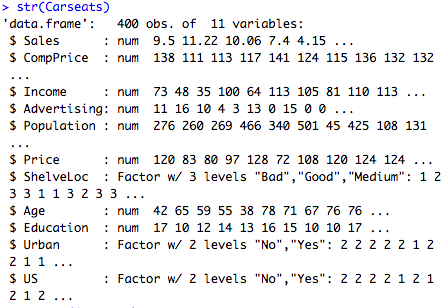
Assignment 2

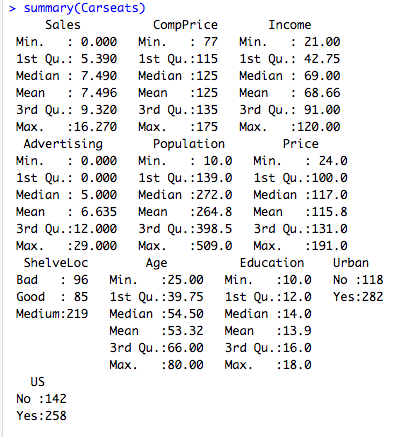
Due: Wednesday, September 25, 2019

In order to answer problems in Assignment 2, you need to use the ‘Carseat’ data, which is part of the ‘ISLR’ library. The goal of this assignment is to predict ‘Sales (child car seat sales)’ in 400 locations based on a number of predictors.

1. Which of the predictors are quantitative, and which are qualitative?

Hint: str() or summary ()



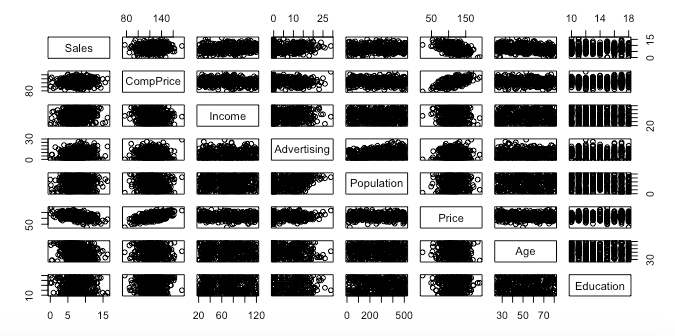


Quantitative are Sales, CompPrice, Income, Advertising, Population, Price, Age, Education

Qualitative are Shelveloc, Urban, US

1. Using Quantitative variables, describe the distributions in terms of shape, symmetry, and potential outlier. Do you think it is required to transform some variable(s)? If so, transform the variable(s) and justify your answer (Since the Advertising includes 1 missing value, please delete the Advertising variable when you compute correlation).

Hint: gpair(), cor()



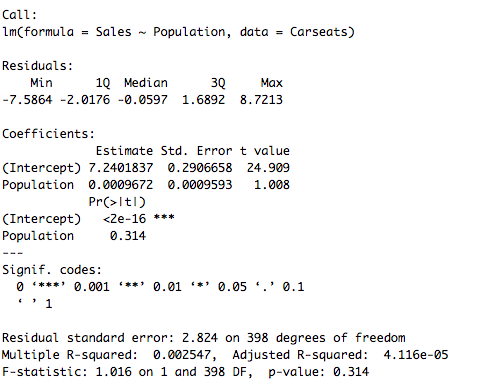
Advertising skew to right and curve linear shape on the pairwise correlation plot.

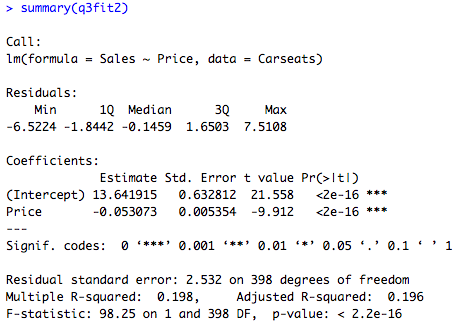
If we improve this variable firstly should consider the transformation logs.

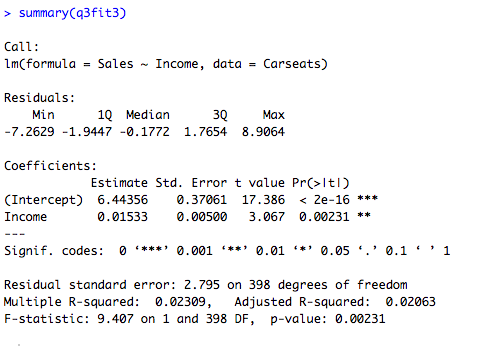
The advertising variable distribution more symmetric.

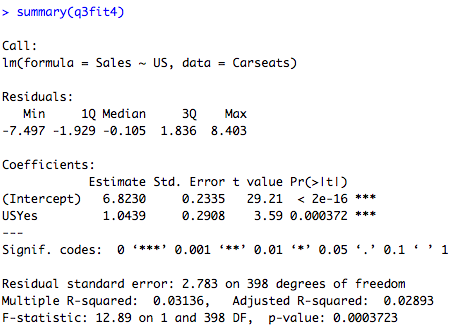
1. Fit four **separate** simple regression models to predict ‘Sales’ using ‘Income’, ‘Population’ and ‘Price’ and US. Then, **write out the estimated model in equation form**.

Hint: lm ()









Sale hat = 7.2401837+0.0009672 Population.

Eery additional increase 1 unit on Population.

Sales will increase 0.0009672, on average.

Sale hat = 13.641915 -0.053073Price.

Every additional decrease 1 unit on Price

Sales will decrease 0.053073, on average.

Sale hat = 6.44356+ 0.01533 Income.

Every additional increase 1 unit on Income.

Sales will increase 0.01533, on average.

Sale hat = 6.8230+ 1.0439 US.

Every additional increase 1 unit on US.

Sales will increase 1.0439, on average.

1. Provide an interpretation of coefficients in each separate model. Be careful-some of the variables in the model are qualitative!

Population: Every additional increase 1 unit on Population

Sales will increase 0.0009672, on average.

This is not expected to have a statistically significant impact.

Price: Every additional increase 1 unit on Price.

Sales will decrease 0.053073, on average.

This value cannot be predicted as a statistically significant effect.

Income: Every additional increase 1 unit on Income.

Sales will increase 0.01533, on average.

This value cannot be predicted as a statistically significant effect.

US Yes: Every additional increase 1 unit on US.

Sales will increase 1.0439, on average.

This is expected to be important and represents a sales unit.

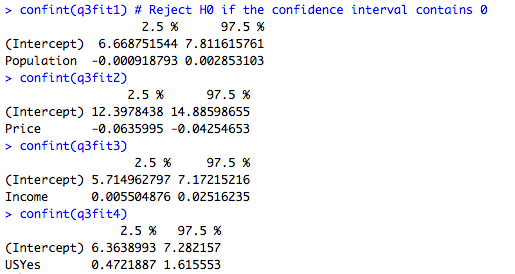
1. For which of the predictors can you reject the null hypothesis ?

In the United States and in terms of income, we can reject the zero hypothesis.

1. Using the models Question 3, obtain 95% confidence intervals for the coefficient(s). Using the confidence intervals, test the null hypothesis .

Hint: confint()

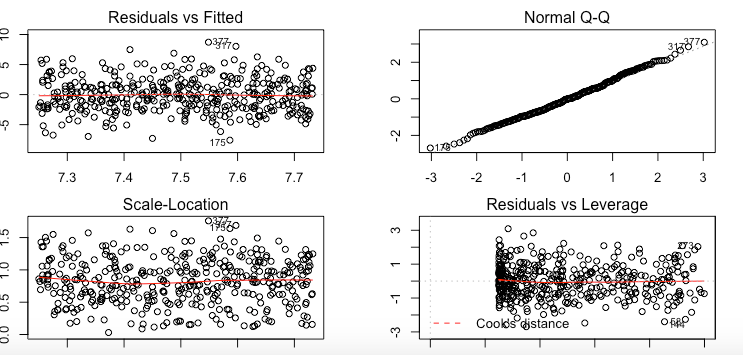
Coindence interral (CI) if “0” is NOT in the CI 🡪reject H0

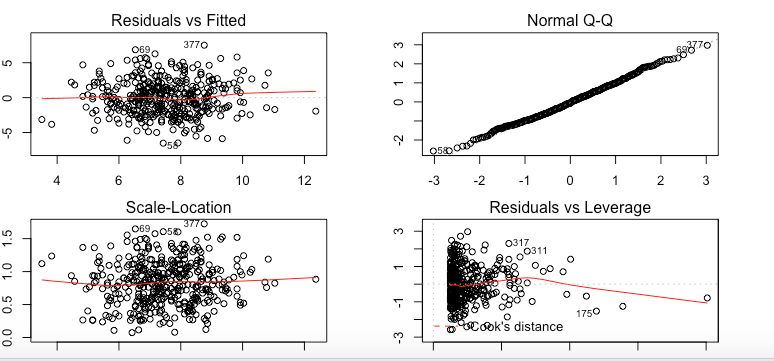


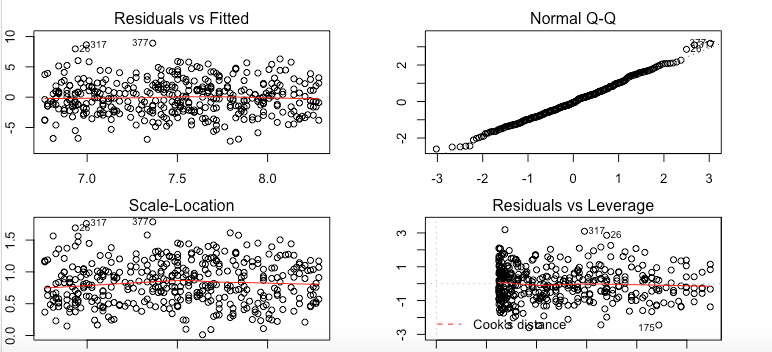
* In Price (-0.0635995, -0.04252653), in Income (0.005504876, 0.02526235) and USYes (0.4721887, 1.615553) are not overlap with 0, so they reject null hypothesis.
* In Population CI contain 0 is fail to reject H0

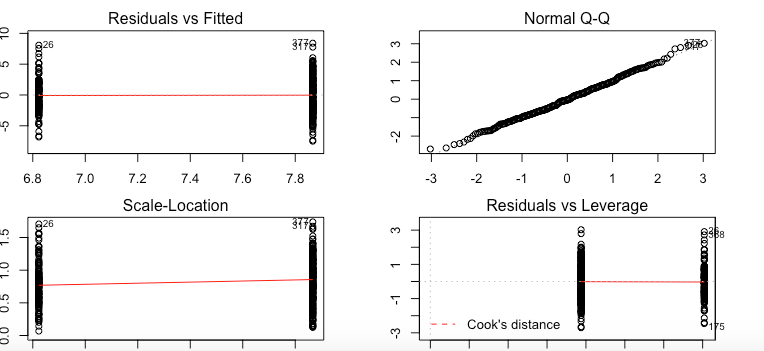
1. Check the assumptions of the models using plot(). Is there evidence of outliers in the models? If so, please inspect the outliers.

Hint: plot()









These points do not seem to be particularly leveraged.

There are some outliers, but not many.

Normal Q-Q - Linear

Residual Plot - Random