

# Formative

Zehao Qian

2024-02-28

## Part 0: Prepare for the Runtime and Dataset

gridExtra is a good library for combining multiple ggplots graphs in one graph, while par() can't make it.

```
# -----  
## clear the environment var area  
# rm(list = ls())  
## clear all plots  
# graphics.off()  
## clear the console area  
# cat("\014")  
# -----  
# install.packages("gridExtra")  
# -----  
require(lme4)  
require(lmerTest)  
require(ggplot2)  
require(sjPlot)  
require(gridExtra)
```

Download data set from GitHub and show the first lines.

```
CRT <- read.csv("https://andygolightly.github.io/teaching/MATH43515/CRT.csv", header=TRUE)  
head(CRT)
```

```
##   Pupil School Posttest Intervention Pretest FSM class  
## 1     1     1      16             1      1  0    1  
## 2     2     1      13             1      4  1    1  
## 3     3     1      18             1      5  1    1  
## 4     4     1      14             1      4  1    1  
## 5     5     1      25             1      5  1    1  
## 6     6     1      13             1      2  0    1
```

```
# dim(CRT)
```

---

## Part 1: Introduction

- randomized controlled trials and cluster randomized trials

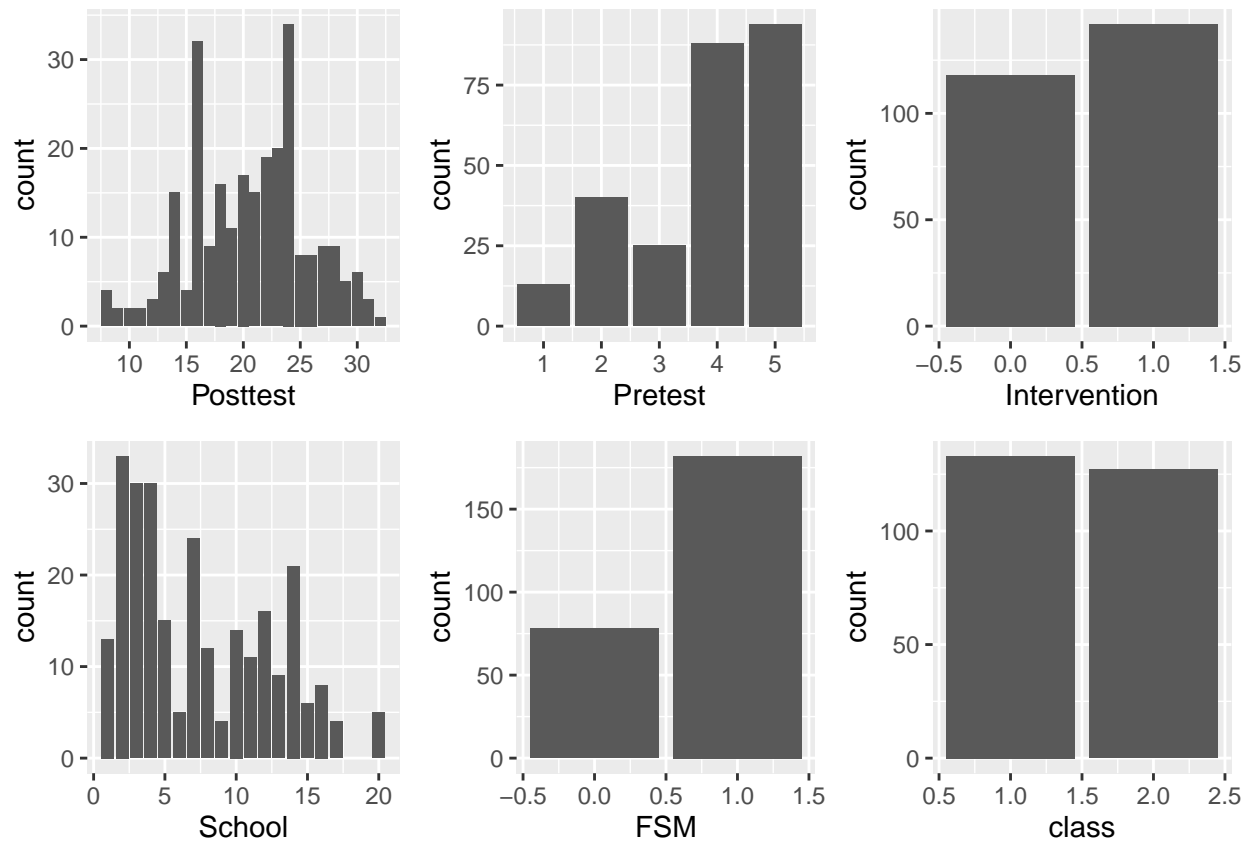
- Intro to the CRT data set
  - Pupil: anonymized student ID;
  - School: anonymized school ID {1, ..., 20};
  - Class: anonymized class ID {1, 2};
  - Intervention: Intervention indicator (0=control; 1=treatment);
  - FSM: pupil eligible for free school meal (0=not eligible; 1 =eligible)
  - Pretest: A pre-test score for each pupil;
  - Posttest: A post-test score for each pupil (**response**).
- Histograms / Bar Plots / EDA with ggplots
- Check missing values
- Target: predict the Post-test

```
summary(CRT)
```

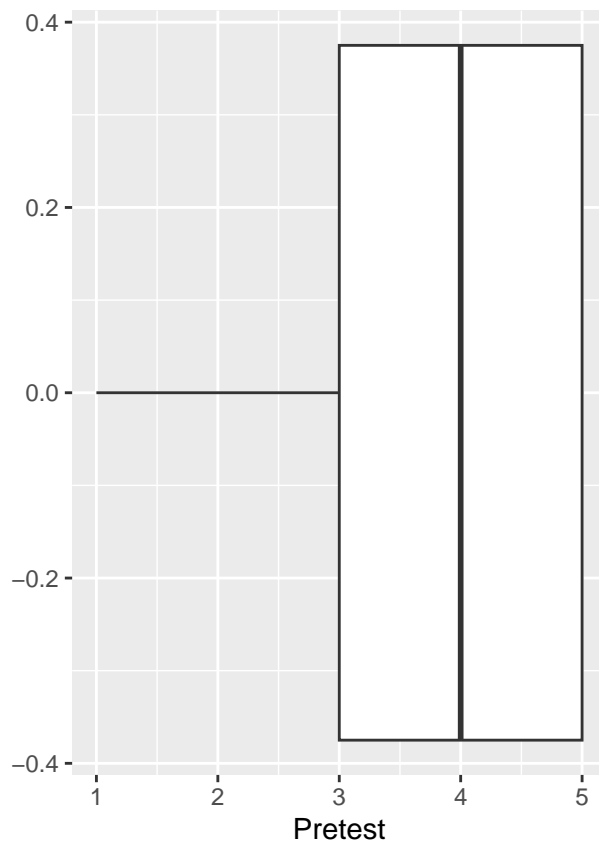
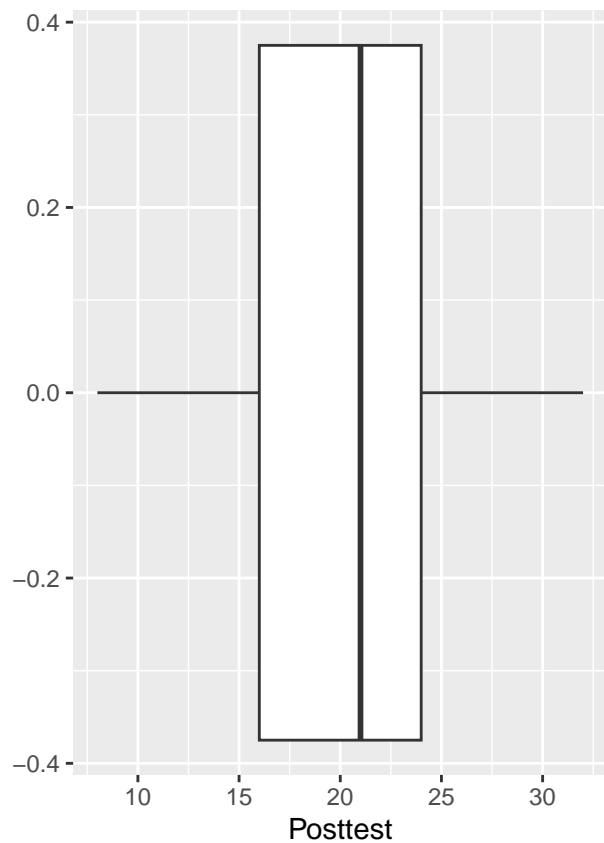
```
##      Pupil      School      Posttest      Intervention
## Min.   : 1.00   Min.   : 1.000   Min.   : 8.00   Min.   :0.0000
## 1st Qu.: 65.75   1st Qu.: 3.000   1st Qu.:16.00   1st Qu.:0.0000
## Median :130.50   Median : 7.000   Median :21.00   Median :1.0000
## Mean   :130.56   Mean    : 7.477   Mean    :20.54   Mean    :0.5462
## 3rd Qu.:195.25   3rd Qu.:12.000   3rd Qu.:24.00   3rd Qu.:1.0000
## Max.   :263.00   Max.    :20.000   Max.    :32.00   Max.    :1.0000
##      Pretest      FSM      class
## Min.   :1.000   Min.   :0.0   Min.   :1.000
## 1st Qu.:3.000   1st Qu.:0.0   1st Qu.:1.000
## Median :4.000   Median :1.0   Median :1.000
## Mean   :3.808   Mean    :0.7   Mean    :1.488
## 3rd Qu.:5.000   3rd Qu.:1.0   3rd Qu.:2.000
## Max.   :5.000   Max.    :1.0   Max.    :2.000
```

```
# -----
bar.Posttest = ggplot(data = CRT, aes(Posttest)) +
  geom_bar()
bar.Pretest = ggplot(data = CRT, aes(Pretest)) +
  geom_bar()
bar.Intervention = ggplot(data = CRT, aes(Intervention)) +
  geom_bar()
bar.School = ggplot(data = CRT, aes(School)) +
  geom_bar()
bar.FSM = ggplot(data = CRT, aes(FSM)) +
  geom_bar()
bar.class = ggplot(data = CRT, aes(class)) +
  geom_bar()
# -----
# boxplot
boxplot.Posttest = ggplot(data = CRT, aes(Posttest)) +
  geom_boxplot(outlier.colour = "red", outlier.shape = 1)
boxplot.Pretest = ggplot(data = CRT, aes(Pretest)) +
  geom_boxplot(outlier.colour = "red", outlier.shape = 1)
# -----
# put these bar charts together
```

```
grid.arrange(bar.Posttest,
              bar.Pretest,
              bar.Intervention,
              bar.School,
              bar.FSM,
              bar.class,
              ncol = 3)
```

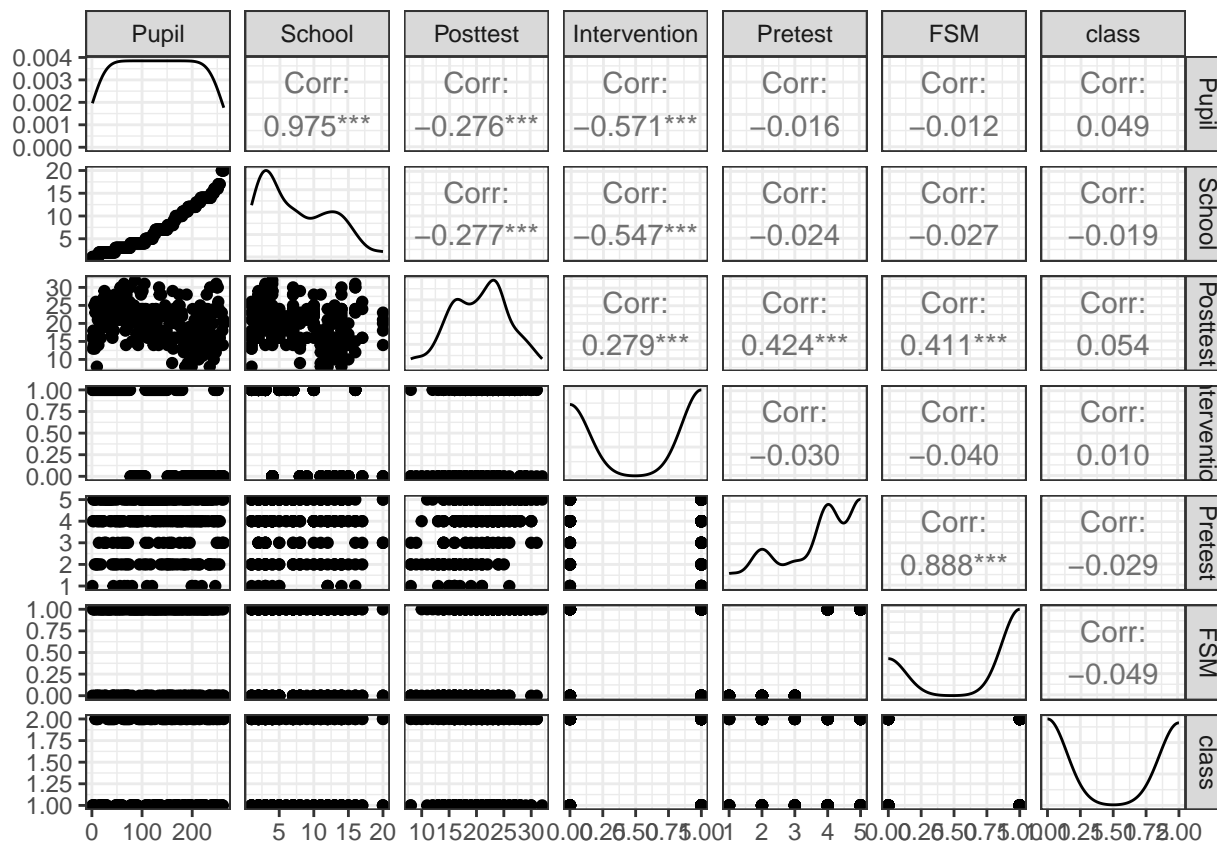


```
grid.arrange(boxplot.Posttest,
              boxplot.Pretest,
              ncol = 2)
```



```
# -----
# remove the plot cache memory
remove(bar.Posttest)
remove(bar.Pretest)
remove(bar.Intervention)
remove(bar.School)
remove(bar.FSM)
remove(bar.class)
remove(boxplot.Posttest)
remove(boxplot.Pretest)
```

```
# Correlation between
library("GGally")
ggpairs(CRT)+theme_bw()
```



```
## Not good to read and analyse
# tapply(CRT$School, CRT$class, table)
```

## Part 2: Methods

## Part 3: Analysis

```
# empty model / intercept-only model
Model.0 = lmer(Posttest ~ 1 + (1 | School)
               + (1 | School:class),
               data = CRT)

# summary(Model.0)
# -----
# REssummary <- as.data.frame(VarCorr(Model.0))
# REssummary
# summary(Model.0)$varcor
# -----
Model.1 = lmer(Posttest ~ Pretest + (1 | School)
               + (1 | School:class),
               data = CRT)

# summary(Model.1)
# -----
anova(Model.0)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ (1 | School) + (1 | School:class)
##           npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>           4 -768.73 1545.5
## (1 | School)       3 -772.30 1550.6 7.1322  1  0.007571 **
## (1 | School:class) 3 -768.93 1543.9 0.3933  1  0.530550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ranova(Model.1)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ Pretest + (1 | School) + (1 | School:class)
##           npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>           5 -732.59 1475.2
## (1 | School)       4 -736.23 1480.5 7.2737  1  0.006997 **
## (1 | School:class) 4 -733.93 1475.9 2.6846  1  0.101321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.1, Model.0)
```

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

## Data: CRT
## Models:
## Model.0: Posttest ~ 1 + (1 | School) + (1 | School:class)
## Model.1: Posttest ~ Pretest + (1 | School) + (1 | School:class)
##           npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## Model.0      4 1546.5 1560.7 -769.24  1538.5
## Model.1      5 1474.8 1492.6 -732.41  1464.8 73.66  1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Model.2 = lmer(Posttest ~ Pretest + (Pretest | School)
               + (1 | School:class),
               data = CRT)
ranova(Model.2)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ Pretest + (Pretest | School) + (1 | School:class)
##           npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>           7 -732.33 1478.7
## Pretest in (Pretest | School) 5 -732.59 1475.2 0.52937  2  0.76745
## (1 | School:class)           6 -733.77 1479.5 2.88093  1  0.08963 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.2, Model.1)
```

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

```
## Data: CRT
```

```
## Models:
```

```
## Model.1: Posttest ~ Pretest + (1 | School) + (1 | School:class)
```

```
## Model.2: Posttest ~ Pretest + (Pretest | School) + (1 | School:class)
```

```
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
```

```
## Model.1      5 1474.8 1492.6 -732.41   1464.8
```

```
## Model.2      7 1478.4 1503.3 -732.18   1464.4 0.4659  2    0.7922
```

```
Model.3 = lmer(Posttest ~ Pretest + (1 | School)
               + (Pretest | School:class),
               data = CRT)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
ranova(Model.3)
```

```
## ANOVA-like table for random-effects: Single term deletions
```

```
##
```

```
## Model:
```

```
## Posttest ~ Pretest + (1 | School) + (Pretest | School:class)
```

```
##          npar logLik    AIC    LRT Df Pr(>Chisq)
```

```
## <none>          7 -730.79 1475.6
```

```
## (1 | School)      6 -735.43 1482.8 9.2654  1  0.002335 **
```

```
## Pretest in (Pretest | School:class)  5 -732.59 1475.2 3.5970  2  0.165546
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.3, Model.1)
```

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

```
## Data: CRT
```

```
## Models:
```

```
## Model.1: Posttest ~ Pretest + (1 | School) + (1 | School:class)
```

```
## Model.3: Posttest ~ Pretest + (1 | School) + (Pretest | School:class)
```

```
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
```

```
## Model.1      5 1474.8 1492.6 -732.41   1464.8
```

```
## Model.3      7 1475.6 1500.5 -730.78   1461.6 3.2727  2    0.1947
```

```
Model.4 = lmer(Posttest ~ Pretest + FSM + (FSM | School)
               + (1 | School:class),
               data = CRT)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
ranova(Model.4)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ Pretest + FSM + (FSM | School) + (1 | School:class)
##
##          npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>          8 -729.98 1476.0
## FSM in (FSM | School)    6 -730.50 1473.0 1.0355 2    0.59587
## (1 | School:class)      7 -731.74 1477.5 3.5268 1    0.06038 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.4, Model.1)
```

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

## Data: CRT
## Models:
## Model.1: Posttest ~ Pretest + (1 | School) + (1 | School:class)
## Model.4: Posttest ~ Pretest + FSM + (FSM | School) + (1 | School:class)
##
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## Model.1    5 1474.8 1492.6 -732.41  1464.8
## Model.4    8 1477.7 1506.2 -730.87  1461.7 3.0975 3    0.3768
```

```
Model.5 = lmer(Posttest ~ Pretest + Intervention + (1 | School)
               + (1 | School:class),
               data = CRT)
ranova(Model.5)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ Pretest + Intervention + (1 | School) + (1 | School:class)
##
##          npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>          6 -728.74 1469.5
## (1 | School)      5 -731.09 1472.2 4.7040 1    0.03009 *
## (1 | School:class) 5 -730.06 1470.1 2.6459 1    0.10381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.5, Model.1)
```

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

## Data: CRT
## Models:
## Model.1: Posttest ~ Pretest + (1 | School) + (1 | School:class)
## Model.5: Posttest ~ Pretest + Intervention + (1 | School) + (1 | School:class)
##
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
```



```
## Model.1      5 1474.8 1492.6 -732.41   1464.8
## Model.5      6 1470.9 1492.3 -729.46   1458.9 5.9035  1    0.01511 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Model.6 = lmer(Posttest ~ Pretest + Intervention + (1 | School)
               + (Intervention | School:class),
               data = CRT)
ranova(Model.6)
```

```
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## Posttest ~ Pretest + Intervention + (1 | School) + (Intervention | School:class)
##
##          npar  logLik   AIC    LRT Df
## <none>          8 -728.74 1473.5
## (1 | School)      7 -731.07 1476.1 4.6678  1
## Intervention in (Intervention | School:class)  6 -728.74 1469.5 0.0025  2
##
##          Pr(>Chisq)
## <none>
## (1 | School)      0.03073 *
## Intervention in (Intervention | School:class)  0.99873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(Model.6, Model.5)
```

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

```
## Data: CRT
## Models:
## Model.5: Posttest ~ Pretest + Intervention + (1 | School) + (1 | School:class)
## Model.6: Posttest ~ Pretest + Intervention + (1 | School) + (Intervention | School:class)
##          npar   AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## Model.5      6 1470.9 1492.3 -729.46   1458.9
## Model.6      8 1474.9 1503.4 -729.46   1458.9 0.0019  2    0.9991
```

```
Model.7 = lmer(Posttest ~ Pretest + Intervention + (Intervention | School)
               + (1 | School:class),
               data = CRT)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

```
## Warning: Model failed to converge with 1 negative eigenvalue: -3.8e-04
```

```
ranova(Model.7)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## unable to evaluate scaled gradient  
  
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues  
  
## Warning: Model failed to converge with 1 negative eigenvalue: -4.2e-05  
  
## ANOVA-like table for random-effects: Single term deletions  
##  
## Model:  
## Posttest ~ Pretest + Intervention + (Intervention | School) + (1 | School:class)  
##  
##          npar  logLik   AIC    LRT Df  
## <none>          8 -728.70 1473.4  
## Intervention in (Intervention | School)  6 -728.74 1469.5 0.06634  2  
## (1 | School:class)          7 -730.03 1474.1 2.65001  1  
##  
##          Pr(>Chisq)  
## <none>  
## Intervention in (Intervention | School)    0.9674  
## (1 | School:class)          0.1035
```

```
anova(Model.7, Model.5)
```

```
## refitting model(s) with ML (instead of REML)  
  
## Data: CRT  
## Models:  
## Model.5: Posttest ~ Pretest + Intervention + (1 | School) + (1 | School:class)  
## Model.7: Posttest ~ Pretest + Intervention + (Intervention | School) + (1 | School:class)  
##  
##          npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)  
## Model.5     6 1470.9 1492.3 -729.46  1458.9  
## Model.7     8 1474.9 1503.4 -729.44  1458.9 0.0472  2    0.9767
```

```
require(performance)
```

```
##      performance
```

```
# https://easystats.github.io/performance/reference/icc.html  
icc(Model.0)
```

```
## # Intraclass Correlation Coefficient  
##  
##      Adjusted ICC: 0.258  
##      Unadjusted ICC: 0.258
```

```
icc(Model.1)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.360
##     Unadjusted ICC: 0.292
```

```
icc(Model.2)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.363
##     Unadjusted ICC: 0.292
```

```
icc(Model.3)
```

```
## Warning: Can't compute random effect variances. Some variance components equal
## zero. Your model may suffer from singularity (see `?lme4::isSingular`
## and `?performance::check_singularity`).
## Solution: Respecify random structure! You may also decrease the
## `tolerance` level to enforce the calculation of random effect variances.
```

```
## [1] NA
```

```
icc(Model.4)
```

```
## Warning: Can't compute random effect variances. Some variance components equal
## zero. Your model may suffer from singularity (see `?lme4::isSingular`
## and `?performance::check_singularity`).
## Solution: Respecify random structure! You may also decrease the
## `tolerance` level to enforce the calculation of random effect variances.
```

```
## [1] NA
```

```
icc(Model.5)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.294
##     Unadjusted ICC: 0.216
```

```
icc(Model.6)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.293
##     Unadjusted ICC: 0.216
```

```
icc(Model.7)
```

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.298
##     Unadjusted ICC: 0.220
```

## Part 4: Discussion of results

---

### References

Evaluating Intervention Programs with a Pretest-Posttest Design: A Structural Equation Modeling Approach

### Word count

```
# install.packages("devtools")
# devtools::install_github("benmarwick/wordcountaddin", type = "source", dependencies = TRUE)
require(wordcountaddin)
word_count()
```

```
## [1] 151
```

```
text_stats()
```

Method	koRpus	stringi
Word count	151	133
Character count	1031	1142
Sentence count	24	Not available
Reading time	0.8 minutes	0.7 minutes