Philippine Journal of Psychology
1969, Vol. 2, No. 1, 31—33

MANN-WHITNEY'S *U* AS AN INDICATOR OF RELATIONSHIP[1]

LEONEL CAMPOS AND JOSEFINA SANTOS
Ateneo de Manila University

It is here suggested that Mann-Whitney's *U* can be used to derive a measure of relationship, r^{u}, which is the rank-analogue of the product-moment point biserial, r_{pb}. It is further suggested that Mann-Whitney's *U* itself can be used to evaluate the significance of r_u. An empirical comparison between r_u and r_{pb} yields $r = .875$ suggesting a basic similarity in their behavior. This fact is interpreted to indicate that r_u is an adequate substitute for r_{pb} in situations where the latter does not apply.

Maurice Kendall introduced the principle that the difference between two complementary proportions can, under appropriate circumstances, provide an index of relationship analogous—though not conceptually equivalent—to what is ordinarily obtained through product-moment techniques (Kendall, 1962, pp. 3-7). The main deficiency of this principle resides in the fact that it leads to formulas which fail to display the mathematical versatility of product-moment indices. From a practical viewpoint, however, Kendall's principle, as exemplified in rank correlation, *tau*, possesses at least two distinct advantages: (*a*) the outcome of Kendall's counting technique is amenable to simple interpretations, and, (*b*) the sampling problems involved here are greatly minimized as compared for instance, to the sampling problems which complicate a straightforward interpretation of Spearman's *rho* (Johnson, 1949; Hays, 1963). Added to these is the possibility that, in situations where product-moment applications are impracticable, Kendall's principle may provide an adequate alternative.

Assuming that sampling difficulties related to tests of significance can be satisfactorily worked out, it would seem that the most reasonable demand to make of a Kendall-type indicator, is that it behave to some extent like a correlation of the Pearson brand. That is, given data which naturally possess metric properties, and which can be artificially reduced to an ordinal scale, we want both our measures to co-vary to the extent that, when one is small, the other one is small also; when one is positive, the other one is positive also; when one is negative, the other is negative also. Their range ought to be, in each case, between —1 and 1. Whether there exists numerical equivalence between these two types of measures is a desirable, but not an absolutely necessary, attribute. Again, it must be pointed out that product-moment techniques have gained their present status on the basis of their mathematical versatility, and not because they may be indicators of "true" relationships, since the problem of "true" relationships is a metaphysical preoccupation, not an empirical one.

A situation where Kendall's principle may be used profitably is exemplied as follows: Suppose we have rated the quality of drawings made by a group of boys and girls, and we desire to ascertain the extent to which children of one sex are superior to children of the other sex in the quality of their drawings. If the ratings are scores with metric properties, our purpose is served if we use a point-biserial correlation, but

[1] This communication represents Report No. 3 of the Laboratory of Experimental Psychology of the Ateneo de Manila University.

if the drawings have simply been ranked from "poorest" to "best" regardless of the group membership of their authors, then the point-biserial correlation cannot be used, and measures of correlations estimated from ranks presently available are of limited practical application, e.g., Wald-Wolfowitz' Circular Serial Correlation Coefficient (Walsh, 1962).

It is of course well known that if we wish to know whether, when members of one group have gained higher ranks than the members of another group, when the ranking has been done in the manner indicated above, the statistical significance of this difference can be evaluated by means of Mann-Whitney's statistic, U (Siegel, 1956). Now Mann and Whitney utilize a counting technique similar to that prescribed by Kendall for the computation of *tau* (Kendall, 1962). The former authors have suggested that for two sets of independent observations, each with n_i and n_j observations respectively, it is possible to get $0 \leq U_i \leq n_i n_j$, or, $0 \leq U_j \leq n_i n_j$ depending on whether U is computed with reference to the group possessing attribute i (e.g., male), or with reference to the group possessing attribute j (e.g., female). In any case, $U_i + U_j = n_i n_j$, always, indicating that U_i and U_j are complementary quantities. This fact suggests that without further elaboration we define

$$r_u = \frac{U_i - U_j}{U_i + U_j} \qquad (1)$$

Clearly, the differernce between two complementary proportions, as required by Kendall's principle. Notice that if $U_j = 0$, $r_u = 1.000$; if $U_i = 0$, $r_u = -1.000$, and, if $U_i = U_j$, $r_u = 0.000$. That is, the numerical limits of r_u conform to the conventional range of a correlation index.

To illustrate the computation of r_u and its use, we may continue the example started above. Suppose now that the drawings made by six boys ($n_b = 6$) and six girls ($n_g = 6$) have been ordered, in terms of their quality, as follows: GBGGGBBBGBGB, that is, the best drawing was made by a boy, the second best by a girl, and so on. To get U_b, we simply count the number of G's preceding every B, whether a G has been counted already or not, and add up these frequencies. The result is U_b. Similarly, to obtain U_g, we count the B's preceding every G regardless of whether any B has been counted before. The sum of these frequencies is U_g. More concretely, we order B's and G's:

G	B	G	G	G	B	B	B	G	B	G	B	Sums:
	1				4	4	4		5		6	$24 = U_b$
0		1	1	1				4		5		$12 = U_g$

then, we notice that the first B is preceded only by one single G, and we place the number 1 under it; the second B from the left is preceded by four G's, and we place the number 4 under it, and so on. The sum of the first row of numbers is $U_b = 24$. The sum of the second row is $U_g = 12$. These two quantities represent U computed with respect to the boys, and with respect to the girls, respectively. Now we desire to estimate r_u with respect to the boys and so we let $U_i = U_b$, and $U_j = U_g$. Applying formula (1) gives, $r_u = (24 - 12)/(24 + 12) = .333$. If we wanted to estimate r_u with respect to the girls, we would let $U_i = U_g$, and $U_j = U_b$. Since we have simply reversed the labels, the magnitude of r_u will be identical, but of opposite sign.

The significance of r_u. Since r_u is entirely dependent on the quantities U_i and U_j, it is enough to use elementary algebra to arrive at

$$U = \frac{n_i n_j}{2} (1 - r_u) \qquad (2)$$

which is distributed as Mann-Whitney's U and is subject to the same interpretation. For our example,

$$U = \frac{6 \times 6}{2} (1 - .333)$$
$$= 12$$

the smaller out of two U's, with probability $p > .15$, suggesting that $r_u = .333$

is not significantly different from what we would have obtained from a random arrangement of the letters G and B, as a group.

An empirical comparison between r_{pb} *and* r_u. In order to determine whether r_u behaves at all like r_{pb}, the point biserial, to which it is presumably analogous, we used a 75-item test which had been administered to 904 applicants for admission to the Ateneo de Manila University. The data were intended for item analysis, and making use of the fact that cumulative frequencies behave like ranks ,we decided to obtain both r_{pb} and r_u. A Pearson's r between these two measures came up to $r = .875$, a rather high degree of agreement, suggesting that both statistics are measuring basically the same thing. Figure 1 illustrates the relationship in a graphic manner. As can be seen, the regression of r_u on r_{pb} is essen-

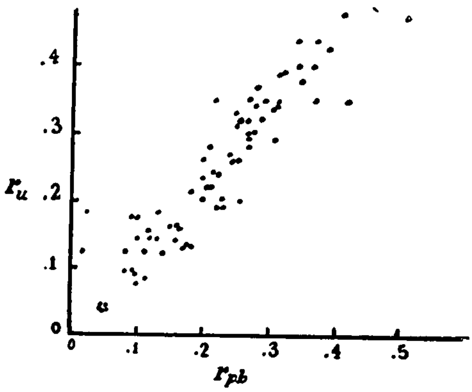

FIGURE 1. REGRESSION OF r_u ON r_{pb}. PEARSON'S r IS, $r = .875$ $p < .01$ WITH $N = 75$.

tially linear. In general, r_u tends to be consistently larger than r_{pb}. From data gathered from smaller samples, it looks like this difference is inversely proportional to the size of the sample. Although we have not studied the matter systematically, it appears like r_u is ordinarily larger — as they both depart from zero — by a factor $1/m$, where m is the number of cases in the smaller of two groups with n_i and n_j cases respectively. Otherwise, there is little doubt that at least for this set of data (75 correlations in all), r_u mimics rather well the behavior of r_{pb}.

It seems, then, that r_u can be considered as a rank-biserial coefficient of correlation, analogous to an r_{pb} and it offers the following advantages: it is not subject to the assumption of normality prescribed for a correct interpretation of r_{pb}; r_u can be used both with data which come naturally on an ordinal scale, or which can be reduced artificially to the form of ranks; formula (2) provides a useful transformation, so that the significance of r_u can be evaluated against a specified sampling distribution — the distribution of U. All these characteristics ought to make it an attractive substitute for r_{pb} in situations where the latter does not apply.

REFERENCES

HAYS, W. G. *Statistics for psychologists.* New York: Holt-Rinehart, 1962.

JOHNSON, P. O. *Statistical methods in research.* Englewood Cliffs, N.J.: Prentice-Hall, 1949.

KENDALL, M. G. *Rank correlation methods.* (3rd ed.) New York: Hafner, 1962.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

WALSH, J. E. *Handbook of nonparametric statistics.* New York: Van Nostrand, 1962.