1. In your own words, discuss (in less than one page) the differences between Multiple Regression Analysis and Multiple Discriminant Analysis.

Multiple regression describes predicting a metric dependent variable (or variables) from independent variables. Multiple Discriminant Analysis is quite similar, but instead of predicting a metric variable, it predicts a nonmetric or categorical variable. In other words, we are performing a classification rather than a regression task.

2. For the data set associated with this homework (HBAT and HBAT_Test. Using X4 as the non-metric variable and (X6 up to X18) as the metric variables:

a. What does each variable represent? (go back to Week # 2)
X4 is the customer's region, X6-X18 are customer perceptions.
X6: Product quality
X7: E-commerce activities/website
X8: Technical support
X9: Complaint resolution
X10: Advertising
X11: Product line
X12: Salesforce image
X13: Competitive pricing
X14: Warranty and claims
X15: New products
X16: Ordering and billing
X17: Price flexibility
X18: Delivery speed

b. How many groups does X4 have?
2
c. Apply linear discriminant analysis to the data (HBAT) and find:

• The linear discriminant function for X4.
Z =-0.67548538X6 - 3.00383934X7 - 0.0348183X8 + 0.02132907X9 - 0.41070965X10
-0.33681933X11 + 3.00340929X12 + 0.58597043X13-0.161654X14 - 0.28139853X15
0.59308282X16 + 2.08495472X17 - 2.17076208X18
• By applying the LDF to the training data (HBAT): How many observations were misclassified? What are they? Find the confusion matrix and the probability of (error) Misclassification.

39	0
5	56

Prob misclassification = 0.05

• By applying the LDF to the test data (HBAT_Test): How many observations were misclassified? What are they? Find the confusion matrix and the probability of (error) misclassification.

37	2
11	50

Prob misclassification = 0.13