1) Andrew James Tillmann

2) Statement from Andrew James Tillmann

"I have completed this work independently. The solutions given are entirely my own work."

3a)

The article looks at the analyst of the data in observational studies and states that results found in these studies should be repeatable while often they are not. The article gives a few possible reasons why.

Firstly, the manages or the ones paying for the study were getting to toss out the final report if the results were not to their liking. The solution with this was to have those in charge be able to question the process and change the way things were being done. However, they were not given power to toss out the final report. The problem with this solution is that no way to make sure the study was publish if it goes against the desires of the those funding the study. No sane company would let a report release if they found out that the study's findings were harmful to them.

Secondly, those that clean the data are often the very ones that analyze it. This may contribute to analyzing the data while trying to clean it thus creating a unknown bias. The solution is to have two different groups of people one group to clean the data and then another to analyze it. This is a rational argument and I see no problem with it.

Thirdly, right now the data for these studies are often not released or if so not given in full disclosure. Since the data and its collection and analyzing methods are not share they cannot be put to questioned. The solution would be to open up the data to the public and let other question the data if they so desire. Opening up the data may seem like a great idea but what if other things could be found in the data such as something harmful to those that funded the study. If that was the case those that funded the study would not want that data to get out. Thus, unless this was somehow enforceable to release the data many may not see it in their best interest.

This is important for those that work in researching since it gives a better way to conduct studies than currently being done. However, these ideas are in the standpoint of what should be done for the best interest of the science but not from the stand point of what is best interest for those funding the studies. Until, those two standpoints are the same I feel those that "Have the gold make the rules" ,thus science will lose out in real life. Nonetheless it is important to theorize how best to do something so if the conditions were ever in place to act upon it the best methods can be applied.

3b)

If the regression model is of a single variable it would mean that 79% of the variability in response variable is explained by explanatory variable. However, if the model is of multiple variables we would need to use the adjusted-R-squared. Moreover, you would need to use a p-value for this test to see if the null hypothesis can be rejected or not. Then, one would need to use a T-test to see if the alternative hypothesis can be accepted or not. Thus **more data would be needed**.

3c)

Regression fallacy is where you draw the wrong conclusion from a regression analyst. The data had a bias in it causing the incorrect conclusion.

Example: Whenever there is police presence in the neighborhood there is a crime reported to the data. When there is no police presence no crime is reported to data. Therefore, the police cause crime.

4a) Use R to perform a regression analysis on the QUASAR dataset(found in MenDenhall content). Done

4b) Paste your model into your submission.

Before you can build a model you need to know a little about the data like what is the response variable is and the explanatory variables are. In this case the response variable is "Rest Frame Equivalent Width" while the explanatory variables are; "Redshift", "Line Flux", "Line Luminosty", "AB", "Abolute Magnitude".

H0=all betas equal 0
HA = at least one beta is not 0
This is a two sided test.

```
Call:
lm(formula = RFEWIDTH ~ poly(LUMINOSITY, ABSMAG, degree = 2),
   data = QUASAR)

Residuals:
   Min      1Q  Median      3Q     Max
-7.7610 -2.6917 -0.0723  2.2503  5.6840

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                          105.272      1.102   95.57  < 2e-16 ***
poly(LUMINOSITY, ABSMAG, degree = 2)1.0  191.352      4.412   43.37  < 2e-16 ***
poly(LUMINOSITY, ABSMAG, degree = 2)2.0   82.270      7.342   11.21 8.16e-10 ***
poly(LUMINOSITY, ABSMAG, degree = 2)0.1  258.154      4.353   59.31  < 2e-16 ***
poly(LUMINOSITY, ABSMAG, degree = 2)1.1  684.143     35.327   19.37 5.72e-14 ***
poly(LUMINOSITY, ABSMAG, degree = 2)0.2   94.459      4.995   18.91 8.78e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.343 on 19 degrees of freedom
Multiple R-squared:  0.9961,    Adjusted R-squared:  0.995
F-statistic: 960.8 on 5 and 19 DF,  p-value: < 2.2e-16
```

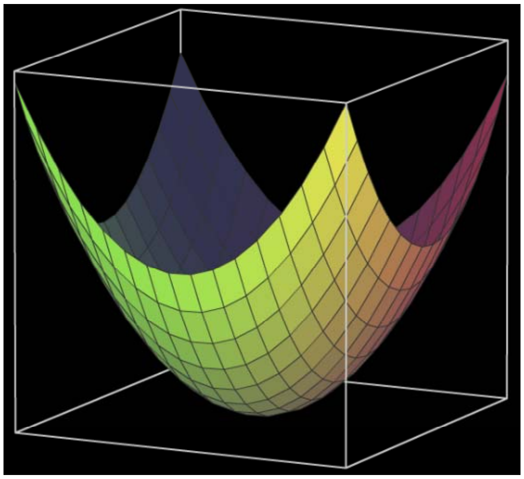
4c) Describe your model

Rest Frame Equivalent Width = 105.272 + 191.352*Luminosity+258.143*Absolute Mangnitude+ 684.143*Luminosity*Absolute Magnitude+82.270*(Luminosity^2)+94.459*(Absolute Mangnitude^2)

My model is a Complete Socend-Order Modle with Two Quantitative Independent Variables. This was the final model choosen because with aejusted R-squared of 99.5 it accounts for everything but .5% of the Rest Frame Equivalent Widith. Forthermore, it has a F-staticstic of 960.8 this is extremely high with a very low p-value almost 0. Moverover, even on the indivual states they had very high T-values with the lowest being 18.91 and the highest being 59.31 all this while having basically 0 p-values.

This model rejects the h0 and accepts HA.

Here is a visual of how the Rest Frame Equivalent Width looks with input form -100 to 100 for both Abslotue Mangnitude and Luminosity.

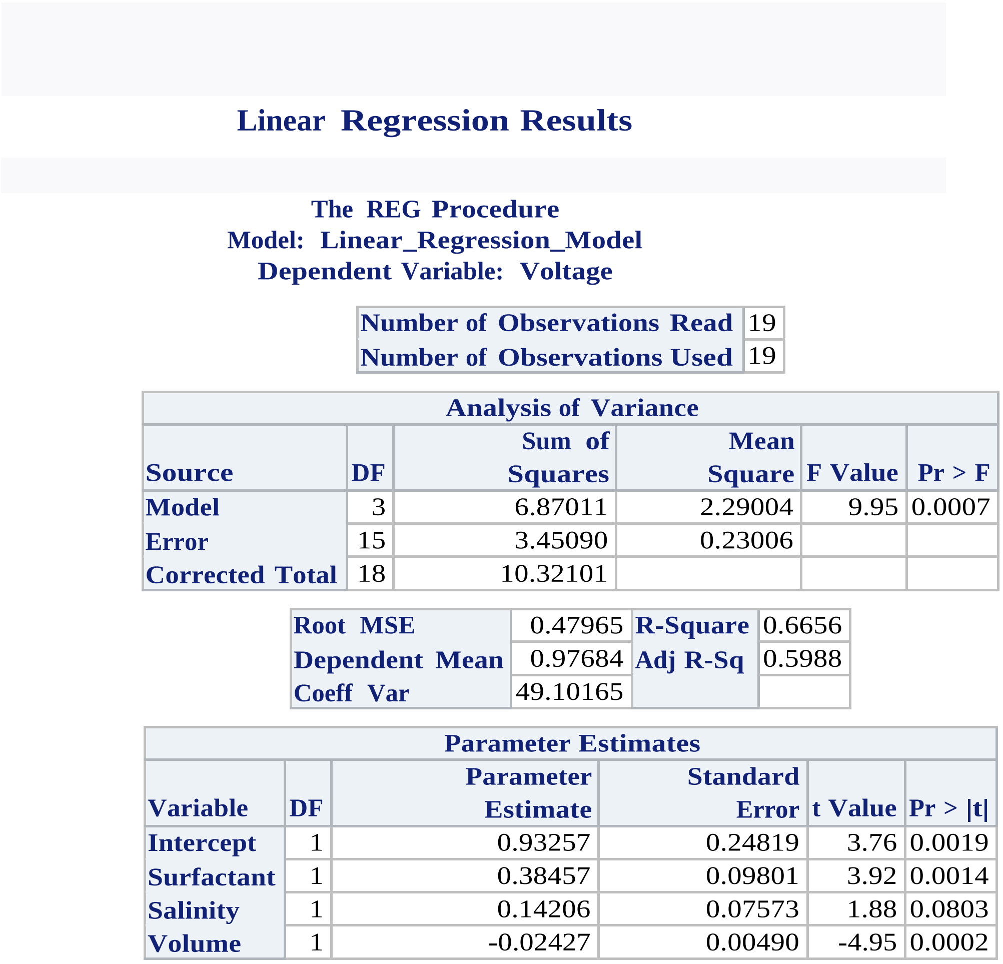


5a) Use SAS to perform a regression analysis on the WATEROIL dataset . Done

Before you can build a model you need to know a little about the data like what is the response variable is and the explanatory variables are. In this case the response variable is "Voltage" while the explanatory variables are; "Volume", "Salinity", "Temperature", "Time Delay", "Surfactant Concentration" , "Span: Triton", and "Solid Particles" .

H0=all betas equal 0
HA = at least one beta is not 0
This is a two sided test.

5b)Paste your model into your submission.



Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: Voltage

Number of Observations Read	19
Number of Observations Used	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6.87011	2.29004	9.95	0.0007
Error	15	3.45090	0.23006		
Corrected Total	18	10.32101			

Root MSE	0.47965	R-Square	0.6656
Dependent Mean	0.97684	Adj R-Sq	0.5988
Coeff Var	49.10165		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.93257	0.24819	3.76	0.0019
Surfactant	1	0.38457	0.09801	3.92	0.0014
Salinity	1	0.14206	0.07573	1.88	0.0803
Volume	1	-0.02427	0.00490	-4.95	0.0002

5c) Describe your model.

This model is a first order model with three variables.
Voltage = .93257+.38457*Surfactant+.14206*Salinity- .02427*Volume

The model accounts for 59.88% of the Voltage. It has a F-Value of 9.95 and a low P-value . Since F-value is above 2.093 and the P-Value is less than .025 the model overall can reject the HO and accept HA. However, for the individual inputs things do not go as well. Surfactant is the only one that can reject HO. Nevertheless, salinity and volume is needed to improve the overall value of the model without them the F-value and P-value for the model drop.