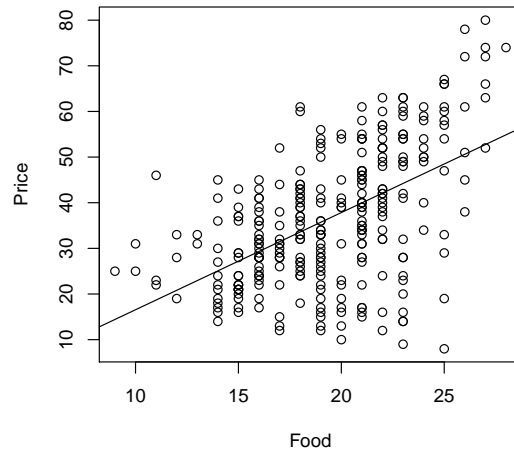


Regression Diagnostics

STAT-UB.0003: Regression and Forecasting Models

Linear regression model

1. Here is the least squares regression fit to the Zagat restaurant data:



Here is the Minitab output from the fit:

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.5559	27.93%	27.68%	26.86%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-4.74	3.95	-1.20	0.232	
Food	2.129	0.200	10.64	0.000	1.00

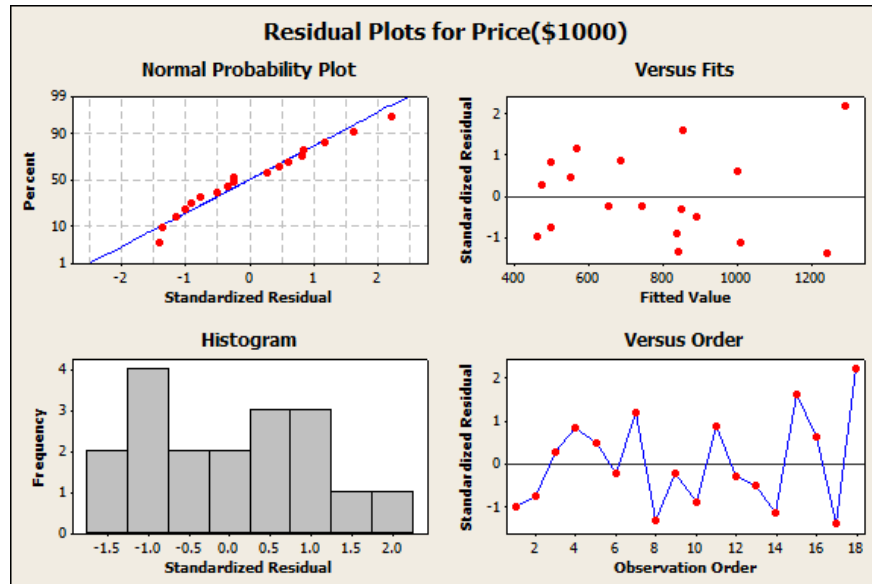
Regression Equation

Price = -4.74 + 2.129 Food

What proportion of the variability in the response is explained by the regression model (this is the “coefficient of determination”, commonly referred to as the R^2 value)? What is the meaning of this number?

Model assumptions

2. 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.

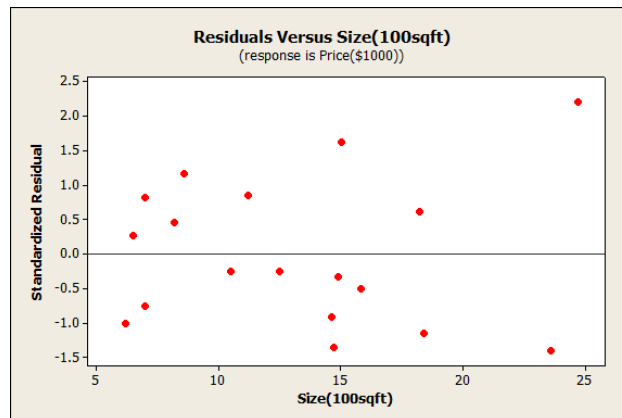
(a) Do the residual errors look approximately normal?

(b) Does the error variance look constant?

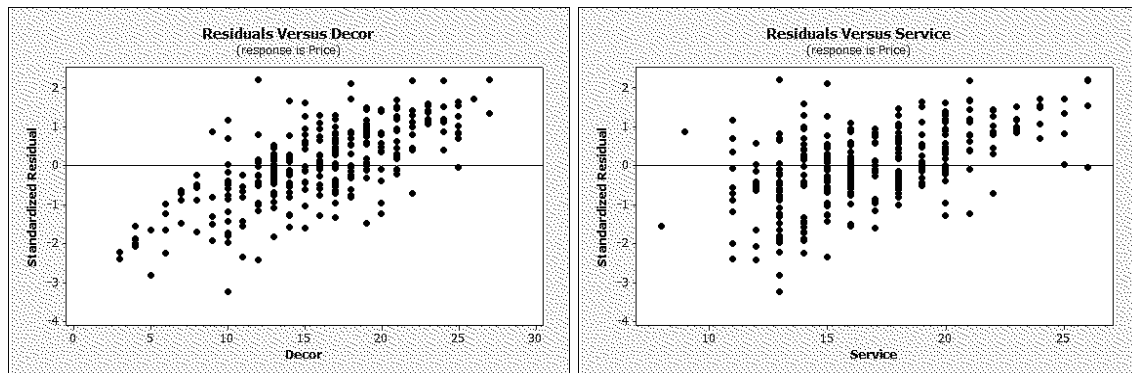
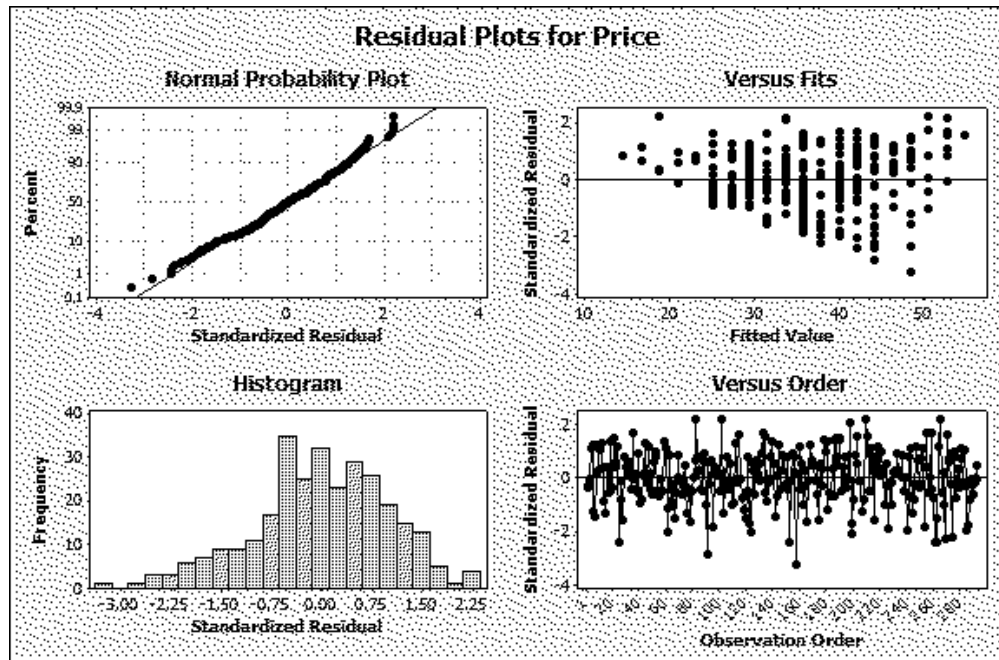
(c) Is there any apparent dependence in the residuals?

3. What is a “standardized” residual? Why is it sometimes easier to interpret a standardized residual than an ordinary residual?

4. Here is a plot of the residuals versus Size (x). Why is this plot nearly identical to the plot of residuals versus fits?



5. 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.