
COMMON ERRORS IN STATISTICS (AND HOW TO AVOID THEM)

Phillip I. Good
James W. Hardin

 **WILEY-INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

Contents

Preface ix

PART I FOUNDATIONS 1

1. Sources of Error 3
 Prescription 4
 Fundamental Concepts 4
 Ad Hoc, Post Hoc Hypotheses 7

2. Hypotheses: The Why of Your Research 11
 Prescription 11
 What Is a Hypothesis? 11
 Null Hypothesis 14
 Neyman–Pearson Theory 15
 Deduction and Induction 19
 Losses 20
 Decisions 21
 To Learn More 23

3. Collecting Data 25
 Preparation 25
 Measuring Devices 26
 Determining Sample Size 28
 Fundamental Assumptions 32
 Experimental Design 33
 Four Guidelines 34
 To Learn More 37

PART II HYPOTHESIS TESTING AND ESTIMATION	39
4. Estimation	41
Prevention	41
Desirable and Not-So-Desirable Estimators	41
Interval Estimates	45
Improved Results	49
Summary	50
To Learn More	50
5. Testing Hypotheses: Choosing a Test Statistic	51
Comparing Means of Two Populations	53
Comparing Variances	60
Comparing the Means of K Samples	62
Higher-Order Experimental Designs	65
Contingency Tables	70
Inferior Tests	71
Multiple Tests	72
Before You Draw Conclusions	72
Summary	74
To Learn More	74
6. Strengths and Limitations of Some Miscellaneous Statistical Procedures	77
Bootstrap	78
Bayesian Methodology	79
Meta-Analysis	87
Permutation Tests	89
To Learn More	90
7. Reporting Your Results	91
Fundamentals	91
Tables	94
Standard Error	95
p Values	100
Confidence Intervals	101
Recognizing and Reporting Biases	102
Reporting Power	104
Drawing Conclusions	104
Summary	105
To Learn More	105
8. Graphics	107
The Soccer Data	107
Five Rules for Avoiding Bad Graphics	108

One Rule for Correct Usage of Three-Dimensional Graphics	115
One Rule for the Misunderstood Pie Chart	117
Three Rules for Effective Display of Subgroup Information	118
Two Rules for Text Elements in Graphics	121
Multidimensional Displays	123
Choosing Effective Display Elements	123
Choosing Graphical Displays	124
Summary	124
To Learn More	125
PART III BUILDING A MODEL	127
9. Univariate Regression	129
Model Selection	129
Estimating Coefficients	137
Further Considerations	138
Summary	142
To Learn More	143
10. Multivariable Regression	145
Generalized Linear Models	146
Reporting Your Results	149
A Conjecture	152
Building a Successful Model	152
To Learn More	153
11. Validation	155
Methods of Validation	156
Measures of Predictive Success	159
Long-Term Stability	161
To Learn More	162
Appendix A	163
Appendix B	173
Glossary, Grouped by Related but Distinct Terms	187
Bibliography	191
Author Index	211
Subject Index	217

Preface

ONE OF THE VERY FIRST STATISTICAL APPLICATIONS ON which Dr. Good worked was an analysis of leukemia cases in Hiroshima, Japan following 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 Scheffe's *Analysis of Variance*. The work took several months and the results were somewhat inconclusive, 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 the necessary calculations. Yet, statistical software will no more make one a statistician than would a scalpel turn one into a neurosurgeon. Allowing these tools to do our thinking for us is a sure recipe for disaster.

Pressed by management or the need for funding, too many research workers have no choice but to go forward with data analysis regardless of the extent of their statistical training. Alas, while a semester or two of undergraduate statistics may suffice to develop familiarity with the names of some statistical methods, it is not enough to be aware 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 concepts in statistics as null and alternative

hypotheses, p value, significance level, and power. Assisted by reprints from the statistical literature, we reexamine sample selection, linear regression, the analysis of variance, maximum likelihood, Bayes' Theorem, meta-analysis, and the bootstrap.

Now the good news: Dr. Good's articles on women's sports have appeared in the *San Francisco Examiner*, *Sports Now*, and *Volleyball Monthly*. So, if you can read the sports page, you'll find 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 Vance 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 we detail the procedures to follow when making two- and k -sample comparisons. In addition, we cover the procedures for analyzing contingency tables and two-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. Practical significance is distinguished from statistical significance, and induction is distinguished from deduction.

Twelve rules for more effective graphic presentations are given in Chapter 8 along with numerous examples of the right and wrong ways to maintain reader interest while communicating essential statistical information.

Chapters 9 through 11 are devoted to model building and to the assumptions and limitations of standard 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, a 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.

We hope you soon put this textbook to practical use.

Phillip Good

Huntington Beach, CA

brother_unknown@yahoo.com

James Hardin

College Station, TX

jhardin@stat.tamu.edu