# ใบความรู้ที่ 7
## เรื่อง การวัดการกระจายสัมบูรณ์ (ส่วนเบี่ยงเบนมาตรฐาน/ความแปรปรวน)

การวัดการกระจายของข้อมูลโดยใช้ส่วนเบี่ยงเบนมาตรฐาน เป็นวิธีที่นักสถิติยอมรับว่าเป็นวิธีที่ใช้วัดการกระจายได้ดีที่สุด เมื่อเปรียบเทียบกับวิธีวัดการกระจายทั้งสามวิธีที่ได้กล่าวมาแล้ว คือ พิสัย ส่วนเบี่ยงเบนควอร์ไทล์ และส่วนเบี่ยงเบนเฉลี่ย ทั้งนี้ เนื่องจากการวัดการกระจายโดยวิธีนี้ใช้ข้อมูลทุกๆค่า หรือมีตัวแทนของข้อมูลทุกค่ามาคำนวณ และขจัดปัญหาในการที่ต้องใช้ค่าสัมบูรณ์ให้หมดไป การวัดการกระจายโดยวิธีนี้ นอกจากจะให้ค่าการกระจายที่มีความละเอียดถูกต้อง และเชื่อถือได้มากที่สุดแล้ว ยังสามารถนำไปใช้ในการวิเคราะห์ข้อมูลสถิติในชั้นสูงต่อไป ซึ่งการวัดการกระจายแบบอื่นนำไปใช้ไม่ได้

**การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่ได้แจกแจงความถี่**

ส่วนเบี่ยงเบนมาตรฐาน คือ ค่าที่ใช้วัดการกระจายที่ได้จากการหารรากที่สอง ของค่าเฉลี่ยกำลังสอง ของผลต่างระหว่างค่าของข้อมูลแต่ละค่า กับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น

ถ้าให้ $x_1, x_2, x_3, \ldots, x_N$ เป็นข้อมูลของประชากร N หน่วย และมีค่าเฉลี่ยเลขคณิต เป็น $\mu$ แล้ว ส่วนเบี่ยงเบนมาตรฐานของประชากรหรือ $\sigma$ สามารถคำนวณได้ดังนี้

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} \quad \text{หรือ} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2}$$

โดยที่ $\mu$ แทนค่าเฉลี่ยเลขคณิตของประชากร

N แทนจำนวนข้อมูลทั้งหมดของประชากร

กรณีที่ไม่สามารถศึกษาข้อมูลทั้งหมดของประชากร ส่วนเบี่ยงเบนมาตรฐานใช้สัญลักษณ์ S และ S.D.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \quad \text{หรือ} \quad s = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}{n-1}}$$

โดยที่ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของตัวอย่าง

n แทนจำนวนข้อมูลทั้งหมดของตัวอย่าง

**หมายเหตุ** :: ในหนังสือเรียนของ สสวท. ใช้สูตรที่มีตัวหาร n-1

ลำดับขั้นตอนการหาส่วนเบี่ยงเบนมาตรฐานโดยทั่วไป มีดังนี้

1. หาค่าเฉลี่ยเลขคณิต $(\overline{x})$
2. หาผลต่างระหว่างข้อมูลกับค่าเฉลี่ยเลขคณิต $\left(x_i - \overline{x}\right)$
3. หารกำลังสองของผลต่าง ในข้อ $\left(x_i - \overline{x}\right)^2$
4. หาผลบวกของกำลังสองของผลต่าง ในข้อ 3 $\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
5. หารผลบวกในข้อ 4 ด้วยจำนวนข้อมูล $\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}}$
6. หารากที่สองที่ไม่ใช้จำนวนลบของผลหารในข้อ 5

**ตัวอย่างที่ 1** จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูล 5 , 6 , 6 , 6 , 7 , 8

**วิธีทำ**

| $x_i$ | $x_i - \overline{x}$ | $(x_i - \overline{x})^2$ |
|---|---|---|
| 5 | - 1.3 | 1.69 |
| 6 | - 0.3 | 0.09 |
| 6 | - 0.3 | 0.09 |
| 6 | 0.3 | 0.09 |
| 7 | 0.7 | 0.49 |
| 8 | 1.7 | 2.89 |
| $\sum_{i=1}^{N} x_i = 38$ | | $\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 = 5.34$ |

จากสูตร $\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$

แทนค่า จะได้ $\overline{x} = \frac{38}{6} = 6.3$

จากสูตร $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}}$

แทนค่า จะได้ $S = \sqrt{\frac{5.34}{6}}$

$S = \sqrt{0.89} = 0.94$

**ตัวอย่างที่ 2** จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูล 1 , 2 , 4 , 7 , 9 , 12

**วิธีทำ**

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| 1 | - 4.8 | 23.04 |
| 2 | - 3.8 | 14.44 |
| 4 | - 1.8 | 3.24 |
| 7 | 1.2 | 1.44 |
| 9 | 3.2 | 10.24 |
| 12 | 6.2 | 38.44 |
| $\sum_{i=1}^{N} x_i = 35$ | | $\sum_{i=1}^{N}(x_i - \bar{x})^2 = 90.84$ |

จากสูตร $\bar{x} = \sum_{i=1}^{N} x_i$

แทนค่า จะได้ $\bar{x} = \frac{35}{6} = 5.8$

จากสูตร $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}}$

แทนค่า จะได้ $S = \sqrt{\frac{90.84}{6}}$

$S = 3.90$

**การหาส่วนเบี่ยงเบนมาตรฐานของประชากรของข้อมูลที่แจกแจงความถี่แล้ว**

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่แจกแจงความถี่แล้ว เป็นการหาค่าโดยประมาณ เพราะข้อมูลไม่ใช่ข้อมูลที่แน่นอน ซึ่งจะใช้วิธีเดียวกันกับการหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่แจกแจงความถี่ ดังนี้

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

เมื่อ $x_i$ แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

$f_i$ แทนความถี่ของอันตรภาคชั้นที่ i

k แทนจำนวนอันตรภาคชั้น

N แทนจำนวนของข้อมูลทั้งหมด หรือผลรวมของความถี่ของทุกๆอันตรภาคชั้น

$\mu$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลทั้งหมด

**การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างที่แจกแจงความถี่แล้ว**

การหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างของข้อมูลที่มีจำนวนมากหรือน้อยก็ตาม ทำได้ทำนองเดียวกันกับกรณีของข้อมูลที่ไม่แจกแจงความถี่ และในปัจจุบัน ถ้าเรามีข้อมูลดิบทุกหน่วย จะสามารถใช้เครื่องคอมพิวเตอร์ช่วยคำนวณได้สะดวก อย่างไรก็ตาม ถ้ามีกรณีที่ข้อมูลของเราไม่ใช่ข้อมูลดิบทุกหน่วย แต่เป็นข้อมูลที่อาจมาจากแหล่งทุติยภูมิอื่นๆ ซึ่งข้อมูลมีการแจกแจงความถี่เป็นอันตรภาคชั้นหรือเป็นกลุ่มมาแล้ว สามารถใช้สูตรการหาส่วนเบี่ยงเบนมาตรฐาน เป็นดังนี้

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

เมื่อ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลตัวอย่าง

n แทนจำนวนตัวอย่างทั้งหมด

k เป็นจำนวนอันตรภาคชั้นหรือจำนวนกลุ่ม

$x_i$ แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

ทั้งนี้เราสามารถใช้ N แทน n-1 ได้

**ตัวอย่างที่ 3** จงหาส่วนเบี่ยงเบนมาตรฐาน จากตารางแจกแจงความถี่ต่อไปนี้

| คะแนน | 10 – 14 | 15 – 19 | 20 – 24 |
|---|---|---|---|
| ความถี่ | 3 | 5 | 2 |

**วิธีทำ** ใช้สูตร $S = \sqrt{\frac{\sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2}{N}}$

| คะแนน | x | f | fx | $(x - \bar{x})^2$ | $f(x - \bar{x})^2$ |
|---|---|---|---|---|---|
| 10 – 14 | 15 | 3 | 36 | 20.25 | 60.75 |
| 15 – 19 | 17 | 5 | 85 | 0.25 | 1.25 |
| 20 – 24 | 22 | 2 | 44 | 30.25 | 60.25 |
| | | N = 10 | $\sum fx = 165$ | | $\sum f(x - \bar{x})^2 = 122.50$ |

จากสูตร $\bar{x} = \frac{\sum fx}{N} = \frac{165}{10} = 16.5$

$\therefore \quad S = \sqrt{\frac{122.50}{10}} = \sqrt{12.25} = 3.5$

**ความแปรปรวนของข้อมูล**

**ความแปรปรวนของประชากร (population variance)**

ความแปรปรวน คือ ส่วนเบี่ยงเบนมาตรฐานยกกำลังสอง ความแปรปรวนของข้อมูลประชากรที่ไม่ได้แจกแจงความถี่ หาได้โดยใช้สูตร

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

และความแปรปรวนของข้อมูลที่แจกแจงความถี่แล้ว หาได้โดยใช้สูตร

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

เมื่อ $k$ แทนจำนวนอันตรภาคชั้น

$x_i$ แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

**ความแปรปรวนของตัวอย่าง (sample variance)**

ความแปรปรวนตัวอย่างทั้งกรณีไม่แจกแจงความถี่และแจกแจงความถี่คือกำลังสองของส่วนเบี่ยงเบนมาตรฐานของข้อมูลตัวอย่างในทั้งสองกรณีตามลำดับ

ดังนั้น ความแปรปรวนของตัวอย่างคำนวณได้ดังนี้

**กรณีไม่แจกแจงความถี่**

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X}_i)}{n-1} = \frac{\sum_{i=1}^{N} x_i^2 - n\overline{X}^2}{n-1}$$

**กรณีมีการแจกแจงความถี่**

หรือ $$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{X})^2}{n-1}$$

$$= \frac{\sum_{i=1}^{N} f_i x_i^2 - n\overline{X}^2}{n-1}$$

โดยที่ $f_i$ แทนความถี่ของอันตรภาคชั้นที่ i

$x_i$ แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

$k$ แทนจำนวนอันตรภาคชั้น

$n$ แทนจำนวนตัวอย่างทั้งหมด

**ตัวอย่างที่ 4** จงหาส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนจากข้อมูลที่เกี่ยวกับปริมาณน้ำฝนของจังหวัดต่าง ๆ ในภาคตะวันออกเฉียงเหนือในปี พ.ศ. 2545 ดังตาราง

| **จังหวัด** | **ปริมาณน้ำฝน** |
|---|---|
| ขอนแก่น | 1,402.6 |
| ชัยภูมิ | 927.5 |
| นครพนม | 2,995.9 |
| มุกดาหาร | 1,901.7 |
| ร้อยเอ็ด | 1,357.2 |
| เลย | 1,414.8 |
| สกลนคร | 1,888.6 |
| สุรินทร์ | 1,857.9 |
| หนองคาย | 2,247.5 |
| อุดรธานี | 1,777.0 |

**ที่มา** : กรมอุตุนิยมวิทยา กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร

**วิธีทำ** ปริมาณน้ำฝนแยกตามจังหวัด

| **จังหวัด** | **ปริมาณน้ำฝน** | $x_i - \overline{X}$ | $(x_i - \overline{X})^2$ |
|---|---|---|---|
| ขอนแก่น | 1,402.6 | –374.47 | 140227.78 |
| ชัยภูมิ | 927.5 | –849.57 | 721769.19 |
| นครพนม | 2,995.9 | 1218.83 | 1485546.57 |
| มุกดาหาร | 1,901.7 | 124.63 | 15532.64 |
| ร้อยเอ็ด | 1,357.2 | –419.87 | 176290.82 |
| เลย | 1,414.8 | –362.27 | 131239.55 |
| สกลนคร | 1,888.6 | 111.53 | 12438.94 |
| สุรินทร์ | 1,857.9 | 80.83 | 6533.49 |
| หนองคาย | 2,247.5 | 470.43 | 221304.39 |
| อุดรธานี | 1,777.0 | –0.07 | 0.005 |
| **รวม** | 17770.7 | | 2910883.36 |

จากตัวอย่างที่ 3 จะได้ $\overline{X}$ = 1,777.07 มิลลิเมตร

เนื่องจากความแปรปรวน $s^2 = \dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n-1} = \dfrac{2910883.36}{10-1}$

$= 323{,}431.49$ มิลลิเมตร

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 568.71 มิลลิเมตร

**ตัวอย่างที่ 5** จงหาความแปรปรวนของราคาสินค้าชนิดหนึ่งที่ขายตามร้านต่าง ๆ ในสองท้องที่

ท้องที่ที่หนึ่ง (บาท) 50 52 45 55 54 48 53

ท้องที่ที่สอง (บาท) 40 50 51 52 51 51 62 53 49

**วิธีทำ** ราคาสินค้าชนิดหนึ่งที่ขายตามร้านต่าง ๆ ในสองท้องที่

| | **ราคา (บาท)** |
|---|---|
| **ท้องที่ที่หนึ่ง** | 50 52 45 55 54 48 53 |
| **ท้องที่ที่สอง** | 40 50 51 52 51 51 62 53 49 |

หา $\overline{X}$ และ $s^2$ ของทั้งสองท้องที่รวมกัน

$$\overline{X} = \frac{50+52+45+48+53+40+50+53+49}{7+9} = \frac{816}{16} = 51 \text{ บาท}$$

$$s^2 = \frac{(-1)^2+1^2+(-6)^2\ +4^2+3^2+(-3)^2+2^2+(-11)^2+(-1)^2+1^2+11^2+2^2+(-2)^2}{7+9}$$

$$= \frac{328}{15} = 21.87$$

ดังนั้นความแปรปรวนของสินค้าในสองท้องที่ เท่ากับ 21.87 บาท

**สมบัติที่สำคัญของส่วนเบี่ยงเบนมาตรฐาน**

(1) ส่วนเบี่ยงเบนมาตรฐานมีค่าเป็นบวกเสมอ

(2) ถ้าคำนวณหาส่วนเบี่ยงเบนมาตรฐานโดยใช้ค่ากลางของข้อมูลชนิดอื่น ๆ ที่ไม่ใช่ค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานที่ได้จะมีค่ามากกว่าส่วนเบี่ยงเบนมาตรฐานที่ใช้ค่าเฉลี่ยเลขคณิตเสมอ นั่นคือ

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

อสมการนี้เป็นจริงเสมอ เมื่อ a เป็นจำนวนจริงใด ๆ ที่ไม่เท่ากับค่าเฉลี่ยเลขคณิต

สรุปสัญลักษณ์ที่เกี่ยวกับค่าเฉลี่ยเลขคณิต ส่วนเบี่ยงเบนมาตรฐานและจำนวนข้อมูลที่ใช้เป็นดังนี้

| | **ประชากร (พารามิเตอร์** | **ตัวอย่าง (ตัวประมาณ)** |
|---|---|---|
| ค่าเฉลี่ยเลขคณิต | $\mu$ | $\overline{X}$ |
| ส่วนเบี่ยงเบนมาตรฐาน | $\sigma$ | s หรือ S.D. |
| จำนวนข้อมูล | N | n |

(3) ถ้าข้อมูล 2 ชุด ซึ่งประกอบด้วยข้อมูล N และ N จำนวน มีค่าเฉลี่ยเลขคณิตเท่ากัน แต่ความแปรปรวนของข้อมูลชุดที่ 1 เป็น $S_1^2$ และ ข้อมูลชุดที่ 2 เป็น $S_2^2$ ตามลำดับ แล้วความแปรปรวนรวมของข้อมูลทั้งสองชุด เท่ากับ $\dfrac{N_1S_1^2+N_2S_2^2}{N_1+N_2}$

**การนำสมบัติของส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนไปใช้แก้ปัญหา**

**ตัวอย่างที่ 6** นักเรียนห้องที่หนึ่งจำนวน 40 คน มีส่วนเบี่ยงเบนมาตรฐานของอายุเท่ากับ 2.5 ปี นักเรียนห้องที่สองจำนวน 45 คน มีส่วนเบี่ยงเบนมาตรฐานของอายุเท่ากับ 2.2 ปี ถ้าทั้งสองห้องมีค่าเฉลี่ยเลขคณิตของอายุเท่ากับ 16 ปี จงหาส่วนเบี่ยงเบนมาตรฐานรวมของอายุของนักเรียนทั้งสองห้อง

**วิธีทำ** จากสูตร $S_R^2 = \dfrac{N_1S_1^2+N_2S_2^2}{N_1+N_2}$ $(\because \overline{x}_1 = \overline{x}_2)$

จากโจทย์ จะได้ $N_1 = 40$ , $N_2 = 45$

$S_1 = 2.5$ , $S_2 = 2.2$

แทนค่า $S_R^2 = \dfrac{40(2.5)^2+45(2.2)^2}{40+45}$

$= \dfrac{250+217.80}{85} = \dfrac{467.8}{85} = 5.50$

$\therefore$ $S_R = \sqrt{5.5} = 2.35$ ปี

$\therefore$ ส่วนเบี่ยงเบนมาตรฐานรวมของอายุของนักเรียนทั้งสองห้องเท่ากับ 2.35 ปี

---