STATS506 Project Proposal

Fan Zhang zff@umich.edu

Dataset and Topic

- Dataset: The Commercial Buildings Energy Consumption Survey (CBECS)
- Research Topic: What does the energy consumption-number of floors ratio change as the number of floors increases?

Research Plan

- **Data cleaning and wrangling**

  The dataset comes from real-world survey, there may probably be some missing values and outliers. To ensure the reliability of my analysis, I will do some data cleaning and wrangling work such as dealing with missing values, outliers, and data types.

- **Ordinary Least Squares (OLS) regression**

  To evaluate the relationship between the energy consumption-number of floors ratio and the number of floors, the basic and most primary method is to fit a linear regression model. I will use OLS regression to fit the model and get the coefficients. This is the primary method I will use in this project. It helps to get into the data and get a general idea of the relationship.

- **Model Evaluation**

  I will explore the model assumptions and evaluate the model performance by plotting the fitted values and residuals, and by calculating some metrics such as R-squared, adjusted R-squared, and AIC. Probably the simple OLS model does not work well. I will try some other complex models and do model selection. Furthermore, the relationship may be influenced by other factors. I will evaluate whether there is any difference of the pattern of change among different building types or different regions.