บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression) เป็นวิธีทางสถิติที่ใช้ในการวิจัย ซึ่งปัจจัยที่ทำการศึกษา เรียกว่า ตัวแปรตาม (dependent variable) และปัจจัยที่ส่งผลกระทบต่อปัจจัยที่ทำการศึกษาเรียกว่า ตัวแปรอิสระ(independent variable) โดย " การวิเคราะห์ความถดถอยพหุนาม (polynomial regression analysis) " เป็นกรณีหนึ่งของการวิเคราะห์ความถดถอยพหุคูณซึ่งพบว่ามีการนำมาใช้วิเคราะห์ปัญหาในหลายสาขา เช่น ทางด้านเศรษฐศาสตร์ ทางด้านประชากรศาสตร์ เป็นต้น

วิธีการคัดเลือกตัวแปรอิสระจะพิจารณาจากตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามสูงที่สุดมาใช้เพียงตัวเดียว ซึ่งวิธีการของโพลิโนเมียลเหมาะสำหรับสถานการณ์ที่พบว่าเส้นกราฟความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระมีความผันผวน (Fluctuation) มาก เมื่อความผันผวนยิ่งมากเพียงใดก็จะต้องเพิ่มตัวแปรอิสระกำลังสูงเข้าไปในตัวแบบเพิ่มขึ้นด้วย แต่การเพิ่มตัวแปรอิสระกำลังต่าง ๆ เข้าไปในตัวแบบจะกระทำโดยปราศจากเหตุผลมิได้เพราะการเพิ่มตัวแปรอิสระเข้าไปในตัวแบบหนึ่งตัวก็จะมีผลให้ระดับขั้นความเสรีของส่วนเหลือ (Degree of Freedom of Residual) ลดลงหนึ่งหน่วย และ MSR (Mean Square Regression) จะค่อย ๆ ต่ำลงส่งผลให้การวิเคราะห์มีประสิทธิภาพต่ำลงด้วย การวิเคราะห์ความถดถอยพหุนามเชิงเส้นทั่วไปมีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \varepsilon_i \quad , \quad i = 1, 2, 3, \ldots, n$$

เมื่อ y_i แทนตัวแปรตามที่ i

x_i^p แทนตัวแปรอิสระที่ i ยกกำลัง p

β_j แทนพารามิเตอร์ของตัวแปรอิสระตัวที่ j ; $j = 1, 2, \ldots, p$

ε_i แทนความคลาดเคลื่อนที่ i

n แทนขนาดตัวอย่าง

และ p แทนจำนวนพจน์พหุนามในสมการ

การวิเคราะห์ความถดถอยพหุนามเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กัน เรียกว่าพหุสัมพันธ์ (multicolinearity) ส่งผลให้ความคลาดเคลื่อนของตัวแบบเชิงเส้นที่ได้มีค่าเฉลี่ยความ

คลาดเคลื่อนยกกำลังสองสูงมากกว่าปกติ ดังนั้นเพื่อทำการแก้ปัญหาดังกล่าวที่เกิดขึ้นนี้เราจึงควรพิจารณาเลือกตัวประมาณค่าพารามิเตอร์ที่ใช้ประมาณค่าพารามิเตอร์ที่สามารถแก้ปัญหาความสัมพันธ์ของตัวแปรอิสระ เพื่อลดค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสองของตัวแบบเชิงเส้น

ในการวิจัยครั้งนี้ผู้วิจัยได้ทำการศึกษาเกี่ยวกับการวิเคราะห์ความถดถอยพหุนามพบว่ามีการวิเคราะห์คล้ายกับการวิเคราะห์ความถดถอยหลายตัวแปร ดังนั้นควรคัดเลือกตัวแบบที่ใช้ในการวิเคราะห์ความถดถอยก่อนซึ่งทำได้หลายวิธีด้วยกัน และเมื่อได้ตัวแบบที่ต้องการวิเคราะห์ที่เหมาะสมจะทำให้การประมาณค่าพารามิเตอร์ในขั้นตอนต่อมาเป็นการประมาณพารามิเตอร์ที่ดีและเหมาะสมด้วย โดยมีงานวิจัยที่มีความเกี่ยวข้องดังนี้

ในปีการศึกษา 2539 นพมาศ อัครจันทโชติ ได้ทำการศึกษาเกี่ยวกับการเปรียบเทียบวิธีการสร้างตัวแบบในการวิเคราะห์ความถดถอยพหุนามกรณีมี 2 ตัวแปรอิสระ ซึ่งเกิดอันตรกิริยา(interaction) โดยทำการเปรียบเทียบวิธีการสร้างตัวแบบ 4 วิธีได้แก่ วิธีการสร้างตัวแบบด้วยวิธีกำลังสองน้อยสุดสามัญ (ordinary least squares method (OLS)) การสร้างตัวแบบด้วยวิธีกำจัดตัวแปรอิสระย้อนหลัง (backward elimination method (BE)) การสร้างตัวแบบด้วยวิธีความถดถอยขั้นบันได(stepwise regression method (SW)) และการสร้างตัวแบบด้วยวิธีตัวแบบหลักเกณฑ์ดี (well – formulated method (WF))

ในปีการศึกษา 2546 ธิฏิรัตน์ เมฆบัณฑิตกุล ได้ทำการศึกษาเกี่ยวกับการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยพหุนาม กรณีที่มีความคลาดเคลื่อนเกิดขึ้นในตัวแปรอิสระ (PARAMETERS-ESTIMATION OF POLYNOMIAL REGRESSION MODELS WITH ERRORS IN INDEPENDENT VARIABLES) โดยจะเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ 3 วิธี ได้แก่ วิธีกำลังสองน้อยสุดสามัญ วิธีกำลังสองน้อยสุดปรับปรุง (adjusted least squares method (ALSM)) และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (weighted least squares method (WLS))

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าพารามิเตอร์ความถดถอยพหุนามโดยใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของพารามิเตอร์ที่ได้ โดยตัวประมาณค่าพารามิเตอร์ความถดถอยพหุนามที่ใช้ในการวิจัยมี 3 ตัวประมาณ ดังนี้

1) ตัวประมาณกำลังสองน้อยสุดสามัญ (ordinary least squares estimator (OLS))
2) ตัวประมาณริดจ์สามัญ (ridge ordinary least squares estimator (ROLS))
3) ตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด (ridge least absolute value estimator (RLAV))

1.3 สมมติฐานของการวิจัย

สมมติฐานการวิจัยในครั้งนี้คือตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด (ridge least absolute value estimator (RLAV)) มีประสิทธิภาพมากกว่าตัวประมาณกำลังสองน้อยสุด (ordinary least squares estimator (OLS)) และตัวประมาณริดจ์สามัญ (Ridge Ordinary Least Squares Estimator (ROLS)) (เพราะตัวประมาณค่าสัมบูรณ์น้อยสุดเป็นตัวประมาณที่ได้รับอิทธิพลจากค่าผิดปกติ และพหุสัมพันธ์น้อยกว่าตัวประมาณกำลังสองน้อยสุดสามัญ ดังนั้นเมื่อนำตัวประมาณค่าสัมบูรณ์น้อยสุดมาพัฒนาเป็นตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุดจึงเป็นตัวประมาณที่ดีที่สุด)

1.4 ขอบเขตของการวิจัย

1. ตัวแบบของความถดถอยพหุนามที่ใช้ในการวิจัยมีรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + ... + \beta_p x_i^p + \varepsilon_i \quad , \quad i = 1 , 2 , 3 , ... , n$$

เมื่อ y_i แทนตัวแปรตามตำแหน่งที่ i

x_i^p แทนตัวแปรอิสระตำแหน่งที่ i ยกกำลัง p

β_j แทนพารามิเตอร์ของตัวแปรอิสระตำแหน่งที่ j , $j = 1,...,p$

ε_i แทนความคลาดเคลื่อนตำแหน่งที่ i

n แทนขนาดตัวอย่าง

p แทนจำนวนพจน์พหุนามในสมการดังนี้ คือ 2 , 3 , 4 , 5 และ 6

และ ค่าลำดับชั้น (rank) ของเมทริกซ์ X เท่ากับ $p+1$; $((p+1) < n)$

2. ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและมีการแจกแจงเดียวกัน ที่มีค่าเฉลี่ย $E(\varepsilon) = 0$ และเมทริกซ์ความแปรปรวนร่วม $E(\varepsilon\varepsilon') = \sigma^2 I_n$

3. กำหนดให้ $\beta = (\beta_0 , \beta_1 , \dots , \beta_p)'_{(p+1)\times 1}$ เป็นค่าใด ๆ ในการวิจัยครั้งนี้กำหนดให้ $\beta' = (1 , 1 , \dots , 1)_{(p+1)\times 1}$ ในประชากรทุกรูปแบบที่ศึกษา โดยที่ p เป็นจำนวนตัวพยากรณ์ (จำนวนพจน์พหุนามในตัวแบบโดยไม่นับพจน์ค่าคงที่) สำหรับสร้างตัวแปรตาม

4. ขนาดตัวอย่าง (n) ที่ใช้ 15 , 30 , 60 , 120 และ 240

5. กำหนดตัวแปรอิสระเริ่มต้น x_i ที่ใช้สร้างสมการพหุนามมีการแจกแจงเป็น $N(\mu,\sigma^2)$ โดยค่าพารามิเตอร์ μ , σ^2 เป็นค่าใด ๆ ในที่นี้กำหนดให้ $\mu = 5$ และ $\sigma^2 = 4$

6. ในการวิจัยครั้งนี้ศึกษาเมื่อความคลาดเคลื่อนสุ่มในตัวแปรตามมีการแจกแจง $N(0,\sigma^2)$ โดยกำหนด $\sigma^2 = 4 , 6 , 8$ และ 10 เพื่อศึกษาว่าเมื่อความแปรปรวนเพิ่มขึ้นส่งผลต่อตัวประมาณหรือไม่

7. ตัวแปรตาม (y) ที่ใช้สำหรับคำนวณค่ารากของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ สร้างจากตัวแปรอิสระ (X) และความคลาดเคลื่อน (ε) โดยที่ตัวแปรตาม (y) มีการแจกแจงแบบ $N_n(X\beta , \sigma^2 I_n)$

8. ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์

1.5 คำจำกัดความที่ใช้ในการวิจัย

1. ตัวแปรตาม (dependent variable (y)) หมายถึง เรื่องหรือสิ่งที่ต้องการศึกษา

2. ตัวแปรอิสระ (independent variables (x)) หมายถึง สิ่งที่ส่งผลต่อตัวแปรตาม หรือมีผลกระทบต่อสิ่งที่ต้องการศึกษา

3. ความคลาดเคลื่อนยกกำลังสอง (MSE) หมายถึง ผลต่างระหว่างค่าพารามิเตอร์จริงและค่าพารามิเตอร์ที่ได้จากตัวประมาณค่าพารามิเตอร์

4. การวิเคราะห์ความถดถอยพหุนาม (Polynomial Regression Analysis) หมายถึง การวิเคราะห์สมความถดถอยเมื่อตัวแปรอิสระที่ใช้ในการวิเคราะห์นั้นเกิดจากการที่นำตัวแปรอิสระตัวก่อนหน้านั้นมาทำการยกกำลังเพิ่มขึ้นเท่าที่ต้องใช้ในการวิเคราะห์ในแต่ละรูปแบบ

1.6 ขั้นตอนในการดำเนินการวิจัย

1. ศึกษาโปรแกรม Delphi6 ซึ่งเป็นโปรแกรมที่จะใช้ในการวิเคราะห์ในครั้งนี้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. สร้างโปรแกรมที่ใช้ในการสร้างข้อมูลที่ใช้ในการทดสอบ โดยวิธีที่เลือกใช้สร้างข้อมูล คือ วิธีมอนติคาร์โล (Simulation by Monte Carlo Method)

3. ศึกษาวิธีการสร้างตัวประมาณที่ต้องการเปรียบเทียบ และเขียนโปรแกรมการสร้างตัวประมาณ ดังนี้

1) ตัวประมาณกำลังสองน้อยสุดสามัญ (OLS)

2) ตัวประมาณริดจ์สามัญ (ROLS)

3) ตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด (RLAV)

4. เขียนโปรแกรมสำหรับใช้เป็นเกณฑ์ในการตัดสินผลที่ได้จากข้อมูลที่ใช้ในการวิเคราะห์ในแต่ละตัวประมาณ

5. วิเคราะห์ผลจากข้อมูลที่ได้จากโปรแกรมที่เขียน

1.7 เกณฑ์ในการตัดสินใจ

เกณฑ์การตัดสินใจว่าตัวประมาณค่าพารามิเตอร์วิธีใดมีความถูกต้องมากที่สุดพิจารณาจากค่าเฉลี่ยของรากที่สองของค่าคลาดเคลื่อนกำลังสอง (average root mean squares error ($AMSE$)) และเปรียบเทียบประสิทธิภาพโดยพิจารณาเพิ่มเติมจากค่าอัตราส่วนผลต่างของรากของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (ratio of different average root mean squares error ($Diff$)) มีสูตรดังนี้

$$RMSE = \sqrt{\frac{\sum_{i=0}^{p}(\beta_{ij} - \hat{\beta}_{ij})^2}{p+1}}$$

$$AMSE = \frac{\sum_{j=1}^{1,000} RMSE_j}{1,000}$$

$$DIFF = \frac{AMSE_i - AMSE_{min}}{AMSE_{min}}$$

เมื่อ β_i แทนค่าจริงของพารามิเตอร์ตัวที่ i สมการความถดถอยในการจำลองรอบที่ j

$\hat{\beta}_i$ แทนค่าประมาณของพารามิเตอร์ตัวที่ i ในสมการความถดถอยในการจำลองรอบที่ j

$RMSE_j$ แทนรากของค่าคลาดเคลื่อนกำลังสองของตัวประมาณค่าพารามิเตอร์ที่ i ของการทำซ้ำรอบที่ j

$AMSE_i$ แทนค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ i

และ $AMSE_{min}$ แทนค่าเฉลี่ยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำสุดจากทั้ง 3 ตัวประมาณ

ผู้วิจัยพิจารณาว่าตัวประมาณค่าพารามิเตอร์ใดให้ค่า $AMSE$ และค่า $DIFF$ ต่ำสุด จะเป็นตัวประมาณค่าพารามิเตอร์ที่ดีที่สุด และมีประสิทธิภาพมากที่สุด

1.8 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการประมาณค่าสัมประสิทธิ์ของตัวแบบถดถอยพหุนามกรณีที่ใช้ตัวประมาณค่าสัมประสิทธิ์ที่สามารถแก้ปัญหาพหุสัมพันธ์ (multicolinearity) และทำให้ได้สมการที่ใช้พยากรณ์ค่าของตัวแปรตามที่มีความคลาดเคลื่อนลดลง