Introduction to Statistical Inference Final Project Description

Mohammad-Reza A. Dehaqani
Amirali Soltani Tehrani

University of Tehran
Department of Computer and Electrical Engineering

January 15, 2024

1 Visualization and Summarization of Dataset Variables

- Perform exploratory data analysis by visualizing categorical, numerical, and mixed variables. Use appropriate plots like bar charts for categorical data and histograms for numerical data.

- Analyze the distribution and relationship of variables. Create scatter plots, box plots, and QQ-plots to understand variable dependencies.

- Identify and report any dependent or independent factors discovered during analysis.

- If the dataset comes from a published source, include and discuss any visualizations provided in the publication.

2 Parametric Inference and Estimation

- Conduct parametric inference methods where applicable, explaining the choice of parametric tests.

- Apply estimation techniques to various variables, such as point estimation or maximum likelihood estimation.

- Calculate and interpret confidence intervals for the estimated parameters.

- Utilize interpolation methods to predict missing or unobserved data points, if necessary.

3 Hypothesis Testing and Statistical Analysis

- Formulate and test hypotheses based on the dependency of variables. Clearly state the hypothesis, the rationale for testing, and the method used.

- Report all statistical findings, including test statistics, p-values, and conclusions drawn.

- Critically analyze and discuss any tests presented in associated publications, and compare them with your findings.

- Employ bootstrap or resampling methods for hypothesis testing, especially in cases of non-normal data distributions.

- Conduct Analysis of Variance (ANOVA) tests to compare means of different groups. Use correction methods like Bonferroni or Tukey's HSD for multiple comparisons, if applicable.

4 Regression Analysis and Reporting

- Perform regression analysis to investigate relationships between variables. Explain the choice of regression model and interpret the results.

- Include a detailed analysis of regression findings from the publication, if available.

- Prepare a comprehensive report summarizing all tests, analyses, and findings. Highlight the significance and implications of the results in the context of the dataset and research question.

Additional Information

- Students can produce a maximum of 8 different results for each of the four sections above, totaling a maximum of 32 results.

- The maximum achievable grade is 32, but the final grade will be calculated out of 20. Any grade beyond 20 will be considered as bonus grades.

- Students are required to submit a document (draft) detailing their project work, methodologies, and results. In academic terms, a 'draft' refers to a preliminary version of your work. It should comprehensively outline your project's progress and findings, though it may not be in its final, fully polished state. This draft serves as an essential step in developing and refining academic reports.

- Additionally, students must create a video presentation summarizing their findings and interpretations (i.e., record their screen while presenting their presentation). The presentation should be submitted along with the project document. To avoid having to upload large files, you can (but are not obligated to) use compression software.

- **Final Word:** In addition to adhering to the instructions in the four outlined sections, students are expected to engage in further research and decision-making as data scientists on their own. Should certain aspects within these sections not be fully explained, you are encouraged to independently seek out necessary information and decide on appropriate approaches. While support will be available, your initiative in exploring beyond the provided guidelines and justifying your methodologies is essential and highly valued in this project. Think of this as your first work beyond the realm of "Courses and Grades". Do not worry about grades, and focus on doing work you are interested in. But remember, no effort shall go unrewarded (or ungraded!)

Good luck on your journey with the data,
Your Teaching Team