การศึกษาประสิทธิภาพวิธีประมาณค่าสูญหายเพื่อใช้ในการวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์

ปริญญานิพนธ์
ของ
วิสุทธิดา ศรีดวงโชติ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางการศึกษา
มีนาคม 2556

การศึกษาประสิทธิภาพวิธีประมาณค่าสูญหายเพื่อใช้ในการวางแผนการทดลองแบบ
แฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์

ปริญญานิพนธ์
ของ
วิสุทธิดา ศรีดวงโชติ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางการศึกษา
มีนาคม 2556

การศึกษาประสิทธิภาพวิธีประมาณค่าสูญหายเพื่อใช้ในการวางแผนการทดลองแบบ
แฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์

บทคัดย่อ

ของ

วิสุทธิดา ศรีดวงโชติ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและสถิติทางการศึกษา
มีนาคม 2556

วิสุทธิดา ศรีดวงโชติ. (2556). *การศึกษาประสิทธิภาพวิธีประมาณค่าสูญหายเพื่อใช้ในการวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์.* ปริญญานิพนธ์ กศ.ม. (การวิจัยและสถิติทางการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม : อ.ดร.สุวิมล กฤชคฤหาสน์, อ.ดร.อิทธิฤทธิ์ พงษ์ปิยะรัตน์.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายเคเนียเรส เนย์บอร์อิมพิวเทชันในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) จำลองข้อมูลด้วยเทคนิคมอนติคาร์โลซิมูเลชัน โดยใช้โปรแกรม R กำหนดให้ปัจจัย A มี 3 ลักษณะคือ 3, 4และ5 ระดับ, ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, จำนวนบล็อกเท่ากับ 3,4 และ5 บล็อก, สัมประสิทธิ์ความผันแปรเท่ากับ 5%,25% และ45% และข้อมูลสูญหายเท่ากับ 10%,20%และ30% ในรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ(Arbitrary) ทำการเปรียบเทียบวิธีประมาณค่าสูญหาย โดยพิจารณาจากค่าความคลาดเคลื่อนสมบูรณ์สูงสุด (Maximum Absolute Error) วิธีใดให้ค่าความคลาดเคลื่อนสูงสุดมีค่าต่ำสุดแสดงว่าวิธีนั้นเหมาะสมกับการประมาณค่าสูญหาย

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลที่เกิดจากการจำลองด้วยรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ กำหนดให้ประชากรมีขนาดเท่ากับ 512 ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ จำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 1,000 รอบ

ผลการวิจัยสรุปได้ว่า เมื่อเปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่ามากขึ้น ค่าความคลาดเคลื่อนสมบูรณ์สูงสุดจะมีค่าเพิ่มขึ้นด้วย ในสถานการณ์ที่กำหนด การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน ให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำ และวิธีประมาณค่าสูญหายของวิลคินซัล ดังนั้น ถ้าทำการทดลองในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ ตามสถานการณ์ต่างๆ ที่กำหนด และมีรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ ควรเลือกใช้การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนเบอร์อิมพิวเทชัน เพราะให้ค่าใกล้เคียงกับค่าจริงมากที่สุดและสะดวกต่อการใช้งาน

A STUDY ON THE EFFICIENCY OF MISSING DATA ESTIMATION METHODS FOR TWO FACTORS FACTORIAL EXPERIMENT IN RANDOMIZED COMPLETE BLOCK DESIGN

AN ABSTRACT

BY

WISUTTIDA SRIDUANGCHOT

Present in Partial Fulfillment of the Requirements for the
Master of Education Degree in Educational Research and Statistics
at Srinakharinwirot University
March 2013

Wisuttida Sriduangchot. (2013). *A study on the efficiency of missing data estimation methods for two factors factorial experiment in randomized complete block design.* Master thesis, M.Ed. (Educational Research and Statistics). Bangkok: Graduate School, Srinakharinwirot University. Advisor committee: Dr. Suwimon Kritkharuehart, Dr.Ittirith Phongpiyaratana.

The objectives of this research were to study and compare three method of missing value estimation for two factors factorial experiment in randomized complete block design (RCB) such as Iterative Method, Wilkinson Method, and K-Nearest Neighbor Imputation Method. The datasets used in this study were simulated by the Monte Carlo technique; used for data generated through R program code. Prescriptive factor A has 3 different levels 3,4 and 5 ; factor B has 3 different levels 3,4 and 5 and the number of level for block is 3,4 and 5 ; the coefficient of variation (C.V.) for the observed data varied from 5%, 25% and 45% and the level of missing data is 10%, 20% and 30%. The pattern of missing data was arbitrary. The Maximum Absolute Error (MAE) was used as criterion to compare the three methods to determine the most efficient method for missing data estimation.

The population used in this study, were data obtained from the simulation with the linear model of two factors factorial experiment in randomized complete block design. The population size 512 at random sample size of 27, 36, 45, 48, 60, 64, 75, 80, 100 and 125 were tested for every 1,000 simulation times.

The study revealed that once the coefficient of variation and the percentage of missing data was greater, the accuracy of MAE would be increased at the maximum. Under such given situation, the estimated missing data using K-Nearest Neighbor Imputation Method, proved to have the MAE lesser than missing value estimated by Iterative Method and Wilkinson Method. The study therefore concludes that the missing value estimation methods for two factors factorial experiment in randomized complete block design under different scenarios to determine the format and pattern of missing data in arbitrary pattern, could be efficiently estimated using K-Nearest Neighbor Imputation Method as the most efficient and accurate method to determine missing data closest to the real data.

ปริญญานิพนธ์

เรื่อง
การศึกษาประสิทธิภาพวิธีประมาณค่าสูญหายเพื่อใช้ในการวางแผนการทดลองแบบ
แฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์

ของ
วิสุทธิดา ศรีดวงโชติ

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต สาขาการวิจัยและสถิติทางการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนกุล)
วันที่ เดือน พ.ศ. 2556

อาจารย์ที่ปรึกษาปริญญานิพนธ์

.... ที่ปรึกษาหลัก
(ดร.สุวิมล กฤชคฤหาสน์)

.... ที่ปรึกษาร่วม
(ดร.อิทธิฤทธิ์ พงษ์ปิยะรัตน์)

คณะกรรมการสอบปากเปล่า

.... ประธาน
(รองศาสตราจารย์ ชูศรี วงศ์รัตนะ)

....กรรมการ
(ดร.สุวิมล กฤชคฤหาสน์)

....กรรมการ
(ดร.อิทธิฤทธิ์ พงษ์ปิยะรัตน์)

....กรรมการ
(รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์)

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จาก
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดีโดยได้รับความกรุณาเป็นอย่างยิ่งจากอาจารย์ ดร.สุวิมล กฤชคฤหาสน์ ประธานควบคุมปริญญานิพนธ์ และอาจารย์ดร.อิทธิฤทธิ์ พงษ์ปิยะรัตน์ กรรมการควบคุมปริญญานิพนธ์ ที่ได้สละเวลาอันมีค่าในการให้คำปรึกษา และแนะนำการทำงานวิจัยนี้ทุกขั้นตอนจนเสร็จสมบูรณ์ ผู้วิจัยซาบซึ้งถึงความกรุณาดังกล่าว และขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

กราบขอบพระคุณรองศาสตราจารย์ ชูศรี วงศ์รัตนะและรองศาสตราจารย์ดร.องอาจ นัยพัฒน์ ที่กรุณาเป็นกรรมการในการสอบปริญญานิพนธ์และได้ให้ข้อเสนอแนะเพิ่มเติมที่เป็นประโยชน์ในการทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

กราบขอบพระคุณอาจารย์คุณาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้วิจัยในการศึกษาตามหลักสูตรการวิจัยและสถิติทางการศึกษา อันเป็นพื้นฐานสำคัญในการวิจัยครั้งนี้

กราบขอบพระคุณบิดา มารดา และครอบครัวที่ช่วยส่งเสริมและสนับสนุนการเรียนของผู้วิจัยตลอดมา และเป็นกำลังใจที่สำคัญยิ่งซึ่งมีส่วนช่วยให้ปริญญาเล่มนี้เสร็จลุล่วงด้วยดี

ขอขอบคุณพี่ๆ กรมวิชาการเกษตร เพื่อนๆ ปริญญาโทสาขาการวิจัยและสถิติทางการศึกษา และผู้ที่เกี่ยวข้องทุกท่านที่มิได้กล่าวนามไว้ ณ ที่นี้ ที่กรุณาให้ความช่วยเหลือและเป็นกำลังใจในการทำปริญญานิพนธ์ฉบับนี้

ปริญญานิพนธ์ได้รับทุนการศึกษาเพื่อทำปริญญานิพนธ์จากบัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ ประจำปีงบประมาณ 2555 ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

สุดท้ายนี้ คุณประโยชน์และความดีที่ได้รับจากปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบให้แด่บิดา มารดา ครูอาจารย์ และผู้มีอุปการคุณทุกท่านที่คอยอบรมสั่งสอน ให้ความช่วยเหลือ และให้กำลังด้วยดีตลอดมา

วิสุทธิดา ศรีดวงโชติ

สารบัญ

บทที่	หน้า
1 บทนำ………………………………………………………………………………..	1
ภูมิหลัง……………………………………………………………………………	1
ความมุ่งหมายของการวิจัย…………………………………………………………	5
ความสำคัญของการวิจัย……………………………………………………………	5
ขอบเขตของการวิจัย……………………………………………………………...	5
นิยามศัพท์เฉพาะ…………………………………………………………………....	7
กรอบแนวคิดในการวิจัย……………………………………………………………	9
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง…………………………………………………………..	12
ทฤษฎีที่เกี่ยวข้อง……………………………………………………………………..	12
ลักษณะของข้อมูลสูญหาย……………………………………………………………	38
ประชากรและการสุ่มตัวอย่าง………………………………………………………...	43
การจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล…………………………………………..	48
เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีประมาณค่าสูญหาย……………………………….....	51
งานวิจัยที่เกี่ยวข้อง……………………………………………………………………	51
3 วิธิดำเนินการวิจัย……………………………………………………………………..	63
การกำหนดขนาดประชากรและกลุ่มตัวอย่าง………………………………………..	63
การจำลองข้อมูลด้วยโปรแกรม R……………………………………………………..	64
ขั้นตอนการทำงานของโปรแกรม…………………………………………………….....	69
4 ผลการดำเนินงานวิจัย…………………………………………………………………	77
สัญลักษณ์และอักษรย่อในการวิเคราะห์ข้อมูล………………………………………	77
การเสนอผลการวิเคราะห์ข้อมูล……………………………………………………..	78
ผลการวิเคราะห์ข้อมูล………………………………………………………………..	78

สารบัญ (ต่อ)

บทที่	หน้า
5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ……………………………………..	100
สังเขปจุดมุ่งหมายและวิธีดำเนินการวิจัย……………………………….....................	100
สรุปผลการวิจัย…………………………………………………………………………	101
อภิปรายผล……………………………………………………………………………...	105
ข้อเสนอแนะ……………………………………………………………………………..	109
บรรณานุกรม………………………………………………………………………………	110
ภาคผนวก…………………………………………………………………………………	116
ภาคผนวก ก. ตัวอย่างการใช้โปรแกรม R การประมาณค่าสูญหายวิลคินซัล..………....	117
ภาคผนวก ข. สัญลักษณ์ของ Flowchart…………………………………………………	125
ภาคผนวก ค. โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูล………………………………………	126
ประวัติย่อผู้วิจัย…………………………………………………………………………	137

บัญชีตาราง

ตาราง	หน้า
1 ลักษณะของข้อมูลจากแผนการทดลองแบบแฟคทอเรียลใน RCB	15
2 ตารางข้อมูลคะแนนเฉลี่ย วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย	17
3 ตารางข้อมูลที่สุ่มตัดข้อมูลสูญหาย จำนวน10%	25
4 ตารางข้อมูลสูญหาย 3 ค่า คือ Y_1,Y_2 และY_3	26
5 จัดตารางข้อมูลใหม่ เพื่อคำนวณวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน.....................	33
6 คำนวณหาค่าเฉลี่ย	34
7 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_1........................	35
8 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_2........................	36
9 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_3........................	37
10 ข้อมูลสูญหายหนึ่งตัวแปร..	39
11 ข้อมูลขาดหายมากกว่าหนึ่งตัวแปร..	39
12 ข้อมูลสูญหายเป็นไปในทิศทางเดียวกัน..	39
13 ข้อมูลสูญหายแบบไม่เป็นระบบ..	40
14 สรุปงานวิจัยที่เกี่ยวข้องทั้งในประเทศ และต่างประเทศ......................................	57
15 จำนวนข้อมูลสูญหายในแต่ละสถานการณ์...	79
16 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 10%	81
17 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 20%	82
18 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 30%	84
19 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	88

บัญชีตาราง

ตาราง	หน้า
20 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	89
21 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	90
22 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	91
23 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	92
24 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	93
25 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	94
26 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	95
27 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก	96

บัญชีภาพประกอบ

ภาพประกอบ **หน้า**

1 กรอบแนวคิดในการวิจัย………………………………………………………………… 11
2 แผนผังแสดงขนาดตัวอย่าง…………………………………………………………... 47
3 ตรวจสอบรูปแบบข้อมูลสูญหาย 10%……………………………………………….. 66
4 ตรวจสอบรูปแบบข้อมูลสูญหาย 20%……………………………………………….. 66
5 ตรวจสอบรูปแบบข้อมูลสูญหาย 30%……………………………………………….. 67
6 ผังแสดงขั้นตอนการเขียนโปรแกรมหลัก (MainProgram)…………………………….. 69
7 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดยการวนซ้ำ……………. 72
8 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายของวิลคินซัล…………….. 73
9 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์ อิมพิวเทชัน ……………………………………………………………………………….. 74
10 ผังแสดงสรุปขั้นตอนการดำเนินงานวิจัย…………………………………………..... 75

บทที่ 1
บทนำ

ภูมิหลัง

การศึกษาวิจัยด้วยแบบแผนการวิจัยเชิงทดลอง เป็นการวางแผนการค้นหาคำตอบจากการวิจัยด้วยวิธีที่มีแบบแผน มีวัตถุประสงค์เพื่อต้องการศึกษาค้นคว้าหาความจริงใหม่ๆ ตามที่นักวิจัยออกแบบไว้ โดยจะทำการหาความสัมพันธ์ระหว่างปัจจัยกับผลลัพธ์ทั้งหมดของการทดลอง ดังนั้นการวางแผนการทดลองได้ถูกนำมาใช้อย่างกว้างขวางในทุกๆ วงการ ทั้งวงการการศึกษา ด้านเกษตรกรรม อุตสาหกรรม และการแพทย์ เป็นต้น แบบแผนการทดลองจึงนับว่ามีความสำคัญต่อการทดลองเป็นอย่างมาก การใช้แบบแผนการทดลองที่เหมาะสมจะช่วยลดขนาดของความคลาดเคลื่อนในการทดลอง (Experimental error) ให้น้อยลง นอกจากนี้จะช่วยให้ผู้ทดลองสามารถศึกษาถึงอิทธิพลของปัจจัยต่างๆ ที่สำคัญๆ รวมทั้งความสัมพันธ์ระหว่างปัจจัยที่ศึกษาได้ในการทดลองเพียงครั้งเดียว หรือการทดลองที่มีการวางแผนอย่างต่อเนื่องและมีเป้าหมายชัดเจน (จิราวัลย์ จิตรถเวช. 2552: 1)

แผนการทดลอง(Experiment Design) ใช้ในกรณีที่ต้องการทราบผลการทดลองและการสรุปผลที่ถูกต้อง เพื่อเป็นทางเลือกในการตัดสินใจ หรือใช้ในการแก้ปัญหาต่างๆ ตลอดจนการคาดคะเนความเสี่ยงและความเป็นไปได้ของเหตุการณ์ที่อาจเกิดขึ้น สำหรับทางการศึกษานั้นได้นำแผนการทดลองมาประยุกต์ใช้กับงานวิจัยทางการศึกษา มีทั้งแผนการทดลองแบบปัจจัยเดียว (Single factor experiment) และการทดลองแบบหลายปัจจัย (Multi factor experiment) ซึ่งแผนการทดลองแบบปัจจัยเดียว (Single factor experiment)เป็นการศึกษาถึงผลตอบสนองของปัจจัยใดปัจจัยหนึ่งโดยเฉพาะ มีการควบคุมปัจจัยอื่นๆ ที่เกี่ยวข้องให้คงที่ แผนการทดลองแบบปัจจัยเดียวมี 3 รูปแบบที่เป็นมาตรฐานคือ 1) แผนการทดลองแบบสุ่มตลอด (CRD) เป็นแผนการทดลองที่มีลักษณะง่ายและสะดวก เหมาะสำหรับหน่วยการทดลองที่มีความสม่ำเสมอกัน 2) แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (RCB) เป็นแผนการทดลองที่มีการจัดกลุ่มของหน่วยทดลอง ให้มีความแตกต่างกันน้อยที่สุด เพื่อลดความแปรปรวนในการทดลอง 3) แผนการทดลองแบบจัตุรัสลาติน (LS) เป็นแผนการทดลองที่ใช้กับหน่วยการทดลองที่มีความแปรปรวนในการทดลอง 2 ทิศทาง และใช้จำนวนทรีตเมนต์เท่ากับจำนวนแถวและจำนวนคอลัมน์ ซึ่งจำนวน ทรีตเมนต์ที่มาก ส่งผลให้การทดลองมีขนาดใหญ่และต้องทำซ้ำมากให้เท่ากัน ตรงกันข้ามกับจำนวน

ทรีตเมนต์น้อย ก็จะมีความน่าเชื่อถือน้อยซึ่งแผนการทดลองแบบจัตุรัสลาตินนี้ไม่นิยมกันอย่างแพร่หลาย สำหรับงานวิจัยทางการศึกษาที่ใช้แผนการทดลองแบบปัจจัยเดียว เช่น สุวัฒน์ นิยมไทย (2531: 37) ได้ศึกษาการศึกษาผลการเรียนรู้วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ซึ่งมีขนาดกลุ่มต่างกัน โดยใช้แบบแผนการทดลองแบบสุ่มในบล็อคสมบูรณ์ (Randomized Complete Block Designs) เลิศลักษณ์ กลิ่นหอม (2532: 6 – 7) ได้ศึกษาการศึกษาความคลาดเคลื่อนชนิดที่ 1 และอำนาจของการทดสอบในแผนการทดลองแบบสุ่มภายในบล็อกที่มีปฏิสัมพันธ์ระหว่างสิ่งทดลองกับบล็อก

ส่วนการทดลองแบบหลายปัจจัย (Multi factor experiment)นั้น เป็นแผนการทดลองที่ศึกษาปัจจัยที่มีผลต่อการตอบสนองของการทดลองมากกว่าหนึ่งปัจจัย ทั้งนี้เพื่อให้ทราบเรื่องราวของสิ่งที่กำลังศึกษาให้กว้างขวางขึ้น แผนการทดลองแบบหลายปัจจัยมีหลายรูปแบบ เช่น การทดลองแบบแฟคทอเรียล (Factorial experiment) เป็นการจัดกลุ่มของทรีตเมนต์ที่ทำการทดลอง ซึ่งแต่ละทรีตเมนต์เกิดจากการรวมตัวกันของปัจจัยที่ต้องการทดสอบ แต่ไม่ได้บอกให้ทราบว่าเป็นแผนการทดลองแบบใด ดังนั้น จึงต้องทำการจัดกลุ่มของทรีตเมนต์ที่ต้องการให้เข้ากับลักษณะของแผนการทดลองมาตรฐานแบบใดแบบหนึ่งใน 3 แบบดังกล่าว แต่ในงานวิจัยแบบแฟคทอเรียลที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) จะเป็นแบบที่นิยมใช้มากที่สุด การทดลองแบบแฟคทอเรียลมีทั้ง สองปัจจัย, สามปัจจัย เป็นต้น ซึ่งจะเป็นการศึกษาปัจจัยที่มีผลต่อการตอบสนองของการทดลองทั้งหมดเท่าๆกัน และต้องการทดสอบด้วยความเที่ยงตรงเท่าๆ กัน นอกจากนี้ยังมีแผนการทดลองแบบสปริท พรอต (Split plot design) เป็นแผนการทดลองแบบหลายปัจจัยที่ใช้ในการทดลองที่มีหน่วยการทดลอง 2 ขนาด สำหรับปัจจัย 2 ชนิด จะแบ่งเป็นหน่วยการทดลองที่มีขนาดใหญ่ (Main plot) และหน่วยการทดลองขนาดเล็ก (Sub plot) รูปแบบการทดลองนี้เหมาะสำหรับใช้ในกรณีที่หน่วยทดลองของปัจจัยหนึ่งต้องใหญ่กว่าอีกปัจจัยหนึ่ง นอกจากนี้ยังมีแผนการทดลองแบบสตริฟ พรอต (Strip plot design) เป็นแผนการทดลองที่ดัดแปลงมาจากแผนการทดลองแบบสปริท พรอต ที่มีหน่วยการทดลองที่มีขนาดใหญ่ (Main plot) จัดในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) และจัดหน่วยการทดลองขนาดเล็ก (Sub plot) เดียวกันของแต่ละหน่วยการทดลองที่มีขนาดใหญ่ (Main plot) ให้ตรงกันเป็นสตริฟ (Strip) มีงานวิจัยทางการศึกษาที่ใช้แผนการทดลองแบบหลายปัจจัย เช่น งานวิจัยของพรวุฒิ คำแก้ว (2546: 67) ได้ศึกษาผลการใช้บทเรียนคอมพิวเตอร์มัลติมีเดีย 3 รูปแบบ ที่มีต่อผลสัมฤทธิ์ทางการเรียน ความคงทนในการเรียนรู้และเจตคติต่อบทเรียนของนักเรียนที่มีระดับความสามารถต่างกัน 3 ระดับโดย ใช้แบบแผนการทดลองแบบแฟคทอเรียล (Factorial Designs) ส่วนงานวิจัยของอรุณี เต๊ะอ้วน (2550:

101–105) ได้ศึกษาเปรียบเทียบผลการเรียนรู้ สาระดนตรี ด้วยบทเรียนคอมพิวเตอร์ มัลติมีเดียสองรูปแบบ คือ แบบแผนการทดลองแบบ 2×2 แฟคทอเรียล กับแบบแผนการทดลองแบบกลุ่มสุ่ม (Randomize block design) ผลการศึกษาพบว่า เมื่อใช้แบบแผนการทดลองแบบแฟคทอเรียล ผลสัมฤทธิ์ทางการเรียนสาระดนตรีของนักเรียนที่เรียนด้วยบทเรียนคอมพิวเตอร์มัลติมีเดียแบบควบคุมบทเรียน และแบบ ไม่ควบคุมบทเรียน มีค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ.01 แต่เมื่อใช้แบบแผนการทดลองแบบกลุ่มสุ่ม จะไม่พบความแตกต่าง อาจเป็นผลจากรูปแบบการทดลองแบบแฟคทอเรียล (สายชล สินสมบูรณ์. 2546: 102) เพราะการทดลองแบบแฟคทอเรียลสามารถศึกษาได้หลายปัจจัยในการทดลองเดียว รวมทั้งสามารถศึกษาปฏิกิริยาสัมพันธ์ระหว่างปัจจัยต่างๆ ได้ เป็นเหตุให้สามารถสรุปผลได้อย่างกว้างขวาง จะเห็นได้ว่าการวางแผนการทดลองแบบแฟคทอเรียลมีประสิทธิภาพมากกว่าการวางแผนการทดลองแบบปัจจัยเดียว เนื่องจากได้ใช้หน่วยการทดลองทุกหน่วยในการวัด ซึ่งงานวิจัยทางการศึกษานำแผนการทดลองแบบแฟคทอเรียลมาใช้ในการเปรียบเทียบวิธีการสอนของนักเรียน

ปัญหาอย่างหนึ่งที่พบในระหว่างการดำเนินการทดลอง คือ การเก็บรวบรวมข้อมูลไม่ได้ครบตามจำนวนที่ต้องการ เนื่องมาจากหลายสาเหตุ เช่น ข้อมูลสูญหาย ข้อมูลเสียหาย ข้อมูลที่ไม่เป็นไปตามข้อกำหนดเบื้องต้นของการใช้วิธีวิเคราะห์ความแปรปรวน ฯลฯ การทดลองบางการทดลองไม่สามารถที่จะหาข้อมูลจริงมาทดแทนหรือทำการทดลองใหม่ได้ โดยเฉพาะการทดลองที่ต้องใช้เวลานาน ใช้ทุนทรัพย์และบุคลากรมาก อาจแก้ปัญหาโดยวิธีตัดข้อมูลที่ไม่สมบูรณ์ออกและเก็บข้อมูลเพิ่มเติม หรืออาจจะใช้วิธีเก็บข้อมูลมากกว่าขนาดตัวอย่างที่ต้องการ การดำเนินงานด้วยวิธีการเหล่านี้ก่อให้เกิดปัญหาทางสถิติที่มีนัย 2 ประการ คือ อำนาจทางสถิติในการทดสอบสมมติฐาน (Statistical power) และความถูกต้องของพารามิเตอร์ที่ต้องการประมาณค่า (Accuracy of estimated parameters) ทั้งนี้เพราะอาจจะมี "ความเอนเอียง (Bias)" เกิดขึ้น (สุนันทา วีรกุลเทวัญ. 2544: 17; อ้างอิงจาก Roth.1994) การอ้างอิงข้อค้นพบที่ได้จากกลุ่มตัวอย่างไปยังกลุ่มประชากรมีความคลาดเคลื่อนสูง สิ่งที่สำคัญคือทำให้สูญเสียรายละเอียดบางอย่างไป ซึ่งอาจจะมีผลต่อผลสรุปของการวิเคราะห์นั้นๆ (วารุณี ตรีบำรุงศักดิ์. 2538: 2) เพื่อหลีกเลี่ยงการสูญเสียรายละเอียดของข้อมูลบางชุดไป ควรมีการประมาณค่าสูญหายเสียก่อน แล้วจึงนำไปวิเคราะห์ข้อมูล ดังนั้น วิธีการประมาณค่าที่สูญหายไปในแผนการทดลองแบบต่างๆ จึงมีความสำคัญมาก

ปัจจัยที่ส่งผลต่อการประมาณค่าสูญหายมีหลายปัจจัยด้วยกัน เช่น จำนวนทรีตเมนต์ที่ใช้ในการทดลองและจำนวนบล็อก ในกรณีที่จำนวนทรีตเมนต์ที่ใช้น้อย ควรเพิ่มจำนวนบล็อกให้มากขึ้น เพื่อทำให้จำนวนองศาความเป็นอิสระของความคลาดเคลื่อนเพิ่มขึ้น ในการประมาณค่าสูญหายนั้น

จำนวนข้อมูลสูญหายก็มีผลต่อการประมาณค่าสูญหาย ถ้าจำนวนข้อมูลสูญหายมีค่ามากจะส่งผลกระทบต่อผลการยอมรับของงานวิจัย ส่วนความน่าเชื่อถือของการทดลองนั้นประเมินได้จากค่าสัมประสิทธิ์ความผันแปร เพราะเป็นค่าที่บ่งชี้ถึงคุณภาพของงานทดลองที่จะตัดสินใจว่างานนั้นได้ผลเป็นที่น่าเชื่อถือมากน้อยเพียงใด ข้อมูลที่นำมาทดลองต้องมีการแจกแจงแบบปกติ กล่าวคือมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 แต่ในการทดลองนี้กำหนดค่าคงที่ h 3 ระดับ คือ 1, 2 และ3 เพื่อศึกษาขนาดความแปรปรวนของความคลาดเคลื่อนของการทดลองให้มีประสิทธิภาพในทางปฏิบัติ และวิธีการประมาณค่าสูญหายมีหลายวิธีจึงต้องเลือกให้เหมาะสมในแต่ละการทดลองจากปัจจัยดังกล่าวซึ่งมีหลายปัจจัยที่ส่งผลต่อการประมาณค่าสูญหาย เพื่อให้ได้สารสนเทศที่ครบถ้วนงานวิจัยนี้จึงต้องทำการจำลองสถานการณ์ต่างๆ ให้ใกล้เคียงกับสถานการณ์จริงมากที่สุด

จากประโยชน์ของการวิจัยแบบแฟคทอเรียลที่ให้ประโยชน์ในการศึกษาปัจจัยที่มีผลต่อการตอบสนองของการทดลองทั้งหมดเท่าๆกัน ทำให้ประหยัดเวลา และประหยัดค่าใช้จ่ายที่จะต้องทำการทดลองทีละปัจจัยหลายๆ ครั้ง และเพื่อแก้ไขปัญหาการวิเคราะห์ข้อมูลที่มีการขาดหายเพราะในสภาพปัจจุบัน ข้อมูลมีการสูญหายในขณะทำการทดลองเป็นอย่างมาก การศึกษาถึงผลกระทบที่เกิดขึ้นและการดำเนินการวิจัยด้วยวิธีที่เหมาะสม จะให้สารสนเทศแก่นักวิจัย นักสถิตินักวัดผล ได้คำนึงถึงการเก็บข้อมูลและวิเคราะห์ข้อมูลด้วยสถิติที่เหมาะสม ดังนั้น ผลจากการวิจัยนี้จะนำมาซึ่งวิธีการวิเคราะห์ข้อมูลที่เหมาะสมกับรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) สำหรับการวิจัยครั้งนี้ ผู้วิจัยสนใจที่ศึกษาการประมาณค่าสูญหาย 3 วิธี คือ 1) วิธีประมาณค่าสูญหายโดยการวนซ้ำ 2) วิธีประมาณค่าสูญหายของวิลคิลซัล และ3) วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน ซึ่งทั้ง 3 วิธีนี้เป็นวิธีที่ง่าย สะดวกในการปฏิบัติและให้ผลที่ใกล้เคียงกับวิธีที่ต้องใช้สถิติขั้นสูง

ความมุ่งหมายของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งความมุ่งหมายไว้ดังนี้

1. เพื่อศึกษาวิธีการประมาณค่าสูญหาย ในแผนการทดลองแบบแฟคทอเรียล 3 วิธี คือ
   - 1.1 วิธีประมาณค่าสูญหายโดยการวนซ้ำ (Iterative Method : IM)
   - 1.2 วิธีประมาณค่าสูญหายของวิลคินซัล (Wilkinson Method : WM)
   - 1.3 วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน (K-Nearest Neighbor Imputation Method : KNN)
2. เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าสูญหาย โดยใช้วิธีประมาณค่าสูญหาย 3 วิธี

ความสำคัญของการวิจัย

จากการศึกษาค้นคว้าในครั้งนี้ทำให้ทราบถึงประสิทธิภาพของวิธีประมาณค่าสูญหาย ทั้ง 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน ในแผนการทดลองแฟคทอเรียล 2 ปัจจัย ซึ่งสามารถศึกษาได้ 2 ปัจจัยในการทดลองเดียวกัน โดยแต่ละปัจจัยจะต้องมีความสำคัญเท่าๆ กัน การทดลองมีประสิทธิภาพสูง เนื่องจากได้ใช้หน่วยทดลองทุกหน่วยในการวัดผลจากการศึกษาในครั้งนี้จะเป็นแนวทางสำหรับผู้ที่สนใจการวางแผนการทดลองแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ เมื่อข้อมูลสูญหาย ให้เหมาะสมสำหรับการทดลองในแต่ละการทดลองเพื่อข้อมูลที่ได้จะมีความน่าเชื่อถือ และแม่นยำมากขึ้น

ขอบเขตของการวิจัย

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลที่ได้จากการจำลองสถานการณ์การทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน (Monte carlo simulation technique) ข้อมูลที่ได้เป็นข้อมูลที่เกิดจากการจำลองด้วยรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย เพื่อให้ได้ข้อมูลใกล้เคียงกับจำนวนจริงที่ใช้ในการปฏิบัติทดลองในการแผนการทดลองมากที่สุด ผู้วิจัยจึงกำหนดให้ประชากรมีขนาดเท่ากับ 512 สำหรับเทคนิคมอนติคาร์โลซิมูเลชัน เป็นการจำลองตัวแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน เขียนด้วยโปรแกรม R

กลุ่มตัวอย่าง

ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ โดยจำแนกตาม ปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ และจำนวน บล็อกเท่ากับ 3,4 และ5 บล็อก เพราะฉะนั้นจำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 รายละเอียดขนาดตัวอย่างดัง ภาพประกอบ 2 สุ่มตัวอย่างให้มีขนาดตามแต่ละ สถานการณ์ การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 1,000 รอบ ซึ่งมีความเพียง พอที่จะศึกษาค่าประมาณพารามิเตอร์

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ

1.1 ปัจจัย 2 ปัจจัย คือ ปัจจัย A และปัจจัย B

- ปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ
- ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ

ดังนั้น จำนวนการทดลองมีเท่ากับ 9 การทดลอง คือ (3×3), (3×4), (3×5), (4×3), (4×4), (4×5), (5×3), (5×4) และ(5×5)

1.2 จำนวนบล็อก เท่ากับ 3, 4 และ 5 ทุกการทดลอง

1.3 จำนวนข้อมูลสูญหาย 10%, 20% และ 30% อย่างสุ่ม

1.4 ค่าสัมประสิทธิ์ความผันแปร [Coefficient of variation : C.V. (%)] ในระดับต่างๆ กัน คือ 5%, 25% และ 45%

1.5 ค่าคงที่ h มี 3 ระดับคือ 1, 2 และ3

1.6 วิธีประมาณค่า 3 วิธี ได้แก่

1.6.1 วิธี Iterative

1.6.2 วิธี Wilkinson

1.6.3 วิธี K-Nearest Neighbor Imputation

2. ตัวแปรตาม คือ ประสิทธิภาพของวิธีประมาณค่าที่สูญหายทั้ง 3 วิธี โดยพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error) ที่ต่ำที่สุด ถือว่าเป็นวิธีที่มีประสิทธิภาพมากที่สุด

นิยามศัพท์เฉพาะ

1. วิธีประมาณค่าสูญหาย หมายถึง วิธีการทางสถิติที่ใช้หลักการประมาณค่าขึ้นมาแทนค่าที่สูญหาย ในการวิจัยครั้งนี้ผู้วิจัยได้กำหนดวิธีการประมาณค่าสูญหาย 3 วิธี ได้แก่

1.1 วิธีประมาณค่าสูญหายโดยการวนซ้ำ เป็นการประมาณค่าขึ้นมาแทนค่าที่ สูญหายโดยใช้สูตรของเยทส์ (Yates's)

1.2 วิธีประมาณค่าสูญหายของวิลคินซัล เป็นการแก้สมการในการประมาณค่าขึ้นมาแทนค่าที่สูญหาย จำนวนสมการทั้งหมดจะเท่ากับจำนวนค่าสูญหายที่ต้องการประมาณ

1.3 วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน เป็นวิธีที่ประมาณค่าข้อมูลสูญหายจากค่าที่ใกล้ที่สุด โดยคำนวณค่าแทนข้อมูลสูญหายจากระยะห่างที่สนใจกับข้อมูลทั้งหมด วิธีการ Euclidian distance แล้วเลือกมาจำนวน k ตัวที่อยู่ใกล้ที่สุด เพื่อนำมาหาค่าเฉลี่ยในการแทนค่าข้อมูลสูญหาย ซึ่งเป็นวิธีที่ง่ายและมีประสิทธิภาพ

2. ประสิทธิภาพของวิธีประมาณค่า หมายถึง ค่าที่ได้จากการเปรียบเทียบระหว่างค่าจริงกับค่าที่ได้จากการประมาณค่าของข้อมูลจำลองทั้ง 3 วิธี ในการจำลองข้อมูล 1 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์จำนวน 1 ค่า การศึกษาในครั้งนี้ทำการจำลองข้อมูล 1,000 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ 1,000 ค่า พิจารณาเอาค่าที่มีค่ามากที่สุดใน 1,000 ค่า เป็นตัวแทนของค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด(Maximum Absolute Error : MAE) ในแต่ละสถานการณ์และแต่ละวิธีมาเปรียบเทียบกัน วิธีใดให้ค่าความคลาดเคลื่อนสูงสุดมีค่าต่ำที่สุด แสดงว่า ค่าประมาณที่ได้มีค่าใกล้เคียงกับค่าจริงที่สูญหายไปมากที่สุด จะเป็นวิธีการประมาณค่าสูญหายที่มีประสิทธิภาพมากที่สุด

3. แผนการทดลอง หมายถึง การวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย คือ ปัจจัย A และปัจจัย B ในรูปแบบ Randomized Complete Block Design (RCB) โดยที่ปัจจัย A มี 3 ลักษณะ คือ 3,4,5 ระดับ และปัจจัย B มี 3 ลักษณะ คือ 3,4,5 ระดับ ดังนั้น จำนวนการทดลองมีเท่ากับ 9 การทดลอง คือ (3×3), (3×4), (3×5), (4×3), (4×4), (4×5), (5×3), (5×4) และ (5×5)

4. หน่วยทดลอง หมายถึง หน่วยหนึ่งหรือกลุ่มหนึ่งของวัตถุทดลองที่ได้รับทรีตเมนต์ต่าง ๆ ที่นำมาศึกษา และนำผลที่ได้มาวิเคราะห์หาข้อสรุปของอิทธิพลของสิ่งทดลองนั้น ๆ

5. บล็อก หมายถึง การที่สิ่งทดลองหนึ่ง ๆ ปรากฏในการทดลองมากกว่า 1 ครั้ง โดยจำนวนครั้งที่สิ่งทดลองนั้น ๆ ปรากฏในการทดลองจะถูกเรียกว่าจำนวนซ้ำ(number of replications) ซึ่งใน

การจัดรูปแบบ Randomized Complete Block Design (RCB) กำหนดให้จำนวนซ้ำคือจำนวนบล็อก เท่ากับ 3, 4 และ 5 บล็อก

6. ค่าสัมประสิทธิ์ความผันแปร [Coefficient of variation : C.V. (%)] หมายถึง ค่าที่บ่งชี้ถึงคุณภาพของงานทดลองที่จะตัดสินใจว่างานนั้นได้ผลเป็นที่น่าเชื่อถือได้มากน้อยเพียงใด ค่าสัมประสิทธิ์ความผันแปรจะเป็นตัวบอกความแปรปรวนของการทดลองในเชิงสัมพัทธ์ที่แสดงค่าเป็นร้อยละของอัตราส่วนระหว่างค่าเบี่ยงเบนมาตรฐานกับค่าเฉลี่ย ถ้าค่า C.V. มาก แสดงว่าการทดลองนั้นมีความแปรปรวนมาก ผลการทดลองจะมีความน่าเชื่อถือน้อย แต่ถ้าค่า C.V. น้อย แสดงว่าการทดลองนั้นมีความแปรปรวนน้อย ผลการทดลองจะมีความน่าเชื่อถือมากกว่า และค่า C.V. สามารถใช้เปรียบเทียบการทดลอง 2 การทดลองที่มีหน่วยแตกต่างกันได้ ค่า C.V. แบ่งออกเป็น 3 ระดับ คือ 5%, 25% และ45%

7. ค่าคงที่ h หมายถึง ค่าที่กำหนดขึ้นเพื่อใช้ในการศึกษาขนาดความแปรปรวนของความคลาดเคลื่อนของการทดลอง ในการทดลองทุกการทดลองจะต้องมีความคลาดเคลื่อนการทดลองเกิดขึ้น ความคลาดเคลื่อนชนิดนี้ไม่สามารถควบคุมแหล่งที่มาได้ เมื่อควบคุมแหล่งที่มาไม่ได้ก็ควบคุมความคลาดเคลื่อนไม่ได้ ความคลาดเคลื่อนชนิดนี้จึงเป็นความคลาดเคลื่อนสุ่ม (Random error) ซึ่งความผันแปรของความคลาดเคลื่อนในการทดลองนั้น เรียกว่า ความแปรปรวนของความคลาดเคลื่อนในการทดลอง (Experimental error variance) มีลักษณะเป็นผลที่เกิดขึ้นโดยไม่ได้ตั้งใจ (Random effect) หรือเกิดขึ้นแบบสุ่มๆ มากบ้างน้อยบ้าง ในงานวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาค่าการกระจายของความแปรปรวนของความคลาดเคลื่อน จึงกำหนดให้ h มีค่า เท่ากับ 1,2 และ 3

กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับวิธีการประมาณค่าสูญหายในการวางแผนการทดลองผู้วิจัยทำการศึกษาจำนวนทรีตเมนต์ที่ใช้ในการทดลอง การทดลองที่มีจำนวนทรีตเมนต์มากจะใช้จำนวนบล็อกน้อยกว่าการทดลองที่มีจำนวนทรีตเมนต์น้อย ถ้าจำนวนทรีตเมนต์ที่ใช้มีมาก อาจใช้จำนวนบล็อกเพียง 2 – 3 บล็อกก็เพียงพอแล้ว แต่ถ้าจำนวนทรีตเมนต์ที่ใช้น้อย ควรเพิ่มจำนวนบล็อกให้มากขึ้น เพื่อทำให้จำนวนองศาความเป็นอิสระของความคลาดเคลื่อนเพิ่มขึ้น เนื่องจากถ้าจำนวนองศาความเป็นอิสระของความคลาดเคลื่อนมีค่าสูง จะทำให้ค่า F ที่คำนวณได้มีค่าเพิ่มขึ้น ซึ่งทำให้ประสิทธิภาพในการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของทรีตเมนต์เพิ่มขึ้น (สายชล สินสมบูรณ์ทอง. 2549: 14) ในการประมาณค่าสูญหายนั้นจำนวนข้อมูลสูญหายก็มีผลต่อการประมาณค่าสูญหาย ถ้าข้อมูลสูญหายจำนวนน้อยประมาณ 5% การตัดหน่วยตัวอย่างออกไปดูเหมือนว่าจะมีเหตุผลในการแก้ปัญหาข้อมูลสูญหาย แต่ถ้ามีการสูญหายมากการตัดข้อมูลออกจะไม่มีประสิทธิภาพข้อมูลที่เหลืออยู่จะไม่เป็นตัวแทนของประชากรซึ่งมีเป้าหมายในการอ้างอิง (Schafer.1997: 1; Little;& Rubin.1987 : 5) อีกทั้ง รูท (Roth. 1994: 551) ได้กล่าวว่า การเลือกวิธีการจัดการข้อมูลสูญหายจะมีความสำคัญเมื่อจำนวนข้อมูลสูญหายอยู่ระหว่าง 15 – 20% และจะมีความสำคัญมากที่สุดเมื่อจำนวนข้อมูลสูญหายระหว่าง 30 – 40% ที่ระดับนี้การเลือกใช้วิธีการจัดการข้อมูลสูญหายจะให้ผลลัพธ์ที่แตกต่างกัน จะเห็นได้ว่าจำนวนข้อมูลสูญหายมีค่าอยู่ระหว่าง 5 – 40% จากงานวิจัยและเอกสารดังกล่าวข้อมูลสูญหายจะอยู่ระหว่าง 5% ถึง 20% อยู่ในช่วงที่ผลการวิจัยที่ผ่านมาพบว่า เป็นช่วงที่ยอมรับได้ ผู้วิจัยจึงกำหนดจำนวนข้อมูลสูญหายให้มีจำนวนมากกว่า 20% โดยกำหนดจำนวนข้อมูลสูญหายสูงสุดเท่ากับ 30% เพื่อให้ได้ข้อค้นพบที่เป็นประโยชน์มากยิ่งขึ้นในเรื่องของจำนวนข้อมูลสูญหาย ส่วนความน่าเชื่อถือของการทดลองนั้นประเมินได้จากค่าสัมประสิทธิ์ความผันแปร [Coefficient of variation : C.V. (%)] เพราะเป็นค่าที่บ่งชี้ถึงคุณภาพของงานทดลองที่จะตัดสินใจว่างานนั้นได้ผลเป็นที่น่าเชื่อถือมากน้อยเพียงใด ข้อมูลที่นำมาทดลองต้องมีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 แต่ในการทดลองนี้กำหนดค่าคงที่ h 3 ระดับ คือ 1, 2 และ3 เพื่อศึกษาขนาดความแปรปรวนของความคลาดเคลื่อนของการทดลองให้มีประสิทธิภาพในทางปฏิบัติ และวิธีการประมาณค่าสูญหาย มีหลายวิธีจึงต้องเลือกให้เหมาะสมในแต่ละการทดลอง

จะเห็นว่าการศึกษาวิธีประมาณค่าสูญหายมีตัวแปรอื่นๆ เข้ามาเกี่ยวข้องหลายตัวแปร เพื่อให้การทดลองมีความน่าเชื่อถือ และเพื่อเป็นแนวทางให้กับผู้ทดลองในการเลือกใช้วิธีการประมาณค่าที่เหมาะสม ผู้วิจัยจึงกำหนดตัวแปรอิสระ 6 ตัวแปร ได้แก่

1. จำนวนปัจจัย
2. จำนวนบล็อก
3. จำนวนข้อมูลสูญหาย
4. ค่าสัมประสิทธิ์ความผันแปร
5. ค่าคงที่ h
6. วิธีการประมาณค่าสูญหาย

รายละเอียดตัวแปรที่ใช้ในการวิจัยสรุปได้ดังภาพประกอบ 1

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้กำหนดกรอบแนวคิดที่ใช้ในการวิจัย ดังภาพประกอบ 1

จำนวนปัจจัย 2 ปัจจัย คือ
- ปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5
- ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5

จำนวนบล็อก
3 บล็อก, 4 บล็อก และ 5 บล็อก

จำนวนข้อมูลสูญหาย
10%, 20% และ 30% อย่างสุ่ม

ค่าสัมประสิทธิ์ความผันแปร
5%, 25% และ 45%

ค่าคงที่ h
1, 2 และ 3

วิธีประมาณค่า 3 วิธี ได้แก่
- วิธี Iterative
- วิธี Wilkinson
- วิธี K-Nearest Neighbor Imputation

→

ประสิทธิภาพของวิธีประมาณค่าสูญหายทั้ง 3 วิธี

ภาพประกอบ 1 กรอบแนวคิดที่ใช้ในการวิจัย

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับการประมาณค่าสูญหาย และงานวิจัยที่เกี่ยวข้องต่างๆ ที่เป็นความรู้พื้นฐาน โดยผู้วิจัยขอนำเสนอเป็น 5 ตอน ดังนี้

ตอนที่ 1 ทฤษฎีที่เกี่ยวข้อง

1.1 ตัวแบบที่ศึกษา

1.2 ความแปรปรวนของความคลาดเคลื่อนของการทดลอง

1.3 วิธีประมาณค่าสูญหาย 3 วิธี

1.4 ตัวอย่างการคำนวณเมื่อมีข้อมูลสูญหาย

ตอนที่ 2 ลักษณะของข้อมูลสูญหาย

ตอนที่ 3 การจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล

ตอนที่ 4 ประชากรและการสุ่มตัวอย่าง

ตอนที่ 5 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีประมาณค่าสูญหาย

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ทฤษฎีที่เกี่ยวข้อง

เนื้อหาในส่วนนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับวิธีการประมาณค่าสูญหายในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ มีรายละเอียดดังนี้

1.1 ตัวแบบที่ศึกษา

1.2 ความแปรปรวนของความคลาดเคลื่อนของการทดลอง

1.3 วิธีประมาณค่าสูญหาย 3 วิธี คือ

1.3.1 วิธีประมาณค่าสูญหายโดยวิธีวนซ้ำ

1.3.2 วิธีประมาณค่าสูญหายของวิลคินซัล

1.3.3 วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน

1.4 ตัวอย่างการคำนวณเมื่อมีข้อมูลสูญหาย
1.4.1 วิธีประมาณค่าสูญหายโดยวิธีวนซ้ำ
1.4.2 วิธีประมาณค่าสูญหายของวิธีวิลคินซัล
1.4.3 วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน

1.1 ตัวแบบที่ศึกษา

ในงานวิจัยครั้งนี้ได้ทำการศึกษาวิธีประมาณค่าสูญหายทั้งหมด 3 วิธี เพื่อเปรียบเทียบหาวิธีการประมาณค่าสูญหายที่เหมาะสมที่สุดสำหรับแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ซึ่งในการวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัยนั้น จะต้องมีการจัดรูปแบบของทรีตเมนต์ตามรูปแบบมาตรฐาน 3 รูปแบบ คือ แผนการทดลองแบบสุ่มสมบูรณ์ (Completely Randomize Design : CRD), แบบสุ่มในบล็อกสมบูรณ์ (Randomize Complete Block Design : RCB) และแบบจัตุรัสลาติน (Latin Square Design : LS) แต่ละรูปแบบเหมาะสำหรับงานทดลองที่แตกต่างกัน ผู้วิจัยต้องการที่จะศึกษาสำหรับหน่วยทดลองที่ไม่มีความสม่ำเสมอกันจึงเลือกแบบสุ่มในบล็อกสมบูรณ์ (Randomize Complete Block Design : RCB) เพราะการจัดหน่วยทดลองออกเป็นบล็อก จะทำกลุ่มของหน่วยทดลองภายในกลุ่มมีความแตกต่างกันน้อยที่สุด แต่ภายนอกกลุ่มมีความแตกต่างกันมากที่สุด ทำให้ได้ผลที่เที่ยงตรง สูงกว่าแบบสุ่มสมบูรณ์ ส่วนแบบจัตุรัสลาตินนั้นจำนวนทรีตเมนต์ที่จะใช้ต้องมีจำนวนแถวและคอลัมน์เท่ากัน ถ้าไม่เท่ากันก็จะใช้รูปแบบนี้ไม่ได้ รูปแบบนี้จึงยังไม่กว้างขวางนัก สำหรับตัวแบบของแบบแผนการทดลองแบบแฟคทอเรียลมี 3 รูปแบบ คือ

1. ตัวแบบอิทธิพลกำหนด (fixed effect model)
2. ตัวแบบอิทธิพลสุ่ม (random effect model)
3. ตัวแบบอิทธิพลผสม (mixed effect model)

ในงานวิจัยครั้งนี้ผู้วิจัยเลือกศึกษาตัวแบบอิทธิพลกำหนด มี 2 ลักษณะด้วยกันคือ ปัจจัย A (Factor A) และปัจจัย B (Factor B) เป็นตัวแปรเจาะจง ซึ่งตัวแบบอิทธิพลกำหนดมีรูปแบบเชิงเส้นตรง ดังนี้

รูปแบบเชิงเส้นตรงของแผนการทดลองแบบ a×b แฟคทอเรียลใน RCB (พิสมัย หาญมงคลพิพัฒน์. 2546: 196)

$$Y_{ijk} = \mu + B_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad ; \quad \begin{aligned} i &= 1,2,\ldots,a \\ j &= 1,2,\ldots,b \\ k &= 1,2,\ldots,r \end{aligned}$$

โดยที่

Y_{ijk} แทนค่าของข้อมูลจากหน่วยทดลองของปัจจัย A ระดับที่ i ของปัจจัย B ระดับที่ j และบล็อกที่ k

μ แทนค่าเฉลี่ยของประชากร

B_k แทนอิทธิพลของบล็อก k

α_i แทนอิทธิพลของปัจจัย A ระดับที่ i

β_j แทนอิทธิพลของปัจจัย B ระดับที่ j

$(\alpha\beta)_{ij}$ แทนอิทธิพลร่วมกันของของปัจจัย A ระดับที่ i และของปัจจัย B ระดับที่ j

ε_{ijk} แทนความคลาดเคลื่อนของการทดลอง

ข้อกำหนดเบื้องต้นของแผนการทดลอง

1. $\sum \alpha_i = 0$
2. $\sum \beta_j = 0$
3. $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$
4. $\varepsilon_{ijk} \sim N(0, \sigma^2)$

ลักษณะของข้อมูลจากแผนการทดลองแบบแฟคทอเรียลใน RCB

แบบแผนการทดลองแบบแฟคทอเรียลที่มี 2 ปัจจัย คือ ปัจจัย A ซึ่งมี i ระดับ และปัจจัย B มี j ระดับ และบล็อกที่ k แสดงดังตาราง 1

ตาราง 1 ลักษณะของข้อมูลจากแผนการทดลองแบบแฟคทอเรียลใน RCB

ปัจจัย A	ปัจจัย B	Block 1	Block 2	Block …	Block k	รวม
a_1	b_1	Y_{111}	Y_{112}	…	Y_{11k}	$Y_{11.}$
	b_2	Y_{121}	Y_{122}	…	Y_{12k}	$Y_{12.}$
	⋮	⋮	⋮	…	⋮	⋮
	b_j	Y_{1j1}	Y_{1j2}	…	Y_{1jk}	$Y_{1j.}$
a_2	b_1	Y_{211}	Y_{212}	…	Y_{21k}	$Y_{21.}$
	b_2	Y_{221}	Y_{222}	…	Y_{22k}	$Y_{22.}$
	⋮	⋮	⋮	…	⋮	⋮
	b_j	Y_{2j1}	Y_{2j2}	…	Y_{2jk}	$Y_{2j.}$
⋮	⋮	⋮	⋮	…	⋮	⋮
a_i	b_1	Y_{i11}	Y_{i12}	…	Y_{i1k}	$Y_{i1.}$
	b_2	Y_{i21}	Y_{i22}	…	Y_{i2k}	$Y_{i2.}$
	⋮	⋮	⋮	…	⋮	⋮
	b_j	Y_{ij1}	Y_{ij2}	…	Y_{ijk}	$Y_{ij.}$
รวม		$Y_{..1}$	$Y_{..2}$	…	$Y_{..k}$	$Y_{...}$

โดยที่

Y_{ijk} แทนค่าของข้อมูลจากหน่วยทดลองของปัจจัย A ระดับที่ i ของปัจจัย B ระดับที่ j และบล็อกที่ k

$Y_{ij.}$ แทนผลรวมของข้อมูลทั้งหมดของปัจจัย A ระดับที่ i และปัจจัย B ระดับที่ j

$Y_{i..}$ แทนผลรวมของข้อมูลทั้งหมดของปัจจัย A ระดับที่ i

$Y_{.j.}$ แทนผลรวมของข้อมูลทั้งหมดของปัจจัย B ระดับที่ j

$Y_{...}$ แทนผลรวมของข้อมูลทั้งหมด

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} Y_{ijk} = \sum_{i=1}^{a} \sum_{j=1}^{b} Y_{ij.}$$

$$= \sum_{i=1}^{a} Y_{i..} = \sum_{j=1}^{b} Y_{.j.}$$

ตัวอย่างข้อมูลจากแผนการทดลองแบบแฟคทอเรียลใน RCB

ตัวอย่าง ผู้วิจัยขอจำลองข้อมูล โดยดัดแปลงข้อมูลมาจากสุวัฒน์ นิยมไทย (2531 : 41) ศึกษาผลการเรียนรู้วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ซึ่งมีขนาดของกลุ่มแตกต่างกัน โดยการทดลองแฟคทอเรียล 2 ปัจจัย ที่วางแผนแบบ RCB จำนวน 3 ห้องเรียน

ความมุ่งหมายของการวิจัย

เพื่อศึกษาเปรียบเทียบผลการเรียนรู้ของนักเรียน ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ที่มีขนาดของกลุ่มต่างกัน 5 ขนาด คือ ขนาดกลุ่มละ 2 คน, 3 คน, 4 คน, 5 คน และ 6 คน

ปัจจัยที่ทดสอบ - ปัจจัย A คือ เพศ

A_1 ชาย

A_2 หญิง

- ปัจจัย B คือ ขนาดของกลุ่มย่อย มี 5 ขนาด คือ

B_1 ขนาดกลุ่มละ 2 คน

B_2 ขนาดกลุ่มละ 3 คน

B_3 ขนาดกลุ่มละ 4 คน

B_4 ขนาดกลุ่มละ 5 คน

B_5 ขนาดกลุ่มละ 6 คน

ตาราง 2 ตารางข้อมูลคะแนนเฉลี่ย วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจาก คอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย

ปัจจัย A	ปัจจัย B	คะแนนของแต่ละห้องเรียน			รวม
		1	2	3	
ชาย	กลุ่มละ 2 คน	31	33	29	93
	กลุ่มละ 3 คน	36	32	36	104
	กลุ่มละ 4 คน	38	27	26	91
	กลุ่มละ 5 คน	53	43	49	145
	กลุ่มละ 6 คน	56	63	54	173
หญิง	กลุ่มละ 2 คน	43	35	41	119
	กลุ่มละ 3 คน	46	43	48	137
	กลุ่มละ 4 คน	47	47	41	135
	กลุ่มละ 5 คน	61	58	53	172
	กลุ่มละ 6 คน	64	46	50	160
รวม		475	427	427	1329

1.2 ความแปรปรวนของความคลาดเคลื่อนของการทดลอง

ความคลาดเคลื่อนของการทดลอง (Experimental error) หมายถึง ค่าการวัดผลการทดลองที่มีความผันแปรจากแหล่งต่างๆ ที่ไม่สามารถควบคุมได้ ซึ่งส่วนประกอบความผันแปรดังกล่าวมีผลกระทบต่อค่าการวัดผลการทดลองที่แท้จริง โดยปกติแล้วการควบคุมความคลาดเคลื่อนจะควบคุมจากแหล่งที่มาของความคลาดเคลื่อนโดยการทำให้เป็นระบบแต่ความคลาด เคลื่อนชนิดนี้ไม่สามารถควบคุมแหล่งที่มาได้ เมื่อควบคุมแหล่งที่มาไม่ได้ก็ควบคุมความคลาดเคลื่อนไม่ได้ ความคลาดเคลื่อนชนิดนี้จึงเป็นความคลาดเคลื่อนสุ่ม (Random error) ซึ่งความผันแปรของความคลาดเคลื่อนในการทดลองนั้น เรียกว่า ความแปรปรวนของความคลาดเคลื่อนในการทดลอง (Experimental error variance) มีลักษณะเป็นผลที่เกิดขึ้นโดยไม่ได้ตั้งใจ (Random effect) หรือเกิดขึ้นแบบสุ่มๆ มากบ้างน้อยบ้าง และเมื่อวัดแล้วหาค่าเฉลี่ย(μ) ทุกหน่วยทดลองจะได้ค่าเป็น 0 และค่าการกระจายมีค่าเป็นความแปรปรวนของความคลาดเคลื่อน (σ_ε^2) ความแปรปรวนชนิดนี้เป็นกลุ่มของความแปรปรวนของตัวแปรตามในการทดลองที่มีสาเหตุมาจากแหล่งที่ไม่ต้องการนั่นเอง (ชัยลิขิต สร้อยเพชรเกษม: online)

จากตัวแบบที่ใช้ในการศึกษาในครั้งนี้ อยู่ในรูปแบบเชิงเส้นตรงของแผนการทดลองแบบ a×b แฟคทอเรียลใน RCB

$$Y_{ijk} = \mu + B_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

กำหนดให้ μ เป็นพารามิเตอร์ที่ไม่ทราบค่า, B_k เป็นอิทธิพลของบล็อก k, α_i เป็นอิทธิพลของปัจจัย A ระดับที่ i, β_j เป็นอิทธิพลของปัจจัย B ระดับที่ j, $(\alpha\beta)_{ij}$ เป็นอิทธิพลร่วมกันของของปัจจัย A ระดับที่ i และของปัจจัย B ระดับที่ j

ε_{ijk} เป็นอิทธิพลของความคลาดเคลื่อนจากการทดลองซึ่งเป็นความคลาดเคลื่อนของหน่วยทดลองของปัจจัย A ระดับที่ i, ของปัจจัย B ระดับที่ j และของบล็อกที่ k มีความเป็นอิสระจาก ε_{ijk} อื่นๆ มีการกระจายตัวเป็นปกติและมีค่าเฉลี่ยเป็น 0 การกระจายเป็นความแปรปรวนเท่ากับ σ_ε^2

จากการจำลองข้อมูล กำหนดให้ $\sigma_B^2 = \sigma_\alpha^2 = \sigma_\beta^2 = \sigma_{\alpha\beta}^2 = h\sigma_\varepsilon^2$ เมื่อ σ_ε^2 คือความแปรปรวนของความคลาดเคลื่อนของการทดลอง ในงานวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาค่าการกระจายของความแปรปรวนของความคลาดเคลื่อน จึงกำหนดให้ h มีค่า เท่ากับ 1,2 และ 3

1.3 วิธีประมาณค่าสูญหาย มี 3 วิธี ดังนี้

1.3.1 วิธีประมาณค่าสูญหายโดยการวนซ้ำ

ใช้หลักของการประมาณค่าขึ้นมาแทนค่าที่สูญหาย หลังจากนั้นจึงทำการวิเคราะห์ผลด้วยเทคนิคตามแผนการทดลองที่วางไว้ตามปกติ การประมาณค่าสูญหาย 1 ค่า จะใช้สูตรของเยทส์ (Yates.1937) คือ

$$Y = \frac{bB + kT - G}{(b-1)(k-1)} \qquad \text{.........................(1)}$$

เมื่อ Y แทน ค่าที่สูญหาย

b แทน จำนวนบล็อก

k แทน จำนวนทรีตเมนต์

B แทน ผลรวมของค่าข้อมูลในบล็อกที่มีค่าสูญหาย

T แทน ผลรวมของค่าข้อมูลในทรีตเมนต์ที่มีค่าสูญหาย

G แทน ผลรวมของค่าข้อมูลทั้งหมด

ที่มาของสูตรเยทสเป็นดังนี้

จาก $SS_{error} = SS_{tatal} - SS_{block} - SS_{treatment}$

$$= \sum_i \sum_j (y_{ij})^2 + Y^2 - \frac{(G+Y)^2}{bk} - \frac{1}{k}\sum_i [(y_{i.})^2] - \frac{1}{k}(B+Y)^2$$

$$+ \frac{(G+Y)^2}{bk} - \frac{1}{b}\sum_j [(y_{.j})^2] - \frac{1}{b}(T+Y)^2 + \frac{(G+Y)^2}{bk}$$

ต้องการหาค่า Y เพื่อทำให้ SS_{error} มีค่าต่ำสุด จะได้

$$\frac{\partial SS_{error}}{\partial Y} = 2Y - \frac{2}{k}(B+Y) - \frac{2}{b}(T+Y) + \frac{2}{bk}(G+Y) = 0$$

$$bkY - bY - kY + Y = bB + kT - G$$

$$(b-1)(k-1)Y = bB + kT - G$$

$$Y = \frac{bB + kT - G}{(b-1)(k-1)}$$

กรณีมีค่าสูญหาย 2 ค่าขึ้นไป ใช้หลักการวนซ้ำ โดยคำนวณค่าที่สูญหายไปทุกค่าจากสูตร

$$\overline{Y}_i = \frac{(\overline{M}_t + \overline{M}_b)}{2} \qquad \text{........................(2)}$$

โดยเว้นไว้เพียงค่าเดียว เมื่อ

$\overline{Y}_i$ แทน ค่าเฉลี่ยของค่าสูญหาย เมื่อ i =1,2,…,m

$\overline{M}_t$ แทน ค่าเฉลี่ยของค่าข้อมูลทั้งหมดของทรีตเมนต์ 2 ปัจจัย

$\overline{M}_b$ แทน ค่าเฉลี่ยของค่าข้อมูลทั้งหมดของบล็อก

ขั้นตอนการประมาณค่าสูญหาย 2 ค่า

1. ประมาณตัวเลขมาทดแทนค่าข้อมูลที่สูญหายเป็นการชั่วคราว โดยให้เหลือค่าข้อมูลที่สูญหายไว้เพียงค่าเดียว แทนด้วยค่าที่ได้จากสูตร (2)
2. แทนค่า $\overline{Y}_1$ ที่คำนวณได้ ในตารางข้อมูล แล้วคำนวณหาค่าสูญหายตัวที่สอง Y_2 จากสูตร (1)
3. ทำซ้ำหาค่า Y_1 ใหม่โดยใช้สูตร (1) ทำเช่นนี้จนค่าประมาณที่ได้ไม่เปลี่ยนแปลง นำค่าประมาณที่ได้แทนในตารางข้อมูล
4. คำนวณหาค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error : MAE)

1.3.2 วิธีประมาณค่าสูญหายของวิลคินซัล

วิธีการประมาณค่าสูญหายของวิลคินซัลได้เริ่มต้นโดยเยทส์ (Yates.1937) ข้อมูลจะถูกทำให้สมบูรณ์โดยการใส่ค่าประมาณของค่าที่หายไป ที่พิจารณาจากค่า Residual sum of square ที่น้อยที่สุด วิลคินซัลเสนอวิธีการหาค่าประมาณกำลังสองต่ำสุด โดยพิจารณาจากสมการเชิงซ้อนของค่าที่สูญหาย วิธีการอื่นที่ไม่ใช่การวนซ้ำ สิ่งที่สำคัญสำหรับการประมาณค่าสูญหายของวิลคินซัล คือ การตั้งสมการสำหรับหาค่าที่หายไป จะต้องเป็นรูปแบบที่พิจารณาได้ง่าย และการแก้ไขปัญหาหรือหาคำตอบจากสมการด้วยการใช้ แมทริกซ์อินเวอร์ชั่นที่ง่ายที่สุด (matrix inversion)

ขั้นตอนของการประมาณค่าสูญหาย

จากหลักการของกำลังสองน้อยที่สุดการประเมินของพารามิเตอร์ τ ของข้อมูลที่สูญหายต้องทำให้ลดค่า error sum of square ลงให้มากที่สุด

$$S_1(\tau) = \left[y - \varepsilon_y(\tau)\right]'\left[y - \varepsilon_y(\tau)\right]$$

$$S(\tau,u) = \left[y - \varepsilon_y(\tau)\right]'\left[y - \varepsilon_y(\tau)\right] + \left[u - \varepsilon_u(\tau)\right]'\left[u - \varepsilon_u(\tau)\right]$$

$$= \quad S_1(\tau) \quad + \quad S_2(\tau,u)$$

$$u = \varepsilon_u(\tau) \qquad \text{................(1)}$$

$$\tau = t(y,u) \qquad \text{................(2)}$$

$$S(t,u) = \left[y - E_y(y,u)\right]'\left[y - E_y(y,u)\right] + \left[u - E_u(y,u)\right]'\left[u - E_u(y,u)\right] \qquad \text{................(3)}$$

สมการของการประมาณค่าสูญหาย

$$u = E_u(y,u) \qquad \text{................(4)}$$

ค่าคาดหวังในสมการเชิงเส้นของการทดลองที่สมบูรณ์ ($E(z)$)

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ u \end{bmatrix}$$

$$u = E_u(y,0) + E_u(0,u) \qquad \text{................(5)}$$

$$\begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ u \\ v \\ w \end{bmatrix} = u\begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + v\begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + w\begin{bmatrix} 0 \\ 0 \\ . \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= u\delta_u + v\delta_v + w\delta_w$$

เพราะฉะนั้น เขียนเป็นสมการประมาณค่าสูญหายได้ดังนี้

$$u = E_u(y,0) + uE_u(\delta_u) + vE_u(\delta_v) + wE_u(\delta_w)$$

$$v = E_v(y,0) + uE_v(\delta_u) + vE_v(\delta_v) + wE_v(\delta_w) \qquad \text{................(6)}$$

$$w = E_w(y,0) + uE_w(\delta_u) + vE_w(\delta_v) + wE_w(\delta_w)$$

หาคำตอบโดยใช้แมทริกซ์ กำหนดให้ A คือ แมทริกซ์สัมประสิทธิ์ของสมการ B คือ แมทริกซ์คำตอบของสมการ Y คือ แมทริกซ์ค่าที่สูญหาย

$$Au = E_u^0$$

$$u = A^{-1}E_u^o$$

โดยที่

$$A^{-1} = \frac{1}{|A|}\left[A_{ij}\right]$$

เพราะฉะนั้น

$$Y = A^{-1}B$$

จำนวนสมการทั้งหมดจะเท่ากับจำนวนค่าสูญหายที่ต้องการประมาณ ให้ M_i แทนค่าสูญหาย เมื่อ i = 1, 2, …,m

สมการที่ได้ คือ $nY_i = bB_{Y_i} + kT_{Y_i} - G$ (Wilkinson.1958: 266)

โดยที่ n แทน จำนวนข้อมูลทั้งหมด $(n = bk)$

b แทน จำนวนบล็อก

k แทน จำนวนทรีตเมนต์ของ 2 ปัจจัย

B_{Y_i} แทน ผลรวมของค่าข้อมูลในบล็อกที่มี Y_i สูญหาย

T_{Y_i} แทน ผลรวมของค่าข้อมูลในทรีตเมนต์ที่มี Y_i สูญหาย

G แทน ผลรวมของค่าข้อมูลทั้งหมดที่มี Y_i สูญหาย

1.3.3 วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน (K – Nearest Neighbor Imputation Method)

วิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน เป็นวิธีที่ได้รับความนิยมอย่างมาก (Troyanskaya; el al. 2001: 520 – 525) เนื่องจากเป็นวิธีที่ง่ายและมีประสิทธิภาพอีกวิธีหนึ่งที่นำมาใช้ประมาณค่าข้อมูลสูญหาย โดยข้อมูลที่จะนำมาประมาณค่ากับข้อมูลที่สูญหายจะต้องมีความสัมพันธ์กัน เพื่อนำมาสร้างเป็นแบบจำลอง ผู้วิจัยต้องกำหนดค่า k เพื่อใช้ในการพิจารณาข้อมูลที่อยู่ใกล้ที่สุด ซึ่งจะต้องเป็นจำนวนเต็มบวก เช่น k = 3 คือพิจารณาเฉพาะข้อมูล 3 แรกที่มีค่า Euclidian Distance น้อยที่สุด ซึ่งมีขั้นตอนดังนี้

1. กำหนดค่า k
2. คำนวณหาระยะห่างระหว่างจุดด้วยวิธี Euclidian Distance ระหว่างข้อมูลที่เกิดค่าสูญหายที่ต้องการพิจารณากับข้อมูลที่มีความสมบูรณ์ ดังสมการ

$$dist(R_i, R_j) = \sqrt{\sum_{k=1}^{n} \left(y_{i,k} - y_{j,k}\right)^2}$$ (ณรงค์ โพธิ. 2554: 13)

โดยที่

$dist(R_i, R_j)$ แทน ระยะห่างระหว่างข้อมูลแถวที่ i และข้อมูลแถวที่ j

n แทน จำนวนข้อมูลทั้งหมด

$y_{i,k}$ แทน ค่าข้อมูลที่เกิดการสูญหาย แถวที่ i คอลัมน์ k

$y_{j,k}$ แทน ค่าข้อมูลที่มีความสมบูรณ์ แถวที่ j คอลัมน์ k

3. เรียงลำดับระยะห่างระหว่างจุดโดยพิจารณาจากข้อมูลที่ใกล้ที่สุดตามจำนวน k
4. ประมาณค่าข้อมูลสูญหายจากค่าเฉลี่ยของข้อมูลที่อยู่ใกล้ที่สุด ดังสมการ

$$\hat{y}_i = \frac{\sum_{i=1}^{k} y_i}{k}$$ (ณรงค์ โพธิ. 2554: 14)

โดยที่

$\hat{y}_i$ แทน ค่าข้อมูลสูญหายที่ได้จากการประมาณค่าใหม่

k แทน จำนวนที่กำหนดไว้เพื่อพิจารณาค่าที่อยู่ใกล้ที่สุด

y_i แทน ค่าของข้อมูลที่มีความสมบูรณ์ที่ตรงกับข้อมูลสูญหาย

หลังจากนั้น แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลที่อยู่ใกล้ที่สุด

1.4 ตัวอย่างการคำนวณเมื่อมีข้อมูลสูญหาย

ตัวอย่าง ผู้วิจัยขอจำลองข้อมูล โดยดัดแปลงข้อมูลมาจากสุวัฒน์ นิยมไทย (2531: 41) ศึกษาผลการเรียนรู้วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ซึ่งมีขนาดของกลุ่มแตกต่างกัน โดยการทดลองแฟคทอเรียล 2 ปัจจัย ที่วางแผนแบบ RCB จำนวน 3 ห้องเรียน

ความมุ่งหมายของการวิจัย

เพื่อศึกษาเปรียบเทียบผลการเรียนรู้ของนักเรียน ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ที่มีขนาดของกลุ่มต่างกัน 5 ขนาด คือ ขนาดกลุ่มละ 2 คน, 3 คน, 4 คน, 5 คน และ 6 คน

ปัจจัยที่ทดสอบ - ปัจจัย A คือ เพศ

A_1 ชาย

A_2 หญิง

- ปัจจัย B คือ ขนาดของกลุ่มย่อย มี 5 ขนาด คือ

B_1 ขนาดกลุ่มละ 2 คน

B_2 ขนาดกลุ่มละ 3 คน

B_3 ขนาดกลุ่มละ 4 คน

B_4 ขนาดกลุ่มละ 5 คน

B_5 ขนาดกลุ่มละ 6 คน

ข้อมูลคะแนนเฉลี่ยวิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ ช่วยสอนในลักษณะกลุ่มย่อย แสดงดังตาราง 2

ทำการสุ่มตัดข้อมูลสูญหาย โดยกำหนดจำนวนข้อมูลสูญหาย 10% เพราะฉะนั้นจากข้อมูลตัวอย่างจะมีข้อมูลสูญหายทั้งหมด 3 ตำแหน่ง ดังตาราง 3

ตาราง 3 ตารางข้อมูลที่สุ่มตัดข้อมูลสูญหาย จำนวน 10%

ปัจจัย A	**ปัจจัย B**	**Block**			**รวม**
		1	**2**	**3**	
ชาย	กลุ่มละ 2 คน	31	33	29	**93**
	กลุ่มละ 3 คน	36	32	36	**104**
	กลุ่มละ 4 คน	38	missing	26	**64**
	กลุ่มละ 5 คน	53	43	49	**145**
	กลุ่มละ 6 คน	56	63	missing	**119**
หญิง	กลุ่มละ 2 คน	43	35	41	**119**
	กลุ่มละ 3 คน	46	43	48	**137**
	กลุ่มละ 4 คน	47	47	41	**135**
	กลุ่มละ 5 คน	missing	58	53	**111**
	กลุ่มละ 6 คน	64	46	50	**160**
รวม		**414**	**400**	**373**	**1187**

ตาราง 4 ตารางข้อมูลสูญหาย 3 ค่า คือ Y_1, Y_2 และ Y_3

ปัจจัย A	ปัจจัย B	Block			รวม
		1	2	3	
ชาย	กลุ่มละ 2 คน	31	33	29	93
	กลุ่มละ 3 คน	36	32	36	104
	กลุ่มละ 4 คน	38	Y_1	26	$64+Y_1$
	กลุ่มละ 5 คน	53	43	49	145
	กลุ่มละ 6 คน	56	63	Y_2	$119+Y_2$
หญิง	กลุ่มละ 2 คน	43	35	41	119
	กลุ่มละ 3 คน	46	43	48	137
	กลุ่มละ 4 คน	47	47	41	135
	กลุ่มละ 5 คน	Y_3	58	53	$111+Y_3$
	กลุ่มละ 6 คน	64	46	50	160
รวม		$414+Y_3$	$400+Y_1$	$373+Y_2$	$1187+Y_1+Y_2+Y_3$

ประมาณค่าสูญหายของข้อมูลทั้ง 3 วิธี ดังนี้

1.4.1 วิธีประมาณค่าสูญหายโดยการวนซ้ำ

$$Y = \frac{bB + kT - G}{(b-1)(k-1)} \qquad \text{........................(1)}$$

เมื่อ Y แทน ค่าที่สูญหาย

b แทน จำนวนบล็อก

k แทน จำนวนทรีตเมนต์ของ 2 ปัจจัย

B แทน ผลรวมของค่าข้อมูลในบล็อกที่มีค่าสูญหาย

T แทน ผลรวมของค่าข้อมูลในทรีตเมนต์ที่มีค่าสูญหาย

G แทน ผลรวมของข้อมูลทั้งหมด

ขั้นตอนการประมาณข้อมูลสูญหาย 3 ค่า

รอบที่ 1

1) ประมาณตัวเลขมาทดแทนค่าข้อมูลที่สูญหายเป็นการชั่วคราว โดยให้เหลือค่าข้อมูลที่สูญหายไว้เพียงค่าเดียว คำนวณจากสูตร

$$\overline{Y}_i = \frac{(\overline{M}_t + \overline{M}_b)}{2} \qquad \text{……………………(2)}$$

โดยเว้นไว้เพียงค่าเดียว เมื่อ

$\overline{Y}_i$ แทน ค่าเฉลี่ยของค่าสูญหาย เมื่อ i=1,2,…,m

$\overline{M}_t$ แทน ค่าเฉลี่ยของค่าข้อมูลทั้งหมดของทรีตเมนต์ 2 ปัจจัย

$\overline{M}_b$ แทน ค่าเฉลี่ยของค่าข้อมูลทั้งหมดของบล็อก

คำนวณข้อมูลสูญหาย ค่าที่ 1

$$\overline{M}_{t_1} = \frac{26+38}{2} = \frac{64}{2} = 32$$

$$\overline{M}_{b_1} = \frac{33+32+43+63+35+43+47+58+46}{9} = \frac{400}{9} = 44.44$$

$$\therefore \quad \overline{Y}_1 = \frac{32+44.44}{2} = 38.22$$

คำนวณข้อมูลสูญหาย ค่าที่ 2

$$\overline{M}_{t_2} = \frac{56+63}{2} = \frac{119}{2} = 59.5$$

$$\overline{M}_{b_2} = \frac{29+36+26+49+41+48+41+53+50}{9} = \frac{373}{9} = 41.44$$

$$\therefore \quad \overline{Y}_2 = \frac{59.5+41.44}{2} = 50.47$$

1.) แทนค่า $\overline{Y}_1 = 38.22$, $\overline{Y}_2 = 50.47$ แล้วคำนวณหาค่า Y_3

$$\therefore Y_3 = \frac{3(414)+10(111)-(1187+38.22+50.47)}{(3-1)(10-1)}$$

$$= \frac{1242+1110-1275.69}{(2)(9)}$$

$$= \frac{1076.31}{18}$$

$$= 59.76$$

2.) แทนค่า $\overline{Y}_2 = 50.47$, $Y_3 = 59.76$ แล้วคำนวณหาค่า Y_1

$$\therefore Y_1 = \frac{3(400) + 10(64) - (1187 + 50.47 + 59.76)}{(3-1)(10-1)}$$

$$= \frac{1200 + 640 - 1297.23}{(2)(9)}$$

$$= \frac{542.77}{18}$$

$$= 30.15$$

3.) แทนค่า $Y_3 = 59.76$, $Y_1 = 30.15$ แล้วคำนวณหาค่า Y_2

$$\therefore Y_2 = \frac{3(373) + 10(119) - (1187 + 59.76 + 30.15)}{(3-1)(10-1)}$$

$$= \frac{1119 + 1190 - 1276.91}{(2)(9)}$$

$$= \frac{1032.09}{18}$$

$$= 57.33$$

รอบที่ 2

1.) แทนค่า $Y_2 = 57.33$, $Y_3 = 59.76$ แล้วคำนวณหาค่า Y_1

$$\therefore Y_1 = \frac{3(400) + 10(64) - (1187 + 57.33 + 59.76)}{(3-1)(10-1)}$$

$$= \frac{1200 + 640 - 1304.09}{(2)(9)}$$

$$= \frac{535.91}{18}$$

$$= 29.77$$

2.) แทนค่า $Y_1 = 29.77$, $Y_3 = 59.76$ แล้วคำนวณหาค่า Y_2

$$\therefore Y_2 = \frac{3(373) + 10(119) - (1187 + 29.77 + 59.76)}{(3-1)(10-1)}$$

$$= \frac{1119 + 1190 - 1276.53}{(2)(9)}$$

$$= \frac{1032.47}{18}$$

$$= 57.36$$

3.) แทนค่า $Y_1 = 29.77$, $Y_2 = 57.36$ แล้วคำนวณหาค่า Y_3

$$\therefore Y_3 = \frac{3(414) + 10(111) - (1187 + 29.77 + 57.36)}{(3-1)(10-1)}$$

$$= \frac{1242 + 1110 - 1274.13}{(2)(9)}$$

$$= \frac{1077.87}{18}$$

$$= 59.88$$

รอบที่ 3

1.) แทนค่า $Y_2 = 57.36$, $Y_3 = 59.88$ แล้วคำนวณหาค่า Y_1

$$\therefore Y_1 = \frac{3(400) + 10(64) - (1187 + 57.36 + 59.88)}{(3-1)(10-1)}$$

$$= \frac{1200 + 640 - 1304.24}{(2)(9)}$$

$$= \frac{535.76}{18}$$

$$= 29.76$$

2.) แทนค่า $Y_1 = 29.76$, $Y_3 = 59.88$ แล้วคำนวณหาค่า Y_2

$$\therefore Y_2 = \frac{3(373) + 10(119) - (1187 + 29.76 + 59.88)}{(3-1)(10-1)}$$

$$= \frac{1119 + 1190 - 1276.64}{(2)(9)}$$

$$= \frac{1031.34}{18}$$

$$= 57.35$$

3.) แทนค่า $Y_1 = 29.76$, $Y_2 = 57.35$ แล้วคำนวณหาค่า Y_3

$$\therefore Y_3 = \frac{3(414) + 10(111) - (1187 + 29.76 + 57.35)}{(3-1)(10-1)}$$

$$= \frac{1242 + 1110 - 1274.11}{(2)(9)}$$

$$= \frac{1077.89}{18}$$

$$= 59.88$$

รอบที่ 4

1.) แทนค่า $Y_2 = 57.35$, $Y_3 = 59.88$ แล้วคำนวณหาค่า Y_1

$$\therefore Y_1 = \frac{3(400) + 10(64) - (1187 + 57.35 + 59.88)}{(3-1)(10-1)}$$

$$= \frac{1200 + 640 - 1304.22}{(2)(9)}$$

$$= \frac{535.77}{18}$$

$$= 29.76$$

2.) แทนค่า $Y_1 = 29.76$, $Y_3 = 59.88$ แล้วคำนวณหาค่า Y_2

$$\therefore Y_2 = \frac{3(373) + 10(119) - (1187 + 29.76 + 59.88)}{(3-1)(10-1)}$$

$$= \frac{1119 + 1190 - 1276.64}{(2)(9)}$$

$$= \frac{1031.34}{18}$$

$$= 57.35$$

3.) แทนค่า $Y_1 = 29.76$, $Y_2 = 57.35$ แล้วคำนวณหาค่า Y_3

$$\therefore Y_3 = \frac{3(414) + 10(111) - (1187 + 29.76 + 57.35)}{(3-1)(10-1)}$$

$$= \frac{1242 + 1110 - 1274.11}{(2)(9)}$$

$$= \frac{1077.89}{18}$$

$$= 59.88$$

ดังนั้น ค่าข้อมูลที่สูญหาย 3 ค่า คือ $Y_1 = 29.76$, $Y_2 = 57.35$ และ $Y_3 = 59.88$

1.4.2 วิธีประมาณค่าสูญหายของวิลคินซัล ใช้ข้อมูลตาราง 4

จากสูตร

$$nY_i = bB_{Y_i} + kT_{Y_i} - G$$

โดยที่ n แทน จำนวนข้อมูลทั้งหมด $(n = bk)$

b แทน จำนวนบล็อก

k แทน จำนวนทรีตเมนต์ของ 2 ปัจจัย

B_{Y_i} แทน ผลรวมของค่าข้อมูลในบล็อกที่มี Y_i สูญหาย

T_{Y_i} แทน ผลรวมของค่าข้อมูลในทรีตเมนต์ที่มี Y_i สูญหาย

G แทน ผลรวมของค่าข้อมูลทั้งหมดที่มี Y_i สูญหาย

แทนค่าในสมการ หาค่า Y_1

$$\begin{aligned} 30Y_1 &= 3B_{Y_1} + (2)(5)T_{Y_1} - G \\ &= 3(400 + Y_1) + 10(64 + Y_1) - (1187 + Y_1 + Y_2 + Y_3) \\ &= 653 + 12Y_1 - Y_2 - Y_3 \end{aligned}$$

$\therefore\ 18Y_1 + Y_2 + Y_3 = 653$

แทนค่าในสมการ หาค่า Y_2

$$\begin{aligned} 30Y_2 &= 3B_{Y_2} + (2)(5)T_{Y_2} - G \\ &= 3(373 + Y_2) + 10(119 + Y_2) - (1187 + Y_1 + Y_2 + Y_3) \\ &= 1122 - Y_1 + 12Y_2 - Y_3 \end{aligned}$$

$\therefore\ Y_1 + 18Y_2 + Y_3 = 1122$

แทนค่าในสมการ หาค่า Y_3

$$\begin{aligned} 30Y_3 &= 3B_{Y_3} + (2)(5)T_{Y_3} - G \\ &= 3(414 + Y_3) + 10(111 + Y_3) - (1187 + Y_1 + Y_2 + Y_3) \\ &= 1165 - Y_1 - Y_2 + 12Y_3 \end{aligned}$$

$\therefore\ Y_1 + Y_2 + 18Y_3 = 1165$

แก้สมการหาค่า Y_1, Y_2 และ Y_3 โดยใช้แมทริกซ์อินเวอร์ส

$$A^{-1} = \frac{1}{|A|}\left[A_{ij}\right]$$

จากสูตร

$$Y_{ijk} = A^{-1}B$$

กำหนดให้ A คือ แมทริกซ์สัมประสิทธิ์ของสมการ

$$A = \begin{bmatrix} 18 & 1 & 1 \\ 1 & 18 & 1 \\ 1 & 1 & 18 \end{bmatrix}$$

แมทริกซ์ A^{-1} คือ

$$A^{-1} = \begin{bmatrix} 0.05 & -2.94\times10^{-3} & -2.94\times10^{-3} \\ -2.94\times10^{-3} & 0.05 & -2.94\times10^{-3} \\ -2.94\times10^{-3} & -2.94\times10^{-3} & 0.05 \end{bmatrix}$$

กำหนดให้ B คือ แมทริกซ์คำตอบของสมการ

$$B = \begin{bmatrix} 653 \\ 1122 \\ 1165 \end{bmatrix}$$

แทนค่าในสมการ

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 0.05 & -2.94\times10^{-3} & -2.94\times10^{-3} \\ -2.94\times10^{-3} & 0.05 & -2.94\times10^{-3} \\ -2.94\times10^{-3} & -2.94\times10^{-3} & 0.05 \end{bmatrix} \begin{bmatrix} 653 \\ 1122 \\ 1165 \end{bmatrix}$$

$$= \begin{bmatrix} 29.76 \\ 57.35 \\ 59.88 \end{bmatrix}$$

1.4.3 วิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน ใช้ข้อมูลตาราง 4 เพื่อให้ง่ายต่อการคำนวณ จัดตารางใหม่ได้ดังตาราง 5

ตาราง 5 จัดตารางข้อมูลใหม่ เพื่อคำนวณวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน

	V1	V2	V3
R1	31	33	29
R2	36	32	36
R3	38	Y_1	26
R4	53	43	49
R5	56	63	Y_2
R6	43	35	41
R7	46	43	48
R8	47	47	41
R9	Y_3	58	53
R10	64	46	50

ข้อมูลในตัวอย่าง มีค่าสูญหายทั้งหมด 3 ค่า คือ Y_1, Y_2 และ Y_3 ซึ่งอยู่ในตำแหน่ง R3V2, R5V3 และ R9V1 ตามลำดับ มีขั้นตอนในการประมาณค่าสูญหายดังนี้

1. กำหนดค่า K = 3
2. คำนวณหาระยะห่างระหว่างจุดด้วยวิธี Euclidian Distance ระหว่างข้อมูลที่มีค่าสูญหายและข้อมูลที่มีความสมบูรณ์

2.1 คำนวณหาค่า Y_1 โดยแทนค่า Y_2 และ Y_3 ด้วยค่าเฉลี่ย ดังตาราง 6

ตาราง 6 คำนวณหาค่าเฉลี่ย

	V1	**V2**	**V3**	**ค่าเฉลี่ย**
R1	31	33	29	
R2	36	32	36	
R3	38	Y_1	26	
R4	53	43	49	
R5	56	63	Y_2	**59.5**
R6	43	35	41	
R7	46	43	48	
R8	47	47	41	
R9	Y_3	58	53	**55.5**
R10	64	46	50	

2.1.1 คำนวณหาระยะห่างระหว่างจุดด้วยวิธี Euclidian Distance ได้ดังนี้

$$dist(R3, R1) = \sqrt{(38-31)^2 + (26-29)^2} = 7.6$$

$$dist(R3, R2) = \sqrt{(38-36)^2 + (26-36)^2} = 10.1$$

$$dist(R3, R4) = \sqrt{(38-53)^2 + (26-49)^2} = 27.4$$

$$dist(R3, R5) = \sqrt{(38-56)^2 + (26-59.5)^2} = 38.0$$

$$dist(R3, R6) = \sqrt{(38-43)^2 + (26-41)^2} = 15.8$$

$$dist(R3, R7) = \sqrt{(38-46)^2 + (26-48)^2} = 23.4$$

$$dist(R3, R8) = \sqrt{(38-47)^2 + (26-41)^2} = 17.4$$

$$dist(R3, R9) = \sqrt{(38-55.5)^2 + (26-53)^2} = 32.1$$

$$dist(R3, R10) = \sqrt{(38-64)^2 + (26-50)^2} = 35.3$$

2.1.2 เรียงลำดับระยะห่างระหว่างจุดโดยพิจารณาจากข้อมูลที่ใกล้ที่สุดตามจำนวน k ที่ได้กำหนดไว้ คือ k=3 ได้ผลลัพธ์ดังตาราง 7

ตาราง 7 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_1

	V1	V2	V3	dist	sort
R1	31	33	29	7.6	1
R2	36	32	36	10.1	2
R6	43	35	41	15.8	3

2.1.3 ประมาณค่าข้อมูลสูญหาย Y_1 ซึ่งอยู่ในตำแหน่ง R3V2 ได้ดังนี้

$$\hat{R}3\hat{V}2 = \frac{\sum_{i=1}^{k}(R1V2 + R2V2 + R6V2)}{k}$$

$$= \frac{(33 + 32 + 35)}{3}$$

$$= 33.33$$

เพราะฉะนั้น Y_1 มีค่าเท่ากับ 33.33

2.2 คำนวณหาค่า Y_2 โดยแทนค่า Y_1 = 33.33 และ Y_3 ด้วยค่าเฉลี่ย

2.2.1 คำนวณหาระยะห่างระหว่างจุดด้วยวิธี Euclidian Distance ได้ดังนี้

$$dist(R5, R1) = \sqrt{(56 - 31)^2 + (63 - 33)^2} = 39.0$$

$$dist(R5, R2) = \sqrt{(56 - 36)^2 + (63 - 32)^2} = 36.8$$

$$dist(R5, R3) = \sqrt{(56 - 38)^2 + (63 - 33.33)^2} = 34.7$$

$$dist(R5, R4) = \sqrt{(56 - 53)^2 + (63 - 43)^2} = 20.2$$

$$dist(R5, R6) = \sqrt{(56 - 43)^2 + (63 - 47)^2} = 30.8$$

$$dist(R5, R7) = \sqrt{(56 - 46)^2 + (63 - 43)^2} = 22.3$$

$$dist(R5, R8) = \sqrt{(56 - 47)^2 + (63 - 47)^2} = 18.3$$

$$dist(R5, R9) = \sqrt{(56 - 55.5)^2 + (63 - 58)^2} = 5.0$$

$$dist(R5, R10) = \sqrt{(56 - 64)^2 + (63 - 46)^2} = 18.7$$

2.2.2 เรียงลำดับระยะห่างระหว่างจุดโดยพิจารณาจากข้อมูลที่ใกล้ที่สุดตามจำนวน k ที่ได้กำหนดไว้ คือ k=3 ได้ผลลัพธ์ดังตาราง 8

ตาราง 8 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_2

	V1	V2	V3	dist	sort
R8	47	47	41	18.3	2
R9	55.5	58	53	5.0	1
R10	64	46	50	18.7	3

2.2.3 ประมาณค่าข้อมูลสูญหาย Y_2 ซึ่งอยู่ในตำแหน่ง R5V3 ได้ดังนี้

$$\hat{R}5\hat{V}3 = \frac{\sum_{i=1}^{k}(R9V3 + R8V3 + R10V3)}{k}$$

$$= \frac{(53+41+50)}{3}$$

$$= 48$$

เพราะฉะนั้น Y_2 มีค่าเท่ากับ 48

2.3 คำนวณหาค่า Y_3 โดยแทนค่า Y_1 = 33.33 และ Y_2 = 48

2.3.1 คำนวณหาระยะห่างระหว่างจุดด้วยวิธี Euclidian Distance ได้ดังนี้

$$dist(R9, R1) = \sqrt{(58-33)^2 + (53-29)^2} = 34.6$$

$$dist(R9, R2) = \sqrt{(58-32)^2 + (53-36)^2} = 31.0$$

$$dist(R9, R3) = \sqrt{(58-33.33)^2 + (53-26)^2} = 36.5$$

$$dist(R9, R4) = \sqrt{(58-43)^2 + (53-49)^2} = 15.5$$

$$dist(R9, R5) = \sqrt{(58-63)^2 + (53-48)^2} = 7.0$$

$$dist(R9, R6) = \sqrt{(58-35)^2 + (53-41)^2} = 25.9$$

$$dist(R9, R7) = \sqrt{(58-43)^2 + (53-48)^2} = 15.8$$

$$dist(R9, R8) = \sqrt{(58 - 47)^2 + (53 - 41)^2} = 16.2$$
$$dist(R9, R10) = \sqrt{(58 - 64)^2 + (53 - 50)^2} = 12.3$$

2.3.2 เรียงลำดับระยะห่างระหว่างจุดโดยพิจารณาจากข้อมูลที่ใกล้ที่สุดตามจำนวน k ที่ได้กำหนดไว้ คือ k=3 ได้ผลลัพธ์ดังตาราง 9

ตาราง 9 เรียงลำดับระยะห่างระหว่างจุด เพื่อประมาณค่าข้อมูลสูญหาย Y_3

	V1	**V2**	**V3**	**dist**	**sort**
R4	53	43	49	15.5	3
R5	56	63	48	7.0	1
R10	64	46	50	12.3	2

2.3.3 ประมาณค่าข้อมูลสูญหาย Y_3 ซึ่งอยู่ในตำแหน่ง R9V1 ได้ดังนี้

$$\hat{R}9\hat{V}1 = \frac{\sum_{i=1}^{k}(R5V1 + R10V1 + R4V1)}{k}$$

$$= \frac{(56 + 64 + 53)}{3}$$

= 57.67

เพราะฉะนั้น Y_3 มีค่าเท่ากับ 57.67

ตอนที่ 2 ลักษณะของข้อมูลสูญหาย

ข้อมูลที่ได้จากการทดลองนั้น ก่อนที่จะนำข้อมูลไปวิเคราะห์จะต้องตรวจสอบความสมบูรณ์ของข้อมูล ศุภลักษณ์ กรรณิการ์(2549: 8 – 9) ได้กล่าวถึงสิ่งสำคัญที่ควรตรวจสอบก่อนการนำข้อมูลไปวิเคราะห์ ได้แก่

1. ข้อมูลสูญหาย (Missing data) หมายถึง ข้อมูลที่มีบางหน่วยหรือเซลล์ของเมตริกซ์หายไป

2. ข้อมูลไม่คงเส้นคงวา (Inconsistent) หรือมีความคลาดเคลื่อน (Non sampling Error) หมายถึง ข้อมูลที่ค่าหรือความหมายของตัวแปรหนึ่งหรือหลายตัวแปลขัดแย้งกับค่าหรือความหมายของตัวแปรอื่น

3. ข้อมูลมีสิ่งรบกวน (Noisy) หรือมีข้อมูลผิดปกติ (Outliers) ซึ่งข้อมูลที่มีสิ่งรบกวนหมายถึง ความคลาดเคลื่อนเชิงสุ่ม หรือความแปรปรวนในการวัดค่าของตัวแปร ส่วนข้อมูลผิดปกติหรือมีสิ่งไม่พึงประสงค์ หมายถึง ข้อมูลที่มีลักษณะอย่างใดอย่างหนึ่ง หรือหลายลักษณะต่อไปนี้

- มีค่าสูงหรือต่ำมาก ๆ (Extreme Value)
- มีค่าแตกต่างจากข้อมูลส่วนใหญ่
- มีค่าอยู่นอกช่วงที่กำหนด

2.1 รูปแบบของข้อมูลสูญหาย (Patterns of Missing Data) (ดวงภรณ์ โปทาวี. 2552: 9 – 10 อ้างอิงจาก Chantala and Suchindran. n.d. : online)
สามารถจำแนกได้ 4 ลักษณะ คือ

1. ข้อมูลสูญหายหนึ่งตัวแปร
2. ข้อมูลขาดหายมากกว่าหนึ่งตัวแปร
3. ข้อมูลสูญหายเป็นไปในทิศทางเดียวกัน
4. ข้อมูลสูญหายแบบไม่เป็นระบบ

ดังรายละเอียดต่อไปนี้

1. ข้อมูลสูญหายหนึ่งตัวแปร (Univariate Nonresponse) คือ ตัวแปร 1 ตัว มีข้อมูลสูญหาย

ตาราง 10 ข้อมูลสูญหายหนึ่งตัวแปร

Case	Y_1	Y_2	Y_3
A	4	7	8
B	7	6	
C	5	8	
D	6	6	8

2. ข้อมูลขาดหายมากกว่าหนึ่งตัวแปร (Multivariate Two Patterns) คือ มีข้อมูลสูญหายมากกว่าหนึ่งตัวแปรในหน่วยตัวอย่างเดียวกัน

ตาราง 11 ข้อมูลขาดหายมากกว่าหนึ่งตัวแปร

Case	Y_1	Y_2	Y_3
A	4	7	8
B	7	6	6
C	5		
D	6		

3. ข้อมูลสูญหายเป็นไปในทิศทางเดียวกัน (Monotone) คือ อันดับของตัวแปรมีความสำคัญ นิยามคือ ให้เซตของตัวแปรคือ Y_1, $Y_2,\ldots, Y_p$ ถ้า Y_i มีค่าสูญหาย แล้ว Y_{i+1}, $Y_{i+2},\ldots, Y_p$ จะมีค่าสูญหายด้วย

ตาราง 12 ข้อมูลสูญหายเป็นไปในทิศทางเดียวกัน

Case	Y_1	Y_2	Y_3	Y_4
A	4	7	4	8
B	7	5	6	
C	5	6		
D	6	7	8	5

4. ข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) โดยข้อมูลสูญหายสามารถเกิดขึ้นตรงจุดไหนก็ได้และอันดับของตัวแปรไม่มีความสำคัญ

ตาราง 13 ข้อมูลสูญหายแบบไม่เป็นระบบ

Case	Y_1	Y_2	Y_3	Y_4
A	4	7	4	8
B		5	6	
C	5	6		7
D	6		8	5

การวิจัยที่มีกิจกรรมยุ่งยากหรือซับซ้อนและมีช่วงเวลาที่ยาวนาน โดยทั่วไปมักเกิดสภาวการณ์เกี่ยวกับการสูญเสียหรือขาดหายของตัวอย่างสำหรับการศึกษาวิจัยในระหว่างการดำเนินงานวิจัย ซึ่งอาจเกิดจากการถอนตัวและการไม่ให้ความร่วมมือของหน่วยตัวอย่างเนื่องจากเกิดความเครียด เหนื่อยล้า เจ็บป่วย เสียชีวิต หรืออพยพย้ายถิ่น (องอาจ นัยพัฒน์. 2554: 101) ซึ่งตัวแปรเหล่านี้ตรงกับงานวิจัยทางการศึกษา เพราะกลุ่มตัวอย่างส่วนใหญ่จะเป็นนักเรียน บางครั้งก็ไม่สามารถจัดเก็บข้อมูลได้ ลักษณะของข้อมูลที่สูญหายไม่สามารถที่จะระบุจุดที่ข้อมูลจะสูญหายได้และอันดับของตัวแปรก็ไม่มีความสำคัญ เพราะฉะนั้น งานวิจัยในครั้งนี้ ผู้วิจัยจึงกำหนดให้ข้อมูลสูญหายอยู่ในรูปแบบข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) ซึ่งเป็นรูปแบบข้อมูลสูญหายที่เกิดขึ้นบ่อยในแผนการทดลองมากที่สุด

2.2 สาเหตุของข้อมูลสูญหาย

ลิทเทิลและรูบิน(เชาว์ อินใย. 2547: 15; อ้างอิงจาก Little; & Rubin. 1987: 6 – 7) ได้กล่าวถึงสาเหตุของข้อมูลสูญหาย (missing data mechanisms) มี 3 ประเภท คือ

1. ข้อมูลสูญหายแบบสุ่มสมบูรณ์ (missing completely at random : MCAR) เช่น ถ้ากลุ่มตัวอย่างได้รับการคัดเลือกขึ้นมาแบบสุ่มเพื่อให้ได้รับการวัดอย่างใดอย่างหนึ่งจะมีกลุ่มตัวอย่างบางคนที่ไม่สามารถวัดได้อย่างสมบูรณ์ ข้อมูลที่สูญหายไปจะเป็นแบบ MCAR ถึงแม้ว่าการสูญหายจะเกิดขึ้นจากตัวแปรบางตัว แต่ไม่สัมพันธ์กับตัวแปรที่มีข้อมูลสูญหาย ข้อมูลสูญหายยังคงมีลักษณะเป็น MCAR ข้อดีของการสูญหายแบบ MCAR ก็คือ สาเหตุของการสูญหายไม่ต้องนำไปเป็นส่วนหนึ่งของการวิเคราะห์เพื่อควบคุมอคติของการสูญหายมีวิธีการแบบเก่าในการจัดการข้อมูล

สูญหาย เช่น การตัดข้อมูลออกแบบลิสท์ไวส์ ให้ผลการวิเคราะห์ที่ไม่มีอคติ เมื่อข้อมูลสูญหายเป็นแบบ MCAR แต่ก็เป็นวิธีการที่ไม่ดีเพราะอำนาจการทดสอบมีค่าต่ำ

2. ข้อมูลสูญหายแบบมีระบบ (systematic missing data) การสูญหายของข้อมูลเกิดขึ้นจากตัวแปรอื่น ๆ ที่มีความสัมพันธ์กับตัวแปรที่มีข้อมูลสูญหาย สาเหตุของการสูญหายจะไม่เป็นแบบ MCAR ถ้าสาเหตุของการสูญหายวัดได้แล้วนำมาวิเคราะห์จะเรียกว่าเป็น accessible missing data mechanisms ดังนั้น อคติทั้งหมดที่เกี่ยวข้องกับข้อมูลสูญหายจะถูกปรับ ลิทเทิลและรูบิน (Little; & Rubin.1987) เรียกสถานการณ์นี้ว่า ignorable (missing at random)

3. มีสถานการณ์อื่น ๆ ที่เป็นสาเหตุของการสูญหาย เช่น สาเหตุของการสูญหายไม่สามารถวัดได้ และสาเหตุของการสูญหายสัมพันธ์กับตัวแปรที่มีข้อมูลสูญหาย จะเรียกว่าเป็น inaccessible missing data mechanisms (nonignorable mechanisms) สถานการณ์นี้เกิดขึ้นเมื่อค่าของตัวแปรสูญหายเป็นสาเหตุให้เกิดการสูญหาย เช่น คนที่ดื่มสุรามาก ๆ จะหลีกเลี่ยงการตรวจแอลกอฮอล์มากกว่าคนที่ดื่มน้อย หรือเด็กที่ดื้อรั้นจะต่อต้านการตรวจปัสสาวะทั้ง ๆ ที่เขาอาจจะใช้ยาหรือไม่ใช้ยาก็ได้ การสูญหายของข้อมูลก็จะเกิดขึ้นและเป็นสาเหตุที่เข้าถึงไม่ได้ inaccessible mechanisms

2.3 วิธีการจัดการข้อมูลสูญหาย

ลิทเทิลและรูบิน (Little; & Rubin. 1987: 6 – 7) ได้แบ่งวิธีการจัดการข้อมูลสูญหายไว้ 4 วิธี คือ

1. วิธีที่ใช้ข้อมูลสมบูรณ์ (Procedures based on completely recorded units) วิธีนี้เป็นวิธีที่ง่ายที่สุดโดยการตัดข้อมูลที่สูญหายออกไปแล้ววิเคราะห์ข้อมูลที่สมบูรณ์เท่านั้น เช่น การตัดข้อมูลแบบลิสท์ไวส์ (listwise deletion) การตัดข้อมูลแบบแพร์ไวส์ (pairwise deletion) เป็นวิธีที่มีอยู่ในโปรแกรมคอมพิวเตอร์ เช่น โปรแกรม SPSS

2. วิธีการแทนค่า (Imputation based procedures) วิธีนี้ใช้การแทนข้อมูลสูญหายด้วยค่าที่ได้จากวิธีการต่าง ๆ เช่น การแทนค่าแบบฮอทเดค (hot deck imputation) การแทนค่าโดยใช้ค่าเฉลี่ย (mean imputation) และการแทนค่าโดยวิธีการถดถอย (regression imputation)

3. วิธีการถ่วงน้ำหนัก (Weighting procedures) เป็นวิธีที่ใช้การถ่วงน้ำหนักนำไปปรับข้อมูลที่สูญหาย การถ่วงน้ำหนักมีความสัมพันธ์กับการแทนค่าด้วยค่าเฉลี่ย ถ้าน้ำหนักที่กำหนดไว้เป็นค่าคงที่ในกลุ่มตัวอย่างย่อยแล้ว ทั้งการแทนค่าด้วยค่าเฉลี่ยของกลุ่มย่อย และการให้น้ำหนักหน่วยที่ตอบเป็นสัดส่วนของการตอบในแต่ละกลุ่มย่อย จะทำให้การแทนค่าข้อมูลสูญหายเหมือนกัน

4. วิธีการที่ได้จากการนิยมโมเดล (model – based procedures) วิธีนี้ได้จากการนิยามโมเดลของข้อมูลสูญหายบางส่วนและใช้หลักการอ้างอิงเกี่ยวกับไลค์ลิฮูด (likelihood) ในโมเดลด้วยวิธีการประมาณค่าพารามิเตอร์ที่เรียกว่าแม๊กซิมัมไลค์ลิฮูด (maximum likelihood) โดยวิธีการทำซ้ำ

วิธีการประมาณค่าสูญหายนั้นมีหลายวิธี นักวิจัยต้องเลือกวิธีประมาณค่าให้เหมาะสมในแต่ละสถานการณ์ จึงจะทำให้ข้อมูลที่ได้มีความน่าเชื่อถือ แม่นยำ สำหรับงานวิจัยนี้ได้ทำการเปรียบเทียบวิธีการประมาณค่าสูญหายเพียง 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายวิลคินซัล และวิธีประมาณค่าสูญหายเคเนียเรส เนย์บอร์อิมพิวเทชัน ดังรายละเอียดได้กล่าวไปแล้วนั้น

2.4 จำนวนของข้อมูลสูญหาย

เชาว์ อินใย (2546: 58) ได้กล่าวว่า การประมาณค่าข้อมูลสูญหาย จำนวนข้อมูลสูญหายก็มีความสำคัญเป็นอย่างมาก เพราะจะทำให้ผลที่ได้มีความน่าเชื่อถือมากหรือน้อย ขึ้นอยู่กับจำนวนของข้อมูลสูญหาย ถ้ามีข้อมูลสูญหายจำนวนน้อยประมาณ 5% การตัดหน่วยตัวอย่างออกไปดูเหมือนว่าจะมีเหตุผลในการแก้ปัญหาข้อมูลสูญหาย แต่ถ้ามีการสูญหายจำนวนมาก การตัดข้อมูลออกจะไม่มีประสิทธิภาพข้อมูลที่เหลืออยู่จะไม่เป็นตัวแทนของประชากรซึ่งมีเป้าหมายในการอ้างอิง (Schafer.1997: 1 ; Little; & Rubin. 1987: 5) แต่ยังมีงานวิจัยของเรมอนด์และโรเบิร์ท (Roth. 1994: 545 ; citing Raymond; & Robert.1987) ที่กล่าวว่า ถ้าใช้วิธีการถดถอยจัดการข้อมูลสูญหายจะมีความเหมาะสมเมื่อมีข้อมูลสูญหายมากกว่า 20% ส่วนเฟรดและลี (Fred; & Lii. 1998: online) ได้ศึกษาประเมินความถูกต้องของวิธีการจัดการข้อมูลสูญหายหลายวิธีภายใต้รูปแบบของการสูญหายที่แตกต่างกันโดยกำหนดจำนวนของข้อมูลสูญหายแบบสุ่มมีค่าต่ำสุดเป็น 10% และสูงสุดเท่ากับ 20% นอกจากนี้มีนักวิจัยคนอื่นๆได้กำหนดจำนวนข้อมูลสูญหายแตกต่างออกไป เช่น รูท (Roth. 1994: 551) ได้กล่าวว่า การเลือกวิธีการจัดการข้อมูลสูญหายจะมีความสำคัญเมื่อจำนวนข้อมูลสูญหายอยู่ระหว่าง 15 – 20% และจะมีความสำคัญมากที่สุดเมื่อจำนวนข้อมูลสูญหายมีค่าเท่ากับ 30 – 40% ที่ระดับนี้การเลือกใช้วิธีการจัดการข้อมูลสูญหายจะให้ผลลัพธ์ที่แตกต่างกัน จะเห็นได้ว่าจำนวนข้อมูลสูญหายมีค่าอยู่ระหว่าง 5 – 40% จากงานวิจัยและเอกสารดังกล่าวข้อมูลสูญหายจะอยู่ระหว่าง 5% ถึง 20% อยู่ในช่วงที่ผู้วิจัยยอมรับและสามารถนำไปวิเคราะห์ต่อไปได้

ในการวิจัยครั้งนี้ ผู้วิจัยกำหนดจำนวนข้อมูลสูญหายให้มีจำนวนมากกว่า 20% โดยกำหนดจำนวนข้อมูลสูญหายสูงสุดเท่ากับ 30% เพื่อให้ได้ข้อค้นพบที่เป็นประโยชน์มากยิ่งขึ้นในเรื่องของจำนวนข้อมูลสูญหาย ซึ่งจำนวนข้อมูลสูญหายดังกล่าวสอดคล้องกับงานวิจัยของรูท (Roth) ที่กล่าวว่า การเลือกวิธีการจัดการข้อมูลสูญหายมีความสำคัญมากที่สุดเมื่อข้อมูลสูญหายมีค่าอยู่ระหว่าง 30 – 40% (เชาว์ อินใย. 2547: 58)

ดังนั้น ในการศึกษาวิจัยครั้งนี้ผู้วิจัยจึงกำหนดจำนวนข้อมูลสูญหายออกเป็น 3 ระดับ คือ 10%, 20% และ 30%

ตอนที่ 3 ประชากรและการสุ่มตัวอย่าง

ความหมายของประชากร

กัลยา วานิชย์บัญชา (2549: 3) ให้ความหมายของประชากรว่า ประชากรในทางสถิติ ประชากรหมายถึงทุกหน่วยในเรื่องที่สนใจศึกษา หน่วยต่างๆในประชากร อาจหมายถึง บุคคล กลุ่มบุคคล องค์กรต่างๆ สัตว์ สิ่งของ ฯลฯ

ชูศรี วงศ์รัตนะ(2544: 2) ให้ความหมายของประชากรว่า ประชากรหมายถึง กลุ่มทั้งหมดของคน สัตว์หรือสิ่งของ ซึ่งมีคุณลักษณะร่วมกันตรงกับผู้วิจัยสนใจจะศึกษา เช่น ถ้าต้องการศึกษาความคิดเห็นของนิสิตมหาวิทยาลัยเกษตรที่มีต่อการปกครองแบบประชาธิปไตย นิสิตทุกคนในมหาวิทยาลัยเกษตรศาสตร์เป็นประชากร

นวลอนงค์ บุญฤทธิพงศ์ (2552: 131) ให้ความหมายของประชากรว่า ประชากรหมายถึง ทุกหน่วยของสิ่งที่ผู้วิจัยต้องการศึกษา อาจเป็นคน สัตว์ สิ่งของ สถานที่ ฯลฯ ที่อยู่ในขอบข่ายการวิจัยของผู้วิจัย เช่น ถ้าผู้วิจัยต้องการศึกษาความรู้ของชาวนา ประชากรคือ ชาวนาทุกคน

จากการศึกษางานวิจัยทางด้านการศึกษาที่วางแผนงานวิจัยเป็นแบบแผนการทดลองพบว่า ประชากรส่วนใหญ่จะเป็นนักเรียน ผู้วิจัยแบ่งประชากรออกเป็น 3 ขนาด คือขนาดใหญ่มีประชากรมากกว่า 1,000 คน ขนาดกลางมีประชากร 501 – 999 คน และขนาดเล็กมีประชากรน้อยกว่า 500 คน งานวิจัยที่มีประชากรขนาดใหญ่ เช่น รุ่งโรจน์ ศรีจันทร์แก้ว (2547: 60) ศึกษาเปรียบเทียบอำนาจการทดสอบ (Power of test) ของผลการวิเคราะห์ข้อมูลด้วยแบบแผนการวิเคราะห์แบบกลุ่มสุ่ม(RBD) กับแบบแผนการวิเคราะห์ความแปรปรวนร่วม (ANCOVA) ประชากรคือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2546 โรงเรียนสังกัดในกรุงเทพมหานคร จำนวน 8,980 คน งานวิจัยที่มีประชากรขนาดกลาง เช่น พรวุฒิ คำแก้ว (2546: 64) ศึกษาผลของ

การใช้บทเรียนคอมพิวเตอร์มัลติมีเดีย 3 รูปแบบที่มีผลสัมฤทธิ์ทางการเรียนความคงทนในการเรียนรู้และเจตคติต่อบทเรียนที่มีระดับความสามารถต่างกัน 3 ระดับ ประชากรคือ นักเรียนชั้นประถมศึกษาปีที่ 5 โรงเรียนอนุบาลมุกดาหารและโรงเรียนสตรีวรนาถ บางเขน จำนวน 800 คน และงานวิจัยที่มีประชากรขนาดเล็ก เช่น อรุณี เต๊ะอ้วน (2550: 71) ศึกษาเปรียบเทียบผลการเรียนรู้สาระดนตรี บทเรียนคอมพิวเตอร์มัลติมีเดียสองรูปแบบที่ใช้แผนการทดลองที่ต่างกัน ประชากรคือ นักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนสังกัดกรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2549 จำนวน 334 คน พัลลภ พิริยะสุรวงศ์ (2542: 77) ศึกษาการออกแบบและพัฒนามัลติมีเดียแบบฝึกโดยใช้รูปแบบการควบคุมการเรียนต่างกัน ประชากรคือ นักศึกษาระดับประกาศนียบัตรวิชาชีพ(ปวช.) ชั้นปีที่ 3 ภาคเรียนที่ 1/2542 ของวิทยาลัยเทคโนโลยีอุตสาหกรรม สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ จำนวน 279 คน สุวัตน์ นิยมไทย (2533: 33) ศึกษาผลการเรียนรู้วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ซึ่งมีขนาดของกลุ่มต่างกัน ประชากรคือ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2531 โรงเรียนสตรีวิทยา 2 เขตบางเขน กรุงเทพมหานคร จำนวน 132 คน

ส่วนใหญ่ประชากรที่ใช้ศึกษาในแผนการทดลองมีจำนวนไม่มาก เพราะเป็นการทดลองขนาดเล็ก ซึ่งมีจำนวนน้อยกว่า 500 คน เพื่อให้การวิจัยในครั้งนี้ให้ผลใกล้เคียงกับการปฏิบัติทดลองจริงมากที่สุด ผู้วิจัยจึงทำการจำลองข้อมูลประชากรตามสถานการณ์ต่างๆ ที่กำหนด ด้วยเทคนิคมอนติคาร์โล ในโปรแกรม R โดยกำหนดให้ประชากรมีขนาดเท่ากับ 512

ความหมายของตัวอย่าง

กัลยา วานิชย์บัญชา (2544: 3) ให้ความหมายของตัวอย่างว่า ตัวอย่างเป็นเซ็ทย่อยของประชากร หมายถึงบางส่วนของประชากร เช่น ต้องการหาอายุเฉลี่ยของคนไทย ประชากรคือคนไทยทุกคน ตัวอย่างคือคนไทยบางคนที่ถูกเลือกเป็นตัวอย่าง

ชูศรี วงศ์รัตนะ (2544: 2) ให้ความหมายของตัวอย่างว่า ตัวอย่างหมายถึง บางส่วนของกลุ่มประชากร

นวลอนงค์ บุญฤทธิพงศ์ (2552: 131) ให้ความหมายของตัวอย่างว่า ตัวอย่างหมายถึงหน่วยของประชากรที่ถูกเลือก เพื่อให้เป็นตัวแทนของประชากร กลุ่มตัวอย่างจึงเป็นส่วนหนึ่งของประชากรที่ใช้ในการวิจัย

ลักษณะของตัวอย่างที่ดี

ตัวอย่างที่ใช้ในการวิจัยต้องเป็นตัวอย่างที่ดี ลักษณะของตัวอย่างที่ดี มี 2 ประการ ดังนี้ (นวลอนงค์ บุญฤทธิพงศ์. 2552: 132)

1. เป็นตัวอย่างที่ดีของประชากร ในการวิจัยส่วนใหญ่ จะศึกษาจากกลุ่มตัวอย่างที่มีจำนวนน้อย เพื่อสรุปอ้างอิงไปยังประชากร ตัวอย่างที่ดีจึงต้องมีลักษณะต่างๆ ที่คล้ายคลึงกับลักษณะของประชากรมากที่สุด ดังนั้น ในการทำวิจัย ก่อนจะได้เลือกตัวอย่าง ผู้วิจัยต้องกำหนดประเภท ขอบเขต และลักษณะของประชากรที่จะศึกษาก่อน จะได้เลือกตัวอย่างที่เหมาะสมกับลักษณะของประชากรที่จะศึกษา

2. มีขนาดพอเหมาะ การสรุปผลทางการวิจัย มีการใช้สถิติเพื่อทดสอบความเชื่อมั่นทางสถิติ ซึ่งจะต้องมีข้อมูลมากพอที่จะทำการทดสอบทางสถิติ เพื่อให้สามารถสรุปอ้างอิงไปยังประชากรได้ กลุ่มตัวอย่างที่จะนำมาใช้ในการวิจัยจึงต้องมีขนาดที่พอเหมาะกับการทดสอบทางสถิติด้วย

หลักการเลือกตัวอย่างเพื่อให้เป็นตัวแทนที่ดีของประชากร

การเลือกตัวอย่างให้เป็นตัวแทนที่ดีของประชากร มีหลักการ ดังนี้ (บุญใจ ศรีสถิตนรากูร. 2544: 171)

1. ใช้หลักความน่าจะเป็น เป็นการเลือกที่พยายามหาวิธีให้ทุกหน่วยของประชากรมีโอกาสถูกเลือกเป็นตัวอย่างโดยเท่าเทียมกัน เช่น การสุ่มอย่างง่าย การสุ่มอย่างมีระบบ การสุ่มแบบชั้นภูมิ เป็นต้น

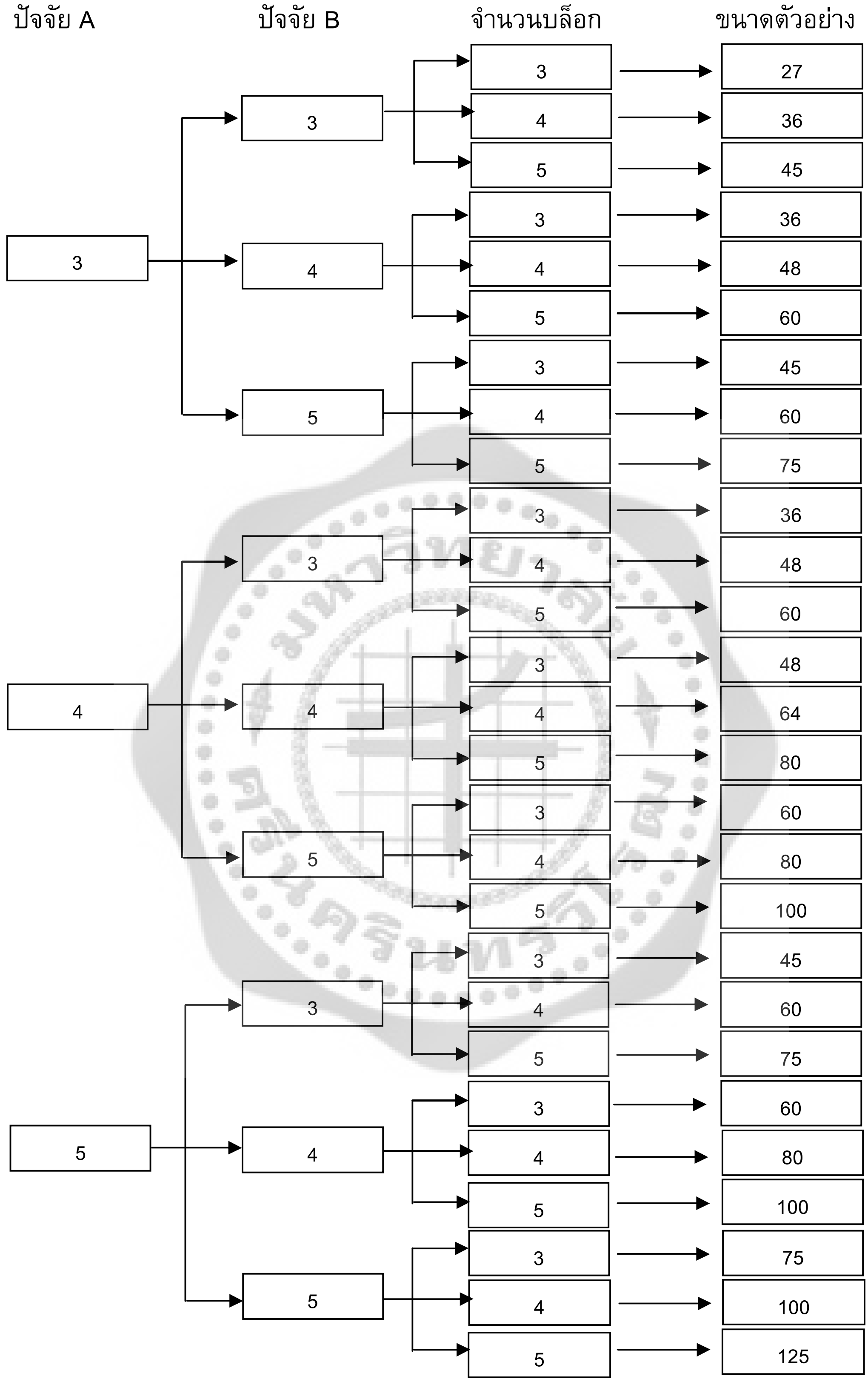
2. ใช้วิธีที่เหมาะสม ในการเลือกตัวอย่าง นอกจากจะใช้หลักความน่าจะเป็นแล้ว ผู้วิจัยควรพิจารณาด้วยว่าวิธีใดที่มีความเหมาะสมที่สุดกับงานวิจัยที่ศึกษา และเป็นวิธีที่จะทำให้กลุ่มตัวอย่างมีคุณลักษณะครอบคลุมคุณลักษณะของประชากรมากที่สุด

3. มีขนาดที่เหมาะสม เนื่องจากขนาดกลุ่มตัวอย่างแปรผกผันกับค่าความคลาดเคลื่อนมาตรฐาน กล่าวคือ ถ้ากลุ่มตัวอย่างมีขนาดใหญ่จะทำให้ค่าความคลาดเคลื่อนลดลงได้ แต่สิ้นเปลืองงบประมาณและเวลา ดังนั้น การกำหนดขนาดของกลุ่มตัวอย่าง ผู้วิจัยควรพิจารณาจากปัจจัยอื่นๆ ที่เกี่ยวข้อง

การหาขนาดตัวอย่างสำหรับงานวิจัยเชิงทดลองนั้น นักวิธีวิทยาการวิจัยบางท่าน แนะนำให้ใช้ตัวอย่างขนาด 15 คนต่อกลุ่ม สำหรับการวิจัยเชิงทดลอง/กึ่งทดลอง (องอาจ นัยพัฒน์. 2548: 116; อ้างอิงจาก Borg; & Gall.1989) หรือให้ใช้การเปิดจากตารางสำเร็จรูป เหตุผลที่ทำให้กำหนดขนาดตัวอย่างด้วยวิธีดังกล่าวไม่ได้ เพราะมีปัจจัยหลายประการเข้ามาเกี่ยวข้องกับการกำหนดตัวอย่างเช่น ความถูกต้องแม่นยำของผลการศึกษาจากตัวอย่าง ความเป็นเอกพันธ์ของประชากร งบประมาณในการใช้จ่ายและเวลา ระดับนัยสำคัญ และขนาดผลการจัดกระทำทางการวิจัย จากการทบทวนเอกสารและงานวิจัยทางด้านการศึกษาที่วางแผนงานวิจัยเป็นแบบแผนการทดลอง มีขนาดตั้งแต่ 30 – 180 (รุ่งโรจน์ ศรีจันทร์แก้ว. 2547: 60, พรวุฒิ คำแก้ว.2546: 64,พัลลภ พิริยะสุรวงศ์. 2542: 77, สุวัฒน์ นิยมไทย, 2533: 33, อรุณี เต๊ะอ้วน. 2550: 4) ในงานวิจัยครั้งนี้ ผู้วิจัยกำหนดขนาดตัวอย่าง โดยมีขั้นตอน ดังนี้

1. ทำการสุ่มประชากรการทดลองแบบแฟคทอเรียล ตัวแบบอิทธิพลกำหนด รูปแบบเชิงเส้นตรงของแผนการทดลองแบบ a×b แฟคทอเรียลใน RCB มีขนาดเท่ากับ 512

2. สุ่มขนาดตัวอย่างในประชากร จำแนกตามปัจจัย A ปัจจัย B และบล็อก ให้มีขนาดตามแต่ละสถานการณ์ และในแต่ละเซลล์ของแผนการทดลองแบบแฟคทอเรียล มักจะมีความแปรปรวนเกิดขึ้น เพื่อลดความแปรปรวน ผู้วิจัยจึงกำหนดให้แต่ละเซลล์มีค่าเท่ากับ 1 ผู้วิจัยสุ่มขนาดของตัวอย่างดังภาพประกอบ 2



ภาพประกอบ 2 แผนผังแสดงขนาดตัวอย่าง

ตอนที่ 4 การจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล

ความเป็นมาของวิธีมอนติคาร์โล

เทคนิควิธีมอนติคาร์โล (เชาว์ อินใย. 2547: 32 – 34) เป็นสาขาหนึ่งของคณิตศาสตร์ ใช้คอมพิวเตอร์ช่วยในการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่ม (Random Number) มาสร้างตัวแปรให้เหมือนกับสถานการณ์จริงและมีการทดลองซ้ำหลายๆ ครั้ง เพื่อให้ได้ค่าที่แน่นอนที่จะใช้เป็นข้อสรุปหรืออธิบายปรากฏการณ์ต่างๆ ในสถานการณ์จริง (ต่าย เซี่ยงฉี. 2534: 62 – 68) หรือช่วยหาคำตอบในเรื่องราวต่างๆ ที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น

เทคนิควิธีมอนติคาร์โล ได้มีการใช้มานานแล้วแต่ในสมัยก่อนๆไม่ได้เรียกว่ามอนติคาร์โล โดยนำมาใช้พัฒนาทฤษฎีความน่าจะเป็น (Probability Theory) ในราวปี ค.ศ. 1753 จอร์ส หลุ่ยส์ เลคเลอร์ และบุฟฟอง (Georges Louis Leclere and Comte de Buffon) ทำการทดลองหาค่า (π) โดยการโยนเข็มที่มีความยาว k หน่วย อย่างสุ่มลงมาบนพื้นราบที่มีเส้นขนานอยู่ โดยให้ระยะห่างระหว่างเส้นขนานแต่ละเส้นห่างกัน d หน่วย และกำหนดให้ $d>k$ จะได้ความน่าจะเป็นที่เข็มจะตัดกับเส้นขนาน $P = 2k / \pi d$ ซึ่งถ้าความน่าจะเป็น (P) เป็นค่าสุ่ม ก็จะหาค่า π ได้ ในราวปี ค.ศ. 1908 กอสเซท (W.S.Gosset) ได้ศึกษาหารแจกแจงความถี่ของความสูงของนักโทษอาชญากรรม จำนวน 3,000 คน โดยเทียบกับการแจกแจงความถี่ของกลุ่มตัวอย่างที่สุ่มมาครั้งละ 4 คน จำนวน 750 กลุ่มตัวอย่าง ผลการศึกษาพบว่า การแจกแจงความถี่ทั้งสองมีลักษณะเหมือนกัน กอสเซทได้ตั้งชื่อการแจกแจงความถี่ที่ค้นพบนี้ว่า การแจกแจงค่าที (t – distribution) ซึ่งถือได้ว่าเป็นจุดเริ่มต้นของเทคนิควิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคมอนติคาร์โล ได้รับการพัฒนาอย่างจริงจังในปี ค.ศ. 1944 ช่วงสงครามโลกครั้งที่ 2 อูลาม และออน นิวแมน (Ulam and Von Neumann) เป็นผู้ตั้งชื่อ มอนติคาร์โล เป็นรหัสลับของงานที่ทำใน ลอส อาลามอส (Los Alemos) ได้นำเทคนิคนี้มาหาผลของการเผยแพร่อย่างสุ่มของนิวตรอน (Neutron diffusion) ในวัสดุเชื้อเพลิงซึ่งเป็นการทดลองทางคณิตศาสตร์เพื่อหาผลของคำตอบก่อนที่จะทำการทดลองจริง ซึ่งทำให้ไม่เกิดอันตรายและช่วยประหยัดค่าใช้จ่ายก่อนการทดลองจริง หลังจากนั้นเทคนิคมอนติคาร์โล ได้มีการนำมาใช้อย่างกว้างขวางทั้งทางด้านฟิสิกส์ คณิตศาสตร์ สถิติ และการวิจัย นับได้ว่าเทคนิควิธีมอนติคาร์โล มีประโยชน์อย่างมากในการขยายความรู้เชิงทฤษฎี เช่น การนำมาศึกษาค่าความคลาดเคลื่อนทางสถิติ เปรียบเทียบประสิทธิภาพการทดสอบต่างๆ เป็นต้น

ขั้นตอนของระเบียบวิธีมอนติคาร์โล

หลักการที่สำคัญของวิธีมอนติคาร์โล ก็คือ การนำเอาตัวเลขสุ่ม (Random Number) มาประยุกต์แก้ปัญหาต่างๆ มีขั้นตอนที่สำคัญ ดังนี้

1. สร้างตัวเลขสุ่ม (Generate random number) ระยะแรกทำได้โดยอาศัยเครื่องมือทางกายภาพ เช่น ล้อรูเล็ต ลูกเต๋า ไพ่ กระดาษเขียนเบอร์ เป็นต้น แต่ได้ตัวเลขสุ่มไม่มาก ต่อมาจึงหันมาใช้เครื่องมืออิเล็กทรอนิกส์ เช่น เครื่องสร้างตัวเลขสุ่มที่สร้างขึ้นโดยบริษัท แรนด์ (Rand) โดยสร้างตัวเลขสุ่มจากเครื่องกำเนิดพัลส์ (Pulse) ซึ่งสามารถสร้างตัวเลขสุ่มได้เป็นล้านตัว

การสร้างหรือเลือกใช้ตัวเลขสุ่มดังกล่าวกับเครื่องคอมพิวเตอร์ยังมีปัญหา 2 ประการ คือ เป็นการยากที่จะทำให้คอมพิวเตอร์สามารถเรียกใช้ได้เมื่อมีความต้องการ และยากที่จะทำให้เครื่องมือดังกล่าวสร้างตัวเลขชุดเดิม เมื่อต้องการใช้เปรียบเทียบวิธีการต่างๆ ภายใต้เงื่อนไขของระบบเลขสุ่มชุดเดียว หรือถ้าจะเก็บเลขสุ่มเหล่านี้ไว้ในหน่วยความจำหรือจานแม่เหล็กก็จะทำให้สูญเสียหน่วยความจำหรือเสียเวลาในการค้นหา ฉะนั้นการสร้างตัวเลขสุ่มในคอมพิวเตอร์จึงนิยมสร้างตัวเลขสุ่มเทียม (Pseudo random number) โดยอาศัยสูตรทางคณิตศาสตร์

ในปัจจุบันมีโปรแกรมสำหรับสร้างตัวเลขสุ่มในเครื่องคอมพิวเตอร์ เช่น ในภาษาเบสิก (Basic) มีคำสั่งเรียกใช้ตัวเลขสุ่มได้คือ RANDOMIZED และ RND ในภาษาฟอร์แทน (FORTRAN) ก็มีคำสั่งเรียกตัวเลขสุ่มได้ คือ RANDUM ส่วนภาษาฟอกโปร (FOXPRO) มีคำสั่งเรียกตัวเลขสุ่ม คือ RAND()

คุณสมบัติของตัวเลขสุ่มที่ดี

1. ตัวเลขสุ่มที่ได้ต้องมีลักษณะการกระจายความน่าจะเป็นแบบสม่ำเสมอ (Uniform Distribution)
2. ตัวเลขสุ่มที่ได้ต้องเป็นอิสระต่อกัน
3. อนุกรมของตัวเลขสุ่มที่ได้ต้องสามารถสร้างซ้ำเดิมได้
4. อนุกรมตัวเลขสุ่มที่ได้ต้องไม่ซ้ำเดิมในช่วงที่ต้องการใช้ตัวเลขสุ่ม
5. ต้องใช้เวลาน้อยในการสร้างตัวเลขสุ่ม
6. ต้องใช้หน่วยความจำในคอมพิวเตอร์น้อย

2. นำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาต่างๆ เป็นการนำตัวเลขสุ่มไปสร้างตัวแปรตามลักษณะการแจกแจงของปัญหาที่จะศึกษาเพื่อเป็นข้อมูลของปัญหานั้น เช่น สร้างตัวเลขสุ่มแล้วนำตัวเลขสุ่มไปสร้างเป็นคะแนนการสอบของผู้เรียน แต่บางครั้งตัวแปรของปัญหาที่จะศึกษาไม่ได้สร้างจากตัวเลขสุ่มโดยตรง แต่ใช้ตัวเลขสุ่มเป็นพื้นฐานก็ได้

3. ทำการทดลองซ้ำหลายๆ ครั้ง การศึกษาด้วยวิธีมอนติคาร์โล ต้องมีการทดลองซ้ำหลายๆ ครั้ง เพื่อลดความคลาดเคลื่อนของคำตอบที่จะได้ และสามารถสรุปเป็นความน่าจะเป็นของการเกิดเหตุการณ์ในปัญหานั้นๆ

จุดเด่นของการใช้เทคนิคมอนติคาร์โล

เทคนิคมอนติคาร์โล ใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้างตัวแปรของปัญหาโดยอาศัยทฤษฎี สูตร หรือกฎเกณฑ์ต่างๆ ที่มีอยู่ และมีการทดลองซ้ำหลายๆ ครั้ง เพื่อลดความคลาดเคลื่อนต่างๆ ซึ่งนับว่ามีประโยชน์ที่สำคัญดังนี้

1. สามารถควบคุมตัวแปรแทรกซ้อนและสามารถสังเกตได้อย่างสมบูรณ์และทำการทดลองซ้ำภายใต้สภาพแวดล้อม (Context) เดิมหลายๆ ครั้งได้ ส่วนในการทดลองจริงนั้นทำไม่ได้เพราะไม่สามารถรักษาสภาพแวดล้อมให้เหมือนเดิมทุกอย่างได้เมื่อเวลาเปลี่ยนไป

2. ถ้ามีสูตรหรือหรือกฎเกณฑ์ต่างๆ ที่ถูกต้องรองรับในการสร้างตัวแปรของปัญหาในการทดลองแล้วจะให้ผลที่ถูกต้องแม่นยำกว่านี้เมื่อใช้ทดลองในสถานการณ์จริง เพราะสามารถลดตัวแปรแทรกซ้อนเชิงจิตวิทยาได้

3. สิ้นเปลืองเวลา แรงงาน และค่าใช้จ่ายน้อยกว่าเมื่อเทียบกับการทดลองในสถานการณ์จริง

ตอนที่ 5 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีประมาณค่าสูญหาย

เกณฑ์ที่ใช้ในการเปรียบเทียบว่าวิธีประมาณค่าสูญหายวิธีใดจะให้ค่าประมาณดีที่สุด จะพิจารณาจากการคำนวณหาค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error : MAE) หาค่าประมาณจากวิธีการต่าง ๆ และนำมาเปรียบเทียบกับค่าจริง ดังนี้

$$MAE = Max[Max|y_i - \hat{y}_i|]_n$$ (ศุภลักษณ์ กรรณิกา. 2549: 10)

เมื่อ y_i แทน ค่าจริงที่ i ได้จากการจำลอง

$\hat{y}_i$ แทน ค่าประมาณ ค่าที่ i ที่ได้จากการใช้วิธีประมาณค่า

n แทน รอบที่ n จากการทดลอง (กำหนดให้ $n=1000$ รอบ)

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด คำนวณจากค่าประมาณและค่าจริงที่ได้จากการจำลอง กล่าวคือ ในการจำลองข้อมูล 1 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ระหว่างค่าจริงที่ได้จากการจำลองกับค่าที่ได้จากการประมาณค่าสูญหายจำนวน 1 ค่า การศึกษาในครั้งนี้ทำการจำลองข้อมูล 1,000 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ 1,000 ค่า พิจารณาเอาค่าที่มีค่ามากที่สุดใน 1,000 ค่า เป็นตัวแทนของค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ในแต่ละสถานการณ์และแต่ละวิธีมาเปรียบเทียบกัน วิธีใดให้ค่าความคลาดเคลื่อนสูงสุดมีค่าต่ำที่สุด แสดงว่า ค่าประมาณที่ได้มีค่าใกล้เคียงกับค่าจริงที่สูญหายไปมากที่สุด จะเป็นวิธีการประมาณค่าสูญหายที่ดีที่สุด

ตอนที่ 6 งานวิจัยที่เกี่ยวข้อง

ปราณี ศรีภา (2532: 1 – 106) ศึกษาเปรียบเทียบวิธีวิเคราะห์เมื่อมีค่าทุกค่าสูญหายในบล็อกใดบล็อกหนึ่งในแผนการทดลองแบบบล็อกไม่สมบูรณ์ที่สมดุลย์ 3 วิธี คือ วิธีของพูริ วิธีของวิลคินซัน และวิธีประมาณค่าสูญหายโดยให้ค่าความคลาดเคลื่อนกำลังสองต่ำสุดเมื่อประชากรมีการแจกแจงและแผนการทดลองแบบต่าง ๆ ผลการศึกษาพบว่า ทั้ง 3 วิธีให้ค่าความเคลื่อนสัมบูรณ์เฉลี่ยไม่แตกต่างกัน และทั้ง 3 วิธีให้ค่าอำนาจทดสอบสูงที่สุดทุก ๆ สถานการณ์ ดังนั้น เมื่อเกิดปัญหาข้อมูลสูญหายทั้งบล็อกในแผนการทดลองแบบบล็อกไม่สมบูรณ์สมดุลขึ้น ควรเลือกใช้วิธีประมาณค่าสูญหายโดยให้ค่าความคลาดเคลื่อนกำลังสองต่ำสุด

ชุติมา ชัยมุสิก (2533: 1 – 91) ศึกษาเปรียบเทียบการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอยเชิงซ้อน 4 วิธี คือ วิธีสมการถดถอย วิธีเม็กซิมั่มไลลิฮูด วิธีค่าเฉลี่ย และวิธีค่ามัธยฐาน จากกลุ่มตัวอย่าง 30 70 และ 100 การกระจายข้อมูล 3 ระดับ โดยใช้ C.V. เป็นตัวกำหนด คือ .05 .20 และ1.00 จำนวนตัวแปรอิสระ 4 ระดับ คือ 2 3 5 และ 7 ค่าเบี่ยงเบนมาตรฐาน 4 ระดับ คือ 5 10 20 และ 25 และสัดส่วนข้อมูลที่สูญหายของตัวแปรอิสระ 3 ระดับ คือ 5% 10% และ 15% ทำการศึกษาโดยใช้วิธีมอนติคาร์โล เกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของสมการถดถอยของวิธีที่ไม่มีข้อมูลสูญหาย ผลการศึกษาพบว่าวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอยเชิงซ้อนทั้ง 4 วิธี ให้ผลต่างกันตามสถานการณ์ต่าง ๆ ซึ่งโดยส่วนใหญ่วิธีค่าเฉลี่ยให้ผลดีที่สุด ยกเว้นเมื่อมีขนาดตัวอย่างน้อยและจำนวนตัวแปรอิสระมาก วิธีสมการถดถอยจะให้ผลดีที่สุด แต่ถ้าตัวอย่างมีขนาดใหญ่และจำนวนตัวแปรอิสระมีน้อย การตัดชุดของข้อมูลสูญหายทิ้งจะไม่มีผลกระทบต่อการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด

วารุณี ตรีบำรุงศักดิ์ (2538: 1 – 125) ศึกษาการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณเมื่อตัวแปรตามมีค่าสูญหาย วิธีที่ใช้ประมาณค่าตัวแปรตามเมื่อมีข้อมูลสูญหาย คือ วิธีค่าเฉลี่ย วิธีสมการถดถอย วิธีอีเอ็ม (EM algorithm) และวิธีของฮันท์ (Hunt's Method) ใช้กลุ่มตัวอย่างขนาด 10 20 30 50 70 ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน 5 10 15 20 และ 25 สัดส่วนการสูญหายของตัวแปรตาม 10% 20% 30% 40% 50% 60% และ 70% ทำการศึกษาด้วยวิธีมอนติคาร์โล และหารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของค่าพยากรณ์ ผลการศึกษาพบว่าวิธีการของฮันท์เป็นวิธีการที่ดี เมื่อกลุ่มตัวอย่างมีขนาดเล็ก ความคลาดเคลื่อนน้อย และสัดส่วนการสูญหายมาก แต่ถ้าความคลาดเคลื่อนสูง วิธีค่าเฉลี่ยจะเป็นวิธีที่ดีในทุกสัดส่วนการสูญหายของตัวแปรตาม ส่วนในสถานการณ์ที่ขนาดตัวอย่างมีขนาดใหญ่ วิธีสูญหายจะเหมาะสมเกือบทุกกรณี

รุ่งกานต์ กาใจคำ (2540: 1 – 64) ศึกษาการประมาณค่าสูญหายโดยใช้กระบวนการ R ซึ่งเป็นกระบวนการที่จะได้มาซึ่งเมตริกซ์สุดท้าย โดยมีข้อสมมติของการทดลองว่าการทดลองต้องไม่มีผลกระทบร่วมเกิดขึ้นในแต่ละแถว และในแต่ละหลักของการทดลองจะต้องมีค่าสังเกตอย่างน้อย 1 ค่าสังเกต ผลการศึกษาพบว่า เมื่อเปรียบเทียบวิธีประมาณค่าสูญหายโดยใช้ค่าเฉลี่ย และวิธีประมาณค่าสูญหายโดยใช้กระบวนการ R ในแผนการทดลองแบบบล็อกสุ่มและแผนการทดลองแบบบล็อกไม่สมบูรณ์ ซึ่งจัดว่าเป็นแผนการทดลองแบบสองทางแล้ว ปรากฏว่าค่าประมาณที่ได้จากวิธีการทั้ง 2 วิธีมีค่าใกล้เคียงกัน ดังนั้นผู้วิจัยสามารถเลือกใช้วิธีประมาณค่าสูญหายวิธีใดก็ได้ใน 2 วิธีนี้

ดวงฤดี เหง้าพรมนิล (2545: 1 – 582) ศึกษาคุณสมบัติและวิธีการวิเคราะห์ความแปรปรวนปัจจัยเดียวเมื่อมีค่าสูญหายในแผนการทดลองแบบบล็อกสุ่มสมบูรณ์ 5 วิธี ได้แก่ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวินคิลซัล วิธีทดสอบนัยสำคัญโดยวิธีของการถดถอยทั่วไป วิธีวิเคราะห์ความแปรปรวนร่วม และวิธีวิเคราะห์ของแดสและกีริ กำหนดให้จำนวนบล็อกเท่ากับ 3 5 7 และ 9 และจำนวนสิ่งทดลองเท่ากับ 3 5 และ7 รูปแบบการแจกแจงของความคลาดเคลื่อนของการทดลองแบบปกติ ที และลอกนอร์มอล จากการศึกษาพบว่า เมื่อความคลาดเคลื่อนของการทดลองมีการแจกแจงแบบปกติและที สำหรับทุกค่าของค่าเฉลี่ยรวมและขนาดของความแตกต่างระหว่างอิทธิพลของบล็อก กรณีที่มีค่าสูญหายเท่ากับ 1 ค่า ควรเลือกใช้วิธีประมาณค่าสูญหายโดยการวนซ้ำหรือวิธีประมาณค่าสูญหายของวิลคินซัล แต่เมื่อมีค่าสูญหายเท่ากับ 2 และ 3 ค่า ควรเลือกใช้วิธีทดสอบนัยสำคัญโดยวิธีของการถดถอยทั่วไป วิธีวิเคราะห์ความแปรปรวนร่วม และวิธีวิเคราะห์ของแดสและกีริ ซึ่งจะใช้วิธีการวิเคราะห์ใดนั้นขึ้นอยู่กับข้อกำหนดของแต่ละวิธี เมื่อความคลาดเคลื่อนของการทดลองมีการแจกแจงแบบลอกนอร์มอล สำหรับทุกค่าของค่าสูญหาย ค่าเฉลี่ยรวมและขนาดของความแตกต่างระหว่างอิทธิพลของบล็อก ควรใช้วิธีประมาณค่าสูญหายโดยการวนซ้ำ หรือวิธีประมาณค่าสูญหายของวิลคินซัล

ประพจน์ ดำรงค์สุทธิพงศ์ (2546: 1 – 149) ศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ 3 วิธี ได้แก่ การประมาณค่าวิธีกำลังสองน้อยที่สุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และการประมาณค่าวิธีอิมพิวเทชัน จำลองด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม S – PLUS 2000 โดยศึกษาภายใต้ความคลาดเคลื่อนมีการแจกแจงแบบปกติ การเปรียบเทียบกระทำเมื่อจำนวนระดับปัจจัยของวิธีการทดลอง เท่ากับ 3 4 และ 5 ระดับ จำนวนบล็อค เท่ากับ 2 4 และ 6 บล็อค กำหนดให้จำนวนเปอร์เซ็นต์ข้อมูลสูญหาย 10% 20% และ 30% กำหนดให้สัมประสิทธิ์การแปรผัน (C.V.) เป็น 5% 25% และ 45% และทำการเปรียบเทียบโดยใช้เกณฑ์ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE (Mean Square Error) ผลการศึกษาพบว่า เปอร์เซ็นต์ข้อมูลสูญหายเมื่อมีค่ามากขึ้นค่าเฉลี่ยกำลังสองของค่าคลาดเคลื่อนจะมีค่าสูงขึ้น โดยวิธีอิมพิวเทชัน ให้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน MSE (Mean Square Error) ต่ำกว่าวิธีค่าคาดหวังสูงสุด และวิธีกำลังสองน้อยที่สุดในทุกสถานการณ์ของการทดลองที่ทำการศึกษา

เชาว์ อินใย (2547: 1 – 332) ศึกษาการพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอี และตรวจสอบความแม่นยำ และอำนาจการทดสอบที่ได้จากวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีกับแบบอีเอ็มและแบบลิสท์ไวส์ ตามวิธีการสุ่มตัวอย่างแบบแบ่งชั้น แบบ กลุ่ม และแบบหลายขั้นตอน ที่ความสัมพันธ์ระหว่างตัวแปรระดับต่ำ (r = .30) ปานกลาง (r = .50) และสูง (r = .70) และจำนวนข้อมูลสูญหาย 5% 10% 20% และ 30% และศึกษาปฏิสัมพันธ์ระหว่างวิธีการสุ่มตัวอย่าง วิธีการประมาณค่าสูญหาย จำนวนข้อมูลสูญหาย และความสัมพันธระหว่างตัวแปร ที่มีต่อความแม่นยำของค่าเฉลี่ยเลขคณิต ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์ ข้อมูลที่ใช้ศึกษามีลักษณะการแจกแจงแบบปกติสองตัวแปร และใช้วิธีมอนติคาร์โล จำลองการทดลองด้วยเครื่องคอมพิวเตอร์ ผลการศึกษาพบว่าวิธีการประมาณค่าสูญหายโดยการแทนค่าแบบอีพีเอสเอสอีได้ค่าความแม่นยำของค่าเฉลี่ยเลขคณิต ไม่แตกต่างจากวิธีอีเอ็มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 วิธีประมาณค่าสูญหายโดยการตัดออกแบบลิสท์ไวส์ ได้ค่าความแม่นยำของความแปรปรวนแตกต่างจากวิธีอื่น ๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และวิธีการประมาณค่าสูญหายแบบลิสท์ไวส์ ได้ค่าความแม่นยำของสัมประสิทธิ์สหสัมพันธ์แตกต่าง จากวิธีการประมาณค่าสูญหายแบบอื่นๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ศุภลักษณ์ กรรณิกา (2549: 1 – 96) ศึกษาและเปรียบเทียบวิธีประมาณค่าสูญหายในการวางแผนการทดลองแบบจตุรัสละติน 3 วิธี คือ การประมาณค่าวิธีกำลังสองน้อยที่สุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และ การประมาณค่าวิธีมัลลติเพิล อิมพิวเทชัน จำลองข้อมูลด้วยเทคนิคมอนติคาร์โล กำหนดให้จำนวนวิธีทดลองที่ใช้ทดลองเท่ากับ 3 4 5 6 และ 7 สัมประสิทธิ์ความผันแปรเท่ากับ 5% 25% และ45% และจำนวนข้อมูลสูญหายเท่ากับ 10% 20% และ 30% ทำการทดลองซ้ำๆ กัน 500 ครั้งในแต่ละสถานการณ์ที่กำหนด และทำการเปรียบเทียบโดยใช้เกณฑ์ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ผลการศึกษาพบว่า เมื่อเปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น ค่าความคาดเคลื่อนสัมบูรณ์สูงสุดจะมีค่าเพิ่มขึ้น สำหรับกรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่ามาก พบว่าการประมาณค่าสูญหายวิธีมัลติเพิล อิมพิวเทชัน จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด กรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่าน้อย พบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นควรเลือกใช้วิธีกำลังสองน้อยที่สุดในการประมาณค่าสูญหายเนื่องจากสะดวกและรวดเร็วกว่า

จริยา แสงสุวรรณ (2551: 1 – 57) ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยพหุคูณ โดยทำการประมาณค่าสูญหายของตัวแปรตาม 4 วิธี คือ วิธีสูญหาย วิธีค่าเฉลี่ย วิธีสมการถดถอย และวิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่า (วิธีเอ็มไอ) เกณฑ์ที่ใช้เปรียบเทียบคือ ค่าประมาณของรากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ใช้ขนาดตัวอย่างเท่ากับ 50 70 100 และ 200 ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1 5 และ15 เปอร์เซนต์การสูญหายของตัวแปรตามเท่ากับ 5% 10% 20% และ30% ตามลำดับ และตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร ซึ่งระดับความสัมพันธ์ระหว่างตัวแปรอิสระมี 3 ระดับ คือ ระดับต่ำ (0.20) ระดับปานกลาง (0.50) และระดับสูง (0.70) ทำการจำลองด้วยวิธีมอนติคาร์โล ซึ่งกระทำซ้ำ 5,000 ครั้ง ในแต่ละสถานการณ์ ผลการศึกษาพบว่า เมื่อเปอร์เซ็นต์การสูญหายเพิ่มขึ้น วิธีสมการถดถอยและวิธีเอ็มไอ ให้ค่าประมาณของ RMSE ลดลง และวิธีการประมาณค่าสูญหายทั้ง 4 วิธี ให้ค่าประมาณค่าของ RMSE แตกต่างกัน วิธีสมการถดถอยและวิธีเอ็มไอ ให้ค่าประมาณของ RMSE ใกล้เคียงกัน แต่เนื่องจากวิธีสมการถดถอยเป็นวิธีอย่างง่ายและไม่ซับซ้อน ดังนั้น วิธีสมการถดถอย จึงเป็นวิธีที่เหมาะสมสำหรับการประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยพหุคูณ

อุษณีย์ วงศ์อามาตย์ (2555: 1 – 93) ศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยเชิงเส้นพหุ เมื่อข้อมูลตัวแปรตามมีการสูญหายแบบนอนอิกนอร์เรเบิล วิธีการประมาณค่าสูญหายที่ใช้ในงานการวิจัยนี้คือ วิธี EM Algorithm (EM) วิธี K-Nearest Neighbor Imputation (KNN) และวิธี Predictive Mean Matching Imputation (PMM)ข้อมูลที่ใช้ในการศึกษาได้จากการจำลองโดยมีสัดส่วนของการสูญหาย 3 ระดับคือ 10%, 20%, 30% และมีระดับของการสูญหายแบบนอนอิกนอร์เรเบิล 3 ระดับคือ ไม่มี, ปานกลาง, สูง จากการเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average mean square error; AMSE) พบว่า i) วิธีการประมาณทุกวิธีสามารถประมาณได้ดีขึ้นเมื่อขนาดตัวอย่างมีขนาดใหญ่ขึ้น ii) วิธีการประมาณทุกวิธีประมาณได้แย่ลงเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อน สัดส่วนของการสูญหาย และ ระดับของการสูญหายแบบอิกนอร์เรเบิล มีค่าเพิ่มขึ้น iii) โดยรวมแล้ววิธี EM ประมาณค่าได้ดีที่สุดเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนมีค่าไม่สูง (10 – 30) และ iv) วิธี KNN ประมาณค่าได้ดีที่สุดเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนมีค่าสูง(90)

คราวฟอร์ดและคณะ (Crawford; et al. 1995: 209 – 219) ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลที่สูญหาย 4 วิธี คือวิธีการใส่ค่าเพียงค่าเดียว วิธีเอ็มไอ วิธีค่าเฉลี่ย และวิธีสูญหาย ข้อมูลที่ใช้ในการศึกษามาจากโครงการสุขภาพคนชราของมลรัฐแมชซะ – ซูเซ็ท วัตถุประสงค์เพื่อเปรียบเทียบวิธีการจัดการข้อมูลสูญหายทั้ง 4 วิธี กระทำภายใต้แบบแผนการสูญหายต่างกัน ผลการศึกษาพบว่า วิธีสูญหาย และวิธีค่าเฉลี่ย จะให้ค่าประมาณที่มีความเอนเอียง ยกเว้น เฉพาะกรณีการสูญหายของผลลัพธ์ไม่มีความสัมพันธ์กับตัวแปรร่วม วิธีค่าเฉลี่ย และ วิธีการใส่ค่าเพียงค่าเดียวจะให้ค่า ความแปรปรวนต่ำกว่าความเป็นจริง ส่วนวิธีเอ็มไอ จะให้ค่าเฉลี่ยของประชากรไม่มีความเอนเอียง

ทรอยันสกายะและคณะ (Troyanskaya; et al. 2001: 520 - 525) ศึกษาเปรียบเทียบวิธีประมาณค่าสูญหายของยีนในDNA เป็นข้อมูลไมโครอาร์เรย์ โดยเปรียบเทียบวิธีประมาณค่า 3 วิธี คือ วิธี Singular Value Decomposition (SVD) วิธี K – nearest neighbor algorithm (KNN impute) และ Row Average ซึ่งเปอร์เซ็นต์การสูญหายอยู่ในช่วง 1 - 20% เกณฑ์ที่ใช้เปรียบเทียบคือ RMS Error ประมาณค่าข้อมูลสูญหายทั้ง 3 วิธี ภายใต้เปอร์เซ็นต์การสูญหายต่างกัน ผลการศึกษาพบว่า วิธี KNN impute มีประสิทธิภาพในการประมาณค่าดีกว่าวิธี Singular Value Decomposition (SVD) และ Row Average

ฮุวอง และแคร์เรียริ (Huang; & Carriere 2006: 235 – 247) ศึกษาเปรียบเทียบวิธีการจัดการข้อมูลสูญหายในข้อมูลแบบวัดซ้ำในกลุ่มตัวอย่างขนาดเล็กโดยมีวัตถุประสงค์เพื่อพัฒนาวิธีเอ็มไอของข้อมูลแบบวัดซ้ำในกลุ่มตัวอย่างขนาดเล็ก กระทำภายใต้ข้อตกลงเบื้องต้นข้อมูลมีการแจกแจงปกติหลายตัวแปร (multivariate normal distribution) ทำการศึกษาโดยใช้สถานการณ์จำลอง (simulation) และเปรียบเทียบวิธีการจัดการข้อมูลที่สูญหายระหว่างวิธีไม่ใส่ค่าแทนข้อมูลที่สูญหาย โดยใช้วิธีเม็กซิมั่มไลลิฮูด กับวิธีใส่ค่าแทนข้อมูลที่สูญหายด้วย วิธีเอ็มไอ (ใช้จำนวนค่าที่ใส่แทนข้อมูลที่สูญหาย M = 5) เพื่อทดสอบสมมติฐานของอิทธิพลของทรีทเมนต์ และอิทธิพลของความคลาดเคลื่อนของทรีทเมนต์ ผลการศึกษาปรากฏว่า วิธีการใส่ค่าแทนข้อมูลที่สูญหายด้วยวิธีเอ็มไอจะให้ผลการทดสอบไม่แตกต่างจากวิธีไม่ใส่ค่าแทนข้อมูลที่สูญหายโดยวิธีเม็กซิมั่มไลลิฮูด

จากการศึกษางานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศ เพื่อให้เห็นภาพรวมของสิ่งที่เกี่ยวข้องกับการศึกษาครั้งนี้ ผู้วิจัยจึงสรุปดังตาราง 14 ซึ่งมีรายละเอียดดังนี้

ตาราง 14 สรุปงานวิจัยที่เกี่ยวข้องทั้งในประเทศ และต่างประเทศ

ลำดับ	ชื่อวิทยานิพนธ์	วิธีประมาณค่าสูญหาย	ลักษณะการจำลองข้อมูล	เกณฑ์การเปรียบเทียบ	ประเด็นศึกษาต่อ
1	ศึกษาวิธีเปรียบเทียบวิธีวิเคราะห์เมื่อมีค่าสูญหายในบล็อคใดบล็อคหนึ่งในแผนการทดลองบล็อคไม่สมบูรณ์ (ปราณี ศรีภา : 2532) จุฬาลงกรณ์มหาวิทยาลัย	- PURI - LSD - Wilkinson	- แผนการทดลอง : BIB - จำนวนข้อมูลสูญหาย : 1 บล็อก - จำนวนทรีตเมนท์ : 4,6,7,8,9,10 - จำนวนบล็อก : 4,7,10,12,15,18 - จำนวนซ้ำ : 3,4,5,6,9 - ระดับนัยสำคัญ : 0.01, 0.05, 0.10 - การแจกแจง : แบบปกติ - การแจกแจงของความคลาดเคลื่อน : 1. แบบปกติแบบโลจิสติค 2. แบบดับเบิ้ลเอ็กซ์โปเนนเชียล 3.แบบปกติปลอมปน	ค่าความคลาดเคลื่อนเฉลี่ย	การวิจัยครั้งนี้ศึกษาเฉพาะสถานการณ์ที่กำหนด และการสูญหายของข้อมูลมีเพียง 1 บล็อกเท่านั้น สิ่งที่น่าสนใจต่อไปคือถ้าข้อมูลสูญหายมากกว่านี้จะให้ผลอย่างไร
2	การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย (ชุติมา ชัยมุสิก : 2533) จุฬาลงกรณ์มหาวิทยาลัย	- วิธีการถดถอย - Maximum likelihood - วิธีค่าเฉลี่ย - วิธีค่ามัธยฐาน	- กลุ่มตัวอย่าง : 30,70,100 - C.V. : 0.05,0.20,1.00 - ตัวแปรอิสระ 4 ระดับ : 2,3,5,7 - ค่าเบี่ยงเบนมาตรฐาน : 5,10,20,25 - สัดส่วนข้อมูลสูญหาย: 5%,10%,15%	ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE)	ในการวิจัยต่อๆไป ควรศึกษาในกรณีที่ข้อมูลไม่ได้มีการแจกแจงแบบปกติและควรคำนึงถึงความสัมพันธ์ระหว่างตัวแปรอิสระและสัดส่วนการสูญหายของข้อมูลไม่เท่ากันทุกตัวแปร และควรจะนำค่าสัมประสิทธิ์ความถดถอยที่ได้จากสมการถดถอยที่ประมาณด้วยวิธีต่างๆ เปรียบเทียบกับวิธีที่ไม่มีข้อมูลสูญหาย
3	การพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุเมื่อตัวแปรตามมีค่าสูญหาย (วารุณี ตรีบำรุง : 2538) จุฬาลงกรณ์มหาวิทยาลัย	- วิธีค่าเฉลี่ย - วิธีการถดถอย - EM algorithm - Hunt's method	- กลุ่มตัวอย่าง : 10,20,30,50,70 - ค่าเบี่ยงเบนมาตรฐาน : 5,10,15,20,25 - สัดส่วนข้อมูลสูญหาย: 10%,20%,30%,40%,50%,60%, 70%	รากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนกำลังสองของค่าพยากรณ์	- สำหรับกรณีที่ความคลาดเคลื่อนของข้อมูลมีการแจกแจงรูปแบบอื่น วิธีการเหล่านี้อาจจะไม่มีประสิทธิภาพ จึงควรศึกษาวิจัยในปัญหาดังกล่าว - ในการวิจัยครั้งต่อไปอาจจะทำการศึกษาในกรณีของการวิเคราะห์การถดถอยที่ไม่ใช่เชิงเส้น(non-linear regression)

ตาราง 14 (ต่อ)

ลำดับ	ชื่อวิทยานิพนธ์	วิธีประมาณค่าสูญหาย	ลักษณะการจำลองข้อมูล	เกณฑ์การเปรียบเทียบ	ประเด็นศึกษาต่อ
4	การพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและลิสไวท์ (เชาว์ อินใย : 2540) มหาวิทยาลัยนเรศวร	- Listwise deletion - EM algorithm or Expectation maximization - Estimated parameter and Smallest standard error	- ความสัมพันธ์ระหว่างตัวแปร (r) : .30, .50, .70 - จำนวนข้อมูลสูญหาย: 5%,10%,20%,30% - วิธีการสุ่ม 1. การสุ่มแบบแบ่งชั้น 2. การสุ่มแบบกลุ่ม 3. การสุ่มแบบหลายขั้นตอน - ข้อมูลมีการแจกแจงแบบปกติ 2 ตัวแปร	ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนระหว่างค่าสถิติที่ได้จากวิธีการจัดการข้อมูลสูญหาย วิธีสุ่มตัวอย่าง ความสัมพันธ์ระหว่างตัวแปร และจำนวนข้อมูลสูญหายที่แตกต่างกันกับค่าพารามิเตอร์	1. ควรศึกษาถึงความคลาดเคลื่อนในการสร้างสมการถดถอยแล้วหาวิธีนำมารวมในสมการเพื่อแทนค่าได้ถูกต้องมากยิ่งขึ้น 2. ควรเปรียบเทียบข้อมูลที่ใช้ในงานวิจัยครั้งนี้กับการสูญหายแบบเป็นระบบ 3. ควรใช้จำนวนกลุ่มตัวอย่าง 200,300,400,500,600 ถึง 1,000 โดยกำหนด r=.70 4. ควรศึกษาวิธีการจัดการข้อมูลสูญหายกับค่าสถิติอื่นๆ
5	การประมาณค่าสูญหายจากแผนการทดลอง (รุ่งกานต์ กาใจคำ : 2540) มหาวิทยาลัยเชียงใหม่	- LSD - Mean method - กระบวนการ R	- การแจกแจง : แบบปกติ - การแจกแจงของความคลาดเคลื่อน : แบบปกติ	ค่าประมาณจากวิธีใช้กระบวน R เปรียบเทียบกับวิธีใช้ค่าเฉลี่ย	ในการศึกษาครั้งนี้ศึกษาการประมาณค่าข้อมูลสูญหายในแผนการทดลองแบบบวกจัดสองทาง และแผนทดลองแบบบวกจัดสามทางเท่านั้น ดังนั้นควรทำการศึกษากรณีเกิดข้อมูลสูญหายในแผนทดลองแบบบวกจัดหลายทางต่อไปได้อีก นอกจากนี้ควรประยุกต์กับโปรแกรมคอมพิวเตอร์ก็จะเป็นการสะดวก และง่ายต่อการใช้งานยิ่งขึ้น

ตาราง 14 (ต่อ)

ลำดับ	ชื่อวิทยานิพนธ์	วิธีประมาณค่าสูญหาย	ลักษณะการจำลองข้อมูล	เกณฑ์การเปรียบเทียบ	ประเด็นศึกษาต่อ
6	การวิเคราะห์ความแปรปรวนแบบปัจจัยเดียวเมื่อมีค่าสูญหายในแผนการทดลอง (ดวงฤดี เหง้าพรมมิล : 2545) มหาวิทยาลัยขอนแก่น	- Wilkinson - IM Iterative method - General regression significant test - Analysis of covariance Das and Giri	- แผนการทดลอง : RCB - จำนวนข้อมูลสูญหาย : 1,2,3 ค่า - จำนวนทรีตเมนท์ : 3,5,7 - จำนวนบล็อก : 3,5,7,9 - ระดับนัยสำคัญ : 0.01, 0.05 - การแจกแจง : แบบปกติ - การแจกแจงของความคลาดเคลื่อน : 1. แบบปกติ 2. แบบที 3. แบบลอกนอร์มอล	1. ความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก Ho เมื่อ Hoเป็นจริง 2.ความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก Ho เมื่อ Hoไม่เป็นจริง 3.ร้อยละของจำนวนครั้งที่วิเคราะห์	1.ควรศึกษาเพิ่มเติมในแผนการทดลองแบบอื่นๆ 2.ควรทำการศึกษาเปรียบเทียบทั้ง 5 วิธีสำหรับสถานการณ์อื่นๆ เช่น ขนาดจำนวนบล็อก จำนวนสิ่งทดลอง จำนวนค่าสูญหายมากขึ้น 3.ในแต่ละสถานการณ์อาจทำการทดลองซ้ำมากกว่า 1,000 รอบ เพื่อให้ได้ข้อสรุปที่ชัดเจนยิ่งขึ้น 4.เมื่อมีค่าสูญหาย 1 ค่า ควรเลือกใช้วิธีIm หรือวิธีWm แต่เมื่อมีค่าสูญหายเท่ากับ 2,3 ค่า ควรเลือกใช้วิธีRm,วิธีAm, วิธีDm เมื่อความคลาดเคลื่อนของการทดลองมีการแจกแจงแบบปกติและที ส่วนความคลาดเคลื่อนของการทดลองมีการแจกแจงแบบลอกนอร์มอล ควรเลือกใช้วิธี Im หรือวิธี Wm
7	การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มบล็อคสมบูรณ์ (ประพจน์ ดำรงค์สุทธิพงศ์ : 2546) จุฬาลงกรณ์มหาวิทยาลัย	- LSD - Imputation - EM algorithm	- แผนการทดลอง : RCB - จำนวนข้อมูลสูญหาย : 10%,20%,30% - จำนวนทรีตเมนท์ : 3,4,5 - จำนวนบล็อก : 2,4,6 - การแจกแจง : แบบปกติ - การแจกแจงของความคลาดเคลื่อน : แบบปกติ	ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE)	1.การศึกษาครั้งต่อไปอาจจะศึกษาวิธีประมาณค่าสูญหายแบบอื่นๆอีก 2.การศึกษาครั้งต่อไปอาจจะศึกษาตัวแบบอื่นๆ เช่น ตัวแบบจตุรัสลาติน

ตาราง 14 (ต่อ)

ลำดับ	ชื่อวิทยานิพนธ์	วิธีประมาณค่าสูญหาย	ลักษณะการจำลองข้อมูล	เกณฑ์การเปรียบเทียบ	ประเด็นศึกษาต่อ
8	การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน (ศุภลักษณ์ กรรณิกาพงศ์ : 2549) จุฬาลงกรณ์มหาวิทยาลัย	- LSD - EM algorithm - MI Multiple imputation	- แผนการทดลอง : LS - จำนวนข้อมูลสูญหาย : 10%,20%,30% - จำนวนทรีตเมนท์ : 3,4,5,6,7 - จำนวนปัจจัยแถว : 3,4,5,6,7 - จำนวนปัจจัยคอลัมน์ : 3,4,5,6,7 - การแจกแจง : แบบปกติ - การแจกแจงของความคลาดเคลื่อน : แบบปกติ	ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) มีค่าต่ำสุด	1. ศึกษาวิธีประมาณค่าสูญหายแบบอื่นๆ เพื่อให้เหมาะสมกับแผนการทดลองแบบจตุรัสลาติน 2.ทำการศึกษาในตัวแบบอื่นๆ
9	การศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ (จริยา แสงสุวรรณ : 2550) มหาวิทยาลัยเกษตรศาสตร์	- loss method - Mean method - Regression method - Multiple imputation method	- ข้อมูลมีการแจกแจงแบบปกติหลายตัวแปร - ระดับความสัมพันธ์ (r) : .20, .50, .70 - การแจกแจงของความคลาดเคลื่อน : แบบปกติ - เปอร์เซ็นต์การสูญหาย : 5, 10, 20 , 30 - ขนาดตัวอย่าง : 50, 70, 100, 200	ค่าประมาณของรากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE)	1.ศึกษาวิธีประมาณค่าสูญหายกรณีการวิเคราะห์การถดถอยของข้อมูลลักษณะเชิงคุณภาพ 2. ศึกษาในกรณีที่ตัวแปรอิสระมีจำนวนมากกว่า 3 ตัวแปร 3. ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายด้วยวิธีการอื่นๆ นอกเหนือจากที่ใช้ศึกษาในครั้งนี้
10	การเปรียบเทียบวิธีการประมาณค่าสูญหายแบบนอนอิกนอร์เรเบิล ในการวิเคราะห์การถดถอยเชิงเส้นพหุ (อุษณีย์ วงศ์อามาตย์: 2555) จุฬาลงกรณ์มหาวิทยาลัย	- วิธี EM Algorithm (EM) - วิธี K-Nearest Neighbor Imputation (KNN) - วิธี Predictive Mean Matching Imputation (PMM)	- การสูญหายแบบนอนอิกนอร์เรเบิล - สัดส่วนข้อมูลสูญหาย : 10%,20%,30% - ระดับของการสูญหาย 3 ระดับคือ ไม่มี, ปานกลาง, สูง	ค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (average mean square error; AMSE)	1.วิธี EM ประมาณค่าได้ดีที่สุดเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนมีค่าไม่สูง (10-30) 2.วิธี KNN ประมาณค่าได้ดีที่สุดเมื่อส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนมีค่าสูง (90)

ตาราง 14 (ต่อ)

ลำดับ	ชื่อวิทยานิพนธ์	วิธีประมาณค่าสูญหาย	ลักษณะการจำลองข้อมูล	เกณฑ์การเปรียบเทียบ	ประเด็นศึกษาต่อ
11	Comparison of Analytic Method for Non-Random Missingness of Outcome Data. (Crawford et al. : 1995) Journal of Clinical Epidemiology	- วิธีการใส่ค่าเพียงค่าเดียว - Mean method - MI Multiple imputation - loss method	- ข้อมูลมาจากโครงการสุขภาพคนชราของมลรัฐแมซซะ-ซูเซ็ท	ค่าประมาณที่มีความเอนเอียง	ศึกษาการแจกแจงข้อมูลรูปแบบอื่นๆ และวิธีการประมาณค่าด้วยวิธีอื่น
12	Missing value estimation methods for DNA microarrays (Troyanskaya et al. : 2001) Bioinfomatics.	- Singular Value Decomposition (SVD) - K – nearest neighbor (KNN) - Row average	- ข้อมูลยีนส์ DNA เป็นแบบไมโครอาร์เรย์ - เปอร์เซ็นต์ข้อมูลสูญหายอยู่ในช่วง 1 – 20%	RMS Error	วิธี KNN มีความแม่นยำและมีความละเอียดอ่อนต่อการประมาณค่าข้อมูลสูญหายสำหรับ microarrays
13	Comparison of Methods for Incomplete Repeated Measures Data Analysis in Small Samples. (Huang and Carriere : 2006) Statistical Planning and Inference.	- Maximum likelihood - MI Multiple impution	- กลุ่มตัวอย่างขนาดเล็ก - ข้อมูลมีการแจกแจงปกติหลายตัวแปร	อิทธิพลของทรีทเมนต์ และอิทธิพลของความคลาดเคลื่อนของทรีทเมนต์	1. ควรศึกษาข้อมูลที่มีการแจกแจง non-normal 2. ควรศึกษาข้อมูลสูญหายด้วย non-ignorable

จากการศึกษางานวิจัยในประเทศและต่างประเทศ พบว่า งานวิจัยเกี่ยวกับการประมาณค่าสูญหายในแผนการทดลองนั้น ส่วนใหญ่จะศึกษาในแบบแผนการทดลองแบบปัจจัยเดียว (Singer factor experiment) เช่น แผนการทดลองแบบสุ่มตลอด (CRD) แผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (RCB) แผนการทดลองแบบจัตุรัสลาติน (LS) แผนการทดลองแบบบล็อกไม่สมบูรณ์สมดุล (Balanced Incomplete Block Design : BIB) ซึ่งแผนการทดลองนี้มีการจัดบล็อกและทรีตเมนต์มาจากแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ (RCB) และแผนการทดลองแบบจัตุรัสละติน (LS)เป็นส่วนใหญ่ แต่ยังไม่พบว่ามีการศึกษาเปรียบเทียบวิธีประมาณค่าสูญหายในแผนการทดลองแบบแฟคทอเรียล ซึ่งเป็นที่นิยมใช้ในการวางแผนการทดลอง เพราะสามารถศึกษาปัจจัยหลายปัจจัยในการทดลองเดียวกันได้ ลดทุนในการทดลอง และสะดวกต่อการปฏิบัติ นอกจากแผนการทดลองแล้วยังมีตัวแปรอื่นๆ เข้ามาเกี่ยวข้องทำให้ผลของการประมาณค่าสูญหายแตกต่างกัน เช่น จำนวนปัจจัย, จำนวนบล็อก, จำนวนข้อมูลสูญหาย, ค่าสัมประสิทธิ์ความผันแปร, ความแปรปรวนของความคลาดเคลื่อนของการทดลองและวิธีการประมาณค่าสูญหาย ดังนั้น จึงควรที่จะศึกษาตัวแปรอื่นๆ ด้วย ซึ่งจะทำให้ข้อค้นพบที่สามารถนำไปประยุกต์ใช้ได้อย่างมีประสิทธิภาพ

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาวิธีประมาณค่าสูญหาย และเปรียบเทียบประสิทธิภาพของวิธีประมาณประมาณค่าสูญหาย ในการวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) โดยทำการเปรียบเทียบวิธีการประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยวิธีวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน ทำการประมาณค่าสูญหายจากข้อมูลที่ได้ และทำการเปรียบเทียบว่าวิธีการใดให้ค่าประมาณใกล้เคียงกับค่าจริงมากที่สุด โดยจะพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด วิธีการประมาณค่าสูญหายวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด มีค่าต่ำที่สุดเป็นวิธีที่ดีที่สุด แสดงว่าวิธีนั้นเหมาะสมสำหรับนำมาใช้ในการประมาณค่าสูญหายมากที่สุด ซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอนดังต่อไปนี้

1. การกำหนดประชากรและกลุ่มตัวอย่าง
2. การจำลองข้อมูลด้วยโปรแกรม R
   - 2.1 การจำลองข้อมูลให้มีการแจกแจงแบบปกติ
   - 2.2 การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง
   - 2.3 การตรวจสอบรูปแบบของข้อมูลสูญหาย
   - 2.4 การประมาณค่าข้อมูลสูญหาย
   - 2.5 การเปรียบเทียบวิธีประมาณค่าสูญหาย
3. ขั้นตอนการทำงานของโปรแกรม R

การกำหนดประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลที่ได้จากการจำลองสถานการณ์การทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน ข้อมูลที่ได้เป็นข้อมูลที่เกิดจากการจำลองด้วยรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย เพื่อให้ได้ข้อมูลใกล้เคียงกับจำนวนจริงที่ใช้ในการปฏิบัติทดลองในการแผนการทดลองมากที่สุด ผู้วิจัยจึงกำหนดให้ประชากรมีขนาดเท่ากับ 512 สำหรับเทคนิคมอนติ

คาร์โลซิมูเลชัน เป็นการจำลองตัวแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน เขียนด้วยโปรแกรม R

กลุ่มตัวอย่าง

ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน จำแนกตามปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ และจำนวนบล็อกเท่ากับ 3,4 และ5 บล็อก เพราะฉะนั้นจำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 รายละเอียดขนาดตัวอย่างดังภาพประกอบ 2 สุ่มตัวอย่างให้มีขนาดตามแต่ละสถานการณ์ การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 1,000 รอบ ซึ่งมีความเพียงพอที่จะศึกษาค่าประมาณพารามิเตอร์

การจำลองข้อมูลด้วยโปรแกรม R

การสร้างข้อมูลในงานวิจัยครั้งนี้ ได้จากการจำลองสถานการณ์ข้อมูลการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์(RCB) ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน ผู้วิจัยทำการจำลองข้อมูลด้วยโปรแกรม R ดำเนินการตามขั้นตอน ดังนี้

1. การจำลองข้อมูลให้มีการแจกแจงแบบปกติ
2. การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง
3. การตรวจสอบรูปแบบของข้อมูลสูญหาย
4. การประมาณค่าข้อมูลที่สูญหาย
5. การเปรียบเทียบวิธีประมาณค่าสูญหาย

1. การจำลองข้อมูลให้มีการแจกแจงแบบปกติ

การจำลองข้อมูลประชากรให้มีการแจกแจงแบบปกติ ด้วยเทคนิคมอนติคาร์โล ซึ่งใช้ในการจำลองตัวแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน โดยหลักของมอนติคาร์โลนั้น ต้องจำลองตัวเลขสุ่ม (Random Number) ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบันแบ่งได้เป็น 2 ขั้นตอน ดังนี้

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในวิธีมอนติคาร์โล เนื่องจากหลักของมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา

2. การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่มโดยตรงนั้น มีบางขั้นตอนที่จะต้องใช้ตัวเลขสุ่ม การเขียนโปรแกรมในงานวิจัยครั้งนี้ ใช้โปรแกรม R ซึ่งการสร้างการแจกแจง

แบบปกติจะใช้ตัวเลขสุ่มในฟังก์ชัน rnorm กำหนดให้ค่าเฉลี่ย(μ) เท่ากับ 50 (เลิศลักษณ์ กลิ่นหอม. 2532: 84; และ ศุภลักษณ์ กรรณิกา. 2549: 4) สำหรับกรณีที่ค่าเฉลี่ยและความแปรปรวนของประชากรที่มีค่าอื่นผลสรุปที่ได้จะเหมือนกัน กล่าวคือ ค่าสถิติที่คำนวณได้จะมีค่าเท่ากันไม่ว่าจะศึกษา ณ จุดที่มีค่าเฉลี่ยและความแปรปรวนค่าใดๆ (สมชัย ยืนนาน. 2528: 30 – 32) โดยมีรายละเอียดในการแจกแจงแบบปกติดังนี้

กำหนดให้

$$Y_{ijk} = \mu + B_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

และ $B_k, \alpha_i, \beta_j, (\alpha\beta)_{ij}, \ \varepsilon_{ijk}$ เป็นตัวแปรสุ่มที่อิสระจากกัน และมีการแจกแจงแบบปกติที่มี $E(B_k) = E(\alpha_i) = E(\beta_j) = E(\alpha\beta)_{ij} = E(\varepsilon_{ijk}) = 0$

และ $Var(B_k) = \sigma_B^2, Var(\alpha_i) = \sigma_\alpha^2, Var(\beta_j) = \sigma_\beta^2, Var(\alpha\beta)_{ij} = \sigma_{\alpha\beta}^2, Var(\varepsilon_{ijk}) = \sigma_\varepsilon^2$

โดยที่ $\sigma_B^2 = \sigma_\alpha^2 = \sigma_\beta^2 = \sigma_{\alpha\beta}^2 = h\sigma_\varepsilon^2$

ซึ่ง h เป็นค่าจำนวนเต็มคงที่ ดังนั้นจะได้ค่า Y_{ijk} ซึ่งเป็นค่าข้อมูลในการทดลองนั้นๆ

2. การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง

ลักษณะของข้อมูลสูญหายในการวิจัยครั้งนี้ จะกำหนดเป็นข้อมูลสูญหายแบบสุ่ม (Randomly missing data) หมายถึง ค่าของข้อมูลสูญหายที่เกิดขึ้นไม่ได้ขึ้นอยู่กับข้อมูลค่าใดค่าหนึ่ง เมื่อข้อมูลถูกสร้างเสร็จเรียบร้อยแล้วในขั้นตอนนี้จะทำการสุ่มตัดข้อมูลออก และข้อมูลที่ถูกตัดออกไปนั้นจะเก็บไว้เพื่อนำมาเปรียบเทียบกับค่าใหม่ที่ประมาณขึ้นมาทั้ง 3 วิธี

3. การตรวจสอบรูปแบบของข้อมูลสูญหาย

ข้อมูลที่ถูกสุ่มตัดออกไปแล้วนั้น จะต้องนำข้อมูลมาตรวจสอบรูปแบบของข้อมูลสูญหาย ผู้วิจัยขอยกตัวอย่างเป็นการข้อมูลจำลองในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3 ระดับและจำนวนบล็อกเท่ากับ 3 เปอร์เซ็นต์ข้อมูลสูญหายเป็น 10%, 20% และ30% ดังนี้

ตัวอย่างที่ 1 ข้อมูลสูญหาย 10%

```
> Y
         [,1]     [,2]     [,3]
 [1,] 48.35157 49.24147 60.98875
 [2,] 47.27535 64.49631 61.77155
 [3,] 56.19844 46.15048 47.75705
 [4,] 45.43503 47.72569 38.93998
 [5,] 48.19986 49.36460 46.93178
 [6,] 51.30779 56.13313 59.86041
 [7,] 52.33343 52.21921 51.46942
 [8,] 41.24546 54.56281 57.82126
 [9,] 54.40819 61.78886 46.92837
> YM
         [,1]     [,2]     [,3]
 [1,] 48.35157       NA 60.98875
 [2,] 47.27535 64.49631 61.77155
 [3,] 56.19844 46.15048 47.75705
 [4,] 45.43503 47.72569 38.93998
 [5,] 48.19986 49.36460 46.93178
 [6,]       NA 56.13313 59.86041
 [7,] 52.33343 52.21921 51.46942
 [8,] 41.24546 54.56281 57.82126
 [9,] 54.40819       NA 46.92837
> |
```

ภาพประกอบ 3 ตรวจสอบรูปแบบข้อมูลสูญหาย 10%

ตัวอย่างที่ 2 ข้อมูลสูญหาย 20%

```
> Y
         [,1]     [,2]     [,3]
 [1,] 52.18102 47.14988 45.88542
 [2,] 52.11786 48.80443 46.88406
 [3,] 45.93622 51.61679 49.10109
 [4,] 51.33742 46.09356 48.01473
 [5,] 48.58472 49.35247 46.48615
 [6,] 47.41253 54.47138 47.92655
 [7,] 44.40100 50.83475 54.06642
 [8,] 41.58126 45.76691 45.87626
 [9,] 46.61429 49.52452 49.49537
> YM
         [,1]     [,2]     [,3]
 [1,]       NA 47.14988 45.88542
 [2,] 52.11786 48.80443       NA
 [3,] 45.93622 51.61679 49.10109
 [4,]       NA 46.09356 48.01473
 [5,] 48.58472 49.35247 46.48615
 [6,] 47.41253 54.47138       NA
 [7,] 44.40100 50.83475 54.06642
 [8,] 41.58126       NA 45.87626
 [9,] 46.61429 49.52452 49.49537
> |
```

ภาพประกอบ 4 ตรวจสอบรูปแบบข้อมูลสูญหาย 20%

ตัวอย่างที่ 3 ข้อมูลสูญหาย 30%

```
> Y
         [,1]     [,2]     [,3]
 [1,] 48.68822 44.30245 51.36115
 [2,] 47.49173 48.20268 54.69530
 [3,] 50.54759 53.36546 46.13910
 [4,] 47.61588 49.97376 50.38662
 [5,] 50.59504 47.48458 51.62465
 [6,] 51.61854 49.96907 48.74517
 [7,] 49.76113 54.86175 48.09409
 [8,] 47.20326 48.74770 52.31170
 [9,] 43.68088 48.71502 48.71960
> YM
         [,1]     [,2]     [,3]
 [1,] 48.68822       NA       NA
 [2,]       NA 48.20268 54.69530
 [3,] 50.54759 53.36546 46.13910
 [4,] 47.61588       NA 50.38662
 [5,] 50.59504 47.48458       NA
 [6,]       NA 49.96907 48.74517
 [7,] 49.76113 54.86175       NA
 [8,] 47.20326       NA 52.31170
 [9,] 43.68088 48.71502 48.71960
> |
```

ภาพประกอบ 5 ตรวจสอบรูปแบบข้อมูลสูญหาย 30%

จากการจำลองข้อมูลทั้ง 3 สถานการณ์ พบว่า ข้อมูลมีการสูญหายแบบสุ่ม ไม่สามารถที่จะระบุจุดที่ข้อมูลจะสูญหายได้ และอันดับของตัวแปรก็ไม่มีความสำคัญ เพราะฉะนั้น การจำลองข้อมูลงานวิจัยในครั้งนี้ ข้อมูลสูญหายจัดอยู่ในรูปแบบข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary)

4. การประมาณค่าข้อมูลที่สูญหาย

เมื่อสร้างข้อมูล Y_{ijk} ที่เป็นไปตามข้อกำหนดข้างต้นและได้ทำการสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริงเรียบร้อยแล้ว นำข้อมูลที่เหลือไปคำนวณหาค่าประมาณด้วยวิธีการประมาณค่าสูญหาย 3 วิธี การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 1,000 รอบ

5. การเปรียบเทียบวิธีประมาณค่าสูญหาย

เกณฑ์ที่ใช้ในการเปรียบเทียบว่าวิธีประมาณค่าสูญหายวิธีใดจะให้ค่าประมาณดีที่สุด จะพิจารณาจากการคำนวณหาค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error : MAE) หาค่าประมาณจากวิธีการต่างๆ และนำมาเปรียบเทียบกับค่าจริง ดังนี้

$$MAE = Max[Max|y_i - \hat{y}_i|]_n$$ (ศุภลักษณ์ กรรณิกา.2549: 10)

เมื่อ y_i แทน ค่าจริงที่ i ได้จากการจำลอง

$\hat{y}_i$ แทน ค่าประมาณ ค่าที่ i ที่ได้จากการใช้วิธีประมาณค่า

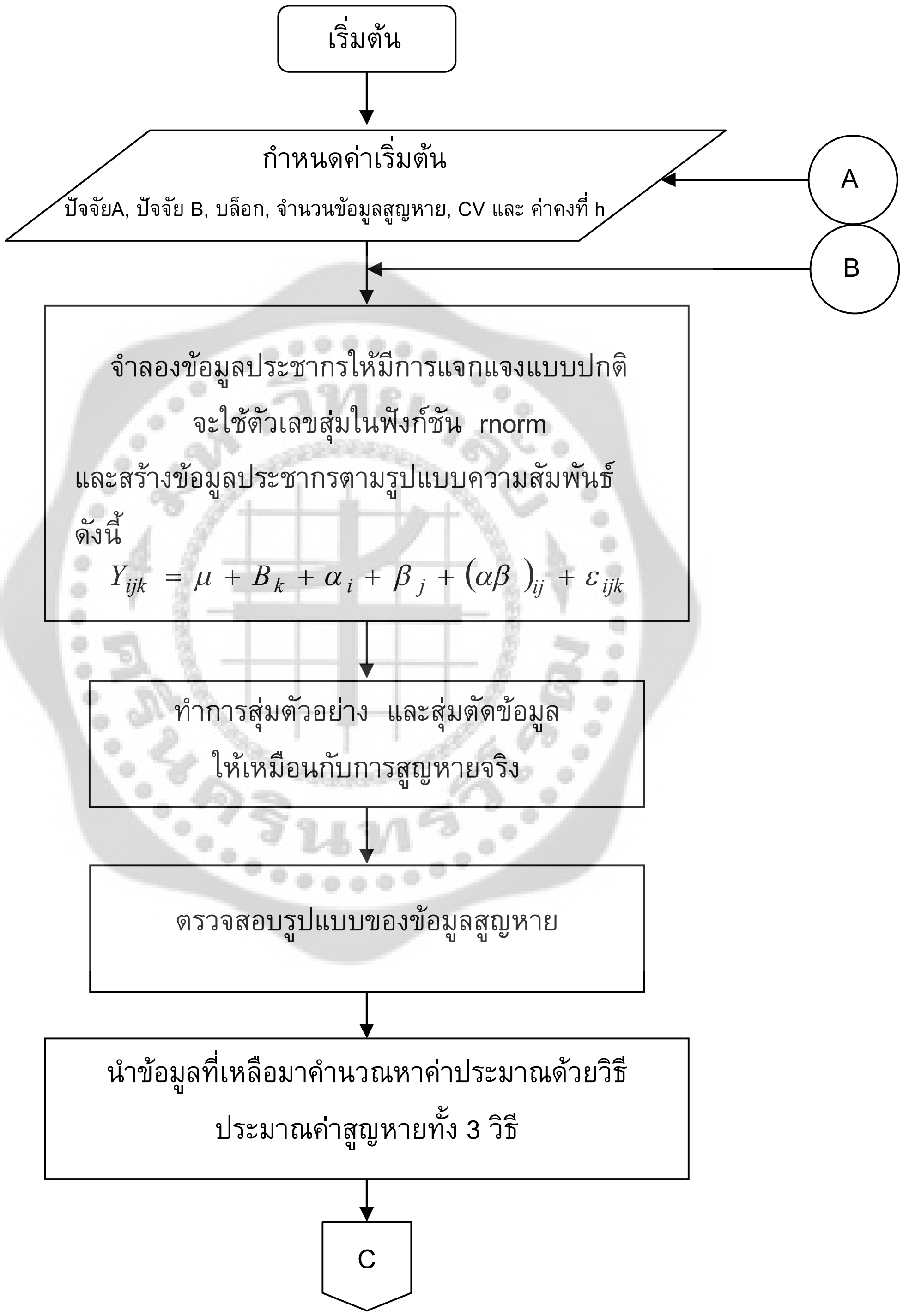
n แทน รอบที่ n จากการทดลอง (กำหนดให้ n=1000 รอบ)

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด คำนวณจากค่าประมาณและค่าจริงที่ได้จากการจำลอง กล่าวคือ ในการจำลองข้อมูล 1 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ระหว่างค่าจริงที่ได้จากการจำลองกับค่าที่ได้จากการประมาณค่าสูญหายจำนวน 1 ค่า การศึกษาในครั้งนี้ทำการจำลองข้อมูล 1,000 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ 1,000 ค่า พิจารณาเอาค่าที่มีค่ามากที่สุดใน 1,000 ค่า เป็นตัวแทนของค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ในแต่ละสถานการณ์และแต่ละวิธีมาเปรียบเทียบกัน วิธีใดให้ค่าความคลาดเคลื่อนสูงสุดมีค่าต่ำที่สุด แสดงว่า ค่าประมาณที่ได้มีค่าใกล้เคียงกับค่าจริงที่สูญหายไปมากที่สุด จะเป็นวิธีการประมาณค่าสูญหายที่ดีที่สุด

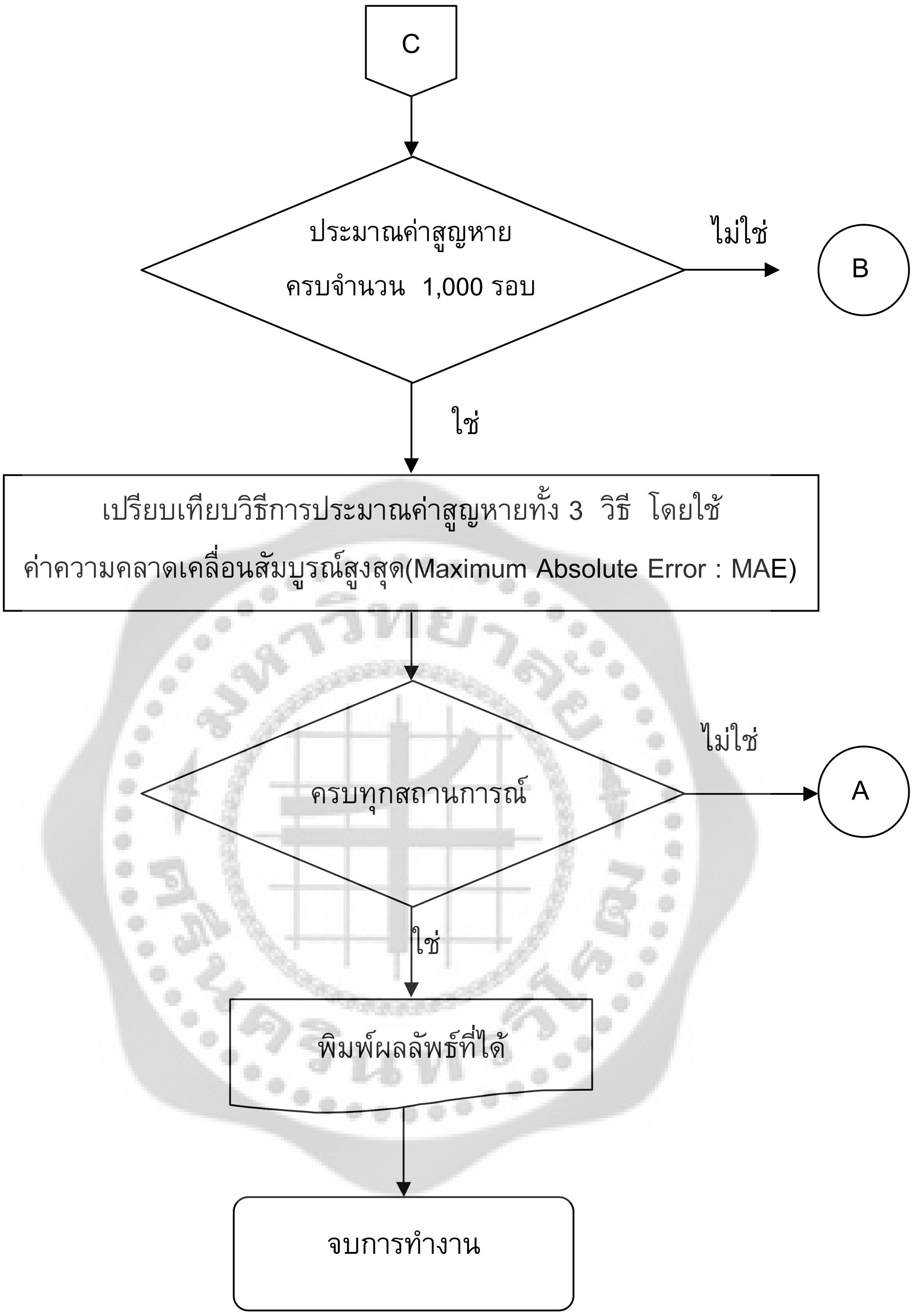
ขั้นตอนการทำงานของโปรแกรม

ขั้นตอนการเขียนโปรแกรมในครั้งนี้ แบ่งออกเป็น 2 ขั้นตอน คือ

1. ขั้นตอนการเขียนโปรแกรมหลัก ชื่อ Main Program



ภาพประกอบ 6 ผังแสดงขั้นตอนการเขียนโปรแกรมหลัก (Main Program)

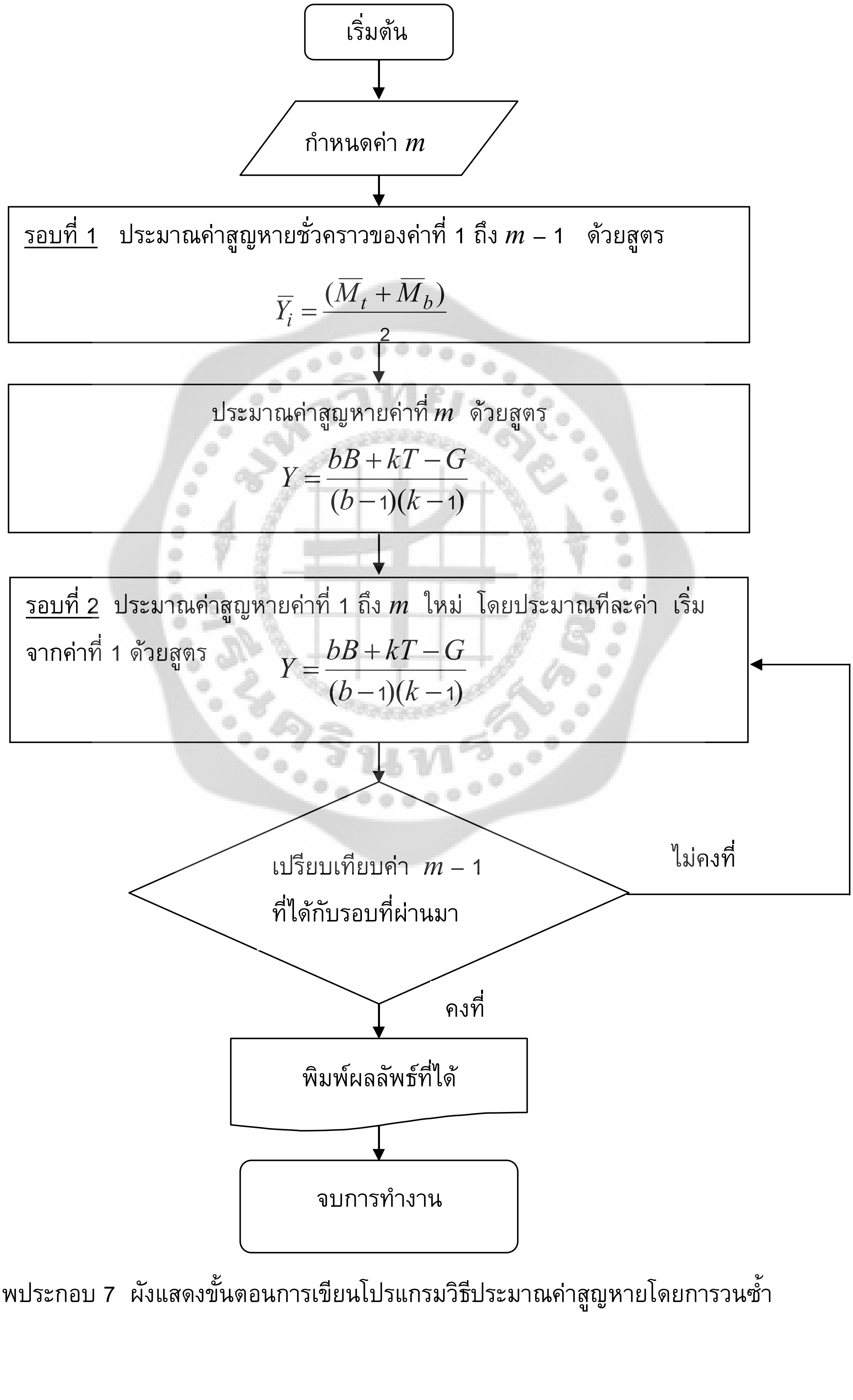


ภาพประกอบ 6 (ต่อ)

จากภาพประกอบ 6 ขั้นตอนการเขียนโปรแกรมหลัก เริ่มต้นโดยการกำหนดค่าเริ่มต้นให้กับตัวแปรอิสระ คือ ปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, จำนวนบล็อกเท่ากับ 3,4 และ 5 บล็อก, จำนวนข้อมูลสูญหาย มีค่าเท่ากับ 10%,20% และ 30%, ค่าสัมประสิทธิ์ความผันแปร มีค่าเท่ากับ 5%, 25% และ45% และค่าคงที่ เท่ากับ 1,2 และ 3 จากนั้นทำการจำลองข้อมูลประชากรให้มีการแจกแจงแบบปกติ จำลองโดยใช้เลขสุ่มในฟังก์ชัน rnorm และสร้างข้อมูลตามรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์แบบสุ่มในบล็อกสมบูรณ์ (RCB) ให้มีจำนวนเท่ากับ 512 ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ โดยจำแนกตามปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, จำนวนบล็อกเท่ากับ 3,4 และ 5 บล็อก เพราะฉะนั้นจำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 รายละเอียดขนาดตัวอย่างดังภาพประกอบ 2 เมื่อได้ขนาดตัวอย่างแล้วก็ทำการสุ่มตัดให้เหมือนกับการสูญหายจริงตามจำนวนข้อมูลสูญหาย 10%,20% และ30% ทำการตรวจสอบรูปแบบของข้อมูลสูญหาย เพราะในงานวิจัยครั้งนี้จะทำการจำลองข้อมูลสูญหายจัดอยู่ในรูปแบบข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) จากนั้นเก็บค่าที่สุ่มตัดไว้เพื่อจะนำมาเปรียบเทียบกับค่าที่ได้จากการประมาณ นำข้อมูลที่เหลือมาคำนวณหาค่าประมาณด้วยวิธีประมาณค่าทั้ง 3 วิธี ทำการเปรียบเทียบวิธีประมาณค่าสูญหายทั้ง 3 วิธี ในขั้นตอนนี้ต้องเข้าสู่โปรแกรมย่อยทั้ง 3 วิธี เปรียบเทียบโดยใช้ค่าความคลาดเคลื่อนสมบูรณ์สูงสุด ตรวจสอบว่าในแต่ละสถาการณ์ทำการจำลองข้อมูลครบ 1,000 รอบหรือไม่ ถ้าไม่ให้กลับไปจำลองข้อมูลใหม่ที่จุด B ถ้าใช่ก็ไปจำลองข้อมูลสถานการณ์ต่อไป ตรวจสอบสถานการณ์ว่าทำการจำลองข้อมูลครบทุกสถานการณ์หรือไม่ ถ้าไม่ให้กลับไปที่จุด A ถ้าใช่ก็ให้แสดงผลลัพธ์ที่ได้เป็นค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ที่มีค่าต่ำที่สุดของแต่ละวิธีจนครบทุกสถานการณ์ จบการทำงานของโปรแกรมหลัก

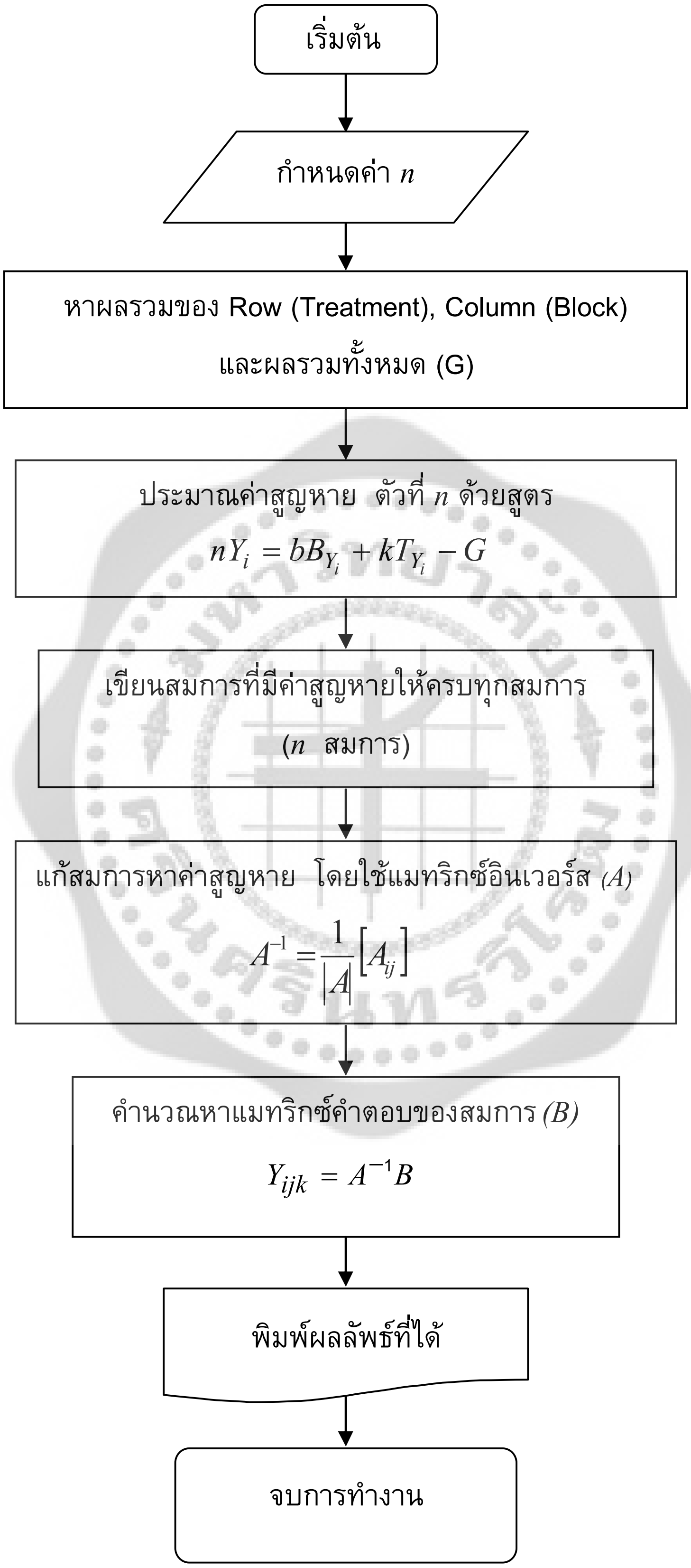
2. ขั้นตอนการเขียนโปรแกรมย่อย ดังนี้

2.1 ขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดยการวนซ้ำ ชื่อ Iterative Method



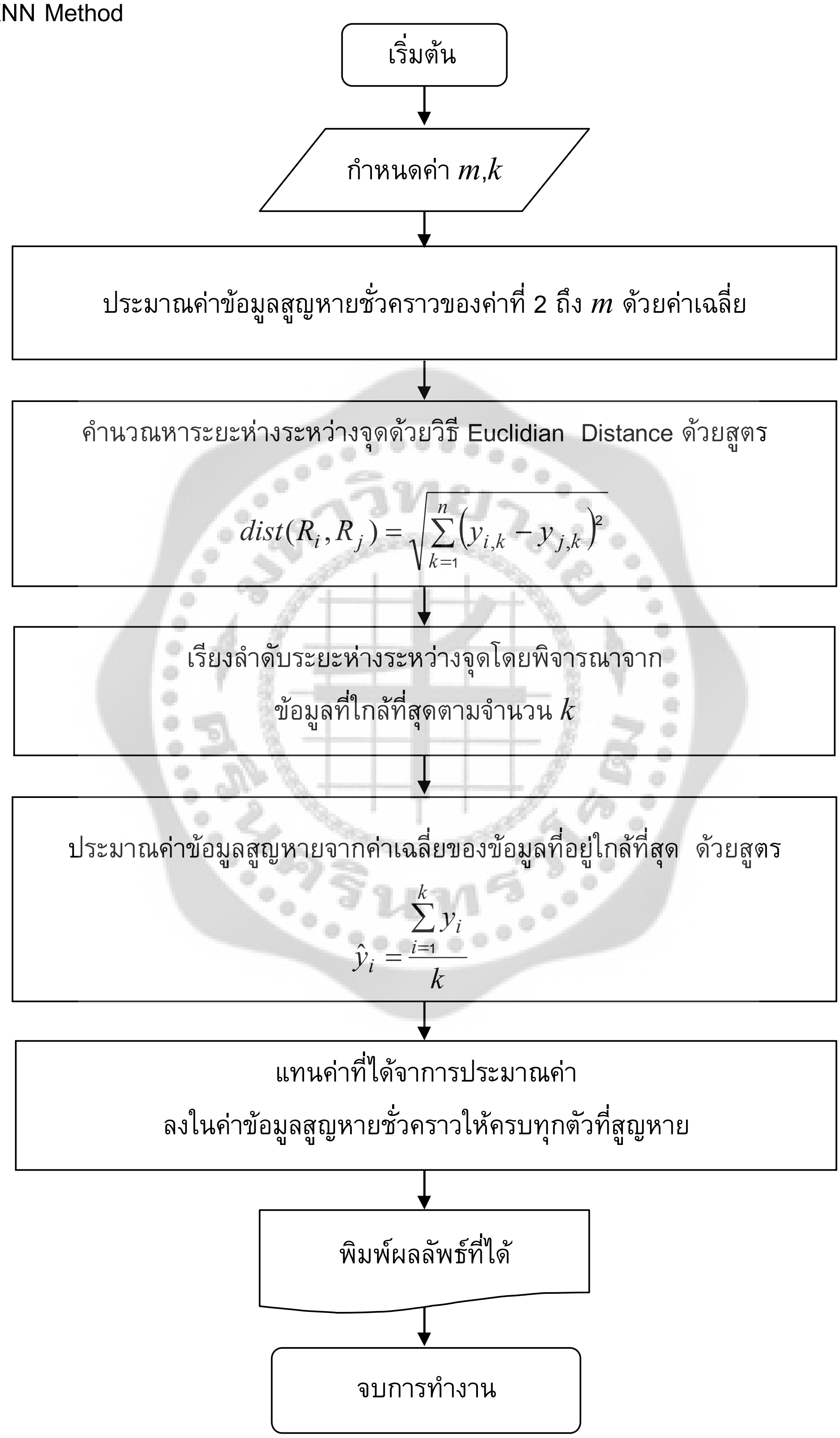
ภาพประกอบ 7 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดยการวนซ้ำ

2.2 ขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายของวิลคินซัล ชื่อ Wilkinson Method



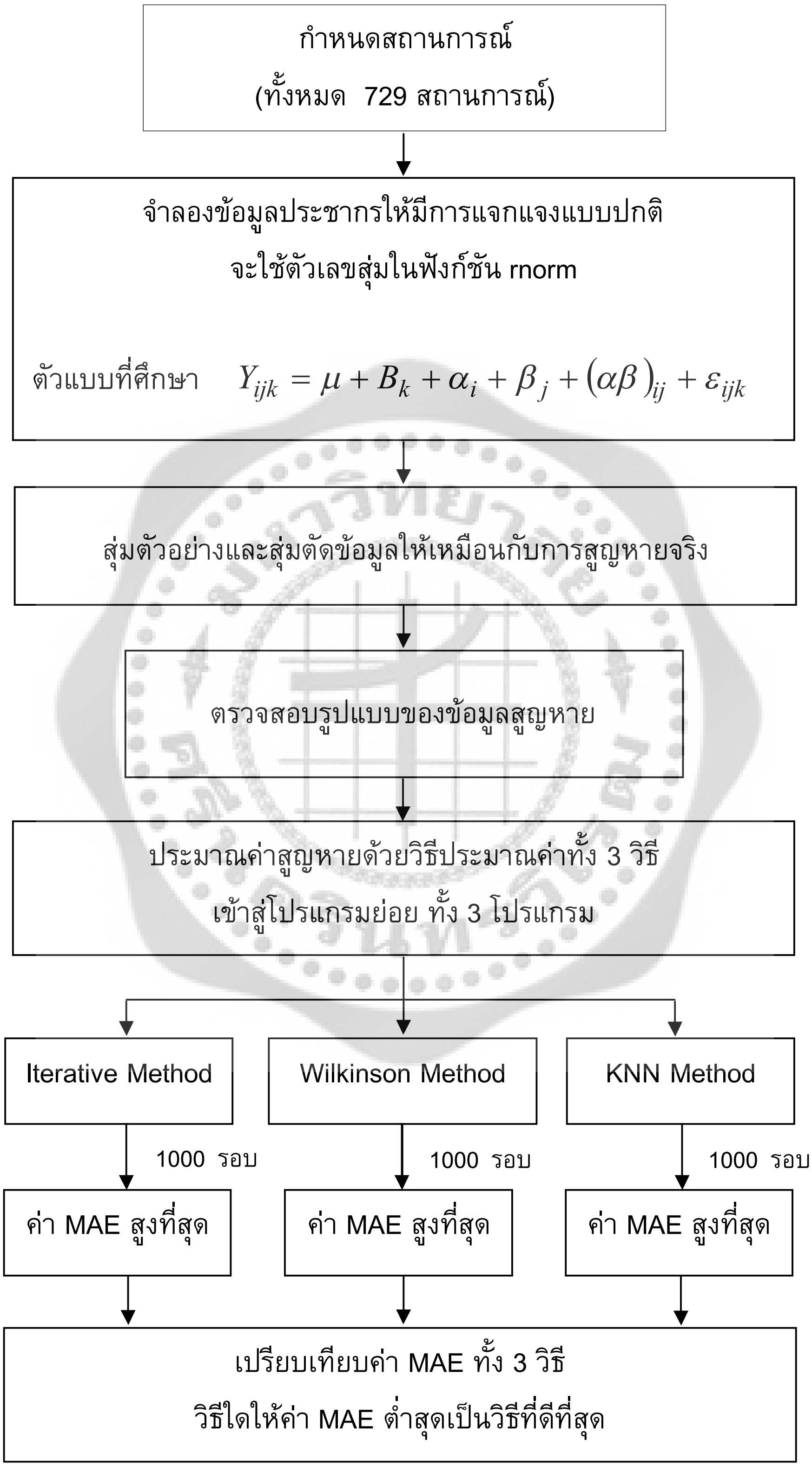
ภาพประกอบ 8 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายของวิลคินซัล

2.3 ขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน ชื่อ KNN Method



ภาพประกอบ 9 ผังแสดงขั้นตอนการเขียนโปรแกรมวิธีประมาณค่าสูญหายโดย K-Nearest Neighbor

ขั้นตอนการดำเนินงานวิจัย ผู้วิจัยสรุปเป็นแผนภาพได้ดังนี้



ภาพประกอบ 10 ผังแสดงสรุปขั้นตอนการดำเนินงานวิจัย

จากภาพประกอบ 10 สรุปขั้นตอนการดำเนินงานวิจัย เริ่มต้นโดยการกำหนดขนาดประชากรและกลุ่มตัวอย่าง ทั้งหมด 729 สถานการณ์ค่าเริ่มต้นให้กับตัวแปรอิสระ คือ ปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, จำนวนบล็อกเท่ากับ 3,4 และ5 บล็อก, จำนวนข้อมูลสูญหายมีค่าเท่ากับ 10%,20%และ30%, ค่าสัมประสิทธิ์ความผันแปรมีค่าเท่ากับ 5%, 25% และ45% และค่าคงที่ h เท่ากับ 1,2 และ3 จากนั้นทำการจำลองข้อมูลประชากรให้มีการแจกแจงแบบปกติ จำลองโดยใช้เลขสุ่มในฟังก์ชัน rnorm และสร้างข้อมูลตามรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์แบบสุ่มในบล็อกสมบูรณ์ (RCB) ให้มีจำนวนเท่ากับ 512 ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ โดยจำแนกตามปัจจัย A มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, ปัจจัย B มี 3 ลักษณะ คือ 3,4 และ5 ระดับ, จำนวนบล็อกเท่ากับ 3,4 และ 5 บล็อก เพราะฉะนั้นจำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 รายละเอียดขนาดตัวอย่างดังภาพประกอบ 2 เมื่อได้ขนาดตัวอย่างแล้วก็ทำกี่สุ่มตัดให้เหมือนกับการสูญหายจริงตามจำนวนข้อูลสูญหาย 10%,20% และ30% ทำการตรวจสอบรูปแบบของข้อมูลสูญหาย เพราะในงานวิจัยครั้งนี้จะทำการจำลองข้อมูลสูญหายจัดอยู่ในรูปแบบข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) จากนั้นเก็บค่าที่สุ่มตัดไว้เพื่อจะนำมาเปรียบเทียบกับค่าที่ได้จากการประมาณ นำข้อมูลที่เหลือมาคำนวณหาค่าประมาณด้วยวิธีประมาณค่าทั้ง 3 วิธี ทำการเปรียบเทียบวิธีประมาณค่าสูญหายทั้ง 3 วิธี ในขั้นตอนนี้ต้องเข้าสู่โปรแกรมย่อยทั้ง 3 วิธี เปรียบเทียบโดยใช้ค่าความคลาดเคลื่อนสมบูรณ์สูงสุด ทำการจำลองข้อมูลครบ 1,000 รอบ พิจารณาค่าความคลาดเคลื่อนสมบูรณ์สูงสุดของแต่ละวิธี เลือกค่าความคลาดเคลื่อนสมบูรณ์สูงสุดที่มีค่าต่ำสุดใน 1,000 รอบ ของแต่ละวิธี แล้วนำค่าต่ำสุดของแต่ละวิธีมาเปรียบเทียบกับกัน วิธีที่ให้ค่าความคลาดเคลื่อนสมบูรณ์ต่ำที่สุด เป็นวิธีที่ดีที่สุดในแต่ละสภานการณ์ เพราะเป็นวิธีที่ให้ค่าประมาณใกล้เคียงค่าจริงมากที่สุด

บทที่ 4
ผลการวิเคราะห์ข้อมูล

ผลการวิจัยในครั้งนี้ได้จากการจำลองข้อมูลด้วยโปรแกรม R โดยมีวัตถุประสงค์เพื่อศึกษาวิธีประมาณค่าสูญหาย และเปรียบเทียบประสิทธิภาพของวิธีประมาณค่า ทั้ง 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ (Iterative Method) วิธีประมาณค่าสูญหายของวิลคินซัล (Wilkinson Method) และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน (Nearest Neighbor Imputation Method) ในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ทำการศึกษาตามสถานการณ์ต่างๆ ดังนี้ จำนวนปัจจัย A มี 3 ลักษณะ คือ 3, 4 และ5 ระดับ จำนวนปัจจัย B มี 3 ลักษณะ คือ 3, 4 และ5 ระดับ จำนวนบล็อกมี 3, 4 และ5 บล็อก กำหนดสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% จำนวนข้อมูลสูญหายเป็น 10%, 20% และ30% ในรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ(Arbitrary) กำหนดให้ค่าคงที่ h เป็น 1, 2 และ3 ได้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล โดยทำการจำลองซ้ำสถานการณ์ละ 1,000 รอบ โดยประสิทธิภาพในการประมาณค่าสูญหายในแต่ละวิธีนั้นพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error) ซึ่งวิธีใดที่ให้ค่าความคาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด จะเป็นวิธีการประมาณค่าสูญหายที่ดีที่สุด เนื่องจากค่าที่ได้จากการประมาณค่าสูญหายนั้นมีค่าใกล้กับค่าจริงที่สูญหายไปมากที่สุด

สัญลักษณ์และอักษรย่อในการวิเคราะห์ข้อมูล

Missing	แทน	จำนวนข้อมูลสูญหาย
C.V.	แทน	สัมประสิทธิ์ความผันแปร
h	แทน	ค่าคงที่
MAE	แทน	ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด
IM	แทน	วิธีประมาณค่าสูญหายโดยการวนซ้ำ
WM	แทน	วิธีประมาณค่าสูญหายของวิลคินซัล
KNN	แทน	วิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน

การนำเสนอผลการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูลจะอยู่ในรูปของตาราง ผู้วิจัยขอนำเสนอผลการวิเคราะห์ข้อมูลตามลำดับดังนี้

1. การวิเคราะห์คุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์

   1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30%

   1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง

2. เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3

ผลการวิเคราะห์ข้อมูล

ผู้วิจัยขอเสนอผลการวิเคราะห์ข้อมูล แบ่งออกเป็น 2 ตอนซึ่งมีรายละเอียดดังต่อไปนี้

ตอนที่ 1 ผลการวิเคราะห์คุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์

1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30%

1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง

ตอนที่ 2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3

ตอนที่ 1 ผลการวิเคราะห์คุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์

การนำเสนอผลการวิเคราะห์ข้อมูลในตอนนี้ เป็นการบรรยายคุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์ โดยนำเสนอเป็น 2 ส่วน ดังนี้ 1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30% 1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง โดยมีราย ละเอียด ดังต่อไปนี้

1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30%

การวิเคราะห์ข้อมูลเบื้องต้นของคุณลักษณะของข้อมูลที่ได้จากการจำลอง จำนวนข้อมูลสูญหายผู้วิจัยคิดคำนวณจากขนาดตัวอย่าง ดังตารางต่อไปนี้

ตาราง 15 จำนวนข้อมูลสูญหายในแต่ละสถานการณ์

ปัจจัย A	ปัจจัย B	จำนวนบล็อก	ขนาดตัวอย่าง	จำนวนข้อมูลสูญหายของ 10%	20%	30%
3	3	3	27	3	5	8
		4	36	4	7	11
		5	45	5	9	14
	4	3	36	4	7	11
		4	48	5	10	14
		5	60	6	12	18
	5	3	45	5	9	14
		4	60	6	12	18
		5	75	8	15	23
4	3	3	36	4	7	11
		4	48	5	10	14
		5	60	6	12	18
	4	3	48	5	10	14
		4	64	6	13	19
		5	80	8	16	24
	5	3	60	6	12	18
		4	80	8	16	24
		5	100	10	20	30
5	3	3	45	5	9	14
		4	60	6	12	18
		5	75	8	15	23
	4	3	60	6	12	18
		4	80	8	16	24
		5	100	10	20	30
	5	3	75	8	15	23
		4	100	10	20	30
		5	125	13	25	38

จากตาราง 15 พบว่า จำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 จำนวนข้อมูลสูญหายคิดคำนวณจากจำนวนตัวอย่างเป็น 10%, 20% และ30% ดังนั้น จำนวนตัวอย่างที่มีค่าน้อยที่สุดคือ 27 มีข้อมูลสูญหาย เท่ากับ 3 ค่า, 5 ค่า และ 8 ค่า ตามลำดับ จำนวนตัวอย่างที่มีค่ามากที่สุดคือ 125 มีข้อมูลสูญหาย เท่ากับ 13 ค่า, 25 ค่า และ 38 ค่า ตามลำดับ

1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง

การวิเคราะห์ข้อมูลเป็นการบรรยายคุณลักษณะของข้อมูลการจำลอง กำหนดค่าเริ่มต้น ให้ค่าเฉลี่ยมีค่าเท่ากับ 50 ทำการจำลองข้อมูลตามรูปแบบเชิงเส้นตรงของแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบ RCB และประมาณค่าจากวิธีประมาณค่า 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายวิลคินซัล และวิธีประมาณค่าโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน ตามสถานการณ์ที่กำหนดทั้งหมด 729 สถานการณ์ โดยจำแนกตามปัจจัย A, ปัจจัย B, จำนวนบล็อก และจำนวนข้อมูลสูญหาย ข้อมูลที่ได้จากการจำลองจะผันแปรไปตามค่าสัมประสิทธิ์ความผันแปร และค่าคงที่ h ตัวอย่างข้อมูลแสดงดังต่อไปนี้

ตาราง 16 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 10% ในสถานการณ์นี้จะมีข้อมูลสูญหายทั้งหมด 3 ค่า

จำนวนข้อมูลสูญหาย	C.V.	ค่าคงที่	ข้อมูลที่ได้จากการจำลอง					
			IM		WM		KNN	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 10%	C.V. 5%	h = 1	50.73254	49.49199	50.73254	49.49199	50.73254	50.74364
			52.53830	51.86784	52.53830	51.86784	52.53830	50.84932
			50.61655	53.30708	50.61655	53.30708	50.61655	48.99945
		h = 2	46.70998	49.41725	46.70998	49.41725	46.70998	44.68117
			48.98497	43.23223	48.98497	43.23223	48.98497	45.60411
			52.16752	47.15717	52.16752	47.15717	52.16752	52.12145
		h = 3	48.64803	49.83694	48.64803	49.83694	48.64803	40.85894
			44.51698	39.69813	44.51698	39.69813	44.51698	54.24165
			64.85477	46.53604	64.85477	46.53604	64.85477	48.08832
	C.V. 25%	h = 1	23.93561	63.09925	23.93561	63.09925	23.93561	49.22620
			64.40088	88.77670	64.40088	88.77670	64.40088	73.78637
			18.43629	74.95540	18.43629	74.95540	18.43629	51.99331
		h = 2	81.88398	86.48347	81.88398	86.48347	81.88398	95.09964
			71.60161	99.42011	71.60161	99.42011	71.60161	65.16714
			44.16731	89.56294	44.16731	89.56294	44.16731	65.16714
		h = 3	11.22746	53.54468	11.22746	53.54468	11.22746	36.79792
			101.9736	85.99396	101.9736	85.99396	101.9736	79.98828
			23.66961	78.52368	23.66961	78.52368	23.66961	69.93152
	C.V. 45%	h = 1	93.31850	83.85521	93.31850	83.85521	93.31850	87.09024
			4.11307	32.80957	4.11307	32.80957	4.11307	54.69715
			73.60996	115.23320	73.60996	115.23320	73.60996	57.29951
		h = 2	62.29151	165.34540	62.29151	165.34540	62.29151	72.09271
			203.34130	106.86940	203.34130	106.86940	203.34130	67.34897
			53.43706	114.34500	53.43706	114.34500	53.43706	97.35277
		h = 3	448.53330	303.85920	448.53330	303.85920	448.53330	122.14200
			17.77242	246.08900	17.77242	246.08900	17.77242	179.51050
			33.72513	310.50400	33.72513	310.50400	33.72513	0.76335

จากตาราง 16 พบว่า การจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 10% มีข้อมูลสูญหาย 3 ค่า ข้อมูลที่จำลองมีค่าไม่

แตกต่างจากค่าเฉลี่ยมากในค่าสัมประสิทธิ์ความผันแปรเท่ากับ 5% และจะมีค่าแตกต่างจากค่าเฉลี่ยมากขึ้นเมื่อความผันแปรมีค่ามากขึ้น ณ สัมประสิทธิ์ความผันแปรเท่ากับ45% ข้อมูลจะมีการกระจายตัวมาก ทำให้ค่าที่ประมาณได้มีค่าแตกต่างจากค่าจริงมาก ค่าคงที่ h ในระดับค่าสัมประสิทธิ์ความผันแปรเดียวกันจะมีค่าใกล้เคียงกัน และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

ตาราง 17 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 20% ในสถานการณ์นี้จะมีข้อมูลสูญหายทั้งหมด 5 ค่า

จำนวนข้อมูลสูญหาย	C.V.	ค่าคงที่	ข้อมูลที่ได้จากการจำลอง					
			IM		WM		KNN	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 20%	C.V. 5%	h = 1	57.24168	45.44190	57.24168	45.44190	57.24168	51.00469
			51.16563	50.50192	51.16563	50.50192	51.16563	48.27484
			48.64825	48.71168	48.64825	48.71168	48.64825	49.83964
			48.11500	49.39433	48.11500	49.39433	48.11500	50.73150
			49.75670	44.18343	49.75670	44.18343	49.75670	51.06548
		h = 2	50.80701	53.68125	50.80701	53.68125	50.80701	49.83576
			51.70815	47.86477	51.70815	47.86477	51.70815	48.96613
			54.38406	44.93484	54.38406	44.93484	54.38406	46.03501
			53.60685	43.63168	53.60685	43.63167	53.60685	46.55931
			44.71730	48.87648	44.71730	48.87648	44.71730	47.80653
		h = 3	62.41058	61.89813	62.41058	61.89813	62.41058	50.09233
			58.37967	50.00917	58.37967	50.00917	58.37967	49.98377
			46.09071	36.30899	46.09071	36.30898	46.09071	37.84974
			45.11983	46.09955	45.11983	46.09955	45.11983	45.53763
			61.05352	40.86789	61.05352	40.86788	61.05352	51.18944
	C.V. 25%	h = 1	47.44850	125.27450	47.44850	125.27450	47.44850	95.79416
			54.53122	60.45334	54.53122	60.45334	54.53122	47.14991
			21.64632	116.50460	21.64632	116.50460	21.64632	22.80447
			100.95990	31.06825	100.95990	31.06825	100.95990	50.48263
			42.62226	44.14906	42.62226	44.14906	42.62226	52.75895

ตาราง 17 (ต่อ)

จำนวนข้อมูลสูญหาย	C.V.	ค่าคงที่	ข้อมูลที่ได้จากการจำลอง					
			IM		WM		KNN	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 20%	C.V. 25%	h = 2	55.89605	51.98406	55.89605	51.98406	55.89605	22.83418
			46.53511	5.06666	46.53511	5.06666	46.53511	59.04163
			16.15674	49.06968	16.15674	49.06968	16.15674	99.57602
			60.45704	34.71949	60.45704	34.71949	60.45704	12.89452
			48.11292	136.19570	48.11292	136.19570	48.11292	54.22734
		h = 3	29.41023	69.20518	29.41023	69.20518	29.41023	96.77878
			47.31345	111.18220	47.31345	111.18220	47.31345	56.84820
			9.19161	73.06898	9.19161	73.06898	9.19161	96.77878
			33.95606	111.17410	33.95606	111.17410	33.95606	72.43808
			26.89631	59.05408	26.89631	59.05409	26.89631	4.48874
	C.V. 45%	h = 1	37.43394	37.53664	37.43394	37.53664	37.43394	32.19664
			2.80300	81.97008	2.80300	81.97008	2.80300	40.77191
			89.50413	12.03577	89.50413	12.03577	89.50413	34.98976
			41.92207	34.06059	41.92207	34.06059	41.92207	34.98976
			19.49876	17.49322	19.49876	17.49322	19.49876	60.73142
		h = 2	-15.16565	177.71740	-15.16565	177.71740	-15.16565	25.00664
			31.73239	190.08710	31.73239	190.08710	31.73239	-33.03735
			-19.11301	72.69369	-19.11301	72.69369	-19.11301	52.87366
			-17.70248	18.25150	-17.70248	18.25150	-17.70248	44.38565
			42.19122	136.89120	42.19122	136.89120	42.19122	70.15260
		h = 3	505.08780	267.27440	505.08780	267.27440	505.08780	123.12370
			82.60383	96.97675	82.60383	96.97675	82.60383	217.76460
			52.29056	86.62100	52.29056	86.62100	52.29056	138.63580
			66.64505	-20.50283	66.64505	-20.50283	66.64505	7.11176
			178.56650	44.01551	178.56650	44.01551	178.56650	-55.83199

จากตาราง 17 พบว่า การจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 20% มีข้อมูลสูญหาย 5 ค่า ข้อมูลที่จำลองมีค่าแตกต่างจากค่าเฉลี่ยมากขึ้นเมื่อความผันแปรมีค่ามากขึ้น เพราะข้อมูลมีการกระจายตัวมากขึ้นค่าที่ประมาณได้จากทั้ง 3 วิธีมีค่าไม่แตกต่างกันเมื่อสัมประสิทธิ์ความผันแปรเท่ากับ 5 แต่จะมีค่าแตกต่างกันมากขึ้นเมื่อสัมประสิทธิ์ความผันแปรมากขึ้น ค่าคงที่ h ในระดับค่าสัมประสิทธิ์ความผันแปรเดียวกันจะมี

ค่าใกล้เคียงกัน และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

ตาราง 18 ตัวอย่างแสดงค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลองสถานการณ์ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 30% ในสถานการณ์นี้จะมีข้อมูลสูญหายทั้งหมด 8 ค่า

จำนวนข้อมูลสูญหาย	**C.V.**	**ค่าคงที่**	**ข้อมูลที่ได้จากการจำลอง**					
			IM		**WM**		**KNN**	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 30%	C.V. 5%	h = 1	50.36863	48.40580	50.36863	48.40580	50.36863	49.07995
			47.52886	51.46796	47.52886	51.46797	47.52886	48.98713
			51.55130	53.45469	51.55130	53.45469	51.55130	48.98713
			50.04237	53.29006	50.04237	53.29007	50.04237	50.48241
			49.74866	52.95606	49.74866	52.95606	49.74866	50.24991
			47.69284	54.94278	47.69284	54.94279	47.69284	50.24991
			49.76743	48.08904	49.76743	48.08904	49.76743	48.64982
			52.24285	50.03380	52.24285	50.03380	52.24285	49.26397
		h = 2	57.26671	55.65014	57.26671	55.65014	57.26671	47.80317
			43.12171	52.94154	43.12171	52.94154	43.12171	52.54114
			49.07744	49.41534	49.07744	49.41533	49.07744	47.11922
			43.30015	49.15190	43.30015	49.15190	43.30015	47.11922
			48.67015	47.59885	48.67015	47.59885	48.67015	46.36337
			49.89781	47.89940	49.89781	47.89940	49.89781	46.36337
			44.30000	46.84247	44.30000	46.84246	44.30000	46.36337
			46.40127	41.44439	46.40127	41.44439	46.40127	48.46681
		h = 3	47.53761	54.88212	47.53761	54.88212	47.53761	50.81841
			50.32298	48.25044	50.32298	48.25044	50.32298	49.43762
			48.88264	49.38327	48.88264	49.38327	48.88264	51.20772
			53.32364	45.49873	53.32364	45.49873	53.32364	48.27718
			48.48781	47.85053	48.48781	47.85053	48.48781	48.27718
			46.12593	45.71524	46.12593	45.71524	46.12593	48.27718
			49.63160	46.45237	49.63160	46.45237	49.63160	48.27718
			51.34081	44.42378	51.34081	44.42378	51.34081	48.27718

ตาราง 18 (ต่อ)

จำนวนข้อมูลสูญหาย	C.V.	ค่าคงที่	ข้อมูลที่ได้จากการจำลอง					
			IM		WM		KNN	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 30%	C.V. 25%	h = 1	38.16454	78.52034	38.16454	78.52034	38.16454	50.52400
			86.58716	39.82672	86.58716	39.82673	86.58716	65.28894
			30.66873	85.57662	30.66873	85.57662	30.66873	52.89891
			70.56252	10.99994	70.56252	10.99993	70.56252	28.62666
			69.31825	7.16537	69.31825	7.16537	69.31825	25.48248
			40.97441	61.53488	40.97441	61.53488	40.97441	42.97868
			65.17726	49.44459	65.17726	49.44459	65.17726	71.28080
			12.32206	31.43165	12.32206	31.43165	12.32206	64.81710
		h = 2	75.04807	54.88181	75.04807	54.88181	75.04807	37.72888
			63.02223	44.12149	63.02223	44.12149	63.02223	60.45309
			93.40737	45.10284	93.40737	45.10284	93.40737	70.37685
			32.42435	21.33571	32.42435	21.33571	32.42435	70.37685
			71.97736	52.72427	71.97736	52.72427	71.97736	56.65715
			57.31966	74.92805	57.31966	74.92805	57.31966	56.65715
			88.02785	8.55996	88.02785	8.55995	88.02785	27.56569
			68.72492	45.95509	68.72492	45.95509	68.72492	35.85221
		h = 3	6.136155	44.41849	6.136155	44.41849	6.136155	40.78059
			150.6101	-7.87865	150.6101	-7.878652	150.6101	34.86268
			-33.96243	164.7422	-33.96243	164.7422	-33.96243	81.65216
			200.7844	60.76245	200.7844	60.76245	200.7844	89.40704
			18.64827	153.0165	18.64827	153.0165	18.64827	81.65216
			92.01793	148.6509	92.01793	148.6509	92.01793	110.9779
			-28.48645	183.9391	-28.48645	183.9391	-28.48645	121.3019
			-3.175322	172.2134	-3.175322	172.2134	-3.175322	121.3019

ตาราง 18 (ต่อ)

จำนวนข้อมูลสูญหาย	C.V.	ค่าคงที่	ข้อมูลที่ได้จากการจำลอง					
			IM		WM		KNN	
			ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ	ค่าจริง	ค่าประมาณ
Missing 30%	C.V. 45%	h = 1	38.89681	16.81495	38.89681	16.81495	38.89681	18.94251
			13.41163	117.41080	13.41163	117.41080	13.41163	26.57206
			36.96195	91.08362	36.96195	91.08363	36.96195	26.57206
			31.76339	114.84460	31.76339	114.84460	31.76339	48.26198
			40.14342	136.49160	40.14342	136.49160	40.14342	6.85345
			10.28842	110.16440	10.28842	110.16440	10.28842	6.85345
			20.90712	17.35395	20.90712	17.35394	20.90712	63.59223
			55.34308	77.16237	55.34308	77.16237	55.34308	36.16310
		h = 2	69.60597	276.40780	69.60597	276.40780	69.60597	66.84308
			366.05230	538.93890	366.05230	538.93890	366.05230	66.84308
			304.57500	133.63860	304.57500	133.63860	304.57500	122.44900
			-16.75158	66.93549	-16.75158	66.93549	-16.75158	47.49130
			-23.53900	41.65449	-23.53900	41.65449	-23.53900	47.49130
			405.33990	-52.53065	405.33990	-52.53065	405.33990	39.79964
			165.12030	445.36840	165.12030	445.36840	165.12030	73.76995
			112.11080	-37.08650	112.11080	-37.08650	112.11080	46.28073
		h = 3	-5.69077	-47.93314	-5.69077	-47.93314	-5.69077	-49.68149
			-41.94476	269.95460	-41.94476	269.95460	-41.94476	-46.11171
			-8.99755	144.45260	-8.99755	144.45260	-8.99755	-46.11171
			-18.60183	237.44420	-18.60183	237.44420	-18.60183	13.34222
			22.43527	319.12470	22.43527	319.12470	22.43527	-96.10699
			-48.56194	193.62280	-48.56194	193.62280	-48.56194	-96.10699
			-56.27622	-34.06663	-56.27622	-34.06663	-56.27622	108.02860
			27.28461	144.53250	27.28461	144.53250	27.28461	-7.91975

จากตาราง 18 พบว่า การจำลองสถานการณ์ ปัจจัย A มี 3 ระดับ, ปัจจัย B มี 3 ระดับ, จำนวนบล็อก 3 บล็อก และข้อมูลสูญหาย 30% มีข้อมูลสูญหาย 8 ค่า ณ สัมประสิทธิ์ความผันแปรเท่ากับ 5% ค่าคงที่ h ทั้ง 3 ระดับมีค่าใกล้เคียงกัน มีค่าไม่แตกต่างจากค่าเฉลี่ย ข้อมูลมีการกระจายตัวไม่มาก ค่าที่ประมาณได้กับค่าจริงมีค่าไม่แตกต่างกันมาก สัมประสิทธิ์ความผันแปรเท่ากับ 25% ค่าคงที่ h ทั้ง 3 ระดับ มีค่าแตกต่างกัน และจะมีค่ามากขึ้นเมื่อค่าคงที่ h เท่ากับ 3 ข้อมูลมีการกระจายตัวมากขึ้น ค่าที่ประมาณได้กับค่าจริงมีค่าแตกต่างกันมากขึ้น และสัมประสิทธิ์

ความผันแปรเท่ากับ 45% ช่วงของข้อมูลกว้างมากขึ้น ทำให้ค่าที่จำลองได้มีการกระจายตัวมาก ค่าที่ประมาณได้กับค่าจริงมีค่าแตกต่างกันมากขึ้น และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

ตอนที่ 2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ3 ผู้วิจัยขอนำเสนอผลการเปรียบเทียบในรูปแบบตาราง ค่าในตารางเป็นค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ที่มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำสุดในแต่ละสถานการณ์ โดยมีรายละเอียดดังนี้

ตาราง 19 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				3								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 10 %	C.V. 5%	h = 1	IM	17.10412985	15.01171457	***14.35067281***	15.82171194	13.66733070	23.00144885	***18.49360224***	***13.72661995***	15.86001816
			WM	17.10413109	15.01171463	14.35067294	15.82171200	13.66733057	23.00144878	18.49360425	13.72661996	15.86001801
			KNN	***13.56077507***	***12.11799731***	14.83005262	***14.88742418***	***12.30889891***	***21.87244105***	24.61125266	14.11341118	***14.32396331***
		h = 2	IM	25.94582075	26.64221476	31.97181667	28.07829710	***31.00210931***	***27.80195961***	***23.46858383***	34.65311135	***24.67438661***
			WM	25.94582094	26.64221484	31.97181668	28.07829709	31.00210930	27.80195963	23.46858393	34.65311164	24.67438674
			KNN	***22.34607772***	***26.30898664***	***26.85306969***	***21.76168992***	35.05275841	30.65184893	27.99293638	***31.91857236***	24.73562661
		h = 3	IM	35.07324056	44.21800802	39.63938477	37.78062699	***34.15675795***	49.69322782	46.38928824	47.17176382	42.22407158
			WM	35.07324057	44.21800880	39.63938420	37.78062702	34.15675798	49.69322795	46.38928825	47.17176359	42.22407137
			KNN	***34.42068949***	***35.85148474***	***38.38009249***	***32.74627067***	47.99877064	***48.62771148***	***40.25514704***	***43.53969725***	***38.18550148***
	C.V. 25%	h = 1	IM	***211.81757552***	244.25153161	***245.88802480***	297.49971107	262.86879964	***242.39947620***	299.03164865	329.66177826	***220.49476775***
			WM	***211.81757552***	244.25153238	245.88802483	297.49971117	262.86880087	242.39947621	299.03165409	329.66177860	220.49476789
			KNN	255.74389501	***207.25166338***	294.61913028	***241.13732002***	***258.63087103***	274.46606525	***210.88929150***	***278.79688100***	227.16214471
		h = 2	IM	607.48757709	687.83071674	***539.51198422***	601.56636040	695.43332780	509.05226324	626.88459151	636.08710796	760.15246173
			WM	607.48757734	687.83071640	539.51198439	601.56636042	695.43332796	509.05226280	626.88459125	636.08710863	760.15246180
			KNN	***539.15594780***	***575.43012358***	546.99436780	***504.97144489***	***529.96927683***	***457.05501293***	***617.12487933***	***632.73653220***	***737.80827076***
		h = 3	IM	***1000.35522619***	874.33972507	***855.82447976***	1001.94318611	984.54009200	882.18429016	***795.91023542***	***831.34833635***	813.37467874
			WM	1000.35522657	874.33972508	855.82447978	1001.94318605	984.54009188	***882.18429013***	795.91023569	831.34833858	***813.37467872***
			KNN	1000.58283557	***792.58893375***	953.09310568	***852.09782592***	***744.33140462***	997.71678201	831.99060400	870.93607142	864.51197969
	C.V. 45%	h = 1	IM	824.74404320	989.02885656	754.76931118	939.83683948	***792.77592896***	964.12469781	854.74946180	827.06090073	756.46186446
			WM	824.74404322	989.02885701	754.76931121	939.83683951	792.77592905	***964.12469773***	***854.74946156***	827.06090082	***756.46186418***
			KNN	***745.59517536***	***750.03254300***	***716.24263333***	***716.84189846***	849.29040204	1005.32050549	1029.53541787	***817.70568263***	763.20121291
		h = 2	IM	2868.98331497	1585.87280570	1819.32562388	1552.25831372	2225.55768227	1890.20742459	2508.40953084	2257.26660734	2171.41511999
			WM	2869.02428238	***1585.87280567***	1819.32562390	***1552.25831361***	2225.55768229	***1890.20742457***	2508.40953423	2257.26660759	2171.41512024
			KNN	***1997.52634666***	1732.08895054	***1580.88805650***	1742.91387395	***2033.95637720***	1892.59755077	***2454.16780248***	***2012.80847340***	***2049.54640918***
		h = 3	IM	2875.84784747	2543.12210333	***2784.09076937***	***2721.27908369***	***4201.89732153***	2452.58696516	2822.68057623	***2883.48553808***	3226.88804320
			WM	2875.84784748	***2543.12210328***	2784.09076953	2721.27908738	4201.89732158	***2452.58696506***	2822.68057617	2883.48553809	3226.88804321
			KNN	***2735.08433508***	2627.80993084	3192.89669538	2737.63258767	4282.65004270	2566.47199713	***2375.51584505***	3039.94246077	***2672.53166492***

ตาราง 20 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				4								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 10 %	C.V. 5%	h = 1	IM	19.23331631	***14.64445252***	***13.12152652***	18.78032720	19.55079034	17.73693291	19.36909417	16.29626684	***13.66906286***
			WM	19.23331629	14.64445259	13.12152659	18.78033227	19.55079033	17.73693301	19.36910286	16.29626713	13.66908120
			KNN	***16.54883043***	14.97245915	14.45674573	***14.88121047***	***19.47825987***	***16.24480558***	***15.48442706***	***14.59607187***	15.05189220
		h = 2	IM	38.71541966	28.25722734	27.64657772	24.77360816	***29.89597298***	28.91102364	30.54060133	32.32833254	29.96882820
			WM	38.71541987	28.25722741	27.64657761	24.77360902	29.89597299	28.91104051	30.54060126	32.32833267	29.96882784
			KNN	***30.51685114***	***27.24895300***	***22.12679489***	***24.31646956***	30.36246645	***28.15179529***	***29.54829201***	***30.09833741***	***27.84988472***
		h = 3	IM	49.71641088	***39.50900009***	***39.84198697***	42.35670455	***37.58470563***	48.38086592	47.36781508	44.64319197	45.69517167
			WM	49.71641102	39.50900040	39.84198709	42.35670463	37.58470566	48.38086602	47.36781482	44.64319220	45.69517168
			KNN	***42.86525431***	40.44622536	52.13626224	***33.46615293***	38.91091186	***43.62724314***	***45.38065933***	***39.02207062***	***44.40486050***
	C.V. 25%	h = 1	IM	308.23862138	***241.63800665***	247.76312056	288.84341936	331.93734059	***352.80996742***	251.11450783	264.26369758	421.23702068
			WM	308.23862140	241.63800729	***247.76312043***	288.84342624	331.93734072	352.80996744	251.11450786	264.26369832	421.23702881
			KNN	***240.19185829***	249.30515979	254.38799570	***283.26401285***	***289.54160274***	358.93088947	***195.38466675***	***252.90203109***	***318.61966258***
		h = 2	IM	514.34979323	458.63284713	579.06859985	514.46051412	470.62667468	708.48438677	***586.33150337***	563.09015581	646.01679847
			WM	514.34979321	***458.63284711***	579.06859984	514.46076195	***470.62667451***	708.48438675	***586.33150337***	563.09015590	646.01679842
			KNN	***472.77943561***	508.29416896	***499.80136878***	***453.03885475***	484.38902174	***618.38115761***	657.27286881	***531.73894458***	***558.17299703***
		h = 3	IM	1118.21671930	***773.90931459***	889.15495482	***764.87478704***	***848.48334817***	828.51234255	1001.53949921	1377.61042887	900.97199866
			WM	1118.21671926	***773.90931459***	889.15495482	764.87478706	848.48335088	***828.51234227***	1001.53949922	1377.61042890	900.97200060
			KNN	***1027.23796525***	840.69730545	***713.27568984***	877.79586137	945.86215689	832.39009542	***833.08496444***	***1223.81844658***	***777.00709521***
	C.V. 45%	h = 1	IM	850.57663142	***768.68913413***	1121.81717133	***868.20168398***	798.67784615	985.74173800	738.98141432	1079.04473212	***946.26715815***
			WM	850.57663121	768.68913414	1121.81717133	868.20168561	***798.67784573***	***985.74173794***	738.98141437	1079.04473225	946.26715826
			KNN	***815.54004425***	806.00489252	***1095.93391166***	1054.04838981	1136.62130800	994.87443147	***655.03300281***	***1057.70289780***	1010.65793105
		h = 2	IM	2015.83323581	1981.69560617	2361.40780602	1868.23912727	1806.98584994	2062.97822084	2097.56949475	2130.75768645	***2077.69136182***
			WM	2015.83324007	1981.69560620	2361.40780557	1868.23912728	1806.98584993	2062.97822099	2097.56949398	2130.75768701	2077.69136228
			KNN	***1850.22205369***	***1817.60878857***	***1837.00321093***	***1745.21021768***	***1574.50677319***	***1986.29959940***	***1873.33560153***	***1887.30611675***	2140.03182155
		h = 3	IM	2954.86340512	3038.15134663	***2346.49437020***	2637.86415128	3552.70334816	***3501.99739336***	2736.18197385	***3064.61013061***	***2722.90627754***
			WM	2954.86340781	3038.15134778	2346.49437502	2637.86415093	3552.70334822	3501.99739342	2736.18197410	3064.61013066	2722.90627795
			KNN	***2428.51247679***	***2410.12411422***	2670.23926021	***2383.67545676***	***2869.89187668***	3739.97046797	***2724.55304936***	3737.62630415	3108.53585316

ตาราง 21 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 10% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก3,4,5 บล็อก

ปัจจัย A				5								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 10 %	C.V. 5%	h = 1	IM	15.09619304	18.20575586	***17.11105511***	16.53146539	17.67106636	18.29668817	***14.17804357***	15.69037372	14.27400170
			WM	15.09619313	18.20575607	17.11105514	16.53146521	17.67106681	***18.29668796***	14.17804408	15.69042901	14.27400365
			KNN	***13.87325197***	***16.08266194***	18.32786318	***15.85984606***	***14.94565507***	20.33504997	15.04339311	***15.39294265***	***13.65679911***
		h = 2	IM	25.46636950	***31.47646419***	29.70043642	25.57141822	27.38425289	25.50932226	29.18159535	27.60069570	29.70037067
			WM	***25.46636949***	31.47646422	***29.70043641***	25.57141838	27.38425288	***25.50932224***	29.18159541	27.60069583	29.70037073
			KNN	28.52343801	32.15305403	30.36084137	***25.32667749***	***26.17306219***	31.36421254	***25.60435257***	***25.57208425***	***28.86492737***
		h = 3	IM	42.86094048	34.08128079	46.61173481	42.16057679	37.24285124	35.89870971	51.79458275	49.00082538	37.20736460
			WM	42.86094050	34.08127997	***46.61173438***	42.16057677	37.24285179	35.89870924	51.79458281	49.00082539	***37.20736430***
			KNN	***39.60952375***	***33.05085460***	50.29969190	***40.54023046***	***37.07335667***	***34.55279403***	***40.11875921***	***46.63614938***	39.37091734
	C.V. 25%	h = 1	IM	***267.85231411***	271.48213236	259.72982411	232.34442048	***293.98879627***	278.97336954	241.56259906	318.08215112	271.72332562
			WM	267.85260182	271.48213250	259.72982388	***232.34442041***	***293.98879627***	278.97336959	241.56260180	***318.08215111***	271.72332567
			KNN	284.16234773	***252.37488304***	***243.62746207***	243.80308645	333.08928494	***260.38981070***	***223.33751356***	347.49883810	***258.22905807***
		h = 2	IM	***561.65645304***	585.00625032	767.54800409	579.69905210	691.89147249	***662.16370625***	584.74176790	631.66678633	***615.11329796***
			WM	***561.65645304***	585.00625038	767.54800435	579.69906985	691.89147296	662.16370629	584.74176823	631.66684295	615.11329808
			KNN	674.57770542	***481.24092640***	***580.25046760***	***529.18851612***	***672.11769523***	686.39102566	***466.84337156***	***514.96273903***	676.32683221
		h = 3	IM	844.71973241	***838.24399884***	***998.62185764***	999.27988985	***969.65604746***	937.03498633	1004.30227337	850.04935613	***1027.53258884***
			WM	***844.71973213***	838.24399892	998.62185796	999.27988977	969.65604754	937.03498671	1004.30227337	***850.04935033***	1027.53258925
			KNN	920.73386628	910.40167297	1080.01086212	***898.56844661***	1001.85999646	***929.74578304***	***849.27378394***	856.78196837	1053.18420902
	C.V. 45%	h = 1	IM	959.19197780	747.35791217	833.54205493	1023.98981082	761.45758842	***965.79586080***	879.32218128	***869.43365178***	947.67998172
			WM	959.19197786	***747.35791207***	833.54205372	1023.98981088	761.45758835	965.79586149	879.32218127	869.43365192	947.67998155
			KNN	***866.04509319***	1057.24705882	***809.27021569***	***775.64589927***	***705.55117537***	975.73917211	***771.86009489***	904.92393383	***942.84937733***
		h = 2	IM	2495.70525058	2036.68989572	2020.93778943	2369.05198239	2343.76705069	1839.21773873	1841.50175396	2153.09364626	***1607.29929551***
			WM	2495.70525072	2036.68989678	2020.93778903	2369.05198238	2343.76704770	1839.21773897	1841.50175388	2153.09364679	1607.29929563
			KNN	***2244.30556747***	***1671.50011568***	***1825.42285397***	***2232.30894977***	***2230.90633482***	***1678.31591961***	***1822.16389050***	***1855.45599287***	2096.14550533
		h = 3	IM	2843.48959861	3287.21435236	3157.52254785	3014.11731176	3527.16419716	3354.41129605	3884.24147060	***2668.02555890***	***2838.72162788***
			WM	***2843.48959855***	3287.21435300	3157.52254795	3014.11731198	3527.16419710	3354.41129605	3884.24147308	2668.02555918	2838.72162802
			KNN	3262.37043221	***2523.47292555***	***2639.35664444***	***2984.21048078***	***2899.21740352***	***3021.61243846***	***2792.90594193***	2769.19688049	3236.41118930

ตาราง 22 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				3								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 20 %	C.V. 5%	h = 1	IM	15.38245495	15.92705723	16.31596940	19.09609907	17.26663794	**21.55438725**	16.13191145	**18.67728351**	19.01375643
			WM	15.38245388	15.92705628	16.31596975	19.09609889	17.26664452	21.55438826	16.13191258	18.67728585	19.01375243
			KNN	**14.28896538**	**14.75177478**	**15.45024023**	**17.24478451**	**15.03239713**	21.73902136	**15.15696165**	18.84836780	**17.06004389**
		h = 2	IM	37.89775524	38.79689815	33.05226893	29.41928116	31.00395611	36.15148819	**28.87102327**	**31.12196533**	**28.41613275**
			WM	37.89775804	38.79689870	33.05226928	29.41928288	31.00396837	36.15148716	28.87102456	31.12208520	28.41613291
			KNN	**34.13581644**	**28.48511832**	**32.93679830**	**24.00771509**	**27.98956923**	**31.97699735**	31.73979014	34.60053841	30.49971338
		h = 3	IM	38.94843645	50.30709016	36.97454422	44.84514326	44.17082395	38.60212353	46.90082930	49.01708843	51.29048848
			WM	38.94843936	50.30708979	36.97454078	44.84514340	44.17082481	38.60212949	46.90082767	49.01708881	51.29049085
			KNN	**36.75805099**	**45.95136414**	**41.03568369**	**39.30417955**	**42.12074898**	**38.40444718**	**40.61106596**	**45.39142187**	**50.10325830**
	C.V. 25%	h = 1	IM	286.00056926	275.88163681	339.39765097	304.67239331	373.06612064	**307.64947584**	265.88231678	361.86381416	325.43582089
			WM	286.00056981	275.88164727	339.39765254	304.67239890	373.06612582	307.64947900	265.88231266	361.86394579	**325.43581996**
			KNN	**268.49464463**	**270.71892271**	**276.31469911**	**276.68732377**	**345.61680784**	351.71240448	**251.34561760**	**246.41413783**	346.93746810
		h = 2	IM	563.87093344	635.50255045	**600.34370899**	909.74399080	617.15895581	**602.90935266**	669.68765895	660.74400929	666.84580256
			WM	563.87093392	**635.50254978**	600.34370990	909.74399305	617.15895684	602.90935332	669.68766814	660.74400875	666.84580336
			KNN	**501.45643659**	680.58131765	649.92742579	**611.99032070**	**561.78591959**	647.17272749	**580.49461466**	**571.86140199**	**575.59523701**
		h = 3	IM	892.09332941	952.33961242	844.88184946	1586.35270845	**933.39737094**	923.27702230	**897.83306412**	1185.72357121	1253.82835815
			WM	892.09333129	952.33961230	844.88194672	1586.35270928	933.39737169	923.27702364	897.83306875	1185.72357282	1253.82835739
			KNN	**837.52182804**	**837.93097494**	**792.16495212**	**1157.38246677**	987.12185626	**837.63939259**	942.79160800	**885.13191181**	**1137.07318602**
	C.V. 45%	h = 1	IM	937.98726242	1062.98775314	836.01619237	982.93625047	921.01285542	**839.59171220**	872.56097861	890.55596321	1026.60402301
			WM	937.98726245	1062.98775870	836.01619267	982.93625379	921.01285541	839.59171243	872.56098128	890.55599454	1026.60414326
			KNN	**800.37697435**	**768.53542215**	**770.24946397**	**955.94520655**	**765.08299325**	889.50034029	**802.12054364**	**877.36835877**	**836.78207014**
		h = 2	IM	1926.26716027	2431.85277371	1835.03121131	2291.06100345	2074.03530787	1986.08974680	2609.24231506	**1972.01480403**	1782.12729106
			WM	1926.26715990	2431.85283371	1835.03121210	2291.06100673	**2074.03530690**	**1986.08974658**	2609.24231510	1972.01480430	**1782.12729013**
			KNN	**1840.91325802**	**1868.51573206**	**1735.35951459**	**1924.73982275**	2170.93259526	2372.66478962	**1929.47790604**	2053.59276321	1959.87728827
		h = 3	IM	2715.97111301	3203.34664988	**2646.68472368**	**3318.74182133**	4499.46499637	3148.92696846	3894.62626026	4127.92653658	2820.76400717
			WM	2715.97111581	3203.38493121	2646.68472369	3318.74182424	4499.46500097	3148.92697109	3894.62626148	4127.92654015	2820.76400723
			KNN	**2666.99103218**	**2525.12822200**	2822.96536212	3476.10941784	**3068.79813500**	**2994.59484321**	**3053.18404080**	**2608.83880245**	**2814.01220124**

ตาราง 23 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				4								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 20 %	C.V. 5%	h = 1	IM	17.49991770	22.62742134	23.56270680	14.87151138	20.34570661	16.66037407	18.32068157	16.20041114	16.85710389
			WM	17.49992386	22.62742148	23.56271082	14.87151231	20.34570751	16.66037503	18.32067913	16.20041211	16.85722619
			KNN	***13.53985845***	***18.86400223***	***20.96054781***	***14.42746602***	***17.67675047***	***15.94247056***	***17.50567903***	***15.07042772***	***16.04042655***
		h = 2	IM	40.96373549	***36.71152553***	27.86212941	27.95271756	32.19930638	31.82706499	38.51510254	29.41924591	35.70854881
			WM	40.96373775	36.71152601	***27.86212914***	27.95272261	32.19930685	31.82706553	38.51510286	***29.41924517***	35.70854965
			KNN	***29.94965071***	37.35548822	32.97870979	***25.37154595***	***29.11896513***	***29.51612496***	***28.78633828***	30.27006531	***34.45089566***
		h = 3	IM	49.87404705	45.11391604	44.88840454	47.49069066	51.04440372	44.71855485	51.33330345	62.20290242	***44.87727397***
			WM	49.87404611	45.11391513	***44.88840387***	***47.49069028***	51.04440330	44.71856617	51.33330464	62.20290250	44.87727564
			KNN	***49.66102831***	***40.94349770***	47.93971881	49.38353078	***42.20112844***	***36.03713126***	***37.94274513***	***57.36731381***	50.64812696
	C.V. 25%	h = 1	IM	274.27286838	312.89290003	300.57338757	293.55048068	247.53017045	369.79276313	316.48462977	264.77201756	324.16318828
			WM	274.27286839	312.89292044	300.57338662	293.55048004	***247.53016760***	369.79276402	316.48463261	264.77208248	324.16318857
			KNN	***267.89004778***	***260.71842502***	***249.52708089***	***258.62988195***	324.51751816	***323.91512595***	***277.60498669***	***252.37871044***	***295.44811460***
		h = 2	IM	676.51297141	682.29631755	***692.14422565***	660.48867507	921.10307823	818.91617740	771.87896626	685.36498793	666.31534942
			WM	676.51326863	682.29631692	692.14422635	660.48867483	921.10308870	818.91617698	771.87897107	685.36498830	666.31535023
			KNN	***482.29281099***	***599.60377634***	758.60521777	***605.56167516***	***764.95195628***	***722.03661728***	***703.63324317***	***615.50264924***	***654.39787951***
		h = 3	IM	***1399.34685687***	1003.06868030	***896.93567223***	***977.44222351***	971.17638672	1151.26477135	940.32246584	1084.37300206	927.47726137
			WM	1399.34685786	1003.06868207	896.93567224	977.44222581	***971.17638646***	1151.26477196	940.32246997	1084.37300660	927.47726151
			KNN	1438.89639261	***834.67013447***	1134.15499456	1007.27920856	1038.54991490	***1020.97193808***	***893.12635566***	***908.46350146***	***856.87362024***
	C.V. 45%	h = 1	IM	***888.95157497***	980.01319940	933.91613868	1083.96756585	863.34953341	***913.64353890***	907.79571142	854.97591973	***903.54851176***
			WM	888.95157572	980.01320021	933.91613901	1083.96756512	***863.34953326***	913.64354087	907.79571364	854.97637764	903.54851414
			KNN	967.31526214	***835.42056458***	***788.64101673***	***978.62318078***	900.46530041	921.73620864	***740.31771408***	***758.96406014***	959.38638893
		h = 2	IM	***2002.24524616***	2346.74660996	***1779.14111680***	2337.35498960	2801.95007627	1962.16023342	***2188.22594870***	***2124.02546971***	2092.75283718
			WM	2002.24524651	2346.74663209	1779.14111842	2337.35499491	2801.95007925	***1962.16023313***	2188.22594998	2124.02549558	2092.75298541
			KNN	2199.90664074	***2175.84380047***	1814.25074459	***1962.19983593***	***2494.54625491***	2009.70066655	2279.85492271	2264.80346315	***1967.15610723***
		h = 3	IM	3930.12188231	3525.18899947	***3135.64827876***	4120.63273246	3172.47765360	4024.60139547	3483.94459452	***2999.19928686***	4304.09111327
			WM	3930.12188282	3525.18899895	3135.64835763	4120.63273193	3172.47765350	4024.60152132	***3483.94458913***	2999.19928737	4304.09111443
			KNN	***2823.67026037***	***3103.23461663***	3160.70092437	***3789.70133767***	***2983.02597678***	***3677.63501279***	3493.24349322	3183.27951140	***3054.67310112***

ตาราง 24 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 20% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				5								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 20 %	C.V. 5%	h = 1	IM	15.51402791	16.09184554	16.56385057	19.72339209	17.13513453	***18.79830136***	17.44696872	17.30190181	***14.79687415***
			WM	15.51402828	16.09184447	16.56385625	19.72339975	17.13513413	18.79830179	17.44697122	17.30190111	14.79687882
			KNN	***15.41294744***	***14.80339868***	***15.12675447***	***14.99967623***	***18.52935827***	21.24093288	***15.71802693***	***16.81073828***	17.38925630
		h = 2	IM	38.64043241	33.45243512	29.92997594	33.07274521	45.79007045	34.53094553	42.02290548	***31.19775477***	40.83262769
			WM	38.64043169	33.45243461	29.92997763	33.07274485	45.79009524	34.53094540	42.02290588	31.19775979	40.83262744
			KNN	***29.17318140***	***32.39816008***	***27.14137023***	***30.48618890***	***40.76848144***	***33.73934469***	***33.53894554***	31.68996047	***31.97658760***
		h = 3	IM	***46.95974471***	43.48220267	45.59841587	49.62427233	48.38813044	41.53361016	50.04117601	47.12806911	54.85201591
			WM	46.95974512	43.48220302	45.59841575	49.62427493	48.38812881	***41.53360985***	50.04117827	***47.12816615***	54.85201619
			KNN	51.90655950	***42.00527364***	***40.53498561***	***47.08967984***	***43.02069497***	42.99457650	***39.96692800***	52.99640217	***50.86791584***
	C.V. 25%	h = 1	IM	349.58342863	335.93308648	297.06133757	325.06387774	***257.81801824***	263.21676910	***247.37694242***	277.94474382	322.23493866
			WM	349.58342980	335.93308729	297.06134238	325.06387824	257.81801878	263.21676955	247.37694388	***277.94474015***	322.23493724
			KNN	***275.67473954***	***287.26512977***	***280.53972654***	***307.54332960***	263.90253548	***251.44182991***	263.43809254	291.20158528	***302.82836979***
		h = 2	IM	752.13063928	613.86728513	636.38810031	687.00647807	***688.46406523***	604.38417406	881.16630861	575.72711331	789.35156440
			WM	752.13063897	***613.86728477***	636.38809972	687.00647925	688.46406576	604.38417781	881.16630767	575.72718089	789.35156514
			KNN	***592.33195847***	721.70784607	***609.04380330***	***562.09960666***	858.92340242	***554.76110734***	***623.09210221***	***518.34217003***	***758.37531215***
		h = 3	IM	920.04265345	***945.74912719***	1245.23322069	1129.12071369	946.30182187	1192.50469224	1296.28014720	1080.32281229	1342.06476891
			WM	920.04265583	945.74918366	1245.23322641	1129.12071568	946.30182180	1192.50469898	1296.28014870	1080.32282957	1342.06492031
			KNN	***897.83110300***	1032.34250182	***1131.57760086***	***854.80676164***	***839.40304037***	***1129.02520679***	***833.36528850***	***1014.09468360***	***1115.00248191***
	C.V. 45%	h = 1	IM	877.15868412	859.00684352	744.23308275	***1240.15022236***	979.83646866	974.90924578	1663.50440919	1038.54973733	1081.93022015
			WM	***877.15868203***	859.00684385	744.23309077	1240.15022247	***979.83628969***	***974.90924500***	1663.50440545	1038.54973743	***1081.93022007***
			KNN	909.57383695	***853.59678385***	***738.50504538***	1376.43744586	1007.47493060	1135.06829623	***914.24408711***	***1024.12881001***	1217.18126657
		h = 2	IM	1980.45772074	***2049.49814174***	2038.20196474	2782.85175180	2353.12599004	2387.86721444	1993.12181181	2444.15710745	2209.93577722
			WM	1980.45772162	2049.49820279	2038.20196517	2782.85175356	2353.12602170	2387.86721631	1993.12181096	2444.15710640	***2209.93577567***
			KNN	***1729.40731889***	2234.82123478	***1821.38944873***	***2596.20726296***	***2222.66057583***	***2350.24091222***	***1831.41945352***	***2312.47037202***	2462.53077868
		h = 3	IM	3936.83226785	***3255.27756168***	***3098.55116906***	3303.75379756	3031.04813412	3532.08148678	3120.05772128	3917.01819652	3447.37789464
			WM	3936.83226723	3255.27756193	3098.55116948	3303.75379987	***3031.04813324***	3532.08148761	3120.05772081	3917.01821375	3447.37789607
			KNN	***2276.35745561***	3267.99795693	3689.20341802	***2610.97125402***	3215.28130239	***2741.41908706***	***3096.06430632***	***3819.82727510***	***3184.37343784***

ตาราง 25 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 3 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				3								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 30 %	C.V. 5%	h = 1	IM	20.80774772	20.93422366	***15.92806552***	18.58521053	20.29066013	21.42429723	20.57933439	16.64116401	19.57920035
			WM	20.80774697	20.93423608	15.92806670	18.58521272	20.29069109	21.42429831	20.57933897	16.64125722	19.57920572
			KNN	***19.57254472***	***16.93927916***	16.37790126	***17.45586159***	***17.40993653***	***18.53862814***	***18.70524915***	***15.16085118***	***16.38565996***
		h = 2	IM	35.39316199	35.97257015	31.70747586	33.51021231	34.10953486	45.41841274	31.71853284	29.51330957	***32.34584418***
			WM	35.39316426	35.97257909	31.70747212	33.51021287	34.10953551	45.41840462	31.71853073	29.51330352	32.34584986
			KNN	***27.57923179***	***25.07809050***	***27.68396709***	***26.86046166***	***23.66097696***	***43.86300131***	***31.71851938***	***29.45275720***	35.13671586
		h = 3	IM	***55.05915192***	42.97773406	41.00308523	52.07465049	55.64395417	44.46955045	52.83907016	53.47949189	46.14984201
			WM	55.05915540	42.97775153	41.00310518	***52.07464841***	55.64396188	44.46955491	52.83907274	53.47949583	46.14983825
			KNN	57.78012108	***39.05555963***	***40.35376579***	57.06655474	***42.74546495***	***43.19513289***	***61.88325697***	***52.46386870***	***42.00899489***
	C.V. 25%	h = 1	IM	321.82360438	326.67203979	***316.42692062***	333.69519012	368.32366247	257.65006098	372.58507457	290.47189417	***280.51588885***
			WM	321.82360753	326.67204386	316.42694669	333.69519323	368.32368306	257.65006018	372.58507834	290.47189535	280.51589067
			KNN	***266.46088965***	***291.66341601***	342.46624708	***238.44116647***	***347.41457912***	***250.01928962***	***301.09113237***	***288.97350231***	303.22561781
		h = 2	IM	665.34925828	763.45360264	810.52534649	625.47404969	648.05797782	***700.45059352***	***945.24635201***	818.11168390	607.23819012
			WM	665.34925951	763.45363294	***810.52534591***	625.47404624	648.05797566	700.45060194	945.24635260	818.11167298	607.23832880
			KNN	***604.38349521***	***542.91296996***	945.39743060	***618.01299639***	***575.15692367***	790.80937275	1081.07399169	***694.95032526***	***581.51728732***
		h = 3	IM	968.33464996	1207.25095814	1179.09942036	1087.27504191	1094.64547850	***868.61196605***	***1073.37022340***	1156.73825760	***1060.26653268***
			WM	968.33465528	1207.25096023	***1179.09941962***	1087.27504524	1094.64547984	868.61196848	1073.37022757	1156.73826942	1060.26664687
			KNN	***786.29040979***	***1119.70806025***	1195.48583624	***839.31563161***	***1062.34190788***	1013.41536979	1140.45250914	***912.56523656***	1178.99348003
	C.V. 45%	h = 1	IM	983.95622180	1340.13536829	1085.13818771	964.30006348	991.15120254	1142.98249269	944.33080815	960.34250237	1203.89935312
			WM	983.95622475	1340.13536698	1085.13819062	964.30005995	991.15120483	1142.98250571	944.33081209	960.34253189	1203.89935724
			KNN	***704.31290305***	***1254.22346108***	***858.76921572***	***767.34930211***	***757.98731423***	***1015.28430693***	***839.15892923***	***787.52970366***	***1089.93174270***
		h = 2	IM	3478.66963071	***2319.57250802***	2228.14675930	2165.58983367	2159.21082060	2858.86501466	2293.33536246	2538.36898098	2375.40170083
			WM	3478.66963451	2319.57251413	***2228.14675838***	2165.58983590	2159.21087627	2858.86502205	2293.33537178	2538.36898608	2375.40170083
			KNN	***1657.64992840***	2438.99224605	2476.24932932	***2032.25765477***	***2120.61337983***	***2567.35936953***	***1873.74371272***	***2279.84877497***	***2100.33756736***
		h = 3	IM	2821.37200239	3372.91734319	3436.78738103	3841.12749767	***3232.01915902***	***3047.18205757***	3516.23570078	***2822.75602777***	4286.79090172
			WM	***2821.37200172***	3372.91736402	***3436.78737697***	3841.12749453	3232.01917806	3047.18205785	3516.23570517	2822.75603391	4286.79090011
			KNN	3032.56763709	***2864.94753624***	4055.70190651	***3355.80081922***	3406.51933200	3365.13525385	***3340.97345780***	3694.14116259	***3861.64861751***

ตาราง 26 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 4 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				4								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 30 %	C.V. 5%	h = 1	IM	22.18077606	15.32886124	15.44024480	23.58483726	17.77277139	17.95343429	17.43234124	***18.79329545***	***18.22565052***
			WM	22.18078291	***15.32886012***	15.44032251	23.58484443	17.77282033	17.95343811	17.43234363	18.79330116	18.22583863
			KNN	***16.00746854***	16.94357318	***14.52323822***	***15.83155720***	***15.72580845***	***17.14391762***	***15.89847022***	18.89987858	20.62150725
		h = 2	IM	39.85292910	31.79655187	29.59095438	***35.36367002***	***31.17922454***	36.47172158	40.49833228	31.38750628	40.45777378
			WM	39.85293064	***31.79654975***	29.59101227	35.36367364	31.17924077	36.47171556	40.49833411	31.38750797	40.45777449
			KNN	***31.52870735***	36.24886361	***29.35307350***	37.35188615	32.55755095	***31.36117870***	***28.31359699***	***27.51748568***	***37.18102560***
		h = 3	IM	44.04419964	44.93007874	44.67045164	58.87876263	44.18881391	44.99432786	68.66451659	61.67516438	65.59005093
			WM	44.04420355	44.93008045	44.67044945	58.87876414	44.18882944	44.99433543	68.66451002	61.67519078	65.59005303
			KNN	***38.74657434***	***42.37523109***	***41.49839432***	***57.42462415***	***39.78010433***	***44.86882859***	***62.55110963***	***41.75588392***	***60.01428612***
	C.V. 25%	h = 1	IM	374.04098056	411.75460304	309.47871107	326.43324779	327.39118631	269.43077951	304.92678494	303.58498259	276.98140639
			WM	374.04098184	411.75462277	309.47879784	326.43324680	327.39118720	269.43077654	304.92678870	303.58498557	***276.98140389***
			KNN	***295.58516207***	***297.72839937***	***267.02635962***	***255.11976640***	***323.27917301***	***269.13987835***	***227.02257113***	***283.83677719***	320.56888944
		h = 2	IM	751.13173065	618.40459786	623.59106881	730.55086465	658.69489142	686.32469991	760.44332039	725.77027017	714.75390752
			WM	751.13172861	618.40459628	***623.59106855***	730.55086768	***658.69490037***	686.32466065	760.44332458	725.77031146	714.75391017
			KNN	***707.98317653***	***616.15624527***	641.27080657	***530.12094880***	673.86894613	***586.89527159***	***657.75089576***	***587.68072994***	***638.61159866***
		h = 3	IM	1069.93015275	1036.54902991	1066.95392638	941.49280824	917.49621602	***912.08539572***	1130.11946223	1082.01427598	1283.05739908
			WM	1069.93015597	1036.54903535	1066.95401005	***941.49280583***	***917.49621553***	912.08539808	1130.11946085	1082.01427767	1283.05740066
			KNN	***939.78909935***	***925.18871528***	***884.20239328***	1127.26346184	956.25663901	925.91763314	***990.60999513***	***942.91742958***	***1131.19467396***
	C.V. 45%	h = 1	IM	***955.01380221***	998.75908829	***1100.14640379***	***1280.03421792***	1024.35218857	896.52935087	1052.84679548	979.00578935	958.91545474
			WM	955.01380530	***998.75908710***	1100.14640407	1280.03422019	1024.35219293	***896.52935022***	1052.84680073	979.00581691	***958.91544156***
			KNN	972.46419580	1068.52537117	1176.46251320	1281.05022588	***975.63098866***	1116.26203404	***869.77274056***	***831.08914259***	1084.90045361
		h = 2	IM	2351.68904732	2380.90957939	2167.00941130	***2243.95118253***	***1861.02987790***	***1990.67285245***	2623.70973892	2393.14705145	***1901.19101156***
			WM	2351.68904672	2380.90957702	2167.00941036	2243.95118675	1861.02991151	1990.67293036	2623.70974235	2393.14705184	1901.19101487
			KNN	***1782.01563410***	***1890.35209456***	***2080.99801345***	2641.05801599	2012.86166293	2054.88559516	***2114.60931122***	***2069.02812212***	2039.39674834
		h = 3	IM	3690.88207389	5565.51215390	4089.50629093	3110.26797885	4068.01049754	3827.48805139	3850.52809262	4430.84139729	3493.71427628
			WM	3690.88207416	5565.51219886	4089.50629318	3110.26798301	***4068.01049542***	3827.48806886	3850.52809625	4430.84140315	***3493.71427477***
			KNN	***2754.21325519***	***3472.63090458***	***3121.87388214***	***2902.12784356***	4169.92421786	***3313.63965863***	***3125.18411423***	***3354.77314144***	3571.91516406

ตาราง 27 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ทั้ง 3 วิธี จำนวนข้อมูลสูญหาย 30% ในสถานการณ์ปัจจัย A มี 5 ระดับ ปัจจัย B มี 3,4,5 ระดับ และจำนวนบล็อก 3,4,5 บล็อก

ปัจจัย A				5								
ปัจจัย B				3			4			5		
จำนวนบล็อก				3	4	5	3	4	5	3	4	5
Missing 30 %	C.V. 5%	h = 1	IM	20.00732789	***19.46795575***	22.27307748	19.77567424	20.34231930	16.23574439	20.27570662	***20.91559883***	17.34956704
			WM	20.00732132	19.46795988	22.27316829	19.77567008	20.34231733	16.23573170	20.27570744	20.91562717	17.34957024
			KNN	***16.36229735***	20.65874828	***19.40279155***	***16.74413412***	***17.26735678***	***15.11408103***	***17.76492307***	20.97082382	***17.16624249***
		h = 2	IM	30.75019133	31.52018776	31.09376219	37.94566719	***32.70236734***	32.27328911	36.60501340	35.77062225	***35.36274122***
			WM	30.75019658	31.52019281	***31.09374481***	37.94567042	32.70237121	32.27329001	36.60501388	35.77061770	35.36291108
			KNN	***27.81166965***	***25.79133660***	31.27274134	***28.43123994***	34.05076248	***28.95298847***	***32.99899570***	***30.67729968***	36.51751045
		h = 3	IM	***46.58215046***	***48.23014250***	43.46179162	64.64874123	49.95552526	***46.60041416***	***49.76073068***	52.62990097	47.26467100
			WM	46.58215148	48.23015083	43.46179837	64.64874405	49.95553988	46.60042249	49.76073811	52.62996359	47.26467496
			KNN	55.13760581	55.59321755	***41.38164895***	***46.43228703***	***54.60261775***	47.78417283	51.13414686	***47.55681542***	***46.85081862***
	C.V. 25%	h = 1	IM	298.02017262	327.94757414	312.63438229	347.72601671	370.15143544	398.67010217	356.63566053	314.28788037	348.90231027
			WM	298.02017348	327.94758470	312.63438505	347.72602267	370.15145990	398.67016499	356.63566896	314.28790024	348.90234884
			KNN	***263.26015122***	***244.30088263***	***260.05579521***	***264.00421369***	***257.18361777***	***299.31752149***	***289.16347954***	***296.21967274***	***291.38064859***
		h = 2	IM	792.17685007	998.24861084	642.00209586	719.93630184	660.45003400	710.71890951	732.55867699	722.15434753	783.09301945
			WM	792.17684319	998.24861572	642.00210144	719.93630328	660.45006949	710.71891463	732.55867997	722.15438341	783.09302179
			KNN	***706.59578447***	***982.17781308***	***607.34468694***	***577.37341437***	***610.14453495***	***651.17494203***	***690.74371358***	***635.83619038***	***691.64905562***
		h = 3	IM	959.20433455	1154.56274149	1051.20219332	1182.23548297	***1047.24269553***	1221.98615544	1076.29863784	1482.70198871	1282.32362514
			WM	959.20434059	1154.56275663	1051.20219408	1182.23548301	1047.24271145	***1221.98613589***	1076.29864583	1482.70198554	1282.32362380
			KNN	***849.64076085***	***931.92017377***	***933.78186600***	***1012.23101757***	1209.03637269	1398.77116147	***1004.41147130***	***936.74050845***	***1256.32006657***
	C.V. 45%	h = 1	IM	1032.57267160	983.78075529	878.80052994	1716.40147484	1822.78544240	949.72604001	939.15076040	1044.73634269	888.13303080
			WM	1032.57267055	983.78075537	***878.80052923***	1716.40148189	1822.78545898	949.72604034	939.15076271	1044.73635480	***888.13302597***
			KNN	***980.39263906***	***925.48901385***	920.64145864	***1143.96519463***	***1049.39946523***	***863.43003858***	***863.29312056***	***909.30717952***	920.57591412
		h = 2	IM	***2234.59625153***	***1975.69558431***	2355.25585085	2255.63370990	3563.43611777	2198.12283017	2934.06417029	2361.79613040	2034.63052265
			WM	2234.59625439	1975.69558598	2355.25585654	2255.63370495	3563.43615920	2198.12288811	2934.06417526	2361.79617512	***2034.63052162***
			KNN	2288.79242509	2634.56559104	***2040.20180227***	***1996.89582742***	***2052.76105422***	***2124.90366589***	***2425.58523806***	***2025.22687151***	2082.71903345
		h = 3	IM	3330.95121226	3619.00221683	3477.46606612	4218.25861461	3030.22393337	4130.29021417	4035.54514028	4233.75347633	3493.65937393
			WM	3330.95121490	3619.00228487	3477.46608546	4218.25861920	3030.22397547	4130.29021780	4035.54514115	4233.75348448	3493.65935981
			KNN	***3252.49379509***	***3598.82235198***	***3219.20307250***	***3608.17780132***	***2657.53481863***	***2944.52450588***	***3468.55782846***	***3497.26550025***	***2928.21197503***

จากตาราง 19 – 27 แสดงค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนเบอร์อิมพิวเทชัน ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3 สามารถสรุปผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี จำแนกตามสถานการณ์ต่างๆ ได้ดังนี้

1. จำแนกตามจำนวนปัจจัย
2. จำแนกตามเปอร์เซ็นต์ข้อมูลสูญหาย
3. จำแนกตามค่าสัมประสิทธิ์ความผันแปร
4. จำแนกตามค่าคงที่

1. จำแนกตามจำนวนปัจจัย

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน จำแนกตามจำนวนปัจจัย พบว่า ระดับปัจจัยต่างๆ ในทุกสถานการณ์ วิธีประมาณค่าสูญหายโดยเคเนียเรส เนเบอร์อิมพิวเทชัน มีจำนวนค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล ในทุกระดับปัจจัย

2. จำแนกตามเปอร์เซ็นต์ข้อมูลสูญหาย

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน จำแนกตามเปอร์เซ็นต์ข้อมูลสูญหาย พบว่า วิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำในช่วงข้อมูลสูญหาย 10% และจะมีค่าเพิ่มขึ้นเรื่อยๆ เมื่อข้อมูลสูญหายเป็น 30% แสดงว่า วิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัลเหมาะสมกับข้อมูลที่มีค่าสูญหายอยู่ในช่วง 10% ส่วนการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดที่ใกล้เคียงกันทั้งจำนวนข้อมูลสูญหายเป็น 10%, 20%, 30% และมีค่าความคลาดเคลื่อน

สัมบูรณ์สูงสุดต่ำกว่าทั้ง 2 วิธีดังกล่าว นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่า วิธีประมาณค่าสูญหายโดยการวนซ้ำ และวิธีประมาณค่าสูญหายของวิลคินซัล และเหมาะสมกับจำนวนข้อมูลสูญหายเป็น 10%, 20% และ 30%

3. จำแนกตามค่าสัมประสิทธิ์ความผันแปร

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน จำแนกตามค่าสัมประสิทธิ์ความผันแปร พบว่า วิธีประมาณค่าสูญหายของวิลคินซัล มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำในช่วงค่าสัมประสิทธิ์ความผันแปรเท่ากับ 45% และมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดเพิ่มขึ้นเรื่อยๆ เมื่อค่าสัมประสิทธิ์ความผันแปรเท่ากับ 5% แสดงว่า วิธีประมาณค่าสูญหายของวิลคินซัลเหมาะสมกับข้อมูลที่มีค่าสัมประสิทธิ์ความผันแปรเท่ากับ 45% วิธีประมาณค่าสูญหายโดยการวนซ้ำให้ค่าเหมือนกับวิธีประมาณค่าสูญหายของวิลคินซัล แต่ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่า แสดงว่า วิธีประมาณค่าสูญหายโดยการวนซ้ำให้ค่าประมาณที่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายของวิลคินซัล ส่วนการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดที่ใกล้เคียงกันทั้งค่าสัมประสิทธิ์ความผันแปรเท่ากับ 5% ,25%, 45% และมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าทั้ง 2 วิธีดังกล่าว นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล และเหมาะสมกับจำนวนข้อมูลที่มีค่าสัมประสิทธิ์ความผันแปรเท่ากับ 5% ,25% และ45%

4. จำแนกตามค่าคงที่

ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหาย 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล และวิธีประมาณค่าสูญหายโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน จำแนกตามค่าคงที่ h พบว่า วิธีประมาณค่าสูญหายของวิลคินซัล มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในทุกระดับ วิธีประมาณค่าสูญหายโดยการวนซ้ำมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในค่าคงที่ h เท่ากับ1, 2 และมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดเพิ่มขึ้นเล็กน้อยเมื่อค่าคงที่ h เท่ากับ 3 ส่วนการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์

บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในทุกระดับ แสดงว่า ทั้ง 3 วิธี มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดของค่าคงที่ h ใกล้เคียงกันในทุกระดับ แต่การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล และเหมาะสมกับจำนวนข้อมูลที่มีค่าคงที่ h เท่ากับ 1,2 และ3

จากการจำลองข้อมูลในสถานการณ์ทั้งหมด 729 สถานการณ์ พบว่า ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าที่ไม่แตกต่างกันมาก โดยเฉพาะวิธีการประมาณค่าสูญหายโดยการวนซ้ำ และการประมาณค่าสูญหายวิลคินซัล ส่วนการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุด ดังนั้น ถ้าทำการทดลองในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ตามสถานการณ์ต่างๆ ที่กำหนด และมีรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) ควรเลือกใช้การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน เพราะให้ค่าใกล้เคียงกับค่าจริงมากที่สุด และสะดวกต่อการใช้งาน

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ในบทนี้ผู้วิจัยได้นำเสนอสาระสำคัญในภาพรวมของการวิจัยครั้งนี้ ในส่วนของสังเขปจุดมุ่งหมายและวิธีดำเนินการวิจัย สรุปผลการวิจัย อภิปรายผลการวิจัย และข้อเสนอแนะตามลำดับ โดยมีรายละเอียดดังนี้

สังเขปจุดมุ่งหมายและวิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาวิธีประมาณสูญหาย และเปรียบเทียบประสิทธิภาพของวิธีประมาณค่า ทั้ง 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ (Iterative Method) วิธีประมาณค่าสูญหายของวิลคินซัล (Wilkinson Method) และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน (K - Nearest Neighbor Imputation Method) สำหรับการวิจัยครั้งนี้ ผู้วิจัยทำการจำลองข้อมูลตามแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) และมีรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) โดยทำการศึกษาตามสถานการณ์ต่างๆ ดังนี้

1. จำนวนปัจจัย A มี 3 ลักษณะ คือ 3,4 และ 5 ระดับ
2. จำนวนปัจจัย B มี 3 ลักษณะ คือ 3,4 และ 5 ระดับ
3. จำนวนบล็อกเท่ากับ 3,4 และ 5 บล็อก
4. กำหนดสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ 45%
5. จำนวนข้อมูลสูญหายเป็น 10%, 20% และ 30%
6. กำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3

ได้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte carol simulation) โดยทำการจำลองซ้ำสถานการณ์ละ 1,000 รอบ โดยประสิทธิภาพในการประมาณค่าสูญหายในแต่ละวิธีนั้นพิจารณาจากค่าความคาดเคลื่อนสัมบูรณ์สูงสุด (Maximum absolute error) ซึ่งวิธีใดที่ให้ค่าความคาดเคลื่อน

สัมบูรณ์สูงสุดมีค่าต่ำสุด จะเป็นวิธีการประมาณค่าสูญหายที่ดีที่สุด เนื่องจากค่าที่ได้จากการประมาณค่าสูญหายนั้นมีค่าใกล้กับค่าจริงที่สูญหายไปมากที่สุด

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นข้อมูลที่ได้จากการจำลองสถานการณ์การทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน กำหนดให้ประชากรมีขนาดเท่ากับ 512 จำลองตัวแบบทางคณิตศาสตร์ เขียนด้วยโปรแกรม R

ดำเนินการสุ่มขนาดตัวอย่างในประชากรที่ได้จากการจำลองสถานการณ์ จำแนกตามปัจจัย A มี 3 ระดับคือ 3, 4 และ5 ปัจจัย B มี 3 ระดับ คือ 3, 4 และ5 และจำนวนบล็อก เท่ากับ 3, 4 และ5 เพราะฉะนั้นจำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 1,000 รอบ ซึ่งมีความเพียงพอที่จะศึกษาค่าประมาณพารามิเตอร์

การสร้างข้อมูลในงานวิจัยครั้งนี้ ได้จากการจำลองสถานการณ์ข้อมูลการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์(RCB) ด้วยเทคนิคมอนติคาร์โลซิมูเลชัน ผู้วิจัยทำการจำลองข้อมูลด้วยโปรแกรม R ดำเนินการตามขั้นตอน ดังนี้

1. การจำลองข้อมูลให้มีการแจกแจงแบบปกติ
2. การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง
3. การตรวจสอบรูปแบบของข้อมูลสูญหาย
4. การประมาณค่าที่สูญหาย
5. การเปรียบเทียบวิธีประมาณค่าสูญหาย

สรุปผลการวิจัย

การนำเสนอสรุปผลการวิจัย แบ่งออกเป็น 2 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์คุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์ ดังนี้

1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30%

1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง

ตอนที่ 2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3 จำแนกตามสถานการณ์ต่างๆ ดังนี้

2.1 จำแนกตามจำนวนระดับของปัจจัย

2.2 จำแนกตามเปอร์เซ็นต์ข้อมูลสูญหาย

2.3 จำแนกตามค่าสัมประสิทธิ์ความผันแปร

2.4 จำแนกตามค่าคงที่ h

ตอนที่ 1 ผลการวิเคราะห์คุณลักษณะของข้อมูลจำลองในแต่ละสถานการณ์

1.1 จำนวนข้อมูลสูญหาย 10%,20% และ30%

พบว่า จำนวนตัวอย่างมีขนาดเท่ากับ 27, 36, 45, 48, 60, 64, 75, 80, 100 และ125 จำนวนข้อมูลสูญหายคิดคำนวณจากจำนวนตัวอย่างเป็น 10%,20% และ30% ดังนั้น จำนวนตัวอย่างที่มีค่าน้อยที่สุดคือ 27 มีข้อมูลสูญหาย เท่ากับ 3 ค่า, 5 ค่า และ 8 ค่า ตามลำดับ จำนวนตัวอย่างที่มีค่ามากที่สุดคือ 125 มีข้อมูลสูญหาย เท่ากับ 13 ค่า, 25 ค่า และ 38 ค่า ตามลำดับ

1.2 ค่าจริงและค่าประมาณจากข้อมูลที่ได้จากการจำลอง

ทำการจำลองข้อมูลตามรูปแบบเชิงเส้นตรงของแผนการทดลองแบบ แฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบ RCB ทำการจำลองข้อมูลโดยกำหนดค่าเริ่มต้นให้ค่าเฉลี่ยมีค่าเท่ากับ 50 และประมาณค่าจากวิธีประมาณค่า 3 วิธี คือ วิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายวิลคินซัล และวิธีประมาณค่าโดยเคเนียเรส เนย์บอร์อิมพิวเทชัน ตามสถานการณ์ที่กำหนดทั้งหมด 729 สถานการณ์ พบว่า

กรณีข้อมูลสูญหาย 10% ข้อมูลที่จำลองมีค่าไม่แตกต่างจากค่าเฉลี่ยมากในค่าสัมประสิทธิ์ความผันแปรเท่ากับ 5% และจะมีค่าแตกต่างจากค่าเฉลี่ยมากขึ้นเมื่อความผันแปรมีค่ามากขึ้น ณ สัมประสิทธิ์ความผันแปรเท่ากับ 45% ข้อมูลจะมีการกระจายตัวมาก ทำให้ค่าที่ประมาณได้มีค่าแตกต่างจากค่าจริงมาก ค่าคงที่ h ในระดับค่าสัมประสิทธิ์ความผันแปรเดียวกันจะมีค่าใกล้เคียงกัน และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

กรณีข้อมูลสูญหาย 20% ข้อมูลที่จำลองมีค่าแตกต่างจากค่าเฉลี่ยมากขึ้นเมื่อความผันแปรมีค่ามากขึ้น เพราะข้อมูลมีการกระจายตัวมากขึ้น ค่าที่ประมาณได้จากทั้ง 3 วิธีมีค่าไม่แตกต่างกันเมื่อสัมประสิทธิ์ความผันแปรเท่ากับ 5 แต่จะมีค่าแตกต่างกันมากขึ้นเมื่อสัมประสิทธิ์ความผันแปรมากขึ้น

ค่าคงที่ h ในระดับค่าสัมประสิทธิ์ความผันแปรเดียวกันจะมีค่าใกล้เคียงกัน และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

กรณีข้อมูลสูญหาย 30% ณ สัมประสิทธิ์ความผันแปรเท่ากับ 5% ค่าคงที่ h ทั้ง 3 ระดับมีค่าใกล้เคียงกัน มีค่าไม่แตกต่างจากค่าเฉลี่ย ข้อมูลมีการกระจายตัวไม่มาก ค่าที่ประมาณได้กับค่าจริงมีค่าไม่แตกต่างกันมาก สัมประสิทธิ์ความผันแปรเท่ากับ 25% ค่าคงที่ h ทั้ง 3 ระดับ มีค่าแตกต่างกัน และจะมีค่ามากขึ้นเมื่อค่าคงที่ h เท่ากับ 3 ข้อมูลมีการกระจายตัวมากขึ้น ค่าที่ประมาณได้กับค่าจริงมีค่าแตกต่างกันมากขึ้น และสัมประสิทธิ์ความผันแปรเท่ากับ 45% ช่วงของข้อมูลกว้างมากขึ้น ทำให้ค่าที่จำลองได้มีการกระจายตัวมาก ค่าที่ประมาณได้กับค่าจริงมีค่าแตกต่างกันมากขึ้น และค่าที่ได้จากวิธีประมาณค่าโดยการวนซ้ำและจากวิธีวิลคินซัล มีค่าไม่แตกต่างกัน ในทุกสถานการณ์

ตอนที่ 2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย 10%, 20% และ 30% ค่าสัมประสิทธิ์ความผันแปรเป็น 5%, 25% และ45% และกำหนดให้ค่าคงที่ h เป็น 1, 2 และ 3 จำแนกตามสถานการณ์ต่าง ๆ ดังนี้

2.1 จำแนกตามจำนวนระดับของปัจจัย

เมื่อจำนวนระดับของปัจจัยเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดไม่ได้เพิ่มขึ้นตามจำนวนระดับของปัจจัยทั้ง 3 วิธี แสดงว่าระดับของปัจจัยไม่ส่งผลต่อค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ถ้าดูจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำสุด พบว่า ระดับปัจจัยต่างๆ ในทุกสถานการณ์ วิธีประมาณค่าสูญหายโดยเคเนียเรส เนเบอร์อิมพิวเทชัน มีจำนวนค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล ในทุกระดับปัจจัย

2.2 จำแนกตามเปอร์เซ็นต์ข้อมูลสูญหาย

เมื่อเปอร์เซ็นต์ข้อมูลสูญหายมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคิลซัล และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันมีค่าเพิ่มขึ้น เนื่องจากเมื่อข้อมูลมีการสูญหายเพิ่มมากขึ้น จะทำให้การประมาณค่าสูญหายมีความผิดพลาดมากขึ้น และเมื่อข้อมูลสูญหายเป็น 10%, 20% หรือ 30% วิธีประมาณค่าสูญหาย

โดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันจะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคิลซัลในสถานการณ์ต่างๆ มากที่สุด นั่นคือการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันจะให้ค่าส่วนใหญ่ใกล้ เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคิลซัล

2.3 จำแนกตามค่าสัมประสิทธิ์ความผันแปร

เมื่อค่าสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคิลซัล และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันมีค่าเพิ่มขึ้น เนื่องจากเมื่อสัมประสิทธิ์ความผันแปรของข้อมูลเพิ่มมากขึ้น จะทำให้ข้อมูลมีการกระจายมากขึ้น การประมาณค่าสูญหายจึงมีความผิดพลาดมากขึ้น เมื่อค่าสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้นวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันจะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคิลซัลในสถานการณ์ต่างๆ มากที่สุด ค่าสัมประสิทธิ์ความผันแปรที่ให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุดคือ ค่าสัมประสิทธิ์ความผันแปร เท่ากับ 5%

2.4 จำแนกตามค่าคงที่ h

เมื่อค่าคงที่ h มีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคิลซัล และวิธีประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันมีค่าเพิ่มขึ้นด้วย ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำสุดของวิธีประมาณค่าสูญหายทั้ง 3 วิธี จำแนกตามค่าคงที่ h พบว่า พบว่า วิธีประมาณค่าสูญหายของวิลคินซัล มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในทุกระดับ วิธีประมาณค่าสูญหายโดยการวนซ้ำมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในค่าคงที่ h เท่ากับ 1,2 และมีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดเพิ่มขึ้นเล็กน้อยเมื่อค่าคงที่ h เท่ากับ 3 ส่วนการประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดเท่ากันในทุกระดับ แสดงว่า ทั้ง 3 วิธี มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำที่สุดของค่าคงที่ h ใกล้เคียงกันในทุกระดับ แต่การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำ วิธีประมาณค่าสูญหายของวิลคินซัล นั่นคือ การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์

อิมพิวเทชัน จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าวิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายของวิลคินซัล และเหมาะสมกับจำนวนข้อมูลที่มีค่าคงที่ h เท่ากับ 1,2 และ3

จากการจำลองข้อมูลในสถานการณ์ทั้งหมด 729 สถานการณ์ พบว่า การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชัน มีค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด ดังนั้น ถ้าทำการทดลองในแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ตามสถานการณ์ต่างๆ ที่กำหนดและมีรูปแบบของข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) ควรเลือกใช้การประมาณค่าสูญหายโดยวิธีเคเนียเรส เนย์บอร์อิมพิวเทชันเพราะให้ค่าใกล้เคียงกับค่าจริงมากที่สุด และสะดวกต่อการใช้งาน

อภิปรายผล

การวิจัยครั้งนี้ได้ข้อค้นพบเกี่ยวกับประสิทธิภาพของวิธีประมาณค่าสูญหายในการวางแผนการทดลองแบบแฟคทอเรียล 2 ปัจจัยและลักษณะของข้อมูลที่ได้จากการจำลอง โดยการจำลองข้อมูลตามตัวแบบทางคณิตศาสตร์ด้วยเทคนิคมอนติคาร์โล เขียนด้วยโปรแกรม R ดังรายละเอียดต่อไปนี้

1. การเปรียบเทียบประสิทธิภาพของวิธีประมาณค่าสูญหายทั้ง 3 วิธี

หลักการสำคัญของการเปรียบเทียบวิธีประมาณค่าสูญหายทั้ง 3 วิธี คือ พิจารณาจากการคำนวณหาค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error : MAE) จากทั้ง 3 วิธี โดยค่าที่ประมาณได้จะถูกนำมาเปรียบเทียบกับค่าจริงที่ได้จากการจำลอง กล่าวคือ ในการจำลองข้อมูล 1 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ระหว่างค่าจริงที่ได้จากการจำลองกับค่าที่ได้จากการประมาณค่าสูญหายจำนวน 1 ค่า การศึกษาในครั้งนี้ทำการจำลองข้อมูล 1,000 รอบ จะได้ค่าสูงสุดของค่าสัมบูรณ์ 1,000 ค่า พิจารณาค่ามากที่สุดใน 1,000 ค่า เป็นตัวแทนของค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด ในแต่ละสถานการณ์และแต่ละวิธีมาเปรียบเทียบกัน วิธีใดให้ค่าความคลาดเคลื่อนสูงสุดมีค่าต่ำที่สุด แสดงว่า ค่าประมาณที่ได้มีค่าใกล้เคียงกับค่าจริงที่สูญหายไปมากที่สุด จะเป็นวิธีการประมาณค่าประมาณค่าสูญหายที่ดีที่สุด

ผลจากการวิเคราะห์พบว่า เมื่อพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) สามารถแบ่งกลุ่มวิธีประมาณค่าสูญหายได้เป็น 2 กลุ่ม คือ กลุ่ม 1 วิธีประมาณค่าสูญหายโดยการวนซ้ำและวิธีประมาณค่าสูญหายวิลคินซัล ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 2 วิธีมีค่าใกล้เคียง

กันมาก เป็นเพราะว่าทั้ง 2 วิธีนี้ใช้หลักการของวิธีกำลังสองน้อยที่สุดในการประมาณค่า วิธีนี้เป็นวิธีการสำคัญมีรากฐานมาจากทฤษฎีการประมาณเชิงเส้น (Theory of linear estimation) โดยไม่จำเป็นต้องทราบรูปแบบการแจกแจงความน่าจะเป็น แต่อาศัยผลต่างระหว่างค่าสังเกตและค่าคาดหมายเป็นสำคัญ (ประชุม สุวัตถี. 2553 หน้า 144) ซึ่งวิธีการประมาณค่าโดยการวนซ้ำเป็นการทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด

จาก $SS_{error} = SS_{tatal} - SS_{block} - SS_{treatment}$

วิธีการประมาณค่าโดยการวนซ้ำต้องการหาค่า Y เพื่อทำให้ SS_{error} มีค่าต่ำสุด วิธีนี้เริ่มต้นการประมาณค่าโดยการหาค่าเฉลี่ยของทรีตเมนต์และบล็อก แล้วนำค่าเฉลี่ยที่ได้ไปแทนค่าในสูตรทำซ้ำจนได้ค่าประมาณที่ไม่เปลี่ยนแปลง

ส่วนวิธีวิลคินซัลเป็นการประมาณค่ากำลังสองน้อยที่สุด โดยพิจารณาจากสมการเชิงซ้อนของค่าที่สูญหาย สิ่งที่สำคัญของวิธีนี้คือ การตั้งสมการสำหรับหาค่าที่หายไป จะต้องเป็นรูปแบบที่พิจารณาได้ง่าย และการแก้ไขปัญหาหรือหาคำตอบจากสมการด้วยการใช้แมทริกซ์อินเวอร์ชันที่ง่ายที่สุด (Matrix inversion) จากหลักการของกำลังสองน้อยที่สุดในการประมาณค่าของทั้ง 2 วิธีทำให้ค่าที่ประมาณได้มีค่าใกล้เคียงกัน

กลุ่ม 2 การประมาณค่าสูญโดยวิธีเคเนียเรส เนเบอร์อิมพิวเทชันเป็นวิธีที่ให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธีการประมาณค่าโดยการวนซ้ำและวิธีวิลคินซัล เป็นเพราะว่าวิธีนี้ใช้หลักการหาค่าประมาณโดยคำนวณจากระยะทางที่สั้นที่สุด ในการประมาณค่าต้องเลือกข้อมูลที่มีคุณสมบัติใกล้เคียงที่สุด K ตัว จากข้อมูลทั้งหมด โดยการคำนวณหาค่า Euclidian distance ระหว่างข้อมูลพิจารณากับข้อมูลที่มีความสมบูรณ์ เมื่อได้ค่าระยะห่างระหว่างจุดน้อยที่สุด K ตัว แล้วนำมาหาค่าเฉลี่ย แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลที่อยู่ใกล้ที่สุด

2. ลักษณะของข้อมูลที่ได้จากการจำลอง

2.1 จำนวนข้อมูลสูญหาย

ในการวิจัยในครั้งนี้ผู้วิจัยสร้างข้อมูลสูญหายจัดอยู่ในรูปแบบข้อมูลสูญหายแบบไม่เป็นระบบ (Arbitrary) คือไม่สามารถที่จะระบุจุดที่ข้อมูลสูญหายได้ และอันดับของตัวแปรก็ไม่มีสำคัญ การประมาณค่าข้อมูลสูญหาย จำนวนข้อมูลสูญหายมีความเป็นอย่างมาก เพราะจะทำให้ผลที่ได้มีความ

น่าเชื่อถือมากหรือน้อย ผลการวิจัยพบว่า ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของจำนวนข้อมูลสูญหาย 10% และ20% มีค่าใกล้เคียงกัน แต่เมื่อจำนวนข้อมูลสูญหายเพิ่มมากขึ้นเป็น 30% ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดเพิ่มขึ้นจากเดิมมาก แสดงว่า เมื่อข้อมูลสูญหายเป็น 30% ค่าที่ประมาณได้จะห่างจากค่าจริงมาก ซึ่งจะทำให้ค่าที่ประมาณได้ไม่เป็นตัวแทนของพารามิเตอร์ที่ต้องการศึกษา เพราะฉะนั้นจำนวนข้อมูลสูญหายที่ผู้วิจัยยอมรับได้ต้องอยู่ในช่วง 10% – 20% ซึ่งตรงกับงานวิจัยของเฟรดและลี (Fred; & Lii. 1998: online) ได้ศึกษาประเมินความถูกต้องของวิธีการจัดการข้อมูลสูญหายหลายวิธีภายใต้รูปแบบของการสูญหายที่แตกต่างกันโดยกำหนดจำนวนของข้อมูลสูญหายแบบสุ่มมีค่าต่ำสุดเป็น 10% และสูงสุดเท่ากับ 20% นอกจากนี้ยังมีงานวิจัยและเอกสารอื่นๆ ที่กล่าวว่าข้อมูลสูญหายจะอยู่ระหว่าง 5% ถึง 20% อยู่ในช่วงที่ผู้วิจัยยอมรับและสามารถนำไปวิเคราะห์ต่อไปได้

2.2 ค่าสัมประสิทธิ์ความผันแปร

ค่าสัมประสิทธิ์ความผันแปร เป็นค่าที่บ่งชี้ถึงคุณภาพของงานทดลองที่จะตัดสินใจว่างานนั้นได้ผลเป็นที่น่าเชื่อถือได้มากน้อยเพียงใด ค่าสัมประสิทธิ์ความผันแปรจะเป็นตัวบอกความแปรปรวนของการทดลองในเชิงสัมพัทธ์ที่แสดงค่าเป็นร้อยละของอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ย

จาก $$C.V. = \frac{\sigma}{\mu} \times 100$$

จะเห็นได้ว่า ส่วนเบี่ยงเบนมาตรฐานมีความสำคัญมาก เพราะส่วนเบี่ยงเบนมาตรฐานเป็นการวัดการกระจายของข้อมูลโดยใช้ผลรวมกำลังสองของส่วนเบี่ยงเบนมาตรฐานจากค่าเฉลี่ย แล้วถอดรากที่สอง ถ้าค่า C.V. มาก แสดงว่าส่วนเบี่ยงเบนมาตรฐานมีค่ามาก การทดลองนั้นมีการกระจายตัวของข้อมูลมาก ผลการทดลองจะมีความน่าเชื่อถือน้อย แต่ถ้าค่า C.V. น้อย แสดงว่าการทดลองนั้นมีการกระจายตัวของข้อมูลน้อย ผลการทดลองจะมีความน่าเชื่อถือมากกว่า ผลการวิจัยพบว่า ค่า C.V. ที่ระดับ 5% มีค่าความคลาดเคลื่อนสมบูรณ์สูงสุดต่ำสุด ที่ระดับ 25% และที่ระดับ 45% ค่าความคลาดเคลื่อนสมบูรณ์สูงสุดมีค่าเพิ่มขึ้นจากเดิมมาก แสดงว่าในการจำลองข้อมูลในแต่ละครั้ง ข้อมูลมีการกระจายตัวจากค่าเฉลี่ยมาก การทดลองมีความแปรปรวนมาก ทำให้ค่าที่ประมาณได้มีค่าห่างจากค่าจริงมาก ซึ่งเกินกว่าที่นักวิจัยจะยอมรับและให้ความน่าเชื่อถือ

2.3 ค่าคงที่ h

จากตัวแบบที่ใช้ในการศึกษาในครั้งนี้ อยู่ในรูปแบบเชิงเส้นตรงของแผนการทดลองแบบ a×b แฟคทอเรียลใน RCB

$$Y_{ijk} = \mu + B_k + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

กำหนดให้ μ เป็นพารามิเตอร์ที่ไม่ทราบค่า, B_k เป็นอิทธิพลของบล็อก k, α_i เป็นอิทธิพลของปัจจัย A ระดับที่ i, β_j เป็นอิทธิพลของปัจจัย B ระดับที่ j, $(\alpha\beta)_{ij}$ เป็นอิทธิพลร่วมกันของของปัจจัย A ระดับที่ i และของปัจจัย B ระดับที่ j โดยที่ $B_k, \alpha_i, \beta_j, (\alpha\beta)_{ij}, \varepsilon_{ijk}$ เป็นตัวแปรสุ่มที่เป็นอิสระกันและมีการแจกแจงแบบปกติ

เพราะฉะนั้น $E(B_k) = E(\alpha_i) = E(\beta_j) = E(\alpha\beta)_{ij} = E(\varepsilon_{ijk}) = 0$

และ $Var(B_k) = \sigma_B^2, Var(\alpha_i) = \sigma_\alpha^2, Var(\beta_j) = \sigma_\beta^2, Var(\alpha\beta)_{ij} = \sigma_{\alpha\beta}^2, Var(\varepsilon_{ijk}) = \sigma_\varepsilon^2$

ดังนั้น $E(y_{ijk}) = \mu, Var(y_{ijk}) = \sigma_B^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2$

นั่นคือ $y_{ijk} \sim N(\mu, \sigma_B^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2)$

จาก

$$c.v. = \frac{\sigma}{\mu} = \frac{\sqrt{\sigma_B^2 + \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2}}{\mu}$$

กำหนดให้

$\sigma_B^2 = \sigma_\alpha^2 = \sigma_\beta^2 = \sigma_{\alpha\beta}^2 = h\sigma_\varepsilon^2$ โดยที่ h เป็นจำนวนเต็มคงที่เท่ากับ 1, 2 และ 3

นั่นคือ

$$c.v. = \frac{\sqrt{h\sigma_B^2 + h\sigma_\alpha^2 + h\sigma_\beta^2 + h\sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2}}{\mu} = \frac{\sigma_\varepsilon\sqrt{4h+1}}{\mu}$$

ดังนั้นจะได้ว่า

$$\sigma_\varepsilon^2 = \frac{(c.v.(y_{ijk})\mu)^2}{4h+1}$$

คำนวณค่าความแปรปรวนจากค่าเฉลี่ยเท่ากับ 50 และค่า C.V. เท่ากับ 5,25,และ 45 การทดลองนี้จะมีความแปรปรวน $(\sigma_{y_{ijk}}^2)$ เท่ากับ 6.25, 156.25 และ 506.25 ตามลำดับ ค่า σ_ε^2 คือค่าความแปรปรวนของความคลาดเคลื่อนของการทดลอง และกำหนดให้ $h\sigma_\varepsilon^2$ โดยที่ h มีค่าเท่ากับ 1,2 และ3 เพื่อที่จะ

ศึกษาดูว่า ถ้าขนาดของค่าความแปรปรวนของความคลาดเคลื่อนของการทดลองเปลี่ยนไปจะให้ผลเป็นอย่างไร ผลจากการศึกษา พบว่า เมื่อค่าคงที่ h เพิ่มขึ้น ค่าข้อมูลที่จำลองก็มีความแปรปรวนของข้อมูลมากขึ้น ทำให้ค่าที่ประมาณได้ห่างจากค่าจริง เพราะฉะนั้น หากทำการทดลองจะต้องคำนึงค่าความคลาดเคลื่อนของการทดลองด้วย ถ้าค่าความคลาดเคลื่อนของการทดลองมีค่ามาก จะทำให้ผลที่ได้ไม่มีความน่าเชื่อถือ

ข้อเสนอแนะ

1. ข้อเสนอะแนะในการนำผลการวิจัยไปใช้ประโยชน์

1.1 สามารถนำวิธีการประมาณค่าสูญหายที่มีประสิทธิภาพตามเงื่อนไขการทดลองไปประยุกต์ใช้กับงานวิจัยที่วางแผนการทดลองการทดลองแบบแฟคทอเรียล 2 ปัจจัย ที่จัดทรีตเมนต์ในรูปแบบสุ่มในบล็อกสมบูรณ์ (RCB) ได้ทั้งทางด้านศึกษา เกษตรกรรม อุตสาหกรรม และการแพทย์ เป็นต้น

1.2 วิธีประมาณค่าโดยการวนซ้ำและการประมาณค่าสูญหายวิลคินซัล มีค่าใกล้เคียงกันมาก ดังนั้น ในการเลือกใช้วิธีประมาณค่าสูญหาย 2 วิธีนี้ จะเลือกวิธีใดก็ได้ เพราะให้ผลไม่ต่างกัน โดยที่วิธีประมาณค่าโดยการวนซ้ำ มีข้อดี คือ คิดคำนวณง่าย เพราะค่าเริ่มต้นจะคำนวณจากค่าเฉลี่ย ส่วนวิธีประมาณค่าสูญหายวิลคินซัล มีข้อดี คือ ประหยัดเวลาในการคำนวณ เพราะใช้สมการแมทริกซ์มาช่วยในการคำนวณ

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรทำการศึกษาการประมาณค่าสูญหายในแผนการทดลองแบบอื่นๆ เพราะงานวิจัยบางงานไม่ได้วางแผนการทดลองแบบแฟคทอเรียล ยังมีแผนการทดลองอื่นอีกมากที่เหมาะสมกับงานวิจัยอื่นๆ

2.2 ควรศึกษาตัวแปรอื่นๆ ที่ส่งผลต่อวิธีประมาณค่าสูญหาย เพื่อให้ค่าที่ประมาณได้มีค่าใกล้เคียงค่าจริงมากที่สุด

2.3 ควรศึกษาเปรียบเทียบผลการประมาณค่าสูญหายระหว่างค่าจริงและค่าที่ประมาณด้วยวิธีการอื่นๆ เพราะหากผลการวิเคราะห์ไม่แตกต่างกัน แล้วทำให้สะดวกและประหยัด เวลาในการดำเนินการ ก็จะเป็นประโยชน์สำหรับผู้ที่ต้องการวางแผนการทดลองแบบนี้

บรรณานุกรม

บรรณานุกรม

กนกทอง มหาวงศนันท์. (2550). *การวิเคราะห์ความสัมพันธ์ของตัวแปรที่มีอิทธิพลต่อความสามารถในการคิดอย่างมีวิจารณญาณของนักเรียนชั้นมัธยมศึกษาปีที่ 2*. ปริญญานิพนธ์ กศ.ม. (การวิจัยและสถิติทางการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

กัลยา วาณิชย์ปัญชา. (2549). *หลักสถิติ*. พิมพ์ครั้งที่ 8. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

จริยา แสงสุวรรณ. (2551). *การศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์การถดถอยพหุคูณ*. วิทยานิพนธ์ วท.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์. ถ่ายเอกสาร.

จิราวัลย์ จิตรถเวช. (2552). การวางแผนและ*การวิเคราะห์การทดลอง Design and Analysis of Experimental*. พิมพ์ครั้งที่ 1. กรุงเทพฯ : ไทยพัฒนรายวันการพิมพ์.

ชะไมพร ธรรมวัฒน์ไพศาล. (2522). *วิธีการประมาณค่าที่ขาดหายไปในการวิเคราะห์การถดถอย*. วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ชัยลิขิต สร้อยเพชรเกษม. (2556). *มุมมองความคลาดเคลื่อนของการทดลอง ANOVA ใน CR_p Design*. สืบค้นเมื่อ 1 กุมภาพันธ์ 2556, จาก www.edu.tsu.ac.th/major/old-eva/journal/ANOVA.pdf

ชุติมา ชัยมุสิก. (2533). *การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย*. วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ชูศักดิ์ จอมพุก. (2552). *สถิติ : การวางแผนการทดลองและการวิเคราะห์ข้อมูลในงานวิจัยด้านพืชด้วย R*. พิมพ์ครั้งที่ 1. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาตร์.

ชูศรี วงศ์รัตนะ. (2544). *เทคนิคการใช้สถิติเพื่อการวิจัย*. พิมพ์ครั้งที่ 8. กรุงเทพฯ : เทพเนรมิตรการพิมพ์

เชาว์ อินใย. (2547). *การพัฒนาวิธีการจัดการข้อมูลสูญหายแบบอีพีเอสเอสอีและการตรวจสอบความแม่นยำและอำนาจการทดสอบเปรียบเทียบกับวิธีอีเอ็มและลิสท์ไวท์: เทคนิคมอนติคาร์โล*. วิทยานิพนธ์ กศ.ด. (วิจัยและประเมินผลการศึกษา). พิษณุโลก : บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. ถ่ายเอกสาร.

ดวงฤดี เหง้าพรมนิล. (2545). *การวิเคราะห์ความแปรปรวนแบบปัจจัยเดียวเมื่อมีค่าสูญหายในแผนการทดลอง*. วิทยานิพนธ์ วท.ม. (สถิติประยุกต์). ขอนแก่น : บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น. ถ่ายเอกสาร.

ต่าย เซี่ยงฉี. (2534). *การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบเทเลอร์รูปปิรามิดที่มี รูปแบบ จำนวนขั้น และวิธีการให้คะแนนที่แตกต่างกัน โดยใช้วิธีมอนติ คาร์โล.* ปริญญานิพนธ์ กศ.ด. (วิจัยและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ถวัลย์ จันทร์เพ็ง. (2531). *การเปรียบเทียบความแม่นยำของการประมาณค่าข้อมูลสูญหายสามวิธีในกลุ่มตัวอย่างขนาดเล็ก.*วิทยานิพนธ์ ค.ม.(วิจัยการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

นราทิพย์ จั่นสกุล. (2552). *แผนแบบการทดลองและการวิเคราะห์ข้อมูลโดยใช้โปรแกรม R.* พิมพ์ครั้งที่ 1. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นวลอนงค์ บุญฤทธิพงศ์. (2552).*ระเบียบวิจัยทางการศึกษา.* พิมพ์ครั้งที่ 1. กรุงเทพฯ : บริษัทจุดทอง จำกัด.

ณรงค์ โพธิ. (2554). *การประมาณค่าข้อมูลสูญหายด้วยวิธีแบบจำลองสมการโครงสร้าง.* วิทยานิพนธ์ ปรัชญาดุษฎีบัณฑิต (เทคโนโลยีสารสนเทศ). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. ถ่ายเอกสาร.

บวรวรรณ ดิเรกโภค. (2543). *การประยุกต์ใช้วิธีการใส่ค่าหลายค่าแทนข้อมูลที่สูญหายแต่ละค่าในการวิเคราะห์ข้อมูลอุบัติเหตุผู้ขับขี่จักรยานยนต์.* วิทยานิพนธ์ วท.ม. (ชีวสถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล. ถ่ายเอกสาร.

ประชุม สุวัตถี. (2553). *ทฤษฎีการอนุมานเชิงสถิติ.* พิมพ์ครั้งที่ 3. กรุงเทพฯ : สำนักงานกิจการโรงพิมพ์ องค์การสงเคราะห์ทหารผ่านศึก.

ประพจน์ ดำรงสุทธิพงศ์.(2546). *การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มบล็อคสมบูรณ์.* วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ปราณี ศรีภา. (2532). *ศึกษาวิธีเปรียบเทียบวิธีวิเคราะห์เมื่อมีค่าสูญหายในบล็อคใดบล็อคหนึ่งในแผนการทดลองบล็อคไม่สมบูรณ์.* วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ฝ่ายวิชาการสถิติ กรมวิชาการเกษตร. (2544). *เอกสารประกอบคำบรรยายการฝึกอบรมสถิติ หลักสูตรการใช้สถิติในงานวิจัยเกษตร เล่ม 1.* กรุงเทพฯ : กรมวิชาการเกษตร กระทรวงเกษตรและสหกรณ์

พรวุฒิ คำแก้ว. (2546). *ผลการใช้บทเรียนคอมพิวเตอร์มัลติมีเดีย 3 รูปแบบ ที่มีต่อผลสัมฤทธิ์ทางการเรียน ความคงทนในการเรียนรู้ และเจตคติต่อบทเรียน ของ นักเรียนที่มีระดับความสามารถต่างกัน 3 ระดับ.* ปริญญานิพนธ์ กศ.ม.(เทคโนโลยีการศึกษา).กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

พรศิริ หมื่นไชยศรี. (2529). *การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวิเคราะห์ตัวแปรพหุ.* วิทยานิพนธ์ สต.ม.(สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

พิสมัย หาญมงคลพิพัฒน์. (2546). *หลักสถิติ1.* พิมพ์ครั้งที่ 4. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.

พิศิษฐ์ ตัณฑวณิช. (2553). *สถิติเพื่องานวิจัยทางการศึกษา.* พิมพ์ครั้งที่ 4. กรุงเทพฯ : บุ๊ค พอยท์.

พัลลภ พิริยะสุรวงศ์. (2542). *การออกแบบและพัฒนามัลติมีเดียแบบฝึกโดยใช้รูปแบบการควบคุมการเรียนต่างกัน.* ปริญญานิพนธ์ กศ.ด.(เทคโนโลยีการศึกษา).กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

รุ่งกานต์ กาใจคำ. (2540). *การประมาณค่าสูญหายจากแผนการทดลอง.* การค้นคว้าเชิงอิสระเชิงวิทยานิพนธ์ วท.ม. (สถิติประยุกต์). เชียงใหม่ : บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่. ถ่ายเอกสาร.

รุ่งโรจน์ ศรีจันทร์แก้ว.(2547).*ศึกษาเปรียบเทียบอำนาจการทดสอบ (Power of test) ของผลการวิเคราะห์ข้อมูลด้วยแบบแผนการวิเคราะห์แบบกลุ่มสุ่ม(RBD) กับแบบแผนการวิเคราะห์ความแปรปรวนร่วม (ANCOVA).*ปริญญานิพนธ์ กศ.ม. (การวิจัยและสถิติทางการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ลำปาง แสนจันทร์. (2549). *การควบคุมคุณภาพเชิงสถิติ.* พิมพ์ครั้งที่ 1. เชียงใหม่ : สถานบริการวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเชียงใหม่.

เลิศลักษณ์ กลิ่นหอม. (2532). *การศึกษาความคลาดเคลื่อนชนิดที่ 1 และอำนาจของการทดสอบในแบบแผนการทดลองแบบสุ่มภายในบล็อกที่มีปฏิสัมพันธ์ระหว่างสิ่งทดลองกับบล็อก.* ปริญญานิพนธ์ กศ.ด. (การวิจัยและพัฒนาหลักสูตร). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

วารุณี ตรีบำรุงศักดิ์. (2538). *การพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุเมื่อตัวแปรตามมีค่าสูญหาย.* วิทยานิพนธ์ สต.ม.(สถิติ) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ศุภลักษณ์ กรรณิกา. (2549). *การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน.* วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

สมชัย ยืนนาน. (2528). *การศึกษาโดยวิธีมอนติคาร์โลเปรียบเทียบอำนาจของการทดสอบของการเท่ากันของความแปรปรวนระหว่างประชากรสองกลุ่ม.* วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

สายชล สินสมบูรณ์ทอง. (2546). *สถิติวิเคราะห์ : Statistical Analysis.* พิมพ์ครั้งที่ 2. กรุงเทพฯ : โครงการตำรา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.

-------------. (2549). *สถิติกับการวางแผนการทดลองทางการเกษตร.* พิมพ์ครั้งที่ 4. กรุงเทพฯ : จามจุรีโปรดักท์.

สุนันทา วีรกุลเทวัญ. (2544). *การวิเคราะห์ข้อมูลทางสถิติกรณีที่ข้อมูลไม่ครบถ้วน.* วารสารศึกษาศาสตร์มหาวิทยาลัยขอนแก่น. 25(1), 16 – 20.

สุวัฒน์ นิยมไทย. (2531). *ผลการเรียนรู้วิชาฟิสิกส์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจากคอมพิวเตอร์ช่วยสอนในลักษณะกลุ่มย่อย ซึ่งมีขนาดของกลุ่มต่างกัน.* ปริญญานิพนธ์ กศ.ม. (เทคโนโลยีทางการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

องอาจ นัยพัฒน์. (2548). *วิธีวิทยาการวิจัยเชิงปริมาณและเชิงคุณภาพทางพฤติกรรมศาสตร์และสังคมศาสตร์.*พิมพ์ครั้งที่ 1. กรุงเทพฯ : สามลดา.

-------------. (2554). *การออกแบบการวิจัย : วิธีการเชิงปริมาณ เชิงคุณภาพ และผสมผสานวิธีการ.* พิมพ์ครั้งที่ 2. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

อรุณี เต๊ะอ้วน.(2550). *การศึกษาเปรียบเทียบผลการเรียนรู้ สาระดนตรี ด้วยบทเรียนคอมพิวเตอร์มัลติมีเดียสองรูปแบบที่ใช้แบบแผนการทดลองต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวิจัยและสถิติทางการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อุษณีย์ วงศ์อามาตย์.(2555). *การเปรียบเทียบวิธีการประมาณค่าสูญหายแบบนอนอิกนอร์เรเบิลในการวิเคราะห์การถดถอยเชิงเส้นพหุ.* วิทยานิพนธ์ สต.ม. (สถิติ). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

Cochran, William G.; & Cox, Gertrude M. (1987). *Expermental Designs.* 2nd ed. New York: john Wiley & Sons.

Crawford, L., L. Tennstedt; & B. Mckinlay. (1995). Comparison of Analytic Method for Non-Random Missingness of Outcome Data. *J Clin Epidemiol*. 48: 209-219.

Dodge Yadolah. (1958). *Analysis of Experiments with Missing Data.* New York: john Wiley&Sons.

Enders, C.K. (2001). The Performance of the Full Information Maximum Likelihoods Estimator in Multiple Regression Models with Missing Data. *Educational and Psychological Measurement*. 61: 713-740.

Fred, S.; & Lii, S. (1998). *Systematic data loss in HRM seting : a Monte Carlo analysis human respurces management.* Retrived May 22, 2002, from : http://www.findarticles.com

Heitjan, D.B.; & R.J.A. Little. (1991). Multiple Imputation for the Fetal Accident Reporting System. *Appl Stat*. 40: 13-29.

Huang, R.; & K.C. Carriere. (2006). *Comparison of Methods for Incomplete Repeated Measures Data Analysis in Small Samples.* Statistical Planning and Inference. 136: 235-247.

Little, R. J.A.; & Rubin, D. B. (1987). *Statistical Analysis with Missing Data* . New York: Wiley.

Montgomery, Douglas C. (2005). *Design and analysis of experiments.* 6th ed. New York: john Wiley & Sons.

Roth, P.L. (1994). Missing Data : A conceptual review for applied psychologists. *Personnel Psychology*, 47 : 537-560.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data* . New York: Chapman&Hall.

Viragoontavan, S. (September 15,2000). *Comparing six missing data method within the discriminant analysis context : A Monte Carlo Study.* Retrieved September 26, 2002, from : http://thailis.uni.net.th/doa/detail.nsp.

Wilkinson, G.N. (June 1958). *Estimation of Missing Value for the Analysis of Incomplete data.* Biometrics 14 : 257 – 286.

Troyanskaya O; et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001,17 : 520 – 525.

ภาคผนวก

ภาคผนวก ก

ตัวอย่างการใช้โปรแกรม R การประมาณค่าสูญหายวิธีวิลคินซัล

กำหนดตัวแปร

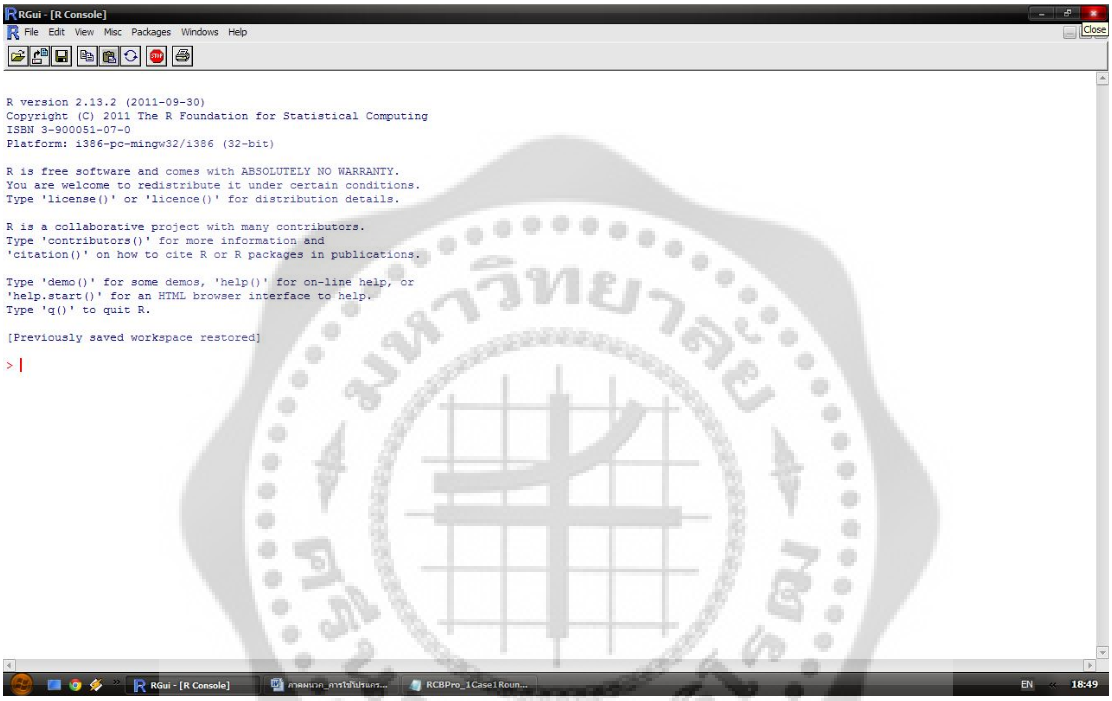
n_alpha.arr	แทน	จำนวนปัจจัย A	มีค่าเท่ากับ 3,4,5
n_beta.arr	แทน	จำนวนปัจจัย B	มีค่าเท่ากับ 3,4,5
nb.arr	แทน	จำนวน block	มีค่าเท่ากับ 3,4,5
iter	แทน	จำนวนรอบ	มีค่าเท่ากับ 1000 รอบ
h.arr	แทน	ค่าคงที่ h	มีค่าเท่ากับ 1,2,3
mu	แทน	ค่าเฉลี่ยของประชากร	มีค่าเท่ากับ 50
mi.arr	แทน	จำนวนข้อมูลสูญหาย	มีค่าเท่ากับ 5%,10%,15%
cv.arr	แทน	สัมประสิทธิ์ความแปรผัน	มีค่าเท่ากับ 5%,25%,45%
numofcase	แทน	จำนวนการทดลองทั้งหมด	
nm	แทน	จำนวนข้อมูลทั้งหมดของการประมาณค่า 1 วิธี	
MAETABLE	แทน	แมทริกซ์ที่เก็บข้อมูลที่ได้จากการจำลองข้อมูล	
Alpha	แทน	อิทธิพลของปัจจัย A	
Beta	แทน	อิทธิพลของปัจจัย B	
Block	แทน	อิทธิพลของบล็อก	
e	แทน	ความคลาดเคลื่อนของการทดลอง	
Y_pop	แทน	จำนวนประชากร	
Y	แทน	จำนวนตัวอย่าง	
miss	แทน	สุ่มตำแหน่งข้อมูลสูญหาย	
YM	แทน	ข้อมูลที่ตัดออกเพื่อเก็บไว้เปรียบเทียบกับข้อมูลที่ได้จากการประมาณค่า	
X	แทน	เก็บตำแหน่งข้อมูลสูญหายในแถวและคอลัมน์	
coef_x	แทน	แมทริกซ์สัมประสิทธิ์ของสมการ	
crow	แทน	ตัวแปรทำซ้ำ เพื่อหาผลรวมด้านแถว	
ccol	แทน	ตัวแปรทำซ้ำ เพื่อหาผลรวมด้านคอลัมน์	
ct	แทน	ผลรวมของทรีตเมนต์ที่มีค่าสูญหาย	
cb	แทน	ผลรวมของบล็อกที่มีค่าสูญหาย	
cg	แทน	ผลรวมทั้งหมด	
A	แทน	แมทริกซ์สัมประสิทธิ์ของสมการ	
B	แทน	แมทริกซ์คำตอบของสมการ	
Resulte	แทน	แมทริกซ์ค่าประมาณ	

Estimation แทน ค่าประมาณโดยวิธีวิลคิลซัล

sigma2 แทน ค่าความแปรปรวน

ตัวอย่างขั้นตอนการใช้โปรแกรม R

1. เปิดโปรแกรม R หน้าจอของโปรแกรม R ดังรูป



2. กำหนดค่าเริ่มต้นให้กับโปรแกรม ค่าที่กำหนดให้กับโปรแกรมเป็นตัวอย่างในการคำนวณเท่านั้น เพื่อให้ง่ายต่อความเข้าใจโปรแกรม R

```
>
> iter=1000;
> mi=c(0.1);
> cv=c(0.05);
> h=c(1);
> n_alpha=c(3);
> n_beta=c(3);
> nb=c(3);
> mu=50;
> p=8;
> nt=n_alpha*n_beta;
> sigma2=(cv*mu)^2/(4*h+1);
> Alpha=rnorm(p,0,h*sqrt(sigma2));
> Beta=rnorm(p,0,h*sqrt(sigma2));
> Block=rnorm(p,0,h*sqrt(sigma2));
> e=rnorm( p*p*p,0,sqrt(sigma2));
> y=matrix(0,p*p,p);
> k=1;
```

กำหนดค่าเริ่มต้น

3. การจำลองข้อมูล (Simulate data)

```
> Alpha=rnorm(p,0,h*sqrt(sigma2));
> Beta=rnorm(p,0,h*sqrt(sigma2));
> Block=rnorm(p,0,h*sqrt(sigma2));
> e=rnorm( p*p*p,0,sqrt(sigma2));
> y=matrix(0,p*p,p);
> k=1;
> for(i in 1:p){
+ for(ii in 1:p){
+ for(j in 1:p){
+ y[k,j]=mu+Alpha[i]+Beta[ii]+Alpha[i]*Beta[ii]+Block[j];
+ }
+ k=k+1;
+ }
+
+ }
>
> Pop<-t(y)+e;
> Y_pop = t(Pop);
> ns=n_alpha*n_beta*nb;
> sam_y=sample(Y_pop,ns,replace=TRUE);
> Y =matrix(sam_y,n_alpha*n_beta,nb);
> nm=round(n_alpha*n_beta*nb*mi);
>
```

กำหนดการแจกแจงแบบปกติให้กับตัวแปร โดยใช้คำสั่ง rnorm

ตัวแบบที่ศึกษา

4. ข้อมูลประชากรที่ได้จากการจำลอง

```
RGui (64-bit) - [R Console]
File  Edit  View  Misc  Packages  Windows  Help

> Y_pop
          [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]
 [1,] 51.68959 51.32336 53.40892 51.65687 52.85720 52.88750 53.59186 50.06901
 [2,] 51.45018 52.31151 49.48649 52.85324 50.38942 50.63102 50.89036 46.28479
 [3,] 46.32165 44.41789 44.39311 42.80405 44.19935 43.05864 46.34991 41.68045
 [4,] 51.10037 49.96501 48.95012 49.52476 50.91373 49.45593 48.60358 46.68217
 [5,] 51.74005 52.33371 50.30864 52.36758 53.00167 51.39233 52.62263 49.59006
 [6,] 48.07792 47.47371 48.03405 49.63483 47.33359 49.29524 47.73986 48.43007
 [7,] 50.59591 52.30992 50.72469 55.57863 52.89900 50.20455 54.42912 49.83816
 [8,] 46.40640 46.16319 46.80813 44.61890 48.41735 46.18544 46.53925 43.26995
 [9,] 49.83996 50.71410 48.34446 49.16954 49.39576 48.30979 50.37556 48.15369
[10,] 50.67964 49.43196 48.74464 49.07539 49.12130 48.31650 50.43425 47.36485
[11,] 49.62186 49.95225 48.75139 49.31892 48.83267 48.49108 48.22088 45.60728
[12,] 50.42720 49.56055 47.87751 48.11023 50.44058 50.44385 50.56085 49.15660
[13,] 50.69388 49.68266 47.73926 49.43408 47.97191 50.47586 50.28362 46.65246
[14,] 50.04531 48.18115 47.51578 48.84503 48.86988 47.72474 49.52223 46.03257
[15,] 53.27734 50.25979 50.88960 49.87481 47.58053 50.59133 51.57868 46.61254
[16,] 48.68290 48.13854 46.86876 48.41677 49.90962 47.01804 50.17195 47.19191
[17,] 52.05048 54.04739 50.30101 50.75397 52.25625 53.01192 51.69399 48.01928
[18,] 50.43879 49.33656 52.28746 48.58984 50.15633 49.60671 49.98593 48.53855
[19,] 44.89676 43.89189 43.65197 46.06853 44.77419 45.09771 46.51878 42.05077
[20,] 50.22112 52.50264 49.58277 52.20470 51.46618 50.30622 52.81762 49.16523
[21,] 51.10765 52.38235 48.79134 51.85943 53.28866 50.56254 53.99921 49.22380
[22,] 47.58894 48.10830 46.35920 48.09523 47.24733 49.14224 48.40017 46.43933
[23,] 51.53914 52.05752 52.31032 53.42983 51.68402 52.00978 54.37420 51.40116
[24,] 44.44789 44.95328 44.91110 45.95903 47.62378 46.36687 46.40242 43.80284
[25,] 52.09761 50.49052 53.05054 52.36285 51.97498 51.54346 53.37081 48.51582
[26,] 50.73934 49.87901 50.30185 50.93931 49.62996 50.08905 52.26251 48.93789
[27,] 44.20767 44.13172 44.84253 45.02637 45.85934 45.06816 48.06960 43.25147
[28,] 49.58532 49.98276 49.87253 49.13942 49.92214 47.49044 49.56453 47.92510
[29,] 52.71961 48.10528 48.79732 51.42206 51.10228 51.33198 49.74125 50.03093
[30,] 48.24858 48.54715 49.26882 48.04141 47.96482 48.15954 48.80102 47.45381
[31,] 53.37336 50.36273 50.88122 52.81824 53.37477 52.81735 52.61008 50.55621
[32,] 46.49539 44.81607 45.97180 46.90592 48.14009 46.20934 47.65328 42.86195
[33,] 50.49249 52.00900 49.97075 52.08820 51.56470 51.10982 51.76377 48.68483
[34,] 50.73925 48.82201 48.01123 49.93257 51.58349 48.88624 50.64604 49.19846
[35,] 43.34018 48.59256 47.06651 45.38626 46.99652 46.35218 45.92293 43.35633
[36,] 47.28868 48.59664 47.08402 50.52089 52.89901 49.21358 50.93597 47.88641
[37,] 49.56507 48.46801 50.43660 50.40873 49.44509 50.55897 52.51153 47.57545
```

5. สุ่มตัวอย่างจากข้อมูลประชากร ตามแต่ละสถานการณ์

```
> Y
         [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765 43.85466
 [7,] 49.33419 52.51321 49.08462
 [8,] 45.46650 48.54174 46.61416
 [9,] 49.32943 50.65631 43.26966
> 
```

6. สุ่มตัดข้อมูลสูญหายและตรวจสอบรูปแบบข้อมูลสูญหาย

```
>
> miss <- sample( seq_len(nt*nb ), nm );
> miss
[1] 17  8 24  9
> YM <- Y;
> YM[miss]
[1] 48.54174 45.46650 43.85466 49.32943
> YM[miss] <- NA;
> X=which(is.na(YM),arr.ind=TRUE);
> X
     row col
[1,]   8   1
[2,]   9   1
[3,]   8   2
[4,]   6   3
> print(YM);
          [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765       NA
 [7,] 49.33419 52.51321 49.08462
 [8,]       NA       NA 46.61416
 [9,]       NA 50.65631 43.26966
>
> |
```

สุ่มตำแหน่งที่ข้อมูลสูญหาย

ตำแหน่งที่ข้อมูลหายคือ ตำแหน่ง 10, 12, 9 ,18

ข้อมูลจริงที่ตัดออก แต่เก็บไว้เพื่อเปรียบเทียบกับข้อมูลที่ได้จากการประมาณค่า

แถวและคอลัมน์ ในตำแหน่งที่ข้อมูลสูญหาย

ข้อมูลที่สุ่มตัด เป็นข้อมูลสูญหายเรียบร้อยแล้ว

7. ประมาณค่าสูญหายวิธีวิลคินซัล

```
>
> #Wilkinson Method
> coef_x=matrix(0,nm,nm);
> c=array(0,nm);
> for(i in 1:nm){
+ crow=X[i,'row'];
+ ccol=X[i,'col'];
```

กำหนดตัวแปร

```
+
+ ct=nt*sum(YM[crow,],na.rm = TRUE);
```

ผลรวมของข้อมูลในทรีตเมนต์ที่มีค่าสูญหาย

```
+ cb=nb*sum(YM[,ccol],na.rm = TRUE);
```

ผลรวมของข้อมูลในบล็อกทีมีค่าสูญหาย

```
+ cg=sum(YM,na.rm = TRUE);
```

ผลรวมทั้งหมด

```
+ c[i]=cb+ct-cg;
```

สูตรที่ใช้ในการคำนวณ

```
+ x_index=i;
+ coef_x[i,x_index]=(nt*nb)-nt-nb+1;
+ for(j in 1:nm){
+     if(j!=x_index){
+ if(X[j,'row']==crow){
+    coef_x[i,j]=nt;
+ }
+ if(X[j,'col']==ccol){
+    coef_x[i,j]=coef_x[i,j]+nb;
+ }
+ coef_x[i,j]=(coef_x[i,j]-1)*(-1);
+     }
+ }
+ }
```

ขั้นตอนการทำซ้ำ เพื่อสร้าง แมทริกซ์ สัมประสิทธิ์ของสมการ

8. แสดงค่าของตัวแปรแต่ละตัว

```
>
> A=coef_x;
> A
     [,1] [,2] [,3] [,4]
[1,]   16    1    1    1
[2,]    1   16    1    1
[3,]    1    1   16   -2
[4,]    1    1   -2   16
```

แมทริกซ์ A คือ แมทริกซ์สัมประสิทธิ์ของสมการ

```
> B=as.matrix(c);
> B
          [,1]
[1,]  966.6475
[2,] 1069.4636
[3,]  798.3896
[4,]  744.7852
```

แมทริกซ์ B คือ แมทริกซ์คำตอบของสมการ

```
> Results=try(solve(A)%*%B,TRUE);
```

แมทริกซ์ A^{-1} คูณ แมทริกซ์ B

```
> Results
         [,1]
[1,] 50.88734
[2,] 57.74174
[3,] 48.84318
[4,] 45.86516
```

ค่าที่ได้จากการประมาณโดยวิธีวิลคินซัล

```
> Estimation=YM;
> Estimation
          [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765       NA
 [7,] 49.33419 52.51321 49.08462
 [8,]       NA       NA 46.61416
 [9,]       NA 50.65631 43.26966
> Estimation[X]=Results;
> Estimation[X]
[1] 50.88734 57.74174 48.84318 45.86516
>
```

9. แสดงผลรันโปรแกรม R

```
>
> #Show Results;
> print(Y);
         [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765 43.85466
 [7,] 49.33419 52.51321 49.08462
 [8,] 45.46650 48.54174 46.61416
 [9,] 49.32943 50.65631 43.26966
```

ข้อมูลที่ได้จากการจำลองข้อมูล

```
> print(YM);
         [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765       NA
 [7,] 49.33419 52.51321 49.08462
 [8,]       NA       NA 46.61416
 [9,]       NA 50.65631 43.26966
```

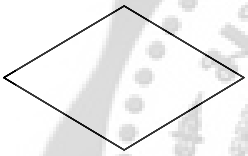
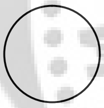
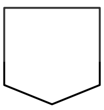
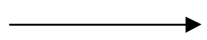
ตำแหน่งที่ข้อมูลสูญหาย

```
> print(Estimation);
         [,1]     [,2]     [,3]
 [1,] 45.30865 49.02350 44.92077
 [2,] 55.15693 54.67817 50.72299
 [3,] 55.69416 58.77031 53.73152
 [4,] 52.62105 51.29998 47.08237
 [5,] 45.80658 51.70853 46.09136
 [6,] 48.10733 49.85765 45.86516
 [7,] 49.33419 52.51321 49.08462
 [8,] 50.88734 48.84318 46.61416
 [9,] 57.74174 50.65631 43.26966
>
```

ข้อมูลที่ได้จากการประมาณค่าโดยวิธี
วิลคิลซัล

ภาคผนวก ข

สัญลักษณ์ของ Flowchart

สัญลักษณ์	ความหมาย
	จุดเริ่มต้น และจุดสิ้นสุดของโปรแกรม
	การประมวลผล การทำงาน การคิดคำนวณ
	รับข้อมูล/ส่งออกข้อมูล
	เงื่อนไข เป็นจุดที่มีเงื่อนไขให้เลือกทำ หรือเช็คค่าตามเงื่อนไข
	จุดเชื่อมต่อ
	การแสดงผลทางเครื่องพิมพ์
	ขึ้นหน้าถัดไป
	เส้นทางการไหลของโปรแกรม เพื่อช่วยในการเชื่อมแต่ละขั้นของโปรแกรม

ภาคผนวก ค

โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูล

```
#Main Programm
rm(list=ls(all=TRUE));
start=date();

#กำหนด path ที่จะเขียนไฟล์ Excel
path_iter = "D:\\DATA\\iterative";
path_wil = "D:\\DATA\\wilkinson";
path_knn = "D:\\DATA\\knn";

#กำหนดค่าตัวแปรสำหรับการสุ่มจำนวนประชากรและเก็บค่าผลลัพธ์
mi.arr=c(0.1,0.2,0.3);
cv.arr=c(0.05,0.25,0.45);
h.arr=c(1,2,3);
n_alpha.arr=c(3,4,5);
n_beta.arr=c(3,4,5);
nb.arr=c(3,4,5);
mu=50;
p=8;

numofcase=length(mi.arr)*length(cv.arr)*length(h.arr)*length(n_alpha.arr)*length(n_beta.a
rr)*length(nb.arr);
#สร้าง Matrix เพื่อเก็บค่าผลลัพธ์
MAETABLE1=matrix(0,length(mi.arr)*length(cv.arr)*length(h.arr),length(n_alpha.arr)*length
(n_beta.arr)*length(nb.arr));
MAETABLE2=matrix(0,length(mi.arr)*length(cv.arr)*length(h.arr),length(n_alpha.arr)*length
(n_beta.arr)*length(nb.arr));
MAETABLE3=matrix(0,length(mi.arr)*length(cv.arr)*length(h.arr),length(n_alpha.arr)*length
(n_beta.arr)*length(nb.arr));

count=1;
row_index=1;

#เริ่มต้นหาค่า simulate โดยทำการวนซ้ำเพื่อแทนค่า MI, CV และ h
for(i5 in 1: length(mi.arr)){
        for(i4 in 1: length(cv.arr)){
                for(i3 in 1: length(h.arr)){
                column_index=1;
                colname='';
                #ทำการวนซ้ำเพื่อแทนค่า Alpha, Beta และ Block
                for(i22 in 1: length(n_alpha.arr)){
                        for(i21 in 1: length(n_beta.arr)){
                                for(i1 in 1: length(nb.arr)){
                                #สุ่มจำนวนลงใน block
                                #กำหนดให้ทำการSimulate1000  รอบ
                                iter=1000;
```

```
nb=nb.arr[i1];#แทนค่าจำนวน Block
n_alpha=n_alpha.arr[i22];#แทนค่า Alpha
n_beta=n_beta.arr[i21];#แทนค่า Beta
nt=n_alpha*n_beta;
h=h.arr[i3];#แทนค่า H
cv=cv.arr[i4];#แทนค่า CV
mi=mi.arr[i5];#แทนค่าMi
sigma2=(cv*mu)^2/(4*h+1);  #คำนวณหาค่า Sigma
nm=round(n_alpha*n_beta*nb*mi);#ประมาณค่าจำนวนที่สูญหาย

if(count==1){
cat("====================START
RUNNING....=======================\n");
cat("Time:", start,"\n");
}
#กำหนดค่า Array 1000 ตัวเพื่อเก็บค่าจากการ Simulate ของแต่ละวิธีในแต่ละรอบของ Case นั้นๆ
MAEi_1=array(0,iter);
MAEi_2=array(0,iter);
MAEi_3=array(0,iter);
#ทำการ Simulate วนซ้ำ 1000 รอบ
for(iter_index in 1:iter){
Alpha=rnorm(p,0,h*sqrt(sigma2));#สุ่ม  Alpha
Beta=rnorm(p,0,h*sqrt(sigma2));#สุ่ม  Beta
Block=rnorm(p,0,h*sqrt(sigma2));#สุ่ม Block
e=rnorm( p*p*p,0,sqrt(sigma2));#สุ่ม e
y=matrix(0,p*p,p);#สร้าง Matrix ขนาด 8x8 เพื่อเก็บค่าประชากร
k=1;
for(i in 1:p){
for(ii in 1:p){
for(j in 1:p){
#ตัวแบบที่ศึกษา

y[k,j]=mu+Alpha[i]+Beta[ii]+Alpha[i]*Beta[ii]+Block[j];

}
k=k+1;
}

}

Pop<-t(y)+e
Y_pop = t(Pop);
ns=n_alpha*n_beta*nb;#คำนวณหาค่าประชากรที่ต้องสุ่ม
```

```
#สุ่มจำนวนตัวอย่างจากประชากร
sam_y=sample(Y_pop,ns,replace=TRUE);ทั้งหมด
Y =matrix(sam_y,n_alpha*n_beta,nb);#เก็บจำนวนตัวอย่างที่สุ่มได้ไว้ใน Matrix
nm=round(n_alpha*n_beta*nb*mi);#คำนวณหาค่าตัวอย่างที่สูญหาย

rand = 0;
#วนลูบเพื่อตรวจสอบแถวหรือคอลัมน์ ที่มีค่าสูญหายว่าแถวหรือคอลัมน์นั้นที่เป็นค่าสูญหายทั้งหมด หากพบการสุ่มค่าสูญหายครั้งใด ได้ผลว่าแถวหรือคอลัมน์นั้นที่เป็นค่าสูญหายทั้งหมด   ต้องทำการสุ่มค่าสูญหายใหม่
while(rand == 0){
	Check = 0;
	NA_Row =array(0,nt);
	NA_Col =array(0,nb);
	#สุ่มตำแหน่งที่สูญหายตามจำนวนที่สูญหาย
	miss<- sample( seq_len(nt*nb), nm );
	YM <- Y;
	#ระบุตำแหน่งที่สูญหาย
	YM[miss] <- NA;
	X=which(is.na(YM),arr.ind=TRUE);#เก็บค่าตำแหน่งที่สูญหายไว้ใน Array
#เริ่มวนลูบตรวจสอบค่าสูญหาย
for(a in 1:nb){
	for(b in 1:nt){
		if (is.na(YM[b,a])==FALSE){
		NA_Row[b] = NA_Row[b] + YM[b,a];
		NA_Col[a] = NA_Col[a] + YM[b,a];
		}
	}
}
#วนลูบตรวจสอบคอลัมน์
for(a1 in 1:nb){
	if (NA_Col[a1]==0){
		Check = Check +1;
	}
}
#วนลูบตรวจสอบแถว
for(b1 in 1:nt){
	if (NA_Row[b1]==0){
		Check = Check +1;
	}
}
# check all missing not same row and same column
x.nrow = array(0,nm);
x.ncol = array(0,nm);
x.nrow2 = array(0,nm);
x.ncol2 = array(0,nm);
x.nrowdiff = 0;
```

```
        x.ncoldiff = 0;
        samerowcol = 0;
        for(g in 1:nm){
                x.nrow[g] = X[g];
                x.ncol[g] = X[nm+g];
        }
        for(f in 1:nm){
                x.nrow2[f] = X[nm-(f-1)];
                x.ncol2[f] = X[(nm*2)-(f-1)];
        }

        x.nrowdiff=  sum(abs(x.nrow- x.nrow2));
        x.ncoldiff=  sum(abs(x.ncol- x.ncol2));

        if(x.nrowdiff == 0 || x.ncoldiff == 0) {
                samerowcol = 1;
        }

# check random again or not
        if(Check == 0 &&samerowcol == 0) {
                rand = 1;
        }

}#จบการวนลูบตรวจสอบค่าสูญหาย
```

```
#----------Iterative Method Start------------#
#การประมาณค่าด้วยวิธีวนซ้ำ(Iterative Method)

        SUM_Row =array(0,nt);
        SUM_Col =array(0,nb);
        NO_Row =array(0,nt);
        NO_Col =array(0,nb);

        for(c1 in 1:nb){
                for(d1 in 1:nt){
                        if (is.na(YM[d1,c1])==FALSE){
                        SUM_Row[d1] = SUM_Row[d1] + YM[d1,c1];
                        NO_Row[d1] = NO_Row[d1] + 1;
                        SUM_Col[c1] = SUM_Col[c1] + YM[d1,c1];
                        NO_Col[c1] = NO_Col[c1] + 1;
                        }
                }
        }

        miss_row =array(0,nm);
        miss_col =array(0,nm);
        Est_Y = array(0,nm);

        for(k1 in 1:nm-1){
                miss_row[k1] = SUM_Row[X[k1,'row']] / NO_Row[X[k1,'row']];
                miss_col[k1] = SUM_Col[X[k1,'col']] / NO_Col[X[k1,'col']];
                Est_Y[k1] =  (miss_row[k1] + miss_col[k1]) / 2 ;
        }

        chk_final = 0;
        ctl_nm = nm;
        Div_val = (nt-1)*(nb-1);
        Y_Bar=  array(0,nm);
        Y_Chk=  array(0,nm);
        Match_Y = 0;
        Y_Chk_ = 0;
        Y_Bar_ = 0;
        cn = 0;

        #วนซ้ำเพื่อหาค่าประมาณของค่าที่สูญหายแต่ละตัว
        while(chk_final==0){
                cn = cn+1;
                Sum_G = sum(SUM_Row[]);
                Total_G = Sum_G;
                SUM_Block = SUM_Col[X[ctl_nm,'col']];
                SUM_Treatment = SUM_Row[X[ctl_nm,'row']];

        for(k in 1:nm){
        if(k != ctl_nm ) {
                if(Y_Bar[k] == 0 ){
```

```
				Total_G=  Total_G + Est_Y[k];
				if(X[ctl_nm,'col'] == X[k,'col'] ) {
					SUM_Block = SUM_Block + Est_Y[k];
				}
				if(X[ctl_nm,'row'] == X[k,'row'] ) {
					SUM_Treatment = SUM_Treatment + Est_Y[k];
				}
			}
		else{
			Total_G=  Total_G + Y_Bar[k];
			if(X[ctl_nm,'col'] == X[k,'col'] ){
					SUM_Block = SUM_Block + Y_Bar[k];
			}
			if(X[ctl_nm,'row'] == X[k,'row'] ){
				SUM_Treatment = SUM_Treatment + Y_Bar[k];
			}
		}
		}
		}

		Y_Bar[ctl_nm] = (((nb*SUM_Block) + (nt*SUM_Treatment) ) - Total_G ) / Div_val;
		Y_Bar_ = Y_Bar[ctl_nm];
		Y_Chk_ = Y_Chk[ctl_nm];
		if(round(Y_Chk_,5) == round(Y_Bar_,5)){
			Match_Y = Match_Y + 1;
			Y_Chk[ctl_nm] = Y_Bar[ctl_nm];
		}
		else
		{
			Y_Chk[ctl_nm] = Y_Bar[ctl_nm];
			Match_Y = Match_Y + 0;
		}

		if(Match_Y == nm){
			chk_final=  1;
		}
		else
		{
			if(ctl_nm-1 < 1 ){
				ctl_nm = nm ;
			}
			Else
			{
				ctl_nm = ctl_nm -1;
			}
		}
	}#จบการวนซ้ำเพื่อหาค่าประมาณของค่าที่สูญหายแต่ละตัว

		if(is.na(Y_Bar)==FALSE){
			X.real=Y[X];#แทนค่าจริง
```

```
            #หาค่าสูงสุดของค่าจริงลบด้วยค่าประมาณ
            MAEi_1[iter_index]=max(abs(X.real-Y_Bar));
            MAE1=max(MAEi_1);
        }
        else
        {   #หากค่าประมาณหาไม่ได้หรือได้ค่า N/A ให้ย้อนกลับไปทำใหม่ 1Step
            iter_index=iter_index-1;
            MAE1 <- NA;
        }
#จบการประมาณค่าด้วยวิธีวนซ้ำ(Iterative Method)
#Iterative Method End



#--------------Wilkinson Method Start----------------#
#การประมาณค่าด้วยวิธีWilkinson(Wilkinson Method)

coef_x=matrix(0,nm,nm);
c=array(0,nm);
for(i in 1:nm){
        crow=X[i,'row'];
        ccol=X[i,'col'];
        ct=nt*sum(YM[crow,],na.rm = TRUE);
        cb=nb*sum(YM[,ccol],na.rm = TRUE);
        cg=sum(YM,na.rm = TRUE);
        c[i]=cb+ct-cg;
        x_index=i;
        coef_x[i,x_index]=(nt*nb)-nt-nb+1;

        for(j in 1:nm){
            if(j!=x_index){
                if(X[j,'row']==crow){
                    coef_x[i,j]=nt;
                }
                if(X[j,'col']==ccol){
                    coef_x[i,j]=coef_x[i,j]+nb;
                }
            coef_x[i,j]=(coef_x[i,j]-1)*(-1);
            }
        }
}
        A=coef_x;
        B=as.matrix(c);
        X.estimation=try(solve(A)%*%B,TRUE);#ประมาณค่าที่สูญหาย
        if (inherits(X.estimation, "try-error")==FALSE){
            X.real=Y[X];    #แทนค่าจริง
            #หาค่าสูงสุดของค่าจริงลบด้วยค่าประมาณ
```

```
            MAEi_2[iter_index]=max(abs(X.real-X.estimation));
            MAE2=max(MAEi_2);
        }
        Else
        {
        #หากค่าประมาณหาไม่ได้หรือได้ค่า N/A ให้ย้อนกลับไปทำใหม่ 1Step
            iter_index=iter_index-1;
            MAE2 <- NA;
        }
#การประมาณค่าด้วยวิธี Wilkinson(Wilkinson Method)
#Wilkinson Method End



#------------------------------KNN Method Start----------------------#
#การประมาณค่าด้วยวิธีImputationแบบ KNN (KNN Method)

if(nm > 1){
        k=3;
        verbose=F;
        YMI = YM;
        missing.matrix = is.na(YMI)
        numMissing = sum(missing.matrix)
        x.dist = as.matrix(dist(YMI, upper=T))

        missing.rows.indices = which(
                                apply(missing.matrix, 1, function(i) {
                                                any(i)
                                        }
                                )
                        )
        x.missing = (cbind(1:nrow(YMI),YMI))[missing.rows.indices,]
            x.missing.imputed = t(apply(x.missing, 1,
                function(i) {
                        rowIndex = i[1]
                        i.original = i[-1]

                if(verbose) print(paste("Imputing row", rowIndex,sep=" "))
                        missing.cols = which(missing.matrix[rowIndex,])
                        if(length(missing.cols) == ncol(YMI))
                        warning( paste("Row",rowIndex,"is completely
missing",sep=" ") )
                imputed.values = sapply(missing.cols, function(j) {
                        neighbor.indices = which(!missing.matrix[,j])

                        knn.ranks = order(x.dist[rowIndex,neighbor.indices])

                        knn = neighbor.indices[(knn.ranks[1:k])]
```

```
                    mean(YMI[knn,j])
                })
            i.original[missing.cols] = imputed.values
            i.original
             }))

            YMI[missing.rows.indices,] = x.missing.imputed
            X.imputed = YMI[X]
    }
    Else
    {
    X.imputed<- na;
    }


    if (is.na(X.imputed)==FALSE){
            X.real=Y[X];
            MAEi_3[iter_index] = max(abs(X.real-X.imputed));
            MAE3=max(MAEi_3);
    }
    Else
    {
            iter_index=iter_index-1;
            MAE3 <- 0;
    }

}
#จบการประมาณค่าด้วยวิธีImputation แบบ KNN (Imputation KNN Method)
#Imputation KNN Method Start


#Write table
#เขียนค่าที่ได้จากวิธีวนซ้ำ (Interactive Method) ในแต่ละรอบจากการวน1000 รอบลงใน Excel
write.csv(MAEi_1, file =
paste(path_iter,"\\MAE_iter(Case=",count,",alpha=",n_alpha,",beta=",n_beta,",b=",nb,",h
=",h,",cv=",cv,",mi=",mi,",mu=",mu,",sigma2=",sigma2,").csv",sep=""));
#เก็บค่าผลลัพธ์มากที่สุดของวิธีวนซ้ำ (Interactive Method)ใน Case นั้นไว้ใน Matrix
MAETABLE1[row_index,column_index]=MAE1;

#เขียนค่าที่ได้จากวิธีWilkinson(Wilkinson Method) ในแต่ละรอบจากการวน 1000 รอบลงใน Excel
write.csv(MAEi_2, file =
paste(path_wil,"\\MAE_wil(Case=",count,",alpha=",n_alpha,",beta=",n_beta,",b=",nb,",h
=",h,",cv=",cv,",mi=",mi,",mu=",mu,",sigma2=",sigma2,").csv",sep=""));

#เก็บค่าผลลัพธ์มากที่สุดของวิธีวิธีWilkinson(Wilkinson Method) ใน Case นั้นไว้ใน Matrix
MAETABLE2[row_index,column_index]=MAE2;
```

#เขียนค่าที่ได้จากวิธีImputationKNN(Imputation KNN Method)ในแต่ละรอบจากการวน 1000 รอบลงใน Excel

```
write.csv(MAEi_3, file =
paste(path_knn,"\\MAE_knn(Case=",count,",alpha=",n_alpha,",beta=",n_beta,",b=",nb,",h
=",h,",cv=",cv,",mi=",mi,",mu=",mu,",sigma2=",sigma2,").csv",sep=""));
```

#เก็บค่าผลลัพธ์มากที่สุดของวิธีวิธีImputationKNN(Imputation KNN Method)ใน Case นั้นไว้ใน Matrix

```
MAETABLE3[row_index,column_index]=MAE3;

cat("CASE:",count,"alpha:",n_alpha,"beta:",n_beta,"b:",nb,"h",h,"cv",cv,"mi",mi,"mu",mu,"sigma2",sigma2,"\n");
cat("MAE_Iter:",MAE1,"\n");
cat("CASE:",count,"alpha:",n_alpha,"beta:",n_beta,"b:",nb,"h",h,"cv",cv,"mi",mi,"mu",mu,"sigma2",sigma2,"\n");
cat("MAE_Wil:",MAE2,"\n");
cat("CASE:",count,"alpha:",n_alpha,"beta:",n_beta,"b:",nb,"h",h,"cv",cv,"mi",mi,"mu",mu,"sigma2",sigma2,"\n");
cat("MAE_Knn:",MAE3,"\n");
if(count==numofcase){
cat("==============END RUNNING....====================\n");
        cat("Time end:",date(),"\n");
}
        count=count+1;
        column_index=column_index+1;
        }
        }
        }
        row_index=row_index+1;
}
}
}
```

#กำหนดชื่อคอลัมน์ให้กับ Matrix

```
ind=1;
for(i22 in 1: length(n_alpha.arr)){
for(i21 in 1: length(n_beta.arr)){
        for(i1 in 1: length(nb.arr)){
        if(ind==1){
        colname=paste("alpha",n_alpha.arr[i22],":beta",n_beta.arr[i21],":block",nb.arr[i1],
sep='');
        }
        else
        {
        colname=c(colname,paste("alpha",n_alpha.arr[i22],":beta",n_beta.arr[i21],":block",nb.arr[i1],sep="
"));
        }
        ind=ind+1;
}
}
}
```

#กำหนดชื่อแถวและคอลัมน์ให้กับ Matrixที่เก็บค่าผลลัพธ์วิธีวนซ้ำ (Interactive Method)

```
colnames(MAETABLE1)=colname;
rownames(MAETABLE1)=c('mi10%cv5%h1','mi10%cv5%h2','mi10%cv5%h3','mi10%cv25%h1','mi10%cv25%h2','mi10%cv25%h3','mi10%cv45%h1','mi10%cv45%h2','mi10%cv45%h3','mi20%cv5%h1','mi20%cv5%h2','mi20%cv5%h3','mi20%cv25%h1','mi20%cv25%h2','mi20%cv25%h3','mi20%cv45%h1','mi20%cv45%h2','mi20%cv45%h3','mi30%cv5%h1','mi30%cv5%h2','mi30%cv5%h3','mi30%cv25%h1','mi30%cv25%h2','mi30%cv25%h3','mi30%cv45%h1','mi30%cv45%h2','mi30%cv45%h3');
```

#เขียนค่า Matrix ที่เก็บค่าผลลัพธ์วิธีวนซ้ำ (Interactive Method) ลงในExcel ตาม Path ที่กำหนดไว้

```
write.csv(MAETABLE1,file = paste(path_iter,"\\MAETABLE_iter.csv",sep=""));
```

#กำหนดชื่อแถวและคอลัมน์ให้กับ Matrixที่เก็บค่าผลลัพธ์วิธีWilkinson(Wilkinson Method)

```
colnames(MAETABLE2)=colname;
rownames(MAETABLE2)=c('mi10%cv5%h1','mi10%cv5%h2','mi10%cv5%h3','mi10%cv25%h1','mi10%cv25%h2','mi10%cv25%h3','mi10%cv45%h1','mi10%cv45%h2','mi10%cv45%h3','mi20%cv5%h1','mi20%cv5%h2','mi20%cv5%h3','mi20%cv25%h1','mi20%cv25%h2','mi20%cv25%h3','mi20%cv45%h1','mi20%cv45%h2','mi20%cv45%h3','mi30%cv5%h1','mi30%cv5%h2','mi30%cv5%h3','mi30%cv25%h1','mi30%cv25%h2','mi30%cv25%h3','mi30%cv45%h1','mi30%cv45%h2','mi30%cv45%h3');
```

#เขียนค่า Matrix ที่เก็บค่าผลลัพธ์วิธีWilkinson(Wilkinson Method)ลงในExcel ตาม Path ที่กำหนดไว้

```
write.csv(MAETABLE2,file = paste(path_wil,"\\MAETABLE_will.csv",sep=""));
```

#กำหนดชื่อแถวและคอลัมน์ให้กับ Matrixที่เก็บค่าผลลัพธ์วิธีImputationKNN(Imputation KNN Method)

```
colnames(MAETABLE3)=colname;
rownames(MAETABLE3)=c('mi10%cv5%h1','mi10%cv5%h2','mi10%cv5%h3','mi10%cv25%h1','mi10%cv25%h2','mi10%cv25%h3','mi10%cv45%h1','mi10%cv45%h2','mi10%cv45%h3','mi20%cv5%h1','mi20%cv5%h2','mi20%cv5%h3','mi20%cv25%h1','mi20%cv25%h2','mi20%cv25%h3','mi20%cv45%h1','mi20%cv45%h2','mi20%cv45%h3','mi30%cv5%h1','mi30%cv5%h2','mi30%cv5%h3','mi30%cv25%h1','mi30%cv25%h2','mi30%cv25%h3','mi30%cv45%h1','mi30%cv45%h2','mi30%cv45%h3');
```

#เขียนค่า Matrix ที่เก็บค่าผลลัพธ์วิธีImputationKNN(Imputation KNN Method)ลงในExcel ตาม Pathที่กำหนด

```
write.csv(MAETABLE3,file = paste(path_knn,"\\MAETABLE_knn.csv",sep=""));
```

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางวิสุทธิดา ศรีดวงโชติ
วันเดือนปีเกิด	20 กุมภาพันธ์ 2525
สถานที่เกิด	อำเภอเมือง จังหวัดยโสธร
สถานที่อยู่ปัจจุบัน	999/10 ถ.เพิ่มสิน 20 แยก 6 แขวงคลองถนน เขตสายไหม กรุงเทพมหานคร 10220
ประวัติการศึกษา	
พ.ศ. 2549	วิทยาศาสตรบัณฑิต (วท.บ.) สาขาสถิติ จากมหาวิทยาลัยนเรศวร
พ.ศ. 2556	การศึกษามหาบัณฑิต (กศ.ม.) สาขาการวิจัยและสถิติทางการศึกษา จากมหาวิทยาลัยศรีนครินทรวิโรฒ