

Cambridge University Press  
978-0-521-44554-2 - Statistics: Concepts and Applications  
Harry Frank and Steven C. Althoen  
Frontmatter  
[More information](#)

**STATISTICS**

Cambridge University Press  
978-0-521-44554-2 - Statistics: Concepts and Applications  
Harry Frank and Steven C. Althoen  
Frontmatter  
[More information](#)

---

# STATISTICS

## CONCEPTS AND APPLICATIONS

**Harry Frank**

Department of Psychology  
The University of Michigan – Flint

**Steven C. Althoen**

Department of Mathematics  
The University of Michigan – Flint



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press  
978-0-521-44554-2 - Statistics: Concepts and Applications  
Harry Frank and Steven C. Althoen  
Frontmatter  
[More information](#)

Published by the Press Syndicate of the University of Cambridge  
The Pitt Building, Trumpington Street, Cambridge CB2 1RP  
40 West 20th Street, New York, NY 10011-4211, USA  
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1994

First published 1994

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*Library of Congress Cataloging-in-Publication Data*

Frank, Harry, 1942–  
Statistics: concepts and applications / Harry Frank, Steven C.  
Althoen.  
p. cm.  
ISBN 0-521-44554-X  
1. Statistics. I. Althoen, Steven C. II. Title.  
QA276.12.F73 1994  
519.5 – dc20 93-27502

A catalog record for this book is available from the British Library.

ISBN 0-521-44554-X hardback

Cambridge University Press  
978-0-521-44554-2 - Statistics: Concepts and Applications  
Harry Frank and Steven C. Althoen  
Frontmatter  
[More information](#)

---

To  
Dr. David A. August, M.D.,  
Division of Surgical Oncology  
Dr. Robert L. Cody, M.D., and Dr. Max S. Wicha, M.D.,  
Division of Hematology and Oncology,  
Department of Internal Medicine  
Dr. Allen S. Lichter, M.D.,  
Department of Radiation Oncology  
The University of Michigan Medical School  
frontline soldiers in the war on cancer, where statistical  
decisions are truly matters of life and death  
Harry Frank

To  
Marcia, Michael, and Tosca  
for all their patience during the preparation of this text  
Steven C. Althoen

# Contents

---

<i>Preface</i>	xv
<i>Acknowledgments</i>	xxv

**Part I:** Organization and description of data

<b>1</b>	The organization of data	2
	The meaning of data	3
	Quantitative data	3
	Types of quantitative data	3
	Representations of data	4
	Tabular representation	4
	Graphic representation	10
	Exercises 1.1	13
	Summary	16
<b>2</b>	Describing distributions	18
	Descriptive statistics	19
	Desirable properties of descriptive statistics	20
	Statistical shorthand	21
	Exercises 2.1	24
	Describing the location of a distribution	26
	The mode: <i>Mo.</i>	27
	The median: $\tilde{x}$	28
	Exercises 2.2	37
	The mean: $\bar{x}$	39
	Exercises 2.3	44
	Describing the dispersion of a distribution	48
	The range	49
	The mean difference among observations	50
	The mean deviation from the mean	51

**viii Contents**

---

The variance and standard deviation	52
Exercises 2.4	59
Describing the shape of a distribution	62
Skewness	63
Kurtosis	68
Exercises 2.5	71
Summary	73
<b>3 Describing individuals in distributions</b>	<b>78</b>
Percentile ranks and percentiles	79
A special case: the median	82
Percentiles with ungrouped scores	83
Exercises 3.1	84
Standardization	85
Measurement in the sciences	86
Establishing common scales of measurement	87
The standard score, $z$	92
Exercises 3.2	95
Summary	96
<b>4 Describing joint distributions of data</b>	<b>98</b>
The scatter diagram	99
The concept of a statistical relation	101
Exercises 4.1	104
Quantitative description of a statistical relation	106
The covariance $C_{XY}$	107
Exercises 4.2	110
The correlation coefficient $r_{XY}$	110
Exercises 4.3	121
Linear regression	125
The regression equation: prediction and error	126
Interpreting regression	131
Exercises 4.4	134
Summary	137
Postscript to Part I	140
Graphs and statistics: two sides of a coin	140
Outliers	142
Intervals and scales	143
Statistical description as data exploration	147
<b>Part II: Probability</b>	
<b>5 Introduction to probability</b>	<b>150</b>
The classical approach to probability	152
The probability of $E_1$ and $E_2$ : the multiplication theorem	156

	Contents	ix
The probability of $E_1$ or $E_2$ : the addition theorem		160
Exercises 5.1		163
Counting results of more complicated experiments		165
Exercises 5.2		177
Other approaches to probability		180
The relative frequency approach to probability		180
The mathematical definition of probability		182
Summary		184
<b>6 Discrete probability distributions</b>		<b>186</b>
Introduction: the notion of a probability distribution		187
Discrete and continuous random variables		187
The uniform distribution		189
Describing probability distributions		189
Standardized random variables		198
Exercises 6.1		200
The binomial distribution		203
The binomial experiment		204
The expected value and variance of a binomial random variable		209
The standardized binomial random variable		213
Exercises 6.2		214
Representation of probability distributions of discrete random variables		217
Tabular representation		217
Graphic representation		218
Functional representation		220
Exercises 6.3		220
Summary		221
<b>7 Continuous probability distributions</b>		<b>224</b>
The uniform distribution revisited: probability as area and the notion of probability density $\phi$		225
Cumulative probability $\mathbb{P}$		231
Exercises 7.1		231
The normal distribution		232
The effects of sample size on the histogram for the binomial distribution		234
The probability histogram for the standardized binomial random variable		235
The standard normal random variable		240
The table of cumulative probabilities for the standard normal curve		244
Exercises 7.2		250
The normal approximation to the binomial		251
The nonstandard normal random variable		260
Exercises 7.3		264

**x Contents**

---

Representation of probability distributions of continuous random variables	268
Tabular representation	268
Graphic representation	268
Functional representation	269
Summary	269
 <b>Part III: Introduction to statistical inference</b>	
<b>8 Sampling distributions and estimation</b>	<b>274</b>
Populations and samples	275
Sampling as a random experiment	278
Statistics as random variables	279
Exercises 8.1	281
Sampling distributions	281
The sampling distribution of the mean $\bar{X}$	282
Exercises 8.2	292
Estimation	295
Point estimation	296
Interval estimation	305
Exercises 8.3	318
Summary	321
 <b>9 Hypothesis testing</b>	
Outline and example of hypothesis testing	327
Formulation	328
Decisions	329
Data collection	330
Conclusions	330
Foundations of hypothesis testing	330
Formulation	330
Decisions	334
Exercises 9.1	341
Data collection	352
Conclusions	356
Exercises 9.2	360
Evaluation of the statistical test	363
The concept of power	363
The power of a test against a composite alternative hypothesis	364
Scientific versus statistical significance	366
Postscript: decisions concerning hypotheses about proportions	368
Test statistics: $Z_X$ and $Z_{X/N}$	368
Upper-tailed tests, lower-tailed tests, and two-tailed tests	369



	Contents xi
Significance level and adjustments for continuity	370
Exercises 9.3	372
Summary	375
<b>10 Testing hypotheses about population means</b>	<b>380</b>
A “textbook” example: the Neanderthal problem	382
Formulation	383
Decisions	384
Data collection	386
Conclusions	387
The power of the statistical test	387
Testing hypotheses about one population mean $\mu$	388
Large samples ( $N \geq 30$ )	389
Exercises 10.1	398
Small samples ( $N < 30$ ): Student’s $t$	401
Exercises 10.2	408
Exercises 10.3	413
Testing hypotheses about two population means $\mu_1$ and $\mu_2$	417
Formulation	418
Decisions	419
Exercises 10.4	424
Data collection	436
Testing hypotheses about two population proportions $p_1$ and $p_2$	439
Exercises 10.5	441
Summary	448
<b>11 Testing hypotheses about population variances</b>	<b>454</b>
The $\chi^2$ family of distributions	455
Expected value and expected variance of $\chi^2(\nu)$	456
Properties of $\chi^2$ distributions	457
The addition and subtraction of $\chi^2$ random variables	460
Testing hypotheses about one population variance $\sigma^2$	461
Formulation	461
Decisions	462
Exercises 11.1	467
Fisher’s $F$ distribution	470
Testing hypotheses about two population variances $\sigma_1^2$ and $\sigma_2^2$	471
Formulation	471
Decisions	471
Exercises 11.2	474
Summary	478

**xii Contents**

---

<b>12 Testing hypotheses about many population means</b>	
$\mu_1, \mu_2, \dots, \mu_J$ : introduction to analysis of variance	482
Double summation notation	483
The algebra of double summations	486
Exercises 12.1	487
Logic and overview of analysis of variance (ANOVA)	488
Formulation	488
Decisions	490
Exercises 12.2	495
Foundations of analysis of variance	498
Assumptions underlying analysis of variance	498
The linear model for analysis of variance	500
Variability as sums of squares	504
Decisions: test statistic and rejection rule for ANOVA	509
Mean squares	509
Rejection rule for ANOVA	513
Results: the ANOVA summary table	514
Exercises 12.3	516
Evaluation of analysis of variance: post hoc comparisons	524
Formulation: hypotheses about contrasts	524
Scheffé's test statistic	526
Exercises 12.4	528
Summary	532
 <b>Part IV: Hypothesis testing: intermediate techniques</b>	
 <b>13 More complex analyses of variance</b>	538
Two-way analysis of variance	539
Interaction and the linear model for two-way ANOVA	542
Fixed effects and random effects	544
Hypotheses and test statistics for two-way ANOVA	544
Analysis of variance for one experimental treatment with repeated measures	554
Exercises 13.1	559
Summary	565
 <b>14 Testing hypotheses about covariance: correlation and regression</b>	570
The joint distribution of two variables	571
Joint discrete probability distributions	571
Jointly distributed continuous random variables	577
Exercises 14.1	580

	Contents	xiii
The population correlation coefficient $\rho$		583
Testing hypotheses about $\rho$		583
Testing hypotheses about $\rho_1$ and $\rho_2$		588
Interpreting tests about $\rho$ : linearity, reduction of uncertainty, and independence		590
Exercises 14.2		593
Regression		596
The population regression equation		596
Formulating hypotheses about regression		597
Test statistics: analysis of variance for regression		598
Exercises 14.3		607
Summary		609
<b>15 Testing hypotheses about entire distributions:</b>		
Pearson's <i>chi</i> -square		616
Testing hypotheses about one distribution: goodness-of-fit tests		619
Goodness of fit as a test of proportions		626
Exercises 15.1		628
Testing hypotheses about two or more distributions: tests of association		635
Tests of homogeneity		636
Tests of independence		644
The fourfold table		647
Exercises 15.2		649
Assumptions underlying Pearson's <i>chi</i> -square		656
Special applications of Pearson's <i>chi</i> -square		657
Goodness-of-fit test for normality		657
The median test		661
Fourfold point correlation: the <i>phi</i> coefficient		663
A test for correlated proportions		667
A test of homogeneity when expected frequencies are small		668
Exercises 15.3		670
Summary		674
<b>Appendices</b>		
I. The algebra of summations		682
II. Linear interpolation		686
III. Regression coefficients		688
IV. Scales of measurement		692
V. The algebra of expectations		696
VI. Fundamentals of research design		704
VII. Sums of squares		708

**xiv Contents**

---

<b>VIII. Tables</b>	<b>716</b>
1. Uniformly distributed random numbers	716
2. The binomial probability distribution	716
3. Cumulative probabilities for the standard normal random variable	724
4. Student's <i>t</i> -values for selected cumulative probabilities	726
5. <i>Chi</i> -square values for selected cumulative probabilities	727
6. Fisher's <i>F</i> -values for selected cumulative probabilities	729
7. The Poisson probability distribution for selected values of $\mu$	744
<b>IX. Solutions to odd-numbered exercises</b>	<b>746</b>
 <i>Index</i>	 <b>840</b>

# Preface

---

It is trite and a little pompous to describe writing a book as a journey of discovery. In the present case it happens to be literally true. The authors didn't set out to write a book. We set out to be running partners. In the course of our weekly 10-mile runs, we discovered that despite our wide differences in training and teaching experience, we had come to very similar conclusions about how undergraduate statistics should be taught. Much of the book was written, or at least drafted, during our runs. Running does strange things to the mind. Lack of oxygen forces a runner toward mental simplicity. (Try a few simple arithmetic problems after you've run three or four miles; it is a humbling exercise.) Sometimes this leaves the runner in a hopeless muddle. Sometimes it results in startling clarity by paring away everything that is not essential. We hope that the final result—this book—has captured the clarity and excised the muddle.

## TO THE INSTRUCTOR

A textbook is written for students, but students don't have a chance to read it unless the instructor buys it. That's why the traditional note "To the Instructor" usually precedes the note "To the Student." It is the authors' opportunity to make a genteel sales pitch directly to the potential buyer.

Since you are teaching undergraduate courses, you probably don't have enough time to examine all the new books that cross your desk. We'll make things easy. This book is probably *not* for you if you want a book

1. for students who have *not* had at least one good, solid course in college algebra,
2. that is tied to a specific academic discipline,
3. with minimal coverage of probability theory.

**xvi Preface**

---

If any of the preceding apply to you, give this complimentary copy to a colleague or a talented student or return it to the publisher. Better yet, send it to one of the authors. We have lots of relatives who are expecting gift copies, and (unlike the publisher) *we* will refund your postage.

---

If you are still reading, you are probably teaching the sort of course we had in mind when we wrote this book.

The course in which the book evolved is housed in a department of psychology, but about a third of the students who enroll are majoring in nursing, health care, biology, chemistry, or computer science. The book is most ideally suited, therefore, to the sort of introductory general statistics course frequently offered by departments of statistics or mathematics for students in the behavioral and natural sciences.

**Student preparation**

The expected level of mathematical preparation is modest, but uncompromising. Your students should be familiar with roots, powers, absolute values, inequalities, graphs, and equations. It is helpful if they have had exposure to functions, variables, and factorial notation. Familiarity with calculus is neither assumed nor required. In Chapter 7 (Box 7.1, p. 241) we *briefly* explain the integral sign  $\int$ , but only as a sidebar to show that the definitions of expected value, variance, and cumulative probability for continuous random variables are analogous to the definitions given for the discrete case.

**Notation**

Contemporary statistical notation derives from a host of traditions, disciplines, and historical precedents, and as a consequence there is only a handful of symbols that may truly be thought of as “standard notation.” Where such standard notation exists, we use it. In cases where conventions vary widely and we had to make arbitrary decisions, we tried to choose notation that is most consistent with the established conventions and, at the same time, tried to avoid notational distinctions that are difficult to render on a blackboard.

Unfortunately, even widely used notation is occasionally ambiguous, inconsistent, contrary to more general convention, or conceptually misleading. In such cases we have sometimes chosen notation that may be less familiar, but is by no means unique to this text.

A glossary of the most frequently used symbols appears in the end papers, and our few departures from common practice are as follows:

1. The number of *observations*, or experimental trials, is denoted with a capital  $N$ . The number of *distinct values* in a collection of observations is denoted with a lowercase  $n$ . We find that this notational distinction helps avoid confusion between summations for raw data and summations using frequency notation. By reserv-

ing  $N$  for the former, we give precedence to the most traditional notation for the number of observations (trials) in binomial experiments.

- 2. Random variables, or measurements, are sometimes treated as value sets and unless otherwise indicated are therefore denoted with capital letters (e.g.,  $V, W, X, Y$ ). Corresponding lowercase letters ( $v, w, x, y$ ) are *values* of random variables.
- 3. We observe the standard practice of denoting sample statistics with Roman letters and denoting population values and parameters with Greek letters.<sup>1</sup> The quantity

$$\frac{\sum (x - \bar{x})^2}{N}$$

is the variance of a collection of  $N$  observations and is therefore denoted  $s^2$ . The quantity

$$\frac{\sum (x - \bar{x})^2}{N - 1}$$

is the estimate of the *population* variance and is therefore denoted  $\hat{\sigma}^2$ .

- 4. We have found that introductory statistics students often have difficulty with the notion that *probability* in a confidence statement, such as

$$P(x - 1.96\sigma \leq \mu \leq x + 1.96\sigma) = .95$$

ceases to be meaningful as soon as the value of  $x$  is specified. To avoid awkward and often counterintuitive semantics, we therefore use  $C$  (for confidence), rather than  $P$ . For example,

$$C(x - 1.96\sigma \leq \mu \leq x + 1.96\sigma) = .95$$

- 5. Nomenclature for sums (and means) of squares in analysis of variance now differs widely. We retain the traditional sum of squares *within* groups (rather than adopting the sum of squares *error* favored by some authors), but we have taken advantage of the recent liberalization to eliminate a long-standing grammatical flaw: We refer to the sum of squares *among* groups (denoted  $SS_A$ ), rather than the sum of squares *between* groups.

Probability theory

We devote 123 pages to our three chapters on probability. This is a sizable fraction of the book, but it must be emphasized that much of this is exercises, worked-out examples, and graphic illustrations. Since many of the students for whom the text is intended find probability uncomfortably abstract, we present the material at a more gradual pace than might

<sup>1</sup>There are two notable exceptions. We use *cov* to denote the population covariance. We also follow the nearly universal convention of denoting binomial probability of success as  $p$ .

**xviii Preface**

---

be optimal for a more mathematically oriented readership, developing concepts in a methodical, step-by-step fashion and building in ample redundancy.

You will also find that many of the probability “story problems” involve scenarios from the popular fantasy role-playing game, *Advanced Dungeons & Dragons*®. No familiarity with the game is necessary. We use these problems only because the game makes use of a variety of polyhedral dice (4-sided, 6-sided, 8-sided, etc.) in combinations that generate a refreshing variety of discrete probability distributions. Our students have found these exercises a welcome respite from the usual run of balls-and-urns scenarios or gambling problems. Incidentally, most hobby shops sell packets of these dice for a few dollars. They can be used in class to generate data on the spot.

**Boxed displays**

Thin-line boxes are used to highlight important definitions and definitional equations. Bold-line, numbered boxes are used for a number of proofs and derivations, supplementary explanations, and other such material that we consider too important to bury in an appendix, but which would be intrusive if incorporated into the main flow of text. We have included these in the same spirit in which hikers and cross-country skiers tuck an extra pair of socks in the outside pocket of a day pack: Better to have them and not need them than to need them and not have them.

**Topical coverage and instructional strategies**

Our selection of topics is predicated on the belief that an introductory general statistics course should not be a superficial survey or compendium of techniques. It should teach *less* and do it *thoroughly*, with deliberate emphasis on fundamental principles. Consequently, a number of topics that have largely disappeared from introductory texts are covered in some detail. Our treatment of statistical description, for example, is more extensive than you might expect in a contemporary book. We do this for three reasons. First, many instructors simply want more emphasis on description. If you are looking for a “traditional” treatment of topics like skew and kurtosis you will find it. If you feel that these topics are a waste of time, they comprise a self-contained unit (at the end of Chapter 2), and you can omit them with no loss of continuity. The same is true for the treatment of percentiles (Chapter 3).

Second, a strictly descriptive context allows earlier introduction of some topics, such as correlation and regression, that would have to be deferred if description and inference were interwoven.

Third, relative frequency distributions are tangible and real and therefore make it easier to grasp many concepts that can be very elusive if first introduced in the context of probability distributions. For example, even though the availability of computers has made it unnecessary as a practical matter to calculate the mean using frequency notation, early introduction of this technique makes it easier to present the expected



value later on. Likewise, our rather detailed coverage of the median with grouped data paves the way for introducing the notion of probability as the area under a density function.

(Incidentally, our treatment of descriptive statistics as a topic in its own right means that you will not find anything about *sampling* under the rubric of description. In Part I we talk about “data collections,” not “samples.” A sample implies a parent population, which begs the question of inference. Random sampling is therefore introduced with inference in Chapter 8.)

Because of its heuristic usefulness, discrete probability is also covered more deliberately than has become customary. Since discrete variables are more palpable than continuous variables, we find that fewer students “hit the wall” when introduced to methods based on the normal curve if much of the architecture of hypothesis testing is first developed in the context of the binomial distribution. From there, it is an easy progression to normal approximation methods for testing hypotheses about proportions and, then, to hypotheses about means.

The other side of this pedagogical coin is that there are many things you will *not* find in this book. We exclude a number of introductory-level topics that might be desirable in books geared to specific disciplines. For example, there is no discussion of time series, which is a standard topic in business statistics, or measurement models, a staple of educational statistics.

We have also limited our presentation of more advanced material. Our treatment of experimental design (Appendix VI) is confined to fundamentals. There is no discussion of such topics as randomized blocks, Latin squares, nested factors, partial  $F$  tests, and so on. Nor do we discuss multiple correlation or multiple regression, and there is no coverage of nonparametric statistics per se. Nevertheless, Chapters 13–15 offer a selection of those advanced topics that are fundamental to the widest range of research traditions and which follow with least elaboration from the core material in Chapters 1–12 (e.g., without introducing new distributions).

Furthermore, Chapters 13–15 are organized to permit either a broad survey or narrower, in-depth coverage of selected techniques. None of the last three chapters assumes background beyond Chapter 12, and more specialized applications are treated at the end of each chapter.

### Examples, exercises, and computer software packages

As suggested by the title, one of the principle goals of this text is mastery of applications. Almost every new topic is therefore illustrated with worked-through examples. In addition:

- Exercises are placed at the end of each topical section in order to promote step-by-step consolidation of problem-solving skills.
- Routine calculational drill problems comprise 30 to 40 percent of the exercises. This is no longer fashionable in disciplines outside of mathematics, but it lets students become familiar with the basic

## xx Preface

mechanics of problem solving without the additional task of translating textual situations into numbers. The balance between drill exercises and content problems changes in favor of content problems as one proceeds through each set of exercises, as one proceeds through the exercise sets in any particular chapter, and as one proceeds from chapter to chapter through the book.

- To demonstrate the diversity of content to which statistical techniques are applied and to appeal to a broad range of student interests, we have drawn content problems and examples from a wide variety of academic areas (e.g., paleoanthropology, agriculture, epidemiology, psychology, animal behavior, and environmental chemistry) as well as from the everyday statistical environment—sports, public opinion surveys, census data, civil rights litigation, etc.
- The personal computer has become the basic tool of the statistical workplace, and use of statistical software packages is becoming a common feature of introductory statistics curricula. Accordingly, we have prepared a companion workbook (*User-Friendly*) that includes a tutorial for the ASP (A Statistical Package for Business, Economics and the Social Sciences<sup>®</sup>) system and supplementary exercises. In addition, many of the content problem in *this* book include data sets that can be used easily with ASP or any other software package. However, the text is intended to stand independently of the software-driven workbook.

Instead of gearing the text to computer protocols, we develop and provide calculational formulae in terms of basic computational quantities (e.g., sums of squares and sums of products) that can be obtained with a single entry of data values on most “student” models of hand calculators. We have taken this approach because it is our experience that immersion in the mechanics of calculation has fundamental instructional importance. One cannot become a world-class tennis player by watching videotapes of Boris Becker or become a superb downhill racer by watching replays of Franz Klammer defy gravity at Innsbruck. You have to hit the ball and you have to take your falls. Similarly, one cannot acquire a statistical sense for data without *doing* statistics. *Watching* numbers come up on a computer screen is no substitute. The computational algorithms eliminate much of the drudgery of routine arithmetic without depriving the student of the instructional benefits that derive from working through the intermediate steps and seeing raw data evolve into interpretable statistical results.

- Appendix IX provides the *solutions* to the odd-numbered problems, not just the *answers*. The detail of explanation for the more challenging exercises is about the same level as the explanations given in the text for worked-out examples. We consider the solutions to be an integral elaboration of the text. Furthermore, the odd- and even-numbered problems are usually paired in such a way that the solutions to odd-numbered exercises offer implicit guidelines for solving neighboring even-numbered exercises.

**Permission to photocopy materials from this book**

A number of the appendices and chapter sections are self-contained presentations that might be useful additional reading for students in other courses. We have asked Cambridge University Press to grant all requests to photocopy materials from this book if copies are to be distributed free of charge to students.

**TO THE STUDENT**

Galileo is supposed to have said that mathematics is the language in which God wrote the universe. Since Galileo was a fluent speaker of mathematics, his claim may seem a little self-serving. Nevertheless, it is probably safe to say that mathematics—or that body of mathematical applications called statistics—is the language in which Man *reads* the universe. Statistics in one form or another has become the common language by which agronomists, chemists, educators, geneticists, medical researchers, political scientists, psychologists, and sociologists—not to mention attorneys, corporate executives, baseball managers, public policy advocates, political strategists, military planners, and the insurance industry—read some portion of the world in which we all live.

Statistics is a language with a numerical vocabulary, a mathematical grammar, and, like any language, its own, distinctive way of shaping the speaker’s view of the world. In many respects an introductory statistics course is therefore like an introductory course in German, Hebrew, Japanese, or any other language. Language texts provide glossaries for quick reference to unfamiliar words. A glossary of mathematical and statistical notation likewise appears in the end papers of this book. A quick glance at the glossary reveals that the symbol system of statistics shares an important property with the symbol system of other languages: Meaning often depends on context. For example, the lowercase Greek alpha ( $\alpha$ ) can mean the probability of a Type I error or it can mean the *y*-intercept of a regression equation. Native speakers of English are seldom confused about the meaning of the word “draw,” but the *Oxford English Dictionary* lists over 75 meanings (and does not include the American English use of “draw” as a noun meaning a shallow gully).

Textbooks for introductory language courses almost invariably advise the student to buy a dictionary. Similarly, you will need a calculator for the exercises in this book. The text provides computational formulae that reduce both the time required to perform calculations and the chances of making a mistake. To use most of these formulae, you need a calculator that computes sums, sums of squares, and sums of products. These functions are found on almost all but the least expensive calculators, and the symbols that most commonly appear on calculator keys to indicate these operations are  $\Sigma x$ ,  $\Sigma y$ ,  $\Sigma x^2$ ,  $\Sigma y^2$ , and  $\Sigma xy$ . More specialized (and only slightly more expensive) calculators allow you to enter your data and then obtain a variety of statistics directly, each with a single keystroke. Check with your instructor before purchasing a calculator. He or she may want you to use a model that has (or *doesn’t* have!) certain capabilities.

**xxii Preface**

---

Every statistics course, like every language course, also presupposes some particular level of background experience. If you have read the preface section “To the Instructor” you know that this book assumes that you have more or less mastered the fundamentals of college-level algebra. Courses in mathematics beyond the prerequisite level are helpful only insofar they leave a residue of mathematical maturity. There are very few places in the book where past experience with the material found in, say, a calculus or analytic geometry course will be helpful.

Not surprisingly, the first encounter with statistics is often like the first encounter with any foreign language. It *can* be fraught with apprehension, frustration, and bewilderment. Statistics can also be richly rewarding, offering fresh perspectives on daily experience and aesthetic appreciation of the elegant logic of scientific reasoning. Professional golfer Walter Hagen once said, “We pass this way only once, so don’t hurry, don’t worry, and be sure to smell the flowers on the way.” We can’t improve on this advice.

**Tips for using this book**

By the time you have been in school long enough to be reading this book, you have already had many years to develop your own highly individual strategy for using textbooks and completing assignments. On the other hand, authors are also individuals and necessarily approach the task of writing books with certain assumptions about how their books can be used most effectively. These assumptions begin with authors’ own experience, but it is very difficult to recall with any authenticity one’s first encounter with something that has become familiar. (Try conjuring up a vivid recollection of the first time you rode a bicycle or drove a car or tasted a hot fudge sundae.) Authors therefore subject their own students to endless manuscript drafts before a book is published, and they find that their assumptions must inevitably be molded and adapted to the reality of student experience. The following advice is based on how we planned for the book to be used and on feedback from our many students who have described the strategies they found successful—and not so successful.

- Don’t try to work backward. The backward approach goes something like this: Begin with the exercises. When you are stumped by a problem, scramble madly through the chapter to find an example that looks like the problem you’re working on. If you don’t find one, look in the back of the book for the solution to a similar problem. If all else fails, read just enough of the chapter to find out how to do the problem.
- Read the entire chapter first. This ordinarily means that you should try to stay ahead of your instructor’s lectures, and this requires a measure of discipline. Very few people (and the authors are *not* numbered among this lucky few) can absorb mathematics from a book alone. Most people need, in addition, to have mathematics *told* to them and *demonstrated*, so there is some natural reluctance to

venture alone into unknown mathematical territory. The lectures will be much more enlightening if you have already read the material. Trust us.

- Use a variety of reading strategies. We have said that introductory statistics is like an introductory language course. The fundamental skills in learning a language are *reading* skills. The same is true for statistics. One important reading skill is to use different strategies at different stages of the learning process.

The first time you read new material, read for *passive understanding*, or *familiarity*. Don't try to put it all together. Read through the exercises, but don't do them. Skip the boxed proofs and derivations. *Follow* the examples, but don't work them out unless you can't see how the authors got their answers. Just make sure you understand each sentence *as you read it*. This reading serves a number of purposes. First, it gives you confidence. If you can understand every sentence, then it can truly be said that there's nothing in the chapter that you *don't* understand. Second, even a light reading begins the unconscious process of organizing the new material—making connections with material you've learned earlier and creating the mental pegs on which you'll hang material you pick up on the next reading and the points explained by your instructor. Third, it shows where the chapter leads. Getting there is easier if you know where you're going.

The second time you read is for *comprehension*. This time, you actively work toward putting the pieces together, seeing how the various sections and subsections (and even sentences) form a single body of information and how the new topics relate to earlier topics. Start with the outline at the beginning of the chapter *and* with the summary at the end of the chapter. Then, read with a pencil in your hand and work through the examples and the derivations. Taking a crack at some of the *drill* exercises can be useful now. You've *seen* the material, you've probably had some of it *told* to you; now it's time to *do* it.

The third reading is for *mastery*. At this stage of the game you can get maximum benefit from the more challenging content exercises (a euphemism for story problems). Students have repeatedly told us that if they read the material thoroughly *before* attempting the exercises, they actually spend less *total* time (reading plus exercises) than they spend on *just the problems* when they try doing the exercises cold (i.e., when they do things backward). More important, *this* is the stage of learning when the exercises serve to consolidate your *understanding* of what you've been reading.

- How to approach story problems without fear and loathing. The routine, computational drill exercises at the beginning of almost every exercise set are intended to help you master the mechanics of doing problems, but in the real world, there is no divine agency (or instructor or author) to set up problems for you. You have to figure out what to do before you set about applying the mechanics of doing it. This is why we have story problems.



**xxiv Preface**

---

When you do computational drills, you begin with the specifics of the problem and apply general techniques. With story problems it's better to begin with general principles and move to the specifics. That is, *don't* start with the problem. There is an infinity of potential problems out there, and if you approach each one individually, all you will do is get lost in the details, drink too much coffee, and get an ulcer. On the other hand, there are only a few *kinds* of problems, and for each kind of problem, there are only a few *principles* to be applied. So, make it easy on yourself. Read the story, sit back, and ask yourself *what* you are supposed to do (e.g., describe a collection of data, compute a probability, estimate a parameter, test an hypothesis, etc.). Then, ask yourself what are the general principles involved in this sort of task. What information do you need, what assumptions have to be satisfied, etc.? *Then* get down to the specifics. Ask yourself how the principles can be applied to this particular problem. Where is the information you need given or implied in the story, what assumptions can you make, what operations do you perform on the data to get your answer, etc.?

- We have saved the most important tip for last: *It is almost always easier to figure out how to get from what you know to what you want if you draw a picture.*

P.S. Don't forget the flowers.

**TO ALL READERS**

The corrected proof pages that we returned to the publisher were, of course, perfect. However, it is a well-known fact that typographical and other errors are spontaneously generated by drafty warehouses. If you find any such errors, no matter how trivial, please let us know by snail mail or e-mail:

salthoen@umich.edu

Your help will be acknowledged in the next edition, and we will keep you on an e-mailing list and send you cumulative errata as they are detected by other readers. Errors are sometimes corrected between printings, so please indicate the printing of your copy, as shown on the copyright page—the final digit in the sequence beginning with 10.

# Acknowledgments

---

Two names appear on the cover of this book, but many people contributed to what is between the covers. First, we want to express appreciation to the University of Michigan–Flint undergraduate students who, with relative good cheer, used various versions of the manuscript as their principal text in Psychology 301. Most of you detected errors in the exercises, many of you identified passages that were perfectly sensible to the writers but incomprehensible to anyone else, and some of you accepted our challenge to rewrite sections you didn’t like. All of you can rest easier knowing that the single most nearly universal complaint about the manuscript has been rectified: The published text *does* have an index!

We also want to thank the Psychology 301 laboratory/teaching assistants, the book’s intellectual midwives: Mark Stefanski, Tom Bowyer, Mark Siefert, and Amy Collins Siefert. The last two, first as students and later as assistants, saw the text through at least three full revisions and undertook to rewrite a number of especially troublesome passages. The present treatment of the binomial random variable and binomial experiments in Chapter 6 is the only slightly revised work of these two remarkable students.

Family, friends, and University of Michigan–Flint (UM–F) colleagues enriched the content and improved our presentation in many ways. Melvin J. Warrick, former Associate Director of the Human Engineering Division, USAF Aerospace Medical Laboratory, read Part I with scrupulous attention, furnished the references from which we obtained the anthropometry data in Exercises 8.2, and made a number of valuable suggestions that inspired our Postscript to Part I. Martha G. Frank patiently explained how chemists use statistics, set up several of the analytic chemistry exercises, and furnished suitable data. Dave Dvorak, Deputy Principal of Mott Adult High School (and running partner), used the manuscript to supplement the statistics course required in his Ph.D. program and copyedited every line. Krista Hansen (UM–F Department of Mathematics) simplified cumbersome algebra in several places, and

**xxvi Acknowledgments**

---

Ed Bowden and Paul Bronstein (UM-F Department of Psychology) furnished raw data from their own research for exercises.

Two former colleagues also contributed to the project. Cay Horstmann taught mathematics at UM-F while attending graduate school and now combines an academic career with private enterprise. He is the father of ChiWriter™, the multifont word processor in which we wrote the manuscript and produced both the companion workbook, *User-Friendly*, and the instructor's manual. George Blackford, who once chaired the UM-F Department of Economics, left academia altogether for the world of software. He is the creator of ASP®, the statistical software package bundled with *User-Friendly*. His generous collaboration in this project prompted a number of changes in the present text.

A word about our reviewers: Reviewers are underpaid, underloved, unacknowledged collaborators. Underpaid because publishers could not possibly afford fair market value for their time and expertise; underloved because, if they do their job well, they make work for the authors and delay production; and unacknowledged because they traditionally work in anonymity. At our request, Cambridge University Press departed from this tradition and invited open reviews. We can therefore thank George Poole (Eastern Tennessee State University, Department of Mathematics) by name. Our treatment of probability runs a briefer and smoother course because of his thoughtful and detailed comments, and his suggestions concerning percentiles of ungrouped data furnished a much-needed link to our treatment of percentiles in *User Friendly*. Editors, of course, are never anonymous. Nor are they ordinarily qualified to act as reviewers. Alan Harvey is an exception. He is a mathematician, and his initial comments persuaded us to draft him for full-scale review duties. It was a good choice. He has a keen eye for devilish flaws that are really quite serious but so subtle that readers—and authors—glide over them with only a vague sense of uneasiness.