การวัดความสัมพันธ์ : Pearson's Sample Correlation Coefficient

ฉัตรศิริ ปิยะพิมลสิทธิ์*

แผนภาพกระจัดกระจายแสดงความสัมพันธ์ระหว่างข้อมูล 2 ชุดคือ X และ Y ดังแสดงในภาพประกอบ 1 จะบ่งบอกถึงความสัมพันธ์ในรูปแบบต่าง ๆ อย่างในภาพประกอบ 1(a) จะเป็นความสัมพันธ์ทางบวกระหว่าง X และ Y นั่นคือในทุก ๆ คู่ของค่า X และ Y ที่พล็อตลงไปนั้น เมื่อค่า X มีค่าเพิ่มขึ้นแล้วค่า Y ก็จะมีค่าเพิ่มขึ้นด้วย ส่วนในภาพประกอบ 1(b) จะแสดงการเพิ่มของค่า X และ Y ในทิศทางตรงกันข้าม ก็คือเมื่อ X มีค่าเพิ่มขึ้นแล้ว Y จะมีค่าลดลง จะเรียกว่ามีความสัมพันธ์ทางลบ ส่วนในภาพประกอบ 1(c) แสดงความไม่สัมพันธ์กันระหว่าง X และ Y ก็คือเมื่อ X เพิ่มขึ้น Y อาจจะลดหรือเพิ่มขึ้นก็ได้

ให้ $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ แทนคู่ของตัวแปร (X, Y) ในภาพประกอบ 1(**a**) จะมีเส้นตรงตั้งฉากกับแกนนอนที่ $\bar{X}$ และเส้นตรงตั้งฉากกับแกนตั้งที่ $\bar{Y}$ แบ่งแผนภาพออกเป็น 4 ส่วน ในส่วนที่ 1 ทั้งค่า X และ Y จะมีค่าสูงกว่าค่าเฉลี่ย ดังนั้นส่วนเบี่ยงเบน $X - \bar{X}$ และ $Y - \bar{Y}$ จะมีค่าเป็นบวก ดังนั้น $(X - \bar{X})(Y - \bar{Y})$ จึงมีค่าเป็นบวก เหมือนกับในส่วนที่ 3 ที่มีค่าเป็นบวกเช่นเดียวกันเพราะส่วนเบี่ยงเบน $X - \bar{X}$ และ $Y - \bar{Y}$ มีค่าติดลบ เมื่อนำมาคูณกันจึงมีค่าเป็นบวก ในส่วนที่ 2 และ 4 จะมีค่า $(X - \bar{X})(Y - \bar{Y})$ เป็นลบ ในภาพประกอบ 1(**a**) นี้เกือบทุกจุดจะอยู่ในส่วนที่ 1 และ 3 ซึ่งจะให้ผลเป็นบวก ดังนั้น $\Sigma(X - \bar{X})(Y - \bar{Y})$ จึงมีค่าเป็นบวก

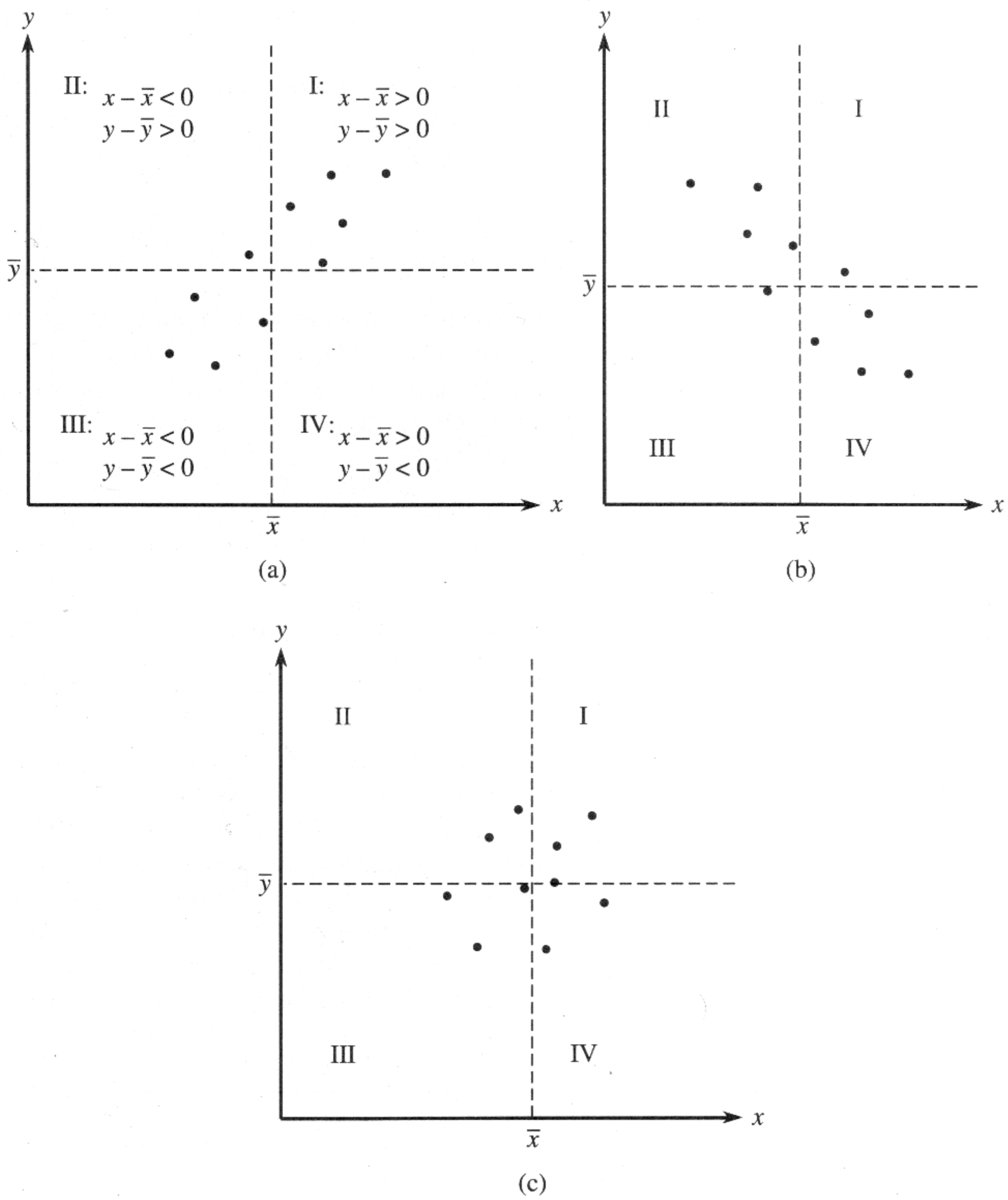
เหตุผลเดียวกันในภาพประกอบ 1(b) ซึ่งจะแสดงความสัมพันธ์ทางลบ เกือบทุกจุดจะอยู่ในส่วนที่ 2 และ 4 ซึ่งจะมีค่า $(X - \bar{X})(Y - \bar{Y})$ เป็นลบ ดังนั้น $\Sigma(X - \bar{X})(Y - \bar{Y})$ จึงมีค่าเป็นลบ ส่วนในภาพประกอบ 1(**c**) ทุก ๆ จุดจะอยู่ในส่วนที่ 1, 2, 3 และ 4 ในปริมาณที่พอกัน ดังนั้นผลของ $\Sigma(X - \bar{X})(Y - \bar{Y})$ จึงมีค่าเข้าใกล้ศูนย์ โดยสรุปแล้ว $\Sigma(X - \bar{X})(Y - \bar{Y})$ สามารถบอกระดับของความสัมพันธ์ระหว่างตัวแปร x และ y ได้ว่ามีความสัมพันธ์ทางบวก มีความสัมพันธ์ทางลบ หรือไม่มีความสัมพันธ์

ต่อมาได้มีการนำ $\Sigma(X - \bar{X})(Y - \bar{Y})$ มาพัฒนาเป็นสมการในการคำนวณหาความสัมพันธ์ที่เป็นที่นิยมกันอย่างแพร่หลาย เรียกว่า "สัมประสิทธิ์สหสัมพันธ์เพียร์สัน"

เมื่อ $\bar{X}$ และ S_X แทนค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่า X ในคู่ของ $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ และให้ $\bar{Y}$ และ S_Y แทนค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่า Y แล้ว สัมประสิทธิ์สหสัมพันธ์เพียร์สัน (r) เขียนเป็นสมการว่า

$$\mathbf{r} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{(N-1)S_X S_Y}$$

กศ.ม. (การวัดผลการศึกษา) http://www.watpon.com



ภาพประกอบ 1 แผนภาพกระจัดกระจายแสดงความสัมพันธ์ระหว่างตัวแปร x และ y

ตัวอย่างการคำนวณ 1

เจ้าของโรงงานแห่งหนึ่งต้องการหาความสัมพันธ์ระหว่างผลผลิตทางการเกษตร 2 ชนิด คือข้าวเหนียว และข้าวสาลี ที่เขาผลิตได้ในแต่ละวัน เขาได้บันทึกผลผลิตของเขาเป็นจำนวน 8 วัน ได้ผลผลิตดังตารางต่อไปนี้

วันที่	1	2	3	4	5	6	7	8
ข้าวเหนียว (X)	2.4	3.4	4.6	3.7	2.2	3.3	4.0	2.1
ข้าวสาลี (Y)	1.33	2.12	1.80	1.65	2.00	1.76	2.11	1.63

คำนวณหาค่าต่าง ๆ ได้ดังนี้

$\Sigma X = 25.7$ $\quad \Sigma X^2 = 88.31$ $\quad S_X^2 = 0.821250$ $\quad S_X = 0.906228$

$\Sigma Y = 14.40$ $\quad \Sigma Y^2 = 26.4324$ $\quad S_Y^2 = 0.073200$ $\quad S_Y = 0.270555$

ค่าเฉลี่ย $\bar{X} = 25.7/8 = 3.2125$ และ $\bar{Y} = 14.40/8 = 1.8000$

$$\begin{aligned}\Sigma(X-\bar{X})(Y-\bar{Y}) &= (X_1-\bar{X})(Y_1-\bar{Y}) + \ldots + (X_8-\bar{X})(Y_8-\bar{Y}) \\ &= (2.4-3.2125)(1.33-1.8000) + \ldots \\ &\quad + (2.1-3.2125)(1.63-1.8000) \\ &= 0.3819 + \ldots + 0.1891 \\ &= 0.5960\end{aligned}$$

ดังนั้น

$$\mathbf{r} = \frac{0.5960}{7(0.9062)(0.2706)} = 0.347$$

ผลของค่าสหสัมพันธ์ก็คือ $\mathbf{r} = 0.347$ บ่งบอกสหสัมพันธ์ทางบวกที่ค่อนข้างต่ำ

สมการในการคำนวณค่าสหสัมพันธ์อีกสมการหนึ่งที่ไม่จำเป็นต้องใช้ส่วนเบี่ยงเบนในการคำนวณก็คือ

$$\mathbf{r} = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}\sqrt{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}}$$

หรือเขียนในอีกรูปหนึ่งได้ว่า

$$\mathbf{r} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

ตัวอย่างการคำนวณ 2

จากข้อมูลในตัวอย่างการคำนวณ 1 สามารถคำนวณหาค่า **r** ได้ดังนี้

X	Y	X^2	Y^2	XY
2.4	1.33	5.76	1.7689	3.192
3.4	2.12	11.56	4.4944	7.208
4.6	1.80	21.16	3.2400	8.280
3.7	1.65	13.69	2.7225	6.105
2.2	2.00	4.84	4.0000	4.400
3.3	1.76	10.89	3.0976	5.808
4.0	2.11	16.00	4.4521	8.440
2.1	1.63	4.41	2.6569	3.423
$\Sigma X = 25.7$	$\Sigma Y = 14.40$	$\Sigma X^2 = 88.31$	$\Sigma Y^2 = 26.4324$	$\Sigma XY = 46.856$

แทนค่าในสมการได้

$$\mathbf{r} = \frac{46.856 - \frac{(25.7)(14.40)}{8}}{\sqrt{88.31 - \frac{(25.7)^2}{8}}\sqrt{26.4324 - \frac{(14.40)^2}{8}}}$$

$$= \frac{0.5960}{(2.3977)(0.7158)}$$

$$= 0.347$$

หรืออาจจะเขียนสมการในรูปของคะแนนมาตรฐานได้ว่า

$$\mathbf{r} = \frac{\Sigma Z_X Z_Y}{N}$$

และเขียนอยู่ในรูปของความแปรปรวนร่วมได้ว่า

$$\mathbf{r} = \frac{COV_{XY}}{S_X S_Y}$$

เมื่อ $COVxy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}$

การทดสอบนัยสำคัญ

ในการทดสอบนัยสำคัญของสหสัมพันธ์ใช้เมื่อต้องการอ้างอิงผลการคำนวณที่ได้จากกลุ่มตัวอย่างไปยังประชากร เขียนเป็นสมมติฐานได้ว่า

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$

สามารถทดสอบสมมติฐานได้ด้วย **t-test** มีสมการคำนวณดังนี้

$$\mathbf{t} = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \;; \quad \mathbf{df} = N - 2$$

แทนค่าในสมการจะได้

$$\mathbf{t} = \frac{0.347\sqrt{8-2}}{\sqrt{1-(0.347)^2}}$$

$$= 0.84997/.937865$$

$$= 0.906$$

เปิดตาราง **t** ที่ระดับ .05 **df** = 6 ได้ค่า 2.447 ซึ่งมากกว่าค่าที่คำนวณได้ ผลก็คือไม่มีนัยสำคัญทางสถิติ หรือยอมรับ H_0 หรือผลผลิตของข้าวสาลีและข้าวเหนียวไม่มีความสัมพันธ์กัน

หรืออาจจะเปรียบเทียบค่าสหสัมพันธ์ที่คำนวณได้กับค่าวิกฤติที่เปิดจากตารางสหสัมพันธ์แบบเพียร์สันก็ได้ โดยค่าจากตารางสหสัมพันธ์ที่ระดับ .05 **df** = 6 ได้ค่า 0.7067 ซึ่งมากกว่าสหสัมพันธ์ที่คำนวณ 0.347 นั่นคือไม่มีนัยสำคัญทางสถิติ

หมายเหตุ : **r** เป็นค่าสหสัมพันธ์ที่คำนวณจากกลุ่มตัวอย่าง
ρ เป็นค่าสหสัมพันธ์ที่คำนวณจากประชากร

คุณสมบัติของ r

1. ค่าของ **r** ไม่ขึ้นกับหน่วยในการวัดของตัวแปรทั้งสอง ถ้า X เป็นความสูง ซึ่งอาจจะมีหน่วยเป็นเมตร ถ้าหากเปลี่ยนหน่วยมาเป็นนิ้ว หรือ เซนติเมตรแล้ว ค่าสหสัมพันธ์ที่คำนวณได้จะไม่เปลี่ยนแปลง หรือ y คืออุณหภูมิ อาจจะเป็นองศาเซลเซียสหรือเปลี่ยนมาเป็นองศาฟาเรนไฮท์ ค่าสหสัมพันธ์ที่คำนวณได้ก็ยังคงเดิม

2. ค่าของ **r** อยู่ระหว่าง −1.00 ถึง 1.00 ถ้าหากค่า **r** มีค่ามากกว่า 0 แล้วจะเป็นความสัมพันธ์ทางบวก ถ้าหากมีค่าน้อยกว่า 0 แล้วจะเป็นความสัมพันธ์ทางลบ ตัวแปรจะสัมพันธ์กันสูง ปานกลางหรือต่ำมีเกณฑ์ดังนี้

สัมพันธ์กันสูง $\mathbf{r} \geq 0.80$ หรือ $\mathbf{r} \leq -0.80$

สัมพันธ์กันปานกลาง $0.50 < \mathbf{r} < 0.80$ หรือ $-0.80 < \mathbf{r} < -0.5$

สัมพันธ์กันต่ำ $-0.50 \leq \mathbf{r} \leq 0.50$

บรรณานุกรม

Runyon, Richard P. and Other. (1996). ***Fundamentals of Behavioral Statistics.*** U.S.A. : McGraw-Hill,

Devore, Jay and Peck, Roxy. (1993). ***Statistics : The Exploration and Analysis of Data.*** California : Wadsworth, Inc.,

Howell, David C. (1992). ***Statistical Methods for Psychology.*** California : Wadsworth, Inc.,