Assignment 06

Analysis of Covariance Using Multiple Regression

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.

Unadjusted Group Differences Model: ANOVA

Use the `lm()` function to fit a regression model with `avg_eval` as the outcome and `native_english` as the predictor. Use the `glance()` and `tidy()` functions to examine the output.

1. Write the fitted regression equation.
2. Interpret the intercept coefficient.
3. Interpret the slope coefficient.
4. In terms of means (not betas), write the null hypothesis (using mathematical notation) associated with the t-test of the slope? Be specific.
5. Based on results of the t-test for the slope, what do you conclude about differences in evaluation scores between native and non-native English speaking professors?
6. Use the fitted regression equation to estimate (a) the mean course rating for native and (b) the mean course rating for non-native English speakers. Show your work.

Adjusted Group Differences Model: ANCOVA

Now, suppose you want to examine differences between native and non-native English speakers, but this time you want to control for differences in beauty (`beauty`) and the number of courses (`num_courses`). Fit this model using `lm()` and use the `glance()` and `tidy()` functions to examine the output.

7. Write the fitted regression equation.
8. Interpret the fitted regression coefficient for `native_english`.
9. Compare the `native_english` coefficient and standard error in the adjusted model to those from the unadjusted model. How do they compare?
10. Write the fitted regression equation for native English speakers. (Note: This equation should only include the predictors `beauty` and `num_courses`.)
11. Write the fitted regression equation for non-native English speakers. (Note: This equation should only include the predictors `beauty` and `num_courses`.)
12. Compute the *adjusted* mean course rating for native and non-native English speakers (based on professors having an average beauty and an average number of courses). Show your work.
13. Create the density plot of the marginal distribution of the standardized residuals from the ANCOVA model. Add the confidence envelope for the normal distribution. Explain whether or not this plot suggests problems about meeting the normality assumption.
14. Create the scatterplot of the standardized residuals versus the fitted values from the ANCOVA model. In the plot identify observation with extreme residuals (≤ -3 or ≥ 3) by indicating the row number of that observation in the plot.
15. Explain whether or not this plot suggests problems about meeting the linearity and homogeneity of variance assumptions.