# Homework 4

**Name: Caleb Thian Jia Le 田家樂**

**Student ID: NN6114035**

1. Using the ”Carseats” data set to answer the following questions:
   1. Fit a multiple regression model to predict Sales using Price, Urban, and US.
   2. Provide an interpretation of each coefficient in the model.
   3. Write out the model in equation form, being careful to handle the qualitative variables properly.
   4. For which of the predictors can you reject the null hypothesis ?
   5. Based on (d), fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.
   6. How well do the models in (a) and (e) fit the data? Give the reason.
   7. Try to fit a better regression model using more predictors in data set? What is the adjusted R2 ? The analysis should provide the diagnostic figures of residuals showing the model satisfies the assumptions.

**Ans:**

1. Implemented in main.r.

一張含有 桌 的圖片

自動產生的描述

1. The coefficients are

|  |  |
| --- | --- |
| Features | Coefficients |
| Price | -0.05 |
| UrbanYes | -0.02 |
| USYes | 1.20 |

The fitted coefficient of “Price” is -0.05, and it is significant. It means the values of Sales is affected by Price. The larger the different in Price, the more significance difference in Sales.

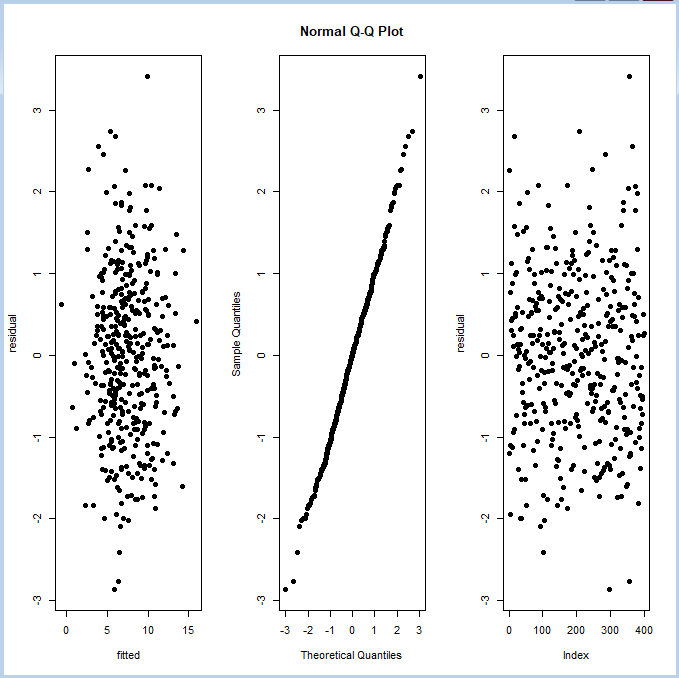
The fitted coefficient of “UrbanYes” is -0.02, but it is not significant, indicates that the values of Sales for car seats in urban and car seats in non-urban do not have significant difference. Specifically, assume that not in US, the intercept of car seats in urban is 13.04-0.02 = 13.02, and the intercept of car seats in non-urban is 13.04.

The fitted coefficient of “USYes” is 1.20, and it is significant, indicates the values of Sales for car seats in US and car seats in non-US have significant difference. Specifically, assume that not in urban, the intercept of car seats in US is 13.02+1.20 = 14.22, and the intercept of car seats in non-US is -493.73.

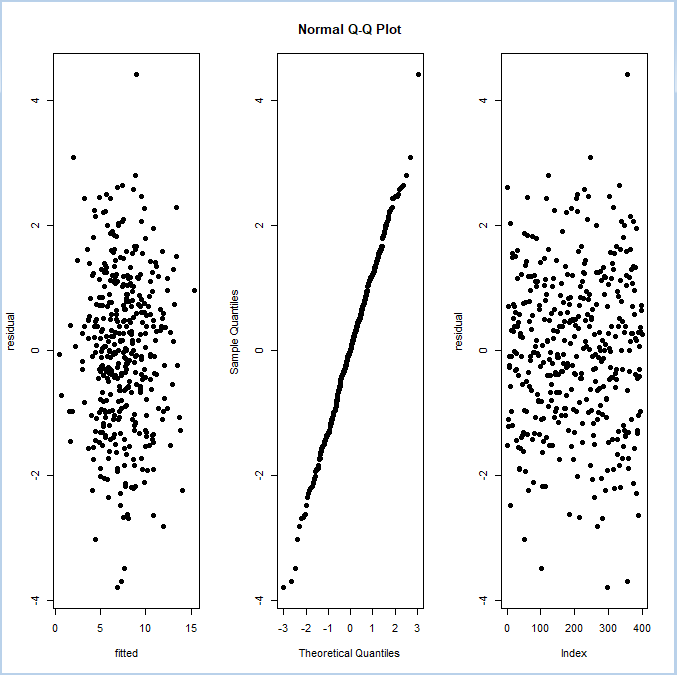
1. From summary at (a), Price and US are significant variables. Thus, for these predictors, null hypothesis can be rejected.
2. Fit a regression model is fitted by Price and US.

一張含有 文字 的圖片

自動產生的描述

1. The model in (e) is slightly better than the model in (a) as the adjusted R2 is 0.2354 and 0.2335 respectively, which is slightly higher. This is because the model in (e) uses less variables but only slightly reduce the value of R2 compare with the model in (a).
2. 一張含有 桌 的圖片

   自動產生的描述By fitting the regression model by all predictors, we can observe that CompPrice, Income, Advertising, Price, ShelveLoc, Age are significant variables.

Fit the model with significant variables.

一張含有 桌 的圖片

自動產生的描述

Both 2 models hold the assumption of multiple regressions, which are constant variance, normality and independent. Then, the best model is using all the predictors as it has a higher adjusted R2 = 0.8698 than the model only with significant variables which adjusted R2=0.8001.

1. Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Level (1 for College and 0 for High School), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get
   1. True or False
      1. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.
      2. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates.
      3. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.
      4. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough
      5. Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.
   2. Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.

**Ans:**

Assume

* + 1. Under the assumption, , let assume that GPA is under normal distribution, than on average, , thus, the statement is false.
    2. True, please refer to the explanation in i.
    3. Under the assumption, , provide GPA is high enough, than , the statement is true.
    4. False, please refer to the explanation in iii.

This indicates that the maximum of the interaction effect is 6.235, compare with the intercept term, , it is very little evidence. The statement is true.

The predicted salary is 137.1 thousands of dollars.