

การเปรียบเทียบตัวประมาณอัตราความผิดพลาดในการวิเคราะห์การจำแนกกลุ่ม



นางสาว ลลนา ทวีรุ่งโรจน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2534

ISBN 974-578-559-8

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

017445j1๗๘๖๙๑๐

A COMPARISON ON ERROR RATE ESTIMATORS IN DISCRIMINANT ANALYSIS

Miss Lalana Thaveerungroj

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Statistics

Graduate School

Chulalongkorn University

1991

ISBN 974-578-559-8

หัวข้อวิทยานิพนธ์ การเปรียบเทียบตัวประมาณอัตราความผิดพลาดในการวิเคราะห์  
การจำแนกกลุ่ม

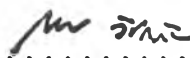
โดย นางสาว ลลนา ทวีรุ่งโรจน์

ภาควิชา สถิติ


อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. สรชัย นิคาลบุตร

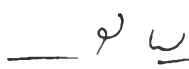


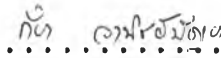
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วนหนึ่ง  
ของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต


  
..... คณะบดีบัณฑิตวิทยาลัย  
(ศาสตราจารย์ ดร. ถาวร วัชรภักย์)

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร. ชีระพร วิระถาวร)

  
..... กรรมการ  
(รองศาสตราจารย์ ดร. สรชัย นิคาลบุตร)

  
..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร. กัญญา วานิชย์บัญชา)

  
..... กรรมการ  
(รองศาสตราจารย์ นพรัตน์ รุ่งอุทัยศิริ)

ผลงาน ทวีรุ่งโรจน์ : การเปรียบเทียบตัวประมาณอัตราความผิดพลาดในการวิเคราะห์การจำแนกกลุ่ม (A COMPARISON ON ERROR RATE ESTIMATOR IN DISCRIMINANT ANALYSIS)

อ. ที่ปรึกษา รศ.ดร.สรชัย ทิคำลบุตร , 128 หน้า

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะศึกษา และนำเอาตัวประมาณ 4 วิธี ดังนี้

1. วิธี DS หรือ SHRUNKEN-D ESTIMATOR 2. วิธี R หรือ RESUBSTITUTION ESTIMATOR 3. วิธี U หรือ Leave-one-out ESTIMATOR 4. วิธี B หรือ BOOTSTRAP ESTIMATOR มาประมาณอัตราความผิดพลาดที่มีเงื่อนไขที่เกิดจากการใช้กฎการจำแนกกลุ่ม สำหรับการเปรียบเทียบประสิทธิภาพของตัวประมาณนั้นจะพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองด้วยโปรแกรมคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล เมื่อขนาดตัวอย่างมีขนาด 10, 20, 25, 50 และ 100 ขนาดของตัวแปรอิสระ (k) 3, 5, 7 และ 9 ลักษณะการแยกจากกันของประชากร หรือค่ารากที่สองของ Mahalanobis distance ( $\Delta$ ) มีขนาด 0.0, 0.5, 1.0, 1.5, 2.0, 2.5 และ 3.0 โดยในแต่ละการทดลองกระทำซ้ำ 500 ครั้ง และการสุ่มตัวอย่างแบบใส่คืนโดยวิธี Bootstrap กระทำ 100 ครั้ง ผลการศึกษาสรุปได้ดังนี้

1. ขนาดตัวอย่าง จำนวนตัวแปรอิสระ และลักษณะการแยกจากกันของประชากรมีผลต่อความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ
2. ในกรณีที่ขนาดตัวแปรอิสระในค่าสังเกตเพิ่มขึ้น วิธี U มีประสิทธิภาพดีกว่าวิธี R วิธี DS และวิธี B
3. โดยส่วนใหญ่วิธี R ให้ประสิทธิภาพในการประมาณต่ำกว่าตัวประมาณตัวอื่น
4. ค่ารากที่สองของ Mahalanobis distance เป็นปัจจัยหลักที่สำคัญในการเลือกใช้ตัวประมาณอัตราความผิดพลาดที่มีเงื่อนไข
5. เมื่อ  $\Delta \leq 0.5$  ควรเลือกใช้ตัวประมาณโดยวิธี B
6. เมื่อ  $\Delta > 0.5$  ถ้า  $k \leq 7$  ควรใช้ตัวประมาณโดยวิธี DS แต่ถ้า  $k > 7$  ควรใช้ตัวประมาณโดยวิธี U



ภาควิชา ..... สถิต  
สาขาวิชา ..... สถิต  
ปีการศึกษา ..... 2532

ลายมือชื่อนิติ .....

ลายมือชื่ออาจารย์ที่ปรึกษา .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

LALANA THAVEERUNGROJ : A COMPARISON ON ERROR RATE ESTIMATORS IN DISCRIMINANT ANALYSIS. THESIS ADVISOR : ASSO PROF. SORACHAI BHISALBUTRA, Ph.D. , 128 pp.

The purpose of this study is to investigate the four estimation methods in Discriminant Analysis namely, 1. Shrunken-D' estimator or DS method 2. Leave-one-out estimator or U method 3. Resubstitution estimator or R method 4. Bootstrap estimator or B method, using mean square errors as means of comparison.

The data for each experiment were obtained through simulation using the Monte Carlo technique. The computer program was designed to calculate the mean square error for each methods in different situations with varying sample size 10, 20, 25, 50, 100; number of independent variables (k) 3, 5, 7, 9 and square root of Mahalanobis distance ( $\Delta$ ) 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0. For each case the experiment was repeated 500 times and bootstrap sampling was repeated 100 times.

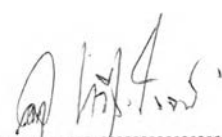
Results of the study are as follows :

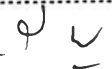
1. Sample size number of independent variables and square root of Mahalanobis distance had affected to mean square errors.
2. When number of independent variable was increased, U method is better than R method DS method and B method.
3. For almost cases, R method gave low efficiency.
4. The importance factor, which estimators should estimate the conditional error rate, is square root of Mahalanobis distance.
5. When  $\Delta \leq 0.5$  , B method should be used to estimate the conditional error rate.
6. When  $\Delta > 0.5$  , if  $k \leq 7$  then DS method should be used to estimate the conditional error rate, otherwise using U method.

ภาควิชา ..... สอศ.....

สาขาวิชา ..... สอศ.....

ปีการศึกษา ..... 2533.....

ลายมือชื่อนิสิต .....  .....

ลายมือชื่ออาจารย์ที่ปรึกษา .....  .....

ลายมือชื่อคณาจารย์ที่ปรึกษาอื่น ..... .....



## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้โดยได้รับความช่วยเหลือตลอดทั้งคำแนะนำและ  
ตรวจแก้ไขข้อบกพร่องของการเขียนวิทยานิพนธ์จาก รองศาสตราจารย์ ดร.สรชัย  
นาคาลบุตร ผู้เขียนขอกราบขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. อธิระพร วีระถาวร  
ผู้ช่วยศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา และ รองศาสตราจารย์ นพรัตน์  
รุ่งอุทัยศิริ ในฐานะกรรมการสอบวิทยานิพนธ์ที่กรุณาอ่านและตรวจแก้วิทยานิพนธ์  
ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้นและขอขอบพระคุณอาจารย์ทุกท่านในภาควิชาสถิติที่ได้  
อบรมและสอนวิชาความรู้ต่างๆ แก่ผู้เขียน

ท้ายนี้ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งให้กำลังใจแก่ผู้วิจัย  
เสมอมาจนสำเร็จการศึกษา

ลลนา ทวีรุ่งโรจน์



## สารบัญ

	หน้า
บทคัดย่อภาษาไทย . . . . .	ง
บทคัดย่อภาษาอังกฤษ . . . . .	จ
กิตติกรรมประกาศ . . . . .	ฉ
สารบัญตาราง . . . . .	ณ
สารบัญภาพ . . . . .	ญ
บทที่	
1. บทนำ . . . . .	1
1.1 ความสำคัญและความเป็นมาของปัญหา . . . . .	1
1.2 วัตถุประสงค์ของการวิจัย . . . . .	3
1.3 สมมติฐานของการวิจัย . . . . .	3
1.4 ข้อยกเว้นเบื้องต้น . . . . .	3
1.5 ขอบเขตการวิจัย . . . . .	4
1.6 ประโยชน์ที่คาดว่าจะได้รับ . . . . .	5
1.7 ตัวแปรตามอัตราความผิดพลาด . . . . .	5
2. ระเบียบวิธีที่ใช้ในการวิจัย . . . . .	8
2.1 การวิเคราะห์การจัดกลุ่มและการจำแนกกลุ่ม . . . . .	9
2.2 อัตราความผิดพลาดที่มีเงื่อนไข . . . . .	18
2.3 การประมาณค่าอัตราความผิดพลาด . . . . .	22
3. ระเบียบวิธีวิจัย . . . . .	34
3.1 วิธีมอนติคาร์โล . . . . .	34
3.2 แผนการทดลอง . . . . .	35
3.3 ขั้นตอนในการวิจัย . . . . .	35
3.4 โปรแกรมที่ใช้ในการวิจัย . . . . .	36

สารบัญ (ต่อ)

	หน้า
4. ผลการทดลอง .....	52
4.1 การเปรียบเทียบประสิทธิภาพของตัวประมาณ โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย .....	53
4.2 ผลการเปรียบเทียบประสิทธิภาพของตัวประมาณ โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย .....	73
5. สรุปผลการวิจัยและข้อเสนอแนะ .....	105
5.1 สรุปผลการวิจัย .....	105
5.2 ข้อเสนอแนะ .....	107
บรรณานุกรม .....	110
ภาคผนวก .....	112
ประวัติผู้เขียน .....	128





## สารบัญตาราง (ต่อ)

หน้า

ตาราง 4.1.12	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 25$ , $P = 9$ . . . . .	64
ตาราง 4.1.13	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 50$ , $P = 3$ . . . . .	65
ตาราง 4.1.14	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 50$ , $P = 5$ . . . . .	66
ตาราง 4.1.15	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 50$ , $P = 7$ . . . . .	67
ตาราง 4.1.16	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 50$ , $P = 9$ . . . . .	68
ตาราง 4.1.17	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 100$ , $P = 3$ . . . . .	69
ตาราง 4.1.18	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 100$ , $P = 5$ . . . . .	70
ตาราง 4.1.19	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 100$ , $P = 7$ . . . . .	71
ตาราง 4.1.20	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ อัตราความผิดพลาด ที่ได้จาก วิธี R วิธี U วิธี B และวิธี DS เมื่อ $n_1 = n_2 = 100$ , $P = 9$ . . . . .	72

## สารบัญภาพ

	หน้า
รูปที่ 3.1 แสดงแผนผังงานสำหรับการหาค่าความคลาดเคลื่อน กำลังสองเฉลี่ยของตัวประมาณโดยวิธี R วิธี U วิธี B และวิธี DS .....	43
รูปที่ 3.2 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตรา ความผิดพลาดที่มีเงื่อนไขโดยวิธี R .....	45
รูปที่ 3.3 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตรา ความผิดพลาดที่มีเงื่อนไขโดยวิธี U .....	46
รูปที่ 3.4 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตรา ความผิดพลาดที่มีเงื่อนไขโดยวิธี B .....	48
รูปที่ 3.5 แสดงแผนผังงานสำหรับการหาค่าประมาณของอัตรา ความผิดพลาดที่มีเงื่อนไขโดยวิธี DS.....	51
รูปที่ 4.1 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อน กำลังสองเฉลี่ยกับค่ารากที่สองของค่ามาหาลาโนบิส ดิสแทนท์ เมื่อขนาดตัวอย่างเป็น 10 และขนาด ตัวแปรอิสระเป็น 3 .....	78
รูปที่ 4.2 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อน กำลังสองเฉลี่ยกับค่ารากที่สองของค่ามาหาลาโนบิส ดิสแทนท์ เมื่อขนาดตัวอย่างเป็น 10 และขนาด ตัวแปรอิสระเป็น 5 .....	78
รูปที่ 4.3 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อน กำลังสองเฉลี่ยกับค่ารากที่สองของค่ามาหาลาโนบิส ดิสแทนท์ เมื่อขนาดตัวอย่างเป็น 10 และขนาด ตัวแปรอิสระเป็น 7 .....	79
รูปที่ 4.4 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อน กำลังสองเฉลี่ยกับค่ารากที่สองของค่ามาหาลาโนบิส ดิสแทนท์ เมื่อขนาดตัวอย่างเป็น 10 และขนาด ตัวแปรอิสระเป็น 9 .....	79
รูปที่ 4.5 แสดงความสัมพันธ์ระหว่างค่าความคลาดเคลื่อน กำลังสองเฉลี่ยกับค่ารากที่สองของค่ามาหาลาโนบิส ดิสแทนท์ เมื่อขนาดตัวอย่างเป็น 20 และขนาด ตัวแปรอิสระเป็น 3 .....	80









