Working Paper

विद्याविनियोगाद्विकासः
IIM
AHMEDABAD

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD

AN APPLICATION OF THE MANN-WHITNEY 'U' TEST

By

C.D. Bhattacharya

W P No. 356
April 1981

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380015
INDIA

AN APPLICATION OF THE MANN-WHITNEY 'U' TEST

The private sector industrial giants in India are dominated by Engineering units. The top thirty industrial giant units are constituted by approximately half number of engineering units. This of course depends on the way engineering units are defined. At the outset are there certain general engineering units like TELCO, Escorts, Ashok Leyland, Hindustan Motors, etc. Besides, there are units manufacturing steel, aluminium, metals, alloys, metal products and structurals. Included in this list are units like TISCO, Metal Box, Indian Aluminium, Guest Keen Williams, etc. One may debate about the inclusion of the units relating to coal, mining, cement, electricity and transport under this class. The units other than engineering are of varied nature. It includes units relating to chemicals, dyes, pharmaceuticals, refineries, plastics, cotton spinning and weaving, synthetic fibres, textiles, food products, etc. Though there cannot be a consensus in this respect, the units can be classified between engineering and non-engineering units. Once we have drawn these two classes it is possible to determine whether the two independent sample of units have been drawn from the same population or from two different populations having the same distribution. The test employed in this connection are Rank Sum tests which are in fact a whole family of tests. The Mann-Whitney 'U' test is just one member of this family.[1]

The Mann-Whitney 'U' test has been applied in this paper to two groups of top industrial units in the Indian private sector - engineering and non-engineering. The two groups are given in Exhibit I. There are 16 units in each group; they have been ranked in terms of their net sales as per the Economic Times dated March 18, 1980. The following symbols are used in a Mann-Whitney 'U' test:

n_1 = number of units in group 1.

n_2 = number of units in group 2.

R_1 = Sum of the ranks of the units in group 1.

R_2 = Sum of the ranks of the units in group 2.

In the existing case both n_1 and n_2 are equal to 16. It is, however, not necessary that both the groups should be of the same size. Now, by adding the ranks assigned to different units, we get the total ranks for each group of units. In our example, therefore,

n_1 = 16 R_1 = 281

n_2 = 16 R_2 = 247

---

[1] Richard I. Levin, Statistics for Management, Prentice Hall of India Private Limited, New Delhi - 110001.

Exhibit I

Sl No.	Engineering Units	Rank	Sl. No.	Non-engineering Units	Rank
1.	Tata Engineering	1	1.	Hindustan Lever	3
2.	Tata Steel	2	2.	Delhi Cloth	4
3.	Voltas	6	3.	Gwalior Rayon	5
4.	A C C	9	4.	Brooke Bond	7
5.	Escorts	12	5.	I T C	8
6.	Ashok Leyland	13	6.	Dunlop	10
7.	Hindustan Motors	15	7.	Rallis	11
8.	Metal Box	17	8.	Century Spinning	14
9.	Larsen & Toubro	20	9.	Reliance Textiles	16
10.	Calico Indl. Engg.	21	10.	Godrej Soaps	18
11.	G K W	24	11.	Tata Oil	19
12.	Mahindra & Mahindra	26	12.	EID Parry	22
13.	Calcutta Electric	27	13.	Shah Wallace	23
14.	Scindia Steam	28	14.	Union Carbide	25
15.	Indian Aluminium	29	15.	Bombay Dyeing	30
16.	Tata Power	31	16.	Madura Coats	32
	Total ranks	281		Total ranks	247

Calculating the 'U' Statistic

The 'U' statistic which is a measurement of the difference between the ranked observations of the two group of units can be determined by using the values for n_1 and n_2 and the ranked sums of R_1 and R_2.

$$U = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1$$

$$= (16)(16) + \frac{16\ (17)}{2} - 281$$

$$= 256 + 136 - 281$$

$$= 392 - 281$$

$$= 111 \Longleftarrow\text{---- U Statistic.}$$

If the hypothesis that the two groups of observations came from the same population is true, then this 'U' statistic has a distribution with a mean of

$$\mu u = \frac{n_1\ n_2}{2}$$

$$= \frac{(16)\ (16)}{2}$$

$$= 128 \Longleftarrow\text{---- mean of the 'U' statistic.}$$

The Standard error of the 'U' Statistic is given by

$$\sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

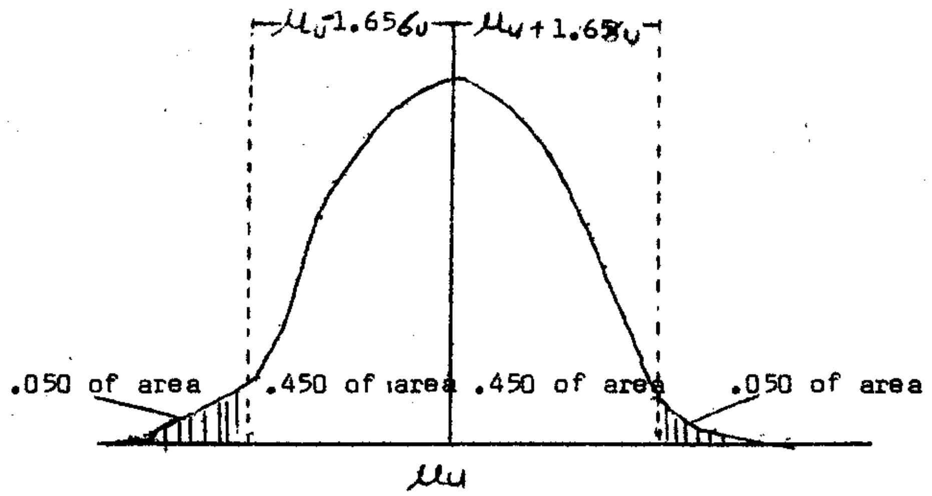
$$= \sqrt{\frac{(16)\ (16)\ (16 + 16 + 1)}{12}}$$

$$= \sqrt{\frac{(16)\ (16)\ (33)}{12}}$$

$$= 26.4$$

In the present case as n_1 and n_2 are both larger than 10, the distribution of the 'U' statistic can be approximated by the normal **distribution.** Suppose we want to test at the 10 per cent level of significance the hypothesis that these two groups have been drawn from the same population. Exhibit II illustrates this method diagramatically. The two shaded areas represent the 10 per cent level of significance.

Exhibit II



Using the normal distribution as our sample distribution in this test, it may be noted that the approximate value for an area of .450 is 1-65[2]. The two limits of the acceptance region can be calculated as under :

$\mu_u + 1.65\,\sigma_u$

$= 128 + 1.65 \times 26.4$

$= 128 + 43.6$

$= 171.6$ <---- Upper limit

and $\mu_u + 1.65\,\sigma_u$

$= 128 - 1.65 \times 26.4$

$= 128 - 43.6$

$= 84.4$ <---- Lower limit

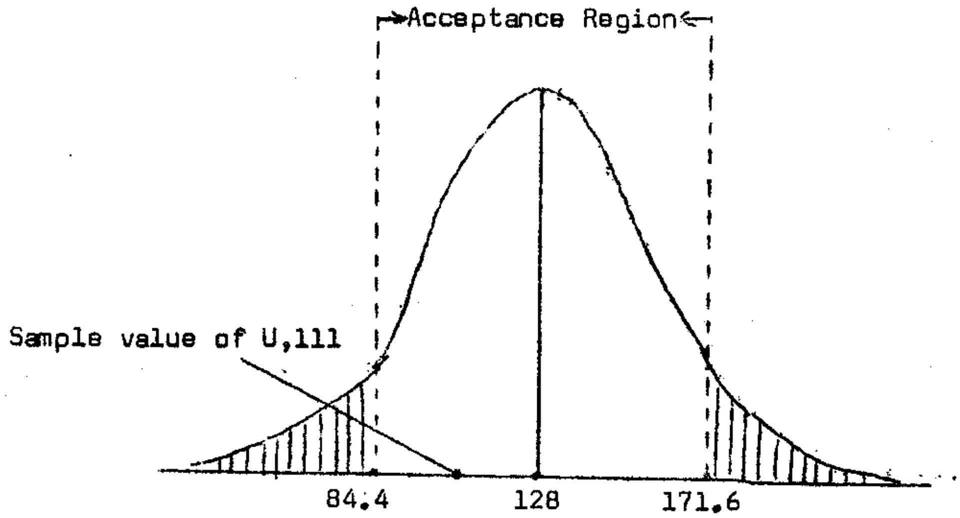
Exhibit III shows the limits of the acceptance region. The lower and the upper limits are 84.4 and 171.6, and the 'U' value is 111. The sample U statistic therefore lie within the acceptance region. The null hypothesis of no difference between the two group of units is therefore accepted. It may also be concluded that the distributions are equal.

The U statistic computed above were based on two groups of equal sizo. It is possible to undertake the same exerciso when the groups

---

[2] From Robert D. Mason, Essentials of Statistics, Prentice Hall Inc. 1976.

Exhibit III



are not equal to each other. Likewise, though the U statistic were arrived at by making use of R_1 values, it is possible to do the same exercise by using R_2 values. In that case

$$U = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2$$

$$= (16)(16) + \frac{(16)(17)}{2} - 247$$

$$= 256 + 136 - 247$$

$$= 392 - 247$$

$$= 145.$$

This value is just as far above the mean of 128 as 111 was below it. Regardless of whether we use R_1 or R_2 in order to calculate U we arrive at the same conclusion. It may be noted that in our example 145 falls under the acceptance region just as 111 did.

Conclusion

The present study makes use of the Mann-Whitney U test in order to verify whether the two independent group of units - engineering and non-engineering - have been drawn from the same population of private sector industrial giants in India or from two different populations having the same distribution. The units selected are ranked in terms of their net sales. The mean of the 'U' statistic comes to 128 and the standard error 26.4. At 10 per cent level of significance, the two limits of the acceptance region are 84.4 and 171.6. The sample 'U' statistic lie within the acceptance region. There is therefore no difference between the population of two group of units - engineering and non-engineering and the distributions are equal. It is possible to extend the analysis when the two groups are not equal to each other.