

# ความรู้เบื้องต้นเกี่ยวกับสถิติ

## คำศัพท์ที่ควรรู้เกี่ยวกับสถิติ

- **สถิติ** มีความหมายกว้างๆ ได้ 2 ประการ

**ประการแรก** หมายถึง ข้อความจริงหรือตัวเลขซึ่งได้จากการรวบรวมไว้เพื่อหาความหมายที่แน่นอน เช่น สถิติพลเมือง สถิตินักท่องเที่ยว ฯลฯ

**ประการที่สอง** หมายถึง ศาสตร์แขนงหนึ่งว่าด้วยวิธีการเก็บรวบรวมข้อความจริง (ข้อมูล) การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการตีความตลอดจนสรุปผลข้อมูล

- **ประชากร** หมายถึง ขอบเขตของข้อมูลทั้งหมดที่เรากำลังศึกษา หรืออาจหมายถึงกลุ่มของสิ่งทั้งหมดที่ให้ข้อมูลตามที่เราต้องการศึกษา เช่น ต้องการหาอายุเฉลี่ยของคนไทย ประชากร คือ คนไทยทุกคน

- **ค่าพารามิเตอร์** หมายถึง ค่าที่ประมวลได้จากข้อมูลทั้งหมดของประชากรโดยวิธีการทางสถิติ นิยมใช้สัญลักษณ์กรีกแทน

เช่น ค่าเฉลี่ย แทนด้วย  $\mu$  อ่านว่า มิว

ค่าความแปรปรวน แทนด้วย  $\sigma^2$  อ่านว่า ซิกมากำลังสอง

ส่วนเบี่ยงเบนมาตรฐาน แทนด้วย  $\sigma$  อ่านว่า ซิกมา

- **ตัวอย่าง** หมายถึง ส่วนหนึ่งของประชากรซึ่งถูกเลือกมาศึกษา

- **ค่าสถิติ** หมายถึง ค่าที่ประมวลได้จากข้อมูลตัวอย่างโดยวิธีการทางสถิติ

เช่น ค่าเฉลี่ย แทนด้วย  $\bar{x}$

ค่าความแปรปรวน แทนด้วย  $s^2$

ส่วนเบี่ยงเบนมาตรฐาน แทนด้วย  $s$

● **การศึกษาสถิติแบ่งเป็น 2 สาขา คือ**

**สถิติพรรณนา** หมายถึง การศึกษาลักษณะของข้อมูลเฉพาะกลุ่มในรูปของการแจกแจงความถี่ การหาค่าสัดส่วนหรือร้อยละ การหาค่ากลางและการวัดการกระจาย ผลการศึกษาก็จะสรุปเฉพาะกลุ่มนั้นๆ จะไม่นำไปอ้างอิงถึงกลุ่มอื่นๆ

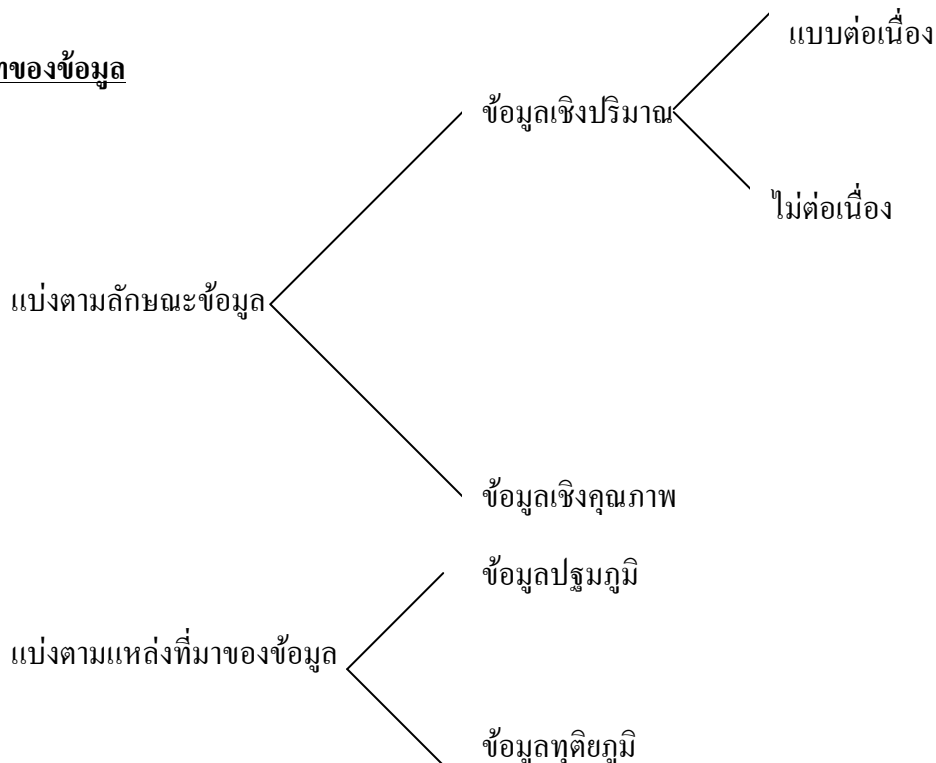
**สถิติอนุมาน** หมายถึง การศึกษาสรุปลักษณะของประชากรโดยใช้ข้อมูลตัวอย่าง จัดเป็นการวิเคราะห์ข้อมูลขั้นสูง ได้แก่ การประมาณค่า การทดสอบสมมติฐาน การวิเคราะห์ถดถอย และสหสัมพันธ์ การวิเคราะห์ความแปรปรวน เป็นต้น

## การจัดเก็บข้อมูล

### มาตราการวัด

1. มาตรานามบัญญัติ (Nominal Scale) เป็นมาตรวัดที่ไม่มีความละเอียด เป็นการแบ่งกลุ่มหรือจัดประเภทตามคุณลักษณะของสิ่งต่างๆ
2. มาตราเรียงอันดับ (Ordinal Scale) เป็นมาตรวัดที่แบ่งกลุ่มตามลำดับความสำคัญ
3. มาตราอันตรภาค (Interval Scale) เป็นมาตรวัดที่มีความละเอียด หน่วยการวัดคงที่สามารถระบุความแตกต่างระหว่างสิ่งต่างๆได้ เป็นมาตรวัดที่มีความหมายเชิงปริมาณ แต่ไม่มีศูนย์ที่แท้จริง (true zero) ข้อมูลที่ได้จากการวัดระดับนี้ สามารถนำมาบวกลบกันได้ แต่ไม่สามารถนำมาคูณและหารกันได้
4. มาตราอัตราส่วน (Ratio Scale) เป็นมาตรวัดในระดับที่มีความละเอียดมากที่สุดเท่าที่มีอยู่ มีศูนย์ที่แท้จริงซึ่งเป็นจุดเริ่มต้น

### ประเภทของข้อมูล



### การเก็บรวบรวมข้อมูล แบ่งเป็น 3 วิธี ดังนี้

1. การเก็บรวบรวมข้อมูลจากงานทะเบียน เช่นมหาวิทยาลัยจะมีการบันทึกหมายเหตุรายวัน ในเรื่องจำนวนอาจารย์ เจ้าหน้าที่ที่มาปฏิบัติราชการ , จำนวนอาจารย์ เจ้าหน้าที่ที่ลา , จำนวนอาจารย์ที่ไปราชการ เป็นต้น
2. การเก็บรวบรวมข้อมูลโดยการสำรวจ เป็นการเก็บรวบรวมข้อมูลจากหน่วยที่ศึกษาโดยตรง เช่น การสำรวจความคิดเห็นของประชาชนในการร่างรัฐธรรมนูญ ซึ่งหน่วยที่ศึกษาคือ

ประชาชนคนไทย การเก็บรวบรวมข้อมูลโดยการสำรวจจะแบ่งออกเป็น 2 ประเภท คือ การสำมะโน (Census) และการสำรวจตัวอย่าง (Sample Survey) กล่าวคือ การสำมะโนหมายถึง การเก็บรวบรวมข้อมูลจากทุกๆหน่วยศึกษาของประชากร ส่วนการสำรวจด้วยตัวอย่าง หมายถึง การเก็บรวบรวมข้อมูลจากบางหน่วยศึกษาของประชากร

3. การเก็บข้อมูลจากการทดลอง ข้อมูลบางประเภทไม่สามารถหาได้จากการสำรวจ แต่จัดทำได้จาก การทดลอง เช่น การศึกษาวิธีการปลูกพืชที่แตกต่างกัน 3 วิธี การศึกษาเปรียบเทียบผลผลิตข้าว 4 สายพันธุ์ เป็นต้น

**วิธีการสุ่มตัวอย่าง (Sampling)** แบ่งเป็น 2 ประเภทใหญ่ ๆ คือ

1. การสุ่มตัวอย่างโดยอาศัยหลักความน่าจะเป็น (Probability sampling) เป็นการสุ่มตัวอย่างที่คำนึงถึงความน่าจะเป็นของแต่ละหน่วยตัวอย่างที่จะได้รับเลือก โดยไม่จำเพาะเจาะจงเพื่อนำผลไปใช้อ้างอิง (Inference) ถึงประชากรมี 5 รูปแบบดังนี้

- 1.1 การสุ่มตัวอย่างแบบง่าย (Simple random Sampling)
- 1.2 การสุ่มตัวอย่างแบบเป็นระบบ (Systematic sampling)
- 1.3 การสุ่มตัวอย่างแบบสุ่มกลุ่ม (Cluster sampling)
- 1.4 การสุ่มตัวอย่างแบบแบ่งชั้นภูมิ (Stratified sampling)
- 1.5 การสุ่มตัวอย่างแบบหลายขั้นตอน (Multi – stage sampling)

2. การสุ่มตัวอย่างโดยไม่อาศัยหลักความน่าจะเป็น (Non probability sampling) เป็นการสุ่มตัวอย่างที่ไม่คำนึงถึงความน่าจะเป็นที่หน่วยตัวอย่างจะได้รับเลือกเป็นการสุ่มตัวอย่างแบบเจาะจง (purposive sampling) ส่วนมากมักจะใช้กับการวิจัยที่ไม่ต้องมีการกำหนดกรอบของประชากร เช่น การให้ลูกค้าทดสอบรสชาติของเครื่องดื่มชนิดใหม่ อาจไปตั้งโต๊ะแจกเครื่องดื่มตามศูนย์การค้า หรือ ซูเปอร์มาร์เก็ต แล้วให้ลูกค้าชิมเครื่องดื่มแล้วสอบถามความพอใจในรสชาติ กรณีนี้ไม่ต้องมีกรอบตัวอย่าง

#### การนำเสนอข้อมูล

1. การนำเสนอแบบบทความ
2. การนำเสนอแบบตาราง
  - ตารางทางเดียว (one – way table)
  - ตารางสองทาง (two – way table)
  - ตารางหลายทาง (multi – way table)
  - ตารางแจกแจงความถี่ (frequency distribution table)

โดยการสร้างตารางแจกแจงความถี่ มีขั้นตอน ดังนี้

ขั้นตอนที่ 1 หาพิสัยของข้อมูล (R)

$$\text{พิสัย (Range)} = \text{ค่าสูงสุด} - \text{ค่าต่ำสุด}$$

ขั้นตอนที่ 2 กำหนดจำนวนชั้น (K)

$$K = 1 + 3.3 \log N$$

ขั้นตอนที่ 3 คำนวณหาความกว้างของชั้น (Class Interval : I)

$$I = \text{ความกว้างของชั้น} = \frac{\text{พิสัย}}{\text{จำนวนชั้น}}$$

ขั้นตอนที่ 4 คำนวณหาขีดจำกัดชั้น (Class limit)

$$\text{ขีดจำกัดล่างของชั้นแรก} = \text{ค่าต่ำสุด} - (I \times K - R)/2$$

ขั้นตอนที่ 5 นับจำนวนค่าของข้อมูล (ความถี่) ในแต่ละชั้น

- ฮิสโตแกรม (Histogram) เป็นการนำเสนอข้อมูลที่ได้แจกแจงความถี่ในตารางแจกแจงความถี่แสดงเป็นภาพ
- โพลีกอน (Polygon) เป็นการนำเสนอข้อมูลให้เห็นเด่นชัดขึ้นโดยการลากเส้นตรงเชื่อมต่อระหว่างค่ากึ่งกลางชั้นของฮิสโตแกรม

### การสรุปข้อมูล

จากข้อมูลที่เก็บรวบรวมได้ นอกจากการนำเสนอด้วยรูปแบบต่างๆแล้ว การสรุปลักษณะของประชากรด้วยการวิเคราะห์ขั้นต้นนั้น โดยทั่วไปจะดูที่ค่ากลางและค่าวัดการกระจาย การคำนวณค่าดังกล่าวมีดังนี้

- การวัดแนวโน้มเข้าสู่ส่วนกลาง เป็นการคำนวณหาค่ากลางๆของข้อมูลว่ามีค่าเท่าใด ค่ากลางที่นิยมใช้กันอยู่ มี 3 ชนิด คือ

ก. ค่าเฉลี่ย (Mean) ซึ่งจำแนกได้ดังนี้

- ค่าเฉลี่ยเลขคณิต หรือมัธมิมเลขคณิต (Arithmetic Mean) เป็นค่าเฉลี่ยที่นิยมใช้กันมากที่สุดเพราะสามารถสื่อความหมายและทำความเข้าใจได้ง่าย และยังมีสมบัติทางสถิติที่ดี

ประชากร แทนด้วย  $\mu$   
 โดยที่ ข้อมูลไม่ได้จัดกลุ่ม  $\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{(X_1 + X_2 + \dots + X_N)}{N}$

ข้อมูลที่จัดกลุ่ม  $\mu = \frac{\sum_{i=1}^R X_i f_i}{N}$  เมื่อ  $X_i$  คือค่ากึ่งกลางของชั้นที่  $i$   
 $f_i$  คือความถี่ของชั้นที่  $i$

ตัวอย่าง แทนด้วย  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$

2. ค่าเฉลี่ยเรขาคณิต (Geometric Mean)

3. ค่าเฉลี่ยฮาร์โมนิก (Harmonic Mean)

ข. มัชฐาน (Median, Me) คือ ค่าของข้อมูลที่มีตำแหน่งอยู่ตรงกลางของชุดข้อมูล เมื่อนำชุดข้อมูลเรียงจากน้อยไปหามาก

- กรณีข้อมูลเป็นเลขคี่ ( $n$  เป็นเลขคี่)

ค่ามัชฐาน คือ ค่าของข้อมูลที่อยู่ตำแหน่ง  $\frac{(n+1)}{2}$

- กรณีข้อมูลเป็นเลขคู่ ( $n$  เป็นเลขคู่)

ค่ามัชฐาน คือ ค่าเฉลี่ยระหว่างข้อมูลที่อยู่ตำแหน่ง  $\frac{n}{2}$  และ  $\frac{n+2}{2}$

ในกรณีข้อมูลจัดกลุ่ม

$$M_e = L + \left( \frac{\frac{n}{2} - \sum f_L}{f} \right)$$

โดยที่  $L$  แทน ขอบเขตล่างของชั้นมัชฐาน

$I$  แทน ความกว้างของชั้นมัชฐาน

$n$  แทน จำนวนค่าสังเกตทั้งหมด

$\sum f_L$  แทน ผลรวมความถี่ของอันตรภาคชั้นทุกชั้นที่มีค่าสังเกตต่ำกว่าชั้นมัชฐาน

$f$  แทนความถี่ของชั้นมัชฐาน

ค. ฐานนิยม (Mode,  $M_o$ )

ในกรณีที่ข้อมูลไม่ได้จัดกลุ่ม ฐานนิยม คือ ค่าของข้อมูลที่เกิดขึ้นบ่อยที่สุดหรือมีความถี่สูงสุด

ในกรณีที่ข้อมูลจัดกลุ่ม

$$M_0 = L + I \left( \frac{d_1}{d_1 + d_2} \right)$$

โดยที่ L แทน ขอบเขตล่างของชั้นฐานนิยม

$d_1$  แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นก่อนฐานนิยม

$d_2$  แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นหลังฐานนิยม

I แทน ความกว้างของชั้นฐานนิยม

## 2. การวัดการกระจาย

การสรุปลักษณะของข้อมูล โดยทั่วไปมักดูที่ค่ากลาง และค่าการกระจาย เพราะข้อมูล 2 ชุด อาจมีค่ากลางเท่ากัน แต่ค่าการกระจายไม่เท่ากัน ดังนั้นการดูแต่ค่ากลาง อาจไม่ช่วยให้เห็นลักษณะของข้อมูลที่ชัดเจน เช่น มีข้อมูล 2 ชุด คือ A และ B

$$\text{ชุด A : } 4 \quad 5 \quad 7 \quad 9 \quad 10 ; \bar{X} = 7$$

$$\text{ชุด B : } 2 \quad 4 \quad 7 \quad 10 \quad 12 ; \bar{X} = 7$$

จะเห็นว่าข้อมูลชุด A และ B มีค่ากลางเท่ากันแต่การกระจายของข้อมูลไม่เท่ากัน ดังนั้นการเปรียบเทียบข้อมูล 2 ชุด ก็ควรพิจารณาทั้งค่ากลางและค่าวัดการกระจาย ค่ากลางที่นิยมใช้มีดังนี้

ก.พิสัย (Range)

เป็นความแตกต่างระหว่างค่าสังเกตสูงสุดและค่าสังเกตต่ำสุดของข้อมูล

$$R = X_{\max} - X_{\min}$$

ข. ส่วนเบี่ยงเบนเฉลี่ย (MD)

เป็นค่าเฉลี่ยของความแตกต่างระหว่างค่าสังเกตและค่าเฉลี่ย

ประชากร

$$MD = \frac{\sum |x - \mu|}{N}$$

ตัวอย่าง

$$MD = \frac{\sum |x - \bar{x}|}{n}$$

ค. ส่วนเบี่ยงเบนมาตรฐาน และเมื่อยกกำลังสองจะเป็นความแปรปรวน

ประชากร

$$\sigma^2 = \frac{\left[ \sum_{i=1}^N X_i^2 - \frac{\left( \sum_{i=1}^N X_i \right)^2}{N} \right]}{N} \quad (\text{ข้อมูลไม่ได้จัดกลุ่ม})$$

$$= \frac{\sum_{i=1}^R f_i X_i^2 - \frac{\left[ \sum_{i=1}^R f_i X_i \right]^2}{N}}{N} \quad (\text{ข้อมูลจัดกลุ่ม})$$

ตัวอย่าง

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - \frac{\left( \sum_{i=1}^n X_i \right)^2}{n}}{n - 1}$$

ง. ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variance) หรือสัมประสิทธิ์การกระจาย

เป็นค่าที่ใช้วัดการกระจายของข้อมูลที่ไม่มีหน่วย ถ้าข้อมูลชุดใดมีค่า C.V. มาก จะมีการกระจายมากกว่าข้อมูลที่มีค่า C.V. น้อย

ประชากร

$$C.V. = \frac{\sigma}{\mu} \times 100$$

ตัวอย่าง

$$C.V. = \frac{S}{\bar{X}} \times 100$$

\*\*\*\*\*