Stats 539 Homework 4: Due Tuesday Feb. 21 by 10:50am

1. Agresti Exercise 5.17 (p. 196-7)

2. Agresti Exercise 5.30 (p. 199)

3. Using the data given in Agresti Exercise 5.35 (p. 201), answer the following questions:

   (a) Enter the data into R in grouped format. Show the R code used to create the data frame and the first few rows.

   (b) Calculate and interpret the sample marginal odds ratio of developing AIDS for those taking AZT immediately to those who wait until their T cells showed severe immune weakness.

   (c) Calculate and interpret the sample conditional odds ratio of developing AIDS for those taking AZT immediately to those who wait until their T cells showed severe immune weakness, for black subjects.

   (d) Do these data exhibit Simpson's Paradox? Why or why not?

   (e) Fit a logistic regression model to these data using AZT treatment and race as predictors, with no interaction.

      i. Write the equation of the fitted model, clearly defining any variables and symbols used. Write a sentence interpreting each of the estimated coefficients.

      ii. Express the conditional odds ratio of developing AIDS for those taking AZT immediately to those who wait until their T cells showed severe immune weakness, for black subjects, in terms of the estimated model coefficients. Calculate and interpret a 95% confidence interval for this quantity. How does your estimate compare to the sample conditional odds ratio found in part (c)?

      iii. Use the deviance of this model to perform a goodness of fit test.

      iv. Plot the deviance residuals versus the fitted values. Which covariate groups contribute most to lack of fit? how?

4. Agresti Exercise 5.38 (p. 201). Your analysis should include a short description of how you chose your model with relevant tests, model checking methods including Hosmer-Lemeshow goodness of fit test and residual analysis, interpretation of the coefficients in the final model, and a short paragraph summarizing the results. Your write-up should be no more than two pages. Upload a well-commented .R file of the relevant R code you used for the data analysis process to the "Homework 4 R Code" D2L Assignment folder.