The Korean Communications
in Statistics Vol. 4, No. 2, 1997
pp. 421 - 426

Comparison of Confidence Subsets for Umbrella Orderings[1)]

Dong Hee Kim[2)] and Young Cheol Kim[3)]

Abstract

This paper proposes a distribution-free procedure that obtain confidence subset for umbrella orderings. We compare the proposed confidence procedure with Pan's(1996) confidence procedure.

1. Introduction

There are many practical situations for testing umbrella orderings in biology, agriculture, economics, etc.. As an example, we could suspect that with increasing dose, the response effect tends to improve, but that after some point tends to mean diminishing the response effect. An umbrella ordering contains all the unknown peaks. Our aim is to identify the treatments that correspond to the optimal effects.

Some nonparametric tests for umbrella orderings have been discussed by many authors. Mack and Wolfe(1981) proposed a distribution-free test for umbrella alternatives by combining Jonckheere(1954) statistic and a reverse Jonckheere statistic. Simpson and Margolin(1986) proposed a recursive nonparametric test for dose-response relationship subject to downturns at high doses. On the other hand, Hettmansperger and Norton(1987) proposed a nonparametric method based on linear rank statistic and offered a comparison with Mack and Wolfe(1981) test by umbrella pattern. Shi(1988a,1988b) derived the likelihood ratio test in the normal case and proposed an optimal rank test which was called a maximin efficient linear rank test. Chen and Wolfe(1990) proposed a natural generalization of Chacko's(1963) statistic to obtain a test for umbrella alternatives when the peak is known. Pan and Wolfe(1996) compared two different populations with umbrella orderings and developed likelihood ratio tests for testing whether two groups with umbrella orderings have the same peaks.

We consider a test for the null hypothesis that the treatment effects are all the same and

1) 본 연구는 부산대학교 기성회 재원 학술연구조성비(95. 5-99. 4)에 의한 연구임.
2) Professor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea.
3) Part Time Instructor, Department of Statistics, Pusan National University, Pusan, 609-735, Korea.

the general alternatives that the treatment effects are not all the same or the umbrella alternatives that the peak is unknown. When a test is statistically significant, i.e. the null hypothesis is rejected, we want to identify the peaks. Pan(1996) constructed a subset of the k treatments that the subset contains all the unknown optimal treatment level with confidence level $1-\alpha$. This subset is referred to as a "confidence subset". But it is not considered the informations of between the treatment at the peak and the other treatments in the Pan(1996)'s statistic.

We propose a nonparametric distribution-free confidence subset selection procedure using the informations of between the treatment at the peak and the other treatments. The statistic of this type is simple and its distributional properties are easily derived.

2. Notations and Statistics

Let $X_i=(X_{i1},\cdots,X_{in_i})$, $i=1,\cdots,t$ be t independent random samples from an absolutely continuous distribution function $F(x-\beta_i)$ where $F(0)=1/2$. Assume that the location parameters $\beta=(\beta_1,\cdots,\beta_t)$ satisfy an umbrella ordering $\beta_1\le\cdots\le\beta_p\ge\cdots\ge\beta_t$ for some unknown peak $p\in\{1,\cdots,t\}$. Also assume that an umbrella ordering have at least one peak.

Let $P(\beta)=\{p\in\{1,\cdots,t\}:\beta_1\le\cdots\le\beta_p\ge\cdots\ge\beta_t\}$ be the set of all unknown peaks of an umbrella ordering β. Let R_{ij} be the rank of X_{ij} among all the observations and let R be the rank vector of all R_{ij}. Denote a confidence subset by $S(R)$. We find $S(R)$ of all the treatment levels $\{1,\cdots,t\}$ such that $S(R)$ contains all the unknown peak in $P(\beta)$ with confidence level $1-\alpha$.

To construct a confidence subset $S(R)$ for all the unknown peaks, for a fixed $p(=1,\cdots,t)$, Pan(1996) considered the parameter and the corresponding U-statistic

$$\gamma_p=\sum_{i=1}^{p-1}\Pr\{X_i<X_{i+1}\}+\sum_{i=p}^{t-1}\Pr\{X_{i+1}<X_i\}$$

and

$$U_p(X_1,\cdots,X_t)=\sum_{i=1}^{p-1}\frac{U(X_i,X_{i+1})}{n_in_{i+1}}+\sum_{i=p}^{t-1}\frac{U(X_{i+1},X_i)}{n_{i+1}n_i},$$

where $U(X_i,X_j)=\sum_{\alpha_i=1}^{n_i}\sum_{\alpha_j=1}^{n_j}\Psi(X_{j\alpha_j}-X_{i\alpha_i})$ is the Mann-Whitney counting form between the ith and jth samples. As mentioned earlier, Pan's U-statistic is not including the informations among the pth sample and the other except $(p-1)$th and $(p+1)$th samples. Hence we propose the U-statistic that consider this informations.

For a fixed $p(=1,\cdots,t)$, we consider the parameter δ_p that is an estimable parameter of degree $(1,\cdots,1)$

$$\delta_p = \sum_{i=1,\, i\neq p}^{t} \Pr\{X_i < X_p\}.$$

An appropriate symmetric kernel is

$$h_p(X_{11}, \cdots, X_{t1}) = \sum_{i=1,\, i\neq p}^{t} \Psi(X_p - X_i)$$

where $\Psi(x) = 1$ if $x > 0$ and 0 otherwise. Therefore, the corresponding U-statistic is

$$\begin{aligned} U_p(X_1, \cdots, X_t) &= \frac{1}{n_1 \cdots n_t} \sum_{\alpha_1=1}^{n_1} \cdots \sum_{\alpha_t=1}^{n_t} h_p(X_{1\alpha_1}, \cdots, X_{1\alpha_t}) \\ &= \sum_{i=1,\, i\neq p}^{t} \frac{U(X_i, X_p)}{n_i n_p} \end{aligned}$$

where $U(X_i, X_p)$ is the Mann-Whitney counting form between the ith sample and pth sample. We use the statistics $(U_1, \cdots, U_t)$ to estimate $P(\beta)$ that is the set of all unknown peaks of an umbrella ordering β. See Pan(1996) for a details.

Under the null hypothesis $H_0: \beta_1 = \cdots = \beta_t$, $\delta_1 = \cdots = \delta_t = (t-1)/2$. The null variance of U_p for $p = 1, \cdots, t$ is

$$\begin{aligned} \sigma_0^2(U_p) &= \sum_{i=1,\, i\neq p}^{t} Var\left[\frac{U(X_i, X_p)}{n_i n_p}\right] + 2 \sum_{i=1,\, i\neq p}^{t-1} \sum_{j=i+1,\, j\neq p}^{t} Cov\left[\frac{U(X_i, X_p)}{n_i n_p}, \frac{U(X_j, X_p)}{n_j n_p}\right] \\ &= \sum_{i=1,\, i\neq p}^{t} \left[\frac{1}{12 n_i} + \frac{1}{12 n_p} + \frac{1}{12 n_i n_p}\right] + (t-1)(t-2)\frac{1}{12 n_p}. \end{aligned}$$

Theorem. Let $N = \sum_{i=1}^{t} n_i$. Under H_0, if $n_i/N \to \lambda_i$, $0 < \lambda_i < 1$ as $N \to \infty$, then $\sqrt{N}(U_p - (t-1)/2)/\sigma_0(U_p)$ is asymptotically standard normal distributed.

Proof. It can be shown by U-statistic theorem and Slutsky's theorem.

3. Confidence Subset and Example

Let $\varphi_p = \{(\beta_1, \cdots, \beta_t) \in R^t : \beta_1 \le \cdots \le \beta_p \ge \cdots \ge \beta_t\}$. Let $P_w(U \mid \varphi_p)$ be the isotonic regression of a vector $U = (U_1, \cdots, U_t) \in R^t$ onto the φ_p with respect to nonnegative weights $w = (w_1, \cdots, w_t)$ and let $w_i = 1/\sigma_0^2(U_i)$ for $i = 1, \cdots, t$. The confidence subset $S(R)$ for the unknown peaks $P(\beta)$ of an umbrella ordering β is

$$S(R) = \{i \in \{1, \cdots, t\} : \| U - P_w(U \mid \varphi_i) \|_w^2 \le c_i\},$$

where c_i is critical point such that

$$\Pr_{H_0}\{\| U - P_w(U \mid \varphi_i) \|_w^2 \le c_i,\ i = 1, \cdots, t\} = 1 - \alpha.$$

Exact critical point is obtained by exhausting all the possible permutations of the rank vector

Table 1. Critical points $c_i = c$ for balanced designs $n_i = n$

(The values in parentheses are α-values close to 0.01, 0.05 or 0.1)

	t=3	t=4	t=5	t=6
$n=1$		3.733(0.75)	7.083(0.0083) 5.056(0.05)	10.641(0.01) 8.443(0.05) 7.029(0.097)
$n=2$	5.571(0.0667) 5.250(0.1778) 5.143(0.2)	8.222(0.0095) 7.630(0.055) 7.185(0.1)	12.156(0.01) 9.477(0.05) 8.694(0.1)	
$n=3$	9.633(0.01) 8.133(0.046) 6.533(0.094)	11.08(0.01) 8.99(0.05) 7.86(0.1)		

R and computing the squared distance between the observed vector U and the φ_t for each permutation. Table 1 gives critical values of $c_1 = \cdots = c_t = c$ for a balanced design, $t=3(1)6$ and $n=1(1)3$. This constant c is determined by

$$P_{H_0}\{\max_{i=1,\cdots,t} \| U - P_w(U \mid \varphi_i) \|_w^2 \le c\} = 1-\alpha.$$

Then the corresponding confidence subset is

$$S(R) = \{i \in \{1,\cdots,t\} : \| U - P_w(U \mid \varphi_i) \|_w^2 \le c\}.$$

Pan(1996) proved that the confidence subset $S(R)$ is distribution-free and $P_\beta\{P(\beta) \subseteq S(R)\} = 1-\alpha$ for any umbrella ordering β and any continuous distribution function $F(x)$.

For $c_1 = \cdots = c_t = c$, we can reject the null hypothesis if

$$\max_{i=1,\cdots,t} \| U - P_w(U \mid \varphi_i) \|_w^2 > c,$$

Table 2. Artificial data I

n \ t	1	2	3	4	5
1	8.60	9.90	10.0	9.10	4.90
2	9.90	10.4	10.7	9.90	9.10

Table 3. Artificial data II

n \ t	1	2	3	4	5
1	9.90	10.4	10.7	9.90	9.10
2	10.1	11.3	11.4	10.6	9.30

Table 4. Artificial data III

n \ t	1	2	3	4	5
1	9.90	10.4	10.7	5.90	4.10
2	10.1	11.3	11.4	7.60	5.30

where the critical value c is the upper α percentile of the null distribution of $\max_{i=1,\cdots,t} \| U-P_w(U \mid \varphi_i) \|_w^2$. To compare the procedure using the proposed U-statistic with Pan's procedure, as an example, we consider the artificial data for $t=5$ and $n=2$.

We computed the distances D_F and D_K, where D_F and D_K are the distances $(\| U-P_w(U \mid \varphi_1) \|_w^2, \cdots, \| U-P_w(U \mid \varphi_5) \|_w^2)$ for Pan's procedure and the proposed procedure, respectively. From Pan's results, the critical values for $\alpha=0.103, 0.053$ are 9.00 and 11.81, respectively. From Table 1, we obtain that the critical values for $\alpha=0.1, 0.05$ are 8.694 and 9.477, respectively. In artificial data I, we conclude that 90% confidence subset have the peak among treatments 1 to 5 for the both procedures. Also, in artificial data II, we conclude that 90% confidence subset have the peak among treatments 1 to 4 for the both procedures. In artificial data III, 95% confidence subset using Pan's results have the peak among treatments 1 to 5, but 95% confidence subset using our results have the peak among treatments 1 to 4. We conclude with 95% confidence subset that our procedure is more informative than Pan's procedure. But our procedure may be not always more informative.

Table 5. The D_F and D_K for artificial data I, II and III

I	D_F	3.375	0.375	0.000	1.500	7.125
	D_K	4.227	0.545	0.000	3.455	7.909
II	D_F	5.063	0.375	0.000	1.500	9.600
	D_K	3.591	0.307	0.000	2.227	8.795
III	D_F	5.063	0.375	0.000	1.500	9.600
	D_K	1.773	0.136	0.000	4.682	10.50

References

[1] Chacko, V. J. (1963). Testing Homogeneity against Ordered Alternatives, *The Annals of Mathematical Statistics*, 34, 945-956.

[2] Chen, Y. I. and Wolfe, D. A. (1990). A Study of Distribution-Free Tests for Umbrella Alternatives, *Biometrical Journal*, 32(1), 47-57.

[3] Hettmansperger, T. P. and Norton, R. M. (1987). Tests for Patterned Alternatives in K-sample Problems, *Journal of the American Statistical Associations*, 82, 292-299.

[4] Jonckheere, A. P. (1954). A Distribution-Free K-sample Test against Ordered Alternatives, *Biometrika*, 41, 133-145.

[5] Mack, G. A. and Wolfe, D. A. (1981). K-sample Rank Tests for Umbrella Alternatives, *Journal of the American Statistical Associations*, 76, 175-181.

[6] Pan, G. (1996). Distribution-Free Confidence Procedure for Umbrella Ordering, The *Australian Journal of Statistics*, 38(2), 161-172.

[7] Pan, G. and Wolfe, D. A. (1996). Comparing Groups with Umbrella Orderings, *Journal of the American Statistical Association*, 91(433), 311-317.

[8] Shi, N. Z. (1988a). A Test of Homogeneity for Umbrella Alternatives and Tables of the Level Probability, *Communications in Statistics-Theory and Methods*, A17(3), 657-670.

[9] Shi, N. Z. (1988b). Rank Test Statistics for Umbrella Alternatives, *Communications in Statistics-Theory and Methods*, A17(6), 2059-2073.

[10] Simpson, D. G. and Margolin, B. H. (1986). Recursive Nonparametric Testing for Dose-Response Relationship Subject to Downturns at High Doses, *Biometrika*, 73, 589-596.