**Table of Contents**

APPENDIX A: PDF Page Numbers of Data - iii

APPENDIX B: List of Major Groupings Per Year - iv

APPENDIX C: STEM Versus Non STEM Major Classifications - v

APPENDIX D: Mean Mathematical SAT Scores Tables - vi

APPENDIX E: Note on Intervention Analysis R Code - vii

APPENDIX F: Time Series Plots and Interpretations - viii

APPENDIX G: ACF/PACF Plots and Interpretations - x

APPENDIX H: Determination of Time Series Structures - xii

APPENDIX I: Time Series Residual Analysis Plots - xiv

APPENDIX J: Paired T Test Normality Verification and Interpretation - xvi

**APPENDIX A: PDF Page Numbers of Data**

Source of All Data Used in Study from the College Board



**APPENDIX B: List of Major Groupings Per Year**



**APPENDIX C: STEM Versus Non STEM Major Classifications**

Based on the lists below, a column was added to each record of data in this study to indicate if the major grouping within a particular record was STEM-related or non STEM-related, with a “1” signifying a STEM-related degree, and a “0” signifying a non STEM-related degree.



**APPENDIX D: Mean Mathematical SAT Scores Tables**

**APPENDIX E: Note on Intervention Analysis R Code**

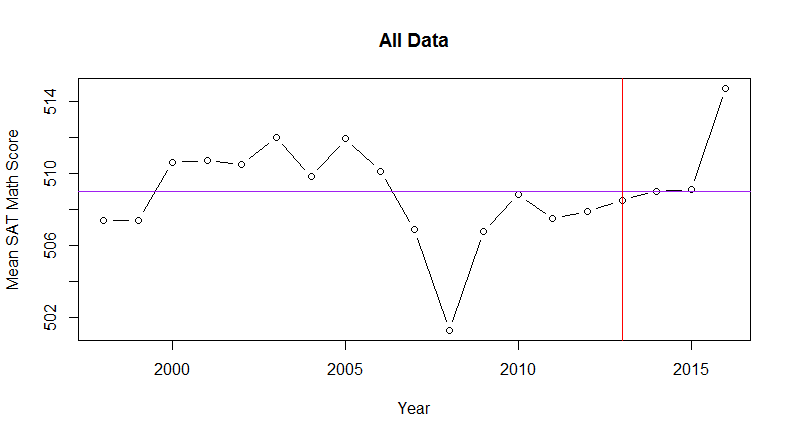
Code utilized via R Studio is available through an attachment called “Jacobs Capstone R Script.R”. The code includes commentary for those who might not know how to use R Studio who would like to replicate this experiment. If you are unfamiliar with R Studio, you will need to install both R Studio and R in order to run this code. Below references websites where you can install both programs:

R Studio: https://www.rstudio.com/products/rstudio/download/

R: https://cran.cnr.berkeley.edu/

The three .dat files used for this study are also available through attachments, titled “NYALL.dat”, “NYSTEM.dat”, and “NYNONSTEM.dat” respectively.

**APPENDIX F:** **Time Series Plots and Interpretations**



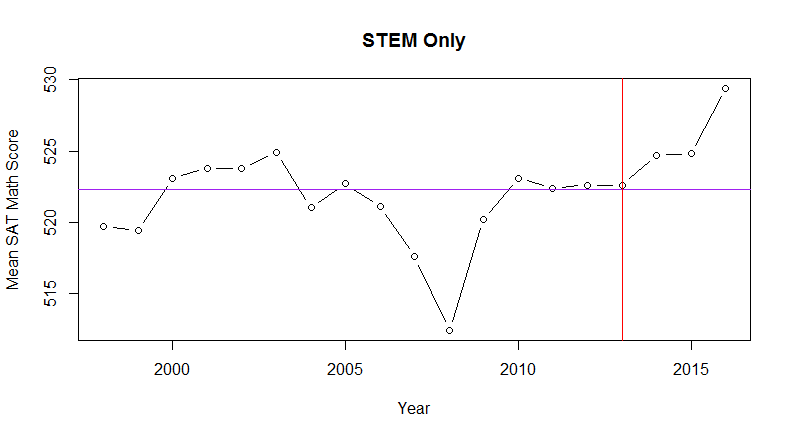
Trend: There appears to be a positive trend between 1998 and 2003, as well as from 2011 to 2016. There appears to be a negative trend from 2005 to 2008.

Seasonality: There is no seasonality in this data, as the average score is reflected on a yearly basis.

Long-Run Cycle: There does not seem to be a consistent pattern unrelated to seasonality.

Constant Variance: The variance looks relatively constant with the exception of 2008 and 2016 – values are changing at a consistent rate otherwise.

Abrupt Changes: There is a stark decrease in average mathematical SAT score in 2008



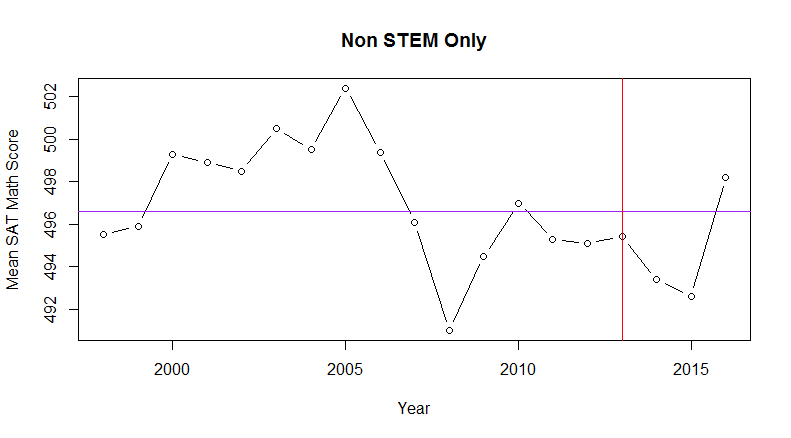
Trend: With the exception of 2008, there appears to be a relatively positive trend in mathematical SAT performance.

Seasonality: There is no seasonality in this data, as the average score is reflected on a yearly basis.

Long-Run Cycle: There does not seem to be a consistent pattern unrelated to seasonality.

Constant Variance: The variance looks relatively constant with the exception of 2008 – values are changing at a consistent rate otherwise.

Abrupt Changes: There is a stark decrease in average mathematical SAT score in 2008



Trend: There appears to be a positive trend between 1998 and 2005. There appears to be a negative trend from 2005 to 2008, as well as from 2010 to 2015.

Seasonality: There is no seasonality in this data, as the average score is reflected on a yearly basis.

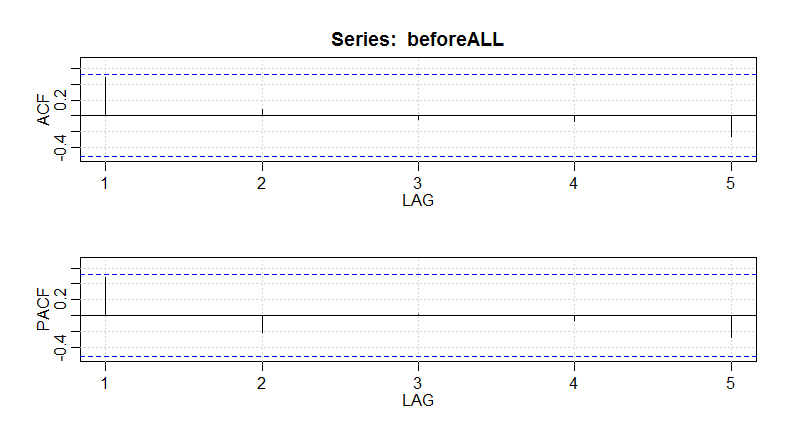
Long-Run Cycle: There does not seem to be a consistent pattern unrelated to seasonality.

Constant Variance: The variance looks relatively constant with the exception of 2008 – values are changing at a consistent rate otherwise.

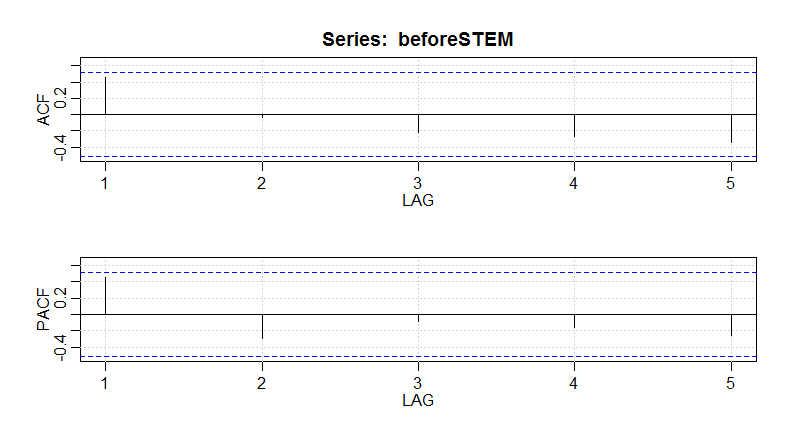
Abrupt Changes: It appears that 2005, the maximum average mathematical score for a calendar year, was followed by a sharp decrease in average mathematical SAT score, shortly hitting the minimum in 2008.

**APPENDIX G: ACF/PACF Plots and Interpretations**

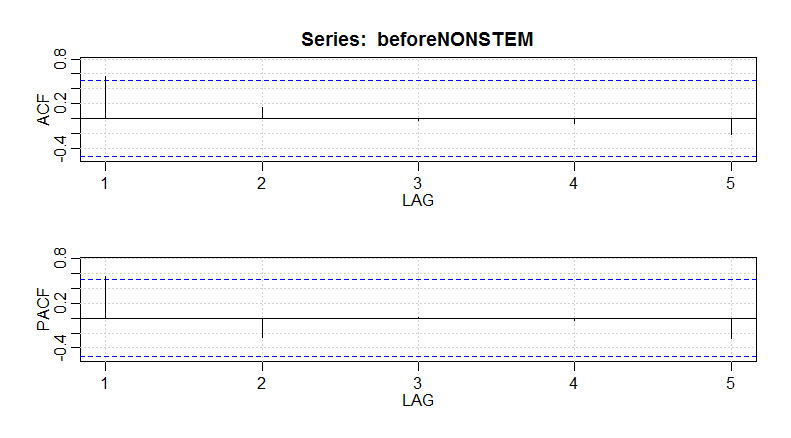
An ACF/PACF plot is used to detect patterns in data relationships between different points of time within a data set (called “lags”). These plots are used to determine which time series structures may be the best fit for a data set at hand.



Interpretation: The first lag of both the ACF and PACF have values that are very close to crossing the blue dashed line. A line exceeding the blue dashed line for the ACF is indicative of a potential Moving Average (MA) component, while a line exceeding the blue dashed line for the PACF is indicative of a potential Autoregressive (AR) component. There are no seasonal components in this plot, which could be observed if there were lags at 12, 24, and 36. Because the remainder of lines for the ACF and PACF taper toward zero in some sort of fashion, differencing is not needed. If differencing were needed, almost all of the lines would cross the blue dashed line. It seems best to test for an ARMA model.



Interpretation: The first lag of both the ACF and PACF have values that are very close to crossing the blue dashed line. A line exceeding the blue dashed line for the ACF is indicative of a potential Moving Average (MA) component, while a line exceeding the blue dashed line for the PACF is indicative of a potential Autoregressive (AR) component. There are no seasonal components in this plot, which could be observed if there were lags at 12, 24, and 36. Because the remainder of lines for the ACF and PACF taper toward zero in some sort of fashion, differencing is not needed. If differencing were needed, almost all of the lines would cross the blue dashed line. It seems best to test for an ARMA model.



Interpretation: The first lag of both the ACF and PACF cross the blue dashed line on the plot. A line exceeding the blue dashed line for the ACF is indicative of a potential Moving Average (MA) component, while a line exceeding the blue dashed line for the PACF is indicative of a potential Autoregressive (AR) component. There are no seasonal components in this plot, which could be observed if there were lags at 12, 24, and 36. Because the remainder of lines for the ACF and PACF taper toward zero in some sort of fashion, differencing is not needed. If differencing were needed, almost all of the lines would cross the blue dashed line. It seems best to test for an ARMA model.

**APPENDIX H: Determination of Time Series Structures**

Below are summaries of the results of trying to fit different Time Series structures to each data set, as well as interpretations and conclusions on the best model to select. When selecting a model, we are looking for the smallest values of AIC, AICc, BIC, and Variance as possible. AIC, AICc, and BIC are all different standardized calculations that help to determine the best fit for a model. The variance, simply put, is a quantitative measurement on how different all of the data points are within a pattern or structure; since we are looking for a best fit, we want the variance to be small and our data points to be similar within a pattern or structure.



Interpretation: Two models gave promise to proceed with. To determine between the two, we will use the smallest AICc to make the decision, as the AICc criterion can be helpful to emphasize on for a sample size smaller than 30 (which we have). Therefore, we will select model ARMA (0,0,1) because the AICc and BIC are minimized in this model.



Interpretation: Two models gave promise to proceed with, and one model produced an error and had a poor residual output (highlighted in red). Between the two potential models, we will use the smallest AICc to make the decision, as the AICc criterion can be helpful to emphasize on for a sample size smaller than 30 (which we have). Therefore, we will select model ARMA (0,0,1) because the AICc and BIC are minimized in this model.



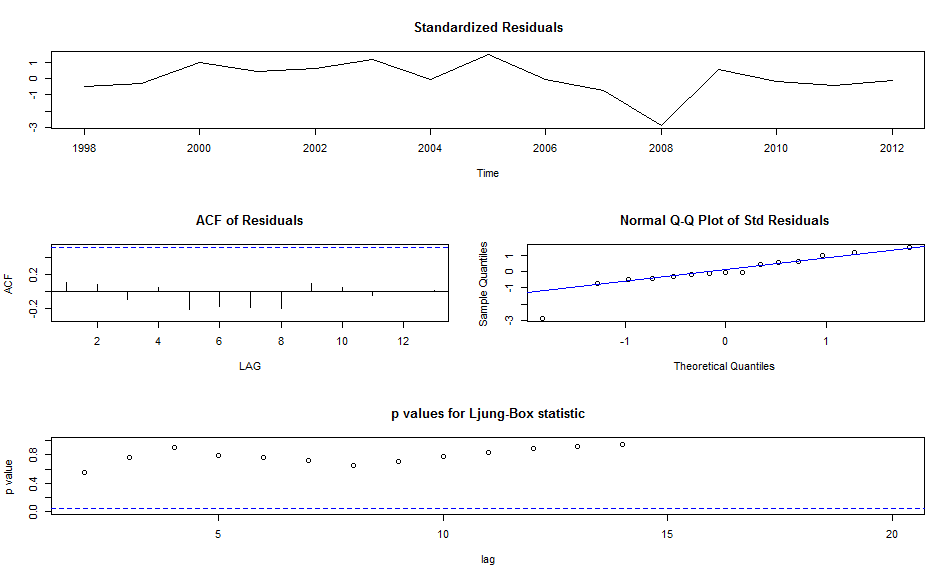
Interpretation: We will select model ARMA (1,0,2) because the AIC, BIC, and Variance are all minimized within this model. While the AICc is smallest within model ARMA (1,0,0), the variance is the second largest in this model.

**APPENDIX I: Time Series Residual Analysis Plots and Interpretations**

When interpreting these plots, the following qualities indicate a good fit:

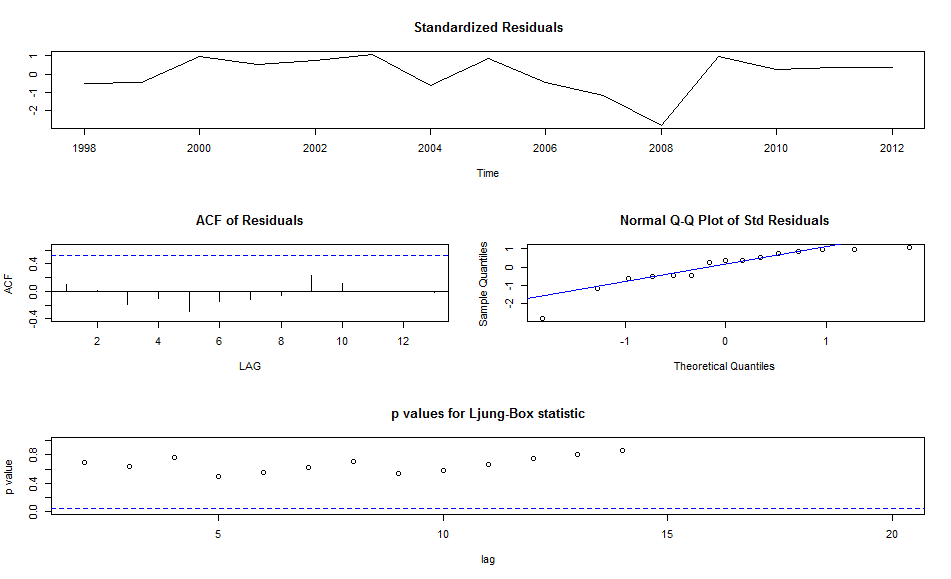
* Standardized Residuals: Plot bounces up and down, but is ultimately fairly consistent.
* ACF of Residuals: All lines need to be between blue dashed lines
* Normal Q-Q Plot of Std Residuals: Points need to adhere as closely as possible to the line.
* P Values for Ljung-Box Statistic: Points need to be above dashed line

All Data:



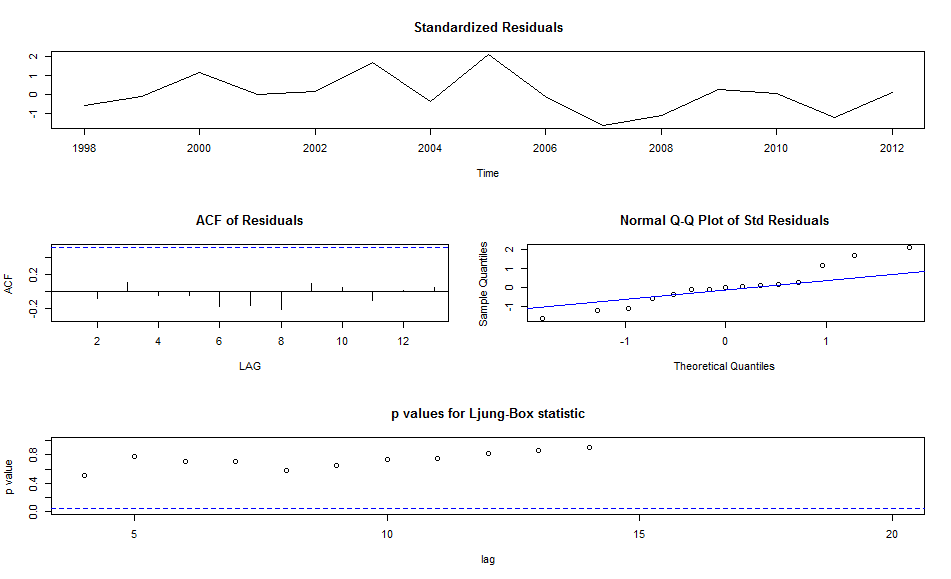
Interpretation: With the exception of 2008, which is an outlier accounted for when starting this analysis, the standardized residuals are fairly consistent. The normal Q-Q plot displays a fairly good fit, with the exception of one point (which is 2008). The ACF of residuals appears to be a good fit, as all the lines are small and between blue dashed lines. The p values for Ljung-Box statistic shows all points as not significant, which is what we would desire from a model.

STEM Data:



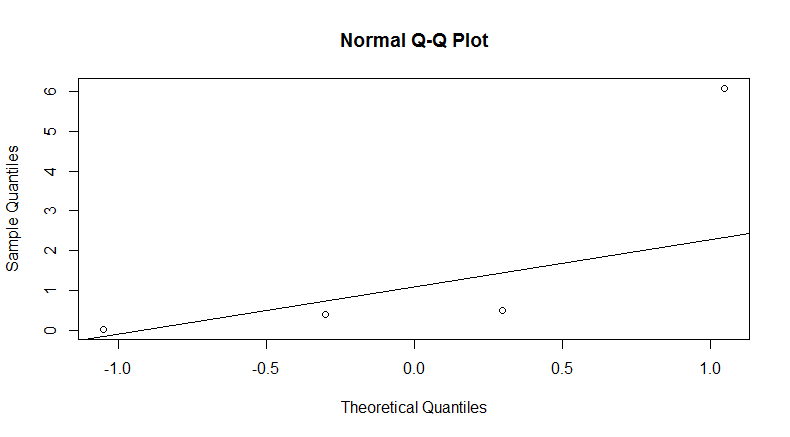
Interpretation: With the exception of 2007 and 2008, which was detected in prior analysis, the standardized residuals are fairly consistent. The normal Q-Q plot displays a decent fit, as the points appear to swerve about the line. There are two points on the normal Q-Q that don’t adhere closely to the line (2007 and 2008). The ACF of residuals appears to be a good fit, as all the lines are small and between blue dashed lines. The p values for Ljung-Box statistic shows all points as not significant, which is what we would desire from a model.

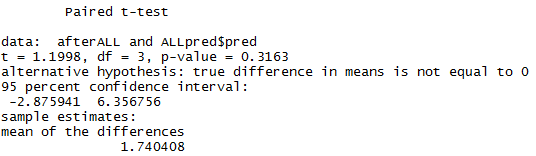
Non STEM Data:



Interpretation: With the exception of 2005 and 2008, which was detected in prior analysis, the standardized residuals are fairly consistent. The normal Q-Q plot displays an adequate fit, due to the swerving but not aligning with the line. There are six points on the normal Q-Q that don’t adhere closely to the line (which is not preferred), although this will need to be acceptable since this model was the best fit in all other aspects. The ACF of residuals appears to be a good fit, as all the lines are small and between blue dashed lines. The p values for Ljung-Box statistic shows all points as not significant, which is what we would desire from a model.

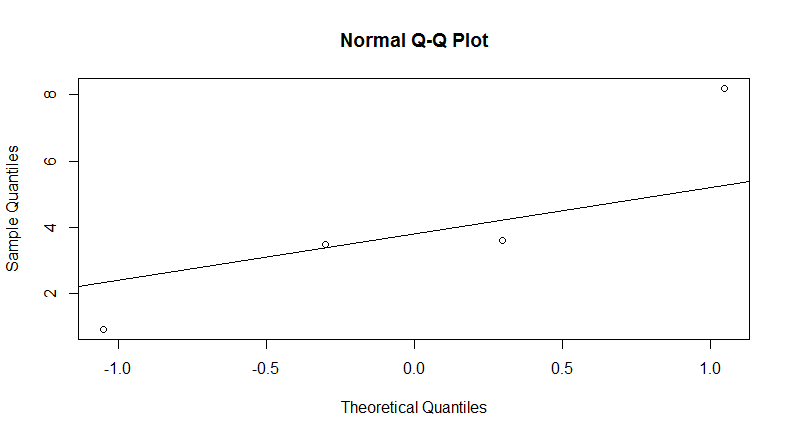
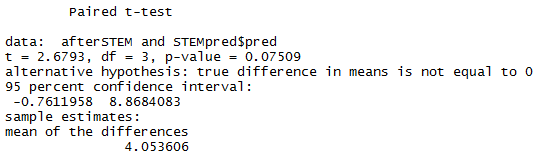
**APPENDIX J: Paired T Test Normality Verification and Interpretation**

All Data



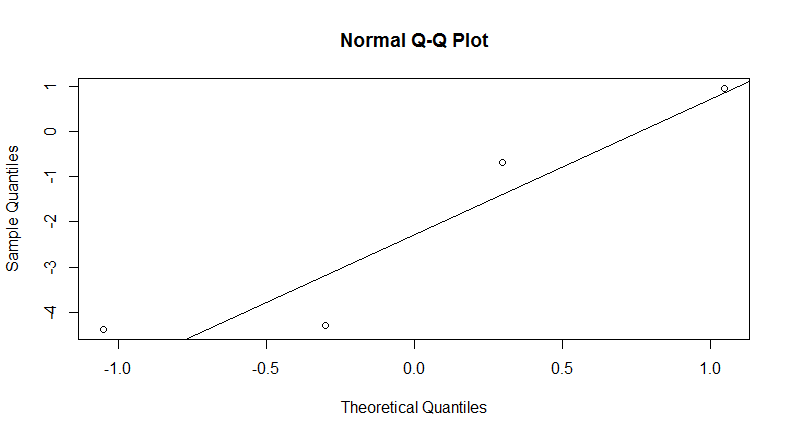
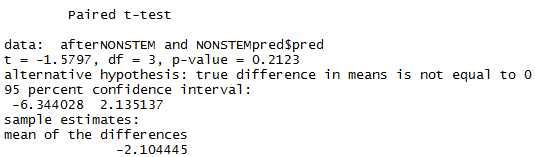
Interpretation: Plotted points do not follow the probability plot line closely. The sample size is too small for a paired t test. Regardless, with a p value of 0.3163, the result wouldn’t be statistically significant even if used (with a confidence level of .05).

STEM-Related Degrees



Interpretation: Plotted points do not follow the probability plot line closely. The sample size is too small for a paired t test. Regardless, with a p value of 0.07509, the result wouldn’t be statistically significant even if used (with a confidence level of .05).

Non STEM-Related Degrees



Interpretation: Plotted points do not follow the probability plot line closely. The sample size is too small for a paired t test. Regardless, with a p value of 0.2123, the result wouldn’t be statistically significant even if used (with a confidence level of .05).