

Making Everything Easier!™

2nd Edition

Statistics

FOR

DUMMIES®

Learn to:

- Grasp statistical ideas, techniques, formulas, and calculations

Interpret and critique graphs and charts, determine probability, and work with confidence intervals

Critique and analyze data from polls and experiments

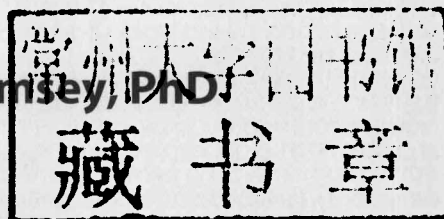
Deborah J. Rumsey, PhD

*Auxiliary Professor of Statistics,
The Ohio State University*



Statistics FOR DUMMIES® 2ND EDITION

by Deborah J. Rumsey, PhD



WILEY

Wiley Publishing, Inc.

Statistics For Dummies®, 2nd Edition

Published by
Wiley Publishing, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2011 by Wiley Publishing, Inc., Indianapolis, Indiana

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 either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. 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, the Wiley Publishing logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. Wiley Publishing, 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 www.wiley.com/techsupport.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Control Number: 2011921775

ISBN: 978-0-470-91108-2

ISBN 978-0-470-91108-2 (pbk); ISBN 978-1-118-01204-8 (ebk); ISBN 978-1-118-01205-5 (ebk); ISBN 978-1-118-01206-2 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4



About the Author

Deborah J. Rumsey, PhD, is a Statistics Education Specialist and Auxiliary Professor in the Department of Statistics at The Ohio State University. Dr. Rumsey is a Fellow of the American Statistical Association. She has won the Presidential Teaching Award from Kansas State University and has been inducted into the Wall of Inspiration at her high school alma mater, Burlington High School, in Burlington, Wisconsin. She is also the author of *Statistics II For Dummies*, *Statistics Workbook For Dummies*, *Probability For Dummies*, and *Statistics Essentials For Dummies*. She has published numerous papers and given many professional presentations and workshops on the subject of statistics education. She is the original conference designer of the biennial United States Conference on Teaching Statistics (USCOTS). Her passions include being with her family, camping and bird watching, getting seat time on her Kubota tractor, and cheering the Ohio State Buckeyes on to their next national championship.

Dedication

To my husband Eric: My sun rises and sets with you. To my son Clint: I love you up to the moon and back.

Author's Acknowledgments

My heartfelt thanks to Lindsay Lefevere and Kathy Cox for the opportunity to write *For Dummies* books for Wiley; to my project editors Georgette Beatty, Corbin Collins, and Tere Drenth for their unwavering support and vision; to Marjorie Bond, Monmouth College, for agreeing to be my technical editor (again!); to Paul Stephenson, who also provided technical editing; and to Caitie Copple and Janet Dunn for great copy editing.

Special thanks to Elizabeth Stasny, Joan Garfield, Kythrie Silva, Kit Kilen, Peg Steigerwald, Mike O'Leary, Tony Barkauskas, Ken Berk, and Jim Higgins for inspiration and support along the way; and to my entire family for their steadfast love and encouragement.

Publisher's Acknowledgments

We're proud of this book; please send us your comments at <http://dummies.custhelp.com>. For other comments, 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.

Some of the people who helped bring this book to market include the following:

Acquisitions, Editorial, and Media Development

Project Editor: Corbin Collins

(Previous Edition: Tere Drenth)

Senior Project Editor: Georgette Beatty

Executive Editor: Lindsay Sandman Lefevere

Copy Editor: Caitlin Cople

(Previous Edition: Janet S. Dunn, PhD)

Assistant Editor: David Lutton

Technical Editors: Marjorie E. Bond,
Paul L. Stephenson III

Editorial Manager: Michelle Hacker

Editorial Supervisor and Reprint Editor:
Carmen Krikorian

Editorial Assistant: Jennette ElNaggar

Cover Photo: © iStockphoto.com/Norebbo

Cartoons: Rich Tennant
(www.the5thwave.com)

Composition Services

Project Coordinator: Sheree Montgomery

Layout and Graphics: Carrie A. Cesavice,
Corrie Socolovitch

Proofreaders: Dwight Ramsey,
Shannon Ramsey

Indexer: Christine Karpeles

Publishing and Editorial for Consumer Dummies

Kathleen Nebenhaus, Vice President and Executive Publisher

Kristin Ferguson-Wagstaffe, Product Development Director

Ensley Eikenburg, Associate Publisher, Travel

Kelly Regan, Editorial Director, Travel

Publishing for Technology Dummies

Andy Cummings, Vice President and Publisher

Composition Services

Debbie Stailey, Director of Composition Services

Contents at a Glance



Introduction	1
Part I: Vital Statistics about Statistics	7
Chapter 1: Statistics in a Nutshell.....	9
Chapter 2: The Statistics of Everyday Life	23
Chapter 3: Taking Control: So Many Numbers, So Little Time	33
Chapter 4: Tools of the Trade	43
Part II: Number-Crunching Basics.....	65
Chapter 5: Means, Medians, and More.....	67
Chapter 6: Getting the Picture: Graphing Categorical Data	91
Chapter 7: Going by the Numbers: Graphing Numerical Data	103
Part III: Distributions and the Central Limit Theorem	129
Chapter 8: Random Variables and the Binomial Distribution.....	131
Chapter 9: The Normal Distribution.....	143
Chapter 10: The <i>t</i> -Distribution	157
Chapter 11: Sampling Distributions and the Central Limit Theorem.....	163
Part IV: Guesstimating and Hypothesizing with Confidence.....	179
Chapter 12: Leaving Room for a Margin of Error.....	181
Chapter 13: Confidence Intervals: Making Your Best Guesstimate.....	193
Chapter 14: Claims, Tests, and Conclusions	215
Chapter 15: Commonly Used Hypothesis Tests: Formulas and Examples.....	227
Part V: Statistical Studies and the Hunt for a Meaningful Relationship	243
Chapter 16: Polls, Polls, and More Polls	245
Chapter 17: Experiments: Medical Breakthroughs or Misleading Results?	261
Chapter 18: Looking for Links: Correlation and Regression	279
Chapter 19: Two-Way Tables and Independence	295

<i>Part VI: The Part of Tens</i>	317
Chapter 20: Ten Tips for the Statistically Savvy Sleuth	319
Chapter 21: Ten Surefire Exam Score Boosters	331
<i>Appendix: Tables for Reference</i>	347
<i>Index</i>	357

Table of Contents

Introduction 1

About This Book	1
Conventions Used in This Book.....	2
What You're Not to Read	2
Foolish Assumptions.....	3
How This Book Is Organized	3
Part I: Vital Statistics about Statistics	3
Part II: Number-Crunching Basics.....	4
Part III: Distributions and the Central Limit Theorem	4
Part IV: Guesstimating and Hypothesizing with Confidence.....	4
Part V: Statistical Studies and the Hunt for a Meaningful Relationship.....	5
Part VI: The Part of Tens	5
Icons Used in This Book	6
Where to Go from Here	6

Part I: Vital Statistics about Statistics 7

Chapter 1: Statistics in a Nutshell.9

Thriving in a Statistical World	9
Designing Appropriate Studies	11
Surveys	11
Experiments.....	12
Collecting Quality Data	12
Selecting a good sample.....	13
Avoiding bias in your data.....	13
Creating Effective Summaries	14
Descriptive statistics.....	14
Charts and graphs	15
Determining Distributions	16
Performing Proper Analyses	17
Margin of error and confidence intervals	17
Hypothesis tests	18
Correlation, regression, and two-way tables.....	20
Drawing Credible Conclusions.....	21
Reeling in overstated results.....	21
Questioning claims of cause and effect.....	21
Becoming a Sleuth, Not a Skeptic	22

Chapter 2: The Statistics of Everyday Life23

Statistics and the Media: More Questions than Answers?	23
Probing popcorn problems	24
Venturing into viruses	24
Comprehending crashes	25
Mulling malpractice	25
Belaboring the loss of land	26
Scrutinizing schools	26
Studying sports	27
Banking on business news	28
Touring the travel news	28
Surveying sexual stats	29
Breaking down weather reports	29
Musing about movies	30
Highlighting horoscopes	30
Using Statistics at Work	31
Delivering babies — and information	31
Posing for pictures	31
Poking through pizza data	32
Statistics in the office	32

Chapter 3: Taking Control: So Many Numbers, So Little Time.33

Detecting Errors, Exaggerations, and Just Plain Lies	34
Checking the math	34
Uncovering misleading statistics	35
Looking for lies in all the right places	41
Feeling the Impact of Misleading Statistics	42

Chapter 4: Tools of the Trade43

Statistics: More than Just Numbers	43
Grabbing Some Basic Statistical Jargon	45
Data	45
Data set	47
Variable	47
Population	47
Sample, random, or otherwise	48
Statistic	50
Parameter	50
Bias	51
Mean (Average)	51
Median	52
Standard deviation	52
Percentile	53
Standard score	54

Distribution and normal distribution.....	54
Central Limit Theorem	55
z-values	56
Experiments.....	56
Surveys (Polls)	58
Margin of error.....	58
Confidence interval.....	59
Hypothesis testing.....	60
p-values	61
Statistical significance.....	61
Correlation versus causation	63

***Part II: Number-Crunching Basics* 65**

Chapter 5: Means, Medians, and More 67

Summing Up Data with Descriptive Statistics.....	67
Crunching Categorical Data: Tables and Percents.....	68
Measuring the Center with Mean and Median	71
Averaging out to the mean	71
Splitting your data down the median	73
Comparing means and medians: Histograms.....	74
Accounting for Variation	76
Reporting the standard deviation.....	77
Being out of range.....	80
Examining the Empirical Rule (68-95-99.7)	81
Measuring Relative Standing with Percentiles.....	84
Calculating percentiles.....	84
Interpreting percentiles	85
Gathering a five-number summary	89
Exploring interquartile range.....	90

Chapter 6: Getting the Picture: Graphing Categorical Data 91

Take Another Little Piece of My Pie Chart	92
Tallying personal expenses	92
Bringing in a lotto revenue	92
Ordering takeout.....	94
Projecting age trends	95
Raising the Bar on Bar Graphs.....	97
Tracking transportation expenses.....	97
Making a lotto profit	99
Tipping the scales on a bar graph	100
Pondering pet peeves.....	101

Chapter 7: Going by the Numbers: Graphing Numerical Data103

Handling Histograms.....	103
Making a histogram	104
Interpreting a histogram	106
Putting numbers with pictures	110
Detecting misleading histograms	112
Examining Boxplots.....	115
Making a boxplot.....	115
Interpreting a boxplot	117
Tackling Time Charts	123
Interpreting time charts.....	123
Understanding variability: Time charts versus histograms.....	124
Spotting misleading time charts	124

Part III: Distributions and the Central Limit Theorem... 129

Chapter 8: Random Variables and the Binomial Distribution131

Defining a Random Variable.....	131
Discrete versus continuous	132
Probability distributions.....	133
The mean and variance of a discrete random variable	134
Identifying a Binomial	135
Checking binomial conditions step by step	135
No fixed number of trials	136
More than success or failure.....	136
Trials are not independent	137
Probability of success (p) changes	137
Finding Binomial Probabilities Using a Formula	137
Finding Probabilities Using the Binomial Table	140
Finding probabilities for specific values of X	140
Finding probabilities for X greater-than, less-than, or between two values.....	141
Checking Out the Mean and Standard Deviation of the Binomial	142

Chapter 9: The Normal Distribution143

Exploring the Basics of the Normal Distribution.....	143
Meeting the Standard Normal (Z -) Distribution.....	146
Checking out Z	146
Standardizing from X to Z	147
Finding probabilities for Z with the Z -table	148
Finding Probabilities for a Normal Distribution	149
Finding X When You Know the Percent.....	152
Figuring out a percentile for a normal distribution.....	152
Translating tricky wording in percentile problems.....	154
Normal Approximation to the Binomial	155

Chapter 10: The t -Distribution157

Basics of the t -Distribution	157
Comparing the t - and Z -distributions	157
Discovering the effect of variability on t -distributions	159
Using the t -Table	159
Finding probabilities with the t -table	160
Figuring percentiles for the t -distribution	160
Picking out t^* -values for confidence intervals	161
Studying Behavior Using the t -Table	162

**Chapter 11: Sampling Distributions
and the Central Limit Theorem.163**

Defining a Sampling Distribution	163
The Mean of a Sampling Distribution	164
Measuring Standard Error	166
Sample size and standard error	166
Population standard deviation and standard error	168
Looking at the Shape of a Sampling Distribution	169
Case 1: The distribution of X is normal	170
Case 2: The distribution of X is not normal — enter the Central Limit Theorem	170
Finding Probabilities for the Sample Mean	173
The Sampling Distribution of the Sample Proportion	175
Finding Probabilities for the Sample Proportion	177

**Part IV: Guesstimating and Hypothesizing
with Confidence 179****Chapter 12: Leaving Room for a Margin of Error.181**

Seeing the Importance of That Plus or Minus	181
Finding the Margin of Error: A General Formula	183
Measuring sample variability	183
Calculating margin of error for a sample proportion	184
Reporting results	186
Calculating margin of error for a sample mean	187
Being confident you're right	188
Determining the Impact of Sample Size	189
Sample size and margin of error	189
Bigger isn't always (that much) better!	189
Keeping margin of error in perspective	190

**Chapter 13: Confidence Intervals:
Making Your Best Guesstimate193**

Not All Estimates Are Created Equal	193
Linking a Statistic to a Parameter	194

Getting with the Jargon.....	195
Interpreting Results with Confidence	196
Zooming In on Width.....	197
Choosing a Confidence Level	198
Factoring In the Sample Size	199
Counting On Population Variability	201
Calculating a Confidence Interval for a Population Mean	201
Case 1: Population standard deviation is known.....	202
Case 2: Population standard deviation is unknown and/or n is small.....	203
Figuring Out What Sample Size You Need.....	204
Determining the Confidence Interval for One Population Proportion	206
Creating a Confidence Interval for the Difference of Two Means.....	207
Case 1: Population standard deviations are known.....	208
Case 2: Population standard deviations are unknown and/or sample sizes are small	210
Estimating the Difference of Two Proportions	211
Spotting Misleading Confidence Intervals	213
Chapter 14: Claims, Tests, and Conclusions	215
Setting Up the Hypotheses	216
Defining the null	216
What's the alternative?	216
Gathering Good Evidence (Data).....	217
Compiling the Evidence: The Test Statistic	218
Gathering sample statistics	218
Measuring variability using standard errors.....	218
Understanding standard scores.....	219
Calculating and interpreting the test statistic	219
Weighing the Evidence and Making Decisions: p -Values.....	220
Connecting test statistics and p -values	220
Defining a p -value.....	221
Calculating a p -value.....	222
Making Conclusions	223
Setting boundaries for rejecting H_0	224
Testing varicose veins.....	224
Assessing the Chance of a Wrong Decision	225
Making a false alarm: Type-1 errors	225
Missing out on a detection: Type-2 errors.....	226
Chapter 15: Commonly Used Hypothesis Tests: Formulas and Examples	227
Testing One Population Mean.....	228
Handling Small Samples and Unknown Standard Deviations: The t -Test.....	229
Putting the t -test to work	230
Relating t to Z	231

Handling negative t -values	231
Examining the not-equal-to alternative	232
Testing One Population Proportion	232
Comparing Two (Independent) Population Averages	234
Testing for an Average Difference (The Paired t -Test)	236
Comparing Two Population Proportions	240

***Part V: Statistical Studies and the Hunt for a Meaningful Relationship* 243**

Chapter 16: Polls, Polls, and More Polls 245

Recognizing the Impact of Polls	245
Getting to the source	246
Surveying what's hot	248
Impacting lives	248
Behind the Scenes: The Ins and Outs of Surveys	250
Planning and designing a survey	250
Selecting the sample	254
Carrying out a survey	256
Interpreting results and finding problems	259

Chapter 17: Experiments: Medical Breakthroughs or Misleading Results? 261

Boiling Down the Basics of Studies	262
Looking at the lingo of studies	262
Observing observational studies	263
Examining experiments	264
Designing a Good Experiment	264
Designing the experiment to make comparisons	265
Selecting the sample size	267
Choosing the subjects	268
Making random assignments	269
Controlling for confounding variables	270
Respecting ethical issues	272
Collecting good data	273
Analyzing the data properly	274
Making appropriate conclusions	275
Making Informed Decisions	277

Chapter 18: Looking for Links: Correlation and Regression 279

Picturing a Relationship with a Scatterplot	280
Making a scatterplot	281
Interpreting a scatterplot	281
Quantifying Linear Relationships Using the Correlation	282
Calculating the correlation	283
Interpreting the correlation	284
Examining properties of the correlation	286

Working with Linear Regression.....	286
Figuring out which variable is X and which is Y	287
Checking the conditions	287
Calculating the regression line.....	288
Interpreting the regression line	290
Putting it all together with an example:	
The regression line for the crickets.....	291
Making Proper Predictions.....	292
Explaining the Relationship: Correlation versus Cause and Effect.....	293
Chapter 19: Two-Way Tables and Independence	295
Organizing a Two-Way Table	296
Setting up the cells	297
Figuring the totals.....	297
Interpreting Two-Way Tables.....	298
Singling out variables with marginal distributions.....	299
Examining all groups — a joint distribution.....	302
Comparing groups with conditional distributions	305
Checking Independence and Describing Dependence.....	308
Checking for independence	308
Describing a dependent relationship.....	311
Cautiously Interpreting Results	312
Checking for legitimate cause and effect	313
Projecting from sample to population	314
Making prudent predictions	315
Resisting the urge to jump to conclusions	315
 Part VI: The Part of Tens.....	 317
Chapter 20: Ten Tips for the Statistically Savvy Sleuth	319
Chapter 21: Ten Surefire Exam Score Boosters	331
 Appendix: Tables for Reference.....	 347
 Index.....	 357

Introduction

You get hit with an incredible amount of statistical information on a daily basis. You know what I'm talking about: charts, graphs, tables, and headlines that talk about the results of the latest poll, survey, experiment, or other scientific study. The purpose of this book is to develop and sharpen your skills in sorting through, analyzing, and evaluating all that info, and to do so in a clear, fun, and pain-free way. You also gain the ability to decipher and make important decisions about statistical results (for example, the results of the latest medical studies), while being ever aware of the ways that people can mislead you with statistics. And you see how to do it right when it's your turn to design the study, collect the data, crunch the numbers, and/or draw the conclusions.

This book is also designed to help those of you out there who are taking an introductory statistics class and can use some back-up. You'll gain a working knowledge of the big ideas of statistics and gather a boatload of tools and tricks of the trade that'll help you get ahead of the curve when you take your exams.

This book is chock-full of real examples from real sources that are relevant to your everyday life — from the latest medical breakthroughs, crime studies, and population trends to the latest U.S. government reports. I even address a survey on the worst cars of the millennium! By reading this book, you'll understand how to collect, display, and analyze data correctly and effectively, and you'll be ready to critically examine and make informed decisions about the latest polls, surveys, experiments, and reports that bombard you every day. You even find out how to use crickets to gauge temperature!

You also get to enjoy poking a little fun at statisticians (who take themselves too seriously at times). After all, with the right skills and knowledge, you don't have to be a statistician to understand introductory statistics.

About This Book

This book departs from traditional statistics texts, references, supplemental books, and study guides in the following ways:

- ✓ It includes practical and intuitive explanations of statistical concepts, ideas, techniques, formulas, and calculations found in an introductory statistics course.
- ✓ It shows you clear and concise step-by-step procedures that explain how you can intuitively work through statistics problems.

- ✓ It includes interesting real-world examples relating to your everyday life and workplace.
- ✓ It gives you upfront and honest answers to your questions like, “What does this really mean?” and “When and how will I ever use this?”

Conventions Used in This Book

You should be aware of three conventions as you make your way through this book:

- ✓ **Definition of sample size (n):** When I refer to the size of a sample, I mean the final number of individuals who participated in and provided information for the study. In other words, n stands for the size of the final data set.
- ✓ **Dual-use of the word *statistics*:** In some situations, I refer to statistics as a subject of study or as a field of research, so the word is a singular noun. For example, “Statistics is really quite an interesting subject.” 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.”
- ✓ **Use of the word *data*:** You’re probably unaware of the debate raging amongst statisticians about whether the word *data* should be singular (“data is . . .”) or plural (“data are . . .”). It got so bad that recently one group of statisticians had to develop two different versions of a statistics T-shirt: “Messy Data Happens” and “Messy Data Happen.” At the risk of offending some of my colleagues, I go with the plural version of the word *data* in this book.
- ✓ **Use of the term *standard deviation*:** When I use the term *standard deviation*, I mean s , the sample standard deviation. (When I refer to the population standard deviation, I let you know.)

Here are a few other basic conventions to help you navigate this book:

- ✓ I use *italics* to let you know a new statistical term is appearing on the scene.
- ✓ If you see a **boldfaced** term or phrase in a bulleted list, it’s been designated as a keyword or key phrase.
- ✓ Addresses for Web sites appear in `mono font`.

What You’re Not to Read

I like to think that you won’t skip anything in this book, but I also know you’re a busy person. So to save time, feel free to skip anything marked with the