

บทที่ 9

การวิเคราะห์ความสัมพันธ์สำหรับข้อมูลเชิงปริมาณ

1. สัมประสิทธิ์ความสัมพันธ์ (The coefficient of Correlation)

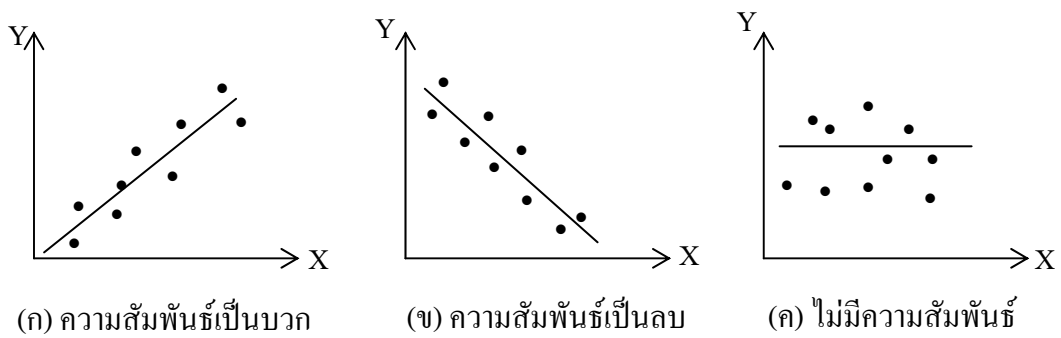
สถิติที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรที่มีสเกลการวัดแบบช่วง (interval scale) ซึ่งข้อมูลมีลักษณะเป็นตัวเลขเชิงปริมาณ คือ ค่าสถิติเพียร์สัน โพรดัก โมเมนต์ (Pearson Product Moment) สัญลักษณ์ที่ใช้คือ r มีสูตรคือ

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$
$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

เมื่อ r คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y

\bar{x} และ \bar{y} คือค่าเฉลี่ยของข้อมูล x และ y ตามลำดับ

เราเรียก r ว่า สหสัมพันธ์เพียร์สัน (Pearson Correlation) ซึ่งมีค่าอยู่ระหว่าง -1 ถึง $+1$ เมื่อ r มีค่าเข้าใกล้ $+1$ หมายถึง ตัวแปร 2 ตัวมีความสัมพันธ์กันอย่างสมบูรณ์และมีทิศทางเดียวกัน คือ ถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น ตัวแปรอีกตัวหนึ่งก็มีค่าเพิ่มขึ้นด้วย หรือถ้าตัวแปรตัวหนึ่งมีค่าลดลง ตัวแปรอีกตัวหนึ่งก็มีค่าลดลงด้วย ดังแสดงในภาพที่ 9.1 (ก) เมื่อ r มีค่าเข้าใกล้ -1 หมายถึง ตัวแปร 2 ตัวมีความสัมพันธ์กันอย่างสมบูรณ์ แต่มีทิศทางกลับกันคือ ถ้าตัวแปรตัวหนึ่งมีค่าเพิ่มขึ้น ตัวแปรอีกตัวหนึ่งจะมีค่าลดลง ดังแสดงในภาพที่ 9.1 (ข) และเมื่อ r มีค่า 0 หมายถึง ตัวแปร 2 ตัวไม่มีความสัมพันธ์กัน และถ้าค่าเข้าใกล้ 0 หมายถึง มีความสัมพันธ์กันน้อย ดังแสดงในภาพที่ 9.1 (ค)



ภาพที่ 9.1 แสดงความสัมพันธ์ระหว่างตัวแปร X และ Y

การวัดความสัมพันธ์นี้เป็นการวัดความสัมพันธ์ในเชิงเส้นตรง (linear) ดังนั้นถ้าค่า r เข้าใกล้ 0 อาจหมายความว่า ตัวแปร 2 ตัวนี้อาจมีความสัมพันธ์กันในรูปแบบอื่นที่ไม่ใช่เส้นตรงก็ได้ ดังนั้นจึงควรพล็อตกราฟของความสัมพันธ์ดูก่อนว่ามีรูปแบบของความสัมพันธ์เป็นเส้นตรงหรือไม่

2. การทดสอบความสัมพันธ์ของ 2 ตัวแปรที่มีสเกลการวัดแบบช่วง

r คือ สหสัมพันธ์เพียร์สัน เป็นการวัดขนาดของความสัมพันธ์เชิงเส้นของค่าสังเกตของตัวอย่างระหว่างตัวแปร 2 ตัว ที่มีสเกลการวัดแบบช่วง

เมื่อ r มีค่าเข้าใกล้ 0 ความสัมพันธ์ระหว่างตัวแปร X และ Y จะเข้าใกล้เส้นระนาบ เนื่องจาก r เป็นค่าดัชนีที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่างขนาด n ที่ใช้ในการประมาณค่าพารามิเตอร์ของประชากร ที่เรียกว่า สัมประสิทธิ์สหสัมพันธ์ของประชากร เขียนแทนด้วยสัญลักษณ์ ρ_{XY} หรือ ρ มีสูตรคือ

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

เมื่อ σ_X, σ_Y คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรของตัวแปรสุ่ม X และ Y

σ_{XY} คือ ความแปรปรวนร่วมระหว่างตัวแปรสุ่ม X และ Y

โดยทั่วไปเราจะไม่ทราบค่า ρ เราจึงสุ่มตัวอย่างจากประชากรที่สนใจศึกษา แล้วคำนวณหาค่า r ซึ่งเป็นค่าประมาณของ ρ

สมมติฐานที่ต้องการทดสอบมีหลายรูปแบบ ตัวอย่างเช่น

$$H_0 : \rho = 0 \quad \text{คู่กับ} \quad H_1 : \rho \neq 0$$

$$H_0 : \rho = \rho_0 \quad \text{คู่กับ} \quad H_1 : \rho \neq \rho_0$$

ในที่นี้จะอธิบายเฉพาะการทดสอบที่มีสมมติฐาน $H_0 : \rho = 0$ คู่กับ $H_1 : \rho \neq 0$

ข้อตกลงเบื้องต้นคือ

1. สำหรับแต่ละค่าของ X ค่าของ Y จะมีการแจกแจงแบบปกติ
2. สำหรับแต่ละค่าของ Y ค่าของ X จะมีการแจกแจงแบบปกติ
3. การแจกแจงร่วมของ X และ Y มีการแจกแจงแบบปกติ
4. ทุก ๆ ประชากรของค่า Y มีความแปรปรวนเท่ากัน
5. ทุก ๆ ประชากรของค่า X มีความแปรปรวนเท่ากัน

สถิติทดสอบความสัมพันธ์คือ

$$\begin{aligned} t &= \frac{r - \rho}{S_r} \\ &= \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \end{aligned}$$

ซึ่งมีการแจกแจงแบบ t ที่มีจำนวนชั้นอิสระ $df = n - 2$

โดยที่ $\rho = 0$

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

เมื่อ n คือ ขนาดตัวอย่าง

3. การใช้คำสั่ง Correlate แบบ Bivariate

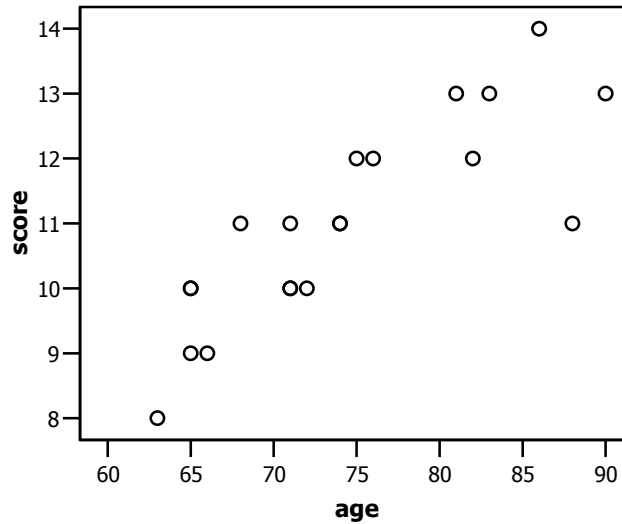
ตัวอย่างเช่น ผู้วิจัยสนใจศึกษาความสัมพันธ์ระหว่างอายุและคะแนนสอบเกี่ยวกับคำศัพท์ กำหนดให้อายุแทนด้วยตัวแปร age และคะแนนสอบเกี่ยวกับคำศัพท์แทนด้วยตัวแปร score ดังตารางที่ 9.1

ตารางที่ 9.1 อายุและคะแนนสอบเกี่ยวกับคำศัพท์ของผู้สอบ 20 คน

age	score	age	score
66	9	65	9
81	13	72	10
83	13	74	11
65	10	76	12
71	11	71	10
82	12	88	11
86	14	68	11
71	10	74	11
63	8	65	10
90	13	75	12

สมมติฐานที่ต้องการทดสอบคือ H_0 : อายุกับคะแนนสอบเกี่ยวกับคำศัพท์ไม่มีความสัมพันธ์กัน หรือ $\rho = 0$ คู่กับ H_1 : อายุกับคะแนนสอบเกี่ยวกับคำศัพท์มีความสัมพันธ์กัน หรือ $\rho \neq 0$ ต้องการหาความสัมพันธ์ระหว่างตัวแปร age และตัวแปร score ซึ่งมีสเกลการวัดแบบช่วงทั้ง 2 ตัวแปร บันทึกข้อมูลในแฟ้มข้อมูล Vocab.sav เราสามารถใช้โปรแกรม SPSS ช่วยในการคำนวณ มีขั้นตอนในการทดสอบความสัมพันธ์ดังนี้

1. ไปที่เมนูบาร์ คลิกที่ Graphs, Scatter... จะได้นหน้าต่าง Scatterplot คลิกเลือกที่ Simple เพื่อดูความสัมพันธ์ระหว่างตัวแปร 2 ตัวเท่านั้น แล้วคลิกที่ปุ่ม Define จะได้นหน้าต่าง Simple Scatterplot
2. ในหน้าต่าง Simple Scatterplot ในช่องซ้ายมือคลิกที่ตัวแปร score ให้ย้ายเข้าไปอยู่ในช่อง Y Axis : แล้วคลิกที่ตัวแปร age ให้ย้ายเข้าไปอยู่ในช่อง X Axis : แล้วคลิกปุ่ม OK จะได้ผลลัพธ์เป็นกราฟของ age และ score ดังภาพที่ 9.2



ภาพที่ 9.2

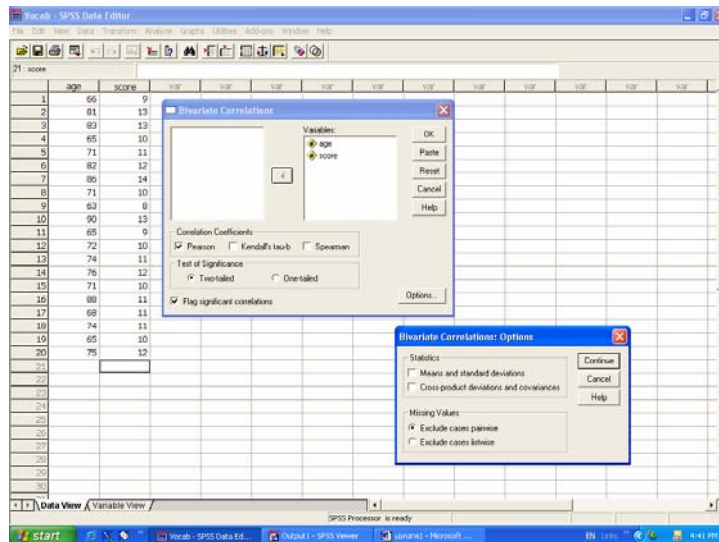
3. ไปที่เมนูบาร์ คลิกที่ Analyze, Correlate, Bivariate จะได้นหน้าต่าง Bivariate Correlations

4. ในหน้าต่าง Bivariate Correlations ในช่องซ้ายมือ คลิกที่ตัวแปร age และ score ให้ย้ายเข้าไปอยู่ในช่อง Variables :

ในกรอบ Correlation Coefficient คลิกเลือกที่ Pearson โดยปกติโปรแกรมจะเลือกให้อยู่แล้ว เพื่อให้คำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

ในกรอบ Test of Significance คลิกเลือกที่ Two-tailed เพื่อทำการทดสอบแบบสองทาง โดยปกติโปรแกรมจะเลือกให้อยู่แล้ว

คลิกที่ปุ่ม Options... จะได้นหน้าต่าง Bivariate Correlations : Options ดังภาพที่ 9.3



ภาพที่ 9.3

5. ในหน้าต่าง Bivariate Correlations : Options

ในกรอบ Statistics คลิกที่ Means and standard deviations ถ้าต้องการผลลัพธ์ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของตัวแปร age และ score

ในกรอบ Missing Values คลิกที่ Exclude cases pairwise หมายถึง จะไม่นำตัวอย่างที่มีค่า Missing ของตัวแปร age และ score มาหาค่าสหสัมพันธ์

แล้วคลิกที่ปุ่ม Continue หน้าต่างนี้จะถูกปิดไป

6. ในหน้าต่าง Bivariate Correlations

คลิกที่ปุ่ม OK จะได้ผลลัพธ์ดังภาพที่ 9.4

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
age	74.30	8.234	20
score	11.00	1.556	20

Correlations

		age	score
age	Pearson Correlation	1	.838**
	Sig. (2-tailed)		.000
	N	20	20
score	Pearson Correlation	.838**	1
	Sig. (2-tailed)	.000	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

ภาพที่ 9.4

จากภาพผลลัพธ์ในตาราง Descriptive Statistics แสดงค่า N คือ ขนาดของกลุ่มตัวอย่างเท่ากับ 20 ที่มีอายุเฉลี่ยเท่ากับ 74.3 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8.234 และมีคะแนนเฉลี่ยเท่ากับ 11 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.556

ในตาราง Correlations ได้ค่า Pearson Correlation ของตัวแปร age และ score เท่ากับ .838 และค่า Sig. (2-tailed) เท่ากับ .000 ซึ่งน้อยกว่าระดับนัยสำคัญที่กำหนด ($\alpha = .05$) สรุปได้ว่าปฏิเสธ $H_0 : \rho = 0$ นั่นคือ อายุและคะแนนสอบ มีความสัมพันธ์กันอย่างมีนัยสำคัญ

