Stats 500 Homework 8

Online Submission to Canvas, Due date: 11:59pm, November 14, 2025

1. (Exercise 4 in Chapter 10 of Linear Models with R, 2nd edition) Using the *trees* data, fit a model with *log(Volume)* as the response and a second-order polynomial (including the interaction term) in *Girth* and *Height*. Determine whether the model may be reasonably simplified.

2. Question 5.7 from **Applied Linear Regression book by Sanford Weisberg.** (see below for the question description).

5.7 Suppose X_1 were a continuous predictor, and F is a factor with three levels, represented by two dummy variables X_2 with values equal to 1 for the second level of F and X_3 with values equal to 1 for the third level of F. The response is Y. Consider three mean functions:

$$\mathrm{E}(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (5.19)$$

$$\mathrm{E}(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 \quad (5.20)$$

$$\mathrm{E}(Y|\mathbf{X}=\mathbf{x}) = \beta_0 + \beta_1 (x_1 - \delta) + \beta_{12} (x_1 - \delta) x_2 + \beta_{13} (x_1 - \delta) x_3 \quad (5.21)$$

Equation (5.21) includes an additional unknown parameter δ that may need to be estimated.

All of these mean functions specify that for a given level of F the plot of $\mathrm{E}(Y|X_1, F)$ is a straight line, but in each the slope and the intercept changes. For each of these three mean functions, determine the slope(s) and intercept(s), and on a plot of Y on the vertical axis and X_1 on the horizontal axis, sketch the three fitted lines.

The model (5.21) is a generalization of (5.20). Because of the extra parameter δ that multiplies some of the βs, this is a nonlinear model; see Saw (1966) for a discussion.

3. Question 5.17 from **Applied Linear Regression book by Sanford Weisberg.** (see below for the question description).

5.17 Sex discrimination (Data file: `salary`) The data file concerns salary and other characteristics of all faculty in a small Midwestern college collected in the early 1980s for presentation in legal proceedings for which discrimination against women in salary was at issue. All persons in the data hold tenured or tenure track positions; temporary faculty are not included. The variables include `degree`, a factor with levels PhD and MS; `rank`, a factor with levels Asst, Assoc, and Prof; `sex`, a factor with levels Male and Female; `Year`, years in current rank; `ysdeg`, years since highest degree, and `salary`, academic year salary in dollars.

5.17.1 Get appropriate graphical summaries of the data and discuss the graphs.

5.17.2 Test the hypothesis that the mean salary for men and women is the same. What alternative hypothesis do you think is appropriate?

5.17.3 Assuming no interactions between `sex` and the other predictors, obtain a 95% confidence interval for the difference in salary between males and females.

5.17.4 Finkelstein (1980), in a discussion of the use of regression in discrimination cases, wrote, "[a] variable may reflect a position or status bestowed by the employer, in which case if there is discrimination in the award of the position or status, the variable may be 'tainted.'" Thus, for example, if discrimination is at work in promotion of faculty to higher ranks, using rank to adjust salaries before comparing the sexes may not be acceptable to the courts.

Exclude the variable `rank`, refit, and summarize.

Data link: http://users.stat.umn.edu/~sandy/alr4ed/data/