VIDEO: “More on sampling and confidence intervals” (3:11)

(00:00)
For me to understand difficult concepts I might execute MATLAB code to mimic statistical functions and see if my results match the results of the build in function. I will show you a way to proximate using code you have already learned. Imagine you wanted to know something about all morning science student grades. Like maybe the mean- is it possible to find the mean of this population? No way, we would need to record all of those grades. So, what do we do? We might sample a thousand grades and find the mean and standard deviation of the sample, but does that tell us the mean of the population? No. For example, if I sample another thousand grades and sample a second mean the two means would probably not be the same due to natural variation. To account for the variation we stimulation counting thousands of samples. In MATLAB I can do it by creating ten thousand similar samples by using the random function, finding normal distributions and using the mean and standard deviation of our original sample. I want a thousand numbers in each sample to match our original sample size and ten thousand samples to test for much variation.

(01:15)
Next, I can find the mean of each of the ten thousand samples. Each will be similar to the mean of our initial sample but with some variation due to the random function. If I were to create a histogram of the means of the samples, it might it might range from 84.5 to 86. Hopefully this is not obvious, but this is where many people get confused. These are the means of grades not the grades themselves. I can use the percentile function to find the numbers at the 2.5 and 97.5 percent’s which correspond to the middle 95% of the numbers and this gives us a good approximation to the t-test 5% confidence interval. The hurtle to understanding hypotheses testing is missing the conversion from a sample of grades to many means of sample of grades that are like our initial sample.

(02:10)
Now that we have constructed our confidence interval for our population let me show you the process to find our p value of our test case of 85. Remember we have already seen the codes to find the means of all ten thousand samples. Also remember the percentile function gives me a number at a specific percentile of a list of numbers but I don’t know a function to do the opposite. Specifically, I need to find where 85 falls among all of the numbers, so I wrote my own function to do this. You have not learned this yet, so I won’t confuse you with the code. Here is what it does- first we put the sample means in order from low to high then we find the position of a number that is very close to our test case. Its percentile is relative to the size of its sample. In this case 13.8% from the left since this is two sided, we multiply this percentile by two and we have a close approximation of the p value of 85 given the sample. I hope this video helped you understand using MATLAB for hypotheses testing.