

2018.DA.5

Daniel Pinedo

December 13th, 2018

Load libraries

A local school district realized that none of their elementary teachers taught science (true story). In an attempt to remedy the situation, they developed a professional development course designed to teach their teachers science and how to teach science to their classes. This district sent their teachers to this course in cohorts. In order to determine if the Professional Development course was working they had the first cohort of teachers complete a survey which asked them to rate, among other things, how comfortable they were teaching science to their elementary school class on a scale of 1 to 7. This survey was administered just before the Professional Development course began, during the course, and one month after the course was completed. A CGU student begged them to get a comparison group and so, the same survey was administered to the second cohort of teachers at the same time as it was administered to the first cohort of teachers. This second cohort still has not received the PD course.

Import data

```
science <- read.csv("https://www.dropbox.com/s/2vt8wls2f1yckfq/PSY.308b.DA5.csv?dl=1")
science$Subject <- as.factor(science$Subject)
head(science)

## Subject condition time value
## 1    1 treatment pre    4
## 2    2 treatment pre    1
## 3    3 treatment pre    2
## 4    4 treatment pre    6
## 5    5 treatment pre    4
## 6    6 treatment pre    1

#order IV's correctly
levels(science$time)

## [1] "during" "post"  "pre"
```

```

science$time <- factor(science$time, levels = c("pre", "during", "post"))
levels(science$time)

## [1] "pre" "during" "post"

levels(science$condition)

## [1] "control" "treatment"

# Assumptions:

# cast to wide format from long format [Control and Treatment = between, Time = Value]
science.wide <- cast(science, Subject + condition ~ time, value = "value")

# Normal distribution for RM variable - calculate difference scores
science.wide$pre_during <- (science.wide$pre - science.wide$during)
science.wide$pre_post <- (science.wide$pre - science.wide$post)
science.wide$during_post <- (science.wide$during - science.wide$post)

message("RM descriptives")

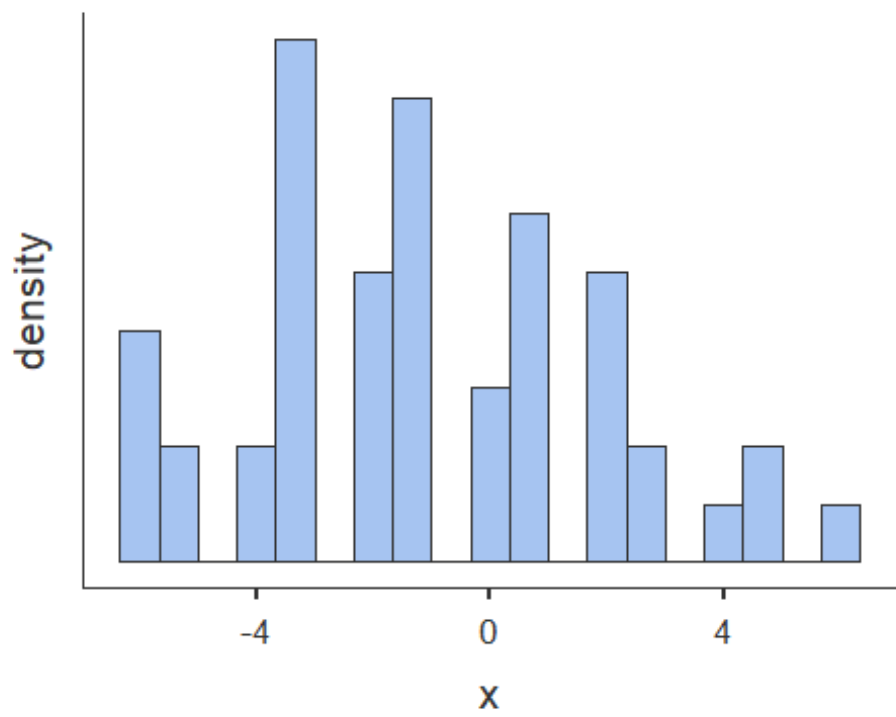
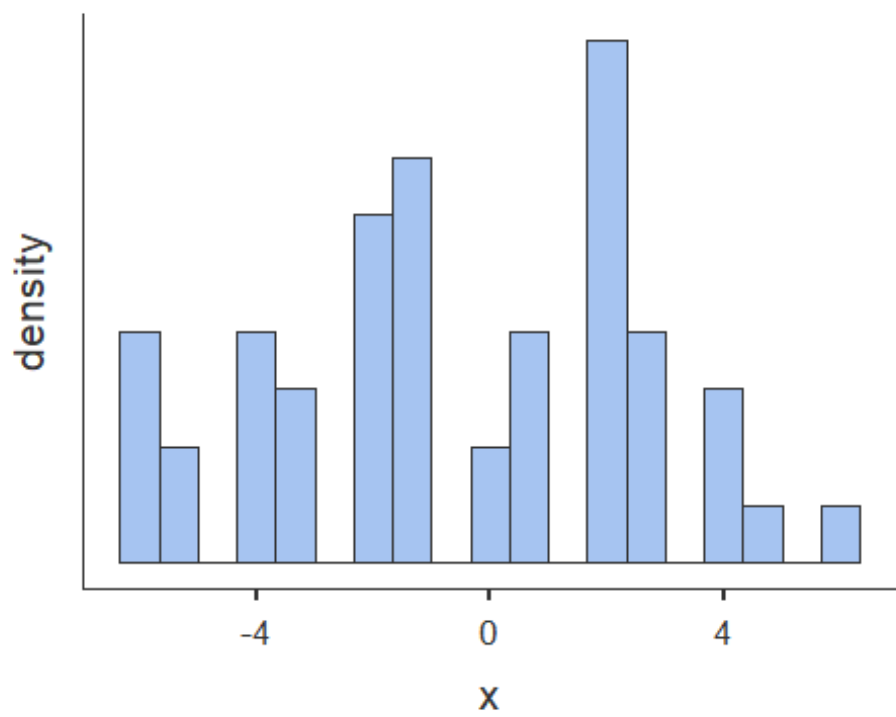
## RM descriptives

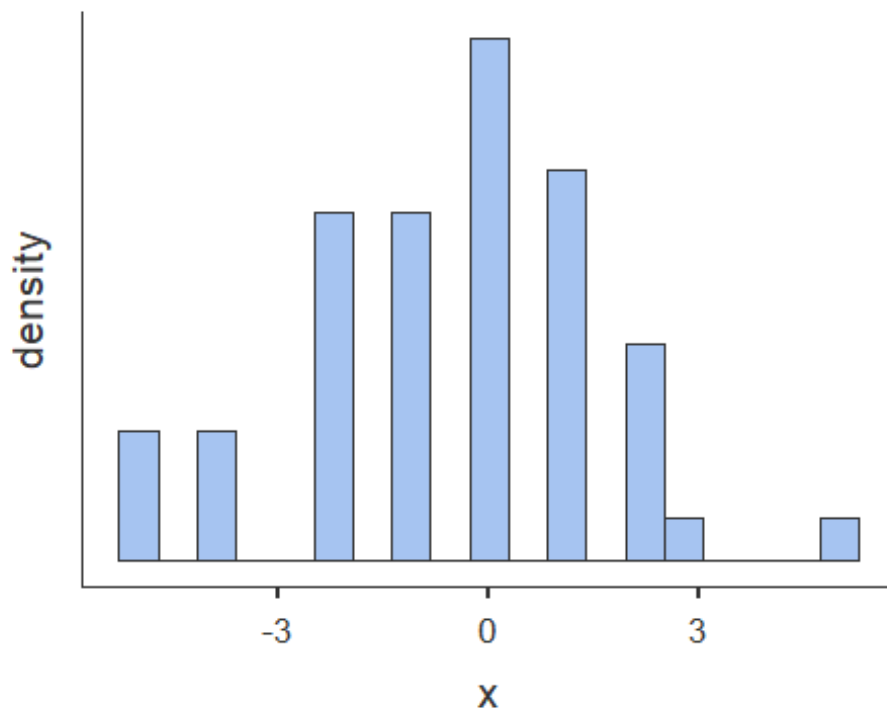
descriptives(science.wide, vars = c('pre_during', 'pre_post', 'during_post'), skew = TRUE, kurt =
TRUE, hist = TRUE)

##
## DESCRIPTIVES
##
## Descriptives
## -----
##               pre_during  pre_post  during_post
## -----
## N                50      50      50
## Missing            0       0       0
## Mean             -0.420  -0.900  -0.480
## Median            -1.00   -1.00    0.00

```

##	Minimum	-6	-6	-5
##	Maximum	6	6	5
##	Skewness	-0.110	0.255	-0.287
##	Std. error skewness	0.337	0.337	0.337
##	Kurtosis	-0.869	-0.395	0.464
##	Std. error kurtosis	0.662	0.662	0.662
##	-----			





#Normal distribution for BS variable

#Need to find the average time across RM conditions for each BS condition before being able to check the distribution with descriptives.

```
science.wide$averagevalue <- (science.wide$pre + science.wide$during + science.wide$post) /
3
```

```
message("BS descriptives")
```

```
## BS descriptives
```

```
descriptives(science.wide, vars = c('averagevalue'), splitBy = c('condition'), skew = TRUE, kurt
= TRUE, hist = TRUE)
```

```
##
```

```
## DESCRIPTIVES
```

```
##
```

```
## Descriptives
```

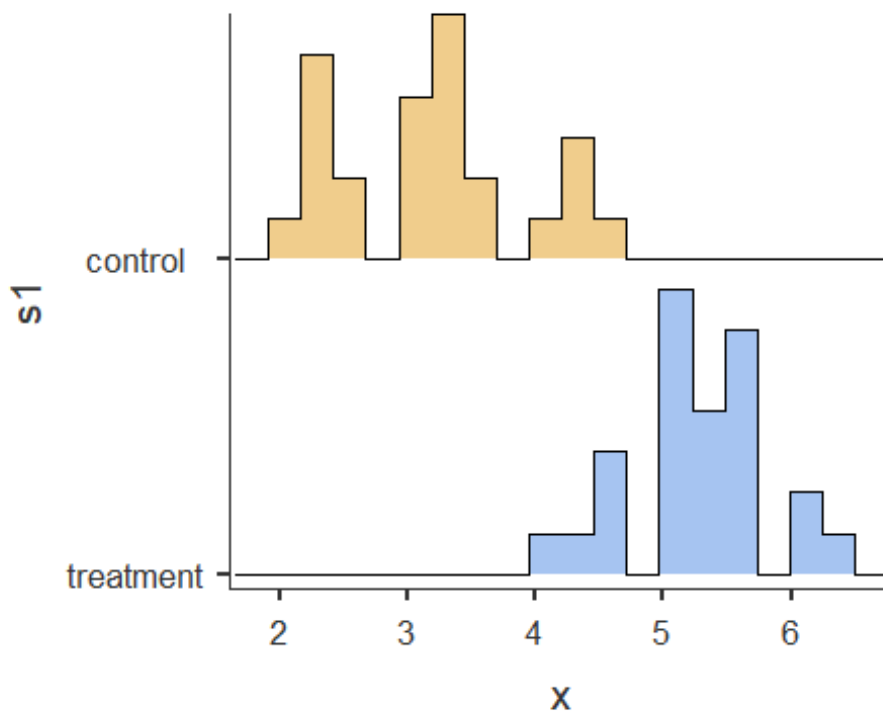
```
## -----
```

```
##           condition  averagevalue
```

```

## -----
## N          control      25
##          treatment      25
## Missing    control      0
##          treatment      0
## Mean       control      3.20
##          treatment      5.24
## Median     control      3.33
##          treatment      5.33
## Minimum    control      2.00
##          treatment      4.00
## Maximum    control      4.67
##          treatment      6.33
## Skewness   control      0.329
##          treatment      -0.181
## Std. error skewness control      0.464
##          treatment      0.464
## Kurtosis   control      -0.699
##          treatment      -0.0603
## Std. error kurtosis control      0.902
##          treatment      0.902
## -----

```



Levene's Test of Homogeneity of Variance for BS variable

```
leveneTest(science.wide$averagevalue, science.wide$condition, center = mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
```

```
##      Df F value Pr(>F)
```

```
## group 1  1.9823 0.1656
```

```
##      48
```

#dfa, dfbetween = # of groups - 1 OR a - 1

#dfb, dfwithin = # of conditions - 1 OR b - 1

#dfAxB = (a - 1)(b - 1)

#dfS/A = a(n - 1) ---> n = # in each group; error term for between subjects condition that is systematic

#dfTOTAL = (a)(b)(n - 1)

#add bs and bsTerms

```
model.rm <- anovaRM(data = science.wide,
```

```
rm = list(list(label = 'Time',
```

```

        levels = c('pre', 'during', 'post'))),
rmCells = list(list(measure = 'pre', cell = 'pre'),
               list(measure = 'during', cell = 'during'),
               list(measure = 'post', cell = 'post')),
rmTerms = list('Time'),
bs = 'condition',
bsTerms = list('condition'),
effectSize = c('partEta'),
spherTests = TRUE,
spherCorr = c('none','GG'),
postHoc = list('Time'),
postHocCorr = 'holm',
emMeans = list(NULL))

```

NOTE: Results may be misleading due to involvement in interactions

model.rm

##

REPEATED MEASURES ANOVA

##

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p
--	-----------------------	----------------	----	-------------	---	---

partial <U+03B7>²

## Time	None	20.3	2	10.14	3.96	0.022	0.076
---------	------	------	---	-------	------	-------	-------

##	Greenhouse-Geisser	20.3	1.92	10.55	3.96	0.024	
----	--------------------	------	------	-------	------	-------	--

0.076

##

## Time:condition	None	136.4	2	68.22	26.63	< .001	
-------------------	------	-------	---	-------	-------	--------	--

0.357

##	Greenhouse-Geisser	136.4	1.92	71.01	26.63	< .001	
----	--------------------	-------	------	-------	-------	--------	--

0.357

##


```

## Residual      None                245.9   96      2.56
##              Greenhouse-Geisser      245.9  92.23      2.67
## -----
## Note. Type 3 Sums of Squares
##
##
## Between Subjects Effects
## -----
##              Sum of Squares  df  Mean Square  F    p      partial <U+03B7>²
## -----
## condition          156.1    1    156.06   123  < .001    0.719
## Residual           61.0    48     1.27
## -----
## Note. Type 3 Sums of Squares
##
##
## ASSUMPTIONS
##
## Tests of Sphericity
## -----
##              Mauchly's W    p      Greenhouse-Geisser e  Huynh-Feldt e
## -----
## Time          0.959  0.375          0.961          1.000
## -----
##
##
## POST HOC TESTS
##
## Post Hoc Comparisons - Time
## -----
## Time      Time  Mean Difference  SE    df    t      p-holm
## -----
## pre      -  during      -0.420  0.320  96.0  -1.31  0.274

```

```
##          - post          -0.900  0.320  96.0  -2.81  0.018
## during - post          -0.480  0.320  96.0  -1.50  0.274
## -----
```

Subset your data by time in Long format

```
pre <- subset(science, science$time == "pre")
```

```
during <- subset(science, science$time == "during")
```

```
post <- subset(science, science$time == "post")
```

#difference between the BS conditions (room) at each level of RM measure (young, middle, old).

#Welch's only if Levene's test is significant

```
message("pre")
```

```
## pre
```

```
ttestlS(data=pre, vars = 'value', group = 'condition', eqv = T, effectSize = T)
```

```
##
```

```
## INDEPENDENT SAMPLES T-TEST
```

```
##
```

```
## Independent Samples T-Test
```

```
## -----
```

```
##          statistic  df    p    Cohen's d
```

```
## -----
```

```
## value Student's t      1.45  48.0  0.152      0.411
```

```
## -----
```

```
##
```

```
##
```

```
## ASSUMPTIONS
```

```
##
```

```
## Test of Equality of Variances (Levene's)
```

```
## -----
```

```
##          F      df    p
```

```
## -----
## value 1.67 1 0.202
## -----
## Note. A low p-value suggests a
## violation of the assumption of
## equal variances

message("during")

## during

ttestlS(data=during, vars = 'value', group = 'condition', eqv = T, effectSize = T)

##
## INDEPENDENT SAMPLES T-TEST
##
## Independent Samples T-Test
## -----
##          statistic  df    p    Cohen's d
## -----
## value Student's t   -9.42  48.0 < .001   -2.66
## -----
##
##
## ASSUMPTIONS
##
## Test of Equality of Variances (Levene's)
## -----
##          F    df    p
## -----
## value 2.76 1 0.103
## -----
## Note. A low p-value suggests a
## violation of the assumption of
## equal variances
```

```

message("post")

## post

ttestIS(data=post, vars = 'value', group = 'condition', welchs = TRUE, eqv = T, effectSize = T)

##
## INDEPENDENT SAMPLES T-TEST
##
## Independent Samples T-Test
## -----
##               statistic   df    p      Cohen's d
## -----
## value Student's t      -6.88  48.0 < .001      -1.95
##      Welch's t       -6.88  36.2 < .001      -1.95
## -----
##
##
## ASSUMPTIONS
##
## Test of Equality of Variances (Levene's)
## -----
##           F      df    p
## -----
## value  10.7    1  0.002
## -----
## Note. A low p-value suggests a
## violation of the assumption of
## equal variances

#Another simple effects analysis would be difference between RM condition (time) at each level
of BS measure (control, treatment)

```
#pre
message("pre")
```


```

```
## pre

descriptives(pre, vars = c('value'), splitBy = c('condition'))

##
## DESCRIPTIVES
##
## Descriptives
## -----
##          condition  value
## -----
## N          control    25
##          treatment    25
## Missing    control     0
##          treatment     0
## Mean       control    4.08
##          treatment    3.48
## Median     control     4
##          treatment     4
## Minimum    control     1
##          treatment     1
## Maximum    control     7
##          treatment     6
## -----

#during
message("during")

## during

descriptives(during, vars = c('value'), splitBy = c('condition'))

##
## DESCRIPTIVES
##
## Descriptives
```

```
## -----
##          condition  value
## -----
##  N          control    25
##          treatment    25
##  Missing    control     0
##          treatment     0
##  Mean       control    2.28
##          treatment    6.12
##  Median     control     2
##          treatment     7
##  Minimum    control     1
##          treatment     3
##  Maximum    control     6
##          treatment     7
## -----
```

#post

message("post")

post

descriptives(post, vars = **c**('value'), splitBy = **c**('condition'))

##

DESCRIPTIVES

##

Descriptives

```
## -----
##          condition  value
## -----
##  N          control    25
##          treatment    25
##  Missing    control     0
##          treatment     0
```

```
## Mean      control    3.24
##          treatment    6.12
## Median    control     3
##          treatment     6
## Minimum    control     1
##          treatment     3
## Maximum    control     7
##          treatment     7
## -----
```

Visualization

again, we need to find the standard errors here so we can have beautiful error bars --- always have error bars...always.

```
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
                      conf.interval=.95, .drop=TRUE) {
```

```
  library(plyr)
```

New version of length which can handle NA's: if na.rm==T, don't count them

```
length2 <- function (x, na.rm=FALSE) {
```

```
  if (na.rm) sum(!is.na(x))
```

```
  else    length(x)
```

```
}
```

This does the summary. For each group's data frame, return a vector with

N, mean, and sd

```
datac <- ddply(data, groupvars, .drop=.drop,
               .fun = function(xx, col) {
                 c(N    = length2(xx[[col]], na.rm=na.rm),
                   mean  = mean  (xx[[col]], na.rm=na.rm),
                   sd    = sd    (xx[[col]], na.rm=na.rm)
                 )
               },
               measurevar)
```

```

)

# Rename the "mean" column
datac <- rename(datac, c("mean" = measurevar))

datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error of the mean

# Confidence interval multiplier for standard error
# Calculate t-statistic for confidence interval:
# e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1
ciMult <- qt(conf.interval/2 + .5, datac$N-1)
datac$ci <- datac$se * ciMult

return(datac)
}

sumdat <- summarySE(science, measurevar="value", groupvars=c("condition", "time"))

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:reshape':
##
##   rename, round_any

```

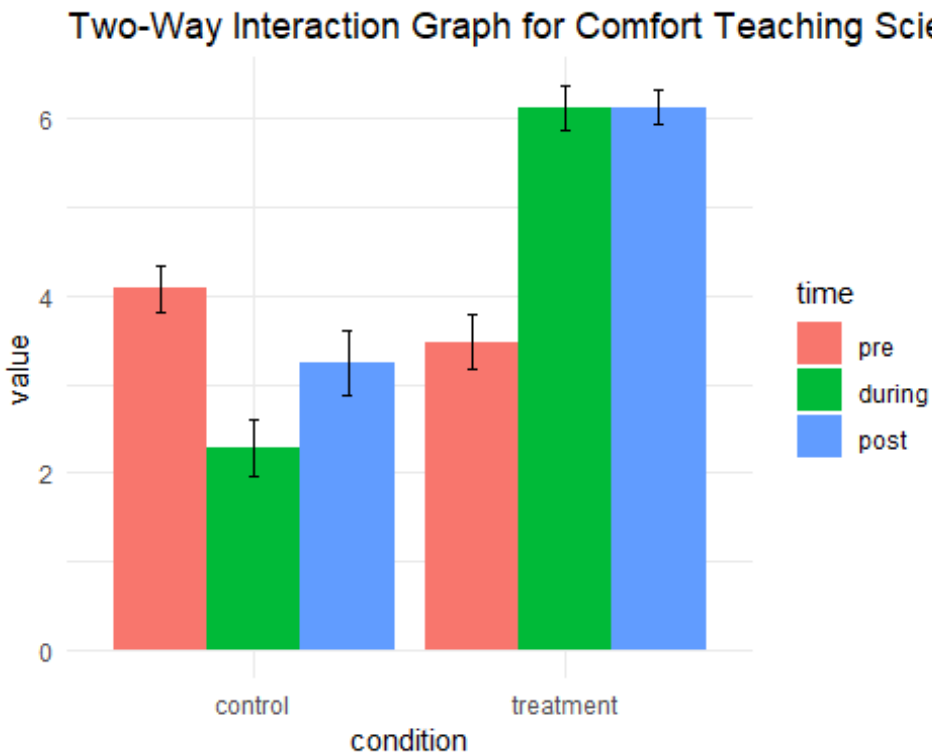
Weird plot doesn't say much

```

#this plot will group them on the x-axis by condition and the cluster by time (three groups, with two bars)
plot <- ggplot(sumdat, aes(x = condition, y = value, fill = time))+
  geom_bar(stat='identity', position= 'dodge') +
  geom_errorbar(aes(ymin=value-se, ymax=value+se), position = position_dodge(.9), width = .1) +
  theme_minimal()

plot + ggtitle('Two-Way Interaction Graph for Comfort Teaching Science by Condition and Time')

```

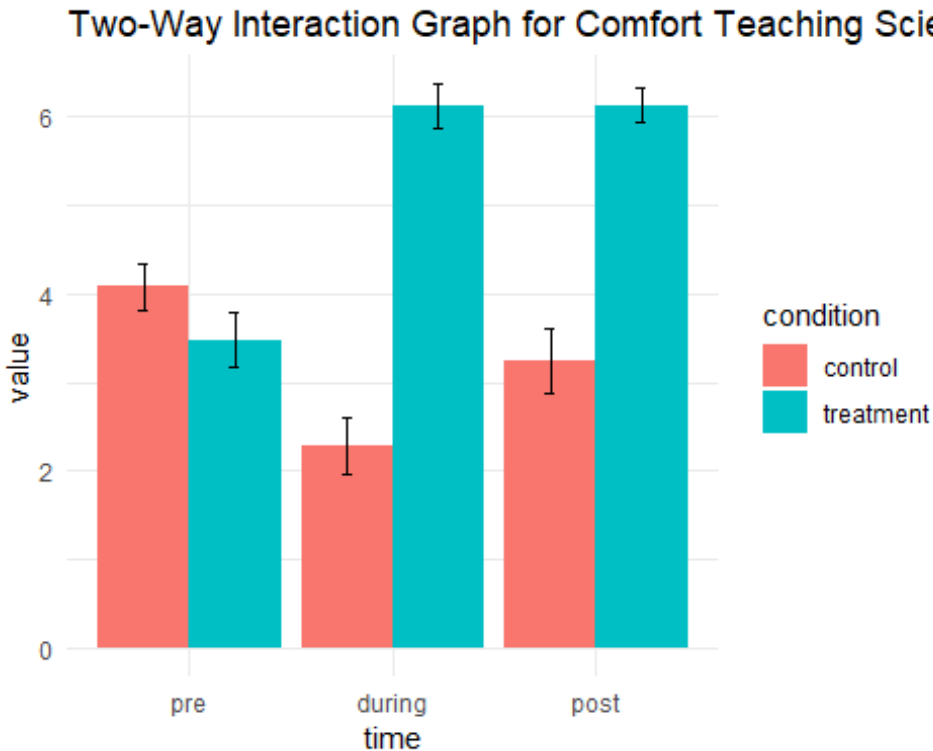



Simple effects plot, shows what is going on

#this plot will group them on the x-axis by time and the cluster by condition (two groups, with three bars)

```
plot <- ggplot(sumdat, aes(x = time, y = value, fill = condition)) +
  geom_bar(stat='identity', position= 'dodge') +
  geom_errorbar(aes(ymin=value-se, ymax=value+se), position = position_dodge(.9), width =
.1) +
  theme_minimal()
```

```
plot + ggtitle('Two-Way Interaction Graph for Comfort Teaching Science by Time and
Condition')
```



Question 1: As the description states, we are looking at two different groups over three time points on their comfort levels for teaching science. As this is the case, (a) what type of test will we run? (b) Make sure the assumptions are met for this test. What did you test? Are any assumptions violated? Report any violations in an orderly manner.

(a) a 2 x 3 Mixed Factorial ANOVA was used to compare comfort levels in teaching science classes (value) between treatment conditions (control, treatment) over time (pre-, during intervention, post)

(b) Assumptions

Repeated Measures Assumption 1. IV is categorical, and DV is continuous. IV is time pre, during, post and are discrete **Pass** DV is value measured on interval level **Pass**

Assumption 2. DV is normally distributed (across each difference score) **Pass** Skew (-3 to +3) and kurtosis (-10 to +10) are within normal limits for each difference score

	pre_during	pre_post	during_post
Skewness	-0.11	0.26	-0.29
Kurtosis	-0.869	-0.395	0.464

Visual inspection of histograms of difference scores across all conditions do not look normal

Assumption 3. Mauchly's test for sphericity **Pass** $W = 0.96, p = .375$

Between Subjects Assumption 1. IV is categorical, and DV is continuous. IV is control, treatment and are discrete **Pass** DV is value measured on interval level **Pass**

Assumption 2. DV is normally distributed (across conditions) - **Pass** Skew (-3 to +3) and kurtosis (-10 to +10) are within normal limits for each difference score

Skewness control 0.33 treatment -0.18

Kurtosis control -0.70 treatment -0.06

Visual inspection of histograms of difference scores across all conditions do not look normal

Assumption 3. Independent observations - each group observation is separate for all groups - **Pass**

Assumption 4. Levene's test of homogeneity of variance - **Pass** $F(1, 48) = 1.98, p = .166$.

Question 2: (a) Which sphericity correction is the most conservative? (b) Report the adjusted p value for this correction. (c) Do we need to use this correction for this test (why/why not)?

- (a) *Greenhouse-Geisser is the most conservative correction for sphericity.* The first step in each sphericity test is to estimate something called epsilon, a descriptive statistic indicating the degree to which sphericity has been violated. If sphericity is met perfectly then epsilon will be exactly 1. If epsilon is below 1 then sphericity is violated. The further epsilon gets away from 1 the worse the violation. Lower bound of epsilon = $1/(k-1)$. So for 3 levels epsilon can go as low as 0.5, for 6 levels it can go as low as 0.2 and so forth. The more levels on the repeated measures factor the worse the potential for violations of sphericity. The Greenhouse-Geisser correction is a conservative correction (it tends to underestimate epsilon when epsilon is close to 1 and therefore tends to over-correct). Huynh-Feldt produced a modified version for use when the true value of epsilon is thought to be near or above 0.75.
- (b) Using Greenhouse-Geisser correction, there is a main effect for time on value, $F(1.92, 92.23) = 3.96, p = .024, \eta_p^2 = .08$. There was a medium effect, with time accounting for 8% of the variance in value.
- (c) We do not need to use this correction for this test as Mauchly's test for sphericity passed, $W = 0.96, p = .375$

Question 3: Is there a main effect for time? In other words, did the teachers comfort level change across the time points? Report all relevant statistics according to APA format.

There was a main effect for time on value, $F(2, 96) = 3.96, p = .022, \eta_p^2 = .08$. There was a medium effect, with time accounting for 8% of the variance in value.

Question 4: Is there a main effect for condition? In other words, are treatment teachers more comfortable teaching science than control teachers? Report all relevant statistics according to APA format.

There was a main effect for condition on value, $F(1, 48) = 12.3, p < .001, \eta_p^2 = .72$. There was a large effect, with condition accounting for 72% of the variance in value.

Question 5: Does comfort teaching science differ across time points depending on what condition the teacher is in (i.e., is there an interaction)? Report all relevant statistics according to APA format.

There was a significant interaction of time and condition on value, $F(2, 96) = 56.63, p < .001, \eta_p^2 = .36$. There was a large effect, with 36% of the variance accounted for in the interaction.

Question 6: Test the assumptions for the simple effect analyses between conditions. What did you test? Are any assumptions violated? Report any violations using APA format. (Hint: You should be testing assumptions for three simple effect analyses)

Simple effects assumptions were run for the simple effect of condition (control, treatment) on each level of time (pre, during, post). The assumptions run were for homogeneity of variance (Levene's test). The assumption of homogeneity of variance was violated for the simple effect of condition on post level of time, $F(1) = 10.7, p = .002$.

Question 7: Test the simple effects between conditions. Report all relevant statistics for each simple effect according to APA format. (Hint: You should be running three simple effect analyses).

The simple effects of condition (control, treatment) was calculated at each level of time (pre, during, post) using independent samples t-tests.

Across the pre level of time, there was no significant difference between control ($M = 4.08$) and treatment ($M = 3.48$) conditions, $t(48) = 1.45, p = .152, d = 0.41$. There was a small effect size ($d = 0.41$) on the difference in value scores between control and treatment conditions.

Across the during level of time, there was a significant difference between control ($M = 2.28$) and treatment ($M = 6.12$) conditions, $t(48) = -9.42, p < .001, d = -2.66$. There was a large effect size ($d = -2.66$) on the difference in value scores between control and treatment conditions.

Across the post level of time, there was a significant difference between control ($M = 3.24$) and treatment ($M = 6.12$) conditions, Welch's $t(36.2) = -6.88, p < .001, d = -1.95$. There was a large effect size ($d = -1.95$) on the difference in value scores between control and treatment conditions.

Question 8: Do you need to follow up these simple effects with post hoc pairwise comparisons? If so report your findings in APA format. If not, explain why not.

No. a t-test is a pairwise comparison, so that means we are done.

Question 9: (a) Bob, a member of the school board and a trained statistician, pulls you aside and asks for your interpretation of the analyses. Write up an interpretation to send to him. (b) The rest of the school board are not trained in statistics but still want to know what is going on with their teachers. Explain the results of your analyses to them.

(a) A 2 x 3 Mixed Factorial ANOVA was used to compare comfort levels in teaching science classes (value) between treatment conditions (control, treatment) over time (pre, during intervention, post).

There was a main effect for time, $F(2, 96) = 3.96, p = .022, \eta_p^2 = .08$. There was a medium effect, with time accounting for 8% of the variance in value.

There was a main effect for condition on value, $F(1, 48) = 123, p < .001, \eta_p^2 = .72$. There was a large effect, with condition accounting for 72% of the variance in value.

There was a significant interaction of time and condition on value, $F(2, 96) = 56.63, p < .001, \eta_p^2 = .36$. There was a large effect, with 36% of the variance accounted for in the interaction.

The simple effects of condition (control, treatment) was calculated at each level of time (pre, during, post) using independent samples t-tests.

Across the pre level of time, there was no significant difference between control ($M = 4.08$) and treatment ($M = 3.48$) conditions, $t(48) = 1.45, p = .152, d = 0.41$. There was a small effect size ($d = 0.41$) on the difference in value scores between control and treatment conditions.

Across the during level of time, there was a significant difference between control ($M = 2.28$) and treatment ($M = 6.12$) conditions, $t(48) = -9.42, p < .001, d = -2.66$. There was a large effect size ($d = -2.66$) on the difference in value scores between control and treatment conditions.

Across the post level of time, there was a significant difference between control ($M = 3.24$) and treatment ($M = 6.12$) conditions, Welch's $t(36.2) = -6.88, p < .001, d = -1.95$. There was a large effect size ($d = -1.95$) on the difference in value scores between control and treatment conditions.

It appears that the professional development course helped teachers increase their comfort levels in teaching above and beyond teachers that received no training. Teachers in both groups started off at about the same level of comfort, but during the training as well as the time period we measured after the training, the teachers that received the training were far more comfortable teaching science classes than teachers that did not receive the training. In fact, teachers that did not receive the training felt less comfortable during the time the other group was in training, although their comfort levels did increase again once the training period had ended. Our recommendations are to send your next cohort of teachers to the training. You may want to test for ability to teach science courses next, as comfort levels to teach science may have no actual bearing on actual tests scores of students, whereas competence may have.

Question 10: How do the assumptions of homogeneity of variance and sphericity relate to each other? How do they differ?

Homogeneity of variance and sphericity are similar in that they are indicators of similarity of variance between conditions. They are different in that homogeneity of variance measures similarity of variance between groups for a between subjects condition, whereas sphericity measures similarity of variance between all combinations of difference scores for a repeated measures condition.