LAB 5: MULTIPLE LINEAR REGRESSION II[1]

Problem 1: 7.3, 7.12 Brand preference, chapter 7 textbook

Problem 2. (P8.16, P8.20, textbook) Refer to Grade point average Problem 1.19. An assistant to the director of admissions conjectured that the predictive power of the model could be improved by adding information on whether the student had chosen a major field of concentration at the time the application was submitted. Assume that linear regression model (8.33) is appropriate,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad \textbf{(8.33)}$$

where $X1$ is entrance test score and $X2 = 1$ if student had indicated a major field of concentration at the time of application and 0 if the major field was undecided. Data for $X2$ were as follows:

i:	1	2	3	...	118	119	120
X_{i2}:	0	1	0	...	1	1	0

a. Explain how each regression coefficient in model (8.33) is interpreted here.

b. Fit the regression model and state the estimated regression function.

c. Test whether the $X2$ variable can be dropped from the regression model; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. Fit regression model (8.49) and state the estimated regression function.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i \qquad \textbf{(8.49)}$$

e. Test whether the interaction term can be dropped from the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. If the interaction term cannot be dropped from the model, describe the nature of the interaction effect.

HOMEWORK:

Problem 3: (P7.38, textbook) Refer to the SENIC data set in Appendix C.1. For predicting the average length of stay of patients in a hospital (Y), it has been decided to include age ($X1$) and infection risk ($X2$) as predictor variables. The question now is whether an additional predictor variable would be helpful in the model and, if so,

[1] Reference: Chapters 7-8, Kutner's book.

which variable would be most helpful. Assume that a first-order multiple regression model is appropriate.

a. For each of the following variables, calculate the coefficient of partial determination given that $X1$ and $X2$ are included in the model: routine culturing ratio ($X3$), average daily census ($X4$), number of nurses ($X5$), and available facilities and services ($X6$).

b. On the basis of the results in part (a), which of the four additional predictor variables is best? Is the extra sum of squares associated with this variable larger than those for the other three variables?

c. Using the F^* test statistic, test whether or not the variable determined to be best in part (b) is helpful in the regression model when $X1$ and $X2$ are included in the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. Would the F^* test statistics for the other three potential predictor variables be as large as the one here? Discuss.

Problem 4: (P8.43, textbook) Refer to University admissions data set in Appendix C.4. The director of admissions at a state university wished to determine how accurately students' grade-point averages at the end of their freshman year (Y) can be predicted from the entrance examination (ACT) test score ($X2$); the high school class rank ($X1$, a percentile where 99 indicates student is at or near the top of his or her class and 1 indicates student is at or near the bottom of the class); and the academic year ($X3$). The academic year variable covers the years 1996 through 2000. Develop a prediction model for the director of admissions. Justify your choice of model. Assess your model's ability to predict and discuss its use as a tool for admissions decisions.

-- END --