BST 760: Advanced Regression
Breheny

Assignment 6

Due: Tuesday, March 26

1. Duchenne Muscular Dystrophy (DMD) is a sex-linked genetic disease. Boys with the disease usually die at a young age, while affected girls usually do not suffer symptoms and may unknowingly carry the disease and pass it to their offspring. It is desirable to have some kind of test to detect whether or not a woman is a carrier of the disease. The dataset `dystrophy.txt` contains information from a 1981 study attempting to develop such a test based on two serum enzymes, creatine kinase (`CK`) and hemopexin (`H`) for 38 known DMD carriers (`Case`) and 82 women who are not carriers (`Control`). (Note: In the last 30 years, advances in DNA sequencing technology has made it possible obtain definitive answers; however, tests based on the above proteins are still used as rapid and inexpensive alternatives).

   (a) Use logistic regression to model the way in which case/control status depends on creatine kinase and hemopexin. Construct (using the Wald approach) a table containing the estimated odds ratios and p-values for the two enzymes. Provide confidence intervals for the odds ratios, and give some thought as to what would constitute a meaningful difference (δ_j) for the two enzymes when calculating the odds ratios.

   (b) Can you calculate confidence intervals for the odds ratios in part (a) using the likelihood ratio approach? If so, calculate them. If not, explain why you can't do so.

   (c) Can you carry out the hypothesis testing in part (a) using the likelihood ratio approach? If so, perform the tests. If not, explain why you can't do so.

   (d) Describe (quantitatively) the relationship between creatine kinase levels and the likelihood that a woman is a carrier without using the phrase "odds ratio" (you can use "odds", just not "odds ratio").

   (e) Suppose a woman randomly selected from the population has a hemopexin level of 100 and a creatine kinase level of 150. Can you estimate the probability that she is a carrier? If so, estimate it. If not, explain why you can't do so.

   (f) It is estimated that 1 in 3,300 women are carriers. Treating this as a known constant, calculate the sampling ratio τ_1/τ_0.

   (g) Based on your answer to (f), calculate the probability from part (e).

   (h) Compare[1] the following three numbers: (i) the probability you calculated in (g), (ii) the probability you obtain from naïvely using the logistic regression model and its intercept as if it were a cohort study, and (iii) the marginal probability of being a carrier (*i.e.*, if you don't know a woman's hemopexin/creatine kinase levels).

[1] Do not just say that A is bigger than B. Discuss why A is bigger than B and what it means.