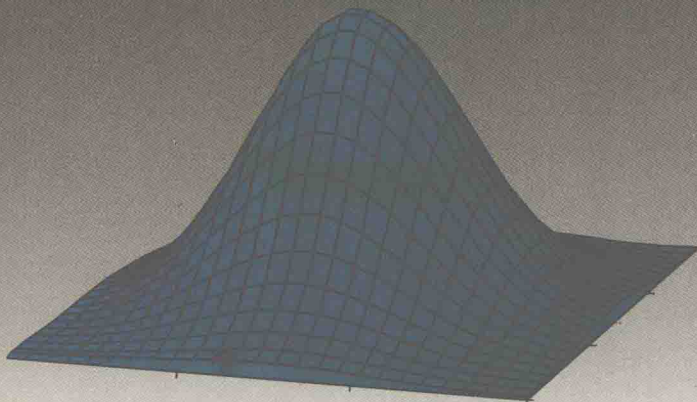
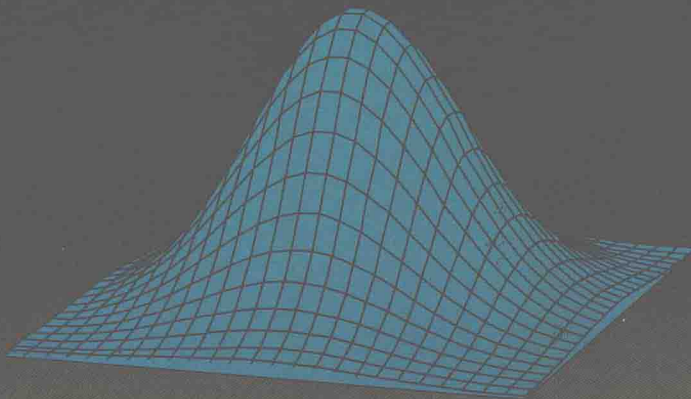
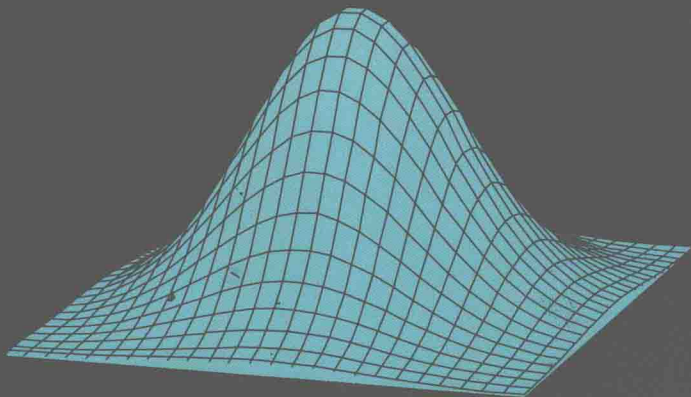


A M I R D. A C Z E L

*Complete
Business
Statistics*



Complete Business Statistics

AMIR D. ACZEL
Bentley College

1989

IRWIN

Homewood, IL 60430
Boston, MA 02116

© RICHARD D. IRWIN, INC., 1989

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Sponsoring editor: Richard T. Hercher, Jr.
Developmental editor: James E. Diestler
Project editor: Margaret S. Haywood
Production manager: Stephen K. Emry
Designer: Paula Lang
Compositor: Interactive Composition Corporation
Typeface: 10/12 Times Roman
Printer: R. R. Donnelley & Sons Company

Library of Congress Cataloging-in-Publication Data

Aczel, Amir D.

Complete business statistics / Amir D. Aczel.

p. cm.

Bibliography: p.

Includes index.

ISBN 0-256-05716-8

1. Commercial statistics. 2. Statistics. I. Title.

HF1017.A26 1989

519.5—dc19

88-13535

CIP

Printed in the United States of America

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

*Complete
Business
Statistics*

The Irwin Series in Quantitative Analysis for Business
Consulting Editor Robert Fetter *Yale University*

*To my father, Captain E. L. Aczel,
and to the memory of my mother, Miriam Aczel*

Preface

In writing this book, I have tried to achieve several important goals. First, I wanted a book that *really* explained statistical thinking and the ideas of statistical inference rather than just giving formulas and examples as do many business statistics books. I also wanted to stress an *intuitive* understanding of statistical procedures and an understanding of what statistics is all about. For example, this book explains the logical principle behind the analysis of variance rather than just the computations. Once the student understands this principle, the complicated sums of squares suddenly make sense and the final result is not just a number—it is a complete picture of what the data are trying to tell us about an actual problem. As another example, this is one of very few statistics books at this level that actually explain the meaning of *degrees of freedom*.

My second goal was to write a book that is firmly rooted in the real world. I carefully chose exceptional problems and examples that I came across in actual business situations; in doing so, I drew on my experience in the field of marketing research. I also chose a wide selection of real problems and examples from the fields of finance, management, transportation, tourism, accounting, public administration, economics, production, the fashion industry, advertising, and other areas. Many of the examples and problems originate in the professional business literature; other problems are company-specific and originate in statistical consulting projects. The book also contains nine cases where the student must apply several statistical concepts, use statistical reasoning, and consider different methods of solution for a complicated business problem.

My third goal was to present a wide range of statistical topics. I sought to write a *complete* book of business statistics, and I included several important topics that are avoided by other authors. This is the only book I know of at this level that includes a chapter on multivariate statistical methods as well as a complete and thorough introduction to the Box-Jenkins forecasting methodology. Since both of these areas of modern statistics are indispensable in many situations in business, economics, and other fields, the lack of an explanation of these topics in other books is surprising and limits their usefulness.

After a brief introduction to descriptive statistics (Chapter 0), the book moves on to a thorough treatment of probabilistic reasoning and the concepts of probability theory—the basis for all of statistical inference. The book then provides the student with a detailed explanation of random variables, expected values, the variance and the standard deviation. There is an entire chapter devoted to the normal distribution—including some optional advanced topics not commonly covered as well as a

wide range of problems demonstrating the usefulness of the normal probability model. Once the student is comfortable with probability and random variables, the ideas of sampling are carefully explained. Estimation and confidence intervals are next, followed by a very thorough discussion of statistical hypothesis testing. Hypothesis tests are explained at several different levels, including the philosophy behind a statistical hypothesis test, when and how tests should be conducted, and how to carefully interpret their results. This is one book that truly explains the meaning of statistical *significance*.

Every chapter in this book introduces the basics of the theory and then expands the discussion to include more advanced material. Chapter 10 on multiple regression, for example, contains many advanced topics. At the beginning undergraduate level, however, an instructor may choose the earlier sections, leaving out more complicated topics such as transformations. He or she is left with a thorough, clear, and easy-to-understand discussion of multiple regression models. The presentation is complete and includes a discussion of the *pitfalls* to avoid when conducting a regression analysis.

Books written specifically for an introductory course often fail to discuss the perceived “difficult” topics in regression and other areas. This is one reason why statisticians so often encounter misuse of statistical methods and the consequent distortion of results and implications. This is especially true nowadays that computers and software packages for statistical analysis are so prevalent that everyone has access to “statistics.” It is now more important than ever before to teach business practitioners and others the *correct* use of statistics. Computing power should go hand in hand with—not replace—statistical reasoning, correct selection of a statistical model, careful evaluation of the underlying assumptions, and cautious interpretation of results. It is my hope that this book stresses these aspects of any analysis more completely and understandably than other statistics texts. It is, in fact, the development and implementation of new statistical software that made me realize just how important it is that we teach the advanced topics. When a microcomputer can be used to perform complicated multivariate analysis, statistical instruction that does not cover such topics is behind the times.

The book integrates theory and examples with use of the computer. The main computer package demonstrated is MINITAB. This package was chosen because it is very easy to use. Each chapter where data analysis is conducted contains a section on the use of the computer. MINITAB is explained in some detail to allow the student to use it with little or no prior preparation. Two other packages are explained and used in the book: SPSSX and SAS. These are used especially in the advanced sections, such as those on multivariate analysis where MINITAB lacks the appropriate capabilities.

Many people were helpful to me in preparing this book. I would like to thank Professor Robert Fetter of Yale University for his many helpful comments and suggestions. I gratefully acknowledge the help of the following reviewers: Professor Michael Sklar, Emory University; Professor John Sennetti, Texas Tech University; Professor Jamie Eng, San Francisco State University; Professor Chaim M. Ehrman, Loyola University of Chicago; Professor Samuel Kotz, University of Maryland; Professor Andrew Seila, University of Georgia; Professor Paul Rubin, Michigan State

University; Professor Amitava Mitra, Auburn University; Professor Randall Anderson, California State University; Professor Charles Feinstein, University of Santa Clara; Professor Peter Seagle, State University of New York—Albany; Professor Michael Broida, Miami University of Ohio; Professor Gary Franko, Siena College; Professor Mary Sue Younger, University of Tennessee; Professor Robert Des Jardines, University of North Carolina at Chapel Hill; Professor Marvin Puterman, University of British Columbia; Professor Ralph St. John, Bowling Green State University; Professor Charles Warnock, Hong Kong University; Professor Thomas Knowles, Illinois Institute of Technology; Professor Donald Robinson, Illinois State University; and Professor Peg Young, George Mason University.

I would like to thank the following colleagues, friends, and students, who also helped with the project: Professor Sherman Chottiner, Syracuse University; Dr. Beverly Beeton, Vice Chancellor of the University of Alaska; Dr. Michael Paradise, AACJC; Professor Gary Anders and Professor Richard Hacker, University of Alaska; Professor Paul D. Berger, Boston University; Nancy Koon Barto, Dennis Burns, Linda Dwyer, Wayne Powers, Lyn Egolf, E. J. Caplan, Rev. Lorin Bradbury, Kathryn Daughettee, Janice McLean, Anne Chi, Dennis Harris, Calvin Crumrine. I would also like to thank many other people at the University of Alaska, where I wrote much of this book.

I would like to thank Bruce Powell of Richard D. Irwin, Inc., who first suggested that I write this book. Furthermore, I am indebted to my editor, Richard T. Hercher Jr., for all his help, encouragement, enthusiasm, and support throughout this project. I would also like to thank James E. Diestler, developmental editor; Margaret Haywood, project editor; Karen Eichorst, editorial assistant; David Mason, copy editor; Paula Lang, designer; and many others of the highly professional staff at Irwin for all their help and hard work.

Among others, I am indebted to the *Biometrika* trustees for permission to reprint various statistical tables from *Biometrika* publications, and to the American Marketing Association for permission to reprint many tables and other information from the *Journal of Marketing* and the *Journal of Marketing Research*. I would also like to acknowledge the New York Times Company for its kind permission to reprint several articles. Finally, I thank my wife, Debra, for all her help in editing the manuscript and for her many suggestions and ideas.

Amir D. Aczel

Contents

- 0 Introduction and Descriptive Statistics** **2**
- Introduction: What Is Statistics? Percentiles and Quartiles. Measures of Central Tendency. Measures of Variability. Grouped Data and the Histogram. Skewness and Kurtosis. Scales of Measurement: *Nominal Scale. Ordinal Scale. Interval Scale. Ratio Scale.* Using the Computer. Summary and Review of Terms.
- 1 Probability** **30**
- Introduction. Basic Definitions: Events, Sample Space, and Probabilities. Basic Rules for Probability: *Mutually Exclusive Events.* De Morgan's Rules. Conditional Probability. Independence of Events. Combinatorial Concepts. The Law of Total Probability and Bayes' Theorem: *The Law of Total Probability. Bayes' Theorem.* Summary and Review of Terms.
- 2 Random Variables** **74**
- Introduction. Discrete Random Variables: *Cumulative Distribution Function.* Expected Values of Discrete Random Variables: *The Expected Value of a Function of a Random Variable. The Variance and the Standard Deviation of a Random Variable. Chebyshev's Theorem. Variance of a Linear Function of a Random Variable.* The Binomial Distribution: *Binomial Probabilities. Mean, Variance, and Shape of the Binomial Distribution. Sampling with or without Replacement.* Other Discrete Probability Distributions: *The Poisson Distribution. The Hypergeometric Distribution. The Multinomial Distribution. The Geometric Distribution.* Continuous Random Variables. Summary and Review of Terms.
- 3 The Normal Distribution** **114**
- Introduction. The Normal Probability Distribution: *Notation.* The Standard Normal Distribution: *Finding Probabilities of the Standard Normal Distribution. Finding Values of Z Given a Probability.* The Transformation of Normal Random Variables: *Using the Normal Transformation.* The

Relationship between X and Z and the Use of the Inverse Transformation. More Complex Problems. The Normal Distribution as an Approximation to Other Probability Distributions. Summary and Review of Terms.

4 *Sampling and Sampling Distributions* 146

Introduction. Sample Statistics as Estimators of Population Parameters: *Obtaining a Random Sample*. Sampling Distributions: *The Central Limit Theorem*. *The History of the Central Limit Theorem*. *The Standardized Sampling Distribution of the Sample Mean When σ Is Not Known*. *The Sampling Distribution of the Sample Proportion, P* . Estimators and Their Properties: *Applying the Concepts of Unbiasedness, Efficiency, Consistency, and Sufficiency*. Degrees of Freedom. Summary and Review of Terms.

5 *Confidence Intervals* 179

Introduction. Confidence Interval for the Population Mean When the Population Standard Deviation Is Known. Confidence Intervals for μ When σ Is Unknown—The t Distribution: *The t Distribution*. Large-Sample Confidence Intervals for the Population Proportion, p . The Finite-Population Correction Factor. Confidence Intervals for the Population Variance. Sample Size Determination. One-sided Confidence Intervals. Using the Computer. Summary and Review of Terms.

6 *Hypothesis Testing* 214

Introduction. Statistical Hypothesis Testing. A Two-Tailed, Large-Sample Test for the Population Mean: *Standardized Form of the Statistical Hypothesis Test*. A Two-Tailed, Small-Sample Test for the Population Mean. A Two-Tailed, Large-Sample Test for the Population Proportion. One-Tailed Tests. The p -Value: *The t Distribution, the Chi-Square Distribution, and Others*. *Two-Tailed Tests*. Tests Involving Finite Populations. Tests of Hypotheses about the Population Variance. The Probability of a Type II Error and the Power of the Test. Sample Size Determination for Hypothesis Tests. How Hypothesis Testing Works: *Some Comments and Caveats*. Using the Computer. Summary and Review of Terms.

Case One The Alaska Marine Highway 292

7 *The Comparison of Two Populations* 296

Introduction. Paired-Observations Comparisons: *Confidence Intervals*. A Test for the Difference between Two Population Means Using Independent Random Samples: *Confidence Intervals*. *A Small-Sample Test for the Difference between Two Population Means*. A Test for the Difference between

Two Population Means, Assuming Equal Population Variances: *Confidence Intervals*. A Large-Sample Test for the Difference between Two Population Proportions: *Confidence Intervals*. The *F*-Distribution and a Test for Equality of Two Population Variances: *A Statistical Test for Equality of Two Population Variances*. *Confidence Intervals*. Using the Computer. Summary.

Case Two Program Trading 340

8 Analysis of Variance 342

Introduction. The Hypothesis Test of Analysis of Variance: *The Test Statistic*. The Theory and the Computations of ANOVA: *The Sum of Squares Principle*. *The Degrees of Freedom*. *The Mean Squares*. *The Expected Values of the Statistics $MSTR$ and MSE under the Null Hypothesis*. *The F Statistic*. The ANOVA Table and Examples. Further Analysis: *The Tukey Pairwise Comparisons Test*. *Conducting the Tests*. *The Case of Unequal Sample Sizes, and Alternative Procedures*. Models, Factors, and Designs: *One-Factor versus Multifactor Models*. *Fixed-Effects versus Random-Effects Models*. *Experimental Design*. Two-Way Analysis of Variance: *The Two-Way ANOVA Model*. *The Hypothesis Tests in Two-Way ANOVA*. *Sums of Squares, Degrees of Freedom, and Mean Squares*. *The F Ratios and the Two-Way ANOVA Table*. *The Overall Significance Level*. *The Tukey Method for Two-Way Analysis*. *Extension of ANOVA to Three Factors*. *Two-Way ANOVA with One Observation per Cell*. Blocking Designs: *Randomized Complete Block Design*. *Latin Square Design*. Using the Computer.

Case Three New Coke, Old Coke, and Pepsi 417

9 Simple Linear Regression and Correlation 420

Introduction. The Simple Linear Regression Model. Estimation: The Method of Least Squares. Error Variance and the Standard Errors of Regression Estimators: *Confidence Intervals for the Regression Parameters*. Correlation. Hypothesis Tests about the Regression Relationship: *Other Tests*. How Good Is the Regression? Analysis of Variance Table and an *F* Test of the Regression Model. Residual Analysis and Checking for Model Inadequacies: *A Check for the Equality of Variance of the Errors*. *Testing for Missing Variables*. *Detecting a Curvilinear Relationship between Y and X* . *Detecting Deviations from the Normal Distribution Assumption*. Use of the Regression Model for Prediction: *Point Predictions*. *Prediction Intervals*. *A Confidence Interval for the Average Y , Given a Particular Value of X* . Using the Computer. Summary and Review of Terms.

Case Four The January Indicator 480

10	<i>Multiple Regression</i>	482
	Introduction. The k -Variable Multiple Regression Model: <i>The Estimated Regression Relationship</i> . The ANOVA Test of a Multiple Regression Model. How Good Is the Regression? Tests of the Significance of Individual Regression Parameters: <i>Some Quick and Simple Rules</i> . Testing the Validity of the Regression Model: <i>Residual Plots</i> . <i>Standardized Residuals</i> . <i>Outliers and Influential Observations</i> . <i>Lack of Fit and Other Problems</i> . Using the Multiple Regression Model for Prediction. Qualitative Independent Variables: <i>Interactions between Qualitative and Quantitative Variables</i> . Polynomial Regression: <i>Other Variables and Cross-Product Terms</i> . Nonlinear Models and Transformations: <i>Variance-Stabilizing Transformations</i> . <i>Regression with Dependent Indicator Variable</i> . Multicollinearity: <i>Causes of Multicollinearity</i> . <i>Detecting the Existence of Multicollinearity</i> . <i>Solutions to the Multicollinearity Problem</i> . Residual Autocorrelation and the Durbin-Watson Test. Partial F Tests and Variable Selection Methods: <i>Variable Selection Methods</i> . Using the Computer: <i>MINITAB</i> . <i>SAS</i> . <i>SPSSX</i> . The Matrix Approach to Multiple Regression Analysis: <i>What about Multicollinearity?</i> Summary and Review of Terms.	
	<i>Case Five Diamond State Telephone Company</i>	591
11	<i>Time Series and Forecasting</i>	594
	Introduction. Trend Analysis. Seasonality and Cyclical Behavior. The Ratio-to-Moving-Average Method: <i>The Cyclical Component of the Series</i> . <i>Forecasting a Multiplicative Series</i> . Exponential Smoothing Methods. A General Discussion of Forecasting: <i>The Forecast Function</i> . <i>Criteria for Forecast Evaluation</i> . Introduction to the Box-Jenkins Methodology: <i>The Backshift Operator Notation</i> . ARIMA Models and the Concept of Autocorrelation: <i>Autoregressive Models</i> . <i>Moving-Average Models</i> . <i>Mixed ARMA Processes</i> . Stationarity of Time Series and Integrated Models. Seasonal ARIMA Models. Identification of ARIMA Models: The First Stage of Box-Jenkins Analysis: <i>Standard Errors of the Estimators</i> . <i>The Identification of Autoregressive Processes</i> . <i>The Identification of Moving-Average Processes</i> . <i>The Identification of Mixed ARMA (p, q) Processes</i> . <i>Nonstationary Processes and Differencing</i> . <i>The Identification Procedure</i> . <i>Identification of Seasonal Time Series</i> . <i>Identification of Multiplicative ARIMA Models</i> . <i>Identification of Nonmultiplicative ARIMA Models</i> . Model Estimation. Model Diagnostic Checking: <i>Chi-Square Tests</i> . Forecasting: <i>How Are ARIMA Forecasts Produced?</i> Further Ramifications of ARIMA Models. Using the Computer. Summary and Review of Terms.	
	<i>Case Six Forecasting the Demand for Ships</i>	692

- 12 *More Descriptive Statistics: Index Numbers and Exploratory Data Analysis* 694**
- Introduction. Simple Index Numbers. Composite Index Numbers: *The Laspeyres Index. The Paasche-Index. The Consumer Price Index.* Methods of Displaying Data: *Pie Charts. Bar and Column Graphs. Frequency Polygons and Ogives. A Caution About Graphs.* Exploratory Data Analysis: *Stem-and-Leaf Displays. Box Plots.* Other Statistics: *The Coefficient of Variation. The Geometric Mean.* Using the Computer. Summary and Review of Terms.
- 13 *Nonparametric Methods and Chi-Square Test* 736**
- Introduction. The Sign Test: *The McNemar Test. The Cox and Stuart Test. Extension of the Cox and Stuart Test as a Test for Correlation.* The Runs Test—A Test for Randomness: *Large-Sample Properties. The Wald-Wolfowitz Test.* The Mann-Whitney *U* Test: *The Computational Procedure.* The Wilcoxon Signed-Rank Test: *The Paired-Observations Two-Sample Test. Large-Sample Version of the Test. A Test for the Mean or Median of a Single Population.* The Kruskal-Wallis Test—A Nonparametric Alternative to One-Way ANOVA: *Further Analysis.* The Friedman Test for a Randomized Block Design. The Spearman Rank Correlation Coefficient. Chi-Square Test for Goodness of Fit: *A Goodness-of-Fit Test for the Multinomial Distribution.* Contingency Table Analysis—A Chi-Square Test for Independence: *Degrees of Freedom. Measuring the Degree of Dependence. Further Analysis. Higher-Order Contingency Tables.* Chi-Square Test for Equality among Proportions: *The Median Test.* Using the Computer. Summary and Review of Terms.
- Case Seven *The Nine Nations of North America* 824**
- 14 *Bayesian Statistics and Decision Analysis* 829**
- Introduction. Bayes' Theorem and Discrete Probability Models. Bayes' Theorem and Continuous Probability Distributions: *The Normal Probability Model. Credible Sets.* The Evaluation of Subjective Probabilities: *Assessing a Normal Prior Distribution. The de Finetti Game.* Decision Analysis: An Overview: *Actions. Chance Occurrences. Probabilities. Final Outcomes. Additional Information. Decision.* Decision Trees: *The Payoff Table.* Handling Additional Information Using Bayes' Theorem: *A Method of Assessing Utility.* The Value of Information. Using the Computer. Summary and Review of Terms.
- Case Eight *Getting Bumped* 890**

15	<i>Multivariate Analysis</i>	894
	Introduction. The Multivariate Normal Distribution: <i>Evaluating Probabilities. Sample Estimates. The Sampling Distribution of the Sample Centroids. Hypothesis Tests about the Population Centroid μ. The Shape and the Orientation of the Bivariate Normal Density.</i> Multivariate Analysis of Variance: <i>MANOVA Assumptions. Wilks' Lambda. Other Statistics. Further Analysis. Two-Way MANOVA.</i> Discriminant Analysis: <i>Developing a Discriminant Function. Evaluating the Performance of the Model. Discriminant Analysis with More than Two Groups.</i> Principal Components and Factor Analysis: <i>Principal Components. The Extraction of the Components. Factor Analysis. The Extraction of Factors. The Rotation of Factors.</i> A Brief Description of Other Multivariate Methods: <i>Cluster Analysis. Conjoint Analysis. Canonical Correlation Analysis. Multidimensional Scaling.</i> Using the Computer. Summary and Review of Terms.	
16	<i>Sampling Methods</i>	966
	Introduction. Nonprobability Sampling and Bias. Stratified Random Sampling: <i>Practical Applications. Confidence Intervals. Stratified Sampling for the Population Proportion. What Do We Do When the Population Strata Weights Are Unknown? How Many Strata Should We Use? Post-Sampling Stratification. Optimum Allocation.</i> Cluster Sampling: <i>Single-Stage Cluster Sampling for the Population Mean. Single-Stage Cluster Sampling for the Population Proportion. Two-Stage Cluster Sampling. The Relation with Stratified Sampling.</i> Systematic Sampling: <i>The Advantages of Systematic Sampling. Estimation of the Population Mean in Systematic Sampling.</i> Other Methods. Nonresponse. Summary and Review of Terms.	
	<i>Case Nine The Boston Redevelopment Authority</i>	992
	<i>Appendix A: References</i>	994
	<i>Appendix B: Answers to Most Odd-Numbered Problems</i>	999
	<i>Appendix C: Statistical Tables</i>	1013
	<i>Appendix D: The Federal Reserve Data Set</i>	1047
	<i>Index</i>	1051

*Complete
Business
Statistics*

*Introduction
and
Descriptive
Statistics*

