Kaggle Project - Team Fat Tails

Grant Bourzikas
Quinton Nixon
Brock Friedrich

Contents

Introduction

```
#setup, echo=FALSE, results="hide"}
# read setup script
#source("Setup.R")
print("hello world!!!")

print("Verbose will be printed first")
```

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. However, it is essential to review the data because it proves that there are many other influences in price negotiations than the number of bedrooms or a white-picket fence.

Data Synopsis

The Ames House dataset was compiled by Dean De Cock and contains 79 explanatory variables describing almost every aspect of residual home in Ames Iowa from 2006 to 2010. The data set contains 2930 observations involved in assessing home values.

Analysis Question 1

Restatement of Problem

To build and fit a model, an analysis must be performed to identify features of the dataset that are statistically significant in their relation to, and prediction of, the sales price.

Build and Fit the Model

Interrogate the Data

To build and fit a model, an analysis must be performed to identify features of the dataset that are statistically significant in their relation to, and prediction of, the sales price.

- Plot the data.
- Develop a tentative model(s).
    - Using the question(s) of interest (QOI).
    - Accounting for confounders.
    - Accounting for relationships (X^2,X^3, etc).
- Fit the model(s).
- Evaluate residual plots.
    - Constant SD.
    - Normality and zero me an.
    - Identify any influential observations.

Assumptions

Homogeneity of Variances

Neighborhooods: $x_1 = BrkSide$ $x_2 = NAmes$ $x_3 = Edwards$

• Ames^SalesPrice = B0 + B1$BrkSide$ + B2Edwards + B3$NAmes$ + B4(LogLivingArea*BrkSide) + B5(LogLivingArea*Edwards)

$\hat{\mu}\{SalesPrice\} = \beta_0 + \beta_1 \, x \, BrkSide + \beta_2 \, Edwards + \beta_3 * NAmes + \beta_4(LivingArea_{log} \, x \, BrkSide) + \beta_5 \, x \, (LivingArea_{log} \, x \, Edwards)$
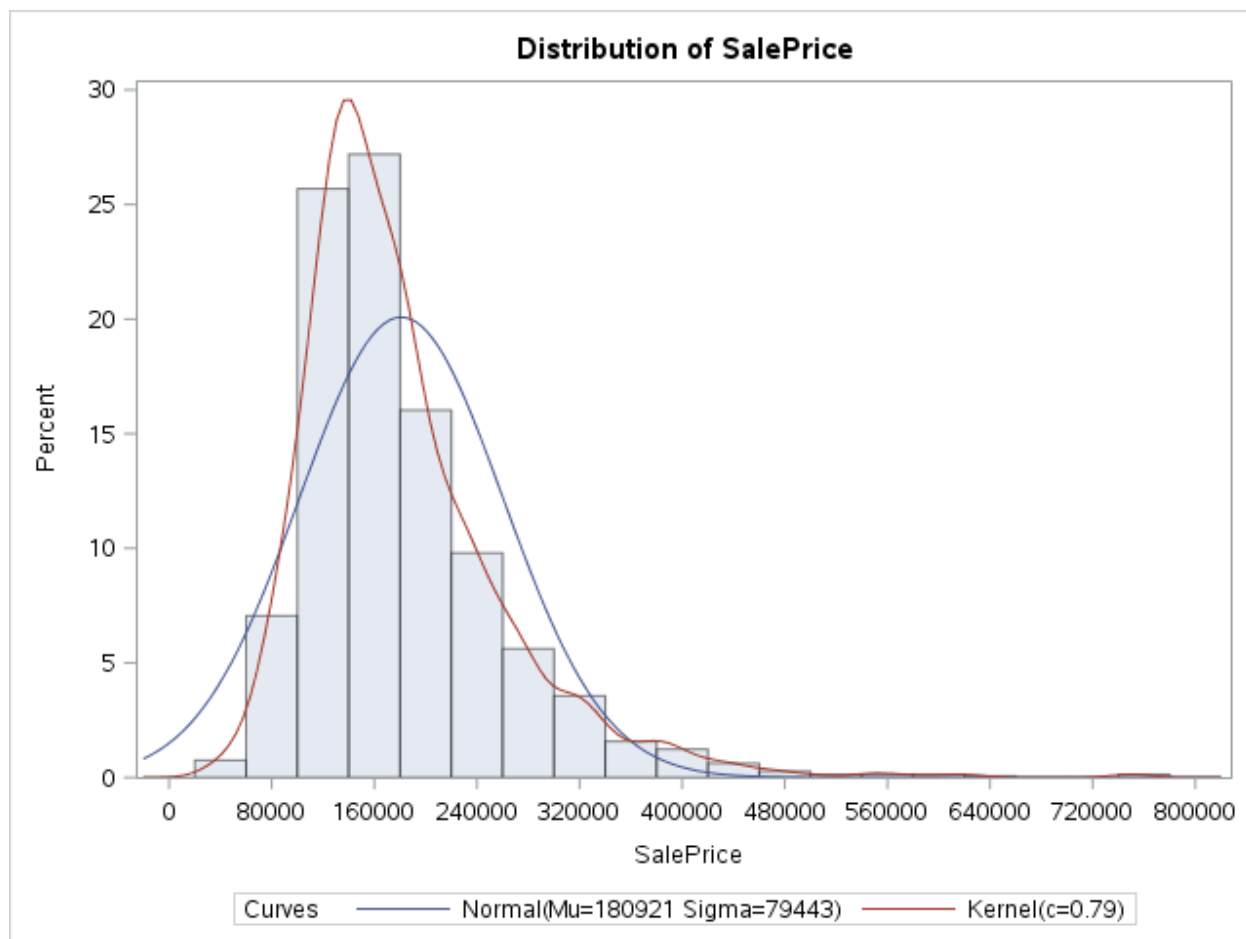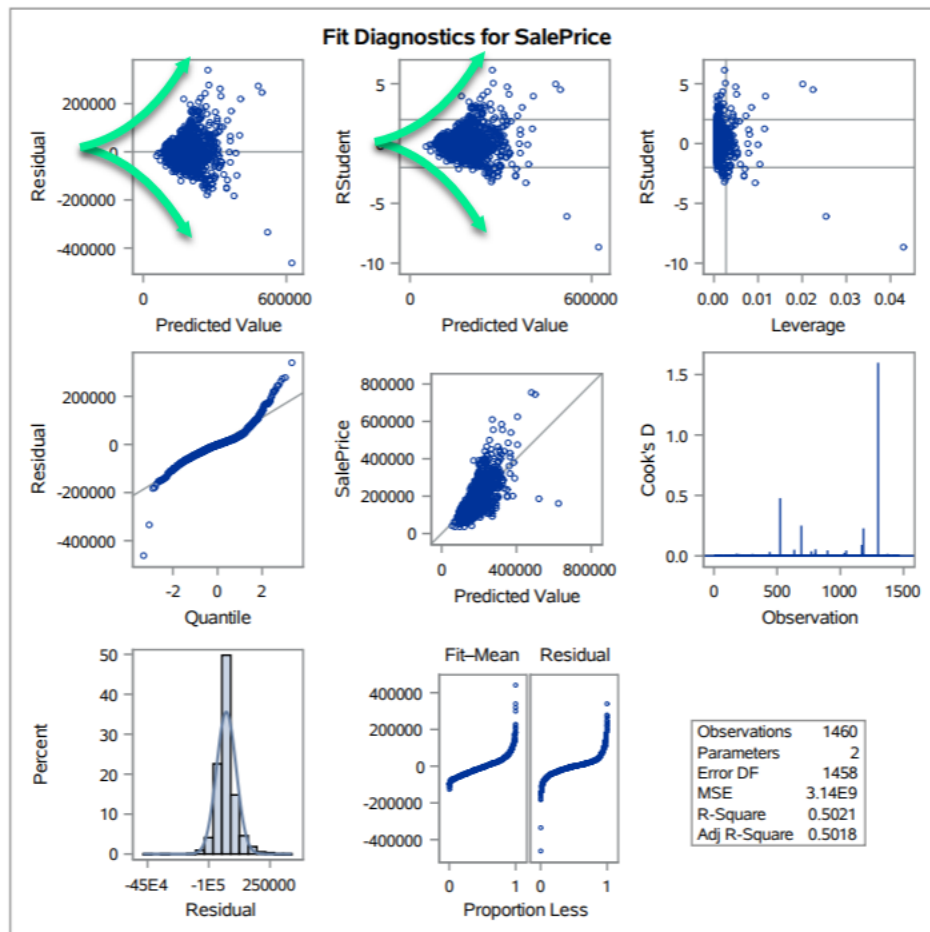
Figure 1: train_original_histogram_salesprice

Figure 2: train_original_histogram_salesprice

$$\mu\{\widehat{SalesPrice}_{Ames}\} = \beta_0 + \beta_1 \, x \, BrkSide + \beta_2 \, Edwards + \beta_3 * NAmes + \beta_4(LivingArea_{log} \, x \, BrkSide) + \beta_5 \, x \, (LivingArea_{log} \, x \, Edwards)$$

% • Ames^SalesPrice = 8.49 + (-2.58*BrkSide)* + *(-0.49*Edwards) + (0.0 * NAmes) + (B3*0.0)* + *B4(0.47*BrkSide) + B5(0.47*Edwards)

Analysis Question 2