

การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อคสมบรูณ์

นาย ประพจน์ ดำรงค์สุทธิพงศ์

ศูนย์วิทยพัทยากร

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2546

ISBN 974-17-4625-3

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR RANDOMIZED
COMPLETE BLOCK DESIGN



Mr. PRAPHOT DAMRONGSUTTIPONG

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย
A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2003

ISBN 974-17-4625-3

หัวข้อวิทยานิพนธ์

การเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการ
ทดลองแบบสุ่มในบล็อกผสมบูรณ์

โดย

นาย ประพจน์ ดำรงค์สุทธิพงษ์

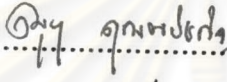
สาขาวิชา

สถิติ

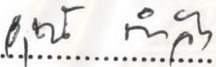
อาจารย์ที่ปรึกษา

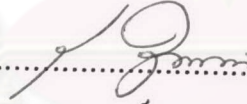
รองศาสตราจารย์ ดร. สุพล คุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้
วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต


.....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(ผู้ช่วยศาสตราจารย์ ดร. ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์


.....ประธานกรรมการ
(อาจารย์ ดร. อรุณี กำลั้ง)


.....อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. สุพล คุรงค์วัฒนา)


.....กรรมการ
(รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ)

คณบดีคณะพาณิชยศาสตร์และการบัญชี
จุฬาลงกรณ์มหาวิทยาลัย

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

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

วัตถุประสงค์ของการวิจัยครั้งนี้ เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสูญหายในการวางแผนการทดลองแบบสุ่มในบล็อกสมบูรณ์ 3 วิธีได้แก่ การประมาณค่าวิธีกำลังสองน้อยสุด การประมาณค่าวิธีค่าคาดหวังสูงสุด และการประมาณค่าวิธีอิมพิวเทชัน ซึ่งตัวแบบสำหรับแผนแบบการทดลองแบบสุ่มในบล็อกสมบูรณ์ที่ไม่มีการทำซ้ำเป็นดังนี้

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

เมื่อ Y_{ij} แทนค่าสังเกตจากหน่วยทดลองที่ได้รับวิธีการทดลองที่ i และอยู่ใน บล็อกที่ j μ แทนค่าเฉลี่ยรวม τ_i แทนอิทธิพลของวิธีการทดลองที่ i โดยที่ τ_i มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน σ_{τ}^2 β_j แทนอิทธิพลของบล็อกที่ j โดยที่ β_j มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน σ_{β}^2 ε_{ij} แทนความคลาดเคลื่อนสุ่มของค่าสังเกตจากหน่วยทดลองที่ได้รับวิธีการทดลองที่ i และอยู่ในบล็อกที่ j โดยที่ ε_{ij} มีการแจกแจงอิสระจากกันซึ่งมีค่าเฉลี่ย 0 ความแปรปรวน σ_{ε}^2 τ_i , β_j และ ε_{ij} เป็นอิสระซึ่งกันและกัน, t แทนจำนวนวิธีการทดลอง b แทนจำนวนบล็อกในการทดลอง

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

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

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2546

ลายมือชื่อนิติศ. ประพจน์ คำรงค์สุทธิพงษ์

ลายมือชื่ออาจารย์ที่ปรึกษา

4482297026 : MAJOR STATISTICS

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

PRAPHOT DAMRONGSUTTIPONG : A COMPARISON OF MISSING VALUE ESTIMATION METHODS FOR RANDOMIZED COMPLETE BLOCK DESIGN.

THESIS ADVISOR : ASSOCIATE PROFESSOR . SUPOL DURONGWATANA, Ph.D.

163pp. ISBN 974-17-4625-3

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

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

when Y_{ij} is observation for the i^{th} level of treatment and j^{th} level of block; μ is the grand mean; τ_i is the i^{th} random effect of treatment factor where τ_i is independently distributed with mean 0 and variance σ_τ^2 ; β_j is the j^{th} random effect of blocking factor where β_j is independently distributed with mean 0 and variance σ_β^2 ; ε_{ij} is the random error for the observed data at the i^{th} level of treatment factor, the j^{th} level of blocking factor where ε_{ij} is independently distributed with mean 0 and variance σ_ε^2 ; τ_i , β_j and ε_{ij} are independent, t is number of levels for treatment factor and b is number of levels for blocking factor.

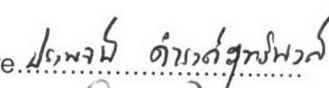
Monte Carlo simulation is used for data generated through S-plus 2000 code, The random error term in the model is generated with normal distribution. The number of levels for treatment factor is specified at $t = 3$, $t = 4$ and $t = 5$; the number of levels for blocking factor is specified at $b = 2$,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

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

Department Statistics

Field of study Statistics

Academic year 2003

Student's signature.....

Advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของรองศาสตราจารย์ ดร.สุพล คุงศ์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ คำปรึกษา ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จลุล่วงไปด้วยดี

ผู้วิจัยขอกราบขอบพระคุณ อ.ดร. อรุณี กำลิ่ง รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ ในฐานะประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาตรวจแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น ทั้งนี้ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติ ที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ท้ายนี้ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งให้การสนับสนุน และเพื่อน ๆ ที่ให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	2
1.3 สมมติฐานการวิจัย.....	2
1.4 ขีดตกตั้งเบื้องต้น.....	2
1.5 ขอบเขตของการวิจัย.....	3
1.6 เกณฑ์ที่ใช้ในการตัดสินใจ.....	4
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	4
2. ระเบียบวิธีการวิจัย.....	5
2.1 การประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด (least square method).....	5
2.2 วิธีประมาณค่าสูญหายโดยวิธี EM algorithm (Expectation Maximization).....	11
2.3 วิธีประมาณค่าสูญหายโดยวิธี Imputation (Imputation Method).....	12
2.4 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณ.....	13
3. วิธีดำเนินการวิจัย.....	14
3.1 การสร้างรูปแบบการแจกแจงของประชากรแบบปกติ.....	14
3.2 สร้างข้อมูลให้เกิดการสูญหาย.....	15
3.3 การคำนวณค่าที่สูญหาย.....	15
3.3.1 วิธีประมาณค่าสูญหายโดยวิธีกำลังสองน้อยสุด.....	15
(least square method)	
3.3.2 วิธีประมาณค่าสูญหายโดยวิธี EM algorithm.....	16
(Expectation Maximization)	

สารบัญ(ต่อ)

	หน้า
3.3.3 วิธีประมาณค่าสูญหายโดยวิธี Imputation (Imputation Method).....	17
3.4 ขั้นตอนในการทำงานของโปรแกรม.....	19
4. ผลการวิเคราะห์ข้อมูล.....	21
4.1 การเปรียบเทียบค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนทั้ง 3 วิธี ณ เปอร์เซนต์.....	21
ข้อมูลสูญหาย เมื่อกำหนดให้ค่าคงที่ h และสัมประสิทธิ์การแปรผันคงที่	
4.2 การเปรียบเทียบค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนทั้ง 3 วิธี ณ สัมประสิทธิ์..	59
การแปรผัน เมื่อกำหนดให้ค่าคงที่ h และ เปอร์เซนต์ข้อมูลสูญหายคงที่	
4.3 การเปรียบเทียบค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนทั้ง 3 วิธี ณ ค่าคงที่ h.....	96
เมื่อกำหนดให้เปอร์เซนต์ข้อมูลสูญหาย และสัมประสิทธิ์การแปรผันคงที่	
4.4 การวิเคราะห์ห่อทธิพลร่วมระหว่างปัจจัย.....	133
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	147
5.1 สรุปผลการวิจัย.....	147
5.2 ข้อเสนอแนะ.....	149
รายการอ้างอิง.....	150
ภาคผนวก.....	151
ประวัติผู้เขียนวิทยานิพนธ์.....	163

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ(ต่อ)

ภาพประกอบ

หน้า

รูปที่ 4.243 แสดงการเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองทั้ง 3 วิธี ณ ค่าคงที่ h ต่าง ๆ
 เมื่อเปอร์เซ็นต์ข้อมูลสูญหายมีค่า 30% วิธีการทดลอง 5 วิธีการทดลองจำนวนบิตออก 6 บิตออก....
 และสัมประสิทธิ์การแปรผัน 0.45

131



ศูนย์วิทยทรัพยากร
 จุฬาลงกรณ์มหาวิทยาลัย