การเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่เมื่อข้อมูลมีค่าผิดปกติ

Robustness Comparison of the Rank Correlation Coefficients for Data Containing Outliers

ประชาชาติ อารีชาติ*, ชนิกานต์ ตั้งตระกูล และจุฑาภรณ์ สินสมบูรณ์ทอง

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

Parchachart Areechart*, Chanikarn Tangtrakul and Juthaphorn Sinsomboonthong

Department of Statistics, Faculty of Science, Kasetsart University,

Ladyao, Chatuchak, Bangkok 10900

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่เมื่อข้อมูลมีค่าผิดปกติ 4 วิธี คือ สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา และสัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง สัมประสิทธิ์สหสัมพันธ์เบรสท์ ซึ่งเกณฑ์ในการเปรียบเทียบความแกร่ง คือ ค่าความเอนเอียงสัมบูรณ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย โดยจำลองข้อมูลที่มีการแจกแจงปรกติมาตรฐานสองตัวแปร กำหนดระดับความสัมพันธ์ระหว่างสองตัวแปร 6 ระดับ คือ 0, 0.2, 0.4, 0.6, 0.8 และ 1.0 ขนาดตัวอย่าง (n) เท่ากับ 10, 20, 50, 100 และ 200 ร้อยละของค่าผิดปกติจากกลุ่มในตัวแปรเพียงตัวเดียวเท่ากับ 0, 10, 20 และ 30 ของขนาดตัวอย่าง รวมสถานการณ์ที่ศึกษาทั้งหมด 120 สถานการณ์ และมีการทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ผลการวิจัยพบว่าเมื่อมีค่าผิดปกติในข้อมูลและตัวแปรทั้งสองไม่มีความสัมพันธ์กัน ส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งมีความแกร่งสูงสุดเกือบทุกขนาดตัวอย่าง แต่เมื่อตัวแปรทั้งสองมีความสัมพันธ์กันในระดับสูง สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทามีแนวโน้มให้ค่า MSE ต่ำที่สุดในทุกขนาดตัวอย่าง และค่า ABS ของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์เบรสท์ส่วนใหญ่จะให้ค่าต่ำที่สุดในกรณีที่ตัวแปรทั้งสองมีความสัมพันธ์กันในระดับสูงเช่นกัน นอกจากนี้หากตัวแปรทั้งสองมีความสัมพันธ์กันในระดับ 0.6 ถึง 0.8 สัมประสิทธิ์สหสัมพันธ์สเปียร์แมนจะให้ค่า MSE ต่ำที่สุดในทุกขนาดตัวอย่าง

คำสำคัญ : สัมประสิทธิ์สหสัมพันธ์; ค่าผิดปกติ; ความแกร่ง; ค่าความเอนเอียงสัมบูรณ์; ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

***ผู้รับผิดชอบบทความ :** parchachart.a@hotmail.com

Abstract

The objective of this research is to compare the robustness of four rank correlation coefficients–Spearman, Kendall Tau, Blest and Hoeffding correlation coefficients–for data containing outliers. The criteria for this comparison are the absolute bias and mean square error. The simulations data in the form of standard bivariate normal distribution are generated with six levels of the correlation between two variables are set at 0, 0.2, 0.4, 0.6, 0.8 and 1.0. In addition, the sample sizes (n) used in this study are determined at 10, 20, 50, 100 and 200, and the percentages of mild outliers for one variable are set at 0, 10, 20 and 30 of the sample size. The totals of 120 situations are studied. The experiment is repeated 1,000 times for each situation. The conclusions of this research are as follows: when the data are contaminated with outliers and two variables are not correlated, the most robust correlation coefficient for almost all sample sizes is Hoeffding correlation coefficient. However, when two variables are highly correlated, the MSE of Kendall tau correlation coefficient and the ABSs of Spearman and Blest correlation coefficients tend to have a smallest value for all sample sizes. Furthermore, the MSE of Spearman correlation coefficient tends to have a smallest value when the correlation of two variables is between 0.6 and 0.8 for all sample sizes.

Keywords: correlation coefficient; outliers; robustness; absolute bias; mean square error

1. บทนำ

ปัจจุบันพบว่าสถิติเข้ามามีบทบาทในงานวิจัยมากขึ้น กล่าวคือ เป็นเครื่องมือที่ช่วยในการวิเคราะห์ข้อมูล ซึ่งสัมประสิทธิ์สหสัมพันธ์เป็นหนึ่งในค่าสถิติที่ช่วยบอกได้ว่าตัวแปรของข้อมูลนั้นมีความสัมพันธ์กันมากหรือน้อยเพียงใด ในการเก็บรวบรวมข้อมูลที่จะนำมาวิเคราะห์นั้น ผู้วิจัยไม่สามารถเก็บรวบรวมข้อมูลมาได้ทุกหน่วยของประชากรเนื่องจากต้องใช้ค่าใช้จ่ายสูงและใช้ระยะเวลามากในการเก็บข้อมูล จึงต้องมีการเก็บข้อมูลจากตัวอย่างเพื่อนำมาวิเคราะห์ ภายหลังจากการเก็บรวบรวมข้อมูลจากหน่วยตัวอย่างแล้ว จะใช้ผลวิเคราะห์ข้อมูลที่ได้จากการศึกษาข้อมูลตัวอย่างไปอนุมานหรือสรุปผลของประชากร ซึ่งเป็นเป้าหมายหลักของการศึกษาวิจัยต่อไปได้ ในการเก็บรวบรวมข้อมูลบ่อยครั้งที่พบว่าข้อมูลที่มีค่าต่ำเกินไปหรือสูงเกินไปจากข้อมูลส่วนใหญ่จะเรียกว่าค่าผิดปกติ (outlier) ซึ่งสาเหตุของการเกิดค่าผิดปกติ [1] อาจเกิดจากความคลาดเคลื่อนของการวัด (measurement error) หรือความคลาดเคลื่อนจากการปฏิบัติการ (execution error) เช่น การบันทึกข้อมูลผิด รวมไปถึงความคลาดเคลื่อนจากการผันแปรที่มีอยู่ในประชากรที่ศึกษา (inherent variability) ซึ่งไม่ได้เกิดจากการวัดหรือการบันทึกข้อมูลที่ผิดพลาดแต่เป็นความผิดพลาดที่หลีกเลี่ยงไม่ได้เนื่องจากไม่ว่าจะควบคุมการวัดหรือการเก็บข้อมูลได้ดีเพียงใด ความผันแปรในประชากรก็ยังคงมีอยู่ ดังนั้นการนำข้อมูลที่มีค่าผิดปกติมาใช้อาจทำให้การวิเคราะห์ข้อมูลมีความผิดพลาดเกิดขึ้น ซึ่งอาจแก้ปัญหานี้โดยการตัดข้อมูลที่มีค่าผิดปกติออกหรือปรับค่าให้เหมาะสม ซึ่งขึ้นอยู่กับดุลยพินิจของผู้ทำวิจัย [2]

การเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์เมื่อข้อมูลมีค่าผิดปกติปลอมปนนั้นได้มีการศึกษากันอย่างกว้างขวาง โดยวิธีที่นำมาเปรียบเทียบประสิทธิภาพมีหลากหลายวิธี เช่น งานวิจัยของ Croux และ Dehon [3] ได้ศึกษาเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา และสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน ซึ่งเป็นวิธีที่ได้รับความนิยมอย่างแพร่หลายในด้านวิทยาศาสตร์ประยุกต์ โดยพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา และสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนเป็นวิธีที่มีความแกร่งและมีประสิทธิภาพสูง เมื่อข้อมูลมีค่าผิดปกติปลอมปนและ ρ=0 แต่เมื่อพิจารณาค่าความเอนเอียงของสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา Evandt และคณะ [4] พบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทามีค่าความเอนเอียงค่อนข้างมากในเกือบทุกระดับของ ρ ยกเว้น ρ=0 หรือ 1 เมื่อเทียบกับค่าความเอนเอียงของสัมประสิทธิ์สหสัมพันธ์เพียร์สันและสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน ในงานวิจัยของนพดล [5] ได้เปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ 4 วิธี คือ สัมประสิทธิ์สหสัมพันธ์เพียร์สัน สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา และสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนัก เมื่อข้อมูลมีค่าผิดปกติจากกลุ่มทีละตัวแปร โดยพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ผลการศึกษาพบว่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์แบบถ่วงน้ำหนักมีความแกร่งเมื่อข้อมูลมีค่าผิดปกติ สำหรับงานวิจัยของ Devlin และคณะ [6] ศึกษาอิทธิพลของค่าผิดปกติจากกลุ่มที่มีต่อการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ประชากร ผลการศึกษาพบว่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีประสิทธิภาพดีเมื่อข้อมูลไม่มีค่าผิดปกติปลอมปน แต่ถ้าข้อมูลมีค่าผิดปกติปลอมปนเกิดขึ้นจะมีอิทธิพลต่อการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ประชากรอย่างมาก ซึ่งผลกระทบเหล่านี้เกิดขึ้นน้อยเมื่อใช้สถิติแบบไม่อิงพารามิเตอร์ นอกจากนี้ Hoeffding [7] ได้นำเสนอสัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งซึ่งเป็นสถิติแบบไม่อิงพารามิเตอร์ ซึ่งมีประสิทธิภาพดีกว่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันเมื่อข้อมูลมีค่าผิดปกติ และ Blest [8] ได้เสนอสัมประสิทธิ์สหสัมพันธ์เบรสท์ซึ่งเป็นสถิติแบบไม่อิงพารามิเตอร์ โดยพัฒนาวิธีนี้มาจากวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่เป็นที่นิยมอย่างแพร่หลายอย่างสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา ซึ่งเป็นพื้นฐานของการพัฒนาวิธีใหม่เพื่อให้ได้ตัวประมาณที่มีประสิทธิภาพที่ดียิ่งขึ้นกว่าวิธีที่ใช้กันในปัจจุบัน ในการวิเคราะห์ข้อมูลส่วนใหญ่ผู้วิจัยมักจะมองข้ามข้อมูลที่มีค่าผิดปกติ ซึ่งจะส่งผลให้การสรุปผลการวิเคราะห์ข้อมูลมีความคลาดเคลื่อนเกิดขึ้น

ในการวิจัยครั้งนี้ผู้วิจัยสนใจศึกษาเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่ซึ่งเป็นสถิติแบบไม่อิงพารามิเตอร์เมื่อข้อมูลมีค่าผิดปกติจากกลุ่มในตัวแปรเพียงตัวเดียวเท่านั้นโดยอีกตัวแปรไม่มีค่าผิดปกติปลอมปนอยู่เนื่องจากการเกิดค่าผิดปกติปลอมปนในข้อมูลส่วนใหญ่พบว่าสาเหตุอาจจะมาจากความคลาดเคลื่อนของการวัดหรือการบันทึกผล รวมไปถึงความคลาดเคลื่อนจากการผันแปรที่มีอยู่ในประชากรที่ศึกษาดังได้กล่าวถึงในข้างต้น ซึ่งการเกิดเหตุการณ์ดังกล่าวนี้อาจไม่ได้เกิดขึ้นในทั้ง 2 ตัวแปรกล่าวคืออาจเกิดค่าผิดปกติกับตัวแปรใดตัวแปรหนึ่งเท่านั้น จึงทำให้ผู้วิจัยสนใจศึกษาในกรณีเกิดค่าผิดปกติดังกล่าว โดยวิธีที่สนใจศึกษาในที่นี้มี 4 วิธี คือ สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง และสัมประสิทธิ์สหสัมพันธ์เบรสท์ โดยจำลองข้อมูล

เพื่อเปรียบเทียบประสิทธิภาพทั้งหมด 120 สถานการณ์ เพื่อเป็นแนวทางให้แก่ผู้วิจัยในงานด้านต่าง ๆ เลือกใช้วิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ให้เหมาะสมกับข้อมูลที่มีอยู่ และให้ความสำคัญกับค่าผิดปกติของข้อมูลมากขึ้น เพื่อให้ได้ผลลัพธ์ของงานวิจัยที่มีความแม่นยำมากขึ้น

2. วิธีการวิจัย

การเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่เมื่อข้อมูลมีค่าผิดปกติในตัวแปรเพียงตัวเดียว ส่วนอีกตัวแปรไม่มีค่าผิดปกติปลอมปนอยู่ โดยวิธีที่สนใจศึกษาในที่นี้มี 4 วิธี คือ สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง และสัมประสิทธิ์สหสัมพันธ์เบรสท์ ซึ่งขั้นตอนการวิจัยเป็นดังนี้

2.1 จำลองข้อมูลที่มีการแจกแจงปรกติสองตัวแปรด้วยค่าเฉลี่ย (μ_x,μ_y) เท่ากับ (0,0) และความแปรปรวน (σ_x^2,σ_y^2) เท่ากับ (1,1) และกำหนดค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) ระหว่างตัวแปรสองตัวทั้งหมด 6 ระดับ ได้แก่ 0, 0.2, 0.4, 0.6, 0.8 และ 1.0

2.2 กำหนดขนาดตัวอย่าง (n) ทั้งหมด 5 ระดับ คือ 10, 20, 50, 100 และ 200

2.3 กำหนดร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y ในระดับปานกลาง (mild outlier) ทางด้านมาก โดยร้อยละของค่าผิดปกติจากกลุ่มมีทั้งหมด 4 ระดับ คือ 0, 10, 20 และ 30 ของขนาดตัวอย่าง ซึ่งสร้างค่าผิดปกติจากกลุ่มในตัวแปร Y ตามหลักการที่ Barnett และ Lewis [9] กล่าวถึง โดยสร้างค่าของตัวแปร Y ซึ่งผิดปกติอยู่ในช่วง (Q_3 + 1.5 IQR, Q_3 + 3 IQR) เมื่อ Q_1 คือ ค่าควอไทล์ที่ 1; Q_3 คือค่าควอไทล์ที่ 3; และ IQR คือ ระยะระหว่างควอไทล์ซึ่งมีค่าเท่ากับ Q_3 - Q_1

2.4 การตรวจสอบค่าผิดปกตินั้นสามารถตรวจสอบได้หลายวิธี คือ การตรวจสอบโดยวิธีกราฟซึ่งเป็นวิธีการตรวจสอบที่ง่ายและสะดวก หรือการตรวจสอบด้วยการทดสอบทางสถิติซึ่งตัวสถิติทดสอบมีหลายวิธี เช่น ตัวสถิติทดสอบ T_{N1} [9] ที่ใช้ในกรณีไม่ทราบค่าเฉลี่ยและความแปรปรวนของประชากร หรือตัวสถิติทดสอบของ Furguson [10] ที่ใช้สัมประสิทธิ์ความเบ้และสัมประสิทธิ์ความโด่งเป็นพื้นฐานในการทดสอบค่าผิดปกติ สำหรับงานวิจัยนี้ได้ตรวจสอบค่าผิดปกติโดยวิธีกราฟซึ่งตรวจสอบด้วยแผนภาพกล่อง (box plot) โดยค่าสังเกตที่มีค่าผิดปกติในระดับปานกลางจะมีค่ามากกว่า Q_3 ระหว่าง 1.5 ถึง 3 เท่าของ IQR

2.5 ใช้โปรแกรม SAS 9.4 จำลองข้อมูลด้วยเทคนิคมอนติคาร์โล โดยในแต่ละสถานการณ์ทดลองซ้ำ 1,000 ครั้ง

2.6 สหสัมพันธ์ (correlation) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัว ขึ้นไป ในการพิจารณาตัวแปรว่ามีความสัมพันธ์กันมากน้อยเพียงใดนั้น จะพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ซึ่งเป็นวิธีการทางสถิติที่มีอยู่หลายวิธี การใช้สถิติตัวใดจึงขึ้นอยู่กับมาตรวัดของตัวแปร ดังนั้นสัมประสิทธิ์สหสัมพันธ์จึงมีทั้งแบบที่เป็นสถิติอิงพารามิเตอร์และสถิติไม่อิงพารามิเตอร์ กล่าวคือ ค่าสัมประสิทธิ์สหสัมพันธ์ในสถิติอิงพารามิเตอร์เป็นการหาความสัมพันธ์สำหรับตัวแปรที่มีมาตรวัดช่วง (interval scale) หรือมาตรวัดอัตราส่วน (ratio scale) และมีการแจกแจกความน่าจะเป็นปรกติ ซึ่งตัวอย่างสัมประสิทธิ์สหสัมพันธ์ในสถิติอิงพารามิเตอร์ เช่น สัมประสิทธิ์สหสัมพันธ์เพียร์สัน โดยสัมประสิทธิ์สหสัมพันธ์เพียร์สันนี้จะใช้ได้อย่างเหมาะสมกับข้อมูลที่มีความสัมพันธ์เชิงเส้นเท่านั้น และข้อมูลไม่มีค่าผิดปกติปลอมปน ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ใน

สถิติไม่อิงพารามิเตอร์ใช้สำหรับหาความสัมพันธ์ระหว่างตัวแปรที่มีมาตรวัดตั้งแต่นามบัญญัติขึ้นไป และไม่เจาะจงชนิดของการแจกแจงความน่าจะเป็นของข้อมูล เช่นสัมประสิทธิ์สหสัมพันธ์ลำดับที่ 4 วิธี ซึ่งจะศึกษาต่อไปในงานวิจัยนี้ โดยสัมประสิทธิ์สหสัมพันธ์ลำดับที่ใช้วัดความสัมพันธ์ระหว่างลำดับที่ของตัวแปรอันดับ (ordinal variable) ที่แตกต่างกัน ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่จะขึ้นอยู่กับค่าที่จัดอันดับสำหรับแต่ละตัวแปรแทนที่จะเป็นข้อมูลดิบ ดังนั้นสัมประสิทธิ์สหสัมพันธ์ลำดับที่จึงเหมาะกับการวัดระดับความสัมพันธ์ระหว่างลำดับที่ของตัวแปรถึงแม้ข้อมูลจะมีค่าผิดปกติปลอมปนก็ตาม โดยสัมประสิทธิ์สหสัมพันธ์ลำดับที่ ที่ใช้ในการวิจัยนี้ มี 4 วิธี คือ

2.6.1 สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (Spearman correlation coefficient)

Spearman [11] ได้เสนอสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน ซึ่งได้รับการพัฒนามาจากสัมประสิทธิ์สหสัมพันธ์เพียร์สันมาเป็นเวลานานแล้ว แต่ก็ยังเป็นที่รู้จักกันดีในปัจจุบัน โดยใช้ในการหาความสัมพันธ์เมื่อตัวแปรทั้งสองอยู่ในมาตรวัดอันดับ (ordinal scale) หรือเมื่อตัวแปรหนึ่งอยู่ในมาตรวัดอันดับและอีกตัวแปรหนึ่งอยู่ในมาตรวัดช่วง (interval scale) หรือมาตรวัดอัตราส่วน (ratio scale) โดยข้อมูลในแต่ละชุดจะต้องมีความเป็นอิสระต่อกันและตัวแปรทั้งสองมีขนาดเท่ากัน การคำนวณวิธีนี้จะคำนวณจากตำแหน่งที่ของข้อมูลที่มีการเรียงลำดับข้อมูลที่มีค่าสูงสุดถึงข้อมูลที่มีค่าต่ำสุด หรือเรียงลำดับข้อมูลต่ำสุดถึงสูงสุด ถ้าข้อมูลใดมีค่าเท่ากันให้ใช้ลำดับที่เฉลี่ยของข้อมูลเป็นลำดับที่ของข้อมูลค่านั้น

สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน คำนวณได้จากสูตร

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(p_i - q_i)^2}{n^3 - n}$$

โดย p_i = ลำดับของ x_i; q_i = ลำดับของ y_i; n = ขนาดตัวอย่าง

2.6.2 สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา (Kendall Tau correlation coefficient)

Kendall [12] ได้นำเสนอสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา ใช้วัดความสัมพันธ์ระหว่างข้อมูล 2 ชุด ที่อยู่ในระดับอันดับ หรือเมื่อตัวแปรหนึ่งอยู่ในมาตรวัดอันดับและอีกตัวหนึ่งอยู่ในมาตรวัดช่วงหรือมาตรวัดอัตราส่วน โดยข้อมูลทั้งสองชุดมาจากการพิจารณาจัดอันดับสิ่งเดียวกัน

สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา สามารถคำนวณได้ดังนี้

$$r_k = \frac{2S}{n(n-1)}$$

เมื่อ S = ผลต่างระหว่างผลบวกของอันดับที่ที่สอดคล้องกัน (concordant pairs) กับผลบวกของอันดับที่ที่ไม่สอดคล้องกัน (discordant pairs); $S = C\text{-}D$; C = จำนวนอันดับที่มีค่าสูงกว่าเมื่อนับจากจำนวนที่อยู่ใต้ลงมา; D = จำนวนอันดับที่มีค่าต่ำกว่าเมื่อนับจากจำนวนที่อยู่ใต้ลงมา; n = ขนาดตัวอย่าง

2.6.3 สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง (Hoeffding correlation coefficient)

Hoeffding [7] ได้นำเสนอสัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง ซึ่งเป็นสถิติไม่อิงพารามิเตอร์ข้อมูลแต่ละชุดมีความเป็นอิสระกัน โดยกำหนดค่า X และ Y สำหรับการจัดกลุ่มของข้อมูล

สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง คำนวณได้จากสูตร

$$r_h = 30\frac{(n-2)(n-3)D_1 + D_2 - 2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)}$$

โดยที่ $D_1 = \sum_{i=1}^{n}(r_i-1)(r_i-2)$; $D_2 = \sum_{i=1}^{n}(p_i-1)(p_i-2)(q_i-1)(q_i-2)$; $D_3 = \sum_{i=1}^{n}(p_i-2)(q_i-1)$; p_i = ลำดับของ x_i; q_i = ลำดับของ y_i; r_i = ลำดับของสองตัวแปร

2.6.4 สัมประสิทธิ์สหสัมพันธ์เบรสท์ (Blest correlation coefficient)

Blest [8] ได้นำเสนอสัมประสิทธิ์สหสัมพันธ์เบรสท์ ซึ่งเป็นสัมประสิทธิ์สหสัมพันธ์ที่ใช้วัดความสัมพันธ์ระหว่างตัวแปร 2 ตัว โดยข้อมูลอยู่ในมาตรวัดอันดับ และพิจารณากรณีลำดับที่ไม่ซ้ำกัน (tied rank) เท่านั้น โดย Blest พัฒนาสัมประสิทธิ์สหสัมพันธ์วิธีนี้มาจากสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน และสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา

สัมประสิทธิ์สหสัมพันธ์เบรสท์ คำนวณได้จากสูตร

$$r_b = 1 - \frac{2W}{W_{\max}} = 1 - \frac{24W}{n(n+1)^2(n-1)}$$

โดยที่ $W_{\max} = \frac{n(n+1)^2(n-1)}{12}$;

$$W = \frac{1}{2}\sum_{i=1}^{n}(n+1-i)^2 q_i - \frac{n(n+1)^2(n+2)}{24}$$

เมื่อ i = ลำดับของ x_i; q_i = ลำดับของ y_i; n = ขนาดตัวอย่าง

2.7 ความแกร่งของตัวประมาณค่า

เกณฑ์การตัดสินว่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่แบบใดที่มีความแกร่งในการประมาณค่า จะพิจารณาจากค่าความเอนเอียงสัมบูรณ์ (absolute bias, ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error, MSE) ของตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี ซึ่งตัวประมาณค่าวิธีใดที่ให้ค่าความเอนเอียงสัมบูรณ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด จะถือว่าเป็นสัมประสิทธิ์สหสัมพันธ์ลำดับที่ที่มีความแกร่ง

ค่าความเอนเอียงสัมบูรณ์ สามารถประมาณได้จากสูตร

$$\widehat{ABS} = \left| \frac{\sum_{j=1}^{m} r_j}{m} - \rho \right|$$

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสามารถประมาณได้จากสูตร

$$\widehat{MSE} = \frac{1}{m}\sum_{j=1}^{m}(r_j - \rho)^2$$

เมื่อ r_j = ค่าประมาณสัมประสิทธิ์สหสัมพันธ์ในการทำซ้ำรอบที่ j ในแต่ละสถานการณ์; ρ = ค่าพารามิเตอร์สัมประสิทธิ์สหสัมพันธ์ที่กำหนด; m = จำนวนรอบที่ทำซ้ำในที่นี้เท่ากับ 1,000

3. ผลการวิจัย

งานวิจัยนี้ได้ทำการเปรียบเทียบความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่ 4 วิธี คือ สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง และสัมประสิทธิ์สหสัมพันธ์เบรสท์ เมื่อข้อมูลมีค่าผิดปกติจากกลุ่มในตัวแปรเพียงตัวเดียว ส่วนอีกตัวแปรไม่มีค่าผิดปกติปลอมปนอยู่ โดยเกณฑ์ในการเปรียบเทียบประสิทธิภาพ ได้แก่ ค่าความเอนเอียงสัมบูรณ์ (ABS) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ซึ่งวิธีใดที่ให้ค่า ABS และ MSE ต่ำกว่าจะมีประสิทธิภาพมากกว่า และถือว่าเป็นสัมประสิทธิ์สหสัมพันธ์ลำดับที่ที่มีความแกร่ง

จำลองข้อมูลประชากร (X,Y) ที่มีการแจกแจงปรกติสองตัวแปรด้วยค่าเฉลี่ย (μ_x,μ_y) เท่ากับ (0,0) และความแปรปรวน (σ_x^2,σ_y^2) เท่ากับ (1,1) กำหนดระดับความสัมพันธ์ระหว่าง X และ Y ทั้งหมด 6 ระดับ คือ 0, 0.2, 0.4, 0.6, 0.8 และ 1.0 โดยใช้เทคนิคมอนติคาร์โล ด้วยโปรแกรม SAS 9.4 จากนั้นสุ่มตัวอย่างขนาดเท่ากับ 10, 20, 50, 100 และ 200 จากประชากรดังกล่าว นอกจากนี้ยังกำหนดร้อยละค่าผิดปกติจากกลุ่มในตัวแปร Y ในระดับปานกลาง (mild outlier) ทางด้านมากในระดับที่แตกต่างกันตั้งแต่ร้อยละ 0, 10, 20 และ 30 ของขนาดตัวอย่าง ในงานวิจัยนี้ได้

ตรวจสอบค่าผิดปกติของข้อมูลที่จำลองทุกสถานการณ์ โดยใช้แผนภาพกล่อง (box plot) โดยค่าสังเกตที่มีค่าผิดปกติในระดับปานกลางจะมีค่ามากกว่า Q_3 ระหว่าง 1.5 ถึง 3 เท่าของ IQR รวมสถานการณ์ที่จำลองข้อมูลทั้งหมด 120 สถานการณ์และในแต่ละสถานการณ์ทดลองซ้ำ 1,000 ครั้ง ผลการวิจัยจำแนกตามร้อยละของค่าผิดปกติในตัวแปร Y ของขนาดตัวอย่าง แสดงในตารางที่ 1 ถึงตารางที่ 4

เมื่อพิจารณาจากตารางที่ 1 คือ ร้อยละของค่าผิดปกติในตัวแปร Y เท่ากับ 0 กรณีที่ข้อมูลไม่มีความสัมพันธ์กัน กล่าวคือพารามิเตอร์ ρ มีค่าเท่ากับ 0 และตัวอย่างมีขนาดเล็ก (n=10) สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา จะมีแนวโน้มให้ทั้งค่า MSE และ ABS ต่ำที่สุด แต่เมื่อตัวอย่างมีขนาดใหญ่ขึ้น (n=50, 100 และ 200) สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งจะมีแนวโน้มให้ทั้งค่า MSE และ ABS ต่ำกว่าวิธีอื่นอย่างชัดเจน กรณีที่ข้อมูลมีความสัมพันธ์กัน กล่าวคือ พารามิเตอร์ ρ มีค่าตั้งแต่ 0.2 ถึง 0.8 ในทุกขนาดตัวอย่าง พบว่าส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์เบรสท์และสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนมีแนวโน้มให้ทั้งค่า MSE และ ABS ต่ำกว่าวิธีอื่น แต่จะมีประสิทธิภาพใกล้เคียงกัน นอกจากนี้ในบางกรณีพบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทามีค่า MSE ต่ำที่สุดแต่ค่า ABS จะมีค่าค่อนข้างสูงกว่าของสัมประสิทธิ์สหสัมพันธ์เบรสท์และสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนดังเช่นกรณีที่ (n=10 และ ρ =0.4) หรือ (n=20 และ ρ =0.2, 0.4) หรือ (n=50 และ ρ =0.2) สำหรับกรณีที่ข้อมูลมีความสัมพันธ์กันในระดับสูง กล่าวคือ พารามิเตอร์ ρ มีค่าเท่ากับ 1 ในทุกขนาดตัวอย่าง พบว่าแนวโน้มทั้งค่า MSE และ ABS ของ 4 วิธี จะมีค่าต่ำที่สุดเท่ากัน โดยมีค่าเท่ากับ 0

เมื่อพิจารณาจากตารางที่ 2 ถึงตารางที่ 4 คือ ค่าผิดปกติในตัวแปร Y เกิดขึ้นร้อยละ 10, 20 และ 30 ของขนาดตัวอย่าง จะพบว่าแนวโน้มค่าของ MSE และ ABS ของทั้ง 4 วิธี ส่วนใหญ่เป็นไปในทิศทางเดียวกันทั้ง 3 ตาราง ดังนี้ กรณีที่ข้อมูลไม่มีความสัมพันธ์กัน กล่าวคือพารามิเตอร์ ρ มีค่าเท่ากับ 0 ส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งจะมีแนวโน้มให้ทั้งค่า MSE และ ABS ต่ำกว่าวิธีอื่น โดยเฉพาะอย่างยิ่งเมื่อขนาดตัวอย่างตั้งแต่ 50 ขึ้นไปในทุกระดับของค่าผิดปกติในตัวแปร Y ที่ปลอมปนในข้อมูล กรณีที่ข้อมูลมีความสัมพันธ์กัน กล่าวคือ พารามิเตอร์ ρ มีค่าตั้งแต่ 0.2 ถึง 0.8 ในทุกขนาดตัวอย่าง ส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์เบรสท์และสัมประสิทธิ์สหสัมพันธ์สเปียร์-แมน จะมีแนวโน้มให้ค่า MSE และ ABS ต่ำกว่าวิธีอื่น โดยจะมีประสิทธิภาพใกล้เคียงกัน ยกเว้นในบางกรณีพบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทามีค่า MSE ต่ำที่สุดแต่ค่า ABS จะมีค่าค่อนข้างสูงกว่าของสัมประสิทธิ์สหสัมพันธ์เบรสท์และสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนดังเช่นกรณีที่ (n=10 และ ρ = 0.4) หรือ (n=20 และ ρ =0.2) หรือ (n=50 และ ρ = 0.2) โดยมีค่าผิดปกติในตัวแปร Y ปลอมปนอยู่ร้อยละ 10 เป็นต้น นอกจากนี้ในบางกรณียังพบว่าค่า MSE ของสัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง ต่ำที่สุดจริงแต่ค่า ABS กลับมีค่าสูงที่สุดดังเช่นกรณีที่ (n=10 และ ρ = 0.2)) โดยมีค่าผิดปกติในตัวแปร Y ปลอมปนอยู่ร้อยละ 10 สำหรับกรณีร้อยละค่าผิดปกติอื่น ๆ เป็นไปในทำนองเดียวกันดังปรากฏในตารางที่ 3 และ 4 สำหรับกรณีที่ข้อมูลมีความสัมพันธ์กันในระดับสูง กล่าวคือ พารามิเตอร์ ρ มีค่าเท่ากับ 1.0 พบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทาให้ค่า MSE ต่ำที่สุดในขณะที่ค่า ABS ไม่ได้มีค่าต่ำที่สุดโดยส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์สเปียร์แมนหรือสัมประสิทธิ์สหสัมพันธ์เบรสท์จะให้ค่า ABS ต่ำที่สุดในสถานการณ์ดังกล่าวนี้ดังปรากฏผลในตารางที่ 2 ถึงตารางที่ 4

ตารางที่ 1 ค่า MSE และ ABS เมื่อร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y เท่ากับ 0

n	ρ	สเปียร์แมน		เคนดอลล์เทา		โฮฟดิ้ง		เบรสท์	
		MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS
10	0	0.1131	0.0052	0.0615*	0.0021*	0.1168	0.0044	0.1168	0.0023
	0.2	0.1058	0.0327	0.0649	0.0773	0.0451*	0.1937	0.1101	0.0305*
	0.4	0.0937	0.0575	0.0752*	0.1434	0.1419	0.3585	0.0992	0.0553*
	0.6	0.0688*	0.0683	0.0816	0.1912	0.2617	0.4868	0.0732	0.0652*
	0.8	0.0346*	0.0672	0.0727	0.2099	0.3358	0.5456	0.0376	0.0653*
	1	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
20	0	0.0565	0.0026	0.028	0.0001*	0.0010*	0.0006	0.059	0.0025
	0.2	0.0533	0.0238	0.0325*	0.0752	0.0381	0.1904	0.0559	0.0235*
	0.4	0.0445	0.0386	0.0436*	0.1407	0.1315	0.3565	0.0468	0.0383*
	0.6	0.0314*	0.0461	0.0561	0.1922	0.247	0.4878	0.0332	0.0459*
	0.8	0.0136*	0.0432	0.0554	0.21	0.3114	0.5451	0.0144	0.0428*
	1	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
50	0	0.0203	0.0076	0.0095	0.0046	0.0001*	0.0001*	0.0218	0.009
	0.2	0.0195	0.0211*	0.0152*	0.0777	0.0367	0.1907	0.0211	0.0222
	0.4	0.0168*	0.0317*	0.029	0.1441	0.1292	0.3578	0.0181	0.0327
	0.6	0.0119*	0.0350*	0.0449	0.1958	0.243	0.4904	0.0131	0.036
	0.8	0.0053*	0.0295*	0.0496	0.2134	0.3054	0.5485	0.0058	0.0302
	1	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
100	0	0.0102	0.004	0.0046	0.0023	0.0000*	0.0001*	0.0111	0.0044
	0.2	0.0097*	0.0154*	0.01	0.0748	0.0363	0.1902	0.0106	0.0156
	0.4	0.0083*	0.0243*	0.024	0.1416	0.1281	0.3572	0.0091	0.0245
	0.6	0.0058*	0.0274*	0.0406	0.1938	0.2406	0.4892	0.0063	0.0274*
	0.8	0.0025*	0.0222*	0.0468	0.2121	0.3019	0.5475	0.0027	0.0223
	1	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
200	0	0.005	0.0035	0.0022	0.0023	0.0000*	0.0001*	0.0053	0.003
	0.2	0.0049*	0.0132	0.0077	0.0743	0.0362	0.1901	0.0052	0.0127*
	0.4	0.0043*	0.0206	0.0217	0.1405	0.1275	0.3567	0.0045	0.0201*
	0.6	0.0030*	0.023	0.0385	0.1924	0.2389	0.4882	0.0032	0.0226*
	0.8	0.0013*	0.0183	0.0454	0.211	0.2995	0.5463	0.0014	0.0180*
	1	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*

หมายเหตุ : * คือ กรณีที่สัมประสิทธิ์สหสัมพันธ์ลำดับที่ให้ค่า MSE และ ABS ต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 2 ค่า MSE และ ABS เมื่อร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y เท่ากับ 10

n	ρ	สเปียร์แมน		เคนดอลล์เทา		โฮฟดิ้ง		เบรสท์	
		MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS
10	0	0.1139	0.003	0.0621	0.0022*	0.0062*	0.0028	0.1183	0.0047
	0.2	0.1134	0.057	0.0706	0.0962	0.0449*	0.1934	0.1207	0.0555*
	0.4	0.1132	0.1184	0.0941*	0.1895	0.1468	0.3668	0.1232	0.1168*
	0.6	0.1093*	0.1655	0.1231	0.2667	0.286	0.5147	0.1214	0.1632*
	0.8	0.0970*	0.202	0.1415	0.319	0.4101	0.6114	0.1106	0.2005*
	1	0.0675	0.1899	0.0533*	0.1925	0.1255	0.291	0.0834	0.1897*
20	0	0.055	0.0012	0.0272	0.0004	0.0009*	0.0007	0.0582	0.0003*
	0.2	0.0555	0.0552	0.0360*	0.0978	0.0385	0.1923	0.0599	0.0534*
	0.4	0.0561*	0.1045	0.0593	0.1885	0.1386	0.368	0.0614	0.1026*
	0.6	0.0577*	0.1491	0.0923	0.2686	0.2752	0.5185	0.0638	0.1472*
	0.8	0.0563*	0.1842	0.1173	0.3209	0.3895	0.6156	0.0626	0.1822*
	1	0.0499	0.1886	0.0433*	0.1915	0.0924	0.2825	0.057	0.1866*
50	0	0.0193	0.0027	0.009	0.0012	0.0001*	0.0002*	0.0209	0.0043
	0.2	0.0215	0.0525*	0.0187*	0.099	0.0376	0.1931	0.0238	0.0541
	0.4	0.0263*	0.0993*	0.0447	0.1908	0.1371	0.369	0.0292	0.1007
	0.6	0.0328*	0.1406*	0.0807	0.2711	0.2722	0.5199	0.0363	0.1421
	0.8	0.0394*	0.1748*	0.1103	0.3238	0.3839	0.6166	0.0429	0.176
	1	0.0412	0.1889*	0.0383*	0.1894	0.0785	0.2727	0.0444	0.1893
100	0	0.0103	0.0028	0.0046	0.0016	0.0000*	0.0000*	0.0111	0.0033
	0.2	0.0121*	0.0503*	0.0141	0.0985	0.0373	0.1929	0.0132	0.0505
	0.4	0.0172*	0.0956*	0.0403	0.1907	0.1366	0.3691	0.0185	0.0956*
	0.6	0.0246*	0.1361	0.0767	0.271	0.271	0.5198	0.0259	0.1359*
	0.8	0.0330*	0.1701	0.1074	0.3241	0.3822	0.6169	0.0343	0.1699*
	1	0.0392	0.1905	0.0376*	0.1907	0.077	0.2736	0.0406	0.1902*
200	0	0.0052	0.0032	0.0023	0.0022	0.0000*	0.0000*	0.0055	0.0028
	0.2	0.0074*	0.0491	0.012	0.0985	0.0372	0.1928	0.0078	0.0487*
	0.4	0.0131*	0.0931	0.0383	0.1901	0.136	0.3686	0.0135	0.0927*
	0.6	0.0211*	0.1329	0.0747	0.2701	0.2696	0.5188	0.0216	0.1326*
	0.8	0.0301*	0.167	0.1057	0.3231	0.3797	0.6155	0.0308	0.1667*
	1	0.0376	0.1901*	0.0368*	0.1902	0.0752	0.2724	0.0384	0.1901*

หมายเหตุ : * คือ กรณีที่สัมประสิทธิ์สหสัมพันธ์ลำดับที่ให้ค่า MSE และ ABS ต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 3 ค่า MSE และ ABS เมื่อร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y เท่ากับ 20

n	ρ	สเปียร์แมน		เคนดอลล์เทา		โฮฟดิ้ง		เบรสท์	
		MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS
10	0	0.1093	0.0006*	0.0608	0.0007	0.0062*	0.0014	0.1133	0.0045
	0.2	0.1159	0.0914	0.0756	0.1211	0.0457*	0.1959	0.1232	0.0866*
	0.4	0.134	0.1829	0.1145*	0.2384	0.1539	0.3795	0.1455	0.1784*
	0.6	0.1575*	0.2639	0.1696	0.3429	0.3143	0.547	0.1726	0.2593*
	0.8	0.1832*	0.3336	0.2255	0.4263	0.4848	0.6795	0.2019	0.3304*
	1	0.194	0.3697	0.1695*	0.3743	0.3298	0.5356	0.2177	0.3671*
20	0	0.0525	0.0037	0.0262	0.002	0.0009*	0.0000*	0.0559	0.0028
	0.2	0.0588	0.0914	0.0408	0.1234	0.0397*	0.1957	0.0641	0.0905*
	0.4	0.0758*	0.1737	0.0805	0.238	0.1465	0.3795	0.0831	0.1728*
	0.6	0.1035*	0.2521	0.141	0.3453	0.3039	0.5471	0.1126	0.2514*
	0.8	0.1368*	0.3234	0.2032	0.4302	0.4675	0.6779	0.1478	0.3231*
	1	0.1624	0.3703	0.1494*	0.3711	0.2772	0.5138	0.175	0.3701*
50	0	0.0207	0.0098	0.0096	0.0059	0.0001*	0.0000*	0.0219	0.0112
	0.2	0.0281	0.0898*	0.0249*	0.1245	0.0385	0.1956	0.0302	0.0911
	0.4	0.0469*	0.1685*	0.0662	0.2392	0.1446	0.3794	0.0501	0.1698
	0.6	0.0748*	0.2424*	0.1266	0.3441	0.299	0.5455	0.0788	0.2437
	0.8	0.1094*	0.3109*	0.1895	0.4274	0.457	0.674	0.1141	0.3122
	1	0.1433	0.3649	0.1372*	0.3642*	0.2522	0.4974	0.1485	0.3656
100	0	0.01	0.0013	0.0046	0.0006	0.0000*	0.0001*	0.0106	0.0016
	0.2	0.0164*	0.0805*	0.0187	0.119	0.0381	0.195	0.0175	0.0809
	0.4	0.0343*	0.1585*	0.059	0.2337	0.1432	0.378	0.0359	0.1589
	0.6	0.0619*	0.2326*	0.1189	0.3392	0.2958	0.5433	0.064	0.233
	0.8	0.0973*	0.3018*	0.1824	0.4233	0.4514	0.671	0.0997	0.3022
	1	0.1353	0.3607*	0.1327*	0.3611	0.2441	0.4918	0.138	0.361
200	0	0.0046	0.004	0.0021	0.0028	0.0000*	0.0001*	0.0049	0.0038
	0.2	0.0113*	0.0823	0.0167	0.1208	0.0381	0.1951	0.0118	0.0822*
	0.4	0.0294*	0.1589*	0.0573	0.2351	0.1432	0.3782	0.0302	0.1589*
	0.6	0.0575*	0.2324*	0.1176	0.3403	0.2956	0.5434	0.0585	0.2325
	0.8	0.0935*	0.3011*	0.1814	0.4242	0.4511	0.6712	0.0947	0.3014
	1	0.1329	0.3613	0.1314*	0.3611*	0.2417	0.4906	0.1343	0.3617

หมายเหตุ : * คือ กรณีที่สัมประสิทธิ์สหสัมพันธ์ลำดับที่ให้ค่า MSE และ ABS ต่ำที่สุดในแต่ละสถานการณ์

ตารางที่ 4 ค่า MSE และ ABS เมื่อร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y เท่ากับ 30

n	ρ	สเปียร์แมน		เคนดอลล์เทา		โฮฟดิ้ง		เบรสท์	
		MSE	ABS	MSE	ABS	MSE	ABS	MSE	ABS
10	0	0.1069	0.0065	0.0586	0.0024	0.0062*	0.0054	0.112	0.0021*
	0.2	0.1193	0.1237	0.0774	0.1433	0.0479*	0.204	0.1286	0.1202*
	0.4	0.1569	0.238	0.1332*	0.2786	0.1616	0.3924	0.1703	0.2357*
	0.6	0.2128*	0.3462	0.2173	0.4052	0.3366	0.5705	0.2306	0.3441*
	0.8	0.2882*	0.4517	0.3197	0.5206	0.5475	0.7291	0.3105	0.4508*
	1	0.3488	0.5267*	0.3170*	0.5294	0.5374	0.7133	0.3775	0.5269
20	0	0.054	0.001	0.0272	0.0023	0.0011*	0.0008*	0.0577	0.0017
	0.2	0.0657	0.112	0.0455	0.1367	0.0399*	0.196	0.0718	0.1109*
	0.4	0.1002	0.2233	0.0996*	0.2719	0.1503	0.385	0.1086	0.2225*
	0.6	0.1560*	0.3303	0.1846	0.4002	0.3198	0.5622	0.1671	0.3299*
	0.8	0.2284*	0.4328	0.2851	0.5131	0.5196	0.7168	0.2413	0.4327*
	1	0.3053	0.5184*	0.2865*	0.5187	0.4743	0.6806	0.3205	0.5185
50	0	0.0204	0.0052	0.0095	0.003	0.0001*	0.0001*	0.0216	0.0044
	0.2	0.0332	0.114	0.0293*	0.141	0.0389	0.197	0.0353	0.1131*
	0.4	0.0683*	0.222	0.0852	0.276	0.1494	0.3861	0.0712	0.2213*
	0.6	0.1235*	0.326	0.1715	0.4039	0.3186	0.5638	0.1274	0.3255*
	0.8	0.1965*	0.4259	0.2731	0.5154	0.5152	0.7168	0.2014	0.4257*
	1	0.2777	0.5142	0.2708*	0.5146	0.4496	0.6681	0.2831	0.5137*
100	0	0.01	0.0033	0.0045	0.0021	0.0000*	0.0001*	0.0105	0.0039
	0.2	0.0223*	0.1111*	0.024	0.1396	0.0387	0.1967	0.0234	0.1116
	0.4	0.0569*	0.2177*	0.0798	0.2746	0.1488	0.3856	0.0585	0.2182
	0.6	0.1119*	0.3214*	0.166	0.4023	0.3169	0.5626	0.1141	0.3219
	0.8	0.1847*	0.4209*	0.2672	0.5133	0.5115	0.7146	0.1874	0.4213
	1	0.2695	0.5124	0.2655*	0.5123*	0.4399	0.6619	0.2727	0.5128
200	0	0.005	0.001	0.0023	0.0006	0.0000*	0.0001*	0.0054	0.0001*
	0.2	0.0165*	0.1075	0.0212	0.1377	0.0387	0.1966	0.0169	0.1063*
	0.4	0.0499*	0.2126	0.0761	0.2719	0.1484	0.3852	0.0503	0.2116*
	0.6	0.1037*	0.3153	0.1613	0.3991	0.3156	0.5616	0.1043	0.3144*
	0.8	0.1761*	0.4151	0.2622	0.5103	0.5088	0.7131	0.1767	0.4143*
	1	0.2622	0.5087	0.2614*	0.5098	0.435	0.6589	0.263	0.5079*

หมายเหตุ : * คือ กรณีที่สัมประสิทธิ์สหสัมพันธ์ลำดับที่ให้ค่า MSE และ ABS ต่ำที่สุดในแต่ละสถานการณ์

นอกจากนี้เมื่อร้อยละของค่าผิดปกติจากกลุ่มมีค่าเพิ่มขึ้น ค่า MSE และ ABS ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี จะมีแนวโน้มเพิ่มขึ้นในทุกขนาดตัวอย่างและ ρ โดยเฉพาะอย่างยิ่ง เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่ามากจะสังเกตเห็นแนวโน้มการเพิ่มขึ้นของ MSE และ ABS อย่างชัดเจน กรณีที่ร้อยละของค่าผิดปกติจากกลุ่มและสัมประสิทธิ์สหสัมพันธ์คงที่ เมื่อตัวอย่างมีขนาดใหญ่ขึ้น ค่า MSE และ ABS ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี จะมีแนวโน้มลดลง กรณีร้อยละของค่าผิดปกติจากกลุ่มและขนาดตัวอย่างคงที่ เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าเพิ่มขึ้น ค่า MSE และ ABS ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี จะมีแนวโน้มเพิ่มขึ้น

4. วิจารณ์

เมื่อค่าสัมประสิทธิ์สหสัมพันธ์มีค่าระหว่าง 0.6 ถึง 0.8 ในทุกขนาดตัวอย่างและทุกขนาดร้อยละของค่าผิดปกติในตัวแปร Y สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน จะเป็นตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ที่มีความแกร่งในด้านค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งสอดคล้องกับงานวิจัยของนพดล [5] และยังให้ประสิทธิภาพใกล้เคียงกับสัมประสิทธิ์สหสัมพันธ์เบรสท์ เนื่องจาก Blest [8] ได้พัฒนาสัมประสิทธิ์สหสัมพันธ์เบรสท์มาจากสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนและสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา เพื่อให้ได้วิธีใหม่ที่มีประสิทธิภาพดีกว่าวิธีทั้งสองดังกล่าวนี้ นอกจากนี้ผลการศึกษาพบว่าเมื่อ ρ=0 และทุกขนาดร้อยละของค่าผิดปกติในตัวแปร Y สัมประสิทธิ์เคนดอลล์เทาให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยในระดับต่ำ ทุกขนาดตัวอย่าง n ซึ่งสอดคล้องกับการศึกษาของ Croux และ Dehon [3] แต่เมื่อพิจารณาจากค่า ABS ของสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทาในกรณีที่ ρ มีค่าตั้งแต่ 0.2 ถึง 0.8 และทุกขนาดร้อยละของค่าผิดปกติในตัวแปร Y พบว่ามีค่า ABS ค่อนข้างมากกว่าของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนทุกขนาดตัวอย่าง n ซึ่งสอดคล้องกับการศึกษาของ Evandt และคณะ [4]

5. สรุปผลการวิจัย

เกณฑ์ที่ใช้ในการพิจารณาความแกร่งของสัมประสิทธิ์สหสัมพันธ์ลำดับที่ คือ พิจารณาจากค่าความเอนเอียงสัมบูรณ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย โดยสัมประสิทธิ์สหสัมพันธ์ลำดับที่วิธีใดให้ค่าความเอนเอียงสัมบูรณ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดจะถือว่าวิธีนั้นมีความแกร่งสูงที่สุดในสถานการณ์นั้น ๆ ซึ่งสัมประสิทธิ์สหสัมพันธ์ลำดับที่ที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าความเอนเอียงสัมบูรณ์ต่ำที่สุด สำหรับสถานการณ์ต่าง ๆ สามารถสรุปได้ดังตารางที่ 5 ซึ่งสัมประสิทธิ์สหสัมพันธ์ลำดับที่วิธีใดที่ให้ทั้งค่าความเอนเอียงสัมบูรณ์และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเป็นสิ่งที่พึงปรารถนาอย่างยิ่งสำหรับการตัดสินใจเลือกใช้วิธีการดังกล่าวเพื่อหาสหสัมพันธ์ระหว่างลำดับที่ของตัวแปรต่อไป

จากตารางที่ 5 พบว่าขนาดตัวอย่าง (n) ตั้งแต่ 50 ขึ้นไปและข้อมูลไม่มีค่าผิดปกติปลอมปนหรือมีค่าผิดปกติในตัวแปร Y ปลอมปนร้อยละ 10 เมื่อค่าพารา-มิเตอร์ ρ เท่ากับ 0 สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งมีแนวโน้มเป็นตัวประมาณที่มีความแกร่งที่สุด กล่าวคือ สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งให้ทั้งค่า MSE และ ABS ต่ำที่สุด แต่เมื่อร้อยละของค่าผิดปกติในตัวแปร Y เพิ่มขึ้นเป็นร้อยละ 20 ถึง 30 และขนาดตัวอย่างตั้งแต่ 20 ขึ้นไป สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งยังคงมีแนวโน้มเป็นตัวประมาณที่มีความแกร่งที่สุด โดยพิจารณาจากค่า MSE และ ABS ที่มีค่าต่ำที่สุดทั้งคู่เช่นกัน แต่ถ้าขนาดตัวอย่างมีค่าน้อย (n = 10) และข้อมูลมีค่า

ผิดปกติปลอมปนในตัวแปร Y ร้อยละ 10 ถึง 30 สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งยังคงให้ค่า MSE ต่ำที่ สุด ในขณะที่ค่า ABS ของสัมประสิทธิ์สหสัมพันธ์โฮฟดิ้งไม่ได้มีค่าต่ำที่สุดดังปรากฏผลในตารางที่ 5

ตารางที่ 5 สัมประสิทธิ์สหสัมพันธ์ลำดับที่ซึ่งให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความเอนเอียงสัมบูรณ์ต่ำที่สุดในสถานการณ์ต่าง ๆ

ร้อยละของค่าผิดปกติจากกลุ่มในตัวแปร Y	n	ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 2 ตัว (ρ)											
		0		0.2		0.4		0.6		0.8		1	
		MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด	MSE ต่ำสุด	ABS ต่ำสุด
0	10	r_k	r_k	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	all	all
	20	r_h	r_k	r_k	r_b	r_k	r_b	r_s	r_b	r_s	r_b	all	all
	50	r_h	r_h	r_k	r_s	r_s	r_s	r_s	r_s	r_s	r_s	all	all
	100	r_h	r_h	r_s	r_s	r_s	r_s	r_s	r_s,r_b	r_s	r_s	all	all
	200	r_h	r_h	r_s	r_b	r_s	r_b	r_s	r_b	r_s	r_b	all	all
10	10	r_h	r_k	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	r_k	r_b
	20	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_k	r_b
	50	r_h	r_h	r_k	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_k	r_s
	100	r_h	r_h	r_s	r_s	r_s	r_s,r_b	r_s	r_b	r_s	r_b	r_k	r_b
	200	r_h	r_h	r_s	r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_k	r_s,r_b
20	10	r_h	r_s	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	r_k	r_b
	20	r_h	r_h	r_h	r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_k	r_b
	50	r_h	r_h	r_k	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_k	r_k
	100	r_h	r_h	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_k	r_s
	200	r_h	r_h	r_s	r_b	r_s	r_s	r_s	r_s	r_s	r_s	r_k	r_k
30	10	r_h	r_b	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	r_k	r_s
	20	r_h	r_h	r_h	r_b	r_k	r_b	r_s	r_b	r_s	r_b	r_k	r_s
	50	r_h	r_h	r_k	r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_k	r_b
	100	r_h	r_h	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_s	r_k	r_k
	200	r_h	r_h,r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_s	r_b	r_k	r_b

หมายเหตุ : r_s คือ สัมประสิทธิ์สหสัมพันธ์สเปียร์แมน; r_k คือ สัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทา; r_h คือ สัมประสิทธิ์สหสัมพันธ์โฮฟดิ้ง; r_b คือ สัมประสิทธิ์สหสัมพันธ์เบรสท์; all คือ สัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธี

สำหรับกรณีที่พารามิเตอร์ ρ มีค่าตั้งแต่ 0.2 ถึง 0.4 ส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์เบรสท์ให้ค่า ABS ต่ำที่สุดเมื่อ n = 10, 20, 200 ในทุกระดับของค่าผิดปกติที่เกิดขึ้นในตัวแปร Y แต่ถ้า n = 50, 100 ส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์สเปียร์แมนให้ค่า ABS ต่ำที่สุด ในขณะที่ค่า MSE ของสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนจะมีค่าต่ำที่สุดเมื่อ ($n \geq 100$ และ ρ = 0.2) หรือ ($n \geq 50$ และ ρ = 0.4) ในทุกระดับของค่าผิดปกติที่เกิดขึ้นในตัวแปร Y

สำหรับทุกขนาดตัวอย่างและทุกระดับของค่าผิดปกติที่เกิดขึ้นในตัวแปร Y กรณีที่พารามิเตอร์ ρ มีค่าตั้งแต่ 0.6 ถึง 0.8 สัมประสิทธิ์สหสัมพันธ์สเปียร์แมนให้ค่า MSE ต่ำที่สุดในขณะที่ค่า ABS ไม่ได้มีค่าต่ำที่สุดในทุกสถานการณ์ดังกล่าวนี้ดังปรากฏผลในตารางที่ 5

และเมื่อตัวแปรทั้งสองมีความสัมพันธ์กันในระดับสูง กล่าวคือ พารามิเตอร์ ρ มีค่าเท่ากับ 1.0 และมีค่าผิดปกติปลอมปนร้อยละ 10 ถึง 30 ในทุกขนาดตัวอย่าง พบว่าสัมประสิทธิ์สหสัมพันธ์เคนดอลล์เทาให้ค่า MSE ต่ำที่สุดในขณะที่ค่า ABS ไม่ได้มีค่าต่ำที่สุดโดยส่วนใหญ่สัมประสิทธิ์สหสัมพันธ์สเปียร์แมนหรือสัมประสิทธิ์สหสัมพันธ์เบรสท์จะให้ค่า ABS ต่ำที่สุดในสถานการณ์ดังกล่าวนี้ดังปรากฏผลในตารางที่ 5 ส่วนกรณีที่ไม่มีค่าผิดปกติเกิดขึ้นในตัวแปร Y และตัวแปรทั้งสองมีความสัมพันธ์กันในระดับสูง กล่าวคือ พารามิเตอร์ ρ มีค่าเท่ากับ 1.0 พบว่าสัมประสิทธิ์สหสัมพันธ์ทั้ง 4 วิธีมีแนวโน้มเป็นตัวประมาณที่มีความแกร่งเท่ากัน

6. ข้อเสนอแนะ

ควรมีการเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์สหสัมพันธ์วิธีอื่น ๆ อีก เช่น สัมประสิทธิ์สหสัมพันธ์โซเมอร์ [13] สัมประสิทธิ์สหสัมพันธ์แพลน-ตาจีเนท [14] ทั้งนี้เพื่อเป็นแนวทางในการเลือกใช้ตัวประมาณค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละกรณีตามความต้องการของผู้วิจัยและได้ผลลัพธ์ที่แม่นยำขึ้น

7. กิตติกรรมประกาศ

ขอขอบคุณ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ที่สนับสนุนค่าใช้จ่ายในการทำวิจัยครั้งนี้

8. เอกสารอ้างอิง

[1] Anscombe, F.J., 1960, Rejection of outliers, J. Am. Stat. Assoc. Am. Soc. Qual. 2: 123-147.

[2] Baragona, R., Battaglia, F. and Poli, I., 2011, Outliers, Available Source: http://link.springer.com/chapter/10.1007/978-3-642-16218-3_6?no-access=true, October 6, 2016.

[3] Croux, C. and Dehon, C., 2005, Robustness Versus Efficiency for Nonparametric Correlation Measures, Available Source: https://lirias.kuleuven.be/bitstream/123456789/164226/1/kbi_0803.pdf, October 6, 2016.

[4] Evandt, O., Coleman, S., Ramalhoto, M. F. and Lottum, C.V., 2004, A little-known robust estimator of the correlation coefficient and its use in a robust graphical test for bivariate normality with applications in the aluminium industry, Qual. Reliab. Eng. Int. 20: 433-456.

[5] นพดล พิมพ์จันทร์, 2550, ผลของข้อมูลที่ค่าผิดปกติจากกลุ่มต่อความแกร่งของสัมประสิทธิ์สหสัมพันธ์, วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัย

ขอนแก่น, ขอนแก่น.

[6] Devlin, S.J., Gnanadesikan, R. and Kettenring, J., 1975, Robust estimation and outlier detection with correlation coefficients, Biometrika 62: 531-545.

[7] Hoeffding, W., 1948, A Non-parametric test of independence, Ann. Math. Stat. 19: 546-557.

[8] Blest, D.C., 2000, Rank correlation – an alternative measure, Aust. N.Z. J. Stat. 42: 101-111.

[9] Barnett, V. and Lewis, T., 1995, Outlier in Statistical Data, 3th Ed., John Wiley & Sons., New York.

[10] Ferguson, T.S., 1961, On the rejection of outliers, pp. 253-287, In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.

[11] Spearman, C., 1904, The proof and measurement of association between two things, Am. J. Psychol. 15: 72-101.

[12] Kendall, M.G., 1938, A new measure of rank correlation, Biometrika 30: 81-93.

[13] Somers, R.H., 1962, A new asymmetric measure of association for ordinal variables, Am. Soc. Rev. 27: 799-811.

[14] Genest, C. and Plante, J.F., 2003, On Blest's measure of rank correlation, Can. J. Stat. 31: 35-52.