# Chapter 6: Non-parametric models

## Oliver Twisted

## Please Sir, can I have some more … jonck?

Jonckheere's test is based on the simple, but elegant, idea of taking a score in a particular condition and counting how many scores in subsequent conditions are smaller than that score. So, the first step is to order your groups in the way that you expect your medians to change. If we take the soya example, then we expect sperm counts to be highest in the no soya meals condition, and then decrease in the following order: 1 meal per week, 4 meals per week, 7 meals per week. So, we start with the no meals per week condition, and we take the first score and ask 'How many scores in the next condition are bigger than this score?'.

You'll find that this is easy to work out if you arrange your data in ascending order in each condition. Table 1 shows a convenient way to lay out the data. Note that the sperm counts have been ordered in ascending order and the groups have been ordered in the way that we expect our medians to decrease. So, starting with the first score in the no soya meals group (this score is 0.35), we look at the scores in the next condition (1 soya meal) and count how many are greater than 0.35. It turns out that all 19 of the 20 scores are greater than this value, so we place the value of 19 in the appropriate column and move onto the next score (0.58) and do the same. When we've done this for all of the scores in the no meals group, we go back to the first score (0.35) again, but this time count how many scores are bigger in the next but one condition (the 4 soya meal condition). It turns out that 18 scores are bigger so we register this in the appropriate column and move onto the next score (0.58) and do the same until we've done all of the scores in the 7 meals group. We basically repeat this process until we've compared the first group with all subsequent groups.

At this stage we move onto the next group (the 1 soya meal condition). Again, we start with the first score (0.33) and count how many scores are bigger than this value in the subsequent group (the 4 meals group). In this case there all 20 scores are bigger than 0.33, so we register this in the table and move onto the next score (0.36). Again, we repeat this for all scores and then go back to the beginning of this group (i.e. back to the first score of 0.33) and repeat the process until this category has been compared with all subsequent categories. When all categories have been compared with all subsequent categories in this way, we simply add up the counts as I have done in the table. These sums of counts are denoted by $U_{ij}$.

Table 1: Data to show Jonckheere's test for the soya example

| No Soya Meals | | | | 1 Soya Meal | | | 4 Soya Meals | | 7 Soya Meals |
|---|---|---|---|---|---|---|---|---|---|
| Sperm | How Many Scores Are Bigger in the… | | | Sperm | How Many Scores Are Bigger in the… | | Sperm | How Many Scores Are Bigger in the… | Sperm |
| | 1 Meal | 4 Meals | 7 Meals | | 4 Meals | 7 Meals | | 7 Meals | |
| 0.35 | 19 | 20 | 18 | 0.33 | 20 | 18 | 0.40 | 18 | 0.31 |
| 0.58 | 18 | 19 | 16 | 0.36 | 20 | 18 | 0.60 | 16 | 0.32 |
| 0.88 | 18 | 18 | 13 | 0.63 | 18 | 16 | 0.96 | 13 | 0.56 |
| 0.92 | 15 | 18 | 13 | 0.64 | 18 | 16 | 1.20 | 12 | 0.57 |
| 1.22 | 15 | 15 | 12 | 0.77 | 18 | 15 | 1.31 | 11 | 0.71 |
| 1.51 | 15 | 14 | 7 | 1.53 | 14 | 7 | 1.35 | 9 | 0.81 |
| 1.52 | 15 | 14 | 7 | 1.62 | 14 | 7 | 1.68 | 7 | 0.87 |
| 1.57 | 14 | 14 | 7 | 1.71 | 13 | 7 | 1.83 | 7 | 1.18 |
| 2.43 | 10 | 11 | 6 | 1.94 | 12 | 7 | 2.10 | 6 | 1.25 |
| 2.79 | 9 | 11 | 4 | 2.48 | 11 | 6 | 2.93 | 3 | 1.33 |
| 3.40 | 9 | 6 | 1 | 2.71 | 11 | 5 | 2.96 | 3 | 1.34 |
| 4.52 | 8 | 5 | 0 | 4.12 | 6 | 0 | 3.00 | 3 | 1.49 |
| 4.72 | 8 | 5 | 0 | 5.65 | 4 | 0 | 3.09 | 2 | 1.50 |
| 6.90 | 6 | 3 | 0 | 6.76 | 3 | 0 | 3.36 | 1 | 2.09 |
| 7.58 | 4 | 3 | 0 | 7.08 | 3 | 0 | 4.34 | 0 | 2.70 |
| 7.78 | 4 | 3 | 0 | 7.26 | 3 | 0 | 5.81 | 0 | 2.75 |
| 9.62 | 2 | 3 | 0 | 7.92 | 3 | 0 | 5.94 | 0 | 2.83 |
| 10.05 | 2 | 3 | 0 | 8.04 | 3 | 0 | 10.16 | 0 | 3.07 |
| 10.32 | 2 | 2 | 0 | 12.10 | 1 | 0 | 10.98 | 0 | 3.28 |
| 21.08 | 0 | 0 | 0 | 18.47 | 0 | 0 | 18.21 | 0 | 4.11 |
| Total $(U_{ij})$ | 193 | 187 | 104 | | 195 | 122 | | 111 | |

The test statistic, *J*, is simply the sum of these counts:

$$J = \sum_{i<j}^{k} U_{ij}$$

which for these data is simply:

$$J = \sum_{i<j}^{k} U_{ij} = 193 + 187 + 104 + 195 + 122 + 111$$
$$= 912$$

For small samples there are specific tables to look up critical values of *J*; however, when samples are large (anything bigger than about 8 per group would be large in this case) the sampling distribution of *J* becomes normal with a mean and standard deviation of:

$$\bar{J} = \frac{N^2 - \sum n_k^2}{4}$$

$$\sigma_j = \sqrt{\frac{1}{72}\left[N^2(2N+3) - \sum n_k^2(2n_k+3)\right]}$$

in which $N$ is simply the total sample size (in this case 80) and $n_k$ is simply the sample size of group $k$ (in each case in this example this will be 20 because we have equal sample sizes). So, we just square each group's sample size and add them up, then subtract this value from the total sample size squared. We then divide the result by 4. Therefore, we can calculate the mean for these data as:

$$\bar{J} = \frac{80^2 - \left(20^2 + 20^2 + 20^2 + 20^2\right)}{4}$$
$$= \frac{4800}{4}$$
$$= 1200$$

The standard deviation can similarly be calculated using the sample sizes of each group and the total sample size:

$$= \sqrt{\frac{1}{72}\left[80^2(2\times 80+3) - 20^2(2\times 20+3) + 20^2(2\times 20+3) + 20^2(2\times 20+3) + 20^2(2\times 20+3)\right]}$$

$$= \sqrt{\frac{1}{72}(1043200 - 68800)}$$

$$= 116.33$$

We can use the mean and standard deviation to convert $J$ to a $z$-score (see Chapter 1) using the standard formulae:

$$z = \frac{x - \bar{x}}{s} = \frac{J - \bar{J}}{\sigma_j} = \frac{1488 - 1200}{116.33} = 2.476$$

This $z$ can then be evaluated using the critical values in the Appendix of the book. This test is always one-tailed because we have predicted a trend to use the test. So we're looking at $z$ being above 1.65 (when ignoring the sign) to be significant. In fact, the sign of the test tells us whether the medians ascend across the groups (a positive $z$) or descend across the groups (a negative $z$) as they do in this example!

### Does it matter how I order my groups?

I have just shown how to use the test when the groups are ordered by descending medians (i.e. we expect sperm counts to be highest in the no soya meals condition, and then decrease in the following order: 1 meal per week, 4 meals per week, 7 meals per week; so we ordered the groups: No soya, I meal, 4 meals and 7 meals). Certain books will tell you to order the groups in ascending order (start with the group that you expect to have the lowest median). For the soya data this would mean arranging the groups in the opposite order to how I did above; that is, 7 meals, 4 meals, 1 meal and no meals. The purpose of this section is to show you what happens if we order the groups the opposite way around!

The process is similar to that used in the Appendix, only now we start with start with the 7 meals per week condition, and we take the first score and ask 'How many scores in the next condition are bigger than this score?' You'll find that this is easy to work out if you arrange your data in ascending order in each condition. Table 2 shows a convenient way to lay out the data. Note that the sperm counts have been ordered in ascending order and the groups have been ordered in the way that we expect our medians to increase. So, starting with the first score in the 7 soya meals group (this score is 0.31), we look at the scores in the next condition (4 soya meals) and count how many are greater than 0.31. It turns out that all 20 scores are greater than this value, so we place the value of 20 in the appropriate column and move on to the next score (0.32) and do the same. When we've done this for all of the scores in the 7 meals group, we go back to the first score (0.31) again, but this time count how many scores are bigger in the next but one condition (the 1 soya meal condition). It turns out that all 20 scores are bigger so we register this in the appropriate column and move on to the next score (0.32) and do the same until we've done all of the scores in the 7 meals group. We basically repeat this process until we've compared the first group to all subsequent groups.

Table 2: Data to show Jonckheere's test for the soya example, groups in descending order

| 7 Soya Meals | | | | 4 Soya Meals | | | 1 Soya Meal | | No Soya Meals |
|---|---|---|---|---|---|---|---|---|---|
| Sperm | How Many Scores are Bigger in the… | | | Sperm | How Many Scores are Bigger in the… | | Sperm | How Many Scores are Bigger in the… | Sperm |
| | 4 Meals | 1 Meal | No Meals | | 1 Meal | No Meals | | No Meals | |
| 0.31 | 20 | 20 | 20 | 0.40 | 18 | 19 | 0.33 | 20 | 0.35 |
| 0.32 | 20 | 20 | 20 | 0.60 | 18 | 18 | 0.36 | 19 | 0.58 |
| 0.56 | 19 | 18 | 19 | 0.96 | 15 | 16 | 0.63 | 18 | 0.88 |
| 0.57 | 19 | 18 | 19 | 1.20 | 15 | 16 | 0.64 | 18 | 0.92 |
| 0.71 | 18 | 16 | 18 | 1.31 | 15 | 15 | 0.77 | 18 | 1.22 |
| 0.81 | 18 | 15 | 18 | 1.35 | 15 | 15 | 1.53 | 13 | 1.51 |
| 0.87 | 18 | 15 | 18 | 1.68 | 13 | 12 | 1.62 | 12 | 1.52 |
| 1.18 | 17 | 15 | 16 | 1.83 | 12 | 12 | 1.71 | 12 | 1.57 |
| 1.25 | 16 | 15 | 15 | 2.10 | 11 | 12 | 1.94 | 12 | 2.43 |
| 1.33 | 15 | 15 | 15 | 2.93 | 9 | 10 | 2.48 | 11 | 2.79 |
| 1.34 | 15 | 15 | 15 | 2.96 | 9 | 10 | 2.71 | 11 | 3.40 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1.49 | 14 | 15 | 15 | 3.00 | 9 | 10 | 4.12 | 9 | 4.52 |
| 1.50 | 14 | 15 | 15 | 3.09 | 9 | 10 | 5.65 | 7 | 4.72 |
| 2.09 | 12 | 11 | 12 | 3.36 | 9 | 10 | 6.76 | 7 | 6.90 |
| 2.70 | 11 | 10 | 11 | 4.34 | 8 | 9 | 7.08 | 6 | 7.58 |
| 2.75 | 11 | 9 | 11 | 5.81 | 7 | 7 | 7.26 | 6 | 7.78 |
| 2.83 | 11 | 9 | 10 | 5.94 | 7 | 7 | 7.92 | 4 | 9.62 |
| 3.07 | 8 | 9 | 10 | 10.16 | 2 | 2 | 8.04 | 4 | 10.05 |
| 3.28 | 7 | 9 | 10 | 10.98 | 2 | 1 | 12.10 | 1 | 10.32 |
| 4.11 | 6 | 9 | 9 | 18.21 | 1 | 1 | 18.47 | 1 | 21.08 |
| **Total ($U_{ij}$)** | **289** | **278** | **296** | | **204** | **212** | | **209** | |

At this stage we move on to the next group (the 4 soya meals). Again, we start with the first score (0.40) and count how many scores are bigger than this value in the subsequent group (the 1 meal group). In this case there are 18 scores bigger than 0.40, so we register this in the table and move on to the next score (0.60). Again, we repeat this for all scores and then go back to the beginning of this group (i.e. back to the first score of 0.40) and repeat the process until this category has been compared to all subsequent categories. When all categories have been compared to all subsequent categories in this way, we simply add up the counts as I have done in the table. These sums of counts are denoted by $U_{ij}$. As before, test statistic $J$ is simply the sum of these counts:

$$J = \sum_{i<j}^{k} U_{ij}$$

which for these data is simply:

$$J = \sum_{i<j}^{k} U_{ij} = 289 + 278 + 296 + 204 + 212 + 209$$
$$= 1488$$

As I said earlier, for small samples there are specific tables to look up critical values of $J$; however, when samples are large (anything bigger than about 8 per group would be large in this case) the sampling distribution of $J$ becomes normal with a mean and standard deviation of:

$$\bar{J} = \frac{N^2 - \sum n_k^2}{4}$$

$$\sigma_j = \sqrt{\frac{1}{72}\left[N^2(2N+3) - \sum n_k^2(2n_k+3)\right]}$$

in which $N$ is simply the total sample size (in this case 80) and $n_k$ is simply the sample size of group $k$ (in each case in this example this will be 20 because we have equal sample sizes). So, we just square each group's sample size and add them up, then subtract this value from the total sample size squared. We then divide the result by 4. Therefore, we can calculate the mean for these data as we did earlier:

$$\bar{J} = \frac{80^2 - \left(20^2 + 20^2 + 20^2 + 20^2\right)}{4}$$
$$= \frac{4800}{4}$$
$$= 1200$$

The standard deviation can similarly be calculated using the sample sizes of each group and the total sample size (again this is the same as earlier):

$$= \sqrt{\frac{1}{72}\left[80\left(2\times80+3\right) - 20\left(2\times20+3\right) + 20\left(2\times20+3\right) + 20\left(2\times20+3\right) + 20\left(2\times20+3\right)\right]}$$

$$= \sqrt{\frac{1}{72}\left(1043200 - 68800\right)}$$

$$= 116.33$$

We can use the mean and standard deviation to convert *J* to a *z*-score (see Chapter 1) using the standard formulae. The mean and standard deviation are the same as before, but we now have a different test statistic (it is 1488 rather than 912). So, let's see what happens when we plug this new test statistic into the equation:

$$z = \frac{x - \bar{x}}{s} = \frac{J - \bar{J}}{\sigma_j} = \frac{1488 - 1200}{116.33} = 2.476$$

Note that the *z*-score is the same value as when we ordered the groups in descending order, except that it now has a positive value rather than a negative one! This goes to prove what I wrote earlier: the sign of the test tells us whether the medians ascend across the groups (a positive *z*) or descend across the groups (a negative *z*)! Earlier we ordered the groups in descending order and so got a negative *z*, and now we ordered them in ascending order and so got a positive *z*.