Preface

This book has two objectives. The first is to teach how to fit a multivariable statistical model to cross-sectional safety data using a simple spreadsheet. The second is to promote the understanding that is at the core of good modeling.

These twin objectives determine the flow and the structure of the book. After some preliminaries, a real data set is introduced. From here on and all the way to the last chapter, the same data are used to gradually build up a regression model. Along the way there are the "what and how" sections: what is an exploratory data analysis, how to use pivot tables, what is a curve-fitting spreadsheet and how to build one, how to use the Solver for parameter estimation, how to examine the quality of a fit by a CURE plot, what functions look like, etc. These are in support of the first objective, that of teaching how to fit a model to data. Interspersed between these are sections of a reflective nature. These support the "understanding" objective and speak about the "why, when, and whether" issues of modeling: why do we need to curve-fit, whether models be used in a cause–effect manner, when should a variable be added to the model equation, why it is important to know what function hides behind the data, etc.

Data about accidents, about the road, and about the traffic on it are routinely collected and maintained in databases. We know where reported accidents occurred and much about their circumstances. We also have information about many features of the road (grade, curve radius, lane width, speed limit, parking control, etc.) and of its traffic (daily volumes, percent of trucks, etc.). Inasmuch as these data pertain to what is observed to exist on a cross section of "units" (road segments, intersections, etc.) they are called "observational cross-sectional" data.

To be of use for evidence-based road safety management, the data need to be summarized and cast into the form of statistical models. Models serve three purposes:

1. To estimate how (un)safe are certain road segments, intersections, ramps, crossings, etc. thereby determining the size of the road safety problem that could perhaps be altered by interventions and design changes.

2. To estimate by how much safety has changed following an intervention or a design change.
3. To estimate by how much safety might be altered due to a change in some variable value.

The book focuses on the development of regression models for purposes 1 and 2. While the use of regression models for purpose 3 is commonplace, the trustworthiness of the result is in doubt. Still, it is possible that even this vexed purpose will be well served by the approach to modeling that is advocated here.

Who will use this book? Perhaps they will be graduate students with interest in road safety, perhaps professionals with responsibilities in data analysis, or perhaps others. I do not know how much math, probability, and statistics I can rely on. Some such knowledge is required by the very nature of the subject matter. There will be parts which some readers may find taxing. The hope is that judicious skipping will make the book accessible and useful for a variety of audiences. I tried to make the narration succinct; diversions, elaborations, and detail were relegated to footnotes. A glossary is provided to serve as a refresher of notation and acronyms; the index at the end is for finding the page on which a topic or concept is first introduced.

The book evolved from lecture notes used in a series of hands-on workshops and is richly laced with data-based illustrations, tables, and figures. The supporting materials are available for downloading. To access and download them, go to http://extras.springer.com/ and enter the ISBN of this book. The ISBN (International Standard Book Number) is found just after the title page. The information is in five folders: Data, PowerPoint presentations, Problems, Solutions, and Spreadsheets. Together with the book these materials will be of interest to the reader, student, and instructor.

The book, of course, can be read as is. However, unlike textbooks in the past, it can also be used actively and creatively. The reader is invited to use the downloadable materials to see how results were obtained; to modify, expand, and enrich the analysis; and to use the same spreadsheets with other data.

Modeling in this book is built around the use of Excel spreadsheets and readers are assumed to have facility in their use. Information about less commonly known spreadsheet functionalities is provided where needed. Of course there are specialized and sophisticated statistical software packages which, once acquired and mastered, will do a good, perhaps a superior, job of model development. However, I find that the spreadsheet provides all the essentials; it makes for intimate contact with the data, it has adequate and flexible visualization, the "pivot tables" serve for exploration, an optimization tool does the curve-fitting, and it is a hospitable environment for writing custom pieces of code.

Model development is often presented as if it was a nearly algorithmic sequence of steps, an ordered progression of activities from "Start" to "End." In my opinion such an approach tends to produce inferior results. It is better to think of model development as detective work with clues embedded in data. Like in a game of snakes and ladders, there are advances and setbacks whereby the modeler gradually moves towards a satisfactory outcome. Such work is well served by the atmosphere

of a spreadsheet. For all these reasons, the spreadsheet is my environment of choice for both instruction and creative modeling.

The modeling approach described in these pages may be thought old-fashioned, perhaps unsophisticated. Emphasis is on what is of essence. Papers describing novel modeling ideas and newfangled statistical techniques are being published daily and I make no attempt to capture the latest. In defense I only say that the quality of a meal depends more on the skill of the cook and the time spent on its preparation than on the modernity of the food processor. As has been said: "... second-rate minds grappling with first-rate problems can teach you more than first-rate minds lost in the shrubbery." (Lilla, 2013, p. xii).

Toronto, ON, Canada Ezra Hauer

Reference

Lilla M in Foreword for Berlin I (2013) Against the current, 2nd ed. Princeton University Press, Princeton

http://www.springer.com/978-3-319-12528-2

The Art of Regression Modeling in Road Safety

Hauer, E.

2015, XVII, 233 p. 116 illus., 112 illus. in color.,
Hardcover
ISBN: 978-3-319-12528-2