

STAT 374: NONPARAMETRIC INFERENCE

Data Analysis Project

As noted in the syllabus, the graduate section of the course, Stat 37400, involves a data analysis project in addition to the written exam. This project requires analysis of a data set, to be chosen by the student. The analysis should include a comparison of parametric against nonparametric methods. A 4- to 6-page report is expected, with complete description of the data, the statistical problem, the methods used, and the findings. As an example of such a data set and analysis, recall the “Capital Bikeshare” problem from Assignment 1.

Completing this project gives enrolled graduate students an extra opportunity to demonstrate their mastery of the nonparametric inference methods. This normally reduces the bias of the letter grade as an estimator for evaluating their course performances. You are welcome to analyze datasets encountered in your own research, or datasets acquired from another statistics course with a focus different from nonparametric statistics. In these scenarios, please include sufficient descriptions of the source of the dataset (e.g. the particular project you are working on, or the course from which you acquired this dataset).

Your report should include the following parts.

1. ***Description of the data.*** Describe the data, including where it came from, how it was collected and the meaning of the variables. Clearly state the objective of analyzing the data, and any preprocessing that was carried out on the data, such as handling of outliers. Give plots that summarize and visualize the data.
2. ***Parametric analysis.*** Describe the application of a parametric model to the data. What assumptions are made by the model? Fit the model to the data, and indicate the level of uncertainty in all of your estimates through confidence intervals. Describe alternative models, and how you chose the particular parametric model you analyzed. Include descriptive plots.
3. ***Nonparametric analysis.*** Describe the application of a nonparametric model to the data. What assumptions are made by the model? Fit the model to the data, and include a measure of uncertainty (e.g., confidence bands) for the model. As above, describe the strengths and weaknesses of alternative nonparametric models. Include descriptive plots.
4. ***Discussion.*** Discuss your results, comparing the parametric and nonparametric approaches. What conclusions can you draw? Which is more effective or appropriate? If your analysis is inconclusive, describe additional work that could be carried out.

Students are encouraged to thoughtfully consider many different aspects of statistical modeling and theory in their writeups. Grading will assign equal weight 25% to each of these four parts,

based on the quality and completeness of each section. The written exam and project will together make up 30% of the final exam grade, with the written exam weighted $\frac{3}{4}$.

Pass in your report to Jones 316 by 4:00 pm on Tuesday, December 3, 2019.