Assignment 02

Bootstrapping

This goal of this assignment is to give you experience using bootstrapping in a regression analysis. Create an R Markdown document to respond to each of the questions below. Turn in a printed version of the knitted file. This assignment is worth 10 points. Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be set to an appropriate aspect ratio and sized in the knitted document so that they do not take up more room than necessary. They should also have an appropriate caption. Learn how to do this in a code chunk using knitr syntax.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Markdown's display equations. See here for some examples of how mathematics can be typeset in R Markdown.
- All syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden.

For each question, specify the question number using a level-2 (or smaller) header.

Part I: Fitting a Regression Model

Use the *homework-education-gpa.csv* data set to fit a regression model that uses time spent on homework (`homework`) and parent education level (`parentEd`) to explain variation in students' GPAs (`gpa`).

1. Indicate the value of the estimated R^2 statistic from the observed data.
2. Indicate the values for each of the estimated normal-theory coefficients and standard errors from the fitted regression model.

Part II: Bootstrapping Model-Level Statistics

Carry out a bootstrap analysis with 5,000 replications to re-sample R^2 based on the same multiple regression model you fitted in Part I.

3. Compute the confidence interval for R^2 using the percentile method.
4. Interpret the CI you computed in Question #4 using the context of the data.
5. Plot the distribution of the bootstrapped R^2 values. Annotate your plot to indicate the R^2 value from the observed data, and the lower and upper limits of the percentile interval.

Part III: Bootstrapping Coefficient-Level Statistics

Carry out a bootstrap analysis with 5,000 replications to re-sample all three regression coefficients based on the same multiple regression model you fitted in Part I.

6. Plot the bootstrap distribution for each of the three regression coefficients. Annotate each plot to indicate the appropriate observed coefficient value. In your document, all three plots should be printed side-by-side (in a single row). Be sure that the aspect ratio of this side-by-side plot is set appropriately.
7. Based on the three bootstrap distributions, will the add/subtract method of computing the CI give approximately the same limits as using the percentile method? Explain.
8. Compute and report the bootstrapped SEs for each of the three regression coefficients.
9. Compute and report the bootstrap-based CI for each of the three regression coefficients. Also report which method you used.
10. Using a completely nonparametric-based approach, compute and report the simulation-adjusted p-values for each of the three regression coefficients.