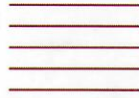


# Statistics

Concepts and Applications

---



A m i r D . A c z e l



# Statistics

Concepts and Applications



PADI  
INSTRUCTOR 33913  
Albrecht Salm

del 2000





## Preface

We live in the information age. The half-life of human knowledge is currently ten years overall, and only two years in science. This means that new knowledge accumulates so fast that in ten years (and a mere two years in science), half of what we now know will no longer be valid and will be replaced with new knowledge. The methods for gathering, assessing, understanding and interpreting, and drawing inferences from data constitute the field of *statistics*.

Yet, despite the great importance of the field of statistics, it seems that we have not been able to teach this discipline well. In almost two decades of dealing with students of different levels, backgrounds, and prior statistical education, I have rarely met a student who—a semester or more after completing an introductory statistics course—could demonstrate a true understanding of basic statistical concepts. I believe that one of the culprits is poorly written statistics textbooks.

Most statistics textbooks have fallen into one of two categories. In the first group are books that are theoretically correct, but abstract, mathematical, and intimidating to the student. They have no direct relevance to, or bearing on, the real world. These books perpetuate the well-known “Oh no! Not statistics!” mentality among beginning students. In the second category are books designed to be very user friendly. These books aim at what their authors perceive as the level of understanding of today’s student—they focus on simple “cookbook” procedures and calculations. In my opinion, such books miss the boat completely. They tend to avoid the important concepts, their explanations tend to be superficial, and ultimately they do not really motivate

or teach statistics. Students who use such books come out of the introductory course with no lasting understanding of this important discipline or appreciation of its applications. Days after taking the final exam, students cannot remember what a  $p$ -value is, and in the future they are likely to misuse statistical methods and misinterpret results.

### WHAT’S DIFFERENT, AND WHY?

The key features of this book that distinguish it from the competition are

- Superior treatment of the topics that usually cause the most difficulty for students, such as sampling, confidence intervals, and hypothesis testing.
- Optimal, research-based choice of topics to be covered.
- Careful selection of stimulating, relevant, and instructive examples that motivate the student and clarify key concepts.
- Wide variety of problems covering all levels and application areas.
- A flexible, modern introduction to statistical use of the computer, which stresses application rather than “syntax.”
- A unique cross-referencing system for key concepts.
- A friendly writing style and an approach that makes statistics understandable and interesting.

In writing this book, I felt that I could make statistics both interesting and useful to students—I have been doing this for many years, teaching stu-

---



## Contents in Brief

---

CHAPTER 1 Information Everywhere 2

*Interview with Melissa Anderson 67*

CHAPTER 2 What Are the Chances? 70

*Interview with Alan Dershowitz 107*

CHAPTER 3 Chance Quantities 112

*Interview with Arnold Zellner 151*

CHAPTER 4 The Bell-Shaped Curve 154

*Interview with Bud Goode 185*

CHAPTER 5 Let's Take a Sample 190

*Interview with Bradley Efron 239*

CHAPTER 6 Trial by Probability 244

*Interview with Lee Wilkinson 309*

CHAPTER 7 Making Comparisons 314

*Interview with Toni Falvo 351*

CHAPTER 8 Making Extended  
Comparisons 356

*Interview with Nancy Kirkendall 401*

CHAPTER 9 Is There a Relationship? 406

*Interview with Mary Guinan 461*

CHAPTER 10 The Quest for Quality 464

*Interview with Phil Crosby 481*



---



# Contents

---

## CHAPTER 1 Information Everywhere 2

- 1.1 Introduction 2
  - Samples and Populations 5
  - Data and Data Collection 5
  - Data Distributions 7
- 1.2 Methods of Displaying Data 9
  - Pie Charts 9
  - Bar Graphs 12
  - The Histogram 13
  - Stemplots (Stem-and-Leaf Displays) 15
  - Frequency Polygons and Ogives 19
  - A Caution about Graphs 21
- 1.3 Percentiles, Quartiles, and the Median 35
  - Percentiles 35
  - Quartiles 35
  - The Median 36
- 1.4 Box Plots 42
- 1.5 Measures of Centrality 45
- 1.6 Measures of Variability 50
  - The Empirical Rule 55
- 1.7 Caution, Care with Numbers, and Ethics 56
  - Ethics 57

*Interview with Melissa Anderson 67*

## CHAPTER 2 What Are the Chances? 70

- 2.1 Introduction 72
- 2.2 Basic Definitions: Events, Sample Space, and Probabilities 75
- 2.3 Basic Rules for Probability 81
  - The Range of Values 81
  - The Rule of Complements 82
  - The Rule of Unions 83
  - Mutually Exclusive Events 83

- 2.4 Conditional Probability 87
- 2.5 Independence of Events 92
  - Product Rules for Independent Events 93
- 2.6 Bayes' Theorem 97

*Interview with Alan Dershowitz 107*

## CHAPTER 3 Chance Quantities 112

- 3.1 Random Variables 115
  - Discrete and Continuous Random Variables 120
- 3.2 Expected Values of Discrete Random Variables 124
  - The Variance and the Standard Deviation of a Random Variable 127
  - Some Properties of Means and Variances 129
- 3.3 The Binomial Distribution 130
  - The Binomial Table 132
  - Mean, Variance, and Shape of the Binomial Distribution 135
  - Sampling with or without Replacement 137
- 3.4 Continuous Random Variables 140

*Interview with Arnold Zellner 151*

## CHAPTER 4 The Bell-Shaped Curve 154

- 4.1 The Normal Probability Distribution 156
  - The Shape of the Normal Distribution 158
  - Notation 159
- 4.2 The Standard Normal Distribution 159
  - Finding Probabilities of the Standard Normal Distribution 160
  - Finding Values of Z Given a Probability 164

- 4.3 The Transformation of Normal Random Variables 167
  - Using the Normal Transformation 169
- 4.4 The Relationship between X and Z and the Use of the Inverse Transformation 173
- 4.5 The Normal Distribution as an Approximation to Other Probability Distributions 176
- 4.6 Normal Data 178

*Interview with Bud Goode 185*

## CHAPTER 5 Let's Take a Sample 190

- 5.1 Statistical Inference 192
  - Sampling Methods and Statistical Inference 196
- 5.2 Sampling Distributions 199
- 5.3 The Sampling Distribution of the Sample Mean 200
- 5.4 Confidence Intervals 205
  - Confidence Interval for the Population Mean When the Population Standard Deviation Is Known 205
- 5.5 Confidence Intervals for  $\mu$  When  $\sigma$  Is Unknown—the  $t$  Distribution 214
  - The  $t$  Distribution 214
- 5.6 The Sampling Distribution of the Sample Proportion 222
  - Large-Sample Confidence Intervals for the Population Proportion 223
- 5.7 Sample Size Determination 228
- 5.8 Epilogue: The Polls 234

*Interview with Bradley Efron 239*

## CHAPTER 6 Trial by Probability 244

- 6.1 Introduction 247
- 6.2 Statistical Hypothesis Testing 257
- 6.3 Tests about the Population Mean,  $\mu$ , When the Population Standard Deviation,  $\sigma$ , Is Known 257
  - Standardizing the Test 257
  - One-Tailed Tests 260
  - The  $p$ -Value 265
- 6.4 Tests about the Population mean,  $\mu$ , When the Population Standard Deviation,  $\sigma$ , Is Unknown 274
- 6.5 Large-Sample Tests for the Population Proportion 284

- 6.6 How Hypothesis Testing Works 291
  - Some Comments and Caveats 294
- 6.7 The Probability of a Type II Error and the Power of the Test 297

*Interview with Lee Wilkinson 309*

## CHAPTER 7 Making Comparisons 314

- 7.1 Introduction 316
- 7.2 Paired-Observations Comparisons 316
  - Experimental Design 316
  - Confidence Intervals 318
- 7.3 A Test for the Difference between Two Population Means, Using Independent Random Samples 322
  - Confidence Intervals 325
- 7.4 A Test for the Difference between Two Population Means, Assuming Equal Population Variances 330
  - Confidence Intervals 333
- 7.5 A Large-Sample Test for the Difference between Two Population Proportions 336
  - Confidence Intervals 338

*Interview with Toni Falvo 351*

## CHAPTER 8 Making Extended Comparisons 356

- 8.1 Introduction 358
- 8.2 The Hypothesis Test of Analysis of Variance (ANOVA) 359
  - The Test Statistic and the  $F$  Distribution 361
- 8.3 The Theory and the Computations of ANOVA 365
  - The Sum of Squares Principle 369
  - The Degrees of Freedom 372
  - The Mean Squares 373
  - The Expected Values of the Statistics MSTR and MSE under the Null Hypothesis 373
  - The  $F$  Statistic 374
  - The ANOVA Table 374
  - More on Experimental Design 379
  - Randomized Complete Block Design 379
- 8.4 Contingency Table Analysis: A Chi-Square Test for Independence 384
  - The Chi-Square Distribution 386
  - A Chi-Square Test for Equality of Proportions 389

*Interview with Nancy Kirkendall 401*



**CHAPTER 9 Is There a Relationship? 406**

- 9.1 Introduction 408
- 9.2 Statistical Models 409
- 9.3 The Simple Linear Regression Model 410
- 9.4 Estimation: The Method of Least Squares 414
- 9.5 Error Variance and the Standard Errors of Regression Estimators 421
  - Confidence Intervals for the Regression Parameters 424
  - A Hypothesis Test for the Existence of a Linear Regression Relationship 425
- 9.6 How Good Is the Regression? 431
  - Analysis of Variance Table and an  $F$  Test of the Regression Model 435
- 9.7 Residual Analysis and Checking for Model Inadequacies 437
  - A Check for the Equality of Variance of the Errors 438
  - Testing for Missing Variables 438
  - Detecting a Curvilinear Relationship between  $Y$  and  $X$  439
  - Detecting Deviations from the Normal Distribution Assumption 439
- 9.8 Use of the Regression Model for Prediction 444
  - Point Predictions 444
  - Prediction Intervals 445
  - A Confidence Interval for the Average  $Y$ , Given a Particular Value of  $X$  446
- 9.9 Correlation 448

*Interview with Mary Guinan 461*

**CHAPTER 10 The Quest for Quality 464**

- 10.1 Control Charts 466
- 10.2 The  $\bar{x}$  Chart 468
- 10.3 The  $R$  Chart and the  $s$  Chart 472
  - The  $R$  Chart 472
  - The  $s$  Chart 473
- 10.4 The  $p$  Chart 474
- 10.4 The  $c$  Chart 476

*Interview with Phil Crosby 481*

**GLOSSARY 483****BIBLIOGRAPHY 487****APPENDIX A: DATA TABLES 489****APPENDIX B: STATISTICAL TABLES 493****APPENDIX C: ANSWERS TO MOST ODD-NUMBERED PROBLEMS 519****INDEX 531**