Laboratory Tutorial 1-7: Non-parametric tests

In this laboratory tutorial you will:

1. Learn the basics of how non-parametric tests work and when to use them
2. Compare the output of Spearman's rho to that of Pearson's r
3. Compute Mann-Whitney U-test as an alternative to independent samples t-test
4. Compute Wilcoxon Signed Ranks test as an alternative to paired samples t-test

This is a mandatory tutorial. In order to pass the coursework, you must achieve a score of 50% or higher on the associated Blackboard quiz (Lab Quiz 1-7).

***Note: We strongly suggest you do not start the Lab Quiz for this tutorial before you have all your answers ready.**

Preamble: Non-parametric statistics

In the previous lab tutorial, you applied statistical tests of correlation (Pearson r) and difference (t-test) to answer a range of questions. All of these tests had one thing in common: they were parametric tests. Whilst parametric tests tend to be the most powerful (sensitive) tests available to a statistician and normally the first choice, they carry some stringent requirements, most notably:

- The underlying distribution of scores in the population being sampled should be (approximately) normal i.e. not strongly skewed, bi-modal etc.
- The data level of the dependent variable should be quantitative (equal interval or ratio) level, not ordinal or nominal

Non-parametric statistics have less stringent requirements because they do not make assumptions about the distribution of scores. Essentially, they transform raw scores (or differences in the case of the Wilcoxon test) into ranks (i.e. lowest score = 1, next lowest score = 2, etc.). For this reason, they are sometimes referred to as distribution-free statistics and as the rank-transformation operation renders the tests blind to the actual distribution of the real values.

Both correlation and difference tests have their own additional, specific requirements. For instance, for parametric correlation (Pearson's r) the variability in scores should be homoscedastic i.e. the variability in X scores should be more or less equal for all levels of Y. For instance, if you can see a "carrot" shaped distribution (as opposed to the normal cigar shape) in the scatter-plot of X and Y, then this assumption has been violated and you should consider using the non-parametric **Spearman's rho** or **rank** correlation test instead. This avoids problems with the distribution by converting values to ranks and then testing the co-variance of these rank scores, rather than the original values.

The paired t-tests works by examining the distribution of differences between two related measures (e.g. before and after scores) taken from the same set of cases. For this reason, in addition to the general parametric requirements, these difference values must also be normally distributed. The non-parametric alternative for related data is called the **Wilcoxon Signed Ranks** test. Like Spearman's, this works by converting the values for each variable to a rank distribution. The test is then computed using the differences between the two rank scores.

Finally, an alternative to the independent t-test is the **Mann-Whitney U**-test. This test involves replacing original values with ranks across both groups. Therefore, any difference between the groups shows up as a difference in the sum of their rank scores.

All these assumptions can be checked using either domain knowledge (i.e. is the DV ordinal or quantitative scale?), graphs and descriptive statistics as covered in Lab 6. If there is ever any doubt over the right choice, you should compute both parametric and non-parametric statistics and accept the least significant result (i.e. the one with the higher p-value). Generally, this is the non-parametric result because as they rely on the median, rather than the mean, these tests tend to be less affected by extreme cases (outliers). The table below summarises the non-parametric equivalents of the parametric tests used so far:

Parametric	***Non-parametric***
Pearson r	**Spearman's rho**
Independent t-test	**Mann-Whitney U**
Paired t-test	**Wilcoxon Signed Ranks**

In summary, non-parametric tests are distribution-free statistics that work by converting values to rank scores. They are frequently used when data is either not normally distributed and/or involves ordinal level measures. An important thing to remember is that they are more conservative than parametric tests. This means that a non-parametric test is less likely than a parametric test to return a false positive caused, for instance, by extreme values or differences in the size or shape of the distribution. This also means, of course, that there is a greater risk of rejecting a hypothesis that is actually true (false negative), although most statisticians would argue that it is always better to err on the side of caution. If you aren't sure which one to use, a good tip is to compute both the parametric and the non-parametric test and compare their significance (p) values. If they differ wildly (e.g. $p = 0.03$ vs $p = 0.001$) then it is best to take the least significant of the results.

In the remainder of this tutorial you will practice using each of these three non-parametric tests. The data file you need for the following three exercises is "Sleep3ED.sav". Before attempting these exercises, be sure you have already completed Laboratory Tutorial 1-5.

Exercise 1: Computing Spearman's rho in SPSS

Computing a Spearman rho correlation test is very easy as it involves almost the same procedure as that for Pearson's r. The only extra thing you just need to check the box labelled "Spearmans" in the "Bivariate Correlations" dialogue.

In this exercise, we are interested to know if depression, anxiety or daytime sleepiness levels are related to smoking behaviour. The variable "smokenum" contains the responses to the question "How many cigarettes do you smoke per day?". You'll notice that the number of non-zero cases is quite small (n=35) as the majority of of the sample (87%) declared themselves to be non-smokers, so the effect of any violation of parametric assumptions when computing correlations, with this small sample, could quite severe. In fact, this is a good reason in itself why we should choose a non-parametric correlation test if we are in any way unsure about the parametric assumptions being met.

To test these assumptions, let's run a descriptive analysis comprising the full range of averages and measures of dispersion. When you have done this, interpret the results in order to answer Q1.

Q1: Which of the following is NOT a valid reason why "smokenum" fails to meet the parametric assumptions required for a Pearson correlation test?

- **Distribution is skewed**
- **Distribution is kurtotic**
- **Data level is not scale**

We will use the variables "depress", "anxiety" and "totsas" as measures of depression, anxiety and sleepiness respectively. If you generate the descriptive statistics, you'll see that all of these variables are relatively normally distributed, compared with smokenum, although certainly not perfectly normal. One could argue that as these (psychometric) scales measure psychological, not physical constructs, they are actually more ordinal than scale in level – does a depress score of 20 make the person twice as depressed as one with a score of 10 (can you quantify something like depression or anxiety)? On the other hand, these scores are certainly not ordered categories in the sense that say educ or agegp3 are.

Let's examine the practical effect of these different parametric violations by computing and comparing both Pearson and Spearman correlation coefficients between "smokenum" and the other three psychometric variables.

Run a correlation test with all four variables in the variables box. Make sure both Pearson and Spearman boxes are checked. When you interpret the results, you should be looking at the relative extent to which each correlation coefficient changes from one test to the other.

Q2: Overall, which one of the four variables was most sensitive to the choice of Pearson or Spearman?

Q3: For which pair of variables did the correlation coefficient differ the most between tests?

- totsas and anxiety
- anxiety and depress
- smokenum and anxiety
- smokenum and depress

Based on these results, think about the relative impact of skew, kurtosis and data level respectively on the decision to adopt a parametric vs non-parametric test. Feel free to discuss your thoughts with a tutor.

Exercise 2: Computing Mann-Whitney U in SPSS

The next question to answer is *“Is there a difference in quality of sleep reported by men and women?”*

Like many self-report “Likert” type scales, it is questionable whether this is indeed a quantitative (scale) measure at all. Say person A rates their sleep quality as excellent (coded as 6) and person B rates theirs as fair (coded as 3) – it is one thing to say that person A has better sleep quality, but is their sleep truly twice as good as that of person B? In this sense, the variable is based on an ordinal level of measurement. Strictly speaking, a non-parametric difference test should be used instead of a t-test.

The non-parametric equivalent of Independent Samples T-test is Mann-Whitney U. To compute Mann-Whitney test:

- Go to: Analyze → Non-parametric tests → Legacy Dialogs → 2-Independent Samples
- Transfer qualslp and sex to the appropriate lists (the terminology is the same as for t-test)
- Make sure **Mann-Whitney U** is checked in **Test Type**
- Click OK

You should see the following set of results tables. We can see that the mean rank is higher for males than for females – remember there is a positive (but not necessarily linear) correlation between original score and rank. We can find out if this difference is significant by looking at the last row of the test statistics table.

Mann-Whitney Test

Ranks

	sex	N	Mean Rank	Sum of Ranks
quality of sleep	female	150	131.41	19712.00
	male	118	138.42	16334.00
	Total	268		

Test Statistics[a]

	quality of sleep
Mann-Whitney U	8387.000
Wilcoxon W	19712.000
Z	-.762
Asymp. Sig. (2-tailed)	[redacted]

a. Grouping Variable: sex

Q4: Is the difference in reported sleep quality between males and females significant (according to the Mann-Whitney U test)?

When dealing with ordinal/rank data it usually more meaningful to compare median, rather than mean average. How would you compute and compare median sleep quality between males and females using SPSS?

Q5: What is the median of qualslp for the male group?

Exercise 3: Computing Wilcoxon Signed Ranks in SPSS

Finally, this section will show you how to compute the non-parametric equivalent of a paired-samples t-test. This is called Wilcoxon Signed Ranks test. The SPSS procedure is as follows:

- Go to: Analyze → Non-parametric tests → Legacy Dialogs → 2 Related samples
- Transfer your two variables into the same 'pair' row in the **Test Pairs** list
- Check that **Wilcoxon** is checked in the **Test Type** frame

Because it is based on a single administration of a questionnaire, the Sleep data is a little short on paired-samples (the same measures observed on different occasions/under different conditions). Hence, we will just use the same variables that we used in the last tutorial with paired t-test (Hours slept on weeknights and weekend nights: "hourwnit" and "hourwend"). Enter this variable pair into the test dialogue and run the Wilcoxon test. The output should look like this:

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
hours sleep/ week ends - hours sleep/ week nights	Negative Ranks	16[a]	51.28	820.50
	Positive Ranks	152[b]	88.00	13375.50
	Ties	[redacted][c]		
	Total	269		

a. hours sleep/ week ends < hours sleep/ week nights

b. hours sleep/ week ends > hours sleep/ week nights

c. hours sleep/ week ends = hours sleep/ week nights

Test Statistics[a]

	hours sleep/ week ends - hours sleep/ week nights
Z	-10.144[b]
Asymp. Sig. (2-tailed)	.000

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

If you compare the results with those from the paired t-test, you will see they are very similar, in that a highly significant difference is reported. We would expect this because the data does in fact meet the criteria for a parametric test quite well. The variables contain scale (interval/ratio) level values (it would make sense to multiply and divide them) as opposed to ordinal, and if you check the distributions of both variable you will see they are quite normal. The Wilcoxon output (above) also shows us a graph of the distribution of the differences between value pairs, which we can also see is fairly normal.

Like Mann-Whitney, Wilcoxon converts raw scores into ranks before computing the test result. The algorithm works by assigning ranks based on the absolute (unsigned) value of the difference between values for each case. Then, the sum/mean of ranks is found separately for all negative and all positive differences. We can see from the Ranks table that there are many more positive differences. If we check the key (for a, b and c superscripts) we see this confirms what we found from the t-test: that people tend to sleep somewhat longer at the weekend. However, it is also interesting to note how often the difference is negative and that in many cases people said they slept exactly the same number of hours on weeknights as they did on weekends. This extra information about the distribution of differences is not available from a t-test and can sometimes be useful, providing a different perspective for interpretation of the data.

Q6: In what percentage of cases was the difference negative (i.e. slept longer on weeknights)?

Q7: How many cases were tied (no difference)?

Summary and further work

In this tutorial you have extended your knowledge of correlation and difference tests by learning about non-parametric tests. We have only scratched the surface of the topic here. You are encouraged to read further into non-parametric tests by reading Chapter 16 of Pallant. In addition, Chapter 11 discusses Spearman's rho in greater detail.

Further Reading

Cohen, J.W. (1988) *Statistical power analysis for the behavioural sciences* (2nd Ed.). New York: Erlbaum.

Pallant, J. (2007) SPSS survival manual : a step by step guide to data analysis using SPSS for Windows (Version 15), Chapters 11 and 16.