การวิเคราะห์ข้อมูลเชิงสถิติ

⇒ เป็นการอธิบายและสรุปข้อมูลที่รวบรวมได้จากกลุ่มตัวอย่าง / ประชากร

⇒ อธิบายว่าอะไรเป็นลักษณะแบบอย่างของกลุ่ม “mean”

⇒ ชี้ว่าแต่ละบุคคลในกลุ่มผันแปรไปเพียงใด “S.D”

⇒ แสดงความสัมพันธ์ระหว่างตัวแปร/ทิศทาง/ระดับความสัมพันธ์

⇒ อธิบายความแตกต่างระหว่างกลุ่ม F–Test/T–Test

สถิติสำหรับการวิเคราะห์ข้อมูล

สถิติ

⇒ เป็นศาสตร์ประยุกต์/สามารถเอาวิธีการทางสถิติไปใช้ในกรณีต่าง ๆ ได้

หน้าที่ของสถิติ

- **Descriptive/summarize (สรุป/บรรยายข้อมูล)**

⇒ สรุปข้อมูลต่าง ๆ

เช่น A=10 / B=15 ส่วนใหญ่ B

⇒ บอกการกระจาย เช่น S.D.

⇒ สรุปทีละตัวแปร เช่น Mean/Mode

⇒ สรุปจากหลายตัวแปร เพื่อหาความสัมพันธ์

หน้าที่ของสถิติ (ต่อ)

- **Inductive/Inference (อุปมา/คาดคะเน/อ้างอิง)**

⇒ **ทดสอบความสัมพันธ์/ความแตกต่าง**

⇒ **Estimation/คาดคะเน**

ประเภทของสถิติ

1. Descriptive Stat. (สถิติบรรยาย)
2. Inference Stat. (สถิติอ้างอิง)

1. Descriptive Stat. (สถิติบรรยาย)

ใช้บรรยายข้อมูลที่มีเท่านั้น

1.1 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central Tendency Measure)

1) มัชฌิมเลขคณิต (Arithematic mean)

$\Rightarrow$ Mean หรือค่าเฉลี่ย = $\frac{EX}{N}$

Descriptive Stat. (ต่อ)

2) มัฐยฐาน (Median)

⇒ นำสิ่งที่จะศึกษามาเรียงตามความมากน้อย แล้วแบ่งครึ่ง ค่าที่แบ่งครึ่งคือ Me

3) ฐานนิยม (Mode)

⇒ คือค่าที่มีความถี่สูงสุด Mo

Descriptive Stat. (ต่อ)

1.2 การวัดการกระจาย (Disperson)

✳ บอกความแตกต่างของข้อมูล

1) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) S.D

⇒ บอกความแตกต่างอย่างเฉลี่ยของข้อมูล

Descriptive Stat. (ต่อ)

2) ความแปรปรวน (Variance)

⇒ บอกการกระจายอีกลักษณะหนึ่ง

⇒ มีค่าเท่ากับ S.D. ยกกำลังสอง

3) พิสัย (Range)

⇒ บอกความแตกต่างอย่างหยาบ สูง–ต่ำ

Descriptive Stat. (ต่อ)

1.3 การนำเสนอข้อมูล (Data Presentation)

⇒ แสดงด้วยตารางแจกแจง (Table)

⇒ แสดงด้วยรูปภาพ กราฟ ฯลฯ

1.4 การเปรียบเทียบข้อมูล

⇒ แปลงคะแนนดิบเป็นคะแนนมาตรฐาน

2. Inference Stat. (สถิติอ้างอิง)

- เป็นการอ้างอิงจากค่าสถิติ (ที่คำนวณจากกลุ่มตัวอย่าง) ไปยังค่าพารามิเตอร์ (ที่คำนวณจากประชากร)

⇒ โดยอาศัยเทคนิคการสรุปอ้างอิงทางสถิติศาสตร์ ภายใต้ความน่าจะเป็น (Probability)

การอ้างอิงทางสถิติ มี 2 ลักษณะ

1. การทดสอบสมมติฐานทางสถิติ (Hypothesis testing)
2. การประมาณค่าพารามิเตอร์ (ประชากร)

สมมติฐาน

- **สมมติฐาน** ⇒ การคาดเดาคำตอบในการวิจัย ซึ่งนักวิจัยต้องการทดสอบว่าเป็นความจริงหรือไม่

1) **สมมติฐานวิจัย** ⇒ เขียนเชิงบรรยายความสัมพันธ์ระหว่างตัวแปรที่นักวิจัยคาดคะเนไว้

2) **สมมติฐานทางสถิติ** ⇒ แปลงสมมุติฐานวิจัยมาเขียนในรูปสัญลักษณ์และตัวเลข เพื่อการทดสอบทางสถิติ

สมมติฐานทางสถิติ

- สมมุติฐานสูญ (Null Hyp. : H_0)

  ⇒ ตั้งขึ้นเพื่อถูกปฏิเสธ

  ⇒ มีข้อตกลงเบื้องต้นว่า น่าจะเป็นจริง

  ⇒ ไม่แตกต่าง/ไม่สัมพันธ์

- สมมุติฐานอื่น (Alternative Hyp. : H_1)

  ⇒ ใช้เมื่อ H_0 ถูกปฏิเสธ

การทดสอบสมมุติฐาน

- ต้องทดสอบ H_0 (Accept/Reject)

⇒ H_0 ควรได้รับการปฏิเสธภายใต้ความคลาดเคลื่อนที่น้อยที่สุด

แบบที่ 1 / Type Ⅰ error

⇒ ในบางกรณีต้องยอมรับ H_0 ก็ควรจะยอมรับภายใต้(ความคลาดเคลื่อนแบบที่ 2 / Type Ⅱ error)

ความคลาดเคลื่อนที่เกิดขึ้น

สมมุติฐาน H_0

	ถ้าจริง	ถ้าไม่จริง
Accept	ไม่มี	2 (β)
Reject	1 (α)	ไม่มี

Type Ⅰ error เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้วิจัย Reject H_0

Type Ⅱ error เป็นความคลาดเคลื่อนที่เกิดขึ้นเมื่อผู้วิจัย Accept H_0 ที่ไม่เป็นจริง (รุนแรงกว่า Type Ⅰ)

ตัวอย่าง การทดสอบ Hyp.

H_0 : นักศึกษาไม่แตกต่างจากโจร

H_1 : นักศึกษาและโจรแตกต่างกัน

✹การตัดสินใจ (จากผลการทดสอบ)

⇒ ถ้า Reject H_0= นักศึกษาไม่ใช่โจร

(เป็นความคลาดเคลื่อนแบบที่ 1/ปฏิเสธ H_0ที่เป็นจริง)

⇒ ถ้า Accept H_0= นักศึกษาคือโจร

(เป็นความคลาดเคลื่อนแบบที่ 2/ยอมรับ H_0 ที่ไม่เป็นจริง)

ตัวอย่าง การทดสอบ Hyp.

$H_0 : \mu_1 = \mu_2$

H_1 : เขียนได้ 2 ลักษณะ

$H_1 : \mu_1 \neq \mu_2$ (ไม่แสดงทิศทาง)

$H_1 : \mu_1 < \mu_2$ (แสดงทิศทาง)

การทดสอบสมมุติฐาน (H_0)

- ความน่าจะเป็น หรือ โอกาสเสี่ยง ดังนี้

1. ถ้ายอมรับ H_0ที่ไม่จริง จะเป็น β (รุนแรงมาก)

2. ถ้าปฏิเสธ H_0ที่เป็นจริง จะเป็น $\propto$ (รุนแรงน้อย)

- ในทางปฏิบัติ เราจะตั้ง $\propto$ ไว้ (ระดับความมีนัยสำคัญ)

$\Rightarrow$ เพราะความน่าจะเป็นด้านการปฏิเสธ H_0 ที่เป็นจริงทำได้ง่ายกว่าการยอมรับ H_0ที่ไม่จริง

$\Rightarrow$ ทั้งนี้ เพราะเราไม่ทราบว่า Hyp. นั้นเป็นจริงหรือไม่

ระดับความมีนัยสำคัญ (Level of significance) (α)

- หมายถึงความน่าจะเป็นในการปฏิเสธ (reject) H_0 ที่เป็นจริง
- การตั้งระดับความมีนัยสำคัญ (α) ให้ความหมายคือ

$\Rightarrow$ โอกาสที่ α ครั้งที่ปฏิเสธ H_0 ที่เป็นจริง

เช่น $\alpha = 0.05$

$\Rightarrow$ การยอมให้ข้อสรุปเกิดความคลาดเคลื่อน 5 ครั้ง ใน 100 ครั้ง ของการทดลอง/ทดสอบทั้งหมด (Type I error)

การประมาณค่าพารามิเตอร์

- เป็นการอ้างอิงทางสถิติ จากกลุ่มตัวอย่างที่ศึกษาไปยังประชากร

1. การประมาณค่าที่เป็นจุด (point estimation)

⇒ เป็นการใช้ค่าสถิติประมาณค่าพารามิเตอร์ เช่น $\bar{X}$ ประมาณค่า μ

การประมาณค่าพารามิเตอร์ (ต่อ)

2. การประมาณค่าที่เป็นช่วง (Interval estimation)

$\Rightarrow$ เป็นการใช้ค่าสถิติ ประมาณค่าพารามิเตอร์เป็นช่วง เช่น ค่า μ จะอยู่ระหว่าง k_1 และ k_2 ภายใต้ความเชื่อมั่น 95 %

สถิติอ้างอิง (Inference statistics)

มีเทคนิคทางสถิติ 2 ประเภท

1) Parametric statistic

2) Non-Parametric statistic

Parametric statistic (พาราเมตริก)

1) ประมาณค่า Parameter จากค่าสถิติที่ศึกษาจากกลุ่มตัวอย่าง โดยการสุ่ม (Sampling)

2) การแจกแจงของประชากรเป็นแบบโค้งปกติ

3) ค่าความแปรปรวน (Variance) คงที่

สถิติแบบพาราเมตริก

- เป็นเทคนิคทางสถิติ (Inference statistic)ที่ใช้อ้างอิงจากกลุ่มตัวอย่าง ไปหาค่าประชากร มี 4 อย่างที่สำคัญ

  ⇒T – test

  ⇒ Z – test

  ⇒ X^2 –test

  ⇒ F – test

หลักในการใช้สถิติทดสอบ

1) ศึกษาข้อตกลงเบื้องต้นของสถิติแต่ละตัวก่อน

2) เลือกสถิติทดสอบให้เหมาะสมกับเงื่อนไขของการวิจัย

3) ต้องรู้ลักษณะการแจกแจงของสถิติแต่ละตัว

Non-Parametric statistic (นันพาราเมตริก)

1) กลุ่มตัวอย่างที่ศึกษา ไม่ได้มาจากการสุ่ม

2) การแจกแจงของประชากร ไม่เป็นแบบโค้งปกติ

ปัจจัยในการพิจารณาเลือกใช้สถิติวิเคราะห์ข้อมูล

1. วัตถุประสงค์ของการวิเคราะห์ข้อมูลคืออะไร

* บรรยายข้อมูลตามสภาพที่เป็นอยู่

⇒ Descriptive statistic

* สรุปอ้างอิงไปยังประชากร

⇒ Inference statistic

ปัจจัยในการพิจารณาเลือกใช้สถิติวิเคราะห์ข้อมูล(ต่อ)

2. จำนวนกลุ่มตัวอย่างที่ศึกษา

- 1 กลุ่ม
- 2 กลุ่ม
- มากกว่า 2 กลุ่ม

ปัจจัยในการพิจารณาเลือกใช้สถิติวิเคราะห์ข้อมูล(ต่อ)

3. ระดับของข้อมูลที่จะวิเคราะห์

- จัดพวก Norminal scale
- จัดลำดับ Ordinal scale
- ช่วง Interval scale
- อัตราส่วน Ratio scale

ปัจจัยในการพิจารณาเลือกใช้สถิติวิเคราะห์ข้อมูล(ต่อ)

4. จำนวนกลุ่มตัวอย่างที่ศึกษา

- **1 ตัวแปร Univariate Analysis**
- **2 ตัวแปร Bivariate Analysis**
- **มากกว่า 2 ตัวแปร Multi – Variate Analysis**