CHAPTER OVERVIEW

11: Simulating Data

You may have noticed that throughout this book so far we have analyzed a lot of fake data. We used R to simulate pretend numbers, and then we analyzed those numbers. We also, from time to time, loaded in some “real” data, and analyzed that. In your labs each week, you have been analyzing a lot of real data. You might be thinking that the simulations we ran were just for educational purposes, to show you how things work. That’s partly true, that’s one reason we ran so many simulations. At the same time, conducting simulations to understand how data behaves is a legitimate branch of statistics. There are some problems out there where we don’t have really good analytic math formulas to tell us the correct answer, so we create and run simulations to approximate the answer.

I’m going to say something mildy controversial right now: If you can’t simulate your data, then you probably don’t really understand your data or how to analyze it. Perhaps, this is too bold of a statement. There are many researchers out there who have never simulated their data, and it might be too much too claim that they don’t really understand their data because they didn’t simulate. Perhaps. There are also many students who have taken statistics classes, and learned how to press some buttons, or copy some code, to analyze some real data; but, who never learned how to run simulations. Perhaps my statement applies more to those students, who I believe would benefit greatly from learning some simulation tricks.
