Exercises 4 · Quantifying uncertainty

(1) Bootstrapping practice

(A) Return to the data set "ut2000.csv" on SAT scores from UT students across all 10 undergraduate colleges. Calculate an approximate 95% confidence interval for the difference in mean SAT math (SAT.Q) scores between students in the colleges of architecture and liberal arts. I can think of at least two ways you could accomplish this, so make sure you describe precisely what you did and why, and report the interval.

(B) Fit a regression model for graduating GPA in terms of SAT combined score (SAT.C) and College (with no interaction term), and provide a 95% confidence interval for the slope of the SAT score.

(C) In your own words, briefly describe the idea of bootstrapping (both what we do and why we do it). Think of this as like a practice essay question for the midterm; if you want to replicate midterm conditions, set yourself a timer (e.g. 30 minutes), don't look at your notes while answering, and write an answer by hand. Then revisit your answer by comparing with the course notes, and improve it to learn from your inaccuracies.

(2) Gas prices in Austin

Visit the case study on gas prices in Austin at `https://github.com/jgscott/learnR/blob/master/cases/gasprices/gasprices.md`. Use the data set provided to assess the evidence for six different "pet theories" about what explains variation in gas prices from one gas station to another. Turn in a concise write-up summarizing your conclusions and the evidence used to support them.