WJ
204490

Research Tools: Statistic

สถิติเพื่อการวิจัย

จุดมุ่งหมายของการวิจัย

- เพื่อนำไปสู่การแก้ปัญหา การวางแผนและตัดสินใจ โดยผู้วิจัยจะต้องตระหนัก ในประเด็นต่อไปนี้
  - จะใช้ข้อมูลอะไรบ้าง
  - จะหาข้อมูลเหล่านั้นมาได้อย่างไร
  - จะใช้วิธีการอะไรในการวิเคราะห์ข้อมูล
    - ✓สถิติศาสตร์
    - ✓ศาสตร์อื่น ๆ ที่เกี่ยวข้อง

บทบาทของสถิติต่องานวิจัย

- ทำไมงานวิจัยต้องใช้สถิติ?
  - เพื่อหาข้อเท็จจริง ผ่านระบบการเก็บรวบรวมข้อมูลที่เป็นตัวแทนของข้อมูลส่วนใหญ่
  - เพื่อให้ได้งานวิจัยที่มีคุณภาพ น่าเชื่อถือ
- งานวิจัยที่ต้องใช้สถิติ
  - งานวิจัยโดยใช้การทดลอง (Experimental Research)
  - งานวิจัยโดยการสำรวจ (Survey Research)

งานวิจัยโดยใช้การทดลอง

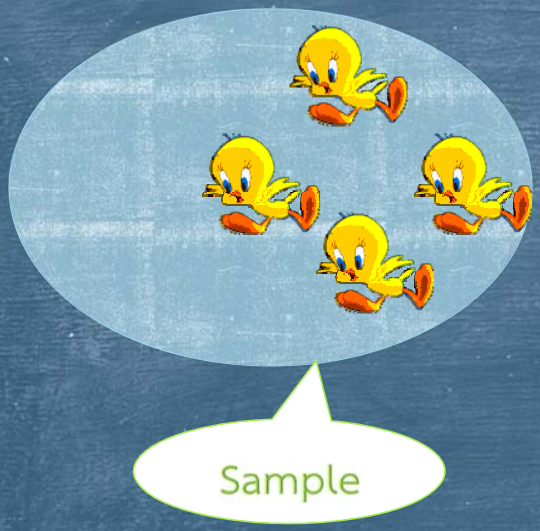
- สามารถกำหนดตัวแปรไว้ได้ล่วงหน้า และพิจารณาว่าจะควบคุมหน่วยทดลอง (Experiment Unit) ได้อย่างไร จึงจะไม่มีผลกระทบต่อหน่วยทดลองอื่น ๆ

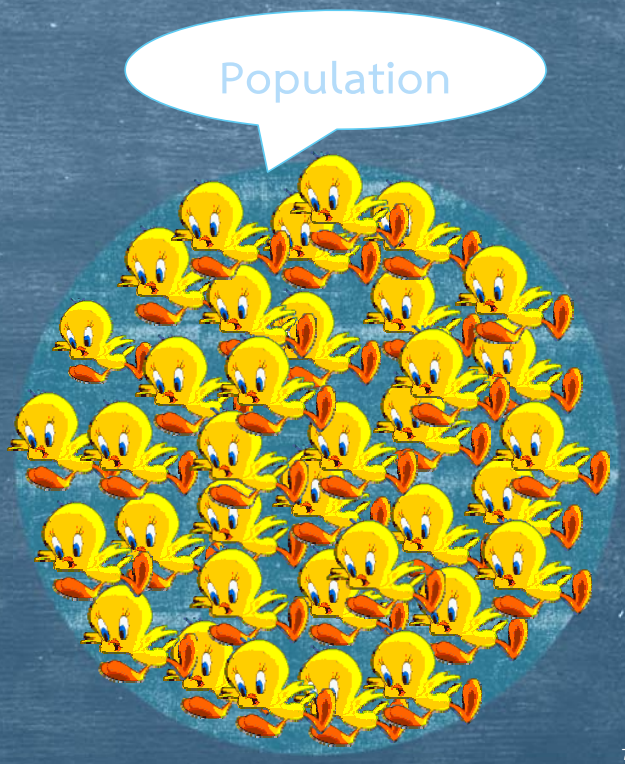
งานวิจัยโดยการสำรวจ

- จะต้องพิจารณาว่า จะเลือกตัวอย่างเพื่อเป็นตัวแทนที่ดีของประชากรได้อย่างไร
- แผนการเลือกตัวอย่าง
- วิธีการเลือกตัวอย่าง
- วิธีประมาณตัวอย่างที่เหมาะสม

ค่าพารามิเตอร์ (Parameter) หมายถึง
ค่าที่ใช้บอกคุณลักษณะใดๆของประชากร

ค่าสถิติ (Statistic) หมายถึงค่า
คำนวณได้จากข้อมูลที่เก็บจาก
กลุ่มตัวอย่าง ซึ่งเป็นค่าที่ใช้
ประมาณค่าพารามิเตอร์





การกำหนดกลุ่มตัวอย่าง

- เป็นการพิจารณาว่า จะเก็บรวบรวมข้อมูลจากแหล่งใด โดยใช้วิธีการใด ใช้แผนการเลือกตัวอย่างใด และใช้ขนาดตัวอย่างเท่าไร
- วิธีการเลือกตัวอย่างในการวิจัย มี 4 วิธี
  - การเลือกตัวอย่างอย่างง่าย (Simple Random Sampling) : SRS
  - การเลือกตัวอย่างแบบมีระบบ (Systematic Sampling)
  - การเลือกตัวอย่างแบบแบ่งชั้นภูมิ (Stratified Sampling)
  - การเลือกตัวอย่างแบบแบ่งกลุ่ม (Cluster Sampling)

(๑) การเลือกตัวอย่างอย่างง่าย

- เป็นวิธีการเลือกตัวอย่างขนาด n จากประชากร N หน่วย
- ใช้วิธีการจับฉลาก หรือใช้เลขสุ่ม
- เหมาะกับประชากรมีขนาดไม่มาก และไม่แตกต่างกันมากนัก
- ข้อสำคัญ ประชากรต้องมีกรอบที่แน่นอน

(๒) การเลือกตัวอย่างแบบมีระบบ

- เป็นวิธีการเลือกตัวอย่างขนาด n หน่วยจากประชากร N หน่วย
- แต่ละหน่วยตัวอย่างมีโอกาสถูกเลือกเท่าเทียมกัน
- ใช้วิธีการกำหนดตัวอย่าง ตามขั้นตอนต่อไปนี้
  - หาช่วงกว้างของการสุ่ม (Sampling interval) $I = \frac{N}{n}$
  - กำหนดเลขที่เหมาะแก่ประชากรทุกหน่วย
  - เลือกเลขสุ่ม R ที่มีค่าระหว่าง $1 < R < I$
  - หน่วยตัวอย่างที่ถูกเลือก คือ R, R+I, R+2I, . . .

วิธีการเลือกตัวอย่างแบบมีระบบ

N = 100

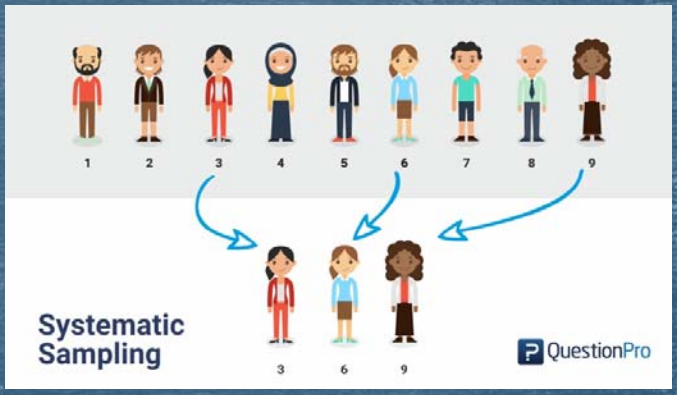
want n = 20

N/n = 5

select a random number from 1-5: chose 4

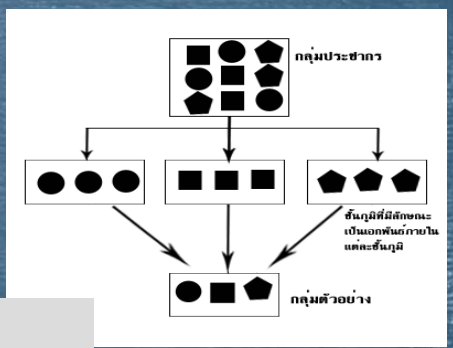
start with #4 and take every 5th unit

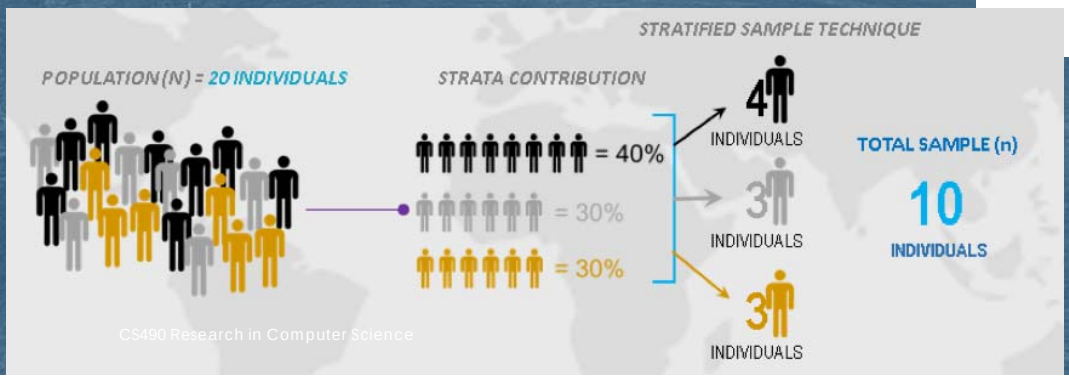
1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100



(๓) วิธีการเลือกตัวอย่างแบบแบ่งชั้นภูมิ

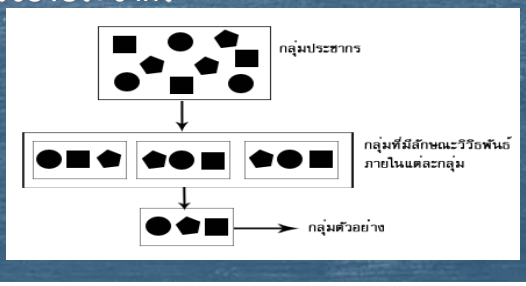
- ประชากรถูกแบ่งเป็นชั้นภูมิ (Stratum)
- ภายในชั้นภูมิเดียวกันต้องมีความคล้ายคลึงกันมากที่สุด
- ต่างชั้นภูมิกัน จะต้องมีความแตกต่างกันมากที่สุด
- เลือกตัวอย่างจากประชากร ในแต่ละชั้นภูมิ

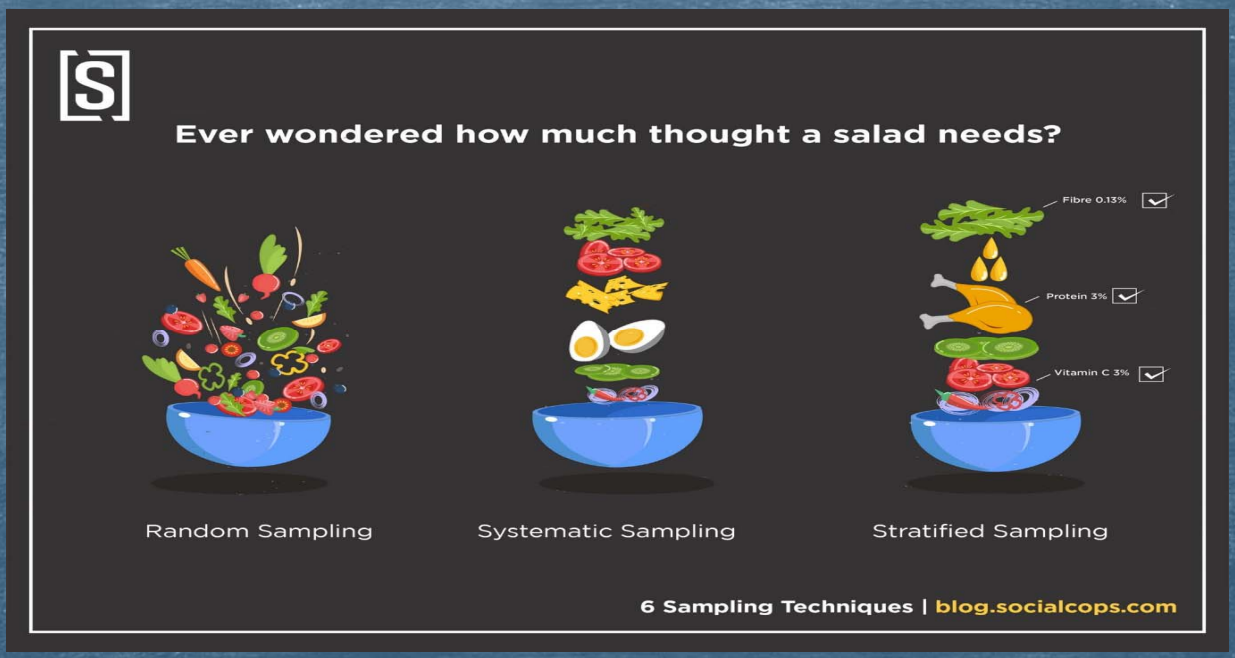




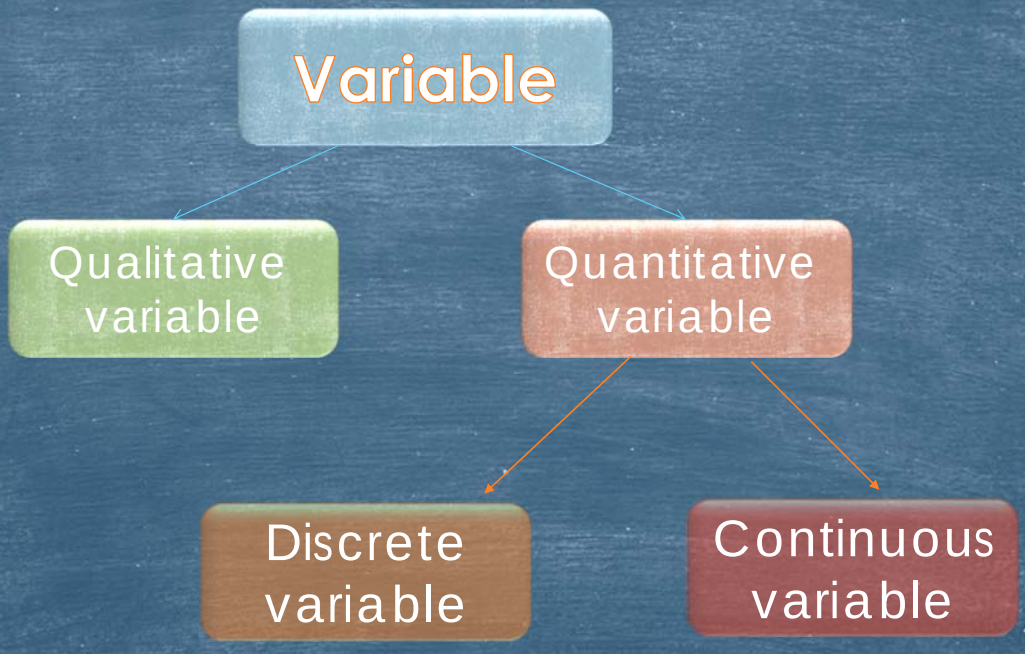
(๔) วิธีการเลือกตัวอย่างแบบแบ่งกลุ่ม

- พบมากในการวิจัยขนาดใหญ่ ข้อมูลมีการกระจายมาก
- การสร้างกรอบตัวอย่างทำได้ยาก, ค่าใช้จ่ายสูง
- การเลือกตัวอย่างแบบแบ่งกลุ่ม มีขั้นตอนสำคัญคือ
  - แบ่งประชากรออกเป็นกลุ่มย่อย ๆ
  - แต่ละกลุ่มย่อยมีความแตกต่างกันมาก ๆ
  - เลือกกลุ่มย่อยมาเพียงบางกลุ่มเป็นตัวแทนของประชากร





ตัวแปร (Variable)



Qualitative variable

ตัวแปรที่มีค่าได้ต่างๆกัน แต่ค่าดังกล่าวไม่ได้อยู่ในรูปของจำนวน หรือขนาด ส่วนใหญ่อยู่ในรูปของคุณภาพ หรือชนิด ซึ่งเรียกว่า คุณลักษณะ (Attribute) เช่น

ตัวแปรเพศ มีค่าได้ 2 ชนิด คือ ชาย กับ หญิง

ตัวแปรการศึกษา มีค่าได้แตกต่างกัน

เช่น ประถมศึกษา มัธยมศึกษา ปริญญาตรี

Quantitative variable

ตัวแปรที่ค่าต่างๆ อยู่ในรูปของขนาด

โดยที่ขนาดสามารถบอกความมากน้อยได้

เช่น อายุ ความสูง เป็นต้น

ตัวแปรเชิงปริมาณ แบ่งเป็น Continuous variable

และ Discrete variable

Continuous variable คือ ตัวแปรที่มีค่าใดๆก็ได้ในพิสัยหนึ่งๆที่กำหนดให้ ค่าที่อยู่ในพิสัยนั้นมากมายนับไม่ถ้วน

Discrete variable คือ ตัวแปรที่ไม่สามารถมีค่าทุกค่าในพิสัยหนึ่งที่กำหนดให้ ค่าที่อยู่ในพิสัยนี้ไม่ต่อเนื่องกันและนับจำนวนได้ว่ามีกี่ค่า

Scale of measurement

1.ระดับมาตรานามบัญญัติ (Nominal Scale)

2.ระดับมาตราเรียงอันดับ (Ordinal Scale)

3.ระดับมาตรอันตรภาค (Interval Scale)

4.ระดับมาตราอัตราส่วน (Ratio Scale)

1. มาตรานามบัญญัติ (Nominal scale)

เป็นมาตรา ที่กำหนดเป็นตัวเลขหรือสัญลักษณ์ต่างๆ ที่กำหนดขึ้นเพื่อเรียกชื่อหรือเพื่อจำแนกสิ่งต่างๆ ออกจากกัน เช่น เพศ เบอร์ของนักฟุตบอล ศาสนา
สิ่งต่างๆ ที่ได้รับการกำหนดชื่อหรือจัดจำแนก มีคุณสมบัติเท่าเทียมกัน ทุกประการ
ไม่มีความหมายในเชิงปริมาณ
ไม่สามารถนำมาบวก ลบ คูณหารได้
แต่สามารถแจกแจงความถี่

2. มาตราเรียงอันดับ (Ordinal scale)

เป็นระดับการวัดที่คุณสมบัติการวัดสูงกว่ามาตรานามบัญญัติ
ตัวเลขหรือสัญลักษณ์ที่ใช้แทนคุณลักษณะแต่ละหน่วย
นอกจากบอกชื่อและจำแนกแล้ว
ยังสามารถแสดงปริมาณความมากน้อย (Magnitude)
จัดเรียงลำดับได้ แต่ช่วงของความห่างของแต่ละอันดับไม่
เท่ากัน

3. มาตราอันตรภาค (Interval scale)

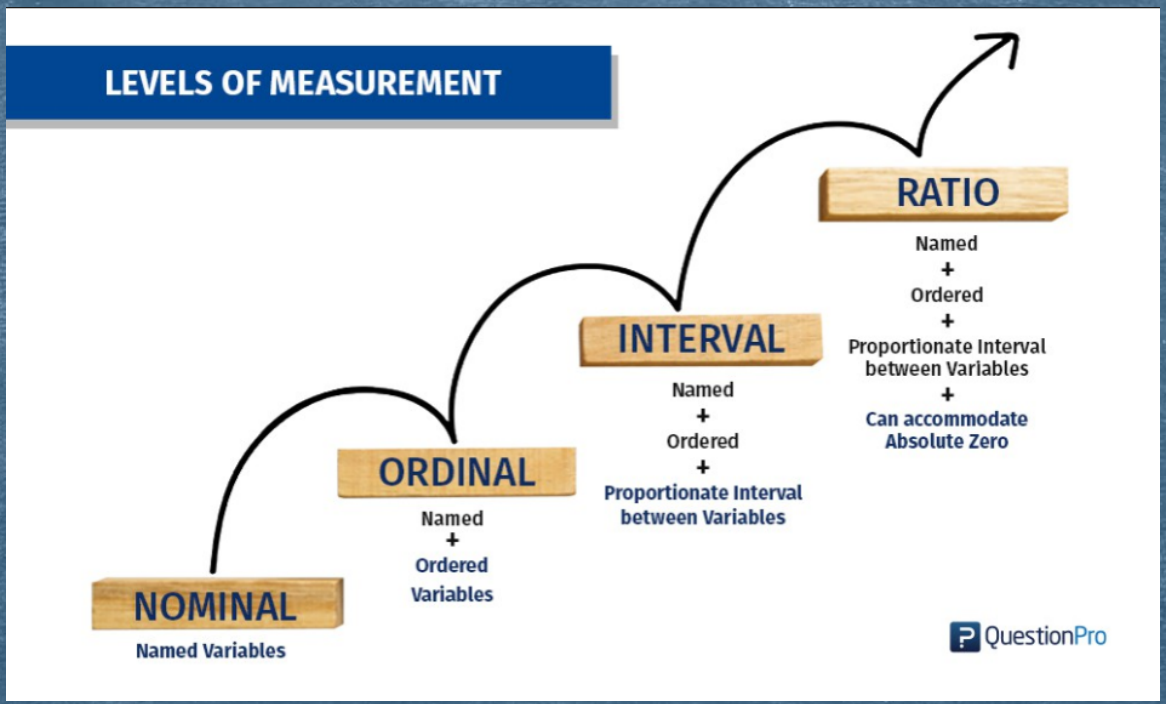
เป็นข้อมูลที่มีระดับการวัดที่มีคุณสมบัติสูงขึ้นจากมาตรา
เรียงอันดับ
สามารถแสดงปริมาณความมากน้อยได้ (magnitude)
ความแตกต่างระหว่างแต่ละหน่วยมีค่าเท่ากัน (Equal
interval)
ไม่มีศูนย์แท้ (non absolute zero)

4.ข้อมูลระดับอัตราส่วน (Ratio scale)

เป็นข้อมูลที่มีระดับการวัดที่มีคุณสมบัติทางคณิตศาสตร์ครบถ้วน 3 ประการคือ

- magnitude
- Equal interval
- Absolute zero

สามารถบวก ลบ คูณหาร ได้



สถิติเชิงพรรณา (descriptive statistics)

เป็นสถิติที่ใช้บรรยายคุณลักษณะของสิ่งที่ต้องการศึกษา

1. การวัดแนวโน้มเข้าสู่ส่วนกลาง

2. การวัดการกระจาย

1 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measurement of central tendency)

1.1 มัชฌิมเลขคณิต (Arithmetic mean)
หรือ ค่า Mean (ค่าเฉลี่ย)

สำหรับประชากร

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

สำหรับกลุ่มตัวอย่าง

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

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f}$$

สำหรับข้อมูลที่เป็นความถี่

1.2 มัธยฐาน (Median : Md)

หมายถึงค่าที่อยู่ตำแหน่งกึ่งกลางของข้อมูลแต่ละชุด

Md = 7

ข้อมูลชุดที่ 1 : 6 8 9 10 7 5 3 2 11

ข้อมูลชุดที่ 2 : 12 15 10 17 18 11

เรียงลำดับข้อมูล แล้วหาข้อมูล ณ ตำแหน่งกลาง Md = (12+15)/2=13.5

การหามัธยฐานข้อมูลที่มีการแจกแจงความถี่

$$Md = L + i\left[\frac{\frac{N}{2} - F}{f}\right]$$

L แทน ขีดจำกัดล่างที่แท้จริงของชั้นที่มีมัธยฐานอยู่

N แทน จำนวนคะแนนทั้งหมด

F แทน ความถี่สะสมตั้งแต่คะแนนต่ำสุดถึงชั้นก่อนชั้นที่มัธยฐาน

f แทน ความถี่ของคะแนนในชั้นที่มีมัธยฐาน

i แทน อันตรภาคชั้น

ตัวอย่าง

คะแนน	ความถี่	ความถี่สะสม
30-34	2	25
25 -29	3	23
20-24	5	20
15-19	7	15
10 -14	4	8
5 - 9	3	4
0 - 4	1	1
	∑f = 25	

- Md = L + I (N/2 – F)/f
- Md = 14.5 + 5 (25/2 – 8)/7
  = 17.71

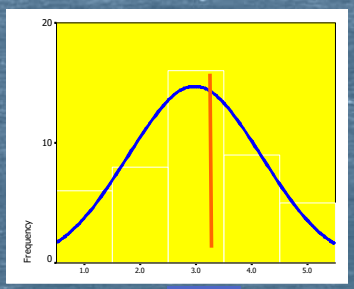
1.3 ฐานนิยม (Mode : Mo)
หมายถึงค่าที่ความถี่สูงสุดของข้อมูลชุดหนึ่งๆ

Mo = 16

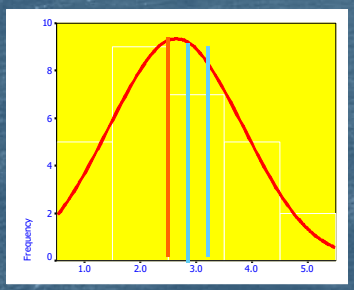
ข้อมูลชุดที่ 1 : 12 13 15 15 16 16 16
ข้อมูลชุดที่ 2 : 10 12 15 15 16 17 17 19 20

Mo = 15 และ 17

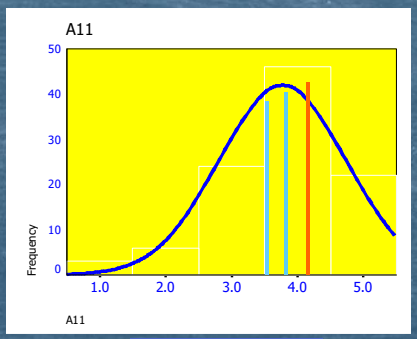
ความสัมพันธ์ระหว่าง X , Md, Mo



X
Md
Mo



Mode,Md,X



X, Md, Mo

2. การวัดการกระจาย (Measurement of dispersion)

2.1 พิสัย (Range)
2.2 ความเบี่ยงเบนควอไทล์(Quartile deviation)
2.3 ความแปรปรวน(variance)
2.4 ความเบี่ยงเบนมาตรฐาน (standard deviation)

2.1 พิสัย(Range)

พิสัย คือ ความแตกต่างระหว่างค่าสูงสุด
กับค่าต่ำสุดของข้อมูลชุดหนึ่งๆ

พิสัย = ค่าสูงสุด - ค่าต่ำสุด
ถ้า ค่าพิสัยเป็น 0 แสดงว่า ข้อมูลไม่มีการกระจาย
ถ้าพิสัยน้อยแสดงว่ากระจายน้อย
ถ้าค่าพิสัยมากแสดงว่าข้อมูลชุดนั้นมีการกระจายมาก

ตัวอย่าง

จงหาพิสัยของคะแนน 10 คน ซึ่งมีดังนี้ :

15,12,18,20,17,11,15,13,19,8

พิสัย = 20 – 8 = 12

ข้อสังเกตเกี่ยวกับพิสัย

1. ถ้าข้อมูลมีจำนวนมาก ค่าพิสัยมีแนวโน้มที่จะสูงด้วย ดังนั้นจึงไม่เหมาะในการเปรียบกรณีที่จำนวนข้อมูลไม่เท่ากัน
2. การคำนวณพิสัยมาจากตัวเลข 2 ตัว อาจเป็นตัวแทนที่ดีหรือไม่ดีก็ได้
3. พิสัยเหมาะกับกลุ่มตัวอย่างขนาดเล็ก เพราะให้ค่าคงที่มากกว่ากลุ่มตัวอย่างขนาดใหญ่
4. พิสัยเหมาะกับการหาค่าการกระจายของข้อมูลแบบคร่าวๆ

ชูศรี วงศ์รัตนะ.2544 ; Chase.1967 ; Huntsberger.1967

2.2 ความเบี่ยงเบนควอไทล์ (Quartile deviation)

ความเบี่ยงเบนควอไทล์ คือ ค่าที่ใช้วัดการกระจายของข้อมูลซึ่งหาได้จากครึ่งหนึ่งของความแตกต่างระหว่างควอไทล์ที่ 3 (Q_3) กับ ควอไทล์ที่ 1 (Q_1)

$$Q.D = \frac{Q_3 - Q_1}{2}$$

ในกรณีที่ข้อมูลไม่ถูกแจกแจงความถี่

☺ **การหาค่าควอไทล์ เดไซล์ และเปอร์เซ็นต์ไทล์**

เมื่อข้อมูลไม่ได้แจกแจงความถี่

สามารถดำเนินการตามขั้นตอนดังนี้

1. เรียงลำดับข้อมูลจากน้อยไปหามาก
2. หาตำแหน่งของข้อมูลที่ต้องการ โดย

$$Q_k = \frac{k(N+1)}{4} \quad , \quad D_k = \frac{k(N+1)}{10} \quad \text{และ} \quad P_k = \frac{k(N+1)}{100}$$

3. นำค่าของตำแหน่งที่ได้ไปเปรียบเทียบกับข้อมูลว่าตรงกับข้อมูลค่าใด

ตัวอย่างการคำนวณ (๑)

ตัวอย่าง เด็กกลุ่มหนึ่งจำนวน 7 คน มีอายุดังนี้
14, 13, 19, 12, 17, 14 และ 16 ปี จงหา Q_1, Q_3 และ D_5

วิธีทำ ดำเนินตามขั้นตอนดังนี้

- ขั้นตอนที่ 1 เรียงลำดับข้อมูลจากน้อยไปหามากได้ ดังนี้
  12, 13, 14, 14, 17, 19
- ขั้นตอนที่ 2 หาตำแหน่งที่ต้องการ

ตำแหน่ง Q_1 คือ $\frac{1}{4}(7+1)=2$

ตำแหน่ง Q_3 คือ $\frac{3}{4}(7+1)=6$

ตำแหน่ง D_5 คือ $\frac{1}{10}(7+1)=4$

- ขั้นตอนที่ 3 คำนวณค่าในตำแหน่งที่ต้องการ

ค่าที่อยู่ในตำแหน่งที่ 2 ตรงกับ 13 พอดี ดังนั้น Q_1 =13 ปี
ค่าที่อยู่ในตำแหน่งที่ 6 ตรงกับ 17 พอดี ดังนั้น Q_3 =17 ปี
ค่าที่อยู่ในตำแหน่งที่ 4 ตรงกับ 14 พอดี ดังนั้น D_5 =14 ปี

ตัวอย่างการคำนวณ (๒)

ผลการชั่งการน้ำหนัก (หน่วยเป็นกิโลกรัม) ของนักเรียนชั้น ม. 5 ห้องหนึ่งจานวน 31 คน เป็นดังนี้

- 42 53 68 49 68 56 44 38 60 51 48 45 44 58 62 45 50 66 54 62 43 57 65 70 52 57 69 65 64 48 62

1. จงหาว่าน้ำหนักจะต้องตรงกับกี่กิโลกรัมจึงจะทำให้นักเรียนประมาณสามในสี่ของห้องมีน้ำหนักมากกว่า
2. จงหาว่าน้ำหนักจะต้องตรงกับกี่กิโลกรัมจึงจะทำให้นักเรียนประมาณหกในสิบของห้องมีน้ำหนักมากกว่า

- วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ได้ดังนี้
- 38 42 43 44 44 45 45 48 48 49 50 51 52 53 54 56 57 57 58 60 62 62 62 64 65 65 66 68 68 69 70

1. น้ำหนักที่มีนักเรียนประมาณสามในสี่ของห้องที่มีน้ำหนักมากกว่า แสดงว่าต้องการหาน้ำหนักที่มีนักเรียนประมาณหนึ่งในสี่ของห้องที่มีน้ำหนักน้อยกว่า นั่นคือ ต้องการหาน้ำหนักที่ตรงกับ P_{25} นั่นเอง

ตำแหน่ง $P_k = \frac{k(N+1)}{100}$

ดังนั้น $P_{25} = \frac{25(31+1)}{100} = 8$

น้ำหนักที่ตรงกับตำแหน่งที่ 8 คือ 48 กิโลกรัม

2. น้ำหนักที่มีนักเรียนประมาณหกในสิบของห้องที่มีน้ำหนักน้อยกว่า มีความหมายเช่นเดียวกับ น้ำหนักที่มีนักเรียนประมาณ 60 ใน 100 ของห้องที่มีน้ำหนักน้อยกว่า ซึ่งตรงกับน้ำหนักที่ P_{60}

น้ำหนักในลำดับที่ 19 และ 20 คือ 58 และ 60 ตามลำดับ
ตำแหน่งต่างกัน 20 - 19 = 1 น้ำหนักต่างกัน 60 - 58 = 2 กิโลกรัม
ตำแหน่งต่างกัน 19.2 - 19 = 0.2 น้ำหนักต่างกัน 2x0.2=0.4 กิโลกรัม
ดังนั้น P_{60} = 58 +0.4 = 58.4 กิโลกรัม

ตำแหน่ง $P_k = \frac{k(N+1)}{100}$

ดังนั้น $P_{60} = \frac{60(31+1)}{100} = 19.2$

ในกรณีที่ข้อมูลอยู่ในรูปของการแจกแจงความถี่แบบจัดกลุ่ม

เมื่อข้อมูลแจกแจงความถี่แล้ว สามารถดำเนินการตามขั้นตอนดังนี้

1. หาความถี่สะสม
2. หาตำแหน่งของข้อมูล โดยใช้สูตร

$$Q_k = \frac{kN}{4} \qquad D_k = \frac{kN}{10} \qquad \text{และ} \qquad P_k = \frac{kN}{100}$$

3. หาคำตอบโดยใช้สูตร

$$Q_k = L + I\left(\frac{\frac{kN}{4} - \sum f_L}{f_{me}}\right), \quad D_k = L + I\left(\frac{\frac{kN}{10} - \sum f_L}{f_{me}}\right) \quad \text{และ} \quad P_k = L + I\left(\frac{\frac{kN}{100} - \sum f_L}{f_{me}}\right)$$

เมื่อ L คือ ขอบล่างของชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไทล์

I คือ ความกว้างของอันตรภาคชั้น

$\sum f_L$ คือ ความถี่สะสมของชั้นที่ต่ำกว่าชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไ

f_{me} คือ ความถี่ของชั้นชั้นควอไทล์ เดไซล์ เปอร์เซนต์ไทล์

ในกรณีที่ข้อมูลอยู่ในรูปของการแจกแจงความถี่แบบจัดกลุ่ม

$$Q_X = L_0 + i\left[\frac{\frac{NX}{4} - F}{f}\right]$$

Q_x แทน ควอไทล์ที่ต้องการหา

L_0 แทน ขีดจำกัดล่างที่แท้จริงของชั้นคะแนนที่ควอไทล์นั้นอยู่

i แทน อันตรภาคชั้น

N แทน จำนวนคะแนนทั้งหมด

X แทน ตำแหน่งที่ของควอไทล์นั้น

F แทน ความถี่สะสมก่อนถึงชั้นคะแนนที่ควอไทล์นั้นอยู่

f แทน ความถี่ของชั้นคะแนนที่ควอไทล์นั้นอยู่

ตัวอย่างการหาความเบี่ยงเบนควอไทล์

คะแนน	f	cf
32-34	1	70
29-31	2	69
26-28	7	67
23-25	18	60
20-22	21	42
17-19	18	21
14-16	2	3
11-13	1	1

$$Q_1 = 16.5 + 3\left[\frac{\frac{70x1}{4} - 3}{18}\right] = 18.92$$

$$Q_3 = 22.5 + 3\left[\frac{\frac{70x3}{4} - 42}{18}\right] = 24.5$$

$$QD = \frac{24.5-18.92}{2} = 2.79$$

ที่มาของข้อมูล : ชูศรี วงศ์รัตนะ. 2544

3. ความเบี่ยงเบนเฉลี่ย (Mean deviation)

คือ ผลเฉลี่ยของความเบี่ยงเบนของคะแนนแต่ละตัวในข้อมูลชุดหนึ่งจากตัวกลางเลขคณิตของข้อมูลชุดนั้น

$$M.D = \frac{\sum_{i=1}^{N} \left| X_i - \bar{X} \right|}{N}$$

ข้อมูล : 5 7 9 10 15 ค่าเฉลี่ย 9.2

X	|X - X|
5	4.2
7	2.2
9	0.2
10	0.8
15	5.8
∑|X - X| = 13.2	

$$M.D = \frac{13.2}{5} = 2.64$$

Questions & Discussion