8. การสุ่มตัวอย่าง และ การประมาณค่า

8.1 ประชากร (Populations)

- ประชากรในทางสถิติ หมายถึง ตัวเลขหรือหน่วยของข้อมูล (Data) ที่จะใช้ในการศึกษาอาจเป็นตัวบุคคล ครอบครัว หรือกลุ่มต่าง ๆของสังคม ปริมาณสิ่งของต่าง ๆ ปริมาณน้ำฝน ปริมาณข้าวที่ผลิตได้ในแต่ละไร่ น้ำหนัก ส่วนสูง เราต้องกำหนดขอบเขตของการศึกษาว่าจะครอบคลุมแค่ไหน เราเรียกทุก ๆหน่วยที่อยู่ในขอบเขตนั้นว่า ประชากร (populations)

8.2 ประชากร (Samples)

- ในการดำเนินงานเก็บข้อมูลทางสถิติ_มักจะพบบ่อยๆ ว่าการเก็บข้อมูลจากทุกๆหน่วยในประชากรนั้นทำได้ยาก เสียค่าใช้จ่ายมาก และไม่จำเป็นต้องทำถึงขนาดนั้น ในทางสถิติเรานิยมใช้ตัวอย่างเป็นตัวแทนของประชากร ซึ่งตัวอย่างที่ใช้จะต้องเป็นตัวแทนที่ดีของประชากรได้เพียงไรนั้น

8.3 ค่าพารามิเตอร์และค่าสถิติ (Parameter and Statistics)

ค่าพารามิเตอร์ในทางสถิติ หมายถึง ค่าอันที่จริง (True Value) ที่หาได้โดยวิธีทางสถิติเพื่ออธิบายถึงลักษณะต่าง ๆของประชากร เป็นค่าคงที่ เช่น มัชฌิมเลขคณิต (Arithmetic mean) สัดส่วน (Proportion) หรือค่าความแปรปรวน (Variance) ของประชากร

- ลักษณะของข้อมูลที่ดี ที่จะใช้เป็นตัวแทนของประชากรทั้งหมดต้องไม่มีความเอนเอียง (Unbiased) และให้ความคลาดเคลื่อน (error) น้อยที่สุด
- **<u>สาเหตุที่ทำให้เกิดความเอนเอียง (Biased) มี 2 ประการใหญ่ ๆ คือ</u>**
- 1. เกิดจากการใช้สูตรในการประมาณค่าไม่ถูกต้อง คือ หาค่า parameters เพื่อคำนวณค่าสถิติไม่ถูกต้อง
- 2. เกิดจากการเลือกตัวอย่าง ตัวอย่างที่เลือกมานั้นไม่ได้เป็นตัวแทนที่ดีของประชากร

- **ลักษณะตัวอย่างที่ดี** ตัวอย่างที่ดีควรได้มาจากการสุ่ม ตามทฤษฎีความน่าจะเป็น กล่าวคือ ทุกๆหน่วยในประชากรมีโอกาสที่จะได้รับเลือกมาเป็นตัวอย่าง เรียกว่า probability sample กล่าวคือ ตัวอย่างที่เลือกไม่ได้เฉพาะเจาะจง คือ ไม่มีความเอนเอียงในการเลือก
- **ข้อดีของการสุ่มตัวอย่าง**
- ใช้ดีในกรณีที่ประชากรมีขนาดใหญ่เกินไป ไม่สามารถตรวจสอบได้ทั้งหมด
- เป็นการสิ้นเปลืองค่าใช้จ่ายมากเกินไปหากต้องเก็บข้อมูลจากทุกหน่วยในประชากร
- เป็นการประหยัดค่าใช้จ่ายโดยไม่ต้องเสียเวลาศึกษาประชากรโดยตรง
- หากเลือกตัวอย่างได้ดีแล้ว การศึกษาตัวอย่างจะให้ผลที่เที่ยงตรงมากกว่าการศึกษาจากประชากรโดยตรง

- **8.4 วิธีการสุ่มตัวอย่าง (Sampling Methodology)** ที่ใช้กันแพร่หลายมี 5 วิธี
- การสุ่มตัวอย่างอย่างง่าย (Sample Random Sampling)
- การสุ่มตัวอย่าง แบบขั้นภูมิ (Stratified Random Sampling)
- การสุ่มตัวอย่าง แบบมีระบบ (Systematic Random Sampling)
- การสุ่มตัวอย่าง แบบแบ่งเป็นกลุ่ม (Cluster Random Sampling)
- การสุ่มตัวอย่างหลายขั้น (Multi-Stage Sampling)

8.5 การสุ่มตัวอย่างแบบแทนที่ และไม่แทนที่ (Sampling with and without replacement)

- การสุ่มตัวอย่างแบบแทนที่ หมายความว่า
- การสุ่มตัวอย่างแบบไม่แทนที่ หมายความว่า
- ผลที่ได้จากแต่ละหน่วยตัวอย่างจะเป็นอิสระต่อกัน (Statistical independent) เมื่อเป็นการสุ่มตัวอย่างแบบแทนที่ แต่ผลที่ได้จากแต่ละหน่วย ตัวอย่างจะขึ้นต่อกัน (Statistical dependent) เมื่อเป็นการสุ่มตัวอย่างแบบไม่แทนที่ คือความน่าจะเป็นที่หน่วยตัวอย่างหน่วยใด จะถูกสุ่มในครั้งต่อไป จะไปอยู่กับผลของการสุ่มครั้งก่อนหน้า

- ในกรณีที่ประชากรมีขนาดใหญ่มากเมื่อเทียบกับขนาดของตัวอย่าง ไม่ว่าจะเป็นสุ่มตัวอย่างแบบแทนที่หรือไม่แทนที่ ความแตกต่างของผลที่ได้จะมีไม่มากนัก
- โดยทั่วไป ความแตกต่างระหว่างการสุ่มตัวอย่างแบบแทนที่กับการสุ่มตัวอย่างแบบไม่แทนที่จะน้อยมาก ถ้าขนาดของประชากรใหญ่กว่าของตัวอย่างไม่ต่ำกว่า 10 เท่า

8.6 การประมาณค่า (ESTIMATION)

การประมาณค่าต่าง ๆในประชากร (Points and interval estimators)

ความแตกต่างระหว่างค่าที่แท้จริงในประชากร (parameters) และค่าที่ประมาณได้จากตัวอย่าง (statistics) ยกตัวอย่างเช่น ค่าอาหารกลางวันของนิสิตในมหาวิทยาลัยแห่งหนึ่ง จำนวน 5,000 คน หากนำข้อมูลตัวเลขของนิสิตทั้งหมดรวมกันแล้วหารด้วย 5,000 จะได้ค่าเฉลี่ยประชากร (Population mean)

- โดยทั่วไปในทางสถิติเมื่อหาค่าเฉลี่ยได้แล้ว เราก็ต้องการทราบว่า โดยเฉลี่ยแล้วค่าเหล่านั้นจะแตกต่างไปจากค่าเฉลี่ยที่หาได้มากน้อยเพียงใด ซึ่งตัวที่ใช้วัดการกระจายของค่าต่างๆ แต่ละค่าจากค่าเฉลี่ยของประชากร เรียกว่า ส่วนเบี่ยงเบนมาตรฐานของประชากร (Population Standard Deviation)
- นิสิตจำนวน 100 คน ถูกเลือกมาเรียกว่า ขนาดของตัวอย่าง (Size of Sample) จากนั้นก็สามารถหาค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานได้เช่นเดียวกับประชากร ซึ่งค่าที่ได้มักเรียกว่า ค่าเฉลี่ยจากตัวอย่าง และส่วนเบี่ยงเบนมาตาฐานจากตัวอย่าง หรือที่เรียกว่า ค่าสถิติ

สำหรับประชากร

μ แทน ค่าเฉลี่ย หรือมัชฌิมเลขคณิต

σ แทน ส่วนเบี่ยงเบนมาตรฐาน

N แทน ขนาดของประชากร

π แทน สัดส่วน

σ_p แทน ความคลาดเคลื่อนมาตรฐานของสัดส่วน

สำหรับตัวอย่าง

$\bar{x}$ แทน ค่าเฉลี่ย หรือมัชฌิมเลขคณิต

s แทน ส่วนเบี่ยงเบนมาตรฐาน

n แทน ขนาดของประชากร

p แทน สัดส่วน

s_p แทน ความคลาดเคลื่อนมาตรฐานของสัดส่วน

ถ้าสมมติว่า ตัวอย่างจำนวน n และ $x_1, x_2, \ldots, x_n$ คือค่าที่วัดได้จาก n หน่วย นั้นถูกเลือกมา โดยสุ่มจากประชากรซึ่งมีการแจกแจงแบบปกติ มีมัชฌิมเลขคณิตเป็น μ และส่วนเบี่ยงเบนมาตรฐาน σ ซึ่งยังไม่ทราบค่า โดยทั่วไปมักใช้ค่าเฉลี่ยจากตัวอย่าง $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ เป็นตัวประมาณค่าเฉลี่ยของประชากร (μ)

และใช้ส่วนเบี่ยงเบนมาตรฐานจากตัวอย่าง $s = \dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$ เป็นตัวประมาณค่า ส่วนเบี่ยงเบนมาตรฐานของประชากร (σ)

8.7 การแจกแจงของค่าเฉลี่ยที่ได้จากตัวอย่าง (The Sampling Distribution of)

การแจกแจงของค่าเฉลี่ยจากตัวอย่างที่สำคัญที่สุดคือ mean ถ้าเราทำการสุ่มตัวอย่างขนาดเท่า ๆกันมาหลาย ๆครั้งจากประชากรเดียวกัน จะเห็นว่าค่าเฉลี่ยจากตัวอย่างก็จะเปลี่ยนไปจากตัวอย่างชุดหนึ่งไปยังตัวอย่างอีกชุดหนึ่ง ซึ่งค่าเหล่านี้จะเป็นตัวกำหนดการแจกแจง

<u>ตัวอย่าง</u> สมมติว่ามีตัวเลขอยู่ 18 ตัว คือ

3	6	9	6	4	7
4	2	8	1	1	4
5	6	5	0	0	6

จากค่าเหล่านี้จะคำนวณค่าเฉลี่ย ได้ , ค่าความแปรปรวน ถ้าแบ่งข้อมูลทั้งหมดออกเป็น 6 กลุ่ม กลุ่มละ 3 จำนวน จากนั้นก็เอาจำนวนค่าเฉลี่ยของแต่ละกลุ่มซึ่งจะได้ค่าเฉลี่ยดังนี้ คือ 4 , 4.67 , 7.33 , 2.33 และ 5.67 ซึ่งค่าทั้ง 6 ค่านี้ เป็นค่าเฉลี่ยของตัวอย่างแต่ละชุด ซึ่งมีขนาดเท่ากับ 3 และค่าเฉลี่ยของค่าเฉลี่ยที่ได้จากตัวอย่างก็ยังคงมีค่าเป็น 4.39 แต่ค่าความแปรปรวนของค่าเฉลี่ยที่ได้จากตัวอย่างเพียง 3.21 ซึ่งมีค่าน้อยกว่าความแปรปรวนของข้อมูลเดิมมาก

- แสดงให้เห็นว่า ค่าเฉลี่ยที่มาจากแต่ละกลุ่มจะใกล้เคียงกันมากกว่าข้อมูลเดิม และถ้าขนาดของตัวอย่างยิ่งใหญ่มากเท่าไร ค่าเฉลี่ยที่ได้จากตัวอย่างก็จะมีค่าใกล้เคียงกับค่าเฉลี่ยจริงๆมากขึ้นเท่านั้น
- ในกรณีนี้เนื่องจาก เป็นตัวแปรสุ่ม จึงจำเป็นต้องทราบลักษณะการกระจายของค่าเฉลี่ยจากตัวอย่าง การกระจายของตัวแปรสุ่มจัดด้วยค่าของความแปรปรวน หรือส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวนของค่าเฉลี่ย จากตัวอย่างขึ้นอยู่กับความแปรปรวนในประชากรของตัวแปรเดิม

8.8 ทฤษฎีที่สำคัญเกี่ยวกับการสุ่มตัวอย่าง

ทฤษฎีที่ 1 : ถ้าตัวอย่างขนาด ถูกสุ่มมาจากประชากรที่มีการแจกแจงซึ่งมีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน แล้ว ค่าเฉลี่ยของตัวอย่างก็จะมีการแจกแจงซึ่งมีค่าเฉลี่ยเท่ากับ เหมือนกัน แต่ส่วนเบี่ยงเบนมาตรฐานมีค่าเพียง ถ้าเป็นการสุ่มแบบแทนที่ และจะมีค่าเป็น เมื่อเป็นการสุ่มตัวอย่างแบบไม่แทนที่

- **ทฤษฎีที่2:** ตัวอย่างขนาด ถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ ซึ่งมีค่าเฉลี่ย และความแปรปรวน การแจกแจงของ ก็จะเป็นแบบปกติเหมือนกัน ซึ่งมีค่าเฉลี่ยเท่าเดิม และค่าความแปรปรวนคลาดเคลื่อนมาตรฐานเป็น