

Transcript

Video – 0421 Unit4B Inference for Relationships

01. 00:00 / 00:05 - We are now ready to talk about inference for relationships which I think is one of the
02. 00:05 / 00:11 - most fun parts of the course. It's where we begin to start to apply statistics to real
03. 00:11 / 00:18 - life situations in the previous modules, we learned to perform inference for a single
04. 00:19 / 00:25 - categorical or quantitative variable in the form of point estimation, confidence intervals,
05. 00:25 / 00:31 - and hypothesis tests. We used inference about the population proportion when the variable
06. 00:31 / 00:36 - of interest was categorical and about the population mean when the variable of interest
07. 00:36 / 00:41 - was quantitative. The main goal was to introduce you to the concepts of confidence intervals,
08. 00:41 / 00:47 - to the concepts of hypothesis tests, and to do that using two fairly simple examples where
09. 00:47 / 00:53 - we could go through of all the theory behind why it works and walk you through the steps
10. 00:53 / 01:00 - one at a time. Now that we've gotten through that material, our final goal is to look at
11. 01:01 / 01:06 - performing inference on relationships. We will focus more on what type of test we should
12. 01:06 / 01:12 - conduct, what do the results mean, and we will use the computer to calculate the p-values
13. 01:12 / 01:18 - and confidence intervals for us when we need them. The idea here is we have a population
14. 01:18 / 01:23 - of interest and we're going to measure two variables on that population, variable X and
15. 01:23 / 01:30 - Y and we're going to get a sample from our population, have the variables X and Y measured,
16. 01:30 / 01:36 - and then ask ourselves based upon the observed data do we have enough evidence that X and
17. 01:36 / 01:43 - Y are related. Across all the influential methods that we will learn, our hypotheses
18. 01:43 / 01:50 - will really, underneath, be the same hypotheses. We might have some specific symbols relating
19. 01:50 / 01:57 - to means or proportions in certain tests but they all boil down into the same kind of hypothesis.
20. 01:57 / 02:02 - The null hypothesis is going to be that there is no relationship between X and Y and the
21. 02:02 / 02:08 - alternative hypothesis is going to be that there is a relationship between X and Y. No
22. 02:08 / 02:14 - matter what symbols we might use you can always boil your hypotheses back down into these
23. 02:14 / 02:20 - two statements where X and Y are going to be whatever variables it is that you're discussing in the
24. 02:20 / 02:27 - problem. And when you write your conclusions you can always come back to these hypotheses
25. 02:27 / 02:32 - and state that they're either was enough evidence that there is a relationship between X and
26. 02:32 / 02:39 - Y or there was not enough evidence of a relationship between X and Y. Most of what we're going
27. 02:39 / 02:45 - to do is learn hypothesis tests but we will supplement some of these with confidence intervals.
28. 02:45 / 02:52 - Recall our four cases in our table here. We have our explanatory variable which could
29. 02:53 / 02:57 - be categorical or quantitative. We have our response variable which could be categorical
30. 02:57 / 03:04 - or quantitative. And in the next three modules were going to analyze the cases CQ, CC, and
31. 03:05 / 03:12 - QQ. Recall that we will not talk about using QC in this course where we have a quantitative
32. 03:12 / 03:18 - explanatory variable that we are using to predict a categorical response variables.
33. 03:18 / 03:22 - That's the only case that we will not talk about. We will also look back on and we'll
34. 03:22 / 03:28 - see popping up the same things that we did. For two categorical variables, we are going
35. 03:28 / 03:34 - to start with a two-way table. For a categorical predictor or explanatory variable and a quantitative
36. 03:34 / 03:39 - response, we are going to start with boxplots by groups. For two quantitative variables
37. 03:39 / 03:44 - were going to start with a scatterplot. But we are going to turn all of those into, now
38. 03:44 / 03:49 - that we know about inferential statistics, the relationship that we see when we look
39. 03:49 / 03:54 - at the data descriptively, can we say that that is statistically significant. We are
40. 03:54 / 04:00 - just really adding one extra piece to what we learned, we're coming back and being able
41. 04:00 / 04:06 - to test, is what we see happening in this descriptive summary indicative of what we
42. 04:06 / 04:12 - would expect to happen in the population. In this material we're not going to focus
43. 04:12 / 04:18 - on the detail or the theory much behind each of the tests that we talk about. We will mention
44. 04:18 / 04:23 - the test statistics and we might mention the null distribution of the test statistic under
45. 04:23 / 04:27 - which the p-values are calculated but we're not going to talk about the theory behind
46. 04:27 / 04:32 - how we know that that is true. We did that with proportions based upon the central limit

47. 04:32 / 04:38 - theorem basically and the normal approximations that we talked about. So here our goal is
48. 04:38 / 04:43 - going to be when is the inferential method appropriate. Under what conditions is the
49. 04:43 / 04:50 - procedure safe to be used. What's the conceptual idea behind the test, in other words sort
50. 04:50 / 04:57 - of what is going on behind the idea of the test but again no details about how we prove
51. 04:57 / 05:02 - the sampling distribution or anything like that. We're going to use software to carry
52. 05:02 / 05:07 - out the procedures and get the p-values. Reading the output will also be important and then
53. 05:07 / 05:11 - interpreting the results in the context out the problem.