การตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้น โดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์

น.ส. พรรณวดี โงวหจินตนา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

ภาควิชาสถิติ

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2530

ISBN 974-568-436-8

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

013035

I10292214

VALIDATING LINEAR REGRESSION MODEL BY SPLITTING DATA WITH DUPLEX ALGORITHM

Ms. Panvadee Ngowhajintana

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Statistics
Graduate School
Chulalongkorn University
1987

ISBN 974-568-436-8

หัวข้อวิทยานิพนธ์ การตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์

โดย น.ส. พรรณวดี โงวหจินตนา

ภาควิชา สถิติ

อาจารย์ที่ปรึกษา อาจารย์ ดร. สุพล ดุรงค์วัฒนา

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย

(ศาสตราจารย์ ดร. ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ มณฑา พัววิไล)

..... กรรมการ

(รองศาสตราจารย์ มัลลิกา บุนนาค)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ธีระพร วีระถาวร)

..... กรรมการ

(อาจารย์ ดร. สุพล ดุรงค์วัฒนา)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พรรณวดี โงวหจินตนา : การตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ (VALIDATING LINEAR REGRESSION MODEL BY SPLITTING DATA WITH DUPLEX ALGORITHM) อ. ที่ปรึกษา : อ. ดร. สุพล ดุรงค์วัฒนา, 109 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์ที่จะเสนอวิธีการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้นโดยการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ ซึ่งความเหมาะสมในที่นี้จะพิจารณาที่ข้อมูล นอกจากนี้ยังต้องการที่จะศึกษาความแกร่งของการทดสอบเชา (Chow test) ซึ่งเป็นการทดสอบเกี่ยวกับความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้นเมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้นคือ ความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และปกติปลอมปนที่กำหนดค่าสเกลแฟคเตอร์เป็น 3 และ 10 มีเปอร์เซ็นต์การปลอมปนเป็น 1% 5% 10% และ 25% ตามลำดับ ขนาดตัวอย่างที่ศึกษาคือ 30 40 50 60 70 80 90 และ 100 ส่วนจำนวนตัวแปรอิสระที่ศึกษาคือ 1 2 4 6 8 และ 10 ณ ระดับนัยสำคัญ .01 .05 และ .10 ทั้งนี้จะศึกษาความแกร่งของการทดสอบเชาจากค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 สำหรับข้อมูลลักษณะต่าง ๆ ที่ใช้ในการทดลองครั้งนี้ จำลองด้วยโปรแกรมคอมพิวเตอร์โดยใช้เทคนิคมอนติคาร์โล และกระทำซ้ำกัน 100 ครั้งในแต่ละกรณี

ผลจากการศึกษาความแกร่งของการทดสอบเชาสรุปผลได้ดังนี้

1. เมื่อความคลาดเคลื่อนมีลักษณะการแจกแจงแบบต่าง ๆ ที่ไม่เป็นไปตามข้อตกลงเบื้องต้นคือ มีการแจกแจงแบบโลจิสติค ดับเบิ้ลเอ็กซ์โปเนนเชียล และปกติปลอมปนที่กำหนดค่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนขนาดต่าง ๆ ผลปรากฏว่า ลักษณะการแจกแจงของความคลาดเคลื่อนมีอิทธิพลต่อความแกร่งของการทดสอบเชา คือ เมื่อความคลาดเคลื่อนมีลักษณะผิดปกติมากขึ้น ความสามารถในการควบคุม α ของการทดสอบจะลดลง

2. สำหรับแต่ละการแจกแจงของความคลาดเคลื่อนที่ไม่เป็นไปตามข้อตกลงเบื้องต้นของการทดสอบเชา ผลปรากฏว่า จำนวนตัวแปรอิสระมีอิทธิพลต่อความสามารถในการควบคุมความแกร่งของการทดสอบคือ เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น จะทำให้ความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 ของการทดสอบมีมากขึ้น

3. ณ ระดับนัยสำคัญต่าง ๆ กัน ผลปรากฏว่า ระดับนัยสำคัญมีอิทธิพลต่อความสามารถในการควบคุมความแกร่งของการทดสอบเชาคือ เมื่อระดับนัยสำคัญเพิ่มขึ้น จะทำให้ความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 ของการทดสอบมีมากขึ้น

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ปีการศึกษา 2530 ลายมือชื่ออาจารย์ที่ปรึกษา .....

PANVADEE NGOWHAJINTANA : VALIDATING LINEAR REGRESSION MODEL BY SPLITTING DATA WITH DUPLEX ALGORITHM. THESIS ADVISOR : DR. SUPOL DURONGWATTANA, Ph.D., 109 PP.

The objective of this study was to propose the method of validating linear regression model by splitting data with DUPLEX algorithm. It was the data validity. The study also investigated the robustness of Chow test, the test of equality in linear regression coefficients when the assumption of error was violated for the case of logistic, double exponential and scale contaminated normal distributions. As for scale contaminated normal distribution, the scale factor of 3 and 10 and contaminated percent of 1, 5, 10 and 25 were used in this case. Considering the robustness of Chow test, it was measured by probability of type I error.

This study was carried out using the sample size of 30, 40, 50, 60, 70, 80, 90 and 100 with number of independent variables of 1, 2, 4, 6, 8 and 10. The data were obtained through simulation using Monte Carlo technique, repeated 100 times, at the significance level of .01, .05 , and .10.

The results of this study were as follows:

1. The exchangeability assumption of error distribution; such as, logistic, double exponential and scale contaminated normal, had influence on the robustness of Chow test.

2. Under the same circumstance stated above, the ability to control the robustness of Chow test was affected by the number of independent variables. The more independent variables were being used the greater the ability to control probability of type I error of Chow test was.

3. At different level of significance, robustness of Chow test was out of control. In other word, the ability to control probability of type I error of Chow test increased significantly as the higher level of significance was used.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ

ปีการศึกษา 2530 ลายมือชื่ออาจารย์ที่ปรึกษา .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงลงได้ด้วยความกรุณาของอาจารย์ ดร. สุพล ดุรงค์วัฒนา อาจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำ ปรึกษา ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดีมาโดยตลอด ซึ่งผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบคุณผู้ที่ให้ความช่วยเหลือตลอดจนคำแนะนำต่าง ๆ ไว้ ณ ที่นี้ด้วย ทุก ๆ ท่าน

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และพี่ ๆ ที่ส่งเสริมและสนับสนุนการเรียนของผู้วิจัยตลอดมา และขอกราบขอบพระคุณท่านอาจารย์ทุกท่าน ที่ได้ประสิทธิประสาทความรู้แก่ผู้วิจัยมาโดยตลอด

พรรณวดี โงวหลินตนา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูป ..... ฏ

บทที่ 1 บทนำ 1

1.1 ความสำคัญและความเป็นมาของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 5

1.3 สมมติฐานของการวิจัย 5

1.4 ข้อตกลงเบื้องต้น 5

1.5 ขอบเขตของการวิจัย 6

1.6 คำจำกัดความ 10

1.7 ประโยชน์ของการวิจัย 10

2 สถิติทดสอบและผลงานวิจัยที่เกี่ยวข้อง 11

2.1 การแบ่งข้อมูลด้วยวิธีดูเพล็กซ์ 11

2.2 การประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้นโดยวิธีกำลังสองน้อยที่สุด 16

2.3 การประมาณค่ารากที่ P ของผลต่างระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์ 18

สารบัญ (ต่อ)

หน้า

2.4 การทดสอบความคงที่ของสัมประสิทธิ์ถดถอยเชิงเส้น โดยใช้การทดสอบเชา (Chow test) 18

2.5 ผลงานวิจัยที่เกี่ยวข้อง 26

บทที่ 3 วิธีดำเนินการวิจัย 28

3.1 วิธีมอนติคาร์โล (Monte Carlo Method)..... 28

3.2 แผนการทดลอง 28

3.3 ขั้นตอนในการวิจัย 29

3.4 โปรแกรมที่ใช้ในการวิจัย 37

4 ผลการวิจัย 40

4.1 ค่ารากที่ p ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์.. 40

4.2 ความน่าจะเป็นของความผิดพลาดประเภทที่ 1 41

5. สรุปผลการวิจัย และอภิปราย 84

5.1 ผลสรุปความสามารถในการควบคุมความผิดพลาดประเภทที่ 1. 84

5.2 อภิปราย 85

5.3 ข้อเสนอแนะ 86

บรรณานุกรม 87

ภาคผนวก 90

ประวัติผู้เขียน 109

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
3.1	แสดงค่าสเกลแฟคเตอร์และเปอร์เซ็นต์การปลอมปนทั้งหมดที่ใช้ในการวิจัย	29
3.2	แสดงขนาดตัวอย่างและจำนวนตัวแปรอิสระทั้งหมดที่ใช้ในการวิจัย	29
3.3	แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย	38
4.1	แสดงค่ารากที่ p ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์ จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	40
4.2	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบโลจิสติค ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	43
4.3	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียลที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ.....	47
4.4	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 3 และเปอร์เซ็นต์การปลอมปนเป็น 1% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	51

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.5	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 3 และเปอร์เซ็นต์การปลอมปนเป็น 5% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	54
4.6	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 3 และเปอร์เซ็นต์การปลอมปนเป็น 10% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	58
4.7	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 3 และเปอร์เซ็นต์การปลอมปนเป็น 25% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	62
4.8	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 10 และเปอร์-เซ็นต์การปลอมปนเป็น 1% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	66
4.9	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 10 และเปอร์-เซ็นต์การปลอมปนเป็น 5% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	70

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.10	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 10 และเปอร์เซ็นต์การปลอมปนเป็น 10% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	74
4.11	แสดงค่าความน่าจะเป็นของความผิดพลาดประเภทที่ 1 ของการทดสอบเชาจากการทดลอง เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน ที่กำหนดค่าสเกลแฟคเตอร์เป็น 10 และเปอร์เซ็นต์การปลอมปนเป็น 25% ที่ระดับนัยสำคัญ .01 .05 และ .10 จำแนกตามขนาดตัวอย่างและจำนวนตัวแปรอิสระ	78
4.12	แสดงจำนวนครั้งที่การทดสอบเชาสามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ และควบคุมไม่ได้จากการทดลอง 44 กรณี สำหรับแต่ละรูปแบบการแจกแจงของความคลาดเคลื่อนที่ระดับนัยสำคัญ .01 .05 และ .10	82

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญรูป

รูปที่		หน้า
3.1	แสดงฝังงานในการแบ่งข้อมูลด้วยวิธีดูเพล็กซ์	34
3.2	แสดงฝังงานสำหรับการหาค่ารากที่ p ของสัดส่วนระหว่างดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการประมาณค่ากับดีเทอร์มิแนนท์ของตัวแปรอิสระที่ได้จากข้อมูลที่ใช้ในการพยากรณ์และการหาค่าความน่าจะเป็นที่จะเกิดความผิดพลาดประเภทที่ 1 ของการทดสอบเชา	36

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย