UNIT 2 MANN WHITNEY ‘U’ TEST FOR TWO SAMPLE TEST

Structure

2.0 INTRODUCTION

Non-parametric statistics are distribution free statistics and can be used for small samples as well as any kind of distribution. It has many tests which are equivalent to the parametric tests. For instance for the tests like mean, we have Mann Whitney U test, for Pearson ‘r’ we have Kendall tau test and so on. The non-parametric tests are available for single sample, matched pair sample, two samples and k samples. In this unit we will be dealing with Two sample tests and the various non-parametric tests that we can use analyse data if we have two samples. We will initially start with the definition of what is two sample test and go on to present different non-parametric statitistics that could be applied to analyse such data and then finally present how to solve problems based on such data.

2.1 OBJECTIVES

After reading this unit, you will be able to:

- Define two sample data;
- Explain what are two sample tests;

- Present the various non-parametric tests that can be used to analyse two sample data;
- Explain the significance levels and interpretation of such data; and
- Solve problems in two sample data.

2.2 DEFINITION OF TWO SAMPLE TESTS

Two sample tests are those which we call as tests of independence rather than goodness of fit tests. We are testing to see whether or not 2 variables are "related" or "dependent". Thus, the Ho , that is the Null Hypothesis, takes the general form

Ho: x and y are independent.

In a parametric test, we have seen earlier that to find out if two groups differ in performance , we used the t-test and if the t value was significant at .05 level, we rejected the null hypothesis and concluded that the two groups differed in their performance. In this type of t test we required the performance to be nornmally distributed and the sample size to be more than 30 and such other parametric test conditions. However if the sample size is less than 30 and the data is not normally distributed we would use the non-parametric test to find out if the two groups differed in their performance.

Let us say the 2 samples are males and females . We are comparing their marks in History at the final examination. Here one sample is gender, categorised into male and female. This is a nominal scale measurement. The other is 'scores obtained in History', a continuous variable which can range from zero to 100 or more depending on out of how many marks a score has been taken as performance.

Thus one variable , let us say Gender , that is X variable is in nominal scale of measurement and dichotomous (takes on only 1 of 2 possible values – that is male or female). The marks in History Y variable is treated as continuous (can take on a whole range of values on a continuum). To find out if males scored significantly higher than females in history in the final examination, we may if the sample size was more than 30, apply t-test and if the t value is significant, and also males average or the mean score is higher than that of female students, then we will conclude that males have scored significantly higher in history as compared to female students.

To take another example, let us say 2 groups matched in many respects each receive a different teaching method, and their final exam scores are compared.

Let us say X = Teaching method categorised into 1 and 2 methods.(X1 and X2)

Y = the final exam scores

Ho: X1 and X2 will not differ in terms of marks obtained

Here X = the independent variable and

Y = the dependent variable. (Marks in history)

We are trying to find out which of the two teaching methods is producing higher marks in final exam performance. The null hypothesis states there will be no difference in the marks obtained irrespective of X1 or X2 method of teaching.

Now all the subjects marks are taken for the two groups of persons undergoing two different methods of teaching. For this the ideal non-parametric test will be the t-test.

If the t-value goes beond the value given in the table at .05 level, we reject the null hypothesis and state that the two teaching methods do bring about a change in the performance of the students.

However if the sample size is small and there is no assumption of normality, then we would apply non-parametric test. These tests help in rejecting or accepting null hypothesis depending on the analysis.

Now let us see what are the various tests do we have under the non-parametrict test that can be applied.

2.3 MANN WHITNEY 'U' TEST

The Mann-Whitney (Wilcoxon) rank-sum test is a non-parametric analog of the two-sample *t* test for independent samples. The *Mann-Whitney U test* is a non-parametric test that can be used to analyse data from a two-group independent groups design when measurement is at least ordinal. It analyses the *degree of separation* (or the amount of overlap) between the Experimental (E) and Control (C) groups.

The *null hypothesis* assumes that the two sets of scores (E and C) are samples from the same population; and therefore, because sampling was random, the two sets of scores *do not differ systematically* from each other.

The *alternative hypothesis*, on the other hand, states that the two sets of scores *do* differ systematically. If the alternative is directional, or one-tailed, it further specifies the direction of the difference (i.e., Group E scores are systematically higher or lower than Group C scores).

The statistic that is calculated is either U or U'.

U1 = the number of Es less than Cs

U2 = the number of Cs less than Es

U = the smaller of the two values calculated above

U' = the larger of the two values calculated above

When you perform these tests, your data should consist of a random sample of observations from two different populations. Your goal is to compare either the location parameters (medians) or the scale parameters of the two populations. For example, suppose your data consist of the number of days in the hospital for two groups of patients: those who received a standard surgical procedure and those who received a new, experimental surgical procedure. These patients are a random sample from the population of patients who have received the two types of surgery. Your goal is to decide whether the median hospital stays differ for the two populations.

2.4 RELEVANT BACKGROUND INFORMATION ON 'U' TEST

The Mann-Whitney *U* test is employed with ordinal (rank-order) data in a hypothesis testing situation involving a design with two independent samples. If the result of the Mann-Whitney *U* test is significant, it indicates there is a significant difference between the two sample medians, and as a result of the latter the researcher can conclude there is a high likelihood that the samples represent populations with different median values.

Two versions of the test to be described under the label of the Mann-Whitney *U* test were independently developed by Mann and Whitney (1947) and Wilcoxon (1949).

The version to be described here is commonly identified as the Mann-Whitney *U* test, while the version developed by Wilcoxon (1949) is usually referred to as the Wilcoxon-Mann-Whitney test.' Although they employ different equations and different tables, the two versions of the test yield comparable results.

In employing the Mann-Whitney *U* test, one of the following is true with regard to the rank order data that are evaluated:

a) The data are in a rank order format, since it is the only format in which scores are available; or

b) The data have been transformed into a rank order format from an interval ratio format, since the researcher has reason to believe that the normality assumption (as well as, perhaps, the homogeneity of variance assumption) of the t test for two independent samples (which is the parametric analog of the Mann-Whitney U test) is saliently violated.

It should be noted that when a researcher elects to transform a set of interval/ratio data into ranks, information is sacrificed. This latter fact accounts for the reluctance among some researchers to employ non-parametric tests such as the Mann-Whitney *U* test, even if there is reason to believe that one or more of the assumptions of the t test for two independent samples have been violated.

Various sources (e.g., Conover (1980, 1999), Daniel (1990), and Marascuilo and McSweeney (1977)) note that the Mann-Whitney U test is based on the following assumptions:

a) Each sample has been randomly selected from the population it represents;

b) The two samples are independent of one another;

c) The original variable observed (which is subsequently ranked) is a continuous random variable. In truth, this assumption, which is common to many non-parametric tests, is often not adhered to, in that such tests are often employed with a dependent variable which represents a discrete random variable; and

d) The underlying distributions from which the samples are derived are identical in shape. The shapes of the underlying population distributions, however, do not have to be normal.

Maxwell and Delaney (1990) pointed out the assumption of identically shaped distributions implies equal dispersion of data within each distribution. Because of this, they note that like the *t* test for two independent samples, the Mann-Whitney *U* test also assumes homogeneity of variance with respect to the underlying population distributions.

Because the latter assumption is not generally acknowledged for the Mann-Whitney *U* test, it is not uncommon for sources to state that violation of the homogeneity of variance assumption justifies use of the Mann-Whitney *U* test in lieu of the *t* test for two independent samples.

It should be pointed out, however, that there is some empirical evidence which suggests that the sampling distribution for the Mann-Whitney *U* test is not as affected by violation of the homogeneity of variance assumption as is the sampling distribution for *t* test for two independent samples. One reason cited by various sources for employing the Mann-

Whitney *U* test is that by virtue of ranking interval/ratio data, a researcher will be able to reduce or eliminate the impact of outliers.

Self Assessment Questions

1) Which non-parametric test should we use when the data is obtained from two different samples (Independent of each other) and we wish to see the difference between the two samples on a particular variable?

....

....

....

....

2) What is the underlying assumption of Mann-Whitney U test?

....

....

....

....

2.5 STEP BY STEP PROCEDURE FOR 'U' TEST FOR SMALL SAMPLE

Step-by-step procedure

Mann Whitney U Test for Small Sample case (not more than 20 items in each set), use U if the data is

a) in the form of ranks or

b) not normally distributed

c) there is an obvious difference in the variance of the two groups.

STEP 1: Rank the data (taking both groups together) giving rank 1 to the lowest score, and the highest rank to then highest score.

STEP 2: Find the sum of the ranks for the smaller sample

STEP 3: Find the sum of the ranks for the larger sample

STEP 4: Find U applying the formula given below:

$U = N_1N_2 + [N_1(N_1 + 1) / 2] - \Sigma R_1$

and

$U' = N_1N_2 + [N_2(N_2 + 1) / 2] - \Sigma R_2$

STEP 5: Look up the smaller of U and U' in Table H. There is a significant difference if the observed value is equal to or more than the table value.

STEP 6: Translate the results of the test back in the terms of experiment.

Worked Up Example:

Assessment Center Rating By Two Teams: Officers Randomly Assigned to Teams

Team A		Team B	
Score	Rank (R_1)	Score	Rank (R_2)
72	13	97	25
67	10	76	16
87	21	83	19
46	2	69	12
58	6	56	5
63	8	68	11
84	20	92	24
53	3	88	22
62	7	74	15
77	17	73	14
82	18	65	9
89	23	54	4
		43	1
	$\sum R1 = 148$		$\sum R2 = 177$

Step 1: Rank the ratings from lowest to highest regardless of assessment team.

Step 2: Sum the ranks in either group

$\Sigma (R_1) = 148$

$\Sigma (R_2) = 177$

Step 3: Calculate U

$U = N_1N_2 + [N_1(N_1 + 1) / 2] - \Sigma R_1$

$U = (12)(13) + [12 (12 + 1) / 2] - 148$

$U = 156 + 78 – 148 = 86$

And Calculate U'

$U' = N_1N_2 + [N_2(N_2 + 1) / 2] - \Sigma R_2$

$U' = (12)(13) + [13(13 + 1) / 2] - 177$

$U' = 156 + 91 – 175 = 70$

Step 4: Determine the significance of U

Decide whether you are making a one- or a two-tailed decision

Compare the smaller value of U to the appropriate critical table value for N_1 and N_2

If the observed U is smaller than the table value, the result is significant.

Step 5: The critical value of U for $N_1 = 12$ and $N_2 = 13$, two-tailed $\alpha = 0.05$, is 41.

Since the smaller obtained value of U ($U' = 70$) is larger than the table value, the null hypothesis is accepted. And we conclude that there is no significant difference in the ratings given by the two assessment teams.

2.6 STEP BY STEP PROCEDURE FOR 'U' TEST FOR LARGE SAMPLE

When both sample sizes are greater than about 20, the sampling distribution of U is for practical purposes, normal. Therefore, under these conditions, one can perform a z-test as follows:

The procedure to obtain U is similar as in small sample case (Step 1 to 3). Then the formula for Z is applied as:

$$Z = [U - (N_1N_2) / 2] / \quad N_1N_2 \ \sqrt{(N_1 + N_2 + 1)/12}$$

If we are dealing with a two-tailed test, then the observed z is significant at the 5 per cent level if it exceeds 1.96. For one tailed test, 5 per cent significance is attained if z exceeds 1.64 (Check these in table D in Statistics book original).

The ranking procedure can become quite laborious in large samples. Partly for this reason and partly because violation of the assumptions behind parametric statistics become less important for large sample, the Mann Whitney U test tends to be restricted to use with relatively small samples.

Self Assessment Questions

1) What unit of sample is considered as an appropriate sample for Mann Whitney U test for small sample?

....

....

....

2) What is the rationale of applying Z test be applied in a non-parametric setting?

....

....

....

2.7 COMPUTING MANN-WHITNEY U TEST IN SPSS

Step 1.Choose Analyse

Step 2. Select Non-parametric Tests

Step 3. Select 2 Independent Samples

Step 4. Highlight your test variable (in our example this would be age) and click on the arrow to move this into the Test Variable List box

Step 5. Highlight the grouping variable and click on the arrow to move this into the Grouping Variable box.

Step 6. Click on Define Groups and type in the codes that indicate which group an observation belongs to (in our example, the codes which indicate whether a subject is male or female). Click on Continue

Step 7. Under Test Type make sure that Mann-Whitney U is selected

Step 8. If you want exact probabilities, click on Exact, choose Exact, then Continue

Click on OK

2.8 WILCOXON MATCHED PAIR SIGNED RANK TEST

The Wilcoxon Matched Pair signed-ranks test is a non-parametric test that can be used for 2 repeated (or correlated) measures when measurement is at least ordinal. But unlike the sign test, it *does* take into account the magnitude of the difference.

In using this test, the difference is obtained between each of N pairs of scores observed on matched objects, for example, the difference between pretest and post-test scores for a group of students.

The difference scores obtained are then ranked.

The ranks of negative score differences are summed and the ranks of positive score differences are summed.

The test statistic T is the smaller of these two sums.

Difference scores of 0 are eliminated since a rank cannot be assigned.

If the null hypothesis of no difference between the groups of scores is true, the sum of positive ranks should not differ from the sum of negative ranks beyond that expected by chance.

2.9 RELEVANT BACKGROUND INFORMATION ON WILCOXON TEST

The Wilcoxon matched-pairs signed-ranks test (Wilcoxon (1945, 1949)) is a non-parametric procedure employed in a hypothesis testing situation involving a design with two dependent samples. Whenever one or more of the assumptions of the t test for two dependent samples are saliently violated, the Wilcoxon matched-pairs signed-ranks test (which has less stringent assumptions) may be preferred as an alternative procedure.

The Wilcoxon matched-pairs signed-ranks test is essentially an extension of the Wilcoxon signed-ranks test (which is employed for a single sample design) to a design involving two dependent samples.

In order to employ the Wilcoxon matched-pairs signed ranks test, it is required that each of n subjects (or n pairs of matched subjects) has two interval/ratio scores (each score having been obtained under one of the two experimental conditions).

A difference score is computed for each subject (or pair of matched subjects) by subtracting a subject's score in Condition 2 from his score in Condition 1.

The hypothesis evaluated with the Wilcoxon matched-pairs signed-ranks test is whether or not in the underlying populations represented by the sampled experimental conditions, the median of the difference scores equals zero.

If a significant difference is obtained, it indicates that there is a high likelihood that the two sampled conditions represent two different populations.

The Wilcoxon matched-pairs signed-ranks test is based on the following assumptions:

a) The sample of n subjects has been randomly selected from the population it represents;

b) The original scores obtained for each of the subjects are in the format of interval/ratio data; and

c) The distribution of the difference scores in the populations represented by the two samples is symmetric about the median of the population of difference scores.

As is the case for the t test for two dependent samples, in order for the Wilcoxon matched pairs signed ranks test to generate valid results, the following guidelines should be adhered to:

a) To control for order effects, the presentation of the two experimental conditions should be random or, if appropriate, be counterbalanced; and

b) If matched samples are employed, within each pair of matched subjects each of the subjects should be randomly assigned to one of the two experimental conditions

As is the case with the t test for two dependent samples, the Wilcoxon matched-pairs signed-ranks test can also be employed to evaluate a "one-group pretest-posttest" design. The limitations of the one group pretest posttest design are also applicable when it is evaluated with the Wilcoxon matched pairs signed ranks test.

It should be noted that all of the other tests in this text that rank data (with the exception of the Wilcoxon signed-ranks test), ranks the original interval/ratio scores of subjects.

The Wilcoxon matched-pairs signed-ranks test, however, does not rank the original interval/ratio scores, but instead ranks the interval/ratio difference scores of subjects (or matched pairs of subjects).

For this reason, some sources categorise the Wilcoxon matched-pairs signed-ranks test as a test of interval/ratio data.

Most sources, however, categorise the Wilcoxon matched-pairs signed-ranks test as a test of ordinal data, by virtue of the fact that a ranking procedure is part of the test protocol.

Self Assessment Questions

1) Which non-parametric test should we use when the data is obtained from two related sample and we wish to see the difference between the two samples on a particular variable?

....

....

....

....

2) Which one assumption does not apply to Wilcoxon Matched Pair Test, which applies to Mann Whitney U test?

....

....

....

....

3) What is the difference between t Test for Matched Pair sample and Wilcoxon Matched Pair Test?

....

....

....

....

2.10 STEP BY STEP PROCEDURE FOR WILCOXON TEST FOR SMALL SAMPLE

Step-by-step procedure

Wilcoxon Test-Small Sample case (not more than 25 pairs of scores).

For matched pairs or repeated measures designs: use instead of a correlated *t*-test if either

a) the differences between treatments can only be ranked in size or

b) the data is obviously non-normal or

c) there is an obvious difference in the variance of the two groups.

STEP 1: Obtain the difference between each pair of reading, taking sign into account

STEP 2: Rank order these differences (ignoring the sign), giving rank 1 to the smallest difference

STEP 3: Obtain *T,* the sum of the ranks for differences with the less frequent sign

STEP 4: Consult Table J. If the observed *T* is equal to or less than the table value then there is a significant difference between two conditions

STEP 5: Translate the result of the test back in terms of the experiment

Worked Up Example:

Eight pairs of twins were tested in complex reaction time situations; one member of each pair was tested after drinking 3 double whiskies, the other member was completely sober. The following reaction times were recorded:

Sober Group	Whisky Group	Step 1: Differences	Step 2:Ranks
310	300	-10	1
340	320	-20	2
290	360	70	5
270	320	50	4
370	540	170	6
330	360	30	3
320	680	360	7
320	1120	800	8

STEP 3: Less frequent sign of difference is negative,

$T = 1 + 2 = 3$

STEP 4: From Table J, when N = 8, T = 4. As the observed value of *T* is less than the table value, there is a significant difference between the two conditions.

STEP 5: Complex reaction time scores are significantly higher after drinking 3 double whiskies than when sober.

2.11 STEP BY STEP PROCEDURE FOR WILCOXON TEST FOR LARGE SAMPLE

When both sample sizes are greater than about 20, the sampling distribution of U is (for practical purposes) normal.

As with the Mann Whitney U test, the sampling distribution of the statistics (In this case *T*) approaches the normal distribution as the sample size becomes large. Therefore, under these conditions, again one can perform a z-test as follows:

$Z= \{T- (N(N+1) / 4)\}/ N(N+1)(2N+1) / \sqrt{24}$

The significance decisions are identical to those for the Mann Whitney largesample case. Thus, if we have a two tailed test, the observed z is significant at the 5 per cent level if it exceeds 1.96. For the one-tailed test, significance is attained if z exceeds 1.64. However, as with the Mann-Whitney test, and for the same reasons, the Wilcoxon test tends to be restricted to use with relatively small samples.

Self Assessment Questions

What unit of sample is considered as an appropriate sample for Mann Whitney U test for small sample?Give the underlying assumptions of Mann Whitney U test?

....

....

....

....

2.12 COMPUTING THE WILCOXON SIGNED RANK SUM TEST IN SPSS

- Choose Analyse
- Select Non-parametric Tests
- Select 2 Related Samples
- Specify which two variables comprise your pairs of observation by clicking on them both then clicking on the arrow to put them under Test Pair(s) List.
- Under Test Type select Wilcoxon

If you want exact probabilities (i.e. based on the binomial distribution), click on Exact, choose Exact, then Continue

Click on OK

2.13 COMPARISON OF MANN-WHITNEY 'U' TEST AND WILCOXON MPSR TEST WITH T-TEST

The power efficiency of the Mann-Whitney and Wilcoxon tests, whilst usually somewhat lower than the corresponding *t*-test, compares very favourably with it. The Mann-Whitney and Wilcoxon tests can be used in situations where the *t* –Test would be in-appropriate (e.g. where the assumptions of the *t*-test obviously do not apply). In other words, they are capable of wider application.

Different statisticians give different advice as to the relative merits of parametric and non-parametric tests. The non-parametric camp claims that their tests are simpler to compute, have fewer assumptions and can be used more widely. The parametric camp claims that their tests are robust with respect to violations of their assumptions and have greater power efficiency.

The strategy recommended here is to use the *t*-test unless the data is in form of ranks, or where the sample is small and either the distribution is obviously non-normal or there are obviously large differences in variance.

However, if you are particularly pressed for time or have a large number of analyses to do there is particularly nothing inappropriate about using non-parametric statistics, even in cases where *t*-tests might have been used.

2.14 LET US SUM UP

Two Sample test can be of two types independent sample test (two different samples being tested on one variable wherein one sample does not affect the other sample) or paired or dependent sample (same sample being tested twice or sample have some relation with each other).

The *t*-test is the parametric test for a two sample test, in non-parametric tests, Mann-Whitney 'U' test and Wilcoxon test are used for independent and paired sample respectively.

Both these tests have their own advantages, and can be used for a smaller sample size, do not have too many assumptions and can be used more widely.

2.15 UNIT END QUESTIONS

1) A researcher had an experimental group of m = 3 cases and a control group of n = 4 cases. The scores were as following:

   Exprimental scores: 9, 11, 15

   Control scores: 6, 8, 10, 13

2) In the problem 1 above, Assume these groups are independent, apply appropriate statistics and state whether the experimental condition and control conditions differ or not.

3) In the problem 1 above, Assume these groups are correlated, apply appropriate statistics and state whether the experimental condition and control conditions differ or not.

4) Doctor Radical, a math instructor at Logarithm University, has two classes in advanced calculus. There are six students in Class 1 and seven students in Class 2.

The instructor uses a programmed textbook in Class 1 and a conventional textbook in Class 2. At the end of the semester, in order to determine if the type of text employed influences student performance, Dr. Radical has another math instructor, Dr. Root, to rank the 13 students in the two classes with respect to math ability. The rankings of the students in the two classes follow:

Class 1: 1, 3, 5, 7, 11, 13

Class 2: 2, 4, 6, 8, 9, 10, 12

(Assume the lower the rank the better the student).

5) To 4 above, Apply appropriate statistics and tell if the type of text employed influenced students performance?

6) Why should you not use the large-sample *z*-test version of a non-parametric test when you have samples small enough to allow the use of small sample version?

7) Identify the non-parametric test that ought to be used.

8) You have 5 independent groups of subjects, with different numbers per group. There is also substantial departure from homogeneity of variance. The null hypothesis states that there are no differences between the groups.

You have the same situation described in question 4 (a); and in addition, the alternative hypothesis states that when the mean ranks for the 5 groups are listed from smallest to largest, they will appear in a particular *pre-specified* order.

2.16 SUGGESTED READINGS

Daniel, W. W. (1990) *Applied Non-parametric Statistics*, 2d ed. Boston: PWS-Kent.

Johnson, Morrell, and Schick (1992), Two-Sample Non-parametric Estimation and Confidence Intervals Under Truncation, *Biometrics*, 48, 1043-1056.

Siegel S. and Castellan N.J. (1988) *Non-parametric Statistics for the Behavioral Sciences* (2nd edition). New York: McGraw Hill.

Wampold BE & Drew CJ. (1990) *Theory and Application of Statistics*. New York: McGraw-Hill.