

STA 303H1S / STA 1002HS: Logistic Regression Practice Problems 3

1. Logistic regression is one possible method to find a combination of explanatory variables to best classify observations into two groups. An observation is classified as having estimated response of 1 if the estimated probability of 1 from the logistic regression model is greater than a cut-off probability; else it is classified as having an estimated response of 0. The R output display the overall percent correct, the false positive rate, and the false negative rate.

On the practice problems website, output including Classification Tables is given for both the Donner Party and Krunnit Islands examples.

- (a) For each example, choose the optimal cut-off probability considering the percentage correct and the percentages of false positives and false negatives.
- (b) Without any additional information, you might guess that 0.5 is a reasonable choice for the cut-off probability. Are your answers to part (a) close to 0.5? Why or why not?
- (c) Why should you not trust the percentage of correct classifications?