Scientific Method and Application of Statistics in Business

By: Seyedeh Fateme Moezzi

In the name of God

A NOTE TO READERS, TEACHERS and …..

Welcome to our book. Learn about methods application and research design with stories from researchers in the field. Hundreds of titles provide a quick definition of methods to comprehensive exploration of concepts and methods. A Little Book, which summarizes many of the recent advances in the **Scientific Method and Application of Statistics in Business.** We have a Little Green Book for that!

Chapter1: The Scientific Method in agricultural and environment education

Lesson1: Scientific Method

The scientific method is a set of principles and procedures followed to gain knowledge through making questions and observations, performing experiments, and formulating and testing a hypothesis. Explore the key elements and real-life applications of the scientific method. Is There Only One Scientific Method?

When you first took science class in school, you probably learned the basic steps of a scientific investigation. You've likely heard of words like 'hypothesis,' 'experiment,' and 'observation.' You may have even memorized a prescribed set of steps. The scientific method is a set of procedures that scientists follow in order to gain knowledge about the world.

However, the steps involved in the scientific method vary widely among the different scientific disciplines. Chemists follow the method a bit differently than psychologists. Geologists and botanists have their own unique methods. So, is there really one scientific method that encompasses all of science? To find out, we'll need to learn more about the scientific process.

Key Elements of the Scientific Method

There are six key steps that tend to characterize the scientific method. The first step is the question. This is the part where a scientist proposes the problem that he or she wants to solve. A well-conceived question usually leads to a hypothesis, a potential answer to the question at hand. Sometimes, hypotheses look more like predictions. The scientist predicts what the outcome will be when he or she tests the hypothesis. The scientist's test is also called the experiment. Experiments are ordered investigations that are intended to prove or disprove a hypothesis. Important data comes from performing an experiment.

The scientist has to make observations of the results that he or she gets from the experiment. An observation is a statement of knowledge gained through the senses or through the use of scientific equipment. Observations are crucial for collecting data. Once the results are in, the scientist must begin the analysis. Data analysis involves comparing the results of the experiment to the prediction posed by the hypothesis. Based on the observations he or she made, the scientist has to determine whether the hypothesis was correct. He or she then sums up his or her findings with a conclusion. The conclusion of a scientific process is a statement of whether the original hypothesis was supported or refuted by the observations gathered.

The six steps of the scientific method do not always occur in the same order. The elements of the scientific method.

The scientific method usually employs all six of the steps I mentioned, but the steps don't always occur in the same order. Real scientists may go back and repeat steps many times before they come to any conclusions. It's actually better to use the word 'elements' to describe the steps, since the first step, question, does not always come first. Sometimes, for example, it's an observation that came first and spawned the initial question. Likewise, observations that are made during an experiment can inspire more questions that scientists have to answer. The scientific method is much more fluid than you might think. Let me show you how the steps can feed back and branch out from one another with an example from my own experience.

Environmental problems can be investigated in a systematic way using the scientific method. Explore the steps of the scientific method: observations and questions, hypothesis, experiment, interpreting results and making conclusions. Discover the definition and applications of the scientific method in this lesson.

The Scientific Method

All science begins with making observations and asking questions. These questions are often developed because people want to learn more about what they are observing. To find the answers to their questions, researchers use the scientific method, which is a structured technique used to test ideas and potential answers to scientific questions. Although the specific steps of the scientific method can vary by discipline, it is widely agreed upon that the fundamental elements of the scientific method include five steps. The scientific method starts with:

- Making observations, and then moves on to
- Asking questions
- Formulating a hypothesis
- Conducting an experiment, and ends with
- Interpreting results and making conclusions

Observations and Questions

Most scientific evaluations begin with someone making observations about something that is occurring in nature or the world around them. Observations are the first step in the scientific method, and they set the foundation that the proceeding steps are based on. Once a scientist has made observations, they are often curious about what they have seen and want to learn more. At this point, scientists begin to ask questions about their observations, which is the second step in the scientific method. These questions can be very broad or specific depending on the topic being investigated.

Hypothesis

After a question has been asked, the next step in the scientific method can occur. The third step is to formulate a hypothesis. A hypothesis is a proposed explanation that aims to answer the question formulated, and the hypothesis must be testable through experimentation. The development of a hypothesis is a very important step in the scientific method because it is what the remaining steps are based on, and it will be returned to after more steps are completed. Once the hypothesis is created, scientists make predictions about what should happen if the hypothesis is found valid, and therefore supported, and also what should happen if the hypothesis is not valid and is rejected.

Experiment

Conducting an experiment is the fourth step in the scientific method. An experiment is an activity designed to gather data that will be used to support or reject the hypothesis. An experiment involves multiple variables, which are specific factors that can be manipulated. Most experiments include an independent variable, which is the factor that the scientist alters, and a dependent variable, which is the factor that is being measured. When an experiment includes these aspects, it is referred to as a controlled experiment because the scientist is in control of how each variable influences the experiment. This is an important part of the scientific method because by controlling the variables in the experiment, the scientists can determine which variable is causing the predicted result.

What is the next step if the data doesn't support the hypothesis?

- ○ The hypothesis is rejected and a new hypothesis is developed.
- ○ The hypothesis is rejected and the experiment is over.
- ○ The hypothesis is accepted and the research is published.
- ○ The hypothesis is accepted and new hypothesis is developed.
- ○ The hypothesis is accepted and the research is repeated.

Which type of experimental variable is the factor that the scientist is altering?

- Independent variable
- Dependent variable
- Control variable
- Altered variable
- Changed variable

What is the correct order of the steps of the scientific method?

- Observations, questions, hypothesis, experiment, interpreting results and making conclusions
- Observations, questions, predictions, experiment, interpreting results and making conclusions
- Observations, hypothesis, experiment, interpreting results and making conclusions
- Questions, hypothesis, experiment, interpreting results and making conclusions
- Questions, hypothesis, predictions, experiment, and interpreting results and making conclusions

Lesson2: Formulating the Research Hypothesis and Null Hypothesis

Generally, research projects are designed to test a hypothesis, which is a statement of the condition to be studied or the question to be answered. Research projects actually prove or disprove a null hypothesis, which states the opposite of the hypothesis. Learn how to formulate the research hypothesis and null hypothesis, including research questions and if/then statements, and understand the concept of invalid hypothesis.

Research Question

After determining a specific area of study, writing a hypothesis and a null hypothesis is the second step in the experimental design process. But before you start writing a hypothesis and a null hypothesis, which we will get to, you have to have a question. This is the bottom, or base, which you will build up from.

What are you interested in? What are you curious about? This is a good place to start because your research should answer the question. I am curious about the effects of bright lights on studying. You take the thing you are interested in and turn it into a question. Here is mine: 'What is the effect of bright light on studying?' That's how easy it is to write a research question. Next we will explore how to formulate a research hypothesis based on your research question, then we'll look into what a null hypothesis is and how to write one of these.

Formulating a Hypothesis

You have a question and now you need to turn it into a hypothesis. A hypothesis is an educated prediction that provides an explanation for an observed event. An observed event is a measurable result or condition. If you can't measure it, then you can't form a hypothesis about it because you can't confirm or reject it. In addition, a hypothesis typically takes the form of an if-then statement so you can test it with your research. What does our hypothesis look like?

'If we increase the amount of light during studying, then the participant's performance on test scores will decrease.'

Let's break down our hypothesis. First off, it is an if-then statement: 'If we increase..., then the participant's...' This creates a prediction that we can test by increasing the light on participants as they study and then see if their test scores go down. It also means that the hypothesis can be proven correct or incorrect based on what happens to the test scores. If test scores don't change, then our hypothesis was incorrect and we will reject it.

You probably also noticed that we changed 'studying' to 'test scores' and the vague term about 'bright light' into 'amount of light.' This is an example of operationalizing, which is finding a way to measure or quantify a variable. Studying can't really be researched, but test scores can. And they are basically the same thing since studying typically increases test scores. Also, simply saying 'light' is too vague to be useful or researched, so it was turned into 'amount of light.'

Null Hypothesis

After you formulated your research hypothesis, what if there isn't a connection between light and studying? That is kind of what a null hypothesis is; a null hypothesis is defined as a prediction that there will be no effect observed during the study. The reason researchers develop a null hypothesis is to ensure that their research can be proven false. So whenever you are conducting an experiment with a hypothesis, you will create a null hypothesis. Research typically includes a hypothesis, and when this is the case you will form a null hypothesis as

Starting off with a research question provides a researcher with:

- A strong foundation and a question to answer
- Their hypothesis
- Their null hypothesis
- Accurate information about how to accomplish the experiment

A hypothesis is:

- A research question
- A description of an event
- An educated prediction
- A sentence proving a phenomenon
- A psychological diagnosis

Which of the following has NOT been operationalized?

- Test scores
- Popularity
- Weight
- Amount of light

Lesson3: Research Variables: Dependent, Independent, Control, Extraneous & Moderator

In experimental research, factors that have any varying quality or quantity are known as a Variable. Discover the uses of Independent, Dependent, Control, Extraneous, and Moderator variables in conducting research.

As a researcher, you're going to perform an experiment. I'm kind of hungry right now, so let's say your experiment will examine four people's ability to throw a ball when they haven't eaten for a specific period of time - 6, 12, 18 and 24 hours.

We can say that in your experiment, you are going to do something and then see what happens to other things. But, that sentence isn't very scientific. So, we're going to learn some new words to replace the unscientific ones, so we can provide a scientific explanation for what you're going to do in your experiment.

The starting point here is to identify what a variable is. A variable is defined as anything that has a quantity or quality that varies. Your experiment's variables are not eating and throwing a ball.

Now, let's science up that earlier statement. 'You are going to manipulate a variable to see what happens to another variable.' It still isn't quite right because we're using the blandest term for variable, and we didn't differentiate between the variables. Let's take a look at some other terms that will help us make this statement more scientific and specific.

Dependent and Independent Variables

A moment ago, we discussed the two variables in our experiment - hunger and throwing a ball. But, they are both better defined by the terms 'dependent' or 'independent' variable.

The dependent variable is the variable a researcher is interested in. The changes to the dependent variable are what the researcher is trying to measure with all their fancy techniques. In our example, your dependent variable is the person's ability to throw a ball. We're trying to measure the change in ball throwing as influenced by hunger.

An independent variable is a variable believed to affect the dependent variable. This is the variable that you, the researcher, will manipulate to see if it makes the dependent variable change. In our example of hungry people throwing a ball, our independent variable is how long it's been since they've eaten.

To reiterate, the independent variable is the thing over which the researcher has control and is manipulating. In this experiment, the researcher is controlling the food intake of the participant. The dependent variable is believed to be dependent on the independent variable.

Your experiment's dependent variable is the ball throwing, which will hopefully change due to the independent variable. So now, our scientific sentence is, 'You are going to manipulate an independent variable to see what happens to the dependent variable.'

Unwanted Influence

Sometimes, when you're studying a dependent variable, your results don't make any sense. For instance, what if people in one group are doing amazingly well while the other groups are doing about the same. This could be caused by a confounding variable, defined as an interference caused by another variable. In our unusually competent group example, the confounding variable could be that this group is made up of players from the baseball team.

In our original example of hungry people throwing the ball, there are several confounding variables we need to make sure we account for. Some examples would be:

Metabolism and weight of the individuals (for example, a 90 lb woman not eating for 24 hours compared to a 350 lb man not eating for 6 hours)

Ball size (people with smaller hands may have a difficult time handling a large ball)

Age (a 90-year-old person will perform differently than a 19-year-old person)

Confounding variables are a specific type of extraneous variable. Extraneous variables are defined as any variable other than the independent and dependent variable. So, a confounding variable is a variable that could strongly influence your study, while extraneous variables are weaker and typically influence your experiment in a lesser way. Some examples from our ball throwing study include:

If I release 10 spiders into a room and measure how long it takes people to get out of the room, what is the independent variable?

- The spiders
- The people
- The amount of time it takes the people to leave the room
- The room

A researcher wants to figure out if uncomfortable seats in a classroom will increase the research participant's attention to the lesson and decrease their fidgeting. Why are attention level and fidgeting the dependent variable?

- Because it is the variable that the researcher changes to see how the other variables will respond.
- Because it is the variable that the researcher is manipulating.
- Because it is the variable the researcher is primarily interested in.
- It isn't; the dependent variable is the uncomfortableness of the seats.

Which of the following best describes control variables?

- They provide the researcher with control over the independent variable.
- They are kept the same over multiple trials.
- They increase the variability of the experiment.
- They increase the participants' control over the experiment.

Lesson4: True Experimental Design

Scientists in almost every field of study use experiments to answer research questions. Imagine you are a psychologist, and you want to investigate whether caffeine has an effect on student behaviors and performance in the classroom. How would you go about finding out the answer to this question? The answer is that you would do an experiment. This lesson covers all of the different aspects of an experiment you would want to consider.

Independent and Dependent Variables

The first thing any experimenter needs to decide is what variables you are studying. Let's imagine your hypothesis is that when students in school consume caffeine, their performance on tests is affected. You might hypothesize that caffeine increases test performance because it causes the students to be less sleepy and more focused, or you might hypothesize that caffeine decreases test performance because it makes the students jumpy and hyper. Either way, you have two variables involved in this study.

The independent variable in an experiment is the variable that you control as the experimenter and the one that creates two or more groups in the study. In order to study caffeine, you might give half of the students a caffeinated drink and the other half of the students simply get water. The difference between the two groups is whether they have caffeine or not. So, the independent variable is the variable that you, as the experimenter, have manipulated.

The dependent variable in an experiment is the outcome variable or the one you are simply measuring. Here, you guessed that caffeine might affect test performance. So, in this example, your dependent variable is test performance.

Another way to think about independent variables and dependent variables is in terms of cause and effect. This study is testing whether caffeine (the cause) has an effect on test performance. All experiments are testing if whatever makes the groups different has an effect on some outcome variable. The independent variable is always the cause. Here, that's the caffeine. The dependent variable is always the effect. Here, that's test performance. So, the independent variable always happens first, and the dependent variable always happens second.

Experimental vs. Control Groups

Now, let's talk about why we need more than one group in an experiment. Imagine you went into a classroom, gave every student caffeine and then tested them on some kind of performance measure, such as the number of times they can jump a rope. You can see how these students performed after having caffeine. But how can you know if their performance was increased or decreased compared to what they would have done without caffeine? With only one group in your study, you can't be sure what the effects of caffeine were.

So, in an experiment we always need at least two groups to compare. Let's go back to the example of giving half of the students caffeine and half of the students water to drink. When we're testing for the effect of the independent variable, we want to make sure that one of the groups in our study can serve as the natural, or baseline, group. That natural or baseline group is called a control group. In our example, the control group would be all of the children who only drank water.

We then compare the control group to the group of children who received caffeine. In an experiment, the group that receives some kind of change to their natural environment is called the experimental group. In our example, the experimental group would be all of the children who drank caffeine.

We need at least two groups so that we can compare the experimental group to the control group. Control groups are especially necessary when testing for the effect of drugs, like caffeine, because we want to make sure the group doesn't change simply because they think they are supposed to. When you change your behavior just because you expect a change, that's called the placebo effect. For example, let's say we give all of the children soda, but half of the sodas have caffeine and half are caffeine-free. We wouldn't want to tell the children which kind of soda they got, because they might change their behavior simply due to expectations. This type of problem can be avoided with a good control group.

Random Assignment

When we divide the class up into the control group versus the experimental group, it's important to make sure that this division occurs at random. When each person in the study has an equal chance of being in either the control group or the experimental group, that's called random assignment. You might decide which group each person is in by flipping a coin, as an example.

Why is random assignment important? We want to make sure that the groups are as identical as possible in every way except for the independent variable. Let's go through an example of why this matters.

Imagine that you decided that all the boys in the class would get the caffeine drink, while all the girls in the class got the no-caffeine drink. Then, you tested for the effects of caffeine using the dependent variable of jumping rope. Now, imagine that you see a difference! The caffeine group is better at jumping rope. But, you can't actually conclude that caffeine was the cause. It could be that boys are better at jumping rope. Unless the groups are identical in every possible way except for the independent variable, you can't be sure what caused any difference in the dependent variable. But if the independent variable really is the only difference between groups, then you can be sure, because there's no other explanation. This is why random assignment is so important: only with random assignment can you be sure of a cause-effect relationship within an experiment.

Dr. Patel does an experiment in which she shows people either a happy movie, a sad movie, or no movie in a classroom. Then, she asks them to fill out a survey assessing their feelings about whether they want to adopt a pet from the animal shelter. In this experiment, what is the INDEPENDENT variable?

- The participants
- The classroom location
- Their feelings about pet adoption
- How long the movie lasts
- The type of movie viewed

Dr. Patel does an experiment in which she shows people either a happy movie, a sad movie, or no movie in a classroom. Then, she asks them to fill out a survey assessing their feelings about whether they want to adopt a pet from the animal shelter. In this experiment, what is the DEPENDENT variable?

- The participants
- The classroom location
- Their feelings about pet adoption
- How long the movie lasts
- The type of movie viewed

Dr. Patel does an experiment in which she shows people either a happy movie, a sad movie, or no movie in a classroom. Then, she asks them to fill out a survey assessing their feelings about whether they want to adopt a pet from the animal shelter. In this experiment, what was the control group?

- The group of pets who are currently in the animal shelter
- The larger population of people who live in that town
- The people who watched no movie
- The people who watched the sad movie
- The people who watched the happy movie

Chapter2: Application of Statistics in Business

Lesson5: What Is Business Statistics?

Business statistics is a specialty area of statistics which are applied in the business setting. It can be used for quality assurance, financial analysis, production and operations, and many other business areas. Just as in general statistics, there are two categories: descriptive and inferential. Descriptive statistics are used to describe the total group of numbers. Inferential statistics infers relationships from the population of numbers.

Use of Descriptive Statistics

Descriptive statistics are used to summarize and describe total numbers. Looking at statistical numbers such as mean, or the average number, mode, or the most frequent number, or median, or the middle number, helps managers monitor business activities and make decisions. Often numbers themselves do not show the big picture, so ratios, or numbers representing relationships are used.

Perhaps you're a regional manager who oversees 15 different car dealerships. You will keep track of sales per month, number of vehicles sold, number of salesmen, sales per person, operational costs, delivery times, and other information. You use this statistical information to look at trends, understand relationships between numbers, and make sound business decisions. Perhaps you need to shift a sales member to a different store, or realize that you need more available stock during certain months, all of which helps maximize the company's resources and profit.

Or, perhaps, you're a regional store manager and are concerned about how long customers have to wait to check out. Currently, customers queue in multiple lines with a single cashier. You measure the mean length of time for customers to complete their transaction for a month. Then the next month, you line up customers in a single line with multiple cashiers and measure the mean length of time for the transaction. You notice that even though customers take longer to walk to the next available cashier, the overall transactions are reported faster. When the company expands and builds the next store, you specifically request it to be designed with one queuing line, using your findings to substantiate the request.

Use of Inferential Statistics

Inferential statistics help managers draw conclusions based on limited data. When predicting the future, we don't have a magic crystal ball, but we do have statistical strategies, such as sampling, probability, and models.

Marketing departments often use inferential statistics. A company might issue a survey and ask questions about their products. However, it's impossible to survey every individual customer. The marketing department will determine the appropriate sample size, or the number of people to ask. Based on the results, statisticians can infer the responses are representative of the larger group of customers.

Finance departments use statistical modeling for predicting budgets and capital expenditures, when there are many unknown factors. A statistical model is a representation of what will probably happen. Unfortunately, economists can't predict everything so they use probability. Probability is simply the likelihood of something happening.

For example, let's say a company is taking out a loan for construction of a plant expecting it to be operational in 3 years. They are relying on the income from the plant to repay the loan. The finance team will run a statistical model that will include any possible delays, changes to building regulations, and catastrophic events. It's best to be prepared so it doesn't bankrupt the company if something goes wrong.

How do descriptive and inferential statistics differ?

- Descriptive statistics only attempt to describe data, while inferential statistics attempt to make predictions based on data.
- Inferential statistics only attempt to describe data, while descriptive statistics attempt to make predictions based on data.
- Descriptive statistics are more computationally sophisticated than inferential statistics.
- Inferential statistics are more computationally sophisticated than descriptive statistics.

Which two are examples of descriptive statistics?

- Mean and standard deviation.
- Median and correlation.
- Variance and regression analysis.
- Hypothesis testing and histograms.

What is statistical estimation?

- Methods for reducing errors in inferential statistics.
- Methods for reducing errors in descriptive statistics.
- Methods for rounding answers in statistical calculations.
- Methods to determine the best graph to represent statistical data.

Lesson6: Defining the Difference between Parameters & Statistics

Using Statistics to Understand Populations

Lenae is campaigning for town mayor. Today, she is doing a little market research to understand the concerns of the people in her town. In order to collect this information, Lenae will have to understand parameters and statistics when working with populations.

In this lesson, you will learn about the differences between parameters and statistics when working with data. But first, let's review populations and samples.

Understanding Populations & Samples

A population is all members of a specified group. For example, in Lenae's case she is collecting data on a very literal population, the population of her town. All members of this town would be included in the population. However, most of the time it isn't practical to get information from every member of a population.

When this happens, we have to find a different way of getting information that represents the population without actually asking the whole population. Lenae will probably not have the time and resources to collect information from the entire town; therefore, she will need another approach to getting the information she needs. She will need a sample to gather this information.

A sample is a part of a population used to describe the whole group. For example, Lenae could go to a local event or mail out surveys to the people in her town. As long as the information she collects is from the town and not a neighboring community, then this can count as Lenae's sample group.

Samples are used to help the data researcher, in this case Lenae, to understand the entire population. There are many different ways you can get a sample from your population. These include:

- Random sampling
- Simple random sampling
- Cluster sampling
- Stratified sampling
- Systematic sampling

Now that you understand population and samples, let's discuss how they relate to parameters and statistics.

Parameters & Statistics

Lenae has collected the information she needs from her sample. Now it is time to analyze this information using the concepts of parameters and statistics.

A parameter is the characteristics used to describe a population. For example, Lenae knows one parameter of her population is that they all live in the same town. This is a 100% known parameter of the population. If Lenae was able to see a 100% accurate census in the town, she could probably find parameters such as the number of people in a certain age range in the population. It's hard to have 100% proof for particular characteristics of a population. This is when we use statistics.

A statistic is the characteristics of a sample used to infer information about the population. For example, Lenae is using a sample to analyze data about her town's population. Lenae has found that 64% of the people she surveyed are concerned about the safety of the town's parks. Lenae can use this statistic to infer that approximately 64% of the town is also concerned about the safety of the town's parks.

Now that you understand parameters and statistics, let's practice by identifying the two.

What are all members of a specified group?

Population Sample Parameter Statistics Variance

What are the characteristics of a sample used to infer information about the population?

- Population
- Sample
- Statistic
- Parameter
- Variance

What do we call a part of a population used to describe the whole group?

- Population
- Sample
- Statistic
- Parameter
- Variance

Lesson7: What is Quantitative Data? - Definition & Examples

What Is Quantitative Data?

What's the difference between having seven apples and saying that they are delicious? Well, for one, we can count or measure the seven apples, but we can't put a number to how delicious they are. Those apples might be delicious to one person and be completely sour to another person.

What does this have to do with quantitative data? It has everything to do with quantitative data because it shows you what is considered quantitative data and what is not. Saying you have seven apples, because they can be represented numerically, is a piece of quantitative data. But, saying that they are delicious is not because you can't write that using numbers.

There are two types of data that quantitative data covers. They are data that can be counted and data that can be measured. Let's talk about what each data type looks like.

Data That Can Be Counted

Another way of saying that the data can be counted is to call it discrete data. Having the seven apples, for example, is discrete data because you can count seven apples. If you were to count the number of apples each tree produced in an apple orchard, that data is quantitative since the apples can be counted.

Other examples of discrete data include the number of girls in a math class, the number of boys who come to eat ice cream at three pm, and the number of kittens that a particular mom cat has. All of these are discrete and quantitative data because they can be represented by a mathematical number, and you can physically count them.

Data That Can Be Measured

Quantitative data is also data that you can measure. In math lingo, this is called continuous data. The weight of seven apples is continuous data because you can put the apples on a scale and weigh or measure them.

Other examples of continuous data include the height of your mom, the length of a football field, and the weight of a wolf. All of these are continuous data because you can measure them and represent them in a numerical manner.

Brown school has decided to review their ratio of girls to boys in the Second grade class. They decided to count the number of boys in Second grade. This is an example of what type of data?

- Qualitative and continuous
- Discrete and continuous
- Quantitative and continuous
- Measured and graphed
- Quantitative and discrete

Which of the following answers is an example of discrete data?

- 58 km
- 16 miles
- 120 pounds
- 8 puppies
- 1.12 inches

Which of the following answers is an example of continuous data?

3.14 cm 1 apple 3 pies 10 tests 4 zebras

Lesson8: Nominal, Ordinal, Interval & Ratio Measurements: Definition & Examples

Understanding Data

Chloe is working on a project for her agriculture class. She is growing and classifying different types of flowers. So far, she has collected data on the different types of flowers, the color of the flowers, the height of the flowers, and the number of petals on each flower. In order to analyze all of the data that Chloe has collected, she will need to understand the different types of data.

First, data is just information that is collected for analysis. Data is generally used to either disprove or prove a hypothesis. Data can be grouped into two different types of information: categorical and quantitative. In this lesson, we will be talking about two types of categorical data: nominal and ordinal. We will also be talking about two other types of measuring data: interval and ratio measurements.

Nominal and Ordinal Data

First Chloe collects information about the different types of flowers she's growing, and the colors of each. She counts 4 roses, 6 daisies, 3 sunflowers, and 4 lilies. The colors of the flowers are 3 red, 4 yellow, and 10 white. She decides to assign a code to each type and color of flower to keep track of this information in her records.

According to this table, a red rose would have the numerical code 15, with the first number representing the flower type and the second number representing the color. As far as the data is concerned, the number 15 does not need to be added, subtracted, multiplied, divided, or ordered in any way. It does not have a value of 15; it is simply a code to indicate a red rose. This is an example of nominal data, or categorical data that assigns numerical values as an attribute to an object, animal, person or any other non-number. This is similar to the numbers that are given to horses during a race. The numbers themselves have no value, they simply identify the horses.

Next, Chloe has entered her roses into a local gardening competition. She had one rose get 4th place, one got 7th place, and one got 2nd place. These rankings are an example of ordinal data, which is data that can be ordered and ranked, but not measured, such as levels of achievement, prizes, rankings, and placements. Similar to nominal data, ordinal data cannot be multiplied, divided, added, or subtracted. However, the difference between ordinal data and nominal data is that the data can be ordered. Because ordinal data has to do with placement and rankings, you can order the data in descending or ascending order. When remembering ordinal data, think ordinal=order.

Interval and Ratio Measurements

Chloe has entered another rose competition. This time, her flowers have to fit in certain categories. As she is filling out the paperwork, she notices some unusual questions. Take a look:

The second question represents a form of measurement called interval measurement, which is data that is grouped in evenly distributed values and measured based on the group to which that variable is attributed. In other words, interval data is measured in groups, rather than individually. For example, Chloe can measure the distance between each leaf and come up with 0.7. However, if the competition managers are trying to collect and group data in a more manageable way, then they will ask her to circle the option that is 0.5-1 inch. If you are collecting this information, you can really get a good visual representation by using a bar graph to display the data like this:

The next question asks about how many petals there are on Chloe's flower. You can take data like this and make a comparison. For example, what if the competition managers wanted to know how many flower petals there are in comparison to the leaves? They can learn this by using another form of measurement called a ratio, which is a mathematical comparison between two numbers. This can be represented with a colon; however, we can take a ratio and manipulate it to gather more information about our data. For example, if Chloe has 6 leaves and 12 petals, then we could write that ratio like this: 6:12. We can also easily see from this that there are twice as many petals as there are leaves. However, how would you find this comparative information if the ratio was more complicated? Let's take another look at Chloe's survey.

What is information that is collected for analysis?

- Data
- Nominal Data
- Ordinal Data
- Ratio
- Interval measurement

What is data that is grouped in evenly distributed values and measured based on the group to which the variable is attributed?

- Data
- Nominal Data
- Ordinal Data
- Interval measurement
- Ratio

What is a mathematical comparison between two numbers?

- Data
- Nominal Data
- Ordinal Data
- Interval measurement
- Ratio

Lesson9: The Purpose of Statistical Models

Purpose of Statistics

Benjamin is working on a project for his agriculture class. He found research that shows that under the right conditions, plants will grow a consistent amount every day. Benjamin wants to test this information and see if he can predict the height of his plants after 10 days.

Benjamin needs to understand statistical models and the purpose of statistics before he can properly analyze this information. In this lesson, we will discuss the purpose of statistics and how you can use statistical models to achieve this purpose.

First, let's discuss the purpose of statistics. The purpose of statistics is to describe and predict information. This can be divided into descriptive statistics and inferential statistics. Sometimes we collect data in an attempt to describe the characteristics of a population. For example, Benjamin can collect data on the colors of the flowers of certain types of plants. Over time, he may have enough information to say that the plant produces a white flower 56% of the time, a purple flower 34% of the time, and a blue flower 10% of the time. This is an example of how Benjamin used statistics to describe the plant.

Statistics is also used to predict information. Benjamin can use the same information that he collected to predict the color of flower that the plant would produce. If Benjamin has a plant that hasn't yet produced a flower, he can say that it is most likely to produce a white flower and least likely to produce a blue flower.

Now that you understand that the purpose of statistics is to describe and/or predict, let's discuss the role that statistical model plays in that purpose.

A statistical model is a combination of inferences based on collected data and population understanding used to predict information in an idealized form. This means that a statistical model can be an equation or a visual representation of information based on research that's already been collected over time. Notice that the definition mentions the words 'idealized form'. This means that there are always exceptions to the rules. For example, let's say that Benjamin waters his plants for 10 days with the correct amount of water under the correct conditions. However, what if someone accidentally knocks over one of the plants? Or what if an animal breaks into the greenhouse and starts feeding on the plant? These are extreme examples, but often unexpected conditions can interfere with collecting data.

Now let's talk about types of statistical models and how they are used.

Types of Statistical Models

Before you can understand the types of statistical models, you must first understand the reason these models exist. Statistical models exist because we are looking for a relationship between two, or sometimes more, variables. For example, in Benjamin's case there are two variables: the number of days the plants grow and the height of the plants. We know from earlier that the more days the plants grow, the taller they get. Of course, there is the matter of the condition of the plants, the amount of water, the amount of light, etc. These are all other variables that could affect the experiment. But for now, let's limit these two variables, just to keep things simple. The relationship between the height of the plants and the number of days the plants grow is known as a correlation, which is the relationship between two variables or sets of data. A correlation test is one type of statistical model.

Essentially, all statistical models exist to find inferences between different types of variables and because there are different types of variables, there are different types of statistical models. For example, let's say that Benjamin was collecting information about the different types of plants that grow in his region. He would be collecting data that would be grouped into categories, which is known as categorical data. In this experiment, Benjamin would have to use a different statistical model to analyze his data than the one he used to find a correlation between the height of the plants and the number of days they spent growing.

Some of the types of models, or statistical tests, include regression, analysis of variance, analysis of covariance, and chi-square. These are just a few examples of statistical models; there are many different ways we can analyze data depending on the variables. We will discuss many of these models in depth in future lessons.

Now let's talk more about the types of variables involved in different statistical models.

Types of Variables

Benjamin has been experimenting with his plants. He has added a different type of fertilizer, different amounts of water, and different amounts of humidity and sunlight to some of the plants. Now one of the plants has started to bloom only blue flowers, which is very rare. Unfortunately, Benjamin isn't sure which of the changes or combination of changes caused the plant to bloom blue flowers. To understand this phenomenon, Benjamin needs to understand two types of variables: response and explanatory.

A response variable is the observed variable, or variable in question. In Benjamin's case, the blue flowers would be the response variable. This is similar to a dependent variable, which is a condition or piece of data in an experiment that is controlled or influenced by an outside factor, most often the independent variable. However, sometimes data can be collected without doing an experiment and in these cases, there is still a response variable.

When analyzing data, we often ask, 'What is causing the response variable?' Benjamin has been asking the same question: 'What is causing the blue flowers?' To answer his question, you'll need to understand explanatory variables.

An explanatory variable is a variable or set of variables that can influence the response variable. In Benjamin's case, this refers to all of the things he did to his plants, such as watering, adding fertilizer, and changing humidity and sunlight. All of these factors could have influenced the blue flower's appearance.

Why do different types of statistical models exist?

- because there is no way to find relationships between all of the variables that exist
- to explain why a set of variables incur a response in the variable in question
- because there are many different types of variables and the models provide ways to analyze them
- to determine which variables should be controlled or changed

Variables are collected in the form of data. What are the two major forms of data?

- measurable and non-measurable
- numerical and ordinal
- explanatory and response
- categorical and quantitative

The _____ is a condition or piece of data in an experiment that is controlled or influenced by an outside factor, most often the independent variable.

- statistical model
- explanatory variable
- independent variable
- dependent variable

Lesson10: Random Selection & Random Allocation: Differences, Benefits & Examples

Random Selection

Aubree is conducting an experiment. She wants to find out if oranges consumed on a regular basis will help improve the chances of someone staying healthy during the winter months. Aubree will have to design a research experiment to find the answer to this question. As she is designing her research, she will need to understand random selection and random allocation.

In this lesson, you will learn about random selection and random allocation, how to use them and the differences between the two. First, let's discuss random selection.

Random selection is the method of selecting a sample from the population to participate in a study. Basically, random selection is the way Aubree will choose who will be a part of her study. Most studies use some sort of random sampling to select participants. There are different ways you can select participants for a study.

First, Aubree will need to decide on her population. A population is all members of a specified group. For example, let's say that Aubree wants to study the effects of oranges on college students. This means that her population will be all college students in the world. Of course, it's hard to conduct a study of this size. We can solve this problem using random selection. Aubree can get a sample of her population by selecting students at a local college for her study.

A sample is a part of a population used to describe the whole group. For example, Aubree can conduct her study with a random selection of students in certain classes at the college, or she can select every other student that is willing to be a part of the study.

There are many different ways you can get a sample from your population. These include random sampling, simple random sampling, cluster sampling, stratified sampling, and systematic sampling. Now that we have covered random selection, let's move on to random allocation.

Random Allocation

Remember, Aubree is studying the effects of orange consumption on college students. She wants to know if consuming oranges on a regular basis will help improve the chances of someone staying healthy during the winter months.

As such, Aubree has to compare two groups of students: those that consume oranges on a regular basis and those that consume oranges on an irregular basis or not at all. Aubree will have to use a control group and a treatment group because she wants to see the effects of orange consumption and will need to compare two groups.

The control group is the group that remains untreated throughout the duration of an experiment. For example, in Aubree's experiment the students that do not consume oranges on a regular basis would be the control group.

The treatment is the variable in an experiment that is used on an experimental group. For example, the oranges in Aubree's experiment would be the treatment, and the experimental group would be whomever is selected to receive the treatment. This is where random allocation comes in.

Random allocation is the method used to select members of a sample to receive the treatment in an experiment. Random allocation is the way Aubree will select her experimental group, or the group that will consume oranges in the experiment. She can select this group using similar methods that she used with random sampling. For example, she can write all of the participants' names on a piece of paper and randomly select half of the names from a hat. The names selected would be the experimental group.

What is the method used to choose members of a group to participate in an experiment?

- Random population
- Random selection
- Random sample
- Random allocation
- Random control

All members of a specified group are referred to as what?

- Selection
- Population
- Control
- Allocation
- Sample

What is a part of a population used to describe the whole group?

- Selection
- Population
- Sample
- Allocation
- Control

Lesson11: What is the Center in a Data Set? - Definition & Options

The center of a data set is represented by the typical value that represents the data set. Explore the definition of the center in a data set, and review options for finding it, including the mean and median.

Finding Your Center

Elizabeth is working on a science project. She is testing the effectiveness of bleach on certain types of stains. After trying the bleach on each stain five times, Elizabeth records how many times the bleach removes the stain completely. Check out this chart to see how the project turned out.

Data representing how many washes it takes to remove a stain with bleach data chart

Elizabeth has a data set of 4, 2, 5, 4, 1, 2 and 3. She wants to find a way to summarize the information. Elizabeth needs to find the center of data, a single number that summarizes the entire data set. You can find the center of data using either the mean or the median of the data set.

Using Mean

The mean is the sum of the numbers in a data set divided by the total number of values in the data set. The mean is also commonly known as the average. The mean can be used to get an overall idea or big picture of the data set. Mean is best used for a data set with numbers that are closer together.

Since all of Elizabeth's numbers are close together, she can use mean to find the center of her data set. Simply add all of the numbers together and divide by how many numbers there are in the data set: 4 + 2 + 5 + 4 + 1 + 2 + 3 = 21 / 7 = 3. Elizabeth's center of data for this data set is 3. She can summarize that, on average, bleach will effectively remove a stain 3 out of 5 times.

Using Median

Elizabeth is now experimenting with the effectiveness of plain soap on stains. After trying the soap on stains eight times, Elizabeth records how many times the soap removes the stain completely. Check out the chart she has made showing the effectiveness of soap on the stains.

From this chart you can see that our data set is 6, 7, 8, 8, 6, 0 and 7. You might notice that we have several numbers that are close together and one number that is a bit off. This number is referred to as the outlier. An outlier is a value that is much larger or smaller than the other values in a data set, or a value that lies outside the given data set.

What is a single number that summarizes an entire data set?

- Center of data
- Mean
- Median
- Outlier
- Range

What is the sum of the numbers in a data set divided by the total number of values in the data set?

- Center of data
- Mean
- Median
- Outlier
- Range

For the following data set, select the best method for summarizing the data.

3, 3, 4, 5, 2, 9, 5, 2, 3, 4

- ○ Mean
- ○ Median
- ○ Outlier
- ○ Mode
- ○ Range

Lesson12: Calculating the Mean, Median, Mode & Range: Practice Problems

The mean, median, mode, & range are fundamental statistical calculations necessary to evaluate and comprehend the significance of a set of numbers. Learn how to calculate mean, medium, mode, & range through real-world practice problems.

Mean, Median, Mode & Range Definitions

The mean is the arithmetic average of a data set. This is found by adding the numbers in a data set and dividing by how many numbers there are. The median is the middle number in a data set when the numbers are listed in either ascending or descending order. The mode is the value that occurs the most often in a data set, and the range is the difference between the highest and lowest values in a data set.

we will be looking at practice problems to help you find the mean, median, mode, and range of a data set. Since these are practice problems, feel free to pause the video during each example and try to work out the problems on your own. Then, play the video to check your answers.

Gasoline Prices

According to the U.S. Bureau of Labor Statistics, the gas prices for each month of the year in 2000 were as follows, rounded to the nearest hundredth of a decimal:

1.30, 1.37, 1.54, 1.51, 1.50, 1.62, 1.59, 1.51, 1.58, 1.56, 1.56, 1.49.

Let's start with the mean. Pause the video here to see if you can find the mean of this data set. The mean of a data set tells us on average how much gas cost in the year 2000. We can find the mean by adding all of the numbers up and dividing by 12, which is the number of months in the year and how many numbers we have in this data set.

1.30 + 1.37 + 1.54 + 1.51 + 1.50 + 1.62 + 1.59 + 1.51 + 1.58 + 1.56 + 1.56 + 1.49 = 18.13 / 12 = 1.51

1.51 is the mean for this data set. This number tells us on average the price of gas for the entire year. You will notice that 1.51 appears in the data set. Sometimes you will have an average that does not appear in the data set, but will still show you the big picture of the numbers given.

Okay, let's move on to median. Pause the video here to see if you can find the median of this data set. The median of a data set tells us what number falls directly in the middle. This is useful if you have one or two numbers that are greatly larger or smaller than the rest of the numbers in the data set. If the numbers are all pretty close together, then the mean and the median will be very close to the same number.

First, arrange the numbers in either ascending or descending order.

1.30, 1.37, 1.49, 1.50, 1.51, 1.51, 1.54, 1.56, 1.56, 1.58, 1.59, 1.62

Now, eliminate each number until you are down to the middle.

The following data set represents the average number of children per family for 10 countries in 2014. Calculate the mean number of children per family. 5.8, 6.1, 1.9, 1.4, 2.6, 2.8, 1.3, 4.4, 4.4, 1.7

- 4.4
- 4.8
- 5
- 3.24

Which of the following describes the best approach to finding the median of an even set of numbers?

- Select whichever of the numbers appears to fall in the middle of the data set
- Place numbers in ascending or descending order, calculate the average of the middle terms
- Calculate the average of the two highest and two lowest numbers in the set
- Determine which of the numbers in the set is found most frequently

The value that occurs the most frequently in a data set is known as the _____.

- Mean
- Median
- Mode
- Range

Lesson13: Simple Random Samples: Definition & Examples

Simple Random Sampling

Adrian is gathering information for a trip he plans to take. He is thinking about moving permanently to a new town. However, he wants to get an idea of how the people in the town feel about the safety of the town. Unfortunately, Adrian does not have the resources to ask every person in the town how they feel about the safety. How should Adrian go about collecting this data?

In this lesson, you will learn about how to use and recognize simple random sampling in statistics. First, let's discuss the meaning of simple random sampling by defining a few key terms.

When you are doing an experiment, you want to gather information about a population. A population is all members of a specific group. For example, the population of Adrian's research will be, quite literally, the population of the town. Sometimes a population is not that geographically contained. If Adrian wanted to know the typical income of a person over 30, then anyone with a job over 30 in the entire world would be in his population. Since Adrian does not have the resources to ask everyone in the town how they feel about safety, he will have to use a sample of the population.

A sample is a part of the population used to describe the whole group. Adrian will want to make sure all demographics are represented in his sample. He doesn't just want the opinion of the teenagers in his town or just the men over 50. He wants to get all demographics equally represented in his sample. To do this, Adrian will need to use random sampling, which is a method of choosing an equally distributed subset from a larger population. This takes us to simple random sampling or SRS, which is a type of random sampling where the variables have an equal and unsystematic chance of selection.

For example, if you were to toss a handful of 6-sided dice on a table, you would have an equal and unsystematic likelihood of getting a one, two, three, four, five, or six. When I say unsystematic, I mean you aren't throwing the dice and then only choosing the number off of every other dice on the table. When you use a system to randomize the selection instead of just taking the random selections as they fall, then you are not using simple random selection.

In Adrian's experiment, Adrian can use a phone book with all of the names of the people in the town as his population group. He can then put each name on a piece of paper and put the papers into a bag. Adrian can blindly select a certain number of names from the bag as part of his simple random sampling.

Now that you understand the meaning of simple random sampling, let's discuss how to use simple random sampling.

Using Simple Random Sampling

Simple random sampling is meant to be a balanced representation of the demographics of the population. When you are discussing a population of people, that means all of the demographics: age, race, religion, ethnicity, socioeconomic status, education level, etc. that are all currently present in the given population**.**

What is the name for all the members of a specified group?

Sample   Subset   Simple random sampling   Population

What is a part of a population used to describe the whole group?

- Sample
- Subset
- Simple random sampling
- Event

What is a type of random sampling where the variables have an equal, and unsystematic, chance of selection?

- Subset
- Simple random sampling
- Population
- Event

Lesson14: Stratified Random Samples: Definition, Characteristics & Examples

Stratified Random Sampling

Jackie is the president of the party planning committee of her school. Right now, the party planning committee is planning a winter formal. She is researching the different DJs that are available to work at the winter formal. Each DJ has a different percentage of music that he or she is able to play.

Jackie needs to find the right DJ for the winter formal so that the students of her school are happy with the music choices. Jackie can't ask every member of the school his or her music preferences. How can she figure out which DJ to hire?

In this lesson, you will learn about stratified random sampling, what it is, and how to use it.

What Is Stratified Random Sampling?

Stratified random sampling is a random sampling method where you divide members of a population into 'strata,' or homogeneous subgroups.

This is the percentage of each music genre that each DJ will play. DJ Thunder Cat will only play 60% rock, 20% pop, 15% hip hop, and 5% country. DJ Xtreme Mix will only play 75% hip hop, 20% pop, 5% rock, and no country. DJ Midnight will only play 50% pop, 35% country, 10% hip hop, and 5% rock. Each of these genres is an example of a strata, or a homogeneous subgroup. The group or population is music, while the strata is each kind of music.

Stratified random sampling works the same way. Jackie doesn't have the ability to question every student in the school. However, she can use stratified random sampling to get an understanding of the music tastes of the students in the school. Jackie can divide the student body into different strata, or subgroups, and then ask each of these subgroups what types of music they prefer.

Stratified random sampling is different from other types of sampling because you are separating the population into groups first. You must be very familiar with the demographics of your population if you intend on using stratified random sampling. Let's discuss how to use stratified sampling and the ways you can use this sampling in an experiment.

Using Stratified Random Sampling

Jackie decides that a stratified random sample may be the best way to collect her information. When dividing the school into stratified random samples, she must keep two things in mind:

- Stratified random samples cannot have crossover.
- Stratified random samples must include all members of a population.

Stratified random samples cannot have crossover. In other words, each of the strata must be mutually exclusive. In Jackie's case, she must choose some type of group in which each student is a part of one group but not more than one. Jackie could use something like age or school classification, such as freshman, sophomore, junior, and senior.

Which type of random sampling divides members of a population into homogeneous subgroups?

- ○ Simple random sampling
- ○ Cluster random sampling
- ○ Stratified random sampling
- ○ Systematic random sampling

Which of the following conditions must be met for a sample to be a stratified random sample?

- ○ The strata must occur naturally.
- ○ Each group must be equally separated.
- ○ The strata must have a crossover.
- ○ It must include all members of a population.

When selecting strata groups, it is recommended you have:

- More than 7 strata.
- No more than 4 to 6 strata.
- Less than 3 strata.
- Between 6 and 8 strata.

Lesson15: Systematic Random Samples: Definition, Formula & Advantages

Systematic Random Sampling

Lucas is a new manager at the local movie theater. The owner of the movie theater wants to find out how the customers feel about the new renovations they've done at the theater. Lucas can't ask every customer that comes in how they feel, especially when the movie theater gets busiest on Friday and Saturday nights. In this lesson, you will learn about systematic random sampling and how to use it when collecting data.
What Is Systematic Random Sampling?
Systematic random sampling is the random sampling method that requires selecting samples based on a system of intervals in a numbered population. For example, Lucas can give a survey to every fourth customer that comes in to the movie theater. The fact that Lucas is giving the survey to every fourth customer is what makes the sampling systematic because there is an interval system. Likewise, this is a random sample because Lucas cannot control what type of customer comes through the movie theater.
Additionally, remember that systematic random sampling must still ensure that all outcomes are given equal chance of getting selected in the sample. Therefore, Lucas cannot only select every fourth customer that comes through the door during the evenings or on the weekends. He must select every fourth customer every time the theater is open.
Lucas must also ensure that by choosing every fourth customer he does not include any sort of pattern in the selection. We will talk about this more when discussing the pros and cons of systematic random sampling.
Now that you understand the definition of systematic random sampling, you can learn when and how to use systematic random sampling.

How to Use Systematic Random Sampling

Let's discuss when and how to use systematic random sampling. Lucas's boss wants to send his employees to a weeklong training session that is out of town. Due to limited funding, Lucas's boss, Alex, cannot send all of his employees; he must choose a group to go to the training. Alex owns 12 movie theaters and employs 200 people. He has 12 managers out of the 200 employees. Alex can use systematic random sampling to select the group of employees that will attend the training.
Alex can follow these steps to create a group from systematic random sampling:

- Create a list of employees
- Select a beginning number
- Select an interval
- Gather a list of employees based on the interval number

First, Alex will need to create a list of his employees. Then, he will need to randomly decide which number to start his selection process. For this, Alex uses a random number generator to select which employee he will begin with. The random number generator produces the number 34. Now Alex needs to create an interval. First, he needs to decide how many employees he wants to send to the training. After reviewing his budget, Alex decides he can afford to send 20 employees to the training. To find the interval he needs, Alex can divide the total number of employees he has (the population size) by the number of employees he wants to send to the training (the sample size), like this:
200 / 20 = 10
This would make his interval 10, meaning that every 10th person after the 34th person would be selected until he had a total of 20 people.
Therefore, the following people would be selected:
34, 44, 54, 64, 74, 84, 94, 104, 114, 124, 134, 144, 154, 164, 174, 184, 194, 14, 24, 35

The numbers 14, 24, and 35 are included here because in order to select 20 people, Alex will have to continue selecting every 10th person, even if that means starting back at the beginning of the list. The number 35 is included because the 34th person has already been selected at this point.
What if the interval number happened to be a fraction? What if Alex decided he wanted to select 23 people to go to the training?
He would still use this method:
200 / 23 = 8.69565….
Obviously, you cannot select .69565….. of a person. To compensate for this, Alex will need to pick every 8th person then every 9th person and continue to rotate this pattern until he has 23 people.
Alex would end up with the following people:
34, 42, 51, 59, 68, 76, 85, 93, 102, 110, 119, 127, 136, 144, 153, 161, 170, 178, 187, 195, 4, 12, 21

What is the method that requires selecting samples based on a system of intervals in a numbered population?

- Simple random sampling
- Cluster sampling
- Stratified random sampling
- Systematic random sampling
- Law of large numbers

When doing systematic sampling, you must remember that all outcomes must be given _____ chance of being selected.

- equal
- twice the
- less than half of the
- more than double

To use systematic random sampling, you must first create a list of your population, and then:

- Create a list of employees
- Select an interval
- Select a beginning number
- Gather a list of your sample
- Put all names in a hat

Lesson16: Cluster Random Samples: Definition, Selection & Examples

Cluster random samples refer to the participants including all members of a population, which are selected through cluster sampling. Learn about cluster sampling and cluster random sampling, explore the definition, selection, and examples of cluster random samples, and understand how and when to use cluster sampling in research.

Cluster Random Sampling

Donna is running for student class president. She has some key issues in her campaign, such as bullying in the school, the prom theme, and fixing the water fountains in the school. Her campaign manager, Lulu, wants to collect information about how the students feel about the different issues in the campaign. She specifically wants to know about the different interest groups in the school.

In this lesson, you will learn about cluster sampling, including what it is, how to use it, and some of the advantages and disadvantages of using this sampling method.

What Is Cluster Sampling?

Cluster sampling is the sampling method where different groups within a population are used as a sample. This is different from stratified sampling in that you will use the entire group, or cluster, as a sample rather than a randomly selected member of all groups.

For example, Lulu wants to conduct some marketing research for Donna's campaign. She specifically wants information from the different interest groups of the school. In the school, she has found that 30% of the students are involved in athletics, 25% of the students are involved in an academic club, 20% of the students are involved in an art or theater club, and 25% are involved in a music club. None of the students are involved in more than one club and all of the students are involved in a club. Lulu knows that Donna is a member of an athletics club and the athletics students support Donna's campaign.

How to Use Cluster Sampling

When conducting research and using cluster sampling, you must keep a few things in mind:

- Cluster random samples cannot have crossover.
- Cluster random samples must include all members of a population.

Let's start with the rule that cluster random samples cannot have crossover. In other words, each of the clusters must be mutually exclusive. In this case, Lulu cannot divide the school population into interest groups where students are involved in more than one club or are a member of more than one cluster. Lulu has decided to conduct research with only the students that are in the arts, theater, and music clubs. Since Lulu already knows that no students are involved in more than one of each of the club categories, she knows her cluster sample does not have crossover.

We also know that cluster samples must include all members of a population, meaning that all of the students in the school must be a member of one of the four clubs. If there are students that are not a member of one of the clubs, then the cluster sample does not work.

If Lulu doesn't have time to give a survey to all the students in the club, she can use two-way cluster sampling, which is a sampling method that involves separating the population into clusters, then selecting random samples from each of the clusters. Lulu can use simple random sampling to select members of each cluster, or club, to give a survey.

You may be wondering how two-way cluster sampling is different from stratified sampling.

Stratified random samples must have an equal selection from each group that is proportionate to the population. Cluster sampling selection does not have to be equal; however, the clusters should be as close to the same size as possible.

Stratified random samples should not divide the population into more than six groups and are usually organized by demographic. Cluster sampling can be many groups and can be based on anything, including interests, hobbies, political views, geographical location, etc.

Because Lulu is using cluster sampling, she can select as many or as few people from each group as she wants.

When to Use Cluster Sampling

You should use cluster sampling when:

- ✓ The entire population is unclear or unknown
- ✓ The sample clusters are geographically convenient

_____ is a sampling method where different groups within a population are used as a sample.

- ○ Random sampling
- ○ Simple random sampling
- ○ Cluster sampling
- ○ Systematic sampling

Which is a sampling method that involves separating the population into clusters, and then selecting random samples from each of the clusters?

- ○ Two-way cluster sampling.
- ○ Simple random sampling.
- ○ Stratified sampling.
- ○ Systematic sampling.

How do cluster sampling and stratified random sampling differ?

- ○ In cluster sampling, the sample must match the proportion of the groups in the population, but in stratified random sampling this is not necessary.
- ○ Stratified random sampling requires selection from each group proportional to their share of the population, but cluster sampling does not.
- ○ Stratified random sampling allows groups to mix together, but cluster sampling does not.
- ○ Cluster sampling allows groups to mix together, but stratified random sampling does not.

Lesson17: How to Interpret Correlations in Research Results

Correlations acknowledge some relationship between two variables. In this lesson, learn how to graphically represent and statistically interpret correlational data.

Introduction to Correlation

Imagine you're reading the newspaper, and you see an article that says that a study was done on whether reading books about vampires makes children want to turn into vampires themselves. The article says that there's a correlation between reading vampire books and desire to be a vampire, and that the correlation is -1.5. The reporter concludes that vampire books should be banned, because they are causing children to turn into vampires! What do you think of this reporter's conclusion? If you understand the theory and statistics behind correlational studies, you'll know that this reporter needs to go back to school to learn about how correlations really work. That's the topic of this lesson.

The Purpose of Correlations

A correlation is a simple statistic that explains whether there's a relationship or association between any two variables. Correlations are probably the most common statistic used in the field of psychology, so it's important to understand how they work.

Let's start with how we might do a basic correlational study before we get to the meaning behind the actual numbers from the statistics. In a correlational study, researchers pick two variables they think might be associated with each other. For this lesson, let's think about a student who wants to go from high school to college. The admissions office at each college will want to know what that student's high school grades were like, because they believe that high school grades can predict college grades. In other words, they believe that high school grades are associated with college grades. Why would they make this conclusion? They could make a graph showing all of the students they have accepted in the past, and this graph could show both variables.

On the y-axis, we could plot each student's high school grade point average. On the x-axis we could plot that same student's overall college grade point average. We put a dot on the graph showing where these two variables intersect. We keep going until we have a dot for every student. Each dot represents one person. This type of graph is called a scatter plot. A scatter plot shows a dot for each person of interest, where each dot represents one person's scores on the two variables of interest. Here, each dot shows one person's high school grades and college grades. You can remember the name 'scatter plot' because after we plot, or mark, the graph with each person, it looks like a bunch of dots have been scattered all over it.

After all of the dots have been plotted, we can look for the general pattern, or trend, that is a representation of most people. In other words, if you were to draw a single, straight line on this graph, where would you draw the line? It would probably be right here. This line is a quick summary of the general pattern of dots we see on the scatter plot.

Now, how does this graph relate to correlations? A correlation is simply a number that is assigned to represent this scatter plot and this line. The equation for how to calculate the number you end up with is complicated, and you don't need to know it until you take a statistics class in college. For now, all you need to know is that the equation gives you a number that's like a code, and you can interpret this number, or code, to know what the graph looks like that resulted in this number. How to read the number is what we'll cover next in this lesson.

Interpreting the Statistic: Direction

The resulting statistic you get from a correlation equation is called a correlation coefficient. There will always be two parts to a correlation coefficient. The first part is the sign, or direction, meaning whether the coefficient is a positive number or a negative number. That sign is the first part of the code you need to know.

The second part of the correlation coefficient will be a number. The number will always be between zero and one. That means that the correlation coefficient will always be somewhere between negative one and positive one, but it could be anywhere in between. Let's go over each part of the correlation coefficient and discuss what that part means.

We'll start with the sign, or direction. Unless your coefficient is exactly zero, you'll have a number that's either positive or negative. The sign of positive or negative is simply a code that indicates how the line appeared on the scatter plot. Remember our example before? We plotted high school grades and college grades, and we ended up with a line that looked like this. Notice that the line goes from the bottom left corner of the graph to the upper right corner of the graph. That means that as one of our variables went up in value, so did the other variable. In other words, if a student had a high GPA in high school, he or she is likely to also have a high GPA in college. As one variable gets higher, the other variable also gets higher.

Whenever we graph two variables that move in the same direction, the line we draw will generally go from the bottom left to the upper right of the graph. We call this a positive correlation. A positive correlation means that both variables move in the same direction - as one goes up, the other goes up, or vice versa. We call this a positive correlation because when we do the equation to come up with our correlation coefficient, the result is going to be a positive number. It can be anywhere from +0.01 all the way up to +1.00.

The only other option for a correlation will be that it's got a negative sign in front of the coefficient. You won't be surprised to learn that we call this a negative correlation. A negative correlation means that the two variables move in the opposite direction from each other - as one goes up, the other goes down. What would that look like on a scatter plot? Where would we draw the line?

Imagine that we plotted two variables we think move in opposite directions. Let's pick college grade point average and the number of hours a student spends partying instead of studying. We might imagine that the more you party at college, the lower your grades might be. So the two variables move in opposite directions; as one goes up (that would be the partying), the other goes down (that would be the GPA).

If we made a scatter plot of several students who were already in college, we could put number of hours partying on the y-axis, and keep college grades on the x-axis. Now, the scatter plot might look like this. Where would we draw the line representing the general pattern? Here, it goes from the upper left to the bottom right. This will always be where the line goes for a negative correlation.

So, now you know what the positive or negative sign means on a correlation coefficient. It tells you whether the two variables move in the same direction or opposite directions. It also tells you the general direction of the line you would see on a scatter plot showing all of the people used to calculate the correlation**.**

Interpreting the Statistic: the Number

The second part of any correlation coefficient is the number that appears after the sign. Remember that a correlation coefficient will always range from zero to one. So, you might see a correlation of -0.85, or +0.14, or +0.98. You already know what the negative or positive sign means. What does the number behind it mean?

The number you see in a correlation tells you the strength of the association between the two variables of interest. In other words, are these two variables very strongly related, or not? Let's go through some examples to make this clear.

If the number you get is a perfect zero, that means that the two variables are not related to each other at all. You can imagine some variables that simply have nothing to do with each other. For example, college grades are not correlated with how tall you are, or what color your eyes are or the average number of pizza slices you eat in any given week. These variables are not correlated, meaning the correlation coefficient you would get would be zero.

As a correlation moves from zero to one, it means that the relationship becomes stronger and stronger. A low correlation means that the two variables are a little bit related to each other, but not much. A high correlation, meaning one that's closer to the value of one, means that the two variables are very strongly related to each other. For example, high school grades and college grades are generally related to each other.

Which one of the following coefficients indicates the strongest relationship between two variables while remaining within the possible range of correct calculations?

- +4.50
- 0.00
- -0.83
- -2.78

Which one of the following coefficients indicates that the two variables of interest move in opposite directions while remaining within the possible range of correct calculations?

- +4.50
- +0.32
- -0.83
- -2.78

José wants to calculate a correlation; which of the following graphs should he use?

- Scatter plot
- Line graph
- Histogram
- Pie chart

Lesson18: Inferential Statistics

Imagine a teacher is interested in studying several aspects of her class, such as the personality of her students, whether boys are different from girls or whether different teaching styles lead to different results in her students. In order to understand any of these aspects of the children in her class, the teacher must understand some basic statistics so that she can quantify her understanding, or, in other words, put it into numerical form.

In another lesson for Educational Psychology, you can learn about ideas such as the mean, median and mode to describe people, or how a standard bell curve works. This lesson focuses on a slightly different type of statistic, called inferential statistics.

Inferential statistics are ways of analyzing data that allow the researcher to make conclusions about whether a hypothesis was supported by the results. You can remember the term inferential because it comes from the word 'inference,' meaning 'to draw a conclusion from clues in the environment.' How do inferential statistics work?

Two Types of Inferential Statistics

To make things easier, let's think about an example from a classroom. Imagine a teacher suspects that the boys in her class are more extroverted - or more talkative, energetic and social - than the girls in her class. The teacher's guess about the difference between boys and girls is what we call a hypothesis. In psychology, a hypothesis is an educated guess about a trend, group difference or association believed to exist.

Her hypothesis is that boys are more extroverted than girls. How would she test this hypothesis? The teacher would probably do something to measure extroversion, such as give the students a personality survey to complete, or simply observe them and keep track of extroverted behaviors. Either way, she can measure the level of extroversion in every boy and every girl. Then she can compare the scores across the two groups.

The first type of inferential statistic we need to discuss is called a t-test. A t-test is used to compare the average scores between two different groups in a study to see if the groups are different from each other.

In our example, the teacher would use a t-test to compare the average level of extroversion in the group of boys versus the group of girls. T-tests are very common in psychology because they can be used to compare any two groups in an experiment. If you do an experiment where you ask some people to eat healthy food and some people to eat unhealthy food, such as candy, you could then test them on some variable, such as whether they get a stomachache. A t-test would again be used here, because you are comparing the two different groups. You can use t-tests to compare two groups that occur naturally, such as boys versus girls, or you can compare two groups that you have created in an experiment.

So a t-test compares two groups. You can remember the term t-test by pretending that the letter 't' stands for the word 'two,' meaning the two groups you are comparing. But what if you want to compare more than two groups? Imagine that the teacher thinks that as children age, they become more extroverted. Now she might give personality tests to children in each grade level in the school, such as all the way from kindergarten to sixth grade. How could she compare all of these different groups, now that we have more than two?

The second basic type of inferential statistics is called an analysis of variance. Researchers usually use the nickname ANOVA for this test. An analysis of variance is a test that compares the average scores between three or more different groups in a study to see if the groups are different from each other.

In other words, an ANOVA is exactly the same as a t-test, but it can analyze multiple groups at once. The difference is simply how the equation works to analyze the groups, which you can learn more about in a statistics class if you're interested. For now, all you need to know is that the ANOVA compares multiple groups, while a t-test can only compare two groups.

Let's go through one more example of when you might use each test. Imagine a teacher believes that different teaching styles result in different scores when children take a test over the material. He might try lecturing for one group of students, versus worksheets with a second group of students. He then gives everyone the same test, and wants to compare the results.
If he only had these two groups, he would use a t-test to compare the scores. However, now let's say that he wanted to add a third teaching style, which was having the students learn the material on their own and then teach it to each other. If he now wants to compare all three teaching styles to each other, he would use an analysis of variance, or an ANOVA test.

A teacher wants to compare the reading comprehension skills of children with learning difficulties and children without learning difficulties. Which statistical test should she use to make this comparison?

- Regression
- p-test
- Analysis of Variance (ANOVA)
- t-test

A teacher wants to compare children of high, medium, and low family incomes on their satisfaction with the school lunch program. Which statistical test should be used to make this comparison?

- Regression
- p-test
- Analysis of Variance (ANOVA)
- t-test

A researcher asks a group of students to complete a survey that measures maturity. She finds that the average score for girls is higher than the average score for boys, with a p-value of 0.03. What should she conclude from this p-value?

- Three percent of the girls are more mature than the boys, on average.
- There's a three percent chance that the scores in the study are due to random chance.
- There's a ninety-three percent chance that the scores in the study are due to random chance.
- Ninety-three percent of the girls are more mature than the boys, on average.

Lesson19: One-Tailed Vs. Two-Tailed Tests: Differences & Examples

Significance

Some of my friends are currently applying for their doctoral internship. This process involves writing several essays about yourself that must be shorter than 500 words, tabulating several hundred hours of therapy with clients, figuring out what assessment procedures they have completed, and no matter how hard and how fast you work you always feel like you're a week behind.

This has, understandably, caused the level of stress in these doctoral students to rise. The question I have is, 'Are their stress levels significantly different now than as compared to their stress levels last year?'

To fully test this will require me to go back in time with my time machine and have them take several stress tests. Then I jump back to the present and give the same people more stress tests. With all of this data, I am looking to see if there is a significant difference between last year and this year.

Statistically significant means the difference in the results did not occur by random chance. This is almost always represented by a lower case p, which stands for probability. Another term you may also hear is 'alpha,' and it may be represented by the alpha symbol (that little one that looks like a fish).

In psychological research, random chance typically means these results would only occur by chance less than 1 in 20 times, or .05. If you've done any reading of psychological research articles, you may have seen $p < .05$, which means the probability of these results being a fluke is less than 1 in 20 times. The ways something could be a fluke are by data collection error or just by the numbers being too similar. The significance tells you that the scores are so far apart that even with some variations they are telling us something.

This significance is taken from a normal distribution. It says that the numbers you are comparing are so different that something is going on, whether the data set is way higher or way lower. What the tailed test does is manipulate how we interpret the probability.

Two-Tailed Test

A two-tailed test, also known as a non directional hypothesis, is the standard test of significance to determine if there is a relationship between variables in either direction. Two-tailed tests do this by dividing the .05 in two and putting half on each side of the bell curve.

Why would someone do this? To determine if there is an interaction.

Remember those stress levels I went back in time to get? Let's say I do a simple test called a t-test, which compares two averages. I have the average stress levels from last year compared to the average stress levels of this year.

After some probability calculations, I learn that there is no significant difference between last year's and this year's stress levels. This tells me that this year's stress levels are neither higher nor lower than last years.

What if I jump in my time machine again and go back 15 years. The average age of my subjects is currently 26, so I will talk to them when they are about 11. Wow. I suddenly am starting to feel really old. Anyway, I collect their stress levels and then jump back to the present and do another t-test, and I find out that their stress levels are lower now than when they were younger. The beauty of the two-tailed test is that when you run your numbers, the math will tell you if it's significantly higher or lower.

One-Tailed Test

A one-tailed test, also known as a directional hypothesis, is a test of significance to determine if there is a relationship between the variables in one direction. A one-tailed test is useful if you have a good idea, usually based on your knowledge of the subject, that there is going to be a directional difference between the variables.

Directional difference is my fancy way of saying that you know one of the set of scores will be higher or lower than the other. Looking back at our original example of the stressed out graduate students, I think we can make a good guess on whether adding additional stressors will cause a person to be more or less stressed.

A civil engineer is trying to prove that his new method of laying blocks is more effective than the traditional method. The engineer will conduct a:

- One-tailed test
- Two-tailed test
- Random test
- Point estimate

Why is a two-tailed test sometimes referred to as the non-directional hypothesis?

- Because this test only verifies if there is a statistically significant difference between groups or not.
- Because researchers are unsure of what to do with their data.
- Because this test determines if there is a statistically significant difference between groups in either direction.
- Because two-tailed tests are conducted when the results are both positive and negative.

In statistical equations, probability is usually represented by _____.

- a lower case p
- the word probability
- the abbreviation py
- the alpha symbol

Lesson20: What is a Chi-Square Test? - Definition & Example

Definitions Involved in Chi-Square Test

I've been reading a lot about undercover officers lately, and it made me start wondering how many police officers work undercover versus how many apply to be in the program. I mean, not everyone who applies can work undercover because they may not fit a need or their scores on psychological tests just don't measure up.

If the numbers were really close between those who applied and those who got in, we would need to know if there is a statistically significant difference. Statistically significant means the difference in the results did not occur by random chance. This is almost always represented by a lowercase p, which stands for probability.

If you have read any psychological research articles, you may have seen $p < .05$, which means that the probability of these results being a fluke is less than 1 in 20 times. This has been the agreed upon level of chance that results can be wrong for quite a while. We'll get into how you figure it out for a chi-square in just a moment.

What we need is a specific statistical test to allow us to take categorical data, like those who did make it into the undercover program and those who did not. What we need is a chi-square, which is a statistical test used to compare expected data with what we collected.

What a chi-square will tell us is if there is a large difference between collected numbers and expected numbers. If the difference is large, it tells us that there may be something causing a significant change. A significantly large difference will allow us to reject the null hypothesis, which is defined as the prediction that there is no interaction between variables. Basically, if there is a big enough difference between the scores, then we can say something significant happened. If the scores are too close, then we have to conclude that they are basically the same.

Statistics

The actual formula for running a chi-square is actually very simple:

(o-e)^2 / e

You take your observed data (o), and subtract what you expected (e). You square the results, and then divide by the expected data in all the categories.

To use the number we find, we refer to the degrees of freedom, usually labeled as df for short, and is defined for the chi-square as the number of categories minus 1. Due to the nature of the chi-square test, you will always use the number of categories minus 1 to find the degrees of freedom. The reason this is done is because there is an assumption that your sample data is biased, and this helps shift your scores to allow for error.

You will then locate a chi-square distribution table, which is found in almost every statistical textbook printed. Using your degrees of freedom, you will locate the p-value you're interested in using the process below; typically the p-value is .05. If you can, see if your number is greater than .01, which means that your results could only happen by chance 1 in 100 times. Because of copyright restriction issues, we won't be able to provide a full image of the chi-square distribution table, but below is basically what they look like and how you find the digit you're lo

Which of the following types of data are required in order to run a chi-square?

- Categorical data
- Interval data
- Ratio data
- Experienced data

You are observing a psychologist run a chi-square. You see the psychologist take his number of categories and subtract 1. Which of the following is the psychologist calculating?

- Degrees of freedom
- The chi of chi-square
- How to reduce the chi-square to useability
- How to make a chi-cube

Between which of the following data does a chi-square measure the difference?

- Collected and expected
- Assumed and expected
- Recorded and implied
- Implied and population

Lesson21: What Are t-Tests? - Assessing Statistical Differences between Groups

Definitions

I've often wondered if having two classes combined is a good thing. No. Sorry. Not some kind of anti-Marxist/Marxist idea - I meant school classes. When you combine two classes you have to teach to the lowest level in the room. For example, if you had a class for high schoolers where you combined seniors and freshman, than you have to teach it at the freshman level. This will bore the seniors out of their minds because they are well above that level.

However, if you taught at the senior level, then you would confuse the freshman because they haven't had as much experience. The same issues happen at college and graduate school, probably even more. So the question for those designing the classes and those who are evaluating whether or not to combine two different levels of students is, 'Are the experience and knowledge levels of the classes so different that one group will not get any use out of the class?'

The t-Test is a statistical test to determine if there is a measured difference between two groups. The t-Test always has a lower case letter 't,' which seems to have been chosen arbitrarily. The t-Test can also be referred to as the Student's t-Test because it was printed under the pseudonym 'Student' by William Gossett to avoid issues with an employer at a brewery. While you don't need to know this to do the statistics, sometimes it's interesting to remember that the purpose of statistics is to solve real world problems. For instance, is beer A, on average, better than beer B?

Null & Alt

T-Tests use both a null and an alternative hypothesis. A null hypothesis is typically the standard assumption and is defined as the prediction that there is no interaction between the variables. This is opposed by the alternative hypothesis, also known as the research hypothesis, defined as the prediction that there is a measured interaction between the variables.

Because t-Tests are specifically looking to see if there is a difference between the two groups of scores, the null hypothesis is that there is no difference between the groups. So, if we have a 4th and 5th grade combination class and if we failed to reject the null hypothesis, then that would mean there is no significant difference between the 4th and the 5th graders. The t-Test can reject the null hypothesis, which would mean that we would look to the alternative hypothesis. With the alternative hypothesis, we would be able to say that there is a significant difference between the groups.

T-Test Stats

Let's illustrate how a t-Test works by looking at two groups of graduate students who have been put into the same class. We gave everyone a test to determine their level of knowledge on the subject. First, we will list out the data scores that we collected:

Group 1	Group 2
5	2
6	3
4	4
5	3
2	7

Our total number of participants is 10 and we will represent this by the letter n, so $n = 10$. We need to figure out how much variance there is within all of the scores, so we will calculate the mean of each group. The mean of Group 1 is 4.4, and the mean of Group 2 is 3.8. To calculate the variance, we need to calculate their standard deviation, which is defined as a number corresponding to a bell curve describing how spread out the data is.

The process of calculating standard deviation involves finding how much each score differs from the average, squaring them, adding them up, and then square rooting them. Many calculators and online cheats will let you do it quickly, and few statisticians will calculate it by hand. If you're interested in how to do a more in depth calculation, please view the lesson detailing it. The standard deviation of Group 1 is 1.52, and the standard deviation of Group 2 is 1.92.

The first step in the t-Test is calculating how much variability is in the sample. This part of the formula will help inform us if the differences between the samples might be explained by how spread out the data is. The formula is simply squaring each standard deviation, and then dividing by the number of numbers in the group, and then adding them up.

So, (1.52^2 / 5) + (1.92^2 / 5). This becomes 0.462 + 0.737, which equals 1.199.

When are degrees of freedom lost?

- When calculating the mean.
- When calculating the t-value.
- When calculating standard deviation.
- When calculating variance.

Identify the statement that appropriately describes the null hypothesis:

- There is no difference between the groups.
- There is no difference between the individuals.
- There is a difference between the groups.
- There is a difference between the individuals.

Some high school students were divided into two groups, group A and group B, and given the same test. Group A consisted of 6 students and the average score was 12 with a standard deviation of 2. Group B consisted of 8 students and the average score was 11 with a standard deviation of 3. The variability within the groups is:

- 1.339
- 1.792
- 0.667
- 1.125

Lesson22: Using ANOVA to Analyze Variances between Multiple Groups

Definitions

I don't want to get too much into my own history, but I attend (or did attend, depending on when you're hearing this) a clinical program that emphasized a particular type of theory over others. This is, in fact, not uncommon; if you progress you'll hear more and more of the different camps that exist in psychology. Some of the larger ones include psychoanalytic, cognitive-behavioral therapy, and person-centered therapy. But these are just a few of the several hundred styles.

Eventually, you'll arrive at something called evidence-based practice. This is where a therapist uses techniques that have been demonstrated to work through statistical and experimental evidence. In my own opinion, if your therapist isn't using evidence-based practice, then you should run far and fast.

To create evidence-based practice, we have to look at, well, evidence. We need to compare the effectiveness of treating different psychological maladies. What's more, we need to be able to compare a whole bunch of different modalities because sitting there and doing ten statistical tests is no fun at all. We want to do just one.

The analysis of variance, usually shortened to ANOVA, is a statistical procedure for locating a difference between multiple levels of a single independent group mean. For those of you who are familiar with the T-tests, this is basically a way of running a whole bunch of those in one go. But why run an ANOVA instead of several T-tests?

It is unethical to run multiple statistical tests on the same data because you will eventually find something due to the sheer probability and not an actual relationship.

You'll be able to examine a large amount of data and look for interactions without having to do a whole lot more statistical analyses.

ANOVAs come in many flavors, so let's look at each one independently so your brain doesn't start trying to make them all smash together.

One-Way Between Subjects

We start with a one-way between subjects ANOVA, which is a ridiculous mouthful to say. One-way has several specific components that identify it as such, and they are:

One independent variable

An independent variable that has multiple levels

One dependent variable

Things are clearer with an example. Let's say you're interested in seeing how effective therapy is for a depressed patient. Your independent variable here will be your patient's amount of time in therapy, while the dependent variable is your patient's depression level.

Your multiple levels will be the amount of time in therapy. We will have our different levels in three-month intervals, so three, six, nine, and twelve months. But wait! We forgot about the between subjects aspect.

Between subjects is defined as a study in which the subjects are placed in mutually exclusive groups and will be compared to each other. So in our experiment, we will have four groups of participants, and each of them will remain separate.

Following our experiment, we will compute to see if there is a difference between the different levels of the independent variable. This will give us an F-ratio, which is defined as a score to determine the level of difference between the means. This score is checked in a specific table, or, if you're lucky enough, you will do the math on a statistical computer program and be given the significance value. If your F-ratio is significant, you will know that there is a difference between one of your variable levels.

All the F-ratio has told us is if there is a significant difference between any of the groups. If our F-ratio is significant, then we will run a follow-up test to determine what the difference is. The follow-up test will tell us exactly which groups are different. There are several iterations, and they require a bit more than we're going into here.

One-Way Repeated Measure

Very similar to a one-way between measures is a one-way repeated measure. Repeated measure indicates that the study uses the same group of participants for each level of the variable. Looking at a similar example, let's say we have one group of participants and check in with them every three months. This would give us the same number of scores, but we would be repeating the measures instead of comparing between them.

If you were doing the same testing on the same group of people, you may have what is called a carryover effect. Carryover effect is defined as previous levels or conditions that may cause subsequent assessments to be altered. Basically, if you're assessing the same group of people every three months, and most of the people got better in the first three months, then their 'already better-ness' will carry over to the next set of scores.

When calculating your F-ratio, the statistics will be slightly different when taking the carryover effect into account. This makes the F-ratio less likely to be statistically significant because you're looking for larger changes because smaller changes can be explained by the carryover effect.

Two-Way

Ready for stuff to get a little more complicated? A two-way ANOVA has the following characteristics:

Two independent variables

Each independent variable has multiple levels

There's one dependent variable

The purpose of running an ANOVA is to find the _____ .

- differences between means
- similarities between means
- characterizations shared between means
- most effective score

Brian is conducting a study to determine whether meditation affects body weight. Brian plans to have three different groups, where each group will be assigned a different length of meditation time. When Brian gathers his data, he will compare the body weights of the participants between groups. Which type of ANOVA is Brian using?

- Between subject
- Repeated measure
- Mixed method
- Multivariate

In order to conduct a study to determine whether meditation affects body weight, Brian plans to have one group participate in different lengths of meditation time. After each length of meditation time, Brian will measure the body weight of each individual in the group, which is _____ form of ANOVA.

- between subject
- repeated measure
- mixed method
- multivariate

Scientific Method and Application of Statistics in Business

By: Seyedeh Fateme Moezzi