BZAN 535 - HW#10 – Due Nov. 23 (not Nov. 21)
This assignment concerns logistic regression.

1. Using the attached file from the journey data, fit a logistic regression model for Y = Produce (based on the basket containing one item or more from the Produce Department) and X = basket $ total.

a) Report the slope coefficient and interpret this corresponding odds.

b) Does this model have significant lack of fit? Are their any outliers evident in the deviance residual plot? What do you conclude? Is this model generally satisfactory?

2. Fit a logistic regression model for Y = Soft drinks and X= basket $ total. Report the equation of the model for logit(Soft drinks). Then re-write this equation for the probability of a basket containing soft drinks.

a) Contrast the model for predicting purchase of soft drinks with the model for predicting purchase of produce. In particular, which are more common in baskets and how is this reflected in the logistic regression model? What does the difference in the slopes tell you about shopping behaviors related to produce and soft drinks?

b) Show and interpret the lift plots for the model in 1a and 2. Explain what about the fitted logistic regression models cause the differences you see in lift.

3. When we discussed the odds ratios in Lectures 20-22, we noted that ignoring basket size made commodities all appear to have affinity for one another.

a) What is the odds ratio for produce and soft drinks? Interpret this quantity.

b) Use a logistic regression to show that the answer in a) is misleading, and that when we control for basket size, the correct value is less than 1. Show the results of such an analysis, and interpret the resulting odds ratio. [Hint: Think about how a model containing basket size (in $) provides estimates for other effects in the model, controlling for basket size.]

4. Write the query for producing the data I provided for questions 1-3. Be sure to exclude baskets that involve gas, as we did in a query for HW9.

5. Using logistic regression for the buytest data, which factors affect the odds of buying? Answer this question by:

a) Reporting a logistic regression model that includes all the explanatory variables that have a statistically significant effect on P(Purchase).

b) Describing the most likely buyer according to your model; include the estimated probability of a purchase for such individuals.