Biostatistics 515/518 Winter 2020
HW6
(2 problems)

The class website contains the 'WCGS' dataset that was collected to identify risk factors for coronary heart disease (CHD).

In 1960-1961, 3,154 healthy, middle-aged men were entered into the Western Collaborative Group Study, a long-term study of coronary heart disease (CHD). The risk of coronary heart disease mortality was studied for several variables measured at baseline, i.e., Type A/B behavior, systolic blood pressure, serum cholesterol level, cigarette smoking status, and age. Unfortunately, there is no codebook for the dataset. Fortunately, the variable descriptions in the dataset are largely informative.

The response variable we will use in this assignment is chd69, which indicates whether a study subject experienced a CHD event during the study. Since study subjects were in the study for differing lengths of time, this is not really a scientifically well-defined variable. For the purposes of this assignment, we will ignore that issue. For example, we can pretend that chd69 records whether a study subject experienced a CHD event within 5 years of enrolling in the study.

1.

a. Use logistic regression to investigate an association between age and having a CHD event using the binary response variable chd69. Write out the fitted model. Interpret the intercept and slope of the fitted model, and comment on whether these interpretations are reasonable or likely of scientific interest.

b. Examine the fitted model by making a plot that has age on the horizontal axis and the **log odds** of having a CHD event on the vertical axis. Restrict your plot to the range of ages in the dataset.

c. Examine the fitted model by making a plot that has age on the horizontal axis and the **probability** of having a CHD event on the vertical axis. Restrict your plot to the range of ages in the dataset.

d. Using the fitted model, make a table that gives the probability of having a CHD event and also the odds of having a CHD event for age 40, 45, 50, 55, and 60. For each age>40, use the fitted model to compute a risk ratio and an odds ratio comparing the groups 5 years apart. For example, in the row for age 45, you table should give P(CHD|age=45)/P(CHD|age=40) and in another column odds(CHD|age=45)/odds(CHD|age=40). Comment on these risk ratios and odds ratios. (You do not need to include confidence intervals in your table.)

e. Using your statistical software, use your fitted model to estimate the odds ratio comparing men 5 years apart in age (include a confidence interval). Summarize the results in language suitable for a scientific publication. Compare this odds ratio to the odds ratios in part d.

f. Using your simple logistic regression model, perform a Wald test that age is not associated with having a CHD event. Report the results using language suitable for a scientific publication.

g. Using your simple logistic regression model and a restricted model, perform a likelihood ratio test that age is not associated with having a CHD event. How does the p-value compare to your answer to part f?

2. Consider the binary personality "type" variable dibpat and the binary variable of having a CHD event chd69.

a. Make a 2x2 table to summarize these variables. Perform a chi-square test and summarize the results, including the p-value. Use your 2x2 table to estimate P(CHD|dibpat=A) and P(CHD|dibpat=B) (do not use regression). Also use your 2x2 table to estimate odds(CHD|dibpat=A) and odds(CHD|dibpat=B) and the ratio of these odds (again, do not use regression).

b. Fit a logistic regression model with CHD as the response and dibpat as the predictor. Summarize the results in terms of an estimated odds ratio (with confidence interval) and p-value using language suitable for a scientific publication.

c. Fit a logistic regression model with dibpat as the response and CHD as the predictor. Summarize the results in terms of an estimated odds ratio (with confidence interval).

d. Compare the odds ratio in a and b and compare the p-values in a and b. Comment on why they are similar or different.

e. Compare the odds ratio in b and c. Comment on why they are similar or different.