Slide 1 – Title Slide

Hello and welcome to Week 8, Part 5 of EGM101: Non-parametric tests. In this lesson, we'll see how even if our data don't fit the assumptions required of parametric tests, we can still do some hypothesis testing.

Slide 2 – When do I use non-parametric tests?

When we introduced parametric tests earlier this week, we made the following assumptions about the data we were using: first, that the data are approximately normally-distributed, or that the sample size was large enough that we could get by with the central limit theorem, and second, that we have continuous data.

This isn't always going to be the case, though. For starters, we might have ordinal data, our data might be heavily skewed or otherwise non-normal, and we might have to deal with small sample sizes. If any of these things are true, we won't be able to use parametric tests reliably (or at all).

So, when our data don't fit the assumptions of the parametric tests, we use non-parametric tests. One of the big reasons for using non-parametric tests is because we can still get reliable results using non-normal data; but also, we will see some types of tests that allow us to use ordinal data.

One other thing to note is that with non-parametric tests, we normally use the median, denoted using the lowercase Greek letter eta, rather than the mean, to calculate the test statistic.

Slide 3 – Wilcoxon signed-rank tests

The first example of a non-parametric test that we will look at are actually a pair of tests, called Wilcoxon signed-rank tests. We use the Wilcoxon signed-rank tests when we have data that are continuous and symmetric, though not necessarily normally-distributed.

We'll introduce this further on the next slide, but the test statistic we use for the Wilcoxon signed-rank is the sum of the positive difference ranks, W+. Alternative formulations use the sum of the negative difference ranks, W-, or the minimum of the two values.

If we have a large enough sample size – if n times n +1, divided by 2 is greater than 20, the W-statistics are approximately normally distributed, with a mean and a variance given by the formulas on the slide here. We can then calculate the z-score and use the standard normal distribution to calculate the critical value for the test. If the sample size is small, though, we can't use the normal approximation; in that case, we would need to use a lookup table to find the critical value of our test statistic.

The formulas for the mean and variance shown here assume that there are no ties in the data – if we do have ties, the variance is slightly more complicated. If that's the case, for each group of tied observations of size t, we subtract t cubed minus t divided by 48 from the variance, to account for the reduction in variance caused by tied values.

Slide 4 – Signed ranks

Now that we've introduced the test procedure, we need to discuss the test statistics W+ and W- in more detail. Similar to Spearman's rank correlation, the Wilcoxon tests use ranks, rather than the actual values of the data. Specifically, we use the ranks of differences – for the one sample Wilcoxon, we use the differences between the data and the mean or median that we're testing against; for the two-sample test, we use the difference between matched pairs of the data.

Starting with the values of two variables, and y, we take the differences between each pair of values. Then, we rank the absolute values of the differences, starting from the smallest value to the largest value. So, the smallest absolute difference gets a rank of 1, and we work our way up from there.

And, in this example, you can see that we have a number of tied differences – if we do have ties, then we assign the average rank for each group of tied values. Here, we have two ones, so we assign an average rank of 1.5 to these; we also have three sixes, which get an average rank of 6; and two sevens, which get an average rank of 8.5.

Finally, we use the sign of the difference to sum the ranks of the positive differences, though for this example, the W+ and W- statistics have the same value: 27.5. From here, because we have such a small sample size, we would use a lookup table to find the critical value for our chosen alpha and sample size.

The final question is, what do we do if we have differences of zero? The Wilcoxon test is not really made to handle these cases, so the classical answer is to just ignore them. But, this is not the only option – we could also include them when assigning ranks, and exclude them from calculating the test statistic.

Slide 5 – One-sample Wilcoxon

Now, we'll look at a one-sample Wilcoxon example, to see how this works in practice. Let's say that in a study of salmon in a particular river, we went out and caught only 7 fish. We still want to be able to answer the question, is the population median equal to 150 cm? Because of the small sample size, we can't reliably use the student's t-test or another parametric test – instead, we have to use a non-parametric test.

The hypotheses for the test are then as follows: the null hypothesis is that the population median is equal to the hypothesized value of 150 cm; the alternative hypothesis is that the value is less than 150 cm.

To proceed with the test, we first subtract the hypothesized median value of 150 cm from each value of our dataset. Then, we take the absolute values of those differences, and rank the values. After that, we calculate the test statistic as the sum of the positive values – for this dataset, we have a value of 13.

Now, our sample size is small, which means that we can't just use the normal approximation that we introduced earlier. Instead, we have to use a lookup table of the exact values of the distribution, based on the values of n and alpha. For $n = 7$ and $\alpha = 0.05$, the critical value is equal to 4. Because the value of our test statistic, 13, is greater than this critical value, 4, we fail to reject the null hypothesis –

our observations do not provide sufficient evidence to conclude that the population median is less than 150 cm.

Slide 6 – Two-sample Wilcoxon

The two-sample Wilcoxon signed-rank test is the non-parametric counterpart to the paired sample t-test. I won't go through an example here, but I will outline the basic steps of the test.

First, we take the differences of the paired samples, as we saw earlier when I introduced signed ranks. Then, we rank the absolute values of those differences, and calculate the test statistic by summing the ranks of the positive differences.

Finally, if our sample size was big enough we could use the normal approximation, or we would need to use a lookup table to find the critical value of the test statistic, based on the sample size and chosen significance level.

Slide 7 – Mann-Whitney U-Test

The final non-parametric test that we'll learn about in this lesson is the Mann-Whitney U-test, which is a non-parametric counterpart to the independent samples t-test. The question that we're attempting to answer with the Mann-Whitney U-test is: is there a difference in the rank sum between two groups?

To use the Mann-Whitney test, we have a few requirements: first, we require random, independent samples – just like with the independent samples t-test. We also need at least one sample to have a sample size larger than 5. And finally, we need the dependent variable to be either ordinal or numeric, since we have to rank the values. This means, though, that we do not have to have numeric data for the Mann-Whitney U-test – it can also be used for ordinal data.

The test proceeds like the others that we have seen – first, we calculate the test statistic, U; then, we either look up the critical value in a lookup table, or, if we have a large enough sample size, we can use a normal approximation for the U statistic.

Slide 8 – Mann-Whitney U-Test

Let's say that we have two small samples of some data, each drawn from a different population: some red data, drawn from population 1; and some blue data, drawn from population 2. We are interested in determining if the two populations have different median values.

Our hypotheses for the Mann-Whitney U-test are then: the null hypothesis is that the two population medians are equal to each other; the alternative hypothesis is that the two population medians are not equal to each other.

The test proceeds as follows: first, we have to rank all of the values, from smallest to largest. The table here shows the ranks for each value, with the color of each cell corresponding to whether the value is part of the red sample or the blue sample.

For each variable, we then sum the ranks. For the red variable, the sum of the ranks is equal to 44; for the blue variable, it is 22.

Next, we have to calculate the two test statistics, U1 and U2. U1 is calculated according to the formula shown here – n1 and n2 are the sample sizes of the two variables, and R2 is the sum of the ranks of the blue variable. Plugging in our values, we get a value of 23 for U1. U2 is calculated using a similar formula – again, plugging in our values, we get a value of 7 for U2.

If the total sample size, n, is greater than 20, the U-statistic is approximately normally-distributed, with an expected value and standard error given by the formulas shown here. We can then calculate the z-score using the minimum of U1 and U2 – plugging in the different values for n1 and n2 here, we calculate a z-score of -1.461.

Finally, from the standard normal distribution, the p-value corresponding to a z-score of -1.461 is equal to 0.144 – because this is greater than our significance level of 0.05, we do not reject the null hypothesis – our observations do not provide sufficient evidence to conclude that the two populations have a different median value.

Slide 9 – Summary

In this lesson, we've seen how non-parametric tests are tests where we don't make any assumptions about the distribution of the population. Because of this, we can use non-parametric tests with a wider range of data, including data that are non-normally distributed, and even non-numeric data in some cases.

We also saw some non-parametric counterparts to parametric tests, such as the one-sample Wilcoxon signed rank test as a counter part to the one-sample t-test; the two-sample Wilcoxon signed rank test as a counterpart to the paired sample t-test; and the Mann-Whitney U-test as a counterpart to the independent samples t-test.

As before, focus more on the question of "why" you would use a particular non-parametric test, rather than the formulas. Most of the time, you will end up using some software package such as SPSS to actually perform the test. Instead, you should try to understand when to apply which test, based on the data that you are using.

Slide 10 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Weiss, Chapters 9.6, 10.4, and 10.6. I've also included links to an article about parametric and non-parametric tests, and when to use them, as well as links to two YouTube videos that cover the same topic.

That's all for this lesson – I hope you found it interesting, and if you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!