บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การศึกษาวิจัยดำเนินงานต่างๆ โดยทั่วไปจะใช้ระเบียบวิธีทางสถิติในการหาข้อสรุปเกี่ยวกับประชากร ระเบียบวิธีทางสถิติที่สำคัญอย่างหนึ่ง คือ การอนุมานเชิงสถิติ (inference) เช่น การทดสอบสมมติฐาน (hypothesis testing) และการประมาณค่าพารามิเตอร์ (parameter estimation) ซึ่งการทดสอบสมมติฐานจะเป็นการนำค่าสถิติที่ได้จากตัวอย่างมาคำนวณค่าของตัวสถิติที่สอดคล้องกับเรื่องที่สนใจศึกษาเพื่อใช้ในการตัดสินใจว่าควรจะปฏิเสธหรือไม่ปฏิเสธสมมติฐานนั้น ส่วนการประมาณค่าพารามิเตอร์จะเป็นการนำค่าสถิติที่ได้จากตัวอย่างไปประมาณค่าพารามิเตอร์ของประชากร การประมาณค่าพารามิเตอร์จะสามารถแบ่งออกได้เป็นสองประเภท ได้แก่ การประมาณค่าแบบจุด (point estimation) และการประมาณค่าแบบช่วง (interval estimation)

ในการประมาณค่าแบบจุดจะเป็นการประมาณค่าพารามิเตอร์ด้วยค่าๆหนึ่งซึ่งความคลาดเคลื่อนที่เกิดจากการประมาณนี้จะขึ้นอยู่กับการเลือกใช้ตัวประมาณ ส่วนการประมาณค่าแบบช่วงจะเป็นการประมาณค่าพารามิเตอร์โดยอาศัยการประมาณค่าแบบจุดและการแจกแจงของตัวประมาณแบบจุด การประมาณค่าแบบช่วงนี้จะทำให้มั่นใจได้ในระดับหนึ่งว่าค่าพารามิเตอร์ที่สนใจอยู่ในช่วงความเชื่อมั่นที่ประมาณได้ ซึ่งสิ่งที่มีผลกระทบต่อความแคบหรือความกว้างของช่วงการประมาณ คือ สัมประสิทธิ์ความเชื่อมั่น การแจกแจงของตัวสถิติ และขนาดตัวอย่าง ดังนั้นจึงควรเลือกใช้ตัวสถิติที่มีความเหมาะสมกับเรื่องที่สนใจศึกษาและสอดคล้องกับประชากรเพื่อให้ช่วงความเชื่อมั่นที่ได้มีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น การประมาณค่าแบบช่วงที่สำคัญแบบหนึ่ง คือ การประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากร โดยทั่วไปจะใช้ตัวสถิติ Z (Z-statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อทราบส่วนเบี่ยงเบนมาตรฐานของประชากร และตัวสถิติที (t-statistic) ในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรเมื่อไม่ทราบส่วนเบี่ยงเบนมาตรฐานของประชากร โดยจะประมาณค่าส่วนเบี่ยงเบนมาตรฐานของประชากรด้วยส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง ซึ่งการใช้ตัวสถิติดังกล่าวข้างต้นนี้จะอยู่ภายใต้ข้อตกลงเบื้องต้น คือ ประชากรมีการแจกแจงแบบปกติ แต่บางครั้งในทางปฏิบัติประชากรที่สนใจศึกษาอาจไม่มีการแจกแจงเป็นแบบปกติ ซึ่งส่วนใหญ่ข้อมูลต่างๆจะมีการแจกแจงแบบเบ้ขวาและมักจะพบมากกว่าข้อมูลที่มีลักษณะเบ้ซ้าย เช่น ระยะเวลาที่ลูกค้ารอคอยจนกระทั่งได้รับบริการ อายุการใช้งานของเครื่องจักรและอุปกรณ์ ข้อมูล

เกี่ยวกับสิ่งแวดล้อม ได้แก่ ปริมาณสารพิษในน้ำ ปริมาณธาตุต่างๆในดิน เป็นต้น ในกรณีที่ขนาดตัวอย่างมีขนาดใหญ่ก็จะสามารถประมาณได้ว่าข้อมูลที่ได้มีการแจกแจงใกล้เคียงกับการแจกแจงแบบปกติโดยอาศัยทฤษฎีค่าจำกัดสู่ส่วนกลาง (The Central Limit Theorem) แต่ในกรณีที่ขนาดตัวอย่างมีขนาดเล็กซึ่งอาจเป็นเพราะการเก็บข้อมูลทำได้ยาก ดังนั้นการใช้ตัวสถิติทีดังกล่าวในการสร้างช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยอาจจะทำให้ผลสรุปที่ได้มีความผิดพลาด

ในปี ค.ศ. 1928 เนย์แมน (Neyman) และ เพียร์สัน (Pearson) ได้แสดงให้เห็นถึงผลกระทบจากการใช้ตัวสถิติที เมื่อประชากรไม่ได้มีการแจกแจงเป็นแบบปกติและขนาดตัวอย่างมีขนาดเล็ก และเขาพบว่าความเบ้ (skewness) ของประชากรจะมีผลต่อการแจกแจงของตัวสถิติทีมากกว่าความโด่ง (kurtosis) และประชากรที่มีลักษณะเบ้ขวา (positively skewed distribution) มีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ซ้าย (negatively skewed distribution) ในทางกลับกัน ประชากรที่มีลักษณะเบ้ซ้ายจะมีผลกระทบต่อการแจกแจงของตัวสถิติทีในทางเบ้ขวา

ในปี ค.ศ. 1978 จอห์นสัน (Johnson) ได้เสนอตัวสถิติ (modified t-tests) ซึ่งได้ทำการแปลงจากตัวสถิติทีเพื่อให้มีความเหมาะสมสำหรับการประมาณค่าเฉลี่ยของการแจกแจงที่ไม่สมมาตรโดยใช้ Cornish-Fisher Expansion แต่วิธีการของจอห์นสันนี้ยังคงไม่เหมาะสมในกรณีที่ขนาดตัวอย่างมีขนาดเล็กและประชากรมีความเบ้มาก ต่อมาในปี ค.ศ. 1992 ฮอลล์ (Hall) ได้เสนอว่าวิธีการที่ปรับปรุงแก้ไขเพิ่มความถูกต้องจากวิธีการของจอห์นสันโดยทำการแปลงตัวสถิติทีด้วย Edgeworth Expansion เขาพบว่าช่วงความเชื่อมั่นด้านล่างที่ได้ให้ความแม่นยำสูงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงไม่สมมาตรต่างๆ แต่ช่วงความเชื่อมั่นด้านบนยังให้ความแม่นยำไม่ค่อยดีนักสำหรับตัวอย่างขนาดเล็ก

ในปี ค.ศ. 1995 เชน (Chen) ได้เสนอตัวสถิติทดสอบใหม่สำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาด้วยกลุ่มตัวอย่างขนาดเล็กซึ่งทำการแปลงจากตัวสถิติทีโดยใช้ Edgeworth Expansion และ Taylor Expansion นอกจากนั้น ในปี ค.ศ. 2000 ซู (Zhou) และ เกาว์ (Gao) ได้แสดงให้เห็นว่าการนำวิธีการบูตสแตรป (Bootstrap method) มาช่วยในการหาค่าของตัวสถิติที่แปลงจากตัวสถิติทีทั้งของจอห์นสันและฮอลล์จะทำให้ผลที่ได้มีความแม่นยำมากยิ่งขึ้น

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงสนใจที่จะทำการศึกษาเพื่อเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาที่ระดับความเบ้ต่างกันในการแจกแจงหนึ่งๆจำแนกตามขนาดตัวอย่างและสัมประสิทธิ์ความเชื่อมั่น โดยพิจารณาวิธีการประมาณค่าแบบช่วง 4 วิธี คือ

1. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที (T)

2. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสัน (J)
3. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์ (H)
4. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน (C)

ผลการวิจัยจะทำให้ทราบว่าในสถานการณ์ที่แตกต่างกัน วิธีการประมาณค่าแบบช่วงวิธีการใดให้ช่วงความเชื่อมั่นที่เหมาะสม

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากร โดยมีวิธีการประมาณค่าแบบช่วง 4 วิธี คือ

1. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติที
2. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสัน
3. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของฮอลล์
4. วิธีการประมาณค่าแบบช่วงด้วยตัวสถิติของเชน

ภายใต้ประชากรที่มีการแจกแจงแบบเบ้ขวา ได้แก่

1. การแจกแจงไคกำลังสอง (Chi-Square Distribution)
2. การแจกแจงลอกนอร์มอล (Log-normal Distribution)
3. การแจกแจงแกมมา (Gamma Distribution)
4. การแจกแจงไวบูลล์ (Weibull Distribution)

1.3 สมมติฐานของการวิจัย

ในการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงเบ้ขวา การประมาณค่าแบบช่วงด้วยตัวสถิติของเชนให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงกว่าวิธีการประมาณอื่น และให้ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่ามีค่าสูงกว่าวิธีการประมาณอื่น การประมาณค่าแบบช่วงด้วยตัวสถิติของจอห์นสันให้ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนสำหรับการทดสอบสมมติฐานทางเดียวด้านมากกว่าและค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นสำหรับการทดสอบสมมติฐานสองทางมีค่าต่ำกว่าวิธีการประมาณอื่น

1.4 ขอบเขตของการวิจัย

ในการวิจัยนี้จะศึกษาเปรียบเทียบวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวา โดยที่

1.4.1 การแจกแจงของประชากรที่ใช้ในการศึกษานี้ 4 การแจกแจง ได้แก่

1.4.1.1 การแจกแจงไคกำลังสอง (Chi-Square Distribution)

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงไคกำลังสอง ด้วยระดับขั้นความเสรี (ร.ส.) n ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{2^{-n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-x/2} \qquad , x > 0$$

เมื่อสัมประสิทธิ์ความเบ้เท่ากับ $\sqrt{\frac{8}{n}}$

และสัมประสิทธิ์ความโด่งเท่ากับ $3 + \frac{12}{n}$

1.4.1.2 การแจกแจงลอกนอร์มอล (Log-normal Distribution)

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงลอกนอร์มอล ด้วยพารามิเตอร์ μ และ σ^2 ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right) \qquad , x > 0 \quad , \sigma^2 > 0$$

เมื่อ μ และ σ^2 เป็นค่าเฉลี่ยและความแปรปรวนของ Y โดยที่ Y = ln X

Y มีการแจกแจงแบบปกติ

และสัมประสิทธิ์ความเบ้เท่ากับ $(\omega + 2)\sqrt{\varpi - 1}$

ส่วนสัมประสิทธิ์ความโด่งเท่ากับ $\omega^4 + 2\omega^3 + 3\omega^2 - 3$ โดยที่ $\omega = \exp(\sigma^2)$

1.4.1.3 การแจกแจงแกมมา (Gamma Distribution)

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงแกมมา ด้วยพารามิเตอร์ α และ β ดังนั้นฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta} \right) \qquad , x > 0 \quad , \alpha > 0 \quad , \beta > 0$$

เมื่อสัมประสิทธิ์ความเบ้เท่ากับ $\frac{2}{\sqrt{\alpha}}$

และสัมประสิทธิ์ความโด่งเท่ากับ $3 + \frac{6}{\alpha}$

1.4.1.4 การแจกแจงไวบูลล์ (Weibull Distribution)

ให้ X เป็นตัวแปรสุ่มต่อเนื่องที่มีการแจกแจงไวบูลล์ ด้วยพารามิเตอร์ α และ β ฟังก์ชันความหนาแน่นของ X จะอยู่ในรูปของ

$$f(x)=\frac{1}{\beta^{\alpha}}\alpha x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)^{\alpha} \qquad ,x>0 \quad ,\alpha>0 \quad ,\beta>0$$

เมื่อ สัมประสิทธิ์ความเบ้เท่ากับ $\dfrac{c-3ab+2a^3}{\left(b-a^2\right)^{3/2}}$

และ สัมประสิทธิ์ความโด่งเท่ากับ $\dfrac{d+6a^2b-3a^4-4ac}{\left(b-a^2\right)^2}$

โดยที่ $a=\Gamma\left(1+\frac{1}{\alpha}\right)$, $b=\Gamma\left(1+\frac{2}{\alpha}\right)$, $c=\Gamma\left(1+\frac{3}{\alpha}\right)$, $d=\Gamma\left(1+\frac{4}{\alpha}\right)$

ดังนั้นพารามิเตอร์และสัมประสิทธิ์ความโด่งของการแจกแจงต่างๆจะกำหนดจากสัมประสิทธิ์ความเบ้ดังนี้

ตารางที่ 1.4 แสดงการแจกแจงที่ใช้ในการวิจัย ค่าพารามิเตอร์ ค่าสัมประสิทธิ์ความเบ้และค่าสัมประสิทธิ์ความโด่ง

สัมประสิทธิ์ความเบ้	การแจกแจง	พารามิเตอร์	สัมประสิทธิ์ความโด่ง
0.5	การแจกแจงไคกำลังสอง	N=32	3.375
	การแจกแจงลอกนอร์มอล	$\mu=0$, $\sigma=0.1641$	3.448
	การแจกแจงแกมมา	$\beta=1$, $\alpha=16$	3.375
	การแจกแจงไวบูลล์	$\beta=1$, $\alpha=2.2156$	3.028
1.0	การแจกแจงไคกำลังสอง	N=8	4.500
	การแจกแจงลอกนอร์มอล	$\mu=0$, $\sigma=0.3143$	4.830
	การแจกแจงแกมมา	$\beta=1$, $\alpha=4$	4.500
	การแจกแจงไวบูลล์	$\beta=1$, $\alpha=1.5639$	4.159
1.5	การแจกแจงลอกนอร์มอล	$\mu=0$, $\sigma=0.4435$	7.251
	การแจกแจงแกมมา	$\beta=1$, $\alpha=1.7778$	6.375
	การแจกแจงไวบูลล์	$\beta=1$, $\alpha=1.2111$	6.133
2.0	การแจกแจงไคกำลังสอง	N=2	9.000
	การแจกแจงลอกนอร์มอล	$\mu=0$, $\sigma=0.5514$	10.864
	การแจกแจงแกมมา	$\beta=1$, $\alpha=1$	9.000
2.5	การแจกแจงลอกนอร์มอล	$\mu=0$, $\sigma=0.6409$	15.850
	การแจกแจงแกมมา	$\beta=1$, $\alpha=0.64$	12.375
	การแจกแจงไวบูลล์	$\beta=1$, $\alpha=0.8631$	12.823

ตารางที่ 1.4 (ต่อ) แสดงการแจกแจงที่ใช้ในการวิจัย ค่าพารามิเตอร์ ค่าสัมประสิทธิ์ความเบ้และค่าสัมประสิทธิ์ความโด่ง

สัมประสิทธิ์ความเบ้	การแจกแจง	พารามิเตอร์	สัมประสิทธิ์ความโด่ง
3.0	การแจกแจงลอกนอร์มอล	$\mu = 0$, $\sigma = 0.7156$	22.404
	การแจกแจงแกมมา	$\beta = 1$, $\alpha = 0.4444$	16.501
	การแจกแจงไวบูลล์	$\beta = 1$, $\alpha = 0.7686$	17.656
5.0	การแจกแจงลอกนอร์มอล	$\mu = 0$, $\sigma = 0.9202$	68.262
	การแจกแจงแกมมา	$\beta = 1$, $\alpha = 0.16$	40.500
	การแจกแจงไวบูลล์	$\beta = 1$, $\alpha = 0.5737$	48.253

1.4.2 กำหนดขนาดตัวอย่างเท่ากับ 10 20 30 และ 50

1.4.3 กำหนดค่าสัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ เท่ากับ 0.90 0.95 และ 0.99

1.4.4 ในการวิจัยนี้จะศึกษาการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาโดยสมมติฐานทางเดียวของการทดสอบค่าเฉลี่ยมีรูปแบบดังนี้

ก) $H_0 : \mu = \mu_0$

เทียบกับ $H_1 : \mu < \mu_0$

หรือ ข) $H_0^* : \mu = \mu_0$

เทียบกับ $H_1^* : \mu > \mu_0$

และสมมติฐานสองทางของการทดสอบค่าเฉลี่ยมีรูปแบบดังนี้

ค) $H_0^{**} : \mu = \mu_0$

เทียบกับ $H_1^{**} : \mu \neq \mu_0$

1.4.5 การวิจัยครั้งนี้จะสร้างแบบจำลองข้อมูลตามสถานการณ์ต่างๆ โดยใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) ทำการทดลองซ้ำ 3,000 ครั้ง และใช้ขนาดตัวอย่างสำหรับวิธีบูตสแตรป (Bootstrap method) เท่ากับ 2,000 ด้วยโปรแกรมภาษาฟอร์แทรน (Fortran)

1.5 ข้อตกลงเบื้องต้น

ในการหาช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของประชากรจะพิจารณาช่วงความเชื่อมั่นล่างสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่า ช่วงความเชื่อมั่นบนสำหรับการทดสอบ

สมมติฐานทางเดียวด้านมากกว่าและช่วงความเชื่อมั่นสำหรับการทดสอบสมมติฐานสองทางภายใต้สัมประสิทธิ์ความเชื่อมั่น $(1-\alpha)$ 100% ที่กำหนด

1.6 คำจำกัดความที่ใช้ในการวิจัย

สัมประสิทธิ์ความเชื่อมั่น (confidence coefficient) หมายถึง ความน่าจะเป็นที่ช่วงสุ่มจะครอบคลุมค่าจริงของพารามิเตอร์ของประชากร

ช่วงความเชื่อมั่น (confidence interval) หมายถึง ช่วงค่าประมาณพารามิเตอร์ด้วยค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการศึกษาและเลือกวิธีการประมาณค่าแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาในสถานการณ์ต่างๆได้อย่างเหมาะสม
2. เพื่อเป็นแนวทางในการศึกษาวิธีการประมาณแบบช่วงสำหรับค่าเฉลี่ยของประชากรที่มีการแจกแจงแบบเบ้ขวาโดยใช้ตัวสถิติทดสอบอื่นๆต่อไป

1.8 วิธีดำเนินการวิจัย

1.8.1 สร้างข้อมูลให้มีลักษณะตามที่กำหนดในแผนการทดลอง

1.8.2 คำนวณช่วงความเชื่อมั่นสำหรับค่าเฉลี่ยของแต่ละวิธี

1.8.3 ตรวจสอบว่าช่วงความเชื่อมั่นที่ได้จากการคำนวณของแต่ละวิธีให้ค่าสัมประสิทธิ์ความเชื่อมั่นเท่ากับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดหรือไม่

1.8.4 เปรียบเทียบค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นล่างสำหรับการทดสอบสมมติฐานทางเดียวด้านน้อยกว่า ค่าเฉลี่ยของขีดจำกัดความเชื่อมั่นบนสำหรับการทดสอบสมมติฐานทางเดียวด้านมากกว่า และค่าเฉลี่ยความยาวของช่วงความเชื่อมั่นสำหรับการทดสอบสมมติฐานสองทาง โดยจะพิจารณาเฉพาะวิธีที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นเท่ากับค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดเท่านั้น

1.8.6 สรุปผลการวิจัย