



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

การวิเคราะห์ความสัมพันธ์ (Correlation Analysis) เป็นการศึกษาเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ว่ามีมากน้อยเพียงใด และเป็นการศึกษาถึงระดับของความสัมพันธ์ระหว่างตัวแปร ว่ามีทิศทางของความสัมพันธ์ ไปในทิศทางใด

การวัดความสัมพันธ์ โดยวิธีการทางสถิติมีอยู่หลายวิธี แตกต่างกันไป ขึ้นอยู่กับลักษณะของตัวแปรหรือมาตรการวัดค่าตัวแปร ลักษณะของข้อมูลเป็นแบบต่อเนื่อง หรือแบบไม่ต่อเนื่อง คุณสมบัติของการแจกแจงของข้อมูล ซึ่งมาตรวัดความสัมพันธ์ระหว่างตัวแปร เรียกว่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) และวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 2 ตัว ที่ใช้อย่างกว้างขวางที่สุดคือ สัมประสิทธิ์สหสัมพันธ์เชิงเส้นแบบ Pearson (Pearson Product Moment Correlation Coefficient) ที่ใช้สัญลักษณ์ ρ ซึ่งสามารถคำนวณหาได้โดย

$$\begin{aligned}\rho &= E \left[\frac{(X - \mu_x)(Y - \mu_y)}{\sigma_x \sigma_y} \right] \\ &= \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\} \{N\sum Y^2 - (\sum Y)^2\}}}\end{aligned}$$

โดยมีตัวประมาณค่าของ ρ คือ r ซึ่งคำนวณหาได้จาก

$$r = \frac{1}{n-1} E \left[\left(\frac{X - \bar{X}}{s_x} \right) \left(\frac{Y - \bar{Y}}{s_y} \right) \right]$$

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

เมื่อ	ρ	คือ สัมประสิทธิ์สหสัมพันธ์ของประชากรระหว่างตัวแปร X และ Y
	r	คือ สัมประสิทธิ์สหสัมพันธ์ของตัวอย่างระหว่างตัวแปร X และ Y
	ΣX	คือ ผลรวมของค่าตัวแปร X
	ΣY	คือ ผลรวมของค่าตัวแปร Y
	ΣXY	คือ ผลรวมของผลคูณระหว่างค่าตัวแปร X และตัวแปร Y
	ΣX^2	คือ ผลรวมของกำลังสองของค่าตัวแปร X
	ΣY^2	คือ ผลรวมของกำลังสองของค่าตัวแปร Y
	N	คือ จำนวนคู่ของค่าตัวแปร X และ Y
	n	คือ จำนวนคู่ของค่าตัวแปร X และ Y จากตัวอย่าง
	σ_x	คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรของตัวแปร X
	σ_y	คือ ส่วนเบี่ยงเบนมาตรฐานของประชากรของตัวแปร Y
	S_x	คือ ส่วนเบี่ยงเบนมาตรฐานจากตัวอย่างของตัวแปร X
	S_y	คือ ส่วนเบี่ยงเบนมาตรฐานจากตัวอย่างของตัวแปร Y

โดยค่าของ ρ หรือ r จะอยู่ระหว่าง -1 และ $+1$ เสมอ ถ้า ρ หรือ r มีค่าเป็นบวก แสดงว่า ตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นในทิศทางเดียวกัน นั่นคือ ถ้าตัวแปรตัวหนึ่งมีค่ามาก (หรือน้อย) อีกตัวหนึ่งจะมีค่ามาก (หรือน้อย) เช่นเดียวกัน แต่ถ้ามีค่าเป็นลบ แสดงว่า ตัวแปรทั้งสองมีความสัมพันธ์กันในทิศทางตรงกันข้าม นั่นคือ ถ้าตัวแปรตัวหนึ่ง

มีค่ามากอีกตัวหนึ่งจะมีค่าน้อย ถ้าค่า ρ หรือ r เท่ากับศูนย์ แสดงว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กันเลย แต่ถ้าค่า ρ หรือ r เท่ากับ ± 1 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์กันมาก หรือ มีความสัมพันธ์กันอย่างสมบูรณ์ (Perfect Correlation)

ถ้าค่า ρ หรือ r มีค่าเข้าใกล้ '+ 1' หรือ '-1' แสดงว่า ขนาดของความสัมพันธ์มีมาก แต่ถ้าค่า ρ หรือ r เข้าใกล้ 0 แสดงว่าขนาดความสัมพันธ์มีน้อย

การหาความสัมพันธ์ระหว่างตัวแปรที่สนใจศึกษานั้น นับได้ว่ามีความสำคัญมากในการนำไปใช้ประโยชน์ เพื่อการสรุปผลการวิเคราะห์วิจัยสำหรับงานด้านต่าง ๆ ได้อย่างดี โดยที่ลักษณะข้อมูลที่จะหาความสัมพันธ์นั้น จะต้องได้จากสิ่งเดียวกันเป็นคู่ ๆ หรือ เป็นข้อมูลที่มีความสัมพันธ์กัน ซึ่งตามหลักความจริงแล้วข้อมูลที่จะวัดความสัมพันธ์กันนั้นอาจจะได้มาจากแหล่งข้อมูลคนละแหล่งกัน การสกัดเก็บรวบรวมข้อมูลที่เป็นคู่ ๆ อย่างถูกต้องสมบูรณ์ ไม่อาจจะกระทำได้อย่างชัดเจน ทั้งนี้อาจเนื่องจาก ข้อมูลนั้นมีความสำคัญมาก และเป็นความลับไม่ควรให้รู้วไปยังหน่วยงาน หรือบุคคลอื่นที่ไม่เกี่ยวข้องทราบ ซึ่งเป็นปัญหาที่นักสถิติหรือนักวิเคราะห์ทั่ว ๆ ไป ต้องประสบอยู่เสมอ คือ ต้องใช้ข้อมูลที่ถูกต้องสกัดเก็บรวบรวมความสัมพันธ์กันเป็นคู่ ๆ อย่างครบถ้วนแล้วมาวิเคราะห์หาความสัมพันธ์ระหว่างข้อมูลเพื่อทำรายงานเสนอต่อไป แต่โดยความจริงแล้วบางครั้งนักวิเคราะห์ไม่สามารถจะกระทำได้ เนื่องจากลักษณะของข้อมูลอาจจะถูกสกัดเก็บรวบรวมไว้เป็นแบบกลุ่ม การวิเคราะห์หาความสัมพันธ์วิธีเดิมอาจจะไม่เหมาะสมนัก

วิธีที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร โดยการสกัดกลุ่มข้อมูลที่นำสนใจศึกษา มีด้วยกัน

4 วิธี ดังนี้

- 1.1.1 วิธีการจัดกลุ่มข้อมูล (Grouped data)
- 1.1.2 วิธีกำลังสองน้อยที่สุด (Least square)
- 1.1.3 วิธีของบาเรตและวาลท์ (Bartlett and Wald)
- 1.1.4 วิธีของคราเมอร์ (Cramer)

โดยจะทำการเปรียบเทียบ วิธีทั้ง 4 ดังกล่าวข้างต้นกับวิธีเดิมที่ใช้โดยทั่วไป คือ วิธีของเพียร์สัน (Pearson) และเปรียบเทียบวิธีวัดความสัมพันธ์ของข้อมูลแบบจัดกลุ่ม วิธีต่าง ๆ วิธีใดจะให้ผลการทดสอบที่มีความถูกต้อง เชื่อถือได้มากกว่ากัน เพื่อจะได้วิธีที่เหมาะสมที่จะนำไปหาความสัมพันธ์ เมื่อข้อมูลอยู่ในลักษณะของการจัดกลุ่ม

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อเปรียบเทียบวิธีประมาณค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อจัดกลุ่มข้อมูลแบบต่าง ๆ

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลแบบไม่จัดกลุ่ม กับแบบที่จัดกลุ่ม

1.3 สมมติฐานของการวิจัย

ตัวสถิติที่ใช้วัดความสัมพันธ์ระหว่างตัวแปรโดยการจัดกลุ่มข้อมูลทั้ง 4 วิธี ให้ผลแตกต่างกัน โดยตัวสถิติที่ได้จากวิธีกำลังสองน้อยที่สุด จะมีประสิทธิภาพที่สุด

1.4 ข้อตกลงเบื้องต้น

1.4.1 ตัวแปรทั้งสองต้องเป็นค่าต่อเนื่อง และมีกาแจกแจงแบบปกติสองตัวแปร (bivariate normal distribution)

1.4.2 ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นแบบเส้นตรง (linear relationship)

1.4.3 ข้อมูล (X_i, Y_i) ($i = 1, 2, \dots, n$) เป็นตัวอย่างเชิงกลุ่ม

1.5 ขอบเขตของการวิจัย

1.5.1 กำหนดให้จำนวนข้อมูลทั้งหมด (N) มี 4 ค่า คือ 100 200 500 และ 1,000 โดยแต่ละค่า N จะจัดแบ่งจำนวนข้อมูลในแต่ละกลุ่ม (m) ร้อยละ 5 10 20 และ 25 ของจำนวนข้อมูลทั้งหมดตามลำดับ

1.5.2 กำหนดให้ประชากรมีค่าสัมประสิทธิ์สหสัมพันธ์ RHO (ρ) เท่ากับ 0.10
0.25 0.50 0.75 0.90 และ 1.00 ตามลำดับ

1.5.3 จำลองข้อมูลในแต่ละลักษณะจำนวน 50 ครั้ง

1.6 คำจำกัดความ

Grouped data คือ ข้อมูลที่ถูกจัดกลุ่ม

Ordered data คือ ข้อมูลที่จัดลำดับแล้ว

Concomitants of order statistics คือ ค่าของตัวแปรหนึ่งที่ สัมพันธ์กับ
ค่าของอีกตัวแปรหนึ่งที่จัดลำดับแล้ว

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 ทราบวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลโดยการ จัดกลุ่มข้อมูล

1.7.2 ทราบวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลวิธีที่เหมาะสม เมื่อ
จัดแบ่งจำนวนกลุ่มและจำนวนข้อมูลในแต่ละกลุ่มแบบต่าง ๆ