บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์ความถดถอยเป็นวิธีการทางสถิติที่นิยมใช้กันมากในงานวิจัยทั้งทางด้านสังคมศาสตร์และวิทยาศาสตร์ เราใช้เทคนิคนี้ในการคาดคะเนหรือพยากรณ์ค่าของตัวแปรที่สนใจ โดยอาศัยรูปแบบความสัมพันธ์ระหว่างตัวแปรที่ต้องการพยากรณ์ซึ่งเราเรียกว่าตัวแปรตาม (dependent variable) กับตัวแปรที่มีอิทธิพลกับตัวแปรตามซึ่งเราเรียกว่า ตัวแปรอิสระ (independent variables) โดยทั่วไปจะมีตัวแปรอิสระมากกว่าหนึ่งตัวแปร ซึ่งเรียกการถดถอยในกรณีนี้ว่าการวิเคราะห์ความถดถอยพหุคูณ (multiple regression analysis) และถ้าตัวแปรอิสระแต่ละตัวกับตัวแปรตามมีลักษณะความสัมพันธ์ในรูปแบบเชิงเส้น จะเรียกการวิเคราะห์ความถดถอยดังกล่าวว่า การวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (multiple linear regression) ซึ่งสามารถแสดงในรูปของสมการได้ดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \qquad ; \; i = 1,2,\ldots,n$$

หรือเขียนในรูปเมทริกซ์ได้เป็น

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

เมื่อ $\underset{\sim}{y}$ คือเวกเตอร์ของตัวแปรตามขนาด n×1

X คือเมทริกซ์ของตัวแปรอิสระขนาด n×(p+1)

$\underset{\sim}{\beta}$ คือเวกเตอร์ของค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยขนาด (p+1)×1

$\underset{\sim}{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อนขนาด n×1

n คือจำนวนค่าสังเกต

และ p คือจำนวนตัวแปรอิสระ

ในการวิเคราะห์ความถดถอยจะมีเงื่อนไขหรือข้อสมมติเบื้องต้น คือ

1. ความคลาดเคลื่อนมีการแจกแจงแบบปกติ
2. ความคลาดเคลื่อนมีค่าเฉลี่ยเป็นศูนย์ กล่าวคือ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$
3. ความคลาดเคลื่อนมีความแปรปรวนคงที่ กล่าวคือ $Var(\varepsilon_i) = \sigma^2 , \forall i$
4. ความคลาดเคลื่อนแต่ละค่าต้องเป็นอิสระซึ่งกันและกัน
   กล่าวคือ $cov(\varepsilon_i , \varepsilon_j) = 0 \ , i \neq j$
5. ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันเอง (multicollinearity)

ถ้าข้อมูลมีลักษณะสอดคล้องกับข้อตกลงเบื้องต้นแล้ว เราจะทำการประมาณค่าสัมประสิทธิ์ความถดถอย ซึ่งโดยปกติเราจะใช้วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS)) วิธีการนี้จะหาตัวประมาณ $\underset{\sim}{\beta}$ ที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนมีค่าน้อยที่สุด กล่าวคือตัวประมาณอยู่ในรูป $\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X' \underset{\sim}{y}$ และ $\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{p} \hat{\beta}_j \bar{x}_j$ ตัวประมาณนี้มีคุณสมบัติที่สำคัญคือเป็นตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator (B.L.U.E.)) ตัวประมาณ OLS จะเป็นตัวประมาณที่มีรูปแบบเดียวกันกับตัวประมาณความควรจะเป็นสูงสุด (Maximum Likelihood Estimator (M.L.E.)) ถ้า ε_i มีการแจกแจงปกติ (normal distribution)

ในทางปฏิบัตินั้นพบว่า การทำให้ข้อมูลมีข้อสมมติดังกล่าวเป็นไปได้ยาก ซึ่งหากข้อมูลไม่เป็นไปตามข้อสมมติดังกล่าวแล้ว จะทำให้การนำสมการถดถอยที่ได้ไปใช้พยากรณ์มีความผิดพลาดและมีความน่าเชื่อถือน้อย ปัญหาสำคัญอย่างหนึ่งที่นักวิจัยมักพบในการนำวิธีการวิเคราะห์ความถดถอยไปใช้ คือข้อมูลตัวอย่างมีขนาดเล็ก ทฤษฎีค่าจำกัดสู่ส่วนกลาง (Central Limit Theorem (C.L.T.)) กล่าวว่าเมื่อข้อมูลมีขนาดตัวอย่างใหญ่ขึ้นการแจกแจงของตัวอย่างสุ่มจะลู่เข้าสู่การแจกแจงปกติ ดังนั้นเมื่อตัวอย่างมีขนาดเล็กจึงส่งผลในเรื่องการแจกแจงของข้อมูลตัวอย่าง กล่าวคือการที่ตัวอย่างมีขนาดเล็กอาจทำให้ค่าความคลาดเคลื่อน (ε) มีการแจกแจงไม่เป็นปกติ (non-normality) ซึ่งกระทบต่อเงื่อนไขของการวิเคราะห์ความถดถอย เมื่อเกิดค่าที่ผิดปกติเช่นเกิดค่าที่สูงหรือต่ำจนเกินไป จะทำให้การแจกแจงของข้อมูลมีลักษณะเบ้หรือโด่งไปจากความเป็นจริงมาก ซึ่งถ้าความคลาดเคลื่อนไม่มีการแจกแจงเป็นปกติ การประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุดนั้นจะไม่เหมาะสม เพราะจะทำให้ตัวประมาณที่ได้ไม่มีคุณสมบัติเป็นตัวประมาณที่มีค่าความแปรปรวนต่ำสุดในบรรดาตัวประมาณที่ไม่เอนเอียง (Uniformly Minimum

Variance Unbiased Estimator (U.M.V.U.)) ซึ่งอาจทำให้เกิดข้อผิดพลาดได้มากขึ้นเมื่อนำไปวิเคราะห์ความถดถอย

ดังนั้นในงานวิจัยนี้จะพิจารณาประสิทธิภาพของตัวประมาณในกรณีที่ตัวอย่างมีขนาดแตกต่างกันซึ่งไม่ใหญ่นักโดยจะเริ่มศึกษาจากในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงเป็นปกติ จากนั้นจึงศึกษาในกรณีที่ข้อมูลมีค่าผิดปกติเกิดขึ้นหรือมีการแจกแจงไม่เป็นปกติ วิธีการประมาณที่ใช้คือวิธีการประมาณค่าสัมประสิทธิ์ที่เรียกว่าวิธีการถดถอยที่แกร่ง (robust regression method) ซึ่งเป็นวิธีการลดอิทธิพลของข้อมูลที่มีค่าความคลาดเคลื่อนผิดปกติหรือมีการแจกแจกไม่เป็นปกติลงตามเงื่อนไขของเกณฑ์ความแกร่ง (robust criterion) ในปี พ.ศ.2507 ฮิวเบอร์ (Huber) ได้ศึกษาตัวประมาณที่แกร่งซึ่งเรียกว่าตัวประมาณ M (M-estimator) โดยพัฒนาตัวประมาณดังกล่าวมาจากหลักการพื้นฐานของการประมาณแบบความควรจะเป็นสูงสุด (maximum likelihood estimation) ฮิวเบอร์ได้สร้างฟังก์ชันความแกร่งของฮิวเบอร์ (Huber's robust criterion function) ขึ้นมาเป็นฟังก์ชันของความคลาดเคลื่อน ต่อมาได้มีผู้ศึกษาฟังก์ชันความแกร่งอีกหลายท่าน ได้แก่ ในปี พ.ศ.2514 แฮมเพล (Hampel) ได้สร้างฟังก์ชันความแกร่งของแฮมเพล (Hampel's robust criterion function) ซึ่งฟังก์ชันนี้จะมีค่าเท่ากับศูนย์เมื่อค่าผิดปกติมีขนาดใหญ่ ปี พ.ศ.2515 แอนดริว (Andrew) ได้สร้างฟังก์ชันความแกร่งของแอนดริว (Andrew's robust criterion function) เป็นลักษณะฟังก์ชันคลื่นแบบไซน์ (sine function) ในปี พ.ศ.2517 ตูกี (Tukey) ได้สร้างฟังก์ชันความแกร่งของตูกี (Tukey's robust criterion function) อยู่ในรูปฟังก์ชันพหุนาม (polynomial function) ซึ่งจากการศึกษาพบว่าฟังก์ชันความแกร่งของตูกีเหมาะสำหรับข้อมูลที่ผิดปกติที่รุนแรงมากๆ และในปี พ.ศ.2520 แรมเซย์ (Ramsay) ได้สร้างฟังก์ชันความแกร่งของแรมเซย์ (Ramsay's robust criterion function) ซึ่งมีลักษณะเป็นฟังก์ชันชี้กำลัง (exponential function) นั่นคือฟังก์ชันความแกร่งของแรมเซย์ จะเข้าใกล้ศูนย์เมื่อค่าผิดปกติมีขนาดใหญ่เช่นเดียวกัน ในงานวิจัยนี้จึงได้ศึกษาตัวประมาณ M โดยใช้ฟังก์ชันความแกร่งของตูกีและแรมเซย์

นอกจากนี้ยังมีวิธีการประมาณค่าไม่อิงพารามิเตอร์ (Non-parametric estimation) เนื่องจากวิธีการดังกล่าวไม่จำเป็นต้องทราบลักษณะการแจกแจงของประชากร วิธีการประมาณที่จะใช้ในงานวิจัยนี้คือวิธีบูตสแตรป (Bootstrap) เป็นวิธีที่เสนอขึ้นโดยเบรดลี่ เอฟรอน (Bradley Efron) ในปี พ.ศ.2520 เนื่องจากพบว่าวิธีบูตสแตรปเป็นวิธีที่ให้ผลดีที่สุดในบรรดาวิธีไม่อิงพารามิเตอร์อื่นๆ เช่น วิธีแจคไนฟ์ (Jackknife) ของ Quenouille ในปี พ.ศ.2492 ,วิธี half-sampling หรือวิธี sub-sampling เป็นต้น เพราะการหาค่าประมาณด้วยวิธีบูตสแตรปนี้จะได้ตัว

ประมาณความควรจะเป็นสูงสุดไม่อิงพารามิเตอร์ (nonparametric maximum likelihood estimater) หลักเกณฑ์ของวิธีบูตสแตรป คือ ทำการสุ่มตัวอย่างจากข้อมูลที่เก็บรวบรวมมาแบบคืนที่ (with replacement) ขนาดเท่ากับจำนวนตัวอย่างหรือข้อมูลที่มีอยู่แล้วนั้น เพื่อสร้างข้อมูลชุดใหม่แล้วนำมาใช้ในการประมาณค่าพารามิเตอร์ที่สนใจ

ดังนั้นผู้วิจัยจึงสนใจที่จะทำการศึกษาการประมาณค่าด้วยวิธีกำลังสองน้อยสุด วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของตูกีและของแรมเซย์ และวิธีบูตสแตรป ในการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็ก

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาวิธีการประมาณพารามิเตอร์หรือสัมประสิทธิ์ความถดถอยในสมการการถดถอยเชิงเส้นพหุคูณ เมื่อตัวอย่างมีขนาดเล็ก โดยจะศึกษาวิธีการประมาณค่าด้วยวิธีกำลังสองน้อยสุด วิธีตัวประมาณ M ซึ่งใช้เกณฑ์ความแกร่งของตูกีและของแรมเซย์ และวิธีบูตสแตรป
2. เพื่อเปรียบเทียบวิธีการประมาณทั้งสี่วิธี โดยการเปรียบเทียบค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (average of mean square error (AMSE)) ของการประมาณค่าสัมประสิทธิ์ความถดถอยที่ได้จากแต่ละวิธี

1.3 สมมติฐานของการวิจัย

1. การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีตัวประมาณ M จะให้ตัวประมาณที่มีประสิทธิภาพสูงกว่าวิธีบูตสแตรป และวิธีกำลังสองน้อยสุด (เพราะตัวประมาณ M เป็นตัวประมาณที่มีความแกร่งและเป็นตัวประมาณอิงพารามิเตอร์)
2. เนื่องจากวิธีบูตสแตรปเป็นวิธีการที่ไม่อิงพารามิเตอร์ ดังนั้นเมื่อขนาดตัวอย่างน้อย ตัวประมาณที่ได้จากวิธีบูตสแตรปน่าจะมีประสิทธิภาพสูงกว่าวิธีตัวประมาณ M

1.4 ข้อตกลงเบื้องต้น

1. ค่าความคลาดเคลื่อน (ε_i) มีการแจกแจงเหมือนกันและเป็นอิสระซึ่งกันและกัน (i.i.d)
2. ตัวแปรอิสระเป็นอิสระซึ่งกันและกันและเป็นอิสระจากค่าความคลาดเคลื่อน
3. ค่าความคลาดเคลื่อนมีความแปรปรวนคงที่
4. ขนาดตัวอย่างต้องมากกว่าจำนวนตัวแปรอิสระ กล่าวคือ $n > p$

1.5 ขอบเขตของการวิจัย

1. ศึกษาวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงเส้นพหุคูณและหาค่าความแปรปรวนของตัวประมาณด้วยวิธีกำลังสองน้อยสุด วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของตูกีและของแรมเซย์ และวิธีบูตสแตรป
2. จำนวนตัวแปรอิสระที่ใช้ศึกษา p = 3 , 6
3. ขนาดตัวอย่างที่ศึกษา (n) มีค่าเท่ากับ 10 , 15 , 20 , 25 และ 30 ตามลำดับ*
4. ลักษณะการแจกแจงของค่าความคลาดเคลื่อนที่ศึกษา มีดังนี้

4.1 การแจกแจงปกติ (Normal distribution) มีฟังก์ชันความน่าจะเป็น คือ

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

ในที่นี้จะศึกษาในกรณีที่ค่าเฉลี่ย (μ) = 0 และ 1 สำหรับค่าความแปรปรวนจะกำหนดจากค่าสัมประสิทธิ์การแปรผัน (Coefficient Variation ($C.V.$)) โดยที่

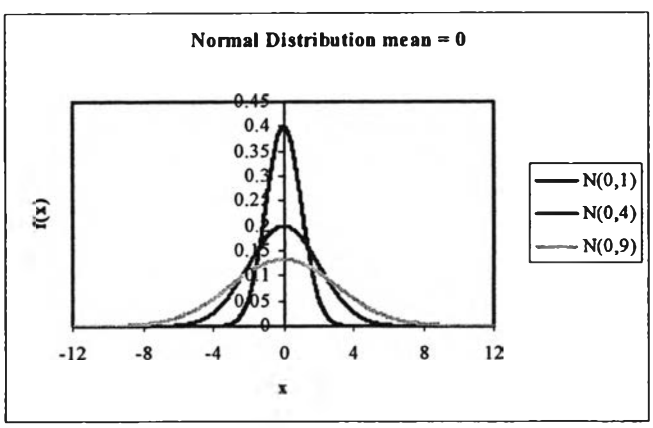
$\sigma^2 = 1$ $\quad C.V.(X) = 100\%$

$\sigma^2 = 4$ $\quad C.V.(X) = 200\%$

$\sigma^2 = 9$ $\quad C.V.(X) = 300\%$

เมื่อความแปรปรวนมีค่ามากขึ้น การกระจายจะมีค่าสูงขึ้น แต่จะมีความโด่งน้อยลง

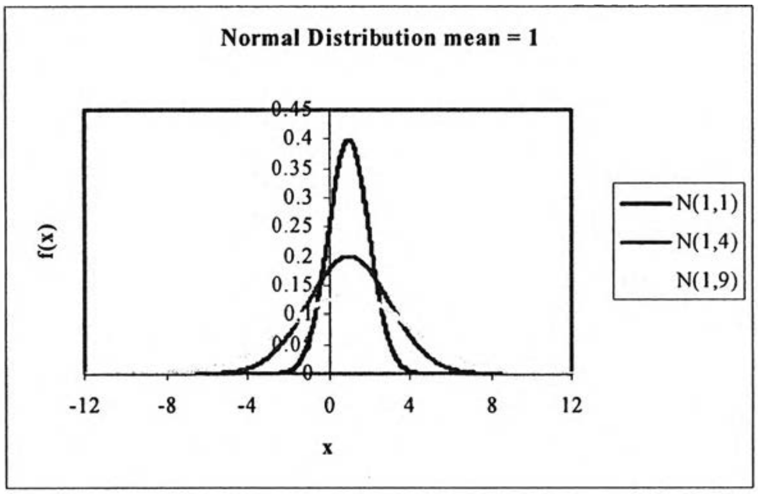
รูปที่ 1.1 การแจกแจงปกติเมื่อค่าเฉลี่ยเท่ากับ 0



---

* การกำหนดตัวอย่างขนาดเล็กควรกำหนดให้ขนาดตัวอย่างมีค่าน้อยกว่า 50+8p เมื่อ p คือจำนวนตัวแปรอิสระ ตามทฤษฎีของกรีน (Green) ในปี พ.ศ.2534 หรือมีค่าน้อยกว่าจำนวนตัวแปรอิสระอย่างน้อย 5 เท่าตามทฤษฎีของเรคอฟ (Raykov) ในปี พ.ศ.2538 ซึ่งในที่นี้มีจำนวนตัวแปรอิสระมากที่สุดเท่ากับ 6 จึงได้กำหนดขนาดตัวอย่างให้มีขนาดน้อยกว่า 6 × 5 = 30 ตามทฤษฎีของเรคอฟ

รูปที่ 1.2 การแจกแจงปกติเมื่อค่าเฉลี่ยเท่ากับ 1



4.2 <u>การแจกแจงสมมาตร</u> (Symmetric distribution) ในที่นี้จะใช้ **การแจกแจงปกติปลอมปน (Scale-contaminated normal distribution)** มีฟังก์ชันความน่าจะเป็นที่แปลงมาจากการแจกแจงปกติ คือ

$$f(x) = (1 - p) \cdot X + p \cdot Y$$

เมื่อ X มีการแจกแจงเป็นปกติด้วยค่าเฉลี่ย μ และความแปรปรวน σ^2

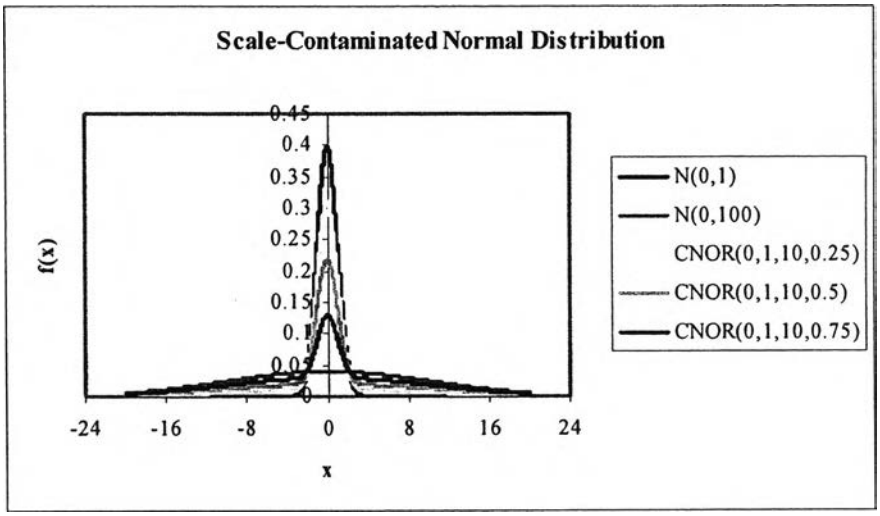
Y มีการแจกแจงเป็นปกติด้วยค่าเฉลี่ย μ และความแปรปรวน $C^2\sigma^2$; $C > 0$

p คือค่าเปอร์เซ็นต์การปลอมปน

และ C คือค่าเสกลแฟคเตอร์

โดยในที่นี้จะศึกษากรณีที่ค่าเฉลี่ย (μ) = 0 ค่าความแปรปรวน (σ^2) = 1 มีค่าสเกลแฟคเตอร์ (C) เท่ากับ 10 และมีค่าเปอร์เซ็นต์การปลอมปน (p) 3 ระดับเป็น 25%, 50% และ 75% ตามลำดับ ซึ่งเมื่อค่าเปอร์เซนต์การปลอมปนสูงขึ้นจะทำให้เกิดค่าผิดปกติสูง

รูปที่ 1.3 การแจกแจงปกติปลอมปน



4.3 <u>การแจกแจงเบ้</u> (Skewed distribution) ในที่นี้จะใช้ **การแจกแจงแกมมา (Gamma distribution)** มีฟังก์ชันความน่าจะเป็น คือ

$$f(x;\alpha,\beta)=\frac{1}{\beta\cdot\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\cdot e^{-\frac{x}{\beta}} \qquad ; x\geq 0,\alpha>0,\beta>0$$

โดยจะศึกษาในกรณีที่ β = 1 และ α = 0.5, 1, 2 และ 4 เนื่องจากความเบ้ของการแจกแจงขึ้นอยู่กับค่า α ดังนั้นจึงกำหนดให้ค่า β คงที่ ซึ่งจะพิจารณาจากค่าสัมประสิทธิ์การแปรผัน ดังนี้

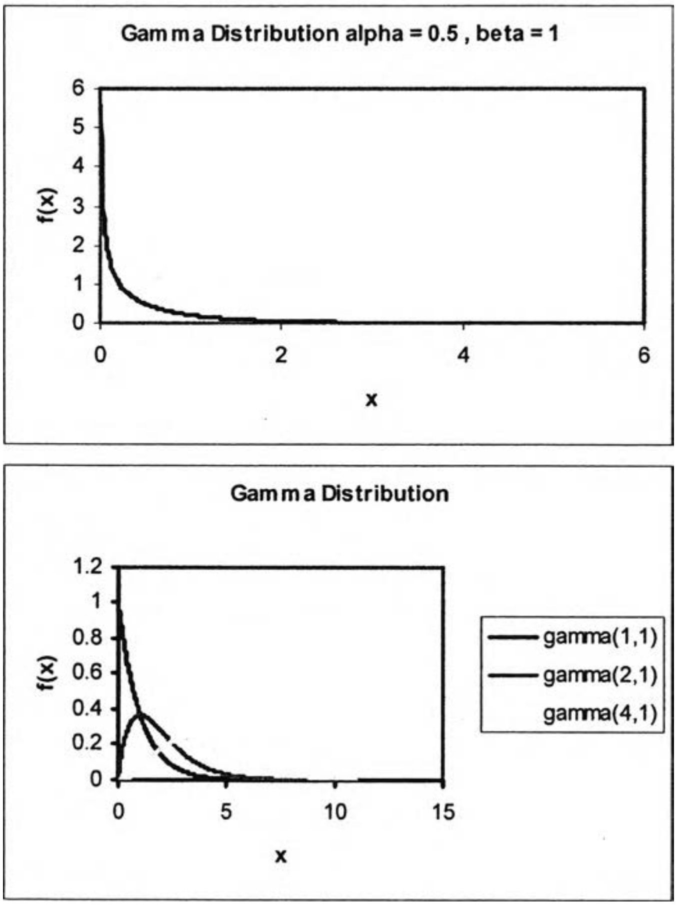
$$E(X)=\beta\alpha \qquad C.V.(X)=141\% \quad (\beta=1\,,\alpha=0.5)$$

$$Var(X)=\beta^2\alpha \qquad C.V.(X)=100\% \quad (\beta=1\,,\alpha=1)$$

$$C.V.(X)=\frac{1}{\sqrt{\alpha}} \qquad C.V.(X)=70\% \quad (\beta=1\,,\alpha=2)$$

$$C.V.(X)=50\% \quad (\beta=1\,,\alpha=4)$$

รูปที่ 1.4 การแจกแจงแกมมา



ถ้าค่า $C.V.$ มากข้อมูลจะยิ่งมีการกระจายมากขึ้นเมื่อตัวอย่างมีขนาดเล็ก และถ้ามีค่าผิดปกติเกิดขึ้นการกระจายจะยิ่งสูงมาก ในที่นี้กำหนดให้ค่า $C.V.(X) = 141\%$ เป็นค่าที่สูงที่สุดเพื่อที่จะได้เห็นชัดเจนว่าตัวประมาณใดเหมาะสมกับข้อมูลที่มีค่าผิดปกติมากที่สุด

1.6 เกณฑ์ที่ใช้ในการตัดสินใจ

ในงานวิจัยนี้จะใช้ค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (average of mean square error (AMSE)) เป็นเกณฑ์ที่ใช้ในการตัดสินใจ โดยมีสูตรในการคำนวณคือ

$$AMSE = \frac{1}{1,000}\sum_{i=1}^{1,000} MSE_i$$

เมื่อ $MSE_i = \dfrac{\sum_{j=0}^{p}\left(\hat{\beta}_j - \beta_j\right)^2}{p+1}$

โดยที่ β_j คือสัมประสิทธิ์การถดถอยที่กำหนดตัวที่ j

$\hat{\beta}_j$ คือค่าประมาณของสัมประสิทธิ์การถดถอยตัวที่ j

p คือจำนวนตัวแปรอิสระ6

และ MSE_j คือค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำซ้ำรอบที่ j

นอกจากนี้ยังใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (ratio of different average mean square error (RDAMSE)) เป็นค่าที่ใช้ประกอบการตัดสินใจซึ่งมีสูตรในการคำนวณคือ

$$RDAMSE_i = \frac{(AMSE_i - AMSE_{\min})}{AMSE_{\min}} \times 100\%$$

เมื่อ $AMSE_i$ เป็นค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ i

และ $AMSE_{\min}$ เป็นค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของวิธีที่มีค่าต่ำสุดในสถานการณ์นั้นๆ

1.7 วิธีดำเนินการวิจัย

งานวิจัยนี้ได้แบ่งวิธีดำเนินการออกเป็นขั้นตอนหลักๆ ดังนี้

1. ศึกษาวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงเส้นพหุคูณ ได้แก่ วิธีกำลังสองน้อยสุด วิธีตัวประมาณ M และวิธีบูตสแตรป
2. สร้างโปรแกรมย่อยสำหรับการแจกแจงของความคลาดเคลื่อนตามที่กำหนดไว้ในขอบเขตของการวิจัย
3. สร้างข้อมูลเพื่อใช้ในการวิจัย โดยในงานวิจัยครั้งนี้ได้ข้อมูลจากการจำลองด้วยเทคนิคมอนติ-คาร์โล (Monte Carlo technique) กระทำซ้ำ 1,000 ครั้ง
4. คำนวณหาค่า MSE เพื่อใช้ในการเปรียบเทียบตัวประมาณดังกล่าว พร้อมทั้งบันทึกผลเพื่อนำไปทำสรุปการวิจัย

1.8 คำจำกัดความต่างๆที่ใช้ในการวิจัย

1. การประมาณค่า (Estimation) หมายถึงการใช้ข้อมูลจากตัวอย่างสุ่มที่มีอยู่มาประมาณค่าหรือสรุปผลเกี่ยวกับค่าพารามิเตอร์ว่ามีค่าเป็นอย่างไร ซึ่งการประมาณค่ามีทั้งการประมาณค่าแบบจุด (point estimation) คือการประมาณค่าที่ได้ค่าประมาณเป็นตัวเลขซึ่งคำนวณได้จากตัวอย่างเพียงค่าเดียว และการประมาณค่าแบบช่วง (interval estimation) คือการประมาณค่าที่ได้ค่าประมาณเป็นช่วงของตัวเลขในช่วงใดช่วงหนึ่ง

2. ความแกร่งของตัวประมาณ (Robustness) หมายถึงคุณสมบัติของตัวประมาณที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การที่วิธีการประมาณไม่เป็นไปตามข้อตกลงเบื้องต้น หรือมีค่าผิดปกติเกิดขึ้น เป็นต้น

3. ฟังก์ชันความควรจะเป็น (Likelihood function) หมายถึงฟังก์ชันที่บอกว่าค่าพารามิเตอร์ $\underset{\sim}{\theta}$ มีความสัมพันธ์อย่างไรกับตัวแปรสุ่ม X ถ้า $\underset{\sim}{\theta}$ ที่กำหนดเป็นค่าจริงของ Θ ($\underset{\sim}{\theta} \in \Theta$) หรือ $\underset{\sim}{\theta}$ จะมีภาวะน่าจะเป็นอย่างไร ณ ค่าสังเกต X หนึ่งๆ

   ถ้าให้ $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)'$ เป็นตัวแปรสุ่มขนาด n ที่เป็นอิสระซึ่งกันและกัน โดยมีฟังก์ชันความหนาแน่น $f(x;\underset{\sim}{\theta})$ เหมือนกัน ในกรณีนี้จะได้ฟังก์ชันความควรจะเป็น $\ell(\underset{\sim}{\theta};x) = \prod_{i=1}^{n} f(x_i;\underset{\sim}{\theta})$

4. สัมประสิทธิ์ความถดถอย (Regression coefficient (β)) หมายถึงค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของตัวแปรตาม (Y) เมื่อตัวแปรอิสระ(X) เปลี่ยนไปหนึ่งหน่วยหรือก็คือความชัน (slope) ของเส้นตรงนั่นเอง

5. การแจกแจงหางยาวหรือหางหนา (Long-tailed distribution or Heavy-tailed distribution) หมายถึง การแจกแจงซึ่งมีฟังก์ชันการแจกแจงความหนาแน่น $f(x)$ เข้าใกล้ศูนย์ ในอัตราที่ช้ากว่าฟังก์ชันการแจกแจงความน่าจะเป็นแบบปกติ เมื่อ X เข้าใกล้ $-\infty$ หรือ ∞

1.9 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อให้ผู้วิจัยได้ทราบถึงวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็ก และสามารถเลือกใช้ตัวประมาณที่เหมาะสมกับขนาดตัวอย่างแต่ละขนาดได้
2. เพื่อให้ผู้วิจัยได้ทราบถึงผลอันอาจเกิดจากการที่ตัวอย่างมีขนาดเล็ก และสามารถแก้ไขปัญหาดังกล่าวได้