Assignment 04

Introduction to Multiple Regression

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 and 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 15 points. Each question is worth one point unless otherwise noted.

For this assignment, you will use the file *evaluations.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_id`: Professor ID number
- `avg_eval`: Average course rating
- `num_courses`: Number of courses for which the professor has evaluations
- `num_students`: Number of students enrolled in the professor's courses
- `perc_evaluating`: Average percentage of enrolled students who completed an evaluation
- `beauty`: Measure of the professor's beauty composed of the average score on six standardized beauty ratings
- `tenured`: Is the professor tenured? (0 = non-tenured; 1 = tenured)
- `native_english`: Is the professor a native English speaker? (0 = non-native English speaker; 1 = native English speaker)
- `age`: Professor's age (in years)
- `female`: Is the professor female? (0 = male; 1 = female)

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 models using R. You will use the output from the fitted models to answer the questions in the assignment.

- Model 1: `avg_eval ~ 1 + beauty`
- Model 2: `avg_eval ~ 1 + beauty + num_courses`
- Model 3: `avg_eval ~ 1 + beauty + num_courses + perc_evaluating`

Part I

1. Create a correlation table that indicates the pairwise correlations between the outcome and each of the predictors, and between each set of predictors. Model this table after Table 7.2 (p. 43) in the *Presenting Your Findings* book. Make sure the table you create also has an appropriate caption.

2. Examine the structure and formatting of Table 2 in the article: Garcia, D. R., McIlroy, L., & Barber, R. T. (2008). Starting behind: A comparative analysis of the academic standing of students entering charter schools. *Social Science Quarterly, 89*(1), 199–216. Notice that models are presented in columns (in the article 6 models are presented). Predictors used in the models are presented in rows, and so are the model-level summaries (e.g., N, F, R^2). Also note that the intercept ('Constant') is the last term presented in the table, generally because it is the least important coefficient. Blank cells indicate that the model does not include that particular predictor. Mimic the format and structure of this table to create a table to present the numerical information from the three models you fitted in this assignment. Re-create the formatting of Table 1 as closely as you can. Instead of giving the adjusted R^2 value, provide the unadjusted R^2 value. Make sure the table you create also has an appropriate caption. If the table is too wide, change the page orientation in your word processing program to "Landscape", rather than changing the size of the font. **(2pts.)**

3. Create a coefficient plot that graphically presents the coefficient estimates and the uncertainty (as 95% confidence intervals) for the coefficients included in the three fitted models. Present the three different models using different colors (or, if you are printing in greyscale, using differetn linetypes). Be sure that the figure is appropriately captioned. **(2pts.)**

4. Based on results presented in the regression table and from the coefficient plot, explain why it would be appropriate to adopt Model 2.

Use the results from Model 2 to answer the remainder of the questions on this assignment.

Part II

5. Report the regression equation from fitting Model 2. Use Equation Editor (or some other program that correctly types mathematical expressions) to typeset the equation correctly.

6. Using output from the ANOVA table, compute and report the value for the model R^2. Show your work for full credit.

7. Interpret the value of the model R^2 using the context of the data.

8. Using symbols, write the omnibus null hypothesis that is tested by the F-statistic in this analysis in two different manners: (1) using the coefficient parameters used in the regression model, and (2) using the variance accounted for parameter.

9. Based on the results of the F-test, does the model explain variation in course evaluations? Explain.

10. Interpret the estimated coefficient value associated with the beauty predictor.

11. Based on the results of the t-test, what do the data suggest about the partial effect of beauty on course evaluations? Explain.

12. Create a publication quality plot that displays the results from Model 2. For this plot, put the beauty predictor on the x-axis. Show two separate lines to show the effect of the number of courses (a small and large number of courses based on the data). The two lines should have different linetypes or use different colors so that they can easily be differentiated. **(2pts.)**