Descriptive Statistics and Regression Analysis with R

Overview and Rationale

It is important for you to be able to describe data numerically and graphically and using multiple regression to predict influential variables. In this assignment you will use R in a hands-on experience on data analytics as a review.

Course Outcomes

This assignment is directly linked to the following key learning outcomes from the course syllabus:

- Describe data numerically and graphically and predict influential variables for real world business problems

Assignment Summary

There are two parts of this assignment:

Part A: Use R functions to describe data numerically and graphically.

Part B: Use R functions to build a multiple regression model for real world data.

You will then report your work and findings in a 1000 word paper.

Use the following supporting materials for R syntax, data sets and tools:

- [Using R for Data Analysis and Graphics](#) by J H Maindonald.

- [Quick R](#)

Follow the instructions below for each part of the assignment:

Part A

Use the "*Trees*" data or another data set that is part of R. Then, use the functions in sections 2.5, 3.5 and 3.6 of "Using R for Data Analysis and Graphics" to describe data numerically and construct the graphs to describe data graphically. Follow the steps below.

1. Invoke R and use the "*Tree*" dataset

2. Find the 5 summary numbers in the data

3. Graph a straight line regression

4. Create Histograms and density plots

5. Create Boxplots

6. Normal probability plots

Include your code and results in your report.

Part B

Use the "*Rubber*" and "*oddbooks*" data sets, or choose two use other appropriate data sets, in R. Then use the functions in section 5.4 of "Using R for Data Analysis and Graphics" to build multiple regression models.

In addition, you need to install the `DAAG` package before you can complete this part of the assignment. Follow the steps below:

1. Load the MASS and ggplot2 libraries and use the "*Rubber*" data set

2. Load the DAAG library and use the "*oddblocks*" data set

3. Build multiple regression models using `summary(), log(), lm()` and `ggcorrplot()`

Include your code and results in your report. Be sure to show the model with insights, correlation matrix and explanations.

Report

Your assignment/project should have a good cover/title page, introduction of what the goals of the project and the methods you use. It also should follow APA format with at least 1000 words (excluding title page and references page) and references page. In the body of your project you should incorporate the R codes and R outputs with interpretation of your results. You need to make sense of your results to make good points to show your understanding of the course material and its application to the dataset.

Graphs, figures, charts, tables are very useful to increase visual effects to impress your readers. You also should do your best to give insight and understanding to the project with

a good conclusion. Please use subtitles to make your assignment more reader friendly as well.

Format & Guidelines

The report should follow the following format:

(i)   Title page

(ii)   Introduction

(iii)   Analysis

(iv)   Conclusion/Interpretations

(v)   References

And be 1000 words in length and presented in the APA format

Assignment Rubric

Category	Meets Standards	Approaching Standards	Below Standards
Introduction	Introduction provides a brief and intelligible overview of the goals and methods of the assignment	Introduction provides an overview of the goals and methods of the assignment, but is ambiguous or not concise	Does not introduce project goals, project questions or methods.
Analysis	Provides all R code and the outputs. Includes interpretation of the output, graphs, figures, charts, and tables and the significance of the results in the analysis.	Provides R codes and outputs, but the R code does not match the outputs or is missing some code or outputs. Includes limited interpretations, charts, and tables and the significance of the results in the analysis.	Does not provide R code or its outputs or minimal R code is provided. Includes few interpretations, charts, or tables. Does not identify the significance of the results in the analysis
Data Visualizations	Data visualizations are appropriate for the level and type of analysis. Graphs, figures and tables communicate insights and significance to the reader.	Data visualization are useful for the level and type of analysis, but graphs, figures and tables do not clearly communicate significance of the results to the reader.	Data visualization are used minimally or not at all. If graphs, figures and tables are used, it is unclear what they are intended to communicate or why.
Interpretation & Conclusions	The conclusion summarizes and makes sense of the results, making good points that reflect clear understanding of the assignment material.	The conclusion summarizes and makes sense of the results, making good points that reflect a basic understanding of the assignment material.	The conclusion does not summarize or attempt to make sense of the results. Conclusions do not reflect an understanding or reflect a misunderstanding of the material
Report: Writing Mechanics, Title Page, & References	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains approximately of 1000 words	There are very few errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains approximately 1000 words	There are more than five errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are incomplete; or the report contains far less than 1000 words

Northeastern University
College *of* Professional Studies