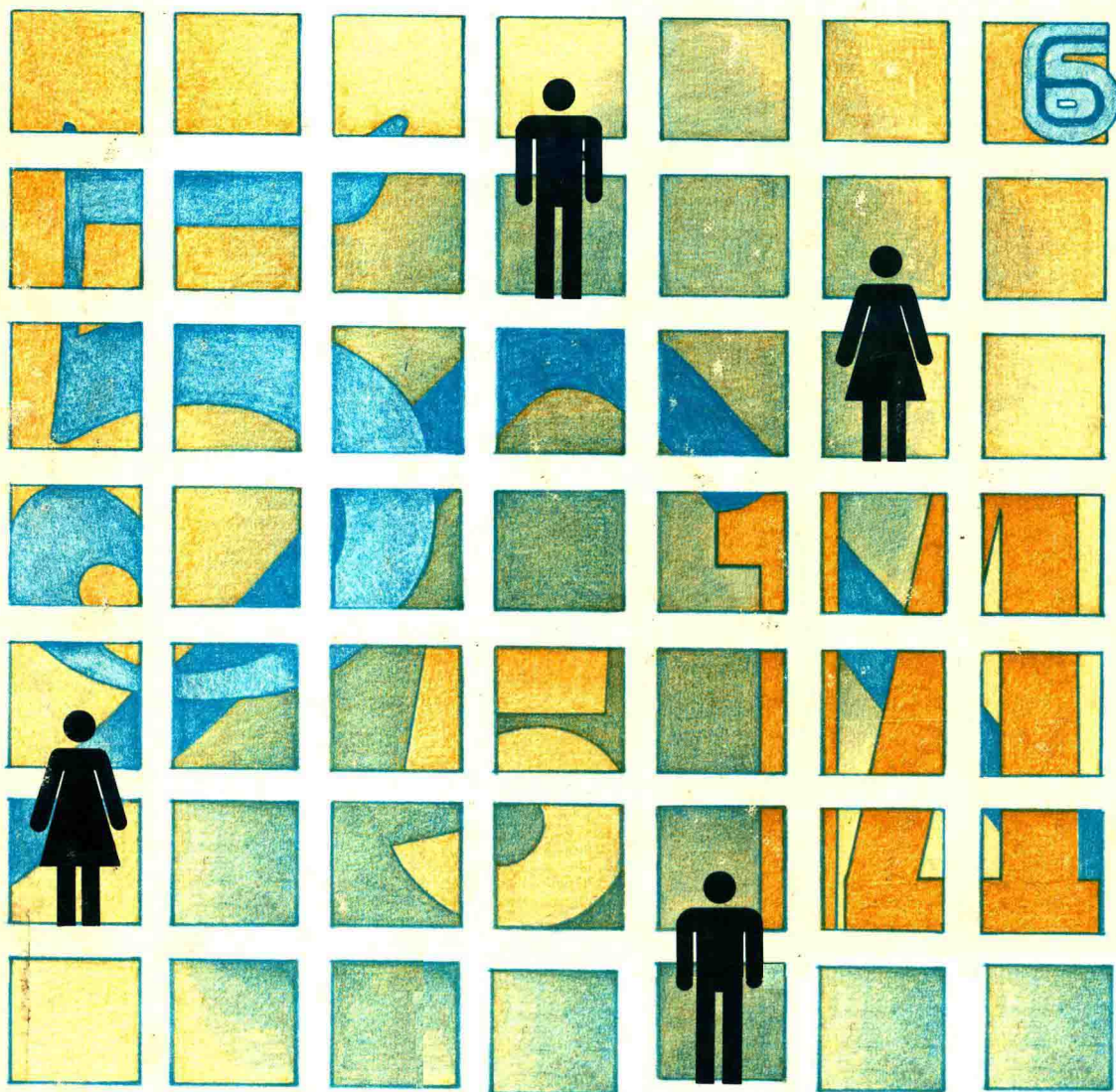


INTRODUCTION TO
**SOCIAL
STATISTICS**

NORMAN R. KURTZ



INTRODUCTION TO **SOCIAL STATISTICS**

NORMAN R. KURTZ

Professor of Social Research
Brandeis University

McGRAW-HILL BOOK COMPANY

New York St. Louis San Francisco Auckland Bogotá Hamburg
Johannesburg London Madrid Mexico Montreal New Delhi
Panama Paris São Paulo Singapore Sydney Tokyo Toronto

INTRODUCTION TO SOCIAL STATISTICS

Copyright © 1983 by McGraw-Hill, Inc.

All rights reserved. Printed in the United States of America.

Except as permitted under the United States Copyright Act of 1976,
no part of this publication may be reproduced or distributed in any form
or by any means, or stored in a data base or retrieval system,
without the prior written permission of the publisher.

234567890 DOCDOC 89876543

ISBN 0-07-035676-9

This book was set in Souvenir Light by A Graphic Method Inc.

The editors were Eric M. Munson, Christina Mediate, and Claudia Tantillo;

the designer was Anne Canevari Green;

the production supervisor was Charles Hess.

The drawings were done by VIP Graphics.

The cover illustration was done by Anne Canevari Green.

R. R. Donnelley & Sons Company was printer and binder.

Library of Congress Cataloging in Publication Data

Kurtz, Norman R.

Introduction to social statistics.

Includes bibliographies and indexes.

1. Social sciences—Statistical methods.

2. Statistics. I. Title. II. Title: Social
statistics.

HA29.K935 1983 519.5'024301 82-13056

ISBN 0-07-035676-9

INTRODUCTION TO SOCIAL STATISTICS

To
Loretta, Richard, and Jeannie

PREFACE

The purpose of this book is to introduce you to statistical procedures used by social researchers. Inasmuch as the book is intended for the introductory level, do not expect too much, since you simply cannot learn everything there is to know about statistics from an introductory text. If your ultimate goal is to become competent in the use of sophisticated statistical techniques, you will have to go beyond what is offered here. However, the material presented here will give you the kind of foundation that will make it easy for you to master more advanced material.

While you should not expect too much from this book, do not expect too little. The book is designed to provide a self-contained set of ideas and techniques that should be of substantial use even if you have no plans to go beyond introductory-level statistics. It will give you sufficient background to engage in statistical descriptions and carry out analyses of social data that are of interest to you. The material is also designed to assist you in reading and evaluating social research reports more critically. It is also true that more and more of the information we must digest to make intelligent everyday choices in our roles as consumers, employers, investors, voters, and community members is based on statistically oriented reports. The material in this text should better enable you to make these decisions.

Whatever your objectives in studying introductory statistics, a text should do two things for you. First, it should familiarize you with the rationale around which statistical work and decision making is organized. Statistics, like other formal analytic approaches for interpreting the world, represents a particular way of thinking. In order to use and understand statistics properly you must be acquainted with the conceptual foundations on which statistical thinking is based. Special effort has been made here to elaborate concepts and show how they fit together to make up the logical models of statistics.

The second thing an introductory text should do is help you learn the mechanics of applying and interpreting statistics. This means that you should learn how various statistics are calculated, which statistics are most appropriate for a given set of data, and what the precise meaning of the statistical result is. While the calculation aspect of statistics is perhaps most often seen as responsi-

ble for students' lack of enthusiasm for enrolling in statistics courses, you will find that in this text calculations have been reduced to relatively simple arithmetic. And care has been taken to illustrate in detail the steps that must be followed in calculating a statistic. Emphasis has also been placed on how to determine what the most appropriate statistic is for a data set. And concrete examples have been developed which help to illustrate the use and meaning of different statistical procedures.

The chapters in the text have been designed to ease mastery of the material. A topic outline has been placed at the opening of each chapter to give you a road map of the topics the chapter will cover. You can also look at the outline as a summary of the new ideas and techniques that are discussed. At the end of each chapter you will find a section titled Key Terms, Symbols, and Formulas that contains a summary and definitions of new terms, symbols, and formulas introduced in the chapter. These summaries provide a convenient reference to the material in the chapter as well as a guide in preparing for examinations. Each chapter also contains a set of exercises you should complete to help internalize the material. Answers to the exercises are presented at the end of the book. Finally, each chapter has a list of references to other texts on statistics. The references are accompanied by brief comments indicating the particular contribution that is to be gained from them.

The text consists of 13 chapters. The first chapter has the dual purpose of showing how statistical work fits into the overall process of research, as well as presenting some preliminary comments and reviews of the basic arithmetic that will be used in the book. Chapters 2 and 3 contain techniques for descriptive statistics. Chapter 2 discusses measurement and various methods for displaying data, whereas Chapter 3 presents summary descriptions including measures of central tendency, the mode, median, and mean, and measures of variation, that is, measures showing the differences among sample members.

Chapters 4 through 6 provide an introduction to statistical inference or the process whereby the results of studies of small samples can be generalized to large populations. Chapters 4 and 5 present two different models for determining the probability of a given sample result, and Chapter 6 details the basis for generalizing the observations on relatively small samples to large populations.

Chapter 7 details the procedures for testing hypotheses with statistical measures, and Chapter 8 shows how to determine whether a sample represents a good fit to the population it is presumed to represent. Chapters 9 and 10 show how to compare two samples to find out whether the populations they represent are similar. Chapter 11 extends the concept of comparing samples to three or more samples using methods of analysis of variance.

The last two chapters introduce measures of association. Regression analysis and measures of correlation are presented in Chapter 12, and Chapter 13 presents a variety of additional measures of association.

Writing a text like this obviously requires the support and interest of a great many people. Most important on that list is my wife, Loretta, who provided the

encouragement and concern that is needed on a day-to-day basis. My children, Richard and Jeannie, also played a vital supportive role. It would be impossible to detail all those who were instrumental in the completion of the work, but the foremost influence was that of my students at the Florence Heller Graduate School at Brandeis University, who throughout the years have been patient and consistent in providing the challenge and atmosphere necessary to encourage the development of clear expression and meaningful organization of the subject matter. I also wish to thank Judson Pearson of the University of Colorado who first introduced me to statistics, and Howard E. Freeman of the University of California, Los Angeles, who first encouraged me to write a statistics text. The comments and criticisms of my reviewers were most helpful, including those of Frank D. Bean, University of Texas, Austin; Sheldon Eklund-Olson, University of Texas, Austin; Howard Fradkin, California State University, Long Beach; Rebecca F. Guy, Memphis State University; Anita S. Harbert, San Diego State University; Anthony A. Hickey, George Mason University; Sidney Kronus, University of North Carolina, Asheville; David D. McFarland, University of California, Los Angeles; James R. Marshall, State University of New York, Buffalo; Kathleen Natalino, Cleveland State University; David E. Payne, University of North Dakota; David J. Pratto, University of North Carolina, Greensboro; Kenneth M. Rothrock, Bowling Green State University; Marsha M. Seltzer, Boston University; and Kenneth Southwood, University of Illinois, Urbana.

I am also indebted to the editorial staff of McGraw-Hill Book Company, who were most diligent in guiding me through the process including Eric Munson, Claudia Tantillo, Christina Mediate, and James W. Bradley. Finally, I wish to thank Frances Buckley who went above and beyond the call of duty in typing, assembling, and accomplishing all the other necessary work that transforms a semilegible draft into a finished product.

Norman R. Kurtz

CONTENTS

PREFACE

xiii

Chapter 1	THE SUBJECT OF SOCIAL STATISTICS	1
	Introduction	1
	What Is Statistics About?	1
	How Important Is Statistics to Social Research?	4
	How Important Is Mathematics to Statistics?	7
	How Does the Computer Deal with Data?	7
	A Review of Basic Arithmetic Procedures	10
	Summary	16
	Key Terms, Symbols, and Formulas	17
Chapter 2	MEASUREMENT AND DATA DISPLAY	21
	Introduction	21
	What Is Measurement?	21
	Displaying Data	27
	Summary	46
	Key Terms, Symbols, and Formulas	48
Chapter 3	NUMERICAL SUMMARIES OF DISTRIBUTIONS	56
	Introduction	56
	Measures of Central Tendency	57
	Measures of Variation	66
	Summary	77
	Key Terms, Symbols, and Formulas	78

Chapter 4	A CONTINUOUS PROBABILITY DISTRIBUTION: THE NORMAL DISTRIBUTION	84
	Introduction	84
	The Role of Probability	85
	The Normal Probability Distribution	88
	Skewness and Kurtosis	95
	Summary	98
	Key Terms, Symbols, and Formulas	99
Chapter 5	A DISCRETE PROBABILITY DISTRIBUTION	105
	Introduction	105
	Characteristics of the Binomial Distribution	105
	The Model and Its Application	106
	Transforming the Binomial to Continuous Form	112
	Summary	115
	Key Terms, Symbols, and Formulas	116
Chapter 6	THE BASIS OF STATISTICAL INFERENCE	119
	Introduction	119
	Random Sampling	119
	Accuracy and Error in Sampling	121
	The Logic of Statistical Inference	123
	Point Estimates of μ	129
	Interval Estimates of μ	131
	Confidence Limits and Sample Size	134
	Confidence Limits for Proportions	136
	Summary	137
	Key Terms, Symbols, and Formulas	138
Chapter 7	TESTING HYPOTHESES	144
	Introduction	144
	What Is a Hypothesis?	144
	The Level of Significance	146

Testing the Hypotheses	148
The Risk of Error	151
Summary	155
Key Terms, Symbols, and Formulas	155
 Chapter 8 TESTING THE GOODNESS OF FIT	 159
Introduction	159
Parametric and Nonparametric Tests	160
The Goodness of Fit for Continuous Variables: Parametric Tests	160
The Goodness of Fit of Discrete Variables: Nonparametric Tests	167
Summary	173
Key Terms, Symbols, and Formulas	173
 Chapter 9 COMPARING TWO CONTINUOUS MEASURES: PARAMETRIC TEST	 178
Introduction	178
Some Background on Group Comparisons	179
Comparing Two Independent Measures	181
Comparing Two Dependent Measures	188
Summary	193
Key Terms, Symbols, and Formulas	193
 Chapter 10 COMPARING TWO DISCRETE MEASURES: NONPARAMETRIC TEST	 200
Introduction	200
Comparing Two Independent Ordinal Measures	201
Comparing Two Dependent Ordinal Measures	209
Comparing Two Independent Nominal Measures	212
Comparing Two Dependent Nominal Variables	218
Summary	220
Key Terms, Symbols, and Formulas	220

Chapter 11	COMPARING MORE THAN TWO MEASURES: ANALYSIS OF VARIANCE	227
	Introduction	227
	One-Way ANOVA: The F Test	228
	Two-Way ANOVA	234
	Kruskal-Wallis One-Way ANOVA	243
	The Friedman Two-Way ANOVA	245
	Summary	247
	Key Terms, Symbols, and Formulas	247
Chapter 12	LINEAR REGRESSION AND CORRELATION	253
	Introduction	253
	Regression Analysis	254
	Correlation Analysis	270
	Summary	282
	Key Terms, Symbols, and Formulas	283
Chapter 13	TESTING THE ASSOCIATION OF DISCRETE DATA	289
	Introduction	289
	Measures of Association	289
	Measuring the Association of Ordinal Data	293
	Measuring the Association of Nominal Data	300
	Summary	306
	Key Terms, Symbols, and Formulas	307
	SELECTED ANSWERS TO EXERCISES	312
Appendix	Table 1: Areas under the Normal Curve	318
	Table 2: Areas under the Binomial Curve	319
	Table 3: Critical Values of τ	322
	Table 4: Critical Value of D for the Kolmogorov-Smirnov One-Sample Test	323

Table 5: Critical Values of Chi-Square	324
Table 6a: Critical Values for the Mann-Whitney U Test (For Samples Where neither n_1 nor $n_2 < 8$)	325
Table 6b: Critical Values for Mann-Whitney U Test (For Samples Where $n_1 \leq 20$ and $n_2 > 9 \leq 20$)	328
Table 7: Critical Values for the Kolmogorov-Smirnov Two-Sample D	332
Table 8: Critical Values for Wilcoxon T	333
Table 9: Distribution of F	334
Table 10a: Critical Values for Correlation Coefficients	338
Table 10b: Critical Values of r_{rho} (Rank-Order Correlation Coefficient) n , Number of Pairs	339
Table 11: Transformation of r to z	340
Table 12: Random Numbers	342
Table 13: Squares and Square Roots	346
 INDEXES	 359
Name Index	
Subject Index	

CHAPTER 1

THE SUBJECT OF SOCIAL STATISTICS

INTRODUCTION

WHAT IS STATISTICS ABOUT?

HOW IMPORTANT IS STATISTICS TO SOCIAL RESEARCH?

HOW IMPORTANT IS MATHEMATICS TO STATISTICS?

HOW DOES THE COMPUTER DEAL WITH DATA?

A REVIEW OF BASIC ARITHMETIC PROCEDURES

- BASIC SYMBOLS
- THE USE OF SIGNED NUMBERS
- EXPONENTS
- SQUARES AND SQUARE ROOTS
- FACTORIALS
- ROUNDING DECIMALS

SUMMARY

INTRODUCTION The purpose of this book is to introduce you to the subject of social statistics. Before launching into the specifics of the subject, however, you should have a general idea of what statistics is about and the role it plays in the process of social research. This chapter is intended to provide you with such an overview. In addition the chapter will discuss the use of mathematics in statistics so that you will know what to expect, and it will provide you with a brief review of the basic mathematical symbols and procedures you should be familiar with.

WHAT IS STATISTICS ABOUT?

Statistics is a body of methods; these methods are used to assemble, describe, and analyze numerical data pertaining to various aspects of social life. So statisticians are concerned with the same topics that are of interest to the social sciences in general. What makes statistical work unique is its focus on numerical representations of social life, that is, on things that can be counted, measured, or quantified. For example, statistics are used to describe such things as the number of members in an average family, the relationship between family size and

amount of income, and the correspondence between *amount* of family income and *measures* of attitudes.

The methods of statistics can be divided into two categories, descriptive statistics and inferential statistics. *Descriptive* statistics consist of exactly what the name implies: *They are methods for describing social phenomena* and, more specifically, they provide *numerical* descriptions. As you likely know from your own experience, providing an exact description can be deceptively difficult unless you follow a well-organized set of procedures. Suppose, for example, that you were asked to describe your home community in terms of three basic characteristics—sex, age, and family income. The purpose of the description is to provide a good understanding of the number of males and females, the age structure of the community, and the proportion of families that fall into various income categories.

Each of the three characteristics is a *variable*, that is, each is an entity that *can take different values* for the members of your community. Thus sex is a variable that has one of two possible values, male or female; age can have values ranging from 1 day to more than 100 years; and income can take on a wide range of values from very low to very high earnings. The *variation* in values for a given characteristic is the primary concern of statistical description. The total set of values for a particular characteristic is known as the *distribution of the variable*. Each variable has a specific distribution for a given population. For example, the members of your home community are distributed across the values of male and female in a specific form, and this form is referred to as the *distribution of males and females*. If your home community consists of 39 percent males and 61 percent females, this is its distribution on the variable of sex. Similarly, the members of your community represent a specific distribution on the variables of age and income. These distributions differ from that of sex because they represent more than two values. If age, for example, is reported in years, the distribution would have values all the way from 1 year and less to the age of the oldest community member.

Descriptive statistics is the body of methods used to assemble, organize, and display distributions of variables (also referred to as data). An important goal of this book is to familiarize you with these descriptive techniques. The presentation is intended not only to enable you to better understand the statistical descriptions presented by others but will also prepare you to develop your own descriptions of those aspects of social life that interest you.

The second category of statistical methods is *inferential statistics* which are used to *infer the characteristics of a population from observations made on a sample*. Inferential statistics make it possible to describe large populations using information from relatively small samples representing the populations. The great majority of social research is based on inferential statistics, because ordinarily the study of large populations is too difficult and expensive.

Returning to the study of your community, suppose it consists of 4000 people. Even for such a relatively small community it would be almost impossible for

you to contact every one of the 4000 members and ask them about sex, age, and family income. A much easier course of action would be to select a small sample, perhaps as few as 250, collect the required information from them, and based on this information infer what the distributions for the three variables are for the total population. If the sample is selected properly and correct procedures followed to derive the information, an accurate description of the population will be obtained.

While numerical descriptions of samples and populations serve a somewhat similar purpose, it is important to maintain a distinction between them. As we shall see in studying statistical inference, population values are fixed. For example, the number of males and females in your community at any point in time is a fixed value, such as 39 percent male. Such numerical descriptions of populations are referred to as *parameters*.

Sample descriptions, however, are not fixed values in that, if you draw several samples from the same population at the same point in time, the value provided by each sample may differ slightly. Thus one sample may show 38.2 percent males, another 39.8 percent, and so on. The techniques of statistical inference are specifically concerned with evaluating the meaning of the variation in sample estimates. The numerical descriptions derived from samples are referred to as *statistics* and are *estimates of population parameters*.

Researchers are concerned not only with estimating the values for variables describing one population, but often wish to compare the values of variables for two or more populations. Thus, you may wish to compare the sex ratio of your community to that of two neighboring communities. Suppose you did that and found the following data.

	Your community, %	South community, %	North community, %
Male	39.0	47.1	51.2
Female	61.0	52.9	48.8
	100.0	100.0	100.0

If the data were based on samples from the three communities, methods of statistical inference would enable you to determine whether the populations represented by the three samples really differed in terms of their sex distributions.

The use of inferential procedures perhaps most familiar to the general public relates to reports of election outcomes in which pollsters predict elections results with amazing accuracy when 5 percent or fewer of the ballots have been counted. They are able to do this because the use of small samples makes it possible to estimate accurately how the population at large will vote.

A major goal of this text is to introduce you to the methods and logic that make statistical inference possible. You will not only learn about the procedures that undergird the inferential process, but you will also find that methods are

available for determining the margin of error that is likely to accompany a particular inference.

HOW IMPORTANT IS STATISTICS TO SOCIAL RESEARCH?

While the general purpose of statistics is to describe characteristics and relationships within and between populations, statistical methods represent a small part of a broader set of activities referred to as social research. But the part that statistics plays is important. An example can serve to illustrate the role of statistics in the research process. While the example simplifies the actual way in which research is done, it will give you a better idea of why statistics is so important.

Reference group theory suggests that membership in a group furnishes individuals with frames of reference that among other things provide them with a basis for self-evaluation. A researcher, Lorenz, wondered whether membership in a racial group and the frame of reference it provides affect the way in which individuals evaluate their potential for future economic success.* Her theory suggests that members of reference groups experiencing lower overall success, even if it is due to an external force such as racial discrimination, are likely to have lower aspirations for future economic achievements.

An important component of social research is developing and evaluating theoretical schemes that provide explanations for social behaviors. Lorenz, who was curious about potential differences in the aspirations of blacks and whites, found that reference group theory provided a possible explanation.

However, the proposition that reference group membership affects future economic aspirations is an abstract statement, and in order to evaluate its efficacy, it must be reduced to elements that can be measured. This means that variables must be identified that correspond in some way to the abstract terms of the theoretical proposition. Thus, for example, Lorenz defined the abstract term *reference groups* in terms of the variable of membership in a racial group. Only two racial groups, blacks and whites, were involved, thus the variable of race had only two values. Aspirations were defined in terms of whether individuals felt they would own a home someday. (Actually, Lorenz used several measures, but for the sake of our example only one measure, aspirations for future home ownership, will be considered.)

The process of moving from abstract concepts to measurable variables is called *operationalization*. It consists of specifying the operations that must be performed to observe or measure referents of abstract concepts. In our example,

*The example used here is based on a study reported by Gerda Lorenz, "Aspirations of Low-Income Blacks and Whites: A Case Study of Reference Group Processes," *American Journal of Sociology*, **78**:371–398 (1972). While the discussion which follows draws on the material from the article, the presentation represents a considerable simplification of the study.