中國統計學報
第 34 卷第 1 期
八十五年三月
66-78 頁

當某些對照處理之觀測值漏失時集區設計之穩健性之探討

丁兆平　　郎冰瑩　　蔡風順
國立政治大學　國立政治大學　中央研究院
統計系　　應用數學系　　統計研究所

摘　要

本文主要是在探討集區設計對觀測值漏失時之穩健性。文中求出任一集區設計，當某一集區內漏失 t 個對照處理之觀測值時，該設計仍維持穩健之充要條件；一個計算上較簡單之充份條件亦在文中給出。殘留設計之 A 效率值之上下限也在文中求出。一些特殊形式之 $BTBD$ 之 A 效率值可準確算出，實例亦在最後一節中列出。

關鍵詞： 集區設計，殘留設計，$BTBD$，穩健設計，A 效率值。

美國數學會分類索引： 主要62K10; 次要62K05, 05B05。

1. 前言

本文旨在研究，在比較試驗處理(test treatment)與對照處理(control treatment)之問題下，當某些觀測值漏失時，集區設計(block designs)之穩健性(robustness)。探討集區設計穩健性之文章相當多，如Hedayat and John (1974), John (1976), Ghosh (1982), Ghosh, Rao, and Singhi (1983), Baksalary and Tabis (1987), Dey and Dhall (1988), Srivastava, Gupta, and Dey (1990), Mukerjee and Kageyama (1990), and Dey (1993)。有興趣對此問題做廣泛了解的讀者，可參閱Kageyama (1990)。上述之文章中，有些是在不同的架構之下，從不同的觀點來探討當試驗處理之觀測值漏失時，集區設計之穩健性。本文則主要是在探討，當某些對照處理之觀測值漏失時，集區設計之穩健性。

文中將針對二種穩健性來探討，第一種是連接(connectedness)穩健性。此穩健性最初是由Ghosh (1982)所定義。由於本文是在比較試驗處理與對照處理之架構下，與Ghosh之試驗處理間相互比較略有不同，故將Ghosh之穩健性之觀念衍生如下：令$D(v+1,b,k)$表具有b個大小爲k之集區，且有v個試驗處理與一個對照處理配置其中之所有集區設計之集合。對任一設計$d \in D(v+1,b,k)$，令d_R表在任一集區中漏失t個對照處理之觀測值後之殘留設計(residual design)。若當所有對照處理與試驗處理之對比(contrast)在d_R中均是可估計的，則設計d稱之爲具有連接穩健性。

第二個在文中將探討之設計穩健性，是和A最適性(A-optimality)相當有關之A效率(A-efficiency)穩健性。設計d稱之爲具有A效率穩健性，係指當d漏失對照處理之觀測值到d_R時所損失之A效率很小而言。

當漏失t個對照處理之觀測值時，我們對殘留設計最基本之要求爲：所有對照處理與試驗處理之對比仍然是可以估計的。故尋找連接穩健性之條件，爲本文之首要興趣。第三節中將給出，任一集區設計滿足連接穩健性之充要條件，以及一較簡單之充份條件。當設計具連接穩健性之後，接下來我們便想探究，相較於原來之設計，殘留設計之良窳。故在第四節中我們將求出，殘留設計之A效率值之上下限。由於平衡處理集區設計(balanced treatment block designs, BTBD)不但具有A最適性，且具有“對稱”之性質，故其A效率值可準確求出，並將在文中之第五節中給出，範例亦將列在第五節。

2. 背景

本節將列出文中所需之定義、定理、以及引理。對任一矩陣M,令M'表M之轉置(transpose)矩陣。

定義:設計$d \in D(v+1,b,k)$符合下列條件,則稱之爲平衡處理集區設計:

(i) $\sum_{j=1}^{b} n_{d0j} n_{dij} = \gamma_0, 1 \le i \le v$,

(ii) $\sum_{j=1}^{b} n_{dij} n_{di'j} = \gamma_1, 1 \le i \neq i' \le v$,

其中n_{dij}表處理i, $0 \le i \le v$,在第j個, $1 \le j \le b$,集區中出現之次數。

本文將僅針對A最適性已被證實之平衡處理集區設計做探討,亦即符合下列條件之平衡處理集區設計:$|n_{d0j} - n_{d0j'}| \le 1, 1 \le j, j' \le b$,和$|n_{dij} - n_{di'j'}| \le 1$, $1 \le i, i' \le v, 1 \le j, j' \le b$,標示爲BTBD$(v,b,k;r_0)$,其中$r_0 = \sum_{j=1}^{b} n_{d0j}$表對照處理在實驗中出現之總次數。

定理2.1: 令M_1爲$p \times p$之正定(positsive definite)對稱矩陣,M_2爲$p \times p$之非負定(nonegative definite)對稱矩陣,且$M_1 = M_2 + AA'$,其中A爲$p \times q$之矩陣,且A之行空間爲M_1行空間之子空間。則M_2爲正定之充要條件爲: $I_q - A'M_1^{-1}A$爲正定矩陣。

證明:請參閱Dey (1993)。

定理2.2: 令M_1、M_2爲二個$p \times p$之正定對稱矩陣,A如定理2.1。則

$$M_2^{-1} = M_1^{-1}(I_p + A(I_q - A'M_1^{-1}A)^{-1}A'M_1^{-1})$$

證明:請參閱Rao (1973)。

引理2.1: 令M_1爲$p \times p$之正定對稱矩陣,令M_2爲$p \times p$之對稱矩陣。令$\pi_1 \le \cdots \le \pi_p$表$M_1^{-1}M_2$之特徵值(eigenvalues)。則對任一不爲$\vec{0}$之向量$\vec{x}$,

$$\pi_1 \le \vec{x}'M_2\vec{x}/\vec{x}'M_1\vec{x} \le \pi_p.$$

證明:請參閱Magnus and Neudecker (1988)。

3. 連接穩健之條件

令C表設計d之估計對照處理－試驗處理差比之係數矩陣(coefficient matrix), 根據Bechhofer and Tamhane (1981),

$$C=\begin{pmatrix} r_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_v \end{pmatrix}-\frac{1}{k}NN',$$

其中$r_i=\sum_{j=1}^{b}n_{dij}$表試驗處理$i$在實驗中出現之總次數, $N=((n_{dij})), 1\leq i\leq v, 1\leq j\leq b$, 為$v\times b$之關聯矩陣(incidence matrix)。令C_R表所對應之d_R之係數矩陣。不失一般性的, 我們可假設漏失t個對照處理之觀測值發生在第一個集區, 則C和C_R之關係如下：

$$\begin{aligned} C_R &= \begin{pmatrix} r_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_v \end{pmatrix}-N\begin{pmatrix} \frac{1}{k-t} & 0_{1,b-1} \\ 0_{b-1,1} & \frac{1}{k}I_{b-1} \end{pmatrix}N' \\ &= \begin{pmatrix} r_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_v \end{pmatrix}-N\left(\frac{1}{k}I_b+\begin{pmatrix} \frac{t}{k(k-t)} & 0_{1,b-1} \\ 0_{b-1,1} & 0_{b-1,b-1} \end{pmatrix}\right)N' \\ &= C-\frac{t}{k(k-t)}(n_{d11},\cdots,n_{dv1})'(n_{d11},\cdots,n_{dv1}), \end{aligned}$$

則,

$$C=C_R+QQ',$$

其中$Q=(t/(k(k-t)))^{1/2}(n_{d11},\cdots,n_{dv1})', 1\leq t\leq n_{d01}$, $0_{a,b}$表$a\times b$之零矩陣。

根據定理2.1, 我們可得下列之定理。

<u>**定理 3.1:**</u> 在同一集區中漏失t個對照處理之觀測值時, 設計$d\in D(v+1,b,k)$, 具有連接穩健性之充要條件為$Q'C^{-1}Q<1$。

上述定理3.1之充要條件, 在實際運用上不甚方便, 因為我們必須確實知道v個試驗處理在第一個集區中之配置情形(由Q中可看出)。下面之系理3.1應可解決這個問題。

系理3.1: 在同一集區中漏失t個對照處理之觀測值時，若C之最小特徵值大於$(t(k-t))/k$，則d具有連接穩健性。

證明: 令$(\lambda_i, \vec{e_i}), i = 1, \cdots, v$，表$C$之特徵值與特徵向量組，且$0 < \lambda_1 \le \cdots \le \lambda_v$。則$C = \sum_{i=1}^{v} \lambda_i \vec{e_i}\vec{e_i}'$，且$Q$可寫成$\vec{e_i}$之線性組合，$Q = \sum_{i=1}^{v} a_i \vec{e_i}$，其中$a_i, i = 1, \cdots, v$，爲實數。則

$$
\begin{aligned}
Q'C^{-1}Q &= (\sum_{i=1}^{v} a_i \vec{e}_i)'(\sum_{i=1}^{v} \lambda_i^{-1} \vec{e}_i \vec{e_i}')(\sum_{i=1}^{v} a_i \vec{e}_i) \\
&= \sum_{i=1}^{v} (a_i^2/\lambda_i) \text{。}
\end{aligned}
$$

因爲

$$
\begin{aligned}
Q'Q &= \sum_{i=1}^{v} a_i^2 = \frac{t}{k(k-t)} \sum_{i=1}^{v} n_{i1}^2 \le \frac{t}{k(k-t)} (k - n_{01})^2 \\
&\le \frac{t}{k(k-t)} (k-t)^2 = \frac{t(k-t)}{k},
\end{aligned}
$$

系理得證。

所以對任一設計d，我們不需要知道$v+1$個處理在d中之實際配置情形，僅需要檢視C之最小特徵值是否大於$(t(k-t))/k$，即可知該設計是否具有連接穩健性。

4. 殘留設計效率值之上下限

相較於原設計，當殘留設計之效率值相當低時，就算該設計已具連接穩健性，但實驗者依然不會使用它。所以在本節中，我們將定義殘留設計之效率值EFF如下：

$$EFF = tr(C^{-1})/tr(C_R^{-1}) \text{。}$$

若EFF接近1時，我們稱該設計具A效率穩健性。

由於設計必須先具有連接穩健性之後，才有繼續探討其A效率穩健性之必要，故在本節中，我們僅將探討符合系理2.1之設計之EFF值，亦即，係數矩陣C之最小特徵值大於$(t(k-t))/k$之設計。

定理4.1: 在同一集區中漏失t個對照處理之觀測值時,殘留設計d_R效率值之上下限爲

$$(1+\frac{\varphi_v}{tr(C^{-1})})^{-1} \le EFF \le (1+\frac{\varphi_1}{tr(C^{-1})})^{-1},$$

其中$\varphi_1 \le \cdots \le \varphi_v$爲$((Q'Q)^{-1}I-C^{-1})^{-1}C^{-2}$之特徵值。

證明: 根據定理2.2可得

$$C_R^{-1}=C^{-1}+(1-Q'C^{-1}Q)^{-1}C^{-1}QQ'C^{-1} \text{。}$$

則

$$tr(C_R^{-1})=tr(C^{-1})+(1-Q'C^{-1}Q)^{-1}Q'C^{-2}Q \text{。}$$

於是

$$\begin{aligned} EFF &= tr(C^{-1})/tr(C_R^{-1}) \\ &= \frac{tr(C^{-1})}{tr(C^{-1})+(Q'C^{-2}Q)(1-Q'C^{-1}Q)^{-1}} \text{。} \end{aligned}$$

現因

$$1-Q'C^{-1}Q=Q'((Q'Q)^{-1}I-C^{-1})Q.$$

且C^{-2}爲正定,並因$\lambda_1 \ge (t(k-t))/k$,故$((Q'Q)^{-1}I-C^{-1})^{-1}C^{-2}$亦爲正定,則根據引理2.1,

$$\varphi_1 \le \frac{Q'C^{-2}Q}{Q'((Q'Q)^{-1}I-C^{-1})Q} \le \varphi_v,$$

其中$\varphi_1 \le \cdots \le \varphi_v$爲$((Q'Q)^{-1}I-C^{-1})^{-1}C^{-2}$之特徵值,且

$$\varphi_i=\lambda_{v+1-i}^{-1}((Q'Q)^{-1}\lambda_{v+1-i}-1)^{-1}, i=1,\cdots,v \text{。}$$

於是

$$\frac{tr(C^{-1})}{tr(C^{-1})+\varphi_v} \le EFF \le \frac{tr(C^{-1})}{tr(C^{-1})+\varphi_1} \text{。}$$

故定理得證。

因上述之上下限均含Q,故計算此上下限時亦必須確實知道v個試驗處理在第一個集區之配置。但因

$$Q'Q=(t/(k(k-t))\sum_{i=1}^{v} n_{i1}^2,$$

且

$$(k-t)^2/v \le \sum_{i=1}^{v} n_{i1}^2 \le (k-t)^2,$$

所以我們可得

$$\varphi_1 \ge \lambda_v^{-1}(\frac{vk}{t(k-t)}\lambda_v - 1)^{-1} = \varphi_1^*,$$

$$\varphi_v \le \lambda_1^{-1}(\frac{k}{t(k-t)}\lambda_1 - 1)^{-1} = \varphi_v^* \text{ 。}$$

根據上面二不等式，我們便可求得一個計算較簡單，且只需知道t值和C之特徵值之上下限，如下式：

$$(1+\frac{\varphi_v^*}{tr(C^{-1})})^{-1} \le EFF \le (1+\frac{\varphi_1^*}{tr(C^{-1})})^{-1} \text{ 。}$$

5. BTBD之A效率值和範例

$BTBD$在最適實驗設計之領域中扮演非常重要的角色。當最適之r_0值求得時，且所對應之$BTBD$存在時，該設計之A最適性已被證實。故在本節中，我們將針對$BTBD$來探討其穩健性。

令$d^* \in D(v+1,b,k)$爲一BTBD$(v,b,k;r_0)$，且令C^*表d^*之係數矩陣，則

$$C^* = (r+\frac{\omega_1-\omega_2}{k})I_v - \frac{\omega_1}{k}J_{v,v}, \tag{5.1}$$

其中$r=(bk-r_0)/v, \omega_1=\sum_{j=1}^b n_{ij}n_{i'j}, \omega_2=\sum_{j=1}^b n_{ij}^2, \forall i \ne i', i=1,\cdots,v$，$J_{a,b}$表所有元素均爲1之$a\times b$矩陣，且$n_{ij}=[(bk-r_0)/vb]$或$[(bk-r_0)/vb]+1, i=1,\cdots,v, j=1,\cdots,b$，此處$[\cdot]$表最大整數函數。

經過簡單的計算，我們可求得殘留設計之係數矩陣如下：

$$C_R^* = \begin{pmatrix} xI_m - yJ_{m,m} & -uJ_{m,v-m} \\ -uJ_{v-m,m} & xI_{v-m} - zJ_{v-m,v-m} \end{pmatrix},$$

其中$x=r+\frac{1}{k}(\omega_1-\omega_2), y=\frac{1}{k}(\omega_1+\frac{t\alpha^2}{k-t}), z=\frac{1}{k}(\omega_1+\frac{t(\alpha+1)^2}{k-t}), u=\frac{1}{k}(\omega_1+\frac{t\alpha(\alpha+1)}{k-t}), \alpha=[(bk-r_0)/vb]$，且$m=n_{di1}$爲$\alpha$之個數。因爲$C_R^*$爲具特殊形式之矩陣，故其特徵值可求得如後：$C_R^*$有$v-2$個特徵值爲$x$，其餘二個特徵值$\rho_1$和$\rho_2$則滿足下列二條件，$\rho_1+\rho_2=2x-my-(v-m)z, \rho_1\rho_2=(x-my)(x-(v-m)\delta)$,

其中$\delta = z + (mu^2/(x - my))$。

於是

$$tr(C_R^*) = (v-2)/x + 1/\rho_1 + 1/\rho_2$$
$$= \frac{v-2}{x} + \frac{2x - my - (v-m)z}{(x-my)(x-(v-m)\delta)} \text{。}$$

由(5.1)可得

$$tr(C^*) = \frac{v(x - \frac{1}{k}(v-1)\omega_1)}{x(x - \frac{1}{k}v\omega_1)} \text{。}$$

所以

$$EFF = \frac{\frac{v(x-\frac{1}{k}(v-1)\omega_1)}{x(x-\frac{1}{k}v\omega_1)}}{\frac{v-2}{x} + \frac{2x-my-(v-m)z}{(x-my)(x-(v-m)\delta)}} \text{。}$$

且當$m = v$時,此值可簡化爲:

$$EFF = \frac{x - \frac{1}{k}(v-1)\omega_1}{x - \frac{1}{k}v\omega_1} \frac{x - vy}{x-(v-1)y} \text{。}$$

下面之範例中,0代表對照處理,$1,\cdots,v$分別代表v個試驗處理。

範例1: 當$v = 4, b = 4, k = 4$,下列之BTBD(4,4,4;4)爲$r_0 = 4, r = 3, \omega_1 = 2, \omega_2 = 3, \alpha = 0, m = 1, t = 1$。則$x = 11/4, y = u = 0.5, z = 7/12, \delta = 25/36$,且EFF=0.8376963。

0	0	0	0
1	1	1	2
2	2	3	3
3	4	4	4

範例2: 當$v = 5, b = 5, k = 5$,下列之BTBD(5,5,5;5)爲$r_0 = 5, r = 4, \omega_1 = 3, \omega_2 = 4, \alpha = 0, m = 1, t = 1$。則$x = 3.8, y = u = 0.6, z = 0.65, \delta = 0.7625$,且EFF=0.8779264。

0	0	0	0	0
1	1	1	1	2
2	2	2	3	3
3	3	4	4	4
4	5	5	5	5

範例3: 當 $v = 4, b = 6, k = 6$, 下列之 BTBD(4,6,6;12) 爲 $r_0 = 12, r = 6, \omega_1 = \omega_2 = 6, \alpha = 1, m = 4$ 。

0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4

情況1: 當 $t = 1$, 則 $x = 6, y = 31/30$, 且 EFF=0.9655172 。

情況2: 當 $t = 2$, 則 $x = 6, y = 13/12$, 且 EFF=0.9090909 。

範例4: 當 $v = 4, b = 10, k = 6$, 下列之 BTBD(4,10,6;16) 爲 $r_0 = 16, r = 11, \omega_1 = 12, \omega_2 = 13, \alpha = 1$ 。

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	1	1
1	1	1	1	1	1	2	2	2	2
2	2	2	2	2	2	3	3	3	3
3	3	3	3	3	3	4	4	4	4
4	4	4	4	4	4	1	2	3	4

情況1: 當 $m = 4, t = 1$ (亦即僅漏失一個對照處理, 且漏失之情形發生在前六個集區中之任一集區), 則 $x = 65/6, y = 61/30$, 且 EFF=0.9730735 。

情況2: 當 $m = 4, t = 2$ (亦即漏失兩個對照處理,且漏失之情形發生在前六個集區中之任一集區),則 $x = 65/6, y = 25/12$, 且 EFF=0.9303813 。

情況3: 當 $m = 3, t = 1$ (亦即僅漏失一個對照處理,且漏失之情形發生在後四個集區中之任一集區),則 $x = 65/6, y = 61/30, u = 62/30, z = 64/30, \delta = 4.8403756$, 且 EFF=0.9569847 。

由上面之四個範例中我們可看出,當 v, b, k 和 r_0 值小時, BTBD$(v, b, k; r_0)$ 之 EFF 值並不是非常接近1。(到目前爲止我們所試過之範例中,大部份之EFF值均大於0.8)。這是可以理解的,因爲當 v, b, k 和 r_0 值小時,任一觀測值相對的變得非常具有代表性,所以當某些觀測值漏失時,其所損失之效率值也就相對的提高。但當 v, b, k 和 r_0 值稍微增加時, BTBD$(v, b, k; r_0)$ 之EFF值則迅速接近1。所以我們可以總括來說, BTBD$(v, b, k; r_0)$ 是相當具有A效率穩健性之設計。

至於其他形式之漏失觀測值之情形,以及組可分處理設計(group divisible treatment design, GDTD)(其最適性亦已被證實)之A效率穩健性之探討,將爲我們後續研究之重點之一。

參考文獻

Bechhofer, R. E. and Tamhane, A. C. (1981). Incomplete block designs for comparing treatments with a control: general theory. *Technometrics* **23**, 45-47.

Baksalary, J. K. and Tabis, Z. (1987). Conditions for the robustness of block designs against the unavailability of data. *J. Statist. Plann. Infer.* **16**, 49-54.

Dey, A. (1986). *Theory of Block Designs.* A Halsted Press Book. John Wiley, New York.

Dey, A. (1993). Robustness of block designs against missing data. *Statistica Sinica* **3**, 219-231.

Dey, A. and Dhall, S. P. (1988). Robustness of augmented BIB designs. *Sankhyā Ser.* **B** 50, 376-381.

Ghosh, S. (1982). Robustness of BIBD against the unavailability of data. *J. Statis. Plann. Infer.* **6**, 29-32.

Ghosh, S., Kageyama, S. and Mukerjee, R. (1992). Efficiency of connected binary block designs with a single observation is unavailable. *Ann. Inst. Statist. Math.* **44**, 593-603.

Ghosh, A., Rao, S. B., and Singhi, N. M. (1983). On a robustness property of PBIBD. *J. Statist. Plann. Infer.* **8**, 355-363.

Hedayat, A. and John, P. W. M. (1974). Resistant and susceptible BIB designs. *Ann. Statist.* **2**, 148-158.

John, P. W. M. (1976). Robustness of balanced incomplete block designs. *Ann. Statist.* **4**, 960-962.

Kageyama, S. (1990). Robustness of block designs. *Probability, Statistics and Design of Experiments* (R. R. Bahadur, ed.) 425-438. Wiley Eastern, New Delhi.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley, New York.

Mukerjee, R. and Kageyama, S. (1990). Robustness of group divisible designs. *Comm. Statist. Theory Methods* **19**, 3189-3203.

Pringle, R. M. and Rayner, A. A. (1971). *Generalized Inverse Matrices: with Applications to Statistics.* Griffin, London.

Rao, C. R. (1973). *Linear Statistical Inference and It's Applications.* Wiley, New York.

Srivastava, R., Gupta, V. K, and Dey, A. (1990). Robustness of some designs against missing observations. *Comm. Statist. Theory Methods* **19**, 121-126.

[民國84年8月1日收稿,民國85年1月16日修正。]

JCSA, March 1996, Vol. 34, No. 1, page 66-78

Robustness of Block Designs When Some Observations of the Control Treatment Are Missing

Chao-Ping Ting
Department of Statistics
National Chengchi University
Taipei, Taiwan 11623, R.O.C.

Bing-Ying L. Lin
Department of Mathematical Sciences
National Chengchi University
Taipei, Taiwan 11623, R.O.C.

Feng-Shun Chai
Institute of Statistical Science
Academia Sinica
Taipei, Taiwan 11529, R.O.C.

ABSTRACT

Robustness of block designs with missing data is investigated. A necessary and sufficient condition for the robustness of an arbitrary block design is derived when t observations of the control treatment in a block are missing. A simpler sufficient condition is also provided. A lower bound and an upper bound to the efficiency of the residual designs are obtained as well. The robustness of some special BTBD's is also discussed.

Key words and phrases: Block designs, residual designs, BTBD, criteria of robustness, A-efficiency.

AMS 1991 subject classifications: Primary 62K10; secondary 62K05, 05B05.