Research Journal of Mathematics and Statistics 1(1): 27-29, 2009
ISSN: 2040-7505

Submitted Date: July 16, 2009 Accepted Date: August 17, 2009 Published Date: September 23, 2009

Simple K_Sample Rank Tests for Umbrella Alternatives

Sharada V. Bhat
Department of Statistics, Karnatak University, Dharwad-580003, India

Abstract: Distribution-free tests are proposed to test the homogeneity of k-samples against umbrella alternatives when the point of umbrella is known. The tests are simple to compute and hence are useful for practical purposes. They are based on two-sample Mann-Whitney U-statistics and can be treated as the competitors to a test due to Mack and Wolfe. The distributional properties of the tests have been derived and their performances in terms of Pitman asymptotic relative efficiency are discussed.

Key words: Asymptotic relative efficiency, distribution-free, efficacy, homogeneity, U-statistics

INTRODUCTION

Let X_{ij} be k independent random samples with X_{ij}, $j=1,\ldots,n_i$ having absolutely continuous distribution function $F_i(x) = F(x-\theta_i)$. For such a data, we are often interested in testing the hypothesis that all the k-samples come from a single common distribution, that is, the null hypothesis is

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k \tag{1}$$

against the alternative

$$H_1 : \theta_1 \le \theta_2 \le \ldots \le \theta_{l-1} \le \theta_l \ge \theta_{l+1} \ge \ldots \ge \theta_{k-1} \ge \theta_k \tag{2}$$

with atleast one strict inequality and 'l' is the peak or point of umbrella. These kinds of alternatives are one way analogs to a quadratic regression settings and special case where 1 is known to be at the kth population corresponds to the ordered alternatives. They are appropriate while evaluating marginal gain in performance efficiency as a function of degree of training, effectiveness of a drug as a function of time, growth of human beings as a function of age, yield of a crop as a function of quantity of fertilizer applied, reaction to increasing dosage of drug, etc.

Archambault, Mack and Wolfe (1978) introduced these kinds of alternatives for the first time and considered them for the application of their results proposed for the ordered alternatives. Mack and Wolfe (1981) proposed k-sample rank tests based on Mann-Whitney (1947) two-sample U-Statistics, Simpson and Margolin (1986) suggested a general recursive test procedure, Hettmansperger and Norton (1987) proposed the tests based on ranks, Shi (1988) developed a maximin efficient linear rank test, Chen and Wolfe (1990) proposed the tests which are natural extension of Chacko (1963) rank statistics and Mack and Wolfe (1981) test statistics, Shetty and Bhat (1997) proposed a three-sample distribution-free test, Kossler (2006) developed tests using arbitrary scores and Bhat and Patil (2008) proposed a class of k-sample distribution-free tests based on the weighted linear combination of consecutive two-sample U-statistic having a kernel of sub-sample medians.

In this research, k-sample distribution-free tests based on two-sample Mann-Whitney U-statistics are proposed and it is assumed that the '1' is known.

MATERIALS AND METHODS

This study is an attempt to develop simple k-sample rank tests for umbrella alternatives and has been carried out during the first few months of 2009. Here, two tests are proposed, their null mean and null variances are obtained.

The proposed tests: For the above problem of testing (1) against (2), the following test statistics are proposed.

$$A = \sum_{i=1}^{l-1} a_i U_{i,i+1} + \sum_{i=1}^{k-1} a_i U_{i+1,i} \tag{3}$$

where $a_1, \ldots, a_{k-1}$ are some real constants to be chosen suitably. U_{ij}. is the two-sample Mann-Whitney U-statistic for the ith and jth samples, where

$$U_{ij} = (1/n_i n_j) \sum_C h(X_i, Y_j). \tag{4}$$

C denotes summation over all $\binom{n_i}{1}\binom{n_j}{1}$ combinations of ith and jth sample observations and

$$h_{ij}(x,y) = \begin{cases} 1 & if\ x \le y \\ 0 & otherwise \end{cases} \tag{5}$$

By putting $a_i=1$ for i=1,2,…,k-1 in (3), we have

$$B = \sum_{i=1}^{l-1} U_{i,i+1} + \sum_{i=1}^{k-1} U_{i+1,i} \tag{6}$$

The distribution of tests: Under H_0, the mean of the test statistics are given by

$$\mu_A = EA = \left(\tfrac{1}{2}\right)\sum_{i=1}^{l-1} a_i + \left(\tfrac{1}{2}\right)\sum_{i=1}^{k-1} a_i = \left(\tfrac{1}{2}\right)\sum_{i=1}^{k-1} a_i \qquad (7)$$

and $\mu_A = EA = {(k-1)}/{2}$. (8)

From Lehmann (1963) and Puri (1965), the null distribution of $\sqrt{N}(A-\mu_A)$ follows asymptotic normal distribution with mean zero and variance ξ_A, where

$$\xi_A = \mathrm{Var}(A) = \xi_1 + \xi_2 + 2\xi_{12}, \qquad (9)$$

$$\xi_1 = Var(A_1) = \underline{a}_1' \textstyle\sum_1 \underline{a}_1,\ \xi_2 = Var(A_2) = \underline{a}_2' \textstyle\sum_2 \underline{a}_2 \qquad (10)$$

$$A_1 = \sum_{i=1}^{l-1} a_i U_{i,i+1}, \quad A_2 = \sum_{i=1}^{k-1} a_i U_{i+1,i} \qquad (11)$$

$$\underline{a}_1 = a_1, \dots, a_{l-1}, \quad \underline{a}_2 = a_l, \dots, a_{k-1}$$

$$\sum\nolimits_1 = \begin{cases} ((1/\lambda_i) + (1/\lambda_{i+1}))\xi_{10} & i=j=1,\dots,l-1 \\ -(1/\lambda_{i+1})\xi_{10} & i=1,\dots,l-2,\ j=i+1 \\ -(1/\lambda_i)\xi_{10} & i=1,\dots,l-1,\ j=i-1 \\ 0 & Otherwise \end{cases}$$

$$\sum\nolimits_2 = \begin{cases} ((1/\lambda_i) + (1/\lambda_{i+1}))\xi_{10} & i=j=1,\dots,k-1 \\ -(1/\lambda_{i+1})\xi_{10} & i=l,\dots,k-2,\ j=i+1 \\ -(1/\lambda_i)\xi_{10} & i=l+1,\dots,k-1,\ j=i-1 \\ 0 & Otherwise \end{cases}$$

$$\xi_{12} = \mathrm{Cov}(A_1, A_2) = a_{l-1}\, a_l\, \mathrm{Cov}(U_{l-1,l}, U_{l+1,l}) = (a_{l-1}\, a_l/\lambda_l)\,\xi_{10}, \qquad (12)$$

$$\lambda_i = \lim_{N\to\infty} (n_i/N),\ N = \sum_{i=1}^{k} n_i,\ 0<\lambda_i<1\ and\ \xi_{10} = (1/12).$$

When sample sizes are equal, that is $\lambda_1 = \lambda_2 = \dots = \lambda_k = (1/k)$, substituting (10) and (12) in (9), we have

$$\xi_A = (k/6)\left(\sum_{i=1}^{k-1} a_i^2 - \sum_{i=1}^{k-1} a_i a_{i+1} + 2a_{l-1}a_l\right) \qquad (13)$$

And on similar lines, the null distribution of $\sqrt{N}(B-\mu_B)$ follows asymptotic normal distribution with mean zero and variance $\xi_B = (k/2)$. .

Optimal choice of weights: Under the sequence of Pitman alternatives, the efficacy of the test statistic A is given by

$$eff(A) = \left(\sum_{i=1}^{k-1} a_i^2\right)\left(\int_{-\infty}^{\infty} f^2(y)\,dy\right)\Big/\xi_A. \qquad (14)$$

For efficiency comparisons, equal sample sizes and equally spaced alternatives of the type

$$\theta_i = \begin{cases} i\theta & i=1,\dots,l \\ (2l-i)\theta & i=l+1,\dots,k \end{cases},\ \theta>0 \qquad (15)$$

are considered. From Rao (1973), the optimum weights a_i^* for which A has maximum efficacy is given by

$$a_i^* = \begin{cases} (i/k)(k(l-i)-(l-1)) & i=1,\dots,l-1 \\ ((k-i)/k)(k(i+1-l)-(l-1)) & i=l,\dots,k-1 \end{cases}$$

for odd k and $l=(k+1)/2$. (16)

The square of the efficacy of the test with optimal choice of weights is given by

$$e^*(A) = \left((k^4+2k^2-3)/4k^2\right)\left(\int_{-\infty}^{\infty} f^2(y)\,dy\right)^2 \qquad (17)$$

And the square of the efficacy of B is given by

$$e(B) = \left(2(k-1)^2/k\right)\left(\int_{-\infty}^{\infty} f^2(y)\,dy\right)^2 \qquad (18)$$

RESULTS AND DISCUSSIONS

The asymptotic relative efficiency of A with respect to any other test * is calculated using

$$\mathrm{ARE}(A,*) = e(A)/e(*). \qquad (19)$$

As the test due to Mack and Wolfe (1981), say M has been compared with many other tests in the literature and the new tests developed here too depend on two-sample Mann-Whitney test statistics, my interest lies in comparing A,B with M. It is observed that,

$$ARE(A,M) = (k^4 + 2k^2 - 3)/8k(k-1)^2 \qquad (20)$$

Table 1: ARE (A, M) for different values of k

k	ARE(A,M)
3	1.000000
5	1.050000
7	1.238095
9	1.458333
11	1.690909
13	1.929487
15	2.171429
17	2.415441
19	2.660819
21	2.907143
23	3.154150
25	3.401667

$$ARE(B,M) = 1 \tag{21}$$

and $ARE(A,B) = ARE(A,M)$ (22)

The $ARE(A,M)$ depends only on k but not on the underlying distribution and these values for different values of k are given in Table 1. From Table 1, one can see that, the test A performs better than M for $k \geq 3$ and ARE increases as k increases. From (21) and (22), it is observed that, the performance of B and M are asymptotically equivalent and A performs better than B too.

CONCLUSIONS

From the above discussions, it is clear that the newly proposed test A is more efficient than M and B is equally efficient to M, for $k \geq 3$ from Pitman's asymptotic relative efficiency approach.

From practical view point, one may note that, the usage of A and B tests are easier, simple and less time consuming when compared to M for the following reason:

"To use A and B tests, it is enough if (k-1) two-sample Mann-Whitney U-statistics(U_{ij}) are computed, where as, to use M test, one has to compute 2(k-2) U_{ij}s. For example, when k=7, to use A and B tests, 6 U_{ij}s have to be computed, where as, to use M, 10 U_{ij}s will have to be computed. As k becomes larger, the computation of U_{ij}s to use M will be much more than those to use A and B".

Hence from both computational and Pitman asymptotic relative efficiency aspects, the newly proposed tests are preferable to Mack and Wolfe's test for testing homogeneity of k-samples against umbrella alternatives when the point of umbrella is known.

REFERENCES

Archambault, W.A., G.A. Mack and D.A. Wolfe, 1978. K-sample rank tests using pair-specific scoring functions. Can. J. Stat., 5: 195-207.

Bhat, S.V. and A.B. Patil, 2008. A class of k-sample distribution-free tests for umbrella alternatives. Int. J. Agric. Stat. Sci., 4: 53-61.

Chacko, V.J., 1963. Testing homogeneity against ordered alternatives. Ann. Math. Stat., 34: 945-956.

Chen, Y.I. and D.A. Wolfe, 1990. A study of distribution-free tests for umbrella alternatives. Biometric. J., 32: 47-57.

Hettmansperger, T.P. and R.M. Norton, 1987. Tests for patterned alternatives in k-sample problems. J. Am. Stat. Assoc., 82: 292-299.

Kossler, W., 2006. Some c-sample rank tests of homogeneity against umbrella alternatives with unknown peak. J. Stat. Comp. Simul., 76: 57-74.

Lehmann, E.L., 1963. Robust estimation in analysis of variance. Ann. Math. Stat., 34: 957-966.

Mack, G.A. and D.A. Wolfe, 1981. K-sample rank tests for umbrella alternatives, J. Am. Stat. Assoc., 76: 175-181.

Mann, H.B. and D.R. Whitney, 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat., 18: 50-60.

Puri, M.L., 1965. Some distribution-free k-sample rank order test of homo- geneity against ordered alternatives. Comm. Pure Applied Math., 18: 51-53.

Rao, C.R., 1973. Linear Statistical Inference and Its Applications. John Wiley and Sons, New York.

Shetty, I.D. and S.V. Bhat, 1997. A distribution-free test for umbrella alternatives based on sub-sample medians. J. Karnatak University Science, pp: 86-94.

Shi, N.Z., 1988. Rank test statistics for umbrella alternatives. Comm. Stat. Theory and Methods, 17: 2059-2073.

Simpson, D.G. and B.H. Margolin, 1986. Recursive nonparametric testing for dose-response relationships subject to downturns at high doses. Biometrika, 73: 589-596.