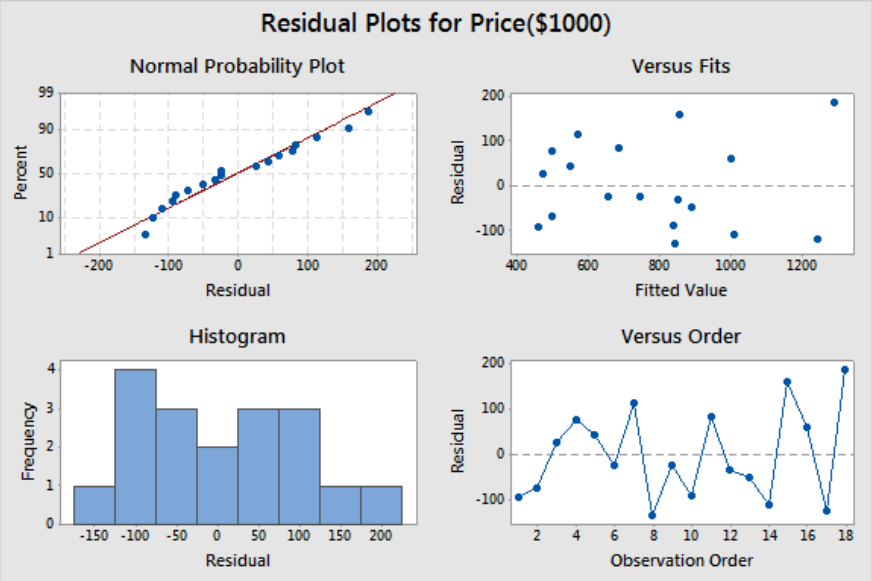
Regression Diagnostics
COR1-GB.1305 – Statistics and Data Analysis

Model Assumptions

1. Here are plots of the residuals from the least squares fit to the housing data.



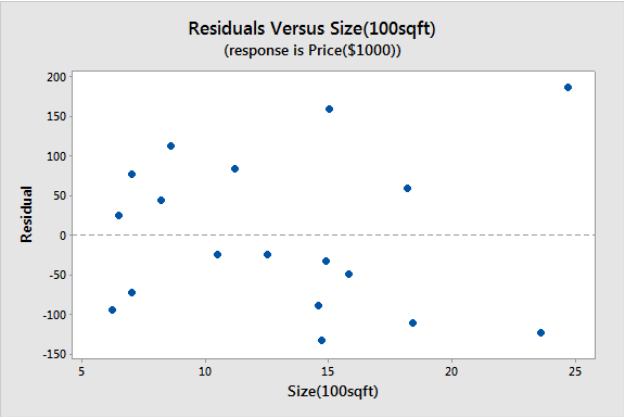
Do the plots indicate any potential violations in assumptions? Specifically, answer the following questions.

(a) Do the residual errors look approximately normal?

(b) Does the error variance look constant?

(c) Is there any apparent dependence in the residuals?

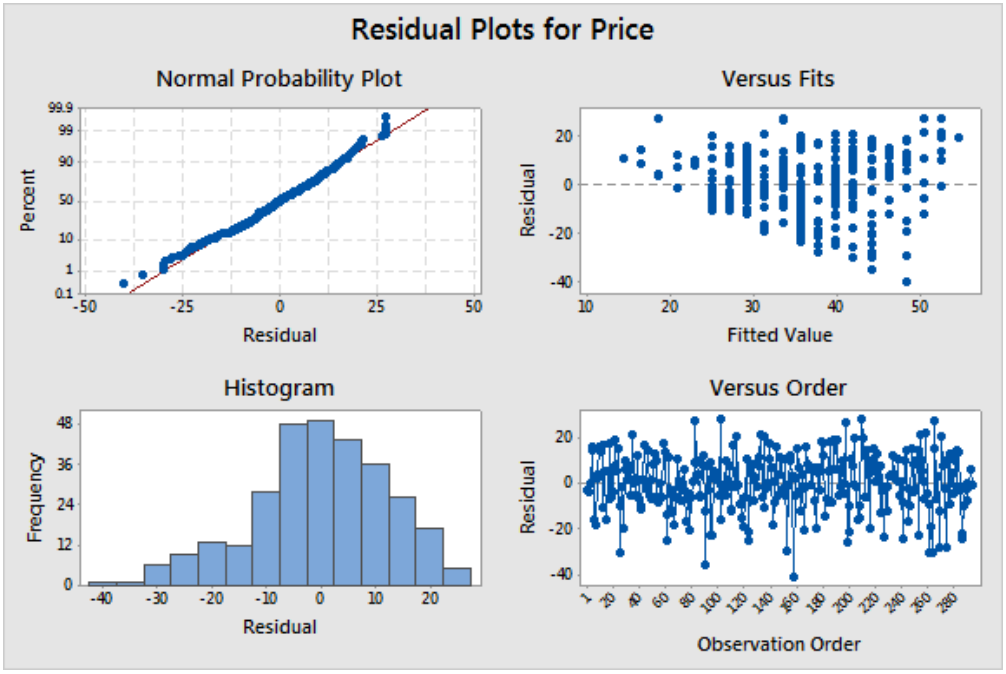
2. Here is a plot of the residuals versus Size (x).

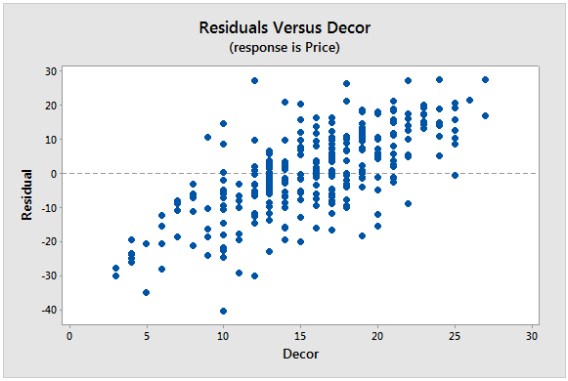


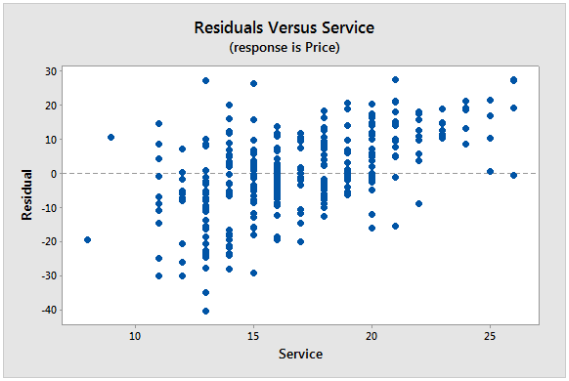
(a) Why is this plot nearly identical to the plot of residuals versus fits?

(b) Does the plot of residuals versus fit always look like the plot of residuals versus x?

3. Here are some plots of the residuals from the fit of Price to Food for the Zagat data:



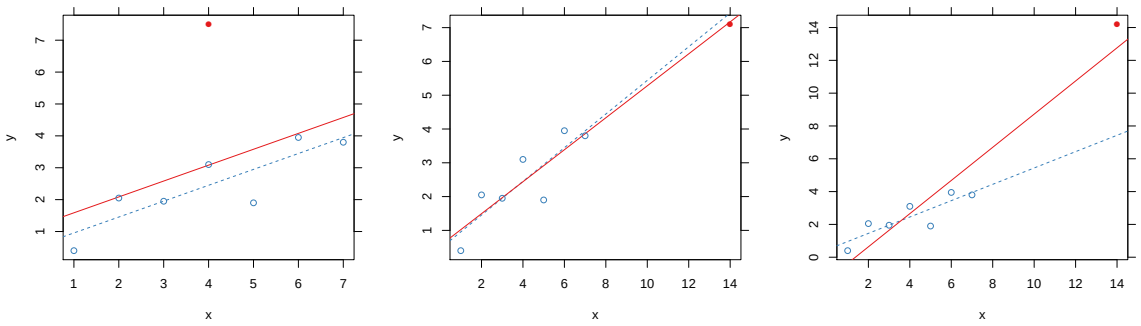




Use the plots to assess whether or not the four regression assumptions hold.

Outliers and Leverage Points

4. Each of the following scatterplots show two regression lines: the solid line is fitted to all of the points, and the dashed line is fitted to just the hollow points.



(a) (b) (c)

(a) In plot (c), including the solid point has a big effect on the regression fit. Why is this?

(b) Does a leverage point always have a big influence on the regression fit?

(c) Can a point that is not a leverage point have a big influence on the regression fit?

(d) In each of the above three cases, should we include the solid point in the regression analysis? If not, what should we do with the point?