บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

ในการศึกษาวิจัยโดยทั่วไป เช่น การศึกษาทางด้านสังคมศาสตร์ จะเห็นว่าผู้วิจัยมีจุดมุ่งหมายที่ต้องการข้อสรุปจากข้อมูลทั้งหมดเกี่ยวกับสิ่งที่สนใจศึกษา แต่ไม่สามารถศึกษาจากข้อมูลทั้งหมดได้ เนื่องจากอาจมีข้อจำกัดในเรื่องของเวลา งบประมาณ หรือกำลังคน จึงจำเป็นต้องทำการศึกษาจากข้อมูลเพียงบางส่วนหรือที่เรียกว่า ข้อมูลตัวอย่าง (sample data) และนำผลของการศึกษาข้อมูลตัวอย่างนั้นไปใช้ในการอธิบายลักษณะของข้อมูลทั้งหมดที่สนใจศึกษา ที่เรียกว่าประชากร (population) การอาศัยข้อมูลเพียงบางส่วนหรือข้อมูลตัวอย่าง เพื่อนำไปอธิบายลักษณะของประชากรที่สนใจศึกษานี้ เปรียบเสมือนการสรุปผลจากกรณีเฉพาะไปสู่กรณีทั่วไป ซึ่งเราเรียกกันว่า การอนุมาน (Inductive inference) ดังนั้นการศึกษาข้อมูลตัวอย่างและใช้วิธีการทางสถิติมาทำการหาข้อสรุปเกี่ยวกับประชากรจากตัวอย่างนั้น จึงเรียกกันว่า การอนุมานเชิงสถิติ (Statistical Inference)

จากที่กล่าวมาจะเห็นได้ว่า เมื่อผู้วิจัยต้องการทราบข้อสรุปเกี่ยวกับสิ่งที่สนใจศึกษา ซึ่งเป็นความต้องการทราบลักษณะของประชากรทั้งหมด ไม่ใช่แต่เพียงตัวอย่างเท่านั้น แต่เมื่อมีความจำเป็นต้องศึกษาจากข้อมูลตัวอย่างเนื่องจากมีข้อจำกัดหลายประการดังที่ได้กล่าวมาแล้ว ดังนั้นจึงมีความจำเป็นต้องอาศัยวิธีการอนุมานเชิงสถิติเพื่อศึกษาลักษณะต่าง ๆ ของประชากร หรือที่เรียกกันว่า พารามิเตอร์ (parameter) ที่ผู้วิจัยต้องการทราบจากข้อมูลตัวอย่างที่มีอยู่ โดยทั่วไปในการวิเคราะห์ข้อมูลเบื้องต้น ค่าพารามิเตอร์ที่สนใจศึกษามีหลายตัวเช่น ค่าเฉลี่ย ผลต่างระหว่างค่าเฉลี่ย ค่าสัดส่วน และค่าความแปรปรวน เป็นต้น ซึ่งในการอนุมานเชิงสถิติเพื่อศึกษาเกี่ยวกับค่าพารามิเตอร์ดังกล่าวสามารถกระทำได้ในหลายลักษณะเช่น การประมาณค่า (Estimation) การทดสอบสมมติฐาน (Hypothesis testing) เป็นต้น สำหรับการวิจัยครั้งนี้สนใจศึกษาการประมาณค่า ซึ่งโดยทั่วไปการประมาณค่าพารามิเตอร์สามารถทำการประมาณได้ในสองรูปแบบคือการประมาณแบบจุด (Point Estimation) และการประมาณแบบช่วง (Interval Estimation) สำหรับการประมาณแบบจุดเป็นการประมาณค่าพารามิเตอร์ด้วยค่า ๆ หนึ่งหรือจุด ๆ หนึ่ง ส่วนการประมาณค่าแบบช่วงเป็นการประมาณที่จะให้ช่วง ๆ หนึ่ง ซึ่งมีคุณสมบัติว่าค่าที่แท้จริงของประชากรจะอยู่ในช่วงที่ประมาณได้ด้วยความเชื่อมั่นระดับหนึ่ง ซึ่งการประมาณค่าแบบจุด ค่าประมาณที่ได้จะคลาดเคลื่อนไปจากค่าจริงหรือพารามิเตอร์เพียงใด ขึ้นอยู่กับการเลือกใช้ตัวประมาณที่เหมาะสม ซึ่งโดยทั่วไปในการเลือกตัวประมาณมักเลือกตามหลักเกณฑ์ที่ใช้ในการพิจารณาคัดเลือกตัวประมาณ ซึ่งมีหลายประการด้วยกันยกตัวอย่างเช่น ความไม่เอนเอียง (Unbiasedness) ความคงเส้นคงวา (Consistency) ความมีประสิทธิภาพ (Efficiency) เป็นต้น ส่วนการประมาณค่าแบบช่วงเป็นการประมาณโดยอาศัยตัวประมาณแบบจุด และการแจกแจงความน่าจะเป็นของตัวประมาณนั้น ซึ่งผลจากการประมาณจะทำให้ผู้วิจัยเชื่อมั่นได้ในระดับหนึ่งว่า ช่วงที่ประมาณได้คลุมค่าพารามิเตอร์ที่สนใจศึกษา จะเห็นได้ว่าการประมาณค่าแบบช่วงสามารถบอกขอบเขตของค่าประมาณได้ดีกว่าการประมาณค่าแบบจุด ซึ่งค่าประมาณที่ได้เป็นเพียงค่า ๆ เดียว นอกจากนี้ความกว้างของช่วงที่ประมาณได้ ยังมีประโยชน์ในการที่จะบ่งบอกถึงคุณภาพของตัวประมาณแบบจุดว่าเป็นตัวประมาณที่เหมาะสมหรือไม่ ทั้งนี้เพราะถ้าช่วงที่ประมาณได้กว้างมากแสดงว่าค่าประมาณที่ได้ต่างจากค่าจริงของพารามิเตอร์มากซึ่งอาจเป็นผลมาจากตัวประมาณแบบจุดที่ใช้อาจไม่เหมาะสม ผู้วิจัยจึงอาจต้องย้อนกลับไปพิจารณาคุณสมบัติของตัวประมาณแบบจุดที่ใช้ใหม่ หรือหาตัวประมาณแบบจุดในรูปแบบอื่นที่อาจทำให้ค่าประมาณที่ได้ดีขึ้น

การศึกษาการประมาณค่าแบบช่วง โดยทั่วไปค่าพารามิเตอร์ที่นำมาศึกษากันมีหลายตัว เช่น ค่าเฉลี่ย สัดส่วน ความแปรปรวน เป็นต้น แต่ในการวิจัยครั้งนี้ผู้วิจัยสนใจที่จะศึกษาการประมาณแบบช่วงสำหรับความแปรปรวน (σ^2) ของประชากรที่มีการแจกแจงแบบปกติ ซึ่งเป็นพารามิเตอร์ตัวหนึ่งที่มีความสำคัญไม่น้อยในการอธิบายลักษณะของประชากร เช่น เป็นค่าที่ใช้วัดการกระจายของข้อมูลแต่ละชุด และสำหรับในกรณีที่ต้องการเปรียบเทียบการกระจายของประชากรตั้งแต่สองชุดขึ้นไป และประชากรแต่ละชุดมีหน่วยวัดที่แตกต่างกัน เรายังใช้อัตราส่วนของค่าส่วนเบี่ยงเบนมาตรฐาน (σ) ต่อค่าเฉลี่ย หรือที่เรียกว่า สัมประสิทธิ์ความแปรปรวน (Coefficient of Variation) เป็นค่าที่ใช้เปรียบเทียบการกระจายของข้อมูลอีกด้วย ซึ่งถ้าข้อมูลชุดใดมีการกระจายสูง ก็อาจเป็นสัญญาณบ่งบอกให้ผู้วิจัยทราบว่า อาจมีสิ่งผิดปกติในขั้นตอนการสุ่มตัวอย่าง หรือขนาดตัวอย่างที่ใช้อาจเล็กเกินไป

ในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติโดยทั่วไปเราใช้ $S^2 = [\sum_{i=1}^{n}(X_i-\bar{X})^2]/(n-1)$ เป็นตัวประมาณแบบจุดของ σ^2 เนื่องจากเป็นตัวประมาณที่ไม่เอนเอียงและมีความแปรปรวนต่ำสุดของ σ^2 และ ทราบว่า $(n-1)S^2/\sigma^2$ มีการแจกแจงแบบ χ^2 ที่มีองศาอิสระ เท่ากับ n-1 ดังนั้นในการประมาณแบบช่วง จะได้ว่า

$$Pr[a < (n-1)S^2/\sigma^2 < b] = 1-\alpha$$

หรือ $$Pr[(n-1)S^2/b < \sigma^2 < (n-1)S^2/a] = 1-\alpha$$

นั่นคือ การประมาณแบบช่วงของ σ^2 จะอยู่ในรูป $(n-1)S^2/b < \sigma^2 < (n-1)S^2/a$ โดยปกติเพื่อความสะดวกในการหาค่า a และ b เรามักจะกำหนดให้ $Pr(\chi^2_{(n-1)} < a)$ มีค่าเท่ากับ $Pr(\chi^2_{(n-1)} > b) = \alpha/2$ จากนั้นหาค่า a และ b จากตารางค่าวิกฤตของไคสแควร์จึงได้ช่วงความเชื่อมั่น $100(1-\alpha)$ เปอร์เซ็นต์ของ σ^2 ที่ปรากฏในหนังสือสถิติพื้นฐานโดยทั่วไปคือ $(n-1)s^2/\chi^2_{\alpha/2} < \sigma^2 < (n-1)s^2/\chi^2_{(1-\alpha/2)}$ และในที่นี้จะขอเรียก การประมาณช่วงความเชื่อมั่นด้วยวิธีนี้ว่า วิธีไคสแควร์ (Chi-square Confidence Interval) ซึ่งการประมาณช่วงความเชื่อมั่นด้วยวิธีดังกล่าวจะให้ "ช่วงความเชื่อมั่นที่สั้นที่สุด ถ้าการแจกแจงมีลักษณะสมมาตร" ดังนั้นวิธีไคสแควร์จะเป็นวิธีที่เหมาะสม ถ้าตัวอย่างมีขนาดใหญ่ เพราะการแจกแจงไคสแควร์จะมีลักษณะเข้าใกล้การแจกแจงแบบปกติ ส่วนในกรณีที่ตัวอย่างมีขนาดเล็ก การแจกแจงไคสแควร์เป็นการแจกแจงที่ไม่สมมาตร จะมีผลทำให้การประมาณช่วงความเชื่อมั่นของ σ^2 ด้วยวิธีไคสแควร์ ไม่ได้ช่วงความเชื่อมั่นที่สั้นที่สุด

จากปัญหาที่ได้กล่าวมาแล้ว จึงได้มีผู้ศึกษาเกี่ยวกับ วิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติ เพื่อให้ได้ช่วงความเชื่อมั่นที่เหมาะสมกว่าการประมาณช่วงความเชื่อมั่นด้วยวิธีไคสแควร์ที่พบในหนังสือสถิติพื้นฐานทั่วไป ซึ่งผลงานของนักสถิติที่น่าสนใจมีดังต่อไปนี้

ในปี ค.ศ.1959 เทต (Tate, R. F.) และ เคล็ท (Klett, G. W.) ทำการศึกษาเกี่ยวกับช่วงความเชื่อมั่นที่เหมาะสมที่สุด สำหรับความแปรปรวนของการแจกแจงแบบปกติซึ่งในการศึกษาครั้งนั้น เทตและเคล็ท นำวิธีการทางคณิตศาสตร์เพื่อหาความยาวที่สั้นที่สุดของช่วงความเชื่อมั่นที่มีรูปแบบ $[(n-1)s^2/b, (n-1)s^2/a]$ ซึ่งผลจากการศึกษาจะได้ค่า a และ b ซึ่งในที่นี้จะขอเรียกว่า ตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีช่วงความเชื่อมั่นที่สั้นที่สุด (Divisors for the Confidence Interval of Minimum Length) และจะเรียกการประมาณช่วงความเชื่อมั่นด้วยวิธีนี้ว่า วิธีช่วงความเชื่อมั่นที่สั้นที่สุด (Confidence Interval of Minimum Length)

ต่อมาในปี ค.ศ.1960 ลีนเลย์ (Lindley, D. V.) อีส (East, D. A.) และแฮมมิลตัน (Hamilton, P. A.) ศึกษาเกี่ยวกับการอนุมานค่าความแปรปรวนของการแจกแจงแบบปกติอีกวิธีการหนึ่ง โดยใช้หลักการของเบส์หาค่าของ a และ b เพื่อนำไปใช้ในการคำนวณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ ซึ่งในที่นี้จะขอเรียกค่า a และ b ว่าตัวหารสำหรับช่วงความเชื่อมั่นโดยวิธีของเบส์ (Divisors for the Bayesian Confidence Interval) และเรียกการประมาณช่วงความเชื่อมั่นด้วยวิธีนี้ว่าวิธีของเบส์ (Bayesian Confidence Interval)

จากผลงานของนักสถิติหลายท่านที่ได้กล่าวมาแล้ว ผู้วิจัยจึงเห็นว่าเป็นที่น่าสนใจที่จะศึกษาเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติ 3 วิธีการคือ

1. วิธีไคสแควร์ (Chi-square Confidence Interval)
2. วิธีช่วงความเชื่อมั่นที่สั้นที่สุด (Confidence Interval of Minimum Length)
3. วิธีของเบส์ (Bayesian Confidence Interval)

ซึ่งผลที่ได้จากการวิจัย นอกจากจะทำให้ทราบวิธีการประมาณช่วงความเชื่อมั่นที่เหมาะสมที่สุดในระหว่าง 3 วิธีการข้างต้น สำหรับความความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติแล้ว ยังสร้างตารางสำเร็จรูปแสดงค่าตัวหารสำหรับช่วงความเชื่อมั่น ของวิธีการประมาณที่เหมาะสมนั้นด้วย โดยมีค่าระดับนัยสำคัญและขนาดตัวอย่างให้เลือกใช้หลายระดับ

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของการแจกแจงแบบปกติ 3 วิธีการ คือ

1.2.1 วิธีไคสแควร์

1.2.2 วิธีช่วงความเชื่อมั่นที่สั้นที่สุด

1.2.3 วิธีของเบส์

1.3 สมมติฐานของการวิจัย

1.3.1 ในกรณีตัวอย่างขนาดเล็ก วิธีช่วงความเชื่อมั่นที่สั้นที่สุด จะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่า วิธีไคสแควร์ และ วิธีของเบส์

1.3.2 ในกรณีขนาดตัวอย่างใหญ่ ช่วงความเชื่อมั่นทั้ง 3 วิธี จะให้ค่าระดับความเชื่อมั่น และค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นใกล้เคียงกัน

1.4 ข้อตกลงเบื้องต้น

$X_1, X_2, \ldots, X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีการแจกแจงปกติ ซึ่งมีพารามิเตอร์ที่ไม่ทราบค่า μ และ σ^2 โดยที่

μ คือ ค่าเฉลี่ยประชากร

σ^2 คือ ความแปรปรวนประชากร

$\bar{X}$ คือ ค่าเฉลี่ยตัวอย่าง เท่ากับ $(1/n)\sum_{i=1}^{n} X_i$

S^2 คือ ความแปรปรวนตัวอย่าง เท่ากับ $[\sum_{i=1}^{n}(X_i-\bar{X})^2]/(n-1)$

n คือ ขนาดตัวอย่าง

r คือ องศาอิสระ เท่ากับ $n-1$

1.5 ขอบเขตของการวิจัย

1.5.1 กำหนดขนาดตัวอย่างที่ใช้คือ n มีค่าตั้งแต่ 2 ถึง 50

1.5.2 กำหนดให้ข้อมูลมีค่า สัมประสิทธิ์ความแปรปรวน 5%,10%,15% และ 20% (พารามิเตอร์ของประชากรที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย เท่ากับ 50 และความแปรปรวนเท่ากับ 6.25,25,56.25 และ 100)

1.5.3 กำหนดค่าสัมประสิทธิ์ความเชื่อมั่น 4 ระดับคือ $(1-\alpha)$ เท่ากับ 0.90,0.95 ,0.99 และ 0.995

1.5.4 ในการวิจัยครั้งนี้สร้างแบบจำลองข้อมูล โดยใช้เทคนิคมอนติคาร์โลซิมูเลชัน (Monte Carlo Simulation Technique) เขียนโปรแกรมด้วยภาษา FORTRAN 77 ทำการทดลองซ้ำ 2000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

1.6.1 สัมประสิทธิ์ความเชื่อมั่น (Confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าของพารามิเตอร์ของประชากร

1.6.2 ช่วงความเชื่อมั่น (Confidence interval) หมายถึง ช่วงตัวอย่างที่คำนวณจากข้อมูลตัวอย่างหนึ่งชุดใด ๆ ซึ่งใช้ในการประมาณค่าพารามิเตอร์แบบช่วง

1.7 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1.7.1 ทราบวิธีการประมาณช่วงความเชื่อมั่นที่เหมาะสม ในการประมาณช่วงความเชื่อมั่นสำหรับความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติ เพื่อเป็นแนวทางในการศึกษาวิจัยต่อไป

1.7.2 สร้างตารางแสดงค่าตัวหารสำหรับช่วงความเชื่อมั่น สำหรับวิธีการประมาณที่ให้ผลดีที่สุดจากการวิจัยครั้งนี้ เพื่อเป็นประโยชน์แก่ผู้ที่จะนำไปใช้ในการคำนวณช่วงความเชื่อมั่นสำหรับความแปรปรวนของประชากรที่มีการแจกแจงแบบปกติ