Regression Graphics: Ideas for Studying Regression Through Graphics.

R. Dennis COOK. New York: Wiley, 1998. ISBN 0-471-19365-8. xviii + 349 pp. $79.95 (H).

Statistical plots have many uses in regression analysis. They can suggest which models to fit and are powerful in detecting model discrepancies. Yet despite such well-recognized contributions, most still consider the role of graphics in regression supplementary. This is quite evident by noting the relatively insignificant space devoted to the discussion of graphics in most regression books. Clearly, there are so many different patterns that plots can show. So how can one wish for a systematic way of telling others what to do with them? Well, reading this ground-breaking book by Dennis Cook ought to shatter one's preconceptions about regression graphics. Nowadays, high-quality graphics are so easy to produce by modern computer. The important issues lagging behind are what to plot, what to learn from plotting, and what to do afterward. The reader will find definitive answers in *Regression Graphics*.

The subtitle of the book has already told us much of what to anticipate—ideas for studying regressions through graphics. The author's intention is clear—he wants graphics to take the leading rather than the supplementary role. Philosophically, this change of viewpoint is already remarkable. But the real accomplishment is that the ideas as projected in the book are not just a collection of empirical wisdom. They are indeed fully backed by statistical properties that can be mathematically proved.

In contrast to most regression books, which discuss one regressor variable in great length, this book opens the first chapter with "many" predictor variables. The fact is that many good ideas behind cases with only one predictor simply do not work in higher dimensions. Recognizing this, Cook brings out the theme of *dimension reduction* to weave together the various nice ideas he offers.

The book's main context is quite new; most results are developed only after the late 1980s. Quoting from the Preface, "this new context is intended to blend with rather than replace more traditional paradigms for regression analysis... It imposes few scope-limiting restrictions on the nature of the regression" (p. xvi). In such carefully phrased terms, Cook keeps a good distance from the context of nonparametric regression. Although the framework is intrinsically nonparametric and some smoothing ideas are used, no elaboration on the rate or constant of convergence is involved. In fact, throughout most of the book, curves are fitted only to help interpret the data.

The book is comprised of 15 chapters. The first three chapters set the stage for later developments. Modern graphical facilities and their dynamic features, including scatterplot matrices, three-dimensional (3-D) plots, slicing, brushing, linking, rotation, rescaling, and highlighting, are discussed. Notational conventions are introduced. Although introductory, these chapters contain plenty of interesting materials to gain the reader's immediate interest. For example, via a cleverly constructed example, Cook is able to illustrate simultaneously two types of pitfalls in interpreting residual plots after multiple linear regression: failing to indicate model deficiency when there is nonlinearity in one variable, and falsely attributing the model failure to a variable that is not needed. There is an illuminating historical account of the works of Mordecai Ezekiel and Louis H. Bean in the 1920s. It is really amazing to see how these pioneers, little-known to the modern nonparametric regression world, could actually succeed in fitting data with the so-called additive models of the 1980s (models taking the form of $E(y|x_1, x_2, x_3) = g_1(x_1) + g_2(x_2) + g_3(x_3)$, for example) without the aid of a modern computer. I also find the notation of conditional distribution of a random variable y given a random variable x, and the notation "$u \perp\!\!\!\perp v|z$" for the conditional independence of u and v given z, extremely handy in describing the more complex situations developed in later chapters.

The concept of dimension reduction begins with Chapter 4 when the discussion turns to the interpretation of 3-D scatterplots. This is motivated by a fundamental question: "Can one replace a 3-D plot with a 2-D plot without loss of information?" Conditional independence is naturally introduced for formalizing the problem. By going through several real datasets step-by-step, Cook shows how to use dynamic graphics to answer this question. In Chapter 5 the discussion is devoted to binary output response. This is more challenging, because the scatterplot of y on x would not be very informative, and so other devices must be introduced.

The central theme of dimension reduction receives a formal treatment in Chapter 6, where the key notion of dimension-reduction space is introduced. This is a thoughtful extension of the so-called effective dimension (edr) directions that first appeared in the literature of sliced inverse regression (SIR) and principal Hessian directions (PHD), which deals with the general regression model of the form $y = f(\beta_1' x, \ldots, \beta_k' x, \varepsilon)$ with unknown functional form f and unknown distribution of the error ε. The goal is to estimate the β vectors so that the relationship between y and a p-dimensional predictor x can be described through the relationship between y and k variables $\beta_1' x, \ldots, \beta_k' x$ without loss of information. The goal of dimension reduction is fulfilled when k is smaller than p. In Chapters 4–12, Cook covers most of the advances in this area since 1989, including works by Cook himself and his long-term collaborator Sanford Weisberg and several colleagues/students, by me and my collaborators, and by many other statisticians.

Under the somewhat mysterious title "Graphical Regression," Chapter 7 deals with the question of whether or not one can piece together information from plots of y against selected coordinate variates. Chapter 8 discusses how standard methods like multiple linear regression can be used to find a dimension-reduction space. Chapter 9 provides an opportunity to see how the new ideas introduced so far can make a difference. This is done by reanalyzing three datasets that had been well studied by traditional methods in the literature. The idea of using inverse regression to aid forward regression is introduced in Chapter 10, leading eventually to the discussion of the SIR dimension-reduction method in Chapter 11. Chapter 12 is devoted to an in-depth account of the PHD dimension-reduction method.

How are the new developments related to the traditional models? Chapter 13 studies the issue of how to assess the contribution of a selected predictor to a regression. *Net-effect plots* are introduced as natural extensions of the classical added-variable plots (also called adjusted-variable plots). Chapter 14 explores graphical methods for studying predictor transformations. Chapter 15 suggests graphical methods for assessing model adequacy.

This is an ideal graduate-level textbook. Because most materials are built up from scratch, readers would not need much statistical background beyond an undergraduate level of exposure to regression analysis. Each chapter includes a few exercise problems for further exploration of ideas and methodology. The author also provides a web site for easy access to computer programs, data, and other supplemental information.

The appearance of *Regression Graphics* signifies the emergence of a new era in regression. This comes in time to meet new challenges posed by the abundance of large-dimensional data. Thus, although the presentation itself is aimed primarily at statisticians, data miners might find ideas in this book especially relevant. Because it is fair to say that regression is one of the most popular techniques in statistics and that graphics is one of the most elementary tools in data analysis, I strongly recommend this book to all JASA readers—students, instructors, researchers, and practitioners. Whether or not you are a statistician, this truly wonderful book should have an impact on how you think of regression and graphics—and thus on how you think of statistics.

Ker-Chau LI
University of California-Los Angeles

Information Graphics: A Comprehensive Illustrated Reference: Visual Tools for Analyzing, Managing, and Communicating.

Robert L. HARRIS. Atlanta: Management Graphics, 1996. ISBN 0-9646925-0-3. 448 pp. $14.95.

Statistics is an extraordinary discipline. People with no training or qualifications attempt to teach or write with authority on the subject. The results are usually statistically dismal. This book is another example of this phenomenon. Subtitled *A Comprehensive Illustrated Reference*, this self-published work aspires to be the standard reference for graphics that convey information. In its large format and pages filled with many graphics, this book evokes other well-known large-format privately published comprehensive graphics volumes. But Harris is not Tufte.