STA 303H1S / STA 1002HS: Logistic Regression 2 Practice Problems
SOLUTIONS

1. (a) For the Donner Party example, the cut-off probability of 0.50 correctly classifies the status of 78% of people. Associated with this cut-offs are fairly low false positive (8%) and false negative (40%) rates, so 0.50 is a reasonable choice for a cut-off probability balancing the criteria equally. (.44 would be even better)

   For the Krunnit Islands example, the cut-off probability of 0.14 classifies 37% of species correctly, and a cutoff of 0.08 classifies 27% currectly. The cut-off probability of 0.22 classifies all of the non-extinct species correctly and all of the extinct species incorrectly.

   (b) 0.50 is a reasonable cut-off probability for the Donner Party example but not for the Krunnit Islands example. This is because we have observed close to an equal number of events/non-events for the Donner Party example but not for the Krunnit Islands example. Since we observed fewer non-events for the Krunnit Islands example, we achieve better classification rates by choosing a cut-off probability that is larger.

   (c) These classification rates are calculated from the data that were used to fit the model and thus will be overly optimistic. It would be better to fit the model on a training data set, and estimate the classification rates from a test data set.