Real Estate Pricing Analysis

Instructions for Lab #6

Statistics 111 - Probability and Statistical Inference

Lab Objective

The purpose of the lab is to perform an exploratory data analysis.

Lab Procedures

What is the most important factor in determining the selling price of a house? Is it the size of the house, the location, or the number of rooms? In this lab, we examine a survey of real estate data from across the country. The data comprise 6,068 sales of detached single-family homes. They were collected by the Department of Housing and Urban Development. The variables on the data set are described in the code book appended to this lab.

Open the data file soc11.dta.

Data Analysis Caveat: The answers you get here are incomplete, in the sense that we’re only looking at bivariate relationships. It is possible to predict sales price from more than one variable at a time using multiple regression. This is the preferred approach and will be discussed later.

Questions:

When appropriate, provide number summaries that justify your answers. Include brief interpretations of those numbers (e.g., “the correlation between xxx and yyy is -0.87, which means the two variables have a very strong, inverse relationship”). You don’t need to provide graphical displays, but you can allude to them when appropriate. Keep your total writing to one page if possible.

All mentions of prices below refer to the Sales Price.

Further, there are various flags for missing values.

1. Relations of prices with housing characteristics

   (a) Describe the relationship between price and number of bedrooms.

   (b) Is the ratio of number of bathrooms to number of bedrooms a good predictor of sales price?

   (c) Describe the relationship between price and square foot area of house (not final area).
(d) Describe the relationship between price and lot size.

2. Relations of prices with amenities

(a) What percentage of houses have central air-conditioning? How does this vary between states in the South versus the rest of the country?

(b) How much do houses with central air-conditioning cost relative to houses without it? How does this vary between states in the South versus the rest of the country?

(c) What percentage of houses have a garage (i.e. all sizes of garages), or no garage at all (i.e. other options)? How does this vary between states in the South versus the rest of the country?

(d) How much do houses with garages cost relative to houses without them? How does this vary between states in the South versus the rest of the country?

3. Relation of prices with geographic characteristics

(a) Describe the relationship between prices and whether the house is in a community association or not.

(b) Describe the relationship between prices and census division.

(c) Describe the relationship between prices and whether the house is in a MSA or not.