LEARNING MADE EASY

2nd Edition

Statistics

WORKBOOK

WITH ONLINE PRACTICE

for dummies®

A Wiley Brand

Find FREE quizzes for every chapter online

Slay statistics with hundreds of practice problems

Review key concepts

Deborah J. Rumsey, PhD

Professor of Statistics,
The Ohio State University

This book comes with access to more content online.

Test your knowledge with a quiz
for every chapter!

Register your book or ebook at
www.dummies.com/go/getaccess.

Select your product, and then follow the prompts
to validate your purchase.

You'll receive an email with your PIN and instructions.

Statistics Workbook

2nd Edition with Online Practice

by Deborah J. Rumsey, PhD

for dummies®
A Wiley Brand

Statistics Workbook For Dummies®, 2nd Edition with Online Practice

Published by: **John Wiley & Sons, Inc.,** 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2019 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

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, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permissions.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc., and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit https://hub.wiley.com/community/support/dummies.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2019931550

ISBN 978-1-119-54751-8 (pbk); ISBN 978-1-119-54767-9 (ebk); ISBN 978-1-119-54768-6

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

Introduction 1

Part 1: Getting Off to a Statistically Significant Start 5

CHAPTER 1: Summarizing Categorical Data: Counts and Percents 7
CHAPTER 2: Summarizing Quantitative Data: Means, Medians, and More 17
CHAPTER 3: Organizing Categorical Data: Charts and Graphs 27
CHAPTER 4: Organizing Quantitative Data: Charts and Graphs 43

Part 2: Probability, Distributions, and the Central Limit Theorem (Are You Having Fun Yet?) 73

CHAPTER 5: Understanding Probability Basics 75
CHAPTER 6: Measures of Relative Standing and the Normal Distribution 83
CHAPTER 7: The Binomial Distribution 105
CHAPTER 8: The *t*-Distribution 117
CHAPTER 9: Demystifying Sampling Distributions and the Central Limit Theorem 123

Part 3: Guesstimating and Hypothesizing with Confidence 137

CHAPTER 10: Making Sense of Margin of Error 139
CHAPTER 11: Calculating Confidence Intervals 151
CHAPTER 12: Deciphering Your Confidence Interval 169
CHAPTER 13: Testing Hypotheses 177
CHAPTER 14: Taking the Guesswork Out of p-Values and Type I and II Errors 197

Part 4: Statistical Studies and the Hunt for a Meaningful Relationship 211

CHAPTER 15: Examining Polls and Surveys 213
CHAPTER 16: Evaluating Experiments 223
CHAPTER 17: Looking for Links in Categorical Data: Two-Way Tables 233
CHAPTER 18: Searching for Links in Quantitative Data: Correlation and Regression 259

Part 5: The Part of Tens 277

CHAPTER 19: Math Review: Ten Steps to a Better Grade 279
CHAPTER 20: Top Ten Statistical Formulas 289
CHAPTER 21: Ten Ways to Spot Common Statistical Mistakes 301

Appendix: Tables for Reference 309

Index 319

Table of Contents

INTRODUCTION 1
About This Book 1
Foolish Assumptions 2
Icons Used in This Book 3
Beyond the Book 3
Where to Go from Here 4

PART 1: GETTING OFF TO A STATISTICALLY SIGNIFICANT START 5

CHAPTER 1: Summarizing Categorical Data: Counts and Percents 7
Counting On the Frequency 7
Relating with Percentages 9
Interpreting Counts and Percents with Caution 11
Answers to Problems in Summarizing Categorical Data 13

CHAPTER 2: Summarizing Quantitative Data: Means, Medians, and More 17
Finding and Interpreting Measures of Center 18
Finding and Interpreting Measures of Spread 20
Using Percentiles and the Interquartile Range 22
Answers to Problems in Summarizing Quantitative Data 24

CHAPTER 3: Organizing Categorical Data: Charts and Graphs 27
Making, Interpreting, and Evaluating Pie Charts 27
Making, Interpreting, and Evaluating Bar Graphs 32
Answers to Problems in Organizing Categorical Data 37

CHAPTER 4: Organizing Quantitative Data: Charts and Graphs 43
Creating a Histogram 44
Making Sense of Histograms 47
Straightening Out Skewed Data 51
Spotting a Misleading Histogram 53
Making Box Plots 55
Interpreting Box Plots 56
Looking at Line Graphs 58
Understanding the Empirical Rule 60
Answers to Problems in Organizing Quantitative Data 63

PART 2: PROBABILITY, DISTRIBUTIONS, AND THE CENTRAL LIMIT THEOREM (ARE YOU HAVING FUN YET?) 73

CHAPTER 5: Understanding Probability Basics 75
Grasping the Rules of Probability 75
Avoiding Probability Misconceptions 78
Making Predictions Using Probability 79
Answers to Problems in Probability 81

CHAPTER 6: **Measures of Relative Standing and the Normal Distribution** 83
Mastering the Normal Distribution 83
Finding and Interpreting Standard (Z) Scores 86
Knowing Where You Stand with Percentiles 88
Finding Probabilities for a Normal Distribution 90
Finding the Percentile (Backwards Normal) 92
Answers to Problems in Normal Distribution 95

CHAPTER 7: **The Binomial Distribution** 105
Characterizing the Binomial Distribution 105
Finding Probabilities Using the Binomial Formula for small n 107
Finding Probabilities Using the Binomial Table for Medium-Sized n 109
Calculating the Mean and Variance of the Binomial 110
Estimating Probabilities in Large Cases — the Normal Approximation 112
Answers to Problems in the Binomial Distribution 114

CHAPTER 8: **The *t*-Distribution** 117
Getting to Know the *t*-Distribution 117
Working with the *t*-Table and Degrees of Freedom 120
Answers to Problems in the *t*-Distribution 122

CHAPTER 9: **Demystifying Sampling Distributions and the Central Limit Theorem** 123
Exactly What Is a Sampling Distribution? 124
Clearing Up the Central Limit Theorem (Once and for All) 126
Finding Probabilities with the Central Limit Theorem 129
When Your Sample's Too Small: Employing the *t*-Distribution 131
Answers to Problems in Sampling Distributions and the Central Limit Theorem 133

PART 3: GUESSTIMATING AND HYPOTHESIZING WITH CONFIDENCE 137

CHAPTER 10: **Making Sense of Margin of Error** 139
Reviewing Margin of Error 139
Calculating the Margin of Error for Means and Proportions 142
Increasing and Decreasing Margin of Error 144
Interpreting Margin of Error Correctly 146
Answers to Problems in Making Sense of Margin of Error 148

CHAPTER 11: **Calculating Confidence Intervals** 151
Walking through a Confidence Interval 151
Deriving a Confidence Interval for a Population Mean 154
Figuring a Confidence Interval for a Population Proportion 156
Calculating a Confidence Interval for the Difference of Two Means 158
Computing a Confidence Interval for the Difference of Two Proportions 160
Answers to Problems in Calculating Confidence Intervals 163

CHAPTER 12: **Deciphering Your Confidence Interval** 169
Interpreting Confidence Intervals the Right Way 169
Evaluating Confidence Interval Results: What the Formulas Don't Tell You 173
Answers to Problems in Confidence Intervals 175

CHAPTER 13: **Testing Hypotheses** 177
Walking Through a Hypothesis Test 177
Testing a Hypothesis about a Population Mean 181
Testing a Hypothesis about a Population Proportion 183
Testing for a Difference between Two Population Means 185
Testing for a Mean Difference (Paired *t*-Test) 188
Testing a Hypothesis about Two Population Proportions 190
Answers to Problems in Testing Hypotheses 192

CHAPTER 14: **Taking the Guesswork Out of p-Values and Type I and II Errors** 197
Understanding What p-Values Measure 198
Test (Statistic) Time: Figuring Out p-Values 199
The Value Breakdown: Interpreting p-Values Properly 201
Deciphering Type I Errors 204
Deciphering and Distinguishing Type II Errors 205
Answers to Problems in p-Values and Type I and II Errors 208

PART 4: STATISTICAL STUDIES AND THE HUNT FOR A MEANINGFUL RELATIONSHIP 211

CHAPTER 15: **Examining Polls and Surveys** 213
Planning and Designing a Survey 214
Selecting a Random Sample 215
Carrying Out a Survey Properly 217
Interpreting and Evaluating Survey Results 218
Answers to Problems in Polls and Surveys 220

CHAPTER 16: **Evaluating Experiments** 223
Distinguishing Experiments from Observational Studies 223
Designing a Good Experiment 225
Looking for Cause and Effect: Interpreting Experiment Results 228
Answers to Problems in Evaluating Experiments 230

CHAPTER 17: **Looking for Links in Categorical Data: Two-Way Tables** 233
Understanding Two-Way Tables Inside and Out 234
Working with Intersection, Unions, and the Addition Rule 237
Figuring Marginal Probabilities 240
Nailing Down Conditional Probabilities and the Multiplication Rule 242
Inspecting the Independence of Categorical Variables 246
Answers to Problems in Two-Way Tables 250

CHAPTER 18: **Searching for Links in Quantitative Data: Correlation and Regression** 259
Relating X and Y with a Scatterplot 259
Toeing the Line of Correlation 262
Picking Out the Best Fitting Regression Line 265
Interpreting the Regression Line and Making Predictions 267
Checking the Fit of the Regression Line 269
Answers to Problems in Correlation and Regression 272

PART 5: THE PART OF TENS 277

CHAPTER 19: **Math Review: Ten Steps to a Better Grade** 279
Know Your Math Symbols 279
Uproot Roots and Powers 280
Treat Fractions with Extra Care 280
Obey the Order of Operations 281
Avoid Rounding Errors 282
Get Comfortable with Formulas 283
Stay Calm When Formulas Get Tough 283
Feel Fine about Functions 285
Know When Your Answer Is Wrong 286
Show Your Work 287

CHAPTER 20: **Top Ten Statistical Formulas** 289
Mean (or Average) 289
Median 290
Sample Standard Deviation 291
Correlation 292
Margin of Error for the Sample Mean 293
Sample Size Needed for Estimating μ 294
Test Statistic for the Mean 295
Margin of Error for the Sample Proportion 296
Sample Size Needed for Estimating p 297
Test Statistic for the Proportion 298

CHAPTER 21: **Ten Ways to Spot Common Statistical Mistakes** 301
Scrutinizing Graphs 301
Searching for and Specifying Bias 302
Marking the Margin of Error 303
Scanning for Sample Size 303
Studying Sample Selection (Gotta Be Random) 304
Checking for Confounding Variables 305
Considering Correlation 305
Doing the Math 306
Detecting Selective Reporting 306
Avoiding the Anecdote 307

APPENDIX: TABLES FOR REFERENCE 309

INDEX 319

Introduction

Perhaps you're taking a statistics class, or you're about to take one. You may understand some of the basic ideas, but you have questions and want a place to go for a little extra help to give you an edge. And you also want a heads-up as to what instructors really think about when they write their exams. Well, look no further; help has arrived in the form of *Statistics Workbook For Dummies*, 2nd Edition.

This workbook helps you become more comfortable with and confident about statistics. Through plenty of practical problems that take you from step one all the way to your final exam, you review the concepts you know, identify areas where you need to focus more work, and address the little things that can make the difference between a B and an A.

As a statistics professor who has taught tens of thousands of students over the years, I have noticed that certain problems keep cropping up and causing my pen to take points off exams over and over again. And believe me: I want nothing more than to put my red pen away. So I give you all my secrets about what professors really want you to know, the kinds of questions they ask, and the types of answers they love and hate to see (so you can avoid the latter). And I focus only on the topics that you absolutely need to know, with minimal background information.

About This Book

The major objectives of this workbook are for you to understand, calculate, and interpret the most common statistical formulas and techniques; get a handle on basic probability; gain confidence with difficult statistical topics such as the central limit theorem and *p*-values; know which statistical technique to use in different situations (for example, when to employ what kind of confidence interval); and evaluate and pinpoint problems with studies, polls, and experiments.

Although I wrote this workbook to serve as a companion to *Statistics for Dummies*, 2nd Edition (also published by Wiley and written by yours truly), this workbook works quite well with any introductory statistics textbook.

You may be asking how this workbook is different from other workbooks on the shelf. Well, here are a few ways, listed in order of importance:

- **Plenty of excellent practice problems** to lead you down the path of examination success, chosen by me, a card-carrying member of the "million statistics exam question writing and grading" club. I provide all answers at the end of each chapter.

- **Workspace** for you to work through the problems directly in the section you're working on, so you can easily refer to your notes later when you need them.
- **Not only answers, but also clear, complete explanations to go with them.** Explanations help you know exactly how to approach a problem, what information you need to solve it, and common problems you need to avoid.
- **A view inside a professor's mind** to help you determine the most popular questions, the answers we look for, and the answers that make us pull our hair out.
- **Tips, strategies, and warnings** based on my vast experience with students of all backgrounds and learning styles (and my grading experience).
- **An example accompanying each section,** directly followed by the solution. Use the example as a reference when you work the other problems.
- **A focus on problem-solving skills** to help you develop a problem-solving strategy when you take exams. I don't show you how I would do the problems; I help you see how you can do the problems. And believe me, there's a big difference!
- **The nonlinear approach** allows you to skip around in the workbook and still have easy access to and understanding of any given topic.
- **Understandable language** to help you process, remember, and put into practice statistical definitions, techniques, and processes.
- **Clear and concise step-by-step procedures** that intuitively explain how to work through statistics problems and remember the process.

I also used a few conventions while writing this book that you should be aware of:

- The most important convention that you need to be aware of deals with my dual use of the word "statistics." In some situations, I refer to statistics as a subject of study or as a field of research. For example, "Statistics is really quite an interesting subject!" (Note I said statistics "is" in this case.) In other situations, I refer to statistics as the plural of statistic, in a numerical sense. For example, "The most common statistics are the mean and the standard deviation." (Notice my use of the word "are" in this case.)
- I also use data in a plural form ("the data are" rather than "the data is"). The battle rages on between statisticians over which way is right, but I go with the plural form.
- I use Ho to represent the null hypothesis in a hypothesis test. Although this is a commonly used notation, others might use the notation H_0 to mean the same thing.
- I use * to indicate a multiplication sign.

Foolish Assumptions

This book is for you if you have some exposure to statistics already and want more opportunities to enjoy success through additional practice of the skills and techniques. Or perhaps you're taking a statistics class and could use some extra support (and insider information). Or maybe you just really want to understand *p*-values because they keep you awake at night (been there, done that).

Note: If you're totally new to the subject of statistics, I suggest that you first read *Statistics for Dummies*, 2nd Edition, (Wiley), because I cover the various concepts of statistics in much more detail in that book (but any introductory text will suffice). After you feel comfortable and confident with the material, you can try the problems in this workbook. Or, as an alternative, you can use this workbook to practice along with what you read in *Statistics For Dummies*, 2nd Edition.

Icons Used in This Book

Icons in this workbook draw your attention to certain features that occur on a regular basis. Think of them as road signs that you encounter on a trip. Here are the road signs you encounter on your journey through this workbook.

EXAMPLE

Each section of this workbook begins with a brief overview of the topic. After the intro, you see an example problem with a fully worked solution for use as a reference as you work the practice problems. You can quickly locate the example problems by looking for this icon.

REMEMBER

I use this icon for particular ideas that I hope you'll remember long after you read this workbook.

TIP

This icon points out helpful hints, ideas, or shortcuts that save you time or give you alternative ways to think about a particular concept. I also use this icon to "get down to the nitty-gritty" discussing the types of questions your instructor may ask you and why, revealing what instructors really look for in your answers, and giving you a heads-up on the types of errors that really make them nuts (so you can avoid them at all costs).

WARNING

This icon refers to specific ways that you may get tripped up while working a certain kind of problem and how to avoid those problems. Commit these items to memory while it still doesn't cost you any points (in other words, before the exam takes place).

Beyond the Book

Be sure to check out the free Cheat Sheet for a handy guide that covers tips and tricks for answering statistics questions. To get this Cheat Sheet, simply go to `www.dummies.com` and enter "Statistics Workbook For Dummies" in the Search box.

You also have the opportunity to complete online quizzes for Chapters 1 through 18 that test your knowledge of the concepts in each chapter. To gain access to the online practice, all you have to do is register by following these simple steps:

1. Register your book or ebook at Dummies.com to get your PIN. Go to `www.dummies.com/go/getaccess`.
2. Select your product from the dropdown list on that page.
3. Follow the prompts to validate your product, and then check your email for a confirmation message that includes your PIN and instructions for logging in.

If you do not receive this email within two hours, please check your spam folder before contacting us through our Technical Support website at `http://support.wiley.com` or by phone at 877-762-2974.

Now you're ready to go! You can come back to the practice material as often as you want — simply log on with the username and password you created during your initial login. No need to enter the access code a second time.

Your registration is good for one year from the day you activate your PIN.

Where to Go from Here

I wrote this workbook in a nonlinear way, so you can start anywhere and still understand what's happening. However, I can make some recommendations to readers who are interested in knowing where to start:

- If you want to get right into the number-crunching aspects of statistics (finding the mean, median, standard deviation, and so on), I suggest starting with Part 1.
- If you want to break down the normal distribution or the central limit theorem, go to Part 2.
- If you're most worried about confidence intervals and hypothesis tests, jump to Part 3.
- If you want to develop your skills evaluating and making sense of the results of medical studies, polls, surveys, and experiments, start with Part 4.
- If you want to nail down data collected on two variables (correlation and the like), head directly to Part 4.
- If you want tips on math, common statistical formulas, or ways to spot statistical mistakes, head to Part 5.

1 Getting Off to a Statistically Significant Start

IN THIS PART . . .

Get down to the basics of number crunching.

Make and interpret charts and graphs.

Crank out and understand descriptive statistics.

Develop important skills for critiquing others' statistics.

IN THIS CHAPTER

- Making tables to summarize categorical data
- Highlighting the difference between frequencies and relative frequencies
- Interpreting and evaluating tables

Chapter 1

Summarizing Categorical Data: Counts and Percents

Categorical data is data in which individuals are placed into groups or categories — for example gender, region, or type of movie. Summarizing categorical data involves boiling down all the information into just a few numbers that tell its basic story. Because categorical data involves pieces of data that belong in categories, you have to look at how many individuals fall into each group and summarize the numbers appropriately. In this chapter, you practice making, interpreting, and evaluating frequency and relative frequency tables for categorical data.

Counting On the Frequency

One way to summarize categorical data is to simply count, or tally up, the number of individuals that fall into each category. The number of individuals in any given category is called the frequency (or count) for that category. If you list all the possible categories along with the frequency for each, you create a frequency table. The total of all the frequencies should equal the size of the sample (because you place each individual in one category).

See the following for an example of summarizing data by using a frequency table.

EXAMPLE

Q. Suppose that you take a sample of 10 people and ask them all whether they own a cellphone. Each person falls into one of two categories: yes or no. The data are shown in the following table.

Person #	Cellphone	Person #	Cellphone
1	Y	6	Y
2	N	7	Y
3	Y	8	Y
4	N	9	N
5	Y	10	Y

a. Summarize this data in a frequency table.

b. What's an advantage of summarizing categorical data?

A. Data summaries boil down the data quickly and clearly.

a. The frequency table for this data is shown in the following table.

b. A data summary allows you to see patterns in the data, which aren't clear if you look only at the original data.

Own a Cellphone?	Frequency
Y	7
N	3
Total	10

1 You survey 20 shoppers to see what type of soft drink they like best, Brand A or Brand B. The results are: A, A, B, B, B, B, B, B, A, A, A, B, A, A, A, A, B, B, A, A. Which brand do the shoppers prefer? Make a frequency table and explain your answer.

2 A local city government asks voters to vote on a tax levy for the local school district. A total of 18,726 citizens vote on the issue. The yes count comes in at 10,479, and the rest of the voters said no.

a. Show the results in a frequency table.

b. Why is it important to include the total number at the bottom of a frequency table?

3 A zoo asks 1,000 people whether they've been to the zoo in the last year. The surveyors count that 592 say yes, 198 say no, and 210 don't respond.

a. Show the results in a frequency table.

b. Explain why you need to include the people who don't respond.

4 Suppose that instead of showing the number in each group, you show just the percentage (called a *relative frequency*). What's one advantage a relative frequency table has over a frequency table?

Relating with Percentages

Another way to summarize categorical data is to show the percentage of individuals who fall into each category, thereby creating a relative frequency. The *relative frequency* of a given category is the frequency (number of individuals in that category) divided by the total sample size, multiplied by 100 to get the percentage. For example, if you survey 50 people and 10 are in favor of a certain issue, the relative frequency of the "in-favor" category is $10 \div 50 = 0.20$ times 100, which gives you 20 percent. If you list all the possible categories along with their relative frequencies, you create a *relative frequency table.* The total of all the relative frequencies should equal 100 percent (subject to possible round-off error).

See the following for an example of summarizing data by using a relative frequency table.

EXAMPLE

Q. Using the cellphone data from the following table, make a relative frequency table and interpret the results.

Person #	Cellphone	Person #	Cellphone
1	Y	6	Y
2	N	7	Y
3	Y	8	Y
4	N	9	N
5	Y	10	Y

A. The following table shows a relative frequency table for the cellphone data. Seventy percent of the people sampled reported owning cellphones, and 30 percent admitted to being technologically behind the times.

Own a Cellphone?	Relative Frequency
Y	70%
N	30%

You get the 70 percent by taking $7 \div 10 * 100$, and you calculate the 30 percent by taking $3 \div 10 * 100$.

5 You survey 20 shoppers to see what type of soft drink they like best, Brand A or Brand B. The results are: A, A, B, B, B, B, B, B, A, A, A, B, A, A, A, A, B, B, A, A. Which brand do the shoppers prefer?

a. Use a relative frequency table to determine the preferred brand.

b. In general, if you had to choose, which is easier to interpret: frequencies or relative frequencies? Explain.

6 A local city government asked voters in the last election to vote on a tax levy for the local school district. A record 18,726 voted on the issue. The yes count came in at 10,479, and the rest of the voters checked the no box. Show the results in a relative frequency table.

7 A zoo surveys 1,000 people to find out whether they've been to the zoo in the last year. The surveyors count that 592 say yes, 198 say no, and 210 don't respond. Make a relative frequency table and use it to find the *response rate* (percentage of people who respond to the survey).

8 Name one disadvantage that comes with creating a relative frequency table compared to using a frequency table.

Interpreting Counts and Percents with Caution

Not all summaries of categorical data are fair and accurate. Knowing what to look for can help you keep your eyes open for misleading and incomplete information.

Instructors often ask you to "interpret the results." In this case, your instructor wants you to use the statistics available to talk about how they relate to the given situation. In other words, what do the results mean to the person who collects the data?

REMEMBER

With relative frequency tables, don't forget to check whether all categories sum to 1 or 100 percent (subject to round-off error), and remember to look for some indicator as to total sample size.

See the following for an example of critiquing a data summary.

EXAMPLE

Q. You watch a commercial where the manufacturer of a new cold medicine ("Nocold") compares it to the leading brand. The results are shown in the following table.

How Nocold Compares	Percentage
Much better	47%
At least as good	18%

a. What kind of table is this?

b. Interpret the results. (Did the new cold medicine beat out the leading brand?)

c. What important details are missing from this table?

A. Much like the cold medicines I always take, the table about "Nocold" does "Nogood."

a. This table is an incomplete relative frequency table. The remaining category is "not as good" for the Nocold brand, and the advertiser doesn't show it. But you can do the math and see that $100\% - (47\% + 18\%) = 35\%$ of the people say that the leading brand is better.

b. If you put the two groups together, 65% of the patients say that Nocold is at least as good as the leading brand, and almost half of the patients say Nocold is much better.

c. What's missing? The remaining percentage (to keep all possible results in perspective). But more importantly, the total sample size is missing. You don't know whether the surveyors sampled 10 people, 100 people, or 1,000 people. This means that the precision of the results is unknown. (Precision means how consistent the results will be from sample to sample; it's related to sample size, as you see in Chapter 10.)

9 Suppose that you ask 1,000 people to identify from a list of five vacation spots which ones they've already visited. The frequencies you receive are Disney World: 216; New Orleans: 312; Las Vegas: 418; New York City: 359; and Washington, D.C.: 188.

a. Explain why creating a traditional relative frequency table doesn't make sense here.

b. How can you summarize this data with percents in a way that makes sense?

10 If you have only a frequency table, can you find the corresponding relative frequency table? Conversely, if you have only a relative frequency table, can you find the corresponding frequency table? Explain.

Answers to Problems in Summarizing Categorical Data

1 Eleven shoppers prefer Brand A, and nine shoppers prefer Brand B. The frequency table is shown in the following table. Brand A got more votes, but the results are pretty close.

Brand Preferred	Frequency
A	11
B	9
Total	20

2 Frequencies are fine for summarizing data as long as you keep the total number in perspective.

a. The results are shown in the following table. Because the total is 18,726, and the yes count is 10,479, the no count is the difference between the two, which is $18{,}726 - 10{,}479 = 8{,}247$.

b. The total is important because it helps keep the frequencies in perspective when you compare them to each other.

Vote	Frequency
Y	10,479
N	8,247
Total	18,726

3 This problem shows the importance of reporting not only the results of participants who respond but also what percentage of the total actually respond.

a. The results are shown in the following table.

b. If you don't show the nonrespondents, the total doesn't add up to 1,000 (the number surveyed). An alternative way to show the data is to base it on only the respondents, but the results would be biased. You can't definitively say that the nonrespondents would respond the same way as the respondents.

Gone to the Zoo in the Last Year?	Frequency
Y	592
N	198
Nonrespondents	210
Total	1,000

4 Showing the percents rather than counts means making a relative frequency table rather than a frequency table. One advantage of a relative frequency table is that everything sums to 100 percent, making it easier to interpret the results, especially if you have a large number of categories.

5 Relative frequencies do just what they say: They help you relate the results to each other (by finding percentages).

a. Eleven shoppers out of the 20 prefer Brand A, and nine shoppers out of the 20 prefer Brand B. The relative frequency table is shown in the following table. Brand A got more votes, but the results are pretty close, with 55 percent of the shoppers preferring Brand A, and 45 percent preferring Brand B.

Brand Preferred	Relative Frequency
A	55%
B	45%

b. You often have an easier time interpreting percents, because when you need to interpret counts, you have to put them in perspective in terms of "out of how many?"

6 The results are shown in the following table. The yes percentage is $10,479 \div 18,726 = 55.96\%$. Because the total is 100%, the no percentage is $100\% - 55.96\% = 44.04\%$.

Vote	Relative Frequency
Y	55.96%
N	44.04%

7 You can see the relative frequency table that follows this answer. Knowing the response rate is critical for interpreting the results of a survey. The higher the response rate, the better. The response rate is $59.2\% + 19.8\% = 79.0\%$ – the total percentage of people who responded in any way (yes or no) to the survey. (Note that 21% is the nonresponse rate.)

Gone to the Zoo in the Last Year?	Relative Frequency
Y	$592 \div 1,000 = 0.592 = 59.2\%$
N	19.8%
Nonrespondents	21.0%

8 One disadvantage of a relative frequency table is that if you see only the percents, you don't know how many people participated in the study; therefore, you don't know how precise the results are. You can get around this problem by putting the total sample size somewhere at the top or bottom of your relative frequency table.

REMEMBER

When making a relative frequency table, include the total sample size somewhere on the table.

9 Be careful about how you interpret tables where an individual can be in more than one category at the same time.

a. The frequencies don't sum to 1,000, because people have the option to choose multiple locations or none at all, so each person doesn't end up in exactly one group. If you take the grand total of all the frequencies (1,493) and divide each frequency by 1,493 to get a relative frequency, the relative frequencies sum to 1 (or 100 percent). But what does that mean? It makes it hard to interpret these percents because they don't account for the total number of people.

b. One way you can summarize this data is by showing the percentage of people who have been at each location separately (compared to the percentage who haven't been there before). These percents add up to 1 for each location. The following table shows the results summarized with this method. ***Note:*** The table isn't a relative frequency table; however, it uses relative frequencies.

Location	% Who Have Been There	% Who Haven't Been There
Disney World	$216 \div 1{,}000 = 21.6\%$	$100\% - 21.6\% = 78.4\%$
New Orleans	$312 \div 1{,}000 = 31.2\%$	68.8%
Las Vegas	$418 \div 1{,}000 = 41.8\%$	58.2%
New York City	$359 \div 1{,}000 = 35.9\%$	64.1%
Washington, D.C.	$188 \div 1{,}000 = 18.8\%$	81.2%

REMEMBER

Not all tables involving percents should sum to 1. Don't force tables to sum to 1 when they shouldn't; do make sure you understand whether each individual can fall under more than one category. In those cases, a typical relative frequency table isn't appropriate.

10 You can always sum all the frequencies to get a total and then find each relative frequency by taking the frequency divided by the total. However, if you have only the percents, you can't go back and find the original counts unless you know the total number of individuals. Suppose that you know that 80 percent of the people in a survey like ice cream. How many people in the survey like ice cream? If the total number of respondents is 100, $80 = (100 * 0.80)$ people like ice cream. If the total is 50, you're looking at $40 = (50 * 0.80)$ positive answers. If the total is 5, you deal only with $4 = (5 * 0.80)$. This illustrates why relative frequency tables need to have the total sample size somewhere.

REMEMBER

Watch for total sample sizes when given a relative frequency table. Don't be misled by percentages alone, thinking they're always based on large sample sizes, because many are not.

IN THIS CHAPTER

- Calculating measures for center and spread
- Interpreting your results properly
- Compiling the results for a data discussion

Chapter 2

Summarizing Quantitative Data: Means, Medians, and More

Before data are organized in a chart or graph, the first step is to summarize them — that is, find a few numbers and/or words that can tell the story of the data in a nutshell. For quantitative data, the most important characteristics are the shape of the data (which you see in Chapter 4), where the center is located, and how much variability or spread is in the data. You may also want to point out any *outliers* in the data (numbers that appear far from the rest). And like everything else in statistics, there's room for people to stretch the truth in how they choose to summarize their data (or in what they choose not to tell you). So it's good to know the big ideas of how data are summarized and what to look for in terms of interpreting and evaluating data summaries. That's what you practice in this chapter.

Finding and Interpreting Measures of Center

The most common way to summarize quantitative data is to describe where the center is. One way of thinking about what *center* means is to ask, "What's a typical value in this data set?" You can measure the center of a data set in different ways, and the method you choose can greatly influence the conclusions people make about the data.

The mean of a data set is also known as the *average.* To find the mean, add all the numbers in the data set and divide by the number of numbers. The notation for the sample mean is $\bar{x}$, and the formula for the sample mean is $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. In this case the capital sigma stands for sum, and the subscript "i" starts at 1 and ends with n, so you are summing each value in the data set from x_1 to x_n. Then divide by n.

The *median* of a data set is the true middle value when the data are ordered from smallest to largest. To find the median, order the data and pick the middle number(s). If you have an odd number of numbers, only one value is in the middle. If you have an even number of numbers, you pinpoint two values in the middle and determine the average of the two to get the median.

See the following for an example of calculating the mean and median.

EXAMPLE

Q. Find the mean and the median of the following data set: 1, 6, 5, 7, 3, 2.5, 2, –2, 1, 0.

A. The mean is $1+6+5+7+3+2.5+2+-2+1+0=25.5$, divided by 10 (because you have 10 numbers), which equals 2.55. To find the median, order the numbers: –2, 0, 1, 1, 2, 2.5, 3, 5, 6, 7. Now find the middle number. In this, case there are two middle values: 2 and 2.5. Take the average: $(2+2.5)\div 2 = 4.5 \div 2 = 2.25$.

1 Does the mean have to be one of the numbers in the data set? Explain.

2 Does the median have to be one of the numbers in the data set? Explain.