Home Price Regression Project

Using your home price data set, do the following.

1. Provide the data set and summary statistics of all of the variables.

2. Provide plots of the response variable versus the individual predictor variables.

3. Provide the correlations between all pairs of variables.

4. Regress price on square footage.
   (a) Find the best model (using transformations or polynomial regression if needed).
   (b) What proportion of the variation in price is explained by your model?
   (c) Perform residual analyses to assess the model assumptions of independent, normally distributed errors with constant error variance. In particular, do a modified Levene test for constancy of error variance and a correlation test for normality of errors.
   (d) Is square footage a statistically significant predictor of price? Use $\alpha = .05$.
   (e) If there are duplicate square feet measurements, perform a lack of fit test for your model.
   (f) Estimate the mean price for a 2000 sq. ft. home with a 95% confidence interval.
   (g) Predict the price of an individual 2000 sq. ft. home with a 95% prediction interval.
   (h) Estimate the change in price for a 100 sq. ft. increase in area with a 95% confidence interval.

5. Determine which of the remaining variables should be added to your model.
   Use the best subsets Minitab command to provide 3 possible subsets of variables on which to regress price. Use $C_p$ criterion and adjusted $R^2$. Also use stepwise regression.
   (a) Find the best model.
   (b) What proportion of the variation in price is explained by your model?
   (c) Perform residual analyses to assess the model assumptions of independent, normally distributed errors with constant error variance.
   (d) Estimate the mean price for a 2000 sq. ft. home (with the other variables at their average level) with a 95% confidence interval.
   (e) Estimate the change in price for a 100 sq. ft. increase in area (holding the other variables constant) with a 95% confidence interval.
   (f) What other unobserved predictor variables do you think may improve your model if the data were available? Explain.

6. Write a two-to-three page typed summary of your findings.

You will be graded on neatness, clarity, accuracy, and completeness.