

# ***Introduction to Statistics for Community College Students***

(First Edition)

**By Matt Teachout**

**College of the Canyons  
Santa Clarita, CA, USA**



*This material is from [Introduction to Statistics for Community College Students](#),  
1<sup>st</sup> Edition, by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed  
under a "CC-By" [Creative Commons Attribution 4.0 International license](#) – 10/1/18*

**Special thanks to all of the people that made this book possible.**

**I would like to thank all of the math and statistics teachers at College of the Canyons. Your work to improve statistics education continues to inspire educational reform.**

**I would also like to thank James Glapa-Grossklag, Brian Weston and the College of the Canyons OER staff for their tireless work to support free high quality materials for students through the OER movement.**

**Special thanks to Randy Ades, Gevork Demirchyan, Kathy Kubo, Joe Gerda, Udani Ranasinghe and Rupa Sinha for supporting me through this project.**

**I would also like to thank the fall 2019 pilot team: Alvard Adamyan, Randy Ades, Steve Brown, Roy Erickson, Jason Moss, Said Najafi, Linda Newland, Eric Pham, Udani Ranasinghe, Amar Singh, Rupa Sinha and Juliana Yankey. Your feedback and suggestions were fabulous.**

**Most especially, I would like to thank my wife Link for her love, support, and patience with me during the endless hours of writing.**



## Introduction

We live in a world of big data. Our children grow up with technology, computers, the internet, and a vast amount of information at their fingertips. How can we make sense of this vast amount of information? How can we know when we are being deceived or if data is biased? How can businesses or hospitals make sound decisions based on data? The answer to all of these questions lies in the study of statistics and data science.

The study of statistics and data science is vital in our modern age. Statisticians, data scientists and data analysts are now in high demand. Every company, hospital, sports team or college needs trained employees who can collect and analyze data and help make good decisions based on data. In the U.S., there is a huge deficit in the number of people trained in this area. Not only do we not have enough statisticians, data scientists and data analysts, but also we do not have enough statistics teachers and statistics tutors.

Statistics is a deep well of knowledge that men and women have devoted their lives to studying. In this book, I will attempt to give you some useful tools and an overarching picture of statistics, but we will only be playing in a puddle compared to that deep well of statistical knowledge. I do hope to whet your appetite though and encourage you to study statistics and data science past this initial class. Maybe we can add a few more statisticians, data scientists and data analysts to our ranks. The world desperately needs you.

## Vocabulary

In many ways, studying statistics is like learning a new language. Statisticians and data scientists are experts in collecting and analyzing data. Yet they also need to be able to explain their findings to a world that has very little understanding of statistical reasoning. Statistical terms often have a very different meaning than we would find in a dictionary. New statistics students often find the vocabulary overwhelming. For this reason, you will find definitions of important terms at the beginning of each section. It is important to read these carefully and commit them to memory. Many of my past students found it very helpful to would write each term and definition on a 3x5 card. They would then study their 3x5 cards throughout the semester. Here are some terms and definitions to get you started.

Data: Information in all forms.

Statistics: The science of collecting, preparing and analyzing data.

Statistician: An expert in the science of statistics. The leaders in the field.

Data Mining: The process of collecting and storing data.

Data Science: Many feel this is just another word for statistics, but with maybe less emphasis on high-level statistical analysis and more emphasis on data mining and using computer science.

Data Scientist: A specialist in data science.

Data Analyst: A specialist in analyzing data to make good decisions.



### Notes about OER and Creative Commons Licensing

There are many fabulous books on statistics and analyzing data. Unfortunately, they are extremely expensive and many college students cannot afford the cost. I wrote this book to help people learn to analyze data. It is free to use the material in this book, update it, add to it, print it or just read it. It is an open educational resource (OER) and so anyone can use it.

Many college students struggle to balance work and family with their education. One of the biggest roadblocks for many students is the cost of textbooks. Many students simply cannot afford the cost of textbooks. They attend classes without purchasing books and materials needed for the class. Of course, this is a major impediment to passing their classes, but the students have no choice. They simply cannot afford \$100-\$200 textbooks. For this reason, I believe strongly in open educational resources (OER). Open source materials like this book are available and are virtually free for students.

This textbook is licensed through Creative Commons as “Attribution CC-BY”. Creative Commons describes this license as follows: “This license lets others distribute, remix, tweak, and build upon (the author’s) work, even commercially, as long as they (give) credit (to the author) for the original creation.” This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.” If you need to see the license deed or legal code, please go to <https://creativecommons.org/licenses/> and look under the “CC-By” section.

### Supplementary Materials

Links to the individual sections and problem sets can be found at [www.matt-teachout.org](http://www.matt-teachout.org). Just click on the “statistics” tab and pick what chapter you wish to study. (Posted summer 2019)

There are also additional supplementary materials under the “statistics” tab at [www.matt-teachout.org](http://www.matt-teachout.org). (Posted summer 2019)

- Teaching notes for instructors.
- Data sets.
- Affective domain assignments.

The following supplementary materials may be found under the “Instructor Resources” and “Statistics Resources” tabs at [www.matt-teachout.org](http://www.matt-teachout.org). (Posted summer 2019)

- Suggested teaching schedules.
- Example projects.
- Practice Problem Answer Keys.



### Important Note about Technology

We live in the age of computers, internet and a huge volume of data. No practicing statistician or data scientist uses a calculator or tables to analyze data. You cannot even begin to analyze a data set of 100,000 values by hand with a calculator. You need high-powered computer software. There are many statistics software programs on the market, but very few of them are free.

If you read the history of statistics, you will find brilliant scientists, mathematicians and people in business who were trying to figure out data, but had no access to a computer. (Computers had not been invented yet.) Our pioneers of statistics dreamed of the day that they could compute statistics and graphs and analyze data with the touch of a button. They invented complicated techniques for analyzing data because they had no choice. Today, computers can calculate statistics and graphs directly.

Here is the problem. Many statistics classes taught in high schools, community colleges and even some universities are teaching statistics as if computers have not been invented yet. They are teaching the techniques developed by our pioneers of statistics before the computer age. This is a terrible approach to the subject, especially for the thousands of students that actually want to work in the field. A statistics class should be a study of how to practically collect and analyze data with a computer, not a class on what to do if computers had not been invented yet.

Are formulas important in statistics? Yes. We look at formulas to understand what they tell us about the data and the world around us. The pioneers of statistics did an amazing job of addressing the major ideas of statistics with formulas and inventive calculations. However, we should not use a formula and a calculator to calculate a statistic for a data set with 10,000 values or use charts that list critical values and P-values. High-powered computers with statistics software can calculate the statistic and make graphs directly. Then students can focus on the analysis part, and explore and discover the meaning behind the data.

This book will show students how to use statistics software to calculate statistics and graphs. I want students to learn to explore and analyze the data and not spend all their time just trying to calculate something. Remember, no one pays a data scientist to calculate something a computer can already do. A data scientist is paid to explore the data and explain what the data may be telling us. The key question is not “how is this calculated?” The key question is “what can I learn from this data?” Computers are tools to explore the data.

### StatKey

Teaching statistics with computer software is very important, but many schools and students cannot afford to pay for software. For this reason, I prefer to use software that is free for students, but is also relatively easy to use. My favorite software by far is StatKey. StatKey is hosted online at [www.lock5stat.com](http://www.lock5stat.com). StatKey is free, well organized, easy to use, and has a nice blend of modern and traditional statistics capabilities.

### Statcato

Another free software that I sometimes use is Statcato. Statcato is a traditional JAVA based software found at [www.statcato.org](http://www.statcato.org). It can be saved to your computer or it can be opened with the “Java Web Start” function. Some students have difficulty getting Statcato to open on their home computers. For this reason, the homework sets will already have Statcato printouts. That way, if students have any technical issues, they can still complete the homework.

### Other Programs

You can of course use the book with any statistics software. Most basic statistics software programs are very similar to Statcato, but may not be free for students. Instructors that use a program other than StatKey or Statcato will need to give software directions to their students.



## Data Sets

The national (GAISE) guidelines for teaching statistics recommend that you use real data. Allowing students to learn statistics principles through analysis of real data is key. With that being said, there are many places where raw data can be found and used. The key data sets throughout this book are located at [www.matt-teachout.org](http://www.matt-teachout.org) under “Statistics” and “Data Sets”. It is wise to save the book and the data sets on your computer. That way you have them when you need them and do not have to keep going to the website.

## The Computer Dilemma

A statistics or pre-statistics class should be taught in a computer lab. It is important to allow the computers to do the difficult calculations. Students need to focus on interpretation and discovering the meaning behind the data. They cannot do that if they spend all their time trying to calculate with a formula or making graphs by hand.

If your school wants to teach statistics or pre-statistics, but you cannot teach in a computer lab, here are some suggestions for you.

1. Reserve unused computer labs. Some schools may have a couple computer labs that are not always in use. Schedule your statistics and pre-statistics classes in order to use the computer lab. Even if you can only reserve the lab once a week or once every two weeks, it will be a huge help to students.
2. Have groups of students share computers. If you do have a few computers in your classroom, you can divide the class up into groups and share computers. This has many advantages like encouraging explanations to one another and teamwork.
3. Teachers can use their own computer or laptop to project statistics software on a screen and have the class analyze the data with you. Teachers without any computer can make printed copies of the software printouts for your class and for exams. It is a poor substitute for a computer lab, but it is much better than teaching statistics as if computers have not been invented.

## Organization of Chapters and Sections

Many statistics books organize the material into four or five large units with several chapters in each unit. For example, a book may have five units with thirty chapters and give one exam for each unit. I prefer to organize the material into four large chapters with several sections in each chapter. This book has twenty-eight sections broken into four chapters. A good rule of thumb is to plan on one major exam for each chapter.

## Pedagogy

Instructors often ask me about pedagogy and the difficulty in teaching statistics to underprepared students. I thought I would share my core principles that define how I like to teach statistics classes.

1. Active Learning and Classwork
  - It is vital for students to practice problem solving and using the statistics software during class. When teaching a new section, I like to give a short lecture that focuses on vocabulary, understanding the main ideas, and using technology. Then I have the students work on the problems in the section with my help. They can work in groups or individually, but they need to work with the data themselves and get used to explaining concepts to others. I then give them a break and repeat the process (short lecture and work on problems).



## 2. Alternative Approaches to Homework

- I try to minimize the number of problems students do outside of class and maximize the number of problems done in class with me to tutor and help them. Homework usually consists of them finishing the problems they did not finish during class. I also like assigning alternative homework assignments that focusses on main concept understanding and explaining. Small writing assignments work well this. For example, I like to have students write a paragraph explaining the main concepts they learned during class. It is also a way for me to check for understanding. Statistical concepts can be very difficult, so I also like to supplement my students' knowledge for homework. For example, I have the students watch a video online, take notes and turn in their notes to me.

## 3. Use Technology

- In a way, I see my statistics classes as job training. No working statistician uses a calculator and a formula especially with big data. The job of a working data scientist is to collect data, use computer software to explore the data, make good data driven decisions, and be able to explain it to others. In my class, I like to mimic that process. I do not spend time teaching students how to calculate something by hand with a formula. Instead, I focus on using and interpreting the computer software. The computer software does the calculations. The students have the more important job of exploring, understanding and explaining.
- I tend not to front-load when teaching technology. I like to teach the theory and show the software together. It is difficult to find a program that does everything, especially one that is free. In this book, I used a blend of Excel, Statcato and StatKey. Data sets are managed in Excel spreadsheets.

## 4. Understanding over Calculating

- Formulas are very important in statistics, but not to calculate. As I have said, no one working in the field calculates statistics by hand with a formula and calculator. We use computer software to calculate. I like to study the formulas with my students. These formulas are programmed into statistical software like Statcato and StatKey. We need to understand what the formula is calculating and what it tell us about the data. What are the limitations of this formula? When can we use it and when can we not use it? It is about understanding the theory behind the formula.

## 5. Blend Traditional and Modern Statistics

- Computer technology drastically changed the landscape of statistics. While it is important for student to understand traditional formulas and techniques, I have found that modern computer-based techniques like bootstrapping and randomization have benefits also, especially in conceptual understanding. I like to blend the traditional and modern statistics techniques. In this book, I use Statcato for traditional calculations. I use StatKey for data exploration and exposing students to more modern computer-based techniques.

## 6. Vocabulary

- Statistics is like learning a new language. Statistical terms have their own definitions that are rarely the same as the one in the dictionary. Students need to learn to explain statistical ideas and understand statistical reasoning. I constantly cover and review the key terms and ideas in each section. Without understanding the vocabulary, students will really struggle with the theory.

