Boston University
Department of Biostatistics
Department of Mathematics and Statistics

BS755/MA575 – Fall 2023

Linear Models

Lab Session 4: Ordinary Least Squares

Project Report 2 Due: October 5, 2023, by 11:59pm (on Blackboard)

Regression models are used to estimate the relationships among variables. Regression analysis helps us to understand the changes in one noisy variable given that the other variables are changed. Regression is commonly used for the purposes of prediction, inference and to determine statistical relationships between the variables (statistical significance).

The purpose of this lab session to learn how to use R to perform ordinary least squares. You will then build linear regression models for your project data set.

Pre-lab activities

1. Download files from Lab 4 item on Blackboard. Make sure that you have AirQualityData.csv file downloaded from Lab 3. We will be using this data set to implement ordinary least squares.

Lab activities

1. Run the Lab4.R file to perform ordinary least squares on the air quality data set.

2. Perform ordinary least squares on your project dataset.

3. Discuss feedback on project report 1.

Post-lab activities

1. Continue exploring and visualizing your project data and discuss with your team.
2. You have project report 2 due. See instructions below.

Project Deliverable: Ordinary Least Squares
Due Date: October 5, 2023, by 11:59 pm
Report Details:

- Lab Section: C1/C2/C3/C4
- Team Members: [List all team members' names in the alphabetical order of the last names]

Report Content:

Part I

1. Updates to *Project Report 1: Description of Potential Topic Areas.* Based on your exploratory data analysis for your selected project data set and the feedback from the teaching team,
    a. Contemplate revisions/updates you would like to make to your *Project Report 1*.
    b. Specifically, consider any updates required for the three questions of interest and the motivation behind considering them.

Format: One page, 12-point Times New Roman font, 0.5-inch margins on all sides.

Part II

Considering the three questions from your chosen project data set (*Bike Sharing*, *BMW Pricing Challenge* or *Facebook Social Media Metrics*):

2. Choose a response variable (Y) and one covariate (X). Deliberate thoughtfully and explain the rationale in choosing these variables for Ordinary Least Squares (OLS).
3. Plot Y vs X (i.e., a scatterplot) from your data set using R. Include your code, the scatterplot along with your interpretation of it.
4. Perform OLS using R on your response Y and covariate X. Submit the R code and its output of OLS (coefficient estimate, t values, p values) and interpret the results.
5. Overlay the linear regression fit on the scatterplot using R. Include your code and the scatterplot. Is this a good model? Explain your findings.
6. Author Contribution Statement: Include an author contribution statement indicating your roles/contributions in preparing this lab report. See example provided in the `ProjectDescriptionFall2023.pdf` file under Lab 2.

Format: Four pages maximum. You can use the R Markdown to help you create this (`.pdf`) document. See `Lab3.Rmd` and/or `Lab4.Rmd` for example.

Submission Instructions:

- Combine the .pdf files from Part I and Part II into a single .pdf file to be submitted.
- Designate one team member to submit this combined document through Blackboard: `Content → Project and Lab Materials → Lab Reports (Upload)`
- Click on `Lab Report #2` and upload the document.

Ensure the report adheres to the specified format and contains all the required information.