ใบความรู้ที่ 8

เรื่อง การวัดการกระจายสัมพัทธ์

ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไป เพื่อตัดสินว่า ชุดใดมีการกระจายมาก ชุดใดมีการกระจายน้อย ถ้าใช้ค่าที่ได้จากการวัดการกระจายสัมบูรณ์ของข้อมูลแต่ละชุดมาเปรียบเทียบกันย่อมตัดสินใจได้ยาก เช่น ข้อมูลชุดหนึ่งมีค่าตั้งแต่ 0 ถึง 10 มีส่วนเบี่ยงเบนมาตรฐาน 2.2 อีกชุดหนึ่งมีค่าตั้งแต่ 200 ถึง 800 มีส่วนเบี่ยงเบนมาตรฐาน 60.5 ถ้าจะตัดสินใจว่าข้อมูลที่หนึ่ง มีการกระจายน้อยกว่าข้อมูลชุดที่สองก็อาจจะไม่ถูกต้องนัก เพราะค่าของข้อมูลสองชุดนี้ต่างกันมาก ค่ากลางและค่าแสดงการกระจายก็จะต่างกันมากด้วย

เพื่อให้การเปรียบเทียบมีความหมาย เรานิยมหาอัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้นๆ (แต่ไม่ใช่กรณีที่หาพิสัย หรือส่วนเบี่ยงเบนควอไทล์) แล้วจึงนำอัตราส่วนที่หาได้มาเปรียบเทียบกัน

การกระจายสัมพัทธ์ (relative variation) คือ การหาค่าเพื่อเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่า 1 ชุด โดยใช้อัตราส่วน เช่น อัตราส่วนระหว่างค่าการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้น ๆ การวัดการกระจายสัมพัทธ์ของการกระจายสัมพัทธ์ของข้อมูลแต่ละชุด เพื่อนำไปใช้ในการเปรียบเทียบการกระจายของข้อมูลระหว่างชุดมีอยู่ 4 ชนิด คือ

1. สัมประสิทธิ์ของพิสัย (coefficient)
2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ (coefficient of quartile deviation)
3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย (coefficient of average deviation)
4. สัมประสิทธิ์ของความแปรผัน (coefficient of variation)

ทั้งนี้ สัมประสิทธิ์ที่ใช้วัดการกระจายสัมพัทธ์ มีดังนี้

1. สัมประสิทธิ์ของพิสัย $= \dfrac{X_{\max} - X_{\min}}{X_{\max} + X_{\min}}$
2. สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$
3. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของตัวอย่าง $= \dfrac{M.D.}{\overline{X}}$
4. สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ยของประชากร $= \dfrac{M.D.}{\mu}$
5. สัมประสิทธิ์ของความแปรผันของตัวอย่าง $= \dfrac{s}{\overline{X}}$
6. สัมประสิทธิ์ของความแปรผันของประชากร $= \dfrac{\sigma}{\mu}$

สัมประสิทธิ์ของการแปรผัน เป็นวิธีวัดการกระจายสัมพัทธ์ที่นิยมใช้กันมากที่สุด และมักเขียนในรูปเปอร์เซ็นต์ โดยคูณด้วย 100 เช่น ข้อมูลชุดหนึ่งมี s = 10 และ $\overline{X} = 30$ สัมประสิทธิ์ของการแปรผันของข้อมูลชุดนี้ คือ 0.33 หรือ 33%

ตัวอย่างที่ 1 กำหนดข้อมูล 2 ชุด ซึ่งเป็นข้อมูลเกี่ยวกับน้ำหนักของนักเรียน ชั้น ม.6 ซึ่งสุ่มมาจำนวนหนึ่งของโรงเรียนนารายณ์คำผงวิทยา ปีการศึกษา 2553 ดังนี้

- ห้องเรียนที่หนึ่ง (ก.ก.) 52 , 56 , 57 , 59 , 64
- ห้องเรียนที่สอง (ก.ก.) 54 , 60 , 65 , 67 , 68 , 58

จงเปรียบเทียบการกระจายของน้ำหนักนักเรียน ชั้น ม.6 โรงเรียนนารายณ์คำผงวิทยา โดยใช้สัมประสิทธิ์ของพิสัย

วิธีทำ

ข้อมูลชุด	น้ำหนักสูงสุด	น้ำหนักต่ำสุด	สัมประสิทธิ์ของพิสัย	การเปรียบเทียบการกระจายของข้อมูล
1	64	52	$\frac{64-52}{64+52} = \frac{12}{116}$ $= 0.103$	น้ำหนักของนักเรียนห้องเรียนที่ 2 มีการกระจายมากกว่าน้ำหนักของนักเรียนห้องเรียนที่ 1
2	68	54	$\frac{68-54}{68+54} = \frac{14}{122}$ $= 0.115$	

ตัวอย่างที่ 2 จงเปรียบเทียบการกระจายของอายุบุตรสองครอบครัว โดยที่อายุบุตรทั้งสองครอบครัวเป็นดังนี้

อายุของบุตรในครอบครัวที่หนึ่ง (ปี) 6, 5, 3, 1

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

(1) ใช้สัมประสิทธิ์ของพิสัย

(2) ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์

(3) ใช้สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย

(4) ใช้สัมประสิทธิ์ของการแปรผัน

ผลของการเปรียบเทียบที่ได้จากการใช้วิธีทั้ง 4 นี้ เหมือนกันหรือไม่อย่างไร

วิธีทำ

อายุของบุตรในครอบครัวที่หนึ่ง (ปี)	6	5	3	1	
อายุของบุตรในครอบครัวที่สอง (ปี)	25	24	22	21	17

(1) เนื่องจากสัมประสิทธิ์ของพิสัย = $\frac{x_{max} - x_{min}}{x_{max} + x_{min}}$

ดังนั้น สัมประสิทธิ์ของพิสัยของครอบครัวที่หนึ่ง $= \frac{6-1}{6+1} = 0.714$

สัมประสิทธิ์ของพิสัยของครอบครัวที่สอง $= \frac{25-17}{25+17} = 0.190$

จะได้ อายุของบุตรครอบครัวที่หนึ่งมีการกระจายมากกว่าอายุของบุตรครอบครัวที่สอง

(2) เนื่องจากสัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

ตำแหน่งที่ของ Q_1 ของครอบครัวที่หนึ่งเท่ากับ $\frac{4+1}{4}$ = 1.25

จะได้ Q_1 ของครอบครัวที่หนึ่งเท่ากับ $1 + (2 \times 0.25) = 1.5$

ตำแหน่งที่ของ Q_3 ของครอบครัวที่หนึ่งเท่ากับ $\frac{3}{4}(4+1) = 3.75$

จะได้ Q_3 ของครอบครัวที่หนึ่งเท่ากับ $5 + (1 \times 0.75) = 5.75$

สัมประสิทธิ์ส่วนเบี่ยงเบนควอร์ไทล์ครอบครัวที่หนึ่งเท่ากับ $\frac{5.75-1.5}{5.75+1.5} = 0.586$

ตำแหน่งที่ของ Q_1 ของครอบครัวที่สองเท่ากับ $\frac{5+1}{4} = 1.5$

จะได้ Q_1 ของครอบครัวที่สองเท่ากับ $17 + (4 \times 0.5) = 19$

ตำแหน่งที่ของ Q_3 ของครอบครัวที่สองเท่ากับ $\frac{3}{4}(5+1) = 4.5$

จะได้ Q_3 ของครอบครัวที่สองเท่ากับ $24 + (1 \times 0.5) = 24.5$

สัมประสิทธิ์ส่วนเบี่ยงเบนควอร์ไทล์ที่สองเท่ากับ $\frac{24.5-19}{24.5+19} = 0.126$

จะได้ อายุของบุตรครอบครัวที่หนึ่งมีการกระจายมากกว่าอายุของบุตรครอบครัวที่สอง

(3) เนื่องจากสัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย = $\frac{M.D.}{\overline{X}}$

$\overline{X}$ ของครอบครัวที่หนึ่งเท่ากับ $\frac{6+5+3+1}{4} = 3.75$

M.D. ของครอบครัวที่หนึ่งเท่ากับ $\frac{2.25+1.25+0.75+2.75}{4} = 1.75$

สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ยครอบครัวที่หนึ่งเท่ากับ $\frac{1.75}{3.75} = 0.467$

$\overline{X}$ ของครอบครัวที่สองเท่ากับ $\frac{25+24+22+21+17}{5} = 21.8$

M.D. ของครอบครัวที่สองเท่ากับ $\frac{3.2+2.2+0.2+0.8+4.8}{5} = 2.24$

สัมประสิทธิ์ส่วนเบี่ยงเบนเฉลี่ยครอบครัวที่สองเท่ากับ $\frac{2.24}{21.8} = 0.103$

จะได้อายุของบุตรครอบครัวที่หนึ่งมีการกระจายมากกว่าอายุของบุตรครอบครัวที่สอง

(4) เนื่องจากสัมประสิทธิ์ของการแปรผัน = $\frac{s}{\overline{X}}$

s ของครอบครัวที่หนึ่งเท่ากับ $\sqrt{\frac{(2.25)^2 + (1.25)^2 + (-0.75)^2 + (-2.75)^2}{4-1}}$

$= \sqrt{\frac{14.75}{3}} = 2.217$

สัมประสิทธิ์การแปรผันของครอบครัวที่หนึ่งเท่ากับ $\frac{2.217}{3.75} = 0.591$

s ของครอบครัวที่สองเท่ากับ $\sqrt{\dfrac{(3.2)^2+(2.2)^2+(0.2)^2+(-0.8)^2+(-4.8)^2}{5-1}}$

$$= \sqrt{\frac{38.8}{4}} = 3.114$$

สัมประสิทธิ์การแปรผันของครอบครัวที่สองเท่ากับ $\dfrac{3.114}{21.8} = 0.143$

จะได้ อายุของบุตรครอบครัวที่หนึ่งมีการกระจายมากกว่าอายุของบุตรครอบครัวที่สอง ผลของการเปรียบเทียบที่ได้จากข้อ (1) – (4) เหมือนกัน

สรุปได้ว่า อายุของบุตรครอบครัวที่หนึ่งมีการกระจายมากกว่าอายุของบุตรครอบครัวที่สอง

ตัวอย่างที่ 3 ถ้าราคาข้าวเปลือกและราคาข้าวสารต่อถังของร้านค้าข้าวที่สำรวจมาเป็นตัวอย่าง 6 ร้าน ในท้องที่แห่งหนึ่ง เป็นดังนี้

ราคาข้าวเปลือก (บาท)	72	75	73	74	76	71
ราคาข้าวสาร (บาท)	115	118	112	114	117	110

จงหาสัมประสิทธิ์ของการแปรผันและสัมประสิทธิ์ของพิสัยของราคาข้าวเปลือกและราคาข้าวสาร พร้อมทั้งอธิบายความหมายของค่าที่หาได้

วิธีทำ จากโจทย์เรียงลำดับข้อมูลจากน้อยไปหามากได้ดังนี้

ราคาข้าวเปลือก (บาท)	71	72	73	74	75	76
ราคาข้าวสาร (บาท)	110	112	114	115	117	118

$\overline{X}$ ของราคาข้าวเปลือก เท่ากับ $\dfrac{71+72+73+74+75+76}{6} = \dfrac{441}{6} = 73.5$

s ของราคาข้าวเปลือกเท่ากับ $\sqrt{\dfrac{(-2.5)^2+(-1.5)^2+(-0.5)^2+(1.5)^2+(2.5)^2}{6-1}} = \sqrt{\dfrac{17.5}{5}} = 1.871$

สัมประสิทธิ์การแปรผันของราคาข้าวเปลือกเท่ากับ $\dfrac{1.871}{73.5} = 0.025$

$\overline{X}$ ของราคาข้าวสาร เท่ากับ $\dfrac{110+112+114+115+117+118}{6} = \dfrac{686}{6} = 114.33$

s ของราคาข้าวสารเท่ากับ $\sqrt{\dfrac{(-2.33)^2+(-2.35)^2+(-0.33)^2+(0.67)^2+(2.67)^2+(3.67)^2}{6-1}}$

$$= \sqrt{\frac{45.3334}{5}} = 3.011$$

สัมประสิทธิ์การแปรผันของราคาข้าวสารเท่ากับ $\dfrac{3.011}{114.33} = 0.026$

สัมประสิทธิ์ของพิสัยของราคาข้าวเปลือก เท่ากับ $\dfrac{76-71}{76+71} = 0.034$

สัมประสิทธิ์ของพิสัยของราคาข้าวสารเท่ากับ $\dfrac{118-110}{118+110} = 0.035$

จากค่าที่ได้จะสรุปได้ว่า ราคาของข้าวเปลือกต่อถังมีการกระจายน้อยกว่าราคาข้าวสารต่อถัง

ความสัมพันธ์ระหว่างการแจกแจงความถี่ ค่ากลาง และการกระจายของข้อมูล

จากเรื่องการแจกแจงความถี่ของข้อมูลที่กล่าวมาแล้ว จะเห็นว่าโดยทั่วๆ ไป รูปแบบของเส้นโค้งของความถี่ อาจแบ่งออกได้เป็น 3 แบบ คือ

1) เส้นโค้งปกติ หรือรูประฆัง (normal or bell-shaped curve)

2) เส้นโค้งเบ้ลาดทางขวาหรือทางบวก (positively skewed curve)

3) เส้นโค้งเบ้ลาดทางซ้ายหรือทางลบ (negatively skewed curve)

เส้นโค้งของความถี่ของข้อมูลที่พบเสมอ ๆ ไม่ว่าจะเป็นข้อมูลทางด้านประชากร เกษตร สังคม เศรษฐกิจ หรือวิทยาศาสตร์ ส่วนใหญ่เป็นข้อมูลที่เกิดขึ้นหรือเป็นไปตามธรรมชาติ และจะมีเส้นโค้งความถี่เป็นรูปเส้นโค้งปกติ เช่น ข้อมูลเกี่ยวกับส่วนสูง น้ำหนัก ราคา และผลผลิตทางการเกษตร มักจะมีลักษณะเป็นเส้นโค้งปกติ

ถ้าข้อมูลชุดใดมีเส้นโค้งของความถี่เป็นเส้นโค้งปกติ จะมีค่าเฉลี่ยเลขคณิตมัธยฐานและฐานนิยมอยู่ที่จุดเดียวกัน คือ จุดที่มีความถี่สูงสุด ถ้าเส้นโค้งของความถี่เป็นเส้นโค้งเบ้ลาดทางขวาหรือส่วนของเส้นโค้งที่มีความชันน้อยอยู่ทางด้านขวาของรูปแล้ว ค่าเฉลี่ยเลขคณิตจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐานและฐานนิยมตามลำดับ ในกรณีที่เส้นโค้งของความถี่เป็นเส้นโค้งเบ้ลาดทางซ้ายหรือส่วนของเส้นโค้งที่มีความชันน้อยอยู่ทางด้านซ้ายของรูป ฐานนิยมจะมีค่ามากที่สุด รองลงมาเป็นมัธยฐาน และค่าเฉลี่ยเลขคณิตมีค่าน้อยที่สุด

เส้นโค้งปกติจะมีความโด่งมากหรือน้อยขึ้นอยู่กับการกระจายของข้อมูล ถ้าข้อมูลมีการกระจายมาก เส้นโค้งปกติจะมีความโด่งน้อยหรือค่อนข้างแบน แต่ถ้าข้อมูลมีการกระจายน้อย เส้นโค้งปกติจะมีความโด่งมากหรือค่อนข้างสูง
