

Transcript

Video – 0424 Unit4B Case CQ Two Independent Samples A

01. 00:01 / 00:06 - Now we move on to independent samples. Here is the summary of the tests that we are learning
02. 00:06 / 00:12 - for the case where k equals 2. We've completed our discussion on dependent samples where
03. 00:12 / 00:18 - we covered the paired t-test, the Sign test, and the Wilcoxon Signed-Rank test, and now
04. 00:18 / 00:23 - we will move on to independent samples where we will cover the two-sample t-test and the
05. 00:23 / 00:30 - nonparametric alternative the Mann-Whitney U or Wilcoxon Rank-Sum test. We've discussed
06. 00:30 / 00:36 - the case for dependent samples. In this case we have observations in one sample which are
07. 00:36 / 00:43 - matched, paired, or linked with observations in the other sample. The matching could be
08. 00:43 / 00:48 - because the measurements are taken on the same individual or they could be taken on
09. 00:48 / 00:54 - two separate individuals and matched based upon other characteristics. To analyze, this
10. 00:54 / 00:59 - data we simply took the differences between the observations for each pair and analyzed
11. 00:59 / 01:05 - the differences using one sample techniques. Now we will discuss the case of independent
12. 01:05 / 01:10 - samples. In this case all individuals are independent of all other individuals in their
13. 01:10 / 01:16 - sample as well as all individuals in the other sample. This is most often accomplished by
14. 01:16 / 01:22 - one of two methods the first method is taking a random sample from each of the two groups
15. 01:22 / 01:28 - under study. The second method is taking a random sample from the entire population and
16. 01:28 / 01:34 - then dividing it into two subgroups based upon the grouping variable of interest. So
17. 01:34 / 01:39 - our example where we discussed talking about comparing heights of males and females, we
18. 01:39 / 01:44 - could consider taking random samples from each group individually. In other words taking
19. 01:44 / 01:50 - a random sample of females and then taking a random sample of males. Or we could take
20. 01:50 / 01:57 - a random sample from the entire population and sub-classify them into males and females
21. 01:57 / 02:03 - after the sample is taken. Recall that we don't have to have equal sample sizes in our
22. 02:03 / 02:09 - two groups in the case where we're looking at independent samples. Now we want to move
23. 02:09 / 02:15 - on to the two-sample t-test. Here the idea is we have a categorical variable X which
24. 02:15 / 02:21 - has two possible groups, two possible levels, and a quantitative response variable Y . So
25. 02:21 / 02:27 - we have our two independent samples, drawn in one of the two ways that we just discussed,
26. 02:27 / 02:34 - and then we want to see if the two population means in the two groups are different. This
27. 02:35 / 02:40 - also translates into asking the question is there a relationship between our quantitative
28. 02:40 / 02:47 - response variable Y and our categorical explanatory variable X . The test that we're going to begin
29. 02:47 / 02:54 - with is commonly known as the two-sample t-test and p-values for this test will be calculated
30. 02:54 / 03:00 - under some t-distribution which we will discuss as we go through the steps. So here we have
31. 03:00 / 03:06 - two figures that illustrate two of the examples that are covered in the material. On the left
32. 03:06 / 03:12 - we have age group and we're comparing the weights of males in two age groups, a young
33. 03:12 / 03:19 - age group is aged 20 to 29, and the older age group is age 75 and above. And we're asking
34. 03:19 / 03:25 - the question is there a significant difference in the mean weight of males 20 to 29 years
35. 03:25 / 03:32 - old and the mean weight of males 75 years or older. On the right we have a study concerned
36. 03:33 / 03:38 - with the length of pregnancy and comparing whether the length of the pregnancy is the
37. 03:38 / 03:43 - same or different among the pregnant women who smoke versus pregnant women who do not
38. 03:43 / 03:49 - smoke. In this case we had a simple random sample of the same size (35) from each group. And
39. 03:49 / 03:54 - we're comparing the mean length of pregnancy between pregnant women who smoke and
40. 03:54 / 04:00 - pregnant women who don't smoke. As we go through this material it is important to think about
41. 04:00 / 04:06 - the cases that we are in. We have these three scenarios depending on what kind of explanatory
42. 04:06 / 04:11 - variable and what kind of response variable. This is a big part of what we want to learn
43. 04:11 / 04:16 - in this class. You want to be able to take two variables that we've been given, classify
44. 04:16 / 04:22 - them quickly as categorical or quantitative figure out which is the explanatory variable
45. 04:22 / 04:28 - and which is the response variable. And then from there, decide which methods are appropriate.
46. 04:28 / 04:33 - Once we've decided on a method then we have to decide if that really is applicable. And

47. 04:33 / 04:38 - we're also providing you with some alternative methods when the standard methods don't apply.
48. 04:38 / 04:45 - So as you go through these last modules always ask yourself, how do I know I'm in case, in
49. 04:46 / 04:52 - this case CQ. Well I know I'm in case CQ because we have the response variable of interest
50. 04:52 / 04:58 - for weight or length of pregnancy being a quantitative response and we have two groups
51. 04:58 / 05:03 - in this case that represent the categorical variable that we want to compare. So on the
52. 05:03 / 05:09 - left we have age group young or older and we have on the right smoking group, yes or
53. 05:09 / 05:16 - no. On the left, if we had age measured quantitatively that would be a different situation. So think
54. 05:18 / 05:23 - about how the variable is being represented. Is the variable the explanatory variable or
55. 05:23 / 05:28 - the response variable? And then that helps lead you to what are the appropriate methods
56. 05:28 / 05:34 - to try to analyze this data. Once we sort of know what method we are using then we can
57. 05:34 / 05:37 - dig into the details of how those methods work.