STA302 Lec5101, Methods of Data Analysis 1
Module 3: Problem set

September 18, 2024

1 Basic learning objective practice

Problem.

Use the "nyc.txt" dataset from Sheather's textbook. Fit the model in R using:

$$Price \sim East + Food + Decor$$

Write a brief report using RMarkdown as follows:

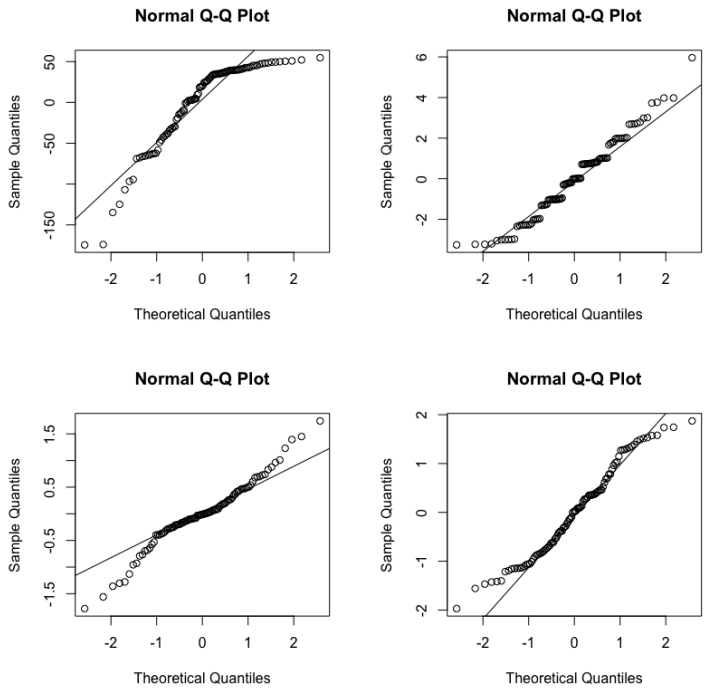
- Fit the model in R
- Check the regression modeling assumptions using plots.
- Briefly explain / interpret each plot.
- Interpret each coefficient in the context of the problem.
- According to this model, provide a recommendation to the new restaurant so they may charge a premium price for their new menu. Where might they want to allocate the most money: Location (East or West of Fifth Avenue), food, decor, or service.
- Compare the fitted coefficients with the following alternative model:

$$Price \sim East + Food + Decor + Service$$

Would you recommend the new restaurant focus on high-quality service to increase the price of their menu? Briefly explain why or why not.

Problem.

Describe what evidence these QQ plots provide for the standardized residuals.



Problem.

For each statement below, state whether it is a correct or incorrect description of a linear regression assumption. If it is incorrect, try to write it correctly.

- The linearity assumption means we assume that the true population response is linear in the coefficients.
- The constant variance assumption means we assume the variance of the of the responses given the predictor (s) does not change with the value of the predictor(s).

Problem.

A sample of 50 patients was randomly selected by a hospital to study whether a new medication can reduce cholesterol (Y) and thus the risk of heart disease. Each patient had their cholesterol measured at baseline and again every 2 weeks for 3 months (for a total of 8 cholesterol measurements per person). Descriptive plots of cholesterol show that the distribution appears to be highly left-skewed and that, when plotted against each possible predictor, all points appear to be equally spread out as the predictor increases.

Given the above setup, which of the assumptions of linear regression would we most likely be concerned could violated if a model was fit to these data, assuming our literature suggests we included all the right predictors in the correct way and have not omitted any important predictor information?

2 Basic learning objective practice from textbooks

Problem.

Sheather:

Chapter 3, Exercise 1(b), 4(a)

Chapter 6, Exercise 3(a), (b), 4(a) (b)

3 Advanced learning objective practice from textbooks

Problem.

Sheather:

Chapter 3, Exercise 2

Exercise 3 Part A (a) and c) only