**Problem 3 - (10 points)**

Using the “Cereals” data set, we are interested in predicting nutritional ratings based on shelf.

Use the following SAS code snippet to create the ‘Shelf2’ indicator.

data cereal\_dsb;

set cereal\_ds;

if shelf=2 then shelf2=1;

else shelf2=0;

run;

1. Construct the appropriate scatter plot.

Ans: **In the respective pdf**

1. Perform the appropriate regression analysis.

Ans**: In the respective PDF**

1. Interpret the results.

Ans: **Even though we have a few outliers, Residual graph is almost following the straight-line pattern, which is well to do. The variable "shelf2" and F Value are significant. Unfortunately, we have a small R-Square value. In sum, even though this is not a great model, overall, it's doing a good job, but it has a lot of space for improvement. Additionally, according to our univariate analysis, we have several influential and high leverage data points, and since they have huge impact on our model, we need to check whether they are collected correctly.**

1. Does cereal location (e.g. Shelf2) cause a change in the cereal rating?

Ans: **There is a statistical relationship between the two but since the information is not enough, we don’t know the extent of the relation.**