

## PREFACE

**S***tats: Data and Models*, fifth edition, has been especially exciting to develop. The book you hold steps beyond our previous editions in several important ways. Of course, we've kept our conversational style and anecdotes,<sup>1</sup> but we've enriched that material with tools for teaching about randomness, sampling distribution models, and inference throughout the book. And we've expanded discussions of models for data to introduce models with more than two variables earlier in the text. We've taken our inspiration both from our experience in the classroom and from the 2016 revision of the Guidelines for Assessment and Instruction in Statistics Education (GAISE) report adopted by the American Statistical Association. As a result, we increased the text's innovative uses of technology to encourage more statistical thinking, while maintaining its traditional core concepts and coverage. You'll notice that, to expand our attention beyond just one or two variables, we've adjusted the order of some topics.

## Innovations

### Technology

One of the new GAISE guidelines states: *Use technology to explore concepts and analyze data*. We think a modern statistics text should recognize from the start that statistics is practiced with technology. And so should our students. You won't find tedious calculations worked by hand. You *will* find equation forms that favor intuition over calculation. You'll find extensive use of real data—even large datasets. Throughout, you'll find a focus on statistical thinking rather than calculation. The question that motivates each of our hundreds of examples is not “How do you calculate the answer?” but “How do you think about the answer?”

For this edition of *Stats: Data and Models* we've taken this principle still further. We have harnessed technology to improve the learning of two of the most difficult concepts in the introductory course: the idea of a sampling distribution and the reasoning of statistical inference.

### Multivariable Thinking and Multiple Regression

GAISE's first guideline is to give students experience with multivariable thinking. The world is not univariate, and relationships are not limited to two variables. This edition of *Stats: Data and Models* introduces a third variable as early as Chapter 3's discussion of contingency tables and mosaic plots. Then, following the discussion of correlation and regression as a tool (that is, without inference) in Chapters 6, 7, and 8, we introduce multiple regression in Chapter 9.

Multiple regression may be the most widely used statistical method, and it is certainly one that students need to understand. It is easy to perform multiple regressions with any statistics program, and the exercise of thinking about more than two variables early in the course is worth the effort. We've added new material about interpreting what regression models say. The effectiveness of multiple regression is immediately obvious and makes the reach and power of statistics clear. The use of real data underscores the universal applicability of these methods.

When we return to regression in Chapters 23 and 24 to discuss inference, we can deal with both simple and multiple regression models together. There is nothing different to discuss. (For this reason we set aside the  $F$ -test until the chapter on ANOVA.)

<sup>1</sup>And footnotes

Innovative ways to teach the logic of statistical inference have received increasing attention. Among these are greater use of computer-based simulations and resampling methods (randomization tests and bootstrapping) to teach concepts of inference.

## Bootstrap

The introduction to the new GAISE guidelines explicitly mentions the bootstrap method. The bootstrap is not as widely available or as widely understood as multiple regression. But it follows our presentation naturally. In this edition, we introduce a new feature, **Random Matters**. Random Matters elements in early chapters draw small samples repeatedly from large populations to illustrate how the randomness introduced by sampling leads to both sampling distributions and statistical reasoning for inference. But what can we do when we have only a sample? The bootstrap provides a way to continue this line of thought, now by resampling from the sample at hand.

Bootstrapping provides an elegant way to simulate sampling distributions that we might not otherwise be able to see. And it does not require the assumption of Normality expected by Student's  $t$ -based methods. However, these methods are not as widely available or widely used in other disciplines, so they should not be the only—or even the principal—methods taught. They may be able to enhance student understanding, but instructors may wish to downplay them if that seems best for a class. We've placed these sections strategically so that instructors can choose the level that they are comfortable with and that works best with their course.

## Real Data

GAISE recommends that instructors integrate real data with a context and purpose. More and more high school math teachers are using examples from statistics to demonstrate intuitively how a little bit of math can help us say a lot about the world. So our readers expect statistics to be about real-world insights. *Intro Stats* keeps readers engaged and interested because we show statistics in action right from the start. The exercises pose problems of the kind likely to be encountered in real life and propose ways to think about making inferences almost immediately—and, of course, always with real, up-to-date data.

Let us be clear. *Stats: Data and Models* comes with an archive of more than 500 datasets used in more than 700 applications throughout the book. The datasets are available online at the student resource site, in the DASL archive and in MyLab Statistics. Examples that use these datasets cite them in the text. Exercises are marked when they use one of them; exercise names usually indicate the name of the dataset. We encourage students to get the datasets and reproduce our examples using their statistics software, and some of the exercises require that.

## Streamlined Content

Following the GAISE recommendations, we've streamlined several parts of the course: Introductory material is covered more rapidly. Today's students have seen a lot of statistics in their K–12 math courses and in their daily contact with online and news sources. We still cover the topics to establish consistent terminology (such as the difference between a histogram and a bar chart). Chapter 2 does most of the work that previously took two chapters.

The Random Matters features show students that statistics vary from sample to sample, show them (empirical) sampling distributions, note the effect of sample size on the shape and variation of the sampling distribution of the mean, and suggest that it looks Normal. As a result, the discussion of the Central Limit Theorem is transformed from the most difficult one in the course to a relatively short discussion (“What you think is true about means really is true; there's this theorem.”) that can lead directly to the reasoning of confidence intervals.

Finally, introducing multiple regression doesn't really add much to the lesson on inference for multiple regression because little is new.

## GAISE 2016

As we've said, all of these enhancements follow the new Guidelines for Assessment and Instruction in Statistics Education (GAISE) 2016 report adopted by the American Statistical Association:

1. Teach statistical thinking.
  - ◆ Teach statistics as an investigative process of problem solving and decision making.
  - ◆ Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

The result is a course that is more aligned with the skills needed in the 21st century, one that focuses even more on statistical thinking and makes use of technology in innovative ways, while retaining core principles and topic coverage.

The challenge has been to use this modern point of view to improve learning without discarding what is valuable in the traditional introductory course. Many first statistics courses serve wide audiences of students who need these skills for their own work in disciplines where traditional statistical methods are, well, traditional. So we have not reduced our emphasis on the concepts and methods you expect to find in our texts.

## Chapter Order

---

We've streamlined the presentation of basic topics that most students have already seen. Pie charts, bar charts, histograms, and summary statistics all appear in Chapter 2. Chapter 3 introduces contingency tables, and Chapter 4 discusses comparing distributions. Chapter 5 introduces the Normal model and the 68–95–99.7 Rule. The four chapters of Part II then explore linear relationships among quantitative variables—but here we introduce only the models and how they help us understand relationships. We leave the inference questions until later in the book. Part III discusses how data are gathered by survey and experiment.

Part IV provides background material on probability, random variables, and probability models. In Part V, Chapter 16 introduces confidence intervals for proportions as soon as we've reassured students that their intuition about the sampling distribution of proportions is correct. Chapter 17 formalizes the Central Limit Theorem and introduces Student's *t*-models. Chapter 18 is then about testing hypotheses, and Chapter 19 elaborates further, discussing alpha levels, Type I and Type II errors, power, and effect size. The chapters in Part VI deal with comparing groups (both with proportions and with means), paired samples, chi-square. Finally, Part VII discusses inferences for regression models (both simple and multiple), intelligent uses of multiple regression, and Analysis of Variance, both one- and two-way. A final chapter on data mining looks to the future.

We've found that one of the challenges students face is how to know what technique to use when. In the real world, questions don't come at the ends of the chapters. So, as always, we've provided summaries at the end of each part along with a series of exercises designed to stretch student understanding. These Part Reviews are a mix of questions from all the chapters in that part. The final set are "book-level" review problems that ask students to integrate what they've learned from the entire course. The questions range

from simple questions about what method to use in various situations to a more complete data analyses from real data. We hope that these will provide a useful way for students to organize their understanding at the end of the course.

## Our Approach

We've discussed how this book is different, but there are some things we haven't changed.

- ◆ **Readability.** This book doesn't read like other statistics texts. Our style is both colloquial and informative, engaging students to actually read the book to see what it says.
- ◆ **Humor.** You will find quips and wry comments throughout the narrative, in margin notes, and in footnotes.
- ◆ **Informality.** Our informal diction doesn't mean that we treat the subject matter lightly or informally. We try to be precise and, wherever possible, we offer deeper explanations and justifications than those found in most introductory texts.
- ◆ **Focused lessons.** The chapters are shorter than in most other texts so that instructors and students can focus on one topic at a time.
- ◆ **Consistency.** We try to avoid the "do what we say, not what we do" trap. Having taught the importance of plotting data and checking assumptions and conditions, we model that behavior through the rest of the book. (Check out the exercises in Chapter 24.)
- ◆ **The need to read.** Statistics is a consistent story about how to understand the world when we have data. The story can't be told piecemeal. This is a book that needs to be read, so we've tried to make the reading experience enjoyable. Students who start with the exercises and then search back for a worked example that looks the same but with different numbers will find that our presentation doesn't support that approach.

## Mathematics

Mathematics can make discussions of statistics concepts, probability, and inference clear and concise. We don't shy away from using math where it can clarify without intimidating. But we know that some students are discouraged by equations, so we always provide a verbal description and a numerical example as well.

Nor do we slide in the opposite direction and concentrate on calculation. Although statistics calculations are generally straightforward, they are also usually tedious. And, more to the point, today, virtually all statistics are calculated with technology. We have selected the equations that focus on illuminating concepts and methods rather than for hand calculation. We sometimes give an alternative formula, better suited for hand calculation, for those who find that following the calculation process is a better way to learn about the result.

## Technology and Data

We assume that computers and appropriate software are available—at least for demonstration purposes. We hope that students have access to computers and statistics software for their analyses.

We discuss generic computer output at the end of most chapters, but we don't adopt any particular statistics software. The **Tech Support** sections at the ends of chapters offer guidance for seven common software platforms: Data Desk, Excel, JMP, Minitab, SPSS, StatCrunch, and R. We also offer some advice for TI-83/84 Plus graphing calculators, although we hope that those who use them will also have some access to computers and statistics software.

We don't limit ourselves to small, artificial datasets, but base most examples and exercises on real data with a moderate number of cases. Machine-readable versions of the data are available at the book's website, [pearsonhighered.com/dvb](http://pearsonhighered.com/dvb) and at [dasl.datadescription.com](http://dasl.datadescription.com).

## Features

---

### Enhancing Understanding

**Where Are We Going?** Each chapter starts with a paragraph that raises the kinds of questions we deal with in the chapter. A chapter outline organizes the major topics and sections.

**New! Random Matters.** This new feature travels along a progressive path of understanding randomness and our data. The first Random Matters element begins our thinking about drawing inferences from data. Subsequent Random Matters draw histograms of sample means, introduce the thinking involved in permutation tests, and encourage judgment about how likely the observed statistic seems when viewed against the simulated sampling distribution of the null hypothesis (without, of course, using those terms).

**Margin and in-text boxed notes.** Throughout each chapter, boxed margin and in-text notes enhance and enrich the text.

**Reality Check.** We regularly remind students that statistics is about understanding the world with data. Results that make no sense are probably wrong, no matter how carefully we think we did the calculations. Mistakes are often easy to spot with a little thought, so we ask students to stop for a reality check before interpreting their result.

**Notation Alert.** Throughout this book, we emphasize the importance of clear communication, and proper notation is part of the vocabulary of statistics. We've found that it helps students when we are clear about the letters and symbols statisticians use to mean very specific things, so we've included Notation Alerts whenever we introduce a special notation that students will see again.

Each chapter ends with several elements to help students study and consolidate what they've seen in the chapter.

- ◆ **What Can Go Wrong?** sections highlight the most common errors that people make and the misconceptions they have about statistics. One of our goals is to arm students with the tools to detect statistical errors and to offer practice in debunking misuses of statistics, whether intentional or not.
- ◆ **Connections** specifically ties the new topics to those learned in previous chapters.
- ◆ The **Chapter Review** summarizes the story told by the chapter and provides a bullet list of the major concepts and principles covered.
- ◆ A **Review of Terms** is a glossary of all of the special terms introduced in the chapter. In the text, these are printed in **bold** and underlined. The Review provides page references, so students can easily turn back to a full discussion of the term if the brief definition isn't sufficient.

The **Tech Support** section provides the commands in each of the supported statistics packages that deal with the topic covered by the chapter. These are not full documentation, but should be enough to get a student started in the right direction.

### Learning by Example



**Step-by-Step Examples.** We have expanded and updated the examples in our innovative Step-by-Step feature. Each one provides a longer, worked example that guides students through the process of analyzing a problem. The examples follow our three-step Think, Show, Tell organization for approaching a statistics task. They are organized with general explanations of each step on the left and a worked-out solution on the right. The right side of the grid models what would be an "A" level solution to the problem. Step-by-Steps illustrate the importance of thinking about a statistics question (What do we know? What do we hope to learn? Are the assumptions and conditions satisfied?) and reporting our findings (the Tell step). The Show step contains the mechanics of calculating results and conveys our belief that it is only one part of the process. Our emphasis is on statistical thinking, and the pedagogical result is a better understanding of the concept, not just number crunching.

**Examples.** As we introduce each important concept, we provide a focused example that applies it—usually with real, up-to-the-minute data. Many examples carry the discussion through the chapter, picking up the story and moving it forward as students learn more about the topic.

**Just Checking.** Just Checking questions are quick checks throughout the chapter; most involve very little calculation. These questions encourage students to pause and think about what they’ve just read. The Just Checking answers are at the end of the exercise sets in each chapter so students can easily check themselves.

## Assessing Understanding

Our **Exercises** have some special features worth noting. First, you’ll find relatively simple, focused exercises organized by chapter section. After that come more extensive exercises that may deal with topics from several parts of the chapter or even from previous chapters as they combine with the topics of the chapter at hand. All exercises appear in pairs. The odd-numbered exercises have answers in the back of student texts. Each even-numbered exercise hits the same topic (although not in exactly the same way) as the previous odd exercise. But the even-numbered answers are not provided. If a student is stuck on an even exercise, looking at the previous odd one (and its answer) can often provide the help needed.

More than 600 of our exercises have a  tag next to them to indicate that the dataset referenced in the exercise is available electronically. The exercise title or a note provides the dataset title. Some exercises have a  tag to indicate that they call for the student to generate random samples or use randomization methods such as the bootstrap. Although we hope students will have access to computers, we provide ample exercises with full computer output for students to read, interpret, and explain.

We place all the exercises—including section-level exercises—at the end of the chapter. Our writing style is colloquial and encourages reading. We are telling a story about how to understand the world when you have data. Interrupting that story with exercises every few pages would encourage a focus on the calculations rather than the concepts.

**Part Reviews.** The book is partitioned into seven conceptual parts; each ends with a Part Review. The part review discusses the concepts in that part of the text, tying them together and summarizing the story thus far. Then there are more exercises. These exercises have the advantage (for study purposes) of not being tied to a chapter, so they lack the hints of what to do that would come from that identification. That makes them more like potential exam questions and a good tool for review. Unlike the chapter exercises, these are not paired.

**Parts I-VII Cumulative Review Exercises.** Cumulative Review exercises are longer and cover concepts from the book as a whole.

## Additional Resources Online

---

Most of the supporting materials can be found online:

At the book’s website at [pearsonhighered.com/dvb](http://pearsonhighered.com/dvb)

Within the MyLab Statistics course at [pearson.com/mylab/statistics](http://pearson.com/mylab/statistics)

Datasets are also available at [dasl.datadesk.com](http://dasl.datadesk.com).

**Data desk** is a statistics program with a graphical interface that is easy to learn and use. A student version is available at [datadesk.com](http://datadesk.com). Click on the **Teachers & Students** tab at the top of the page.

New tools that provide interactive versions of the distribution tables at the back of the book and tools for randomization inference methods such as the bootstrap and for repeated sampling from larger populations can be found online at [astools.datadesk.com](http://astools.datadesk.com).



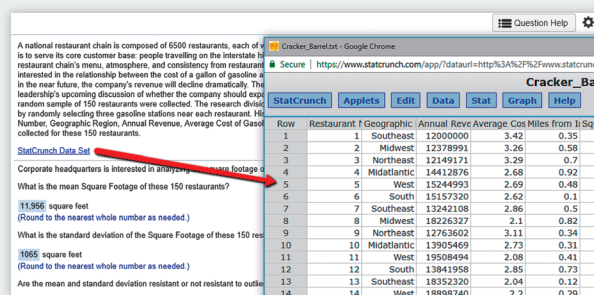
## MyLab Statistics for *Stats: Data & Models, 5e*

(access code required)

MyLab Statistics is available to accompany Pearson's market-leading text offerings. To give students a consistent tone, voice, and teaching method, each text's flavor and approach are tightly integrated throughout the accompanying MyLab course, making learning the material as seamless as possible.

### NEW! StatCrunch Projects

StatCrunch Projects provide opportunities for students to explore data beyond the classroom. In each project, students analyze a data set in StatCrunch® and answer assignable MyLab questions for immediate feedback. StatCrunch Projects span the entire curriculum or focus on certain key concepts. Questions from each project can also be assigned individually.



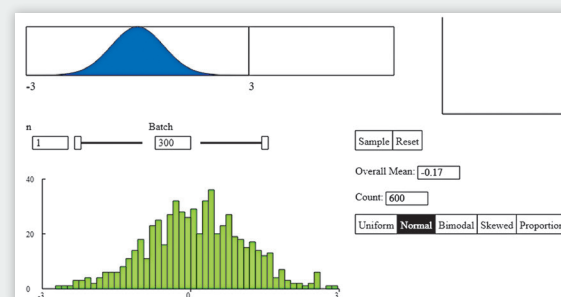
Row	Movie	Run Time (minutes)	Budget (\$M)	USGross (\$)	Genre	Dramas	var7
1	White Noise	101	30	56094360	Horror	0	
2	Coach Carter	136	45	67264877	Drama	1	
3	Elektra	100	65	24409722	Action	0	
4	Racing Stripes	110	30	49772522	Comedy	0	
5	Assault on Precinct 13	109	30	20040895	Action	0	
6	Are We There Yet?	94	20	82674398	Comedy	0	
7	Alone in the Dark	96	20	5178569	Horror	0	
8	Indigo	105	25	51100486	Drama	1	
9	Boogeyman	88	20	46752382	Horror	0	
10	The Wedding Date	90	40	31726995	Comedy	0	
11	Hitch	113	55	1.77784268	Comedy	0	
12	Pooh's Heffalump Movie	68	20	18081628	Adventure	0	
13	Because of Winn-Dixie	106	15	32647042	Comedy	0	
14	Constantine	122	75	75976178	Action	0	
15	Son of the Mask	86	100	17018422	Comedy	0	
16	Cursed	105	35	19294901	Horror	0	
17	Man of the House	100	50	19699706	Comedy	0	
18	Be Cool	150	75	55849401	Comedy	0	
19	The Jacket	103	28.5	6301131	Thriller	0	
20	The Pacifier	97	56	1.13006888	Comedy	0	
21	Hostage	108	75	34626443	Action	0	
22	Robots	89	80	1.28200018	Adventure	0	
23	The Upside of Anger	117	12	18761993	Drama	1	
24	Ice Princess	99	25	24381134	Comedy	0	
25	The Ring Two	111	50	75941727	Horror	0	
26	Miss Congeniality 2: Armed at Heart	115	60	48478006	Comedy	0	

### UPDATED! Real-World Data

Statistical concepts are applied to everyday life through the extensive current, real-world data examples and exercises provided throughout the text.

### NEW! Interactive Applets

Author-created interactive applets take an experiential approach to helping students learn statistical concepts. They are available in MyLab Statistics and at [astools.datadesk.com](http://astools.datadesk.com).



# Resources for Success

## Instructor Resources

### Annotated Instructor's Edition

Includes answers to all text exercises, as well as a set of Instructor Resource Pages that offer chapter-by-chapter teaching suggestions and commentary. (ISBN-13: 978-0-13-516396-2; ISBN-10: 0-13-516396-X)

### Instructor's Solutions Manual (Download Only)

This manual contains detailed solutions to all of the exercises. These files can be downloaded from within MyLab Statistics or from [www.pearson.com](http://www.pearson.com).

### Instructor Resource Guide (Download Only)

This resource guide includes chapter-by-chapter comments on the major concepts, tips on presenting topics, extra teaching examples, and sample quizzes, tests, and projects. These files can be downloaded from within MyLab Statistics or from [www.pearson.com](http://www.pearson.com).

### TestGen

TestGen® ([www.pearson.com/testgen](http://www.pearson.com/testgen)) enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text. TestGen is algorithmically based, allowing instructors to create multiple but equivalent versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The software and test bank are available for download from Pearson's online catalog, [www.pearson.com](http://www.pearson.com). The questions are also assignable in MyLab Statistics.

### PowerPoint Lecture Slides

PowerPoint Lecture Slides provide an overview of each chapter, stressing important definitions and offering additional examples. They can be downloaded from MyLab Statistics or from [www.pearson.com](http://www.pearson.com).

### Learning Catalytics

Now included in all MyLab Statistics courses, this student response tool uses students' smartphones, tablets, or laptops to engage them in more interactive tasks and thinking during lecture. Learning Catalytics™ fosters student engagement and peer-to-peer learning with real-time analytics. Access pre-built exercises created specifically for statistics.

### Question Libraries

In addition to Statcrunch Projects, MyLab Statistics also includes a Getting Ready for Statistics library that contains more than 450 exercises on prerequisite topics and a Conceptual Question Library with 1000 questions that assess conceptual understanding.

### Minitab and Minitab Express™

Bundling Minitab and Minitab Express software with the text ensures students have access to the software they need for the duration of their course. ISBN 13: 978-0-13-445640-9; ISBN 10: 0-13-445640-8

### JMP Student Edition

An easy-to-use, streamlined version of JMP desktop statistical discovery software from SAS Institute, Inc. is available for bundling with the text. ISBN-13: 978-0-13-467979-2; ISBN-10: 0-13-467979-2

### XLSTAT™

An Excel add-in that enhances the analytical capabilities of Excel, XLSTAT is used by leading businesses and universities around the world. It is available to bundle with this text. For more information go to [www.pearsonhighered.com/xlstatupdate](http://www.pearsonhighered.com/xlstatupdate). ISBN-13: 978-0-321-75932-0; ISBN-10: 0-321-75932-X

### Accessibility

Pearson works continuously to ensure our products are as accessible as possible to all students. We are working toward achieving WCAG 2.0 Level AA and Section 508 standards, as expressed in the Pearson Guidelines for Accessible Educational Web Media, [www.pearson.com/mylab/statistics/accessibility](http://www.pearson.com/mylab/statistics/accessibility).

## Student Resources

### Video Resources

Step-by-Step Example videos guide students through the process of analyzing a problem using the "Think, Show, and Tell" strategy from the textbook. StatTalk Videos, hosted by fun-loving statistician Andrew Vickers, demonstrates important statistical concepts through interesting stories and real-life events. StatTalk videos come with accompanying MyLab assessment questions.

### StatCrunch

StatCrunch® is powerful web-based statistical software that allows users to collect, crunch, and communicate with data. The vibrant online community offers tens of thousands of shared datasets for students and instructors to analyze, in addition to all of the datasets in the text or online homework. StatCrunch is integrated directly into MyLab Statistics or it can be purchased separately. Learn more at [www.statcrunch.com](http://www.statcrunch.com).

### Datasets Available Online

Data sets can be found at [pearsonhighered.com/dvb](http://pearsonhighered.com/dvb) and (DASL) at [www.DASL.datadescription.com](http://www.DASL.datadescription.com) which holds all of the datasets for the book as well as many others. Search by dataset name to find datasets for the exercises in the book. Search by statistics method to find examples for lessons or for assignments. Datasets can be easily transferred to any statistics program.

### Statistical Software Support

Instructors and students can copy datasets from the text, DASL, and MyLab exercises directly into software such as StatCrunch, Data desk, or Excel®. Students can also access instructional support tools including tutorial videos, Study Cards, and manuals for a variety of statistical software programs including StatCrunch, Excel, Minitab®, JMP®, R, SPSS, and TI 83/84 calculators.