



ประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์
เมื่อข้อมูลแจกแจงแบบไม่ปกติ

Efficiency of Pearson, Spearman and Kendall's Correlation Coefficients When Data is
Non-normal Distributed

ชนิรัตน์ พุกษ์ประมool *

Chaninan Pruekpramool *

นัฐริกา เจริญตะคุ **

Nattika Jaroentaku **

สิวะโชติ ศรีสุทธิยากร ***

Siwachoat Srisuttiyakorn ***

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการประมาณและการทดสอบสมมติฐานของการวิเคราะห์ด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ โดยใช้ข้อมูลจำลองด้วยเทคนิคมอนติคาร์โลจากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอล ทั้งสิ้น 288 สถานการณ์ กระทำซ้ำสถานการณ์ละ 1,000 รอบ พิจารณาประสิทธิภาพด้านการประมาณจากค่าความเอนเอียงสัมพัทธ์ (RB) และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (MCSE) ประสิทธิภาพด้านการทดสอบสมมติฐานจากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่งและอำนาจการทดสอบสมมติฐานผลการวิจัยพบว่า (1) สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์ต่ำที่สุด ในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์มีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์สูงที่สุด (2) สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมนมีแนวโน้มให้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลใกล้เคียงกันในกรณีที่ข้อมูลไม่สัมพันธ์กัน และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ให้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลต่ำที่สุด (3) สัมประสิทธิ์สหสัมพันธ์ทั้งสามมีแนวโน้มควบคุมความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่ .05 และ (4) สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้อำนาจการทดสอบสูงที่สุด

คำสำคัญ: การแจกแจงไม่ปกติ, ประสิทธิภาพ, สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน, สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน, สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์

*นิสิตมหาบัณฑิตสาขาวิชาสถิติและสารสนเทศการศึกษ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

*Graduate Student of Educational Statistics and Information Division, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University. E-mail Address: chaninan_p@hotmail.com

**นิสิตมหาบัณฑิตสาขาวิชาสถิติและสารสนเทศการศึกษ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

**Graduate Student of Educational Statistics and Information Division, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University. E-mail Address: pampam_pkpk@hotmail.com

**อาจารย์ประจำสาขาวิชาสถิติและสารสนเทศการศึกษ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

**Lecturer of Educational Statistics and Information Division, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University. E-mail Address: choat.cu@gmail.com

Abstract

This research aimed to compare the efficiency in estimation and hypothesis testing of correlation analysis among Pearson, Spearman and Kendall's correlation coefficients when the data were non-normal distributed. The Monte Carlo data simulation was used in this study to generate the data from a log-normal probability distribution. The total numbers of simulations were 288 situations repeating 1,000 rounds in each situation. The efficiency in estimation was considered from Relative Bias (RB) and Monte Carlo Standard Errors (MCSE). The efficiency in hypothesis testing was considered from the probability of a type I error and power of testing. The results revealed that (1) Pearson's correlation coefficients had tended to present the estimating values with the lowest relative bias while Kendall's correlation coefficients had tended to present the estimating values with the highest relative bias. (2) Pearson's and Spearman's correlation coefficients had tended to present the values of Monte Carlo standard errors close to each other when there was no relationship between two sets of the data and Kendall's correlation coefficients had tended to have the lowest values of the Monte Carlo standard errors. (3) The significance testing of these three correlation coefficients had tended to be able to control the type I error under .05 significant level, and (4) Pearson's correlation coefficients had tended to present the highest power of testing.

Keywords: non-normal distribution, efficiency, Pearson's correlation coefficient, Spearman's correlation coefficient, Kendall's correlation coefficient

บทนำ

สัมประสิทธิ์สหสัมพันธ์เป็นการวัดทางสถิติอย่างง่ายเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวขึ้นไป ตัวอย่างเช่น อาจเป็นการวัดความสัมพันธ์ระหว่างตัวแปรตาม 1 ตัวกับตัวแปรต้นตั้งแต่ 1 ตัว เป็นต้น (Bolboacă & Jäntschi, 2006; Chok, 2010) สัมประสิทธิ์สหสัมพันธ์เป็นดัชนีชี้วัดปริมาณของ ความเกี่ยวข้องของตัวแปร ในการวิเคราะห์ความสัมพันธ์ของตัวแปร 2 ตัว เมื่อกำหนดให้ตัวแปร X มีค่ามากขึ้น ค่าของตัวแปร Y มีแนวโน้มมากขึ้นด้วย ความสัมพันธ์ในลักษณะนี้ถือว่าเป็นความสัมพันธ์เชิงบวก (positive correlation) แต่ถ้าตัวแปร X มีค่ามากขึ้น แต่ค่าของตัวแปร Y มีแนวโน้มลดลงหรือตัวแปร X มีค่าน้อยลง แต่ค่าของตัวแปร Y มีแนวโน้มเพิ่มขึ้น ความสัมพันธ์ในลักษณะนี้ถือว่าเป็นความสัมพันธ์เชิงลบ (Negative correlation) (Tukey, 1977) โดยความสัมพันธ์ระหว่างตัวแปรนั้นสามารถพิจารณาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ที่ต่างกันไปตามระดับการวัดของตัวแปร สัมประสิทธิ์สหสัมพันธ์ที่นิยมที่สุดในการศึกษาเรื่องความสัมพันธ์ระหว่างตัวแปร ได้แก่ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson correlation coefficient), สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน (Spearman's rank correlation coefficient) และ สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ (Kendall's rank correlation coefficients) (Hauke & Kossowski, 2011; Wang, 2012)

การวิจัยเพื่อศึกษาเปรียบเทียบระหว่างสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์เป็นที่สนใจศึกษาของนักสถิติมาเป็นเวลานาน อย่างไรก็ตามการวิจัยศึกษาเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์ทั้งสาม ยังไม่ปรากฏการศึกษาเปรียบเทียบในกรณีที่ข้อมูลแจกแจงแบบไม่ปกติด้วยการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอล (lognormal distribution) ดังนั้น ผู้วิจัยจึงมีความสนใจศึกษาเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์ทั้งสาม โดยพิจารณาในด้านการประมาณค่าและการทดสอบสมมติฐาน ประสิทธิภาพด้านการประมาณค่าพิจารณาจากค่าความเอนเอียงสัมพัทธ์ (relative bias หรือ RB) และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (Monte Carlo standard errors หรือ MCSE) และประสิทธิภาพด้านการทดสอบสมมติฐานพิจารณาจากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่ง (type I error rate) และอำนาจการทดสอบสมมติฐาน (power of testing) โดยมุ่งหวังให้งานวิจัยนี้เป็นประโยชน์ต่อผู้สนใจในการเลือกใช้สัมประสิทธิ์สหสัมพันธ์ได้อย่างถูกต้องและเหมาะสมในการทดสอบความสัมพันธ์ของตัวแปรต่อไป

แนวคิดเกี่ยวกับสัมประสิทธิ์สหสัมพันธ์และการเปรียบเทียบประสิทธิภาพ สหสัมพันธ์ (correlation) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป โดยการวัดระดับความสัมพันธ์ระหว่างตัวแปรนั้นต้องอาศัยค่าสถิติ เรียกว่า สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ทั้งนี้ การเลือกใช้สัมประสิทธิ์สหสัมพันธ์ให้เหมาะสมขึ้นอยู่กับระดับการวัดของตัวแปรที่ผู้วิจัยสนใจศึกษา โดยสัมประสิทธิ์สหสัมพันธ์ที่พบมากในการศึกษาเรื่องความสัมพันธ์ระหว่างตัวแปร ได้แก่ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ ซึ่งมีแนวคิดหลักการ ดังนี้

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เป็นการศึกษาความสัมพันธ์เชิงเส้น (linear relationship) ระหว่างตัวแปร 2 ตัวหรือข้อมูล 2 ชุด โดยมีข้อตกลงเบื้องต้น คือ ตัวแปรทั้งสองต้องเป็นตัวแปรต่อเนื่อง อยู่ในมาตราอันตรภาคชั้น (interval scale) ขึ้นไป และมีการแจกแจงปกติสองตัวแปร (bivariate normal distribution) โดยค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน จะมีค่าอยู่ระหว่าง - 1 ถึง + 1 ถ้า เข้าใกล้ ± 1 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันอยู่ในระดับมาก ถ้าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันในระดับน้อยหรือไม่สัมพันธ์กันเลย เครื่องหมายบวก (+) จะแสดงทิศทางของความสัมพันธ์ว่ามีทิศทางเป็นแบบใด โดยถ้าสัมประสิทธิ์สหสัมพันธ์มีเครื่องหมายบวก (+) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไปทิศทางเดียวกัน หากมีเครื่องหมายลบ (-) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไม่ไปทิศทางเดียวกันหรือตรงกันข้าม (Bolboacă & Jäntschi, 2006; Chok, 2010; Hauke & Kossowski, 2011; Wang, 2012) สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สามารถคำนวณได้จากสูตร (Pearson, 1920) ดังนี้

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

เมื่อ	r_{xy}	คือ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
	$\sum x$	คือ ผลรวมของข้อมูลที่วัดจากตัวแปรที่ 1 หรือตัวแปร x
	$\sum y$	คือ ผลรวมของข้อมูลที่วัดจากตัวแปรที่ 1 หรือตัวแปร y
	$\sum xy$	คือ ผลรวมของผลคูณของตัวแปรที่ 1 และ 2
	$\sum x^2$	คือ ผลรวมของกำลังสองของข้อมูลที่วัดจากตัวแปรที่ 1
	$\sum y^2$	คือ ผลรวมของกำลังสองของข้อมูลที่วัดจากตัวแปรที่ 1
	n	คือ ขนาดตัวอย่าง

นอกจากสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน หากตัวแปรทั้งสองอยู่ในระดับการวัดที่เป็นมาตราวัดแบบเรียงอันดับ (ordinal scale) หรือไม่มีข้อตกลงเบื้องต้นในเรื่องของระดับการวัดของตัวแปร สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เป็นค่าสัมประสิทธิ์สหสัมพันธ์ที่สามารถนำมาใช้เป็นทางเลือกให้กับข้อมูลในมาตรวัดนี้ (Chok, 2010)

สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนเป็นการศึกษาความสัมพันธ์แบบไม่อิงพารามิเตอร์ (non-parametric correlation) (Bolboacă & Jäntschi, 2006) ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนไม่ได้ใช้ในการวัดความสัมพันธ์เชิงเส้นระหว่างสองตัวแปรหรือข้อมูล 2 ชุด แต่สามารถอธิบายความสัมพันธ์ระหว่างสองตัวแปรได้โดยไม่ต้องทำตามข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงความถี่ของตัวแปรซึ่งจะแตกต่างจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน อย่างไรก็ตาม ข้อมูลที่จะนำมาวิเคราะห์สหสัมพันธ์แบบสเปียร์แมน จะต้องเป็นตัวอย่างสุ่มของตัวแปร 2 ตัว และอยู่ในมาตราเรียงอันดับ แต่ไม่จำเป็นต้องมีการแจกแจงแบบปกติ (Bolboacă & Jäntschi, 2006; Hauke & Kossowski, 2011) สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนสามารถคำนวณได้จากสูตร (Spearman, 1904) ดังนี้

$$r_s = 1 - \frac{6 \sum (p_i - q_i)^2}{n(n^2 - 1)}$$

เมื่อ	r_s	คือ สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน
	$\sum (p_i - q_i)^2$	คือ ผลรวมของกำลังสองของผลต่างของอันดับคะแนนในแต่ละคู่
	n	คือ ขนาดตัวอย่าง

สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เป็นสัมประสิทธิ์สหสัมพันธ์แบบไม่อิงพารามิเตอร์เช่นเดียวกับสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน และเหมาะกับการวัดความสัมพันธ์ของตัวแปรที่เป็นมาตราวัดแบบเรียงอันดับที่ไม่เป็นช่วง สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ถูกพิจารณาให้มีค่าเท่ากับสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน ในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนมีความเหมือนกับสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันเพียงแต่คำนวณจากการจัดลำดับ (rank) (Bolboacă & Jäntschi, 2006) สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ สามารถคำนวณได้จากสูตร (Kendall, 1938) ดังนี้

$$\tau = \frac{(N_c - N_d)}{\frac{1}{2}n(n-1)}$$

เมื่อ τ	คือ สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์
N_c	คือ จำนวนของอันดับที่สอดคล้องกัน (Number of concordant)
N_d	คือ จำนวนของอันดับที่ไม่สอดคล้องกัน (Number of discordant)
n	คือ ขนาดตัวอย่าง

ในการเปรียบเทียบประสิทธิภาพการประมาณค่าและการทดสอบสมมติฐานของการวิเคราะห์ด้วยสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ ประสิทธิภาพด้านการประมาณพิจารณาได้จากค่าความเอนเอียงสัมพัทธ์ (RB) และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (MCSE) ส่วนประสิทธิภาพด้านการทดสอบสมมติฐานพิจารณาได้จากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่งและอำนาจการทดสอบสมมติฐาน จากการทบทวนวรรณกรรม สรุปได้ดังนี้

ความเอนเอียงสัมพัทธ์ (RB) เป็นเกณฑ์ที่ใช้วัดความเอนเอียงหรือความแตกต่างของพารามิเตอร์ θ_j ในการประมาณค่าพารามิเตอร์ของประชากร θ_j โดยสามารถคิดเป็นเป็นสัดส่วนหรือค่าร้อยละของความแตกต่าง (Chumney, 2012; Trierweiler, 2010) ความเอนเอียงสัมพัทธ์สามารถคำนวณได้ดังนี้

$$RB(\hat{\theta}_j) = \frac{|\overline{\hat{\theta}_j} - \theta_j|}{\theta_j} \quad \text{โดยที่ } j = 1, 2, 3, \dots, n$$

เมื่อ $\overline{\hat{\theta}_j}$	คือ ค่าเฉลี่ยของการประมาณค่าพารามิเตอร์ที่ได้จากการจำลองข้อมูล n รอบ
θ_j	คือ ค่าประมาณพารามิเตอร์รอบที่ j

ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (MSCE) เป็นเกณฑ์ที่ใช้ในการวัดความคลาดเคลื่อนในการประมาณค่าความคลาดเคลื่อนมาตรฐานของการจำลองข้อมูลในแต่ละรอบ (Boos & Osborne, 2014) สามารถคำนวณได้ดังนี้

$$MSCE = \sqrt{\frac{\sum_{ij=1}^n (\hat{\theta}_j - \overline{\hat{\theta}_j})^2}{n}}$$

เมื่อ $\overline{\hat{\theta}_j}$ คือ ค่าเฉลี่ยของค่าประมาณพารามิเตอร์ที่ได้จากการจำลองข้อมูล n รอบ

ความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่ง (probability of a type I error) หมายถึง โอกาสในการปฏิเสธสมมติฐานหลัก (null hypothesis) เมื่อสมมติฐานหลักเป็นจริง

อำนาจการทดสอบสมมติฐาน (power of testing) หมายถึง โอกาสในการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นเท็จ

วัตถุประสงค์

1. เพื่อเปรียบเทียบประสิทธิภาพด้านการประมาณค่าของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ เมื่อข้อมูลแจกแจงแบบไม่ปกติ

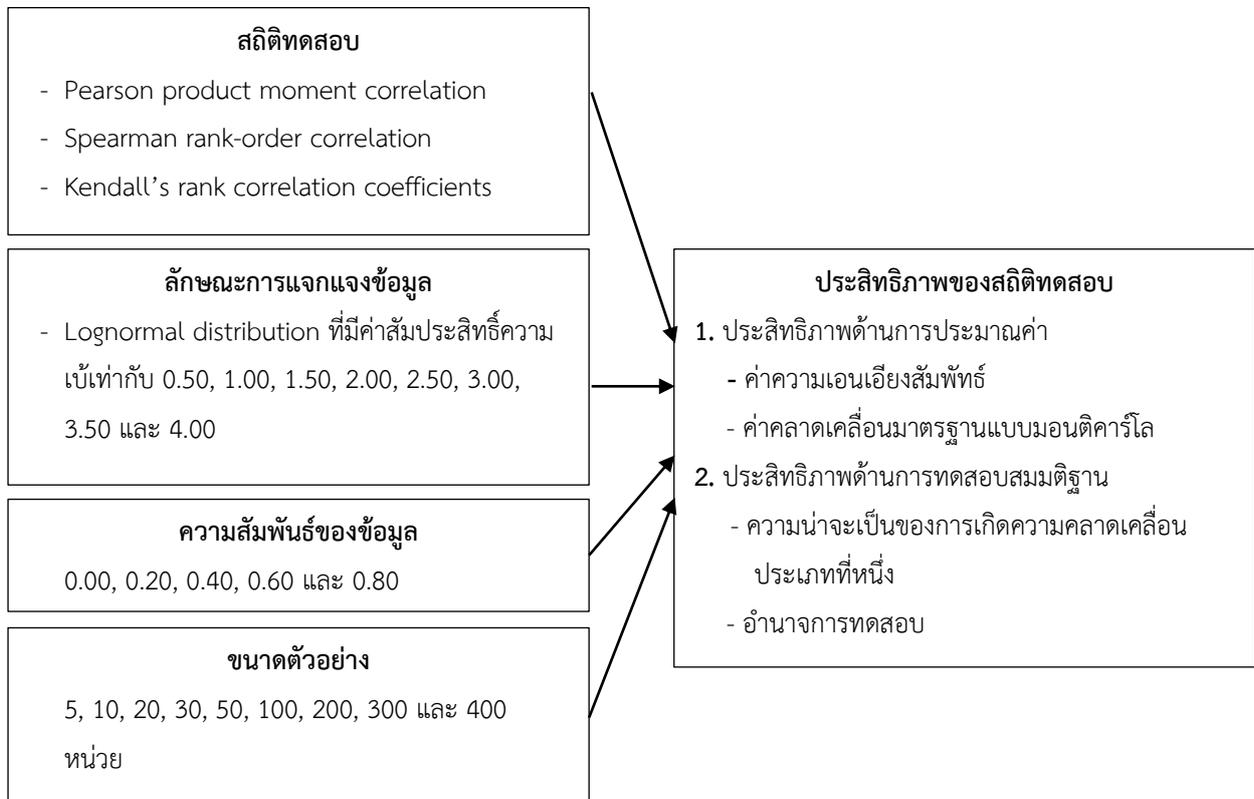
2. เพื่อเปรียบเทียบประสิทธิภาพด้านการทดสอบสมมติฐานของสหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ เมื่อข้อมูลแจกแจงแบบไม่ปกติ

วิธีการวิจัย

ในการวิจัยครั้งนี้ใช้ข้อมูลจำลองด้วยเทคนิคมอนติคาร์โลด้วยโปรแกรม R โดยจำลองจากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอล (lognormal distribution) ที่มีพารามิเตอร์ $\mu = 1$ และ $\sigma = 0.162, 0.314, 0.444, 0.552, 0.641, 0.716, 0.779$ และ 0.833 ซึ่งเทียบเท่าการแจกแจงของข้อมูลแบบเบ้ขวาที่มีค่าสัมประสิทธิ์ความเบ้เท่ากับ $0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50$ และ 4.00 ตามลำดับ โดยจำลองข้อมูลที่มีค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรเท่ากับ $0.00, 0.20, 0.40, 0.60$ และ 0.80 และมีขนาดตัวอย่างเท่ากับ $5, 10, 20, 30, 50, 100, 200, 300$ และ 400 หน่วย ตามลำดับ คิดเป็นสถานการณ์จำลองทั้งสิ้น 288 สถานการณ์ กระทำซ้ำสถานการณ์ละ 1,000 รอบประสิทธิภาพด้านการประมาณค่าพิจารณาจากค่าความเอนเอียงสัมพัทธ์ และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล และประสิทธิภาพด้านการทดสอบสมมติฐานพิจารณาจากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่ง และอำนาจการทดสอบ ในการวิจัยครั้งนี้ได้กำหนดกรอบแนวคิดในการวิจัย แสดงดังภาพประกอบ 1

ภาพ 1

กรอบแนวคิดการวิจัย



งานวิจัยนี้เป็นเปรียบเทียบประสิทธิภาพด้านการประมาณค่าและการทดสอบสมมติฐานของการวิเคราะห์สหสัมพันธ์ระหว่างการใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เมื่อข้อมูลแจกแจงแบบไม่ปกติ มีวิธีดำเนินงานวิจัยดังนี้

1. ทำการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลโดยใช้โปรแกรม R ด้วยฟังก์ชัน rlnorm โดยจำลองจากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอลที่มีพารามิเตอร์ค่าเฉลี่ย $\mu = 1$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma = 0.162, 0.314, 0.444, 0.552, 0.641, 0.716, 0.779$ และ 0.833 ซึ่งเทียบเท่าการแจกแจงของข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้เท่ากับ $0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50$ และ 4.00 ตามลำดับ
2. กำหนดค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรทั้งหมด 5 ระดับ ได้แก่ $0.00, 0.20, 0.40, 0.60$ และ 0.80
3. กำหนดขนาดตัวอย่างเท่ากับ $5, 10, 20, 30, 50, 100, 200, 300$ และ 400 หน่วย ตามลำดับ คิดเป็นสถานการณ์จำลองทั้งสิ้น 288 สถานการณ์
4. วิเคราะห์ประสิทธิภาพด้านการประมาณพิจารณาจากค่าความเอนเอียงสัมพัทธ์ (RB) และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (MCSE) และประสิทธิภาพด้านการทดสอบสมมติฐานพิจารณาจากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่งและอำนาจการทดสอบสมมติฐาน โดยใช้โปรแกรม R กระทำซ้ำสถานการณ์ละ 1,000 รอบ

ผลการวิจัย

ในการศึกษาครั้งนี้ผู้วิจัยขอเสนอผลการวิเคราะห์ข้อมูลออกเป็น 2 ส่วน ได้แก่ 1. ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณค่า และ 2. ผลการวิเคราะห์ประสิทธิภาพด้านการทดสอบสมมติฐาน ดังนี้

1. ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณค่า

ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณค่าพิจารณาจากค่าความเอนเอียงสัมพัทธ์ (RB) และค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โล (MCSE) ทั้งนี้ ผู้วิจัยนำเสนอตัวอย่างผลการวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลจากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอลที่มีพารามิเตอร์ค่าเฉลี่ย $\mu = 1$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma = 0.162$ และ 0.833 ซึ่งเทียบเท่าการแจกแจงของข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 0.50 และ 4.00 พบว่า

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์ต่ำที่สุดในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์มีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์สูงที่สุด เช่นพิจารณาข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 0.50 ค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.20 ที่ $n = 50$ พบว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีค่าความเอนเอียงสัมพัทธ์เท่ากับ 0.004 ในขณะที่ สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนมีค่าความเอนเอียงสัมพัทธ์เท่ากับ 0.038 และสัมประสิทธิ์สหสัมพันธ์แบบ

เคนดอลล์มีค่าความเอนเอียงสัมพันธ์เท่ากับ 0.341 เป็นต้น โดยที่ขนาดตัวอย่างเป็นปัจจัยที่มีอิทธิพลทางลบต่อความเอนเอียงสัมพันธ์ของค่าประมาณสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมน นั่นคือค่าความเอนเอียงสัมพันธ์ลดลงเมื่อขนาดของกลุ่มตัวอย่างใหญ่ขึ้น

นอกจากนี้ยังพบว่า สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมนมีแนวโน้มให้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลใกล้เคียงกันในกรณีที่ข้อมูลไม่มีความสัมพันธ์กัน ($\rho = 0.00$) และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ให้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลต่ำที่สุด ในทุกกรณีที่ข้อมูลมีและไม่มี ความสัมพันธ์กัน เช่น พิจารณาข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 4.00 ค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.20 ที่ $n = 20$ พบว่า สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมน มีค่า MCSC เท่ากับ 0.235 และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ มีค่า MCSC เท่ากับ 0.166 เป็นต้น ผลการวิเคราะห์แสดงดังตาราง 1 และตาราง 2

ตาราง 1

ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณของข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 0.50

N	ค่าความสัมพันธ์				RB			MCSE		
	rho	Pearson	Spearman	Kendall	RB.P	RB.S	RB.K	MCSE.P	MCSE.S	MCSE.K
5	0.00	0.01	0.01	0.01	Inf	Inf	Inf	0.496	0.492	0.397
	0.20	0.18	0.16	0.13	0.095	0.189	0.335	0.483	0.491	0.407
	0.40	0.38	0.35	0.28	0.061	0.137	0.299	0.447	0.447	0.378
	0.60	0.55	0.51	0.43	0.079	0.158	0.291	0.390	0.411	0.367
	0.80	0.77	0.70	0.60	0.043	0.128	0.250	0.261	0.300	0.300
10	0.00	0.01	0.01	0.01	Inf	Inf	Inf	0.330	0.333	0.247
	0.20	0.18	0.17	0.12	0.095	0.171	0.395	0.327	0.332	0.249
	0.40	0.39	0.35	0.26	0.028	0.118	0.341	0.284	0.293	0.228
	0.60	0.59	0.55	0.42	0.012	0.088	0.299	0.231	0.249	0.211
	0.80	0.79	0.74	0.59	0.014	0.080	0.256	0.145	0.179	0.174
20	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.235	0.233	0.167
	0.20	0.20	0.19	0.13	0.014	0.058	0.336	0.219	0.217	0.156
	0.40	0.38	0.36	0.25	0.048	0.106	0.364	0.203	0.206	0.153
	0.60	0.58	0.55	0.41	0.031	0.081	0.324	0.164	0.175	0.142
	0.80	0.79	0.76	0.59	0.011	0.055	0.263	0.094	0.111	0.109

ตาราง 1 (ต่อ)

ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณของข้อมูลที่มีค่าสัมประสิทธิ์ความแปร 0.50

n	ค่าความสัมพันธ์				RB			MCSE		
	rho	Pearson	Spearman	Kendall	RB.P	RB.S	RB.K	MCSE.P	MCSE.S	MCSE.K
30	0.00	0.00	-0.01	-0.01	Inf	Inf	Inf	0.184	0.181	0.126
	0.20	0.19	0.18	0.12	0.066	0.106	0.385	0.183	0.180	0.126
	0.40	0.40	0.38	0.27	0.008	0.056	0.336	0.154	0.158	0.117
	0.60	0.59	0.56	0.41	0.019	0.063	0.319	0.128	0.133	0.106
	0.80	0.80	0.77	0.59	0.005	0.042	0.261	0.074	0.085	0.082
50	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.143	0.144	0.098
	0.20	0.20	0.19	0.13	0.004	0.038	0.341	0.140	0.140	0.097
	0.40	0.39	0.38	0.26	0.013	0.050	0.340	0.120	0.120	0.087
	0.60	0.60	0.57	0.41	0.005	0.043	0.313	0.094	0.101	0.080
	0.80	0.80	0.77	0.59	0.006	0.037	0.265	0.057	0.066	0.064
100	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.096	0.095	0.064
	0.20	0.20	0.19	0.13	0.014	0.052	0.358	0.100	0.102	0.070
	0.40	0.40	0.38	0.26	0.009	0.044	0.342	0.086	0.087	0.062
	0.60	0.60	0.58	0.41	0.006	0.040	0.316	0.071	0.073	0.058
	0.80	0.80	0.78	0.59	0.001	0.029	0.263	0.040	0.046	0.044
200	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.071	0.071	0.048
	0.20	0.20	0.19	0.13	0.001	0.026	0.344	0.068	0.069	0.047
	0.40	0.40	0.39	0.26	0.003	0.035	0.338	0.062	0.063	0.045
	0.60	0.60	0.58	0.41	0.003	0.028	0.311	0.047	0.049	0.038
	0.80	0.80	0.78	0.59	0.002	0.026	0.264	0.028	0.032	0.031
300	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.057	0.057	0.038
	0.20	0.20	0.20	0.13	0.015	0.012	0.335	0.058	0.058	0.039
	0.40	0.40	0.38	0.26	0.010	0.040	0.343	0.049	0.049	0.035
	0.60	0.60	0.58	0.41	0.004	0.036	0.318	0.037	0.039	0.030
	0.80	0.80	0.78	0.59	0.002	0.025	0.265	0.023	0.026	0.025

ตาราง 2

ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณของข้อมูลที่มีค่าสัมประสิทธิ์ความแปร 4.00

n	ค่าความสัมพันธ์				RB			MCSE		
	rho	Pearson	Spearman	Kendall	RB.P	RB.S	RB.K	MCSE.P	MCSE.S	MCSE.K
5	0.00	-0.01	-0.01	0.00	Inf	Inf	Inf	0.519	0.520	0.428
	0.20	0.22	0.21	0.17	0.120	0.054	0.143	0.505	0.495	0.413
	0.40	0.39	0.37	0.30	0.022	0.085	0.244	0.466	0.446	0.379
	0.60	0.54	0.50	0.43	0.098	0.172	0.291	0.433	0.422	0.366
	0.80	0.71	0.65	0.57	0.117	0.189	0.293	0.349	0.342	0.321
10	0.00	-0.02	-0.01	0.00	Inf	Inf	Inf	0.317	0.319	0.236
	0.20	0.22	0.24	0.18	0.119	0.183	0.103	0.334	0.316	0.238
	0.40	0.40	0.41	0.32	0.012	0.020	0.212	0.332	0.297	0.230
	0.60	0.58	0.55	0.43	0.038	0.090	0.276	0.294	0.260	0.216
	0.80	0.75	0.69	0.56	0.065	0.144	0.295	0.236	0.218	0.195
20	0.00	-0.01	-0.01	-0.01	Inf	Inf	Inf	0.235	0.235	0.166
	0.20	0.22	0.24	0.17	0.105	0.204	0.135	0.245	0.222	0.157
	0.40	0.42	0.43	0.32	0.042	0.070	0.207	0.250	0.197	0.148
	0.60	0.60	0.57	0.44	0.008	0.052	0.273	0.227	0.176	0.141
	0.80	0.76	0.72	0.58	0.045	0.098	0.280	0.168	0.134	0.118
30	0.00	0.00	-0.01	-0.01	Inf	Inf	Inf	0.185	0.189	0.132
	0.20	0.21	0.25	0.18	0.058	0.243	0.121	0.205	0.178	0.125
	0.40	0.41	0.42	0.31	0.018	0.056	0.230	0.208	0.163	0.120
	0.60	0.59	0.57	0.43	0.021	0.046	0.281	0.193	0.142	0.113
	0.80	0.78	0.72	0.57	0.027	0.101	0.287	0.145	0.115	0.100
50	0.00	0.00	-0.01	0.00	Inf	Inf	Inf	0.150	0.143	0.098
	0.20	0.20	0.24	0.17	0.018	0.218	0.146	0.154	0.135	0.093
	0.40	0.40	0.42	0.31	0.007	0.055	0.237	0.167	0.123	0.090
	0.60	0.59	0.57	0.43	0.020	0.043	0.285	0.167	0.111	0.087
	0.80	0.78	0.72	0.57	0.029	0.095	0.291	0.123	0.085	0.074
100	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.099	0.102	0.068
	0.20	0.20	0.25	0.17	0.003	0.241	0.140	0.112	0.098	0.067
	0.40	0.41	0.44	0.31	0.014	0.089	0.220	0.124	0.083	0.060
	0.60	0.60	0.58	0.43	0.008	0.036	0.284	0.130	0.075	0.059
	0.80	0.79	0.73	0.57	0.012	0.089	0.291	0.095	0.062	0.054

ตาราง 2 (ต่อ)

ผลการวิเคราะห์ประสิทธิภาพด้านการประมาณของข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 4.00

n	ค่าความสัมพันธ์				RB			MCSE		
	rho	Pearson	Spearman	Kendall	RB.P	RB.S	RB.K	MCSE.P	MCSE.S	MCSE.K
200	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.071	0.071	0.047
	0.20	0.21	0.26	0.18	0.034	0.289	0.112	0.081	0.065	0.044
	0.40	0.40	0.43	0.31	0.004	0.080	0.230	0.093	0.060	0.043
	0.60	0.60	0.58	0.43	0.001	0.029	0.282	0.097	0.053	0.041
	0.80	0.79	0.73	0.57	0.010	0.085	0.290	0.072	0.042	0.036
300	0.00	0.00	0.00	0.00	Inf	Inf	Inf	0.057	0.059	0.039
	0.20	0.20	0.20	0.13	0.006	0.004	0.329	0.054	0.054	0.037
	0.40	0.40	0.40	0.27	0.004	0.009	0.318	0.051	0.050	0.036

2. ผลการวิเคราะห์ประสิทธิภาพด้านการทดสอบสมมติฐาน

ผลการวิเคราะห์ประสิทธิภาพด้านการทดสอบสมมติฐานพิจารณาจากความน่าจะเป็นของการเกิดความคลาดเคลื่อนประเภทที่หนึ่งและอำนาจการทดสอบสมมติฐาน โดยผู้วิจัยขอนำเสนอตัวอย่างผลการวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลจากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอลที่มีพารามิเตอร์ค่าเฉลี่ย $\mu = 1$ และส่วนเบี่ยงเบนมาตรฐาน $\sigma = 0.162$ และ และ 0.833 ซึ่งเทียบเท่าการแจกแจงของข้อมูลที่มีค่าสัมประสิทธิ์ความเบ้ 0.50 และ 4.00 พบว่า

การทดสอบนัยสำคัญของสหสัมพันธ์ด้วยสัมประสิทธิ์สหสัมพันธ์ทั้งสามแบบมีแนวโน้มควบคุมอัตราความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่ .05 และสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้อำนาจการทดสอบสูงที่สุด ในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนและเคนดอลล์มีแนวโน้มให้อำนาจการทดสอบไม่แตกต่างกัน เช่น พิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ที่ 0.20 ที่ $n = 20$ พบว่า พบว่า สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีอำนาจการทดสอบเท่ากับ 0.14 ในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนและเคนดอลล์ อำนาจการทดสอบเท่ากับ 0.12 เป็นต้น ผลการวิเคราะห์แสดงดังตาราง 3

ตาราง 3

ผลการวิเคราะห์ความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบของข้อมูล

		Type I error rate และ Power of the test ตามระดับความสัมพันธ์ (กำหนด rep = 1,000)									
n	การทดสอบ	ค่าสัมประสิทธิ์ความเบ้ 0.50					ค่าสัมประสิทธิ์ความเบ้ 4.00				
		Type I		Power of the test			Type I		Power of the test		
		0.00	0.20	0.40	0.60	0.80	0.00	0.20	0.40	0.60	0.80
5	Pearson	0.04	0.07	0.10	0.19	0.44	0.06	0.12	0.16	0.27	0.45
	Spearman	0.01	0.03	0.03	0.08	0.15	0.03	0.03	0.04	0.07	0.14
	Kendall	0.01	0.03	0.03	0.08	0.15	0.03	0.03	0.04	0.07	0.14
10	Pearson	0.04	0.08	0.21	0.50	0.88	0.04	0.14	0.29	0.51	0.76
	Spearman	0.05	0.07	0.18	0.41	0.78	0.04	0.10	0.25	0.42	0.66
	Kendall	0.04	0.07	0.16	0.40	0.77	0.03	0.09	0.25	0.44	0.67
20	Pearson	0.06	0.14	0.41	0.80	1.00	0.06	0.20	0.48	0.75	0.95
	Spearman	0.06	0.12	0.36	0.73	0.99	0.05	0.18	0.49	0.78	0.96
	Kendall	0.06	0.12	0.35	0.74	0.99	0.05	0.18	0.51	0.80	0.97
30	Pearson	0.05	0.17	0.62	0.95	1.00	0.05	0.23	0.59	0.87	0.99
	Spearman	0.06	0.16	0.55	0.92	1.00	0.05	0.28	0.65	0.92	0.99
	Kendall	0.05	0.15	0.55	0.92	1.00	0.06	0.29	0.67	0.94	1.00
50	Pearson	0.06	0.29	0.83	1.00	1.00	0.07	0.32	0.77	0.95	1.00
	Spearman	0.05	0.28	0.80	1.00	1.00	0.05	0.40	0.88	0.99	1.00
	Kendall	0.05	0.28	0.80	1.00	1.00	0.05	0.42	0.90	1.00	1.00
100	Pearson	0.04	0.51	0.99	1.00	1.00	0.04	0.50	0.96	1.00	1.00
	Spearman	0.04	0.48	0.98	1.00	1.00	0.05	0.70	1.00	1.00	1.00
	Kendall	0.04	0.47	0.98	1.00	1.00	0.05	0.73	1.00	1.00	1.00
200	Pearson	0.05	0.81	1.00	1.00	1.00	0.04	0.80	1.00	1.00	1.00
	Spearman	0.05	0.79	1.00	1.00	1.00	0.04	0.96	1.00	1.00	1.00
	Kendall	0.05	0.79	1.00	1.00	1.00	0.04	0.97	1.00	1.00	1.00

อภิปรายผลการวิจัย

ผลจากการศึกษาเปรียบเทียบประสิทธิภาพด้านการประมาณค่าและการทดสอบสมมติฐานของการวิเคราะห์สหสัมพันธ์ระหว่างการใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ เมื่อข้อมูลไม่มีการแจกแจงแบบปกติ มีสาระสำคัญของผลการวิเคราะห์ข้อมูล ดังนี้

1. สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์ต่ำที่สุดในขณะที่สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์มีแนวโน้มให้ค่าประมาณที่มีความเอนเอียงสัมพัทธ์สูงที่สุด

สอดคล้องกับงานวิจัยของ Zimmerman et al. (2003) ที่ทำการศึกษาความเอนเอียงในการประมาณค่าของ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมน พบว่า สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันความเอนเอียง ในการประมาณค่าต่ำกว่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน เช่นเดียวกับงานวิจัยของ Bishara and Hitter (2015) ที่พบว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันสามารถลดความเอนเอียงในการประมาณค่าได้ ดีกว่าสัมประสิทธิ์สหสัมพันธ์แบบอื่น ๆ ส่วนกรณีที่ประสิทธิสหสัมพันธ์แบบเคนดอลล์มีแนวโน้มให้ค่าประมาณ ที่มีความเอนเอียงสัมพัทธ์สูงที่สุดทั้งนี้อาจเป็นเพราะ สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ เหมาะสำหรับ ศึกษาความสัมพันธ์ระหว่าง 2 ตัวแปร ที่อยู่ในมาตรวัดแบบเรียงอันดับแบบไม่เป็นช่วง (non-interval scaled ordinal variables) (Bolboacă & Jäntschi, 2006) ในการจำลองข้อมูลในงานวิจัยครั้งนี้เป็นการจำลอง ข้อมูลที่มีความต่อเนื่องจึงน่าจะไม่เหมาะกับการประมาณค่าด้วยสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ทำให้ได้ ค่าประมาณมีความเอนเอียงสัมพัทธ์สูงที่สุด นอกจากนี้พบว่า ขนาดตัวอย่างเป็นปัจจัยที่มีอิทธิพลทางลบต่อ ความเอนเอียงสัมพัทธ์ของค่าประมาณสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมน ทั้งนี้เนื่องมาจาก ขนาดตัวอย่างที่ใหญ่ขึ้น ทำให้ค่าประมาณพารามิเตอร์ θ_j ไม่แตกต่างจากค่าเฉลี่ยของการประมาณ ค่าพารามิเตอร์ที่ได้จากการจำลองข้อมูล $\overline{\theta_j}$ มากนัก จึงส่งผลต่อความเอนเอียงสัมพัทธ์ที่ลดลง สอดคล้องกับ งานวิจัยของ Zimmerman et al. (2003) ที่พบว่า ขนาดตัวอย่างที่ใหญ่ค่าความเอนเอียงจะมีค่าต่ำ

2. สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสเปียร์แมนมีแนวโน้มให้ค่าคลาดเคลื่อนมาตรฐานแบบ มอนติคาร์โลใกล้เคียงกันในกรณีที่ข้อมูลไม่มีความสัมพันธ์กัน และสัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ให้ค่า คลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลต่ำที่สุดในทุกกรณี สอดคล้องกับงานวิจัยของ ประชาชาติ อารีชาติ และคณะ (2560) และ Croux and Dehon (2008) พบว่า เมื่อศึกษาตัวแปร 2 ตัวแปรที่มีความสัมพันธ์กัน โดยข้อมูลมีการแจกแจงแบบไม่ปกติ ประสิทธิสหสัมพันธ์แบบเคนดอลล์มีแนวโน้มให้ค่าความคลาดเคลื่อน กำลังสองเฉลี่ย (mean square error) ต่ำที่สุดในทุกขนาดของกลุ่มตัวอย่าง อาจเนื่องมาจากการกำหนดการ จำลองข้อมูลในงานวิจัยนี้เป็นการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอล ที่มีพารามิเตอร์ค่าเฉลี่ยและส่วน เบี่ยงเบนมาตรฐานต่ำ ซึ่งอาจส่งผลให้ข้อมูลมีความสอดคล้องกับมาตรวัดแบบอันดับที่ไม่เป็นช่วง จึงทำให้ สัมประสิทธิ์สหสัมพันธ์แบบเคนดอลล์ให้ค่าคลาดเคลื่อนมาตรฐานแบบมอนติคาร์โลต่ำที่สุด

3. การทดสอบนัยสำคัญของสหสัมพันธ์ด้วยสัมประสิทธิ์สหสัมพันธ์ทั้งสามตัวมีแนวโน้มควบคุมอัตรา ความคลาดเคลื่อนประเภทที่หนึ่งให้อยู่ภายใต้ระดับนัยสำคัญที่ .05 โดยที่ขนาดตัวอย่าง $n = 5$ สหสัมพันธ์ แบบสเปียร์แมนมีการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่าสหสัมพันธ์แบบเพียร์สัน ทั้งนี้อาจเป็น เพราะสหสัมพันธ์แบบเพียร์สันเหมาะสำหรับการวิเคราะห์ข้อมูลแบบต่อเนื่อง การวิเคราะห์ความสัมพันธ์ที่ ขนาดตัวอย่างเล็กนั้นอาจจะทำให้ข้อมูลที่จำลองมาไม่เป็นข้อมูลต่อเนื่องที่ชัดเจน (Croux & Dehon, 2008) สหสัมพันธ์แบบสเปียร์แมนเหมาะสำหรับการวิเคราะห์ความสัมพันธ์ของข้อมูลมาตรวัดเรียงอันดับซึ่งอาจมอง ว่าข้อมูลที่จำลองนั้นอยู่ในมาตรวัดเรียงอันดับและไม่จำเป็นต้องแจกแจงแบบปกติ จึงทำให้สหสัมพันธ์ของ สเปียร์แมนควบคุมการเกิดความคลาดเคลื่อนประเภทที่ 1 ได้ดีกว่า (Zimmerman et al., 2003) ในลักษณะ

ข้อมูลที่แจกแจงแบบไม่เป็นปกติ ซึ่งสอดคล้องกับงานวิจัยของ ทองดี แยมสรवल (2530) ได้ทำการศึกษา การควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจการทดสอบระหว่างสัมประสิทธิ์สหสัมพันธ์แบบ สเปียร์แมน เคนดอลล์ และक्रमเมอวีร์ พบว่า อำนาจการทดสอบที่ใช้ทดสอบค่าสัมประสิทธิ์สหสัมพันธ์แบบ สเปียร์แมน มีอำนาจในการทดสอบสูงที่สุด ซึ่งสอดคล้องกับงานวิจัยของ นภดล พิมพ์จันทร์ และคณะ (2550) เมื่อข้อมูลไม่มีค่าที่ผิดปกติ ค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมนมีอำนาจในการทดสอบสูงที่สุดในกรณีที กลุ่มตัวอย่างขนาดเล็ก และสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีอำนาจในการทดสอบสูงที่สุดถ้ากลุ่มตัวอย่าง มีขนาดใหญ่

4. สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมีแนวโน้มให้อำนาจการทดสอบสูงที่สุดในขณะที่สัมประสิทธิ์ สหสัมพันธ์แบบสเปียร์แมนและเคนดอลล์มีแนวโน้มให้อำนาจการทดสอบไม่แตกต่างกัน ซึ่งสอดคล้องกับ งานวิจัยของ Chok (2008) ที่ได้ทำการศึกษาอำนาจการทดสอบทางสถิติ โดยได้เสนอว่าอำนาจการทดสอบ ทางสถิติได้รับผลกระทบจากโครงสร้างของค่าสัมประสิทธิ์สหสัมพันธ์และชนิดของการทดสอบทางสถิติ โดย ทำการศึกษา 4 รูปแบบของ การแจกแจงสองตัวแปร ประกอบด้วยคู่ของการแจกแจงแบบปกติ, การแจกแจง แบบล็อกปกติ (log-normal), ดับเบิลเอ็กซ์โพเนนเชียล (double exponential) และการแจกแจงแบบที (t-distribution) พบว่า การแจกแจงเหล่านี้ช่วยให้สามารถสร้างโมเดลสถานการณ์กับความแตกต่างระดับของ การฝ่าฝืนในแง่ของความเบ้ (skewness) และความโด่ง (kurtosis) และได้สรุปผลการตรวจสอบแสดงให้เห็น ว่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันอาจมีข้อได้เปรียบที่สำคัญสำหรับข้อมูลการกระจายของข้อมูลต่อเนื่องที่ ไม่ใช่โค้งปกติ (non-normal distribution) ซึ่งไม่มีค่าผิดปกติที่เห็นได้ชัด จึงส่งผลให้สัมประสิทธิ์สหสัมพันธ์ แบบเพียร์สันมีแนวโน้มให้อำนาจการทดสอบสูงที่สุด นอกจากนี้จากงานวิจัยยังพบว่า ขนาดตัวอย่างเป็นปัจจัย ที่มีอิทธิพลทางบวกต่ออำนาจการทดสอบของสัมประสิทธิ์สหสัมพันธ์ทั้งสามตัว ทั้งนี้เนื่องจาก ขนาดของ กลุ่มตัวอย่างที่มากขึ้นส่งผลให้การวิจัยและการวิเคราะห์ข้อมูลมีความคลาดเคลื่อนลดลง (บุญชม ศรีสะอาด, 2553) จึงส่งผลให้อำนาจการทดสอบของสัมประสิทธิ์สหสัมพันธ์ทั้งสามตัวมีค่าสูงขึ้นเมื่อตัวอย่างมีขนาดใหญ่ ขึ้น

ข้อเสนอแนะ

การเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ ใน งานวิจัยนี้ ทำให้ค้นพบว่า ถึงแม้ในการวิจัยได้มีการจำลองข้อมูลให้มีการแจกแจงแบบไม่เป็นปกติโดยจำลอง จากการแจกแจงความน่าจะเป็นแบบล็อกนอร์มอล สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันยังคงมีแนวโน้มให้ ค่าประมาณที่มีความเอนเอียงสัมพัทธ์ต่ำที่สุดและอำนาจการทดสอบสูงที่สุด อย่างไรก็ตามสัมประสิทธิ์ สหสัมพันธ์แบบเพียร์สันยังไม่ใช่ตัวเลือกที่เหมาะสมที่สุดในการศึกษาความสัมพันธ์ในทุกรูปแบบของการแจก แจงของข้อมูล จำเป็นต้องพิจารณาร่วมกับค่าความคลาดเคลื่อนอื่น ๆ ร่วมด้วย นอกจากนี้ ควรมีการศึกษา เปรียบเทียบประสิทธิภาพด้านการประมาณค่าและการทดสอบสมมติฐานของสหสัมพันธ์แบบอื่น ๆ เช่น

สัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก เป็นต้น หรือกับการแจกแจงข้อมูลในรูปแบบอื่น ๆ เพื่อเป็นแนวทางในการเลือกใช้สหสัมพันธ์อย่างเหมาะสมต่อไป

รายการอ้างอิง

ภาษาไทย

ทองดี แยมสรवल. (2530). การศึกษาลักษณะการแจกแจงการควบคุมความคลาดเคลื่อนประเภทที่ 1 และอำนาจ การทดสอบของค่าสัมประสิทธิ์สหสัมพันธ์แบบสเปียร์แมน แคนดัลเทาและเครมเมอร์วี [วิทยานิพนธ์ปริญญาโทบัณฑิต ไม่ได้ดีพิมพ์]. จุฬาลงกรณ์มหาวิทยาลัย.

นภดล พิมพ์จันทร์, จิราพร เขียวอยู่, นิคม ถนอมเสียง, และ ยุภาพร ตงประสิทธิ์. (2550). ผลของข้อมูลที่มีค่าผิดปกติจากกลุ่มต่อความแกร่งของสัมประสิทธิ์สหสัมพันธ์. *วารสารวิจัยมหาวิทยาลัยขอนแก่นระดับบัณฑิตศึกษา*, 7(2), 111-120.

บุญชม ศรีสะอาด. (2553). *การวิจัยเบื้องต้น* (พิมพ์ครั้งที่ 8). สุวีริยาสาส์น.

ประชาชาติ อารีชาติ, ชนิกันต์ ตั้งตระกูล, และ จุฑาภรณ์ สิ้นสมบูรณ์ทอง. (2560). การเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่เมื่อข้อมูลมีค่าผิดปกติ. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(6), 930-943.

ภาษาอังกฤษ

Bishara, A. J., & Hitter, J. B. (2015). Reducing bias and error in the correlation coefficient due to nonnormality. *Educational and Psychological Measurement*, 75(5), 785-804.

Bolboacă, S. D., & Jäntschi, L. (2006). Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationships of biologic active compounds. *Leonardo Journal of Sciences*, 9, 179-200.

Boos, D. D., & Osborne, J. A. (2014). Assessing variability of complex descriptive statistics in Monte Carlo studies using resampling methods. *International Statistical Review*, 83(2), 228–238. <https://doi.org/10.1111/insr.12087>

Chok, N. S., (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data* [Unpublished master's thesis]. University of Pittsburgh.

Chumney, F. L. (2012). *Comparison of maximum likelihood, Bayesian, partial least square and generalized structured component analysis methods for estimation of structural equation models with small sample: An exploratory study* [Unpublished master's thesis]. University of Nebraska - Lincoln.

- Croux, C., & Dehon, C. (2008). *Robustness versus efficiency for nonparametric correlation*. KU Leuven Association. https://lirias.kuleuven.be/bitstream/123456789/164226/1/KBI_0803.pdf
- Hauke, J., & Kossowski, T. (2011). Compare of value of Pearson's and Spearman's correlation coefficient on the same sets of data. *Quaestiones Geographicae*, 30(2), 1-2.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81-93.
- Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13(1), 25-45.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 100(3/4), 441-471.
- Trierweiler, T. (2010). *An evaluation of estimation methods in confirmatory factor analytic model with ordered categorical data in LISREL*. ETD Collection for Fordham University. <https://research.library.fordham.edu/dissertations/AAI3416004/>
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.
- Wang, J. (2012). On the relationship between Pearson correlation coefficient and Kendall's Tau under bivariate homogeneous shock model. *International Scholarly Research Notices*, 2012, 1-7. <https://doi.org/10.5402/2012/717839>
- Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicológica*, 24(1), 133-158.