บทที่ 3
การวิเคราะห์ข้อมูลเบื้องต้น

การวิเคราะห์ข้อมูลเป็นขั้นตอนต่อจากการเก็บรวบรวมข้อมูล หลังจากเราเก็บรวบรวมข้อมูลมาได้แล้วจะต้องทำการวิเคราะห์ข้อมูลที่ได้มา โดยในบทนี้จะกล่าวถึงการวิเคราะห์ข้อมูลเบื้องต้นคือ การวัดตำแหน่งเปรียบเทียบ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจาย ซึ่งการวัดตำแหน่งเปรียบเทียบ การวัดแนวโน้มเข้าสู่ส่วนกลาง และการวัดการกระจายมีความสำคัญเป็นอย่างยิ่ง เนื่องจากข้อมูลที่เก็บรวบรวมมามีจำนวนค่อนข้างมากหรือมีจำนวนมาก ในบทนี้จะกล่าวถึงวัดตำแหน่งเปรียบเทียบได้แก่ อัตราส่วน สัดส่วน ร้อยละ ควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ ซึ่งจะทำให้ทราบถึงลักษณะตำแหน่งของข้อมูลแต่ละตัว การวัดแนวโน้มเข้าสู่ส่วนกลางได้แก่ ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม เพื่อทำให้ทราบถึงค่ากลางของข้อมูลชุดนั้น และการวัดการกระจายได้แก่ พิสัย ความแปรปรวน ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์การกระจาย เพื่อทำให้เห็นความแตกต่างหรือการกระจายของข้อมูลชุดนั้น ๆ ได้อย่างชัดเจน

การวัดตำแหน่งเปรียบเทียบ

ข้อมูลชุดหนึ่งควรที่จะนำมาบรรยาย เปรียบเทียบ และแปลความหมายเพื่อนำไปใช้ประโยชน์ วิธีการที่จะทำให้ทราบถึงลักษณะตำแหน่งของข้อมูลแต่ละตัว คือ การวัดตำแหน่งเปรียบเทียบ ซึ่งประกอบด้วย อัตราส่วน สัดส่วน ร้อยละ ควอไทล์ เดไซล์ เปอร์เซ็นไทล์ และความสัมพันธ์ระหว่างควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ ซึ่งมีรายละเอียดดังนี้ (ชัยวิชิต เชียรชนะ. 2562 : 36-37)

1. **อัตราส่วน** (Ratio)

อัตราส่วนเป็นการเปรียบเทียบความถี่ระหว่างรายการย่อยกับรายการย่อยของตัวแปร คำนวณได้จาก

ความถี่รายการย่อย 1 : ความถี่รายการย่อย 2 : ... : ความถี่รายการย่อย n

ตัวอย่าง 3.1 อัตราส่วนของบุคลากรในสำนักงานต่าง ๆ คือ 200 : 150 : 100 : 50 : 250

คิดเป็นอัตราส่วนคือ 4 : 3 : 2 : 1 :5

2. สัดส่วน (Proportion)

สัดส่วนเป็นการเปรียบเทียบความถี่ระหว่างรายการย่อยกับความถี่ทั้งหมด คำนวณได้จาก

สัดส่วน = ความถี่รายการย่อย/ความถี่ทั้งหมด

ตัวอย่าง 3.2 มีนักศึกษาทั้งหมดจำนวน 7,500 คน ซึ่งเป็นนักศึกษาชั้นปีที่ 1 จำนวน 1,270 คน

ดังนั้น สัดส่วนของนักศึกษาชั้นปีที่ 1 กับนักศึกษาทั้งหมด คือ $\frac{1,270}{7,500} = 0.169$

3. ร้อยละ (Percentage)

ร้อยละเป็นการเปรียบเทียบความถี่ระหว่างรายการย่อยกับความถี่ทั้งหมด โดยปรับเทียบให้เป็น 100 นั่นคือ เป็นการนำค่าสัดส่วนมาปรับฐานให้เป็น 100 คำนวณได้จาก

ร้อยละ = (ความถี่รายการย่อย/ความถี่ทั้งหมด) x 100

ตัวอย่าง 3.3 มีนักศึกษาทั้งหมดจำนวน 7,500 คน ซึ่งเป็นนักศึกษาชั้นปีที่ 1 จำนวน 1,270 คน

ดังนั้น ร้อยละของนักศึกษาชั้นปีที่ 1 กับนักศึกษาทั้งหมด คือ $\frac{1,270}{7,500} = 0.169$

คิดเป็นร้อยละ 16.9 หรือ 16.9%

4. ควอไทล์ (Quartile)

เป็นค่าที่แบ่งข้อมูลออกเป็น 4 ส่วนเท่า ๆ กัน เมื่อข้อมูลถูกเรียงจากน้อยไปหามาก ค่าที่แบ่งแต่ละส่วนเรียกว่า Q_1, Q_2, Q_3

ตัวอย่าง 3.4 นายปฏิมากรณ์ ทำคะแนนสอบวิชาหลักสถิติได้ 55 คะแนน ซึ่งตรงกับควอไทล์ที่ 2 แสดงว่า จากผู้เข้าสอบทั้งหมด มีจำนวน 2 ส่วนใน 4 ส่วน ที่สอบได้คะแนนต่ำกว่าหรือเท่ากับ 55 คะแนน

5. เดไซล์ (Decile)

เป็นค่าที่แบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน เมื่อข้อมูลถูกเรียงจากน้อยไปหามาก ค่าที่แบ่งแต่ละส่วนเรียกว่า $D_1, D_2, \ldots, D_9$

ตัวอย่าง 3.5 นางสาวฮานะ ทำคะแนนสอบวิชาหลักสถิติได้ 80 คะแนน ซึ่งตรงกับเดไซล์ที่ 7 แสดงว่า จากผู้เข้าสอบทั้งหมด มีจำนวน 7 ส่วนใน 10 ส่วน ที่สอบได้คะแนนต่ำกว่าหรือเท่ากับ 80 คะแนน

6. เปอร์เซ็นไทล์ (Percentile)

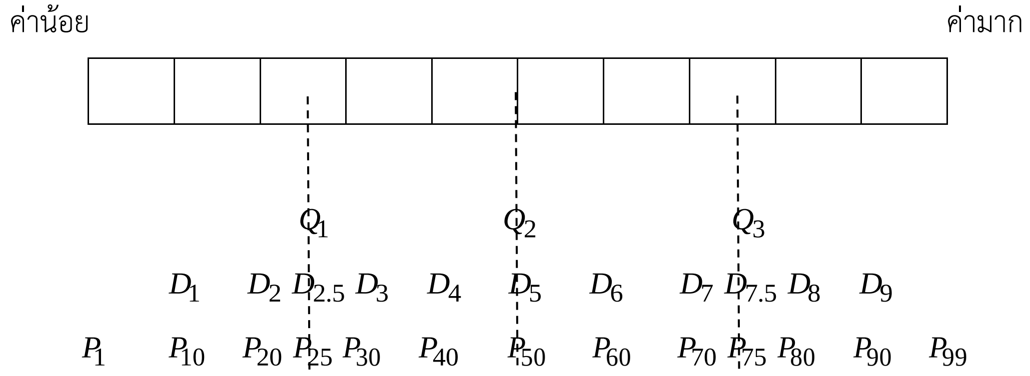
เป็นค่าที่แบ่งข้อมูลออกเป็น 100 ส่วนเท่า ๆ กัน เมื่อข้อมูลถูกเรียงจากน้อยไปหามาก ค่าที่แบ่งแต่ละส่วนเรียกว่า $P_1, P_2, \ldots, P_{99}$

ตัวอย่าง 3.6 นางสาวเอวา ทำคะแนนสอบวิชาหลักสถิติได้ 65 คะแนน ซึ่งตรงกับเปอร์เซ็นไทล์ที่ 80 แสดงว่า จากผู้เข้าสอบทั้งหมด มีจำนวน 80 ส่วนใน 100 ส่วน ที่สอบได้คะแนนต่ำกว่าหรือเท่ากับ 65 คะแนน หรือแสดงว่า ถ้ามีผู้เข้าสอบทั้งหมด 100 คน มีจำนวน 80 คน ที่ทำคะแนนได้ต่ำกว่าหรือเท่ากับ 65 คะแนน หรือแสดงว่า มีผู้สอบร้อยละ 80 ของผู้สอบทั้งหมดที่สอบได้คะแนนต่ำกว่าหรือเท่ากับ 80 คะแนน

7. ความสัมพันธ์ระหว่างควอไทล์ เดไซล์ และเปอร์เซ็นไทล์

ความสัมพันธ์ระหว่างควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ จะพบว่า เป็นการหาตำแหน่งเพื่อเปรียบเทียบเช่นกัน ซึ่งควอไทล์ที่ 1 เดไซล์ที่ 2.5 และเปอร์เซ็นไทล์ที่ 25 จะมีค่าตรงกัน ควอไทล์ที่ 2 เดไซล์ที่ 5 และเปอร์เซ็นไทล์ที่ 50 จะมีค่าตรงกัน และควอไทล์ที่ 3 เดไซล์ที่ 7.5 และเปอร์เซ็นไทล์ที่ 75 จะมีค่าตรงกัน เช่นกัน แสดงดังภาพประกอบ 3.1 ตลอดจนในการแบ่งระดับ

คะแนนด้วยเปอร์เซ็นไทล์ตามประเพณีนิยมก็สามารถนำมาแบ่งให้สอดคล้องกับควอไทล์ดังแนวคิดของ Clark-Carter (2005 : 1539-1540) ได้เสนอหลักการแบ่งเกณฑ์ที่น่าเชื่อถือคือ เปอร์เซ็นไทล์ที่ 25 (หรือควอไทล์ที่ 1) เปอร์เซ็นไทล์ที่ 50 (หรือควอไทล์ที่ 2) และเปอร์เซ็นไทล์ที่ 75 (หรือควอไทล์ที่ 3)



ภาพประกอบ 3.1 ความสัมพันธ์ระหว่างควอไทล์ เดไซล์ และเปอร์เซ็นไทล์
ที่มา : ชัยวิชิต เชียรชนะ (2562 : 38)

8. การคำนวณหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์

ปรีดาภรณ์ กาญจนสำราญวงศ์ (2560 : 56) อำนาจ วังจีน และพรรณี บุญสุยา (2553 :) และอนุวัติ คูณแก้ว (2560 :) ได้กล่าวถึง การคำนวณหาควอไทล์ เดไซล์ และเปอร์เซ็นไทล์ มีวิธีการคำนวณแบ่งเป็น 2 กรณี คือ

กรณีที่ 1 ข้อมูลที่ไม่ได้แจกแจงความถี่ มีขั้นตอนในการคำนวณดังนี้

1. เรียงลำดับข้อมูลจากน้อยไปหามาก
2. หาตำแหน่งของค่าที่เราต้องการจากสูตรต่อไปนี้

ตำแหน่งของ $Q_r = \frac{r}{4} N+1$

ตำแหน่งของ $D_r = \frac{r}{10} N+1$

ตำแหน่งของ $P_r = \frac{r}{100} N+1$

เมื่อ N คือ จำนวนข้อมูลทั้งหมด และ r คือ ตำแหน่งของข้อมูล

3. หาค่าที่ต้องการ ถ้าตำแหน่งที่คำนวณได้เป็นจำนวนเต็ม ให้นับหาตำแหน่งได้เลย แต่ถ้าตำแหน่งที่คำนวณได้เป็นทศนิยมให้เปรียบเทียบหาค่าโดยการเทียบบัญญัติไตรยางศ์

ตัวอย่าง 3.7 จากข้อมูลต่อไปนี้

10 11 22 17 10 9 21 15 12 14 15 จงหา Q_2 , D_6 และ P_{80}

วิธีทำ 1. เรียงลำดับข้อมูลจากน้อยไปหามาก

9 10 10 11 12 14 15 15 17 21 22

2. หาตำแหน่งของค่าที่เราต้องการจากสูตรต่อไปนี้

ตำแหน่งของ $Q_r = \frac{r}{4}\ N+1$

ตำแหน่งของ $D_r = \frac{r}{10}\ N+1$

ตำแหน่งของ $P_r = \frac{r}{100}\ N+1$

แทนค่าจะได้

ตำแหน่งของ $Q_2 = \frac{2}{4}\ 11+1 = \frac{2}{4}\ 12 = 6$

ตำแหน่งของ $D_6 = \frac{6}{10}\ 11+1 = \frac{6}{10}\ 12 = 7.2$

ตำแหน่งของ $P_{80} = \frac{80}{100}\ 11+1 = \frac{80}{100}\ 12 = 9.6$

3. หาค่าที่ต้องการ

ตำแหน่ง Q_2 อยู่ที่ 6 เป็นจำนวนเต็ม การหาคำตอบใช้การนับตำแหน่งได้เลย

ดังนั้น Q_2 อยู่ในตำแหน่งที่ 6 ซึ่งมี $=$ 14 หมายความว่า มีอยู่ 50% ของจำนวนข้อมูลทั้งหมดที่มีค่าน้อยกว่าหรือเท่ากับ 14

ตำแหน่ง D_6 อยู่ที่ 7.2 ไม่ใช่จำนวนเต็ม การหาคำตอบจึงใช้การเทียบบัญญัติไตรยางศ์เพื่อหาค่าตำแหน่งที่ 7.2 มีค่าเท่าใด

ตำแหน่งที่ 7 และ 8 ต่างกัน $8-7 = 1$ ค่าต่างกัน $15-15 = 0$

ตำแหน่งที่ 7 และ 7.2 ต่างกัน $7.2-7 = 0.2$ ค่าต่างกัน $\frac{0\times 0.2}{1} = 0$

ดังนั้น $D_6 = 15+0 = 15$ หมายความว่า มีอยู่ 60% ของจำนวนข้อมูลทั้งหมดที่มีค่าน้อยกว่าหรือเท่ากับ 15

ตำแหน่ง P_{80} อยู่ที่ 9.6 ไม่ใช่จำนวนเต็ม การหาคำตอบจึงใช้การเทียบบัญญัติไตรยางศ์เพื่อหาค่าตำแหน่งที่ 9.6 มีค่าเท่าใด

ตำแหน่งที่ 9 และ 10 ต่างกัน $10-9 = 1$ ค่าต่างกัน $21-17 = 4$

ตำแหน่งที่ 9 และ 9.6 ต่างกัน $9.6-9 = 0.6$ ค่าต่างกัน $\frac{4\times 0.6}{1} = 2.4$

ดังนั้น $P_{80} = 17+2.4 = 19.4$ หมายความว่า มีอยู่ 80% ของจำนวนข้อมูลทั้งหมดที่มีค่าน้อยกว่าหรือเท่ากับ 19.4

กรณีที่ 2 ข้อมูลที่แจกแจงความถี่ มีขั้นตอนในการคำนวณดังนี้

1. สร้างตารางความถี่สะสม
2. หาตำแหน่งของค่าที่เราต้องการจากสูตรต่อไปนี้

ตำแหน่งของ $Q_r = \frac{rN}{4}$

ตำแหน่งของ $D_r = \frac{rN}{10}$

ตำแหน่งของ $P_r = \frac{rN}{100}$

เมื่อ N คือ จำนวนข้อมูลทั้งหมด และ r คือ ตำแหน่งของข้อมูล

3. หาค่าที่อยู่ในตำแหน่งดังกล่าวจาก

$$Q_r = L+\left[\frac{\frac{rN}{4}-\sum f_L}{fQ_r}\right]I$$

$$D_r = L+\left[\frac{\frac{rN}{10}-\sum f_L}{fD_r}\right]I$$

$$P_r = L+\left[\frac{\frac{rN}{100}-\sum f_L}{fP_r}\right]I$$

โดยที่ N คือ จำนวนข้อมูลทั้งหมด

L คือ ขีดจำกัดล่างของชั้นที่มี Q_r, D_r, P_r อยู่

$\sum f_L$ คือ ความถี่สะสมจากชั้นต่ำสุดถึงก่อนชั้นที่มี Q_r, D_r, P_r อยู่

I คือ ความกว้างของชั้น

fQ_r คือ ความถี่ของชั้นที่มี Q_r อยู่

fD_r คือ ความถี่ของชั้นที่มี D_r อยู่

fP_r คือ ความถี่ของชั้นที่มี P_r อยู่

ตัวอย่าง 3.8 จากตารางแจกแจงความถี่ของคะแนนสอบรายวิชาหลักสถิติของนักศึกษา 100 คน จงหา Q_1, D_7, P_{60}

คะแนน	**ความถี่**
10-19	3
20-29	5
30-39	10
40-49	28
50-59	25
60-69	16
70-79	13
รวม	**100**

วิธีทำ สร้างตารางความถี่สะสม

คะแนน	**ความถี่**	**ความถี่สะสม**
10-19	3	3
20-29	5	8
30-39	10	18
40-49	28	46
50-59	25	71
60-69	16	87
70-79	13	100
รวม	**100**	

จากตารางหาตำแหน่ง Q_1

ตำแหน่ง $Q_r = \frac{rN}{4}$

ตำแหน่ง $Q_1 = \frac{1\ 100}{4} = 25$

จากโจทย์ทราบค่า $L = 39.5, \quad \sum f_L = 18, \quad fQ_1 = 28, \quad I = 10$

หาค่าของ Q_1 จากสูตร

$$Q_r = L + \left[\frac{\frac{rN}{4} - \sum f_L}{fQ_r}\right] I$$

$$Q_1 = 39.5 + \left[\frac{\frac{1\ 100}{4} - 18}{28}\right] 10$$

$$= 39.5 + 2.5$$

$$= 42$$

ดังนั้น $Q_1 = 42$ หมายความว่า มีนักศึกษาอยู่ 1 ใน 4 ที่มีคะแนนสอบน้อยกว่าหรือเท่ากับ 42

จากตารางหาตำแหน่ง D_7

ตำแหน่ง $D_r = \frac{rN}{10}$

ตำแหน่ง $D_7 = \frac{7\ 100}{10} = 70$

จากโจทย์ทราบค่า $L = 49.5, \quad \sum f_L = 46, \quad fD_7 = 25, \quad I = 10$

หาค่าของ D_7 จากสูตร

$$D_r = L + \left[\frac{\frac{rN}{10} - \sum f_L}{fD_r} \right] I$$

$$D_7 = 49.5 + \left[\frac{\frac{7\ 100}{10} - 46}{25} \right] 10$$

$$= 49.5 + 9.6$$

$$= 59.1$$

ดังนั้น D_7 หมายความว่า มีนักศึกษาอยู่ 7 ใน 10 ที่มีคะแนนสอบน้อยกว่าหรือเท่ากับ 59.1

จากตารางหาตำแหน่ง P_{60}

ตำแหน่ง $P_r = \frac{rN}{100}$

ตำแหน่ง $P_{60} = \frac{60\ 100}{100} = 60$

จากโจทย์ทราบค่า $L = 49.5, \quad \sum f_L = 46, \quad fP_{60} = 25, \quad I = 10$

หาค่าของ P_{60} จากสูตร

$$P_r = L + \left[\frac{\frac{rN}{100} - \sum f_L}{fP_r} \right] I$$

$$P_{60} = 49.5 + \left[\frac{\frac{60\ 100}{100} - 46}{25} \right] 10$$

$$= 49.5 + 5.6$$

$$= 55.1$$

ดังนั้น $P_{60} = 55.1$ หมายความว่า มีนักศึกษาอยู่ 60 ใน 100 ที่มีคะแนนสอบน้อยกว่าหรือเท่ากับ 55.1

การวัดแนวโน้มเข้าสู่ส่วนกลาง

การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of Central Tendency) เป็นการนำเสนอข้อมูลเบื้องต้นเพื่อให้ทราบถึงค่ากลางของข้อมูลชุดนั้น ซึ่งการคำนวณต้องพิจารณาว่าเป็นข้อมูลไม่ได้แจกแจงความถี่ หรือข้อมูลที่แจกแจงความถี่ โดยค่าวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้มี 3 ชนิด คือ ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม (มนตรี สังข์ทอง. 2557 : 38) ซึ่งแต่ละวิธีมีข้อดี ข้อเสีย และความเหมาะสมในการนำไปใช้งานไม่เหมือนกันขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลชุดนั้น โดยมีรายละเอียดดังนี้ (ปรีดาภรณ์ กาญจนสำราญวงศ์. 2560 : 56 และประสพชัย พสุนนท์. 2555 : 269-270)

1. **ค่าเฉลี่ยเลขคณิต** (Arithmetic Mean)

ค่าเฉลี่ยเลขคณิต หรือเรียกสั้น ๆ ว่า ค่าเฉลี่ย (Mean) คือ ค่าที่ได้จากการนำข้อมูลทั้งหมดมารวมกันแล้วหารด้วยจำนวนข้อมูลทั้งหมด ซึ่งใช้สัญลักษณ์ μ (อ่านว่า มิว) แทนค่าเฉลี่ยของข้อมูลที่ได้จากประชากร และ $\bar{X}$ (อ่านว่า เอกซ์บาร์) แทนค่าเฉลี่ยของข้อมูลที่ได้จากตัวอย่าง การคำนวณหาค่าเฉลี่ยเลขคณิตสามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ข้อมูลไม่มีการแจกแจงความถี่และกรณีที่ข้อมูลมีการแจกแจงความถี่ (ปรีดาภรณ์ กาญจนสำราญวงศ์. 2560 : 56 ; พิศมัย หาญมงคลพิพัฒน์. 2560 : 20 และ Manikandan, S. 2011 : 140)

1.1 การหาค่าเฉลี่ยเลขคณิตในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

การหาค่าเฉลี่ยเลขคณิตในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่ แบ่งออกเป็น 2 ชนิด คือ ค่าเฉลี่ยเลขคณิตของประชากร และค่าเฉลี่ยเลขคณิตของตัวอย่าง ซึ่งมีสูตรในการคำนวณดังนี้

ค่าเฉลี่ยเลขคณิตของประชากร

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

โดยที่ μ คือ ค่าเฉลี่ยเลขคณิตของประชากร

$\sum_{i=1}^{N} X_i$ คือ ผลรวมทั้งหมดของข้อมูล

N คือ จำนวนข้อมูลทั้งหมดของประชากร

ค่าเฉลี่ยเลขคณิตของตัวอย่าง

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

โดยที่ $\bar{X}$ คือ ค่าเฉลี่ยเลขคณิตของตัวอย่าง

$\sum_{i=1}^{n} X_i$ คือ ผลรวมทั้งหมดของข้อมูล

n คือ จำนวนข้อมูลทั้งหมดของตัวอย่าง

ตัวอย่าง 3.9 ในร้านแห่งหนึ่งมีพนักงานขาย 10 คน ขายสินค้าได้จำนวนดังนี้ (หน่วย : ชิ้น)

22 23 24 24 25 25 21 23 15 20

จงหาค่าเฉลี่ยของจำนวนสินค้าที่ขายได้ของพนักงานขายทั้ง 10 คน

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$\begin{aligned}
\mu &= \frac{\sum_{i=1}^{N} X_i}{N} \\
&= \frac{22+23+24+24+25+25+21+23+15+20}{10} \\
&= \frac{222}{10} \\
&= 22.2
\end{aligned}$$

ดังนั้น พนักงานขายขายสินค้าได้เฉลี่ย เท่ากับ 22.2 ชิ้น

ตัวอย่าง 3.10 จากการสุ่มตัวอย่างเพื่อสอบถามอายุของนักศึกษาคณะวิทยาศาสตร์ชั้นปีที่ 1 จำนวน 9 คน พบว่าอายุของแต่ละคนเป็นดังนี้ (หน่วย : ปี) 18 19 19 18 20 18 21 19 18 จงหาอายุเฉลี่ยของนักศึกษาคณะวิทยาศาสตร์ชั้นปีที่ 1

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$
\begin{aligned}
\bar{X} &= \frac{\sum_{i=1}^{n} X_i}{n} \\
&= \frac{18+19+19+18+20+18+21+19+18}{9} \\
&= \frac{170}{9} \\
&= 18.89
\end{aligned}
$$

ดังนั้น อายุเฉลี่ยของนักศึกษาคณะวิทยาศาสตร์ชั้นปีที่ 1 เท่ากับ 18.89 ปี

1.2 การหาค่าเฉลี่ยเลขคณิตในกรณีที่ข้อมูลมีการแจกแจงความถี่

การหาค่าเฉลี่ยเลขคณิตในกรณีที่ข้อมูลมีการแจกแจงความถี่ แบ่งออกเป็น 2 ชนิด คือ ค่าเฉลี่ยเลขคณิตของประชากร และค่าเฉลี่ยเลขคณิตของตัวอย่าง ซึ่งมีสูตรในการคำนวณดังนี้

ค่าเฉลี่ยเลขคณิตของประชากร

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

โดยที่ μ คือ ค่าเฉลี่ยเลขคณิตของประชากร

f_i คือ ความถี่ของข้อมูลแต่ละชั้น

X_i คือ จุดกึ่งกลางของแต่ละชั้น

k คือ จำนวนชั้นของข้อมูล

N คือ จำนวนข้อมูลทั้งหมดของประชากร

ค่าเฉลี่ยเลขคณิตของตัวอย่าง

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

โดยที่ $\bar{X}$ คือ ค่าเฉลี่ยเลขคณิตของตัวอย่าง

f_i คือ ความถี่ของข้อมูลแต่ละชั้น

X_i คือ จุดกึ่งกลางของแต่ละชั้น

k คือ จำนวนชั้นของข้อมูล

n คือ จำนวนข้อมูลทั้งหมดของตัวอย่าง

ตัวอย่าง 3.11 ค่าจ้างรายวันของคนงาน 50 คน แสดงข้อมูลดังต่อไปนี้ จงหาค่าเฉลี่ยของค่าจ้างรายวันของคนงาน

ค่าจ้าง (บาท)	**ความถี่** f_i
100-120	6
121-141	7
142-162	12
163-183	15
194-214	6
215-235	9
รวม	**55**

วิธีทำ สูตรที่ใช้ในการคำนวณคือ $\mu = \frac{\sum_{i=1}^{k} f_i X_i}{N}$

ค่าจ้าง (บาท)	ความถี่ f_i	จุดกึ่งกลาง X_i	$f_i X_i$
100-120	6	110	660
121-141	7	131	917
142-162	12	152	1,824
163-183	15	173	2,595
194-214	6	204	1,224
215-235	9	225	2,025
รวม	$n = 55$		$\sum_{i=1}^{k} f_i X_i = 9,245$

จะได้ว่า $\mu = \dfrac{\sum_{i=1}^{k} X_i f_i}{N} = \dfrac{9,245}{55} = 168.901$

ดังนั้น ค่าเฉลี่ยของค่าจ้างรายวันของคนงาน เท่ากับ 168.901 บาท

ตัวอย่าง 3.12 สุ่มตัวอย่างรถยนต์จำนวน 40 คัน สอบถามเวลาที่รถยนต์รอสัญญาณไฟจราจรก่อนที่จะผ่านสี่แยกแสงรุ้ง มีข้อมูลดังนี้

เวลาที่รอ (นาที)	ความถี่ f_i
5-10	5
11-16	6
17-22	9
23-28	7
29-34	5
35-40	8
รวม	**40**

จงหาเวลาเฉลี่ยที่รถยนต์ 40 คัน ใช้ในการรอสัญญาณไฟจราจร

วิธีทำ สูตรที่ใช้ในการคำนวณคือ $\bar{X} = \dfrac{\sum_{i=1}^{k} f_i X_i}{n}$

เวลาที่รอ (นาที)	ความถี่ f_i	จุดกึ่งกลาง X_i	$f_i X_i$
4-10	5	7	35
11-17	6	14	84
18-24	9	21	189
25-31	7	28	196
32-38	5	35	175
39-45	8	42	336
รวม	$n = 40$		$\sum_{i=1}^{k} f_i X_i = 1,015$

จะได้ว่า $\bar{X} = \dfrac{\sum_{i=1}^{k} X_i f_i}{n} = \dfrac{1,015}{40} = 25.375$

ดังนั้น เวลาเฉลี่ยที่รถยนต์ 40 คัน ใช้ในการรอสัญญาณไฟจราจร เท่ากับ 25.375 นาที

1.3 การหาค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก

การหาค่าเฉลี่ยเลขคณิตโดยทั่วไปให้ความสำคัญกับข้อมูลแต่ละค่าเท่ากัน ในกรณีที่ข้อมูลแต่ละค่ามีน้ำหนักความสำคัญไม่เท่ากัน จะการหาค่าเฉลี่ยเลขคณิตด้วยค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนัก เช่น การคำนวณคะแนนเฉลี่ยแต่ละภาคการศึกษา ซึ่งจะเห็นได้ว่ามีการให้ความสำคัญกับวิชาแต่ละวิชาไม่เท่ากัน โดยกำหนดเป็นจำนวนหน่วยกิตเพื่อใช้ในการถ่วงน้ำหนัก วิชาที่มีความสำคัญมากก็จะมีจำนวนหน่วยกิตมากตามไปด้วย ถ้าใช้การคำนวณค่าเฉลี่ยเลขคณิตแบบธรรมดาจะทำให้ได้ค่าเฉลี่ยที่คลาดเคลื่อนไปจากที่ควรจะเป็น (ปรีดาภรณ์ กาญจนสำราญวงศ์. 2560 : 62 และเกรียง กิจบำรุงรัตน์. 2562 : 50)

ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักของประชากร

$$\mu = \frac{\sum_{i=1}^{k} w_i X_i}{\sum_{i=1}^{k} w_i}$$

โดยที่ μ คือ ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักของประชากร

$\sum_{i=1}^{k} w_i X_i$ คือ ผลรวมระหว่างผลคูณของน้ำหนักกับข้อมูลลำดับที่ 1 ถึง k

$\sum_{i=1}^{k} w_i$ คือ ผลรวมของน้ำหนักลำดับที่ 1 ถึง k

ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักของตัวอย่าง

$$\bar{X} = \frac{\sum_{i=1}^{k} w_i X_i}{\sum_{i=1}^{k} w_i}$$

โดยที่ $\bar{X}$ คือ ค่าเฉลี่ยเลขคณิตแบบถ่วงน้ำหนักของตัวอย่าง

$\sum_{i=1}^{k} w_i X_i$ คือ ผลรวมระหว่างผลคูณของน้ำหนักกับข้อมูลลำดับที่ 1 ถึง k

$\sum_{i=1}^{k} w_i$ คือ ผลรวมของน้ำหนักลำดับที่ 1 ถึง k

ตัวอย่าง 3.13 ถ้านักศึกษาลงทะเบียนเรียน 3 รายวิชา คือ วิชาชีววิทยาพื้นฐาน 2 หน่วยกิต ได้เกรด 3 วิชาแคลคูลัส 3 หน่วยกิต ได้เกรด 2 และวิชาหลักสถิติ 3 หน่วยกิต ได้เกรด 4 จงหาเกรดเฉลี่ยของนักศึกษา

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$\mu = \frac{\sum_{i=1}^{k} w_i X_i}{\sum_{i=1}^{k} w_i}$$

โดยที่ $w_1 = 2, \quad w_1 = 3, \quad w_1 = 3$

$X_1 = 3, \quad X_1 = 2, \quad X_1 = 4$

แทนค่า

$$\mu = \frac{w_1X_1 + w_2X_2 + w_3X_3}{w_1 + w_2 + w_3}$$

$$= \frac{2\quad 3 + 3\quad 2 + 3\quad 4}{2+3+3}$$

$$= \frac{6+6+12}{8}$$

$$= \frac{24}{8}$$

$$= 3$$

ดังนั้น นักศึกษามีเกรดเฉลี่ย เท่ากับ 3

ตัวอย่าง 3.14 สุ่มพนักงานขายสินค้าของบริษัทแห่งหนึ่งมา 3 ทีม พบว่า ทีมที่ 1 ขายสินค้าได้ 20 ชิ้น มีพนักงานขาย 6 คน ทีมที่ 2 ขายสินค้าได้ 13 ชิ้น มีพนักงานขาย 5 คน และทีมที่ 3 ขายสินค้าได้ 8 ชิ้น มีพนักงานขาย 4 คน จงหาจำนวนสินค้าเฉลี่ยที่ขายได้

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$\bar{X} = \frac{\sum_{i=1}^{k} w_iX_i}{\sum_{i=1}^{k} w_i}$$

โดยที่ $w_1 = 6, \quad w_1 = 5, \quad w_1 = 4$

$X_1 = 20, \quad X_1 = 13, \quad X_1 = 8$

แทนค่า

$$\bar{X} = \frac{w_1X_1 + w_2X_2 + w_3X_3}{w_1 + w_2 + w_3}$$

$$= \frac{6\quad 20 + 5\quad 13 + 4\quad 8}{6+5+4}$$

$$= \frac{120+65+32}{15}$$

$$= \frac{217}{15}$$

$$= 14.467$$

ดังนั้น จำนวนสินค้าเฉลี่ยที่ขายได้ เท่ากับ 14.467

2. มัธยฐาน (Median)

มัธยฐาน คือ ค่าของข้อมูลที่อยู่ตรงกลาง เมื่อมีการเรียงข้อมูลตามลำดับจากมากไปหาน้อย หรือจากน้อยไปหามาก โดยมี 50% ของข้อมูลทั้งหมดที่มีสูงกว่ามัธยฐาน และมี 50% ของข้อมูลทั้งหมดที่มีต่ำกว่ามัธยฐาน ในการคำนวณหามัธยฐานสามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ข้อมูลไม่มีการแจกแจงความถี่ และกรณีที่ข้อมูลมีการแจกแจงความถี่ (มนตรี สังข์ทอง. 2557 : 48 และ ปรีดาภรณ์ กาญจนสำราญวงศ์. 2560 : 62)

2.1 การหาค่ามัธยฐานในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

ปรีดาภรณ์ กาญจนสำราญวงศ์ (2560 : 63) ได้กล่าวถึงขั้นตอนในการคำนวณการหาค่ามัธยฐานในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่ ไว้ดังนี้

1. เรียงลำดับค่าของข้อมูลจากมากไปหาน้อยหรือจากน้อยไปหามาก
2. นับจำนวนข้อมูลที่ต้องการหามัธยฐานว่าเป็นจำนวนคี่หรือจำนวนคู่
3. หาตำแหน่งของมัธยฐาน ซึ่งขึ้นอยู่ว่าเป็นจำนวนคี่หรือจำนวนคู่ มีสูตรดังนี้

กรณีเป็นจำนวนคี่ ตำแหน่งของมัธยฐาน คือ $\frac{n+1}{2}$

กรณีเป็นจำนวนคู่ ตำแหน่งของมัธยฐาน คือ $\frac{n}{2}$ และ $\frac{n}{2}+1$

4. หาค่ามัธยฐาน ซึ่งขึ้นอยู่ว่าเป็นจำนวนคี่หรือจำนวนคู่ ถ้าเป็นจำนวนคี่ คือ ข้อมูลที่ตรงกับตำแหน่งมัธยฐาน ถ้าเป็นจำนวนคู่ คือ ค่าเฉลี่ยของข้อมูลที่อยู่ตำแหน่งมัธยฐานทั้งสอง

ตัวอย่าง 3.15 สุ่มนักศึกษาที่เรียนรายวิชาหลักสถิติมาจำนวน 9 คน พบว่า มีคะแนนสอบรายวิชาสถิติดังต่อไปนี้

32 39 45 12 40 27 31 41 17

จงหามัธยฐานของคะแนนสอบของนักศึกษาทั้ง 9 คน

วิธีทำ 1. เรียงลำดับค่าของข้อมูลจากน้อยไปหามาก ดังนี้

12 17 27 31 32 39 40 41 45

2. นับจำนวนข้อมูลที่ใช้หามัธยฐาน พบว่า มี 9 จำนวน ซึ่งเป็นจำนวนคี่

3. ตำแหน่งมัธยฐาน คือ $\frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5$

4. ค่าของมัธยฐาน คือ ข้อมูลที่อยู่ในตำแหน่งที่ 5 เท่ากับ 32

ดังนั้น มัธยฐานของคะแนนสอบของนักศึกษาทั้ง 9 คน เท่ากับ 32

ตัวอย่าง 3.16 สุ่มตัวอย่างผู้มาใช้บริการที่ธนาคารแห่งหนึ่งจำนวน 10 คน พบว่าอายุแต่ละคนเป็นดังนี้

22 30 25 31 43 38 35 29 26 42

จงหามัธยฐานของผู้ที่มาใช้บริการที่ธนาคารแห่งนี้

วิธีทำ 1. เรียงลำดับค่าของข้อมูลจากน้อยไปหามาก ดังนี้

22 25 26 29 30 31 35 38 42 43

2. นับจำนวนข้อมูลที่ใช้หามัธยฐาน พบว่า มี 10 จำนวน ซึ่งเป็นจำนวนคี่

3. ตำแหน่งมัธยฐาน คือ $\frac{n}{2} = \frac{10}{2} = 5$ และ $\frac{n}{2}+1 = \frac{10}{2}+1 = 5+1 = 6$

4. ค่าของมัธยฐาน คือ ข้อมูลที่อยู่ในตำแหน่งที่ 5 เท่ากับ 30 และข้อมูลที่อยู่ในตำแหน่งที่ 6 เท่ากับ 31 จะได้ว่ามัธยฐาน คือ $\frac{30+31}{2} = \frac{61}{2} = 30.5$

ดังนั้น มัธยฐานของผู้ที่มาใช้บริการที่ธนาคารแห่งนี้ เท่ากับ 30.5

2.2 การหาค่ามัธยฐานในกรณีที่ข้อมูลมีการแจกแจงความถี่

เกรียง กิจบำรุงรัตน์ (2562 : 52) ได้กล่าวถึงขั้นตอนในการคำนวณการหาค่ามัธยฐานในกรณีที่ข้อมูลมีการแจกแจงความถี่ ไว้ดังนี้

1. หาความถี่สะสมจากชั้นที่ข้อมูลมีค่าน้อยจนถึงชั้นที่ข้อมูลมีค่ามาก
2. หาตำแหน่งของมัธยฐาน จากสูตร $\frac{n}{2}$
3. หาชั้นที่มัธยฐานอยู่
4. คำนวณหาค่ามัธยฐาน จากสูตรดังนี้

$$Med = L + \left(\frac{\frac{n}{2} - \sum f_L}{f_M} \right) I$$

โดยที่ L คือ ขอบเขตล่างของชั้นที่มีมัธยฐานอยู่

n คือ จำนวนข้อมูลทั้งหมด

$\sum f_L$ คือ ผลรวมของความถี่ทุกชั้นก่อนถึงชั้นที่มีมัธยฐานอยู่

f_M คือ ความถี่ของชั้นที่มัธยฐานอยู่

I คือ ความกว้างของชั้นที่มัธยฐานอยู่

ตัวอย่าง 3.17 สุ่มตัวอย่างผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี จำนวน 60 คน จำแนกตามอายุ ได้ข้อมูลดังนี้

อายุ (ปี)	จำนวนผู้ซื้อสินค้า
20-29	9
30-39	25
40-49	18
50-59	8

จงหามัธยฐานอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี

วิธีทำ

1. หาความถี่สะสมและพิจารณาความถี่สะสมของชั้นแรกที่สูงกว่าหรือเท่ากับจำนวนข้อมูลหารสอง จะได้ว่า $\frac{N}{2} = \frac{60}{2} = 30$

อายุ (ปี)	จำนวนผู้ซื้อสินค้า	ความถี่สะสม
20-29	9	9
30-39	25	34
40-49	18	52
50-59	8	60
รวม	$n = 60$	

2. คำนวณหาค่ามัธยฐาน จากสูตรดังนี้

$$Med = L + \left(\frac{\frac{n}{2} - \sum f_L}{f_M} \right) I$$

เมื่อ $L = 29.5, \quad \sum f_L = 9, \quad f_M = 25, \quad I = 10$

แทนค่า

$$\begin{aligned} Med &= 29.5 + \left(\frac{30 - 9}{25} \right) 10 \\ &= 29.5 + \left(\frac{21}{25} \right) 10 \\ &= 29.5 + 0.84 \quad 10 \\ &= 29.5 + 8.4 \\ &= 37.9 \end{aligned}$$

ดังนั้น มัธยฐานอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี เท่ากับ 37.9 ปี

3. ฐานนิยม (Mode)

ฐานนิยม คือ ค่าของข้อมูลที่มีจำนวนซ้ำมากที่สุด หรือมีความถี่สูงสุด ซึ่งอาจมีหรือไม่มีก็ได้ และบางครั้งข้อมูลชุดหนึ่งอาจมีมากกว่า 1 ค่าก็ได้ ซึ่งสามารถหาได้ทั้งข้อมูลเชิงคุณภาพและข้อมูลเชิงปริมาณ สำหรับการคำนวณหาฐานนิยม สามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ข้อมูลไม่มีการแจกแจงความถี่ และกรณีที่ข้อมูลมีการแจกแจงความถี่ (มนตรี สังข์ทอง. 2557 : 52 และ เกรียง กิจบำรุงรัตน์. 2562 : 54)

3.1 การหาค่าฐานนิยมในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

ฐานนิยม คือ ค่าของข้อมูลที่มีจำนวนซ้ำมากที่สุดหรือมีความถี่สูงสุด

ตัวอย่าง 3.18 จงหาฐานนิยมของข้อมูลแต่ละชุดต่อไปนี้

1. 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18

   ดังนั้น ฐานนิยม คือ 9

2. 3, 5, 8, 10, 12, 15, 16

   ดังนั้น ไม่มีฐานนิยม เนื่องจากไม่มีข้อมูลค่าใดมีความถี่สูงสุด

3. 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9

   ดังนั้น ฐานนิยม คือ 4 และ 7

4. ในการสำรวจความนิยมเลี้ยงสัตว์ของผู้รักสัตว์จำนวน 12 คน พบว่า มีข้อมูลสัตว์เลี้ยงไว้ ดังนี้

สุนัข	แมว	กระต่าย	ปลา	ปลา	สุนัข	ปลา	สุนัข
แมว	กระต่าย	สุนัข	ปลา	กระต่าย	ปลา	แมว	ปลา

ดังนั้น ฐานนิยม คือ ปลา

3.1 การหาค่าฐานนิยมในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

มนตรี สังข์ทอง (2557 : 52) ได้กล่าวถึงขั้นตอนในการคำนวณการหาค่าฐานนิยมในกรณีที่ข้อมูลมีการแจกแจงความถี่ ไว้ดังนี้

1. พิจารณาความถี่โดยชั้นที่มีความถี่สูงสุดเป็นชั้นที่มีฐานนิยมอยู่
2. คำนวณหาค่าฐานนิยม จากสูตรดังนี้

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

โดยที่ L คือ ขอบเขตล่างของชั้นที่มีฐานนิยมอยู่

d_1 คือ ผลต่างของความถี่ชั้นที่มีฐานนิยมกับความถี่ชั้นติดกันที่มีขีดจำกัดต่ำกว่า

d_2 คือ ผลต่างของความถี่ชั้นที่มีฐานนิยมกับความถี่ชั้นติดกันที่มีขีดจำกัดสูงกว่า

I คือ ความกว้างของชั้นที่ฐานนิยมอยู่

ตัวอย่าง 3.19 สุ่มตัวอย่างชายไทยอายุ 20-35 ปี จำนวน 60 คน เพื่อวัดปริมาณน้ำตาลในเลือด ได้ข้อมูลดังนี้

ปริมาณน้ำตาล (กรัม)	จำนวน
70-79	6
80-89	5
90-99	13
100-109	15
110-119	9
120-129	4
130-139	8

จงหาฐานนิยมของปริมาณน้ำตาลในเลือดในชายไทยอายุ 20-35 ปี

วิธีทำ 1. ฐานนิยมจะอยู่ในชั้นที่ 4 (ความถี่มากที่สุด)

2. คำนวณหาค่าฐานนิยม จากสูตรดังนี้

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

เมื่อ $L = 99.5, \quad d_1 = 15 - 13 = 2, \quad d_1 = 15 - 9 = 6, \quad I = 10$

แทนค่า

$$\begin{aligned} Mo &= 99.5 + \left(\frac{2}{2+6}\right) 10 \\ &= 99.5 + \left(\frac{2}{8}\right) 10 \\ &= 99.5 + \ 0.25 \ \ 10 \\ &= 99.5 + 2.5 \\ &= 102 \end{aligned}$$

ดังนั้น ฐานนิยมของปริมาณน้ำตาลในเลือดในชายไทยอายุ 20-35 ปี เท่ากับ 102 กรัม

การวัดการกระจาย

มนตรี สังข์ทอง (2557 : 56-57) และพิสมัย หาญมงคลพิพัฒน์ (2560 : 29-30) ได้กล่าวถึงการวัดแนวโน้มเข้าสู่ส่วนกลาง เพื่อใช้เป็นตัวแทนของข้อมูลในการสรุปถึงลักษณะของข้อมูล บางครั้งยังไม่เพียงพอเพราะอาจทำให้ตีความหมายของข้อมูลผิดพลาดไปได้ โดยเฉพาะอย่างยิ่งถ้ามีข้อมูลหลายชุดซึ่งมีค่าเฉลี่ยเท่ากัน อาจเข้าใจว่าข้อมูลที่มีค่าเฉลี่ยเท่ากันจะต้องมีลักษณะเหมือนกันทุกประการ ดังนี้

ข้อมูลชุดที่ 1 ประกอบด้วย 6, 6, 6, 6

ข้อมูลชุดที่ 2 ประกอบด้วย 5, 6, 5, 8

ข้อมูลชุดที่ 3 ประกอบด้วย 1, 9, 3, 11

จากข้อมูลทั้ง 3 ชุด จะเห็นได้ว่า ข้อมูลชุดที่ 1 ประกอบด้วยค่าที่มีขนาดเท่ากัน ข้อมูลชุดที่ 2 ประกอบด้วยค่าที่แตกต่างกันไม่มาก และข้อมูลชุดที่ 3 ประกอบด้วยค่าที่แตกต่างกันมาก ซึ่งเราจะเห็นได้ว่าข้อมูลทั้ง 3 ชุด ให้ค่าเฉลี่ยเท่ากัน คือเท่ากับ 6 ดังนั้นถ้าพิจารณาเฉพาะค่าวัดแนวโน้มเข้าสู่ส่วนกลางเพียงค่าเดียวเป็นตัวแทนของข้อมูลทั้งหมดอาจทำให้ได้สารสนเทศไม่ครบถ้วน เพื่อขจัดปัญหาการอธิบายคุณลักษณะของข้อมูลชุดหนึ่ง ๆ นั้น จะต้องวัดทั้งค่าแนวโน้มเข้าสู่ส่วนกลางและค่าวัดการกระจายของข้อมูลชุดนั้นควบคู่กันไป

ค่าวัดการกระจาย แสดงให้เห็นว่า ถ้าข้อมูลมีค่าใกล้เคียงหรือแตกต่างกันไม่มาก ค่าวัดการกระจายจะมีค่าค่อนข้างต่ำ แต่ถ้าข้อมูลมีความแตกต่างกันมาก ค่าวัดการกระจายจะมีค่าค่อนข้างสูง ซึ่งในการวัดการกระจายของข้อมูลที่นิยมใช้มี 4 วิธี ได้แก่ พิสัย (Range) ความแปรปรวน (Variance) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) และสัมประสิทธิ์การกระจาย (Coefficient of Variation)

1. พิสัย (Range)

พิสัย หมายถึง ความแตกต่างระหว่างค่าสูงสุดและกับต่ำสุดของข้อมูล ถ้าพิสัยมีค่ามากแสดงว่าข้อมูลมีการกระจายมาก แต่ถ้าพิสัยมีค่าน้อยแสดงว่าข้อมูลมีการกระจายน้อย (ปรีดาภรณ์ กาญจนสำราญวงศ์. 2560 : 83) พิสัยเป็นค่าวัดการกระจายของข้อมูลที่คำนวณง่ายสุด แต่ค่อนข้างหยาบ เนื่องจากใช้ข้อมูลเพียง 2 ค่า คือ ค่าสูงสุดและค่าต่ำสุดของข้อมูล ดังนั้นการนำมาใช้ต้องพิจารณาข้อมูลชุดนั้นมีค่าผิดปกติหรือไม่ เนื่องจากถ้ามีค่าผิดปกติคือ ข้อมูลมีค่าสูงหรือต่ำเกินไปจะทำให้พิสัยคลาดเคลื่อนจากที่เป็นจริง ดังนั้นพิสัยจึงเหมาะสำหรับการพิจารณาการกระจายของข้อมูล

อย่างคร่าว ๆ สำหรับการคำนวณหาพิสัย สามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ข้อมูลไม่มีการแจกแจงความถี่ และกรณีที่ข้อมูลมีการแจกแจงความถี่ (มนตรี สังข์ทอง. 2557 : 57)

1.1 การหาค่าพิสัยในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

พิสัย = ค่าสูงสุดของข้อมูล – ค่าต่ำสุดของข้อมูล

ตัวอย่าง 3.20 จงหาพิสัยของน้ำหนักของนักศึกษาของคณะวิทยาศาสตร์ชั้นปีที่ 1 ซึ่งมีน้ำหนัก ดังนี้

43 45 51 49 63 55 61 52 46

วิธีทำ พิสัย = ค่าสูงสุดของข้อมูล – ค่าต่ำสุดของข้อมูล

= 63 – 43

= 20

ดังนั้น พิสัยของน้ำหนักของนักศึกษาของคณะวิทยาศาสตร์ชั้นปีที่ 1 เท่ากับ 20

1.2 การหาค่าพิสัยในกรณีที่ข้อมูลมีการแจกแจงความถี่

พิสัย = ขีดจำกัดชั้นบนที่แท้จริงชั้นสูงสุด – ขีดจำกัดชั้นล่างที่แท้จริงชั้นต่ำสุด

ตัวอย่าง 3.21 สุ่มตัวอย่างผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี จำนวน 60 คน จำแนกตามอายุ ได้ข้อมูลดังนี้

อายุ (ปี)	**จำนวนผู้ซื้อสินค้า**
20-29	9
30-39	25
40-49	18
50-59	8

จงหาพิสัยอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี

วิธีทำ ขีดจำกัดชั้นล่างที่แท้จริงชั้นต่ำสุด = 20 – 0.5 = 19.5

ขีดจำกัดชั้นบนที่แท้จริงชั้นสูงสุด = 59 + 0.5 = 59.5

พิสัย = ขีดจำกัดชั้นบนที่แท้จริงชั้นสูงสุด – ขีดจำกัดชั้นล่างที่แท้จริงชั้นต่ำสุด

= 59.5 – 19.5

= 40

ดังนั้น พิสัยอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี เท่ากับ 40 ปี

2. ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน (Variance and Standard Deviation)

ปรีดาภรณ์ กาญจนสำราญวงศ์ (2560 : 85) และ พิศมัย หาญมงคลพิพัฒน์ (2560 : 31) ได้กล่าวว่า ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดการกระจายที่นิยมใช้ในทางสถิติมากที่สุด และถือว่าเป็นวิธีที่ดีที่สุดวิธีหนึ่ง เพราะเป็นวิธีการวัดการกระจายที่นำทุกค่าของข้อมูลมาร่วมในการคำนวณ

ความแปรปรวน คือ ค่าเฉลี่ยของความแตกต่างระหว่างค่าข้อมูลแต่ละค่ากับค่าเฉลี่ยของข้อมูลชุดนั้นยกกำลังสอง โดยใช้สัญลักษณ์ σ^2 แทนความแปรปรวนของประชากร และ S^2 แทนความแปรปรวนของตัวอย่าง

ส่วนเบี่ยงเบนมาตรฐาน คือ รากที่สองของความแปรปรวน โดยใช้สัญลักษณ์ σ แทนส่วนเบี่ยงเบนมาตรฐานของประชากร และ S แทนส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

สำหรับในการคำนวณความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน สามารถแบ่งออกเป็น 2 กรณี คือ กรณีที่ข้อมูลไม่มีการแจกแจงความถี่ และกรณีที่ข้อมูลมีการแจกแจงความถี่

2.1 การหาคำนวณความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่

การหาคำนวณความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานในกรณีที่ข้อมูลไม่มีการแจกแจงความถี่ แบ่งออกเป็น 2 ชนิด ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของประชากร และความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง ซึ่งมีสูตรในการคำนวณดังนี้

ความแปรปรวนของประชากร

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

หรือ

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

ส่วนเบี่ยงเบนมาตรฐานของประชากร

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

หรือ

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

โดยที่	σ^2	คือ	ความแปรปรวนของประชากร
	σ	คือ	ส่วนเบี่ยงเบนมาตรฐานของประชากร
	μ	คือ	ค่าเฉลี่ยของประชากร
	N	คือ	จำนวนข้อมูลทั้งหมดของประชากร

ความแปรปรวนของตัวอย่าง

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

หรือ

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

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

หรือ

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

โดยที่ S^2 คือ ความแปรปรวนของตัวอย่าง

S คือ ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

$\bar{X}$ คือ ค่าเฉลี่ยของตัวอย่าง

n คือ จำนวนข้อมูลทั้งหมดของตัวอย่าง

ตัวอย่าง 3.22 จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลประชากรต่อไปนี้

2 3 4 4 5 6

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

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

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

$$= \frac{2+3+4+4+5+6}{6}$$

$$= \frac{24}{6}$$

$$= 4$$

แทนค่า

$$\sigma^2 = \frac{(2-4)^2 + (3-4)^2 + (4-4)^2 + (4-4)^2 + (5-4)^2 + (6-4)^2}{6}$$

$$= \frac{4+1+0+0+1+4}{6}$$

$$= \frac{10}{6}$$

$$= 1.67$$

หรือ

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

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

$$= 4+9+16+16+25+36$$

$$= 106$$

แทนค่า

$$\sigma^2 = \frac{106}{6} - (4)^2$$

$$= 17.67 - 16$$

$$= 1.67$$

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

$$= \sqrt{1.67}$$

$$= 1.29$$

ดังนั้น ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลประชากร เท่ากับ 1.67 และ 1.29 ตามลำดับ

ตัวอย่าง 3.23 สุ่มตัวอย่างนักศึกษาสาขาวิชาสถิติ จำนวน 6 คน และสอบถามจำนวนรายวิชาที่ลงทะเบียนเรียนในภาคการศึกษาที่ 1/2564 ได้ข้อมูลดังนี้

5 6 3 4 2 5

จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของจำนวนรายวิชาที่ลงทะเบียนเรียนของนักศึกษา

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$S^2 = \frac{\sum_{i=1}^{n} \; X_i - \bar{X} \;^2}{n-1}$$

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

$$= \frac{5+6+3+4+2+5}{6}$$

$$= \frac{25}{6}$$

$$= 4.17$$

แทนค่า

$$S^2 = \frac{\; 5-4.17 \;^2 + \; 6-4.17 \;^2 + \; 3-4.17 \;^2 + \ldots + \; 5-4.17 \;^2}{6-1}$$

$$= \frac{0.6889+3.3489+1.3689+\ldots+0.6889}{5}$$

$$= \frac{10.8334}{5}$$

$$= 2.17$$

หรือ

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

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

$$= 25+36+9+16+4+25$$

$$= 115$$

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

$$= 25$$

แทนค่า

$$S^2 = \frac{6\ 115\ -\ 25^{\ 2}}{6\ 6-1}$$

$$= \frac{690-625}{6\ 5}$$

$$= \frac{690-625}{30}$$

$$= \frac{65}{30}$$

$$= 2.17$$

$$S^2 = \frac{6\ 115\ -\ 25^{\ 2}}{6\ 6-1}$$

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

$$= \sqrt{2.17}$$

$$= 1.47$$

ดังนั้น ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของข้อมูลประชากร เท่ากับ 2.17 และ 1.47 ตามลำดับ

2.2 การหาคำนวณความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานในกรณีที่ข้อมูลมีการแจกแจงความถี่

การหาคำนวณความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานในกรณีที่ข้อมูลมีการแจกแจงความถี่ แบ่งออกเป็น 2 ชนิด ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของประชากร และความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง ซึ่งมีสูตรในการคำนวณดังนี้

ความแปรปรวนของประชากร

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

หรือ

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

ส่วนเบี่ยงเบนมาตรฐานของประชากร

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

หรือ

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

โดยที่	σ^2	คือ	ความแปรปรวนของประชากร
	σ	คือ	ส่วนเบี่ยงเบนมาตรฐานของประชากร
	f_i	คือ	ความถี่ของชั้นที่ i
	X_i	คือ	จุดกึ่งกลางของชั้นที่ i
	μ	คือ	ค่าเฉลี่ยของประชากร
	N	คือ	จำนวนข้อมูลทั้งหมดของประชากร

ความแปรปรวนของตัวอย่าง

$$S^2 = \frac{\sum_{i=1}^{k} f_i \left(X_i - \bar{X}\right)^2}{n-1}$$

หรือ

$$S^2 = \frac{n\sum_{i=1}^{k} f_i X_i^2 - \left(\sum_{i=1}^{k} f_i X_i\right)^2}{n(n-1)}$$

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i \left(X_i - \bar{X}\right)^2}{n-1}}$$

หรือ

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

โดยที่ S^2 คือ ความแปรปรวนของตัวอย่าง

S คือ ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง

f_i คือ ความถี่ของชั้นที่ i

X_i คือ จุดกึ่งกลางของชั้นที่ i

$\bar{X}$ คือ ค่าเฉลี่ยของตัวอย่าง

n คือ จำนวนข้อมูลทั้งหมดของตัวอย่าง

ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐาน เป็นค่าที่บอกให้ทราบว่าค่าต่าง ๆ ของข้อมูลภายในชุดนั้นมีความแตกต่างมากน้อยเพียงใด ถ้าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานที่คำนวณได้มีค่ามาก แสดงว่าค่าต่าง ๆ ในข้อมูลชุดนั้นมีค่าแตกต่างกันมากหรือมีการกระจายมาก ถ้าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานที่คำนวณได้มีค่าน้อย แสดงว่าค่าต่าง ๆ ในข้อมูลชุดนั้นมีค่าใกล้เคียงกันหรือมีการกระจายน้อย และถ้าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานที่คำนวณได้มีเท่ากับศูนย์ แสดงว่าค่าต่าง ๆ ในข้อมูลชุดนั้นมีค่าเท่ากันหมดหรือไม่มีการกระจาย (พิศมัย หาญมงคลพิพัฒน์. 2560 : 34-35)

ตัวอย่าง 3.24 จากการสำรวจอายุของเด็กที่มาใช้บริการในโรงพยาบาลแห่งหนึ่ง ได้ข้อมูลดังนี้

อายุ (ปี)	**จำนวนเด็ก**
1-3	3
4-6	3
7-9	4
10-12	2
13-15	3

จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของอายุเด็กที่มาใช้บริการในโรงพยาบาลแห่งหนึ่ง

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

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

$$\mu = \frac{\sum_{i=1}^{k} X_i f_i}{N} = \frac{117}{15} = 7.8$$

อายุ (ปี)	f_i	X_i	X_i^2	$X_i f_i$	$\left(X_i - \mu\right)^2 f_i$	$X_i^2 f_i$
1-3	3	2	4	6	100.92	12
4-6	3	5	25	15	23.52	75
7-9	4	8	64	32	0.16	256
10-12	2	11	121	22	20.48	242
13-15	3	14	196	42	115.32	588
รวม	15		410	117	260.4	1,173

แทนค่า

$$\sigma^2 = \frac{260.4}{15} = 17.36$$

หรือ

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

$$= \frac{1,173}{15} - (7.8)^2$$

$$= 78.20 - 60.84$$

$$= 17.36$$

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

$$= \sqrt{17.36}$$

$$= 4.17$$

ดังนั้น ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของอายุของอายุเด็กที่มาใช้บริการในโรงพยาบาลแห่งหนึ่งเท่ากับ 17.36 ปี^{2} และ 4.17 ปี ตามลำดับ

ตัวอย่าง 3.25 สุ่มตัวอย่างผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี จำนวน 60 คน จำแนกตามอายุ ได้ข้อมูลดังนี้

อายุ (ปี)	จำนวนผู้ซื้อสินค้า
20-29	9
30-39	25
40-49	18
50-59	8

จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวี

วิธีทำ สูตรที่ใช้ในการคำนวณคือ

$$S^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \bar{X})^2}{n-1}$$

$$\bar{X} = \frac{\sum_{i=1}^{k} X_i f_i}{n}$$

$$= \frac{2,320}{60}$$

$$= 38.67$$

อายุ (ปี)	f_i	X_i	X_i^2	$X_i f_i$	$(X_i - \bar{X})^2 f_i$	$X_i^2 f_i$
20-29	9	24.5	600.25	220.5	1,807.10	5,402.25
30-39	25	34.5	1,190.25	862.5	434.72	29,756.25
40-49	18	44.5	1,980.25	801	611.80	35,644.5
50-59	8	54.5	2,970.25	436	2,004.71	23,762.0
รวม	60		6,741	2,320	4,858.33	94,565

แทนค่า

$$S^2 = \frac{4,858.33}{60-1}$$

$$= \frac{4,858.33}{59}$$

$$= 82.34$$

หรือ

$$S^2 = \frac{n\sum_{i=1}^{k} f_i X_i^2 - \left(\sum_{i=1}^{k} f_i X_i\right)^2}{n(n-1)}$$

$$= \frac{60(94,565) - (2,320)^2}{60(60-1)}$$

$$= \frac{5,673,900 - 5,382,400}{60(59)}$$

$$= \frac{5,673,900 - 5,382,400}{3,540}$$

$$= \frac{291,500}{3,540}$$

$$= 82.34$$

$$S = \sqrt{S^2}$$
$$= \sqrt{82.34}$$
$$= 9.07$$

ดังนั้น ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของอายุของผู้ซื้อสินค้าภายหลังรับชมรายการส่งเสริมการขายผ่านเคเบิลทีวีเท่ากับ 82.34 ปี^{2} และ 9.07 ปี ตามลำดับ

สัมประสิทธิ์การผันแปร

การวัดการกระจายของข้อมูลโดยใช้ค่าพิสัย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐาน ที่กล่าวมาแล้วเป็นการวัดการกระจายของข้อมูลชุดเดียวกัน ในกรณีที่ต้องการเปรียบเทียบการกระจายของข้อมูลตั้งแต่ 2 ชุดขึ้นไป โดยที่ข้อมูลแต่ละชุดนั้นที่หน่วยการวัดที่ต่างกันหรือมีค่าเฉลี่ยแตกต่างกันมาก จะใช้สัมประสิทธิ์การผันแปรในการเปรียบเทียบการกระจายของข้อมูลชุดนั้น ๆ การคำนวณสัมประสิทธิ์การผันแปรคือ อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ย โดยคิดเป็นร้อยละ เขียนแทนด้วยสัญลักษณ์ CV แบ่งออกเป็น 2 ชนิด คือ สัมประสิทธิ์การผันแปรของประชากร และสัมประสิทธิ์การผันแปรของตัวอย่าง ดังนี้ (มนตรี สังข์ทอง. 2557 : 71-72 ; พิศมัย หาญมงคลพิพัฒน์. 2560 : 36-37 และเกรียง กิจบำรุงรัตน์. 2562 : 66)

สัมประสิทธิ์การผันแปรของประชากร

$$CV = \frac{\sigma}{\mu} \times 100$$

สัมประสิทธิ์การผันแปรของตัวอย่าง

$$CV = \frac{S}{\bar{X}} \times 100$$

ถ้าข้อมูลชุดใดที่มีสัมประสิทธิ์การผันแปรมากกว่า แสดงว่าข้อมูลชุดนั้นมีการกระจายมาก แต่ถ้าต้องการเปรียบเทียบว่าข้อมูลชุดใดดีกว่าหรือมีประสิทธิภาพมากกว่า จะต้องดูข้อมูลชุดที่ค่าสัมประสิทธิ์การผันแปรน้อยกว่า

ตัวอย่าง 3.26 จากการสำรวจส่วนสูงและน้ำหนักของนักเรียนกลุ่มหนึ่ง พบว่า ส่วนสูงของนักเรียนมีค่าเฉลี่ย 160 เซนติเมตร ส่วนเบี่ยงเบนมาตรฐาน 25 เซนติเมตร ส่วนน้ำหนักของนักเรียนมีค่าเฉลี่ย 55 กิโลกรัม ส่วนเบี่ยงเบนมาตรฐาน 6 กิโลกรัม จงเปรียบเทียบการกระจายของส่วนสูงและน้ำหนักของนักเรียนกลุ่มนี้

วิธีทำ ส่วนสูง : $\mu = 160$ เซนติเมตร และ $\sigma = 25$ เซนติเมตร

น้ำหนัก : $\mu = 55$ กิโลกรัม และ $\sigma = 6$ กิโลกรัม

จะได้ว่า สัมประสิทธิ์การผันแปรของส่วนสูง

$$\begin{aligned} CV &= \frac{\sigma}{\mu} \times 100 \\ &= \frac{25}{160} \times 100 \\ &= 15.625 \end{aligned}$$

สัมประสิทธิ์การผันแปรของน้ำหนัก

$$\begin{aligned} CV &= \frac{\sigma}{\mu} \times 100 \\ &= \frac{6}{55} \times 100 \\ &= 10.909 \end{aligned}$$

ดังนั้น สัมประสิทธิ์การผันแปรของส่วนสูงมากกว่าสัมประสิทธิ์การผันแปรของน้ำหนัก แสดงว่า ส่วนสูงมีการกระจายมากกว่าน้ำหนัก

สรุป

การวิเคราะห์ข้อมูลเบื้องต้นเป็นการนำข้อมูลที่เก็บรวบรวมได้ทั้งเชิงปริมาณและคุณภาพนำมาวิเคราะห์เพื่อสรุปลักษณะโดยรวมของข้อมูลเกี่ยวกับเรื่องนั้น ๆ ซึ่งในบทนี้จะกล่าวถึง การวัดตำแหน่งเปรียบเทียบจะทำให้ทราบถึงลักษณะตำแหน่งของข้อมูลแต่ละตัว ซึ่งได้แก่ อัตราส่วน สัดส่วน ร้อยละ ควอไทล์ เดไซล์ เปอร์เซ็นไทล์ การวัดแนวโน้มเข้าสู่ส่วนกลางเป็นการหาค่ากลางของข้อมูล เพื่อนำไปใช้เป็นตัวแทนของข้อมูลชุดนั้น ๆ ซึ่งได้แก่ ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม โดยแต่ละค่าจะมีข้อดีและข้อเสียที่แตกต่างกันไป ดังนั้นควรเลือกใช้ค่ากลางที่เหมาะสมกับข้อมูลชุดนั้น ๆ ในการคำนวณค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม จะพิจารณาจากข้อมูลทั้งกรณีที่ข้อมูลไม่มีการแจกแจงความถี่และข้อมูลมีการแจกแจงความถี่ นอกจากนี้ยังกล่าวถึงการวัดการกระจายเพื่อใช้อธิบายลักษณะการกระจายของข้อมูลว่าข้อมูลกระจายมากหรือกระจายน้อย ซึ่งได้แก่ พิสัย ความแปรปรวน ส่วนเบี่ยงเบนมาตรฐาน โดยการหาค่าความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานจะแบ่งออกเป็น 2 กรณี คือข้อมูลไม่มีการแจกแจงความถี่และข้อมูลมีแจกแจงความถี่ และท้ายสุดได้กล่าวถึงสัมประสิทธิ์การกระจายด้วย