A STATISTICS PRIMER FOR MANAGERS

How to Read a Statistical Report
Or a Computer Printout and Get the Right Answers

John J. Clark
Margaret T. Clark

A STATISTICS PRIMER FOR MANAGERS

How to Read a Statistical Report
Or a Computer Printout and Get the Right Answers

John J. Clark
Margaret T. Clark

THE FREE PRESS

The Free Press
A Division of Macmillan Publishing Co., Inc.
866 Third Avenue, New York, N.Y. 10022

Collier Macmillan Canada, Inc.

Printed in the United States of America

printing number

1 2 3 4 5 6 7 8 9 10

Library of Congress Cataloging in Publication Data

Clark, John J.
A statistics primer for managers.

Bibliography p
Includes index.
1 Statistical decision. 2 Management—Statistical methods. I. Clark, Margaret T II Title
HD30 23 C575 1982 519 5'024658 82-70999
ISBN 0-02-905800-7 AACR2

Preface

Managerial statistics treats the contribution of statistics to decision making in typical business situations. Business decisions generate new data and managers in turn rely upon the analysis of data to provide a basis for taking decisions. Yet the analysis of data, however sophisticated, does not ineluctably point to preferred decisions. Decisions ultimately result from the experience and acumen of managers. This book attempts to help in the formation of *informed* management judgments.

The position of managers vis-a-vis statistics is analagous to that of the accounting system. Managers need not be accountants but few managers would realize their potential lacking the ability to interpret intelligently the firm's financial statements or to raise pertinent questions concerning accounting data. Similarly, managers need not be statisticians. Yet the successful manager should possess an understanding of statistical methods sufficient to question the appropriateness of the data presented in support of decisions under consideration. Herein lies the central theme of this volume. Each section and chapter offers an introduction to one of the topics that comprise managerial statistics and concludes with a series of questions that managers can pose when presented with similar data. The resulting dialogue between manager and statistician should foster enhanced appreciation by the executive of the data's strong points and limitations. As a corollary, such give and take helps to define the areas of uncertainty where managerial judgment plays a crucial role.

Statistics can lead or mislead. The latter is generally a result of inaccurate data, erroneous interpretation of valid data, or the application of statistical techniques without regard to underlying assumptions. Statistical methods assume particular characteristics about the data under study. For example, if a firm plans to forecast sales, the input data should exhibit a behavior consistent with the assumptions underlying the statistical method applied or the forecast will fall wide of the mark. Statistical techniques rest upon assumptions that should agree with the reality to be described. To the extent that assumptions

deviate from reality, justifiable suspicion attaches to the conclusions of the investigation. Therefore, in each chapter, we have delineated the underlying assumptions of the method discussed in order to alert the manager of the necessity to examine the assumptions behind the statistical analysis supporting a decision.

To accomplish these purposes, the illustrating problems, or "mini-cases," in each chapter typify business applications of statistical techniques. Thus the assumptions underlying statistical methods are seen in the light of business data and the questions that should be raised by the manager are made concrete in business situations. Hopefully, some situations will be immediately relevant to the reader, while others will only require modification to fulfill present or future needs.

The raw material of statistics is numerical data; this conjures up an image of countless formulae to be mastered by rote without comprehension. We have attempted to minimize the use of formulae. In all instances, whether a formula is stated or not, we have solved the problem in familiar column fashion plus written instructions. The approach rests upon the notion that if the reader can understand what is being done and why—if he can verbalize the method illustrated—then he can construct his own formulae. After all, a formula represents a kind of shorthand comprising a set of instructions to accomplish some end. The end is always: what is to be done; why it is to be done; how it is to be done; and the significance of the results.

The emergence of managerial statistics has coincided with the growing complexity of business operations and the size of the marketplace. The demand for information and the advent of the computer began the "information revolution." Formerly the province of big business, the appearance of microcomputers and time-sharing arrangements have made these capabilities available to all business, large or small, at cost-effective rates. Consequently, it is hardly possible to discuss managerial statistics without reference to the computer. The statistical methods described in our chapters can be solved by computer. Moreover, in business situations, computer capabilities add new dimensions to statistical analysis. It behooves every manager, therefore, to ascertain how the computer might assist him in tackling those decisions that confront him in the performance of his duties. By the same token, the installation of computer facilities imposes on the manager the responsibility to acquire some of the basics about computer performance. Essentially, the manager should be able to interpret computer outputs, intelligently question computer solutions, and comprehend the limitations that attend use of computer data in decision making. The last chapter, dealing with managerial statistics and the computer, has these objectives.

The *Primer* is not a text in the conventional sense. It is plainly an introduction to those topics which are more immediately useful in business situations. Other topics could have been added and the discussion of the topics chosen could also have been expanded. These are matters of selectivity with which some will disagree. However, Appendix A contains a bibliography of academic and business publications. The sources cited introduce other topics in managerial statistics and/or offer advanced treatments of the topics included in the *Primer*.

The *Primer* contains many illustrating problems. To facilitate retention and understanding of the concept illustrated by the problem, the reader will do well to follow the solution using a pocket calculator. Appendix B, containing a glossary of key terms, should also prove useful in the learning process.

Contents

A STATISTICS PRIMER FOR MANAGERS

How to Read a Statistical Report
Or a Computer Printout and Get the Right Answers

John J. Clark
Margaret T. Clark

THE FREE PRESS
A Division of Macmillan Publishing Co., Inc.
NEW YORK
Collier Macmillan Publishers
LONDON

chapter 1

What Is Statistics . . . and Managerial Statistics?

Statistics is essentially concerned with mass phenomena, with aggregates, not with individual observations. The characteristics of some aggregate are sought, such as average family income in the United States, price movements on the New York Stock Exchange, business expenditures on new capital equipment, or changes in the size of the nation's money supply. The investigator looks for a *statistic*—a total, an average, a rank, or a percentage—to typify a mass of data. This perspective also conveys the idea that a statistic is a fact stated numerically.

Statistics focuses on aggregates because the stability of mass data makes it feasible to describe and/or predict. Although chance factors can shape the behavior of individuals, thus eroding the ability to predict, valid statements can still be formulated about aggregate behavior. We may not be able to predict next year's capital expenditures for a given company but we do, with a degree of success, annually forecast capital expenditures for *all* business firms. Statistical theory postulates the *stability of mass data*.

However, although definable, and to some degree predictable, the uniformities appearing from the analysis of mass data are not always perfect. An element of uncertainty is involved in the statistician's assertions about the aggregate under study. Thus statistics can also be viewed as a method of decision making in the face of uncertainty, on the basis of numerical data, and at calculated risks. Now, the statistician couches his generalizations in terms of probabilities and may employ methodologies congenial to managerial situations. When these

methodologies are applied in a business context, we have identified the field of managerial statistics.

The uncertainty coloring statistical generalization emanates from the concept of randomness. A random variable is one that assumes different values because of chance. Consequently, random samples taken from the same population (for example, samples on business capital expenditures) may each show different results. Nonetheless, such chance fluctuations in sampling outcomes, while they create uncertainty, do not prevent the possibility of statistical generalization. Probability theory offers a method of dealing with these uncertainties and allows us to make inferences about the population from which the sample was taken.

Another source of uncertainty can arise from a deficiency of knowledge concerning the "state of nature," or how the laws of randomness apply in a given situation. In the case of capital expenditures, there already exists a body of statistical literature on the past behavior of the variable and the interpretation of changes in the numerical data. This provides a frame of reference when judging future sample results. For other variables, there may not exist a similar set of signposts pointing, tentatively at least, to the parameters of the problem. The investigator may have to formulate and test his own hypotheses aided by the sample information to arrive at a statistical generalization.

The difference is this: if you toss a fair coin a given number of times, you know that in the long run the number of heads and tails will *tend* to be equal; any given toss could go either way (hence the presence of risk) but the state of nature is known to be a 50–50 distribution. On the other hand, if the coin is biased, more heads or tails will appear; if you are not sure whether the coin is biased, you can run a sample number of tosses and test the results against the outcome that would have resulted were the coin a fair one. What is the *probability* of a similar result using a fair coin? You now have a basis for stating whether the coin is fair or biased—but you will never be sure. Herein are the elements of managerial statistics: decision making under conditions of uncertainty; use of numerical datas; the presence of risk expressed in probabilities.

It is convenient to define a few key terms used in the preceding paragraphs. *Population* refers to the totality of data that can be collected on the problem being studied, the group or aggregate about which the investigator will state relevant generalizations. For determining average family income in the United States, the population consists of all family units (as defined by the investigator). Thus average family income might be measured simply by looking at the income of each family unit—*a complete enumeration.*

This is likely to involve substantial expense, however. Instead of a complete enumeration of all family units, the investigator might survey

only a portion of the population, that is, draw a *sample*. The size of the sample, the type of data collected, and the method of collection are appropriately left for later discussion. For the moment, we shall define a *random sample* as one in which every item in the population (that is, family unit) has an equal chance of being selected for the survey. *Only random samples permit the use of probabilities and create the basis for inferring statistical generalizations about the population surveyed.*

On the other hand, a complete enumeration of the population, where possible, eliminates or vastly reduces the risk of incorrectly describing the characteristics of the population. With a complete enumeration it is no longer necessary to make assumptions concerning the population in order to formulate statistical generalizations based on sample information.

原书缺页

原书缺页

SECTION II

The Several Varieties of Averages

chapter 2

Averages: Ungrouped Data

The world of business is beset by averages of all kinds. The daily proliferation of economic, business, and financial data makes it necessary for management to grasp quickly and concisely, the significance of volumes of statistical information. As a result there is great reliance by management staffs on the use of averages that summarize and typify large arrays of individualized information. Administrative services may use averages for comparative wage and salary studies; marketing may use averages in divisional sales analysis; in the financial area analysis is replete with averages of earnings, stock prices, income, operating statistics, and so forth. Undoubtedly, averages form the basis for much "communication" between management, staff, and stockholders concerning the company's profile.

However, although averages are widely used in business and are relatively uncomplicated, it behooves the manager to use caution in interpreting an average as typical of a data group. There are many types of averages, each with a distinct meaning and set of limitations.

TYPES OF AVERAGES

A valid average is representative of the individual values from which it is calculated. It aggregates numerical values into a single capsule. Hence, *an average measures the central tendency of the data, the point where the values cluster to typify the series.* But how typical is the average and how much reliance can be placed in this one number? The answer depends upon the type of average and the method of calculation.

Averages can be broadly classified as computational or positional:

1. *Computational averages.* These include the arithmetic mean, the geometric mean, and the harmonic mean. They are determined by computation of all the values in the data series and are subject to further manipulation.