.8823
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  8.152799   0.008065 1010.914   <2e-16 ***
## score       -0.022187   0.012423   -1.786   0.0741 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.308 on 26313 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.0001212,  Adjusted R-squared:  8.321e-05
## F-statistic:  3.19 on 1 and 26313 DF,  p-value: 0.07411
```

```
plot(data$score,data$effect)
abline(mod_ols)
```
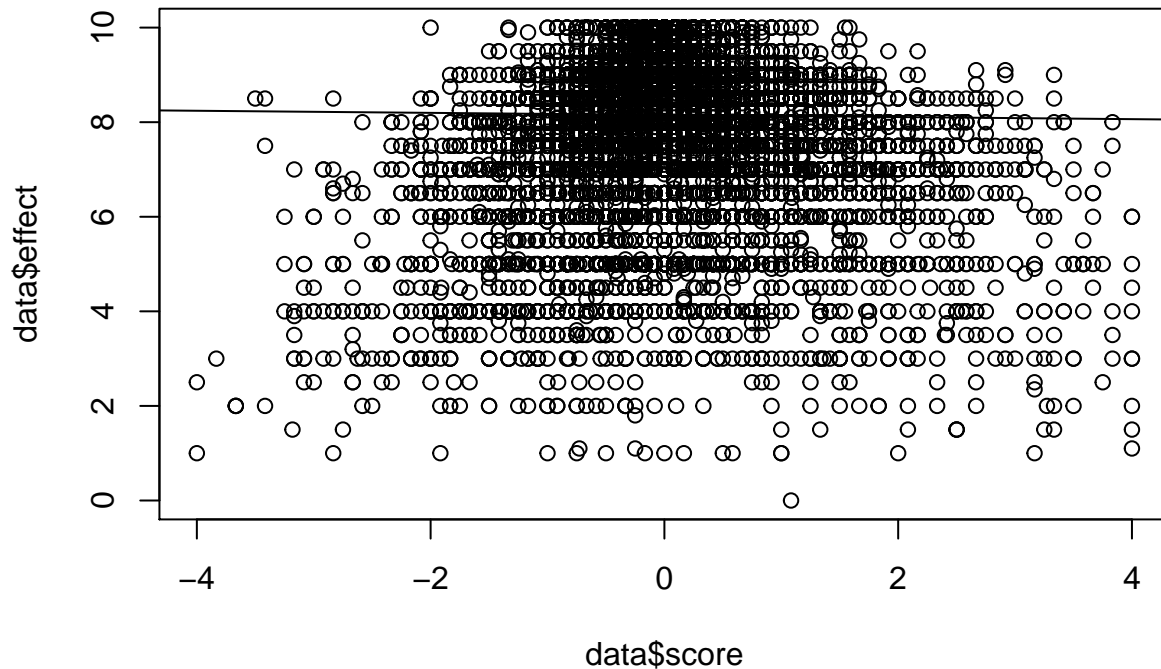


```
mod_null <- lmer(effect~ (1|leaderID),data = data)
summary(mod_null)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: effect ~ (1 | leaderID)
##    Data: data
##
## REML criterion at convergence: 86838.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.0465 -0.4027  0.1451  0.6009  3.0024
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  leaderID (Intercept) 0.3783   0.615
##  Residual             1.3425   1.159
## Number of obs: 26325, groups:  leaderID, 4886
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
```

```
## (Intercept) 8.138e+00  1.154e-02 4.660e+03    704.9   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#calculating the ICC
0.3782/(.3782+1.3425)
```

```
## [1] 0.2197943
```

```
#mixed model
mod_mixed <- lmer(effect ~ score + (1|leaderID), data = data)
summary(mod_mixed)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: effect ~ score + (1 | leaderID)
##    Data: data
##
## REML criterion at convergence: 86776
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.0280 -0.4027  0.1445  0.6001  3.0066
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  leaderID (Intercept) 0.3781   0.6149
##  Residual             1.3405   1.1578
## Number of obs: 26315, groups:  leaderID, 4886
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  8.138e+00  1.154e-02 4.661e+03 705.192   <2e-16 ***
## score       -2.694e-02  1.275e-02 2.601e+04  -2.113   0.0346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## score 0.007
```

```
#random slope model
mod_rs <- lmer(effect ~ score + (score|leaderID), data = data)
summary(mod_rs)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: effect ~ score + (score | leaderID)
##    Data: data
##
## REML criterion at convergence: 83682.6
##
```

```
## Scaled residuals:
##     Min     1Q  Median      3Q     Max
## -6.7705 -0.4441  0.1001  0.5850  3.2880
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  leaderID (Intercept) 0.2820   0.5310
##           score       0.8751   0.9355   -0.02
##  Residual             1.0573   1.0283
## Number of obs: 26315, groups:  leaderID, 4886
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 8.292e+00  1.062e-02 4.632e+03 780.909   <2e-16 ***
## score       4.407e-02  2.063e-02 3.743e+03   2.136   0.0327 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## score 0.032
```

```
anova(mod_mixed,mod_rs)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
## Models:
## mod_mixed: effect ~ score + (1 | leaderID)
## mod_rs: effect ~ score + (score | leaderID)
##           npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## mod_mixed    4 86770 86803 -43381    86762
## mod_rs       6 83681 83730 -41835    83669 3092.6  2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#3 Level model
mod_3L <- lmer(effect ~ score + (1|OrganizationID/leaderID), data = data)
summary(mod_3L)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: effect ~ score + (1 | OrganizationID/leaderID)
##    Data: data
##
## REML criterion at convergence: 86614.1
##
## Scaled residuals:
##     Min     1Q  Median      3Q     Max
## -6.1176 -0.3989  0.1434  0.5982  3.0449
##
## Random effects:
```

```
##  Groups                   Name          Variance Std.Dev.
##  leaderID:OrganizationID (Intercept) 0.33755  0.5810
##  OrganizationID          (Intercept) 0.08079  0.2842
##  Residual                             1.34048  1.1578
## Number of obs: 26315, groups:
## leaderID:OrganizationID, 4886; OrganizationID, 131
##
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  8.163e+00  3.220e-02  8.961e+01 253.517   <2e-16 ***
## score       -2.781e-02  1.272e-02  2.591e+04  -2.186   0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## score 0.005
```

```r
ICC_Lead <- (0.3377+0.0807)/(0.3377+0.0807+1.3425)
ICC_Org <- (0.0807)/(0.3377+0.0807+1.3425)
ICC_Lead
```

```
## [1] 0.2376058
```

```r
ICC_Org
```

```
## [1] 0.04582884
```