**Lab 9: Bootstrapping and Confidence Intervals**

**Instructions**

* Create a Quarto file called “Lab 9: Bootstrapping and Confidence Intervals”
* Copy the questions/prompts with the numbers/letters into the markdown file as text (i.e., in between code chunks, without any #). Use a header for each question #.
* Provide the code responses into code chunks directly beneath the questions (or beneath the text if the question requires both verbal and code answers).
* Submit both a knitted HTML or DOCX file and your .qmd file to ELMS before 11:59pm.
* *See ‘lab assignment demo’ file (.qmd) on ELMS or Jupyter for an example.* *Do not directly edit this file, instead create your own markdown file, copy the content from the demo and edit that.*

The data we are going to explore come from a survey of Indian children during the COVID-19 lockdowns. Responses cam from students of different ages at various educational institutions in the Delhi National Capital Region. The survey consisted of assessments of constructs like health, diet, and screen time.

Although there are more variables in the dataset, the ones we will be specifically looking at are:

* age – The age of the student in years
* class\_hours – The number of hours the child spent in class
* class\_rating – The overall rating of the class by the student (higher scores = more favorable impression)
* health\_problems – Whether the child experienced health issues during lockdown

The ultimate question we’re interested in is whether class hours impacted class rating controlling for health problems and student age.

**Question 1**

Create a scatterplot visualizing the relationship between class\_hours and class\_rating. What do you see? Do you suspect there is a relationship?

**Question 2**

Fit a regression model with class\_rating as the outcome and class\_hours, health\_problems, and age as predictors. Is the effect of class\_hours significant at the level? Report the results in APA format and interpret the coefficient and *p*-value for the effect of class\_hours. Do you reject or fail-to-reject the null?

*Bonus Challenge:* Manually calculate this *p* -value.

**Question 3**

Obtain the 99% confidence interval for the effect of class\_hours and interpret the confidence interval. Do you reject or fail-to-reject the null? Does this agree with your decision from the *p*-value? Why?

*Bonus Challenge:* Manually calculate this confidence interval.

**Question 4**

Construct a 99% bootstrap confidence interval for this effect. Set the seed to 1001 before creating the confidence interval and use 10,000 resamples. Interpret the confidence interval. Do you reject or fail-to-reject the null? Does this “agree” with the decision from the *p*-value and normal theory confidence intervals from earlier? Is this always the case?

**Question 5**

Visualize the empirical bootstrap distribution and plot the endpoints of the confidence interval on it.