

Final Project | STAT 419

Abby Drongpa, Alex Sullivan, Chris Liu, Jett Palmer, Sumanth Thokala

Introduction

Our data set is a subset of home in Cincinnati, Ohio, from the year 2002. It includes 552 observations and seven variables, each representing an individual house. This data set is meant to identify factors that influence and may predict the selling of homes.

Our ordinal grouping variable, “Pricerank”, categorizes home based on their selling prices. Houses sold for less than \$190,000 are coded as 1, those sold between \$190,000 and \$285,000 are coded as 2, and houses sold for more than \$285,000 are coded as 3.

The remaining six variables are explanatory:

- Area: The size of the house in square feet.
- BR: The number of bedrooms.
- BA: The number of bathrooms.
- Garage: The number of cars the garage(s) can accommodate.
- Quality: An index of construction quality, where 1 indicates high quality, 2 indicates medium quality, and 3 indicates low quality.
- Age: The age of the house as of 2002.

Graphs and Summary Statistics

Discriminant and Classification Analysis

Correlated Quantitative Variables

Discriminant Analysis

Classification Analysis

Summary

R Code