

Assignment 04

Confidence and Prediction Intervals

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 L^AT_EX.

This assignment is worth 10 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 models using R. You will use the output from the fitted models to answer the questions in the assignment.

- Model 1: Use the beauty predictor to explain variation in course evaluation ratings;
- Model 2: Use the beauty predictor and the percentage of students completing an evaluation to explain variation in course evaluation ratings.

Model Uncertainty

1. Report the RMSE for both models.
2. Based on the RMSE values, which model has less uncertainty? Explain why this model would be expected to have less uncertainty.

Computing Prediction Intervals

3. Consider a teacher that has a `btystdave` value of 0.75. Use the `predict()` function to estimate the 95% prediction interval based on Model 1.
4. Interpret the interval you computed in the previous question.
5. Explain how the width of the 95% prediction interval is related to the RMSE value you reported for Model 1 in Question 1.
6. 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 compute the 95% prediction interval, but this time base the interval on the results from Model 2. Report the interval.

Computing Confidence Intervals

7. Using Model 2, compute and report an 80% confidence interval for teachers that have a `btystdave` value of 0.75 and have 80% of the enrolled students complete an evaluation.
8. Interpret the interval you computed in the previous question.
9. The 80% CI you computed in Question 8 has less uncertainty than the 80% PI you computed in Question 6. Explain (by referring to the sources of variation) why a PI should have more uncertainty than a CI.

Displaying Results

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% prediction 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).