2017.DA.5

Michael Espero

December 12th, 2017

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 PD 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 PD 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.

### R output is included first, followed by Q & A.

## Descriptives

# Clear working directory  
rm(list = ls())  
  
# Load up our packages  
library(RCurl)  
library(psych)  
library(ez)  
library(jmv)  
library(tidyverse)  
library(ggpubr)  
library(tidyverse)  
library(ggjoy)  
library(ggthemes)  
library(reshape)  
  
# I put the data in a private Github folder to make it easy to grab.   
x <- getURL(  
   
"https://raw.githubusercontent.com/michaelespero/r/master/PSY.308b.DA5.csv?token=AFx\_festDNU0NqBH7PPVf9BoICzf8Uynks5aMCgdwA%3D%3D"  
  
)  
  
# Import the data from the link  
science <- read.csv(text = x)  
  
# Let's make the subject variable a factor  
science$Subject <- as.factor(science$Subject)  
  
# Let's reorder the levels of the time variable in the science dataset.   
science$time = factor(science$time,levels=c("pre", "during", "post"),ordered=TRUE)  
  
# Peek at the data  
glimpse(science)

Observations: 150  
Variables: 4  
$ Subject <fctr> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ...  
$ condition <fctr> treatment, treatment, treatment, treatment, treatme...  
$ time <ord> pre, pre, pre, pre, pre, pre, pre, pre, pre, pre, pr...  
$ value <int> 4, 1, 2, 6, 4, 1, 1, 3, 1, 4, 5, 4, 5, 3, 4, 6, 3, 1...

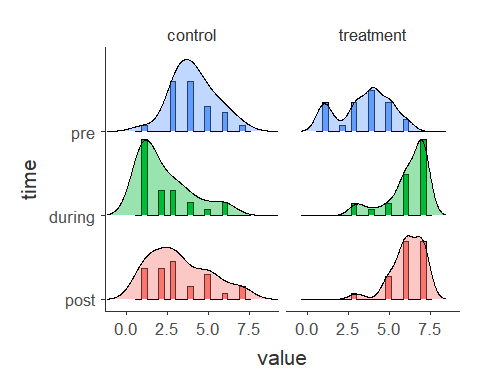
head(science)

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

# Let's take a look at those distributions   
dists\_together <- ggplot(data = science, aes(x = value, y = time, col = condition)) +  
 geom\_joy() + theme\_tufte() +  
 xlab("Comfort Level") + ylab("When") +   
 ggtitle("Comfort Level (Teaching Science) Distributions of Elementary School Teachers")  
  
describe\_comfort <- descriptives(science, vars = 'value', bar = T, n = T, dens = T,   
 kurt = T, skew = T, se = T, hist = T, splitBy = c('time', 'condition'))  
  
#Call up our descriptive statistics including those for skew and kurtosis by time and condition.   
describe\_comfort$descriptives

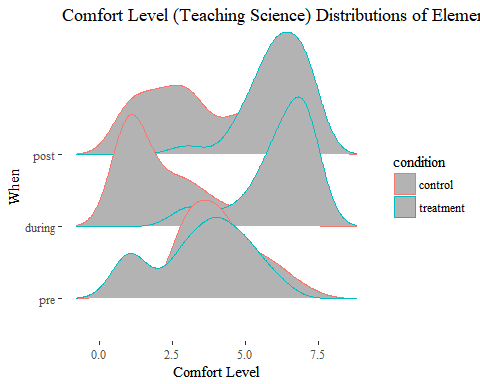
Descriptives   
 ---------------------------------------------------   
 time condition value   
 ---------------------------------------------------   
 N pre control 25   
 treatment 25   
 during control 25   
 treatment 25   
 post control 25   
 treatment 25   
 Missing pre control 0   
 treatment 0   
 during control 0   
 treatment 0   
 post control 0   
 treatment 0   
 Mean pre control 4.08   
 treatment 3.48   
 during control 2.28   
 treatment 6.12   
 post control 3.24   
 treatment 6.12   
 Median pre control 4   
 treatment 4   
 during control 2   
 treatment 7   
 post control 3   
 treatment 6   
 Minimum pre control 1   
 treatment 1   
 during control 1   
 treatment 3   
 post control 1   
 treatment 3   
 Maximum pre control 7   
 treatment 6   
 during control 6   
 treatment 7   
 post control 7   
 treatment 7   
 Standard error pre control 0.264   
 treatment 0.317   
 during control 0.324   
 treatment 0.247   
 post control 0.371   
 treatment 0.194   
 Skewness pre control 0.173   
 treatment -0.297   
 during control 1.03   
 treatment -1.36   
 post control 0.528   
 treatment -1.27   
 Kurtosis pre control 2.86   
 treatment 1.92   
 during control 2.83   
 treatment 3.79   
 post control 2.14   
 treatment 4.81   
 ---------------------------------------------------

#Call up a plot of the distributions.  
describe\_comfort$plots



TRUE

dists\_together



TRUE

## Factorial ANOVA

# Let's make a model. First we'll make an object called 'fma' that keeps the  
# contents of the formula we'll use in the factorial ANOVA.  
  
# fma contains a term for two IVs (condition & time) as well as an  
# interaction term (condition \* time). fma <- value ~ condition + time +  
# condition \* time  
  
# Now we'll feed fma into the lm function and specify that we're using the  
# science dataset. comfort\_mod1 <- lm(fma, science)  
  
# If we use the anova\_stats function from the sjstats package we'll get  
# effect sizes in addition to SS, MS, F, and P. anova\_for\_comfort <-  
# anova\_stats(comfort\_mod1) knitr::kable(anova\_for\_comfort)  
  
# Levene's Test: Test the assumption of equal variances  
dat.wide <- cast(science, Subject + condition ~ time, value = "value")  
dat.wide$averagetime <- (dat.wide$pre + dat.wide$during + dat.wide$post)/3  
car::leveneTest(dat.wide$averagetime, dat.wide$condition, center = mean)

Levene's Test for Homogeneity of Variance (center = mean)  
 Df F value Pr(>F)  
group 1 1.9823 0.1656  
 48

# check\_assumptions from the sjstats package gives us a few assumption  
# checks. knitr::kable(sjstats::check\_assumptions(comfort\_mod1, as.logical =  
# T))  
  
# Let's do the same with the jmv package. jmv\_anova <- jmv::anova(data =  
# science, dep = 'value', factors = c('condition', 'time'), effectSize =  
# 'partEta', plotHAxis = 'time', plotSepLines = 'condition', descStats = T,  
# postHocCorr = 'tukey', postHoc = c('condition', 'time'), plotError = 'se',  
# qq = T, homo = T)  
  
# jmv\_anova$main  
  
# We'll need to test for sphericity using Mauchley's Test as well. Let's let  
# ez do the work.  
ez <- ezANOVA(data = science, dv = value, wid = Subject, within = time, between = condition,   
 detailed = T) #detailed=T is necessary for effect size  
ez$`Mauchly's Test for Sphericity` #This gives us a test for the assumption of sphericity.

Effect W p p<.05  
3 time 0.9590914 0.3747229   
4 condition:time 0.9590914 0.3747229

ez #print the all of the ANOVA table

$ANOVA  
 Effect DFn DFd SSn SSd F p p<.05  
1 (Intercept) 1 48 2671.26 61.01333 2101.515734 2.737651e-41 \*  
2 condition 1 48 156.06 61.01333 122.774476 7.925362e-15 \*  
3 time 2 96 20.28 245.94667 3.957931 2.229926e-02 \*  
4 condition:time 2 96 136.44 245.94667 26.628212 6.312968e-10 \*  
 ges  
1 0.89693172  
2 0.33704808  
3 0.06197286  
4 0.30771313  
  
$`Mauchly's Test for Sphericity`  
 Effect W p p<.05  
3 time 0.9590914 0.3747229   
4 condition:time 0.9590914 0.3747229   
  
$`Sphericity Corrections`  
 Effect GGe p[GG] p[GG]<.05 HFe p[HF]  
3 time 0.9606992 2.385392e-02 \* 0.9999058 2.230287e-02  
4 condition:time 0.9606992 1.266981e-09 \* 0.9999058 6.323519e-10  
 p[HF]<.05  
3 \*  
4 \*

# Effect Size (those wicked smart TAs gave us this sweet function)  
  
eta.2 = function(aov.mdl, ret.labels = FALSE) {  
 eta.2vector = c()  
 labels = c()  
 for (table in summary(aov.mdl)) {  
 # each block of factors  
 SS.vector = table[[1]]$"Sum Sq" #table is a list with 1 entry; [[1]]   
 last = length(SS.vector)  
 labels = c(labels, row.names(table[[1]])[-last]) #all but last (error term)  
 for (SS in SS.vector[-last]) {  
 # all but last entry (error term)  
 new.etaval = SS/(SS + SS.vector[last])  
 eta.2vector = c(eta.2vector, new.etaval)  
 }  
 }  
 if (ret.labels)   
 return(data.frame(eta.2 = eta.2vector, row.names = labels))  
 return(eta.2vector)  
}  
  
comfort\_mod2 <- aov(science$value ~ science$condition + science$time + science$condition \*   
 science$time + Error(factor(science$Subject)/science$time))  
# <- eta.2(comfort\_mod2, ret.labels=T)  
  
# Let's use the pipe operator (%>%) to call up our effect sizes.  
effect\_sizes <- eta.2(comfort\_mod2, ret.labels = T)  
  
effect\_sizes %>% arrange(desc(eta.2)) %>% round(., digits = 2) %>% knitr::kable(.,   
 caption = "Main Effect Sizes")

|  |
| --- |
| eta.2 |
| 0.72 |
| 0.36 |
| 0.08 |

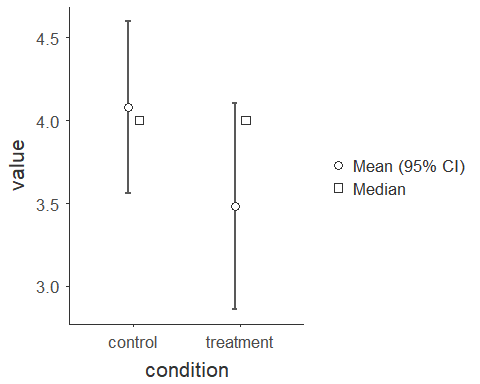
## Simple Effects Analyses

# There's a significant interaction between condition and time. Let's take a  
# closer look by performing a simple effects analyses.  
  
# We'll subset the data by condition and see how each condition moves across  
# the three time points.  
  
# Simple Effects: First we'll subset the data by the three levels of \*time\*.  
data\_pre <- subset(science, time == "pre")  
data\_during <- subset(science, time == "during")  
data\_post <- subset(science, time == "post")  
  
# Is there a simple effect at the \*pre\* measurement?  
  
describe(data\_pre$value)

vars n mean sd median trimmed mad min max range skew kurtosis se  
X1 1 50 3.78 1.47 4 3.83 1.48 1 7 6 -0.23 -0.37 0.21

ttestIS(data = data\_pre, vars = "value", group = "condition", welchs = T, eqv = T,   
 norm = T, plots = T, ci = T, effectSize = T, desc = T)

INDEPENDENT SAMPLES T-TEST  
  
 Independent Samples T-Test   
 --------------------------------------------------------------------------------------   
 statistic df p Cohen's d Lower Upper   
 --------------------------------------------------------------------------------------   
 value Student's t 1.45 48.0 0.152 0.411 -0.229 1.43   
 Welch's t 1.45 46.5 0.152 0.411 -0.230 1.43   
 --------------------------------------------------------------------------------------   
  
  
 ASSUMPTIONS  
  
 Test of Normality (Shapiro-Wilk)   
 --------------------------------   
 W p   
 --------------------------------   
 value 0.929 0.005   
 --------------------------------   
 Note. A low p-value  
 suggests a violation of  
 the assumption of  
 normality  
  
  
 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  
  
  
 Group Descriptives   
 ---------------------------------------------------------------   
 Group N Mean Median SD SE   
 ---------------------------------------------------------------   
 value control 25 4.08 4.00 1.32 0.264   
 treatment 25 3.48 4.00 1.58 0.317   
 ---------------------------------------------------------------



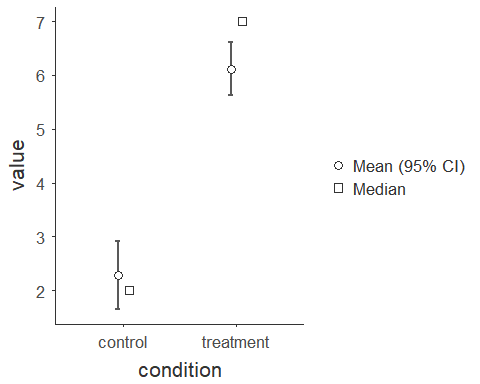
TRUE

# Is there a simple effect of condition at the \*during\* measurement?  
  
describe(data\_during$value)

vars n mean sd median trimmed mad min max range skew kurtosis se  
X1 1 50 4.2 2.41 4.5 4.25 3.71 1 7 6 -0.15 -1.65 0.34

ttestIS(data = data\_during, vars = "value", group = "condition", welchs = T,   
 eqv = T, norm = T, plots = T, ci = T, effectSize = T, desc = T)

INDEPENDENT SAMPLES T-TEST  
  
 Independent Samples T-Test   
 --------------------------------------------------------------------------------------   
 statistic df p Cohen's d Lower Upper   
 --------------------------------------------------------------------------------------   
 value Student's t -9.42 48.0 < .001 -2.66 -4.66 -3.02   
 Welch's t -9.42 44.9 < .001 -2.66 -4.66 -3.02   
 --------------------------------------------------------------------------------------   
  
  
 ASSUMPTIONS  
  
 Test of Normality (Shapiro-Wilk)   
 --------------------------------   
 W p   
 --------------------------------   
 value 0.838 < .001   
 --------------------------------   
 Note. A low p-value  
 suggests a violation of  
 the assumption of  
 normality  
  
  
 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  
  
  
 Group Descriptives   
 ---------------------------------------------------------------   
 Group N Mean Median SD SE   
 ---------------------------------------------------------------   
 value control 25 2.28 2.00 1.62 0.324   
 treatment 25 6.12 7.00 1.24 0.247   
 ---------------------------------------------------------------



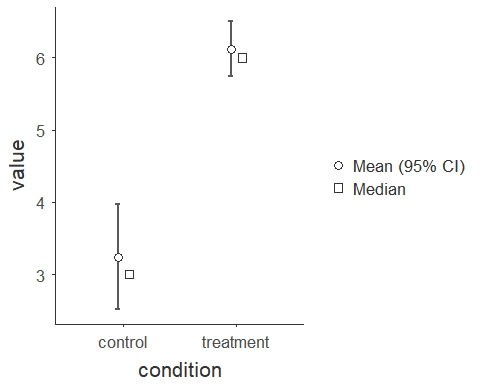
TRUE

# Is there a simple effect of condition at the \*post\* measurement?  
  
describe(data\_post$value)

vars n mean sd median trimmed mad min max range skew kurtosis se  
X1 1 50 4.68 2.06 5 4.85 2.97 1 7 6 -0.46 -1.23 0.29

ttestIS(data = data\_post, vars = "value", group = "condition", welchs = T, eqv = T,   
 norm = T, plots = T, ci = T, effectSize = T, desc = T)

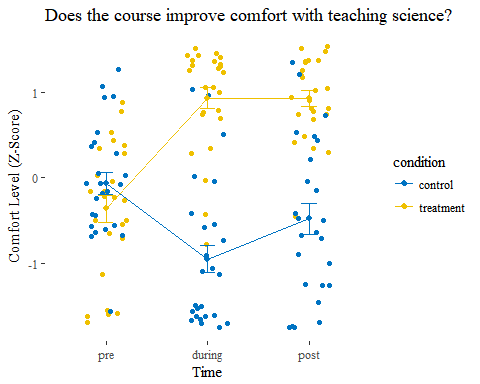
INDEPENDENT SAMPLES T-TEST  
  
 Independent Samples T-Test   
 --------------------------------------------------------------------------------------   
 statistic df p Cohen's d Lower Upper   
 --------------------------------------------------------------------------------------   
 value Student's t -6.88 48.0 < .001 -1.95 -3.72 -2.04   
 Welch's t -6.88 36.2 < .001 -1.95 -3.73 -2.03   
 --------------------------------------------------------------------------------------   
  
  
 ASSUMPTIONS  
  
 Test of Normality (Shapiro-Wilk)   
 --------------------------------   
 W p   
 --------------------------------   
 value 0.874 < .001   
 --------------------------------   
 Note. A low p-value  
 suggests a violation of  
 the assumption of  
 normality  
  
  
 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  
  
  
 Group Descriptives   
 ----------------------------------------------------------------   
 Group N Mean Median SD SE   
 ----------------------------------------------------------------   
 value control 25 3.24 3.00 1.85 0.371   
 treatment 25 6.12 6.00 0.971 0.194   
 ----------------------------------------------------------------



TRUE

## Line Plot with Standard Error Bars

science$value\_centered <- (science$value - mean(science$value)) / sd(science$value)  
  
p <- ggline(science, x = "time", y = "value\_centered",   
 color = "condition", #indicate our variables  
 add = c("mean\_se", "jitter"), #Get error bars, indiv. obs.  
 palette = "jco", #color scheme  
 xlab = "Time", ylab = "Comfort Level (Z-Score)", #Labels for our axes  
 title = "Does the course improve comfort with teaching science?",   
 position = position\_dodge()) + #adjusts obs. horizontally  
 theme\_tufte() #Adds a visual theme   
p



TRUE

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.

1. A Mixed Factorial ANOVA is required to test for a main effect of condition, a main effect of time, as well as for an interaction between condition and time. The omnibus tests provides us with evidence to make a decision regarding *any* significant differences.
2. The Mixed Factorial ANOVA approach depends on several parametric assumptions. First, the assumption of normal distributions was retained using skewness and kurtosis values, each less than 3 and 10, respectively. Next, the assumption of equal variances was retained following results from Levene’s Test, *F*(1, 48) = 1.98, *p* = .1656. The assumption of sphericity (equal variance across difference scores) was assessed with Mauchley’s Test and was retained as well, *W* = .96, *p* = .3747229.

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)?

1. Of the three sphericity corrections commonly used, the Lower Bound Correction is the most conservative [Lower bound of epsilon = 1/(k-1); Here, 1/(3-1) = .5].
2. If we chose to use the Lower Bound Correction we would adjust degrees of freedom by multiplying degrees of freedom by .5.
3. No, it is not necessary to apply a correction because sphericity was not violated, *W* = .96, *p* = .3747229.

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.

* Yes, there is a main effect for time, *F* (2, 96) = 3.96, *p* < .001, 2p = .08. This means that the teachers’ comfort level did change significantly across the time points, however the size of the gross effect of time was *small*, making up 6% of the variance in comfort level scores.

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.

* Yes, there is a main effect for condition, *F* (1, 48) = 122.77, *p* < .001, 2p = .72. This means that the teachers’ comfort level did differ significantly between those assigned to the treatment and control groups and that there was a *whopping* effect size.

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.

* Yes, a significant interaction between time and condition was revealed, *F* (2, 96) = 26.63, *p* < .001, 2p = .36. With regards to the interaction between time and condition, there was a *whopping* effect size.

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)

* Each independent t-test assumed a normally distributed DV and homogeneity (equal) of variances. However, as you can see, an adjustment to the third test will be prompted by an assumption violation:
* With regards to measurements before the course, assumption of the normal, continuus DV was retained provided descriptives within local bounds, skew = -0.23 and kurtosis = -0.37. Levine’s Test was retained, *F* (1) = 1.67, *p* = .202.
* With regards to measurements during the course, assumption of the normal, continuus DV was retained provided descriptives within local bounds, skew = -0.15 and kurtosis = -1.65. Levine’s Test was also retained, *F* (1) = 2.75, *p* = .103.
* With regards to post course measurements, assumption of the normal, continuus DV was retained provided descriptives within local bounds, skew = -0.46 and kurtosis = -1.23. For post course measurements Levine’s Test showed hetereogeneity of variances, therefore, here, the null hypothesis of equal variances should be rejected, *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).

* Simple effects analyses were performed in a series of independent t-tests to show mean differences by condition (between IV: treatment or control) at each measurement phase (IV within: time and its three levels, in our dataset). Pre measurements showed no significant mean difference in teacher comfort level between treatment and controls, *t* (48) = 1.45, *p* = .152, Cohen’s *d* = .41. Measurements taken during the course showed a significant mean difference between treatment and controls, *t* (48) = -9.42, *p* < .001, Cohen’s *d* = -2.66. Post course measurements showed significant mean difference between treatment and controls, Welch’s (1938) *t* (36.2) = -6.88, *p* < .001, Cohen’s *d* = -1.95. Post course measurements appeared to have a nonnormal, negatively skewed distribution, as did the DV for previous tests if we doubt skew and kurtosis values and take into account the appearance of the histograms above and empirical tests (Shapiro-Wilk).

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, the condition variable only contains 2 levels, so there is no need to make follow-up comparisons.

Question 9: 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.

Dear Bob,

* In short, I think the data tells us that teachers in the control group became less comfortable with the prospect of teaching science to their students while those who trained in the course to teach science became more comfortable over time. From the factorial ANOVA we ran, there is a main effect for time, *F* (2, 96) = 3.96, *p* < .001, 2p = .08. There is a main effect for condition, *F* (1, 48) = 122.77, *p* < .001, and a *whopping* 2p = .72. There was a significant interaction between time and condition, *F* (2, 96) = 26.63, *p* < .001, another *whopping* 2p = .36. Simple effects analyses demonstrated significant between group mean differences in during and post measurements (see above), however there was no significant difference at pre course measurements; notice the mostly overlapping “pre” distributions above.
* While several parametric tests may be *robust* to minor deviations from underlying assumptions and these reported p values can still indicate significance of the data after dividing alpha by the number of tests in this family (adjusts alpha to 0.01666667 for each), some statisticians have called for alternative procedures (Yuen, 1974; Keselman, Othman, Wilcox, & Fradette, 2004; Lakens, 2015). For the third t-test in these simple analyses, one might take an approach that includes any or all of the following steps: 1) Transforming the DV applying Box-Cox (1964) and a -1 lambda setting to “pull in” deviations from the mean and reflect a negatively skewed distribution; another option of this sort might be to perform transformation of the percentile-*t* interval (Hall, 1989). 2) Trim the mean by 10% to reduce the influence of “extreme” scores on the distribution 3) Perform a non-parametric Mann-Whitney *U* test and/or bootstrapping procedure to estimate standard errors with repeated random sampling.

Question 10: Most 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.

* The data shows that teachers who trained to teach science became more comfortable over time while those who are left to take the course appear to have become less comfortable with their own comfort in teaching science to their elementary students. The evidence seems pretty tight, however there’s more we can do to deal with any assumptions we may take for granted regarding your teachers. I suggest the next step to include giving the course to those teachers who haven’t had it yet and see if, at a minimum, we can bring back their comfort levels to where they were before the start of *this* administration. Finally, it will be meaningful to see if the teachers who’ve trained in the course maintain their comfort as new science classes begin and whether any of this is related to how the elementary students eventually perform as scientists. Let’s continue learning more, but rest assured that there is at least some evidence that the professional development course works.