Stats 539 Homework 3: Due Tuesday Feb. 14 by 10:50am

1. Recall from equation (4.15) on p. 133 that the deviance of a model M is equal to:

$$D(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n} w_i[y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)].$$

Let $\hat{\mu}_i$ be the fitted values for model M. Use this expression of the deviance to show that

   (a) the deviance for a Normal GLM with identity link is equal to:

$$\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2.$$

   (b) the deviance for a Poisson GLM with log link is equal to:

$$2\sum_{i=1}^{n}(y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)).$$

2. Consider the simple case where $nY \sim Bin(n,\pi)$ and we would like to test the null hypothesis $H_0 : \pi = \pi_0$ versus $H_0 : \pi \neq \pi_0$.

   (a) Show that the chi-squared forms of the test statistics are:

$$\text{Likelihood-ratio: } -2(\ell_0 - \ell_1) = -2\log\left[\frac{\pi_0^{ny}(1-\pi_0)^{n(1-y)}}{y^{ny}(1-y)^{n(1-y)}}\right]$$

$$\text{Wald: } z^2 = \frac{(y-\pi_0)^2}{[y(1-y)]/n]}$$

$$\text{Score: } z^2 = \frac{(y-\pi_0)^2}{[\pi_0(1-\pi_0)/n]}$$

   (These are also given on p. 131 of our textbook.)

   (b) A recent study examined expressions of commitment between two partners in a committed romantic relationship. One aspect of the study involved 47 heterosexual couples who are part of an online pool of people willing to participate in surveys. These 47 couples were asked about which person was the first to say "I love you." In 26 of the 47 couples, the male said "I love you" first. Set up a hypothesis test to test the null hypothesis that males and females in a committed romantic relationship are equally likely to say "I love you" first, defining the parameter of interest in context. Calculate the likelihood-ratio test statistic, the Wald test statistic, and the score test statistic, and the corresponding p-value for each. Do the three tests yield different conclusions? Write a conclusion of the study in context of the problem (using the LRT p-value).

3. Agresti Exercise 4.16 (p. 161)

4. Consider again the Framingham heart study discussed in Lecture 7. A description of the data set can be found here: http://www.ics.uci.edu/∼staceyah/111-202/data/framingham.html. Read the data into R using the following command:

```
fram <- read.table("http://www.math.montana.edu/shancock/courses/stat539/
data/Framingham.txt", header=TRUE)
```

   (a) Write an R function to run the iterated weighted least squares algorithm for a generalized linear model that uses the logit link. (You may just modify the IWLS R code from class.) Use this function to calculate the maximum likelihood estimates of the coefficients for a logistic regression model with response `chdfate` and predictors `sex` and `age`. (Be sure to first re-code `sex` as a factor.) Turn in a well-commented .R script file with your IWLS R function and the R commands used to run the IWLS algorithm on the Framingham data to the "Homework 3 R Code" Assignment submission folder in D2L.

   (b) For each of the following questions,

   - Create at least one well-labeled, informative plot that helps illuminate the question of interest.
   - Clearly state the reduced model and full model for a model comparison test that addresses the question.
   - Use the `anova` function in R to carry out the appropriate likelihood ratio test. Write a conclusion of the test that addresses the question.
   - Calculate and interpret an approximate 95% confidence interval(s) that addresses the question.

   i. How do the odds of coronary heart disease differ between men and women, adjusting for age?

   ii. Controlling for age and sex, what effect does serum cholesterol level have on the odds of coronary heart disease?

   iii. Controlling for age and sex, does the effect of serum cholesterol level on the odds of coronary heart disease differ among different levels of body mass index? If so, how?