ISyE 3030 – Basic Statistical Methods

Project Guideline: Data Analytics Project

The goal of this project is to use the statistical methods learned in this class to solve real-world problems. It is expected that students will work in a team, laying out a clear goal of the project (e.g., what the project aims to find out). To support the accomplishment of the goal, the team needs to collect a dataset or use a public dataset (in the latter case, the source needs to be cited), then use summary statistics and graphical methods to describe the data, perform statistical inference and regression analysis, finally draw conclusions.

Some examples are:

* To find out if there is a difference in daily coffee consumption between males and females, and if the daily coffee consumption is a predictor for sleeping quality, and further if there is a difference in the coffee consumption-sleeping quality relationship between males and females.
* To estimate the difference in the average number of hours studied per week at the second half of the semester and that at the second half of the semester, and which (the average number of hours studied per week at the second half of the semester, or that at the first half of the semester, or the difference between the two averages) is a better predictor for the end-of-semester grade.

**Team**: Each project is done by a team of **4-5 students**. Please form teams by yourselves and pick a topic on your own.

**Deliverables**:

* A project report submitted by each team to Canvas. The report should include the sections in the project outline below. **Format**: The report should be a pdf file, ≤3 pages, single-line space, 11 pt. font size, Times New Roman font style, 1’’ margin.

**Project outline**:

**ISyE 3030 Project: title of the project**

**Team members’ names:**

(Each report must include section I-VI below using the provided heading for each section. However, the bullet points in each section are only suggested and you can choose to include more or other contents.)

I. Introduction

* Explain the background and purpose/goals of the project.

II. Data Collection and Preprocessing

* Describe the data sources and collection methods (you can collect a dataset yourself or use a public dataset. If you use a public dataset, the source should be cited).
* Discuss the steps taken to preprocess the data (e.g. cleaning, filtering, formatting).

III. Descriptive Statistics

* Present the descriptive statistics of the dataset.
* Use graphical tools to visualize the data.

IV. Statistical Inference

* Formulate a hypothesis to test.
* Explain the chosen test (e.g. t-test, chi-squared test) and why it is appropriate.
* Perform the test and interpret the results.

(Alternatively)

* Estimate a parameter of interest using a point estimator, or confidence interval
* Discuss properties of the estimator if applicable (bias, variance, MSE)
* Interpret the results

V. Regression Analysis

* Perform a regression analysis to model the relationship of the variables in the dataset.
* Estimate parameters of the regression and perform statistical inference of the parameters
* Assess model adequacy
* Discuss implications of the findings

VI. Conclusion

* Summarize the findings and their significance.
* Discuss limitations of the project and potential future research directions.

**Project due date**:

* **Project report is due by the end of your final exam date**. Submit your project report to Canvas. Each team only submits one report.