[MUSIC PLAYING]

JOHN WILLOUGHBY: Hello. And welcome. My name is John Willoughby. And I'll be teaching this course on the basics of statistical analysis for use in natural resource management.

The course is in four modules. This is module 1, Introduction to Statistics. You can see the other modules that I'll also be doing. One is on descriptive statistics, the second one. Third one is inferential statistics, where we estimate a single population parameter from a sample. And then the last module is another inferential statistics. But here, we compare two or more population parameters to each other to see if they are different or not, if the populations differ.

Some pertinent references are two books that I'm a coauthor on,*Measuring and Monitoring Plant Populations,* with Dan Salzer and Caryl Elzinga, and *Monitoring Plant and Animal Populations,* again, with Caryl and Dan, also with James Gibbs.

Most of, but certainly not all of, the material that I'll cover in these modules can be found in these two publications. The first one is available free, both as a digital download and as a hard-copy book. And instructions on ordering the book from the BLM are given on this slide.

The other book is copyrighted and, therefore, has to be purchased. It's available from Amazon, Barnes & Noble, and other booksellers.

So again, this is the first of four modules. This is the Introduction to Statistics module.

So what is statistics? Well, very simply, it's the science of data. Data are numerical values containing some information. More specifically, it's a collection of methods for collecting, displaying, analyzing, and drawing conclusions from data.

Statistics is often divided into two branches, descriptive statistics, which involves organizing, displaying, and describing data, and inferential statistics, which involve drawing conclusions about a population based on information contained in a sample taken from that population-- usually, just a single sample.

And here are sources of the definitions that I just gave you. And these are actually very good resources on statistical analysis in general.

So let's look at a wildlife example. Let's say we're interested in determining the average chest height of male moose in a large area incorporating Nova Scotia and New Brunswick, Canada.

Chest heights were of interest here because the author of the study was interested in comparing the success of moose versus white-tailed deer in the snow. And moose are taller than white-tailed deer. And so the author was trying to see if they function better in the snow than white-tail.

So how do we go about this? Well, if we could somehow find and measure the chest height of every male moose, we could simply record all of the heights and take the average. No problem, right?

This would be the true population mean, designated by the Greek symbol mu-- looks like a U. But it's pronounced "mu." And mu is a population parameter. It's what we'd like to know. We'd like to know what the average chest height of every male moose in this area of study is.

So obviously, we can't somehow track down thousands of moose in a 49,000 square mile area and measure all of their chest heights. So what do we do instead? We're going to take a sample.

If we take a sample and measure the chest heights of that sample, we're hopeful that that will be also representative of the population. So we do-- the sample average is referred to as the sample mean. And it's designated by x, this x with a bar over it. And it's called "x bar." And x bar is a sample statistic.

Now, you might wonder, well, how do we even take this sample of moose? Do we go out and capture them? I guess we could do that. But in the study itself, which is referenced here at the bottom of the slide, the author measured moose that had been killed at hunter check stations. So that's the source of the data.

So here are the results. So small n, which is the sample size-- we had 94 male moose. And the mean chest height was 100.46 centimeters.

And describing properties of the data, such as the number of observations and the mean, is an operation of descriptive statistics. If we use the sample mean to make conclusions about the population, this is an operation of inferential statistics. And in fact, that's what this author was about-- was trying to make inferences to the moose population as a whole. And here, for example, we could use the sample mean, x bar, as an estimate of the true population mean, mu.

So parameters versus statistics-- a parameter is a number that summarizes some aspect of the population as a whole. A statistic is a number computed from the sample data. And we saw the sample mean in the case of the moose. But there are other statistics that we could use that we will talk about, including the median or an estimate of the population total or a proportion. We'll talk about those during this-- rest of this course.

So here are some example parameters. Here's mu, which we've already talked about, which is the population mean, the true population mean. That's what we'd like to know. But typically, we're not going to know the true population mean. Instead, we're going to take a sample. And that is a statistic, the sample mean. And that's x bar.

Another population parameter is sigma, this Greek symbol here. And that's the population standard deviation. Again, we're not going to know the true population standard deviation in most cases. And so we calculate the standard deviation from a sample. And that is designated by the letter s, or sometimes you'll see SD.

p is the population proportion. That's another parameter we could be interested in, denoted here just by the letter-- lowercase letter p, although sometimes you'll see uppercase letter P used. And that's the true population proportion. Again, we'll typically take a sample. And the sample proportion is designated by this what's called p hat, p with this little symbol on top of it that looks like a hat.

And finally, another important parameter is big N, uppercase N, which is the number of members of the population. Lowercase n is the number of members of the sample. Sometimes, big N-- we'll know what that is, particularly if we're sampling area. But oftentimes, like in the case of the moose example, we won't have a clue because we have to know the true number of moose in this very large area. And that's something we'll just never know.

So some cases-- big N, the number of members of the population, is known. In other cases, it isn't. But we'll always know how many members we have in our sample. That's small n.

So another thing I want to talk about here is the difference between an experiment and an observational study. And these are important differences, although, as we'll see, many of the statistics we use are exactly the same.

So an experiment-- the focus of the design is on the random assignment of treatments to groups. We'll look about that-- look at that here in a minute in more detail.

Levels of particular treatments are manipulated by the researcher. The treatments, which also usually include a control in which no treatment is applied, are replicated several to many times on experimental units.

Experiments can be used to determine cause and effect. They're very important in that regard. So here's a couple of examples involving plants and vegetation.

So let's say that we want to find out-- we have two treatments that we're going to apply to a particular crop. And we want to know if these two treatments improve the production of the crop. So the two treatments are treatment A and treatment B. And then we have a control, where we do nothing.

So over here on the right, you can see we have four blocks of three fields each. And in each block, we have a control and two treatments, treatment A and treatment B. In each of those blocks, we have randomly assigned the treatment to one of the three fields. That's very important, random assignment of treatment and control.

Typically, we block because fields that are close together are going to respond more similarly to one another than fields that are further and further apart. And so each block consists of three fields that are very close together. The blocks could be quite a ways apart. It doesn't matter in terms of the analysis because the blocking is taken into account. But we're not going to go into the details of that here.

But if-- the bottom line is if statistical analysis determines that the production is higher under Treatment B, then we can attribute that increased production to Treatment B In other words, Treatment B caused this increased production, cause and effect.

Here's another example. So we have, in this case, 10 blocks. And here we've just got a single treatment and control. And we're interested in whether burning affects a particular rare plant. We think that this plant responds favorably to fire.

We know that-- we see it a lot after wildfires. And before the wildfires, we don't see it very much. So it looks like it could be a fire follower.

So in order to determine this, we set out these 10 blocks. We randomly assign a treatment to one of those-- of the two plots in each of the blocks. And we burn that one. And we leave the other one unburned. And then we look at the response of this rare plant.

So if we do a statistical test and we determine that the rare plant increases in density in the treatment plots, but not in the controls, then we can say that fire is the reason, is the cause, of this increase.

So that's an experiment. Most of-- well, really, all of what we're going to be talking about in the rest of this course involves observational studies, mostly monitoring, although much of what we say can involve situations that don't involve monitoring the two, where you want to compare two populations not necessarily over time. But these are observational studies.

So what's an observational study? Well, again, I said we'll focus on using statistics to analyze data by sampling natural plant communities, particularly over time. And a monitoring study looking at change over time can be considered a type of observational study.

In this case, causation is much more difficult to prove because we haven't randomly assigned treatments to plots. Rather, we're tracking change over time in a given area. So it's hard to prove cause and effect. But that doesn't mean that management decisions can't be made. They are routinely made, actually, based on observational data.

So most of the statistical methods that we'll talk about in this course are the same whether we're talking about an experiment or an observational study. Some methods, though, such as the finite population correction factor, are unique to studies involving sampling. And I'll talk about what this beast, the finite population correction factor, is later.

A focus of some monitoring studies is to estimate population totals for a rather large area. And that's something that's usually not the focus of an experiment. So that's a couple of ways where observational studies and experiments differ.

But the design of both experiments and observational studies requires a random component. This is really important. So if we're doing an experiment, this randomness is provided by the random assignment of treatment to experimental units. In an observational study, it's provided by the random sampling of the population of interest.

So let's look at an example here where we're sampling to detect change over time using temporary quadrats. So here we have a macroplot, a large plot, within which we're going to do some kind of sampling. And in this case, we decided that we're going to sample with quadrats. And we're going to sample with 25 quadrants. Here they are in a grid.

And so because this is an area, we can grid off the area, at least in our minds, and then go to these quadrat locations by taking coordinates. So we can randomly choose a coordinate on the x-axis and on the-- a y-axis and then go to those randomly selected quadrats and measure whatever it is we're measuring that's of interest to us.

So here's the first year. And we've taken a random sample of 25 square quadrats. And we, again, randomly position them. So here's the next year. And here we've taken another sample of 25 random quadrats. But they're in different locations.

So the quadrats are said to be temporary in this case. They're within the same macroplot because that's the area that we're sampling. But they're in different positions because we took a separate random sample in the second year. So maybe we made cover estimates in this sample of quadrats.

I want to talk briefly about qualitative versus quantitative methods because oftentimes when we do monitoring, an important part of monitoring is often qualitative. We'll go out and we'll simply look and see if things have changed over time or we'll take photos.

Very important part of monitoring-- in fact, it's more common than taking quantitative measurements. But we're focusing on the quantitative measurement part of the issue here because it's important to understand how to handle quantitative data once you have it in hand.

But qualitative methods typically don't collect numerical data. So what are some examples? Well, photographs, field journal entries, subjective assessments of ecosystem health-- those are all examples of qualitative methods. I'm sure you could come up with several more.

Quantitative methods, on the other hand, involve the collection and/or analysis, usually both, of numerical data. Some examples-- the number of plants or animals in a sample of randomly positioned plots, the height of trees in a sample of trees from a forest, the number of "hits" on bare ground in a sample of 200 randomly position points. All of those are examples of quantitative type of data.

Statistical analysis-- you don't need it for qualitative methods-- not appropriate, don't need it. Everybody's saying, yay, let's just do qualitative. Well, qualitative-- it doesn't always get us what we need.

Any time we collect quantitative data, we need something, some sort of statistical analysis. It may not be very involved. But we need to somehow-- to do something, at least summarize the data that we've collected.

And even in those rare cases where we completely measure the entire population where the-- maybe we've got a population of a rare plant species and there's only 100 individuals. It's really rare. And so we can count them all. So we can do a complete census.

But even in those cases, we still are-- often are interested in population parameters, such as the mean and standard deviation. So we still would apply the techniques that we're going to describe under descriptive statistics.

But in most cases, a collection of quantitative data involves sampling from a population of interest. And in that case, we apply both descriptive and inferential statistics, which brings us to descriptive versus inferential statistics. What's the difference?

Well, descriptive statistics describe the characteristics of a sample or in-- as I mentioned, in rare cases, this could be the entire population. Inferential statistics use the data from a sample to make inferences about the larger population from which the sample was drawn.

And statistical procedures quantify the uncertainty associated with the sample. Some have called statistics the science of uncertainty because any time we sample, we're hopeful that the sample represents the true population. But we need to assess whether that's true or not using statistical analysis. And that's what statistics does.

So statistical techniques are also required to determine if two or more populations are different from one another. And module 2-- we're going to focus on descriptive statistics. And in module 3 and 4, we'll cover inferential statistics.

[MUSIC PLAYING]

So that's the conclusion of module 1-- look forward to seeing you in module 2.