Stat 220 Lab. 3: Neighborhood Homeowners Association

September 19, 2023

Background

In the quiet and rapidly evolving neighborhood of Willow Creek, a group of proactive residents have decided to establish a homeowners association. The primary goal of this association is to foster a close-knit community that promotes the well-being and prosperity of its residents. As a part of their initiatives, the association is keen to gather detailed insights about the residents to tailor community programs and amenities that cater to the needs and preferences of everyone in the neighborhood.

Scenario

To achieve this, the association has launched a comprehensive survey to collect data on various aspects including the income levels, house prices, ages, heights, and other variables from the residents. This data, they believe, will help them in understanding the current demographic and economic makeup of the neighborhood. The result of this survey can be found at `https://richardson.byu.edu/220/neighborhood_survey.csv`.

The association is particularly interested in knowing which variables follow a bell curve, which we know is a normal distribution, as this would assist in making informed decisions on policy formulation and community development projects. For instance, understanding the distribution of house prices could help in setting fair and reasonable association fees, while knowing the age distribution might assist in planning community events that cater to the majority age group.

Your Role

As budding statisticians and data scientists, your task is to assist the association in this endeavor. You are required to analyze the data to find the maximum likelihood estimates of each variable under the assumption of a normal distribution. Furthermore, you are to verify the empirical rule for normally distributed data, which states that:

- Approximately 68% of the data falls within 1 standard deviation of the mean.
- Approximately 95% of the data falls within 2 standard deviations of the mean.

Your analysis will help in determining which variables can reasonably be assumed to follow a normal distribution and which cannot.

Instructions

1. Load the provided data set from `https://richardson.byu.edu/220/neighborhood_survey.csv`

2. For each variable in the data set:
    (a) Calculate the maximum likelihood estimates (MLE) for the mean and standard deviation assuming a normal distribution.
    (b) Calculate the percentage of data falling within 1 and 2 standard deviations of the mean.
    (c) Create graphical representations (e.g., histograms or density plots) to visualize the distribution of the data.
3. Based on your analysis, identify which variables adhere to a normal distribution and which do not. Justify your conclusions with the calculations and graphical representations you've created.

Submission

Please submit a detailed report of your analyses, including graphs and conclusions. In your report, also include a section where you discuss the potential implications of your findings for the homeowners association, suggesting how the information might be used to foster a better community living experience.