STA-590

1. The data (prob1.txt) provided in this problem are the result of a study involving the analysis of performance degradation data from accelerated tests, published in IEEE Transactions on Reliability. The response variable (Y) is dialectric breakdown strength in kilo-volts, and the predictor variables are time in weeks (X_1) and temperature in degrees Celcius (X_2). A three parameter exponential class model is considered as follows:

```
Y = γ₀*exp(γ₁*X₁)*X₂**γ2 +  e
```

```
 where exp is the exponential function.
```

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

   A.   Using an appropriate method, obtain the starting values for γ_0, 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.   Consider the following data set (prob2.txt) which attempts to investigate allegations of sex discrimination in the hiring practices of a particular company. The variables are coded as follows:

```
 Y= 1 if hired  Y=0 if not hired
 Edu: Education in years
 Exp: Years of Work Experience
 Gen =1 if a male applicant  Gen =0 if female applicant
 Edu_Exp Edu_Gen  Exp_Gen  :  Pairwise interaction terms.
```

Based on this information, answer the following questions:
   A.   If possible, construct a valid model that can estimate the probability of being hired by this firm.
   B.   Evaluate the validity of your model using appropriate diagnostics. Provide the necessary output and your comments on the diagnostics you have selected.
   C.   Determine if the company in question practices sex discrimination.
   D.   Using your chosen model, estimate the probability of a female with 1 year of work experience and 4 years of education being hired. Also provide the odds of being hired.
   E.   Using your chosen model, estimate the probability of a male with 1 year of work experience and 4 years of education being hired.
   F.   If possible, interpret the influence experience has on being hired.

3. Scientists are concerned with the effects of acid rain on aquatic life. Studies have shown that at very low pH levels (around pH 3, which is very acidic), many fish die. In the environment, however, the decrease in pH from naturally occurring levels (of around 7) is rather gradual and rarely are very low pH values seen. Scientists are interested in the effects of this lowered pH on fish, particularly in how the lowered pH adversely changes the parental behavior in fish. Fish offspring that are not cared for adequately have a lower survival rate.

   In one study, they observed several parental behaviors of a certain species of fish. Guarding the nest is an important activity because it ensures that the fish eggs and young are safe from predators. Randomly selected fish were exposed to water of varying pH levels and the percentage of time each fish spent guarding and fanning its young was recorded.

   Consider the data (prob3.txt) as follows:

   Guarding: Dependent Variable. Percent of time fish spend guarding their eggs;

   pH: Four categories (1=Neutral,2= Slightly Acidic,3= More Acidic, 4=Highly Acidic)

   Gender: Male or Female.

   Consider this a two-factor balanced and completely crossed ANOVA study. Based on this data, perform the following analysis:

   A. Create an appropriate statistical model with diagnostics included. Discuss your results and diagnostics. Included any necessary output.
   B. Determine if the two factors interact. Based on your results, interpret the meaning of the presence or lack of interaction.
   C. Determine if either factor has a statistically significant effect on Guarding.
   D. The research objective is the pairwise comparison of 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 comparisons (underlining) of the treatment means. Justify the multiple comparison procedure used.