

การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการความถดถอยเชิงเส้น
ด้วยวิธีกำลังสองต่ำสุดและวิธีบูตสเตรป

นาง มาลี ตระการศิรินนท์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษิตตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2532

ISBN 974-569-856-3

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

016075

I 1749932x

A COMPARISON OF PARAMETER ESTIMATION IN LINEAR REGRESSION MODEL
BETWEEN LEAST SQUARES METHOD AND BOOTSTRAP METHOD

Mrs. Malee Trakarnsirinont

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 974-569-856-3


หัวข้อวิทยานิพนธ์ : การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการ
ความถดถอยเชิงเส้น ด้วยวิธีกำลังสองต่ำสุดและวิธีบูตสเตรป
ชื่อนิสิต : นาง มาลี ตระการศิรินนท์
ภาควิชา : สถิติ
อาจารย์ที่ปรึกษา : อาจารย์ ดร. สุพล ดุรงค์วัฒนา

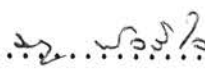
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

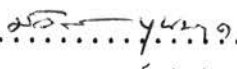

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วัชรากัญ)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. สรชัย พิศาลบุตร)


..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร. สุพล ดุรงค์วัฒนา)


..... กรรมการ
(รองศาสตราจารย์ มณฑา หัววิไล)


..... กรรมการ
(รองศาสตราจารย์ มัลลิกา บุนนาค)



มาลี ตระการศิรินนท์ : การเปรียบเทียบการประมาณค่าพารามิเตอร์ของรูปแบบสมการถดถอยเชิงเส้น ด้วยวิธีกำลังสองต่ำสุดและวิธีบูตสเตรป (A COMPARISON OF PARAMETER ESTIMATION IN LINEAR REGRESSION MODEL BETWEEN LEAST SQUARES METHOD AND BOOTSTRAP METHOD) อ.ที่ปรึกษา : ดร.สุพล ดุรงค์วัฒนา, 126 หน้า.

การวิจัยในครั้งนี้มีวัตถุประสงค์ที่จะศึกษาและนำเอาวิธีบูตสเตรปมาใช้ในการประมาณค่าพารามิเตอร์ในสมการความถดถอยเชิงเส้น และศึกษาคุณสมบัติของตัวประมาณ นอกจากนั้นยังทำการศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณที่ได้จากวิธีกำลังสองต่ำสุด และวิธีบูตสเตรป โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นตัวเปรียบเทียบ ข้อมูลที่ใช้ได้จากการจำลองด้วยโปรแกรมคอมพิวเตอร์ โดยใช้เทคนิคมอนติคาร์โล เมื่อความคลาดเคลื่อน (ϵ) มีการแจกแจงแบบชุกติ แบบยูนิฟอร์ม แบบโลจิสติก แบบดับเบิลเอ็กซ์โพเนนเชียล และแบบปกติปโลมปน ซึ่งทำการศึกษาเมื่อสเกลแพคเตอร์ (c) มีค่าเป็น 3 และ 10 และมีเปอร์เซ็นต์การปโลมปน(p) เป็น 1%, 5%, 10% และ 25% ตามลำดับ และขนาดตัวอย่างและจำนวนตัวแปรอิสระที่ใช้มี 3 ชุดคือ

- ชุดที่ 1 มีขนาด (n_1) = 50 จำนวนตัวแปรอิสระ (p_1) = 5
- ชุดที่ 2 มีขนาด (n_2) = 10 จำนวนตัวแปรอิสระ (p_2) = 4
- ชุดที่ 3 มีขนาด (n_3) = 5 จำนวนตัวแปรอิสระ (p_3) = 3

โดยในแต่ละการทดลองกระทำซ้ำ 200 ครั้ง และจำนวนการสุ่มตัวอย่างแบบใส่คืนในวิธีบูตสเตรปกระทำ 100 ครั้ง

เนื่องจากขนาดตัวอย่างแต่ละชุดมีจำนวนตัวแปรอิสระแตกต่างกันจึงพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยทั้งในรูปตัวประมาณแต่ละตัวและในรูปฟังก์ชันเชิงเส้นของตัวประมาณ ผลการศึกษาสรุปได้ดังนี้

1. วิธีบูตสเตรปสามารถนำมาใช้ประมาณค่าพารามิเตอร์ใด ๆ ได้ ในกรณีที่ไม่ทราบลักษณะการแจกแจงของประชากรหรือค่าประมาณนั้นไม่สามารถหาได้โดยตรงจากสูตรทั่วไป
2. การนำวิธีบูตสเตรปมาใช้ในการประมาณค่าพารามิเตอร์ในสมการความถดถอยเชิงเส้นที่มีรูปแบบทั่วไป คือ $Y = X\beta + \epsilon$ เมื่อไม่ทราบลักษณะการแจกแจงของความคลาดเคลื่อน (ϵ) ผลการเปรียบเทียบประสิทธิภาพของตัวประมาณพบว่า ขนาดตัวอย่าง จำนวนตัวแปรอิสระ จำนวนครั้งในการสุ่มตัวอย่างในวิธีบูตสเตรป และรูปแบบการแจกแจงของความคลาดเคลื่อน มีผลต่อความคลาดเคลื่อนกำลังสองเฉลี่ย
3. วิธีกำลังสองต่ำสุดจะมีประสิทธิภาพดีเมื่อ ความคลาดเคลื่อนมีการแจกแจงที่ไม่เบี่ยงเบนไปจากแบบปกติมากนักและมีขนาดตัวอย่างมากพอ ในกรณีที่ความแตกต่างระหว่างขนาดตัวอย่างกับจำนวนตัวแปรอิสระมีค่าน้อย วิธีบูตสเตรปจะมีประสิทธิภาพดีกว่า แต่จะต้องเสียค่าใช้จ่ายเพิ่มขึ้นในการประมวลผลโดยใช้คอมพิวเตอร์

ดังนั้นในทางปฏิบัติ เมื่อสงสัยว่าข้อมูลที่จะนำมาวิเคราะห์หาค่าผิดปกติ คือมีค่าสูงหรือต่ำมาก ๆ เมื่อเทียบกับค่าอื่น ๆ อาจมีสาเหตุเนื่องมาจากความคลาดเคลื่อนที่มีการแจกแจงแบบอื่นที่ไม่ใช่การแจกแจงแบบปกติ ก็ให้ทำการทดสอบดูและเมื่อพบว่าเป็นค่าผิดปกติจริง ควรทำการแก้ไขก่อนแล้วจึงใช้วิธีกำลังสองต่ำสุดในการประมาณค่าพารามิเตอร์ จะทำให้ประหยัดเวลาและทุนค่าใช้จ่าย

ภาควิชา สถิติ
 สาขาวิชา สถิติ
 ปีการศึกษา 2531
 ลายมือชื่อนิติ มาลี ตระการศิรินนท์
 ลายมือชื่ออาจารย์ที่ปรึกษา

MALEE TRAKARNSIRINONT : A COMPARISION OF PARAMETER ESTIMATION IN
LINEAR REGRESSION MODEL LEAST SQUARES METHOD AND BOOTSTRAP METHOD.
THESIS ADVISOR : SUPOL DURONGWATANA, Ph.D. 126 PP.

This research has the following objectives (i) To study and employ Bootstrap method for parameter estimation ; (ii) To study the characteristics of estimators ; and (iii) To make a comparative study between the efficiency of the estimators resulting from the Least Squares method and the Bootstrap method by means of mean square errors as the relatives.

The data from this experiment was obtained through simulation using Monte Carlo technique. The distributions of error for this were normal, uniform, logistic, double exponential and scaled contaminated normal using scale factor 3,10 with 1%, 5%, 10% and 25% contaminated. The sample size and independent variables used in this study are as follows :

- sample size (n_1) = 50 ; independent variable (p_1) = 5
- sample size (n_2) = 10 ; independent variable (p_2) = 4
- sample size (n_3) = 5 ; independent variable (p_3) = 5

In each experiment, which was conducted 200 times total, the number of sampling with replacement in Bootstrap method was done 100 times.

Since each sample size has different numbers of independent variables, the mean square error were thus considered in two types, (i) the individual estimators and (ii) the linear function of estimators. The results of the study are as follows : (1) The Bootstrap method can be used to estimate any parameter when the distribution of population is unknown or the estimators can not becalculated from normal formula. (2) The Bootstrap method is used for parameter estimation inordinary linear regression equation, $Y = X\beta + \epsilon$, when the distribution of error is unknown. The results of comparative efficiency of estimators reveals that the sample size, the number of independent variables, the number ofbootstrap sampling and the distribution of error effect mean square errors. (3) The Least Squares method is more efficient when the distributions of error is not much deviated from normal distribution and there is sufficient sample size. In addition, the Bootstrap method is more efficient when the samplesize and the number of independent variables are not much different. However, the computer processing for this purpose would increase the expense.

Practically, suspicious data should be tested when the data appears to be much higher or lower than other values, resulting from the distribution of errors is not the normal distribution. If the data is not the normal value, it should be corrected before the Least Squares method for parameter estimation is used. It would then be more economical.

ภาควิชา
สาขาวิชา
ปีการศึกษา 2531

ลายมือชื่อผู้นิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงลงได้ด้วยความช่วยเหลืออย่างดียิ่งจาก อาจารย์ ดร. สุปล คุรงค์วัฒนา อาจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่กรุณาให้คำปรึกษา แนะนำ ตลอดจนดูแล แก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด ผู้เขียนขอกราบขอบพระคุณด้วยความรู้สึกซาบซึ้งและสำนึกในพระคุณยิ่ง

นอกจากนี้ผู้เขียนขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ ซึ่งประกอบด้วย รองศาสตราจารย์ ดร.สรชัย พิศาลบุตร รองศาสตราจารย์มณฑา พัววิไล รองศาสตราจารย์มัลลิกา บุณนาค ที่ได้ช่วยตรวจและแก้ไขให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ถูกต้องยิ่งขึ้น

ขอกราบขอบพระคุณ คุณพ่อ - คุณแม่ ที่กรุณาให้ความช่วยเหลือ และเป็นกำลังใจในการศึกษาครั้งนี้ ขอขอบคุณเพื่อน ๆ และน้อง ๆ ที่ช่วยดำเนินการติดต่อประสานงานต่าง ๆ มาโดยตลอด

ท้ายที่สุดนี้ผู้เขียนขอขอบคุณสำนักบริการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่ ที่ได้ให้โอกาสลาศึกษาต่อเพื่อเพิ่มพูนความรู้ และยังได้ให้ใช้บริการเครื่อง VAX-11/750 ในการประมวลผล เพื่อศึกษาผลการวิจัยในครั้งนี้อีกด้วย

มาลี ตระการศิรินนท์

สารบัญ

หน้า

บทคัดย่อภาษาไทย	ค
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ช
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	4
1.3 ขอบเขตของเบื้องต้น	4
1.4 ขอบเขตของการวิจัย	5
1.5 คำจำกัดความ	11
1.6 ประโยชน์ของการวิจัย	12
บทที่ 2 การประมาณค่าสัมประสิทธิ์ความถดถอย	13
2.1 การประมาณค่าพารามิเตอร์โดยวิธีกำลังสองต่ำสุด	13
2.2 คุณสมบัติที่สำคัญของตัวประมาณกำลังสองต่ำสุด	15
2.3 การประมาณค่าพารามิเตอร์โดยวิธีบุคคลแปร	20
2.4 คุณสมบัติที่สำคัญของตัวประมาณที่ได้จากวิธีบุคคลแปร	25
2.5 ผลงานวิจัยที่เกี่ยวข้อง	30
บทที่ 3 วิธีดำเนินการวิจัย	31
3.1 วิธีมอนติคาร์โล	32
3.2 แผนการทดลอง	33

3.3	ขั้นตอนในการวิจัย	34
3.4	โปรแกรมที่ใช้ในการวิจัย	42
3.5	ข้อมูลที่ใช้ในการวิจัย	43
บทที่ 4	ผลการวิเคราะห์	45
4.1	การเปรียบเทียบประสิทธิภาพของตัวประมาณแต่ละตัวโดยใช้ ความคลาดเคลื่อนกำลังสองเฉลี่ย	46
4.2	การเปรียบเทียบประสิทธิภาพสัมพัทธ์ของตัวประมาณแต่ละตัว	60
4.3	การเปรียบเทียบประสิทธิภาพในรูปฟังก์ชันเชิงเส้นของตัวประมาณ โดยใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย	64
4.4	การเปรียบเทียบประสิทธิภาพสัมพัทธ์ในรูปฟังก์ชันเชิงเส้น ของตัวประมาณ	69
บทที่ 5	สรุปและอภิปรายผล	73
5.1	ผลสรุปการเปรียบเทียบประสิทธิภาพของตัวประมาณแต่ละตัว โดยพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย	73
5.2	ผลสรุปการเปรียบเทียบประสิทธิภาพในรูปฟังก์ชันเชิงเส้นของ ตัวประมาณแต่ละตัว โดยพิจารณาจากค่าความคลาดเคลื่อน กำลังสองเฉลี่ย	81
5.3	การอภิปรายผล	84
5.4	ข้อเสนอแนะ	86
บรรณานุกรม	88
ภาคผนวก	93
ประวัติผู้เขียน	126