APPLIED STATISTICS
TUTORIAL 8

Question 1 (ex from Chapter 20 of the class text "The Statistical Sleuth")

Duchenne Muscular Dystrophy (DMD) is a genetically transmitted disease, passed from a mother to her children. Boys with the disease usually die at a young age; but affected girls who usually do not suffer symptoms, may unknowingly carry the disease, and may pass it to their offspring. It is believed that 1 in 3300 women are DMD carries. A woman might suspect she is a carrier when a related male child develops the disease. Doctors must rely on some kind of test to detect the presence of the disease. The file "DMD.csv" contains levels of two enzymes in the blood, creatine kinase (CK) and hemopexin (H) for 38 known DMD carries and 82 women who are not carriers (controls). It is desired to use these data to obtain an equation for indicating whether a woman is a likely carrier.

a) Make a scatterplot of H versus log(CK); use one plotting symbol to represent the controls on the plot and another to represent the carriers. Does it appear that these enzymes might be useful predictors of whether a woman is a carrier?
b) Fit the logistic regression of carrier on CK and CK^2. Next fit the logistic regression of carrier on log(CK) and $[\log(CK)]^2$.
c) Fit the logistic regression of carrier on log(CK) and H.
e) Typical values of CK and H are 80 and 85. Suppose that a suspected carrier has values of 300 and 100. What are the odds she is a carrier relative to the odds that a woman with typical values (80 and 85) is a carrier?