

STATISTICS FOR BUSINESS AND ECONOMICS

Paul Newbold

University of Illinois, Urbana-Champaign

Library of Congress Cataloging in Publication Data Newbold, Paul.

Statistics for business and economics.

Includes index.

1. Commercial statistics. 2. Economics—Statistical methods. 3. Statistics. I. Title. HF1017.N48 1983 519.5 83-22992

ISBN 0-13-845140-0

Editorial/production supervision by Margaret Rizzi Interior design by Suzanne Behnke Cover photograph by Paul Silverman Cover design by Suzanne Behnke Manufacturing buyer: Ed O'Dougherty

© 1984 by Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-845140-0

Prentice-Hall International, Inc., London
Prentice-Hall of Australia Pty. Limited, Sydney
Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro
Prentice-Hall Canada Inc., Toronto
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Prentice-Hall of Southeast Asia Pte. Ltd., Singapore
Whitehall Books Limited, Wellington, New Zealand

preface

Statistics courses are offered in virtually all college business programs, feared by students in most, and rarely enjoyed. At least at the outset, students tend to suspect that such courses will be both difficult and boring. I would like to reassure the minority of students who actually read prefaces by asserting that the subject matter of business statistics is both easily grasped and invariably entertaining. However, honesty prohibits my doing so.

Experience suggests that students may have problems in immediately grasping some of the material in this book. Many new procedures for the analysis of business data will be met and, while the temptation is to try to memorize the ingredients of each ("cookbook" style), this can eventually lead to confusion. To avoid such problems, I have tried to set out carefully the rationale for each procedure, keeping in mind that the vast majority of readers of this book will be consumers rather than producers of statistical analyses. To facilitate understanding of the material, I have included a large number of numerical examples. Nevertheless, in the final analysis, statistics is a subject in which "learning by doing" is crucially important. This being the case, the text also contains a great many exercises. It is hoped that, by working through several exercises on each topic, the student's understanding of the subject matter will be sharpened.

In writing this text, I have been conscious of the importance of persuading students to invest the effort needed to absorb the methods of statistical analysis. My strategy has been to incorporate many realistic examples and exercises taken from the business literature, including, for example, the fields of accounting, economics, finance, marketing, and industrial organization. It is hoped that the accumulation of this illustrative material will convince the reader of the relevance of statistics to the modern business environment, and therefore heighten his or her interest in the subject.

This text is suitable for a one- or two-semester introductory course aimed at business or economics majors. I have included more than enough material for a two-semester course, and many instructors will not wish to cover every chapter in detail, particularly if a large amount of time is to be devoted to project work. All, or parts, of Chapters 10, 11, and 14–19 can be omitted without loss of continuity. Also, Chapter 19 can be covered at any stage after Chapter 4, and Chapter 18 at any time after Chapter 8. One of the many possibilities for a one-semester course is to cover Chapters 1–9, 12, and perhaps 13. In such a course, instructors may wish to omit Sections 4.4, 4.6, 4.7, 7.3, 9.9, and 12.2.

This book developed from courses I have taught to students of business and economics in a number of universities, most recently the University of Illinois at Urbana-Champaign. I am very grateful to my colleagues, to graduate teaching assistants, and, perhaps most of all, to students for helpful discussions, suggestions, and criticisms. All, or parts, of earlier drafts were read by a number of reviewers. including Professor Craig F. Ansley, University of Chicago; Professor Paul D. Berger, Boston University; Professor Frank P. Buffa, Texas A&M University; Professor Virginia Knight, Western New England College; Professor Zenon S. Malinowski, University of Connecticut; Professor Glenn W. Milligan, Ohio State University; Professor Charles F. Mott, Bowling Green State University; Professor Chester Piascik, Bryant College; Professor Leonard Presby, William Paterson College; Professor Susan L. Reiland, North Carolina State University; Professor Terry G. Seaks, University of North Carolina; Professor Ronald E. Shiffler, University of Louisville; and Professor Craig W. Slinkman, University of Texas. Their comments were invaluable and greatly enhanced the presentation in many places. I must also acknowledge the help and constructive prodding of two Prentice-Hall editors, Doug Thompson and Margaret Rizzi.

The final manuscript and its earlier drafts were expertly typed by Dixie Trinkle. I must thank Alice Newbold for preparing the graphs, and also for her patience and understanding throughout this long project.

Finally, to any student who has read this far through the preface, I wish you success in your studies of business statistics. In particular, I hope that I will be able to convince you of the importance and relevance of statistics in many branches of business.

Paul Newbold

xiv

applications index

Throughout this text, statistical methods are illustrated through real and realistic business examples. Many examples and exercises are drawn from the published literature in different areas of business. These applications are indexed below.

ACCOUNTANCY

Confidence interval for difference between population means, 310

Confidence interval for population mean, 287

Confidence interval for population proportion, 327

Contingency tables, 431, 433, 437

Grouped data and histograms, 35

Hypothesis test for equality of population variances, 387

Hypothesis test for population mean, 352

Mann-Whitney test, 412

One-way analysis of variance, 618

Rank correlation, 454

Wilcoxon test, 403

ADVERTISING, MARKETING AND RETAILING

Confidence interval for difference between population means, 311

Confidence interval for difference between population proportions, 317, 320, 321

proportion, 299, 300, 305, 327 Contingency tables, 426, 432, 434, 435, 438 Correlation, 444 Goodness of fit test, 415 Hypothesis test for difference between population means, 361, 368, 373, 374, 391, 392 Hypothesis test for difference between population proportions, 371, 374, 375, Hypothesis test for population proportion, 361 Jointly distributed discrete random variables, 151 Kruskal-Wallis test, 623, 627 Lagged dependent variable regression, 556 Linear regression, 456 Log linear regression, 579 Multiple regression, 517 One-way analysis of variance, 651 Probability, 91, 107, 132 Rank correlation, 453, 454, 488

Confidence interval for population

Sampling distribution of sample proportion, 242, 248

Time series analysis, 692, 693

Two-way analysis of variance, 653, 655

BANKING AND FINANCE

Confidence interval for difference between population means, 314

Confidence interval for difference between population proportions, 330

Contingency tables, 439

Correlation, 487

Hypothesis test for difference between population means, 366

Hypothesis test for population mean, 347

Hypothesis test for population proportion, 359

Lagged dependent variable regression, 607

Linear regression, 489

Mann-Whitney test, 408

Multiple regression, 511, 519, 536, 537

Probability, 127

ECONOMICS AND INDUSTRIAL ORGANIZATION

Confidence interval for difference between population means, 320
Confidence interval for population mean, 294

Dummy variable regression, 552, 560, 561, 564, 606

Hypothesis test for equality of population variances, 378

Lagged dependent variable regression, 565, 566

Linear regression, 485, 490

Log linear regression, 570, 578, 608

Multiple regression, 494, 529, 530

MANAGEMENT AND ORGANIZATIONAL BEHAVIOR

Confidence interval for difference between population means, 319

Confidence interval for population proportion, 326, 327

Correlation, 449, 487

Grouped data and histograms, 78

Hypothesis test for difference between population means, 391

Hypothesis test for difference between population proportions, 393

Multiple regression, 541

One-way analysis of variance, 626, 627

Poisson distribution, 176

Probability, 127

Two-way analysis of variance, 646

contents

PREFACE xiii

APPLICATIONS INDEX XV

1	WHAT	IS	STATISTICS?	1
---	------	----	-------------	---

1.1 Mak	king Sense	of Numerical	Information	- 1
---------	------------	--------------	-------------	-----

1.2 Dealing with Uncertainty 2

3

- 1.3 Sampling
- 1.4 Analyzing Relationships 4
- 1.5 Forecasting 5
- 1.6 Decision-Making in an Uncertain Environment 5

SUMMARIZING NUMERICAL INFORMATION 6

- 2.1 Populations and Samples 6
- 2.2 Numerical Summary: Measures of Central Tendency 8
- 2.3 Numerical Summary: Measures of Dispersion 14
- 2.4 Grouped Data and Histograms 30

3	PROBABILITY 82 3.1 Introduction 82 3.2 Random Experiment, Outcomes, Events 83 3.3 What Is Probability? 91 3.4 Probability and Its Postulates 94 3.5 Permutations and Combinations 99 3.6 Probability Rules 104 3.7 Bivariate Probabilities 115 3.8 Bayes' Theorem 120
4	DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS 134 4.1 Random Variables 134 4.2 Probability Distributions for Discrete Random Variables 136 4.3 Expectations for Discrete Random Variables 141 4.4 Jointly Distributed Discrete Random Variables 151 4.5 The Binomial Distribution 164 4.6 The Hypergeometric Distribution 171 4.7 The Poisson Distribution 173
5	CONTINUOUS RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS 185 5.1 Continuous Random Variables 185 5.2 Probability Distributions for Continuous Random Variables 186 5.3 Expectations for Continuous Random Variables 193 5.4 Jointly Distributed Continuous Random Variables 197 5.5 The Normal Distribution 201 5.6 The Central Limit Theorem 213

5.7 The Normal Distribution as an Approximation to the Binomial and Poisson

36

2.5 Numerical Summary of Grouped Data

66

2.6 Some Other Graphical Methods

2.7 Lying with Statistics

Distributions

218

6	SAMPLING AND SAMPLING DISTRIBUTIONS 228					
	6.1 6.2 6.3 6.4	Sampling Distribution of a Sample Proportion 239 Sampling Distribution of the Sample Variance 249				
7		INT ESTIMATION 261 Introduction 261 Unbiased Estimators and Their Efficiency 264				
0	7.3	Consistent Estimators 270 TERVAL ESTIMATION 276				
ŏ	8.1 8.2 8.3 8.4 8.5 8.6 8.7 8.8	Confidence Intervals 276 Confidence Intervals for the Mean of a Normal Distribution: Population Variance Known 278 The Student's t Distribution 288 Confidence Intervals for the Mean of a Normal Population: Population Variance Unknown 291 Confidence Intervals for the Population Proportion (Large Samples) 297 Confidence Intervals for the Variance of a Normal Population 301 Confidence Intervals for the Difference Between the Means of Two Normal Populations 306 Confidence Intervals for the Difference Between Two Population Proportions (Large Samples) 315 Estimating the Sample Size 321				
9	9.1 9.2 9.3	POTHESIS TESTING 331 Concepts of Hypothesis Testing 331 Tests of the Mean of a Normal Distribution: Population Variance Known 338 Tests of the Mean of a Normal Distribution: Population Variance Unknown 348				
	9.4	Tests of the Mean of a Normal Distribution: Population Variance Unknown 348 Tests of the Variance of a Normal Distribution 353				

9.5 Tests of the Population Proportion (Large Samples)

9.7 Test for the Difference Between Two Population Proportions (Large

9.6 Tests for the Difference Between Two Means

369

Samples)

357

	9.9 Measuring the Power of a Test 3799.10 Some Comments on Hypothesis Testing 386
10	SOME NONPARAMETRIC TESTS 394
10	10.1 Introduction 394 10.2 The Sign Test 395 10.3 The Wilcoxon Test 400 10.4 The Mann-Whitney Test 404 10.5 Discussion 409
11	GOODNESS OF FIT TESTS AND CONTINGENCY TABLES 415
	 11.1 Goodness of Fit Tests 415 11.2 Goodness of Fit Tests: Population Parameters Unknown 421 11.3 Contingency Tables 426
19	LINEAR CORRELATION AND REGRESSION 440 12.1 Correlation 440
12	12.2 Rank Correlation 44912.3 The Linear Regression Model 45612.4 Least Squares Estimation 460
	12.5 Standard Assumptions for the Linear Regression Model 464 12.6 The Gauss-Markov Theorem 465
	12.7 The Explanatory Power of a Linear Regression Equation 467 12.8 Confidence Intervals and Hypothesis Tests 474 12.9 Prediction 479 Appendix A12.1 492
13	MULTIPLE REGRESSION 494
10	 13.1 The Multiple Regression Model 494 13.2 Least Squares Estimation 502 13.3 Standard Assumptions for the Multiple Regression Model 504
viii	Contents

9.8 Testing the Equality of the Variances of Two Normal Populations

376

	 13.5 The Explanatory Power of a Multiple Regression Equation 507 13.6 Confidence Intervals and Hypothesis Tests for Individual Regression Parameters 513 13.7 Tests on Sets of Regression Parameters 520 13.8 Prediction 526 13.9 Computer Packages for Regression Calculations 533 Appendix A13.1 545
1/	ADDITIONAL TOPICS IN REGRESSION ANALYSIS 549
14	14.1 Model Building Methodology 14.2 Dummy Variables 552 14.3 Lagged Dependent Variables 555 14.4 Nonlinear Models 567 14.5 Specification Bias 571 14.6 Multicollinearity 574 14.7 Heteroscedasticity 582 14.8 Autocorrelated Errors 588 14.9 Summary 601
15	ANALYSIS OF VARIANCE 609 15.1 Comparison of Several Population Means 609 15.2 One-Way Analysis of Variance 612 15.3 The Kruskal-Wallis Test 621 15.4 Two-Way Analysis of Variance: One Observation per Cell, Randomized Blocks 628 15.5 Two-Way Analysis of Variance: More than One Observation per Cell 639
16	INDEX NUMBERS 657 16.1 The Index Number Problem 657 16.2 Price Index for a Single Item 659 16.3 Unweighted Aggregate Price Indices 661 16.4 Weighted Aggregate Price Indices 663 16.5 Weighted Aggregate Quantity Indices 671 16.6 A Value Index 673 16.7 Change in Base Period 681

505

13.4 The Gauss-Markov Theorem

17	TII	ME SERIES ANALYSIS AND FORECASTING	688
1 /	17.1 17.2 17.3 17.4	Components of a Time Series 693 Moving Averages 698	
	17.5 17.6 17.7 17.8 17.9	Simple Exponential Smoothing 713 The Holt-Winters Exponential Smoothing Forecasting Model 717 Autoregressive Models 730	711
10	SU	RVEY SAMPLING METHODS 745	
10	18.1 18.2 18.3 18.4 18.5 18.6	, , , , , , , , , , , , , , , , , , , ,	
10	ST	ATISTICAL DECISION THEORY 795	
19	19.1 19.2 19.3 19.4 19.5 19.6	Decision-Making under Uncertainty 795 Solutions Not Involving Specification of Probabilities 799 Expected Monetary Value 805 Use of Sample Information: Bayesian Analysis 813 The Value of Sample Information 820 Allowing for Risk: Utility Analysis 837	
APPEN	DIX:	: TABLES 849	
	1	Probability Function of the Binomial Distribution 849	
	2 3 4 5 6	Values of $e^{-\lambda}$ 854 Cumulative Distribution Function of the Standard Normal Distribution Some Uniformly Distributed Random Numbers 857 Cutoff Points of the Chi-Square Distribution Function 858 Cutoff Points for the Student's t Distribution 860	855

- 7 Cutoff Points for the F Distribution 862
- 8 Cutoff Points for the Distribution of the Wilcoxon Test Statistic 865
- 9 Cutoff Points for the Distribution of Spearman's Rank Correlation Coefficient 865
- 10 Cutoff Points for the Distribution of the Durbin-Watson Test Statistic 866
- 11 Cumulative Distribution Function of the Runs Test Statistic 869

ANSWERS TO SELECTED EVEN-NUMBERED EXERCISES 870

INDEX 888

Contents

what is statistics?

CHAPTERONE

Since the majority of readers tend to skip the introductory chapter of a text, it is tempting to answer the question of the title with, "Statistics is what statisticians do," and to proceed immediately to the next chapter. Unfortunately, my editor argued forcibly that rather more might be expected. In the end, I decided to compromise with convention and to keep these introductory remarks as brief as possible.

Compulsory statistics courses have received a bad press. Indeed, on the typical college campus, enrollment in such a course ranks somewhere in popularity between laws establishing a minimum drinking age and mandatory draft registration. In part, this reputation is deserved. The concepts involved are not always easy to grasp at first, and it is certainly necessary to work hard to keep up with any worthwhile course in the subject. However, one complaint heard occasionally is definitely unjustified. **Statistics is not irrelevant.** The remainder of this chapter is devoted to an expansion of this point, particularly with respect to business and economic problems.

Statistics *is* what statisticians do. What is remarkable is the range of activities in which statisticians are involved. These activities impinge on virtually every aspect of daily business and economic life. We will group these activities under six broad headings.

1.1 MAKING SENSE OF NUMERICAL INFORMATION

Any manager operating in the business environment requires as much information as possible about the characteristics of that environment. In the modern era, much of the available information is quantitative. For example, it may be necessary to assimilate movements in interest rates, stock market prices, money supply, or unemployment.

Market research surveys are carried out to determine the strength of product demand. An auditor is concerned about the number and size of errors found in accounts receivable. A personnel manager may be able to use aptitude test scores, in addition to subjective judgment of candidates for employment. The list is virtually endless.

The common features of these examples are that the information to be absorbed is *numerical*, and that the sheer amount of that information renders it, in its raw form, virtually impossible to comprehend fully. The statistician's role involves the extraction and synthesis of the important features of a large body of numerical information. One objective is to try to make sense of numerical data by summarizing it in such a way that a readily understood picture emerges while little of importance is lost.

Many issues are involved in a comprehensive analysis and synthesis of numerical data. The most appropriate method will depend on the nature of the numerical information and how it is to be used. In some circumstances it will be desirable to employ some of the heavy artillery of formal techniques to be discussed in the later chapters of this book. On other occasions, a relatively straightforward numerical or graphical summary may be sufficient and should, in any event, provide a good basis for a deeper analysis. In Chapter 2, we will consider some useful techniques for summarizing numerical information.

1.2 DEALING WITH UNCERTAINTY

A second answer to the question posed at the beginning of this chapter is, "Statistics is the science of uncertainty." In statistics, we do not deal with questions of what *is*, but of what *could be*, what *might be*, or what *probably is*. Consider the following statements:

"The price of IBM stock will be higher in six months than it is now."

"If the federal budget deficit is as high as predicted, interest rates will remain high for the rest of the year."

"If a bid of this level is submitted, it will be lower than competitors' bids and the contract will be secured."

"The best opportunities for improvement in market share for this product lie in an advertising campaign aimed at the 18-25 year age group."

Each of these statements contains language suggesting a spurious amount of certainty. At the time the assertions were made, it would have been impossible to be *sure* of their truth. Although an analyst may believe that anticipated developments over the next few months are such that the price of IBM stock is likely to rise over the period, he or she will not be certain of this. Thus, from a purely semantic point of view, the above statements should be modified, as indicated by the following examples:

"The price of IBM stock is likely to be higher in six months than it is now."

"If a bid of this level is submitted, it is probable that it will be lower than competitors' bids and the contract will be secured."

However, our concern about uncertainty is not merely semantical. All we have done so far is to replace unwarrantedly precise statements by unnecessarily vague statements. After all, what is meant by "is likely to be" or "it is probable that"? Perhaps the two modified statements could be interpreted as assertions that the events of interest are more likely than not to occur. But, *how much* more likely? The English language is rich in words that describe uncertainty, and, indeed, some of these suggest a gradation from the impossible to the certain. Nevertheless, language alone is inadequate to provide a satisfactory description of the degree of uncertainty attached to the occurrence of a particular event. Rather, we need a more formal structure for this purpose.

In the majority of this book, we will be discussing procedures for attacking problems where the conclusion will necessarily be couched in the formal language of uncertainty. As a prelude, that language—probability—is introduced in Chapter 3.

1.3 SAMPLING

Before bringing a new product to market, a manufacturer wants to arrive at some assessment of the likely level of demand, and a market research survey may be undertaken. The manufacturer is, in fact, interested in the **population** of all potential buyers. However, it is prohibitively expensive, if not impossible, for a typical market research survey to contact every member of that population. Rather, a small subset—or **sample**—of population members will be contacted, and any conclusions about the population will be based on information obtained from the sample.

The technique of sampling large populations is commonly used in business. For example, decisions about whether a production process is operating correctly are based on the quality of a sample from its output. Again, an audit of accounts receivable will generally be based on a sample of all accounts.

When we have information on a sample from a population, it is generally straightforward to summarize the numerical sample data. However, taking a sample is merely a means to an end. The objective is not to make statements about the sample, but rather to draw conclusions about the wider population. Thus, an important problem for the statistician involves the extent to which it is possible to generalize about a population, based on results obtained from a sample.

Of course, if a sample is taken from a population, we will not be able to learn *precisely* the population characteristics. For example, suppose that a sample of accounts receivable is examined, and it is found that 8.2% of these are in error. It does not follow that *exactly* 8.2% of all the accounts receivable in the population are in error. We will have learned something about this population percentage, but we will not know its exact value. Some uncertainty will remain. Hence, in making inferences about a population based on the results of a sample, any conclusions will naturally involve the language of uncertainty, as discussed in the previous section.

We begin our exploration of procedures for the analysis of sample data in Chapter 6, postponing to Chapter 18 a comprehensive discussion of methods for selecting samples.