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 29

7 *The Comparison of Two Populations* 29

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 Regresion Estimators: *Confidence Intervals for the Regresion 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 Multiple Regression 482

Introduction. The *k*-Variable Multiple Regression Model: *The Estimated Regression Relationship*. The ANOVA Test of a Multiple Regression Model. How Good Is the Regression? Tests of the Significance of Individual Regression Parameters: *Some Quick and Simple Rules*. Testing the Validity of the Regression Model: *Residual Plots*. *Standardized Residuals*. *Outliers and Influential Observations*. *Lack of Fit and Other Problems*. Using the Multiple Regression Model for Prediction. Qualitative Independent Variables: *Interactions between Qualitative and Quantitative Variables*. Polynomial Regression: *Other Variables and Cross-Product Terms*. Nonlinear Models and Tranformations: *Variance-Stabilizing Transformations*. *Regression with Dependent Indicator Variable*. Multicollinearity: *Causes of Multicollinearity*. *Detecting the Existence of Multicollinearity*. *Solutions to the Multicollinearity Problem*. Residual Autocorrelation and the Durbin-Watson Test. Partial *F* Tests and Variable Selection Methods: *Variable Selection Methods*. Using the Computer: *MINITAB*. *SAS*. *SPSSX*. The Matrix Approach to Multiple Regression Analysis: *What about Multicollinearity?* Summary and Review of Terms.

Case Five Diamond State Telephone Company 591

11 Time Series and Forecasting 594

Introduction. Trend Analysis. Seasonality and Cyclical Behavior. The Ratio-to-Moving-Average Method: *The Cyclical Component of the Series*. *Forecasting a Multiplicative Series*. Exponential Smoothing Methods. A General Discussion of Forecasting: *The Forecast Function*. *Criteria for Forecast Evaluation*. Introduction to the Box-Jenkins Methodology: *The Backshift Operator Notation*. ARIMA Models and the Concept of Autocorrelation: *Autoregressive Models*. *Moving-Average Models*. *Mixed ARMA Processes*. Stationarity of Time Series and Integrated Models. Seasonal ARIMA Models. Identification of ARIMA Models: The First Stage of Box-Jenkins Analysis: *Standard Errors of the Estimators*. *The Identification of Autoregressive Processes*. *The Identification of Moving-Average Processes*. *The Identification of Mixed ARMA (p, q) Processes*. *Nonstationary Processes and Differencing*. *The Identification Procedure*. *Identification of Seasonal Time Series*. *Identification of Multiplicative ARIMA Models*. *Identification of Nonmultiplicative ARIMA Models*. Model Estimation. Model Diagnostic Checking: *Chi-Square Tests*. Forecasting: *How Are ARIMA Forecasts Produced?* Further Ramifications of ARIMA Models. Using the Computer. Summary and Review of Terms.

Case Six Forecasting the Demand for Ships 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 Freidman 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 Multivariate Analysis 894

Introduction. The Multivariate Normal Distribution: *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.* Multivariate Analysis of Variance: *MANOVA Assumptions. Wilks' Lambda. Other Statistics. Further Analysis. Two-Way MANOVA.* Discriminant Analysis: *Developing a Discriminant Function. Evaluating the Performance of the Model. Discriminant Analysis with More than Two Groups.* Principal Components and Factor Analysis: *Principal Components. The Extraction of the Components. Factor Analysis. The Extraction of Factors. The Rotation of Factors.* A Brief Description of Other Multivariate Methods: *Cluster Analysis. Conjoint Analysis. Canonical Correlation Analysis. Multidimensional Scaling.* Using the Computer. Summary and Review of Terms.

16 Sampling Methods 966

Introduction. Nonprobability Sampling and Bias. Stratified Random Sampling: *Practical Applications. Confidence Intervals. Stratified Sampling for the Population Proportion. What Do We Do When the Population Strata Weights Are Unknown? How Many Strate Should We Use? Post-Sampling Stratification. Optimum Allocation.* Cluster Sampling: *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.* Systematic Sampling: *The Advantages of Systematic Sampling. Estimation of the Population Mean in Systematic Sampling.* Other Methods. Nonresponse. Summary and Review of Terms.

Case Nine The Boston Redevelopment Authority 992

Appendix A: References 994

Appendix B: Answers to Most Odd-Numbered Problems 999

Appendix C: Statistical Tables 1013

Appendix D: The Federal Reserve Data Set 1047

Index 1051