HomeWorkCh4.Rmd

F.A. Barrios

11/7/2020

Problem 1

Using the WCGS data for middle-aged men at risk for heart disease, fit a multipredictor model for total cholesterol (chol) that includes the binary predictor arcus, which is coded 1 for the group with arcus senilis, a milky ring in the iris associated with high cholesterol levels, and 0 for the reference group. Save the fitted values. Now refit the model with the code for the reference group changed to 2. Compare the coefficients, standard errors, P-values, and fitted values from the two models. The WCGS data are available at DataRegressBok (in GitHub).

Problem 2

Using the WCGS data referenced in Problem 1, extract the fitted values from the multipredictor linear regression model for cholesterol and show that the square of the sample correlation between the fitted values and the outcome variable is equal to R2. Look at the variables in the model data frame estimated by lm.

Problem 3

If we "center" age—that is, replace it with a new variable defined as the deviation in age from the sample mean, what would be the interpretation of the intercept in the model for SBP (3.2)? If BMI had not been centered, how would the interpretation of the statin use variable change in the model in Sect. 4.6.2 allowing for interaction in predicting LDL?