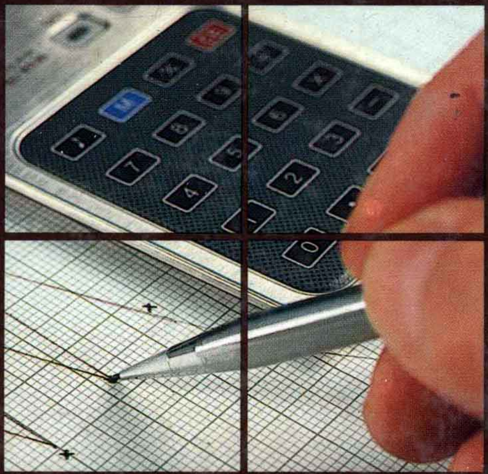
Runyon / Haber

Business Statistics



Business Statistics

Richard P. Runyon

Audrey Haber

1982

RICHARD D. IRWIN, INC. Homewood, Illinois 60430
Irwin-Dorsey Limited Georgetown, Ontario L7G 4B3

ISBN 0-256-02655-6

Library of Congress Catalog Card No. 81-81549

Printed in the United States of America

1 2 3 4 5 6 7 8 9 0 H 9 8 7 6 5 4 3 2

Dedicated to
Lois Runyon
and
Jerome Jassenoff

Preface

It is our belief that statistics should be one of the most enjoyable and mind-expanding courses in the curriculum. After all, statistics and probability relate to and impact on virtually every aspect of our daily lives, professional and personal. We are continuously assessing information (perceptual as well as numerical), summarizing it, drawing conclusions, and acting upon these conclusions in the countless decisions we make each and every day. Statistics, the discipline, attempts to bring some order, coherence, and a degree of formality to procedures we routinely use to lessen the degree of uncertainty in our knowledge base, the validity of our conclusions, and the cogency of our daily decisions and actions.

In other words, statistics is relevant in our lives. But more than that, the subject is itself inherently interesting. It deserves being presented in a lively and vibrant manner without sacrificing the mathematical and logical integrity of the subject. This we have attempted to do. How well we have succeeded is for others to judge.

Following are some of the features of this book that make it different from most other statistical texts:

1. The writing style is informal. We attempt to "talk" directly with the student. When our teaching experience shows us that students typically encounter difficulties, we attempt to act as the students' surrogate. We raise the questions that frequently crop up in class and we attempt to answer them.

2. Many chapters are introduced by a scenario, drawn from real life, that anticipates the main instructional points in the chapter. Thus, in Chapter 1, when we introduce the basic concepts of descriptive and inferential statistics, we focus on a serious contemporary problem, namely, computer fraud. Noting that computer criminals are often brilliant practitioners of their trade, we observe that the usual techniques for uncovering white-collar crimes (i.e., exhaustive auditing of the books) are frequently unable to detect the numerical skullduggery that takes place in the complex world of the floppy discs and random access. We introduce a character, Montana Rincon, who attempts to use sampling techniques as a means of uncovering the presence of computer

fraud. We return to Montana Rincon's statistical sleuthing throughout the book.

3. We have included many within-chapter and end-of-chapter exercises. We have given many of these exercises a context so that the student can appreciate the fact that statistical analyses do not take place in the abstract. Rather they are methods of dealing with real problems in the real world. Consequently, many of the data sets are obtained from the literature of business, economics, finance, and management.

4. Each key term is defined in the margin of the page in which the given concept is introduced. This technique assures that the definitions are given a degree of prominence in accordance with their importance. Moreover, reading the marginal definitions provides a useful chapter summary.

5. We regard the various tables within the text as an integral part of the instructional program. Thus, the captions are usually quite detailed, reinforcing concepts and principles introduced in the textual materials themselves.

6. The statistical tables appearing in the Appendixes were also given much thought and care. Perhaps nothing is more frustrating to a student than to analyze the problem correctly but to falter when using the decision table. This is often not the student's fault since many tables fail to indicate whether or not the values shown are one- or two-tailed, whether the critical region is equal or greater to the tabled values, or equal to or less than this value, etc. To obviate this difficulty, most tables are preceded by a brief explanation of their use, along with an illustrative example.

All in all, we have attempted an intuitive and conceptual approach to statistical analysis. A student with a satisfactory background in high school algebra should have no difficulty in mastering the statistical procedures presented in this text. An accompanying Study Guide provides additional guided practice in each of the statistical procedures appearing in the text. The study guide also augments the text by introducing additional explanatory materials and topics or, in some cases, by presenting statistical procedures that are useful in specialized applications. In addition, the answers to the exercises provided in the text are given in greater detail in the study guide. Thus the student is able to see more than the bottom line, since various intermediate steps are included.

In an undertaking of this sort, there are many others besides the authors who have actively participated. We wish to express the deepest appreciation to the following colleagues who provided invaluable comments and criticisms during the evolution of the text: (Needless to say, we take full responsibility for the final form of the manuscript.)

Paul D. Berger, Boston University; Richard Green, University of California; Ronald Koot, Pennsylvania State University; R. Burt Madden, Shippensburg State College; Donald S. Miller, Emporia State University; Vasanth Solomon, Drake University; Paul Van Ness, Rochester Institute of Technology; Charles Warnock, Colorado State University; Richard J. Westfall, Cabrillo Community College; Howard J. Williams, Wilkes College; and Jack Yurkiewicz, Hofstra University.

A special note of thanks to Teri French who labored cheerfully over our chicken scratches and managed to arrange them into a semblance of coherent language. We love you, Teri.

Thank you also, Laurie Jassenoff, for your invaluable help with some of the more difficult calculations.

Then there are our spouses—Lois Runyon and Jerome Jassenoff—and our children. They endured (or perhaps welcomed?) many long separations, both physical and psychological, as we concentrated on bringing this project to a conclusion. We are profoundly grateful for their patience and understanding during this time.

Richard P. Runyon
Audrey Haber

Contents

What Is Statistics?

1

1.1 STATISTICS AND UNCERTAINTY

Montana Rincon usually awakened in the morning with a surge of good spirits. He liked his work as a CPA, found it challenging, and enjoyed interacting with his many clients. On this particular morning, however, he recoiled at the strident summons of the alarm. He did not wish to relinquish sleep. It had been that bad a night. Unlike his family and friends, his idea of an evening of relaxation involved reading lengthy treatises on the latest developments in his field. But something went wrong last evening. Somehow, he had selected a horror story. Entitled innocently enough, "Computer Crime," the paper delved into the many ways that brilliant minds had corrupted that marvelous tool, the computer, and had put it to work for their own criminal purposes. While he found the names used to describe the crimes both fanciful and humorous, the crimes themselves made his flesh crawl. Data Diddling, the Trojan Horse, the Salami Technique, Logic Bombs, Superzapping, Scavenging—these were but a few of the many types of computer crimes. What they all had in common was a devious scheme to deprive the victims of something that was rightfully theirs.

The diabolic part was that the victims were totally unaware of the crimes against them and the crimes were all executed in but a fraction of a second. A person or a corporation could be financially wiped out in the blink of an eye. What made the whole thing absolutely terrifying was the fact that computer crimes were rarely discovered as a result of internal controls, dogged investigation, or audits of the books. Their discovery was usually an accident, plain and simple. He shuddered to think of the computer crimes that were being executed at that very moment in which the chances of detection were remote. Could it be happening to any of his clients?

He was now involved in auditing the books of a large credit card company. What if someone had used the Salami Technique to take thin slices off selected accounts? It would be simple to do, but almost impossible to catch. All told, the company handled about 5 million accounts. What if someone had built a subroutine into the computer program that randomly selected 1 percent of these accounts and added a small amount (averaging, perhaps, 10 cents) to each bill, which was then deposited into a favored account (his own)? That's 50,000 people swindled out of 10 cents each, on the average. The books would balance and the culprit would be $5,000 richer each month. Would he get caught? How many people would notice an overpayment of 10 or 15 cents? If they did, how many would complain? The postage alone would cost as much as or more than the overcharge. It simply wouldn't be worth the time and effort.

The thought was frightening. He could easily certify the books as correct while a thief drained company profits. Admittedly, not much of a drain in this case. But there is a potential for ripoffs that would make the Brink's robbery look like a fraternity prank. Again, a cold chill passed through his body. How could he avoid being an unwilling accessory in a computer crime? There was no way he could examine the computer program and all of its many ramifications. He simply wasn't sufficiently knowledgeable. There was also no way he could examine all 5 million accounts.

What about a statistical approach? The more he thought about the idea, the better he liked it. He could select, say, 1,000 accounts for a detailed analysis. He then would locate those in which there were errors. He could quickly rule out the most common errors, as for example, a key punch operator entering the wrong number. For those remaining, he would raise a number of penetrating questions. Of the thousand selected, what percentage of these were in favor of the company; what percentage in favor of the customer? What was the average amount of error?

Could he consider the thousand accounts he selected as representative of the entire 5 million accounts? If so, could he estimate the percentage of explained errors in *all* 5 million accounts? Could he estimate the total dollar volume involved in these errors? Finally, could he venture an informed guess as to whether these errors represented haphazard and honest mistakes, or did they represent a systematic effort to commit fraud?

Let's pause for a moment and reflect on what Montana Rincon has been doing. The scenario started with uncertainty. Reading the report on computer crime had made him aware of his own professional vulnerability. Following a restless night, he decided to pursue a course that would resolve some of the uncertainty. After considering various alternatives, he decided to *collect, organize, analyze, interpret, and present information.* This is, in fact, a good working definition of what *statistics* is all about. Note that he would not have raised these questions if he had all the information he required. Statistics is called into play when there are doubts and when we are exploring the unknown.

Once he decided to collect *data* as a means of answering his questions, he realized that there were some serious logistical problems. He was interested in drawing conclusions that would apply to all 5 million accounts. These accounts represented the *population* of interest. Like so many populations that interest statisticians, it was inconceivable that Montana could ever study every member of the population. There simply weren't sufficient time, personnel, and resources to do so. This is, in fact, a common problem. Except in those rare instances when a *census* is undertaken, populations are rarely knowable. Some are infinite and others are so large that they might as well be infinite. When we wish to predict the outcome of a national election, the population is all registered voters who will cast ballots in the election. It is inconceivable that we could poll all of them prior to an election. Likewise, if we are considering marketing a new product with a mass appeal, we would not attempt to survey all possible users.

What would we do? In a word, we would collect a *sample.* However, a sample is useless unless it mirrors the population to which we wish to generalize. In a sense, we want our sample to be a miniature model of the population. One way we may approach this model is to select a *random sample.* This is what Montana Rincon did. He activated a computer program that selected 1,000 accounts in such a way that any of the 5 million accounts in the population could have been selected in the sample. Incidentally, judging the population from a sample is not the exclusive domain of statisticians. We all do it at one time or another; in fact, more frequently than we may realize.

statistics

The collection, organization, analysis, interpretation, and presentation of numerical information.

data

Numbers obtained by measuring, observing, or counting real objects or events in the physical world.

population

The complete set of measurements or observations that interest the person collecting a sample.

census

The process by which we collect data on every element of the population.

sample

A subset of a population.

random sample

A subset of a population selected in such a way that each member has an equal chance of being chosen; each sample of a given size has an equal chance of being selected.

Many of us form opinions about a minority group by talking to only a few members. We judge an actor after watching a couple of his movies. We may estimate weather conditions in a particular locale by visiting there once or twice. Similarly, a few sips are usually sufficient to tell you whether the taste of a wine is satisfactory. It is not necessary to chug-a-lug the whole decanter. We return to the subject of random samples in Chapter 6.

variable

A characteristic of an individual or object that is measurable and takes on different values.

Having decided on the method of selecting his sample, Montana next had to decide what *variables* to study. Since he was interested in assessing error, he collected the following information on each account: the debits (including interest) as of the billing date (variable 1) and the actual amount billed to each customer (variable 2). It would then be a simple matter to subtract the value of variable 2 from the value of variable 1. A difference of zero would mean no error; a positive difference would be an error in favor of the customer; and a negative difference, an error favoring the credit card company.

statistic

A summary numerical value calculated from a sample.

After Montana had collected the samples, he organized the numerical information and conducted some calculations that yielded summary statements, such as the proportion of accounts in the sample that contained errors and the average size of the error. Each of these summary numerical values represents a *statistic.* A statistic is calculated from a sample and yields summary descriptive statements about some aspect of that sample. Some commonly used statistics describe the central values in a set of measurements (e.g., the arithmetic mean and median) or the spread of these measurements (e.g., the range and the standard deviation). Descriptive statistics are traditionally symbolized by italic letters (e.g., $\overline{X}$ for mean). We explain and demonstrate these various statistics in Chapter 3.

parameter

A summary numerical value calculated from a population.

For each descriptive statistic there is a corresponding *parameter.* Thus, there is a population mean, median, range, and standard deviation. Traditionally, parameters are symbolized by Greek letters (e.g., μ = the mean of the population).

Table 1.1 illustrates several populations, variables measured, sample statistics of interest, and corresponding population parameters.

Example A

An electronics firm mass-produces chips for use by the computer industry. Occasionally, the daily output of the plant (the *population*) contains so many defective components that shipping them out to customers invites the risk of losing all future sales. The problems the firm faces are as follows: The daily production of the plant involves hundreds of thousands of chips. Testing each chip is a time-consuming process. To hire and train sufficient personnel to run tests on every chip would be so expensive that the company could not remain competitive. Nevertheless, it is clear that some procedures must be undertaken to estimate the percentage of defective components in the daily output (the *parameter*). If this estimate

Illustration of several populations, variables measured, sample statistics of interest, and corresponding population parameters. **Table 1.1**

Purpose	Population of interest	Variable measured	Sample selected from	Sample statistics	Parameter
To predict outcome of a national election	Registered voters	Choice of candidate	Voter registration lists	Proportion favoring each candidate	Actual proportion in the population favoring each candidate
To estimate the resistance to stress of a metallic casting	All castings made of this alloy	The breaking point (time to break under continued stress or amount of total stress applied before breaking)	Actual castings of this alloy	Mean time or mean amount of stress	The actual mean time or stress in the population
To estimate retail cost of unleaded gasoline in a state	Cost of unleaded gasoline at retail outlets in the state	Price per gallon	List of retail gasoline outlets in state	The mean or average cost per gallon	The actual mean cost per gallon in the state
To decide on the packaging of a new product	Potential purchasers of the product	Ratings of each of several purchase options	List of potential users	Average rating of each packaging option	Actual consumer ratings of each option throughout the population
To ascertain possible employee reaction to a wage freeze	Employees in a given manufacturing plant	Favor or oppose the anticipated freeze	List of all employees in the plant	Proportion favoring the freeze	Actual proportion favoring the freeze among all the employees

exceeds a certain critical value (e.g., 2 percent defective), the entire day's output must be scrapped. How would you go about estimating the percentage of defective components?

Since there is no economical way to study the characteristics of each and every chip, we randomly select a small portion of the total output (the *sample*). We then subject each chip in this sample to a standard performance test. For our purposes we'll assume that performance on this test reveals the presence or absence of a defect (the *variable*). The *data* consist of the number of defective and nondefective chips in the sample. The data are then analyzed according to certain rules to yield *descriptive statistics*. In this example, the appropriate statistic is the percentage of defectives. We then use *inferential statistics* to make estimates of the parameter, i.e., the true percentage of defectives in the population.

Example B

A flour mill receives numerous orders for large sacks of bleached flour from its customers. Because of the high costs of labor each sack is not individually weighed; rather, they are filled by automated equipment set to deliver approximately 250 pounds. For billing purposes, it is decided to select 50 sacks at random from the warehouse, weigh them, and then use the mean as the best estimate of the average weight per sack.

In this example, the *variable* of interest is the weights of flour sacks. The *sample* consists of the 50 randomly selected sacks. The *data* are the weights of the sacks. The *population* is the weights of all sacks actually packaged by the mill. The *parameter* is the actual mean weight of the population.

1.2 THE TWO FACES OF STATISTICS: DESCRIPTIVE AND INFERENTIAL

descriptive statistics

Procedures used to summarize information about samples in a convenient and understandable form.

When Montana manipulated the data according to certain rules to yield summary statements, he was engaging in an important function of statistical analysis: the *descriptive* function. Some of the activities involve computation, as when he calculates a statistic, such as an arithmetic mean or proportion. Others may involve the use of graphic techniques, which are visual aids to thinking about data.

It is only rare that the investigator's interest focuses on the descriptive function as such. To illustrate, Montana was interested in drawing conclusions about the population rather than simply describing the results obtained from his sample. In other words, he was interested in the inferences he could make about the proportion of errors in the population based on the proportion of errors he found in the sample. He was also interested in formulating

Figure 1.1 A flow diagram of various stages in a statistical investigation