

## Transcript

### Video – 0424 Unit4B Case CQ Two Independent Samples D

01. 00:01 / 00:05 - Now we'll do something a little more exciting, look at an example. We will actually look  
02. 00:05 / 00:10 - at two examples. The first example we're going to look at is, what is more important to you  
03. 00:10 / 00:16 - personality or looks. This question was asked of a random sample of 239 college students  
04. 00:16 / 00:23 - and they were to answer on a scale from 1 to 25 where an answer of 1 meant personality  
05. 00:23 / 00:29 - had a maximum importance and looks no importance at all and an answer of 25 was the other way  
06. 00:29 / 00:36 - around, looks had a maximum importance and personality no importance at all. The purpose  
07. 00:36 / 00:42 - was to examine whether males or females differ with respect to importance of looks versus  
08. 00:42 / 00:48 - personality. The data are provided and have the following format: on the left we have  
09. 00:48 / 00:53 - their score  $Y$ , which is our quantitative response variable, and on the right we have their gender,  
10. 00:53 / 00:59 -  $X$ , which is our categorical explanatory variable. Here is the picture that represents the situation.  
11. 00:59 / 01:06 - We have population 1 is females, population 2 is males. That is our categorical explanatory  
12. 01:06 / 01:10 - variable. We have a sample, which in this case we've taken a big sample. And then we've  
13. 01:10 / 01:17 - split them into the males and females based upon what gender they were. So that we have  
14. 01:17 / 01:24 - resulted in a simple random sample of 150 females and a simple random sample of 85 males.  
15. 01:24 / 01:30 - So there were four missing observations in the dataset and this is where those four individuals  
16. 01:30 / 01:36 - went. We have the  $Y$ , which is our score,  $X$ , which is our gender, and we have a sample  
17. 01:36 / 01:42 - from each of our two genders. This is clearly an independent sample case. Each female is  
18. 01:42 / 01:47 - not matched, paired, or linked in any way with a male in the other sample. And inside  
19. 01:47 / 01:51 - each sample they are independent due to the fact that we took a random sample to begin  
20. 01:51 / 01:58 - with. We begin by stating our hypotheses. In this case we just want to see whether the  
21. 01:58 / 02:03 - opinions of females and males differ with respect to the importance of looks versus  
22. 02:03 / 02:10 - personality. The hypotheses are going to be that  $\mu_1 - \mu_2$  is zero versus  $\mu_1 - \mu_2$  is not zero, where  $\mu_1$   
23. 02:10 / 02:16 - is going to be the mean looks versus personality score for  
24. 02:16 / 02:22 - females and  $\mu_2$  is going to represent the mean score for males. We can also word this  
25. 02:22 / 02:29 - in terms of the concepts and say the null hypothesis is that score of looks versus personality  
26. 02:29 / 02:35 - is not related to gender for the null hypothesis versus score is related to gender for the  
27. 02:35 / 02:41 - alternative hypothesis. So in step 2 we are going to obtain our data, check conditions, and  
28. 02:41 / 02:48 - summarize the data. Here we have 239 students. Once we break them into the two samples, there  
29. 02:48 / 02:55 - are 150 and 85. We were told it was a random sample. So we are satisfying that condition  
30. 02:55 / 03:00 - and then the sample sizes in both groups are large, so we don't need to necessarily check  
31. 03:00 / 03:06 - whether the populations are normal or not, although we still can. In this case we will  
32. 03:06 / 03:10 - see in the software that we're going to be in case A where we are going to be able to  
33. 03:10 / 03:16 - assume the variances are equal. This is going to give us a test statistic of negative 4.58.  
34. 03:16 / 03:21 - So we will find that value when we go into the output in a moment. We're going to let  
35. 03:21 / 03:27 - the software find the p-value for us. In this case the p-value is less than 0.001 and in  
36. 03:27 / 03:33 - fact it is practically zero so that's going to say that it's almost impossible to get  
37. 03:33 / 03:39 - data like that observed, if the null hypothesis had been true, if there were no differences  
38. 03:39 / 03:44 - between males and females with respect to whether they value looks versus personality  
39. 03:44 / 03:49 - more. It would have been almost impossible to get data where the difference between the  
40. 03:49 / 03:56 - sample means of females and males was negative 2.6. This value, negative 2.6, is our estimate  
41. 03:57 / 04:02 - of the difference between the means. But what we are asking, is this statistically significant,  
42. 04:02 / 04:08 - and we are using the fact that what we see is that our estimator was 4.58 standard errors  
43. 04:08 / 04:15 - below the null value. Once we have our p-value, we can write our conclusion. In this case  
44. 04:15 / 04:19 - we're going to reject the null hypothesis and that's going to say that the data provide  
45. 04:19 / 04:24 - strong evidence against the null hypothesis. So we reject it and conclude that the mean  
46. 04:24 / 04:31 - importance score of looks versus personality of males differs from that of females. In

47. 04:31 / 04:37 - other words males and females differ with respect to how they value looks versus personality.
48. 04:37 / 04:41 - As a follow-up to this conclusion, we can construct the confidence interval for the
49. 04:41 / 04:46 - difference between population means and in this case we're going to let it be  $\mu_1$ ,
50. 04:46 / 04:53 - which was females, minus  $\mu_2$ , which was males, and using the software we get roughly negative
51. 04:53 / 05:00 - 3.7 to negative 1.5 and this lets us be able to say with 95 percent confidence that the
52. 05:01 / 05:08 - population mean of looks versus personality score for females is between 3.7 and 1.5 points lower
53. 05:10 / 05:17 - than that of males. The confidence interval therefore quantifies the effect that the explanatory
54. 05:17 / 05:24 - variable gender has on the response, looks versus personality score. Since low values
55. 05:24 / 05:29 - correspond to the personality being more important and high values correspond to looks being
56. 05:29 / 05:36 - more important the result of our investigation suggest that on average females place personality
57. 05:36 / 05:43 - higher than do males or alternatively we could say that males place looks higher than do
58. 05:43 / 05:43 - females.