INSTRUCTIONS

Create a new R Script named Assignment7\_FirstnameLastname eg Assignment7\_SusanConnolly .

When you have completed the tasks, save your R Script and upload a ZIPPED VERSION OF THE SCRIPT to Blackboard in Assignments: Lab7. Make sure to save regularly during the lab so as not to lose your work.

Assignments submitted more than two hours after the scheduled end of the lab will not be graded and will receive 0 marks.

You may make multiple attempts (uploads), however note that only the latest attempt will be graded.

For the avoidance of doubt, if multiple attempts are made, and some are uploaded after the deadline, the latest attempt before the deadline will be the only attempt graded.

In a single file, write code to conduct the following tasks and comment the code clearly. Especially comment where the different parts begin in your program

TASKS

A bank has asked you to help them with finding a way to decide whether an applicant for a loan is likely to default.

**Part 1**

Read in the file credit.csv that contains bank records detailing whether or not people paid back their loans (Loan\_Status 1= paid back, 0=defaulted), an id (Loan\_ID) and 11 potential explanatory values. Split your data into training (80%) and test (20%) subsets. Fit a logistic regression model to the test data. (Use all variables in the dataset except for the ID)

**Part 2**

Plot a ROC curve for your model to the test data and choose a threshold to classify an applicant as “repay” or “default”. Explain in comments why you chose the threshold you did.

**Part 3**

Apply your model to the test data and report on its performance in the comments. You can choose the performance measures, but explain in comments why you chose the measures you did. Compare the performance in the test data to that in the training data and explain any differences.

Some hints:

* Make sure that categorical variables are correctly encoded as factors
* Remove any observations with missing data