

Stat 5102 (Geyer) Spring 2013  
Homework Assignment 12  
Due Wednesday, May 8, 2013

Solve each problem. Explain your reasoning. No credit for answers with no explanation. If the problem is a proof, then you need words as well as formulas. Explain why your formulas follow one from another.

**12-1.** The data set in the URL

<http://www.stat.umn.edu/geyer/5102/data/prob12-1.txt>

has two variables  $\mathbf{x}$  (the predictor variable) and  $\mathbf{y}$  (the response variable). The response is zero-or-one-valued, so these data are suitable for logistic regression.

- (a) Fit a GLM with logit link having natural parameter (linear predictor) that is quadratic in  $\mathbf{x}$ , that is

$$\theta_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$$

Report the regression coefficients table for this fit.

- (b) Give a 95% confidence interval for the natural parameter value for an individual with  $x_i = 30$ .
- (c) Give a 95% confidence interval for the mean parameter value for an individual with  $x_i = 30$  that is of the form point estimate  $\pm$  critical value  $\times$  standard error.
- (d) Give a 95% confidence interval for the mean parameter value for an individual with  $x_i = 30$  that is made by transforming the endpoints of the interval from part (b) from the natural parameter scale to the mean value parameter scale.

**12-2.** Redo problem 12-1 substituting probit link for logit link.

**12-3.** Redo problem 12-1 substituting cauchit link for logit link.

**12-4.** The data set in the URL

<http://www.stat.umn.edu/geyer/5102/data/prob12-4.txt>

has three variables  $\mathbf{x1}$  and  $\mathbf{x2}$  (the predictor variables) and  $\mathbf{y}$  (the response variable). The response is zero-or-one-valued, so these data are suitable for logistic regression.

- (a) Fit a GLM with logit link having natural parameter (linear predictor) that is linear in  $x_1$  and  $x_2$ , that is

$$\theta_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i}$$

Report the regression coefficients table for this fit.

- (b) Same as part (a) but with natural parameter quadratic in  $x_1$  and  $x_2$  (there will be six regression coefficients).
- (c) Same as part (a) but with natural parameter cubic in  $x_1$  and  $x_2$  (there will be ten regression coefficients).
- (d) Compare the three models fit in parts (a) through (c) using likelihood ratio tests. There will be two tests, one comparing (a) to (b) and one comparing (b) to (c).
- (e) Interpret these tests. What do they say about the correctness of each of the models?

**12-5.** The data set in the URL

<http://www.stat.umn.edu/geyer/5102/data/prob12-5.txt>

has twenty-one variables  $x_1$  through  $x_{20}$  (the predictor variables) and  $y$  (the response variable). The response is normal, so these data are suitable for ordinary least squares regression.

- (a) Use the R function `regsubsets` in the R package of the same name to examine all  $2^{20}$  (more than a million) models each containing a subset of the predictor variables plus an intercept.
- (b) Which model fits best according to AIC?
- (c) Which model fits best according to BIC?