Page Type Test for Ordered Alternatives
on Multiple Ranked Set Samples.

Donghee Kim[1], Youngcheol Kim[2], Hyungee Kim[3]

Abstract

In this paper, we propose the test statistic for ordered alternatives on multiple ranked set samples. Since the proposed test statistic is Page type, its asymptotic properties are easily obtained. From the simulation works, we calculate the power of test statistic(P_{RSS}) under the underlying distributions, such as uniform, normal, double exponential, logistic and Cauchy distributions.

1. Introduction

Ranked-set sampling(RSS) method is useful where measurements are destructive or expensive while ranking of the observations is relatively easy.

McIntyre(1952) used the RSS technique to assess yields of pasture plots without actually carrying out the time consuming process of moving and weighting the hay for a large number of plots. Dell and Clutter(1972) considered a useful technique for improving estimates of the mean in the situation that measurements of the observation are difficult but ranking of the observations is relatively easy. Stokes(1980) investigated the estimation of the population variance and the asymptotic relative efficiency on RSS. In one sample location problem, Hettmansperger(1995) studied the sign test on RSS and Kim and Kim(1996) considered the Wilcoxon signed rank test on RSS. In two sample location problem, Bohn and Wolfe(1992, 1994) proposed Mann-Whitney-Wilcoxon statistic and investigated the properties of test procedures based on RSS, for perfect and imperfect judgement, respectively.

In this paper, we deal with multiple(c-sample) ranked set samples. For convenience, the sample structure is as follows.

1) Professor, Department of Statistics, (Research Institute of Computer, Information and Communication),
Pusan National University, Pusan, 609-735, Korea.
2) Instructor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
3) Department of Statistics, Pusan National University, Pusan, 609-735, Korea.

population	cycle					p.d.f

population 1, cycle columns: $1 \quad 2 \quad \cdots \quad n_1$

$$\left(\begin{array}{cccc} X_{1(1)1} & X_{1(1)2} & \cdots & X_{1(1)n_1} \\ X_{1(2)1} & X_{1(2)2} & \cdots & X_{1(2)n_1} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1(k_1)1} & X_{1(k_1)2} & \cdots & X_{1(k_1)n_1} \end{array} \right) \quad \begin{array}{c} \sim \\ \sim \\ \vdots \\ \sim \end{array} \quad \begin{array}{c} f_{(1)} \\ f_{(2)} \\ \vdots \\ f_{(k_1)} \end{array}$$

population 2, cycle columns: $1 \quad 2 \quad \cdots \quad n_2$

$$\left(\begin{array}{cccc} X_{2(1)1} & X_{2(1)2} & \cdots & X_{2(1)n_2} \\ X_{2(2)1} & X_{2(2)2} & \cdots & X_{2(2)n_2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{2(k_2)1} & X_{2(k_2)2} & \cdots & X_{2(k_2)n_2} \end{array} \right) \quad \begin{array}{c} \sim \\ \sim \\ \vdots \\ \sim \end{array} \quad \begin{array}{c} f_{(1)} \\ f_{(2)} \\ \vdots \\ f_{(k_2)} \end{array}$$

population c, cycle columns: $1 \quad 2 \quad \cdots \quad n_c$

$$\left(\begin{array}{cccc} X_{c(1)1} & X_{c(1)2} & \cdots & X_{c(1)n_c} \\ X_{c(2)1} & X_{c(2)2} & \cdots & X_{c(2)n_c} \\ \vdots & \vdots & \vdots & \vdots \\ X_{c(k_c)1} & X_{c(k_c)2} & \cdots & X_{c(k_c)n_c} \end{array} \right) \quad \begin{array}{c} \sim \\ \sim \\ \vdots \\ \sim \end{array} \quad \begin{array}{c} f_{(1)} \\ f_{(2)} \\ \vdots \\ f_{(k_c)} \end{array}$$

In population j, $j = 1, \cdots, c$, we assume that each sample has a continuous distribution with cdf $F_j(x) = F(x - \theta_j)$ and its pdf $f_j(x) = f(x - \theta_j)$, where θ_j is a location parameter of the j-th population. Our goal of this paper is to propose the test statistic for testing $H_0: \theta_1 = \cdots = \theta_c (= \theta_0)$ against $H_1: \theta_1 \leq \cdots \leq \theta_c$ with at least one strict inequality. Our proposed test statistic is to use the correlation between population number as a natural order and the rank of median in each cycle and population, so it is Page type test. In section 2, we propose the test statistic under null distribution. Section 3, we deal with the asymptotic properties of the proposed test statistic and efficacies of the proposed test statistic. Finally, section 4 has simulation design and results under several underlying distribution. Short conclusions are in section 5.

2. Test Statistic

Let $X_{j(1)1}, \cdots, X_{j(1)n_j}; \cdots; X_{j(k_j)1}, \cdots, X_{j(k_j)n_j}$ be a ranked set sample of size $k_j n_j$ for a continuous distribution with cdf $F_j(x) = F(x - \theta_j)$ and pdf $f_j(x) = f(x - \theta_j)$, $j = 1, \cdots, c$, where θ_j is a location parameter of the $j - th$ population, and let k_j and n_j be the sample size and independent times cycle size to get $k_j n_j$ observations from $k_j^2 n_j$ preranking sample observations.

In this paper, we consider the testing problem for testing $H_0: \theta_1 = \cdots = \theta_c (= \theta_0)$ against $H_1: \theta_1 \le \cdots \le \theta_c$ with at least one strict inequality.

Our proposed test statistic is of the form,

$$P_{RSS} = \sum_{j=1}^{c} \sum_{i=1}^{n_i} jR_{ij} ,$$

where R_{ij} is rank of $Med(X_{j(1:k_j)i})$ among $Med(X_{j(1:k_j)i})$, $j = 1, \cdots, c$ and $Med(X_{j(1:k_j)i})$ is Median of $(X_{j(1)i}, \cdots, X_{j(k_j)i})$.

We can easily find that it is Page type test, we denote the pdf of $Med(X_{j(1:k_j)i})$, $i = 1, \cdots, n_j$, $j = 1, \cdots, c$, by $f_{(l)}(t)$, where l is median of $(1, \cdots, k_j)$.

We easily obtain the mean and variance of the proposed test statistic by Randle and Wolfe(1979). We consider here the case when cycle size and sample size are $n_j = n$, $k_j = k$, $j = 1, \cdots, c$ and specially k is odd.

Corollary. (Randles and Wolfe(1979))

Under $H_0: \theta_1 = \cdots = \theta_c (= \theta_0)$, the limiting $(N \to \infty)$ null distribution of $(P_{RSS} - E_0(P_{RSS})) / \sqrt{Var_0(P_{RSS})}$ is standard normal,

where $E_0(P_{RSS}) = \dfrac{nc(c+1)^2}{4}$, $Var_0(P_{RSS}) = \dfrac{nc^2(c+1)^2(c-1)}{144}$, $N = nc$.

3. The Asymptotic Properties

To obtain the asymptotic properties of the proposed test statistic, we consider when cycle size and sample size are $n_j = n$, $k_j = k$, $j = 1, \cdots, c$, specially k is odd, a sequence of translation alternative of the form, $H_{1N}: \theta_j = \theta_0 + j\triangle$, $j = 1, \cdots, c$, where $\triangle = \theta / \sqrt{N}$. Under the sequence of translation alternatives, the efficacy of the proposed test statistic is obtained as follows.

$$E_\triangle(P_{RSS}) = E_\triangle(\sum_{j=1}^{c} \sum_{i=1}^{n} jR_{ij})$$

$$= E_\triangle(\sum_{j=1}^{c} j \sum_{i=1}^{n} \sum_{j'=1}^{c} \Psi(Med(X_{j(1:k)i}) - Med(X_{j'(1:k)i}))) ,$$

where $\Psi(t) = 1$ or 0 if $\ge$, < 0.

$$= \sum_{j=1}^{c} j \sum_{i=1}^{n} \sum_{j'=1}^{c} P_\triangle \left(Med(X_{j(1:k)i}) \geq Med(X_{j'(1:k)i}) \right)$$

$$= \sum_{j=1}^{c} j \sum_{i=1}^{n} \sum_{j'=1}^{c} \int_{-\infty}^{\infty} (1 - F_{(l)}(t - (\theta_j - \theta_{j'}))) dF(t) ,$$

where l is median of sample in each cycle within $j-th$ population.
The derivative of $E_\triangle(R_{RSS})$ is

$$\frac{\partial E_\triangle(P_{RSS})}{\partial \triangle} \mid_{\triangle = 0} = \sum_{j=1}^{c} j \sum_{i=1}^{n} \sum_{j'=1}^{c} (j - j') \int_{-\infty}^{\infty} f_{(l)}(t - (j - j')\triangle) dF_{(l)}(t) \mid_{\triangle = 0}$$

$$= \sum_{j=1}^{c} j \sum_{i=1}^{n} \sum_{j'=1}^{c} (j - j') \int_{-\infty}^{\infty} f_{(l)}^2(t) dt$$

$$= n \left(\sum_{j=1}^{c} \sum_{j'=1}^{c} j^2 - \sum_{j=1}^{c} j \sum_{j'=1}^{c} j' \right) \int_{-\infty}^{\infty} f_{(l)}^2(t) dt$$

$$= n \frac{c^2(c+1)(c-1)}{12} \int_{-\infty}^{\infty} f_{(l)}^2(x) dx.$$

By the definition 5.2.14 of Randles and Wolfe(1979), we have the following efficacy of P_{RSS} in the case of $n_j = n$, i.e , $N = nc$.

$$eff^2(P_{RSS}) = \lim_{N \to \infty} \left(\frac{\frac{\partial}{\partial \triangle} E_\triangle(P_{RSS}) \mid_{\triangle = 0}}{\sqrt{N Var(P_{RSS})}} \right)^2$$

$$= c(c-1) \left(\int_{-\infty}^{\infty} f_{(l)}^2(x) dx \right)^2.$$

4. Simulation Design and Results.

We obtain the results of a small-sample Monte-Carlo simulation study of the powers in P_{RSS}. The powers are obtained under the uniform, normal, double exponential, logistic and Cauchy distribution on which these simulations are based.   The shift parameter θ has 0.0 (0.2) 1.0 . Also, we consider the design with the sample sizes $k=3, 5$, the cycle sizes $n = 3, 9$ and populations $c = 3, 5$. The results of this simulation have 1,000 replications. The IMSL(International Mathematics and Statistical Library) procedures RNUN, RNNOA and RNCHY are used to generate the required uniform, normal and Cauchy pseudo-random variates, respectively. And the double exponential and logistic random variates are generated by using the probability integral transformation.

Table 1 and 2 give empirical powers of test for $k=3, 5$, $n = 3, 9$, $c=3$ and $\alpha=0.05$ in underlying distributions, at $\theta = 0.0$, the powers are the empirical significance levels.   Table

3 and 4 give empirical powers of test for $k=3,5$, $n = 3,9$, $c=5$ and $\alpha=0.05$ in underlying distributions. When the population number c equals 3 or 5 , from the simulation results, the empirical powers of P_{RSS} increase as n and k increase. And the proposed test statistic is not more complex in mean and variance under any population sizes, sample sizes and cycle sizes.

If we consider RSS in the imperfect judgement, the proposed test statistic using median may be robust.

5. Conclusions

Our proposed test statistic is very simple, so easy to use under any population sizes, sample sizes and cycle sizes. The powers of P_{RSS} increase as n and k increase. If we apply our proposed test statistic to the imperfect judgement, since we use the median in each cycle, we think that it may be robust.

For further works, we now study the test statistic and its properties using pairwise comparision of the median in each cycle.

Table 1. Empirical powers of the proposed test statistic ($c=3$, $k=3$)

n	distribution	θ	0.0	0.2	0.4	0.6	0.8	1.0
3	Uniform		0.024	0.093	0.174	0.304	0.482	0.643
	Normal		0.031	0.098	0.223	0.427	0.628	0.771
	D.E		0.036	0.136	0.352	0.581	0.755	0.897
	Logistic		0.039	0.124	0.277	0.439	0.639	0.833
	Cauchy		0.036	0.117	0.317	0.466	0.639	0.760
9	Uniform		0.057	0.258	0.559	0.872	0.966	0.998
	Normal		0.059	0.300	0.718	0.941	0.996	1.000
	D.E		0.072	0.438	0.887	0.984	0.999	1.000
	Logistic		0.056	0.359	0.787	0.965	0.996	0.999
	Cauchy		0.053	0.431	0.769	0.950	0.996	0.996

Table 2. Empirical powers of the proposed test statistic ($c=3$, $k=5$)

n	distribution	θ	0.0	0.2	0.4	0.6	0.8	1.0
3	Uniform		0.033	0.089	0.229	0.406	0.595	0.772
	Normal		0.038	0.126	0.304	0.556	0.773	0.890
	D.E		0.035	0.188	0.470	0.777	0.912	0.969
	Logistic		0.028	0.162	0.380	0.635	0.819	0.929
	Cauchy		0.036	0.182	0.409	0.671	0.822	0.921
9	Uniform		0.065	0.315	0.709	0.949	0.993	1.000
	Normal		0.070	0.434	0.870	0.991	1.000	1.000
	D.E		0.068	0.613	0.970	0.999	1.000	1.000
	Logistic		0.067	0.457	0.897	0.997	1.000	1.000
	Cauchy		0.071	0.562	0.930	0.997	0.999	1.000

Table 3. Empirical powers of the proposed test statistic ($c=5$, $k=3$)

n	distribution	θ	0.0	0.2	0.4	0.6	0.8	1.0
3	Uniform		0.050	0.100	0.192	0.304	0.434	0.593
	Normal		0.040	0.117	0.227	0.407	0.568	0.734
	D.E		0.046	0.155	0.332	0.568	0.762	0.869
	Logistic		0.052	0.136	0.269	0.441	0.605	0.796
	Cauchy		0.043	0.122	0.275	0.472	0.636	0.766
9	Uniform		0.055	0.176	0.399	0.679	0.859	0.960
	Normal		0.057	0.231	0.529	0.801	0.950	0.998
	D.E		0.059	0.325	0.727	0.952	0.995	0.999
	Logistic		0.047	0.243	0.596	0.882	0.979	0.998
	Cauchy		0.057	0.273	0.640	0.880	0.959	0.991

Table 4. Empirical powers of the proposed test statistic ($c=5$, $k=5$)

n	distribution	θ	0.0	0.2	0.4	0.6	0.8	1.0
3	Uniform		0.046	0.119	0.223	0.396	0.579	0.745
	Normal		0.049	0.131	0.308	0.520	0.741	0.883
	D.E		0.039	0.198	0.485	0.759	0.914	0.973
	Logistic		0.044	0.148	0.345	0.618	0.819	0.912
	Cauchy		0.061	0.185	0.412	0.691	0.831	0.913
9	Uniform		0.041	0.223	0.499	0.806	0.945	0.990
	Normal		0.049	0.288	0.697	0.913	0.992	1.000
	D.E		0.058	0.449	0.880	0.995	1.000	1.000
	Logistic		0.059	0.335	0.766	0.960	0.999	0.999
	Cauchy		0.050	0.374	0.831	0.985	1.000	1.000

References

[1] Bohn, L. L. and Wolfe, D. A.(1992). Nonparametric two-sample procedures for ranked-set samples data, *Journal of the American Statistical Association*, 87(418), 552-561.

[2] Bohn, L. L. and Wolfe, D. A.(1994). The effect of imperfect judgement rankings on properties of procedures based on the ranked-set samples analog of Mann-Whitney-Wilcoxon statistic, *Journal of the American Statistical Association*, 89(425), 168-176.

[3] Dell, T. R. and Clutter, J. L. (1972). Ranked-set sampling theory with order statistics background, *Biometrics*, 28, 545-553.

[4] Hettmansperger, T. P. (1995). The ranked-set sample sign test, *Journal of the Nonparametric Statistics*, 4, 271-282.

[5] Jonckheere, A. P. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika*, 41, 133-145.

[6] Kim, D. H. and Kim, Y. C.(1996). Wilcoxon signed rank test using ranked-set sample, *The Korean Journal of Computational and Applied Mathematics*, 3(2), 235-243.

[7] Kim, D. H. and Kim, Y. C.(1998). Two-sample comparison using sign test on ranked-set samples, *The Korean Journal of Computational and Applied   Mathematics*, 5(1), 262-268.

[8] McIntyre, G. A.(1952). A method of unbiased selective sampling using ranked sets, *Australian Journal of Agricultural Research*, 3, 385-390.

[9] Randles, R. H. and Wolfe, D. A.(1979). *Introduction to the Theory of Nonparametric Statistics*, New York : John Wiley.

[10] Stokes, S. L.(1980). Estimation of variance using judgement ordered ranked-set samples, *Biometrics*, 36, 35-42.

[11] Stokes, S. L. and Sager(1988). Characterization of a ranked-set sample with application to estimating distribution function, *Journal of the American Statistical Association*, 83, 374-381.