

Preface



ONE OF Dr. Good's very first statistical applications was an analysis of leukemia cases in Hiroshima, Japan after World War II; on August 7, 1945 this city was the target site of the first atomic bomb dropped by the United States. Was the high incidence of leukemia cases among survivors the result of exposure to radiation from the atomic bomb? Was there a relationship between the number of leukemia cases and the number of survivors at certain distances from the atomic bomb's epicenter?

To assist in the analysis, Dr. Good had an electric (not an electronic) calculator, reams of paper on which to write down intermediate results, and a prepublication copy of Scheffé's *Analysis of Variance*. The work took several months, and the results were somewhat inclusive, mainly because he could never seem to get the same answer twice—a consequence of errors in transcription rather than the absence of any actual relationship between radiation and leukemia.

Today, of course, we have high-speed computers and prepackaged statistical routines to perform necessary calculations. Yet access to statistical software will no more make one a statistician, than access to a chainsaw will make one a lumberjack. Allowing these tools to do our thinking for us is a sure recipe for disaster—just ask any emergency room physician.

Pressed by management or by funding needs, too many research workers have no choice but to go forward with data analysis regardless of the extent of their statistical training. Alas, although a semester or two of undergraduate statistics may suffice to develop familiarity with the names of some statistical methods, it is not enough to ensure awareness of all the circumstances under which these methods may be applicable.

The purpose of the present text is to provide a mathematically rigorous but readily understandable foundation for statistical procedures. Here for the second time are such basic statistical concepts as null and alternative hypotheses, p -value, significance level, and power. Reprints from the statis-

tical literature provide illustration as we reexamine sample selection, linear regression, the analysis of variance, maximum likelihood, Bayes' theorem, meta-analysis and the bootstrap.

For the second edition, we've added material from online courses we offer at statistics.com. This new material is devoted to unbalanced designs, report interpretation, and alternative modeling methods.

More good news for technophobes: Dr. Good's articles on women's sports have appeared in the *San Francisco Examiner*, *Sports Now*, and *Volleyball Monthly*. Twenty-two of his short stories are also in print. If you can read the sports page, you'll find the presentation of material in this text easy to read and to follow. Lest the statisticians among you believe this book is too introductory, we point out the existence of hundreds of citations in statistical literature calling for the comprehensive treatment we have provided. Regardless of past training or current specialization, this book will serve as a useful reference; you will find applications for the information contained herein whether you are a practicing statistician or a well-trained scientist who just happens to apply statistics in the pursuit of other science.

The primary objective of the opening chapter is to describe the main sources of error and provide a preliminary prescription for avoiding them. The hypothesis formulation—data gathering—hypothesis testing and estimate cycle is introduced, and the rationale for gathering additional data before attempting to test after-the-fact hypotheses is detailed.

Chapter 2 places our work in the context of decision theory. We emphasize the importance of providing an interpretation of each and every potential outcome in advance of consideration of actual data.

Chapter 3 focuses on study design and data collection, for failure at the planning stage can render all further efforts valueless. The work of Berger and his colleagues on selection bias is given particular emphasis.

Desirable features of point and interval estimates are detailed in Chapter 4 along with procedures for deriving estimates in a variety of practical situations. This chapter also serves to debunk several myths surrounding estimation procedures.

Chapter 5 reexamines the assumptions underlying testing hypotheses. We review the impacts of violations of assumptions and detail the procedures to follow when making 2- and k -sample comparisons. In addition, we cover the procedures for analyzing contingency tables and 2-way experimental designs if standard assumptions are violated.

Chapter 6 is devoted to the value and limitations of Bayes' theorem, meta-analysis, and resampling methods.

Chapter 7 lists the essentials of any report that will utilize statistics, debunks the myth of the "standard" error, and describes the value and limitations of p -values and confidence intervals for reporting results. Prac-

tical significance is distinguished from statistical significance and induction is distinguished from deduction. Chapter 8 covers much the same material, but the viewpoint is that of the report reader rather than the report writer. Of particular importance is a section on interpreting computer output.

Twelve rules for more effective graphic presentations are given in Chapter 9 along with numerous examples of the right and wrong ways to maintain reader interest while communicating essential statistical information.

Chapters 10 through 13 are devoted to model building and to the assumptions and limitations of a multitude of regression methods and data mining techniques. A distinction is drawn between goodness of fit and prediction, and the importance of model validation is emphasized. Seminal articles by David Freedman and Gail Gong are reprinted.

Finally, for the further convenience of readers, we provide a glossary grouped by related but contrasting terms, an annotated bibliography, and subject and author indexes.

Our thanks to William Anderson, Leonardo Auslender, Vance Berger, Peter Bruce, Bernard Choi, Tony DuSoir, Cliff Lunneborg, Mona Hardin, Gunter Hartel, Fortunato Pesarin, Henrik Schmiediche, Marjorie Stinespring, and Peter A. Wright for their critical reviews of portions of this text. Doug Altman, Mark Hearnden, Elaine Hand, and David Parkhurst gave us a running start with their bibliographies. Brian Cade, David Rhodes, and, once again, Cliff Lunneborg helped us complete the second edition.

We hope you soon put this text to practical use.

Sincerely yours,

Phillip Good

brother_unknown@yahoo.com
Huntington Beach CA.

James Hardin

jhardin@gwm.sc.edu
Columbia, SC.
July 2003/2005

