

## Syllabus

### Course Overview

The course's objective is to provide you with both the theory and practices of descriptive, inferential, and predictive statistics with the focus on applications in data mining and business analytics. What we study here should enable you to apply your skill set to your current or future work on business analytics. Statistical methods use data to produce descriptive, inferential, and predictive result. It is imperative to understand that the results obtained through these methods are as good as the data from which the conclusions are drawn. In data mining we use descriptive statistics as analytical tools to add to our current knowledge.

Traditionally, "statistics is the science of collecting, describing, and analyzing the data" (Lock, Lock, Lock, Lock, and Lock, 2013, p. 4). According to Mirkin, this is done by creating a mathematical model that fits and tests the data in a predefined way. Data mining, on the other hand, uses data to add new knowledge and insights by finding patterns that may be useful to the organization (Mirkin, 2011).

In this course we will be using two types of variables—categorical (C) and quantitative (Q). Categorical variables divide cases into one or more groups. Quantitative variables are numeric and additive. Data mining and analytics tends to focus on group cases, whereas statistics tends to focus mostly on quantitative summarization. Both methods and variables are important to the business analyst.

In addition to descriptive statistics, we will be deploying inferential and predictive statistical methods for business analytics. When a business analyst encounters a problem, he or she goes through the sequences of problem definition, data requirements, methodology, testing, and solution. The three last steps involve the use of statistical methods for analytics. Sometimes the process of making sense of data is embedded in the problem definition.

Through the application of statistical methods in analytics, we can examine what happened (descriptive statistics); why it happened (inferential and diagnostics statistics); what will happen in the future (predictive statistics); and what should happen (prescriptive statistics). In all statistical methods, some degree of likelihood and confidence level should be attached, which is designed to reduce the degree of uncertainty (Albright & Winston, 2013). During this course, you will learn statistical methods which are applied in business analytics. We will study descriptive, diagnostic, inferential, and predictive statistics by learning about data, measurement, probability, distribution, estimation, confidence Interval, test of hypothesis, predictive models (regression, multiple regression), ANOVA, logistic regression, and statistical applications and assessment.

### About Vila Health

In this program, a simulation of a fictitious hospital organization called Vila Health allows you to solve real problems with real analytical solutions. The media interactions can help you understand a business problem and steps you may take to solve it. It also lets you practice the role of analyst and suggests how you can articulate a solution in a manner that others can understand.

### Technology Resources

If you require the use of assistive technology or alternative communication methods to participate in course activities, please contact [DisabilityServices@Capella.edu](mailto:DisabilityServices@Capella.edu) to request accommodations.

#### References

- Albright, S. C., & Winston, W. L. (2013). *Business analytics: Data analysis and decision making*. Stamford, CT: Cengage Learning.
- Lock, R. H., Lock, P. F., Lock, M. K., Lock, E. F., & Lock, D. F. (2013). *Statistics: Unlocking the power of data*. Hoboken, NJ: Wiley.
- Mirkin, B. (2011). *Core concepts in data analysis: Summarization, correlation, and visualization*. London, England: Springer.

### Course Competencies

(Read Only)

To successfully complete this course, you will be expected to:

- 1 Explain the data requirements for statistical methods.
- 2 Apply descriptive statistical analysis in tabular and graphical methods.
- 3 Apply statistical methods to interpret data points and present estimation and confidence levels.
- 4 Use appropriate tools to obtain a result using statistical methods.

5 Present results to stakeholders in a succinct and relevant manner.

**Course Prerequisites**

***Prerequisite(s): Completion of or concurrent registration in ANLT5020.***

## Syllabus >> Course Materials

### Required

The materials listed below are required to complete the learning activities in this course.

### Integrated Materials

Many of your required books are available via the VitalSource Bookshelf link in the courseroom, located in your Course Tools. Registered learners in a Resource Kit program can access these materials using the courseroom link on the Friday before the course start date. Some materials are available only in hard-copy format or by using an access code. For these materials, you will receive an email with further instructions for access. Visit the [Course Materials](#) page on Campus for more information.

#### Book

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2018). *Modern business statistics with Microsoft Excel* (6th ed.). Boston, MA: Cengage Learning. ISBN: 9781337115186

### Library

The following required readings are provided in the Capella University Library or linked directly in this course. To find specific readings by journal or book title, use [Journal and Book Locator](#). Refer to the [Journal and Book Locator library guide](#) to learn how to use this tool.

- Delaney, L. (2010). [Descriptive statistics: Simply telling a story](#). *African Journal of Midwifery & Women's Health*, 4(1), 43–48.
- Fullerton, A. S. (2009). [A conceptual framework for ordered logistic regression models](#). *Sociological Methods & Research*, 38(2), 306–342.
- Kim, H.-Y. (2014). [Analysis of variance \(ANOVA\) comparing means of more than two groups](#). *Restorative Dentistry & Endodontics*, 39(1), 74–77.
- Panik, M. J. (2012). [Statistical inference: A short course](#). Hoboken, NJ: Wiley.
- Ranganathan, P., Pramesh, C. S., & Buyse, M. (2015). [Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals](#). *Perspectives in Clinical Research*, 6(2), 116–117.
- Sarmukaddam, S. B. (2012). [Interpreting "statistical hypothesis testing" results in clinical research](#). *Journal of Ayurveda and Integrative Medicine*, 3(2), 65–69.
- Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). [Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XL Miner \(2nd ed.\)](#). Hoboken, NJ: Wiley.
- Van De Steeg, L., Langelaan, M., & Wagner, C. (2014). [Can preventable adverse events be predicted among hospitalized older patients? The development and validation of a predictive model](#). *International Journal for Quality in Health Care*, 26(5), 547–552.
- Wills, M. J. (2014). [Decisions through data: Analytics in healthcare](#). *Journal of Healthcare Management*, 59(4), 254–262.
- Zhou, S., & Zhu, N. (2013). [Multiple regression models for energy consumption of office buildings in different climates in China](#). *Frontiers in Energy*, 7(1), 103–110.

### External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- [Cengage Brain](#). (n.d.). Retrieved from <http://www.cengagebrain.com>
- [SAS Proceedings](#). (n.d.). Retrieved from <http://www.lexjansen.com>
- [SAS](#). (n.d.). Retrieved from [http://www.sas.com/en\\_us/home.html](http://www.sas.com/en_us/home.html)

### Suggested

The following materials are recommended to provide you with a better understanding of the topics in this course. These materials are not required to complete the course, but they are aligned to course activities and assessments and are highly recommended for your use.

### Optional

The following optional materials are offered to provide you with a better understanding of the topics in this course. These materials are not required to complete the course.

#### Library

The following optional readings may be available in the Capella University Library. To find specific readings by journal or book title, use [Journal and Book Locator](#). Refer to the [Journal and Book Locator library guide](#) to learn how to use this tool. If the full text is not available, you may be able to request a copy through the [Interlibrary Loan](#) service.

- Ofner, M. H., Otto, B., & Österle, H. (2012). [Integrating a data quality perspective into business process management](#). *Business Process Management Journal*, 18(6), 1036–1067.

#### External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Kaggle. (n.d.). Retrieved from <https://www.kaggle.com/>
- Python. (n.d.). Retrieved from <http://www.python.org>
- Python.org. (n.d.). [Beginner's Guide to Python](#). Retrieved from <https://wiki.python.org/moin/BeginnersGuide>
- Python.org. (n.d.). [Downloading Python](#). Retrieved from <https://wiki.python.org/moin/BeginnersGuide/Download>
- R. (n.d.). Retrieved from <https://www.r-project.org/>
- SAS. (2014). [SAS product documentation](#). Retrieved from <http://support.sas.com/documentation/>
- [World Bank Open Data](#). (n.d.). Retrieved from <https://data.worldbank.org/>

## Unit 1 >> Data and Data Sources

### Introduction

Accurate data analysis depends on the quality of the data. To understand data quality, you must first understand the kinds of data that you may encounter in the workplace and the kinds of statistical procedures appropriate for each data type. For example, an analyst may code a state numerically: 1=Alabama, 2=Alaska, and so on. In this situation, calculating an average of this number does not make sense; what does average state mean? The numeric code of this value is considered nominal data, where a number represents a name and nothing else. A code of 1 is not better or worse than 2; 2 is not better or worse than 3. It is critical for an analyst to understand what kinds of statistical functions are available, given the data available.

Data, when properly processed, means information. It may produce false information, too. In today's business environments, we are encountering data-driven activities. Data characteristics may be described by the "6Vs": volume, velocity, variety, value, variability, and veracity (Normandeau, 2013).

From production points of view, data may be categorized as follows:

- Structured data: transactional and time phased.
- Unstructured data: social, channel, customer service, and warranty.
- Sensor data: temperature, RFID, QR codes, and GPS.
- New data types: mapping and GPS, video, voice, and digital images (Cecere, 2012, p. 281).

Data generated in the above categories with the types listed, if stored in disparate systems or repositories, should be integrated with a unified format to be used in data mining and analytics. Fortunately, the majority of data used in analytics applications are stored in data warehouses and data marts.

#### References

Cecere, L. M., & Chase, C. W. (2013). *Bricks matter: The role of supply chains in building market-driven differentiation*. Somerset, NJ: Wiley.

Normandeau, K. (2013). Beyond volume, variety, and velocity is the issue of big data veracity. Retrieved from <http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>

## Learning Activities

### u01s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 1, "Data and Statistics."
  - This chapter covers data, data sources, descriptive and inferential statistics, data mining, and ethical guidelines for statistical practice.

Use the Capella library to complete the following readings:

- Delaney, L. (2010). Descriptive statistics: Simply telling a story. *African Journal of Midwifery & Women's Health*, 4(1), 43–48.
  - This article describes the application of descriptive statistics.

## Optional Reading

- Ofner, M. H., Otto, B., & Österle, H. (2012). Integrating a data quality perspective into business process management. *Business Process Management Journal*, 18(6), 1036–1067.
  - This article examines the impact of data quality in decision making.

## Assignment Preparation Resources

### Multimedia

The following optional walkthroughs may help you understand concepts addressed in your assignments:

- [SAS OnDemand: Import Data Files Walkthrough](#).
- [SAS OnDemand: Task and Utilities Walkthrough](#).
- [SAS OnDemand: Variables Walkthrough](#).
- [SAS OnDemand: Characterize Data Walkthrough](#).
- [SAS OnDemand: One-Way Frequency Table Walkthrough](#).
- [SAS OnDemand: Scatterplot Walkthrough](#).
- [SAS OnDemand: Cross-Tabulation Table Walkthrough](#).
- [SAS OnDemand: Distribution Analysis Walkthrough](#).
- [SAS OnDemand: Table Attributes Walkthrough](#).
- [SAS OnDemand: Cross-tabulation Tables Fuel Data Walkthrough](#).
- [SAS Enterprise Guide: Vila Health – One-Way Frequencies Walkthrough](#).
- [SAS Enterprise Guide: Vila Health – Graph Highlights Walkthrough](#).
- [SAS Enterprise Guide: Scatterplot Walkthrough](#).
- [SAS Enterprise Guide: Excel File – One-Way Frequencies Walkthrough](#).
- [SAS Enterprise Guide: Excel File – Cross Tabulation Tables Walkthrough](#).

### u01s2 - CengageBrain Data Sets

To practice statistics in SAS and complete some of the assignments in this course, you will need to access data sets provided for free with your textbook. To access these files:

1. Visit [Cengage Brain](#).
2. Enter **ISBN: 9781337115186** in the Enter Your Course or Register a Product search bar.
  - You will need to create an account if you do not have one.
  - Complete your profile to create a new account or login if you have previously registered on this site. For new users, you do not need to use a Capella email. To Find Your Institution, select U.S., 2 Year College, within 25 miles, zip code 55402, and select Capella University.
3. Click **Search**. The textbook will be listed under My Products in My Dashboard.
4. Click on the textbook from My Dashboard to access the supplemental files.
5. Select the book chapter you are working in for the exercise or assignment.
6. Click the DATAfiles link to download the zip files for that chapter.

The Excel files provided allow you to work with real data sets that accompany the exercises in the text.

### u01s3 - Software Preparation and Technology Access

In this course, you will be using software and technology that is needed to complete designated activities and assignments. There is no additional cost for this software and technology. Some software packages will be made available to you at no additional cost through Capella's subscription with Microsoft, while other software packages are available for free download through open-source licensing.

Capella University requires learners to meet certain minimum [computer requirements](#). Please note that some software required for a course may exceed these minimum requirements. Check the requirements for the software you may need to download and install to make sure it will work on your device. Most software will require a Windows PC. If you use a Mac, refer to [Installing a Virtual Environment and Windows on a Mac](#).

The software and technologies below are strongly recommended to support you in completing the course objectives. If you have access to other tools that you believe may still meet the requirements of this course, please discuss your selected alternatives with your instructor.

If you use assistive technology or any alternative communication methods to access course content, please contact [DisabilityServices@Capella.edu](mailto:DisabilityServices@Capella.edu) with any access-related questions or to request accommodations.

For this course, follow the instructions provided through the links below to download and install software or register for an account, as required.

### SAS Statistical Software

1. Visit [SAS OnDemand for Academics \(SODA\)](#).
2. Download the SAS data files for use in the assignments.
3. Download the SAS data files for the assignments:
  - [ANLT5030 Data Files \[ZIP\]](#).
  - [Strike Reports 2010-Current \[XLSX\]](#).
4. Open the files and take some time to explore the data sets to see how they were constructed.

### Open Source Statistical Software

R and Python are two open source software that can be applied as an alternative to SAS to complete the assignments in this course.

- [Python](#): Go to the [Download](#) section of the [Python Beginners Guide](#) to download the latest version of Python.
- [R](#): Follow the instructions on the page to download the latest version of R.

#### Selecting a Statistical Software

On the one hand, R and Python are open source software that are free and commonly used. On the other hand, there are several reasons why an analyst would choose SAS over R or Python. SAS has better reference documentation. R and Python are getting better, but official documentation is not well-written and somewhat opaque. SAS is supported by a massive infrastructure of books. SAS has single-source support. If it is broken, you know who to call. If R or Python are broken, there is nobody to call. R and Python packages vary in quality, some are written by experts, and some are not. Whereas, SAS is only written by experts. SAS is very much the standard for many industries, especially the pharmaceutical industry, where SAS is required by the FDA. The name SAS carries a lot of inertia and is used predominantly in institutions and academia.

### u01s4 - Programming Instructions

#### Using R

Access the [Using R](#) page on Campus for resources on:

- How to get started with R.
- How to read a file in R.
- How to get started with Descriptive Statistics in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on:

- How to get started with Python.
- How to read a file in Python.
- How to get started with Descriptive Statistics in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on:

- How to get started with SAS.
- How to read a file in SAS.
- How to get started with Descriptive Statistics in SAS.

### u01s5 - Analytics Internship: Assessing Data

An integral part of Capella University's data analytics program is your virtual internship with Vila Health, a fictional health care system that operates hospitals and other health care facilities throughout the upper midwestern United States. This internship is designed to allow you to apply the skills and knowledge you are acquiring in a realistic setting. In addition to the technical requirements of the assignment, the internship also provides a context for you to develop the collaborative, interpersonal skills that employers are looking for in new data analysts.

In this first activity, you will continue your Vila Health internship experience by checking in with your mentor, who has a new case study for you and your study group to focus on throughout this course. You will use the information about the case study as the context for your group discussions (Study Group Tasks) and the related assignments in the coming units. Click **Analytics Internship: Assessing Data** to view the Vila Health scenario.

Course Resources

---

[Analytics Internship: Assessing Data](#)

### u01s6 - Group Work and Personal Effectiveness

## Study Group Tasks

For each unit in the course, you will find suggested study group tasks describing topics to discuss with your group within a social media setting. These discussions are opportunities for you to collaborate with your group members to help you solve the Vila Health related assignments. It is important that you attend these study group sessions to help you develop your professional skills in areas like collaboration, decision making, oral and written communication, prioritizing, working with people from diverse backgrounds, et cetera.

Review the most recent Vila Health challenge and any of the materials you created or acquired during project-related or challenge activities prior to your study group meetings. Downloaded notes from the notebook can provide useful reminders and questions for your group discussion.

You will encounter study group tasks throughout the course, and in every course that you encounter as you journey through your program.

The questions provided are meant to provide you and your group members with some guidance for your assignment deliverables. These questions should not be overlooked, as they are critical to you completing the work in the course.

## Study Group Meetings

Capella recommends that you use a social media channel to collaborate and connect with your group members. Work with your teammates to identify a platform you are all comfortable using (such as Basecamp, Yammer, Google Hangouts, or Facebook) and check to make sure that this platform is also reasonably private and secure.

For example, if you decide to use Facebook, create a secret group and work within it. A secret group is invisible to the outside world, and only members can find the group and see posts. This social media group will allow you to stay connected to your team and allow you to discuss the study group topics assigned.

## LinkedIn

For networking with past, current, and future learners in this program, we recommend that you use LinkedIn, which is a great way to stay connected as you move through the program and beyond.

Course Resources

---

[Analytics Internship: Assessing Data](#) | Transcript

---

Group Collaboration Skills

### u01d1 - Types of Variables in a Data Set

As a data analyst, it is important that you understand the variables within your data set before you run any analysis. This includes understanding the meaning of each variable and type of each variable, whether it is a nominal, interval, ratio, or ordinal variable.

For this discussion, browse the Vila Health data provided by your instructor, or in SAS OnDemand. Create a list or chart that specifies the nominal, interval, ratio and ordinal variables in the Vila Health data set. Also identify any variables that are considered flag variables, which are also called binary or boolean variables, and only take on the values of 0 and 1.

This discussion will help you prepare for the Unit 1 assignment. In the assignment, you will be asked to create a frequency distribution or a statistical summary table, depending on the type of variable. Knowing the type of each variable beforehand will help you choose the correct procedure in SAS to create either a frequency distribution or a statistical summary table.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

### u01a1 - Analytics Internship: Assessing Data

Descriptive statistics provide the foundation for a data analyst's work. They provide the basic formulas that data analysts use to make meaning out of raw data as well as the prerequisite understanding for more complex statistical computations. Before completing descriptive statistics on any data set, you must first understand the types of data you have to work with and the conditions of your data set. For this assignment, examine the data set presented in the Vila Health media piece to orient yourself to the scenario, tasks, and data presented.

In your assignment, address the following elements:

- Identify the number of observations in the data set.
- Identify any missing values in the provided data set and any variables missing a lot of values.
- Identify the variables in the data set using flag logic. For each one, describe the logic used to populate it.
- Create frequency tables for the nominal variables in the data set.

- Create a statistical summary table of the interval and ratio data in the data set (include average, median, standard deviation, and minimum/maximum). **Hint:** Try using the UNIVARIATE or the MEANS procedures in SAS to generate your measures.
- Create a frequency table of the ordinal data in the data set. Include the percent frequency and the cumulative percent frequency.
- Identify two variables that could be used, with statistics, to track and address the business problem in the Vila Health scenario.
- Summarize your output briefly, using headings to organize your work.
- Include your SAS output in your submission.

**Note:** Using SAS or an alternative statistical analysis software is required to complete your work for this assignment.

Review the Analytics Internship: Assessing Data Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[SAS Product Documentation](#)

## Unit 2 >> Descriptive Statistics

### Introduction

This unit covers descriptive statistics, which includes mean, median, standard deviation, variance, and mode. Descriptive statistics can be used to give an overall picture of both quantitative and qualitative data, which can be presented in a table of numbers containing these statistics or graphically through histograms, boxplots, bar charts or line charts. It includes numerical values representing the measures of central tendency and variability. In tabular format, descriptive statistics are arranged in columns and rows, and contain the descriptive statistic value, including averages, standard deviations, medians, minimums, maximums and frequencies. Histograms, pie charts, bar charts, polygons, box plots, and stem and leaf diagrams are graphical means of displaying descriptive statistics.

### Learning Activities

#### u02s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 2, "Descriptive Statistics: Tabular and Graphical Displays," sections 2.1, 2.2, 2.3, 2.4, and Figure 2.29.
  - Skim through the Excel demonstrations in each section and skip the exercises. This chapter provides an introduction to visual presentation of data.
- Read Chapter 3, "Descriptive Statistics: Numerical Measures," sections 3.1–3.5.
  - Read all sections and skim through the Excel demonstrations; skip the exercises. This chapter introduces numerical measures of data.

Use the Capella library to complete the following reading:

- Wills, M. J. (2014). [Decisions through data: Analytics in healthcare](#). *Journal of Healthcare Management*, 59(4), 254–262. This article describes an application of descriptive statistics in the health care sector.

## Assignment Preparation Resources

### Multimedia

The following optional walkthroughs may help you understand concepts addressed in your assignments:

- [SAS OnDemand: Distribution Analysis Walkthrough](#).
- [SAS OnDemand: Scatterplot Walkthrough](#).
- [SAS Enterprise Guide: Excel File – Cross Tabulation Tables Walkthrough](#).
- [SAS Enterprise Guide: Excel File – One-Way Frequencies Walkthrough](#).
- [SAS Enterprise Guide: Vila Health – Graph Highlights Walkthrough](#).

- [SAS Studio: Import Excel File – Scatterplots Walkthrough](#).

## u02s2 - Programming Instructions

### Using R

Access the [Using R](#) page on Campus for resources on Graphical Methods in R.

### Using Python

Access the [Using Python](#) page on Campus for resources on Graphical Methods in Python.

### Using SAS

Access the [Using SAS](#) page on Campus for resources on Graphical Methods in SAS.

## u02d1 - Planning Your Group

Throughout the course, you will be required to work collaboratively outside of the courseroom. By now, your instructor should have assigned you into groups of 3 to 5 members. With your group, be sure to:

- Introduce yourself to your group members. Share information about your background and experience and include your experience using and working with analytics software.
- Decide on the social media channels and collaborative tools that your group plans to use throughout this course.
- Meet with your group at your agreed-upon social media location and discuss logistics related to when and how you will meet.
- Consider how your group will work together to collaborate and support each other throughout the course.

Complete your group set up and role selection by Wednesday of this week.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[Analytics Internship: Assessing Data](#) | [Transcript](#)

---

[Group Collaboration Skills](#)

## u02d2 - Comparing Data Presentation Methods

The way in which data is presented can have a significant impact on how users and stakeholders interpret the data. In order to make decisions about how to present data, you must first understand the basic presentation methods that are available and the kinds of meaning that can be conveyed in each.

For this discussion, provide a use case and example of tabular data, bar chart, pie chart, histogram, and box plot relevant to the use case. Also provide business use examples of a cross-tabulation table and a scatter diagram.

You can use an existing report or create your own charts using a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

Discuss when you would use each type of presentation method and the pros and cons of each method.

In your responses, comment and provide support on your classmates' presentation methods in this discussion.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

---

[Kaggle](#)

---

[World Bank Open Data](#)

---

### u02a1 - Descriptive Statistics for Customer Profile

For this assignment, you will practice using descriptive statistics on a case in your textbook. Read Case Problem 1, Pelican Stores, on pages 102–103 of your text, and download the accompanying data set from CengageBrain, to create the following in your report:

1. Create a percent frequency distribution for the key variables in the data and depict in tabular form.
2. Create a bar chart or pie chart showing the number of customer purchases attributable to the method of payment. **Hint:** Use the GCHART procedure in SAS to create these charts.
3. Create a cross tabulation of type of customer (regular or promotional) versus net sales. Comment on any similarities or differences present. **Hint:** Use the FREQ procedure in SAS to create cross tabulations.
4. Create a scatter diagram to explore the relationship between net sales and customer age. **Hint:** Use the SGPLOT in SAS to create a scatter diagram.

Remember, to access the PelicanStores data set Web file from the Chapter 2 content, follow the directions in Unit 1 to access the free content at CengageBrain, if you have not yet signed up on the site.

**Note:** The data sets included with the course text are in Excel file format. You can use the Excel tools presented in the book to complete the assignment, but you must import the data sets into SAS and complete the exercises there. The hints in the questions above give you a place to start in researching the procedures used in SAS to get to the desired result. To import the data in SAS go to File> Import Data. Navigate to where you downloaded, extracted, and saved the files from the ANLT5030 Data Files (ZIP) package. Be sure to check the box titled, "rename columns to comply with SAS naming conventions," which will make programming much easier.

Review the Descriptive Statistics for Customer Profile Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

---

[SAS Product Documentation](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 3 >> Probability, Distribution, Estimation, and Confidence Interval

### Introduction

Life is probabilistic. Probability has been applied in areas as diverse as medicine, gambling, weather forecasting, finance, and the law. We have learned to cope with uncertainty about the weather, our food supply, and other aspects of our environment, and have tried to reduce uncertainty and its impact on many aspects of our lives.

The probability distribution of a random variable is a set of probabilities associated with each of the possible values the random variable may take. The classic example involves rolling a fair die. With each roll of the die, the only possible value the die can display is a number between 1 and 6. The probability distribution is therefore probability of rolling a 1, a 2, a 3, a 4, a 5, or a 6.

By definition, statistical inference is a procedure that produces conclusions about a group or a population based upon statistical analysis of a set of data that we have. It is generally not possible to analyze all members of a population, so we analyze a subset of the population, called a sample. Based upon the result of the analysis observed in the sample, such as the average value of a variable, we can use probability distributions to come up with a range of values that we can expect the analysis outcome to fall within if we were able to analyze the whole population. The range of values that we get depends upon how sure the researcher wants to be that their range encompasses the true value of the population. This is known as the confidence interval and the levels most commonly used include the 90% level, in other words, we are 90% sure that an average lies between  $x$  and  $y$ , and the 95% interval, meaning that we are 95% sure that the true population value lies between  $x$  and  $y$ . Levels of confidence exist, at its most confident, at the 99% level.

### Learning Activities

#### u03s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 8, "Interval Estimation."
  - This chapter covers methods of interval estimates with cases of known and unknown standard deviation.
- Skim Chapters 4, 5, and 6.
  - Skim through the Excel demonstrations in each section and skip the exercises. These chapters cover probability concepts as well as discrete and continuous probability distributions.

Use the Capella library to complete the following readings:

- Ranganathan, P., Pramesh, C. S., & Buyse, M. (2015). [Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals](#). *Perspectives in Clinical Research*, 6(2), 116–117. This article reviews the "p" value in statistical analysis.

#### u03s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on using Confidence Intervals in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on using Confidence Intervals in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on using Confidence Intervals in SAS.

### u03d1 - Confidence Interval Estimates

For this discussion, provide two real-world examples of confidence interval: one example for a population mean and another example for a population proportion.

Use SAS to run the confidence interval on your data set. Interpret the meaning of your intervals in the context of the scenario you presented.

You can use a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

Discuss how you could either make the interval narrower or wider in terms of the confidence level, sample size, and margin of error.

In your responses, comment and provide support on your classmates' methods in this discussion.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

---

[Kaggle](#)

---

[World Bank Open Data](#)

---

### u03a1 - Descriptive Statistics and Confidence Intervals

For this assignment, you will continue your practice of descriptive statistics and add confidence intervals to your skill set. For this assignment, read Case Problem 2, Gulf Real Estate Properties, on page 402–403 of your text, and download the accompanying GulfProp data set from CengageBrain. Use SAS to create the following elements in your report:

- Use appropriate descriptive statistics to summarize each of the three variables for the 40 Gulf View condominiums.
- Use appropriate descriptive statistics to summarize each of the three variables for the 18 No Gulf View condominiums.
- Compare your summary results. Discuss any specific statistical results that would help a real estate agent understand the condominium market.
- Develop a 95% confidence interval estimate of the population mean sales price and population mean number of days to sell for Gulf View condominiums and interpret your results.
- Develop a 95% confidence interval estimate of the population mean sales price and population mean number of days to sell for No Gulf View condominiums and interpret your results.
- Assume the branch manager requested estimates of the mean selling price of Gulf View condominiums with a margin of error of \$40,000 and the mean selling price of No Gulf View condominiums with a margin of error of \$15,000. Using 95% confidence, how large should the sample sizes be?
- Gulf Real Estate Properties just signed contracts for two new listings: a Gulf View condominium with a list price of \$589,000 and a No Gulf View condominium with a list price of \$285,000. What is your estimate of the final selling price and number of days required to sell each of these units?

Remember, to download the GulfProp data set web file from the Chapter 8 content, access the free content at CengageBrain.

Review the Descriptive Statistics and Confidence Intervals Scoring Guide for detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

## Unit 4 >> Test of Hypothesis

### Introduction

When we are interested in whether or not a parameter is greater, less than, or equal to some other value, we conduct a hypothesis test. Hypotheses are generally tested in pairs, but there can be more than two items. The first hypothesis is called the null hypothesis, which states that the distribution parameter is equal to the other value. The second hypothesis is called the alternate hypothesis, which states that the parameter is either greater than or less than some other value. There can be as many alternate hypotheses as needed to adequately address the overall question. This framework is generally used to compare groups.

It is important to remember that hypothesis tests are not without their limitations. Sometimes, in situations where a lot of data is being analyzed, the amount of difference needed to produce a significant result is very small. From a business perspective, such a small change produces a significant difference but not a functional difference. That is, the difference is so small that business leaders do not find it actionable or worthy of action. Data analysts must be mindful of this tendency of hypothesis tests and adjust their interpretation accordingly.

### Learning Activities

#### u04s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 7, "Sampling and Sampling Distributions."
  - This chapter introduces the concept and process of random sampling and its distribution.
- Review Chapter 8, "Interval Estimation."

Use the Capella library to complete the following readings:

- Sarmukaddam, S. B. (2012). [Interpreting "statistical hypothesis testing" results in clinical research](#). *Journal of Ayurveda and Integrative Medicine*, 3(2), 65–69. This article examines the importance and application of hypothesis testing in a medical setting. Keep in mind that a medical setting is only one example of the use of hypothesis testing. Consider additional applications and contexts as you read.

#### u04s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on Hypothesis Testing in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on Hypothesis Testing in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on Hypothesis Testing in SAS.

### u04s3 - Analytics Internship: Using Data to Drive Performance

In this Vila Health activity, you will have the opportunity to speak with key stakeholders throughout the Vila Health organization. These interviews will help you determine how to apply the statistical analyses you are conducting on the Vila Health data set to the performance issues identified by Vila Health leaders in the initial business problem. You will be able to learn more about the fields and variables and the data set in general by interviewing various end users and stakeholders. Click **Analytics Internship: Using Data to Drive Performance** to view the Vila Health scenario.

Course Resources

---

[Analytics Internship: Using Data to Drive Performance](#)

### u04s4 - Study Group Task

By Unit 2 your instructor should have assigned you to a group and your group should have chosen an area to meet outside of the courseroom to discuss the course topics. See the study in Unit 1, Group Work and Personal Effectiveness, for more information on roles and social media.

Your study group tasks in this course are intended to provide you an organized structure through which you can learn from your peers, get help on your assignment questions, and develop your professional development skills. Do not skip any of these discussions as it will negatively impact your grade.

This week's topic for group discussion:

- Discuss the Vila Health media piece from this unit and the assignment you need to complete in Unit 5. Discuss any questions you have with your group.

Keep in mind the role you are playing within your group. You may find that some information is more useful to your role than others.

Course Resources

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[Analytics Internship: Using Data to Drive Performance](#) | Transcript

---

Group Collaboration Skills

### u04d1 - Hypothesis Testing

For this discussion, give two real-world examples of hypothesis testing for two situations: one example for testing a population mean and another example for testing a population proportion.

You can use a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

In your initial post, include answers to these questions:

- What is your hypothesis?
- What are you testing within the population?
- What SAS procedures would you run for the hypothesis test?

In your responses, comment and provide support on your classmates' methods in this discussion.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

---

[Kaggle](#)

---

[World Bank Open Data](#)

---

### u04a1 - Applying Hypothesis Tests

For this assignment, you will continue practicing with confidence intervals in a more ambiguous context. For this assignment, read Case Problem 2, Ethical Behavior of Business Students at Bayview University, on pages 452–454 of your text, and download the accompanying Bayview data set from CengageBrain. Use SAS to create the following elements in your report:

- Use descriptive statistics to summarize the data and comment on your findings.
- Develop 95% confidence intervals for the proportion of all students, the proportion of male students, and the proportion of female students who were involved in some type of cheating.
- Conduct a hypothesis test to determine if the proportion of business students at Bayview University who were involved in some type of cheating is less than that of business students at other institutions, as reported by the *Chronicle of Higher Education*.
- Conduct a hypothesis test to determine if the proportion of business students at Bayview University who were involved in some type of cheating is less than that of nonbusiness students at other institutions, as reported by the *Chronicle of Higher Education*.
- Describe the advice you would give to the dean based on your analysis of the data related to cheating at Bayview University.

Remember, to download the Bayview data set Web file from the Chapter 9 content, access the free content at CengageBrain.

Review the Applying Hypothesis Tests Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

---

[SAS Product Documentation](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 5 >> Inference About Means and Variances

### Introduction

There are a number of statistical procedures that perform hypothesis testing, and it is important to distinguish between all of them and how they relate to business data. In this unit, you examine the construction of interval estimates and make hypotheses about means and variance of two populations. In previous units, you practiced developing interval estimates and conducting tests of hypothesis for situations involving only one population. In this unit, you learn about statistical inference by showing how interval estimates and tests of hypothesis are designed for business cases concerning two populations when the difference between the population's means and standard deviation matters. For instance, differentiating between the number of

defective items produced by two suppliers is an example of inference about means. Analysis and inference about variances is very important in the decision-making process. For example, for the concept of weight uniformity, we will examine the variance of two filling machines in order to make sure that their variances are within acceptable levels.

## Learning Activities

### u05s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 10, "Inference About Means and Proportions With Two Populations," pages 442–492.
- Read Chapter 11, "Inferences About Population Variances," pages 504–524.

Use the Capella library to complete the following readings:

- Panik, M. J. (2012). *Statistical Inference: A Short Course*. Hoboken, NJ: Wiley.
  - Read Chapter 8, "Confidence Interval Estimation of  $\mu$ ," pages 152–169.

### u05s2 - Study Group Task

Share the work you have completed so far for this unit's assignment with your group. Ideally, you should have most of it completed by the time you meet with your group so you can share your report and get ideas for ways to improve your work before submitting it for a grade in this unit.

Course Resources

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[Analytics Internship: Using Data to Drive Performance](#) | Transcript

### u05d1 - Hypothesis Testing for Two Populations

For this discussion, find an example that uses confidence intervals and hypothesis testing for the difference of two population means.

Use the readmission data for Vila Health. Look at two populations within that data set.

In your initial post:

- Explain how you could apply two-sample hypothesis testing for the Vila Health data.
- Run a SAS example and share with the class.

In your responses, comment and provide support to your classmates.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

Graduate Discussion Participation Scoring Guide

### u05a1 - Analytics Internship: Using Data to Drive Performance

This is the second assignment related to Vila Health. For the first assignment, you were asked to identify basic information and statistics about the data set provided for Vila Health. For this assignment, you will run various descriptive statistics on the Vila Health data set to determine significant variables that may then support Key Performance Indicators (KPIs) to address readmission rates and length of stay. The hospital leadership is concerned about readmission rates and length of stay. Your job, as the intern, is to discover what the data is telling you and present it to leadership.

- Provide summary descriptive statistics by DRG for length of stay and readmittance.
- Determine DRGs across hospitals with significant length of stay and readmission rates.
- Provide a narrative based on your calculations and other data in the data set related to the DRGs you found interesting.
- Present two cross-tabulation tables, pairing a dimension of your choice with the frequency of readmission within 30 days. You may need to experiment with various dimensions to find ones with significance.
- Provide any additional findings you discovered in the data set. Include visuals to support your findings.
- Create a report for business leaders that shares the data and insights from your results in an informative, ethical, and easy to consume format. Include your KPI recommendations for both readmission rates and length of stay. You should include a minimum of 2 visuals in your report.
- Include an appendix with samples of the programs that you used in SAS and screenshots of the SAS output used to generate your outcomes.

Remember to review the Vila Health media pieces from this unit and Unit 1, provided in the Resources. **Note:** You should use R, and/or SAS to complete this assignment. Review the Analytics Internship: Using Data to Drive Performance Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[SAS Product Documentation](#)

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[Analytics Internship: Using Data to Drive Performance](#) | Transcript

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 6 >> ANOVA

### Introduction

In the previous unit, you studied how to conduct inferences about means and variances of two populations. In this unit, you learn how to compare means of three populations or more. As an example, assume that you want to compare the performance of three high schools. For that purpose, you may use the SAT scores of the seniors at those schools. The same principles may be applied in studying the income of households in given census districts. This analysis method, ANOVA (analysis of variances), is used for conducting such an analysis.

ANOVA can be very useful when we want to compare averages over three or more groups; however, there are more strict data requirements with ANOVA and it does not necessarily tell us, depending how the test is structured, if there are differences between specific groups. For example, if I was comparing successful course completion ratios across standard courses, first courses and capstones, I would know from an ANOVA that there may be differences in the ratios across the three groups, but I would not know if there were differences between standard courses and first courses, or standard courses and capstones, or first courses and capstones. I would have to run additional analysis for that and be careful of compounding my selected error rate, which is a common issue with ANOVA.

Course Resources

---

Kim, H.-Y. (2014). [Analysis of variance \(ANOVA\) comparing means of more than two groups](#). *Restorative Dentistry & Endodontics*, 39(1), 74–77.

## Learning Activities

### u06s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 13, "Experimental Design and Analysis of Variance."
  - This chapter examines how the analysis of variance may be used for testing the equality of many population means.

Use the Capella library to complete the following reading:

- Kim, H.-Y. (2014). Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative Dentistry & Endodontics*, 39(1), 74–77.

### u06s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on using Analysis of Variance in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on using Analysis of Variance in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on using Analysis of Variance in SAS.

### u06d1 - Collaboration Reflection

The collaboration in this course has probably looked slightly different from your experience in previous courses. You have been asked to use your group as more of a resource and to serve as a resource to them as well, rather than completing assignments collaboratively.

For this discussion:

- Review the report you provided in the Unit 5 assignment and assess its limitations for business decisions. Discuss other types of analysis that can be done to help business leaders determine the actions they can take now to influence future results.
- Reflect on your group collaboration experience working on the Vila Health problem and completing the Unit 5 assignment.
- Describe and self-assess your contributions to the group study tasks so far, in terms of both content and your role in the group.
  - What challenges do you feel you are facing as you work together?
  - How did your ideas differ from others' approach to the assignment, if at all?
- Talk about what you learned from others in your group and how you plan to incorporate what you learned.

## Response Guidelines

Respond to at least two of your peers, providing your insight from the group collaboration experience in this course so far.

### u06d2 - Using ANOVA With a Randomized Block Design Experiment

For this discussion, find a real-world ANOVA example. Explain what is being tested in your example and how you would go about conducting that test.

You can use an existing report or create your own charts using a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

Based on your example, run the ANOVA using a five-step approach:

1. Set up hypotheses and choose a level of significance.
2. Select the appropriate test statistic.
3. Set up the decision rule.
4. Compute the test statistic.
5. Present your conclusion.

In your initial post, include answers to these questions:

- What is your null hypothesis?
- What is the purpose of ANOVA?
- Why would you run an ANOVA on data?
- What does ANOVA tell you about the data?

In your responses, comment and provide support on your classmates' methods in this discussion.

### Response Guidelines

Respond to at least two of your peers to discuss differences that you may have found in your responses or questions your peers may have on how to complete the necessary calculations.

Course Resources

---

Graduate Discussion Participation Scoring Guide

---

[Cengage Brain](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

---

[Kaggle](#)

---

[World Bank Open Data](#)

### u06a1 - Analysis of Variance

For this assignment, you will be assessed on your use and interpretation of ANOVA through a case in your textbook. Read Case Problem 1, Wentworth Medical Center, on page 616 of your text, and download the accompanying data sets from CengageBrain to create the following in your report:

- Use descriptive statistics to summarize the data from the two studies.
- Provide your preliminary observations about the depression scores for both studies.
- Using analysis of variance (ANOVA) on both data sets, state the hypothesis being tested in each case.

- Provide your conclusions based on your ANOVA calculations.
- Use inferences about individual treatment means where appropriate and state your conclusions.

Review the Analysis of Variance Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 7 >> Linear Regression

### Introduction

Predictive models are data driven analytics methods that look for patterns within data and then identify the probability that a selected outcome, or *target*, will occur. Predictive models draw from several related disciplines including statistics, pattern recognition, machine learning, artificial intelligence, and data mining. One of the more common tools for predictive modeling is the regression, which you will examine in this unit. In regression modeling for predictive purposes, we subset our data randomly into a training set and a testing set. The training set is used to build the regression equation by selecting variables to be included in the model and assigning them a weight based upon the data. Then, the equation is tested for accuracy on the testing data. The way in which the predictive model generates output, and its key characteristics, are determined by the regression equation.

### Learning Activities

#### u07s1 - Studies

## Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 14, "Simple Linear Regression."
  - In regression, we try to predict dependent variables as a function of other variables.

#### u07s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on using Simple Linear Regression in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on using Simple Linear Regression in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on using Simple Linear Regression in SAS.

### u07d1 - Assumptions in a Regression Model

In order to present a valid regression model to management, an analyst needs to check a number of important assumptions for the regression to be useful. Make a list of each of those assumptions and explain how you would check each assumption using graphical techniques.

Provide a business use example of simple linear regression. You can use a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

In your initial post:

- Discuss how you could check those assumptions in SAS. What procedures are useful for this?
- What SAS procedure options can you add to test the assumptions?
- List the assumptions for the regression model and how to check each one using SAS.

In your response, comment on your classmates' methods in this discussion.

### Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

---

[Kaggle](#)

---

[World Bank Open Data](#)

### u07a1 - Regression Equations

For this assignment, you will practice creating regression equations using a case in your textbook. Read Case Problem 1, Measuring Stock Market Risk, on pages 698–699 of your text, and download the accompanying data set from CengageBrain to create the following in your report:

- Compute descriptive statistics for each stock and the S&P 500.
- Analyze your descriptive statistics for the stocks and the S&P 500, including which stocks are the most volatile.
- Compute the value of beta for each stock.
- Determine which stocks would be expected to perform best in an up market and which would hold their value best in a down market.
- Evaluate how much of the return for the individual stocks are explained by the market.

Remember, to access the Beta data set Web file from the Chapter 14 content, access the free content at CengageBrain.

Review the Regression Equations Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 8 >> Logistic Regression

### Introduction

A logistic regression model is the extension of the idea behind linear regression. It applies to scenarios where the dependent variable, Y, is categorical, that is the outcome is divided into classes; for example, if a dependent variable indicates for a yes or no answer. Another example may be in a way that Y implies a recommendation of go/no go, or in the case of three variables; hold/sell/buy, in the case of stock recommendation. A logistics regression model may be used in cases such as:

- Prediction on financial matters, approval, or disapproval.
- Profiling; studying the factors that differentiate between two classes, such as male and female.
- Classifying returning or non-returning customers.

### Learning Activities

#### u08s1 - Studies

## Readings

Use the Reserved Reading provided for you in this course to complete this reading:

- Read Chapter 10, "[Logistic Regression](#)," in Shmueli, G., Patel, N. R., & Bruce, P. C. (2010). *Data mining for business intelligence*. Hoboken, NJ: Wiley.

Use the Capella library to complete the following reading:

- Fullerton, A. S. (2009). [A conceptual framework for ordered logistic regression models](#). *Sociological Methods & Research*, 38(2), 306–342.

#### u08s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on using Logistic Regression in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on using Logistic Regression in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on using Logistic Regression in SAS.

#### u08s3 - Analytics Internship: Applying Predictive Analytics

In this Vila Health activity, you will learn that the scope for the Vila Health project has been changed in ways that will affect the work and recommendations your team has been preparing. Your group will need to discuss what the changes mean and how they will impact the analytic approach. As you gather information from your mentor, consider what questions you and your team will need to address in order to make your final recommendations. Click **Analytics Internship: Applying Predictive Analytics** to view the Vila Health scenario.

Analytics Internship: Applying Predictive Analytics

**u08s4 - Study Group Task**

This week's group topic for discussion:

- Discuss the process of building and evaluating a logistic regression model. As part of your discussion, explain how you would select predictors and evaluate model fit once the regression is run.
- Discuss as well how to present the output to business leaders, as output from regression is cryptic even for analysts.

**u08d1 - Logistic Regression as a Predictive Model**

For this discussion, provide or create a business question or problem for which logistic regression can be applied as a predictive model to answer the question. How is this model different from a linear regression model?

In SAS, the procedure PROC LOGISTIC is used to conduct a logistic regression. Find or create an output of PROC LOGISTIC for a business example and explain the output to your peers. You can use a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

In your initial post, answer these questions:

- How would you know if the output of your logistic regression model is valid?
- What are the key statistics to determine the validity of your regression model?

In your responses, comment on your classmates' methods in this discussion.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and providing answers to others' questions or offering alternative approaches.

**u08a1 - Real World Logistic Regression**

For this assignment, your instructor will provide you with a scenario and question to investigate and a data set containing current data that supports the scenario. For this assignment:

1. Prepare the data set for analysis and identify variables that could be used in analysis.
2. Run a logistic regression on the data, as appropriate, in order to respond to the question.

The data provided to you is real data collected by a government agency or independent researchers. The data are not perfect; they may contain missing data or missing columns that you need to do your analysis. You may be required to build these columns using formulas, logic, or programming; and you may need to exclude observations from analysis that contain too many missing values. Use what you have learned so far in the course to make judgments on what data you choose to include in your analysis, and be sure to include explanations supported by your course readings or other scholarly sources.

Include the following elements in your deliverable:

- Include a brief summary introduction of the business problem your instructor has asked your analysis to address.
- Describe the process of preparing the data set for analysis.
  - How many observations did you exclude due to missing data?
  - Are there patterns in the data that you excluded? How does this impact analysis?
  - Did you build any new columns? If yes, how were they built?
- Perform a logistic regression on the data that would help investigate the business problem presented by your instructor. Present your output in a way that can be shared with organizational decision makers.
- Assess the quality of the model that you generated. How well does the model fit?
- Assess your presentation of the regression results. Could decision makers act on this information? If so, how? If not, why not?
- Include in an appendix the programs that you wrote to perform the analysis, along with any statistical software output and visuals resulting from the logistic regression analysis. You may use SAS, R or any other statistical software of your choosing.

Review the Real World Logistic Regression Scoring Guide for the detailed grading expectations of this assignment.

## Unit 9 >> Multiple Regression

### Introduction

In previous units, you studied simple linear regression. In that case, you used one independent variable (explanatory) to come up with an estimate of dependent variable (response). In many cases, a more accurate and effective prediction model may be constructed by employing more than one independent variable. As an example, more accurate sales for a corporation may be estimated by considering the number of sales staff as well as advertising dollars as two independent variables that impact the dependent variable. Models that have more than one independent variable are called multiple regression models. In model development of multiple regression analysis, the same concepts that were introduced in simple regression are applicable to multiple regression.

### Learning Activities

#### u09s1 - Studies

### Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 15, "Multiple Regression," to learn about multiple regression.

Use the Capella library to complete the following reading:

- Zhou, S., & Zhu, N. (2013). [Multiple regression models for energy consumption of office buildings in different climates in China](#). *Frontiers in Energy*, 7(1), 103–110. This article illustrates an additional example of multiple regression modeling.

#### u09s2 - Programming Instructions

## Using R

Access the [Using R](#) page on Campus for resources on using Multiple Regression in R.

## Using Python

Access the [Using Python](#) page on Campus for resources on using Multiple Regression in Python.

## Using SAS

Access the [Using SAS](#) page on Campus for resources on using Multiple Regression in SAS.

### u09s3 - Study Group Task

With your group, discuss the process of developing a multiple regression model.

- How would you approach building the regression?
- What assumptions would you check for?
- What data would you need?
- How would you evaluate the accuracy of the model?
- How would you explain the output to a business audience?

Consider these questions in the context of your Unit 10 assignment. Discuss any additional questions you have about approaches for your Unit 10 assignment.

Course Resources

---

[Analytics Internship: Applying Predictive Analytics](#) | Transcript

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[Analytics Internship: Using Data to Drive Performance](#) | Transcript

### u09d1 - Multiple Regression Models

When applying a multiple regression model, describe how you can determine the most influential variables on the dependent variable, or target in the multiple linear regression result.

Use an applicable business problem to support your explanation, if useful. Find an example of an SAS output from a multiple linear regression model and describe how you would explain this output to a business leader. You can use a data set from any of the following items in the Resources:

- The textbook.
- World Bank Open Data.
- Kaggle.

In your initial post, focus on these questions and objectives:

- Which SAS procedure would you run?
- Comment on the overall F test and individual T tests.
- Comment on the P value for each of these tests.
- How would you determine which variable is the most influential?

In your response, comment on your classmates' methods in this discussion.

## Response Guidelines

Respond to at least two of your peers, providing your insight and professional experience and supporting your perspective with APA-cited resources where applicable.

Course Resources

---

[Graduate Discussion Participation Scoring Guide](#)

---

[APA Style and Format](#)

---

[Kaggle](#)

---

[World Bank Open Data](#)

### u09a1 - Regression Analysis

For this assignment, you will practice multiple regression using a case in your textbook. Read Case Problem 1, Consumer Research, Inc., on pages 755–756 of your text, and download the accompanying data set from CengageBrain to create the following in your report:

- Use methods of descriptive statistics to summarize the data and describe the findings.
- Develop an estimated regression equation using annual income as the independent variable.
- Develop an estimated regression equation using household size as the independent variable.
- Analyze whether annual income or household size is the better predictor of annual credit card charges and provide a rationale.
- Develop an estimated regression equation with annual income and household size as the independent variables and analyze your findings.
- Determine the predicted annual credit card charge for a three-person household with an annual income of \$40,000.
- Assess the need for additional independent variables that could be added to an estimated regression model and the value they would add.

Remember, to access the Consumer data set Web file from the Chapter 15 content, access the free content at CengageBrain.

Review the Regression Analysis Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Cengage Brain](#)

---

Modern Business Statistics with Microsoft Excel (6th ed.)

## Unit 10 >> Statistical Application

### Introduction

A multiple regression model is the extension of an idea behind simple linear regression. It applies to the scenarios where the dependent variable,  $Y$ , is a function of many attributes and variables. For example, if a dependent variable indicates for the number of patients visiting an emergency room, it is influenced by many attributions including epidemics, weather, number of 24/7 clinics, and the rate of vaccination, to name a few. The model is being used as a predictive tool and preparation of necessary services to the customer. It has application in health care, finance, operations management, and customer services, among others.

### Learning Activities

## u10s1 - Studies

### Readings

Use your *Modern Business Statistics* text to complete the following readings:

- Read Chapter 16, "Regression Analysis: Model Building," pages 760–796, to learn about regression and model building.

Use the Capella library to complete the following reading:

- Van De Steeg, L., Langelaan, M., & Wagner, C. (2014). Can preventable adverse events be predicted among hospitalized older patients? The development and validation of a predictive model. *International Journal for Quality in Health Care*, 26(5), 547–552.

## u10a1 - Analytics Internship: Applying Predictive Analytics

In Unit 8, Vila Health leadership threw you a curveball; asking you to investigate whether there is a way to predict the length of stay or readmittance rate overall. Now that we know that there are differences in KPIs (over time, between DRGs, or between facilities), leadership is interested in seeing if there is a way in which we can predict readmittance or length of stay. With this information, we can develop a regression model to determine if additional research or a pilot project is needed to influence these KPIs into the future.

- Identify your target (readmittance rate or length of stay).
- Identify the variables that could serve as predictors of length of stay or readmittance rate, including an explanation of why you selected the variables you did.
- Build the appropriate regression model based on your chosen target. Explain your choice of regression.
- Examine your regression results and assess the fit of your model.
- Explain how your regression model could be used to predict your target.
- Identify additional questions you would ask of stakeholders as a result of examining your regression results.
- Create a report for business leaders that shares the data and insights from your results in an informative, ethical, and easy to consume format.
- Include an appendix with samples of the programs that you used in SAS and screenshots of the SAS output used to generate your outcomes.

Review the Analytics Internship: Applying Predictive Analytics Scoring Guide for the detailed grading expectations of this assignment.

Course Resources

---

[Analytics Internship: Applying Predictive Analytics](#) | Transcript

---

[Analytics Internship: Assessing Data](#) | Transcript

---

[Analytics Internship: Using Data to Drive Performance](#) | Transcript

## u10s2 - Personal Effectiveness Assessment

For this activity, you will review your peers—and they you—on personal effectiveness skills, based on contributions to the group discussions. Please complete one survey for each member of your group. The assessment addresses the following characteristics:

- Interpersonal skills.
- Integrity.
- Professionalism.
- Initiative.
- Dependability and reliability.
- Willingness to learn.

At certain points in your educational journey, you will meet with your department head for a 1:1 interview. The reviewer will gather your peer's assessments and your own self-reflections and use the data from these, in conjunction with the insight obtained during the 1:1 interview, to assess your skills. Your department head will provide you with feedback specifically intended to help you improve upon these skills.

During your capstone course the capstone panel of professionals will conduct a final review of your personal effectiveness skills. This final review will be assessed and your scores will be calculated as part of your final grade.

Course Resources

---

### Personal Effectiveness Assessment

#### **u10d1 - Course Reflection**

Explain to your peers and your instructor what you will be taking away from this course:

- What did you learn that surprised you?
- What did you find challenging to understand or grasp?
- What aspects of the course did you enjoy? Which did you not enjoy?
- What would you like to see added to the course for future sessions?

### **Response Guidelines**

Respond to your peers if you would like, but responses for this discussion are not required.

Course Resources

---

Graduate Discussion Participation Scoring Guide