BIOSTATISTICS 767, Spring 2017
Homework 5

1. Problem 13.3.

   Add 13.3.0: Present simple descriptive statistics and describe what they show.

   Files: insomnia.*

2. This problem was explained in class. For logistic regression with n independent Bernoulli 0/1 outcomes, after fitting by maximum-likelihood, the deviance (G^2) can be expressed as a function of the estimates $\hat{\beta}$ (or the fitted values) only, without the responses $\{y_i\}$ appearing in the formula. i.e. If I give you ONLY the fitted values (just one column of numbers), and nothing else, you will be able to compute the deviance. The deviance is

$$G^2 = 2\sum_{i=1}^{n}[y_i \log(y_i/\hat{\mu}_i) + (1-y_i)\log\{(1-y_i)/(1-\hat{\mu}_i)\}],$$

   where $\hat{\mu}_i$ is the fitted value for the i-th observation.

   What does G^2/n estimate?

   What is the maximum possible value of G^2/n?

   Compute G^2 for these vectors:
   $\hat{\mu}^\top = (0.4, 0.4, 0.4, 0.4, 0.4)$
   $\hat{\mu}^\top = (0.05, 0.05, 0.1, 0.9, 0.9)$
   $\hat{\mu}^\top = (0.01, 0.01, 0.01, 0.98, 0.99)$
   Comment.

Exercise (not homework, do not turn in):
For logistic regression with independent Bernoulli 0/1 outcomes, after fitting the model $E[Y_i] = \mu$ (intercept only) by maximum-likelihood, compute the value of Pearson's X^2. Surprised?!