การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน

นางสาวศุภลักษณ์ กรรณิกา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2549

A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR LATIN SQUARE DESIGN

Miss Suphalak Kannika

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2006

หัวข้อวิทยานิพนธ์	การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน
โดย	นางสาวศุภลักษณ์ กรรณิกา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ
(รองศาสตราจารย์ ผกาวดี ศิริรังษี)

.... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... กรรมการ
(อาจารย์ ดร. อรุณี กำลัง)

ศุภลักษณ์ กรรณิกา : การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน (A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR LATIN SQUARE DESIGN) อ.ที่ปรึกษา : รศ. ดร.สุพล ดุรงค์วัฒนา, 144 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน 3 วิธี คือ การประมาณค่าวิธีกำลังสองน้อยสุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และการประมาณค่าวิธีมัลติเพิล อิมพิวเทชัน ซึ่งตัวแบบสำหรับแผนแบบการทดลองแบบจัตุรัสละตินที่ไม่มีการทำซ้ำเป็นดังนี้

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

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

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

ผลการวิจัยสรุปได้ว่า เมื่อเปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดจะมีค่าเพิ่มขึ้น สำหรับกรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรมีค่ามาก พบว่าการประมาณค่าสูญหายวิธีมัลติเพิล อิมพิวเทชัน จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำกว่าวิธีค่าคาดหวังสูงสุด และวิธีกำลังสองน้อยสุดในทุกสถานการณ์ที่ทำการศึกษา ดังนั้นในกรณีนี้จึงควรเลือกใช้วิธีมัลติเพิล อิมพิวเทชัน ในการประมาณค่าสูญหาย แต่สำหรับกรณีที่เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความแปรมีค่าน้อย พบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นจึงควรเลือกใช้วิธีกำลังสองน้อยสุดในการประมาณค่าสูญหายเนื่องจากสะดวกและรวดเร็วกว่า

ภาควิชา................สถิติ..................... ลายมือชื่อนิสิต.........ศุภลักษณ์ กรรณิกา.........

สาขาวิชา..............สถิติ..................... ลายมือชื่ออาจารย์ที่ปรึกษา..............................

ปีการศึกษา...........2549....................

4782401226 : MAJOR STATISTIC

KEYWORD : MISSING / LATIN SQUARE DESIGN / LEAST SQUARE / EM ALGORITHM / MULTIPLE IMPUTATION

SUPHALAK KANNIKA : A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR LATIN SQUARE DESIGN. THESIS ADVISOR : ASSOC.PROF. SUPOL DURONGWATANA, Ph.D., 144 pp.

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

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

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

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

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

Department..........Statistics..........	Student's signature...... Suphalak K. ......
Field of study........Statistics..........	Advisor's signature...... [signature] ......
Academic year......2006..............	

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ปรึกษา ตลอดจนแก้ไขข้อบกพร่องต่างๆ ด้วยดีเสมอมา จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี ผู้ทำวิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ผกาวดี ศิริรังษี ในฐานะประธานกรรมการ และอาจารย์ ดร. อรุณี กำลัง กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำชี้แนะอันเป็นประโยชน์ในการแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอน้อมรำลึกถึงพระคุณครูอาจารย์ทุกท่านที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่ช่วยส่งเสริมและสนับสนุนด้านการเรียนของผู้วิจัยและเป็นกำลังใจให้จนสำเร็จการศึกษา และขอขอบคุณ พี่สาว น้องสาว เพื่อนๆ รุ่นพี่และรุ่นน้องที่ให้กำลังใจและให้ความช่วยเหลือด้วยดีเสมอมา

สารบัญ

หน้า

สารบัญตาราง

สารบัญภาพ

รูป หน้า

รูป หน้า

รูป หน้า

รูป หน้า

รูป หน้า

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การเก็บรวมรวมข้อมูล เพื่อนำมาใช้ในการวิเคราะห์นั้น สามารถทำได้สองลักษณะคือ การเก็บรวบรวมข้อมูลจากการสังเกต และการเก็บรวบรวมข้อมูลจากการทดลอง (Experimental Study) ซึ่งในการเก็บรวบรวมข้อมูลจากการสังเกตนั้นเราจะไม่สามารถควบคุมตัวแปรรบกวน (Nuisance Variable) ได้เหมือนกับข้อมูลที่เก็บรวบรวมจากการวางแผนการทดลอง นอกจากนี้ข้อมูลที่ได้จากการสังเกตเมื่อนำมาวิเคราะห์จะได้เฉพาะความสัมพันธ์ของข้อมูลเท่านั้น แต่ข้อมูลที่ได้จากการทดลองนั้นเราสามารถที่จะวิเคราะห์เหตุและผลของข้อมูลได้ด้วย

ดังนั้นการวางแผนการทดลอง (Experimental Design) จึงเป็นที่นิยมใช้กันมากในหลายสาขา เช่น ทางด้านการแพทย์ ด้านการเกษตร ด้านการศึกษา ด้านวิศวกรรม ฯลฯ เนื่องจากการทดลอง (Experiment) เป็นกระบวนการศึกษาค้นคว้าหาข้อเท็จจริงอย่างมีระบบ เพื่อให้ได้ความรู้ความเข้าใจใหม่ๆ หรือเพื่อยืนยันความรู้เดิมที่มีอยู่แล้ว ผลของการทดลองจะทำให้ทราบข้อเท็จจริงซึ่งเป็นประโยชน์ต่อการตัดสินใจได้อย่างถูกต้อง การออกแบบการทดลองที่เหมาะสมและถูกต้องจะทำให้ได้ข้อสรุปที่มีความน่าเชื่อถือและใกล้เคียงกับความเป็นจริง แผนแบบการทดลองนั้นจึงมีหลายแผนแบบขึ้นอยู่กับ ลักษณะงานที่ต้องการทำการทดลอง และลักษณะของหน่วยทดลอง (Experimental Unit) แผนการทดลองหนึ่งที่นิยมใช้กันมาก คือ แผนการทดลองจัตุรัสละติน (Latin Square Design : LSD) แผนแบบการทดลองนี้เหมาะสมที่จะนำมาใช้เมื่อมีแหล่งความแปรปรวน 2 แหล่งที่ต้องการควบคุม ซึ่งแผนการทดลองจัตุรัสละตินสามารถควบคุมความแปรปรวนของการทดลองที่เกิดจากปัจจัยรบกวนได้ 2 ทาง (two nuisance factors) โดยการจัดกลุ่มของหน่วยทดลองออกเป็น 2 กลุ่ม ตามแหล่งของความแปรปรวน เรียกกว่า ปัจจัยแถว (row factor) และปัจจัยคอลัมน์ (column factor) แผนการทดลองจัตุรัสละตินจะสามารถควบคุมหรือขจัดความผันแปรที่ไม่ใช่อิทธิพลของวิธีทดลองออกไปจากความคลาดเคลื่อนของการทดลอง ทำให้ผู้ทดลองสามารถวัดอิทธิพลของแต่ละวิธีทดลองมีประสิทธิภาพมากขึ้น

ในการวางแผนการทดลองจัตุรัสละตินนี้ นอกจากจำนวนปัจจัยแถวจะต้องเท่ากับจำนวนปัจจัยคอลัมน์แล้วยังต้องเท่ากับจำนวนวิธีทดลองด้วย มักนิยมใช้เมื่อมีจำนวนวิธีทดลองไม่เกิน 10 ในกรณีที่มีจำนวนวิธีทดลองมากๆ จะทำให้ขนาดการทดลองใหญ่มาก และยากที่จะหาปัจจัยแถวและปัจจัยคอลัมน์ขนาดใหญ่ที่มีความสม่ำเสมอทั่วทั้งปัจจัยแถวและปัจจัยคอลัมน์ และในการวางแผนการทดลองแบบจัตุรัสละตินยังแบ่งออกเป็น 2 ลักษณะคือ การวางแผนการทดลองแบบมีซ้ำและการวางแผนการทดลองแบบไม่มีซ้ำ ซึ่งในการวางแผนการทดลองแบบมีซ้ำ เมื่อมีค่าสังเกตสูญ

หายเราสามารถตัดค่าสังเกตบางค่าทิ้งไป ส่วนกรณีไม่มีซ้ำ เมื่อมีค่าสังเกตสูญหาย จะทำให้ไม่สามารถวิเคราะห์ข้อมูลตามวิธีการของการวางแผนจัตุรัสละตินได้ตามปกติ และโดยทั่วไปแผนแบบการทดลองจะแบ่งออกเป็น 3 ลักษณะคือ ตัวแบบคงที่ (Fix Model) สำหรับระดับปัจจัยที่ศึกษานั้นมาจากระดับของประชากรที่สนใจทั้งหมด หรือระดับปัจจัยที่ถูกระบุอย่างชัดเจนโดยผู้ทำการทดลองโดยที่ระดับปัจจัยนั้นไม่มีการเปลี่ยนแปลงไม่ว่าจะทำการทดลองเมื่อใด ในการอนุมานนั้นเราสามารถสรุปผลได้เพียงระดับปัจจัยที่เรานำมาศึกษาเท่านั้น ตัวแบบที่สองคือ ตัวแบบเชิงสุ่ม (Random Model) สำหรับกรณีที่ระดับปัจจัยที่นำมาศึกษาถูกสุ่มมาจากระดับของประชากรทั้งหมด ในการอนุมานเราสามารถสรุปผลถึงระดับปัจจัยทั้งหมดที่เป็นไปได้ ส่วนตัวแบบสุดท้ายคือ ตัวแบบผสม (Mixed Model) ซึ่งจะมีทั้งตัวแบบคงที่และตัวแบบเชิงสุ่มอยู่ในแผนแบบการทดลองเดียวกัน ดังนั้นในการวิจัยครั้งนี้จึงได้ศึกษาการประมาณค่าสังเกตสูญหายในการวางแผนการทดลองจัตุรัสละตินกรณีตัวแบบเชิงสุ่มที่ไม่มีการทำซ้ำ ซึ่งวิธีวิเคราะห์ข้อมูลโดยการประมาณค่าสังเกตสูญหายสามารถทำได้หลายวิธี แต่ในที่นี่จะศึกษาเปรียบเทียบวิธีการประมาณเพียง 3 วิธี คือ

1. วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (Least square method) เป็นการประมาณค่าที่ทำให้ผลรวมของค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด
2. วิธีประมาณค่าสูญหายโดยวิธี Expectation Maximization algorithm (EM algorithm) ในการประมาณค่าพารามิเตอร์ และหาค่าตัวสถิติเพียงพอสำหรับวิธีนี้จะอาศัยฟังก์ชันภาวะน่าจะเป็นสูงสุดเข้ามาช่วย
3. วิธีประมาณค่าสูญหายโดยวิธี Multiple Imputation (MI method) วิธีนี้เป็นการสร้างชุดข้อมูลขึ้นมาใหม่ โดยชุดข้อมูลที่สูญหายจะถูกแทนที่โดยชุดของค่าที่เป็นไปได้มากกว่า 1 ($m > 1$) เพื่อที่จะสร้างชุดข้อมูลที่สมบูรณ์ m ชุด แนวคิดพื้นฐานของ MI คือการหา Predictive distribution สำหรับค่าสังเกตที่สูญหายจากค่าสังเกตที่มีอยู่

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายโดยใช้วิธีการประมาณค่า 3 วิธีคือ

1. วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (Least square method)
2. วิธีประมาณค่าสูญหายโดยวิธี Expectation Maximization algorithm (EM algorithm)
3. วิธีประมาณค่าสูญหายโดยวิธี Multiple Imputation (MI method)

1.3 สมมติฐานของการวิจัย

การศึกษาวิจัยครั้งนี้ การประมาณค่าสูญหายในการวางแผนการทดลองด้วย วิธี Multiple Imputation Method จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) มีค่าต่ำสุดในทุกวิธีที่ทำการศึกษา

1.4 ข้อตกลงเบื้องต้น

1. ในการศึกษาครั้งนี้ศึกษาภายใต้การวางแผนการทดลองแบบจัตุรัสละติน สมมติว่าปัจจัยที่ศึกษา 3 ด้านคือ วิธีทดลอง มี p ระดับ ปัจจัยแถว (Row factor) มี p ระดับ และ ปัจจัยคอลัมน์ (Column factor) มี p ระดับ วิธีทดลองแต่ละวิธีจะปรากฏขึ้นเพียงครั้งเดียวในแต่ละแถวและคอลัมน์ ดังนั้น จะมีจำนวนหน่วยทดลองในการทดลองหนึ่งๆ เท่ากับ p x p หน่วย มีตัวแบบเป็นดังนี้

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

เมื่อ Y_{ijk} คือ ค่าสังเกตของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

μ คือ ค่าเฉลี่ยรวมของประชากร

τ_i คือ อิทธิพลของวิธีทดลอง ที่ i

β_j คือ อิทธิพลของปัจจัยแถว ที่ j

α_k คือ อิทธิพลของปัจจัยคอลัมน์ ที่ k

ε_{ijk} คือ ความคลาดเคลื่อนของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

และ p คือ จำนวนวิธีทดลอง, จำนวนปัจจัยแถว และจำนวนปัจจัยคอลัมน์

2. สมมติให้ μ เป็นพารามิเตอร์ที่ไม่ทราบค่า และ τ_i เป็นอิทธิพลของวิธีทดลอง ที่ i, β_j เป็นอิทธิพลของปัจจัยแถว ที่ j, α_k เป็นอิทธิพลของปัจจัยคอลัมน์ ที่ k โดยที่ $\tau_i, \beta_j, \alpha_k, \varepsilon_{ijk}$ เป็นตัวแปนสุ่มที่เป็นอิสระกันและมีการแจกแจงแบบปกติที่มี

$E(\tau_i) = E(\beta_j) = E(\alpha_k) = E(\varepsilon_{ijk}) = 0$ และ $Var(\tau_i) = \sigma_\tau^2$, $Var(\beta_j) = \sigma_\beta^2$, $Var(\alpha_k) = \sigma_\alpha^2$, $Var(\varepsilon_{ijk}) = \sigma_\varepsilon^2$

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

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

1.5 ขอบเขตของการวิจัย

1 ศึกษาเปรียบเทียบวิธีประมาณค่าสังเกตสูญหาย โดยวิธีกำลังสองน้อยสุด วิธี EM algorithm และวิธี Multiple Imputation Method

2 กำหนดตัวแบบเป็นตัวแบบเชิงสุ่มและไม่มีการทำซ้ำ

3 กำหนดจำนวนของปัจจัยในแผนการทดลองจัตุรัสละติน ดังนี้

- จำนวนวิธีทดลอง (i) เท่ากับ 3, 4, 5, 6 และ 7
- จำนวนปัจจัยแถว (j) เท่ากับ 3, 4, 5, 6 และ 7
- จำนวนปัจจัยคอลัมน์ (k) เท่ากับ 3, 4, 5, 6 และ 7

ดังนั้นจะมีจำนวนหน่วยตัวอย่าง 9, 16, 25, 36 และ 49 หน่วย ตามลำดับ

4 กำหนดความคลาดเคลื่อนเป็นตัวแปรสุ่มที่มีการแจกแจงเป็นแบบปกติและเป็นอิสระซึ่งกันและกัน มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ σ^2

5 กำหนดให้จำนวนข้อมูลสูญหาย 10%, 20% และ 30% อย่างสุ่ม

6 กำหนดให้ข้อมูลมีค่าสัมประสิทธิ์ความผันแปร (Coefficient of variation : C.V. (%))* ในระดับต่างๆ กัน คือ 5%, 25% และ 45% และกำหนดให้ค่าเฉลี่ยของประชากรเท่ากันทุกกลุ่ม (μ) เท่ากับ 50

จาก

$$c.v.(y_{ijk}) = \frac{S.D.(y_{ijk})}{\mu} = \frac{\sqrt{\sigma_\tau^2 + \sigma_\beta^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}}{\mu}$$

กำหนดให้

$\sigma_\tau^2 = \sigma_\beta^2 = \sigma_\alpha^2 = h\sigma_\varepsilon^2$ โดยที่ h** เป็นค่าจำนวนเต็มคงที่เท่ากับ 1, 2 และ 3

นั่นคือ

$$c.v.(y_{ijk}) = \frac{\sqrt{h\sigma_\varepsilon^2 + h\sigma_\varepsilon^2 + h\sigma_\varepsilon^2 + \sigma_\varepsilon^2}}{\mu} = \frac{\sigma_\varepsilon\sqrt{3h+1}}{\mu}$$

ดังนั้นจะได้ว่า

$$\sigma_\varepsilon^2 = \frac{(c.v.(y_{ijk})\mu)^2}{3h+1}$$

โดยมีความแปรปรวน $(\sigma_{y_{ijk}}^2)$ เท่ากับ 6.25, 156.25 และ 506.25 ตามลำดับ

---

* เนื่องจากในงานวิจัยครั้งนี้ได้ศึกษาในแผนแบบการทดลองจัตุรัสละติน ข้อมูลที่นำมาศึกษาในแผนแบบการทดลองนี้จึงเป็นข้อมูลที่ได้มาจากการทดลองซึ่งมีความผันแปรไม่มากนัก เนื่องจากเราสามารถควบคุมปัจจัยรบกวนต่างๆ ได้ ดังนั้นผู้วิจัยจึงได้กำหนดค่าสัมประสิทธิ์ความผันแปรออกเป็น 3 ระดับคือ ระดับต่ำ ระดับกลาง และระดับสูง มีค่าเท่ากับ 5%, 25% และ 45% ตามลำดับ

** เนื่องจากเพื่อสะดวกต่อการวิจัยและทำให้การจำลองแบบข้อมูลที่นำมาใช้ในการศึกษามีหลักเกณฑ์มากขึ้น จึงกำหนดให้ $\sigma_\tau^2 = \sigma_\beta^2 = \sigma_\alpha^2 = h\sigma_\varepsilon^2$

7 การจำลองในแต่ละสถานการณ์ของการทดลองกระทำซ้ำ 500 รอบ

8 ในการวิจัยครั้งนี้สร้างแบบการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Simulation) เขียนด้วยโปรแกรม S-PLUS 2000

1.6 เกณฑ์ในกาตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการประมาณค่าสังเกตสูญหายวิธีใดจะให้ค่าประมาณที่ดี จะพิจารณาจากเกณฑ์ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) ซึ่งมีสูตรดังนี้

$$MAE_t = Max\left[Max\left|y_i - \hat{y}_i\right|\right]_j$$

เมื่อ

y_i แทน ค่าจริงค่าที่ i ที่ได้จากการจำลอง

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

j แทน รอบที่ j จากการทดลอง (ในที่นี้กำหนดให้ j = 500 รอบ)

MAE_t แทน ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดจากวิธีที่ t

จะพิจารณาว่าวิธีการประมาณค่าสูญหายวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) มีค่าต่ำสุด จะเป็นวิธีการประมาณค่าสังเกตสูญหายที่ดีที่สุด

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ผลจากการวิจัย จะเป็นแนวทางในการเลือกวิธีประมาณค่าสูญหายในการวางแผนการทดลองแบบจัตุรัสละติน เพื่อให้ได้ผลสรุปจากการวางแผนการทดลองมีความน่าเชื่อถือมากที่สุด

1.8 วิธีดำเนินการวิจัย

1 จำลองข้อมูลโดยใช้วิธีมอนติคาร์โล (Simulation by Monte Carlo Method) โดยให้รูปแบบประชากรมีการแจกแจงแบบปกติ

2 สุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง

3 ประมาณค่าสูญหายโดยใช้วิธีการประมาณค่าสูญหายทั้ง 3 วิธี ดังนี้

3.1 วิธีกำลังสองน้อยสุด (Least square method)

3.2 วิธี Expectation Maximization algorithm (EM algorithm)

3.3 วิธี Multiple Imputation (MI method)

4 คำนวณค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE))

5 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) ที่ได้จากวิธีการประมาณทั้ง 3 วิธี

บทที่ 2

แนวคิดและทฤษฎี

2.1 ตัวแบบที่ทำการศึกษา

ในงานวิจัยครั้งนี้ได้ทำการศึกษาวิธีการประมาณค่าสังเกตสูญหายทั้งหมด 3 วิธี เพื่อพิจารณาเปรียบเทียบหาวิธีการประมาณค่าสูญหายที่เหมาะสมที่สุดสำหรับแผนการทดลองจัตุรัสละติน วิธีที่เลือกมาทำการศึกษาเปรียบเทียบในงานวิจัยครั้งนี้ได้แก่ วิธีกำลังสองน้อยสุด วิธีที่สองคือวิธี EM algorithm และวิธี Multiple Imputation Method ในการวางแผนการทดลองแบบจัตุรัสละตินนี้แบ่งออกเป็น 2 ลักษณะคือ การวางแผนการทดลองแบบมีซ้ำ และการวางแผนการทดลองแบบไม่มีซ้ำ ซึ่งในการวางแผนการทดลองแบบมีซ้ำ เมื่อมีค่าสังเกตสูญหายเราสามารถตัดค่าสังเกตบางค่าทิ้งไป ส่วนกรณีไม่มีซ้ำ เมื่อมีค่าสังเกตสูญหาย จะทำให้ไม่สามารถวิเคราะห์ข้อมูลตามวิธีการของการวางแผนจัตุรัสละตินได้ตามปกติ ดังนั้นในการวิจัยครั้งนี้จึงได้ศึกษาการประมาณค่าสังเกตสูญหายในการวางแผนการทดลองจัตุรัสละตินกรณีที่ไม่มีการทำซ้ำ ซึ่งลักษณะของข้อมูลได้แสดงไว้ในตารางที่ 2.1 และมีตัวแบบที่ทำการศึกษาดังนี้

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

เมื่อ

Y_{ijk} คือ ค่าสังเกตของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

μ คือ ค่าเฉลี่ยรวมของประชากร

τ_i คือ อิทธิพลของวิธีทดลอง ที่ i

β_j คือ อิทธิพลของปัจจัยแถว ที่ j

α_k คือ อิทธิพลของปัจจัยคอลัมน์ ที่ k

ε_{ijk} คือ ความคลาดเคลื่อนของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

และ p คือ จำนวนวิธีทดลอง, จำนวนปัจจัยแถว และจำนวนปัจจัยคอลัมน์

ตารางที่ 2.1 ลักษณะของข้อมูลจากแผนการทดลองจัตุรัสละติน

Row Factor j = 1, 2, …, p	Column Factor k = 1, 2, …, p						รวม $y_{.j.}$	ค่าเฉลี่ย $\bar{y}_{.j.}$
	1	2	.	.	.	p		
1	y_{111}	y_{p12}	.	.	.	y_{21p}	$y_{.1.}$	$\bar{y}_{.1.}$
2	y_{221}	y_{222}	.	.	.	y_{32p}	$y_{.2.}$	$\bar{y}_{.2.}$
3	y_{331}	y_{232}	.	.	.	y_{43p}	$y_{.3.}$	$\bar{y}_{.3.}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
p	y_{pp1}	$y_{(p-1)p2}$	.	.	.	y_{1pp}	$y_{.p.}$	$\bar{y}_{.p.}$
รวม $y_{..k}$	$y_{..1}$	$y_{..2}$	.	.	.	$y_{..p}$	$y_{...}$	
ค่าเฉลี่ย $\bar{y}_{..k}$	$\bar{y}_{..1}$	$\bar{y}_{..2}$	.	.	.	$\bar{y}_{..p}$		$\bar{y}_{...}$

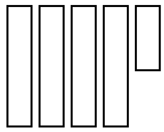
2.2 ลักษณะของข้อมูลสูญหาย

สิ่งสำคัญ 3 ประการที่ควรตรวจสอบข้อมูลก่อนการนำไปวิเคราะห์ ได้แก่

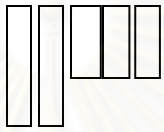
1) ข้อมูลสูญหาย (Missing Data) หมายถึง ข้อมูลที่มีบางหน่วยหรือบางเซลล์ของเมตริกซ์หายไป
2) ข้อมูลไม่คงเส้นคงวา (Inconsistent) หรือมีความคลาดเคลื่อน (Nonsampling Error) หมายถึง ข้อมูลที่ค่าหรือความหมายของตัวแปรหนึ่งหรือหลายตัวแปรขัดแย้งกับค่าหรือความหมายของตัวแปรอื่น
3) ข้อมูลมีสิ่งรบกวน (Noisy) หรือมีข้อมูลผิดปกติ (Outliers) ซึ่งข้อมูลที่มีสิ่งรบกวน หมายถึง ความคลาดเคลื่อนเชิงสุ่ม หรือ ความแปรปรวนในการวัดค่าของตัวแปร ส่วนข้อมูลผิดปกติหรือมีสิ่งไม่พึงประสงค์ หมายถึง ข้อมูลที่มีลักษณะใดลักษณะหนึ่ง หรือหลายลักษณะต่อไปนี้
    - มีค่าสูงงหรือต่ำมากๆ (Extreme Value)
    - มีค่าแตกต่างจากข้อมูลส่วนใหญ่
    - มีค่าอยู่นอกช่วงที่กำหนด

รูปแบบของข้อมูลสูญหายมี 5 แบบ ดังนี้

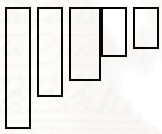
1. Univariate Nonresponse คือ ชุดข้อมูลที่มีข้อมูลสูญหายใน 1 ตัวแปรเท่านั้น เช่น ในการทดลอง ผลผลิตบางแปลงหายไป



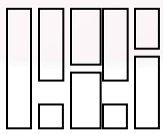
2. Multivariate Two Patterns ส่วนใหญ่มักจะเกิดขึ้นในการสำรวจความคิดเห็น ซึ่งมีคำถามบางข้อที่อาจมีบางคนไม่ประสงค์จะตอบ ไม่ทราบ หรือ ไม่ตอบด้วยเหตุผลบางประการ



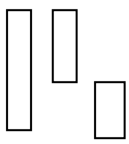
3. Monotone มักจะพบในการศึกษาประเภท Longitudinal Study ซึ่งหน่วยตัวอย่างหรือหน่วยทดลองบางหน่วยอาจตายไปก่อนที่การศึกษาจะจบลง



4. General เป็นลักษณะของข้อมูลสูญหายที่เกิดขึ้นโดยทั่วไปในการสำรวจ โดยเฉพาะอย่างยิ่งการสำรวจความคิดเห็น



5. File Matching เกิดขึ้นในกรณีที่ข้อมูลมีจำนวนมาก ตัวแปรบางตัวไม่สามารถสังเกตได้พร้อมๆ กัน หรือประมาณค่าพารามิเตอร์เกี่ยวกับความสัมพันธ์ ไม่ได้ ถ้าทำ อาจเป็นผลให้แปลความหมายผิดๆ



ลักษณะการเกิดข้อมูลสูญหาย มี 3 ประเภทดังนี้

1. Missing Completely at Random (MCAR) คือความน่าจะเป็นที่หน่วยตัวอย่างของตัวแปรจะเกิดการสูญหายมีค่าเท่าๆ กันทุกหน่วยตัวอย่าง เช่น สมมติให้ตัวแปรที่ศึกษามี 2 ตัว คือ อายุ และรายได้ ในกรณีนี้ ความน่าจะเป็นที่รายได้จะสูญหาย มีค่าเท่าๆ กันทุกหน่วยตัวอย่างไม่ว่าอายุและรายได้จะเป็นอะไร
2. Missing at Random (MAR) คือความน่าจะเป็นที่หน่วยตัวอย่างของตัวแปรหนึ่งจะเกิดการสูญหายขึ้นอยู่กับตัวแปรอีกตัวแปรหนึ่ง แต่ไม่ขึ้นอยู่กับตัวแปรที่มีค่าสูญหายนั้นๆ เช่น ความน่าจะเป็นที่รายได้จะสูญหายแปรผันตามอายุ แต่ไม่แปรผันตามรายได้ของคนที่มีอายุเท่าๆ กัน
3. Not Missing at Random (NMAR) คือความน่าจะเป็นที่หน่วยตัวอย่างของตัวแปรหนึ่งจะเกิดการสูญหายขึ้นอยู่กับตัวแปรนั้นๆ ที่มีลักษณะของหน่วยตัวอย่างเหมือนกัน เช่น ความน่าจะเป็นที่รายได้จะสูญหายแปรผันตามรายได้ของคนที่มีอายุเท่าๆ กัน

วิธีการประมาณค่าสังเกตที่สูญหายนั้น ปัจจุบันมีหลายวิธี สำหรับงานวิจัยครั้งนี้ได้ทำการศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายเพียง 3 วิธี วิธีแรกคือ วิธีกำลังสองน้อยสุด ซึ่งเป็นการประมาณค่าที่ทำให้ผลรวมของค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด วิธีที่สองคือ วิธี EM algorithm เป็นการประมาณค่าพารามิเตอร์ และหาค่าตัวสถิติเพียงพอโดยอาศัยฟังก์ชันภาวะน่าจะเป็นสูงสุดเข้ามาช่วย และวิธีสุดท้ายคือ วิธี Multiple Imputation Method วิธีนี้เป็นการสร้างชุดข้อมูลขึ้นมาใหม่ โดยชุดข้อมูลที่สูญหายจะถูกแทนที่โดยชุดของค่าที่เป็นไปได้มากกว่า 1 ($m > 1$) เพื่อที่จะสร้างชุดข้อมูลที่สมบูรณ์ m ชุด แนวคิดพื้นฐานของ MI คือการหา Predictive distribution สำหรับค่าสังเกตที่สูญหายจากค่าสังเกตที่มีอยู่ โดยอาศัยการประมาณภาวะน่าจะเป็นสูงสุดเข้ามาช่วย ซึ่งรายละเอียดทั้ง 3 วิธีดังกล่าวมีดังต่อไปนี้

2.3 การประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (Least Square Method)

การวิเคราะห์ข้อมูลโดยการประมาณค่าสังเกตสูญหาย ทำได้โดยประมาณค่าสังเกตสูญหายขึ้นมาแล้ววิเคราะห์ผลด้วยเทคนิคตามแผนแบบการทดลองที่วางไว้ตามปกติ

การประมาณค่าสังเกตสูญหายเมื่อมีค่าสังเกตสูญหายเพียง 1 ค่า ได้ริเริ่มนำมาใช้โดย อลันและวิสชาร์ท (Allan and Wishart ; 1930) ต่อมาเยทส์ (Yates ; 1937) ได้แสดงให้เห็นว่าสูตรดังกล่าวได้มาจากวิธีกำลังสองน้อยสุด คือค่าที่ประมาณขึ้นเป็นค่าที่ทำให้ผลรวมของความคลาดเคลื่อนกำลังสอง (SSE) มีค่าน้อยที่สุด และวิธีนี้ยังสามารถนำไปใช้หาค่าประมาณของค่าสังเกตสูญหายหลายค่าโดยวิธีวนซ้ำได้อีกด้วย

การประมาณค่าสังเกตสูญหาย 1 ค่า สำหรับจัตุรัสละตินขนาด p x p

ให้ M_{ijk} แทน ค่าสังเกตสูญหายของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

T_i แทน ผลรวมของวิธีทดลองที่มีค่าสังเกตสูญหาย

R_j แทน ผลรวมของปัจจัยแถวที่มีค่าสังเกตสูญหาย

C_k แทน ผลรวมของปัจจัยคอลัมน์ที่มีค่าสังเกตสูญหาย

T แทน ผลรวมทั้งหมดของค่าสังเกตเมื่อมีค่าสังเกตสูญหาย

เนื่องจากผลรวมของค่าความคลาดเคลื่อนกำลังสอง

$$SSE = \sum_{j=1}^{p}\sum_{k=1}^{p}\left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}\right)^2$$

เมื่อ C เป็นค่าคงที่ใดๆ สามารถเขียนผลรวมของค่าความคลาดเคลื่อนกำลังสอง อยู่ในรูปฟังก์ชันของ M_{ijk} ได้ดังนี้

$$SSE = M_{ijk}^2 - \frac{1}{p}\left(T_i + M_{ijk}\right)^2 - \frac{1}{p}\left(R_j + M_{ijk}\right)^2 - \frac{1}{p}\left(C_k + M_{ijk}\right)^2 + \frac{2}{p^2}\left(T + M_{ijk}\right)^2 + T$$

ค่าของ M_{ijk} ที่ทำผลรวมของค่าความคลาดเคลื่อนกำลังสองน้อยที่สุด หาได้โดยให้ $\frac{\partial SSE}{\partial M_{ijk}} = 0$

จะได้

$$2M_{ijk} - \frac{2}{p}\left(T_i + M_{ijk}\right) - \frac{2}{p}\left(R_j + M_{ijk}\right) - \frac{2}{p}\left(C_k + M_{ijk}\right) + \frac{4}{p^2}\left(T + M_{ijk}\right) = 0$$

$$M_{ijk} - \frac{1}{p}\left(T_i + M_{ijk}\right) - \frac{1}{p}\left(R_j + M_{ijk}\right) - \frac{1}{p}\left(C_k + M_{ijk}\right) + \frac{2}{p^2}\left(T + M_{ijk}\right) = 0$$

$$M_{ijk} - \frac{1}{p}\left(T_i + R_j + C_k\right) - \frac{3}{p}\left(M_{ijk}\right) + \frac{2T}{p^2} + \frac{2}{p^2}\left(M_{ijk}\right) = 0$$

$$p^2 M_{ijk} - p\left(T_i + R_j + C_k\right) - 3p\left(M_{ijk}\right) + 2T + 2M_{ijk} = 0$$

$$p^2 M_{ijk} - 3p\left(M_{ijk}\right) + 2M_{ijk} = p\left(T_i + R_j + C_k\right) - 2T$$

$$M_{ijk}\left(p^2 - 3p + 2\right) = p\left(T_i + R_j + C_k\right) - 2T$$

$$M_{ijk} = \frac{p(T_i + R_j + C_k) - 2T}{(p^2 - 3p + 2)}$$

ดังนั้นจะได้ $$M_{ijk} = \frac{p(T_i + R_j + C_k) - 2T}{(p-2)(p-1)}$$

ในกรณีที่มีค่าสังเกตสูญหายมากกว่า 1 ค่า จะประมาณค่าสูญหายได้โดยวิธีเดียวกัน คือ เขียนผลรวมของค่าความคลาดเคลื่อนกำลังสองในรูปฟังก์ชันของค่าสังเกตที่สูญหายแล้วหาอนุพันธ์เทียบกับค่าสังเกตสูญหายที่ต้องการ แล้วให้เท่ากับศูนย์ จะได้จำนวนสมการเท่ากับจำนวนค่าสังเกตที่สูญหาย เมื่อแก้สมการจะได้สูตรการประมาณค่าสังเกตสูญหายนั้นๆ ดังนี้

เช่นให้ M_{ijk} และ $M_{i^*j^*k^*}$ แทน ค่าสังเกตสูญหาย 2 ค่า

โดย M_{ijk} แทน ค่าสังเกตสูญหายของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

$M_{i^*j^*k^*}$ แทน ค่าสังเกตสูญหายของวิธีทดลองที่ i* ปัจจัยแถวที่ j* และปัจจัยคอลัมน์ที่ k*

T_i และ T_{i^*} แทน ผลรวมของวิธีทดลองที่ i และ i* ที่มีค่าสังเกตสูญหาย

R_j และ R_{j^*} แทน ผลรวมของปัจจัยแถวที่ j และ j* ที่มีค่าสังเกตสูญหาย

C_k และ C_{k^*} แทน ผลรวมของปัจจัยคอลัมน์ที่ k และ k* ที่มีค่าสังเกตสูญหาย

T แทน ผลรวมทั้งหมดของค่าสังเกตเมื่อมีค่าสังเกตสูญหาย

เมื่อ C เป็นค่าคงที่ใดๆ สามารถเขียนผลรวมของค่าความคลาดเคลื่อนกำลังสอง อยู่ในรูปฟังก์ชันของ M_{ijk} และ $M_{i^*j^*k^*}$ ได้ดังนี้

$$SSE = M_{ijk}^2 + M_{i^*j^*k^*}^2 - \frac{1}{p}(T_i + M_{ijk})^2 - \frac{1}{p}(T_{i^*} + M_{i^*j^*k^*})^2 - \frac{1}{p}(R_j + M_{ijk})^2 - \frac{1}{p}(R_{j^*} + M_{i^*j^*k^*})^2$$
$$- \frac{1}{p}(C_k + M_{ijk})^2 - \frac{1}{p}(C_{k^*} + M_{i^*j^*k^*})^2 + \frac{2}{p^2}(T + M_{ijk} + M_{i^*j^*k^*})^2 + T$$

หาอนุพันธ์เทียบกับ M_{ijk} และ $M_{i^*j^*k^*}$ แล้วให้เท่าศูนย์

$$M_{ijk} - \frac{1}{p}(T_i + M_{ijk}) - \frac{1}{p}(R_j + M_{ijk}) - \frac{1}{p}(C_k + M_{ijk}) + \frac{2}{p^2}(T + M_{ijk} + M_{i^*j^*k^*}) = 0$$

……….(1)

$$M_{i^*j^*k^*} - \frac{1}{p}\left(T_{i^*} + M_{i^*j^*k^*}\right) - \frac{1}{p}\left(R_{j^*} + M_{i^*j^*k^*}\right) - \frac{1}{p}\left(C_{k^*} + M_{i^*j^*k^*}\right) + \frac{2}{p^2}\left(T + M_{ijk} + M_{i^*j^*k^*}\right) = 0$$

……….(2)

ซึ่งถ้ามีค่าสังเกตสูญหายเพียง 1 ค่า สมมติว่าเป็น M_{ijk} ดังนั้นค่าประมาณของ M_{ijk} หาได้โดยการแทนค่า $M_{ijk} = 0$ ในสมการที่ (1) แต่ถ้ามีค่าสังเกตสูญหาย 2 ค่า การประมาณค่าสังเกตสูญหายดังกล่าวจะทำได้โดยการแก้สมการที่ (1) และ (2) ซึ่งวิธีการประมาณด้วยวิธีดังกล่าวนี้ค่อนข้างจะยุ่งยาก ดังนั้นจึงมักหาค่าประมาณโดยวิธีการวนซ้ำตามแบบของเยทส์ (Yates) ตามขั้นตอนดังนี้

1) ประมาณค่าสังเกตสูญหายเบื้องต้นของค่าที่ 1 ถึง n-1 โดยใช้สูตร

$$M_{ijk} = \frac{\bar{T}_i + \bar{R}_j + \bar{C}_k}{3}$$

เมื่อ M_{ijk} แทน ค่าสังเกตสูญหายของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

$\bar{T}_i$ แทน ค่าเฉลี่ยของวิธีทดลองที่มีค่าสังเกตสูญหาย

$\bar{R}_j$ แทน ค่าเฉลี่ยของปัจจัยแถวที่มีค่าสังเกตสูญหาย

$\bar{C}_k$ แทน ค่าเฉลี่ยของปัจจัยคอลัมน์ที่มีค่าสังเกตสูญหาย

2) ประมาณค่าสังเกตสูญหายค่าที่ n ด้วยสูตร

$$M_{ijk} = \frac{p\left(T_i + R_j + C_k\right) - 2T}{(p-2)(p-1)}$$

เมื่อ M_{ijk} แทน ค่าสังเกตสูญหายของวิธีทดลองที่ i ปัจจัยแถวที่ j และปัจจัยคอลัมน์ที่ k

T_i แทน ผลรวมของวิธีทดลองที่มีค่าสังเกตสูญหาย

R_j แทน ผลรวมของปัจจัยแถวที่มีค่าสังเกตสูญหาย

C_k แทน ผลรวมของปัจจัยคอลัมน์ที่มีค่าสังเกตสูญหาย

T แทน ผลรวมทั้งหมดของค่าสังเกตเมื่อมีค่าสังเกตสูญหาย

3) ประมาณค่าสังเกตสูญหายค่าที่ 1 ถึง n ใหม่ โดยประมาณทีละค่าเริ่มจากค่าที่ 1 ด้วยสูตรในข้อที่ 2) ทำการวนซ้ำไปมาจนค่าที่ประมาณได้ทุกค่าลู่เข้าสู่ค่าคงที่

2.4 การประมาณค่าสูญหายโดยวิธี Expectation Maximization algorithm (EM algorithm)

การประมาณค่าสูญหายโดยวิธี EM algorithm ได้ถูกเสนอเป็นครั้งแรกโดยเด็มเสตอร์ ลายด์และรูบิน (Dempster Laird and Rubin ; 1977) วิธีนี้จะหาค่าประมาณภาวะน่าจะเป็นสูงสุดของพารามิเตอร์โดยการวนซ้ำ ซึ่งขั้นตอนของ EM algorithm แบ่งออกเป็น 2 ขั้นตอน ขั้นแรกคือ Expectation step (E-step) เป็นขั้นตอนการหาค่าคาดหวัง และขั้นที่สองคือ Maximization step (M-step) เป็นขั้นตอนการหาค่ามากที่สุด

สมมติให้ Y มีข้อมูลทั้งหมด p^2 ค่า มีค่าสังเกตที่เก็บมาได้ m ค่า และมีค่าสังเกตที่สูญหาย $p^2 - m$ ค่า ดังนั้น Y จึงประกอบด้วย 2 ส่วน คือ $Y = (Y_{mis}, Y_{obs})$

ในการประมาณค่าสูญหายโดยวิธี EM algorithm นั้นก่อนอื่นจะต้องพิจารณาการแจกแจงของข้อมูลของค่าสังเกตที่เหลืออยู่ว่ามีการแจกแจงแบบปกติหรือไม่ ถ้าหากมีการแจกแจงแบบปกติจะเริ่มประมาณค่าสูญหายโดยวิธี EM algorithm ตามขั้นตอนดังต่อไปนี้

1. ขั้นตอนการหาค่าคาดหวัง E-step

$$E\left(\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk} \middle| \theta^{(t)}, Y_{obs}\right) = \sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk} + (p^2 - m)\mu^{(t)} \qquad \ldots\ldots(1)$$

$$E\left(\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 \middle| \theta^{(t)}, Y_{obs}\right) = \sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}^2 + (p^2 - m)[(\mu^{(t)})^2 + (\sigma^{(t)})^2] \qquad \ldots\ldots(2)$$

เมื่อ $\mu^{(t)}$ และ $(\sigma^{(t)})^2$ เริ่มต้นคือ

$$\mu^{(t)} = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}}{p^2}$$

และ

$$(\sigma^{(t)})^2 = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}^2}{p^2} - (\mu^{(t)})^2$$

2. ขั้นตอนการหาค่ามากที่สุด M-step

$$\mu^{(t+1)} = \frac{E\left(\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk} \middle| \theta^{(t)}, Y_{obs}\right)}{p^2} \qquad \text{……(3)}$$

$$(\sigma^{(t+1)})^2 = \frac{E\left(\sum_{j=1}^{p}\sum_{k=1}^{p} y_{ijk}^2 \middle| \theta^{(t)}, Y_{obs}\right)}{p^2} - (\mu^{(t+1)})^2 \qquad \text{……(4)}$$

กำหนดให้ $\mu^{(t)} = \mu^{(t+1)} = \hat{\mu}$

และ $\sigma^{(t)} = \sigma^{(t+1)} = \hat{\sigma}$

ทำการวนซ้ำสมการ (1) ถึง (4) จนลู่เข้าสู่ค่าคงที่

$$\hat{\mu} = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}}{m}$$

และ

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}^2}{m} - \hat{\mu}^2$$

2.5 การประมาณค่าสูญหายโดยวิธี Multiple Imputation Method (MI)

วิธี Multiple Imputation เป็นวิธีหนึ่งของวิธี Imputation ซึ่งวิธีนี้จะประมาณค่าสูญหายคล้ายกับวิธี Single Imputation แต่ได้ดำเนินการมากกว่า 1 ครั้ง ในการประมาณค่าสูญหายด้วยวิธี Imputation นี้มีอยู่ด้วยกันหลายวิธี แต่วิธีอื่นๆ มักจะมีข้อด้อยกว่า เช่น วิธี Hot Deck Imputation มีข้อด้อยคือ ได้ตัวประมาณที่เอนเอียนเอียง (Bias) และการหาความเที่ยงตรงแม่นยำมีข้อจำกัดด้านทฤษฎี และอีกวิธีหนึ่งคือ วิธี Single Imputation ซึ่งวิธีนี้มีข้อด้อยกว่าคือ ข้อมูลที่ถูกประมาณค่าขึ้นมาใหม่ ไม่ได้สะท้อนให้เห็นการกระจายสุ่ม (Sampling Variability) ของค่าจริงของ Y เมื่อตัวแบบเป็นอย่างที่เลือก และไม่สะท้อนให้เห็นความแตกต่างหรือความไวของตัวแบบ ส่วนข้อดีของวิธี Multiple Imputation นั้นมีดังนี้

1. เพิ่มประสิทธิภาพของตัวประมาณ
2. อนุมาณเกี่ยวกับการกระจายสุ่มของค่าสูญหายได้ เพราะได้ประมาณค่าสูญหายมากกว่า 1 ค่า
3. ศึกษาเกี่ยวกับความไวของตัวแบบที่ใช้ เมื่อใช้ตัวแบบต่างกันมากกว่า 1 ตัวแบบ

วิธี MI ได้ถูกเสนอเป็นครั้งแรกโดยรูบิน (Rubin, 1978) เพื่อนำมาใช้แก้ไขปัญหากรณีที่ข้อมูลมีการสูญหายซึ่งข้อมูลที่สูญหายนั้นเป็นไปอย่างสุ่ม (Missing at Random : MAR) ซึ่งวิธี MI นี้ ข้อมูลที่สูญหายจะถูกแทนที่โดยชุดของค่าที่เป็นไปได้มากกว่า 1 (m > 1) เพื่อที่จะสร้างชุดข้อมูลที่สมบูรณ์ m ชุด ซึ่งจำนวน m ที่เหมาะสมที่จะได้ชุดข้อมูลที่ดีคือ m ตั้งแต่ 3 ถึง 5 (Sandip Sinharay, Hal S. Stern and Daniel Russell ; 2001) ซึ่งประสิทธิภาพของการประมาณค่าจะขึ้นอยู่กับค่า m โดยหาได้จาก

$$(1+\upsilon/m)^{-1}$$

เมื่อ υ เป็นอัตราส่วนของข้อมูลที่สูญหาย

เช่น ถ้าข้อมูลมีค่าสังเกตที่สูญหาย 40% และ m = 5 imputation จะได้ประสิทธิภาพของการประมาณค่าเท่ากับ 93% ในขณะที่ถ้า m = 10 ประสิทธิภาพของการประมาณค่าจะเพิ่มขึ้นอีกเพียงเล็กน้อยเป็น 96% เท่านั้น

แนวคิดพื้นฐานของ MI คือการหา Predictive distribution สำหรับค่าสังเกตที่สูญหายจากค่าสังเกตที่มีอยู่ โดยให้ $Y=(Y_{obs,}Y_{mis})$ ซึ่ง Y_{obs} แทนค่าสังเกตที่มีอยู่ และ Y_{mis} แทนค่าสังเกตที่สูญหาย และสมมติให้ Y มีการแจกแจง $p(Y|\theta)$ ซึ่ง θ เป็นพารามิเตอร์ทั้งหมดของตัวแบบ Predictive distribution จะหาได้จาก

$$p(Y_{mis}|Y_{obs}) = \int p(Y_{mis}, \theta|Y_{obs})d\theta$$

$$= \int p(Y_{mis}|Y_{obs}, \theta)p(\theta|Y_{obs})d\theta$$

จะเห็นได้ว่าเราสามารถประมาณค่าสูญหาย จากการจำลองค่าพารามิเตอร์จากค่าสังเกตที่มีอยู่ ซึ่งมีการแจกแจงโดยประสบการณ์ $p(\theta|Y_{obs})$ แล้วจำลองค่าสูญหายจากการแจกแจงโดยประสบการณ์ที่มีเงื่อนไข $p(Y_{mis}|Y_{obs}, \theta)$ ซึ่ง θ จะหาได้จากขั้นตอนแรก

นอกจากนี้วิธี MI นี้จะอยู่บนพื้นฐานของภาวะน่าจะเป็นสูงสุด (Maximum Likelihood : ML) ซึ่งวิธีการประมาณค่าภาวะน่าจะเป็นสูงสุดมี 3 วิธี คือ 1) วิธี Factoring the likelihood 2) วิธี EM algorithm และ 3) วิธี Direct ML หรือ Full Information ML แต่ในที่นี้เราจะกล่าวเฉพาะการประมาณค่าภาวะน่าจะเป็นสูงสุดโดยวิธี EM algorithm เท่านั้น

ดังนั้นการประมาณค่าสูงสุดโดยวิธี MI ในที่นี้จะแบ่งเป็น 2 ขั้นตอนคือ

1. EM algorithm เป็นการประมาณหาค่าภาวะน่าจะเป็นสูงสุด ซึ่งจะใช้เป็นค่าเริ่มต้นสำหรับการประมาณค่าสูญหาย
2. Data augmentation (DA) algorithm จะแบ่งเป็น 2 ขั้นตอนคือ

   2.1 Imputation Step (I step) : $Y_{mis}^{(r)} \sim p(Y_{mis}|Y_{obs}, \theta^{(r-1)})$

   2.2 Posterior step (P step) : $\theta^{(r)} \sim p(\theta|Y_{obs}, Y_{mis}{}^{(r)})$

   โดยที่ $(Y_{mis}^{(r)}, \theta^{(r)}); r = 1,2,...$ ทำการวนซ้ำจนค่าคู่เข้าสู่การแจกแจง $p(Y_{mis}, \theta|Y_{obs})$

1. Imputation Step

กำหนดให้ $\mu = [\mu_1', \mu_2']$ โดยที่ μ_1 เป็นพารามิเตอร์ของ Y_{obs} ส่วน μ_2 เป็นพารามิเตอร์ของ Y_{mis} และให้ Σ เป็นเมตริกซ์ของความแปรปรวนร่วม ซึ่ง

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}$$

โดยที่ Σ_{11} เป็นเมตริกซ์ความแปรปรวนของ Y_{obs} ส่วน Σ_{22} เป็นเมตริกซ์ความแปรปรวนของ Y_{mis} และ Σ_{12} เป็นเมตริกซ์ความแปรปรวนร่วมระหว่าง Y_{obs} และ Y_{mis}

เมื่อใช้วิธี Sweep Operator (Goodnight, 1979) จะได้

$$\Sigma = \begin{bmatrix} \Sigma_{11}^{-1} & \Sigma_{11}^{-1}\Sigma_{12} \\ -\Sigma_{12}'\Sigma_{11}^{-1} & \Sigma_{22.1} \end{bmatrix}$$

สำหรับข้อมูลที่มีค่าสังเกตสูญหาย การแจกแจงแบบมีเงื่อนไขของ Y_{mis} เมื่อให้ $Y_{obs} = y_1$ มีการแจกแจงแบบปกติหลายตัวแปรด้วยเวกเตอร์ค่าเฉลี่ยดังนี้

$$\mu_{2.1} = \mu_2 + \Sigma_{12}'\Sigma_{11}^{-1}(y_1 - \mu_1)$$

และได้เมตริกซ์ความแปรปรวนแบบมีเงื่อนไขดังนี้

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}$$

ซึ่งจะได้

$$y_{mis} \sim N(\mu_{2.1}, \Sigma_{22.1})$$

นำค่า y_{mis} ที่ประมาณได้ใหม่แทนลงในค่าสังเกตที่สูญหาย แล้วคำนวณเวกเตอร์ค่าเฉลี่ยและเมตริกซ์ความแปรปรวนร่วมเพื่อใช้ในขั้นตอนต่อไป

<u>ตัวประมาณเบส์ของเวกเตอร์ค่าเฉลี่ยและเมตริกซ์ความแปรปรวนร่วม :</u>

เมื่อ y_{ijk} เป็นค่าสังเกตจากการทดลองที่มีการแจกแจงแบบปกติ แล้วเมตริกซ์ความแปรปรวนร่วมหาได้จาก

$$A = \sum_{ijk}(y_{ijk} - \bar{y})(y_{ijk} - \bar{y})'$$

ซึ่งเรียกว่า เมตริกซ์ CSSCP ซึ่งมีการแจกแจงแบบ Wishart Distribution $W(n-1, \Lambda)$
ถ้า A มีการแจกแจงแบบ Wishart Distribution $W(n, \Lambda)$ แล้ว จะได้ว่า $B = A^{-1}$ มีการแจกแจงแบบ Inverted Wishart distribution $W^{-1}(n, \Psi)$ ซึ่ง n เป็นอันดับความเป็นอิสระ และ $\Psi = \Lambda^{-1}$ (Anderson, 1984)

สมมติว่าแต่ละค่าสังเกตในเมตริกซ์ค่าสังเกตมีการแจกแจงแบบปกติหลายตัวแปรซึ่งมีค่าเฉลี่ย μ และเมตริกซ์ความแปรปรวนร่วม Σ แล้ว การแจกแจงแบบ Prior Inverted Wishart Distribution สำหรับ Σ และการแจกแจงแบบปกติก่อนหน้า (Prior normal distribution) สำหรับ μ หาได้จาก

$$\Sigma \sim W^{-1}(m, \Psi)$$

$$\mu \mid \Sigma \sim N\left(\mu_0, \frac{1}{\tau}\Sigma\right)$$

ซึ่ง $\tau > 0$ เป็นจำนวนเต็มคงที่ แล้วการแจกแจงแบบหลัง (Posterior distribution : Anderson 1984, p.270; Schafer 1997, p.152) หาได้จาก

$$\Sigma \mid Y \sim W^{-1}\left(n+m, (n-1)S + \Psi + \frac{n\tau}{n+\tau}(\bar{y}-\mu_0)(\bar{y}-\mu_0)'\right)$$

$$\mu \mid (\Sigma, Y) \sim N\left(\frac{1}{n+\tau}(n\bar{y}+\tau\mu_0), \frac{1}{n+\tau}\Sigma\right)$$

ซึ่ง $(n-1)S$ เป็นเมตริกซ์ CSSCP

2. Posterior Step

ในขั้นตอนนี้ได้จำลองเวกเตอร์ค่าเฉลี่ยของประชากรแบบหลัง (The Posterior population mean vector) μ และเมตริกซ์ความแปรปรวนร่วม Σ จากการแจกแจงก่อนหน้าของ μ และ Σ และค่าประมาณตัวอย่างที่สมบูรณ์ (The Complete sample estimates)

ในการกำหนดข้อมูลพารามิเตอร์ก่อนหน้า (The prior parameter information) นั้นสามารถหาได้ 4 วิธี ได้แก่ A Noninformative Prior, An Informative Prior for μ and Σ, An Informative Prior for Σ และ A Ridge Prior ในการวิจัยครั้งนี้ได้ใช้วิธี A Noninformativer Prior ซึ่งได้การแจกแจงแบบหลัง (Posterior distribution) เหมือนที่ได้อธิบายไว้ในหัวข้อ “ตัวประมาณเบส์ของเวกเตอร์ค่าเฉลี่ยและเมตริกซ์ความแปรปรวนร่วม” ดังนี้

$$\Sigma^{(t+1)} \mid Y \sim W^{-1}\left(n-1, (n-1)S\right)$$

$$\mu^{(t+1)} \mid (\Sigma^{(t+1)}, Y) \quad \sim \quad N\left(\bar{y}, \frac{1}{n}\Sigma^{(t+1)}\right)$$

ซึ่งจะได้

$$y_{mis} \sim N\big(\mu \mid (\Sigma, Y), \Sigma \mid Y\big)$$

นำค่า y_{mis} ที่ประมาณได้ใหม่แทนลงในค่าสังเกตที่สูญหายอีกครั้งเหมือนกับค่า EM Algorithm ที่กำหนดให้เป็นค่าเริ่มต้นในการประมาณค่าสูญหาย และทำการวนซ้ำขั้นตอน I-Step และ P-Step จนลู่เข้าสู่การแจกแจง $p\left(Y_{mis}, \theta \middle| Y_{obs}\right)$ ทำการประมาณค่าสูญหาย 3* ครั้ง และหาค่าเฉลี่ยของค่าประมาณที่ได้เพื่อนำค่าที่ได้ไปใช้เป็นค่าประมาณที่สูญหาย

---

* เนื่องจากประสิทธิภาพของการประมาณค่าจะอยู่ในช่วง 3 ถึง 5 ครั้ง และเมื่อได้ทำการวิจัยแล้วพบว่าประสิทธิภาพในการประมาณมีค่าไม่แตกต่างกันมากนัก ดังนั้นจึงได้กำหนดให้ทำการประมาณค่าสูญหายเพียง 3 ครั้ง

บทที่ 3
วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายในการวางแผนการทดลองแบบจัตุรัสละติน โดยทำการเปรียบเทียบวิธีการประมาณค่าสูญหาย 3 วิธี คือ วิธีกำลังสองน้อยสุด วิธี Expectation Maximization algorithm (EM algorithm) และวิธี Multiple Imputation Method (MI) มีข้อกำหนด คือ ข้อมูลมีการแจกแจงปกติ ขนาดของแผนแบบการทดลองแบ่งเป็น 5 ขนาด คือ 3x3 4x4 5x5 6x6 และ 7x7 กำหนดสัมประสิทธิ์ความผันแปรเป็น 3 ระดับ คือ 5% 25% และ 45% กำหนดจำนวนข้อมูลสูญหาย คือ 10% 20% และ 30% และได้ทำการจำลองข้อมูลตามสถานการณ์ต่างๆ ที่กำหนด แล้วทำการประมาณค่าสูญหายจากข้อมูลที่ได้ และทำการเปรียบเทียบว่าวิธีการประมาณค่าวิธีใดให้ค่าประมาณใกล้เคียงกับค่าจริงมากที่สุด โดยจะพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) วิธีการประมาณค่าสูญหายวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด แสดงว่าวิธีนั้นเหมาะสมสำหรับนำมาใช้ในการประมาณค่าสูญหายมากที่สุด โดยรายละเอียดในบทนี้จะกล่าวถึงวิธีการดำเนินการวิจัยตามลำดับขั้นตอนดังนี้

1. การจำลองข้อมูลให้มีการแจกแจงแบบปกติ
2. การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง
3. การประมาณค่าสังเกตที่สูญหาย
4. การเปรียบเทียบวิธีการประมาณค่าสูญหาย
5. ขั้นตอนการทำงานของโปรแกรม

3.1 การจำลองข้อมูลให้มีการแจกแจงแบบปกติ

ในการวิจัยครั้งนี้ได้ทำการจำลองข้อมูลประชากรให้มีการแจกแจงแบบปกติ ด้วยเทคนิคมอนติคาร์โล ซึ่งใช้ในการจำลองตัวแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน โดยหลักของมอนติคาร์โลนั้นต้องจำลองตัวเลขสุ่ม (Random Number) ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบันแบ่งได้เป็น 2 ขั้นตอนดังนี้

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โลเนื่องจากหลักของมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา
2. การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่มโดยตรงนั้นมีบางขั้นตอนที่ต้องใช้ตัวเลขสุ่ม การเขียนโปรแกรมในงานวิจัยครั้งนี้ใช้ภาษา S-PLUS 2000 ซึ่งการสร้าง

การแจกแจงแบบปกติจะใช้ตัวเลขสุ่มในฟังก์ชัน rnorm โดยมีรายละเอียดในการแจกแจงแบบปกติดังนี้

ในงานวิจัยครั้งนี้กำหนดให้

$$Y_{ijk} = \mu + \tau_i + \beta_j + \alpha_k + \varepsilon_{ijk}$$

และ $\tau_i, \beta_j, \alpha_k, \varepsilon_{ijk}$ เป็นตัวแปนสุ่มที่เป็นอิสระกันและมีการแจกแจงแบบปกติที่มี $E(\tau_i) = E(\beta_j) = E(\alpha_k) = E(\varepsilon_{ijk}) = 0$ และ $Var(\tau_i) = \sigma_\tau^2$, $Var(\beta_j) = \sigma_\beta^2$, $Var(\alpha_k) = \sigma_\alpha^2$, $Var(\varepsilon_{ijk}) = \sigma_\varepsilon^2$ โดยที่ $\sigma_\tau^2 = \sigma_\beta^2 = \sigma_\alpha^2 = h\sigma_\varepsilon^2$ ซึ่ง h เป็นค่าจำนวนเต็มคงที่ ดังนั้นเราจะได้ค่า y_{ijk} ซึ่งเป็นค่าสังเกตในการทดลองนั้นๆ

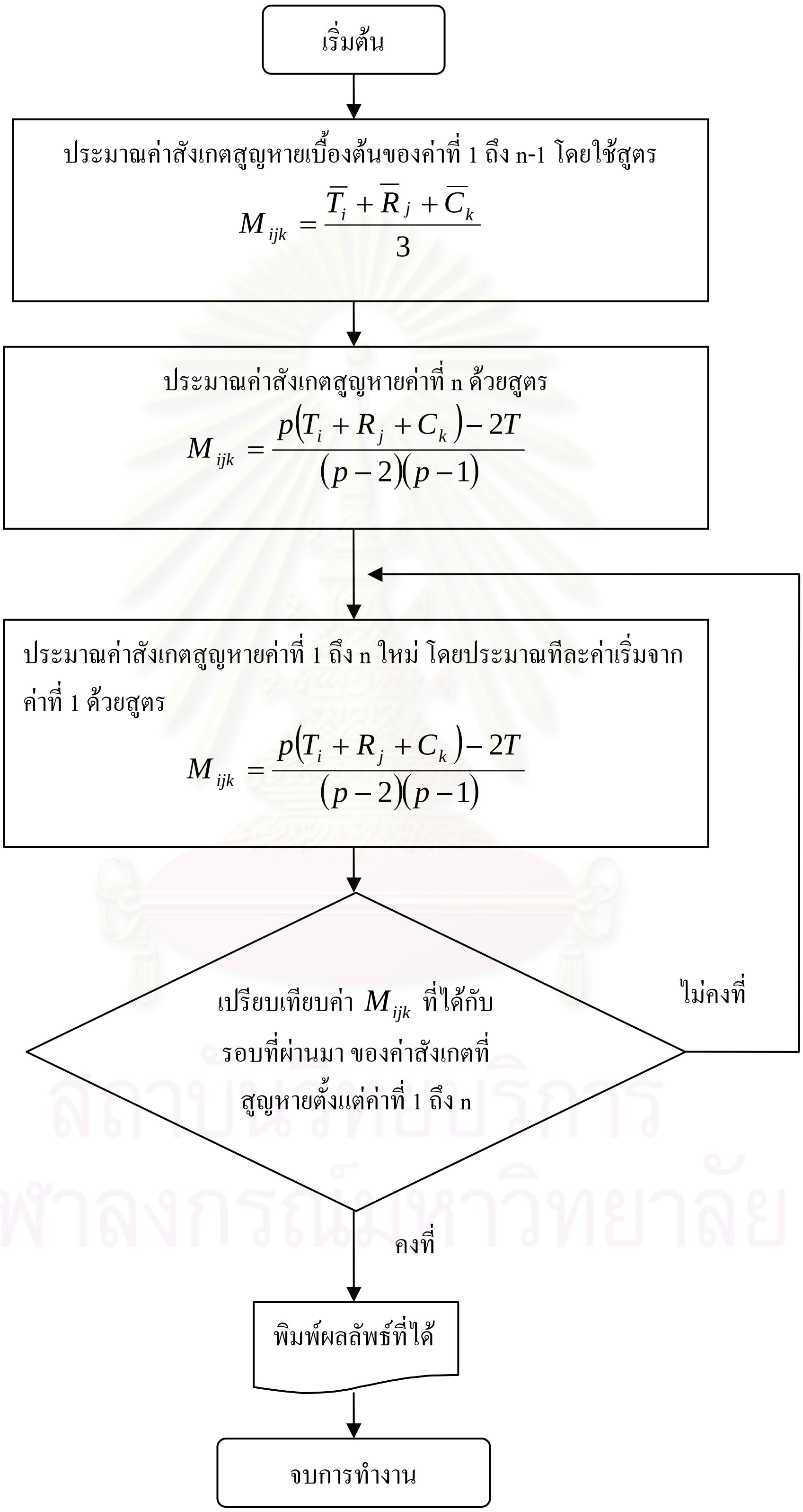
3.2 การสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริง

เมื่อข้อมูลถูกสร้างเสร็จเรียบร้อยแล้ว ในขั้นตอนนี้จะทำการสุ่มตัดข้อมูลออกด้วยฟังก์ชันการแจกแจงแบบยูนิฟอร์ม และข้อมูลที่ถูกตัดออกไปนั้นจะเก็บไว้เพื่อนำมาเปรียบเทียบกับค่าใหม่ที่จะประมาณขึ้นมาทั้ง 3 วิธี

3.3 การประมาณค่าสังเกตที่สูญหาย

เมื่อสร้างข้อมูล y_{ijk} ที่เป็นไปตามข้อกำหนดข้างต้นและได้ทำการสุ่มตัดข้อมูลให้เหมือนกับการสูญหายจริงเรียบร้อยแล้ว นำข้อมูลที่เหลือไปคำนวณหาค่าประมาณด้วยวิธีการประมาณค่าสูญหาย 3 วิธีได้แก่ วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด วิธีประมาณค่าสูญหายโดยวิธี Expectation Maximization algorithm และวิธีประมาณค่าสูญหายโดยวิธี Multiple Imputation ดังรายละเอียดต่อไปนี้

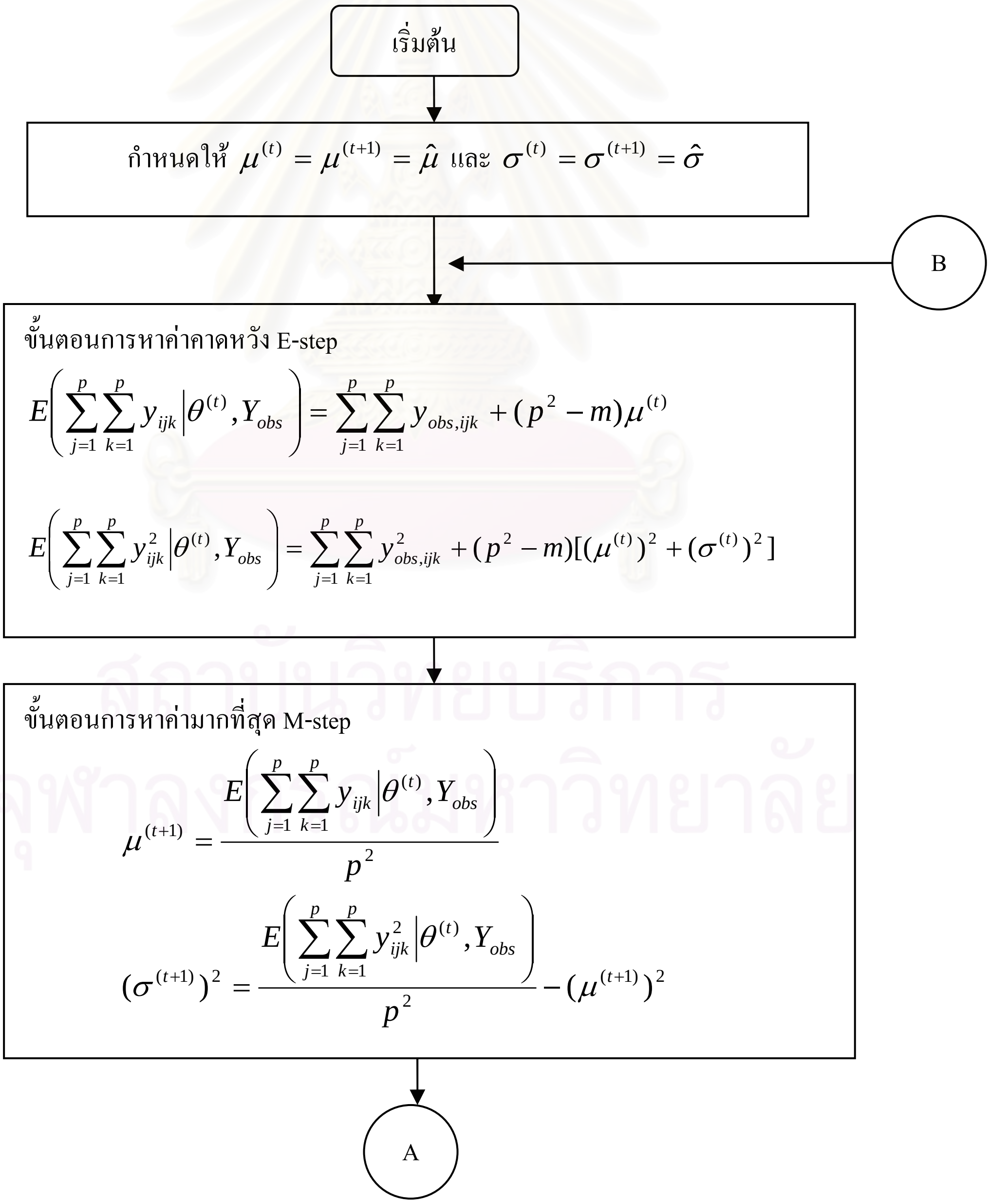
3.3.1 วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (Least square method)

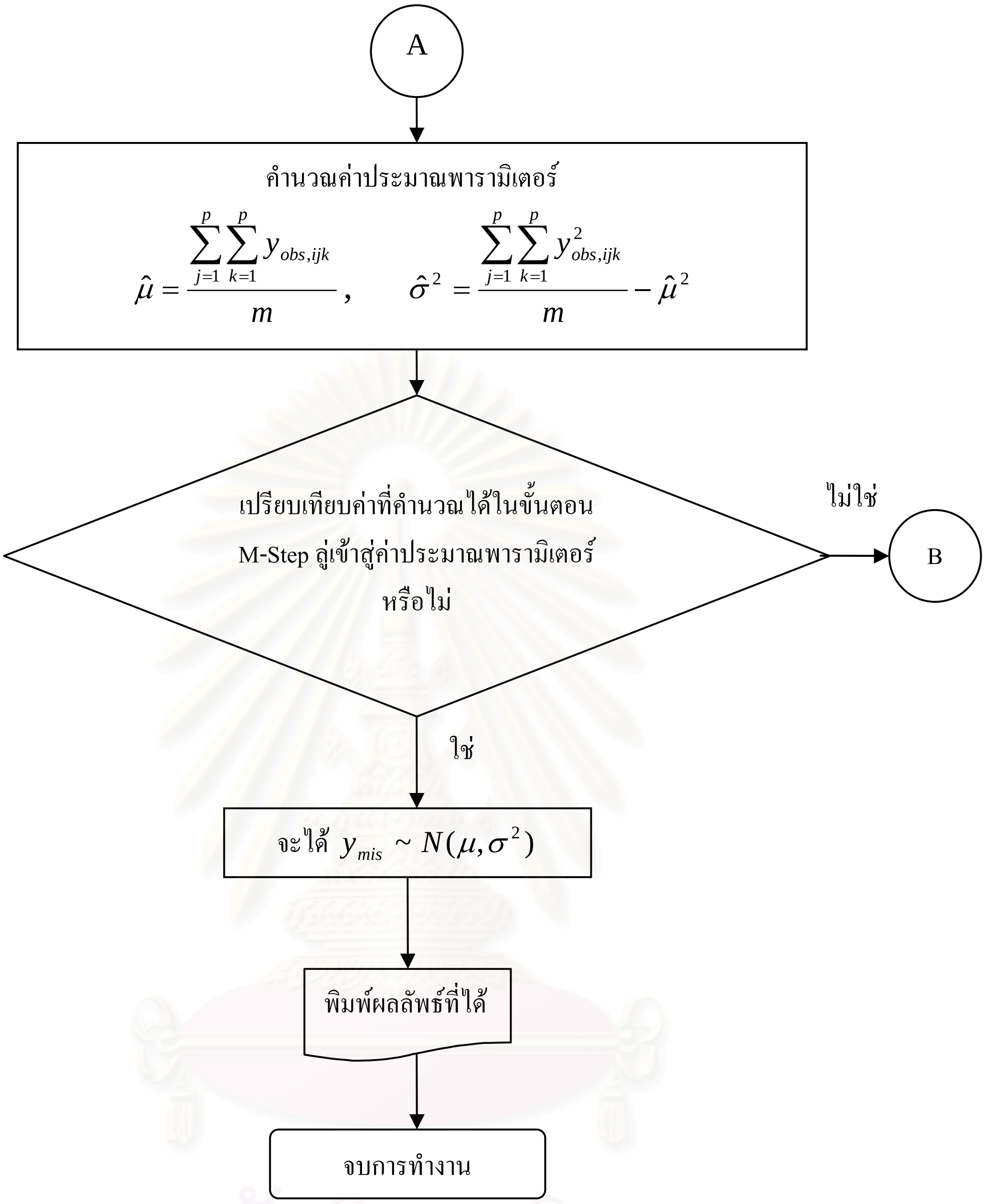


3.3.2 วิธีประมาณค่าสูญหายโดยวิธี Expectation Maximization algorithm (EM algorithm)

EM algorithm จะแบ่งออกเป็น 2 ขั้นตอน ขั้นตอนแรกคือ Expectation step (E-step) เป็นขั้นตอนการหาค่าคาดหวังหรือเรียกว่าการหาค่าตัวสถิติพอเพียง และขั้นตอนที่สองคือ Maximization step (M-step) เป็นขั้นตอนการหาค่ามากที่สุดของพารามิเตอร์ที่จะไม่เพิ่มขึ้นอีกเมื่อเทียบกับรอบที่ t-1 โดยในการวิจัยครั้งนี้ สมมติให้ Y มีข้อมูลทั้งหมด p^2 ค่า มีค่าสังเกตที่เก็บมาได้ m ค่า และมีค่าสังเกตที่สูญหาย $p^2 - m$ ค่า ดังนั้น Y จึงประกอบด้วย 2 ส่วน คือ $Y = (Y_{mis}, Y_{obs})$

สำหรับการประมาณค่าสูญหายโดยวิธี EM algorithm นั้นก่อนอื่นจะต้องพิจารณาการแจกแจงของข้อมูลของค่าสังเกตที่เหลืออยู่ว่ามีการแจกแจงแบบปกติหรือไม่ ถ้าหากมีการแจกแจงแบบปกติ จะเริ่มประมาณค่าสูญหายตามขั้นตอนดังต่อไปนี้




A
คำนวณค่าประมาณพารามิเตอร์
$\hat{\mu} = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}}{m}$, $\hat{\sigma}^2 = \frac{\sum_{j=1}^{p}\sum_{k=1}^{p} y_{obs,ijk}^2}{m} - \hat{\mu}^2$
เปรียบเทียบค่าที่คำนวณได้ในขั้นตอน M-Step ลู่เข้าสู่ค่าประมาณพารามิเตอร์หรือไม่
ไม่ใช่
B
ใช่
จะได้ $y_{mis} \sim N(\mu, \sigma^2)$
พิมพ์ผลลัพธ์ที่ได้
จบการทำงาน

3.3.3 วิธีประมาณค่าสูญหายโดยวิธี Multiple Imputation

เริ่มต้น

C

กำหนดให้ค่าประมาณที่คำนวณได้จากวิธี EM Algorithm
เป็นค่าเริ่มต้นที่ถูกแทนลงในค่าสังเกตที่สูญหาย

B

Imputation Step (I step) :

คำนวณค่าเฉลี่ยใหม่จากสูตร

$$\mu_{2.1} = \mu_2 + \Sigma_{12}'\Sigma_{11}^{-1}(y_1 - \mu_1)$$

และได้เมตริกซ์ความแปรปรวนจากสูตร

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}$$

โดยที่

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}$$ เป็นเมตริกซ์ของความแปรปรวนร่วม

ซึ่งจะได้

$$y_{mis} \sim N(\mu_{2.1}, \Sigma_{22.1})$$

นำค่า y_{mis} ที่ประมาณได้ใหม่แทนลงในค่าสังเกตที่สูญหาย

Posterior Step (P step) :

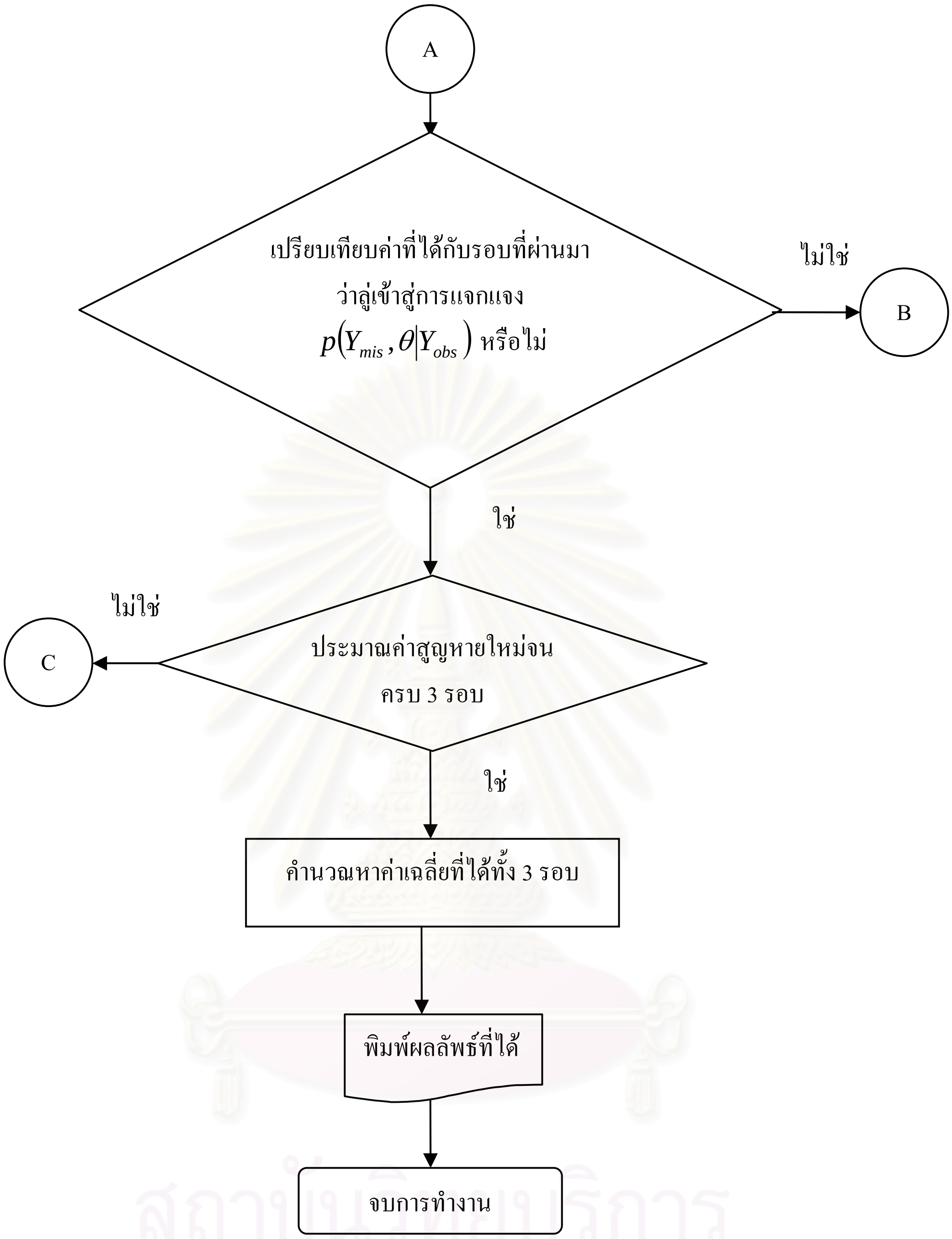
$$\Sigma^{(t+1)} \mid Y \quad \sim \quad W^{-1}\left(n-1, (n-1)S\right)$$

$$\mu^{(t+1)} \mid (\Sigma^{(t+1)}, Y) \quad \sim \quad N\left(\bar{y}, \frac{1}{n}\Sigma^{(t+1)}\right)$$

ซึ่งจะได้ค่าประมาณตัวใหม่ดังนี้

$$y_{mis} \sim N\left(\mu \mid (\Sigma, Y), \Sigma \mid Y\right)$$

A


A
เปรียบเทียบค่าที่ได้กับรอบที่ผ่านมา
ว่าลู่เข้าสู่การแจกแจง
$p(Y_{mis},\theta|Y_{obs})$ หรือไม่
ไม่ใช่
B
ใช่
ไม่ใช่
C
ประมาณค่าสูญหายใหม่จน
ครบ 3 รอบ
ใช่
คำนวณหาค่าเฉลี่ยที่ได้ทั้ง 3 รอบ
พิมพ์ผลลัพธ์ที่ได้
จบการทำงาน

3.4 การเปรียบเทียบวิธีการประมาณค่าสูญหาย

วิธีการประมาณค่าสังเกตสูญหายวิธีใดจะให้ค่าประมาณที่ดี จะพิจารณาจากเกณฑ์ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) ซึ่งมีสูตรดังนี้

$$MAE_t = Max\left[Max\left|y_i - \hat{y}_i\right|\right]_j$$

เมื่อ

y_i แทน ค่าจริงค่าที่ i ที่ได้จากการจำลอง

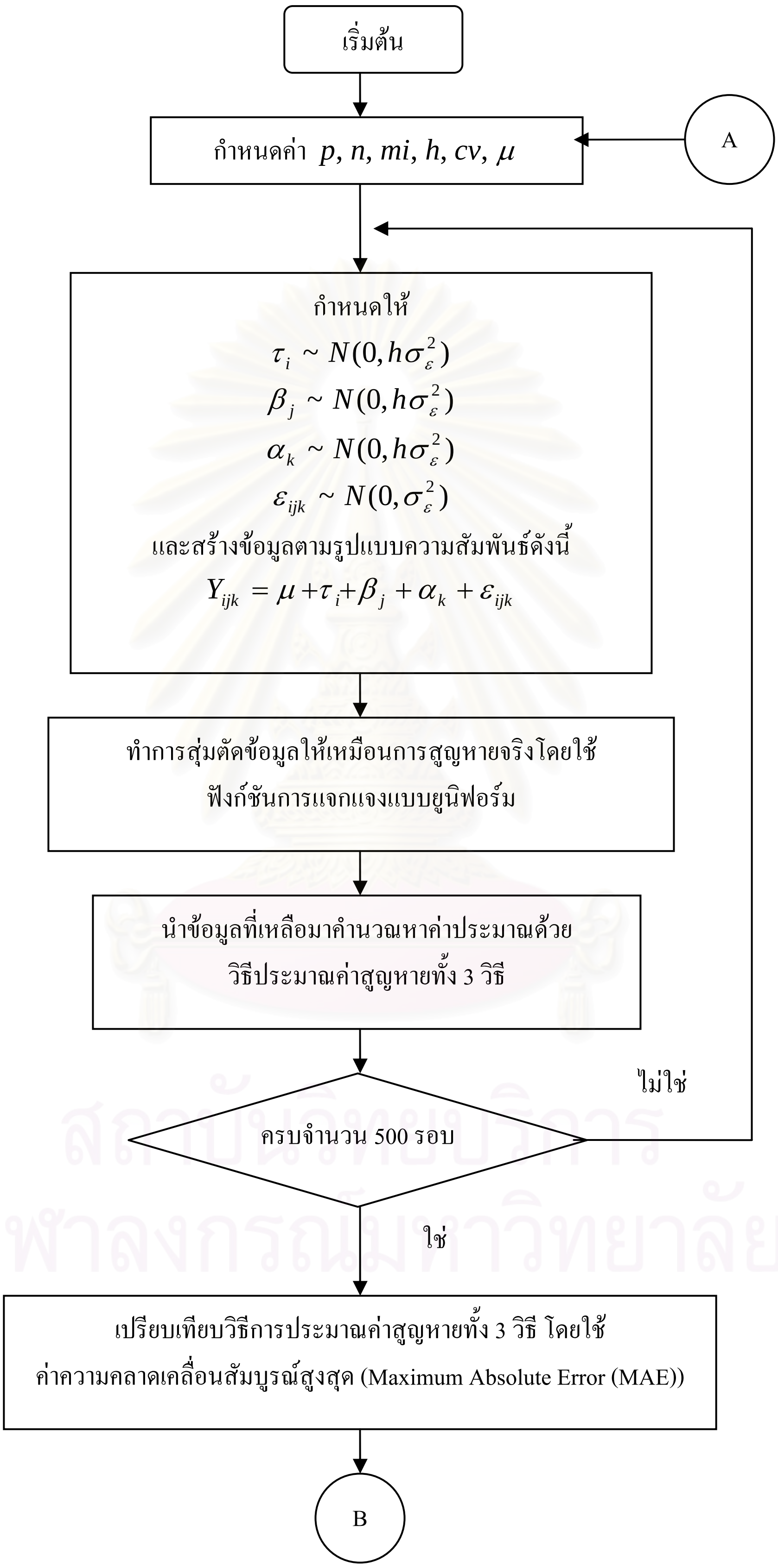
$\hat{y}_i$ แทน ค่าประมาณ ค่าที่ i ที่ได้จากการใช้วิธีการประมาณค่า

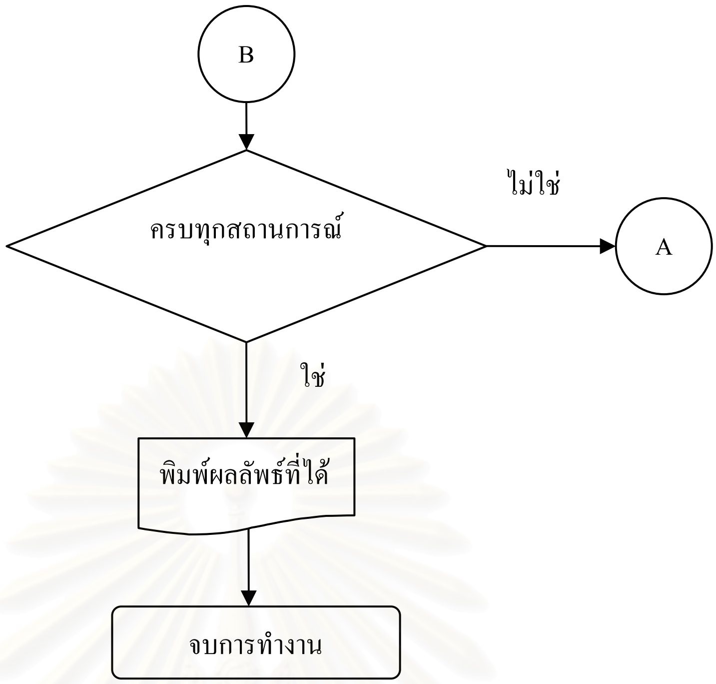
j แทน รอบที่ j จากการทดลอง (ในที่นี้กำหนดให้ j = 500 รอบ)

MAE_t แทน ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดจากวิธีที่ t

จะพิจารณาว่าวิธีการประมาณค่าสูญหายวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) มีค่าต่ำสุด จะเป็นวิธีการประมาณค่าสังเกตสูญหายที่ดีที่สุด

3.5 ขั้นตอนการทำงานของโปรแกรม




B
ครบทุกสถานการณ์
ไม่ใช่
A
ใช่
พิมพ์ผลลัพธ์ที่ได้
จบการทำงาน

บทที่ 4
ผลการวิจัย

ผลการวิจัยครั้งนี้ได้จากการจำลองข้อมูลด้วยโปรแกรม S-PLUS 2000 โดยมีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายในการวางแผนการทดลองแบบจัตุรัสละติน โดยทำการประมาณค่าสังเกตที่สูญหาย 3 วิธี คือ วิธีกำลังสองน้อยสุด วิธี Expectation Maximization algorithm (EM algorithm) และวิธี Multiple Imputation Method (MI) ทำการศึกษาตามสถานการณ์ต่างๆ ดังนี้ จำนวนระดับปัจจัยทดลองเป็น 3 4 5 6 และ 7 กำหนดสัมประสิทธิ์ความผันแปรเป็น 5% 25% และ 45% จำนวนข้อมูลสูญหายเป็น 10% 20% และ 30% กำหนดให้ค่าคงที่ h เป็น 1 2 และ 3 ได้ทำการจำลองข้อมูลโดยใช้เทคนิคมอนติคาร์โล (Monte Carol Simulation) โดยทำการจำลองซ้ำสถานการณ์ละ 500 รอบ โดยประสิทธิภาพในการประมาณค่าสังเกตที่สูญหายในแต่ละวิธีนั้นจะพิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (Maximum Absolute Error (MAE)) ซึ่งวิธีใดที่ให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำสุด จะเป็นวิธีการประมาณค่าสังเกตสูญหายที่ดีที่สุด เนื่องจากค่าสังเกตที่ได้จากการประมาณนั้นมีค่าใกล้เคียงกับค่าสังเกตจริงที่สูญหายไปมากที่สุด

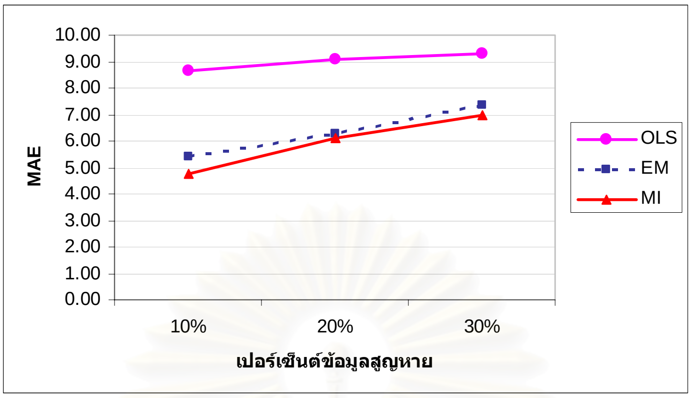
การนำเสนอผลของการวิจัยจะอยู่ในรูปตารางและกราฟ ค่าในตารางเป็นค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด แต่ละสถานการณ์ใช้ทศนิยม 4 ตำแหน่ง สำหรับเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหาย ส่วนการนำเสนอด้วยกราฟจะอธิบายปัจจัยที่มีผลต่อการประมาณค่าสังเกตที่สูญหาย และแนวโน้มของการประมาณค่าสังเกตที่สูญหาย ซึ่งผู้วิจัยได้นำเสนอผลการวิจัยดังนี้

4.1 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ณ เปอร์เซ็นต์ข้อมูลสูญหาย เมื่อกำหนดให้ค่าคงที่ h และสัมประสิทธิ์ความผันแปรคงที่

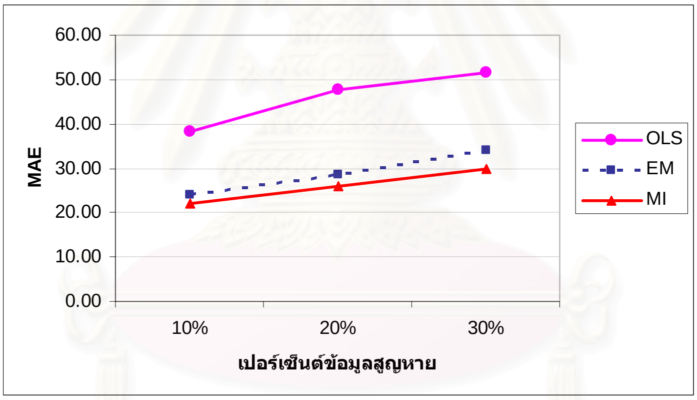
ตารางที่ 4.1.1 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน โดยพิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

h	c.v	missing	MAE OLS	MAE EM	MAE MI
1	0.05	10%	8.6599	5.3896	4.7439
		20%	9.0970	6.2450	6.1115
		30%	9.3009	7.3435	6.9981
	0.25	10%	38.3618	24.0602	22.0340
		20%	47.6126	28.4760	25.9879
		30%	51.5757	34.0191	29.9957
	0.45	10%	71.7228	45.9674	39.8221
		20%	82.5229	57.2253	50.4489
		30%	92.5253	73.2552	69.0453
2	0.05	10%	6.2909	6.1418	5.4974
		20%	8.5862	6.2905	6.0110
		30%	9.2088	7.9365	6.5043
	0.25	10%	29.5559	28.1828	24.5889
		20%	38.0925	30.7788	26.6395
		30%	42.2902	32.5284	30.9707
	0.45	10%	60.7241	49.3142	41.9943
		20%	72.2710	50.1778	46.9934
		30%	81.5236	73.3290	62.6600
3	0.05	10%	5.6183	4.8229	4.0854
		20%	5.6828	5.6306	5.4458
		30%	6.7541	6.4981	5.8000
	0.25	10%	24.9180	24.7828	21.6091
		20%	30.0515	27.4357	24.0875
		30%	35.2088	28.9352	24.4002
	0.45	10%	49.1727	46.6222	41.7862
		20%	59.1564	52.1900	49.2983
		30%	65.9787	58.8536	54.5361

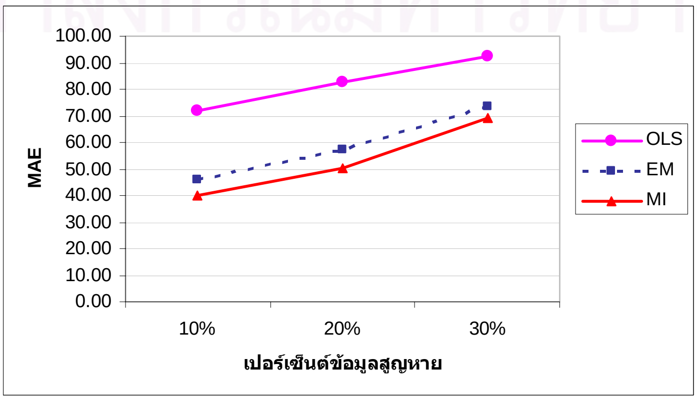
รูปที่ 4.1.1 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.05



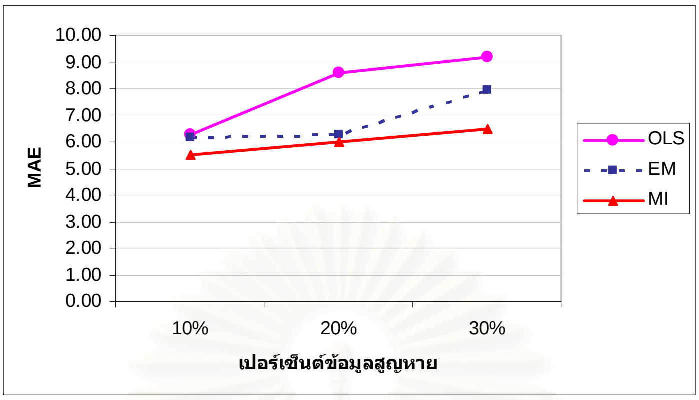
รูปที่ 4.1.2 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.25



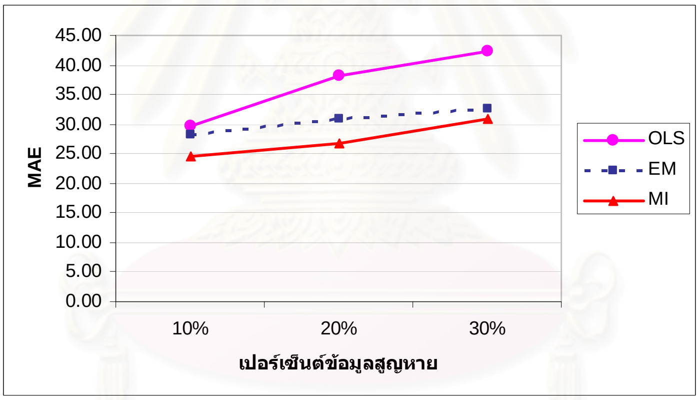
รูปที่ 4.1.3 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.45



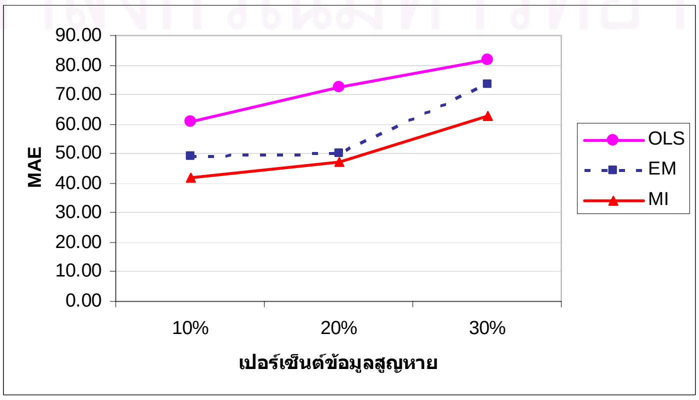
รูปที่ 4.1.4 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.05



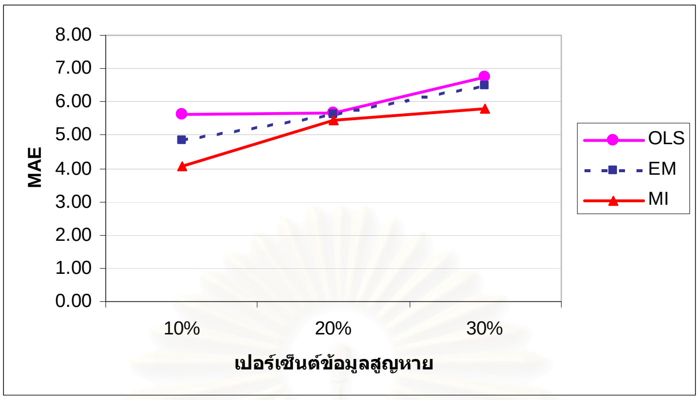
รูปที่ 4.1.5 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.25



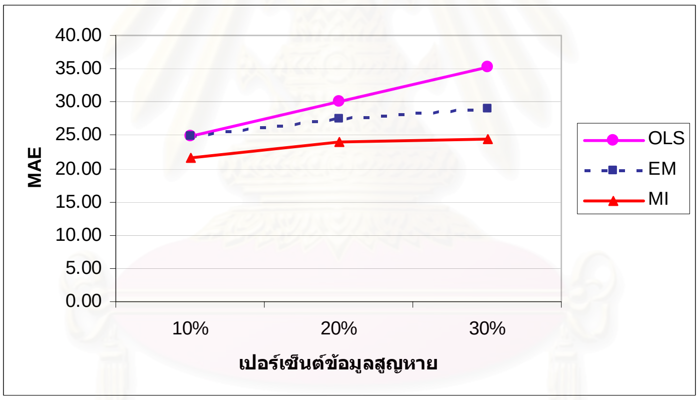
รูปที่ 4.1.6 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.45



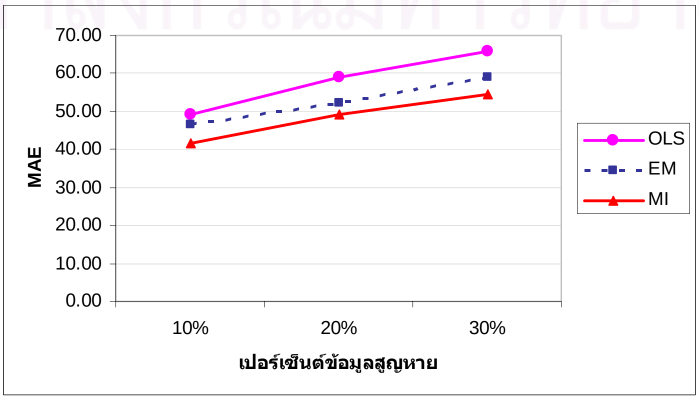
รูปที่ 4.1.7 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.1.8 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.25



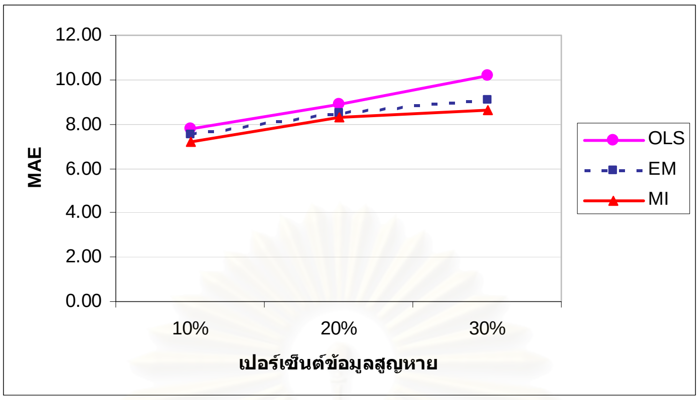
รูปที่ 4.1.9 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.45



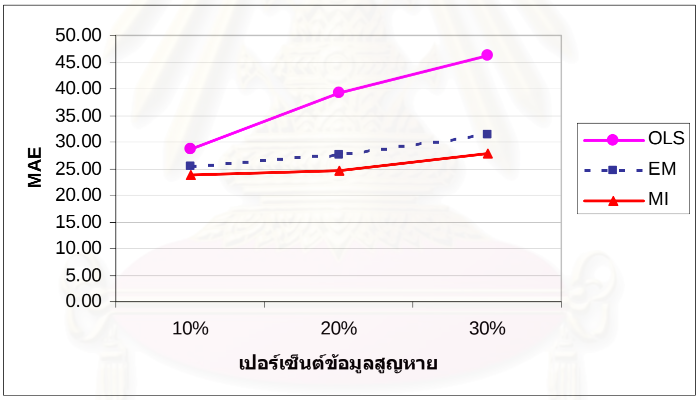
ตารางที่ 4.1.2 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน โดยพิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

h	c.v	missing	MAE OLS	MAE EM	MAE MI
		10%	7.7791	7.5409	7.2314
	0.05	20%	8.8793	8.4767	8.2876
		30%	10.2139	9.0966	8.6274
		10%	28.6270	25.5202	23.7884
1	0.25	20%	39.1272	27.4871	24.6940
		30%	46.1516	31.4099	27.8118
		10%	54.8865	43.4449	35.2345
	0.45	20%	73.4797	47.4956	42.6461
		30%	80.9840	48.5235	42.6717
		10%	6.2162	6.0463	5.8739
	0.05	20%	6.6138	6.1800	6.1218
		30%	9.7351	8.6344	7.9183
		10%	23.7620	22.4726	21.5473
2	0.25	20%	29.0753	25.8020	22.1026
		30%	41.5674	25.9717	22.9555
		10%	45.8605	38.1680	36.7966
	0.45	20%	52.2904	40.0499	37.7904
		30%	59.6147	47.6554	39.6471
		10%	4.6076	4.5990	4.5138
	0.05	20%	5.4724	5.4314	5.3789
		30%	7.7195	7.0991	6.5956
		10%	24.2024	20.3031	19.5846
3	0.25	20%	28.5308	23.9133	21.9964
		30%	30.2066	24.1183	22.1260
		10%	32.5013	31.7206	30.6984
	0.45	20%	47.2045	37.5362	34.3447
		30%	53.9431	43.4060	35.4169

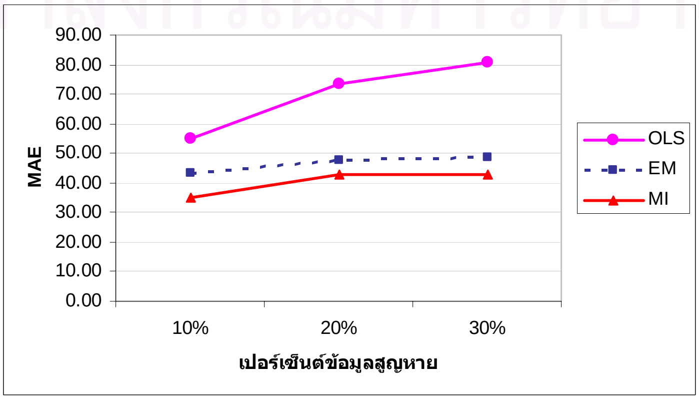
รูปที่ 4.1.10 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.05



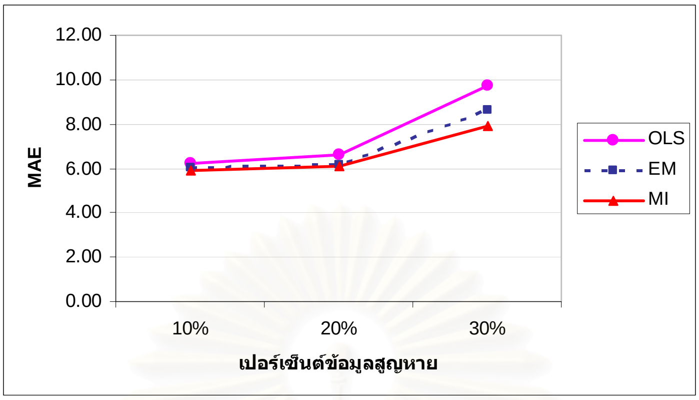
รูปที่ 4.1.11 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.25



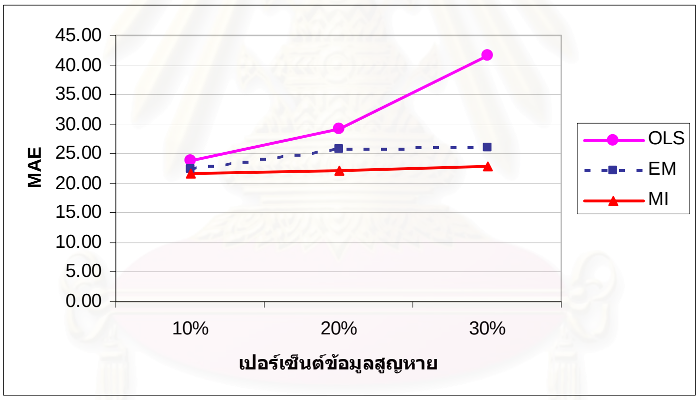
รูปที่ 4.1.12 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.45



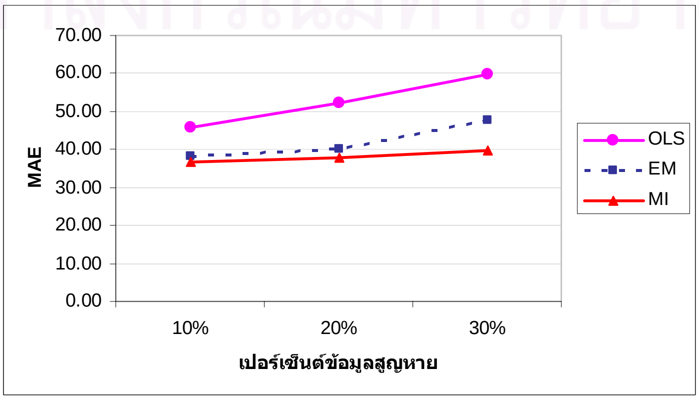
รูปที่ 4.1.13 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.05



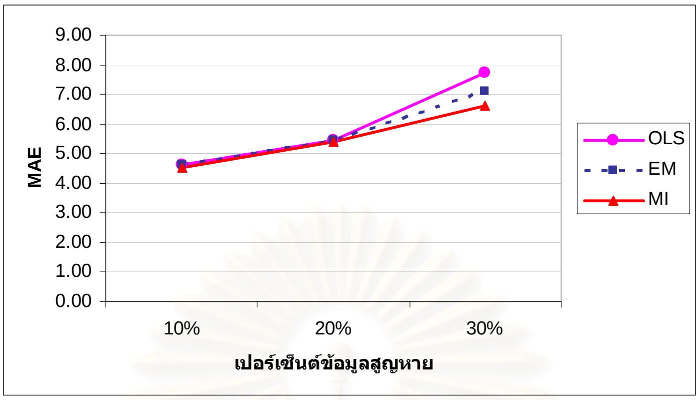
รูปที่ 4.1.14 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.25



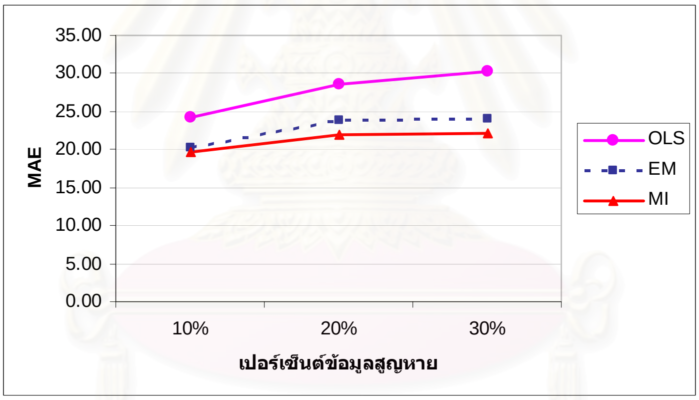
รูปที่ 4.1.15 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.45



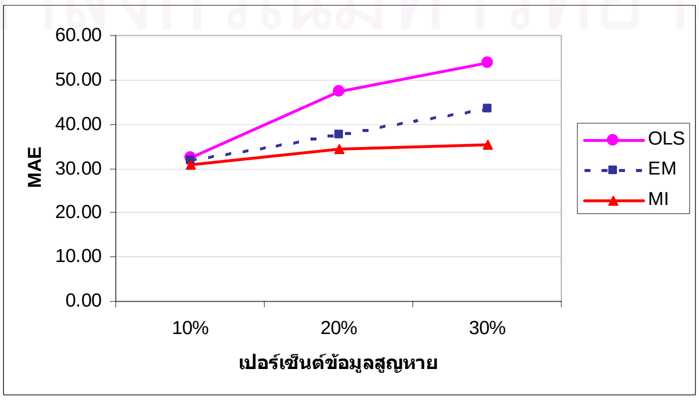
รูปที่ 4.1.16 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.1.17 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.25



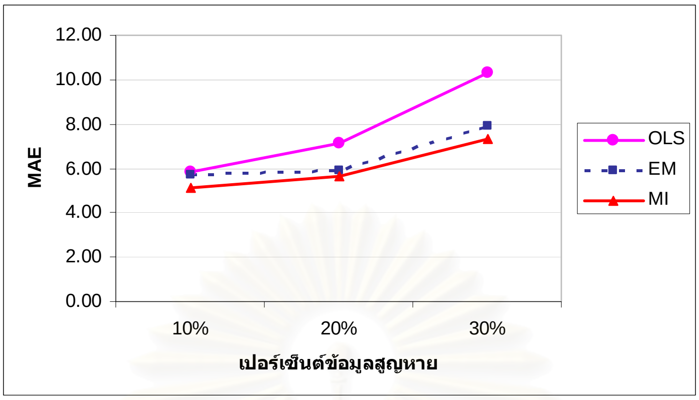
รูปที่ 4.1.18 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.45



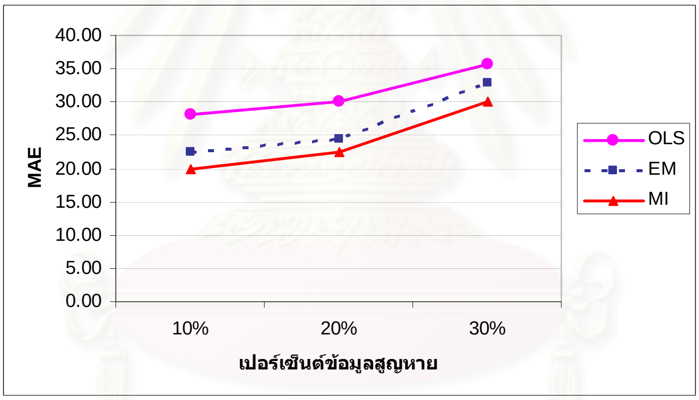
ตารางที่ 4.1.3 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน โดยพิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

h	c.v	missing	MAE OLS	MAE EM	MAE MI
1	0.05	10%	5.8682	5.6968	5.0957
		20%	7.1298	5.8971	5.6700
		30%	10.3068	7.9281	7.3466
	0.25	10%	28.1094	22.3784	19.7854
		20%	30.0478	24.3331	22.5674
		30%	35.7777	32.9384	30.0074
	0.45	10%	50.9904	45.5139	40.9688
		20%	53.2704	50.5609	48.5021
		30%	61.3002	54.4515	49.9025
2	0.05	10%	4.8969	4.7649	4.1803
		20%	5.3906	5.0056	4.2326
		30%	6.8966	6.4575	5.9949
	0.25	10%	21.3309	19.3479	17.3448
		20%	22.0811	20.7711	20.6736
		30%	30.9132	29.6125	26.0193
	0.45	10%	36.4578	31.9861	28.9385
		20%	41.7682	32.1911	29.1245
		30%	50.1699	46.2479	39.7025
3	0.05	10%	4.0798	4.0432	3.7023
		20%	5.6796	5.3929	4.0337
		30%	6.5185	6.3165	5.9189
	0.25	10%	16.3850	15.7351	13.4896
		20%	20.5194	19.1290	16.9755
		30%	24.2990	24.1145	22.4991
	0.45	10%	31.7296	31.7072	27.6233
		20%	37.5360	35.7913	32.6195
		30%	42.9053	39.9780	37.0511

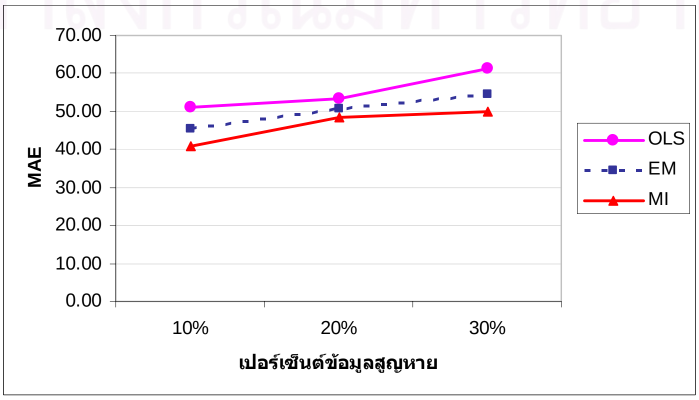
รูปที่ 4.1.19 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.05



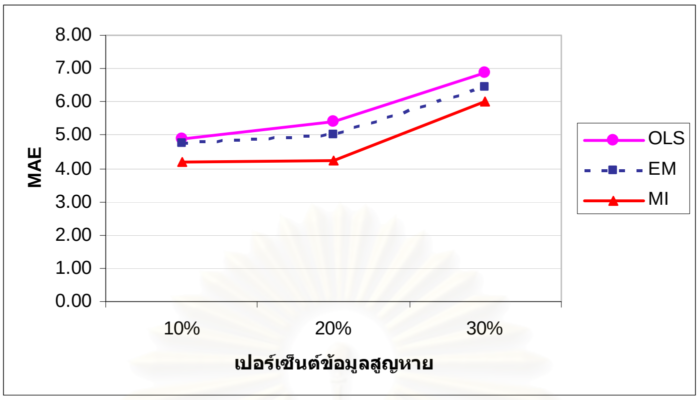
รูปที่ 4.1.20 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.25



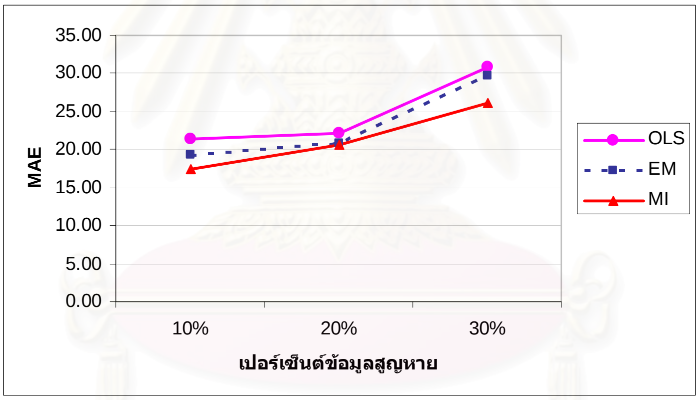
รูปที่ 4.1.21 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.45



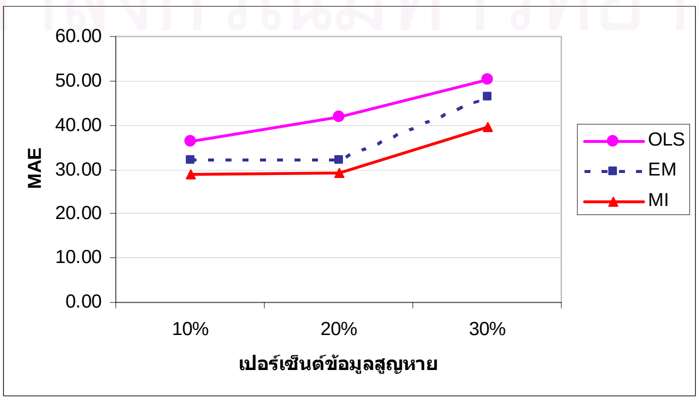
รูปที่ 4.1.22 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.05



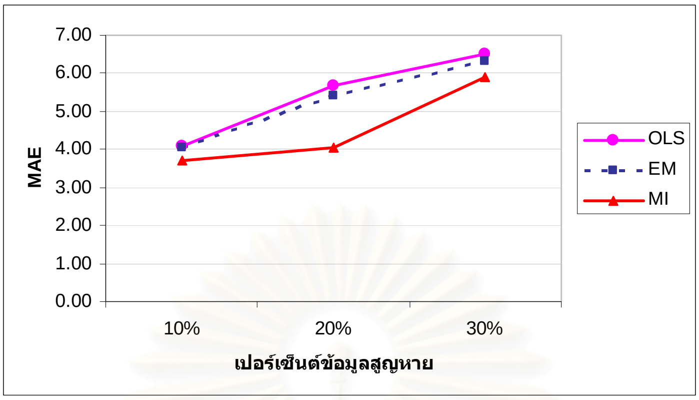
รูปที่ 4.1.23 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.25



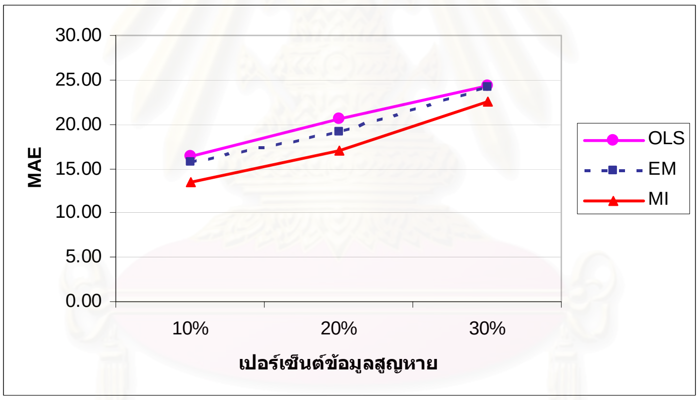
รูปที่ 4.1.24 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.45



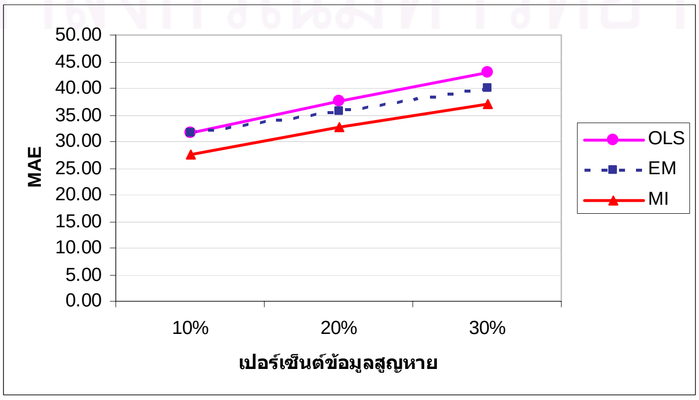
รูปที่ 4.1.25 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.1.26 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.25



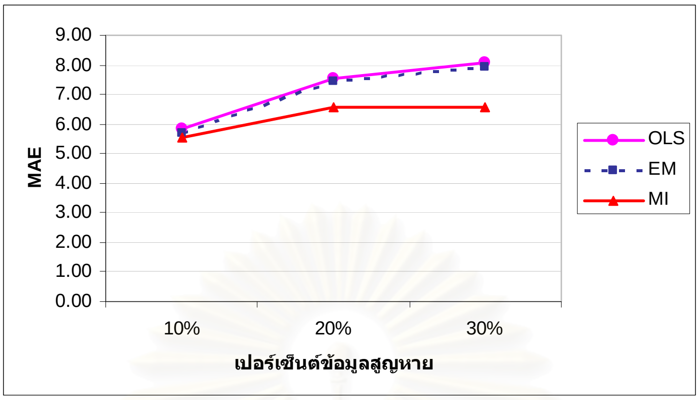
รูปที่ 4.1.27 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.45



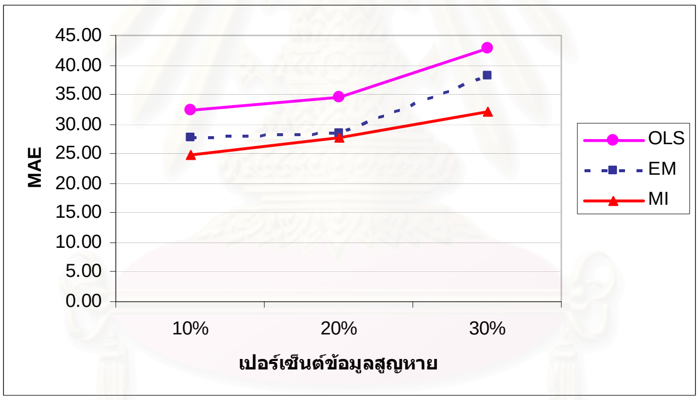
ตารางที่ 4.1.4 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน โดยพิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

h	c.v	missing	MAE OLS	MAE EM	MAE MI
1	0.05	10%	5.8172	5.6917	5.5372
		20%	7.5641	7.4550	6.5478
		30%	8.0636	7.9331	6.5879
	0.25	10%	32.2936	27.7421	24.9148
		20%	34.5393	28.4741	27.7647
		30%	42.7861	38.1336	32.1790
	0.45	10%	49.3574	45.6812	42.9947
		20%	54.7191	46.9111	43.5965
		30%	79.0797	62.2092	53.9579
2	0.05	10%	4.7243	4.6947	3.9671
		20%	6.0893	6.1567	5.4906
		30%	6.2787	6.2280	5.8936
	0.25	10%	19.7685	19.3838	18.6600
		20%	25.1802	22.9325	20.5171
		30%	25.3932	24.9744	20.8245
	0.45	10%	41.5621	39.5780	36.9268
		20%	47.9628	44.9490	40.6498
		30%	51.7270	49.8744	40.8616
3	0.05	10%	5.0199	5.0004	4.4029
		20%	5.9174	5.9066	5.6390
		30%	7.0444	6.8368	6.1050
	0.25	10%	21.1072	21.1000	19.2046
		20%	24.8663	21.9596	19.8567
		30%	25.1659	25.0756	21.4612
	0.45	10%	36.4664	33.5128	32.4316
		20%	45.1973	39.7520	36.4955
		30%	45.9428	44.8748	38.5298

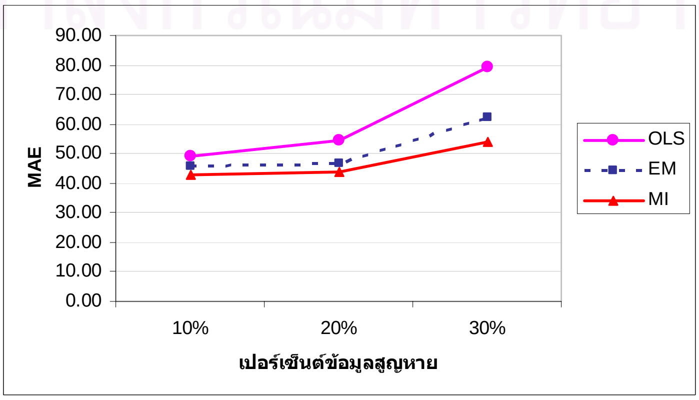
รูปที่ 4.1.28 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.05



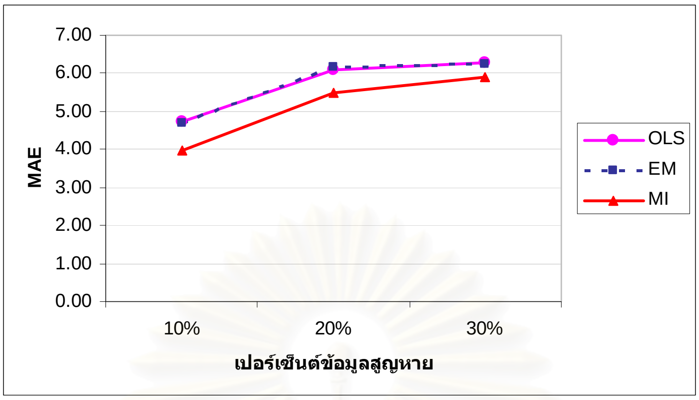
รูปที่ 4.1.29 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.25



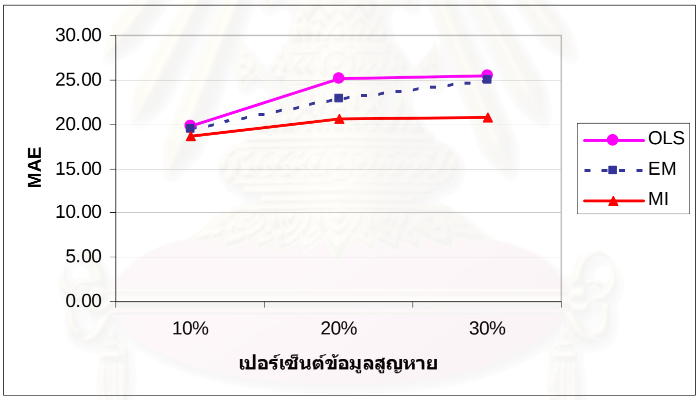
รูปที่ 4.1.30 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.45



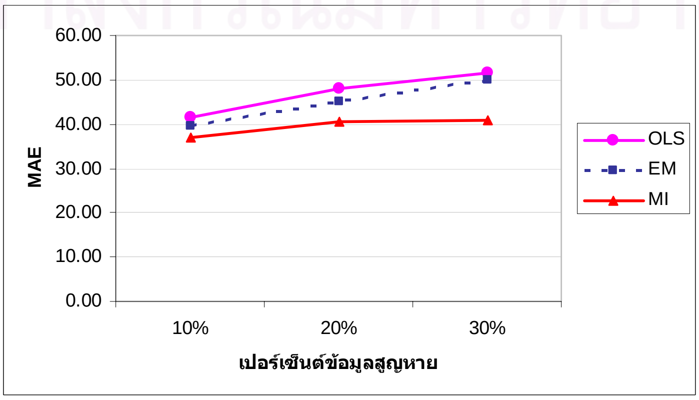
รูปที่ 4.1.31 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.05



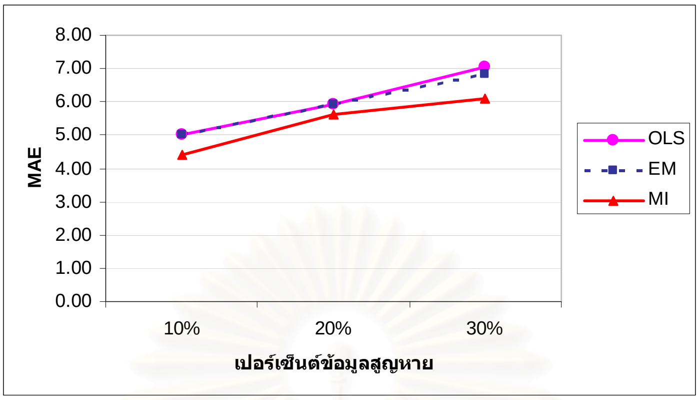
รูปที่ 4.1.32 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.25



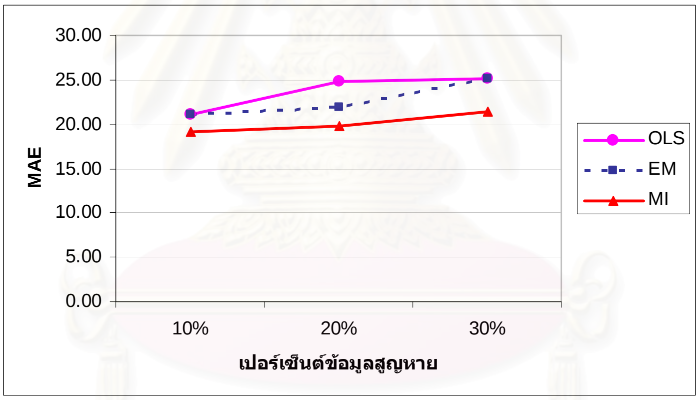
รูปที่ 4.1.33 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.45



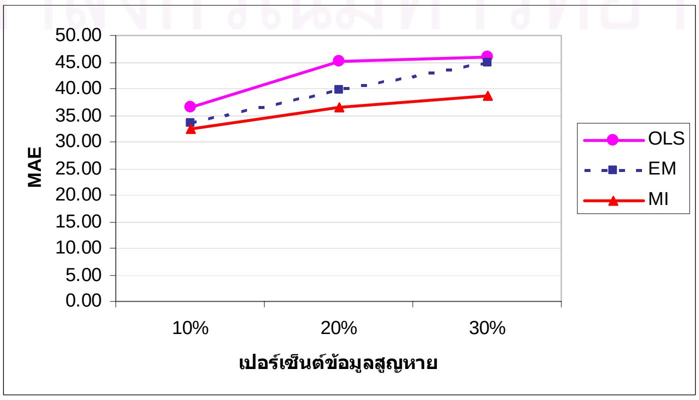
รูปที่ 4.1.34 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.1.35 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.25



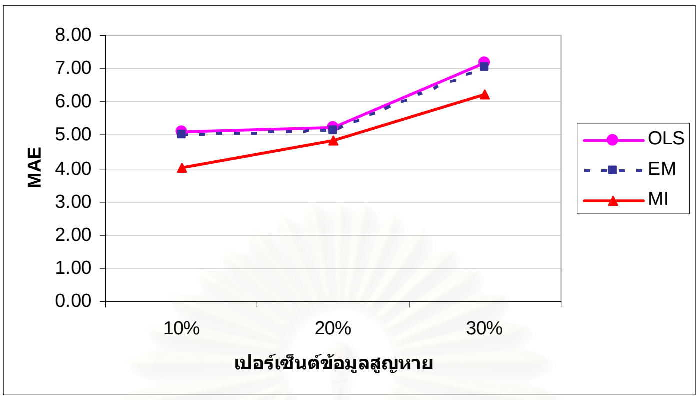
รูปที่ 4.1.36 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.45



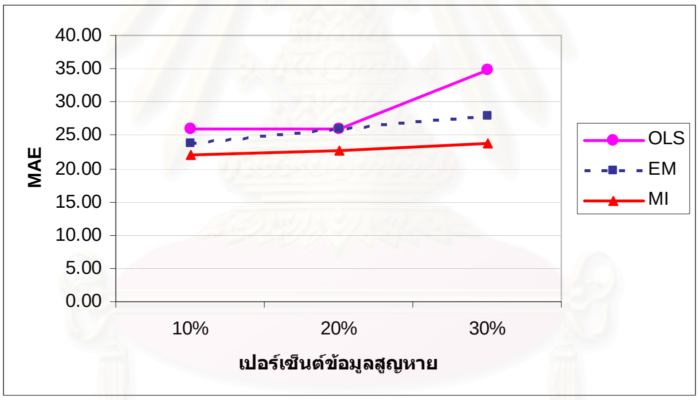
ตารางที่ 4.1.5 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน โดยพิจารณาจากเปอร์เซ็นต์ข้อมูลสูญหาย

h	c.v	mi	MAE OLS	MAE EM	MAE MI
1	0.05	10%	5.1068	5.0130	4.0236
		20%	5.2391	5.1356	4.8602
		30%	7.1953	7.0362	6.2244
	0.25	10%	25.8864	23.7677	22.1253
		20%	25.9979	25.8966	22.6826
		30%	34.7477	27.9877	23.8433
	0.45	10%	48.0682	41.0144	39.2961
		20%	50.5838	49.7984	40.2521
		30%	63.3971	51.8287	44.3837
2	0.05	10%	4.4951	4.4726	3.9210
		20%	4.7138	4.6997	3.9301
		30%	6.5039	6.5012	6.1110
	0.25	10%	21.2013	19.2644	17.2893
		20%	23.0790	22.8575	20.6931
		30%	29.0086	28.6546	21.9417
	0.45	10%	40.5607	36.2060	32.8648
		20%	46.2886	45.1359	38.3525
		30%	50.6945	49.9759	42.5841
3	0.05	10%	3.7407	3.6758	3.0719
		20%	3.9915	3.9866	3.1730
		30%	6.2119	6.1702	5.9378
	0.25	10%	19.5979	19.2842	16.0675
		20%	20.0283	19.9914	18.4019
		30%	26.3185	23.8886	18.8652
	0.45	10%	33.7431	31.6420	23.3490
		20%	35.0983	34.6465	25.7382
		30%	48.8774	47.9632	41.6794

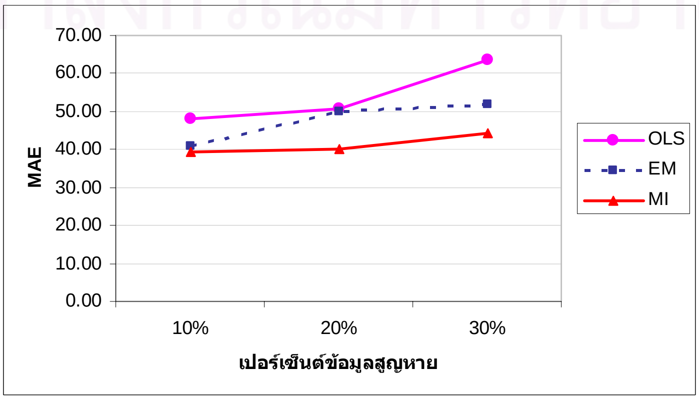
รูปที่ 4.1.37 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.05



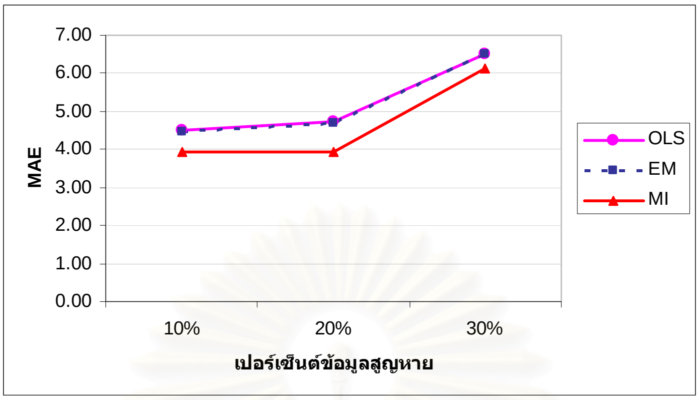
รูปที่ 4.1.38 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.25



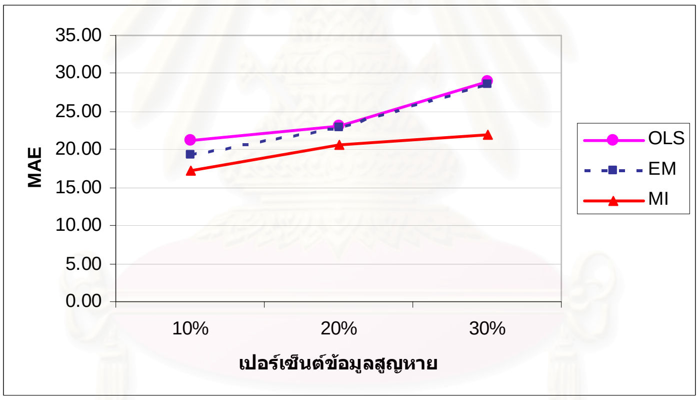
รูปที่ 4.1.39 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และค่าสัมประสิทธิ์ความผันแปร 0.45



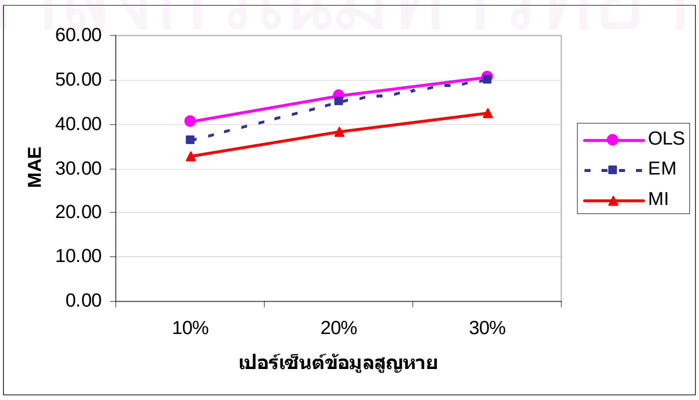
รูปที่ 4.1.40 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.05



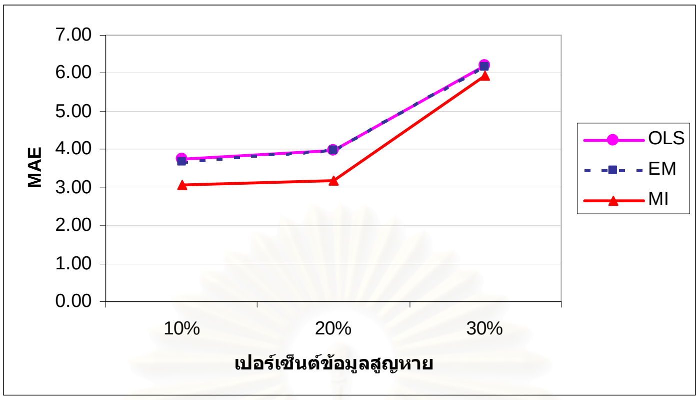
รูปที่ 4.1.41 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.25



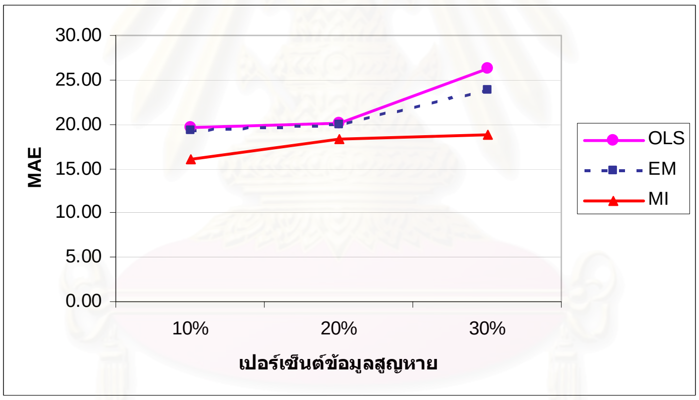
รูปที่ 4.1.42 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และค่าสัมประสิทธิ์ความผันแปร 0.45



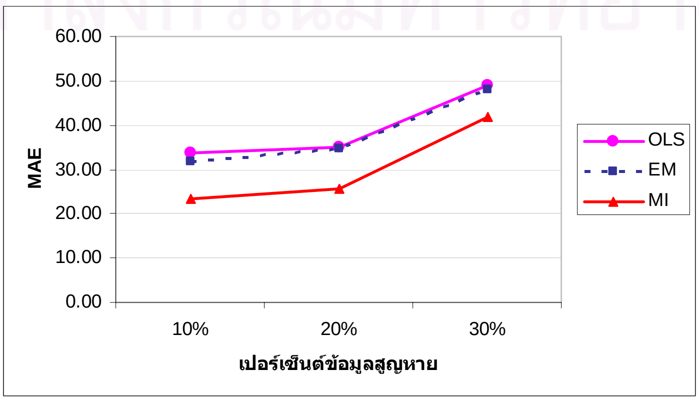
รูปที่ 4.1.43 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.1.44 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.25



รูปที่ 4.1.45 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและเปอร์เซ็นต์ข้อมูลสูญหายในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และค่าสัมประสิทธิ์ความผันแปร 0.45



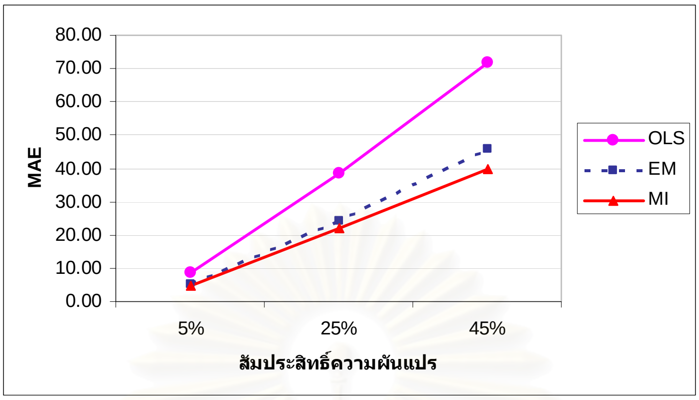
จากตารางที่ 4.1.1 – 4.1.5 และรูปที่ 4.1.1 – 4.1.45 แสดงการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี ในแผนแบบการทดลองจัตุรัสละติน ณ เปอร์เซ็นต์ข้อมูลสูญหาย เมื่อกำหนดให้ค่าคงที่ h และสัมประสิทธิ์ความผันแปรคงที่ พบว่าเมื่อเปอร์เซ็นต์ข้อมูลสูญหายมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าเพิ่มขึ้น เนื่องจากเมื่อข้อมูลมีการสูญหายเพิ่มมากขึ้นจะทำให้การประมาณค่าสูญหายมีความผิดพลาดมากขึ้น และจากการวิจัยพบว่าวิธี Multiple Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุดทุกเปอร์เซ็นต์ข้อมูลสูญหาย ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

4.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ณ สัมประสิทธิ์ความผันแปร เมื่อกำหนดให้ค่าคงที่ h และเปอร์เซ็นต์ข้อมูลสูญหายคงที่

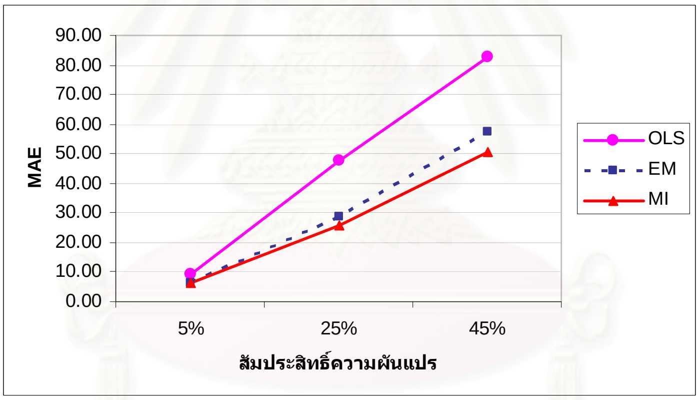
ตารางที่ 4.2.1 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน โดยพิจารณาจากสัมประสิทธิ์ความผันแปร

h	missing	c.v	MAE OLS	MAE EM	MAE MI
1	10%	0.05	8.6599	5.3896	4.7439
		0.25	38.3618	24.0602	22.0340
		0.45	71.7228	45.9674	39.8221
	20%	0.05	9.0970	6.2450	6.1115
		0.25	47.6126	28.4760	25.9879
		0.45	82.5229	57.2253	50.4489
	30%	0.05	9.3009	7.3435	6.9981
		0.25	51.5757	34.0191	29.9957
		0.45	92.5253	73.2552	69.0453
2	10%	0.05	6.2909	6.1418	5.4974
		0.25	29.5559	28.1828	24.5889
		0.45	60.7241	49.3142	41.9943
	20%	0.05	8.5862	6.2905	6.0110
		0.25	38.0925	30.7788	26.6395
		0.45	72.2710	50.1778	46.9934
	30%	0.05	9.2088	7.9365	6.5043
		0.25	42.2902	32.5284	30.9707
		0.45	81.5236	73.3290	62.6600
3	10%	0.05	5.6183	4.8229	4.0854
		0.25	24.9180	24.7828	21.6091
		0.45	49.1727	46.6222	41.7862
	20%	0.05	5.6828	5.6306	5.4458
		0.25	30.0515	27.4357	24.0875
		0.45	59.1564	52.1900	49.2983
	30%	0.05	6.7541	6.4981	5.8000
		0.25	35.2088	28.9352	24.4002
		0.45	65.9787	58.8536	54.5361

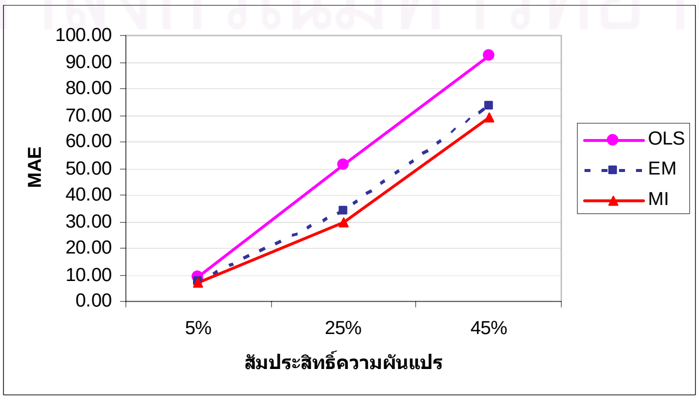
รูปที่ 4.2.1 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



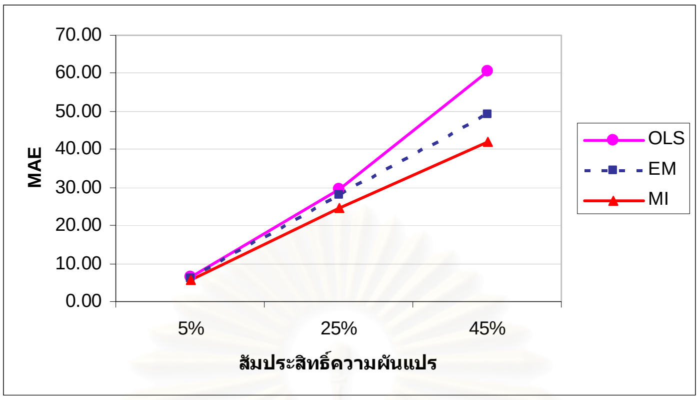
รูปที่ 4.2.2 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



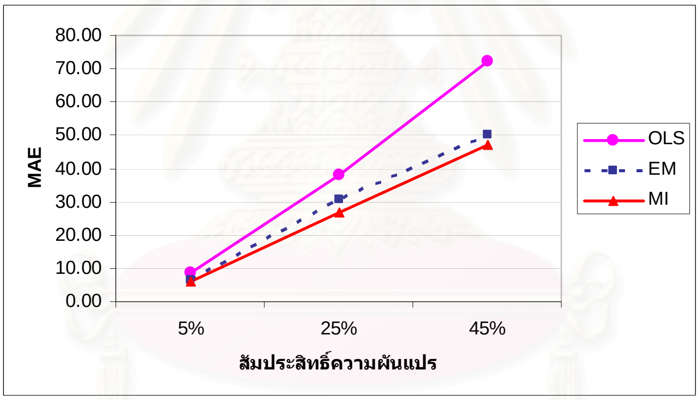
รูปที่ 4.2.3 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



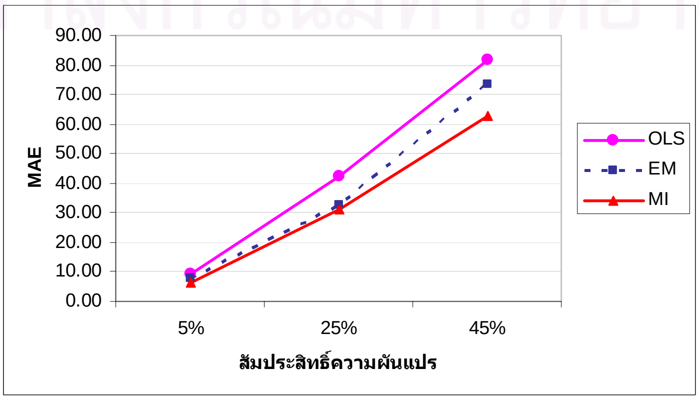
รูปที่ 4.2.4 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



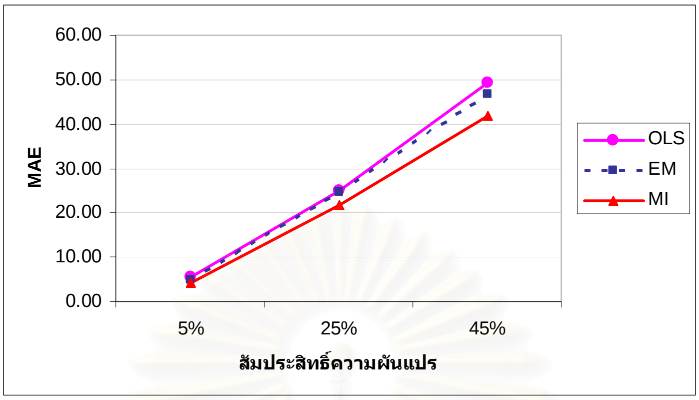
รูปที่ 4.2.5 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



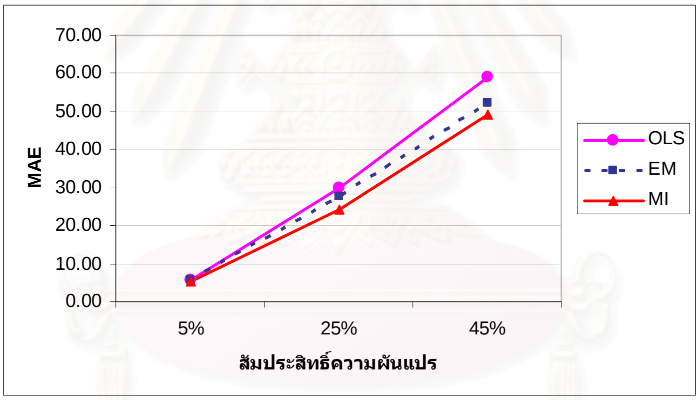
รูปที่ 4.2.6 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



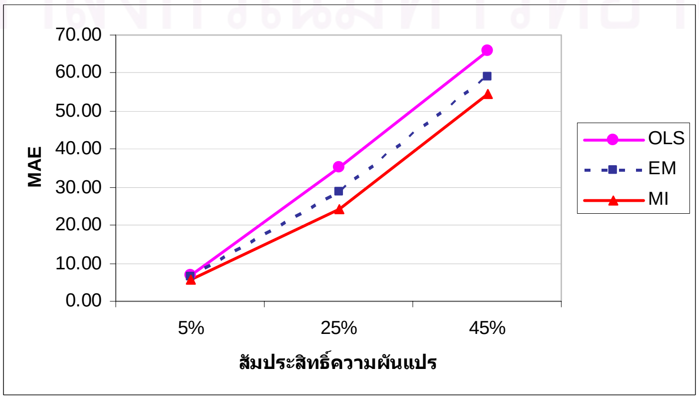
รูปที่ 4.2.7 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



รูปที่ 4.2.8 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



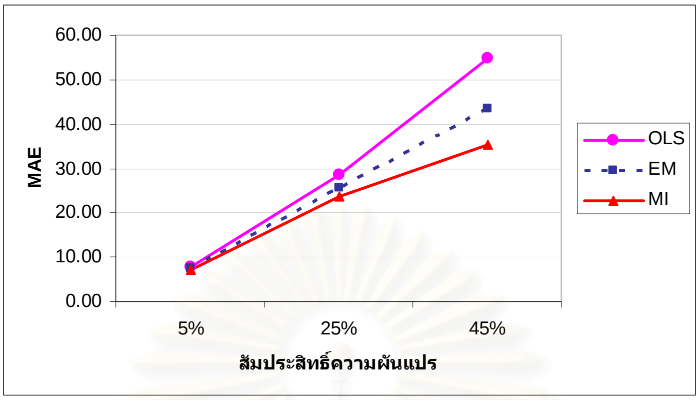
รูปที่ 4.2.9 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



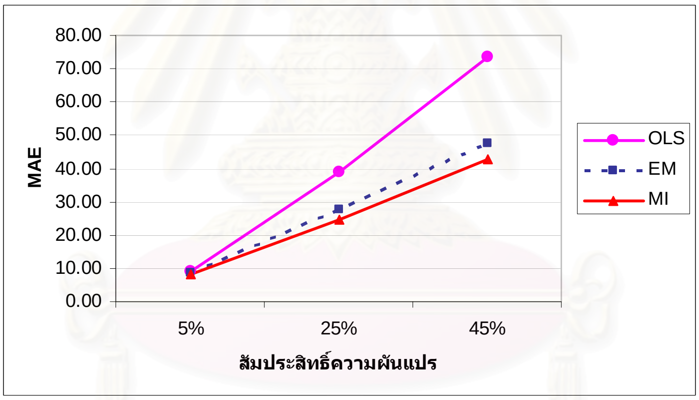
ตารางที่ 4.2.2 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน โดยพิจารณาจากสัมประสิทธิ์ความผันแปร

h	missing	c.v	MAE OLS	MAE EM	MAE MI
1	10%	0.05	7.7791	7.5409	7.2314
		0.25	28.6270	25.5202	23.7884
		0.45	54.8865	43.4449	35.2345
	20%	0.05	8.8793	8.4767	8.2876
		0.25	39.1272	27.4871	24.6940
		0.45	73.4797	47.4956	42.6461
	30%	0.05	10.2139	9.0966	8.6274
		0.25	46.1516	31.4099	27.8118
		0.45	80.9840	48.5235	42.6717
2	10%	0.05	6.2162	6.0463	5.8739
		0.25	23.7620	22.4726	21.5473
		0.45	45.8605	38.1680	36.7966
	20%	0.05	6.6138	6.1800	6.1218
		0.25	29.0753	25.8020	22.1026
		0.45	52.2904	40.0499	37.7904
	30%	0.05	9.7351	8.6344	7.9183
		0.25	41.5674	25.9717	22.9555
		0.45	59.6147	47.6554	39.6471
3	10%	0.05	4.6076	4.5990	4.5138
		0.25	24.2024	20.3031	19.5846
		0.45	32.5013	31.7206	30.6984
	20%	0.05	5.4724	5.4314	5.3789
		0.25	28.5308	23.9133	21.9964
		0.45	47.2045	37.5362	34.3447
	30%	0.05	7.7195	7.0991	6.5956
		0.25	30.2066	24.1183	22.1260
		0.45	53.9431	43.4060	35.4169

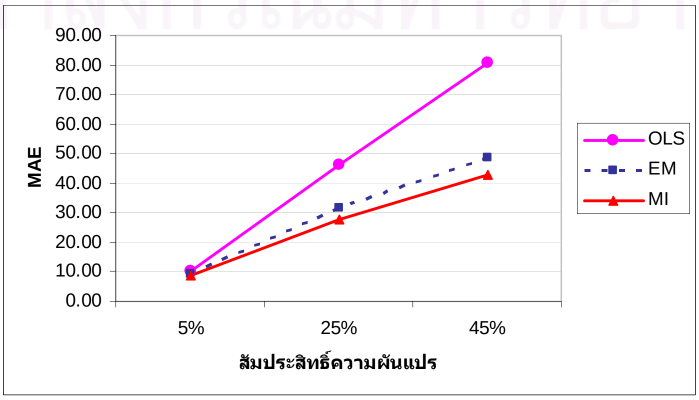
รูปที่ 4.2.10 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



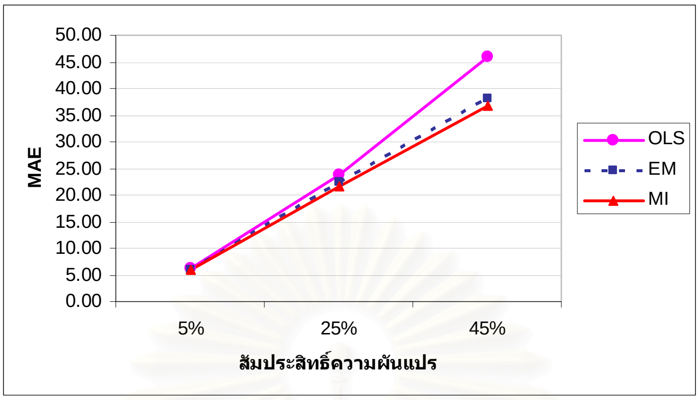
รูปที่ 4.2.11 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



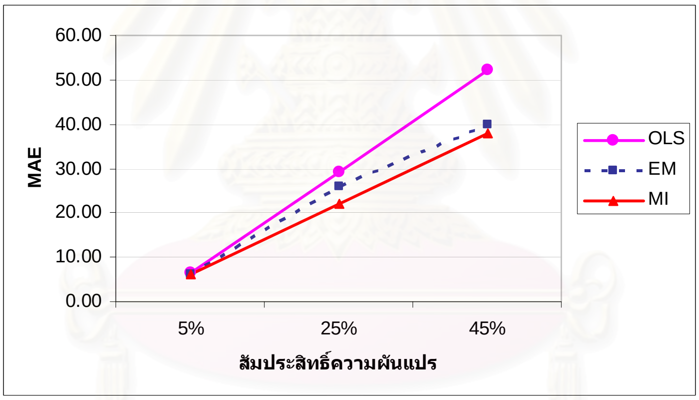
รูปที่ 4.2.12 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



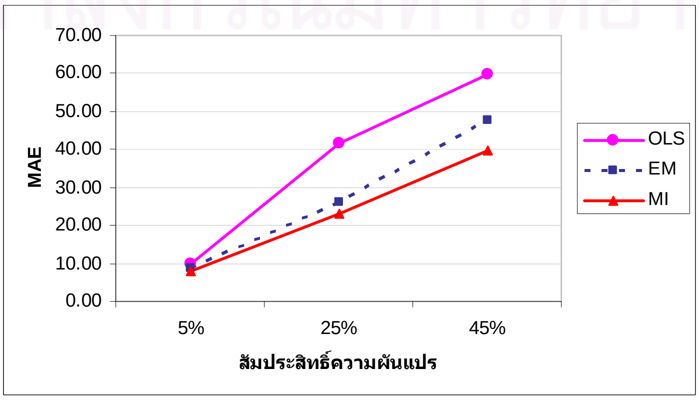
รูปที่ 4.2.13 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



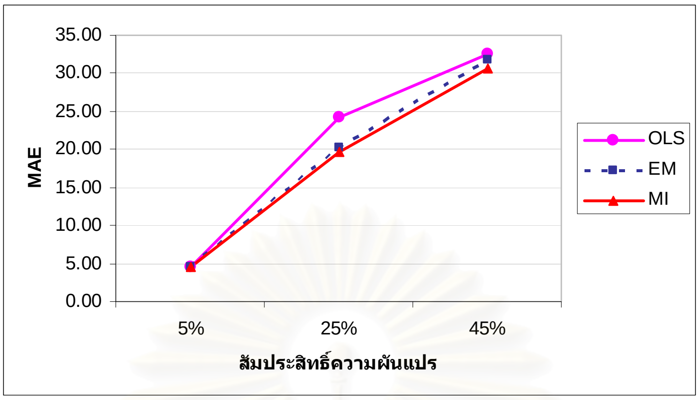
รูปที่ 4.2.14 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



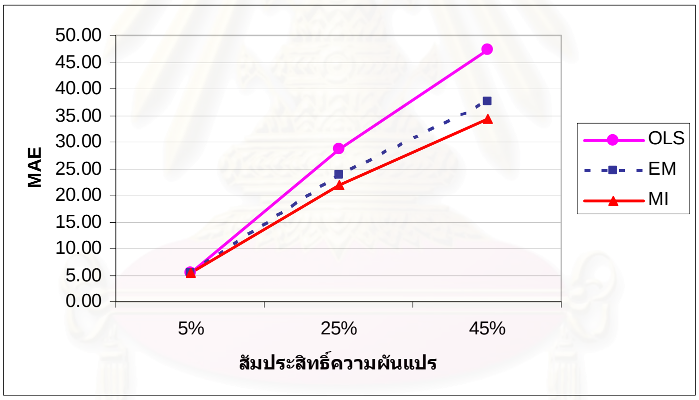
รูปที่ 4.2.15 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



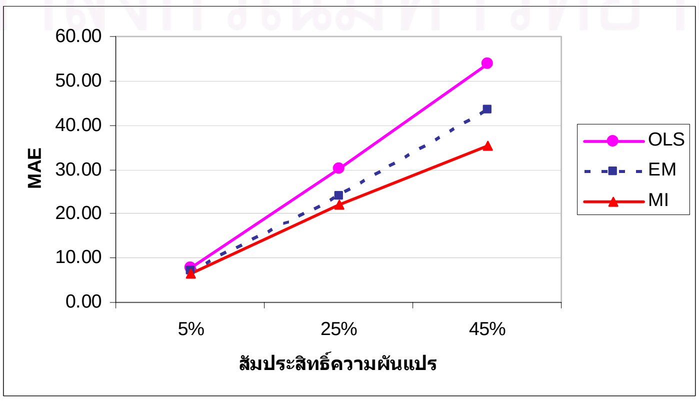
รูปที่ 4.2.16 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



รูปที่ 4.2.17 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



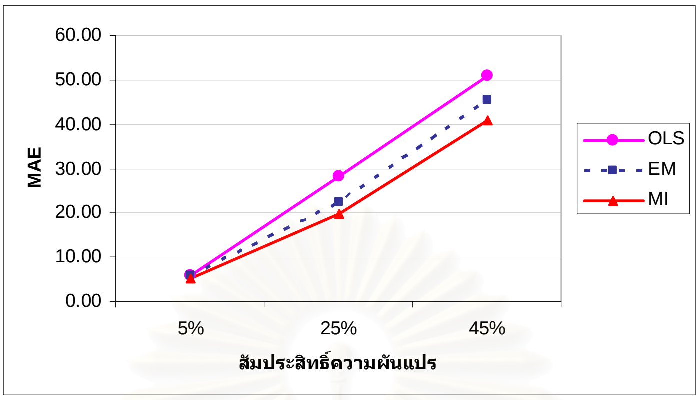
รูปที่ 4.2.18 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



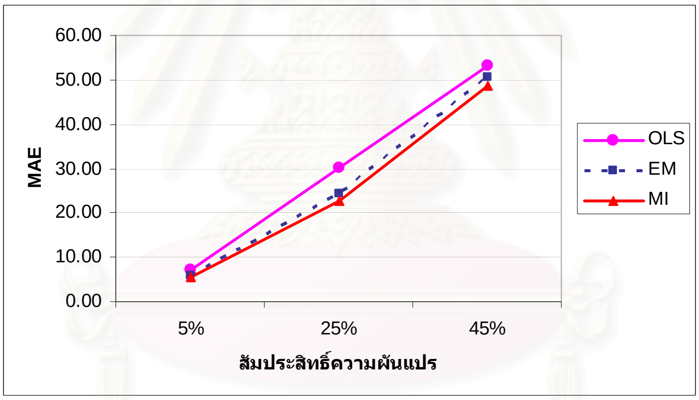
ตารางที่ 4.2.3 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน โดยพิจารณาจากสัมประสิทธิ์ความผันแปร

h	missing	c.v	MAE OLS	MAE EM	MAE MI
1	10%	0.05	5.8682	5.6968	5.0957
		0.25	28.1094	22.3784	19.7854
		0.45	50.9904	45.5139	40.9688
	20%	0.05	7.1298	5.8971	5.6700
		0.25	30.0478	24.3331	22.5674
		0.45	53.2704	50.5609	48.5021
	30%	0.05	10.3068	7.9281	7.3466
		0.25	35.7777	32.9384	30.0074
		0.45	61.3002	54.4515	49.9025
2	10%	0.05	4.8969	4.7649	4.1803
		0.25	21.3309	19.3479	17.3448
		0.45	36.4578	31.9861	28.9385
	20%	0.05	5.3906	5.0056	4.2326
		0.25	22.0811	20.7711	20.6736
		0.45	41.7682	32.1911	29.1245
	30%	0.05	6.8966	6.4575	5.9949
		0.25	30.9132	29.6125	26.0193
		0.45	50.1699	46.2479	39.7025
3	10%	0.05	4.0798	4.0432	3.7023
		0.25	16.3850	15.7351	13.4896
		0.45	31.7296	31.7072	27.6233
	20%	0.05	5.6796	5.3929	4.0337
		0.25	20.5194	19.1290	16.9755
		0.45	37.5360	35.7913	32.6195
	30%	0.05	6.5185	6.3165	5.9189
		0.25	24.2990	24.1145	22.4991
		0.45	42.9053	39.9780	37.0511

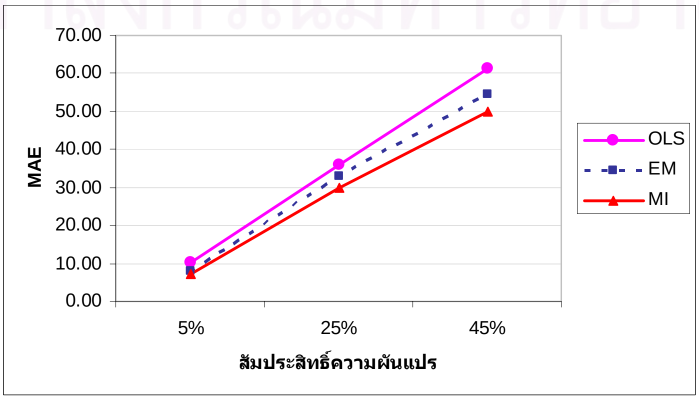
รูปที่ 4.2.19 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



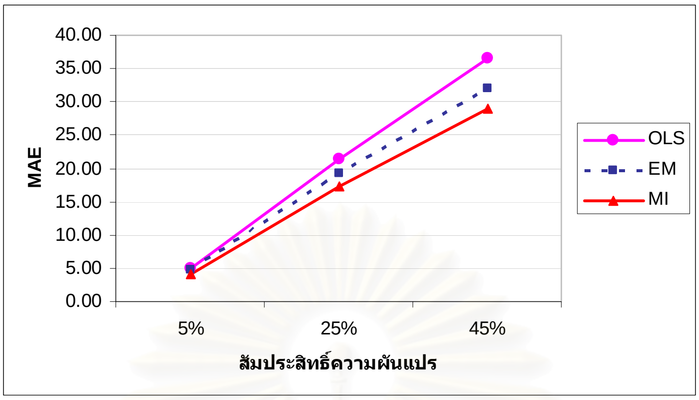
รูปที่ 4.2.20 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



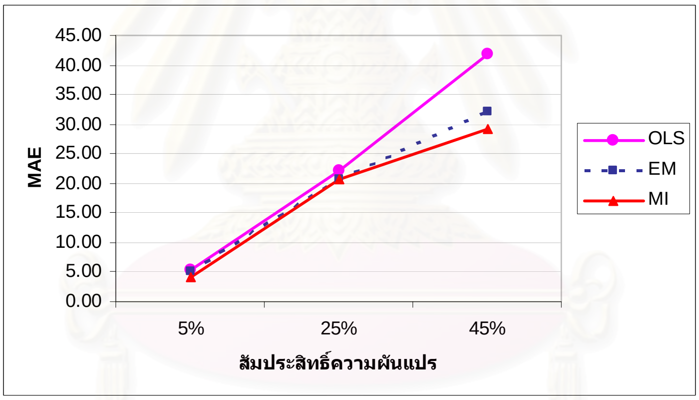
รูปที่ 4.2.21 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



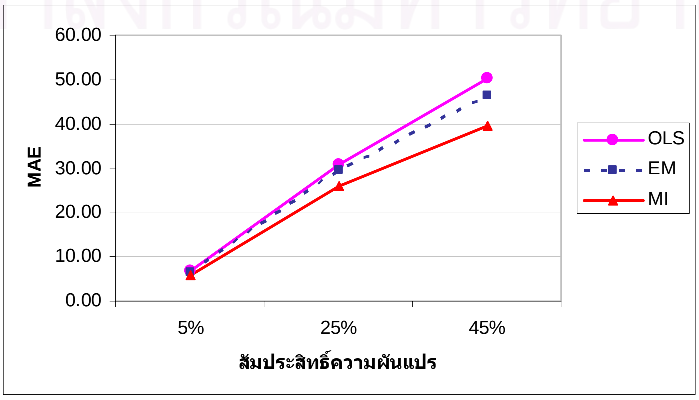
รูปที่ 4.2.22 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



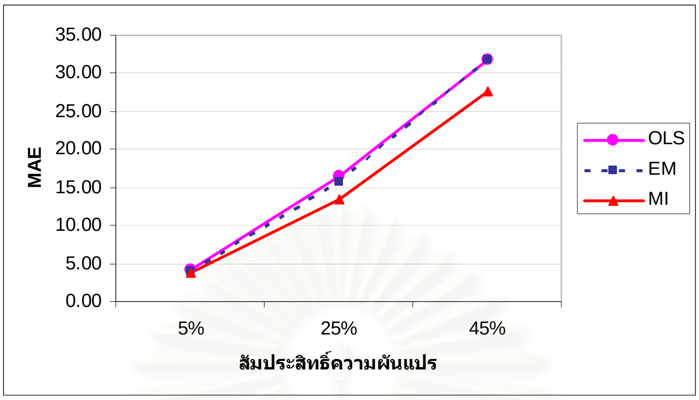
รูปที่ 4.2.23 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



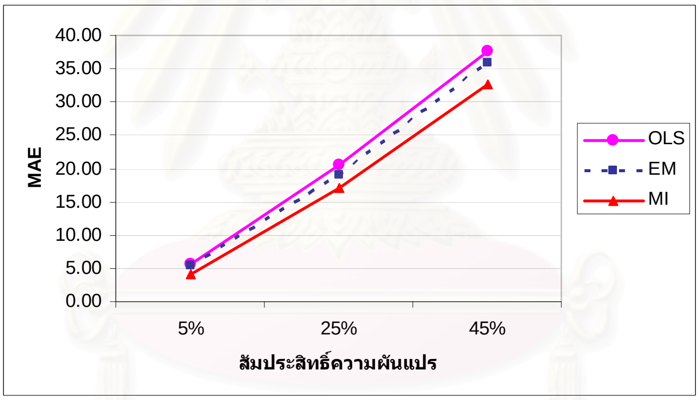
รูปที่ 4.2.24 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



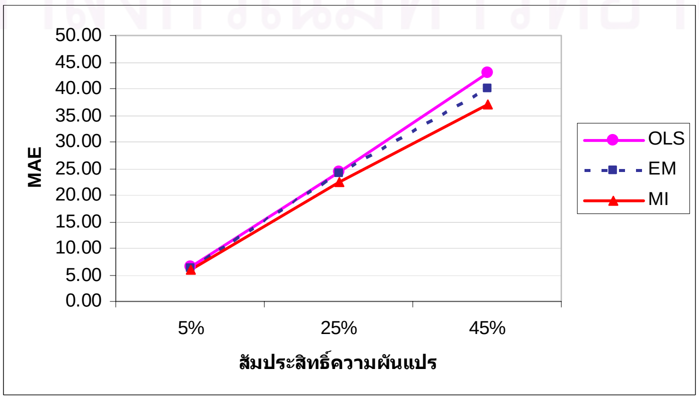
รูปที่ 4.2.25 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



รูปที่ 4.2.26 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



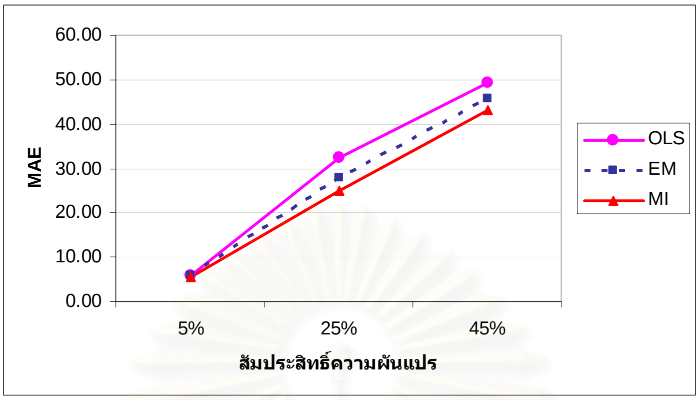
รูปที่ 4.2.27 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



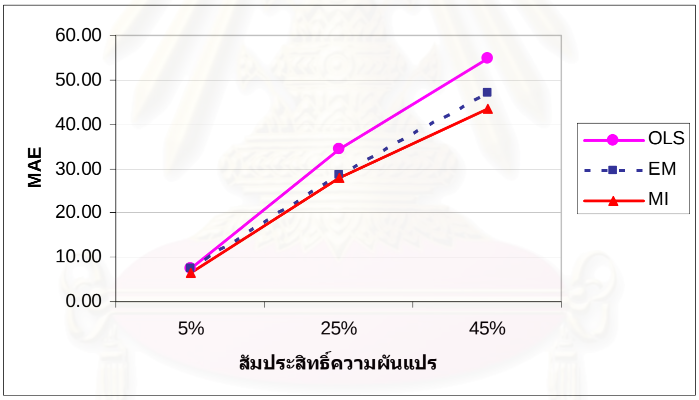
ตารางที่ 4.2.4 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน โดยพิจารณาจากสัมประสิทธิ์ความผันแปร

h	missing	c.v	MAE OLS	MAE EM	MAE MI
1	10%	0.05	5.8172	5.6917	5.5372
		0.25	32.2936	27.7421	24.9148
		0.45	49.3574	45.6812	42.9947
	20%	0.05	7.5641	7.4550	6.5478
		0.25	34.5393	28.4741	27.7647
		0.45	54.7191	46.9111	43.5965
	30%	0.05	8.0636	7.9331	6.5879
		0.25	42.7861	38.1336	32.1790
		0.45	79.0797	62.2092	53.9579
2	10%	0.05	4.7243	4.6947	3.9671
		0.25	19.7685	19.3838	18.6600
		0.45	41.5621	39.5780	36.9268
	20%	0.05	6.0893	6.1567	5.4906
		0.25	25.1802	22.9325	20.5171
		0.45	47.9628	44.9490	40.6498
	30%	0.05	6.2787	6.2280	5.8936
		0.25	25.3932	24.9744	20.8245
		0.45	51.7270	49.8744	40.8616
3	10%	0.05	5.0199	5.0004	4.4029
		0.25	21.1072	21.1000	19.2046
		0.45	36.4664	33.5128	32.4316
	20%	0.05	5.9174	5.9066	5.6390
		0.25	24.8663	21.9596	19.8567
		0.45	45.1973	39.7520	36.4955
	30%	0.05	7.0444	6.8368	6.1050
		0.25	25.1659	25.0756	21.4612
		0.45	45.9428	44.8748	38.5298

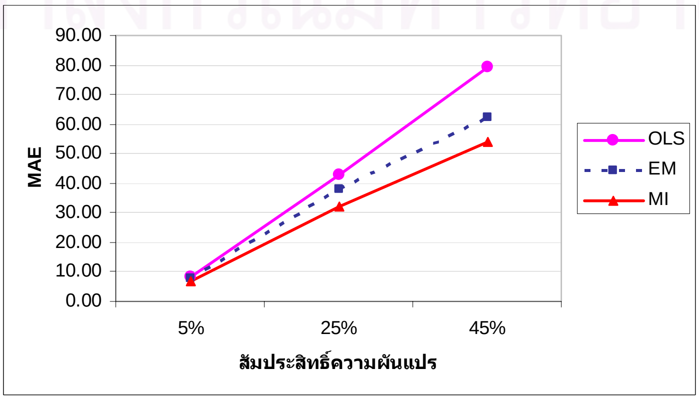
รูปที่ 4.2.28 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



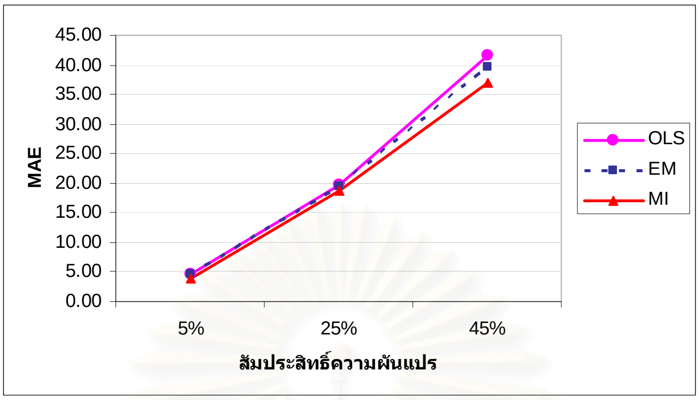
รูปที่ 4.2.29 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



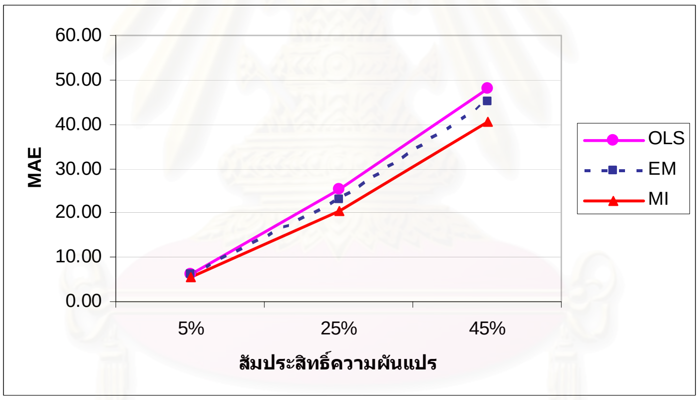
รูปที่ 4.2.30 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



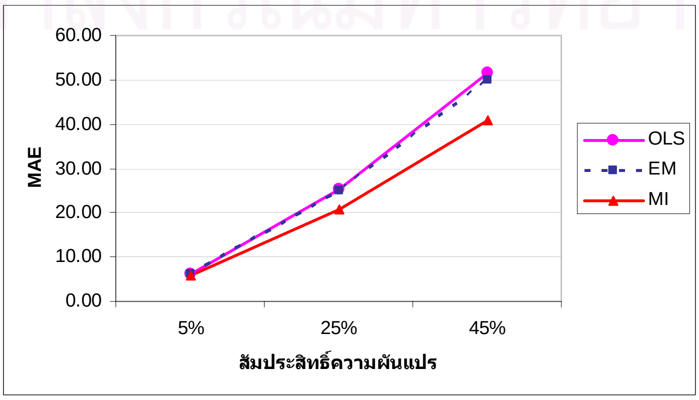
รูปที่ 4.2.31 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



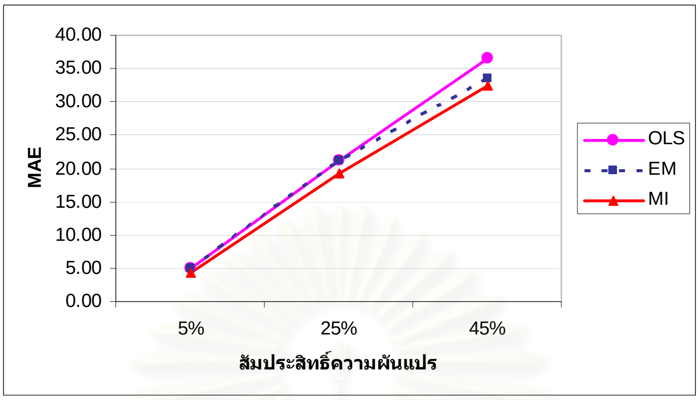
รูปที่ 4.2.32 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



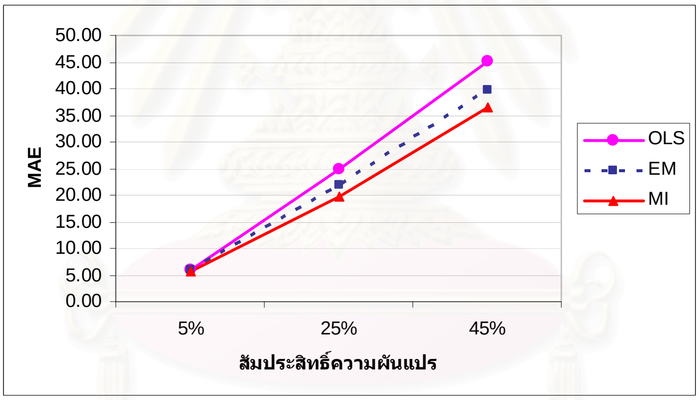
รูปที่ 4.2.33 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



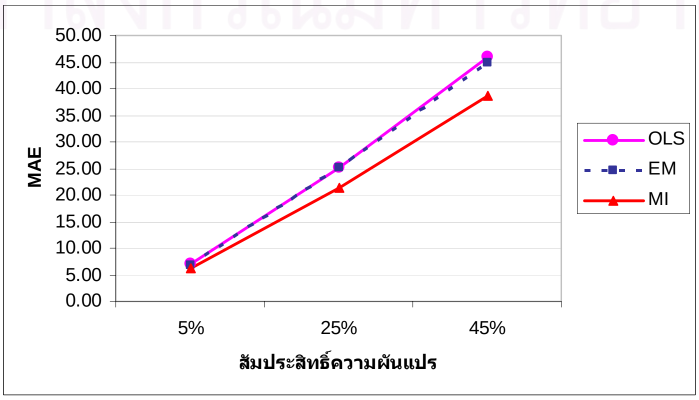
รูปที่ 4.2.34 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



รูปที่ 4.2.35 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



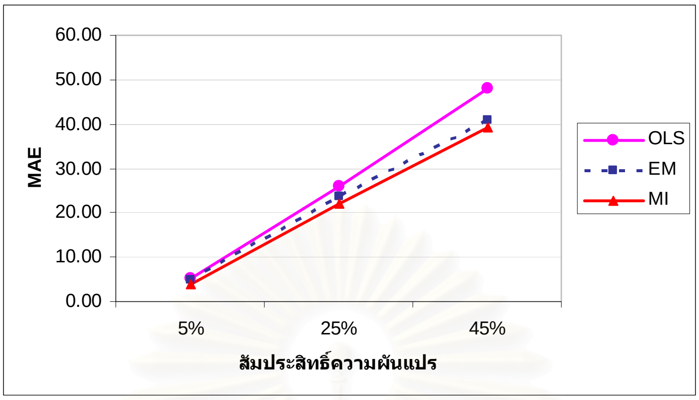
รูปที่ 4.2.36 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



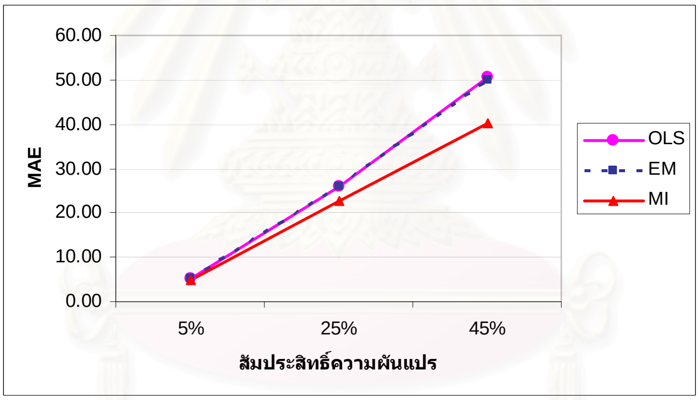
ตารางที่ 4.2.5 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน โดยพิจารณาจากสัมประสิทธิ์ความผันแปร

h	missing	c.v	MAE OLS	MAE EM	MAE MI
1	10%	0.05	5.1068	5.0130	4.0236
		0.25	25.8864	23.7677	22.1253
		0.45	48.0682	41.0144	39.2961
	20%	0.05	5.2391	5.1356	4.8602
		0.25	25.9979	25.8966	22.6826
		0.45	50.5838	49.7984	40.2521
	30%	0.05	7.1953	7.0362	6.2244
		0.25	34.7477	27.9877	23.8433
		0.45	63.3971	51.8287	44.3837
2	10%	0.05	4.4951	4.4726	3.9210
		0.25	21.2013	19.2644	17.2893
		0.45	40.5607	36.2060	32.8648
	20%	0.05	4.7138	4.6997	3.9301
		0.25	23.0790	22.8575	20.6931
		0.45	46.2886	45.1359	38.3525
	30%	0.05	6.5039	6.5012	6.1110
		0.25	29.0086	28.6546	21.9417
		0.45	50.6945	49.9759	42.5841
3	10%	0.05	3.7407	3.6758	3.0719
		0.25	19.5979	19.2842	16.0675
		0.45	33.7431	31.6420	23.3490
	20%	0.05	3.9915	3.9866	3.1730
		0.25	20.0283	19.9914	18.4019
		0.45	35.0983	34.6465	25.7382
	30%	0.05	6.2119	6.1702	5.9378
		0.25	26.3185	23.8886	18.8652
		0.45	48.8774	47.9632	41.6794

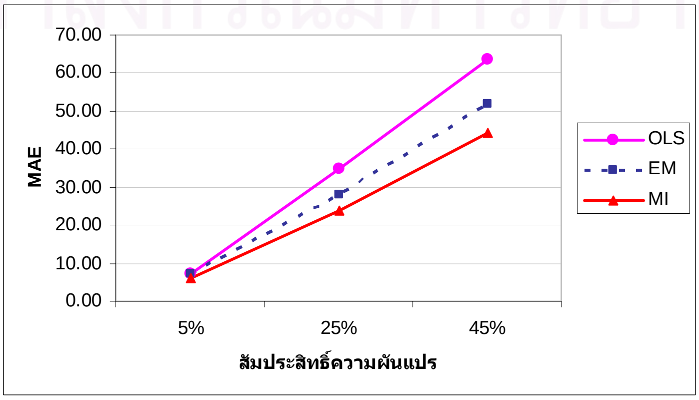
รูปที่ 4.2.37 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



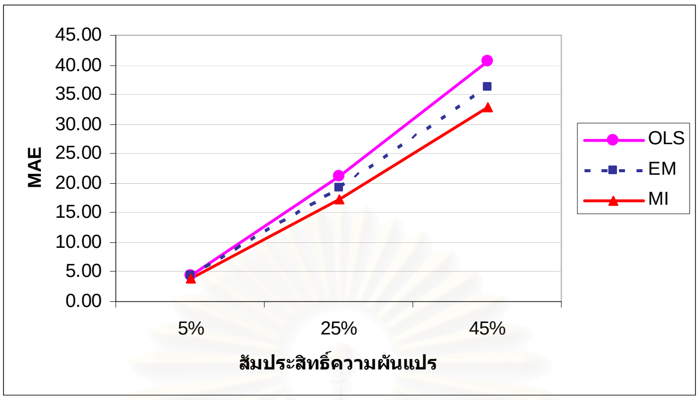
รูปที่ 4.2.38 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



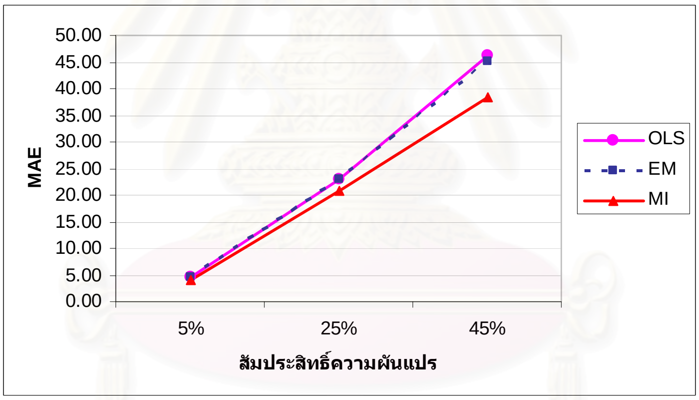
รูปที่ 4.2.39 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 1 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



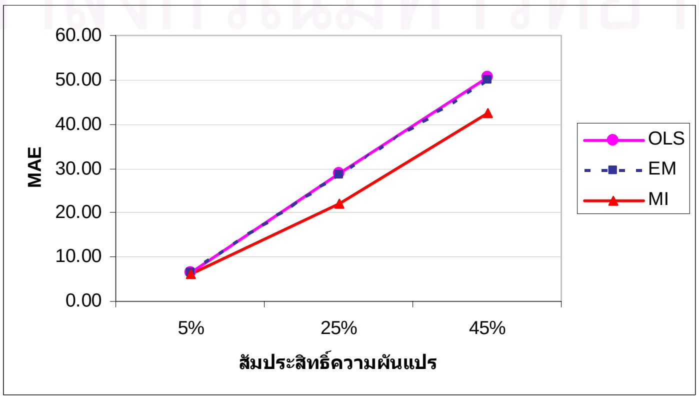
รูปที่ 4.2.40 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



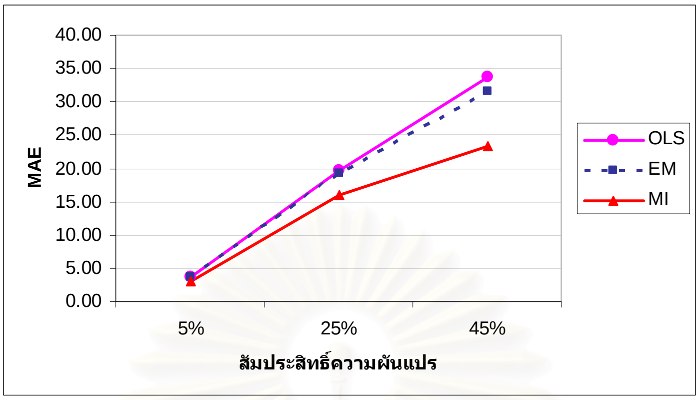
รูปที่ 4.2.41 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



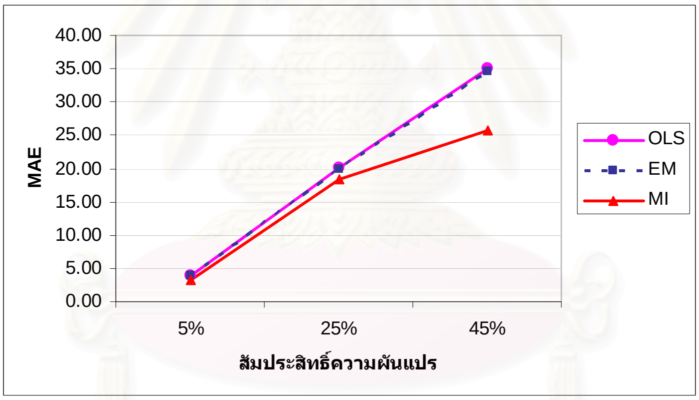
รูปที่ 4.2.42 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 2 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



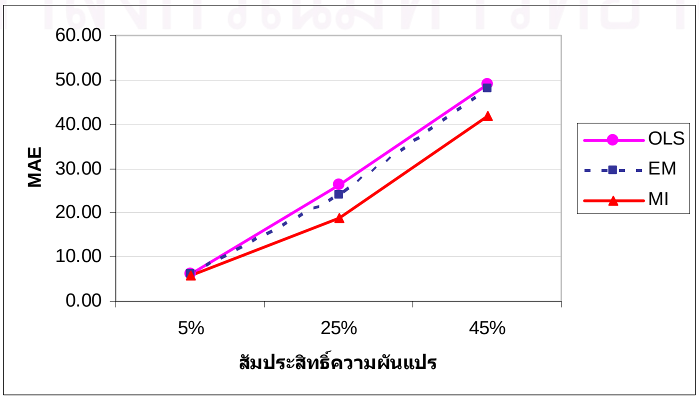
รูปที่ 4.2.43 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10%



รูปที่ 4.2.44 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20%



รูปที่ 4.2.45 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าสัมประสิทธิ์ความผันแปรในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีค่าคงที่ h = 3 และเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30%



จากตารางที่ 4.2.1 – 4.2.5 และรูปที่ 4.2.1 – 4.2.45 แสดงการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี ในแผนแบบการทดลองจัตุรัสละติน ณ สัมประสิทธิ์ความผันแปร เมื่อกำหนดให้ค่าคงที่ h และเปอร์เซ็นต์ข้อมูลสูญหายคงที่ พบว่าเมื่อค่าสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าเพิ่มขึ้น เนื่องจากเมื่อสัมประสิทธิ์ความผันแปรของข้อมูลเพิ่มมากขึ้น จะทำให้ข้อมูลมีการกระจายมากขึ้น การประมาณค่าสูญหายจึงมีความผิดพลาดมากขึ้น และจากการวิจัยพบว่าวิธี Multiple Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุด ทุกสัมประสิทธิ์ความผันแปร ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

เมื่อสัมประสิทธิ์ความผันแปรมีค่าเพิ่มขึ้น จะพบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธี Multiple Imputation ยิ่งห่างจากวิธี EM Algorithm และวิธีกำลังสองน้อยสุดมากขึ้น นั่นคือในกรณีที่ชุดข้อมูลตัวอย่างที่สุ่มมามีความผันแปรมาก ควรจะประมาณค่าสังเกตที่สูญหายโดยใช้วิธี Multiple Imputation ถึงแม้ว่าจะมีวิธีการที่ซับซ้อนมากกว่า แต่ค่าประมาณที่ได้นั้นมีค่าใกล้เคียงกับค่าจริงมากกว่า

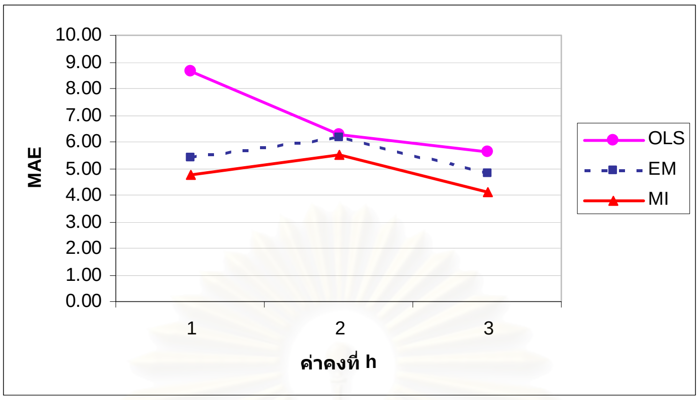
และสำหรับกรณีที่ชุดข้อมูลตัวอย่างที่สุ่มมามีความผันแปรน้อย พบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นควรเลือกใช้การประมาณค่าสังเกตที่สูญหายโดยวิธีกำลังสองน้อยสุดมากกว่า เพราะสะดวกและรวดเร็วกว่า

4.3 ผลการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ณ ค่าคงที่ h เมื่อกำหนดให้เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรคงที่

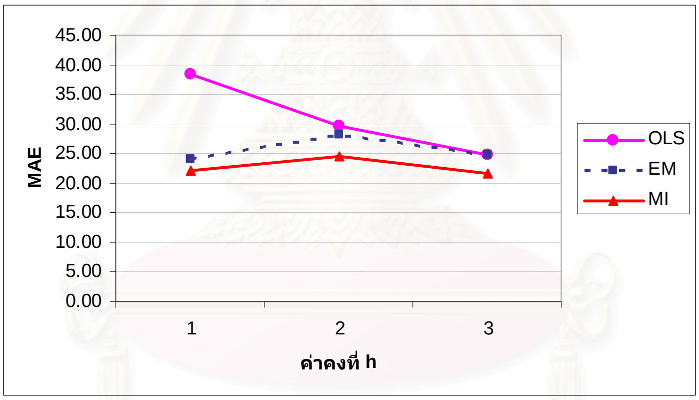
ตารางที่ 4.3.1 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน โดยพิจารณาจากค่าคงที่ h

missing	c.v	h	MAE OLS	MAE EM	MAE MI
10%	0.05	1	8.6599	5.3896	4.7439
		2	6.2909	6.1418	5.4974
		3	5.6183	4.8229	4.0854
	0.25	1	38.3618	24.0602	22.0340
		2	29.5559	28.1828	24.5889
		3	24.9180	24.7828	21.6091
	0.45	1	71.7228	45.9674	39.8221
		2	60.7241	49.3142	41.9943
		3	49.1727	46.6222	41.7862
20%	0.05	1	9.0970	6.2450	6.1115
		2	8.5862	6.2905	6.0110
		3	5.6828	5.6306	5.4458
	0.25	1	47.6126	28.4760	25.9879
		2	38.0925	30.7788	26.6395
		3	30.0515	27.4357	24.0875
	0.45	1	82.5229	57.2253	50.4489
		2	72.2710	50.1778	46.9934
		3	59.1564	52.1900	49.2983
30%	0.05	1	9.3009	7.3435	6.9981
		2	9.2088	7.9365	6.5043
		3	6.7541	6.4981	5.8000
	0.25	1	51.5757	34.0191	29.9957
		2	42.2902	32.5284	30.9707
		3	35.2088	28.9352	24.4002
	0.45	1	92.5253	73.2552	69.0453
		2	81.5236	73.3290	62.6600
		3	65.9787	58.8536	54.5361

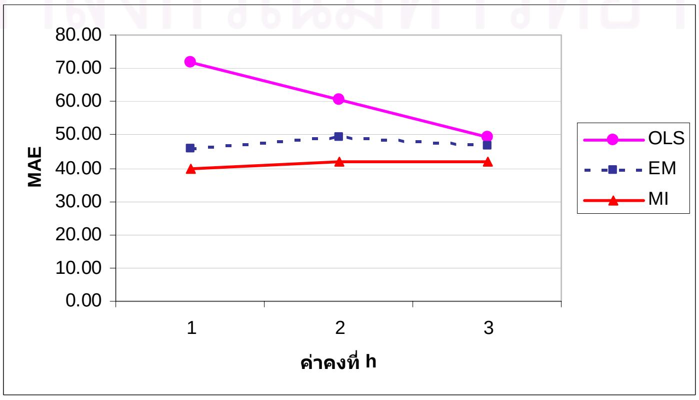
รูปที่ 4.3.1 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.05



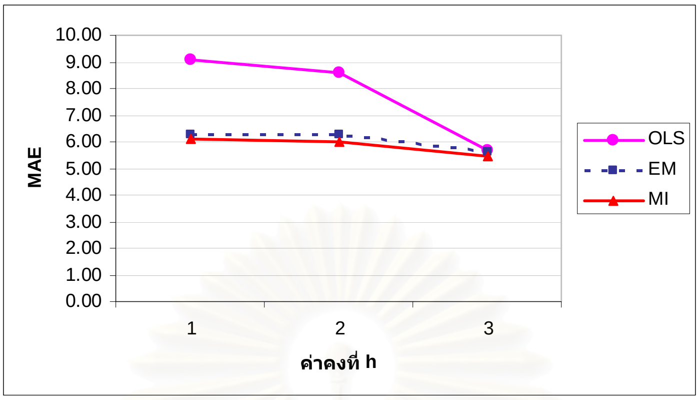
รูปที่ 4.3.2 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.25



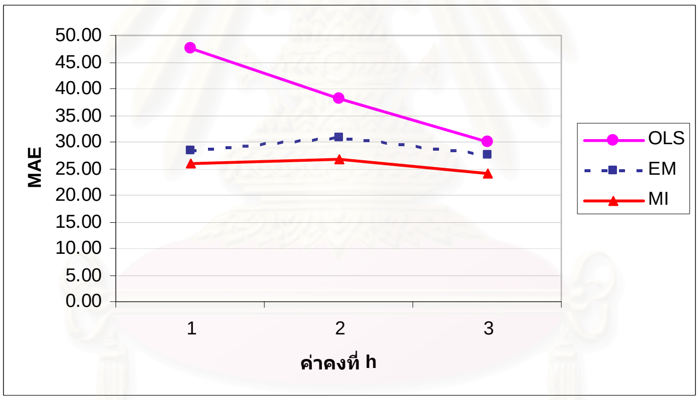
รูปที่ 4.3.3 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.45



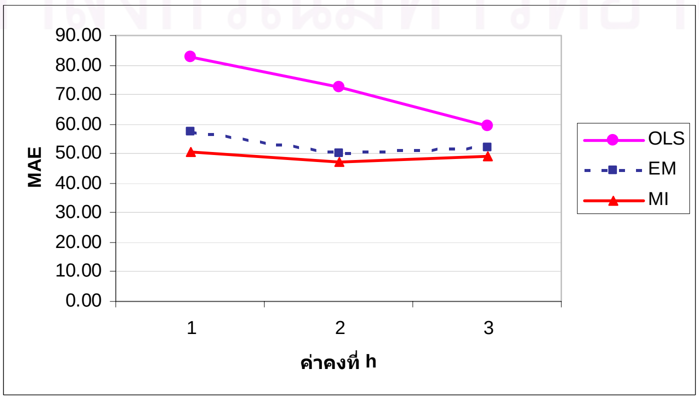
รูปที่ 4.3.4 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.05



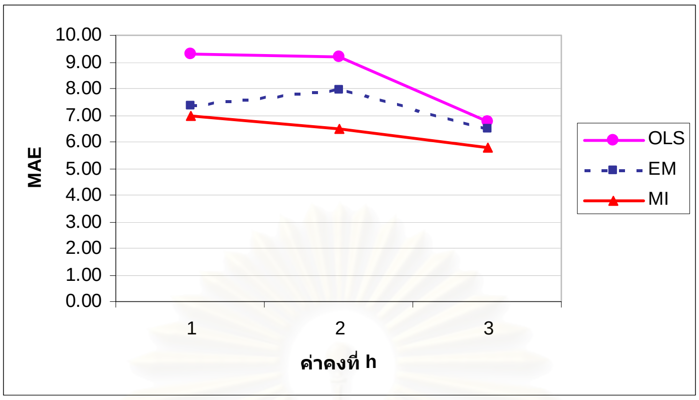
รูปที่ 4.3.5 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.25



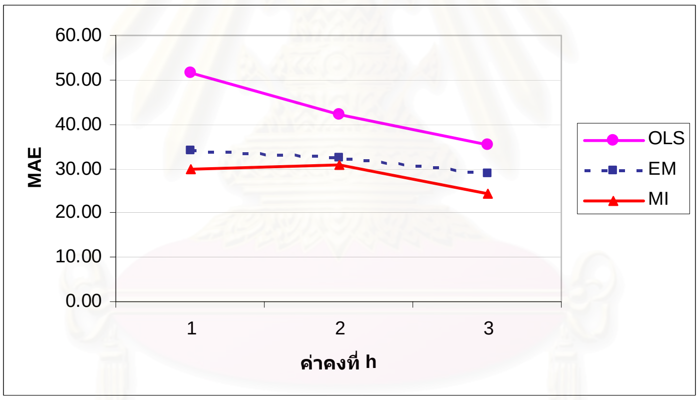
รูปที่ 4.3.6 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.45



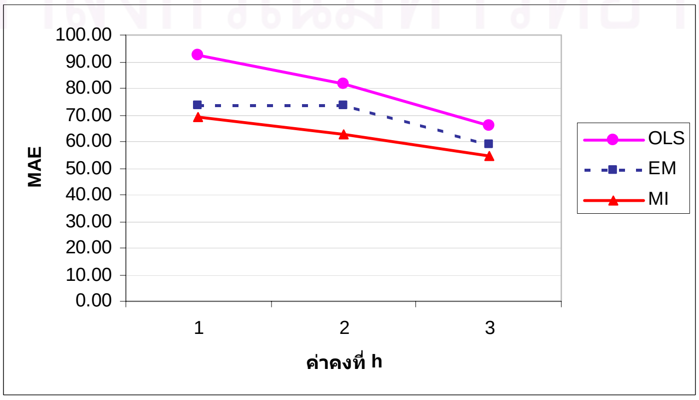
รูปที่ 4.3.7 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.3.8 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.25



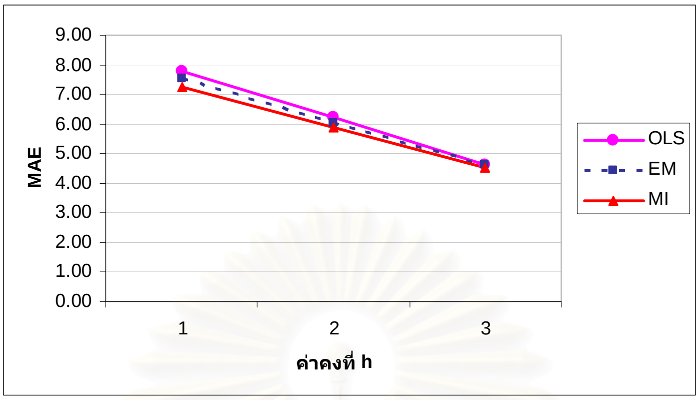
รูปที่ 4.3.9 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 3x3 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.45



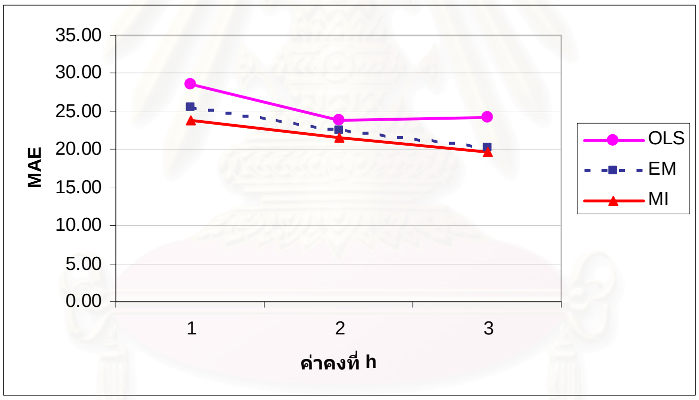
ตารางที่ 4.3.2 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน โดยพิจารณาจากค่าคงที่ h

missing	c.v	h	MAE OLS	MAE EM	MAE MI
		1	7.7791	7.5409	7.2314
	0.05	2	6.2162	6.0463	5.8739
		3	4.6076	4.5990	4.5138
		1	28.6270	25.5202	23.7884
10%	0.25	2	23.7620	22.4726	21.5473
		3	24.2024	20.3031	19.5846
		1	54.8865	43.4449	35.2345
	0.45	2	45.8605	38.1680	36.7966
		3	32.5013	31.7206	30.6984
		1	8.8793	8.4767	8.2876
	0.05	2	6.6138	6.1800	6.1218
		3	5.4724	5.4314	5.3789
		1	39.1272	27.4871	24.6940
20%	0.25	2	29.0753	25.8020	22.1026
		3	28.5308	23.9133	21.9964
		1	73.4797	47.4956	42.6461
	0.45	2	52.2904	40.0499	37.7904
		3	47.2045	37.5362	34.3447
		1	10.2139	9.0966	8.6274
	0.05	2	9.7351	8.6344	7.9183
		3	7.7195	7.0991	6.5956
		1	46.1516	31.4099	27.8118
30%	0.25	2	41.5674	25.9717	22.9555
		3	30.2066	24.1183	22.1260
		1	80.9840	48.5235	42.6717
	0.45	2	59.6147	47.6554	39.6471
		3	53.9431	43.4060	35.4169

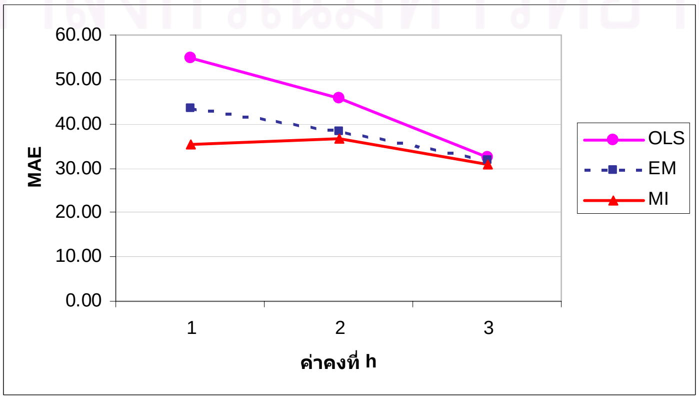
รูปที่ 4.3.10 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.05



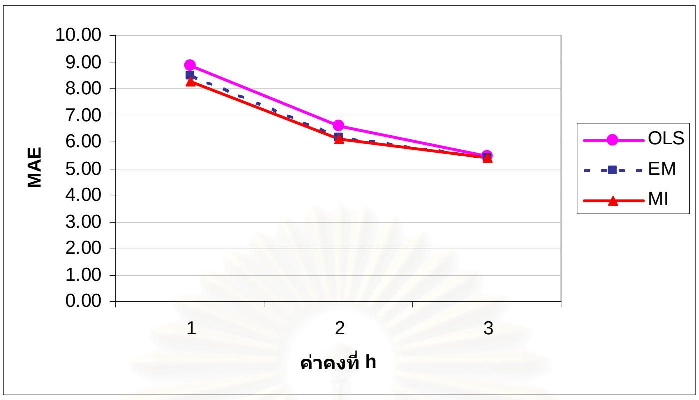
รูปที่ 4.3.11 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.25



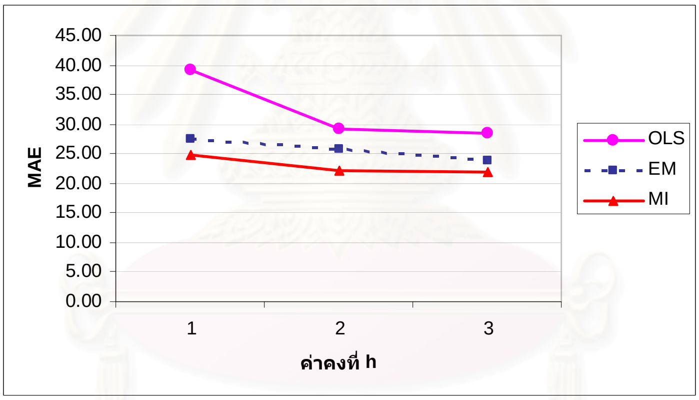
รูปที่ 4.3.12 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.45



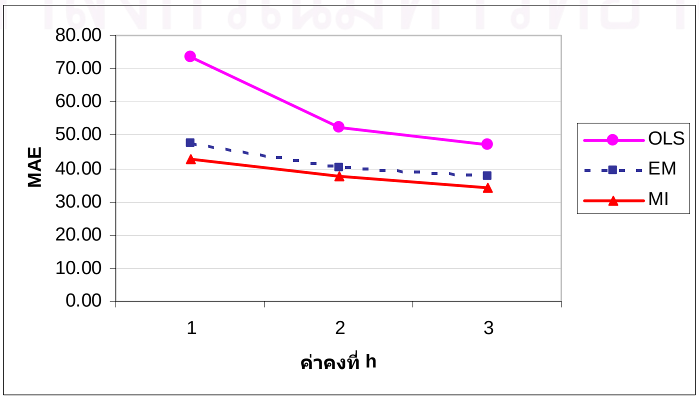
รูปที่ 4.3.13 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.05



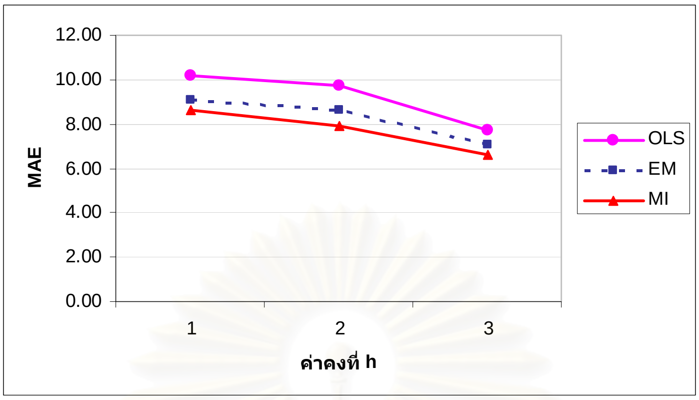
รูปที่ 4.3.14 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.25



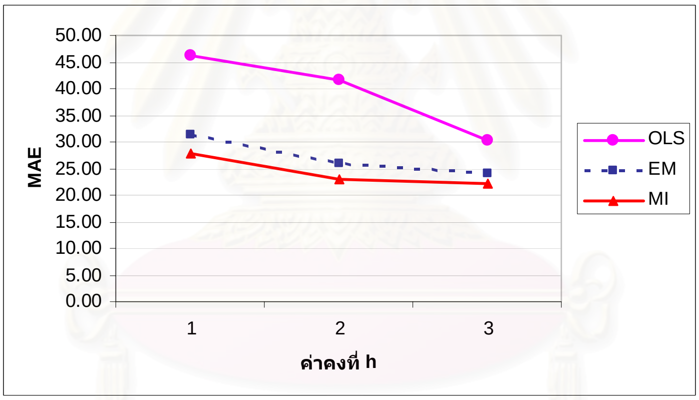
รูปที่ 4.3.15 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.45



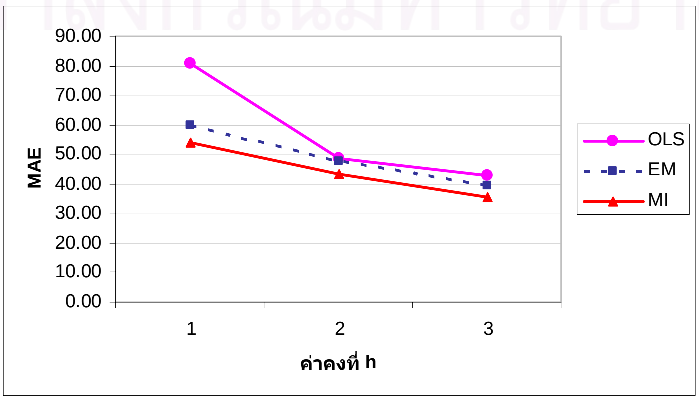
รูปที่ 4.3.16 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.3.17 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.25



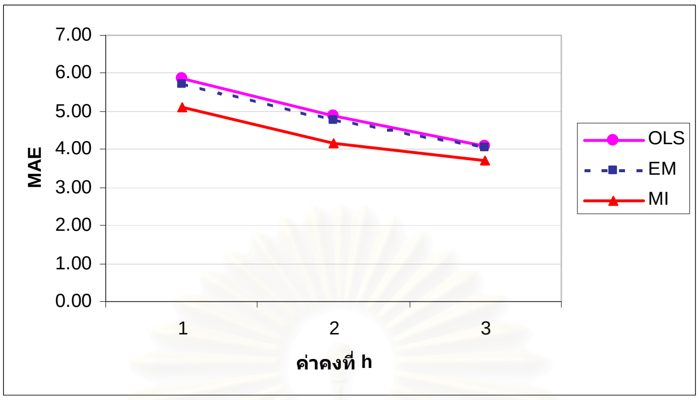
รูปที่ 4.3.18 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 4x4 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.45



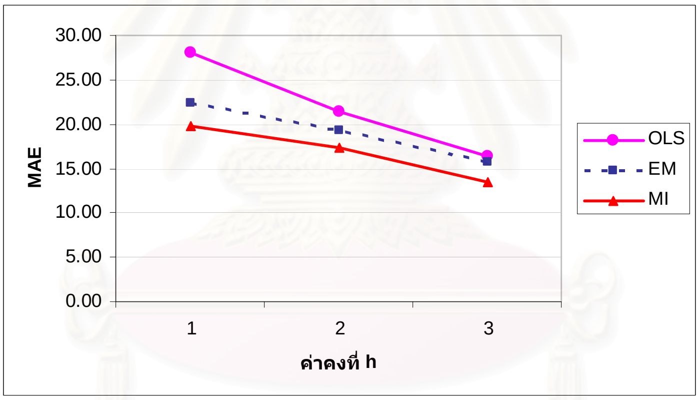
ตารางที่ 4.3.3 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน โดยพิจารณาจากค่าคงที่ h

missing	c.v	h	MAE OLS	MAE EM	MAE MI
10%	0.05	1	5.8682	5.6968	5.0957
		2	4.8969	4.7649	4.1803
		3	4.0798	4.0432	3.7023
	0.25	1	28.1094	22.3784	19.7854
		2	21.3309	19.3479	17.3448
		3	16.3850	15.7351	13.4896
	0.45	1	50.9904	45.5139	40.9688
		2	36.4578	31.9861	28.9385
		3	31.7296	31.7072	27.6233
20%	0.05	1	7.1298	5.8971	5.6700
		2	5.3906	5.0056	4.2326
		3	5.6796	5.3929	4.0337
	0.25	1	30.0478	24.3331	22.5674
		2	22.0811	20.7711	20.6736
		3	20.5194	19.1290	16.9755
	0.45	1	53.2704	50.5609	48.5021
		2	41.7682	32.1911	29.1245
		3	37.5360	35.7913	32.6195
30%	0.05	1	10.3068	7.9281	7.3466
		2	6.8966	6.4575	5.9949
		3	6.5185	6.3165	5.9189
	0.25	1	35.7777	32.9384	30.0074
		2	30.9132	29.6125	26.0193
		3	24.2990	24.1145	22.4991
	0.45	1	61.3002	54.4515	49.9025
		2	50.1699	46.2479	39.7025
		3	42.9053	39.9780	37.0511

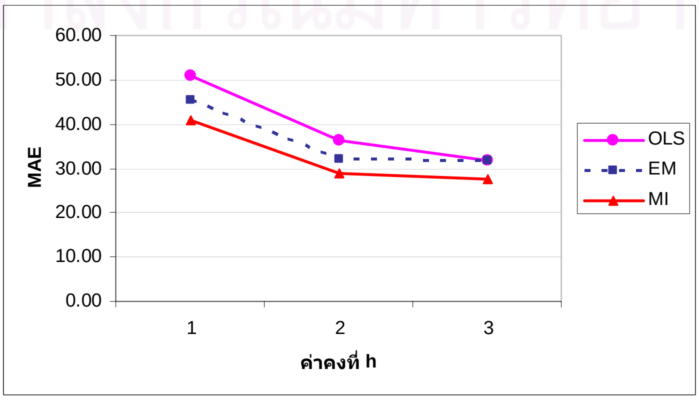
รูปที่ 4.3.19 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.05



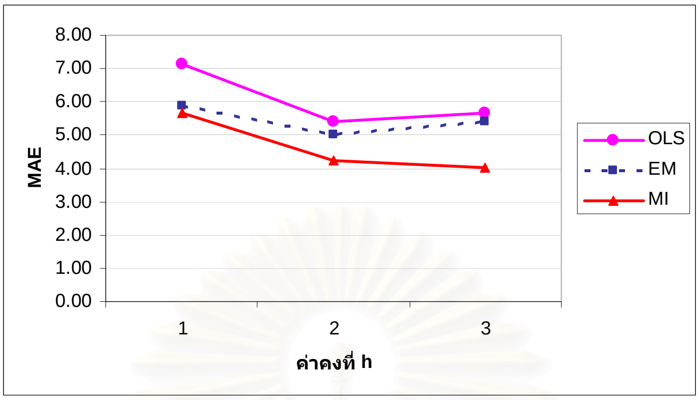
รูปที่ 4.3.20 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.25



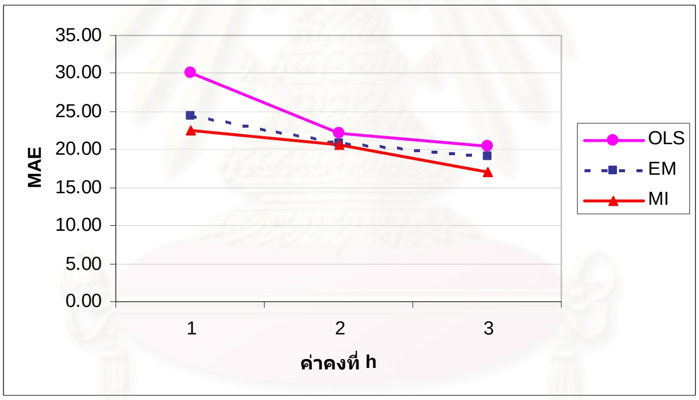
รูปที่ 4.3.21 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.45



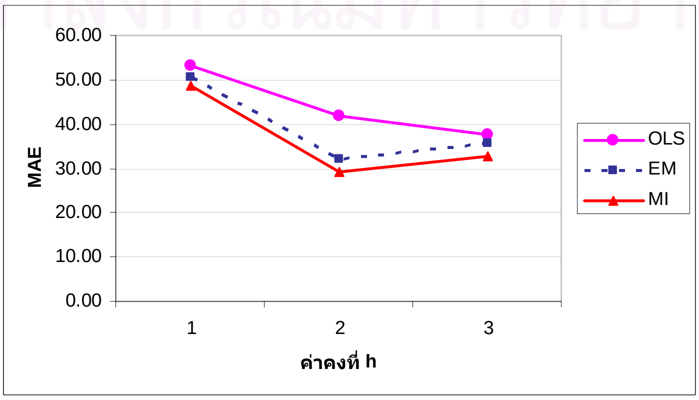
รูปที่ 4.3.22 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.05



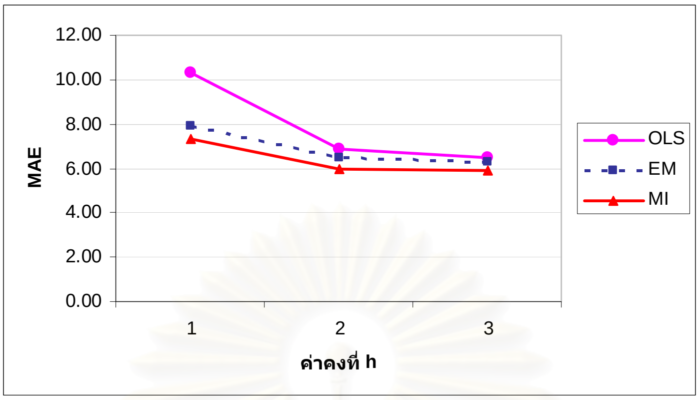
รูปที่ 4.3.23 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.25



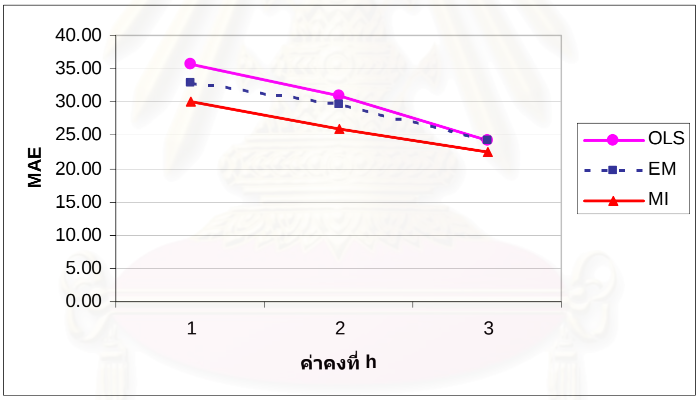
รูปที่ 4.3.24 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.45



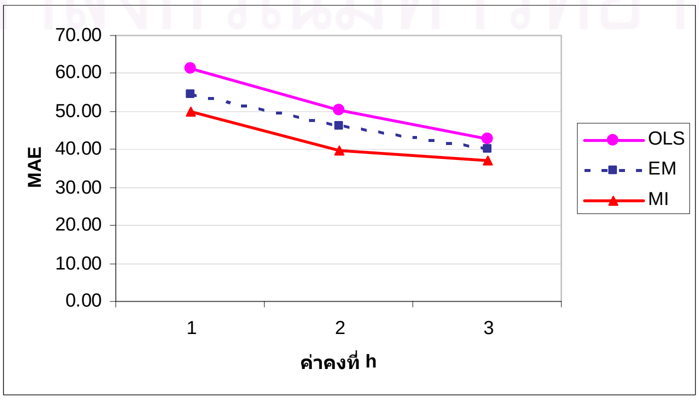
รูปที่ 4.3.25 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.3.26 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.25



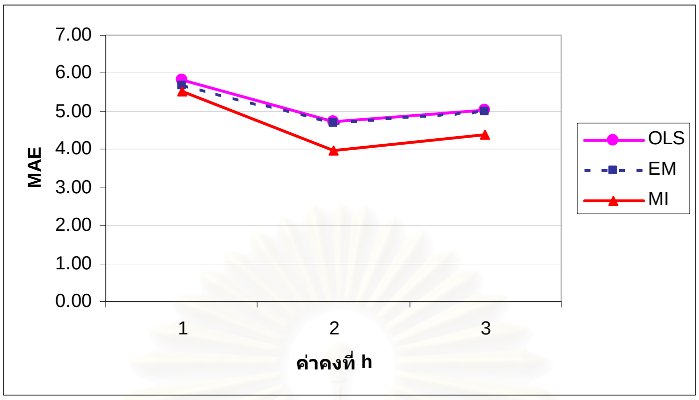
รูปที่ 4.3.27 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 5x5 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.45



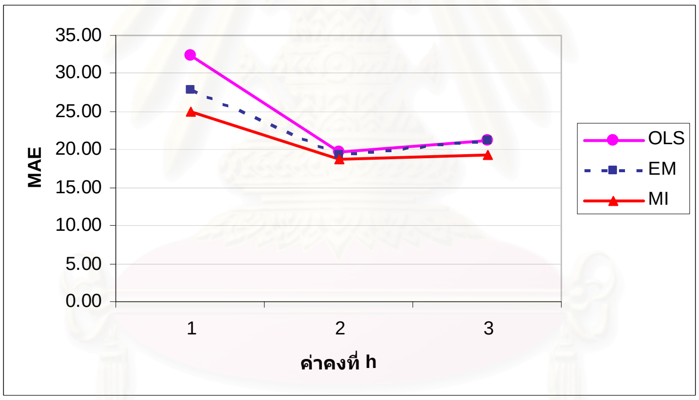
ตารางที่ 4.3.4 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน โดยพิจารณาจากค่าคงที่ h

missing	c.v	h	MAE OLS	MAE EM	MAE MI
10%	0.05	1	5.8172	5.6917	5.5372
		2	4.7243	4.6947	3.9671
		3	5.0199	5.0004	4.4029
	0.25	1	32.2936	27.7421	24.9148
		2	19.7685	19.3838	18.6600
		3	21.1072	21.1000	19.2046
	0.45	1	49.3574	45.6812	42.9947
		2	41.5621	39.5780	36.9268
		3	36.4664	33.5128	32.4316
20%	0.05	1	7.5641	7.4550	6.5478
		2	6.0893	6.1567	5.4906
		3	5.9174	5.9066	5.6390
	0.25	1	34.5393	28.4741	27.7647
		2	25.1802	22.9325	20.5171
		3	24.8663	21.9596	19.8567
	0.45	1	54.7191	46.9111	43.5965
		2	47.9628	44.9490	40.6498
		3	45.1973	39.7520	36.4955
30%	0.05	1	8.0636	7.9331	6.5879
		2	6.2787	6.2280	5.8936
		3	7.0444	6.8368	6.1050
	0.25	1	42.7861	38.1336	32.1790
		2	25.3932	24.9744	20.8245
		3	25.1659	25.0756	21.4612
	0.45	1	79.0797	62.2092	53.9579
		2	51.7270	49.8744	40.8616
		3	45.9428	44.8748	38.5298

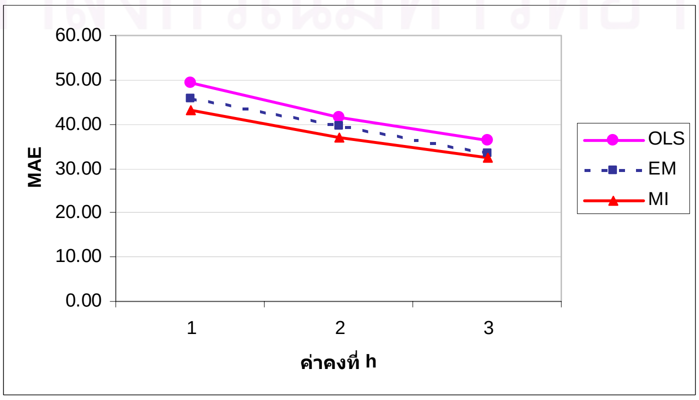
รูปที่ 4.3.28 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.05



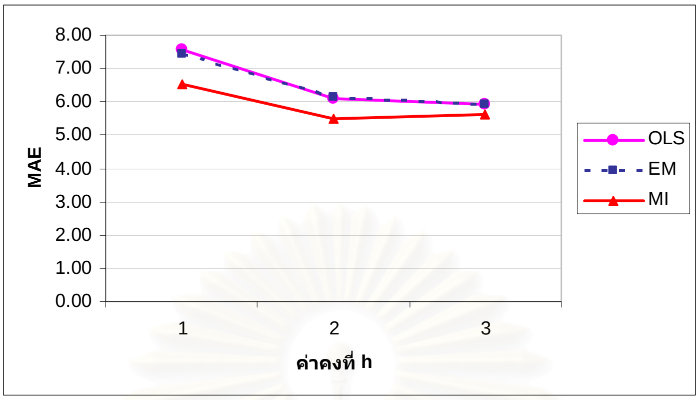
รูปที่ 4.3.29 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.25



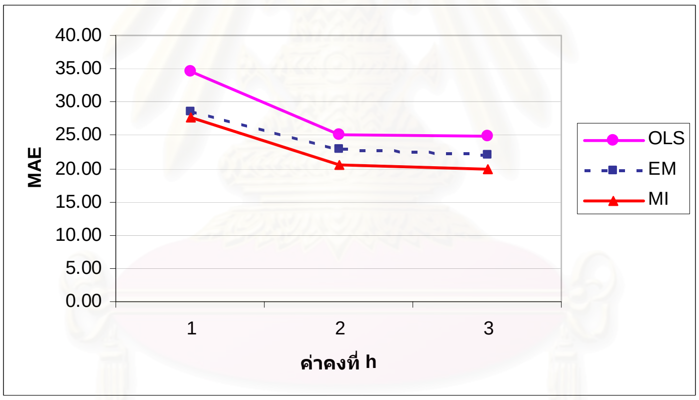
รูปที่ 4.3.30 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.45



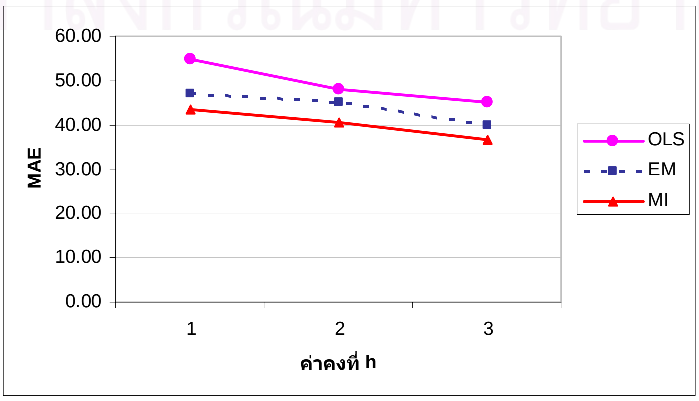
รูปที่ 4.3.31 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.05



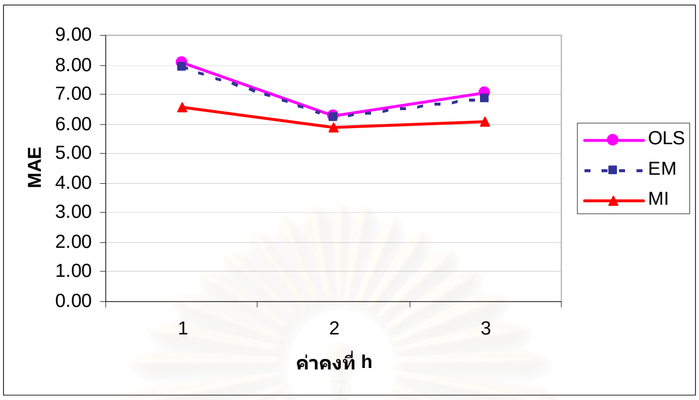
รูปที่ 4.3.32 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.25



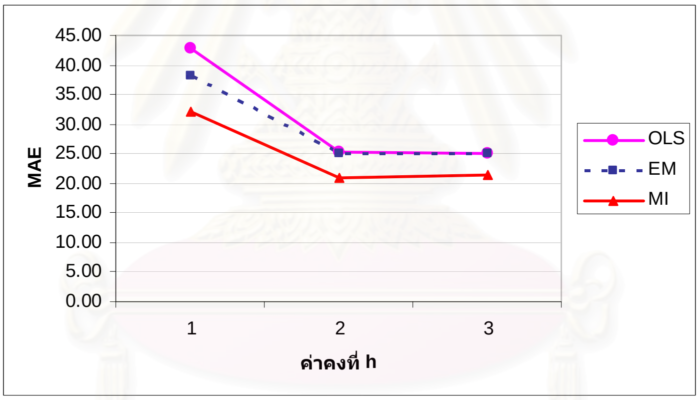
รูปที่ 4.3.33 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.45



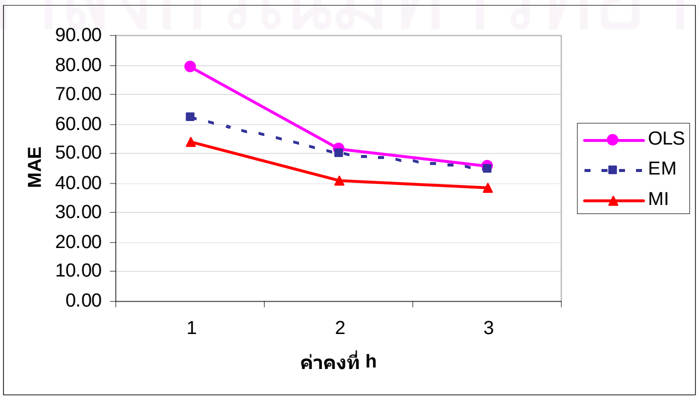
รูปที่ 4.3.34 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.3.35 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.25



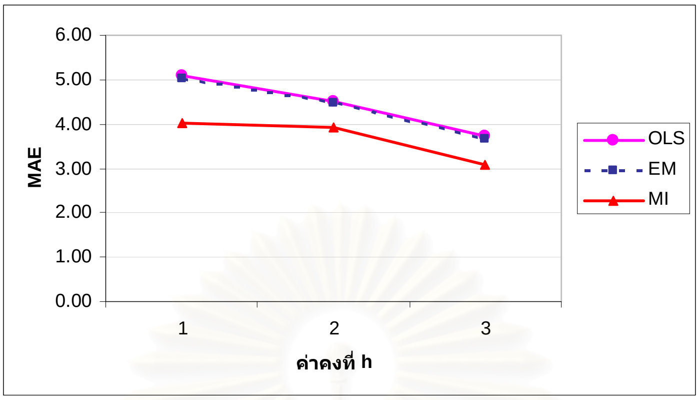
รูปที่ 4.3.36 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 6x6 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.45



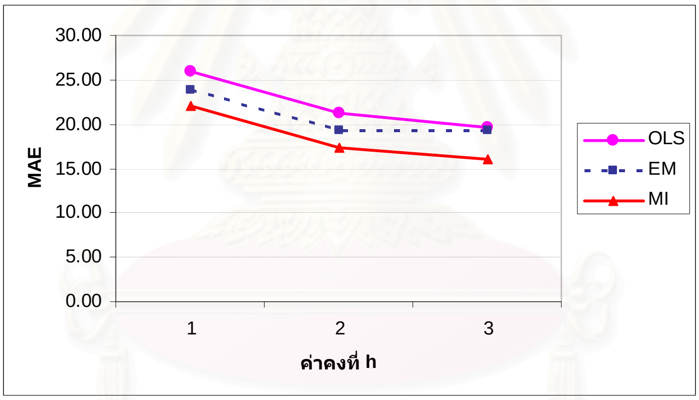
ตารางที่ 4.3.5 เปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้ง 3 วิธี ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน โดยพิจารณาจากค่าคงที่ h

missing	c.v	h	MAE OLS	MAE EM	MAE MI
10%	0.05	1	5.1068	5.0130	4.0236
		2	4.4951	4.4726	3.9210
		3	3.7407	3.6758	3.0719
	0.25	1	25.8864	23.7677	22.1253
		2	21.2013	19.2644	17.2893
		3	19.5979	19.2842	16.0675
	0.45	1	48.0682	41.0144	39.2961
		2	40.5607	36.2060	32.8648
		3	33.7431	31.6420	23.3490
20%	0.05	1	5.2391	5.1356	4.8602
		2	4.7138	4.6997	3.9301
		3	3.9915	3.9866	3.1730
	0.25	1	25.9979	25.8966	22.6826
		2	23.0790	22.8575	20.6931
		3	20.0283	19.9914	18.4019
	0.45	1	50.5838	49.7984	40.2521
		2	46.2886	45.1359	38.3525
		3	35.0983	34.6465	25.7382
30%	0.05	1	7.1953	7.0362	6.2244
		2	6.5039	6.5012	6.1110
		3	6.2119	6.1702	5.9378
	0.25	1	34.7477	27.9877	23.8433
		2	29.0086	28.6546	21.9417
		3	26.3185	23.8886	18.8652
	0.45	1	63.3971	51.8287	44.3837
		2	50.6945	49.9759	42.5841
		3	48.8774	47.9632	41.6794

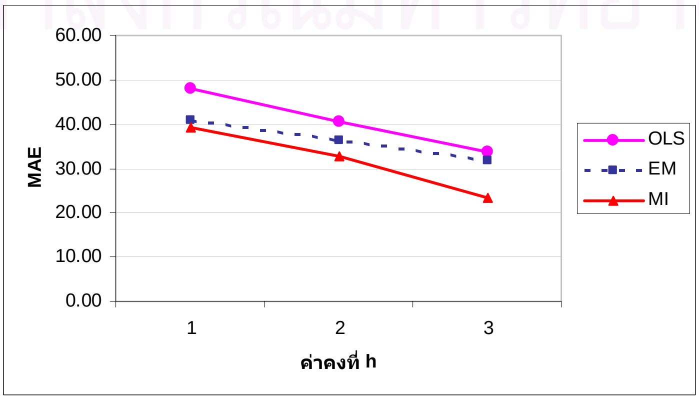
รูปที่ 4.3.37 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.05



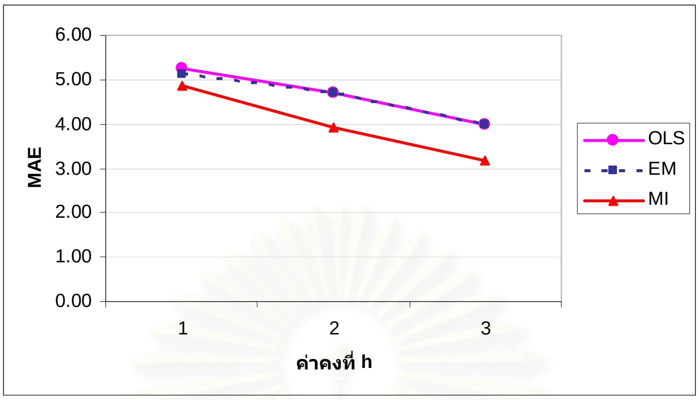
รูปที่ 4.3.38 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.25



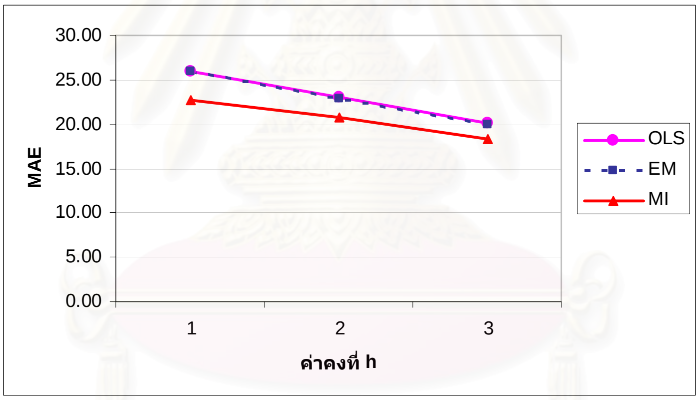
รูปที่ 4.3.39 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 10% และสัมประสิทธิ์ความผันแปร 0.45



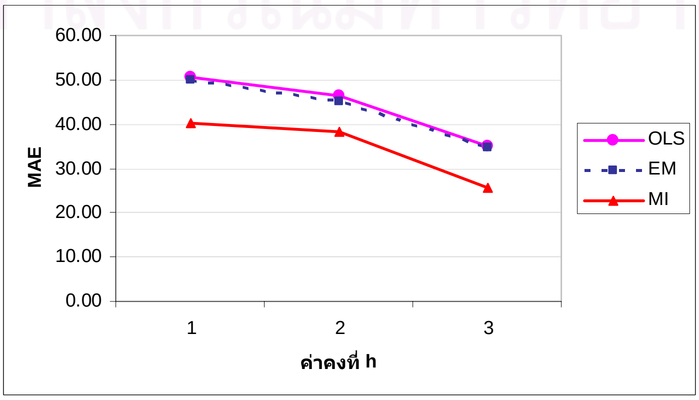
รูปที่ 4.3.40 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.05



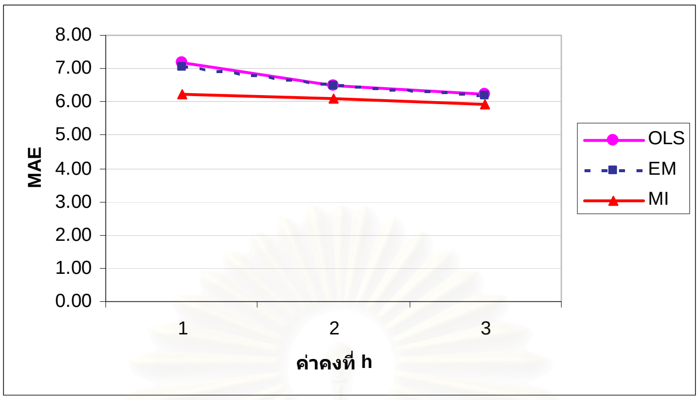
รูปที่ 4.3.41 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.25



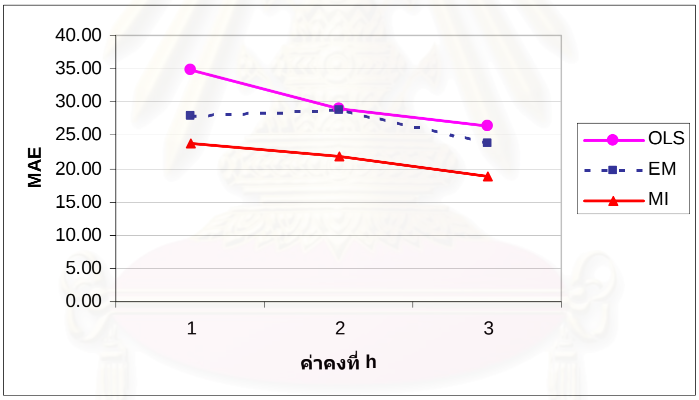
รูปที่ 4.3.42 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 20% และสัมประสิทธิ์ความผันแปร 0.45



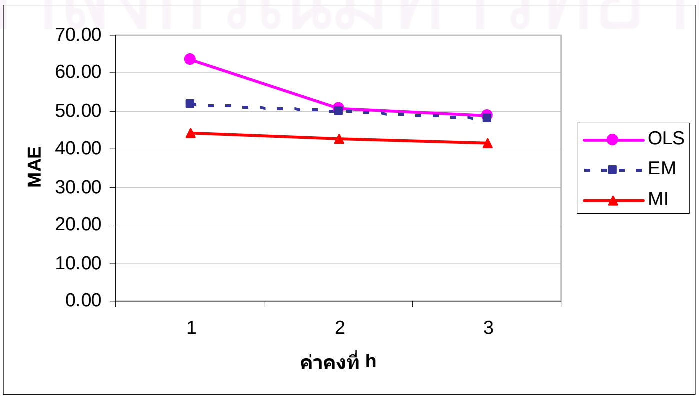
รูปที่ 4.3.43 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.05



รูปที่ 4.3.44 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.25



รูปที่ 4.3.45 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดและค่าคงที่ h ในแผนแบบการทดลองขนาด 7x7 จัตุรัสละติน กรณีเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% และสัมประสิทธิ์ความผันแปร 0.45



จากตารางที่ 4.3.1 – 4.3.5 และรูปที่ 4.3.1 – 4.3.45 แสดงการเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) ทั้ง 3 วิธี ในแผนแบบการทดลองจัตุรัสละติน ณ ค่าคงที่ h เมื่อกำหนดให้เปอร์เซ็นต์ข้อมูลสูญหายและสัมประสิทธิ์ความผันแปรคงที่ พบว่าเมื่อค่าคงที่ h มีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าแตกต่างกันเล็กน้อย และจากการวิจัยพบว่าวิธี Multiple Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุด ทุกค่าคงที่ h ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสังเกตที่สูญหายในการวางแผนการทดลองแบบจัตุรัสละติน โดยทำการเปรียบเทียบวิธีการประมาณค่าสูญหาย 3 วิธี คือ วิธีกำลังสองน้อยสุด เป็นการประมาณค่าที่ทำให้ผลรวมของค่าความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด วิธี EM algorithm เป็นการประมาณค่าพารามิเตอร์และหาค่าตัวสถิติเพียงพอโดยอาศัยฟังก์ชันภาวะน่าจะเป็นสูงสุดเข้ามาช่วย และวิธี Multiple Imputation สำหรับการวิจัยครั้งนี้ ผู้วิจัยทำการจำลองข้อมูลตามสถานการณ์ต่างๆ ที่กำหนดดังนี้

1. แผนแบบการทดลองจัตุรัสละตินขนาด 3x3 4x4 5x5 6x6 และ 7x7
2. กำหนดสัมประสิทธิ์ความผันแปรเป็น 3 ระดับ คือ 5% 25% และ 45%
3. กำหนดจำนวนข้อมูลสูญหาย คือ 10% 20% และ 30%

การเปรียบเทียบว่าวิธีการประมาณค่าสูญหายนั้นจะพิจารณาจาก ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุด (MAE) โดยวิธีใดให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดมีค่าต่ำกว่า แสดงว่าวิธีนั้นเหมาะสมสำหรับนำมาใช้ในการประมาณค่าสูญหายมากที่สุด

5.1 สรุปผลการวิจัย

5.1.1 เมื่อเปอร์เซ็นต์ข้อมูลสูญหายมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าเพิ่มขึ้น เนื่องจากเมื่อข้อมูลมีการสูญหายเพิ่มมากขึ้น จะทำให้การประมาณค่าสูญหายมีความผิดพลาดมากขึ้น และจากการวิจัยพบว่าวิธี Multiple Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุดทุกเปอร์เซ็นต์ข้อมูลสูญหาย ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

5.1.2 เมื่อค่าสัมประสิทธิ์การแปรผันมีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าเพิ่มขึ้น เนื่องจากเมื่อสัมประสิทธิ์การแปรผันของข้อมูลเพิ่มมากขึ้น จะทำให้ข้อมูลมีการกระจายมากขึ้น การประมาณค่าสูญหายจึงมีความผิดพลาดมากขึ้น และจากการวิจัยพบว่าวิธี Multiple

Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุด ทุกสัมประสิทธิ์การแปรผัน ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

เมื่อสัมประสิทธิ์การแปรผันมีค่าเพิ่มขึ้น จะพบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของวิธี Multiple Imputation ยิ่งห่างจากวิธี EM Algorithm และวิธีกำลังสองน้อยสุดมากขึ้น นั่นคือในกรณีที่ชุดข้อมูลตัวอย่างที่สุ่มมามีความผันแปรมาก ควรจะประมาณค่าสังเกตที่สูญหายโดยใช้วิธี Multiple Imputation ถึงแม้ว่าจะมีวิธีการที่ซับซ้อนมากกว่า แต่ค่าประมาณที่ได้นั้นมีค่าใกล้เคียงกับค่าจริงมากกว่า

และสำหรับกรณีที่ชุดข้อมูลตัวอย่างที่สุ่มมามีความผันแปรน้อย พบว่าค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดของทั้ง 3 วิธี มีค่าใกล้เคียงกันมาก ดังนั้นควรเลือกใช้การประมาณค่าสังเกตที่สูญหายโดยวิธีกำลังสองน้อยสุดมากกว่า เพราะสะดวกและรวดเร็วกว่า

5.1.3 เมื่อค่าคงที่ h มีค่าเพิ่มขึ้น ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดทั้งวิธีกำลังสองน้อยสุด วิธี EM Algorithm และวิธี Multiple Imputation มีค่าแตกต่างกันเล็กน้อย และจากการวิจัยพบว่าวิธี Multiple Imputation จะให้ค่าความคลาดเคลื่อนสัมบูรณ์สูงสุดต่ำกว่าวิธี EM Algorithm และวิธีกำลังสองน้อยสุด ทุกค่าคงที่ h ณ ระดับต่างๆ ในทุกสถานการณ์ นั่นคือการประมาณค่าสังเกตที่สูญหายด้วยวิธี Multiple Imputation จะให้ค่าโดยส่วนใหญ่ใกล้เคียงกับค่าจริงมากกว่าการประมาณค่าสังเกตที่สูญหายโดยวิธี EM Algorithm และวิธีกำลังสองน้อยสุด

5.2 ข้อเสนอแนะ

5.2.1 ในการประมาณค่าสูญหายนั้นมีหลายวิธีที่ถูกคิดค้นขึ้นมา แต่ละวิธีนั้นมีความเหมาะสมสำหรับแต่ละตัวแบบแตกต่างกันไป และในการวิจัยครั้งนี้ ผู้วิจัยได้เลือกวิธีการประมาณค่าสูญหายเพื่อนำมาศึกษาเปรียบเทียบเพียง 3 วิธี ดังนั้นในการวิจัยครั้งต่อไป อาจจะนำวิธีการประมาณค่าสูญหายแบบอื่นๆ มาศึกษาเพื่อหาวิธีการประมาณค่าสูญหายที่เหมาะสมที่สุดสำหรับแผนแบบการทดลองจัตุรัสละติน

5.2.2 ในการวิจัยครั้งนี้ ได้ทำการศึกษาการประมาณค่าสูญหายในแผนแบบการทดลองจัตุรัสละติน ดังนั้นในการวิจัยครั้งต่อไป อาจทำการศึกษาในตัวแบบอื่นๆ

รายการอ้างอิง

ภาษาไทย

ประพจน์ ดำรงสุทธิพงศ์. การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2546.

ปรีชา วิจิตรธรรมรส, โอม ศรนิล, สุรพงค์ เอื้อวัฒนามงคล, พาชิตชนัต ศิริพานิช. ความรู้พื้นฐานทางการทำเหมืองข้อมูล. เอกสารประกอบโครงการอบรมเชิงปฏิบัติการ รุ่นที่ 8 คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์, 2549.

ปารเมศ ชุติมา. การออกแบบการทดลองทางวิศวกรรม. พิมพ์ครั้งที่1. กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2545.

สุพล ดุรงค์วัฒนา. การวางแผนการทดลองขั้นสูง. เอกสารประกอบการสอนวิชาการวางแผนการทดลองขั้นสูง ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย, 2547.

สุรพล อุปดิสสกุล. สถิติ การวางแผนการทดลอง เล่ม 1. พิมพ์ครั้งที่3. กรุงเทพฯ : สหมิตรออฟเซท, 2536.

ภาษาอังกฤษ

Alberto Garcia – Diaz and Don T. Phillips. Principles of Experimental Design and Analysis. London : Chapman & Hall, 1995.

Andreas Krause and Melvin Olson. The basics of S and S-Plus. 2nd ed. New York : Springer Veriag, 2000.

Cary, N.C. The MI Procedure. SAS Institute, 2001.

Douglas C. Montgomery. Design and Analysis of Experiment. New York : John Wiley & Son, 1997

Geoffrey J. McLachlan & Thriyambakam Krishnan. The EM Algorithm and Extensions. New York : John Wiley & Son, 1997

Geoge Snedecor and William B. Cochran. Statistic Methods. The Iowa State Univerity Press, 1972.

Paul D. Allison. Missing Data Techniques for Structural Equation Modeling. The American Psychological Association. (2003) : 545 – 557.

Roderick J.A. Little and Donald B. Rubin. Statistical Analysis with Missing Data. New York : John Wiley & Son, 1987

Sandip Sinharay, Hal S. Stern and Daniel Russell. The Use of Multiple Imputation for the Analysis of Missing Data. The American Psychological Association. (2001) : 317 – 329.

ภาคผนวก

โปรแกรมการจำลองข้อมูลและการประมาณค่าสูญหายในแผนการทดลองจัตุรัสละติน

```
p_3
n_p^2
mi_3
m_n-mi
u_50
cv_0.05
h_1
loops_500
var.e_((cv*u)^2)/((3*h)+1)
var.t_(h*var.e)
var.r_(h*var.e)
var.c_(h*var.e)
v_round((var.e+var.t+var.r+var.c),digits=4)

###keep MMAE

MMAE_array(,dim=c(1,loops))

for(l in 1:loops)
{
er_array(rnorm(n,0,sqrt(var.e)),dim=c(p,p))
rblock_array(p,0,sqrt(var.r)),dim=c(p))
cblock_array(p,0,sqrt(var.c)),dim=c(p))

y_array(,dim=c(p,p))
for(j in 1:p)
    for(k in 1:p)
    {
        if(p==3)
        {
            if((j-k)==0)
            i_1
            else
                if(((j-k)==1)||((j-k)==-2))
                i_2
                else
                    i_3
        }

        if(p==4)
        {
            if((j-k)==0)
            i_1
```

```
    else
        if(((j-k)==1)||((j-k)==-3))
        i_2
        else
            if(((j-k)==2)||((j-k)==-2))
            i_3
            else
                i_4
}

if(p==5)
{
    if((j-k)==0)
    i_1
    else
        if(((j-k)==1)||((j-k)==-4))
        i_2
        else
            if(((j-k)==2)||((j-k)==-3))
            i_3
            else
                if(((j-k)==3)||((j-k)==-2))
                i_4
                else
                    i_5
}

if(p==6)
{
    if((j-k)==0)
    i_1
    else
        if(((j-k)==1)||((j-k)==-5))
        i_2
        else
            if(((j-k)==2)||((j-k)==-4))
            i_3
            else
                if(((j-k)==3)||((j-k)==-3))
                i_4
                else
                    if(((j-k)==4)||((j-k)==-2))
                    i_5
                    else
                        i_6
}
```

```
            if(p==7)
            {
                  if((j-k)==0)
                  i_1
                  else
                        if(((j-k)==1)||((j-k)==-6))
                        i_2
                        else
                              if(((j-k)==2)||((j-k)==-5))
                              i_3
                              else
                                    if(((j-k)==3)||((j-k)==-4))
                                    i_4
                                    else
                                          if(((j-k)==4)||((j-k)==-3))
                                          i_5
                                          else
                                                if(((j-k)==5)||((j-k)==-2))
                                                i_6
                                                else
                                                      i_7
            }
                        y[j,k]_round(u+trt[i]+rblock[j]+cblock[k]+er[j,k],digits=4)
      }

# random to delete data

y1_array(,dim=c(p,p))
y1_y

if((p==3)||(p==4))
{
reloop_0
repeat
{
      reloop_reloop+1
r_runif(9)
s_round(r,digit=2)
h_matrix(s,nrow=p,ncol=p,byrow=T)

if(((abs(h[1]-h[2]))>(abs(1/n-2/n)))&((abs(h[1]-h[3]))>(abs(1/n-2/n)))&((abs(h[1]-h[4]))>(abs(1/n-
      2/n)))&((abs(h[1]-h[5]))>(abs(1/n-2/n)))&((abs(h[2]-h[3]))>(abs(1/n-2/n)))&((abs(h[2]-h[4]))>(abs(1/n-
      2/n)))&((abs(h[2]-h[5]))>(abs(1/n-2/n)))&((abs(h[3]-h[4]))>(abs(1/n-2/n)))&((abs(h[3]-h[5]))>(abs(1/n-
      2/n)))&((abs(h[4]-h[5]))>(abs(1/n-2/n))))
      break;
}
}
```

```
if(p==3)
{
# delete data1

for(j in 1:p)
for(k in 1:p)
{
      if(h[1]>=0&h[1]<(1/n))
            y1[1]_0
            else
      if(h[1]>=(1/n)&h[1]<(2/n))
            y1[2]_0
            else
      if(h[1]>=(2/n)&h[1]<(3/n))
            y1[3]_0
            else
      if(h[1]>=(3/n)&h[1]<(4/n))
            y1[4]_0
            else
      if(h[1]>=(4/n)&h[1]<(5/n))
            y1[5]_0
            else
      if(h[1]>=(5/n)&h[1]<(6/n))
            y1[6]_0
            else
      if(h[1]>=(6/n)&h[1]<(7/n))
            y1[7]_0
            else
      if(h[1]>=(7/n)&h[1]<(8/n))
            y1[8]_0
            else
      if(h[1]>=(8/n)&h[1]<(9/n))
            y1[9]_0
            else
                  y1[j,k]_y1[j,k]
}

# delete data2
if(mi>1)
{
for(j in 1:p)
for(k in 1:p)
{
      if(h[2]>=0&h[2]<(1/n))
            y1[1]_0
            else
```

```
	if(h[2]>=(1/n)&h[2]<(2/n))
		y1[2]_0
		else
	if(h[2]>=(2/n)&h[2]<(3/n))
		y1[3]_0
		else
	if(h[2]>=(3/n)&h[2]<(4/n))
		y1[4]_0
		else
	if(h[2]>=(4/n)&h[2]<(5/n))
		y1[5]_0
		else
	if(h[2]>=(5/n)&h[2]<(6/n))
		y1[6]_0
		else
	if(h[2]>=(6/n)&h[2]<(7/n))
		y1[7]_0
		else
	if(h[2]>=(7/n)&h[2]<(8/n))
		y1[8]_0
		else
	if(h[2]>=(8/n)&h[2]<(9/n))
		y1[9]_0
		else
			y1[j,k]_y1[j,k]
}
}


#delete data3
if(mi>2)
{
for(j in 1:p)
for(k in 1:p)
{
	if(h[3]>=0&h[3]<(1/n))
		y1[1]_0
		else
	if(h[3]>=(1/n)&h[3]<(2/n))
		y1[2]_0
		else
	if(h[3]>=(2/n)&h[3]<(3/n))
		y1[3]_0
		else
	if(h[3]>=(3/n)&h[3]<(4/n))
		y1[4]_0
		else
```

```
    if(h[3]>=(4/n)&h[3]<(5/n))
        y1[5]_0
        else
    if(h[3]>=(5/n)&h[3]<(6/n))
        y1[6]_0
        else
    if(h[3]>=(6/n)&h[3]<(7/n))
        y1[7]_0
        else
    if(h[3]>=(7/n)&h[3]<(8/n))
        y1[8]_0
        else
    if(h[3]>=(8/n)&h[3]<(9/n))
        y1[9]_0
        else
            y1[j,k]_y1[j,k]
}
}
}

# Determine sum treatment rblock and cblock

mt_array(0,dim=c(p,1))
mrb_array(0,dim=c(p,1))
mcb_array(0,dim=c(p,1))

if(p==3)
{
    for(j in 1:p)
        for(k in 1:p)
    {
        ### sum trt
        {
            if((j-k)==0)
                mt[1]_mt[1]+y1[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt[2]_mt[2]+y1[j,k]
                        else
                            mt[3]_mt[3]+y1[j,k]
        }
        ### sum rblock
        {
            if(j==1)
                mrb[1]_mrb[1]+y1[1,k]
                else
                    if(j==2)
```

```
                    mrb[2]_mrb[2]+y1[2,k]
                    else
                        mrb[3]_mrb[3]+y1[3,k]
        }
        ### sum cblock
        {
            if(k==1)
                mcb[1]_mcb[1]+y1[j,1]
                else
                    if(k==2)
                        mcb[2]_mcb[2]+y1[j,2]
                        else
                            mcb[3]_mcb[3]+y1[j,3]
        }

    }
}

if(p==4)
{
    for(j in 1:p)
        for(k in 1:p)
    {

        ### sum trt

{
            if((j-k)==0)
            mt[1]_mt[1]+y1[j,k]
            else
                if(((j-k)==1)||((j-k)==-3))
                mt[2]_mt[2]+y1[j,k]
                else
                    if(((j-k)==2)||((j-k)==-2))
                    mt[3]_mt[3]+y1[j,k]
                    else
                        mt[4]_mt[4]+y1[j,k]
        }
        ### sum rblock
        {
            if(j==1)
            mrb[1]_mrb[1]+y1[1,k]
            else
                if(j==2)
                mrb[2]_mrb[2]+y1[2,k]
                else
                    if(j==3)
```

```
                        mrb[3]_mrb[3]+y1[3,k]
                        else
                            mrb[4]_mrb[4]+y1[4,k]
            }

            ### sum cblock
            {
                if(k==1)
                mcb[1]_mcb[1]+y1[j,1]
                else
                    if(k==2)
                    mcb[2]_mcb[2]+y1[j,2]
                    else
                        if(k==3)
                        mcb[3]_mcb[3]+y1[j,3]
                        else
                            mcb[4]_mcb[4]+y1[j,4]
            }

        }##end for
    }##end if

    if(p==5)
    {
    for(j in 1:p)
        for(k in 1:p)
    {
        ## sum trt
        {
            if((j-k)==0)
            mt[1]_mt[1]+y1[j,k]
            else
                if(((j-k)==1)||((j-k)==-4))
                mt[2]_mt[2]+y1[j,k]
                else
                    if(((j-k)==2)||((j-k)==-3))
                    mt[3]_mt[3]+y1[j,k]
                    else
                        if(((j-k)==3)||((j-k)==-2))
                        mt[4]_mt[4]+y1[j,k]
                        else
                            mt[5]_mt[5]+y1[j,k]
        }
        ## sum rblock
        {
            if(j==1)
            mrb[1]_mrb[1]+y1[1,k]
```

```
          else
                if(j==2)
                mrb[2]_mrb[2]+y1[2,k]
                else
                      if(j==3)
                      mrb[3]_mrb[3]+y1[3,k]
                      else
                            if(j==4)
                            mrb[4]_mrb[4]+y1[4,k]
                            else
                                  mrb[5]_mrb[5]+y1[5,k]
    }
    ## sum cblock
    {
          if(k==1)
          mcb[1]_mcb[1]+y1[j,1]
          else
                if(k==2)
                mcb[2]_mcb[2]+y1[j,2]
                else
                      if(k==3)
                      mcb[3]_mcb[3]+y1[j,3]
                      else
                            if(k==4)
                            mcb[4]_mcb[4]+y1[j,4]
                            else
                                  mcb[5]_mcb[5]+y1[j,5]
    }

}##end for
 }

if(p==6)
{
for(j in 1:p)
    for(k in 1:p)
{
    ### sum trt
    {
          if((j-k)==0)
          mt[1]_mt[1]+y1[j,k]
          else
                if(((j-k)==1)||((j-k)==-5))
                mt[2]_mt[2]+y1[j,k]
                else
                      if(((j-k)==2)||((j-k)==-4))
                      mt[3]_mt[3]+y1[j,k]
```

```
                else
                    if(((j-k)==3)||((j-k)==-3))
                    mt[4]_mt[4]+y1[j,k]
                    else
                        if(((j-k)==4)||((j-k)==-2))
                        mt[5]_mt[5]+y1[j,k]
                        else
                            mt[6]_mt[6]+y1[j,k]
}
### sum rblock
{
    if(j==1)
    mrb[1]_mrb[1]+y1[1,k]
    else
        if(j==2)
        mrb[2]_mrb[2]+y1[2,k]
        else
            if(j==3)
            mrb[3]_mrb[3]+y1[3,k]
            else
                if(j==4)
                mrb[4]_mrb[4]+y1[4,k]
                else
                    if(j==5)
                    mrb[5]_mrb[5]+y1[5,k]
                    else
                        mrb[6]_mrb[6]+y1[6,k]
}
### sum cblock
{
    if(k==1)
    mcb[1]_mcb[1]+y1[j,1]
    else
        if(k==2)
        mcb[2]_mcb[2]+y1[j,2]
        else
            if(k==3)
            mcb[3]_mcb[3]+y1[j,3]
            else
                if(k==4)
                mcb[4]_mcb[4]+y1[j,4]
                else
                    if(k==5)
                    mcb[5]_mcb[5]+y1[j,5]
                    else
                        mcb[6]_mcb[6]+y1[j,6]
}
```

```
}
}

if(p==7)
{
for(j in 1:p)
    for(k in 1:p)
{
    ## sum trt
    {
        if((j-k)==0)
        mt[1]_mt[1]+y1[j,k]
        else
            if(((j-k)==1)||((j-k)==-6))
            mt[2]_mt[2]+y1[j,k]
            else
                if(((j-k)==2)||((j-k)==-5))
                mt[3]_mt[3]+y1[j,k]
                else
                    if(((j-k)==3)||((j-k)==-4))
                    mt[4]_mt[4]+y1[j,k]
                    else
                        if(((j-k)==4)||((j-k)==-3))
                        mt[5]_mt[5]+y1[j,k]
                        else
                            if(((j-k)==5)||((j-k)==-2))
                            mt[6]_mt[6]+y1[j,k]
                            else
                                mt[7]_mt[7]+y1[j,k]
    }
    ## sum rblock
    {
        if(j==1)
        mrb[1]_mrb[1]+y1[1,k]
        else
            if(j==2)
            mrb[2]_mrb[2]+y1[2,k]
            else
                if(j==3)
                mrb[3]_mrb[3]+y1[3,k]
                else
                    if(j==4)
                    mrb[4]_mrb[4]+y1[4,k]
                    else
                        if(j==5)
                        mrb[5]_mrb[5]+y1[5,k]
                        else
```

```
                                          if(j==6)
                                          mrb[6]_mrb[6]+y1[6,k]
                                          else
                                                mrb[7]_mrb[7]+y1[7,k]
      }
      ## sum cblock
      {
            if(k==1)
            mcb[1]_mcb[1]+y1[j,1]
            else
                  if(k==2)
                  mcb[2]_mcb[2]+y1[j,2]
                  else
                        if(k==3)
                        mcb[3]_mcb[3]+y1[j,3]
                        else
                              if(k==4)
                              mcb[4]_mcb[4]+y1[j,4]
                              else
                                    if(k==5)
                                    mcb[5]_mcb[5]+y1[j,5]
                                    else
                                          if(k==6)
                                          mcb[6]_mcb[6]+y1[j,6]
                                          else
                                                mcb[7]_mcb[7]+y1[j,7]
      }
}
 }

####### Least Square Method for 3x3
mm_array(0,dim=c(p,p))
m1_array(0,dim=c(p,p))
m2_array(0,dim=c(p,p))
m3_array(0,dim=c(p,p))
m1_y1
m2_y1
m3_y1

#### count cell that not zero

no_array(,dim=c(p,p))
for(j in 1:p)
      for(k in 1:p)
      {
            if(y1[j,k]>0)
                  no[j,k]_1
```

```
            else
                  no[j,k]_0
      }

ntrt_array(0,dim=c(p,1))
nrb_array(0,dim=c(p,1))
ncb_array(0,dim=c(p,1))

if(p==3)
{
      for(j in 1:p)
            for(k in 1:p)
      {
            ##no trt
            {
                  if((j-k)==0)
                        ntrt[1]_ntrt[1]+no[j,k]
                        else
                              if(((j-k)==1)||((j-k)==-2))
                                    ntrt[2]_ntrt[2]+no[j,k]
                                    else
                                          ntrt[3]_ntrt[3]+no[j,k]
            }
            ##no rb
            {
                  if(j==1)
                        nrb[1]_nrb[1]+no[1,k]
                        else
                              if(j==2)
                                    nrb[2]_nrb[2]+no[2,k]
                                    else
                                          nrb[3]_nrb[3]+no[3,k]

            }
            ##no cb
            {
                  if(k==1)
                        ncb[1]_ncb[1]+no[j,1]
                        else
                              if(k==2)
                                    ncb[2]_ncb[2]+no[j,2]
                                    else
                                          ncb[3]_ncb[3]+no[j,3]
            }
      }
}
```

```
nttot_array(,dim=c(p,1))
nrtot_array(,dim=c(p,1))
nctot_array(,dim=c(p,1))

for(i in 1:p)
{
    ##nttot
    {
        if(ntrt[i]>0)
            nttot[i]_1
            else
                nttot[i]_0
    }
    ##nrtot
    {
        if(nrb[i]>0)
            nrtot[i]_1
            else
                nrtot[i]_0
    }
    ##nctot
    {
        if(ncb[i]>0)
            nctot[i]_1
            else
                nctot[i]_0
    }
}

nt_array(,dim=c(p,p))

for(j in 1:p)
    for(k in 1:p)
    {
        if((j-k)==0)
        nt[j,k]_nttot[1]+nrtot[j]+nctot[k]
        else
            if(((j-k)==1)||((j-k)==-2))
            nt[j,k]_nttot[2]+nrtot[j]+nctot[k]
            else
                nt[j,k]_nttot[3]+nrtot[j]+nctot[k]
    }
```

```
##########################################
if(mi==1)
{
for(j in 1:p)
    for(k in 1:p)
    {
        if(y1[j,k]==0)
        {
            if((j-k)==0)
            mm[j,k]_((p*((mt[1])+(mrb[j])+(mcb[k])))-(2*(sum(y1[1:n]))))/((p-2)*(p-1))
            else
                if(((j-k)==1)||((j-k)==-2))
                mm[j,k]_((p*((mt[2])+(mrb[j])+(mcb[k])))-(2*(sum(y1[1:n]))))/((p-2)*(p-1))
                else
                    mm[j,k]_((p*((mt[3])+(mrb[j])+(mcb[k])))-(2*(sum(y1[1:n]))))/((p-2)*(p-1))
        }
        else
        mm[j,k]_y1[j,k]
    }
}## end mi=1

####################################
if(mi==2)
{
### find m1
for(j in 1:p)
    for(k in 1:p)
    {
        if(y1[j,k]==0)
        {
            if((j==1)&(k==1))
            {
            m1[j,k]_((mt[1]/(ntrt[1]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==2)&(k==1)&(y1[1]!=0))
            {
                m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==3)&(k==1)&(y1[1]!=0)&(y1[2]!=0))
            {
                m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==1)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0))
            {
                m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==2)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0))
```

```
                {
                    m1[j,k]_((mt[1]/(ntrt[1]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                }
                if((j==3)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0)&(y1[5]!=0))
                {
                    m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                }
                if((j==1)&(k==3)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0)&(y1[5]!=0)&(y1[6]!=0))
                {
                    m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                }

            if((j==2)&(k==3)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0)&(y1[5]!=0)&(y1[6]!=0)&(y1[7]!=0))
                {
                    m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                }

            }##end y1=0

        }##end for

# Determine sum treatment rblock and cblock
reloop_0
repeat
{

reloop_reloop+1

mt1_array(0,dim=c(p,1))
mrb1_array(0,dim=c(p,1))
mcb1_array(0,dim=c(p,1))

for(j in 1:p)
    for(k in 1:p)
    {
        ### sum trt
        {
            if((j-k)==0)
                mt1[1]_mt1[1]+m1[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt1[2]_mt1[2]+m1[j,k]
                        else
                            mt1[3]_mt1[3]+m1[j,k]
        }
```

```
        ### sum rblock
        {
            if(j==1)
                mrb1[1]_mrb1[1]+m1[1,k]
                else
                    if(j==2)
                        mrb1[2]_mrb1[2]+m1[2,k]
                        else
                            mrb1[3]_mrb1[3]+m1[3,k]
        }
        ### sum cblock
        {
            if(k==1)
                mcb1[1]_mcb1[1]+m1[j,1]
                else
                    if(k==2)
                        mcb1[2]_mcb1[2]+m1[j,2]
                        else
                            mcb1[3]_mcb1[3]+m1[j,3]
        }
    }

###find m2
for(j in 1:p)
    for(k in 1:p)
    {
        if(m1[j,k]==0)
        {
            if((j-k)==0)
            m2[j,k]_((p*((mt1[1])+(mrb1[j])+(mcb1[k])))-(2*(sum(m1[1:n]))))/((p-2)*(p-1))
            else
                if(((j-k)==1)||((j-k)==-2))
                m2[j,k]_((p*((mt1[2])+(mrb1[j])+(mcb1[k])))-(2*(sum(m1[1:n]))))/((p-2)*(p-1))
                else
                    m2[j,k]_((p*((mt1[3])+(mrb1[j])+(mcb1[k])))-(2*(sum(m1[1:n]))))/((p-2)*(p-1))
        }
        else
        m2[j,k]_y1[j,k]
    }

# Determine sum treatment rblock and cblock

mt2_array(0,dim=c(p,1))
mrb2_array(0,dim=c(p,1))
mcb2_array(0,dim=c(p,1))

for(j in 1:p)
```

```
for(k in 1:p)
{
      ### sum trt
      {
            if((j-k)==0)
                  mt2[1]_mt2[1]+m2[j,k]
                  else
                        if(((j-k)==1)||((j-k)==-2))
                              mt2[2]_mt2[2]+m2[j,k]
                              else
                                    mt2[3]_mt2[3]+m2[j,k]
      }
      ### sum rblock
      {
            if(j==1)
                  mrb2[1]_mrb2[1]+m2[1,k]
                  else
                        if(j==2)
                              mrb2[2]_mrb2[2]+m2[2,k]
                              else
                                    mrb2[3]_mrb2[3]+m2[3,k]
      }
      ### sum cblock
      {
            if(k==1)
                  mcb2[1]_mcb2[1]+m2[j,1]
                  else
                        if(k==2)
                              mcb2[2]_mcb2[2]+m2[j,2]
                              else
                                    mcb2[3]_mcb2[3]+m2[j,3]
      }
}

###find m1 again
for(j in 1:p)
      for(k in 1:p)
      {
            if(m2[j,k]==0)
            {
                  if((j-k)==0)
                  m1[j,k]_((p*((mt2[1])+(mrb2[j])+(mcb2[k])))-(2*(sum(m2[1:n]))))/((p-2)*(p-1))
                  else
                        if(((j-k)==1)||((j-k)==-2))
                        m1[j,k]_((p*((mt2[2])+(mrb2[j])+(mcb2[k])))-(2*(sum(m2[1:n]))))/((p-2)*(p-1))
                        else
                              m1[j,k]_((p*((mt2[3])+(mrb2[j])+(mcb2[k])))-(2*(sum(m2[1:n]))))/((p-2)*(p-1))
```

```
            }
            else
            m1[j,k]_y1[j,k]
      }
                  if(reloop==30)
                  break;
}
##calculate mm
mm_m1
for(j in 1:p)
      for(k in 1:p)
      {
            if(mm[j,k]==0)
                  mm[j,k]_m2[j,k]
                  else
                        mm[j,k]_m1[j,k]
      }


}###end mi=2

################################

if(mi==3)
{
for(j in 1:p)
      for(k in 1:p)
      {
            if(y1[j,k]==0)
            {
                  if((j==1)&(k==1))
                  {
                        if(ntrt[1]==0)
                        m1[j,k]_((mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                        else
                              if(nrb[j]==0)
                              m1[j,k]_((mt[1]/(ntrt[1]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                              else
                                    if(ncb[k]==0)
                                    m1[j,k]_((mt[1]/(ntrt[1]))+(mrb[j]/(nrb[j])))/(nt[j,k])
                                    else
                                          m1[j,k]_((mt[1]/(ntrt[1]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                  }
                  if((j==2)&(k==1)&(y1[1]!=0))
                  {
                        if(ntrt[2]==0)
                        m1[j,k]_((mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
```

```
                else
                    if(nrb[j]==0)
                    m1[j,k]_((mt[2]/(ntrt[2]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                    else
                        m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==3)&(k==1)&(y1[1]!=0)&(y1[2]!=0))
            {
                if(ntrt[3]==0)
                m1[j,k]_((mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                else
                    if(nrb[j]==0)
                    m1[j,k]_((mt[3]/(ntrt[3]))+(mcb[k]/(ncb[k])))/(nt[j,k])
                    else
                        m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==1)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0))
            {
                if(ncb[k]==0)
                m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j])))/(nt[j,k])
                else
                    m1[j,k]_((mt[3]/(ntrt[3]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==2)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0))
            {
                m1[j,k]_((mt[1]/(ntrt[1]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==3)&(k==2)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0)&(y1[5]!=0))
            {
                m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }
            if((j==1)&(k==3)&(y1[1]!=0)&(y1[2]!=0)&(y1[3]!=0)&(y1[4]!=0)&(y1[5]!=0)&(y1[6]!=0))
            {
                m1[j,k]_((mt[2]/(ntrt[2]))+(mrb[j]/(nrb[j]))+(mcb[k]/(ncb[k])))/(nt[j,k])
            }

        }##end y1=0
    }##end for

# Determine sum treatment rblock and cblock
mt1_array(0,dim=c(p,1))
mrb1_array(0,dim=c(p,1))
mcb1_array(0,dim=c(p,1))

for(j in 1:p)
    for(k in 1:p)
    {
```

```
        ### sum trt
        {
            if((j-k)==0)
                mt1[1]_mt1[1]+m1[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt1[2]_mt1[2]+m1[j,k]
                        else
                            mt1[3]_mt1[3]+m1[j,k]
        }
        ### sum rblock
        {
            if(j==1)
                mrb1[1]_mrb1[1]+m1[1,k]
                else
                    if(j==2)
                        mrb1[2]_mrb1[2]+m1[2,k]
                        else
                            mrb1[3]_mrb1[3]+m1[3,k]
        }
        ### sum cblock
        {
            if(k==1)
                mcb1[1]_mcb1[1]+m1[j,1]
                else
                    if(k==2)
                        mcb1[2]_mcb1[2]+m1[j,2]
                        else
                            mcb1[3]_mcb1[3]+m1[j,3]
        }
    }

#### count cell that not zero (second round)

no1_array(,dim=c(p,p))
for(j in 1:p)
    for(k in 1:p)
    {
        if(m1[j,k]>0)
            no1[j,k]_1
            else
                no1[j,k]_0
    }

ntrt1_array(0,dim=c(p,1))
nrb1_array(0,dim=c(p,1))
ncb1_array(0,dim=c(p,1))
```

```
if(p==3)
{
	for(j in 1:p)
		for(k in 1:p)
	{
		##no trt
		{
			if((j-k)==0)
				ntrt1[1]_ntrt1[1]+no1[j,k]
				else
					if(((j-k)==1)||((j-k)==-2))
						ntrt1[2]_ntrt1[2]+no1[j,k]
						else
							ntrt1[3]_ntrt1[3]+no1[j,k]
		}
		##no rb
		{
			if(j==1)
				nrb1[1]_nrb1[1]+no1[1,k]
				else
					if(j==2)
						nrb1[2]_nrb1[2]+no1[2,k]
						else
							nrb1[3]_nrb1[3]+no1[3,k]

		}
		##no cb
		{
			if(k==1)
				ncb1[1]_ncb1[1]+no1[j,1]
				else
					if(k==2)
						ncb1[2]_ncb1[2]+no1[j,2]
						else
							ncb1[3]_ncb1[3]+no1[j,3]
		}
	}
}

nttot1_array(,dim=c(p,1))
nrtot1_array(,dim=c(p,1))
nctot1_array(,dim=c(p,1))

for(i in 1:p)
{
	##nttot
```

```
        {
            if(ntrt1[i]>0)
                nttot1[i]_1
                else
                    nttot1[i]_0
        }
        ##nrtot
        {
            if(nrb1[i]>0)
                nrtot1[i]_1
                else
                    nrtot1[i]_0
        }
        ##nctot
        {
            if(ncb1[i]>0)
                nctot1[i]_1
                else
                    nctot1[i]_0
        }
}

nt1_array(,dim=c(p,p))

for(j in 1:p)
    for(k in 1:p)
    {
        if((j-k)==0)
        nt1[j,k]_nttot1[1]+nrtot1[j]+nctot1[k]
        else
            if(((j-k)==1)||((j-k)==-2))
            nt1[j,k]_nttot1[2]+nrtot1[j]+nctot1[k]
            else
                nt1[j,k]_nttot1[3]+nrtot1[j]+nctot1[k]
    }

###find m2
for(j in 1:p)
    for(k in 1:p)
    {
        if(m1[j,k]==0)
        {
            if((j==2)&(k==1))
            {
                m2[j,k]_((mt1[2]/(ntrt1[2]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==3)&(k==1)&(m1[2]!=0))
```

```
            {
                m2[j,k]_((mt1[3]/(ntrt1[3]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==1)&(k==2)&(m1[2]!=0)&(m1[3]!=0))
            {
                m2[j,k]_((mt1[3]/(ntrt1[3]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==2)&(k==2)&(m1[2]!=0)&(m1[3]!=0)&(m1[4]!=0))
            {
                m2[j,k]_((mt1[1]/(ntrt1[1]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==3)&(k==2)&(m1[2]!=0)&(m1[3]!=0)&(m1[4]!=0)&(m1[5]!=0))
            {
                m2[j,k]_((mt1[2]/(ntrt1[2]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==1)&(k==3)&(m1[2]!=0)&(m1[3]!=0)&(m1[4]!=0)&(m1[5]!=0)&(m1[6]!=0))
            {
                m2[j,k]_((mt1[2]/(ntrt1[2]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }
            if((j==2)&(k==3)&(m1[2]!=0)&(m1[3]!=0)&(m1[4]!=0)&(m1[5]!=0)&(m1[6]!=0)&(m1[7]!=0))
            {
                m2[j,k]_((mt1[3]/(ntrt1[3]))+(mrb1[j]/(nrb1[j]))+(mcb1[k]/(ncb1[k])))/(nt1[j,k])
            }

        }##end y1=0

    }##end for

reloop_0
repeat
{

reloop_reloop+1

##calculate m12
m12_m1
for(j in 1:p)
    for(k in 1:p)
    {
        if(m12[j,k]==0)
            m12[j,k]_m2[j,k]
            else
                m12[j,k]_m1[j,k]
    }
```

```
# Determine sum treatment rblock and cblock
mt12_array(0,dim=c(p,1))
mrb12_array(0,dim=c(p,1))
mcb12_array(0,dim=c(p,1))

for(j in 1:p)
    for(k in 1:p)
    {
        ### sum trt
        {
            if((j-k)==0)
                mt12[1]_mt12[1]+m12[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt12[2]_mt12[2]+m12[j,k]
                        else
                            mt12[3]_mt12[3]+m12[j,k]
        }
        ### sum rblock
        {
            if(j==1)
                mrb12[1]_mrb12[1]+m12[1,k]
                else
                    if(j==2)
                        mrb12[2]_mrb12[2]+m12[2,k]
                        else
                            mrb12[3]_mrb12[3]+m12[3,k]
        }
        ### sum cblock
        {
            if(k==1)
                mcb12[1]_mcb12[1]+m12[j,1]
                else
                    if(k==2)
                        mcb12[2]_mcb12[2]+m12[j,2]
                        else
                            mcb12[3]_mcb12[3]+m12[j,3]
        }
    }

###find m3
for(j in 1:p)
    for(k in 1:p)
    {
        if(m12[j,k]==0)
        {
            if((j-k)==0)
```

```
            m3[j,k]_((p*((mt12[1])+(mrb12[j])+(mcb12[k])))-(2*(sum(m12[1:n]))))/((p-2)*(p-1))
            else
                if(((j-k)==1)||((j-k)==-2))
                m3[j,k]_((p*((mt12[2])+(mrb12[j])+(mcb12[k])))-(2*(sum(m12[1:n]))))/((p-2)*(p-1))
                else
                    m3[j,k]_((p*((mt12[3])+(mrb12[j])+(mcb12[k])))-(2*(sum(m12[1:n]))))/((p-2)*(p-1))
        }
        else
        m3[j,k]_y1[j,k]
    }

##calculate m23
m23_m2
for(j in 1:p)
    for(k in 1:p)
    {
        if(m23[j,k]==0)
            m23[j,k]_m3[j,k]
            else
                m23[j,k]_m2[j,k]
    }

# Determine sum treatment rblock and cblock
mt23_array(0,dim=c(p,1))
mrb23_array(0,dim=c(p,1))
mcb23_array(0,dim=c(p,1))

for(j in 1:p)
    for(k in 1:p)
    {
        ### sum trt
        {
            if((j-k)==0)
                mt23[1]_mt23[1]+m23[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt23[2]_mt23[2]+m23[j,k]
                        else
                            mt23[3]_mt23[3]+m23[j,k]
        }
        ### sum rblock
        {
            if(j==1)
                mrb23[1]_mrb23[1]+m23[1,k]
                else
                    if(j==2)
                        mrb23[2]_mrb23[2]+m23[2,k]
```

```
                        else
                              mrb23[3]_mrb23[3]+m23[3,k]
            }
            ### sum cblock
            {
                  if(k==1)
                        mcb23[1]_mcb23[1]+m23[j,1]
                        else
                              if(k==2)
                                    mcb23[2]_mcb23[2]+m23[j,2]
                                    else
                                          mcb23[3]_mcb23[3]+m23[j,3]
            }
      }

###find m1 again
for(j in 1:p)
      for(k in 1:p)
      {
            if(m23[j,k]==0)
            {
                  if((j-k)==0)
                  m1[j,k]_((p*((mt23[1])+(mrb23[j])+(mcb23[k])))-(2*(sum(m23[1:n]))))/((p-2)*(p-1))
                  else
                        if(((j-k)==1)||((j-k)==-2))
                        m1[j,k]_((p*((mt23[2])+(mrb23[j])+(mcb23[k])))-(2*(sum(m23[1:n]))))/((p-2)*(p-1))
                        else
                              m1[j,k]_((p*((mt23[3])+(mrb23[j])+(mcb23[k])))-(2*(sum(m23[1:n]))))/((p-2)*(p-1))
            }
            else
            m1[j,k]_y1[j,k]
      }

##calculate m13
m13_m1
for(j in 1:p)
      for(k in 1:p)
      {
            if(m13[j,k]==0)
                  m13[j,k]_m3[j,k]
                  else
                        m13[j,k]_m1[j,k]
      }

# Determine sum treatment rblock and cblock
mt13_array(0,dim=c(p,1))
mrb13_array(0,dim=c(p,1))
```

```
mcb13_array(0,dim=c(p,1))

for(j in 1:p)
    for(k in 1:p)
    {
        ### sum trt
        {
            if((j-k)==0)
                mt13[1]_mt13[1]+m13[j,k]
                else
                    if(((j-k)==1)||((j-k)==-2))
                        mt13[2]_mt13[2]+m13[j,k]
                        else
                            mt13[3]_mt13[3]+m13[j,k]
        }
        ### sum rblock
        {
            if(j==1)
                mrb13[1]_mrb13[1]+m13[1,k]
                else
                    if(j==2)
                        mrb13[2]_mrb13[2]+m13[2,k]
                        else
                            mrb13[3]_mrb13[3]+m13[3,k]
        }
        ### sum cblock
        {
            if(k==1)
                mcb13[1]_mcb13[1]+m13[j,1]
                else
                    if(k==2)
                        mcb13[2]_mcb13[2]+m13[j,2]
                        else
                            mcb13[3]_mcb13[3]+m13[j,3]
        }
    }

###find m2 again
for(j in 1:p)
    for(k in 1:p)
    {
        if(m13[j,k]==0)
        {
            if((j-k)==0)
            m2[j,k]_((p*((mt13[1])+(mrb13[j])+(mcb13[k])))-(2*(sum(m13[1:n]))))/((p-2)*(p-1))
            else
                if(((j-k)==1)||((j-k)==-2))
```

```
                        m2[j,k]_((p*((mt13[2])+(mrb13[j])+(mcb13[k])))-(2*(sum(m13[1:n]))))/((p-2)*(p-1))
                        else
                            m2[j,k]_((p*((mt13[3])+(mrb13[j])+(mcb13[k])))-(2*(sum(m13[1:n]))))/((p-2)*(p-1))
            }
            else
            m2[j,k]_y1[j,k]
    }
                if(reloop==30)
                break;
}
##calculate mm
mm_m12
for(j in 1:p)
    for(k in 1:p)
    {
        if(mm[j,k]==0)
            mm[j,k]_m3[j,k]
            else
                mm[j,k]_m12[j,k]
    }


}###end mi=3

########################################################

# EM Algorithm

y2_array(0,dim=c(p,p))


y2.square_array(0,dim=c(p,p))

y2_y1

y2.square_y2^2
total1_array(0,dim=c(1,1))

total2_array(0,dim=c(1,1))


for(j in 1:p)
    for(k in 1:p)
    {
        total1_total1+y2[j,k]
        total2_total2+y2.square[j,k]
    }
```

```
u0_total1/(p^2)
sd0.square_(total2/(p^2))-u0^2

Expect.u_total1+(mi*u0)
Expect.var_total2+(mi*((u0^2)+sd0.square))

u1_Expect.u/(p^2)
var.em1_(Expect.var/(p^2))-u1^2

reloop_0
repeat

{

reloop_reloop+1

Expect.u_total1+(mi*u1)
u1_Expect.u/(p^2)

Expect.var_total2+(mi*((u1^2)+var.em1))
var.em1_(Expect.var/(p^2))-u1^2

if(reloop==30)
break;
}
y.em1_rnorm(500,u1,sqrt(var.em1))
a_round(mean(y.em1),digit=4)

for(j in 1:p)
    for(k in 1:p)
    {
        if(y2[j,k]==0)
            y2[j,k]_a
    }

############ MI############(p=3,4)

y3_array(,dim=c(p,p))
y3_y2

x_array(,dim=c(n,n-2))
if(p==3)
{
for(j in 1:n)
    for(k in 1:(n-2))
    {
```

```
if(k==1)
{
      x[j,k]_1
}
if(k==2)
{
      if((j==1)||(j==5)||(j==9))
      x[j,k]_1
      else
            if((j==2)||(j==6)||(j==7))
            x[j,k]_0
            else
                  x[j,k]_-1
}
if(k==3)
{
      if((j==2)||(j==6)||(j==7))
      x[j,k]_1
      else
            if((j==1)||(j==5)||(j==9))
            x[j,k]_0
            else
                  x[j,k]_-1
}
if(k==4)
{
      if((j==1)||(j==4)||(j==6))
      x[j,k]_1
      else
            if((j==2)||(j==5)||(j==8))
            x[j,k]_0
            else
                  x[j,k]_-1
}
if(k==5)
{
      if((j==2)||(j==5)||(j==8))
      x[j,k]_1
      else
            if((j==1)||(j==4)||(j==6))
            x[j,k]_0
            else
                  x[j,k]_-1
}
if(k==6)
{
      if((j==1)||(j==2)||(j==3))
```

```
                x[j,k]_1
                else
                        if((j==4)||(j==5)||(j==7))
                        x[j,k]_0
                        else
                                x[j,k]_-1
        }
        if(k==7)
        {
                if((j==4)||(j==5)||(j==7))
                x[j,k]_1
                else
                        if((j==1)||(j==2)||(j==3))
                        x[j,k]_0
                        else
                                x[j,k]_-1
        }
    }
}##end p=3

if(p==4)
{
for(j in 1:n)
    for(k in 1:(p*3+1))
    {
        if(k==1)
        {
                x[j,k]_1
        }
        if(k==2)
        {
                if((j==1)||(j==6)||(j==11)||(j==16))
                x[j,k]_1
                else
                        x[j,k]_0
        }
        if(k==3)
        {
                if((j==2)||(j==7)||(j==12)||(j==13))
                x[j,k]_1
                else
                        x[j,k]_0
```

```
}
if(k==4)
{
    if((j==3)||(j==8)||(j==9)||(j==14))
    x[j,k]_1
    else
        x[j,k]_0
}
if(k==5)
{
    if((j==4)||(j==5)||(j==10)||(j==15))
    x[j,k]_1
    else
        x[j,k]_0
}
if(k==6)
{
    if((j==1)||(j==5)||(j==9)||(j==13))
    x[j,k]_1
    else
        x[j,k]_0
}
if(k==7)
{
    if((j==2)||(j==6)||(j==10)||(j==14))
    x[j,k]_1
    else
        x[j,k]_0
}
if(k==8)
{
    if((j==3)||(j==7)||(j==11)||(j==15))
    x[j,k]_1
    else
        x[j,k]_0
}
if(k==9)
{
```

```
            if((j==4)||(j==8)||(j==12)||(j==16))
            x[j,k]_1
            else
                x[j,k]_0
        }
        if(k==10)
        {
            if((j==1)||(j==2)||(j==3)||(j==4))
            x[j,k]_1
            else
                x[j,k]_0
        }
        if(k==11)
        {
            if((j==5)||(j==6)||(j==7)||(j==8))
            x[j,k]_1
            else
                x[j,k]_0
        }
        if(k==12)
        {
            if((j==9)||(j==10)||(j==11)||(j==12))
            x[j,k]_1
            else
                x[j,k]_0
        }
        if(k==13)
        {
            if((j==13)||(j==14)||(j==15)||(j==16))
            x[j,k]_1
            else
                x[j,k]_0
        }
    }
}##end p=4
```

```
xt_t(x)
xtx_xt%*%x
inv.xtx_ginverse(xtx)

reloop_0
repeat
{

reloop_reloop+1


###arrang y3###

w_array(0,dim=c(n,1))
for(i in 1:n)
{
	w[i]_y3[i]
}

xty_xt%*%w
para_inv.xtx%*%xty
if(p==3)
{
p.t3_0
p.rb3_0
p.cb3_0
p.mu_para[1]
p.t1_para[2]
p.t2_para[3]
p.t3_-(para[2]+para[3])
p.rb1_para[4]
p.rb2_para[5]
p.rb3_-(para[4]+para[5])
p.cb1_para[6]
p.cb2_para[7]
p.cb3_-(para[6]+para[7])
}
if(p==4)
{
p.mu_para[1]
p.t1_para[2]
p.t2_para[3]
p.t3_para[4]
p.t4_para[5]
```

```
p.rb1_para[6]
p.rb2_para[7]
p.rb3_para[8]
p.rb4_para[9]
p.cb1_para[10]
p.cb2_para[11]
p.cb3_para[12]
p.cb4_para[13]
}
w1_array(0,dim=c(n,1))
if(p==3)
{
w1[1]_p.mu+p.t1+p.rb1+p.cb1
w1[2]_p.mu+p.t2+p.rb2+p.cb1
w1[3]_p.mu+p.t3+p.rb3+p.cb1
w1[4]_p.mu+p.t3+p.rb1+p.cb2
w1[5]_p.mu+p.t1+p.rb2+p.cb2
w1[6]_p.mu+p.t2+p.rb1+p.cb3
w1[7]_p.mu+p.t2+p.rb3+p.cb2
w1[8]_p.mu+p.t3+p.rb2+p.cb3
w1[9]_p.mu+p.t1+p.rb3+p.cb3
}

if(p==4)
{
w1[1]_p.mu+p.t1+p.rb1+p.cb1
w1[2]_p.mu+p.t2+p.rb2+p.cb1
w1[3]_p.mu+p.t3+p.rb3+p.cb1
w1[4]_p.mu+p.t4+p.rb4+p.cb1
w1[5]_p.mu+p.t4+p.rb1+p.cb2
w1[6]_p.mu+p.t1+p.rb2+p.cb2
w1[7]_p.mu+p.t2+p.rb3+p.cb2
w1[8]_p.mu+p.t3+p.rb4+p.cb2
w1[9]_p.mu+p.t3+p.rb1+p.cb3
w1[10]_p.mu+p.t4+p.rb2+p.cb3
w1[11]_p.mu+p.t1+p.rb3+p.cb3
w1[12]_p.mu+p.t2+p.rb4+p.cb3
w1[13]_p.mu+p.t2+p.rb1+p.cb4
w1[14]_p.mu+p.t3+p.rb2+p.cb4
w1[15]_p.mu+p.t4+p.rb3+p.cb4
w1[16]_p.mu+p.t1+p.rb4+p.cb4
```

```
}

er_w-w1
ert_t(er)
s0_er%*%ert

#### separate variance

s11_array(,dim=c(m,m))
for(j in 1:m)
    for(k in 1:m)
    {
        s11[j,k]_s0[j,k]
    }

s12_array(,dim=c(m,mi))
for(j in 1:m)
    for(k in 1:mi)
    {
        s12[j,k]_s0[j,(k+m)]
    }

s21_array(,dim=c(mi,m))
for(j in 1:mi)
    for(k in 1:m)
    {
        s21[j,k]_s0[(j+m),k]
    }

s22_array(,dim=c(mi,mi))
for(j in 1:mi)
    for(k in 1:mi)
    {
        s22[j,k]_s0[j+m,k+m]
    }

inv.s11_ginverse(s11)
s21s11_(s21)%*%(inv.s11)

if(p==3)
{
if(mi==1)
{
    u2_array(mean(y3[9]),dim=c(mi,1))
    u1_array(mean(y3[c(1:8)]),dim=c(n-mi,1))

    y4_array(0,dim=c(m,1))
```

```
        for(i in 1:m)
        {
              y4[i]_y3[i]
        }

}
if(mi==2)
{
        u2_array(mean(y3[c(8,9)]),dim=c(mi,1))
        u1_array(mean(y3[c(1:7)]),dim=c(n-mi,1))

        y4_array(0,dim=c(m,1))
        for(i in 1:m)
        {
              y4[i]_y3[i]
        }
}

if(mi==3)
{
        u2_array(mean(y3[c(6,8,9)]),dim=c(mi,1))
        u1_array(mean(y3[c((1:5),7)]),dim=c(n-mi,1))

        y4_array(0,dim=c(m,1))
        for(i in 1:m)
        {
              if(i!=m)
              y4[i]_y3[i]
              else
                    y4[i]_y3[m+1]
        }
}
}

If(p==4)
{
if(mi==2)
{
        u2_array(mean(y3[c(15:16)]),dim=c(mi,1))
        u1_array(mean(y3[c(1:14)]),dim=c(n-mi,1))
}


if(mi==3)
{
        u2_array(mean(y3[c(14:16)]),dim=c(mi,1))
```

```
    u1_array(mean(y3[c(1:13)]),dim=c(n-mi,1))
}


if(mi==5)
{
    u2_array(mean(y3[c(12:16)]),dim=c(mi,1))
    u1_array(mean(y3[c(1:11)]),dim=c(n-mi,1))
}
}

u21_u2+((s21s11)%*%(y4-u1))
s22.1_round(s22-(s21s11%*%s12),digit=4)

y2.1_array(0,dim=c(mi,1))
for(i in 1:mi)
{
y2.1[i]_0
}

for(i in 1:10)
{
    s _ length(u21)
    Z _  matrix(rnorm(s*1),nrow=s)
    V _ t(chol(s22.1) )
    x _(V%*%Z)+u21
    y2.1_y2.1+x
}

y2.11_y2.1/10

if(p==3)
{
if(mi==1)
{
    y3[9]_y2.11[1]
}##mi=2

if(mi==2)
{
    y3[8]_y2.11[1]
    y3[9]_y2.11[2]

}##mi=2
```

```
if(mi==3)
{
      y3[6]_y2.11[1]
      y3[8]_y2.11[2]
      y3[9]_y2.11[3]

}##mi=3
}

if(p==4)
{
if(mi==2)
{
      y3[15]_ y2.11[1]
      y3[16]_ y2.11[2]
}##mi=2

if(mi==3)
{
      y3[14]_ y2.11[1]
      y3[15]_ y2.11[2]
      y3[16]_ y2.11[3]
}##mi=3

if(mi==5)
{
      y3[12]_ y2.11[1]
      y3[13]_ y2.11[2]
      y3[14]_ y2.11[3]
      y3[15]_ y2.11[4]
      y3[16]_ y2.11[5]
}##mi=5

}
###arrang y3 (step2)###
for(i in 1:n)
{
      w[i]_y3[i]
}
```

```
xty_xt%*%w
para_inv.xtx%*%xty
if(p==3)
{
p.t3_0
p.rb3_0
p.cb3_0
p.mu_para[1]
p.t1_para[2]
p.t2_para[3]
p.t3_-(para[2]+para[3])
p.rb1_para[4]
p.rb2_para[5]
p.rb3_-(para[4]+para[5])
p.cb1_para[6]
p.cb2_para[7]
p.cb3_-(para[6]+para[7])
}

if(p==4)
{
p.mu_para[1]
p.t1_para[2]
p.t2_para[3]
p.t3_para[4]
p.t4_para[5]
p.rb1_para[6]
p.rb2_para[7]
p.rb3_para[8]
p.rb4_para[9]
p.cb1_para[10]
p.cb2_para[11]
p.cb3_para[12]
p.cb4_para[13]
}

for(i in 1:n)
{
    w1[i]_0
}
if(p==3)
{
w1[1]_p.mu+p.t1+p.rb1+p.cb1
w1[2]_p.mu+p.t2+p.rb2+p.cb1
```

```
w1[3]_p.mu+p.t3+p.rb3+p.cb1
w1[4]_p.mu+p.t3+p.rb1+p.cb2
w1[5]_p.mu+p.t1+p.rb2+p.cb2
w1[6]_p.mu+p.t2+p.rb1+p.cb3
w1[7]_p.mu+p.t2+p.rb3+p.cb2
w1[8]_p.mu+p.t3+p.rb2+p.cb3
w1[9]_p.mu+p.t1+p.rb3+p.cb3
}
if(p==4)
{
w1[1]_p.mu+p.t1+p.rb1+p.cb1
w1[2]_p.mu+p.t2+p.rb2+p.cb1
w1[3]_p.mu+p.t3+p.rb3+p.cb1
w1[4]_p.mu+p.t4+p.rb4+p.cb1
w1[5]_p.mu+p.t4+p.rb1+p.cb2
w1[6]_p.mu+p.t1+p.rb2+p.cb2
w1[7]_p.mu+p.t2+p.rb3+p.cb2
w1[8]_p.mu+p.t3+p.rb4+p.cb2
w1[9]_p.mu+p.t3+p.rb1+p.cb3
w1[10]_p.mu+p.t4+p.rb2+p.cb3
w1[11]_p.mu+p.t1+p.rb3+p.cb3
w1[12]_p.mu+p.t2+p.rb4+p.cb3
w1[13]_p.mu+p.t2+p.rb1+p.cb4
w1[14]_p.mu+p.t3+p.rb2+p.cb4
w1[15]_p.mu+p.t4+p.rb3+p.cb4
w1[16]_p.mu+p.t1+p.rb4+p.cb4
}

er_w-w1
ert_t(er)
s0_er%*%ert

#### Posterior Step use inverted Wishart distribution ###
my_array(mean(y3),dim=c(n,1))
invvar.y_ginverse(s0)
prec_round((n-1)*invvar.y,digit=6)

s_n
df_n-1

g_array(0,dim=c(n,n))
for(i in 1:(n^2))
{
```

```
g[i]_0
}

for(i in 1:100)
{
R _ diag(sqrt(2*rgamma(s,(df + s  - 1:s)/2)))
R[outer(1:s, 1:s,  "<")] _ rnorm (s*(s-1)/2)
S _ t(solve(R))%*% chol(prec)
w.var0_(t(S)%*%S)
g_g+w.var0
}
g1_g/100

w.var1_round((g1),digit=6)
w.var2_round((g1)/n,digit=6)

w.mu0_array(0,dim=c(n,1))
for(i in 1:n)
{
w.mu0[i]_0
}

for(i in 1:100)
{
    s1 _ length(w1)
    Z1 _  matrix(rnorm((s1)*1),nrow=(s1))
    V1 _ t(chol(w.var2) )
    x1 _((V1)%*%(Z1))+w1
    w.mu0_w.mu0+x1
}
w.mu_(w.mu0)/100

p.y0_array(0,dim=c(n,1))
for(i in 1:n)
{
p.y0[i]_0
}
for(i in 1:100)
{
    s2 _ length(w.mu)
    Z2 _  matrix(rnorm((s2)*1),nrow=(s2))
    V2 _ t(chol(w.var1) )
    x2 _((V2)%*%(Z2))+w.mu
    p.y0_p.y0+x2
}
p.y_(p.y0)/100
```

```
if(p==3)
{
if(mi==1)
{
      y3[9]_p.y[9]
}
if(mi==2)
{
      y3[8]_p.y[8]
      y3[9]_p.y[9]
}
if(mi==3)
{
      y3[6]_p.y[6]
      y3[8]_p.y[8]
      y3[9]_p.y[9]
}
}
if(p==4)
{
if(mi==2)
{
      y3[15]_p.y[15]
      y3[16]_p.y[16]
}
if(mi==3)
{
      y3[14]_p.y[14]
      y3[15]_p.y[15]
      y3[16]_p.y[16]
}
if(mi==5)
{
      y3[12]_p.y[12]
      y3[13]_p.y[13]
      y3[14]_p.y[14]
      y3[15]_p.y[15]
      y3[16]_p.y[16]
}
}
            if(reloop==40)
            break;
}
```

```
}
    MAE.OLS_array(0,dim=c(p,p))
    MAE.OLS_abs(y-mm)

    MAE.EM_array(0,dim=c(p,p))
    MAE.EM_abs(y-y2)

    MAE.MI_array(0,dim=c(p,p))
    MAE.MI_abs(y-y3)

        MMAE.OLS_max(MAE.OLS[1:n])
        MMAE.EM_max(MAE.EM[1:n])
        MMAE.MI_max(MAE.MI[1:n])

MMAE_array(0,dim=c(3,1))
MMAE[1]_MMAE.OLS
MMAE[2]_MMAE.EM
MMAE[3]_MMAE.MI

print(round(MMAE,digit=4))
}
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศุภลักษณ์ กรรณิกา เกิดเมื่อวันที่ 19 มกราคม พ.ศ. 2522 ที่จังหวัดลำปาง สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต วิชาเอกสถิติ วิชาโทเศรษฐศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ เมื่อปีการศึกษา 2543 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2547