



*Saylor Academy awards*  
**DAN RIMNICEANU**

*this certificate for the prescribed program of study for the*  
**COMPUTER SCIENCE CURRICULUM**

Issue Date: 30 mai 2018

Certificate ID: 11589059



Sean Connor  
Director of Student Affairs  
Saylor Academy



*Saylor Academy awards*

**Dan Rinniceanu**

*this certificate of achievement for*  
**MA121: Introduction to Statistics**

5 aprilie 2018

Issue Date



11427160

Certificate ID

# MA121: Introduction to Statistics

If you invest in financial markets, you may want to predict the price of a stock in six months from now on the basis of company performance measures and other economic factors. As a college student, you may be interested in knowing the dependence of the mean starting salary of a college graduate, based on your GPA. These are just some examples that highlight how statistics are used in our modern society. To figure out the desired information for each example, you need data to analyze.

The purpose of this course is to introduce you to the subject of statistics as a science of data. There is data abound in this information age; how to extract useful knowledge and gain a sound understanding of complex data sets has been more of a challenge. In this course, we will focus on the fundamentals of statistics, which may be broadly described as the techniques to collect, clarify, summarize, organize, analyze, and interpret numerical information.

This course will begin with a brief overview of the discipline of statistics and will then quickly focus on descriptive statistics, introducing graphical methods of describing data. You will learn about combinatorial probability and random distributions, the latter of which serves as the foundation for statistical inference. On the side of inference, we will focus on both estimation and hypothesis testing issues. We will also examine the techniques to study the relationship between two or more variables; this is known as regression.

By the end of this course, you should gain a sound understanding of what statistics represent, how to use statistics to organize and display data, and how to draw valid inferences based on data by using appropriate statistical tools.

## **Unit 1: Statistics and Data**

In today's technologically advanced world, we have access to large volumes of data. The first step of data analysis is to accurately summarize all of this data, both graphically and numerically, so that we can understand what the data reveals. To be able to use and interpret the data correctly is essential to making informed decisions. For instance, when you see a survey of opinion about a certain TV program, you may be interested in the proportion of those people who indeed like the program.

In this unit, you will learn about descriptive statistics, which are used to summarize and display data. After completing this unit, you will know how to present your findings once you have collected data. For example, suppose you want to buy a new mobile phone with a particular type of a camera. Suppose you are not sure about the prices of any of the phones with this feature, so you access a website that provides you with a sample data set of prices, given your desired features. Looking at all of the prices in a sample can sometimes be confusing. A better way to compare this data might be to look at the median price and the variation of prices. The median and variation are two ways out of

several ways that you can describe data. You can also graph the data so that it is easier to see what the price distribution looks like.

In this unit, you will study precisely this; namely, you will learn both numerical and graphical ways to describe and display your data. You will understand the essentials of calculating common descriptive statistics for measuring center, variability, and skewness in data. You will learn to calculate and interpret these measurements and graphs.

Descriptive statistics are, as their name suggests, descriptive. They do not generalize beyond the data considered. Descriptive statistics illustrate what the data shows. Numerical descriptive measures computed from data are called statistics. Numerical descriptive measures of the population are called parameters. Inferential statistics can be used to generalize the findings from sample data to a broader population.

### **Completing this unit should take you approximately 22 hours.**

- Upon successful completion of this unit, you will be able to:
  - describe various types of sampling methods to data collection, and apply these methods;
  - create and interpret frequency tables;
  - display data graphically and interpret the following types of graphs: stem plots, histograms, and boxplots;
  - identify, describe, and calculate the following measures of the location of data: quartiles and percentiles;
  - identify, describe, and calculate the measures of the center of mean, median, and mode; and
  - identify, describe, and calculate the following measures of the spread of data: variance, standard deviation, and range.

- **1.1: The Science of Statistics and Its Importance**

---

- **1.1.1: What is Statistics?**

[http://onlinestatbook.com/2/introduction/what\\_are.html](http://onlinestatbook.com/2/introduction/what_are.html)

- 

---

- **1.1.2: Descriptive and Inferential Statistics**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 1, Section 4 and 5: Descriptive Statistics and Inferential Statistics"URL](#)

Read sections 4 and 5 from chapter 1, and then complete the questions at the end of each section. Section 4 introduces descriptive statistics by using examples and discusses the difference between descriptive and inferential statistics. Section 5 talks about samples and populations, explains how one can identify biased samples, and defines differential statistics.

- [Introductory Statistics: "Chapter 1, Section 1: Basic Definitions and Concepts"URL](#)

Read section 1 from chapter 1 to further enhance your understanding of the elements of descriptive and inferential statistics. This section will introduce some of the key concepts in statistics and has numerous exercise and examples. Complete the odd-numbered exercises before checking the answers.

- **1.1.3: Types of Data and Their Collection**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 1, Section 7: Variables and Chapter 6, Section 4: Data Collection"URL](#)

Read section 7 from chapter 1 and [section 4 from chapter 6](#). Also, complete the questions at the end of each section. Section 7 will introduce several types of data and their distinguishing features. You will also learn about independent and dependent variables. Section 4 will explain how common data can be coded and collected.

- [Introductory Statistics: "Chapter 1, Section 3: Presentation of Data"URL](#)

Study section 3 from chapter 1. This reading talks about ways that data can be presented. Attempt the odd-numbered exercises on page before checking the answers.

- **1.2: Methods for Describing Data**

- **1.2.1: Graphical Methods for Describing Quantitative Data**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 2, Section 3 through 10"URL](#)

Read sections 3-7, 9, and 10 from chapter 2. Also, complete the questions at the end of each section. Section 3 provides an overview of the available methods to portray distributions of quantitative variables. Section 4 introduces you to the stem and leaf plot. In sections 5 and 6, you will learn how to capture the frequency of your data. Section 7 discusses box plots for the purpose of identifying outliers and for comparing distributions. Section 9 discusses bar charts for quantitative variables. Section 10 talks about the method of line graphs, which is based on bar graphs.

- [Introductory Statistics: "Chapter 2, Section 1: Three Popular Data Displays"URL](#)

Read section 1 from chapter 2. This reading further elaborates on ways of describing data. In particular, you will learn about the relative frequency histogram. Complete the odd-numbered exercises on before checking the answers.

- **1.2.2: Numerical Measures of Central Tendency and Variability**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 3, Section 2, 4, 8, 12-13"URL](#)

Read sections 2, 4, 8, 12, and 13 from chapter 3. Also, complete the questions at the end of each section. Section 2 defines the concept of central tendency. Section 4 introduces mean, median, and mode in the context of examples. Section 8 further elaborates on median and mean and discusses their strengths and weaknesses in measuring the central tendency. Section 12 addresses the concept of variability. Section 13 discusses range, interquartile range, variance, and the standard deviation.

- [Introductory Statistics: "Chapter 2, Section 2 and 3"URL](#)

Read sections 2 and 3 from chapter 2. Section 2.2 further elaborates on mean, median, and mode - both at the population level and sample level. This section contains many interesting examples and exercises. Section 2.3 talks about range, variance, and standard deviation using many examples. Complete the odd-numbered problems in the exercise sets for each section before checking the answers.

- [Khan Academy: "Mean, Median, and Mode; Variance"Page](#)

Watch this video series, which begins with a discussion on descriptive statistics and inferential statistics and then talks about mean, median, and mode, as well as sample variance.

- **1.2.3: Methods for Describing Relative Standing**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 1, Section 8: Percentiles"URL](#)

Read section 8 of chapter 1. Also, complete the questions at the end of the section. This reading discusses percentiles, which are useful for describing relative standings of observations in a dataset. This reading presents several definitions, so make sure to take notes.

o **1.2.4: Methods for Describing Bivariate Relationships**

- [Al Greene's "Scatter Plot/Bivariate Data"Page](#)

Watch this video tutorial to learn how to create the scatter plot for bivariate data, using two variables  $x$  and  $y$ . It may be useful to review the definitions on this webpage.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 4, Section 3, 5, and 6"URL](#)

Read sections 3, 5, and 6 from chapter 4. Also, complete the questions at the end of each section. Section 3 introduces Pearson's correlation and explains what the typical values represent. Section 5 further elaborates on the properties of  $r$ , particularly the fact that it is invariant under linear transformation. Section 6 introduces several formulas that can be used to compute Pearson's correlation.

## **Unit 2: Elements of Probability and Random Variables**

Probabilities affect our everyday lives. In this unit, you will learn about probability and its properties, how probability behaves, and how to calculate and use it. You will study the fundamentals of probability and will work through examples that cover different types of probability questions. These basic probability concepts will provide a foundation for understanding more statistical concepts, for example, interpreting polling results. Though you may have already encountered concepts of probability, after this unit, you will be able to formally and precisely predict the likelihood of an event occurring given certain constraints.

Probability theory is a discipline that was created to deal with chance phenomena. For instance, before getting a surgery, a patient wants to know the chances that the surgery might fail; before taking medication, you want to know the chances that there will be side effects; before leaving your house, you want to know the chance that it will rain today. Probability is a measure of likelihood that takes on values between 0 and 1, inclusive, with 0 representing impossible events and 1 representing certainty. The chances of events occurring fall between these two values.

The skill of calculating probability allows us to make better decisions. Whether you are evaluating how likely it is to get more than 50% of the questions correct on a quiz if you

guess randomly; predicting the chance that the next storm will arrive by the end of the week; or exploring the relationship between the number of hours students spend at the gym and their performance on an exam, an understanding of the fundamentals of probability is crucial.

We will also talk about random variables. A random variable describes the outcomes of a random experiment. A statistical distribution describes the numbers of times each possible outcome occurs in a sample. The values of a random variable can vary with each repetition of an experiment. Intuitively, a random variable, summarizing certain chance phenomenon, takes on values with certain probabilities. A random variable can be classified as being either discrete or continuous, depending on the values it assumes. Suppose you count the number of people who go to a coffee shop between 4 p.m. and 5 p.m. and the amount of waiting time that they spend in that hour. In this case, the number of people is an example of a discrete random variable and the amount of waiting time they spend is an example of a continuous random variable.

### **Completing this unit should take you approximately 25 hours.**

- Upon successful completion of this unit, you will be able to:
  - apply simple principles of probability, and use common terminology of probability;
  - calculate conditional probability, and determine whether two events are mutually exclusive and whether two events are independent;
  - calculate probabilities using the addition rules and multiplication rules;
  - construct and interpret Venn diagrams;
  - apply useful counting rules in the context of combinatorial probability;
  - identify and use common discrete probability distribution functions;
  - calculate and interpret expected values;
  - identify the binomial probability distribution, and apply it appropriately;
  - identify the Poisson probability distribution, and apply it appropriately;
  - identify and use continuous probability density functions; and
  - identify the normal probability distribution, and apply it appropriately.

- **2.1: Classical Probability Model**

- **2.1.1: Events, Sample Spaces, and Probability**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 5, Section 2 and 3"URL](#)

Read sections 2 and 3 from chapter 5. Also, complete the questions at the end of each section. Section 2 talks about experiments for which outcomes are equally likely to occur and also discusses the frequency approach to assign probabilities. Section 3 focuses on the concept of events and also touches upon the issue of conditional probability.

- [Introductory Statistics: "Chapter 3: Basic Concepts of Probability"URL](#)

Study chapter 3 to learn about basic concepts of probability. Section 1 discusses spaces, events, and their probabilities using many examples. Section 2 elaborates on sets operations, including complements, intersections, and unions using Venn diagrams. Section 3 introduces conditional probability and talks about independent events. Complete the odd-numbered exercises for each section before checking the answers.

- **2.1.2: Counting Rules**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 5, Section 6: Permutations and Combinations"URL](#)

Read section 6 from chapter 5. Also, complete the questions at the end of this section. Section 6 introduces formulas for combinations and permutations, which are useful to compute probabilities.

- [Khan Academy: "Probability with Playing Cards and Venn Diagrams" and "Addition Rule for Probability"Page](#)

Watch these videos, which introduce Venn diagrams in the context of playing cards and discuss the addition rule for probability.

- **2.2: Random Variables and Distributions**

- **2.2.1: Common Discrete Random Variables**

- [Introductory Statistics: "Chapter 4, Section 1 and 2"URL](#)

Read sections 1 and 2 from chapter 4. Section 1 defines discrete and continuous random variables. Section 2 introduces the distributions for discrete random variables. This section also talks about the mean and variance calculations. Complete the odd-numbered exercises for each section before checking the answers.

- [Khan Academy: "Binomial Distributions"Page](#)

Watch these videos on binomial distributions. The first explains how to compute the mean of a binomial distribution. The next two videos introduce binomial



probabilities and show how to graph them. The remaining videos elaborate on binomial distribution in the context of basketball examples.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 5, Section 8, 10, and 11"URL](#)

Read sections 8, 10, and 11 from chapter 5. Also, complete the questions at the end of each section. Section 8 talks about binomial probabilities, discusses how to compute their cumulatives, and introduces the mean and standard deviation. Section 10 introduces the Poisson probability formula. Section 11 defines multinomial outcomes and discusses how to compute probabilities by using the multinomial distribution.

Note: For Section 8, the link to the Binomial Calculator part way down the page may not work. If so, you can instead use this [Binomial Calculator](#).

### ○ 2.2.2: Normal Distribution

- [Introductory Statistics: "Chapter 5, Section 2: The Standard Normal Distribution"URL](#)

Read section 2 from chapter 5. This section talks about standard normal curve and how to compute certain areas underneath the curve. This section also contains numerous exercises and examples. Complete the odd-numbered exercises for this section before checking the answers.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 7, Section 3, 4, 6, and 7: History"URL](#)

Read sections 3, 4, 6, and 7 from chapter 7. Also, complete the questions at the end of each section. Section 3 briefly talks about the history of both the normal distribution and the central limit theorem, and this section also discusses the relation of normal distributions to errors. Section 4 discusses ways of computing the areas under the normal curve. Section 6 discusses the standard normal distribution and the related areas under the standard normal curve. Regarding the calculation of areas, Section 6 also discusses how to translate from non-standard normal to standard normal. Section 7 addresses how to compute (cumulative) binomial probabilities by using normal approximations.

- [Khan Academy: "Introduction to the Normal Distribution"Page](#)

Watch this video on normal distribution. This video introduces normal distribution and its density curve and explains how to read the areas underneath the normal curve. It also touches on the central limit behavior.

### **Unit 3: Sampling Distributions**

The concept of sampling distribution lies at the very foundation of statistical inference. It is best to introduce sampling distribution using an example here. Suppose you want to estimate a parameter of a population, say the population mean. There are two natural estimators: 1. sample mean, which is the average value of the data set; and 2. median, which is the middle number when the measurements are arranged in ascending (or descending) order. In particular, for a sample of even size  $n$ , the median is the mean of the middle two numbers. But which one is better, and in what sense? This involves repeated sampling, and you want to choose the estimator that would do better on average. It is clear that different samples may give different sample means and medians; some of them may be closer to the truth than the others. Consequently, we cannot compare these two sample statistics or, in general, any two sample statistics on the basis of their performance with a single sample. Instead, you should recognize that sample statistics are themselves random variables; therefore, sample statistics should have frequency distributions by taking into account all possible samples. In this unit, you will study the sampling distribution of several sample statistics. This unit will show you how the central limit theorem can help to approximate sampling distributions in general.

**Completing this unit should take you approximately 15 hours.**

- Upon successful completion of this unit, you will be able to:
  - apply the central limit theorem to approximate sampling distributions;
  - describe the role of sampling distributions in inferential statistics;
  - interpret and create graphs of a probability distribution for the mean of a discrete variable;
  - describe a sampling distribution in terms of repeated sampling;
  - define and compute the mean and standard deviation of the sampling distribution of population proportion  $p$ ;
  - identify or approximate a sampling distribution based on the properties of the population;
  - compare and evaluate the sampling distributions of different sample sizes; and
  - compare and evaluate the performance of different estimators based on their sampling distributions.

- **3.1: The Concept of Sampling Distributions**

- **3.1.1: Continuous Random Variables**

- [Introductory Statistics: "Chapter 5, Section 1, 3, and 4"URL](#)

From chapter 5, read section 1, section 3, and section 4. Section 1 talks about how to describe continuous distributions and compute related probabilities, including some basic facts about the normal distribution. Section 3 talks about how to compute probabilities related to any normal random variable. This section has many examples illustrating the usage of z-score transformations. Section 4 defines tail probabilities and illustrates how to find them. Complete the odd-numbered exercises at the end of each section before checking the answers.

- **3.1.2: Definition and Interpretation**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 9, Section 2: Introduction"URL](#)

Read section 2 from chapter 9. Also, complete the questions at the end. Section 2 introduces sampling distribution by using a concrete, discrete example, followed by a continuous example. This section also discusses sampling distributions' relationship to inferential statistics.

- **3.1.3: Sampling Distributions Properties**

- [Wolfram Demonstrations ProjectPage](#)

Use the information provided on the demonstration pages and interact with the various simulations.

If you have not done so already, you will need to download install the free [Wolfram CDF Player™](#). If using Chrome as your browser, you will also need to download the CDF files from the pages linked to above, and run them through the CDF Player on your desktop. Other browsers will allow you to interact with the demonstrations directly on the webpage.

- **3.2: Sampling Distributions for Common Statistics**

- **3.2.1: The Sampling Distribution of Sample Mean**

- [Rice University: David M. Lane's "Online Statistics Education: "Chapter 9, Section 6 and 7"URL](#)

Read sections 6 and 7 from chapter 9. Also, complete the questions at the end of each section. Section 6 discusses the mean and variance of the sampling distribution of the mean. This section also shows how central limit theorem can help to approximate the corresponding sampling distributions. Section 7 talks about the properties of the sampling distribution for differences between means by giving the formulas of both mean and variance for the sampling distribution.

Using the central limit theorem, it also talks about how to compute the probability of a difference between means being beyond a specified value.

- [Introductory Statistics: "Chapter 6, Section 1 and 2"URL](#)

Read sections 1 and 2 from chapter 6 . Section 1 presents several concrete examples to calculate the exact distributions of the sample mean. Based on these distributions, the corresponding means and standard deviations are computed for demonstrations. Section 2 concerns the sampling distributions of the sample means when the sample size is large. The case when the population is normal is also considered. The central limit theorem is used for large sample approximations. Complete the odd-numbered exercises at the end of each section before checking the answers.

- [Khan Academy's "Sampling Distribution"Page](#)

Watch these videos, which discuss sampling distributions.

- **3.2.2: The Sampling Distribution of Pearson's  $r$**

- [Rice University: David M. Lane's "Online Statistics Education: "Chapter 9, Section 8: Sampling Distribution of  \$r\$ "URL](#)

Read section 8 from chapter 9. Also, complete the questions at the end. Section 8 talks about how the shape of the sampling distribution of Pearson correlation deviates from normality and then discusses how to transform  $r$  to a normally distributed quantity. Furthermore, this section talks about how to calculate the probability of obtaining an  $r$  above a specified value.

- **3.2.3: The Sampling Distribution of the Sample Proportion**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 9, Section 9: Sampling Distribution of  \$p\$ "URL](#)

Read section 9 from chapter 9. Also, complete the questions at the end. Section 9 introduces the mean and standard deviation of the sampling distribution of  $p$ , and this section discusses the relationship between the sampling distribution of  $p$  and the normal distribution.

- [Kham Academy's: "Standard deviation"URL](#)

To enhance your understanding, watch this video on determining standard deviation.

#### **Unit 4: Estimation with Confidence Intervals**

In this unit, you will learn how to use the central limit theorem and confidence intervals, the latter of which enables you to estimate unknown population parameters. The central limit theorem provides us with a way to make inferences from samples of non-normal populations. This theorem states that given any population, as the sample size increases, the sampling distribution of the means approaches a normal distribution. This powerful theorem allows us to assume that given a large enough sample, the sampling distribution will be normally distributed.

You will also learn about confidence intervals, which provide you with a way to estimate a population parameter. Instead of giving just a one-number estimate of a variable, a confidence interval gives a range of likely values for it. This is useful, because point estimates will vary from sample to sample, so an interval with certain confidence level is better than a single point estimate. After completing this unit, you will know how to construct such confidence intervals and the level of confidence.

**Completing this unit should take you approximately 10 hours.**

- Upon successful completion of this unit, you will be able to:
  - explain the central limit theorem, and use it to construct confidence intervals;
  - compare t-distributions and normal distributions;
  - apply and interpret the central limit theorem for sample averages;
  - calculate, describe, and interpret confidence intervals for population averages and one population proportions; and
  - interpret the student-t probability distribution as the sample size changes.

- **4.1: Point Estimators and Their Characteristics**

- **4.1.1: Sample Statistics and Parameters**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 10, Section 2 and 3"URL](#)

Read sections 2 and 3 from Chapter 10. Also, complete the questions in each section. Section 2 explains the basic concepts of sample statistics and population parameters as well as the basic goal of estimation for which point estimates and interval estimates are introduced. Section 3 talks about the degree of freedom, which is defined as the number of independent pieces of information on which a point estimate is based. Section 3 also talks about variance, a quantity depending on the degrees of freedom.

- **4.1.2: Bias and Sampling Variability**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 10, Section 4: Characteristics of Estimators"URL](#)

Read section 4 from Chapter 10. Also, complete the questions at the end. Section 4 discusses two important characteristics used as point estimates of parameters: bias and sampling variability. Bias refers to whether an estimator tends to over or underestimate the parameter. Sampling variability refers to how much the estimate varies from sample to sample.

- **4.2: Confidence Intervals**

- **4.2.1: Confidence Intervals for Mean**

- [Rice University: David M. Lane's "Online Statistics Education: "Chapter 10, Section 7, 8, 9, and 11"URL](#)

Read sections 7, 8, 9, and 11 from Chapter 10. Also, answer the questions at the end of each section. Section 7 explains the need for confidence intervals and why a confidence interval is not the probability the interval contains the parameter. Section 8 explains how to compute a confidence interval on the mean when sigma is unknown and needs to be estimated. For this purpose, it also explains when to use t-distribution or a normal distribution. Section 9 states the difference between the shape of the t distribution and the normal distribution, and this section also explains how this difference is affected by degrees of freedom. Section 11 explains the procedure to compute a confidence interval on the difference between means.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 10, Section 10: Confidence Interval Simulation"URL](#)

This lab will help you develop a basic understanding of the properties of a sampling distribution, based on the properties of the population. Review the illustrated instructions, and follow the general instructions to learn more about confidence interval simulations.

Then, use the Wolfram Demonstration provided in the next resource for a more interactive practice of the topics covered here.

Note: Running the actual *Online Statistics Education* simulation yourself will likely not work, as the Java format that it is is no longer supported by most browsers.

- [Wolfram Demonstrations Project: Chris Boucher's "Confidence Intervals for a Mean"URL](#)

The demonstration provided here is a supplement to the reading above. Use the information provided on the demonstration page and interact with the simulation. If you prefer to watch a video of this simulation, you may do so [here](#).

Note: If you have not done so already, you will need to download install the free [CDF Player](#) from the Wolfram Demonstrations Project. If using Chrome as your browser, you will also need to download the CDF file from the page linked to above, and run it through the CDF Player on your desktop. Other browsers will allow you to interact with the demonstration directly on the webpage.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 12, Section 3: t Distribution Demo"URL](#)

Read the instructions and watch the video demo in order to see how the degrees of freedom affect the difference between t and normal distributions.

Note: Running the actual simulation yourself will likely not work, as the Java format that it is in is no longer supported by most browsers. Instead, you may use the Wolfram Demonstrations provided below for a more interactive practice of the topics covered here.

- [Wolfram Demonstrations Project: Gary McClelland's "Comparing Normal and Student's t-Distributions"URL](#)

The demonstration provided here are a supplement to the readings and videos above. Use the information provided on the demonstration page and interact with the simulation. If you prefer to watch a video of this simulation, you may do so [here](#).

Note: If you have not done so already, you will need to download install the free [CDF Player](#) from the Wolfram Demonstrations Project. If using Chrome as your browser, you will also need to download the CDF file from the page linked to above, and run it through the CDF Player on your desktop. Other browsers will allow you to interact with the demonstration directly on the webpage.

#### ○ **4.2.2: Confidence Intervals for Correlation and Proportion**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 10, Section 12 and 13"URL](#)

Read sections 12 and 13 from Chapter 10. Also, complete the questions at the end of each section. Section 12 shows how to compute a confidence interval for Pearson's correlation; the solution lies in using Fisher's z transformation. Section

13 explains the procedure to compute confidence intervals for population proportions, where the sampling distribution needs a normal approximation.

- [Khan Academy: "Confidence Intervals"Page](#)

Watch these two videos, which discuss confidence intervals.

## **Unit 5: Hypothesis Test**

A hypothesis test involves collecting and evaluating data from a sample. The data gathered and evaluated is then used to make a decision as to whether or not the data supports the claim that is made about the population. This unit will teach you how to conduct hypothesis tests and how to identify and differentiate between the errors associated with them.

Many times, you need answers to questions in order to make efficient decisions. For example, a restaurant owner might claim that his restaurant's food costs 30% less than other restaurants in the area, or a phone company might claim that its phones last at least one year more than phones from other companies. In order to decide whether it would be more affordable to eat at the restaurant that "costs 30% less" or another restaurant in the area, or in order to decide which phone company to choose based on the durability of the phone, you will have to collect data to justify these claims. The process of hypothesis testing is a way of decision-making. In this unit, you will learn to establish your assumptions through null and alternative hypotheses. The null hypothesis is the hypothesis that is assumed to be true and the hypothesis you hope to nullify, while the alternative hypothesis is the research hypothesis that you claim to be true. This means that you need to conduct the correct tests to be able to accept or reject the null hypothesis. You will learn how to compare sample characteristics to see whether there is enough data to accept or reject the null hypothesis.

**Completing this unit should take you approximately 12 hours.**

- Upon successful completion of this unit, you will be able to:
  - differentiate between type I and type II errors, and find the probability of these errors;
  - describe and conduct hypothesis testing, calculate the p-value, and accept or reject the null hypothesis; and
  - explain how to conduct hypothesis tests for a single population mean and population proportion, when the population standard deviation is unknown; perform this task; and interpret the results.

### **5.1: Elements of Hypothesis Testing**

#### **5.1.1: Setting up Hypotheses**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 11, Section 2, 4, and 5"URL](#)



Read sections 2, 4, and 5 from Chapter 11, and complete the questions at the end of each section. Section 2 discusses the logic behind hypothesis testing using concrete examples and explains how to set up null and alternative hypothesis. Section 4 explains what Type I and II errors are and how they can occur. Section 5 introduces one-tailed and two-tailed tests and explains which one should be used for the testing purpose.

### ○ **5.1.2: Interpreting Hypotheses Testing Results**

- [Introductory Statistics: "Chapter 8, Section 3: The Observed Significance of a Test"URL](#)

Read section 3 from Chapter 8. This section explains what the observed significance of a test is; in particular, this reading tells us how to compute it and use it in the p-value approach. Study the examples, and complete the odd-numbered exercises at the end of the section before checking the answers.

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 11, Section 6 and 7"URL](#)

Read sections 6 and 7 from Chapter 11, and complete the questions at the end of each section. Section 6 discusses whether rejection of the null hypothesis should be an all-or-none proposition. Section 7 discusses how to interpret non-significant results; for example, it explains why the null hypothesis should not be accepted, or accepted with caution. This section also describes how a non-significant result can increase confidence that the null hypothesis is false.

- [Boundless: Statistics: "Chapter 12, Section 5: Hypothesis Testing: One Sample"URL](#)

Locate and read the section titled, Type I and Type II Errors, on the linked page, which discusses two types of errors in hypothesis testing, using numerous examples.

- [Khan Academy: "Hypothesis Testing"Page](#)

Watch these videos on hypothesis testing.

### ○ **5.1.3: Steps in Hypothesis Testing and Its Relation to Confidence Intervals**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 11, Section 8 and 9"URL](#)

Read sections 8 and 9 from Chapter 11. Also, complete the questions at the end of each section. Section 8 lists four key steps in hypothesis testing. Section 9 explains the close relationship between confidence intervals and significance tests.

- [Sophia: Al Greene's "Hypothesis Testing"URL](#)

Watch the videos titled "Significance Level in Hypothesis Testing" and "Hypothesis Testing Example." The first video shows you how to make a decision in a hypothesis test based on the significance level of critical values. The second video provides an example of a left tailed hypothesis test.

- **5.2: Tests of Population Means**

---

- **5.2.1: Testing Single Mean**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 12, Section 2: Single Mean"URL](#)

Read section 2 from Chapter 12. This section shows how to test the null hypothesis that the population mean is equal to some hypothesized value, using a very concrete example. In this example, all the main elements of hypothesis testing come in to play a role. There are 9 questions at the end of the section to help your understanding of the material.

- [Introductory Statistics:"Chapter 8, Section 2, 4, and 5"URL](#)

Read section 2, section 4, and 5. Complete the odd-numbered problems at the end of each section before checking your answers.

Section 2 talks about how to use the central limit theorem to test a population mean when the sample size is large. It also addresses how to interpret the test results in the application background. Section 4 discusses testing a population mean when the sample size is small. This section outlines a five-step testing procedure and then illustrates this procedure with an example. Study the example carefully and complete the relevant exercises and applications. Section 5 talks about large sample tests for a population proportion. Both the critical value and p-value approach are introduced based on a standardized test statistic. Once again, this section illustrates the five-step testing procedure in Examples 12-15.

- **5.2.2: Testing the Difference between Two Means**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 12, Section 4: Difference between Two Means"URL](#)

Read section 4 from Chapter 12. Also, answer the questions at the end of this section. This section covers how to test for differences between means from two separate groups of subjects. This reading presents an example of opinions on animal research, and the main interest is to test for gender difference at the population level. The detailed testing procedure is carried out by using the standard steps in hypothesis testing.

- [Khan Academy: "Difference of Means"Page](#)

Watch these videos on the difference of means.

- **5.3: Chi-Square Distribution**

---

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 17, Section 5: Contingency Tables"URL](#)
- 

Read this section, which discusses contingency tables, and answer the questions at the end of the section. While this section is optional, studying it may help you if you wish to take the credit-aligned exam that is linked with this course.

---

- [Rice University: David M. Lane's "Online Statistics Education: "Chapter 17, Section 2 and 3"URL](#)
- 

Read these two sections, which discuss chi-square distributions and how to test goodness of fit. Also, answer the questions at the end of each section. While these sections are optional, studying them may help you if you wish to take the credit-aligned exam that is linked with this course.

---

- [Khan Academy: "Chi-square Distributions"Page](#)
- 

Watch these videos, which discuss chi-square distributions, goodness of fit, and contingency tables.

---

- **5.4: Comparing the Proportions of Populations**

---

- [Khan Academy: "Comparing Population Proportions"Page](#)
- 

Watch these videos, which discuss comparing population proportions. While these videos are optional, studying these topics may help you if you are interested in taking the credit-aligned exam that is linked with this course.

---

## **Unit 6: Linear Regression**

In this unit, we will discuss situations in which the mean of a population, treated as a variable, depends on the value of another variable. One of the main reasons why we conduct such analyses is to understand how two variables are related to each other. The most common type of relationship is a linear relationship. For example, you may want to know what happens to one variable when you increase or decrease the other variable. You want to answer questions such as, "Does one variable increase as the other increases, or does the variable decrease?" For example, you may want to determine how the mean reaction time of rats depends on the amount of drug in bloodstream.

In this unit, you will also learn to measure the degree of a relationship between two or more variables. Both correlation and regression are measures for comparing variables. Correlation quantifies the strength of a relationship between two variables and is a measure of existing data. On the other hand, regression is the study of the strength of a linear relationship between an independent and dependent variable and can be used to predict the value of the dependent variable when the value of the independent variable is known.

### **Completing this unit should take you approximately 12 hours.**

- Upon successful completion of this unit, you will be able to:
  - discuss and apply basic ideas of linear regression and correlation;
  - identify the assumptions that inferential statistics in regression are based on;
  - compute the standard error of a slope;
  - test a slope for significance;
  - construct a confidence interval on a slope; and
  - calculate and interpret the coefficient of determination and the correlation coefficient.

- **6.1: The Regression Model**

---

- **6.1.1: Scatter Plot of Two Variables and Regression Line**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 14, Section 2: Introduction to Linear Regression"URL](#)

Read section 2 from Chapter 14. Also, answer the questions at the end. Section 2 defines simple linear regression, introduces scatter plot to reveal linear patterns, and then talks about prediction error. This section also talks about how to compute regression line by minimizing squared errors.

- [OpenStax College: Susan Dean and Barbara Illowsky's "Collaborative Statistics, 12-1 through 12-5"URL](#)

Read these sections on linear regression. Linear regression, the simplest form of regression, is used to obtain a linear relationship between two variables.

Be sure to click "next" and read each section.

### ○ **6.1.2: Correlation Coefficient**

- [OpenStax College: Susan Dean and Barbara Illowsky's Collaborative Statistics, "12-6 through 12-11"URL](#)

Read these sections on correlation. You will learn the interpretation and calculation of the correlation coefficient, how to test its significance, and the relation between correlation and causation.

Be sure to click "next" and read each section.

- [Introductory Statistics: "Chapter 10, Section 2: The Linear Correlation Coefficient"URL](#)

Read section 2 of Chapter 10 for a discussion on linear correlation. You will learn what the linear correlation coefficient is, how to compute it, and what it tells us about the relationship between two variables  $x$  and  $y$ .

### ○ **6.1.3: Sums of Squares**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 14, Section 4: Partitioning Sums of Squares"URL](#)

Read section 4 from Chapter 14. Also, answer the questions at the end of the section. Section 4 further discusses the sums of squares, including partitioning sum of squares into sums of squares predicted and sum of squares error.

- [Khan Academy: "Regression Lines"Page](#)

Watch these videos, which discuss the regression line.

- **6.2: Fitting the Model**

- **6.2.1: Standard Errors of the Least Squares Estimates**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 14, Section 5: Standard Error of the Estimate"URL](#)

Read section 5 from Chapter 14. Also, answer the questions at the end. Section 5 discusses how to compute the standard error of the estimate based on errors of prediction as well as how to compute the standard error of the estimate based on a sample.

- **6.2.2: Statistical Inference for the Slope and Correlation**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 14, Section 6: Inferential Statistics for b and r"URL](#)

Read section 6 from Chapter 14. Also, answer the questions at the end. Section 6 starts with assumptions on the errors that are necessary for statistical inference. Then, this reading shows an example of a significance test for the slope. This section also talks about constructing confidence intervals for the slope. Then, it closes with a significance test for the correlation.

- [Introductory Statistics: "Chapter 10, Section 5: Statistical Inference about Slope"URL](#)

Read section 5 from Chapter 10. This section further details two types of inferences on the slope parameter, considering both confidence intervals and hypothesis testing. Complete the odd-numbered exercises at the end of the section before checking your answers.

- **6.2.3: Influential Observations**

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 14, Section 7: Influential Observations"URL](#)

Read section 7 from Chapter 14. Also, answer the questions at the end. Section 7 discusses the notion of influence and describes what makes a point influential. It further introduces the concepts of leverage and distance, which are useful to detect influential observations.

- [Introductory Statistics: "Chapter 10, Section 8: A Complete Example"URL](#)

Read section 8 from Chapter 10. This section presents a complete example on linear regression, starting from presenting the data, then proceeds to a scatter plot to identify the linear pattern, and fits a linear model using least squares estimation. This reading also addresses some statistical inferences on both correlation coefficient and slope parameter. Complete the odd-numbered exercises at the end of the section before checking the answers.

- **6.3: ANOVA**

---

This optional subunit will teach you about "Analysis of Variance" (abbreviated ANOVA), which is used for hypothesis tests involving more than two averages. ANOVA is about examining the amount of variability in the y variable and trying to see where that variability is coming from. You will study the simplest form of ANOVA, called single factor or one-way ANOVA. Finally, you will briefly study the F distribution, used for ANOVA, and the test of two variances.

---

- [Khan Academy: "ANOVA"Page](#)
- 

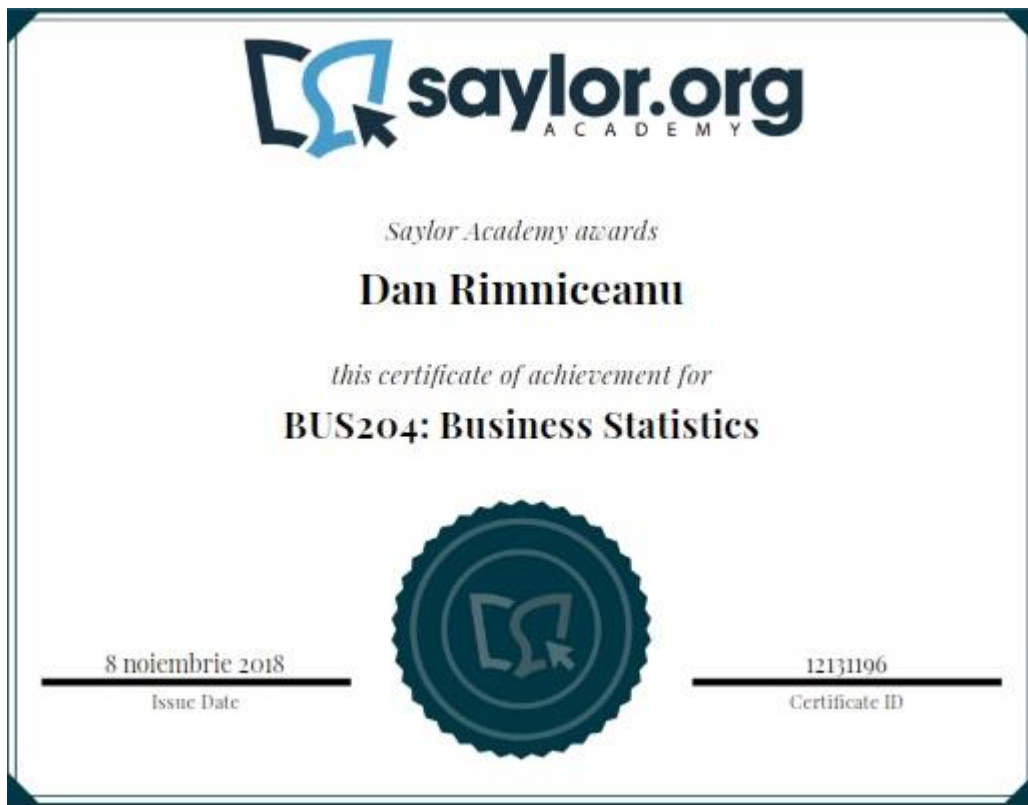
Watch these videos, which discuss each of the steps in ANOVA. While these videos are optional, studying ANOVA may help you if you are interested in taking the credit-aligned exam that is linked with this course.

---

- [Rice University: David M. Lane's "Online Statistics Education: Chapter 15: ANOVA"URL](#)
- 

Read this chapter and complete the questions at the end of each section. While these sections are optional, studying ANOVA may help you if you are interested in taking the credit-aligned exam that is linked with this course.

---



## BUS204: Business Statistics

This course will introduce you to business statistics, or the application of statistics in the workplace. Statistics is how we gather, analyze, and interpret data. If you have taken a statistics course in the past, you may find some of the topics in this course familiar. You can apply statistics to any number of fields, from anthropology to hedge fund management, because many of us interpret data best when it is presented in an organized fashion.

Here, we will look at summary statistics, which give an overview of a data set, such as the average score on an exam. However, the average does not always tell the entire story; since half of the students could have gotten 100 on the exam, and the other half gotten 60. Using statistics, we can learn a lot more about how data is organized. To do that, we will use statistical tools to analyze data, draw conclusions, and make predictions of the future. The course will begin with data distributions, followed by probability analysis, sampling, hypothesis testing, inferential statistics, and regression.

### **Unit 1: Introduction to Statistical Analysis**

Even if you haven't taken a statistics course before, you are already familiar with the fundamentals of statistics from your everyday life. For instance, you already know that the majority of adult males have the same shoe size (which is very close to the average size), and that there are a few adult males on both sides of the average who have smaller or larger shoe sizes. In statistics, we call this phenomenon the "normal distribution".



This unit will introduce you to statistical analysis and how it relates to business. For example, you may be interested in learning about the average price of a 50-inch TV by gathering price data from 30 different stores. You would then take your 30 prices and compute the average price. Given the fact that there are thousands of stores that are selling that particular product, the next question in statistics is: are you confident enough to say that your computed average is reflective of the real average that you would get if you looked at the price for that TV at every possible store?

You are probably familiar with the average of a data set. In this course, we will refer to what most people call the average as the "arithmetic mean". The average is actually any single value used to describe the middle of a data set. The most common averages used in statistics are the arithmetic mean, the median, and the mode. Each describes the middle of a dataset in different ways. The median is the numeric value that separates the upper and lower half of a data set. The mean is the sum of all values divided by the number of values. The mode is the most common value within the dataset.

In many instances, the median and the mean are similar, but we will also talk about many examples where it is not. The distinction between these kinds of summary statistics is important in business statistics. Understanding this vocabulary will be vital to your success in this course and in the business world.

**Completing this unit should take you approximately 6 hours.**

- Upon successful completion of this unit, you will be able to:
  - explain the importance of statistics to business;
  - explain the differences between quantitative and qualitative data, and identify examples of each type of data;
  - define and apply the following terms: data sets, mean, median, mode, standard deviation, variance, population and sample;
  - summarize and interpret data in a tabular format using frequency distributions and visually with histograms; and
  - explain how to use a spreadsheet to describe data sets, create frequency tables, and draw histograms, as well as perform these tasks.

- **1.1: Why Do We Need to Study Statistical Analysis as Part of a Business Program?**

---

- [Why Do We Need to Study Statistical Analysis as Part of a Business Program?Page](#)
- 

Read this article about the different ways that statistics are used in business, and why it is essential that decision-makers have the tools to analyze data as part of their skill set.

---

- **1.2: Measuring Data**

---

- [Definitions of Statistics, Probability, and Key TermsPage](#)

---

This article introduces how we present a summary of data through graphs, tables, and numerical measures such as the average. This will be helpful in terms of analyzing business data in a simple way with the help of the widely-used methods in statistical analysis.

---

- [Kinds of Data in StatisticsPage](#)

---

Watch the first video from 1:03:00 to the end, and the second video until 54:00. These videos explain the foundations of statistics, what data is, and the various types of data we'll be exploring.

---

- **1.3: Measures of Spread and Data**

- [Variance and Standard DeviationPage](#)

---

These videos explain the difference between the variance of a population of a sample, how to estimate the variance of a population based on a sample, and how to find the standard deviation and why it is important.

---

- [Descriptive StatisticsURL](#)

---

Read chapter 2, which discusses how to describe locations within a sample, and how to analyze data from a sample. Make sure you read the introduction as well as sections 2.1 through 2.7. Be sure to attempt the practice problems and homework at the end of each section.

If you would prefer to download this textbook, you may do so [here](#).

---

- **1.4: Spreadsheet Exercises: Measures of Central Tendency and Spread**

- [Measures of Central Tendency: Mode, Median, Mean, and MidrangePage](#)

---

This section explores the measures of central tendency: mode, median, mean and midrange. It shows you the commands for computing these metrics using a spreadsheet program and it gives you the correct format for entering those commands into the spreadsheet program.

---

- [Practice: Measures of Middle and SpreadQuiz](#)

Attempt these questions, which will give you a chance to practice the materials covered in this subunit. You may wish to use a spreadsheet program, like Microsoft Excel, OpenOffice, or Google Sheets to perform calculations on the provided datasets.

---

- **1.5: Spreadsheet Exercises: Graphs of Histograms and Frequency Tables**

---

- [Graphs and ChartsPage](#)
- 

Read this section.

---

- [Shapes of DistributionsPage](#)
- 

Read this section. Note that the instructions given in the text are for version 2.x of OpenOffice. If you have 3.x, some steps are slightly different; you may need to consult the help documents.

---

## **Unit 2: Counting, Probability, and Probability Distributions**

How likely is it that a certain event will occur? What are the chances that a given student will receive a grade of 60–69 on their exam? By studying distributions of data, you can determine the probability that a certain event will occur. By looking at the distribution of grades in a class, you can identify the probability that a student will receive between a 60 and 69. Probability is used in business to predicting profits, determine the chances that a business model will affect regulation, and in many other ways.

Before you can focus on probability, you must first learn how to count. What's that? You already know how to count? Maybe – but in this unit, you will learn how to count the different ways that multiple events can occur together. These are called *combinations* and *permutations*, and they are a fundamental concept of probability.

**Completing this unit should take you approximately 12 hours.**

- Upon successful completion of this unit, you will be able to:
  - identify values of and differentiate between permutations and combinations;
  - explain and apply the different methods for determining probability: equally likely outcomes, frequency theory, and subjective theory;
  - define and apply the axioms of probability theory;
  - apply probability distributions and explain the properties of different distributions;

- solve problems using binomial distribution, and explain when it should be used;
- differentiate between discrete and continuous probability distributions; and
- apply expected value and calculate it for various probability distributions.

- **2.1: Counting**

---

- [TerminologyURL](#)
- 

Read this introduction to some of the most common formulas and terms in probability.

---

- [Counting and ProbabilityPage](#)
- 

Watch this lecture, in which Professor Stark works through several examples of counting and probability.

---

- **2.2: Theories of Probability**

---

- [Independent and Mutually Exclusive EventsURL](#)
- 

Read this section, which explains how to categorize the probability of an event based on what you know about the variables involved.

---

- [Venn DiagramsURL](#)
- 

This section explains some of the most common theories and formulas in probability, and demonstrates set theory by using Venn diagrams.

---

- **2.3: Set Theory**

---

- [Probability with Playing Cards and Venn DiagramsPage](#)
- 

Watch this video, which demonstrates how to use Venn Diagrams to understand probability.

---

- [Addition Rule for ProbabilityPage](#)
-

Watch this video, which discusses Venn diagrams and the addition rule for probability.

---

- [Two Basic Rules of ProbabilityURL](#)
- 

Read this section, which covers the symbols used in set theory. For example, the union of two sets A and B is denoted as  $A \cup B$ . This section also discusses the basic rules of probability and set theory.

---

- [Set TheoryPage](#)
- 

Watch these lectures, in which Professor Stark discusses set theory and works through a series of examples.

---

- **2.4: Probability Fundamentals**

---

- [Properties of Continuous Probability Density FunctionsURL](#)
- 

Read this section on the probability density function, which is the foundation for how we understand probability.

---

- [Probability FundamentalsPage](#)
- 

Watch these lectures, in which Professor Stark works through several examples of how to approach solving problems related to probability.

---

- **2.5: Probability Distributions and the Binomial Distribution**

---

- [Probability Density Functions and Random VariablesPage](#)
- 

Watch these videos, which will introduce you to probability distributions and random variables.

---

- [Discrete Random VariablesURL](#)
- 

Read this chapter, which covers the basic rules of probability and the ways that randomness affects how probabilities are distributed. Make sure you read the introduction and sections 4.1 through 4.4. Attempt the practice problems and homework at the end of each section.

---

- [Probability DistributionsPage](#)
- 

Watch this lecture, in which Professor Stark works through several examples of probability distributions. The second video gives more information on how to find the mean (or expected value) of a discrete random variable.

---

### **Unit 3: The Normal Distribution**

A distribution is a line graph representation of the probability that an event will occur. It is similar to a histogram, but in a distribution, the user does not determine the grouping; instead, data is grouped according to the likelihood that it will occur within the data set. Distributions also allow for analysis of a specific event, whereas a histogram requires events be grouped.

An important type of this distribution is the "normal" distribution. The normal distribution is used to approximate real-world occurrences. If you can make certain assumptions about the occurrence of an event, then you can use the normal distribution to find out the probabilities of that event occurring. Many of the events that are important to business can be approximated using the normal distribution.

**Completing this unit should take you approximately 3 hours.**

- Upon successful completion of this unit, you will be able to:
  - define and apply the properties of the normal distribution while understanding real-world implications and applications of the normal distribution;
  - use the normal distribution to estimate the probability of an event occurring; and
  - explain how the normal distribution relates to the central limit theorem.

#### **3.1: The Normal Distribution**

---

- [Qualitative Sense of Normal DistributionsPage](#)
- 

Watch this video, which gives an intuitive explanation of variables in real life that follow a normal distribution.

---

- [The Central Limit TheoremURL](#)
- 

Read this chapter, which covers some of the most important concepts used in statistics: the central limit theorem and the normal distribution. Make sure you read the introduction and sections 7.1 through 7.4. Attempt the practice problems

and homework at the end of the chapter, which will give you a chance to check your understanding of these concepts.

---

- [Normal Distribution Problems: Z-scorePage](#)
- 

Watch this video, which will walk you through the process of finding Z-scores, an important part of understanding this topic.

---

- [More Empirical Rule and Z-score PracticePage](#)
- 

Watch this video, will walk you through the process of solving problems using the standard normal distribution, an important part of understanding this topic.

---

#### **Unit 4: Sampling and Sampling Distributions**

While you may not become a professional data gatherer, it is likely that you will need to compile data on a regular basis. When gathering data, you will not always have the luxury of collecting all available data. For example, economists cannot measure the entire unemployment of the population, so they must take a random sample instead. Likewise, in a manufacturing facility, quality control managers do not have the resources to test every product that comes off the line; it is simply not feasible. Instead, they take samples at various points during the production process to test the quality of the products the firm produces.

There are a number of methods employed in sampling data. It is important that the sampling method fits the application. For example, marketing managers may wish to test a product on various groups of people. They may define these groups by age, race, geography, income, or any other factors. They then divide the population into these groups and take samples from each group in a process known as cluster sampling. If marketers do not properly divide the population, they may end up marketing to the wrong demographic and achieving poor sales.

**Completing this unit should take you approximately 3 hours.**

- Upon successful completion of this unit, you will be able to:
  - differentiate the population from a sample;
  - define and apply simple random sampling;
  - determine different types of selection bias and sampling errors, and explain how to avoid these errors in survey sampling, such as selection and estimation errors;
  - describe and identify the different sampling methods, including systematic, stratified random, cluster, convenience, panel, and quota sampling, and identify an example of each; and
  - use a point estimator from a sample to estimate the entire population.

- **4.1: Sampling and Sampling Distributions**

---

- [Sampling Distribution of the Sample MeanPage](#)
- 

Watch these videos, which explain the nature of sampling distribution and how it changes as sample size changes.

---

- [Calculating the Sample Size  \$n\$ : Continuous and Binary Random VariablesURL](#)
- 

Read this section, which explains sampling in statistics and discusses some of the possible biases when collecting data for a sample. It also explains what to consider when dealing with continuous versus binary random variables.

---

- [Sampling and Sampling DistributionsPage](#)
- 

Watch the first lecture from 1:19:00 to the end, in which Professor Stark covers some problems related to sampling and sampling distributions. Then, watch the second lecture to see some additional examples.

---

### **Unit 5: Estimation and Hypothesis Testing**

Estimation is the process of making predictions based on the best available information. Businesses employ estimation in order to help managers make decisions regarding the future. For example, if the CFO estimates profits will be lower next year, the CEO will consider cost-cutting measures to make up for the loss. Normally, companies do not want to pursue aggressive cost-cutting because it usually comes in the form of layoffs, which are bad for employee morale.

In order to make accurate estimates, companies use hypothesis testing. For example, assume the CFO thinks profits will be below 5% of revenue next year. His null hypothesis is that profits will be 5% or greater next year. The alternative hypothesis is that profits will not be 5% or greater next year. This seems counter-intuitive, but statistics proposes that a hypothesis cannot be proven true; it can only be rejected, or shown to be not true. Through the hypothesis testing process, the CFO will either reject or accept the null hypothesis. Hypothesis tests are always framed in this manner because, with imperfect information, nothing can be proven.

The best non-business analogy to hypothesis testing comes from the courtroom. In the United States, a defendant is presumed innocent until proven guilty. The null hypothesis in this scenario is innocent or not guilty. The alternative hypothesis is guilty. In order to find the defendant guilty, the jury must be offered enough evidence that suggests the defendant is guilty beyond a reasonable doubt. If the members of the jury make that decision, then they reject the null hypothesis. If the jury members decide they do not have enough evidence to make that judgment, then they must find the defendant not



guilty. Notice not guilty does not mean the jury claims the defendant is innocent. The decision simply means the members of the jury do not have enough information to find the person guilty, so they err on the side of caution and fail to reject the null hypothesis. As an aside, in this example, beyond a reasonable doubt is analogous to the level of significance, which you will learn is crucial to hypothesis testing.

**Completing this unit should take you approximately 11 hours.**

- Upon successful completion of this unit, you will be able to:
  - estimate intervals over which the population parameter could exist;
  - determine and differentiate between the null and alternative hypotheses in hypothesis testing;
  - identify when to use the z and t distributions, and use these distributions to find probabilities;
  - test hypotheses of the population mean and population proportion using one or two samples;
  - define and apply the significance level, and explain its importance to hypothesis testing; and
  - compute a test statistic and determine a region of acceptance based on a test statistic.

- **5.1: Estimation and Confidence Intervals**

---

- [Confidence IntervalsURL](#)
- 

Read this chapter, which discusses how to construct a confidence interval for a given population. Make sure you read the introduction as well as sections 8.1 through 8.4. Attempt the practice problems and homework at the end of the chapter.

---

- [Confidence Intervals and Estimating ParametersPage](#)
- 

Watch the first lecture from 1:05:00 to the end to learn more about confidence intervals and estimating parameters. Then watch the second lecture, which goes into more detail about confidence intervals and how to use them.

---

- [Computing Confidence IntervalsPage](#)
- 

Read this section to learn how to compute confidence intervals for finding a range for the real population parameter using statistics from the sample data.

---

- **5.2: Hypothesis Testing**

---

- [Hypothesis Testing and P-valuesPage](#)

Watch this video, which explains how to test a hypothesis.

- [Hypothesis Testing with One SampleURL](#)

This chapter builds on your knowledge of confidence intervals to introduce you to the concept of hypothesis testing, which is how statisticians use the scientific method to learn more about the populations they are studying. Make sure you read the introduction as well as sections 9.1 through 9.4. Attempt the practice problems and homework at the end of each section.

- [Hypothesis TestingPage](#)

Watch the first lecture from 1:12:00 to the end. In it, Professor Stark covers problems related to hypothesis testing. Then, watch the second lecture, in which Professor Stark goes through additional problems related to hypothesis testing.

### • 5.3: Testing Equality of Two Percentages

- [Comparing Population ProportionsPage](#)

Watch these videos, which explain how to compare the proportion of two different samples.

- [Hypothesis Testing with Two SamplesURL](#)

Read this chapter, which discusses how to compare data from two similar groups. This is useful when, for example, you want to analyze things like how someone's income relates to another sample that you are interested in. Make sure you read the introduction as well as sections 10.1 through 10.6. Attempt the practice problems and homework at the end of the chapter.

### • 5.4: The Chi-Squared Test for Goodness of Fit

- [Introduction to the Chi-Square DistributionPage](#)

Watch these videos, which introduce chi-square tests and show when each kind of chi-squared test is used.

---

- [The Chi-Square DistributionURL](#)
- 

Read this chapter, which introduces you to the three major uses of the chi-squared distribution: the goodness-of-fit test, the test of independence, and the test of a single variance. Make sure you read the introduction as well as sections 11.1 through 11.6. Attempt the practice problems and homework at the end of the chapter.

---

## **Unit 6: Correlation and Regression**

If two data points move in the same direction, does that mean that one causes the other? How are we to analyze their correlation?

Regression is an analysis of the relationship of one variable to another. A regression might identify, for example, the relationship between car speed and the number of fatal accidents. In this example, speed and number of accidents are the two variables; the number of accidents is said to be the dependent variable, because the number of accidents depends on the speed. Speed is considered the independent variable. While regressions can be calculated manually, a statistically significant data set could take a long time to regress.

Regressions not only allow us to determine whether a relationship exists but also to identify how strong that relationship is. The measure of this relationship is known as the regression coefficient. If the regression coefficient is relatively low, then speed may not be the major factor in fatal accidents. Perhaps the major factor is the time of day, whether it rained or not, or if alcohol was involved. With multiple regression, a number of independent variables can be tested against the dependent variable at the same time. The regression coefficient would determine which variables have the strongest relationship with the dependent variable. In business, you will frequently use regression to predict future events. Though not an exact science, regression can be used to make reliable predictions if enough variables are identified. For example, first responders could use regression outputs to predict the number of fatal accidents in a given shift based on average travel speed, time of day, weather, and any other factors deemed significant. This unit will also stress the importance of determining the factors that most likely contribute to a dependent variable.

Regression is often used in finance. Investors often want to know the relationship between a stock's performance and the overall performance of the market. By regressing the period returns of a stock with the returns of the market, investors can see the regression coefficient. This coefficient is known as a stock's beta and is covered extensively in [BUS202: Principles of Finance](#).

**Completing this unit should take you approximately 5 hours.**

- Upon successful completion of this unit, you will be able to:
  - identify the dependent and independent variables in the linear regression model;
  - calculate the equation of the regression line, and plot it;
  - describe the importance of the correlation coefficient and r-squared, and apply these concepts;
  - define outlier, identify examples of outliers, and describe what an outlier can do to summaries of data;
  - estimate a regression line and identify the effect of the independent variable on the dependent variable; and
  - draw a scatter plot, explain how to use a spreadsheet to draw a scatter plot, find the equation of the least-squared line, and draw the line.

## • 6.1 Working with More Than One Variable

---

- [Linear Regression and CorrelationURL](#)

Read this chapter to learn how to use graphs, such as scatter plots, to analyze the relationship between two variables. Two variables may be positively or negatively related when different pairs of data show the same pattern. For example, when incomes of individuals rise so does their consumption of goods and services; thus, income and consumption are considered to be positively related. As a person's income rises, the number of bus rides this person takes falls; thus, income and bus riding are negatively related. Make sure you read the introduction as well as sections 13.1 through 13.7. Attempt the practice problems and homework at the end of the chapter.

---

- [Examples of Univariate and Multivariate DataPage](#)

Watch the first lecture from 0:50:00 to the end. In it, Professor Stark differentiates between univariate and multivariate data. It also covers different data types and how to plot and interpret the correlation between data variables, and works through some examples. Then, watch the second lecture until 0:38:00, in which he works through some additional examples.

---

## • 6.2: Correlation and Association

---

- [The Correlation CoefficientURL](#)

Read this section, which describes the formula for computing the correlation coefficient. It may be useful to save this resource for future reference to this formula. Note that the formula uses the Greek letter sigma,  $\Sigma$ , as the summation symbol. For instance,  $\sum_{i=1,2,3} x_i = x_1 + x_2 + x_3$

---

- **6.3: Regression**

---

- [Linear RegressionPage](#)
- 

Watch this lecture, which discusses how to interpret and understand a linear regression and how regression equations enable you to make predictions.

---

- **6.4: Spreadsheet Activity for Unit 6**

---

- [Paired Data and Scatter DiagramsPage](#)
- 

Read this chapter, which discusses linear regressions and best fit lines.

---

- [Spreadsheet ActivityFolder](#)
- 

For this activity, you will review how a spreadsheet can be used to plot data, determine the slope and intercept of regression line, and draw the regression line. The instructions for creating the scatter graph and regression line are in Section 4.25. However, for this activity, we are solving for the problem presented in Section 4.4. The supporting spreadsheet files (links above to both Excel and Open Office versions) include a tab titled "Starter File", which contains everything you need to get started on the activity. Once you have worked through the activity, you can click on the "Solution File" tab to see how your finished spreadsheet should look.

---