# บทที่ 1

# บทนำ

## 1.1 ความเป็นมาและความสำคัญของปัญหา

ประชากร (Population) ในทางสถิติ หมายถึง เซตที่รวมข้อมูลในเรื่องที่ต้องการจะศึกษาเป็นข้อมูลที่สมบูรณ์ของเรื่องที่สนใจ ตัวอย่างเช่น ในการสำรวจความคิดเห็นของนิสิตจุฬาฯ ที่มีต่อห้องสมุดของมหาวิทยาลัย ความคิดเห็นของนิสิตแต่ละคนเป็นข้อมูลที่จะศึกษา เมื่อรวมเข้าด้วยกันก็จะถือเป็นข้อมูลประชากร ดังนั้นประชากรในที่นี้ ก็คือนิสิตที่แสดงความคิดเห็นนั่นเอง โดยประชากรในทางสถิติแบ่งออกเป็นประชากรอนันต์ (Infinite Population) คือ ประชากรที่ไม่สามารถนับจำนวนหน่วยประชากร (Population Unit) ทั้งหมดได้จริง และประชากรอันตะ (Finite Population) คือ ประชากรที่มีขนาดจำกัดสามารถนับจำนวนหน่วยประชากรได้ เนื่องจากมีการกำหนดคาบเวลาและขอบเขตการทำงานทำให้ทราบลักษณะและภาวะต่าง ๆ ที่เกิดขึ้น ในทางทฤษฎีการสุ่มตัวอย่าง (Sampling Theory) จะพิจารณาเฉพาะประชากรที่เป็นอันตะ เนื่องจากการสุ่มตัวอย่างจะทำในระยะเวลาหนึ่งที่กำหนดเท่านั้น

ทฤษฎีการสุ่มตัวอย่างเป็นการพิจารณาค่าประชากรมี 4 วิธี คือ ค่าเฉลี่ยประชากร (Population Mean) ยอดรวมประชากร (Population Total) สัดส่วนประชากร (Population Proportion) และ อัตราส่วนประชากร (Population Ratio)

สำหรับการเก็บรวบรวมข้อมูล หากสนใจที่จะเก็บข้อมูลจากทุกหน่วยหรือที่เรียกว่า การทำสำมะโน (Census) จะทำให้ได้ข้อมูลที่สมบูรณ์ครบถ้วน แต่ต้องเสียค่าใช้จ่าย เวลา และกำลังคนมาก ส่วนการสำรวจตัวอย่าง (Sample Survey) เป็นการสุ่มเก็บข้อมูลเพียงบางหน่วย แม้ข้อมูลที่ได้จะเป็นข้อมูลเพียงบางส่วนไม่สมบูรณ์ แต่จะทำให้เสียค่าใช้จ่าย เวลา และกำลังคนที่ต่ำกว่าสำมะโน รวมทั้งเกิดความคลาดเคลื่อนน้อยกว่า ดังนั้นจึงต้องอาศัยวิธีการทางสถิติในการสุ่มตัวอย่างขึ้นมาเป็นตัวแทนของประชากร เพราะในความเป็นจริงไม่สามารถเก็บข้อมูลจากทุกหน่วยได้

ในการคำนวณหาค่าลักษณะของประชากร โดยปกติมักเป็นค่าที่ไม่ทราบและไม่สามารถคำนวณได้โดยตรง เนื่องจากต้องเสียค่าใช้จ่าย เวลา และกำลังคนที่สูงเพื่อให้ได้ค่าจริงที่ต้องการ เนื่องจากประชากรมีขนาดใหญ่มากทำให้จำเป็นต้องมีตัวสถิติ เป็นตัวแสดงลักษณะตัวอย่างที่ถูกสุ่มมาจากประชากร ทั้งนี้ตัวสถิติที่จะเป็นตัวแทนประชากรได้จะต้องเป็นตัวประมาณที่ไม่เอนเอียง ส่วนการประมาณค่าประชากร เป็นวิธีการทางสถิติที่ใช้ข้อมูลตัวอย่างไปประมาณค่าประชากร เรียกว่า ตัว

ประมาณ (Estimator) โดยแบ่งได้ 2 แบบ คือ การประมาณค่าแบบจุด (Point Estimation) และ การประมาณค่าแบบช่วง (Interval Estimation)

ปัจจุบันการสุ่มตัวอย่างมีบทบาทมากในงานวิจัยและงานสำรวจ ทั้งด้านการแพทย์ การศึกษา ธุรกิจ บันเทิง การเมือง ฯลฯ วิธีการสุ่มตัวอย่างโดยอาศัยหลักความน่าจะเป็นแบ่งเป็น 4 วิธี คือ วิธีการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) วิธีการสุ่มตัวอย่างแบบมีชั้นภูมิ (Stratified Sampling) วิธีการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) และวิธีการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) ถ้าใช้วิธีการสุ่มตัวอย่างที่ไม่เหมาะสมกับประชากรจะทำให้ผลที่ได้คลาดเคลื่อนนำไปสู่การสรุปที่ผิดพลาดได้ รวมทั้งเป็นการเพิ่มค่าใช้จ่ายและสูญเสียเวลา

แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับ (Adaptive Cluster Sampling) ถูกพัฒนามาจากการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) เป็นแบบแผนที่เหมาะสำหรับการสำรวจสิ่งที่หายากมักจะอยู่บริเวณที่เกาะกลุ่มหรือใกล้เคียงกัน เช่น การประมาณจำนวนสัตว์ที่อยู่อาศัยในพื้นที่หนึ่ง พืชตระกูลต่าง ๆ รวมทั้งแร่หรือเชื้อเพลิงที่เป็นซากพืชและสัตว์ หากพบตัวอย่างที่ต้องการหรือพบสัตว์ประเภทนี้อยู่ในตัวอย่าง จึงสุ่มหน่วยตัวอย่างในบริเวณใกล้เคียงเพิ่มเติม แต่ถ้าบริเวณไหนไม่พบตัวอย่างที่สนใจก็ไม่ต้องสุ่มหน่วยตัวอย่างนั้นเพิ่ม

เห็นได้ว่าแผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับเป็นแผนแบบการสุ่มตัวอย่างที่น่าสนใจและเป็นประโยชน์อย่างยิ่งหากนำไปใช้ให้เหมาะสม โดยปัญหาพบว่าตัวประมาณค่าเฉลี่ยที่ได้ แม้จะไม่เอนเอียงสำหรับค่าเฉลี่ยประชากรเมื่อใช้แผนแบบการสุ่มตัวอย่างแบบที่ยังไม่ได้ปรับ แต่ถ้าใช้กับแผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับจะเกิดความเอนเอียง โดยที่ Thompson (1990) ได้เสนอตัวประมาณค่าเฉลี่ยที่ไม่เอนเอียง และได้สร้างตัวประมาณที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson ดังนั้นเพื่อให้การนำแผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับไปใช้งานได้จริง จึงจำเป็นที่ต้องศึกษาแนวคิดให้เข้าใจอย่างถ่องแท้เพื่อเป็นประโยชน์ในการใช้งานและถูกต้องตามหลักทฤษฎี

## 1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาแนวคิดแผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับ (Adaptive Cluster Sampling)
2. เพื่อศึกษาแนวคิดจากการสุ่มตัวอย่างแบบง่ายโดยไม่ใส่คืน ภายใต้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับ
3. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเฉลี่ยที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่ายโดยไม่ใส่คืนภายใต้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับ และแผนแบบการสุ่มตัวอย่างแบบที่ไม่ได้ปรับ (Nonadaptive)

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

## 1.3 ขอบเขตของการวิจัย

ศึกษาประสิทธิภาพของตัวประมาณค่าเฉลี่ยที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่ายโดยไม่ใส่คืน ภายใต้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับของ Steven K.Thompson (1990) สำหรับตัวประมาณค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz -Thompson และแผนแบบการสุ่มตัวอย่างแบบที่ไม่ได้ปรับ (Nonadaptive) แบ่งพื้นที่ศึกษาออกเป็น 100 หน่วย ประชากรที่ใช้สร้างมาจากกระบวนการพัวซองคลัสเตอร์ ซึ่งตำแหน่งและจำนวนของจุดหลักสร้างจากกระบวนการพัวซองที่มีพารามิเตอร์ $\lambda$ เป็น 20 โดยจำลองประชากรเป็น 3 กรณี และตำแหน่งของบริวารต่าง ๆ สร้างจากตัวแปรสุ่มที่มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal) ตัวอย่างขั้นต้นที่ทำการสุ่มประกอบไปด้วยขนาด 4 หน่วย 8 หน่วย 16 หน่วย และ 32 หน่วย ซึ่งแต่ละหน่วยตัวอย่างแบ่งเป็น 100 รูปแบบ

## 1.4 คำจำกัดความที่ใช้ในงานวิจัย

แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับ ( Adaptive Cluster Sampling) หมายถึง แผนแบบการสุ่มตัวอย่างแบบที่ปรับกระบวนการสุ่มตัวอย่างให้สามารถนำลักษณะของข้อมูลที่เก็บได้มากำหนดหน่วยตัวอย่าง เช่น กำหนดพื้นที่ในการสำรวจสัตว์ หากพื้นที่ใดพบสัตว์ที่เป็นไปตามเกณฑ์ก็จะเพิ่มพื้นที่บริเวณใกล้เคียงเป็นตัวอย่างเพิ่มเติม แต่ถ้าพื้นที่ใดไม่พบสัตว์ที่เป็นไปตามเกณฑ์พื้นที่นั้นก็ไม่ต้องสนใจ

แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับที่ทำการสุ่มตัวอย่างขั้นต้นอย่างง่าย หมายถึง การนำแผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับมาทำการสุ่มตัวอย่าง โดยการสุ่มตัวอย่างเบื้องต้นจะอาศัยหลักการสุ่มตัวอย่างแบบง่าย

การสุ่มตัวแปรสุ่มอย่างง่าย หมายถึง วิธีการสุ่มตัวอย่าง n จากประชากรขนาด N ทุก ๆ ${}^{N}c_{n}$ ตัวอย่างที่แตกต่างกันมีโอกาสถูกสุ่มเท่า ๆ กัน คือ 1 / ${}^{N}c_{n}$

ความไม่เอนเอียงของตัวประมาณ หมายถึง การที่ตัวประมาณนั้นมีค่าคาดหวัง (Expected Value) หรือค่าเฉลี่ยของตัวประมาณมีค่าเท่ากับค่าพารามิเตอร์หรือค่าจริง กล่าวคือ $\hat{Y}$ จะเป็นตัวประมาณไม่เอนเอียงของ Y ถ้าหากว่า

$$E(\hat{Y}) = Y \text{ หรือ } E(\hat{Y}) - Y = 0$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การสุ่มตัวอย่างแบบไม่ใส่คืน (Sampling Without Replacement) หมายถึง ตัวอย่างที่สุ่มไปแล้วไม่สามารถสุ่มได้อีก จะทำให้ขนาดประชากรลดลงไป แตกต่างจากการสุ่มตัวอย่างแบบใส่คืน (Sampling With Replacement) โดยที่ขนาดประชากรจะเท่าเดิม ทำให้ตัวอย่างที่สุ่มแล้วมีโอกาสถูกสุ่มใหม่ โดยปกติมีโอกาสสุ่มตัวอย่างเดิมซ้ำน้อยมาก ดังนั้นในรายงานวิจัยเล่มนี้จึงศึกษาแต่การสุ่มตัวอย่างแบบไม่ใส่คืน

การเปรียบเทียบประสิทธิภาพของตัวประมาณ หมายถึง ตัวประมาณค่าพารามิเตอร์ที่สนใจมีความคลาดเคลื่อนต่ำ โดยวัดได้จากความแปรปรวนของตัวประมาณ หากต้องการเปรียบเทียบประสิทธิภาพของตัวประมาณของ 2 วิธี จะทำได้โดยดูว่าความแปรปรวนของวิธีใดต่ำกว่าก็ถือว่าวิธีนั้นมีประสิทธิภาพดีกว่า

แผนแบบการสุ่มตัวอย่างแบบที่ไม่ได้ปรับ (Nonadaptive) หมายถึง แผนแบบการสุ่มตัวอย่างแบบทั่วไปหรือดั้งเดิม ในที่นี้ทำการศึกษาการสุ่มตัวอย่างโดยอาศัยหลักความน่าจะเป็น โดยวิธีการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling)

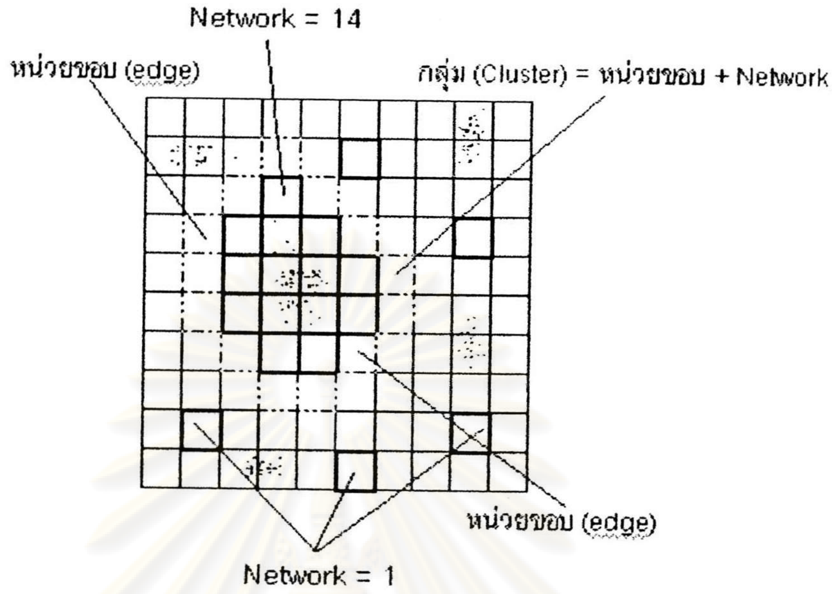
ตัวอย่างขั้นต้น หมายถึง การสุ่มตัวอย่างโดยอาศัยหลักความน่าจะเป็นแบ่งเป็น 4 วิธี คือ วิธีการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) วิธีการสุ่มตัวอย่างแบบมีชั้นภูมิ (Stratified Sampling) วิธีการสุ่มตัวอย่างแบบมีระบบ (Systematic Sampling) และวิธีการสุ่มตัวอย่างแบบกลุ่ม (Cluster Sampling) เช่น ใช้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อตัวอย่างขั้นต้นสุ่มแบบง่าย หรือ ใช้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับเมื่อตัวอย่างขั้นต้นสุ่มแบบมีชั้นภูมิ

หน่วยขอบ (Edge Unit) หมายถึง เป็นตัวบอกขอบเขต หน่วยที่อยู่ภายในหน่วยขอบทั้งหมดจะแสดงถึง Network โดยหน่วยขอบนี้จะไม่ใช้ในการคำนวณ

Network หมายถึง หน่วยย่อยที่เป็นไปตามเงื่อนไขที่ต้องการ แต่ถ้าหน่วยใดไม่เป็นไปตามเงื่อนไขก็ถือว่าให้ Network นั้นมีขนาดเป็น 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กลุ่ม (Cluster) หมายถึง ขอบเขตทั้งหมด กล่าวง่าย ๆ ก็คือ หน่วยขอบ + Network นั่นเอง




**ภาพที่ 1** : ภาพแสดงหน่วยขอบ (Edge Unit) , Network และกลุ่ม (Cluster) เมื่อสุ่มตัวอย่างแบบง่ายโดยมีขนาด n = 6

การแจกแจงแบบพัวซอง (Poisson Distribution) ใช้อธิบายข้อมูลที่เป็นการนับจำนวนครั้งของเหตุการณ์ที่สนใจในช่วงเวลาหรือพื้นที่ที่กำหนด เช่น การนับจำนวนคนไข้ที่เข้ามารักษาในโรงพยาบาลช่วงเวลา 1 ชั่วโมง หรือ ใช้อธิบายข้อมูลที่เป็นการนับจำนวนครั้งของเหตุการณ์ที่สนใจในพื้นที่ที่กำหนด เช่น นับจำนวนสัตว์หรือพืชในพื้นที่ 1 ตารางกิโลเมตร

การแจกแจงแบบเลขชี้กำลัง (Exponential Distribution) ใช้อธิบายช่วงเวลาห่างระหว่างการเกิดเหตุการณ์ เช่น ช่วงเวลาห่างที่ลูกค้าเข้ามารับบริการ หรืออายุการใช้งานของหลอดไฟ

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเกิดความเข้าใจที่ชัดเจนในทฤษฎีการสุ่มตัวอย่างแบบกลุ่มปรับ (Adaptive Cluster Sampling)

2. เพื่อเกิดความเข้าใจที่ชัดเจนของตัวอย่างขั้นต้นจากการสุ่มตัวอย่างแบบง่ายโดยไม่ใส่คืนภายใต้แผนแบบการสุ่มตัวอย่างแบบกลุ่มปรับของ Steven K. Thompson (1990) สำหรับตัวประมาณ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ค่าเฉลี่ยที่ดัดแปลงมาจากตัวประมาณ Horvitz-Thompson และทำให้สามารถนำหลักการทางทฤษฎีไปใช้ในทางปฏิบัติได้ถูกต้อง

3. เพื่อเป็นแนวทางในการศึกษาทฤษฎีทางสถิติอื่นอีกต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย