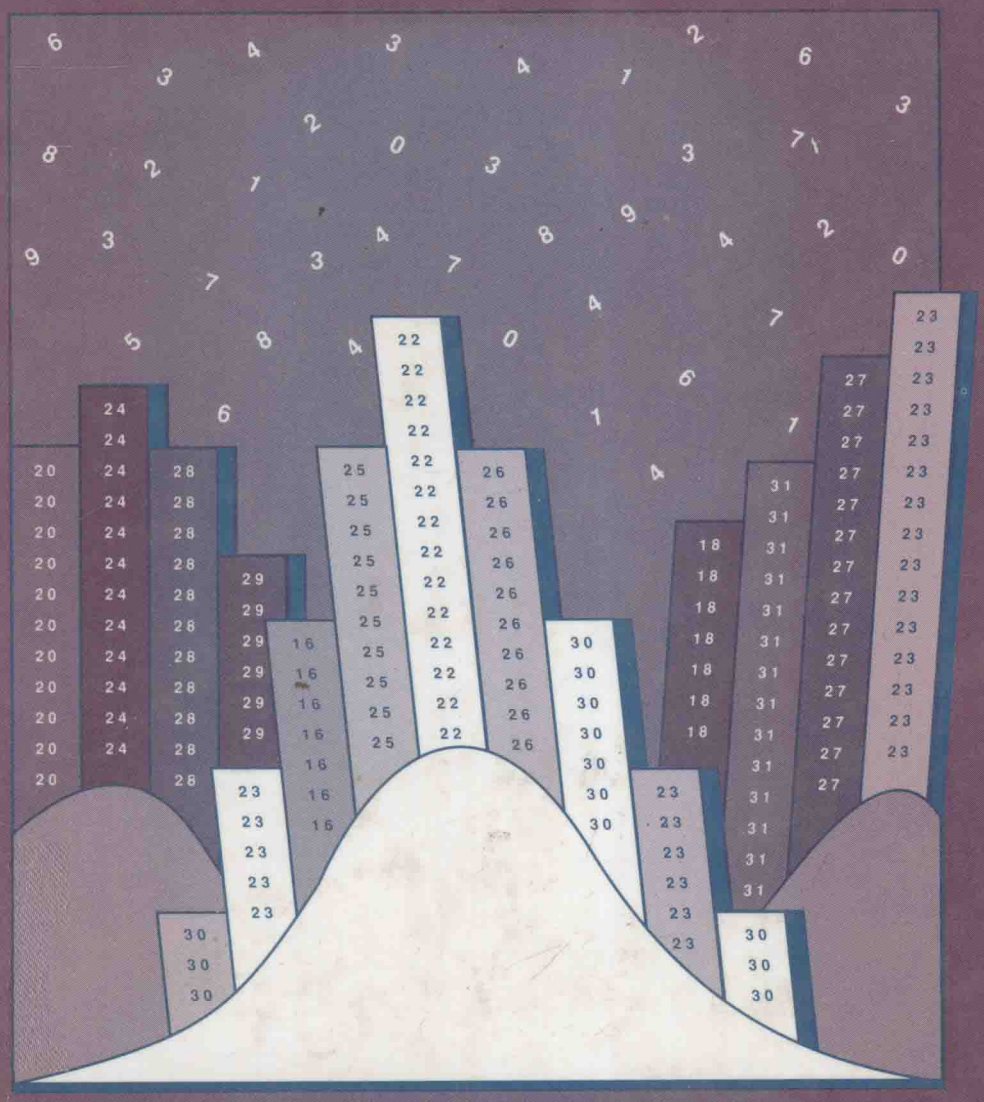
LEON F. MARZILLIER

Elementary Statistics



LEON F. MARZILLIER
Los Angeles Valley College

Elementary Statistics

WCB Wm. C. Brown Publishers

Book Team
Editor *Earl McPeek*
Developmental Editor *Theresa Grutz*
Art Editor *Donna Slade*
Photo Research Editor *Carol M. Smith*
Permissions Editor *Mavis M. Oeth*
Visuals Processor *Vickie Werner*

Wm. C. Brown Publishers

President *G. Franklin Lewis*
Vice President, Editor-in-Chief *George Wm. Bergquist*
Vice President, Director of Production *Beverly Kolz*
Vice President, National Sales Manager *Bob McLaughlin*
Director of Marketing *Thomas E. Doran*
Marketing Communications Manager *Edward Bartell*
Marketing Manager David F. Horwitz
Executive Editor Edward G. Jaffe
Production Editorial Manager *Colleen A. Yonda*
Production Editorial Manager *Julie A. Kennedy*
Publishing Services Manager *Karen J. Slaght*
Manager of Visuals and Design *Faye M. Schilling*

Cover design by Mark Elliot Christianson

Library of Congress Catalog Card Number: 89-61330

ISBN 0-697-05931-3

Printed in the United States of America by Wm. C. Brown Publishers, 2460 Kerper Boulevard, Dubuque, IA 52001

10 9 8 7 6 5 4 3 2 1

List of Formulas

Chapter	Formula	Use
2	(1) $\bar{x} = \frac{\Sigma x}{n}$	To find sample mean from raw data
2	(2) $\mu = \frac{\Sigma x}{N}$	To find population mean from raw data
2	(3) $\bar{x} = \frac{\Sigma fm}{\Sigma f}$ or $\mu = \frac{\Sigma fm}{\Sigma f}$	To find the mean from a frequency distribution
2	(4) $\bar{x}' = \frac{\Sigma fx'}{\Sigma f}$ and $\bar{x} = C + w\bar{x}'$	To find the mean from a frequency distribution using coding
2	(5) $LTCL + r\left(\frac{w}{f+1}\right)$	To find the rth value in an interval, used in calculating a median from a frequency distribution
2	(6) $\bar{x} = \frac{\Sigma xw}{\Sigma w}$	To find the weighted mean
2	(7) $s^2 = \frac{\Sigma d^2}{n}$	To find the variance of a sample
2	(8) $s = \sqrt{\frac{\Sigma d^2}{n}}$	To find the standard deviation of a sample
2	(9) $s = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$	Alternative to formula (8), often easier to apply
2	(10) $s = \sqrt{\frac{\Sigma fm^2}{\Sigma f} - \left(\frac{\Sigma fm}{\Sigma f}\right)^2}$	To find the standard deviation from a frequency distribution
2	(11) $s' = \sqrt{\frac{\Sigma f(x')^2}{\Sigma f} - \left(\frac{\Sigma fx'}{\Sigma f}\right)^2}$ and $s = ws'$	Formulas to find the standard deviation from a frequency distribution using the coding process
2, 3	(12) $z = \frac{x - \mu}{\sigma}$	To find the number of standard deviations a value x in a normal distribution is above or below the mean, also used to convert raw scores to standard scores in a population
2	(13) $z = \frac{x - \bar{x}}{s}$	To convert raw scores to standard scores in a sample
3	(14) $P(E) = \frac{\text{number of simple events in } E}{\text{number of simple events in } S}$	To find the probability (theoretical or empirical) of an event E
3	(15) $P(A') = 1 - P(A)$	To find the probability of the complement of an event; this is sometimes easier than finding the probability of an event directly

3	**(16)** $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ **(17)** $P(A \text{ and } B) = P(A)P(B\|A)$	Compound probability rules
3	**(18)** $\mu = \Sigma x \cdot P(x)$	To find the mean of a probability distribution
3	**(19)** $\sigma = \sqrt{\Sigma x^2 \cdot P(x) - [\Sigma x \cdot P(x)]^2}$	To find the standard deviation of a probability distribution
3	**(20)** $P(x) = w \cdot p^x \cdot q^{n-x}$, where w is found from Pascal's triangle or by using the formula $w = \frac{n!}{x!(n-x)!}$	To find the probability of x successes in n trials of a binomial distribution
3	**(21)** $\mu = np$	To find the mean of a binomial distribution
3	**(22)** $\sigma = \sqrt{npq}$	To find the standard deviation of a binomial distribution
4	**(23)** $SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	To find the standard deviation of a frequency distribution of $\bar{x}$ for samples of size n drawn from a very large, or infinite, population
4	**(24)** $SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$	Same as formula (23), but used when the population size N is not very large in comparison to n, the size of each of the samples
5	**(25)** $\bar{x} \pm zSD_{\bar{x}}$	A confidence interval estimate for μ
5	**(26)** $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$	Formula for the standard error of the sample mean
5	**(27)** $n = \left(\frac{z\sigma}{E}\right)^2$	Formula for determining the sample size when estimating a sample mean; generally rounded up to the next whole number
5	**(28)** $SD_p = \sqrt{\frac{PQ}{n}}$	To find the standard deviation of the frequency distribution of p
5	**(29)** $p \pm zSE_p$, where $SE_p = \sqrt{\frac{pq}{n}}$	A CIE for P, a population proportion
5	**(30)** $n = \frac{z^2}{4E^2}$	To determine a sample size when estimating a proportion and you have no idea of the rough size of P
5	**(31)** $SE_s = \frac{s}{\sqrt{2n}}$	To find the standard error when estimating a standard deviation of a population following a normal distribution

I dedicate this book to my father, Frederick Marzillier, who, while he was still alive, encouraged me to embark on this project in the first place and whose generosity helped me accomplish it.

Preface

"Why another statistics text?" That is the question I anticipate many who hold this book in their hands will ask themselves. It is true, there are many already on the market; but, after seventeen years of teaching the subject at the community college level, I have yet to find a text that does the job satisfactorily. Those texts that have been written for this audience, I have found, fit into one of two categories—those written by mathematicians and those written by behavioral scientists. The former produce texts that are overly mathematical and structured in the way algebra texts are. The latter give nice overviews of the subject but avoid computational material, often to the detriment of understanding.

The background that I possess may be uniquely suited to writing a text that cuts across these two classes. The combination of a bachelor's degree in mathematics and a master's degree in statistics from a social science institution, the London School of Economics, together with extensive experience in teaching students, many of whom dislike the subject, has given me rare insights into methods of presenting the material found in no other text on the market today. I hope the final product, which you now hold in your hands, will meet the objective I had when I started this project—to bring real understanding of statistics to the college student.

The text is designed for use in the standard lower division course in statistics, required of so many disciplines these days. Community college students are, of course, the users I had in mind when I wrote this text. However, it is most certainly appropriate for all lower division university students as well. No particular major field of study that uses statistics is either targeted or excluded by this book. One can find examples of applications of statistics to many diverse fields of study in the text and it would certainly be appropriate for majors in psychology, sociology, biology, economics, business, education, physics, engineering, and health, to name but a few. I have assumed that the student has had at least two years of high school algebra, or the equivalent in college. Without this background,

the reader is too often sidetracked by struggles with manipulations, resulting in frustration, rather than having what I feel should be a pleasurable experience! I have long thought that a logic course would also be a very useful weapon in the student's arsenal, especially when the material on hypothesis testing is reached.

I have included those topics in the text that are found in most introductory statistics texts on the market today. There is, however, too much material to cover in a semester's or quarter's course of standard length, unless the class is above average and is able to move with rapidity. Some of the chapters in Part 3 can, therefore, be omitted without loss of continuity. These chapters are intentionally short compared to the longer, earlier chapters and are designed to give a sample of different inferential techniques to the student, without going into great depth. After all, this is a *first* course in statistics! Some instructors may feel, in this age of computers, that the coverage pertaining to frequency distributions (sections 1–2, 2–2, 2–3, and 2–6) is unnecessarily extensive. I disagree, since I feel it is important for the user of statistics to understand what it is that high-powered calculators and computers do before learning how to push the right buttons. However, I do give alternative plans for covering the first two chapters, in the Instructor's Manual that accompanies the text. I urge you to cover the chapters as written, though, and I think you'll find that they are "user-friendly"!

Another dislike that I have for most statistics texts is that the authors split the treatment of bivariate data, covering some of it with descriptive statistics in Part 1 of the book and the rest at the end, after covering the ideas of inferential statistics. I prefer the more logical approach, which you will find in this text, covering it all together in Part 4, "Analysis of Bivariate Data."

Each statistical technique described in the book is followed by an example or examples with complete solutions, an exercise set follows each section, and a further set of exercises comes at the end of each chapter. These examples and exercises are designed to illustrate for the user how statistical techniques can be applied to many diverse fields and help put the student into the driver's seat, where real learning takes place. Also at the end of each chapter is a summary of terms and formulas, with which a student should be familiar before going on to the next chapter. Each chapter then ends with a set of exercises on all the ideas included in the chapter and an optional "Hands-on Class/Student Project," which is a set of exercises that enables the student to obtain a firmer grasp of the ideas of the chapter by using data that he/she has collected or generated. Answers to the odd-numbered problems appear at the end of the book so that students can self-check their work. To aid the instructor, an Instructor's Manual and TestPak accompany the text. Included in the former are the answers to the even-numbered problems, so that the instructor can use these as homework problems, if desired, as well as my ideas on the structure and teaching of the course. The latter is a resource for testing purposes. A study guide is also available for students but is not a necessary purchase for a complete understanding of the subject.

In conclusion, I would like to take this opportunity to acknowledge the professional way in which this book was brought to the marketplace by the staff

of Wm. C. Brown. In particular, I would like to thank Ed Jaffee, Earl McPeek, and Theresa Grutz. Without their help, the project could never have been completed. Finally, I would like to express my sincere gratitude to my colleague Phil Clarke for working *all* the problems in the text, thereby producing an invaluable error check. I know that nothing irritates students and instructors more than finding numerous errors in the back of the book. As a result of Phil's work, errors should be very few and far between!

Leon F. Marzillier

Reviewers

Michael Bonanno
Suffolk County Community College

Vern Crandall
Novell, Inc.

Mario Fiondella
Manchester Community College

Jim Foster
Weber State College

Mark Goldstein
West Virginia Northern Community College

Carroll Hall
New Mexico State University

Michael Karelius
American River College

Eric Lubot
Bergen Community College

Joseph L. Mazanec
Delta College

Alfred Milligan
Western New Mexico University

Lawrence Ringer
Texas A&M University

Jean Burr Smith
Middlesex Community College

Harold Thomas
Pittsburg State University

Froylan Tiscareno
Mt. San Antonio College

Bruce Vermeulen
Colby College

To the Student—What Is Statistics?

Some of you may have picked up this book simply because you wanted to know more about statistics. If so, I applaud you and am flattered that you chose my book with which to learn it. However, I suspect that the real reason you have this book in your hands is that you have been told you need to know more about this mysterious field to have a complete and thorough knowledge of your chosen course of study.

"But what does statistics have to do with psychology or sociology?" you might ask. "Isn't statistics part of mathematics?"

Statistics is a branch of mathematics, but it is also a branch of psychology and of sociology. Research is the lifeblood of these fields and it would be very difficult, to say the least, to interpret any research without statistical theory. There are, however, many other diverse fields that use statistics: biology, medicine, business, politics, economics, computer technology, weather forecasting, sports, education, and even such unlikely fields as English and history. Any field in which experiments are performed or data is collected needs statistical theory to analyze and interpret the results. It would be useless to perform a psychological experiment and have pages and pages of results if the researcher can't draw any conclusions from them.

Enter statistical theory.

Some questions that might be answered by statistical theory are:

What is the portrait of a city, ethnically and by socio-economic background?

How are the graphs printed in newspapers and magazines read and interpreted?

What is the average score on a psychological test for a group of children?

What is the chance of rain tomorrow?

What percentage of the voters in Chicago think the president is doing a good job? And how can this question be answered if a statistician has time and money to interview only a few of these voters?

How can you check if an advertisement is telling the truth?
Is the number of deaths from cancer related to the distance between victims' homes and a nuclear power plant?

These are the types of questions that can be answered using statistical theory. As you read and study this book, you will learn about methods that will allow you to answer such questions.

It would be impossible, in one course or one book, to discuss all the topics of statistical theory. The material you are about to study is the usual material required for a first course in statistics presented in a clear, comprehensible manner. It is not addressed to any particular group of people, such as psychologists or sociologists. Rather, it presents statistical procedures and techniques anybody can use. Examples and exercises are taken from many fields to illustrate the numerous and diverse applications of statistics.

The book is divided into four parts. The first part presents descriptive statistics—that is, the techniques required for describing and presenting data. This is followed by Part 2, a transition to inferential statistics—methods by which facts about a large population can be inferred or deduced from a small sample of the population, the subject of Part 3. The fourth and final part deals with the study of relationships between variables.

To achieve a complete and thorough understanding of statistics, you would have to know a great deal of mathematics. To understand this text, you need only know algebra, preferably to the intermediate level (i.e., the equivalent of two years of high school algebra). Just as it is possible to learn to drive a car without knowing anything about the inner workings of the engine, so it is possible to learn and apply the techniques of statistical theory without always knowing *why* they work. But be forewarned: when you learn to drive a car, you don't read how to do it in a book. Nor do you simply watch someone else do it. *You* need to get into the driver's seat and try it. And not just once, but as many times as it takes until you get it right. The same is true of learning statistics. It will not be sufficient for you to curl up by the fire and read this book. Nor will it be sufficient for you only to go to class and listen and observe your teacher. You have to get into the driver's seat and actually *do* statistics.

You will find, as you read and follow this book, that the explanation of each statistical technique is followed by one or more examples and that each section concludes with an exercise set to test your understanding. Follow the examples and do the exercises (more than once, if necessary), and you will find yourself beginning to think like a statistician. Statistical thinking is not difficult, but it is different and sometimes requires a little effort to grasp it fully.

To summarize, then, statistics is a branch of mathematics, applicable to many fields, dealing with the presentation, analysis, and interpretation of data. In the following chapters you will examine many ways in which this can be accomplished. With a little effort it is not hard to get into its way of thinking, and you will find that this "mysterious" field will bring new insights to a variety of problems. Good luck in your study of statistics!

CONTENTS

LEON F. MARZILLIER

Elementary Statistics

PART

1

Descriptive Statistics

"Let us sit on this log at the roadside," says I, "and forget the inhumanity and ribaldry of the poets. It is in the glorious columns of ascertained facts and legalized measures that beauty is to be found. In this very log we sit upon, Mrs. Sampson," says I, "is statistics more wonderful than any poem. The rings show it was sixty years old. At the depth of two thousand feet it would become coal in three thousand years. The deepest coal mine in the world is at Killingworth, near Newcastle. A box four feet long, three feet wide, and two feet eight inches deep will hold one ton of coal. If an artery is cut, compress it above the wound. A man's leg contains thirty bones. The Tower of London was burned in 1841."

"Go on, Mr. Pratt," says Mrs. Sampson. "Them ideas is so original and soothing. I think statistics are just as lovely as they can be."

The Handbook of Hymen — O. HENRY