บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันนี้ได้มีการนำความรู้ทางด้านสถิติไปประยุกต์ใช้กับงานต่าง ๆ เป็นอันมาก โดยเฉพาะงานวิจัยในสาขาวิทยาศาสตร์ สังคมศาสตร์และเศรษฐศาสตร์ ทั้งนี้ เนื่องมาจากวิธีการทางสถิติเป็นวิธีดำเนินการที่เป็นระบบ ซึ่งสามารถช่วยในการวิเคราะห์เพื่อหาคำตอบสำหรับงานวิจัยนั้น ๆ ได้ โดยเฉพาะอย่างยิ่ง การหาคำตอบเพื่อคาดคะเนเหตุการณ์ล่วงหน้าหรือการพยากรณ์ ซึ่งผู้วิจัยมักจะเลือกใช้วิธีการวิเคราะห์ความถดถอย (regression analysis)

การวิเคราะห์ความถดถอยพหุเป็นกรณีหนึ่งของการวิเคราะห์ความถดถอยเชิงเส้น เมื่อข้อมูลที่ใช้ศึกษามีความสัมพันธ์กับปัจจัยอื่น ๆ มากกว่า 1 ตัว ความสัมพันธ์ของข้อมูลที่ใช้ศึกษาหรือเรียกว่าตัวแปรตาม กับข้อมูลปัจจัยอื่น ๆ หรือเรียกว่าตัวแปรอิสระ สามารถเขียนให้อยู่ในรูปตัวแบบเชิงเส้น

(1.1) $$Y_i = \beta_o+\beta_1 X_{i1} +\ldots+\beta_m X_{im}+\varepsilon_i \ , \ i = 1,2,\ldots,n$$

ซึ่งสามารถเขียนแทนด้วยเมตริกซ์ดังนี้

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

เมื่อ $\underset{\sim}{Y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$, โดยที่ n เป็นจำนวนค่าสังเกต

X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (m+1)$ ซึ่งได้รวมเวกเตอร์ $\underset{\sim}{1}$ ของสัมประสิทธิ์การถดถอย β_o ด้วย

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอย ขนาด $(m+1) \times 1$

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของค่าผิดพลาดที่เกิดขึ้นขนาด $n \times 1$ ซึ่งมีการแจกแจงแบบปกติ

ในการประมาณค่าสัมประสิทธิ์ความถดถอยพหุของตัวแบบดังกล่าว วิธีการที่นิยมนำมาใช้กันมากคือ วิธีกำลังสองน้อยที่สุด (Least Square method) ซึ่งวิธีการนี้สามารถประมาณค่า

สถิติที่มีคุณสมบัติที่ดี ค่าประมาณของ β ที่ได้คือ $\hat{\beta} = (XX)^{-1} X'y$ ค่าประมาณ $\hat{\beta}$ นี้มีคุณสมบัติของความไม่เอนเอียงเชิงเส้น และมีความแปรปรวนต่ำสุดในบรรดาตัวประมาณที่ไม่เอนเอียงเชิงเส้นทั้งหลาย

แต่อย่างไรก็ตาม เมื่อค่าสังเกตมาจากการแจกแจงของประชากรซึ่งไม่เป็นแบบปกติบางชนิด เช่น มีหางค่อนข้างหนา และหางยาวกว่าปกติ (heavy tailed, long tailed) วิธีกำลังสองน้อยที่สุดอาจจะไม่เหมาะสม เนื่องจากวิธีนี้มีความไวต่อข้อมูลที่ผิดปกติ (outliers) และเกิดการสูญเสียประสิทธิภาพไปเมื่อการแจกแจงของความผิดพลาดไม่เป็นแบบปกติตามข้อสมมติ นั่นคือ วิธีกำลังสองน้อยที่สุดจะไม่แกร่ง (non robust) เนื่องจากอิทธิพลของข้อมูลที่ผิดปกติปรับสมการประมาณการถดถอยไปในทิศทางของมันด้วย

ในปี ค.ศ. 1964 Huber ได้ศึกษาถึงตัวประมาณที่แกร่ง ซึ่งเรียกว่า M-estimator ชุดของตัวประมาณที่แกร่งชนิด M-estimator อยู่ในรูปของค่าน้อยที่สุดของฟังก์ชันของค่าผิดพลาด ซึ่งสามารถเขียนได้ดังนี้

$$\min_{\beta} \sum_{i=1}^{n} \rho\,(\varepsilon_i/s) = \min_{\beta} \sum_{i=1}^{n} \rho\left[(y_i - X_i'\beta)/s\right]$$

เมื่อ ε_i เป็นค่าผิดพลาดของค่าสังเกตที่ i และ s เป็นค่าที่เหมาะสมสำหรับการกระจายของ ε_i สูตรที่เหมือนกันในรูปของ $\psi = \rho'$ คือ

$$\sum_{i=1}^{n} x_{ij}\psi \left|(y_i - X_i'\beta)/s\right. = 0$$

สำหรับฟังก์ชัน $\rho(\varepsilon_i)$ ของวิธีกำลังสองน้อยที่สุดนั้นจะอยู่ในรูปของ ε_i^2 โดยค่า s ไม่จำเป็นต้องใช้เนื่องจาก $\rho(\varepsilon_i)$ เป็นฟังก์ชันที่มีความแปรปรวนเพียงค่าเดียวเท่านั้น (homogeneous) และผลลัพธ์ของการประมาณ β ไม่แปรเปลี่ยนตาม s จากการศึกษาของ Andrews et cil. (ค.ศ. 1972) ตัวประมาณของวิธีกำลังสองน้อยที่สุดไม่มีประสิทธิภาพมากเมื่อเทียบกับตัวประมาณชนิดนอนลิเนียร์ (nonlinear) โดยที่ความผิดพลาดมีการแจกแจงเป็นแบบ คอชชี (Cauchy), ดับเบิ้ลเอ็กซ์โปเนนเชียล (double exponential) และแบบปกติ

ปลอมปน (Scale-contaminated normal distribution) นอกจากนี้ M-estimator ไม่เพียงแต่จะมีประสิทธิภาพสำหรับค่าผิดพลาดที่มีการแจกแจงแบบหางยาวเท่านั้น แต่จะสูญเสียประสิทธิภาพไปเพียงเล็กน้อยเมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบปกติ

การวิเคราะห์ปัญหาของการประมาณในด้านตำแหน่ง (location) ของ Andrews et al. ในปี ค.ศ. 1972 และการทำงานทางด้านทฤษฎีของ Hampel ในปี ค.ศ. 1971 และ ค.ศ. 1974 ได้แนะนำว่าจุดวิกฤตในแง่ของ M-estimator คือ พฤติกรรมของ ψ ที่มีต่อค่าผิดปกติของกลุ่มของตัวอย่าง การประมาณในกรณี ψ ไม่มีขอบเขตของ ε_i มีแนวโน้มที่จะไม่แกร่ง แต่การประมาณในกรณีที่ ψ มีขอบเขตของ ε_i โดยที่ ψ ไม่เป็นศูนย์สำหรับค่า ε_i ที่ใหญ่ มีแนวโน้มที่จะแกร่งสำหรับสัดส่วนเพียงเล็กน้อยของค่าผิดปกติ (outliers) ในขณะที่การประมาณสำหรับ ψ ซึ่งลู่เข้าสู่ศูนย์แสดงถึงความแกร่งสำหรับสัดส่วนที่ใหญ่ของค่าผิดปกติ

Ramsay (ค.ศ. 1977) ได้พิจารณาฟังก์ชันของ ρ และ ψ สำหรับชุดของ M-estimator อีกวิธีหนึ่งคือ $\rho(\varepsilon_i/s) = a^{-2}\left[1-\exp(-a|\varepsilon_i|/s).(1+a|\varepsilon_i|/s)\right]$

และ $\psi(\varepsilon_i/s) = (\varepsilon_i/s)\exp(-a|\varepsilon_i|/s)$

ซึ่งอ้างถึงตัวประมาณนี้ด้วย E_a เมื่อ $a = 0.3$ ตัวประมาณ $E_{0.3}$ ของ Ramsay มีขอบเขตที่ $|\varepsilon_i|/s = 1/a$ ฟังก์ชันของ ρ และ ψ ทำให้ค่าผิดปกติมีอิทธิพลลดลงอย่างสม่ำเสมอ ค่าผิดปกติที่มีค่ามาก ๆ จะอยู่ในสัดส่วนที่ถูกตัดออกจากตัวอย่าง

ในการศึกษาวิจัยนี้สนใจศึกษาการประมาณสัมประสิทธิ์การถดถอย เมื่อความผิดพลาดมีการแจกแจงเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ โดยจะศึกษาเปรียบเทียบการประมาณ 2 วิธีคือ วิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay $E_{0.3}$ ในสถานการณ์ต่าง ๆ โดยใช้การแจกแจงแบบปกติปลอมปน และแบบที

นอกจากนี้ เรายังสนใจกรณีที่ค่าผิดพลาดมีการแจกแจงเป็นแบบเบ้ (skewed distribution) ซึ่งในกรณีนี้วิธีกำลังสองน้อยที่สุดอาจจะไม่เป็นวิธีที่ดีที่สุดสำหรับการประมาณสัมประสิทธิ์การถดถอยพหุ เนื่องจากค่าผิดพลาดไม่มีการแจกแจงเป็นแบบปกติตามข้อสมมุติของวิธีกำลังสองน้อยที่สุด การศึกษานี้ได้อาศัยการแปลงที่ใช้การยกกำลัง (power transformation) ของ Box และ Cox (ค.ศ. 1964) มาใช้สำหรับการแปลงข้อมูลให้มีการแจกแจงเข้าสู่ภาวะปกติ

การแปลงที่ใช้การยกกำลังของ Box และ Cox ในตัวแบบเชิงเส้น เขียนได้เป็น

$$Y^{(\lambda)} = X\beta + \varepsilon \quad , \quad \varepsilon \sim IN(0, \sigma^2 In)$$

เมื่อ $Y_i^{(\lambda)}$ เป็นตัวแปรที่ถูกแปลงแล้ว λ เป็นพารามิเตอร์ของการแปลง โดย

$$Y_i^{\lambda} = \begin{cases} \dfrac{Y_i^{\lambda}-1}{\lambda} & ; \quad \lambda \neq 0 \\ \log_e Y_i & ; \quad \lambda = 0 \end{cases}$$

การแปลงที่ใช้การยกกำลังของ Box และ Cox จะประมาณพารามิเตอร์ λ ซึ่งหาได้จากวิธีกำลังสองน้อยที่สุด เมื่อ $\sigma^2_{\varepsilon|\hat{\lambda}}$ มีค่าน้อยที่สุด ซึ่ง $\sigma^2_{\varepsilon|\lambda}$ เป็นผลรวมกำลังสองของค่าผิดพลาดของการถดถอยของ $Y^{(\lambda)}$ บน X

อนึ่ง ในการศึกษาวิจัยนี้ได้นำเอาวิธี M-estimator มาศึกษาเปรียบเทียบกับวิธีกำลังสองน้อยที่สุดในการประมาณพารามิเตอร์ λ และหา $\sigma^2_{\varepsilon|\hat{\lambda}}$ ที่มีค่าน้อยที่สุดด้วย ภายหลังจากการแปลงข้อมูลแล้ว ε อาจจะมีการแจกแจงเข้าภาวะปกติ ในกรณีที่ ε มีการแจกแจงไม่เป็นแบบปกติ คือ เป็นแบบหางยาว วิธีกำลังสองน้อยที่สุดอาจจะเป็นวิธีการประมาณที่ไม่เหมาะสม เราจะทำการทดสอบความเป็นปกติของการแจกแจงของ $Y_i^{(\lambda)}$ หลังจากการแปลงด้วยโดยใช้วิธีทดสอบความเป็นปกติของ Shapiro และ Wilk สำหรับการแจกแจงแบบเบ้จะศึกษารูปแบบการแจกแจง ลอกนอร์มอล, แกมมา และไวบูลล์

1.2 วัตถุประสงค์ของการวิจัย

1. ศึกษาวิธีการประมาณสัมประสิทธิ์การถดถอยพหุด้วยกระบวนการถดถอยที่แกร่ง (robust regression procedure) เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ
2. ศึกษาวิธีการประมาณสัมประสิทธิ์การถดถอยพหุด้วยกระบวนการแปลงข้อมูลและวิธีการถดถอยที่แกร่งเมื่อความผิดพลาดมีการแจกแจงแบบเบ้

3. ศึกษาเปรียบเทียบประสิทธิภาพของการประมาณสัมประสิทธิ์การถดถอยพหุ ด้วยวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

1.3 ข้อตกลงเบื้องต้น

1. ค่าผิดพลาด (ε_i) เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระซึ่งกันและกัน

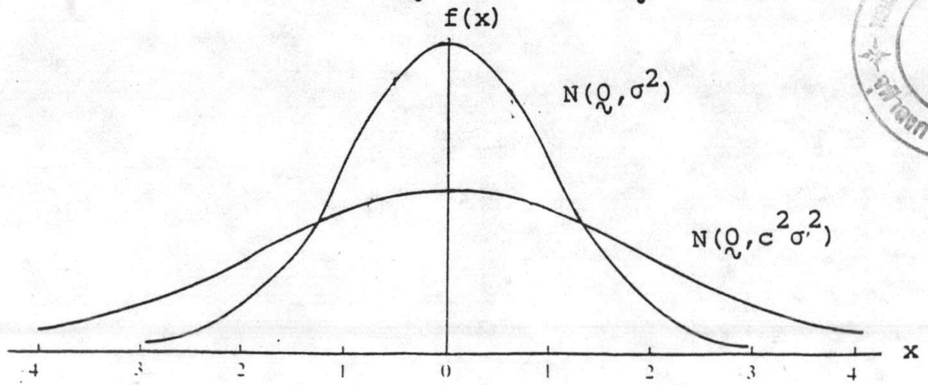
2. การวิจัยครั้งนี้ถือว่า วิธีการประมาณสัมประสิทธิ์การถดถอยพหุภายใต้ลักษณะการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้ ซึ่งให้ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (average of relative mean square error (ARMSE) และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง (average of absolute value of different ratio of mean square error (AADRM) ของการประมาณสูงโดยเฉลี่ย จะเป็นวิธีที่เหมาะสมสำหรับแต่ละสถานการณ์

1.4 ขอบเขตของการวิจัย

1. เมื่อค่าผิดพลาด มีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ จะศึกษาในกรณีของ

1.1) <u>การแจกแจงแบบปกติปลอมปน</u> (scale-contaminated normal distribution)

โดยที่ $F = (1-p)\ N(\underset{\sim}{0}\ ,\ \sigma^2) + p\ N(\underset{\sim}{0}, c^2\ \sigma^2)$

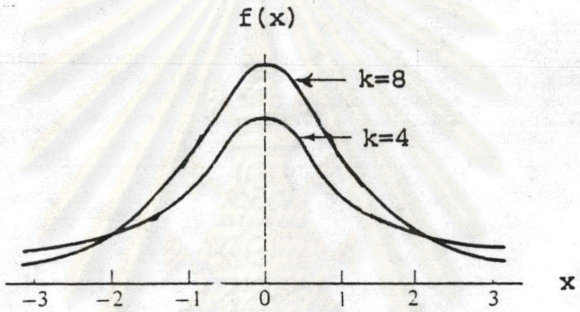


เมื่อ c เป็นค่าสเกลแฟคเตอร์ (scale factor) ถ้าค่าสเกลแฟคเตอร์มีค่าสูงจะทำให้เกิดค่าสังเกตที่ผิดปกติมีค่าสูงด้วย เราจะใช้ $c = 3$ และ $c = 10$

p เป็นเปอร์เซนต์ของการปลอมปน (percent of contamination) เราจะใช้ $p = 1, 5, 10$ และ 25

1.2) <u>การแจกแจงแบบที</u> (t distribution)

โดย $$t = \frac{X}{\sqrt{Y/k}}$$



เมื่อ X มีการแจกแจงเป็นแบบปกติ N(0, 1) Y มีการแจกแจงแบบไคสแควร์ โดยที่ X และ Y เป็นตัวแปรที่เป็นอิสระซึ่งกันและกัน

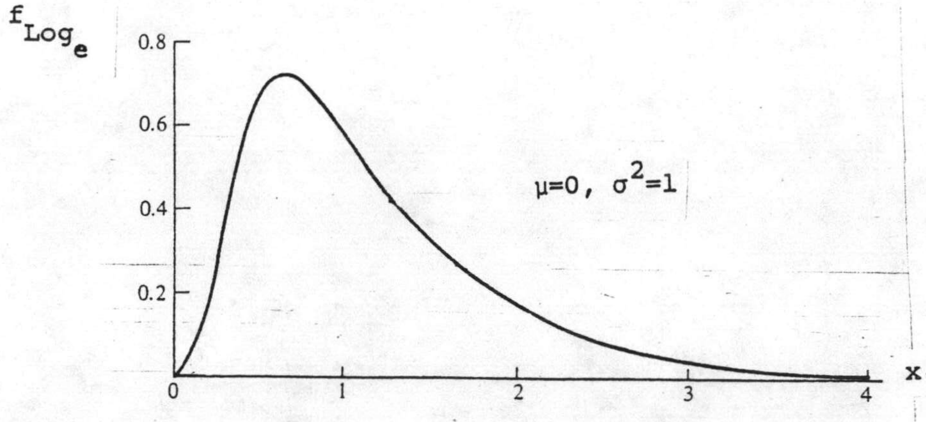
k เป็นระดับความเป็นอิสระ (ร.ส.) โดยศึกษาที่ $k = 4, 8$

2. <u>เมื่อค่าผิดพลาดมีการแจกแจงเป็นแบบเบ้</u> (skewed distribution) จะศึกษาในกรณี

2.1 <u>การแจกแจงแบบลอกนอร์มอล</u> (Lognormal distribution) พังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(X) = \begin{cases} \dfrac{1}{X\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\log_e X-\mu)/\sigma^2} & ; X>0, \sigma>0, -\infty<\mu<\infty \\ 0 & ; \text{อื่น ๆ} \end{cases}$$

เมื่อ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y โดยที่ $Y = \log_e X$ และ Y มีการแจกแจงแบบปกติ ในกรณีนี้จะศึกษาโดยใช้



$$E(x) = \exp\{\mu + \frac{\sigma^2}{2}\}$$

$$V(x) = \exp\{2\mu + \sigma^2\} . \{\exp\{\sigma^2\} - 1\}$$

$$CV(x) = \sqrt{\exp\{\sigma^2\} - 1}$$

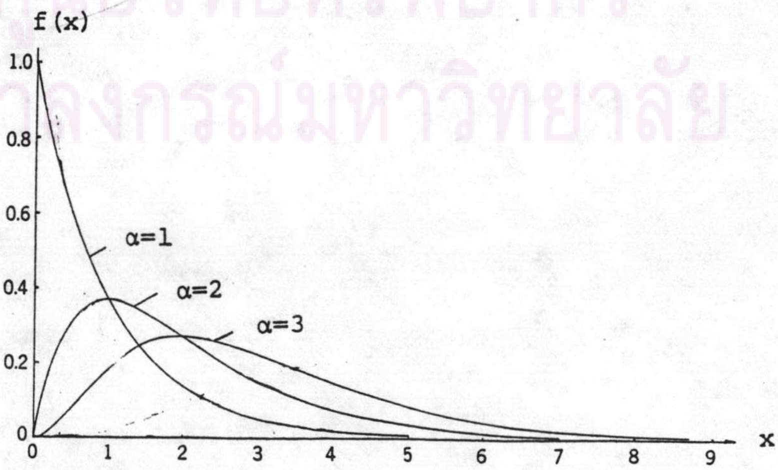
2.2) <u>การแจกแจงแบบแกมมา</u> (Gamma distribution)

พังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(X) = \begin{cases} \dfrac{X^{\alpha-1}\exp\{-X/\beta\}}{\beta^{\alpha}\,\Gamma(\alpha)} & ; \ 0<X \ , \ \alpha>0 \ , \beta>0 \\ 0 & ; \ \text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น scale parameter

α เป็น shape parameter



แสดงเส้นโค้งของการแจกแจงแบบแกมมา ณ $\beta=1$, $\alpha = 1,2,3$

$$E(x) = \beta\alpha$$

$$V(x) = \beta^2\alpha$$

$$CV(x) = \frac{1}{\sqrt{\alpha}}$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\beta = 1$, $\alpha = 1, 2, 3$ และ $\beta = 150$, $\alpha = 10$

$CV(x) = 100\%$ $(\beta = 1, \alpha = 1)$

$CV(x) = 70\%$ $(\beta = 1, \alpha = 2)$

$CV(x) = 58\%$ $(\beta = 1, \alpha = 3)$

และ $CV(x) = 32\%$ $(\beta = 150, \alpha = 10)$

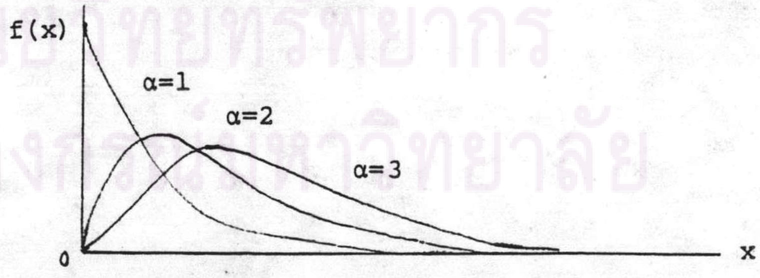
2.3) <u>การแจกแจงแบบไวบูลล์</u> (Weibull distribution)

พังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(X) = \begin{cases} \dfrac{\alpha X^{\alpha-1} \exp\{-X/\beta\}^{\alpha}}{\beta^{\alpha}} & ; \quad 0<X \ , \alpha>0 \ ,\beta>0 \\ 0 & ; \quad \text{อื่น ๆ} \end{cases}$$

เมื่อ β เป็น scale parameter

α เป็น shape parameter



แสดงเส้นโค้งของการแจกแจงแบบไวบูลล์ ณ $\beta = 1, \alpha = 1,2,3$

$$E(x) = \beta\Gamma(1+ \frac{1}{\alpha})$$

$$V(x) = \beta^2\left[\Gamma(1+ \frac{2}{\alpha}) - \Gamma^2(1+ \frac{1}{\alpha})\right]$$

$$CV(x) = \left[\frac{\Gamma(1+2/\alpha)}{\Gamma^2(1+1/\alpha)} -1 \right]^{\frac{1}{2}}$$

ในการวิจัยครั้งนี้จะศึกษาที่ $\beta = 1$, $\alpha = 1, 2, 3$ และ $\beta = 150$, $\alpha = 10$

$CV(x) = 100\%$ $(\beta = 1, \alpha = 1)$

$CV(x) = 52\%$ $(\beta = 1, \alpha = 2)$

$CV(x) = 35\%$ $(\beta = 1, \alpha = 3)$

และ $CV(x) = 17\%$ $(\beta = 150, \alpha = 10)$

3. จำนวนตัวแปรอิสระและขนาดตัวอย่าง

จำนวนตัวแปรอิสระ $m = 3$ จะใช้ขนาดตัวอย่าง $n = 20$

จำนวนตัวแปรอิสระ $m = 5$ และ 10 ใช้ขนาดตัวอย่าง $n = 50, 100$ และ 150

4. จำลองประชากรที่ศึกษาจากตัวแบบเชิงเส้น

4.1) เมื่อการแจกแจงของความผิดพลาดเป็นแบบปกติปลอมปนและแบบที จะจำลองประชากรจากตัวแบบ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมตริกซ์ของ X จะคงที่ในตัวแบบสำหรับการแจกแจงของความผิดพลาดที่กำลังศึกษาทั้งหมด โดยเมตริกซ์ X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (m+1)$ ซึ่งรวมเวกเตอร์ $\underset{\sim}{1}$ ของ intercept ด้วย การสร้างเมตริกซ์ X จะอาศัยการจำลองตัวแปรสุ่มที่มีการแจกแจงแบบปกติ $N(\mu,\sigma^2)$ เพื่อให้คล้ายกับข้อมูลตามธรรมชาติ $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุของประชากรนำมาจากสมการการถดถอยเชิงพหุที่ให้ค่าสหสัมพันธ์ร่วมของ X และ $\underset{\sim}{y}$ สูง $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความผิดพลาดโดยมีการแจกแจงตามที่ต้องการศึกษา

4.2) เมื่อการแจกแจงของความผิดเป็นแบบเบ้ ตัวแปร $\underset{\sim}{y}$ จะสร้างให้มีการแจกแจงเป็นแบบเบ้โดยตรง เมตริกซ์ X และเวกเตอร์ของ $\underset{\sim}{\beta}$ จะสร้างในทำนองเดียวกับข้อ 4.1

1.5 <u>ประโยชน์ที่คาดว่าจะได้รับ</u>

1. ผลจากการศึกษากระบวนการถดถอยที่แกร่งและการแปลงที่ใช้การยกกำลังของ Box และ Cox จะเป็นแนวทางในการศึกษาการประมาณที่แกร่งและการทดสอบที่แกร่งในสถิติหัวข้ออื่น ๆ ต่อไป

2. ผลจากการศึกษาเปรียบเทียบสามารถบอกได้ว่า ถ้ามีข้อมูลอยู่ชุดหนึ่งซึ่งมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ ควรจะใช้วิธีใดในการประมาณสัมประสิทธิ์การถดถอยพหุจึงจะทำให้ผลการประมาณมีค่าผิดพลาดน้อยที่สุด

3. ผลจากการศึกษาเปรียบเทียบสามารถบอกได้ว่า ถ้ามีข้อมูลอยู่ชุดหนึ่ง ซึ่งค่าผิดพลาดมีการแจกแจงแบบเบ้ ควรใช้วิธีใดในการประมาณสัมประสิทธิ์การถดถอยพหุ จึงจะทำให้ผลการประมาณมีค่าผิดพลาดน้อยที่สุด

1.6 วิธีดำเนินการวิจัย

1. ศึกษาวิธีการประมาณสัมประสิทธิ์การถดถอยพหุในกระบวนการถดถอยพหุและเขียนโปรแกรมจำลองค่าสังเกตของตัวแปรในตัวแบบที่ต้องการศึกษา และเขียนโปรแกรมสำหรับคำนวณค่าสัมประสิทธิ์การถดถอยพหุ ของวิธีการแต่ละวิธีดังนี้

   1.1) วิธีกำลังสองน้อยที่สุด

   1.2) วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

   1.3) กระบวนการแปลงข้อมูลของ Box และ Cox

2. ศึกษาเปรียบเทียบวิธีการประมาณสัมประสิทธิ์การถดถอยพหุ ด้วยวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองขึ้นในเครื่องคอมพิวเตอร์ โดยใช้เทคนิค Monte Carlo simulation และจะทำซ้ำ 200 ครั้ง ในแต่ละสถานการณ์ ยกเว้นกรณีเมื่อค่าผิดพลาดมีการแจกแจงแบบแกมมาและไวบูลล์ เมื่อใช้ $\beta = 150$, $\alpha = 10$ จะทำซ้ำ 100 ครั้ง

1.7 คำศัพท์ต่าง ๆ ที่ใช้ในการวิจัย

Outliers หมายถึง ค่าสังเกตที่ตรวจสอบหรือคิดว่าอยู่นอกกลุ่มข้อมูล

Robust หมายถึง อานุภาพการทดสอบหรือความยาวของช่วงความเชื่อมั่นคงที่เป็นกระบวนการที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยภายนอกที่ไม่อยู่ภายใต้การทดสอบ

Long-tailed distribution หรือ heavy-tailed distribution หมายถึง การแจกแจงซึ่งมีฟังก์ชัน การแจกแจงความน่าจะเป็น $f(x)$ เข้าใกล้ 0 ในอัตราที่ช้ากว่าฟังก์ชันการแจกแจงความน่าจะเป็นแบบปกติ เมื่อ X เข้าใกล้ $-\infty$ และ/หรือ $+\infty$

Scale contaminated normal distribution หมายถึง การแจกแจงที่สร้างขึ้นเพื่อให้มีลักษณะเป็น long-tailed distribution โดยส่วนหนึ่งของประชากรมาจาก Normal $N(\mu, \sigma^2)$ และอีกส่วนหนึ่งมาจาก Normal $N(\mu, c^2\sigma^2)$

p contaminated of scale factor หมายถึง ส่วนหนึ่งของประชากร scale contaminated normal distribution มาจากประชากร $N(\mu,\sigma^2)$ ด้วย ความน่าจะเป็น 1-p และอีกส่วนหนึ่งมาจากประชากร $N(\mu,c^2\sigma^2)$ ด้วยความน่าจะเป็น p

ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (Average of relative mean square error (ARMSE)) หมายถึงการเปรียบเทียบค่าเฉลี่ยความผิดพลาดกำลังสอง (MSE) ของการประมาณสัมประสิทธิ์ของการถดถอยพหุระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay ดังนี้

$$\text{ARMSE}^{1} = \begin{cases} \dfrac{MSE_M}{MSE_{OLS}} \times 100 & ; MSE_{OLS} < MSE_M \\ \dfrac{MSE_{OLS}}{MSE_M} \times 100 & ; MSE_M < MSE_{OLS} \end{cases}$$

---

[1] การคำนวณ ARMSE จะใช้ 2 สูตร เนื่องจากในกรณีที่ความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงแบบเบ้ วิธีกำลังสองน้อยที่สุดจะไม่เป็นวิธีมาตรฐานสำหรับใช้เปรียบเทียบกับวิธีอื่นได้โดยตรง ค่า ARMSE เป็นการคำนวณค่าเฉลี่ยตามจำนวนครั้งที่ได้จากการทดลองซึ่งวิธีใดวิธีหนึ่งให้ค่า MSE น้อยกว่า จากการทดลองกระทำซ้ำ 200 ครั้ง

ค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง (Average of absolute value of differrent ratio of mean square error (AADRM)) หมายถึงค่าแตกต่างของการเปรียบเทียบค่าเฉลี่ยความผิดพลาดกำลังสอง (MSE) ของการประมาณสัมประสิทธิ์การถดถอยพหุ ระหว่างวิธีกำลังสองน้อยที่สุด กับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay ดังนี้

$$\text{AADRM}^{1} = \begin{cases} \left| \dfrac{MSE_{OLS} - MSE_{M}}{MSE_{M}} \right| & ;\ MSE_{OLS} < MSE_{M} \\[2ex] \left| \dfrac{MSE_{M} - MSE_{OLS}}{MSE_{OLS}} \right| & ;\ MSE_{M} < MSE_{OLS} \end{cases}$$

---

1 การคำนวณ AADRM จะใช้ 2 สูตร เนื่องจากในกรณีที่ความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงแบบเบ้ วิธีกำลังสองน้อยที่สุดจะไม่เป็นวิธีมาตรฐานสำหรับใช้เปรียบเทียบกับวิธีอื่นได้โดยตรง ค่า AADRM เป็นการคำนวณค่าเฉลี่ยตามจำนวนครั้งที่ได้จากการทดลองซึ่งวิธีใดวิธีหนึ่งให้ค่า MSE น้อยกว่า จากการทดลองกระทำซ้ำ 200 ครั้ง