Assignment 04

Confidence and Prediction Intervals

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 12 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`

Prediction Intervals

1. Report the RMSE for Model 1.
2. Using the RMSE, estimate the width of the 95% prediction interval. Do not use the `predict()` function for this question. (Show your work.)
3. Consider a teacher that has a `btystdave` value of 0.75. Use the `predict()` function to estimate the 95% prediction interval.
4. Interpret the interval you computed in the previous question.
5. Consider a teacher that has a `btystdave` value of 0.75 and also has 80% of the enrolled students complete an evaluation. Use the `predict()` function to estimate the 95% prediction interval, but this time base the interval on the results from Model 2. Report the interval.
6. Explain why the prediction interval for Model 2 has less uncertainty (i.e., is narrower) than the prediction interval for Model 1.

Confidence Intervals

7. Using Model 1, compute an 80% confidence interval for teachers that have a `btystdave` value of 0.75. Report the interval.

8. Interpret the interval.

9. Now consider computing an 80% confidence interval for teachers that have a `btystdave` value of 0, again using Model 1. Why would this interval have less uncertainty (i.e., be narrower) than the confidence interval for teachers with a `btystdave` value of 0.75. Explain.

10. Create a publication quality plot (including title, figure numbering, etc.) that presents the fitted regression line for Model 2, the 95% confidence envelope based on Model 2, and the 95% confidence envelope based on Model 2. Be sure that the prediction envelope and the confidence envelope are drawn using different linetypes (or colors if you are printing in color). **(3pts.)**