

Lab 8: Polynomial Regression, Residual Plots and Normality

Use the below data on cafeteria to answer the following questions.

Obs	Cafeteria	Dispensers	Sales
1	1	0	507.9
2	2	0	498.0
3	3	1	568.3
4	4	1	575.5
5	5	2	651.6
6	6	2	657.1
7	7	3	713.8
8	8	3	699.6
9	9	4	758.4
10	10	4	765.7
11	11	5	797.7
12	12	5	814.3
13	13	6	836.8
14	14	6	825.1

The variables in the data set are:

Cafeteria - cafeteria identification number

Dispensers - number of dispensers at each cafeteria

Sales - coffee sales in hundreds of gallons

1. Perform a simple linear regression with Sales as the response and Dispensers as the predictor variable. What do you see in the residual versus predicted plot? What would you do to fix this problem?
2. Perform a forward selection (by hand) using the AIC criteria (you will need to use the command `AIC(model)` to get the AIC values for each model). The “smallest” model should be the just the intercept. The “biggest” model should be Dispensers up to the power of 4 (be sure to follow model hierarchy). What was the best degree for the polynomial based on AIC?
3. Run the model you selected in #2 and look at the residual versus predicted plot. What do you see?
 - a. In this model, does the Q-Q plot provide evidence of normality?
 - b. Does the histogram of residuals look normally distributed?