All Data files are on the class web page – Note we will be using this data set for the next assignment as well.

1. **Mass Transit:** Through a survey of approx. 600 consumers, the Mass Transit Authority has collected information on consumers' preference for use of Public Transportation as a function of gasoline prices. In addition the Mass Transit Authority collected attitudinal/demographic data on these same consumers.

    There are two data files: (a) *calibration.xls* that has customer id their preference for using public transportation (*x1-x12*) along with their demographic characteristics (b) *prospect.xls* has customer id and demographic information pertaining to a list of prospective users. See data description file for additional details.

    The department is seeking your advice on whom to target. Base your decision on the following analysis:

    (a) Replicate part (a) in assignment II. Create a hold-out sample by randomly sampling 20% of the customers after performing segmentation.

    (b) In part (b) of assignment II you conducted Response Analysis using with Discriminant Analysis. In this assignment you will conduct Response Analysis with Logistic Regression and use the estimates from Logistic Regression to classify consumers in *prospect.xls* into the appropriate groups.

    (c) Use the hold-out sample instead of the prospects to classify the customers using Discriminant Analysis and Logistic Regression and compare the two results.

Interpretation of the Results

(i)   Based on your results in (b) which demographic variables help discriminate potential *users* of public transportation from *non-users*. Why?

(ii)   How good is your logistic regression – what criteria would you use to ascertain this?

(iii)   Which demographic variables are significant and why?

(iv)   Comment on the cross-tabulation in (c). What does it tell you about the effectiveness of your classification procedure?