Homework 7
BIOST 515
Due: March 2, 2004 **in class**
This assignment is worth 30 points.

1. This problem is a continuation of problem 3 from homework 6.

   (a) To assess the appropriateness of the logistic regression function, form four groups of 12 cases each according to their fitted logit values. Plot the observed proportions in each interval against the midpoint of the fitted logit intervals. Is the plot consistent with the S-shaped curve you would expect to see? Explain.

   (b) Obtain the deviance residuals and plot them. Do there appear to be any outlying cases?

   (c) A prediction rule is to be developed. Based on the data, find the overall error (misclassification) rate, the error rate for clients receiving the flu shot and the error rate for clients not receiving the flu shot for the following cutoffs: 0.40, 0.50, 0.60, 0.70.

   (d) Based on your results in part 1c, which cutoff minimizes the total error rate? Are the error rates for clients receiving the flue shot and for clients not receiving the flu shot fairly balanced at this cutpoint? Explain.

   (e) Plot the ROC curve and calculate the area under the ROC curve. You may find the ROC library in R helpful.

2. In this problem, you will revisit the King county birth data from homework 4. Refit the model you selected based on scientific reasoning using logistic regression with low birth weight as the outcome. Perform the appropriate model diagnostics, and examine the predictive ability of your model. Be sure to explain your results as you would to a DSHS investigator.