Transcript

Video – 0426 Unit4B Case CQ Paired Samples - Non-Parametric Tests

01. 00:01 / 00:06 - Before we go through our example, we're going to talk really quickly about two nonparametric0
02. 00:06 / 00:13 - tests that we are going to conduct in the software. Parametric tests are generally defined
03. 00:13 / 00:20 - as tests which require assumptions about the distribution in the population. We might also
04. 00:20 / 00:25 - have requirements to use a certain approximation. We did for the proportion problem, where we
05. 00:25 / 00:32 - had to have n times p be at least 10 and n times one minus p be at least 10. Those requirements
06. 00:33 / 00:37 - are there to say that, if they are satisfied, in the approximations that we're going to do are
07. 00:37 / 00:42 - going to be adequate. So those methods that we've been discussing that require assumptions
08. 00:42 / 00:49 - about the population or requirements to use a certain approximation as the sampling distribution,
09. 00:50 / 00:55 - those methods are called parametric. When those assumptions are not valid, alternative methods
10. 00:55 / 01:02 - often exists to test similar hypotheses. These kinds of tests require minimal distributional
11. 01:02 / 01:09 - assumptions, if any, and are called non-parametric or distribution free tests. In some cases
12. 01:10 / 01:15 - you might see tests called exact tests. This is due to the fact that their method of calculating
13. 01:15 / 01:20 - p values or confidence intervals requires absolutely no mathematical approximation.
14. 01:20 / 01:26 - We're going to look at two nonparametric tests in the paired sample setting. The first is
15. 01:26 / 01:32 - the sign test. The sign test is a very general test to compare paired samples. The idea behind
16. 01:32 / 01:39 - the test is to find the sign of the differences. So we take the differences before and after,
17. 01:39 / 01:45 - pair 1 versus pair 2, and then we just ask ourselves is the difference positive for this
18. 01:45 / 01:51 - pair or is it negative for this pair. We throw away all other information and we just record whether
19. 01:51 / 01:58 - the difference is positive or negative. If the two paired measurements came from populations
20. 01:58 / 02:05 - with equal medians then theoretically we would expect half of the differences to be positive
21. 02:05 / 02:10 - and half of the differences to be negative and so if we count how many positive difference
22. 02:10 / 02:16 - is there are (or how many negative differences there are) we're going to have a binomial
23. 02:16 / 02:20 - distribution. The sampling distribution of our statistic is just a binomial where it's
24. 02:20 / 02:26 - 50-50. It's a 50-50 chance it's going to be positive and a 50-50 chance it's going to
25. 02:26 / 02:33 - be negative if the two populations have equal medians. So what we'll be doing is we will
26. 02:34 / 02:38 - be assuming that the two populations have equal medians, that's going to be our null
27. 02:38 / 02:44 - hypothesis verses our alternate hypothesis is going to be that the medians are not equal.
28. 02:44 / 02:49 - One-sided tests are also possible but we're just going to focus on the two-sided test
29. 02:49 / 02:56 - here. In step 2 we still require random sample or at least a sample that can be considered
30. 02:56 / 03:03 - random in context. But the sign test can actually be used for a very wide range of data as long
31. 03:03 / 03:09 - as we can determine the sign of the difference, we can use this test. So we can certainly
32. 03:09 / 03:13 - use it for quantitative measures which is going to be our case if we're doing a paired
33. 03:13 / 03:18 - t-test we decide we can't use the paired t-test, we're going to be in the case where we have
34. 03:18 / 03:25 - quantitative measurements. They could be continuous or discrete. With the t-test, sometimes
35. 03:26 / 03:31 - discrete variables can be a problem. If I'm counting how many meals I eat per day that's going
36. 03:31 / 03:36 - to be a pretty small number and it's not going to be representative of a normal distribution.
37. 03:36 / 03:41 - So if I have a small sample and a discrete random variable I'm often in trouble when
38. 03:41 / 03:47 - it comes to to the t-test but for the sign test the quantitative measures can be continuous
39. 03:47 / 03:53 - or discrete they don't need to be normally distributed at all. We can also use the Sign
40. 03:53 / 03:59 - test for categorical variables which are ordinal. This would be things like Likert scales, rating
41. 03:59 / 04:05 - scales, or letter grades where I could maybe have your grade before I give you some sort
42. 04:05 / 04:12 - of tutorial, I take your grade after, I can find out whether your grade improved. Even
43. 04:12 / 04:17 - better than that we can use the Sign test when all I can say is whether one pair is
44. 04:17 / 04:22 - larger smaller compared to the other pair. Some examples of this would be: is the left
45. 04:22 / 04:29 - arm more or less sunburned than the right arm in a study of sunscreen effectiveness,
46. 04:29 / 04:36 - was there an improvement in pain after treatment. Pain is very difficult to measure quantitatively

47. 04:36 / 04:43 - but it would be very easy to ask someone is it better, is it worse, is there no difference.
48. 04:43 / 04:47 - So the Sign test can be used in any of those scenarios and for this reason the test is
49. 04:47 / 04:52 - very widely applicable. We are mostly interested in the quantitative case because we're interested
50. 04:52 / 04:59 - in what are our alternative is when we cannot use the paired t-test. The data are summarized
51. 05:01 / 05:07 - by a test statistic which counts the number of positive or the number of negative differences.
52. 05:07 / 05:14 - Any ties are discarded. So any zero differences are thrown out and those observations don't
53. 05:14 / 05:19 - count at all. The p-values are going to be calculated using the binomial distribution.
54. 05:19 / 05:24 - I could actually ask you to calculate the p-values for this test yourself using only
55. 05:24 / 05:30 - the binomial calculator. We could actually calculate this probability fairly easily but
56. 05:30 / 05:34 - we will rely on the software to calculate the p-value for us. The decision is going
57. 05:34 / 05:39 - to be made the same as other tests we can word our conclusion in terms of the medians
58. 05:39 / 05:44 - which was our null and alternative hypothesis, or we can come all the way back and word our
59. 05:44 / 05:50 - conclusion in terms of the relationship between our categorical variable X and our response
60. 05:50 / 05:57 - variable Y, which now does not even have to be quantitative. Again we are in case CQ,
61. 05:58 / 06:05 - so that's what is of interest to us primarily. And the last test is the Wilcoxon Signed-Rand
62. 06:05 / 06:12 - test. This is a little more difficult to comprehend than the sign test so don't worry about it
63. 06:12 / 06:17 - if the little bit of theory that we talk about doesn't make much sense to you. But the idea
64. 06:17 / 06:23 - here is we're going to assume that the distribution of the differences is symmetric. That's our
65. 06:23 / 06:30 - only assumption. This test is based upon the rank of the absolute difference and then we
66. 06:30 / 06:36 - flag those differences with, were they positive or negative. If the two populations were the
67. 06:36 / 06:43 - same, we would expect the ranks to be evenly distributed on the positive side and the negative
68. 06:43 / 06:50 - side. And so that's how the test works. If I see all of the positive ranks are huge and
69. 06:52 / 06:58 - all the negative ranks are small then that tends to tell me that there's something going
70. 06:58 / 07:04 - on here. The hypotheses here are that the means or medians are equal. So we can either
71. 07:04 / 07:09 - word our hypotheses in terms of the means being equal or in terms of the medians being
72. 07:09 / 07:15 - equal and the alternative hypothesis is that they are not equal. One-sided tests are still
73. 07:15 / 07:22 - possible but again we're going to focus on two-sided tests here. We're going to obtain
74. 07:22 / 07:27 - our data, check our conditions, and summarize the data. In this case we do assume that the
75. 07:27 / 07:34 - differences are symmetric. We still need a random sample. We can use to Wilcoxon Signed-Rand
76. 07:34 / 07:40 - test for quantitative data or ordinal data but not binary day. So the Sign test is the
77. 07:40 / 07:44 - only one of the three that we are discussing that can be used in the case where I just
78. 07:44 / 07:51 - know bigger or smaller but the Wilcoxon Signed-Rand test could be used for Likert scales or letter
79. 07:51 / 07:57 - grades. The data are summarized by a test statistic based upon the ranks. So let's talk
80. 07:57 / 08:02 - just a little bit about how this works. To rank the pairs we find the differences exactly
81. 08:02 / 08:09 - as we would in the paired t-test scenario. We take the absolute value of the differences
82. 08:09 / 08:16 - and rank the absolute value of the differences. Rank each pair from one, which is the smallest
83. 08:16 / 08:23 - non-zero difference, to m, which is the largest non-zero difference. Where m is how many
84. 08:24 / 08:30 - non-zero pairs we have. It's not actually very difficult to do this test by hand in terms of calculating
85. 08:30 / 08:36 - the test statistic but it is a little bit tedious. We have to do a lot of sorting, a
86. 08:36 / 08:42 - lot of ranking and then some adding. So we then determine which of the ranks came from
87. 08:42 / 08:47 - the positive differences and add those ranks up. We will NOT be conducting this test by
88. 08:47 / 08:52 - hand at all, again we just want to explain a little bit of the logic behind what's going
89. 08:52 / 08:59 - on in this test. The ranking procedure really does help to eliminate any horrible effect
90. 09:00 / 09:05 - due to outliers, since we're ranking them we're not so much worried about the size of
91. 09:05 / 09:12 - the difference just whether it was the largest or smallest difference. P-values are calculated
92. 09:12 / 09:17 - using a distribution that's actually specific to this test, developed by Wilcoxon when he
93. 09:17 / 09:23 - developed this method. So it's an it's an exact table that relies on the sample size
94. 09:23 / 09:30 - and so forth. We're going to rely on software to obtain our p-value for us and again our
95. 09:30 / 09:36 - conclusions are made exactly the same as we discussed for the sign test. Here we can word
96. 09:36 / 09:41 - our conclusion in terms of the means or the medians or we could word our conclusion in
97. 09:41 / 09:48 - terms of the categorical explanatory variable X and the response variable Y. The Sign test

98. 09:49 / 09:54 - tends to have much lower power than the paired t-test or the Wilcoxon Signed-Rand test. In
99. 09:54 / 10:00 - other words the sign test less chance of being able to detect a true difference than the
100. 10:00 / 10:06 - other tests. It is however more widely applicable since we can use it in the case where we only
101. 10:06 / 10:12 - know better or worse, or smaller or larger, for each pair. The other two methods are not
102. 10:12 / 10:18 - applicable in that case. The Wilcoxon Signed-Rand test is comparable to the paired t-test in
103. 10:18 / 10:24 - power. So it can actually be used pretty well almost all of the time and can even perform
104. 10:24 / 10:29 - better than the paired t-test under certain conditions. In particular this can occur when
105. 10:29 / 10:35 - there are a a few very large outliers as these outliers can greatly affect our estimate of
106. 10:35 / 10:40 - the standard error in the paired t-test, since it is based upon the sample standard deviation
107. 10:40 / 10:47 - which is highly affected by outliers. Both the Sign test and the Wilcoxon Signed-Rand
108. 10:47 / 10:53 - test can also be used for one sample and we'll see how to do that in software. In that case
109. 10:53 / 11:00 - instead of having pair 1 minus pair 2, we have observation minus the null value for
110. 11:00 / 11:06 - every person in our sample. So the differences are calculated, instead of being pair 1 minus
111. 11:06 / 11:12 - pair 2, they are calculated as the observed value minus the null value. So that's our
112. 11:12 / 11:18 - theoretical discussion of what we're going to be talking about in the example but what
113. 11:18 / 11:23 - I'm more excited about is getting into the example and talking about what it means.