**Northeastern University – D’Amore-McKim School of Business**

**BUSN 6324: Predictive Analytics for Managers** - **Fall 24**

**Instructor Information**

|  |  |
| --- | --- |
| Name: | Prof. Nizar Zaarour |
| E-mail address: | [n.zaarour@northeastern.edu](mailto:n.zaarour@northeastern.edu) |
|  |  |

**Course Overview**

1. This course will start by presenting the concepts of correlation and simple linear regression analysis, and then move into multiple regression analysis, where students will learn how to build multiple regression models and use them in forecasting and analyzing data. In addition, we will discuss the pitfalls and the dos and don’ts of using regression analysis incorrectly and the impact of wrongly interpreting the results. We will learn how to read and analyze the ANOVA output tables by using statistical software tools.
2. Students will learn the basics of regression analysis (linear and non-linear) and the practical application to the real world.
3. A software component will be created for this class, where we will be able to demonstrate the work using all the necessary software and online presentation tools.

**Software**

SPSS / Excel / Minitab / R / Python / …

**Course Organization**

* Class sessions will be in-person.
* Class sessions will be in an interactive lecture/discussion format.
* The course web page is located in Canvas.

**In Class Activities**

* You are expected to attend both class sessions and to participate in the discussions in order to get full credit for class participation.

**Determination of Course Grades**

The following provides a percentage allocation of each component:

Class Participation 40%

Project Write-up: 60%

|  |  |
| --- | --- |
| **Topic and Readings** | **Date** |
| **SLR - Project Initial Discussions**  **Software Tools** | **9/21** |
| **MR / NL Regression**  **ANOVA - Project Final Discussions** | **10/05** |
| **Data Analysis Report** | **Due 10/19** |

**SLR**

* New Data collected about any subject that you like (personal interest, work related)
* “y” Variable to be predicted (what are you trying to predict)
* “x” predictors (at least 1 variable that will be used to predict the “y” collected after the first class)
* Sample size needs to be significant enough (more than 30 elements but keep it under 60). We will discuss larger data sets separately.
* Have the idea ready for the first class, since we will be discussing the initial concepts on that day (you cannot use old data collected in previous classes. This must be new and relevant.)
* You can collect the data yourself through research / surveys / experiments, or you can use the Internet as a source as long as it is a reliable source.

**Title page**

In the title page of your project, you should include the name of the topic, student’s name, course name, and professor’s name.

**Executive Summary / Introduction**

The Executive Summary is a one-page summary of all aspects of the report. It usually includes identification of who wrote the report, the intended audience, the objectives of the project, a brief description of the background and the sample, the major findings and recommendations. A clear description of the subject and the context of the project are indicated here, making sure to specify the sources, sample size, the variables, units for data values, etc.…

**Body of Report (Results and Discussions)**

In the main body of the report, you should present your statistical analysis that includes the results of the analysis (findings from analyzing the data set), and especially your thoughts and comments about them. The relevant numbers and statistics generated must be identified and analyzed. Students must intuitively interpret the output generated. What is being sampled and the sample size must be clear. Points are given for initiative if the report contains relevant analysis beyond what is required and for intuition if the report contains insights about the problem that is not obvious from the questions asked.

The body of the report should have (but not limited to) the following results:

1. ***Numerical*** and ***graphical*** summaries of the data.
   1. Scatter Diagram
   2. Descriptive Statistics of the variables selected (‘x1’, ‘x2’, … and ‘y’)
   3. The estimated regression equation: y = f(x)
   4. ANOVA output tables showing SSE, SST, and SSR, *R2*. Comment on the goodness of fit.
2. Testing for significance:
   1. Use the *F* test or the p-value approach to test for significant relationship. Use α = 0.05
3. What is your conclusion? present and discuss your findings.

**Conclusions**

Summarize all of your findings, discuss the limitations of your study, what questions remain unanswered, and make a suggestion to find the answer for unanswered issues in the project. For example, you may consider questions like these: Do the findings make sense? What else would you like to know about the sample data?

What other data would you collect if you could? What other analyses would you want to do then?

**MR**

* Same “y” Variable to be predicted (what are you trying to predict)
* Additional “x” predictors

**Body of Report (Results and Discussions)**

You are not limited to a linear regression model, nor are you limited to quantitative variables. Just make sure you deal with the collinearity issue and the adequate sample size given the increase of independent variables.

Students must intuitively interpret the output generated.

The body of the report should have (but not limited to) the following results:

1. Residual plots.
2. The estimated regression model: y = f (x1, x2, x3, …).
3. ANOVA output tables showing SSE, SST, and SSR, *R2*. Comment on the goodness of fit.
4. Testing for significance:
   1. Use the *F* test or the p-value approach to test for significant relationship. Use different values of α = 0.1, 0.05, 0.01
5. What is your conclusion? Present and discuss your findings.

**Conclusions**

Summarize all of your findings, discuss the limitations of your study, **what questions remain unanswered**, and make a suggestion to find the answer for unanswered issues in the project.

**Appendix**

Relevant charts, statistical details, and output are to be presented in the appendix. The appendix should be annotated to explain the included chart or output.

**The report has the following constraints:**

1. A one-page executive summary
2. An annotated appendix.
3. Please present your findings in plain English, as if it is to be presented to your boss who is not interested in statistical details.
4. Please keep in mind that your report is to present your own thinking in your own words. It should not be written like answers to homework problems. How well the report is written will be a part of its grading.

Please submit you report as a single document in Word, or PDF format.