

## T 152431

ประพจน์ คำรงค์สุทธิพงษ์ :การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์(A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR RANDOMIZED COMPLETE BLOCK DESIGN)อ.ที่ปรึกษา:

รศ. ดร. สุพล คุรงค์วัฒนา, 163หน้า. ISBN 974-17-4625-3

วัตถุประสงค์ของการวิจัยครั้งนี้ เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ 3 วิธี ได้แก่ การประมาณค่าวิธีกำลังสองน้อยสุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และการประมาณค่าวิธีอิมพิวเทชัน ซึ่งตัวแบบสำหรับแผนแบบการทดลองแบบสุ่มในบล็อกสมบูรณ์ที่ไม่มีการทำซ้ำเป็นดังนี้

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad , \quad i = 1, 2, \dots, t ; j = 1, 2, \dots, b$$

เมื่อ  $Y_{ij}$  แทนค่าสังเกตจากหน่วยทดลองที่ได้รับวิธีการทดลองที่  $i$  และอยู่ใน บล็อกที่  $j$   $\mu$  แทนค่าเฉลี่ยรวม  $\tau_i$  แทนอิทธิพลของวิธีการทดลองที่  $i$  โดยที่  $\tau_i$  มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน  $\sigma_\tau^2$   $\beta_j$  แทนอิทธิพลของบล็อกที่  $j$  โดยที่  $\beta_j$  มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน  $\sigma_\beta^2$   $\varepsilon_{ij}$  แทนความคลาดเคลื่อนสุ่มของค่าสังเกตจากหน่วยทดลองที่ได้รับวิธีการทดลองที่  $i$  และอยู่ในบล็อกที่  $j$  โดยที่  $\varepsilon_{ij}$  มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน  $\sigma_\varepsilon^2$   $\tau_i$  ,  $\beta_j$  และ  $\varepsilon_{ij}$  เป็นอิสระซึ่งกันและกัน,  $t$  แทนจำนวนวิธีการทดลอง  $b$  แทนจำนวนบล็อกในการทดลอง

ในการวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม S-PLUS 2000 โดยศึกษาภายใต้ความคลาดเคลื่อนมีการแจกแจงแบบปกติ การเปรียบเทียบกระทำเมื่อจำนวนระดับปัจจัยของวิธีการทดลอง เท่ากับ 3, 4, และ 5 ระดับ จำนวนบล็อก เท่ากับ 2, 4, และ 6 บล็อก กำหนดให้จำนวนเปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20%, และ 30% สำหรับการจำลองสถานการณ์กำหนดให้สัมประสิทธิ์การแปรผัน (Coefficient of Variation : C.V.) เป็น 5%, 25%, 45% หลักเกณฑ์ที่นำมาใช้ในการเปรียบเทียบวิธีการประมาณทั้ง 3 วิธี คือค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE (Mean Square Error)

ผลการวิจัยสรุปได้ว่า สำหรับเปอร์เซ็นต์ข้อมูลสูญหายเมื่อมีค่ามากขึ้นค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนจะมีค่าสูงขึ้น โดยที่วิธีการประมาณค่าสูญหายวิธีอิมพิวเทชัน ให้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE (Mean Square Error) ต่ำกว่าวิธีค่าคาดหวังสูงสุด และวิธีกำลังสองน้อยสุดในทุกสถานการณ์ของการทดลองที่ทำการศึกษา

4482297026 : MAJOR STATISTICS

KEY WORD: MISSING / RANDOMIZED COMPLETE BLOCK / LEAST SQUARE / EM ALGORITHM / IMPUTATION

PRAPHOT DAMRONGSUTTIPONG : A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR RANDOMIZED COMPLETE BLOCK DESIGN.  
 THESIS ADVISOR : ASSOCIATE PROFESSOR . SUPOL DURONGWATANA,Ph.D.  
 163pp.ISBN 974-17-4625-3

The objective of this study is to compare three methods of missing value estimation for randomized complete block design; least square method, expectation maximization (EM) algorithm method and imputation method. The model for randomized complete block design is as follows:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad , \quad i = 1, 2, \dots, t \quad ; \quad j = 1, 2, \dots, b$$

when  $Y_{ij}$  is observation for the  $i^{\text{th}}$  level of treatment and  $j^{\text{th}}$  level of block;  $\mu$  is the grand mean;  $\tau_i$  is the  $i^{\text{th}}$  random effect of treatment factor where  $\tau_i$  is independently distributed with mean 0 and variance  $\sigma_\tau^2$ ;  $\beta_j$  is the  $j^{\text{th}}$  random effect of blocking factor where  $\beta_j$  is independently distributed with mean 0 and variance  $\sigma_\beta^2$ ;  $\varepsilon_{ij}$  is the random error for the observed data at the  $i^{\text{th}}$  level of treatment factor, the  $j^{\text{th}}$  level of blocking factor where  $\varepsilon_{ij}$  is independently distributed with mean 0 and variance  $\sigma_\varepsilon^2$ ;  $\tau_i$ ,  $\beta_j$  and  $\varepsilon_{ij}$  are independent,  $t$  is number of levels for treatment factor and  $b$  is number of levels for blocking factor.

Monte Carlo simulation is used for data generated through S-plus 2000 code, The random error term in the model is generated with normal distribution. The number of levels for treatment factor is specified at  $t = 3$ ,  $t = 4$  and  $t = 5$ ; the number of levels for blocking factor is specified at  $b = 2$ ,  $b = 4$  and  $b = 6$ , the level of missing data is 10%, 20% and 30%, and the coefficient of variation (CV) for the observed data is varied from 5%, 25% and 45%. The mean square error for comparison between three methods is used as criterion for evaluate.

The results for the study show that estimates for missing value in the randomized complete block design model using the imputation method regardless the number of levels for treatment factor, the number of levels for blocking factor, percent of missing data, and the coefficient of variation for the observed data, provide smaller mean square error than the ones from the EM algorithm and the least square method in all situations.