ใบความรู้ที่ 6

เรื่อง การวัดการกระจายสัมบูรณ์ (พิสัย /ส่วนเบี่ยงเบนควอร์ไทล์/ ส่วนเบี่ยงเบนเฉลี่ย)

โดยทั่วไป การวัดการกระจายของข้อมูล แบ่งออกได้เป็น 2 วิธี คือ การกระจายสัมบูรณ์และการกระจายสัมพัทธ์

การกระจายสัมบูรณ์ (absolute variation) คือ การวัดการกระจายของข้อมูลเพียงชุดเดียว เพื่อดูว่าในข้อมูลชุดนั้น แต่ละค่ามีความแตกต่างกันมากหรือน้อยเพียงไร การวัดการกระจายสัมบูรณ์ที่นิยมใช้กันมี 4 ชนิด คือ

(1) พิสัย (rang)

(2) ส่วนเบี่ยงเบนควอไทล์ หรือกึ่งช่วงควอร์ไทล์ (quartile deviation หรือ semi-inter quartile deviation)

(3) ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)

(4) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ---> (ใบความรู้ที่ 7)

1. พิสัย

พิสัย คือ ค่าที่ใช้วัดการกระจายของข้อมูล ที่หาได้จากการนำข้อมูลที่มีค่าสูงสุด ลบด้วยข้อมูลที่มีค่าต่ำสุด และเป็นการวัดการกระจายของข้อมูลที่ค่อนข้างหยาบ เพราะเป็นค่าที่คำนวณจากค่าเพียงสองค่าเท่านั้น แต่การวัดการกระจายโดยใช้พิสัย สามารถวัดได้สะดวก และรวดเร็วกว่าวิธีอื่น ๆ

การวัดการกระจายของข้อมูลที่ไม่แจกแจงความถี่ โดยใช้พิสัย

ถ้าให้ $x_1 , x_2 , x_3 ,\ldots,x_n$ เป็นค่าของข้อมูลชุดหนึ่ง พิสัยของข้อมูลชุดนี้ เท่ากับ $x_{max} - x_{min}$

เมื่อ x_{max} แทน ข้อมูลที่มีค่าสูงสุด

x_{min} แทน ข้อมูลที่มีค่าต่ำสุด

ตัวอย่างที่ 1 จงหาพิสัยของชุดข้อมูลต่อไปนี้

(1) 20 , 35 , 100 , 80 , 10, 9 , 30 , 15

วิธีทำ ค่าสูงสุด คือ 100 และค่าต่ำสุด คือ 9

พิสัยของข้อมูลชุดนี้ คือ ค่าสูงสุด – ค่าต่ำสุด = 100-9 = 91

(2) 26 , 25 , 48 , 57 , 60 , 68 , 73 , 85 , 90 , 92

วิธีทำ ค่าสูงสุด คือ 92 และค่าต่ำสุด คือ 25

พิสัยของข้อมูลชุดนี้ คือ ค่าสูงสุด – ค่าต่ำสุด = 92 – 25 = 67

การวัดการกระจายของข้อมูลที่แจกแจงความถี่ โดยใช้พิสัย

พิสัย คือ ผลต่างระหว่างขอบบนของอันตรภาคชั้นของข้อมูลที่มีค่าสูงสุดและขอบล่างของอันตรภาคชั้นของข้อมูลที่มีค่าต่ำสุด

เขียนแทนพิสัยในรูปสัญลักษณ์ ได้ดังนี้ พิสัย = U – L

เมื่อ U แทน ขอบบนของอันตรภาคชั้นที่มีค่าสูงสุด

L แทน ขอบล่างของอันตรภาคชั้นที่มีค่าต่ำสุด

ตัวอย่างที่ 2 จงหาพิสัยของชุดข้อมูลต่อไปนี้

คะแนน	ความถี่
169 – 171	23
166 – 168	35
163 – 165	50
160 – 162	78
157 – 159	83
154 – 156	62
151 – 153	37

วิธีทำ ขอบบนของอันตรภาคชั้นที่มีค่าสูงสุด คือ 171.5 และ ขอบล่างของอันตรภาคชั้นที่มีค่าต่ำสุด คือ 151.5 ดังนั้น พิสัย = U – L = 171.5 – 150.5 = 21

การวัดการกระจายโดยพิสัยนี้ เป็นวิธีวัดการกระจายอย่างคร่าวๆ เพราะหาได้จากค่าของข้อมูลเพียง 2 ค่าเท่านั้น ค่าอื่นๆ ของข้อมูลไม่ได้นำมาใช้ในการคำนวณพิสัยเลย ดังนั้น ถ้าค่าของข้อมูลค่าใดค่าหนึ่งมีค่ามากหรือน้อยผิดปกติ จากค่าของข้อมูลอื่นๆ ก็อาจมีผลทำให้การวัดการกระจายโดยใช้พิสัยมีค่าสูงกว่าที่ควรจะเป็นมาก แต่การวัดการกระจายโดยใช้พิสัยสามารถวัดได้สะดวกและรวดเร็ว จึงมักใช้การวัดการกระจายของข้อมูลในกรณีซึ่งไม่ต้องการความถูกต้องมากนัก

2. ส่วนเบี่ยงเบนควอไทล์ หรือกึ่งช่วงควอร์ไทล์ (quartile deviation หรือ semi-inter quartile deviation)

คือ ค่าที่ใช้วัดการกระจายที่หาได้จากครึ่งหนึ่งของผลต่างระหว่าง ควอร์ไทล์ที่ 3 กับควอร์ไทล์ที่ 1 นั่นคือ

ส่วนเบี่ยงเบนควอร์ไทล์ (Q.D.) = $\dfrac{Q_3 - Q_1}{2}$

เมื่อ Q_3 แทน ควอร์ไทล์ที่ 3

Q_1 แทน ควอร์ไทล์ที่ 1

ตัวอย่างที่ 3 จงหาส่วนเบี่ยงเบนควอไทล์ จากตารางแจกแจงความถี่แสดงผลผลิตเฉลี่ยต่อไร่ของถั่วเหลือง จาก 47 จังหวัดของประเทศไทย มีดังนี้

ผลผลิตถั่วเหลืองเฉลี่ยต่อไร่ (กิโลกรัม)	ความถี่	ความถี่สะสม
120-138	3	3
139-157	1	4
158-176	6	10
177-195	8	18
196-214	19	37
215-233	2	39
234-252	7	46
253-271	1	47

วิธีทำ จะได้ $Q_1 = 180.66$ กิโลกรัม

$Q_2 = 212.75$ กิโลกรัม

ดังนั้น ส่วนเบี่ยงเบนควอไทล์ (Q.D.) = $\frac{Q_3 - Q_1}{2} = \frac{212.75 - 180.66}{2}$

= 16.045 กิโลกรัม

นั่นคือ ผลผลิตถั่วเหลืองเฉลี่ยต่อไร่ใน 47 จังหวัดของประเทศไทยมีค่าการกระจายของข้อมูลที่วัดโดย ส่วนเบี่ยงเบนควอไทล์ มีค่าประมาณ 16.045 กิโลกรัม

ตัวอย่างที่ 4 จงหาส่วนเบี่ยงเบนควอไทล์ จากตารางแจกแจงความถี่ต่อไปนี้

คะแนน	50-54	55-59	60-64	65-69	70-74	75-79
ความถี่	10	20	11	12	21	6

วิธีทำ จะได้ $Q_1 = 57$

$Q_2 = 71.2$

ดังนั้น ส่วนเบี่ยงเบนควอไทล์ (Q.D.) = $\frac{Q_3 - Q_1}{2} = \frac{71.2 - 57}{2}$

= 7.10

การวัดการกระจายของข้อมูล โดยส่วนเบี่ยงเบนควอร์ไทล์ มีส่วนดี เพราะในกรณีที่มีข้อมูลบางค่าสูงหรือต่ำกว่าข้อมูลอื่นๆ ในชุดเดียวกันมาก จะมีผลกระทบต่อพิสัย แต่จะไม่มีผลกระทบต่อส่วนเบี่ยงเบนควอร์ไทล์ เนื่องจากไม่ได้เอาข้อมูลที่มีค่าต่ำสุดกว่า Q_1 หรือสูงกว่า Q_3 มาคำนวณด้วย ในกรณีที่การแจกแจงความถี่ของข้อมูลมีอันตรภาคชั้นแรก หรืออันตรภาคชั้นสุดท้ายในตารางเป็นอันตรภาคชั้นเปิด ก็สามารถหาส่วนเบี่ยงเบนควอร์ไทล์ได้ เพราะไม่ต้องเกี่ยวข้องกับอันตรภาคชั้นแรกและอันตรภาคชั้นสุดท้ายในตาราง ต่างกับการหาส่วนเบี่ยงเบนชนิดอื่นๆ ที่จะกล่าวต่อไป แต่ส่วนเบี่ยงเบนควอร์ไทล์ มีข้อเสียเนื่องจากไม่ได้ใช้ข้อมูลทั้งหมดที่มีอยู่มาคำนวณ ใช้เฉพาะข้อมูลที่มีค่าใกล้เคียงหรือเท่ากับตำแหน่งที่ของควอร์ไทล์ที่หนึ่งและสามเท่านั้น จึงเป็นการวัดการกระจายที่ไม่ค่อยละเอียดนัก

3. ส่วนเบี่ยงเบนเฉลี่ย (mean deviation หรือ average deviation)

ส่วนเบี่ยงเบนเฉลี่ย คือ ค่าที่ใช้วัดการกระจายของข้อมูลที่ได้จากการเฉลี่ยค่าสัมบูรณ์ของความแตกต่างระหว่างค่าของข้อมูลแต่ละค่า จากค่ากลางของข้อมูลชุดนั้น ค่ากลางที่ใช้อาจเป็นค่าเฉลี่ยเลขคณิตหรือมัธยฐานก็ได้ แต่ส่วนมากนิยมใช้ค่าเฉลี่ยเลขคณิต

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่ไม่ได้แจกแจงความถี่

ถ้าให้ $x_1, x_2, x_3, \ldots, x_n$ เป็นข้อมูลตัวอย่าง n จำนวน และมีค่าเฉลี่ยเลขคณิต เป็น $\bar{x}$ แล้ว

$$\text{ส่วนเบี่ยงเบนเฉลี่ย (M.D.)} = \frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{n}$$

ตัวอย่างที่ 5 แสดงการหาค่าส่วนเบี่ยงเบนเฉลี่ยของข้อมูล

ข้อ	ข้อมูล	n	$\bar{x}$	$\sum_{i=1}^{n}\lvert x_i-\bar{x}\rvert$	ส่วนเบี่ยงเบนเฉลี่ย (M.D.)
1	68 , 50 , 46 , 57 , 49	5	54	34	$\frac{34}{5}=6.8$
2	3 , 5 , 1 , 9 , 7	5	5	12	$\frac{12}{5}=2.4$
3	3 , 7 , 8 , 14	4	8	12	$\frac{12}{4}=3$
4	2 , 8 , 9 , 5	4	6	10	$\frac{10}{4}=2.5$
5	2 , 4 , 5 , 9	4	5	8	$\frac{8}{4}=2$

การหาส่วนเบี่ยงเบนเฉลี่ยของข้อมูลที่แจกแจงความถี่แล้ว

ถ้าให้ $x_1, x_2, x_3, \ldots, x_k$ เป็นจุดกึ่งกลางของอันตรภาคชั้นต่างๆ k ชั้น ซึ่งมีความถี่ $f_1, f_2, f_3, \ldots, f_k$ ตามลำดับ n เป็นจำนวนข้อมูลตัวอย่างทั้งหมด และถ้าข้อมูลชุดนี้มีค่าเฉลี่ยเลขคณิต เป็น $\bar{x}$ แล้ว สามารถคำนวณส่วนเบี่ยงเบนเฉลี่ย (โดยประมาณ) ได้จากสูตร

$$\text{ส่วนเบี่ยงเบนเฉลี่ย (M.D.)} = \frac{\sum_{i=1}^{k} f_i\left|x_i - \bar{x}\right|}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i\left|x_i - \bar{x}\right|}{n}$$

เมื่อ k แทนจำนวนอันตรภาคชั้น

f_i แทนความถี่ของอันตรภาคชั้นที่ i

x_i แทนจุดกึ่งกลางของอันตรภาคชั้นที่ i

ตัวอย่างที่ 6 จงหาส่วนเบี่ยงเบนควอร์ไทล์และส่วนเบี่ยงเบนเฉลี่ยของรายได้ของคนงานหญิง 400 คน จากตารางแจกแจงความถี่ต่อไปนี้ แล้วเปรียบเทียบค่าที่ได้กับพิสัยของข้อมูลชุดนี้

รายได้	จำนวนคนงาน
1500 – 1599	20
1600 – 1699	70
1700 – 1799	120
1800 – 1899	100
1900 – 1999	60
2000 – 2099	20
2100 - 2199	10

วิธีทำ การแจกแจงความถี่ของรายได้

รายได้	**จุดกึ่งกลาง x_i**	**จำนวนคนงาน f_i**	**ความถี่สะสม**	f_ix_i	$\left|x_i - \bar{X}\right|$	$f_i\left|x_i - \bar{X}\right|$
1500 – 1599	1549.5	20	20	30990	252.5	5050
1600 – 1699	1649.5	70	90	115465	152.5	10675
1700 – 1799	1749.5	120	210	209940	52.5	6300
1800 – 1899	1849.5	100	310	184950	47.5	4750
1900 – 1999	1949.5	60	370	116970	147.5	8850
2000 – 2099	2049.5	20	390	40990	247.5	4950
2100 - 2199	2149.5	10	400	21495	347.5	3475
		400		720800		44050

จากตาราง จะได้ $\overline{X} = \frac{720800}{400} = 1802$

ตำแหน่งที่ของ Q_1 เท่ากับ $\frac{400}{4} = 100$

ตำแหน่งที่ของ Q_1 อยู่ระหว่างความถี่สะสม 90 กับ 210

ในอันตรภาคชั้น 1600 – 1699 กับ 1700 – 1799

จะได้ Q_1 เท่ากับ $1699.5 + (\frac{100 \times 10}{120}) = 1707.83$

ตำแหน่งที่ของ Q_3 เท่ากับ $\frac{3}{4}(400) = 300$

ตำแหน่งที่ของ Q_3 อยู่ระหว่างความถี่สะสม 210 กับ 310

ในอันตรภาคชั้น 1700 – 1799 กับ 1800 – 1899

จะได้ Q_3 เท่ากับ $1799.5 + (\frac{100 \times 90}{120}) = 1889.5$

ดังนั้น ส่วนเบี่ยงเบนควอร์ไทล์เท่ากับ $\frac{1889.5 - 1707.83}{2} = \frac{181.67}{2} = 90.835$ บาท

จากตารางส่วนเบี่ยงเบนเฉลี่ยเท่ากับ $\frac{44050}{400} = 110.125$ บาท

พิสัยเท่ากับ 2199.5 – 1499.5 = 700 บาท

เปรียบเทียบค่าของส่วนเบี่ยงเบนควอร์ไทล์ และส่วนเบี่ยงเบนเฉลี่ยกับค่าพิสัย จะพบว่า พิสัยมีค่าสูงกว่าส่วนเบี่ยงเบนควอไทล์และส่วนเบี่ยงเบนเฉลี่ยมาก

เนื่องจากการวัดการกระจายโดยใช้ส่วนเบี่ยงเบนเฉลี่ยนี้ ได้นำค่าของข้อมูลทุกๆ ค่าหรือทุกค่าของข้อมูลมีตัวแทนมาใช้ในการคำนวณ จึงทำให้การวัดการกระจายโดยวิธีนี้ดีกว่าการวัดการกระจายโดยใช้พิสัย หรือส่วนเบี่ยงเบนควอร์ไทล์ ซึ่งใช้เฉพาะค่าของข้อมูลเพียงบางค่าเท่านั้น แต่ในกรณีที่ข้อมูลมีจำนวนมากๆ จะมีปัญหาในการคำนวณ โดยเฉพาะเมื่อต้องใช้เครื่องคำนวณที่มีความสามารถต่ำ ซึ่งจะทำให้หาค่าสัมบูรณ์ของผลต่างค่อนข้างลำบาก จึงนิยมวัดการกระจายโดยใช้ส่วนเบี่ยงเบนมาตรฐาน ซึ่งจะได้กล่าวถึงในลำดับต่อไป
