

การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย
เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง

นางสาวจงดี โรจนประศาสน์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2536

ISBN 974-582-501-8

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

019236 ; 117867260

A COMPARISON ON METHODS OF ESTIMATING PARAMETERS IN SIMPLE LINEAR
REGRESSION WITH A CENSORED RESPONSE VARIABLE

MISS CHONGDEE ROGANAPRASAST

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

Department of Statistics

Graduate School

Chulalongkorn University

1993

ISBN 974-582-501-8.

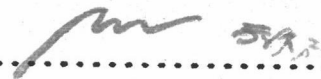
หัวข้อวิทยานิพนธ์ การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้น
อย่างง่ายเมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง

โดย นางสาวจงดี โรจนประศาสน์


ภาควิชา สถิติ

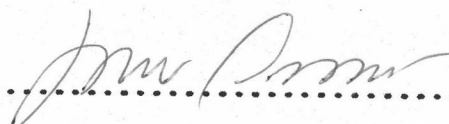
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์


บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

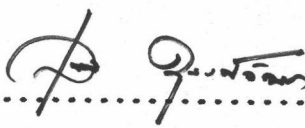

.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วิษราภักดิ์)

คณะกรรมการสอบวิทยานิพนธ์


.....ประธานกรรมการ
(รองศาสตราจารย์ ผกาวัต ศิริรังษี)


.....อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ร้อยเอก มานพ วราภักดิ์)


.....กรรมการ
(รองศาสตราจารย์ ดร. ชีระพร วีระถาวร)


.....กรรมการ
(อาจารย์ ดร. สุนล ชุรงค์วัฒนา)



พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

จงตี โรจนประศาสน์ : การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอย
เชิงเส้นอย่างง่ายเมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง (A COMPARISON ON METHODS OF
ESTIMATING PARAMETERS IN SIMPLE LINEAR REGRESSION WITH A CENSORED
RESPONSE VARIABLE) อ. ที่ปรึกษา : ผศ. ร.อ.มานพ วรภักดิ์, 102 หน้า.
ISBN 974-582-501-8.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของสมการ
ถดถอยเชิงเส้นอย่างง่าย ในกรณีที่ค่าสังเกตของตัวแปรตามเป็นค่าที่ถูกตัดทิ้งทางขวา ซึ่งวิธีการประมาณ
ค่าพารามิเตอร์ที่ใช้ในการศึกษาค้างนี้คือ 1) วิธีการกำลังสองต่ำสุด (Ordinary Least Squares
Method) 2) วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation
Method) 3) วิธีการของบัคเคิลและเจมส์ (Buckley and James Method) การเปรียบเทียบกระทำ
ภายใต้สถานการณ์ของขนาดตัวอย่างเท่ากับ 10, 15, 30, 50 และ 70 สัดส่วนของข้อมูลที่ถูกตัดทิ้ง
เป็น 10%, 20%, 30% และ 40% และการแจกแจงของค่าที่ถูกตัดทิ้งเป็นแบบสมมาตรแบบปกติดัด
ปลายทางซ้าย (Left-truncated normal distribution) แบบไวบูลล์ตัดปลายทางซ้าย (Left-
truncated Weibull distribution) และแบบเชิงเส้น ข้อมูลที่ใช้ในการวิจัยได้จากการจำลอง
ด้วยเทคนิคมอนติคาร์โลและทำการทดลองซ้ำ ๆ กัน 1,000 ครั้ง สำหรับแต่ละสถานการณ์ที่กำหนด
เพื่อประมาณค่าพารามิเตอร์ และหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE)
ของการประมาณค่าตัวแปรตามด้วยวิธีการทั้งสาม

ผลการวิจัยสรุปได้ดังนี้

วิธีการประมาณด้วยภาวะน่าจะเป็นสูงสุด จะให้ค่าความคลาดเคลื่อน RMSE ของการ
ประมาณค่าตัวแปรตามต่ำกว่าวิธีการอื่น ๆ ในทุก ๆ สถานการณ์ที่ทำการศึกษา รองลงมาคือวิธีการ
ของบัคเคิลและเจมส์ และวิธีการกำลังสองต่ำสุดตามลำดับ

ปัจจัยที่มีผลกระทบต่อค่า RMSE มีดังนี้ กรณีที่ขนาดตัวอย่างคงที่ ค่า RMSE ของวิธีการ
ประมาณด้วยภาวะน่าจะเป็นสูงสุดจะไม่แปรผันตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง แต่วิธีการกำลังสองต่ำสุด
และวิธีการของบัคเคิลและเจมส์จะแปรผันโดยตรงตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง และกรณีขนาดตัวอย่าง
เพิ่มขึ้นขณะที่สัดส่วนของข้อมูลที่ถูกตัดทิ้งคงที่ ค่า RMSE ของแต่ละวิธีการจะแปรผันโดยตรงตามขนาด
ตัวอย่าง ส่วนการแจกแจงของค่าที่ถูกตัดทิ้งไม่มีผลกระทบต่อประสิทธิภาพของแต่ละวิธีการ

ภาควิชาสถิติ.....
สาขาวิชาสถิติ.....
ปีการศึกษา 2535

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

##C323207 : MAJOR STATISTICS

KEY WORD : ESTIMATING PARAMETERS/ SIMPLE LINEAR REGRESSION/

A CENSORED RESPONSE VARIABLE. THESIS ADVISOR : ASST. PROF.

CAPT. MANOP VARAPHAUDI, M.S. 102 PP.

ISBN 974-582-501-8.

The objective of this study is to compare methods of estimating a response variable in simple linear regression analysis having a right censored response variable. The methods of estimating parameters under consideration in this study are Ordinary Least Squares method, Maximum Likelihood Estimation method, and Buckley and James method. The comparison was done under conditions of sample sizes 10, 15, 30, 50, and 70, censoring proportions 10%, 20%, 30%, and 40%, and censoring distributions are of uniform distribution, left-truncated normal distribution, left-truncated Weibull distribution, and linear function. The data for this experiment were generated through the Monte Carlo simulation technique. The experiment was repeated 1,000 times under each condition in estimating parameters and evaluating the square root of mean squares error of estimating the response variable (RMSE).

Results of the study are as follows :-

Under all conditions in this study, the RMSE of maximum likelihood estimation method is lower than the RMSE of other methods and the RMSE of ordinary least squares method is largest.

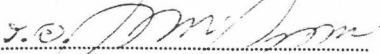
For any sample size, the RMSE of maximum likelihood estimation method does not increase as the censoring proportion increases, but the RMSE of ordinary least squares method and Buckley and James method increase as the censoring proportion increases. For any censoring proportion, the RMSE of each method increase as the sample size increases. The efficiency of all methods are not affected by the censoring distributions.

ภาควิชา.....สถิติ.....

สาขาวิชา.....สถิติ.....

ปีการศึกษา.....2535.....

ลายมือชื่อนิสิต.....

ลายมือชื่ออาจารย์ที่ปรึกษา.....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาของ ผู้ช่วยศาสตราจารย์ ร.อ.มานพ วรารักษ์ดี ที่กรุณาให้คำแนะนำ ปรึกษา ตลอดจนช่วยเหลือแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดียิ่ง จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณด้วยความรู้สึกซาบซึ้งและสำนึก ในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ผกาวัต ศิริรังษี รองศาสตราจารย์ ดร.ธีระพร วีระถาวร อาจารย์ ดร.สุพล ตุงรงค์วัฒนา ในฐานะประธานกรรมการและกรรมการ สอบวิทยานิพนธ์ ที่กรุณาตรวจแก้วิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ให้โอกาสทางการศึกษา และประสิทธิประสาทความรู้ให้แก่ผู้เขียนจนกระทั่งสำเร็จการศึกษา

ผู้วิจัยขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์มะเนาะ ยูเด็น คณบดี ผู้ช่วยศาสตราจารย์ มานพ จิตต์ภุชญา รองคณบดีฝ่ายวิจัยและวางแผน คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี ที่ให้คำปรึกษา แนะนำด้านการศึกษา ขอขอบคุณพี่ ๆ เพื่อน ๆ น้อง ๆ ผู้ร่วมงานที่ให้กำลังใจและให้ความช่วยเหลือมาโดยตลอด

ผู้วิจัยขอระลึกถึงพระคุณ คุณพ่อ คุณแม่ และขอขอบคุณ พี่สาว น้องสาว และน้องชาย ที่ให้ความสนับสนุนด้านการศึกษาและให้กำลังใจกระทั่งสำเร็จการศึกษา

ท้ายนี้ขอขอบคุณเพื่อน ๆ น้อง ๆ ที่ให้กำลังใจและมีส่วนช่วยเหลืออย่างดีตลอดมา

จงดี โรจนประศาสน์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ณ
สารบัญรูป.....	ญ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	4
1.3 สมมติฐานของการวิจัย.....	4
1.4 ข้อตกลงเบื้องต้น.....	5
1.5 ขอบเขตของการวิจัย.....	5
1.6 เกณฑ์การตัดสินใจ.....	7
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 สถิติที่ใช้ในการวิจัย.....	8
2.1 ทฤษฎีพื้นฐานที่เกี่ยวข้อง.....	8
2.2 การประมาณค่าพารามิเตอร์.....	17

	หน้า
บทที่ 3 วิธีดำเนินการวิจัย.....	33
3.1 วิธีการจำลอง โดยใช้เทคนิคมอนติคาร์โล.....	33
3.2 แผนการทดลอง.....	34
3.3 ขั้นตอนการวิจัย.....	35
3.4 โปรแกรมที่ใช้ในการวิจัย.....	43
บทที่ 4 ผลการวิจัย.....	45
4.1 ผลการศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี...	46
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	64
5.1 สรุปผลการวิจัย.....	64
5.2 ข้อเสนอแนะ.....	66
รายการอ้างอิง.....	68
ภาคผนวก.....	70
ประวัติผู้เขียน.....	102

สารบัญตาราง

ตารางที่	หน้า
4.1 แสดงค่า RMSE ของการประมาณ ของวิธีการประมาณค่าพารามิเตอร์ในสมการ ถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทั้ง 3 วิธี เมื่อขนาด ตัวอย่างเท่ากับ 10 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และจำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10%, 20%, 30%, 40% ...	49
4.2 แสดงค่า RMSE ของการประมาณ ของวิธีการประมาณค่าพารามิเตอร์ในสมการ ถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทั้ง 3 วิธี เมื่อขนาด ตัวอย่างเท่ากับ 15 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และจำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10%, 20%, 30%, 40% ...	52
4.3 แสดงค่า RMSE ของการประมาณ ของวิธีการประมาณค่าพารามิเตอร์ในสมการ ถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทั้ง 3 วิธี เมื่อขนาด ตัวอย่างเท่ากับ 30 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และจำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10%, 20%, 30%, 40% ...	55
4.4 แสดงค่า RMSE ของการประมาณ ของวิธีการประมาณค่าพารามิเตอร์ในสมการ ถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทั้ง 3 วิธี เมื่อขนาด ตัวอย่างเท่ากับ 50 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และจำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10%, 20%, 30%, 40% ...	58
4.5 แสดงค่า RMSE ของการประมาณ ของวิธีการประมาณค่าพารามิเตอร์ในสมการ ถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้ง ทั้ง 3 วิธี เมื่อขนาด ตัวอย่างเท่ากับ 70 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และจำแนกตามสัดส่วนของข้อมูลที่ถูกตัดทิ้ง 10%, 20%, 30%, 40% ...	61

สารบัญรูป •

รูปที่	หน้า
2.1 แสดงแผนภาพการทดลองที่มีการตัดทิ้งแบบสุ่ม.....	10
2.2 แสดงอัตราการสูญเสียของข้อมูล.....	13
2.3 แสดงการคำนวณฟังก์ชันการอยู่รอด โดยใช้ตัวประมาณฟีแอล.....	16
3.1 แสดงการแจกแจงแบบปกติ และแบบปกติมาตรฐาน.....	37
3.2 แสดงการแจกแจงแบบสมมาตร.....	37
3.3 แสดงการแจกแจงแบบปกติ และแบบปกติตัดปลายทางซ้าย.....	38
3.4 แสดงการแจกแจงแบบไวบูลล์ และแบบไวบูลล์ตัดปลายทางซ้าย.....	38
3.5 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนจากการประมาณ ทั้ง 3 วิธีการ...	43
4.1 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้งกรณีขนาดตัวอย่างเท่ากับ 10 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P).....	50
4.2 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้งกรณีขนาดตัวอย่างเท่ากับ 15 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P).....	53
4.3 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้งกรณีขนาดตัวอย่างเท่ากับ 30 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P).....	56

สารบัญรูป(ต่อ)

รูปที่

หน้า

- 4.4 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้งกรณีขนาดตัวอย่างเท่ากับ 50 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)..... 59
- 4.5 การเปรียบเทียบ RMSE ของการประมาณ ทั้ง 3 วิธี ในการประมาณค่าพารามิเตอร์ ในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรตามมีค่าที่ถูกตัดทิ้งกรณีขนาดตัวอย่างเท่ากับ 70 จำแนกตามการแจกแจงของค่าที่ถูกตัดทิ้ง (U, TN, TW, LI) และสัดส่วนของข้อมูลที่ถูกตัดทิ้ง (P)..... 62