หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการประมาณค่าและการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ของความถดถอยเชิงเส้นแบบง่าย (Simple Linear Regression) นั้น การเลือกใช้วิธีที่เหมาะสมกับลักษณะของข้อมูลที่ต้องการวิเคราะห์นั้นจะต้องคำนึงถึงข้อตกลงเบื้องต้นของวิธีแต่ละวิธีที่ใช้ด้วย ในการวิจัยโดยทั่ว ๆ ไป ปัญหาที่ผู้วิจัยพบมากก็คือ ลักษณะการแจกแจงของความคลาดเคลื่อน (error) ไม่เป็นไปตามข้อตกลงเบื้องต้น คือไม่เป็นการแจกแจงแบบปกติ เช่น อาจเป็นการแจกแจงที่มีหางยาว (long tails) หรือมีการกระจายไปทางหางมาก (heavy tails) ในกรณีเช่นนี้ผู้วิจัยอาจจะเลือกใช้วิธีการนอนพาราเมตริก (Nonparametric Method) ซึ่งเป็นวิธีที่ไม่มีข้อจำกัดเกี่ยวกับลักษณะการแจกแจงของความคลาดเคลื่อน สามารถคำนวณได้รวดเร็ว ทำความเข้าใจง่าย และสะดวกในการนำไปใช้ นับเป็นอีกวิธีหนึ่งที่ผู้วิจัยสามารถเลือกใช้ได้

สำหรับการศึกษาความถดถอยเชิงเส้นแบบง่ายที่มีรูปแบบทั่วไปเป็นดังนี้คือ

$y_i = \beta_o + \beta_1 x_i + \varepsilon_i$, $i = 1,2,\ldots,n$ โดยที่ x_i เป็นตัวแปรอิสระที่มีค่าคงที่ y_i เป็นตัวแปรตาม β_o และ β_1 เป็นพารามิเตอร์ที่ไม่ทราบค่าแทนจุดตัดบนแกน y และความชันของเส้นถดถอยตามลำดับ ε_i เป็นความคลาดเคลื่อน และ n เป็นขนาดตัวอย่าง โดยปกติการประมาณค่าและการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์นั้น ผู้วิจัยมักจะเลือกใช้วิธีกำลังสองต่ำสุด (Least Squares Method) ซึ่งเป็นวิธีที่ให้ตัวประมาณ (estimator) ที่มีคุณสมบัติเป็น BLUE (Best Linear Unbiased Estimator) และในการทดสอบสมมติฐานก็ให้สถิติทดสอบ (test statistic) ที่มีอำนาจของการทดสอบ (power of the test) สูง ทั้งนี้จะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนดังนี้คือ ความคลาดเคลื่อนจะต้องมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ^2 และ $\varepsilon_i, \varepsilon_j$ ไม่มีสหสัมพันธ์ต่อกัน เมื่อ $i \neq j$ หรือ $\varepsilon_i \sim NID(o, \sigma^2)$ อย่างไรก็ตามในทางปฏิบัติ บ่อยครั้งที่เราพบว่า ข้อมูลที่ถูกนำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว โดยความคลาดเคลื่อนมักจะมีการแจกแจงลักษณะคล้ายแบบปกติ แต่จะมีการกระจายไปทางหางมาก หรือมีหางยาวกว่าปกติ

ซึ่งการแจกแจงลักษณะเช่นนี้จะทำให้เกิดค่าคลาดเคลื่อนบางค่ามีค่าสูงมาก ๆ หรือต่ำมาก ๆ ซึ่งจะไปมีผลทำให้ค่าสังเกต (y) บางค่าเป็นค่าผิดปกติ* (outliers) ลักษณะข้อมูลเช่นนี้จะพบมากทางด้าน ชีววิทยา เคมี และการแพทย์ ดังนั้นในกรณีเช่นนี้ ผู้วิจัยอาจเลือกใช้วิธีการนอนพาราเมตริก ซึ่งใช้ประมาณค่าและทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ได้ เช่นเดียวกับวิธีกำลังสองต่ำสุด โดยไม่จำเป็นต้องมีข้อตกลงเบื้องต้นเกี่ยวกับลักษณะการแจกแจงของความคลาดเคลื่อนดังกล่าว เช่น วิธีของบราวน์และมูด (Brown and Mood's Method) วิธีของเซ็นและทิลล์ (Sen and Theil's Method) วิธีของซีเวอร์ (Sievers' Method) หรือวิธีของแลนคาสเตอร์และเควด (Lancaster and Quade's Method) ซึ่งอาจจะเป็นวิธีที่เหมาะสมกว่า

สำหรับในเรื่องการประมาณค่าของพารามิเตอร์นั้น ผู้วิจัยต่างมุ่งหวังที่จะได้ตัวประมาณที่ "ดี" เสมอ ส่วนการที่จะตัดสินว่าตัวประมาณเช่นไรจึงจะเป็นตัวประมาณที่ "ดี" นั้นขึ้นกับจุดประสงค์ของการวิจัย ซึ่งโดยส่วนใหญ่จะอาศัยคุณสมบัติของตัวประมาณหลายอย่างเป็นเกณฑ์ คุณสมบัติเหล่านั้นได้แก่ ความไม่เอนเอียง (Unbiasedness) ความคงเส้นคงวา (Consistency) ความพอเพียง (Sufficiency) ประสิทธิภาพ (Efficiency) ความแปรปรวนต่ำสุด (Minimum Variance) ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum Mean Square Error) เป็นต้น สำหรับการพิจารณาความเหมาะสมของตัวประมาณในการวิจัยครั้งนี้ จะพิจารณาจากคุณสมบัติ ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด เนื่องจากในการเปรียบเทียบเชิงทฤษฎีเราทราบค่าพารามิเตอร์ ดังนั้นสามารถใช้ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) หรือ MSE เป็นเกณฑ์ในการเปรียบเทียบได้ ซึ่งค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณจะเป็นค่าที่แสดงถึงความเที่ยงตรง (accuracy) ของตัวประมาณ โดยตัวประมาณที่มีความเที่ยงตรงที่สุดก็คือ ตัวประมาณที่มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด

---

*ค่าผิดปกติ หมายถึง ค่าสังเกตที่มีค่ามาก หรือน้อยกว่าค่าสังเกตอื่น ๆ อย่างผิดปกติ

การศึกษาวิธีการประมาณค่าของพารามิเตอร์ โดยวิธีการนอนพาราเมตริกนั้น ได้มีผู้ศึกษาไว้หลายท่าน เช่น การศึกษาของทิลล์ (1950 : 386-392, 521-525 และ 1397-1412) บราวน์และมู้ด (1951 : 159-166) เฮ็น (1968 : 1379-1389) และซีเวอร์ (1978 : 628-631) ซึ่งส่วนใหญ่จะเสนอเพียงวิธีการหาค่าประมาณเท่านั้น แต่ก็มีบางท่านที่ศึกษาเพิ่มเติมในรูปของค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ (Asymptotic Relative Efficiency) หรือ ARE เช่นการศึกษาของซีเวอร์ ซึ่งศึกษาค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์ โดยเทียบกับวิธีกำลังสองต่ำสุด พบว่าถ้าความคลาดเคลื่อนมีการแจกแจงแบบปกติ ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์จะน้อยกว่า 1 กล่าวคือ วิธีกำลังสองต่ำสุดจะมีประสิทธิภาพดีกว่า แต่ถ้าความคลาดเคลื่อนมีการแจกแจงที่มีการกระจายไปทางหางมาก หรือมีหางยาวกว่าปกติ ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์จะมากกว่า 1 กล่าวคือ วิธีของซีเวอร์ จะมีประสิทธิภาพดีกว่า จากแนวคิดในการศึกษาเปรียบเทียบนี้ จึงเป็นสิ่งน่าสนใจที่จะศึกษาต่อไปสำหรับวิธีการอื่น ๆ ที่ยังไม่มีการศึกษาเปรียบเทียบไว้ ซึ่งในการวิจัยครั้งนี้จะเปรียบเทียบในรูปของค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency) หรือ RE ของวิธีการนอนพาราเมตริกเทียบกับวิธีกำลังสองต่ำสุด โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณในแต่ละวิธีเป็นค่าเปรียบเทียบ เพื่อจะได้มีข้อสรุปที่แน่นอนในการเลือกใช้วิธีการประมาณค่าพารามิเตอร์ เมื่อความคลาดเคลื่อนมีการแจกแจงที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว

สำหรับการพิจารณาความเหมาะสมของสถิติทดสอบในการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์นั้น สิ่งที่ควรจะพิจารณาก็คือ อำนาจของการทดสอบ และความแกร่ง (robustness) โดยจะพิจารณาในลักษณะที่ว่า สถิติทดสอบนั้นจะต้องมีความไว (sensitive) ต่อการเปลี่ยนแปลงของปัจจัยที่ต้องการทดสอบ และจะต้องไม่มีความไว (insensitive)ต่อการเปลี่ยนแปลงของสิ่งอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ ทั้งนี้จะพิจารณาตามคำกล่าวของเนย์แมน (Neyman 1950 : 265 อ้างโดย Direk Srisukho 1974 : 38) ซึ่งกล่าวว่า "เมื่อต้องการที่จะเลือกใช้สถิติทดสอบ เราต้องเริ่มพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ก่อน แล้วจึงจะพิจารณาถึงความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 2 โดยมีขั้นตอนดังนี้คือ ให้ความน่าจะเป็นที่ยอมให้เกิดความคลาดเคลื่อนประเภทที่ 1 ไม่เกิน α ที่กำหนดไว้ และเมื่อเป็นไปตามเงื่อนไขดังกล่าวแล้ว สิ่งที่จะใช้ในการพิจารณาเลือกสถิติทดสอบอีกก็คือ เลือกสถิติทดสอบที่มีโอกาสน้อยที่สุดที่จะยอมรับสมมติฐาน H_o เมื่อสมมติฐาน H_o นั้นผิด ซึ่งหมายความว่าให้อำนาจของการทดสอบสูงที่สุด"

การศึกษาวิธีการทดสอบแบบนอนพาราเมตริก สำหรับทดสอบสมมติฐานเฉพาะพารามิเตอร์ β_1 นั้น ได้มีผู้ศึกษาไว้หลายท่าน ซึ่งทำการศึกษาไปพร้อม ๆ กับการประมาณค่า เช่น การศึกษาของทิลล์ (1950:386-392, 521-525 และ 1397-1412) บราวน์และมูด (1951:159-166) เซ็น (1968:1379-1389) และซีเวอร์ (1978:628-631) โดยทุกท่านที่กล่าวมานี้เสนอเพียงวิธีการหาค่าสถิติทดสอบเท่านั้น สำหรับการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_0, β_1 นั้น ก็มีผู้ศึกษาไว้หลายท่านเช่นกันคือ การศึกษาของบราวน์และมูด (1951 159-166) และแลนคาสเตอร์และเควด (1985:393-397) ซึ่งได้ศึกษาเปรียบเทียบไว้ในรูปของค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมน (Pitman Asymptotic Relative Efficiency) ของการทดสอบแบบนอนพาราเมตริก เมื่อเทียบกับการทดสอบแบบพาราเมตริก หรือนอนพาราเมตริกด้วยกัน พบว่า ถ้าความคลาดเคลื่อน มีการแจกแจงลักษณะที่มีการกระจายไปทางหางมาก หรือมีหางยาวกว่าปกติ จะได้ค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมนมากกว่า 1 กล่าวคือ การทดสอบแบบนอน-พาราเมตริกมีประสิทธิภาพดีกว่า การทดสอบแบบพาราเมตริก และการทดสอบของแลนคาส-เตอร์และเควด จะมีค่าประสิทธิภาพสัมพัทธ์ใกล้อนันต์แบบพิทแมนสูงสุด จากแนวคิดในการศึกษาเปรียบเทียบนี้ จึงเป็นสิ่งน่าสนใจที่จะศึกษาต่อไปสำหรับการทดสอบอื่น ๆ ที่ยังไม่มีการศึกษาเปรียบเทียบไว้ ซึ่งในการวิจัยครั้งนี้จะพิจารณาจากอำนาจของการทดสอบและความแกร่ง เพื่อจะได้มีข้อสรุปที่แน่นอนในการเลือกใช้สถิติทดสอบที่เหมาะสม สำหรับการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ เมื่อความคลาดเคลื่อนมีการแจกแจงที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว

ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการศึกษาทั้งวิธีการประมาณค่า และการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ และเนื่องจากการคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ และค่าอำนาจของการทดสอบของสถิติทดสอบแบบนอนพาราเมตริก หรือแบบพารา-เมตริก เมื่อสภาพการแจกแจงไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าวกระทำได้ยากมาก ดังนั้นการวิจัยครั้งนี้จึงได้ทำการศึกษาโดยใช้เทคนิคการจำลองแบบที่เรียกว่า เทคนิคมอนติคาร์โลซีมูเลชั่น (Monte Carlo Simulation Technique) ซึ่งเป็นเทคนิคที่จะทำให้ได้ผลสรุปจากสภาพการณ์ที่เป็นการทดลอง สามารถกำหนดขนาดของกลุ่มตัวอย่าง ลักษณะการแจกแจงของความคลาด-เคลื่อน ค่าเฉลี่ย และความแปรปรวนได้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณ 4 วิธี ซึ่งจำแนกเป็น

1.2.1.1 ตัวประมาณแบบพาราเมตริก โดยวิธีกำลังสองต่ำสุด (Least Squares Method)

1.2.1.2 ตัวประมาณแบบนอนพาราเมตริก โดย

1. วิธีของบราวน์และมูด (Brown and Mood's Method)
2. วิธีของเซ็นและทิลล์ (Sen and Theil's Method)
3. วิธีของซีเวอร์ (Sievers' Method)

1.2.2 เพื่อเปรียบเทียบสถิติที่ใช้ทดสอบพารามิเตอร์ β_1 4 วิธี ซึ่งจำแนกเป็น

1.2.2.1 การทดสอบแบบพาราเมตริก โดยวิธีกำลังสองต่ำสุด (Least Squares Method)

1.2.2.2 การทดสอบแบบนอนพาราเมตริก โดย

1. วิธีของบราวน์และมูด (Brown and Mood's Method)
2. วิธีของเซ็นและทิลล์ (Sen and Theil's Method)
3. วิธีของซีเวอร์ (Sievers' Method)

1.2.3 เพื่อเปรียบเทียบสถิติที่ใช้ทดสอบพารามิเตอร์ β_o , β_1 3 วิธี ซึ่งจำแนกเป็น

1.2.3.1 การทดสอบแบบพาราเมตริก โดยวิธีกำลังสองต่ำสุด (Least Squares Method)

1.2.3.2 การทดสอบแบบนอนพาราเมตริก โดย

1. วิธีของบราวน์และมูด (Brown and Mood's Method)
2. วิธีของแลนคาสเตอร์และเควด (Lancaster and Quade's Method)

1.3 สมมติฐานของการวิจัย

ลักษณะการแจกแจงของความคลาดเคลื่อน ที่มีรูปแบบต่าง ๆ กัน จะมีผลให้

1.3.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณแตกต่างกัน และ

1.3.2 ค่าอำนาจของการทดสอบของสถิติทดสอบแตกต่างกัน

1.4 ข้อตกลงเบื้องต้น

1.4.1 ความคลาดเคลื่อน เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแบบเดียวกัน และเป็นอิสระซึ่งกันและกัน

1.4.2 การวิจัยครั้งนี้ถือว่า ความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการเลือกตัวประมาณ ส่วนอำนาจของการทดสอบและความแกร่งเป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการเลือกสถิติทดสอบ

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ ของวิธีกำลังสองต่ำสุด วิธีของบราวน์และมูด วิธีของเซ็นและทิลล์ และวิธีของซีเวอร์ เมื่อความคลาดเคลื่อนมีการแจกแจงแบบยูนิฟอร์ม แบบโลจิสติค แบบดับเบิลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน

1.5.2 ศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ในการทดสอบเกี่ยวกับพารามิเตอร์ β_1 ของวิธีกำลังสองต่ำสุด วิธีของบราวน์และมูด วิธีของเซ็นและทิลล์ และวิธีของซีเวอร์ เมื่อความคลาดเคลื่อนมีการแจกแจงแบบต่าง ๆ ดังกล่าวใน 1.5.1

1.5.3 ศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ในการทดสอบเกี่ยวกับพารามิเตอร์ β_0, β_1 ของวิธีกำลังสองต่ำสุด วิธีของบราวน์และมูด และวิธีของแลนคาสเตอร์และเควด เมื่อความคลาดเคลื่อนมีการแจกแจงแบบต่าง ๆ ดังกล่าวใน 1.5.1

1.5.4 ในกรณีการแจกแจงแบบปกติปลอมปน จะศึกษาเมื่อเปอร์เซ็นต์การปลอมปนเป็น 1% , 5% , 10% และ 25% สำหรับสเกลแฟคเตอร์* (scale factor) 2 ระดับ คือ 3 และ 10

---

*สเกลแฟคเตอร์ เป็นค่าที่ปรับให้ข้อมูลมีการกระจายมากขึ้น โดยสเกลแฟคเตอร์ที่มีค่าสูง อาจทำให้เกิดค่าผิดปกติขึ้นในข้อมูล และจากการศึกษาที่ผ่านมา สเกลแฟคเตอร์ที่มีค่าน้อยกว่า 3 จะมีโอกาสของการเกิดค่าผิดปกติน้อย ส่วนสเกลแฟคเตอร์ที่มีค่ามากกว่า 10 จะมีโอกาสของการเกิดค่าผิดปกติมากจึงทำการศึกษาเพียง 2 ระดับดังกล่าว

1.5.5 กำหนดค่าเฉลี่ย (μ) เป็น 0 และค่าความแปรปรวน (σ^2) เป็น 100 ทุกรูปแบบที่ศึกษา

1.5.6 ศึกษาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ เมื่อพารามิเตอร์ β_0 มีค่าเป็น 50 และ β_1 มีค่าเป็น 1

1.5.7 ศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบสมมติฐานเฉพาะพารามิเตอร์ β_1 เมื่อพารามิเตอร์ β_0 มีค่าเป็น 50 และ β_1 มีค่าเป็น .1, .3, .5, .6, .7, .8, .9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7 และ 1.9

1.5.8 ศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ β_0, β_1 เมื่อพารามิเตอร์ β_0 มีค่าเป็น 50 และ β_1 มีค่าเป็น .5, .6, .7, .8, .9, 1.0, 1.1, 1.2, 1.3, 1.4,และ 1.5

หมายเหตุ ในการวิจัยครั้งนี้ ที่กำหนดค่าความแปรปรวนของความคลาดเคลื่อน และค่าพารามิเตอร์ β_0 , β_1 เป็นค่าต่าง ๆ ดังกล่าว เนื่องจากการทดลองกระทำที่ขนาดตัวอย่าง 10 แล้วพบว่า ค่าความแปรปรวนและค่าพารามิเตอร์ไม่มีผลต่อค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนประเภทที่ 1 และค่าอำนาจของการทดสอบ อย่างไรก็ตามที่กำหนดค่าความแปรปรวนเป็น 100 เนื่องจากต้องการดูค่าอำนาจของการทดสอบในการทดสอบการพารามิเตอร์ให้กว้างขึ้น ซึ่งค่า β_1 ที่อยู่นอกช่วงดังกล่าวใน 1.5.7 และ 1.5.8 จะให้ค่าอำนาจของการทดสอบเป็น 1 จึงศึกษาค่า β_1 เพียงในช่วงดังกล่าว

1.5.9 ศึกษาอำนาจของการทดสอบเฉพาะการทดสอบแบบสองทาง (two-sided test) ณ ระดับนัยสำคัญ .01, .05 และ .10

1.5.10 การศึกษาทุกรูปแบบดังกล่าว จะทำการศึกษาทั้งกรณีที่ค่าคงที่ x มีช่วงห่างเท่ากัน และไม่เท่ากัน

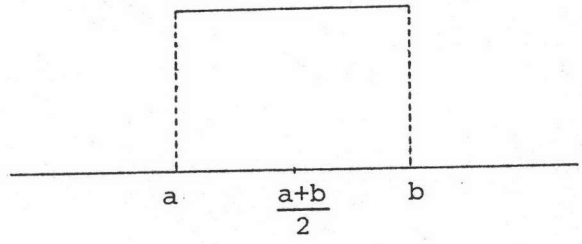
1.5.11 ศึกษาที่ขนาดตัวอย่าง 10, 15 และ 20 สำหรับการศึกษาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ และจะศึกษาที่ขนาดตัวอย่าง 10, 15, 20 และ 50 สำหรับการศึกษาอำนาจของการทดสอบ และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

1.5.12 ในการวิจัยครั้งนี้ จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่นจากเครื่องคอมพิวเตอร์ IBM370/3031 ซึ่งจะศึกษาเมื่อ ความคลาดเคลื่อนมีการแจกแจงแบบยูนิฟอร์ม แบบโลจิสติค แบบดับเบิ้ลเอ็กซ์โปเนนเชียล และแบบปกติปลอมปน โดยมีค่าฟังก์ชั่นความน่าจะเป็น ค่าคาดหวัง ค่าความแปรปรวน ของการแจกแจงแต่ละรูปแบบเป็นดังนี้

1.5.12.1 การแจกแจงแบบยูนิฟอร์ม (Uniform Distribution)

ฟังก์ชั่นความน่าจะเป็น คือ

$$f(x) = \frac{1}{b-a} \quad , \ a < x < b$$



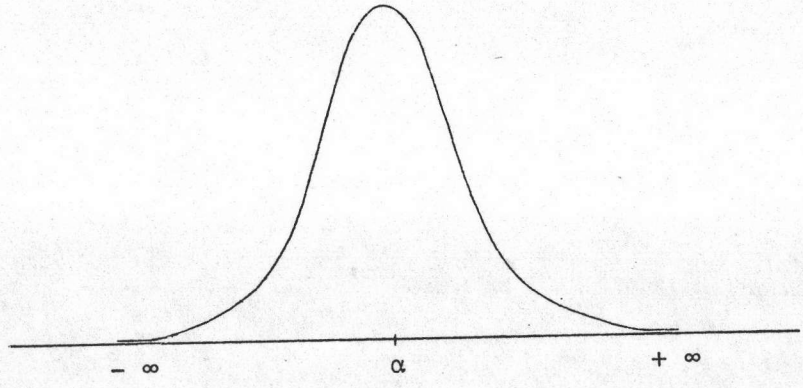
ค่าคาดหวัง $E(X) = \frac{(a+b)}{2}$

ค่าความแปรปรวน $V(X) = \frac{(b-a)^2}{12}$

1.5.12.2 การแจกแจงแบบโลจิสติค (Logistic Distribution)

ฟังก์ชั่นความน่าจะเป็น คือ

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



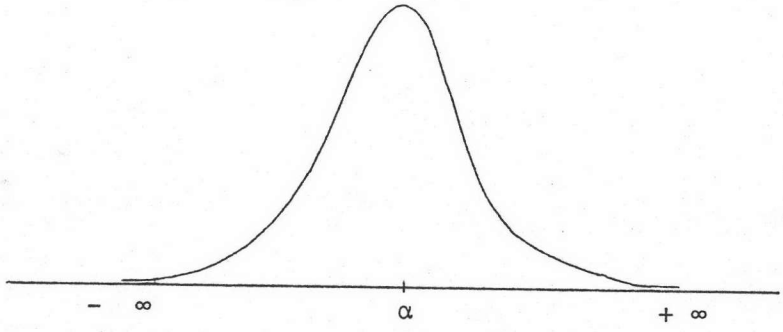
ค่าคาดหวัง $E(X) = \alpha$

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

1.5.12.3 การแจกแจงแบบดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution)

ฟังก์ชั่นความน่าจะเป็น คือ

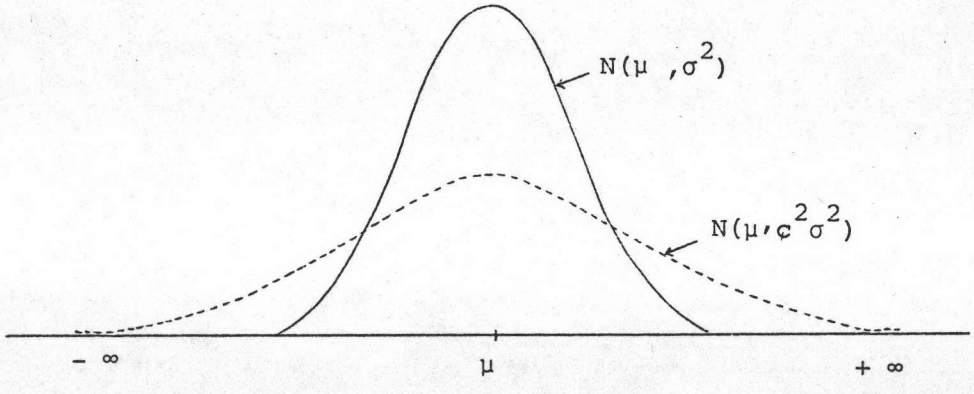
$$f(x) = \frac{1}{2\beta} \cdot e^{-\frac{|x-\alpha|}{\beta}}, \quad -\infty < x < \infty$$



ค่าคาดหวัง $E(X) = \alpha$

ค่าความแปรปรวน $V(X) = 2\beta^2$

1.5.12.4 การแจกแจงแบบปกติปลอมปน (Scale Contaminated Normal Distribution)



ลักษณะการแจกแจงแบบปกติปลอมปนที่พิจารณาในวิทยานิพนธ์นี้ เป็นการแจกแจงที่แปลงมาจากการแจกแจงแบบปกติ ซึ่งมีฟังก์ชั่นการแปลงเป็นดังนี้

$$F = (1-p)\, N(\mu,\sigma^2) + pN(\mu,c^2\sigma^2), \quad c > 0$$

หมายความว่าค่า x จะมาจากการแจกแจง $N(\mu, \sigma^2)$ ด้วยความน่าจะเป็น $1-p$ และจากการแจกแจง $N(\mu, c^2\sigma^2)$ ด้วยความน่าจะเป็น p

μ และ σ^2 เป็นพารามิเตอร์ที่กำหนดค่าเฉลี่ย และค่าความแปรปรวนของความคลาดเคลื่อน

p และ c เป็นพารามิเตอร์ที่กำหนดสัดส่วนการปลอมปน และสเกลแฟคเตอร์

1.5.13 การจำลองการทดลองจะกระทำซ้ำกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

1.6.1 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) หรือ MSE ของตัวประมาณ คือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}$ คือ $E(\hat{\theta} - \theta)^2$

1.6.2 ความแปรปรวน (Variance) ของตัวประมาณ คือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว ความแปรปรวนของ $\hat{\theta}$ คือ $E(\hat{\theta} - E(\hat{\theta}))^2$

1.6.3 ตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง (Best Linear Unbiased Estimator) หรือ BLUE เป็นคุณสมบัติหนึ่งของตัวประมาณ โดยตัวประมาณ $\hat{\theta}$ จะมีคุณสมบัติเป็น BLUE ของพารามิเตอร์ θ ถ้า $\hat{\theta}$ มีคุณสมบัติครบ 3 ข้อต่อไปนี้คือ

1.6.3.1 เป็นฟังก์ชั่นเชิงเส้นของตัวอย่างสุ่ม

1.6.3.2 เป็นตัวประมาณที่ไม่เอนเอียง

1.6.3.3 เป็นตัวประมาณที่มีความแปรปรวนต่ำสุด

1.6.4 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency) หรือ RE ของตัวประมาณ $\hat{\theta}_1$ เมื่อเทียบกับ $\hat{\theta}_2$ เป็นการเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณ 2 ตัว ที่ประมาณค่าพารามิเตอร์เดียวกัน ในรูปอัตราส่วนของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณหนึ่ง ต่อค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของอีกตัวหนึ่ง หรืออาจจะพิจารณาจากสูตร

$$RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}$$

1.6.5 ประสิทธิภาพสัมพัทธ์ใกล้อนันต์ (Asymptotic Relative Efficiency) หรือ ARE ของตัวประมาณ $\hat{\theta}_1$ เมื่อเทียบกับ $\hat{\theta}_2$ เป็นการเปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อ n มีค่าเข้าใกล้อนันต์ ($n \to \infty$) ของตัวประมาณ 2 ตัวที่ประมาณค่าพารามิเตอร์เดียวกัน เช่นเดียวกับประสิทธิภาพสัมพัทธ์

1.6.6 ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมติฐาน H_o เมื่อสมมติฐาน H_o ถูก

1.6.7 ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมติฐาน H_o เมื่อสมมติฐาน H_o ผิด

1.6.8 อำนาจของการทดสอบ (Power of the test) คือความน่าจะเป็นที่จะปฏิเสธสมมติฐาน H_o เมื่อสมมติฐาน H_o ผิด

1.6.9 ความแกร่ง (Robustness) ของการทดสอบ หมายถึง คุณสมบัติของการทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่น การฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น ซึ่งสิ่งที่ใช้พิจารณาความแกร่งของการทดสอบคือ ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1

1.7 ประโยชน์ของการวิจัย

เพื่อช่วยให้ข้อสรุปที่เป็นประโยชน์ต่อผู้ใช้ในการเลือกใช้ตัวประมาณและสถิติทดสอบเกี่ยวกับพารามิเตอร์ ได้อย่างมีประสิทธิภาพและเหมาะสม เมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น