Assignment 05

Regression Assumptions

Research. Teaching. Service. The trifecta upon which that almost every university instructor is evaluated, and, ultimately compensated. There has been substantial research to indicate that higher quality teaching is associated with increased faculty salararies. One way which academic administrators judge teaching quality is through teachers' course evaluations. While we know evaluation scores are not perfectly measures of teaching quality, nonetheless, they do play a role in the tenure and promotion process. Hamermesh & Parker (2005) examined whether, adjusting for other determinants, whether instructor attractiveness explain differences in course evaluation scores—and thus on earnings differences. In this assignment, you will use a sample of their data to examine this question.

Please submit your responses to each of the questions below in a printed document. Also, please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption and labels.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Equation Editor, Markdown, or LaTeX.

This assignment is worth 20 points. Each question is worth one point unless otherwise noted.

For this assignment, you will use the file *beauty.csv*. This file contains data collected from student evaluations of instructors' beauty and teaching quality for several courses at the University of Texas. The teaching evaluations were conducted at the end of the semester, and the beauty judgments were made later, by six students who had not attended the classes and were not aware of the course evaluations. The variables are:

- `prof`: Professor ID number
- `avgeval`: Average course rating
- `btystdave`: Measure of the professor's beauty composed of the average score on six standardized beauty ratings
- `tenured`: 0 = non-tenured; 1 = tenured
- `nonenglish`: 0 = native English speaker; 1 = non-native English speaker
- `age`: Professor's age (in years)
- `female`: 0 = male; 1 = female
- `students`: Number of students enrolled in the course
- `percentevaluating`: Percentage of enrolled students who completed an evaluation

These source of these data is: Hamermesh, D. S. & Parker, A. M. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*, 369–376. The data were made available by: Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Preparation

Fit the following regression model using R. You will use the output from the fitted model to answer the questions in the assignment.

- Model: `avgeval ~ btystdave + percentevaluating + students`

Preliminary Examination of Assumptions

1. Create the density plots for the outcome and each of the predictors (four total). **(3pts.)**
2. Do any of these distributions foreshadow problems for the normality assumption? Explain.
3. Create the scatterplot of the outcome vs. each of the predictors (three total). Include the fitted simple regression line in each plot. **(3pts.)**
4. For each of the three scatterplots describe the structural relationship between the outcome and predictor (e.g., linear, non-linear, etc.). Do any of these relationships foreshadow problems for the linearity assumption? Explain.

Examination of the Standardized Residuals from Each Simple Regression Model

5. Create the residual plots using the standardized residuals for each of the simple regression models (three total). **(3pts.)**
6. Do any of these plots foreshadow problems for the linearity assumption? Explain.
7. Do any of these plots foreshadow problems for the homogeneity of variance assumption? Explain.
8. Do any of these plots point to any extreme observations? If so, identify the observation(s) by marking them in the scatterplot in which they are problematic, and identifying the observation number (row in the data frame).

Examination of the Standardized Residuals from the Multiple Regression Model

9. Create both residual plots using the standardized residuals for the three-predictor multiple regression model (one total).
10. Does this plot suggest problems about meeting the linearity assumption? Explain.
11. Does this plot suggest problems problems about meeting the homogeneity of variance assumption? Explain.
12. Does this plot point to any extreme observations? If so, identify the observation(s) by marking them in the scatterplot in which they are problematic, and identifying the observation number (row in the data frame).

13. Create the density plot of the marginal distribution of the standardized residuals. Add the confidence envelope for the normal distribution.

14. Does this plot suggest problems about meeting the assumption of normality? Explain.