# MAT 303 Project One Summary Report

[Full Name]

[SNHU Email]

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## 1. Introduction

*Discuss the statement of the problem in terms of the statistical analyses that are being performed. Be sure to address the following:*

* *What is the data set that you are exploring?*
* *How will your results be used?*
* *What type of analyses will you be running in this project?*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 2. Data Preparation

*There are some important variables that are used in this project. Identify and explain these variables.*

* *What are the important variables in this data set?*
* *How many rows and columns are present in this data set?*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 3. Model #1 - First Order Regression Model with Quantitative and Qualitative Variables

### Correlation Analysis

* *Create the following scatterplots and include a copy of each in this section:*
  + *Price (price) vs. the living area (sqft\_living)*
  + *Price (price) vs. the age of the home (age)*
* *Describe what trends, if any, exist for each scatterplot.*
* *Report the correlation coefficients between the following variables:*
  + *Price (price) vs. the living area (sqft\_living)*
  + *Price (price) vs. the age of the home (age)*
* *Describe the strength and direction of each correlation coefficient.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Reporting Results

* *Write the general form and the prediction equation of the multiple regression model using price as the response variable and living area, upper level area, age of the home, number of bathrooms, and view as predictor variables. Use (where i* equals *1, 2, ... ) to represent the slope parameters for all predictor variables.*
* *Create the multiple regression model for price as a response variable with living area, upper level area, age of the home, number of bathrooms, and view as predictor variables. Write the prediction model equation using outputs obtained from your R script.   
  Note: You will use the variables living area, upper level area, age of the home, and number of bathrooms as* ***quantitative*** *variables and view as a* ***qualitative*** *variable in this model. Use the equation editor to write the prediction model equation.*
* *What are the values of R2 (R-squared) and Adjusted R-Squared (adjusted R-squared) for the model? Provide your interpretation of these statistics.*
* *Interpret the beta estimates for the living area and lake view.*
* *Obtain the residuals and fitted values to create the following plots. Include these plots and comment on the validity of assumptions. Do not include the residual or fitted values tables.*
  + *Residuals against Fitted Values*
  + *Normal Q-Q plot*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Evaluating Significance of Model

* *Is the model significant at a 5% level of significance? Carry out the overall F-test by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of the test.*
* *Which terms are significant at a 5% level of significance? Carry out individual beta tests by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of each test.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Making Predictions Using Model

* *What is the predicted price for a home that has 2150 sqft living area, 1050 sqft upper level living area, is 15 years old, has 3 bathrooms, and backs out to road? Obtain 90% prediction and confidence intervals for the price of this home. Interpret each interval.*
* *What is the predicted price for a home that has 4250 sqft living area, 2100 sqft upper level living area, is 5 years old, has 5 bathrooms, and backs out to a lake? Obtain 90% prediction and confidence intervals for the price of this home. Interpret each interval.*
* *Why is the prediction interval wider than the confidence interval?*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 4. Model #2 - Complete Second Order Regression Model with Quantitative Variables

### Correlation Analysis

* *Create scatterplots of:*
  + *Price (price) vs. average school rating in the area (school\_rating)*
  + *Price (price) vs. the crime rate per 100,000 people (crime)*
* *Comment on each scatterplot. Is a second order model appropriate using these variables?*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Reporting Results

* *Write the general form and the prediction equation of a complete second order model for price using average school rating in the area and crime rate per 100,000 people as predictors. Use (where i* equals *1, 2, ... ) to represent the slope parameters for all predictor variables.*
* *Create the complete second order model for price using average school rating in the area and crime rate per 100,000 people as predictors. Write the prediction model equation using outputs obtained from your R script.  
  Note: Use average school rating in the area and crime rate as* ***quantitative*** *variables in this model. Use the equation editor to write the prediction model equation.*
* *What are the values of (R-squared) and (adjusted R-squared) for the model? Provide your interpretation of these statistics.*
* *Obtain the residuals and fitted values to create the following plots. Include these plots and comment on the validity of assumptions. Do not include the residual or fitted values tables.*
  + *Residuals against Fitted Values*
  + *Normal Q-Q plot*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Evaluating Significance of Model

* *Is the model significant at a 5% level of significance? Carry out the overall F-test by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of the test.*
* *Which terms are significant at a 5% level of significance? Carry out individual beta tests by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of each test.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Making Predictions Using Model

* *What is the predicted price for a home in an area with average school rating of 9.80 and a crime rate of 81.02 per 100,000 individuals? Obtain 90% prediction and confidence intervals for the price of this home. Interpret each interval.*
* *What is the predicted price for a home in an area with average school rating of 4.28 and a crime rate of 215.50 per 100,000 individuals? Obtain 90% prediction and confidence intervals for the price of this home. Interpret each interval.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 5. Nested Models F-Test

### Reporting Results

* *Write the general form and the prediction equation of a first order model for price using average school rating in the area and crime rate per 100,000 people as predictors. Include the interaction term between average school rating and crime rate. Use (where i* equals *1, 2, ... ) to represent the slope parameters for all predictor variables.*
* *Create the first order regression model for price using average school rating in the area and crime rate per 100,000 people as predictors. Include the interaction term between average school rating and crime rate. Write the prediction model equation using outputs obtained from your R script.  
  Note: Use average school rating and crime rate as* ***quantitative*** *variables in this model. Use the equation editor to write the prediction model equation.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Evaluating Significance of Model

* *Is the model significant at a 5% level of significance? Carry out the overall F-test by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of the test.*
* *Which terms are significant at a 5% level of significance? Carry out individual beta tests by identifying the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of each test.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

### Model Comparison

*You will now compare this model with the second order model for price using average school rating in the area and crime rate per 100,000 people as predictors to test whether the quadratic (squared) terms contribute in predicting the prices of homes. The complete second order model is Model #2, which you created in this project.*

* *In general, what is a reduced and a complete model when comparing two models?*
* *Write general form and prediction equation of the model that is the reduced model in this comparison in terms of and ?.*
* *Write general form* *and prediction equation of the model that is the complete model in this comparison in terms of and ?*
* *Run the nested model F-test at a 5% level of significance to evaluate if the quadratic (squared) terms are needed. Identify the null hypothesis, the alternative hypothesis, the P-value, and the conclusion of the test.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 6. Conclusion

*Describe the results of the statistical analyses clearly, using proper descriptions of statistical terms and concepts. Fully describe what these results mean for your scenario.*

* *Which model would you choose to predict house prices? Briefly summarize your findings in plain language.*
* *What is the practical importance of the analyses that were performed?*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include R code in your report.*

## 7. Citations

*You were* ***not*** *required to use external resources for this report. If you did not use any resources, you should remove this entire section. However, if you did use any resources to help you with your interpretation, you* ***must*** *cite them. Use proper APA format for citations.*

Insert references here in the following format:

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.