

## PREFACE

---

I have been feeling a great sense of sadness while I was working alone on this edition of the book after Professor Samprit Chatterjee, my longtime teacher, mentor, friend, and co-author, passed away in April 2021. Our first paper was published in 1986 (Chatterjee and Hadi, 1986). Samprit and I also co-authored our 1988 book (Chatterjee and Hadi, 1988) as well as several other papers. My sincere condolences to his family and friends. May God rest his soul in peace.

Regression analysis has become one of the most widely used statistical tools for analyzing multifactor data. It is appealing because it provides a conceptually simple method for investigating functional relationships among variables. The standard approach in regression analysis is to take data, fit a model, and then evaluate the fit using statistics such as  $t$ ,  $F$ ,  $R^2$ , and Durbin–Watson test. Our approach is broader. We view regression analysis as a set of data analytic techniques that examine the interrelationships among a given set of variables. The emphasis is not on formal statistical tests and probability calculations. We argue for an informal analysis directed toward uncovering patterns in the data. We have also attempted to write a book for a group of readers with diverse backgrounds. We have also tried to put emphasis on the art of data analysis rather than on the development of statistical theory.

The material presented is intended for anyone who is involved in analyzing data. The book should be helpful to those who have some knowledge of the basic concepts of statistics. In the university, it could be used as a text

for a course on regression analysis for students whose specialization is not statistics, but, who nevertheless use regression analysis quite extensively in their work. For students whose major emphasis is statistics, and who take a course on regression analysis from a book at the level of Rao (1973), Seber (1977), or Sen and Srivastava (1990), this book can be used to balance and complement the theoretical aspects of the subject with practical applications. Outside the university, this book can be profitably used by those people whose present approach to analyzing multifactor data consists of looking at standard computer output ( $t$ ,  $F$ ,  $R^2$ , standard errors, etc.), but who want to go beyond these summaries for a more thorough analysis.

We utilize most standard and some not-so-standard summary statistics on the basis of their intuitive appeal. We rely heavily on graphical representations of the data and employ many variations of plots of regression residuals. We are not overly concerned with precise probability evaluations. Graphical methods for exploring residuals can suggest model deficiencies or point to troublesome observations. Upon further investigation into their origin, the troublesome observations often turn out to be more informative than the well-behaved observations. We notice often that more information is obtained from a quick examination of a plot of residuals than from a formal test of statistical significance of some limited null hypothesis. In short, the presentation in the chapters of this book is guided by the principles and concepts of exploratory data analysis.

As we mentioned in previous editions, the statistical community has been most supportive, and we have benefitted greatly from their suggestions in improving the text. Our presentation of the various concepts and techniques of regression analysis relies on carefully developed examples. In each example, we have isolated one or two techniques and discussed them in some detail. The data were chosen to highlight the techniques being presented. Although when analyzing a given set of data it is usually necessary to employ many techniques, we have tried to choose the various data sets so that it would not be necessary to discuss the same technique more than once. Our hope is that after working through the book, the reader will be ready and able to analyze their data methodically, thoroughly, and confidently.

The emphasis in this book is on the analysis of data rather than on plugging numbers into formulas, tests of hypotheses, or confidence intervals. Therefore no attempt has been made to derive the techniques. Techniques are described, the required assumptions are given and, finally, the success of the technique in the particular example is assessed. Although derivations of the techniques are not included, we have tried to refer the reader in each case to sources in which such discussion is available. Our hope is that

some of these sources will be followed up by the reader who wants a more thorough grounding in theory.

Recently there has been a qualitative change in the analysis of linear models, from model fitting to model building, from overall tests to clinical examinations of data, from macroscopic to the microscopic analysis. To do this kind of analysis a computer is essential and, in previous editions, we have assumed its availability, but did not wish to endorse or associate the book with any of the commercially available statistical packages to make it available to a wider community.

We are particularly heartened by the arrival of the language R, which is available on the Internet under the General Public License (GPL). The language has excellent computing and graphical features. It is also free! For these and other reasons, I decided to introduce and use R in this edition of the book to enable the readers to use R on their own datasets and reproduce the various types of graphs and analysis presented in this book. Although a knowledge of R would certainly be helpful, no prior knowledge of R is assumed.

Major changes have been made in streamlining the text, removing ambiguities, and correcting errors pointed out by readers and others detected by the authors. Chapter 2 is new in this edition. It gives a brief but, what we believe to be, sufficient introduction to R that would enable readers to use R to carry out the regression analysis computations as well as the graphical displays presented in this edition of the book. To help the readers out, we provide all the necessary R code in the new Chapter 2 and throughout the rest of the chapters. Section 5.11 about regression diagnostics in R is new. New references have also been added. The index at the end of the book has been enhanced. The addition of the new chapter increased the number of pages. To offset this increase, data tables that are larger than 10 rows are deleted from the book because the reader can obtain them in digital forms from the Book's Website at <http://www.aucegypt.edu/faculty/hadi/RABE6>. This Website contains, among other things, all the data sets that are included in this book, the R code that are used to produce the graphs and tables in this book, and more. Also, the use of R enabled us to delete the statistical tables in the appendix because the reader can now use R to compute the  $p$ -values as well as the critical values of test statistics for any desired significance level, not just the customary ones such as 0.1, 0.05, and 0.01.

We have rewritten some of the exercises and added new ones at the end of the chapters. We feel that the exercises reinforce the understanding of the material in the preceding chapters. Also new to accompany this edition a

Solution Manual and Power Point files are available only for instructors by contacting the authors at [ahadi@aucegypt.edu](mailto:ahadi@aucegypt.edu) or [ali-hadi@cornell.edu](mailto:ali-hadi@cornell.edu).

Previous editions of this book have been translated to Persian, Korean, and Chinese languages. We are grateful to the translators Prof. H. A. Niromand, Prof. Zhongguo Zheng, Prof. Kee Young Kim, Prof. Myoungshic Jhun, Prof. Hyuncheol Kang, and Prof. Seong Keon Lee. We are fortunate to have had assistance and encouragement from several friends, colleagues, and associates. Some of our colleagues and students at New York University, Cornell University, and The American University in Cairo have used portions of the material in their courses and have shared with us their comments and comments of their students. Special thanks are due to our friend and former colleague Jeffrey Simonoff (New York University) for comments, suggestions, and general help. The students in our classes on regression analysis have all contributed by asking penetrating questions and demanding meaningful and understandable answers. Our special thanks go to Nedret Billor (Cukurova University, Turkey) and Sahar El-Sheneity (Cornell University) for their very careful reading of an earlier edition of this book.

We also appreciate the comments provided by Habibollah Esmaily, Hassan Doosti, Fengkai Yang, Mamunur Rashid, Saeed Hajebi, Zheng Zhongguo, Robert W. Hayden, Marie Duggan, Sungho Lee, Hock Lin (Andy) Tai, and Junchang Ju. We also thank Lamia Abdellatif for proofreading parts of this edition, Dimple Philip for preparing the Latex style files and the corresponding PDF version, Dean Gonzalez for helping with the production of some of the figures, and Michael New for helping with the front and back covers.

*Cairo, Egypt*  
*September 2023*

ALI S. HADI