BIO 226 Homework Assignment 7

This assignment will not be graded and does not need to be turned in.
(see further comments in Instructions below).

Purpose:

1. To review concepts of generalized linear models for a single response.
2. To provide an introduction to PROC GENMOD in SAS for fitting generalized linear models.

Instructions:

1. For each question requiring data analysis, support your conclusions by including only the relevant SAS output in your answer.
2. Include your SAS program as an appendix to your solutions.
3. This assignment will not be graded and does not need to be turned in. A solution is posted on the class web site for self-review. The assignment should, however, be completed and reviewed prior to the class on Tuesday April 24. If you have any questions about this assignment, please come to one of the TA or Instructor Office Hours.

Logistic Regression: Bladder Tumor Study (tumor.txt)

In an experimental study of patients with bladder cancer, patients underwent surgery to remove tumors. Following surgery, patients were randomized to treatment with thiotepa or to receive a placebo. Subsequently patients were examined at 18, 24, 30 and 36 months. For this assignment, the focus is only on data at month 18. The response variable is binary, indicating whether or not there was a new tumor at the 18 month visit (Y=1 if new tumor; Y=0 if no new tumor). The objective of the analysis is to determine the effect of treatment on new tumor recurrence at month 18.

The data are stored in the ASCII file, **tumor.txt**, on the course web page. Each row of the data set contains the following three variables: ID, x (1=thiotepa; 0=placebo), Y.

1. Assuming a Bernouilli distribution for the recurrence of a new tumor at the 18 month visit, fit the following model relating the mean, μ_i, to treatment, x:

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x$$

2. What is the interpretation of the estimates for β_1 and β_2?

3. What is the predicted probability of recurrence of a new tumor among those who received placebo?

4. What is the predicted probability of recurrence of a new tumor among those who received thiotepa?

5. Construct a 95% confidence interval for the log odds ratio comparing thiotepa to placebo.

6. Construct a 95% confidence interval for the odds ratio comparing thiotepa to placebo.

Poisson Regression: Epilepsy Study (seizure4.txt)

The data are from a randomized trial comparing a treatment, Progabide, to placebo among patients suffering from epileptic seizures. The study measured the number of seizures in an 8-week period before any treatment, and in each of four 2-week treatment periods, in which patients received with Progabide or placebo in addition to standard therapy. For this assignment, the focus is on the data from the last 2-week treatment period. The objective is to assess whether or not Progabide is effective in reducing epileptic seizures.

The data are in the ASCII file, **seizure4.txt**, available on the course web page. Each row of the data set contains the four variables: ID, x (1=Progabide; 0=placebo), age, Y, where age is the age in years of the subject at baseline, and Y is the count of seizures in the last 2-week period.

1. Assuming a Poisson distribution for the counts in the last 2-week period, fit the following model relating the mean, μ_i, to treatment, x:

$$\log(\mu_i) = \beta_1 + \beta_2 x$$

2. What is the interpretation of the estimates for β_1 and β_2?

3. Construct a 95% confidence interval for the rate ratio comparing Progabide to placebo.

4. Extend the model in question 1 to adjust for the effect of age.

5. Construct a 95% confidence interval for the age-adjusted rate ratio comparing Progabide to placebo.

6. Extend the model in question 4 to allow for potential overdispersion.

7. Construct a 95% confidence interval for the age-adjusted rate ratio comparing Progabide to placebo after taking account of overdispersion. How does the result obtained differ from that in question 5? Why is this?

Hints:

Bladder Tumor Data

```
PROC GENMOD DATA=tumor DESCENDING;
MODEL y=x / D=BIN LINK=LOGIT;
```

Epilepsy Data

```
PROC GENMOD DATA=epilepsy;
MODEL y=x / D=POISSON LINK=LOG;
```

Note: The scale parameter is held fixed at 1 for the Binomial and Poisson distributions (as in the SAS code above). To allow for overdispersion, we can consider a model where the scale parameter is estimated using Pearson's chi-squared statistic divided by the corresponding degrees of freedom. To achieve this in SAS, use the option SCALE=PEARSON in the MODEL statement.