ISSN 3027-6357 (Print)
ISSN 3027-6365 (Online)

วารสารวิทยาศาสตร์และเทคโนโลยี Homepage : *http://li01.tci-thaijo.org/index.php/tstj*
ปีที่ 32 ฉบับที่ 6 (พฤศจิกายน-ธันวาคม 2567) หน้า [1-11]

การประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์เมื่อทรีตเมนต์และบล็อกเป็นอิทธิพลแบบสุ่ม

Estimation of a Missing Value in Randomized Complete Block Design When Treatment and Block are Random Effects

ธัญวรัตน์ เมืองทรัพย์*, บุญอ้อม โฉมที, อำไพ ทองธีรภาพ
ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร 10900
Tanwarat Muangtrup*, Boonorm Chomtee, Ampai Thongteeraparp
Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900
Received 16 November 2023; Received in revised 30 June 2024; Accepted 8 July 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์ 3 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (LS) วิธีกลุ่มอนุภาค (PSO) และวิธีเชิงพันธุกรรม (GA) ทำการศึกษาทั้งในข้อมูลจริงและข้อมูลจำลอง ซึ่งมีข้อมูลสูญหาย (m) เท่ากับ 1 และ 2 ค่า กำหนดให้ข้อมูลจำลองมีสัมประสิทธิ์ความผันแปร (C.V.) เท่ากับ 10% 30% และ 50% ค่าคงที่ h เท่ากับ 1 2 และ 3 ขนาดของข้อมูล เท่ากับ 9, 12, 15, 20, 21, 25, 28, 35 โดยจำนวนทรีตเมนต์ (t) เท่ากับ 3, 4, 5 และจำนวนบล็อก (b) เท่ากับ 3, 5, 7 ซึ่งทำซ้ำในแต่ละสถานการณ์จำนวน 50000 รอบ โดยใช้ค่า MSE และ AVG MSE เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสูญหายสำหรับข้อมูลจริงและข้อมูลจำลอง ตามลำดับ โดยวิธีการประมาณค่าสูญหายวิธีใดมีค่า MSE หรือ AVG MSE ที่ต่ำกว่าวิธีนั้นจะเป็นวิธีการประมาณค่าสูญหายที่มีประสิทธิภาพดีกว่า ผลการวิจัยพบว่า สำหรับข้อมูลจริง เมื่อ m เท่ากับ 1 วิธี GA ให้ค่า MSE ต่ำที่สุด และเมื่อ m เท่ากับ 2 วิธี LS ให้ค่า MSE ต่ำที่สุด สำหรับข้อมูลจำลอง วิธี LS ให้ค่า AVG MSE ต่ำที่สุดทุกสถานการณ์ และกรณีที่ค่าคงที่ h มีค่าน้อย (h = 1) พบว่า วิธี GA จะให้ค่า AVG MSE ต่ำที่สุดเช่นเดียวกับ วิธี LS

คำสำคัญ: แผนแบบบล็อกสุ่มสมบูรณ์; ค่าสูญหาย; วิธีกำลังสองน้อยที่สุด; วิธีกลุ่มอนุภาค; วิธีเชิงพันธุกรรม

***ผู้รับผิดชอบบทความ:** tanwarat.m@ku.th

Science32-N6_1.indd 1 12/12/2567 BE 13:50

Abstract

This research aims to study and compare the 3 missing value estimation methods: Least Square (LS), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA), in randomized complete block design using both real and simulated data. We consider constraints of missing values (m), such as m = 1, 2 and assume coefficients of variation (C.V.) of 10%, 30%, and 50%. Additionally, we examine this experiment with three constants (h) 1, 2 and 3. The simulated data sizes are 9, 12, 15, 20, 21, 25, 28 and 35 with the number of treatments (t) equal to 3, 4, 5 and the number of blocks (b) equal to 3, 5, or 7. The number of iterations for each situation is 50000. The criteria for comparing the performance of missing value estimation methods for real and simulated data are the mean squared error (MSE) and the average MSE (AVG MSE) values, respectively. Both values from these estimation methods are insignificant. This means the methods are very high efficiency. The results are divided to 2 parts. For the real data set, if m = 1, the lowest MSE value is from the GA method. Nevertheless, the lowest MSE value is from the LS method when m = 2. For the simulated data, the LS method also has the smallest AVG MSE values for all situations. Lastly, the GA and LS methods get the smallest MSE values when h = 1.

Keywords: Randomized complete block design; Missing value; Least square; Particle swarm optimization; Genetic algorithm

Science32-N6_1.indd 2 12/12/2567 BE 13:50

1. บทนำ

ในปัจจุบันการเก็บรวบรวมข้อมูล เพื่อนำมาใช้ในการวิเคราะห์นั้นสามารถทำได้สองลักษณะ คือ การเก็บรวบรวมข้อมูลจากการสังเกตและจากการทดลอง ซึ่งการเก็บรวบรวมข้อมูลจากการทดลองนั้นเราสามารถควบคุมตัวแปรอื่นๆ ที่มีผลต่อการทดลองได้ไม่เหมือนกับการเก็บรวบรวมข้อมูลจากการสังเกต นอกจากนี้ข้อมูลที่ได้จากการสังเกตนั้นเมื่อนำมาวิเคราะห์จะได้เฉพาะความสัมพันธ์ของข้อมูลเท่านั้น แต่ข้อมูลที่ได้จากการทดลองเราสามารถวิเคราะห์เหตุและผลของข้อมูลนั้นได้ด้วย ทำให้การวางแผนการทดลอง (Experimental Design) จึงเป็นที่นิยมใช้กันอย่างแพร่หลายในหลายสาขา การวางแผนการทดลองมีอยู่มากมายหลายแผนแบบ การเลือกนำมาใช้งานจะขึ้นอยู่กับลักษณะงานที่ต้องการทำการทดลองและลักษณะของหน่วยทดลอง ซึ่งในทางปฏิบัติเป็นไปได้ยากที่หน่วยทดลองทุกๆ หน่วยที่นำมาทำการทดลองจะมีลักษณะเหมือนกัน จึงมีการนำแผนการทดลองแบบบล็อกสุ่มสมบูรณ์ (Randomized Complete Block Design) มาใช้โดยการจัดหน่วยทดลองเป็นกลุ่ม ๆ หรือบล็อก (Block) เพื่อลดความผิดพลาดในการวิเคราะห์ข้อมูล โดยกำหนดให้หน่วยทดลองที่อยู่ในกลุ่มหรือบล็อกเดียวกันต้องมีลักษณะเหมือนกันหรือคล้ายคลึงกัน ส่วนหน่วยทดลองที่อยู่ต่างบล็อกกันต้องมีลักษณะแตกต่างกัน และในแต่ละบล็อกมีจำนวนหน่วยทดลองที่ได้รับทรีตเมนต์ครบทุกทรีตเมนต์ [1] ดังนั้นในกรณีที่มีข้อมูลสูญหาย จึงต้องมีการประมาณค่าข้อมูลที่สูญหายก่อนที่จะทำการวิเคราะห์ข้อมูลนั้นๆ

ในปี ค.ศ. 2006 Kannika [2] ได้ศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในแผนแบบจัตุรัสละติน 3 วิธี ได้แก่ วิธีกำลังสองน้อยสุด (LS) วิธีค่าคาดหวังสูงสุด (EM) และวิธีมัลติเพิลอิมพิวเทชัน (MI) โดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ผลการวิจัยสรุปได้ว่า เมื่อเปอร์เซ็นข้อมูลสูญหายและสัมประสิทธิ์ความผันแปร (C.V.) มีค่ามาก พบว่าวิธี MI ให้ค่า MAE ต่ำกว่าวิธีอื่น สำหรับกรณีที่ เปอร์เซ็นข้อมูลสูญหายและ C.V. มีค่าน้อย พบว่าค่า MAE ของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นควรเลือกใช้วิธี LS ในการประมาณค่าสูญหายเนื่องจากเป็นวิธีที่สะดวกและรวดเร็วกว่า ต่อมาใน ค.ศ. 2011 Nootyaskool [3] ได้ศึกษาและนำเสนอวิธีการทำงานของวิธีกลุ่มอนุภาค (PSO) เมื่อเปรียบเทียบกับวิธีเชิงพันธุกรรม (GA) เพื่อให้เห็นถึงกระบวนการทำงานและการใช้งาน โดยยกตัวอย่างการแก้ปัญหาการหาผลลัพธ์ที่ดีที่สุดของฟังก์ชันทางคณิตศาสตร์ ผลจากการศึกษาพบว่าวิธี PSO มีคุณสมบัติที่เหนือกว่าวิธี GA โดยวิธี PSO สามารถนำตำแหน่งที่ได้ผลลัพธ์ดีสุดในรุ่นและตำแหน่งที่ได้ผลลัพธ์ดีที่สุดจากอดีตถึงปัจจุบัน ซึ่งแตกต่างจากวิธี GA ที่มีการแบ่งข้อมูลของตำแหน่งระหว่างประชากรในขอบเขตที่จำกัด การทดสอบการทำงานของทั้ง 2 วิธี สะท้อนให้เห็นว่าวิธีการทั้งสองสามารถลู่เข้าสู่จุดที่ได้ผลลัพธ์ที่ดีได้และใน ค.ศ. 2013 Azadeh et al. [4] ได้ทำการศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายสำหรับแผนแบบบล็อกสุ่มสมบูรณ์ 3 วิธี คือ วิธี GA วิธี PSO และ วิธี ANN ผลการศึกษาพบว่าค่าเฉลี่ยของเปอร์เซ็นความคลาดเคลื่อนสัมบูรณ์ของวิธี GA มีค่าต่ำที่สุด ดังนั้น วิธี GA เหมาะสมที่สุดที่จะใช้ประมาณค่าสูญหายสำหรับงานวิจัยนี้

ดังนั้นผู้วิจัยจึงสนใจศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์เมื่อทรีตเมนต์และบล็อกเป็นอิทธิพลแบบสุ่ม 3 วิธี ได้แก่ วิธี LS วิธี PSO และวิธี GA โดยใช้ข้อมูลจริงและข้อมูลจำลอง โดยกำหนดขอบเขตการศึกษาเปรียบเทียบด้านขนาดตัวอย่าง (n) จำนวนข้อมูลสูญหาย (m) สัมประสิทธิ์ความผันแปร (C.V.) และค่าคงที่ h ให้มีค่าที่หลากหลาย และพิจารณาประสิทธิภาพจากค่า MSE หรือ AVG MSE ในแต่ละสถานการณ์ ถ้าวิธีใดให้ค่า MSE หรือ AVG MSE ต่ำที่สุด แสดงว่าวิธีการประมาณค่าสูญหายวิธีนั้นเป็นวิธีที่มีประสิทธิภาพดีที่สุด

Science32-N6_1.indd 3 12/12/2567 BE 13:50

2. วัสดุ อุปกรณ์และวิธีการวิจัย

ในงานวิจัยนี้เป็นการศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์เมื่อทรีตเมนต์และบล็อกเป็นอิทธิพลแบบสุ่ม 3 วิธี ได้แก่ วิธี LS วิธี PSO และวิธี GA โดยใช้ข้อมูลจริงและข้อมูลจำลอง สำหรับข้อมูลจริงประกอบด้วยชุดข้อมูล 2 ชุด โดยชุดที่ 1 มีจำนวนทรีตเมนต์ (t) เท่ากับ 4 จำนวนบล็อก (b) เท่ากับ 3 และมีจำนวนข้อมูลสูญหาย (m) 1 ค่า (t=4, b=3, m=1) แสดงดังตารางที่ 1 และชุดที่ 2 มีจำนวนทรีตเมนต์ (t) เท่ากับ 6 จำนวนบล็อก (b) เท่ากับ 4 และมีจำนวนข้อมูลสูญหาย (m) 2 ค่า (t=6, b=4, m=2) แสดงดังตารางที่ 2 สำหรับข้อมูลจำลอง ในที่นี้จำลองแต่ละสถานการณ์ 50,000 รอบ โดยกำหนดให้ t=3 b=3, t=3 b=5, t=4 b=3, t=4 b=5, t=5 b=3, t=3 b=7, t=4 b=7, t=5 b=5, t=5 b=7 และ m=1, 2 โดยมีขั้นตอนดังนี้

2.1 ขั้นตอนการวิจัย

2.1.1 เนื่องจากในการศึกษานี้ กำหนดอิทธิพลของทรีตเมนต์และบล็อกเป็นแบบสุ่ม ในการจำลองข้อมูลจึงต้องกำหนดค่า C.V. เท่ากับ 10% 30% และ 50% ค่าเฉลี่ยรวมของข้อมูล (μ) เท่ากับ 60 และค่าคงที่ h เท่ากับ 1, 2 และ 3 เพื่อให้ค่าความแปรปรวนของอิทธิพลทรีตเมนต์และบล็อกมีค่าหลากหลายในแต่ละสถานการณ์ที่ศึกษา

2.1.2 กำหนดให้ $\sigma_\varepsilon^2 = \frac{[C.V.(y_{ij})(60)]^2}{2h+1}$ และจำลองข้อมูลภายใต้ $\tau_i \sim N(0,h\sigma_\varepsilon^2)$ $\beta_j \sim N(0,h\sigma_\varepsilon^2)$ และ $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$ ภายใต้สมมติฐานว่างเป็นจริง

2.1.3 จำลองข้อมูลตามตัวแบบเชิงเส้นของแผนแบบบล็อกสุ่มสมบูรณ์

2.1.4 นำข้อมูลมาทำให้สูญหาย m ค่าอย่างสุ่ม ไม่เกิน 5% ของจำนวนข้อมูลทั้งหมด (n) (ถ้า m<1 จะกำหนดให้ m=1 และปัดให้เป็นเลขจำนวนเต็ม จึงได้ m=1, 2) โดยใช้วิธีการเลือกตัวอย่างสุ่มแบบง่าย

2.1.5 ประมาณค่าข้อมูลสูญหายโดยใช้วิธีการประมาณค่า 3 วิธีที่ศึกษา คือ วิธี LS วิธี PSO และวิธี GA

2.1.6 สำหรับข้อมูลจำลองจะพิจารณาจากค่า AVG MSE

2.2 แผนแบบบล็อกสุ่มสมบูรณ์

กำหนดให้อิทธิพลของทรีตเมนต์และบล็อกมีอิทธิพลสุ่ม ดังนั้น ตัวแบบที่ศึกษาในงานครั้งนี้ คือ

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \; ; i = 1, 2, ..., t \; ; j = 1, 2, ..., b \quad (1)$$

โดยที่ y_{ij} คือ ค่าสังเกตจากทรีตเมนต์ที่ i และบล็อกที่ j

μ คือ ค่าเฉลี่ยรวม (Grand Mean)

τ_i คือ อิทธิพลของทรีตเมนต์ที่ i

β_j คือ อิทธิพลของบล็อกที่ j

ε_{ij} คือ ความคลาดเคลื่อนสุ่ม (Random Error) โดยที่ $\varepsilon_{ij} \sim NID(0,\sigma_\varepsilon^2)$

2.3 วิธีการประมาณค่าข้อมูลสูญหาย

2.3.1 วิธีกำลังสองน้อยที่สุด (LS)

การประมาณค่าสังเกตสูญหายเมื่อมีค่าสังเกตสูญหายเพียง 1 ค่า คือ ค่าที่ประมาณขึ้นเป็นค่าที่ทำให้ผลรวมของความคลาดเคลื่อนกำลังสอง (SSE) มีค่าน้อยที่สุด และวิธีนี้ยังสามารถนำไปใช้หาค่าประมาณของค่าสังเกตสูญหายหลายค่าโดยวิธีวนซ้ำ [2] สามารถคำนวณได้ดังสมการ (2)

$$M_{ij} = \frac{bB_j + tT_i - T}{(b-1)(t-1)} \quad (2)$$

เมื่อ M_{ij} คือ ค่าสังเกตสูญหายซึ่งอยู่ในทรีตเมนต์ที่ i และบล็อกที่ j

B_j คือ ผลรวมของบล็อกที่มีค่าสังเกตสูญหาย

T_i คือ ผลรวมของทรีตเมนต์ที่มีค่าสังเกตสูญหาย

T คือ ผลรวมของค่าสังเกตทั้งหมด

t และ b คือ จำนวนทรีตเมนต์และจำนวนบล็อกตามลำดับ

Science32-N6_1.indd 4 12/12/2567 BE 13:50

2.3.2 วิธีกลุ่มอนุภาค (PSO)

ใน ค.ศ. 1995 Eberhart และ Kennedy [7] ได้พัฒนาวิธี PSO ให้เป็นวิธีการค้นหาคำตอบที่ดีที่สุดสำหรับปัญหาการหาค่าที่ดีที่สุดแบบไม่มีเงื่อนไข (Unconstrained Optimization Problem) ซึ่งต้องการคำตอบที่ทำให้ค่าของฟังก์ชันวัตถุประสงค์ (Objective Function) มีค่าดีที่สุด ซึ่งอาจหมายถึงค่าที่มากที่สุดหรือน้อยที่สุดก็ได้ตามแต่วัตถุประสงค์ของปัญหานั้นๆ

วิธีการค้นหาค่าที่ดีที่สุด PSO แบบมาตรฐาน พาร์ทิเคิลแต่ละตัวจะมีการปรับความเร็วใหม่ด้วยสมการที่ (3) และจากนั้นทำการปรับตำแหน่งใหม่ด้วยสมการที่ (4) โดยสมการทั้งสองเป็นค่าในแต่ละมิติของความเร็วและตำแหน่ง วิธี PSO แบบมาตรฐานนับว่าเป็นวิธีที่นิยมใช้มากที่สุดเนื่องจากความง่ายของสมการ

$$V_D' = (w(t)*V_D) + (c_p*u_p*(Pbest - X_D)) + (c_g*u_g*(Gbest - X_D)) \quad (3)$$

$$X_D' = X_D + V_D' \quad (4)$$

เมื่อ V_D' คือ ค่าความเร็วที่คำนวณได้จากอนุภาครุ่นปัจจุบัน

V_D คือ ค่าความเร็วที่คำนวณได้จากอนุภาครุ่นก่อน

X_D' คือ ค่าตำแหน่งที่คำนวณได้จากอนุภาครุ่นปัจจุบัน

X_D คือ ค่าตำแหน่งที่คำนวณได้จากอนุภาครุ่นก่อน

Pbest คือ ค่าความเหมาะสมที่ดีที่สุดที่กลุ่มอนุภาครุ่นปัจจุบันได้มา

Gbest คือ ค่าตำแหน่งที่อนุภาคนั้นได้ค่าความเหมาะสมสูงสุด

w(t) คือ ค่าการถ่วงน้ำหนัก

c_p คือ ค่าคงที่อัตราเร่งของตำแหน่งที่ดีที่สุดส่วนตัว

c_g คือ ค่าคงที่อัตราเร่งของตำแหน่งที่ดีที่สุดสากล

u_p และ u_g คือ ตัวเลขสุ่มที่มีค่าในช่วง [0, 1]

2.3.3 วิธีเชิงพันธุกรรม (GA)

วิธีเชิงพันธุกรรม (Genetic Algorithm) เป็นอัลกอริทึมที่ใช้ในการค้นหาคำตอบที่ดีที่สุด (Optimization) โดยใช้หลักการคัดเลือกสายพันธ์แบบธรรมชาติ (Natural Selection) จากการจำลองแนวคิดวิวัฒนาการของสิ่งมีชีวิต โดยอิงหลักการคัดเลือกตามธรรมชาติของดาร์วิน (Darwin's theory of natural selection) ซึ่งแนวคิดดังกล่าวถูกนำมาประยุกต์ใช้เป็นวิธีการแก้ปัญหาโดย John Holland ใน ค.ศ. 1975 [8] มีขั้นตอนดังนี้

1. การกำหนดฟังก์ชันความเหมาะสม (Fitness Function) คือ การกำหนดฟังก์ชันความเหมาะสมของแต่ละปัญหา ซึ่งอาจเป็นการวัดค่าความเหมาะสมสูงสุดหรือต่ำสุดก็ได้

2. การกำหนดค่าเริ่มต้น (Initialization) โดยการสุ่มค่าคำตอบซึ่งเป็นคำตอบที่อยู่ในขอบเขตของคำตอบ (Solution space) ขึ้นมา เพื่อที่จะถูกคัดเลือกเป็นต้นแบบสำหรับการสร้างประชากรรุ่นใหม่ โดยในงานวิจัยนี้สร้างข้อมูลจากการประมาณค่าข้อมูลสูญหายมาจาก 3 วิธี คือ

1) วิธี Mean คือ จะใช้ค่าเฉลี่ยของข้อมูลที่ไม่สูญหายทั้งหมด

2) วิธี Mean of Block คือ จะใช้ค่าเฉลี่ยของข้อมูลที่ไม่สูญหายในบล็อกที่มีข้อมูลสูญหายอยู่

3) วิธี Mean of Treatment คือ จะใช้ค่าเฉลี่ยของข้อมูลที่ไม่สูญหายในทรีตเมนต์ที่มีข้อมูลสูญหายอยู่

3. การคัดเลือก (Selection) เป็นวิธีการประเมินค่าความเหมาะสมของข้อมูลที่สร้างขึ้นมา โดยที่ข้อมูลที่สร้างมาจากการประมาณค่าข้อมูลสูญหาย 3 วิธี คือ วิธี Mean วิธี Mean of Block และวิธี Mean of Treatment ข้อมูลที่เหมาะสมจะถูกเลือกให้เป็น Elitism และข้อมูลชุดที่เหลือจะเข้าสู่กระบวนการวิธี GA ต่อไป

Science32-N6_1.indd 5 12/12/2567 BE 13:50

4. การผลิตรุ่นถัดไป (New Population)

4.1 การผสม (Blending) คือ การผสมระหว่างโครโมโซมตั้งต้น สมมติให้ A_a และ B_b เป็นค่าที่มาจากวิธีการคัดเลือกค่า Elitism ถ้าความน่าจะเป็นของการทดลองผ่าน ค่า A_a และ B_b จะผสมกัน จะได้ค่าใหม่ดังนี้

$$A_a^* = \beta_1 A_a + (1-\beta_1) B_b \quad (5)$$

$$B_b^* = (1-\beta_1) A_a + \beta_1 B_b \quad (6)$$

โดย A_a^* และ B_b^* เป็นโครโมโซมใหม่

4.2 การสลับสายพันธุ์ (Crossover) คือ การนำเอาโครโมโซม 2 โครโมโซมมาทำการตามขั้นตอนต่างๆ ซึ่งจะให้ค่าโครโมโซมใหม่ที่จะนำไปใช้ในการคัดเลือกต่อไป หรือการนำโครโมโซม 2 โครโมโซมมาผสมกันเพื่อได้โครโมโซมใหม่ เป็นกระบวนการที่สำคัญของขั้นตอนวิธี GA เมื่อเกิดการสลับสายพันธุ์ขึ้นในทางพันธุกรรมจะทำให้เกิดการเปลี่ยนแปลง [9] ซึ่งในงานวิจัยนี้กำหนดเกณฑ์ในการผสมและการสลับสายพันธุ์ดังนี้

ถ้า $\beta_0 \geq 0.10$ ไม่ต้องทำขั้นตอนนี้

$\beta_0 < 0.10$ ให้ทำขั้นตอนนี้

โดย β_0 คือ ความน่าจะเป็นของเกณฑ์การผสมและการสลับสายพันธุ์

5. การสิ้นสุดการทำงาน โดยเงื่อนไขในการสิ้นสุดการทำงานของวิธี GA จะสิ้นสุดการทำงานก็ต่อเมื่อได้ดำเนินการจนครบตามจำนวนรุ่นของประชากร (Number of generation) ตามที่ผู้ใช้กำหนด [8] หรือผู้ใช้อาจจะกำหนดให้สิ้นสุดการทำงานเมื่อขั้นตอนการเลือก Elitism ของรอบใหม่ได้ประชากรเริ่มต้นเหมือนประชากรเริ่มต้นรอบก่อนหน้านี้ ซึ่งในที่นี้คือ การได้ค่าประมาณของข้อมูลสูญหายคงที่

2.4 ข้อมูลจริงที่ใช้สำหรับศึกษาการประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์

ข้อมูลชุดที่ 1 แสดงข้อมูลเครื่องจักร 4 เครื่องที่นำมาจาก 4 บริษัท (Treatment) เพื่อผลิตสินค้า 3 ชนิด (Block) ได้ปริมาณเท่ากันหรือไม่ โดยการทดลองผู้ทดลองเลือกพนักงานที่มีความสามารถเท่ากันเข้าทำงานกับเครื่องจักรอย่างสุ่มได้ผลผลิตเป็นจำนวนสินค้าอย่างหนึ่ง [5] ดังนี้

Table 1 Real data set 1 (t=4 b=3 m=1)

Block		1	2	3
Treatment	1	17	34	23
	2	15	26	(21)
	3	11	23	18
	4	12	14	13

"()" indicates the missing value.

ข้อมูลชุดที่ 2 แสดงข้อมูลการทดลองในตู้เพาะเลี้ยงของพืชในแต่ละ genotype ที่มาจาก 6 genotypes ถูกสุ่มให้ได้รับช่วงแสง (photoperiod) ที่ต่างกัน 4 ระดับ คือ ช่วงแสงสั้นมาก (Very short) ช่วงแสงสั้น (Short) ช่วงแสงยาว (Long) และช่วงแสงยาวมาก (Very long) [6] ได้ผลดังนี้

Table 2 Real data set 2 (t=6 b=4 m=2)

Genotypes	Very short	Short	Long	Very long
A	(2)	3	3	4
B	3	(4)	5	6
C	1	2	1	2
D	1	1	2	2
E	2	2	2	2
F	1	1	2	3

"()" indicates the missing value.

Science32-N6_1.indd 6 12/12/2567 BE 13:50

2.5 เกณฑ์ในการเปรียบเทียบ

ในการศึกษานี้จะพิจารณาเกณฑ์การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสูญหายด้วยวิธีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) สามารถคำนวณดังสมการ (7)

$$MSE = \frac{\sum_{i=1}^{t}\sum_{j=1}^{b}(y_{ij} - \hat{y}_{ij})^2}{n} \qquad (7)$$

โดยที่ y_{ij} คือ ค่าจริงของข้อมูล

$\hat{y}_{ij}$ คือ ค่าที่ประมาณได้ของข้อมูล

n คือ จำนวนข้อมูลทั้งหมด (รวมข้อมูลที่สูญหาย)

t และ b คือ จำนวนทรีตเมนต์และจำนวนบล็อกตามลำดับ

สำหรับข้อมูลจำลอง เนื่องจากกระทำซ้ำ 50000 รอบ (k) ในแต่ละสถานการณ์ จึงคำนวณค่า Average of Mean Square Error (AVG MSE) ดังนี้

$$AVG\ MSE = \frac{\sum_{s=1}^{k} MSE_s}{k} \qquad (8)$$

โดยพิจารณาว่าวิธีการประมาณค่าสูญหายวิธีใดให้ค่า MSE หรือ AVG MSE ต่ำที่สุด แสดงว่าวิธีการประมาณค่าสูญหายวิธีนั้นเป็นวิธีที่มีประสิทธิภาพดีที่สุด

3. ผลการทดลอง

จากข้อมูลทั้งในชุดข้อมูลจริงและข้อมูลจำลองสุ่มค่าสูญหาย (m) จำนวน 1 และ 2 ค่า แบบการสูญหายอย่างสุ่มสมบูรณ์ (Missing Completely at Random: MCAR) นั่นคือข้อมูลที่สูญหายไม่มีความสัมพันธ์กับค่าอื่นๆ ทั้งที่ทราบค่าและไม่ทราบค่า และประมาณค่าสูญหายจากทั้ง 3 วิธี คือ วิธี LS วิธี PSO และวิธี GA พิจารณาค่า MSE และค่า AVG MSE โดยวิธีใดให้ค่า MSE และค่า AVG MSE น้อยที่สุดจะเป็นวิธีที่มีประสิทธิภาพมากที่สุด

สำหรับข้อมูลจริงชุดที่ 1 ที่มี m เท่ากับ 1 พบว่า วิธี GA ให้ค่า MSE ต่ำที่สุด และข้อมูลจริงชุดที่ 2 ที่มี m เท่ากับ 2 พบว่า วิธี LS ให้ค่า MSE ต่ำที่สุด (Table 3)

สำหรับข้อมูลจำลองที่มี m เท่ากับ 1 ที่ C.V. เท่ากับ 10% สำหรับ h เท่ากับ 1 พบว่า วิธี GA ให้ค่า AVG MSE ต่ำที่สุด สำหรับ h เท่ากับ 2 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุดในแผนแบบ t=4 b=5 และ t=5 b=3 วิธี GA ให้ค่า AVG MSE ต่ำที่สุดในแผนแบบ t=3 b=3 และ t=4 b=3 ส่วนแผนแบบ t=3 b=5 วิธี LS และ GA ให้ค่า AVG MSE เท่ากัน และสำหรับ h เท่ากับ 3 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด ที่ C.V. เท่ากับ 30% และ 50% สำหรับ h เท่ากับ 1 พบว่า วิธี GA ให้ค่า AVG MSE ต่ำที่สุด และสำหรับ h เท่ากับ 2 และ 3 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด (Table 4)

สำหรับข้อมูลจำลองที่มี m เท่ากับ 2 ที่ C.V. เท่ากับ 10% สำหรับ h เท่ากับ 1 พบว่า วิธี GA ให้ค่า AVG MSE ต่ำที่สุด และสำหรับ h เท่ากับ 2 และ 3 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด ที่ C.V. เท่ากับ 30% สำหรับ h เท่ากับ 1 พบว่า วิธี GA ให้ค่า AVG MSE ต่ำที่สุด สำหรับแผนแบบ t=5 b=7 วิธี LS ให้ค่า AVG MSE ต่ำที่สุด และสำหรับ h เท่ากับ 2 และ 3 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด และ ที่ C.V. เท่ากับ 50% สำหรับ h เท่ากับ 1 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด สำหรับแผนแบบ t=3 b=7 วิธี GA ให้ค่า AVG MSE ต่ำที่สุด และสำหรับ h เท่ากับ 2 และ 3 พบว่า วิธี LS ให้ค่า AVG MSE ต่ำที่สุด (Table 5) MSE value of real data set 1 (t=4 b=3 m=1)

Science32-N6_1.indd 7 12/12/2567 BE 13:50

Table 3 MSE values of the real data set 1 (t=4, b=3, m=1) and set 2 (t=6, b=4, m=2)

Data sets	Methods	MSE
1	LS	0.08
	PSO	0.43
	GA	**0.02**
2	**LS**	**0.02**
	PSO	0.21
	GA	0.09

"Bold number" indicates the lowest value.

Table 4 AVG MSE values of simulated data when m=1

No. of Treatment			3	3	4	4	5
No. of Block			3	5	3	5	3
Criteria			AVG MSE	AVG MSE	AVG MSE	AVG MSE	AVG MSE
C.V. = 10%	h = 1	LS	3.02	1.51	2.00	1.00	1.51
		PSO	3.87	2.13	2.87	1.58	2.31
		GA	**1.91**	**1.08**	**1.34**	**0.76**	**1.09**
	h = 2	LS	1.82	**0.90**	1.20	**0.60**	**0.89**
		PSO	3.61	1.96	2.65	1.49	2.14
		GA	**1.65**	**0.90**	**1.16**	0.65	0.90
	h = 3	LS	**1.28**	**0.65**	**0.87**	**0.43**	**0.64**
		PSO	3.50	1.92	2.61	1.44	2.07
		GA	1.53	0.86	1.10	0.60	0.84
C.V. = 30%	h = 1	LS	26.95	13.57	17.92	8.98	13.53
		PSO	34.87	18.97	26.00	14.13	20.70
		GA	**19.94**	**11.27**	**14.17**	**7.94**	**11.19**
	h = 2	LS	**16.20**	**8.13**	**10.80**	**5.33**	**8.04**
		PSO	32.62	17.69	24.04	13.24	19.27
		GA	17.23	9.75	12.43	6.89	9.72
	h = 3	LS	**11.68**	**5.80**	**7.72**	**3.82**	**5.83**
		PSO	31.43	17.07	23.00	12.82	18.70
		GA	16.01	8.99	11.47	6.51	8.93

Science32-N6_1.indd 8 12/12/2567 BE 13:50

Table 4 AVG MSE values of simulated data when m=1 (Continue)

No. of Treatment			3	3	4	4	5
No. of Block			3	5	3	5	3
Criteria			AVG MSE	AVG MSE	AVG MSE	AVG MSE	AVG MSE
C.V. =50%	h = 1	LS	74.77	37.23	49.85	**24.83**	37.15
		PSO	96.30	52.38	71.94	39.70	57.67
		GA	**61.88**	**34.61**	**44.90**	25.03	**34.53**
	h = 2	LS	**45.15**	**22.67**	**29.73**	**14.93**	**22.48**
		PSO	90.95	48.96	66.81	37.44	53.09
		GA	54.51	30.13	38.62	21.99	30.50
	h = 3	LS	**32.28**	**15.97**	**21.47**	**10.71**	**16.07**
		PSO	87.84	47.93	65.20	35.60	52.17
		GA	51.91	28.36	36.67	20.38	29.02

"Bold number" indicates the lowest value.

Table 5 AVG MSE values of simulated data when m=2

No. of Treatment			3	4	5	5
No. of Block			7	7	5	7
Criteria			AVG MSE	AVG MSE	AVG MSE	AVG MSE
C.V. = 10%	h = 1	LS	1.89	1.28	1.46	0.98
		PSO	3.89	2.89	3.27	2.33
		GA	**1.63**	**1.17**	**1.34**	**0.93**
	h = 2	LS	**1.15**	**0.77**	**0.88**	**0.59**
		PSO	3.90	2.91	3.27	2.34
		GA	1.42	1.03	1.16	0.81
	h = 3	LS	**0.83**	**0.56**	**0.63**	**0.42**
		PSO	3.89	2.91	3.26	2.33
		GA	1.34	0.97	1.10	0.76
C.V. = 30%	h = 1	LS	16.94	11.60	13.12	**8.80**
		PSO	31.24	23.50	26.36	18.76
		GA	**15.49**	**11.26**	**12.72**	8.86
	h = 2	LS	**10.36**	**6.94**	**7.92**	**5.31**
		PSO	31.35	23.47	26.51	18.83
		GA	13.63	9.88	11.21	7.75
	h = 3	LS	**7.51**	**5.01**	**5.69**	**3.82**
		PSO	31.46	23.46	26.33	18.79
		GA	12.84	9.29	10.50	7.28

Science32-N6_1.indd 9 12/12/2567 BE 13:50

Table 5 AVG MSE values of simulated data when m=2 (Continue)

No. of Treatment			3	4	5	5
No. of Block			7	7	5	7
Criteria			AVG MSE	AVG MSE	AVG MSE	AVG MSE
C.V. =50%	h = 1	LS	**47.11**	**32.18**	**36.39**	**24.47**
		PSO	86.17	64.79	72.41	51.69
		GA	**44.79**	32.61	37.02	25.80
	h = 2	LS	**28.83**	**19.28**	**22.01**	**14.73**
		PSO	86.43	64.62	72.70	51.72
		GA	40.01	28.99	32.73	22.80
	h = 3	LS	**20.86**	**13.97**	**15.81**	**10.61**
		PSO	86.36	64.63	72.56	51.72
		GA	37.88	27.39	30.97	21.51

"Bold number" indicates the lowest value.

4. อภิปรายผลและสรุป

จากการศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในแผนแบบบล็อกสุ่มสมบูรณ์เมื่อทรีตเมนต์และบล็อกเป็นอิทธิพลแบบสุ่ม 3 วิธี คือ วิธี LS วิธี PSO และวิธี GA ข้อมูลที่ใช้ในการทดสอบเป็นข้อมูลจริงและข้อมูลจำลอง โดยมีค่า m เท่า 1 และ 2 สำหรับข้อมูลจริงและข้อมูลจำลองจะใช้ค่า MSE และ AVG MSE เป็นเกณฑ์ในการเปรียบเทียบ ตามลำดับ ผลการศึกษาพบว่า ถ้าค่า C.V. เพิ่มมากขึ้น ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) จะมีค่ามากขึ้นด้วย สำหรับข้อมูลจริงที่มี m เท่ากับ 1 วิธี GA ให้ค่า MSE ต่ำที่สุด และข้อมูลจริงที่มี m เท่ากับ 2 วิธี LS ให้ค่า MSE ต่ำที่สุด สำหรับข้อมูลจำลองที่มี m เท่ากับ 1 เมื่อค่าคงที่ h เท่ากับ 1 วิธี GA มีแนวโน้มให้ค่า AVG MSE ต่ำที่สุดทุกสถานการณ์ และเมื่อค่าคงที่ h มากขึ้น (h = 2, 3) วิธี LS มีแนวโน้มให้ค่า AVG MSE ต่ำที่สุด และสำหรับข้อมูลจำลองที่มี m เท่ากับ 2 เมื่อค่าคงที่ h เท่ากับ 1 วิธี LS และ GA มีแนวโน้มให้ค่า AVG MSE ต่ำที่สุด และเมื่อค่าคงที่ h มากขึ้น (h = 2, 3) วิธี LS ให้ค่า AVG MSE ต่ำที่สุดทุกสถานการณ์

สำหรับผลการศึกษาข้อมูลจริงและข้อมูลจำลอง เมื่อพิจารณาจากเกณฑ์ MSE และ AVG MSE พบว่า วิธีการประมาณค่าจะให้ผลการศึกษาที่แตกต่างกันในบางสถานการณ์ โดยสำหรับข้อมูลจริงเมื่อ m=1 วิธี GA ให้ค่า MSE ต่ำที่สุด และเมื่อ m=2 วิธี LS ให้ค่า MSE ต่ำที่สุด สำหรับข้อมูลจำลอง วิธี LS ให้ค่า AVG MSE ต่ำที่สุด แต่เมื่อค่าคงที่ h มีค่าน้อย (h =1) ส่วนใหญ่วิธี LS และวิธี GA จะให้ค่า AVG MSE ต่ำที่สุด ซึ่งสอดคล้องกับงานวิจัยของ Azadeh et al. [4]

5. ข้อเสนอแนะ

ในงานวิจัยนี้ ผู้วิจัยศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหาย 3 วิธี คือ วิธี LS วิธี PSO และวิธี GA สำหรับแผนแบบบล็อกสุ่มสมบูรณ์ เมื่อทรีตเมนต์และบล็อกเป็นอิทธิพลแบบสุ่ม โดยในงานวิจัยนี้กำหนดค่าเริ่มต้นของวิธี PSO จากค่า Mean, Mean of Block และ Mean of Treatment ของชุดข้อมูลที่มีค่าสูญหาย ดังนั้นในงานวิจัยต่อไปอาจกำหนดค่าเริ่มต้นของวิธี PSO แบบอื่นๆ เช่น พิจารณาค่าเริ่มต้นด้วยวิธี LS หรือวิธี GA

Science32-N6_1.indd 10 12/12/2567 BE 13:50

6. กิตติประกาศ

ขอขอบคุณโครงการพัฒนาและส่งเสริมผู้มีความสามารถพิเศษทางวิทยาศาสตร์และเทคโนโลยี (พสวท.) และภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ที่ให้การสนับสนุนทุนวิจัย

7. References

[1] Chomtee, B., 2013, Statistical Experimental Design: Theory and Analysis with SAS, Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, 294 p. (in thai)

[2] Kannika, S., 2006, A Comparison of Missing Value Estimation Method for Latin Square Design, Master Thesis, Chulalongkorn University, Bangkok, 167 p. (in thai)

[3] Nootyaskool, S., 2011, Evolutionary Computation between Genetic Algorithm and Particle Swarm Optimization, JIST 2 (2): 13-22. (in thai)

[4] Azadeh, A., Asadzadeh, S.M., Jafari-Marandi, R., Nazari-Shirkouhi, S., Khoshkhou, G.B., Talebi, S. & Naghavi, A., 2013, Optimum Estimation of Missing Values in Randomized Complete Block Design by Genetic Algorithm, Knowledge-based Systems 37: 37-47.

[5] Kaiyawan, Y., 2022, Analysis of Randomized Complete Block Designs: RCBD, FIT SSRU J. 10 (2): 23 - 36. (in thai)

[6] Kittipaakul, P., Guide to Statistical Analysis Using R, Available Source: http://forprod.forest.go.th, August 9, 2023. (in thai)

[7] Pongchairerks, P., 2014, An Explanation of Particle Swarm Optimization's Procedure Via Examples, SWU. Sci. J. 6 (11): 78 - 87. (in thai)

[8] Khadwilard, A., 2011, Application of Genetic Algorithm for Optimisation Problems, RMUTP sci. J. 5 (2): 153 - 163. (in thai)

[9] Kaewchaicharoenkit, C., 2020, A Comparison of Missing Value Estimation in Randomized Complete Block Design, Master Thesis, Kasetsart University, Bangkok, 148 p. (in thai)

Science32-N6_1.indd 11 12/12/2567 BE 13:50