บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การอนุมานทางสถิติเป็นการศึกษาข้อมูลตัวอย่างและใช้วิธีการทางสถิติมาทำการหาข้อสรุปเกี่ยวกับสิ่งที่สนใจศึกษาของประชากร ส่วนหนึ่งที่สำคัญ คือการประมาณค่า (estimation) ซึ่งวิธีการประมาณค่านี้มีวัตถุประสงค์เพื่อนำค่าสถิติที่ได้จากตัวอย่างไปประมาณค่าพารามิเตอร์ของประชากร และนอกจากนี้การประมาณค่ายังมีความสำคัญในทุกๆสาขาวิชาไม่ว่าจะเป็นทางธุรกิจ สังคม เศรษฐกิจ วิทยาศาสตร์ และในสาขาอื่นๆ โดยทั่วไปการประมาณค่าพารามิเตอร์สามารถทำการประมาณได้ในสองรูปแบบคือการประมาณค่าแบบจุด (point estimation) และการประมาณแบบช่วง (interval estimation) สำหรับการประมาณแบบจุดเป็นการประมาณค่าพารามิเตอร์ด้วยค่าๆหนึ่งหรือจุดๆหนึ่ง ในการประมาณแบบจุดมีโอกาสคลาดเคลื่อนจากพารามิเตอร์มากหรือน้อยขึ้นกับการเลือกใช้ตัวประมาณ ซึ่งมีหลักเกณฑ์ที่ใช้ในการพิจารณาคัดเลือกตัวประมาณหลายประการ เช่น ความเพียงพอ ความไม่เอนเอียง ความคงเส้นคงวา และความมีประสิทธิภาพ เป็นต้น ส่วนการประมาณแบบช่วงเป็นการประมาณโดยอาศัยการประมาณแบบจุดและการแจกแจงของตัวประมาณแบบจุด ซึ่งช่วงความเชื่อมั่นจะแคบหรือกว้างขึ้นอยู่กับสัมประสิทธิ์ความเชื่อมั่น ความแปรปรวนของตัวประมาณและการแจกแจงของตัวสถิติทดสอบ ผลจากการประมาณแบบช่วงนี้ ทำให้เราเชื่อมั่นได้ในระดับหนึ่งว่าค่าพารามิเตอร์ที่สนใจอยู่ในขอบเขตของช่วงความเชื่อมั่นที่ประมาณได้

ในการประมาณค่าเฉลี่ยแบบจุดของประชากรหนึ่งที่มีการแจกแจงปกติ ซึ่งมีการใช้ตัวอย่างหลายกลุ่มตัวอย่าง โดยที่มีวิธีการหรือเครื่องมือแตกต่างกันเป็นตัววัดค่าที่สนใจศึกษา เพื่อหาข้อสรุปที่นำไปสู่การอธิบายลักษณะของประชากร เช่น การประมาณคุณภาพเฉลี่ยของผลิตภัณฑ์ที่ได้จากกระบวนการผลิตเดียวกันแต่ในการวัดคุณภาพใช้เครื่องมือตรวจสอบต่างกัน และการประมาณปริมาณสารพิษในแม่น้ำ โดยใช้ห้องทดลองที่แตกต่างกันตรวจสอบ ในกรณีเหล่านี้การประมาณค่าเฉลี่ยแบบจุด โดยไม่คำนึงถึงความแปรปรวนที่แตกต่างกันซึ่งเกิดจากวิธีการหรือเครื่องมือวัดที่ต่างกัน จึงมีโอกาสที่จะเกิดความคลาดเคลื่อนขึ้นได้ และไม่เหมาะสมที่จะนำไปทำการประมาณแบบช่วง จากปัญหาดังกล่าวจึงมีการพิจารณาหารูปแบบของตัวประมาณแบบจุดในกรณีลักษณะเช่นนี้ โดยในปี ค.ศ. 1959 เกรยบิล และดีล (Graybill and Deal) ได้ศึกษาหาตัวประมาณค่าเฉลี่ยของสองกลุ่มตัวอย่างที่เป็นตัวแทนของแต่ละประชากร โดยตัวประมาณแต่ละตัวที่ได้จากกลุ่มตัวอย่างทั้งสองเป็นตัวประมาณไม่เอนเอียง แต่ตัวประมาณดังกล่าวมีความแปรปรวนไม่เท่ากัน ในการศึกษาหาตัวประมาณค่าเฉลี่ยนี้ เกรยบิลและดีล ได้ให้นิยามของตัว

ประมาณที่ได้จากผลรวมเชิงเส้นของตัวประมาณที่มีการถ่วงน้ำหนักด้วยน้ำหนักแบบสุ่ม ซึ่งตัวประมาณที่ได้เป็นตัวประมาณไม่เอนเอียงดีกว่าอย่างสม่ำเสมอ (uniformly better unbiased estimator) กล่าวคือ ตัวประมาณนี้มีความแปรปรวนน้อยกว่าความแปรปรวนของตัวประมาณทั้งสองค่า ตัวประมาณที่เกรยบิลและดีลเสนอนั้นเป็นที่ยอมรับโดยทั่วไปในการใช้เป็นตัวประมาณแบบจุดสำหรับค่าเฉลี่ยของสองกลุ่มประชากรหรือหลายกลุ่มประชากรที่มีการแจกแจงปกติ และได้มีการศึกษาพัฒนาหาตัวประมาณค่าเฉลี่ยโดยพิจารณาหลักเกณฑ์การคัดเลือกตัวประมาณประกอบเพิ่มเติมด้วย ในปี ค.ศ. 1974 โคเฮ็นและแซคโรวิซ (Cohen and Sackrowitz) ได้ศึกษาหาตัวประมาณค่าเฉลี่ยของสองประชากรที่มีการแจกแจงปกติและเสนอตัวประมาณค่าเฉลี่ย ในกรณีที่ใช้ขนาดตัวอย่างเท่ากัน โดยใช้ฟังก์ชันความสูญเสียเป็นเกณฑ์ ซึ่งตัวประมาณที่ได้เป็นตัวประมาณไม่เอนเอียงและเป็นตัวประมาณมินิแมกซ์ เมื่อขนาดตัวอย่างมากกว่าหรือเท่ากับ 5

สำหรับการประมาณแบบช่วงนั้นมักมีปัญหาเกิดขึ้นเนื่องจากไม่ทราบว่าการแจกแจงตัวอย่างของตัวประมาณที่แท้จริงคืออะไร แต่อาจทราบว่าค่าความคาดหวังและความแปรปรวนของตัวประมาณมีลักษณะเป็นอย่างไรเท่านั้น การทำการประมาณแบบช่วงจึงต้องอาศัยการสมมติการแจกแจงตัวอย่างของตัวประมาณขึ้นเสียก่อน ซึ่งส่วนใหญ่การแจกแจงตัวอย่างที่นำมาพิจารณามักเป็นการแจกแจงโดยประมาณและช่วงความเชื่อมั่นที่ได้เป็นช่วงความเชื่อมั่นโดยประมาณ จึงได้มีนักสถิติเสนอวิธีหาช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของสองประชากรหรือมากกว่าสองประชากรที่มีการแจกแจงปกติ ในกรณีความแปรปรวนไม่เท่ากัน โดยใช้การรวมช่วงความเชื่อมั่นที่มีตัวประมาณแบบจุดที่เป็นค่าเฉลี่ยของตัวอย่างจากแต่ละประชากรซึ่งเป็นช่วงความเชื่อมั่นที่เป็นอิสระซึ่งกันและกัน โดยช่วงความเชื่อมั่นที่ได้ต้องมีความน่าจะเป็นที่ครอบคลุมค่าพารามิเตอร์ตามที่กำหนดไว้ ผลงานของนักสถิติที่ได้เสนอวิธีการประมาณแบบช่วงคือในปี ค.ศ.1972 แฟรเวทเทอร์ (Fairweather) ได้ศึกษาหาช่วงความ.เชื่อมั่น โดยใช้ผลรวมเชิงเส้นของตัวสถิติสติวเดนท์ที ซึ่งในที่นี้ผู้วิจัยขอเรียกวิธีดังกล่าวว่า วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติสติวเดนท์ที (interval estimation method with linear combination of t-statistics) และในปี ค.ศ. 1996 จอร์แดนและคริสนามรตี้ (jordan and Krishnamoorthy) ได้ศึกษาหาช่วงความเชื่อมั่น โดยใช้ผลรวมเชิงเส้นของตัวสถิติเอฟ ผู้วิจัยขอเรียกวิธีดังกล่าวว่า วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติเอฟ (interval estimation method with linear combination of F-statistics) นอกจากนี้จอร์แดนและคริสนามรตี้ ได้นำตัวสถิติทดสอบซึ่งเป็นค่าสัมบูรณ์ที่มากที่สุดของตัวสถิติสติวเดนท์ทีมาหาช่วงความเชื่อมั่นโดยใช้ความสัมพันธ์ระหว่างการทดสอบสมมติฐานและการประมาณแบบช่วง ซึ่งผู้วิจัยขอเรียกวิธีดังกล่าวว่า วิธีการประมาณแบบช่วงด้วยค่าสัมบูรณ์ที่มากที่สุดของตัวสถิติสติวเดนท์ที (interval estimation method with absolute maximum of t-statistics)

ในการวิจัยครั้งนี้ผู้วิจัยจะศึกษาเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของสองกลุ่มประชากรที่มีการแจกแจงปกติและเป็นอิสระซึ่งกันและกัน โดยที่ความแปรปรวนไม่เท่ากัน การศึกษาเปรียบเทียบจะใช้วิธีการประมาณแบบช่วงที่เสนอโดยนักสถิติที่กล่าวข้างต้น ซึ่งมี 3 วิธี คือ

1. วิธีการประมาณแบบช่วงด้วยค่าสัมบูรณ์ที่มากที่สุดของตัวสถิติสติวเดนท์ที (Mt)
2. วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติสติวเดนท์ที (Ct)
3. วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติเอฟ (CF)

ผลการวิจัยจะทำให้ทราบว่าในสถานการณ์ที่แตกต่างกันของความแปรปรวนและขนาดตัวอย่าง วิธีการประมาณแบบช่วงวิธีการใดเป็นให้ช่วงความเชื่อมั่นที่เหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของสองกลุ่มประชากรที่มีการแจกแจงปกติและเป็นอิสระซึ่งกันและกัน ในกรณีที่ค่าความแปรปรวนไม่เท่ากัน โดยมีวิธีการประมาณแบบช่วง 3 วิธี คือ

1. วิธีการประมาณแบบช่วงด้วยค่าสัมบูรณ์ที่มากที่สุดของตัวสถิติสติวเดนท์ที
2. วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติสติวเดนท์ที
3. วิธีการประมาณแบบช่วงด้วยผลรวมเชิงเส้นของตัวสถิติเอฟ

1.3 สมมติฐานการวิจัย

1. ภายใต้อัตราส่วนความแปรปรวนของประชากรที่แตกต่างกันมาก ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยที่ได้จากวิธีการประมาณแบบช่วง CF จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าวิธีการประมาณ Mt และวิธีการประมาณ Ct

2. ภายใต้อัตราส่วนความแปรปรวนของประชากรที่แตกต่างกันมาก ช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยที่ได้จากวิธีการประมาณแบบช่วง CF จะให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีการประมาณ Mt และวิธีการประมาณ Ct

1.4 ข้อตกลงเบื้องต้น

1. ให้ X_{ij} $\left(i = 1,2 \ ; \ j = 1,2,...n_i \right)$ เป็นตัวแปรสุ่มที่เป็นอิสระซึ่งกันและกันจากประชากรสองกลุ่มที่มีการแจกแจงปกติ ซึ่งมีค่าเฉลี่ยเดียวกันคือ μ และมีความแปรปรวนต่างกันคือ σ_1^2, σ_2^2 ตามลำดับ โดยที่

μ คือค่าเฉลี่ยประชากร

σ_i^2 คือความแปรปรวนของประชากรที่ i

$\overline{X}_i$ คือค่าเฉลี่ยตัวอย่างจากประชากรที่ i เท่ากับ $\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$

S_i^2 คือความแปรปรวนตัวอย่างจากประชากรที่ i เท่ากับ $\frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2$

n_i คือขนาดตัวอย่างจากประชากรที่ i

และ i คือกลุ่มประชากรที่ i =1,2

2. ให้ $T_i = \frac{\overline{X}_i - \mu}{S_i/\sqrt{n_i}}$; i =1,2 เป็นตัวสถิติที่เป็นอิสระซึ่งกันและกัน และ $F_i = \frac{n_i\left(\overline{X}_i-\mu\right)^2}{S_i^2}$; i =1,2 เป็นตัวสถิติที่เป็นอิสระซึ่งกันและกันเช่นเดียวกัน

1.5 ขอบเขตการวิจัย

ในการวิจัยนี้ ผู้วิจัยศึกษาเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของสองประชากรที่มีการแจกแจงปกติ ในกรณีที่ความแปรปรวนไม่เท่ากัน จึงกำหนดปัจจัยที่คาดว่าจะมีผลต่อการศึกษาดังต่อไปนี้

1.5.1 <u>อัตราส่วนความแปรปรวน</u>

1.5.1.1 <u>อัตราส่วนความแปรปรวนของประชากรที่ 1 น้อยกว่าประชากรที่ 2</u> $\left(\sigma_1^2 < \sigma_2^2\right)$

ผู้วิจัยกำหนดอัตราส่วนความแปรปรวนของประชากรให้มีความแตกต่างเป็น 3 ระดับ คือน้อย ปานกลางและมาก ตามวิธีการของเกมและโปรเบิท (Game and Probert) โดยใช้ค่านอนเซ็นทรัลลิตี้พารามิเตอร์ (noncentrality parameter) ϕ $(\phi > 0)$ เป็นตัววัดระดับความแตกต่างของความแปรปรวนประชากร

$$\phi^2 = \sum_i\left(\sigma_i^2 - \overline{\sigma}^2\right)^2 \Big/ \left(k\sigma_1^2\right) \qquad ; \ i=1,2,\ldots,k$$

เมื่อ $\overline{\sigma}^2$ คือค่าเฉลี่ยความแปรปรวนของ k ประชากร

σ_1^2 คือความแปรปรวนของประชากรกลุ่มที่ 1

และ σ_k^2 คือความแปรปรวนของประชากรกลุ่มที่ k

สำหรับงานวิจัยนี้ ผู้วิจัยกำหนดอัตราส่วนความแปรปรวนของประชากรดังนี้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ก. <u>อัตราส่วนความแปรปรวนแตกต่างกันน้อย</u> $(0 < \phi < 1.5)$

โดยพิจารณาอัตราส่วนความแปรปรวนของประชากร $(\sigma_1^2 : \sigma_2^2)$ ดังนี้

$\sigma_1^2 : \sigma_2^2$	ϕ
1 : 1.5	0.25
1 : 2.5	0.75
1 : 3.5	1.25

ข. <u>อัตราส่วนความแปรปรวนแตกต่างกันปานกลาง</u> $(1.5 \leq \phi < 3)$

โดยพิจารณาอัตราส่วนความแปรปรวนของประชากร $(\sigma_1^2 : \sigma_2^2)$ ดังนี้

$\sigma_1^2 : \sigma_2^2$	ϕ
1 : 4.5	1.75
1 : 5.5	2.25
1 : 6.5	2.75

ค. <u>อัตราส่วนความแปรปรวนแตกต่างกันมาก</u> $(\phi \geq 3)$

โดยพิจารณาอัตราส่วนความแปรปรวนของประชากร $(\sigma_1^2 : \sigma_2^2)$ ดังนี้

$\sigma_1^2 : \sigma_2^2$	ϕ
1 : 7.5	3.25
1 : 10	4.5
1 : 15	7

และกำหนดให้ค่าความแปรปรวนของประชากรกลุ่มที่หนึ่งเท่ากับ 10 $(\sigma_1^2 = 10)$

1.5.1.2 <u>อัตราส่วนความแปรปรวนของประชากรที่ 1 มากกว่าประชากรที่ 2</u> $\left(\sigma_1^2 > \sigma_2^2\right)$

กำหนดอัตราส่วนความแปรปรวนของประชากรในลักษณะเดียวกันกับกรณี 1.5.1.1 โดยเปลี่ยนรูปแบบเป็น $\sigma_2^2 : \sigma_1^2$

1.5.2 <u>ขนาดตัวอย่าง</u>

1.5.2.1 <u>ขนาดตัวอย่างเท่ากัน</u>

กำหนดให้ขนาดตัวอย่างทั้งสองกลุ่มประชากรเท่ากันคือเท่ากับ 10 , 20 , 30 , 50

1.5.2.2 ขนาดตัวอย่างไม่เท่ากัน

กำหนดโดยใช้อัตราส่วนของขนาดตัวอย่างและค่าความแตกต่างของขนาดตัวอย่าง ดังนั้นขนาดตัวอย่างของประชากรที่หนึ่งกับประชากรที่สอง ($n_1 : n_2$) คือ

ค่า $n_1 : n_2$	ความแตกต่างของขนาดตัวอย่าง				
อัตราส่วน($n_1 : n_2$)	5	10	15	20	40
1 : 3	-	-	-	10 : 30	20 : 60
1 : 2	-	10 : 20	15 : 30	20 : 40	40 : 80
1 : 1.5	10 : 15	20 : 30	30 : 45	40 : 60	-
1 : 1.25	20 : 25	40 : 50	60 : 75	-	-

1.5.3 กำหนดค่าเฉลี่ยของประชากร เท่ากับ 10 $(\mu = 10)$

1.5.4 กำหนดค่าสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ เท่ากับ 0.95

1.5.5 การวิจัยครั้งนี้สร้างแบบจำลองข้อมูล โดยใช้เทคนิคมอนติคาร์โล (Monte Caelo Simulation Technique) เขียนโปรแกรมด้วยภาษา FORTRAN 77 ทำการทดลองซ้ำ 2000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 ประโยชน์ของการวิจัย

1. เพื่อเป็นแนวทางในการเลือกวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของสองกลุ่มประชากรและหลายกลุ่มประชากรที่มีการแจกแจงปกติ เมื่อไม่ทราบความแปรปรวน โดยที่ความแปรปรวนไม่เท่ากัน

2. เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของสองกลุ่มประชากรและหลายกลุ่มประชากรที่มีการแจกแจงปกติ ในกรณีที่ความแปรปรวนไม่เท่ากัน โดยใช้ตัวสถิติทดสอบอื่นๆต่อไป