

การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย



ร.ศ.ท.หญิง ชุติมา ชัยมุสิก

ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาคตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

จุฬาลงกรณ์มหาวิทยาลัย
ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2533

ISBN 974-577-243-7

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

016899

117351293

MULTIPLE REGRESSION ANALYSIS WITH MISSING OBSERVATIONS
AMONG THE INDEPENDENT VARIABLES



POL.LT. Chutima Chaimusig

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

Department of Statistics
Graduate School


Chulalongkorn University

1989


ISBN 974-577-243-7

หัวข้อวิทยานิพนธ์ การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย
โดย ร.ศ.ท.หญิง ชุติมา ชัยมุสิก
ภาควิชา สถิติ
อาจารย์ที่ปรึกษา อาจารย์ ดร. อรุณี กำลัง

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต


..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. นวน วิษิตกาน)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. สรชัย นิสาลบุตร)


..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร. อรุณี กำลัง)


..... กรรมการ
(รองศาสตราจารย์ ดร. สุชาดา กิระนันท์)


..... กรรมการ
(รองศาสตราจารย์ วัลภา ประกอบผล)

ศูนย์วิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ชุตินา ชัยมุสิก : การวิเคราะห์การถดถอยเชิงซ้อนเมื่อข้อมูลของตัวแปรอิสระสูญหาย
(MULTIPLE REGRESSION ANALYSIS WITH MISSING OBSERVATIONS AMONG
THE INDEPENDENT VARIABLES) อ. ที่ปรึกษา : คร. อรุณี กาลัง, 139 หน้า

ในการวิเคราะห์การถดถอยเชิงซ้อนด้วยวิธีกำลังสองน้อยที่สุดจะทำได้ถ้ามีข้อมูลบางตัวสูญหาย วิธีการแก้ปัญหาอย่างหนึ่งก็คือ ตัดค่าสังเกตชุดนั้นทิ้งไปแต่การแก้ปัญหาโดยวิธีนี้จะมีผลทำให้จำนวนค่าสังเกตน้อยลงและสูญเสียรายละเอียดบางอย่างไป วิธีการแก้ปัญหาก็วิธีหนึ่งคือทำการประมาณค่าข้อมูลสูญหายด้วยวิธีการต่าง ๆ ก่อนแล้วจึงทำการวิเคราะห์ข้อมูลด้วยวิธีกำลังสองน้อยที่สุด

ในการวิจัยนี้ได้ศึกษาเปรียบเทียบวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอยเชิงซ้อน ซึ่งจะประมาณข้อมูลของตัวแปรอิสระที่สูญหายด้วย วิธีวิเคราะห์ความถดถอย วิธี MAXIMUM LIKELIHOOD วิธีค่าเฉลี่ย และวิธีค่ามัธยฐาน โดยใช้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของสมการถดถอยของวิธีที่ไม่มีข้อมูลสูญหายเป็นเกณฑ์ในการเปรียบเทียบ สถานการณ์ที่ใช้ในการศึกษาจำลองขึ้นโดยใช้เทคนิคมอนติคาร์โล 432 สถานการณ์ จำแนกตามการกระจายข้อมูล 3 ระดับ โดยใช้ C.V. เป็นตัวกำหนดคือ 0.05, 0.20 และ 1.0, ตามขนาดตัวอย่าง (N) 3 ระดับคือ 30, 70 และ 100, ตามจำนวนตัวแปรอิสระ (M) 4 ระดับ คือ 2, 3, 5, และ 7, ตามค่าเบี่ยงเบนมาตรฐาน (σ) 4 ระดับคือ 5, 10, 20 และ 25, ตามการสูญหายของข้อมูล ซึ่งเท่ากันในทุกตัวแปร (MM) 3 ระดับคือ 5%, 10% และ 15%

จากการวิจัยพบว่าวิธีการประมาณข้อมูลสูญหายในการวิเคราะห์การถดถอยเชิงซ้อน ทั้ง 4 วิธี ให้ผลต่างกันตามสถานการณ์ต่าง ๆ ซึ่งโดยส่วนใหญ่วิธีค่าเฉลี่ยให้ผลดีที่สุด ยกเว้นเมื่อมีขนาดตัวอย่างน้อยและจำนวนตัวแปรอิสระมาก วิธีวิเคราะห์การถดถอยจะให้ผลดีที่สุด แต่ถ้าตัวอย่างมีขนาดใหญ่และจำนวนตัวแปรอิสระมีน้อยการตัดชุดของข้อมูลสูญหายทั้งจะไม่มีผลกระทบต่อผลการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด

ศูนย์วิทยุทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2533

ลายมือชื่อนิสิต *ชุตินา ชัยมุสิก*
ลายมือชื่ออาจารย์ที่ปรึกษา *อรุณี กาลัง*

CHUTIMA CHAIMUSIG : MULTIPLE REGRESSION ANALYSIS WITH MISSING OBSERVATIONS AMONG THE INDEPENDENT VARIABLES.

THESIS ADVISOR : DR. ARUNEE KUMLUNG, 139 pp.

Multiple regression analysis using the least squares method cannot be accomplished when some of the independent variable values are missing. One way to overcome this problem is to eliminate the incomplete observations from the analysis, but this would cause a reduction in the number of observed data and thus details of the analysis would be missing. Another alternative is to make an estimation of missing data, which can be done by various methods, and then perform the multiple regression analysis using the least square method.

This research analyses and compares missing data estimation methods in multiple regression obtained from making estimates of missing independent variable values by linear regression method, MLE method, mean method and median method. The MSE of regression equations by complete method was criteria used in the comparison of 432 cases obtained through the MONTE CARLO SIMULATION process. The cases considered are derived from various assumptions on dispersion levels (C.V. = 0.05, 0.20, 1.00), sample sizes (N = 30, 70, 100), number of independent variables (M = 2, 3, 5, 7), standard deviation of error levels ($\sigma = 5, 10, 20, 25$), and percentage of missing data (MM = 5, 10, 15).

From this study the 4 different methods of the multiple regression analysis with missing observations among the independent variables provides different results. The mean method is concluded to be the best except for the small sample size and large number of independent variables, the linear regression method is the best. However, eliminating the incomplete observations from the analysis can be done on the large sample size and small number of independent variables situation.

ภาควิชา สถิติ
สาขาวิชา สถิติ
ปีการศึกษา 2533

ลายมือชื่อนิสิต ชุตติมา ไชยมุสิก
ลายมือชื่ออาจารย์ที่ปรึกษา อรุณี คูมลุง

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ อาจารย์ ดร. อรุณี กำลิ่ง อาจารย์ประจำภาควิชาคหกรรมศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำและข้อคิดเห็นต่าง ๆ ของการวิจัยมาด้วยดีตลอดผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ. โอกาสนี้

นอกจากนี้ ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สรชัย นิสาลบุตร และ รองศาสตราจารย์ ดร. สุชาดา กิระนันท์ ที่ให้คำแนะนำเพิ่มเติมเสมอมา

ขอกราบขอบพระคุณท่านอาจารย์ทุก ๆ ท่านที่ประสิทธิ์ประสาทความรู้ให้แก่ผู้วิจัยมาโดยตลอด

ขอขอบคุณ พ.ศ.อ. เกษียร วรศิริ ผู้กำกับศูนย์ประมวลข่าวสาร กรมตำรวจ พ.ศ.ท.หญิง รังสิมา พันธุ์สุวรรณ และเจ้าหน้าที่ทุกท่านที่ช่วยประมวลผลงานวิจัยตั้งแต่ต้นจนแล้วเสร็จ

ขอขอบคุณ ร.ศ.อ. เพิ่มพูน ชิดชอบ ที่ให้ความอนุเคราะห์ในการพิมพ์วิทยานิพนธ์ฉบับนี้ ลุล่วงไปด้วยดี

ขอขอบคุณ เพื่อน ๆ และพี่ ๆ ทุกคนที่ให้ความช่วยเหลือและคำแนะนำต่าง ๆ
ทำนองนี้ ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และญาติพี่น้องที่ช่วยส่งเสริมและสนับสนุนการเขียนของผู้วิจัยตลอดมา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ชติมา ชัยมุสิก

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ช
บทที่	
1. บทนำ	1
2. ทฤษฎีที่ใช้ในการวิจัย	5
3. ระเบียบวิธีวิจัย	16
4. ผลการวิจัย	25
5. สรุปผลการวิจัยและข้อเสนอแนะ	82
บรรณานุกรม	91
ภาคผนวก ก.	94
ภาคผนวก ข.	97
ประวัติ	139

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย