

Chapter Project

Home Sweet Home: Transforming Data Prior to Regression Analysis

Use the data set named Mount Pleasant Real Estate Data which contains information about properties for sale in three subdivisions of Mount Pleasant, South Carolina in the year 2017.

Data

The data can be found at stat.hawkeslearning.com

Data Sets > Mount Pleasant Real Estate Data.

1. Classify the three variables *List Price*, *Squarefootage*, and *Subdivision* as qualitative or quantitative and provide the level of measurement as described in Section 2.2.
2. Which of the quantitative variable(s) in the previous question should be considered the dependent variable (as described in Section 5.2)? Why?
3. Use technology to make a histogram for *List Price* and describe the distribution.
4. Create a scatterplot of *List Price* vs. *Squarefootage*. That is let $y = \text{List Price}$ and $x = \text{Squarefootage}$ and describe any patterns.
5. Can we use Pearson's correlation coefficient to measure the strength of the linear association between *List Price* and *Squarefootage*? Why or why not?
6. Is the simple linear regression model appropriate for predicting List Price based on *Squarefootage*? Explain your answer by investigating the model assumptions.
7. Take the natural log of List Price and call it *LnPrice*. Make a histogram for this variable
8. Describe the distribution of *LnPrice*.
9. Create a scatterplot of *LnPrice* vs. *Squarefootage* and describe any patterns.
10. Calculate Pearson's correlation coefficient between *LnPrice* and *Squarefootage* and interpret it.
11. Using technology, build a simple linear model to predict *LnPrice* from *Squarefootage*.
12. Use $\alpha=0.05$ and test to see if the model is significant.
13. Predict the price of a 3045 square foot house.
14. Find the residual of the predicted value of the house in the previous question if the house is actually the house from the data set with ID = 27 in Carolina Park with a price of \$575,000.