ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่

โดย
นางสาวปาริชาต ภัทรพานิชชัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
วิทยาศาสตร์ มหาวิทยาลัยศิลปากร
ปีการศึกษา 2559
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่

โดย
นางสาวปาริชาต ภัทรพานิชชัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
วิทยาศาสตร์ มหาวิทยาลัยศิลปากร
ปีการศึกษา 2559
ลิขสิทธิ์ของบัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

THE MODIFIED CONCORDANCE CORRELATION COEFFICIENT ESTIMATORS

By

MISS Parichart PATTARAPANITCHAI

A Thesis Submitted in partial Fulfillment of Requirements

for Master of Science (APPLIED STATISTICS)

Science Silpakorn University

Academic Year 2016

Copyright of Graduate School, Silpakorn University

หัวข้อ	ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่
โดย	ปาริชาต ภัทรพานิชชัย
สาขาวิชา	สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ผู้ช่วยศาสตราจารย์ ดร. ศศิประภา หิริโอตป์

วิทยาศาสตร์ มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.ปานใจ ธารทัศนวงศ์)

พิจารณาเห็นชอบโดย

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. กมลชนก พานิชการ)

..... อาจารย์ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร. ศศิประภา หิริโอตป์)

..... ผู้ทรงคุณวุฒิภายนอก

(รองศาสตราจารย์ ดร. กมล บุษบา)

58304202 : สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง, สัมประสิทธิ์สหสัมพันธ์เพียร์สัน, ตัวประมาณ

นางสาว ปาริชาต ภัทรพานิชชัย: ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร. ศศิประภา หิริโอตป์

งานวิจัยนี้มีวัตถุประสงค์เพื่อเสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยประยุกต์ใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันทางเลือกอื่นที่มีประสิทธิภาพดีขึ้นในเทอมของความเอนเอียง คือ 1. ตัวประมาณของ Olkin และ Pratt และ 2. ตัวประมาณของ Haddad และ Provost และเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่กับตัวประมาณดั้งเดิมในเทอมของความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ย ภายใต้สถานการณ์ที่กำหนดคือข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานที่มีค่าเฉลี่ยและความแปรปรวนระหว่างสองตัวแปรไม่แตกต่างกัน แตกต่างกันเล็กน้อย และแตกต่างกันปานกลาง สัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่า -0.9, -0.8,..., 0.8, 0.9 ด้วยขนาดตัวอย่าง 5, 10, 30 และ 50 แต่ละพารามิเตอร์และขนาดตัวอย่างทำซ้ำ 10,000 ครั้ง

ผลการวิจัยพบว่า

1. ตัวประมาณที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

เมื่อเปรียบเทียบตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมและตัวประมาณที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt พบว่าตัวประมาณที่ปรับใหม่มีประสิทธิภาพดีกว่าในเทอมของความเอนเอียงที่ทุกระดับของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน และเมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีขนาดใหญ่ ตัวประมาณที่ปรับใหม่จะมีประสิทธิภาพดีกว่าในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย ในทุกระดับของความแตกต่างของค่าเฉลี่ยและความแปรปรวนระหว่างสองตัวแปร

2. ตัวประมาณที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost

เมื่อเปรียบเทียบตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมและตัวประมาณที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost พบว่าตัวประมาณที่ปรับใหม่มีประสิทธิภาพดีกว่าในเทอมของความเอนเอียงเมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีขนาดปานกลางถึงสูง ในทุกระดับของความแตกต่างของค่าเฉลี่ยและความแปรปรวนระหว่างสองตัวแปร และในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย ตัวประมาณที่ปรับใหม่มีประสิทธิภาพดีกว่าตัวประมาณเดิมทุกระดับของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

58304202 : Major (APPLIED STATISTICS)
Keyword : The concordance correlation coefficient, The Pearson correlation coefficient, Estimator

MISS Parichart PATTARAPANITCHAI : The Modified Concordance Correlation Coefficient Estimators Thesis advisor : Assistant Professor Dr. Sasiprapa Hiriote

This research proposed the modified concordance correlation coefficient estimators by using the alternative Pearson correlation coefficient estimators that are improved in terms of bias; (1) Olkin and Pratt estimator, (2) Haddad and Provost estimator. The performances of the modified concordance correlation coefficient estimators and the original one were compared in terms of the absolute bias and mean square error under given circumstances, the data were randomly generated from the standard normal distribution with means and variances between the two variables are not different, slightly different and moderately different, the Pearson correlation coefficient are -0.9, -0.8, ..., 0.8, 0.9 with the sample sizes of 5, 10, 30 and 50. Each parameters and sample sizes was reproduced 10,000 replications.

The results of the study indicate that

1. The modified concordance correlation coefficient estimator by using Olkin and Pratt estimator

Compared with the original concordance correlation coefficient estimator, the modified estimator has better performance in terms of bias at all levels of correlation and all levels of means and variances difference between the two variables. For mean square error, the proposed estimator performs better than the original one when the correlation is high.

2. The modified concordance correlation coefficient estimator by using Haddad and Provost estimator

Compared with the original concordance correlation coefficient estimator, the modified estimator performs better than the original estimator in terms of bias when the correlation is moderate to high in all levels of differences of means and variances between the two variables. In terms of mean square error, the proposed estimator performs better than the original one at all levels of correlation.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้เป็นอย่างดีเนื่องจากได้รับความอนุเคราะห์จากผู้ช่วยศาสตราจารย์ ดร.ศศิประภา หิริโอตป์ ซึ่งเป็นอาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ คำปรึกษา แนวคิดวิธีการแก้ปัญหา รวมทั้งช่วยตรวจทานแก้ไขข้อบกพร่องต่างๆเป็นอย่างดียิ่งตลอดเวลาในการทำวิทยานิพนธ์ ตั้งแต่เริ่มต้น จนสำเร็จเรียบร้อย นอกเหนือจากงานวิจัย ท่านยังเป็นอาจารย์ที่ปรึกษาผู้ให้คำแนะนำในการดำรงชีวิตเป็นอย่างดีเยี่ยม เป็นผู้ที่เป็นแรงบัลดาลใจในการทำงานของข้าพเจ้าตลอดเวลา ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างยิ่งในความกรุณาของท่านไว้ ณ ที่นี้ด้วย

นอกจากนี้ผู้วิจัยขอกราบขอบพระคุณผู้ช่วยศาสราจารย์ ดร.กมลชนก พานิชการ ประธานกรรมการตรวจสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.กมล บุษบา กรรมการผู้ทรงคุณวุฒิ สำหรับคำแนะนำ ตรวจสอบ และคำชี้แนะ ทำให้วิทยานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ภาควิชาสถิติ มหาวิทยาลัยศิลปากรทุกท่านที่ช่วยประสิทธิประสาทองค์ความรู้ ความช่วยเหลือ และคำแนะนำต่างๆทั้งในด้านการเรียนและเป็นแนวทางในการดำเนินชีวิต ตลอดจนกำลังใจ แนวคิดในการศึกษา และเป็นแรงบัลดาลใจในการเรียนและการทำงานของข้าพเจ้า

ขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร.รัชฎา บุญเต็ม ผู้เป็นเจ้าของข้อมูลและอนุญาตให้ผู้วิจัยนำมาใช้เป็นตัวอย่าง และให้คำแนะนำเกี่ยวกับข้อมูล นอกจากนี้ยังให้กำลังใจและคำแนะนำทั้งในด้านการทำงานและการใช้ชีวิตร่วมกับผู้อื่น

กราบขอบพระคุณมารดา พี่สาวและพี่ชาย ที่สนับสนุนการศึกษา และการทำวิจัยครั้งนี้ ทั้งกำลังทรัพย์และกำลังใจให้กับผู้วิจัยมาตลอดการศึกษา จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์

ขอบพระคุณโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่ให้การสนับสนุนในด้านทุนทรัพย์ในการทำวิจัย และการไปนำเสนอผลงานภายในประเทศ

ขอบคุณพี่ๆ เพื่อนๆ และน้องๆ สาขาวิชาสถิติและสาขาวิชาอื่นๆที่ให้คำแนะนำ ให้ความช่วยเหลือ และเป็นแรงผลักดันตลอดมา ขอบคุณพี่นงลักษณ์ เอี๊ยวเจริญ ที่ให้ความช่วยเหลือในการอำนวยความสะดวกด้านเอกสารตลอดการศึกษาของผู้วิจัย

ปาริชาต ภัทรพานิชชัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญภาพ ..... ฎ

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ของการศึกษา 2

ขอบเขตการศึกษา 3

ขั้นตอนของการวิจัย 4

ประโยชน์ที่คาดว่าจะได้รับ 4

นิยามศัพท์เฉพาะ 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

ทฤษฎีที่เกี่ยวข้อง 6

1. สหสัมพันธ์ 6

2. สัมประสิทธิ์สหสัมพันธ์เพียร์สัน 7

3. ตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน 9

3.1 ตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน 9

3.2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt 9

3.3 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost 10

4. ความสอดคล้อง 12
5. สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง 14
งานวิจัยที่เกี่ยวข้อง 16
บทที่ 3 วิธีดำเนินการวิจัย 18
1. สร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ 18
2. การศึกษาความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยโดยการจำลองข้อมูล 19
บทที่ 4 ผลการวิเคราะห์ข้อมูล 23
การประมาณความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ยด้วยการจำลองข้อมูล 23
1. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร(กรณีที่ 1) 24
2. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2) 27
3. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3) 30
4. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4) 33
5. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5) 36
6. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6) 39
7. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7) 42
บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ 45
สรุปผลการวิจัย 45
1. ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt 45

2. ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost 46
อภิปรายผล 49
ข้อเสนอะแนะ 50
ตัวอย่างการนำตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องไปใช้ในการวัดความสอดคล้องข้อมูล 51
รายการอ้างอิง 57
ภาคผนวก 58
โปรแกรมที่ใช้ในงานวิจัย 59
ข้อมูลตัวอย่างที่ใช้ในงานวิจัย 66
เอกสารการเผยแพร่วิทยานิพนธ์ 70
ประวัติผู้เขียน 76

มหาวิทยาลัยศิลปากร

สารบัญตาราง

หน้า

ตารางที่ 1 ขนาดของค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง ρ_c ในแต่ละกรณีที่ศึกษา..... 20

ตารางที่ 2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละระดับของ ค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ ρ เมื่อขนาดตัวอย่างเท่ากับ 5..... 47

ตารางที่ 3 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละระดับของ ค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ ρ เมื่อขนาดตัวอย่างเท่ากับ 10..... 48

ตารางที่ 4 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์สหสัมพันธ์เพียร์สันของร้อยละของธาตุหลักในลูกปัดแก้ว ที่วัดด้วยวิธี LA-ICP-MS และวิธี EPMA ทั้งหมด 52 ลูก..... 52

ตารางที่ 5 ค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องระหว่างวิธี LA-ICP-MS และวิธี EPMA จากตัวประมาณทั้ง 3 ชนิด..... 56

ตารางที่ 6 ร้อยละของธาตุหลักในลูกปัดแก้วโบราณที่วัดโดยวิธี LA-ICP-MS จำนวน 52 ลูก..... 67

ตารางที่ 7 ร้อยละของธาตุหลักในลูกปัดแก้วโบราณที่วัดโดยวิธี EPMA จำนวน 52 ลูก..... 69

สารบัญภาพ

หน้า

ภาพที่ 1 สหสัมพันธ์ทางบวก สหสัมพันธ์ทางลบ และไม่มีสหสัมพันธ์ 6

ภาพที่ 2 ความสัมพันธ์เชิงเส้นตรง และความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง 7

ภาพที่ 3 ตัวอย่างแผนภาพการกระจายเมื่อ ρ มีค่าต่างๆ 8

ภาพที่ 4 ความสอดคล้องสมบูรณ์ในทิศทางเดียวกัน ความสอดคล้องสมบูรณ์ในทิศทางตรงกันข้าม และความไม่สอดคล้อง 12

ภาพที่ 5 ตัวอย่างกรณีที่สัมประสิทธิ์สหสัมพันธ์เพียร์สันไม่สามารถตรวจจับความไม่สอดคล้องของข้อมูล 13

ภาพที่ 6 ตัวอย่างกรณีที่สถิติทดสอบ t และการวิเคราะห์การถดถอยไม่สามารถตรวจจับความไม่สอดคล้องของข้อมูล 13

ภาพที่ 7 ตัวอย่างแผนภาพการกระจายเมื่อ ρ_c มีค่าต่างๆ 15

ภาพที่ 8 แผนภาพแสดงลำดับขั้นตอนการศึกษาความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องโดยการจำลองข้อมูล 22

ภาพที่ 9 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร(กรณีที่ 1) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 24

ภาพที่ 10 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลไม่มีความแตกต่างของค่าเฉลี่ยและความแปรปรวน (กรณีที่ 1) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 25

ภาพที่ 11 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ยแต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 27

ภาพที่ 12 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 29

ภาพที่ 13 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 30

ภาพที่ 14 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 32

ภาพที่ 15 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 33

ภาพที่ 16 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 35

ภาพที่ 17 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 36

ภาพที่ 18 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 38

ภาพที่ 19 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$. 39

ภาพที่ 20 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 41

ภาพที่ 21 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 42

ภาพที่ 22 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$ 44

ภาพที่ 23 แผนภาพการกระจายของร้อยละของธาตุหลักในลูกปัดแก้ว ระหว่างการวัดด้วยวิธี LA-ICP-MS และวิธี EPMA ทั้งหมด 52 ลูก เมื่อธาตุหลักคือ (a) อลูมิเนียม, (b) ซิลิกอน, (c) ทองแดง, (d) โซเดียม, (e) แคลเซียม 52

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ทฤษฎี เทคโนโลยี และแนวคิดใหม่ๆยังคงพัฒนาอย่างต่อเนื่อง ซึ่งทำให้เกิดวิธีการ อุปกรณ์และเครื่องมือสำหรับการเข้าถึงข้อมูลที่จำเป็นในการใช้ในหลายสาขาวิชา เช่นในทางสังคม พฤติกรรม กายภาพ ชีวภาพ และทางวิทยาศาสตร์การแพทย์ การเข้าถึงข้อมูลที่น่าเชื่อถือและแม่นยำถูกต้องเป็นจุดเริ่มต้นในการวิเคราะห์ที่ดี ดังนั้นการมีสถิติที่เหมาะสมในการวิเคราะห์ว่าข้อมูลหรือเครื่องมือที่ใช้วัดน่าเชื่อถือหรือสอดคล้องกับเครื่องมือที่มีอยู่แล้วจึงจำเป็นอย่างยิ่ง ซึ่งสถิติที่ถูกนำมาใช้ในการวัดประสิทธิภาพของเครื่องมือเมื่อต้องการเปรียบเทียบเครื่องมือชนิดใหม่กับเครื่องมือที่มีอยู่แล้ว กรณีวัดความสอดคล้องของข้อมูลในมาตรวัดแบบต่อเนื่องสามารถวิเคราะห์ได้หลายวิธี เช่น สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง (Concordance correlation coefficient: CCC), สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson correlation coefficient), สถิติทดสอบ paired-t, การวิเคราะห์การถดถอยเมื่อความชันเท่ากับ 1 และระยะตัดแกนเท่ากับ 0, สัมประสิทธิ์ความผันแปร (Coefficient of variation), และสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass correlation coefficient)[1]

สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง เป็นสถิติที่ใช้วัดความสอดคล้องระหว่างข้อมูลเชิงปริมาณสองตัวแปรที่ถูกใช้มากในวรรณกรรมทางสถิติ สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องถูกคิดค้นโดย Lin(1989)[2] และเนื่องจากสามารถวัดทั้งขนาดและทิศทางของความสอดคล้องจึงทำให้ง่ายต่อการแปลผล นอกจากนี้ในการอนุมานยังไม่จำเป็นต้องทราบการแจกแจงของตัวสถิติในกรณีที่ตัวอย่างมีขนาดใหญ่ จึงสะดวกในการนำไปใช้ สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องสามารถเขียนอยู่ในรูปของค่าเฉลี่ยและความแปรปรวนของทั้งสองตัวแปร และสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

นอกจากนี้สถิติที่ถูกใช้เพื่อวัดความสอดคล้องของเครื่องมือมีอยู่หลากหลายเช่น สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน, สถิติทดสอบ t, การวิเคราะห์การถดถอย, สัมประสิทธิ์ความผันแปร, และสัมประสิทธิ์สหสัมพันธ์ภายในชั้น เมื่อพิจารณาสถิติที่กล่าวมานี้จะพบว่าการใช้เฉพาะสถิติใดสถิติหนึ่งอาจไม่สามารถประเมินความสอดคล้องได้ถูกต้องเพราะการใช้สถิติที่กล่าวมาเพียงชนิดเดียวอาจจะทำให้การแปลผลผิดพลาดไปได้ เช่นกรณีที่ข้อมูลมีสหสัมพันธ์กันสูงและมีค่าเฉลี่ยแตกต่างกันมาก เมื่อพิจารณาสัมประสิทธิ์สหสัมพันธ์เพียร์สันจะมีค่าสูง เพราะเป็นสถิติที่ไม่ขึ้นอยู่กับความแตกต่างของ

ค่าเฉลี่ย ส่งผลให้สรุปว่าข้อมูลมีความสอดคล้องสูง แต่ความแตกต่างของค่าเฉลี่ยที่มากทำให้ทราบว่าข้อมูลไม่สอดคล้องกัน

ในหลายงานวิจัยสนใจศึกษาความสัมพันธ์ระหว่างตัวแปรสองตัว ซึ่งสัมประสิทธิ์สหสัมพันธ์เพียร์สันเป็นสถิติที่นิยมใช้ในการวัดขนาดและทิศทางของความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเชิงปริมาณสองตัวแปร นอกจากนี้ยังถูกใช้เพื่อคำนวณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องด้วย ดังนั้นประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันจะส่งผลต่อประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง

Fisher (1915) ได้แสดงว่าค่าคาดหวังของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของตัวอย่างสุ่มที่มาจากการแจกแจงปกติเป็นตัวประมาณที่เอนเอียง โดยมีแนวโน้มที่จะประมาณได้ต่ำกว่าค่าพารามิเตอร์เมื่อข้อมูลมีสหสัมพันธ์ทางบวก และมีแนวโน้มที่จะประมาณได้สูงกว่าพารามิเตอร์เมื่อข้อมูลมีสหสัมพันธ์ทางลบ[3] ดังนั้นตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทางเลือกจึงถูกคิดค้นขึ้นโดยตัวประมาณสัมประสิทธิ์สหสัมพันธ์ของ Olkin and Pratt (1958) ซึ่งมีความเอนเอียงน้อยลงกว่าตัวประมาณดั้งเดิมบนเงื่อนไขการแจกแจงปกติของตัวอย่างสุ่ม[4] นอกจากนี้ยังมีตัวประมาณที่อยู่บนพื้นฐานของพิสัยนั่นคือตัวประมาณ Range-based คิดค้นโดย Haddad และ Provost (2011)[5] ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงเมื่อตัวอย่างสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์ ρ

งานวิจัยนี้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทางเลือกอื่นนั่นคือ ตัวประมาณของ Olkin และ Pratt และตัวประมาณของ Haddad และ Provost (ตัวประมาณ Range-based) และพิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ย โดยการจำลองข้อมูล (Monte Carlo simulation)

วัตถุประสงค์ของการศึกษา

1. เพื่อเสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่กับตัวประมาณดั้งเดิมในเทอมของความเอนเอียงสัมบูรณ์

3. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่กับตัวประมาณดั้งเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย

ขอบเขตการศึกษา

ในงานวิจัยนี้จะเสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ และเปรียบเทียบประสิทธิภาพของตัวประมาณดั้งเดิมและตัวประมาณที่ปรับใหม่ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ย โดยกำหนดของเขตของการศึกษาครั้งนี้ดังนี้

1. ศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณเดิมและตัวประมาณที่ปรับใหม่ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ย โดยการจำลองข้อมูลด้วยโปรแกรม R

2. การแจกแจงของข้อมูลที่ศึกษาคือ (X,Y) มีการแจกแจงปกติสองตัวแปรภายใต้พารามิเตอร์ดังนี้โดยพารามิเตอร์ที่กำหนดคือ

$\mu_X = 0$

$\mu_Y = 0, 0.5, 1.25$

$\sigma_X^2 = 1$

$\sigma_Y^2 = 0.5, 1, 9$

$\rho = -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9$

ที่ขนาดตัวอย่าง $n = 5, 10, 30, 50$

โดยกรณีที่ศึกษาคือ

กรณีที่ 1: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 2: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 0.5$$

กรณีที่ 3: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 9$$

กรณีที่ 4: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน:

$$\mu_X = 0, \mu_Y = 0.5, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 5: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน:

$$\mu_X = 0, \mu_Y = 1.25, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 6: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย:

$$\mu_X = 0, \mu_Y = 0.5, \sigma_X^2 = 1, \sigma_Y^2 = 0.5$$

กรณีที่ 7: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง:

$$\mu_X = 0, \mu_Y = 1.25, \sigma_X^2 = 1, \sigma_Y^2 = 9$$

ที่ทุกระดับของสัมประสิทธิ์สหสัมพันธ์

3. ขนาดตัวอย่างที่ใช้ในการศึกษาคือ $n = 5, 10, 30, 50$

ขั้นตอนของการวิจัย

1. ศึกษาทฤษฎีและวรรณกรรมที่เกี่ยวข้อง
2. สร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่
3. ประมาณค่าความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องโดยการจำลองข้อมูล
4. เปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิสหสัมพันธ์ความสอดคล้องที่ปรับใหม่ และตัวประมาณเดิม
4. สรุปและอภิปรายผล

ประโยชน์ที่คาดว่าจะได้รับ

1. ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่สามารถนำไปใช้วัดความสอดคล้องของข้อมูล
2. สามารถเลือกใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่เหมาะสมกับข้อมูลมากที่สุดในการวัดความสอดคล้องของข้อมูล

นิยามศัพท์เฉพาะ

1. ความเอนเอียง Bias คือความแตกต่างระหว่างค่าคาดหมายของตัวประมาณกับค่าพารามิเตอร์
2. ความคลาดเคลื่อนกำลังสองเฉลี่ย Mean square error (MSE) คือค่าคาดหมายของกำลังสองของผลต่างระหว่างตัวประมาณ $\hat{\theta}$ กับพารามิเตอร์ θ หาได้จาก $MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2$

3. ตัวประมาณไม่เอนเอียง Unbiased estimator คือตัวประมาณที่มีค่าคาดหมายเท่ากับพารามิเตอร์ กล่าวคือ $\hat{\theta}$ เป็นตัวประมาณไม่เอนเอียงของ θ ก็ต่อเมื่อ $E\left(\hat{\theta}\right)=\theta$ โดยที่ E. เป็นสัญลักษณ์แทนค่าคาดหมาย[6]

4. ความสอดคล้อง Agreement คือความเห็นพ้องต้องกัน ความเหมือนกัน หรือความใกล้เคียงกัน ความสอดคล้องถูกนำมาใช้สำหรับการประเมินการยอมรับในกระบวนการ ระเบียบวิธีและสูตรต่าง ๆ กล่าวคือเมื่อกระบวนการสองชนิดมีความสอดคล้องกันสามารถบอกได้ว่ากระบวนการนั้นใช้แทนกันได้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

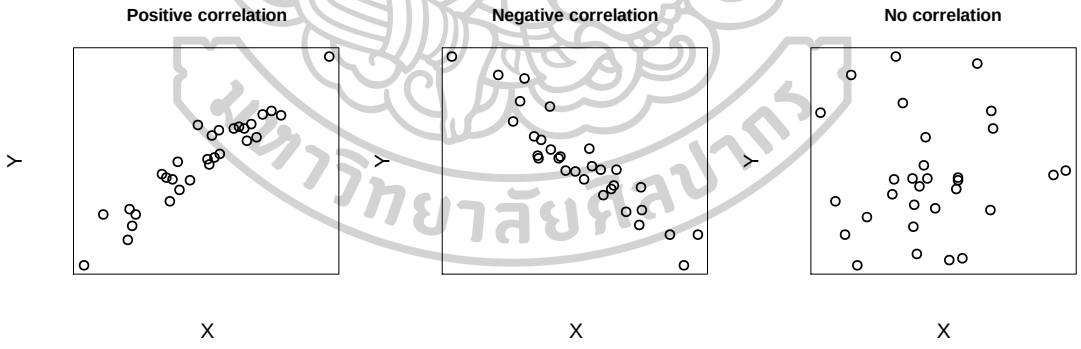
1. สหสัมพันธ์

โดยส่วนมากแต่ละหน่วยสังเกตจะถูกเก็บข้อมูลหลายตัวแปรเพื่อตอบวัตถุประสงค์ในงานวิจัยนั้นๆ ซึ่งในงานวิจัยส่วนใหญ่ ผู้วิจัยสนใจที่จะศึกษาความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณสองตัวแปร นั่นคือ สหสัมพันธ์

ชนิดของสหสัมพันธ์สามารถแบ่งได้หลายรูปแบบดังนี้

การแบ่งชนิดของสหสัมพันธ์แบบแรกคือการพิจารณาแนวโน้มของตัวแปรที่สองเมื่อตัวแปรที่หนึ่งมีค่าเพิ่มขึ้นโดยแบ่งชนิดของสหสัมพันธ์ได้ดังนี้

- สหสัมพันธ์ทางบวก คือ เมื่อตัวแปรหนึ่งมีค่าเพิ่มขึ้น ตัวแปรที่สองมีแนวโน้มที่จะเพิ่มขึ้น
- สหสัมพันธ์ทางลบ คือ เมื่อตัวแปรหนึ่งมีค่าเพิ่มขึ้น ตัวแปรที่สองมีแนวโน้มที่จะลดลง
- ไม่มีสหสัมพันธ์ คือ เมื่อตัวแปรหนึ่งมีค่าเพิ่มขึ้น ตัวแปรที่สองไม่มีแนวโน้มที่จะเพิ่มขึ้นหรือลดลง



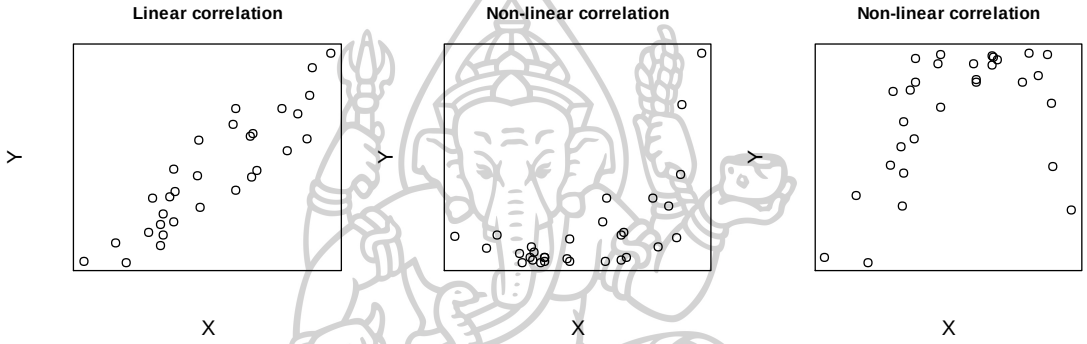
ภาพที่ 1 สหสัมพันธ์ทางบวก สหสัมพันธ์ทางลบ และไม่มีสหสัมพันธ์

การแบ่งชนิดของสหสัมพันธ์แบบที่สองคือจำนวนตัวแปรที่นำมาวิเคราะห์สหสัมพันธ์ โดยแบ่งได้ดังนี้

- Simple correlation คือ สหสัมพันธ์ระหว่างตัวแปรสองตัว
- Multiple correlation คือ สหสัมพันธ์ระหว่างตัวแปรสามตัวขึ้นไป
- Partial correlation คือ สหสัมพันธ์ระหว่างตัวแปรสองตัวเมื่อควบคุมตัวแปรอื่นให้คงที่

การแบ่งชนิดของสหสัมพันธ์แบบที่สามคือลักษณะของความสัมพันธ์ โดยแบ่งชนิดของสหสัมพันธ์ได้ดังนี้

- ความสัมพันธ์เชิงเส้นตรง (Linear correlation) คือ ปริมาณของการเปลี่ยนแปลงของตัวแปรหนึ่งมีแนวโน้มที่จะทำให้อัตราส่วนของปริมาณการเปลี่ยนแปลงของตัวแปรอื่นคงที่ กราฟของตัวแปรที่มีความสัมพันธ์เชิงเส้นจะมีรูปแบบเป็นเส้นตรง
- ความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง (Non-linear correlation) คือ ปริมาณการเปลี่ยนแปลงของตัวแปรหนึ่งไม่มีแนวโน้มที่จะทำให้อัตราส่วนของปริมาณการเปลี่ยนแปลงของตัวแปรอื่นคงที่



ภาพที่ 2 ความสัมพันธ์เชิงเส้นตรง และความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรง

2. สัมประสิทธิ์สหสัมพันธ์เพียร์สัน

สัมประสิทธิ์สหสัมพันธ์เพียร์สันหรือ the Pearson product-moment correlation coefficient ถูกคิดค้นโดย Pearson (1896, 1900) ซึ่งใช้วัดระดับความสัมพันธ์เชิงเส้นระหว่างข้อมูลสองตัวแปรในมาตราอันตรภาคหรือมาตราอัตราส่วน และเป็นสัมประสิทธิ์ที่เป็นที่นิยมใช้ในการวิเคราะห์ความสัมพันธ์ของข้อมูล[7, 8]

ให้ตัวแปรสุ่ม (X,Y) มีการแจกแจงปกติสองตัวแปร สัมประสิทธิ์สหสัมพันธ์เพียร์สันของตัวแปรสุ่ม X และ Y สามารถเขียนได้ดังนี้

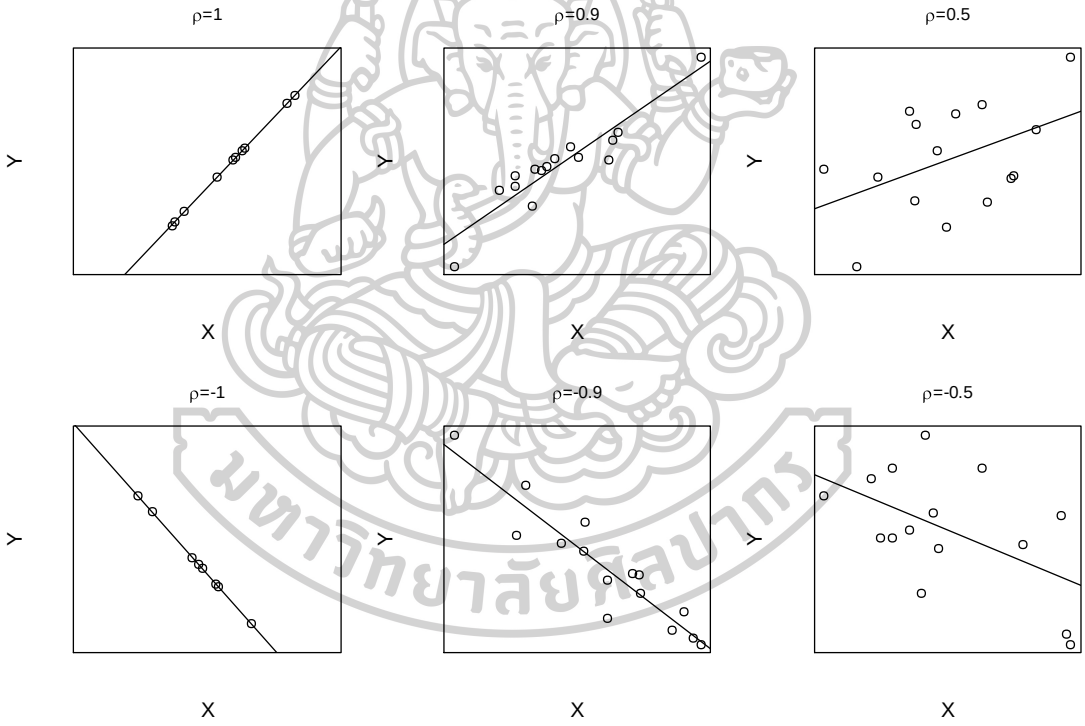
$$\rho = \frac{E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]}{\sqrt{E\left[\left(X-\mu_X\right)^2\right]E\left[\left(Y-\mu_Y\right)^2\right]}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \qquad (1)$$

เมื่อ

$$\mu_X = E\left(X\right)$$
$$\mu_Y = E\left(Y\right)$$
$$\sigma_{XY} = Cov\left(X,Y\right)$$

$$\sigma_X = \sqrt{V(X)}$$
$$\sigma_Y = \sqrt{V(Y)}$$

สัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่าอยู่ระหว่าง -1 ถึง 1 $(-1 \le \rho \le 1)$ ค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์เพียร์สันแสดงถึงระดับความสัมพันธ์ เมื่อค่าสัมบูรณ์ของ ρ มีค่าเข้าใกล้ 1 แสดงว่าระดับความสัมพันธ์เชิงเส้นของ X และ Y แข็งแกร่งมากขึ้น เครื่องหมายของสัมประสิทธิ์สหสัมพันธ์เพียร์สันแสดงถึงทิศทางของความสัมพันธ์เชิงเส้นระหว่าง X และ Y เมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่าเท่ากับ -1 หมายถึง X และ Y มีความสัมพันธ์เชิงเส้นทางลบโดยสมบูรณ์(มีความสัมพันธ์ในทิศทางตรงกันข้าม), 1 หมายถึง X และ Y มีความสัมพันธ์เชิงเส้นทางบวกโดยสมบูรณ์(มีความสัมพันธ์ในทิศทางเดียวกัน), และเมื่อมีค่าเท่ากับ 0 หมายถึง X และ Y ไม่มีความสัมพันธ์เชิงเส้นกัน[9]



ภาพที่ 3 ตัวอย่างแผนภาพการกระจายเมื่อ ρ มีค่าต่างๆ

ข้อควรระวัง

1. สัมประสิทธิ์สหสัมพันธ์เป็นเพียงค่าวัดความสัมพันธ์เชิงเส้นระหว่างตัวแปรว่ามีมากน้อยเพียงใดเท่านั้น แต่ไม่จำเป็นที่ตัวแปรหนึ่งต้องเป็นเหตุ และอีกตัวแปรเป็นผล คือถ้าสหสัมพันธ์ระหว่าง X และ Y มีค่าสูง ไม่ได้หมายความว่าการเปลี่ยนแปลงของตัวแปร X เป็นสาเหตุให้เกิดการเปลี่ยนแปลงของตัวแปร Y

2. การศึกษาความสัมพันธ์ระหว่างตัวแปรด้วยสัมประสิทธิ์สหสัมพันธ์ ควรทำคู่กับการใช้แผนภาพการกระจาย เพราะค่าสัมประสิทธิ์ที่เท่ากัน อาจจะมีลักษณะความสัมพันธ์ที่ต่างกันอย่างมากได้[10]

3. ตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

3.1 ตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

ให้ตัวแปรสุ่ม $(X_i, Y_i), i = 1, ..., n$ มีการแจกแจงปกติสองตัวแปร สัมประสิทธิ์สหสัมพันธ์เพียร์สันของตัวแปรสุ่ม X และ Y สามารถประมาณได้ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันดังนี้

$$R = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X}\right)\left(\frac{Y_i - \bar{Y}}{S_Y}\right) \qquad (2)$$

เมื่อ

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_X^2 = \sum_{i=1}^{n}\frac{\left(X_i - \bar{X}\right)^2}{n-1}$$

$$S_Y^2 = \sum_{i=1}^{n}\frac{\left(Y_i - \bar{Y}\right)^2}{n-1}$$

อย่างไรก็ตาม ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันนี้ถูกพิจารณาว่าเป็นตัวประมาณที่เอนเอียงโดยมีแนวโน้มที่จะประมาณค่าได้ต่ำกว่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของประชากร[11, 12]

3.2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

Olkin และ Pratt (1958) ได้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ไม่เอนเอียงเมื่อตัวอย่างมาจากการแจกแจงแบบปกติ ซึ่งเขียนอยู่ในรูปฟังก์ชันของ R และมีรูปแบบที่ซับซ้อน Olkin และ Pratt จึงปรับตัวประมาณให้อยู่ในรูปที่ง่ายขึ้น แต่ยังมีความเอนเอียงลดลงจากตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันเดิม [4, 12, 13] ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่เอนเอียงน้อยลง คือ

$$\hat{\rho}_{OP} = R\left[1+\frac{1-R^2}{2(n-3)}\right] \qquad (3)$$

3.3 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost

Haddad และ Provost (2011) ได้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันในเทอมของพิสัย $|X_i - Y_i|$ เมื่อ (X_i, Y_i) มีการแจกแจงแบบปกติสองตัวแปร และสามารถจำแนกตัวประมาณได้ดังนี้

กรณีตัวอย่างมาจากการแจกแจงแบบปกติมาตรฐานสองตัวแปร

1. เมื่อสหสัมพันธ์มีค่าเป็นบวก ρ

$$\hat{\rho}_{HP} = 1-\frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{2n}$$

2. เมื่อสหสัมพันธ์มีค่าเป็นลบ กำหนดให้ X และ Y^- มีความสัมพันธ์เชิงลบด้วยสัมประสิทธิสหสัมพันธ์ ρ^- และให้ $Y = -Y^-$ และ $\rho = -\rho^-$

$$\begin{aligned}\hat{\rho}^-_{HP} &= -\hat{\rho}_{HP} \\ &= \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{2n} - 1 \\ &= \frac{\sum_{i=1}^{n}(X_i + Y_i^-)^2}{2n} - 1\end{aligned}$$

กรณีตัวอย่างมาจากการแจกแจงปกติสองตัวแปรเมื่อไม่ทราบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ดังนั้นแปลงข้อมูลเป็นมาตรฐานโดยการให้ $X_i^* = (X_i - \bar{X})/S_X$ และ $Y_i^* = (Y_i - \bar{Y})/S_Y$ ทำให้สามารถเขียนตัวประมาณสัมประสิทธิสหสัมพันธ์ได้ดังนี้

1. เมื่อสหสัมพันธ์มีค่าเป็นบวก ρ

$$\hat{\rho}_{HPs} = 1 - \frac{\sum_{i=1}^{n}\left(X_i^* - Y_i^*\right)^2}{2n}$$

$$= 1 - \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X} - \frac{Y_i - \bar{Y}}{S_Y}\right)^2$$

$$= 1 - \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{S_Y\left(X_i - \bar{X}\right) - S_X\left(Y_i - \bar{Y}\right)}{S_X S_Y}\right)^2$$

$$= 1 - \frac{1}{2n}\left[\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{S_X^2} + \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}{S_Y^2} - \frac{2\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{S_X S_Y}\right]$$

$$= 1 - \frac{1}{2n}\left[\frac{(n-1)S_X^2}{S_X^2} + \frac{(n-1)S_Y^2}{S_Y^2} - 2(n-1)R\right]$$

$$= 1 - \frac{n-1}{n}(1-R)$$

$$= R + \frac{1-R}{n}$$

2. เมื่อสหสัมพันธ์มีค่าเป็นลบ กำหนดให้ X และ Y^- มีความสัมพันธ์เชิงลบด้วยสัมประสิทธิสหสัมพันธ์ ρ^- และให้ $Y = -Y^-$ และ $\rho = -\rho^-$ แปลงข้อมูลเป็นมาตรฐานโดยการให้ $X_i^* = \left(X_i - \bar{X}\right)/S_X$ และ $Y_i^* = \left(Y_i - \bar{Y}\right)/S_Y$

$$\hat{\rho}_{HPs}^- = -\hat{\rho}_{HPs} = \frac{\sum_{i=1}^{n}\left(X_i^* - Y_i^*\right)^2}{2n} - 1$$

$$= \frac{1}{2n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X} - \frac{Y_i - \bar{Y}}{S_Y}\right)^2 - 1$$

$$= \frac{1}{2n}\left[\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{S_X^2} + \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}{S_Y^2} - \frac{2\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{S_X S_Y}\right] - 1$$

$$= \frac{1}{2n}\left[\frac{(n-1)S_X^2}{S_X^2} + \frac{(n-1)S_Y^2}{S_Y^2} - 2(n-1)R\right] - 1$$

$$= \frac{n-1}{n}(1-R) - 1$$

$$= -R - \frac{1-R}{n}$$

$$= R^- - \frac{1+R^-}{n}$$

ดังนั้นตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost คือ

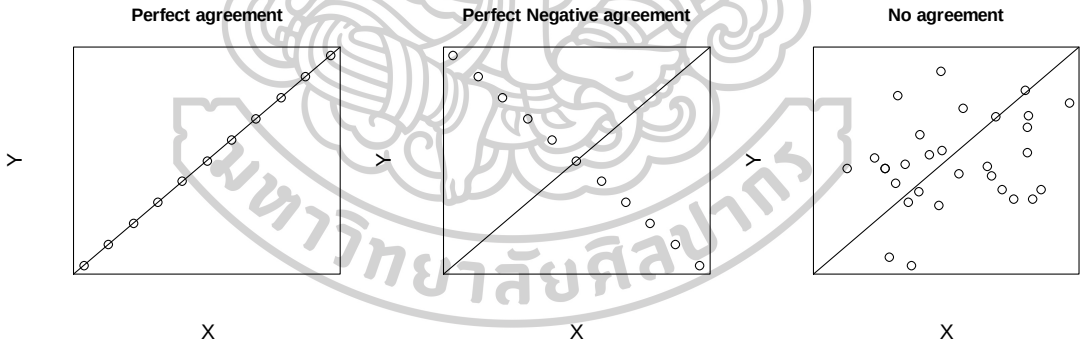
$$\hat{\rho}_{HP}=\begin{cases} 1-\dfrac{\sum_{i=1}^{n}\left(X_i-Y_i\right)^2}{2n} & ,\left(X_i,Y_i\right)\sim N_2\left(0,0,1,1,\rho\geq 0\right) \\ \dfrac{\sum_{i=1}^{n}\left(X_i+Y_i\right)^2}{2n}-1 & ,\left(X_i,Y_i\right)\sim N_2\left(0,0,1,1,\rho< 0\right) \\ R+\dfrac{1-R}{n} & ,\left(X_i,Y_i\right)\sim N_2\left(\left(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2\right)\neq\left(0,0,1,1\right),\rho\geq 0\right) \\ R-\dfrac{1+R}{n} & ,\left(X_i,Y_i\right)\sim N_2\left(\left(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2\right)\neq\left(0,0,1,1\right),\rho< 0\right) \end{cases} \quad (4)$$

Haddad และ Provost ได้แสดงว่าตัวประมาณ Range-based นี้เป็นตัวประมาณที่ไม่เอนเอียง เมื่อตัวอย่างมาจากการแจกแจงแบบปกติมาตรฐานสองตัวแปร[5] $\left(X_i,Y_i\right)\sim N_2\left(0,0,1,1,\rho\right)$

$$E\left(\hat{\rho}_{HP}\right)=\rho$$

4. ความสอดคล้อง

การวัดความสอดคล้องกันของข้อมูล (Agreement) สามารถนำไปใช้ประโยชน์ได้อย่างมาก เพราะการวัดความสอดคล้องกันหรือความใกล้เคียงกันของข้อมูลมีความสำคัญอย่างยิ่งในการยอมรับเครื่องมือ กระบวนการ หรือสูตรใหม่ๆในหลากหลายสาขาวิชา ความสอดคล้องของข้อมูลเชิงปริมาณสองชุดสามารถพิจารณาได้จากกราฟของข้อมูลที่ตกอยู่บริเวณเส้นตรง 45 องศา ที่ผ่านจุดกำเนิด

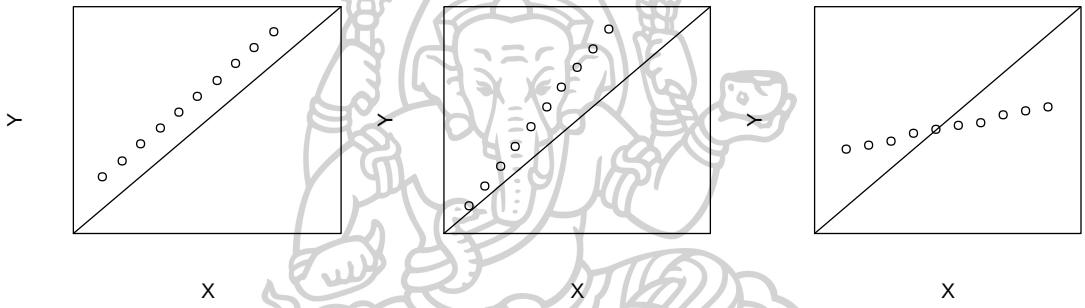


ภาพที่ 4 ความสอดคล้องสมบูรณ์ในทิศทางเดียวกัน ความสอดคล้องสมบูรณ์ในทิศทางตรงกันข้าม และความไม่สอดคล้อง

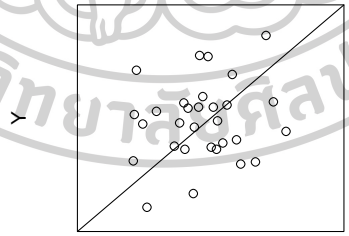
การประเมินเครื่องมือใหม่ที่ถูกคิดค้นขึ้นเทียบกับเครื่องมือที่เป็นมาตรฐานคือการประเมินความสอดคล้องของเครื่องมือ โดยสถิติที่เหมาะสมในการวัดความสอดคล้องคือสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง เนื่องจากสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องวัดการกระจายของข้อมูลที่กระจายออกจากเส้นตรง 45 องศา ที่ผ่านจุดกำเนิด โดยการกระจายของข้อมูลที่ห่างจากเส้นตรง 45 องศา ที่

ผ่านจุดกำเนิด มากขึ้น สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องจะมีค่าลดลง และหมายถึงระดับความสอดคล้องของข้อมูลมีค่าต่ำ

นอกจากนี้สถิติที่ถูกใช้เพื่อวัดความสอดคล้องของเครื่องมือมีอยู่หลากหลายเช่น สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson correlation coefficient), สถิติทดสอบ paired-t, การวิเคราะห์การถดถอยเมื่อความชันเท่ากับ 1 และระยะตัดแกนเท่ากับ 0, สัมประสิทธิ์ความผันแปร (Coefficient of variation), และสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass correlation coefficient) เมื่อพิจารณาสถิติที่กล่าวมานี้จะพบว่าการใช้เฉพาะสถิติใดสถิติหนึ่งจะไม่สามารถประเมินความสอดคล้องได้เพราะการใช้สถิติที่กล่าวมาเพียงชนิดเดียวอาจจะทำให้การแปลผลด้านความสอดคล้องผิดพลาดไปได้ ตัวอย่างการแปลผลที่ผิดพลาดในการวัดความสอดคล้อง คือ



ภาพที่ 5 ตัวอย่างกรณีที่สัมประสิทธิ์สหสัมพันธ์เพียร์สันไม่สามารถตรวจจับความไม่สอดคล้องของข้อมูล



ภาพที่ 6 ตัวอย่างกรณีที่สถิติทดสอบ t และการวิเคราะห์การถดถอยไม่สามารถตรวจจับความไม่สอดคล้องของข้อมูล

จากภาพที่ 5 พบว่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันไม่สามารถวัดความไม่สอดคล้องได้ เนื่องจากข้อมูลมีความสัมพันธ์เชิงเส้นตรงกันอย่างสมบูรณ์ (สัมประสิทธิ์สหสัมพันธ์เพียร์สันเท่ากับ 1) ทำให้เกิดการสรุปผลที่ผิดได้ว่าข้อมูลมีความสอดคล้องกัน แต่จากกราฟพบว่าข้อมูลไม่ได้อยู่บน

เส้นตรง 45 องศาที่ผ่านจุดกำเนิด เนื่องมาจากความแตกต่างกันของค่าเฉลี่ยและความแปรปรวน ซึ่งหมายความว่าข้อมูลไม่สอดคล้องกัน

จากภาพที่ 6 พบว่าสถิติทดสอบ t และการวิเคราะห์การถดถอยไม่สามารถวัดความไม่สอดคล้องของข้อมูลได้ เนื่องจากสถิติทดสอบ t ให้ผลการทดสอบคือค่าเฉลี่ยของทั้งสองตัวแปรเท่ากัน และการวิเคราะห์การถดถอยให้ผลการทดสอบว่าความชันเท่ากับ 1 และระยะตัดแกนเท่ากับ 0 ทำให้เกิดการสรุปผลที่ผิดได้ว่าข้อมูลมีความสอดคล้องกัน แต่จากกราฟพบว่าการกระจายของข้อมูลมีมากทำให้ข้อมูลไม่ได้ตกอยู่บริเวณเส้นตรง 45 องศาที่ผ่านจุดกำเนิด ซึ่งหมายความว่าข้อมูลไม่สอดคล้องกัน[2]

5. สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง

ในการประเมินความสอดคล้องของข้อมูล สถิติที่ถูกใช้มากที่สุดอย่างชนิดหนึ่งคือสัมประสิทธ์สหสัมพันธ์ความสอดคล้อง (Concordance correlation coefficient, CCC)[1]

ให้ตัวแปรสุ่ม (X,Y) มีการแจกแจงสองตัวแปร สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง ρ_c ของตัวแปรสุ่ม X และ Y สามารถเขียนได้ดังนี้

$$\begin{aligned}\rho_c &= \frac{2\sigma_{XY}}{\sigma_X^2+\sigma_Y^2+(\mu_X-\mu_Y)^2} \\ &= \rho C_b\end{aligned} \tag{5}$$

เมื่อ

ρ คือสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

$$C_b = \frac{2\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2+(\mu_X-\mu_Y)^2}$$

$$\mu_X = E(X)$$

$$\mu_Y = E(Y)$$

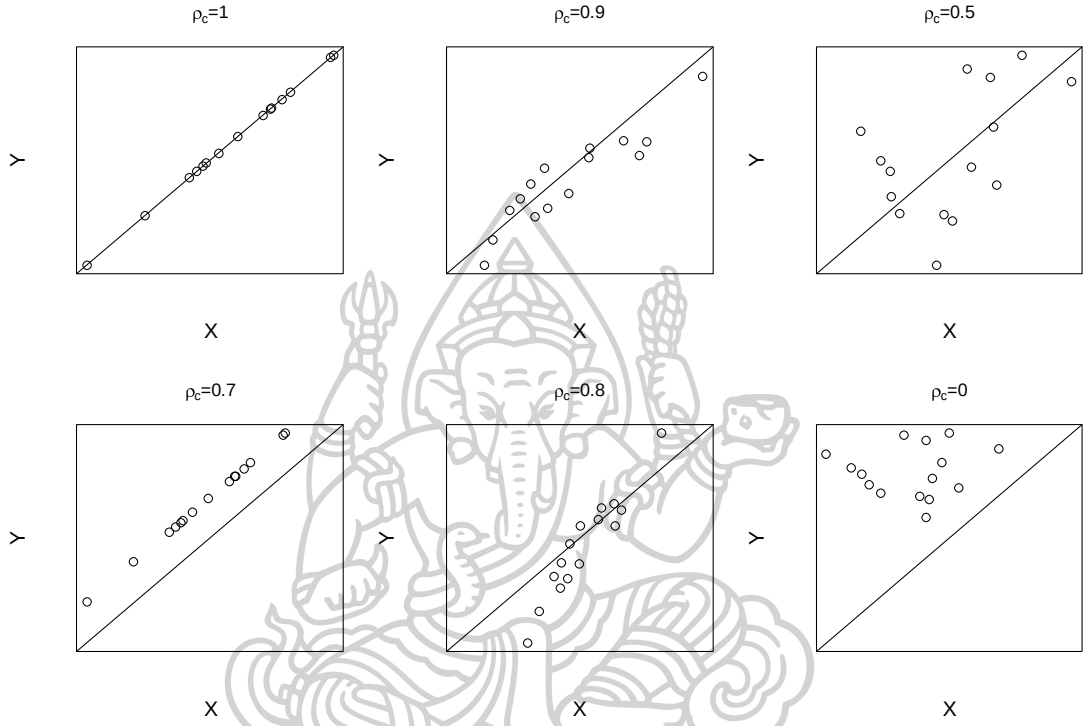
$$\sigma_X^2 = V(X)$$

$$\sigma_Y^2 = V(Y)$$

$$\sigma_{XY} = Cov(X,Y)$$

พิจารณาเส้นถดถอยของข้อมูล X และ Y พบว่า C_b $(0<C_b\le 1)$ หรือค่าวัดความถูกต้อง เป็นค่าวัดระยะห่างที่ เส้นถดถอยเบี่ยงเบนไปจากเส้นตรง 45 องศา ที่ผ่านจุดกำเนิด (วัดความถูกต้อง) กรณีที่เส้นถดถอยไม่ได้เบี่ยงเบนไปจากเส้นตรง 45 องศา ที่ผ่านจุดกำเนิด คือ $C_b=1$ และสัมประสิทธิ์สหสัมพันธ์เพียร์สันวัดระยะห่างที่ข้อมูลแต่ละตัวเบี่ยงเบนไปจากเส้นถดถอย (วัดความแม่นยำ)

สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าอยู่ระหว่าง -1 ถึง 1 เมื่อสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าเท่ากับ -1 หมายถึง X และ Y สอดคล้องกันอย่างสมบูรณ์ในทิศทางตรงกันข้าม , 0 หมายถึง X และ Y ไม่สอดคล้องกัน, และเมื่อมีค่าเท่ากับ 1 หมายถึง X และ Y สอดคล้องกันอย่างสมบูรณ์ในทิศทางเดียวกัน[2]



ภาพที่ 7 ตัวอย่างแผนภาพการกระจายเมื่อ ρ_c มีค่าต่างๆ

ข้อควรระวัง

ข้อมูลที่มีความสอดคล้องกันสูงจะหมายถึงมีความสัมพันธ์กันสูงด้วย แต่ข้อมูลที่มีความสัมพันธ์ที่สูงอาจจะไม่มีความสอดคล้องกัน

ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องคือ

$$\hat{\rho}_c = \frac{2S_{XY}}{S_X^2 + S_Y^2 + \left(\bar{X} - \bar{Y}\right)^2} \tag{6}$$

เมื่อ

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_X^2 = \sum_{i=1}^{n} \frac{\left(X_i - \bar{X}\right)^2}{n}$$

$$S_Y^2 = \sum_{i=1}^{n} \frac{\left(Y_i - \bar{Y}\right)^2}{n}$$

สามารถเขียนตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องในรูปของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันได้ดังนี้

$$\hat{\rho}_c = R\hat{C}_b \tag{7}$$

เมื่อ

$$R = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

$$\hat{C}_b = \frac{2S_X S_Y}{S_X^2 + S_Y^2 + \left(\bar{X} - \bar{Y}\right)^2}$$

ตัวประมาณของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าอยู่ระหว่าง -1 ถึง 1 สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเป็นที่นิยมใช้เนื่องจากสามารถบอกได้ทั้งระดับความสอดคล้อง และทิศทางความสอดคล้องระหว่างค่าวัด อีกทั้งในการอนุมานยังไม่จำเป็นต้องทราบการแจกแจงของตัวสถิติในกรณีที่ตัวอย่างมีขนาดใหญ่ จึงสะดวกในการนำไปใช้[1, 2]

งานวิจัยที่เกี่ยวข้อง

Fisher (1915) ได้แสดงว่าค่าคาดหวังของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของตัวอย่างสุ่มที่มาจากการแจกแจงปกติเป็นตัวประมาณที่เอนเอียง และมีความเอนเอียงเข้าใกล้ 0 เมื่อตัวอย่างมีขนาดใหญ่ขึ้น ซึ่งค่าคาดหวังของตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน คือ

$$E\left(R\right) = \rho - \frac{\rho\left(1-\rho^2\right)}{2n}$$

โดยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีแนวโน้มที่จะประมาณได้ต่ำกว่าค่าพารามิเตอร์เมื่อข้อมูลมีสหสัมพันธ์ทางบวก และมีแนวโน้มที่จะประมาณได้สูงกว่าพารามิเตอร์เมื่อข้อมูลมีสหสัมพันธ์ทางลบ และ Fisher ได้เสนอตัวประมาณที่มีความเอนเอียงเข้าใกล้ 0 เร็วขึ้น เรียกว่า Fisher approximate unbiased estimator [3] นอกจากนี้ Olkin และ Pratt (1958) ได้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่มีความเอนเอียงน้อยกว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันแบบดั้งเดิม[4]

Zimmerman (2003) ได้ทำการเปรียบเทียบตัวประมาณสัมประสิทธิ์สหสัมพันธ์แบบดั้งเดิม ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Fisher และตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt ในเทอมของความเอนเอียง ภายใต้การแจกแจงปกติสองตัวแปร เมื่อตัวอย่างมีขนาดเล็ก และสหสัมพันธ์ทางบวก พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt มีประสิทธิภาพในเทอมของความเอนเอียงดีที่สุด โดยมีความเอนเอียงเข้าใกล้ 0 มาก รองลงมาคือตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Fisher ซึ่งมีความเอนเอียงลดลงจากตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันแบบดั้งเดิม แต่ยังเอนเอียงมากกว่าตัวประมาณของ Olkin และ Pratt เมื่อตัวอย่างมีขนาดเท่ากับ 10[12]

Sinsomboonthong (2013) ได้ทำการเสนอตัวประมาณที่มีประสิทธิภาพดีขึ้นในเทอมของความเอนเอียงโดยใช้เทคนิค jackknife method ในการลดความเอนเอียงของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สัน และเปรียบเทียบตัวประมาณของสัมประสิทธิ์สหสัมพันธ์เพียร์สันแบบดั้งเดิม ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ประยุกต์ใช้วิธี jackknife พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Ptatt และตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ประยุกต์ใช้วิธี jackknife มีความเอนเอียงที่ดีกว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันแบบดั้งเดิมเมื่อตัวอย่างมีขนาดเล็ก[13]

มหาวิทยาลัยศิลปากร

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่มีประสิทธิภาพในด้านการลดความเอนเอียง โดยการประยุกต์ใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทางเลือกอื่นที่มีความเอนเอียงน้อยกว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันเดิมในการสร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับปรุงใหม่ เปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องดั้งเดิมและตัวประมาณที่ปรับใหม่ โดยการศึกษาความเอนเอียงของตัวประมาณดั้งเดิมและตัวประมาณที่ปรับใหม่ภายใต้สถานการณ์ที่แตกต่างกันของพารามิเตอร์ในการแจกแจงปกติสองตัวแปร โดยใช้โปรแกรม R เวอร์ชั่น 3.3.1 สำหรับการจำลองข้อมูล การประมาณค่า และการสร้างกราฟ ผู้วิจัยมีขั้นตอนในการดำเนินงานวิจัยดังนี้

1. สร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่

ให้ตัวแปรสุ่ม $(X_i, Y_i), i = 1, \ldots, n$ มีการแจกแจงสองตัวแปร ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม (สมการที่ 7) คือ

$$\hat{\rho}_c = R\hat{C}_b$$

เมื่อ

$$R = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

$$\hat{C}_b = \frac{2S_X S_Y}{S_X^2 + S_Y^2 + \left(\bar{X} - \bar{Y}\right)^2}$$

สร้างตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยการประยุกต์ใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทางเลือกอื่นที่มีประสิทธิภาพด้านความเอนเอียงดีขึ้นกับตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องได้ดังนี้

1.1 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ประยุกต์ใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

$$\hat{\rho}_{cOP} = \hat{\rho}_{OP}\hat{C}_b \qquad (8)$$

เมื่อ

$\hat{\rho}_{OP} = R\left[1+\frac{1-R^2}{2(n-3)}\right]$ คือตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

1.2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ประยุกต์ใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost

$$\hat{\rho}_{cHP} = \hat{\rho}_{HP}\hat{C}_b \tag{9}$$

เมื่อ

$\hat{\rho}_{HP}$ คือตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost ซึ่งจำแนกเป็น 4 กรณีดังนี้

$$\hat{\rho}_{HP} = \begin{cases} 1-\frac{\sum_{i=1}^{n}(X_i-Y_i)^2}{2n}, (X_i,Y_i) \sim N_2(0,0,1,1,\rho \geq 0) \\ \frac{\sum_{i=1}^{n}(X_i+Y_i)^2}{2n}-1, (X_i,Y_i) \sim N_2(0,0,1,1,\rho < 0) \\ R+\frac{1-R}{n}, (X_i,Y_i) \sim N_2\left((\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2) \neq (0,0,1,1), \rho \geq 0\right) \\ R-\frac{1+R}{n}, (X_i,Y_i) \sim N_2\left((\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2) \neq (0,0,1,1), \rho < 0\right) \end{cases}$$

2. การศึกษาความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยโดยการจำลองข้อมูล

3.1 จำลองข้อมูล $(X_i,Y_i), i=1,\ldots,n$ ที่มีการแจกแจงปกติสองตัวแปร โดยพารามิเตอร์ที่กำหนดคือ

$\mu_X = 0$

$\mu_Y = 0, 0.5, 1.25$

$\sigma_X^2 = 1$

$\sigma_Y^2 = 0.5, 1, 9$

$\rho = -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9$

ที่ขนาดตัวอย่าง $n = 5, 10, 30, 50$

โดยกรณีที่ศึกษาคือ

กรณีที่ 1: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 2: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 0.5$$

กรณีที่ 3: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง:

$$\mu_X = \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 9$$

กรณีที่ 4: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน:

$$\mu_X = 0, \mu_Y = 0.5, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 5: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน:

$$\mu_X = 0, \mu_Y = 1.25, \sigma_X^2 = \sigma_Y^2 = 1$$

กรณีที่ 6: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย:

$$\mu_X = 0, \mu_Y = 0.5, \sigma_X^2 = 1, \sigma_Y^2 = 0.5$$

กรณีที่ 7: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง:

$$\mu_X = 0, \mu_Y = 1.25, \sigma_X^2 = 1, \sigma_Y^2 = 9$$

ที่ทุกขนาดสัมประสิทธิ์สหสัมพันธ์ และทุกขนาดตัวอย่าง

3.2 คำนวณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง ρ_c จาก $\rho_c = \rho C_b$ จะได้ ρ_c ในแต่ละกรณีที่ศึกษาคือ

ตารางที่ 1 ขนาดของค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง ρ_c ในแต่ละกรณีที่ศึกษา

กรณี	C_b	ρ_c
กรณีที่ 1	1	ρ
กรณีที่ 2	0.94	0.94ρ
กรณีที่ 3	0.60	0.6ρ
กรณีที่ 4	0.89	0.89ρ
กรณีที่ 5	0.56	0.56ρ
กรณีที่ 6	0.81	0.81ρ
กรณีที่ 7	0.52	0.52ρ

3.3 คำนวณค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องจากตัวประมาณเดิมและตัวประมาณที่ปรับใหม่ $\hat{\rho}_c, \hat{\rho}_{cOP}, \hat{\rho}_{cHP}$

3.4 ทำซ้ำขั้นตอนที่ 3.1-3.3 ในแต่ละพารามิเตอร์ที่กำหนดจำนวน 10,000 ครั้ง

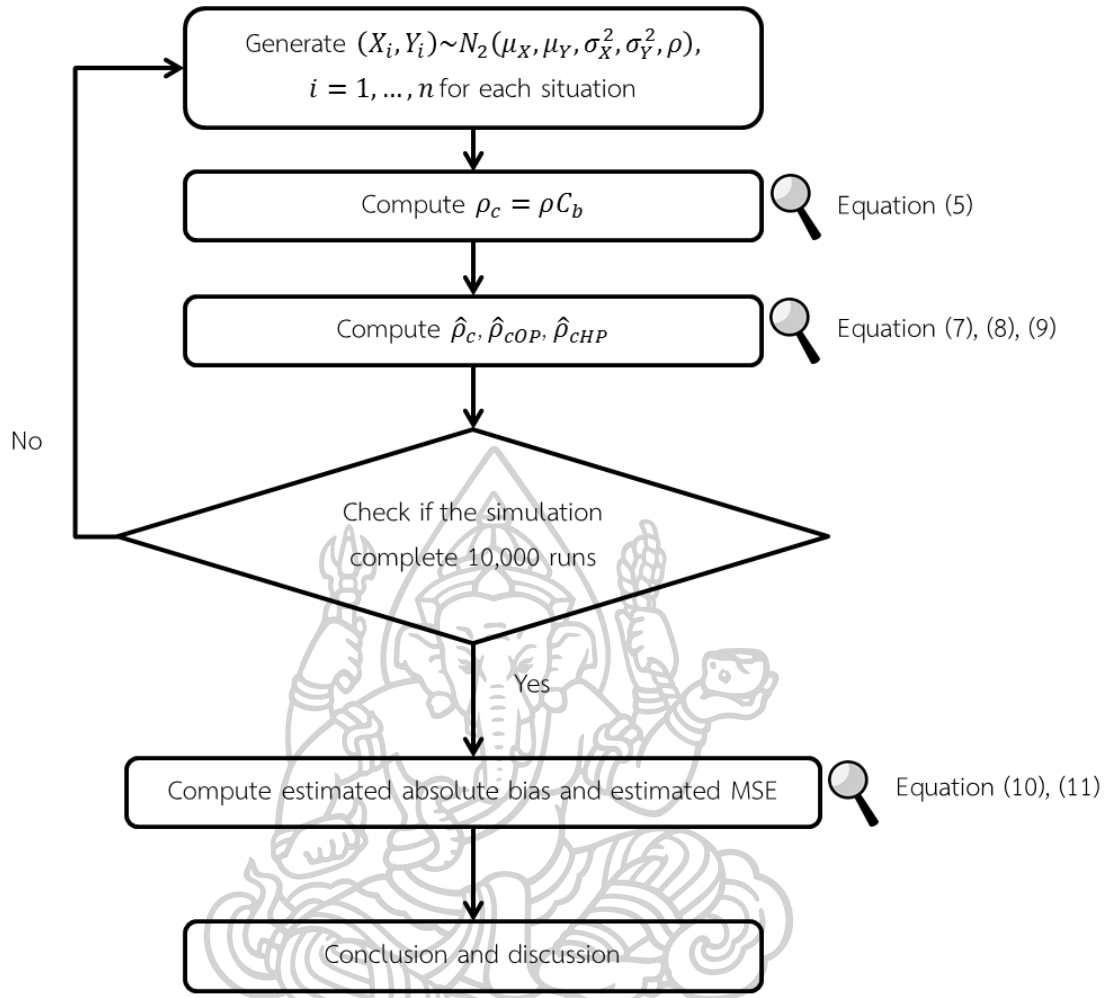
3.5 คำนวณค่าประมาณของความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมและตัวประมาณที่ปรับใหม่

$$\begin{aligned} \left|Bias\left(\hat{\rho}_c\right)\right| &= \left|\frac{\sum\left(\hat{\rho}_c - \rho_c\right)}{10,000}\right| \\ \left|Bias\left(\hat{\rho}_{cOP}\right)\right| &= \left|\frac{\sum\left(\hat{\rho}_{cOP} - \rho_c\right)}{10,000}\right| \\ \left|Bias\left(\hat{\rho}_{cHP}\right)\right| &= \left|\frac{\sum\left(\hat{\rho}_{cHP} - \rho_c\right)}{10,000}\right| \end{aligned} \tag{10}$$

3.6 คำนวณค่าประมาณของความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมและตัวประมาณที่ปรับใหม่จาก

$$\begin{aligned} MSE\left(\hat{\rho}_c\right) &= \frac{\sum\left(\hat{\rho}_c - \rho_c\right)^2}{10,000} \\ MSE\left(\hat{\rho}_{cOP}\right) &= \frac{\sum\left(\hat{\rho}_{cOP} - \rho_c\right)^2}{10,000} \\ MSE\left(\hat{\rho}_{cHP}\right) &= \frac{\sum\left(\hat{\rho}_{cHP} - \rho_c\right)^2}{10,000} \end{aligned} \tag{11}$$

3.7 เปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมและตัวประมาณที่ปรับใหม่ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับในแต่ละพารามิเตอร์และขนาดตัวอย่าง



ภาพที่ 8 แผนภาพแสดงลำดับขั้นตอนการศึกษาความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องโดยการจำลองข้อมูล

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยนี้มีวัตถุประสงค์เพื่อเสนอตัวประมาณที่ปรับใหม่ของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง โดยการประยุกต์ใช้ตัวประมาณทางเลือกอื่นของสัมประสิทธิ์สหสัมพันธ์ เปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ยด้วยการจำลองข้อมูล

การประมาณความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ยด้วยการจำลองข้อมูล

จากแผนการจำลองข้อมูล $(X_i, Y_i), i = 1, \ldots, n$ ที่มีการแจกแจงปกติสองตัวแปร ภายใต้กรณีที่กำหนดคือ

กรณีที่ 1: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร

กรณีที่ 2: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย

กรณีที่ 3: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง

กรณีที่ 4: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน

กรณีที่ 5: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน

กรณีที่ 6: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย

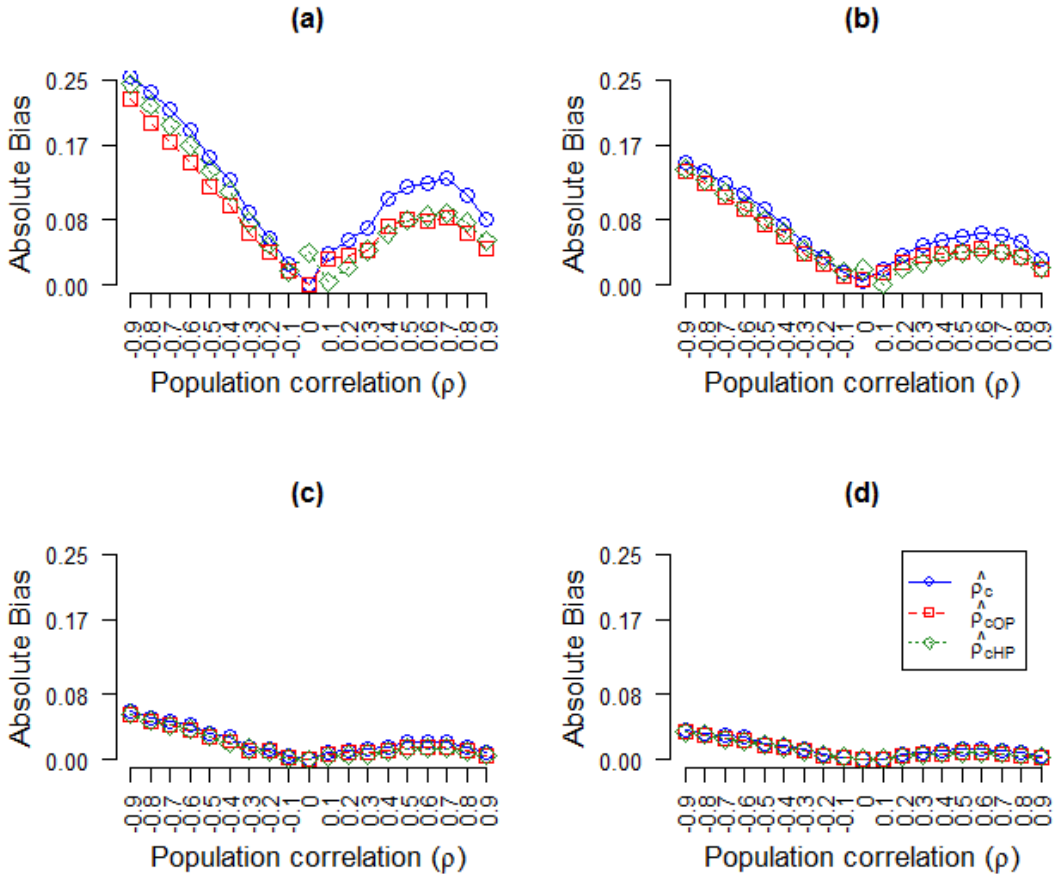
กรณีที่ 7: ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง

ที่ทุกขนาดสัมประสิทธิ์สหสัมพันธ์ และทุกขนาดตัวอย่าง

งานวิจัยนี้จำลองข้อมูลโดยโปรแกรม R มีการทำซ้ำ 10,000 รอบ ผลการวิจัยมีรายละเอียดดังนี้

1. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร(กรณีที่ 1)

1.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร

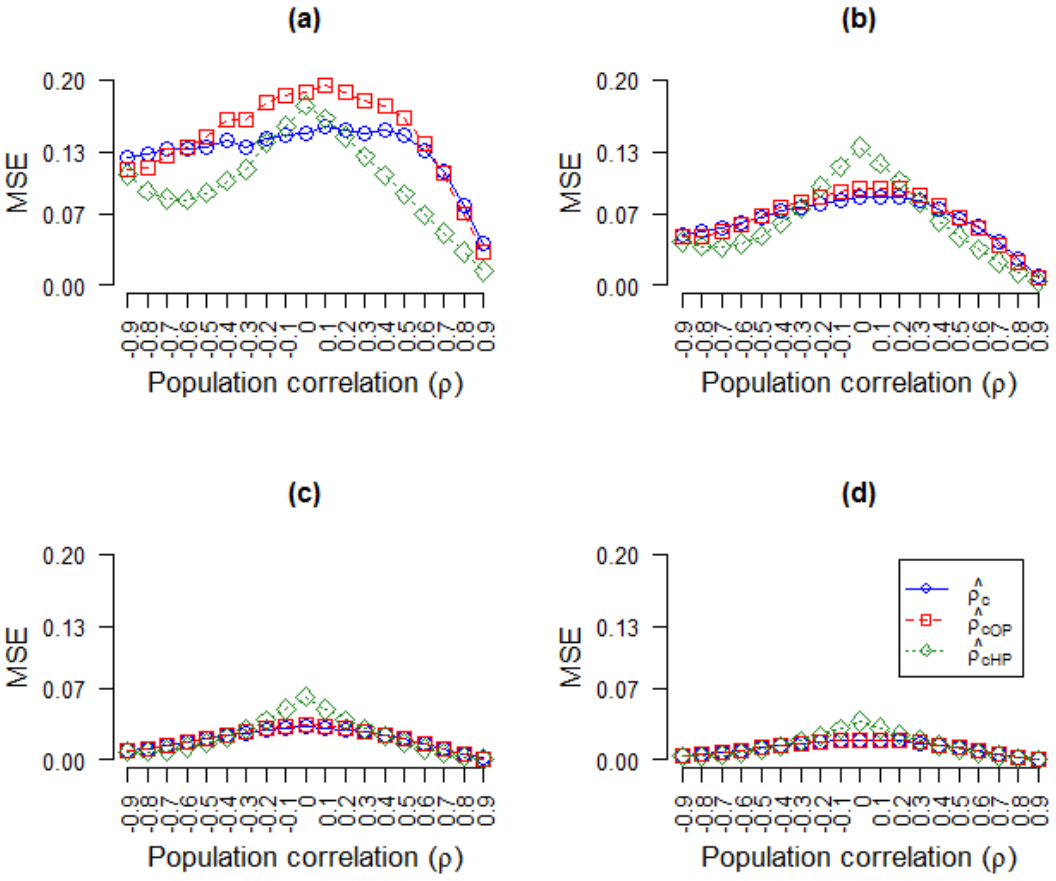


ภาพที่ 9 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร(กรณีที่ 1) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

จากภาพที่ 9 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร หรือค่าวัดความถูกต้องมีค่าเท่ากับ 1 $(C_b=1)$ พบว่าเมื่อตัวอย่างมีขนาดเล็ก $(n=5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และ $\hat{\rho}_{cHP}$ มีประสิทธิภาพดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ยกเว้นเมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ 0 ซึ่งก็คือสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมี

ค่าเท่ากับ 0 $(\rho_c = 0)$ ตัวประมาณ $\hat{\rho}_{cHP}$ จะมีความเอนเอียงสัมบูรณ์สูงกว่าตัวประมาณอื่นๆ เล็กน้อย

เมื่อตัวอย่างมีขนาดใหญ่ $(n = 30, 50)$ พบว่าตัวประมาณทั้ง 3 จะมีประสิทธิภาพในเทอมของความเอนเอียงใกล้เคียงกันและมีความเอนเอียงสัมบูรณ์เข้าใกล้ 0



ภาพที่ 10 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลไม่มีความแตกต่างของค่าเฉลี่ยและความแปรปรวน (กรณีที่ 1) เมื่อขนาดตัวอย่าง (a) $n = 5$, (b) $n = 10$, (c) $n = 30$, (d) $n = 50$

1.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร

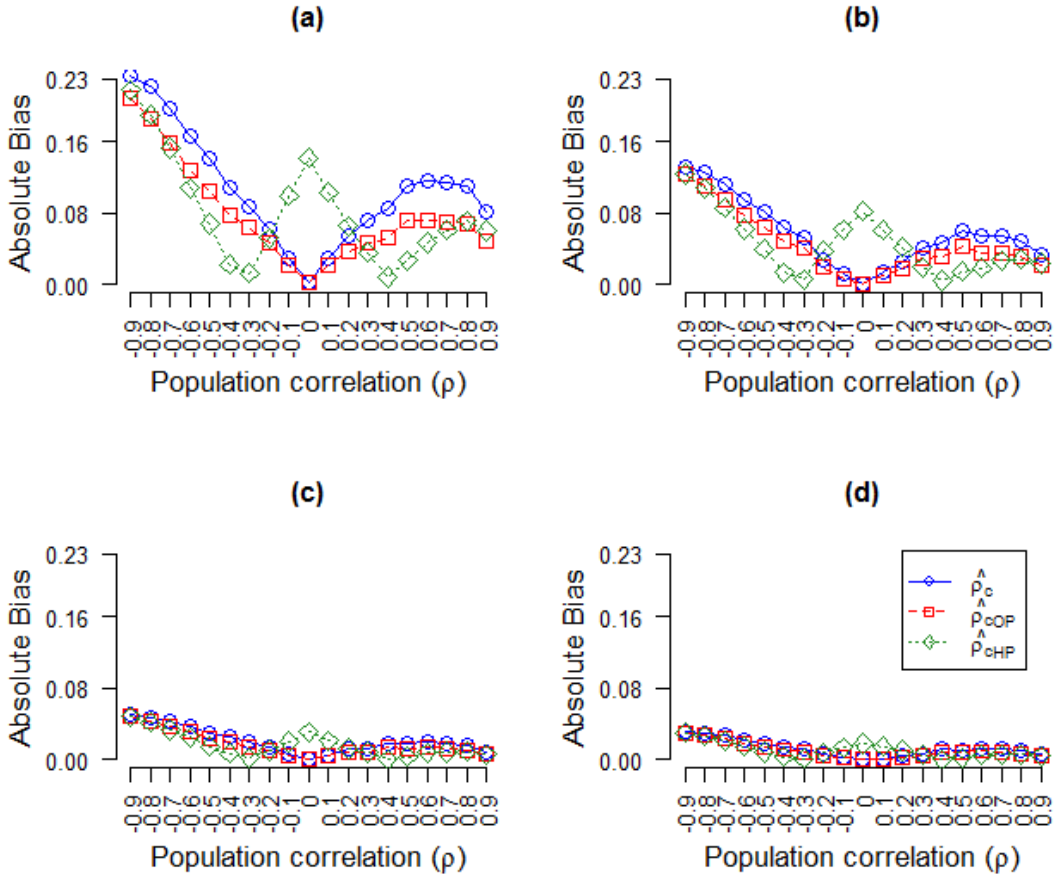
จากภาพที่ 10 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูก

สุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร $(C_b = 1)$ พบว่า เมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho| > 0.6)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าสูง $(|\rho_c| > 0.6)$ และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณ $\hat{\rho}_{cOP}$ และ $\hat{\rho}_c$ มีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกัน และ MSE เข้าใกล้ 0

พิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามี MSE น้อยกว่าตัวประมาณ $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างต่ำถึงสูง $(|\rho| > 0.2)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c| > 0.2)$ ที่ทุกขนาดตัวอย่าง

2. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2)

2.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2)



ภาพที่ 11 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

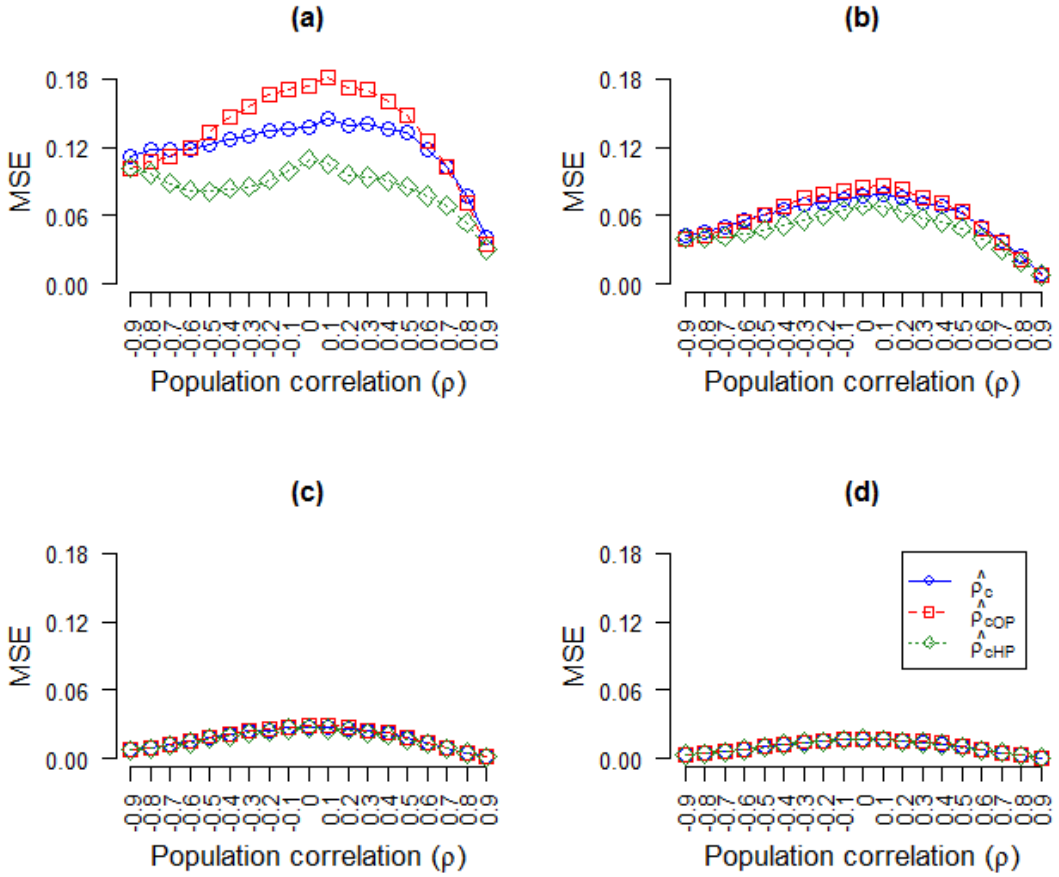
จากภาพที่ 11 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจง

ปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย $(C_b = 0.94)$ พบว่าเมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิมมีความเอนเอียงสัมบูรณ์ไม่แตกต่างกันและเข้าใกล้ 0

เมื่อพิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างต่ำถึงสูง $(|\rho| > 0.2)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c| > 0.188)$ ที่ทุกขนาดตัวอย่าง

2.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2)

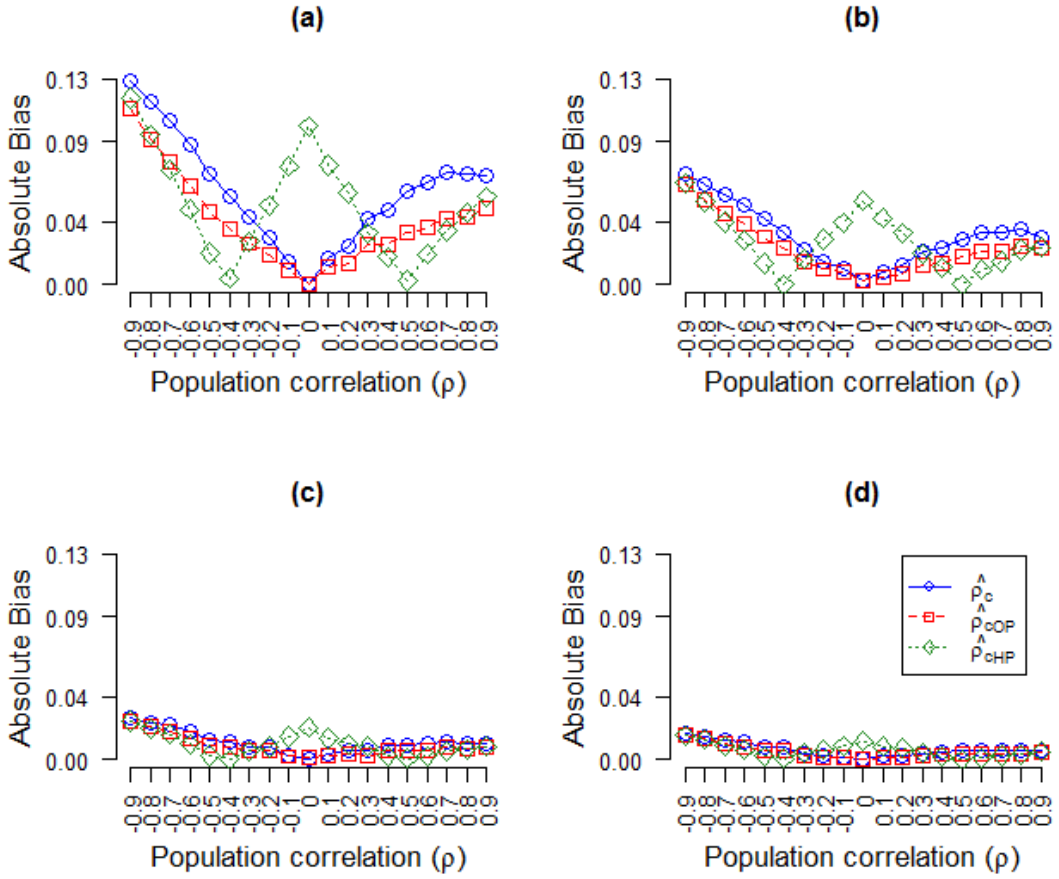
จากภาพที่ 12 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย $(C_b = 0.94)$ พบว่า เมื่อตัวอย่างมีขนาดเล็กมาก $(n = 5)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho| > 0.6)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าสูง $(|\rho_c| > 0.564)$ และเมื่อพิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ หรือในทำนองเดียวกันคือทุกระดับของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง เมื่อตัวอย่างมีขนาด 10 ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ มีประสิทธิภาพดีกว่าตัวประมาณเดิมเล็กน้อย และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 12 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 2) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

3. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3)

3.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง



ภาพที่ 13 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

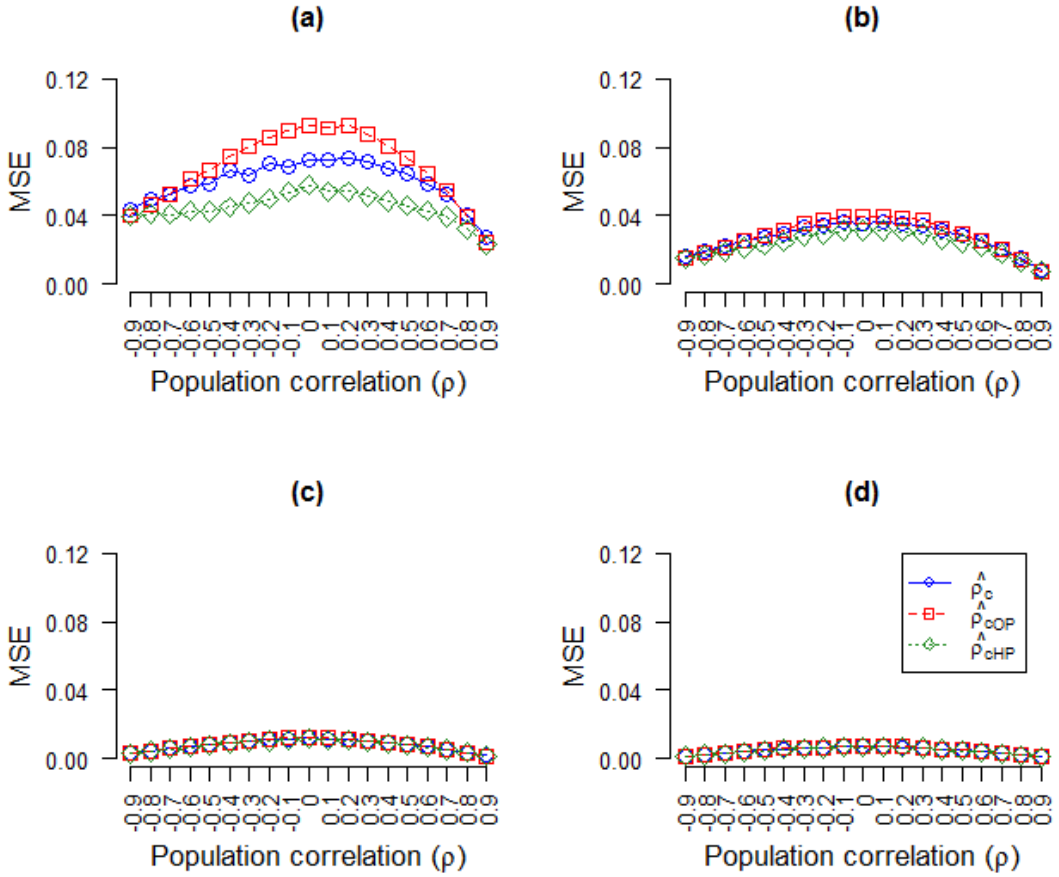
จากภาพที่ 13 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจง

ปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง $(C_b = 0.60)$ พบว่าลักษณะของความเอนเอียงสัมบูรณ์คล้ายกับกรณีที่ 2 นั่นคือ เมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิม มีความเอนเอียงสัมบูรณ์ไม่แตกต่างกันและเข้าใกล้ 0

เมื่อพิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างต่ำถึงสูง $(|\rho| > 0.2)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c| > 0.12)$ ที่ทุกขนาดตัวอย่าง

3.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง

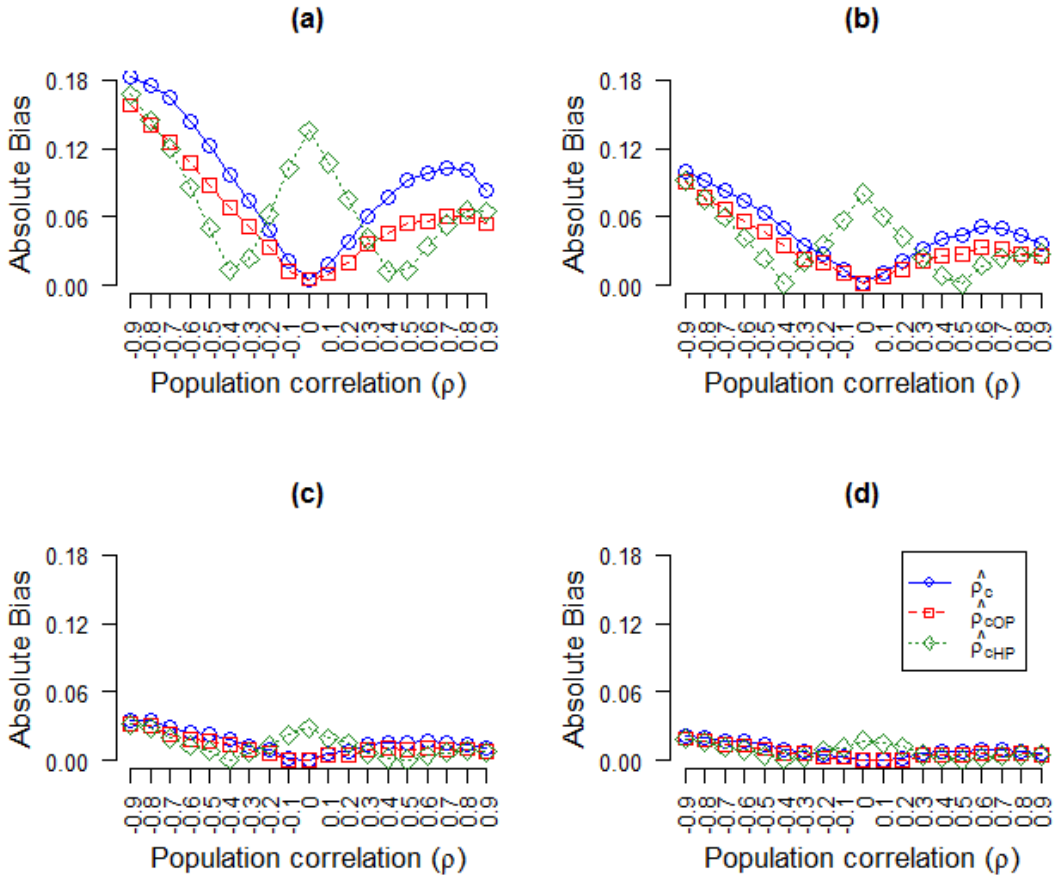
จากภาพที่ 14 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง $(C_b = 0.60)$ พบว่า ลักษณะของ MSE คล้ายกับกรณีที่ 2 นั่นคือ เมื่อตัวอย่างมีขนาดเล็กมาก $(n = 5)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho| > 0.6)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าปานกลางถึงสูง $(|\rho_c| > 0.36)$ และเมื่อพิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c เมื่อตัวอย่างมีขนาด 10 ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ มีประสิทธิภาพดีกว่าตัวประมาณเดิมเล็กน้อย และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 14 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่ไม่มีความแตกต่างของค่าเฉลี่ย แต่ความแปรปรวนต่างกันปานกลาง (กรณีที่ 3) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

4. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4)

4.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน



ภาพที่ 15 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

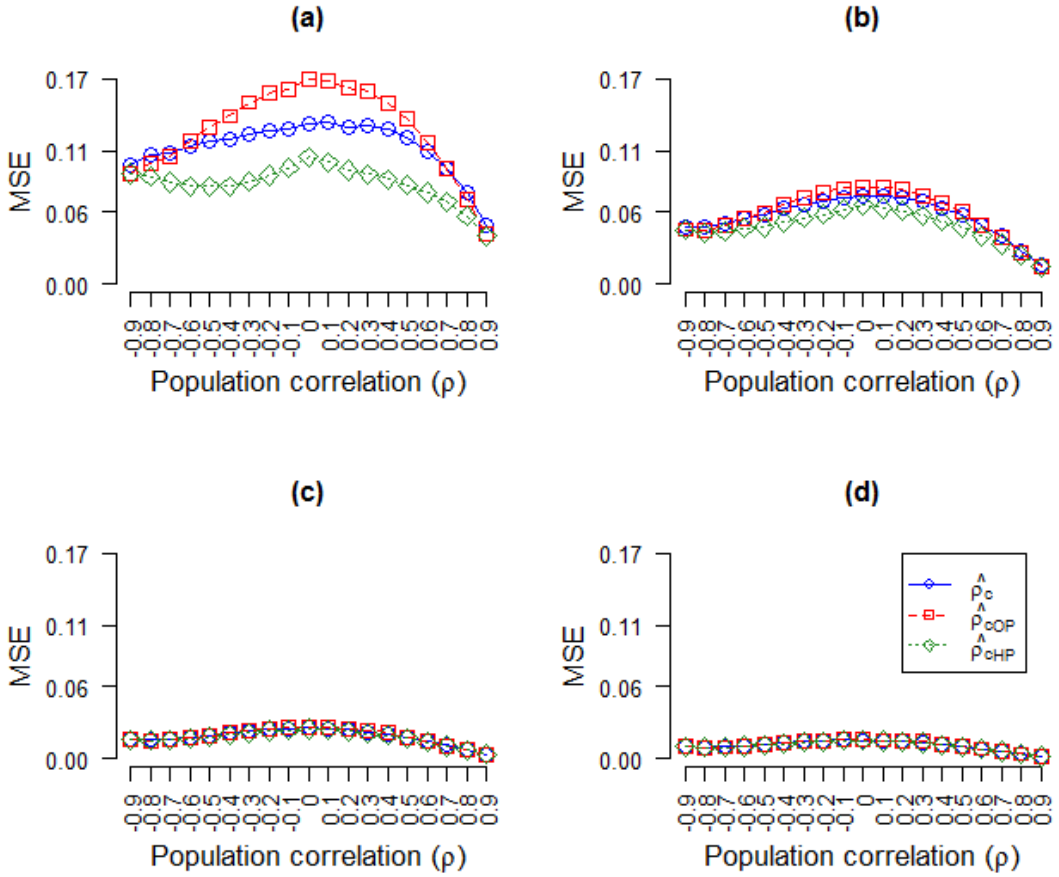
จากภาพที่ 15 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน $(C_b = 0.89)$

พบว่าลักษณะของความเอนเอียงสัมบูรณ์คล้ายกับกรณีที่ 2 และ 3 นั่นคือ เมื่อตัวอย่างมีขนาดเล็ก $(n=5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิม มีความเอนเอียงสัมบูรณ์ไม่แตกต่างกันและเข้าใกล้ 0

เมื่อพิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างต่ำถึงสูง $(|\rho|>0.2)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c|>0.178)$ ที่ทุกขนาดตัวอย่าง

4.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน

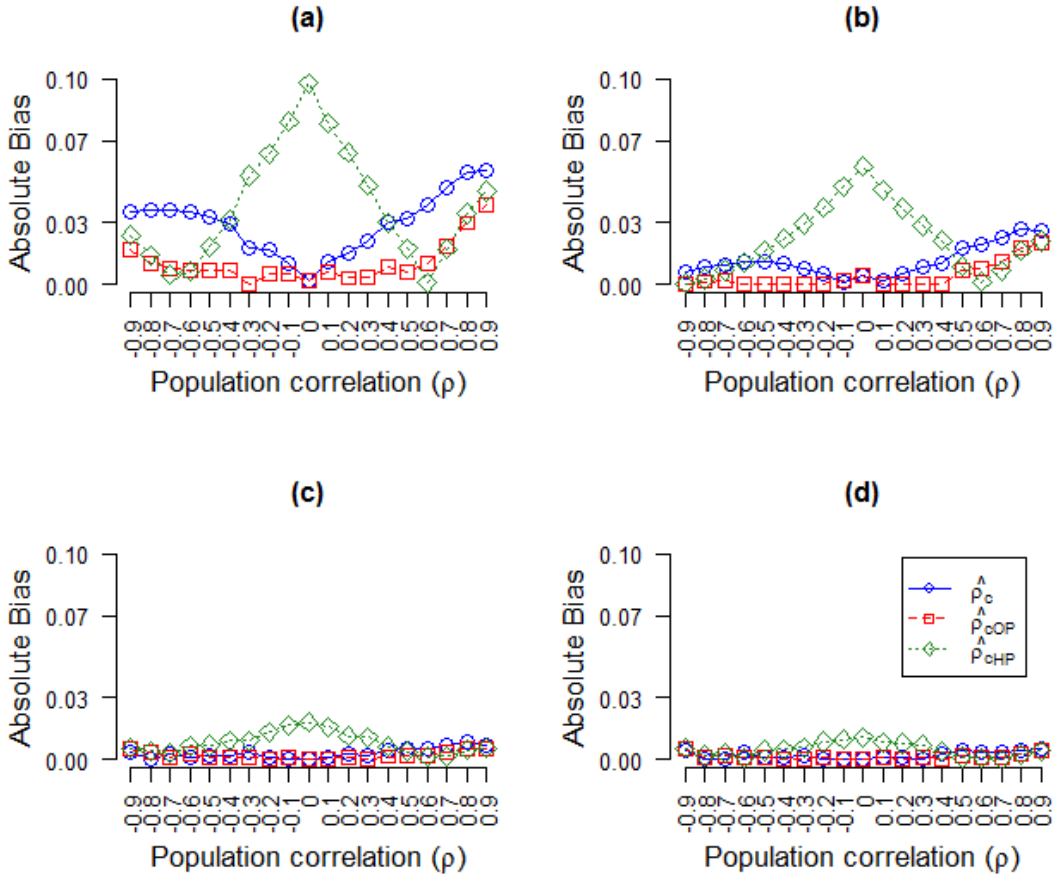
จากภาพที่ 16 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP},\hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน $(C_b=0.89)$ พบว่าประสิทธิภาพของตัวประมาณมีรูปแบบคล้ายกรณีที่ 2 และกรณีที่ 3 นั่นคือ เมื่อตัวอย่างมีขนาดเล็กมาก $(n=5)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho|>0.6)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c|>0.534)$ และเมื่อพิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c เมื่อตัวอย่างมีขนาด 10 ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ มีประสิทธิภาพดีกว่าตัวประมาณเดิมเล็กน้อย และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 16 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันเล็กน้อย แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 4) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

5. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5)

5.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน



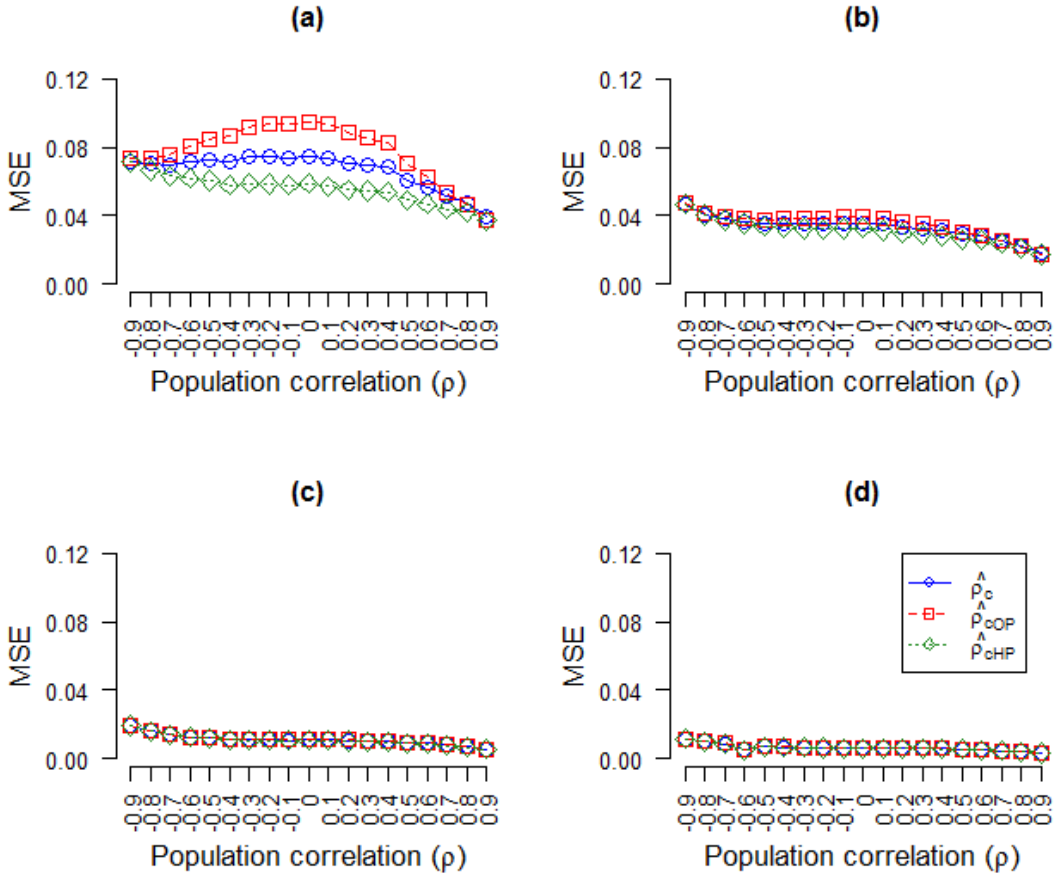
ภาพที่ 17 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

จากภาพที่ 17 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน

$(C_b = 0.56)$ พบว่า เมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และมีความเอนเอียงสัมบูรณ์เข้าใกล้ 0 มาก เมื่อพิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างสูง $(|\rho| > 0.4)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c| > 0.224)$ และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิม มีความเอนเอียงสัมบูรณ์ไม่แตกต่างกันและเข้าใกล้ 0

5.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน

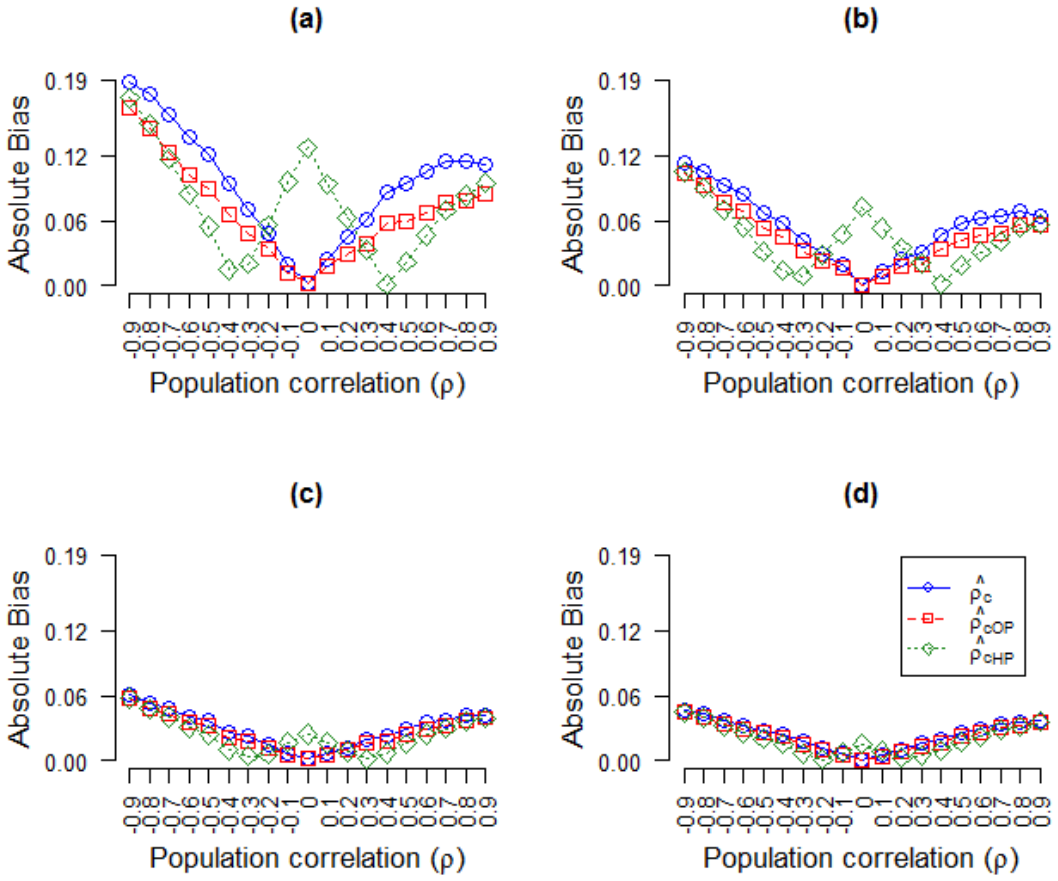
จากภาพที่ 18 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน $(C_b = 0.56)$ พบว่าเมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ มี MSE ต่ำกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ในทำนองเดียวกันคือทุกระดับของ ρ_c และตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพไม่ต่างจากตัวประมาณเดิม เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho| > 0.7)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างสูง $(|\rho_c| > 0.392)$ และเมื่อตัวอย่างมีขนาดใหญ่ $(n = 30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 18 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง แต่ไม่มีความแตกต่างของความแปรปรวน (กรณีที่ 5) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

6. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6)

6.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย



ภาพที่ 19 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

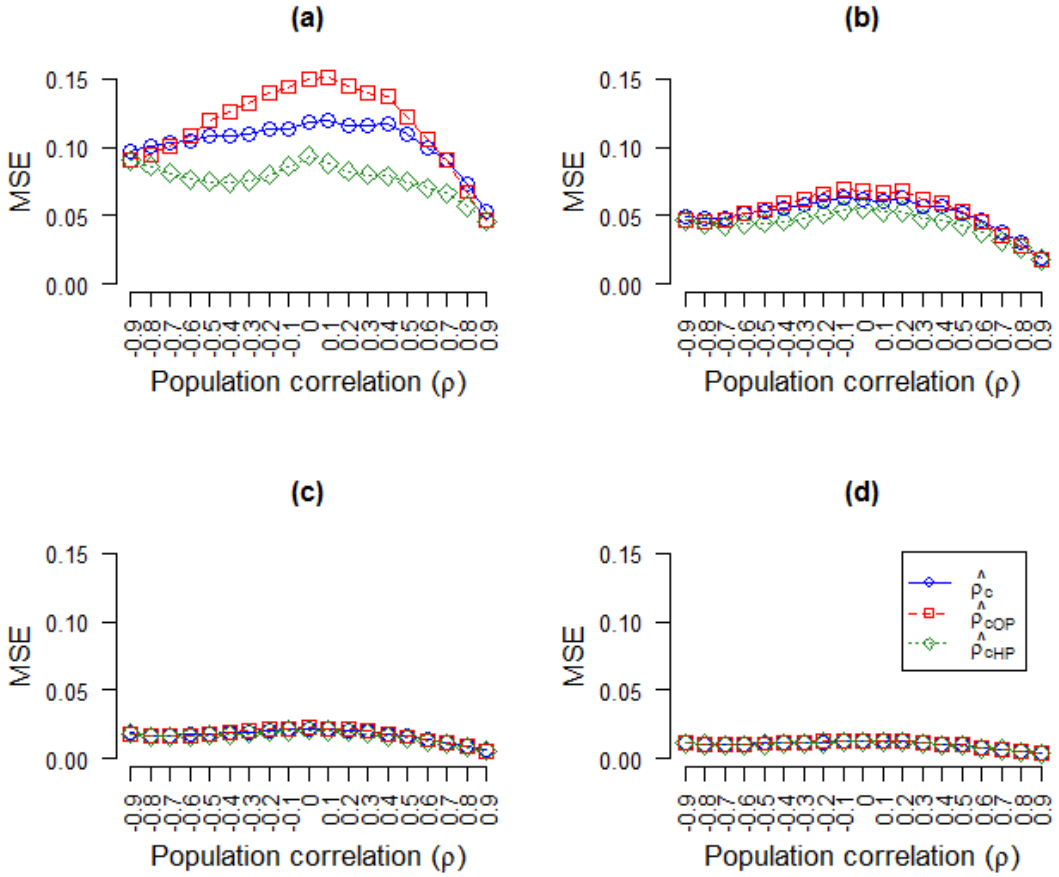
จากภาพที่ 19 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย $(C_b = 0.81)$ พบว่าลักษณะของความเอนเอียงสัมบูรณ์คล้ายกับกรณีที่ 2 กรณีที่ 3 และกรณีที่ 4 นั่นคือ เมื่อตัวอย่างมีขนาดเล็ก

$(n=5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ในทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิม มีความเอนเอียงสัมบูรณ์ไม่แตกต่างกันและเข้าใกล้ 0

เมื่อพิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าค่อนข้างต่ำถึงสูง $(|\rho|>0.2)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างต่ำถึงสูง $(|\rho_c|>0.162)$ ที่ทุกขนาดตัวอย่าง

6.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย

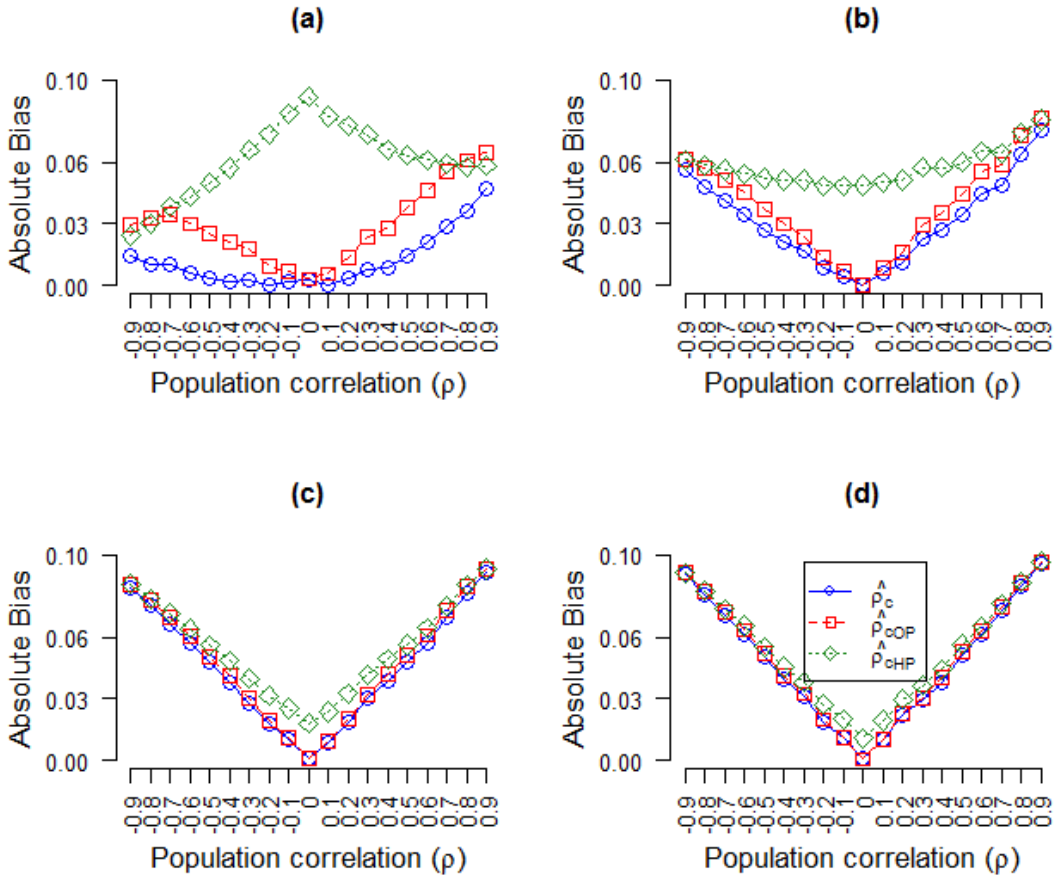
จากภาพที่ 20 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP},\hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย $(C_b=0.81)$ พบว่าประสิทธิภาพของตัวประมาณมีรูปแบบคล้ายกรณีที่ 2 กรณีที่ 3 และกรณีที่ 4 นั่นคือ เมื่อตัวอย่างมีขนาดเล็กมาก $(n=5)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho|>0.6)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าค่อนข้างสูง $(|\rho_c|>0.486)$ และเมื่อพิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามี MSE น้อยกว่าตัวประมาณเดิม $\hat{\rho}_c$ ทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c เมื่อตัวอย่างมีขนาด 10 ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ มีประสิทธิภาพดีกว่าตัวประมาณเดิมเล็กน้อย และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 20 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันเล็กน้อย (กรณีที่ 6) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

7. กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7)

7.1 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความเอนเอียงสัมบูรณ์ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง



ภาพที่ 21 ความเอนเอียงสัมบูรณ์ของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

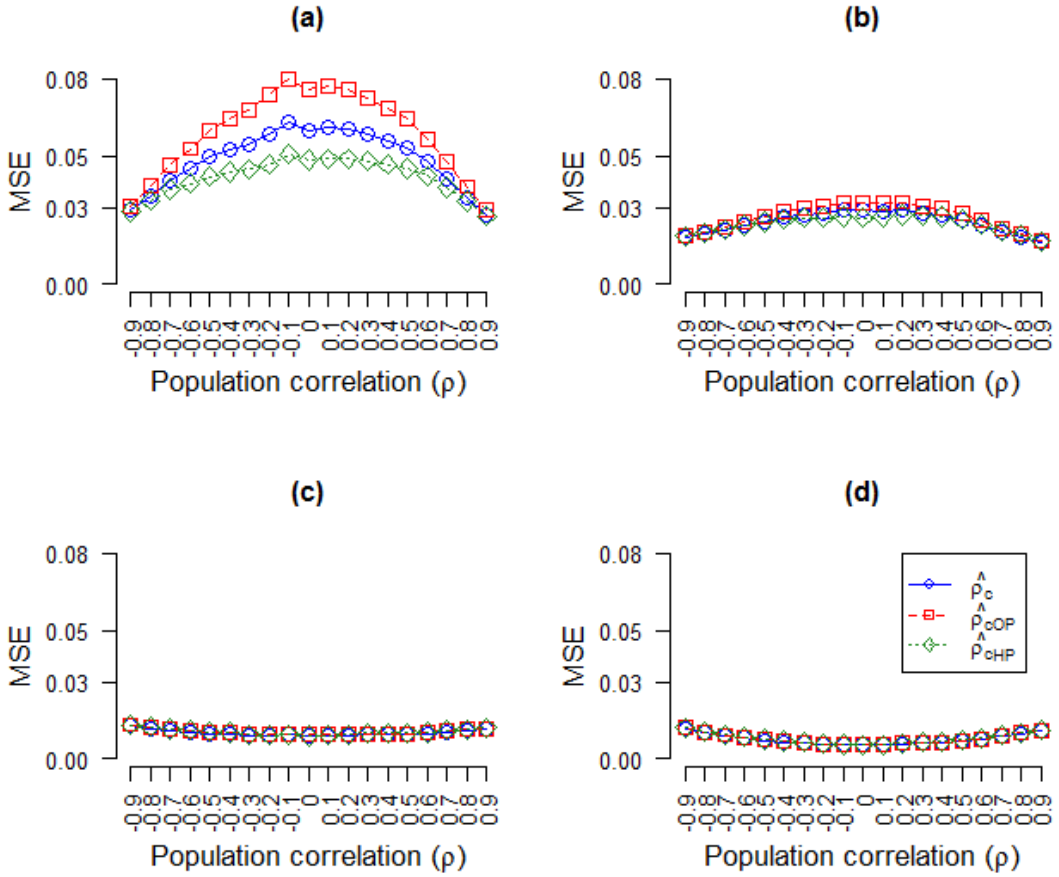
จากภาพที่ 21 เปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP}, \hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง $(C_b = 0.52)$ พบว่าเมื่อตัวอย่างมีขนาดเล็ก $(n = 5,10)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพในเทอมของความเอนเอียง

สัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ทุกระดับของสหสัมพันธ์ ρ ในทำนองเดียวกันคือทุกระดับของ ρ_c และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ และตัวประมาณเดิมมีความเอนเอียงสัมบูรณ์ไม่แตกต่างกัน

เมื่อตัวอย่างมีขนาดเล็กมาก $(n=5)$ พิจารณาตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ พบว่ามีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ดีกว่าตัวประมาณเดิม $\hat{\rho}_c$ ที่ขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho|>0.7)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าปานกลางถึงสูง $(|\rho_c|>0.364)$ สำหรับกรณีที่ขนาดตัวอย่าง $n=10,30,50$ ตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cHP}$ มีประสิทธิภาพในเทอมของความเอนเอียงสัมบูรณ์ไม่แตกต่างจากตัวประมาณเดิม $\hat{\rho}_c$ ที่ขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho|>0.7)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าปานกลางถึงสูง $(|\rho_c|>0.364)$

7.2 พิจารณาประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิมในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ย เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง

จากภาพที่ 22 เปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ย(MSE) ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cOP},\hat{\rho}_{cHP}$ และตัวประมาณเดิม $\hat{\rho}_c$ เมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง $(C_b=0.52)$ พบว่าประสิทธิภาพของตัวประมาณมีรูปแบบคล้ายกรณีที่ 5 นั่นคือ เมื่อตัวอย่างมีขนาดเล็ก $(n=5,10)$ ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ $\hat{\rho}_{cHP}$ มี MSE ต่ำกว่าตัวประมาณเดิม $\hat{\rho}_c$ ทุกระดับของสหสัมพันธ์ ρ ซึ่งก็คือทุกระดับของ ρ_c และตัวประมาณที่ปรับใหม่ $\hat{\rho}_{cOP}$ มีประสิทธิภาพไม่ต่างจากตัวประมาณเดิม เมื่อขนาดของสัมประสิทธิ์สหสัมพันธ์มีค่าสูง $(|\rho|>0.7)$ ซึ่งก็คือขนาดของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าปานกลางถึงสูง $(|\rho_c|>0.364)$ และเมื่อตัวอย่างมีขนาดใหญ่ $(n=30,50)$ ตัวประมาณทั้งสามมีประสิทธิภาพในเทอมของ MSE ไม่แตกต่างกันและ MSE เข้าใกล้ 0



ภาพที่ 22 ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณความสอดคล้องที่ปรับใหม่และตัวประมาณความสอดคล้องเดิม กรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยและความแปรปรวนต่างกันปานกลาง (กรณีที่ 7) เมื่อขนาดตัวอย่าง (a) $n=5$, (b) $n=10$, (c) $n=30$, (d) $n=50$

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ ที่สามารถใช้เป็นเครื่องมือในการวัดความสอดคล้องกันของข้อมูล และเปรียบเทียบประสิทธิภาพของตัวประมาณที่นำเสนอใหม่และตัวประมาณเดิมในเทอมของความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ย

จากผลการเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt, ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost, และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ย สามารถสรุปได้ดังนี้

สรุปผลการวิจัย

1. ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

จากการเปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม พบว่าตัวประมาณที่ปรับใหม่มีความเอนเอียงสัมบูรณ์ต่ำกว่าตัวประมาณเดิมทุกระดับของสัมประสิทธิ์สหสัมพันธ์ $\left(\rho = -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9\right)$ ซึ่งก็คือทุกระดับของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง (ดูตารางที่ 1) เมื่อตัวอย่างมีขนาดเล็ก $\left(n = 5, 10\right)$ โดยเฉพาะในกรณีที่ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยของข้อมูลต่างกันปานกลาง ตัวประมาณที่ปรับใหม่มีความเอนเอียงสัมบูรณ์เข้าใกล้ 0 มาก และเมื่อตัวอย่างมีขนาดใหญ่ ตัวประมาณที่ปรับใหม่และตัวประมาณเดิมมีความเอนเอียงสัมบูรณ์ไม่แตกต่างกัน และมีความเอนเอียงสัมบูรณ์เข้าใกล้ 0 มาก

จากการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม พบว่าตัวประมาณที่ปรับใหม่มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าตัวประมาณเดิม เมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าสูง นั่นคือ

สัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าสูง และขนาดตัวอย่างเล็ก ยกเว้นกรณีที่ข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยของข้อมูลต่างกันปานกลาง ตัวประมาณทั้งสองมีความคลาดเคลื่อนกำลังสองเฉลี่ยไม่แตกต่างกัน และเมื่อตัวอย่างมีขนาดใหญ่ ตัวประมาณที่ปรับใหม่และตัวประมาณเดิมมีความคลาดเคลื่อนกำลังสองเฉลี่ยไม่แตกต่างกัน และมีความคลาดเคลื่อนกำลังสองเฉลี่ยเข้าใกล้ 0 มาก

2. ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost

จากการเปรียบเทียบความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม พบว่าเมื่อข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐานสองตัวแปร ตัวประมาณที่ปรับใหม่มีความเอนเอียงสัมบูรณ์ต่ำกว่าตัวประมาณเดิมเมื่อสัมประสิทธิ์สหสัมพันธ์ไม่เท่ากับ 0 และกรณีข้อมูลถูกสุ่มมาจากการแจกแจงปกติสองตัวแปรที่มีค่าเฉลี่ยต่างกันปานกลาง นั่นคือกรณีที่ ตัวประมาณที่ปรับใหม่มีความเอนเอียงสัมบูรณ์ต่ำกว่าตัวประมาณเดิมเมื่อระดับของสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่าสูงและตัวอย่างขนาดเล็ก สำหรับกรณีที่ 2, 3, 4 และ 6 พบว่าตัวประมาณที่ปรับใหม่มีความเอนเอียงสัมบูรณ์ต่ำกว่าตัวประมาณเดิมเมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีขนาดค่อนข้างต่ำถึงสูง หรือก็คือเมื่อสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีขนาดค่อนข้างต่ำถึงสูง ที่ทุกขนาดตัวอย่าง

จากการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม พบว่าเมื่อตัวอย่างมีขนาดเล็ก ตัวประมาณที่ปรับใหม่มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าตัวประมาณเดิม ทุกระดับของสัมประสิทธิ์สหสัมพันธ์ หรือก็คือทุกระดับของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง และเมื่อตัวอย่างมีขนาดใหญ่ ตัวประมาณที่ปรับใหม่และตัวประมาณเดิมมีความคลาดเคลื่อนกำลังสองเฉลี่ยไม่แตกต่างกัน และมีความคลาดเคลื่อนกำลังสองเฉลี่ยเข้าใกล้ 0 มาก ยกเว้นกรณีที่ข้อมูลถูกสุ่มมาจากการแจกแจงปกติมาตรฐาน พบว่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณที่ปรับใหม่มีค่าต่ำกว่าตัวประมาณเดิม เมื่อระดับของสัมประสิทธิ์สหสัมพันธ์มีขนาดค่อนข้างต่ำถึงสูง หรือก็คือระดับของสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีขนาดค่อนข้างต่ำถึงสูง และตัวอย่างขนาดเล็ก และความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณทั้งสองไม่แตกต่างกันเมื่อระดับของสัมประสิทธิ์สหสัมพันธ์มีขนาดปานกลางถึงสูงและตัวอย่างขนาดใหญ่

สามารถสรุปตัวประมาณที่ดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละกรณี โดยกำหนดสัญลักษณ์

Ori แทน ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม

OP แทน ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt

HP แทน ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost

ได้ดังนี้

ตารางที่ 2 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละระดับของ ค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ ρ เมื่อขนาดตัวอย่างเท่ากับ 5

ρ	C_b						
	1	0.94	0.89	0.81	0.6	0.56	0.52
-0.9	OP HP	OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.8	OP HP	OP HP	HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.7	HP	HP	HP	HP	HP	Ori HP	Ori OP HP
-0.6	HP	HP	HP	HP	HP	Ori HP	Ori HP
-0.5	HP	HP	HP	HP	HP	HP	Ori HP
-0.4	HP	HP	HP	HP	HP	HP	HP
-0.3	HP	HP	HP	HP	HP	HP	Ori HP
-0.2	Ori HP	HP	HP	HP	HP	HP	HP
-0.1	Ori HP	HP	HP	HP	HP	HP	Ori HP
0	Ori	HP	HP	HP	HP	HP	HP
0.1	Ori HP	HP	HP	HP	HP	HP	HP
0.2	Ori HP	HP	HP	HP	HP	HP	HP
0.3	HP	HP	HP	HP	HP	HP	HP
0.4	HP	HP	HP	HP	HP	HP	Ori HP
0.5	HP	HP	HP	HP	HP	HP	Ori HP
0.6	HP	HP	HP	HP	HP	Ori HP	Ori HP

ρ	C_b						
	1	0.94	0.89	0.81	0.6	0.56	0.52
0.7	HP	HP	HP	HP	HP	Ori OP HP	Ori OP HP
0.8	HP	HP	HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
0.9	HP	OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP

*หมายเหตุ กรณีตัวประมาณที่ดีที่สุดมากกว่า 1 ชนิด คือ MSE ต่างกันไม่เกิน 0.01

ตารางที่ 3 ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละระดับของ ค่าวัดความถูกต้อง C_b และสัมประสิทธิ์สหสัมพันธ์ ρ เมื่อขนาดตัวอย่างเท่ากับ 10

ρ	C_b						
	1	0.94	0.89	0.81	0.6	0.56	0.52
-0.9	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.8	HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.7	HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.6	HP	OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.5	HP	HP	Ori HP	Ori HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.4	HP	HP	HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.3	Ori OP HP	HP	Ori HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.2	Ori OP	HP	Ori HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
-0.1	Ori OP	HP	Ori HP	Ori HP	Ori OP HP	Ori OP HP	Ori OP HP
0	Ori OP	HP	Ori HP	Ori HP	Ori OP HP	Ori OP HP	Ori OP HP
0.1	Ori OP	HP	Ori HP	Ori HP	Ori OP HP	Ori OP HP	Ori OP HP
0.2	Ori OP	HP	Ori HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
0.3	Ori OP HP	HP	Ori HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
0.4	HP	HP	HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
0.5	HP	HP	Ori HP	HP	Ori OP HP	Ori OP HP	Ori OP HP
0.6	HP	OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
0.7	HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP
0.8	HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP	Ori OP HP

ρ	C_b																				
	1			0.94			0.89			0.81			0.6			0.56			0.52		
0.9	Ori	OP	HP	Ori	OP	HP	Ori	OP	HP	Ori	OP	HP	Ori	OP	HP	Ori	OP	HP	Ori	OP	HP

*หมายเหตุ กรณีตัวประมาณที่ดีที่สุดมากกว่า 1 ชนิด คือ MSE ต่างกันไม่เกิน 0.01

จากตารางที่ 2 และตารางที่ 3 พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost มีประสิทธิภาพดีที่สุดในเทอมของความคลาดเคลื่อนกำลังสองเฉลี่ยเกือบทุกระดับของสัมประสิทธิ์สหสัมพันธ์เพียร์สันและ C_b อย่างไรก็ตาม เมื่อขนาดตัวอย่างเพิ่มขึ้น และ C_b มีขนาดเล็กลง พบว่าตัวประมาณทั้งสามสามารถใช้ได้ในทุกระดับของสัมประสิทธิ์สหสัมพันธ์

สำหรับกรณีที่ตัวอย่างมีขนาด 30 และ 50 พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องทั้ง 3 รูปแบบ มีความคลาดเคลื่อนกำลังสองเฉลี่ยต่างกันไม่เกิน 0.01 ในทุกๆระดับสัมประสิทธิ์หสหสัมพันธ์และทุกระดับของ C_b ดังนั้นสามารถใช้ได้ทั้ง 3 รูปแบบ ยกเว้นกรณีที่ข้อมูลมีการแจกแจงปกติมาตรฐาน หรือ $C_b=1$ พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost มีความคลาดเคลื่อนกำลังสองเฉลี่ยสูงกว่าตัวประมาณอีกสองชนิดเล็กน้อย (<0.03) เมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีขนาด -0.1, 0 และ 0.1

อภิปรายผล

งานวิจัยนี้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ทางเลือกอื่น คือตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณของ Haddad และ Provost และศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่และตัวประมาณเดิม ในเทอมของความเอนเอียงสัมบูรณ์ และความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งสามารถอภิปรายผลการวิจัยได้ดังนี้

เมื่อพิจารณาความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันเดิม พบว่า ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่มีความเอนเอียงสัมบูรณ์น้อยกว่าตัวประมาณเดิม ที่ทุกๆระดับของสัมประสิทธิ์สหสัมพันธ์ แม้ว่าตัวอย่างมีขนาดเล็ก ซึ่งมีผลการวิเคราะห์มีแนวโน้มไปในทางเดียวกันกับงานวิจัยที่พิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ที่พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ

Olkin และ Pratt มีความเอนเอียงน้อยกว่าตัวประมาณสัมประสิทธิสหสัมพันธ์เพียร์สันเดิม ในกรณีที่ตัวอย่างมีขนาดเล็ก และข้อมูลมีการแจกแจงปกติ โดย Donald W. Zimmerman [12] นอกจากนี้ Juthaphorn Sinsomboonthong ได้ศึกษาความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt และตัวประมาณเพียร์สันเดิม พบว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt มีความเอนเอียงน้อยกว่าตัวประมาณเดิมทุกๆระดับของสัมประสิทธิ์สหสัมพันธ์ และความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณทั้งสองมีค่าใกล้เคียงกัน[13]

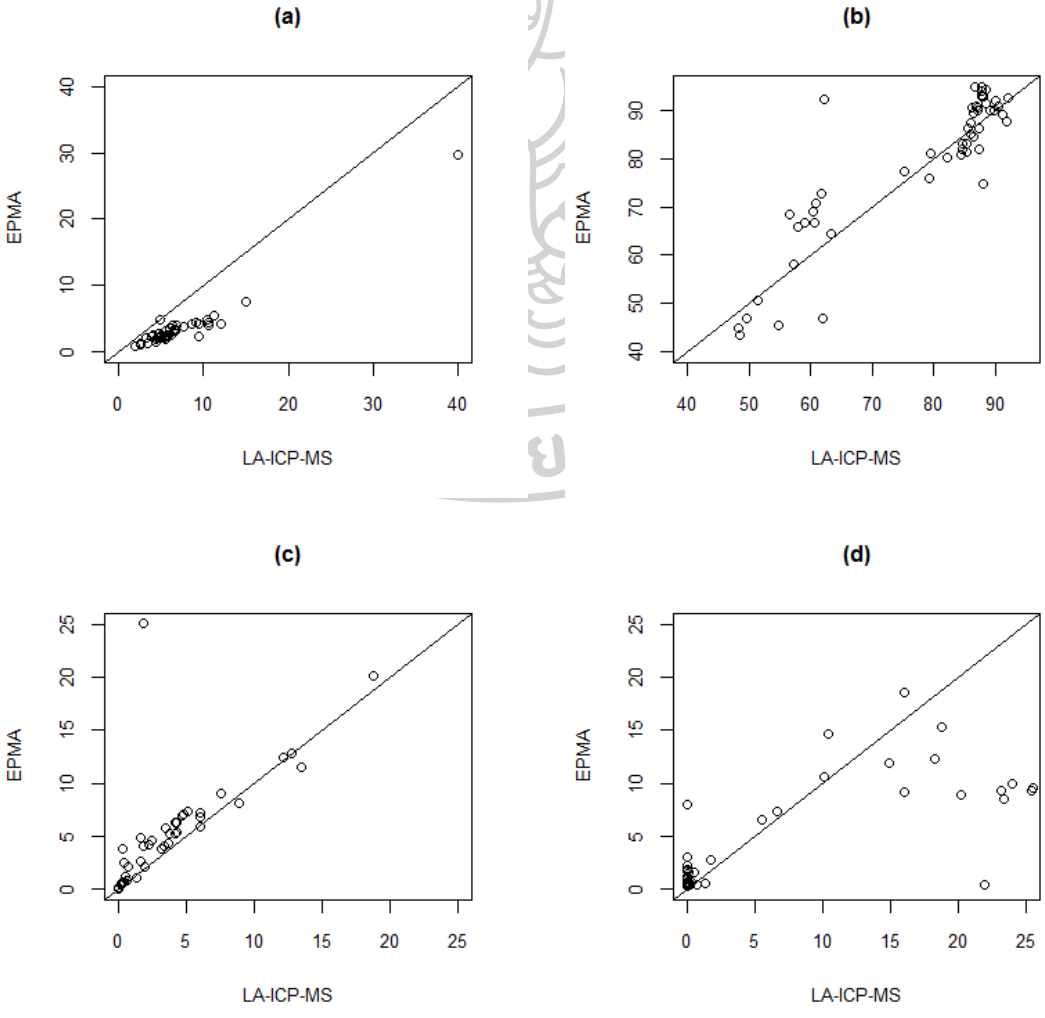
สังเกตว่าความเอนเอียงสัมบูรณ์ของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost จะมีค่าสูงเมื่อสัมประสิทธิ์สหสัมพันธ์เพียรสันมีค่าเข้าใกล้ 0 ทั้งนี้อาจเป็นเพราะว่าตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันเดิมเป็นตัวประมาณที่ไม่เอนเอียงเมื่อสัมประสิทธิ์สหสัมพันธ์เพียร์สันมีค่าเท่ากับ 0 แต่เมื่อพิจารณาตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost พบว่ามีการปรับค่าสัมประสิทธิ์คือ $R+\frac{1-R}{n}$ หรือ $R-\frac{1+R}{n}$ แต่ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt มีการปรับค่าสัมประสิทธิ์คือ $R+\frac{R\left(1-R^2\right)}{2\left(n-3\right)}$ ทำให้ตัวประมาณของ Haddad และ Provost ถูกปรับค่าไปมาก ซึ่งทำให้ค่าประมาณของสัมประสิทธิ์สหสัมพันธ์มีความเอนเอียงเมื่อสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 0 แต่จะส่งผลให้ตัวประมาณของ Haddad และ Provost มีค่าใกล้เคียงกับค่าจริงมากเมื่อระดับของสหสัมพันธ์มีขนาดปานกลาง

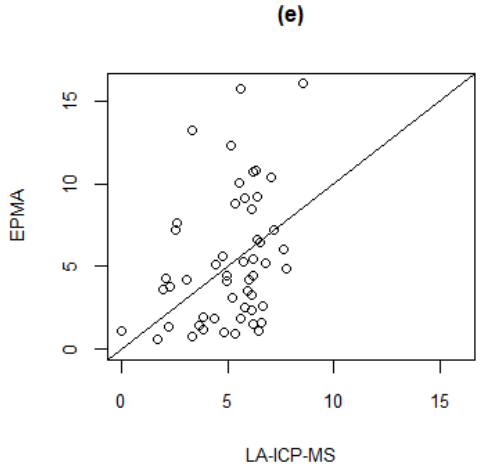
ข้อเสนอแนะ

ในงานวิจัยนี้ ผู้วิจัยได้เสนอตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่โดยใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันทางเลือกอื่นที่มีประสิทธิภาพดีกว่าในเทอมของความเอนเอียง และได้ศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณที่ปรับใหม่และตัวประมาณเดิมในเทอมของความเอนเอียงสัมบูรณ์และความคลาดเคลื่อนกำลังสองเฉลี่ยโดยการจำลองข้อมูล ซึ่งยังมีวิธีการศึกษาประสิทธิภาพของตัวประมาณโดยการประมาณค่าความเอนเอียงในเชิงทฤษฎีเพื่อการเปรียบเทียบตัวประมาณและสามารถบอกได้ว่ากรณีใดควรใช้ตัวประมาณใด นอกจากนี้ยังมีวิธีที่จะปรับปรุงประสิทธิภาพของตัวประมาณได้โดยตรง เช่นการปรับตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องจากการแจกแจงของตัวประมาณ

ตัวอย่างการนำตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องไปใช้ในการวัดความสอดคล้องข้อมูล

การวิเคราะห์ลูกปัดแก้วโบราณจากต่างแหล่งที่อยู่เพื่อหาองค์ประกอบธาตุต่างๆที่อยู่ภายในลูกปัด โดยวิธีการที่จะวัดองค์ประกอบธาตุหลักภายในลูกปัดแก้วนั้นสามารถวัดได้จากเครื่องมือที่เรียกว่า the Electron probe micro analyzer (EPMA) หรือ the laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) ซึ่งการวัดด้วยวิธี LA-ICP-MS มีค่าใช้จ่ายที่สูงกว่า แต่การวัดด้วยวิธี EPMA มีค่าใช้จ่ายต่ำกว่า ในการศึกษาทางเคมีจึงสนใจที่จะทราบว่าเครื่องมือทั้งสองชนิดนี้สามารถใช้ทดแทนกันได้หรือไม่ เพื่อลดค่าใช้จ่ายในการศึกษา ดังนั้นจึงมีการเปรียบเทียบความใกล้เคียงกันของค่าวัดที่ได้จากทั้งสองวิธีโดยใช้สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง วัดความสอดคล้องของร้อยละขององค์ประกอบหลักของลูกปัดแก้วจำนวน 52 ลูก ที่ถูกวัดด้วยวิธี LA-ICP-MS และวิธี EPMA โดยธาตุหลักประกอบด้วย อลูมิเนียม (Al), ซิลิกอน (Si), ทองแดง (Cu), โซเดียม (Na), และแคลเซียม (Ca)





ภาพที่ 23 แผนภาพการกระจายของร้อยละของธาตุหลักในลูกปัดแก้ว ระหว่างการวัดด้วยวิธี LA-ICP-MS และวิธี EPMA ทั้งหมด 52 ลูก เมื่อธาตุหลักคือ (a) อลูมิเนียม, (b) ซิลิกอน, (c) ทองแดง, (d) โซเดียม, (e) แคลเซียม

ตารางที่ 4 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์สหสัมพันธ์เพียร์สันของร้อยละของธาตุหลักในลูกปัดแก้ว ที่วัดด้วยวิธี LA-ICP-MS และวิธี EPMA ทั้งหมด 52 ลูก

ธาตุหลักใน	Means		Standard Deviations		Pearson
ลูกปัดแก้ว	LA-ICP-MS	EPMA	LA-ICP-MS	EPMA	correlations
Al	6.725	3.327	5.407	3.950	0.962
Si	77.520	79.230	14.271	15.138	0.894
Cu	3.077	4.329	4.035	5.065	0.765
Na	5.490	3.864	8.827	4.904	0.770
Ca	5.104	5.297	1.791	3.934	0.335

ตารางที่ 4 แสดงค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์สหสัมพันธ์เพียร์สันของร้อยละของธาตุหลักในลูกปัดแก้ว ที่วัดด้วยวิธี LA-ICP-MS และวิธี EPMA ของลูกปัดแก้วทั้งหมด 52 ลูก สามารถหาค่าประมาณของ C_b ได้ดังนี้

อลูมิเนียม

$$\hat{C}_b = \frac{2(5.407)(3.950)}{5.407^2 + 3.950^2 + (6.725 - 3.327)^2}$$
$$= 0.758$$

ซิลิกอน

$$\hat{C}_b = \frac{2(14.271)(15.138)}{14.271^2 + 15.138^2 + (77.520 - 79.230)^2}$$
$$= 0.992$$

ทองแดง

$$\hat{C}_b = \frac{2(4.035)(5.065)}{4.035^2 + 5.065^2 + (3.077 - 4.329)^2}$$
$$= 0.940$$

โซเดียม

$$\hat{C}_b = \frac{2(8.827)(4.904)}{8.827^2 + 4.904^2 + (5.490 - 3.864)^2}$$
$$= 0.828$$

แคลเซียม

$$\hat{C}_b = \frac{2(1.791)(3.934)}{1.791^2 + 3.934^2 + (5.104 - 5.297)^2}$$
$$= 0.753$$

จากตัวอย่างขนาด 52 ซึ่งมีขนาดใหญ่ และค่าประมาณของ C_b มีค่าไม่เท่ากับ 1 ดังนั้นบนพื้นฐานของความคลาดเคลื่อนกำลังสองเฉลี่ย สามารถใช้ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องได้ทั้งสามรูปแบบ คือตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม $(\hat{\rho}_c)$ ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Olkin และ Pratt $(\hat{\rho}_{cOP})$ และตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณสัมประสิทธิ์สหสัมพันธ์เพียร์สันของ Haddad และ Provost $(\hat{\rho}_{cHP})$ กรณีที่ตัวอย่างไม่ได้มาจากการแจกแจงปกติมาตรฐาน และสหสัมพันธ์มีค่าเป็นบวก ซึ่งสามารถหาค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องได้ดังนี้

อลูมิเนียม

$$\hat{\rho}_c = R\hat{C}_b$$
$$= (0.962)(0.758)$$
$$= 0.729$$

$$\hat{\rho}_{cOP} = R\left(1 + \frac{1 - R^2}{2(n-3)}\right)\hat{C}_b$$
$$= 0.962\left(1 + \frac{1 - 0.962^2}{2(52-3)}\right)(0.758)$$
$$= 0.729$$

$$\hat{\rho}_{cHP} = \left(R + \frac{1-R}{n} \right) \hat{C}_b$$
$$= \left(0.962 + \frac{1-0.962}{52} \right) (0.758)$$
$$= 0.729$$

ซิลิกอน

$$\hat{\rho}_c = R\hat{C}_b$$
$$= (0.894)(0.992)$$
$$= 0.886$$

$$\hat{\rho}_{cOP} = R\left(1 + \frac{1-R^2}{2(n-3)} \right) \hat{C}_b$$
$$= 0.894\left(1 + \frac{1-0.894^2}{2(52-3)} \right) (0.992)$$
$$= 0.888$$

$$\hat{\rho}_{cHP} = \left(R + \frac{1-R}{n} \right) \hat{C}_b$$
$$= \left(0.894 + \frac{1-0.894}{52} \right) (0.992)$$
$$= 0.888$$

ทองแดง

$$\hat{\rho}_c = R\hat{C}_b$$
$$= (0.765)(0.940)$$
$$= 0.719$$

$$\hat{\rho}_{cOP} = R\left(1 + \frac{1-R^2}{2(n-3)} \right) \hat{C}_b$$
$$= 0.765\left(1 + \frac{1-0.765^2}{2(52-3)} \right) (0.940)$$
$$= 0.722$$

$$\hat{\rho}_{cHP} = \left(R + \frac{1-R}{n} \right) \hat{C}_b$$
$$= \left(0.765 + \frac{1-0.765}{52} \right) (0.940)$$
$$= 0.723$$

โซเดียม

$$\hat{\rho}_c = R\hat{C}_b$$
$$= (0.770)(0.828)$$
$$= 0.637$$

$$\hat{\rho}_{cOP} = R\left(1 + \frac{1-R^2}{2(n-3)} \right) \hat{C}_b$$
$$= 0.770\left(1 + \frac{1-0.770^2}{2(52-3)} \right) (0.828)$$
$$= 0.640$$

$$\hat{\rho}_{cHP} = \left(R + \frac{1-R}{n} \right) \hat{C}_b$$
$$= \left(0.770 + \frac{1-0.770}{52} \right) (0.828)$$
$$= 0.641$$

แคลเซียม

$$\hat{\rho}_c = R\hat{C}_b$$
$$= (0.335)(0.753)$$
$$= 0.252$$

$$\hat{\rho}_{cOP} = R\left(1 + \frac{1-R^2}{2(n-3)} \right) \hat{C}_b$$
$$= 0.335\left(1 + \frac{1-0.335^2}{2(52-3)} \right) (0.753)$$
$$= 0.254$$

$$\hat{\rho}_{cHP} = \left(R + \frac{1-R}{n} \right) \hat{C}_b$$
$$= \left(0.335 + \frac{1-0.335}{52} \right) (0.753)$$
$$= 0.262$$

ตารางที่ 5 ค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องระหว่างวิธี LA-ICP-MS และวิธี EPMA จากตัวประมาณทั้ง 3 ชนิด

ธาตุหลักในลูกปัดแก้ว	$\hat{\rho}_c$	$\hat{\rho}_{cOP}$	$\hat{\rho}_{cHP}$
Al	0.729	0.729	0.729
Si	0.886	0.888	0.888
Cu	0.719	0.722	0.723
Na	0.637	0.640	0.641
Ca	0.252	0.254	0.262

จากตารางที่ 5 พบว่าค่าประมาณที่ได้จากตัวประมาณทั้ง 3 ชนิด แตกต่างกันเพียงเล็กน้อยเท่านั้น เนื่องจากบนพื้นฐานของความคลาดเคลื่อนกำลังสองเฉลี่ยพบว่า ข้อมูลนี้เป็นข้อมูลที่อยู่ในกรณีที่สามารถใช้ตัวประมาณได้ทั้ง 3 ชนิดในการหาค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องของข้อมูล

พิจารณาค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องของธาตุอลูมิเนียม ซิลิกอน ทองแดงและโซเดียมพบว่ามีค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องค่อนข้างสูงถึงสูงมาก ดังนั้นเมื่อพิจารณาภายใต้ธาตุทั้ง 4 นี้สามารถสรุปได้ว่า วิธี LA-ICP-MS และวิธี EPMA มีประสิทธิภาพในการวัดปริมาณธาตุทั้ง 4 ได้ใกล้เคียงกัน และสามารถใช้วัดปริมาณธาตุทั้ง 4 แทนกันได้

พิจารณาธาตุแคลเซียมพบว่า ค่าประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องมีค่าต่ำเนื่องมาจากความแตกต่างของส่วนเบี่ยงเบนมาตรฐานของทั้งสองวิธี และสัมประสิทธิสหสัมพันธ์มีค่าต่ำ ดังนั้นสามารถสรุปได้ว่า วิธี LA-ICP-MS และวิธี EPMA มีประสิทธิภาพในการวัดปริมาณธาตุแคลเซียมได้แตกต่างกัน จึงไม่สามารถใช้วัดปริมาณธาตุแคลเซียมแทนกันได้

รายการอ้างอิง

1. Barnhart, H.X., M.J. Haber, and L.I. Lin, *An overview on assessing agreement with continuous measurements.* Journal of Biopharmaceutical Statistics, 2007. **17**: p. 529-569.
2. Lin, L.I., *A concordance correlation coefficient to evaluate reproducibility.* Biometrics, 1989. **45**(1): p. 255-268.
3. Fisher, R.A., *Frequency distribution of the value of the correlation coefficient in samples from an indefinitely large population.* Biometrika, 1915. **10**.
4. Olkin, I. and J.W. Pratt, *Unbiased estimation of certain correlation coefficients.* The Annals of Mathematical Statistics, 1958. **29**.
5. Haddad, J.N. and S.B. Provost, *Approximations to the distribution of the sample correlation coefficient.* International Journal of Mathematical and Computer Sciences, 2011. **7**: p. 1.
6. สำนักงานราชบัณฑิตยสภา, พจนานุกรมศัพท์สถิติศาสตร์ ฉบับราชบัณฑิตยสภา. 2558, สำนักงานราชบัณฑิตยสภา: กรุงเทพฯ, ประเทศไทย. p. 19-184.
7. Huson, L.W., B. Group, and F.H. LaRoche, *Performance of some correlation coeffcients when applied to zero-clustered data.* J. Mod. Appl. Stat. Methods 2007. **6**: p. 530–536.
8. Rodgers, J.L. and W.A. Nicewander, *Thirteen ways to look at the correlation coeffcient.* Am. Stat. , 1988. **42**(1): p. 59–66.
9. Neter, J., et al., *Applied Linear Statistical Models*. Irwin, ed. 4. 1996, Chicago, IL, USA. 1423.
10. ปราณี นิลกรณ์, และคณะ, สถิติพื้นฐาน *1*. 2551: ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร. 384.
11. Gorsuch, R.L. and C.S. Lehmann, *Correlation coefficients: Mean bias and confidence interval distortions.* Journal of Methods and Measurement in the Social Sciences, 2010. **1**(2): p. 52–65.
12. Zimmerman, D.W., B.D. Zumbo, and R.H. Williams, *Bias in Estimation and Hypothesis Testing of Correlation* Psicológica, 2003. **24**: p. 133-158.

13. Sinsomboonthong, J., *Bias correction in estimation of the population correlation coefficient.* Kasetsart J. (Nat. Sci.), 2013. **47**: p. 453-459.

ภาคผนวก

โปรแกรมที่ใช้ในงานวิจัย

1. กรณีข้อมูลมีการแจกแจงปกติมาตรฐาน

```
library(mvtnorm)
library(Matrix)
require(MBESS)
require(matrixcalc)
Nrep <- 10000                          #จำนวนการทำซ้ำ=10,000
MuX <- 0                               #ค่าเฉลี่ยของ X=0
MuY <- 0                               #ค่าเฉลี่ยของ Y=0
SdX <- 1                               #ส่วนเบี่ยงเบนมาตรฐานของ X=1
SdY <- 1                               #ส่วนเบี่ยงเบนมาตรฐานของ Y=1
rholist <- c(seq(-0.9,0.9,0.1))        #ขนาดของสัมประสิทธิ์สหสัมพันธ์ที่สนใจศึกษา
N <- 5                                 #ขนาดตัวอย่างที่สนใจศึกษา n=5, 10, 30, 50
D <- NULL

#Calculate true value of rho c
Real_CCC <- function(rho){
        Cb <- 2/(SdX/SdY+SdY/SdX+(MuX-MuY)^2)
        Rho_c <- rho*Cb
        list(Cb,Rho_c)
}

#Calculate estimated CCC from original and modified estimators
Estimated_CCC <- function(X,Y){

        Cbhat <- 2/(sqrt(var(X)/var(Y))+sqrt(var(Y)/var(X))+((mean(X)-
mean(Y))^2)/sqrt(var(X)*var(Y)))
        Rhohat_c <- cor(X,Y)*Cbhat
```

```
        Rhohat_cOP <- cor(X,Y)*(1+(1-cor(X,Y)^2)/(2*(N-3)))*Cbhat

        if(rho >=0) {Rhohat_HP <- (1-sum((X-Y)^2)/(2*N))
        }else Rhohat_HP <- (sum((X+Y)^2)/(2*N)-1)

        if(Rhohat_HP >= 1) {Rhohat_cHP <- Cbhat
        }else if(Rhohat_HP <= -1) {Rhohat_cHP <- -Cbhat
        }else Rhohat_cHP <- Rhohat_HP*Cbhat

        list(Cbhat,Rhohat_c,Rhohat_cOP,Rhohat_cHP)
}

#Begining of the simulation loop
xyr <- matrix(nrow=N,ncol=2,0)
for(rho in rholist)
{
        mu <- as.vector(cbind(MuX,MuY))
        Sigma <- matrix(c(SdX^2, SdX*SdY*rho, SdX*SdY*rho, SdY^2),2)
        CCC <- Real_CCC(rho)
        Cb <- CCC[[1]]
        Rho_c <- CCC[[2]]
        for(i in 1:Nrep)
        {
                xyr <- rmvnorm(N,mu,Sigma)
                X <- xyr[,1]
                Y <- xyr[,2]
                Est <- Estimated_CCC(X,Y)
                Cbhat <- Est[[1]]
                Rhohat_c <- Est[[2]]
                Rhohat_cOP <- Est[[3]]
                Rhohat_cHP <- Est[[4]]
```

```
                D <-
rbind(D,c(i,rho,Cb,Rho_c,Cbhat,Rhohat_c,Rhohat_cOP,Rhohat_cHP))
        }
}

colnames(D) = c("i","rho","Cb","Rho_c","Cbhat","Rhohat_c","Rhohat_cOP","Rhohat_cHP")
D <- data.frame(D)

#คำนวณค่าความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย
MR <- D[1:10000,]          #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.9
MR <- D[10001:20000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.8
MR <- D[20001:30000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.7
MR <- D[30001:40000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.6
MR <- D[40001:50000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.5
MR <- D[50001:60000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.4
MR <- D[60001:70000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.3
MR <- D[70001:80000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.2
MR <- D[80001:90000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.1
MR <- D[90001:100000,]     #กรณีสัมประสิทธิ์สหสัมพันธ์=0
MR <- D[100001:110000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.1
MR <- D[110001:120000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.2
MR <- D[120001:130000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.3
MR <- D[130001:140000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.4
MR <- D[140001:150000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.5
MR <- D[150001:160000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.6
MR <- D[160001:170000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.7
MR <- D[170001:180000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.8
MR <- D[180001:190000,]    #กรณีสัมประสิทธิ์สหสัมพันธ์=0.9

#ความเอนเอียงของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง
```

```
Bias1 <- mean(MR$Rhohat_c-MR$Rho_c)      #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม
Bias2 <- mean(MR$Rhohat_cOP-MR$Rho_c)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ Olkin และ Pratt
Bias3 <- mean(MR$Rhohat_cHP-MR$Rho_c)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ Hadad และ Provost

#ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง
MSE1 <- mean((MR$Rhohat_c-MR$Rho_c)^2)      #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องเดิม
MSE2 <- mean((MR$Rhohat_cOP-MR$Rho_c)^2)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ Olkin และ Pratt
MSE3 <- mean((MR$Rhohat_cHP-MR$Rho_c)^2)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ Hadad และ Provost
```

2. กรณีข้อมูลไม่ได้มีการแจกแจงปกติมาตรฐาน

```
library(mvtnorm)
library(Matrix)
require(MBESS)
require(matrixcalc)
Nrep <- 10000                            #จำนวนการทำซ้ำ=10,000
MuX <-0                                  #ค่าเฉลี่ยของ X =0
MuY <-                                   #ค่าเฉลี่ยของ Y=0, 0.5, 1.25
SdX <- 1                                 #ส่วนเบี่ยงเบนมาตรฐานของ X=1
SdY <-                                   #ส่วนเบี่ยงเบนมาตรฐานของ Y=1, sqrt(0.5), 3
rholist <- c(seq(-0.9,0.9,0.1))          #ขนาดของสัมประสิทธิ์สหสัมพันธ์ที่สนใจศึกษา
```

```
N <- 5                                  #ขนาดตัวอย่างที่สนใจศึกษา n=5, 10, 30, 50
D <- NULL

#Calculate true value of rho c
Real_CCC <- function(rho){
        Cb <- 2/(SdX/SdY+SdY/SdX+(MuX-MuY)^2)
        Rho_c <- rho*Cb
        list(Cb,Rho_c)
}

#Calculate estimated CCC from original and modified estimators
Estimated_CCC <- function(X,Y){
        Cbhat <- 2/(sqrt(var(X)/var(Y))+sqrt(var(Y)/var(X))+((mean(X)-
mean(Y))^2)/sqrt(var(X)*var(Y)))
        Rhohat_c <- cor(X,Y)*Cbhat
        Rhohat_cOP <- cor(X,Y)*(1+(1-cor(X,Y)^2)/(2*(N-3)))*Cbhat
        if(rho >= 0) {Rhohat_cHP <- (cor(X,Y)+(1-cor(X,Y))/N)*Cbhat
        } else Rhohat_cHP <- (cor(X,Y)-(1+cor(X,Y))/N)*Cbhat
        list(Cbhat,Rhohat_c,Rhohat_cOP,Rhohat_cHP)
}

#Begining of the simulation loop
xyr <- matrix(nrow=N,ncol=2,0)
for(rho in rholist)
{
        mu <- as.vector(cbind(MuX,MuY))
        Sigma <- matrix(c(SdX^2, SdX*SdY*rho, SdX*SdY*rho, SdY^2),2)
        CCC <- Real_CCC(rho)
        Cb <- CCC[[1]]
        Rho_c <- CCC[[2]]
        for(i in 1:Nrep)
```

```
        {
                xyr <- rmvnorm(N,mu,Sigma)
                X <- xyr[,1]
                Y <- xyr[,2]
                Est <- Estimated_CCC(X,Y)
                Cbhat <- Est[[1]]
                Rhohat_c <- Est[[2]]
                Rhohat_cOP <- Est[[3]]
                Rhohat_cHP <- Est[[4]]
                D <-
rbind(D,c(i,rho,Cb,Rho_c,Cbhat,Rhohat_c,Rhohat_cOP,Rhohat_cHP))
        }
}

colnames(D) = c("i","rho","Cb","Rho_c","Cbhat","Rhohat_c","Rhohat_cOP","Rhohat_cHP")
D <- data.frame(D)


#คำนวณค่าความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ย
MR <- D[1:10000,]           #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.9
MR <- D[10001:20000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.8
MR <- D[20001:30000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.7
MR <- D[30001:40000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.6
MR <- D[40001:50000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.5
MR <- D[50001:60000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.4
MR <- D[60001:70000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.3
MR <- D[70001:80000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.2
MR <- D[80001:90000,]       #กรณีสัมประสิทธิ์สหสัมพันธ์=-0.1
MR <- D[90001:100000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0
MR <- D[100001:110000,]     #กรณีสัมประสิทธิ์สหสัมพันธ์=0.1
MR <- D[110001:120000,]     #กรณีสัมประสิทธิ์สหสัมพันธ์=0.2
```

```
MR <- D[120001:130000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.3
MR <- D[130001:140000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.4
MR <- D[140001:150000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.5
MR <- D[150001:160000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.6
MR <- D[160001:170000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.7
MR <- D[170001:180000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.8
MR <- D[180001:190000,]      #กรณีสัมประสิทธิ์สหสัมพันธ์=0.9

#ความเอนเอียงของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง
Bias1 <- mean(MR$Rhohat_c-MR$Rho_c)     #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความ
                                        สอดคล้องเดิม
Bias2 <- mean(MR$Rhohat_cOP-MR$Rho_c)   #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความ
                                        สอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ Olkin
                                        และ Pratt
Bias3 <- mean(MR$Rhohat_cHP-MR$Rho_c)   #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความ
                                        สอดคล้องที่ปรับใหม่ด้วยตัวประมาณของ
                                        Hadad และ Provost

#ความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณสัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง
MSE1 <- mean((MR$Rhohat_c-MR$Rho_c)^2)      #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์
                                            ความสอดคล้องเดิม
MSE2 <- mean((MR$Rhohat_cOP-MR$Rho_c)^2)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์
                                            ความสอดคล้องที่ปรับใหม่ด้วยตัว
                                            ประมาณของ Olkin และ Pratt
MSE3 <- mean((MR$Rhohat_cHP-MR$Rho_c)^2)    #ตัวประมาณสัมประสิทธิ์สหสัมพันธ์
                                            ความสอดคล้องที่ปรับใหม่ด้วยตัว
                                            ประมาณของ Hadad และ Provost
```

ข้อมูลตัวอย่างที่ใช้ในงานวิจัย

ตารางที่ 6 ร้อยละของธาตุหลักในลูกปัดแก้วโบราณที่วัดโดยวิธี LA-ICP-MS จำนวน 52 ลูก

ลูกปัด	Al	Si	Cu	Na	Ca
1	5.8559	86.6548	0.0158	0.7187	6.5842
2	5.6394	87.7910	0.0115	0.2161	6.2077
3	5.3922	88.0697	0.0464	0.1136	6.1589
4	1.9586	90.0141	0.0078	0.2075	7.6091
5	5.7012	87.8225	0.0255	0.0000	6.1507
6	6.2137	86.8576	0.1527	0.0739	6.2129
7	5.5836	87.6958	0.2525	0.2392	5.8333
8	5.4917	79.3811	1.9836	5.5054	7.1661
9	15.0245	75.2755	0.0023	1.7443	6.3956
10	2.6757	86.2908	5.0651	0.0000	5.7444
11	9.4324	63.3809	4.3266	14.8258	6.3703
12	4.8818	87.9064	1.3125	0.0000	5.3275
13	9.4966	62.1087	0.4384	21.9533	5.3296
14	40.0195	56.6222	0.0021	0.0000	3.3545
15	10.5963	51.4699	18.7896	10.4018	6.5282
16	4.3392	86.1567	2.2852	0.0937	6.6558
17	11.1782	61.6841	0.4889	20.2142	5.8213
18	2.6174	92.0537	0.0194	0.0000	4.9585
19	3.3595	87.3761	0.5640	0.0000	7.0452
20	5.6265	88.5153	0.0311	0.0000	5.6102
21	4.7743	85.9105	4.1765	0.0000	4.3478
22	4.4045	86.9964	0.2302	0.0000	7.7544
23	6.7153	87.2505	1.6321	0.0000	3.8272
24	6.7101	87.0102	1.8589	0.0000	3.8323
25	5.3798	87.6683	0.2757	0.0000	6.4260
26	5.5401	90.5202	0.0018	0.0000	3.6775
27	5.1188	88.3453	0.0400	0.0000	5.9946

ลูกปัด	Al	Si	Cu	Na	Ca
28	9.0842	60.4863	0.7280	23.9672	5.1699
29	12.1334	61.9991	1.8767	16.0004	6.1691
30	5.7908	86.2899	0.3753	0.0000	6.8069
31	2.6572	85.8585	4.7567	0.0000	6.2242
32	5.3770	89.6916	1.5780	0.5205	1.6978
33	10.6513	49.5776	13.4877	16.0236	5.5180
34	10.4218	48.2501	12.1356	18.2611	5.6248
35	4.5217	91.8604	0.1932	0.0000	2.2660
36	7.6579	57.9865	4.6816	23.3087	2.5499
37	4.7180	85.1782	7.5591	0.0000	1.9329
38	4.8642	84.3338	6.0042	0.0000	2.1084
39	6.8879	60.5084	3.1153	25.4567	3.3540
40	8.5633	85.5783	2.4509	0.3291	2.2250
41	6.4973	60.9062	3.4174	25.3601	2.6257
42	4.6720	57.2442	8.8271	10.0697	5.9531
43	4.4888	79.1951	0.3446	0.0371	5.2074
44	2.6585	54.8775	6.0530	6.6207	6.1126
45	3.2480	85.2714	5.9805	0.0000	3.0357
46	6.3978	48.5406	12.7748	18.7140	8.5503
47	5.9522	58.9766	4.1307	23.1373	6.2958
48	3.9291	91.1328	3.6632	0.0000	0.0000
49	4.7749	82.0712	4.2984	1.3714	4.9583
50	4.8613	84.5816	3.8145	0.0000	4.4609
51	5.0629	84.6722	3.3881	0.0000	4.7628
52	4.0862	89.1338	0.3377	0.0000	4.8489

ตารางที่ 7 ร้อยละของธาตุหลักในลูกปัดแก้วโบราณที่วัดโดยวิธี EPMA จำนวน 52 ลูก

ลูกปัด	Al	Si	Cu	Na	Ca
1	2.5400	95.0299	0.0356	0.3398	1.5565
2	2.5463	95.0926	0.0388	0.3721	1.4628
3	2.1502	93.0436	0.1301	1.5336	2.3346
4	0.7684	92.1339	0.0463	0.4941	6.0560
5	2.3154	93.0344	0.0800	0.2658	3.2549
6	2.5637	90.8903	0.3793	0.4779	4.4236
7	2.2379	93.2846	0.5142	0.4190	2.4808
8	1.8085	81.2630	2.1336	6.5140	7.1801
9	7.6148	77.4624	0.0982	2.7010	6.5710
10	1.1907	84.6868	7.2958	0.8470	5.2350
11	4.2239	64.3347	6.2384	11.8859	9.1816
12	4.7186	74.7560	1.0913	7.9292	8.7471
13	2.2641	92.4446	2.5264	0.4444	0.9037
14	29.7826	68.5256	0.0254	0.5685	0.7629
15	3.8688	50.6368	20.1523	14.6726	6.4602
16	1.4242	90.5892	4.2325	0.3576	2.5608
17	5.4942	72.7608	1.2088	8.9499	9.1608
18	1.2304	92.7037	0.1035	1.0186	4.0604
19	1.2672	82.1225	0.8197	2.2095	10.3552
20	2.5143	94.3793	0.1490	0.3276	1.8639
21	2.1191	87.5419	6.2246	0.7030	1.8646
22	1.9273	90.5240	0.4780	0.6731	4.8848
23	3.2037	86.3760	4.7929	1.7550	1.9186
24	3.2590	90.1861	4.0575	0.3520	1.1262
25	2.2726	94.1745	0.6315	0.5461	1.1195
26	3.0585	90.9055	0.1636	3.0574	1.4573
27	2.4813	91.4168	0.1362	0.4397	4.1961
28	4.3118	68.9726	2.1040	9.9687	12.3304

ลูกปัด	Al	Si	Cu	Na	Ca
29	4.1755	46.9189	25.0905	9.2124	10.7422
30	2.4171	89.4149	0.6870	0.4690	5.1683
31	1.1427	85.0741	7.0708	0.4508	5.4172
32	2.3846	90.0532	2.5906	1.5145	0.5877
33	4.4009	47.0319	11.5465	18.6286	10.0523
34	4.7299	44.8692	12.4489	12.2803	15.7400
35	2.1638	87.7706	0.4791	1.6286	3.7431
36	3.7389	65.9143	6.8650	8.5414	7.1778
37	2.0193	83.2762	9.0123	1.1690	3.5656
38	2.1331	80.9951	7.2054	0.4994	4.2934
39	3.9255	66.8133	3.8121	9.5399	13.2330
40	4.1333	86.4026	4.5242	0.6048	1.3540
41	3.3318	70.6538	5.7546	9.2808	7.5986
42	2.0656	58.0096	8.1424	10.6259	3.4813
43	2.0558	75.8871	0.5604	1.8662	3.0876
44	0.9498	45.5005	5.8364	7.2605	8.4773
45	2.1833	81.4201	6.8208	0.9915	4.1846
46	4.0205	43.3950	12.8744	15.2406	16.0580
47	3.6121	66.7359	5.2622	9.2869	10.7863
48	2.3882	89.1064	4.3702	0.2539	1.0524
49	2.6229	80.2177	5.3156	0.4687	4.4006
50	2.2822	82.0012	5.2421	0.4764	5.1264
51	2.4208	83.0262	4.0071	0.5745	5.5671
52	2.5676	90.0441	3.7243	0.2542	1.0329

เอกสารการเผยแพร่วิทยานิพนธ์

The 22nd Annual Meeting in Mathematics (AMM 2017)
Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai, Thailand

A Bias Improved Estimator of the Concordance Correlation Coefficient

Parichart Pattarapanitchai[a,†,‡], and Sasiprapa Hiriote[a]

[a]Department of Statistics, Faculty of Science
Silpakorn University, Nakorn Pathom 73000, Thailand

Abstract

Lin's concordance correlation coefficient is one of the most popular indices for measuring agreement between two variables. It can be computed from means, variances and the Pearson correlation coefficient. The alternative estimator of the Pearson correlation coefficient, Olkin and Pratt estimator, is used in the proposed estimator instead of the original sample correlation coefficient. The performances of the new concordance correlation coefficient estimators are empirically studied in terms of bias and mean square error compared to the original estimator by a Monte Carlo simulation. It is found that the absolute bias of the proposed estimator is less than the original estimator even with small sample size. In addition, in terms of mean square error, the proposed estimator performs better than the original one for strong correlation.

Keywords: Agreement, Correlation, Estimator, Monte Carlo simulation.

2010 MSC: Primary 62H20; Secondary 62F12, 62F10, 65C05.

1 Introduction

When a new technique or tool is developed, it is needed to check whether it is usable, i.e., the results of new approach agree with the gold standard one. Also in many researches, a validation of measurement is required. Lin's concordance correlation coefficient [1] is one of the most popular indices for measuring agreement between two continuous variables.

The concordance correlation coefficient evaluates the agreement between two measurements by measuring the variation from the 45° line through the origin. The concordance correlation coefficient can be written in terms of means, variances and the Pearson correlation coefficient of two variables. The Pearson correlation coefficient, ρ is widely used for correlation investigation and also used for evaluate the concordance correlation coefficient.

†Corresponding author.
‡Speaker.
E-mail address: Pattarapanitcha_p@silpakorn.edu (P. Pattarapanitchai), Hiriote_s@silpakorn.edu (S. Hiriote).

The sample correlation coefficient, $\widehat{\rho}$ is well known as the maximum likelihood estimator of ρ. Fisher (1915) [2] proved that the sample correlation coefficients is biased based on random sample from a normal population and provides an asymptotically unbiased estimator of the population correlation. Later, Olkin and Pratt (1958) [3]developed the Pearson correlation estimator which is more nearly unbiased. Zimmerman et al.(2003) [4] and Sinsomboonthong (2013) [5] demonstrated that the biasness is more completely eliminated by an estimator recommended by Olkin and Pratt.

In this study, the modified concordance correlation coefficient estimator using the Olkin and Pratt estimator is proposed. Furthermore, the performances of the proposed estimator are studied in terms of absolute bias and mean square error compared to the original estimator by a simulation study.

2 Research methodology

This study proposes an estimator of the concordance correlation coefficient by applying the alternative estimator of the Pearson correlation coefficient; the Olkin and Pratt estimator. A simulation study is conducted in order to empirically evaluate the validity and reliability of the proposed concordance correlation coefficient estimator along with the sample concordance correlation coefficient estimator. Then, the proposed estimator and original estimator were compared in terms of the absolute bias and mean square error.

2.1 The Concordance Correlation Coefficient

Assume paired sample (X, Y) has a bivariate distribution with means μ_X and μ_Y, variances σ_x^2 and σ_y^2 and the Pearson correlation coefficient ρ. The concordance correlation coefficient, denoted by ρ_c , is given by equation (2.1)

$$\rho_c = \rho C_b \tag{2.1}$$

where $C_b = \frac{2\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2+(\mu_X-\mu_Y)^2}$.

2.2 The Sample Concordance Correlation Coefficient

For n independent pairs of sample $(X_i, Y_i), i = 1, ..., n$, the sample means and variances is $\bar{X}, \bar{Y}, S_x^2$ and S_y^2. The original concordance correlation coefficient estimator, denoted by $\widehat{\rho_c}$, is given by equation (2.2)

$$\widehat{\rho_c} = \widehat{\rho}\widehat{C_b} \tag{2.2}$$

where $\widehat{\rho}$ is the sample correlation coefficient and $\widehat{C_b} = \dfrac{2S_xS_y}{S_x^2 + S_y^2 + (\bar{X} - \bar{Y})^2}$.

2.3 The Proposed Estimator of the Concordance Correlation Coefficient

The proposed estimator of ρ_c applies the nearly unbiased estimator of ρ instead of the maximum likelihood estimator of ρ. The bias corrected estimator of ρ recommended by Olkin and Pratt (1958), denoted by $\widehat{\rho}_{OP}$, is given by

$$\widehat{\rho}_{OP} = \widehat{\rho}\left(1 + \frac{1 - \widehat{\rho}^2}{2(n-3)}\right)$$

The proposed estimator of ρ_c in this paper, denoted by $\widehat{\rho}_{c_{OP}}$, is given by equation (2.3)

$$\widehat{\rho}_{c_{OP}} = \widehat{\rho}_{OP}\widehat{C_b} \tag{2.3}$$

3 Simulation Results

The simulation study was conducted in order to empirically evaluate the validity and reliability of the proposed estimator. The n pairs of x and y were generated from bivariate normal distribution with five cases of the location and scale differences as follows; **Case 1:** No location and scale difference: $\mu_X = \mu_Y = 0, \sigma_x^2 = \sigma_y^2 = 1$, **Case 2:** No location difference and slightly scale difference: $\mu_X = \mu_Y = 0, \sigma_x^2 = 1, \sigma_y^2 = 0.5$, **Case 3:** No location difference and moderate scale difference: $\mu_X = \mu_Y = 0, \sigma_x^2 = 1, \sigma_y^2 = 9$, **Case 4:** Slightly location and scale difference: $\mu_X = 0, \mu_Y = 0.5, \sigma_x^2 = 1, \sigma_y^2 = 0.5$, **Case 5:** Moderate location and scale difference: $\mu_X = 0, \mu_Y = 1.25, \sigma_x^2 = 1, \sigma_y^2 = 9$. In addition, the correlation coefficient (ρ) of X and Y were considered at -0.9 to 0.9 increased by 0.1. In each case, sample sizes of $n = 5, n = 10$ and $n = 30$ were considered. For each of these situations, the absolute bias and mean square error (MSE) of the estimators were computed based on 10,000 runs using R program.

The simulation results are presented in Figure 1 and Figure 2. Figure 1 shows the absolute biases of the original estimator ($\widehat{\rho}_c$) and the proposed estimator ($\widehat{\rho}_{cOP}$). Overall, the trends of absolute biases for both estimators are the same for almost situations. The absolute biases of the proposed estimator are significantly lower than that of the original estimator when the size of the population correlation coefficient are moderate to large except for the case of moderate location and scale difference. However for nearly and no correlation of all situation, the absolute biases of both estimators are almost the same and close to 0. Moreover, the absolute biases of the proposed and original estimator tend to decrease when sample size increased, regardless of the size of the population correlation coefficient. For both estimators, the case of the larger scale difference gives the smaller the absolute bias for all sample sizes but the case of the larger location difference gives the smaller the absolute bias only for small sample sizes ($n = 5, n = 10$), except for $\rho = 0$.

Figure 2 shows the mean square errors of the original estimator ($\widehat{\rho}_c$) and the proposed estimator ($\widehat{\rho}_{cOP}$). For small sample sizes, the mean square errors of the proposed estimator is less than that of the original estimator when the level of the population correlation coefficient is large. The mean square errors of the both estimators are not different and close to 0 for all 5 cases when sample size is large ($n = 30$). The mean square errors of the proposed and original estimator decrease as the sample size increases, for all sizes of the population correlation coefficient. For each sample size, the larger the scale or the location difference, the smaller the mean square error, except for the very high correlations.

4 Conclusion

This paper proposed the estimator of the concordance correlation coefficient for measuring agreement between two variables which have a bivariate normal distribution. The proposed estimator ($\widehat{\rho}_{cOP}$) was modified from the nearly unbiased estimator of the population correlation coefficient recommended by Olkin and Pratt. The simulation results showed that the original sample concordance correlation coefficient ($\widehat{\rho}_c$) is a biased estimator. The biases of the proposed estimator are smaller than those of the original estimator regardless of the level of the population correlation coefficients and sample sizes. However, there are some cases they perform closely and the biases are nearly 0. In addition, in terms of mean square error, the proposed estimator performs bettter than the original estimator for small sample size and the large size of the population correlation coefficient.

Acknowledgment. The authors are thankful to the referees for providing the useful comments of the manuscript.

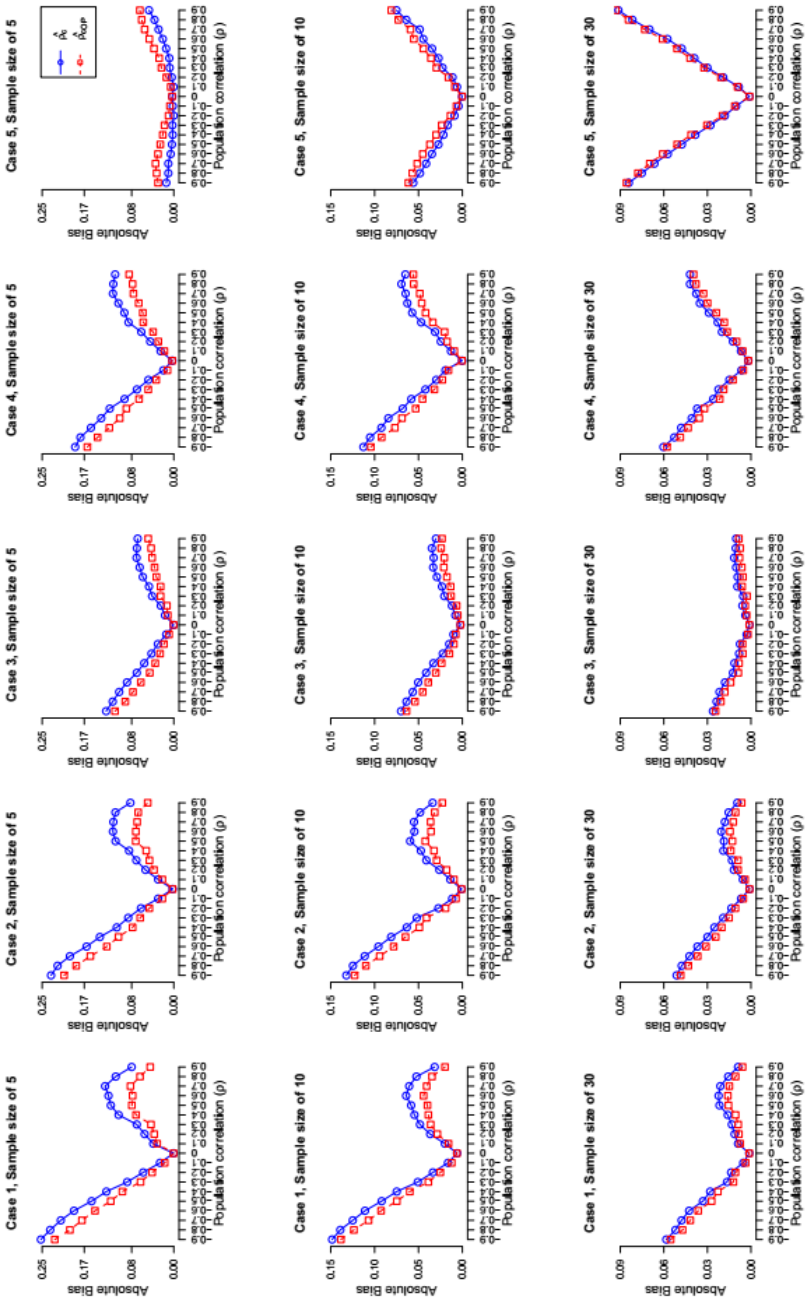


Figure 1: Absolute biases of $\widehat{\rho}_c$ and $\widehat{\rho}_{cOP}$ for 5 different cases and 3 different sample sizes.

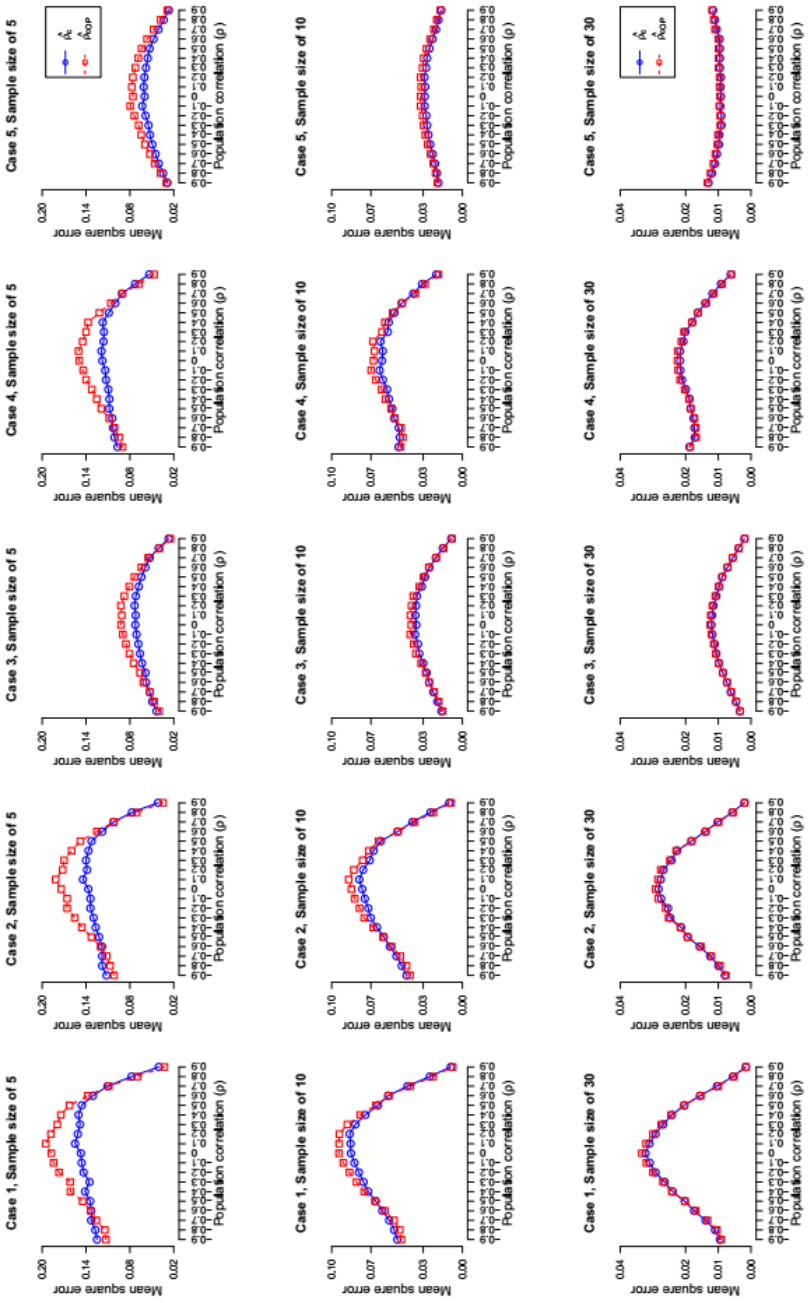


Figure 2: Mean square errors of $\widehat{\rho}_c$ and $\widehat{\rho}_{cOP}$ for 5 different cases and 3 different sample sizes.

References

[1] L.I. Lin, *A concordance correlation coefficient to evaluate reproducibility*, Biometrics **45** (1989), 255–268.

[2] R.A. Fisher, *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population*, Biometrika **10** (1915), 507–521.

[3] I. Olkin and J.W. Pratt, *Unbiased estimation of certain correlation coefficients*, Annals of Mathematical Statistics. **29** (1958), 201–211.

[4] D.W. Zimmerman, B.D. Zumbo and R.H. Williams, *Bias in estimation and hypothesis testing of correlation*, Psicolgica. **24** (2003), 133–158.

[5] J Sinsomboonthong, *Bias correction in estimation of the population correlation coefficient*, Kasetsart J. (Nat. Sci.) **47(3)** (2013), 453–459.

ประวัติผู้เขียน

ชื่อ-สกุล	ปาริชาต ภัทรพานิชชัย
วัน เดือน ปี เกิด	1 พฤษภาคม 2536
สถานที่เกิด	สุพรรณบุรี ประเทศไทย
วุฒิการศึกษา	เกียรตินิยมอันดับหนึ่ง ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยศิลปากร
ที่อยู่ปัจจุบัน	บ้านเลขที่ 63 หมู่ 5 ต.หนองสาหร่าย อ.ดอนเจดีย์ จ.สุพรรณบุรี 72170
ผลงานตีพิมพ์	Hiriote, S. and Pattarapanitchai, P. (2016), Concordance correlation coefficient for assessing agreement between two multivariate measurements, ICAS2016, Phuket, Thailand. Pattarapantichai, P. and Hiriote, S. (2017), A bias improved estimator of the concordance correlation coefficient, The 22nd Annual Meeting in Mathematics (AMM 2017), Chiang Mai, Thailand

มหาวิทยาลัยศิลปากร