

บทนำ

สถิติสำหรับจิตวิทยา 1

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- คำศัพท์พื้นฐาน
- ระดับการวัด
- การสรุปข้อมูลด้วยรูปภาพ / ตาราง
- สถิติเชิงพรรณนา
- รูปแบบของข้อมูล
- การวัดแนวโน้มสู่ศูนย์กลาง (Central Tendency)
- การวัดการกระจาย (Variability)

คำศัพท์พื้นฐาน

- สถิติ (Statistics) เป็นสาขาหนึ่งของคณิตศาสตร์ที่สนใจการจัดกลุ่ม การวิเคราะห์ และการตีความหมายของกลุ่มตัวเลข แบ่งได้เป็น 2 ประเภท
 - สถิติเชิงพรรณนา (Descriptive Statistics)
 - สถิติเชิงอ้างอิง (Inferential Statistics) เป็นสถิติที่ใช้ในการประมาณการข้อมูลของกลุ่มที่ใหญ่กว่า จากข้อมูลจากกลุ่มที่เล็กกว่า
 - เช่น สุ่มกลุ่มตัวอย่างมา 1,000 คนจากประชากรไทย ไม่ดื่มสุรา 60% คนไม่ดื่มสุราในประเทศไทยคิดเป็นร้อยละเท่าไร

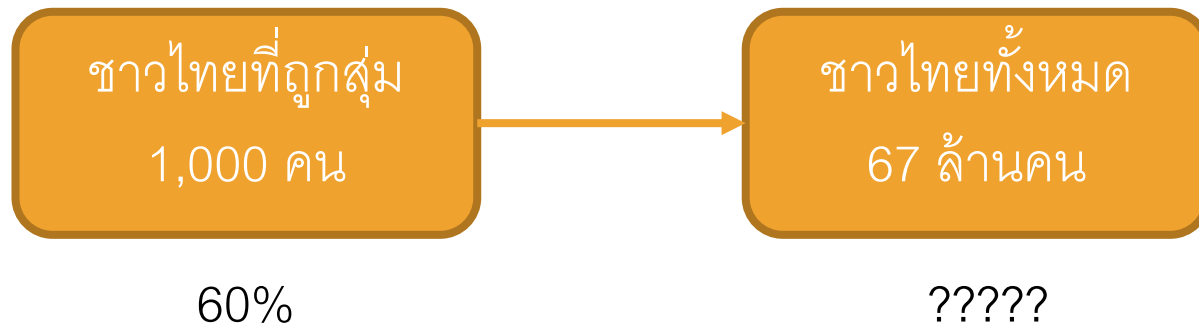
ขออนุญาตทับศัพท์ภาษาอังกฤษ เพราะคำ
แปลภาษาไทย ไม่มีใครใช้กัน

คำศัพท์พื้นฐาน

- Descriptive Statistics เป็นสถิติที่ใช้ในการสรุปกลุ่มของข้อมูล เช่น
 - อายุเฉลี่ยของกลุ่มนิสิตที่เรียนวิชาสถิติในคาบนี้
 - ร้อยละของชาวนาในประเทศไทย

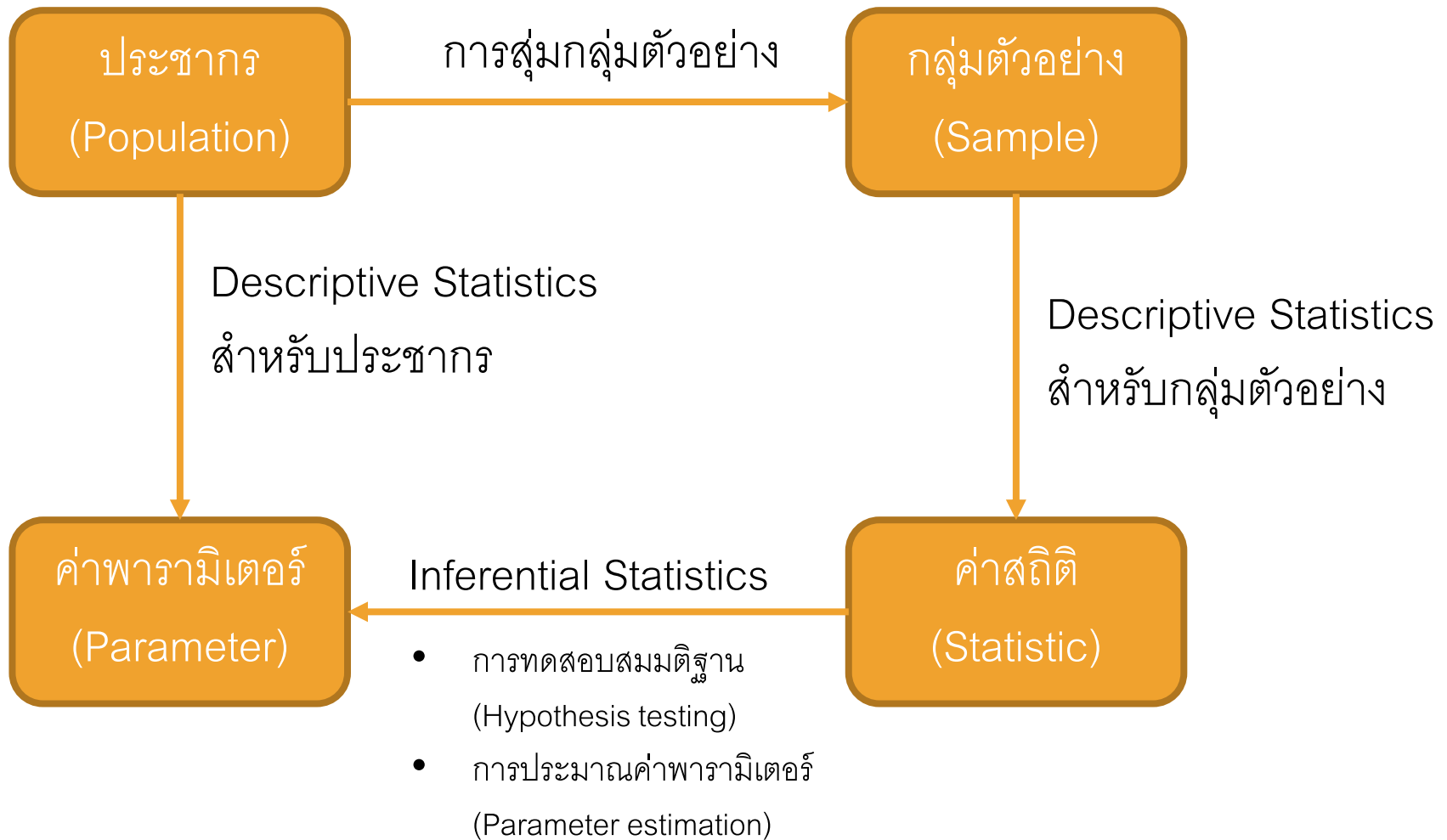
คำศัพท์พื้นฐาน

- Inferential Statistics เป็นสถิติที่ใช้ในการประมาณการข้อมูลของกลุ่มที่ใหญ่กว่า จากข้อมูลจากกลุ่มที่เล็กกว่า
 - เช่น สุ่มกลุ่มตัวอย่างมา 1,000 คนจากประชากรไทย ไม่ดื่มสุรา 60% คนไม่ดื่มสุราในประเทศไทยคิดเป็นร้อยละเท่าไร



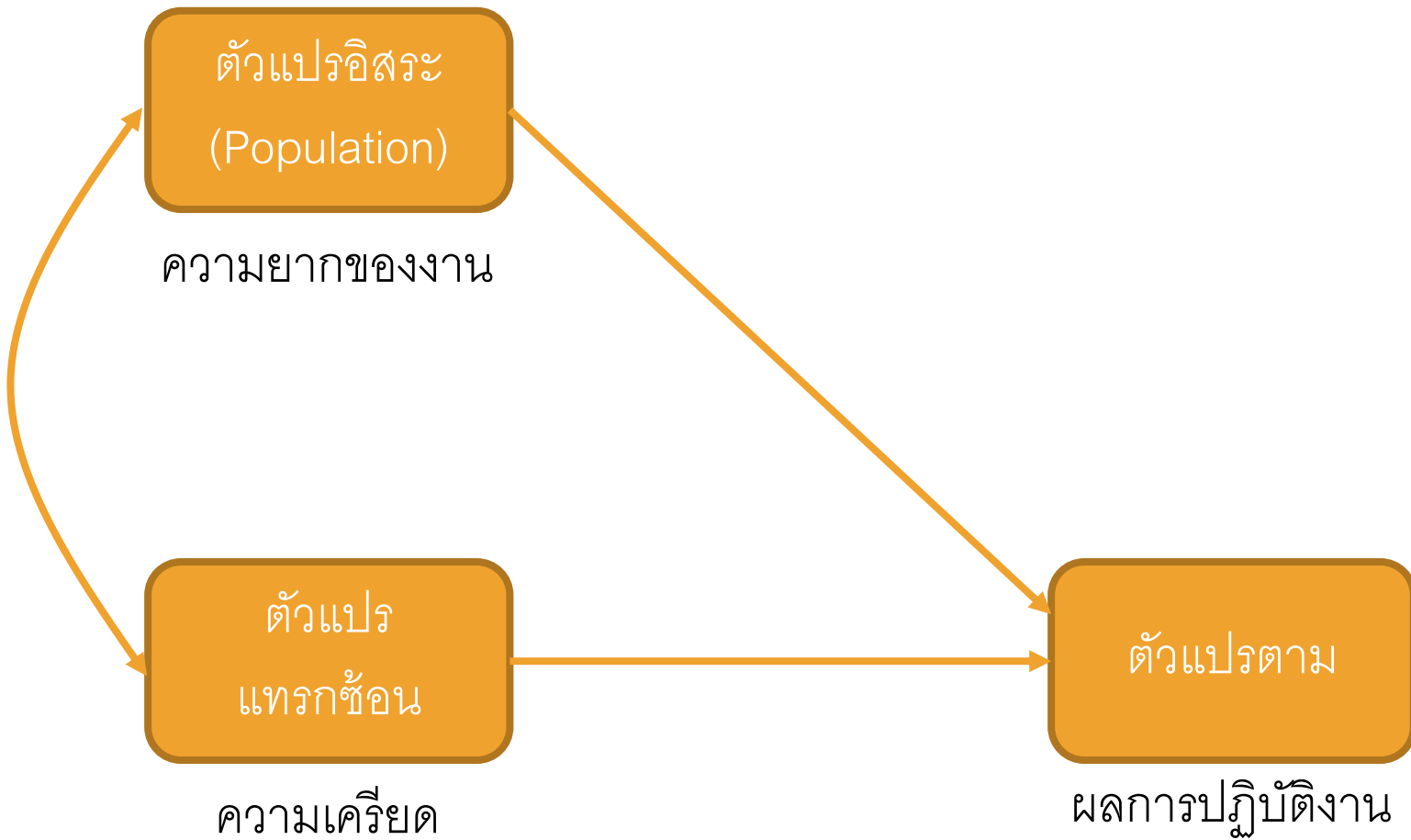
คำศัพท์พื้นฐาน

- ประชากร (Population) คือกลุ่มเป้าหมายที่เราสนใจ เช่น ชาวไทยทั้งหมด
- กลุ่มตัวอย่าง (Sample) คือหน่วยที่เราเลือกออกมาศึกษา เช่น ชายไทย 1,000 คนถูกสุ่มออกมา
- การสุ่มกลุ่มตัวอย่าง (Random Sampling) คือ วิธีการเลือกกลุ่มตัวอย่างจากประชากร ทุกๆ หน่วยมีโอกาสในการถูกเลือกเท่ากัน
- ค่าสถิติ (Statistic) คือ ผลการสรุปข้อมูลจากกลุ่มตัวอย่าง
- ค่าพารามิเตอร์ (Parameter) คือ ผลการสรุปข้อมูลจากประชากร



คำศัพท์พื้นฐาน

- ตัวแปรอิสระ (Independent variable) คือ ตัวแปรในงานวิจัย ที่ไม่ได้รับอิทธิพลจากตัวแปรอื่น
 - อาจถูกจัดกระทำขึ้นมา เช่น สุ่มคนเข้ากลุ่มเป็นกลุ่มทดลองและกลุ่มควบคุม
- ตัวแปรตาม (Dependent variable) คือ ตัวแปรในงานวิจัย ที่ได้รับอิทธิพลจากตัวแปรอื่น
- ตัวแปรแทรกซ้อน (Confounding/Extraneous variable) คือ ตัวแปรอื่นที่มีผลต่อตัวแปรตามนอกเหนือจากตัวแปรอิสระ
 - มีได้ แต่ต้องทำให้แน่ใจว่าตัวแปรแทรกซ้อนจะไม่ทำให้ข้อสรุปอิทธิพลของตัวแปรอิสระต่อตัวแปรตามผิดพลาด



ระดับการวัด (Level of Measurement)

- นามบัญญัติ (Nominal scale)
- จัดอันดับ (Ordinal scale)
- อินตรภาค (Interval scale)
- อัตราส่วน (Ratio scale)

ขออนุญาตทับศัพท์ภาษาอังกฤษ เพราะคำ
แปลภาษาไทย ไม่มีใครใช้กัน

ระดับการวัด (Level of Measurement)

- Nominal scale; Nominal variable; Unordered categorical variable
 - จัดกลุ่ม
 - ไม่มีกลุ่มใดเหนือกว่ากลุ่มใด
 - เช่น เพศ (ชาย/หญิง) ศาสนา (พุทธ/คริสต์/อิสลาม/อื่นๆ)
- ตัวแปรทวินาม (Dichotomous variable) มีการจัดแบ่งแค่สองประเภท
- ตัวแปรพหุนาม (Polytomous variable) มีการจัดแบ่งมากกว่าสองประเภท

ระดับการวัด (Level of Measurement)

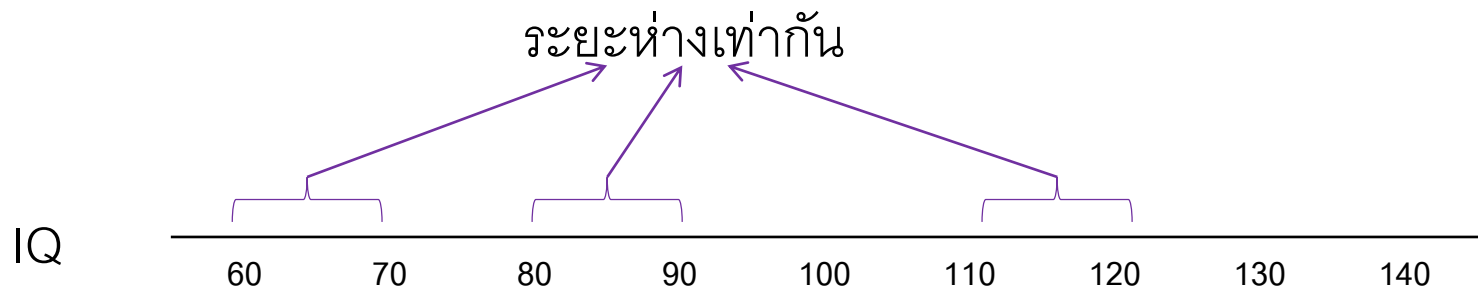
- Ordinal scale; Rank variable; Ordered categorical variable
 - จัดกลุ่มหรือไม่จัดกลุ่มก็ได้
 - มีอันดับชัดเจน
 - อันดับต่างๆ อาจมีความแตกต่างในระดับความสามารถไม่เท่ากัน
 - เช่น สถานะทางเศรษฐกิจและสังคม (สูง/กลาง/ต่ำ)
 - อันดับที่อยู่ในชั้นเรียน



ระดับการวัด (Level of Measurement)

- Interval scale

- มีลักษณะเป็นตัวเลข
- ระยะห่างระหว่างตัวเลขมีค่าเท่ากันตลอดมาตราบ
- เช่น คะแนนสอบเลข (0-100)
- IQ (20-180)
- ค่าศูนย์ไม่มีความหมาย (IQ = 0 ก็ยังมีความสามารถในการคิดอยู่)



ระดับการวัด (Level of Measurement)

- Ratio scale
 - มีลักษณะเป็นตัวเลข
 - ระยะห่างระหว่างตัวเลขเท่ากันตลอดสเกล
 - มีค่าศูนย์ที่แท้จริง
 - เช่น ความสูง (0 cm = ไม่มีความสูงเลย)
 - จำนวนบุตร (0 คน = ไม่มีสักคน)
 - สามารถบอกเป็นเท่าได้ เช่น สูง 100 cm เป็น 2 เท่าของสูง 50 cm

Descriptive Statistics

- เทคนิคทางสถิติที่ใช้ในการรวบรวมข้อมูล สามารถแสดงผลเป็น 2 รูปแบบ คือ
 - การแสดงผลเป็นรูปภาพ เช่น ภาพ แผนภูมิ ตาราง
 - การแสดงผลเป็นตัวเลข เช่น ค่าเฉลี่ย ร้อยละ

Descriptive Statistics

- การแสดงผลเป็นรูปภาพ
 - ตารางแจกแจงความถี่ (Frequency Distribution)
 - ฮิสโทแกรม (Histogram) / แผนภูมิแท่ง (Bar graph)
 - แผนภูมิต้นใบ (Stem and Leaf Display)

ตารางแจกแจงความถี่ (Frequency Distribution)

- ข้อมูล → ตาราง
- ความถี่ (Frequency; f)

1, 1, 1, 1, 2, 2, 2, 2, 2,
3, 3, 3, 3, 4, 4, 4, 5, 5,
5, 6



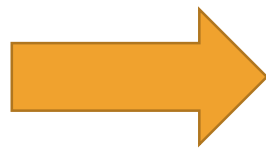
ข้อมูล	f
1	4
2	5
3	4
4	3
5	3
6	1

ตารางแจกแจงความถี่ (Frequency Distribution)

- ให้ N เป็นจำนวนกลุ่มตัวอย่าง
- สัดส่วน (Proportion; p) = f / N ; ร้อยละ (Percentage) = $p \times 100\%$
- ความถี่สะสม (Cumulative Frequency) คือ จำนวนความถี่ของคะแนนที่เท่ากันหรือน้อยกว่า
- ร้อยละสะสม (Cumulative Percentage) หรือ Percentile คือจำนวนร้อยละของคะแนนที่เท่ากันหรือน้อยกว่า

คะแนน	<i>f</i>	<i>p</i>	%	Cum. <i>f</i>	Cum. %
1	4	.20	20	4	20
2	5	.25	25	9	45
3	4	.20	20	13	65
4	3	.15	15	16	80
5	3	.15	15	19	95
6	1	.05	5	20	100

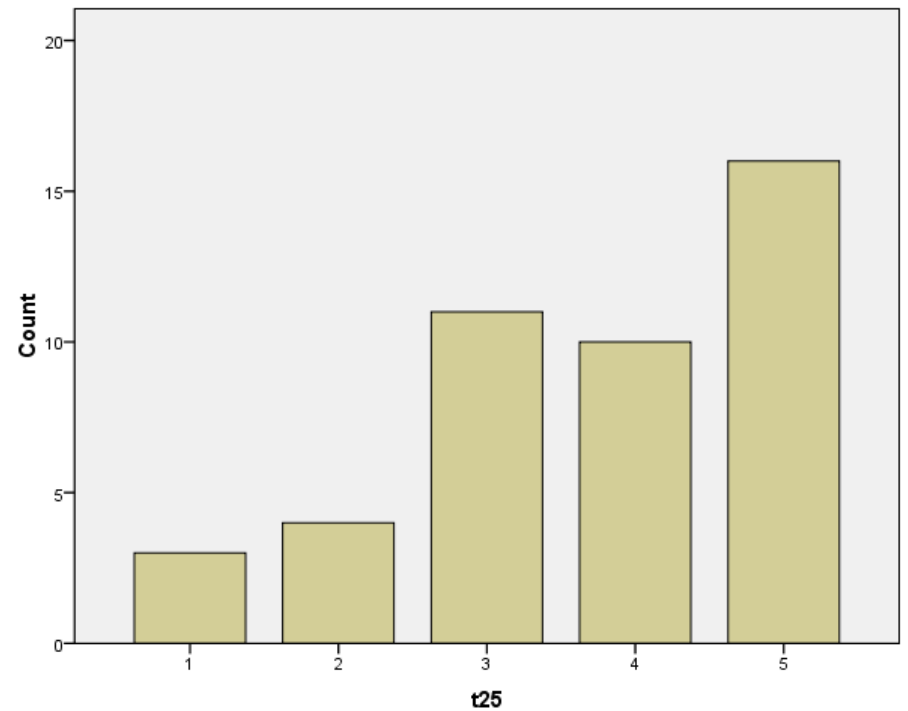
ข้อมูล	f
1	4
2	5
3	4
4	3
5	3
6	1



ข้อมูล	f
1-2	9
3-4	7
5-6	4

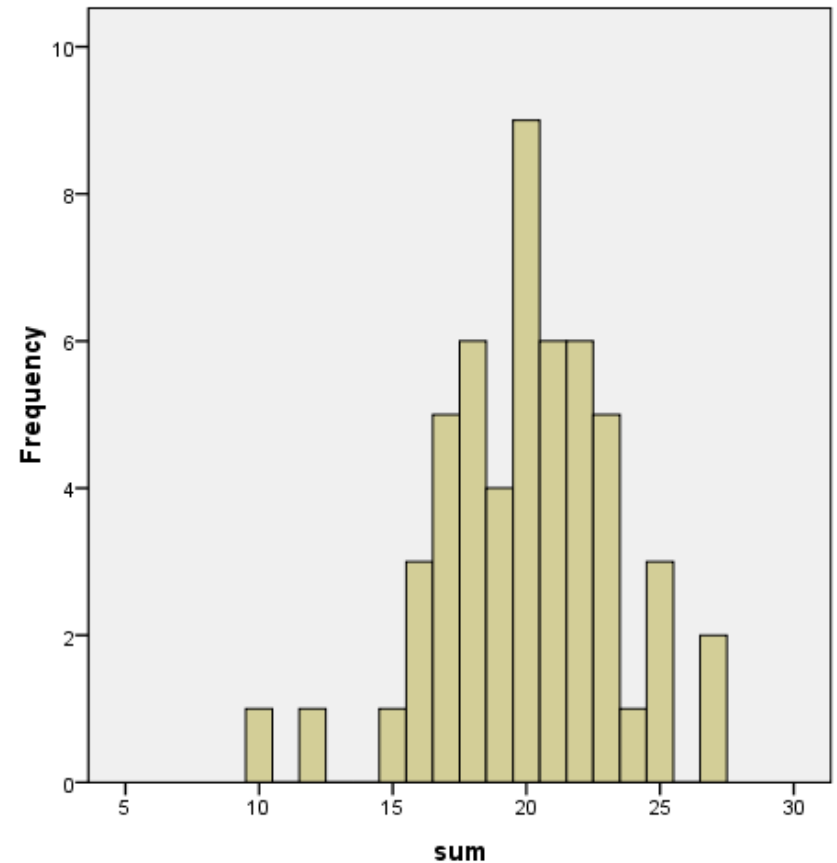
แผนภูมิแท่ง (Bar Graph)

- ใช้กับ Categorical variable (Nominal/Ordinal)
- สังเกตว่าแท่งจะไม่ติดกัน



Histogram

- ใช้กับ Continuous variable (Interval scale/ratio scale)
- สังเกตว่าแท่งจะติดกัน แสดงให้เห็นถึงความต่อเนื่องของตัวแปร



Stem and Leaf Display

- ลักษณะเหมือน Histogram แต่
แสดงข้อมูลดิบ (Raw data) ด้วย

```
Frequency      Stem & Leaf
              1.00 Extremes      (= <10)
                .00          1 .
                1.00          1 . 2
                1.00          1 . 5
                8.00          1 . 66677777
               10.00          1 . 8888889999
               15.00          2 . 000000000111111
               11.00          2 . 22222233333
                4.00          2 . 4555
                2.00          2 . 77

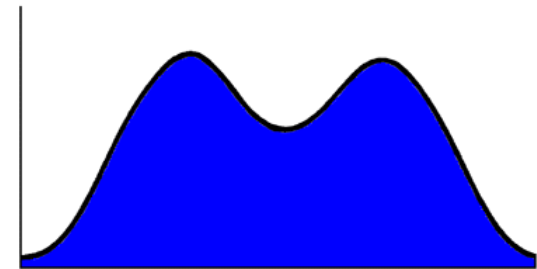
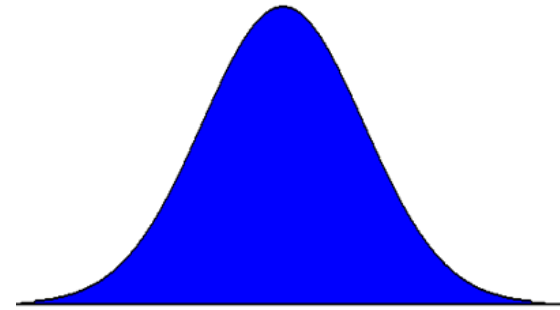
Stem width:           10
Each leaf:           1 case(s)
```

รูปแบบของข้อมูล

- จำนวนฐานนิยม (Modality)
- ความเบ้ (Skewness)
- ความโด่ง (Kurtosis)

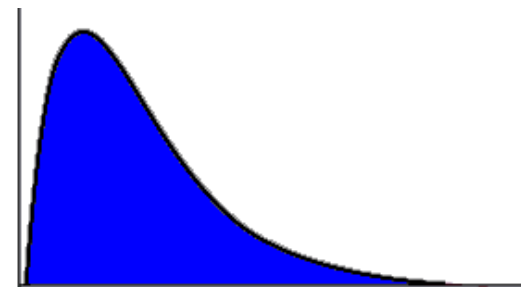
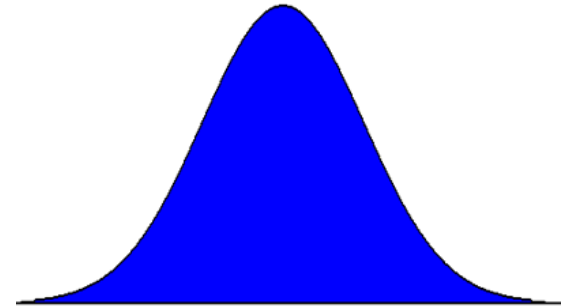
รูปแบบของข้อมูล

- จำนวนฐานนิยม (Modality)
 - ฐานนิยมเดียว (Unimodal)
 - ฐานนิยมสองอัน (Bimodal)



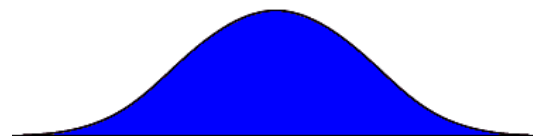
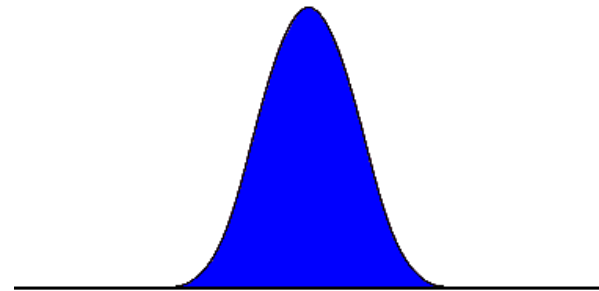
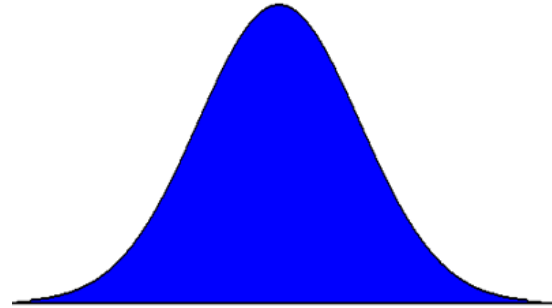
รูปแบบของข้อมูล

- ความเบ้ (Skewness)
 - สมมาตร (ไม่เบ้)
 - เบ้ซ้าย (Negatively skewed)
 - Ceiling effect
 - เบ้ขวา (Positively skewed)
 - Floor effect



รูปแบบของข้อมูล

- ความโด่ง (Kurtosis)
 - ความโด่งเท่าโค้งปกติ (Normal Distribution)
 - ความโด่งมากกว่าโค้งปกติ (Kurtosis > 0)
 - กลางโด่ง หางแบน
 - ความโด่งน้อยกว่าโค้งปกติ (Kurtosis < 0)
 - กลางแบน เป็นรูป U คว่ำ



การแสดงผลเป็นตัวเลข

- การวัดแนวโน้มสู่ศูนย์กลาง (Central Tendency)
- การวัดการกระจาย (Variability)

การวัดแนวโน้มสู่ศูนย์กลาง

- การวัดแนวโน้มสู่ศูนย์กลาง เป็นการหาค่ากลางของกลุ่มคะแนน ในที่นี้จะพูดถึง 3 ชนิด
 - ค่าเฉลี่ย (Mean)
 - มัธยฐาน (Median)
 - ฐานนิยม (Mode)

ค่าเฉลี่ย

- ค่าเฉลี่ย (Mean, Average; M , \bar{X}) คือการนำคะแนนรวมมาหาด้วยจำนวนกลุ่มตัวอย่าง
 - ให้ X_i เป็นคะแนนของตัวแปรที่สนใจ ของหน่วย i
 - ผลรวมของคะแนนทุกหน่วยตั้งแต่ 1 ถึง N แทนค่าด้วย $\sum_{i=1}^N X_i$

$$M = \frac{\sum_{i=1}^N X_i}{N} = \frac{\Sigma X}{N}$$

ค่าเฉลี่ย

- เช่น จำนวนบุตรเฉลี่ย

0, 2, 1, 2, 1, 1, 3, 0, 1, 2

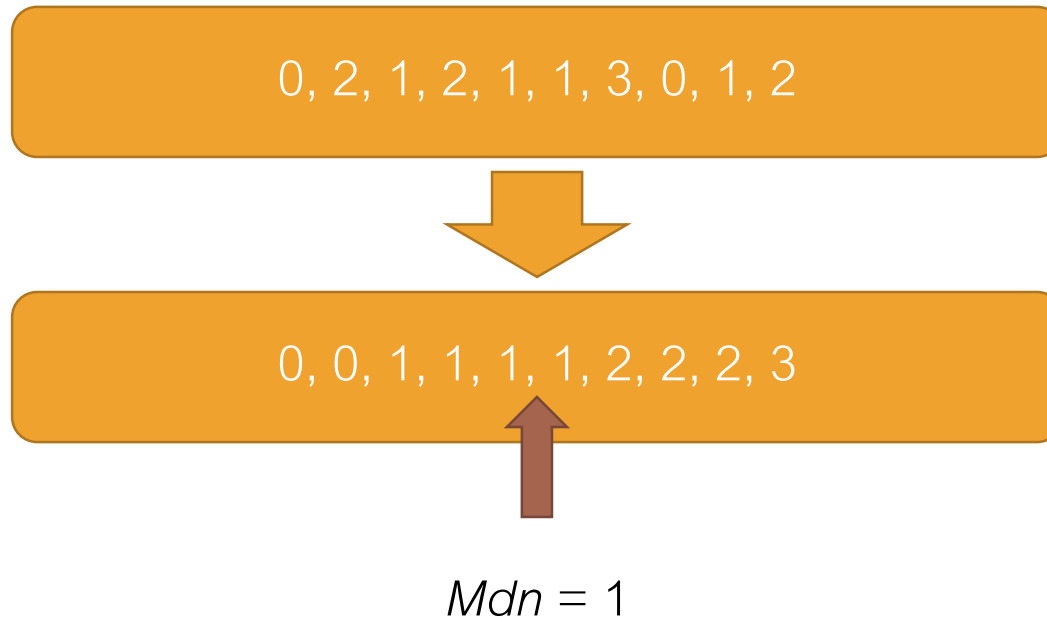
$N = 10$

$$M = (0 + 2 + 1 + 2 + 1 + 1 + 3 + 0 + 1 + 2) / 10 = 13 / 10$$

$$M = 1.3$$

มัธยฐาน

- มัธยฐาน (Median; *Mdn*) คือ ค่าที่อยู่ตำแหน่งตรงกลางของข้อมูล



ฐานนิยม

- ฐานนิยม (Mode) คือ ค่าที่มีความถี่มากที่สุด

0, 2, 1, 2, 1, 1, 3, 0, 1, 2



0, 0, 1, 1, 1, 1, 2, 2, 2, 3



Mode = 1

ปัจจัยที่ผลกระทบต่อการวัดแนวโน้มสู่ศูนย์กลาง

- ค่าสุดโต่ง (Outlier, Extreme Score)

0, 2, 1, 2, 1, 1, 3, 0, 1, 2



0, 0, 1, 1, 1, 1, 2, 2, 2, 3

$$M = 1.3$$

$$Mdn = 1$$

$$Mode = 1$$

0, 2, 1, 2, 1, 1, 3, 0, 1, 20



0, 0, 1, 1, 1, 1, 2, 2, 3, 20

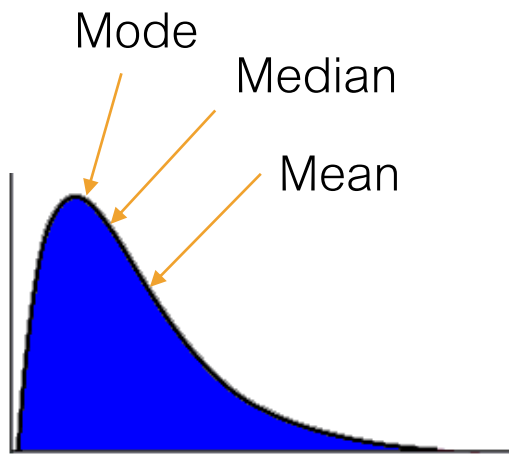
$$M = 3.1$$

$$Mdn = 1$$

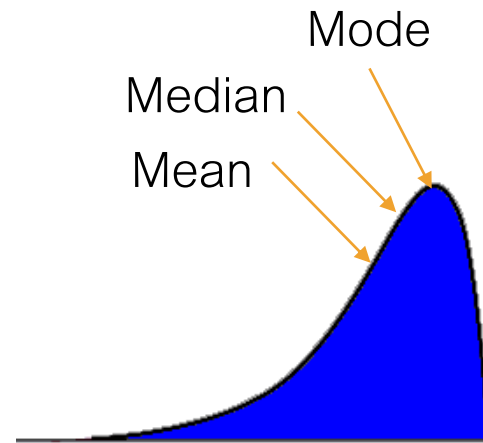
$$Mode = 1$$

ปัจจัยที่ผลกระทบท่อการวัดแนวโน้มสู่ศูนย์กลาง

- การกระจายไม่เป็นโค้งปกติ



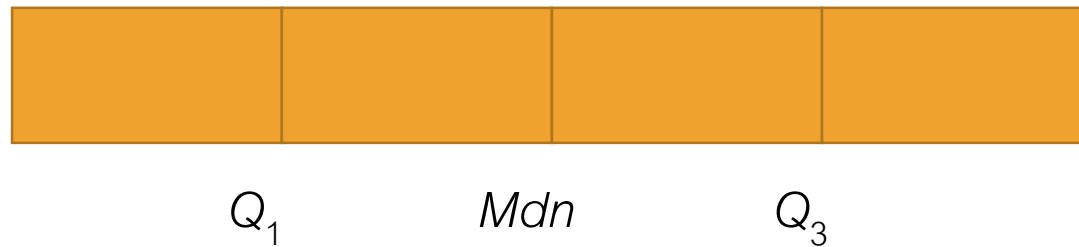
Positively skewed



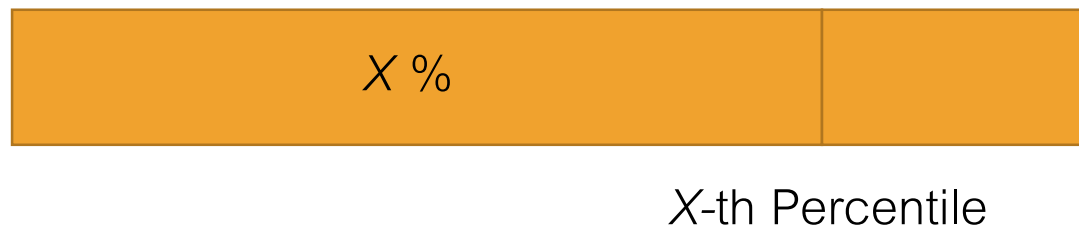
Negatively skewed

การวัดตำแหน่งของข้อมูล (Centile)

- Quartile



- Percentile



การกระจาย

- การกระจายของข้อมูล (Variability) เป็นการวัดขนาดความแตกต่างภายในกลุ่ม
- ตัวแปร (Variable) จะมีการกระจาย แต่ทว่า ค่าคงที่ (Constant) ไม่มีการกระจาย
- ในที่นี้จะพูดถึง
 - ความแปรปรวน (Variance)
 - ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)
 - พิสัย (Range)
 - พิสัยระหว่าง Quartile (Interquartile range)

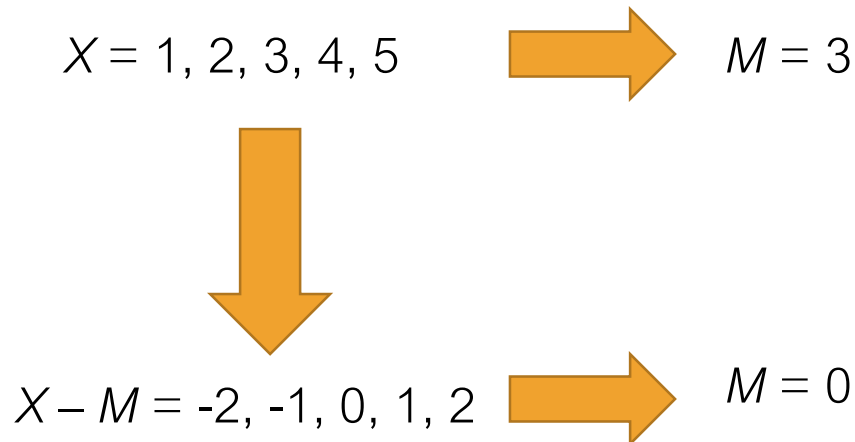
ความแปรปรวน (Variance)

- ความแปรปรวน (Variance; s^2) คือ ค่าเบี่ยงเบนกำลังสองเฉลี่ย
 - ก่อนเข้าใจความแปรปรวนต้องพูดถึงค่าเบี่ยงเบน (Deviation Score)
 - ค่าเบี่ยงเบน คือ ความแตกต่างข้อมูลกับค่าเฉลี่ย
 - เช่น

$$\begin{array}{ccc} X = 1, 2, 3, 4, 5 & \longrightarrow & M = 3 \\ & \downarrow & \\ X - M = -2, -1, 0, 1, 2 & & \end{array}$$

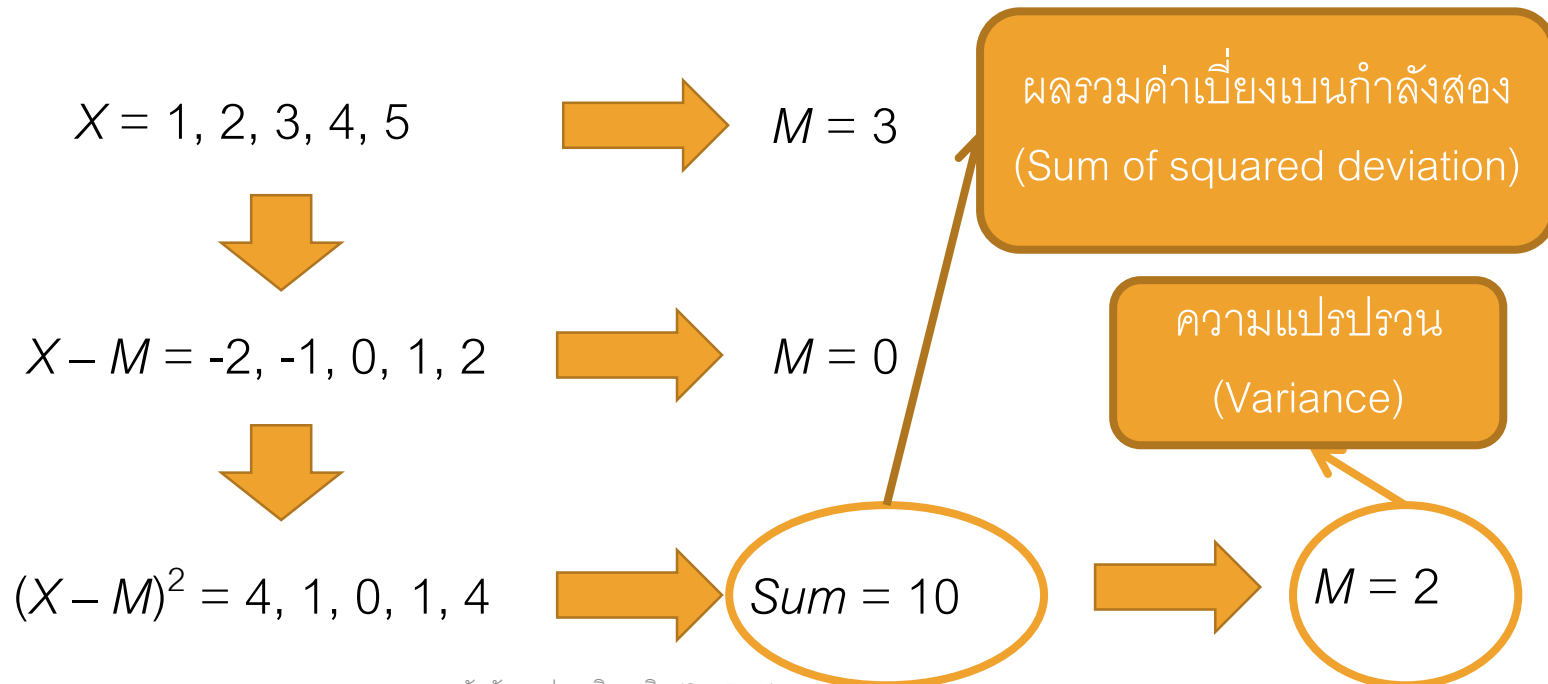
ความแปรปรวน (Variance)

- ความแปรปรวน (Variance; s^2) คือ ค่าเบี่ยงเบนกำลังสองเฉลี่ย
 - หากหาค่าเฉลี่ยของค่าเบี่ยงเบนนี้ได้ น่าจะวัดการกระจายได้
 - แต่ค่าเฉลี่ยของค่าเบี่ยงเบนทั้งหมดจะเท่ากับศูนย์



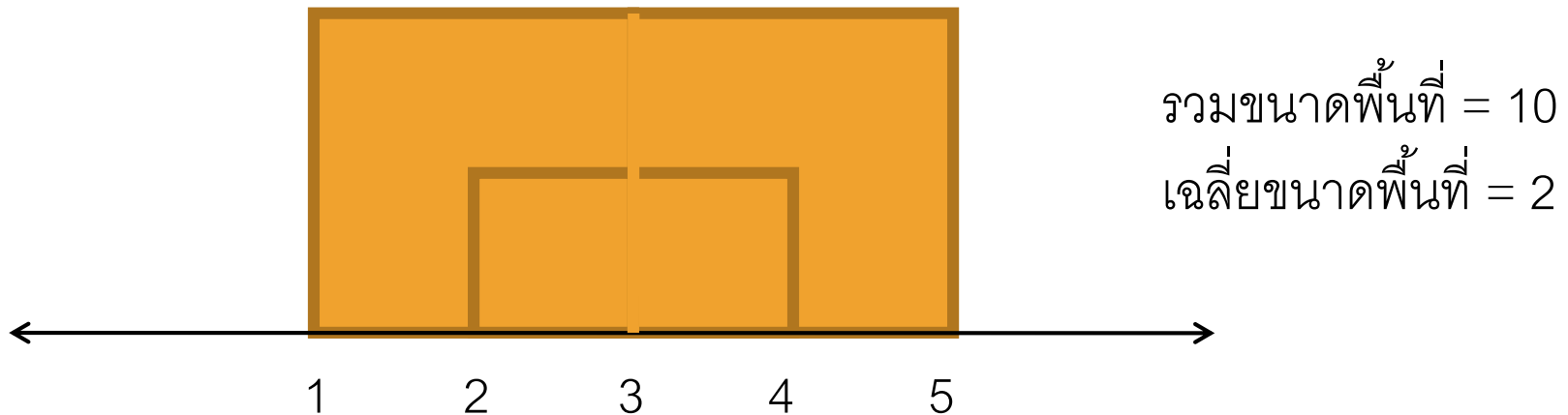
ความแปรปรวน (Variance)

- ความแปรปรวน (Variance; s^2) คือ ค่าเบี่ยงเบนกำลังสองเฉลี่ย
 - หากนำค่าเบี่ยงเบนมายกกำลังสอง จะทำให้ค่าติดลบหายไป
 - หากนำค่าเบี่ยงเบนที่ยกกำลังสองแล้วมาหาค่าเฉลี่ย จะได้ความแปรปรวน



ความแปรปรวน (Variance)

- ความแปรปรวน (Variance; s^2) คือ ค่าเบี่ยงเบนกำลังสองเฉลี่ย
 - หากพิจารณาเป็นภาพ จะเป็นดังนี้



- ด้วยเหตุนี้ความแปรปรวนจึงอาจเรียกว่าเป็นการกระจายรูปพื้นที่

ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation)

- ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation; SD) คือ รากที่สองของความแปรปรวน
 - เป็นการนำการกระจายรูปพื้นที่ กลับมาเป็นเชิงเส้นเช่นเดิม

$$\begin{array}{l} X = 1, 2, 3, 4, 5 \\ \downarrow \\ X - M = -2, -1, 0, 1, 2 \\ \downarrow \\ (X - M)^2 = 4, 1, 0, 1, 4 \\ \rightarrow \sum (X - M)^2 / N = 2 \\ \rightarrow SD = 1.414 \end{array}$$
$$SD = \sqrt{s^2} = \sqrt{\frac{\sum (X - M)^2}{N}}$$

พิสัย (Range)

- พิสัย (Range) คือ ค่าสูงสุดลบด้วยค่าต่ำสุด

$$\text{Range} = \text{Max} - \text{Min}$$

0, 2, 1, 2, 1, 1, 3, 0, 1, 2



Range = 3

พิสัยระหว่าง Quartile (Interquartile Range)

- ค่าพิสัย จะได้รับอิทธิพลจากค่าสุดโต่งมาก
- จึงน่าจะใช้ควอร์ไทล์ที่ 1 และ 3 บอกพิสัยแทน จึงเรียกว่า พิสัยระหว่างควอร์ไทล์ (Interquartile Range; *IQR*)

$$IQR = Q_3 - Q_1$$

0, 2, 1, 2, 1, 1, 3, 0, 1, 2



$$Q_1 = 0.75$$

$$Q_3 = 2.00$$

$$IQR = 1.25$$

ปัจจัยที่ส่งผลกระทบต่อการวัดการกระจาย

- ค่าสุดโต่งมีผลกระทบต่อพิสัย, ความแปรปรวน, และส่วนเบี่ยงเบนมาตรฐาน

1, 2, 3, 4, 5, 6

Range = 4

IQR = 1.5

Variance = 1.67

SD = 1.29

1, 2, 3, 4, 5, 17

Range = 16

IQR = 1.5

Variance = 29.67

SD = 5.45