บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

การศึกษาผลของข้อมูลที่มีค่าห่างผิดปกติจากกลุ่มต่อความแกร่งของสัมประสิทธิ์สหสัมพันธ์ มีแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง ในประเด็นดังต่อไปนี้

1. ความรู้ทั่วไปเกี่ยวกับสหสัมพันธ์

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

1.1 ลักษณะทั่วไปของสหสัมพันธ์

จารุณี ไชยมูล (2542) ได้กล่าวว่า สหสัมพันธ์สามารถนำมาอธิบายหรือแบ่งแยกได้ในวิธีที่แตกต่างกันหลายวิธี ที่สำคัญมี 3 อย่าง ดังนี้

1.1.1 การแบ่งสหสัมพันธ์ออกเป็นแบบทิศทางเดียวกัน (Direct) กับทิศทางกลับกัน (Inverse) หลักเกณฑ์ที่ใช้คือ การพิจารณาทิศทางของการเปลี่ยนแปลงดังกล่าวคือ ถ้าตัวแปรทั้งสองเปลี่ยนแปลงไปในทิศทางเดียวกันทั้งหมดไม่ว่าจะเป็นไปในทางเพิ่มขึ้นหรือลดลงก็ตาม จะเรียกสหสัมพันธ์แบบนี้ว่าเป็นสหสัมพันธ์ที่มีทิศทางเดียวกัน แต่ถ้าตัวแปรทั้งสองตัวมีทิศทางการเปลี่ยนแปลงที่ตรงข้ามกัน สหสัมพันธ์แบบนี้เรียกว่าสหสัมพันธ์แบบมีทิศทางกลับกัน

1.1.2 การแบ่งสหสัมพันธ์ออกเป็นแบบเชิงเส้นตรง (Linear) กับไม่ใช่เชิงเส้นตรง (Non-Linear) หลักเกณฑ์ที่ใช้คือการพิจารณาความคงที่ของอัตราการเปลี่ยนแปลง ถ้าหากว่าอัตราการเปลี่ยนแปลงของตัวแปรหนึ่งมีอัตราส่วนคงที่กับอัตราการเปลี่ยนแปลงของอีกตัวแปรหนึ่ง สหสัมพันธ์แบบนี้เรียกว่าสหสัมพันธ์เชิงเส้นตรง แต่ถ้าหากอัตราส่วนการเปลี่ยนแปลงในตัวแปรหนึ่งไม่ได้เป็นไปในอัตราส่วนที่คงที่ กับจำนวนการเปลี่ยนแปลงของอีกตัวแปรหนึ่ง สหสัมพันธ์แบบนี้เรียกว่า สหสัมพันธ์แบบไม่ใช่เชิงเส้นตรง

1.1.3 การแบ่งสหสัมพันธ์ออกเป็นแบบอย่างง่าย (Simple) แบบบางส่วน (Partial) และแบบพหุ (Multiple) หลักเกณฑ์ที่ใช้คือการพิจารณาจำนวนตัวแปรที่นำมาวิเคราะห์ความสัมพันธ์ ถ้าเป็นสหสัมพันธ์ระหว่างตัวแปร 2 ชุด จะเรียกว่าเป็นสหสัมพันธ์อย่างง่าย (Simple Correlation) ถ้ามีข้อมูลมีตัวแปรมากกว่า 2 ชุดขึ้นไปจะเรียกว่าเป็นสหสัมพันธ์แบบพหุ (Multiple Correlation) และถ้าต้องการศึกษาความสัมพันธ์ระหว่าง 2 ตัวแปรขึ้นไป โดยขจัดอิทธิพลของตัวแปรอื่นๆ ออกด้วยเรียกว่าเป็นสหสัมพันธ์บางส่วน (Partial Correlation)

ส่วนการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีอยู่หลายวิธีด้วยกัน การที่จะพิจารณาเลือกใช้วิธีใดในการคำนวณนั้นขึ้นอยู่กับปัจจัยหลายอย่าง เช่น ประเภทของข้อมูลเป็นแบบต่อเนื่องหรือแบบไม่ต่อเนื่อง การแจกแจงของข้อมูลเป็นการแจกแจงแบบปกติหรือแบบไม่ปกติ ข้อมูลอยู่ในมาตรวัดแบบใด วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ที่นิยมใช้ทั่วไป ได้แก่ วิธีหาค่าสัมประสิทธิ์แบบเพียร์สัน ซึ่งใช้กับข้อมูลที่มีมาตรวัดแบบช่วงและแบบอัตราส่วน วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ซึ่งใช้ได้กับข้อมูลที่มีมาตรวัดตั้งแต่แบบอันดับ แบบช่วงและแบบอัตราส่วน เป็นต้น (Runyon and Haber, 1991)

1.2 การวัดความสัมพันธ์เชิงเส้น

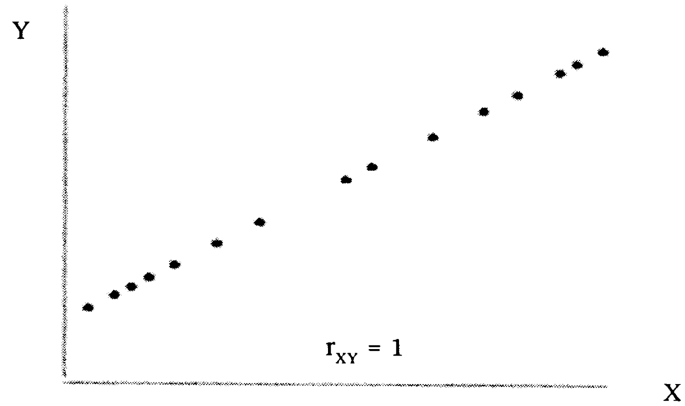
การศึกษาเรื่องสหสัมพันธ์ เป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรที่สนใจว่ามีความสัมพันธ์กันหรือไม่ และความสัมพันธ์ดังกล่าวเป็นไปในทิศทางใด เช่น การศึกษาความสัมพันธ์ระหว่างน้ำหนักกับส่วนสูง คะแนนเฉลี่ยของนักเรียนชั้นมัธยมศึกษาปีที่ 6 กับคะแนนสอบเข้าระดับอุดมศึกษา ในการพิจารณาว่าค่าความสัมพันธ์ระหว่างตัวแปรมากน้อยเพียงใดนั้น ทราบได้โดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์

สัมประสิทธิ์สหสัมพันธ์เป็นค่าที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรโดยมีค่าตั้งแต่ -1.00 ถึง 1.00 ถ้ามีตัวแปร X และตัวแปร Y ความสัมพันธ์ของตัวแปร X และ Y อาจเป็นได้ดังนี้

1.2.1 ตัวแปร X และ Y มีความสัมพันธ์กัน

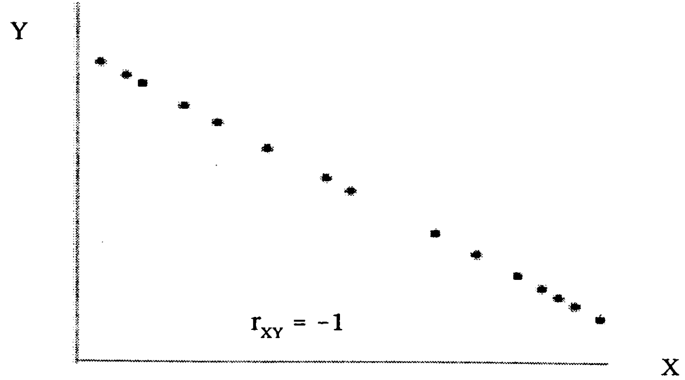
1.2.1.1 ตัวแปร X และ Y มีความสัมพันธ์กันอย่างสมบูรณ์ (Perfect Correlation) ซึ่งมี 2 ลักษณะคือ

(1) ความสัมพันธ์กันทางบวกอย่างสมบูรณ์หรือในทางเดียวกัน (Positive Correlation) กรณีนี้สัมประสิทธิ์สหสัมพันธ์จะมีค่าเป็น 1 ถ้าตัวแปร X เพิ่มขึ้น ตัวแปร Y ก็จะเพิ่มขึ้น ถ้าเขียนแผนภาพการกระจาย (Scatter Diagram) ลักษณะการกระจายของข้อมูลตัวแปร Xและ Y จะเป็นเส้นตรง (ภาพที่1)



ภาพที่ 1 ข้อมูลที่มีความสัมพันธ์ทางบวกอย่างสมบูรณ์

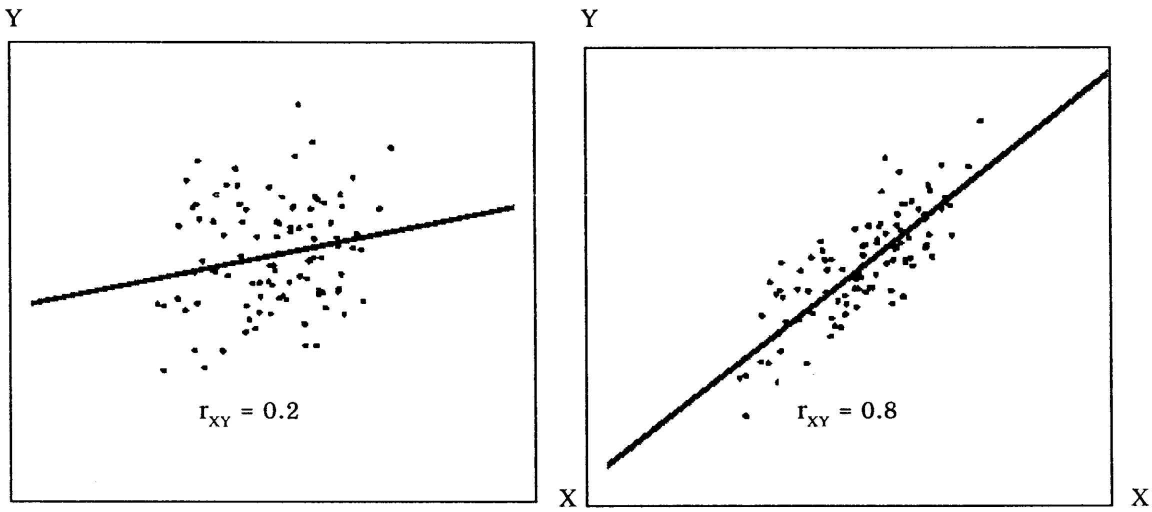
(2) ความสัมพันธ์กันทางลบอย่างสมบูรณ์หรือในทางตรงกันข้าม ในกรณีนี้สัมประสิทธิ์จะมีค่าเป็น -1 ถ้าตัวแปร X เพิ่มขึ้น ตัวแปร Y จะลดลง (ภาพที่ 2)



ภาพที่ 2 ข้อมูลมีความสัมพันธ์ทางลบอย่างสมบูรณ์

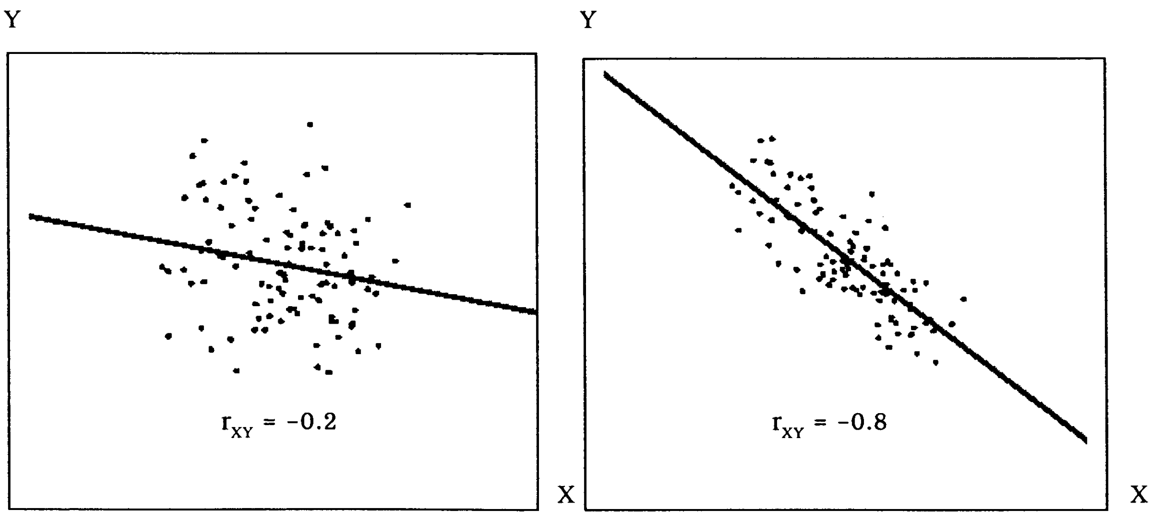
1.2.1.2 ตัวแปร X และ Yมีความสัมพันธ์กันอย่างไม่สมบูรณ์มี 2 ลักษณะดังนี้

(1) มีความสัมพันธ์กันทางบวกอย่างไม่สมบูรณ์ ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง 0 กับ 1 เช่น ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.2 แสดงว่าตัวแปร X และตัวแปร Y มีความสัมพันธ์กันค่อนข้างน้อยแต่เป็นไปในทางเดียวกัน ลักษณะการกระจายของข้อมูลกระจายออกจากกันมาก ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.8 แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์ค่อนข้างมากและเป็นไปในทิศทางเดียวกัน (ภาพที่ 3)



ภาพที่ 3 ข้อมูลมีความสัมพันธ์กันทางบวกอย่างไม่สมบูรณ์

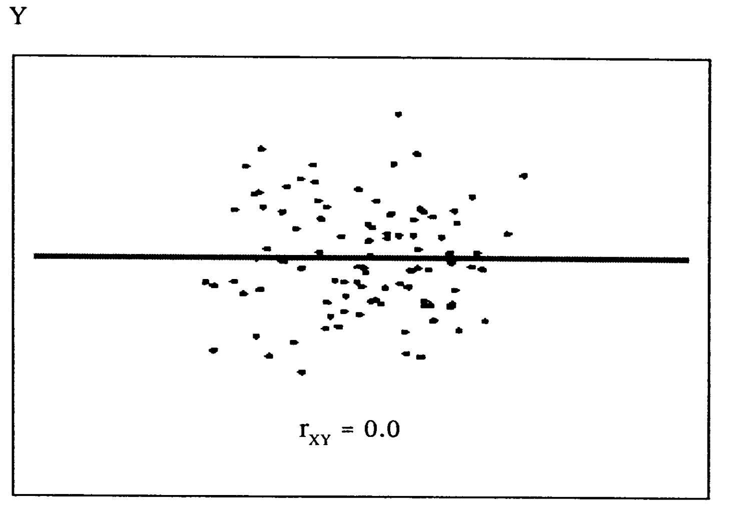
(2) มีความสัมพันธ์ทางลบอย่างไม่สมบูรณ์ ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง -1 กับ 0 เช่น ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.2 และ -0.8 แสดงว่าตัวแปร X และ Y มีความสัมพันธ์กันในทางตรงกันข้ามโดยมีความสัมพันธ์ค่อนข้างน้อยและค่อนข้างมากตามลำดับ (ภาพที่ 4)



ภาพที่ 4 ข้อมูลที่มีความสัมพันธ์ทางลบอย่างไม่สมบูรณ์

1.2.2 ตัวแปร X และ Y ไม่มีความสัมพันธ์กัน

ในกรณีที่ตัวแปร X และ ตัวแปร Y ไม่มีความสัมพันธ์กันเลย สัมประสิทธิ์สหสัมพันธ์มีค่าเป็น 0 ลักษณะการกระจายของข้อมูลมีรูปแบบไม่แน่นอน ทำให้ไม่สามารถระบุได้ว่าถ้าตัวแปรตัวหนึ่งเพิ่มขึ้นหรือลดลงแล้ว จะทำให้ตัวแปรอีกตัวหนึ่งเปลี่ยนแปลงไปในทิศทางใด (ภาพที่ 5)



ภาพที่ 5 ข้อมูลที่ไม่มีความสัมพันธ์กัน

จากลักษณะความสัมพันธ์ที่เป็นไปได้ข้างต้นสามารถสรุปได้ดังนี้ ถ้าค่าสัมประสิทธิ์มีค่าใกล้ ±1 แสดงว่ามีความสัมพันธ์ระหว่างตัวแปรมาก โดยค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็น 1.00 แสดงว่าตัวแปร 2 ตัวแปรนั้น มีค่าความสัมพันธ์สอดคล้องกันอย่างสมบูรณ์ ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็น -1.00 แสดงว่ามีความสัมพันธ์กันในทางตรงกันข้ามอย่างสมบูรณ์ ถ้ามีค่าใกล้ 0 แสดงว่ามีความสัมพันธ์ระหว่างตัวแปรน้อย และค่าสัมประสิทธิ์สหสัมพันธ์เป็นศูนย์ แสดงว่าตัวแปร 2 ตัวนั้นไม่มีความสัมพันธ์กัน (Sprinthall, 2003) ส่วนเครื่องหมายเป็นการแสดงทิศทางของความสัมพันธ์ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก หมายความว่า ตัวแปร 2 ตัวแปรนั้น มีความสัมพันธ์ในทิศทางเดียวกัน กล่าวคือ ถ้าตัวแปรแรกมีคะแนนสูง ตัวแปรอีกตัวแปรหนึ่งก็จะมีคะแนนหรือค่าสูง ถ้าค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบ หมายความว่ามีความสัมพันธ์ในทิศทางกลับกัน กล่าวคือ ถ้าตัวแปรแรกมีคะแนนสูง ตัวแปรอีกตัวแปรหนึ่ง ก็จะมีคะแนนหรือค่าต่ำ (Glass and Hopkins, 1984)

ในการแปลความหมายของค่าสัมประสิทธิ์สหสัมพันธ์นั้น ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่คำนวณได้จะบอกได้เพียงว่าตัวแปรทั้งสองมีความสัมพันธ์มากน้อยเพียงใดและมีความสัมพันธ์ในทิศทางใด ไม่ได้มีเหตุผลอะไรที่จะอ้างว่าตัวแปรทั้งสองนั้นเป็นเหตุเป็นผลต่อกัน หรือแม้จะเป็นสาเหตุต่อกันเราก็ยังบอกไม่ได้ว่า X เป็นสาเหตุของ Y หรือ Y เป็นสาเหตุของ X ดังนั้นแม้จะพบว่าตัวแปรคู่ใดมีความสัมพันธ์กันเราจะยังไม่สรุปไปถึงความเป็นสาเหตุต่อกัน แต่ตัวแปรใดก็ตามที่มีส่วนเป็นสาเหตุต่อกันจะมีความสัมพันธ์อยู่เสมอ ซึ่งการศึกษาความสัมพันธ์นี้จึงเป็นจุดเริ่มต้นที่สำคัญของการศึกษาความเป็นสาเหตุต่อกัน

1.3 การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์

ในการหาค่าสัมประสิทธิ์สหสัมพันธ์โดยทั่ว ๆ ไป หาค่าได้จากสมการพื้นฐาน ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร คือ $\rho = \dfrac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{N}$

ค่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง คือ $r_{XY} = \dfrac{\sum_{i=1}^{n} Z_{X_i} Z_{Y_i}}{(n-1)}$

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

ดังนั้น

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

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

นั่นคือ

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

หรือ

$$r_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}$$

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

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

S_X หมายถึง ส่วนเบี่ยงเบนมาตรฐานของตัวแปร X โดยที่

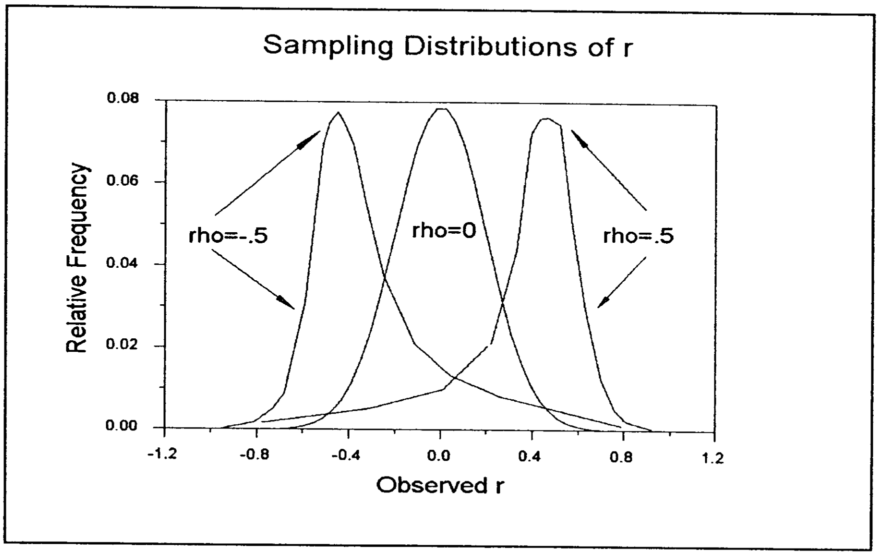
$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

S_Y หมายถึง ส่วนเบี่ยงเบนมาตรฐานของตัวแปร Y โดยที่

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

1.4 การแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์

Fisher (1915 อ้างถึงใน จารุณี ไชยมูล, 2542) ได้ศึกษาลักษณะการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ (r) พบว่า การแจกแจงของค่า r นั้นขึ้นอยู่กับค่าสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) และขนาดของกลุ่มตัวอย่าง (n) เท่านั้น โดยค่า r จะมีการแจกแจกแบบที (t- Distribution) เมื่อ $\rho = 0$ การแจกแจงของ r จะสมมาตร แต่จะมีลักษณะเบ้ ถ้า $\rho \neq 0$ โดยจะเบ้ซ้าย (Negative Skewness) ถ้า $\rho > 0$ และจะเบ้ขวา (Positive Skewness) ถ้า $\rho < 0$ (ภาพที่ 6)



ภาพที่ 6 การแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์

ต่อมา Soper และ คณะ (1916 อ้างถึงใน จารุณี ไชยมูล, 2542) 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} มีลักษณะเบ้ซ้าย เมื่อ $\rho = 0.6$ และ $\rho = 0.8$ ซึ่ง Soper ได้เสนอแผนภาพอย่างชัดเจนว่า ถ้า n มีขนาดเล็ก การแจกแจงของค่า r_{XY} จะมีการกระจายมาก แต่ถ้า n มีขนาดใหญ่ขึ้น การแจกแจงของค่า r_{XY} ก็จะแคบลง และขนาดตัวอย่างที่เพิ่มขึ้นจะทำให้ r_{XY} เข้าสู่การแจกแจงแบบสมมาตรเร็วขึ้น

2. สหสัมพันธ์ในงานวิจัยนี้

2.1 สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product-Moment Correlation Coefficient)

ในปี ค.ศ.1985 คาร์ล เพียร์สัน (Karl Pearson) นักคณิตศาสตร์ชาวอังกฤษได้พัฒนาและคิดค้นเกี่ยวกับการวัดความสัมพันธ์ขึ้น โดยในขณะนั้นเพียร์สันทำงานอยู่กับฟราสซิส กอลตัน (Francis Galton) ซึ่งเป็นบิดาของหลักการเกี่ยวกับลักษณะความแตกต่างระหว่างบุคคล โดยเพียร์สันให้เหตุผลว่า ลักษณะของบุคคลมีความหลากหลาย การที่จะวัดลักษณะเหล่านี้ให้สามารถนำไปใช้ประโยชน์ได้สูงสุด ควรขยายไปถึงการวัดเกี่ยวกับความสัมพันธ์ และเรียกค่าที่ได้จากการวัดความสัมพันธ์ระหว่างตัวแปรว่า ค่าสัมประสิทธิ์สหสัมพันธ์

สหสัมพันธ์แบบเพียร์สันค่อนข้างจะเป็นพื้นฐานของสัมประสิทธิ์สหสัมพันธ์ตัวอื่นๆ สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน คือ r หรือ r_{XY} หรือ r_p โดยมีข้อตกลงเบื้องต้นดังนี้

(1) ใช้ในสถานการณ์ที่ต้องการดัชนีชี้วัดความสัมพันธ์เชิงเส้นตรง (Linear Relationship) ระหว่างตัวแปร 2 ตัวแปร

(2) ข้อมูลของตัวแปรทั้งสองอยู่ในมาตรวัดแบบช่วงหรือแบบอัตราส่วน

(3) ข้อมูลของตัวแปรทั้งสองเป็นอิสระต่อกัน

(4) ข้อมูลของตัวแปรทั้งสองสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ

2.1.1 สูตรคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวคิดมาจากการหาความแปรร่วมระหว่างสองตัวแปร (Covariance) แต่เนื่องจากค่าความแปรปรวนร่วมที่คำนวณได้เป็นตัวเลขที่มีหน่วย ทำให้เกิดความยุ่งยากในการแปลความหมาย เพื่อให้เกิดความเข้าใจง่ายจึงปรับแก้หน่วยให้หายไป โดยนำค่าความแปรปรวนร่วมไปหารด้วยผลคูณของส่วนเบี่ยงเบนมาตรฐานของทั้งสองตัวแปร เรียกว่าสูตรคำนวณสัมประสิทธิ์สหสัมพันธ์ด้วยค่าความแปรปรวนร่วมมาตรฐาน (Correlation as Standardized Covariance) (Rodgers, Nicewander, 1988) โดยมีสูตรในการคำนวณดังนี้

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

$$= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{[\sum_{i=1}^{n}(X_i - \overline{X})^2][\sum_{i=1}^{n}(Y_i - \overline{Y})^2]}} \qquad (2.1)$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์

X_i หมายถึง ค่าคะแนนดิบแต่ละตัวของตัวแปร X

Y_i หมายถึง ค่าคะแนนดิบแต่ละตัวของตัวแปร Y

$\overline{X}$ หมายถึง ค่าเฉลี่ยของชุดตัวแปรในตัวแปร X

$\overline{Y}$ หมายถึง ค่าเฉลี่ยของชุดตัวแปรในตัวแปร Y

นอกจากสูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบสูตร (2.1) แล้ว นักสถิติได้พัฒนาสูตรเพื่อการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันเพื่อให้สอดคล้องกับสถานการณ์ต่าง ๆที่แตกต่างกัน โดย Rodgers and Nicewander (1988) ได้รวบรวมสูตรในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ในการคำนวณเมื่อลักษณะข้อมูลที่แตกต่างกัน 13 สถานการณ์ รวมทั้งสิ้น 15 สูตร

2.1.2 การทดสอบนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

การทดสอบการมีนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน มีการทดสอบสมมติฐานในสองกรณีได้แก่ กรณีที่กำหนดสมมติฐานหลักเป็น H_0: ρ = 0 และกรณีที่กำหนดสมมติฐานหลักเป็น H_0: $\rho = \rho_0$ เมื่อ ρ_0 เป็นค่าคงที่ใด ๆที่มีอยู่ระหว่าง -1ถึง 1 ดังรายละเอียดต่อไปนี้ (ทรงศิริ แต้สมบัติ, 2541)

2.1.2.1 เมื่อต้องการทดสอบว่าตัวแปร X และ Y มีสหสัมพันธ์กันหรือไม่ จะกำหนดสมมติฐานคือ H_0: $\rho = 0$ กับ H_1: $\rho \neq 0$ เมื่อต้องการทดสอบว่าตัวแปร X และ Y มีสหสัมพันธ์ตามกันหรือไม่ จะกำหนดสมมติฐาน H_0: $\rho = 0$ กับ H_1: $\rho > 0$ เมื่อต้องการทดสอบว่าตัวแปร X และ Y มีสหสัมพันธ์ทางตรงกันข้ามหรือไม่ จะกำหนดสมมติฐาน H_0: $\rho = 0$ กับ H_1: $\rho < 0$ เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน มีลักษณะการแจกแจงแบบที (t- Distribution) ที่ชั้นแห่งความเป็นอิสระ (Degree of Freedom) เท่ากับ n – 2 ดังนั้นการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ r_{XY} จึงใช้สูตรการทดสอบค่าที (กานดา พูนลาภทวี, 2539 ; Daniel, 1995) ดังนี้

$$t = \frac{r_{XY} - \rho}{\sqrt{\frac{1 - r_{XY}^2}{n-2}}}$$

$$t = r_{XY} \frac{\sqrt{n-2}}{\sqrt{1 - r_{XY}^2}} \qquad (2.2)$$

เมื่อ r_{XY} หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

ρ หมายถึง สัมประสิทธิ์สหสัมพันธ์ของประชากร

n หมายถึง ขนาดของกลุ่มตัวอย่าง

ขั้นตอนในการทดสอบสมมติฐาน

(1) กำหนดสมมติฐาน

H_0 : $\rho = 0$

H_1 : $\rho \neq 0$ หรือ $\rho > 0$ หรือ $\rho < 0$

(2) เลือกใช้สถิติทดสอบ $t = r_{XY} \frac{\sqrt{n-2}}{\sqrt{1 - r_{XY}^2}}$

(3) กำหนดระดับการมีนัยสำคัญในการทดสอบ ($\alpha = 0.05$)

(4) คำนวณค่า t แล้วนำค่า t ที่ได้ไปหาค่า P-value

(5) การสรุปผลการทดสอบ โดยผลการทดสอบมี 2 กรณีดังนี้

1. มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างมีนัยสำคัญ

2. ไม่มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้มากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ หรือ ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน

2.1.2.2 เมื่อต้องการทดสอบว่าตัวแปร X และ Y มีสหสัมพันธ์กันในระดับหนึ่งหรือไม่ หรือมีสหสัมพันธ์กันมากกว่าระดับหนึ่งหรือไม่ หรือมีสหสัมพันธ์กันน้อยกว่าระดับหนึ่งหรือไม่ จะกำหนดสมมติฐาน H_0: $\rho = \rho_0$ กับ H_1: $\rho \neq \rho_0$ หรือ H_0: $\rho = \rho_0$ กับ H_1: $\rho > \rho_0$ หรือ H_0: $\rho = \rho_0$ กับ H_1: $\rho < \rho_0$ ตามลำดับ โดยใช้ตัวทดสอบสถิติ คือ

$$Z = \frac{Z_{(r)} - Z_{(\rho)}}{SE_{Z_{(r)}}} \qquad (2.3)$$

เมื่อ $$Z_{(r)} = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \qquad (2.4)$$

$$Z_{(\rho)} = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \qquad (2.5)$$

$$SE_{Z_{(r)}} = \frac{1}{\sqrt{n-3}} \qquad (2.6)$$

ขั้นตอนในการทดสอบสมมติฐาน

(1) กำหนดสมมติฐาน

$H_0 : \rho = \rho_0$

$H_1 : \rho \neq \rho_0$

(2) เลือกใช้สถิติทดสอบ $Z = \dfrac{Z_{(r)} - Z_{(\rho)}}{SE_{Z_{(r)}}}$

(3) กำหนดระดับการมีนัยสำคัญในการทดสอบ ($\alpha = 0.05$)

(4) คำนวณค่า Z แล้วนำค่า Z ที่ได้ไปหาค่า P-value

(5) การสรุปผลการทดสอบ โดยผลการทดสอบมี 2 กรณีดังนี้

1. มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่าความสัมพันธ์ของตัวแปรทั้งสองนั้นมีค่าแตกต่างจากค่าสหสัมพันธ์ที่กำหนดอย่างมีนัยสำคัญ

2. ไม่มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้มากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าความสัมพันธ์ของตัวแปรทั้งสองนั้นมีค่าไม่แตกต่างจากค่าสหสัมพันธ์ที่กำหนด

2.1.3 การประมาณค่าช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

การแจกแจงของค่าสถิติ r_{xy} จะเบ้เมื่อ ρ มีค่าสูงๆใกล้ +1 หรือ -1 ดังนั้นในการประมาณค่า ρ จึงไม่สามารถใช้สถิติ t ในการคำนวณช่วงเชื่อมั่นได้ Fisher ได้เสนอวิธีการแปลงค่า r_{xy} ที่มีการแจกแจงเบ้ให้เป็นค่าสถิติ $Z_{(r)}$ ที่มีการแจกแจงแบบปกติ (Fisher's transformation) และใช้สถิติ $Z_{(r)}$ในการคำนวณช่วงเชื่อมั่น ในการแปลงค่าสัมประสิทธิ์สหสัมพันธ์ r ให้เป็นค่าสถิติ $Z_{(r)}$ ทำได้ ดังนี้ (Gardner and Altman, 1989 ; อรุณ จิรวัฒน์กุล, 2547)

$$Z_{(r)} = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

หรือ $$Z_{(r)} = 1.1513\log\left(\frac{1+r}{1-r}\right) \qquad (2.7)$$

ค่า $Z_{(r)}$ ที่ได้จะมีการแจกแจงแบบปกติ คำนวณค่าความคลาดเคลื่อนมาตรฐานของ $Z_{(r)}$ ได้จากสูตร

$$SE_{Z(r)} = \frac{1}{\sqrt{n-3}}$$; เมื่อ n คือ ขนาดตัวอย่าง

การประมาณค่าช่วงเชื่อมั่นคำนวณจากสูตร

$$100(1-\alpha)\%\ \text{CI ของ}\ Z_{(\rho)} = Z_{(r)} \pm Z_{\alpha/2}\frac{1}{\sqrt{n-3}} \qquad (2.8)$$

เมื่อได้ค่า CI ของ $Z_{(\rho)}$ แล้วให้ทำการแปลงค่า $Z_{(\rho)}$ กลับเป็นค่า ρ ด้วยสูตรดังนี้

$$\rho = \frac{e^{2(Z_{(\rho)})}-1}{e^{2(Z_{(\rho)})}+1} \qquad (2.9)$$

2.2 สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman Rank Correlation Coefficient)

ในปี ค.ศ. 1904 Charles Spearman ได้พัฒนาวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์โดยดัดแปลงมาจากวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน เพื่อใช้ในการหาค่าสหสัมพันธ์ของข้อมูลที่ตัวแปรทั้งสองไม่ได้มีการแจกแจงแบบปกติ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนนี้เป็นที่นิยมใช้กันอย่างแพร่หลาย เพราะสามารถคำนวณได้ง่ายและรวดเร็ว ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการคำนวณด้วยวิธีของสเปียร์แมนจะมีค่าใกล้เคียงกับค่าที่คำนวณด้วยวิธีของเพียร์สันในข้อมูลชุดเดียวกัน โดยเฉพาะเมื่อข้อมูลเป็นแบบอันดับไม่ซ้ำกันจะให้ค่าเท่ากับวิธีของเพียร์สัน สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน คือ r_s ในการคำนวณสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน มีข้อตกลงในเบื้องต้นดังนี้

(1) ข้อมูลของตัวแปรทั้งสองอยู่ในมาตรวัดแบบอันดับ หรือสามารถแปลงให้อยู่ในมาตรวัดแบบอันดับได้

(2) ข้อมูลของตัวแปรทั้งสองต้องเป็นอิสระต่อกัน

(3) ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นเส้นตรง

ในการแปลความหมายของค่า r_s นั้น จะแปลความหมายเช่นเดียวกับ r_{XY} และค่าที่คำนวณได้จะมีค่าอยู่ระหว่าง - 1 ถึง +1

2.2.1 สูตรคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน

ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน สามารถคำนวณได้จากข้อมูลในแบบต่างๆ 3 กรณี (พนิตา แก้วกูร, 2539) ดังนี้

2.2.1.1 กรณีข้อมูลไม่มีอันดับซ้ำ (Kendall and Gibbons, 1990; Siegel and Castellan, 1988; Gibbons, 1971)

$$r_{sa} = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad (2.10)$$

เมื่อ d หมายถึง ผลต่างของอันดับในแต่ละคู่

n หมายถึง ขนาดของกลุ่มตัวอย่าง

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}} \tag{2.11}$$

เมื่อ $x = X_i - \overline{X}$

$y = Y_i - \overline{Y}$

2.2.1.2 กรณีข้อมูลมีอันดับซ้ำ (Kendall and Gibbons, 1990)

กรณีที่มีอันดับซ้ำกันไม่ว่าจะเป็นซ้ำเฉพาะข้อมูลในตัวแปร X หรือซ้ำเฉพาะในข้อมูลตัวแปร Y หรือซ้ำทั้งข้อมูลในตัวแปร X และตัวแปร Y โดยมีค่าที่ซ้ำมากกว่า 2 ตำแหน่ง การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนต้องมีการแก้ไขความคลาดเคลื่อนที่เกิดขึ้น โดยใช้สูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์ใหม่ดังนี้

จากสูตร (2.11) $r_s = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}}$

เมื่อ $d = x_i - y_i \quad ; i=1,2,......,n$

$\sum x^2 = \frac{1}{12}(n^3 - n) - T_X$

$\sum y^2 = \frac{1}{12}(n^3 - n) - T_Y$

T_X = T_i ของตัวแปร X

T_Y = T_i ของตัวแปร Y

t_i = จำนวนซ้ำกันของข้อมูลที่มีค่า i

$T_i = \frac{1}{12}\sum (t_i^3 - t_i)$ เป็นค่าปรับแก้เมื่อตัวแปรมีอันดับซ้ำกัน

เมื่อแทนค่าในสูตร (2.11) จะได้

$$r_{sb} = \frac{\frac{1}{6}(n^3 - n) - \sum d^2 - T_X - T_Y}{\sqrt{\frac{1}{6}(n^3 - n) - 2T_X}\sqrt{\frac{1}{6}(n^3 - n) - 2T_Y}} \tag{2.12}$$

2.2.1.3 กรณีข้อมูลมีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร

นอกจากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนใน 2 กรณีดังกล่าวแล้ว ในปี ค.ศ. 1960 Alan Stuart ได้พัฒนาวิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน โดยนำมาประยุกต์ ใช้กับข้อมูลที่มีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร r x c เมื่อ r หมายถึง แถว และ c หมายถึง สดมภ์ ซึ่งการหาค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวใช้เมื่อ การจัดอันดับของข้อมูลในตารางการณ์จรเกิดการซ้ำของอันดับ และการซ้ำของอันดับดังกล่าวจะเกิดการซ้ำเป็นจำนวนมาก การหาค่าสัมประสิทธิ์สหสัมพันธ์ในสูตร (2.12) จึงไม่เหมาะที่จะนำมาใช้ในลักษณะนี้ สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์

วิธีนี้ คือ r_{cs} การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ตามวิธีของ Stuart กรณีที่ข้อมูลจัดอันดับในรูปตารางการณ์จร จะไม่กล่าวรายละเอียดในที่นี้

2.2.2 การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (r_s)

ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้นั้น เป็นค่าที่คำนวณได้จากข้อมูลของกลุ่มตัวอย่าง ดังนั้นการที่จะบอกว่าตัวแปรทั้งสองมีค่าสัมประสิทธิ์สหสัมพันธ์ตามที่คำนวณได้จะต้องแสดงให้ได้ว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรไม่เท่ากับ 0 นั่นคือต้องมีการทดสอบนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์ r_s ก่อน เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (r_s) มีลักษณะการแจกแจงแบบที (t-Distribution) ที่ชั้นแห่งความเป็นอิสระ เท่ากับ n − 2 ดังนั้นการทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์จึงใช้สูตรการทดสอบที (t − t est) สำหรับขนาดตัวอย่างตั้งแต่ 4 ถึง 30 (Daniel, 1995; Kendall and Gibbons, 1990; Zar, 1999) ใช้สูตรดังนี้

$$t = \frac{r_S\sqrt{n-2}}{\sqrt{1-r_S^2}} \quad ; df = n-2 \qquad (2.13)$$

เมื่อ r_s หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน
n หมายถึง ขนาดของกลุ่มตัวอย่าง
df หมายถึง ชั้นแห่งความเป็นอิสระ

ขั้นตอนในการทดสอบสมมติฐาน

(1) กำหนดสมมติฐาน

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$ หรือ $\rho > 0$ หรือ $\rho < 0$

(2) เลือกใช้สถิติทดสอบ $t = r_{XY}\frac{\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$

(3) กำหนดระดับการมีนัยสำคัญในการทดสอบ ($\alpha = 0.05$)

(4) คำนวณค่า t แล้วนำค่า t ที่ได้ไปหาค่า P-value

(5) การสรุปผลการทดสอบ โดยผลการทดสอบมี 2 กรณีดังนี้

1. มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างมีนัยสำคัญ

2. ไม่มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้มากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ หรือ ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน

การทดสอบการมีนัยสำคัญทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ในกรณีที่ตัวอย่างมากกว่า 30 ค่า r_S มีลักษณะการแจกแจงประมาณได้กับการแจกแจงปกติมาตรฐาน จึงทดสอบการมีนัยสำคัญทางสถิติโดยใช้สถิติ Z (นิคม ถนอมเสียง, 2540) โดยใช้สูตร

$$Z = r_s\sqrt{n-1} \qquad (2.14)$$

นำค่า Z ที่คำนวณได้ไปคำนวณค่า P-value เพื่อตัดสินว่ามีนัยสำคัญทางสถิติหรือไม่ โดยใช้วิธีการตัดสินในการสรุปผลการทดสอบเช่นเดียวกับการใช้สถิติทดสอบแบบที กรณีขนาดตัวอย่างน้อยกว่า 30

2.2.3 การประมาณค่าช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน

เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (r_s) มีลักษณะการแจกแจงแบบที (t-Distribution) ดังนั้นในการคำนวณช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน จะใช้วิธีการคำนวณแบบเดียวกับการคำนวณช่วงเชื่อมั่นของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Altman, 1991) ซึ่งรายละเอียดได้นำเสนอไว้ในหัวข้อ 2.1.3 หน้า 13

2.3 สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ (Kendall Rank Correlation Coefficient)

การหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เป็นวิธีการหาความสัมพันธ์ร่วมของการจัดอันดับข้อมูล 2 เซต โดยที่ข้อมูลหรือตัวแปรทั้งสองอยู่ในมาตรวัดขั้นต่ำแบบเรียงอันดับ และการหาค่าสัมประสิทธิ์สหสัมพันธ์วิธีนี้มีข้อตกลงเบื้องต้นเหมือนกับวิธีการหาสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน แต่แตกต่างกันตรงที่ขั้นแรกต้องเรียงลำดับในตัวแปรแรก(ตัวแปร X)เสียก่อน ส่วนลำดับที่ในตัวแปร Y จะแปรผันตามอันดับที่ของตัวแปร X ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์จะให้ค่าประมาณที่ไม่เอนเอียง (Un-bias Estimator) กับค่าพารามิเตอร์ของประชากร ในขณะที่ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนจะไม่นำไปใช้ในการประมาณค่าพารามิเตอร์ของประชากร (Siegel and Castellan, 1988) สัญลักษณ์ที่ใช้แทนสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ คือ τ

2.3.1 สูตรที่ใช้ในการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ สามารถกำหนดใช้ได้ตามเงื่อนไขใน 3 กรณีต่อไปนี้

2.3.1.1 ข้อมูลไม่มีอันดับซ้ำกัน (Kendall and Gibbons, 1990)

การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เมื่อข้อมูลไม่มีอันดับซ้ำกัน โดยกำหนดสัญลักษณ์คือτ_a มีสูตรในการคำนวณดังนี้ (นิภา ศรีไพโรจน์, 2538)

$$\tau_a = \frac{2S}{n(n-1)} \tag{2.15}$$

เมื่อ τ_a หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลล์กรณีข้อมูลไม่มีอันดับซ้ำกัน

S หมายถึง ผลรวมของผลต่างระหว่างความสอดคล้องและไม่สอดคล้องของการจัดอันดับ คำนวณจาก $S = \sum P - \sum Q$

P หมายถึง จำนวนของอันดับที่มีค่าสูงกว่าเมื่อนับจำนวนที่อยู่ใต้ลงมา

Q หมายถึง จำนวนของอันดับที่มีค่าต่ำกว่าเมื่อนับจำนวนที่อยู่ใต้ลงมา

n หมายถึง ขนาดของกลุ่มตัวอย่าง

หรืออาจใช้สูตร

$$\tau_a = 1 - \frac{4M}{n(n-1)} \tag{2.16}$$

เมื่อ M หมายถึง จำนวนสูงสุดของการสลับอันดับ
n หมายถึง ขนาดของกลุ่มตัวอย่าง

2.3.1.2 กรณีข้อมูลมีอันดับซ้ำกัน (Kendall and Gibbons, 1990)

กรณีข้อมูลมีอันดับซ้ำกันในชุดใดชุดหนึ่งหรือทั้งสองชุด การกำหนดอันดับที่ซ้ำกันโดยการเฉลี่ย จะมีผลทำให้ค่าผลรวมสูงสุดของจำนวนคู่ทั้งหมดของตัวแปร มีค่าน้อยกว่า $n(n-1)/2$ ทั้งนี้เพราะคู่ที่มีอันดับเท่ากันจะมีค่าเป็น 0 จึงต้องปรับค่าผลรวมสูงสุดของจำนวนคู่ทั้งหมดของตัวแปรใหม่ ดังนั้นสูตรที่ใช้สำหรับคำนวณค่าสัมประสิทธิ์ (τ_b) ในกรณีนี้ คือ

$$\tau_b = \frac{S}{\sqrt{\frac{1}{2}n(n-1)-U}\sqrt{\frac{1}{2}n(n-1)-V}} \tag{2.17}$$

เมื่อ τ_b หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลล์กรณีข้อมูลมีอันดับซ้ำกัน
U หมายถึง ค่าแก้เมื่อตัวแปร X มีค่าซ้ำกัน

โดยที่ $U = \frac{1}{2}\sum u(u-1)$

V หมายถึง ค่าแก้เมื่อตัวแปร Y มีค่าซ้ำกัน

โดยที่ $V = \frac{1}{2}\sum v(v-1)$

u หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร X
v หมายถึง จำนวนอันดับที่ซ้ำกันในกลุ่มของตัวแปร Y

2.3.1.2 ข้อมูลมีการแยกประเภทและจัดอันดับในรูปตารางการณ์จร

นอกจากการหาค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลล์ ใน 2 กรณีดังกล่าวแล้ว ในปี ค.ศ. 1953 เอเลน (Alan Stuart) ได้พัฒนาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอลล์ มาประยุกต์ใช้กับข้อมูลที่มีการแยกประเภท แบบจัดอันดับในรูปตารางการณ์จร r x c สามารถใช้ได้กับทุก ๆขนาดของ r และ c และสามารถทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าสูงสุดถึง $\pm$ 1 ซึ่งค่าของ τ จะให้ค่าสูงสุดได้ก็ต่อเมื่อ r = c เท่านั้น (Brown, Benedetti, 1977) การคำนวณหาค่าสัมประสิทธิ์ τ ใช้กับข้อมูลที่มีการแยกประเภท แบบจัดอันดับในรูปตารางการณ์จร r x c จะไม่กล่าวรายละเอียดในที่นี้

2.3.2 การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์

คะแนนลำดับที่ของตัวแปร X และ Y จะมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติเมื่อลำดับที่ของตัวแปร X และ Y มีสหสัมพันธ์กันสูงมากพอสมควร ดังนั้นเมื่อจะหาค่าความน่าจะเป็นที่จะทดสอบ H_0 ค่าความน่าจะเป็นที่ได้ต้องน้อยกว่า α เมื่อ α คือระดับนัยสำคัญ หรือถ้าคะแนนลำดับที่ของตัวแปร X และ Y มีความสัมพันธ์กัน นั่นหมายถึงว่า การจัดคะแนนเซต X_i เมื่อ i = 1,2,3,...,n เรียงตามลำดับที่แล้วจะทำให้ลำดับที่ของเซต Y_i เปลี่ยนแปลงตามลำดับที่ของเซต X_i ถ้ากำหนดว่า n คือจำนวนตัวอย่าง ในการทดสอบ

นัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบอันดับที่ของเคนดอลล์ มีแนวคิดในการทดสอบ 2 แบบโดยอิงตามจำนวนตัวอย่าง (นิภา ศรีไพโรจน์, 2538) ดังนี้

2.3.2.1 กรณีตัวอย่างมีขนาดเล็ก ($n \leq 10$)

เมื่อ $n \leq 10$ จะใช้วิธีการเปิดตารางสถิติ เพื่อดูค่านัยสำคัญของ τ ถ้าความน่าจะเป็นในตาราง Kendall's τ มีค่าน้อยกว่าค่าความน่าจะเป็น ณ ระดับนัยสำคัญที่ตั้งไว้ แสดงว่ามีนัยสำคัญทางสถิติ

2.3.2.2 กรณีตัวอย่างมีขนาดใหญ่ ($n > 10$)

ถ้าขนาดตัวอย่างมีค่ามากกว่า 10 การแจกแจงของค่า τ จะใกล้เคียงการแจกแจงแบบปกติ (Normal Distribution) โดยมี

(1) ค่าเฉลี่ยเท่ากับ 0 หรือ $E(\tau) = 0$

(2) ส่วนเบี่ยงเบนมาตรฐาน $S(\tau) = \sqrt{\dfrac{2(2n+5)}{9n(n-1)}}$

ดังนั้นในการทดสอบนัยสำคัญจึงใช้สถิติในการทดสอบ ซี (Z-test) (นิภา ศรีไพโรจน์, 2538) คือ

$$Z = \frac{\tau - E(\tau)}{S(\tau)}$$

$$= \frac{\tau\sqrt{9n(n-1)}}{\sqrt{2(2n+5)}} \qquad (2.18)$$

เมื่อ τ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับที่ของเคนดอลล์

$E(\tau)$ หมายถึง ค่าเฉลี่ยของτ

$S(\tau)$ หมายถึง ส่วนเบี่ยงเบนมาตรฐานของ τ

n หมายถึง จำนวนตัวอย่าง

ขั้นตอนในการทดสอบสมมติฐาน

(1) กำหนดสมมติฐาน

$H_0 : \rho = 0$

$H_1 : \rho \neq 0$ หรือ $\rho > 0$ หรือ $\rho < 0$

(2) เลือกใช้สถิติทดสอบ $Z = \dfrac{\tau\sqrt{9n(n-1)}}{\sqrt{2(2n+5)}}$

(3) กำหนดระดับการมีนัยสำคัญในการทดสอบ ($\alpha = 0.05$)

(4) คำนวณค่า Z แล้วนำค่า Z ที่ได้ไปหาค่า P-value

(5) การสรุปผลการทดสอบ โดยผลการทดสอบมี 2 กรณีดังนี้

1. มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างมีนัยสำคัญ

2. ไม่มีนัยสำคัญ ถ้าค่า P-value ที่คำนวณได้มากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าตัวแปรทั้งสองนั้นมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ หรือ ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน

2.3.3 การประมาณค่าช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์

การประมาณช่วงเชื่อมั่นของ τ Noether (1967 อ้างถึงใน นิคม ถนอมเสียง, 2540) ได้เสนอวิธีการและขั้นตอนในการคำนวณดังนี้

(1) คำนวณค่าความคลาดเคลื่อนมาตรฐาน (Standard Error) โดยพิจารณาจากคู่ของค่าสังเกต (X_i,Y_i) กับคู่ของค่าสังเกตอื่นๆ (X_j,Y_j) ว่ามีลักษณะการเรียงตัวที่เหมือนกัน (Concordant)หรือไม่เหมือนกัน (Discordant)

(2) ลักษณะการเรียงตัวที่เหมือนกัน เป็นลักษณะที่ความสัมพันธ์ระหว่างค่า X's และ Y's เป็นดังนี้

$X_i > X_j$ และ $Y_i > Y_j$ หรือ $X_i < X_j$ และ $Y_i < Y_j$

(3) เมื่อลักษณะการเรียงตัวผกผันจากลักษณะที่กำหนดแสดงว่ามีลักษณะการเรียงตัวที่ไม่เหมือนกัน เช่น ต้องการเปรียบเทียบคู่ (35,25) และ (30,19) พบว่ามีลักษณะการเรียงตัวที่เหมือนกัน เนื่องจากค่า 35 > 30 และ 25>19 ส่วนตัวอย่างลักษณะการเรียงตัวที่ไม่เหมือนกันเช่น (35,10) และ (25,19) พบว่า 35 > 25 และ 10 <19

(4) ให้ C_i คือจำนวนคู่ (X_i,Y_i) มีลักษณะการเรียงตัวที่เหมือนกันกับคู่ของค่าสังเกต (X_j,Y_j) เมื่อ i และ j มีค่าตั้งแต่ 1 ถึง n เมื่อไม่มีค่าซ้ำประมาณค่าความแปรปรวนได้จาก

$$\hat{\sigma}^2 = 4\sum C_i^2 - 2\sum C_i - \frac{2(2n-3)(\sum C_i)^2}{n(n-1)} \qquad (2.19)$$

(5) ประมาณค่า 100(1-α) ช่วงความเชื่อมั่นของพารามิเตอร์ τ เท่ากับ

$$\tau \pm \frac{2}{n(n-1)}\hat{\sigma}Z \qquad (2.20)$$

เมื่อ Z หมายถึง ค่า upper α/2th เปอร์เซนต์ไทล์ของการแจกแจงปกติมาตรฐาน

2.3.4 ข้อสังเกตในการคำนวณสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์

อำนวย มณีศรีวงศ์กูล (2538) ได้ให้ข้อสังเกตในการคำนวณสัมประสิทธิ์สหสัมพันธ์อันดับที่แบบเคนดอลล์ไว้ดังนี้

2.3.3.1 ในกรณีที่มีการจัดอันดับข้อมูลสลับกันอย่างสิ้นเชิง เช่น ตัวแปร X มีอันดับ 1,2,3,4,5 ตัวแปร Y มีอันดับ 5,4,3,2,1 หากใช้สูตร (2.15) ค่า τ จะเป็น -1 หากใช้สูตร (2.16) ค่าτ จะเป็น 0 แต่ถ้าตัวแปร X และ Y มีอันดับเหมือนกัน ค่า τ จากสูตร (2.15) และ (2.16) จะเป็น +1 นั่นหมายถึงว่าสูตร (2.15) จะให้ค่า τ อยู่ระหว่าง -1 กับ +1 แต่สูตร (2.16)ให้ค่า τ อยู่ระหว่าง 0 กับ 1

2.3.3.2 แม้ว่าสูตร (2.16) จะให้ค่า τ ไม่ติดลบ ซึ่งเป็นข้อด้อย แต่มีข้อเด่นตรงที่ว่าการตีความค่า τ ด้วยสูตรนี้อยู่ภายใต้พื้นฐานทฤษฎีความน่าจะเป็น คือ การแสดงโอกาสที่จะจัดอันดับซ้ำของข้อมูลชุดเดียวกันอย่างเป็นลำดับ เช่นค่า τ เท่ากับ 0.87 แสดงให้เห็นว่ามีโอกาสถึง 87% ที่จะจัดอันดับเหมือนกันนั่นหมายถึงว่าถ้าค่า τ สูง โอกาสที่จะจัดอันดับเหมือนกันของสิ่งใดๆก็สูงด้วย (Lehman, 1991)

2.3.3.3 กรณีที่มีข้อมูลซ้ำอาจจะในตัวแปรเดียวหรือทั้งสองตัวแปร ควรคำนวณหา τ โดยใช้คอมพิวเตอร์ซึ่งมีโปรแกรมทางสถิติที่สามารถคำนวณค่า τ อยู่หลากหลายโปรแกรมจะสะดวกกว่า (Lehman, 1991)

สรุป

สัมประสิทธิ์สหสัมพันธ์อันดับที่แบบสเปียร์แมนและแบบเคนดอลล์ มีการนำมาใช้ในการหาความสัมพันธ์ระหว่างตัวแปรอย่างแพร่หลาย ความยากง่ายในการคำนวณขึ้นอยู่กับความเห็นของบุคคล ปัจจุบันมีการพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการคำนวณทั้งในรูปโปรแกรมสำเร็จรูปและโปรแกรมในระบบอินเตอร์เน็ต ทำให้คำนวณได้อย่างรวดเร็วและถูกต้องจึงไม่อาจสรุปได้ว่าแบบใดสะดวกหรือง่ายกว่ากันในลักษณะต่าง ๆเกี่ยวกับ r_s กับ τ ในที่นี้จะเปรียบเทียบประเด็นที่เห็นชัดเจนประกอบการตัดสินใจในการนำไปใช้

(1) เมื่อคำนวณจากข้อมูลชุดเดียวกันแล้ว โดยทั่วไป r_s จะให้ค่าสูงกว่า τ อย่างไรก็ตามเมื่อ X และ Y มีความสอดคล้องกันอย่างสมบูรณ์ หรือไม่สอดคล้องกันอย่างสมบูรณ์ จะให้ค่าเหมือนกันคือ 1 กับ -1 ตามลำดับ และในการทดสอบสมมติฐานโดยทั่วไปแล้วจะให้ผลสรุปเหมือนกัน

(2) ค่าของ r_s กับ τ ไม่สามารถเปรียบเทียบกันได้ เนื่องจากความเป็นมาของสูตรในการคำนวณไม่เหมือนกัน

(3) τ เป็นค่าประมาณที่ไม่เอนเอียงของสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) ในขณะที่ r_s ไม่ได้เป็นตัวประมาณค่าของสัมประสิทธิ์สหสัมพันธ์ของประชากร

(4) การแจกแจงของ τ เข้าสู่การแจกแจงแบบปกติเร็วกว่า r_s ดังนั้นเมื่อ n มีขนาดใหญ่ ค่า P-value จึงเชื่อถือได้มากกว่าการใช้ r_s

(5) เมื่อตัวอย่างถูกสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ r_s จะมีค่าใกล้เคียงกับ r_{XY}

2.4 สหสัมพันธ์แบบถ่วงน้ำหนัก (Biweight Midcorrelation; r_b)

Wilcox (1997) ได้เสนอวิธีการคำนวณสหสัมพันธ์แบบถ่วงน้ำหนักเพื่อหาค่าสหสัมพันธ์ระหว่างสองตัวแปรที่เป็นข้อมูลแบบต่อเนื่องและมีความสัมพันธ์เชิงเส้นตรง ภายใต้หลักการถ่วงน้ำหนักแบบซ้ำหลายรอบ (The Iterative Reweighed) ซึ่งเรียกว่า Biweigth Estimate (Bw) ตามแนวคิดของ Mosteller และ Tukey (1977) เพื่อลดอิทธิพลของค่าผิดปกติจากกลุ่มที่มีต่อสัมประสิทธิ์สหสัมพันธ์ สหสัมพันธ์แบบถ่วงน้ำหนักเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์อีกทางเลือกหนึ่ง ที่มีคุณสมบัติทั้งด้านความต้านทาน และประสิทธิภาพด้านความแกร่งตามเกณฑ์ของ Mosteller และ Tukey (1977) โดยมีข้อตกลงเบื้องต้นในการใช้เหมือนสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

2.4.1 สูตรคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก

ค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก มีวิธีการและขั้นตอนในคำนวณดังสูตรต่อไปนี้ (Herrington, 2001; Wilcox, 1997)

กำหนดให้ $(X_1,Y_1),...,(X_n,Y_n)$ เป็นตัวแปรเชิงสุ่มมาจากการแจกแจงร่วมแบบสองตัวแปร (Bivariate Distribution)

$$U_i = \frac{X_i - M_X}{9MAD_X} \qquad (2.21)$$

$$V_i = \frac{Y_i - M_Y}{9MAD_Y} \quad (2.22)$$

เมื่อ U_i หมายถึง X's Variable Score

V_i หมายถึง Y's Variable Score

M_X หมายถึง ค่ามัธยฐานของตัวแปร X จากตัวอย่าง

M_Y หมายถึง ค่ามัธยฐานของตัวแปร Y จากตัวอย่าง

MAD_X หมายถึง ค่ามัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบนของตัวแปร X

MAD_Y หมายถึง ค่ามัธยฐานของค่าสัมบูรณ์ของความเบี่ยงเบนของตัวแปร Y

โดย MAD_X ของตัวอย่างคำนวณจากสูตร

$$MAD_X = \text{MEDIAN}\left\{\left|X_1 - M_X\right|,\ldots,\left|X_i - M_X\right|\right\} \quad (2.23)$$

ความแปรปรวนร่วมแบบถ่วงน้ำหนักของตัวอย่าง (The Sample Biweight Midcovariance: S_{bxy}) ระหว่างตัวแปร Xและตัวแปร Y คำนวณได้จากสูตร (Wilcox, 1997)

$$S_{bxy} = \frac{n\sum a_i(X_i - M_X)(1-U_i^2)^2 b_i(Y_i - M_y)(1-V_i^2)^2}{(\sum a_i(1-U_i^2)(1-5U_i^2))(\sum b_i(1-V_i^2)(1-5V_i^2))} \quad (2.24)$$

เมื่อ

$a_i = 1$ if $-1 \leq U_i \leq 1$, Otherwise $a_i = 0$

$b_i = 1$ if $-1 \leq V_i \leq 1$, Otherwise $b_i = 0$

$(1-U_i^2)^2$ หมายถึง น้ำหนักถ่วงในตัวแปร x

$(1-V_i^2)^2$ หมายถึง น้ำหนักถ่วงในตัวแปร y

การประมาณค่าสหสัมพันธ์แบบถ่วงน้ำหนักระหว่างตัวแปร X และตัวแปร Y คำนวณจากสูตร (Wilcox, 1997)

$$r_b = \frac{S_{bxy}}{\sqrt{S_{bxx}S_{byy}}} \quad (2.25)$$

เมื่อ S_{bxx} และ S_{byy} คือความแปรปรวนของตัวแปร X และตัวแปร Y หลังจากถ่วงน้ำหนักแล้ว (Biweight Midvariances of X and Y) ตามลำดับ

2.4.2 การทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก

การทดสอบการมีนัยสำคัญทางสถิติของสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก เมื่อต้องการทดสอบว่าตัวแปร X และ Y มีสหสัมพันธ์กันหรือไม่ จะกำหนดสมมติฐานคือ H_0: $\rho = 0$ กับ H_1: $\rho \neq 0$ โดยใช้การคำนวณเช่นเดียวกับสูตรการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน คือ ใช้สถิติทดสอบ t ดังสูตร

$$t = r_b \frac{\sqrt{n-2}}{\sqrt{1-r_b^2}} \qquad (2.26)$$

เมื่อ r_b หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก

n หมายถึง ขนาดของกลุ่มตัวอย่าง

2.2.3 การประมาณค่าช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก

เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก มีแนวคิดพื้นฐานในการคำนวณจากสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โดย r_b มีลักษณะการแจกแจงแบบที (t-Distribution) ดังนั้นในการคำนวณช่วงเชื่อมั่นของค่าสัมประสิทธิ์สหสัมพันธ์ถ่วงน้ำหนัก จึงใช้วิธีการคำนวณแบบเดียวกับการคำนวณช่วงเชื่อมั่นของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Wilcox, 1997) ซึ่งรายละเอียดได้นำเสนอไว้ในหัวข้อ 2.1.3 หน้า 13

3. ค่าผิดปกติจากกลุ่ม (Outlier)

3.1 ความหมายและลักษณะค่าผิดปกติจากกลุ่ม

ค่าผิดปกติจากกลุ่ม หรือค่าผิดปกติ หรือ ค่านอกกลุ่ม มีผู้ที่ให้ความหมายและคำจำกัดความไว้ดังนี้

บุญชะนะ วาราชะนนท์ (2544) กล่าวว่า ค่าสังเกตที่มีค่าแตกต่างจากค่าสังเกตอื่นๆของชุดข้อมูลอย่างมากเรียกว่า ค่านอกกลุ่ม (Outlier) ค่านอกกลุ่มอาจเป็นค่าสังเกตที่ผิดปกติของตัวแปรอิสระ (X) หรือตัวแปรตาม (Y) ก็ได้ โดยเฉพาะอย่างยิ่งค่าปลายสุด (Extremely Value) ของตัวแปรอิสระ และค่าสังเกตที่ทำให้ค่าพยากรณ์เปลี่ยนแปลงอย่างมากเมื่อตัดค่าสังเกตตัวนั้นออกจากชุดข้อมูล เรียกว่า ค่าสังเกตที่มีอิทธิพล (Influential Observation) แต่อย่างไรก็ตามค่านอกกลุ่มอาจเป็นหรือไม่เป็นค่าสังเกตที่มีอิทธิพลก็ได้

อรรมย์คณา แย้มนวล (2545) ให้ความหมายค่าผิดปกติไว้ว่า ค่าผิดปกติหรือค่าที่ห่างผิดปกติจากกลุ่ม หมายถึงข้อมูลที่มีค่ามากอย่างมาก (Extremely Large Value) หรือมีค่าน้อยอย่างมาก (Extremely Small Value) เมื่อเปรียบเทียบกับข้อมูลตัวอื่นๆที่เก็บรวบรวมมา หรือเป็นข้อมูลที่มีค่าแยกออกจากกลุ่มหรือผิดแผกแตกต่างไปจากข้อมูลค่าอื่น ๆ อาจจะสูงกว่าหรือต่ำกว่าข้อมูลในชุดเดียวกัน

ประสพชัย พสุนนท์ (2545) กล่าวว่า บ่อยครั้งที่พบว่าข้อมูลที่ได้จากการเก็บรวบรวมหรือจากการทดลองมีค่าของข้อมูลบางค่าแตกต่างไปจากข้อมูลส่วนใหญ่ กล่าวคือ อาจมีบางค่าที่มีค่าสูงเกินไปหรือต่ำเกินไปเมื่อเทียบกับข้อมูลตัวอื่นๆ เราเรียกค่าเหล่านี้ว่า ค่าผิดปกติ (Outlier)

กุลธารี ภัทรชญานันท์ (2546) กล่าวว่า การวิจัยสาขาต่างๆไม่ว่าจะเป็นด้านการแพทย์ วิทยาศาสตร์ สังคมศาสตร์ หรือบริหารธุรกิจ บางครั้งข้อมูลที่รวบรวมได้มีค่าสังเกตบางค่าสูงหรือต่ำมาก หรือเป็นค่าสังเกตที่ไม่ได้มาจากประชากรเดียวกับค่าสังเกตส่วนใหญ่ ซึ่งพบมากในข้อมูลที่วางแผนการทดลองผิดพลาด ค่าสังเกตที่มีลักษณะดังกล่าวจะเรียกว่าเป็น “ค่าผิดปกติ” (Outliers)

ฉัตรศิริ ปิยะพิมลสิทธิ์ (2547) กล่าวว่า ค่าผิดปกติ (Outlier) เป็นข้อมูลที่มีค่าแยกออกจากกลุ่มหรือผิดแผกแตกต่างไปจากข้อมูลค่าอื่นๆ ตัวอย่างของค่าผิดปกติเช่น ไอคิวของเด็กวัดได้ 195 น้ำหนักของคน 220 กิโลกรัม ความสูงของคน 210 เซนติเมตร ซึ่งค่าผิดปกติมีโอกาสเกิดขึ้นได้บนพื้นฐานของเหตุผล 2 ประการ คือ การจดบันทึกหรือเก็บรวบรวมข้อมูลมีความคลาดเคลื่อน และกลุ่มตัวอย่างที่เก็บรวบรวมมามีความแตกต่างไปจากกลุ่มจริง

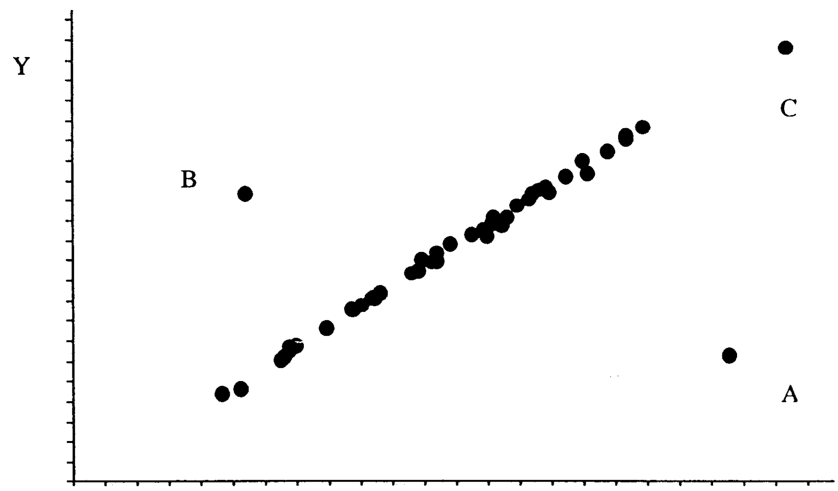
จากความหมายและคำนิยามข้างต้น สรุปได้ว่า ค่าผิดปกติจากกลุ่ม คือค่าสังเกตที่มีค่าแตกต่างจากค่าสังเกตอื่น ๆของชุดข้อมูลอย่างมาก โดยมีค่าสังเกตบางค่าสูงหรือต่ำมาก หรือเป็นค่าสังเกตที่ไม่ได้มาจากประชากรเดียวกับค่าสังเกตส่วนใหญ่ โดยข้อมูลที่มีค่ามากกว่า Q_3 + 1.5(IQR) เป็นค่าที่สูงผิดปกติจากกลุ่ม และข้อมูลที่มีค่าน้อยกว่า Q_1 -1.5(IQR) เป็นค่าที่ต่ำผิดปกติไปจากกลุ่ม

Rousseeuw และ Leroy (1987) ได้แบ่งลักษณะการเกิดค่าผิดปกติในการศึกษาความสัมพันธ์ของตัวแปรสองตัว อาจเกิดค่าผิดปกติจากกลุ่มจากตัวแปรใดตัวแปรหนึ่งหรือทั้งสองตัวแปร เมื่อกำหนดตัวแปรในการศึกษาความสัมพันธ์ของ 2 ตัวแปร คือ X และ Y ค่าผิดปกติจากกลุ่มแบ่งได้ 3 ลักษณะ ดังนี้

(1) X-Outlier เป็นความผิดปกติเนื่องจากตัวแปร X ค่าของตัวแปร X ที่ผิดปกติจะต่างจากค่าตัวแปร Xอื่น ๆอย่างมาก แต่ค่าของตัวแปร Y จะอยู่ในช่วงเดียวกับตัวแปร Y ตัวอื่น ๆ (ดังจุด A ในภาพที่ 7)

(2) Y-Outlier เป็นค่าผิดปกติจากกลุ่มเนื่องจากตัวแปร Y ค่าของตัวแปร Y ที่ผิดปกติจะต่างจากค่าตัวแปร Y อื่น ๆอย่างมาก แต่ค่าของตัวแปร Xจะอยู่ในช่วงเดียวกับตัวแปร Xตัวอื่น ๆ (ดังจุด B ในภาพที่ 7)

(3) X and Y - Outliers เป็นค่าผิดปกติจากกลุ่มเนื่องจากตัวแปร x และy ค่าทั้งสองจะต่างอย่างมากจากข้อมูลตัวอื่น ๆ (ดังจุด C ในภาพที่ 7)



ภาพที่ 7 แผนภาพกระจายกรณีมีค่าผิดปกติจากกลุ่ม

3.2 การแบ่งระดับค่าผิดปกติจากกลุ่ม

Barnett และ Lewis (1995) และ ศิริชัย พงษ์วิชัย (2544) ได้แบ่งระดับค่าผิดปกติจากกลุ่มเป็น 2 ระดับ คือ

3.2.1 ค่าผิดปกติจากกลุ่มระดับปานกลาง (Mild Outliers) คือค่าของตัวแปรที่อยู่ในช่วง (Q_1 - 3(IQR), Q_1 -1.5(IQR)) หรือ (Q_3 + 1.5(IQR), Q_3 + 3(IQR)) เมื่อ Q_1 คือ ค่าควอไทล์ที่ 1, Q_3 คือ ค่าควอไทล์ที่ 3 และ IQR (The Interquartile Range) คือระยะระหว่างควอไทล์ซึ่งมีค่าเท่ากับ Q_3- Q_1

3.2.2 ค่าผิดปกติจากกลุ่มระดับรุนแรง (Extreme Outliers) คือค่าของตัวแปรที่อยู่ในช่วง $(-\infty, Q_1 - 3(IQR))$ หรือ $(Q_3 + 3(IQR), \infty)$

ค่าผิดปกติจากกลุ่มในแต่ละระดับจะมีอิทธิพลต่อการประมาณค่าพารามิเตอร์ที่แตกต่างกัน โดยค่าผิดปกติจากกลุ่มระดับรุนแรงที่มีอิทธิพลจะทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อนสูง แต่มีโอกาสพบได้น้อย โดยทั่วไปข้อมูลตัวอย่างจะพบค่าผิดปกติจากกลุ่มระดับปานกลาง (Iglewicz and Hoaglin, 1993)

3.3 สาเหตุการเกิดค่าผิดปกติจากกลุ่ม

ประสพชัย พสุนนท์ (2545) กล่าวถึงสาเหตุของการเกิดค่าผิดปกติจากกลุ่มไว้ว่า ค่าผิดปกติจากกลุ่มที่เกิดขึ้นสามารถพิจารณาได้ 2 ลักษณะ คือ

3.3.1 ค่าผิดปกติจากกลุ่มที่เกิดขึ้นเป็นค่าที่มีความสนใจในตัวเอง เช่น เป็นค่าที่แทนปริมาณแร่ธาตุที่มีนัยสำคัญต่อการพัฒนาผลิตผลการเกษตร

3.3.2 ค่าผิดปกติจากกลุ่มในกลุ่มตัวอย่าง อาจมีสาเหตุมาจากความคลาดเคลื่อนในข้อมูล ซึ่ง Iglewicz และ Hoaglin (1993) ได้แบ่งประเภทความคลาดเคลื่อนเป็น 3 ประเภท ดังนี้

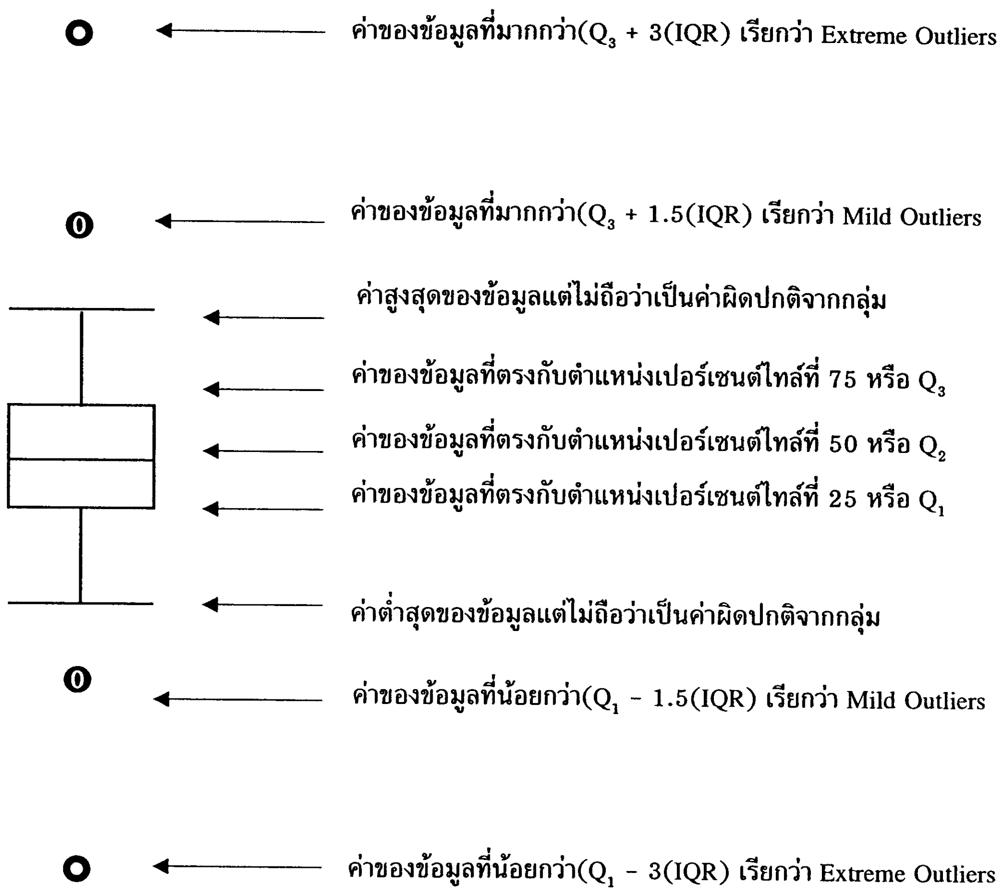
3.3.2.1 ความคลาดเคลื่อนจากความผันแปรที่มีในประชากรที่ทำการศึกษา (Inherent Variability) ไม่ได้เกิดจากการวัดหรือการเก็บข้อมูลที่ผิดพลาด เป็นความคลาดเคลื่อนที่หลีกเลี่ยงไม่ได้ เพราะไม่ว่าจะควบคุมการวัดหรือเก็บข้อมูลให้ดีเพียงใด ความผันแปรในประชากรก็ยังคงอยู่ เช่น มีส่วนสูงมากกว่าปกติ ไอคิวสูงมาก

3.3.2.2 ความคลาดเคลื่อนจากการวัด (Measurement Error) เป็นความคลาดเคลื่อนที่เกิดจากการใช้เครื่องมือในการวัดมีคุณภาพต่ำ หรือ เกิดจากความผิดพลาดของเครื่องมือวัด

3.3.2.3 ความคลาดเคลื่อนจากการปฏิบัติการ (Execution Error) เป็นความคลาดเคลื่อนที่เกิดจากความผิดพลาดของการบันทึกข้อมูลเพื่อประมวลผล เช่น การลงรหัสหรืออาจเกิดจากการเลือกตัวอย่างที่ไม่เหมาะสมกับประชากร

3.4 การตรวจสอบค่าผิดปกติจากกลุ่ม

ค่าผิดปกติจากกลุ่ม อาจมีผลทำให้การวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรมีความผิดพลาดได้ ดังนั้นจึงควรมีการตรวจสอบว่ามีค่าผิดปกติจากกลุ่มปรากฏอยู่ในข้อมูลที่นำมาใช้ในการวิเคราะห์หรือไม่ เพื่อประโยชน์ในการแก้ไขต่อไป ศิริชัย พงษ์วิชัย (2544) ได้กล่าวถึงวิธีการทางสถิติที่นิยมใช้ในการตรวจสอบหาค่าผิดปกติจากกลุ่มมีอยู่หลายวิธีซึ่งเทคนิคทางสถิติอนุมานต่าง ๆ ที่ใช้วิเคราะห์ก็สามารถหาค่าผิดปกติจากกลุ่มได้ แต่วิธีการหาค่าผิดปกติจากกลุ่มเบื้องต้นแบบง่าย ๆ คือ การใช้วิธีการของสถิติพรรณนาด้วยการหาค่าที่แสดงตำแหน่งของข้อมูล (N-tile เช่น ควอไทล์หรือเปอร์เซ็นต์ไทล์) และแสดงในรูปของกราฟรูปสี่เหลี่ยมผืนผ้าที่เรียกว่า Box plot ดังนี้ (ภาพที่ 8)



ภาพที่ 8 Box and Whisker Plots

ฉัตรศิริ ปิยะพิมลสิทธิ์ (2544) ได้สรุปการแปลผลจากกราฟแบบ Box Plots ไว้ว่า Box Plots จะแสดงให้เห็นสิ่งต่าง ๆ ที่สำคัญ 3 ประการ ดังนี้ ประการแรก เส้นตรงกลางจะแบ่งการแจกแจงซึ่งมีลักษณะเกือบสมมาตร ชี้ให้เห็นว่าค่ามัธยฐานเป็นค่าที่อยู่ตรงกลางพอดี ประการที่สอง การพิจารณาลักษณะการกระจายของข้อมูล เช่น ถ้าการแจกแจงมีลักษณะเบ้บวก เห็นได้จากเส้น Whiskers ทางขวาจะยาวกว่าทางซ้าย เมื่อกราฟอยู่ตามแนวนอน หรือเส้น Whiskers ทางด้านบนจะยาวกว่าทางด้านล่าง เมื่อกราฟอยู่ตามแนวตั้ง และประการสุดท้าย จะเห็นจุดที่อยู่นอกขอบเขต 1.5 เท่าของ Q_3- Q_1(Inner Fences) เราจะเรียกจุดเหล่านั้นว่า ค่าผิดปกติจากกลุ่ม (Outliers) คำว่า Outliers เป็นคำที่อนุญาตให้ใช้เรียกค่าที่อยู่ปลายสุดได้ แต่ในความหมายที่แท้จริงของมันจะใช้เฉพาะค่าที่อยู่ถัดจาก Inner Fences เท่านั้น ส่วนค่าที่อยู่ไกลออกไปจะเรียกว่า Extreme Values

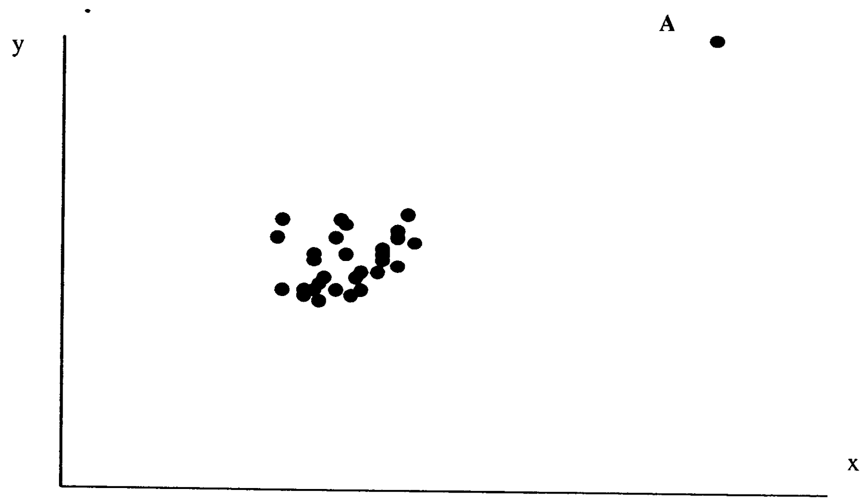
4. ความสัมพันธ์ระหว่างสัมประสิทธิ์สหสัมพันธ์กับค่าผิดปกติจากกลุ่ม

เมื่อตรวจสอบพบว่ามีค่าผิดปกติจากกลุ่มอยู่ในข้อมูลที่ใช้ในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ปัญหาที่ตามมาคือ ควรกำจัดค่าผิดปกติจากกลุ่มนั้นออกไปจากการวิเคราะห์หรือไม่ คำตอบจะขึ้นอยู่กับสาเหตุของการเกิดค่าผิดปกติจากกลุ่ม ถ้าค่าผิดปกติจากกลุ่มเกิดขึ้นเนื่องจากความผิดพลาดในการเก็บรวบรวมข้อมูล การแก้ปัญหาอาจทำได้โดยการตัดค่าผิดปกติจากกลุ่มนั้นออกจากการวิเคราะห์ แต่ถ้าตรวจพบค่าผิดปกติจากกลุ่มนั้นเป็นข้อมูลที่ถูกต้อง ผู้วิเคราะห์ต้องให้ความสนใจกับค่าผิดปกติจากกลุ่มนั้นโดยใช้เป็นข้อมูลเพื่อการวิเคราะห์ด้วย โดยมีความจำเป็นต้องหาข้อมูลเพิ่มเติม และอาจมีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์

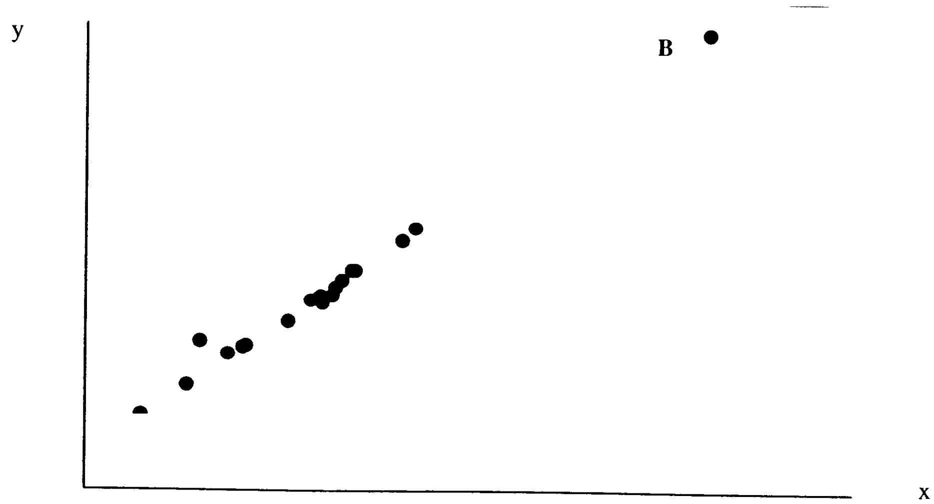
เปลี่ยนแปลงไปจากกรณีที่วิเคราะห์โดยไม่รวมเอาค่าผิดปกติจากกลุ่มไว้ แสดงว่าค่าผิดปกติจากกลุ่มนั้นเป็นค่าที่มีอิทธิพล และอาจต้องแปลงข้อมูลที่ใช้ในการวิเคราะห์

4.1 การตรวจสอบค่าผิดปกติจากกลุ่มที่มีอิทธิพลและการแก้ปัญหาค่าผิดปกติจากกลุ่ม

เนื่องจากค่าผิดปกติจากกลุ่มอาจเป็นค่าที่มีอิทธิพลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์เปลี่ยนแปลงไป ซึ่งการเปลี่ยนแปลงดังกล่าวอาจเกิดขึ้นทั้งขนาดความสัมพันธ์ ทิศทางความสัมพันธ์ หรืออาจไม่เป็นค่าที่มีอิทธิพลก็ได้ จึงควรตรวจสอบว่าค่าผิดปกติจากกลุ่มใดเป็นค่าที่มีอิทธิพล อรรมย์คณา แย้มนวล (2545) ได้เสนอวิธีการตรวจสอบค่าอิทธิพลในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์และการวิเคราะห์การถดถอยอย่างง่ายโดยใช้กราฟ ซึ่งเป็นวิธีการทดสอบง่าย ๆ โดยเขียนแผนภาพกระจาย ตัวอย่างเช่นผลจากการเก็บข้อมูลสองชุดเพื่อนำมาวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ พบว่าเมื่อนำมาเขียนแผนภาพกระจายแสดงดังภาพที่ 9 และ ภาพที่ 10 จะได้ค่าผิดปกติจากกลุ่มในภาพ 9 เป็นค่าที่มีอิทธิพลแต่ในภาพที่ 10 เป็นค่าผิดปกติจากกลุ่มที่ไม่มีอิทธิพลต่อค่าสัมประสิทธิสหสัมพันธ์ ซึ่งอธิบายได้ดังนี้



ภาพที่ 9 แผนภาพการกระจายเมื่อค่าผิดปกติจากกลุ่มเป็นค่าที่มีอิทธิพล



ภาพที่ 10 แผนภาพการกระจายเมื่อค่าผิดปกติจากกลุ่มเป็นค่าที่ไม่เป็นค่าที่มีอิทธิพล

จากภาพที่ 9 แสดงลักษณะข้อมูล x และ y ที่มีจุด A เป็นค่าผิดปกติจากกลุ่มและเป็นค่าที่มีอิทธิพลต่อการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ เนื่องจากเมื่อพิจารณาข้อมูล ณ จุดอื่นๆจากภาพจะพบว่าเมื่อตัดจุด A ออกจากการวิเคราะห์ ตัวแปร x และ y อาจมีความสัมพันธ์กันในทางบวกหรือทางลบก็ได้ เปรียบเทียบกับกรณีวิเคราะห์โดยรวมจุด A ด้วย จุด A จะเป็นค่าที่มีผลทำให้เบี่ยงเบนโดยความสัมพันธ์ระหว่างตัวแปร x และ y มีแนวโน้มเป็นบวก จุด A จึงเป็นค่าผิดปกติจากกลุ่มที่มีอิทธิพลต่อการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ ดังนั้นการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ในกรณีนี้ควรมีการหาข้อมูลมาสนับสนุนเพิ่มเติม หรือใช้สถิติการวิเคราะห์ที่มีความแกร่ง เพื่อผลการวิเคราะห์และการสรุปผลที่ถูกต้อง

จากภาพที่ 10 แสดงลักษณะข้อมูล x และ y ที่มีจุด B เป็นค่าผิดปกติจากกลุ่มแต่ไม่เป็นค่าที่มีอิทธิพลต่อการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ เนื่องจากเมื่อพิจารณาความสัมพันธ์ระหว่างตัวแปร x และ y จะเห็นแนวโน้มความสัมพันธ์ในทางบวกเช่นเดียวกัน ไม่ว่าจะรวมค่าผิดปกติจากกลุ่มจุด B วิเคราะห์ด้วยหรือไม่ก็ตาม ดังนั้นค่าผิดปกติจากกลุ่มในกรณีนี้จึงไม่เป็นค่าผิดปกติจากกลุ่มที่มีอิทธิพลต่อการวิเคราะห์

การแก้ปัญหาค่าผิดปกติจากกลุ่มในการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ จะขึ้นอยู่กับสาเหตุของการเกิดค่าผิดปกติจากกลุ่ม ถ้าเป็นค่าผิดปกติจากกลุ่มเกิดจากความผิดพลาดในการเก็บรวบรวมข้อมูล วิธีการแก้ปัญหาทำได้โดยตัดข้อมูลที่เป็นค่าผิดปกติจากกลุ่มนั้นออกไป แต่ถ้าค่าผิดปกติจากกลุ่มเป็นค่าที่ถูกต้องจะทำการตรวจสอบต่อไปว่าค่าผิดปกนั้นเป็นค่าที่มีอิทธิพลหรือไม่ หากพบว่าค่าผิดปกติจากกลุ่มเป็นค่าที่มีอิทธิพลจะต้องมีการเก็บรวบรวมข้อมูลเพิ่มเติมเพื่อผลการวิเคราะห์ที่มีความถูกต้องมากยิ่งขึ้น

4.2 ผลกระทบของค่าผิดปกติจากกลุ่มต่อสัมประสิทธิ์สหสัมพันธ์

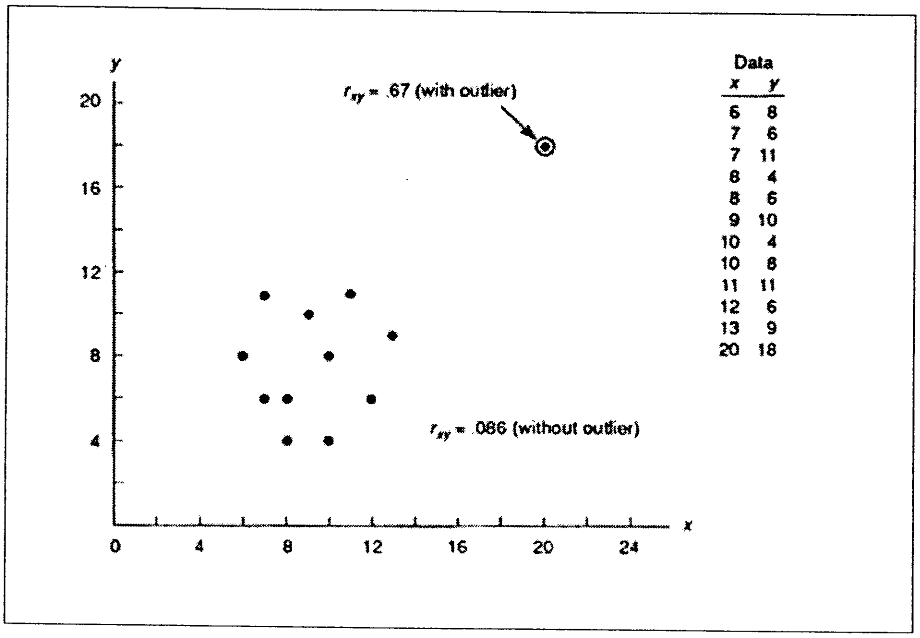
เดฟลิน และคณะ (Devlin et al., 1975) ได้สรุปผลกระทบของค่าผิดปกติจากกลุ่มที่มีต่อสัมประสิทธิ์สหสัมพันธ์ไว้ว่า

(1) ผลกระทบต่อระดับความสัมพันธ์และทิศทางความสัมพันธ์ ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้สูงหรือต่ำกว่าความเป็นจริง และอาจทำให้ทิศทางความสัมพันธ์เปลี่ยนไปเกิดความคลาดเคลื่อนต่อการอนุมานไปยังประชากรที่ศึกษา

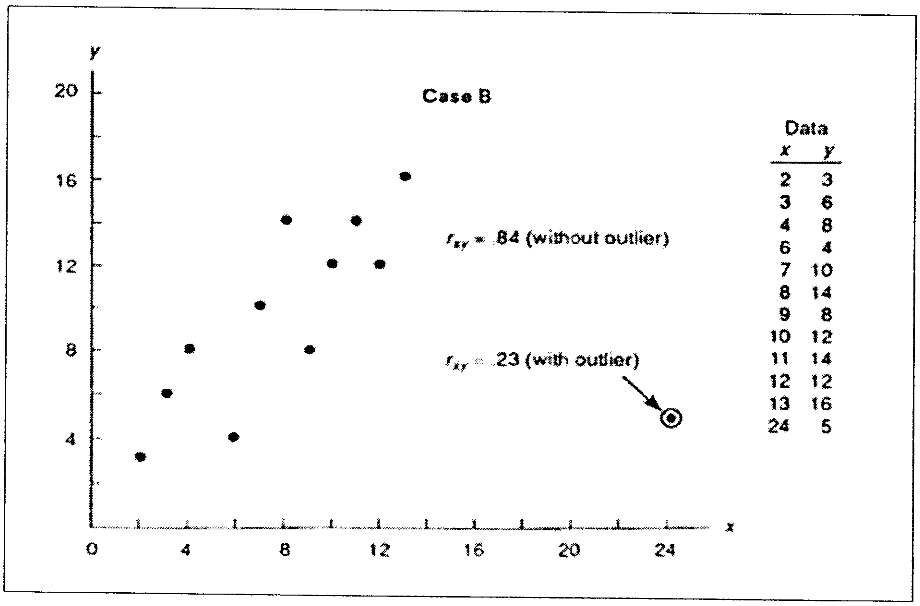
(2) ผลกระทบต่อการสรุปผลการมีนัยสำคัญทางสถิติ (Statistically Significance) เมื่อระดับความสัมพันธ์และทิศทางความสัมพันธ์เปลี่ยนไป ทำให้เกิดความคลาดเคลื่อนต่อการสรุปผลการมีนัยสำคัญทางสถิติ เนื่องจากมีโอกาสเกิดความคลาดเคลื่อนประเภทที่ 1 (Type I Error) สูงขึ้น การสรุปผลการทดสอบสมมติฐานจึงมีโอกาสผิดพลาดสูง

4.2.1 ผลกระทบต่อระดับความสัมพันธ์และทิศทางความสัมพันธ์

ฉัตรศิริ ปิยะพิมลสิทธิ์ (2547) ได้กล่าวถึงผลกระทบของค่าผิดปกติจากกลุ่มที่มีต่อสัมประสิทธิ์สหสัมพันธ์ว่า ค่าผิดปกติจากกลุ่มที่เป็นค่าที่มีอิทธิพลสามารถเปลี่ยนแปลงผลของสหสัมพันธ์ได้ เมื่อวิเคราะห์ข้อมูลโดยรวมค่าผิดปกติจากกลุ่มด้วย ขนาดความสัมพันธ์ที่ได้จะมีความสัมพันธ์กันสูงหรือต่ำกว่าความเป็นจริง และเมื่อขจัดค่าผิดปกติจากกลุ่มออกจากการวิเคราะห์ อาจมีความสัมพันธ์ในลักษณะตรงกันข้าม จากภาพที่ 11 ในกรณี A จะเห็นว่าตัวแปร x และ y มีความสัมพันธ์กันสูงเมื่อรวมค่าผิดปกติจากกลุ่มไว้ด้วย (r_{xy}=0.67) แต่เมื่อขจัดค่าผิดปกติจากกลุ่มออกกลับไม่มีความสัมพันธ์กัน (r_{xy}=0.086) ในขณะที่กรณี B ตัวแปร x และ y มีความสัมพันธ์กันน้อยเมื่อรวมค่าผิดปกติจากกลุ่มไว้ด้วย (r_{xy}=0.23) แต่ความสัมพันธ์จะเพิ่มสูงขึ้นเมื่อขจัดค่าผิดปกติจากกลุ่มออกไป (r_{xy}=0.84)



กรณี A

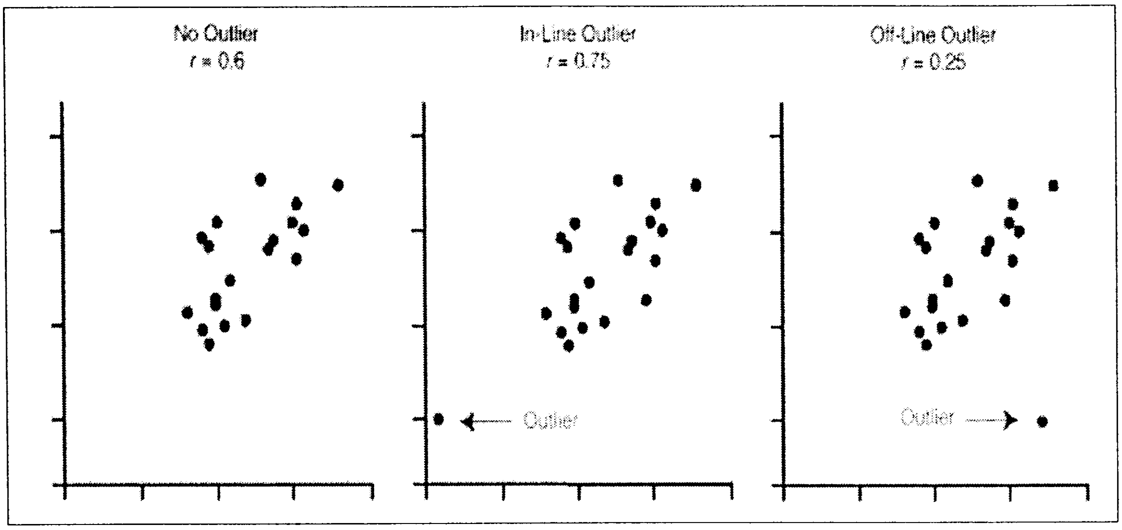


กรณี B

ภาพที่ 11 แสดงอิทธิพลของค่าผิดปกติจากกลุ่มที่มีต่อสัมประสิทธิ์สหสัมพันธ์ (Stevens, 1992)

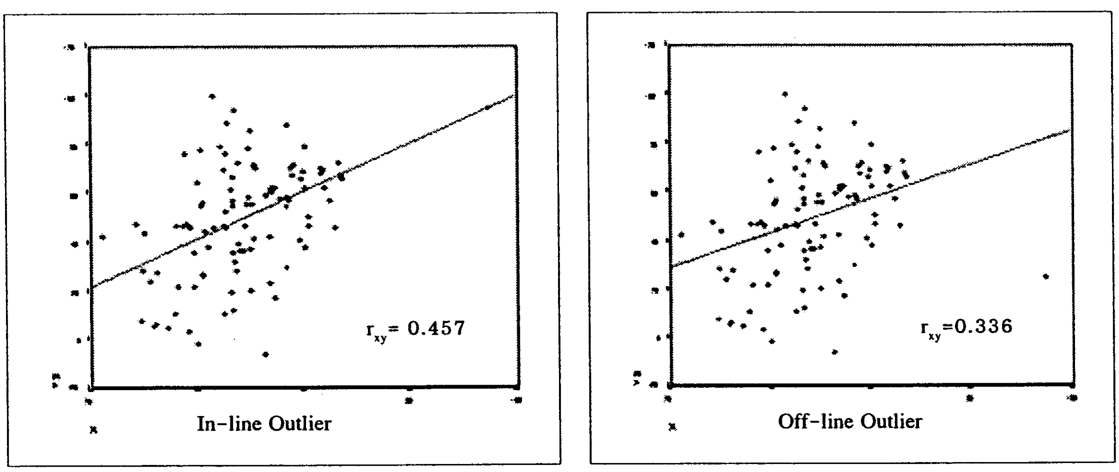
ระดับความสัมพันธ์และทิศทางความสัมพันธ์ จะเปลี่ยนแปลงไปมากน้อยเพียงไรขึ้นอยู่กับตำแหน่งของค่าผิดปกติจากกลุ่มที่เกิดขึ้น เมื่อพิจารณาจากเส้นสมการถดถอย (Regression Line) ตามแผนภาพการกระจายของชุดข้อมูล จะพบว่า ค่าผิดปกติจากกลุ่มที่เกิดขึ้นในตำแหน่งที่ใกล้เส้นสมการถดถอย (In-Line Outlier) จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเพิ่มขึ้น และค่าผิดปกติจากกลุ่มที่อยู่ในตำแหน่งที่

ห่างจากเส้นสมการถดถอย (Off-Line Outlier) จะทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ลดลง (David, 2006) (ภาพที่ 12)



ภาพที่ 12 แสดงระดับความสัมพันธ์กรณีที่เกิดค่าผิดปกติจากกลุ่มในแนวเส้นการถดถอยและนอกเส้นการถดถอย(American Collage of Physicians, 2006)

David (2006) ได้อธิบายถึงผลกระทบจากขนาดตัวอย่างและตำแหน่งที่เกิดค่าผิดปกติจากกลุ่มต่อสัมประสิทธิ์สหสัมพันธ์ว่า ในการศึกษาที่ขนาดตัวอย่างมีจำนวนมากค่าผิดปกติจากกลุ่มจะมีผลต่อค่าสหสัมพันธ์น้อยลง หรืออาจไม่มีผลกระทบใด ๆเลย และถ้าขนาดตัวอย่างมีจำนวนน้อยจะส่งผลกระทบต่อค่าสัมประสิทธิ์สหสัมพันธ์มากขึ้น จากภาพที่ 13 เมื่อขนาดตัวอย่างเท่ากับ 100 และตำแหน่งในการเกิดค่าผิดปกติจากกลุ่มในจุดที่แตกต่างกันคือค่าผิดปกติจากกลุ่มเกิดขึ้นในตำแหน่งที่ใกล้เส้นสมการถดถอย และค่าผิดเกิดขึ้นในตำแหน่งที่ห่างเส้นสมการถดถอย จะพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันมากนัก โดย r_{xy}= 0.457 และ r_{xy}= 0.336 ตามลำดับ



ภาพที่ 13 แสดงผลกระทบจากขนาดตัวอย่างและตำแหน่งที่เกิดค่าผิดปกติจากกลุ่มต่อสัมประสิทธิ์สหสัมพันธ์เมื่อขนาดตัวอย่างมีจำนวนมาก (n=100)

4.2.2 ผลกระทบต่อการสรุปผลการมีนัยสำคัญทางสถิติ

การทดสอบการมีนัยสำคัญทางสถิติ ในการวิเคราะห์ความสัมพันธ์ โดยใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 แบบ ค่าสถิติที่ทดสอบการมีนัยสำคัญคือสถิติทดสอบ t และสถิติทดสอบ Z ซึ่งจะแปรผันตามค่าสัมประสิทธิ์สหสัมพันธ์ (r) และขนาดตัวอย่าง (n) เมื่อมีค่าผิดปกติจากกลุ่มอยู่ด้วย จะทำให้ค่าสถิติทดสอบ t และสถิติทดสอบ Z มีค่าสูงหรือต่ำกว่าความจริง การสรุปผลการทดสอบสมมติฐานจึงมีโอกาสผิดพลาดสูง โดยผลกระทบต่อการมีนัยสำคัญทางสถิติเกิดขึ้นได้ 2 ลักษณะตามการทดสอบสมมติฐาน คือ ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) และความคลาดเคลื่อนประเภทที่ 2 (Type II Error)

5. ความแกร่งของการทดสอบทางสถิติ (Robustness of Statistical Test)

5.1 นิยามของสถิติที่มีความแกร่ง (Robust Statistics)

Huber (1981) ให้ความหมายของสถิติที่มีความแกร่ง คือ สถิติทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงผลการทดสอบแม้มีการเบี่ยงเบนจากข้อตกลงเบื้องต้นเพียงเล็กน้อย

Freund (1992) กล่าวว่า ความแกร่งเป็นตัวชี้วัดในกระบวนการประมาณค่าทางสถิติ และตัวประมาณค่าที่มีความแกร่งจะต้องไม่มีผลกระทบจากการละเมิดข้อตกลงเบื้องต้น ซึ่งอาจมีสาเหตุมาจากค่าผิดปกติจากกลุ่ม เครื่องมือที่ใช้วัด การบันทึกข้อมูล หรือความเสียหายที่เกิดในกระบวนการทดลอง

อโนทัย ตรีวานิช (2542) และ นุกูล ทรงมงคลรัตน์ (2545) ให้ความหมายของความแกร่งของการทดสอบ คือ สถิติที่มีคุณสมบัติของการทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่น ๆที่ไม่ใช่ปัจจัยของการทดสอบ เช่น ฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น

ศิริวรรณ ศิลป์สกุลสุข และ คณะ (2546) กล่าวว่า สถิติแบบไม่อิงพารามิเตอร์และสถิติแบบโรบัส (Non-Parametric and Robust Statistics) เป็นสถิติที่ใช้ข้อมูลที่อยู่ในรูปมาตราแสดงตำแหน่งหรือข้อมูลที่อยู่ในรูปลำดับที่ หรืออยู่ในรูปความถี่ และได้รับการยอมรับว่าสามารถใช้ได้กับข้อมูลที่มีการกระจายแบบไม่ปกติ เพราะจะพบบ่อยว่าข้อมูลหรือผลการทดสอบมีการกระจายแบบไม่ปกติ อีกทั้งข้อมูลที่มีค่าผิดปกติจากกลุ่มไม่มีอิทธิพลต่อการทดสอบ จึงไม่จำเป็นต้องใช้เทคนิคในการหาค่าผิดปกติจากกลุ่ม ซึ่งเป็นข้อได้เปรียบของสถิติแบบนี้

วิบูล วงศ์ภูวรักษ์ (2549) กล่าวว่า Robust Statistics เป็นสถิติที่ทดสอบแล้วจะให้คำตอบซึ่งค่อนข้าง "นิ่ง" (Stable) แม้ว่าจะมีข้อมูลที่มีค่าหลุดจากกลุ่มไปมาก ๆ ก็ตาม เช่น การใช้ค่ามัธยฐาน (Median) ซึ่งจะให้ค่าที่นิ่งกว่าค่าเฉลี่ย (Mean) เพราะค่าที่หลุดจากกลุ่มมักสูงหรือต่ำผิดปรกติ ตัวเลขที่ผิดปรกติเหล่านี้จะดึงค่าเฉลี่ยเข้าหาตัวเองจนเห็นอย่างชัดเจน แต่ค่ามัธยฐานจะยังอยู่ที่เดิม ซึ่งสถิติแขนงนี้จะมีประเด็นที่แปลกใหม่น่าสนใจอยู่มาก เช่น การหาช่วงความเชื่อมั่น โดยวิธี Resampling ซึ่งก็มีหลายรูปแบบ เช่น วิธี Jackknife และ Bootstrap Sampling ปัจจุบันทางการแพทย์จะตื่นตัวมาใช้สถิติประเภทนี้กันมากขึ้น เพราะสามารถช่วยคำนวณค่าช่วงความเชื่อมั่นได้ง่ายกว่าเดิม โดยไม่ต้องอิงทฤษฎีมาก

ดังนั้น สถิตที่มีความแกร่ง หมายถึง สถิติที่ทดสอบแล้วจะให้คำตอบซึ่งค่อนข้าง "นิ่ง" หรือ คงที่ (Stable) และคุณสมบัติของการทดสอบที่ไม่ไวต่อการเปลี่ยนแปลงของปัจจัยอื่น ๆที่ไม่ใช่ปัจจัยของการทดสอบ เช่น ค่าผิดปกติจากกลุ่ม หรือฝ่าฝืนข้อตกลงเบื้องต้นของการทดสอบนั้น

Mosteller และ Tukey (1977) กล่าวว่า สถิติที่มีความแกร่ง ควรมีคุณสมบัติในการทดสอบที่สำคัญ 2 ประการคือ

(1) ความต้านทาน (Resistance) หมายถึง คุณสมบัติของตัวประมาณค่าหรือสถิติทดสอบที่ไม่ทำให้ผลการประมาณค่าและผลการทดสอบทางสถิติเปลี่ยนแปลงมากเมื่อข้อมูลเดิมเปลี่ยนแปลงเพียง

เล็กน้อยหรือเปลี่ยนแปลงไปอย่างมาก เช่น ถ้าต้องการพิจารณาค่ากลางเพื่อเป็นตัวแทนของข้อมูลทั้งชุด ค่าเฉลี่ย (Mean) เป็นค่ากลางที่ไม่มีความต้านทาน ส่วนค่ามัธยฐานเป็นค่ากลางที่มีความต้านทาน การพิจารณาตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความต้านทานมีวิธีการพิจารณาได้หลายวิธี เช่น ค่าความแปรปรวนของตัวประมาณค่า ความเอนเอียงในการประมาณค่า (Bias) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง โดยตัวประมาณค่าใดที่ให้ค่าประมาณที่มีค่าความแปรปรวน ความเอนเอียงในการประมาณค่า หรือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่มีค่าน้อย จะมีคุณสมบัติความต้านทานที่ดี

ในการพิจารณาความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการประมาณเมื่อตัวอย่างมีค่าผิดปกติจากกลุ่มกับไม่มีค่าผิดปกติจากกลุ่ม ว่ามีการเปลี่ยนแปลงไปมากน้อยเพียงใดตามเกณฑ์ความต้านทานของ Mosteller and Tukey (1977) ซึ่งเป็นการวัดซ้ำในแต่ละสถานการณ์จำนวนหลายรอบ เพื่อเปรียบเทียบค่าความคลาดเคลื่อนจากตัวประมาณค่าพารามิเตอร์ในแต่ละแบบ ในการศึกษาครั้งนี้จะเลือกใช้วิธีการวัดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นการเปรียบเทียบความต้านทานของตัวประมาณค่าในแต่ละวิธี

(2) ความแกร่งด้านประสิทธิภาพ (Robustness of Efficiency) หมายถึง สถิติที่มีประสิทธิภาพ (Efficiency) สูงกว่าแบบอื่นในทุกสถานการณ์ที่มีความผันแปร โดยเปรียบเทียบในสถานการณ์เดียวกัน ประสิทธิภาพ หมายถึง ในการประมาณค่าพารามิเตอร์ใด ๆจะให้ค่าประมาณใกล้เคียงค่าพารามิเตอร์มากที่สุด การพิจารณาความแกร่งด้านประสิทธิภาพของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์จะพิจารณาจากค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; R.E.) จากอัตราการเปลี่ยนแปลงของค่าความถี่สัมพัทธ์ในแต่ละสถานการณ์

ตัวประมาณค่าทางสถิติส่วนใหญ่จะมีคุณสมบัติตามเกณฑ์ความแกร่งของ Mosteller และ Tukey เพียงข้อใดข้อหนึ่งเท่านั้น ซึ่งเป็นการยุ่งยากในการทดสอบที่จะสรุปได้ว่าสถิติใดมีคุณสมบัติความแกร่งครบทั้งสองข้อ เช่น ในการวิเคราะห์ความสัมพันธ์เมื่อตัวแปรเป็นข้อมูลต่อเนื่องและมีการแจกแจงเป็นแบบโค้งปกติ การเลือกใช้ตัวประมาณค่าในการวิเคราะห์ความสัมพันธ์ที่เหมาะสมที่สุดคือ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน อย่างไรก็ตามเมื่อชุดข้อมูลมีการแจกแจงปกติแบบปลอมปน (Contaminate Normal Distribution: CN) เช่น การแจกแจงปกติแบบหางยาว ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันจะไม่มีความแกร่งด้านประสิทธิภาพ และไม่มีความต้านทาน ต่อผลกระทบจาก CN ทำให้ผลการประมาณค่าพารามิเตอร์มีความผันแปร แต่สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ซึ่งใช้การจัดอันดับของข้อมูลมาคำนวณจึงได้รับผลกระทบจากข้อมูล CN น้อยกว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนจึงมีความแกร่งด้านประสิทธิภาพดีกว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Statistical Engineering Division, 2006 ; Mosteller and Tukey, 1977)

5.2 วิธีการทดสอบสถิติที่มีความแกร่ง

ในการทดสอบสถิติที่มีความแกร่งมีหลากหลายวิธี ขึ้นอยู่กับวัตถุประสงค์ของการทดสอบและวิธีการทางสถิติในแต่ละชนิด ในการศึกษาครั้งนี้จะใช้วิธีการทดสอบความแกร่ง 2 ลักษณะ คือ

5.2.1 ความแกร่งในการทดสอบทางสถิติ (Robust of Statistical Test)

การทดสอบความแกร่งโดยวิธีการทดสอบทางสถิติ จะพิจารณาจากความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก อำนาจการทดสอบ และค่าประสิทธิภาพสัมพัทธ์ ดังรายละเอียดต่อไปนี้

5.2.1.1 กรณีที่สมมติฐานหลักเป็นจริง (ρ=0) จะพิจารณาจากค่าความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้งหมด กล่าวคือ สำหรับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แต่ละแบบ จะคำนวณหาค่าความถี่สัมพัทธ์ของผลการทดสอบที่

สรุปว่าปฏิเสธสมมติฐานหลัก แล้วนำมาเปรียบเทียบกับค่าระดับนัยสำคัญที่กำหนดไว้(α_0) หากมีค่าอยู่ในช่วงการยอมรับ จะถือว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้นมีความแกร่งของการทดสอบ (Robustness of The Tests) หรือแสดงว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้นสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ ทั้งนี้ค่าความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักจากการทดลอง จะกำหนดให้มีค่าเท่ากับ จำนวนครั้งของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักหารด้วยจำนวนครั้งของการทดสอบสมมติฐานทั้งหมด เมื่อจำนวนครั้งที่ทำการทดลองเพิ่มขึ้นมาก ๆค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง จะมีค่าใกล้เคียงกับค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 หรือค่าระดับนัยสำคัญ นั่นเอง (นุกูล ทรงมงคลรัตน์, 2545)

Bradley (1978) ได้เสนอวิธีการวัดความสามารถในการควบคุมความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริงโดยใช้หลักการทดสอบสมมติฐานเกี่ยวกับความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ดังนี้

กำหนดให้ α_* แทน ความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นจริง

α_0 แทน ระดับนัยสำคัญที่กำหนดในการศึกษาคือ 0.05

สมมติฐานในการทดสอบเป็นดังนี้

$H_0: \quad \alpha_* = \alpha_0$

$H_1: \quad \alpha_* \neq \alpha_0$

ระดับนัยสำคัญที่กำหนดในการศึกษา (α_0) คือ 0.05

$$\alpha_0 - Z_{\alpha/2}\sqrt{\frac{\alpha_0(1-\alpha_0)}{n}} \le \alpha_. \le \alpha_0 + Z_{\alpha/2}\sqrt{\frac{\alpha_0(1-\alpha_0)}{n}}$$

$$0.05 - 1.96\sqrt{\frac{0.05(1-0.05)}{1000}} \le \alpha_* \le 0.05 + 1.96\sqrt{\frac{0.05(1-0.05)}{1000}}$$

$$0.036 \le \alpha_* \le 0.064$$

เมื่อกำหนดระดับนัยสำคัญที่กำหนดในการศึกษา(α_0) เท่ากับ 0.05 แบบทดสอบสามารถควบคุมความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริงไว้ได้ ถ้า α_* อยู่ในช่วง 0.036 ถึง 0.064

กรณีสมมติฐานหลักเป็นจริง เมื่อทำการทดลองจนครบ 1,000 ครั้งในแต่ละสถานการณ์ จะทำการนับจำนวนครั้งที่ปฏิเสธสมมติฐานหลัก แล้วนำไปหาความถี่สัมพัทธ์และนำไปเปรียบเทียบกับช่วงของระดับนัยสำคัญที่กำหนดไว้ข้างต้น ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบใดมีความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริงอยู่ภายในช่วงของระดับนัยสำคัญที่กำหนดไว้ดังกล่าวแสดงว่าประมาณค่าสัมประสิทธิ์นั้นมีความแกร่งหรือควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้

5.2.1.2 กรณีที่สมมติฐานหลักไม่เป็นจริง ($\rho \neq 0$) จะพิจารณา 2 กรณี

(1) พิจารณาจากอำนาจการทดสอบ (Power of Test) ในกรณีนี้จะพิจารณาตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่สามารถควบคุมความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักไม่เป็นจริงได้ โดยจะพิจารณาจากค่าความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลัก

จากการทดลองเมื่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แต่ละแบบ ถ้าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ใดมีค่าความถี่สัมพัทธ์ของผลการทดสอบสูงกว่าตัวประมาณค่าแบบอื่น ๆ แสดงว่าตัวประมาณค่านั้นมีความแกร่งกว่าตัวประมาณค่าอื่น ๆที่เหลือ และถ้าตัวประมาณค่า 2 แบบมีค่าความถี่สัมพัทธ์เท่ากัน แสดงว่าตัวประมาณค่าทั้งสองมีประสิทธิภาพไม่แตกต่างกัน ทั้งนี้หากจำนวนครั้งที่ทำการทดลองเพิ่มขึ้นมาก ๆ ค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักไม่เป็นจริง จะมีค่าใกล้เคียงหรือเทียบเท่ากับค่าอำนาจของการทดสอบ โดยอำนาจการทดสอบคำนวณจาก

$$\text{Power of Test} = \frac{\text{จำนวนครั้งที่ปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักไม่เป็นจริง}}{1{,}000}$$

(2) พิจารณาจากค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency; R.E.)

อโนทัย ตรีวานิช (2542) ได้รายงานว่าการพิจารณาค่าประสิทธิภาพสัมพัทธ์ จะพิจารณาจากอัตราการเปลี่ยนแปลงของค่าความถี่สัมพัทธ์ของการปฏิเสธสมมติฐานหลัก เมื่อมีค่าผิดปกติจากกลุ่มในจำนวน 5%,10%, 20% และ 30% ของขนาดตัวอย่าง เปรียบเทียบกับเมื่อไม่มีค่าผิดปกติจากกลุ่ม โดยกำหนดให้ R.E.มีค่าเป็นดังนี้

$$\text{R.E.} = \frac{f_2}{f_1}$$

เมื่อ f_1 = ความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อใช้ตัวประมาณค่านั้น และไม่มีค่าผิดปกติจากกลุ่ม

f_2 = ความถี่สัมพัทธ์ของผลการทดสอบที่สรุปว่าปฏิเสธสมมติฐานหลักเมื่อใช้ตัวประมาณค่านั้น และมีค่าผิดปกติจากกลุ่มเท่ากับ 5%,10%, 20% และ 30% ของขนาดตัวอย่าง

การแปลความหมายของค่า R.E. เป็นดังนี้

R.E. > 1 หมายถึง การมีค่าผิดปกติจากกลุ่มในตัวอย่างส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น ทำให้มีโอกาสปฏิเสธสมมติฐานหลักมากขึ้น

R.E. < 1 หมายถึง การมีค่าผิดปกติจากกลุ่มในตัวอย่างส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น ทำให้มีโอกาสปฏิเสธสมมติฐานหลักน้อยลง

R.E. $= 1$ หมายถึง การมีค่าผิดปกติจากกลุ่มในตัวอย่างไม่ส่งผลกระทบต่อตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์นั้น

5.2.2 ความแกร่งในการประมาณค่า (Robust of Estimation)

ในการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ เป็นการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรจากค่าสัมประสิทธิ์สหสัมพันธ์ของตัวอย่าง ดังนั้นตัวประมาณค่าที่มีความแกร่งจะให้ค่าประมาณที่เท่ากับค่าพารามิเตอร์หรือใกล้เคียงกัน เกณฑ์การตัดสินว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์ใดมีความแกร่งในการประมาณค่า จะเปรียบเทียบจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error; MSE) จากตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี โดยวิธีที่ให้ค่า MSE ต่ำกว่าจะมีประสิทธิภาพมากกว่า ซึ่งถือว่าเป็นตัวประมาณค่าสัมประสิทธิสหสัมพันธ์ที่มีความแกร่ง MSE คำนวณจากสูตร (Rousseeuw and Leroy, 1987)

$$\mathrm{MSE}\,(\theta_j) = \frac{1}{m}\sum_{k=1}^{m}(\hat{\theta}_j^{(k)} - \theta_j)^2 \qquad (2.27)$$

เมื่อ θ_j หมายถึง ค่าพารามิเตอร์ตัวที่ j ในการทำซ้ำรอบที่ k

$\hat{\theta}_j^{(k)}$ หมายถึง ค่าประมาณพารามิเตอร์ตัวที่ j ในการทำซ้ำรอบที่ k

m หมายถึง จำนวนรอบที่ทำซ้ำ

6. เทคนิคและวิธีมอนติคาร์โล

6.1 ความเป็นมาของวิธีมอนติคาร์โล

อรวรรณ บุญพละ (2539) ได้สรุปว่ามอนติคาร์โล เป็นสาขาหนึ่งของวิชาคณิตศาสตร์ที่ใช้คอมพิวเตอร์ช่วยในการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่ม (Random Number) หรือเลขสุ่มเทียม (Pseudo Random Number) มาสร้างตัวแปรให้มีลักษณะใกล้เคียงกับสถานการณ์จริงและมีการทดลองซ้ำได้หลาย ๆ ครั้ง เพื่อให้ได้ค่าที่แน่นอนสำหรับสรุปหรืออธิบายปรากฏการณ์ต่าง ๆ ในสถานการณ์จริงนั้น ๆ หรือช่วยหาคำตอบในเรื่องราวต่าง ๆ ที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น Rubinstein (1981) และ ต่าย เซี่ยงฉี (2534)ได้รายงานว่าวิธีมอนติ คาร์โลถูกใช้แก้ปัญหาแรกเริ่มที่สุดคือในคริสต์ศตวรรษที่ 17 ในปี ค.ศ. 1733 จอร์ส หลุยส์ เลคเลอร์และบูฟฟอน (Georges Louis Leclere and Comte de Buffon) ได้พัฒนาทฤษฎีความน่าจะเป็น (Probability Theory) ทำการทดลองที่เรียกว่าปัญหาเข็มของบูฟฟอน (The Buffon's Neddle Problem) เพื่อหาคำตอบที่ว่าความน่าจะเป็น (Probability :P) ของเข็มยาว 1 หน่วย ถูกโยนแบบสุ่มลงมาบนพื้นราบที่มีการลากเส้นขนานบนพื้นโดยให้ระยะระหว่างเส้นขนานคู่หนึ่ง ๆ มีค่า d หน่วย โดยกำหนดให้ d มีค่ามากกว่า 1 แล้ว ความน่าจะเป็นที่เข็มตกลงมาตัดกับเส้นขนานนี้มีค่า P=2 $1/\pi$d หรือเป็นค่าประมาณของอัตราส่วนของจำนวนครั้งที่โยนเข็มลงมาบนพื้นให้ตัดกับเส้นขนานต่อจำนวนครั้งของการโยนเข็มลงมาทั้งหมดนั่นเอง การทดลองที่สร้างชื่อเสียงให้กับกอสเซท (W.S. Gosset) มากที่สุดคือในปี ค.ศ. 1908 ได้นำวิธีมอนติคาร์โล มาศึกษาสหสัมพันธ์และการแจกแจงค่าที(t-Distribution) อรวรรณ บุญพละ (2539) ได้กล่าวว่าชื่อมอนติคาร์โลถูกนำมาใช้โดยฟอน นอยมันน์และอูลาม (Von Neuman and Ulam) ระหว่างสงครามโลกครั้งที่สองในปี ค.ศ. 1944 โดยใช้เป็นรหัสลับของงานที่ทำในลอส อาลามอส (Los Alamos) ซึ่งเป็นงานที่เกี่ยวข้องกับการสร้างระเบิดปรมาณูในโครงการแมนฮัทตัน (Manhattan Project) โดยนำวิธีมอนติคาร์โล มาหาผลของการแพร่อย่างสุ่มของนิวตรอน (Neutron Diffusion) ในวัสดุเชื้อเพลิง ซึ่งเป็นการทดลองทางคณิตศาสตร์เพื่อหาผลของคำตอบก่อนที่จะทำการทดลองจริง เพื่อเป็นการลดอันตรายและช่วยประหยัดค่าใช้จ่ายก่อนการทดลองจริง หลังจากนั้นวิธีของ มอนติคาร์โล ถูกนำมาใช้อย่างกว้างขวางทั้งทางด้านฟิสิกส์ คณิตศาสตร์ สถิติ และการวิจัย นับว่าวิธีมอนติคาร์โล มีประโยชน์อย่างมากในการขยายความรู้ในเชิงทฤษฎี เช่น การนำมาศึกษาค่าความคลาดเคลื่อนทางสถิติ และเปรียบเทียบประสิทธิภาพของการทดสอบต่าง ๆ

6.2 ขั้นตอนของเทคนิคมอนติคาร์โล

ศิริจันทร์ ทองประเสริฐ (2544) กล่าวว่าเทคนิควิธีมอนติคาร์โล มีขั้นตอนสำคัญสรุปได้ดังนี้

6.2.1 การสร้างตัวเลขสุ่ม (Generate Random Number) การสร้างตัวเลขสุ่มที่ดีนั้นจะต้องมีลักษณะการแจกแจงความน่าจะเป็นแบบสม่ำเสมอ ดังนั้น การสร้างตัวเลขแต่เดิมจึงอาศัยเครื่องมือทางกายภาพต่าง ๆ เช่น ล้อรูเล็ต การเขียนตัวเลขในแผ่นกระดาษแล้วจับฉลาก ไพ่ เป็นต้น การใช้เครื่องมือดังกล่าวใช้ได้กับการสร้างตัวเลขสุ่มที่มีจำนวนไม่มากนัก ในกรณีที่ต้องการสร้างตัวเลขสุ่มที่มีจำนวนมาก ๆ จึง

ต้องหาเครื่องมือชนิดอื่นช่วยในการสร้างตัวเลขสุ่ม บริษัท RAND จึงสร้างเครื่องมือที่ใช้สร้างตัวเลขสุ่ม ด้วยเครื่องกำเนิดพัลส์อิเลกทรอนิกส์ (Electronic Pluse Generator) ซึ่งทำงานด้วยเสียง เครื่องดังกล่าวสามารถสร้างตัวเลขสุ่มได้เป็นล้านตัว แต่การสร้างตัวเลขสุ่มด้วยเครื่องมือทางกายภาพยังมีปัญหา 2 ประการ คือ เป็นการยากที่จะทำให้คอมพิวเตอร์สามารถเลือกใช้ได้เมื่อมีความต้องการ และเป็นการยากที่จะทำให้เครื่องมือดังกล่าวสร้างตัวเลขสุ่มชุดเดียวกัน หรือถ้าจะเก็บเลขสุ่มที่สร้างขึ้นมาไว้ในหน่วยความจำ หรือจานแม่เหล็กจะทำให้สูญเสียหน่วยความจำหรือเสียเวลาในการค้นหา ดังนั้นการสร้างตัวเลขสุ่มด้วยเครื่องคอมพิวเตอร์จึงนิยมสร้างตัวเลขสุ่มเทียม (Pseudo Random Number) โดยอาศัยสูตรทางคณิตศาสตร์วิธีที่นิยมใช้กันมากมี 2 วิธีคือ

6.2.1.1 วิธีส่วนกลางกำลังสอง (Midsquare Method) เป็นวิธีในยุคแรก ๆ ของการสร้างตัวเลขสุ่มเทียม โดยมีขั้นตอนในการสร้างตัวเลขสุ่มดังนี้

(1) กำหนดตัวเลขสี่หลัก

(2) ยกกำลังสองตัวเลขที่กำหนดนั้น ถ้าเลขที่ได้ไม่เกิน 8หลักให้เติมศูนย์นำหน้า

(3) ใช้เลขที่หลักกลางที่ได้ในขั้น (2) เป็นตัวเลขแบบสุ่ม

(4) ยกกำลังสองของตัวเลขในข้อ (1)

(5) ทำซ้ำในข้อ (3) และข้อ (4) จนได้จำนวนตัวเลขสุ่มตามต้องการ

ตัวอย่าง กำหนดตัวเลขสี่หลักแรกเป็น X = 2519 ดำเนินตามวิธีการที่ 1 – 5

$X = 2519$ จะได้ $X^2 = 06\underline{3453}61$

$X = 3453$ จะได้ $X^2 = 11\underline{9232}09$

$X = 9232$ จะได้ $X^2 = 85\underline{2298}24$

$X = 2298$ จะได้ $X^2 = 05\underline{2808}04$

$X = 2808$ จะได้ $X^2 = 07\underline{8848}64$

$X = 8848$ ฯลฯ

จะได้ตัวเลขสุ่ม คือ 2519 , 3453 , 9232 , 2298 , 2808 , 8848 , ...

การสร้างตัวเลขสุ่มโดยวิธีส่วนกลางกำลังสองนี้มีจุดด้อย คือ จะไม่ทราบช่วงของตัวเลขสุ่มที่จะหมุนกลับมาซ้ำเดิมอีก และถ้ากำหนดตัวเลขเริ่มแรกไม่เหมาะสม จะทำให้ตัวเลขสุ่มที่ไม่เป็นแบบสุ่ม เช่น ถ้ากำหนดตัวเลขสุ่มตัวแรก เป็น X = 2500

$X = 2500$ จะได้ $X^2 = 06\underline{2500}00$

$X = 2500$ จะได้ $X^2 = 06\underline{2500}00$

$X = 2500$ จะได้ $X^2 = 06\underline{2500}00$

$X = 2500$ ฯลฯ

ตัวเลขที่ได้คือ 2500, 2500 , 2500 , 2500 , 2500 , ...ซึ่งไม่ใช่ตัวเลขสุ่ม

6.2.1.2 วิธีเศษเหลือ (Congruent Method) วิธีเศษเหลือที่นิยมใช้กันมากที่สุด คือ การใช้เศษเหลือของผลคูณ (Multiplicative Congruential Method) ซึ่งใช้สูตร ดังนี้

$$X_{i+1} = a X_i (\mathrm{mod}\ m)$$

โดยที่ a , m เป็นจำนวนเต็มที่มากกว่า 0

วิธีการคือ กำหนดค่าเริ่มต้น สมมุติให้เท่ากับ X_i นำค่า X_i คูณด้วยค่าคงที่ a แล้วหารด้วย m เศษเหลือจากการหารคือ ตัวเลขที่ต้องการ ตัวอย่าง เช่นให้ a = 3 , m = 10 และให้ X_0 = 2

X_0 = 2

X_1 = (3 * 2) mod 10 = 6

X_2 = (3 * 6) mod 10 = 8

X_3 = (3 * 8) mod 10 = 4

X_4 = (3 * 4) mod 10 = 2

X_5 = (3 * 2) mod 10 = 6

จะได้เลขสุ่มคือ 6, 8, 4 , 2 , 6, ... จะเห็นว่าเลขสุ่มเริ่มวนกลับมาเลขเดิมเร็วเกินไป ดังนั้นในทางปฏิบัติเมื่อใช้คอมพิวเตอร์ จึงนิยมกระทำดังนี้

(1) กำหนดตัวเลขสุ่มใดๆ ที่น้อยกว่า 9 หลัก ให้เป็น X โดยปกติถ้าเป็นคอมพิวเตอร์ฐานสอง X จะให้เลขคี่ที่เป็นบวก และคอมพิวเตอร์ฐานสิบ X จะให้เป็นเลขบวกที่หารด้วย 2 และ 5 ไม่ลงตัว

(2) คูณเลขในข้อ (1) ด้วยค่า a โดยที่ a ไม่ควรมีค่าน้อยกว่า 5 หลัก ซึ่งปกติแล้วจะกำหนดค่า a ดังนี้

$a = 8T \pm 3$ กรณีใช้คอมพิวเตอร์ฐานสอง

$a = 200T \pm Q$ กรณีใช้คอมพิวเตอร์ฐานสิบ

เมื่อ T คือ จำนวนใด ๆ ที่มากกว่า 0

เมื่อ Q คือ ค่าใดค่าหนึ่งดังต่อไปนี้ ±(3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 77, 83 หรือ 91)

(3) คูณผลคูณในข้อ (2) ด้วย 1/m โดยปกติค่า m จะหาได้จาก

$m = 2^b$ กรณีใช้คอมพิวเตอร์ฐานสอง

$a = 10^d$ กรณีใช้คอมพิวเตอร์ฐานสิบ

เมื่อ b คือ จำนวนบิท (bits) ใน 1 คำ (Word)

d คือ จำนวนหลักใน 1 คำ (Word)

(4) เลือกใช้ตัวเลขหลังทศนิยมของเลขที่ได้ในข้อ(3)เป็นตัวเลขสุ่มที่ต้องการ

(5) ตัดจุดทศนิยมออกจากตัวเลขในข้อ (4) แล้วดำเนินการตามขั้นตอนที่ (2)

(6) ทำซ้ำในขั้นที่ (2) และ(5)จนกว่าจะได้จำนวนตัวเลขสุ่มมากเท่าที่ต้องการ

6.2.2 คุณสมบัติของตัวเลขสุ่มที่ดี

คุณสมบัติที่ใช้ในการพิจารณาว่าการสร้างตัวเลขสุ่มนั้น เหมาะสมหรือดีเพียงใดมีดังนี้

6.2.2.1 ตัวเลขที่ได้ต้องมีลักษณะการแจกแจงของความน่าจะเป็นแบบสม่ำเสมอ

6.2.2.2 ตัวเลขที่ได้ต้องเป็นอิสระต่อกัน

6.2.2.3 อนุกรมของตัวเลขที่ได้ต้องสามารถสร้างซ้ำเดิมได้ (Reproducible)

6.2.2.4 ขนาดความยาวของอนุกรมตัวเลขต้องยาวเพียงพอสำหรับการใช้งาน

6.2.2.5 ต้องใช้เวลาสั้น ๆ ในการสร้างตัวเลขแบบสุ่ม

6.2.2.6 ต้องใช้หน่วยความจำคอมพิวเตอร์น้อย

ในปัจจุบันมีการใช้วิธีการอีกหลายอย่างในการสร้างตัวเลขสุ่มเทียม รวมทั้งมีโปรแกรมสำเร็จรูปให้เรียกใช้งานจากคอมพิวเตอร์ได้เลยโดยผู้ใช้ไม่ต้องสร้างโปรแกรมเอง

6.2.3 การนำตัวเลขสุ่มมาประยุกต์ใช้กับการแก้ปัญหาต่าง ๆ เป็นขั้นตอนของการนำตัวเลขสุ่ม ไปสร้างตัวแปรตามลักษณะการแจกแจงของปัญหาที่จะศึกษา เพื่อเป็นข้อมูลของปัญหานั้น เช่น สร้างตัวเลขสุ่มขึ้นมาจากจำนวนหนึ่ง แล้วนำตัวเลขสุ่มนั้นไปสร้างเป็นคะแนนของตัวแปรของปัญหาที่จะศึกษา

6.2.4 ทำการทดลองซ้ำ ๆ หลาย ๆ ครั้ง เพื่อลดความคลาดเคลื่อนของปัญหาที่ศึกษานั้นหรือสรุปเป็นความน่าจะเป็นของการเกิดเหตุการณ์ในปัญหานั้น

6.3 จุดเด่นของการใช้วิธีมอนติคาร์โล

เนื่อง จากวิธีมอนติคาร์โล ใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้างตัวแปรของปัญหา โดยอาศัยโมเดล ทฤษฎี สูตร หรือกฎเกณฑ์ต่าง ๆที่มีอยู่ และมีการทดลองซ้ำหลายครั้งเพื่อลดความคลาดเคลื่อนต่าง ๆซึ่งนับว่ามีประโยชน์ที่สำคัญดังนี้

6.3.1 วิธีมอนติคาร์โล สามารถควบคุมตัวแปรแทรกซ้อนและสามารถสังเกตได้อย่างสมบูรณ์ นอกจากนี้ยังสามารถทำการทดลองซ้ำภายใต้สภาพแวดล้อมเดิมหลาย ๆครั้งได้ ส่วนในการปฏิบัติจริงนั้นทำไม่ได้ เพราะไม่สามารถรักษาสภาพแวดล้อมให้เหมือนเดิมทุกอย่างได้ เมื่อเวลาได้เปลี่ยนไป

6.3.2 วิธีมอนติคาร์โล ถ้ามีโมเดล ทฤษฎี สูตรหรือกฎเกณฑ์ต่าง ๆที่ถูกต้องรองรับในการสร้างตัวแปรของปัญหาในการศึกษาแล้ว ผลที่ได้จะถูกต้องแม่นยำกว่าเมื่อในศึกษาในสถานการณ์จริง ทั้งนี้เพราะสามารถลดตัวแปรแทรกซ้อนที่เป็นเชิงจิตวิทยาได้

6.3.3 ทำให้สิ้นเปลืองเวลา แรงงาน และค่าใช้จ่ายน้อยกว่า ตลอดทั้งลดการเสี่ยงก่อนการกระทำจริง เมื่อเทียบกับการศึกษาในสถานการณ์จริง

6.4 การวิจัยเชิงจำลอง (Simulation Research)

Creno และ Brewer (1973 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2538) กล่าวว่าการวิจัยเชิงจำลองแบ่งออกเป็น 3 กลุ่ม ตามระยะเวลาการพัฒนา กลุ่มแรก เป็นการวิจัยที่มีการจำลองเลียนแบบสภาพการณ์จริง ตามแนวจิตวิทยาในรูปของการเล่นเกมบทบาทสมมติ (Manned Simulated Role Playing Game Research) ในระยะต่อมาจึงมีการพัฒนาการวิจัยเชิงจำลองในแนวรัฐศาสตร์ เพื่อตรวจสอบทฤษฎีเกี่ยวกับความสัมพันธ์ระหว่างประเทศ และระบบต่าง ๆ ในสังคม เรียกว่า Simulation Research of International Relations ซึ่งเป็นงานวิจัยในกลุ่มที่สอง กลุ่มที่สามเป็นการวิจัยเชิงจำลองที่ใช้คอมพิวเตอร์เพื่อการศึกษาระบบงาน หรือระบบข้อมูลที่มีลักษณะแตกต่างกันตามข้อตกลงเบื้องต้นในการวิเคราะห์ข้อมูล งานวิจัยในกลุ่มนี้เป็นงานวิจัยด้านคณิตศาสตร์ สถิติ และการวิจัยปฏิบัติการ ในเรื่องการรอคอย (Queing) เท่าที่ผ่านมาการวิจัยในประเทศไทย มีการใช้วิธีวิทยาการวิจัยเชิงจำลองในการศึกษาเปรียบเทียบเทคนิควิธีการวิเคราะห์แบบต่าง ๆ เพื่อให้ได้วิธีการที่มีประสิทธิภาพสูงสุด เทคนิคที่ใช้คือ มอนติคาร์โล และนักวิจัยต้องเขียนโปรแกรมเพื่อสร้างข้อมูลจำลองเองแต่ในปัจจุบันมีโปรแกรมคอมพิวเตอร์ที่ช่วยให้นักวิจัยสร้างหรือจำลองข้อมูลได้สะดวก

7. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับสถิติที่มีความแกร่ง มีการศึกษาไว้อย่างแพร่หลายแตกต่างกันไปตามแต่ละชนิดของสถิติทดสอบ ซึ่งในส่วนนี้จะนำเสนอเฉพาะผลการวิจัยที่เกี่ยวข้องกับความแกร่งของสัมประสิทธิ์สหสัมพันธ์ที่ใช้ในการศึกษาครั้งนี้

Devlin et al. (1975) ได้ศึกษาความแกร่งในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์เมื่อพบค่าผิดปกติจากกลุ่มในข้อมูลที่เกิดโดยความบังเอิญจากการสุ่ม โดยใช้วิธีการวัดอิทธิพลค่าผิดปกติจากกลุ่ม พบว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เป็นสถิติที่ใช้ทดสอบสัมประสิทธิ์สหสัมพันธ์ที่ได้รับความนิยมมากที่สุด แต่ถ้ามีค่าผิดปกติจากกลุ่มในข้อมูลเพียงค่าเดียวจะมีอิทธิพลต่อการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ และต้องระวัดระวังในการแปลผลการวิเคราะห์ ซึ่งผลกระทบเหล่านี้เกิดขึ้นน้อยในการทดสอบความสัมพันธ์โดยใช้สถิติแบบไม่อิงพารามิเตอร์

จารุณี ไชยมูล (2542) ได้ศึกษาความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบทางสถิติของค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและแบบสเปียร์แมนเมื่อใช้กลุ่มตัวอย่างขนาดเล็ก คือ 10, 15,20และ 25 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.00,0.50 และ 0.90 ใช้วิธีการวิจัยแบบจำลองข้อมูลด้วยเทคนิคมอนติ คาร์โล ให้ประชากรมีการแจกแจงแบบต่าง ๆ คือเบ้ซ้าย ปกติ และเบ้ขวา ผลการวิจัยพบว่า

(1) ความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 เมื่อสัมประสิทธิ์สหสัมพันธ์ประชากรเท่ากับ 0.00 และกำหนดค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนที่ระบุ α เท่ากับ 0.05 และ 0.01 สถิติทดสอบสัมประสิทธิ์สหสัมพันธ์ทั้ง 2 แบบสามารถควบคุมความคลาดเคลื่อนได้เท่ากับความคลาดเคลื่อนที่ระบุทั้งในระดับ 0.05 และ 0.01ทุกลักษณะการแจกแจงและทุกขนาดกลุ่มตัวอย่าง

(2) อำนาจการทดสอบทางสถิติค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและแบบสเปียร์แมนเมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0.50 และกำหนดค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนที่ระบุ α เท่ากับ 0.05 และ 0.01 พบว่าสถิติทดสอบสัมประสิทธิ์สหสัมพันธ์ทั้ง 2 แบบมีอำนาจการทดสอบที่ไม่แตกต่างกัน และเมื่อกำหนดสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0.90 กำหนดค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนที่ระบุ α เท่ากับ 0.05 สถิติทดสอบสัมประสิทธิ์สหสัมพันธ์ทั้ง 2 แบบมีอำนาจการทดสอบเท่ากับ 1.00 เมื่อขนาดกลุ่มตัวอย่างเท่ากับ 25 ในทุกลักษณะการแจกแจง

Shevlyakov และ Vilchevski (2001) ได้ศึกษาถึงความแกร่งในการประมาณค่าสหสัมพันธ์และการเลือกใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ โดยสังเคราะห์จากบทความวิชาการที่เกี่ยวข้องกับการการประมาณค่าสหสัมพันธ์ พบว่า การใช้สถิติวัดสหสัมพันธ์แบบไม่อิงพารามิเตอร์เช่น สหสัมพันธ์แบบสเปียร์แมน สหสัมพันธ์แบบเคนดอลล์ และสหสัมพันธ์แบบ Qua drant มีการนำไปประยุกต์ใช้ในทางวิทยาศาสตร์ประยุกต์อย่างแพร่หลาย สหสัมพันธ์แบบสเปียร์แมนเป็นสถิติเป็นมาตรฐานในการคำนวณความสัมพันธ์ของตัวแปรที่มีพื้นฐานจากการจัดอันดับ และการกำหนดเครื่องหมายของตัวอย่างและเป็นสถิติที่มีคุณสมบัติด้านความแกร่ง แต่ยังไม่มีข้อสรุปถึงวิธีการวัดความแกร่งที่สมบูรณ์ ในการศึกษาครั้งนี้ Shevlyakov และ Vilchevski ได้เสนอสูตรในการคำนวณสัมประสิทธิ์สหสัมพันธ์แบบ Quadrant ($\hat{r}_Q$) และเคนดอลล์ ($\hat{r}_k$) คือ

$$\hat{r}_Q = \frac{1}{n}\sum_{i=1}^{n}\operatorname{sign}\{(x_i - \operatorname{median}_j(x_j))(y_i - \operatorname{median}_j(y_j))\}$$

$$\hat{r}_k = \frac{2}{n(n-1)}\sum_{i<j}\operatorname{sign}((x_i - x_j)(y_i - y_j))$$

Croux และ Dehon (2005) ได้พัฒนาวิธีการวัดความแกร่งโดยใช้ค่าเฉลี่ยของ Influence Function (IF) ต่อจาก Grize (1978) ซึ่งได้เสนอวิธีการทดสอบแบบ IF เป็นครั้งแรก โดย Croux และ Dehon ได้เปรียบเทียบความแกร่งในการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบไม่อิงพารามิเตอร์กับตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้กับตัวแปรที่มีลักษณะการแจกแจงแบบปกติสองตัวแปร ด้วยวิธีการจำลองข้อมูลให้มีการปลอมปนด้วยค่าผิดปกติจากกลุ่มในระดับสูง พบว่าสหสัมพันธ์แบบสเปียร์แมน และสหสัมพันธ์แบบเคนดอลล์ มีความแกร่งและมีประสิทธิภาพสูง

นอกจากการทดสอบความแกร่งของสถิติวิเคราะห์ความสัมพันธ์ที่เป็นพื้นฐานแล้ว นักสถิติหลายท่านได้พยามยามศึกษาหาวิธีแก้ไขปัญหาผลกระทบในการประมาณค่าความสัมพันธ์ กรณีที่มีค่าผิดปกติจากกลุ่มในตัวอย่าง โดย Mosteller และ Tukey (1977) ได้เสนอวิธีการแปลงข้อมูลที่จะวิเคราะห์ความสัมพันธ์โดยวิธีถ่วงน้ำหนัก ซึ่งใช้หลักการเดียวกับการวิเคราะห์สัมประสิทธิ์การถดถอยอย่างง่าย เรียกสถิติทดสอบนี้ว่า สหสัมพันธ์แบบถ่วงน้ำหนัก (Biweight Midcorrelation) ซึ่งพบว่าเป็นสถิติวิเคราะห์ความสัมพันธ์ที่มีความแกร่งในการทดสอบดีที่สุด

จากผลการศึกษาข้างต้น สรุปได้ว่า การศึกษาเกี่ยวกับความแกร่งของสัมประสิทธิ์สหสัมพันธ์ที่ศึกษาในข้อมูลแบบต่อเนื่อง ให้ผลการศึกษาที่สอดคล้องกันคือสถิติที่ไม่อิงพารามิเตอร์จะมีความแกร่งกว่าสถิติที่อิงพารามิเตอร์และสามารถลดอิทธิพลจากค่าผิดปกติจากกลุ่มได้ นอกจากการทดสอบกับข้อมูลที่เป็นแบบต่อเนื่องแล้ว การทดสอบโดยใช้สถิติสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนและเคนดอลล์เมื่อมีข้อมูลมีค่าซ้ำกันมากสามารถหาความสัมพันธ์ในลักษณะตารางการณ์จร ซึ่งมีการศึกษาถึงประสิทธิภาพและความแกร่งของสหสัมพันธ์แบบสเปียร์แมนและเคนดอลล์ ดังนี้

วีณา เตชะพนาดร (2529) ได้ศึกษาความสัมพันธ์ระหว่างสัมประสิทธิ์ C และสัมประสิทธิ์ของเครมเมอร์กับค่าทดสอบไคสแควร์ โดยพิจารณาถึงอิทธิพลของขนาดกลุ่มตัวอย่างขนาดของตารางและการจัดแบ่งกลุ่มข้อมูล ซึ่งศึกษาจากข้อมูลที่มีการแจกแจงแบบปกติสองตัวแปร เมื่อขนาดของกลุ่มตัวอย่างเป็น 20, 30, 40, 50, 75 และ 100 ขนาดของตารางการณ์จร ดังนี้ 2x2, 2x3, 2x4, 2x5, 3x3, 3x4, 3x5, 4x4, 4x5 และ 5x5 และตัวแปรมีสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ 0.1 ถึง 0.9 ผลปรากฏว่า ขนาดตัวอย่างและขนาดตารางมีผลทำให้ระดับนัยสำคัญจากการจำลอง สูงกว่าระดับนัยสำคัญทางทฤษฎี

พนิตตา แก้วกูร (2539)ได้ศึกษาการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบทางสถิติสำหรับค่าสหสัมพันธ์ แบบกู๊ด-ครัสคัล แบบสเปียร์แมน และแบบเคนดอลล์เทา โดยเทคนิคมอนติคาร์โล โดยพิจารณาถึงอิทธิพลของขนาดกลุ่มตัวอย่าง ขนาดของตาราง ซึ่งศึกษาจากข้อมูลการแจกแจงปกติสองตัวแปร เมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากรที่มีค่าเท่ากับ 0.0,0.5 และ 0.9 ผลสรุปคือ สถิติทดสอบสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนสามารถควบคุมความคลาดเคลื่อนได้เท่ากับความคลาดเคลื่อนที่ระบุทั้งในระดับ 0.05 และ 0.01 ที่ทุกขนาดของกลุ่มตัวอย่างและทุกขนาดของตารางการณ์จร ส่วนสถิติทดสอบสัมประสิทธิ์สหสัมพันธ์แบบกู๊ดแมน – ครัสคัล สามารถควบคุมความคลาดเคลื่อนได้เท่ากับความคลาดเคลื่อนที่ระบุในระดับ .05 เมื่อขนาดของกลุ่มตัวอย่างเท่ากับ 100 และ 200 ส่วนสถิติทดสอบสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลเทาไม่สามารถควบคุมความคลาดเคลื่อนได้เท่ากับความคลาดเคลื่อนที่ระบุ

ทองดี แย้มสรวล (2530) ได้ศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 อำนาจการทดสอบของสถิติทดสอบ ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลล์ และเครมเมอร์วี โดยใช้เทคนิคมอนติคาร์โล เมื่อกลุ่มตัวอย่างมีขนาดใหญ่คือ 150, 200 และ 250 โดยสัมประสิทธิ์สหสัมพันธ์ของประชากรมีค่าเท่ากับ 0.0 ถึง 0.9 เมื่อข้อมูลมีลักษณะเรียงอันดับในรูปตารางการณ์จร 5x5 ผลสรุปได้ดังนี้คือสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรได้ใกล้เคียงมากที่สุด

สถิติของค่าสัมประสิทธิ์สหสัมพันธ์ทั้งสาม สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้เท่าเทียมกัน ส่วนอำนาจการทดสอบของสถิติทดสอบ สถิติแบบสเปียร์แมน อำนาจการทดสอบสูงสุด

คาว และคณะ(Chow et al., 1974) ได้ศึกษาเปรียบเทียบคุณสมบัติบางประการของค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน และแบบเคนดอลล์ โดยวิธีมอนติคาร์โล เมื่อกลุ่มตัวอย่างขนาดเล็ก คือ 5, 10, 15 และ 20 และสัมประสิทธิ์สหสัมพันธ์ของประชากร (ρ) เท่ากับ 0.00, 0.10, 0.20,..., 0.90 ผลปรากฏว่า อำนาจการทดสอบทางสถิติของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนมีค่ามากกว่าแบบเคนดอลล์ ทั้งที่ระดับ และ α = 0.05 และ 0.01

การศึกษาในกรณีที่คำนวณความสัมพันธ์แบบตารางการณ์จรพบว่า ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน มีอำนาจในการทดสอบดีกว่าแบบเคนดอลล์ ซึ่งให้ผลการศึกษาที่สอดคล้องกัน ส่วนความสามารถในควบคุมความคลาดเคลื่อนประเภทที่ 1 พบว่าตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้งสองมีความสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ใกล้เคียงกัน

จากงานวิจัยที่เกี่ยวข้อง พบว่า การวัดความแกร่งของสัมประสิทธิ์สหสัมพันธ์ มีการวัดหลายวิธี ส่วนใหญ่ใช้วิธีการวัดอิทธิพลของค่าผิดปกติจากกลุ่ม การวัดความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบ โดยกำหนดสถานการณ์ในการศึกษาที่แตกต่างกันออกไป ในการศึกษาครั้งนี้ผู้วิจัยจึงสนใจที่จะศึกษาความแกร่งของสัมประสิทธิ์สหสัมพันธ์เมื่อพบมีข้อมูลผิดปกติในจำนวนที่แตกต่างกันในตัวแปร X และ Y โดยกำหนดค่าสหสัมพันธ์ของประชากรให้มีระดับความสัมพันธ์หลายระดับในทิศทางบวก และใช้เกณฑ์การเปรียบเทียบความแกร่งคือวัดความแกร่งในการทดสอบทางสถิติ และวัดความแกร่งในการประมาณค่าโดยตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่ง ควรมีคุณสมบัติทั้งความแกร่งในการทดสอบทางสถิติและความแกร่งในการประมาณค่า