

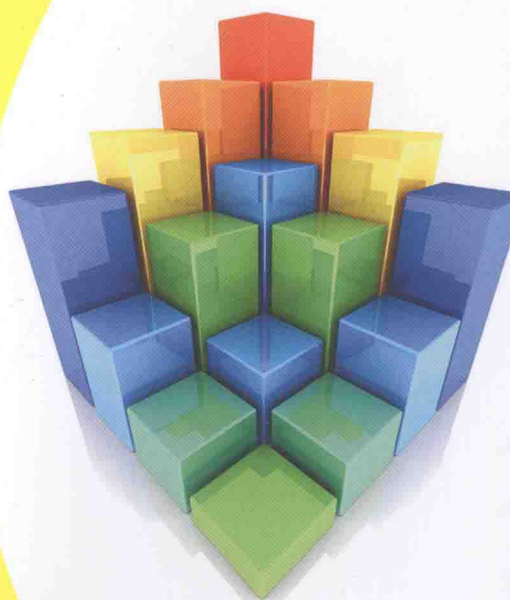
Making Everything Easier!™

Statistics Essentials

FOR
DUMMIES®

Learn:

- Exactly what you need to know about statistical ideas and techniques
- The “must-know” formulas and calculations
- More topics in quick, focused lessons



Deborah Rumsey, PhD

Auxiliary Professor and
Statistics Education Specialist,
The Ohio State University

Statistics Essentials FOR DUMMIES®

by Deborah Rumsey, PhD



WILEY

Wiley Publishing, Inc.

Statistics Essentials For Dummies®

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

Copyright © 2010 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: 2010925241

ISBN: 978-0-470-61839-4

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2



About the Author

Deborah Rumsey is a Statistics Education Specialist and Auxiliary Professor at The Ohio State University. Dr. Rumsey is a Fellow of the American Statistical Association and has won a Presidential Teaching Award from Kansas State University. She has served on the American Statistical Association's Statistics Education Executive Committee and the Advisory Committee on Teacher Enhancement, and is the editor of the Teaching Bits section of the *Journal of Statistics Education*. She is the author of the books *Statistics For Dummies*, *Statistics II For Dummies*, *Probability For Dummies*, and *Statistics Workbook For Dummies*. Her passions, besides teaching, include her family, fishing, bird watching, getting "seat time" on her Kubota tractor, and cheering the Ohio State Buckeyes to another national championship.

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

Senior Acquisitions Editor:
Lindsay Sandman Lefevere

Copy Editor: Corbin Collins

Assistant Editor: Erin Calligan Mooney

Editorial Program Coordinator:
Joe Niesen

Technical Editors: Jason J. Moliterno,
Jon-Lark Kim

Senior Editorial Manager:
Jennifer Ehrlich

Editorial Supervisor and Reprint Editor:
Carmen Krikorian

Editorial Assistants: Rachelle Amick,
Jennette ElNaggar

Senior Editorial Assistant: David Lutton

Cover Photos: © iStockphoto.com/geopaul

Cartoon: Rich Tennant
(www.the5thwave.com)

Composition Services

Project Coordinator: Patrick Redmond

Layout and Graphics: Carl Byers,
Carrie A. Cesavice, Melissa K. Smith

Proofreaders: Laura Albert,
Jennifer Theriot

Indexer: Potomac Indexing, LLC

Publishing and Editorial for Consumer Dummies

Diane Graves Steele, Vice President and Publisher, Consumer Dummies

Kristin Ferguson-Wagstaffe, Product Development Director, Consumer Dummies

Ensley Eikenburg, Associate Publisher, Travel

Kelly Regan, Editorial Director, Travel

Publishing for Technology Dummies

Andy Cummings, Vice President and Publisher, Dummies Technology/
General User

Composition Services

Debbie Stailey, Director of Composition Services

Contents at a Glance

Introduction	1
Chapter 1: Statistics in a Nutshell	5
Chapter 2: Descriptive Statistics	13
Chapter 3: Charts and Graphs	23
Chapter 4: The Binomial Distribution	35
Chapter 5: The Normal Distribution	45
Chapter 6: Sampling Distributions and the Central Limit Theorem	55
Chapter 7: Confidence Intervals	69
Chapter 8: Hypothesis Tests	87
Chapter 9: The t -distribution	107
Chapter 10: Correlation and Regression	113
Chapter 11: Two-Way Tables	127
Chapter 12: A Checklist for Samples and Surveys	137
Chapter 13: A Checklist for Judging Experiments	147
Chapter 14: Ten Common Statistical Mistakes	155
Appendix: Tables for Reference	163
Index	171

Table of Contents

Introduction	1
About This Book	1
Conventions Used in This Book	2
Foolish Assumptions	2
Icons Used in This Book	3
Where to Go from Here	3
Chapter 1: Statistics in a Nutshell	5
Designing Studies	5
Surveys	5
Experiments	6
Collecting Data	7
Selecting a good sample	7
Avoiding bias in your data	8
Describing Data	8
Descriptive statistics	8
Charts and graphs	9
Analyzing Data	10
Making Conclusions	10
Chapter 2: Descriptive Statistics	13
Types of Data	13
Counts and Percents	14
Measures of Center	15
Measures of Variability	17
Percentiles	19
Finding a percentile	19
Interpreting percentiles	20
The Five-Number Summary	21
Chapter 3: Charts and Graphs	23
Pie Charts	23
Bar Graphs	24
Time Charts	26

Histograms	27
Making a histogram	27
Interpreting a histogram	29
The distribution of the data in a histogram	29
Variability in the data from a histogram	29
Center of the data from a histogram	30
Evaluating a histogram	30
Boxplots	31
Making a boxplot	31
Interpreting a boxplot	32
Distribution of data in a boxplot	32
Variability in a data set from a boxplot	34
Center of the data from a boxplot	34
Chapter 4: The Binomial Distribution	35
Characteristics of a Binomial	35
Checking the binomial conditions step by step	36
Non-binomial examples	36
No fixed number of trials	37
More than success or failure	37
Probability of success (p) changes	37
Trials are not independent	38
Finding Binomial Probabilities Using the Formula	38
Finding Probabilities Using the Binomial Table	40
Finding probabilities when $p \leq 0.50$	40
Finding probabilities when $p > 0.50$	41
Finding probabilities for X greater-than, less-than, or between two values	42
The Expected Value and Variance of the Binomial	43
Chapter 5: The Normal Distribution	45
Basics of the Normal Distribution	45
The Standard Normal (Z) Distribution	46
Finding Probabilities for X	48
Finding X for a Given Probability	51
Normal Approximation to the Binomial	53
Chapter 6: Sampling Distributions and the Central Limit Theorem	55
Sampling Distributions	55
The mean of sampling distribution	57
The standard error of a sampling distribution	57

Sample size and standard error.....	58
Population standard deviation and standard error.....	60
The shape	61
Case 1: Distribution of X is normal.....	61
Case 2: Distribution of X is unknown or not normal	61
Finding Probabilities for \bar{X}	62
The Sampling Distribution of the Sample Proportion	63
What proportion of students need math help?.....	64
Finding Probabilities for \hat{p}	66
Chapter 7: Confidence Intervals	69
Making Your Best Guesstimate	69
The Goal: Small Margin of Error.....	71
Choosing a Confidence Level	71
Factoring In the Sample Size.....	73
Counting On Population Variability	75
Confidence Interval for a Population Mean	75
Confidence Interval for a Population Proportion	77
Confidence Interval for the Difference of Two Means.....	78
Confidence Interval for the Difference of Two Proportions	80
Interpreting Confidence Intervals	82
Spotting Misleading Confidence Intervals	84
Chapter 8: Hypothesis Tests	87
Doing a Hypothesis Test	87
Identifying what you're testing	88
Setting up the hypotheses	88
What's the alternative?	88
Knowing which hypothesis is which	89
Finding sample statistics	90
Standardizing the evidence: the test statistic.....	90
Weighing the evidence and making decisions: <p>p-values</p> <p> Finding the p-value</p> <p> Interpreting a p-value.....</p> <p> General steps for a hypothesis test</p>	91
Finding the p -value	92
Interpreting a p -value.....	93
General steps for a hypothesis test	94
Testing One Population Mean	94
Testing One Population Proportion	96
Comparing Two Population Means	97
Testing the Mean Difference: Paired Data	99
Testing Two Population Proportions	102



You Could Be Wrong: Errors in Hypothesis Testing	104
A false alarm: Type-1 error	105
A missed detection: Type-2 error	105
Chapter 9: The <i>t</i>-distribution	107
Basics of the <i>t</i> -Distribution	107
Understanding the <i>t</i> -Table	108
<i>t</i> -distributions and Hypothesis Tests	109
Finding critical values	110
Finding <i>p</i> -values	110
<i>t</i> -distributions and Confidence Intervals.....	112
Chapter 10: Correlation and Regression	113
Picturing the Relationship with a Scatterplot	113
Making a scatterplot.....	114
Interpreting a scatterplot	114
Measuring Relationships Using the Correlation	115
Calculating the correlation	116
Interpreting the correlation	117
Properties of the correlation	118
Finding the Regression Line	119
Which is <i>X</i> and which is <i>Y</i> ?	119
Checking the conditions	119
Understanding the equation.....	120
Finding the slope	121
Finding the <i>y</i> -intercept	121
Interpreting the slope and <i>y</i> -intercept	122
Interpreting the slope	122
Interpreting the <i>y</i> -intercept.....	123
The best-fitting line for the crickets.....	123
Making Predictions	124
Avoid Extrapolation!.....	125
Correlation Doesn't Necessarily Mean Cause-and-Effect.....	125
Chapter 11: Two-Way Tables	127
Organizing and Interpreting a Two-way Table.....	127
Defining the outcomes	128
Setting up the rows and columns	128
Inserting the numbers.....	129
Finding the row, column, and grand totals	130
Finding Probabilities within a Two-Way Table	131
Figuring joint probabilities	131
Calculating marginal probabilities	131
Finding conditional probabilities	132
Checking for Independence	134

Chapter 12: A Checklist for Samples and Surveys . . . 137

The Target Population is Well Defined	138
The Sample Matches the Target Population	138
The Sample Is Randomly Selected	139
The Sample Size Is Large Enough	139
Nonresponse Is Minimized	140
The importance of following up	140
Anonymity versus confidentiality	141
The Survey Is of the Right Type	142
Questions Are Well Worded	142
The Timing Is Appropriate	143
Personnel Are Well Trained	143
Proper Conclusions Are Made	144

Chapter 13: A Checklist for Judging Experiments. . . 147

Experiments versus Observational Studies	147
Criteria for a Good Experiment	148
Inspect the Sample Size	148
Small samples - small conclusions	148
Original versus final sample size	149
Examine the Subjects	149
Check for Random Assignments	150
Gauge the Placebo Effect	150
Identify Confounding Variables	151
Assess Data Quality	152
Check Out the Analysis	152
Scrutinize the Conclusions	153
Overstated results	153
Ad-hoc explanations	154
Generalizing beyond the scope	154

Chapter 14: Ten Common Statistical Mistakes 155

Misleading Graphs	155
Pie charts	155
Bar graphs	156
Time charts	156
Histograms	157
Biased Data	157
No Margin of Error	158
Nonrandom Samples	158
Missing Sample Sizes	159
Misinterpreted Correlations	159
Confounding Variables	160

Botched Numbers	160
Selectively Reporting Results	161
The Almighty Anecdote.....	162
Appendix: Tables for Reference.....	163
Index	171

Introduction

This book is designed to give you the essential, nitty-gritty information typically covered in a first semester statistics course. It's bottom-line information for you to use as a refresher, a resource, a quick reference, and/or a study guide. It helps you decipher and make important decisions about statistical polls, experiments, reports and headlines with confidence, being ever aware of the ways people can mislead you with statistics, and how to handle it.

Topics I work you through include graphs and charts, descriptive statistics, the binomial, normal, and t -distributions, two-way tables, simple linear regression, confidence intervals, hypothesis tests, surveys, experiments, and of course the most frustrating yet critical of all statistical topics: sampling distributions and the Central Limit Theorem.

About This Book

This book departs from traditional statistics texts and reference/supplement books and study guides in these ways:

- ✓ **Clear and concise step-by-step procedures** that intuitively explain how to work through statistics problems and remember the process.
- ✓ **Focused, intuitive explanations** empower you to know you're doing things right and whether others do it wrong.
- ✓ **Nonlinear approach** so you can quickly zoom in on that concept or technique you need, without having to read other material first.
- ✓ **Easy-to-follow examples** reinforce your understanding and help you immediately see how to apply the concepts in practical settings.
- ✓ **Understandable language** helps you remember and put into practice essential statistical concepts and techniques.

Conventions Used in This Book

I refer to statistics in two different ways: as numerical results (such as means and medians); or as a field of study (for example, “Statistics is all about data.”).

The second convention refers to the word *data*. I’m going to go with the plural version of the word data in this book. For example “data are collected during the experiment” — not “data is collected during the experiment.”

Foolish Assumptions

I assume you’ve had some (not necessarily a lot of) previous experience with statistics somewhere in your past. For example, you can recognize some of the basic statistics such as the mean, median, standard deviation, and perhaps correlation; you can handle some graphs; and you can remember having seen the normal distribution. If it’s been a while and you are a bit rusty, that’s okay; this book is just the thing to jog your memory.

If you have very limited or no prior experience with statistics, allow me to suggest my full-version book, *Statistics for Dummies*, to build up your foundational knowledge base. But if you are someone who has not seen these ideas before and either doesn’t have time for the full version, or you like to plunge into details right away, this book can work for you.

I assume you’ve had a basic algebra background and can do some of the basic mathematical operations and understand some of the basic notation used in algebra like x , y , summation signs, taking the square root, squaring a number, and so on. (If you’d like some backup on the algebra part, I suggest you consider *Algebra I For Dummies* and *Algebra II For Dummies* (Wiley)).

Icons Used in This Book

Here are the road signs you'll encounter on your journey through this book:



Tips refer to helpful hints or shortcuts you can use to save time.



Read these to get the inside track on why a certain concept is important, what its impact will be on the results, and highlights to keep on your radar.



These alert you to common errors that can cause problems, so you can steer around them.



These point out things in the text that you should, if possible, stash away somewhere in your brain for future use.

Where to Go from Here

This book is written in a nonlinear way, so you can start anywhere and still be able to understand what's happening. However, I can make some recommendations for those who are interested in knowing where to start.

For a quick overview of the topics to refresh your memory, check out Chapter 1. For basic number crunching and graphs, see Chapters 2 and 3. If you're most interested in common distributions, see Chapters 4 (binomial); 5 (normal); and 9 (*t*-distribution). Confidence intervals and hypothesis testing are found in Chapters 7 and 8. Correlation and regression are found in Ch 10, and two-way tables and independence are tackled in Ch 11. If you are interested in evaluating and making sense of the results of medical studies, polls, surveys, and experiments, you'll find all the info in Chapters 12 and 13. Common mistakes to avoid or watch for are seen in Chapter 14.

The 5th Wave

By Rich Tennant



"Get ready, I think they're starting to drift."

Chapter 1

Statistics in a Nutshell

In This Chapter

- ▶ Getting the big picture of the field of statistics
- ▶ Overviewing the steps of the scientific method
- ▶ Seeing the role of statistics at each step

The most common description of statistics is that it's the process of analyzing data — number crunching, in a sense. But statistics is not just about analyzing the data. It's about the whole process of using the scientific method to answer questions and make decisions. That process involves designing studies, collecting good data, describing the data with numbers and graphs, analyzing the data, and then making conclusions. In this chapter I review each of these steps and show where statistics plays the all-important role.

Designing Studies

Once a research question is defined, the next step is designing a study in order to answer that question. This amounts to figuring out what process you'll use to get the data you need. In this section I overview the two major types of studies: observational studies and experiments.

Surveys

An *observational study* is one in which data are 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 on many