STA-590 Qualifier Spring 2013

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

1. The data (n=6), provided in Microsoft Excel Format (problem #1.xls) concern the energy radiated from a carbon filament lamp in cm**2 per second (the dependent variable) and the predictor variable which is the absolute temperature of the filament in 1000 degrees Kelvin.

   A two parameter model is considered as follows:

   $$Y = \gamma_1 * X ** \gamma_2 + e$$

   Based upon this information you are asked to do the following:

   A. Using an appropriate method, obtain the starting values for γ_1 and γ_2. Provide your program and output to demonstrate how you arrived at these values.
   B. Using the starting values obtained in part A, find the least squares estimates of the parameters.
   C. Evaluate the validity of your model using the appropriate diagnostics. Provide the necessary output and your comments on the diagnostics you have selected.
   D. Assume that large-sample inferences can be employed reasonably here and the diagnostics are acceptable. Obtain an approximate 95% confidence interval for γ_2.

2. The data provided (n=98), in Microsoft Excel format (problem #2.xls) contains information on the outbreak of a contagious disease spread by mosquitos in a large U.S. city. This city also has a major river flowing through the middle of the downtown area. 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 1 mile or more away from the river and 1=lives within 1 mile of the river.)
Disease	Whether or not the subject has the disease (0 = No Disease and 1= Disease Present).

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. **Estimate how each statistically significant predictor influences the likelihood of getting the disease.**

C. **Assume that the cost of incorrectly predicting no disease P(∏ = 0 | Disease present) is 10 times the cost of incorrectly predicting the disease P ((∏ = 1 | No Disease) in a subject. Based on the model chosen, determine the cutoff value that minimizes this misclassification rate.**

3. The data provided (problem #3.xls) concerns analysis of Zooplankton in two different lakes receiving three different types of diet supplements. You set up twelve tanks in your laboratory, six each with water from one of the two lakes. You randomly add one of three nutrient supplements to each of the 6 tanks, with two replicates, and after 30 days you count the zooplankton in a unit volume of water.

   Consider this a two-factor completely crossed Analysis of Variance study. The overall objective of this study is to investigate the influence both Lake and Supplement have on Zooplankton count. 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