

# Saratoga Real Estate

A real estate agent in Saratoga NY wants to know what and how home characteristics impact sales prices. In particular, the agent is interested in how fireplaces affect sales prices and the relationships between the sales price and other variables. The R workspace *saratoga.Rdata* contains home characteristics and sale prices information for homes from Saratoga NY. For the statistical inference, use  $\alpha = 0.05$  for all statistical inference, clearly state the competing hypotheses, and provide interpretations/conclusions.

1. Begin with a thorough exploratory data analysis. For each item presented, provide a discussion of any observations and insights you find.
2. Fit a model that uses size to predict price, denote this as model #1.
  - (a) Is there evidence the line does not pass through the origin? Answer this question using a confidence interval.
  - (b) If the line passes through the origin, then the slope is a proxy for the price per square foot. Is there evidence the price per square foot is less than \$100 per square foot? Answer this question using a hypothesis test.
  - (c) Is there evidence the residuals do not have a Normal distribution? Answer this question with the appropriate visualization and hypothesis test.
  - (d) Is there evidence the fireplace variable is needed in the model? Answer this question with the appropriate visualization and numerical statistics. If you find that the fireplace variable is needed in the model, what condition is violated for model #1?
3. Fit a model that uses the fireplace variable to predict price, denote this as model #2.
  - (a) What is the baseline or reference group?
  - (b) Is there evidence the change in the average price is not zero dollars when changing from homes without a fireplace to homes with a fireplace? Answer this question using a hypothesis test.
  - (c) Refer to the previous part. What statistical procedure is the hypothesis test equivalent to? Specify the corresponding competing hypotheses.
4. Fit a model that uses all of the numeric variables to predict the price, denote this as model #3.
  - (a) Is there evidence of collinear predictors? Answer this question with the appropriate visualization and numerical statistics.
  - (b) Is there evidence at least one of the acreage or age variables are needed in the model? Answer this question using a hypothesis test.
  - (c) Is there evidence the variation of the residuals is heteroskedastic? Answer this question with the appropriate visualization and hypothesis test.

5. Fit a model that uses the size, number of baths, number of bedrooms and the fireplace variable to predict the price, denote this as model #4.
  - (a) Is there evidence the change in the average price is not zero dollars when changing from homes without a fireplace to homes with a fireplace? Answer this question using a hypothesis test.
  - (b) Refer to model #2 and part 3 (b). Explain why the results are different using model #4.
  - (c) From model #4, identify any outliers. Explain what it means for an observation to be an outlier in this context.
6. Fit a model that uses the size, number of baths, number of bedrooms, fireplace variable, and an interaction between the size and fireplace variable to predict the price, denote this as model #5.
  - (a) For homes with a fireplace, what is the slope between size and price.
  - (b) Is there evidence the interaction term is needed in the model? Answer this question using a hypothesis test.
  - (c) Explain what an interaction between the size and fireplace variables means in the context of the problem.
  - (d) Are there any omitted variables that may create endogeneity bias? For each, indicate where the endogeneity bias may appear and if an instrument variable model would be appropriate to use in this situation.
7. How could the sales prices be spatially correlated? Explain.
8. How could the sales prices be temporally correlated? Explain.
9. Are these results useful for a real estate agent in San Luis Obispo CA? Explain.