

# Preface

The path to this book began, as these things so often do, when I was asked to teach a course. In this case, a semester-long class for master's students on epidemiologic analyses using SAS. Over a few years of preparing for and teaching the material I confronted a combination of practical and conceptual considerations that led me to believe that perhaps there was room for another book about SAS.

On a practical level, working with a program like SAS is a skill I consider necessary for all graduating master's epidemiologists. To be honest, the necessity that the program actually be SAS is based on a circular argument. Many employers of epidemiologists use SAS because their current analysts use SAS, and newly minted analysts will compel additional future analysts to use SAS. This reliance on SAS of potential employers of master's-level epidemiology students may change in the future, but my sense is that it will not be anytime soon. While the practical motivation to learn SAS is somewhat self-fulfilling, it does not detract from the capabilities that made SAS an important skill in the first place. And, does it make the choice of SAS any less necessary. As I sit and write this, a quick search on the *New York Times* jobs link returns 15 epidemiology jobs in the New York City area. A search for SAS returns 457 hits. When I do this search on the first day of class, with generally the same results, there is invariably an increased interest among the students in spending a few hours a week learning SAS.

The kinds of SAS-related work that master's-level epidemiologists are called upon to undertake do not exceed some fairly straightforward categorical and continuous data analyses. There was, though, no book that addressed this material in a similarly straightforward fashion. The feedback I've received from the past students is that the procedures covered in this material account for a good majority of their daily activity and that knowing how to do those things helps set the stage for learning more advanced material.

On a conceptual level, the role of statistical software in epidemiologic practice is in a state of flux, and the kinds of data and analyses epidemiologists are being called upon to work with are evolving into what might be called the era of "big data" and the rise of "computational epidemiology." SAS is tailor-made to deal with the kinds of huge data sets that are becoming routine in epidemiology. That there has been

an explosion in the availability of administrative and routinely collected health data, free and open-source data, social media data, and other online data is clear. That the data are amenable to reliable or valid analyses is less clear. The basics, about missing and incorrect values, about confounding, about bias, about study design, are if anything even more important. The data can inform, but we may have to teach them to speak clearly, and in a language that is epidemiologically valid.

Fortunately, SAS is more than up to the task. It has a facility for dealing with extremely large data sets that I have found unsurpassed in other statistical programs. SAS allows epidemiologists to pay special attention to the necessary (though not glamorous) initial steps of reading in, preparing, and cleaning large amounts of data, when early errors or missteps will be amplified throughout the analysis, sometimes in ways that are difficult to trace to their origins. For this reason, fully the first third of this book addresses using SAS to read in and manipulate data to get them into a form that makes epidemiologic sense.

The one aspect of preparing and teaching this material that I did not expect was that it was actually fun. I'm certain this says more about me than it does about the material. But perhaps a kind of geeky enjoyment of some of the practical aspects of epidemiologic methods, like learning how to use SAS, is a sign that you've chosen the right profession. I tried to capture some of what I found interesting and enjoyable by using examples and materials that have practical relevance to epidemiologic practice.

I have come to appreciate that public health practice requires a long-term view, and that you may not always (or even frequently) see the effects of your work. The effects of teaching public health are even farther removed from immediacy. Despite the practical aspects underlying this book, the ultimate motivation is as ephemeral as public health practice itself. In the end, I hope to contribute in some small way to the efforts of someone I haven't met, to improving the health, happiness, and well-being of someone who may not even be born yet.

New York, NY, USA

Charles DiMaggio



<http://www.springer.com/978-1-4614-4853-2>

SAS for Epidemiologists  
Applications and Methods  
DiMaggio, C.  
2013, XVII, 258 p., Hardcover  
ISBN: 978-1-4614-4853-2