บทที่ 2

เอกสาร และวรรณคดีที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยได้นำเสนอ โดยแบ่งเนื้อหาออกเป็น 4 ตอน ดังนี้

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับสหสัมพันธ์

ตอนที่ 2 สหสัมพันธ์ที่ศึกษาในงานวิจัยนี้

ตอนที่ 3 การแจกแจงที่เกี่ยวข้องกับงานวิจัยนี้

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ความรู้ทั่วไปเกี่ยวกับสหสัมพันธ์

สหสัมพันธ์ (Correlation) เป็นการวัดความสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรหรือมากกว่า หรือเป็นการหาขนาดของความสัมพันธ์ของตัวแปร 2 ตัวแปรหรือมากกว่า ซึ่งขนาดของความสัมพันธ์หรือระดับความสัมพันธ์ระหว่างตัวแปรนั้น เรียกว่า สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) (Kiess, 1989)

ลักษณะทั่วไปของสหสัมพันธ์

สหสัมพันธ์สามารถนำมาอธิบายหรือแบ่งแยกได้ในวิธีที่แตกต่างกันหลายวิธี ที่สำคัญมี 3 อย่าง ดังนี้คือ (Riggleman and Frisbee, 1951)

1. การแบ่งสหสัมพันธ์ออกเป็นแบบทิศทางเดียวกัน (Direct) กับทิศทางกลับกัน (Inverse) หลักเกณฑ์ที่ใช้คือการพิจารณาทิศทางของการเปลี่ยนแปลง กล่าวคือ ถ้าตัวแปรเปลี่ยนแปลงไปในทิศทางเดียวกันทั้งหมดไม่ว่าจะเป็นไปในทางเพิ่มขึ้นหรือลดลงก็ตาม จะเรียกสหสัมพันธ์แบบนี้ว่าเป็นสหสัมพันธ์ที่มีทิศทางเดียวกัน (Direct) แต่ถ้าหากว่าตัวแปรสองตัวมีทิศทางการเปลี่ยนแปลงที่ตรงข้ามกัน สหสัมพันธ์แบบนี้เรียกว่าเป็นสหสัมพันธ์แบบที่มีทิศทางกลับกัน (Inverse)

2. การแบ่งสหสัมพันธ์ออกเป็นแบบเชิงเส้นตรง (Linear) กับไม่ใช่เชิงเส้นตรง (Non-Linear) หลักเกณฑ์ที่ใช้คือการพิจารณาความคงที่ของอัตราการเปลี่ยนแปลง ถ้าหากว่าอัตราการเปลี่ยนแปลงของตัวแปรหนึ่ง มีอัตราส่วนคงที่กับอัตราการเปลี่ยนแปลงของอีกตัวหนึ่ง สหสัมพันธ์แบบนี้เรียกว่าเป็นสหสัมพันธ์เชิงเส้นตรง (Linear) แต่ถ้าหากปริมาณการเปลี่ยนแปลงในตัวแปรหนึ่งไม่ได้เป็นไปในอัตราส่วนที่คงที่ กับจำนวนการเปลี่ยนแปลงของอีกตัวแปรหนึ่ง สหสัมพันธ์แบบนี้เรียกว่าสหสัมพันธ์แบบไม่ใช่เชิงเส้นตรง (Non-Linear)

3. การแบ่งสหสัมพันธ์ออกเป็นแบบอย่างง่าย (Simple) แบบบางส่วน (Partial) และแบบพหุ (Mutiple) หลักเกณฑ์ที่ใช้คือการพิจารณาจำนวนชุดที่สัมพันธ์กัน ถ้าเป็นสหสัมพันธ์ระหว่าง 2 คุณลักษณะก็จะเรียกว่าเป็นสหสัมพันธ์อย่างง่าย (Simple) ถ้ามีมากกว่า 2 คุณลักษณะขึ้นไปจะเรียกว่าเป็นสหสัมพันธ์แบบพหุ (Multiple) และถ้าต้องการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวขึ้นไป โดยขจัดอิทธิพลของตัวแปรตัวอื่น ๆ ออกด้วยเรียกว่าเป็นสหสัมพันธ์บางส่วน (Partial)

ส่วนการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีอยู่หลายวิธีด้วยกัน การที่จะพิจารณาเลือกใช้วิธีใดในการคำนวณนั้นขึ้นอยู่กับข้อมูลภายใต้ปัจจัยหลายอย่าง เช่น หน่วยที่ใช้วัดข้อมูลอยู่ในมาตราการวัดใด คุณสมบัติการแจกแจงของข้อมูลเป็นแบบเชิงเส้นหรือไม่ใช่เชิงเส้น การกระจายของข้อมูลเป็นแบบต่อเนื่องหรือไม่ต่อเนื่อง ตัวอย่างเช่น วิธีการหาค่าสัมประสิทธิ์สห-สัมพันธ์ที่นิยมใช้ทั่วไป ได้แก่ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product-Moment Correlation Coefficient) ซึ่งใช้กับข้อมูลที่มีการวัดแบบช่วงหรือแบบอัตราส่วน และวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman Rank Correlation Coefficient) ซึ่งใช้กับข้อมูลที่มีการวัดแบบอันดับ เป็นต้น (Runyon and Haber, 1989)

ค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าอยู่ในช่วง -1 ถึง +1 ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าใกล้ ± 1 แสดงว่ามีความสัมพันธ์ระหว่างตัวแปรมาก แต่ถ้ามีค่าใกล้ 0 แสดงว่ามีความสัมพันธ์ระหว่างตัวแปรน้อย ส่วนเครื่องหมายเป็นการแสดงทิศทางของความสัมพันธ์

ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก หมายความว่า ตัวแปร 2 ตัวนั้นมีความสัมพันธ์ในทิศทางเดียวกัน กล่าวคือ ถ้ามีคะแนนหรือค่าสูงในตัวแปรหนึ่ง ก็จะมีคะแนนหรือค่าสูงในอีกตัวแปรหนึ่งด้วย หรือ ถ้ามีคะแนนหรือค่าต่ำในตัวแปรหนึ่ง ก็จะมีคะแนนหรือค่าต่ำในอีกตัวแปรหนึ่งด้วย ส่วนค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ หมายความว่า ตัวแปร 2 ตัวนั้นมีความสัมพันธ์ในทิศทางกลับกัน กล่าวคือ ถ้ามีคะแนนหรือค่าสูงในตัวแปรหนึ่ง ก็จะมีคะแนนหรือค่าต่ำในอีกตัวแปรหนึ่ง

เราสามารถพิสูจน์ให้เห็นได้ว่าค่าสัมประสิทธิ์สหสัมพันธ์นั้น จะมีค่าไม่น้อยกว่า -1.00 และไม่มากกว่า 1.00 ดังต่อไปนี้ (Glass and Hopkins, 1984)

เนื่องจาก $\sum_{i=1}^{n}\left(Z_{X_i}-Z_{Y_i}\right)^2$ จะมีค่ามากกว่าหรือเท่ากับ 0

นั่นคือ
$$\sum_{i=1}^{n}\left(Z_{X_i}-Z_{Y_i}\right)^2 \geq 0$$
$$\sum_{i=1}^{n}\left(Z_{X_i}^2+Z_{Y_i}^2-2Z_{X_i}Z_{Y_i}\right) \geq 0$$
$$\sum_{i=1}^{n}Z_{X_i}^2+\sum_{i=1}^{n}Z_{Y_i}^2 \geq 2\sum_{i=1}^{n}Z_{X_i}Z_{Y_i}$$

แต่ $\sum_{i=1}^{n}Z_{X_i}^2 = \sum_{i=1}^{n}Z_{Y_i}^2 = n-1$ และจาก $r_{XY}=\dfrac{\sum_{i=1}^{n}Z_{X_i}Z_{Y_i}}{n-1}$

จะได้
$$(n-1)+(n-1) \geq 2r_{XY}(n-1)$$
$$\frac{2(n-1)}{2(n-1)} \geq r_{XY}$$
$$1 \geq r_{XY}$$

และเนื่องจาก
$$\sum_{i=1}^{n}\left(Z_{X_i}+Z_{Y_i}\right)^2 \geq 0$$
$$\sum_{i=1}^{n}\left(Z_{X_i}^2+Z_{Y_i}^2+2Z_{X_i}Z_{Y_i}\right) \geq 0$$
$$\sum_{i=1}^{n}Z_{X_i}^2+\sum_{i=1}^{n}Z_{Y_i}^2 \geq -2\sum_{i=1}^{n}Z_{X_i}Z_{Y_i}$$
$$\frac{2(n-1)}{-2(n-1)} \leq r_{XY}$$
$$-1 \leq r_{XY}$$

ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็น 1.00 แสดงว่าตัวแปร 2 ตัวนั้นมีความสัมพันธ์สอดคล้องกันอย่างสมบูรณ์ ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็น -1.00 แสดงว่ามีความสัมพันธ์กันในทางตรงกันข้ามอย่างสมบูรณ์ และค่าสัมประสิทธิ์สหสัมพันธ์เป็นศูนย์ แสดงว่าตัวแปร 2 ตัวนั้นไม่มีความสัมพันธ์กัน (Sprinthall and Richard C., 1987)

ในการแปลความหมายของค่าสัมประสิทธิ์สหสัมพันธ์นั้น ค่าสัมประสิทธิ์สหสัมพันธ์ของข้อมูลจากตัวแปรสองตัวเป็นค่าที่ศึกษาจากตัวเลขที่วัดได้จากตัวแปรทั้งสองเท่านั้น ไม่ได้มีเหตุผลอะไรที่จะอ้างได้ว่าตัวแปรทั้งสองตัวเป็นเหตุเป็นผลต่อกัน เพราะอาจจะมีสาเหตุอื่นๆ อีกมากมายซึ่งเราไม่ได้นำมาศึกษา หรือแม้จะเป็นสาเหตุต่อกันเราก็ยังบอกไม่ได้ว่า X เป็นสาเหตุของ Y หรือ Y เป็นสาเหตุของ X ดังนั้นแม้จะพบว่าตัวแปรคู่ใดมีความสัมพันธ์กันเราจะยังไม่สรุปไปถึงความเป็นสาเหตุต่อกัน แต่ตัวแปรใดก็ตามที่มีส่วนเป็นสาเหตุต่อกันจะมีความสัมพันธ์อยู่เสมอ ซึ่งการศึกษาความสัมพันธ์นี้จึงเป็นจุดเริ่มต้นที่สำคัญของการศึกษาความเป็นสาเหตุต่อกัน (ทวีวัฒน์ ปิตยานนท์, 2536)

ในการหาค่าสัมประสิทธิ์สหสัมพันธ์โดยทั่วๆไปหาได้จากสมการพื้นฐาน ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร คือ

$$\rho = \frac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{N}$$

ค่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง คือ

$$r_{XY} = \frac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{n-1}$$

เมื่อ $Z_{X_i} = \dfrac{X_i - \overline{X}}{S_X}$, $Z_{Y_i} = \dfrac{Y_i - \overline{Y}}{S_Y}$

ดังนั้น

$$r_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{S_X}\right)\left(\frac{Y_i - \overline{Y}}{S_Y}\right)$$

$$= \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)/n-1}{S_X S_Y}$$

นั่นคือ

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

เมื่อ S_{XY} หมายถึง ความแปรปรวนร่วม (Covariance) ของตัวแปร X และ Y โดยที่

$$S_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1}$$

S_X หมายถึง ส่วนเบี่ยงเบนมาตรฐานของตัวแปร X โดยที่

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}$$

S_Y หมายถึง ส่วนเบี่ยงเบนมาตรฐานของตัวแปร Y โดยที่

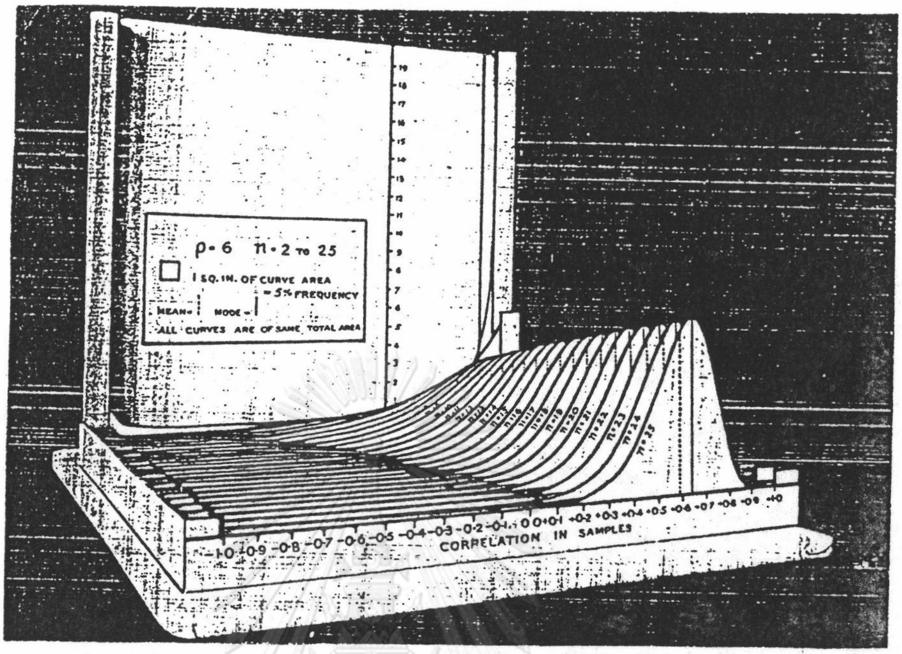
$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n-1}}$$

การแจกแจงของค่าสหสัมพันธ์

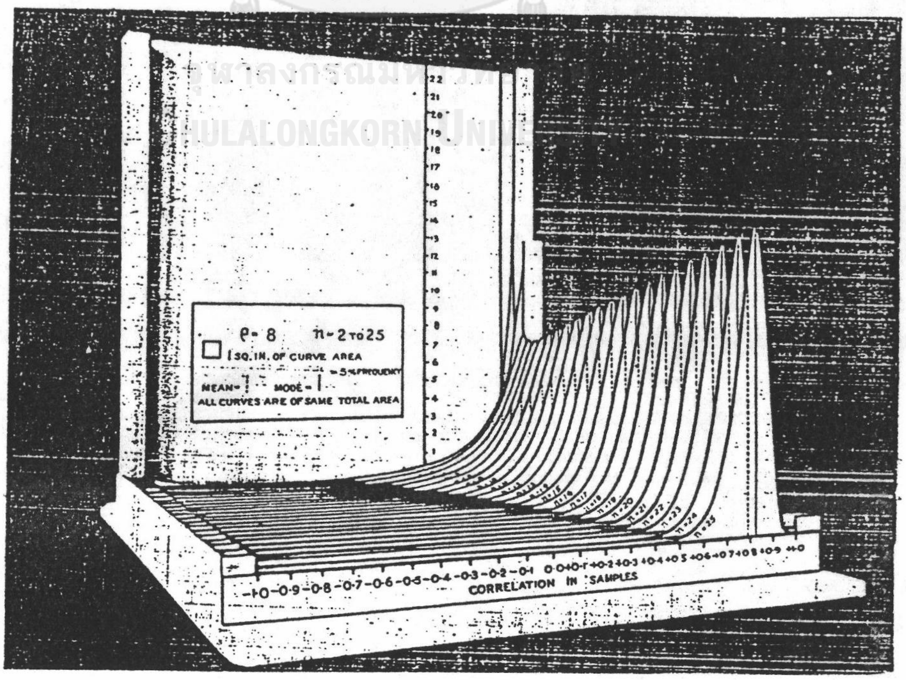
ในปี ค.ศ. 1915 R.A. Fisher (1915 : 507-521) ได้ศึกษาลักษณะการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ (r_{XY}) พบว่าการแจกแจงของค่า r_{XY} นั้นขึ้นอยู่กับค่าของ ρ (ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร) และ n (ขนาดของกลุ่มตัวอย่าง) เท่านั้น การแจกแจงของ r_{XY} จะสมมาตรเมื่อ $\rho = 0$ แต่จะมีลักษณะเบ้ ถ้า $\rho \neq 0$ โดยจะเบ้ซ้าย ถ้า $\rho > 0$ และจะเบ้ขวา ถ้า $\rho < 0$

ต่อมาในปี ค.ศ. 1916 H.E. Soper และคณะ (1916 : 318-413) ได้ศึกษาการแจกแจงของค่า r_{XY} เมื่อ n มีขนาดเล็ก และ $\rho \neq 0$ โดยศึกษาเมื่อ n เท่ากับ 2 ถึง 25 และ $\rho = 0.6$ และ $\rho = 0.8$ พบว่า การแจกแจงของค่า r_{XY} มีลักษณะเบ้ซ้ายเมื่อ $\rho = 0.6$ และ $\rho = 0.8$ ซึ่ง Soper ได้เสนอแผนภาพอย่างชัดเจนว่า ถ้า n มีขนาดเล็ก การแจกแจงของค่า r_{XY} ก็มีการกระจายมาก แต่ถ้า n มีขนาดใหญ่ขึ้น การแจกแจงของค่า r_{XY} ก็จะแคบลง ดังภาพ

แผนภาพที่ 1 การแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ (r_{XY}) เมื่อขนาดของกลุ่มตัวอย่างเป็น 2 ถึง 25 และ ρ = 0.6



แผนภาพที่ 2 การแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ (r_{XY}) เมื่อขนาดของกลุ่มตัวอย่างเป็น 2 ถึง 25 และ ρ = 0.8



ตอนที่ 2 สหสัมพันธ์ที่ใช้ในงานวิจัยนี้

สหสัมพันธ์แบบเพียร์สัน(Pearson Product-Moment Correlation)

ในปลายศตวรรษที่ 19 คาร์ล เพียร์สัน (Karl Pearson) นักคณิตศาสตร์ชาวอังกฤษ ได้พัฒนาและคิดค้นเกี่ยวกับการวัดความสัมพันธ์ขึ้น โดยในขณะนั้นเพียร์สันทำงานอยู่กับ ฟรานซิส กอลตัน (Francis Galton) ซึ่งเป็นบิดาของหลักการเกี่ยวกับลักษณะความแตกต่างระหว่างบุคคล โดยเพียร์สันให้เหตุผลว่า การที่ลักษณะของบุคคลมีความหลากหลาย ดังนั้น การที่จะวัดลักษณะเหล่านี้ให้สามารถนำไปใช้ประโยชน์ได้สูงสุด ควรขยายไปถึงการวัดเกี่ยวกับความสัมพันธ์ และเรียกค่าที่ได้จากการวัดความสัมพันธ์ระหว่างตัวแปรว่า ค่าสัมประสิทธิ์สหสัมพันธ์

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันนี้ ค่อนข้างจะเป็นพื้นฐานของสัมประสิทธิ์สหสัมพันธ์ตัวอื่นๆใช้ในสถานการณ์ที่ต้องการดัชนีชี้ความสัมพันธ์เชิงเส้นตรง (Linear Relationship) ระหว่างตัวแปร 2 ตัวแปร ที่ข้อมูลอยู่ในมาตราการวัดแบบช่วงหรือแบบอัตราส่วน สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน คือ r_{XY}

ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สามารถคำนวณได้จากข้อมูลที่อยู่ในรูปแบบต่าง ๆ ดังนี้

1. สูตรที่คำนวณค่าจากคะแนนส่วนเบี่ยงเบน (Deviation Score)

$$r_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]}}$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์

X_i หมายถึง ค่าคะแนนดิบแต่ละตัวของตัวแปร X

Y_i หมายถึง ค่าคะแนนดิบแต่ละตัวของตัวแปร Y

$\overline{X}$ หมายถึง ค่าเฉลี่ยของชุดในตัวแปร X

$\overline{Y}$ หมายถึง ค่าเฉลี่ยของชุดในตัวแปร Y

หรือ
$$r_{XY} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\left(\sum_{i=1}^{n} x_i^{2}\right)\left(\sum_{i=1}^{n} y_i^{2}\right)}}$$

เมื่อ
$$x_i = \left(X_i - \overline{X}\right)$$
$$y_i = \left(Y_i - \overline{Y}\right)$$

ที่มาของสูตรเป็นดังนี้

จาก
$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

และ
$$S_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1} \quad ; S_X = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}} \quad ; S_Y = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n-1}}$$

ดังนั้น
$$r_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) \Big/ (n-1)}{\sqrt{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \Big/ (n-1)\right]\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \Big/ (n-1)\right]}}$$

$$r_{XY} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right]\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right]}}$$

2. สูตรที่คำนวณค่าจากคะแนนมาตรฐาน (Standard Score)

$$r_{XY} = \frac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{n-1}$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์

Z_{X_i} หมายถึง คะแนนมาตรฐานของข้อมูลแต่ละตัวในตัวแปร X

Z_{Y_i} หมายถึง คะแนนมาตรฐานของข้อมูลแต่ละตัวในตัวแปร Y

n หมายถึง จำนวนข้อมูลในกลุ่มตัวอย่าง

ที่มาของสูตรเป็นดังนี้

จาก $$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

เนื่องจาก $S_{XY} = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$ และ $Z_{X_i} = \dfrac{(X_i - \overline{X})}{S_X}$; $Z_{Y_i} = \dfrac{(Y_i - \overline{Y})}{S_Y}$

จะได้ $$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) \Big/ (n-1)}{S_X S_Y}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\frac{(X_i - \overline{X})}{S_X}\frac{(Y_i - \overline{Y})}{S_Y}$$

$$r_{XY} = \frac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{n-1}$$

3. สูตรการคำนวณค่าจากคะแนนดิบ (Raw Score)

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ตามสูตรนี้จะคำนวณได้สะดวก และง่ายกว่าการคำนวณจากสูตรที่กล่าวมาแล้วข้างต้น ซึ่งในงานวิจัยนี้ผู้วิจัยใช้สูตรนี้ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

$$r_{XY} = \frac{n\left(\sum_{i=1}^{n} X_iY_i\right) - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์

n หมายถึง จำนวนข้อมูลในกลุ่มตัวอย่าง

$\sum_{i=1}^{n} X_iY_i$ หมายถึง ผลรวมของผลคูณระหว่างข้อมูลของตัวแปร X และ Y

$\sum_{i=1}^{n} X_i$ หมายถึง ผลรวมของข้อมูลแต่ละตัวในตัวแปร X

$\sum_{i=1}^{n} Y_i$ หมายถึง ผลรวมของข้อมูลแต่ละตัวในตัวแปร Y

$\sum_{i=1}^{n} X_i^2$ หมายถึง ผลรวมของกำลังสองของข้อมูลแต่ละตัวในตัวแปร X

$\sum_{i=1}^{n} Y_i^2$ หมายถึง ผลรวมของกำลังสองของข้อมูลแต่ละตัวในตัวแปร Y

ที่มาของสูตรเป็นดังนี้

จาก $$r_{XY} = \frac{S_{XY}}{S_XS_Y} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]\left[\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right]}}$$

และ $$\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}) = \sum_{i=1}^{n} X_iY_i - \overline{X}\sum_{i=1}^{n} Y_i - \overline{Y}\sum_{i=1}^{n} X_i + n\overline{X}\overline{Y}$$

เนื่องจาก $\sum_{i=1}^{n} X_i = n\overline{X}$ และ $\sum_{i=1}^{n} Y_i = n\overline{Y}$

แทนค่าจะได้ $\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right) = \sum_{i=1}^{n} X_iY_i - n\overline{X}\overline{Y} - n\overline{X}\overline{Y} + n\overline{X}\overline{Y}$

$$= \sum_{i=1}^{n} X_iY_i - n\overline{X}\overline{Y}$$

แต่ $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ และ $\overline{Y} = \dfrac{\sum_{i=1}^{n} Y_i}{n}$ ดังนั้น

$$\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right) = \sum_{i=1}^{n} X_iY_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)\Big/n$$

และ $\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\Big/n$

$$\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2 = \sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\Big/n$$

นั่นคือ $r_{XY} = \dfrac{\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i \Big/ n}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\Big/n\right]\left[\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\Big/n\right]}}$

$$r_{XY} = \frac{n\left(\sum_{i=1}^{n} X_iY_i\right) - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right]\left[n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right]}}$$

การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ r_{XY}

ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้นั้น เป็นค่าที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่าง ดังนั้นการที่จะบอกว่าตัวแปรทั้งสองมีค่าสัมประสิทธิ์สหสัมพันธ์ตามที่คำนวณได้จะต้องแสดงให้ได้ว่าค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรไม่เท่ากับ 0 นั่นคือต้องมีการทดสอบ

นัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ r_{XY} ก่อน โดยการตั้งสมมติฐานศูนย์ว่าสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0 ($\rho = 0$)

เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ r_{XY} มีลักษณะการแจกแจงเช่นเดียวกับการแจกแจงที (Student's t-distribution) ที่ชั้นแห่งความเป็นอิสระ (Degree of freedom) เท่ากับ n-2 ดังนั้นการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ r_{XY} จึงใช้สูตรการทดสอบที (t-test) ดังนี้

$$t = r_{XY}\frac{\sqrt{n-2}}{\sqrt{1-r_{XY}^{2}}} \quad ; \nu = n-2$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ν หมายถึง ชั้นแห่งความเป็นอิสระ

สหสัมพันธ์แบบสเปียร์แมน (Spearman Rank Correlation)

ในปี ค.ศ.1904 Charles Spearman ได้พัฒนาวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ โดยดัดแปลงมาจากวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เพื่อใช้ในการหาค่าสหสัมพันธ์ของข้อมูลที่อยู่ในมาตราการวัดแบบอันดับ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนนี้เป็นที่นิยมใช้กันอย่างแพร่หลาย เพราะสามารถคำนวณได้ง่ายและรวดเร็ว ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการคำนวณด้วยวิธีของสเปียร์แมนจะมีค่าใกล้เคียงกับค่าที่คำนวณด้วยวิธีของเพียร์สันในข้อมูลชุดเดียวกัน โดยเฉพาะเมื่อข้อมูลเป็นแบบอันดับไม่ซ้ำกันจะให้ค่าเท่ากับวิธีของเพียร์สันสัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน คือ r_S

ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน สามารถคำนวณได้จากข้อมูลในแบบต่างๆ 3 กรณี ดังนี้

1. กรณีข้อมูลไม่มีอันดับซ้ำ (Siegel, 1956)

$$r_S = 1 - \frac{6\sum d^2}{n^3 - n}$$

เมื่อ d หมายถึง ผลต่างของอันดับในแต่ละคู่

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ที่มาของสูตร เป็นดังนี้

เนื่องจากตัวแปร X และ Y เป็นอันดับที่ ดังนั้นตัวแปร X และ Y จะมีค่าเป็นเลขจำนวนเต็ม ได้แก่ 1, 2, 3, ..., n

$$\overline{X} = \overline{Y} = \frac{\sum X}{n} = \frac{1+2+\ldots+n}{n} = \frac{n+1}{2}$$

และ $$\sum X^2 = \sum Y^2 = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

ดังนั้น $$\sum X^2 = \sum\left(X - \overline{X}\right)^2 = \sum X^2 - \frac{\left(\sum X\right)^2}{n} = \sum Y^2 - \frac{\left(\sum Y\right)^2}{n}$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4}$$

$$= n(n+1)\left[\frac{2n+1}{6} - \frac{n+1}{4}\right]$$

$$= n(n+1)\left[\frac{4n+2-3n-3}{12}\right]$$

$$= \frac{n(n+1)(n-1)}{12}$$

$$= \frac{n(n^2-1)}{12}$$

ถ้าให้ d = อันดับที่ของ X — อันดับที่ของ Y ในแต่ละคู่

$$= X_i - Y_i$$

$$= \left(X - \overline{X}\right) - \left(Y - \overline{Y}\right) \qquad \text{เพราะว่า } \overline{X} = \overline{Y}$$

$$d = x - y$$

$$d^2 = (x-y)^2 = x^2 - 2xy + y^2$$

$$\sum d^2 = \sum x^2 + \sum y^2 - 2\sum xy$$

$$\sum xy = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2}$$

จากค่าสัมประสิทธิ์สหสัมพันธ์โดยทั่วไป

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

$$r_S = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}}$$

แต่ $\sum x^2 = \sum y^2$ ดังนั้น $r_S = \dfrac{2\sum x^2 - \sum d^2}{2\sum x^2}$

$$= 1 - \frac{\sum d^2}{2\sum x^2}$$

$$= 1 - \frac{\sum d^2}{\dfrac{2n(n^2-1)}{12}}$$

$$= 1 - \frac{6\sum d^2}{n(n^2-1)}$$

2. กรณีข้อมูลมีอันดับซ้ำ

กรณีที่มีอันดับซ้ำกันไม่ว่าจะเป็นซ้ำเฉพาะข้อมูลในตัวแปร X หรือซ้ำเฉพาะข้อมูลในตัวแปร Y หรือซ้ำทั้งข้อมูลในตัวแปร X และตัวแปร Y การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนต้องมีการแก้ไขความคลาดเคลื่อนที่เกิดขึ้น โดยใช้สูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์ใหม่ดังนี้ (Kendall, 1975)

$$r_S = 1 - \frac{6\left(\sum d^2 + T_X + T_Y\right)}{n^3 - n} \qquad \text{.......................(1)}$$

$$r_S = \frac{\frac{1}{6}\left(n^3 - n\right) - \sum d^2 - T_X - T_Y}{\sqrt{\frac{1}{6}\left(n^3 - n\right) - 2T_X}\sqrt{\frac{1}{6}\left(n^3 - n\right) - 2T_Y}} \qquad \text{................(2)}$$

เมื่อ T_X หมายถึง ค่าแก้เมื่อตัวแปร X มีค่าซ้ำกัน โดยที่

$$T_X = \frac{1}{12}\sum\left(t_x^{\,3} - t_x\right)$$

T_Y หมายถึง ค่าแก้เมื่อตัวแปร Y มีค่าซ้ำกัน โดยที่

$$T_Y = \frac{1}{12}\sum\left(t_y^{\,3} - t_y\right)$$

t_x หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มตัวแปร X

t_y หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มตัวแปร Y

3. กรณีข้อมูลมีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร

นอกจากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนใน 2 กรณีดังกล่าวแล้ว ในปี ค.ศ. 1960 Alan Stuart ได้พัฒนาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนโดยนำมาประยุกต์ใช้กับข้อมูลที่มีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร r X c เมื่อ r หมายถึง แถว และ c หมายถึง สดมภ์ ซึ่งตามปกติวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวเมื่อใช้กับข้อมูลจำนวนมาก การจัดอันดับของข้อมูลในตารางการณ์จรย่อมเกิดการซ้ำของอันดับและการซ้ำของอันดับดังกล่าวจะเกิดการซ้ำเป็นจำนวนมาก การหาค่าสัมประสิทธิ์สหสัมพันธ์ในสูตรกรณีที่ 2 จึงไม่เหมาะที่จะนำมาใช้ในลักษณะนี้ สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์วิธีนี้ คือ r_C

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ r_C ตามวิธีของ Stuart นั้นได้เสนอวิธีการหาค่าต่างๆ เพื่อนำมาใช้ในสูตรการคำนวณ ดังแสดงต่อไปนี้

		แยกประเภทตามสดมภ์			รวมความถี่	ค่าเฉลี่ยของอันดับ
		1	2	 c	(แถว)	(แถว)
	1	n_{11}	n_{12}	.... n_{1c}	$n_{1.}$	$R_{1.} = (n_{1.}+1)/2$
	2	n_{21}	n_{22}	.... n_{2c}	$n_{2.}$	$R_{2.} = n_{1.} + (n_{2.}+1)/2$
	3	n_{31}	n_{32}	.... n_{3c}	$n_{3.}$	$R_{3.} = \sum n_{i.} + (n_{3.}+1)/2$
แยกประเภท	.	.	.	.	.	.
ตามแถว	.	.	.	.	.	.
	.	.	.	.	.	.
	r	n_{r1}	n_{r2}	.... n_{rc}	$n_{r.}$	$R_{r.} = \sum n_{i.} + (n_{r.}+1)/2$
รวมความถี่		$n_{.1}$	$n_{.2}$	.... $n_{.c}$	N (ความถี่รวมทั้งหมด)	

ค่าเฉลี่ยของอันดับ $R_{.1} = \dfrac{n_{.1}+1}{2}$ $R_{.2} = n_{.1} + \dfrac{n_{.2}+1}{2}$ .... $R_{.C} = \sum n_{.i} + \dfrac{n_{.C}+1}{2}$

การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ r_C สามารถคำนวณได้จากสูตรต่อไปนี้

เมื่อ M หมายถึง ค่าต่ำสุดของจำนวนแถว หรือจำนวนสดมภ์

ข้อตกลงเบื้องต้นของ r_C

1. มีความเป็นอิสระในการรวบรวมข้อมูล
2. ข้อมูลจัดอยู่ในรูปตารางการณ์จร r X c
3. ตัวแปรทั้งสองตัวมีระดับการวัดอยู่ในมาตราการวัดแบบอันดับ
4. ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นความสัมพันธ์เชิงเส้นตรง
5. ผลรวมความถี่ของแถวและสดมภ์ใดๆ ต้องไม่เท่ากับศูนย์

การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์ r_S

ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้นั้น เป็นค่าที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่าง ดังนั้นการที่จะบอกว่าตัวแปรทั้งสองมีค่าสัมประสิทธิ์สหสัมพันธ์ตามที่คำนวณได้จะต้องแสดงให้ได้ว่าค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรไม่เท่ากับ 0 นั่นคือต้องมีการทดสอบนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ r_S ก่อน โดยการตั้งสมมติฐานศูนย์ว่าสัมประสิทธิ์สห-สัมพันธ์ของประชากรเท่ากับ 0 ($\rho = 0$)

เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ r_S มีลักษณะการแจกแจงเช่นเดียวกับการแจกแจงที (Student's t-distribution) ที่ชั้นแห่งความเป็นอิสระ (Degree of freedom) เท่ากับ n-2 ดังนั้นการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ r_S จึงใช้สูตรการทดสอบที (t-test) ดังนี้

$$t = \frac{r_{S\sqrt{n-2}}}{\sqrt{1-r_S^{\,2}}} \qquad ; \ \nu = n\text{-}2$$

เมื่อ r_S หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ν หมายถึง ชั้นแห่งความเป็นอิสระ

ตอนที่ 3 การแจกแจงที่เกี่ยวข้องกับงานวิจัยนี้

การแจกแจงแบบปกติสองตัวแปร (Bivariate Normal Distribution)

ถ้า X และ Y เป็นตัวแปรสุ่มแบบต่อเนื่อง และมีการแจกแจงแบบปกติสองตัวแปร ฟังก์ชั่นความน่าจะเป็นของตัวแปรทั้งสอง เขียนเป็นสมการได้ดังนี้

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

เมื่อ $-\infty < x < \infty$ และ $-\infty < y < \infty$

การแจกแจงแบบปกติสองตัวแปรตามสมการดังกล่าว จะประกอบด้วยพารามิเตอร์ $\mu_1, \mu_2, \sigma_1, \sigma_2$ และ ρ โดยที่ $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, $\sigma_1 > 0$, $\sigma_2 > 0$ และ $-1 < \rho < 1$ จะได้ x และ y เป็นตัวแปรที่มีการแจกแจงแบบปกติสองตัวแปร และมี

ค่าเฉลี่ย $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

และเมตริกความแปรปรวน $\Sigma = \begin{bmatrix} \sigma_1^{\,2} & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^{\,2} \end{bmatrix}$

โดยที่ ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และ Y

μ_1 คือ ค่าเฉลี่ยของตัวแปร X

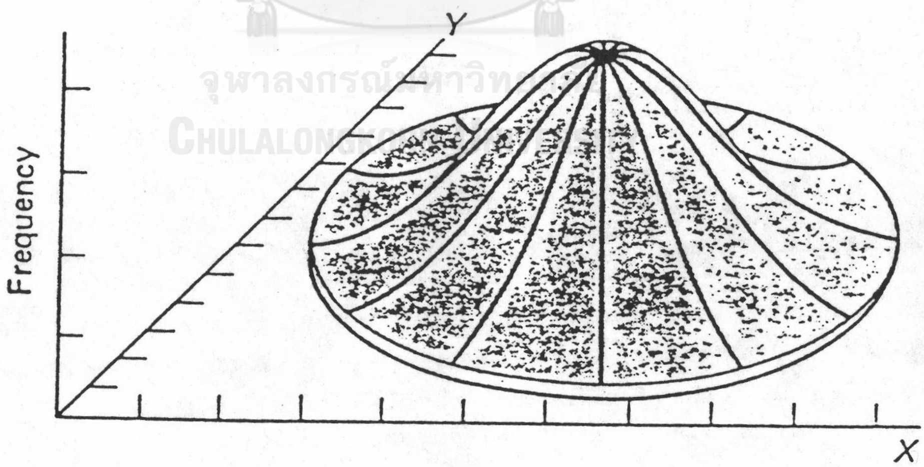
μ_2 คือ ค่าเฉลี่ยของตัวแปร Y

σ_1^2 คือ ค่าความแปรปรวนของตัวแปร X

σ_2^2 คือ ค่าความแปรปรวนของตัวแปร Y

ในกรณีที่การแจกแจงแบบปกติสองตัวแปร มีค่า $\mu_1 = \mu_2 = 0$ และ $\sigma_1^2 = \sigma_2^2 = 1$ เรียกการแจกแจงแบบปกติสองตัวแปรนี้ว่า **การแจกแจงแบบปกติมาตรฐานสองตัวแปร** (Standard Bivariate Normal Distribution)

แผนภาพที่ 1 การแจกแจงแบบปกติสองตัวแปร



การแจกแจงแบบมีเงื่อนไข (Conditional Distribution)

กำหนดให้ตัวแปรสุ่ม X และ Y มีการแจกแจงแบบปกติสองตัวแปร จะได้ว่าการแจกแจงแบบมีเงื่อนไขของตัวแปร X เมื่อกำหนดค่า Y = y ซึ่งเขียนแทนด้วยสัญลักษณ์ X/Y=y จะมีลักษณะการแจกแจงแบบปกติที่มีค่าเฉลี่ย $(\mu_{X/Y=y})$ เท่ากับ $\mu_1 + (\rho\sigma_1|\sigma_2)(Y-\mu_2)$ และค่าความแปรปรวน $(\sigma^2{}_{X/Y=y})$ เท่ากับ $\sigma_1{}^2\,(1-\rho^2)$ และจะได้ว่าการแจกแจงแบบมีเงื่อนไขของตัวแปร Y เมื่อกำหนดค่า X = x ซึ่งเขียนแทนด้วยสัญลักษณ์ Y/X=x จะมีลักษณะเป็นการแจกแจงแบบปกติที่มีค่าเฉลี่ย $(\mu_{\,Y/X=x})$ เท่ากับ $\mu_2 + (\rho\sigma_2|\sigma_1)(X-\mu_1)$ และค่าความแปรปรวน $(\sigma^2{}_{Y/X=x})$ เท่ากับ $\sigma_2{}^2\,(1-\rho^2)$ (Freund, 1992) และเนื่องจาก X และ Y เป็นการแจกแจงแบบปกติสองตัวแปรที่มีค่าเฉลี่ย $\mu_1 = \mu_2 = 0$, ค่าความแปรปรวน $\sigma_1{}^2 = \sigma_2{}^2 = 1$ และค่าสัมประสิทธิ์สหสัมพันธ์ ρ = Cov(X,Y) ดังนั้นจะหาการแจกแจงของตัวแปร Y/X=x ได้ดังนี้

จากสมการ $\mu_{Y/X=x} = \mu_2 + (\rho\sigma_2|\sigma_1)(X-\mu_1)$

แทนค่า $\mu_1 = \mu_2 = 0,\ \sigma_1{}^2 = \sigma_2{}^2 = 1$

$$\mu_{Y/X=x} = \rho X$$

และ $\mu_{Y/X=x} = \sigma_2{}^2\,(1-\rho^2)$

แทนค่า $\sigma_2{}^2 = 1$

$$\sigma^2{}_{Y/X=x} = 1-\rho^2$$

จะได้ว่า $Y \sim N(\rho X,\ 1-\rho^2)$

นั่นคือ ตัวแปร Y จะมีลักษณะเป็นการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ ρX มีค่าความแปรปรวนเท่ากับ $1-\rho^2$ และในการวิจัยนี้ผู้วิจัยได้สร้างประชากร ที่มีลักษณะการแจกแจงแบบปกติสองตัวแปรจากตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐานสองตัวแปร คือ X และ Y ดังสมการ (Brownlee, 1965)

$$Y_i = \rho X_i + \sqrt{1-\rho^2}\cdot W_i$$

เมื่อ W_i คือ ค่าของตัวแปรที่อิสระจากค่าตัวแปร X_i

การแจกแจงที (Student's t - Distribution)

การแจกแจงที คิดค้นโดย William Sealy Gosset ตีพิมพ์ครั้งแรกในปี ค.ศ. 1908 โดยใช้นามปากกาว่า "Student" ดังนั้นการแจกแจงที จึงมีชื่อว่า Student t-Distribution หรือ Student's t-Distribution หรือบางทีเรียกว่า t-Distribution (Freund, 1992)

สมการของการแจกแจงที

ถ้า Y และ Z เป็นตัวแปรสุ่มที่อิสระต่อกัน Y เป็นตัวแปรสุ่มที่มีการแจกแจงแบบไคสแควร์ ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν และ Z เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติ ตัวแปรสุ่ม T จะมีสมการดังนี้

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

ซึ่งได้รับจากฟังก์ชั่นความน่าจะเป็นของตัวแปร T ดังนี้

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\cdot\left(1+\frac{t^2}{\nu}\right)^{\frac{-\nu+1}{2}} \qquad ; -\infty < t < \infty$$

เมื่อ $\Gamma(P)$ คือ ฟังก์ชั่นแกรมม่า มีค่า $= \int_0^\infty x^{P-1}e^{-x}dx \quad ; P > 0$

เรียกตัวแปรสุ่ม T ว่ามีการแจกแจงที ที่ชั้นแห่งความเป็นอิสระเท่ากับ ν

คุณสมบัติของการแจกแจงที

1. มีลักษณะสมมาตร เมื่อเทียบกับแกน $t = 0$ จุดนี้จะเป็นค่าเฉลี่ย มัธยฐาน และฐานนิยม

2. ในกรณีที่ $\nu = 1$ การแจกแจงจะไม่มีค่าเฉลี่ย แต่ถ้า $\nu = 2, 3, \ldots$ ค่าเฉลี่ยจะมีค่าเท่ากับ 0

3. เมื่อ $\nu = 1, 2$ การแจกแจงจะไม่มีค่าความแปรปรวน แต่ถ้า $\nu = 3, 4, \ldots$ ค่าความแปรปรวนจะเท่ากับ $\nu / \nu - 2$

4. ถ้า ν มีค่ามากๆ การแจกแจงทีจะใกล้เคียงเป็นการแจกแจงแบบปกติ

5. กำลังสองของ t ที่ $\nu = k$ จะมีค่าเท่ากับ F ที่ $\nu = 1$ กับ k

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

กรรณิกา เลียงเจริญสิทธิ์ (2527: 48 - 49) ได้ทำการศึกษาการแจกแจงของค่าสหสัมพันธ์แบบปกติสองตัวแปร ที่ระดับความสัมพันธ์ของประชากร (ρ) เท่ากับ 0.1, 0.2, ..., 0.9 ผลการวิจัยพบว่า ลักษณะการแจกแจงของข้อมูลมีความเบ้ ในกรณีที่ $\rho = 0$ เมื่อขนาดของกลุ่มตัวอย่างมีน้อยกว่า 25 แต่เมื่อขนาดของกลุ่มตัวอย่างมีค่าเท่ากับหรือมากกว่า 25 การแจกแจงของค่าสหสัมพันธ์จะมีลักษณะเป็นแบบปกติโดยประมาณ และยืนยันว่าเมื่อแปลงค่าสหสัมพันธ์ โดยวิธี Fisher's Z- trasformation แล้ว Z_F จะมีลักษณะการแจกแจงเป็นแบบปกติโดยประมาณ ข้อสรุปที่สำคัญที่ได้จากการศึกษา คือ ในการทดสอบสมมติฐานกรณีที่ ρ มีค่าอื่น ๆ ที่ไม่เท่ากับ 0 ที่ระดับนัยสำคัญ (α) เท่ากับ .01 ขนาดของกลุ่มตัวอย่างที่เหมาะสมควรใช้ตั้งแต่ 9 ขึ้นไป ที่ระดับนัยสำคัญ เท่ากับ .05 และ .10 ควรใช้ตั้งแต่ 5 ขึ้นไป

ทองดี แย้มสรวล (2530: 87-94) ได้ทำการศึกษาลักษณะการแจกแจง การควบคุมความคลาดเคลื่อนประเภทที่ 1 เมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0 ($\rho = 0$) และอำนาจการทดสอบ เมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0.1, 0.2, ..., 0.9 ของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน แคนดอลเทา และเครมเมอร์วี ศึกษา

ที่ระดับนัยสำคัญ .05 และ .01 ผลการวิจัยพบว่า ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน สามารถประมาณค่า ρ ได้ใกล้เคียงที่สุด ที่ทุกๆค่าของ ρ สถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ทุกกรณีที่ศึกษา ส่วนอำนาจการทดสอบนั้นการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน มีอำนาจในการทดสอบสูงที่สุด

Agresti (1976 : 49-55) ได้ศึกษาเปรียบเทียบความมีเสถียรภาพของการวัด (Stability of the Measure) ความสัมพันธ์แบบอันดับในรูปตารางจำแนกประเภท (Cross Classification Tables) วิธีต่าง ๆ ได้แก่ วิธีหาค่าสัมประสิทธิ์แกมม่า (γ), แคนดอลเทา (τ_b, τ_c), สเปียร์แมน (r_S, r_C) และเพียร์สัน (r_{XY}) เมื่อข้อมูลมีลักษณะการแจกแจงแบบปกติสองตัวแปร โดยศึกษาที่ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร ρ = 0.2, 0.5 และ 0.8 โดยมีการเปลี่ยนค่าของสิ่งที่จะทำการศึกษาบางอย่าง คือ จำนวนแถว จำนวนสดมภ์ และผลรวมของความถี่ตามแนวแถวและสดมภ์ ผลการวิจัยพบว่า เมื่อข้อมูลไม่ได้จัดกลุ่ม (Ungroup Data) ค่าสัมประสิทธิ์τ_b มีเสถียรภาพดีที่สุด โดยพิจารณาจากค่า Mean Square Error แต่ค่าสัมประสิทธิ์ τ_b ก็มีข้อบกพร่องบางประการเกี่ยวกับเรื่องประสิทธิ์ภาพในการทดสอบสมมติฐานและมีความผันแปรไปตามสัดส่วนของข้อมูลที่มีการซ้ำกันอีกด้วย

Brown and Benedetti (1977: 309-315) ได้พัฒนาสูตรการคิดคำนวณค่า Asymptotic Standard Error (ASE) ของค่าสัมประสิทธิ์สหสัมพันธ์ γ, τ_b, τ_c, r_S, r_{XY} ขึ้นมาอีกลักษณะหนึ่งเพื่อนำมาใช้ในการทดสอบสมมติฐานศูนย์ ที่ว่าตัวแปรทั้งสองไม่มีความสัมพันธ์กัน หรือความสัมพันธ์มีค่าเท่ากับ 0 (H_0 : $\rho = 0$) โดยใช้สถิติทดสอบซี (Z - test) ในการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละวิธี แล้วเปรียบเทียบผลที่ได้กับการทดสอบสมมติฐานศูนย์ โดยใช้ค่า ASE ตามสูตรเดิมซึ่งในการศึกษาครั้งนี้ คณะผู้วิจัยได้ทำการทดลองซ้ำในแต่ละกรณีจำนวน 1,000 ครั้ง เมื่อข้อมูลอยู่ในรูปตารางการณ์จรขนาด 4X4 และ 8X8 โดยตารางการณ์จรขนาด 4X4 จะศึกษากับกลุ่มตัวอย่างขนาด 25, 50, 100 และ 200 ส่วนตารางการณ์จรขนาด 8X8 ศึกษากับกลุ่มตัวอย่างขนาด 25, 50, 100, 200 และ 400 ที่ระดับนัยสำคัญทางสถิติ เท่ากับ .20, .10, .05, และ .01 ผลการวิจัยพบว่า การทดสอบซีที่ใช้

ในการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ค่า ASE ที่คณะผู้วิจัยได้พัฒนาขึ้นมาใหม่ มีความสามารถในการทดสอบสมมติฐานศูนย์ได้ดีกว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ค่า ASE แบบเดิม