

การประมาณค่าองค์ประกอบความแปรปรวน
ของแผนแบบการทดลองจุดสุ่มละตินด้วยวิธีการเฉลี่ยตัวแบบ



นางสาวศิริวัลย์ จันทบุตร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2543

ISBN 974-346-894-3

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AN ESTIMATION OF VARIANCE COMPONENTS
FOR LATIN SQUARE DESIGN BY THE MODEL AVERAGING METHOD

Miss Sirivalai Juntabut

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 2000

ISBN 974-346-894-3

ศิริวิทย์ จันทบุตร :การประมาณค่าองค์ประกอบความแปรปรวนของแผนแบบการทดลองจตุรัสละตินด้วยวิธีการเฉลี่ยตัวแบบ (AN ESTIMATION OF VARIANCE COMPONENTS FOR THE DESIGN LATIN SQUARE BY THE MODEL AVERAGING METHOD) อ.ที่ปรึกษา: รศ. ดร. สุพล คุรงค์วัฒนา, 85หน้า. ISBN 974-346-894-3.

วัตถุประสงค์ของการวิจัย เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าองค์ประกอบความแปรปรวนสำหรับตัวแบบจตุรัสละติน 2 วิธี ได้แก่ การประมาณค่าตัวแบบเฉลี่ย (Model Averaging Estimation) และการประมาณค่าวิธีคลาสสิก (Classical Estimation) หรือตัวแบบเต็มรูป การประมาณค่าวิธีคลาสสิกนั้นองค์ประกอบความแปรปรวนทุกตัวถูกประมาณโดยตรงจากตัวแบบเต็มรูป ในขณะที่การประมาณค่าวิธีตัวแบบเฉลี่ยขององค์ประกอบความแปรปรวนเหล่านั้นเกิดจากการลดรูปพารามิเตอร์ของตัวแบบเต็มรูปทีละตัวเพื่อให้ได้ตัวแบบที่เป็นไปได้ทั้งหมด จากนั้นทำการประมาณค่าองค์ประกอบความแปรปรวนของแต่ละปัจจัยตามตัวแบบที่ได้ แล้วนำค่าประมาณขององค์ประกอบความแปรปรวนตามแหล่งที่มาของความแปรปรวนเดียวกันมาทำการเฉลี่ยเพื่อให้ได้องค์ประกอบความแปรปรวนทั้ง 4 องค์ประกอบ

ซึ่งตัวแบบเต็มรูปสำหรับแผนแบบการทดลองจตุรัสละตินที่ไม่มีการทำซ้ำ โดยที่ปัจจัยแบ่งบล็อกตามแถว (Row Factor) และปัจจัยแบ่งบล็อกตามสดมภ์ (Column Factor) เป็นปัจจัยคู่ที่ 2 ปัจจัย มีรูปแบบเป็นดังนี้

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

Y_{ijk} คือ ค่าสังเกตระดับที่ i ของปัจจัยวิธีการทดลอง (Treatment Factor) ระดับที่ j ของปัจจัยแบ่งบล็อกตามแถวและระดับที่ k ของปัจจัยแบ่งบล็อกตามสดมภ์ μ คือ ค่าเฉลี่ยรวม (Grand mean) τ_i คือ ผลกระทบระดับที่ i ของปัจจัยวิธีการทดลองโดยที่ $\tau_i \sim N(0, \sigma_\tau^2)$ α_j คือ ผลกระทบระดับที่ j ของปัจจัยแบ่งบล็อกตามแถวโดยที่ $\alpha_j \sim N(0, \sigma_\alpha^2)$ β_k คือ ผลกระทบระดับที่ k ของปัจจัยแบ่งบล็อกตามสดมภ์โดยที่ $\beta_k \sim N(0, \sigma_\beta^2)$ ε_{ijk} คือ ความคลาดเคลื่อนระดับที่ i ของปัจจัยวิธีการทดลอง ระดับที่ j ของปัจจัยแบ่งบล็อกตามแถว และระดับที่ k ของปัจจัยแบ่งบล็อกตามสดมภ์ และ $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ p เป็นจำนวนระดับของปัจจัยวิธีการทดลอง เท่ากับจำนวนระดับของปัจจัยแบ่งบล็อกทั้งสองปัจจัย โดยเรียกพารามิเตอร์ $\sigma_\tau^2, \sigma_\alpha^2, \sigma_\beta^2$, และ σ_ε^2 ว่าองค์ประกอบความแปรปรวน

ในการวิจัยครั้งนี้ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล โดยทำการทดลองซ้ำ ๆ ด้วยโปรแกรม S-plus 2000 การเปรียบเทียบกระทำภายใต้สถานการณ์ต่าง ๆ ของจำนวนระดับปัจจัยวิธีการทดลอง เท่ากับจำนวนระดับปัจจัยแบ่งบล็อกทั้งสอง (p) โดยที่กำหนดสถานการณ์ให้ $p=3, p=4$ และ $p=5$ ตามลำดับ สำหรับการจำลองสถานการณ์กำหนดให้สัมประสิทธิ์การแปรผัน (Coefficient of Variation : C.V.) เป็น 5%, 15%, 25%, 35%, 45% ถึง 55 % หลักเกณฑ์ที่นำมาใช้ในการเปรียบเทียบวิธีการประมาณทั้ง 2 วิธี คือ ระยะทางยุคคิดเฉลี่ย

ผลการวิจัยสรุปได้ว่า วิธีการประมาณค่าแบบจุดขององค์ประกอบความแปรปรวนวิธีตัวแบบเฉลี่ยให้ค่าระยะทางยุคคิดเฉลี่ยต่ำกว่าการประมาณค่าวิธีคลาสสิกในทุกสถานการณ์ของการทดลองที่ทำการศึกษา

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2543

ลายมือชื่อนิสิต...ศิริวิทย์...จันทบุตร.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....



4182385326 : MAJOR STATISTICS

KEY WORD: Variance Components / Averaging Model Estimation / Latin Square

SIRIVALAI JUNTABUT : ESTIMATION OF VARIANCE COMPONENTS FOR THE

LATIN SQUARE DESIGN BY THE MODEL AVERAGING METHOD. THESIS ADVISOR :

ASSOCIATE PROFESSOR SUPOL DURONGWATANA, Ph. D. 85 pp. ISBN 974-346-894-3.

The objective of this study is to compare two methods of variance component estimation for the Latin Square design; the model averaging method and the classical method. The classical method estimates all variance components directly by the full model while the model averaging method estimates those variance components using all possible reduced models and then averaging all of those estimates. The full model for the Latin Square design is as follows:

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

Y_{ijk} is an observation data for the i^{th} level of treatment factor, the j^{th} level of row blocking factor, and the k^{th} column blocking factor; μ is the grand mean; τ_i is the i^{th} random effect of treatment factor and τ_i is independently and normally distributed with mean 0 and variance σ_τ^2 ; α_j is the j^{th} random effect of row blocking factor and α_j is independently and normally distributed with mean 0 and variance σ_α^2 ; β_k is the k^{th} random effect of column blocking effect and β_k is also independently and normally distributed with mean 0 and variance σ_β^2 ; ε_{ijk} is the random error for the observed data at the i^{th} level of treatment factor, the j^{th} level of row blocking factor, the k^{th} level of column blocking factor, and ε_{ijk} is independently and normally distributed with mean 0 and variance σ_ε^2 ; p is the number of levels for treatment factor, row blocking factor and column blocking factor. The parameters; $\sigma_\tau^2, \sigma_\alpha^2, \sigma_\beta^2$, and σ_ε^2 , are variance components for the model.

Monte Carlo simulation is done through S-plus 2000 code. It is simulated under several situations due to the number of levels for treatment factor, row blocking factor and column blocking factor. In this study, the simulation is specified at $p = 3$, $p = 4$ and $p = 5$. In addition, the coefficient of variation (CV) for the observed data is varied from 5%, 15%, 25%, 35%, 45% to 55%. The average of Euclidean distance between the vector of true values and the estimation vector of variance is a measure for comparison between both methods.

The results for the study show that point estimates for variance components in the Latin Square design model using the model averaging method regardless the number of levels for treatment factor, the number of levels for row blocking factor and the number of levels for column blocking factor, and the coefficient of variation for the observed data, provide shorter averaged Euclidean distance than the ones from the classical method in all simulated situations.

Department Statistics

Field of study Statistics

Academic year 2000

Student's signature.....*Sirivalai Juntabut*.....

Advisor's signature.....*Supol Durongwatana*.....

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้ไม่อาจสำเร็จลุล่วงได้ถ้าขาดความช่วยเหลือจากทุกฝ่าย ผู้วิจัยรู้สึกซาบซึ้งใจในความกรุณาของรองศาสตราจารย์ ดร.สุพล คุรงค์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างยิ่งในการให้คำแนะนำปรึกษาแก่ผู้วิจัย ตลอดจนตรวจแก้ไขข้อบกพร่องต่าง ๆ ในการวิจัยครั้งนี้จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณอาจารย์เป็นอย่างสูง ณ ที่นี้

ผู้วิจัยขอขอบคุณ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัยที่มีส่วนช่วยเหลือผู้วิจัยในการจัดสรรเงินทุนบางส่วนอุดหนุนการวิจัย

ท้ายที่สุดผู้วิจัยขอขอบคุณงามความดีให้กับคุณแม่ที่คอยให้การสนับสนุนในด้านการศึกษาย่างดีมาโดยตลอด รวมทั้งเพื่อน ๆ ผู้ที่คอยให้กำลังใจและรอคอยความดีของผู้อื่นเสมอมา ตลอดจนครูอาจารย์ทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้แก่ผู้วิจัยตั้งแต่ต้นจนสำเร็จการศึกษา

ศิริวัลย์ จันทบุตร

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ฉ
บทที่	
1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	2
1.3 สมมติฐานของการวิจัย.....	2
1.4 ขีดกลางเบื้องต้น.....	3
1.5 ข้อยกเว้นของการวิจัย.....	6
1.6 คำจำกัดความที่ใช้ในงานวิจัย.....	8
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	8
2 ระเบียบวิธีการวิจัย.....	9
2.1 การประมาณค่าองค์ประกอบความแปรปรวนวิธีตัวแบบเต็มรูป.....	9
2.2 การประมาณค่าองค์ประกอบความแปรปรวนวิธีตัวแบบเฉลี่ย.....	
2.2.1 ตัวแบบที่เป็นไปได้ทั้งหมด.....	11
2.2.2 การประมาณค่าแบบจุดขององค์ประกอบความแปรปรวน.....	12
2.2.3 การถ่วงน้ำหนักค่าแบบจุดขององค์ประกอบความแปรปรวน.....	16
2.2.4 การคำนวณค่าแบบจุดขององค์ประกอบความแปรปรวนเฉลี่ย.....	17
2.3 เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณทั้ง 2 วิธี.....	18
3 วิธีดำเนินการวิจัย.....	22
3.1 การสร้างรูปแบบการแจกแจงของประชากรแบบปกติ.....	22
3.2 การคำนวณค่าประมาณแบบจุดขององค์ประกอบความแปรปรวน.....	22

สารบัญ (ต่อ)

	หน้า
3.2.1 วิธีตัวแบบเต็มรูป.....	22
3.2.2 วิธีตัวแบบเฉลี่ย.....	23
3.3 ขั้นตอนในการทำงานของโปรแกรม.....	25
4 ผลการวิเคราะห์ข้อมูล.....	27
4.1 ผลจากการเปรียบเทียบค่าระยะทางยุคลิดทั้งสองวิธี.....	27
4.1.1 การเปรียบเทียบค่าระยะทางยุคลิดเฉลี่ย ณ สัมประสิทธิ์การแปรผัน... ต่าง ๆ เมื่อกำหนดให้ค่าคงที่ m และจำนวนระดับปัจจัยคงที่	27
4.1.2 การเปรียบเทียบค่าระยะทางยุคลิดเฉลี่ย ณ จำนวนระดับปัจจัยต่าง ๆ ... เมื่อกำหนดให้สัมประสิทธิ์การแปรผัน และค่าคงที่ m คงที่	41
4.1.3 การเปรียบเทียบค่าระยะทางยุคลิดเฉลี่ย ณ จำนวนระดับค่าคงที่ m เมื่อกำหนดให้สัมประสิทธิ์การแปรผันและจำนวนระดับปัจจัยคงที่	60
5 สรุปผลการวิจัยและข้อเสนอแนะ.....	73
5.1 สรุปผลการวิจัย.....	73
5.2 ข้อเสนอแนะ.....	74
รายการอ้างอิง.....	75
ภาคผนวก.....	76
ประวัติผู้เขียน.....	85

สารบัญตาราง

ตาราง		หน้า
ตารางที่ 1	แสดงการวิเคราะห์ความแปรปรวนสำหรับแผนแบบการทดลองจตุรัสละติน....	5
ตารางที่ 2	แสดงการเฉลี่ยของค่าประกอบความแปรปรวน.....	17
ตารางที่ 4.1	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 0.5$	28
ตารางที่ 4.2	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 1.0$	29
ตารางที่ 4.3	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 1.5$	30
ตารางที่ 4.4	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 2.0$	31
ตารางที่ 4.5	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 2.5$	32
ตารางที่ 4.6	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ สัมประสิทธิ์การแปรผันต่างๆ เมื่อค่าคงที่ $m = 3.0$	33
ตารางที่ 4.7	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 0.5$	41
ตารางที่ 4.8	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 1.0$	42
ตารางที่ 4.9	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 1.5$	43
ตารางที่ 4.10	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 2.0$	44
ตารางที่ 4.11	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 2.5$	45
ตารางที่ 4.12	แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี..... ณ จำนวนระดับปัจจัยต่างๆ เมื่อค่าคงที่ $m = 3.0$	46

สารบัญตาราง (ต่อ)

ตาราง	หน้า
ตารางที่ 4.13 แสดงการเปรียบเทียบค่าระยะทางซุกติคเฉลี่ยทั้ง 2 วิธี..... ณ ระดับค่าคงที่ m ต่างๆ เมื่อจำนวนระดับปัจจัย $p = 3$	60
ตารางที่ 4.14 แสดงการเปรียบเทียบค่าระยะทางซุกติคเฉลี่ยทั้ง 2 วิธี..... ณ ระดับค่าคงที่ m ต่างๆ เมื่อจำนวนระดับปัจจัย $p = 4$	61
ตารางที่ 4.15 แสดงการเปรียบเทียบค่าระยะทางซุกติคเฉลี่ยทั้ง 2 วิธี..... ณ ระดับค่าคงที่ m ต่างๆ เมื่อจำนวนระดับปัจจัย $p = 5$	62

สารบัญญภาพ (ต่อ)

ภาพประกอบ	หน้า
รูปที่ 4.65 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 4$ เมื่อสัมประสิทธิ์การแปรผันคือ 45%	69
รูปที่ 4.66 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 4$ เมื่อสัมประสิทธิ์การแปรผันคือ 55%	69
รูปที่ 4.67 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 5%	70
รูปที่ 4.68 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 15%	70
รูปที่ 4.69 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 25%	70
รูปที่ 4.70 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 35%	71
รูปที่ 4.71 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 45%	71
รูปที่ 4.72 แสดงการเปรียบเทียบค่าระยะทางยูคลิดเฉลี่ยทั้ง 2 วิธี ณ ค่าคงที่ต่าง ๆ..... เมื่อจำนวนระดับปัจจัยคือ $p = 5$ เมื่อสัมประสิทธิ์การแปรผันคือ 55%	71