บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์การถดถอย (Regression Analysis) เป็นการศึกษาถึงความสัมพันธ์ของตัวแปร 2 กลุ่ม กลุ่มหนึ่งเรียกว่า "ตัวแปรอิสระ" (Independent Variable) ตัวแปรอิสระจะเป็นตัวแทนของปัจจัยต่างๆซึ่งอาจมีหนึ่งหรือหลายตัวแปรและอาจจะมีชื่อเรียกต่างๆกันไปตามลักษณะที่กำหนดเช่น ตัวแปรปัจจัย ตัวแปรดัมมี่ ตัวแปรเวลา ตัวแปรตรีโกณมิติ เป็นต้น อีกกลุ่มหนึ่งเรียกว่า "ตัวแปรตาม" (Dependent Variable) ซึ่งเป็นตัวแปรที่สนใจจะวิเคราะห์และมีค่าขึ้นอยู่กับตัวแปรอิสระ

เมื่อกำหนดตัวแปรอิสระในการวิเคราะห์การถดถอยเพียงหนึ่งตัวแปรจะเรียกการวิเคราะห์การถดถอยว่า การวิเคราะห์การถดถอยเชิงเดียว (Simple Regression Analysis) ส่วนกรณีมีตัวแปรอิสระมากกว่าหนึ่งตัวแปรจะเรียกการวิเคราะห์การถดถอยว่า การวิเคราะห์การถดถอยเชิงพหุ (Multiple Regression Analysis) ส่วนลักษณะความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามอาจจะเป็นแบบเส้นตรง (Linear) หรือแบบไม่เป็นเส้นตรง (Nonlinear)

ตัวแบบการถดถอยเชิงเส้นเชิงเดียว (Simple Linear Regression) มีตัวแบบคือ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

โดยที่ Y_i เป็นตัวแปรตาม (i = 1,2...n)

X_i เป็นตัวแปรอิสระ (i = 1,2...n)

β_0 เป็นระดับค่าเฉลี่ยของ Y_i เมื่อ X_i เท่ากับศูนย์

β_1 เป็นอัตราการเพิ่มหรือลดของตัวแปรตาม Y_i เมื่อเทียบกับ X_i

ε_i เป็นค่าความคลาดเคลื่อน (i = 1,2...n)

n คือขนาดตัวอย่าง

ข้อสมมติของรูปแบบได้แก่ ε_i มีการแจกแจงที่เป็นอิสระกันแบบปกติที่ต่างมีค่าเฉลี่ย 0 และค่าความแปรปรวน σ^2 เขียนแทนด้วย $\varepsilon_i \sim$ Nid($0, \sigma^2$) จากข้อสมมติของ ε_i ดังกล่าวจะได้ว่า Y_i มีการแจกแจงที่เป็นอิสระกัน และที่ X_i ตัวแปร Y_i จะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย $\beta_0 + \beta_1 X_i$ และค่าความแปรปรวน σ^2 เขียนแทนด้วย $Y_i \sim$ Nid($\beta_0 + \beta_1 X_i$, σ^2)

ตัวแบบดังกล่าวเป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรสองตัว หรือลักษณะที่สนใจสองลักษณะ เช่น ในการหาความสัมพันธ์ระหว่างยอดขายและค่าใช้จ่ายในการโฆษณาเพื่อนำมาใช้ในการพยากรณ์ยอดขาย

สินค้าเมื่อทราบค่าใช้จ่ายในการโฆษณา หรือหาความสัมพันธ์ระหว่างปริมาณสินค้าที่โรงงานอุตสาหกรรมแห่งหนึ่งผลิตได้ กับปี พ.ศ. ต่างๆ เพื่อนำมาใช้ในการพยากรณ์ปริมาณสินค้าที่โรงงานแห่งนี้ผลิตได้ในอนาคต การพยากรณ์ที่กล่าวมานี้จะเชื่อถือได้มากน้อยเพียงใด ย่อมขึ้นอยู่กับหลายปัจจัยเช่น รูปแบบของฟังก์ชัน ตัวแปรต่างๆที่มีผลต่อ Y และที่สำคัญคือค่าประมาณของ β_0 กับ β_1 ใกล้เคียงค่าจริงแค่ไหน ในการประมาณค่าของ β_0 และ β_1 จะแยกออกได้เป็นการประมาณแบบจุดและการประมาณแบบช่วง

การประมาณค่าแบบจุดจะเป็นการประมาณค่าพารามิเตอร์ด้วยค่าๆหนึ่งซึ่งความคลาดเคลื่อนที่เกิดจากการประมาณนี้จะขึ้นอยู่กับการเลือกใช้ตัวประมาณ ส่วนการประมาณค่าแบบช่วงจะเป็นการประมาณค่าพารามิเตอร์โดยอาศัยการประมาณค่าแบบจุดและการแจกแจงของตัวประมาณแบบจุด การประมาณค่าแบบช่วงนี้จะทำให้มั่นใจได้ในระดับหนึ่งว่าค่าพารามิเตอร์ที่สนใจอยู่ในช่วงความเชื่อมั่นที่ประมาณได้ สิ่งที่มีผลกระทบต่อความกว้างของช่วงประมาณคือ ระดับความเชื่อมั่นที่ใช้ในการคำนวณ การแจกแจงของตัวสถิติความสอดคล้องกับข้อสมมติ (Assumption) และขนาดตัวอย่างเป็นต้น ดังนั้นจึงควรเลือกใช้ตัวสถิติที่มีความเหมาะสมกับเรื่องที่สนใจศึกษาและสอดคล้องกับประชากรเพื่อให้ช่วงความเชื่อมั่นที่ได้มีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น

เป็นที่ทราบกันดีว่าค่าประมาณแบบช่วงที่สร้างจากตัวประมาณกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Squares) เป็นวิธีการประมาณค่าแบบช่วงที่เหมาะสมกับ β_0 และ β_1 ในตัวแบบการถดถอยเชิงเส้นเชิงเดียวที่มีข้อสมมติตามที่กล่าวไปแล้วข้างต้น ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาวิธีการประมาณค่าแบบช่วงที่เหมาะสมกับ β_0 และ β_1 ในตัวแบบการถดถอยเชิงเส้นเชิงเดียวในกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบเบ้ขวาซึ่งเป็นกรณีที่พบได้บ่อยและไม่เป็นไปตามข้อสมมติข้างต้น แต่ก็จะทำการทดลองในกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบปกติไว้ด้วย เพื่อจะได้รู้ว่าวิธีประมาณแบบช่วงวิธีอื่นใช้ได้ดีหรือไม่เมื่อค่าคลาดเคลื่อนมีการแจกแจงปกติ

ในปี ค.ศ. 1992 กาทเวต (Garthwaite) และ บัคแลนด์ (Buckland) ได้เสนอการใช้กระบวนการรอบบินส์ – มอนโร (Robbins – Monro Search Process) เพื่อหาช่วงความเชื่อมั่นของพารามิเตอร์ซึ่งสร้างมาจากข้อมูลที่ถูกจำลองขึ้นมาด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) งานวิจัยที่กล่าวถึงนี้ได้ทำการทดลองหาช่วงความเชื่อมั่นของพารามิเตอร์เสกล (Scale Parameter) ของการแจกแจงแกมมา (Gamma Distribution) การแจกแจงชี้กำลัง (Exponential Distribution) การแจกแจงชี้กำลังผกผัน (Inverse Exponential Distribution) การแจกแจงโคชี (Cauchy Distribution) การแจกแจงโลจิสติก (Logistic Distribution) การแจกแจงปกติ (Normal Distribution) และการแจกแจงแบบที (T - Distribution) และได้ผลสรุปว่ากระบวนการรอบบินส์ – มอนโรเป็นกระบวนการที่มีความเอนเอียง (Bias) น้อย มีความแกร่ง (Robust) และมีแนวโน้มที่จะได้ช่วงประมาณที่แคบกว่าวิธีแบบฉบับ(Classical Method)ซึ่งเป็นวิธีดั้งเดิมที่ใช้กันอยู่ แต่วิธีนี้ไม่ค่อยมีประสิทธิภาพสำหรับพารามิเตอร์ของการแจกแจงชี้กำลังผกผัน ทั้งนี้ประสิทธิภาพของการใช้กระบวนการรอบบินส์ – มอนโรในการหาช่วงความเชื่อมั่นจะขึ้นอยู่กับการเลือกใช้ค่าปรับแก้ตัวประมาณ ซึ่งในงานวิจัยของกาทเวตนี้ก็ได้เสนอค่าปรับแก้ตัวประมาณที่เหมาะสมเอาไว้ด้วย

ในปี ค.ศ. 1996 กาทเวต (Garthwaite) ได้นำการเรียงสับเปลี่ยน (Permutation) มาใช้ร่วมกับกระบวนการรอบบินส์ – มอนโรเพื่อเพิ่มประสิทธิภาพของการประมาณค่าแบบช่วง กาทเวตได้ยกตัวอย่างการประยุกต์ใช้กับการหาช่วงประมาณของค่าความแตกต่างระหว่างค่าเฉลี่ยของสองประชากร ($\mu_X - \mu_Y$) ภายใต้ข้อสมมติว่าทั้งสองประชากรมีการแจกแจงประเภทเดียวกัน ซึ่งได้ผลสรุปคือวิธีที่เสนอขึ้นมาใหม่นี้ให้ผลดีกว่าวิธีแบบฉบับซึ่งเป็นวิธีดั้งเดิมที่ใช้กันอยู่

ในปี ค.ศ. 2001 โอกอร์แมน (O'Gorman) ได้นำกระบวนการรอบบินส์ – มอนโรและการเรียงสับเปลี่ยนมาประยุกต์ใช้กับการหาช่วงประมาณของ β_1 ในสมการถดถอยเชิงเส้นเชิงเดียวที่มีค่าคลาดเคลื่อนแจกแจงแบบแลมดาของตูกีร์ ที่ระดับความเชื่อมั่น 95% โดยโอกอร์แมนได้เพิ่มการถ่วงน้ำหนัก (Weighted) ให้แก่แต่ละค่าสังเกต และปรับแก้ตัวแบบการถดถอย (Adaptive) ดังนั้นจึงได้วิธีประมาณช่วงความเชื่อมั่นแบบใหม่ซึ่งมีชื่อเรียกว่า "วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักและปรับให้เหมาะสม" (Adaptive Weighted Least Squares) โอกอร์แมนได้ทำการเปรียบเทียบกับวิธีแบบฉบับและได้ผลสรุปว่าวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักและปรับให้เหมาะสมจะให้ผลดีกว่าวิธีแบบฉบับในทุกสถานการณ์ที่ใช้ในการวิจัยเมื่อขนาดตัวอย่างมากกว่าหรือเท่ากับ 40 ($n \geq 40$)

ในปี พ.ศ 2541 วิทยานิพนธ์ของ วีรวรรณ ศักดาจิวะเจิญ เรื่อง "ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา" ได้ทำการเปรียบเทียบค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นที่ได้จากตัวสถิติที ตัวสถิติของจอห์นสัน ตัวสถิติของฮอลล์ และตัวสถิติของเชน โดยได้ข้อสรุปว่าตัวสถิติของจอห์นสันเป็นวิธีที่เหมาะสมกับสถานการณ์เป็นส่วนใหญ่

สำหรับในงานวิจัยนี้จะนำเสนอแนวทางการลดขนาดของช่วงความเชื่อมั่นด้วยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักและปรับให้เหมาะสม (Adaptive Weighted Least Squares) ของโอกอร์แมน โดยนำมาประยุกต์ใช้กับค่า β_0 และ β_1 ของสมการถดถอยเชิงเส้นเชิงเดียวที่มีค่าความคลาดเคลื่อนแจกแจงแบบต่างๆ และจะทำการเปรียบเทียบกับวิธีแบบฉบับและวิธีบูตสแตรป (Bootstrap) ภายใต้สถานการณ์ต่างๆกัน ผลการวิจัยจะทำให้ทราบว่าในสถานการณ์ที่แตกต่างกัน วิธีการประมาณค่าแบบช่วงวิธีการใดให้ช่วงความเชื่อมั่นที่เหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

การวิจัยมีวัตถุประสงค์ดังนี้

1. เพื่อศึกษาและเปรียบเทียบขนาดของช่วงความเชื่อมั่นของสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นเชิงเดียว ที่ได้มาจากวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักและปรับให้เหมาะสม (Adaptive Weighted Least Squares) วิธีแบบฉบับ(Classical Method) และ วิธีบูตสแตรป (Bootstrap)

2. เพื่อศึกษาถึงอิทธิพลของปัจจัยแต่ละชนิดที่มีต่อขนาดของช่วงความเชื่อมั่น ซึ่งปัจจัยดังกล่าวคือ ขนาดตัวอย่าง ระดับของความเชื่อมั่นที่ใช้ในการคำนวณ ความเบ้และความโด่งของค่าคลาดเคลื่อนสุ่ม

3. เพื่อหาข้อสรุปในการเลือกใช้วิธีประมาณแบบช่วงที่เหมาะสมในแต่ละสถานการณ์

1.3 สมมติฐานของการวิจัย

ในการประมาณค่าแบบช่วงสำหรับสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นเชิงเดียวที่มีค่าคลาดเคลื่อนแจกแจงแบบเบ้ขวา "วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนักและปรับให้เหมาะสม" จะให้ช่วงความเชื่อมั่นที่แคบที่สุดภายใต้เงื่อนไขและสถานการณ์ที่กำหนดในการวิจัย

1.4 ข้อตกลงเบื้องต้น

ในการวิจัยนี้จะจำลองค่าความคลาดเคลื่อนที่มีการแจกแจงแบบแลมดาของตูกีร์ การแจงแจงแกมมา การแจกแจงลอกนอร์มอล และการแจกแจงปกติ

1.4.1 <u>การแจกแจงแบบแลมดาของตูกีร์ (Tukey's Lamda Distribution)</u>

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบแลมดาของตูกีร์ด้วยพารามิเตอร์ λ_1 , λ_2 , λ_3 และ λ_4 จะได้ค่าของตัวแปรสุ่ม X เป็นดังนี้

$$f^{-1}(x) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \quad , 0 \le p \le 1$$

เมื่อ p เป็นเลขสุ่มที่มีค่าระหว่าง 0 และ 1

λ_1 เป็นพารามิเตอร์ที่กำหนดตำแหน่ง (Location Parameter)

λ_2 เป็นพารามิเตอร์มาตราส่วน (Scale Parameter)

λ_3 และ λ_4 เป็นพารามิเตอร์สัณฐาน(Shape Parameter) และเป็นฟังก์ชันของ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าความโด่ง และค่าความเบ้

ค่าเฉลี่ย $E(X) = \lambda_1 + A / \lambda_2$

ความแปรปรวน $Var(X) = (B - A^2) / \lambda_2^2$

สัมประสิทธิ์ความเบ้ $\alpha_3 = (C - 3AB + 2A^3) / (\lambda_2 \sigma)^3$

สัมประสิทธิ์ความโด่ง $\alpha_4 = (D - 4AC + 6A^2B - 3A^2) / (\lambda_2 \sigma)^4$

กำหนด $A = 1/(1+\lambda_3) - 1/(1+\lambda_4)$

$B = 1/(1+2\lambda_3) + 1/(1+2\lambda_4) - 2\beta(1+\lambda_3, 1+\lambda_4)$

$C = 1/(1+3\lambda_3) - 3\beta(1+2\lambda_3, 1+\lambda_4) - 1/(1+3\lambda_4) + 3\beta(1+\lambda_3, 1+2\lambda_4)$

$D = 1/(1+4\lambda_3) - 4\beta(1+3\lambda_3, 1+\lambda_4) + 6\beta(1+2\lambda_3, 1+2\lambda_4) - 4\beta(1+\lambda_3, 1+3\lambda_4) + 1/(1+4\lambda_4)$

(โดย β คือค่าเบต้าฟังก์ชัน)

1.4.2 <u>การแจกแจงแกมมา (Gamma Distribution)</u>

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแกมมาด้วยพารามิเตอร์ λ และ γ จะได้ฟังก์ชันความหนาแน่นของ X เป็นดังนี้

$$f(x) = \frac{x^{\gamma-1}\exp(\frac{-x}{\lambda})}{\lambda^{\gamma}\Gamma(\gamma)} \qquad ; x \geq 0,\ \gamma > 0,\ \lambda > 0$$

$$= 0 \qquad \text{อื่นๆ}$$

เมื่อ λ เป็นพารามิเตอร์มาตราส่วน (Scale Parameter)

γ เป็นพารามิเตอร์สัณฐาน(Shape Parameter)

ค่าเฉลี่ย $E(X) = \gamma\lambda$

ความแปรปรวน $Var(X) = \gamma\lambda^2$

สัมประสิทธิ์ความเบ้ $\alpha_3 = 2\gamma^{-\frac{1}{2}}$

สัมประสิทธิ์ความโด่ง $\alpha_4 = 3 + \frac{6}{\gamma}$

1.4.3 <u>การแจกแจงแบบลอกนอร์มอล (Lognormal Distribution)</u>

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงลอกนอร์มอลด้วยพารามิเตอร์ μ และ σ^2 จะได้ฟังก์ชันความหนาแน่นของ X เป็นดังนี้

$$f(x) = \frac{\exp\left(\frac{-(\ln x-\mu)^2}{2\sigma^2}\right)}{x\sqrt{2\pi\sigma^2}} \quad , x \geq 0 , \sigma > 0 , -\infty < \mu < \infty$$

$$= 0 \quad \text{อื่นๆ}$$

เมื่อ μ เป็นค่าเฉลี่ยของ ln x

σ^2 เป็นค่าความแปรปรวนของ ln x

ค่าเฉลี่ย $E(X) = e^{(\mu+\frac{\sigma^2}{2})}$

ความแปรปรวน $Var(X) = e^{(2\mu+\sigma^2)}(e^{\sigma^2}-1)$

สัมประสิทธิ์ความเบ้ $\alpha_3 = (w+2)(w-1)^{1/2}$

สัมประสิทธิ์ความโด่ง $\alpha_4 = w^4 + 2w^3 + 3w^2 - 3$

กำหนด $w = \exp(\sigma^2)$

1.4.4 <u>การแจกแจงแบบปกติ (Normal Distribution)</u>

กำหนดให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงปกติด้วยพารามิเตอร์ μ และ σ^2 จะได้ฟังก์ชันความหนาแน่นของ X เป็นดังนี้

$$f(x) = \frac{\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \quad , -\infty < x < \infty , \sigma > 0 , -\infty < \mu < \infty$$

ค่าเฉลี่ย $E(X) = \mu$

ความแปรปรวน $Var(X) = \sigma^2$

สัมประสิทธิ์ความเบ้ $\alpha_3 = 0$

สัมประสิทธิ์ความโด่ง $\alpha_4 = 3$

1.5 ขอบเขตของการวิจัย

ในการวิจัยครั้งนี้จะทำภายใต้ขอบเขตดังนี้

1. ข้อมูลที่นำมาวิจัยในครั้งนี้จะเป็นข้อมูลจากสมการถดถอยเชิงเส้นเชิงเดียว ($Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$) โดยสร้าง X เป็นค่าใดๆ ในที่นี้กำหนด X เป็น X ~ N(2 , 4) ส่วนค่า Y สร้างมาจากสมการถดถอยที่มีค่าความคลาดเคลื่อนแจกแจงตามที่กำหนด
2. จะสนใจศึกษาเมื่อกำหนดความเบ้และความโด่งของค่าคลาดเคลื่อนตามรายละเอียดดังต่อไปนี้

สัมประสิทธิ์ความเบ้	การแจกแจง	สัมประสิทธิ์ความโด่ง
0	แจกแจงปกติ	3
0.25	แลมดาของตูกีร์	2
		3.2
		6.2
	แกมมา	3.09
	ลอกนอร์มอล	3.11
0.5	แกมมา	3.38
	ลอกนอร์มอล	3.45
1	แลมดาของตูกีร์	4.2
		5.4
		8.4
	แกมมา	4.5
	ลอกนอร์มอล	4.83
1.5	แกมมา	6.37
	ลอกนอร์มอล	7.25
2	แลมดาของตูกีร์	11.4
		12.6
		15.6
	แกมมา	9
	ลอกนอร์มอล	10.86
2.5	แกมมา	12.38
	ลอกนอร์มอล	15.86

ตารางที่ 1.1 แสดงการแจกแจงที่ใช้ในการวิจัย ค่าสปส.ความเบ้ และค่าสปส.ความโด่ง

3. ขนาดตัวอย่างที่ใช้ทำการศึกษาคือ 10 , 20 , 30 , 40 , 50 และ 60
4. ระดับความเชื่อมั่นที่ใช้ในการคำนวณช่วงประมาณคือ 90% , 95% และ 99%
5. เนื่องจาก β_0 กับ β_1 ที่ใช้ในการจำลองข้อมูลไม่มีผลต่อการวิจัย จึงกำหนดค่า β_0 กับ β_1 เป็นค่าใดๆ ในที่นี้กำหนดเป็นสองกรณีดังนี้

    (1) กรณีแรก β_0 = 1 และ β_1 = 2

    (2) กรณีที่สอง β_0 = -1 และ β_1 = -3

    (โดยผลลัพธ์ของกรณีที่ใช้ β_0 = -1 และ β_1 = -3 จะแสดงไว้ในภาคผนวก)
6. การหาค่าประมาณแบบช่วงด้วยวิธีบูทสแตรป กำหนดจำนวนการสุ่มซ้ำเท่ากับ 2,000 ครั้ง
7. จำลองข้อมูลตามสถานการณ์ โดยทำการทดลอง 1,000 ครั้งในแต่ละสถานการณ์

1.6 เกณฑ์ที่ใช้ในการเปรียบเทียบ

ในการเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นเชิงเดียวจะพิจารณาเปรียบเทียบค่าสัมประสิทธิ์ความเชื่อมั่นและค่าเฉลี่ยความยาวของช่วงความเชื่อมั่น (Mean of Confidence Interval Length) ของแต่ละวิธีการประมาณ โดยในขั้นแรกจะต้องตรวจสอบว่าช่วงความเชื่อมั่นที่ได้จากทั้ง 3 วิธีนั้นให้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองมากกว่าหรือเท่ากับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่ แล้วจึงพิจารณาว่าค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นของวิธีการใดให้ค่าต่ำที่สุดก็จะเป็นวิธีการที่เหมาะสมสำหรับการประมาณค่าแบบช่วงภายใต้สถานการณ์นั้นๆ ทั้งนี้จะพิจารณาเฉพาะกรณีที่วิธีการประมาณนั้นให้สัมประสิทธิ์ความเชื่อมั่นจากการทดลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น

1.6.1 <u>การตรวจสอบค่าสัมประสิทธิ์ความเชื่อมั่น</u>

การคำนวณค่าสัมประสิทธิ์ความเชื่อมั่นจะทำการพิจารณาว่าช่วงความเชื่อมั่นที่ได้จากแต่ละวิธีการประมาณนั้นครอบคลุมค่าพารามิเตอร์หรือไม่ โดยทำการนับจำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นที่ประมาณได้ครอบคลุมค่าพารามิเตอร์ นำค่าที่ได้นี้หารด้วยจำนวนครั้งที่ทำซ้ำทั้งหมด ค่าที่ได้นี้คือค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองของแต่ละวิธีการประมาณ ในการตรวจสอบว่าค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองของวิธีการใดมีค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดนั้นจะอาศัยการทดสอบสมมติฐานด้วยตัวสถิติ Z ซึ่งมีรายละเอียดดังนี้

1. คำนวณค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลอง โดยนับสะสมจำนวนครั้งที่ช่วงประมาณคลุมค่าพารามิเตอร์จากการคำนวณซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ และนำผลบวกสะสมที่ได้นี้หารด้วย 1,000 ค่าที่ได้จะเป็นค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลอง (c)

2. เมื่อได้ค่าสัมประสิทธิ์ความเชื่อมั่นจากการทดลองแล้ว จะนำค่าดังกล่าวมาเปรียบเทียบกับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด(c_0) ซึ่งการตรวจสอบได้อาศัยการทดสอบสมมุติฐานโดยใช้ตัวสถิติ Z ซึ่งมีรูปแบบดังนี้

$H_0: c \geq c_0$

$H_1: c < c_0$

ซึ่งจะได้ว่า $$P\left(\frac{c-c_0}{\sqrt{\frac{c_0(1-c_0)}{n}}} > -Z_{1-\alpha_0}\right) = 1-\alpha_0$$

ดังนั้น จะได้ช่วงของการยอมรับสมมุติฐานหลักคือ

$$\left(c_0 - Z_{1-\alpha_0}\sqrt{\frac{c_0(1-c_0)}{n}}, 1\right)$$

และวิธีการประมาณที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่าเกณฑ์ที่กำหนด จะต้องให้ค่า c อยู่ในช่วง

$$\left(c_0 - Z_{1-\alpha_0}\sqrt{\frac{c_0(1-c_0)}{n}}, 1\right)$$

โดยที่ α_0 คือระดับนัยสำคัญหรือความน่าจะเป็นที่จะเกิดความผิดพลาดในการประมาณค่า สำหรับการวิจัยครั้งนี้กำหนดระดับนัยสำคัญของการทดสอบเท่ากับ 0.05

c คือสัมประสิทธิ์ความเชื่อมั่นของวิธีการประมาณ

c คือสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการทดลอง

c_0 คือระดับความเชื่อมั่นที่กำหนด (0.90 , 0.95 และ 0.99)

n คือจำนวนครั้งของการทดลอง (1,000)

ที่ระดับความเชื่อมั่น 90%

$H_0: c \geq 0.9$

$H_1: c < 0.9$

วิธีการประมาณที่ผ่านระดับความเชื่อมั่นจะต้องให้ค่า c อยู่ในช่วง

$$\left(0.9 - 1.645\sqrt{\frac{0.9(0.1)}{1000}}, 1\right)$$

หรือมีค่าอยู่ในช่วง (0.884 , 1)

ที่ระดับความเชื่อมั่น 95%

H_0: $c \geq 0.95$

H_1: $c < 0.95$

วิธีการประมาณที่ผ่านระดับความเชื่อมั่นจะต้องให้ค่า $\hat{c}$ อยู่ในช่วง

$$\left(0.95 - 1.645\sqrt{\frac{0.95(0.05)}{1000}}, 1\right)$$

หรือมีค่าอยู่ในช่วง (0.939 , 1)

ที่ระดับความเชื่อมั่น 99%

H_0: $c \geq 0.99$

H_1: $c < 0.99$

วิธีการประมาณที่ผ่านระดับความเชื่อมั่นจะต้องให้ค่า c อยู่ในช่วง

$$\left(0.99 - 1.645\sqrt{\frac{0.99(0.01)}{1000}}, 1\right)$$

มีค่าอยู่ในช่วง (0.985 , 1)

สรุปคือถ้าวิธีการประมาณแบบใดให้ค่าสัมประสิทธิ์ความเชื่อมั่นไม่ต่ำกว่า 0.884 ที่ระดับความเชื่อมั่น 90% หรือไม่ต่ำกว่า 0.939 ที่ระดับความเชื่อมั่น 95% หรือไม่ต่ำกว่า 0.985 ที่ระดับความเชื่อมั่น 99% จะถือว่าวิธีประมาณนั้นผ่านระดับความเชื่อมั่นที่กำหนด

1.6.2 <u>การเปรียบเทียบความยาวเฉลี่ยของค่าประมาณแบบช่วง</u>

หลังจากตรวจสอบค่าสัมประสิทธิ์ความเชื่อมั่นแล้ว ขั้นต่อไปก็คือการเปรียบเทียบความยาวเฉลี่ยของค่าประมาณแบบช่วงตามขั้นตอนดังต่อไปนี้

1. คำนวณความยาวเฉลี่ยของค่าประมาณแบบช่วง โดยทำการบวกสะสมขนาดของช่วงไว้จนครบ 1,000 ครั้ง และนำค่าที่ได้นี้หารด้วย 1,000 ค่าที่ได้คือ ความยาวเฉลี่ยของช่วงความเชื่อมั่น

2. ทำการเปรียบเทียบความยาวเฉลี่ยของช่วงความเชื่อมั่นที่ได้จากแต่ละวิธี โดยเปรียบเทียบเฉพาะวิธีที่ผ่านระดับความเชื่อมั่นที่กำหนดไว้

1.7 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัยครั้งนี้คือ

1. เป็นแนวทางในการลดขนาดช่วงความเชื่อมั่นของการประมาณค่า
2. ได้ทราบถึงอิทธิพลของปัจจัยต่างๆ ที่มีผลต่อขนาดของช่วงความเชื่อมั่นที่ได้จากแต่ละวิธี
3. เพื่อเป็นแนวทางในการศึกษาและเลือกวิธีการประมาณแบบช่วงสำหรับค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นเชิงเดียวที่มีค่าคลาดเคลื่อนแจกแจงแบบเบ้ขวาและแจกแจงแบบปกติในสถานการณ์ต่างๆได้อย่างเหมาะสม
4. เพื่อเป็นแนวทางในการศึกษาวิธีการประมาณแบบช่วงสำหรับค่าสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นเชิงเดียวที่มีค่าคลาดเคลื่อนแจกแจงแบบเบ้ขวาโดยใช้ตัวสถิติอื่นๆต่อไป