บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวิจัยเป็นกระบวนการการศึกษาค้นคว้าอย่างมีระบบแบบแผนและมีจุดมุ่งหมายที่แน่นอน โดยใช้ระเบียบวิธีการทางวิทยาศาสตร์ในการรวบรวมและวิเคราะห์ข้อมูล เพื่อให้ได้มาซึ่งข้อความรู้ ความจริงที่เชื่อถือได้ ความรู้ที่ได้จากการวิจัยสามารถนำไปใช้ในการพัฒนาทางด้านวิชาการสาขาต่าง ๆ การวางแผนกำหนดนโยบายได้อย่างมีประสิทธิภาพ และเป็นพื้นฐานการพัฒนาให้เกิดสิ่งประดิษฐ์ใหม่ ๆ ขึ้นมา ขั้นตอนในการดำเนินงานวิจัยโดยทั่วไปจะประกอบด้วย การกำหนดปัญหา การตั้งสมมติฐาน การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล และการสรุปผล ขั้นตอนที่สำคัญในการดำเนินงานวิจัยขั้นตอนหนึ่งคือการวิเคราะห์ข้อมูล ซึ่งในการวิเคราะห์ข้อมูลมีความจำเป็นอย่างยิ่งที่จะต้องเลือกใช้สถิติให้เหมาะสมอันจะส่งผลโดยตรงต่อความถูกต้องและความเชื่อถือได้ในข้อความรู้ที่ได้ของงานวิจัยนั้น ๆ

งานวิจัยทางสังคมศาสตร์เป็นจำนวนมากที่การวิเคราะห์ข้อมูลมีจุดมุ่งหมายเพื่อต้องการทราบความสัมพันธ์ระหว่างตัวแปรสองตัวซึ่งอยู่ในประชากรเดียวกันว่ามีความสัมพันธ์กันในลักษณะใด งานการวิเคราะห์ดังกล่าวสามารถดำเนินการได้โดยวิธีการวิเคราะห์ทางสถิติ ซึ่งสามารถส่งผลสรุปได้ทั้งทิศทางและปริมาณความสัมพันธ์ระหว่างตัวแปร การหาความสัมพันธ์โดยวิธีการทางสถิติมีอยู่หลายวิธีขึ้นอยู่กับลักษณะของตัวแปร และมาตราการวัดค่าของตัวแปร

ความสัมพันธ์ระหว่างตัวแปรสองตัวจะมีค่ามากหรือน้อยนั้นจะสรุปได้จากการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งคำนวณได้จากข้อมูลของตัวแปรทั้งสอง สัญลักษณ์ที่ใช้แทนค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรคือ ρ (rho) ถ้า ρ มีค่าใกล้ ± 1 แสดงว่าตัวแปรที่ศึกษาทั้งสองตัวนั้นมีความสัมพันธ์กันมาก

ถ้า ρ มีค่าใกล้ 0 แสดงว่าตัวแปรที่ศึกษาทั้งสองตัวนั้นมีความสัมพันธ์กันน้อย เครื่องหมายของ ρ นั้นเป็นการแสดงถึงทิศทางของลักษณะความสัมพันธ์ กล่าวคือถ้า ρ มีเครื่องหมายบวก แสดงว่าตัวแปรทั้งสองตัวนั้นมีความสัมพันธ์ไปในทางเดียวกัน แต่ถ้า ρ มีเครื่องหมายลบ แสดงว่าตัวแปรทั้งสองตัวนั้นมีความสัมพันธ์ในทางตรงกันข้าม ค่าสัมประสิทธิ์สหสัมพันธ์นอกจากจะเป็นค่าสถิติที่ให้ผลสรุปว่าตัวแปรคู่ใดคู่หนึ่งมีความสัมพันธ์กันหรือไม่ และในทิศทางใดแล้ว ค่ากำลังสองของค่าสัมประสิทธิ์สหสัมพันธ์ยังสามารถอธิบายความสัมพันธ์ระหว่างตัวแปรในรูปของร้อยละได้

สัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปรที่รู้จักกันดีที่สุดคือ สัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Product Moment Correlation Coefficient : r_{xy}) ซึ่งการที่จะตัดสินใจเลือกใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน จะต้องคำนึงถึงข้อตกลงเบื้องต้น (assumption) ดังนี้

1. ตัวแปรทั้งสองเป็นตัวแปรที่มีค่าต่อเนื่อง และมีการแจกแจงของประชากรเป็นแบบปกติสองตัวแปร (Bivariate Normal Distribution)

2. ความสัมพันธ์ระหว่างตัวแปรทั้งสองเป็นความสัมพันธ์เชิงเส้นตรง (Linear Relationship)

3. มีความเป็นอิสระของข้อมูล (Independence Between Pairs)

จะเห็นว่าการหาค่าสัมประสิทธิ์สหสัมพันธ์วิธีนี้ต้องเป็นไปตามข้อตกลงเบื้องต้นดังกล่าวซึ่งไม่ควรที่จะฝ่าฝืน ถ้าฝ่าฝืนข้อตกลงเบื้องต้นข้อใดข้อหนึ่งก็ไม่ควรใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ควรเลือกวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างสองตัวแปรวิธีอื่นแทน (lindeman, Merenda and Gold 1980 : 63)

งานวิจัยทางสังคมศาสตร์มีแนวโน้มที่จะศึกษาตัวแปรครั้งละหลาย ๆ ตัว ซึ่งมักจะมีลักษณะแตกต่างกัน ข้อมูลจากตัวแปรบางตัวอาจวัดค่าได้ในมาตราอันตรภาคหรือมาตราอัตราส่วน แต่ข้อมูลจากตัวแปรบางตัวจะวัดค่าได้เพียงมาตรานามบัญญัติหรือมาตราจัดอันดับเท่านั้น ในบางกรณีแม้ผู้วิจัยจะสามารถสร้างเครื่องมือในการเก็บรวบรวมข้อมูลให้มีความละเอียดได้ถึงระดับมาตราอันตรภาคหรือมาตราอัตราส่วนก็ตาม แต่เนื่องจากความจำกัดทางด้านเวลาที่ใช้ในการสร้างเครื่องมือและการดำเนินการเก็บรวบรวมข้อมูลผู้วิจัยจึงมักแก้ปัญหาด้วยการวัดค่าของตัวแปรในมาตราจัดอันดับเป็นส่วนมาก

งานวิจัยจำนวนมากที่ใช้เครื่องมือในการเก็บรวบรวมข้อมูลในรูปมาตราส่วนประเมินค่า (rating scale) ซึ่งเป็นเครื่องมือที่มีข้อคำถามหรือเรื่องราวที่ถามกำหนดไว้ให้ผู้ตอบได้พิจารณาคำตอบว่าคำตอบนั้น ๆ มีน้ำหนักหรือระดับอยู่ในเกณฑ์ใด มาตราส่วนประเมินค่าโดยมากกำหนดไว้ไม่ต่ำกว่า 3 ระดับ และไม่เกิน 11 ระดับ ส่วนใหญ่จะใช้กันเพียง 5 ระดับ (สุภาพ วาดเขียน 2525 : 34) เช่น

คำถามเกี่ยวกับความคิดเห็น :

เห็นด้วยมาก เห็นด้วย ไม่ออกความคิดเห็น ไม่เห็นด้วย ไม่เห็นด้วยมาก

คำถามเกี่ยวกับความจำเป็นในบางเรื่อง :

จำเป็นมากที่สุด จำเป็นมาก จำเป็น จำเป็นน้อย ไม่จำเป็น

คำถามเกี่ยวกับเรื่องความเหมาะสม :

เหมาะสมมากที่สุด เหมาะสมมาก เหมาะสมปานกลาง เหมาะสมน้อย เหมาะสมน้อยที่สุด

คำถามเกี่ยวกับการประเมินปริมาณ มาตราส่วนประเมินค่าได้รับการนำไปใช้โดยกำหนดตัวเลขแทนข้อความได้ เช่น มากที่สุด = 5 มาก = 4 ปานกลาง = 3 น้อย = 2 และน้อยที่สุด = 1 เป็นต้น ตัวเลขที่ต่างกันนี้แสดงให้เห็นว่าผู้ตอบให้น้ำหนักของคำตอบอยู่ในระดับเกณฑ์ใดเท่านั้น ซึ่งตัวเลขเหล่านั้นไม่สามารถบอกให้ทราบว่าคำตอบนั้นมีคุณสมบัติแตกต่างกันเป็นปริมาณเท่าใด และตัวเลขที่ต่างกัน 1 หน่วย ไม่ได้แทนปริมาณที่เท่ากันเหมือนข้อมูลในมาตราอันตรภาคหรือมาตราอัตราส่วน ด้วยเหตุดังกล่าวจะเห็นว่าข้อมูลประเภทนี้ยังมีผู้วิจัยจำนวนมากที่ไม่สามารถยอมรับว่าอยู่ในระดับมาตราอันตรภาคหรือมาตราอัตราส่วนได้

เมื่อผู้วิจัยต้องการศึกษาความสัมพันธ์ระหว่างตัวแปร โดยที่ตัวแปรทั้งสองวัดจากมาตราส่วนประเมินค่า 5 ระดับ ที่กำหนดค่าของตัวเลขแทนระดับของคำถาม ผู้วิจัยสามารถเลือกวิธีวิเคราะห์หาความสัมพันธ์ได้หลายวิธี ที่พบมากมักใช้วิธีดังต่อไปนี้

1. วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ที่ใช้ไคสแควร์เป็นพื้นฐาน เช่น ค่าสัมประสิทธิ์ฟี (ø) ค่าสัมประสิทธิ์ของเพียร์สันในรูปตารางการ์จร (Pearson's Coefficient of Contingency : C) หรือค่าสัมประสิทธิ์ของเครมเมอร์วี (V) เป็นต้น

2. วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับ (Rank Correlation Coefficient) เช่น ค่าสัมประสิทธิ์ของสเปียร์แมน หรือค่าสัมประสิทธิ์ของเคนดอลเทา เป็นต้น

ลักษณะการศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว โดยมีการวัดค่าของตัวแปรทั้งสองเป็นมาตราส่วนประเมินค่า 5 ระดับ ดังกล่าวมาแล้วนั้นตามปกติผู้วิจัยจะนำค่าที่วัดได้มาจัดอยู่ในรูปตารางการณ์จร (Contingency Table) ซึ่งเมื่อข้อมูลอยู่ในลักษณะเป็นอันดับ และจัดเป็นตารางแบบสองมิติจึงอาจเรียกว่า ตารางแยกประเภทแบบมีอันดับสองทาง (Ordered Two-Way Classification) 5 X 5 ซึ่งแถว (row) และสดมภ์ (column) มีลักษณะเป็นอันดับ ดังนั้นถ้าต้องการจัดอันดับของตัวแปรทั้งสองในรูปตารางการณ์จร 5 X 5 กับกลุ่มตัวอย่างขนาดใหญ่จะจำแนกอันดับได้ยาก ทั้งนี้เพราะมีการซ้ำของอันดับเกิดขึ้นเป็นจำนวนมาก สถานการณ์เช่นนี้จึงเป็นปัญหาในการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับ เช่น สัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เพราะโดยปกติแล้วจะมีข้อตกลงว่าไม่มีค่าซ้ำ ดังนั้นถ้ามีค่าซ้ำเกิดขึ้นจะต้องใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับที่มีค่าแก้ โดยอาศัยการเฉลี่ยค่าที่ซ้ำ (average rank) (Yule and Kendall 1950 : 264-265) ดังนั้นการหาค่าสัมประสิทธิ์สหสัมพันธ์ในรูปตารางการณ์จรจึงมีผู้วิเคราะห์ที่อาศัยการเฉลี่ยค่าซ้ำช่วยในการวิเคราะห์กันได้หลายวิธี ดังได้มีนักสถิติกล่าวไว้ดังนี้

Stuart (1953 : 105) กล่าวว่า เมื่อต้องการวัดความสัมพันธ์ของค่าสังเกตระหว่างสองตัวแปรในรูปตารางการณ์จร ตำราบางเล่มได้เสนอแนะให้ใช้สัมประสิทธิ์ของคาร์ล เพียร์สัน ในรูปตารางการณ์จร (Karl Pearson's coefficient of Contingency : C) ซึ่งตามธรรมดาแล้วค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าอยู่ระหว่าง 0 ถึง 1 แต่การคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ด้วยวิธีนี้ไม่สามารถหาค่าสูงสุดดังกล่าวได้ จึงน่าจะหันมาใช้การหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอลเทา (Kendall's Rank Correlation coefficient : τ_c) ซึ่งใช้กับข้อมูลที่อยู่ในมาตราจัดอันดับ เงื่อนไขสำหรับการนำไปประยุกต์ใช้คือ ใช้กับตารางการณ์จรที่แถวและสดมภ์มีการแยกประเภท และมีลักษณะเป็นอันดับ

Marascuilo and McSweeney (1977 : 447) กล่าวว่า ผู้วิจัยจำนวนมากจะวิเคราะห์ข้อมูลที่มีลักษณะเป็นอันดับในรูปตารางการณ์จรโดยใช้สถิติเพียร์สัน ไคสแควร์ (Pearson's X^2 Statistic) และทฤษฎีสหสัมพันธ์ที่เกี่ยวข้องกับ mean - square contingency table แต่สำหรับกรณีนี้ไม่แนะนำให้ใช้ค่าสถิติดังกล่าว เนื่องจากข้อมูลมีลักษณะเป็นอันดับของการแยกประเภท การสลับลำดับของแถวหรือสดมภ์จะมีผลกระทบต่อความสัมพันธ์ของอันดับที่มีอยู่ในข้อมูล วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ที่สามารถนำมาใช้คือ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอลเทา (τ_c)

จะเห็นว่าเมื่อข้อมูลมีลักษณะเป็นอันดับในรูปตารางการณ์จร ผู้วิจัยสามารถพิจารณาเลือกใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ได้หลายวิธี โดยเฉพาะวิธีที่ใช้กันอยู่ทั่วไป วิธีหนึ่งคือ วิธี Pearson's Mean - Square Contingency Coefficient หรือ Contingency Coefficient : C แต่การหาค่าสัมประสิทธิ์สหสัมพันธ์วิธีนี้มีข้อจำกัดปรากฏอยู่ดังที่ Yule and Kendall (1950 : 53-54) ได้กล่าวสอดคล้องกับ Siegel (1956 : 201) ว่าค่าสัมประสิทธิ์ C จะมีค่าเท่ากับ 0 เมื่อตัวแปรทั้งสองไม่มีความสัมพันธ์กัน แต่ไม่มีโอกาสที่จะมีค่าสูงถึง 1 ได้เลย เมื่อจำนวนแถวและสดมภ์มีค่าเท่ากัน ค่าสูงสุดของสัมประสิทธิ์ C จะมีค่าเท่ากับ $\sqrt{(r-1)/r}$ หรืออาจเขียนได้ว่า $0 \leqslant C \leqslant \sqrt{(r-1)/r} < 1$ เมื่อ r แทนจำนวนแถว ดังเช่น ตารางการณ์จร 4 X 4 ค่าสัมประสิทธิ์ C จะมีค่าสูงสุด = 0.866 ถ้าตารางการณ์จร 5 X 5 ค่าสัมประสิทธิ์ C จะมีค่าสูงสุด = 0.894 เป็นต้น ข้อจำกัดดังกล่าวได้รับการศึกษาและสร้างดัชนีความสัมพันธ์ขึ้นมาใหม่โดยเครมเมอร์ ซึ่งมีผู้นิยมเรียกกันโดยทั่วไปคือ ค่าสัมประสิทธิ์สหสัมพันธ์เครมเมอร์วี (Cramer's V Correlation Coefficient : V) หรือค่าสัมประสิทธิ์เครมเมอร์ฟี (Cramer's Phi Coefficient : ϕ') ซึ่งได้รับคำแนะนำให้ใช้แทนค่าสัมประสิทธิ์ C (Hays 1973 : 745) (Blalock 1981 : 305) (Jacobson 1976 : 438)

การหาค่าสัมประสิทธิ์สหสัมพันธ์เมื่อข้อมูลมีลักษณะเป็นอันดับในรูปตารางการณ์จร ตามที่ Stuart และ Marascuilo and McSweeney ได้แนะนำให้ใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอล (τ_c) ซึ่งค่าสัมประสิทธิ์ τ_c ดังกล่าวนี้ Stuart ได้พัฒนาจากวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอล (τ)

เนื่องจากการจัดอันดับในทางปฏิบัติมักเกิดค่าซ้ำ จึงได้มีการพัฒนาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของเคนดอลโดยการแก้ค่าซ้ำ และเพื่อให้เห็นความแตกต่างของสูตรการคำนวณจึงได้มีการกำหนดสัญลักษณ์ให้แตกต่างกัน โดยเติมอักษรข้างท้ายด้านล่างของสัญลักษณ์ τ (Kohout 1974 : 234) กล่าวคือใช้สัญลักษณ์ τ หรือ τ_a แทนค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณจากข้อมูลที่ไม่มีการซ้ำของข้อมูลในตัวแปรชุดใดชุดหนึ่ง และใช้สัญลักษณ์ τ_b แทนค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณจากข้อมูลที่มีค่าซ้ำในตัวแปรชุดใดชุดหนึ่งหรือทั้งสองชุด สำหรับค่าสัมประสิทธิ์ τ_c นั้นได้รับการพัฒนาขึ้นเพื่อนำมาใช้ในการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรสองตัว ที่ข้อมูลมีลักษณะเป็นอันดับในรูปตารางการณ์จร ซึ่งเมื่อจัดข้อมูลลงในตารางการณ์จรจะเกิดค่าซ้ำจำนวนมาก กรณีเช่นนี้การใช้สัมประสิทธิ์ τ_b ไม่สามารถคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์สูงสุดเท่ากับ 1 ได้ (Stuart 1953 : 106) Stuart จึงแนะนำให้ใช้ค่าสัมประสิทธิ์ τ_c แทน

นอกจากนี้ยังมีวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับอีกวิธีหนึ่งที่เป็นที่รู้จักกันอย่างแพร่หลายคือ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับของสเปียร์แมน (Spearman's Rank Correlation Coefficient or Spearman's Rho ในงานวิจัยนี้จะใช้สัญลักษณ์ r แทน) ซึ่งเป็นวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ที่ดัดแปลงมาจากวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้จะมีค่าเท่ากันเมื่อใช้ข้อมูลชุดเดียวกันที่อยู่ในมาตราจัดอันดับและไม่มีค่าซ้ำ เนื่องจากวิธีคำนวณหาค่าสัมประสิทธิ์ r คำนวณง่ายกว่าวิธีของเพียร์สัน บางครั้งจึงมีผู้ใช้วิธีหาค่าสัมประสิทธิ์ r เพื่อประมาณค่าคร่าว ๆ ของค่าสัมประสิทธิ์ r_{xy} (Kohout 1974 : 213-216) การเลือกใช้สัมประสิทธิ์ r ต้องคำนึงถึงข้อตกลงเบื้องต้น ซึ่งมีข้อตกลงเบื้องต้นเช่นเดียวกับสัมประสิทธิ์ τ ดังนั้นเมื่อผู้วิจัยต้องการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับดังกล่าวผู้วิจัยจะเกิดปัญหาว่าจะเลือกใช้วิธีใดจึงจะเหมาะสมกว่ากัน ซึ่งในปัญหาดังกล่าวมีนักสถิติหลายท่านได้ให้ความคิดเห็นเกี่ยวกับสัมประสิทธิ์ r และ τ ไว้ดังต่อไปนี้

Furguson (1981 : 389) กล่าวว่า ค่าสัมประสิทธิ์ r และ τ แม้ว่าจะใช้ในจุดประสงค์เดียวกัน แต่ค่าที่ได้นั้นแตกต่างกันถึงแม้จะคำนวณจากข้อมูลชุดเดียวกันก็ตาม ค่าสมบูรณ์ของ r จะมีค่ามากกว่าค่าสมบูรณ์ของ τ ค่า r และ τ จะมีความสัมพันธ์กันสูงในกลุ่มตัวอย่างที่มาจากประชากรสองตัวแปรที่มีการแจกแจงแบบปกติ (Bivariate

Normal Population) เมื่อความสัมพันธ์ของประชากรมีค่าเป็น 0 ค่าสหสัมพันธ์เพียร์สัน ระหว่าง r กับ τ มีค่าเท่ากับ 0.980 เมื่อ n = 5 และจะมีค่าเข้าใกล้ 1 เมื่อ n มีค่าเข้าใกล้ ∞ การแจกแจงของทั้ง r และ τ เป็นแบบสมมาตรและมีแนวโน้มที่จะเป็นแบบปกติ เมื่อ n มีขนาดเพิ่มขึ้นการแจกแจงของ τ มีแนวโน้มเป็นแบบปกติได้เร็วกว่า r โดยทั่ว ๆ ไปแล้ว τ เป็นค่าสถิติที่มีความยืดหยุ่นทางด้านคณิตศาสตร์มากกว่า r และปัญหาอันเกิดจากค่าซ้ำของการจัดอันดับนั้น วิธีทั้งสองสามารถแก้ปัญหาได้

Bradley (1968 : 93-95) กล่าวว่า วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ที่สามารถนำมาใช้แทนวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน ภายใต้เงื่อนไขของการทดสอบสมมติฐานสูญ (H_0) เดียวกันได้คือ วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของเคนดอล(τ) ค่าสัมประสิทธิ์ที่ได้จากวิธีทั้งสองมีค่าไม่เท่ากัน แต่ค่า r และ τ มีความสัมพันธ์กันสูง เมื่อใช้กับกลุ่มตัวอย่างที่มาจากประชากรสองตัวแปรที่มีการแจกแจงแบบปกติ สัมประสิทธิ์สหสัมพันธ์ทั้งสองวิธีมีค่า A.R.E. (Asymptotic Relative Efficiency) เท่ากันคือ $9/\pi^2 = 0.912$ เมื่อเปรียบเทียบกับวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน และค่า A.R.E. จะมีค่าเท่ากับ 1 เมื่อเปรียบเทียบกันในแต่ละวิธี สำหรับที่ n ค่าเดียวกัน การแจกแจงค่าสถิติทดสอบค่าสัมประสิทธิ์ τ จะมีแนวโน้มเข้าใกล้ลักษณะการแจกแจงแบบปกติได้ดีกว่าการแจกแจงค่าสถิติทดสอบค่าสัมประสิทธิ์ r

Hays (1973 : 796-797) ได้กล่าวไว้สอดคล้องกับ Jacobson (1976 : 256) ว่าความสัมพันธ์แบบอันดับของสเปียร์แมนเป็นวิธีที่นิยมใช้ในการหาความสัมพันธ์ระหว่างสองตัวแปรที่มีการจัดอันดับ อย่างไรก็ตามมีเหตุผลที่เชื่อถือได้ว่าสัมประสิทธิ์ τ นั้นมีความดีกว่าในหลายๆ ประการ โดยที่ τ สามารถแปลความหมายในสถิติบรรยายได้ง่ายกว่า r นอกจากนี้ τ ยังสามารถประมาณค่าความสัมพันธ์ในประชากรได้อย่างปราศจากอคติ (unbiased estimate) และเหตุผลประการสำคัญคือ การแจกแจงของกลุ่มตัวอย่างมีลักษณะเป็นการแจกแจงแบบปกติได้รวดเร็วกว่าค่า r

จากปัญหาการเลือกใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับวิธีใดระหว่างวิธีของสเปียร์แมนและเคนดอล จึงจะเหมาะสมกว่ากันภายใต้ข้อตกลงเบื้องต้นเดียวกัน จากที่นักสถิติหลายท่านได้กล่าวไว้นั้น ก็ยังไม่สามารถระบุลงไปได้อย่างเด่นชัดว่าวิธีใดดีกว่า

และโดยเฉพาะอย่างยิ่ง ในกรณีที่ข้อมูลมีจำนวนมากจัดอยู่ในรูปตารางการณ์จร 5 X 5 ซึ่งแถวและสดมภ์มีการจัดอันดับ และมีค่าซ้ำเกิดขึ้นเป็นจำนวนมาก ไม่เพียงแต่วิธีของเคนดอลเท่านั้นที่มีการพัฒนาค่าสัมประสิทธิ์มาเป็น τ_c วิธีหาค่าสัมประสิทธิ์ r ก็ได้รับการพัฒนาสูตรขึ้นมาใช้ในการคำนวณเช่นเดียวกัน กล่าวคือ ถ้าข้อมูลมีค่าซ้ำเกิดขึ้น ก็จะใช้วิธีการคำนวณที่มีการแก้ค่าซ้ำ ค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนที่คำนวณตามเงื่อนไขนี้จะใช้สัญลักษณ์ r_b และสำหรับสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมนที่ใช้คำนวณกับข้อมูลในลักษณะเดียวกับ τ_c จะใช้สัญลักษณ์ r_c (Stuart 1963 : 24) ซึ่งแม้แต่จะมีการพัฒนาเทคนิคมาใช้ใหม่ทั้งสองแบบแล้วก็ตาม แต่ผลสรุปการวิจัยในการเปรียบเทียบผลและการนำมาใช้ก็ยังเป็นที่ต้องการอย่างมาก

จากผลการศึกษาและข้อเสนอแนะต่าง ๆ พบว่า ยังไม่สามารถหาข้อสรุปที่มีหลักฐานชัดเจนเกี่ยวกับการเลือกใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ เมื่อข้อมูลของตัวแปรทั้งสองมีลักษณะเป็นอันดับในรูปตารางการณ์จร 5 X 5 ว่า วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลเทา และเครมเมอร์วี ค่าสัมประสิทธิ์สหสัมพันธ์จากวิธีใดจะเหมาะสมกว่ากัน เป็นผลสรุปที่จะมีคุณค่าอย่างยิ่งทั้งด้านทฤษฎีและปฏิบัติ ด้วยเหตุนี้ผู้วิจัยจึงมีความสนใจที่จะทำการศึกษาลักษณะการแจกแจงค่าสัมประสิทธิ์สหสัมพันธ์ที่ปรากฏทั้ง 3 วิธีว่ามีความสอดคล้องกับทฤษฎีอย่างไร นอกจากนี้ยังสนใจศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 (Type I Error) ตามที่ระบุ ซึ่งเป็นการพิจารณาอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I Error Rate) ที่เกิดขึ้นจริงในการทดสอบสมมติฐาน และศึกษาอำนาจการทดสอบของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละวิธี โดยวิธีการซิมูเลชั่น (simulation) อันจะทำให้ได้ผลสรุปที่เด่นชัดภายใต้สภาวะการณ์การทดลอง คือสามารถกำหนดลักษณะการแจกแจงของประชากร ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรและขนาดของกลุ่มตัวอย่างได้ด้วยเครื่องคอมพิวเตอร์ โดยใช้เทคนิคที่เรียกว่า มอนติคาร์โล (Monte Carlo)

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาลักษณะการแจกแจงค่าสัมประสิทธิ์สหสัมพันธ์ที่ปรากฏของค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลเทา และเครมเมอร์วี เมื่อสัมประสิทธิ์สหสัมพันธ์ของประชากร มีค่าเท่ากับ 0.0,0.1,...,0.9

2. เพื่อศึกษาความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 ของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลเทา และเครมเมอร์วี ที่สัมประสิทธิ์สหสัมพันธ์ของประชากร มีค่าเท่ากับ 0.0

3. เพื่อศึกษาอำนาจการทดสอบของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลเทา และเครมเมอร์วี ที่สัมประสิทธิ์สหสัมพันธ์ของประชากร มีค่าเท่ากับ 0.1,0.2,...,0.9

สมมติฐานของการวิจัย

เมื่อการแจกแจงของประชากรเป็นแบบปกติสองตัวแปร และข้อมูลมีลักษณะเป็นอันดับในรูปตารางการณ์จร 5 X 5 ผู้วิจัยได้ตั้งสมมติฐานของการวิจัยไว้ดังนี้

1. สถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน เคนดอลเทา และเครมเมอร์วี สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ไม่แตกต่างกัน

2. สถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน และเคนดอลเทา มีอำนาจการทดสอบไม่แตกต่างกัน แต่สถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของสเปียร์แมน และเคนดอลเทา มีอำนาจการทดสอบแตกต่างจากสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของเครมเมอร์วี

ข้อตกลงเบื้องต้น

1. การวิจัยนี้เป็นการศึกษาวิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งจะทำการศึกษาเฉพาะสหสัมพันธ์เชิงเส้นตรง (Linear Relationship)

2. การวิจัยครั้งนี้ถือว่าค่าสถิติแสดงการแจกแจงค่าสัมประสิทธิ์สหสัมพันธ์ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ เป็นดัชนีสำคัญที่ผู้วิจัยใช้เป็นเกณฑ์ในการเลือกใช้วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัยเป็นประชากรที่สร้างจากเครื่องคอมพิวเตอร์ โดยการเขียนโปรแกรมภาษาฟอร์แทรน 77 (FORTRAN 77) หรือ วีเอส ฟอร์แทรน (VS FORTRAN) ซึ่งใช้กับเครื่องคอมพิวเตอร์ IBM 370/3031 ในระบบ OS/VS1 ของสถาบันบริการคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย เพื่อสร้างประชากรที่มีลักษณะการแจกแจงแบบปกติสองตัวแปร และมีค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรแต่ละคู่ในประชากร ρ = 0.0,0.1,...,0.9 จำนวน 10 ชุด ตามลำดับ

2. ตัวแปรที่ศึกษาในการวิจัยนี้ มีดังต่อไปนี้

2.1 ตัวแปรตาม (Dependent Variables) ได้แก่

ก. ลักษณะการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ที่ปรากฏ

ข. อัตราความคลาดเคลื่อนประเภทที่ 1 จากสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์

ค. อำนาจการทดสอบของสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์

2.2 ตัวแปรอิสระ (Independent Variables) ได้แก่

ก. วิธีหาค่าสัมประสิทธิ์สหสัมพันธ์ ได้แก่ วิธีของสเปียร์แมน วิธีของเคนดอลเทา และวิธีของเครมเมอร์วี

ข. ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรที่ใช้ศึกษา คือ ρ = 0.0,0.1,...,0.9 หรือมีจำนวนประชากรทั้งสิ้น 10 ชุด

3. อัตราความคลาดเคลื่อนประเภทที่ 1 จากสถิติทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของการวิจัยนี้ พิจารณาที่สมมติฐาน $H_o : \rho = 0$ เป็นจริง

4. การวิจัยครั้งนี้ศึกษาจากประชากรที่มีลักษณะการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal Distribution) เท่านั้น

5. การวิจัยครั้งนี้ศึกษาเฉพาะกลุ่มตัวอย่างที่มีขนาดใหญ่ โดยกำหนดกลุ่มตัวอย่างที่ศึกษามีขนาด 150,200 และ 250

6. การวิจัยนี้กำหนดอัตราความคลาดเคลื่อนที่ระบุ (α) 2 ระดับ คือ ที่ α = .05 และ α = .01

7. แต่ละกรณีของการวิจัยครั้งนี้จะทำการทดลองซ้ำจำนวน 4,000 ครั้ง

คำจำกัดความ

ค่าสัมประสิทธิ์สหสัมพันธ์ หมายถึง ระดับความสัมพันธ์ระหว่างตัวแปรสองตัว ซึ่งกลุ่มตัวอย่างสุ่มมาจากประชากรกลุ่มเดียวกัน

การแจกแจงค่าสัมประสิทธิ์สหสัมพันธ์ หมายถึง การแจกแจงสุ่ม (Sampling Distribution) ของค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้จากการทดลอง

ความคลาดเคลื่อนประเภทที่ 1 (Type I Error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐานสูญ เมื่อสมมติฐานสูญเป็นจริง โอกาสที่จะทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 กำหนดด้วย α

ความคลาดเคลื่อนประเภทที่ 2 (Type II Error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการยอมรับสมมติฐานสูญ เมื่อสมมติฐานสูญนั้นผิด โอกาสที่จะทำให้เกิดความคลาดเคลื่อนประเภทที่ 2 กำหนดด้วย β

อำนาจการทดสอบ (Power of Test) หมายถึง ความน่าจะเป็นที่ปฏิเสธสมมติฐานสูญ เมื่อสมมติฐานนั้นผิด ซึ่งมีค่าเท่ากับ $1 - \beta$

อัตราความคลาดเคลื่อนประเภทที่ 1 จากผลการทดลอง หมายถึง สัดส่วนความคลาดเคลื่อนประเภทที่ 1 ที่เกิดขึ้นจริงในการทดสอบสมมติฐานของการทดลอง

อัตราความคลาดเคลื่อนที่ระบุ หมายถึง อัตราความคลาดเคลื่อนประเภทที่ 1 ที่ควบคุมด้วยระดับนัยสำคัญ α ซึ่งผู้วิจัยเป็นผู้กำหนดขึ้น

การซ้ำของอันดับ หมายถึง ค่าที่ได้จากการวัดในมาตราจัดอันดับที่เรียงลำดับจากค่าต่ำสุดไปหาค่าสูงสุด หรือจากค่าสูงสุดไปหาค่าต่ำสุด แต่มีค่าสังเกตบางค่าเท่ากัน

ตารางการณ์จรที่มีลักษณะเป็นอันดับ หมายถึง ตารางที่แต่ละช่องสามารถบรรจุจำนวนหรือความถี่ของคะแนนหรือค่าสังเกตของตัวแปรทั้งสอง โดยที่ตัวแปรทั้งสองอยู่ในมาตราจัดอันดับ ดังตัวอย่างแผนภาพที่ 1

แผนภาพที่ 1 ตารางการณ์จรที่มีการจัดอันดับ 5×5 จากมาตราส่วนประเมินค่า 5 ระดับ

ตัวแปรที่ 1	ตัวแปรที่ 2					
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)	รวม
มากที่สุด (5)						
มาก (4)						
ปานกลาง (3)						
น้อย (2)						
น้อยที่สุด (1)						
รวม						

ประโยชน์ของการวิจัย

เพื่อช่วยให้ผู้ใช้สถิติสามารถเลือกใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์เมื่อข้อมูลมีลักษณะเป็นอันดับในรูปตารางการณ์จรได้อย่างเหมาะสมตามสถานการณ์ อันจะทำให้ผลสรุปหรือข้อความรู้ที่ได้จากงานวิจัยนั้น ๆ มีความเชื่อถือได้สูง