R Notebook

Neil Yetz & Gemma Wallace

Clear environment

```
rm(list=ls())
```

Load Libraries

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## Warning: package 'ggplot2' was built under R version 3.6.3
## Warning: package 'tibble' was built under R version 3.6.3
## Warning: package 'tidyr' was built under R version 3.6.3
## Warning: package 'dplyr' was built under R version 3.6.3
## Warning: package 'forcats' was built under R version 3.6.3
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lme4)
```

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

```
library(psych)
```

```
##
## Attaching package: 'psych'
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(MuMIn)
```

```
## Warning: package 'MuMIn' was built under R version 3.6.3
```

Read in Kim's data: Demo activity

```
teams <- read_csv("mlm_teams.csv")
```

```
## Parsed with column specification:
## cols(
##   kid_id = col_double(),
##   team_id = col_double(),
##   txcond = col_double(),
##   risk = col_double(),
##   score = col_double(),
##   comafrd = col_double()
## )
```

Describe

```
describe(teams)
```

```
##          vars   n   mean     sd median trimmed    mad   min    max  range  skew
## kid_id      1 500 250.50 144.48  250.5  250.50 185.32  1.00 500.00 499.00  0.00
## team_id     2 500  50.50  28.89   50.5   50.50  37.06  1.00 100.00  99.00  0.00
## txcond      3 500   0.50   0.50    0.5    0.50   0.74  0.00   1.00   1.00  0.00
## risk        4 500   2.00   1.42    2.0    2.00   1.48  0.00   4.00   4.00  0.00
## score       5 500  73.22  12.35   74.0   73.55  11.86 21.00 100.00  79.00 -0.43
## comafrd     6 500   4.50   1.27    4.4    4.48   1.24  1.31   8.73   7.41  0.25
##          kurtosis   se
## kid_id      -1.21 6.46
## team_id     -1.21 1.29
## txcond      -2.00 0.02
## risk        -1.31 0.06
## score        0.45 0.55
## comafrd      0.22 0.06
```

Factor txcond

```
teams <- teams %>%
mutate(txcond.f = factor(txcond, levels = c(0,1), labels = c("control", "treatment")),
       team_id.f = factor(team_id))
```

Factor team_id

```
teams <- mutate(teams, team_id.f = factor(team_id))
```

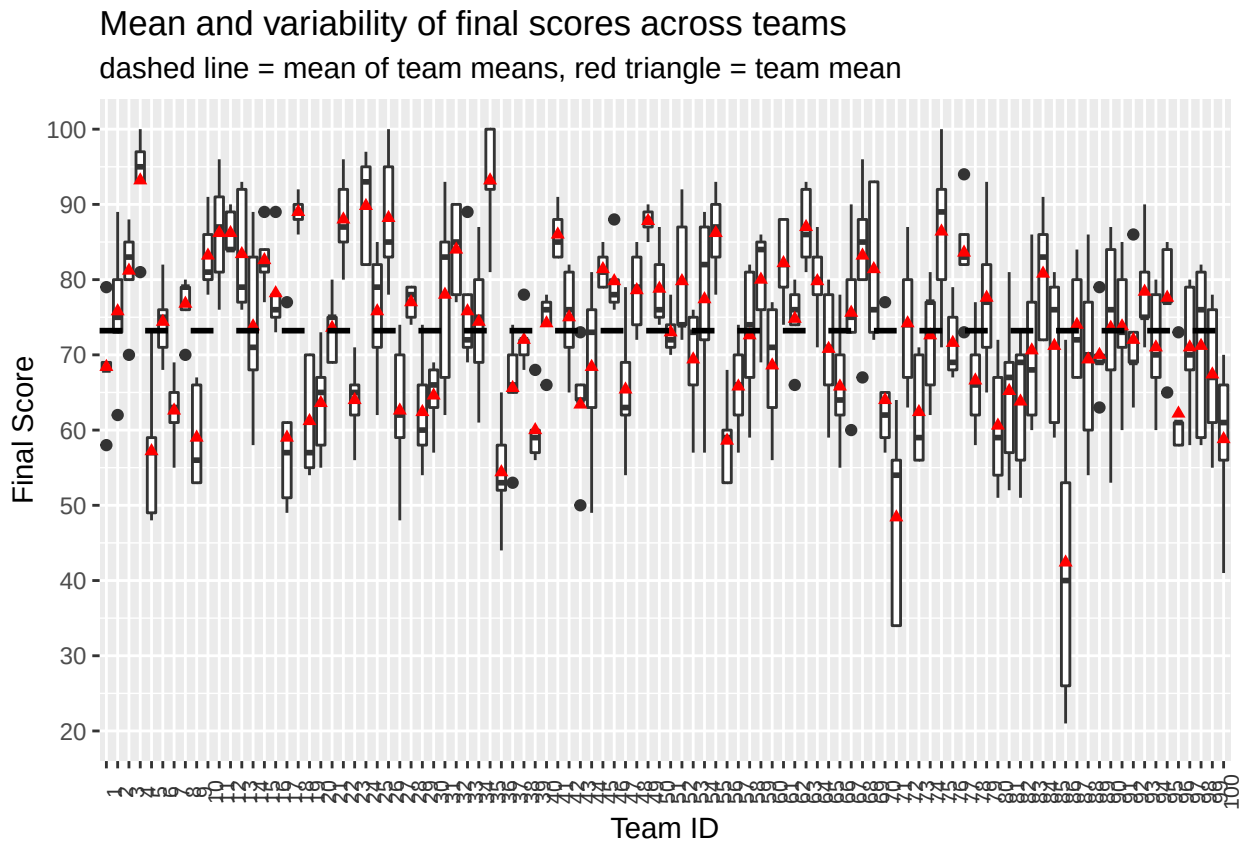
Get Mean of Means

```
team_means  <- group_by(teams, team_id)
team_means  <- summarize(team_means, mean_score = mean(score))
```

```
meanofmeans <- summarize(team_means, meanofmeans = mean(mean_score))
meanofmeans
```

```
## # A tibble: 1 x 1
##   meanofmeans
##         <dbl>
## 1        73.2
```

```
ggplot(data = teams, aes(x = team_id.f, y = score)) +
  geom_boxplot() +
  stat_summary(aes(y = score, group = team_id.f), fun = mean, color = "red", geom = "point", pch = 17, s
  geom_hline(yintercept = meanofmeans$meanofmeans, linetype="dashed", color = "black", size = 1) +
  scale_y_continuous(limits = c(20,100), breaks = seq(20, 100, 10)) +
  labs(title = "Mean and variability of final scores across teams",
       subtitle = "dashed line = mean of team means, red triangle = team mean",
       x = "Team ID", y = "Final Score") +
  theme(axis.text.x = element_text(colour="grey20", size=8, angle=90, hjust=.5))
```



Models

```
# Random intercept
randint <- lmer(score ~ 1 + (1|team_id.f), data = teams, REML = FALSE)
summary(randint)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: score ~ 1 + (1 | team_id.f)
```

```
##    Data: teams
##
##      AIC      BIC   logLik deviance df.resid
##   3749.0   3761.6  -1871.5   3743.0      497
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.08009 -0.59814 -0.01048  0.70383  2.95197
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  team_id.f (Intercept) 80.62    8.979
##  Residual              71.48    8.455
## Number of obs: 500, groups:  team_id.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  73.2160     0.9743   75.15
```

```
(80.62 / (80.62 + 71.48))
```

```
## [1] 0.530046
```

```
library(sjstats)
```

```
## Warning: package 'sjstats' was built under R version 3.6.3
```

```
##
## Attaching package: 'sjstats'
```

```
## The following object is masked from 'package:psych':
##
##     phi
```

```
performance::icc(randint)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.530
##   Conditional ICC: 0.530
```

```
r.squaredGLMM(randint)
```

```
## Warning: 'r.squaredGLMM' now calculates a revised statistic. See the help page.
```

```
##      R2m       R2c
## [1,]   0 0.5300362
```

```
# Random intercept Fixed slope
rifs<-lmer(score ~ 1 + risk + (1|team_id.f),data=teams, REML = FALSE)
summary(rifs)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: score ~ 1 + risk + (1 | team_id.f)
##    Data: teams
##
##      AIC      BIC   logLik deviance df.resid
##   3667.1   3683.9  -1829.5   3659.1      496
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2366 -0.5766 -0.0013  0.6421  2.7827
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  team_id.f (Intercept) 83.33    9.128
##  Residual              57.96    7.613
## Number of obs: 500, groups:  team_id.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  77.8680     1.0867  71.653
## risk         -2.3260     0.2407  -9.662
##
## Correlation of Fixed Effects:
##      (Intr)
## risk -0.443
```

```
performance::icc(rifs)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.590
##   Conditional ICC: 0.548
```

```
83.33 / (83.33 + 57.96)
```

```
## [1] 0.5897799
```

```
r.squaredGLMM(rifs)
```

```
##             R2m       R2c
## [1,] 0.07127094 0.6190137
```

```
# random slopes fixed intercepts
rsfi<- lmer(score ~ 1 + risk + (0 + risk|team_id.f),data=teams, REML = FALSE)
summary(rsfi)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: score ~ 1 + risk + (0 + risk | team_id.f)
##    Data: teams
##
##      AIC      BIC   logLik deviance df.resid
##   3747.5   3764.3  -1869.7   3739.5      496
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3701 -0.5836  0.0391  0.5992  2.5838
##
## Random effects:
##  Groups    Name Variance Std.Dev.
##  team_id.f risk 11.32    3.364
##  Residual       73.37    8.566
## Number of obs: 500, groups:  team_id.f, 100
##
## Fixed effects:
```

```
##             Estimate Std. Error t value
## (Intercept)  77.8680     0.6635 117.359
## risk         -2.3260     0.4319  -5.385
##
## Correlation of Fixed Effects:
##      (Intr)
## risk -0.512
```

```
performance::icc(rsfi)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.481
##   Conditional ICC: 0.446
```

```
11.32 / (11.32 + 73.37)
```

```
## [1] 0.133664
```

```
r.squaredGLMM(rsfi)
```

```
##             R2m       R2c
## [1,] 0.07127094 0.5176855
```

```
# all random
allrand<- lmer(score ~ risk +  (1 + risk|team_id.f),data=teams, REML = FALSE)
summary(allrand)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: score ~ risk + (1 + risk | team_id.f)
##    Data: teams
##
##      AIC      BIC   logLik deviance df.resid
##   3618.7   3644.0  -1803.4   3606.7      494
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.52404 -0.55663  0.02976  0.58267  2.14656
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  team_id.f (Intercept) 79.603   8.922
##            risk         7.101   2.665    -0.22
##  Residual              40.206   6.341
## Number of obs: 500, groups:  team_id.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  77.8680     1.0185  76.456
## risk         -2.3260     0.3335  -6.975
##
## Correlation of Fixed Effects:
##      (Intr)
## risk -0.392
```

```
performance::icc(allrand)
```

```
## # Intraclass Correlation Coefficient
```

```
##
##      Adjusted ICC: 0.715
##   Conditional ICC: 0.664
```

```
(79.603 + 7.213) / (79.603 + 7.213 + 40.206)
```

```
## [1] 0.6834722
```

```
library(modelr)
```

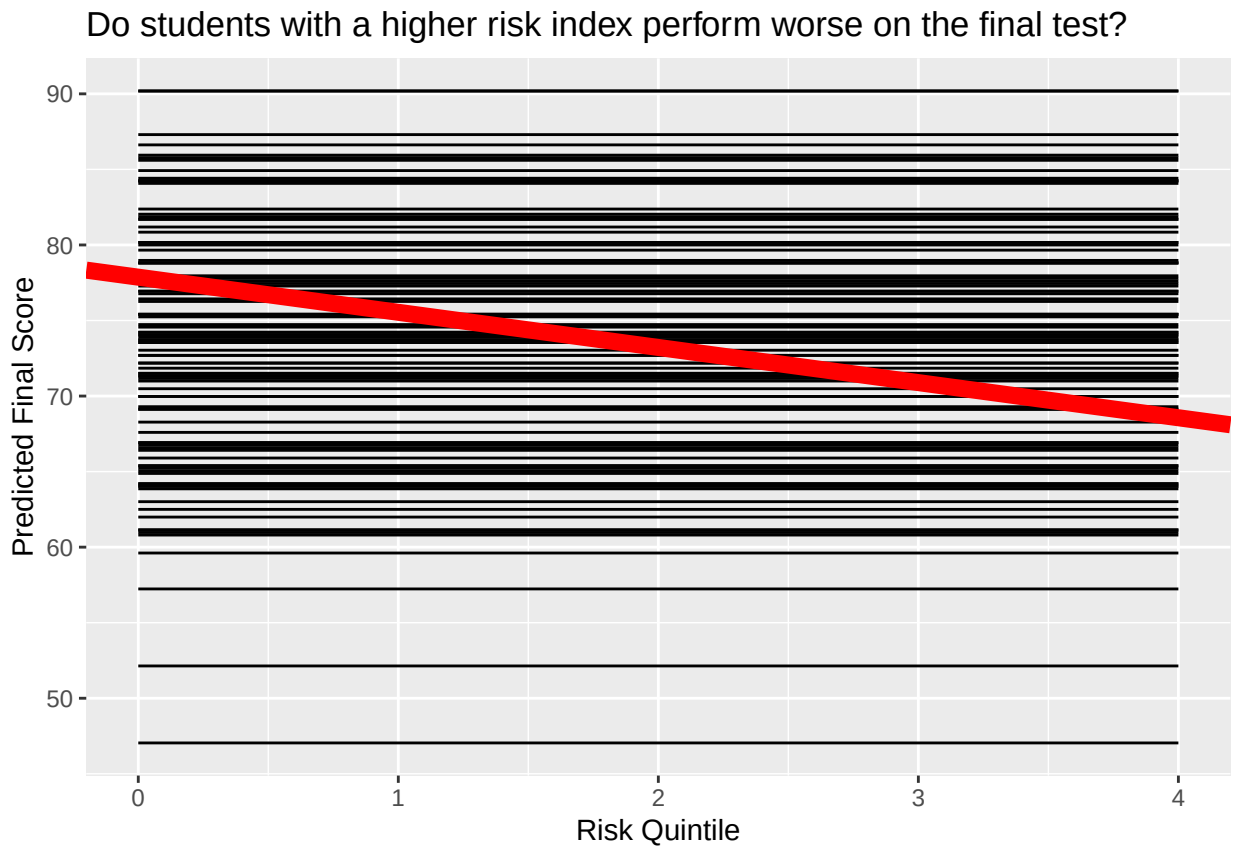
```
## Warning: package 'modelr' was built under R version 3.6.3
```

```
##
## Attaching package: 'modelr'
```

```
## The following objects are masked from 'package:sjstats':
##
##     bootstrap, mse, rmse
```

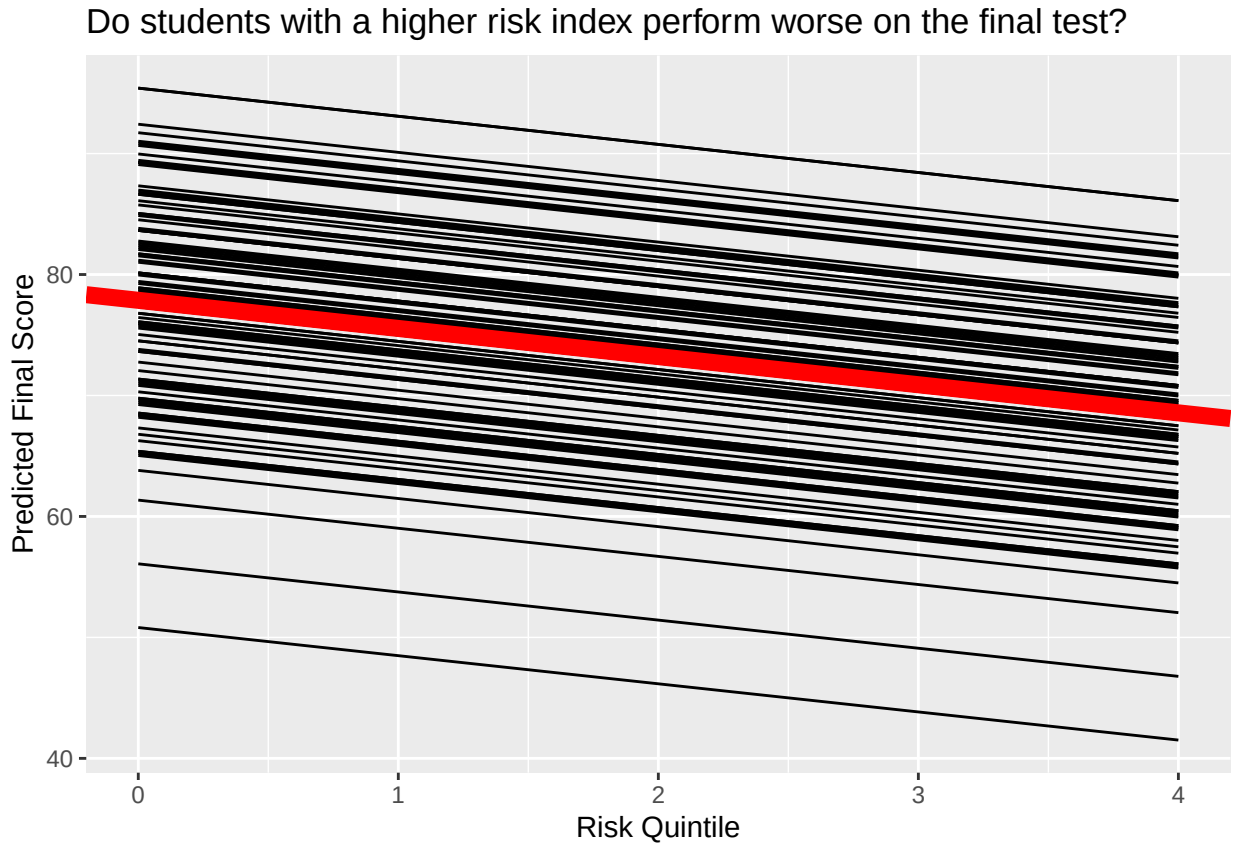
```
# Model without with only random intercept
mod2_b.plot <- add_predictions(data = teams, model = randint)

ggplot(data = mod2_b.plot, aes(x = risk, y = pred, group = team_id.f)) +
  geom_line() +
  geom_abline(intercept = 77.868, slope = -2.326, color="red", size=3) +
  labs(title = "Do students with a higher risk index perform worse on the final test?", x = "Risk Quinti
         "Predicted Final Score")
```



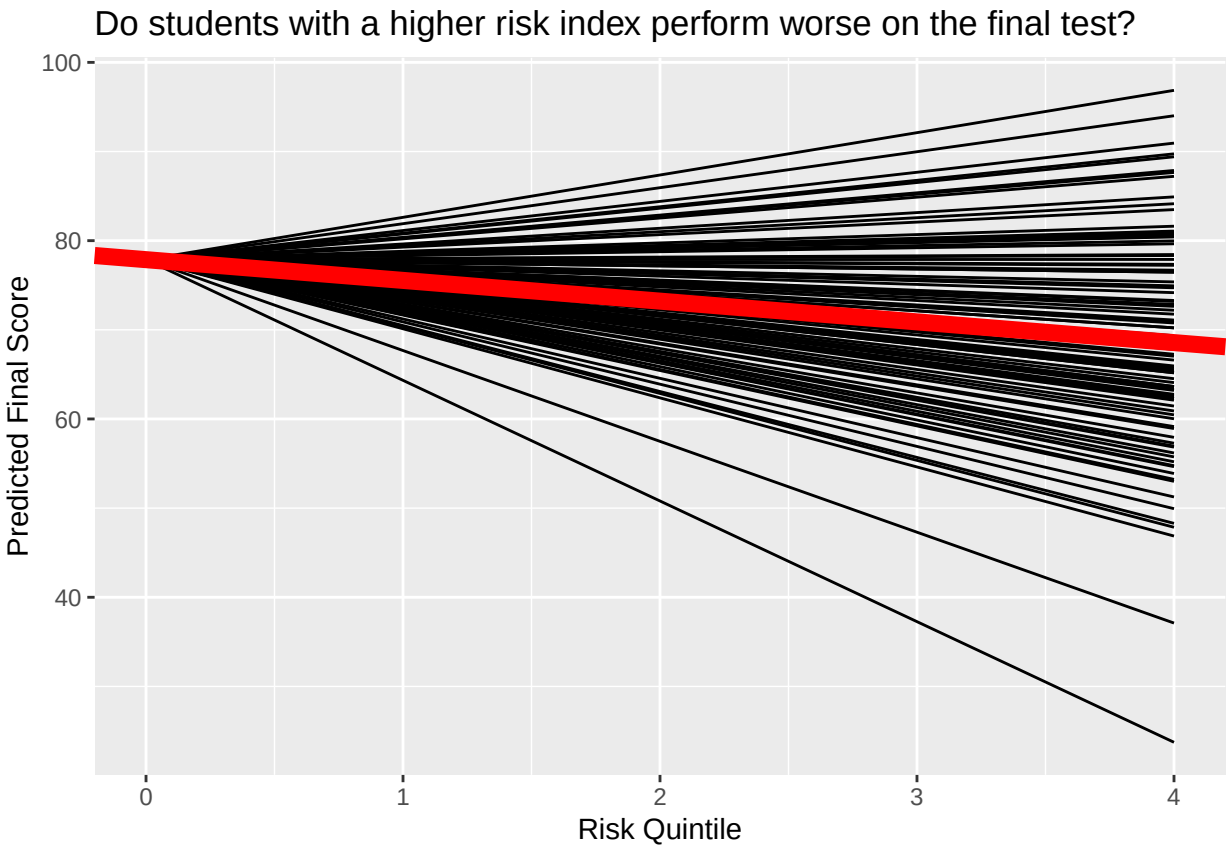

```
# Model without random slope
mod2_b.plot <- add_predictions(data = teams, model = rifs)
```

```
ggplot(data = mod2_b.plot, aes(x = risk, y = pred, group = team_id.f)) +
  geom_line() +
  geom_abline(intercept = 77.868, slope = -2.326, color="red", size=3) +
  labs(title = "Do students with a higher risk index perform worse on the final test?", x = "Risk Quinti
         "Predicted Final Score")
```



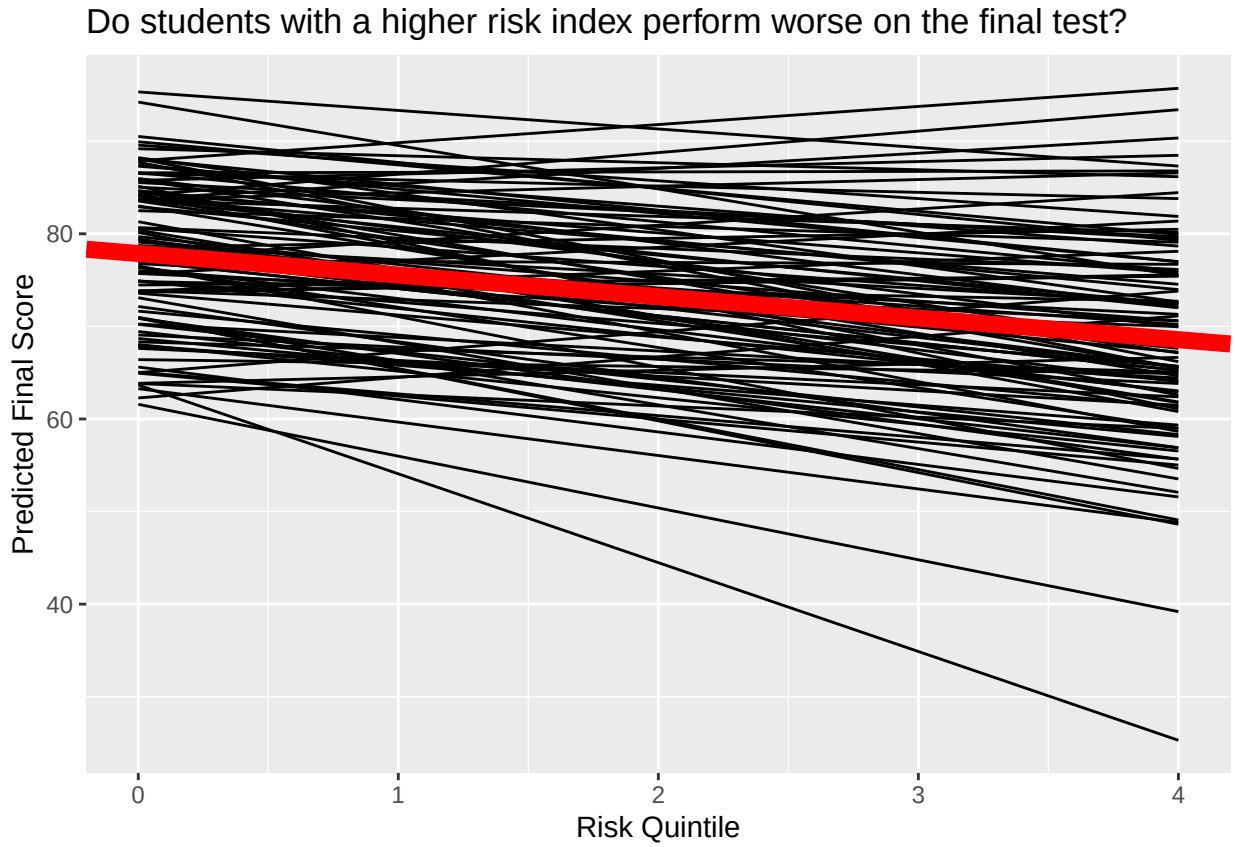

```
# Model without random intercept
mod2_b.plot <- add_predictions(data = teams, model = rsfi)

ggplot(data = mod2_b.plot, aes(x = risk, y = pred, group = team_id.f)) +
  geom_line() +
  geom_abline(intercept = 77.868, slope = -2.326, color="red", size=3) +
  labs(title = "Do students with a higher risk index perform worse on the final test?", x = "Risk Quinti
         "Predicted Final Score")
```




```
# Model with random slope & Intercept
mod2_a.plot <- add_predictions(data = teams, model = allrand)

ggplot(data = mod2_a.plot, aes(x = risk, y = pred, group = team_id.f)) +
  geom_line() +
  geom_abline(intercept = 77.868, slope = -2.326, color="red", size=3) +
  labs(title = "Do students with a higher risk index perform worse on the final test?", x = "Risk Quinti
```




```
r.squaredGLMM(allrand)
```

```
##            R2m       R2c
## [1,] 0.0712698 0.7357103
```

Pairwise comparison

```
anova(randint,allrand)
```

```
## Data: teams
## Models:
## randint: score ~ 1 + (1 | team_id.f)
## allrand: score ~ risk + (1 + risk | team_id.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## randint  3 3749.0 3761.6 -1871.5   3743.0
## allrand  6 3618.7 3644.0 -1803.3   3606.7 136.27      3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(rifs,allrand)
```

```
## Data: teams
## Models:
## rifs: score ~ 1 + risk + (1 | team_id.f)
## allrand: score ~ risk + (1 + risk | team_id.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rifs     4 3667.1 3683.9 -1829.5   3659.1
```

```
## allrand  6 3618.7 3644.0 -1803.3   3606.7 52.371      2  4.243e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(rsfi,allrand)
```

```
## Data: teams
## Models:
## rsfi: score ~ 1 + risk + (0 + risk | team_id.f)
## allrand: score ~ risk + (1 + risk | team_id.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rsfi     4 3747.5 3764.3 -1869.7   3739.5
## allrand  6 3618.7 3644.0 -1803.3   3606.7 132.79      2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

New Dataset: Try it Yourself

Read in data

```
#library(haven)
#
#test <- read_sav("popular2.sav")
#
#
#popular <- test %>%
#  select(pupil, class, Cextrav, popular) %>%
#  as_tibble()
#
#write_csv(popular, "popular.csv", na = "")

popular <- read_csv("popular.csv")
```

```
## Parsed with column specification:
## cols(
##   pupil = col_double(),
##   class = col_double(),
##   Cextrav = col_double(),
##   popular = col_double()
## )
```

Describe data

```
describe(popular)
```

```
##         vars    n  mean    sd median trimmed   mad   min    max range  skew
## pupil      1 2000 10.65  5.97  11.00   10.56  7.41  1.00  26.00  25.0  0.10
## class      2 2000 50.37 29.08  51.00   50.33 37.81  1.00 100.00  99.0  0.01
## Cextrav    3 2000  0.00  1.26  -0.21   -0.04  1.48 -4.21   4.79   9.0  0.37
## popular    4 2000  5.08  1.38   5.10    5.08  1.33  0.00   9.50   9.5 -0.05
##         kurtosis   se
## pupil      -1.06 0.13
## class      -1.22 0.65
## Cextrav     0.66 0.03
```

```
## popular    -0.06 0.03
```

Factor class

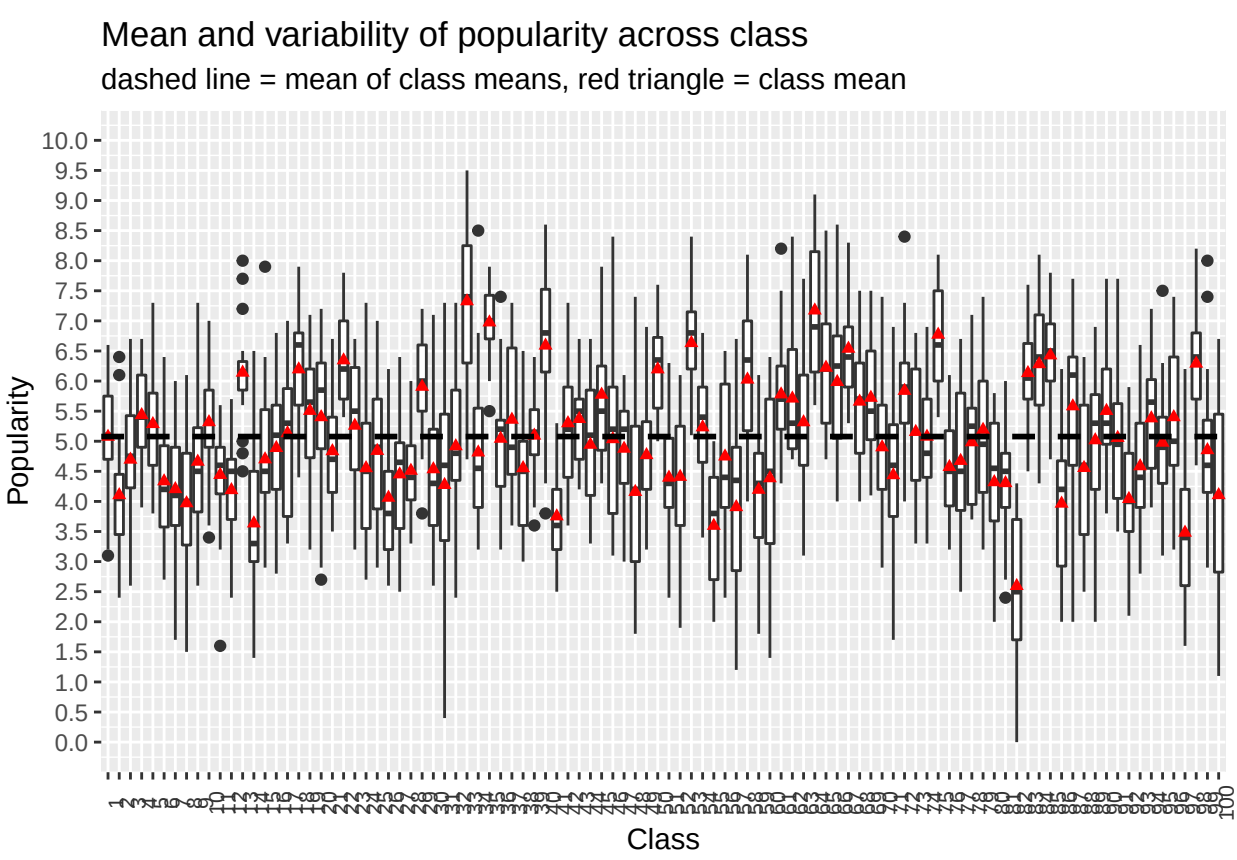
```
popular <- mutate(popular, class.f = factor(class))
```

Get Mean of Means

```
team_means  <- group_by(popular, class)
team_means  <- summarize(team_means, mean_pop = mean(popular))
meanofmeans <- summarize(team_means, meanofmeans = mean(mean_pop))
meanofmeans
```

```
## # A tibble: 1 x 1
##   meanofmeans
##         <dbl>
## 1        5.08
```

```
ggplot(data = popular, aes(x = class.f, y = popular)) +geom_boxplot() +stat_summary(aes(y = popular, gr
```



Random Intercept only model

```
# Random intercept
randint <- lmer(popular ~ 1 + (1|class.f), data = popular, REML = FALSE)
```

```
summary(randint)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: popular ~ 1 + (1 | class.f)
##    Data: popular
##
##      AIC      BIC   logLik deviance df.resid
##   6333.5   6350.3  -3163.7   6327.5     1997
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5662 -0.6983  0.0021  0.6758  3.3173
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class.f  (Intercept) 0.6945   0.8333
##  Residual             1.2218   1.1053
## Number of obs: 2000, groups:  class.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  5.07786    0.08696    58.4
```

ICC

```
# ICC
0.6945/(0.6945+1.2218)
```

```
## [1] 0.3624172
```

Pseudo R^2

```
# R^2
r.squaredGLMM(randint)
```

```
##      R2m       R2c
## [1,]   0 0.3624051
```

Random intercept Fixed Slope model

```
# Random intercept Fixed slope
rifs<-lmer(popular ~ 1 + Cextrav + (1|class.f),data=popular, REML = FALSE)
summary(rifs)
```

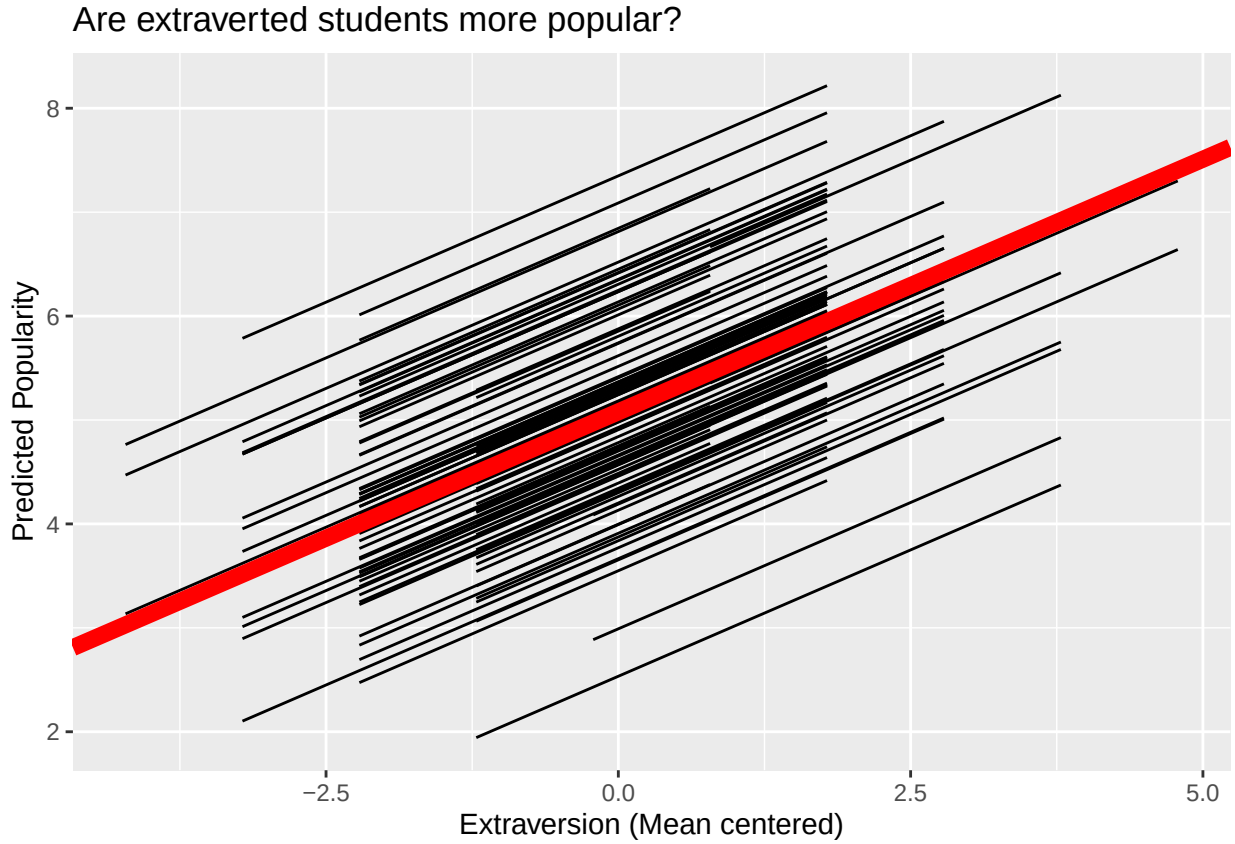
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: popular ~ 1 + Cextrav + (1 | class.f)
##    Data: popular
##
##      AIC      BIC   logLik deviance df.resid
##   5831.8   5854.2  -2911.9   5823.8     1996
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.0653 -0.7266  0.0167  0.7089  3.3597
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  class.f  (Intercept) 0.8314   0.9118
##  Residual             0.9299   0.9643
## Number of obs: 2000, groups:  class.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  5.07824    0.09372   54.18
## Cextrav      0.48620    0.02015   24.13
##
## Correlation of Fixed Effects:
##         (Intr)
## Cextrav 0.000
```

Plot

```
mod2.plot <- add_predictions(data=popular, model=rifs)

ggplot(data=mod2.plot, aes(x=Cextrav, y=pred, group=class.f))+
  geom_line()+
  geom_abline(intercept=5.07824, slope=0.48620, color="red", size=3)+
  labs(title="Are extraverted students more popular?", x="Extraversion (Mean centered)", y ="Predicted
```



ICC

```
# ICC
0.8314/(0.8314 + 0.9299)
```

```
## [1] 0.4720377
```

Pseudo R^2

```
# R^2
r.squaredGLMM(rifs)
```

```
##            R2m       R2c
## [1,] 0.1761901 0.5650694
```

Random Slope, Fixed Intercept model

```
# random slopes fixed intercepts
rsfi<- lmer(popular ~ 1 + Cextrav + (0 + Cextrav|class.f),data=popular, REML = FALSE)
summary(rsfi)
```

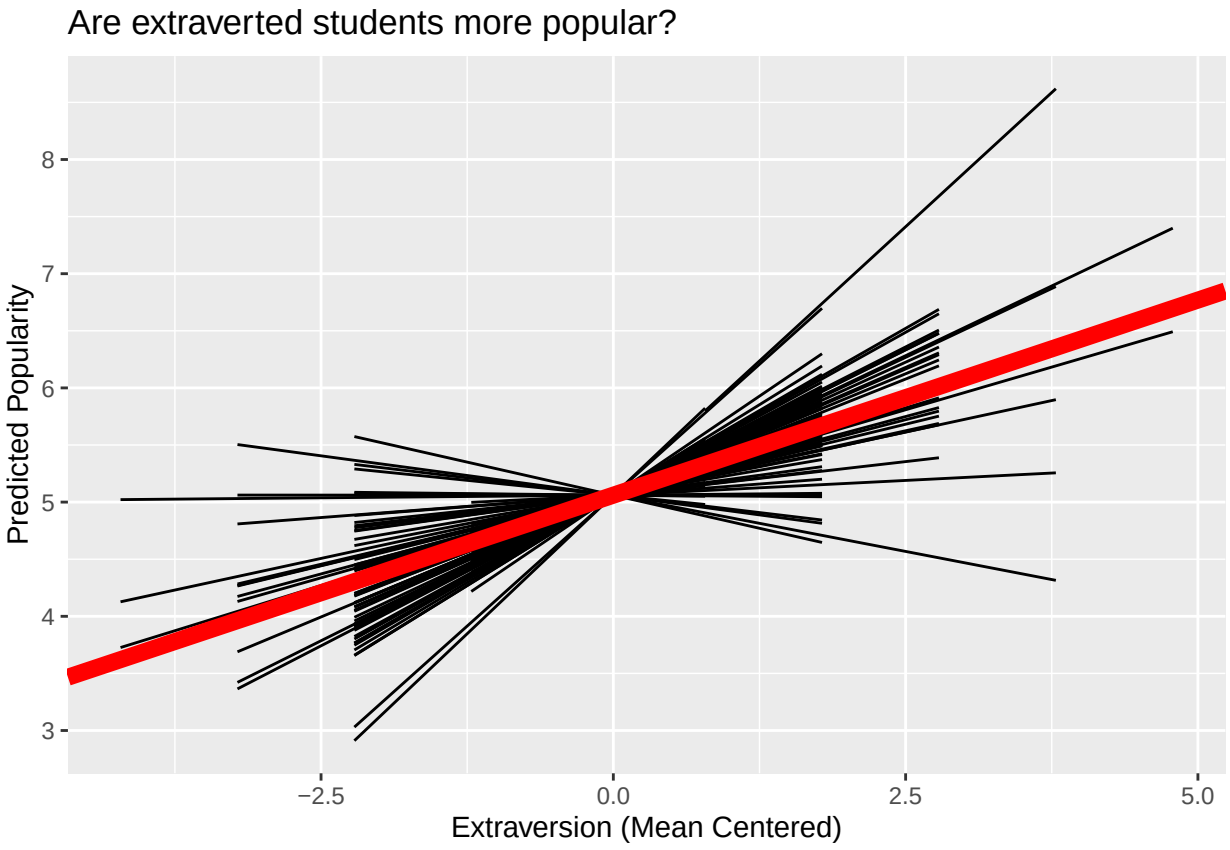
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: popular ~ 1 + Cextrav + (0 + Cextrav | class.f)
##    Data: popular
##
##      AIC      BIC   logLik deviance df.resid
##   6703.6   6726.0  -3347.8   6695.6     1996
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0024 -0.6851  0.0171  0.6630  3.3970
##
## Random effects:
##  Groups   Name    Variance Std.Dev.
##  class.f  Cextrav 0.08356  0.2891
##  Residual         1.59167  1.2616
## Number of obs: 2000, groups:  class.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  5.06055    0.02986 169.469
## Cextrav      0.34167    0.03816   8.955
##
## Correlation of Fixed Effects:
##         (Intr)
## Cextrav 0.053
```

Plot model

```
# Model without random intercept
mod3.plot <- add_predictions(data=popular, model=rsfi)

ggplot(data=mod3.plot, aes(x=Cextrav, y=pred, group=class.f))+
  geom_line()+
```

```
  geom_abline(intercept=5.06055, slope=0.34167, color="red", size=3)+
  labs(title="Are extraverted students more popular?", x="Extraversion (Mean Centered)", y ="Predicted
```



ICC

```
#ICC
0.08356/(0.08356 + 1.59167)
```

```
## [1] 0.04987972
```

Pseudo R^2

```
#R^2
r.squaredGLMM(rsfi)
```

```
##             R2m       R2c
## [1,] 0.09735908 0.1670142
```

All Random Model

```
# all random
allrand<- lmer(popular ~ Cextrav +  (1 + Cextrav|class.f),data=popular, REML = FALSE)
summary(allrand)
```

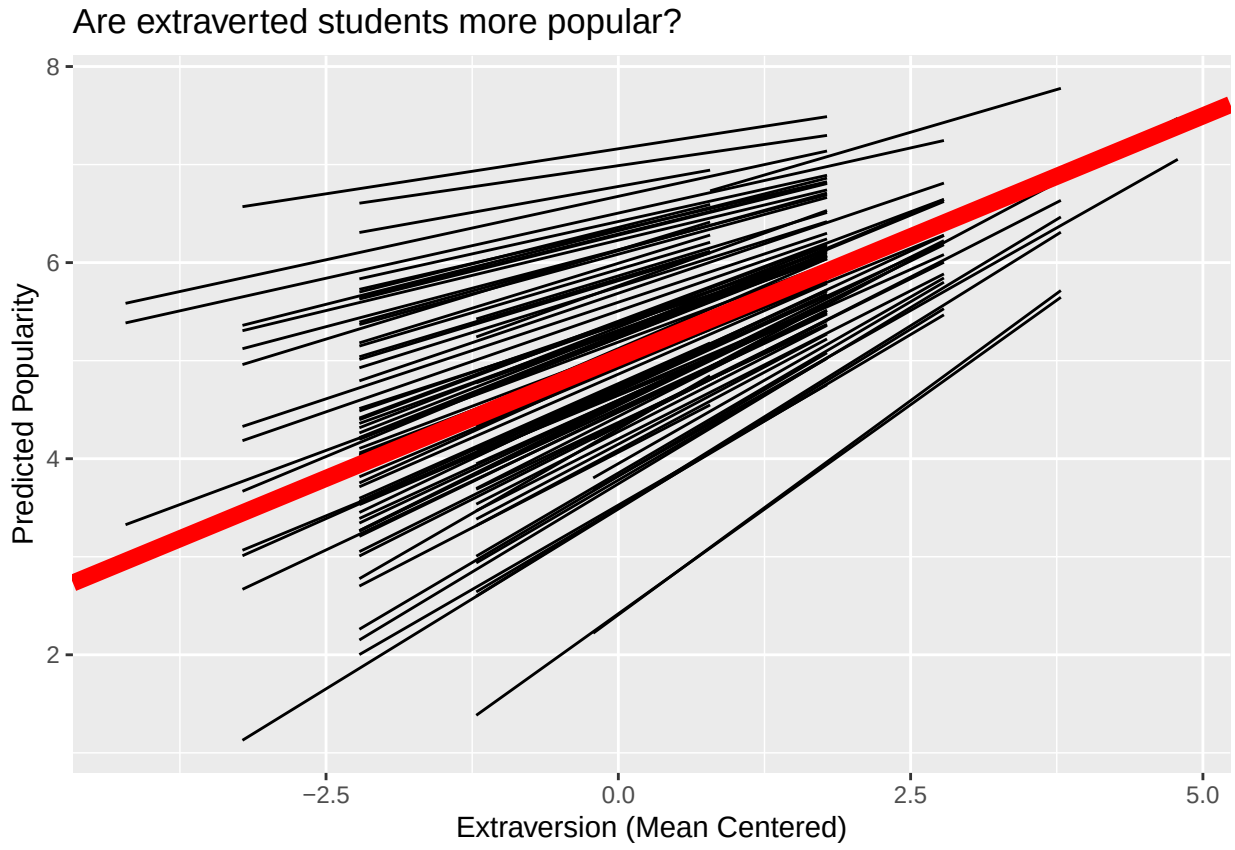
```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: popular ~ Cextrav + (1 + Cextrav | class.f)
```

```
##    Data: popular
##
##      AIC      BIC   logLik deviance df.resid
##   5782.7   5816.3  -2885.3   5770.7     1994
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2041 -0.7288  0.0139  0.6835  3.2224
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  class.f  (Intercept) 0.88225  0.9393
##           Cextrav     0.02522  0.1588   -0.89
##  Residual             0.89505  0.9461
## Number of obs: 2000, groups:  class.f, 100
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  5.03184    0.09652   52.13
## Cextrav      0.49290    0.02529   19.49
##
## Correlation of Fixed Effects:
##         (Intr)
## Cextrav -0.554
```

Plot

```
mod4.plot <- add_predictions(data=popular, model=allrand)

ggplot(data=mod4.plot, aes(x=Cextrav, y=pred, group=class.f))+
  geom_line()+
  geom_abline(intercept=5.03184, slope=0.49290, color="red", size=3)+
  labs(title="Are extraverted students more popular?", x="Extraversion (Mean Centered)", y ="Predicted
```



ICC

```
#ICC
0.88225+0.02522/(0.88225+0.02522+0.89505)
```

```
## [1] 0.8962415
```

Psuedo R^2

```
#R^2
r.squaredGLMM(allrand)
```

```
##            R2m       R2c
## [1,] 0.1756131 0.5940146
```

Pairwise comparison

randint vs all random

```
anova(randint,allrand)
```

```
## Data: popular
## Models:
## randint: popular ~ 1 + (1 | class.f)
## allrand: popular ~ Cextrav + (1 + Cextrav | class.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
```

```
## randint  3 6333.5 6350.3 -3163.7   6327.5
## allrand  6 5782.7 5816.3 -2885.3   5770.7 556.78      3  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

RIFS vs all random

```
anova(rifs,allrand)
```

```
## Data: popular
## Models:
## rifs: popular ~ 1 + Cextrav + (1 | class.f)
## allrand: popular ~ Cextrav + (1 + Cextrav | class.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rifs     4 5831.8 5854.2 -2911.9   5823.8
## allrand  6 5782.7 5816.3 -2885.3   5770.7 53.089      2  2.964e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

RSFI vs all random

```
anova(rsfi,allrand)
```

```
## Data: popular
## Models:
## rsfi: popular ~ 1 + Cextrav + (0 + Cextrav | class.f)
## allrand: popular ~ Cextrav + (1 + Cextrav | class.f)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rsfi     4 6703.6 6726.0 -3347.8   6695.6
## allrand  6 5782.7 5816.3 -2885.3   5770.7 924.93      2  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

OLD DATA

```
org <- read_csv("orgdata.csv")
```

```
## Parsed with column specification:
## cols(
##   nemploy = col_double(),
##   nmale = col_double(),
##   Country = col_double(),
##   HPD = col_double(),
##   HCollect = col_double(),
##   HMF = col_double(),
##   HUA = col_double(),
##   HLong = col_double(),
##   Hind = col_double(),
##   SelfRating = col_double(),
##   PeerRating = col_double(),
##   Pay = col_double(),
##   Develop = col_double()
## )
```

Describe

```
describe(org)
```

```
##            vars   n    mean       sd median trimmed    mad  min    max  range
## nemploy       1 466 2256.95 22542.40    400  607.60 406.23   10 480000 479990
## nmale         2 466 1112.51  9588.87    225  353.09 252.04    4 200000 199996
## Country       3 466    9.62     4.71     10    9.67   4.45    1     19     18
## HPD           4 466   47.63    19.05     46   47.43  16.31   11     94     83
## HCollect      5 466   58.49    65.95     72   62.02  44.48 -126    182    308
## HMF           6 466   55.59    23.04     61   57.16  23.72    5     88     83
## HUA           7 466   66.62    21.68     75   67.74  16.31   23    100     77
## HLong         8 466   51.53    16.74     53   51.37  11.86   21     87     66
## Hind          9 466   51.56    16.29     50   51.87  26.69   20     78     58
## SelfRating   10 451    2.28     1.09      3    2.48   0.00    0      3      3
## PeerRating   11 423    0.46     0.90      0    0.23   0.00    0      3      3
## Pay          12 465    0.78     0.87      1    0.92   0.00   -9      1     10
## Develop      13 465    0.82     0.85      1    0.97   0.00   -9      1     10
##              skew kurtosis      se
## nemploy    20.49   430.06 1044.26
## nmale      19.47   395.87  444.20
## Country    -0.06    -0.77    0.22
## HPD         0.29    -0.07    0.88
## HCollect   -0.61     0.27    3.06
## HMF        -0.54    -0.41    1.07
## HUA        -0.45    -0.87    1.00
## HLong      -0.01    -0.27    0.78
## Hind       -0.15    -1.28    0.75
## SelfRating -1.19    -0.15    0.05
## PeerRating  1.89     2.32    0.04
## Pay        -9.36   101.52    0.04
## Develop    -9.88   109.58    0.04
```

Create measure of gender balance in workforce

```
org <- mutate(org, percmale = nmale/nemploy)
```

Models

```
# Random intercept
randint <- lmer(Develop ~ 1 + (1|HCollect), data = org, REML = FALSE)
summary(randint)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Develop ~ 1 + (1 | HCollect)
##    Data: org
##
##      AIC      BIC   logLik deviance df.resid
##   1178.2   1190.6   -586.1   1172.2      462
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.5216   0.1811   0.1916   0.2309   0.2818
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  HCollect (Intercept) 0.004702 0.06857
##  Residual             0.724004 0.85088
## Number of obs: 465, groups:  HCollect, 19
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.82442    0.04338      19
```

```
# Random intercept Fixed slope
rifs<-lmer(Develop~ 1 + percmale + (1|HCollect),data=org, REML = FALSE)
summary(rifs)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Develop ~ 1 + percmale + (1 | HCollect)
##    Data: org
##
##      AIC      BIC   logLik deviance df.resid
##   1180.1   1196.7   -586.1   1172.1      461
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.5281   0.1791   0.1925   0.2262   0.2992
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  HCollect (Intercept) 0.004331 0.06581
##  Residual             0.724258 0.85103
## Number of obs: 465, groups:  HCollect, 19
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.80259    0.11843   6.777
## percmale     0.03645    0.18656   0.195
##
## Correlation of Fixed Effects:
##          (Intr)
## percmale -0.931
```

```
# random slopes fixed intercepts
rsfi<- lmer(Develop ~ 1 + percmale + (0 + percmale|HCollect),data=org, REML = FALSE)
summary(rsfi)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Develop ~ 1 + percmale + (0 + percmale | HCollect)
##    Data: org
##
##      AIC      BIC   logLik deviance df.resid
##   1180.3   1196.8   -586.1   1172.3      461
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.5183   0.1856   0.2026   0.2229   0.2518
##
## Random effects:
```

```
##  Groups   Name     Variance Std.Dev.
##  HCollect percmale 0.007365 0.08582
##  Residual          0.725696 0.85188
## Number of obs: 465, groups:  HCollect, 19
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.79471    0.11655   6.819
## percmale     0.04418    0.18732   0.236
##
## Correlation of Fixed Effects:
##          (Intr)
## percmale -0.934
```

```
# all random
allrand<- lmer(Develop ~ percmale +   (1 + percmale|HCollect),data=org, REML = FALSE)
summary(allrand)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Develop ~ percmale + (1 + percmale | HCollect)
##    Data: org
##
##      AIC      BIC   logLik deviance df.resid
##   1183.8   1208.7   -585.9   1171.8      459
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.4755   0.1785   0.1921   0.2084   0.4104
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  HCollect (Intercept) 0.04838  0.2199
##           percmale    0.09590  0.3097   -0.97
##  Residual             0.72059  0.8489
## Number of obs: 465, groups:  HCollect, 19
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.81839    0.13118   6.239
## percmale     0.01302    0.20395   0.064
##
## Correlation of Fixed Effects:
##          (Intr)
## percmale -0.946
```

ICC

variance / (variance + residual)

```
library(sjstats)

performance::icc(randint)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.006
```

```
##   Conditional ICC: 0.006
```

```
performance::icc(rifs)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.006
##   Conditional ICC: 0.006
```

```
performance::icc(rsfi)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.004
##   Conditional ICC: 0.004
```

```
performance::icc(allrand)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.011
##   Conditional ICC: 0.011
```

R^2

```
r.squaredGLMM(randint)
```

```
##      R2m         R2c
## [1,]   0 0.006452812
```

```
r.squaredGLMM(rifs)
```

```
##               R2m         R2c
## [1,] 8.303259e-05 0.006026336
```

```
r.squaredGLMM(rsfi)
```

```
##               R2m         R2c
## [1,] 0.0001220062 0.004085227
```

```
r.squaredGLMM(allrand)
```

```
##               R2m        R2c
## [1,] 1.059734e-05 0.01105436
```

Pairwise comparison

```
anova(rifs,allrand)
```

```
## Data: org
## Models:
## rifs: Develop ~ 1 + percmale + (1 | HCollect)
## allrand: Develop ~ percmale + (1 + percmale | HCollect)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rifs     4 1180.1 1196.7 -586.07   1172.1
## allrand  6 1183.8 1208.7 -585.92   1171.8 0.2962      2     0.8623
```

```
anova(randint,allrand)
```

```
## Data: org
## Models:
## randint: Develop ~ 1 + (1 | HCollect)
## allrand: Develop ~ percmale + (1 + percmale | HCollect)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## randint  3 1178.2 1190.6 -586.09   1172.2
## allrand  6 1183.8 1208.7 -585.92   1171.8 0.3332      3     0.9537
```

```
anova(rsfi,allrand)
```

```
## Data: org
## Models:
## rsfi: Develop ~ 1 + percmale + (0 + percmale | HCollect)
## allrand: Develop ~ percmale + (1 + percmale | HCollect)
##         Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## rsfi     4 1180.3 1196.8 -586.13   1172.3
## allrand  6 1183.8 1208.7 -585.92   1171.8 0.4233      2     0.8093
```