

---

Hello. In this episode of Sociology Studies of Society, today's lesson is on statistics. As always, don't be afraid to pause, stop, rewind, or even fast forward to make sure you get the most out of this tutorial.

So what is statistics? Well at its most basic function physics is taking a really disorganized group of raw data and turning it into something more manageable that sociologists can look at and understand and draw meaning from. Now it's not as simple as just organizing these numbers in numerical order, but that can be part of it.

So today, we're going to look at some different parts of statistics. Now when you look at this line here of organized numbers versus the disorganized when we open with, those are the same numbers. And once they're organized, we can start to see that the most common number 7, there's three 7's up there. There's no other number that occurs more often.

And the middle most number is 6. It's in the exact middle between the 1 and the 10. And the average, numerical average is 5.6.

Well, we're going to kind of explain what are some statistical terms for those three things that I just laid out for you. Now, one way to think about statistics is kind of breaking it down into descriptive statistics. Now descriptive statistics is a way to describe the most typical individual of a group.

So a lot of the world would say, describe the average individual for a group, but average really means something a little bit different. So most typical individual is a much better explanation. There's the mean, the median, and the mode.

Now the reason we don't say that descriptive statistics describes the average individual or group is because the mean is the mathematical average. Now the median is the middle most point in the number set. And the most is the most common value in a group.

So in that opening part there, where I describe the most common number 7, the most common number being 7 is can also be called the mode. And the median, so the number that's in the very center, is 6. And that can also be called the median. And the mathematical average of 5.6.

Another term to describe that is the mean. So just to be clear, let's make sure you know how to find these. First one is mode. Mode you just look at the number set. And you find the one that occurred the most.

So there are three 7's. That's more 7's than any other group. There's two 9's, there's two 5's. There's two 2's. And there's one of a bunch of numbers, and there's zero of some numbers as well.

Now, the median is you want to find the exact center number. And now the value of the numbers don't matter, except that they're ranged from highest to lowest or from lowest highest actually, either way it works. So what you do is you just slowly start to go in from the outside one at a time on both sides. And that's how you'll find the exact median.

So as we slowly go together, you'll see, oh, wait, 6 is in the exact middle. Now if there's an even number of numbers, there's not going to be an exact middle, because it might fall between 6 and 7. So then you just add those two together and divide them. And that's how you'd find the median.

And the mean is the mathematical average. And that's why you add up all the numbers together, which equals 32. Then you divide it by the number of numbers there are, which is 11. And you get the answer that the mathematical average is 5.6.

Now let's take those examples and give us a little more complicated data set and use in an actual example. Now this example is number of cans of soda per day. If I were to go around the office here and ask different people how many cans of soda of per day, these are some answers I might get.

Now I've already taken the liberty of organizing these from lowest to highest. Now let's process that data. So the mode, the median, and the mean, let's actually start with the mode. Start with the bottom one, because it's so glaringly obvious in this graph, which one occurred the most?

Well, look, four people said that they had zero cans of soda per day. That one happened to the most. Next one was the median. So there, you count in from the outside to find the one that's in the exact center.

And in this case, that one, the exact center is 1. And then the mean is 2.9. So that's the average.

So when you add up all of those individual cases and divide them by 11, you get 2.9. You can see here these numbers are very different. So when you're saying the typical person in my office, you could use any one of these statistics and say a very different answer.

You could say most people in my office drink zero cans of soda per day. That would be factually correct. You could say the average person, the mathematical average person in my office drinks 2.9 cans of soda per day. That number is very different from zero, but it still is a typical-- it's a descriptive

statistic. You're giving what the typical person in this group is doing.

Now inferential statistics is when you take data from one group and you apply it to a larger population. So if I, instead of taking just this small numbers that I collected here, and I want to say that this is true for all working class individuals in America, that would be inferring, that would be applying the data to another group, a larger population. Now there can be problems with it, but that's what inferential statistics do.

And sociologists really take care to have a really representative sample and a fair sample and normally also a large sample so they can actually infer about society at large or a larger population. Now there are different ways that you can draw connections between two different variables. Now the first one is a correlation.

A correlation is just expressing that this relationship between the likeness of two things happening. That sounds very complicated, but it's not. It's saying that when one thing happens something else is also more likely to happen or also less likely to happen. Correlations can be found everywhere.

They're very different from cause and effect. In fact, one of the famous lines that you've probably heard before is that correlation is not causation. So just because two things are related doesn't mean that the one thing caused the other thing to happen. Cause and effect is when you actually can find that this thing caused the other thing.

So if we look at an example of maybe drinking diet soda-- I mean, there have been studies found that when you're drinking diet soda, you're more likely to be overweight. So there's a correlation between drinking diet soda and having high weight. Now it's probably not fair to say that's a cause and effect, because diet soda may not be the thing making you overweight. Maybe you are already overweight, and so you chose to start drinking diet soda.

There can be another variable in there. And so you really want to try to prove cause and effect. And if you can't prove it, then it's simply a correlation.

There's often the third variable that is the thing that's causing those two things to be related. Now one thing you want to look out for, actually two things you want to look out for, one is a spurious correlation. And that's a false correlation. And that's really what I was talking about with that third variable idea.

That's when two things actually aren't related, but they seem like they are, because there's some other thing that happens that cut ties them together. Now another thing you need to watch out for is

the Hawthorne effect. Now the Hawthorne effect is when subject experiments change their behavior, because they're being studied.

If I'm doing a study about eating habits and I have people write down in their journal what they're eating every day, you have to be careful that the act of writing in the journal isn't changing the way that people are acting, because you're going to get results that are not true to the real world. So today's takeaway message, descriptive statistics are used to describe the most typical member of a group. And you have the mathematical average, which is the mean. You have the middle point, which is the median. And you have the most common value, which is the mode.

Now when you take a data set and you apply that data to a larger population, that's inferential statistics. We also learned about correlations. And that's our relationship expressed between the likeness of two variables changing together. There's spurious correlations or false correlations. This is when two things seem to be correlated, but actually there's some other thing that we haven't found yet that's causing them to be related.

And there's cause and effect, and that's when the relationship in one actually causes the other one to change. Then there's also the Hawthorne effect, which is where subject and experiment behave differently, because they are subjects and experiment. That's it for this lesson. Good work, and hopefully you'll be seeing me on your screen again soon.