

Q1) Write the variable pairs that are not correlated at all to each other.

A1) (Price, Income), (Price, Age).

Q2) Are there any highly correlated variables in this dataset?

A2) No, (Age, Income) are correlated but not highly correlated.

Q3) How many categories are there for the Price variable?

A3) There are 3 Categories (10, 20, 30) → `factor(Mydata$Price)`

Q4): Why is it divided into two entries only in the model?

A4) Because it is a categorical variable with (n=3) values. So we use (n-1) 2 dummy variables instead.

Q5.1) Write the value of AUC (just the number)

A5.1) 0.915272

Q5.2) What is the maximum value of AUC (ideal case)?

A5.2) 1.0

Q6) 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?

A6)) threshold, ROC curve plots (fpr, tpr) as the threshold is varied from 0 (the upper right hand corner) to 1 (the lower left hand corner). As the threshold is raised, the false positive rate decreases, but the true positive rate decreases, too. The curve helps us to choose a threshold so the tpr is as high as possible and fpr is as low as possible.

Q7) How is the predicted probability affected by changing only Price holding all other variables constant?

A7) The probability decreases as the price goes up which makes sense the more expensive the item is the less likely we buy it.

Q8) How is the predicted probability affected by changing only Age holding all other variables constant?

A8) The probability increases as the age goes up. The older we get we become more capable to buy the item I think.

Q9) How is the predicted probability affected by changing only Income holding all other variables constant?

A9) The probability increases as the income goes up. The more money we have the more likely we buy the item.