

# Contents

## Preface xv

## 1 Introduction: Variables and Processes in Statistics 1

Types of Variables: Categorical or Quantitative .....	2
<i>Students Talk Stats</i> : Identifying Types of Variables .....	3
Handling Data for Two Types of Variables .....	5
Roles of Variables: Explanatory or Response .....	7
Statistics as a Four-Stage Process .....	9

Summary 11 / Exercises 11

## PART I Data Production 16

## 2 Sampling: Which Individuals Are Studied 18

Sources of Bias in Sampling: When Selected Individuals Are Not Representative .....	18
Probability Sampling Plans: Relying on Randomness .....	20
The Role of Sample Size: Bigger Is Better If the Sample Is Representative ..	21
From Sample to Population: To What Extent Can We Generalize? .....	22
<i>Students Talk Stats</i> : Seeking a Representative Sample .....	23

Summary 25 / Exercises 25

## 3 Design: How Individuals Are Studied 30

3.1 Various Designs for Studying Variables .....	30
Identifying Study Design .....	32
Observational Studies versus Experiments: Who Controls the Variables? ..	33
Errors in Studies' Conclusions: The Imperfect Nature of Statistical Studies ..	35
3.2 Sample Surveys: When Individuals Report Their Own Values .....	38
Sources of Bias in Sample Surveys .....	38
3.3 Observational Studies: When Nature Takes Its Course .....	46
Confounding Variables and Causation .....	46
Paired or Two-Sample Studies .....	48
Prospective or Retrospective Studies: Forward or Backward in Time .....	49
3.4 Experiments: When Researchers Take Control .....	51
Randomized Controlled Experiments .....	52
Double-Blind Experiments .....	53
"Blind" Subjects .....	53
"Blind" Experimenters .....	54
Pitfalls in Experimentation .....	55
Modifications to Randomization .....	57

<i>Students Talk Stats: Does Watching TV Cause ADHD? Considering Study Design</i> .....	63
Summary 63 / Exercises 65	

## PART II **Displaying and Summarizing Data** 70

### 4 Displaying and Summarizing Data for a Single Variable 72

4.1 Single Categorical Variable .....	72
Summaries and Pie Charts .....	72
The Role of Sample Size: Why Some Proportions Tell Us More Than Others Do .....	74
Bar Graphs: Another Way to Visualize Categorical Data .....	75
Mode and Majority: The Value That Dominates .....	77
Revisiting Two Types of Bias .....	77
<i>Students Talk Stats: Biased Sample, Biased Assessment</i> .....	78
4.2 Single Quantitative Variables and the Shape of a Distribution .....	82
Thinking about Quantitative Data .....	83
Stemplots: A Detailed Picture of Number Values .....	85
Histograms: A More General Picture of Number Values .....	89
4.3 Center and Spread: What's Typical for Quantitative Values, and How They Vary .....	93
Five-Number Summary: Landmark Values for Center and Spread .....	93
Boxplots: Depicting the Key Number Values .....	95
Mean and Standard Deviation: Center and Spread in a Nutshell .....	98
4.4 Normal Distributions: The Shape of Things to Come .....	108
The 68-95-99.7 Rule for Samples: What's "Normal" for a Data Set .....	110
From a Histogram to a Smooth Curve .....	113
Standardizing Values of Normal Variables: Storing Information in the Letter $z$ .....	114
<i>Students Talk Stats: When the 68-95-99.7 Rule Does Not Apply</i> .....	117
"Unstandardizing" $z$ -Scores: Back to Original Units .....	118
The Normal Table: A Precursor to Software .....	119
Summary 125 / Exercises 127	

### 5 Displaying and Summarizing Relationships 133

5.1 Relationships between One Categorical and One Quantitative Variable ...	133
Different Approaches for Different Study Designs .....	133
Displays .....	134
Summaries .....	134
Notation .....	134
Data from a Two-Sample Design .....	134
Data from a Several-Sample Design .....	137
Data from a Paired Design .....	138
<i>Students Talk Stats: Displaying and Summarizing Paired Data</i> .....	139
Generalizing from Samples to Populations: The Role of Spreads .....	141
The Role of Sample Size: When Differences Have More Impact .....	143
5.2 Relationships between Two Categorical Variables .....	150
Summaries and Displays: Two-Way Tables, Conditional Percentages, and Bar Graphs .....	151
The Role of Sample Size: Larger Samples Let Us Rule Out Chance .....	156

Comparing Observed and Expected Counts .....	156
Confounding Variables and Simpson's Paradox: Is the Relationship Really There? .....	157
<b>5.3 Relationships between Two Quantitative Variables .....</b>	<b>165</b>
Displays and Summaries: Scatterplots, Form, Direction, and Strength .....	166
Correlation: One Number for Direction and Strength .....	170
When the Correlation Is 0, +1, or -1 .....	171
Correlation as a Measure of Direction and Strength .....	173
A Closer Look at Correlation .....	174
Correlation Is Unaffected by the Roles of Explanatory and Response Variables .....	175
Correlation Is Unaffected by Units of Measurement .....	176
Least Squares Regression Line: What We See in a Linear Plot .....	177
A Closer Look at Least Squares Regression .....	182
Residuals: Prediction Errors in a Regression .....	182
Spread $s$ about the Line versus Spread $s_y$ about the Mean Response ....	183
The Effect of Explanatory and Response Roles on the Regression Line ..	184
Influential Observations and Outliers .....	185
<i>Students Talk Stats:</i> How Outliers and Influential Observations Affect a Relationship .....	186
Sample versus Population: Thinking Beyond the Data at Hand .....	187
The Role of Sample Size: Larger Samples Get Us Closer to the Truth .....	188
Time Series: When Time Explains a Response .....	189
Additional Variables: Confounding Variables, Multiple Regression .....	191
<i>Students Talk Stats:</i> Confounding in a Relationship between Two Quantitative Variables .....	191
<b>Summary 204 / Exercises 206</b>	

## PART III Probability 224

### 6 Finding Probabilities 226

<b>6.1 The Meaning of "Probability" and Basic Rules .....</b>	<b>226</b>
Permissible Probabilities .....	229
Probabilities Summing to One .....	229
Probability of "Not" Happening .....	231
Probability of One "Or" the Other for Non-overlapping Events .....	231
Probability of One "And" the Other for Two Independent Events .....	233
<b>6.2 More General Probability Rules and Conditional Probability .....</b>	<b>238</b>
Probability of One "Or" the Other for Any Two Events .....	239
Probability of Both One "And" the Other Event Occurring .....	241
<i>Students Talk Stats:</i> Probability as a Weighted Average of Conditional Probabilities .....	245
Conditional Probability in Terms of Ordinary Probabilities .....	246
Checking for Independence .....	247
Counts Expected If Two Variables Are Independent .....	250

**Summary 256 / Exercises 257**

### 7 Random Variables 267

<b>7.1 Discrete Random Variables .....</b>	<b>268</b>
Probability Distributions of Discrete Random Variables .....	269
The Mean of a Random Variable .....	276
The Standard Deviation of a Random Variable .....	278
Rules for the Mean and Standard Deviation of a Random Variable .....	280



7.2	<b>Binomial Random Variables</b> .....	291
	What Makes a Random Variable “Binomial”? .....	291
	The Mean and Standard Deviation of Sample Proportions .....	295
	<i>Students Talk Stats:</i> Calculating and Interpreting the Mean and Standard Deviation of Count or Proportion .....	297
	The Shape of the Distribution of Counts or Proportions: The Central Limit Theorem .....	299
7.3	<b>Continuous Random Variables and the Normal Distribution</b> .....	311
	Discrete versus Continuous Distributions .....	312
	When a Random Variable Is Normal .....	315
	The 68-95-99.7 Rule for Normal Random Variables .....	316
	Standardizing and Unstandardizing: From Original Values to $z$ or Vice Versa .....	319
	Estimating $z$ Probabilities with a Sketch of the 68-95-99.7 Rule .....	319
	Nonstandard Normal Probabilities .....	323
	Tails of the Normal Curve: The 90-95-98-99 Rule .....	326
	<i>Students Talk Stats:</i> Means, Standard Deviations, and Below-Average Heights .....	329
	<b>Summary</b> 335 / <b>Exercises</b> 337	

## 8 Sampling Distributions 344

	Categorical Variables: The Behavior of Sample Proportions .....	344
	Quantitative Variables: The Behavior of Sample Means .....	345
8.1	<b>The Behavior of Sample Proportion in Repeated Random Samples</b> .....	346
	Thinking about Proportions from Samples or Populations .....	346
	Center, Spread, and Shape of the Distribution of Sample Proportion .....	348
8.2	<b>The Behavior of Sample Mean in Repeated Random Samples</b> .....	356
	Thinking about Means from Samples or Populations .....	356
	The Mean of the Distribution of Sample Mean .....	358
	The Standard Deviation of the Distribution of Sample Mean .....	358
	The Shape of the Distribution of Sample Mean: The Central Limit Theorem .....	360
	Center, Spread, and Shape of the Distribution of Sample Mean .....	360
	Normal Probabilities for Sample Means .....	362
	<i>Students Talk Stats:</i> When Normal Approximations Are Appropriate .....	365
	<b>Summary</b> 371 / <b>Exercises</b> 372	

## PART IV Statistical Inference 386

### 9 Inference for a Single Categorical Variable 388

9.1	<b>Point Estimate and Confidence Interval: A Best Guess and a Range of Plausible Values for Population Proportion</b> .....	390
	Probability versus Confidence: Talking about Random Variables or Parameters .....	392
	95% Confidence Intervals: Building around Our Point Estimate .....	394
	The Role of Sample Size: Closing In on the Truth .....	398
	Confidence at Other Levels .....	400
	Deciding If a Particular Value Is Plausible: An Informal Approach .....	403
	The Meaning of a Confidence Interval: What Exactly Have We Found? ...	404
	<i>Students Talk Stats:</i> Interpreting a Confidence Interval .....	405
9.2	<b>Hypothesis Test: Is a Proposed Population Proportion Plausible?</b> .....	413
	Three Forms of Alternative Hypothesis: Different Ways to Disagree .....	416
	One-Sided or Two-Sided Alternative Hypothesis .....	425

How Small Is a “Small” $P$ -Value? .....	429
The Role of Sample Size in Conclusions for Hypothesis Tests .....	430
When to Reject the Null Hypothesis: Three Contributing Factors .....	431
<b>Students Talk Stats:</b> Interpreting a $P$ -Value .....	432
Type I or II Error: What Kind of Mistakes Can We Make? .....	433
<b>Students Talk Stats:</b> What Type of Error Was Made? .....	435
Relating Results of Test with Confidence Interval: Two Sides of the Same Coin .....	435
The Language of Hypothesis Tests: What Exactly Do We Conclude? .....	436
<b>Students Talk Stats:</b> The Correct Interpretation of a Small $P$ -Value .....	437
<b>Students Talk Stats:</b> The Correct Interpretation When a $P$ -Value Is Not Small .....	437
The “Critical Value” Approach: Focusing on the Standard Score .....	438

Summary 451 / Exercises 454

## 10 Inference for a Single Quantitative Variable 461

<b>10.1 Inference for a Mean When Population Standard Deviation Is Known or Sample Size Is Large</b> .....	462
A Confidence Interval for the Population Mean Based on $z$ .....	464
95% Confidence Intervals with $z$ .....	465
<b>Students Talk Stats:</b> Confidence Interval for a Mean: Width, Margin of Error, Standard Deviation, and Standard Error .....	469
Role of Sample Size: Larger Samples, Narrower Intervals .....	471
Intervals at Other Levels of Confidence with $z$ .....	472
Interpreting a Confidence Interval for the Mean .....	473
<b>Students Talk Stats:</b> Correctly Interpreting a Confidence Interval for the Mean .....	473
A $z$ Hypothesis Test about the Population Mean .....	474
<b>10.2 Inference for a Mean When the Population Standard Deviation Is Unknown and the Sample Size Is Small</b> .....	480
A $t$ Confidence Interval for the Population Mean .....	482
95% Confidence Intervals with $t$ .....	482
Intervals at Other Levels of Confidence with $t$ .....	484
A $t$ Hypothesis Test about the Population Mean .....	486
<b>Students Talk Stats:</b> Practical Application of a $t$ Test .....	488
<b>10.3 A Closer Look at Inference for Means</b> .....	491
A One-Sided or Two-Sided Alternative Hypothesis about a Mean .....	491
The Role of Sample Size and Spread: What Leads to Small $P$ -Values? .....	493
Type I and II Errors: Mistakes in Conclusions about Means .....	494
Relating Tests and Confidence Intervals for Means .....	495
Correct Language in Hypothesis Test Conclusions about a Mean .....	496
Robustness of Procedures .....	498

Summary 503 / Exercises 505

## 11 Inference for Relationships between Categorical and Quantitative Variables 520

<b>11.1 Inference for a Paired Design with <math>t</math></b> .....	522
Hypothesis Test in a Paired Design .....	522
Confidence Interval in a Paired Design .....	524
<b>11.2 Inference for a Two-Sample Design with <math>t</math></b> .....	528
The Two-Sample $t$ Distribution and Test Statistic .....	528
Hypothesis Test in a Two-Sample Design .....	530
Confidence Interval in a Two-Sample Design .....	534
The Pooled Two-Sample $t$ Procedure .....	536
<b>Students Talk Stats:</b> Ordinary versus Pooled Two-Sample $t$ .....	537

11.3 Inference for a Several-Sample Design with $F$ : Analysis of Variance . . . . .	543
The $F$ Statistic . . . . .	545
The $F$ Distribution . . . . .	550
Solving Several-Sample Problems . . . . .	552
The ANOVA Table: Organizing What We Know about $F$ . . . . .	555
The ANOVA Alternative Hypothesis . . . . .	557
Assumptions of ANOVA . . . . .	558
Summary . . . . .	566
<i>Students Talk Stats</i> : Reviewing Relationships between Categorical Explanatory and Quantitative Response Variables . . . . .	566
Exercises . . . . .	571
<b>12 Inference for Relationships between Two Categorical Variables</b> . . . . .	591
12.1 Comparing Proportions with a $z$ Test . . . . .	592
12.2 Comparing Counts with a Chi-Square Test . . . . .	598
Relating Chi-Square to $z$ . . . . .	598
The Table of Expected Counts . . . . .	599
Comparing Observed to Expected Counts . . . . .	600
The Chi-Square Distribution . . . . .	602
The Chi-Square Test . . . . .	604
Sample Size and Chi-Square Assumptions . . . . .	604
Summary . . . . .	613 / Exercises 614
<b>13 Inference for Relationships between Two Quantitative Variables</b> . . . . .	628
13.1 Inference for Regression: Focus on the Slope of the Regression Line . . . . .	629
Setting the Stage: Summarizing a Relationship for Sampled Points . . . . .	630
Distinguishing between Sample and Population Relationships . . . . .	631
A Model for the Relationship between Two Quantitative Variables in a Population . . . . .	634
The Distribution of Sample Slope $b_1$ . . . . .	636
The Distribution of Standardized Sample Slope $t$ . . . . .	637
Hypothesis Test about the Population Slope with $t$ : A Clue about the Relationship . . . . .	638
<i>Students Talk Stats</i> : No Evidence of a Relationship . . . . .	643
Confidence Interval for the Slope of the Population Regression Line . . . . .	644
13.2 Interval Estimates for an Individual or Mean Response . . . . .	651
Summary . . . . .	662 / Exercises 664
<b>14 How Statistics Problems Fit into the Big Picture</b> . . . . .	677
14.1 The Big Picture in Problem Solving . . . . .	677
<i>Students Talk Stats</i> : Choosing the Appropriate Statistical Tools: Question 1 . . . . .	678
<i>Students Talk Stats</i> : Choosing the Appropriate Statistical Tools: Question 2 . . . . .	679
<i>Students Talk Stats</i> : Choosing the Appropriate Statistical Tools: Question 3 . . . . .	680
Exercises . . . . .	683
<b>15 Non-parametric Methods</b> ( <i>available online</i> ) . . . . .	
15.1 The Sign Test as an Alternative to the Paired $t$ Test . . . . .	
15.2 The Rank-Sum Test as an Alternative to the Two-Sample $t$ Test . . . . .	
Wilcoxon rank-sum test . . . . .	
15.3 Summary of Non-parametrics . . . . .	
Exercises . . . . .	

## 16 Two-Way ANOVA *(available online)*

Exercises

## 17 Multiple Regression *(available online)*

Exercises

**Solutions to Selected Exercises** 689

**Endnotes** 987

**Index** 1007