PBRU SCIENCE JOURNAL
2024.21(1) : 17-29

การเปรียบเทียบประสิทธิภาพการทดสอบสัมประสิทธิ์สหสัมพันธ์เมื่อข้อมูลมีการแจกแจงแบบผสม

A Comparison of the Efficiency of Correlation Coefficient Tests for Mixture Distribution

บำรุงศักดิ์ เผื่อนอารีย์, อภัสรา เกิดพุ่ม และภัทราภรณ์ กิจผลเจริญ*

Bumrungsak Phuenaree, Aphassara Koedphum and Pattaraporn Kidpholjaroen*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

*Corresponding author; E-mail: pattara@go.buu.ac.th

Received: 21 November 2023 /Revised: 10 February 2024 /Accepted: 27 February 2024

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของการทดสอบสัมประสิทธิ์สหสัมพันธ์พื้นฐาน 3 วิธี ได้แก่ การทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สัน การทดสอบสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และการทดสอบสัมประสิทธิ์สหสัมพันธ์เคนดัล เมื่อข้อมูลมีการแจกแจงปรกติแบบผสม การแจกแจงล็อกนอร์มัลแบบผสม และการแจกแจงแกมมาแบบผสม ซึ่งการแจกแจงเหล่านี้พบได้กับข้อมูลในชีวิตประจำวัน เช่น ข้อมูลปริมาณน้ำฝน ข้อมูลการเคลมประกันรถยนต์ เป็นต้น โดยศึกษาที่ขนาดตัวอย่างเท่ากับ 10, 20, 30, 50 และ 100 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.0, 0.4, 0.6 และ 0.8 ผลการศึกษาพบว่า การทดสอบสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกกรณี และเมื่อพิจารณาจากกำลังการทดสอบ พบว่า การทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุด เมื่อตัวอย่างมีขนาดเล็ก สำหรับกรณีที่ตัวอย่างมีขนาดใหญ่และข้อมูลมีการแจกแจงปรกติแบบผสม การทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สันยังคงให้ค่ากำลังการทดสอบสูงที่สุด แต่เมื่อข้อมูลมีการแจกแจงล็อกนอร์มัลแบบผสม และการแจกแจงแกมมาแบบผสม พบว่ากรณีส่วนใหญ่การทดสอบสัมประสิทธิ์สหสัมพันธ์เคนดัล ให้ค่ากำลังการทดสอบสูงที่สุด

คำสำคัญ: สัมประสิทธิ์สหสัมพันธ์เพียร์สัน สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดัล การแจกแจงปรกติแบบผสม การแจกแจงล็อกนอร์มัลแบบผสม การแจกแจงแกมมาแบบผสม

Abstract

This study aimed to compare the effectiveness of three correlation coefficient tests: the Pearson correlation coefficient test, the Spearman correlation coefficient test, and the Kendall tau correlation coefficient test. Three types of mixture distribution were considered: mixed normal, mixed log-normal, and mixed gamma. These distributions were the distribution of daily life data, such as rainfall, car insurance claim, etc. The sample sizes in this study were 10, 20, 30, 50, and 100. The correlation coefficients were 0.0, 0.4, 0.6, and 0.8, respectively. The results indicated that all tests could control type I error in all cases. For the power of the test in case of small sample size, The Pearson correlation coefficient test showed the greatest power. In case of a large sample size, this test also performed well for mixed normal distribution but for mixed log-normal and mixed gamma distributions, Kendall tau correlation coefficient test was the best.

Keywords: Pearson correlation coefficient, Spearman correlation coefficient, Kendall tau correlation coefficient, mixed normal distribution, mixed log-normal distribution, mixed gamma distribution

บทนำ

การศึกษาความสัมพันธ์ของตัวแปรเชิงปริมาณสองตัวแปรนั้น วิธีการทางสถิติโดยทั่วไปที่ใช้กันอย่างแพร่หลายคือ สัมประสิทธิ์สหพันธ์เพียร์สัน ซึ่งวิธีการดังกล่าวมีข้อสมมุติเบื้องต้นที่สำคัญหลายประการและหนึ่งในข้อสมมุติเบื้องต้นคือ ตัวแปรทั้งสองตัวแปรต้องมีการแจกแจงปรกติและต้องมีความสัมพันธ์กัน (Bivariate normal distribution) หากตัวแปรที่ศึกษาไม่มีลักษณะที่เป็นไปตามข้อสมมุติเบื้องต้นที่กล่าวมา ต้องเปลี่ยนมาใช้สถิติทดสอบที่ไม่อิงพารามิเตอร์แทน ซึ่งวิธีที่รู้จักกันอย่างแพร่หลายและเป็นที่นิยมนำมาใช้ในกรณีนี้คือ สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์เคนดัล Chok[1] ได้ศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าและทดสอบสมมติฐานของสัมประสิทธิ์สหสัมพันธ์ 3 วิธี ได้แก่ สัมประสิทธิ์สหสัมพันธ์เพียร์สัน สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์เคนดัล เมื่อข้อมูลมีการแจกแจงต่อเนื่องที่สัมพันธ์กัน รวมถึง ชนินันท์ พฤกษ์ประมูลและคณะ[2] ได้ทำการศึกษาในลักษณะเดียวกันกับงานวิจัยข้างต้น แต่จำลองข้อมูลที่มีการแจกแจงความน่าจะเป็นแบบล็อกนอมัล ผลลัพธ์จากทั้งสองงานวิจัยพบว่า สัมประสิทธิ์สหสัมพันธ์เพียร์สันมีแนวโน้มให้กำลังการทดสอบสูงที่สุด ซึ่งผลลัพธ์ดังที่กล่าวมานั้นแสดงให้เห็นว่า แม้ในบางกรณีที่ข้อมูลที่นำมาวิเคราะห์ไม่มีการแจกแจงปรกติ สัมประสิทธิ์สหสัมพันธ์เพียร์สัน ยังคงมีแนวโน้มเป็นวิธีการที่มีประสิทธิภาพมากที่สุด ดังนั้นประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์ขึ้นอยู่กับลักษณะการแจกแจงของข้อมูลที่นำมาวิเคราะห์ด้วย

สำหรับข้อมูลบางประเภทโดยเฉพาะที่มาจากแหล่งธรรมชาติ เช่น ความเร็วลม ปริมาณน้ำฝน อีกทั้งในด้านธุรกิจ ด้านอุตสาหกรรม หรือแม้กระทั่งด้านอุตุนิยมวิทยา ที่นำมาใช้เพื่อช่วยในการตัดสินใจวางแผนดำเนินงาน ป้องกันและแก้ไขปัญหาที่อาจจะเกิดได้ โดยข้อมูลที่กล่าวมานี้ส่วนมากเป็นข้อมูลที่ไม่มีการแจกแจงปรกติ Suhali et al.[3] ได้ศึกษาเกี่ยวกับปริมาณน้ำฝนรายวันในคาบสมุทรมาเลเซีย โดยใช้การแจกแจงเอ็กซ์โปเนนเชียล การแจกแจงแกมมาแบบผสม การแจกแจงไวบูลล์แบบผสม การแจกแจงล็อกนอร์มัลแบบผสม พบว่าการแจกแจงล็อกนอร์มัลแบบผสมเป็นการแจกแจงที่เหมาะสมที่สุด Sattayatham and Talangtam[4] ได้ศึกษาการสร้างแบบจำลองชุดข้อมูลเคลมประกันภัยรถยนต์ โดยใช้การแจกแจงล็อกนอร์มัลแบบผสมในการวิเคราะห์ข้อมูล และงานวิจัยของ Williford[5] ได้ศึกษาระยะเวลาการรอเข้ารับการรักษาในโรงพยาบาล โดยใช้การแจกแจงแกมมาแบบผสมในการวิเคราะห์ข้อมูล ซึ่งจากงานวิจัยที่กล่าวมาทั้งหมดแสดงให้เห็นว่าข้อมูลที่สำคัญและเป็นประโยชน์มีลักษณะเป็นการแจกแจงแบบผสม โดยหากต้องการศึกษาความสัมพันธ์ระหว่างตัวแปรดังกล่าวควรเลือกใช้วิธีการพื้นฐานใดจึงจะมีประสิทธิภาพมากที่สุด ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาประสิทธิภาพของวิธีการทดสอบสหสัมพันธ์สามวิธี ได้แก่ การทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สัน การทดสอบสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และการทดสอบสัมประสิทธิ์สหสัมพันธ์เคนดัล ภายใต้ข้อมูลที่มีการแจกแจงปรกติแบบผสม การแจกแจงล็อกนอร์มัลแบบผสม และการแจกแจงแกมมาแบบผสม โดยมีเกณฑ์พิจารณา คือ ความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบ

ข้อมูลที่เกี่ยวข้อง

1. การทดสอบความสัมพันธ์ของตัวแปรสองตัว

ในการศึกษาความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเชิงปริมาณสองตัว สามารถตั้งสมมติฐานของการทดสอบได้เป็น H_0 ตัวแปรทั้งสองไม่มีความสัมพันธ์กัน และ H_a ตัวแปรทั้งสองมีความสัมพันธ์กัน ซึ่งการเลือกใช้วิธีการทดสอบต้องพิจารณาจากข้อสมมุติเบื้องต้นเพื่อให้ได้สถิติทดสอบที่เหมาะสม โดยค่าสถิติพื้นฐานที่นำมาศึกษามีดังต่อไปนี้

2. สัมประสิทธิ์สหสัมพันธ์เพียร์สัน

สัมประสิทธิ์สหสัมพันธ์ของเพียร์สันเป็นการวัดความสัมพันธ์ของตัวแปรเชิงปริมาณสองตัวแปร โดยมีข้อสมมุติเบื้องต้นที่สำคัญคือ ตัวแปรทั้งสองต้องเป็นตัวแปรต่อเนื่องและอยู่ในมาตราอันตรภาคชั้นขึ้นไป กำหนดให้ $X_1, X_2, ..., X_n$ และ $Y_1, Y_2, ..., Y_n$ เป็นตัวแปรสุ่มขนาด n ที่มีการแจกแจงปรกติและมีความสัมพันธ์กัน (Bivariate normal distribution) สามารถคำนวณสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน[6] ได้ดังสมการที่ 1

$$r = \frac{\sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \qquad (1)$$

3. สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน

กำหนดให้สัญลักษณ์ $X_1, X_2, ..., X_n$ และ $Y_1, Y_2, ..., Y_n$ เป็นตัวแปรสุ่มขนาด n ที่มีการแจกแจงแบบต่อเนื่องและมีความสัมพันธ์กัน ทำการเรียงลำดับข้อมูล X และ Y จากน้อยไปหามากตั้งแต่ลำดับที่ 1 ถึง n ของแต่ละตัวแปร และให้ $R(X_i)$ และ $R(Y_i)$ แทนลำดับที่ของข้อมูล X_i และ Y_i ตามลำดับ สามารถคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน[7] ได้ดังสมการที่ 2

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} \qquad (2)$$

โดยที่ $D_i = R(X_i) - R(Y_i)$ คือผลต่างของลำดับคู่ที่ i

4. สัมประสิทธิ์สหสัมพันธ์แบบเคนดัล

สัมประสิทธิ์สหสัมพันธ์แบบเคนดัลเป็นหนึ่งในสัมประสิทธิ์สหสัมพันธ์แบบไม่อิงพารามิเตอร์ เช่นเดียวกันกำหนดให้ $X_1, X_2, ..., X_n$ และ $Y_1, Y_2, ..., Y_n$ เป็นตัวแปรสุ่มขนาด n โดยจะพิจารณาความสอดคล้องของข้อมูล กล่าวคือ หากข้อมูล x และ y เพิ่มขึ้นทั้งคู่ นั่นคือข้อมูลที่สอดคล้องกัน แต่หากข้อมูล x เพิ่มขึ้น y ลดลง นั่นคือข้อมูลไม่สอดคล้องกัน สามารถคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเคนดัล[8] ดังสมการที่ 3

$$\tau = \frac{(N_c - N_d)}{\frac{1}{2}n(n-1)} \qquad (3)$$

โดยที่ N_c แทนจำนวนอันดับที่สอดคล้องกัน

N_d แทนจำนวนอันดับที่ไม่สอดคล้องกัน

5. เกณฑ์ที่ใช้ในงานวิจัย

สำหรับเกณฑ์การพิจารณาประสิทธิภาพของการทดสอบสัมประสิทธิ์สหสัมพันธ์จะพิจารณาจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 และกำลังการทดสอบของสถิติทดสอบ โดยมีรายละเอียดดังต่อไปนี้

5.1. ความสามารถในการควบคุมความผิดพลาดแบบที่ 1

ความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 คือ ความน่าจะเป็นของการปฏิเสธสมมติฐานหลัก เมื่อสมมติฐานหลักเป็นจริง สามารถประมาณค่าได้จาก $\hat{\alpha} = \frac{1}{N}$ โดยที่ I แทนจำนวนครั้งของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นจริง และ N แทนจำนวนครั้งของการทดสอบสมมติฐาน (จำนวนรอบของการทำซ้ำด้วยเทคนิคมอนติคาร์โล) ทั้งนี้ สำหรับระดับนัยสำคัญ $\alpha = 0.05$ ค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ควรอยู่ในช่วง [0.025, 0.075] ตามเกณฑ์ของ Bradley[9] จึงสรุปได้ว่าสถิติทดสอบสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ จากนั้นจึงพิจารณากำลังการทดสอบของสถิติทดสอบเป็นลำดับถัดไป

5.2. กำลังการทดสอบ

กำลังการทดสอบคือความน่าจะเป็นของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นเท็จสามารถประมาณค่าได้จาก $\frac{M}{N}$ เมื่อ M แทนจำนวนครั้งของการปฏิเสธสมมติฐานหลักเมื่อสมมติฐานหลักเป็นเท็จ และ N แทนจำนวนครั้งของการทดสอบสมมติฐาน (จำนวนรอบของการทำซ้ำด้วยเทคนิคมอนติคาร์โล)

6. การแจกแจงที่ศึกษา

การแจกแจงปรกติผสม

$$f(x)=\frac{p}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right)+\frac{(1-p)}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma_2^2}(x-\mu_2)^2\right);x>0 \qquad (4)$$

การแจกแจงล็อคนอร์มัลแบบผสม

$$f(x)=\frac{p}{x\sigma_1\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu_1)^2}{2\sigma_1^2}\right)+\frac{(1-p)}{x\sigma_2\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu_2)^2}{2\sigma_2^2}\right);x>0 \qquad (5)$$

โดยในสมการ (4) และ (5) σ_1,σ_2 แทนพารามิเตอร์บ่งรูปร่าง μ_1,μ_2 แทนพารามิเตอร์บ่งตำแหน่ง และ p แทนพารามิเตอร์ถ่วงน้ำหนัก

การแจกแจงแกมมาแบบผสม

$$f(x)=\frac{p}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}x^{\alpha_1-1}\exp\left(-\frac{x}{\beta_1}\right)+\frac{(1-p)}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}x^{\alpha_2-1}\exp\left(-\frac{x}{\beta_2}\right);x>0 \qquad (6)$$

โดยในสมการ (6) α_1,α_2 แทนพารามิเตอร์บ่งรูปร่าง β_1,β_2 แทนพารามิเตอร์บ่งตำแหน่ง และ p แทนพารามิเตอร์ถ่วงน้ำหนัก

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์สามวิธีที่ใช้ทดสอบความสัมพันธ์ของสองตัวแปร ได้แก่ สัมประสิทธิ์สหสัมพันธ์เพียร์สัน สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์เคนดัล โดยกำหนดขนาดตัวอย่างเท่ากับ 10, 20, 30, 50 และ 100 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.0 (สำหรับศึกษาความสามารถในการควบคุมความผิดพลาดแบบที่ 1), 0.4, 0.6 และ 0.8 (สำหรับศึกษากำลังการทดสอบ) ซึ่งแทนขนาดความสัมพันธ์ของตัวแปร 3 ระดับคือ ต่ำ ปานกลาง และสูง ตามลำดับ จำลองข้อมูลที่มีการแจกแจงแบบผสม 3 การแจกแจงได้แก่ การแจกแจงปรกติผสม การแจกแจงล็อกนอร์มัลแบบผสมและการแจกแจงแกมมาแบบผสม ด้วยโปรแกรม R (ฟังก์ชัน “cor.test”) กำหนดพารามิเตอร์ของแต่ละการแจกแจงในสมการที่ (4) – (6) ดัง Table 1 – 3 ทำซ้ำด้วยเทคนิคมอนติคาร์โล 10,000 รอบ พิจารณาประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์จากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 แสดงดัง Figure 1 - 3 และกำลังการทดสอบ แสดงดัง Figure 4 – 12

Table 1. Parameters of mixed normal distribution

ชุดพารามิเตอร์	องค์ประกอบที่ 1		องค์ประกอบที่ 2		p
	μ_1	σ_1	μ_2	σ_2	
MN1	0	1	1	1	0.5
MN2	0	1	1	1	0.8
MN3	0	1	2	1	0.5
MN4	0	1	2	1	0.8

Table 2. Parameters of mixed log-normal distribution

ชุดพารามิเตอร์	องค์ประกอบที่ 1		องค์ประกอบที่ 2		p
	μ_1	σ_1	μ_2	σ_2	
ML1	- (log 2)/2	log (2)	- (log 3)/2	log (3)	0.5
ML2	- (log 2)/2	log (2)	- (log 3)/2	log (3)	0.8
ML3	- (log 2)/2	log (2)	- (log 5)/2	log (5)	0.5
ML4	- (log 2)/2	log (2)	- (log 5)/2	log (5)	0.8

Table 3. Parameters of mixed gamma distribution

ชุดพารามิเตอร์	องค์ประกอบที่ 1		องค์ประกอบที่ 2		p
	α_1	β_1	α_2	β_2	
MG1	1	1	1/2	2	0.5
MG2	1	1	1/2	2	0.8
MG3	1	1	1/4	4	0.5
MG4	1	1	1/4	4	0.8

ผลการวิจัย

ผลการศึกษาแสดงเป็นแผนภาพค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 และค่าประมาณกำลังการทดสอบ ดังต่อไปนี้

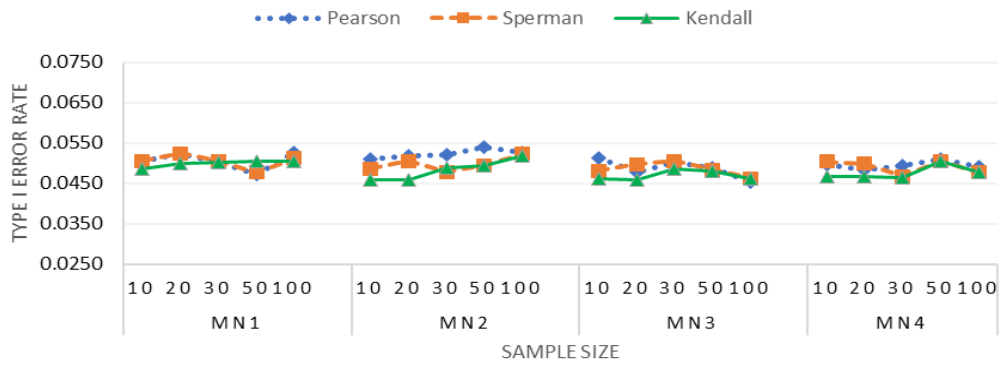

Figure 1. Type I error rates when samples are taken from mixed normal distribution.

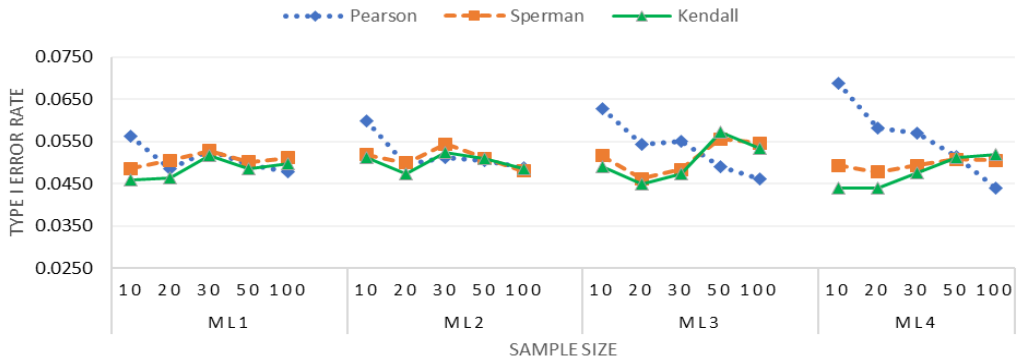

Figure 2. Type I error rates when samples are taken from mixed log-normal distribution.

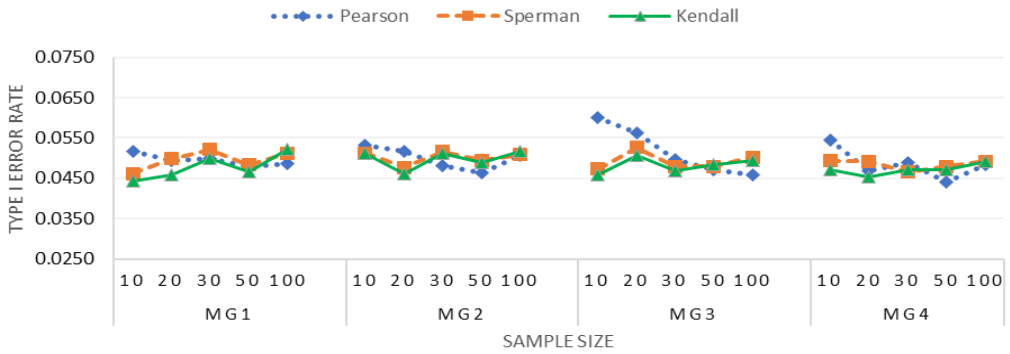

Figure 3. Type I error rates when samples are taken from mixed gamma distribution.

พิจารณาค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 จาก Figure 1 - 3 พบว่า เมื่อข้อมูลมีการแจกแจงปรกติแบบผสม การแจกแจงล็อกนอร์มัลแบบผสม และการแจกแจงแกมมาแบบผสม การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี มีค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 อยู่ในเกณฑ์ของ Bradley ดังนั้นการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทุกวิธีสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกกรณีที่ศึกษา

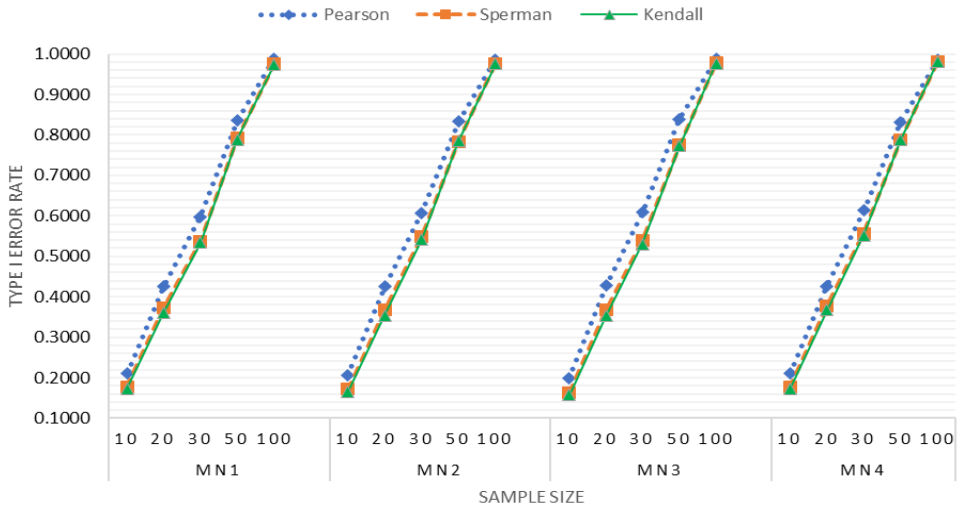

Figure 4. Power of the test when samples are taken from mixed normal distribution with $\rho = 0.4$.

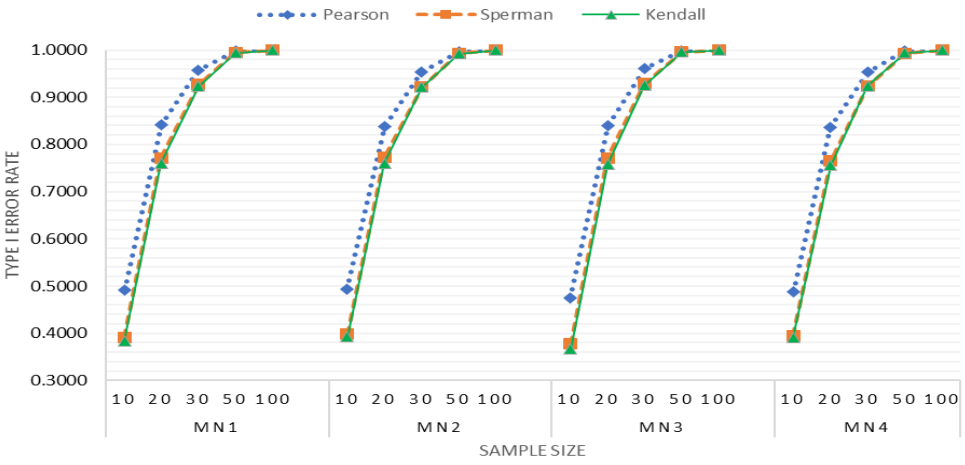

Figure 5. Power of the test when samples are taken from mixed normal distribution with $\rho = 0.6$.

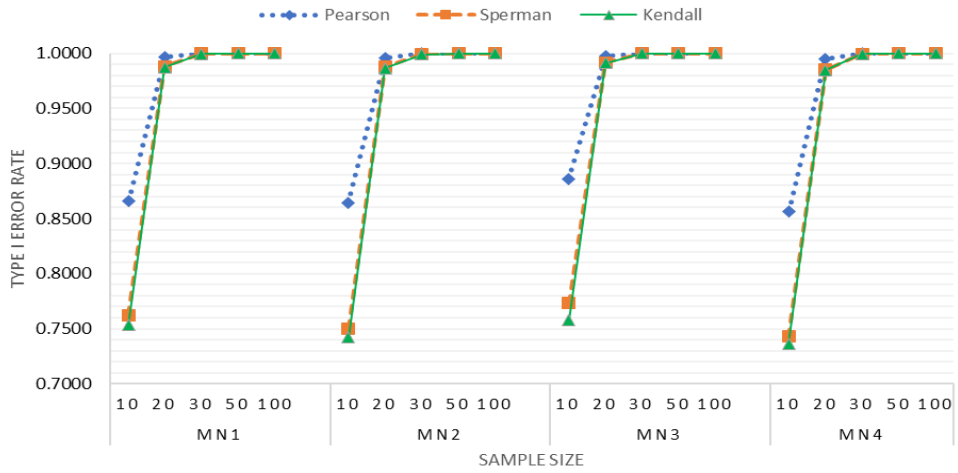

Figure 6. Power of the test when samples are taken from mixed normal distribution with $\rho = 0.8$.

พิจารณาค่ากำลังการทดสอบสำหรับข้อมูลที่มีการแจกแจงปรกติแบบผสม จาก Figure 4 - 6 พบว่า เมื่อกำหนด $\rho = 0.4$ การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุดและเมื่อกำหนด $\rho = 0.6, 0.8$ พบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุดเมื่อขนาดตัวอย่างไม่เกิน 50 นอกจากนี้ยังพบว่าการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี มีค่ากำลังการทดสอบเท่ากันเมื่อขนาดตัวอย่างเท่ากับ 100

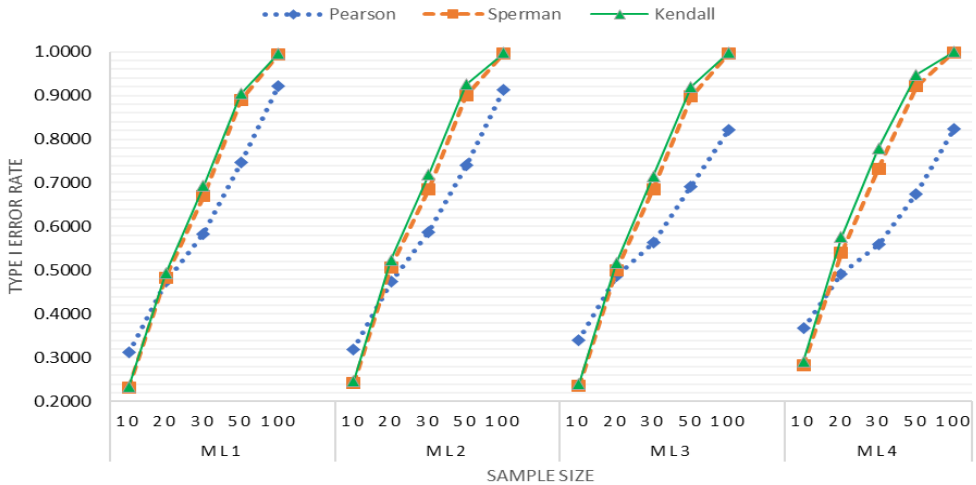

Figure 7. Power of the test when samples are taken from mixed log-normal distribution with $\rho = 0.4$.

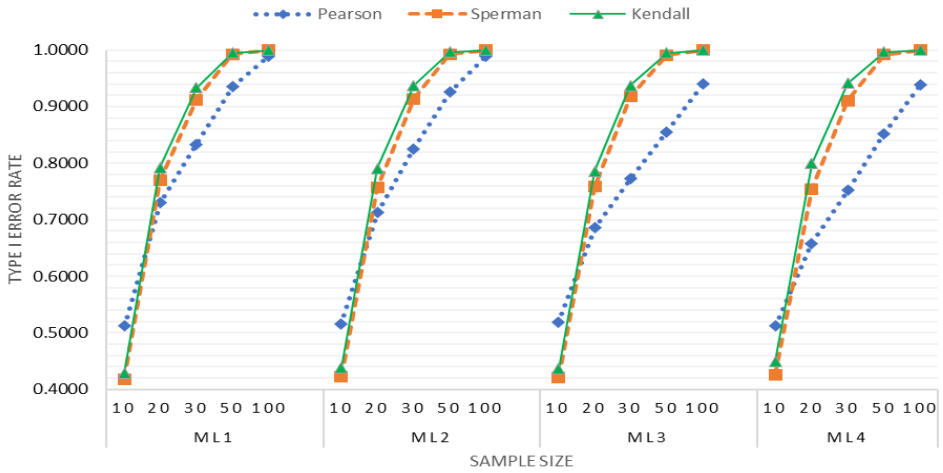

Figure 8. Power of the test when samples are taken from mixed log-normal distribution with $\rho = 0.6$.

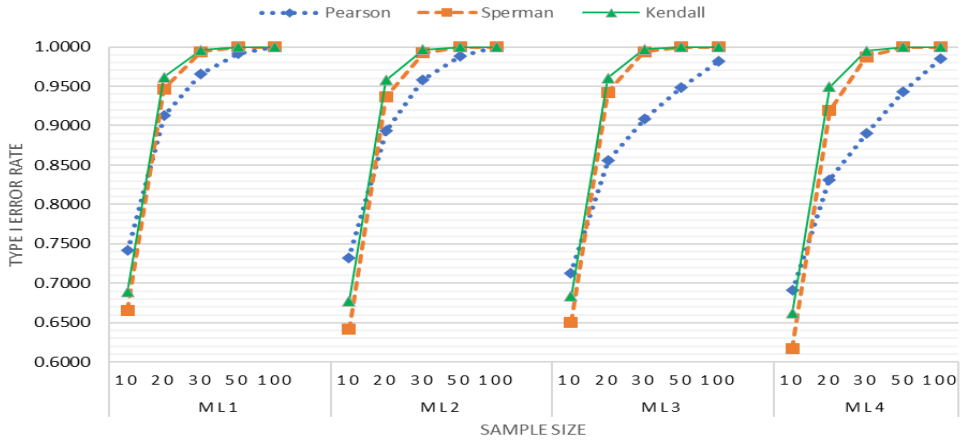

Figure 9. Power of the test when samples are taken from mixed log-normal distribution with $\rho = 0.8$.

สำหรับข้อมูลที่มีการแจกแจงล็อกนอร์มัลแบบผสม (Figure 7 - 9) เมื่อกำหนด $\rho = 0.4$ ขนาดตัวอย่างเท่ากับ 10 การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุด แต่เมื่อมีขนาดตัวอย่างตั้งแต่ 20 ขึ้นไป การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบสูงที่สุด และเมื่อกำหนด $\rho = 0.6, 0.8$ พบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุดเมื่อขนาดตัวอย่างเท่ากับ 10 และเมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 พบว่าการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบสูงที่สุด นอกจากนี้ยังพบว่าเมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมนและค่าสัมประสิทธิ์สหสัมพันธ์เคนดัล ให้ค่ากำลังการทดสอบเท่ากัน

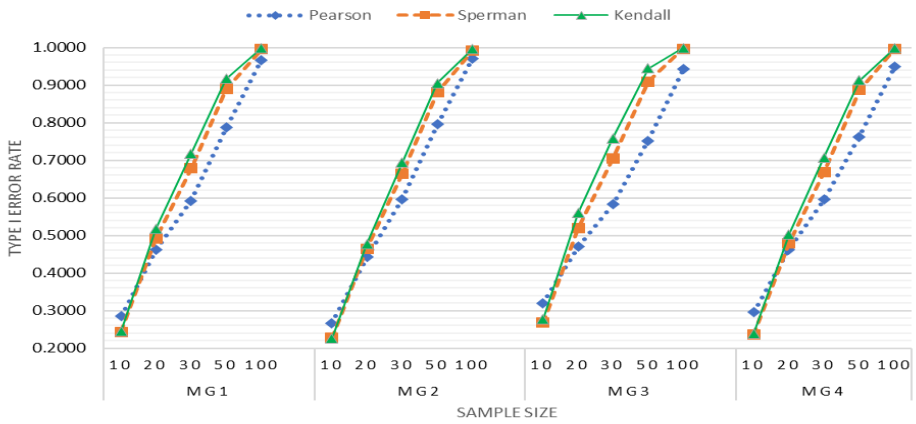

Figure 10. Power of the test when samples are taken from mixed gamma distribution with $\rho = 0.4$.

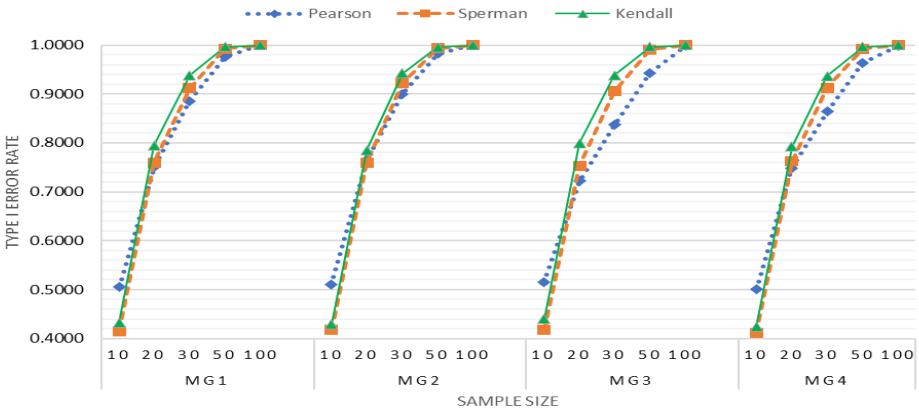

Figure 11. Power of the test when samples are taken from mixed gamma distribution with $\rho = 0.6$.

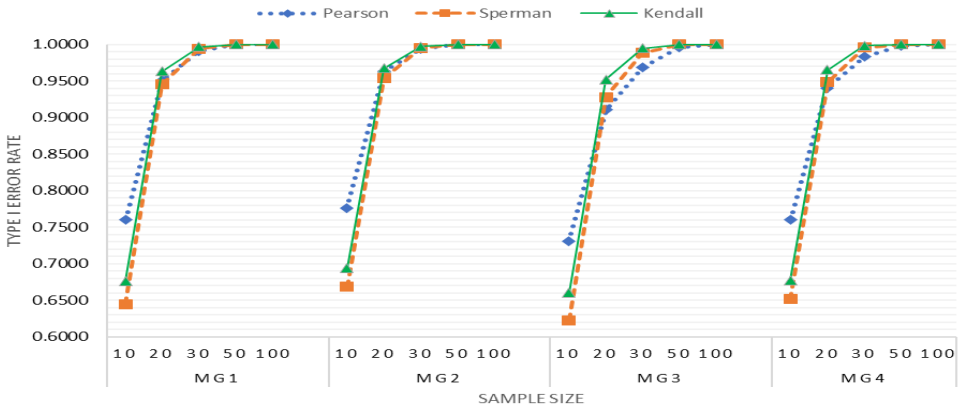

Figure 12. Power of the test when samples are taken from mixed gamma distribution with $\rho = 0.8$.

จาก Figure ที่ 10 - 12 พบว่า สำหรับข้อมูลมีการแจกแจงแกมมาแบบผสม เมื่อกำหนด $\rho = 0.4$ การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุด เมื่อขนาดตัวอย่างเท่ากับ 10 และเมื่อขนาดตัวอย่างเท่ากับ 20, 30, 50 และ 100 พบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบสูงที่สุด

เมื่อกำหนด $\rho = 0.6$ พบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุด เมื่อขนาดตัวอย่างเท่ากับ 10 ในขณะที่การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบดีที่สุด เมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 นอกจากนี้ยังพบว่า เมื่อขนาดตัวอย่างเท่ากับ 100 ค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมนและค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบสูงที่สุด

ในกรณีที่ $\rho = 0.8$ พบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่ากำลังการทดสอบสูงที่สุด เมื่อขนาดตัวอย่างเท่ากับ 10 ในขณะที่การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบดีที่สุด เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 สำหรับการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบเท่ากัน เมื่อขนาดตัวอย่างเท่ากับ 50 นอกจากนี้ยังพบว่า การทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี ให้ค่ากำลังทดสอบเท่ากัน เมื่อขนาดตัวอย่างเท่ากับ 100

อภิปรายผล

จากผลการศึกษาจะเห็นได้ว่าการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 3 วิธี สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ทุกการแจกแจงที่ศึกษา

พิจารณากำลังการทดสอบ พบว่า ทั้ง 3 วิธี มีกำลังการทดสอบเพิ่มมากขึ้น (ลู่เข้าสู่ 1) เมื่อขนาดตัวอย่างใหญ่ขึ้น โดยการทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้ค่าการทดสอบสูงที่สุดเมื่อข้อมูลมีการแจกแจงปรกติแบบผสม สำหรับข้อมูลที่มีการแจกแจงล็อกนอร์มัลแบบผสมและการแจกแจงแกมมาแบบผสม พบว่า กรณีส่วนใหญ่การทดสอบสัมประสิทธิ์สหสัมพันธ์เคนดัลให้ค่ากำลังการทดสอบสูงที่สุด ยกเว้นเมื่อตัวอย่างมีขนาดเล็กการทดสอบสัมประสิทธิ์สหสัมพันธ์เพียร์สันให้กำลังการทดสอบสูงที่สุด อย่างไรก็ตาม เมื่อตัวอย่างมีขนาดใหญ่การทดสอบสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์เคนดัลมีกำลังการทดสอบใกล้เคียงกัน เสมอสอดคล้องกับงานวิจัยของ Tuğran et al.[8] ที่ศึกษากำลังการทดสอบกรณีที่ข้อมูลมีการแจกแจงล็อกนอร์มัลที่มีความสัมพันธ์กัน เนื่องจากข้อมูลมีลักษณะการแจกแจงแบบเบ้เช่นเดียวกับลักษณะข้อมูลที่ศึกษาในงานวิจัยนี้ นอกจากนั้นค่ากำลังการทดสอบจะสูงขึ้นเมื่อตัวอย่างมีขนาดเพิ่มขึ้นสอดคล้องกับงานที่วิจัยหลายงานที่ผ่านมา

โดยวิธีการที่นำมาเปรียบเทียบในงานวิจัยครั้งนี้เป็นวิธีพื้นฐานที่นิยมมาใช้ทดสอบสหสัมพันธ์ เพื่อเป็นแนวทางในการเลือกใช้การทดสอบสัมประสิทธิ์สหสัมพันธ์ที่เหมาะสมกับข้อมูลที่มีลักษณะการแจงแจงผสมที่สัมพันธ์กันซึ่งสอดคล้อง

กับข้อมูลจริงที่กล่าวไว้ในข้างต้นของงานวิจัย อย่างไรก็ตามควรมีการศึกษาวิธีการอื่นเพิ่มเติม หรือข้อมูลที่มีลักษณะการแจกแจงแบบอื่น ๆ ของข้อมูลจริงที่มีความสำคัญต่อไป

เอกสารอ้างอิง

1. Chok NS. Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data [thesis]. Pittsburgh: University of Pittsburgh; 2010.
2. ชนินันท์ พฤกษ์ประมูล, นัฏฐิกา เจริญตะคุ, สิวะโชติ ศรีสุทธิยากร. ประสิทธิภาพของสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สเปียร์แมน และเคนดอลล์ เมื่อข้อมูลแจกแจงแบบไม่ปกติ. OJED 2563;15:1-16.
3. Suhaila J, Ching-Yee K, Fadhilah Y, Hui-Mean F. Introducing the mixed distribution in fitting rainfall data. OJMH 2011;1:11-22. doi:10.4236/ojmh.2011.12002.
4. Sattayatham P, Talangtam T. Fitting of finite mixture distributions to motor insurance claims. J Math Stat 2012;8:49-56.
5. Williford E, Haley V, McNutt LA, Lazariu V. Dealing with highly skewed hospital length of stay distributions: the use of Gamma mixture models to study delivery hospitalizations. PLoS One 2020;15:e0231825. doi:10.1371/journal.pone.0231825.
6. Sprent P, Smeeton NC. Applied nonparametric statistical methods. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC; 2001.
7. Kvam PH, Vidakovic B. Nonparametric statistics with applications to science and engineering. New Jersey, USA: John Wiley & Sons; 2007.
8. Tuğran E, Kocak M, Mirtagioğlu H, Yiğit S, Mendes M. A simulation based comparison of correlation coefficients with regard to type I error rate and power. JDAIP 2015;3:87-101. doi:10.4236/jdaip.2015.33010.
9. Migdadi HS. On the power performance of test statistics for the generalized Rayleigh interval grouped data. OJS 2015;5:474-82. doi:10.4236/ojs.2015.55049.