บทที่ 1

บทนำ

ความสำคัญและความเป็นมาของปัญหา

สำหรับการศึกษาความถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เมื่อข้อมูลถูกแบ่งเป็นกลุ่มที่มีรูปแบบทั่วไปเป็นดังนี้ คือ

$$y_{it} = \beta_{0i} + \beta_{1i} x_{it} + \varepsilon_{it} \qquad ; t=1,2,...,n_i \; ; \; i=1,2,3,4$$

โดยที่ x_{it} แทนตัวแปรอิสระที่มีค่าคงที่

y_{it} แทนตัวแปรตาม

β_{0i} และ β_{1i} แทนพารามิเตอร์ที่ไม่ทราบค่าแทนจุดตัดบนแกน y และความชันของเส้นถดถอยตามลำดับ

ε_{it} แทนความคลาดเคลื่อน

m แทนจำนวนกลุ่ม

n_i แทนขนาดตัวอย่างในแต่ละกลุ่ม ซึ่ง $\sum_{i=1}^{m} n_i = n$

โดยปกติแล้วการประมาณค่าและการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์นั้นผู้วิจัยมักจะเลือกใช้วิธีกำลังสองต่ำสุดสามัญ (Ordinary Least Square Method : OLS) เนื่องจากเป็นวิธีที่ให้ตัวประมาณมีคุณสมบัติเป็น Best Linear Unbiased Estimator : BLUE ทั้งนี้ต้องกระทำภายใต้ข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนดังนี้คือ ความคลาดเคลื่อนมีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่เป็น σ^2 และ e_t , e_k ไม่มีสหสัมพันธ์กัน สำหรับทุกค่า $t \neq k$ อย่างไรก็ตามในทางปฏิบัติบ่อยครั้งที่เราพบว่า ข้อมูลที่ถูกนำมาวิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว ในงานวิจัยนี้ จะศึกษาเฉพาะในกรณีที่ความแปรปรวนของความคลาดเคลื่อนไม่คงที่ เราเรียกสถานะการณ์เช่นนี้ว่า ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ (Heteroscedasticity) ซึ่งมักจะเกิดกับข้อมูลที่มีลักษณะเป็นข้อมูลภาคตัดขวาง (Cross Sectional Data) และพบมากกับข้อมูลทางด้านเศรษฐศาสตร์ และธุรกิจ ดังเช่นปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ที่เกิดขึ้นในการศึกษาระดับของค่าใช้จ่ายในการบริโภคของครอบครัวที่มีรายได้แตกต่างกันพบว่าความแปรปรวนของค่าใช้จ่ายจะไม่คงที่ แต่จะผันแปรค่าไปตามระดับของรายได้ของครอบครัว ผลกระทบตามมาก็คือตัวประมาณจากวิธีกำลังสองต่ำสุดสามัญ $\hat{\beta}_{0i}, \hat{\beta}_{1i}$ ของค่าพารามิเตอร์

β_{0i}, β_{1i} รวมตลอดถึงการตรวจสอบนัยสำคัญของ $\hat{\beta}_{0i}, \hat{\beta}_{1i}$ จะมีความคลาดเคลื่อนมากขึ้นเนื่องจาก $\hat{\beta}_{0i}, \hat{\beta}_{1i}$ ที่ได้จะไม่มีประสิทธิภาพ กล่าวคือ ความแปรปรวนของ $\hat{\beta}_{0i}, \hat{\beta}_{1i}$ จะสูงเกินไป ทำให้ขาดคุณสมบัติของ BLUE ซึ่งมีผลทำให้ได้ช่วงความเชื่อมั่นของ β_{0i}, β_{1i} กว้างเกินไป รวมทั้งมีผลทำให้ค่า t-test ต่ำกว่าความเป็นจริงก่อให้เกิดการยอมรับสมมติฐาน $H_0 : \beta_{1i}=0$ ได้ง่ายกว่าปกติ มีผลสืบเนื่องไปถึงการตัดตัวแปรอิสระที่มี β_{1i} เป็นสัมประสิทธิ์ทิ้งไปทั้ง ๆ ที่ตัวแปรอิสระนั้นอาจจะมีความสำคัญเป็นอย่างยิ่งก็ได้ สมการถดถอยที่ได้ก็จะเป็นสมการถดถอยที่มีคุณภาพต่ำ ดังนั้น นอกเหนือไปจากการประมาณค่าพารามิเตอร์ β_{0i}, β_{1i} ก็คือการตรวจสอบดูว่า ความคลาดเคลื่อนมีความแปรปรวนคงที่หรือไม่ เมื่อตรวจพบว่าความคลาดเคลื่อนมีความแปรปรวนไม่คงที่นักวิจัยควรทำการปรับปรุงแก้ไขโดยการใช้วิธีการอื่น ๆ ที่เหมาะสมกว่าวิธีกำลังสองต่ำสุดสามัญในการประมาณค่าพารามิเตอร์ β_{0i}, β_{1i}

การทดสอบสมมติฐานที่เกี่ยวกับความคลาดเคลื่อนมีความแปรปรวนไม่คงที่มีหลายวิธี เช่น การทดสอบด้วยอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio test) ,การทดสอบของบาร์ตเลตต์ (Bartlette's test) ในการวิจัยครั้งนี้จะทำการเปรียบเทียบวิธีการทดสอบ 3 วิธี คือ การทดสอบแจ็คไนฟ์ (Jackknife test) ,การทดสอบด้วยอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio test) และการทดสอบของบาร์ตเลตต์ (Bartlette's test) ซึ่งจะทำการทดลองภายใต้ความคลาดเคลื่อนมีการแจกแจงปกติ (Normal Distribution) และการแจกแจงปกติปลอมปน (Scale contaminated Normal Distribution)

วัตถุประสงค์ของการวิจัย

ในการวิจัยต้องการศึกษาเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ตรวจสอบปัญหาความคลาดเคลื่อนมีความแปรปรวนไม่คงที่ในการวิเคราะห์ความถดถอย 3 วิธีดังนี้

1. การทดสอบแจ็คไนฟ์ (Jackknife test)
2. การทดสอบด้วยอัตราส่วนภาวะน่าจะเป็น (Likelihood Ratio test)
3. การทดสอบของบาร์ตเลตต์ (Bartlette's test)

สมมติฐานของการวิจัย

ภายใต้ลักษณะการกระจายของข้อมูลที่ความคลาดเคลื่อนมีความแปรปรวนไม่คงที่การทดสอบแจ็คไนฟ์ จะมีค่าอำนาจการทดสอบมีค่าสูงสุดเมื่อความคลาดเคลื่อนมีการแจกแจงปกติ

ข้อตกลงเบื้องต้น

สมการถดถอยที่ใช้ในการศึกษาเป็นแบบ สมการถดถอยเชิงเส้นอย่างง่ายโดยมีรูปแบบเป็น

$$y_{it} = \beta_{0i} + \beta_{1i} x_{it} + \varepsilon_{it} \quad ; t=1,2,....n_i ; i=1,2,3,4$$

โดยที่ x_{it} คือตัวแปรอิสระของกลุ่มที่ i

y_{it} คือตัวแปรตามของกลุ่มที่ i

β_{0i}, β_{1i} คือพารามิเตอร์ของกลุ่มที่ i

ε_{it} คือความคลาดเคลื่อนอิสระที่มีค่าเฉลี่ยและความแปรปรวนไม่เท่ากันในแต่ละกลุ่มของข้อมูล โดยกำหนดให้ความแตกต่างของความแปรปรวนของความคลาดเคลื่อนในแต่ละกลุ่มแตกต่างกันเป็นอัตราส่วน 3 ระดับ คือ แตกต่างกันน้อย แตกต่างกันปานกลาง และแตกต่างกันมาก กำหนดความแตกต่างตามวิธีของ Games และ Probert (1972) โดยใช้ค่านอนเซนทรัลลิตี้พารามิเตอร์ หรือ ϕ (Noncentrality parameter) ซึ่งคำนวณจากสูตร

$$\phi^2 = \sum_{i=1}^{m} \frac{(\sigma_i^2 - \sigma^2)^2}{k\sigma_1^2}$$

เมื่อ σ^2 คือค่าเฉลี่ยเรขาคณิตของความแปรปรวน m กลุ่ม คำนวณจากสูตร

$$\sigma^2 = (\prod \sigma_m^2)^{\frac{1}{k}}$$

เป็นตัววัดความแตกต่างของความแปรปรวนดังต่อไปนี้

1. อัตราส่วนของความแปรปรวนแตกต่างกันน้อยมีค่า ϕ อยู่ระหว่าง (0,1.5)
2. อัตราส่วนของความแปรปรวนแตกต่างกันปานกลางมีค่า ϕ อยู่ระหว่าง (1.5,3.0)
3. อัตราส่วนของความแปรปรวนแตกต่างกันมากมีค่า ϕ มากกว่า 3.0

เกณฑ์ที่ใช้ในการเลือกสถิติทดสอบคือ ค่าความคลาดเคลื่อนประเภทที่ 1 (Type I error) และ อำนาจของการทดสอบ (Power of the test)

ขอบเขตการวิจัย

1. ศึกษาความแกร่งและอำนาจการทดสอบของการทดสอบแจ็คไนฟ์ ,การทดสอบด้วยอัตราส่วนภาวะน่าจะเป็น และการทดสอบของบาร์ตเลตต์ เมื่อสัมประสิทธิ์การถดถอยมีค่าเท่ากันและจำนวนข้อมูลในแต่ละกลุ่มเท่ากัน

2. สร้างค่าของ X จากการแจกแจงแบบปกติ (Normal Distribution) โดยแบ่งข้อมูลเป็นกลุ่ม ๆ แต่ละกลุ่มมีค่าเฉลี่ยและความแปรปรวนไม่เท่ากันโดยกำหนดให้สัมประสิทธิ์ของความแปรผัน (Coefficient of variation) = 0.01 ค่าเฉลี่ยและความแปรปรวนแตกต่างกัน ดังนี้

กลุ่มที่ 1 (i=1) ค่าเฉลี่ย = 200 ความแปรปรวน = 4.0

กลุ่มที่ 1 (i=2) ค่าเฉลี่ย = 300 ความแปรปรวน = 9.0

กลุ่มที่ 1 (i=3) ค่าเฉลี่ย = 400 ความแปรปรวน = 16.0

กลุ่มที่ 1 (i=4) ค่าเฉลี่ย = 500 ความแปรปรวน = 25.0

ได้ทำการทดลองเมื่อ X มีค่าสัมประสิทธิ์ของความแปรผันแตกต่างกันพบ ว่า ไม่มีผลต่อการหาค่าอำนาจการทดสอบของตัวสถิติทั้ง 3 วิธี จึงทำการกำหนดให้ค่าของ X มีค่าสัมประสิทธิ์ความแปรผันเท่ากับค่าดังกล่าวเพื่อความสะดวกในการคำนวณ

3. กำหนดค่าพารามิเตอร์ของสมการถดถอยเป็นดังนี้

$\beta_{0i} = 1$

$\beta_{1i} = 1$; i = 1,2,3,4

เนื่องจากได้ทำการทดลองแล้วพบว่าค่าพารามิเตอร์ของสมการถดถอยไม่มีผลต่อการทดสอบนี้ เมื่อทำการทดลองที่ค่าพารามิเตอร์ของสมการถดถอยมีค่าแตกต่างกันพบว่าผลการทดลองเป็นเช่นเดียวกับการกำหนดให้ค่าพารามิเตอร์ของสมการถดถอยเท่ากับค่าดังกล่าว ดังนั้นในการกำหนดค่าพารามิเตอร์เช่นนี้จึงมีจุดประสงค์เพื่อให้ง่ายต่อการคำนวณ

4. กำหนดการแจกแจงของความคลาดเคลื่อนในแต่ละกลุ่มเป็นดังนี้

4.1 การแจกแจงปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเป็น 0 และ 100 และความแปรปรวนแตกต่างกันเป็น 6 ระดับ

4.2 การแจกแจงปกติปลอมปน (Scale contaminated Normal Distribution) ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนแตกต่างกันเป็น 6 ระดับ เช่นเดียวกับการแจกแจงปกติโดยกำหนดค่าเปอร์เซ็นต์การปลอมปนเป็น 1 , 5 , 10 และค่าสเกลคอนทามิเนตเป็น 3 และ 10

โดยการกำหนดให้ความแปรปรวนของความคลาดเคลื่อนมีอัตราส่วนอย่างต่ำที่มีความแตกต่างกันดังนี้ (กลุ่ม 1: กลุ่ม 2: กลุ่ม 3: กลุ่ม 4)

ความแปรปรวนของความคลาดเคลื่อนแตกต่างกันเล็กน้อย

ชุดที่ 1 1 : 2 : 3 : 4 ค่า ϕ = 1.1540

ชุดที่ 2 1 : 2.5 : 3.5 : 4.5 ค่า ϕ = 1.3449

ความแปรปรวนของความคลาดเคลื่อนแตกต่างกันปานกลาง

ชุดที่ 3 1 : 3 : 4 : 5 ค่า ϕ = 1.5509

ชุดที่ 4 1 : 3 : 5 : 7 ค่า ϕ = 2.3745

ความแปรปรวนของความคลาดเคลื่อนแตกต่างกันมาก

ชุดที่ 5 1 : 4 : 7 : 10 ค่า ϕ = 3.6380

ชุดที่ 6 1 : 5 : 10 : 15 ค่า ϕ = 5.8328

5. ศึกษาในกรณีที่ขนาดตัวอย่างเท่ากันทุกกลุ่ม ดังนี้

$n = 40 , 80 , 120$

$n_i = 10 , 20 , 30$ ตามลำดับ

เมื่อ n แทนขนาดตัวอย่างทั้งหมด

n_i แทนขนาดตัวอย่างในแต่ละกลุ่ม $i = 1,2,3,4$

6. สมมติฐานสำหรับการทดสอบ

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \sigma_4^2 \neq \sigma^2$

7. กำหนดระดับนัยสำคัญเป็น 0.05 , 0.01

8. ในการทดลองจะจำลองการทดลองขึ้นโดยอาศัยเทคนิคมอนติคาร์โลจากเครื่องคอมพิวเตอร์ Acer Power PT560c Pentium Processor โดยจะกระทำซ้ำกัน 500 ครั้ง ในแต่ละสถานะการณ์ของการทดลอง

คำจำกัดความ

1. ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (Probability of Type I Error) คือ ความน่าจะเป็นที่เกิดจากการปฏิเสธ H_0 เมื่อสมมติฐาน H_0 จริง

2. ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 (Probability of Type II Error) คือ ความน่าจะเป็นที่เกิดจากการยอมรับสมมติฐาน H_0 เมื่อสมมติฐาน H_0 ไม่จริง

3. อำนาจการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐาน H_0 เมื่อสมมติฐาน H_0 ไม่จริง

ประโยชน์ของการวิจัย

1. เพื่อให้ผู้ใช้สถิติมีข้อสรุปและหลักฐานเกี่ยวกับสถิติที่ใช้ในการตรวจสอบปัญหาความแปรปรวนไม่คงที่ของความคลาดเคลื่อนในตัวแบบความถดถอยเชิงเส้นเมื่อข้อมูลถูกแบ่งเป็นกลุ่ม และสามารถเลือกใช้สถิติทดสอบได้อย่างถูกต้องและเหมาะสมกับลักษณะข้อมูล

2. เป็นแนวทางในการวิจัยเพื่อศึกษาเปรียบเทียบในกรณีที่สัมประสิทธิ์ของสมการถดถอยแตกต่างกันในแต่ละกลุ่ม

3. เป็นแนวทางในการวิจัยเพื่อศึกษาเปรียบเทียบในกรณีที่ความคลาดเคลื่อนมีการแจกแจงเป็นแบบอื่น ๆ

4. เพื่อนำตัวสถิติที่ได้เปรียบเทียบกับตัวสถิติอื่น ๆ ต่อไป