

Stat 415/615, Regression, Course Project Spring 2021

Project: Data analysis using regression

This project is for all students in Stat 415 and Stat 615. You will work in groups of up to 4 people on a data analysis project. Each group will present the results as well as answer the questions below in class. Also, each group will submit the slides of their presentation on Bb 3 hours before the presentation. Make sure I get the names of all participants.

The project can be one of the following:

- a. Build multiple regression models and analyze data from real world to address research questions of interest.
 - b. Review and provide critiques to a published article that used regression methods for data analysis.
2. You are responsible for choosing the topic of the project and looking for appropriate data sets, articles, reference materials, etc. I will serve as your “consultant” throughout your project.
3. You will discuss your project ideas with me by the end of February.
4. During the scheduled final exam time, each group will give a 20 minutes’ presentation on your project and have 2 min for Q&A. In addition, the following work will be discussed:
- a. The question of interest, the variables under consideration, the methods to collect data, proposed models, etc.
 - b. The data set for your analysis, the articles being reviewed and a list of reference. If you review a paper that comes with the data set, attach the data set as well.
 - c. The presentation material (power point or notes/handouts).
5. You may divide the workload within group, and each group member shall contribute to the project. In addition, each group member is expected be familiar with every aspect of the project even if he/she does not work on that portion. A within group self-evaluation will be conducted and accounted towards the project grade if needed.

In addition to the data analysis project, **graduate students** will work in small groups of 3-4 people to self-study a statistical method. You will give a short lecture to introduce the method to the class. The statistical method of your study can be (1) a regression technique that is not discussed in detail in class, or (2) an alternative method to linear regression. Some possible topics are listed below. Other topics are welcome.

- a. Non-parametric regression, such as classification and regression tree.
- b. Least absolute shrinkage and selection operator (lasso), least-angle regression (lars), elastic net, etc.

You are responsible for choosing the topic of the project and looking for appropriate reference material. I will serve as your “consultant” throughout your project. You will discuss your project ideas with me by end of February.

At the end of the semester, you will give a 20 minutes’ lecture to introduce the method you have studied. You can give the lecture in your own style. However, it should include, but is not limited to, the following:

- a. An introduction of the statistical model, including the assumptions.

- b. A discussion on when and where the method can be applied. Give an example.
- c. The interpretation of analysis results from the method.
- d. A discussion on the benefits and cautions of the method.
- e. Extra credit: example and/or software demonstration.

Guidelines to your project and presentation

1. For a data analysis project, it should focus on solving the applied problem, and it should be accessible to the “educated layperson” audience.
2. Concisely and clearly describe the research question of interest and its background.
3. Explain the design of the study and how the data are collected. This would include, but is not limited to, a discussion of study design, sampling process, target population, etc., where applicable.
4. Concisely and clearly describe and present the data. Use appropriate exploratory statistics and graphs to summarize the data.
5. Present the statistical methods used for analysis, including a brief discussion of advantages, disadvantages and necessary assumptions of methods.
6. Make statistical inference on the data and discuss the statistical results in the context of the problem.
7. Make concise conclusion in the context of the problem. Discuss limitations of the analysis, if applicable.
8. Organize and present your work (semi-)professionally. (Think of a conference or a job interview.)
9. Extra points: Correctly use statistical methods that are not discussed in class. (Data analysis.)