

CONTENTS

| | |
|--|-----------|
| Preface | v |
| 1 Statistics: Randomness and Regularity | 1 |
| 1.1 Statistics: What's in a word? | 3 |
| 1.2 Knowing how statistics is used: Goals for the reader | 4 |
| Understanding what can go wrong | 7 |
| Understanding statistical terms | 9 |
| 1.3 Central ideas in statistics | 9 |
| Randomness and regularity: Twins in tension | 9 |
| Randomness in regularity | 10 |
| Two examples in the study of randomness and regularity | 11 |
| Probability: What are the chances? | 13 |
| Variables: The names we give things | 14 |
| Variables, values, and elements | 14 |
| Theoretical variables and empirical variables | 14 |
| Constants | 17 |
| 1.4 Users of statistics | 17 |
| 1.5 Relationship of statistics to mathematics, pencils, and computers | 20 |
| 1.6 Summary | 21 |
| Additional Readings | 23 |
| Exercises | 23 |
| 2 Collection of Data | 29 |
| 2.1 Defining the variables | 31 |
| 2.2 Observational data: Problems and possibilities | 32 |

| | | |
|---|----|-----------|
| Population versus sample | 32 | |
| Selection of the sample: Making sure the pot is stirred | 33 | |
| Random sample: What is it? | 35 | |
| Convenience sample: How to produce a “poor” sample | 36 | |
| Selecting proper samples | 36 | |
| Selection of variables on which to collect observational data | 37 | |
| 2.3 Errors and “errors” in collecting observational data | | 39 |
| Sampling error: The “error” that is not a mistake | 39 | |
| Nonresponse error: Result of rude, rushed, and reticent respondents | | 41 |
| Response errors | 42 | |
| 2.4 Experimental data: Looking for the causes of outcomes | | 45 |
| Experimental group and control group | 46 | |
| Selecting the experimental and control groups | 47 | |
| Problems with experimenting on people | 47 | |
| Role of statistics in experimentation | 50 | |
| Putting it all together: Does class size affect school performance? | | 53 |
| 2.5 Data matrix/Data file | | 55 |
| 2.6 Summary | | 57 |
| Additional Readings | | 59 |
| Exercises | | 60 |

3 Description of Data: Graphs and Tables **71**

| | | |
|---|----|------------|
| 3.1 Graphs: Picturing data | | 73 |
| Creating statistical graphs | 73 | |
| Types of graphs | 75 | |
| 3.2 Categorical variables: Pie charts and bar graphs | | 75 |
| Graphing one categorical variable | 75 | |
| Graphing two categorical variables | 77 | |
| 3.3 Metric variables: Plots and histograms | | 79 |
| Graphing one metric variable | 80 | |
| Graphing two metric variables | 89 | |
| Time series plot | 91 | |
| 3.4 Creating maps from data | | 94 |
| 3.5 Graphing: Standards for excellence | | 96 |
| “The least ink”: Is the simplest graph best? | 96 | |
| “Chartjunk”: A new name for garbage | 98 | |
| Data density | 98 | |
| “Revelation of the complex” | 99 | |
| 3.6 Tables: Turning can be timely | | 101 |
| 3.7 Summary | | 104 |
| Additional Readings | | 106 |
| Exercises | | 106 |

| | | |
|----------|--|------------|
| 4 | Description of Data: Computing Summary Statistics | 129 |
| 4.1 | Averages: Let us count the ways | 131 |
| | Mode: The hostess with the mostes' 131 | |
| | Median: Counting to the middle 134 | |
| | Mean: Balancing the seesaw 138 | |
| | Mode, median, or mean? 140 | |
| 4.2 | Variety: Measuring the spice of life | 141 |
| | Range: Lassoing the two extreme values 142 | |
| | Standard deviation: The crucial deviant 143 | |
| 4.3 | Standard error of the means | 149 |
| 4.4 | Standard scores: Comparing apples and oranges | 150 |
| 4.5 | Gain in simplicity, loss of information | 153 |
| | Replacing the data with a graph 153 | |
| | Replacing the data with a summary value 154 | |
| 4.6 | Real estate data: Out-of-sight prices | 155 |
| 4.7 | Summary | 156 |
| | Additional Readings | 158 |
| | Formulas | 159 |
| | Exercises | 162 |
| 5 | Probability | 177 |
| 5.1 | How to find probabilities | 180 |
| | Equally likely events 180 | |
| | Relative frequency 180 | |
| | Using subjective probabilities 183 | |
| 5.2 | Computations with probabilities | 184 |
| | Adding probabilities 184 | |
| | Multiplying probabilities 185 | |
| 5.3 | Odds: The opposite of probabilities | 185 |
| 5.4 | Probability distributions for discrete variables | 187 |
| | Binomial distribution 187 | |
| | Poisson distribution 190 | |
| | Hypergeometric distribution 192 | |
| | Displaying probabilities in graphs and tables 192 | |
| | Computations with probabilities 193 | |
| 5.5 | Probability distributions for continuous variables | 194 |
| | Standard normal distribution: The bell curve 194 | |
| | The t -distribution 196 | |
| | Chi-square distribution 199 | |
| | F -distribution 201 | |
| | Need for normally distributed data 202 | |
| 5.6 | Using probabilities to check on assumptions | 202 |
| | Is it a fair coin? 202 | |

| | | |
|--|-----|-----|
| Is it a fair workplace? | 203 | |
| Is it an evenly split electorate? | 205 | |
| 5.7 Decision analysis: Using probabilities to make decisions | | 207 |
| 5.8 Summary | | 210 |
| Additional Readings | | 212 |
| Formulas | | 212 |
| Exercises | | 216 |

6 Drawing Conclusions: Estimation 231

| | | |
|---|-----|-----|
| 6.1 Sample statistic and population parameter | | 233 |
| 6.2 Point estimation | | 235 |
| What is a “good” point estimate? | 235 | |
| A strategic use of the point estimate: How many tanks did the Germans have? | 237 | |
| 6.3 Interval estimation: More room to be correct | | 235 |
| Length of confidence interval | 241 | |
| Confidence intervals for differences | 245 | |
| 6.4 Summary | | 247 |
| Additional Readings | | 248 |
| Formulas | | 249 |
| Exercises | | 251 |

7 Drawing conclusions: Hypothesis testing 265

| | | |
|---|-----|-----|
| 7.1 The hypothesis as a question | | 267 |
| Null hypothesis | 267 | |
| Alternative hypothesis | 268 | |
| Errors in answering the question | 269 | |
| 7.2 How to answer the question posed by the null hypothesis | | 270 |
| Probability: The p -value | 271 | |
| Mechanics of hypothesis testing | 273 | |
| To reject or not to reject the null hypothesis | 274 | |
| Causal effect: A skip too far | 275 | |
| A little statistical theory and a game on the computer | 276 | |
| 7.3 Significance level | | 279 |
| 7.4 Testing a population proportion | | 281 |
| 7.5 Difference between two population proportions | | 284 |
| Testing the null hypothesis | 285 | |
| Estimating the difference | 285 | |
| 7.6 Testing hypotheses versus constructing confidence intervals | | 286 |
| 7.7 Statistical versus substantive significance | | 287 |

| | | |
|----------|--|------------|
| 7.8 | Applications: When to reject the null hypothesis | 289 |
| | Psychology experiment on cooperation and competition | 289 |
| | Community study of blue-collar workers | 290 |
| 7.9 | Summary | 291 |
| | Additional Readings | 294 |
| | Formulas | 294 |
| | Exercises | 299 |
| 8 | Relationships Between Variables | 313 |
| 8.1 | Four questions about two variables and their relationship | 316 |
| | Question 1. Relationship between the variables in the data? | 319 |
| | Question 2. Strength of the relationship? | 320 |
| | Question 3. Relationship in the population? | 321 |
| | Question 4. Causal relationship? | 321 |
| 8.2 | Prediction | 322 |
| 8.3 | Independent and dependent variables | 323 |
| 8.4 | Different types of variables: Categorical, rank, metric | 324 |
| 8.5 | Return to the question of causality | 326 |
| | Role of other variables | 327 |
| | Role of time | 328 |
| | Multiple causal factors | 329 |
| 8.6 | Summary | 330 |
| | Additional Readings | 332 |
| | Exercises | 332 |
| 9 | Chi-square Analysis for Two Categorical Variables | 345 |
| 9.1 | Analysis of the data: Are there trustworthy differences in attitude? | 347 |
| | Bar graphs | 348 |
| | Summary computations with categorical variables | 351 |
| 9.2 | Question 1. Relationship between the variables? | 351 |
| 9.3 | Question 2. Strength of the relationship? | 353 |
| | Phi in the sample | 353 |
| | Phi in the population | 355 |
| 9.4 | Question 3. Relationship in the populations? | 355 |
| | Setting up the null hypothesis | 355 |
| | Testing the null hypothesis | 356 |
| | From chi-square to p -value | 357 |
| | Degrees of freedom for chi-square analysis | 359 |
| 9.5 | Question 4. Causal relationship? | 360 |

| | | |
|------|--|-----|
| 9.6 | Larger tables: A banquet of possibilities | 360 |
| | Question 1. Relationship between the variables? | 362 |
| | Question 2. Strength of the relationship? | 362 |
| | Question 3. Relationship in the populations? | 363 |
| | Question 4. Causal relationship? | 363 |
| 9.7 | Summary | 364 |
| | Additional Reading | 365 |
| | Formulas | 365 |
| | Exercises | 370 |
| | | |
| 10 | Regression and Correlation for Two Metric Variables | 397 |
| 10.1 | Question 1. Relationship between the variables? | 401 |
| | Graphing the data in a scatterplot | 401 |
| | Learning from the scatterplot | 404 |
| | Linear relationships | 404 |
| 10.2 | Question 2a. Strength of the relationship? | 406 |
| | Is r positive or negative? Large or small? | 406 |
| | Four scatterplots: From strong to weak relationships | 407 |
| | Interpretation of r : An issue of inexactness | 410 |
| 10.3 | Question 2b. Form of the relationship? | 411 |
| | A line through the middle of the points | 412 |
| | How to find the regression line: The least squares principle | 414 |
| | Predicting with regression analysis: From fat to calories | 416 |
| | Magnitudes of effects: Interpretation of r -square | 417 |
| | Correlation or regression? The more the merrier | 422 |
| | Regression analysis for data on change | 425 |
| 10.4 | Question 3. Relationship in the population? | 426 |
| | Confidence interval approach | 426 |
| | Hypothesis testing using t | 427 |
| | Hypothesis testing using F | 428 |
| 10.5 | Warning: What you measure is what you get | 429 |
| 10.6 | How to be smart using dummy variables | 431 |
| | Categorical independent variable with two values and metric dependent variable | 431 |
| | Categorical dependent variable with two values and metric independent variable | 434 |
| 10.7 | Question 4. Causal relationship? | 435 |
| 10.8 | Summary | 435 |
| | Additional Readings | 435 |
| | Formulas | 438 |
| | Exercises | 439 |
| | | 442 |

| | | |
|-----------|---|------------|
| 11 | Analysis of Variance for a Categorical and a Metric Variable | 473 |
| 11.1 | Analysis of variance: Comparing the mean-ings of things | 475 |
| 11.2 | Question 1. Relationship between violent crime rate and region? | 478 |
| | Scatterplot | 478 |
| | Boxplot: A simpler view of the data | 479 |
| 11.3 | Question 2. Strength of the relationship? | 481 |
| | Region variable | 481 |
| | Residual variable | 483 |
| | Effect of both region and residual variable: Total sum of squares | 484 |
| | Measuring the strength of the relationship | 485 |
| | Explained amounts of variation | 486 |
| 11.4 | Question 3. Could the relationship have occurred by chance alone? | 490 |
| | The null hypothesis | 490 |
| | p -value from F | 491 |
| | Going beyond the F -test: Making mean comparisons | 494 |
| 11.5 | Question 4. Causal relationship? | 496 |
| 11.6 | Analysis of variance: A bird's-eye review | 496 |
| 11.7 | Matched pair analysis: Two observations per unit | 498 |
| | A t -test | 498 |
| | The sign test: A simple yes or no | 501 |
| 11.8 | Summary | 502 |
| | Additional Readings | 504 |
| | Formulas | 504 |
| | Exercises | 507 |
| 12 | Rank Methods for Two Rank Variables | 525 |
| 12.1 | Two rank variables with words as the values | 527 |
| | Question 1. Relationship between identification and interest? | 528 |
| | Question 2. Strength of the relationship? | 530 |
| | Question 3. Relationship in the population? | 532 |
| | Question 4. Causal relationship? | 533 |
| 12.2 | Ranking numbers as values: How are the Phillies doing? | 534 |
| | Question 1. Relationship in the data? | 534 |
| | Question 2. Strength of the relationship? | 535 |
| | Question 3. Did the relationship occur by chance? | 536 |
| | Question 4. Causal relationship? | 537 |
| 12.3 | Summary | 537 |
| | Additional Readings | 538 |
| | Formulas | 538 |
| | Exercises | 542 |

| | | |
|-------------|---|------------|
| 13 | Multivariate analysis | 559 |
| 13.1 | Partial phis: Three categorical variables | 561 |
| | Control for a third variable: The neutralizing game | 562 |
| | Partial phi | 563 |
| 13.2 | Multiple regression with metric variables | 567 |
| | Question 1. Relationship in the data? | 567 |
| | Question 2b. Form of the relationship? Partial regression coefficients | 568 |
| | Question 2a. Strength of the relationship? Partial correlation coefficients | 571 |
| | Question 3. Relationship in the population? | 574 |
| 13.3 | Multiple regression with a dummy variable | 576 |
| 13.4 | Two-way analysis of variance | 581 |
| | One-way analysis with time of day only | 582 |
| | One-way analysis with route only | 583 |
| | Two-way analysis with time of day and route | 584 |
| | A second study with interaction effects | 589 |
| 13.5 | Establishing causality | 592 |
| 13.6 | Summary | 593 |
| | Additional Readings | 595 |
| | Formulas | 596 |
| | Exercises | 598 |
| | | |
| 14 | Statistics in Everyday Life | 617 |
| 14.1 | Stepping stones to statistical sophistication | 618 |
| 14.2 | Approaching numbers with care | 622 |
| 14.3 | Data and statistical methods | 624 |
| 14.4 | How things can go wrong | 626 |
| | Dangers in the collection of data | 626 |
| | Special problems of survey research | 628 |
| | Misuses of analysis methods | 631 |
| | Misuses of statistical inference | 632 |
| | Misuses in interpretation of numbers | 633 |
| 14.5 | Statistics and Big Brother | 634 |
| 14.6 | Ending on the upbeat | 636 |
| | Additional Readings | 637 |
| | Exercises | 638 |
| | Glossary | 641 |
| | Statistical Tables | 648 |
| | Answers to Odd-Numbered Exercises | 673 |
| | Index | 731 |