

Assignment 04

Introduction to Multiple Regression

The file *beauty.csv* 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. Use these data to answer each of the following questions. Each question is worth one point unless otherwise noted. The entire assignment is worth 10 points.

Fitting Models

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: `avgeval ~ btystdave`
- Model 2: `avgeval ~ btystdave + percentevaluating`

Multiple Regression

1. 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.
2. Compute and report the value for R^2 . Interpret the value of R^2 using the context of the data.
3. Using symbols, write the null hypothesis related to the *regression model coefficients* that is tested by the F -statistic in this analysis.
4. Using symbols, write the null hypothesis related to the *variance accounted for* that is tested by the F -statistic in this analysis.
5. Interpret the value for the estimate of the **btystdave** coefficient from Model 2.
6. Based on the results from Model 2, is there an effect of beauty on course rating? Explain.

Reporting Results from the Analyses

7. Create a table that reports the results for both of the regression models that were fitted in this assignment. This table should be similar in form to the tables of multiple regression results we created in class—with models in columns and predictors in rows. You might also want to consult the book *Presenting Your Findings*. Be sure to include the coefficients, SE, level of significance, and model summaries (RMSE and R^2) for each fitted model. Also, be sure that the table is appropriately captioned and labelled. If the table is too long, change the page orientation in your word processing program to “Landscape”, rather than changing the size of the font. **(2pts.)**
8. Create a publication quality plot that displays the results from Model 2. For this plot, put the beauty predictor on the x -axis. Control out the effects of the percentage of students evaluating the course by setting its value to the mean. Be sure to appropriately number, label, and re-size the plot for publication. **(2pts.)**