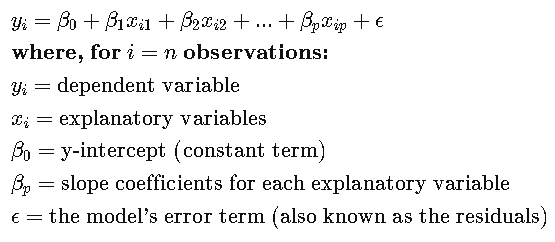
Assignment 1

(Multiple Linear Regression)



**Introduction**

Regression analysis is one of the most widely used techniques in predictive analytics. We use regression analysis for two primary purposes: (1) to assess the relationship between variables and (2) to predict the outcome of a target variable on the basis of several input variables. For example, if a firm increases advertising expenditures by $100,000, then we may want to know the likely impact on sales. Or we may want to predict the price of a house based on its size and location. Regression analysis can be applied in both of these scenarios and many more.

**College Scorecard**

With college costs and student debt on the rise, the choices that families make when searching for and selecting a college have never been more important. Yet, students and parents struggle to find clear, reliable data on critical questions of college affordability and value. For these reasons, the Department of Education (DOE) published a redesigned **College Scorecard** that reports the most reliable national data on college costs and students’ outcomes at specific colleges.

Fiona Schmidt, a college counsellor, believes that the information from the College Scorecard can help her as she advises families. Fiona wonders what college factors influence post-college earnings and wants answers to the following questions:

1. If a college costs more or has a higher graduation rate, should a student expect to earn more after graduation?
2. If a greater percentage of the students are paying down debt after college, does this somehow influence post-college earnings?
3. And finally, does the location of a college affect post-college earnings?

To address these questions, Fiona gathers information from 116 colleges on annual post-college earnings (Earnings in $), the average annual cost (Cost in $), the graduation rate (Grad in %), the percentage of students paying down debt (Debt in %), and whether or not a college is located in a city (City equals 1 if a city location, 0 otherwise).

The table below shows a snapshot of the data.

A picture containing text, screenshot, font, number

Description automatically generated

Using the data:

1. What is the sample regression equation?

2. Interpret the slope coefficients.

3. Predict annual post-college earnings if a college’s average annual cost is $25,000, its graduation rate is 60%, its percentage of students paying down debt is 80%, and it is located in a city.

**Software**

Use 'R' software only to do the analysis. Submissions from using other Softwares are not accepted and evaluated.

**Deliverables**

1. All your answers should be in a compact 4-to-5-page word document report with appropriate graphs, results in tables and subjective interpretation of findings in plain English. Format the document properly. Do not put the 'R' codes in the report.

2. The R code used to derive the results should be in a separate text file only.

**Data**

The dataset is in the file 'college.csv'.

**Do's and Don’ts**

1. Each of your files should have your name, email ID, other relevant details clearly mentioned at the beginning of the report in the word file and the code file.

2. Also name your report file and code file with an appropriate name plus your name.

3. Please upload the report file and code file in LMS only. Email submissions are not considered.

**Hint**

