Advance Praise for *Head First Statistics*

"*Head First Statistics* is by far the most entertaining, attention-catching study guide on the market. By presenting the material in an engaging manner, it provides students with a comfortable way to learn an otherwise cumbersome subject. The explanation of the topics is presented in a manner comprehensible to students of all levels."

— Ariana Anderson, Teaching Fellow/PhD candidate in Statistics, UCLA

"*Head First Statistics* is deceptively friendly. Breeze through the explanations and exercises and you just may find yourself raising the topic of normal vs. Poisson distribution in ordinary social conversation, which I can assure you is not advised!"

— Gary Wolf, Contributing Editor, *Wired* Magazine

"Dawn Griffiths has split some very complicated concepts into much smaller, less frightening, bits of stuff that real-life people will find very easy to digest. Lots of graphics and photos make the material very approachable, and I have developed quite a crush on the attractive lady model who is asking about gumballs on page 458."

— Bruce Frey, author of *Statistics Hacks*

"Head First is an intuitive way to understand statistics using simple, real-life examples that make learning fun and natural."

— Michael Prerau, computational neuroscientist and statistics instructor, Boston University

"Thought Head First was just for computer nerds? Try the brain-friendly way with statistics and you'll change your mind. It really works."

— Andy Parker

"This book is a great way for students to learn statistics—it is entertaining, comprehensive, and easy to understand. A perfect solution!"

— Danielle Levitt

"Down with dull statistics books! Even my cat liked this one."

— Cary Collett

Praise for other *Head First* books

"Kathy and Bert's *Head First Java* transforms the printed page into the closest thing to a GUI you've ever seen. In a wry, hip manner, the authors make learning Java an engaging 'what're they gonna do next?' experience."

—Warren Keuffel, Software Development Magazine

"Beyond the engaging style that drags you forward from know-nothing into exalted Java warrior status, Head First Java covers a huge amount of practical matters that other texts leave as the dreaded "exercise for the reader..." It's clever, wry, hip and practical—there aren't a lot of textbooks that can make that claim and live up to it while also teaching you about object serialization and network launch protocols. "

—Dr. Dan Russell, Director of User Sciences and Experience Research IBM Almaden Research Center (and teaches Artificial Intelligence at Stanford University)

"It's fast, irreverent, fun, and engaging. Be careful—you might actually learn something!"

—Ken Arnold, former Senior Engineer at Sun Microsystems Co-author (with James Gosling, creator of Java), *The Java Programming Language*

"I feel like a thousand pounds of books have just been lifted off of my head."

—Ward Cunningham, inventor of the Wiki and founder of the Hillside Group

"Just the right tone for the geeked-out, casual-cool guru coder in all of us. The right reference for practical development strategies—gets my brain going without having to slog through a bunch of tired stale professor-speak."

—Travis Kalanick, Founder of Scour and Red Swoosh Member of the MIT TR100

"There are books you buy, books you keep, books you keep on your desk, and thanks to O'Reilly and the Head First crew, there is the penultimate category, Head First books. They're the ones that are dog-eared, mangled, and carried everywhere. Head First SQL is at the top of my stack. Heck, even the PDF I have for review is tattered and torn."

— Bill Sawyer, ATG Curriculum Manager, Oracle

"This book's admirable clarity, humor and substantial doses of clever make it the sort of book that helps even non-programmers think well about problem-solving."

— Cory Doctorow, co-editor of Boing Boing Author, *Down and Out in the Magic Kingdom* and *Someone Comes to Town, Someone Leaves Town*

Praise for other *Head First* books

"I received the book yesterday and started to read it...and I couldn't stop. This is definitely très 'cool.' It is fun, but they cover a lot of ground and they are right to the point. I'm really impressed."

— Erich Gamma, IBM Distinguished Engineer, and co-author of *Design Patterns*

"One of the funniest and smartest books on software design I've ever read."

— Aaron LaBerge, VP Technology, ESPN.com

"What used to be a long trial and error learning process has now been reduced neatly into an engaging paperback."

— Mike Davidson, CEO, Newsvine, Inc.

"Elegant design is at the core of every chapter here, each concept conveyed with equal doses of pragmatism and wit."

— Ken Goldstein, Executive Vice President, Disney Online

"I ♥ Head First HTML with CSS & XHTML—it teaches you everything you need to learn in a 'fun coated' format."

— Sally Applin, UI Designer and Artist

"Usually when reading through a book or article on design patterns, I'd have to occasionally stick myself in the eye with something just to make sure I was paying attention. Not with this book. Odd as it may sound, this book makes learning about design patterns fun.

"While other books on design patterns are saying 'Buehler… Buehler… Buehler…' this book is on the float belting out 'Shake it up, baby!'"

— Eric Wuehler

"I literally love this book. In fact, I kissed this book in front of my wife."

— Satish Kumar

Other related books from O'Reilly

Statistics Hacks™

Statistics in a Nutshell

Mind Hacks™

Mind Performance Hacks™

Your Brain: The Missing Manual

Other books in O'Reilly's *Head First* series

Head First Java™

Head First Object-Oriented Analysis and Design (OOA&D)

Head First HTML with CSS and XHTML

Head First Design Patterns

Head First Servlets and JSP

Head First EJB

Head First PMP

Head First SQL

Head First Software Development

Head First JavaScript

Head First Ajax

Head First Physics

Head First PHP & MySQL (2008)

Head First Rails (2008)

Head First Web Design (2008)

Head First Algebra (2008)

Head First Programming (2009)

Head First Statistics

Dawn Griffiths

O'REILLY®

Beijing • Cambridge • Köln • Sebastopol • Taipei • Tokyo

Head First Statistics

by Dawn Griffiths

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly Media books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*safari.oreilly.com*). For more information, contact our corporate/institutional sales department: (800) 998-9938 or *corporate@oreilly.com*.

Series Creators: Kathy Sierra, Bert Bates

Series Editor: Brett D. McLaughlin

Editor: Sanders Kleinfeld

Design Editor: Louise Barr

Cover Designers: Louise Barr, Steve Fehler

Production Editor: Brittany Smith

Indexer: Julie Hawks

Page Viewers: David Griffiths, Mum and Dad

Printing History:

August 2008: First Edition.

No snorers were harmed in the making of this book, although a horse lost its toupee at one point and suffered a minor indignity in front of the other horses. Also a snowboarder picked up a few bruises along the way, but nothing serious.

RepKover. This book uses RepKover™, a durable and flexible lay-flat binding.

ISBN: 978-0-596-52758-7

[M] [9/08]

To David, Mum, Dad, and Carl. Thanks for the support and believing I could do it. But you'll have to wait a while for the car.

Author of Head First Statistics

Dawn Griffiths started life as a mathematician at a top UK university. She was awarded a First-Class Honours degree in Mathematics, but she turned down a PhD scholarship studying particularly rare breeds of differential equations when she realized people would stop talking to her at parties. Instead she pursued a career in software development, and she currently combines IT consultancy with writing and mathematics.

When Dawn's not working on Head First books, you'll find her honing her Tai Chi skills, making bobbin lace or cooking nice meals. She hasn't yet mastered the art of doing all three at the same time. She also enjoys traveling, and spending time with her lovely husband, David.

Dawn has a theory that **Head First Bobbin Lacemaking** might prove to a be a big cult hit, but she suspects that Brett and Laurie might disagree.

Table of Contents (Summary)

Table of Contents (the real thing)

Intro

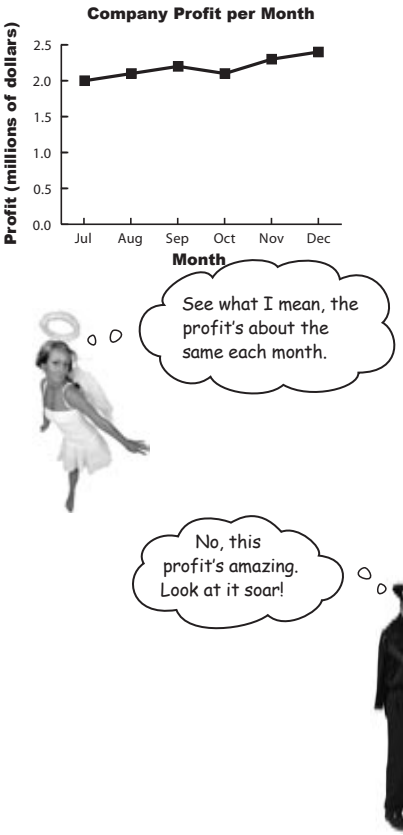
Your brain on statistics. Here *you* are trying to *learn* something, while here your *brain* is doing you a favor by making sure the learning doesn't *stick*. Your brain's thinking, "Better leave room for more important things, like which wild animals to avoid and whether naked snowboarding is a bad idea." So how *do* you trick your brain into thinking that your life depends on knowing statistics?

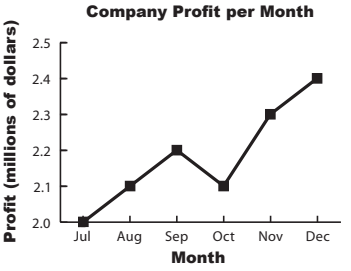
visualizing information

1 First Impressions

Can't tell your facts from your figures?

Statistics help you make sense of confusing sets of data. They **make the complex simple**. And when you've found out what's really going on, you need a way of **visualizing** it and **telling everyone else**. So if you want to pick the best chart for the job, grab your coat, pack your best slide rule, and join us on a ride to Statsville.





measuring central tendency

2 The Middle Way

Sometimes you just need to get to the heart of the matter.

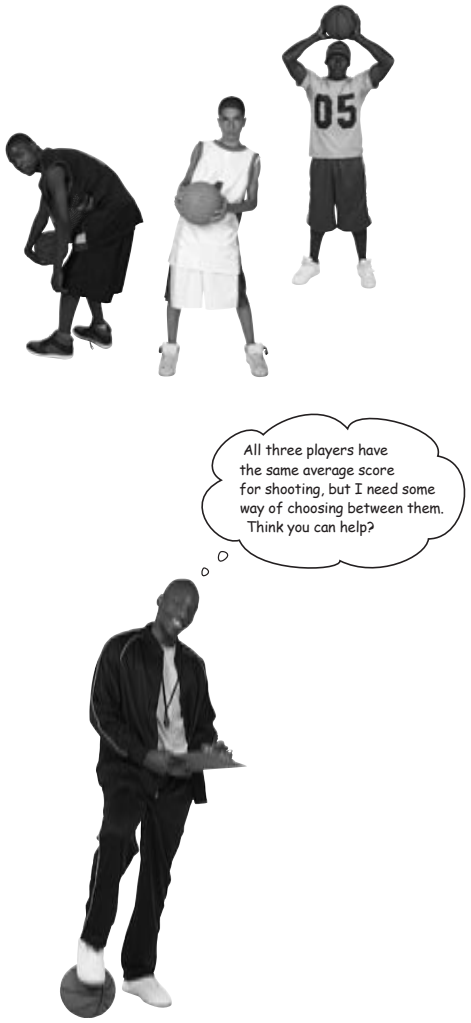
It can be difficult to see patterns and trends in a big pile of figures, and finding the **average** is often the first step towards seeing the bigger picture. With averages at your disposal, you'll be able to quickly find the most representative values in your data and draw important conclusions. In this chapter, we'll look at several ways to calculate one of the most important statistics in town—mean, median, and mode—and you'll start to see how to effectively **summarize data** as concisely and usefully as possible.

measuring variability and spread

3 Power Ranges

Not everything's reliable, but how can you tell?

Averages do a great job of giving you a typical value in your data set, but they **don't tell you the full story**. OK, so you know where the center of your data is, but often the mean, median, and mode alone aren't enough information to go on when you're summarizing a data set. In this chapter, we'll show you how to take your data skills to the next level as we begin to analyze **ranges and variation**.

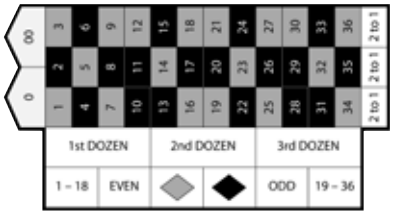


calculating probabilities

4 Taking Chances

Life is full of uncertainty.

Sometimes it can be impossible to say what will happen from one minute to the next. But certain events are more likely to occur than others, and that's where **probability theory** comes into play. Probability lets you **predict the future** by assessing how likely outcomes are, and knowing what could happen helps you make **informed decisions**. In this chapter, you'll find out more about probability and learn how to take control of the future!



using discrete probability distributions

5 Manage Your Expectations

Unlikely events happen, but what are the consequences?

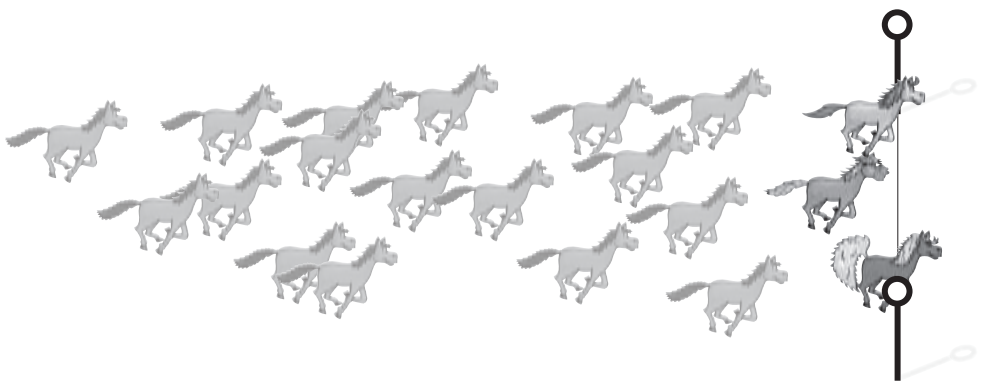
So far we've looked at how probabilities tell you how likely certain events are. What probability *doesn't* tell you is the **overall impact** of these events, and what it means to you. Sure, you'll sometimes make it big on the roulette table, but is it really worth it with all the money you lose in the meantime? In this chapter, we'll show you how you can use probability to **predict long-term outcomes**, and also **measure the certainty** of these predictions.

permutations and combinations

6 Making Arrangements

Sometimes, order is important.

Counting **all the possible ways** in which you can order things is time consuming, but the trouble is, this sort of information is **crucial** for calculating some probabilities. In this chapter, we'll show you a **quick way** of deriving this sort of information without you having to figure out what all of the possible outcomes are. Come with us and we'll show you how to **count the possibilities**.

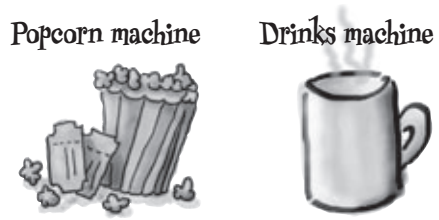


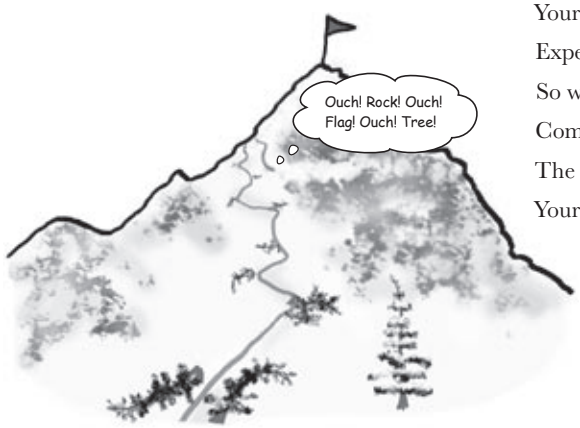
geometric, binomial, and poisson distributions

7 Keeping Things Discrete

Calculating probability distributions takes time.

So far we've looked at how to calculate and use probability distributions, but wouldn't it be nice to have something **easier to work with**, or just **quicker to calculate**? In this chapter, we'll show you some **special probability distributions** that follow very definite patterns. Once you know these patterns, you'll be able to use them to **calculate probabilities, expectations, and variances in record time**. Read on, and we'll introduce you to the geometric, binomial and Poisson distributions.





using the normal distribution

8 Being Normal

Discrete probability distributions can't handle every situation.

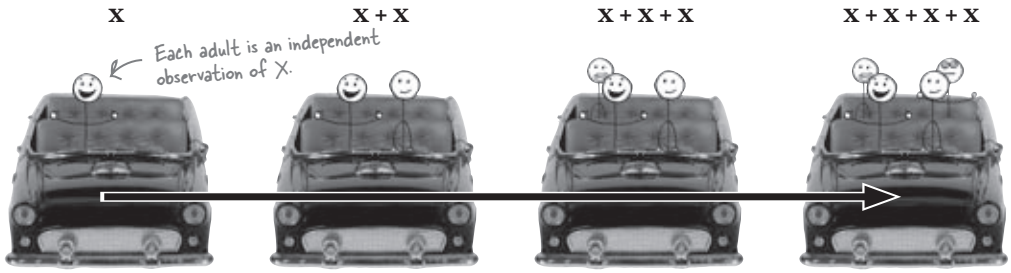
So far we've looked at probability distributions where we've been able to specify exact values, but this isn't the case for every set of data. Some types of data just **don't fit** the probability distributions we've encountered so far. In this chapter, we'll take a look at how **continuous probability distributions** work, and introduce you to one of the most important probability distributions in town—the **normal distribution**.

using the normal distribution ii

9 Beyond Normal

If only all probability distributions were normal.

Life can be so much ***simpler*** with the normal distribution. Why spend all your time working out individual probabilities when you can look up entire ranges in one swoop, and still leave time for game play? In this chapter, you'll see how to **solve more complex problems** in the blink of an eye, and you'll also find out how to bring some of that normal goodness to **other probability distributions**.

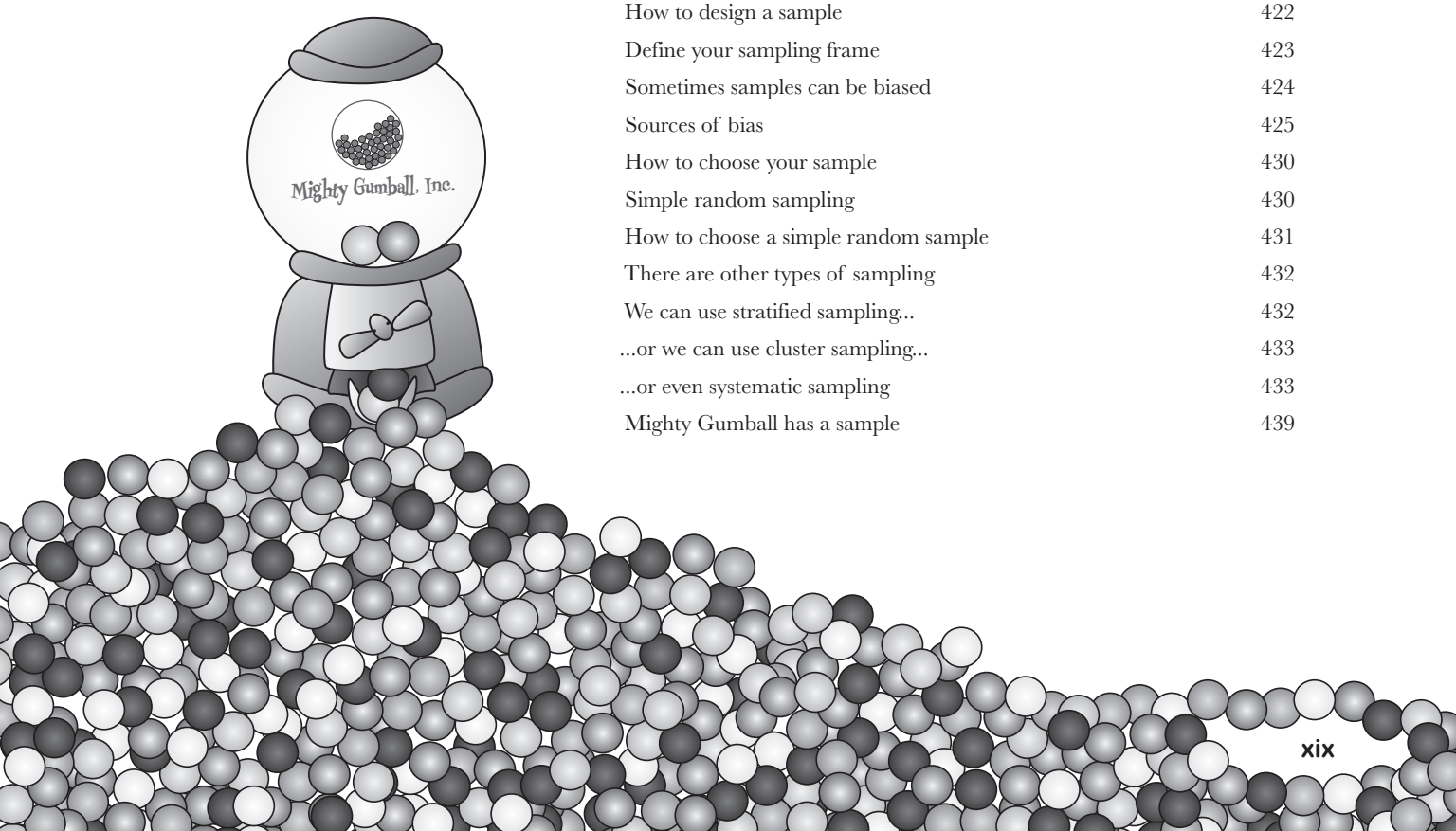


using statistical sampling

10 Taking Samples

Statistics deal with data, but where does it come from?

Some of the time, data's easy to collect, such as the ages of people attending a health club or the sales figures for a games company. But what about the times when data isn't so easy to collect? Sometimes the number of things we want to collect data about are so huge that it's difficult to know where to start. In this chapter, we'll take a look at how you can **effectively gather data** in the real world, in a way that's efficient, accurate, and can also save you time and money to boot. Welcome to the world of sampling.

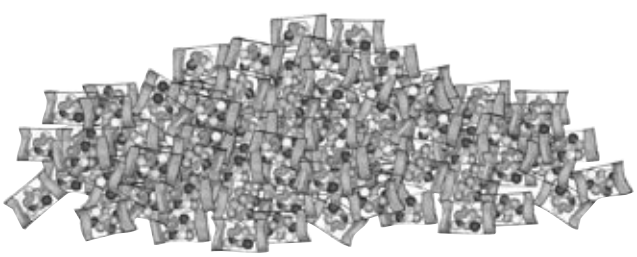


estimating your population

11 Making Predictions

Wouldn't it be great if you could tell what a population was like, just by taking one sample?

Before you can claim **full sample mastery**, you need to know how to use your samples to best effect once you've collected them. This means using them to **accurately predict** what the population will be like and coming up with a way of saying how **reliable** your predictions are. In this chapter, we'll show you how knowing your sample helps you **get to know your population**, and vice versa.

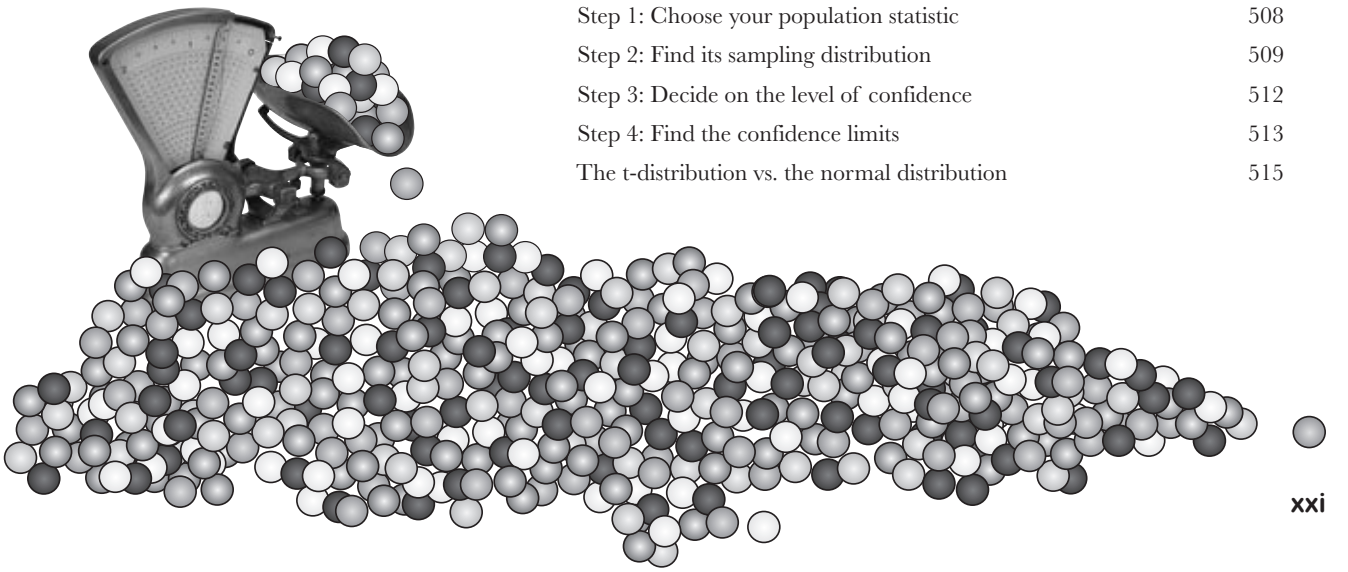


constructing confidence intervals

12 Guessing with Confidence

Sometimes samples don't give quite the right result.

You've seen how you can use point estimators to estimate the **precise value** of the population mean, variance, or proportion, but the trouble is, how can you be certain that your estimate is completely accurate? After all, your assumptions about the population rely on just one sample, and what if your sample's off? In this chapter, you'll see **another way of estimating population statistics**, one that **allows for uncertainty**. Pick up your probability tables, and we'll show you the ins and outs of **confidence intervals**.



using hypothesis tests

13 Look at the Evidence

Not everything you're told is absolutely certain.

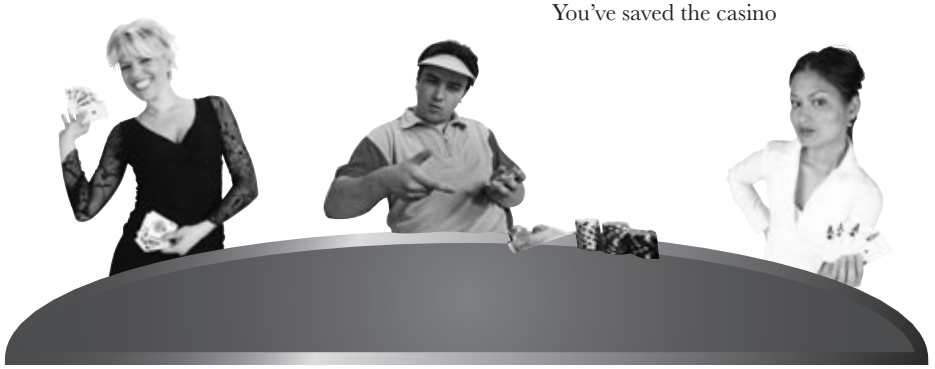
The trouble is, how do you know when what you're being told isn't right? ***Hypothesis tests*** give you a way of using samples to test whether or not statistical claims are likely to be true. They give you a way of ***weighing the evidence*** and testing whether extreme results can be explained by ***mere coincidence***, or whether there are darker forces at work. Come with us on a ride through this chapter, and we'll show you how you can use hypothesis tests to confirm or allay your deepest suspicions.

14 the χ^2 distribution

There's Something Going On...

Sometimes things don't turn out quite the way you expect.

When you model a situation using a particular probability distribution, you have a good idea of how things are likely to turn out long-term. But what happens if there are differences between **what you expect and what you get?** How can you tell whether your discrepancies come down to normal fluctuations, or whether they're a sign of an underlying problem with your probability model instead? In this chapter, we'll show you how you can use the χ^2 distribution to **analyze your results** and sniff out **suspicious results**.



correlation and regression

15 What's My Line?

Have you ever wondered how two things are connected?

So far we've looked at statistics that tell you about just one variable—like men's height, points scored by basketball players, or how long gumball flavor lasts—but there are other statistics that tell you about the **connection between variables**. Seeing how things are connected can give you a lot of information about the real world, information that you can use to your advantage. Stay with us while we show you the **key to spotting connections**: correlation and regression.

leftovers

The Top Ten Things (we didn't cover)

Even after all that, there's a bit more. There are just a few more things we think you need to know. We wouldn't feel right about ignoring them, even though they only need a brief mention. So before you put the book down, take a read through these **short but important statistics tidbits**.

statistics tables

Looking Things up

Where would you be without your trusty probability tables?

Understanding your probability distributions isn't quite enough. For some of them, you need to be able to **look up your probabilities** in standard **probability tables**. In this appendix you'll find tables for the **normal, t and X^2 distributions** so you can look up probabilities to your heart's content.