MTH410

Quantitative Business Analysis

Module 1: Introduction to Statistics

Statistics have many applications in the fields of business and economics. In this module, we will explore those applications, identifying different sources and types of data. In Module 1, we will differentiate between samples and populations and utilize descriptive statistics to evaluate data sets.

Learning Outcomes

1. Apply statistics to business and economics problems.
2. Identify various data types and sources.
3. Evaluate data to derive answers from data sets.

For Your Success & Readings

In this course, it is important that you don't let yourself get intimidated by the formulas. If you carefully review the examples of problems in the textbook and match the information to the variables presented, you will find that the math calculation part is quite easy. Also, as in most quantitative sciences, there is a unique vocabulary incorporated into all that you do, so it may help you to write down the different words and their meaning. Thus, you have less page-flipping to do. You can reference the Key Terms glossary at the end of each chapter as well.

If you run into trouble understanding the concepts presented in the lectures or the textbook, don't hesitate to let your instructor know. The concepts contained in the next eight modules build on each other, so it is important that you have a good grasp of them for each module as you progress through the course.

You will be asked to write four managerial reports during the course. Remember to cite all your references and use APA formatting. An example is provided in the MTH410 Guide to Writing with Statistics. If you have additional questions, consult the **CSU-Global Guide to Writing and APA** (http://csuglobal.libguides.com/apacitations).

There are many tools available to analyze data. For this course, Excel is more than adequate. **Here is a series of Excel tutorials** (http://www.gcflearnfree.org/excelformulas). The first five are good to get started in Excel, but review Module 0 for a concise list of tutorials to review before getting too far into the course materials.

Enjoy the course and the new area of ability and insights you will acquire as you learn to manipulate data, formulas, and outcomes.

To navigate this module successfully, keep the following things in mind:

- Read both the required textbook readings and the module content carefully. You will be quizzed on this information in the Mastery Exercise.
- Be sure to complete the Check Your Understanding Activities in the interactive lectures.
- Actively engage in the required discussion question for the week. The weekly discussions are great opportunities to learn from your instructor and fellow course mates. This week's discussion question asks you to create two real-world survey questions that would be useful to you in a professional application or in your everyday life by doing the following two parts:
    - Your first should be a question associated with a categorical (qualitative) variable. Explain the measurement scale associated with the question and if the data collected are cross-sectional or time series. What might you be able to infer about the data you would collect?
    - Your second should be a question associated with a quantitative variable. Explain the measuring scale associated with the question. Also, determine whether the variable associated with the survey question is discrete or continuous and if the data collected are cross-sectional or time series. What might you be able to infer about the data you would collect?
    - Be sure to support your statements with logic and argument, citing any sources referenced. Post your initial response early and check back often to continue the discussion. Be sure to respond to your peers' and instructor's posts, as well.

Also, remember that there are Live Classroom sessions in Weeks 3 and 6. While optional, this is an excellent opportunity to discuss any questions, concerns, or other course-related topics with your instructor.

Required

- Chapter 1 in *Introductory Business Statistics*

Recommended

- Taylor, C. (2018). Understanding statistics. *ThoughtCo*. Retrieved from **https://www.thoughtco.com/what-is-statistics-3126367** (https://www.thoughtco.com/what-is-statistics-3126367)
- Vaskevich, A., & Luria, R. (2018). **Adding statistical regularity results in a global slowdown in visual search** (https://www-sciencedirect-com.csuglobal.idm.oclc.org/science/article/pii/S0010027718300180). *Cognition, 174*, 19.

1. The Role of Statistics, Definitions

Statistics play an important role in society and in understanding information. With technological advances, we now have an abundance of information and statistics on every facet of our lives. Statistical evaluations can help us make sense of that information.

There are five major reasons to study statistics. *Click on each tab below to discover each of these reasons.*

Conduct Research Effectively

Without the use of statistics, it would be very difficult to make decisions based on the data collected from a research project. Statistical analysis allows us to better understand scenarios from which to make educated decisions. While you may never plan to be involved in research, it is inevitable that it will find its way into your life. If you decide to continue your education, you will likely need to conduct or review statistical information in your course work. More importantly, increased competition in the marketplace has resulted in increased research, so it is very likely that you will find yourself working on or reviewing studies and their statistical findings, no matter what your job.

Comprehend Journals and Articles with Research Information

An understanding of basic statistics will provide you with the fundamental skills necessary to read and evaluate the results found in these articles so that you can evaluate the information critically to enhance your knowledge and understanding.

Develop Critical Thinking and Analytic Skills

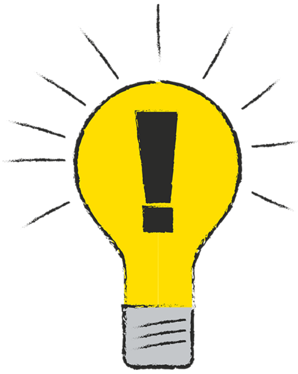
The study of statistics will help to enhance and further develop these skills by developing and using formal logical thinking abilities that are both creative and at an advanced level.

Be an Informed Consumer and Society Participant

Even well-meaning individuals can unintentionally report erroneous statistical conclusions, so your ability to understand the basics of statistical calculations will help you to determine the appropriateness and accuracy of information you are provided. If you know some basic statistical concepts, you will be in a better position to evaluate the information you have been given.

Know When You Need a Professional Statistician

As an organizational leader or manager, you may need to hire a consultant to assist you in gathering and interpreting data. Your basic knowledge of statistics will help you to know if the information you are receiving is accurate and worthwhile.



Learn More

Take some time to read this article about the importance of statistics: **The Importance of Statistics** (http://statisticsbyjim.com/basics/importance-statistics/)

Statistics is especially important in the social sciences, because they rely heavily on data analysis for the advancement of knowledge. In this area, statistics are used to organize, evaluate, and analyze data to make effective use of information.

Statistical analysis incorporates **variables**, which are elements that provide or describe the data. Variables can be hair color, weight, age, blood type, and time. They all describe data.
Data are the actual values of the variable of interest. **Statistics** deals with the collection, analysis, and interpretation of data.

The **population** is the entire group that a researcher is interested in studying. A **sample** is the subset A subset is part of a larger group of related things of the population that one observes. For example, if we are interested in the average weight of an adult in Colorado, then the population is all Colorado adults. If we randomly pick 100 Colorado adults to weigh, then those 100 people form a sample.

Here is an explanation of the population in statistics: **What Is a Population in Statistics?** (https://www.thoughtco.com/what-is-a-population-in-statistics-3126308)

There are two major branches of statistics. Organizing and summarizing data is **descriptive statistics**. **Inferential statistics** involves using probability to generalize research findings from the sample to the population.

Describe variable distribution

When a researcher needs to summarize or describe the distribution (or range) of one or more variables, the values or scores of those variables are arranged so that the relevant information can be quickly understood and appreciated. Percentages, graphs, and charts can all be used to display descriptive statistics.

Example—The percentage of CSU-Global students who eat certain snack foods while studying:

Food Type	**Men**	**Women**
Cookies	10%	20%
Chips	25%	10%
Popcorn	30%	40%
Candy	5%	3%

Understand the relationship between variables

When a researcher wants to understand the relationship between two or more variables. Descriptive statistics allow the researcher to quantify the strength and nature of a relationship. They are useful because they enable us to investigate two matters of central theoretical and practical importance to any science: causation and prediction. These techniques help us determine the connections between variables and to trace the ways in which some variables might have causal influences on others to enable us to predict scores on one variable from the scores on another.

Example—The time students spend studying and their average GPA. Is there a correlation between the two variables?

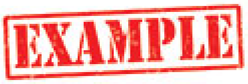
1.1. Parameter Versus Statistic

First determine what the population is. If one is sampling from the entire population, then the average is a parameter. If one is sampling from a proper subset of the population, then the average is a statistic. We use inferential to generalize from the statistic to the parameter.

See the following site describing the difference between a parameter and a statistic:

- **Difference Between a Statistic and a Parameter** (https://www.thoughtco.com/difference-between-a-parameter-and-a-statistic-3126313)

1.2. More on Types of Quantitative Variables



Here are more examples on the types of quantitative variables.

A shoe store manager is interested in the prices of running shoes. Is the price a discrete or continuous variable?

Click "Solution" to check your thinking.

Solution

The price is a discrete variable. One can list prices in dollars and cents (e.g. $25.00, $25.01, $25.02,...).

A shoe store manager is interested in the temperature of the store. Is the temperature a discrete or continuous variable?

Click "Solution" to check your thinking.

Solution

The temperature is a continuous variable. Temperature can be any number on an interval, in an analog thermometer.

A shoe store manager is interested in the time it takes to attend a customer. Is the time a discrete or continuous variable?

Click "Solution" to check your thinking.

Solution

The time taken is a continuous variable. Time can be any number on an interval, in an analog clock.

The following two sites further explain the difference between discrete and continuous variables:

- **Math Is fun** (https://www.mathsisfun.com/data/data-discrete-continuous.html)
- **Statistics How To** (https://www.statisticshowto.datasciencecentral.com/discrete-vs-continuous-variables/)

2. Data and Sampling

The Four Levels of Measurement for Classifying Data

Data can be classified into one of the following four levels of measurement. *Click on each tab to learn more.*

Nominal LevelOrdinal LevelInterval LevelRatio Level

A **nominal level** of measurement describes data that consist of names or labels. For example, gender (male or female), shirt color, and marital status (married, single, divorced, widowed) are all nominal scales.

The level of measurement for a variable is **ordinal level** if it is nominal and the order or rank of the data is meaningful. For example, the grade of a student can be an ordinal level. This is because A can be the highest grade and F can be the lowest. Ordinal data may also be numerical. For example, pain level on a scale of 1 (lowest) to 5 (highest) is an example of an ordinal scale of measure.

Variables have an **interval level** of measurement if they are ordinal and the difference between values is expressed in terms of a fixed unit of measure. For example, the Fahrenheit temperature scale is an example of an interval level. Temperature can be ranked. For example, -10 degrees Fahrenheit is lower (colder) than 50 degrees Fahrenheit. Also, the difference between degrees is meaningful. For example, suppose it was 80 degrees Fahrenheit yesterday and 90 degrees Fahrenheit today. Then it was 90-80=10 Fahrenheit degrees warmer today. The zero temperature is just like any other value, with numbers below and above it.

Variables have a **ratio level** of measurement if they are interval and the ratio between two values is meaningful. An example of a ratio variable is distance in meters. Distance is an interval level, and the ratios of distances is meaningful. For example, suppose Joe is 5 meters away from Sam's house. Suppose Bill is 10 meters away from Sam's house. Then, the ratio of Joe to Bill's distance is 5/10=1/2. This means that Joe is 1/2 as far as Bill from Sam's house. A ratio level has an **absolute zero**. This means that a zero value indicates that nothing exists for the variable at that zero point. In other words, no number exists below this point. For the distance example, there is no distance less than 0. The highest scale of measurement is the ratio scale. This is because ratio scale data incorporate all of the other scales.

Arithmetic operations of addition, subtraction, multiplication, and division can be applied to both Interval and ratio variables. Because one can do more arithmetic, there are more statistical methods for interval and ratio variables than nominal or ordinal variables.

The following is a summary of types of data: **What Is Quantitative Data?** (https://www.thoughtco.com/definition-of-quantitative-data-3126331)

In a **random sample**, each data point has the same chance of being selected. For example, suppose a room of students has 9 males and 9 females, and you will select 9 people by tossing a coin. If the coin toss results in heads, you will select the 9 males. Otherwise, you will select the 9 females. Each of the 20 persons has a 1/2 or 50% chance of being selected.

However, in many statistical techniques or formulas, a random sample is not sufficient. It is often a requirement that one have a simple random sample. In a **simple random sample of n subjects**, one assumes that all *possible* samples of size n have the same chance of being selected from the population. The coin toss example is not a simple random sample. The (n=9) sample consisting of 5 males and 4 females is not possible. A simple random sample is a stronger requirement than a random sample. To obtain a simple random sample from the above example, one could number the students from 1 to 18. Then, randomly select 9 numbers.

There are alternative sampling methods. *Click on the tabs below to learn more about them.*

Stratified Sample

In a **stratified sample**, one subdivides the population into 2 or more subgroups (strata). Then, select a proportional number of subjects from each subgroup. For example, suppose a room has 12 democrats and 10 republicans and the stratum is party affiliation. Suppose the sampling fraction is 1/2. Then we would randomly select 6 Democrats and 5 Republicans.

Cluster Sample

In a **cluster sample**, one divides the population into clusters (groups). Then, randomly select some of those clusters, and then select all members of those clusters. For example, suppose a building has 20 classrooms (clusters). Randomly select three classrooms. Then, sample all students in the three classrooms.

Systematic Sample

In a **systematic sample**, select some starting point. Then, select every k^{th} data point of the population. For example, suppose a room has 20 people, randomly numbered from 1 to 20. We will then start at the first person and select every third person in the room (1, 4, 7, 10, 13, 16, 19).

Convenience Sample

In a **convenience sample**, one uses data that are easy to obtain. For example, the franchise owner samples employees who happen to be working a store during a visit. Another example can be administering a survey only to nearby friends and immediate family.

Sample Errors

In statistics there can be various types of errors. For example, a **sampling error** (or random sampling error) occurs when the method is random, yet there is a discrepancy between the sample and the population. This can be due to randomness or unexpected results. For example, the sample size might not be large enough, or the results are unusual.

A **non-sampling error** is the result of human error. For example, there might be a defective counting device, poorly phrased survey questions, false data reporting, or incorrect statistical methods.

A non-sampling error occurs when the sampling method is not random. For example, there was a convenience sample, or the respondents were voluntary. An example of voluntary responses can be a call-in survey.

Bias in Samples

A sample is **biased** if the results from the sample are not representative of the population. A sample bias can occur due to **sampling bias**. Sampling bias occurs if the sampling method favors one subset of the population over the other. For example, a business owner might conduct a shopping survey to credit card customers only. Those without credit cards might spend less than customers with a credit card. Another reason for a biased sample can be due to a nonresponse bias. A **nonresponse bias** occurs when those who do not respond to a survey tend to have a different opinion than those who do. For example, customers who are unsatisfied with service at a restaurant tend to go out of their way to respond to a satisfaction survey more so than those who felt everything was okay. A third reason for sample bias can be due to response bias. **Response bias** occurs when survey answers do not reflect the true feelings of the respondent. For example, an employee might be afraid of giving his/her boss poor survey scores. Another example of response bias can occur when a survey question is poorly designed or phrased. Here is an example of a poor survey question:

Do you like low prices and are satisfied with our product?

A person might be inclined to respond “yes” because people like low prices.

Besides a biased sample, there are other ways statistics can be misleading. Three ways are outlined below. *Click on each tab to learn more.*

Calculation ErrorSmall Sample SizeOutliers

One error might be a **calculation error**. For example, in a survey, 60 out of 200 people prefer online shoe-shopping. The results might state that one-third of people prefer online shoe-shopping when the correct answer is 30%.

Another source of misleading statistics can be due to a **small sample size**. A manager of a factory with 100 employees might conduct an employee survey regarding benefits. If only 3 employees were surveyed, then that would be an insufficient sample size.

Statistics can be misleading if there are **outliers**. For example, a survey might ask 100 people how much money they make per year. However, two of the respondents are very wealthy. If one were to compute the average wage of those 100 people, the results might be unusually large due to the two wealthy individuals. Another example can be where an instructor gave an exam to 35 students and one student did not even try on the test. The student got a zero. The class average will be lowered substantially because of the one zero score.

When studying how one variable depends on another, one often encounters the **independent** or **explanatory variable**. The researcher often has control over or can manipulate the independent variable. That is why the independent variable is often called the **treatment variable**. The other variable is called the **dependent** or **response variable**. For example, the independent variable can be hours shopping and the dependent variable can be the corresponding amount spent. Sometimes, there might be a linear relationship between the independent and dependent variables, $y = mx + b$. For example, x = hours shopping and y = amount spent. In Module 8, we will study such linear relationships.

See the following site explaining dependent and independent variables: **What Is the Difference Between Independent and Dependent Variables?** (https://www.thoughtco.com/independent-and-dependent-variables-differences-606115)

Types of Study

An **experimental unit** is the subject that is tested. For example, a shopping customer is an experimental unit.

In an **observational study**, the researcher does not manipulate the experimental unit in any way. The researcher records or measures the results of the experiment but does not manipulate the independent variable. A regular survey is an example of an observational study. For example, 1000 customers were surveyed on whether they prefer online shopping. Another example is a study on smoking. Researchers observed the health of both smokers and non-smokers.

However, in an **experiment**, the researcher applies a treatment to the experimental unit. For example, employers added classical music to the work environment to see if productivity increased. Productivity was measured for various days without classical music. Productivity was measured when the classical music was played.

Researchers must be aware of lurking variables. A **lurking variable** is one that affects the relationship between variables but is not part of the study. A famous example is an increase in drownings when there are more ice cream sales. One could erroneously conclude that ice cream contributes to an increase in drownings, or vice versa. However, the lurking variable is temperature. A higher temperature contributes to the increase in both drownings and ice cream sales.

Control vs. Treatment Groups

To understand treatment effects, researchers often separate experimental units into two groups: the control group and the treatment group. A **treatment (or experimental) group** is the one receiving treatment. A **control group** is one that does not receive the treatment. For example, a researcher might be interested if certain new music in stores increases sales. The control group would be the customers listening to regular music. The treatment group can be the customers listening to the new music. A classic example is when some patients are given a new medicine and others receive a placebo. The ones receiving the placebo form the control group.

People are often not aware of being part of the control group, to reduce the power of suggestion. **Blinding** in a randomized experiment occurs when the experimental unit is not aware of being part of the treatment or control group. In a **double-blind experiment**, both the researchers and experimental units are blinded. For example, both the researcher and patient are unaware of who receives the placebo during the experiment. This can prevent the researcher from inadvertently telling the patient whether he/she is receiving a placebo.

2.1. More on Scales of Measurement

Here are some examples on the level of measurement:

Question: Suppose the grade on an exam is A, B, C, D, or F. What is the level of measurement of the grade?

Answer: The grade can be ordinal. A is the highest grade and F is the lowest: A>B>C>D>F.

Question: Suppose the model of car that a customer drives is recorded. What is the level of measurement of car model?

Answer: Car model level of measurement is nominal. Car model is not ordinal since there is no order assigned to a particular model.

See the following explanation on levels of measurement: **The Levels of Measurement in Statistics** (https://www.thoughtco.com/levels-of-measurement-in-statistics-3126349)
Here are many levels of measurement problems and solutions: **Levels of Measurement Worksheet With Solutions** (https://www.thoughtco.com/levels-of-measurement-worksheet-solutions-3126514)

2.2. More on Sampling

Here is an explanation of stratified sampling: **Understanding Stratified Samples and How to Make Them** (https://www.thoughtco.com/stratified-sampling-3026731)

Here is an explanation of cluster sampling: **Cluster Sampling** (https://research-methodology.net/sampling-in-primary-data-collection/cluster-sampling/)

A cluster sample might be useful when it might be difficult to sample all experimental units. For example, some samples might be dispersed geographically. It might be less expensive to randomly choose certain geographic regions and sample from those regions. A disadvantage might be that the experimental units of each geographic region are similar. One difference between a cluster and stratified sample is that the entire population is selected from each of the chosen clusters. With a stratified sample, only a proportion of that population is selected from each stratum. Also, in a stratified sample, elements from all the strata are selected.

For example, suppose a survey is to be conducted to understand soda preferences. The US is divided into four regions. Three hundred people are randomly surveyed from each region. This is an example of a stratified sample since an equal number of people is chosen from each region and all 4 regions are sampled. The strata are the US regions.

Suppose a survey is to be conducted to understand shopping in a grocery store. Three aisles from the store are randomly chosen. Then everyone in those three aisles is surveyed for two hours. This is an example of a cluster sample since three aisles are randomly chosen and all customers from each aisle are chosen. The clusters are the store aisles.

3. Summary

Statistics is both an art and a science, as it involves collecting, analyzing, and interpreting data. Although attempts can be made to make the statistical process purely objective and to provide “the truth,” the injection of human participation into the process can provide varying results. By understanding the process of statistical calculation and evaluation, you will be able to understand the areas that can be subject to human impact so that you can account for them in your decision making as you review calculated outcomes. As a reminder, you may review any of the material if you do not fully understand or reach out to your instructor.

Here is the list of the objectives that we have covered and are part of the Mastery Exercises in Knewton Alta:

- Understand the definitions of population, sampling, statistic, parameter, and data in business applications
- Identify stratified, cluster, systematic, and convenience sampling in business applications
- Identify sampling errors and bias in business applications
- Identify situations in which business statistics can be misleading
- Determine whether a study is observational or an experiment and appropriate use cases
- Identify and describe the steps in the statistical analysis process
- Identify explanatory and response variables in an experiment using business examples
- Identify levels of measurement of business-related data
- Define and distinguish between qualitative, quantitative, discrete, and continuous variables

Check Your Understanding

Embedded Media Content! Please use a browser to view this content.

References

None