PSYCHOMETRIKA—VOL. 48, NO. 3,
SEPTEMBER, 1983
NOTES AND COMMENTS

ASYMPTOTIC CLARIFICATIONS, GENERALIZATIONS, AND CONCERNS REGARDING AN EXTENDED CLASS OF MATCHED PAIRS TESTS BASED ON POWERS OF RANKS

PAUL W. MIELKE, JR. AND KENNETH J. BERRY

COLORADO STATE UNIVERSITY

Recent studies pertaining to an extended class of matched pairs tests based on powers of ranks are discussed. Previous questions regarding the asymptotic properties for this class of tests are clarified and a generalization of this class is described. This generalization raises a previously unanticipated concern about whether or not the analytic comparisons resulting from these tests correspond with an intuitive notion of what is being compared.

Key words: asymptotic nonnormality, congruence principle, nonparametric tests, permutation tests, robust procedures, Wilcoxon signed-ranks test.

Introduction

Mielke and Berry [1976] described a class of linear rank tests based on N nonzero matched-pair differences. The statistic underlying these tests is given by

$$T = \sum_{I=1}^{N} \Psi(I) Z_I$$

where $\Psi(I)$ is the score of the Ith absolute value ordered from below, $\Psi(I) = I^v$, $v > -\frac{1}{2}$, and Z_I is 1 if the corresponding difference is positive and 0 if it is negative. The null hypothesis specifies that $P(Z_I = 1) = P(Z_I = 0) = \frac{1}{2}$ for $I = 1, \ldots, N$. The primary purpose of this note is to relate the findings of recent studies to these tests.

Asymptotic Clarifications

Based on related work involving two-sample tests [Mielke, 1972, 1974], the asymptotic properties were known to be optimum in some vague sense for $v > -1$. However, the results of Mielke and Berry [1976] were confined to $v > -\frac{1}{2}$ because of unresolved questions. These questions have recently been clarified [Mielke & Sen, 1981]. When $v > -1$, each test among this class is a locally most powerful rank test (LMPRT) relative to the distribution parameterized by v at the top of page 94 in Mielke and Berry [1976]. While each test is a LMPRT, this class is partitioned into the following three disjoint cases: (i) when $v > -\frac{1}{2}$, the optimum distribution associated with each test possesses a finite Fisher information and the test is asymptotically most powerful for (Pitman type sequences of) local (contiguous) alternatives (AMPLA) where each test statistic has an asymptotically normal null distribution (ANND); (ii) when $v = -\frac{1}{2}$, the test is not necessarily AMPLA (Fisher information is infinite) but does have the ANND property; and (iii) when $-1 < v < -\frac{1}{2}$, each test is not necessarily AMPLA and also does not possess the ANND property. In particular, the asymptotic skewness ($\gamma_{1,v}$) and the asymptotic

Reprint requests should be addressed to Paul W. Mielke, Jr., Department of Statistics, Colorado State University, Fort Collins, CO., 80523.

0033-3123/83/0900-0583$00.50/0

kurtosis ($\gamma_{2,v}$) of T when $v < -\frac{1}{2}$ are $\gamma_{1,v} = 0$ and

$$\gamma_{2,v} = \frac{-2 \sum_{j=1}^{\infty} j^{4v}}{\left(\sum_{j=1}^{\infty} j^{2v}\right)^2}$$

(when $v \geq -\frac{1}{2}$, $\gamma_{1,v} = \gamma_{2,v} = 0$).

Generalizations and Concerns

A more general class of matched-pairs permutation tests has recently been considered [Mielke & Berry, 1982]. The test statistic underlying this class is given by

$$\delta = \binom{N}{2}^{-1} \sum_{I<J} |(2Z_I - 1)\Psi(I) - (2Z_J - 1)\Psi(J)|^w$$

where $w > 0$ and $\sum_{I<J}$ denotes the sum over all I and J such that $1 \leq I < J \leq N$. In particular, T is a special case of δ when $w = 2$ since

$$\delta = \binom{N}{2}^{-1} \left\{ N \sum_{I=1}^{N} [\Psi(I)]^2 - \left[2T - \sum_{I=1}^{N} \Psi(I) \right]^2 \right\}.$$

This association between δ and T raises a previously unanticipated concern. The intuitive notion of shifts among the signed scores involves comparisons of all $\binom{N}{2}$ pairs of values. For any specific pair of signed scores, the comparison may be visualized as the simple euclidean distance between the two scores (i.e., the comparison is based on $w = 1$ in the expression given for δ). The collection of observed signed scores upon which this comparison is based is the *data space*. On the other hand, the choice of w specifies the type of *analysis space* associated with δ. The data space and the analysis space are congruent only when $w = 1$. Since T is based on $w = 2$, it is not surprising that the intuitive euclidean comparisons among signed scores may differ drastically from those of the test statistic T.

The notion of T being based on $w = 2$ rather than $w = 1$ was completely unanticipated by Mielke and Berry [1976]. This observation is most disturbing since it suggests that many commonly used tests are providing comparisons which may not correspond with the intuitive notion of what is actually being compared. This concern also involves two-sample rank tests and one-way analysis of variance [Mielke, Berry, & Medina, 1982].

The fact that the data space and the analysis space of many matched pairs tests are not compatible is disturbing. Thus a secondary purpose of this note is to raise a question about the choice of w. The principal reason that matched pairs tests are based on $w = 2$, such as the Wilcoxon signed-ranks test, is the assumption of a normal distribution underlying the maximum likelihood approach. When the normality assumption is removed (as it is with permutation tests), the necessity of $w = 2$ is also removed. In the context of permutation tests, the case for another value of w might be stronger than the case for $w = 2$. Mielke and Berry [1982] provide simulated power comparisons which suggest that substantial benefits may be attained when $w = 1$ is used. The implication for statistical inference is that matched pairs tests based on $w = 2$ are only capable of detecting alternatives characterized by mean score location shifts. It is somewhat presumptuous to suppose that the only alternatives of importance are those limited strictly to a mean score location shift. In this light, it should be noted that matched pairs tests based on $w = 1$ are omnibus tests capable of detecting alternatives which include mean score location shifts as well as score concentration changes. A primary reason for developing robust procedures appears to be the restricted alternatives associated with $w = 2$ (a more complex restriction than one might anticipate).

REFERENCES

Mielke, P. W. Asymptotic behavior of two-sample tests based on powers of ranks for detecting scale and location alternatives. *Journal of the American Statistical Association*, 1972, *67*, 850–854.

Mielke, P. W. Squared rank test appropriate to weather modification cross-over design. *Technometrics*, 1974, *16*, 13–16.

Mielke, P. W. & Berry, K. J. An extended class of matched pairs tests based on powers of ranks. *Psychometrika*, 1976, *41*, 89–100.

Mielke, P. W. & Berry, K. J. An extended class of permutation techniques for matched pairs. *Communications in Statistics-Theory and Methods*, 1982, *All*, 1197–1207.

Mielke, P. W., Berry, K. J., & Medina, J. G. Climax I and II: Distortion resistant residual analyses. *Journal of Applied Meteorology*, 1982, *21*, 788–792.

Mielke, P. W. & Sen, P. K. On asymptotic non-normal null distributions for locally most powerful rank test statistics. *Communications in Statistics-Theory and Methods*, 1981, *A10*, 1079–1094.

Manuscript received 4/15/83
Final version received 5/10/83