An Overview of the Median and Mann-Whitney U Tests

E. U. Oti[1*]; M. O. Olusola[2]; W. K. Alvan[3]; J. Ikirigo[4]

[1]Department of Statistics,
Federal Polytechnic, Ekowe, Bayelsa State, Nigeria.

[2]Department of Statistics,
Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

[3,4]Department of Physics with Electronics,
Federal Polytechnic, Ekowe, Bayelsa State, Nigeria.
E-mail: eluchcollections@gmail.com*

Abstract **— The median and Mann-Whitney U tests are non-parametric test methods designed to handle two samples problem. Their data are continuous which consist of two mutually independent random samples. They are used to test whether two (or more) independent samples have been drawn from populations with the same median. In this paper, we discussed and analyzed these two methods using the same illustrative example, testing the null hypothesis at 5 percent significance level and it was observed that both the median and Mann-Whitney U test were statistically significant indicating that the two samples of scores earned by students in Statistics department were drawn from populations with equal median scores in the course.**

Keywords: *Chi-Square, Contingency Table, Population, Rank, Two Samples*.

I. Introduction

The median test and the Mann-Whitney U test are some of the nonparametric techniques developed to handle two samples problem. Their data consist of two mutually independent random samples, i.e., random samples drawn independently from each of two populations. Not only are the elements within each sample independent, but also every element in the first sample is independent of every element in the second sample (Gibbons, 1993; Gibbons and Chakraborti, 2003).

The universe consists of two populations, which we call the X and Y populations, with cumulative distribution functions denoted by F_X and F_Y respectively. We have a random sample of size m drawn from the X population and another random sample of size n drawn independently from the Y population, $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ (Siegel and Castellan, 1988; Corder and Foreman, 2014). Usually, the hypothesis of interest in the two samples problem is that the two samples are drawn from identical populations, that is

$$H_0 : F_X(x) = F_Y(x) \text{ for all } x \text{ versus}$$
$$H_1 : F_X(x) \neq F_Y(x) \text{ for some } x.$$

The purpose of this paper is to analyze the median test and the Mann-Whitney U test and show how they could be applied using the same data set. The rest of this paper is organized as follows: section 2 discussed the median test, its method, decision rule and the test statistic; section 3 is centered on Mann-Whitney U test; furthermore, section 4 is the illustrative example, while section 5 is the conclusion of the paper.

II. The Median Test

Let X_i be the ith $(i = 1, 2, \ldots, m)$ observation in a random sample of size m independently drawn from population X and let Y_i be the jth $(j = 1,2, \ldots, n)$ observation in a random sample of size n independently drawn from population Y. We pool the two samples m and n into one combined sample and determine the common median N for this pooled sample, this combination of samples are done in decreasing (or increasing) order of magnitude. We then find for each sample the number of observations that fall above or below the common median (if there is no tie), and arrange the resulting frequencies in a 2×2 contingency table.

If only few observations, say one or two are exactly equal to the common median, they are discarded and the total sample size is reduced accordingly. If however, many observations are exactly equal to the common median, the 2×2 contingency table is constructed by dichotomizing the data for each sample into those that fall above (>) the common median and those that fall at or below (≤) the common median (Friedlin and Gastwirth, 2000; Oyeka, 2013). In precision, if the two samples are drawn from populations with equal medians, we would expect that approximately one half of the observations in each sample will lie above the common median N and approximately one half will lie below it.

The 2×2 contingency table classifying the observations in each sample according to whether they lie above or below the common median enables us test the null hypothesis that the samples are drawn from populations with the same median. The contingency table is now analyzed by using the chi-square formula

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \quad (1)$$

where n is the total frequency, $n_{i.}$ is the total or marginal frequency for the ith row, while $n_{.j}$ is the total or marginal frequency for the jth column. The degree of freedom is always 1, since $(r-1)(c-1) = (2-1)(2-1) = 1$. The decision rule: if the calculated χ^2 is equal to or greater than the tabulated critical value $\chi^2_{1-\alpha;\,(r-1)(c-1)}$

III. The Mann-Whitney U Test

Mann and Whitney (1947) proposed a method which is based on a comparison of every observation x_i in the first sample with every observation y_j in the other sample. Like in the median test, suppose x_i be the ith $(i = 1,2,\dots,m)$ observation in a random sample of size m independently drawn from population X and let y_j be the jth $(j = 1,2,\dots,n)$ observation in a random sample of size n independently drawn from population Y.

The data in the two samples are combined and then ranked either from the largest to the smallest or from the smallest to the largest. The ranks assigned to the observations in the two samples are then separated and the sums of the ranks are calculated for each sample. We will denote the sum of the ranks for the first sample, with sample size n_x by R_1 and the sum of the ranks for the second sample, with sample size n_y by R_2.

The value of Mann-Whitney U statistic is calculated as

$$U_x = n_x n_y + \frac{n_x(n_x+1)}{2} - R_1 \quad (2)$$

Or

$$U_y = n_x n_y + \frac{n_y(n_y+1)}{2} - R_2 \quad (3)$$

We can count the number of times an x_i from sample 1 is greater than a y_j from sanple 2 which is denoted by U_x. Similarly, the number of times an x_i from sample 1 is smaller than a y_j from sample 2 is denoted by U_y. Under the null hypothesis, we would expect U_x and U_y to be approximately equal. If either U_x and U_y is known, the other can easily be obtained from the expression

$$U_x = n_x n_y - U_y \quad (4)$$

The sampling distribution of either U_x or U_y can be found by listing all of the rank combinations of $n_x + n_y$ ranks, and treating them as equally likely outcomes under the null hypothesis. If the number of observations is such that the samples sizes n_x and n_y are both 8 and more, the statistic U is approximately normally distributed with mean

$$\mu_U = \frac{n_x n_y}{2} \quad (5)$$

and standard deviation

$$\sigma_U = \sqrt{\frac{n_x n_y(n_x+n_y+1)}{12}} \quad (6)$$

Hence, the corresponding z-score for the Mann-Whitney U Statistic is calculated as

$$Z = \frac{U - \mu_u}{\sigma_u} = \frac{U - \frac{n_x n_y}{2}}{\sqrt{\left(\frac{n_x n_y\,(n_x+n_y+1)}{12}\right)}} \quad (7)$$

where U is either U_X or U_y. The z-score is usually compared at a given level of significance with an appropriate critical value obtained from a normal distribution table for a rejection or acceptance of the null hypothesis.

IV. Example

A random sample of 16 students in Federal Polytechnic Ekowe, who were enrolled in an introductory course in Statistics department were taught by lecturer A, while a second sample of 19 students were taught by lecturer B. After the semester examination, the students earned the scores shown below. Test at the 5 percent significance level, the null hypothesis that the students performed equally in the course under both lecturers.

Sample A:
89, 70, 50, 68, 37, 75, 55, 49, 52, 55, 60, 80, 60, 70, 50, 62.
Sample B:
50, 55, 65, 50, 63, 73, 75, 60, 40, 35, 45, 60, 40, 65, 50, 62, 56, 75. 82.

Now, using the Median Test Method to solve the above problem as follows.

The Median test of scores earned in a course in Statistics department, Federal Polytechnic Ekowe which were taught by two lecturers:

H_0: The two samples of students are drawn from student populations with the same median score in the course.
versus
H_1: The two samples of students are drawn from student populations with different median score in the course.

Table 1: Scores earned in a Statistics course taught by two lecturers in Federal Polytechnic, Ekowe Bayelsa State as used for the Median Test.

Scores earned by students taught by lecturer A	Scores earned by students taught by lecturer B	Scores earned in descending order of both samples	Ranks of scores of both samples	Lecturer Labels to scores
89	50	89	1	A
70	55	82	2	B
50	65	80	3	A
68	50	75	5	A
37	63	75	5	B
75	73	75	5	B
55	75	73	7	B
49	60	70	8.5	A
52	40	70	8.5	A
55	35	68	10	A
60	45	65	11.5	B
80	60	65	11.5	B
60	40	63	13	B
70	65	62	14.5	A
50	50	62	14.5	B
62	62	60	17.5	A
	56	60	17.5	A
	75	60	17.5	B
	82	60	17.5	B
		56	20	B
		55	22	A
		55	22	A
		55	22	B
		52	24	A
		50	27	A
		50	27	A
		50	27	B
		50	27	B
		50	27	B
		49	30	A
		45	31	B
		40	32.5	B
		40	32.5	B
		37	34	A
		35	35	B

The common median of the two samples is the score of 60. So dichotomizing the observations for each sample into those scores that fall above the common median score of 60, and those that are equal to or fall below 60 are:

	Lecturer A	Lecturer B	Total
> Median	7	8	15
≤ Median	9	11	20
Total	16	19	35

Calculating the corresponding chi-square test statistic using Equation (1) gives

$$\chi^2 = \frac{35(7\times11-8\times9)^2}{15\times20\times16\times19} = 0.0096.$$

Also $\chi^2_{1-\alpha;\,(r-1)(c-1)} = \chi^2_{0.95;1} = 3.841.$

Since $0.0096 < 3.841 = \chi^2_{0.95;1}$, we do not reject the null hypothesis at the 5 percent significance level and therefore conclude that the two samples of scores earned in Statistics department may have been drawn from populations with equal median scores in the course.

Table 2: Scores earned in a Statistics course taught by two lecturers in Federal Polytechnic, Ekowe Bayelsa State as used for Mann-Whitney U Test

Scores earned by students taught by lecturer A	Scores earned by students taught by lecturer B	Scores earned in descending order	Ranks of scores of students taught by lecturer A	Ranks of scores of students taught by lecturer B
89	50	89	1	
70	55	82		2
50	65	80	3	
68	50	75	5	
37	63	75		5
75	73	75		5
55	75	73		7
49	60	70	8.5	
52	40	70	8.5	
55	35	68	10	
60	45	65		11.5
80	60	65		11.5
60	40	63		13
70	65	62	14.5	
50	50	62		14.5
62	62	60	17.5	
	75	60		17.5
	82	60		17.5
		56		20
		55	22	
		55	22	
		55		22
		52	24	
		50	27	
		50	27	
		50		27
		50		27
		50		27
		49	30	
		45		31
		40		32.5
		40		32.5
		37	34	
		35		35
Total			**271.5**	**358.5**

Also using the Mann-Whitney U test method to solve the same problem as stated above, we go as follows.

Fom Table 2, the sum of the ranks assigned to the observations in sample A, $R_1 = 271.5$, we compute the Mann-Whitney U statistic from Equation (2) as follows;

$$U_x = n_x n_y + \frac{n_x(n_x+1)}{2} - R_1 = 16 \times 19 + \frac{16(16+1)}{2} - 271.5 = 168.5$$

The mean of U statistic is $\mu_u = \frac{n_x n_y}{2} = \frac{16\times19}{2} = 152$, and the standard deviation is

$$\sigma_u = \sqrt{\frac{n_x n_y(n_x + n_y + 1)}{12}} = \sqrt{\frac{16 \times 19(16 + 19 + 1)}{12}}$$

$$= \sqrt{\frac{304(36)}{12}} = 30.199$$

Hence, the normal z-score corresponding to $U_x = 168.5$ is calculated from Equation (7) as

$$z = \frac{U_x - \mu_u}{\sigma_u} = \frac{168.5-152}{30.199} = 0.546$$

According to the limits of acceptance region, keeping in view 5% level of significance. As the z-value for 0.546 of the area under the normal curve is 3, we have the following limits of acceptance region:

Upper limit = $\mu_u + 3\sigma_u = 152 + 3 \times 30.199 = 242.597$

Lower limit= $\mu_u - 3\sigma_u = 152 - 3 \times 30.199 = 61.403$,

so the value of U_x is 168.5 which is in the acceptance region, we accept the null hypothesis and conclude that the two samples of scores earned in Statistics department may have been drawn from populations with equal median scores in the course at 5% level of significance.

V. Conclusion

In this paper, we have presented a statistical analysis of the median and Mann-Whitney U tests for two independent samples. The hypothesis was tested at 5 percent significance level; it was observed that both methods were statistically significant, indicating that the two samples of scores earned by students in a Statistics coursw were drawn from populations with equal median scores in the course.

References

Corder, G. W. and Foreman, D. I. (2014). Nonparametric Statistics: A step-by-step Approach, Wiley. ISBN 978-1118840313.

Friedlin, B. and Gastwirth, J. L. (2000). Should the median test be retired from general use? The American Statistician, 54, 161-164.

Gibbons, J. D. (1993). Nonparametric Statistical; An Introduction, Newbury Park, Sage Publication. Pp. 180-220.

Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric Statistical Inference. Fourth Edition, Revised and Expanded, Marcel Dekker, New York.

Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other, Annals of Mathematical Statistics, 18, 50-60.

Oyeka, C. A. (2013). An Introduction to Applied Statistical Methods. Ninth Edition, Nobern Avocation Publishing Company, Enugu.

Siegel, S. and Castellan, N. J. Jr. (1988). Nonparametric Statistics for the Behavioral Sciences. New York, McGraw-Hill.

.

.