บทที่ 1

บทนำ

ความเป็นมาของปัญหา

ในปัจจุบันนี้ได้มีการพัฒนาทฤษฎีและเทคนิคใหม่ๆ ทางสถิติเพิ่มขึ้นมาเรื่อยๆ ทำให้ขอบข่ายของวิชาสถิติมีความกว้างขวางและมีความเกี่ยวข้องผูกพันกับศาสตร์แขนงอื่นๆ โดยเฉพาะอย่างยิ่งการวิเคราะห์การถดถอย (Regression analysis) ซึ่งเป็นเทคนิคการทำนายที่นิยมใช้กันอย่างกว้างขวางในสาขาวิชาต่างๆ การวิเคราะห์การถดถอยนั้นเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (Dependent variable) นิยมเขียนแทนด้วย y กับตัวแปรอิสระ (Independent variable) นิยมเขียนแทนด้วย x โดยมีวัตถุประสงค์ที่จะประมาณหรือพยากรณ์ค่าตัวแปรตาม โดยรูปแบบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระอาจสัมพันธ์กันในลักษณะเชิงเส้นตรงหรือเส้นโค้งในรูปแบบใดรูปแบบหนึ่งก็ได้ และรูปแบบโดยทั่วไปของสมการถดถอยที่มีความสัมพันธ์ในเชิงเส้นตรง เป็นดังนี้

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{Y}$ แทน เวกเตอร์ของตัวแปรตาม ขนาด $n\times 1$; $\underset{\sim}{Y}=(y_1 \quad y_2 \quad \cdots \quad y_n)'$

$\underset{\sim}{X}$ แทน เมทริกซ์ของตัวแปรอิสระ ขนาด $n\times(k+1)$

$$\underset{\sim}{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}$$

$\underset{\sim}{\beta}$ แทน เวกเตอร์พารามิเตอร์ของสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า ขนาด $(k+1)\times 1$; $\underset{\sim}{\beta} = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_k)'$

$\underset{\sim}{\varepsilon}$ แทน เวกเตอร์ของความคลาดเคลื่อนที่มี ขนาด $n\times 1$; $\underset{\sim}{\varepsilon} = (\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n)'$

k แทน จำนวนตัวแปรอิสระที่ใช้ในการศึกษา และ

n แทน ขนาดตัวอย่างที่ใช้ในการศึกษา

ซึ่งมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อน โดยความคลาดเคลื่อนเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น 0 หรือ $E(\varepsilon_i)=0$ และมีความแปรปรวนเท่ากันหรือ $V(\varepsilon_i)=\sigma^2$ และความคลาดเคลื่อนเป็นอิสระกันหรือ $Cov(\varepsilon_i,\varepsilon_j)=0$ เมื่อ $i\neq j$ โดยที่ $i,j=1,2,\ldots,n$ ซึ่งสามารถเขียนโดยรวมได้ดังนี้ $\varepsilon_i \sim NID(0,\sigma^2),\quad i=1,2,..,n$

ส่วนในกรณีที่ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเป็นเส้นโค้ง รูปแบบการถดถอยแบบเส้นโค้งมีหลายรูปแบบ เช่น รูปแบบการถดถอยพหุนาม (Polynomial regression) เป็นรูปแบบของการถดถอยเชิงเส้นในเทอมของ β จึงสามารถนำหลักการของรูปแบบการถดถอยเชิงเส้นพหุคูณมาประยุกต์ใช้ได้ โดยการถดถอยพหุนามอันดับที่ k มีรูปแบบทั่วไปดังนี้

$$y_i = \beta_0+\beta_1 x_i+\beta_2 x_i^2+\ldots+\beta_k x_i^k+\varepsilon_i \qquad ; \quad i = 1,2,\ldots,n$$

ในการปฏิบัติงานจริงบางครั้ง มักพบปัญหาที่เป็นสาเหตุทำให้ข้อตกลง $Cov(x_i,\varepsilon_i)=0$ ไม่เป็นจริง เพราะตัวแปรอิสระจะมีค่าคลาดเคลื่อนเกิดขึ้นซึ่งเป็นเหตุการณ์ที่ขัดแย้งกับข้อตกลงสมการถดถอยที่กำหนดไว้ว่า ตัวแปรอิสระต้องไม่มีค่าคลาดเคลื่อนทั้งจากการวัดหรือจากการสังเกต ซึ่งมีผลทำให้ตัวประมาณที่ได้จากวิธีกำลังสองน้อยที่สุด (Ordinary least square: OLS) ขาดคุณสมบัติของการเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (The best linear unbiased estimator: BLUE) และทำให้การประมาณค่าด้วยวิธีกำลังสองน้อยที่สุดขาดความแม่นยำและมีประสิทธิภาพต่ำ (ทรงศิริ แต้สมบัติ, 2548) ความคลาดเคลื่อนที่เกิดขึ้นในตัวแปรอิสระสามารถเกิดขึ้นได้ในสถานการณ์ต่างๆ อาทิ ในสถานการณ์ที่ไม่อาจนำค่าตัวแปรที่แท้จริงมาใช้ได้ เพราะไม่อาจวัดค่าได้ทำให้ต้องเลือกใช้ตัวแปรทดแทนจากการใช้ตัวแปรดัมมี่เป็นตัวแปรอิสระ หรือจากการใช้เลขดัชนีแทนที่ค่าที่แท้จริงของตัวแปร เป็นต้น ซึ่งวิธีการประมาณค่าพารามิเตอร์โดยทั่วไปไม่ได้คำนึงถึงความคลาดเคลื่อนในตัวแปรอิสระ โดยความคลาดเคลื่อนในตัวแปรอิสระเป็นสาเหตุให้ค่าของตัวประมาณที่ได้ไม่ใกล้เคียงกับความจริงและเป็นผลทำให้เกิดความเอนเอียง (Bias) และไม่คงเส้นคงวา (Inconsistency) ของตัวประมาณสัมประสิทธิ์การถดถอย (มนตรี พิริยะกุล, 2545) เมื่อพิจารณากรณีที่มีความคลาดเคลื่อนในตัวแปรอิสระในตัวแบบการถดถอยพหุนาม รูปแบบการถดถอยพหุนามเป็นดังนี้

$$y_i = \beta_0 + \beta_1 x_i^* + \beta_2 x_i^{*2} + \ldots + \beta_k x_i^{*k} + \varepsilon_i \qquad ; \quad i = 1,2,\ldots,n$$

หรือ

$$y_i = \beta_0 + \beta_1\left(x_i + u_i\right) + \beta_2\left(x_i + u_i\right)^2 + \ldots + \beta_k\left(x_i + u_i\right)^k + \varepsilon_i \qquad ; \quad i = 1,2,\ldots,n$$

โดยความคลาดเคลื่อนในตัวแปรอิสระคือ $x_i^* = x_i + u_i \quad ; \quad i = 1,2,\ldots,n$
เมื่อ x_i^* เป็นตัวแปรแฝงในตัวแปรอิสระ และ u_i เป็นความคลาดเคลื่อนในตัวแปรอิสระ

การวิเคราะห์การถดถอยสิ่งสำคัญเป็นอย่างมากคือ การสร้างตัวแบบให้มีความเหมาะสมเพื่อให้ได้ค่าประมาณหรือค่าพยากรณ์ที่ใกล้เคียงกับค่าจริงมากที่สุด ซึ่งตัวแบบที่เหมาะสมนั้นควรสร้างมาจากวิธีประมาณค่าพารามิเตอร์ที่เหมาะสม และนักสถิติหลายท่านได้เสนอวิธีการแก้ไขปัญหากรณีตัวแบบการถดถอยพหุนามมีความคลาดเคลื่อนในตัวแปรอิสระ เช่น วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (Weighted least square: WLS) เป็นวิธีประมาณค่าพารามิเตอร์ที่นำค่าความแปรปรวนมาใช้ในการถ่วงน้ำหนักกับสัมประสิทธิ์การถดถอย ทำให้ข้อมูลมีความคงเส้นคงวามากขึ้น (Huang and Huwang, 2001)

Huang and Huwang (2001) ศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยพหุนามอันดับ 2 เมื่อมีความคลาดเคลื่อนในตัวแปรอิสระภายใต้ข้อมูลมีการแจกแจงแบบปกติ โดยแบ่งการศึกษาออกเป็น 2 กรณี ในกรณีที่ 1 กำหนด $\beta_0 = 5$, $\beta_1 = 2$ และ $\beta_2 = 3$ ขนาดตัวอย่างเท่ากับ 100, 200, 400 และ 1,000 ค่า σ_u^2 มีค่าเท่ากับ 0.5 ค่า σ_ε^2 มีค่าเท่ากับ 0.5 และ 1 โดยวิธีประมาณค่า BGLS (Best generalized least square) และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก ในกรณีที่ 2 กำหนด $\beta_0 = 1$, $\beta_1 = 1$ และ $\beta_2 = 0.1$, 0.8 และ 0.5 ขนาดตัวอย่างเท่ากับ 20, 50, 100 และ 500 ค่า σ_u^2 มีค่าเท่ากับ 1 ค่า σ_ε^2 มีค่าเท่ากับ 0.3 โดยวิธีกำลังสองน้อยที่สุด (OLS) วิธีกำลังสองน้อยสุดปรับปรุง (Adjust least square: ALS) และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) โดยทั้ง 2 กรณี ใช้ค่าความคลาดเคลื่อนกำลังเฉลี่ย (Mean square error: MSE) เป็นเกณฑ์ในการเปรียบเทียบ จากผลการศึกษาทั้ง 2 กรณี พบว่า วิธีประมาณค่าพารามิเตอร์วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) ให้ค่าความคลาดเคลื่อนกำลังเฉลี่ย (MSE) น้อยกว่าวิธี BGLS วิธีกำลังสองน้อยที่สุด (OLS) และวิธีกำลังสองน้อยสุดปรับปรุง (ALS)

ธิฏิรัตน์ เมฆบัณฑิตกุล (2546) ศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยพหุนามอันดับ 2, 3, 4, 5 และ 6 เมื่อมีความคลาดเคลื่อนในตัวแปรอิสระภายใต้ข้อมูลมีการแจกแจงแบบปกติ ขนาดตัวอย่างเท่ากับ 15, 30, 50, 100, และ 200 โดย

วิธีกำลังสองน้อยที่สุด (OLS) วิธีกำลังสองน้อยสุดปรับปรุง (ALS) และวิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (Average relative root mean square error: ARSE) จากผลการศึกษาพบว่า วิธีประมาณค่าพารามิเตอร์วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) ให้ค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (ARSE) น้อยกว่าวิธีกำลังสองน้อยที่สุด (OLS) และวิธีกำลังสองน้อยสุดปรับปรุง (ALS)

สำหรับวิธีกำลังสองน้อยสุดทางโครงสร้าง (Strcutural least square: SLS) เป็นวิธีประมาณค่าพารามิเตอร์ที่มีการประยุกต์ใช้ครั้งแรกโดย ทาเมอร์รัส ในปี ค.ศ. 1997 (Thamerus, 1997) ที่ใช้หลักการการปรับตัวแปรอิสระด้วยอัตราส่วนความน่าเชื่อถือของตัวแปรอิสระ (Reliability ratio of x^*: r_{x*}) ทำให้ค่าของตัวแปรอิสระใกล้เคียงกับค่าจริงมากขึ้น แต่การคำนวณค่าอัตราส่วนความน่าเชื่อถือของตัวแปรอิสระมีความยุ่งยากและซับซ้อน ต่อมาในปี ค.ศ. 2000 คูคลัสและชิไนวิส (Kukush and Schneeweiss, 2000) ได้นำเสนอขั้นตอนที่ซับซ้อนน้อยลง

Kukush, Schneeweiss and Wolf (2000) ศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยพหุนามอันดับ 2 และพหุนามอันดับ 4 เมื่อมีความคลาดเคลื่อนในตัวแปรอิสระ ภายใต้ข้อมูลมีการแจกแจงแบบปกติ กำหนด $\beta_0 = 0$, $\beta_1 = 2$ และ $\beta_2 = 3$ ขนาดตัวอย่างเท่ากับ 900 โดยวิธีกำลังสองน้อยที่สุด (OLS) วิธีกำลังสองน้อยสุดปรับปรุง (ALS) และวิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าความเอนเอียง (Bias) และค่าส่วนเบี่ยงเบนมาตรฐาน (Standard deviation: SD) จากผลการศึกษาพบว่า วิธีประมาณค่าพารามิเตอร์วิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) ให้ค่าความเอนเอียง (Bias) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) น้อยกว่าวิธีกำลังสองน้อยที่สุด (OLS) และวิธีกำลังสองน้อยสุดปรับปรุง (ALS)

ณหทัย สระกบแก้ว (2549) ศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยพหุนามอันดับ 2, 3, 4, 5 และ 6 เมื่อมีความคลาดเคลื่อนในตัวแปรอิสระ ภายใต้ข้อมูลมีการแจกแจงแบบปกติ และการแจกแจงแบบเอ็กซ์โปเนนเชียล ขนาดตัวอย่างเท่ากับ 20, 50, 100 และ 200 โดยวิธีกำลังสองน้อยที่สุด (OLS) วิธีกำลังสองน้อยสุดปรับปรุง (ALS) และวิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (ARSE) จากผลการศึกษา พบว่า วิธีประมาณค่าพารามิเตอร์วิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) ให้ค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อน

กำลังสองเฉลี่ยสัมพัทธ์ (ARSE) น้อยกว่าวิธีกำลังสองน้อยที่สุด (OLS) และวิธีกำลังสองน้อยสุดปรับปรุง (ALS)

จากการศึกษางานวิจัยข้างต้นจะเห็นว่า วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) และวิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) สามารถประมาณค่าพารามิเตอร์ได้อย่างมีประสิทธิภาพในกรณีที่มีความคลาดเคลื่อนในตัวแปรอิสระ ดังนั้นผู้วิจัยจึงสนใจศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (OLS) วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก (WLS) และวิธีกำลังสองน้อยสุดทางโครงสร้าง (SLS) ในตัวแบบการถดถอยพหุนามเมื่อมีความคลาดเคลื่อนเกิดขึ้นในตัวแปรอิสระ เกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์จะพิจารณาจากค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ (ARSE) โดยวิธีประมาณค่าที่ให้ค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ต่ำที่สุด จะเป็นวิธีประมาณค่าที่มีประสิทธิภาพมากที่สุด

วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์การถดถอยพหุนามเมื่อมีความคลาดเคลื่อนในตัวแปรอิสระ 3 วิธี คือ วิธีกำลังสองน้อยที่สุด วิธีกำลังสองน้อยสุดถ่วงน้ำหนัก และวิธีกำลังสองน้อยสุดทางโครงสร้าง

ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดขอบเขตการศึกษาไว้ดังนี้

1. ตัวแบบการถดถอยพหุนามที่ใช้ในการศึกษา
2. กำหนดขนาดตัวอย่างมีค่าเท่ากับ 15, 30, 50, 100, 150 และ 200
3. กำหนดตัวแปรอิสระที่ศึกษา คือ 1 ตัว โดยสร้างจากตัวแปรแฝง (x) และความคลาดเคลื่อนในตัวแปรอิสระ (u) โดยตัวแปรแฝงมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย $\mu =$ 0 และความแปรปรวน $\sigma^2 = 1$ และความคลาดเคลื่อนในตัวแปรอิสระมีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ย $\mu = 0$ และความแปรปรวน $\sigma^2 =$ 0.1, 0.25, 0.5, 0.75 และ 1
4. กำหนดความคลาดเคลื่อนอย่างสุ่มมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย $\mu = 0$ และความแปรปรวน $\sigma^2 =$ 0.1, 0.25, 0.5, 0.75 และ 1
5. จำลองข้อมูลด้วยเทคนิคมอนติคาร์โลในแต่ละสถานการณ์ กระทำซ้ำจำนวน 1,000 ครั้ง

เกณฑ์การตัดสินใจ

เกณฑ์การตัดสินใจ สำหรับวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุนามเมื่อมีความคลาดเคลื่อนในตัวแปรอิสระ จะพิจารณาจากค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ โดยวิธีประมาณค่าที่มีค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ต่ำที่สุดจะเป็นวิธีประมาณค่าที่มีประสิทธิภาพมากที่สุด ซึ่งมีรายละเอียดดังนี้

$$ARSE = \frac{\sum_{j=1}^{1000} RSE_j}{1000}$$

โดยที่

$$RSE_j = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_{ij} - \hat{y}_{ij}\right)^2}{n}}}{\bar{y}_j} \times 100\%$$

เมื่อ y_{ij} แทนค่าสังเกตที่ i ในการทำซ้ำรอบที่ j

$\hat{y}_{ij}$ แทนค่าพยากรณ์ที่ i ในการทำซ้ำรอบที่ j

$\bar{y}_j$ แทนค่าเฉลี่ยของค่าสังเกต n ค่าในการทำซ้ำรอบที่ j

n แทนขนาดตัวอย่าง

RSE_j แทนรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์ของการทำซ้ำรอบที่ j

$ARSE$ แทนค่าเฉลี่ยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยสัมพัทธ์จากการทำซ้ำ 1,000 รอบ

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการเลือกใช้วิธีประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุนาม เมื่อมีความคลาดเคลื่อนในตัวแปรอิสระ

2. เพื่อเป็นแนวทางในการศึกษาวิธีประมาณค่าพารามิเตอร์ สำหรับการแจกแจงรูปแบบอื่น