บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันนี้การวิจัยสาขาต่าง ๆ ไม่ว่าจะเป็นทางด้านวิทยาศาสตร์ สังคมศาสตร์ การตลาด อุตสาหกรรมและ ทางด้านธุรกิจต่าง ๆ จำเป็นต้องอาศัยความรู้ทางด้านสถิติเข้ามาช่วยในการวิจัยค้นคว้า และการดำเนินงานที่เป็นระบบ การวิจัยในสาขาวิชาต่าง ๆ เหล่านั้น ส่วนใหญ่แล้วต้องอาศัยวิธีการทางสถิติช่วยในการวิเคราะห์เพื่อหาคำตอบสำหรับการวิจัยนั้น ๆ โดยเฉพาะอย่างยิ่งการค้นคว้าหาคำตอบเพื่อคาดคะเนเหตุการณ์ล่วงหน้าหรือที่เรียกว่าการพยากรณ์ ซึ่งผู้วิจัยเป็นจำนวนมากจะใช้วิธีการวิเคราะห์ความถดถอย (Regression Analysis) เข้าช่วยในการหาคำตอบ

วิธีการวิเคราะห์ความถดถอยที่นักวิจัยทุกสาขาวิชาส่วนใหญ่ นิยมใช้กันอย่างแพร่หลาย ก็คือ การวิเคราะห์ความถดถอยเชิงเส้น (Linear Regression Analysis) ซึ่งมีรูปแบบทั่วไปดังนี้

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

เมื่อ

$\underset{\sim}{Y}$ เป็นเวกเตอร์ของตัวแปรตาม (Dependent Variable) ขนาด (nx1)

X เป็นเมตริกซ์ของตัวแปรอิสระ (Independent Variable) ขนาด(nxp)

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของค่าสัมประสิทธิ์ถดถอยเชิงเส้น ขนาด (px1)

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด (nx1)

n ขนาดตัวอย่าง

p จำนวนพารามิเตอร์สัมประสิทธิ์ถดถอยเชิงเส้น

ตามปกติแล้วเมื่อผู้วิจัยนำเอาข้อมูลมาวิเคราะห์ความถดถอยเชิงเส้น จนกระทั่งภายหลังได้ตัวแบบถดถอยเชิงเส้น (Linear Regression Model) เพื่อนำไปใช้ในการพยากรณ์ นั้นผู้วิจัยส่วนใหญ่ไม่ได้ทำการตรวจสอบความเหมาะสมของตัวแบบถดถอยเชิงเส้น ว่ามีความเหมาะสมเพียงไหนก่อนที่จะนำเอาไปใช้ในการพยากรณ์จริง ๆ ในทางปฏิบัติจริงบางครั้งผู้วิจัยอาจมองข้ามตัวแปรที่สำคัญ ๆ บางตัวที่ต้องบรรจุเข้าไปในตัวแบบถดถอย โดยเฉพาะตัวแปรอิสระที่ใช้ในตัวแบบถดถอยของการวิเคราะห์ความถดถอยนั้น บางครั้งผู้วิจัยอาจจะกำหนดตัวแปรอิสระน้อยเกินไป จนได้ตัวแบบถดถอยเชิงเส้นไม่เหมาะสมที่จะใช้ในการพยากรณ์และบางครั้งผู้วิจัยอาจจะกำหนดตัวแปรอิสระมากเกินไปที่จะใช้ในตัวแบบถดถอยเชิงเส้นจนไม่ทราบว่าจะเลือกใช้ตัวแปรอิสระตัวไหนดีจึงจะทำให้ได้ตัวแบบถดถอยเชิงเส้นที่ดีเพื่อใช้ในการพยากรณ์

ฉะนั้นในการกำหนดหรือการพิจารณาตัวแปรอิสระ เพื่อใช้ในตัวแบบถดถอยเชิงเส้นในการพยากรณ์นั้นถ้าใช้ตัวแปรอิสระเพียงตัวเดียวในตัวแบบถดถอยเชิงเส้นหรือที่เรียกว่า ความถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) อาจจะทำให้การพยากรณ์คลาดเคลื่อนจากความเป็นจริงมากหรืออาจจะเชื่อถือไม่ได้มากนัก เป็นเพราะว่าตัวแปรอิสระที่ใช้ในตัวแบบถดถอยนั้นไม่ดีเท่าที่ควร หรือตัวแปรอิสระตัวที่อยู่ในตัวแบบถดถอยเชิงเส้นนั้นไม่ดีเท่าตัวแปรอิสระที่ยังไม่ได้บรรจุในตัวแบบถดถอยดังนั้นการวิเคราะห์ความถดถอยอย่างง่ายแบบเชิงเส้นจึงไม่เป็นที่นิยมใช้สำหรับการพยากรณ์ ส่วนใหญ่แล้วนักวิจัยจะใช้ตัวแปรอิสระหลายตัวในการวิเคราะห์ความถดถอยเชิงเส้นหรือที่เรียกว่า การวิเคราะห์ความถดถอยพหุเชิงเส้น (Multiple Linear Regression) ซึ่งจะช่วยให้การพยากรณ์แม่นยำและถูกต้องมากขึ้น แต่ในการใช้ตัวแปรอิสระมากเกินไปในตัวแบบถดถอยนั้นไม่ใช่ว่าจะให้ผลดีเสมอไป บางครั้งก็ไม่ช่วยลดความคลาดเคลื่อนในการพยากรณ์ได้เลย ในทางกลับกันอาจจะเพิ่มความคลาดเคลื่อนในการพยากรณ์เพิ่มสูงขึ้นกว่าการใช้ตัวแปรอิสระเพียงตัวเดียวในรูปแบบถดถอยก็เป็นไปได้ การเพิ่มหรือลดความคลาดเคลื่อนในการพยากรณ์นั้นขึ้นอยู่กับตัวแปรอิสระที่เข้าไปอยู่ในตัวแบบถดถอยนั้นเหมาะสมเพียงไหน ดังนั้นจึงจำเป็นจะต้องมีการคัดเลือกตัวแปรอิสระที่มีอยู่มากมายนั้น เข้าในตัวแบบถดถอยเชิงเส้น เพื่อให้ได้ตัวแบบถดถอยเชิงเส้นที่ดีหรือเหมาะสมที่สุดเพื่อใช้ในการพยากรณ์ ในการคัดเลือกตัวแปรอิสระนั้นต้องมีหลักเกณฑ์หรือวิธีการคัดเลือกที่ดีพอจึงจะได้ตัวแปรอิสระ เพื่อใช้ในตัวแบบถดถอยเชิงเส้นได้อย่างเหมาะสม วิธีการคัดเลือกเอาตัวแปรอิสระเข้าสู่สมการถดถอยมีอยู่หลายวิธีด้วยกัน แต่วิธีที่นิยมใช้กันแพร่หลายในปัจจุบันก็คือ การกำจัดตัวแปรแบบ

ถอยหลัง (Backward Elimination) การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) (Efroymson (1966) or Draper and Smith (1966)) และการถดถอยแบบขั้นบันได (Stepwise Regression) (Efroymson (1960)) ซึ่งทั้งสามนี้เป็นวิธีที่ใช้แพร่หลายกันมากในการคัดเลือกเอาตัวแปรอิสระเข้าสู่ตัวแบบถดถอยเชิงเส้น ในแต่ละวิธีนั้นจะมีหลักเกณฑ์และวิธีการที่แตกต่างแต่ยังไม่มีข้อสรุปได้ว่าวิธีไหนให้ผลดีที่สุดในการเลือกตัวแปรอิสระเข้าสู่ตัวแบบถดถอยเชิงเส้น

ดังนั้นในการศึกษาวิจัยครั้งนี้ ผู้วิจัยได้ทำการเปรียบเทียบวิธีที่ใช้ในการคัดเลือกตัวแปรอิสระ เพื่อใช้ในตัวแบบถดถอยเชิงเส้นทั้งหมด 5 วิธีด้วยกันคือ

1) การกำจัดตัวแปรแบบถอยหลัง (Backward Elimination)
2) การเลือกตัวแปรแบบไปข้างหน้า (Forward Selection)
3) การถดถอยแบบขั้นบันได (Stepwise Regression)
4) การถดถอยแบบขั้นตอน (Stagewise Regression)
5) การกำจัดตัวแปรโดยใช้สัมประสิทธิ์สหสัมพันธ์

โดยที่วิธีที่ 5 นี้ ผู้วิจัยได้คิดขึ้นมาเอง โดยใช้หลักการเหมือนกันกับวิธีที่ 1 2 และ 3 คือพยายามลดปัญหาที่เกิดขึ้นจากการที่ตัวแปรอิสระมีความสัมพันธ์ภายในต่อกัน แต่จะมีวิธีการคัดเลือกตัวแปรอิสระแตกต่างจากวิธี 1 2 และ 3 ซึ่งจะกล่าว รายละเอียดในบทที่ 2

1.2 ข้อตกลงเบื้องต้น

ลักษณะของตัวแบบถดถอยเชิงพหุแบบเชิงเส้น ซึ่งมีตัวแปรอิสระมากกว่าหนึ่งตัวในตัวแบบถดถอยและมีรูปแบบดังนี้คือ

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

จากตัวแบบถดถอยจะเห็นว่ามีความจำเป็นต้องกำหนดความเคลื่อนไหวทั้ง $\underset{\sim}{\varepsilon}$ และตัวแปรอิสระ X เหตุที่ต้องกำหนดตัวแปร $\underset{\sim}{\varepsilon}$ ก็เพราะว่าตัวแปร $\underset{\sim}{\varepsilon}$ เป็นที่รวมของความผิดพลาดหรือความคลาดเคลื่อนทั้งหมด ไม่ว่าจะเป็นการกำหนดความสัมพันธ์ทางคณิตศาสตร์ระหว่างตัวแปรตามกับตัวแปรอิสระ หรือความคลาดเคลื่อนที่เกิดจากการวัดค่าของตัวแปรตาม หรือความ

ผิดพลาดที่เกิดจากการปรับข้อมูล และความคลาดเคลื่อนที่เกิดจากการกำหนดตัวแปรอิสระ ความคลาดเคลื่อนเหล่านี้จะส่งผลให้ค่าประมาณของตัวแปรตาม คลาดเคลื่อนจากค่าจริงด้วย เหตุนี้ ε จึงเป็นที่รวมของความคลาดเคลื่อนทั้งหมดเอาไว้ ดังนั้นค่าของ ε จึงไม่แน่นอนและไม่สามารถวัดหรือสังเกตออกมาเป็นค่าได้ จึงถือได้ว่า ε เป็นตัวแปรสุ่มที่ไม่ทราบค่าได้ (Unobservation Random Variable) ดังนั้น ε จึงมีความคลาดเคลื่อนอยู่ตลอดเวลาและเป็นตัวแปรที่มีความไวมาก เพราะฉะนั้นความคลาดเคลื่อนในการพยากรณ์ จึงมาจากความคลาดเคลื่อนของ ε เป็นส่วนใหญ่และเพื่อให้งานวิจัยดำเนินงานอย่างมีประสิทธิภาพจึงจำเป็นต้องกำหนดความคลาดเคลื่อนของ ε ไว้ในขอบเขตจำกัด ซึ่งได้เป็นข้อตกลงของการวิเคราะห์ความถดถอยดังนี้ *

1.2.1 ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่

1.2.2 ความคลาดเคลื่อนเป็นตัวแปรสุ่มแบบต่อเนื่อง

1.2.3 ค่าคาดหวังของความคลาดเคลื่อนเป็นเวกเตอร์ศูนย์ $\left[E(\varepsilon) = 0\right]$

1.2.4 ความคลาดเคลื่อนมีการแจกแจงแบบปกติและเป็นอิสระซึ่งกันและกัน

1.3 <u>วัตถุประสงค์ของการวิจัย</u>

1.3.1 เพื่อศึกษาวิธีการเลือกสมการถดถอยในการวิเคราะห์ตัวแปรพหุ

1.3.2 เพื่อเปรียบเทียบวิธีการเลือกสมการถดถอยที่เหมาะสมในการวิเคราะห์ตัวแปรพหุ

1.3.3 เพื่อหาวิธีการที่เหมาะสมในการคัดเลือกตัวแปรอิสระเข้าสมการถดถอยเมื่อสถานการณ์เปลี่ยนไป

1.3.4 เพื่อทดสอบอำนาจพยากรณ์ของสมการประมาณค่า

---

*มนตรี พิริยะกุล, เทคนิคการวิเคราะห์สมการถดถอย, (โรงพิมพ์มหาวิทยาลัยรามคำแหง กรุงเทพมหานคร, 2525) หน้า 86-90.

1.4 สมมติฐานของการวิจัย

1.4.1 ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่มีรูปแบบต่าง ๆ กัน จะมีผลให้วิธีการเลือกตัวแปรอิสระทั้ง 5 วิธีนั้นให้ค่า ผลรวมของความคลาดเคลื่อนกำลังสองค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง และค่าตัวสถิติธีล (Theil's U Statistics)[1] แตกต่างกันไป

1.4.2 วิธีการถดถอยแบบขั้นบันได จะให้ผลการวิเคราะห์เหมาะสมกว่าวิธีอื่น ๆ

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่มีการแจกแจงแบบปกติ (Normal Distribution) โลจิสติก (Logistic Distribution) ดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution) และปกติปลอมปน (Scale Contaminated normal Distribution)

1.5.2 การวิจัยครั้งนี้กำหนดให้ $\beta' = [100\ 1\ 1\ \ldots\ 1]_{(1xp)}$ ในประชากรทุกรูปแบบที่ศึกษา*

1.5.3 จำนวนข้อมูลที่ศึกษามีขนาดเป็น 15 30 50 และ 100

1.5.4 จำนวนการแปรอิสระที่ศึกษา คือ 3 5 7 และ 9

1.5.5 ในการศึกษาวิจัยครั้งนี้จะศึกษา ณ ระดับนัยสำคัญ 0.01 และ 0.05

1.5.6 ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 100**

---

* การวิจัยครั้งนี้กำหนดค่าพารามิเตอร์เป็นค่าดังกล่าวเนื่องจากการทดลองเปลี่ยนค่าเป็นค่าต่าง ๆ แล้วก็ยังทำให้ผลการเปรียบเทียบในแต่ละวิธีการนั้นไม่เปลี่ยนแปลง

** การวิจัยครั้งนี้กำหนดค่าความแปรปรวนเป็นค่าดังกล่าว เนื่องจากการทดลองเปลี่ยนค่าความแปรปรวนเป็นค่าต่าง ๆ แล้วก็ยังทำให้ผลการเปรียบเทียบในแต่ละวิธีการนั้นไม่เปลี่ยนแปลง

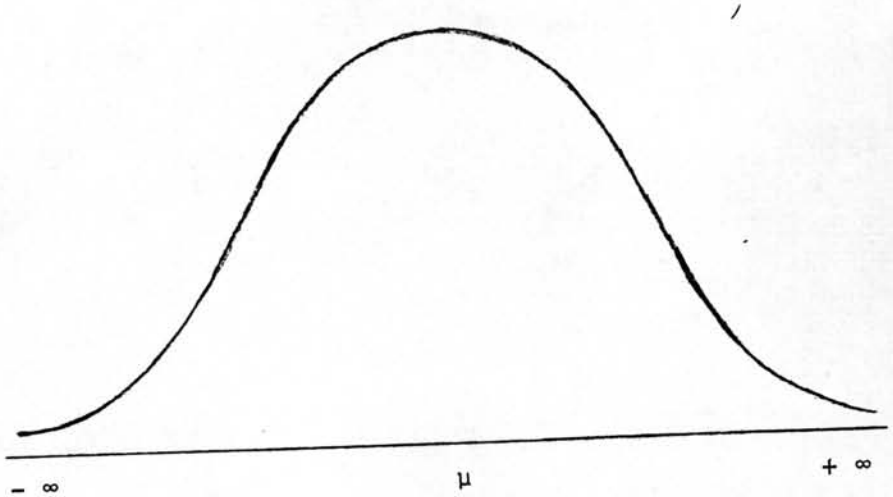
[1] มนตรี พิริยะกุล, เทคนิคการวิเคราะห์สมการถดถอย เล่ม 2, (โรงพิมพ์มหาวิทยาลัยรามคำแหง, กรุงเทพมหานคร, 2526), หน้า 214-216.

1.5.7 ในกรณีการแจกแจงแบบปกติปลอมปนจะศึกษาที่เปอร์เซ็นต์การปลอมปนเป็น 5 10 และ 25 สำหรับสเกลแฟคเตอร์ *** (Scale Factor) มี 2 ระดับคือ 3 และ 10

1.5.8 การวิจัยครั้งนี้จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) จากเครื่องคอมพิวเตอร์ IBM 370/3031 ซึ่งจะศึกษาเมื่อความคลาดเคลื่อน มีการแจกแจงแบบปกติ โลจิสติก ดับเบิ้ลเอ็กซ์โปเนนเชียลและปกติปลอมปน โดยมีฟังก์ชัน ความน่าจะเป็น ค่าคาดหวัง และความแปรปรวนของการแจกแจงแต่ละรูปแบบเป็นดังนี้

1.5.8.1 การแจกแจงแบบปกติ

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2 \right] \quad ; -\infty < X < \infty$$



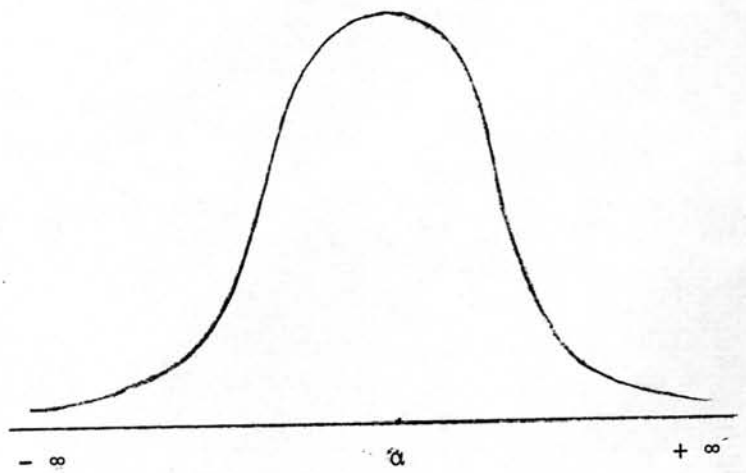
$$E(X) = \mu$$

$$V(X) = \sigma^2$$

*** สเกลแฟคเตอร์ หมายถึง ค่าที่ปรับให้ข้อมูลมีการกระจายมากขึ้นโดยสเกลแฟคเตอร์ที่มีค่าสูงจะทำให้เกิดค่าผิดปกติในข้อมูลและจากการศึกษาที่ผ่านมาพบว่าสเกลแฟคเตอร์ที่มีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลแฟคเตอร์ที่มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมากจึงทำการศึกษาเพียง 2 ระดับดังกล่าว

1.5.8.2 การแจกแจงแบบโลจิสติค

$$f(x) = \frac{1}{\beta} \frac{\exp\left[\frac{-(x-\alpha)}{\beta}\right]}{\left[1+\exp\left[\frac{-(x-\alpha)}{\beta}\right]\right]^2} \quad ; -\infty < x < \infty$$

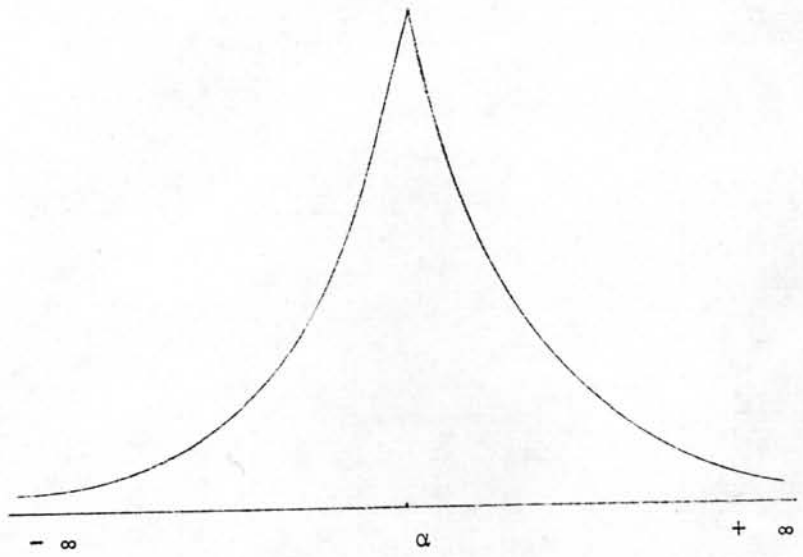


$$E(x) = \alpha$$

$$V(x) = \frac{1}{3}(\pi\beta)^2$$

1.5.8.3 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล

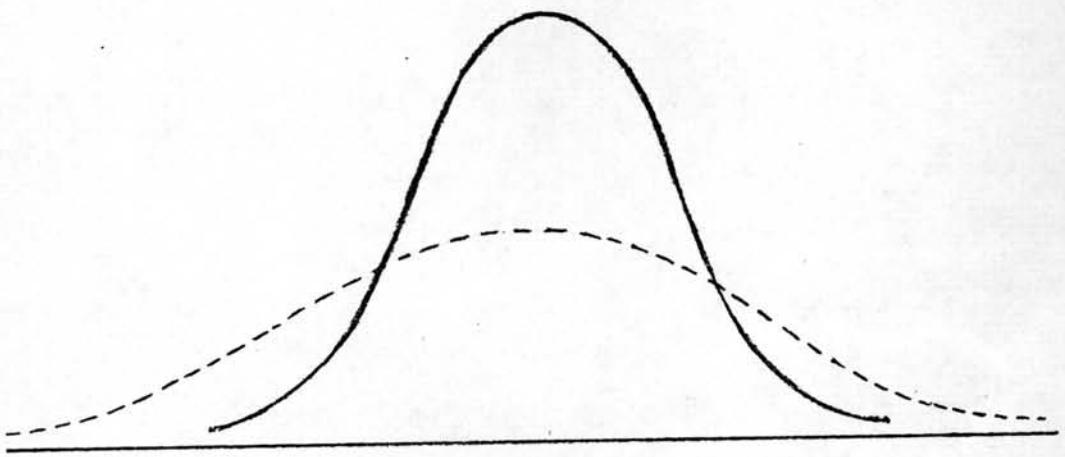
$$f(x) = \frac{1}{2\beta} \exp\left[-\frac{|x-\alpha|}{\beta} \right] \quad ; -\infty < x < \infty$$



$$E(x) = \alpha$$

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

1.5.8.4 การแจกแจงแบบปกติปลอมปน



ลักษณะการแจกแจงแบบปกติปลอมปนที่พิจารณาในการวิจัยครั้งนี้ เป็นการแจกแจงที่แปลงจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชันการแปลงเป็นดังนี้

$$X = (1-P)\ N(\mu, \sigma^2) + P\ N(\mu, c^2\sigma^2) \quad ; c > 0$$

หมายความว่า x มาจากการแจกแจงแบบ $N(\mu,\sigma^2)$ ด้วยความน่าจะเป็น (1-p) และมาจากการแจกแจงแบบ $N(\mu, c^2\sigma^2)$ ด้วยความน่าจะเป็น P ในที่นี้ μ และ σ^2 เป็นค่าเฉลี่ย และความแปรปรวนของความคลาดเคลื่อน ($\mu = 0$, $\sigma^2 = 100$)

P และ c เป็นค่าที่กำหนดเปอร์เซ็นต์การปลอมปนและสเกลแฟคเตอร์ ซึ่งจะทำให้เกิดค่าผิดปกติ

1.5.9 การจำลองการทดลองจะกระทำซ้ำกันไม่ต่ำกว่า 100 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.5.10 ใช้วิธีกำลังสองน้อยที่สุด (Least Squares Method) ในการประมาณค่าสัมประสิทธิ์ถดถอยเชิงเส้น

1.5.11 การวิจัยครั้งนี้ถือว่า ผลรวมของความคลาดเคลื่อนกำลังสอง (Sum Squares of Error) ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Squares of Error) และตัวสถิติของธีล (U) เป็นเกณฑ์ในการพิจารณาความเหมาะสมของตัวแบบถดถอยเชิงเส้น

1.6 ประโยชน์ของการวิจัย

1.6.1 ทำให้ทราบวิธีที่เหมาะสมสำหรับวิธีการเลือกสมการถดถอยในการวิเคราะห์ข้อมูลหลายตัวแปร

1.6.2 สามารถนำเอาวิธีที่เหมาะสมไปประยุกต์กับข้อมูลจริงในการเลือกตัวแปรอิสระเข้าสมการ