ISL Fall 2015 Assignment III. 100 pts.

NAME(s):

**LOKESHWAR REDDY INJA (16207239/ LIRK9)**

**SANTHOSH KUMAR GATTU (16211118/ SG6N6)**

**SANTOSH KUMAR (16207971/ SK7Z9)**

You may submit this assignment in groups of upto three each. Write your names on this sheet and include it as the cover page for your submission.

The objective of this assignment is to practice using R, and gain a fundamental understanding of linear regression. Your submission should include both your code as well your answers to the questions.

Electronic submission on Blackboard is due **latest by 11 pm on Wed, Oct 7th**. You may upload upto **three** submissions **before** the deadline – only the last submission will be graded. Submissions received after the deadline will be graded only for effort for a maximum of 70% of the total grade (Refer to class syllabus for detailed grading policy).

**State any assumptions you make, justify your answers, show intermediate steps and explain your results for maximum credit**. All answers should be in your own words with any sources you refer to cited at the appropriate places. Any knowledge you acquire from the Internet should be written in your own words and be appropriately referenced. Copying and pasting from the Internet, each other or any other source will not count as your effort (Refer to class syllabus for detailed policy on plagiarism).

**Remember that answers need to be word-processed (NOT handwritten) and should use R.**

Answer the following questions from Chapter 4.

Q6

Q7

Q11

Q13

6 .a)

We have the logistic function for two predictors and a dependent variable as,

Probability(X) = (e^B0+B1X1+B2X2) / (1 + e^B0+B1X1+X2)

Here we have B0=-6; B1=0.05; B2=1

Substituting the given values in the above equation:

P = e^ (-6 + 0.05 \* 40 + 1 \* 3.5) / (1 + e ^ (-6 + 0.05 \* 40 + 1 \* 3.5)

P = e ^ (-0.5) / (1+ e ^ (-0.5))

P = 1 / (1 + e ^ 0.5)

P = 1 / 2.65

P= 0.3776

Hence, Probability that a student who studies for 40 hours and has GPA of 3.5 is 0.3776.

6. b)

Here probability of getting A in the class is 0.5

We have the logistic function for two predictors and a dependent variable as,

B0+B1X1+B2X2 = log (P(A) / 1-P(A) )

Substituting the given values in the above equation

-6 + 0.05 \* X1 + 1 \* 3.5 = Log ( 0.5 / (1-0.5) )

-6 + 0.05 \* X1 + 1 \* 3.5 = Log (1)

-6 + 0.05 \* X1 + 1 \* 3.5 = 0

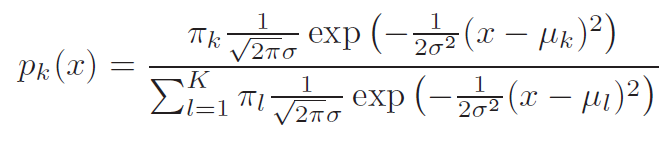
0.05 \* X1 = 6-3.5

0.05 \* X1 = 2.5

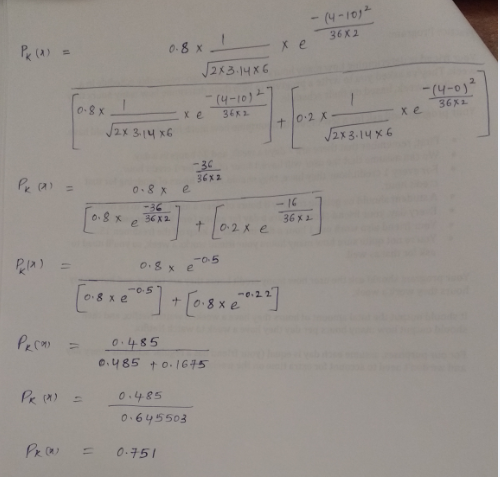
X1 = 2.5 / 0.05 = 50 hours

7)

From Linear Discriminate Analysis we have,



Substitute X=4, 𝜇𝑘 = 10, 𝜇𝑙 = 0, 𝜎 2 = 36, 𝜋𝑘 = 0.8, 𝜋𝑙 = 0.2 in the above equation



Probability that X will issue a dividend is 0.751

11(a)

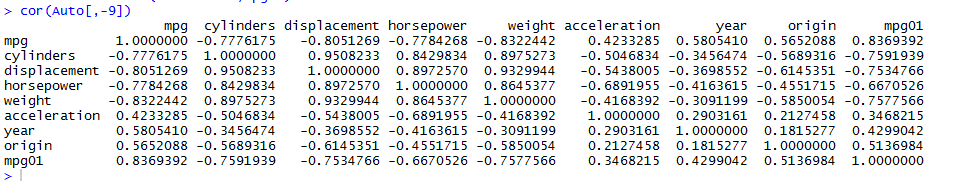
Solution

Please find below the R file

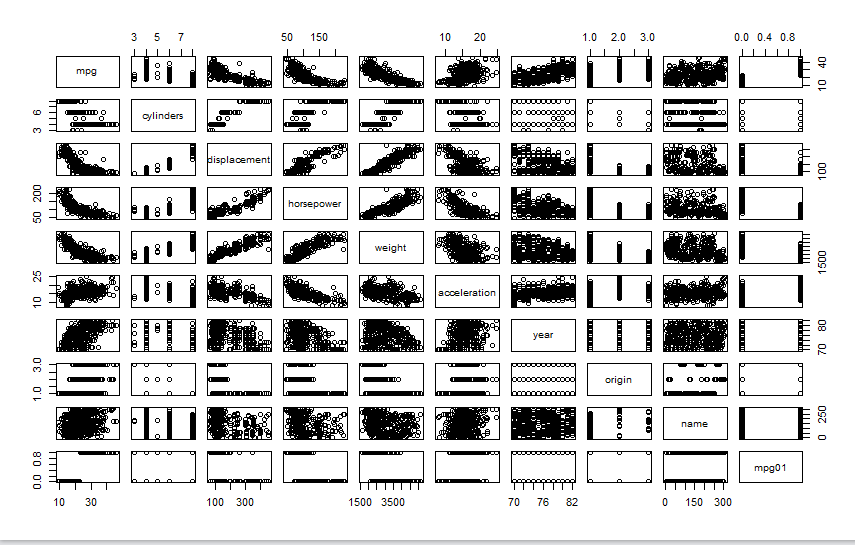


11(b)

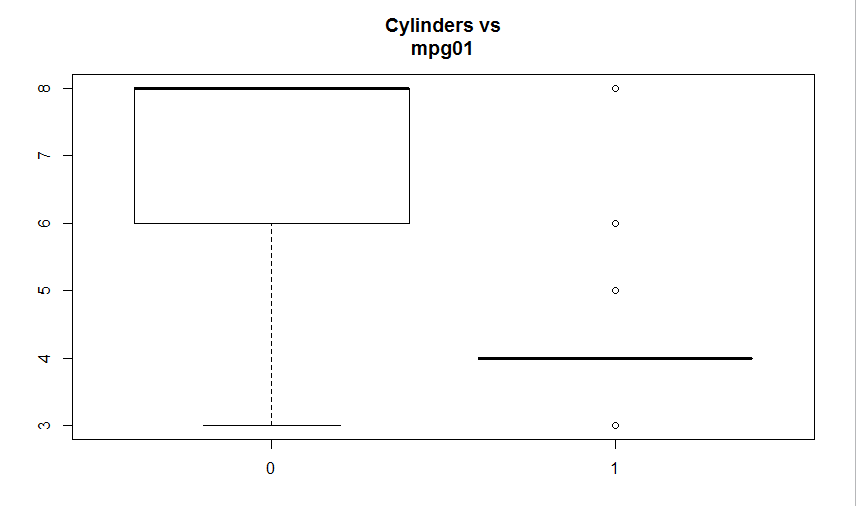




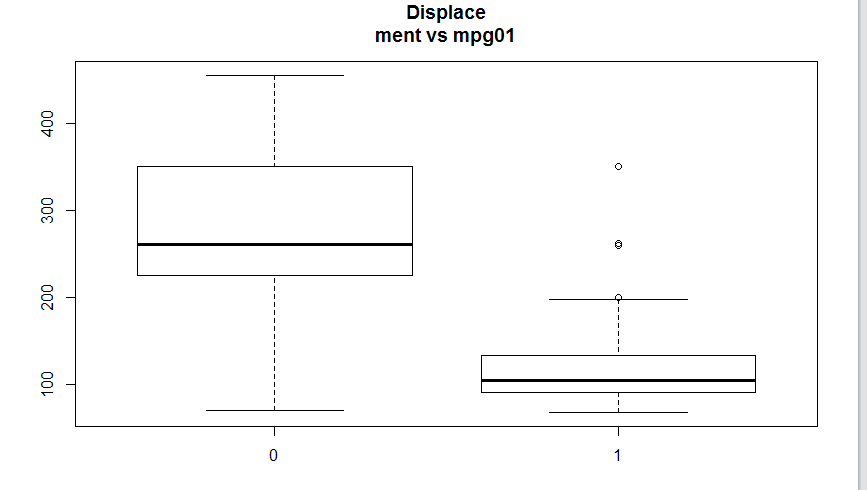
> pairs(Auto)



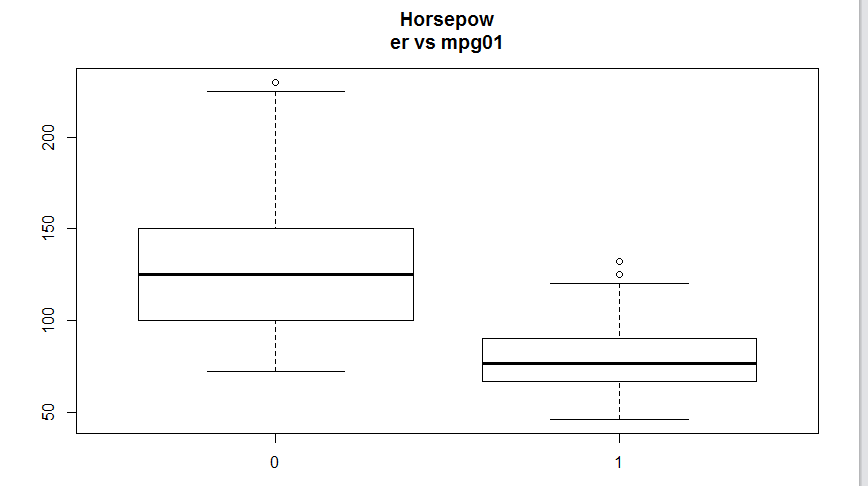
> boxplot(cylinders~mpg01,data=Auto,main="Cylinders vs mpg01")



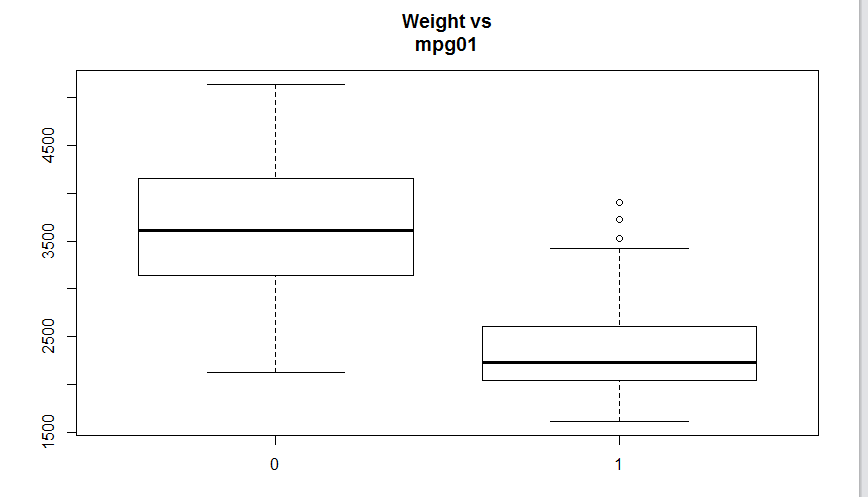
> boxplot(displacement~mpg01,data=Auto,main="Displacement vs mpg01")



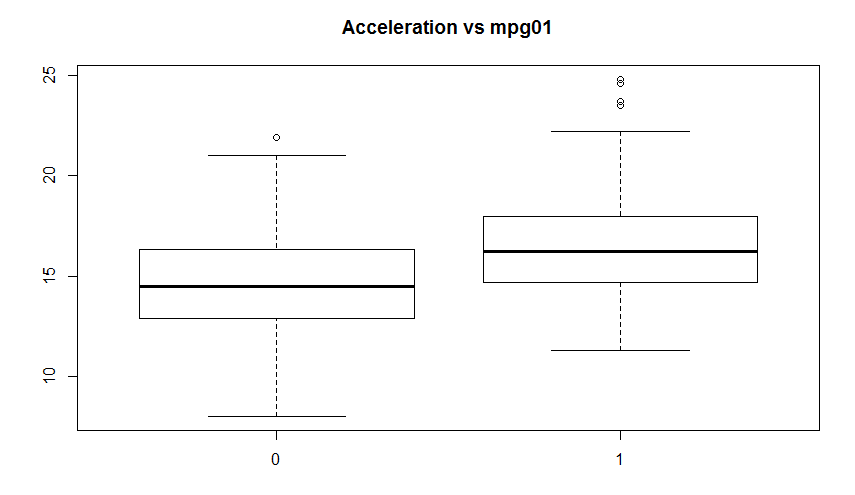
> boxplot(horsepower~mpg01,data=Auto,main="Horsepower vs mpg01")



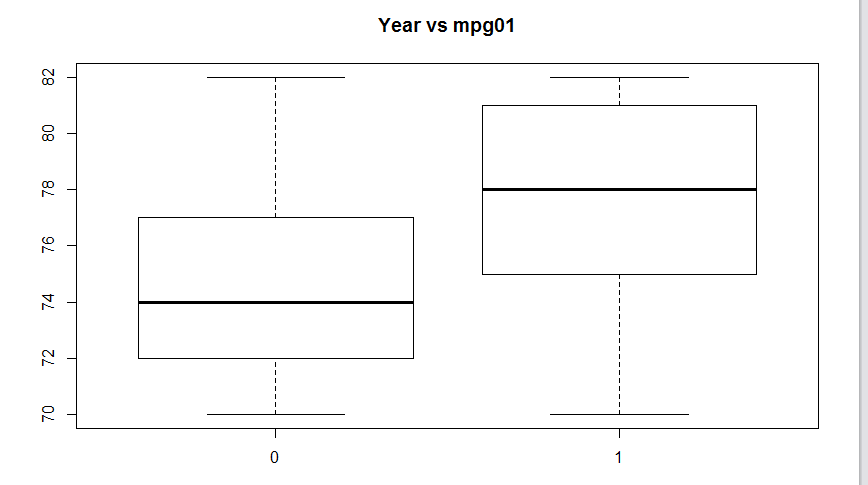
> boxplot(weight~mpg01,data=Auto,main="Weight vs mpg01")



> boxplot(acceleration~mpg01,data=Auto,main="Acceleration vs mpg01")



> boxplot(year~mpg01,data=Auto,main="Year vs mpg01")



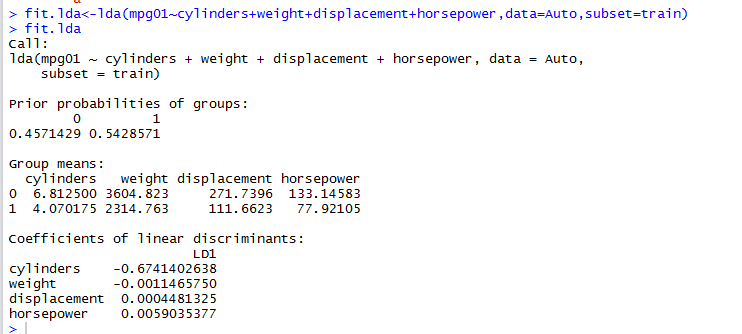
There is a relationship between “mpg01” and “cylinders”, “weight”, “displacement” and “horsepower”

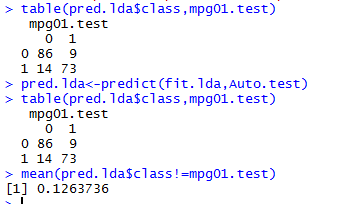
11(C)



11(D)



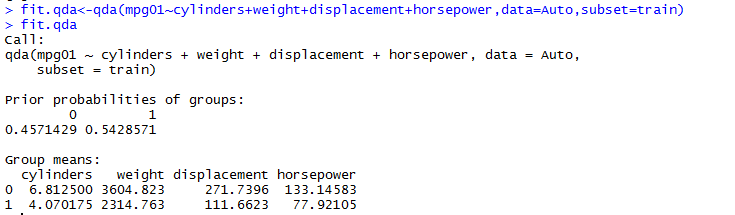


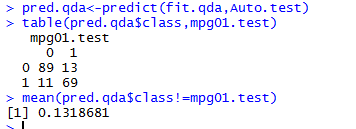


The test error rate is 12.63%

11 (E)



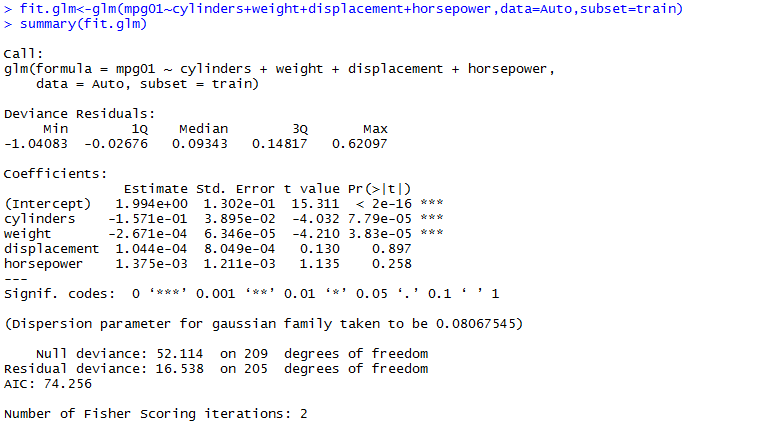


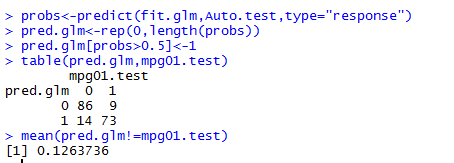


The test error rate is 13.18%

11(F)







The test error rate is 12.63%

11(G)



For k=1, the test error rate is 15.38%

For k=10, the test error rate is 16.48%

For k=100, the test error rate is 14.28%. So for k=100, it seems to perform the best test.

13)

Solution



The test error rate for logistic regression is 18.18%

The test error rate for logistic regression is 15.81%

The test error rate for lda is 13.43%

The test error rate for this lda is 15.01%

For KNN (k=1), the test error rate is 45.84%

For KNN (k=10), the test error rate is 11.85%

For KNN (k=100), the test error rate is 49.01%