CPH 931 — Fall 2009 — Dr. Charnigo

Written Assignment 3

Written Assignment 3 is due on Friday 23 October at the end of class.

[50] 1. Consider the data in {PostOpExcel.xls}, a (substantially) modified version of the post-operative patient data set from {http://archive.ics.uci.edu/ml/datasets/Post-Operative+Patient}. The variables in {PostOpExcel.xls} are HIGHTEMPI (= 1 if internal temperature exceeds 37 degrees Celsius, = 0 otherwise), HIGHTEMPS (= 1 if surface temperature exceeds 36.5 degrees Celsius, = 0 otherwise), HIGHBP (= 1 if blood pressure exceeds 130/90, = 0 otherwise), STATUS (= 2 if patient goes to intensive care, = 1 if patient is admitted to an ordinary hospital room, = 0 if patient is sent home), and COMFORT (patient's subjective comfort level on a scale of 0 to 20). The goal is to predict STATUS from HIGHTEMPI, HIGHTEMPS, HIGHBP, and COMFORT. For convenience in what follows, let the four explanatory variables be denoted by X_1, X_2, X_3, and X_4 respectively.

[10] a. Fit the polytomous regression model

$$\log\left[\frac{p_{1|\mathbf{x}}}{p_{0|\mathbf{x}}}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4, \quad \log\left[\frac{p_{2|\mathbf{x}}}{p_{0|\mathbf{x}}}\right] = \alpha^* + \beta_1^* x_1 + \beta_2^* x_2 + \beta_3^* x_3 + \beta_4^* x_4,$$

where $p_{1|\mathbf{x}}$, $p_{0|\mathbf{x}}$, etc., have the meanings provided in Lecture 5. Report parameter estimates, standard errors, and p-values.

[10] b. Can HIGHBP be removed from the polytomous regression model in part a? Specify an appropriate null hypothesis in terms of model parameters and then state whether the null hypothesis is rejected.

[10] c. Suppose that a post-operative patient with comfort level 12 has an internal temperature of 37.5, a surface temperature of 36.0, and blood pressure of 125/80. Use the polytomous regression model to estimate the probability that such a patient gets admitted to an ordinary hospital room.

Hint: First estimate $p_{1|\mathbf{x}}/p_{0|\mathbf{x}}$ and $p_{2|\mathbf{x}}/p_{0|\mathbf{x}}$. Then use

$$p_{0|\mathbf{x}} + p_{1|\mathbf{x}} + p_{2|\mathbf{x}} = 1 \iff 1 + \frac{p_{1|\mathbf{x}}}{p_{0|\mathbf{x}}} + \frac{p_{2|\mathbf{x}}}{p_{0|\mathbf{x}}} = \frac{1}{p_{0|\mathbf{x}}}$$

to estimate $p_{0|\mathbf{x}}$, $p_{1|\mathbf{x}}$, and $p_{2|\mathbf{x}}$.

[10] d. Fit the ordinal logistic regression model

$$\log[O_{2|\mathbf{x}}] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4, \quad \log[O_{12|\mathbf{x}}] = \alpha^* + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4,$$

where $O_{2|\mathbf{x}}$, $O_{12|\mathbf{x}}$, etc., have the meanings provided in Lecture 5. Report parameter estimates, standard errors, and p-values.

[10] e. Suppose that a post-operative patient with comfort level 12 has an internal temperature of 37.5, a surface temperature of 36.0, and blood pressure of 125/80. Use the ordinal logistic regression model to estimate the probability that such a patient gets admitted to an ordinary hospital room. Compare your answer to that obtained in part c.

Hint: Begin by estimating $O_{2|\mathbf{x}}$ and $O_{12|\mathbf{x}}$ for such an individual. Noting that $O_{2|\mathbf{x}} = p_{2|\mathbf{x}}/(1 - p_{2|\mathbf{x}})$, you can use your estimate of $O_{2|\mathbf{x}}$ to solve for an estimate of $p_{2|\mathbf{x}}$. Then, noting that $O_{12|\mathbf{x}} = (p_{1|\mathbf{x}} + p_{2|\mathbf{x}})/(1 - p_{1|\mathbf{x}} - p_{2|\mathbf{x}})$, you can use your estimates of $O_{12|\mathbf{x}}$ and $p_{2|\mathbf{x}}$ to solve for an estimate of $p_{1|\mathbf{x}}$. Finally, an estimate of $p_{0|\mathbf{x}}$ is determined by the estimates of $p_{1|\mathbf{x}}$ and $p_{2|\mathbf{x}}$.

[50] 2. Consider the SARS data set discussed in Lecture 6 and available at {http://www.richardcharnigo.net/CPH931F09/SARS.xls}.

Note: In all that follows, please confine your attention to the data from SINGAPORE.

[10] a. Fit a Poisson regression model with DAILYINF as the response variable and TIME as the sole explanatory variable. What is the model-based estimate of the incidence rate ratio comparing day 15 to day 10? What about day 60 to day 55?

Hint: For a generic Poisson regression model, we have

$$\frac{\mu_{\mathbf{x}_{new}}}{\mu_{\mathbf{x}_{old}}} = \frac{\exp[\alpha + \beta_1(x_1 + 5) + \beta_2 x_2 + \cdots + \beta_k x_k]}{\exp[\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k]}$$

when $\mathbf{x}_{new}$ and $\mathbf{x}_{old}$ differ only by 5 units on X_1. Note also that the above expression can be simplified.

[10] b. Make a plot of DAILYINF versus TIME. Do your answers to part a <u>seem</u> reasonable?

[10] c. Refit the Poisson regression model by including the quadratic term TIME2 along with the linear term TIME. Now what is the model-based estimate of the incidence rate ratio comparing day 15 to day 10? What about day 60 to day 55?

Hint: Modify the hint for part a by changing the numerator x_2 to $(x_1 + 5)^2$ and the denominator x_2 to x_1^2.

[10] d. Are the two numbers in your answer to part c significantly different from each other? Specify an appropriate null hypothesis and then state whether the null hypothesis is rejected.

[10] e. Does your answer to part d change if you correct for overdispersion? Investigate with both a multiplicative adjustment to the standard errors in Poisson regression and a negative binomial regression.