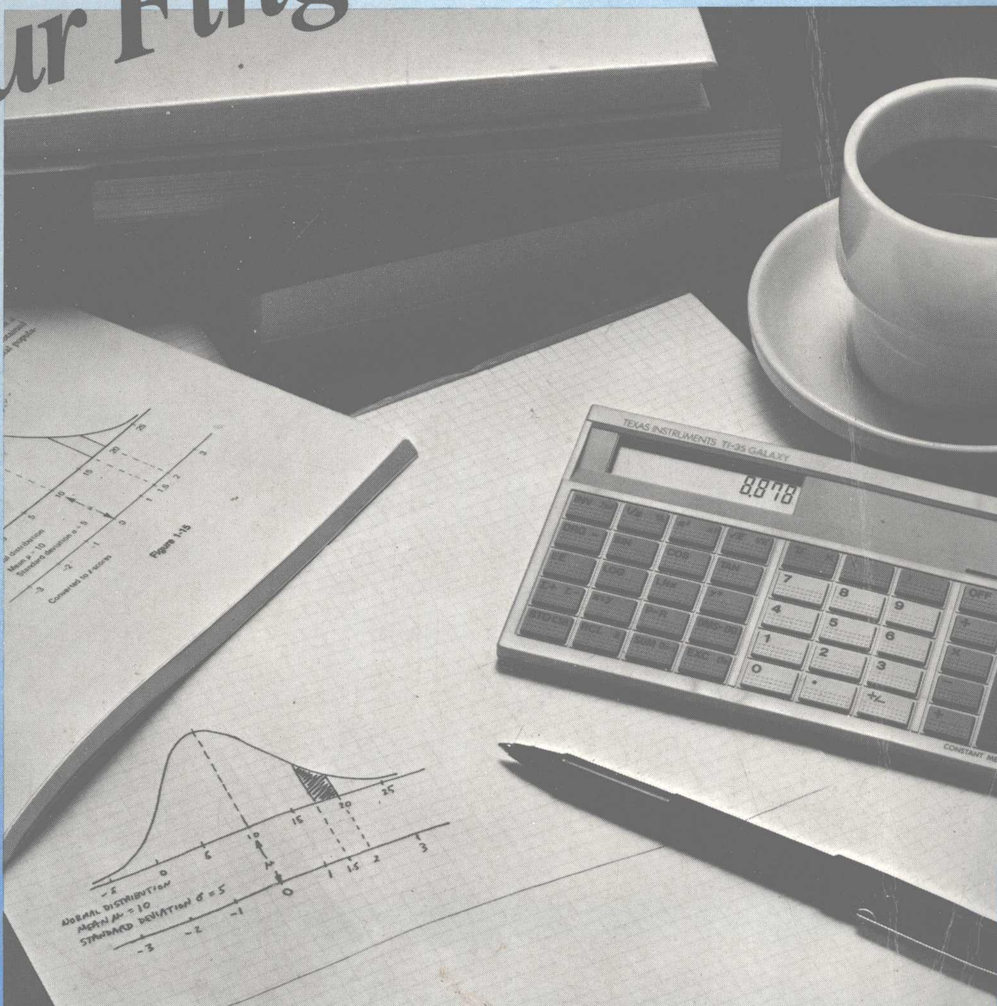
STATISTICS

At Your Fingertips



MARK FINKELSTEIN

Statistics at Your Fingertips

Mark Finkelstein

University of California, Irvine

Wadsworth Publishing Company

Belmont, California

A division of Wadsworth, Inc.

Statistics Editor: Jim Harrison
Signing Representative: Cindy Champagne
Production Editor: Phyllis Niklas
Designer: Rick Chafian
Copy Editor: Carol Dondrea
Cover: Gary Head

Printed in the United States of America
1 2 3 4 5 6 7 8 9 10 89 88 87 86 85
ISBN 0-534-04023-3

Library of Congress Cataloging in Publication Data

Finkelstein, Mark.
Statistics at your fingertips.

Rev. ed. of: Calculate basic statistics/Mark Finkelstein, George McCarty. c1982.
Includes index.
1. Statistics—Data processing. 2. Calculating-machines. I. Finkelstein, Mark. Calculate basic statistics. II. Title.
QA276.4.F54 1985 519.5'028'5 84-19655

ISBN 0-534-04023-3

How to Use this Book

If you've bought this book intending to learn this material on your own, you've taken the right step. I wrote this book with you in mind. (If you're taking a first course in statistics you'll have the guidance of your instructor, so you won't need much guidance from me. But I'm going to give you some anyway.) I wrote this book following many years of experience of teaching "math-phobic" students—students who had the capabilities within themselves of learning this (or more difficult) material, but who somehow thought they didn't. As soon as they saw a symbol they didn't recognize, or an expression they didn't know how to interpret, they threw up their hands and said, "See, I really can't do this, after all!" But some perseverance and, subsequently, some success changed all that: "I really can do this, after all!" Many have been here before you. This material isn't all that difficult, and I've taken special pains to make the explanations clear, and leisurely. It means you have more pages to read through, but you'll find that you will be successful when you do read through them.

THE CALCULATOR YOU NEED TO GET STARTED

In addition to your pencil, you need a calculator. Any model will do—even the cheapest. The only essential functions are addition, subtraction, multiplication, and division. But many brands and models have special features and functions that could be useful to you in your work.

Square-root and *memory* keys are sometimes found on even the calculators sold at drugstores or grocery stores. It will be worth your effort to look for them if you're buying a new machine.

There are *statistical calculators*, such as Texas Instruments's TI-35, which have a special key marked [Σ]. This function can cut your button-punching in half, yet the cost of such a calculator is low. Of course, there are much fancier models, whose cost may run to several hundred dollars, and which are programmable. Such machines can be well worth their cost if you frequently do large or complex statistical calculations. You may decide to get one when you finish your study of this book. But a programmable model is definitely not necessary here.

HOW TO USE A COMPUTER AS A CALCULATOR

Let's say that you're in the unusual position of owning a $1000 home computer but lack a $6.95 drugstore calculator. Your computer is almost surely equipped with BASIC (they all are). The single symbol ? is an instruction to BASIC to print whatever follows. If whatever follows the question mark is an arithmetic expression, the computer will evaluate it and display the result on the screen:

?(3 + 4)/10

will cause the computer to display 0.7 .

?SQR(4 + 6 − 1)

will cause the computer to display 3 , since SQR stands for square root. All the arithmetic symbols are as you expect them to be, except that exponentiation is "up arrow": ↑ . So if you want to compute 3^2, type

? 3 ↑ 2

and the computer will display 9 . You may make the expressions as complicated as you like. Type

? 2 ↑(10*7 − 8*8)

and the computer will display 64 . Be sure to try these expressions on your computer.

THE LAYOUT OF THE BOOK—EXAMPLES, EXERCISES, PROBLEMS, AND QUIZZES

Everybody knows what examples are. I've made it a point to work out at least one example on each topic that I talk about in the book. I work these examples out for you, down to the last detail. I do this because my experience has been that you never can tell which detail will be the one to "hang the student up," and I want you to be successful. Hopefully, there's enough detail included so that you can see your way through each example, from beginning to end. As for the exercises and problems, the difference between them is that I have completely worked out solutions to the exercises, and these are found at the end of each chapter. The problems have no answers included at all, and so they're suitable for assignment by instructors in a course, or possibly for use on exams. There are usually half a dozen to a dozen problems on each topic. At the end of each chapter I've included a chapter quiz, which is just a little self-test to allow you to check if you've understood the main concepts in the text. Since it's a self-test, I've also included the answers at the end of each chapter. By all means, take advantage of the chapter quizzes. It only takes a couple of minutes to work through each quiz and you'll feel really good when you find you got all the answers right.

THERE IS JUST ONE INSTRUCTION

Do all the exercises. There aren't that many of them—usually just one per topic. You can't possibly learn statistics without *doing* statistics, and doing statistics means working out problems with data. You can't learn statistics by watching me do it any more than you could learn to play the violin by watching me do that. So plan now that you're going to do some work. I've included detailed solutions to each of the exercises (as opposed to the problems) so you can do more than check that you were right or wrong—if you were wrong, you can compare your work with mine, line by line, and *see* where you went wrong. Only by a process like this can you actually learn to do statistics, and learning statistics means learning to *do* statistics. I've done my part, and if you do yours, you'll end up learning statistics. I promise you that *you cannot fail.*

WHAT TO DO WHEN YOU ARE BAFFLED

When you make no sense out of an example or an exercise, the reason is always that you do not understand what every one of the words and symbols means. This is no reason for panic—there is a simple method for handling the situation. Starting at the beginning, double-check each word's definition and each symbol's meaning, one by one. This is easy to do. The first time a new word or symbol is used in this book, it is printed in blacker "boldface" type so you'll notice it. Then it is carefully explained and an example is given of its use. If you have trouble finding this definition, check the Index or use the Quick Word Guide beginning on page 281, the Quick Symbol Guide on page 289, or the Quick Formula Guide on page 290. We guarantee that if you check out all the definitions so you know exactly what is being asked for, *you can work every exercise.*

HOW TO USE THIS BOOK TO REVIEW STATISTICS

If you have already studied statistics, and you want to refresh your understanding of just one topic, look it up in the Index to get started. But this method can be discouraging, so be careful. Suppose your topic is ANOVA, and you find in the Index (or the Contents) that Chapter 6 discusses this technique. You might turn to Chapter 6 and make no sense out of what you read. When this happens, heed the advice above for "what to do when you are baffled."

It will help you in checking back for review to know how topics fit together in this book. Chapters 1, 2, 3, and 4 need to be studied in that order—2 depends on 1, 3 on 2, and 4 on 3. But each of the later chapters, 5, 6, 7, 8, and 9, depends only on the first four. In the example above, of ANOVA, if you aren't prepared to read Chapter 6, do not look at Chapter 5 first. Instead, see if you can do the exercises in Chapter 4. If not, go back to Chapter 3, and so on.

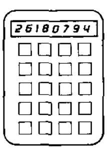
Introduction

YOU CAN'T FAIL AT LEARNING STATISTICS

By buying this book, you have already taken the first step toward mastering statistics. And now, starting with the problem below, you will begin to learn what statistics is all about.

Take the following three numbers

33	45	48



and add them up. Then divide the total by 3. You can work in the space below, if you are going to do it by hand. If you are using a hand calculator, enter the three numbers and add them up; then divide by 3.

Did you get 42 for your answer? Congratulations! You've just observed your first statistic: the *mean*, or average of a *sample*, called simply the *sample mean*. (If you didn't get 42, you either added wrong or divided wrong. Try adding the numbers again and see if you get 126. Then divide 126 by 3 and see if you get 42. If you still don't get the right answer, look at the next page, where I've done the arithmetic for you. Compare your work with mine to see where you went wrong.)

The reason we divide by 3 is that there are 3 numbers in our sample. When we compute the mean, or average, we add up all the numbers in our sample and then divide by the number of numbers. To compute the sample mean of 10 numbers, we add up all 10 numbers and then divide by—you guessed it!—10.

The sample mean is probably the most frequently used statistic. We'll be using it again and again in a variety of settings. But that will come later, after you learn what a statistic is, and what a sample is!

```
   33
   45
 + 48
 ----
  126

     42
   -----
 3 ) 126
     12
     --
      6
      6
      -
```

$$\bar{x} = \text{Sample mean} = \frac{33 + 45 + 48}{3}$$

$$= \frac{126}{3} = 42$$

($\bar{x}$ is the symbol we use for the sample mean.)

WHAT THIS BOOK IS ABOUT

Statistics can be divided into two parts: descriptive statistics and inferential statistics. Descriptive statistics deals with collections of numbers, called populations, and the various ways of describing populations. Populations can be described pictorially, by means of bar charts or pie charts, or numerically, by population parameters. Descriptive statistics deals principally with the numerical descriptions—population parameters.

Inferential statistics is the science of making inference, or drawing conclusions, about population parameters when only a portion of the population is available for consideration. Given information about a portion of the population, called the *sample*, we draw conclusions about the population as a whole or about one or more of the population parameters. The development of this science was stymied for a century by the inescapable fact that the conclusions we draw about the population parameters will occasionally be wrong. For years scientists searched for methodology that would allow *some* conclusions to be drawn that would be correct *all* the time. Finally, in the 1920s, it was determined that error was inherent in the nature of inferential statistics; that the error could always be made as small as desired (by paying some price in time or effort), but could never be eliminated. It took several years for this radical notion to gain acceptance in the scientific community, but today it is recognized as the price one has to pay for inferring something about the whole (the population) by studying only a part (the sample).

In this book I will show you how to make inferences about the population mean from the study of two statistics: the sample mean and the sample variance. This one topic contains the main ideas of inferential statistics.

We will then go on to inferences about other parameters from other statistics: chi-square, the *F*-ratio, and *r*. I will also introduce a different method in inferential statistics—that of hypothesis testing. Instead of directly making inferences about a population parameter from a sample statistic, we formulate a hypothesis about a population parameter and then use the sample statistic to *test* the hypothesis (and accept or reject it).

Contents

1 Descriptive Statistics: Part 1

Terms we'll be learning about in this chapter

fractile
histogram
mean
median
mode
parameter
percentile
population
statistic

1-1 WHAT STATISTICS IS ALL ABOUT

Statistics is a scientific method for both asking and answering questions about the "real world." These questions are usually of the form, "What is the average annual income of families in this city?" or "Do brand *A* tires last longer than brand *B* tires?" or "Is this drug an effective treatment for this disease?"

While these questions appear to be quite different, they all share something in common: they all ask questions about *numbers*. In the first question, the numbers are the dollars of income of each family. In the second question, the numbers are the miles driven on each kind of tire before it wears out. And in the third question, the numbers are 0s attached to each person who didn't get cured and 1s attached to each person who did. (Don't worry if you can't make sense out of the 0s and 1s at this time.)

Getting back to our first question, we can consider the population we are studying as the set of families in this city. Attached to each object in that set (each family) is a *number*: the number of dollars earned by that family in one year. What we really want to study is the collection of numbers attached to the set of families. Actually, the collection of numbers is not exactly a set, because if two families have the same income, that number would be repeated twice in the collection of numbers. Strictly speaking, the collection of numbers (incomes) we are going to study is a set, together with multiplicities (the two or more copies of a single element, or number).

Our questions about family incomes are really questions about the *numbers* attached to the population of families. We will use the term **population** both for the set of families (the objects) and for the set (with multiplicities) of *numbers* attached to the objects. But usually we use the term *population* to refer to a *set of numbers*.*

Since we will formulate all our statistical questions in terms of populations, we need to develop ways of describing these populations. This area of study is called **descriptive statistics**.

1-2 DESCRIBING POPULATIONS

Some populations are so small we can "comprehend" the entire population, such as the population of coin values (in cents) in my pocket right now. That population is

1, 1, 1, 5, 10, 10, 25, 25

It consists of eight coins, with a total value of 78¢.

But some populations are too large to enumerate comfortably, such as the family incomes in this city. There may be as many as 20,000 families, so the population

*To the mathematically sophisticated reader: As you know, a *rule* that attaches a number to each object in some set is really a numerically valued **function** defined on the set. Talking about means, variances, and so on, is really speaking about means and variances of the function. In statistics these functions are called **random variables** (even though there is nothing "random" about them). I won't be using this terminology any more, but if you run into it in another book, you'll know what is being talked about.

would consist of 20,000 elements! And even if we listed all 20,000, we really wouldn't learn much, because all we would have are pages and pages of numbers.

There are many different ways of describing a population without listing all the elements, but basically they fall into two categories: *pictorial descriptions* and *numerical descriptions*. The pictorial descriptions are things like *bar charts* and *pie charts*. The numerical descriptions are called **population parameters**. First we'll deal with the pictorial descriptions. It's a corny old saying, but true: one picture is worth a thousand words.

HISTOGRAMS

Histogram is just another name for *bar chart*, and we'll use the two expressions interchangeably. Just exactly what the heights of the bars mean in Figure 1-1, and what the "cutoff points" for the various classes are, we won't bother specifying at this time. But just from looking at the bar chart in Figure 1-1, we know that the majority of families in that city earn between $15,000 and $30,000 and that few families earn less than $5000. Compare this bar chart (histogram) with the ones shown in Figures 1-2 and 1-3, one for a "rich city" and one for a "poor city." Even

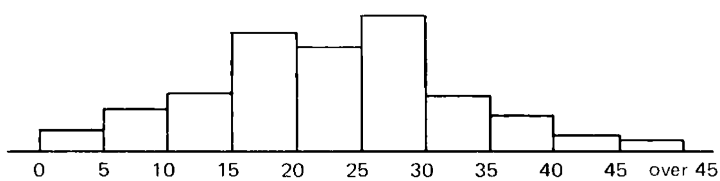


Figure 1-1 Distribution of family incomes in this city, 1982, in thousands of dollars per year.

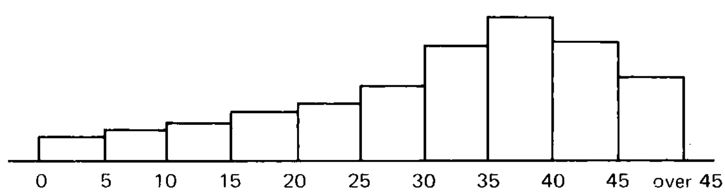


Figure 1-2 "Rich city" incomes in thousands of dollars per year.

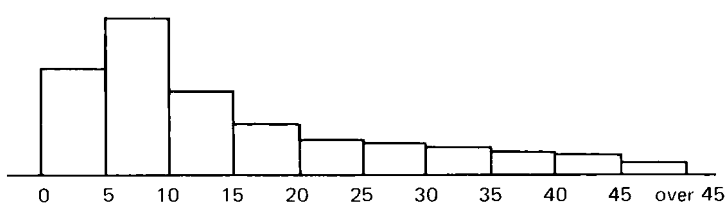


Figure 1-3 "Poor city" incomes in thousands of dollars per year.

without labels, we can tell at a glance which is which. If we were given a list of 20,000 numbers for each of the three cities, however, we would have to spend quite a bit longer figuring out which city was which.

Figures 1-4 and 1-5 show two more examples of histograms; this time the data are taken from the *World Almanac*. Again, we can get quite a bit of information even without looking at the numbers.

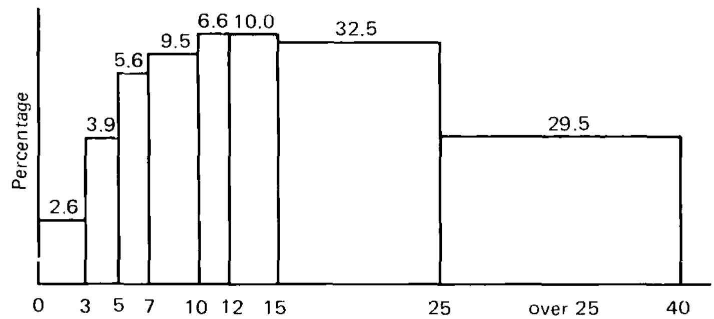


Figure 1-4 White family annual incomes in the United States, 1978, in thousands of dollars. (Data from the *World Almanac*, 1981.)

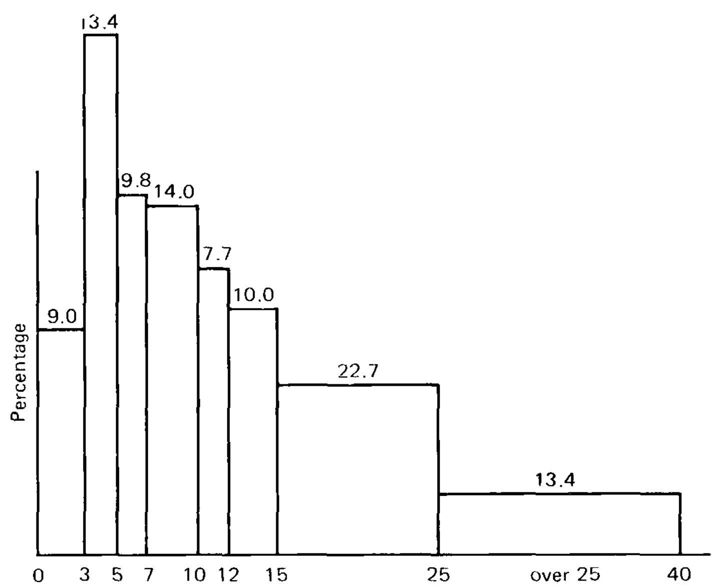


Figure 1-5 Black family annual incomes in the United States, 1978, in thousands of dollars. (Data from the *World Almanac*, 1981.)