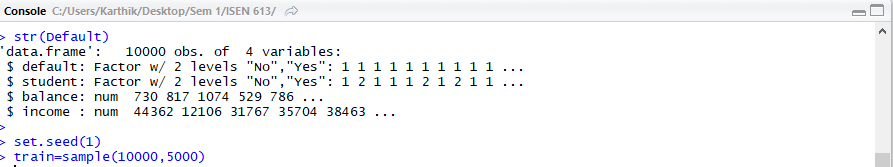
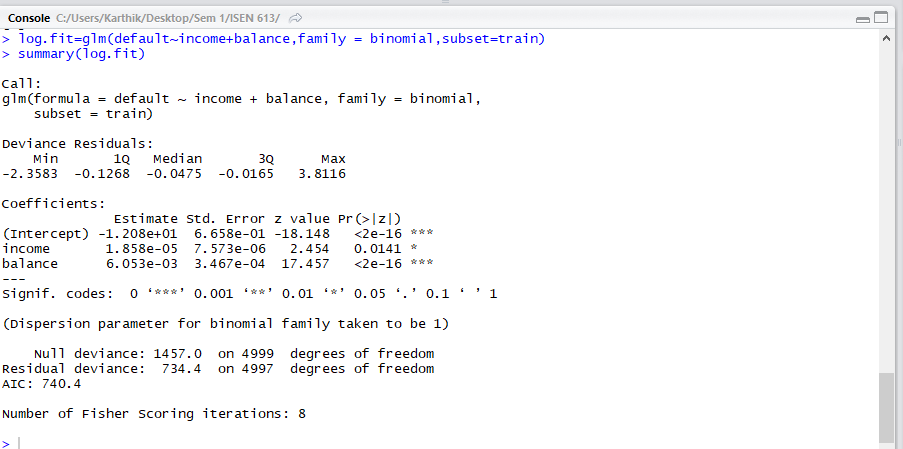




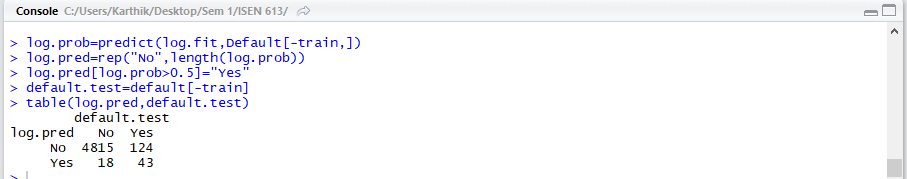
1)



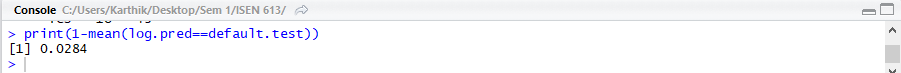
2)



3)

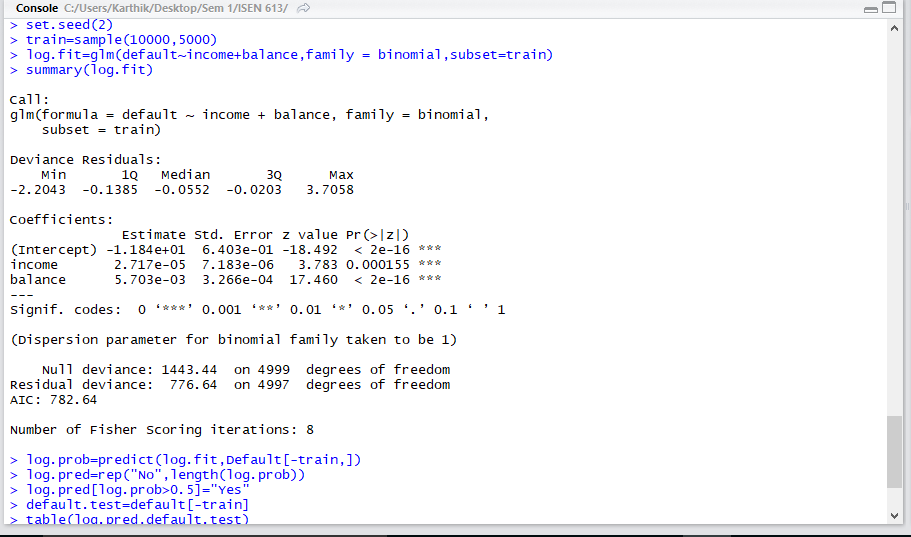


4)



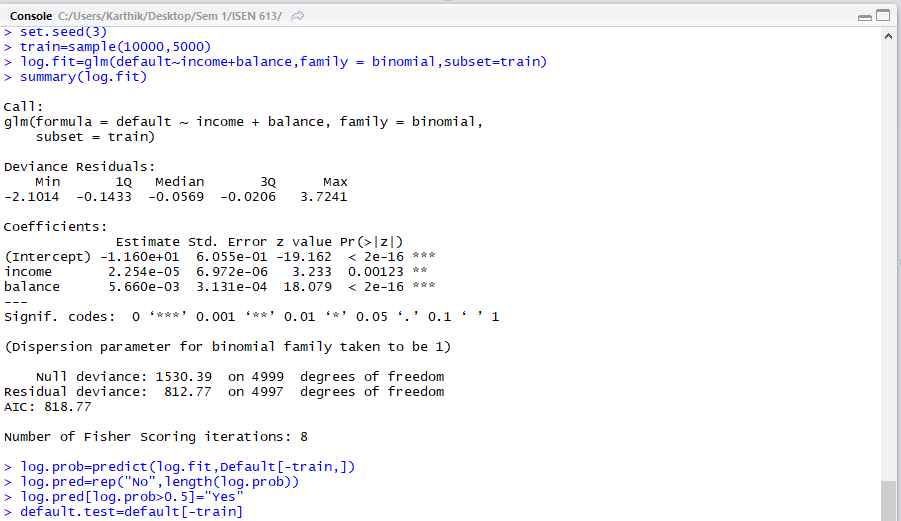
Validation set Error: 2.84%





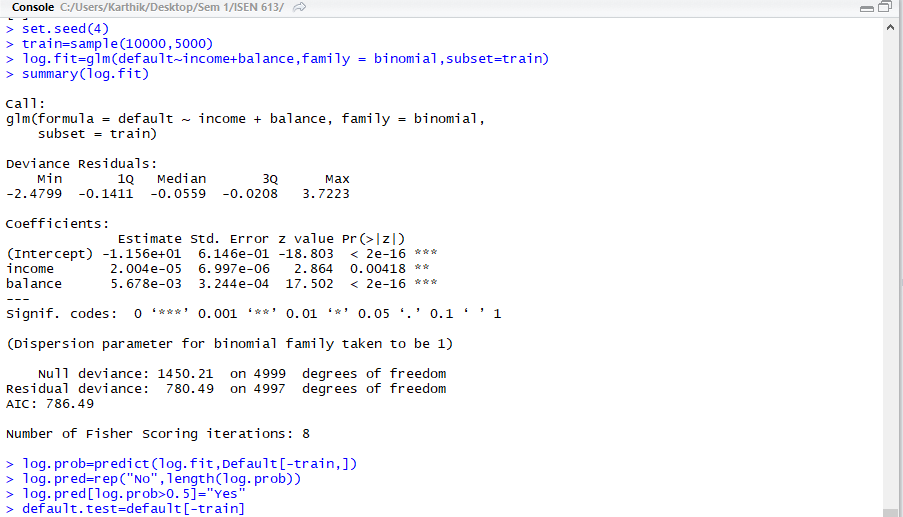


Validation error: 2.78%





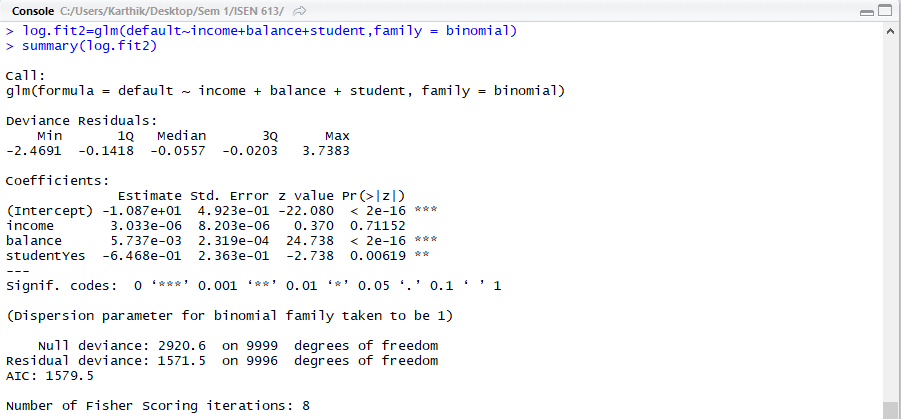
Validation error: 2.54%

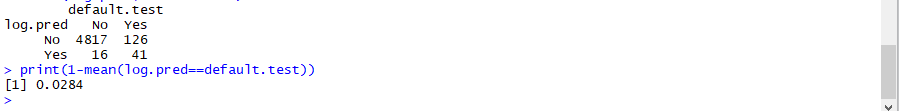
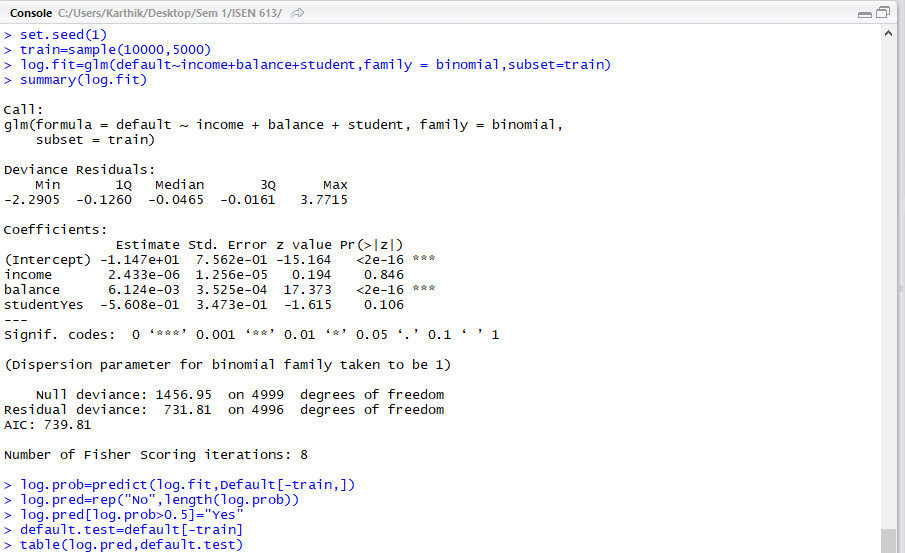




Validation error: 2.76%

Each of the 3 times, the results are different as the seed or the randomness of the split of the data for validating the data is different each time.

1. 



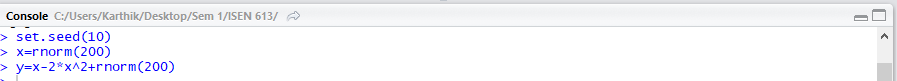
Validation error:2.84%

Comparing the results of seed(1) of the logistic model, before and after

adding the ‘student’ variable, the test error rate does not change much.

2)

a)

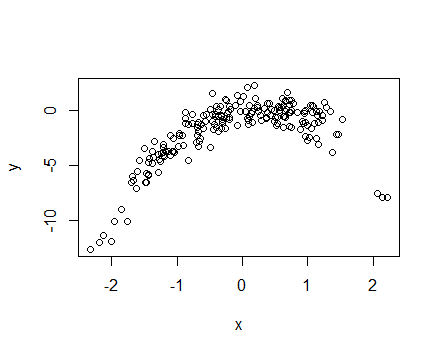


N=200, p=1

True model: Y= X-2\*X^2 + E

E:error term

b)

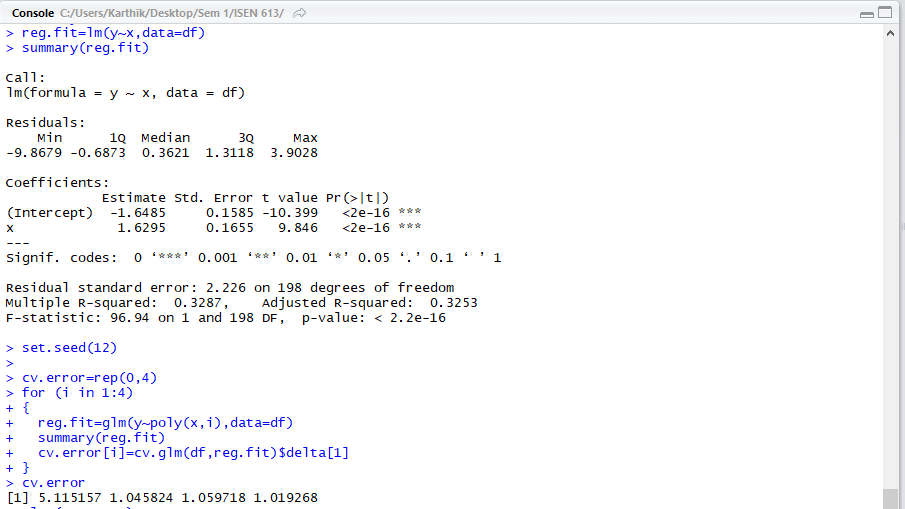


Y seems to have a quadratic relationship with a variation around the fitted curve with an error (which belongs to the one generated from rnorm function.) Also the x axis has a value that has a mean of 0 and +/-3 standard

deviations with many corresponding y values towards the centre as many x values will be generated as near to 0

as possible and lesser y values as one moves away from the mean 0.

c)



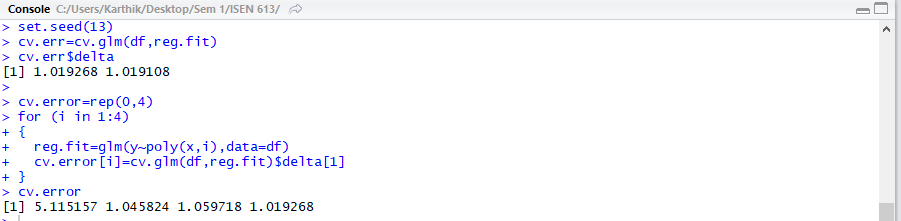
Error for order 1: 5.1151

Error for order 2: 1.0458

Error for order 3: 1.0597

Error for order 4: 1.0193

d)



Error for order 1: 5.1151

Error for order 2: 1.0458

Error for order 3: 1.0597

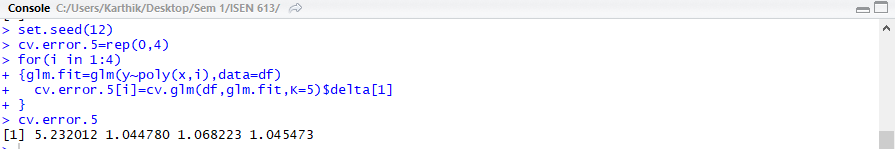
Error for order 4: 1.0193

The results obtained in LOOCV using a different seed is same as the previous result with a different seed as there is no randomness as the data is split into ‘n-1’ and ‘1’ and all possibilities are covered any way the data is split.

e)

Model 2 ( 𝑌 = 𝛽0 +𝛽1𝑋 +𝛽2𝑋2 +𝜖 ) has the smallest test error which is expected as the relation between y and x is quadratic.

f)



Error for order 1: 5.232

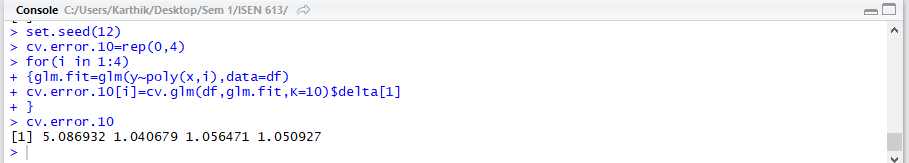
Error for order 2: 1.044

Error for order 3: 1.0682

Error for order 4: 1.045

Model 2 ( 𝑌 = 𝛽0 +𝛽1𝑋 +𝛽2𝑋2 +𝜖 ) has the smallest test error. It is consistent with the LOOCV method, in the sense that even in LOOCV model, Model 2 had the least test error, but the individual test error values, vary a bit for the corresponding models in LOOCV and K-Fold method.

g)



Error for order 1: 5.0869

Error for order 2: 1.0407

Error for order 3: 1.0564

Error for order 4: 1.0509

The individual test errors are different for 5 Fold and 10 Fold Methods, but the conclusion that Model 2 ( 𝑌 = 𝛽0 +𝛽1𝑋 +𝛽2𝑋2 +𝜖 ) has the smallest test error is common and same for both methods.