Use the data in ‘spider_train.csv.’

In this data, the descriptor/predictor is ‘size.’ It represents the size of sand grains on the sea beach. The outcome is ‘spider.’ It represents whether a particular type of spider is detected (marked as 1) or not (marked as 0).

Create a logistic regression model for this data.

Plot the data (size vs. spider) as a scatter plot and overlay the plot with the curve generated by the logistic regression model.

We need to decide suitable probability threshold/cut-off for the classification. Use three different cut-offs P = 0.3, 0.5, 0.7. Show how the sensitivity of the model changes with the cut-off. Use the test dataset (‘spider_test.csv’) to create the confusion matrices and calculate the sensitivities.

Submit:

a) The R script for this problem.

b) A report with the following:

1. Name and roll number.

2. The scatter plot showing the data overlaid with the logistic regression curve. Label the axes correctly.

3. Confusion matrices for all three cut-offs.

4. A bar plot showing the effect of the cut-off P on the sensitivity. Label the axes correctly.