

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัดสุ่มละติน 3 วิธี คือ การประมาณค่าวิธีกำลังสองน้อยสุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และการประมาณค่าวิธีมัลติเพิล อิมพิวเทชัน ซึ่งตัวแบบสำหรับแผนแบบการทดลองแบบจัดสุ่มละตินที่ไม่มีการทำซ้ำเป็นดังนี้

$$Y_{ijk} = \mu + \tau_i + \beta_j + \alpha_k + \varepsilon_{ijk} \quad ; i, j, k = 1, 2, \dots, p$$

เมื่อ Y_{ijk} แทนค่าสังเกตของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k μ แทนค่าเฉลี่ยรวมของประชากร τ_i แทนอิทธิพลของวิธีทดลองที่ i β_j แทนอิทธิพลของปัจจัยแถวที่ j α_k แทนอิทธิพลของปัจจัยคอลัมน์ที่ k ε_{ijk} แทนความคลาดเคลื่อนของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k โดยที่ $\tau_i, \beta_j, \alpha_k, \varepsilon_{ijk}$ เป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 ความแปรปรวน $\sigma_\tau^2, \sigma_\beta^2, \sigma_\alpha^2$ และ σ_ε^2 ตามลำดับ และ p แทนจำนวนวิธีทดลอง, จำนวนปัจจัยแถว และจำนวนปัจจัยคอลัมน์ในการทดลอง

ในการวิจัยครั้งนี้ได้จำลองข้อมูลด้วยเทคนิคมอนติคาร์โลโดยใช้โปรแกรม S-PLUS 2000 โดยกำหนดให้จำนวนวิธีทดลองที่ใช้ทดลองเท่ากับ 3 4 5 6 และ 7 สัมประสิทธิ์ความผันแปรเท่ากับ 5% 25% และ 45% และจำนวนข้อมูลสูญหายเท่ากับ 10% 20% และ 30% ได้ทำการทดลองซ้ำๆ กัน 500 ครั้งในแต่ละสถานการณ์ที่กำหนด และทำการเปรียบเทียบวิธีการประมาณค่าสูญหาย โดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) โดยวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด แสดงว่าเหมาะสมสำหรับการประมาณค่าสูญหาย

ผลการวิจัยสรุปได้ว่า เมื่อเปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดจะมีค่าเพิ่มขึ้น สำหรับกรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่ามาก พบว่าการประมาณค่าสูญหายวิธีมัลติเพิล อิมพิวเทชัน จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำกว่าวิธีค่าคาดหวังสูงสุด และวิธีกำลังสองน้อยสุดในทุกสถานการณ์ที่ทำการศึกษาดังนั้นในกรณีนี้จึงควรเลือกใช้วิธีมัลติเพิล อิมพิวเทชัน ในการประมาณค่าสูญหาย แต่สำหรับกรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่าน้อย พบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นจึงควรเลือกใช้วิธีกำลังสองน้อยสุดในการประมาณค่าสูญหาย เนื่องจากสะดวกและรวดเร็วกว่า

The Objective of this research is to compare three methods of missing value estimation for latin square design which comprise Least square method, Expectation Maximization algorithm (EM algorithm) and Multiple Imputation Method (MI) by using the following model

$$Y_{ijk} = \mu + \tau_i + \beta_j + \alpha_k + \varepsilon_{ijk} \quad ; i, j, k = 1, 2, \dots, p$$

while Y_{ijk} is observation for i^{th} level of treatment factor with the j^{th} level of row factor and k^{th} level of column factor; μ is the population grand mean; τ_i is the i^{th} random effect of treatments factor; β_j is the j^{th} random effect of row factor; α_k is the k^{th} random effect of column factor ; ε_{ijk} is the random error for the observed data at the i^{th} level of treatment factor, the j^{th} level of row factor and the k^{th} level of column factor where $\tau_i, \beta_j, \alpha_k, \varepsilon_{ijk}$ are identically independent normal distribution with zero mean and variance equaled to $\sigma_\tau^2, \sigma_\beta^2, \sigma_\alpha^2$ and σ_ε^2 respectively. Additionally, p is the number of treatment, row factors and column factors.

The data are derived from simulation by Monte Carlo technique with S-PLUS 2000 Professional. The number of treatments are specified as 3, 4, 5, 6 and 7. The coefficients of variation are specified as 5%, 25% and 45%, whereas the numbers of missing data are specified to be 10%, 20% and 30% respectively. Each situation are repeated 500 times and three missing value estimation methods are compared by considering maximum absolute error or MAE. Then, the method resulting the minimum value of MAE will be assessed as an appropriate method for estimating missing values.

It is found, in this study, that in the case of higher missing value percentage and coefficient variation, the MAE is, then, increased. For a case of high missing value percentage and high coefficient variation, an estimation of missing values by the multiple imputation would contribute the lower MAE value as compared to EM algorithm and Least Square method, at all situations in the study. Therefore, the multiple imputation is recommended in the case of high level of missing value and high coefficient variation. Oppositely, when the missing value percentage and coefficients variation are low, the MAE found by the all three methods are nearly the same. Then, the Least Square method is accordingly suggested because it is more convenient and less time consumed.