บทที่ 2
สถิติพรรณนา
(Descriptive Statistics)

2.1 การวัดแนวโน้มสู่ส่วนกลาง (Measures of Central Tendency)

การวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูลนั้น เป็นการหาตัวแทนของข้อมูล โดยคำนวณบนพื้นฐานขอบเขตของข้อมูล ซึ่งตัวแทนของข้อมูลดังกล่าว สามารถทำได้หลายวิธี ซึ่งจะกล่าวถึงบางส่วนดังนี้

- ค่าเฉลี่ยเลขคณิต (Arithematic Mean)
- มัธยฐาน (Median)
- ฐานนิยม (Mode)

2.1.1. ค่าเฉลี่ยเลขคณิต (Arithematic Mean) เป็นค่ากลางของข้อมูลชนิดหนึ่งที่นิยมนำมาเป็นตัวแทนของข้อมูล

1) ข้อมูลไม่แจกแจงความถี่

$$\text{mean}=\frac{\sum_{i=1}^{n} X_i}{N}$$

2) ข้อมูลแจกแจงความถี่

$$\text{mean} = \frac{\sum_{i=1}^{n} f_i X_i}{N}$$

ตัวอย่างที่ 2.1. ในการทดลองยาชนิดหนึ่งโดยให้กระต่าย 5 ตัว ทดลองกินยา พบว่ากระต่ายจะตายหลังกินยาเข้าไปเป็นเวลาดังนี้ 44, 27, 36, 44, 44 (นาที) จงคำนวณเวลาเฉลี่ยที่กระต่ายกินยาเข้าไปแล้วตาย

วิธีทำ

$$\text{mean} = \frac{\sum_{i=1}^{n} X_i}{N}$$

$$= \frac{44+27+36+44+44}{5}$$

$$= 39$$

ตัวอย่างที่ 2.2. ผลการสอบวิชาคณิตศาสตร์ของนักเรียน 20 คน เป็นดังนี้

คะแนนที่สอบได้	60	65	70	75	80	85
จำนวน นร.	1	6	5	3	3	2

จงหาคะแนนเฉลี่ยเลขคณิตของคะแนนสอบครั้งนี้

วิธีทำ

คะแนนที่สอบได้ (x_i)	60	65	70	75	80	85
จำนวน นร. (f_i)	1	6	5	3	3	2
f_iX_i	60	390	350	225	240	170

$$\text{mean} = \frac{\sum_{i=1}^{n} f_iX_i}{N}$$

$$= \frac{60+390+350+225+240+170}{1+6+5+3+3+2} = \frac{1,435}{20} = 71.75$$

$$= 71.75$$

แบบฝึกหัดที่ 2

1. ในการทดลองยาชนิดหนึ่งโดยให้กระต่าย 10 ตัว ทดลองกินยา พบว่ากระต่ายจะตายหลังกินยาเข้าไปเป็นเวลาดังนี้ 4, 2, 2, 3, 6, 4, 4, 2, 2, 6 (นาที) จงหาคำตอบ

1.1 $\sum_{i=1}^{n=10} X_i = 4+2+2+\cdots+6 = 35$

1.2 $\sum_{i=1}^{n=5} X_i = 4+2+2+3+6 = 17$

1.3 $\sum_{i=1}^{n=5} (X_i - 3) = (4-3)+(2-3)+(2-3)+(3-3)+(6-3) = 2$

1.4 $\sum_{i=1}^{n=5} X_i - 3 = (4+2+2+3+6) - 3 = 17-3 = 14$

1.5 $\sum_{i=1}^{n=5} X_i^2 = 4^2+2^2+2^2+3^2+6^2 = 69$

1.6 $\sum_{i=1}^{n=5} (X_i - 1)^2 = (4-1)^2+(2-1)^2+(2-1)^2+(3-1)^2+(6-1)^2 = 40$

1.7 $\sum_{i=1}^{n=5} X_i^2 - 1 = 69-1 = 68$

2. จงคำนวณ ค่าเฉลี่ย

2.1 ข้อมูล คือ 2, 5, 7, 9, 3, 1, 9, 7, 4, 3, 5, 1, 7

วิธีทำ ค่าเฉลี่ย ขั้นที่ 1 สูตรค่าเฉลี่ยคือ = $\mu = \sum_{i=1}^{N} x_i / N$

ขั้นที่ 2 แทนค่า

ค่าเฉลี่ย = $(2+5+\dots+7)/13$

= $63/13$

= 4.8461

ค่าตอบทศนิยม 3 ตำแหน่ง = 4.846

" 2 " = 4.85

" 1 " = 4.8

2.2 **ข้อมูล คือ** 1, 1, 9, 7, 1, 5, 9, 4, 7, 1, 9, 6, 5

วิธีทำ ค่าเฉลี่ย ขั้นที่ 1 สูตรค่าเฉลี่ยคือ = $\mu = \sum_{i=1}^{N} x_i / N$

ขั้นที่ 2 แทนค่า

ค่าเฉลี่ย = $(1+1+ \ldots + 5)/13$

= $65/13$

= 5

2.3 **ข้อมูล คือ** 7, 1, 4, 6, 7, 8, 2, 4, 6, 7

วิธีทำ ค่าเฉลี่ย ขั้นที่ 1 สูตรค่าเฉลี่ยคือ = $\mu = \sum_{i=1}^{N} X_i / N$

ขั้นที่ 2 แทนค่า

ค่าเฉลี่ย = $52/10$

= 5.2

! ถ้าข้อมูลที่กำหนดให้เป็นตัวอย่าง

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

$= 5.2$,

ค่าเท่ากัน แต่ความหมายต่างกัน !

2.2. การวัดการกระจายของข้อมูล (Measure of Dispersion)

เป็นสถิติประเภทหนึ่งที่ใช้อธิบายลักษณะการกระจายของข้อมูล การที่ข้อมูลชุดหนึ่งๆ ประกอบด้วยคะแนนที่มีค่าแตกต่างกันโดยถือว่า เป็นข้อมูลที่มีการกระจาย ถ้าข้อมูลชุดนั้น ประกอบด้วยคะแนนที่มีค่าต่างกันมาก เรียกว่า ข้อมูลที่มีกระจายมาก ถ้าข้อมูลชุดนั้นประกอบด้วย คะแนนที่มีค่าต่างกันน้อย เรียกว่า ข้อมูลที่มีการกระจายน้อย และถ้าข้อมูลชุดนั้นประกอบด้วย คะแนนที่มีค่าเท่ากันหมด เรียกว่า เป็นข้อมูลที่ไม่มีการกระจาย การใช้เฉพาะค่ากลางอย่างเดียวได้จึง ไม่เพียงพอที่จะอธิบายข้อมูลได้อย่างชัดเจน เช่น

ข้อมูลชุดที่ 1	10, 10, 10, 10, 10
ข้อมูลชุดที่ 2	9, 10, 9, 11, 11
ข้อมูลชุดที่ 3	9, 8, 11, 12, 10

ค่าเฉลี่ยทั้งสามชุดเท่ากับ 10 แต่ข้อมูลแต่ละชุดมีลักษณะต่างกัน
ข้อมูลชุดที่ 1 มีค่าเท่ากัน แสดงว่าไม่มีการกระจายเลย
ข้อมูลชุดที่ 1 มีการกระจายน้อยกว่าชุดที่ 2 และ ชุดที่ 3
ข้อมูลชุดที่ 2 มีการกระจายน้อยกว่าชุดที่ 3

การวัดการกระจายนิยมใช้ควบคู่กับการวัดแนวโน้มเข้าสู่ส่วนกลาง เพราะจะช่วยอธิบายลักษณะของข้อมูลได้ชัดเจนขึ้น ทั้งนี้ เนื่องจากการวัดแนวโน้มเข้าสู่กลางเป็นเพียงการบอกค่ากลางของข้อมูลชุดนั้น แต่ก็ยังไม่ทราบชัดเจนถึงลักษณะการกระจายของข้อมูลว่าคะแนนต่างๆ ในชุดข้อมูลนั้นมีค่าใกล้เคียงกัน หรือแตกต่างกันมาก ถ้าเรามีทั้งค่าแนวโน้มเข้าสู่ส่วนกลางและค่าการกระจายก็จะทำให้เข้าใจลักษณะข้อมูลนั้นได้ชัดเจนขึ้นมากกว่ามีแต่ค่าแนวโน้มเข้าสู่ส่วนกลางเพียงอย่างเดียว

2.2.2. ความแปรปรวน (Variance)

ค่าที่ใช้วัดการกระจายของข้อมูล เป็นการวัดความผันแปรเฉลี่ยของข้อมูลที่อธิบายว่าข้อมูลผันแปรจากค่าเฉลี่ยมากน้อยเพียงใด

1) กรณีของข้อมูลไม่แจกแจงความถี่

ประชากร	ตัวอย่าง
$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}$ หรือ $\sigma^2 = \frac{\sum_{i=1}^{N}X_i^2}{N}-\mu^2$	$S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$ หรือ $S^2 = \frac{\sum_{i=1}^{n}X_i^2}{n-1}-\frac{n\bar{X}^2}{n-1}$

2.2.3. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation; S.D.) คือ รากที่ 2 ของค่าความแปรปรวน

ประชากร $\sigma = \sqrt{\sigma^2}$

ตัวอย่าง $S = \sqrt{S^2}$

ตัวอย่างที่ 2.10. ข้อมูลชุดหนึ่งประกอบด้วย 13, 10, 12, 16, 13, 14 จงหาค่าความแปรปรวน และส่วน เบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

วิธีทำ ให้ σ^2 เป็นค่าความแปรปรวนของข้อมูลชุดนี้

หาค่า μ จากสูตร $\mu = \dfrac{\sum X_i}{N}$

$$\mu = \frac{13+10+12+16+13+14}{6}$$

$$= 13$$

หาค่า σ^2 จากตาราง

ข้อมูล	13	10	12	16	13	14	รวม
$x_i - \mu$	0	-3	-1	3	0	1	0
$(x_i-\mu)^2$	0	9	1	9	0	1	20

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}$$

$$= \frac{0+9+1+9+0+1}{6}$$

$$= \frac{20}{6} = 3.33$$

หรือ

$$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N} - \mu^2$$

$$= \frac{13^2 + 10^2 + 12^2 + 16^2 + 13^2 + 14^2}{5} - 13^2$$

$$= 172.33 - 169$$

$$= 3.33$$

และ $\sigma_{.} = \sqrt{3.3} = 1.83$

นั่นคือ ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ เท่ากับ 3.33 และ 1.83

2.2.4. สัมประสิทธิ์ความผันแปร (Coefficient of Variation; C.V.)

การเปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 ชุด ขึ้นไป โดยหน่วยของข้อมูลที่นำมาทำการเปรียบเทียบหรือค่าเฉลี่ยแตกต่างกันสามารถคำนวณได้ดังนี้

$$\text{C.V.}_{\text{ประชากร}} = \frac{\sigma}{\mu}$$

$$\text{C.V.}_{\text{ตัวอย่าง}} = \frac{S}{\overline{X}}$$

ตัวอย่างที่ 2.12. ถ้านาย ก. จะต้องตัดสินใจเลือกซื้อหุ้นบริษัทใดบริษัทหนึ่ง โดยบริษัทที่มีอัตราเงินปันผลดังนี้

บริษัท A เงินปันผลเฉลี่ย 15.6 บาทต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 3.7 บาท

บริษัท B เงินปันผลเฉลี่ย 13.7 บาทต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 2.5 บาท

บริษัท C เงินปันผลเฉลี่ย 18.9 บาทต่อปี และค่าส่วนเบี่ยงเบนมาตรฐาน 5.8 บาท

ถ้าท่านเป็นนาย ก. ท่านจะตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัทใด

วิธีทำ โจทย์ต้องการทราบว่านาย ก. จะตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัทใด สามารถคำนวณได้จากค่าของ C.V.จากสูตร $\mathrm{C.V.} = \dfrac{\sigma}{\mu}$

โจทย์กำหนด
$$\mu_A = 15.6, \sigma_A = 3.7$$
$$\mu_B = 13.7, \sigma_B = 2.5$$
$$\mu_C = 18.9, \sigma_C = 5.8$$

C.V. ของบริษัท A $= \dfrac{3.7}{15.6} = 0.2372$

C.V. ของบริษัท B $= \dfrac{2.5}{13.7} = 0.1825$

C.V. ของบริษัท C $= \dfrac{5.8}{18.9} = 0.3069$

$$\therefore \quad \mathrm{C.V.}_B < \mathrm{C.V.}_A < \mathrm{C.V.}_C$$

ดังนั้น นาย ก. ควรตัดสินใจเลือกลงทุนซื้อหุ้นของบริษัท B

แบบฝึกหัดที่ 3

2. จงหาค่าความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

5, 9, 2, 9, 9, 2, 2, 3

ข้อมูลนี้เป็นของตัวอย่างหรือประชากร เพราะโจทย์ไม่ได้ระบุว่าเลือก สุ่ม, มาจากกลุ่ม, ...

ค่าความแปรปรวน สัญลักษณ์ที่ใช้ คือ σ^2

สูตร ค่าความแปรปรวน = $\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2$

แทนค่า

$\mu = (5+9+\dots+3)/8 = 41/8 = 5.125 \approx 5.13$

$\sum x_i^2 = 5^2+9^2+\dots+3^2 = 289$

$\sigma^2 = (289/8) - 5.13 = 36.125 \approx 36.13 - 5.13$

$= 31$

ค่าเบี่ยงเบนมาตรฐาน สัญลักษณ์ที่ใช้ คือ σ

สูตร ค่าเบี่ยงเบนมาตรฐาน $\sigma = \sqrt{\sigma^2}$

แทนค่า

$\sigma = \sqrt{31}$

$= 5.567 \approx 5.57$

3. จงหาความแปรปรวนและค่าเบี่ยงเบนมาตรฐานของน้ำหนัก (กิโลกรัม) ของนักเรียน โดยทำการสุ่มตัวอย่าง ได้ข้อมูลดังนี้ 2, 4, 8, 4, 6, 6, 8

ข้อมูลนี้เป็นของ**ตัวอย่าง**หรือ**ประชากร** เพราะ: มีคำว่าสุ่มตัวอย่าง

ค่าความแปรปรวน สัญลักษณ์ที่ใช้ คือ $S^2 = \Sigma(x_i-\bar{x})^2/n-1$

สูตร ค่าความแปรปรวน = $\Sigma(x_i-\bar{x})^2/n-1$

แทนค่า $\bar{x} = (2+4+\dots+8)/7 = 38/7 = 5.428 \approx 5.43$

$\Sigma(x_i-\bar{x})^2 = (2-5.43)^2 + (4-5.43)^2 + \dots + (8-5.43)^2 = 29.74 \approx 29.71$

$S^2 = 29.71/7-1 = 29.71/6 = 4.951 \approx 4.95$

ค่าเบี่ยงเบนมาตรฐาน สัญลักษณ์ที่ใช้ คือ S

สูตร ค่าเบี่ยงเบนมาตรฐาน $S = \sqrt{S^2}$

แทนค่า $S = \sqrt{4.95} = 2.22$

4. จงเปรียบเทียบการกระจายของอายุของบุตรของสองครอบครัว โดยที่อายุของบุตรของทั้งสองครอบครัว

เป็นดังนี้ อายุของบุตรในครอบครัวที่หนึ่ง (ปี) 6, 5, 3, 1

อายุของบุตรในครอบครัวที่สอง (ปี) 25, 24, 22, 21, 17

เลือกใช้ ค่าวัดการกระจาย คือ C.V. เพราะ ข้อมูลเป็นประชากร

สูตร คือ σ / μ

ข้อมูลชุดที่ 1

ค่าเฉลี่ย = $(6+5+3+1)/4$

$= 15/4 = 3.75$

ค่าความแปรปรวน = $\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2$

$= 71/4 - 3.75^2 = 17.75 - 14.06$

ค่าเบี่ยงเบนมาตรฐาน = 3.69

$\sigma = \sqrt{3.69} = 1.92$

C.V.

$= 1.92/3.75 = 0.512$

≈ 0.51

ข้อมูลชุดที่ 2

ค่าเฉลี่ย = $(25+24+22+21+17)/5$

$= 109/5 = 21.8$

ค่าความแปรปรวน = $\sigma^2 = \frac{2415}{5} - 475.24 = 7.76$

ค่าเบี่ยงเบนมาตรฐาน σ = $\sqrt{7.76} = 2.785 \approx 2.79$

C.V.

$= 2.79/21.8 = 0.127 \approx 0.13$

เปรียบเทียบค่าวัดการกระจายของข้อมูลชุดที่ 1 และชุดที่ 2 คือ ชุด 2 น้อยกว่า ชุด 1

ดังนั้นข้อมูลชุดที่ 2 มีประสิทธิภาพดีกว่าข้อมูลชุดที่ 1

5. จงเปรียบเทียบการกระจายของข้อมูล 2 ชุดนี้

	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน
ชุดที่ 1	35	6
ชุดที่ 2	40	8

ชุดที่ 1

C.V. = 6/35 = 0.17

ชุดที่ 2

C.V. = 8/40 = 0.2

C.V. ชุดที่ 1 < C.V. ชุดที่ 2

ข้อมูลชุดที่ 1 มีประสิทธิภาพดีกว่า ชุดที่ 2

แบบทดสอบที่ 2

1. จงหาค่าเฉลี่ยเลขคณิตของคะแนนวิชาเคมีของนักเรียน 7 คนต่อไปนี้ 30, 35, 20, 16, 17, 25, 29

2. เมื่อ พ.ศ. 2512 ไข่ไก่ขนาดใหญ่ ขนาดกลาง ขนาดเล็ก ราคาฟองละ 90, 80 และ 75 สตางค์ตามลำดับ ถ้าชายคนหนึ่งซื้อไข่ไก่มาทั้งสิ้น 100 ฟอง เป็นไข่ไก่ขนาดใหญ่ 50 ฟอง ขนาดกลาง 30 ฟอง ขนาดเล็ก 20 ฟอง เฉลี่ยแล้วชายคนนี้ซื้อไข่ไก่มาฟองละเท่าไร

3. จงหาค่าส่วนเบี่ยงเบนมาตรฐานของน้ำหนัก (กิโลกรัม) ของนักเรียนต่อไปนี้ 42, 40, 38, 44, 46, 46, 48

4. เงินเดือนคนงานของโรงงานแห่งหนึ่งเฉลี่ยต่อคนมีค่าเท่ากับ 6,000 บาทต่อเดือน สัมประสิทธิ์ของการแปรผันของเงินเดือนเท่ากับ 12 เปอร์เซ็นต์ ส่วนเบี่ยงเบนมาตรฐานของเงินเดือนของคนงานมีค่าเท่าใด