Northeastern University
College *of* Professional Studies

Module 5 Project

Regression and Correlation Analysis

You must work on teams of **2** students (must be different from previous weeks)

Notice the deadline for this assignment: Monday at 11:59 pm

Overview and Rationale

This assignment is designed to provide you with hands-on experiences in performing regressions and correlation analysis. The data set is provided in an Excel workbook and contains a wide range to data types that you will need to work with.

Remember that for your references you should use **books**, **scientific journals,** or strictly **academic sources**.

Course Outcomes

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

CO1: Explore the use of statistical software in data analysis through hands-on applications,

CO4: Perform estimations of population parameters using confidence intervals based on one sample and perform estimations of the difference between two population parameters of the same kind based on two samples.

CO5: Conduct regression and chi-squared test of independency to study associations between numerical and categorical variables respectively; and justify the legitimacy of the regression model.

CO6: Perform various hypothesis tests, including those for a population parameter (single sample), and the difference between two population parameters of the same kind (two samples), and perform analysis of variance (ANOVA).

CO7: Interpret meaningful relationships and patterns in the data in relation to a given business question

Format & Guidelines

You must submit 3 files:

An R file, which contains all statistical work.

A report, **no longer than five (5) pages**.

Present all relevant graphs showing your work and results, ad make meaningful observations.

The report should follow the following format:

(i) Introduction (Develop ideas and concepts about the statistical tools to be utilized. Be substantial and include references).

(ii) Analysis (Separate sections by title, include all relevant figures and tables, including their corresponding titles and descriptive).

(iii) Conclusion (Include your conclusions about the observed outcomes, what do you learnt from the statistical tools used and the results? Use references to support your conclusions).

(iv) References (Minimum two, proper format will be graded)

Report should follow the **APA format**. With the modifications allowed and explained by the instructor during class.

A Power Point file for part 2. This must be a professional presentation (imagine you prepare this presentation for an important meeting).

Assignment Summary

Part 1

Using the data provided in the attached Excel workbook **M5 Data Part 1**.xlsx, apply regression and correlation analysis on five data sets.

Follow the instructions and data presented in the Excel workbook. Then complete a report summarizing your results and observations. Submit both the report and the R script file.

Important: Explain the benefits of each statistical tool you use, including statistical analyses, tables and figures. Explain why you applied those techniques based on the nature of the data and make substantial observations of the results.

Part 2

There are two datasets on the assignment link:

(1) **M5 Data Cars sales India** and (2) **M5 Middle School Grades**.

Open both datasets, analyze them and **choose one** of them for data analysis and presentation.

First, perform a good descriptive analytics analysis to describe the whole dataset, create one slide that summarize and present this information.

Remember: This is a simulation scenario. You work for a Data Analytics company and this dataset belongs to one of your costumers. Your costumer does not know much of data analytics and that's why they hired you. Your job is to help them find meaningful observations, propose questions based on your findings, perform regression and correlation analyses, perform some hypothesis testing if necessary, and make at least one data-base decision.

It is your whole decision how you are going to analyze the data; you choose the statistical techniques and display tools to present the data.

Prepare:

One title slide: Include the name of the project, this title indicates the dataset you are analyzing, tools and goals, e.g., Regression analysis of cases of pulmonary diseases in the Colombia during winters 2010-2018, and their relationship with increases in household power prices. **Avoid** simple titles such as: "Analysis of car sales in India."

One introduction slide: Use bullet points to very briefly present your audience the main topics of your presentation, including but not limited to: the data set, the statistical tool you will use, the reason why you are using those tools, the questions you want to answer, and what are the outcomes you expect to obtain. In this slide include a short list of your presentation organization.

One Descriptive Analysis slide: This slide describes the whole data set. Use one slide, maximum two if you consider important to highlight some meaningful observations.

One or two inferential statistics slides: These slides are the core of your presentation. This is where you surprise your audience with deep analysis of the dataset, why did you do it? How did you do it? What did you find? What do those findings mean? Use maximum three slides if you consider important to highlight some meaningful findings.

One decision making slide: This is basically the recommendation you will make to your costumer about what they can do with the discoveries you made.

One conclusions slide: Don't forget, BE substantial!

One tools and references slide: List the statistical techniques you used (correlation, hypothesis testing (one-tailed, two-tailed, confidence interval, etc.), the computer software you used (R, excel, Phyton, JMP, etc.), list your academic references are recommended readings.

One slide that says: Thank You.

Grading

Your report has a value of 100 points.

Your R file has a value of 10 points.

Your power point presentation has a value of 50 points.

Please remember: your report is very important, make it look professional, make it as short as possible but containing all the relevant information, tell me what you learnt, and using deep critical thinking, provide examples of practical applications.

Northeastern University
College *of* Professional Studies

Rubric

Category	Above Standard	Meets Standards	Approaching Standards	Below Standards	Not Evident
Excel (or R): Problem Modeling & Set-up	Thoroughly and concisely modeled the problem in Excel (or R) for each method	Accurately modeled the problem in Excel (or R) for each method	Satisfactorily modeled the problem in Excel (or R) for each method.	Partially modeled the problem in Excel (or R) for each method, but there are some gaps in the problem modeling and setup	Did not submit or incompletely modeled the problem in Excel (or R)
Excel (or R): Problem Solution & Accuracy	Thoroughly and efficiently obtained correct and accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Thoroughly obtained accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Satisfactorily obtained correct solutions in Excel (or R) by using the appropriate analytic tools of the software	Partially obtained accurate solutions in Excel (or R) by using the appropriate analytic tools of the software	Did not submit or did not obtain accurate solutions in Excel (or R) using the appropriate analytic tools of the software
Word/Report: Problem Description & Introduction	Thoroughly provided a summary of the problem descriptions and introduced the problem using rich and significant ideas	Thoroughly provided a summary of the problem descriptions and problem introduction	Satisfactorily provided a summary of the problem descriptions and problem introduction	Partially provided a summary of the problem descriptions and problem introduction	Did not submit or did not provide a summary of the problem descriptions and problem introduction
Word/Report: Description of Problem Analysis	Thoroughly and accurately described the analytic concepts and theories used in analyzing the problem	Accurately described the analytic concepts and theories used in analyzing the problem	Satisfactorily described the analytic concepts and theories used in analyzing the problem	Partially described the analytic concepts and theories used in analyzing the problem	Did not submit or did not provide a summary of the problem descriptions and problem introduction
Word/Report: Description of Conclusions	Thoroughly described the conclusions and results obtained in the project using a high level of critical thinking and reasoning	Thoroughly described the conclusions and results obtained in the project	Satisfactorily described the conclusions and results obtained in the project	Partially described the conclusions and results obtained in the project	Did not submit or did not describe the conclusions and results obtained in the project
Word/Report: Writing Mechanics, Title Page, & References	Completely free of errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains a minimum of 1000 words	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains a minimum 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 a minimum of 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 less than 1000 words	Did not submit; or there are many errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are totally incomplete; or the report contains very few words