EXAM 2
APRIL 11, 2006
STAT 8320

Please place your name on the line above. You may use a calculator and equation sheets comprising not more than four sides of an 8.5 × 11 inch sheet of paper. Your exam packet should contain a total of fourteen pages: one title page, four pages of questions, and nine pages of output.
Problem 1: Recall that the Bernoulli distribution has density
\[ f(y; \xi) = \xi^y (1 - \xi)^{1-y}. \]

(10 pts.) Write this distribution in exponential form. Be sure to define the functions \( a(\cdot), b(\cdot), c(\cdot) \).

(6 pts.) Find the mean and variance of this distribution using the exponential form.

(8 pts.) What is the form of the canonical link for this distribution? Show that this form has the desired property of a canonical link.

(6 pts.) Find the form of the deviance for this distribution. Assume that your sample contains \( n \) independent observations.
Problem 2: A major cola manufacturer wishes to determine which factors cause consumers to prefer their beverage to the leading competitor. To study these factors in order to market their product more effectively, the company obtains the following information on each of a variety of individuals:

\[
\begin{align*}
  x_1 & \quad \text{the gender of the individual}, \\
  x_2 & \quad \text{the age of the individual}, \\
  x_3 & \quad \text{the ethnicity of the individual}, \\
  x_4 & \quad \text{the socioeconomic status of the individual, and} \\
  x_5 & \quad \text{the level of education of the individual}. 
\end{align*}
\]

Each individual is also given a sample of the company’s soda and the soda of the major competitor, and asked which of the two sodas is preferred.

(8 pts.) Look at the output provided. Which of the models considered in the output would you consider to be “best”, and why?

(10 pts.) Provide a point estimate and confidence interval for the odds ratio for gender based upon the model you have selected. Interpret these values; make sure to mention whether or not your interval implies that gender significantly effects the odds, and what the point estimate says about the probability of preferring the company’s soda.

(6 pts.) Use the information for your “best model” to estimate the probability of preferring the company’s soda for a 25 year old Hispanic woman with a college degree in the middle socioeconomic group.
Problem 3: Consider the following model for the decay of a substance over time:

\[ Y_i = \exp(-\theta X_i) + \epsilon_i \]

(8 pts.) Show how to use the regression method to find a starting values for this model. Specifically, show the form of the ordinary least regression that you would use. Call the estimates for this regression \( \hat{\beta} \), and write the initial estimates for \( \hat{\theta} \) as a function of the \( \hat{\beta} \).

(12 pts.) Find the forms needed to perform one update of the Gauss-Newton algorithm. Fill in these forms for the three data values \((y, x) = (20, 3), (12, 1), (42, 5)\). Note: you do not need do any simplification, but do write the new parameter estimates as a function of all of the above information.

(6 pts.) Suppose that the constraint \( \theta > 0 \) is fundamental to the model. Unfortunately, when fitting the model to a new data set, it is found that \( \hat{\theta} < 0 \). Suggest a solution; provide complete details to obtain full credit.
Problem 4: Suppose that a nonlinear model of the form

\[ Y_i = X_1^{\theta_1} \theta_2^{X_2} \exp(-\theta_3 X_3) + \epsilon_i \]

is fit using some data. The attached output shows the results of this analysis.

(8 pts.) Use the information provided to predict the response for an observation which has \( x_1 = .5, x_2 = 2, x_3 = 10. \)

(6 pts.) Report and interpret Hougaard’s measure for the three parameters in this study. Do any of the values cause you concern?

(6 pts.) Test the logical hypotheses for all three \( \theta \) parameters. Keep in mind that your desired alternative is no effect of the associated \( x \) variable.