STA 312f10 Assignment 9

Do this assignment in preparation for the quiz on Friday, Nov. 19th. Please bring your R and SAS printouts to the quiz; they may be handed in. Please do *not* write anything on your printouts before the quiz, except possibly your name and student number.

1. Consider a Poisson regression model with just one independent variable.

   (a) Write the log likelihood function $\ell(\beta_0, \beta_1)$, and simplify as much as possible.

   (b) Partially differentiate the log likelihood with respect to β_0 and β_1. Setting both derivatives to zero, obtain two equations in two unknowns. Looking at these, it is clear that a general explicit solution is out of the question.

   (c) Here are some data for $n = 3$: $\begin{array}{lll} x_1 = 0 & x_2 = 0 & x_3 = 1 \\ y_1 = 3 & y_2 = 1 & y_3 = 4 \end{array}$.
   Solve the two equations in two unknowns to obtain the maximum likelihood estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$. Your answer is a set of two numbers. Show your work. I checked my answer with R.

2. The Data Sets link on the course web site will lead you to the Seat Belt Data. For a set of months before and after the introduction of a seat belt law in Great Britain, it has the number of drivers killed in traffic accident, the price of fuel (higher price means less driving and possibly fewer accidents), and an indicator for whether the seat belt law was in effect. This is a time series, but if you accept the independent increments part of the Poisson process model, then the numbers of events in non-overlapping periods (months) are independent, conditionally on the values of the explanatory variables. In general one should *never* accept such a model without checking, but this is just a homework problem so we will proceed with a Poisson regression.

   Oh dear; I told you what to do! Anyway, test whether, controlling for the price of gas, the seat belt law is related to number of fatalities. Do the likelihood ratio test and compare it to the Z test produced by the `summary` function. They are equivalent only as $N \to \infty$. Do they lead to the same conclusion? State the conclusion (or conclusions) in plain, non-statistical language.

3. The Data Sets link on the course web site will lead you to the Florida Death Penalty Data discussed in lecture. Using SAS,

   (a) Read and label the data nicely, and produce frequency distributions (one-dimensional tables) for all three variables.

   (b) Consider a logistic regression model without any interaction between the explanatory variables[1].

      i. For each of the explanatory variables, test whether it is related to the outcome, controlling for the other explanatory variable. Carry out likelihood ratio tests using a calculator on values from your printout. But don't write the results on your printout, and don't include them as comment or title statements either. Can you locate the corresponding Wald tests on your printout? Be able to state the conclusions in plan language based on estimated regression coefficients.

      ii. How many tests can you find for $H_0 : \beta_1 = \beta_2 = 0$? Do they all lead to the same conclusion?

[1] What is the meaning of this in terms of a *log-linear* model? It's different!