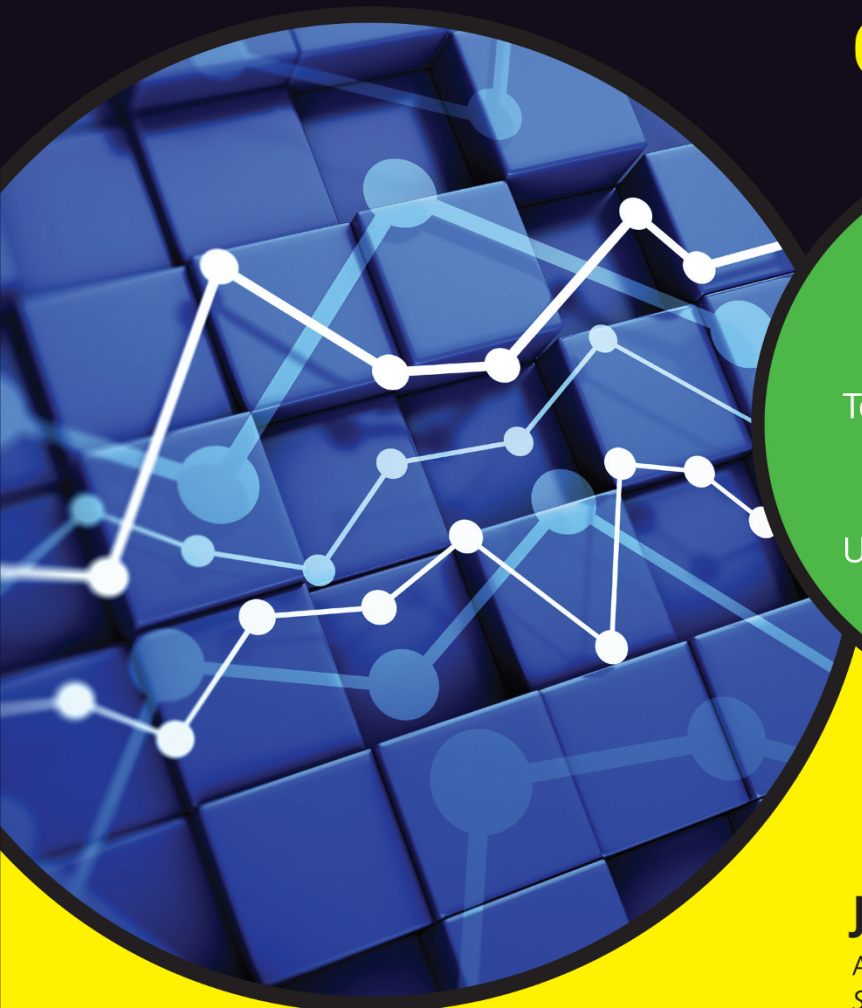
LEARNING MADE EASY

4th Edition

Statistical Analysis with Excel®

for dummies®
A Wiley Brand



Learn all of Excel's statistical tools

Test your hypotheses and draw conclusions

Use Excel to give meaning to your data

Joseph Schmuller, PhD
Author of all previous editions of
Statistical Analysis with Excel For Dummies

Statistical Analysis with Excel®

for dummies®

A Wiley Brand

Statistical Analysis with Excel®

for dummies®
A Wiley Brand

4th edition

by Joseph Schmuller, PhD

for dummies®
A Wiley Brand

Statistical Analysis with Excel® For Dummies®, 4th Edition

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

Library of Congress Control Number: 2016943716

ISBN: 978-1-119-27115-4; 978-1-119-27116-1 (ebk); 978-1-119-27117-8 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents at a Glance

Introduction 1

Part 1: Getting Started with Statistical Analysis with Excel: A Marriage Made in Heaven 7

CHAPTER 1: Evaluating Data in the Real World 9
CHAPTER 2: Understanding Excel's Statistical Capabilities 31

Part 2: Describing Data 61

CHAPTER 3: Show and Tell: Graphing Data 63
CHAPTER 4: Finding Your Center 101
CHAPTER 5: Deviating from the Average 117
CHAPTER 6: Meeting Standards and Standings 135
CHAPTER 7: Summarizing It All 151
CHAPTER 8: What's Normal? 171

Part 3: Drawing Conclusions from Data 183

CHAPTER 9: The Confidence Game: Estimation 185
CHAPTER 10: One-Sample Hypothesis Testing 199
CHAPTER 11: Two-Sample Hypothesis Testing 217
CHAPTER 12: Testing More Than Two Samples 249
CHAPTER 13: Slightly More Complicated Testing 275
CHAPTER 14: Regression: Linear and Multiple 295
CHAPTER 15: Correlation: The Rise and Fall of Relationships 331
CHAPTER 16: It's About Time 351
CHAPTER 17: Non-Parametric Statistics 363

Part 4: Probability 377

CHAPTER 18: Introducing Probability 379
CHAPTER 19: More on Probability 403
CHAPTER 20: A Career in Modeling 417

Part 5: The Part of Tens 437
CHAPTER 21: Ten Statistical and Graphical Tips and Traps 439
CHAPTER 22: Ten Things (Twelve, Actually) That Just Didn't Fit in Any Other Chapter 445
APPENDIX A: When Your Worksheet Is a Database 471
APPENDIX B: The Analysis of Covariance 487

Index 501

Table of Contents

INTRODUCTION 1
About This Book 2
What You Can Safely Skip 2
Foolish Assumptions 2
How This Book Is Organized 3
Part 1: Getting Started with Statistical Analysis with Excel: A Marriage Made in Heaven 3
Part 2: Describing Data 3
Part 3: Drawing Conclusions from Data 3
Part 4: Working with Probability 4
Part 5: The Part of Tens 4
Appendix A: When Your Worksheet Is a Database 4
Appendix B: The Analysis of Covariance 4
Bonus Appendix B1: When Your Data Live Elsewhere 5
Bonus Appendix B2: Tips for Teachers (and Learners) 5
Icons Used in This Book 5
Where to Go from Here 6

PART 1: GETTING STARTED WITH STATISTICAL ANALYSIS WITH EXCEL: A MARRIAGE MADE IN HEAVEN 7

CHAPTER 1: Evaluating Data in the Real World 9
The Statistical (and Related) Notions You Just Have to Know 9
Samples and populations 10
Variables: Dependent and independent 11
Types of data 12
A little probability 13
Inferential Statistics: Testing Hypotheses 14
Null and alternative hypotheses 15
Two types of error 16
What's New in Excel 2016? 18
What's Old in Excel 2016? 19
Knowing the Fundamentals 24
Autofilling cells 24
Referencing cells 26
What's New in This Edition? 28

CHAPTER 2: Understanding Excel's Statistical Capabilities 31
Getting Started 32
Setting Up for Statistics 34
Worksheet functions in Excel 2016 34
Quickly accessing statistical functions 37

Array functions. 40
What's in a name? An array of possibilities 43
Creating your own array formulas 51
Using data analysis tools 52
Accessing Commonly Used Functions 58

PART 2: DESCRIBING DATA 61

CHAPTER 3: Show and Tell: Graphing Data 63
Why Use Graphs?. 63
Some Fundamentals 65
Excel's Graphics (Chartics?) Capabilities. 65
Inserting a Chart 66
Becoming a Columnist 67
Stacking the Columns 71
Slicing the Pie 72
A word from the wise 74
Drawing the Line 75
Adding a Spark. 78
Passing the Bar 80
The Plot Thickens. 82
Finding Another Use for the Scatter Chart 86
Tasting the Bubbly. 87
Taking Stock 88
Scratching the Surface 90
On the Radar 91
Growing a Treemap and Bursting Some Sun 92
Building a Histogram 93
Ordering Columns: Pareto 94
Of Boxes and Whiskers. 95
3D Maps 96

CHAPTER 4: Finding Your Center 101
Means: The Lore of Averages 101
Calculating the mean 102
AVERAGE and AVERAGEA 103
AVERAGEIF and AVERAGEIFS 105
TRIMMEAN 108
Other means to an end. 110
Medians: Caught in the Middle 111
Finding the median 112
MEDIAN. 112
Statistics à la Mode 113
Finding the mode. 113
MODE.SNGL and MODE.MULT 114

CHAPTER 5: **Deviating from the Average** 117
Measuring Variation 118
Averaging squared deviations: Variance and how to calculate it 118
VAR.P and VARPA 121
Sample variance 123
VAR.S and VARA 124
Back to the Roots: Standard Deviation 124
Population standard deviation 125
STDEV.P and STDEVPA 125
Sample standard deviation 126
STDEV.S and STDEVA 126
The missing functions: STDEVIF and STDEVIFS 127
Related Functions 131
DEVSQ 131
Average deviation 132
AVEDEV 133

CHAPTER 6: **Meeting Standards and Standings** 135
Catching Some Zs 135
Characteristics of z-scores 136
Bonds versus the Bambino 137
Exam scores 137
STANDARDIZE 138
Where Do You Stand? 141
RANK.EQ and RANK.AVG 141
LARGE and SMALL 143
PERCENTILE.INC and PERCENTILE.EXC 143
PERCENTRANK.INC and PERCENTRANK.EXC 146
Data analysis tool: Rank and Percentile 148

CHAPTER 7: **Summarizing It All** 151
Counting Out 151
COUNT, COUNTA, COUNTBLANK, COUNTIF, COUNTIFS 151
The Long and Short of It 154
MAX, MAXA, MIN, and MINA 154
Getting Esoteric 156
SKEW and SKEW.P 156
KURT 158
Tuning In the Frequency 160
FREQUENCY 160
Data analysis tool: Histogram 162
Can You Give Me a Description? 164
Data analysis tool: Descriptive Statistics 164
Be Quick About It! 166
Instant Statistics 169

CHAPTER 8: **What's Normal?** 171
Hitting the Curve 171
Digging deeper 172
Parameters of a normal distribution 173
NORM.DIST 175
NORM.INV 176
A Distinguished Member of the Family 177
NORM.S.DIST 178
NORM.S.INV 179
PHI and GAUSS 180
Graphing a Standard Normal Distribution 180

PART 3: DRAWING CONCLUSIONS FROM DATA 183

CHAPTER 9: **The Confidence Game: Estimation** 185
Understanding Sampling Distributions 186
An EXTREMELY Important Idea: The Central Limit Theorem 187
(Approximately) simulating the Central Limit Theorem 189
The Limits of Confidence 193
Finding confidence limits for a mean 193
CONFIDENCE.NORM 195
Fit to a t 197
CONFIDENCE.T 198

CHAPTER 10: **One-Sample Hypothesis Testing** 199
Hypotheses, Tests, and Errors 199
Hypothesis Tests and Sampling Distributions 201
Catching Some Z's Again 203
Z.TEST 205
t for One 206
T.DIST, T.DIST.RT, and T.DIST.2T 208
T.INV and T.INV.2T 209
Visualizing a t-Distribution 210
Testing a Variance 211
CHISQ.DIST and CHISQ.DIST.RT 213
CHISQ.INV and CHISQ.INV.RT 214
Visualizing a Chi-Square Distribution 215

CHAPTER 11: **Two-Sample Hypothesis Testing** 217
Hypotheses Built for Two 217
Revisited 218
Applying the Central Limit Theorem 220
Z's once more 220
Data analysis tool: z-Test: Two Sample for Means 222

t for Two 225
Like peas in a pod: Equal variances 225
Like p's and q's: Unequal variances. 227
T.TEST 227
Data analysis tool: t-Test: Two Sample. 229
A Matched Set: Hypothesis Testing for Paired Samples 232
T.TEST for matched samples 234
Data analysis tool: t-Test: Paired Two Sample for Means 235
Testing Two Variances 237
Using F in conjunction with t 239
F.TEST 240
F.DIST and F.DIST.RT 241
F.INV and F.INV.RT. 243
Data analysis tool: F-test: Two Sample for Variances. 244
Visualizing the F-Distribution. 246

CHAPTER 12: **Testing More Than Two Samples** 249

Testing More Than Two 249
A thorny problem 250
A solution 251
Meaningful relationships 255
After the F-test 256
Data analysis tool: Anova: Single Factor. 259
Comparing the means 261
Another Kind of Hypothesis, Another Kind of Test. 263
Working with repeated measures ANOVA. 263
Getting trendy 266
Data analysis tool: Anova: Two Factor Without Replication 267
Analyzing trend 272

CHAPTER 13: **Slightly More Complicated Testing** 275

Cracking the Combinations 275
Breaking down the variances. 276
Data analysis tool: Anova: Two-Factor Without Replication 277
Cracking the Combinations Again. 279
Rows and columns 280
Interactions. 281
The analysis 281
Data analysis tool: Anova: Two-Factor With Replication 283
Two Kinds of Variables . . . at Once. 285
Using Excel with a Mixed Design. 287
Graphing the Results. 291
After the ANOVA 293

CHAPTER 14: **Regression: Linear and Multiple** 295
The Plot of Scatter 295
Graphing Lines 297
Regression: What a Line! 299
Using regression for forecasting 301
Variation around the regression line 301
Testing hypotheses about regression 303
Worksheet Functions for Regression 308
SLOPE, INTERCEPT, STEYX 309
FORECAST.LINEAR 310
Array function: TREND 311
Array function: LINEST 315
Data Analysis Tool: Regression 317
Tabled output 319
Graphic output 321
Juggling Many Relationships at Once: Multiple Regression 322
Excel Tools for Multiple Regression 323
TREND revisited 323
LINEST revisited 325
Regression data analysis tool revisited 328

CHAPTER 15: **Correlation: The Rise and Fall of Relationships** 331
Scatterplots Again 331
Understanding Correlation 332
Correlation and Regression 334
Testing Hypotheses About Correlation 337
Is a correlation coefficient greater than zero? 337
Do two correlation coefficients differ? 338
Worksheet Functions for Correlation 340
CORREL and PEARSON 340
RSQ 342
COVARIANCE.P and COVARIANCE.S 342
Data Analysis Tool: Correlation 343
Tabled output 344
Data Analysis Tool: Covariance 347
Testing Hypotheses About Correlation 348
Worksheet functions: FISHER, FISHERINV 348

CHAPTER 16: **It's About Time** 351
A Series and Its Components 351
A Moving Experience 352
Lining up the trend 353
Data Analysis tool: Moving Average 353

How To Be a Smoothie, Exponentially 356
One-Click Forecasting! 357

CHAPTER 17: **Non-Parametric Statistics** 363
Independent Samples 364
Two samples: Mann-Whitney U test 364
More than two samples: Kruskal-Wallis one-way ANOVA 366
Matched Samples 367
Two samples: Wilcoxon matched-pairs signed ranks 368
More than two samples: Friedman two-way ANOVA 370
More than two samples: Cochran's Q 371
Correlation: Spearman's r_S 373
A Heads-Up 375

PART 4: PROBABILITY 377

CHAPTER 18: **Introducing Probability** 379
What Is Probability? 379
Experiments, trials, events, and sample spaces 380
Sample spaces and probability 380
Compound Events 381
Union and intersection 381
Intersection again 382
Conditional Probability 383
Working with the probabilities 384
The foundation of hypothesis testing 384
Large Sample Spaces 384
Permutations 385
Combinations 386
Worksheet Functions 387
FACT 387
PERMUT and PERMUTIONA 387
COMBIN and COMBINA 388
Random Variables: Discrete and Continuous 389
Probability Distributions and Density Functions 389
The Binomial Distribution 391
Worksheet Functions 392
BINOM.DIST and BINOM.DIST.RANGE 393
NEGBINOM.DIST 395
Hypothesis Testing with the Binomial Distribution 396
BINOM.INV 396
More on hypothesis testing 398
The Hypergeometric Distribution 399
HYPGEOM.DIST 400

CHAPTER 19: **More on Probability** 403
Discovering Beta 403
BETA.DIST 405
BETA.INV 407
Poisson 408
POISSON.DIST 409
Working with Gamma 411
The gamma function and GAMMA 411
The gamma distribution and GAMMA.DIST 411
GAMMA.INV 414
Exponential 414
EXPON.DIST 415

CHAPTER 20: **A Career in Modeling** 417
Modeling a Distribution 417
Plunging into the Poisson distribution 418
Visualizing the Poisson distribution 419
Working with the Poisson distribution 420
Using POISSON.DIST again 421
Testing the model's fit 421
A word about CHISQ.TEST 424
Playing ball with a model 425
A Simulating Discussion 428
Taking a chance: The Monte Carlo method 428
Loading the dice 428
Simulating the Central Limit Theorem 432

PART 5: THE PART OF TENS 437

CHAPTER 21: **Ten Statistical and Graphical Tips and Traps** 439
Significant Doesn't Always Mean Important 439
Trying to Not Reject a Null Hypothesis Has a Number of Implications 440
Regression Isn't Always Linear 440
Extrapolating Beyond a Sample Scatterplot Is a Bad Idea 441
Examine the Variability Around a Regression Line 441
A Sample Can Be Too Large 441
Consumers: Know Your Axes 442
Graphing a Categorical Variable as Though It's a Quantitative Variable Is Just Wrong 442
Whenever Appropriate, Include Variability in Your Graph 443
Be Careful When Relating Statistics Textbook Concepts to Excel 444

CHAPTER 22: **Ten Things (Twelve, Actually) That Just Didn't Fit in Any Other Chapter** 445
Graphing the Standard Error of the Mean. 446
Probabilities and Distributions 449
PROB 449
WEIBULL.DIST. 450
Drawing Samples. 451
Testing Independence: The True Use of CHISQ.TEST. 452
Logarithmica Esoterica 454
What is a logarithm? 454
What is *e?* 456
LOGNORM.DIST 459
LOGNORM.INV 460
Array Function: LOGEST 461
Array Function: GROWTH. 463
The logs of Gamma 467
Sorting Data 468

APPENDIX A: **When Your Worksheet Is a Database** 471
Introducing Excel Databases 471
The Satellites database. 472
The criteria range 473
The format of a database function 474
Counting and Retrieving. 476
DCOUNT *and* DCOUNTA. 476
DGET 477
Arithmetic 477
DMAX *and* DMIN 477
DSUM. 478
DPRODUCT 478
Statistics 478
DAVERAGE 479
DVAR *and* DVARP 479
DSTDEV *and* DSTDEVP. 479
According to form 480
Pivot Tables 481

APPENDIX B: **The Analysis of Covariance** 487
Covariance: A Closer Look 487
Why You Analyze Covariance. 488
How You Analyze Covariance 489

ANCOVA in Excel 490
Method 1: ANOVA 491
Method 2: Regression 495
After the ANCOVA 498
And One More Thing 499

INDEX 501

Introduction

What? Yet another statistics book? Well . . . this is a statistics book, all right, but in my humble (and thoroughly biased) opinion, it's not *just* another statistics book.

What? Yet another Excel book? Same thoroughly biased opinion — it's not just another Excel book. What? Yet another edition of a book that's not just another statistics book and not just another Excel book? Well . . . yes. You got me there.

So here's the story — for the previous three editions and for this one. Many statistics books teach you the concepts but don't give you a way to apply them. That often leads to a lack of understanding. With Excel, you have a ready-made package for applying statistics concepts.

Looking at it from the opposite direction, many Excel books show you Excel's capabilities but don't tell you about the concepts behind them. Before I tell you about an Excel statistical tool, I give you the statistical foundation it's based on. That way, you understand the tool when you use it — and you use it more effectively.

I didn't want to write a book that's just "select this menu" and "click this button." Some of that is necessary, of course, in any book that shows you how to use a software package. My goal was to go way beyond that.

I also didn't want to write a statistics "cookbook" — when-faced-with-problem-#310-use-statistical-procedure-#214. My goal was to go way beyond that, too.

Bottom line: This book isn't just about statistics or just about Excel — it sits firmly at the intersection of the two. In the course of telling you about statistics, I cover every Excel statistical feature. (Well . . . *almost.* I left one out. I left it out of the first three editions, too. It's called "Fourier Analysis." All the necessary math to understand it would take a whole book, and you might never use this tool, anyway.)

About This Book

Although statistics involves a logical progression of concepts, I organized this book so you can open it up in any chapter and start reading. The idea is for you to find what you're looking for in a hurry and use it immediately — whether it's a statistical concept or an Excel tool.

On the other hand, cover to cover is okay if you're so inclined. If you're a statistics newbie and you have to use Excel for statistical analysis, I recommend you begin at the beginning — even if you know Excel pretty well.

What You Can Safely Skip

Any reference book throws a lot of information at you, and this one is no exception. I intend it all to be useful, but I don't aim it all at the same level. So if you're not deeply into the subject matter, you can avoid paragraphs marked with the Technical Stuff icon.

Every so often, you'll run into sidebars. They provide information that elaborates on a topic, but they're not part of the main path. If you're in a hurry, you can breeze past them.

Because I wrote this book so you can open it up anywhere and start using it, step-by-step instructions appear throughout. Many of the procedures I describe have steps in common. After you go through some of the procedures, you can probably skip the first few steps when you come to a procedure you haven't been through before.

Foolish Assumptions

This is not an introductory book on Excel or on Windows, so I'm assuming:

- You know how to work with Windows. I don't spell out the details of pointing, clicking, selecting, and so forth.
- You have Excel 2016 installed on your Windows computer or on your Mac and you can work along with the examples. I don't walk you through the steps of Excel installation.

- You've worked with Excel, and you understand the essentials of worksheets and formulas.

If you don't know much about Excel, consider looking into Greg Harvey's excellent Excel books in the *For Dummies* series.

How This Book Is Organized

I've organized this book into five parts and four appendixes (including two that you can find on this book's companion website at `www.statisticalanalysiswexcel4e`).

Part 1: Getting Started with Statistical Analysis with Excel: A Marriage Made In Heaven

In Part 1, I provide a general introduction to statistics and to Excel's statistical capabilities. I discuss important statistical concepts and describe useful Excel techniques. If it 's a long time since your last course in statistics or if you've never had a statistics course at all, start here. If you haven't worked with Excel's built-in functions (of any kind), definitely start here.

Part 2: Describing Data

Part of statistics is to take sets of numbers and summarize them in meaningful ways. Here's where you find out how to do that. We all know about averages and how to compute them. But that's not the whole story. In this part, I tell you about additional statistics that fill in the gaps, and I show you how to use Excel to work with those statistics. I also introduce Excel graphics in this part.

Part 3: Drawing Conclusions from Data

Part 3 addresses the fundamental aim of statistical analysis: to go beyond the data and help decision-makers make decisions. Usually, the data are measurements of a sample taken from a large population. The goal is to use these data to figure out what's going on in the population.

This opens a wide range of questions: What does an average mean? What does the difference between two averages mean? Are two things associated? These are only

a few of the questions I address in Part 3, and I discuss the Excel functions and tools that help you answer them.

Part 4: Working with Probability

Probability is the basis for statistical analysis and decision-making. In Part 4, I tell you all about it. I show you how to apply probability, particularly in the area of modeling. Excel provides a rich set of built-in capabilities that help you understand and apply probability. Here's where you find them.

Part 5: The Part of Tens

Part 5 meets two objectives. First, I get to stand on the soapbox and rant about statistical peeves and about helpful hints. The peeves and hints total up to ten. Also, I discuss ten (okay, 12) Excel things I couldn't fit into any other chapter. They come from all over the world of statistics. If it's Excel and statistical, and if you can't find it anywhere else in the book, you'll find it here.

As I said in the first three editions — pretty handy, this Part of Tens.

Appendix A: When Your Worksheet Is a Database

In addition to performing calculations, Excel serves another purpose: recordkeeping. Although it's not a dedicated database, Excel does offer some database functions. Some of them are statistical in nature. I introduce Excel database functions in Appendix A, along with pivot tables that allow you to turn your database inside out and look at your data in different ways.

Appendix B: The Analysis of Covariance

The Analysis of Covariance (ANCOVA) is a statistical technique that combines two other techniques: analysis of variance and regression analysis. If you know how two variables are related, you can use that knowledge in some nifty ways, and this is one of the ways. The kicker is that Excel doesn't have a built-in tool for ANCOVA — but I show you how to use what Excel does have so you can get the job done.

Bonus Appendix B1: When Your Data Live Elsewhere

This appendix is all about importing data into Excel — from the web, from databases, from text, and from PDF documents.

Bonus Appendix B2: Tips for Teachers (and Learners)

Excel is terrific for managing, manipulating, and analyzing data. It's also a great tool for helping people understand statistical concepts. This appendix covers some ways for using Excel to do just that.

Icons Used in This Book

As is the case with all *For Dummies* books, icons appear all over the place. Each one is a little picture in the margin that lets you know something special about the paragraph it's next to.


TIP

This icon points out a hint or a shortcut that can help you in your work and make you an all-around better human being.


REMEMBER

This one points out timeless wisdom to take with you long after you finish this book, young Jedi.


WARNING

Pay attention to this icon. It's a reminder to avoid something that might gum up the works for you.


TECHNICAL STUFF

As I mention earlier, in the section "What You Can Safely Skip," this icon indicates material you can blow right past if statistics and Excel aren't your passion.

Where to Go from Here

You can start the book anywhere, but here are a few hints. Want to learn the foundations of statistics? Turn the page. Introduce yourself to Excel's statistical features? That's Chapter 2. Want to start with graphics? Hit Chapter 3. For anything else, find it in the table of contents or in the index and go for it.

In addition to what you're reading right now, this book also comes with a free, access-anywhere Cheat Sheet that will help you quickly use the tools I discuss. To get this Cheat Sheet, visit `www.dummies.com` and search for "Statistical Analysis with Excel For Dummies Cheat Sheet" in the Search box. And don't forget to check out the bonus content on this book's companion website at `www.dummies.com/go/statisticalanalysiswexcel4e`.

1 Getting Started with Statistical Analysis with Excel: A Marriage Made in Heaven

IN THIS PART . . .

- Find out about Excel's statistical capabilities
- Explore how to work with populations and samples
- Test your hypotheses
- Understand errors in decision making
- Determine independent and dependent variables

IN THIS CHAPTER

- Introducing statistical concepts
- Generalizing from samples to populations
- Getting into probability
- Making decisions
- New and old features in Excel 2016
- Understanding important Excel fundamentals

Chapter 1

Evaluating Data in the Real World

The field of statistics is all about decision-making — decision-making based on groups of numbers. Statisticians constantly ask questions: What do the numbers tell us? What are the trends? What predictions can we make? What conclusions can we draw?

To answer these questions, statisticians have developed an impressive array of analytical tools. These tools help us to make sense of the mountains of data that are out there waiting for us to delve into, and to understand the numbers we generate in the course of our own work.

The Statistical (and Related) Notions You Just Have to Know

Because intensive calculation is often part and parcel of the statistician's tool set, many people have the misconception that statistics is about number crunching.

Number crunching is just one small part of the path to sound decisions, however.

By shouldering the number-crunching load, software increases our speed of traveling down that path. Some software packages are specialized for statistical analysis and contain many of the tools that statisticians use. Although not marketed specifically as a statistical package, Excel provides a number of these tools, which is why I wrote this book.

I said that number crunching is a small part of the path to sound decisions. The most important part is the concepts statisticians work with, and that's what I talk about for most of the rest of this chapter.

Samples and populations

On election night, TV commentators routinely predict the outcome of elections before the polls close. Most of the time they're right. How do they do that?

The trick is to interview a sample of voters after they cast their ballots. Assuming the voters tell the truth about whom they voted for, and assuming the sample truly represents the population, network analysts use the sample data to generalize to the population of voters.

This is the job of a statistician — to use the findings from a sample to make a decision about the population from which the sample comes. But sometimes those decisions don't turn out the way the numbers predicted. History buffs are probably familiar with the memorable picture of President Harry Truman holding up a copy of the *Chicago Daily Tribune* with the famous, but wrong, headline "Dewey Defeats Truman" after the 1948 election. Part of the statistician's job is to express how much confidence he or she has in the decision.

Another election-related example speaks to the idea of the confidence in the decision. Pre-election polls (again, assuming a representative sample of voters) tell you the percentage of sampled voters who prefer each candidate. The polling organization adds how accurate it believes the polls are. When you hear a newscaster say something like "accurate to within 3 percent," you're hearing a judgment about confidence.

Here's another example. Suppose you've been assigned to find the average reading speed of all fifth-grade children in the United States but you haven't got the time or the money to test them all. What would you do?

Your best bet is to take a sample of fifth-graders, measure their reading speeds (in words per minute), and calculate the average of the reading speeds in the sample. You can then use the sample average as an estimate of the population average.

Estimating the population average is one kind of *inference* that statisticians make from sample data. I discuss inference in more detail in the upcoming section "Inferential Statistics: Testing Hypotheses."

REMEMBER

Here's some terminology you have to know: Characteristics of a population (like the population average) are called *parameters*, and characteristics of a sample (like the sample average) are called *statistics.* When you confine your field of view to samples, your statistics are *descriptive.* When you broaden your horizons and concern yourself with populations, your statistics are *inferential.*

REMEMBER

And here's a notation convention you have to know: Statisticians use Greek letters (μ, σ, ρ) to stand for parameters, and English letters ($\overline{X}$, s, r) to stand for statistics. Figure 1-1 summarizes the relationship between populations and samples, and parameters and statistics.

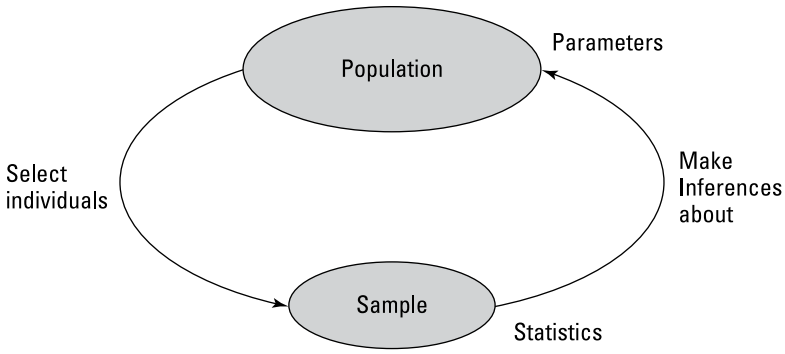


FIGURE 1-1: The relationship between populations, samples, parameters, and statistics.

Variables: Dependent and independent

Simply put, a *variable* is something that can take on more than one value. (Something that can have only one value is called a *constant.*) Some variables you might be familiar with are today's temperature, the Dow Jones Industrial Average, your age, and the value of the dollar against the euro.

Statisticians care about two kinds of variables: *independent* and *dependent.* Each kind of variable crops up in any study or experiment, and statisticians assess the relationship between them.

For example, imagine a new way of teaching reading that's intended to increase the reading speed of fifth-graders. Before putting this new method into schools, it would be a good idea to test it. To do that, a researcher would randomly assign a sample of fifth-grade students to one of two groups: One group receives instruction via the new method, and the other receives instruction via traditional methods. Before and after both groups receive instruction, the researcher measures the reading speeds of all the children in this study. What happens next? I get to that in the upcoming section "Inferential Statistics: Testing Hypotheses."

For now, understand that the independent variable here is Method of Instruction. The two possible values of this variable are New and Traditional. The dependent variable is reading speed — which you might measure in words per minute.

REMEMBER

In general, the idea is to find out if changes in the independent variable are associated with changes in the dependent variable.

REMEMBER

In the examples that appear throughout the book, I show you how to use Excel to calculate various characteristics of groups of scores. Keep in mind that each time I show you a group of scores, I'm really talking about the values of a dependent variable.

Types of data

Data come in four kinds. When you work with a variable, the way you work with it depends on what kind of data it is.

The first variety is called *nominal* data. If a number is a piece of nominal data, it's just a name. Its value doesn't signify anything. A good example is the number on an athlete's jersey. It's just a way of identifying the athlete and distinguishing him or her from teammates. The number doesn't indicate the athlete's level of skill.

Next come ordinal data. *Ordinal* data are all about order, and numbers begin to take on meaning over and above just being identifiers. A higher number indicates the presence of more of a particular attribute than a lower number. One example is the *Mohs scale:* Used since 1822, it's a scale whose values are 1 through 10; mineralogists use this scale to rate the hardness of substances. Diamond, rated at 10, is the hardest. Talc, rated at 1, is the softest. A substance that has a given rating can scratch any substance that has a lower rating.

What's missing from the Mohs scale (and from all ordinal data) is the idea of equal intervals and equal differences. The difference between a hardness of 10 and a hardness of 8 is not the same as the difference between a hardness of 6 and a hardness of 4.