บทที่ 3

การสรุปลักษณะของข้อมูล

3.1 การวัดแนวโน้มเข้าสู่ส่วนกลาง

ในการวิเคราะห์ข้อมูลกรณีที่ข้อมูลมีจำนวนมากบางครั้งอาจทำให้ผู้อ่านเข้าใจยากไม่สะดวกต่อการนำไปใช้ประโยชน์หรือนำข้อมูลไปวิเคราะห์ขั้นสูงต่อไป ดังนั้นจึงมีความจำเป็นที่จะต้องมีการสรุปลักษณะของข้อมูลเสียก่อน โดยการใช้ตัวแทนของข้อมูลเป็นการสรุปลักษณะข้อมูลเบื้องต้น ซึ่งเราจะเรียกว่าการวัดแนวโน้มเข้าสู่ส่วนกลาง โดยการวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้กันมีอยู่ 3 ชนิด คือ

3.1.1 ค่าเฉลี่ย (Mean) ซึ่งจำแนกได้ดังนี้

3.1.1.1 ค่าเฉลี่ยเลขคณิต หรือมัชฌิมเลขคณิต (Arithmetic Mean) เป็นค่าเฉลี่ยที่นิยมใช้กันมากที่สุดเพราะสามารถสื่อความหมายและทำความเข้าใจได้ง่าย และยังมีสมบัติทางสถิติที่ดี

โดยที่ กรณีข้อมูลไม่ได้จัดกลุ่ม ประชากร แทนด้วย

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{(x_1 + x_2 + \ldots + x_N)}{N}$$

ตัวอย่าง แทนด้วย

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{(x_1 + x_2 + \ldots + x_n)}{n}$$

กรณีข้อมูลจัดกลุ่ม

$$\mu = \frac{\sum_{i=1}^{R} f_i X_i}{N}$$

เมื่อ X_i คือค่ากึ่งกลางของชั้นที่ i

f_i คือความถี่ของชั้นที่ i

3.1.1.2 ค่าเฉลี่ยเรขาคณิต (Geometric Mean)

$$G.M. = \frac{N}{\sum \frac{1}{x}}$$

3.1.1.3 ค่าเฉลี่ยฮาร์โมนิค (Harmonic Mean)

$$H.M = \sqrt[N]{x_1 x_2 x_3 \times \ldots x_N}$$

แต่ค่าเฉลี่ยที่นิยมใช้มากที่สุด คือ ค่าเฉลี่ยเลขคณิต เนื่องจากมีคุณสมบัติทางสถิติที่ดี ดังนั้นในบทนี้จะขอยกตัวอย่างเพียงแค่ค่าเฉลี่ยเลขคณิตเท่านั้น

ตัวอย่างที่ 1 จากการบันทึกข้อมูลจำนวนนักศึกษาที่มาใช้บริการห้องคอมพิวเตอร์ของสาขาวิชาวิทยาการคอมพิวเตอร์เป็นเวลา 20 วัน ได้ข้อมูลดังนี้

30	35	27	20	24	32	38	29	19	22
33	37	28	21	38	30	25	27	26	17

จงหาค่าเฉลี่ยเลขคณิต

วิธีทำ จากสูตร

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{(x_1 + x_2 + \ldots + x_N)}{N}$$

$$= \frac{(30 + 35 + 27 + \ldots + 17)}{20}$$

$$= \frac{558}{20} = 27.9$$

จะได้ค่าเฉลี่ยเลขคณิตเท่ากับ 27.9 หมายความว่า จำนวนนักศึกษาเฉลี่ยที่มาใช้บริการห้องคอมพิวเตอร์ของสาขาวิชาวิทยาการคอมพิวเตอร์เป็นเวลา 20 วัน เท่ากับ 28 คน

ตัวอย่างที่ 2 จากการบันทึกเวลาที่นักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ใช้ในการรับประทานอาหารกลางวัน จำนวน 20 คน เป็นดังนี้ หน่วย : นาที

20	25	37	27	28	45	50	55	45
21	57	28	32	38	41	55	40	45
37	46	39						

จงหาค่าเฉลี่ยเลขคณิต

วิธีทำ จากสูตร

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

$$= \frac{(20 + 25 + 37 + \dots + 39)}{20}$$

$$= \frac{790}{20} = 39.5$$ นาที

ตัวอย่างที่ 3 จากการตรวจสอบรอยตำหนิบนชิ้นส่วนไมโครชิพจำนวน 50 ชิ้น เป็นดังนี้

รอยตำหนิ	0 - 2	3 - 5	6 - 8	9 - 11	12 - 14
จำนวนชิ้น	25	14	5	4	2

จงคำนวณหาค่าเฉลี่ย

วิธีทำ

รอยตำหนิ	จำนวนชิ้น (f_i)	จุดกึ่งกลางชั้น (X_i)	$f_i X_i$
0 – 2	25	1	25
3 - 5	14	4	56
6 - 8	5	7	35
9 – 11	4	10	40
12 – 14	2	13	26
รวม	50		$\sum f_i X_i = 182$

จากสูตร

$$\mu = \frac{\sum_{i=1}^{R} f_i X_i}{N}$$

$$= \frac{182}{50} = 3.64$$

จำนวนรอยตำหนิบนชิ้นส่วนไมโครชิพเฉลี่ยเท่ากับ 3.64 หรือประมาณ 4 รอย

3.1.2 มัธยฐาน (Median , Me) แบ่งเป็น 2 กรณี คือ

โดยที่ <u>กรณีที่ข้อมูลไม่ได้จัดกลุ่ม</u> ค่ามัธยฐานคือค่าของข้อมูลที่มีตำแหน่งอยู่ตรงกลางของชุดข้อมูลเมื่อนำชุดข้อมูลเรียงจากน้อยไปหามากหรือมากไปหาน้อย

- ถ้าจำนวนข้อมูลเป็นจำนวนคี่ (n เป็นเลขคี่) ค่ามัธยฐาน คือ ค่าของข้อมูลที่อยู่ที่ตำแหน่ง $\frac{(n+1)}{2}$

- ถ้าจำนวนข้อมูลเป็นจำนวนคู่ (n เป็นเลขคู่) ค่ามัธยฐาน คือ ค่าเฉลี่ยของข้อมูลที่อยู่ที่ตำแหน่ง $\frac{n}{2}$ และ $\frac{(n+2)}{2}$

<u>กรณีข้อมูลจัดกลุ่ม</u> ต้องหาค่ามัธยฐานจากสูตร

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

โดยที่ L แทน ขอบเขตล่างของชั้นมัธยฐาน

I แทน ความกว้างของชั้นมัธยฐาน

N แทน จำนวนค่าสังเกตทั้งหมด

$\sum f_L$ แทน ผลรวมความถี่ของอันตรภาคชั้นทุกชั้นที่มีค่าสังเกตต่ำกว่าชั้นมัธยฐานเมื่อเรียงข้อมูลจากน้อยไปหามาก

f แทน ความถี่ของชั้นมัธยฐาน

ตัวอย่างที่ 1 จงหาค่ามัธยฐานของข้อมูล

24 28 32 39 27 35 33 29 30

วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ดังนี้

24 27 28 29 **30** 32 33 35 39

ค่ามัธยฐานคือค่าของข้อมูลที่อยู่ตำแหน่งที่ $\frac{(n+1)}{2} = \frac{(9+1)}{2} = 5$

ดังนั้นจะได้ค่ามัธยฐานเท่ากับ 30

ตัวอย่างที่ 2 จงหาค่ามัธยฐานของข้อมูล

17 19 12 14 29 22 13 21 20 18

วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ดังนี้

12 13 14 17 **18** **19** 20 21 22 29

ค่ามัธฐานคือค่าเฉลี่ยของข้อมูลที่อยู่ตำแหน่งที่ $\frac{n}{2} = \frac{10}{2} = 5$ และ

$\frac{(n+2)}{2} = \frac{(10+2)}{2} = 6$ ดังนั้นจะได้ค่ามัธยฐานเท่ากับ $\frac{18+19}{2} = 18.5$

ตัวอย่างที่ 3 จากการสำรวจน้ำหนักของนักเรียนชั้นประถมศึกษาปีที่ 4 โรงเรียนสาธิตมหาวิทยาลัยราชภัฏบุรีรัมย์ได้ข้อมูลดังนี้

น้ำหนัก (กก.)	จำนวน
20 – 24	7
25 – 29	10
30 – 34	12
35 – 39	9
40 – 44	2
รวม	40

จงหามัธยฐานของน้ำหนักของนักเรียนกลุ่มนี้

วิธีทำ หาค่าความถี่สะสม ดังนี้

น้ำหนัก (กก.)	จำนวน	ความถี่สะสม
20 – 24	7	7
25 – 29	10	17
30 – 34	12	29
35 – 39	9	38
40 – 44	2	40
รวม	40	

→ ชั้นมัธยฐาน

หา $\frac{N}{2} = \frac{40}{2} = 20$ พิจารณาชั้นที่มีความถี่มากกว่า $\frac{N}{2} = \frac{40}{2} = 20$

เป็นชั้นแรกเมื่อเรียงข้อมูลจากน้อยไปหามาก ดังนั้นจะได้ว่า 30 – 34 คือชั้นมัธยฐาน

จากสูตร

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

ในที่นี้จะได้ว่า $L = 29.5$, $I = 5$, $\frac{N}{2} = \frac{40}{2} = 20,$ $\sum f_L = 17$, $f = 12$

แทนค่าต่างๆลงในสูตร ดังนี้

$$M_e = 29.5 + 5\left(\frac{20 - 17}{12}\right)$$

$$= 29.5 + 5(0.25)$$

$$= 29.5 + 1.25 = 30.75$$

ดังนั้นมัธยฐานของน้ำหนักของนักเรียนกลุ่มนี้ คือ 30.75 กิโลกรัม

3.1.3 ค่าฐานนิยม (Mode) แบ่งเป็น 2 กรณี คือ

โดยที่ กรณีที่ข้อมูลไม่ได้จัดกลุ่ม ค่าฐานนิยม คือ ค่าของข้อมูลที่เกิดขึ้นบ่อยที่สุดหรือมีความถี่สูงสุด

กรณีข้อมูลจัดกลุ่ม ต้องหาค่าฐานนิยมจากสูตร

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

โดยที่ L แทน ขอบเขตล่างของชั้นฐานนิยม

d_1 แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นก่อนฐานนิยมเมื่อเรียงข้อมูลจากน้อยไปหามาก

d_2 แทน ความแตกต่างระหว่างความถี่ของชั้นฐานนิยมกับชั้นหลังฐานนิยมเมื่อเรียงข้อมูลจากน้อยไปหามาก

I แทน ความกว้างของชั้นฐานนิยม

ตัวอย่างที่ 1 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

12 17 19 10 12 19 20

ฐานนิยม คือ 12

ตัวอย่างที่ 2 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

11 10 12 11 13 13 21 25

ฐานนิยม คือ 11 และ 13

ตัวอย่างที่ 3 จากข้อมูลต่อไปนี้ จงหาค่าฐานนิยม

4 7 8 12 11 13 5 9

ฐานนิยม คือ ไม่มีฐานนิยม

ตัวอย่างที่ 4 จากการสำรวจความสูงของนักศึกษาชายคณะวิทยาศาสตร์ จำนวน 100 คน ได้ข้อมูลดังตาราง

ความสูงของนักศึกษา (เซนติเมตร)	จำนวน (คน)
134 – 144	8
145 – 154	27
155 – 164	42
165 – 174	18
175 – 184	5
รวม	100

จงหาฐานนิยมของความสูงของนักศึกษากลุ่มนี้

วิธีทำ จากข้อมูลในตารางจะได้ว่าชั้นที่มีความถี่มากที่สุด คือ 155 – 164 ดังนั้นชั้นฐานนิยมคือชั้นนี้ โดยที่ L = 154.5 , I = 10 , d_1 = 42 – 27 = 15 , d_2 = 42 – 18 = 24

จากสูตร

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

แทนค่าต่างๆลงในสูตร ดังนี้

$$M_0 = 154.5 + 10\left(\frac{15}{39}\right)$$

$$= 154.5 + 10\left(\frac{15}{39}\right)$$

$$= 154.5 + 10(0.385)$$

$$= 154.5 + 3.85 = 158.35$$

ดังนั้นฐานนิยมของความสูงของนักศึกษากลุ่มนี้ คือ 158.35 เซนติเมตร

3.2 การหาตำแหน่งของข้อมูล

การหาค่าเปอร์เซนไทล์ เดไซด์ และควอไทล์ เป็นการสรุปลักษณะของข้อมูลโดยการกำหนดตำแหน่งให้กับข้อมูล โดยที่จะมีวิธีการหาดังนี้

3.2.1 การหาค่าเปอร์เซนไทล์ (Percentile)

การหาค่าเปอร์เซนด์ไทล์ที่ r คือ การหาค่าที่ตรงกับตำแหน่งที่ r เมื่อแบ่งข้อมูลออกเป็น 100 ส่วนเท่าๆกัน สัญลักษณ์เขียนแทนด้วย P_r

โดยที่ <u>กรณีข้อมูลไม่ได้จัดกลุ่ม</u> ขั้นตอนการคำนวณหา P_r มีดังนี้คือ

<u>ขั้นตอนที่ 1</u> เรียงข้อมูลจากน้อยไปหามาก

<u>ขั้นตอนที่ 2</u> หาตำแหน่งของข้อมูลซึ่ง

ตำแหน่งของ $$P_r = \frac{r}{100}(N+1)$$

กรณีข้อมูลจัดกลุ่ม ขั้นตอนการคำนวณหา P_r มีดังนี้คือ

ขั้นตอนที่ 1 หาความถี่สะสมของข้อมูลที่เรียงลำดับจากน้อยไปหามาก

ขั้นตอนที่ 2 หาตำแหน่งของข้อมูล

ตำแหน่งของ $P_r = \frac{rN}{100}$

ขั้นตอนที่ 3 พิจารณาชั้นที่มีความถี่สะสมมากกว่า $\frac{rN}{100}$ เป็นชั้นแรกเมื่อเรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 4 หาค่าเปอร์เซนต์ไทล์ จากสูตร

$$P_r = L + I\left(\frac{\frac{rN}{100} - \sum f_L}{f}\right)$$

3.2.2 การหาค่าเดไซด์ (Decide)

การหาค่าเดไซด์ที่ r คือ การหาค่าที่ตรงกับตำแหน่งที่ r เมื่อแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆกัน สัญลักษณ์เขียนแทนด้วย D_r

โดยที่ กรณีข้อมูลไม่ได้จัดกลุ่ม ขั้นตอนการคำนวณหา D_r มีดังนี้คือ

ขั้นตอนที่ 1 เรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 2 หาตำแหน่งของข้อมูลซึ่ง

ตำแหน่งของ $$D_r = \frac{r}{10}(N+1)$$

กรณีข้อมูลจัดกลุ่ม ขั้นตอนการคำนวณหา D_r มีดังนี้คือ

ขั้นตอนที่ 1 หาความถี่สะสมของข้อมูลที่เรียงลำดับจากน้อยไปหามาก

ขั้นตอนที่ 2 หาตำแหน่งของข้อมูล

ตำแหน่งของ $$D_r = \frac{rN}{10}$$

ขั้นตอนที่ 3 พิจารณาชั้นที่มีความถี่สะสมมากกว่า $\frac{rN}{10}$ เป็นชั้นแรกเมื่อเรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 4 หาค่าเปอร์เซนต์ไทล์ จากสูตร $D_r = L + I\left(\frac{\frac{rN}{10} - \sum f_L}{f}\right)$

3.2.3 การหาค่าควอไทล์ (Quartile)

การหาค่าควอไทล์ที่ r คือ การหาค่าที่ตรงกับตำแหน่งที่ r เมื่อแบ่งข้อมูลออกเป็น 4 ส่วนเท่าๆกัน สัญลักษณ์เขียนแทนด้วย Q_r

โดยที่ กรณีข้อมูลไม่ได้จัดกลุ่ม ขั้นตอนการคำนวณหา Q_r มีดังนี้คือ

ขั้นตอนที่ 1 เรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 2 หาตำแหน่งของข้อมูลซึ่ง

ตำแหน่งของ $Q_r = \frac{r}{4}(N+1)$

กรณีข้อมูลจัดกลุ่ม ขั้นตอนการคำนวณหา Q_r มีดังนี้คือ

ขั้นตอนที่ 1 หาความถี่สะสมของข้อมูลที่เรียงลำดับจากน้อยไปหามาก

ขั้นตอนที่ 2 หาตำแหน่งของข้อมูล

ตำแหน่งของ $Q_r = \frac{rN}{4}$

ขั้นตอนที่ 3 พิจารณาชั้นที่มีความถี่สะสมมากกว่า $\frac{rN}{4}$ เป็นชั้นแรกเมื่อเรียงข้อมูลจากน้อยไปหามาก

ขั้นตอนที่ 4 หาค่าเปอร์เซนต์ไทล์ จากสูตร $Q_r = L + I\left(\frac{\frac{rN}{4} - \sum f_L}{f}\right)$

ตัวอย่างที่ 1 จากข้อมูลต่อไปนี้

5 3 2 1 6 8 5 1 9

จงหา P_{30} , D_1 , Q_3

วิธีทำ เรียงข้อมูลจากน้อยไปหามาก ดังนี้

1 1 2 3 5 5 6 8 9

(1) ตำแหน่งของ $P_r = \frac{r}{100}(N+1)$

$$P_{30} = \frac{30}{100}(9+1) = 3$$

ข้อมูลที่ตรงกับตำแหน่งที่ 3 คือ 2

ดังนั้น P_{30} เท่ากับ 2

(2) ตำแหน่งของ $D_r = \frac{r}{10}(N+1)$

$$D_1 = \frac{1}{10}(9+1) = 1$$

ข้อมูลที่ตรงกับตำแหน่งที่ 1 คือ 1

ดังนั้น D_1 เท่ากับ 1

(3) ตำแหน่งของ $Q_r = \frac{r}{4}(N+1)$

$$Q_3 = \frac{3}{4}(9+1) = 7.5$$

ตำแหน่งที่ 7.5 อยู่ระหว่างตำแหน่งที่ 7 กับ 8 ซึ่งข้อมูลที่ตรงกับตำแหน่งที่ 7 และ 8 คือ 6 และ 8 ตามลำดับ ข้อมูลที่ตรงกับตำแหน่งที่ 7.5 เท่ากับ 7

ดังนั้น Q_3 เท่ากับ 7

ตัวอย่างที่ 2 ในการสอบวิชาสถิติมีนักศึกษาเข้าสอบ 50 คน แจกแจงเป็นอันตรภาคชั้นได้ดังต่อไปนี้

คะแนน	ความถี่
31 – 40	3
41 – 50	5
51 – 60	6
61 – 70	12
71 – 80	14
81 – 90	8
91 – 100	2
รวม	50

จงหา ก. คนที่สอบได้คะแนนในตำแหน่งเปอร์เซ็นไทล์ที่ 39 สอบได้กี่คะแนน

ข. คนที่สอบได้คะแนนในตำแหน่งเดไซด์ที่ 9 สอบได้กี่คะแนน

ค. คนที่สอบได้คะแนนในตำแหน่งค่าควอไทล์ที่ 3 สอบได้กี่คะแนน

วิธีทำ <u>ขั้นตอนที่ 1</u> หาความถี่สะสมของข้อมูลที่เรียงลำดับจากน้อยไปหามาก

คะแนน	ความถี่	ความถี่สะสม
31 – 40	3	3
41 – 50	5	8
51 – 60	6	14
61 – 70	12	26
71 – 80	14	40
81 – 90	8	48
91 – 100	2	50
รวม	50	

ก. หาเปอร์เซ็นไทล์ที่ 39 ดังนี้

พิจารณาตำแหน่งเปอร์เซ็นไทล์ที่ 39 จาก $P_r = \frac{rN}{100}$

$$P_{39} = \frac{(39)(50)}{100} = \frac{1950}{100} = 19.5$$

จะได้ชั้นที่เปอร์เซ็นไทล์ที่ 39 อยู่คือ 61 – 70 ในที่นี้จะได้

L = 60.5 , I = 10 , $\frac{rN}{100} = \frac{(39)(50)}{100} = \frac{1950}{100} = 19.5$, $\sum f_L = 14$, f = 12

จากสูตร

$$P_r = L + I\left(\frac{\frac{rN}{100} - \sum f_L}{f}\right)$$

แทนค่าต่างๆลงในสูตร ดังนี้ $P_{39} = 60.5 + 10\left(\frac{19.5 - 14}{12}\right)$

$$= 60.5 + 10\left(\frac{5.5}{12}\right)$$

$$= 60.5 + 10\left(0.458\right)$$

$$= 60.5 + 4.58 = 65.08$$

ดังนั้นคนที่สอบได้คะแนนในตำแหน่งเปอร์เซ็นไทล์ที่ 39 สอบได้ 65.08 คะแนน

ข. หาเดไซด์ที่ 9 ดังนี้

พิจารณาตำแหน่งเดไซด์ที่ 9 จาก $D_r = \frac{rN}{10}$

$$D_9 = \frac{(9)(50)}{10} = \frac{450}{10} = 45$$

จะได้ชั้นที่เดไซด์ที่ 9 อยู่คือ 81 – 90 ในที่นี้จะได้

$L = 80.5$, $I = 10$, $\frac{rN}{10} = \frac{(9)(50)}{10} = \frac{450}{10} = 45$, $\sum f_L = 40$, $f = 8$

จากสูตร $D_r = L + I\left(\frac{\frac{rN}{10} - \sum f_L}{f}\right)$

แทนค่าต่างๆลงในสูตร ดังนี้

$$D_9 = 80.5 + 10\left(\frac{45 - 40}{8}\right)$$

$$= 80.5 + 10\left(\frac{5}{8}\right)$$

$$= 80.5 + 10(0.625)$$

$$= 80.5 + 6.25 = 86.75$$

ดังนั้นคนที่สอบได้คะแนนในตำแหน่งเดไซด์ที่ 9 สอบได้ 86.75 คะแนน

ค. หาควอไทล์ที่ 3 ดังนี้

พิจารณาตำแหน่งควอไทล์ที่ 3 จาก $Q_r = \frac{rN}{4}$

$$Q_3 = \frac{(3)(50)}{4} = \frac{150}{4} = 37.5$$

จะได้ชั้นที่ควอไทล์ที่ 3 อยู่คือ 71 – 80 ในที่นี้จะได้

L = 70.5 , I =10 , $\frac{rN}{4} = \frac{(3)(50)}{4} = \frac{150}{4} = 37.5$, $\sum f_L = 26$, f = 14

จากสูตร

$$Q_r = L + I\left(\frac{\frac{rN}{4} - \sum f_L}{f}\right)$$

แทนค่าต่างๆลงในสูตร ดังนี้

$$Q_3 = 70.5 + 10\left(\frac{37.5 - 26}{14}\right)$$

$$= 70.5 + 10\left(\frac{11.5}{14}\right)$$

$$= 70.5 + 10(0.821)$$

$$= 70.5 + 8.21 = 78.71$$

ดังนั้นคนที่สอบได้คะแนนในตำแหน่งควอไทล์ที่ 3 สอบได้ 78.71 คะแนน

การวัดการกระจาย

การสรุปลักษณะของข้อมูล โดยทั่วไปมักดูที่ค่ากลาง และค่าวัดการกระจาย เพราะข้อมูล 2 ชุด อาจมีค่ากลางเท่ากัน แต่ค่าวัดการกระจายไม่เท่ากัน ดังนั้นการดูแต่ค่ากลาง อาจไม่ช่วยให้เห็นลักษณะของข้อมูลที่ชัดเจน เช่น มีข้อมูล 2 ชุด คือ A และ B

ชุด A : 4 5 7 9 10 ; $\bar{X} = 7$

ชุด B : 2 4 7 10 12 ; $\bar{X} = 7$

จะเห็นว่าข้อมูลชุด A และ B มีค่ากลางเท่ากันแต่การกระจายของข้อมูลไม่เท่ากัน ดังนั้นการเปรียบเทียบข้อมูล 2 ชุด ก็ควรพิจารณาทั้งค่ากลางและค่าวัดการกระจาย การวัดการกระจายที่นิยมใช้มีดังนี้

1. พิสัย หมายถึง การหาการกระจายของข้อมูลโดยนำข้อมูลที่มีค่าสูงที่สุด ลบกับข้อมูลที่มีค่าต่ำที่สุด เพื่อให้ได้ค่าที่เป็นช่วงของการกระจาย ซึ่งสามารถบอกถึงความกว้างของข้อมูลชุดนั้นๆ สำหรับสูตรที่ใช้ในการหาพิสัยคือ

$$\text{พิสัย (R)} = X_{max} - X_{min}$$

ตัวอย่าง จงหาพิสัยจากข้อมูลชุดนี้

28 30 45 18 27 55 67 58 19 35

วิธีทำ

จากสูตร $\text{พิสัย (R)} = X_{max} - X_{min}$

$= 67 - 18$

$= 49$

ดังนั้นจะได้ค่าพิสัย เท่ากับ 49

2. ค่าส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation หรือ Average Deviation : M.D.)

เป็นค่าเฉลี่ยของความแตกต่างระหว่างค่าสังเกตและค่าเฉลี่ยสำหรับสูตรที่ใช้ในการคำนวณหาค่าส่วนเบี่ยงเบนเฉลี่ยคือ

- กรณีข้อมูลที่เก็บรวบรวมจากประชากร

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

- กรณีข้อมูลที่เก็บรวบรวมจากตัวอย่าง

$$MD = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

ตัวอย่าง จากการตรวจสอบเครื่องคอมพิวเตอร์ที่ต้องทำการซ่อมแซมในแต่ละเดือน จำนวน 12 เดือน พบจำนวนเครื่องคอมพิวเตอร์ที่ต้องทำการซ่อมแซมในแต่ละเดือน ดังนี้

5 12 8 4 13 7 10 9 3 2 1 2

จงหาค่าส่วนเบี่ยงเบนเฉลี่ย

วิธีทำ

$$\mu = \frac{5+12+8+\ldots+2}{12} = \frac{76}{12} = 6.33$$

จากสูตร $$MD = \frac{\sum_{i=1}^{N} |X_i - \mu|}{N}$$

$$= \frac{|5-6.33| + |12-6.33| + |8-6.33| + \ldots + |2-6.33|}{12}$$

$$= \frac{1.33+5.67+1.67+2.33+6.67+0.67+3.67+2.67+3.33+4.33+5.33+4.33}{12}$$

$$= \frac{42}{12} = 3.5$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนเฉลี่ยเท่ากับ 3.5

3. ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.)

ส่วนเบี่ยงเบนมาตรฐานเป็นค่าวัดการกระจายที่สำคัญทางสถิติ เพราะเป็นค่าที่ใช้บอกถึงการกระจายของข้อมูลได้ดีกว่าค่าพิสัย และค่าส่วนเบี่ยงเบนเฉลี่ยและในการนำเสนอข้อมูลเบื้องต้นส่วนใหญ่จะใช้ค่าเฉลี่ยคู่กันกับส่วนเบี่ยงเบนมาตรฐาน สำหรับสูตรที่ใช้ในการคำนวณหาค่าส่วนเบี่ยงเบนมาตรฐานแบ่งออกเป็น 2 กรณีคือ

3.1 กรณีข้อมูลไม่ได้จัดกลุ่ม

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร

สูตรที่ใช้คือ

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

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง

สูตรที่ใช้คือ

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

3.2 กรณีข้อมูลจัดกลุ่ม

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร

สูตรที่ใช้คือ

$$\sigma = \sqrt{\frac{\sum_{i=1}^{R} f_i X_i^2 - \frac{\left(\sum_{i=1}^{R} f_i X_i\right)^2}{N}}{N}}$$

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง

สูตรที่ใช้คือ

$$S = \sqrt{\frac{\sum_{i=1}^{R} f_i x_i^2 - \frac{\left(\sum_{i=1}^{R} f_i x_i\right)^2}{n}}{n-1}}$$

**เมื่อนำค่าส่วนเบี่ยงเบนมาตรฐานมายกกำลังสองจะกลายเป็นค่าความแปรปรวน

ตัวอย่างที่ 1 จากการบันทึกข้อมูลจำนวนนักศึกษาที่เข้ามาใช้บริการห้องคอมพิวเตอร์ สาขาวิชาสถิติประยุกต์ มหาวิทยาลัยราชภัฏบุรีรัมย์เป็นเวลาหนึ่งสัปดาห์ได้ข้อมูลดังนี้

12	15	10	7	9	14	7

จงหาค่าความแปรปรวนและค่าส่วนเบี่ยงเบนมาตรฐาน

วิธีทำ จากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐาน

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

หาค่าต่างๆ ดังนี้

$$\sum_{i=1}^{N} X_i^2 = 12^2 + 15^2 + 10^2 + 7^2 + 9^2 + 14^2 + 7^2 = 844$$

$$\sum_{i=1}^{N} X_i = 12 + 15 + 10 + 7 + 9 + 14 + 7 = 74$$

แทนค่าลงในสูตร ดังนี้

$$\sigma = \sqrt{\frac{844 - \frac{(74)^2}{7}}{7}}$$

$$= \sqrt{\frac{844 - 782.286}{7}}$$

$$= \sqrt{\frac{61.714}{7}} = \sqrt{8.816} = 2.969$$

$$\sigma^2 = 2.969^2 = 8.815$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 2.969 และ 8.815 ตามลำดับ

ตัวอย่างที่ 2 จงหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนผลสัมฤทธิ์ทางการเรียนวิชาสถิติสำหรับนักวิทยาศาสตร์ของนักเรียนกลุ่มตัวอย่างจำนวน 10 คน ซึ่งมีค่าดังต่อไปนี้

75	65	87	77	85	92	83	73	61	58

วิธีทำ จากสูตรการหาค่าส่วนเบี่ยงเบนมาตรฐาน

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

หาค่าต่างๆ ดังนี้

$$\sum_{i=1}^{n} x_i^2 = 75^2 + 65^2 + 87^2 + 77^2 + 85^2 + 92^2 + 83^2 + 73^2 + 61^2 + 58^2 = 58,340$$

$$\sum_{i=1}^{n} x_i = 75 + 65 + 87 + 77 + 85 + 92 + 83 + 73 + 61 + 58 = 756$$

แทนค่าลงในสูตร ดังนี้

$$S = \sqrt{\frac{58,340 - \frac{(756)^2}{10}}{10 - 1}}$$

$$= \sqrt{\frac{58,340 - 57,153.6}{9}}$$

$$= \sqrt{\frac{1186.4}{9}} = \sqrt{131.822} = 11.481$$

$$S^2 = 11.481^2 = 131.822$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 11.481 และ 131.822 ตามลำดับ

ตัวอย่างที่ 3 อาจารย์สอนในรายวิชาปฏิบัติการคอมพิวเตอร์ท่านหนึ่ง ต้องการทราบความสามารถในการใช้โปรแกรมคอมพิวเตอร์ชนิดหนึ่งของนักศึกษาห้องหนึ่ง จึงสุ่มตัวอย่างนักศึกษามา 25 คน และวัดความสามารถในการใช้โปรแกรมของนักศึกษาแต่ละคน คะแนนที่ได้นำมาแจกแจงความถี่ได้ดังต่อไปนี้

ชั้นคะแนน	ความถี่
79 - 81	2
76 - 78	3
73 - 75	4
70 - 72	8
67 - 69	5
64 - 66	2
61 - 63	1
รวม	25

จงคำนวณหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของคะแนนความสามารถในการใช้โปรแกรมของนักศึกษาห้องนี้

วิธีทำ จากสูตร

$$S = \sqrt{\frac{\sum_{i=1}^{R} f_i x_i^2 - \frac{\left(\sum_{i=1}^{R} f_i x_i\right)^2}{n}}{n-1}}$$

คำนวณค่าต่างๆ ดังนี้

ชั้นคะแนน	ความถี่ (f_i)	จุดกึ่งกลางชั้น (X_i)	X_i^2	$f_i X_i^2$	$f_i X_i$
79 - 81	2	80	6,400	12,800	160
76 - 78	3	77	5,929	17,787	231
73 - 75	4	74	5,476	21,904	296
70 - 72	8	71	5,041	40,328	568
67 - 69	5	68	4,624	23,120	340
64 - 66	2	65	4,225	8,450	130
61 - 63	1	62	3,844	3,844	62
รวม	25			128,233	1,787

แทนค่าต่างๆลงในสูตร ดังนี้

$$S = \sqrt{\frac{128,233 - \frac{(1,787)^2}{25}}{25 - 1}}$$

$$= \sqrt{\frac{128,233 - 127,734.76}{24}}$$

$$= \sqrt{\frac{498.24}{24}} = \sqrt{20.76} = 4.556$$

$$S^2 = 4.556^2 = 20.757$$

ดังนั้นจะได้ค่าส่วนเบี่ยงเบนมาตรฐานและค่าความแปรปรวนเท่ากับ 4.556 และ 20.757 ตามลำดับ

3.4 การเปรียบเทียบข้อมูล

การวัดการกระจายของข้อมูลโดยใช้ค่าพิสัย ค่าส่วนเบี่ยงเบนเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐานเป็นการวัดการกระจายของข้อมูลที่อยู่ในชุดเดียวกันและเป็นข้อมูลที่ไม่มีหน่วย แต่ถ้าข้อมูลที่เรานำมาวัดการกระจายเป็นข้อมูลที่อยู่คนละชุดกัน มีหน่วยการวัดที่ไม่เหมือนกันและค่ากลางหรือค่าเฉลี่ยมีความแตกต่างกันมาก หากต้องการเปรียบเทียบการกระจายของข้อมูล 2 ชุดนี้หรือมากกว่า 2 ชุด จะต้องใช้การวัดการกระจายสัมพัทธ์ (relative dispersion) ซึ่งมีอยู่หลายวิธีแต่ที่นิยมใช้กันก็คือ

3.4.1 ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variance) หรือ สัมประสิทธิ์การกระจาย เป็นค่าที่ใช้วัดการกระจายของข้อมูลที่ไม่มีหน่วย ซึ่งถ้าข้อมูลชุดใดมีค่า C.V. มาก จะมีการกระจายมากกว่าข้อมูลที่มีค่า C.V. น้อย สูตรที่ใช้ในการคำนวณเป็นดังนี้

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

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง C.V. $= \frac{S}{\overline{X}} \times 100$

ตัวอย่าง จากการเก็บรวบรวมข้อมูลของน้ำหนักเด็กผู้หญิงที่มีอายุ 12 ปี มาจำนวน 60 คน กับน้ำหนักของผู้หญิงที่อายุ 24 ปี มาจำนวน 60 คน เช่นเดียวกัน ได้ข้อมูลดังนี้

น้ำหนักเด็กผู้หญิงอายุ 12 ปี	น้ำหนักผู้หญิงอายุ 24 ปี
ค่าเฉลี่ย ($\overline{X}$) = 25.7 กิโลกรัม	ค่าเฉลี่ย ($\overline{X}$) = 49.5 กิโลกรัม
ค่าส่วนเบี่ยงเบนมาตรฐาน =7.85 กิโลกรัม	ค่าส่วนเบี่ยงเบนมาตรฐาน = 12.6 กิโลกรัม

จงเปรียบเทียบการกระจายของข้อมูลทั้งสองชุดนี้

วิธีทำ จากโจทย์พบว่าถึงแม้ข้อมูลทั้งสองชุดจะมีหน่วยเดียวกัน แต่ค่าเฉลี่ยของข้อมูลทั้งสองชุดนี้มีความแตกต่างกันมาก ดังนั้นในการเปรียบเทียบการกระจายจึงต้องใช้ค่าสัมประสิทธิ์การกระจาย (C.V.) ในการเปรียบเทียบข้อมูลดังนี้

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

C.V ของเด็กผู้หญิงอายุ 12 ปี = $\frac{7.85}{25.7} \times 100 = 30.545\%$

C.V ของผู้หญิงอายุ 24 ปี = $\frac{12.6}{49.5} \times 100 = 25.455\%$

แสดงว่า การกระจายหรือความแตกต่างของน้ำหนักในกลุ่มเด็กผู้หญิงอายุ 12 ปีมีมากกว่าการกระจายหรือความแตกต่างของน้ำหนักในกลุ่มของผู้หญิงอายุ 24 ปี

3.4.2 แปลงเป็นค่าคะแนนมาตรฐาน (Standard Score)

การเปรียบเทียบข้อมูลที่อยู่ต่างชุดกันการจะนำค่าสังเกตมาเปรียบเทียบกันเลยนั้นไม่ได้เพราะลักษณะของข้อมูลแต่ละชุดจะมีความแตกต่างกันต้องมีการปรับค่าข้อมูลให้เป็นคะแนนมาตรฐานเสียก่อน จากสูตร

- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากประชากร $z = \frac{x - \mu}{\sigma}$
- ถ้าข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลที่นำมาจากตัวอย่าง $z = \frac{x - \bar{x}}{s}$

ตัวอย่าง ในการเรียนการสอนรายวิชาวิทยาศาสตร์และคณิตศาสตร์พื้นฐานในชีวิตประจำวันมีการเรียนการสอนหลายกลุ่ม และมีอาจารย์หลายท่านเป็นผู้สอน มีการให้คะแนนอย่างอิสระ มาดีสอบได้คะแนนในรายวิชานี้ 74 คะแนน โดยลักษณะของกลุ่มนี้คือค่าเฉลี่ยของกลุ่มเท่ากับ 84 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 ส่วนมารวยเรียนอยู่อีกกลุ่มหนึ่ง โดยมารวยสอบได้คะแนนเพียง 57 คะแนน แต่ค่าเฉลี่ยของผู้เรียนในกลุ่มนี้คือ 51 คะแนน และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 จะสามารถบอกได้หรือไม่ว่ามาดีทำคะแนนได้ดีกว่ามารวย

วิธีทำ ในการเปรียบเทียบคะแนนระหว่างมาดีกับมารวย เนื่องจากทั้ง 2 คนมาจากกลุ่มเรียนคนละกลุ่มจึงไม่สามารถนำคะแนนมาเปรียบเทียบกันได้ทันที จึงต้องแปลงคะแนนให้เป็นคะแนนมาตรฐานเสียก่อน จึงจะสามารถนำคะแนนมาเปรียบเทียบกันได้ แปลงคะแนนเป็นคะแนนมาตรฐาน ดังนี้

จากสูตร $$Z = \frac{x - \bar{x}}{s}$$

$$Z_{\text{มาดี}} = \frac{74-84}{10} = -1$$

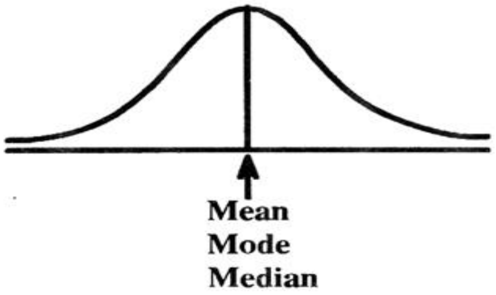
$$Z_{\text{มารวย}} = \frac{57-51}{5} = 1.2$$

จะเห็นได้ว่าคะแนนมาตรฐานของมาดีน้อยกว่าคะแนนมาตรฐานของมารวย นั่นคือ ไม่สามารถบอกได้ว่ามาดีทำคะแนนได้ดีกว่ามารวย

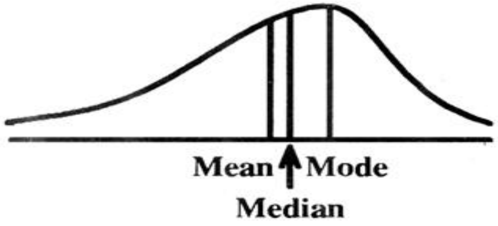
3.5 ความเบ้ (Skewness)

ในการรายงานผลการศึกษาหรือผลการวิจัย การสรุปลักษณะของข้อมูลเราจะใช้ค่ากลางและการวัดการกระจายของข้อมูลเหล่านั้นในการสรุป แต่ถ้าหากผู้ศึกษาต้องการที่จะเลือกวิธีการทางสถิติที่เหมาะสมมาใช้ในการวิเคราะห์ข้อมูลเชิงปริมาณเหล่านี้ เราจำเป็นที่จะต้องทราบว่าการกระจายของข้อมูลเหล่านี้มีความสมมาตรหรือมีความเบ้เกิดขึ้นหรือไม่ เนื่องจากข้อสมมติ (Assumption) ของวิธีการทางสถิติส่วนใหญ่ข้อมูลจะต้องมีความสมมาตรหรือมีการแจกแจงเป็นแบบปกติ สำหรับการตรวจสอบอย่างคร่าวๆก็คือนำข้อมูลมาสร้างเป็นเส้นโค้งความถี่

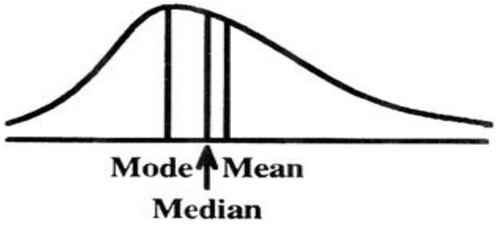
แต่ถ้าต้องการทราบค่าความเบ้ก็จะวัดโดยใช้สัมประสิทธิ์ความเบ้ (Coefficient of skewness) โดยถ้าข้อมูลมีการแจกแจงสมมาตร เส้นโค้งความถี่จะเป็นรูประฆังคว่ำที่มีค่าเฉลี่ย มัธยฐาน และฐานนิยมเท่ากัน และจะมีค่าสัมประสิทธิ์ความเบ้เท่ากับศูนย์ แต่ถ้าข้อมูลมีการแจกแจงไม่สมมาตรคือมีลักษณะเบ้ไปข้างใดข้างหนึ่ง ค่าเฉลี่ย มัธยฐาน และฐานนิยม จะมีค่าต่างกันไม่อยู่ในตำแหน่งเดียวกัน ดังรูป



ก. รูปโค้งปกติ (Normal curve)



ข. รูปโค้งเบ้ทางลบ (Negatively skewed)



ค. รูปโค้งเบ้ทางบวก (Positively skewed)

สำหรับสูตรของสัมประสิทธิ์ความเบ้ซึ่งคาร์ล เพียร์สัน (Karl Pearson) ได้พัฒนาขึ้น

จะเป็นดังนี้ สัมประสิทธิ์ความเบ้ = $\dfrac{3(\text{ค่าเฉลี่ย} - \text{ค่ามัธยฐาน})}{\text{ส่วนเบี่ยงเบนมาตรฐาน}}$

โดยทั่วไปค่าสัมประสิทธิ์ความเบ้จะมีค่าอยู่ระหว่าง -3 กับ 3 ถ้าการแจกแจงของข้อมูลมีไม่มาก ค่าสัมประสิทธิ์ความเบ้จะอยู่ในช่วง ± 1 ถ้าค่าสัมประสิทธิ์ความเบ้มีค่ามากแสดงว่าข้อมูลมีลักษณะเบ้มาก

ตัวอย่าง ศูนย์คอมพิวเตอร์และอินเตอร์เน็ตของมหาวิทยาลัยแห่งหนึ่งได้ทำการบันทึกเวลาเฉลี่ยที่นักศึกษาแต่ละคนเข้ามาใช้บริการในแต่ละวันจากนั้นนำข้อมูลไปแจกแจงความถี่ และคำนวณค่าเฉลี่ยได้เป็น 4.5 ชั่วโมง มีค่ามัธยฐาน 3.4 ชั่วโมง และค่าฐานนิยมเท่ากับ 3 ชั่วโมง โดยมีค่าส่วนเบี่ยงเบนมาตรฐานของจำนวนชั่วโมงที่เข้ามาใช้บริการเท่ากับ 1.89 ชั่วโมง จงตอบคถามต่อไปนี้

ก. การแจกแจงของจำนวนชั่วโมงที่เข้ามาใช้บริการ มีความสมมาตร เบ้ซ้าย หรือเบ้ขวา

ข. จงคำนวณหาค่าสัมประสิทธิ์ความเบ้ พร้อมทั้งสรุปผล

วิธีทำ ก. การแจกแจงของจำนวนชั่วโมงที่เข้ามาใช้บริการที่ศูนย์คอมพิวเตอร์และอินเตอร์เน็ตมีลักษณะเบ้ขวา เพราะค่าเฉลี่ยมีค่ามากกว่าค่ามัธยฐาน

ข. จากสูตร สัมประสิทธิ์ความเบ้ = $\dfrac{3(\text{ค่าเฉลี่ย} - \text{ค่ามัธยฐาน})}{\text{ส่วนเบี่ยงเบนมาตรฐาน}}$

$$\frac{3(4.5-3.4)}{1.89}=\frac{3.3}{1.89}=1.746$$

เนื่องจากค่าสัมประสิทธิ์ความเบ้จะมีค่าไม่เกิน ± 3 ซึ่งในกรณีนี้ค่าสัมประสิทธิ์ความเบ้มีค่า = 1.746 แสดงว่าการกระจายของข้อมูลมีลักษณะเบ้ขวาแต่ไม่มาก ซึ่งหมายความว่านักศึกษาจำนวนไม่มากที่มีจำนวนชั่วโมงที่เข้ามาใช้บริการที่ศูนย์คอมพิวเตอร์และอินเตอร์เน็ตมากกว่า 4.5 ชั่วโมง

3.6 บอกซ์พลอต (Box Plots)

การสร้างบอกซ์พล็อตเป็นการสรุปลักษณะของข้อมูลชุดหนึ่งๆด้วยกราฟ ซึ่งในการสร้างต้องทราบค่าสถิติ 5 ค่า คือ

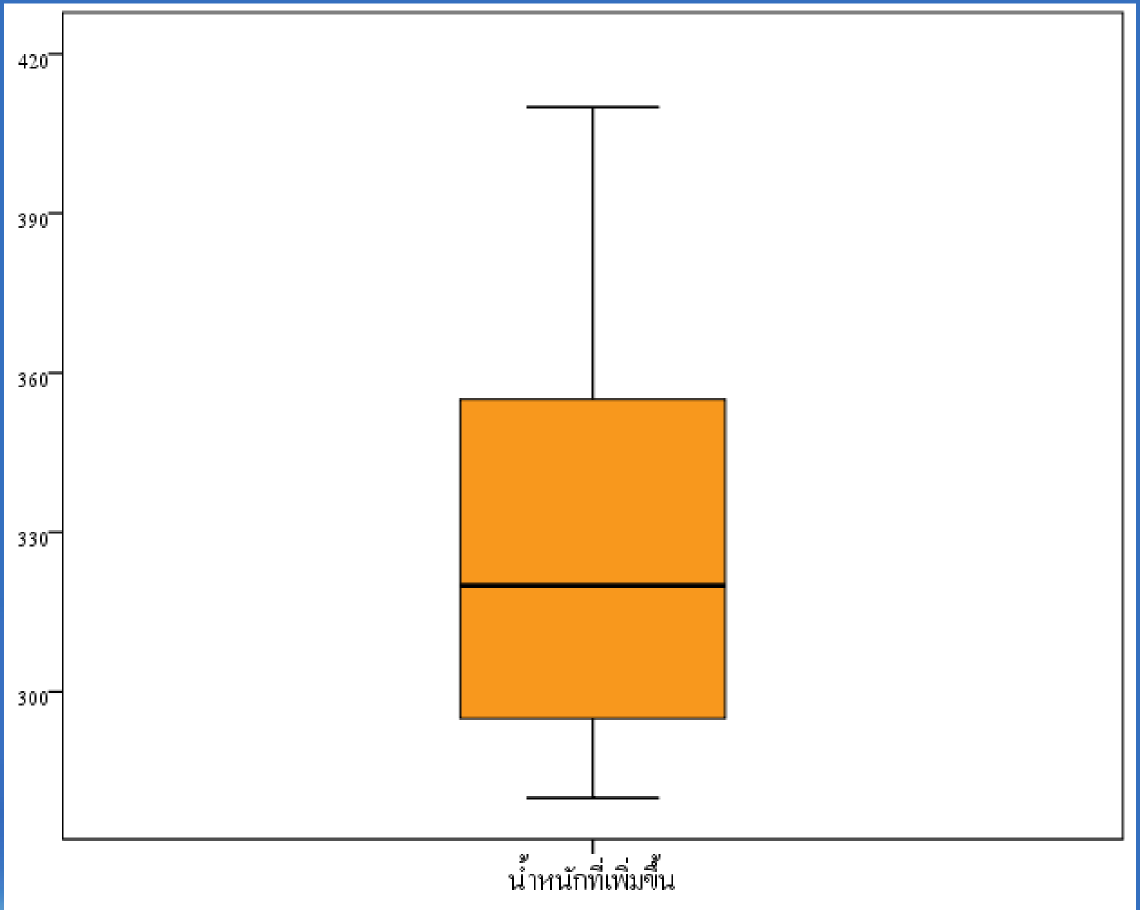
1. ค่าต่ำสุดของข้อมูล
2. ค่าควอไทล์ที่ 1 หรือเปอร์เซ็นต์ไทล์ที่ 25
3. ค่ามัธยฐาน
4. ค่าควอไทล์ที่ 3 หรือเปอร์เซนไทล์ที่ 75
5. ค่าสูงสุดของข้อมูล

ตัวอย่าง จากการสุ่มตัวอย่างลูกไก่จากฟาร์มแห่งหนึ่งขึ้นมา 12 ตัว นำไปชั่งน้ำหนักที่เพิ่มขึ้นในระยะเวลาหนึ่งเป็นดังนี้ (หน่วย : กรัม)

290	330	360	290	410	310
340	370	310	300	280	350

จงสร้างบอกซ์พลอตของน้ำหนักที่เพิ่มขึ้นของลูกไก่ พร้อมทั้งสรุป

บอกซ์พลอตของน้ำหนักที่เพิ่มขึ้นของลูกไก่ เป็นดังนี้


420
390
360
330
300
น้ำหนักที่เพิ่มขึ้น

จากข้อมูลพบว่า ค่าต่ำสุดมีค่าเท่ากับ 280 กรัม

ค่าสูงสุดมีค่าท่ากับ 410 กรัม

เรียงข้อมูลจากน้อยไปหามาก ดังนี้

280 290 290 300 310 310 330 340

280 350 360 370 410

จาก ตำแหน่งของ $Q_r = \frac{rN}{4}$ จะได้ ตำแหน่งของ $Q_1 = \frac{(1)(12)}{4} = 3$

ซึ่งตรงกับค่า 290

จาก ตำแหน่งของ $Q_r = \frac{rN}{4}$ จะได้ ตำแหน่งของ $Q_3 = \frac{(3)(12)}{4} = 9$

ซึ่งตรงกับค่า 350

เมื่อพิจารณาบอกซ์พลอตจะเห็นว่ามัธยฐานไม่ได้อยู่ตรงกึ่งกลางแต่อยู่ค่อนมาทางค่าน้อย แสดงว่าการกระจายของข้อมูลมีลักษณะเบ้ โดยระยะห่างระหว่าง Q_3 กับมัธยฐานจะยาวกว่าระยะห่างระหว่าง Q_1 กับมัธยฐาน นั่นคือการแจกแจงของข้อมูลมีลักษณะเบ้ขวา แสดงว่ามีลูกไก่จำนวนมากที่มีน้ำหนักที่เพิ่มขึ้นน้อยกว่าน้ำหนักที่เพิ่มขึ้นเฉลี่ยของกลุ่ม

3.7 บทสรุป

ในการวิเคราะห์ข้อมูลกรณีที่ข้อมูลมีจำนวนมากบางครั้งอาจทำให้ผู้อ่านเข้าใจยากไม่สะดวกต่อการนำไปใช้ประโยชน์หรือนำข้อมูลไปวิเคราะห์ขั้นสูงต่อไป ดังนั้นจึงมีความจำเป็นที่จะต้องมีการสรุปลักษณะของข้อมูลเสียก่อน โดยการใช้ตัวแทนของข้อมูลเป็นการสรุปลักษณะข้อมูลเบื้องต้น ซึ่งเราจะเรียกว่าการวัดแนวโน้มเข้าสู่ส่วนกลาง แต่ถ้าหากผู้วิเคราะห์ต้องการทราบตำแหน่งของข้อมูลก็อาจจะพิจารณาจากค่าเปอร์เซนไทล์ เดไซด์ และควอไทล์ และหากผู้วิเคราะห์ต้องการให้เห็นลักษณะของข้อมูลที่ชัดเจนขึ้นก็จะต้องมีการวัดการกระจายของข้อมูลโดยมีอยู่หลายค่าให้เลือกใช้ เช่น ค่าพิสัย ค่าเบี่ยงเบนเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน เป็นต้น แต่ถ้าต้องการเปรียบเทียบระหว่างกลุ่มของข้อมูลก็อาจเลือกใช้ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variance) หรือ สัมประสิทธิ์การกระจาย แต่ถ้าจะให้เห็นภาพก็อาจทำการพล็อตเป็นกราฟ หรือบอกซ์พลอตเพื่อดูความเบ้ก็ได้