AD-A088 115   SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS COM--ETC F/G 12/1
ASYMPTOTICALLY DISTRIBUTION-FREE SIMULTANEOUS CONFIDENCE REGION--ETC(U)
FEB 80   L J WEI   F49620-79-C-0140

UNCLASSIFIED   AFOSR-TR-80-0588   NL

1 OF 1
AD
A088115

END
DATE
FILMED
9-80
DTIC

AFOSR-TR- 80-0588

LEVEL II

(5)

ASYMPTOTICALLY DISTRIBUTION-FREE SIMULTANEOUS CONFIDENCE REGION OF TREATMENT DIFFERENCES IN A RANDOMIZED COMPLETE BLOCK DESIGN *

L. J. Wei

University of South Carolina

February, 1980

AD A088115

University of South Carolina
Department of Mathematics,
Computer Science, and Statistics
Columbia, South Carolina 29208

*Research supported by the United States Air Force Office of Scientific Research under Contract No. F49620-79-C-0140.

80 8 14 119

Unclassified ... DOCUMENTATION PAGE | BEFORE COMPLETING FORM

1. REPORT NUMBER (18) (19) AFOSR-TR- 80-0588	2. GOVT ACCESSION NO. AD-A088	3. RECIPIENT'S CATALOG NUMBER 115
4. TITLE (and Subtitle) (6) ASYMPTOTICALLY DISTRIBUTION-FREE SIMULTANEOUS CONFIDENCE REGION OF TREATMENT DIFFERENCES IN A RANDOMIZED COMPLETE BLOCK DESIGN.		5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) L. J. Wei (12) 11		8. CONTRACT OR GRANT NUMBER(s) (15) F49620-79-C-0140
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of South Carolina Dept. of Math, Comp. Sci, & Statistics Columbia, SC 29208		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F (16) 2304/A5 (17) A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/nm Bolling AFB, Washington, D. C. 20332		12. REPORT DATE February, 1980
		13. NUMBER OF PAGES 9
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (11) Feb 80		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public released; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple comparisons; akugned observation; Mann-Whitney two sample statistic.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For a randomized complete block design with additive block effects, as asymptotically distribution-free simultaneous confidence region of pairwise treatment differences is presented. The correspomding confidence bound has an explicit form and is easily obtained. An example is provided for illustration purpose. The case of treatment against control is also discussed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410115

ASYMPTOTICALLY DISTRIBUTION-FREE SIMULTANEOUS CONFIDENCE REGION OF TREATMENT DIFFERENCES IN A RANDOMIZED COMPLETE BLOCK DESIGN

L. J. Wei
Department of Mathematics, Computer Science, and Statistics
University of South Carolina
Columbia, S. C. 29208

A

For a randomized complete block design with additive block effects, an asymptotically distribution-free simultaneous confidence region of pairwise treatment differences is presented. The corresponding confidence bound has an explicit form and is easily obtained. An example is provided for illustration purpose. The case of treatment against control is also discussed.

Key Words and Phrases: Multiple comparisons; aligned observation; Mann-Whitney two sample statistic.

1. INTRODUCTION

Suppose that $K > 2$ treatments are applied once each to n different blocks. Let X_{ij} be the response of the ith treatment in the jth block ($i=1, \ldots, K$; $j=1,\ldots,n$). The model often used for this experimental setting is the linear model in which the observations X_{ij} can be written as

$$X_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \tag{1.1}$$

where the α's are the parameters of interest (treatment effect), $\sum_{i=1}^{K} \alpha_i = 0$, β's are nuisance parameters (block effects) and $\underset{\sim}{e}_j = (e_{1j},\ldots,e_{Kj})'$, $j=1,\ldots,n$ are independent and identically distributed random vectors having a continuous joint distribution function which is symmetric in its K arguments (This relaxes the conventional assumption of having independence and identity of distributions of all the nK error terms.).

Oftentimes a global test for H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_K$ is of less interest and one may feel that the simultaneous inference on *the pairwise treatment* differences $\alpha_i - \alpha_{i'}$, $1 \le i < i' \le K$ is more desirable (c.f. Miller (1966, 1977)). There are several nonparametric pairwise multiple comparisons procedures available for this case (c.f. Puri and Sen (1971), Hollander and Wolfe (1973) and Hettmansperger (1975)). However, they either only utilize the intrablock comparisons and have low efficiency or involve complicated inversion procedures for obtaining simultaneous confidence bound of pairwise treatment differences. Although Sen (1969) has provided a simultaneous confidence bound to $\alpha_i - \alpha_{i'}$ based on two sample Chernoff-Savage rank order statistics, the derivation of his procedure is not obvious (see Puri and Sen (1971), p. 331), and the bound he obtained is not in an explicit form so that numerical method is sometimes required.

In this article, we utilize the information contained in interblock comparisons and provide an asymptotically distribution-free simultaneous confidence region of pairwise treatment differences. The corresponding bound has a simple and explicit form and can be easily obtained. The case of treatment against control is also discussed. An example is presented for illustration purpose in Section 3.

2. ASYMPTOTICALLY DISTRIBUTION-FREE SIMULTANEOUS CONFIDENCE REGION OF PAIRWISE TREATMENT DIFFERENCES

To eliminate the nuisance parameters β's in (1.1), we consider the aligned observations $Y_{ij} = X_{ij} - \tilde{X}_{.j}$, where $\tilde{X}_{.j}$ is a symmetric function of $X_{1j},\ldots,X_{Kj}$, such that $\tilde{X}_{.j} + a$ is the same function of $X_{1j}+ a,\ldots,X_{Kj}+ a$ for all $-\infty < a < \infty$. Typical $\tilde{X}_{.j}$ are the block average ($\bar{X}_{.j}$), the median of $X_{1j},\ldots,X_{Kj}$, the Winsorized or trimmed mean, etc. In this article, we let $\tilde{X}_{.j}$ be $\bar{X}_{.j}$ so that Model (1.1) can be rewritten as $Y_{ij} = \alpha_i + \varepsilon_{ij}$ $(i=1,\ldots,K; j = 1,\ldots,n)$, where $\varepsilon_{ij} = e_{ij} - \bar{e}_{.j}$ and $\bar{e}_{.j}$ is the jth block average of $e_{1j},\ldots,e_{Kj}$. It follows from the interchangeability of $e_{1j},\ldots,e_{Kj}$ that the distribution function of $\varepsilon_{1j},\ldots,\varepsilon_{Kj}$, $j=1,\ldots,n$ is symmetric in its K arguments. Let the marginal distribution function of ε_{ij} be G.

Now, define a scoring function ϕ for comparing Y_{ij} and $Y_{i'j'}$ by

$$\phi(Y_{ij}, Y_{i'j'}) = \begin{cases} 1, & Y_{ij} > Y_{i'j'} \\ -1, & Y_{ij} < Y_{i'j'} \,. \end{cases}$$

Let $p_{ii'} = E\phi(Y_{ij}, Y_{i'j'})$, where $i\neq i'$ and $j\neq j'$. Under H_0, $p_{ii'} = 0$. Consider $U_{ii',n} = \sum_{j=1}^{n} \sum_{j'=1}^{n} (\phi(Y_{ij}, Y_{i'j'}) - p_{ii'})$, $1 \le i < i' \le K$, which is the usual Mann-Whitney two sample statistic (though based on matched samples).

Several lemmas are needed to derive the simultaneous confidence bound to $\alpha_i-\alpha_{i'}$, $1 \le i < i' \le K$.

Lemma 1. As $n \to \infty$, the random vector $\langle n^{-3/2} U_{ii',n}/2 \rangle$ converges in distribution to a K(K-1)/2 dimensional normal random vector with mean $\underset{\sim}{0}$ and covariance matrix $\ddagger$.

(Proof). See Appendix A.

Lemma 2. Under H_0, the asymptotical covariance structure of $\left\langle \frac{n^{-3/2} U_{ii',n}}{2(\sigma^2-\tau)^{1/2}} \right\rangle$ is identical to that of the vector having K(K-1)/2 components $Z_i - Z_{i'}$, $1 \le i < i' \le K$, where $(Z_1, \ldots, Z_K)' \sim N_K(\underset{\sim}{0}, I_K)$, I_K is the K×K identity matrix, $\sigma^2 = E(G(Y_{ij}))^2 = 1/3$ and $\tau = E(G(Y_{ij})G(Y_{i'j}))$.

(Proof). See Appendix A.

Lemma 3. Under H_0, $1/6 \le \tau \le 1/3$.

(Proof). Using the fact that $|\mathrm{cov}(G(Y_{ij}), G(Y_{i'j}))| \le (\mathrm{var}\, G(Y_{ij})\ \mathrm{var}\, G(Y_{i'j}))^{1/2}$ $= 1/12$, the proof is straightforward.

We remark that the bounds of τ in the above Lemma are attainable (c.f. Hollander, Pledger and Lin (1974), p. 180).

Lemma 4. Under H_0, $\lim_{n\to\infty} P(|U_{ii',n}| \le 2q_K^{\gamma}\, n^{2/3}/\sqrt{6},\ 1 \le i < i' \le K) \ge 1-\gamma$, where $0 < \gamma < 1$, q_K^{γ} is the $100(1-\gamma)$ percentage point of the distribution of the range of K independent unit normal random variables.

(Proof). It follows directly from the above lemmas.

Lemma 5. For $i < i'$, suppose that the differences $Y_{ij} - Y_{i'j'}$ are distinct, $j, j' = 1, \ldots, n$. If $D^{ii'}_{(1)} < \ldots < D^{ii'}_{(n^2)}$ denote the ordered differences $Y_{ij} - Y_{i'j'}$, then

$$D^{ii'}_{(\ell)} \le \alpha_i - \alpha_{i'} \quad \text{if and only if} \quad \sum_{j=1}^{n} \sum_{j'=1}^{n} \phi(Y_{ij} - \alpha_i,\ Y_{i'j'} - \alpha_{i'}) < n^2 - 2\ell$$

and

$$D_{(m)}^{ii'} > \alpha_i - \alpha_{i'} \quad \text{if and only if} \quad \sum_{j=1}^{n} \sum_{j'=1}^{n} \phi(Y_{ij}-\alpha_i,\ Y_{i'j'}-\alpha_{i'}) \geq n^2-2m+2.$$

(Proof). See Appendix A.

Now, we are ready to present a simultaneous confidence bound to $\alpha_i - \alpha_{i'}$, $1 \leq i < i' \leq K$.

<u>Theorem 1</u>. $\lim_{n\to\infty} P(D_{(\ell)}^{ii'} \leq \alpha_i - \alpha_{i'} < D_{(m)}^{ii'},\ 1 \leq i < i' \leq K) \geq 1-\gamma$,

where $\ell = [n^2/2 - q_K^\gamma n^{3/2}/\sqrt{6}]$, $m = [n^2/2 + q_K^\gamma n^{3/2}/\sqrt{6}] + 2$, and $[\cdot]$ is the greatest integer function.

(Proof). It follows from the above lemmas.

So far we have assumed that there are no tied observations. When there are ties, Lemma 5 will no longer be valid. If in practice the ties are the result of rounding to the nearest multiple of ε, some modifications can be made to guarantee the validity of Theorem 1. Let the original responses giving rise to the (rounded) observations X_{ij} be X'_{ij} for which Model (1.1) is appropriate, then $|X'_{ij} - X_{ij}| \leq \varepsilon/2$ and hence $|\bar{X}'_{.j} - \bar{X}_{.j}| \leq \varepsilon/2$, where $\bar{X}'_{.j}$ is the jth block average of $X'_{1j},\ldots, X'_{Kj}$. It follows that

$$|Y'_{ij} - Y_{ij}| \leq \varepsilon, \quad \text{where } Y'_{ij} = X'_{ij} - \bar{X}'_{.j}. \tag{2.1}$$

If the ordered differences $Y'_{ij} - Y'_{i'j'}$ are denoted by $E_{(1)}^{ii'} \ldots E_{(n^2)}^{ii'}$, Lemma 5 holds when D replaced by E. However, from (2.1), $|D_{(\ell)}^{ii'} - E_{(\ell)}^{ii'}| \leq 2\varepsilon$. Therefore, if $\prod_{1\leq i<i'\leq K} [E_{(\ell)}^{ii'}, E_{(m)}^{ii'})$ is the $1-\gamma$ simultaneous confidence region

of $\alpha_i-\alpha_{i'}$, $1\le i<i'\le K$ in Theorem 1, then $\prod_{1\le i<i'\le K} [D_{(\ell)}^{ii'} - 2\varepsilon, D_{(m)}^{ii'} + 2\varepsilon)$ is also a $1-\gamma$ simultaneous confidence region of $\alpha_i-\alpha_{i'}$.

3. AN EXAMPLE

We present a numerical example in this section for illustration purpose. The data in Appendix B were obtained by Woodward (1970) to compare three methods of rounding first base to reach the second base. The three methods, "round out", "narrow angle", and "wide angle" are illustrated in Hollander and Wolfe (1973, p. 142).

Each entry in Appendix B is an average time of two runs from a point on the first base line 35 ft. from home plate to a point 15 ft. short of second base. Here, players are blocks and methods of rounding first base are treatments. The observations were rounded to the nearest multiple of $\varepsilon = .01$. For error probability $\gamma = .1$, we obtain $\ell = 119$ and $m = 365$ from Theorem 1. It follows that a 90% simultaneous confidence region of $\alpha_i-\alpha_{i'}$, $1\le i<i'\le 3$ is

$$[D_{119}^{12} - 2\varepsilon, D_{365}^{12} + 2\varepsilon) \times [D_{119}^{13} - 2\varepsilon, D_{365}^{13} + 2\varepsilon) \times [D_{119}^{23} - 2\varepsilon, D_{365}^{23} + 2\varepsilon)$$

which is $[-.07, .09) \times [0, .19) \times [0, .15)$.

4. REMARKS

Suppose that Treatment 1 is a control and the rest (K-1) treatments are under investigation as possible improvements. Then by the same argument as we gave before for all treatment comparisons, a $1-\gamma$ simultaneous confidence region of $\alpha_i-\alpha_1$, $2\le i\le K$ can be obtained as follows.

<u>Theorem 2</u>. $\lim_{n\to\infty} P(D_{(\ell)}^{i1} \le \alpha_i-\alpha_1 < D_{(m)}^{i1},\ 2\le i\le K) \ge 1-\gamma$, where

$$\ell = [n^2/2 - \xi_{K-1}^{\gamma}\, n^{3/2}/\sqrt{3}], \quad m = [n^2/2 + \xi_{K-1}^{\gamma}\, n^{3/2}/\sqrt{3}] + 2$$

and ξ^{γ}_{K-1} is the upper γ percentage point of the maximum absolute value of $(K-1)$ $N(0,1)$ random variables with common correlation $\frac{1}{2}$ (c.f. Hollander and Wolfe (1973), Table A. 14).

APPENDIX A

Proof of Lemma 1. Let $\phi^{0}_{ii'}(y) = E\phi(y, Y_{i'j'}) - p_{ii'}$ and

$\phi^{1}_{ii'}(y) = E\phi(Y_{ij},y) - p_{ii'}$. Also, let

$$g(Y_{ij},Y_{i'j'}) = \phi(Y_{ij},Y_{i'j'}) - p_{ii'} - \phi^{0}_{ii'}(Y_{ij}) - \phi^{1}_{ii'}(Y_{i'j'}) \text{ and}$$

$$U^{*}_{ii',n} = \sum_{j=1}^{n} (\phi^{0}_{ii'}(Y_{ij}) + \phi^{1}_{ii'}(Y_{i'j})) = 2\sum_{j=1}^{n} (G(Y_{ij}-\alpha_{i'})-G(Y_{i'j}-\alpha_{i})-p_{ii'}).$$

Then

$$\text{(A-1)} \qquad E(n^{-3/2} U_{ii',n} - n^{-\frac{1}{2}} U^{*}_{ii',n})^{2} = n^{-3} \sum_{j=1}^{n} \sum_{j'=1}^{n} \sum_{k=1}^{n} \sum_{k'=1}^{n} h(j,j',k,k'),$$

where $h(j,j',k,k') = E[g(Y_{ij},Y_{i'j'})g(Y_{ik},Y_{i'k'})]$. Because g is bounded, we can ignore any u terms of the sum in (A-1) if u is of order $o(n^{3})$. Consider the following cases for which the number of terms is with order larger than or equal to $O(n^{3})$ (where j,j',k and k' represent four distinct indices):

(1) $h(j,j',k,k') = Eg(Y_{ij},Y_{i'j'})\, Eg(Y_{ik},Y_{i'k'}) = 0$;

(2) $h(j,j,k,k') = h(j,j',k,k) = 0$;

(3) $h(j,j',j,k') = h(j,j',k,j') = E\{E[g(Y_{ij},Y_{i'j'})g(Y_{ij},Y_{i'k'})|Y_{ij}]\}$

$$= E\{E[g(Y_{ij},Y_{i'j'})|Y_{ij}]\; E[g(Y_{ij},Y_{i'k'})|Y_{ij}]\} = 0;$$

(4) $h(j,j',k,j) = h(j,j',j',k') = EE[g(Y_{ij},Y_{i'j'})g(Y_{ik},Y_{i'j})|Y_{ij},Y_{i'j}]$

$$= E\{E[g(Y_{ij},Y_{i'j'})|Y_{ij},Y_{i'j}]\; E[g(Y_{ik},Y_{i'j})|Y_{ij},Y_{i'j}]\} = 0.$$

It follows that $E(n^{-3/2} U_{ii',n} - n^{-\frac{1}{2}} U^*_{ii',n})^2 \to 0$, as $n \to \infty$. By Corollary 6 of Lehmann (1975, p. 289) and the fact that the vector $< n^{-\frac{1}{2}} U^*_{ii',n}/2 >$ has a $K(K-1)/2$ dimensional normal limiting distribution $N(\underset{\sim}{0},\ddagger)$, the vector $< n^{-3/2} U_{ii',n}/2 >$ has the same normal limiting distribution $N(0,\ddagger)$. Q.E.D.

<u>Proof of Lemma 2</u>. Under the H_0: $\alpha_1=\alpha_2= \ldots = \alpha_K$, $Y_{1j}, \ldots, Y_{Kj}$ are interchangeable and have a common continuous marginal distribution function G. A typical element $\sigma_{ii',kk'}$ of $\ddagger$ is $E(G(Y_{ij})-G(Y_{i'j}))(G(Y_{kj})-G(Y_{k'j}))$, where $i < i'$ and $k < k'$. It is easy to show that the covariance matrix $\ddagger$ is identical to the vector having $K(K-1)/2$ components $(\sigma^2-\tau)^{\frac{1}{2}}(Z_i-Z_{i'})$, $1 \le i < i' \le K$, where $(Z_1, \ldots, Z_K)' \sim N_K(\underset{\sim}{0},I_K)$, $\sigma^2 = E(G(Y_{ij}))^2 = 1/3$ and $\tau = E(G(Y_{ij})G(Y_{i'j}))$. Q.E.D.

<u>Proof of Lemma 5</u>. The inequality $D^{ii'}_{(\ell)} \le \alpha_i-\alpha_{i'}$ holds if and only if at least ℓ of the differences $(Y_{ij}-\alpha_i) - (Y_{i'j'}-\alpha_{i'})$ are less than or equal to zero and hence $\sum_{j=1}^{n} \sum_{j'=1}^{n} \phi(Y_{ij}-\alpha_i, Y_{i'j'}-\alpha_{i'}) < n^2-2\ell$. The second statement of this lemma can be obtained in a similar manner. Q.E.D.

APPENDIX B

Rounding first base times

Players	Methods: Round out	Narrow Angle	Wide Angle
1	5.40	5.50	5.55
2	5.85	5.70	5.75
3	5.20	5.60	5.50
4	5.55	5.50	5.40
5	5.90	5.85	5.70
6	5.45	5.55	5.60
7	5.40	5.40	5.35
8	5.45	5.50	5.35
9	5.25	5.15	5.00
10	5.85	5.80	5.70
11	5.25	5.20	5.10
12	5.65	5.55	5.45
13	5.60	5.35	5.45
14	5.05	5.00	4.95
15	5.50	5.50	5.40
16	5.45	5.55	5.50
17	5.55	5.55	5.35
18	5.45	5.50	5.55
19	5.50	5.45	5.25
20	5.65	5.60	5.40
21	5.70	5.65	5.55
22	6.30	6.30	6.25

REFERENCES

Hettmansperger, T. P. (1975). Non-parametric inference for ordered alternatives in a randomized block design. Psychometrika 40, 53-62.

Hollander, M., Pledger, G. and Lin, P. E. (1974). Robustness of the Wilcoxon test to a certain dependency between samples. The Annals of Statistics 2, 177-181.

Hollander, M. and Wolfe, D. A. (1973). Nonparametric Statistical Methods, New York: John Wiley & Sons, Inc.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks, San Francisco, Cal.: Holden-Day, Inc.

Miller, R. G., Jr. (1966). Simultaneous Statistical Inference, New York: McGraw-Hill Book Co.

——————— (1977). Developments in multiple comparisons 1966-1976, J. of Amer. Statist. Assoc. 72, 779-788.

Puri, M. L. and Sen, P. K. (1971). Nonparametric Methods in Multivariate Analysis, New York: John Wiley & Sons, Inc.

Sen, P. K. (1969). On nonparametric T-method of multiple comparisons in randomized blocks. Ann. of the Institute of Statist. Math. 21, 329-333.

Woodward, W. F. (1970). A comparison of base running methods in baseball. M. Sc. thesis, Florida State Univ.