## Transcript

**Video – 0424 Unit4B Case CQ Two Independent Samples E**

01. 00:01 / 00:06 - Our next example, we're going to look at BMI versus gender in heart attack patients. This
02. 00:06 / 00:11 - is the WHAS500 study. It is a study that was conducted which enrolled patients following
03. 00:11 / 00:17 - a heart attack and followed them. And these were all individuals that came to the hospital
04. 00:17 / 00:24 - in a certain metropolitan area. So we're going to be looking at, is there a relationship
05. 00:24 / 00:30 - between body mass index and gender among heart attack patients? In our case we're going to
06. 00:30 / 00:36 - be looking at the null hypothesis of the difference is zero versus the difference is not zero.
07. 00:36 / 00:43 - So mu_1 is going to be the mean BMI for males in the population, which represents all heart
08. 00:43 / 00:50 - attack patients, and mu_2 is going to represent the mean BMI for females. Again we can phrase
09. 00:50 / 00:55 - these hypotheses in more general language and say the null hypothesis is that BMI is
10. 00:55 / 01:01 - not related to gender in heart attack patients versus the alternative hypothesis that BMI
11. 01:01 / 01:08 - is related to gender in heart attack patients. We're going to obtain the data, check conditions,
12. 01:08 / 01:14 - and summarize the data. Here the subjects were chosen randomly when they came to the
13. 01:14 / 01:18 - hospital to be enrolled in the study. They are naturally divided into a sample of males
14. 01:18 / 01:25 - and females and so we have independent samples in this case. We are also in the second scenario
15. 01:26 / 01:31 - where the sample sizes are extremely large. We have three hundred of one gender and two
16. 01:31 / 01:36 - hundred of the other and so we can proceed here regardless of whether the populations
17. 01:36 / 01:41 - are normal or not. When we look at the software output we're going to be in the second case
18. 01:41 / 01:46 - where we can not assume the variances are equal this will give us a test statistic of
19. 01:46 / 01:53 - positive 3.21. So that says that our result is 3.21 standard errors above the null value.
20. 01:57 / 02:03 - In this case we're looking at mu_1, which was males, minus mu_2, which was females.
21. 02:03 / 02:07 - To find the p-value we're going to let the software do this. The p-value is going to
22. 02:07 / 02:13 - be reported as 0.001 so it is pretty rare that we would get data like that observed.
23. 02:13 / 02:17 - So we're going to reject the null hypothesis and when we get to our conclusions were going
24. 02:17 / 02:22 - to be able to say the data provide strong evidence against Ho. So we're going to reject
25. 02:22 / 02:27 - the null hypothesis and conclude that the mean BMI of males differs from that of females.
26. 02:27 / 02:34 - We could also say that males and females differ with respect to BMI among heart attack patients.
27. 02:34 / 02:38 - As before we can follow this up with a confidence interval. In this case a confidence interval
28. 02:38 / 02:45 - for mu_1 minus mu_2 is roughly 0.63 to 2.64. We can therefore say that with 95 percent
29. 02:47 / 02:54 - confidence, the population mean BMI for males is between 0.63 and 2.64 units larger than
30. 02:54 / 03:01 - that of females. So females tended to have lower BMI than males. Males tended to have
31. 03:01 / 03:06 - larger BMI than females. The reason that we want to obtain the confidence interval is
32. 03:06 / 03:10 - that it does indeed quantify this effect. If we just said there is a difference, we
33. 03:10 / 03:15 - wouldn't know how much difference there is. It could be 0.02. But we're looking to quantify
34. 03:15 / 03:20 - how much of a difference. Now we can say clearly that the average BMI for male heart attack
35. 03:20 / 03:27 - patients is between 0.63 and 2.64 units larger than the mean BMI of female heart attack patients,
36. 03:29 / 03:36 - based upon the results of this study with 95 percent confidence. It gives us some way
37. 03:36 / 03:42 - to talk about how large the effect is which is usually important in most applied problems.