# Math 70 Homework 8

## Alex Craig

## Problem 1

### Instructions

The file `amazshop.csv` contains the following information on 1000 Amazon.com shoppers: `shop=1` indicates the internal definition of active shoppers and `0` otherwise, `age`, `sex=1` codes male and `0` female, `total` contains the dollar amount spent during the year, and `npurch` indicates the number of items sold.

**(a)** Run logistic regression on the four variables and develop a parsimonious model by selecting only statistically significant variables.

**(b)** Test the hypothesis on the validity of the logistic regression using the likelihood ratio test by testing that all slope coefficients are zero.

**(c)** Plot the probability to be qualified as an active shopper versus total amount for the shopper of age 20 and age 60.

**(d)** Compute and display the 95% confidence interval for the probability of a person who spends $2000 per year and is 60 years old.

**(e)** Compute and display the ROC curves along with AUCs for identification of an inactive shopper for full and parsimonious model using different color.

### Solution

## Problem 2

### Instructions

**(a)** Apply PCA to project the iris data onto the plane, display and color each flower.

**(b)** Compute and display on each axis the proportion of variance explained and two components together on the top of the graph. Use `legend` for flower colors.

**(c)** Display the proportion of variance explained by PCA projections as in `swiss(job=3)`.

### Solution