Assume 95% levels of confidence (α = .05) unless otherwise indicated. You should hand in your computer software output along with your analysis of each problem.

1. The data provided (n=98), in Microsoft Excel format (K:\Departments\MTH\MTH Department\Applied Stats 2012/problem1.xls) contains information on the outbreak of a disease spread by mosquitos in a large U.S. City with a large river flowing through the middle. The data, taken from a random sample of city residents, is coded as follows:

Subject:	The identification number of the subject being examined.
Age:	The age of the subject in the study.
Status:	The economic status of the subject being studied, where (1 = Upper Income, 2=Middle Income, and 3= Lower Income)
Sector:	Subject location (0= lives 2 miles or more away from the river and 1=lives within 2 miles of the river.)
Disease	Whether or not the subject has the disease.

   A. **Using the data set provided, create a valid regression model that can be used to predict the probability that a city resident has the disease. You may need to recode some of the categorical variables. Demonstrate (diagnostics, p-values, etc.) that you have created a valid model. State you final model with the coefficients.**

   B. **Discuss how each significant predictor influences the probability (or odds) of getting the disease.**

   C. **Assume that a cutoff of p=.35 has been decided. In other words, if the model predicts a probability greater than or equal to .35, we conclude the disease is present. If the model predicts a probability less than .35, we conclude the disease is NOT present. Based on this situation, determine the misclassification rate for this model by indicating the number of incorrect predictions.**

2. A U.S. political consultant, located in the State of Montana, is interested in studying the effects of party affiliation (0= Democrat and 1= Republican) and County of Residence (1= Jefferson, 2=Lewis/Clark, 3=Powder River, 4=Stillwater, 5=Sweet Grass, and 6=Yellowstone) on voter participation in a recent election (the dependent variable is the percentage of county residents who actually went to the polls and voted).

Consider this a two-factor completely crossed Analysis of Variance study. The overall objective of this study is to investigate the influence both PARTY and COUNTY have on Voter Percentage. We will employ a simultaneous comparison procedure with the primary objective being the pairwise comparison of factor means.

  A. Create an appropriate and valid statistical model that can examine the pairwise comparison of factor means. Justify the validity of your model in terms of diagnostics using software output and your comments. If any modifications to the data are required, discuss your modifications then demonstrate the validity of your modified model. You can then use your modified model, if necessary, to answer parts B and C.

  B. Discuss the results of your model. Specifically:
     1. **Is there evidence of any statistically significant interaction between the factors? Discuss why or why not. What does the possibility or absence of interaction between the factors actually indicate within the context of this analysis.**
     2. **Is there any evidence of a difference between factor level means (cell means)? Discuss why or why not.**

  C. As previously stated, the research objective is the pairwise comparison of the factor means. We wish to understand the influence of the simple main or main effects on the dependent variable. Using an appropriate multiple comparison procedure, provide simultaneous comparisons (underlining) of the treatment means. Justify the multiple comparison procedure you have chosen.

THE END