

LEARNING MADE EASY



2nd Edition

# Statistics II

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



Learn to analyze  
big data sets

Explore important intermediate  
statistical techniques

Use statistical software for  
real-world applications

**Deborah J. Rumsey, PhD**

Associated Professor of Statistics  
The Ohio State University



# Statistics II

2nd Edition

by Deborah J. Rumsey, PhD

**for  
dummies**  
A Wiley Brand

## **Statistics II For Dummies®**, 2nd Edition

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, [www.wiley.com](http://www.wiley.com)

Copyright © 2022 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

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

<p>LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY 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 ANY IMPLIED</p>
--

WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

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

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included

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

Library of Congress Control Number: 2021946535

ISBN 978-1-119-82739-9 (pbk); ISBN 978-1-119-82740-5 (ebk); ISBN 978-1-119-82741-2 (ebk)

# Statistics II For Dummies®

To view this book's Cheat Sheet, simply go to [www.dummies.com](http://www.dummies.com) and search for “Statistics II For Dummies Cheat Sheet” in the Search box.

## Table of Contents

[Cover](#)

[Title Page](#)

[Copyright](#)

[Introduction](#)

[About This Book](#)

[Foolish Assumptions](#)

[Icons Used in This Book](#)

[Beyond the Book](#)

[Where to Go from Here](#)

[\*\*Part 1: Tackling Data Analysis and Model-Building Basics\*\*](#)

[\*\*Chapter 1: Beyond Number Crunching: The Art and Science of Data Analysis\*\*](#)

[Data Analysis: Looking before You Crunch](#)

[Getting the Big Picture: An Overview of Stats II](#)

[\*\*Chapter 2: Finding the Right Analysis for the Job\*\*](#)

[Categorical versus Quantitative Variables](#)

[Statistics for Categorical Variables](#)

[Statistics for Quantitative Variables](#)

[Avoiding Bias](#)

[Measuring Precision with Margin of Error](#)

[Knowing Your Limitations](#)

## **Chapter 3: Having the Normal and Sampling Distributions in Your Back Pocket**

[Recognizing the VIP Distribution — the Normal](#)

[Finally Getting Comfortable with Sampling Distributions](#)

[Heads Up! Building Confidence Intervals and Hypothesis Tests](#)

## **Chapter 4: Reviewing Confidence Intervals and Hypothesis Tests**

[Estimating Parameters by Using Confidence Intervals](#)

[What's the Hype about Hypothesis Tests?](#)

## **Part 2: Using Different Types of Regression to Make Predictions**

### **Chapter 5: Getting in Line with Simple Linear Regression**

[Exploring Relationships with Scatterplots and Correlations](#)

[Building a Simple Linear Regression Model](#)

[No Conclusion Left Behind: Tests and Confidence Intervals for Regression](#)

[Checking the Model's Fit \(The Data, Not the Clothes!\)](#)

[Knowing the Limitations of Your Regression Analysis](#)

### **Chapter 6: Multiple Regression with Two X Variables**

[Getting to Know the Multiple Regression Model](#)

[Looking at x's and y's](#)

[Collecting the Data](#)

[Pinpointing Possible Relationships](#)

[Checking for Multicollinearity](#)

[Finding the Best-Fitting Model for Two x Variables](#)

[Predicting y by Using the x Variables](#)

[Checking the Fit of the Multiple Regression Model](#)

## **Chapter 7: How Can I Miss You If You Won't Leave? Regression Model Selection**

[Getting a Kick out of Estimating Punt Distance](#)

[Just Like Buying Shoes: The Model Looks Nice, But Does It Fit?](#)

## **Chapter 8: Getting Ahead of the Learning Curve with Nonlinear Regression**

[Anticipating Nonlinear Regression](#)

[Starting Out with Scatterplots](#)

[Handling Curves in the Road with Polynomials](#)

[Going Up? Going Down? Go Exponential!](#)

## **Chapter 9: Yes, No, Maybe So: Making Predictions by Using Logistic Regression**

[Understanding a Logistic Regression Model](#)

[Carrying Out a Logistic Regression Analysis](#)

## **Part 3: Analyzing Variance with ANOVA**

### **Chapter 10: Testing Lots of Means? Come On Over to ANOVA!**

[Comparing Two Means with a t-Test](#)

[Evaluating More Means with ANOVA](#)

[Checking the Conditions](#)

[Setting Up the Hypotheses](#)

[Doing the F-Test](#)

[Checking the Fit of the ANOVA Model](#)

### **Chapter 11: Sorting Out the Means with Multiple Comparisons**

[Following Up after ANOVA](#)

[Pinpointing Differing Means with Fisher and Tukey](#)

[Examining the Output to Determine the Analysis](#)

[So Many Other Procedures, So Little Time!](#)



## **Chapter 12: Finding Your Way through Two-Way ANOVA**

[Setting Up the Two-Way ANOVA Model](#)

[Understanding Interaction Effects](#)

[Testing the Terms in Two-Way ANOVA](#)

[Running the Two-Way ANOVA Table](#)

[Are Whites Whiter in Hot Water? Two-Way ANOVA Investigates](#)

## **Chapter 13: Regression and ANOVA: Surprise Relatives!**

[Seeing Regression through the Eyes of Variation](#)

[Regression and ANOVA: A Meeting of the Models](#)

## **Part 4: Building Strong Connections with Chi-Square Tests and Nonparametrics**

### **Chapter 14: Forming Associations with Two-Way Tables**

[Breaking Down a Two-Way Table](#)

[Breaking Down the Probabilities](#)

[Trying To Be Independent](#)

[Demystifying Simpson's Paradox](#)

### **Chapter 15: Being Independent Enough for the Chi-Square Test**

[The Chi-Square Test for Independence](#)

[Comparing Two Tests for Comparing Two Proportions](#)

### **Chapter 16: Using Chi-Square Tests for Goodness-of-Fit (Your Data, Not Your Jeans)**

[Finding the Goodness-of-Fit Statistic](#)

[Interpreting the Goodness-of-Fit Statistic Using a Chi-Square](#)

### **Chapter 17: Rebels Without a Distribution — Nonparametric Procedures**

[Arguing for Nonparametric Statistics](#)

[Mastering the Basics of Nonparametric Statistics](#)

## **Chapter 18: All Signs Point to the Sign Test**

[Reading the Signs: The Sign Test](#)

## **Part 5: Putting it All Together: Multi-Stage Analysis of a Large Data Set**

### **Chapter 19: Conducting a Multi-Stage Analysis of a Large Data Set**

[Steps Involved in Working with a Large Data Set](#)

[Wrangling Data](#)

[Visualizing Data](#)

[Exploring the Data](#)

[Looking for Relationships](#)

[Building Models and Making Inferences](#)

[Sharing the Story](#)

### **Chapter 20: A Statistician Watches the Movies**

[Examining the Movie Variables and Asking Questions](#)

[Visualizing the Movie Data](#)

[Doing Descriptive Dirty Work](#)

[Looking for Relationships](#)

[Building a Model for Predicting U.S. Revenue](#)

[Writing It Up](#)

### **Chapter 21: Looking Inside the Refrigerator**

[Refrigerator Data — The Variables](#)

[Exploring the Data](#)

[Analyzing the Data](#)

[Writing It Up](#)

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

### **Chapter 22: Ten Common Errors in Statistical Conclusions**

[Claiming These Statistics Prove ...](#)

[It's Not Technically Statistically Significant, But ...](#)

[Concluding That x Causes y](#)

[Assuming the Data Was Normal](#)

[Only Reporting “Important” Results](#)

[Assuming a Bigger Sample Is Always Better](#)

[It’s Not Technically Random, But ...](#)

[Assuming That 1,000 Responses Is 1,000 Responses](#)

[Of Course the Results Apply to the General Population](#)

[Deciding Just to Leave It Out](#)

## **Chapter 23: Ten Ways to Get Ahead by Knowing Statistics**

[Asking the Right Questions](#)

[Being Skeptical](#)

[Collecting and Analyzing Data Correctly](#)

[Calling for Help](#)

[Retracing Someone Else’s Steps](#)

[Putting the Pieces Together](#)

[Checking Your Answers](#)

[Explaining the Output](#)

[Making Convincing Recommendations](#)

[Establishing Yourself as the Statistics Go-To Person](#)

## **Chapter 24: Ten Cool Jobs That Use Statistics**

[Pollster](#)

[Data Scientist](#)

[Ornithologist \(Bird Watcher\)](#)

[Sportscaster or Sportswriter](#)

[Journalist](#)

[Crime Fighter](#)

[Medical Professional](#)

[Marketing Executive](#)

[Lawyer](#)

## **Appendix A: Reference Tables**

[t-Table](#)

[Binomial Table](#)

[Chi-Square Table](#)

[F-Table](#)

[Z-Table](#)

[Index](#)

[About the Author](#)

[Advertisement Page](#)

[Connect with Dummies](#)

[End User License Agreement](#)

## List of Tables

### Chapter 1

[TABLE 1-1 Table Setup for the Breakfast and Age Group Question](#)

### Chapter 2

[TABLE 2-1 Gender and Political Affiliation of 56,735 U.S. Voters](#)

### Chapter 4

[TABLE 4-1 Comparing the Values of Power for  \$n = 10\$  versus  \$n = 100\$  \( \$H\_0\$  is  \$\mu = 0\$ \).](#)

### Chapter 5

[TABLE 5-1 Average Textbook Weight and Student Weight \(Grades 1-12\).](#)

### Chapter 6

[TABLE 6-1 Advertising Dollars and Sales of Plasma TVs](#)

### Chapter 7

[TABLE 7-1 Data Collected for Punt Distance Study.](#)

[TABLE 7-2 Correlations between Distance of a Punt and Other Variables](#)

### Chapter 8

[TABLE 8-1 Number of People Knowing a Secret over a 6-Day Period](#)

[TABLE 8-2 Log\( \$y\$ \) Values for the Secret-Spreading Data](#)

## **Chapter 9**

[TABLE 9-1 Movie Enjoyment \(Yes or No Data\) Based on Age](#)

## **Chapter 10**

[TABLE 10-1 Watermelon Seed-Spitting Distances for Four Age Groups of Children \(M...](#)

[TABLE 10-2 Comparing Variances of Four Data Sets to Check the Equal Variance Con...](#)

## **Chapter 11**

[TABLE 11-1 Cellphone Minutes Used in One Month](#)

## **Chapter 12**

[TABLE 12-1 Six Treatment Combinations for the Drug Study Example](#)

[TABLE 12-2 Data and Its Respective Levels from Two Factors](#)

## **Chapter 13**

[TABLE 13-1 First Ten Observations from the Education and Internet Use Example](#)

## **Chapter 14**

[TABLE 14-1 Two-Way Table for the Cellphone Data](#)

[TABLE 14-2 Completed Two-Way Table for the Cellphone Data](#)

[TABLE 14-3 Marginal and Grand Totals for the Cellphone Data](#)

[TABLE 14-4 Confirmation Calls Related to Showing Up for the Appointment](#)

[TABLE 14-5 Video Games Won and Lost for Males Versus Females](#)

[TABLE 14-6 Three-Way Table for Gender, Game Level, and Game Outcome](#)

## **Chapter 15**

[TABLE 15-1 Gender and House Paint Color Preference: Observed Cell Counts](#)

[TABLE 15-2 Gender and House Paint Color Preference: Expected Cell Counts](#)

[TABLE 15-3 Age Group and Opinion on Cellphone Ban: Observed Cell Counts](#)

[TABLE 15-4 Age Group and Opinion on Cellphone Ban: Expected Cell Counts](#)

## Chapter 16

[TABLE 16-1 Expected Percentage of Each Color of M&M'S Milk Chocolate Candies](#)

[TABLE 16-2 Percentage of M&M'S Observed in One Bag \(1.69 oz.\) versus Percentage ...](#)

[TABLE 16-3 Goodness-of-Fit Statistic for M&M'S Example](#)

## Chapter 18

[TABLE 18-1 Sample of House Prices in a Neighborhood](#)

[TABLE 18-2 Binomial Probabilities to Help Calculate a Confidence Interval for th...](#)

[TABLE 18-3 Testing the Effectiveness of Chewing Gum in Lowering Test Anxiety](#)

## Chapter 20

[TABLE 20-1 The Top Eight Movies of 2018 in Terms of U.S. Box Office Revenue](#)

[TABLE 20-2 Some Descriptive Statistics for Runtime through Audience Ratings](#)

## Chapter 21

[TABLE 21-1 Descriptive Statistics for 41 Refrigerators](#)

## Appendix 1

[TABLE A-1 The  \$t\$ -Table](#)

[TABLE A-2 The Binomial Table](#)

[TABLE A-3 The Chi-Square Table](#)

[TABLE A-4 The  \$F\$ -Table \( \$\alpha = 0.05\$ \)](#)

[TABLE A-5 The  \$Z\$ -Table](#)

# List of Illustrations

## Chapter 1

[FIGURE 1-1: Coffees sold at various air temperatures on football game day.](#)

## Chapter 2

[FIGURE 2-1: The two-dimensional scatterplot helps you look for relationships in...](#)

## Chapter 3

[FIGURE 3-1: Three normal distributions, with means and standard deviations of a...](#)

## Chapter 4

[FIGURE 4-1: Power curve for  \$H\_0: \mu = 0\$  versus  \$H\_a: \mu > 0\$ , for  \$n = 10\$  and  \$\sigma = 2\$ .](#)

[FIGURE 4-2: Power curve for  \$H\_0: \mu = 0\$  versus  \$H\_a: \mu > 0\$ , for  \$n = 100\$  and  \$\sigma = 2\$ .](#)

## Chapter 5

[FIGURE 5-1: Scatterplot of average student weight versus average textbook weigh...](#)

[FIGURE 5-2: Simple linear regression analysis for the textbook-weight example.](#)

[FIGURE 5-3: Regression analysis for study time and quiz score data.](#)

[FIGURE 5-4: Prediction interval of textbook weight for a 100-pound child.](#)

[FIGURE 5-5: Conditions of a simple linear regression model.](#)

[FIGURE 5-6: Standardized residual plots for textbook-weight data.](#)

[FIGURE 5-7: Scatterplot of textbook-weight data minus the outlier.](#)

[FIGURE 5-8: Residual plots for textbook-weight data minus the outlier.](#)

## Chapter 6

[FIGURE 6-1: Scatterplots of a\) TV ad spending and b\) Internet ad spending versu...](#)

[FIGURE 6-2: Correlation values and  \$p\$ -values for the TV sales example.](#)

[FIGURE 6-3: \(Selected\) Regression output for the ads and plasma TV sales exampl...](#)

[FIGURE 6-4: Residual plots for the ads and plasma TV sales example.](#)

[FIGURE 6-5: Standardized residual plots for the ads and plasma TV example.](#)

## Chapter 7

[FIGURE 7-1: A matrix of all scatterplots between pairs of variables in the punt...](#)

[FIGURE 7-2: Best subsets procedure results for the punt distance example.](#)

## **Chapter 8**

[FIGURE 8-1: Learning curve for time performing a new task.](#)

[FIGURE 8-2: A scatterplot showing the spread of a secret over a six-day period.](#)

[FIGURE 8-3: Example of a second-degree polynomial.](#)

[FIGURE 8-4: Example of a third-degree polynomial.](#)

[FIGURE 8-5: Example of a fourth-degree polynomial.](#)

[FIGURE 8-6: Scatterplot showing study time and quiz scores.](#)

[FIGURE 8-7: Trying to fit a straight line to quadratic data.](#)

[FIGURE 8-8: Minitab output for fitting a parabola to the quiz-score data.](#)

[FIGURE 8-9: The parabola appears to fit the quiz-score data nicely.](#)

[FIGURE 8-10: Standardized residual plots for the quiz-score data, using the qua...](#)

[FIGURE 8-11: The exponential regression model for different values of  \$\beta\$ .](#)

[FIGURE 8-12: Minitab fits a line to the  \$\log\(y\)\$  for the secret-spreading data.](#)

[FIGURE 8-13: A scatterplot showing the fit of a straight line to  \$\log\(y\)\$  data.](#)

[FIGURE 8-14: Residual plots showing the fit of a straight line to  \$\log\(y\)\$  data.](#)

## **Chapter 9**

[FIGURE 9-1: Two basic types of S-curves.](#)

[FIGURE 9-2: The model-building part of the movie and age data's logistic regres...](#)

[FIGURE 9-3: The best-fitting S-curve for the movie and age data.](#)

[FIGURE 9-4: The model-fitting part of the movie and age data's logistic regres...](#)

[FIGURE 9-5: Actual observed values \(0 and 1\) compared to the model.](#)

## **Chapter 10**

[FIGURE 10-1: A  \$t\$ -test comparing mean watermelon seed-spitting distances for fem...](#)



[FIGURE 10-2: Checking ANOVA conditions by using histograms and descriptive stat...](#)

[FIGURE 10-3: Examining side-by-side boxplots to check the equal variance condit...](#)

[FIGURE 10-4: ANOVA Minitab output for the watermelon seed-spitting example.](#)

[FIGURE 10-5: F-distribution with \(3, 16\) degrees of freedom.](#)

## **Chapter 11**

[FIGURE 11-1: Basic statistics and confidence intervals for the cellphone data.](#)

[FIGURE 11-2: ANOVA results for comparing cellphone use for four age groups.](#)

[FIGURE 11-3: Output showing Fisher's LSD applied to the cellphone data.](#)

[FIGURE 11-4: Output for Tukey's test used to compare cellphone usage.](#)

[FIGURE 11-5: Multiple comparison results for the cereal example.](#)

## **Chapter 12**

[FIGURE 12-1: Five examples of the results from a two-way ANOVA with interaction...](#)

[FIGURE 12-2: ANOVA tables for the interaction plots from Figure 12-1.](#)

[FIGURE 12-3: ANOVA table for the clothing example.](#)

[FIGURE 12-4: Boxplots showing how brand of detergent and water temperature inte...](#)

## **Chapter 13**

[FIGURE 13-1: Output for simple linear regression applied to education and Inter...](#)

[FIGURE 13-2: Connecting means of populations to the slope of a line.](#)

## **Chapter 14**

[FIGURE 14-1: Pie charts comparing teen versus adult personal cellphone use.](#)

## **Chapter 15**

[FIGURE 15-1: Minitab output for the house paint color preference versus gender ...](#)

[FIGURE 15-2: Chi-square distributions with 1, 2, 4, 6, 8, and 10 degrees of fre...](#)

[FIGURE 15-3: Minitab output comparing the proportion of adult and teenage owner...](#)

[FIGURE 15-4: Minitab output testing independence of age group and latest cellph...](#)

## **Chapter 16**

[FIGURE 16-1: Chi-square distribution with 5 degrees of freedom.](#)

[FIGURE 16-2: Graph of observed versus expected counts for the M&M'S colors data...](#)

[FIGURE 16-3: Minitab output of Chi-square goodness-of-fit test on M&M'S colors ...](#)

## **Chapter 17**

[FIGURE 17-1: Symmetric and skewed histograms.](#)

## **Chapter 18**

[FIGURE 18-1: Histogram of the selling prices of ten houses.](#)

[FIGURE 18-2: Sign test for house prices conducted by Minitab.](#)

## **Chapter 20**

[FIGURE 20-1: Pie charts and bar graphs of a\) release date \(month\), b\) rating, c...](#)

[FIGURE 20-2: Histograms and boxplots of a\) runtime through h\) audience \(ratings...](#)

[FIGURE 20-3: Matrix of scatterplots of runtime through audience ratings.](#)

[FIGURE 20-4: Matrix of correlations for runtime through audience ratings.](#)

[FIGURE 20-5: Stacked bar graph to explore content rating and genre.](#)

[FIGURE 20-6: Side-by-side boxplots to look for relationships.](#)

[FIGURE 20-7: Linear regression analysis using opening weekend revenue to predic...](#)

[FIGURE 20-8: Final regression line on the scatterplot using opening weekend rev...](#)

[FIGURE 20-9: Results of best subsets regression model selection for predicting ...](#)

## **Chapter 21**

[FIGURE 21-1: Histogram of refrigerator prices.](#)

[FIGURE 21-2: Boxplots of refrigerator price by brand.](#)

[FIGURE 21-3: ANOVA for testing equality of mean prices across the refrigerator ...](#)

[FIGURE 21-4: Tukey's multiple comparisons using confidence intervals and by bra...](#)

[FIGURE 21-5: Cubic-foot capacity being used to predict refrigerator price.](#)

[FIGURE 21-6: Matrix of scatterplots for refrigerator data and matrix of correla...](#)

[FIGURE 21-7: Scatterplot and regression line for predicting yearly electric use...](#)

[FIGURE 21-8: a\) Analysis of Variance \(ANOVA\) for comparing yearly electric use ...](#)

[FIGURE 21-9: Two-factor ANOVA with brand and icemaker as the factors and yearly...](#)

[FIGURE 21-10: Interaction plot showing mean yearly electric use for each brand,...](#)

# Introduction

---

So you've gone through some of the basics of statistics. Means, medians, and standard deviations all ring a bell. You know about surveys and experiments and the basic ideas of correlation and simple regression. You've studied probability, margin of error, and a few hypothesis tests and confidence intervals. Are you ready to load your statistical toolbox with a new level of tools?

*Statistics II For Dummies*, 2nd Edition, picks up right where *Statistics For Dummies*, 2nd Edition, (John Wiley & Sons) leaves off and keeps you moving along the road of statistical ideas and techniques in a positive, step-by-step way.

The focus of *Statistics II For Dummies*, 2nd Edition, is on finding more ways of analyzing data. I provide step-by-step instructions for using techniques such as multiple regression, nonlinear regression, one-way and two-way analysis of variance (ANOVA), and Chi-square tests, and I give you some practice with big data sets, which are all the rage right now. Using these new techniques, you estimate, investigate, correlate, and congregate even more variables based on the information at hand, and you see how to put the tools together to create a great story about your data (nonfiction, I hope!).

## ***About This Book***

This book is designed for those who have completed the basic concepts of statistics through confidence intervals and hypothesis testing (found in *Statistics For Dummies*, 2nd Edition) and are ready to plow ahead to get through the final part of Stats I, or to tackle Stats II. However, I do pepper in some brief overviews of Stats I as needed,

just to remind you of what was covered and to make sure you're up to speed. For each new technique, you get an overview of when and why it's used, how to know when you need it, step-by-step directions on how to apply it, and tips and tricks from a seasoned data analyst (yours truly). Because it's very important to be able to know which method to use when, I emphasize what makes each technique distinct and what the results tell you. You will also see many applications of the techniques used in real life.

I also include interpretation of computer output for data analysis purposes. I show you how to use the software to get the results, but I focus more on how to interpret the results found in the output, because you're more likely to be interpreting this kind of information than doing the programming specifically. Because the equations and calculations can get too involved if you are solving them by hand, you often use a computer to get your results. I include instructions for using Minitab to conduct many of the calculations in this book. Most statistics teachers who cover these topics use this approach as well. (What a relief!)

This book is different from the other Stats II books in many ways. Notably, this book features the following:

- » **Full explanations of Stats II concepts.** Many statistics textbooks squeeze all the Stats II topics at the very end of their Stats I coverage; as a result, these topics tend to get condensed and presented as if they're optional. But no worries; I take the time to clearly and fully explain all the information you need to survive and thrive.
- » **Dissection of computer output.** Throughout the book, I present many examples that use statistical software to analyze the data. In each case, I present

the computer output and explain how I got it and what it means.

- » **An extensive number of examples.** I include plenty of examples to cover the many different types of problems you'll face. Some examples are short, and some are quite extensive and include multiple variables.
- » **Lots of tips, strategies, and warnings.** I share with you some trade secrets, based on my experience teaching and supporting students and grading their papers.
- » **Understandable language.** I try to keep things conversational to help you understand, remember, and put into practice statistical definitions, techniques, and processes.
- » **Clear and concise, step-by-step procedures.** In most chapters, you can find steps that intuitively explain how to work through Stats II problems — and remember how to do it on your own later on.

Throughout this book, I've used several conventions that I want you to be aware of:

- » I indicate multiplication by using a times sign, indicated by a lowered asterisk \*.
- » I indicate the null and alternative hypotheses as  $H_0$  (for the null hypothesis) and  $H_a$  (for the alternative hypothesis).
- » The statistical software package I use and display throughout the book is Minitab 18, but I simply refer to it as Minitab.
- » Whenever I introduce a new term, I *italicize* it.
- » Keywords and numbered steps appear in **boldface**.

At times I get into some of the more technical details of formulas and procedures for those individuals who may need to know about them — or just really want to get the full story. These minutiae are marked with a Technical Stuff icon. I also include sidebars along with the essential text, usually in the form of a real-life statistics example or some bonus information you may find interesting. You can feel free to skip those icons and sidebars because you won't miss any of the main information you need (but by reading them, you may just be able to impress your stats professor with your above-and-beyond knowledge of Stats II!).

## ***Foolish Assumptions***

Because this book deals with Stats II, I assume you have one previous course in introductory statistics under your belt (or at least have read *Statistics For Dummies*, 2nd Edition), with topics taking you up through the Central Limit Theorem and perhaps an introduction to confidence intervals and hypothesis tests (although I review these concepts briefly in [Chapter 4](#)). Prior experience with simple linear regression isn't necessary. Only college algebra is needed for the math details. Some experience using statistical software is also a plus, but not required.

As a student, you may be covering these topics in one of two ways: either at the tail end of your Stats I course (perhaps in a hurried way, but in some way nonetheless); or through a two-course sequence in statistics in which the topics in this book are the focus of the second course. If so, this book provides you the information you need to do well in those courses.

You may simply be interested in Stats II from an everyday point of view, or perhaps you want to add to

your understanding of studies and statistical results presented in the media. If this sounds like you, you can find plenty of real-world examples and applications of these statistical techniques in action, as well as cautions for interpreting them.

## *Icons Used in This Book*

I use icons in this book to draw your attention to certain text features that occur on a regular basis. Think of the icons as road signs that you encounter on a trip. Some signs tell you about shortcuts, and others offer more information that you may need; some signs alert you to possible warnings, while others leave you with something to remember.



**COMPUTER  
OUTPUT**

When you see this icon, it means I'm explaining how to carry out that particular data analysis using Minitab. I also explain the information you get in the computer output so you can interpret your results.



**REMEMBER**

I use this icon to reinforce certain ideas that are critical for success in Stats II, such as things I think are important to review as you prepare for an exam.





TECHNICAL  
STUFF

When you see this icon, you can skip over the information if you don't want to get into the nitty-gritty details. They exist mainly for people who have a special interest or obligation to know more about the technical aspects of certain statistical issues.



TIP

This icon points to helpful hints, ideas, or shortcuts that you can use to save time; it also includes alternative ways to think about a particular concept.



WARNING

I use warning icons to help you stay away from common misconceptions and pitfalls you may face when dealing with ideas and techniques related to Stats II.

## ***Beyond the Book***

In addition to all the great content included in the book itself, you can find even more content online. Check out this book's online Cheat Sheet on [dummies.com](http://dummies.com). It covers the major formulas needed for Statistics II. You can access it by going to [www.dummies.com](http://www.dummies.com) and then typing "Statistics II For Dummies Cheat Sheet" into the search bar.

I've also included two major data sets that are analyzed in [Chapters 20](#) and [21](#), so you can follow along with me

or do your own analysis (not required!). Go to [www.dummies.com/go/statisticsIIfd2e](http://www.dummies.com/go/statisticsIIfd2e) to access these files.

## *Where to Go from Here*

This book is written in a nonlinear way, so you can start anywhere and still understand what's happening. However, I can make some recommendations if you want some direction on where to start.

If you're thoroughly familiar with the ideas of hypothesis testing and simple linear regression, start with [Chapter 5](#) (multiple regression). Use [Chapter 1](#) if you need a reference for the jargon that statisticians use in Stats II.

If you've covered all topics up through the various types of regression (simple, multiple, nonlinear, and logistic) or a subset of those as your professor deemed important, proceed to [Chapter 10](#), the basics of analysis of variance (ANOVA).

[Chapter 15](#) is the place to begin if you want to tackle categorical (qualitative) variables before hitting the quantitative stuff. You can work with the Chi-square test there.

Nonparametric statistics are presented starting in [Chapter 17](#). Start there if you want the full details on the most common nonparametric procedures, used when you do not necessarily have an assumed distribution (for example, a normal).

If you want to see a bunch of Stats II ideas put into practice right off the bat, head to [Chapter 19](#) where I discuss a multi-stage approach to analyzing a big data set, or [Chapter 21](#), where you look into a big data set on refrigerators and see how it's analyzed in a multi-stage approach.

## **Part 1**

# **Tackling Data Analysis and Model-Building Basics**

## **IN THIS PART ...**

Understand why data analysis is both a science and an art.

Make sure you use the right type of analysis for the job.

Work with the normal and binomial distributions.

Reacquaint yourself with confidence intervals and hypothesis tests.

# Chapter 1

## Beyond Number Crunching: The Art and Science of Data Analysis

---

### IN THIS CHAPTER

- » Realizing your role as a data analyst
  - » Avoiding statistical faux pas
  - » Delving into the jargon of Stats II
- 

Because you're reading this book, you're likely familiar with the basics of statistics and you're ready to take it up a notch. That next level involves using what you know, picking up a few more tools and techniques, and finally putting it all to use to help you answer more realistic questions by using real data. In statistical terms, you're ready to enter the world of the *data analyst*.

In this chapter, you review the terms involved in statistics as they pertain to data analysis at the Stats II level. You get a glimpse of the impact that your results can have by seeing what these analysis techniques can do. You also gain insight into some of the common misuses of data analysis and their effects.

## ***Data Analysis: Looking before You Crunch***

It used to be that statisticians were the only ones who really analyzed data because the only computer programs available were very complicated to use, requiring a great deal of knowledge about statistics to set up and carry out analyses. The calculations were tedious and at times unpredictable, and they required a thorough understanding of the theories and methods behind the calculations to get correct and reliable answers.

Today, anyone who wants to analyze data can do it easily. Many user-friendly statistical software packages are made expressly for that purpose — Microsoft Excel, Minitab, and SAS are just a few. Free online programs are available, too, such as R, which helps you do just what it says — crunch your numbers and get an answer.

Each software package has its own pros and cons (and its own users and protesters). My software of choice and the one I reference throughout this book is Minitab, because it's very easy to use, the results are precise, and the software's loaded with all the data-analysis techniques used in Stats II. Although a site license for Minitab isn't cheap, the student version is available for rent for only a few bucks a semester.



**REMEMBER** The most important idea when applying statistical techniques to analyze data is to know what's going on behind the number crunching so you (not the computer) are in control of the analysis. That's why knowledge of Stats II is so critical.