

LEARNING MADE EASY



2nd Edition

# Statistics

for  
**dummies**<sup>®</sup>  
A Wiley Brand



Grasp statistical ideas,  
formulas, and calculations

Interpret graphs and polls  
and determine probability

Analyze data using  
many techniques

**Deborah J. Rumsey, PhD**

*Professor of Statistics, The Ohio  
State University*





# Statistics

for  
**dummies**<sup>®</sup>  
A Wiley Brand

2nd edition

**by Deborah J. Rumsey, PhD**  
Professor of Statistics, The Ohio State University

for  
**dummies**<sup>®</sup>  
A Wiley Brand

## Statistics For Dummies®, 2nd Edition

Published by: **Wiley Publishing, Inc.**, 111 River St., Hoboken, NJ 07030-5774, [www.wiley.com](http://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](http://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: 978-1-119-29352-1

ISBN 978-1-119-29352-1 (pbk); ISBN 978-1-119-29749-9 (ebk); ISBN 978-1-119-29751-2 (ebk)

Statistics For Dummies, 2nd Edition (9781119293521) was previously published as Statistics For Dummies, 2nd Edition (9780470911082). While this version features a new Dummies cover and design, the content is the same as the prior release and should not be considered a new or updated product.

Manufactured in the United States of America

20 19 18 17 16 15 14 13 12 11

# Contents at a Glance

<b>Introduction</b> .....	1
<b>Part 1: Vital Statistics about Statistics</b> .....	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 .....	35
CHAPTER 4: Tools of the Trade .....	47
<b>Part 2: Number-Crunching Basics</b> .....	69
CHAPTER 5: Means, Medians, and More .....	71
CHAPTER 6: Getting the Picture: Graphing Categorical Data .....	95
CHAPTER 7: Going by the Numbers: Graphing Numerical Data .....	107
<b>Part 3: Distributions and the Central Limit Theorem</b> .....	133
CHAPTER 8: Random Variables and the Binomial Distribution .....	135
CHAPTER 9: The Normal Distribution .....	149
CHAPTER 10: The $t$ -Distribution .....	165
CHAPTER 11: Sampling Distributions and the Central Limit Theorem .....	171
<b>Part 4: Guesstimating and Hypothesizing with Confidence</b> .....	187
CHAPTER 12: Leaving Room for a Margin of Error .....	189
CHAPTER 13: Confidence Intervals: Making Your Best Guesstimate .....	201
CHAPTER 14: Claims, Tests, and Conclusions .....	223
CHAPTER 15: Commonly Used Hypothesis Tests: Formulas and Examples .....	237
<b>Part 5: Statistical Studies and the Hunt for a Meaningful Relationship</b> .....	255
CHAPTER 16: Polls, Polls, and More Polls .....	257
CHAPTER 17: Experiments: Medical Breakthroughs or Misleading Results? .....	275
CHAPTER 18: Looking for Links: Correlation and Regression .....	293
CHAPTER 19: Two-Way Tables and Independence .....	311
<b>Part 6: The Part of Tens</b> .....	333
CHAPTER 20: Ten Tips for the Statistically Savvy Sleuth .....	335
CHAPTER 21: Ten Surefire Exam Score Boosters .....	349
<b>Appendix: Tables for Reference</b> .....	365
<b>Index</b> .....	375



# Table of Contents

<b>INTRODUCTION</b> .....	1
About This Book .....	1
Conventions Used in This Book .....	2
What You're Not to Read .....	3
Foolish Assumptions .....	3
How This Book Is Organized .....	3
Part 1: Vital Statistics about Statistics .....	3
Part 2: Number-Crunching Basics .....	4
Part 3: Distributions and the Central Limit Theorem .....	4
Part 4: Guesstimating and Hypothesizing with Confidence .....	4
Part 5: Statistical Studies and the Hunt for a Meaningful Relationship .....	5
Part 6: The Part of Tens .....	5
Icons Used in This Book .....	6
Where to Go from Here .....	6
<b>PART 1: VITAL STATISTICS ABOUT STATISTICS</b> .....	7
<b>CHAPTER 1: Statistics in a Nutshell</b> .....	9
Thriving in a Statistical World .....	10
Designing Appropriate Studies .....	11
Surveys .....	11
Experiments .....	12
Collecting Quality Data .....	13
Selecting a good sample .....	13
Avoiding bias in your data .....	14
Creating Effective Summaries .....	14
Descriptive statistics .....	15
Charts and graphs .....	15
Determining Distributions .....	16
Performing Proper Analyses .....	17
Margin of error and confidence intervals .....	18
Hypothesis tests .....	19
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

<b>CHAPTER 2:</b>	<b>The Statistics of Everyday Life</b> .....	23
	Statistics and the Media: More Questions than Answers? .....	24
	Probing popcorn problems .....	24
	Venturing into viruses .....	24
	Comprehending crashes .....	25
	Mulling malpractice .....	26
	Belaboring the loss of land .....	26
	Scrutinizing schools .....	27
	Studying sports .....	28
	Banking on business news .....	28
	Touring the travel news .....	29
	Surveying sexual stats .....	29
	Breaking down weather reports .....	30
	Musing about movies .....	30
	Highlighting horoscopes .....	31
	Using Statistics at Work .....	31
	Delivering babies — and information .....	31
	Posing for pictures .....	32
	Poking through pizza data .....	32
	Statistics in the office .....	33
<b>CHAPTER 3:</b>	<b>Taking Control: So Many Numbers, So Little Time</b> .....	35
	Detecting Errors, Exaggerations, and Just Plain Lies .....	36
	Checking the math .....	36
	Uncovering misleading statistics .....	37
	Looking for lies in all the right places .....	44
	Feeling the Impact of Misleading Statistics .....	44
<b>CHAPTER 4:</b>	<b>Tools of the Trade</b> .....	47
	Statistics: More than Just Numbers .....	47
	Grabbing Some Basic Statistical Jargon .....	49
	Data .....	50
	Data set .....	51
	Variable .....	51
	Population .....	51
	Sample, random, or otherwise .....	52
	Statistic .....	54
	Parameter .....	54
	Bias .....	55
	Mean (Average) .....	55
	Median .....	56
	Standard deviation .....	56
	Percentile .....	57



	Standard score . . . . .	57
	Distribution and normal distribution . . . . .	58
	Central Limit Theorem . . . . .	59
	z-values . . . . .	60
	Experiments . . . . .	60
	Surveys (Polls) . . . . .	62
	Margin of error . . . . .	62
	Confidence interval . . . . .	63
	Hypothesis testing . . . . .	64
	p-values . . . . .	65
	Statistical significance . . . . .	66
	Correlation versus causation . . . . .	67
	<b>PART 2: NUMBER-CRUNCHING BASICS . . . . .</b>	<b>69</b>
<b>CHAPTER 5:</b>	<b>Means, Medians, and More . . . . .</b>	<b>71</b>
	Summing Up Data with Descriptive Statistics . . . . .	71
	Crunching Categorical Data: Tables and Percents . . . . .	72
	Measuring the Center with Mean and Median . . . . .	75
	Averaging out to the mean . . . . .	75
	Splitting your data down the median . . . . .	77
	Comparing means and medians: Histograms . . . . .	78
	Accounting for Variation . . . . .	80
	Reporting the standard deviation . . . . .	81
	Being out of range . . . . .	84
	Examining the Empirical Rule (68-95-99.7) . . . . .	85
	Measuring Relative Standing with Percentiles . . . . .	87
	Calculating percentiles . . . . .	88
	Interpreting percentiles . . . . .	89
	Gathering a five-number summary . . . . .	93
	Exploring interquartile range . . . . .	94
<b>CHAPTER 6:</b>	<b>Getting the Picture: Graphing Categorical Data . . . . .</b>	<b>95</b>
	Take Another Little Piece of My Pie Chart . . . . .	96
	Tallying personal expenses . . . . .	96
	Bringing in a lotto revenue . . . . .	97
	Ordering takeout . . . . .	98
	Projecting age trends . . . . .	99
	Raising the Bar on Bar Graphs . . . . .	101
	Tracking transportation expenses . . . . .	101
	Making a lotto profit . . . . .	103
	Tipping the scales on a bar graph . . . . .	104
	Pondering pet peeves . . . . .	105

<b>CHAPTER 7: Going by the Numbers: Graphing Numerical Data</b> .....	107
Handling Histograms .....	108
Making a histogram .....	108
Interpreting a histogram .....	111
Putting numbers with pictures .....	115
Detecting misleading histograms .....	117
Examining Boxplots .....	120
Making a boxplot .....	120
Interpreting a boxplot .....	121
Tackling Time Charts .....	127
Interpreting time charts .....	127
Understanding variability: Time charts versus histograms .....	128
Spotting misleading time charts .....	128

**PART 3: DISTRIBUTIONS AND THE CENTRAL LIMIT THEOREM** .....

<b>CHAPTER 8: Random Variables and the Binomial Distribution</b> .....	135
Defining a Random Variable .....	136
Discrete versus continuous .....	136
Probability distributions .....	137
The mean and variance of a discrete random variable .....	138
Identifying a Binomial .....	139
Checking binomial conditions step by step .....	140
No fixed number of trials .....	140
More than success or failure .....	141
Trials are not independent .....	141
Probability of success (p) changes .....	141
Finding Binomial Probabilities Using a Formula .....	142
Finding Probabilities Using the Binomial Table .....	144
Finding probabilities for specific values of X .....	145
Finding probabilities for X greater-than, less-than, or between two values .....	146
Checking Out the Mean and Standard Deviation of the Binomial .....	146
<b>CHAPTER 9: The Normal Distribution</b> .....	149
Exploring the Basics of the Normal Distribution .....	150
Meeting the Standard Normal (Z-) Distribution .....	152
Checking out Z .....	153
Standardizing from X to Z .....	153
Finding probabilities for Z with the Z-table .....	155
Finding Probabilities for a Normal Distribution .....	156

Finding X When You Know the Percent . . . . .	158
Figuring out a percentile for a normal distribution . . . . .	159
Translating tricky wording in percentile problems. . . . .	161
Normal Approximation to the Binomial. . . . .	162
<b>CHAPTER 10: The t-Distribution</b> . . . . .	165
Basics of the t-Distribution. . . . .	165
Comparing the t- and Z-distributions . . . . .	165
Discovering the effect of variability on t-distributions. . . . .	167
Using the t-Table . . . . .	167
Finding probabilities with the t-table . . . . .	168
Figuring percentiles for the t-distribution . . . . .	168
Picking out $t^*$ -values for confidence intervals. . . . .	169
Studying Behavior Using the t-Table. . . . .	170
<b>CHAPTER 11: Sampling Distributions and the Central Limit Theorem</b> . . . . .	171
Defining a Sampling Distribution . . . . .	172
The Mean of a Sampling Distribution. . . . .	174
Measuring Standard Error . . . . .	174
Sample size and standard error . . . . .	175
Population standard deviation and standard error. . . . .	176
Looking at the Shape of a Sampling Distribution. . . . .	178
Case 1: The distribution of X is normal. . . . .	178
Case 2: The distribution of X is not normal—enter the Central Limit Theorem . . . . .	178
Finding Probabilities for the Sample Mean . . . . .	181
The Sampling Distribution of the Sample Proportion . . . . .	183
Finding Probabilities for the Sample Proportion . . . . .	185
<b>PART 4: GUESSTIMATING AND HYPOTHESIZING WITH CONFIDENCE</b> . . . . .	187
<b>CHAPTER 12: Leaving Room for a Margin of Error</b> . . . . .	189
Seeing the Importance of That Plus or Minus. . . . .	190
Finding the Margin of Error: A General Formula. . . . .	191
Measuring sample variability. . . . .	191
Calculating margin of error for a sample proportion. . . . .	193
Reporting results . . . . .	194
Calculating margin of error for a sample mean . . . . .	195
Being confident you're right. . . . .	197
Determining the Impact of Sample Size. . . . .	197
Sample size and margin of error. . . . .	198
Bigger isn't always (that much) better! . . . . .	198
Keeping margin of error in perspective . . . . .	199

<b>CHAPTER 13: Confidence Intervals: Making Your Best Guesstimate</b>	201
Not All Estimates Are Created Equal	202
Linking a Statistic to a Parameter	203
Getting with the Jargon	203
Interpreting Results with Confidence	204
Zooming In on Width	205
Choosing a Confidence Level	206
Factoring In the Sample Size	208
Counting On Population Variability	209
Calculating a Confidence Interval for a Population Mean	210
Case 1: Population standard deviation is known	210
Case 2: Population standard deviation is unknown and/or n is small	212
Figuring Out What Sample Size You Need	213
Determining the Confidence Interval for One Population Proportion	214
Creating a Confidence Interval for the Difference of Two Means	216
Case 1: Population standard deviations are known	216
Case 2: Population standard deviations are unknown and/or sample sizes are small	218
Estimating the Difference of Two Proportions	219
Spotting Misleading Confidence Intervals	221
<b>CHAPTER 14: Claims, Tests, and Conclusions</b>	223
Setting Up the Hypotheses	224
Defining the null	224
What's the alternative?	225
Gathering Good Evidence (Data)	226
Compiling the Evidence: The Test Statistic	226
Gathering sample statistics	227
Measuring variability using standard errors	227
Understanding standard scores	228
Calculating and interpreting the test statistic	228
Weighing the Evidence and Making Decisions: p-Values	229
Connecting test statistics and p-values	229
Defining a p-value	230
Calculating a p-value	230
Making Conclusions	231
Setting boundaries for rejecting $H_0$	232
Testing varicose veins	233
Assessing the Chance of a Wrong Decision	233
Making a false alarm: Type-1 errors	234
Missing out on a detection: Type-2 errors	234

<b>CHAPTER 15: Commonly Used Hypothesis Tests: Formulas and Examples</b> .....	237
Testing One Population Mean .....	238
Handling Small Samples and Unknown Standard Deviations: The t-Test .....	240
Putting the t-test to work .....	241
Relating t to Z .....	241
Handling negative t-values .....	242
Examining the not-equal-to alternative .....	242
Testing One Population Proportion .....	243
Comparing Two (Independent) Population Averages .....	245
Testing for an Average Difference (The Paired t-Test) .....	247
Comparing Two Population Proportions .....	251

**PART 5: STATISTICAL STUDIES AND THE HUNT  
FOR A MEANINGFUL RELATIONSHIP** .....

<b>CHAPTER 16: Polls, Polls, and More Polls</b> .....	257
Recognizing the Impact of Polls .....	258
Getting to the source .....	258
Surveying what's hot .....	260
Impacting lives .....	260
Behind the Scenes: The Ins and Outs of Surveys .....	262
Planning and designing a survey .....	263
Selecting the sample .....	266
Carrying out a survey .....	268
Interpreting results and finding problems .....	271

<b>CHAPTER 17: Experiments: Medical Breakthroughs or Misleading Results?</b> .....	275
Boiling Down the Basics of Studies .....	276
Looking at the lingo of studies .....	276
Observing observational studies .....	277
Examining experiments .....	278
Designing a Good Experiment .....	278
Designing the experiment to make comparisons .....	279
Selecting the sample size .....	281
Choosing the subjects .....	283
Making random assignments .....	283
Controlling for confounding variables .....	284
Respecting ethical issues .....	286
Collecting good data .....	287
Analyzing the data properly .....	289
Making appropriate conclusions .....	290
Making Informed Decisions .....	292

<b>CHAPTER 18: Looking for Links: Correlation and Regression</b> . . . . .	293
Picturing a Relationship with a Scatterplot . . . . .	294
Making a scatterplot . . . . .	295
Interpreting a scatterplot . . . . .	296
Quantifying Linear Relationships Using the Correlation . . . . .	297
Calculating the correlation . . . . .	297
Interpreting the correlation . . . . .	298
Examining properties of the correlation . . . . .	300
Working with Linear Regression . . . . .	301
Figuring out which variable is X and which is Y . . . . .	301
Checking the conditions . . . . .	302
Calculating the regression line . . . . .	302
Interpreting the regression line . . . . .	304
Putting it all together with an example: The regression line for the crickets . . . . .	306
Making Proper Predictions . . . . .	306
Explaining the Relationship: Correlation versus Cause and Effect . . . . .	308
<b>CHAPTER 19: Two-Way Tables and Independence</b> . . . . .	311
Organizing a Two-Way Table . . . . .	312
Setting up the cells . . . . .	313
Figuring the totals . . . . .	314
Interpreting Two-Way Tables . . . . .	315
Singling out variables with marginal distributions . . . . .	315
Examining all groups — a joint distribution . . . . .	317
Comparing groups with conditional distributions . . . . .	321
Checking Independence and Describing Dependence . . . . .	324
Checking for independence . . . . .	324
Describing a dependent relationship . . . . .	327
Cautiously Interpreting Results . . . . .	329
Checking for legitimate cause and effect . . . . .	329
Projecting from sample to population . . . . .	330
Making prudent predictions . . . . .	331
Resisting the urge to jump to conclusions . . . . .	332
<b>PART 6: THE PART OF TENS</b> . . . . .	333
<b>CHAPTER 20: Ten Tips for the Statistically Savvy Sleuth</b> . . . . .	335
Pinpoint Misleading Graphs . . . . .	335
Pie charts . . . . .	336
Bar graphs . . . . .	336
Time charts . . . . .	337
Histograms . . . . .	339

Uncover Biased Data . . . . .	339
Search for a Margin of Error . . . . .	340
Identify Non-Random Samples . . . . .	341
Sniff Out Missing Sample Sizes . . . . .	342
Detect Misinterpreted Correlations . . . . .	343
Reveal Confounding Variables. . . . .	344
Inspect the Numbers . . . . .	344
Report Selective Reporting. . . . .	345
Expose the Anecdote . . . . .	346
<b>CHAPTER 21: Ten Surefire Exam Score Boosters. . . . .</b>	<b>349</b>
Know What You Don't Know, and then Do Something about It . . . . .	350
Avoid "Yeah-Yeah" Traps . . . . .	351
Yeah-yeah trap #1 . . . . .	352
Yeah-yeah trap #2 . . . . .	352
Make Friends with Formulas . . . . .	354
Make an "If-Then-How" Chart . . . . .	355
Figure Out What the Question Is Asking . . . . .	357
Label What You're Given. . . . .	358
Draw a Picture . . . . .	360
Make the Connection and Solve the Problem. . . . .	361
Do the Math — Twice . . . . .	362
Analyze Your Answers. . . . .	363
<b>APPENDIX: TABLES FOR REFERENCE. . . . .</b>	<b>365</b>
<b>INDEX . . . . .</b>	<b>375</b>





# Introduction

**Y**ou 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 monofont.

# 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 Technical Stuff icon as well as text in sidebars (the shaded gray boxes that appear throughout the book). These items feature information that's interesting but not crucial to your basic knowledge of statistics.

## Foolish Assumptions

I don't assume that you've had any previous experience with statistics, other than the fact that you're a member of the general public who gets bombarded every day with statistics in the form of numbers, percents, charts, graphs, "statistically significant" results, "scientific" studies, polls, surveys, experiments, and so on.

What I do assume is that you can do some of the basic mathematical operations and understand some of the basic notation used in algebra, such as the variables  $x$  and  $y$ , summation signs, taking the square root, squaring a number, and so on. If you need to brush up on your algebra skills, check out *Algebra I For Dummies*, 2nd Edition, by Mary Jane Sterling (Wiley).

I don't want to mislead you: You do encounter formulas in this book, because statistics does involve a bit of number crunching. But don't let that worry you. I take you slowly and carefully through each step of any calculations you need to do. I also provide examples for you to work along with this book, so that you can become familiar and comfortable with the calculations and make them your own.

## How This Book Is Organized

This book is organized into five parts that explore the major areas of introductory statistics, along with a final part that offers some quick top-ten nuggets for your information and enjoyment. Each part contains chapters that break down each major area of statistics into understandable pieces.

### Part 1: Vital Statistics about Statistics

This part helps you become aware of the quantity and quality of statistics you encounter in your workplace and your everyday life. You find out that a great deal of that statistical information is incorrect, either by accident or by design. You take

a first step toward becoming statistically savvy by recognizing some of the tools of the trade, developing an overview of statistics as a process for getting and interpreting information, and getting up to speed on some statistical jargon.

## Part 2: Number-Crunching Basics

This part helps you become more familiar and comfortable with making, interpreting, and evaluating data displays (otherwise known as charts, graphs, and so on) for different types of data. You also find out how to summarize and explore data by calculating and combining some commonly used statistics as well as some statistics you may not know about yet.

## Part 3: Distributions and the Central Limit Theorem

In this part, you get into all the details of the three most common statistical distributions: the binomial distribution, the normal (and standard normal, also known as  $Z$ -distribution), and the  $t$ -distribution. You discover the characteristics of each distribution and how to find and interpret probabilities, percentiles, means, and standard deviations. You also find measures of relative standing (like percentiles).

Finally, you discover how statisticians measure variability from sample to sample and why a measure of precision in your sample results is so important. And you get the lowdown on what some statisticians describe as the “Crowning Jewel of all Statistics”: the Central Limit Theorem (CLT). I don’t use quite this level of flourishing language to describe the CLT; I just tell my students it’s an MDR (“Mighty Deep Result”; coined by my PhD adviser). As for how my students describe their feelings about the CLT, I’ll leave that to your imagination.

## Part 4: Guesstimating and Hypothesizing with Confidence

This part focuses on the two methods for taking the results from a sample and generalizing them to make conclusions about an entire population. (Statisticians call this process *statistical inference*.) These two methods are confidence intervals and hypothesis tests.

In this part, you use confidence intervals to come up with good estimates for one or two population means or proportions, or for the difference between them (for example, the average number of hours teenagers spend watching TV per week or the percentage of men versus women in the United States who take arthritis

medicine every day). You get the nitty-gritty on how confidence intervals are formed, interpreted, and evaluated for correctness and credibility. You explore the factors that influence the width of a confidence interval (such as sample size) and work through formulas, step-by-step calculations, and examples for the most commonly used confidence intervals.

The hypothesis tests in this part show you how to use your data to test someone's claim about one or two population means or proportions, or the difference between them. (For example, a company claims their packages are delivered in two days on average — is this true?) You discover how researchers (should) go about forming and testing hypotheses and how you can evaluate their results for accuracy and credibility. You also get detailed step-by-step directions and examples for carrying out and interpreting the results of the most commonly used hypothesis tests.

## **Part 5: Statistical Studies and the Hunt for a Meaningful Relationship**

This part gives an overview of surveys, experiments, and observational studies. You find out what these studies do, how they are conducted, what their limitations are, and how to evaluate them to determine whether you should believe the results.

You also get all the details on how to examine pairs of numerical variables and categorical variables to look for relationships; this is the object of a great number of studies. For pairs of categorical variables, you create two-way tables and find joint, conditional, and marginal probabilities and distributions. You check for independence, and if a dependent relationship is found, you describe the nature of the relationship using probabilities. For numerical variables you create scatter-plots, find and interpret correlation, perform regression analyses, study the fit of the regression line and the impact of outliers, describe the relationship using the slope, and use the line to make predictions. All in a day's work!

## **Part 6: The Part of Tens**

This quick and easy part shares ten ways to be a statistically savvy sleuth and root out suspicious studies and results, as well as ten surefire ways to boost your statistics exam score.

Some statistical calculations involve the use of statistical tables, and I provide quick and easy access to all the tables you need for this book in the appendix. These tables are the  $Z$ -table (for the standard normal, also called the  $Z$ -distribution), the  $t$ -table (for the  $t$ -distribution), and the binomial table (for — you guessed it — the binomial distribution). Instructions and examples for using these three tables are provided in their corresponding sections of this book.

# Icons Used in This Book

Icons are used in this book to draw your attention to certain features that occur on a regular basis. Here's what they mean:



TIP

This icon refers to helpful hints, ideas, or shortcuts that you can use to save time. It also highlights alternative ways to think about a particular concept.



REMEMBER

This icon is reserved for particular ideas that I hope you'll remember long after you read this book.



WARNING

This icon refers to specific ways that researchers or the media can mislead you with statistics and tells you what you can do about it. It also points out potential problems and cautions to keep an eye out for on exams.



TECHNICAL  
STUFF

This icon is a sure bet if you have a special interest in understanding the more technical aspects of statistical issues. You can skip this icon if you don't want to get into the gory details.

## Where to Go from Here

This book is written in such a way that you can start anywhere and still be able to understand what's going on. So you can take a peek at the table of contents or the index, look up the information that interests you, and flip to the page listed. However if you have a specific topic in mind and are eager to dive into it, here are some directions:

- »» To work on finding and interpreting graphs, charts, means or medians, and the like, head to Part 2.
- »» To find info on the normal,  $Z$ -,  $t$ -, or binomial distributions or the Central Limit Theorem, see Part 3.
- »» To focus on confidence intervals and hypothesis tests of all shapes and sizes, flip to Part 4.
- »» To delve into surveys, experiments, regression, and two-way tables, see Part 5.

Or if you aren't sure where you want to start, you may just go with Chapter 1 for the big picture and then plow your way through the rest of the book. Happy reading!

# **1**

# **Vital Statistics about Statistics**

## IN THIS PART . . .

When you turn on the TV or open a newspaper, you're bombarded with numbers, charts, graphs, and statistical results. From today's poll to the latest major medical breakthroughs, the numbers just keep coming. Yet much of the statistical information you're asked to consume is actually wrong — by accident or even by design. How is a person to know what to believe? By doing a lot of good detective work.

This part helps awaken the statistical sleuth that lies within you by exploring how statistics affect your everyday life and your job, how bad much of the information out there really is, and what you can do about it. This part also helps you get up to speed with some useful statistical jargon.



Finding out what the process of statistics is all about

Gaining success with statistics in your everyday life, your career, and in the classroom

# Chapter 1

# Statistics in a Nutshell

**T**he world today is overflowing with data to the point where anyone (even me!) can be overwhelmed. I wouldn't blame you if you were cynical right now about statistics you read about in the media — I am too at times. The good news is that while a great deal of misleading and incorrect information is lying out there waiting for you, a lot of great stuff is also being produced; for example, many studies and techniques involving data are helping improve the quality of our lives. Your job is to be able to sort out the good from the bad and be confident in your ability to do that. Through a strong understanding of statistics and statistical procedures, you gain power and confidence with numbers in your everyday life, in your job, and in the classroom. That's what this book is all about.

In this chapter, I give you an overview of the role statistics plays in today's data-packed society and what you can do to not only survive but thrive. You get a much broader view of statistics as a partner in the scientific method — designing effective studies, collecting good data, organizing and analyzing the information, interpreting the results, and making appropriate conclusions. (And you thought statistics was just number-crunching!)

# Thriving in a Statistical World

It's hard to get a handle on the flood of statistics that affect your daily life in large and small ways. It begins the moment you wake up in the morning and check the news and listen to the meteorologist give you her predictions for the weather based on her statistical analyses of past data and present weather conditions. You pore over nutritional information on the side of your cereal box while you eat breakfast. At work you pull numbers from charts and tables, enter data into spreadsheets, run diagnostics, take measurements, perform calculations, estimate expenses, make decisions using statistical baselines, and order inventory based on past sales data.

At lunch you go to the No. 1 restaurant based on a survey of 500 people. You eat food that was priced based on marketing data. You go to your doctor's appointment where they take your blood pressure, temperature, weight, and do a blood test; after all the information is collected, you get a report showing your numbers and how you compare to the statistical norms.

You head home in your car that's been serviced by a computer running statistical diagnostics. When you get home, you turn on the news and hear the latest crime statistics, see how the stock market performed, and discover how many people visited the zoo last week.

At night, you brush your teeth with toothpaste that's been statistically proven to fight cavities, read a few pages of your *New York Times* Best-Seller (based on statistical sales estimates), and go to sleep — only to start it all over again the next morning. But how can you be sure that all those statistics you encounter and depend on each day are correct? In Chapter 2, I discuss in more depth a few examples of how statistics is involved in our lives and workplaces, what its impact is, and how you can raise your awareness of it.



WARNING

Some statistics are vague, inappropriate, or just plain wrong. You need to become more aware of the statistics you encounter each day and train your mind to stop and say “wait a minute!”, sift through the information, ask questions, and raise red flags when something's not quite right. In Chapter 3, you see ways in which you can be misled by bad statistics and develop skills to think critically and identify problems before automatically believing results.

Like any other field, statistics has its own set of jargon, and I outline and explain some of the most commonly used statistical terms in Chapter 4. Knowing the language increases your ability to understand and communicate statistics at a higher level without being intimidated. It raises your credibility when you use precise terms to describe what's wrong with a statistical result (and why). And your presentations involving statistical tables, graphs, charts, and analyses will be

informational and effective. (Heck, if nothing else, you need the jargon because I use it throughout this book; don't worry though, I always review it.)

In the next sections, you see how statistics is involved in each phase of the scientific method.

## Designing Appropriate Studies

Everyone's asking questions, from drug companies to biologists; from marketing analysts to the U.S. government. And ultimately, everyone will use statistics to help them answer their questions. In particular, many medical and psychological studies are done because someone wants to know the answer to a question. For example,

- » Will this vaccine be effective in preventing the flu?
- » What do Americans think about the state of the economy?
- » Does an increase in the use of social networking Web sites cause depression in teenagers?

The first step after a research question has been formed is to design an effective study to collect data that will help answer that question. This step amounts to figuring out what process you'll use to get the data you need. In this section, I give an overview of the two major types of studies — surveys and experiments — and explore why it's so important to evaluate how a study was designed before you believe the results.

### Surveys

An *observational study* is one in which data is collected on individuals in a way that doesn't affect them. The most common observational study is the survey. *Surveys* are questionnaires that are presented to individuals who have been selected from a population of interest. Surveys take many different forms: paper surveys sent through the mail, questionnaires on Web sites, call-in polls conducted by TV networks, phone surveys, and so on.



WARNING

If conducted properly, surveys can be very useful tools for getting information. However, if not conducted properly, surveys can result in bogus information. Some problems include improper wording of questions, which can be misleading, lack of response by people who were selected to participate, or failure to include an entire group of the population. These potential problems mean a survey has to be well thought out before it's given.



REMEMBER

Many researchers spend a great deal of time and money to do good surveys, and you'll know (by the criteria I discuss in Chapter 16) that you can trust them. However, as you are besieged with so many different types of surveys found in the media, in the workplace, and in many of your classes, you need to be able to quickly examine and critique how a survey was designed and conducted and be able to point out specific problems in a well-informed way. The tools you need for sorting through surveys are found in Chapter 16.

## Experiments

An *experiment* imposes one or more treatments on the participants in such a way that clear comparisons can be made. After the treatments are applied, the responses are recorded. For example, to study the effect of drug dosage on blood pressure, one group may take 10 mg of the drug, and another group may take 20 mg. Typically, a control group is also involved, in which subjects each receive a fake treatment (a sugar pill, for example), or a standard, nonexperimental treatment (like the existing drugs given to AIDS patients.)



REMEMBER

Good and credible experiments are designed to minimize bias, collect lots of good data, and make appropriate comparisons (treatment group versus control group). Some potential problems that occur with experiments include researchers and/or subjects who know which treatment they got, factors not controlled for in the study that affect the outcome (such as weight of the subject when studying drug dosage), or lack of a control group (leaving no baseline to compare the results with).

But when designed correctly, an experiment can help a researcher establish a cause-and-effect relationship if the difference in responses between the treatment group and the control group is statistically significant (unlikely to have occurred just by chance).



WARNING

Experiments are credited with helping to create and test drugs, determining best practices for making and preparing foods, and evaluating whether a new treatment can cure a disease, or at least reduce its impact. Our quality of life has certainly been improved through the use of well-designed experiments. However, not all experiments are well-designed, and your ability to determine which results are credible and which results are incredible (pun intended) is critical, especially when the findings are very important to you. All the info you need to know about experiments and how to evaluate them is found in Chapter 17.

# Collecting Quality Data

After a study has been designed, be it a survey or an experiment, the individuals who will participate have to be selected, and a process must be in place to collect the data. This phase of the process is critical to producing credible data in the end, and this section hits the highlights.

## Selecting a good sample



REMEMBER

Statisticians have a saying, “Garbage in equals garbage out.” If you select your *subjects* (the individuals who will participate in your study) in a way that is *biased* — that is, favoring certain individuals or groups of individuals — then your results will also be biased. It’s that simple.

Suppose Bob wants to know the opinions of people in your city regarding a proposed casino. Bob goes to the mall with his clipboard and asks people who walk by to give their opinions. What’s wrong with that? Well, Bob is only going to get the opinions of a) people who shop at that mall; b) on that particular day; c) at that particular time; d) and who take the time to respond.

Those circumstances are too restrictive — those folks don’t represent a cross section of the city. Similarly, Bob could put up a Web site survey and ask people to use it to vote. However, only people who know about the site, have Internet access, and want to respond will give him data, and typically only those with strong opinions will go to such trouble. In the end, all Bob has is a bunch of biased data on individuals that don’t represent the city at all.



REMEMBER

To minimize bias in a survey, the key word is *random*. You need to select your sample of individuals *randomly* — that is, with some type of “draw names out of a hat” process. Scientists use a variety of methods to select individuals at random, and you see how they do it in Chapter 16.

Note that in designing an experiment, collecting a random sample of people and asking them to participate often isn’t ethical because experiments impose a treatment on the subjects. What you do is send out requests for volunteers to come to you. Then you make sure the volunteers you select from the group represent the population of interest and that the data is well collected on those individuals so the results can be projected to a larger group. You see how that’s done in Chapter 17.

After going through Chapters 16 and 17, you’ll know how to dig down and analyze others’ methods for selecting samples and even be able to design a plan you can use to select a sample. In the end, you’ll know when to say “Garbage in equals garbage out.”

## Avoiding bias in your data

*Bias* is the systematic favoritism of certain individuals or certain responses. Bias is the nemesis of statisticians, and they do everything they can to minimize it. Want an example of bias? Say you're conducting a phone survey on job satisfaction of Americans; if you call people at home during the day between 9 a.m. and 5 p.m., you miss out on everyone who works during the day. Maybe day workers are more satisfied than night workers.

You have to watch for bias when collecting survey data. For instance: Some surveys are too long — what if someone stops answering questions halfway through? Or what if they give you misinformation and tell you they make \$100,000 a year instead of \$45,000? What if they give you answers that aren't on your list of possible answers? A host of problems can occur when collecting survey data, and you need to be able to pinpoint those problems.



WARNING

Experiments are sometimes even more challenging when it comes to bias and collecting data. Suppose you want to test blood pressure; what if the instrument you're using breaks during the experiment? What if someone quits the experiment halfway through? What if something happens during the experiment to distract the subjects or the researchers? Or they can't find a vein when they have to do a blood test exactly one hour after a dose of a drug is given? These problems are just some examples of what can go wrong in data collection for experiments, and you have to be ready to look for and find these problems.

After you go through Chapter 16 (on samples and surveys) and Chapter 17 (on experiments), you'll be able to select samples and collect data in an unbiased way, being sensitive to little things that can really influence the results. And you'll have the ability to evaluate the credibility of statistical results and to be heard, because you'll know what you're talking about.

## Creating Effective Summaries

After good data have been collected, the next step is to summarize them to get a handle on the big picture. Statisticians describe data in two major ways: with numbers (called *descriptive statistics*) and with pictures (that is, charts and graphs).