*Name: Hagar Haytham*

*Sec: 2*

*B.N: 28*

***Data Mining, Big Data and Analytics***

***Lab 6 – Logistic Regression***

**1.Write the variable pairs that are not correlated at all to each other.**

(Price, Income)

(Price, Age)

**2.Are there any highly correlated variables in this dataset?**

(Income, Age)

**3.How many categories are there for the Price variable?**

3 categories

**4.Why is it divided into two entries only in the model?**

Cause in categorical variables w represent them by n-1 Variables if we have n categories. In our case we have 3 categories so we can represent them with 2 variables. If the example falls in one of the two variables then this variable is set to one, if both variables are set to 0 that means this example falls in the third category.

**5.1. Write the value of AUC.**

0.915272

**2. What is the maximum value of AUC (ideal case)?**

1

**Note: For this part, you may need to search and read about the ROC curve.**

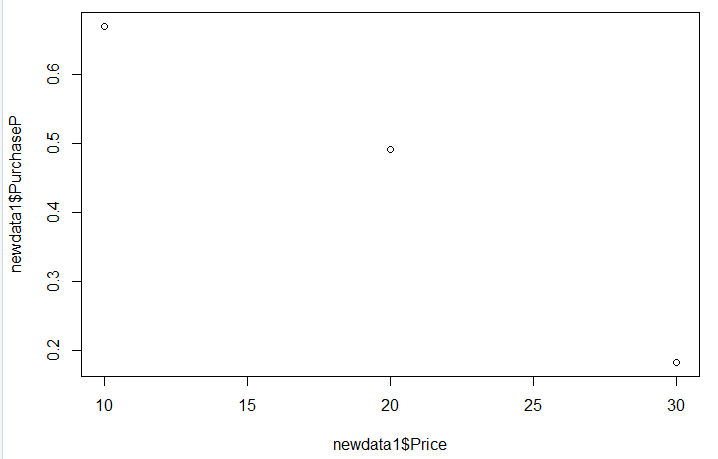
**6.What does each point in the ROC graph represent? In other words, what is the value that changes and drives TPR and FPR to change too from one point to another in the graph?**

These represent a number of confusion matrices each attached to a different threshold for the logistic regression

**7.How is the predicted probability affected by changing only Price holding all other variables constant?**

The more the price increased the more the probability of purchase decreases.

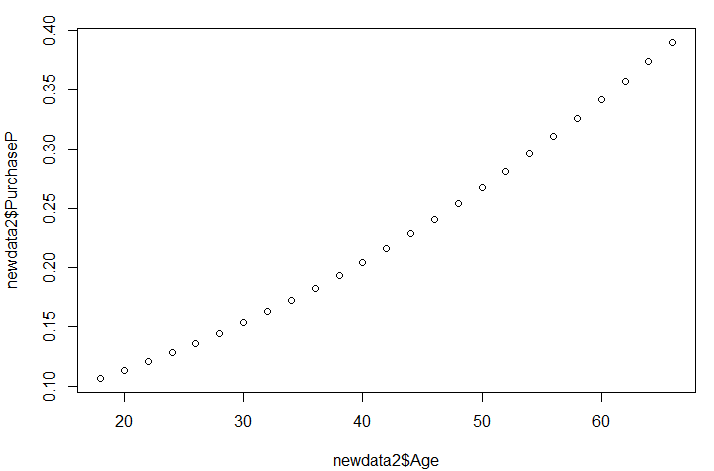
(negatively correlated)



**8.How is the predicted probability affected by changing only Age holding all other variables constant?**

The more the age increases the more the probability of purchasing increases.

(positively correlated)



**9.How is the predicted probability affected by changing only Income holding all other variables constant?**

The more the age increases the more the probability of purchasing increases.

(positively correlated)

