TRANSCRIPT:

Video: Case CQ – More than Two Independent Samples

1. 00:00 / 00:05 - Let's review what we've covered so far. We've covered the dependent sample case where we've
2. 00:05 / 00:12 - done the standard test which is the paired t-test as well as the two non-parametric alternatives
3. 00:12 / 00:17 - the Sign test and the Wilcoxon Signed Rank test. And then we've talked about independent
4. 00:17 / 00:23 - samples where we've covered the two-sample t-test which has two sub-classifications depending
5. 00:23 / 00:30 - on whether we can assume equal variances or not. And we covered the nonparametric alternative
6. 00:30 / 00:37 - in that case which is the Mann-Whitney U or Wilcoxon Rank-Sum test. Now we're going to
7. 00:37 / 00:42 - move on to the case where k is greater than 2, when we have independent samples, and we're
8. 00:42 / 00:47 - going to learn two new tests in this area. We're going to learn the standard test which
9. 00:47 / 00:54 - is the one-way analysis of variance or ANOVA and we are going to learn the non-parametric
10. 00:54 / 01:01 - alternative to that which is the Kruskal-Wallis one-way ANOVA. There are tests for dependent
11. 01:01 / 01:06 - samples when you have k greater than 2. One of those is the repeated measures ANOVA where
12. 01:06 / 01:13 - you measure, instead of just before and after, you might measure before, during, and after,
13. 01:13 / 01:19 - so that you have three measurements over time on the same individuals. So our goal now is
14. 01:19 / 01:26 - to look at how to compare means, how to compare groups when we have more than two groups.
15. 01:26 / 01:33 - And the idea is basically to extend the two-sample t-test to allow for more than two independent
16. 01:33 / 01:38 - samples. Here is a picture describing this process. We're going to have our explanatory
17. 01:38 / 01:45 - categorical variable, X, we're still in case CQ where we have a categorical explanatory
18. 01:45 / 01:51 - variable and a quantitative response variable. So our explanatory variable X is going to
19. 01:51 / 01:59 - have k levels now and we're going to be taking a random sample from each of the k populations.
20. 01:59 / 02:05 - And our goal is to compare the population means between these k populations using the
21. 02:05 / 02:11 - samples that we obtain. The sample sizes don't have to be the same in each group but the
22. 02:11 / 02:17 - samples do need to be independent. The test that we're going to develop is different in
23. 02:17 / 02:22 - two ways from the tests that we've talked about previously. The first of these is that
24. 02:22 / 02:28 - unlike previous tests where we had three possible alternative hypotheses to choose from, not
25. 02:28 / 02:34 - equal to, less than, or greater than, in the ANOVA test there's only one alternative. So
26. 02:34 / 02:38 - we don't even have to worry about what possible alternative hypothesis we are going to state,
27. 02:38 / 02:44 - it's always going to be the same alternative hypothesis. Also the second thing, the test
28. 02:44 / 02:48 - statistic will not have the same structure as the test statistics we've seen so far.
29. 02:48 / 02:54 - This test is developed in a completely different manner. We previously had the test statistic
30. 02:54 / 02:59 - being of the form the estimator minus the null value divided by the standard error of
31. 02:59 / 03:05 - the estimator, and that's very common, but the structure of the F-test will be different
32. 03:05 / 03:11 - and when we do talk about the structure it will sort of clarify where the name analysis
33. 03:11 / 03:16 - of variance is coming from. So just remember this test statistic is not going to look like
34. 03:16 / 03:22 - the ones that we've seen previously. So as always our first step is to state our null
35. 03:22 / 03:27 - and alternative hypotheses. In this case the null hypothesis is that there's no relationship

36. 03:27 / 03:33 - between our categorical explanatory variable X and our quantitative response variable Y.

37. 03:33 / 03:40 - We've already mentioned and seen the idea of comparing these populations that what we're

38. 03:40 / 03:46 - going to look at is comparing the means of Y in the populations defined by the values

39. 03:46 / 03:53 - of X. So we have k population means, one for each of the groups defined by X and if there

40. 03:53 / 04:00 - was no relationship between this X and Y then all those population means would be equal.

41. 04:00 / 04:05 - The null hypothesis of our F-test is just that, the null hypothesis is that mu_1 is

42. 04:05 / 04:11 - equal to mu_2 is equal to ... all the way up to mu_k. All of the population means are

43. 04:11 / 04:18 - equal. We could also say there is no relationship between our X, whatever it might be and our

44. 04:18 / 04:24 - Y. We only have one alternative hypothesis and that's the opposite of what we just said.

45. 04:24 / 04:30 - So either all the means are equal or it's not true that all the means are equal which

46. 04:30 / 04:37 - just says exactly that. The alternative hypothesis is that not all mu's are equal. We could also

47. 04:37 / 04:45 - say that there is a relationship between our variables X and Y. You have to be very careful,

48. 04:45 / 04:51 - all this tells us is that at least two of the means are different. It doesn't tell us

49. 04:51 / 04:57 - that all of the means are different. It just says not all means are equal. All we need

50. 04:57 / 05:03 - is two means to be different and us to statistically detect that and we will be able to reject

51. 05:03 / 05:08 - the null hypothesis. So then we'll often want to go in and figure out, well which means

52. 05:08 / 05:12 - are different? Are they all different? Are only two of them different? Or is there some

53. 05:12 / 05:20 - other combination of differences that we see. The alternative simply states that not all

54. 05:20 / 05:25 - of our means are equal. And it is not specific about in what way that they are different.

55. 05:25 / 05:30 - We could also say the alternative hypothesis by saying at least two means are different.

56. 05:30 / 05:36 - That emphasizes the fact that we don't know a whole lot when we reject the null hypothesis.

57. 05:36 / 05:42 - There's just something going on here. At least two means are different. But be very careful.

58. 05:42 / 05:48 - It is incorrect to say that the alternative is just take the not equal, mu_1 is not equal

59. 05:48 / 05:55 - to mu_2 is not equal to mu_3 is not equal to mu_4, if you only had k equals 4. That

60. 05:55 / 05:59 - statement is much stronger. That says all of the means are different from all of the

61. 05:59 / 06:05 - other means and that's a very strong statement which we are not proving with this test. So

62. 06:05 / 06:11 - we just have, all of the means are equal versus there's something going on here, not all of

63. 06:11 / 06:16 - the means to equal, at least two means are different. But it could be more than two means

64. 06:16 / 06:21 - are different, we will have to investigate to figure that out. So that's the null hypothesis

65. 06:21 / 06:29 - and alternative hypothesis. Always the same for the analysis of variance F-test. The second

66. 06:29 / 06:34 - part is going to be to obtain our data, check our conditions and summarize our data. We

67. 06:34 / 06:38 - have conditions that are fairly similar to what we've seen in other tests that we've

68. 06:38 / 06:42 - conducted. The first is that the samples that are drawn from each of the populations are

69. 06:42 / 06:48 - independent. We must have independent samples. Here, we have to be in one of the two scenarios

70. 06:48 / 06:55 - that we've discussed before. Each of the populations are normal or the populations are not normal

71. 06:55 / 07:00 - but the sample sizes are large enough. In addition we should have random samples or

72. 07:00 / 07:05 - samples that can at least be considered random. We can check normality in the same way; we

73. 07:05 / 07:11 - can use histograms and normal probability plots, looking for signs of extreme skewness

74. 07:11 / 07:18 - or outliers. The ANOVA F test has one additional condition and it's similar to what we saw
75. 07:18 / 07:24 - in the two-sample t-test but in this case we only have one option, we assume that the
76. 07:24 / 07:30 - populations all have the same standard deviation. For the t-test we have the option of handling
77. 07:30 / 07:35 - it when the populations do not have the same standard deviation but for the analysis of
78. 07:35 / 07:41 - variance we are assuming that all the populations have equal variation. They all have the same
79. 07:41 / 07:47 - standard deviation. We can check this condition in two ways. We can use a test similar to
80. 07:47 / 07:53 - the one that we used for the t-test or we can check this condition using sort of a quick
81. 07:53 / 07:58 - rule of thumb and that is that the ratio between the largest sample standard deviation and
82. 07:58 / 08:04 - the smallest sample standard deviation is less than two and if that's the case then
83. 08:04 / 08:09 - you can consider that it's reasonable to assume equal variances or at least if the variances
84. 08:09 / 08:15 - are unequal they're not unequal enough to cause any major problems with the overall
85. 08:15 / 08:21 - result. So let's talk a little bit before we go into the next step of calculating the
86. 08:21 / 08:26 - test statistic to try to give you a little bit of the logic behind how this test is working.
87. 08:26 / 08:31 - So what is the idea behind comparing more than two means? And the question we need to
88. 08:31 / 08:38 - answer is, are the differences among the sample means due to true differences among the mu's,
89. 08:38 / 08:43 - the population means? That's our alternative hypothesis. Or are they merely due to sampling
90. 08:43 / 09:05 - variability or random chance, that's our null hypothesis. What's the chance that we could
91. 09:05 / 09:11 - see the results that we see, if in fact these all came from the same population. How likely
92. 09:11 / 09:17 - is it that we could see the data we have? So we're going to look at two scenarios one
93. 09:17 / 09:23 - of each basically. And in each of these cases we want to discuss what's going on here and
94. 09:23 / 09:28 - whether we believe we should reject the null hypothesis or not. So in scenario 1, here
95. 09:28 / 09:34 - we see that we have four boxplots. There are differences between the means, which are denoted
96. 09:34 / 09:41 - by red dots, but there's so much variation inside each of the four groups, they overlap,
97. 09:41 / 09:47 - these boxplots overlap a lot. So we're trying to see what's the chance that we could have
98. 09:47 / 09:52 - seen these plots if indeed the true mean of the population from which these groups were
99. 09:52 / 09:58 - drawn are really the same. So this shows data with a large amount of variation within each of
100. 09:58 / 10:05 - the four groups and relatively small variation between them the means that we see there indicated
101. 10:05 / 10:11 - as red dots. So looking at how much the means vary between each other versus how much the
102. 10:11 / 10:17 - data varies within each group. So the boxplots again, they overlap a lot, and its it's pretty
103. 10:17 / 10:22 - reasonable, hopefully you believe that it's pretty reasonable that we could have gotten
104. 10:22 / 10:28 - these four boxplots if we had say a population mean that was somewhere in the middle of those
105. 10:28 / 10:33 - red dots. If we had a population mean that was somewhere in the middle of those red dots
106. 10:33 / 10:39 - and we had a fairly large variation in the population then when we sample, it's reasonable
107. 10:39 / 10:44 - that we could have gotten one sample where that mean was fairly low. But the idea is
108. 10:44 / 10:50 - that the variation that we see here between these plots is reasonable to what we would
109. 10:50 / 10:55 - have expected if we just took four random samples from the same population. In this
110. 10:55 / 11:02 - case we may not find enough evidence to reject the null hypothesis in that analysis of variance
111. 11:02 / 11:09 - test. In scenario 2, we see that we basically have the same means drawn, the red dots are

112. 11:09 / 11:16 - in the same place in both pictures, but now we have much less variation inside each group.

113. 11:16 / 11:22 - Now it would be very unlikely that the first boxplot would have come from a population

114. 11:22 / 11:27 - with a mean equal to, let's even say the other three came from the same population, the mean

115. 11:27 / 11:33 - is much larger if we consider those three boxplots compared to the first. So the boxplots

116. 11:33 / 11:38 - overlap very little. There is very small amount of spread within the groups. It would be very

117. 11:38 / 11:44 - hard to imagine that we are sampling from four groups that have equal population means

118. 11:44 / 11:49 - and so in this case what we would expect is that there is a difference in the population

119. 11:49 / 11:55 - means of these four groups and we would like to reject the null hypothesis. So these two

120. 11:55 / 12:01 - pictures illustrate that what we really need to do is take into account not just the variation

121. 12:01 / 12:06 - between means, which is how much the means differ from each other, but we need to somehow

122. 12:06 / 12:13 - incorporate how much variation we see in the samples individually. So what is the variation

123. 12:13 / 12:20 - within each of the groups? And so this brings us to our test statistic. If our conditions

124. 12:20 / 12:25 - are satisfied and we are able to continue with this test, then our F-statistic is going

125. 12:25 / 12:30 - to be a measure of the variation among the sample means, so how much to the red dots

126. 12:30 / 12:37 - differ from each other, over the variation within groups, how much do individuals within

127. 12:37 / 12:41 - a group vary from the mean of that group. And it is a measure that's very similar to

128. 12:41 / 12:46 - the standard deviation in the numerator and the denominator but we're not going to talk

129. 12:46 / 12:51 - about any further details of what's going on with the calculation of that test statistic.

130. 12:51 / 12:57 - So this statistic does follow an F-distribution, hence its name. It has two sets of degrees

131. 12:57 / 13:03 - of freedom. It has k -1 degrees of freedom in the numerator, based upon how many sample

132. 13:03 / 13:10 - means we're comparing and it has n-k degrees of freedom in the denominator based upon how

133. 13:10 / 13:16 - many observations we have in our dataset. So n is the total combined sample size of

134. 13:16 / 13:22 - all of our k groups and k is the number of groups being compared. And we will rely on

135. 13:22 / 13:29 - software entirely to calculate the test statistic and p-value for us for this test. So for the

136. 13:29 / 13:35 - p-value again we're going to let the software do this but what does the p-value represent?

137. 13:35 / 13:42 - We can go back to our idea of hypothesis testing and try to come up with what does the p-value

138. 13:42 / 13:47 - mean in this case. So the p-value up the ANOVA F-test is the probability of getting an F

139. 13:47 / 13:56 - statistic as large as we obtained or larger had the null hypothesis been true, all k population

140. 13:56 / 14:03 - means are equal. So it tells us, how surprising is it to find data like those observed assuming

141. 14:03 / 14:12 - that there is no difference among the population means. Our conclusion will be similar to what

142. 14:12 / 14:16 - we've done before. We're going to base it upon our p-value. A small p-value will tell

143. 14:16 / 14:22 - us that our data contain evidence against the null hypothesis and in this case that's

144. 14:22 / 14:26 - going to mean that the difference between the sample means are statistically significant.

145. 14:26 / 14:32 - Which says that they were unlikely to have happened by chance. And so we reject Ho and

146. 14:32 / 14:38 - conclude that there are some differences between the population means in our groups. If the

147. 14:38 / 14:42 - p-value is not small then the data do not provide enough evidence to reject the null

148. 14:42 / 14:48 - hypothesis and we continue to believe that it may be true. Doesn't mean that it is true

149. 14:48 / 14:51 - we are unable to say there is a difference between the population means.